

A Distributed Multi-Sensor Machine Learning Approach to Earthquake Early Warning

Kevin Fauvel,¹ Daniel Balouek-Thomert,² Diego Melgar,³ Pedro Silva,¹ Anthony Simonet,²
Gabriel Antoniu,¹ Alexandru Costan,¹ Véronique Masson,¹
Manish Parashar,² Ivan Rodero,² Alexandre Termier¹

¹Univ Rennes, Inria, CNRS, IRISA, Rennes, France

²Rutgers Discovery Informatics Institute, Rutgers University, New Jersey, USA

³Department of Earth Sciences, University of Oregon, Oregon, USA

{kevin.fauvel, gabriel.antoniu}@inria.fr, {pedro.silva, alexandru.costan, veronique.masson, alexandre.termier}@irisa.fr,
{daniel.balouek, anthony.simonet, parashar, irodero}@rutgers.edu, dmelgarm@uoregon.edu

Abstract

Our research aims to improve the accuracy of Earthquake Early Warning (EEW) systems by means of machine learning. EEW systems are designed to detect and characterize medium and large earthquakes before their damaging effects reach a certain location. Traditional EEW methods based on seismometers fail to accurately identify large earthquakes due to their sensitivity to the ground motion velocity. The recently introduced high-precision GPS stations, on the other hand, are ineffective to identify medium earthquakes due to its propensity to produce noisy data. In addition, GPS stations and seismometers may be deployed in large numbers across different locations and may produce a significant volume of data consequently, affecting the response time and the robustness of EEW systems.

In practice, EEW can be seen as a typical classification problem in the machine learning field: multi-sensor data are given in input, and earthquake severity is the classification result. In this paper, we introduce the Distributed Multi-Sensor Earthquake Early Warning (DMSEEW) system, a novel machine learning-based approach that combines data from both types of sensors (GPS stations and seismometers) to detect medium and large earthquakes. DMSEEW is based on a new stacking ensemble method which has been evaluated on a real-world dataset validated with geoscientists. The system builds on a geographically distributed infrastructure, ensuring an efficient computation in terms of response time and robustness to partial infrastructure failures. Our experiments show that DMSEEW is more accurate than the traditional seismometer-only approach and the combined-sensors (GPS and seismometers) approach that adopts the rule of relative strength.

1 Introduction

Earthquakes cause substantial loss of life and damage to the built environment across areas spanning hundreds of kilometers from their origins. These large ground motions often lead to hazards such as tsunamis, fires and landslides. To mitigate the disastrous effects, a number of Earthquake Early Warning (EEW) systems have been built around the world (Allen and Melgar 2019). These critical systems, operating 24/7, are expected to automatically detect and char-

acterize earthquakes as they happen, and to deliver alerts before the ground motion actually reaches sensitive areas so that protective measures could be taken.

A recent review by (Allen and Melgar 2019) identified the detection of the whole spectrum of earthquakes with damaging potential and particularly large earthquakes as an outstanding problem in the field of EEW. An EEW system needs to be able to detect both medium ($5 \leq \text{magnitude} < 6$, Richter scale) and large earthquakes ($6 \leq \text{magnitude}$, Richter scale). Depending on the distance from the origin of the earthquake, both of these can cause serious damages. Seismometers, which have long been the bulwark of seismology to detect earthquakes, have a difficulty to detect and characterize large earthquakes (Melgar et al. 2013) due to a well-known saturation issue caused by their sensitivity to ground motion velocity. As a result, earthquakes over magnitude 7.5 tend to be underestimated. A promising solution to this issue (Melgar et al. 2015) emerged with novel high-precision Global Positioning System (GPS) sensors, with their millimeter to centimeter accuracy when measuring high ground motion velocity. However, GPS are unable to characterize medium earthquakes, as they are prone to containing significant signals from a variety of noise sources, mostly of atmospheric origin. Consequently, multi-sensor solutions (leveraging both GPS and seismometers) appear as a promising approach. As EEW can be assimilated as a classification problem, where the input is sensor data and the output is a class (normal activity/medium earthquake/large earthquake); recent machine learning approaches designed to combine large volumes of data from multiple data sources can be applied.

Integrating and processing high-frequency data streams from multiple sensors scattered over a large territory in a timely manner requires high-performance computing techniques and equipments. Thus, a machine learning earthquake detection solution has to be designed jointly with experts in distributed computing and cyberinfrastructure to enable real-time alerts. Because of the large number of sensors and their high sampling rate, a traditional centralized approach which transfers all data to a single point may be impractical. We therefore design a distributed machine learning-based approach.

In this paper, we use machine learning methods to address the most urgent challenges faced by EEW systems, i.e. integrating multiple data sources in real-time to cover the whole spectrum of potentially damaging earthquakes (medium and large). Our solution relies on two complementary types of sensors (GPS stations and seismometers). We introduce a new machine learning technique specifically tailored to allow efficient computation on large-scale distributed cyberinfrastructures.

The contributions of this paper are the following:

- We propose a new EEW approach to characterize the whole spectrum of earthquakes with damaging potential (both medium and large) using a real-world dataset collected and validated with geoscientists combining two complementary data sources;
- We present DMSEEW (Distributed Multi-Sensor Earthquake Early Warning), a new stacking ensemble method jointly designed with cyberinfrastructure experts, which enables real-time earthquake detection and robustness to partial infrastructure failures;
- We show that DMSEEW is more accurate than both the seismometer-only baseline approach, the combined sensors (GPS and seismometers) baseline approach that adopts the rule of relative strength and detects all large earthquakes with a precision of 100%.

