doi:10.26443/seismica.v4i2.1655

Masked graph neural network for rapid ground motion prediction in Italy

Danele Trappolini * [1], **Ilaria Oliveti** [2], **Licia Faenza** [3], **Dario Jozinović** [4], **Alberto Michelini** [2]

[1]Department of Computer, Control and Management Engineering, Sapienza Università di Roma, Via Ariosto, 25, 00185 Rome, Italy, [2]Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti, Via di Vigna Murata 605, 00143 Rome, Italy , [3]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy, [4]Swiss Seismological Service (SED), ETH Zurich, Zurich, Switzerland

Author contributions: *Conceptualization*: Alberto Michelini, Dario Jozinović, Daniele Trappolini, Licia Faenza, Ilaria Oliveti. *Methodology*: Daniele Trappolini, Ilaria Oliveti, Dario Jozinović. *Software*: Daniele Trappolini. *Formal Analysis*: Ilaria Oliveti, Licia Faenza. *Investigation*: Ilaria Oliveti, Licia Faenza. *Resources*: Ilaria Oliveti, Licia Faenza. *Writing - Original draft*: Daniele Trappolini, Ilaria Oliveti. *Writing - Review & Editing*: Alberto Michelini, Licia Faenza, Dario Jozinović, Daniele Trappolini, Ilaria Oliveti. *Visualization*: Daniele Trappolini, Ilaria Oliveti. *Supervision*: Alberto Michelini. *Project administration*: Alberto Michelini, Licia Faenza. *Funding acquisition*: Alberto Michelini, Licia Faenza.

Abstract This study presents an updated version of TISER-GCN, a graph neural network (GCN) designed to predict maximum intensity measurements (IMs) from 10-second seismic waveforms starting at the earthquake origin time, without prior knowledge of location, distance, and magnitude. The improved model was applied to nearly 600 seismic stations from the INSTANCE benchmark dataset, significantly expanding the original TISER-GCN setup, which was limited to 39 stations in a smaller area of central Italy. Input data consist of three-component waveforms selected to ensure high quality and minimize saturation. Results show that masking stations where the P-wave arrives within the first 10 seconds , combined with the integration of additional information, reduces the mean squared error (MSE) by up to 6% for peak ground acceleration (PGA) and 5.5% for peak ground velocity (PGV), compared to the unmasked baseline. Moreover, the proposed approach yields near-zero median residuals across all IMs, mitigating the systematic underestimation observed when using a ground motion model specifically developed for Italy. These findings indicate that the model provides accurate predictions of ground motions, comparable to those obtained with the original TISER-GCN, which, however, requires a fixed seismic network geometry.

Production Editor: Gareth Funning
Handling Editor: Carmine Galasso
Copy & Layout Editor: Théa Ragon

Received: March 19, 2025
Accepted: August 11, 2025
Published: September 15, 2025

1 Introduction

The rapid and accurate assessment of earthquake-generated ground shaking can significantly improve emergency response times and public awareness, mitigating the impact of seismic events. As a core topic in seismology and earthquake engineering, it focuses on estimating the intensity or time series of shaking at a target location. Ground-motion prediction techniques are essential for seismic hazard assessment, early warning, and post-event damage evaluation.

In recent years, artificial intelligence (AI) approaches have shown significant potential in enhancing early warning systems by improving the speed and accuracy of ground motion predictions. Böse et al. (2012) applied artificial neural networks (ANNs) to estimate key seismic parameters, including peak ground velocity (PGV), epicentral distance, and earthquake magnitude using the acceleration, velocity, and displacement waveforms. In a complementary approach, Hsu et al. (2013) utilized support vector regression (SVR) to predict peak ground acceleration (PGA) by extracting six critical features from the initial P-wave's vertical component, such as peak acceleration, peak velocity, peak displacement, effective predominant period, cumulative absolute velocity, and the integral of the square velocity. Gandomi et al. (2011) used a hybrid method coupling genetic programming and orthogonal least squares to model ground motion prediction from source parameters, paving the way for more sophisticated approaches like those explored by Derras et al. (2014), who integrated ANNs for enhanced ground shaking predictions. Hu et al. (2023) employed Random Forest, which uses an ensemble of decision trees, to predict ground-motion intensity by analyzing the initial P-wave data, taking advantage of the model's ability to handle noisy data and avoid overfitting. Iaccarino et al. (2024) used Gradient Boosting Regressor (GBR) to predict PGA, and hypocentral distance starting from P-wave features.

