Septor: Seismic Depth Estimation using Hierarchical Neural Networks

M Ashraf Siddiquee
The University of New Mexico
Albuquerque, USA
siddiquee@unm.edu

Vinicius M. A. Souza
Pontifícia Universidade Católica do Paraná
Curitiba, Brazil
vinicius@ppgia.pucpr.br

Glenn Eli Baker
Air Force Research Lab
Albuquerque, USA
glenn.baker.3@us.af.mil

Abdullah Mueen
The University of New Mexico
Albuquerque, USA
mueen@unm.edu

ABSTRACT

The depth of a seismic event is an essential feature to discriminate natural earthquakes from events induced or created by humans. However, estimating the depth of a seismic event with a sparse set of seismic stations is a daunting task, and there is no globally usable method. This paper focuses on developing a machine learning model to accurately estimate the depth of arbitrary seismic events directly from seismograms. Our proposed deep learning architecture is not-so-deep compared to commonly found models in the literature for related tasks, consisting of two loosely connected levels of neural networks, associated with the seismic stations at the higher level and the individual channels of a station at the lower level. Thus, the model has significant advantages, including a reduced number of parameters for tuning and better interpretability to geophysicists. We evaluate our solution on seismic data collected from the SCEDC (Southern California Earthquake Data Center) catalog for regional events in California. The model can learn waveform features specific to a set of stations, while it struggles to generalize to completely novel sets of event sources and stations. In a simplified setting of separating shallow events from deep ones, the model achieved an 86.5% F1-score using the Southern California stations.

CCS CONCEPTS

• **Applied computing** → **Earth and atmospheric sciences**; • **Information systems** → *Data mining*; • **Computing methodologies** → *Neural networks.*

KEYWORDS

Machine Learning, Seismology, Depth Prediction, Regression, Classification

ACM Reference Format:
M Ashraf Siddiquee, Vinicius M. A. Souza, Glenn Eli Baker, and Abdullah Mueen. 2022. Septor: Seismic Depth Estimation using Hierarchical Neural Networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22), August 14–18, 2022, Washington, DC, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3534678.3539166

KDD '22, August 14–18, 2022, Washington, DC, USA

ACM ISBN 978-1-4503-9385-0/22/08…$15.00
https://doi.org/10.1145/3534678.3539166

1 INTRODUCTION

Accurate depth estimation of seismic events is a critical procedure to discriminate between man-made and natural events. While anthropogenic seismic sources are overwhelmingly less than 1 km depth, nearly all earthquakes nucleate below a 2 or 3 km depth. Most earthquakes considered shallow occur between a few and 70 kilometers depth, while deep earthquakes can occur as deep as 700 kilometers [24]. Distinguishing man-made events from natural events has several applications, for example, in nuclear non-proliferation [6], and seismic hazard monitoring.

Theoretically, the depth of a seismic event is estimated by inverting the travel time equations to individual observing stations. However, the correctness of the estimation largely depends on the locations and the number of observing stations. When stations are far from an event (e.g., more than 100 km), and the number of stations observing the event is small (e.g., three or less), the uncertainty in the estimated depth grows beyond tolerance. In contrast, the depth is most accurately estimated when a seismic station is located exactly above the event's origin. Unfortunately, no single seismic network can guarantee global coverage, leading to inaccurate depth estimation for novel seismic hot spots.

This research considers estimating seismic depth directly from the waveforms generated by the events (i.e., time series, or seismograms) employing modern machine learning (ML) techniques. Such an approach to evaluating ML methods and potentially replacing physics-based estimation methods is gaining significant interest among geophysicists [6]. In this paper, we develop the first hierarchical neural network model named **Septor** (Seismic depth estimator) aimed to estimate the depth of seismic events from waveforms of multiple channels at multiple stations. Besides its novelty in automated depth estimation, our model can potentially be a support tool to distinguish seismic events (i.e., man-made vs. natural), and has several applications in automated seismic monitoring. We train Septor using a set of 8,359 highly calibrated (by human analysts) events from the Southern California Earthquake Data Center (SCEDC) spanning over forty years of monitoring data. We have achieved an impressive root-mean-squared-error (RMSE) of 2.89 km in predicting the depth (with 70.1% correlation to actual depth) of these events using only a few close-by stations from the same network. We have also considered a binary classification problem to

distinguish shallow from deep events. Our model achieves a 86.5% F1-score in shallow-deep discrimination, promising a step closer to fully automatic seismic monitoring. Even though the model can learn waveform features from a specific network of station, we neither expect nor claim the model learning the underlying travel-time inversion process. In an attempt to evaluate the generalizability, we perform three experiments on disjoint train-test sets with no common source and station. The model shows a gradual decrease in performance as the distances between source-station pairs increase.

Septor architecture is not very deep compared to modern ML models used in computer vision or speech processing. The reason is the lack of labeled data to train a deeper network. The calibrated events used to train Septor were labeled by well-trained analysts working on daily shifts for years. Hence, we tailored the model to fit our data instead of fitting a model to tailored data. Our network architecture has two hierarchies, each consisting of Convolutional Neural Network (CNN) layers followed by Long-Short Term Memory (LSTM) layers. Such simple architecture provides a great deal of efficiency and interpretability for geophysicists. We demonstrate that our model conforms to the general scientific understanding of depth estimation and can be employed in at least two different seismic regions.

Why is depth estimation from waveforms challenging? *(i)* Physics-based depth estimation suffers from uncertainty due to noise in the signal and the lack of nearby stations. For example, in the SCEDC original catalog [1, 19], the mean quantified uncertainty is 6.89 km. Such uncertain labels in training data hardly lead to accurate models, making seismologists skeptical about ML-based systems for depth estimation. However, we consider the highly calibrated SCEDC catalog [11] for training our model with a very low uncertainty of 0.357 km (due to high station density in the network). Thus, the uncertainty in training labels cannot accumulate into the validation error. *(ii)* Physics-based estimated depth for the same earthquake event can be different in different earthquake catalogs. For example, the root-mean-squared difference of depths between the SCEDC original catalog and the SCEDC highly calibrated catalog is **3.81** km for the same set of earthquakes. Moreover, different algorithms may calculate the depth from different reference points. For example, in Northern California, NCSN uses depth relative to the geoid (essentially sea level)[4], whereas the double-difference catalog uses depth from the surface[28]. Therefore, ambiguity among data sources makes it hard to evaluate learned models on new catalogs. We contribute experiments to test our model in multiple geographic regions to demonstrate equivalent performance; *(iii)* Only experts in geophysics and seismology can produce confident labeled information. Unfortunately, it is very difficult (or even impossible) for a non-expert to spot any pattern in raw waveforms or other visual data representation. To better illustrate this difficulty, consider the Continuous Wavelet Transform (CWT) representation of a deep earthquake captured by different stations at varying distances illustrated in Figure 1 (a) and a shallow earthquake shown in Figure 1 (b). Therefore, crowd-sourced annotations are not feasible for this application, ruling out the option to train deep models on a large-scale labeled dataset. For this reason, we propose a model based on a small two-hierarchy network, just enough to learn from the set of available well-calibrated events.