2 Background and Related Work

Before discussing in detail our approach, it is necessary to introduce some key concepts and to understand how our strategy fits in the state-of-the-art literature on EEW systems.

2.1 Earthquake Early Warning

An earthquake is the shaking of the surface of the Earth caused by seismic waves. Among these seismic waves, two types stand out: Primary waves (P-waves) and Secondary waves (S-waves). Both waves have the same origin - most commonly an abrupt movement of tectonic plates. However, P-waves travel through Earth's crust around 1.7 times faster than S-waves which propagate through Earth's interior. In addition, only S-waves are responsible for the severe damages. P-waves cause soft shaking due to their longitudinal shape (they move sideways), whereas S-waves are transverse waves (they move up and down). Therefore, an Earthquake Early Warning (EEW) system, which aims to provide an alert before the damaging effects reach sensitive areas, relies on the detection of the P-wave before the S-wave arrives. This gives communities, organizations and governments a time window of seconds to minutes to take protective actions.

Traditionally, inertial seismometers are used to detect primary waves. The inertial mass is designed to remain stationary following sudden movements while the frame and drum move with the ground to record waves. However, during large earthquakes, ground motion velocity causes the inertial mass to be displaced above the allowed span. This effect

is called saturation. As a result, earthquakes over magnitude 7.5 (Richter scale) tend to be underestimated. On the other hand, GPS satellites are not affected by earthquakes, so a GPS receiver station on Earth can be used to assess strong ground motion based on the station displacement. However, GPS is sensitive to a variety of noise sources, mostly of atmospheric origin, and is unable to characterize moderate earthquakes. Both sensors produce data in the form of 3D time series indicating the direction of a ground motion (east-west, north-south and up-down) at a frequency of around 20Hz.

P-waves follow a propagation model (IASP91 (Kennet 1991)) which we use for labeling the time series (sequences of measurements) corresponding to an earthquake. Based on the distance below Earth's surface where each earthquake happened, we estimate the P-wave arrival time on each sensor (seismometers and GPS stations) according to its distance to the epicenter with the propagation model.

2.2 Multivariate Time Series Classification

A time series is multivariate when a sequence of measurements from multiple variables are present. Multivariate Time Series (MTS) collected from GPS stations (3 dimensions: east-west, north-south and up-down) and seismometers (3 dimensions: east-west, north-south and up-down) are labeled in 3 classes according to the the potential damage of the ground motion: normal activity, medium earthquakes, large earthquakes. Therefore, earthquake detection can be formulated as a MTS classification problem.

MTS classifiers are composed of 3 categories: similarity-based, feature-based and deep learning methods.

Similarity-based methods make use of similarity measures (e.g. Euclidean distance) to compare two MTS. Dynamic Time Warping (DTW) has been shown to be the best similarity measure to use along k-Nearest Neighbors (k-NN) (Seto, Zhang, and Zhou 2015) and is called kNN-DTW. There are two versions of kNN-DTW for MTS: dependent (DTW_D) and independent (DTW_I). Neither dominates over the other (Shokoohi-Yekta et al. 2017). DTW_I measures the cumulative distances of all dimensions independently measured under DTW. DTW_D uses a similar calculation with single-dimensional time series; it considers the squared Euclidean cumulated distance over the multiple dimensions.

Feature-based methods include shapelets and bag-of-words (BoW) models. Shapelets models use subsequences (shapelets) to transform the original time series into a lower-dimensional space that is easier to classify. gRSF (Karls-son, Papapetrou, and Boström 2016) and UFS (Wistuba, Grabocka, and Schmidt-Thieme 2015) are the current state-of-the-art shapelets models in MTS classification. They relax the major limiting factor of the time to find discriminative subsequences in multiple dimensions (shapelet discovery) by randomly selecting shapelets. gRSF creates decision trees over randomly extracted shapelets and shows better performance than UFS on average (14 MTS datasets) (Karls-son, Papapetrou, and Boström 2016). On the other hand, BoW models (LPS (Baydogan and Runger 2016), mv-ARF (Tuncel and Baydogan 2018), SMTS (Baydogan and Runger 2014) and WEASEL+MUSE (Schäfer and Leser

2017)) convert time series into a bag of discrete words, and use a histogram of words representation to perform the classification. WEASEL+MUSE shows the best results compared to gRSF, LPS, mv-ARF and SMTS on average (20 MTS datasets) (Schäfer and Leser 2017). WEASEL+MUSE generates a BoW representation by applying various sliding windows with different sizes on each discretized dimension (Symbolic Fourier Approximation) to capture features (unigrams, bigrams, dimension identification).

Finally, deep learning methods use Long-Short Term Memory (LSTM) and/or Convolutional Neural Networks (CNN) to extract latent features. The current state-of-the-art model (MLSTM-FCN) is proposed in (Fazle, Majumdar, and Harford 2018) and consists of a LSTM layer and a stacked CNN layer along with Squeeze-and-Excitation blocks to generate latent features. MLSTM-FCN is shown to be better than WEASEL+MUSE on large datasets on average (relative to the 20 MTS datasets tested).

Therefore, in this work we choose to experiment with the best-in-class for each category (similarity-based, feature-based and deep learning): DTW_D , DTW_I , WEASEL+MUSE and MLSTM-FCN classifiers.

