IL NUOVO CIMENTO **47 C** (2024) 265
DOI 10.1393/ncc/i2024-24265-x

COMMUNICATIONS: SIF Congress 2023

# P-wave polarity determination via ensemble deep learning models

G. MESSUTI([1])([2])([*])

([1]) *Dipartimento di Fisica "E. R. Caianiello", Università degli studi di Salerno - Fisciano, Italy*
([2]) *INFN, Sezione di Napoli, Gruppo collegato di Salerno - Salerno, Italy*

received 20 January 2024

**Summary.** — P-wave first-motion polarities play a central role in understanding earth dynamics. Manual or classical automated procedures for determining polarities face several challenges. To address these issues, recent advanced studies leverage deep learning techniques, particularly Convolutional Neural Networks (CNNs). This paper explores the efficacy of ensemble deep learning approach, combining predictions from multiple CNN models. Ensemble methods exhibit improved overall performance and enhanced capabilities in managing waveforms showing no polarity. Additionally, a specific augmentation procedure known as time-shift, enhances the ability to evaluate the uncertainty on noise-only waveforms.

## 1. – Introduction

The determination of P-wave first-motion polarities represents a fundamental aspect of seismology, playing a pivotal role in unraveling the intricacies of earthquake dynamics. The term "first-motion polarity" denotes the initial direction of ground motion recorded by seismometers when P-waves propagate through the Earth. The analysis of P-wave first-motion polarities serves as a key tool for discerning the earthquake's focal mechanism [1, 2]. This process, known as inversion, involves deducing details about the orientation of fault planes. The information derived from this analysis are crucial for understanding the source rupture process and studying the regional stress field [3]. For this reason, having a precise estimation of first-motion polarities is of utmost importance.

Determining P-wave first-motion polarities manually or through classical automated procedures poses several challenges. The process of visually interpreting waveforms and identifying the initial ground motion direction can be subjective, introducing potential inconsistencies and errors, especially when dealing with small earthquakes or low Signal-to-Noise Ratio waveforms, leading to a lack of standardization. On the other hand,

([*]) E-mail: gmessuti@unisa.it

Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0) 

classical automated procedures face difficulties in accurately capturing the complexities of seismic signals [4]. Noisy signals, overlapping events, or variations in local geological conditions can affect the reliability of automated methods, making it challenging to obtain consistent results across diverse seismic events. Furthermore, most automated algorithms rely on different parameters that require *ad hoc* adjustments based on the specific dataset or region to be analyzed.

In recent years, there has been a shift towards utilizing advanced deep learning techniques to improve the accuracy and efficiency of P-wave first-motion polarity determinations. Several studies have shown the remarkable effectiveness of Convolutional Neural Networks (CNNs [5]) for this purpose [4, 6, 7]. CNNs work by using convolutional layers to extract spatial patterns from multidimensional matrices or matrix-like data. In the convolutional process, these layers apply multiple filters with adjustable weights to extract relevant features while scanning the data. The stacking of multiple convolutional layers allows the network to independently learn and recognize abstract features crucial to the task. These modern approaches aim to overcome the limitations of manual and classical automated methods, offering enhanced precision and reliability in processing seismic signals.

This study proposes an ensemble deep learning approach [8], a powerful technique that involves combining predictions from multiple deep learning models to improve overall performance and generalization ability. Ensemble learning is often employed to mitigate overfitting and manage the uncertainty arising from out-of-sample distribution data [9] (*i.e.,*, data deriving from different classes than the training set). The studies presented emphasize that utilizing an ensemble of CNNs improves uncertainty estimation in noise-only waveforms (without seismic events). This improvement arises from the reduced inclination of the ensemble model to assign polarity to waveforms lacking polarity information. Additionally, it is found that a specific augmentation procedure [10], referred to as time-shift, can significantly help to discriminate waveforms with evident polarity from waveforms carrying no polarity information.

## 2. – Methods

**2**˙1. *Dataset.* – In this study, the data originates from INSTANCE [11], an extensive seismic dataset curated specifically for applications in machine learning. The dataset comprises 1.2 million three-component waveforms, capturing data from 54000 earthquakes spanning January 2005 to January 2020. The earthquakes range in magnitude from 0.0 to 6.5. A manual review has been conducted for each earthquake waveform, assigning them P-wave arrival times and either upward, downward, or undefined polarity. The dataset also includes over 130000 noise-only traces. The networks are trained on the vertical component of waveforms with defined polarity (upward or downward) and utilize noise-only traces to evaluate uncertainty in out-of-distribution data.

For training purposes, as is commonly done in machine learning contexts, the dataset consisting of approximately 160000 waveforms with defined polarity, has been divided into three subsets. Specifically, 142000 seismic traces represent the training waveforms, 10000 are assigned to the validation set, and the remaining 8000 to the test set. To double the size of the training data and balance the upward/downward polarity classes, each training waveform is used twice, once in its unmodified form and once flipped upside-down, with the flipped waveform assigned with the opposite polarity, resulting in 284000 waveforms useful for training.