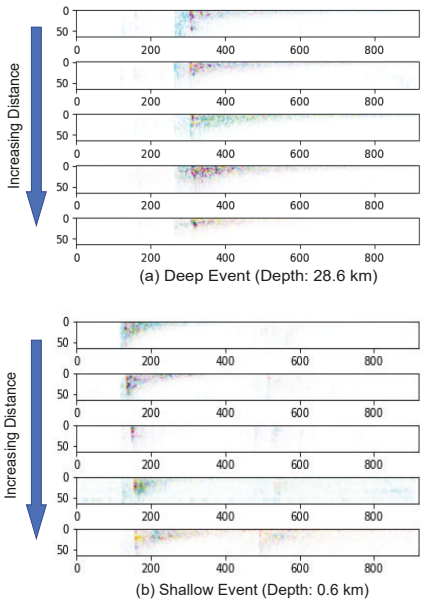

Figure (1) CWT images from the 5 closest stations for a deep (top) and a shallow (bottom) earthquake. For the same earthquake, CWT images are ordered ascending according to the *epicenter to station distance* from top to bottom.

The remainder of this paper is organized as follows. Section 2 introduces essential concepts of seismology related to the task of depth prediction. Section 3 discusses related work. We describe Septor in Section 4. Data description and preprocessing are discussed in Section 5. The experimental setup and our evaluations considering regression and classification settings are presented in Section 6 and Section 7. The use of Septor at a different region is discussed in Section 8. Finally, our conclusions are presented in Section 9.

2 BACKGROUND

In this section, we discuss the main steps of a typical seismic data processing pipeline illustrated in Figure 2. This pipeline refers to the process of transforming a set if seismic signals into a bulletin of seismic events. (i.e earthquakes, explosions, etc). The data processing starts when a network of stations detects a seismic signal that could be from a man-made or a natural event.

Signal Detection: This step consists of monitoring a continuous waveform obtained by seismometers to detect events. Typically, a seismometer captures ground motion in multiple directions (up-down, north-south, and east-west) to obtain a three-dimensional ensemble. Any seismic event must excite at least one dimension or channel (e.g., BHZ, BHN, and BHE) of a seismometer. When all channels simultaneously show signals, one can be confident about the arrival of a seismic wave at the station. Once an event is detected, the remaining steps are executed in sequence.

Phase Identification: Any movement of earth materials generates multiple types of waves and each wave is affected differently by the unique velocity structure along it's own propagation path. Common phase types are compressional (P) and shear (S) waves, which travel through the earth; and surface waves, which travel

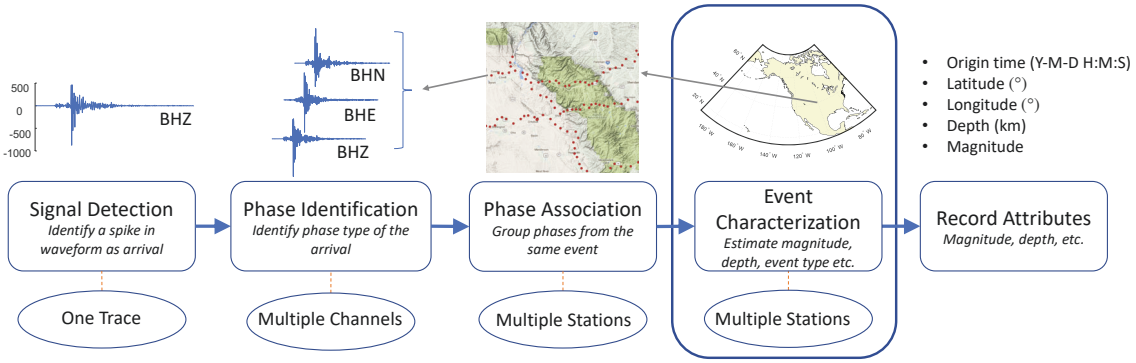

Figure (2) Seismic data processing pipeline. We perform this task of depth prediction in the "Event Characterization" step without the human supervision.

along the earth's surface. A phase picker detects different types of signals and labels them as P, S, or other types of waves, as illustrated in Figure 3. For example, when an earthquake or underground explosion occurs, seismic waves propagate away from the source. The scattered energy immediately following the P and S waves is called coda. Characteristics of each part of the waveform carry information about the source, including about its depth. Accurate phase picking is essential to obtain parameters that can constrain location, depth, and event type.

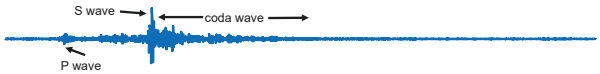

Figure (3) A 230 second long seismic waveform shows the P wave arrives first for an earthquake, followed by the S wave. From 30 seconds before the P arrival time to 200 seconds after is sufficient to capture the entire earthquake.

Phase Association: Association of phases across multiple stations is done by a phase associator. Phase associators aggregate multiple waveforms generated from the same earthquake and triangulate to compute the epicenter[1] and depth of the earthquake. Phase associators work on multiple stations' data and may generate other meta information such as an initial location of the earthquake.

Event Characterization: Typically, the depth estimate of an event may be refined in this stage of the pipeline. Depth can be estimated by minimizing the propagation time residuals relative to some prediction across all the observing stations. However, considering the complexity of the earth's structure, depth estimation involves many geophysical priors, including local, regional, and global velocity models. Misassociated phases are common in automated associators and at the very least can bias depth estimates (and even lead to false events being formed), so human supervision of output is necessary. In this work, we propose to perform the depth estimation after the initial event formation to refine the initial depth estimate. The goal is to estimate and exploit the depth of an event automatically without human supervision. After the event characterization step, an event, with its location, depth, magnitude, etc. are listed in the earthquake catalog.