In particular, Machine Learning (ML) has been applied to the prediction of the final ground-motion intensity at a target site using the first few seconds of its observations (e.g., Spallarossa et al., 2019). Otake et al. (2020) used a deep learning model based on surrounding stations to improve real-time intensity prediction. Jozinović et al. (2020, 2022) and Zhang et al. (2022) applied Convolutional Neural Networks (CNNs) for the rapid prediction of earthquake ground shaking intensity using only the first few seconds of raw waveform data. Fornasari et al. (2022) combined CNNs

*Corresponding author: daniele.trappolini@uniroma1.it

with Voronoi tessellation to reconstruct ground shaking fields. CNNs are well-known for their ability to extract meaningful patterns directly from raw data, without relying on handcrafted features. This makes them particularly effective for processing seismic waveforms, which are often complex and noisy. CNNs can learn time-frequency features linked to shaking intensity, enabling fast and accurate predictions of ground-motion parameters, even without prior knowledge of source location or magnitude. Münchmeyer et al. (2021) developed a transformer-based earthquake alert model, a multi-station approach that predicts final PGA values at multiple target stations using initial raw records at multiple input stations. Wang et al. (2022, 2023) demonstrated the power of Long Short-Term Memory (LSTM) networks, to improve the prediction process by capturing both short-term changes and longer-term trends in the seismic data, ensuring that the model continuously refines its predictions as more data becomes available. Graph Convolutional Networks (GCNs) have recently emerged as a powerful tool for modeling spatially distributed systems, including seismic networks. Unlike traditional machine learning models that treat input stations independently, GCNs leverage the spatial relationships between stations by representing them as nodes in a graph, where edges encode physical proximity or other meaningful connections. This structure allows the model to integrate information across neighboring stations, enabling more spatially coherent and accurate predictions. Moreover, GCNs are inherently robust to missing data and can generalize across varying station geometries, making them highly suitable for real-time applications such as earthquake early warning and rapid response systems. In the context of earthquake ground motion prediction, Bloemheuvel et al. (2023) demonstrated that GCNs can outperform conventional methods in both speed and accuracy, by using early waveform data to estimate shaking even at stations where the peak ground motion has not yet been recorded. In this context, we start from the work by Bloemheuvel et al. (2023) as a departure point and illustrate how to go beyond the limitation of a fixed set of seismic stations that needs to be defined at training time, using an updated GCN model for our task. Our implementation uses a dynamic masking mechanism, first introduced by Yang et al. (2019) in the context of image recognition with incomplete inputs. Instead of propagating all node attributes, the Masked GCN selectively propagates only a subset, controlled by an event-based mask assigned to each node. A binary mask tensor dynamically determines which stations contribute to the prediction, based on whether the P-wave arrives within a 10-second window from the first triggered station. This selective inclusion ensures that only early, informative waveform data are used, thereby reducing noise and improving prediction accuracy. The masking mechanism also allows the model to adapt to changes in network geometry—stations can be added, removed, or deactivated without retraining—thus overcoming the fixed-graph limitations of standard GCNs and enhancing the model's robustness and scalability for earthquake early warning applications. A more detailed explanation of the masking strategy is provided in Section 3.2.

Our work articulates along the following main steps:

1. We build the dataset using a multi-step, pre-processing approach suitable for real-time applications to identify high-quality waveforms and improve the predictive performance of the model.

2. We present a novel architecture utilizing convolutional and graph convolutional layers to perform multivariate regression on time series originating from graph-structured data.

3. We evaluate our updated model thoroughly on the pre-processed dataset featuring 565 stations, evidencing the generality and potential of the proposed GCN-based architecture in this task. We discuss our results in detail and perform a Masked GCN to prioritize stations where the P wave arrives within the first 10 seconds.

4. We systematically evaluate the capabilities of our model through comprehensive experiments, incorporating additional information (e.g., station distances to a reference station) into the final fully connected layer of our proposed workflow.

5. Finally, we compare our results with predictions generated using the Ground Motion Model (GMM) by Bindi et al. (2011), calibrated for Italy, which requires an earthquake location and magnitude as input.