TABLE I. – *Architecture of the convolutional networks used in the study.*

| Layer | Stage | Kernel size | Activation function | Layer shape |
|---|---|---|---|---|
| 0 | Input | – | – | $1 \times 160$ |
| 1 | Conv+Dropout | 5 | ReLU | $32 \times 160$ |
| 2 | Conv+Pooling | 4 | ReLU | $64 \times 78$ |
| 3 | Conv+Pooling | 3 | ReLU | $128 \times 38$ |
| 4 | Conv+Dropout | 5 | ReLU | $256 \times 38$ |
| 5 | Conv+Pooling | 3 | ReLU | $128 \times 18$ |
| 6 | Flatten | – | – | 2304 |
| 7 | Fully connected | – | ReLU | 50 |
| 8 | Fully connected (Output) | – | Sigmoid | 1 |

**2˙2.** *Networks description.* – The architecture of all the networks utilized is identical and is presented in table I. The networks are fed with waveforms sampled at 100 Hz, subsequently demeaned and normalized in the range $[-1, 1]$. The waveforms are provided in time windows of 160 samples. A single network's output is a scalar number lying in the range [0, 1]. The output can be interpreted as the network's estimation of the probability that the input waveform exhibits an upward polarity. Consequently, output values approaching 1 (0) indicate a high level of confidence for the networks to assign an upward (downward) polarity. On the other hand, output values placed towards the center of the interval [0, 1] correspond to waveforms for which networks exhibit uncertainty, *i.e.*, lower confidence in assigning a polarity.

Each network is trained using the Stochastic Gradient Descent (SGD, [12]) algorithm, utilizing the default learning rate, while setting the momentum parameter to 0.8. The batch-size parameter is set to 512. Furthermore, the early stopping technique is employed, which stops the training process if the loss value exhibits no improvement over 10 consecutive epochs.

**2˙3.** *Ensembles construction.* – For this research, each network constituting a single ensemble shares the same architecture, as described in sect. **2˙2**, along with identical training settings. Despite this uniformity, the intrinsic stochastic nature of the training process and the randomness in parameters initialization lead to diversity among the networks. Consequently, when utilized to predict the polarity of the same waveform, they may yield slightly different outcomes. The ensemble method leverages on these diversifications to enhance the overall predictive performance of the model [13].

In this study, the outputs of the networks are aggregated using the Unweighted Model Averaging [8] strategy. It means that for each datum, the output of an ensemble model is given by the mean of the outputs of the N networks that constitute it, as expressed in eq. (1)

$$\frac{1}{N}\sum_{i=1}^{N} O_i, \tag{1}$$

where $O_i$ denotes the output of the $i$-th network and in this study $N = 8$.

Two ensembles of networks are prepared and compared in this analysis. For one of the ensembles, the time windows of 160 samples are centered around the respective

P-wave arrival time specified in the dataset. In the case of the other ensemble, a time-shift augmentation procedure is employed. The time-shift perturbs the centering of time windows contained in the training set, translating the center of an arbitrary number of samples forward or backward with respect to the declared arrival time. Specifically, half of the 284000 training waveforms (comprising original traces and the flipped ones) are selected, and two independent random values of time-shift are applied to each of them. The first value is selected from the range $[-10, -1]$, and the second from $[1, 10]$. The original waveform and the two shifted versions are then included in the training set.

## 3. – Results

As previously stated, trainings are performed on data with defined polarity, and networks constituting the same ensemble are trained independently. Once the training of each network that constitutes an ensemble is complete, the ensemble model can be evaluated both on the test set and on the noise-only waveforms.

**3˙1.** *Performances comparison.* – The test set comprises approximately 8000 defined polarity waveforms. The generic $i$-th waveform is assigned with a true label $y_i$ (either 0 or 1 to distinguish downward and upward polarities). After the evaluation of a model, we can also assign a prediction $p_i$, represented by a number between 0 and 1. Subsequently, a confidence threshold can be set to accept the suggested predicted polarity. The performances of the ensemble models can be evaluated varying the threshold, in terms of precision and recall.

Precision, defined in eq. (2a), represents the accuracy rate of the models' predictions, while recall, defined in eq. (2b), represents the fraction of recognized polarities,

$$Precision = \frac{TP}{TP + FP}, \tag{2a}$$

$$Recall = \frac{TP}{TP + FN}. \tag{2b}$$

In eq. (2), $TP$ stands for True Positive and refers to correct predictions; $FP$ is False Positive, indicating waveforms for which the polarity assigned by the model is wrong; $FN$ is False Negative, referring to waveforms for which the model either assigns an incorrect polarity or does not assign any polarity.