[1]The epicenter is the location of an earthquake on the earth surface. The epicenter does not provide information about depth.

3 RELATED WORK

State-of-the-art seismic event depth prediction techniques require considerable manual oversight in data annotation and expertise in data analysis. Depths are determined using 3D earth models and seismic wave travel-times. They can be refined by consideration of travel times relative to other seismic events, providing extremely accurate relative locations, which can in turn improve absolute locations. The range of uncertainty associated with the estimated depth though often makes it hard to consider them as "ground truth". For these reasons, the number of examples used to train and test predicting models found in the literature has been limited to less than one thousand. Such a limited number of examples make the use of modern machine learning models, such as deep learning, challenging. Besides, those works consider the information from a single station for prediction, as described further below.

Seismologists have used physics based approaches to identify waveform features that indicate source depth, and have applied metrics based on those features to estimate depth. Kafka [14] used the existence of higher frequency fundamental mode Rayleigh waves, or Rg, as an indication of shallow source depth, typically less than 3 or 4 km. Rg detectors are useful tools for identifying shallow seismic sources in some areas, but especially in tectonically active areas such as southern California, Rg may only propagate 10 km before it is attenuated to below background levels of the S wave coda, which arrives at the same time. The ratio of the S wave peak amplitude to the duration of the S wave coda also is affected by and can be used to roughly estimate the source depth, with longer coda durations occurring for shallow sources [15]. This may be due to trapping of high frequency shear waves in shallow low velocity waveguides, where they propagate more slowly, so extend the wavetrain. Alternately, scattering of Rg at or near the surface can also extend the wavetrain. Both mechanisms may be important, but the physical basis of this observation is still not fully settled. The ratio of P to S wave energy is often also indicative of depth, as S waves are more likely to be trapped in highly attenuative near surface waveguides than are P waves, thus increasing the ratio of P to S wave energy for shallow events relative to deeper ones [10]. While the accuracy and broad applicability of these methods are limited, we are encouraged that there are such features in seismic waveforms that ML methods can learn, and optimize the use of to improve estimations. Further, we expect that other features may

exist that we haven't discovered using physics-based approaches, but that algorithms can learn with sufficient training data.

Ochoa et al. [18] proposed to use Support Vector Regression (SVR) for depth estimation of events collected by a single station at Bogota, Colombia. The original catalog has 2,164 events observed between 1998 and 2008. However, events with a lower magnitude and possibly related to man-made events were discarded in the experimental evaluation, leaving only 863 events. A set of 25 features related to the magnitude, epicenter distance, and source location were extracted from data for training the SVR model.

Recently, Yang et al. [29] compared the performance of conventional feature-based classification models (e.g., Support Vector Machine, Random Forest, Naive Bayes, and k-Nearest Neighbors) with a 10 layer convolutional neural network using CWT representation for distinguishing between deep and shallow seismic events. The comparisons were performed using 444 micro-earthquake events associated with an underground collapse of a cavern in South Louisiana. The signals were sampled at 200 Hz by 8 broad-band three-component monitoring devices, and the hypocenter depths ranged from 1 km to 2 km for deeper events and depths between 40 m and 400 m for shallow events.

From a machine learning perspective, classification of deep and shallow earthquakes from waveforms can be described as a *Time Series Classification* problem (TSC) [2, 3, 7], in which the goal is to predict a discrete label for a series from a finite set of categories. In the past decade, TSC problems have been of great interest among data mining researchers, and have been applied in many different domains. In seismology, TSC has been used in the context of phase detection and identification [6, 17, 20]. However, there is a lack of solutions for depth prediction as a regression of seismic events using waveforms from multiple stations, as proposed in this paper.

Depth estimation from seismic waveform falls under the category of *Time Series Extrinsic Regression* (TSER) [25] where a single scalar continuous value is predicted based on the whole time series. Unlike *Time Series Forecasting* [13], where the prediction *mostly* depends on recent values, TSER considers the whole time series *equally* for prediction. Although TSER has been used in a wide range of domains [8, 25], it has not been employed in seismology. In this paper, we compare our solution with Rocket [8], a state-of-the-art TSER method.

4 SEPTOR: HIERARCHICAL NETWORK

Convolutional Neural Networks (CNN) and Long-Short Term Memory (LSTM) networks are the building blocks of our model Septor. A CNN is a class of deep neural networks widely employed on image mining problems. CNNs perform convolutions on images with multiple fixed-sized kernels. A convolution operation can be seen as sliding the kernel over the image and computing the dot product at each step. Each convolution extracts different higher-order representations from the feature map. A convolutional layer is usually followed by a nonlinear activation function (such as ReLU) and a max-pooling function.

Long Short-Term Memory (LSTM) networks are an improved variant of traditional Recurrent Neural Networks (RNNs) [12]. RNNs can model temporal dependencies in the data by feedback connections considering both the input at the current time step as well as the output of the last time step's hidden state. However, vanilla RNNs suffer from the vanishing gradient problem, which prevents the model from learning long range dependencies. LSTM tackles this problem by introducing three gating mechanisms to update the memory cell c_t and hidden state h_t at each step t based on the current time step input x_t and the previous time step's hidden state output h_{t-1}. The input gate i_t, forget gate f_t, output gate o_t, memory cell c_t and hidden state h_t at step t are computed as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (1)$$
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f \quad (2)$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o \quad (3)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$
$$h_t = o_t \odot tanh(c_t) \quad (5)$$

Here, σ is the logistic sigmoid function, tanh is the hyperbolic tangent function, and $\odot$ denotes the element wise multiplication. Each LSTM unit is composed of a memory cell and three main gates: input, output and forget. By this architecture, the LSTM manages to create a controlled information flow by deciding which information it must forget and which information to remember. To understand the mechanism behind the architecture, we can view f_t as the function that controls to what extent the information from the old memory cell is going to be thrown away, i_t controls how much new information is going to be stored in the current memory cell, and o_t controls what to output based on the memory cell c_t.