2.3 Targeting a Distributed Cyberinfrastructure

A cyberinfrastructure is the set of logical and physical computational systems onto which a scientific application is deployed. In the context of EEW, a cyberinfrastructure has to support the processing of large amounts of data produced by geographically distributed seismic sensors, such as GPS stations and seismometers, in real-time.

Current approaches to EEW in the literature make use of centralized data processing strategies: all sensors send their data, through a network, to a data center where processing will take place (Fischer et al. 2012). This strategy implicitly depends on cyberinfrastructures supporting almost insignificant network latency and very high bandwidth, often provided by costly fiber networks. Furthermore a lot of trust is placed in the reliability of the network and telemetry paths (Allen and Melgar 2019). In the case of large earthquakes, power failures and wiring disconnections can frequently lead to regional shutdowns, as happened after the magnitude-9 earthquake in Japan in 2011 (Hoshiba and Ozaki 2012).

In this work, we target a distributed cyberinfrastructure for executing the proposed EEW system, meaning that data processing tasks can be performed in different parts of the infrastructure and at different locations. In particular, we favor processing part of the data at the edge of the network, i.e. as close as possible to the sources of data, in order to reduce the amount of data transferred to the main data center (Yang et al. 2010).

2.4 EEW Machine Learning Solutions

Machine learning in seismology is still a developing field. There are a couple of studies (Yoon et al. 2015; Li et al. 2018; Perol, Gharbi, and Denolle 2018) using machine learning methods for earthquake characterization based on P-wave detection (EEW). However, none of them used a combination of GPS and seismometers data so the whole

spectrum of earthquakes with damaging potential is not appropriately covered. Additionally, none of them used a distributed approach.

The three studies (Yoon et al. 2015; Li et al. 2018; Perol, Gharbi, and Denolle 2018) adopted a binary classification approach (earthquakes vs. noise) with no distinction between medium and large earthquakes. Moreover, the detection is only based on seismometers data, so the saturation issue on large earthquakes is present. Two of these studies (Perol, Gharbi, and Denolle 2018; Yoon et al. 2015) limit their scope to medium earthquakes. Yoon et al. (2015) proposed a waveform similarity-based method optimized by locality-sensitive hashing search using seismometers data from California. It presented a precision of 88.1% and a recall of 87.5%. Li et al. (2018) developed a generative adversarial network with a random forest on seismometers data from Southern California and Japan. It obtained an accuracy of 99.2%. Perol, Gharbi, and Denolle (2018) trained a convolutional neural network using seismometers data from Oklahoma and showed a precision of 94.8% and a recall of 100%.

Lastly, these studies use a centralized data processing strategy and implicitly assume that data arrive almost instantly at the processing cluster. This assumption implies an insignificant latency and an overall network reliability which is unfeasible in real-life scenarios having high latency and/or low bandwidth networks.

Consequently, a distributed EEW machine learning-based solution that can be generalized to the whole spectrum of earthquakes with damaging potential is necessary.

3 A Distributed Machine Learning Approach to Earthquake Early Warning

In this section, we present the Distributed Multi-Sensor Earthquake Early Warning algorithm (DMSEEW), a new two-step stacking ensemble method for earthquake detection. A stacking ensemble is a method which takes the predictions of sub-models as inputs and then attempts to learn how to best combine the input predictions to make a better output prediction. DMSEEW takes sensor-level class predictions (normal activity, medium earthquake or large earthquake) based on the data gathered by each individual sensor (GPS stations and seismometers). It then aggregates those sensor-level class predictions using a bag-of-words representation in order to calculate a final prediction for the earthquake category.

The rest of this section explains the algorithm steps and introduces the cyberinfrastructure onto which DMSEEW would be deployed in a real-life scenario.

3.1 Algorithm

Step 1 – Predicting the MTS Category at the Sensor-Level: We have two types of sensors - GPS stations and seismometers, and we train one MTS classifier per sensor type. The classifiers are trained using a dataset composed of time series of 3 dimensions (east-west, north-south and up-down) and fixed time length (60 seconds, defined in Section 4.1). We illustrate this first step of our approach in the

upper part of Figure 1.

In order to predict the earthquake category at the individual sensor level, we employ the WEASEL+MUSE (Schäfer and Leser 2017) MTS classifier. WEASEL+MUSE creates a symbolic representation of the MTS (a Symbolic Fourier Approximation - SFA) on each dimension, then generates a set of features (multiple window lengths, unigrams, bigrams, dimension identification), and finally performs the classification based on a one-hot encoding representation of the MTS (bag-of-words, feature selection). WEASEL+MUSE fits our approach because (i) its symbolic representation filters out noise (related to GPS and seismometers sensors) from the dataset; (ii) it is phase invariant, i.e. features generated do not have to appear at the same time across different MTS belonging to the same class, which improves generalization; (iii) it keeps the interplay of dimensions since features generated by WEASEL+MUSE contain the identifier of the dimension, which allows the characterization of co-occurrence of events on different dimensions. As we further discuss in the Section 5, WEASEL+MUSE outperforms other MTS classifiers on both GPS and seismometers data.

Step 2 – Detecting Earthquakes by Combining Sensor-level Predictions: We collect the class predictions from the different sensors (GPS stations and seismometers) and perform a bag-of-words representation. Each sensor-predicted class is considered to be a word and the relative frequency vector of the words from each earthquake is used to classify its category. This frequency vector is normalized by the number of instances (number of MTS per earthquake, i.e. number of sensors) to obtain the relative frequency vector. The last step consists of combining the bag-of-words of GPS stations and seismometers to characterize the whole spectrum of earthquakes with damaging potential. We illustrate this second step of our approach in the lower part of Figure 1. For example, 80% of seismometers and 100% of GPS stations for event 1 predict that the activity recorded is normal. Finally, we train a classifier on this bag-of-words representation to perform the combined class prediction. As presented in Section 5, 1-nearest neighbor outperforms other classifiers.