2 Data

The input data consist of three-component earthquake waveform traces recorded in Italy by the Italian National Seismic Network (Michelini et al., 2016; Margheriti et al., 2021). The target data comprise intensity measures (IMs) associated with each recording: peak ground acceleration (PGA), peak ground velocity (PGV), and spectral acceleration (SA) at periods of 0.3, 1, and 3 seconds: SA(0.3), SA(1.0), and SA(3.0). In our approach, the waveform inputs serve as the foundational data for the GCN model, designed to capture complex spatial and temporal patterns within seismic signals. Through this model, we aim to provide accurate predictions of ground motion intensities at stations (locations) further away from the epicentre of the earthquake.

The waveform data were downloaded from the INSTANCE dataset (Michelini et al., 2021) and include both HN$*$ and HL$*$ (acceleration) and HH$*$ and EH$*$ (velocity) channels, where $* \in$ [E,N,Z]. All the waveform traces have a length of 120 s, are sampled at 100 Hz, and are provided in ground motion physical units after deconvolution of the instrument transfer functions. We selected earthquakes in the same area of central Italy (bounded by latitude [42°, 43.75°] and longitude [12.3°, 14°]) used by Jozinović et al. (2020, 2022); Bloemheuvel et al. (2023). The earthquakes occurred between January 2005 and January 2020. Using these criteria, 975 earthquakes with magnitude M $\geq$ 3 and depth $<$ 30 km have been used (Figs 1, 2).