4.1 Architecture of Septor

Septor consists of two loosely connected hierarchies, as illustrated in Figure 4: (1) the waveform aggregator, and (2) the station aggregator. The waveform aggregator (Figure 5) is a CNN-LSTM based network and receives a 3 dimensional linearly spaced Continuous Wavelet Transform (CWT) image for a single station and outputs a 2D feature array. Multiple output from the waveform aggregator is fed into the station aggregator (Figure 6). The station aggregator receives the features from the waveform aggregator in a distance (distance from the epicenter of the earthquake to the station) preserving ordering. The station aggregator is another CNN-LSTM based network which finally predicts the depth of the earthquake.

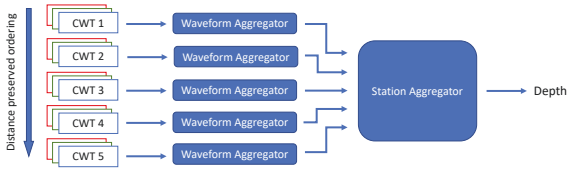

Figure (4) General view of Septor architecture.

Our *waveform aggregator* consists of four CNN layers followed by two LSTM layers followed by three fully connected (FC) layers. Each of the CNN layers is followed by a batch normalization layer and a ReLU activation layer. 30% dropout is applied after ReLU activation layers to prevent overfitting the model. We used 8, 16, 32 and 64 size kernels of size 3×3 for the convolution layers. The output from the fourth CNN layer is fed to the two stacked LSTM layers after a time preserving transformation. Extracted features from the LSTM layers are then passed through three FCN layers. After each FCN layer, batch normalization and ReLU activation is applied. Figure 5 contains the major parts of the waveform aggregator.

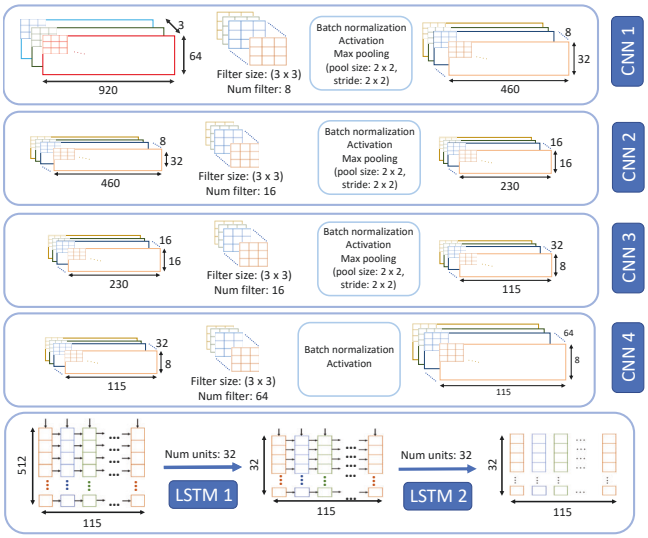

Figure (5) ***Waveform aggregator*** **consists of four CNN and two LSTM layers.**

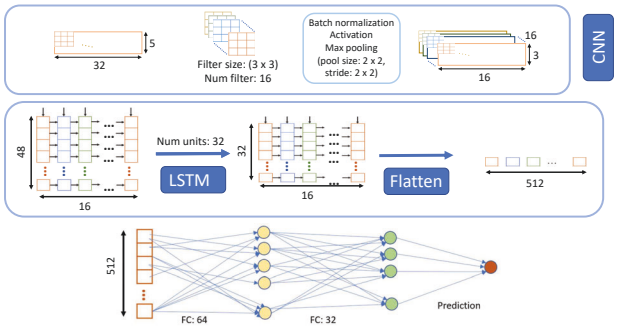

Figure (6) ***Station aggregator*** **consists of a CNN and an LSTM layers. Three fully connected layers are used to get the final prediction.**

The *station aggregator* consists of one CNN layer, one LSTM layer, and a stack of three FCN layers. The CNN layer in the *station aggregator* has 16 kernels of 3×3 size and is followed by a ReLU activation and a max pooling layer. Output from the CNN layer is fed into the LSTM layer which has 32 hidden units. The *station aggregator* hierarchy ends with FCN layers with 64, 32 and 1 output units. Figure 6 contains the details of the *station aggregator*. For both the waveform and station aggregator, we define the number of layers according to preliminary results on training sets.

5 DATA DESCRIPTION

We carry out our experimental evaluation with a highly calibrated earthquake catalog from Southern California [11]. This catalog used either single station locations with a 3D velocity model [16] or a multiple event location method, GrowClust [26] for depth calculation, both of which are more accurate than the baseline Southern California Earthquake Data Center (SCEDC) depth calculation. The catalog includes uncertainty bounds for the depths. The reported median vertical uncertainty is 0.4 km, which is considered very good for many seismic monitoring tasks. Therefore, this dataset is a close-to-ideal candidate for our experimental evaluation. We collected more than 650,000 multi-channel waveforms that are associated with earthquakes having magnitude 2.0 to 4.0, recorded by 423 densely located seismic stations in Southern California region.

We filter out the earthquake events for which we did not find any station with a distance less than 1.2 times the reported depth. Such close proximity of an observing station ensures greater accuracy of the depth estimate. After selecting events measured by at least five stations and containing waveforms from all three broadband channels (BHZ, BHN, BHE), we build a dataset with 8,359 earthquake events. In Figure 7, we show the SCEDC station map and depth distribution of our collected earthquakes.

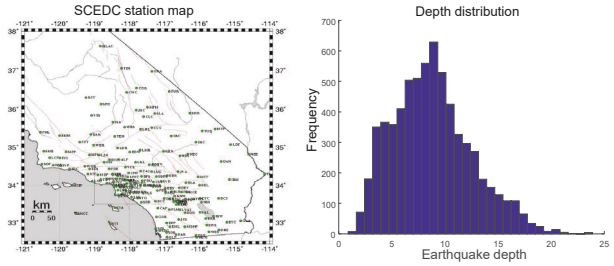

Figure (7) (left) Distribution of Broadband stations in Southern California shows a dense seismic network [21]. (right) Distribution of earthquake depth (in km.) of our dataset.

For each of the 3-channels, we collected 230 seconds of waveforms, starting 30 seconds prior to the first P arrival time and ending at 200 seconds after the P arrival time. This time window is large enough to capture seismic waves generated by any regional earthquake. Since the waveforms are sampled at 40Hz, the length of each waveform is 9,200 data points. In summary, our dataset consists of:

- A total of 8,359 earthquake events;
- Each event is recorded at 40Hz by five observing stations;
- Each station records three broadband waveforms;
- Each waveform contains 9,200 numeric observations.