3.2 Cyberinfrastructure

We propose a distributed cyberinfrastructure composed of geographically distributed data sources. Its objective is to support the processing of high volumes of data produced by GPS stations or seismometers, meanwhile meeting the real-time requirements of EEW applications. Furthermore, in disaster situations, such as partial network shutdown, this type of architecture allows for the redirection of data produced at sensor-level (i.e. individual sensor predictions) to other processing data centers. Figure 2 provides an overview of the cyberinfrastructure.

In summary, the cyberinfrastructure has two main levels: the sensor-level and the central level. The sensor-level is composed of sensing devices (i.e. GPS stations and seismometers) with limited computing capabilities. The central level is composed of well-provisioned computing sys-

tems that can accommodate large computing demands (e.g. cloud datacenters). This type of architecture has been the subject of several contributions in the context of Edge computing (Shi et al. 2016; Satyanarayanan 2017), an emerging paradigm adapted for Internet of Things (IoT) scenarios. Due to space constraints, we leave an in-depth discussion of cyberinfrastructure in a future work. In this paper, we focus on the distributed algorithm in charge of analyzing the data produced by seismic sensors and on its interaction with those main levels of the infrastructure.

Distributed execution. The first step of the algorithm is performed on the sensor-level part of the infrastructure. There, a MTS classifier is ran on each individual sensor (GPS stations and seismometers) in order to generate sensor-level class predictions based on data produced by each sensor. Then, the output of the MTS classifier from each sensor is transferred over the network to the central level part of the cyberinfrastructure. There, the second part of the algorithm is ran, i.e., a machine learning method combines all the class predictions from GPS stations and seismometers to form a final class prediction. It is important to highlight that, in comparison to the traditional centralized EEW cyberinfrastructure, this approach drastically reduces the amount of data over the network since most of data produced by a sensor is not related to an earthquake event and thus can be filtered out. Moreover, a sensor-level prediction is, in fact, an aggregation of data, hence, it also helps reduce on the amount of data sent to central to level data centers.

4 Experiments

In this section, we introduce the methodology and datasets used for evaluating our work, as well as our preprocessing routines and experimental setting. In addition, we render public our real-world dataset collected and validated with geoscientists and we make public reference to the code of our machine learning algorithms used.

4.1 Dataset

We employ a real-world dataset¹ (Fauvel et al. 2019) composed of GPS and seismometers data on normal activity/medium earthquakes/large earthquakes collected and validated with geoscientists. There are two main difficulties to construct such a dataset: *i*) large earthquakes are rare and *ii*) GPS data is not continuously recorded. The dataset has been built around the most complete GPS/seismometers dataset of large earthquakes (29 earthquakes worldwide) which occurred between 2001 and 2018 (Ruhl et al. 2019) with the corresponding metadata (time, magnitude, and location). We adopted a differentiated approach between GPS and seismometers data to augment the dataset and present it in the following two paragraphs. MTS length is set to 60 seconds for both GPS and seismometers. This value reflects the relevant time window to distinguish primary waves from noise across geographical regions, as recognized by the geoscience community.

¹https://figshare.com/articles/Earthquake_Early_Warning_Dataset/9758555

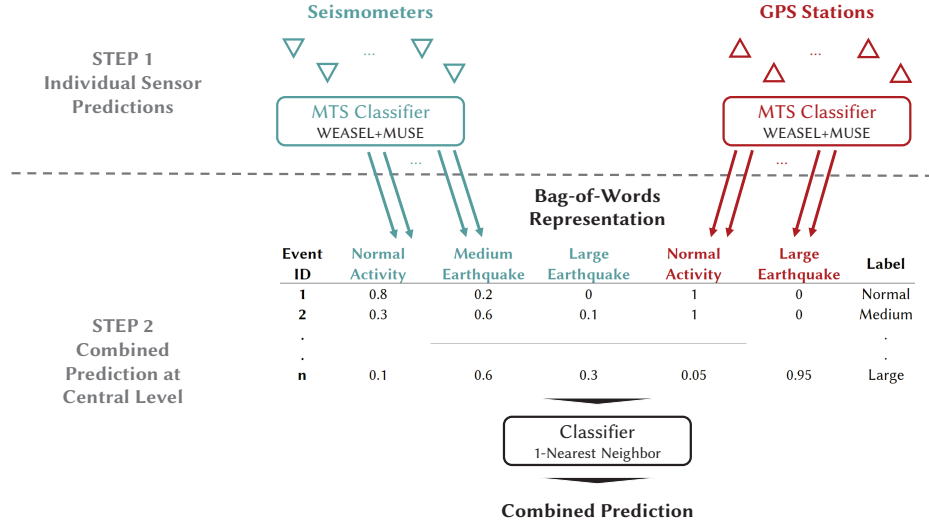


Figure 1: Distributed Multi-Sensor Earthquake Early Warning Algorithm (DMSEEW).