Figure 1 displays the performances of the models, presenting curves of precision as a function of recall. The better a model performs, the closer its precision-recall curve will be to the top-right corner. Comparing the curves, one can conclude that ensemble models slightly outperform the typical respective single network model. It is worth noting that precision values in each case exceed 0.975, indicating high reliability for each model. Models with no time-shift show better precision-recall curves in any case.

**3˙2.** *Distributions of the predictions.* – In this section we analyze the distribution of predictions made by the models. To achieve this, it is useful to consider the outputs of each ensemble and construct relative histograms of predictions. Figure 2 shows the histograms for the predictions on the defined polarity test set. It is evident that the models place most predictions towards the extremes of the interval [0, 1], resulting in high confidence among the models in assigning a polarity to waveforms. Additionally, it is worth noting that there are no substantial differences among the various distributions.

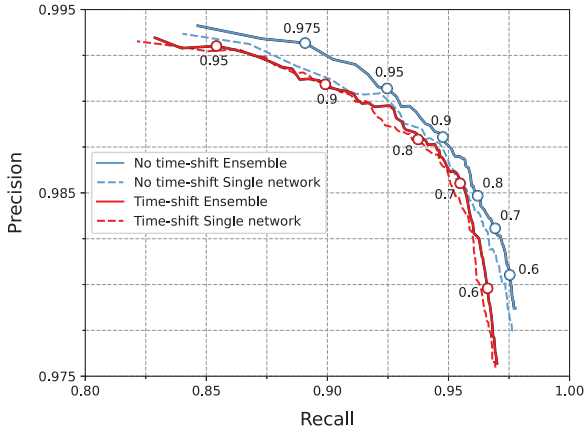


Fig. 1. – Precision-recall curves. The red solid and light blue solid lines represent ensemble models trained with no augmentation procedure and with time-shift augmentation, respectively. The dashed lines depict the typical performances of single networks. Circles correspond to fixed confidence thresholds. A threshold value $\theta \in (0.5, 1)$ indicates that all output values above $\theta$ are accepted as upward predictions, all values below $1 - \theta$ are accepted as downward predictions, and other values interpreted as no polarity prediction.

Now we focus our attention on the behavior of ensembles when dealing with noise-only traces. For this purpose, a time window of 160 samples is randomly selected for each of the 130000 INSTANCE waveforms of noise, resulting in 130000 input data used to feed the models. Optimal behavior would be reflected in perfect uncertainty by models, resulting in an output equal to 0.5 for each noise-only waveform. Figure 3 shows histograms of the predictions corresponding to noise-only waveforms. The results suggest that ensemble methods can enhance the ability to manage waveforms with no polarity, as in each right panel of fig. 3 the ensembles place fewer data in the bins at the edge, in favor of the central ones, compared to the respective panel on the left. In addition, it is evident that the time-shift augmentation procedure causes the distributions of outputs on data

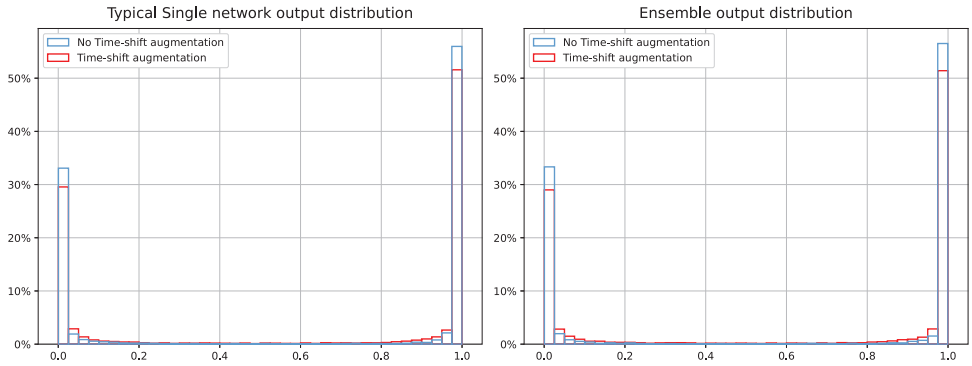


Fig. 2. – Histograms of the distributions of models' outputs when applied to defined polarity test set. On the left, typical behaviors of single networks are shown, while on the right, outputs of ensembles are displayed.