5.1 Waveform Preprocessing

Following conventional seismic signal preprocessing techniques, we remove the instrument responses associated with the station from the waveforms. Then we convert the vertical, and the two horizontal (usually north-south and east-west) components (Z, N, E) to horizontal, radial, and tangential components (Z, R, T). We pass the waveforms through a 0.4Hz to 10Hz bandpass filter, detrend each sample, and remove the mean. We max-normalize the data across each channel, thus retaining the relative amplitudes among components of a station. Afterward, we use a 64-scale CWT to obtain a spectral-temporal representation, as previously illustrated in Figure 1. The final RGB linearly spaced CWT image has 64×920×3 dimensions, where the vertical, radial and tangential components are represented by red, green, and blue colors, respectively.

6 EXPERIMENTAL SETUP

In this work, depth prediction is a regression task from three dimensional spectral-temporal images. Several existing methods can be used to produce depth estimates. We compare the results of Septor against two classic ML-based models: CNN, LSTM; and two state-of-the-art algorithms for time series regression: Rocket [8] and XGBoost [5]. The algorithms are briefly described below.

CNN based models have been widely used in image classification, time series classification, and seismic data classification [9]. We use linearly-spaced CWT images of waveforms as input to a simple four layer CNN network to train towards actual depths.

LSTM based models can capture long-term temporal dependencies, hence, it is very effective for time series classification and regression [12]. We flatten the CWT images to feed into a two layer LSTM network and train towards actual depth.

XGBoost (Extreme Gradient Boosting) is a decision-tree based ensemble algorithm that uses boosting technique to sequentially add new trees to the classification or regression model [5].

Rocket (Random Convolutional KErnel Transform) is originally a classifier for time series that transforms the data using a large number of random convolutional kernels that can capture basic patterns or shapes from the series. Recently, Rocket was adapted for extrinsic regression and achieves the highest overall accuracy in a comprehensive experimental evaluation [25].

Rocket and XGBoost were trained using the vertical component (Z-axis) of the waveforms in case of single-channel resolution; by concatenating the three components (Z, R, T) waveforms into a single vector in the case of multi-channel resolution; and by concatenating the three-channel waveforms from five stations for multi-station resolution. For Rocket, the number of kernels was set to 10,000.

We consider Root-Mean-Square Error (RMSE) as a loss function and Stochastic Gradient Descent (SGD) as the optimizer to train our model. The learning rate is set to 0.01 with a 10% decay per epoch. Both dropout ratio and recurrent dropout ratio for LSTM were set to 0.3 for all models. These values were set based on preliminary evaluations on training sets. In all experiments, we consider a split of 80/10/10 for training, validation, and testing after random shuffling. The results are the average of five separate training sets using five-fold cross-validation of 300 epochs each.

In addition to RMSE to measure the predictions' errors, we calculate Pearson's correlation coefficient to qualitatively show the relationship between the predicted and actual depth. Note that, Pearson's correlation has no unit while RMSE is measured in km. The two metrics complement each other to demonstrate the robustness of performance evaluation.

Pearson's correlation coefficient (ρ) is defined according to Equation 6, in which x_i is the predicted depth and y_i is the actual depth ($1 \leq i \leq number\ of\ test\ events$). Whereas, $\overline{y}$ and $\overline{x}$ are the average of predicted depth and average of actual depth respectively.

$$\rho = \frac{\Sigma(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\Sigma(x_i - \overline{x})^2 \Sigma(y_i - \overline{y})^2}} \tag{6}$$

To demonstrate the performance of Septor as a binary classifier, we modify the output layer of the station aggregator and use a softmax activation function. We use the Adam optimizer instead of SGD and binary cross-entropy as the loss function. In the evaluation, we consider accuracy, precision, recall, and F1-score as measures.

We train Septor using a GPU server with four Nvidia RTX2080 GPUs (total of 44GB GPU memory), 256 GB of RAM, and a 32-core CPU. Septor has 8.2 million trainable parameters and for 300 epochs, the server takes 1 hour and 48 minutes to train. All of our experiments are reproducible. The source code of data preprocessing and model training, waveforms and meta-data, figures, and additional results can be found in our supporting website [23].

7 EMPIRICAL EVALUATION

7.1 Regression performance

In Table 1, we compare the performance of Septor with different rival methods and data resolution. The RMSE of Septor for depth prediction is 2.89 km, which is lower than one standard deviation (3.71 km) of depth values in the test data. Furthermore, we achieve an impressive 70.1% Pearson's correlation coefficient.

Table (1) Performance comparison of Septor with baseline multiple methods.

Model	Data resolution	RMSE (km)	Corr. (%)
CNN	Multi-channel	3.26	56.0
LSTM	Multi-channel	3.38	52.0
XGBoost	Single-channel	3.53	37.0
XGBoost	Multi-channel	3.58	36.0
XGBoost	Multi-station	3.39	44.3
Rocket	Single-channel	3.11	46.2
Rocket	Multi-channel	3.12	46.0
Rocket	Multi-station	3.51	36.5
Septor	**Multi-station**	**2.89**	**70.1**

To confirm the statistical significance of our predictions, we run a t-test where we represent predicted depth with variable X, true depth with variable with Y, and our null hypothesis H_0 is: *there is no significant linear correlation between X and Y*. From the test, we found a p-value lower than the threshold value ($\alpha = 0.05$). Therefore, we reject the null hypothesis H_0 and conclude that the correlation between X and Y is statistically significant and did not occur by chance. Our model, Septor, significantly outperforms all other data-driven methods in terms of RMSE and Pearson's Correlation. In Figure 8, the scatter plot shows the dots representing the predictions close to the expected $y = x$ line.

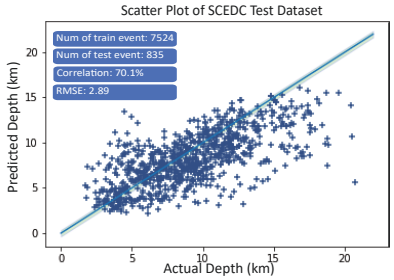

Figure (8) Scatter plot shows 70.1% correlation between actual vs. predicted depth on SCEDC dataset. The blue straight line represents $y = x$.