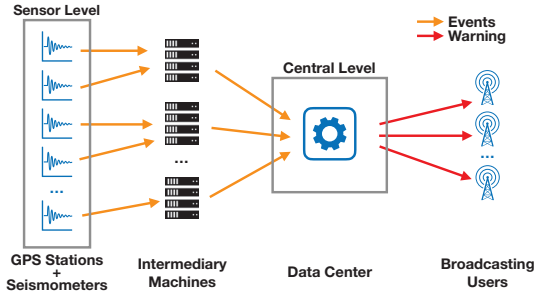


Figure 2: High-level architecture and data workflow for the Earthquake Early Warning System.

First, the two main seismometers data repositories worldwide are the American Incorporated Research Institutions for Seismology (IRIS) and the Japanese National Research Institute for Earth science and Disaster Resilience (NIED). Earthquake origins are defined differently between the two repositories, preventing a direct comparison of P-wave arrival time on each seismometer. Therefore, in order to be able to adopt a homogeneous labeling method, we limited our study to the data available from IRIS (14 large earthquakes remaining over 29). Seismometers data corresponding to medium earthquakes are sampled from medium earthquakes occurring in the same region as large earthquakes ($-179 \leq \text{longitude} \leq 25$, $-62 \leq \text{latitude} \leq 73$). The number of medium earthquakes is calculated by the ratio of medium over large earthquakes during the past 10 years in the region. Then, we keep a ratio above 30% between the number of MTS corresponding to earthquakes (medium + large) and total (earthquakes + normal activity) number of MTS to prevent a class imbalance issue during the training phase. So, we collect two normal activity MTS for each medium earthquake MTS (9 and 7 minutes before each medium

earthquake) to respect this ratio. IRIS data (normal activity, medium earthquakes) is collected with the international Federation of Digital Seismograph Networks (FDSN) client available in Python package ObsPy². Based on geoscience expertise, the relevant region of seismometers is set to 1,000 kilometers around the earthquake epicenter.

Second, unlike seismometers data, GPS displacement data is not continuously recorded. Furthermore, GPS data outside of large earthquake periods can be considered as normal activity (noise). Hence, our approach based on GPS sensors characterizes only normal activity and large earthquakes. We collected GPS normal activity data from an archive of real-time GPS positions maintained by the University of Oregon³ which stores a representative extract of GPS noise. Normal activity MTS are randomly sampled from the archive to match the number of seismometers events (255, normal activity and medium earthquakes) and to keep a ratio above 30% between the number of large earthquakes MTS and normal activity in order to avoid class imbalance issues.

The number of sensor records available varies between earthquakes according to the location and the magnitude of the earthquake. The full dataset composition is presented in Table 1.

Table 1: Dataset Composition

# 60s MTS	Seismometers (# Events)	GPS (# Events)
Normal Activity	7,718 (170)	1,424 (255)
Medium Earthquakes	3,859 (85)	None
Large Earthquakes	1,688 (14)	648 (14)
Total	13,265 (269)	2,072 (269)

²<https://docs.obspy.org/packages/obspy.clients.fdsn.html>

³<http://tunguska.uoregon.edu/rtgnss/data/cwu/mseed/>

4.2 Preprocessing

First, seismometers data are available as digital signal, which is specific for each sensor. Therefore, we converted each instrument digital signal to its physical signal (acceleration) to obtain comparable seismometers data. Second, we performed standardization (StandardScaler (Pedregosa et al. 2011)) of the GPS and seismometers data (fitted on train sets and applied on test sets) to harmonize the different scales. Standardization procedure allows us to keep outliers, which are fundamental in P-wave detection, as compared to the normalization procedure. Finally, we perform data aggregation by second (mean) which permits a common time scale between sensors (frequency between sensors can differ) without deteriorating the P-wave signal.

4.3 Experimental Setting

In this section, we present the algorithms evaluated and the methods used to assess them.

Algorithms We evaluate different algorithms at sensor-level and central level according to the data type in order to define the two blocks of our machine learning solution (Figure 1).

At sensor-level, we have a multivariate time series classification task. Therefore, as presented in section 2.2 on MTS classification, we compare the following algorithms:

- DTW_D and DTW_I : we use the public implementation⁴ based on the original paper (Shokoohi-Yekta et al. 2017);
- WEASEL+MUSE: we use the public implementation⁵ with the recommended settings (SFA word lengths 1 in [2,4,6], windows length in [4:60], $\chi=2$, $\text{bias}=1$, $p=0.1$, $c=5$ and a solver equals to L2R LR DUAL) (Schäfer and Leser 2017);
- MLSTM-FCN, we test the public implementation⁶ based on the original paper (Fazle, Majumdar, and Harford 2018), using the recommended settings (128-256-128 filters, 250 training epochs, a dropout of 0.8 and a batch size of 128);

At central level, we have a classification task on a bag-of-words representation (relative frequency vector) for each earthquake based on individual class predictions of GPS stations and seismometers. We compare the state-of-the art classifiers with the following implementations: K-Nearest Neighbors⁷; Elastic Net⁷; Support Vector Machine⁷ with a radial basis function kernel due to the lower number of features than the number of samples in our dataset; Random Forest⁷ and Extreme Gradient Boosting⁸.