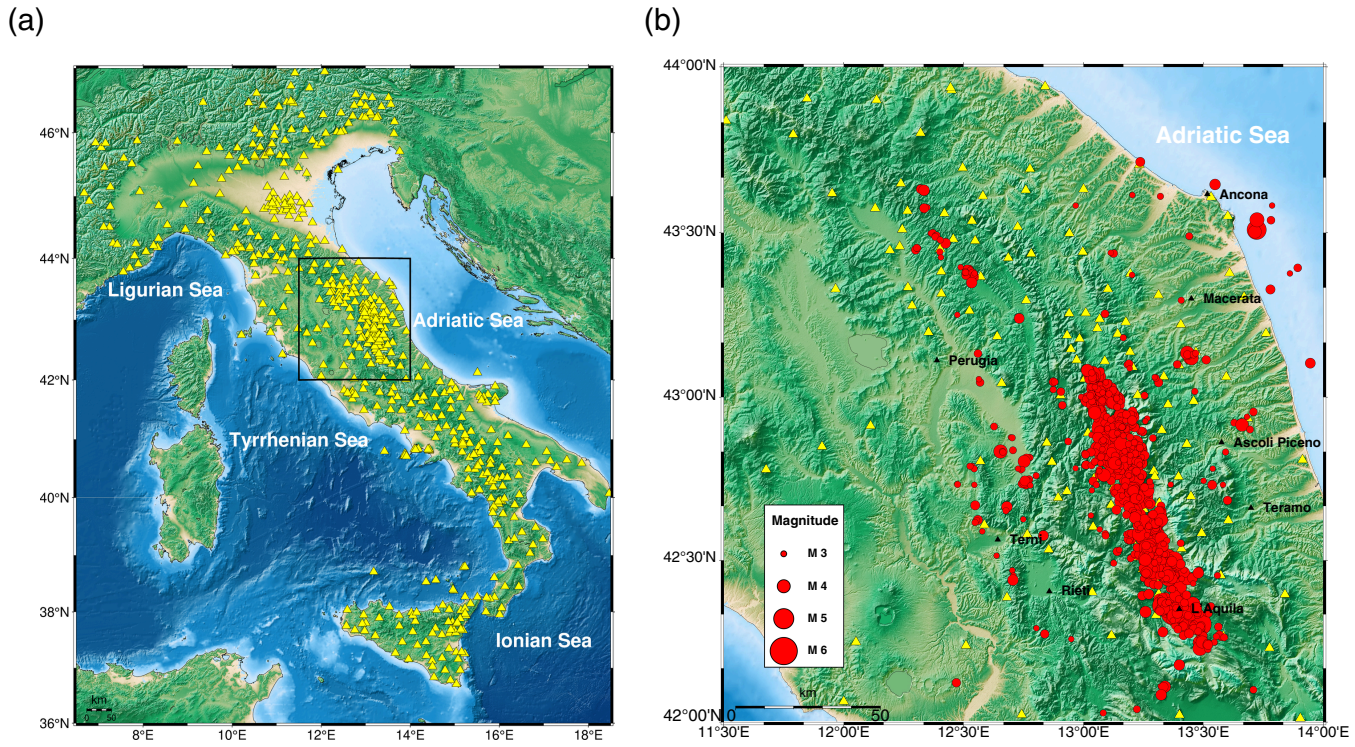


Figure 1 Spatial distribution of (a) the 565 stations (yellow triangles) together with (b) the 975 earthquakes (red dots) used in this study. We have selected earthquakes with magnitude M $\geq$ 3 and depth $<$ 30 km occurred in the study area bounded by latitude [42°, 43.75°] and longitude [12.3°, 14°] between January 2005 and January 2020.

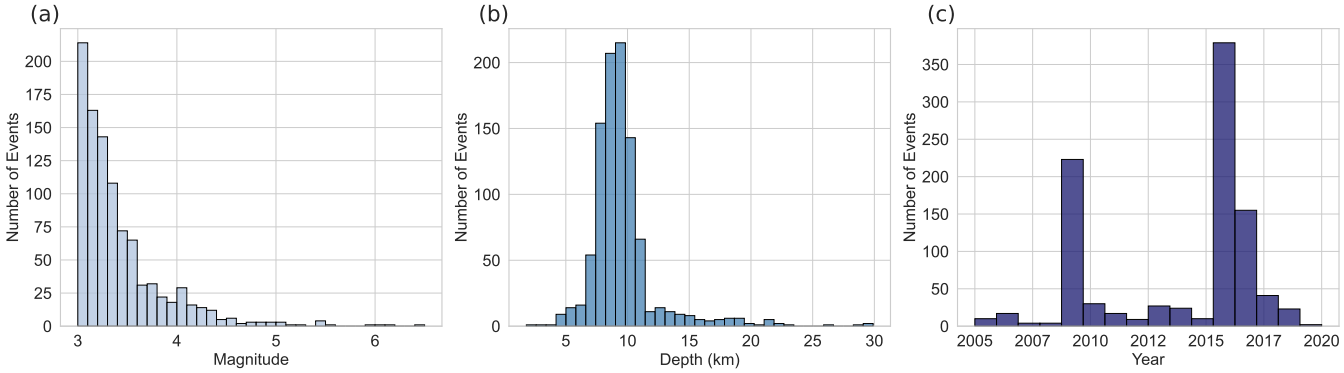


Figure 2 Distribution of earthquake (a) magnitudes, (b) depths and (c) years for the selected events.

2.1 Quality Control and Filtering

To ensure the selection of high-quality waveform records, we applied a set of selection criteria to the metadata provided in INSTANCE, focusing on both accelerometric and velocimetric data characteristics:

- For acceleration data: we retained only traces with PGA values exceeding $10^{-3} ms^{-2}$. This threshold filters out records with extremely low accelerations, minimizing noise-dominated traces.

- For velocimetric Data Acquired by broadband sensors with 24-bit Data Loggers: we included only traces that featured maximum values within ±7,000,000, 17% below full scale, to reduce the risk of clipping and account for possible amplification or preprocessing issues. Additionally, traces with derived acceleration exceeding 100 cms^{-2} were excluded to filter out implausible or distorted records. These conservative thresholds enhance data reliability and improve model stability by limiting the influence of noisy inputs.

- For co-located accelerometric and velocimetric data, and for lower values of ground motions selected as explained in the previous steps, we included only the velocimetric data. Co-located instruments allowed us to do a further layer of quality control - excluding large outliers, which we defined as data where the PGA ratio between the accelerometric and velocimetric data fell outside the range [0.5,2].

- For data from non-co-located Instruments: to determine which non-colocated instrument data should be prioritized, we used the Bindi et al. (2011) GMM, comparing observed PGA values against

predictions. We calculated the difference (i.e. $\log_{10} \frac{PGAobs}{PGApred}$) and selected the instrument data with the observed value closest to this prediction. Specifically, given that Bindi's standard deviation is approximately 0.35 $\log_{10} PGA$, we excluded data for which $\left|\log_{10} \frac{PGAobs}{PGApred}\right| > 0.7$, as they were likely inconsistent or of lower quality. This filtering approach enhanced the dataset's reliability and alignment with expected ground motion.