7.1.1 Sensitivity to the magnitudes of earthquakes. The observed noise in seismic signals depends on the earthquake's magnitude. Given that all other factors are the same, high magnitude earthquakes produce better signals than low magnitude ones. For this reason, we evaluate whether our model is sensitive to different magnitudes range, which could potentially decrease its utility.

We evaluated our model considering depth values in two distinct magnitude ranges: from 2 to 3 and from 3 to 4. The results are

shown in Table 2. For both ranges, Pearson's correlation coefficients are more than 86% with negligible (<< 0.05) p-value. We achieve an RMSE of 2.216 km for magnitude range 2-3 and 2.217 km for magnitude 3-4, where the standard deviation is 3.831 km and 3.712 km, respectively. The relationship between actual and predicted depths is shown in Figure 9. We conclude that the model is invariant to magnitude ranges.

Table (2) Performance evaluation on earthquakes of two magnitude ranges. RMSE and standard deviation are in km.

Magnitude	Test event Std Dev.	RMSE	Corr. (%)
2 - 3	3.831	2.216	86.5
3 - 4	3.712	2.217	88.2

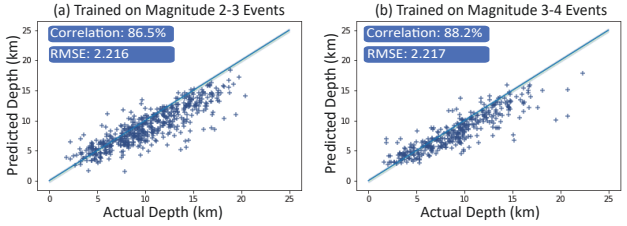

Figure (9) Scatter plots of predicted and actual depths for two magnitude ranges.

7.1.2 Sensitivity to origin location. Earth's non-uniformity is the greatest challenge to generalizable model development for seismic monitoring applications. Southern California has numerous dramatic lateral variations in seismic velocity structure, for example, with rapid transitions from the San Jacinto mountains with their deep crustal roots, to the Salton Trough, with its very thin crust, thick sediments, and high heat flow, and from areas where the crust is being sheared, to areas where it is under compression. Because of such geographic structural variability we evaluate our model on smaller sets of sub-regional events. As a simple test we divide our dataset into north and south sub-regions of Southern California and train and test using these smaller subsets. This is intended to test the performance of our model with earthquakes generated and captured within regions with at least somewhat less variation in earth structure. In Table 3, we show the performances on *North* and *South* datasets. RMSE values are lower than those obtained using the full dataset, and the correlation coefficients are higher for both sub-regions (Figure 10). This indicates that geographic restrictions help the model to perform better.

Table (3) Performance evaluation after splitting the dataset into two geographic locations. RMSE and standard deviation are in km.

Location	Test event Std Dev.	RMSE	Corr. (%)
North	3.065	1.907	85.4
South	3.804	2.264	85.9

7.1.3 Effects of different data preprocessing choices. In this experiment, we evaluate the effects of our choices in the preprocessing pipeline. First, we use spectrograms instead of linearly-spaced CWTs, since this data representation is widely employed in seismology. For the same SCEDC dataset, we observe a 5% decline in

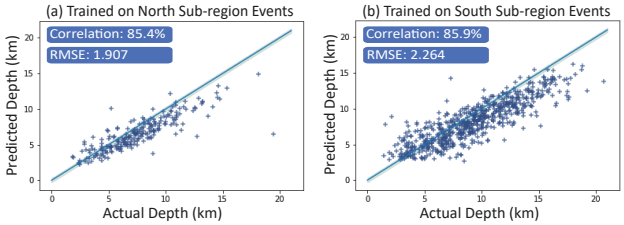

Figure (10) The model performs better for both North (left) and South (right) splitted dataset than whole dataset.

Pearson's Correlation coefficient when using spectrograms. Second, we use horizontal east-west, horizontal north-south components (Z, N, E channels) instead of rotating into radial and transverse components (Z, R, T channels). In this case, we observe Pearson's Correlation coefficient declines around 4% for the same set of earthquakes. Both results are shown in Figure 11. This empirical evaluation confirms that our choices work better than the alternatives.

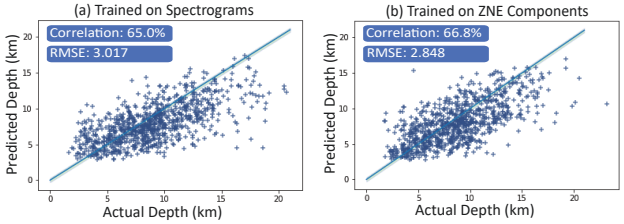

Figure (11) (a) Using spectrograms instead of CWTs results in a 5% performance decrease. (b) Using ZNE components of seismograms instead of rotating into ZRT components results in a 4% performance decrease.

7.1.4 Sensitivity to distance between stations and epicenters. The performance of depth prediction depends on the distance between the source event and observing stations. Information regarding depth in a waveform may vary with distance from the epicenter to the station. In this section, we demonstrate the performance of our model based on stations that fall into a fixed distance range. For example, if the distance range is 20-40 km, all the five stations for each earthquake are between 20 km and 40 km from the epicenter. This example is illustrated in Figure 12, in which we draw a doughnut shape around each earthquake epicenter with an inner circle of 20 km and an outer circle of 40 km. Then, we use the stations that fall within the doughnut-shaped region. Since Septor works after the Phase Association step, we have the epicenter distance information beforehand. Table 4 shows the performance in different distance ranges. Performance improves with distance up to the 40 to 60 km range, likely because the different seismic phases have a great deal of overlap at short distances. Performance then generally decreases with distance, possibly due to decreased signal-to-noise ratios and the accumulation of propagation effects on the waveforms.

7.2 Classification performance

In this section, we evaluate the performance of our network to classify deep and shallow earthquakes. Natural earthquakes can originate anywhere from the surface up to 700 km deep into the earth. United States Geological Survey (USGS) [27] defines three

Table (4) Performance evaluation after separating stations for train and test dataset. RMSE and standard deviation are in km.

Distance band (km)	Test event Std Dev.	RMSE	Corr. (%)
0 - 20	4.258	4.766	30.3
20 - 40	3.977	4.506	14.1
40 - 60	3.889	3.665	38.8
60 - 80	4.014	4.600	36.1
80 - 100	3.882	4.088	8.5
100 - 120	3.803	5.969	18.1

Table (5) Performance of Septor as a binary classifier for a subset of training data based on source to origin distance. The results are shown in percentage (%).