Hyperparameters Optimization Firstly, at sensor-level, classifier hyperparameters setting is presented in previous section with the public implementations of the algorithms

⁴<https://github.com/DavideNardone/MTSS-Multivariate-Time-Series-Software>

⁵<https://github.com/patrickzib/SFA>

⁶<https://github.com/titu1994/MLSTM-FCN>

⁷<https://scikit-learn.org/stable/>

⁸<https://xgboost.readthedocs.io/en/latest/python/>

used. Next, hyperparameters of classifiers at central level are set by hyperopt, a sequential model-based optimization using a tree of Parzen estimators search algorithm (Bergstra, Yamins, and Cox 2013). Hyperopt chooses the next hyperparameters decision from the previous choices and a tree-based optimization algorithm. Tree of Parzen estimators meet or exceed grid search and random search performance for hyperparameters setting (Bergstra et al. 2011). We use the implementation available in the Python package hyperopt⁹. Optimization is undertaken to maximize accuracy score considering our multiclass study.

Classification Performance Classifiers are trained (at sensor and central level) with a 3 class labeling (see section 4.1) on seismometers data and a 2 class labeling on GPS data. We performed a stratified k-fold cross-validation which kept the same proportion of earthquakes of different categories for each fold. K is set to 3 considering the number of large earthquakes (14 earthquakes). We present the dataset split in Table 2. Therefore, the results presented corresponds to the 3-fold performance on the test sets. Classifiers are trained to optimize the accuracy score.

Table 2: Cross-Validation Split

# Events	Fold 1	Fold 2	Fold 3	Total
Seismometers	90	90	89	269
Normal Activity	56	57	57	170
Medium Earthquakes	29	28	28	85
Large Earthquakes	5	5	4	14
GPS	90	90	89	269
Normal Activity	85	85	85	255
Medium Earthquakes	None	None	None	None
Large Earthquakes	5	5	4	14

5 Results and Discussions

This section first presents the results at sensor-level (DM-SEEW step 1). Then, we evaluate the performance of our combined approach (DMSEEW steps 1 and 2) compared to the traditional seismometers baseline approach and the combined sensors (GPS and seismometers) baseline approach that adopts the rule of relative strength.

5.1 Sensor-Level Predictions

We present the accuracy results of the different MTS classifiers on GPS stations (2,072 MTS, 2 classes) and seismometers (13,265 MTS, 3 classes) in Table 3.

Table 3: Accuracy Score on Test Sets of the MTS Classifiers Trained on GPS or Seismometers Data

Accuracy (%)	DTW_D	DTW_I	MLSTM - FCN	WEASEL + MUSE
Seismometers	35.3	35.5	54.6	63.6
GPS	97.9	97.8	98.9	99.5

⁹<https://github.com/hyperopt/hyperopt>

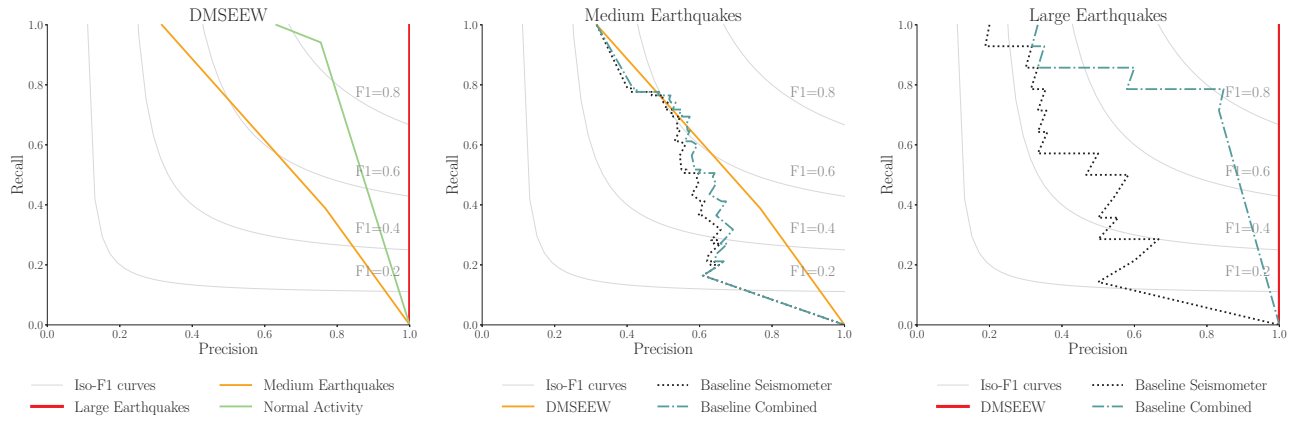


Figure 3: Precision-Recall Curves of DMSEEW.

We observe that WEASEL+MUSE outperforms MLSTM-FCN and similarity-based classifiers (DTW_D and DTW_I) on both GPS and seismometers data. The difference between WEASEL+MUSE and other classifiers is particularly important on seismometers data. We can infer that the noise reduction performed by the truncated Fourier Transform and discretization of WEASEL+MUSE led to a better exploitation of the P-wave information.

The detection results obtained from both GPS stations and seismometers data confirm the complementary performance of these sensors. GPS data distinguishes large earthquakes while the detection based on seismometers data performs poorly (F1 score on large earthquakes: GPS 99%, seismometers 28%). In the next section, we present the results of a combined prediction benefiting from the complementary performance of these sensors following the transfer of all sensor-level class predictions to a central computation facility.

5.2 Combined Predictions at Central Level

DMSEEW benefits from the complementary performances through the combination of sensor-level class predictions (GPS and seismometers) using a bag-of-words representation, followed by the training of a classifier on this representation. There is no state-of-the-art method covering the whole spectrum of earthquakes with damaging potential (medium and large). In order to evaluate the performance of DMSEEW, we define two baselines at central level.