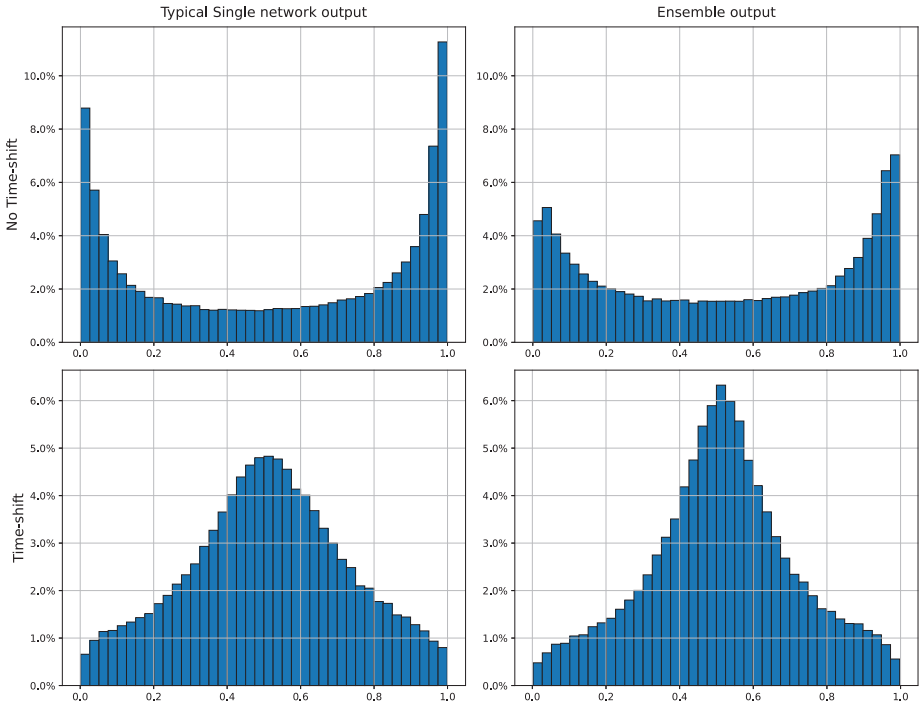


Fig. 3. – Histograms showing the distributions of models' outputs when applied to noise-only waveforms. The first row refers to the training session without the augmentation procedure, while the second row represents the session with time-shift augmentation. On the left, typical behaviors of single networks are displayed, while on the right, outputs of ensembles are shown.

lacking polarity to improve and significantly differ from the distributions of predictions on data with defined polarity.

## 4. – Conclusion

This study highlights the robustness and versatility of ensemble deep learning models in the context of P-wave first motion polarity determination. The observed enhancements in both precision-recall metrics and output distribution characteristics underscore the potential of ensemble methods to address challenges associated with seismic waveform analysis. Additionally, the time-shift procedure is investigated. Models with no time-shift augmentation exhibited superior precision-recall curves. Despite this, several advantages can derive from the time-shift procedure. For example, as noted in [6], greater robustness is achieved when dealing with waveforms with a reduced precision in arrival times. This study highlights the significant role of time-shift in enhancing the models' ability to manage waveforms with no discernible polarity, as explained in sect. **3**˙2. This property becomes particularly advantageous when models are tested on a mixture of waveforms belonging to both classes. In conclusion, the ensemble methods, complemented by the integration of time-shift, prove to be a valuable asset for effectively managing waveform variability, thereby enhancing the overall reliability and effectiveness of seismic waveform analysis.

REFERENCES


[1] Reasenberg P. A., Report 85-739 (U.S. Geological Survey) 1985.
[2] Snoke J. A. *et al.*, *FOCMEC: Focal mechanism determinations*, in *International Handbook of Earthquake and Engineering Seismology*, edited by Lee W. H. K. (Academic Press, San Diego) 2003, pp. 1629–1630.
[3] Vavryčuk V., *Geophys. J. Int.*, **199** (2014) 69.
[4] Ross Z. E. *et al.*, *J. Geophys. Res.*, **123** (2018) 5120.
[5] LeCun Y. *et al.*, *Nature*, **521** (2015) 436.
[6] Messuti G. *et al.*, *Front. Earth Sci.*, **11** (2023) 1223686.
[7] Uchide T., *Geophys. J. Int.*, **223** (2020) 1658.
[8] Ganaie M. A. *et al.*, *Eng. Appl. Artif. Intell.*, **115** (2022) 105151.
[9] Lakshminarayanan *et al.*, *Simple and scalable predictive uncertainty estimation using deep ensembles*, in *Advances in Neural Information Processing Systems*, edited by Von Luxburg U. *et al.* (NeurIPS, Long Beach, CA) 2017.
[10] Alomar K. *et al.*, *J. Imaging*, **9** (2023) 46.
[11] Michelini A. *et al.*, *Earth Syst. Sci. Data*, **13** (2021) 5509.
[12] Bottou L., *Stochastic gradient descent tricks*, in *Neural Networks: Tricks of the Trade: Second Edition*, edited by G. Montavon *et al.* (Springer, Berlin, Heidelberg) 2012, pp. 421–436.
[13] Dietterich T. G., *Ensemble methods in machine learning*, in *International Workshop on Multiple Classifier Systems*, edited by Kittler J. and Roli F. (Springer) 2000, pp. 1–15.