Distance (km)	Accuracy	Precision	Recall	F1-score
0-20	88.4	91.8	89.3	90.5
20-40	75.1	85.2	89.6	87.3
40-60	74.1	92.0	74.4	82.3
60-80	73.1	84.4	92.6	88.3
80-100	70.6	91.0	88.6	89.8
100-120	73.7	89.3	93.3	91.3
>120	67.8	67.1	68.9	68.0
Full dataset	73.0	78.2	75.5	86.5

categories of earthquakes depending on depth: *(i)* Shallow earthquakes originate between 0 to 70 km; *(ii)* Earthquakes that occur between 70 and 300 km are intermediate; *(iii)* Deep earthquakes originate deeper than 300 km into the ground. However, the cutoff depth to separate shallow and deep earthquakes can vary depending on the application. For most seismic monitoring purposes, it is sufficient to find out whether the earthquake occurred on the earth's surface or not. Therefore, an earthquake with a depth of 10 km can be labeled as a deep earthquake as man-made earthquakes (i.e., mining blasts, borehole shots, or nuclear explosions) can never originate 10 km below the surface.

In our experiments, we divide our full dataset into two balanced subsets of shallow and deep earthquake events considering the median depth (8.73 km) as the cutoff depth for shallow and deep earthquakes. We consider accuracy as a performance measure. The results are shown in the last row of Table 5. The confusion matrix is shown in Figure 13 (left).

7.2.1 Sensitivity to distance between station and epicenter. In this section, we demonstrate the performance of our binary classifier based on stations that fall into a fixed distance range.

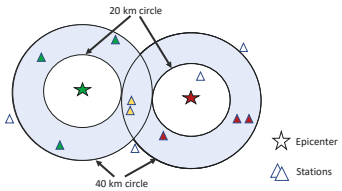

Figure (12) Example of the selection of earthquakes based on distance radius.

Based on our full dataset, we take seven different subsets depending on distance ranges. In Table 5, we show detailed results of our binary classifier for various distance ranges. Our results in Figure 13 (right) show that the accuracy for both deep and shallow earthquakes drop as the distance increases.

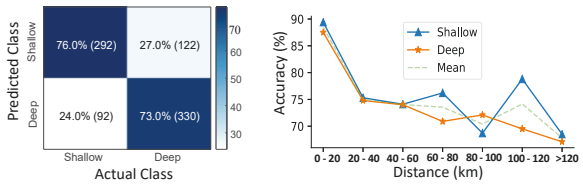

Figure (13) (left) Confusion matrix for binary classification for deep and shallow earthquakes. (right) Accuracy of the binary classifier drops with the source to origin distance of training data.

8 CASE STUDY: NOVEL GEOGRAPHICAL REGION

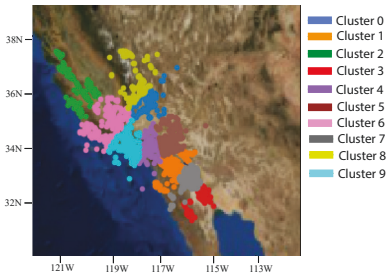

Figure (14) Locations of ten clusters found by DBSCAN based on epicenters of the events.

The ultimate goal of an ML-based depth estimator is to perform at a new location with a new set of stations. We ask if Septor can generalize to a novel geographical region without seeing any training instance from that region. We perform three different tests to evaluate generalizability of Septor.

8.0.1 No common source or station between train and test set. To simulate novel source-station scenarios, we sample the events and and stations in our dataset uniformly to create a test set, and use the remaining events and stations for training. In Figure 15(a) we show the scatter plot results for separated stations.

The results show a clear decrease in correlation to 44.8% from 70.1%, while the RMSE increases to 3.53 km. Although we create disjoint training and test sets with no common station and event source, the overall dataset is limited within the Southern California region, resulting a positive correlation.

8.0.2 Spatial separation between train and test sets. We cluster the events from SCEDC using the DBSCAN algorithm [22]. First, the earthquake events are clustered into ten regions based on epicenter (i.e., latitude and longitude), and then we perform a leave-one-out test. The clusters are shown in Figure 14 on the Southern California map.

In our test, we take each of these ten clusters in turn for testing, and use the remaining clusters for training; excluding the clusters adjacent to the test cluster. For example, if we take cluster 2 (green dots in Figure 14) as our test cluster, we train on clusters 0, 1, 3, 4, 5, 7, and 9. We exclude clusters 6 and 8 from this experiment as they are adjacent to our test cluster (2). The combined result for all test clusters are shown in Figure 15(b). RMSE of actual and predicted depth increases to 4.836 after spatial separation.

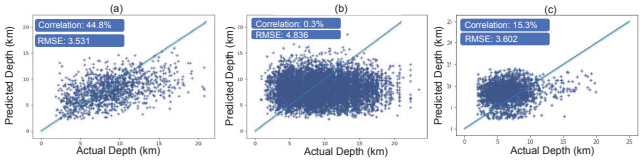

Figure (15) (a) Scatter plot for the model trained and tested on Southern California data after uniformly separating the stations in training and testing datasets. (b) Scatter plot of ten test cluster combined after spatial separation. (c) Scatter plot for the model trained on Southern California dataset and tested on Northern California dataset.

8.0.3 Test set from completely new region. We collected 1,777 natural earthquake events, each with seismograms from the five closest stations from the Northern California (NCEDC) seismic data center. We perform similar preprocessing to the seismograms and test our original model trained using the events from the SCEDC. The results are shown in Figure 15 (c). The model underperforms on this test dataset achieving a RMSE of 3.602 km with Pearson's Correlation coefficient of 15.3%. This is not surprising since Southern California and Northern California have different earth structures. Due to geographical dissimilarity, SCEDC and NCEDC seismograms generate different feature spaces, leading to poor model transferability across regions.

9 CONCLUSION

We present a two-level neural network model (Septor) to predict the depth of a seismic event using the waveforms recorded at multiple nearby stations. Septor achieves an impressive RMSE of 2.89 km, suitable for discriminating man-made events from most earthquakes when calibrated for the region. We demonstrate Septor's effectiveness in predicting depths of low magnitude earthquakes, which will enable fine-grained monitoring of underground events. When working in a classification scenario, our model can identify deep and shallow seismic events with an 86.5% F1-score.