The first is the traditional seismometer approach which relies on seismometers data only. We compute the performance of the traditional seismometer approach by calculating the arg max directly on the seismometers bag-of-words representation for each of the 269 events in our dataset. The second baseline corresponds to the combined sensors (GPS and seismometers) baseline approach that adopts the rule of relative strength in order to assess the value added by DMSEEW combining approach. It is defined based on the strengths of each sensor type: if the GPS bag-of-words representation indicates that the event is a large earthquake, a large earthquake is predicted. Else, we calculate the arg max

Table 4: Performances on Test Sets of DMSEEW and the Two Baselines. Standard Errors are Presented in Parentheses

	Baseline Seismometer	Baseline Combined	DMSEEW
Normal Activity			
Precision (%)	76.2 (1.8)	76.6 (1.5)	75.5 (1.8)
Recall (%)	94.1 (3.2)	94.1 (3.2)	94.1 (3.2)
F1 (%)	84.2 (0.6)	84.4 (0.7)	83.8 (0.9)
Medium Earthquakes			
Precision (%)	65.9 (12.2)	70.7 (10.5)	76.7 (9.7)
Recall (%)	34.1 (11.8)	34.1 (11.8)	38.8 (7.3)
F1 (%)	45.0 (11.5)	46.0 (12.0)	51.6 (6.1)
Large Earthquakes			
Precision (%)	53.3 (17.9)	63.2 (16.2)	100.0 (0.0)
Recall (%)	57.1 (19.2)	85.7 (13.3)	100.0 (0.0)
F1 (%)	55.2 (11.6)	72.7 (6.1)	100.0 (0.0)
Total			
Accuracy (%)	73.2 (1.5)	74.7 (1.8)	76.9 (1.6)

between normal activity and medium earthquake on the seismometers bag-of-words representation. Table 4 shows the performance of the two baselines and DMSEEW. The traditional seismometer approach is indicated as “Baseline Seismometer” and the combined rule-based approach as “Baseline Combined”. The DMSEEW scores correspond to average results on test sets of the 1-Nearest Neighbors (1NN) trained on the combined bag-of-words representation (GPS and seismometers representations). 1NN outperforms other classifiers (Elastic Net, Support Vector Machine, Random Forest, Extreme Gradient Boosting) on the 3-fold cross-validation.

Our dataset has a class imbalance (normal activity 63%/medium earthquakes 32%/large earthquakes 5%), but it does not affect the detection performance on the least represented class. DMSEEW detects all the large earthquakes (100.0% recall) without false alert (100.0% precision). The 1-NN of DMSEEW is always able to closely match an existing typical distribution of GPS predictions in case of large earthquakes, which allows the correct 1-NN classification. It is critical for an EEW system to detect all the large earthquakes with a precision of 100%. The decisions subsequent to a large earthquake alert imply major mitigation measures for the population possibly impacted. We observe in Table 4 that DMSEEW outperforms both baselines (accuracy score:

76.9% versus 74.7% and 73.2%). Moreover, DMSEEW outperforms both baselines on medium and large earthquakes detection. Figure 3 shows the precision-recall curves of DMSEEW versus both baseline on medium and large earthquakes (second and third plots). Firstly, DMSEEW obtains an average F1-score on test sets for medium earthquakes detection of 51.6% versus 45.0% for the baseline seismometer and 46.0% for the baseline combined. The higher F1-score of DMSEEW on medium earthquakes compared to both baselines is driven by higher performances on both precision and recall (precision: 76.7% versus 65.9% baseline seismometer versus 70.7% baseline rule-based, recall: 38.8% versus 34.1% baselines). Lastly, DMSEEW obtains an average F1-score on test sets for large earthquake detection of 100.0% versus 55.2% for the baseline seismometer and 72.7% for the baseline combined. The higher F1-score of DMSEEW on large earthquakes compared to both baselines is also driven by higher performances on both precision and recall (precision: 100.0% versus 53.3% baseline seismometer versus 63.2% baseline rule-based, recall: 100.0% versus 57.1% baseline seismometer versus 85.7% baseline rule-based). These performances confirm the interest of combining GPS stations and seismometers data to cover the whole spectrum of earthquakes with damaging potential (medium and large). In addition, it reveals the benefit of DMSEEW combined approach instead of the combined sensors (GPS and seismometers) baseline approach that adopts the rule of relative strength.

6 Conclusion

The use of machine learning methods in seismology is still in an early phase. One area of development where it demonstrated promising results is earthquake early warning (EEW), i.e. the characterization of an earthquake before it reaches sensitive areas. Current state-of-the-art methods based on seismometers data only demonstrated an applicability limited to medium earthquakes. In contrast, GPS-based methods are only suitable for large earthquake detection.

We propose DMSEEW, a novel stacking ensemble approach for characterizing the whole spectrum of earthquakes with damaging potential by combining both GPS and seismometer data. Our evaluation on a real-world dataset collected with domain experts demonstrates that the proposed distributed stacking ensemble approach improves the detection of both medium and large earthquakes compared to traditional seismometer only approach and the combined sensors (GPS and seismometers) baseline approach that adopts the rule of relative strength (F1 score: +7% and +6% on medium earthquakes, +45% and +27% on large earthquakes). In addition, DMSEEW detects all large earthquakes with a precision of 100%.

While existing solutions rely on fully centralized processing of the sensor data, our approach assumes distributed data processing based on a geographically distributed cyberinfrastructure. This design significantly reduces the volume of data transmitted in the network, meets the real-time requirements while increasing reliability of the EEW system. With

regards to future work, we plan to simulate different scenarios in an existing EEW execution platform to evaluate DMSEEW response time and robustness. Then, we plan to work on the evaluation of DMSEEW on the other main seismic network through collaboration with the Japanese National research Institute for Earth science and Disaster resilience (NIED).