10 ACKNOWLEDGEMENTS

This material is based upon work supported in part by the National Science Foundation EPSCoR Cooperative Agreement OIA-1757207 and Grant RI-2104537.

REFERENCES

[1] 2018. SCEDC (2013): Southern California Earthquake Center. Caltech.Dataset. doi:10.7909/C3WD3xH1. http://scedc.caltech.edu
[2] Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. 2017. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min. Knowl. Discov.* 31, 3 (2017), 606–660.
[3] Anthony Bagnall, Jason Lines, Jon Hills, and Aaron Bostrom. 2015. Time-Series Classification with COTE: The Collective of Transformation-Based Ensembles. *IEEE Trans. on Knowl. and Data Eng.* 27, 9 (Sept. 2015), 2522–2535. https://doi.org/10.1109/TKDE.2015.2416723
[4] U Berkeley. 2015. The NCSS now reports earthquake depth relative to the geoid (sea level). https://ncedc.org/blog/2015/10/07/the-ncss-now-reports-earthquake.html
[5] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *SIGKDD*. ACM, 785–794.
[6] Farhan Asif Chowdhury, M Ashraf Siddiquee, Glenn Eli Baker, and Abdullah Mueen. 2021. FASER: Seismic Phase Identifier for Automated Monitoring. In *SIGKDD*. 2714–2721.
[7] Hoang Anh Dau, Eamonn Keogh, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, Yanping, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen, Gustavo Batista, and Hexagon-ML. 2018. The UCR Time Series Classification Archive. https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.
[8] Angus Dempster, François Petitjean, and Geoffrey I Webb. 2020. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Min. Knowl. Discov.* 34, 5 (2020), 1454–1495.
[9] Ramin MH Dokht, Honn Kao, Ryan Visser, and Brindley Smith. 2019. Seismic event and phase detection using time–frequency representation and convolutional neural networks. *Seismological Research Letters* 90, 2A (2019), 481–490.
[10] Peter Goldstein. 1995. Slopes of P- to S-wave spectral ratios—A broadband regional seismic discriminant and a physical model. *Geophysical Research Letters* 22, 23 (1995), 3147–3150. https://doi.org/10.1029/95GL03315 arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/95GL03315
[11] Egill Hauksson, Wenzheng Yang, and Peter M Shearer. 2012. Waveform relocated earthquake catalog for southern California (1981 to June 2011). *BSSA* 102, 5 (2012), 2239–2244.
[12] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9 (1997), 1735–1780.
[13] Rob J Hyndman. 2020. A brief history of forecasting competitions. *International Journal of Forecasting* 36, 1 (2020), 7–14.
[14] Alan L. Kafka. 1990. Rg as a depth discriminant for earthquakes and explosions: A case study in New England. *Bulletin of the Seismological Society of America* 80, 2 (04 1990), 373–394. https://doi.org/10.1785/BSSA0800020373 arXiv:https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/80/2/373/5339628/bssa0800020373.pdf
[15] K. D. Koper, J. C. Pechmann, R. Burlacu, K. L. Pankow, J. R. Stein, J. M. Hale, P. Roberson, and M. K. McCarter. 2016. Magnitude Based Discrimination of Manmade Seismic Events From Naturally Occurring Earthquakes in Utah, USA. In *AGU Fall Meeting Abstracts*, Vol. 2016. Article S31A-2721, S31A-2721 pages.
[16] Guoqing Lin, Peter M Shearer, and Egill Hauksson. 2007. Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern California seismicity from 1981 to 2005. *Journal of Geophysical Research: Solid Earth* 112, B12 (2007), 1–14.
[17] S Mostafa Mousavi, Weiqiang Zhu, Yixiao Sheng, and Gregory C Beroza. 2019. CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Scientific reports* 9, 1 (2019), 1–14.
[18] Luis H Ochoa, Luis F Niño, and Carlos A Vargas. 2018. Fast determination of earthquake depth using seismic records of a single station, implementing machine learning techniques. *Ingeniería e Investigación* 38, 2 (2018), 97–103.
[19] Division of Geological and California Institute of Technology Planetary Sciences. 2020. SCSN Catalog Change History. https://scedc.caltech.edu/eq-catalogs/change-history.html.
[20] Thibaut Perol, Michaël Gharbi, and Marine Denolle. 2018. Convolutional neural network for earthquake detection and location. *Science Advances* 4, 2 (2018), e1700578.
[21] SCEDC. 2021. Station metadata and maps. Southern California Earthquake Data Center at Caltech. Available at https://scedc.caltech.edu/data/station/index.html.
[22] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)* 42, 3 (2017), 1–21.
[23] M. Ashraf Siddiquee, V. M. A. Souza, and Abdullah Mueen. 2022. Public github repository to download. https://github.com/mashrafsiddiq/Septor.
[24] William Spence, Stuart A Sipkin, and George L Choy. 1989. Measuring the size of an earthquake. *Earthquake Information Bulletin (USGS)* 21, 1 (1989), 58–63.
[25] Chang Wei Tan, Christoph Bergmeir, François Petitjean, and Geoffrey I Webb. 2021. Time series extrinsic regression. *Data Mining and Knowledge Discovery* 35, 3 (2021), 1032–1060.
[26] Daniel T. Trugman and Peter M. Shearer. 2017. GrowClust: A Hierarchical Clustering Algorithm for Relative Earthquake Relocation, with Application to the Spanish Springs and Sheldon, Nevada, Earthquake Sequences. *Seismological Research Letters* 88, 2A (02 2017), 379–391. https://doi.org/10.1785/0220160188 arXiv:https://pubs.geoscienceworld.org/ssa/srl/article-pdf/88/2A/379/2745678/379.pdf
[27] USGS. 2021. United States Geological Survey. Available at https://www.usgs.gov/ (September 2, 2021).
[28] Felix Waldhauser and David P. Schaff. 2008. Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods. *Journal of Geophysical Research: Solid Earth* 113, B8 (2008). https://doi.org/10.1029/2007JB005479 arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007JB005479
[29] De-He Yang, Xin Zhou, Xiu-Ying Wang, and Jian-Ping Huang. 2021. Mirco-earthquake source depth detection using machine learning techniques. *Information Sciences* 544 (2021), 325–342. https://doi.org/10.1016/j.ins.2020.07.045