7 Acknowledgments

This research is supported by the NSF under grants numbers OAC 1640834, OAC 1835661, OAC 1835692, OCE 1745246 and the French ANR OverFlow project (ANR-15-CE25-0003).

References

- Allen, R. M., and Melgar, D. 2019. Earthquake Early Warning: Advances, Scientific Challenges, and Societal Needs. *Annual Review of Earth and Planetary Sciences* 47:361–388.
- Baydogan, M. G., and Runger, G. 2014. Learning a Symbolic Representation for Multivariate Time Series Classification. *Data Mining and Knowledge Discovery* 29:400–422.
- Baydogan, M. G., and Runger, G. 2016. Time Series Representation and Similarity Based on Local Autopatterns. *Data Mining and Knowledge Discovery* 30:476–509.
- Bergstra, J.; Bardenet, R.; Bengio, Y.; and Kégl, B. 2011. Algorithms for Hyper-Parameter Optimization. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*.
- Bergstra, J.; Yamins, D.; and Cox, D. 2013. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In *Proceedings of the 30th International Conference on Machine Learning*.
- Fauvel, K.; Balouek-Thomert, D.; Melgar, D.; Silva, P.; Simonet, A.; Antoniu, G.; Costan, A.; Masson, V.; Parashar, M.; Rodero, I.; and Termier, A. 2019. Earthquake Early Warning Dataset. figshare.
- Fazle, K.; Majumdar, S. and Darabi, H.; and Harford, S. 2018. Multivariate LSTM-FCNs for Time Series Classification. *Neural Networks* 116:237–245.
- Fischer, J.; Redlich, J.; Zschau, J.; Milkereit, C.; and Picozzi, M. 2012. A Wireless Mesh Sensing Network for Early Warning. *Journal of Network and Computer Applications* 35:538–547.
- Hoshiba, M., and Ozaki, T. 2012. Earthquake Early Warning and Tsunami Warning of JMA for the 2011 off the Pacific Coast of Tohoku Earthquake. *Journal of the Seismological Society of Japan* 64:155–168.
- Karlsson, I.; Papapetrou, P.; and Boström, H. 2016. Generalized Random Shapelet Forests. *Data Mining and Knowledge Discovery* 30:1053–1085.
- Kennet, B. L. N. 1991. Iaspei 1991 Seismological Tables. *Terra Nova* 3:122–122.
- Li, Z.; Meier, M. A.; Hauksson, E.; Zhan, Z.; and Andrews, J. 2018. Machine Learning Seismic Wave Discrimination:

Application to Earthquake Early Warning. *Geophysical Research Letters* 45:4773–4779.

Melgar, D.; Bock, Y.; Sanchez, D.; and Crowell, B. W. 2013. On Robust and Reliable Automated Baseline Corrections for Strong Motion Seismology. *Journal of Geophysical Research: Solid Earth* 118:1177–1187.

Melgar, D.; Crowell, B. W.; Geng, J.; Allen, R. M.; Bock, Y.; Riquelme, S.; Hill, E. M.; Protti, M.; and Ganas, A. 2015. Earthquake Magnitude Calculation Without Saturation from the Scaling of Peak Ground Displacement. *Geophysical Research Letters* 42:5197–5205.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Perol, T.; Gharbi, M.; and Denolle, M. 2018. Convolutional Neural Network for Earthquake Detection and Location. *Science Advances*.

Ruhl, C. J.; Melgar, D.; Chung, A. I.; Grapenthin, R.; and Allen, R. M. 2019. Quantifying the Value of Real-Time Geodetic Constraints for Earthquake Early Warning Using a Global Seismic and Geodetic Data Set. *Journal of Geophysical Research: Solid Earth* 124:3819–3837.

Satyanarayanan, M. 2017. The Emergence of Edge Computing. *Computer* 50:30–39.

Schäfer, P., and Leser, U. 2017. Multivariate Time Series Classification with WEASEL+MUSE. *ArXiv*.

Seto, S.; Zhang, W.; and Zhou, Y. 2015. Multivariate Time Series Classification Using Dynamic Time Warping Template Selection for Human Activity Recognition. In *Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence*.

Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; and Xu, L. 2016. Edge Computing: Vision and Challenges. *IEEE Internet of Things Journal* 3:637–646.

Shokoohi-Yekta, M.; Hu, B.; Jin, H.; Wang, J.; and Keogh, E. 2017. Generalizing DTW to the Multi-Dimensional Case Requires an Adaptive Approach. *Data Mining and Knowledge Discovery* 31:1–31.

Tuncel, K. S., and Baydogan, M. G. 2018. Autoregressive Forests for Multivariate Time Series Modeling. *Pattern Recognition* 73:202–215.

Wistuba, M.; Grabocka, J.; and Schmidt-Thieme, L. 2015. Ultra-Fast Shapelets for Time Series Classification. *ArXiv*.

Yang, C.; Raskin, R.; Goodchild, M.; and Gahegan, M. 2010. Geospatial Cyberinfrastructure: Past, Present and Future. *Computers, Environment and Urban Systems* 34:264–277.

Yoon, C.; O'Reilly, O.; Bergen, K. J.; and Beroza, G. C. 2015. Earthquake Detection Through Computationally Efficient Similarity Search. *Science Advances*.