Following the above-mentioned criteria, a subset of 565 stations has been selected (Fig 1). The dominant portion of the data ($\sim$ 99%) has been acquired by the Italian National Seismic Network (IV code) and by the MedNet (MN code), both operated by INGV (Michelini et al., 2016; Danecek et al., 2021). The remaining records included in the dataset come from the Emersito Seismic Network (XO code), which was deployed in Amatrice and surrounding areas (Central Italy) to conduct field surveys shortly after the mainshock event on August 24th, 2016 (Mw 6.0).

2.2 Preparation for Model Training

The data preparation and processing for the training phase followed the criteria outlined in Jozinović et al. (2020, 2022). Based on their findings, a 10-s window offers a good trade-off between prediction accuracy and timeliness, which we also adopted. Specifically, to align waveforms, all records were trimmed to a 10-s window starting from a reference time defined as 1 s before the P-wave arrival at the first recording station for each earthquake. Velocity traces were then differentiated to obtain acceleration.

Since recordings are not available across the entire national network for each earthquake, waveform data were entirely missing for many stations. These waveforms were filled with zeros following the approach of Jozinović et al. (2020, 2022). This solution required overcoming technical challenges to adapt to a sparse graph with null values for numerous stations, marking a substantial advancement over the previous work by Bloemheuvel et al. (2023), which was limited to a fixed network of only 39 stations. When the waveforms (i.e. the model input) for a station were missing, the missing intensity measures (i.e. the model output) were estimated using the USGS ShakeMap software with the latest configuration for Italy (Michelini et al., 2020) to ensure complete output data (target variables), following Jozinović et al. (2020, 2022). Although using ShakeMap estimates introduces additional assumptions into the model, we consider it the most effective method to handle missing data, allowing the model to generate approximations for a site even in the absence of input data—an important feature, for instance, in EEW applications. As a result, the target values are composed of 11% observed data and 89% ShakeMap-calculated data (a consequence of INSTANCE having 21 3-C waveforms per earthquake on average, most of them from stations closest to the epicenter). ShakeMap is a widely used software system that generates interpolated ground motion values by combining recorded data and/or macroseismic information, GMM-based estimates, and seismological parameters such as event location, magnitude, and finite-fault models (when available). As such, it may introduce systematic biases—particularly in areas with sparse station coverage, near-fault regions, or complex site conditions. These biases can propagate into the model during training, potentially affecting its ability to generalize to purely observational data. Nevertheless, the results presented in Section 4 — and in particular in Section 4.2 — confirm that using ShakeMap-derived data as targets for the majority of the training set does not negatively affect the performance of the proposed model.

Following Jozinović et al. (2020), the waveform data are normalized by the input maximum (i.e. the largest amplitude observed across all stations within the time window), and this maximum is saved as the normalization value which is later inserted into the final fully connected layer of our proposed Masked GCN model.

3 Method and Training

3.1 Model Proposed

The Masked GCN model used here is a neural network architecture that combines CNNs and GCNs to capture both spatial and temporal information. Specifically designed for seismic signal analysis across a network of stations, the model predicts earthquake IMs (PGA, PGV, SA(0.3), SA(1.0) and SA(3.0)) at recording stations using multistation waveforms.

The architecture is based on Bloemheuvel et al. (2023), with several modifications, the most relevant being the introduction of a deeper network (i.e., an increased number of layers) and the implementation of the dynamic masking mechanism within the GCN component (Fig 3). Specifically, the model uses three primary components: convolutional layers, graph convolutional layers, and fully connected (FC) layers. The convolutional layers, designed to process 1D wave input data, consist of three layers with increasing output channels (32, 64, and 128). Each convolutional layer uses a kernel size of (1, 125) and a stride of (1, 2) to extract temporal features, followed by a ReLU activation function. The output from the convolutional layers is then reshaped for subsequent graph processing. The graph convolutional layers leverage spatial relationships across stations, with the first GCN layer projecting the input features into a 64-dimensional space, with the second layer further refining the learned representation.

In particular, spatial information is provided to the model through an adjacency matrix, which represents a graph structure based on interstation distances. The adjacency matrix was computed following the methodology described in Bloemheuvel et al. (2023). All pairwise geodesic distances between stations are computed, then normalized and inverted so that closer station pairs receive higher edge weights. To control graph sparsity, a threshold of 0.3 is applied to remove weak (i.e., distant) connections. This means that all edges with a weight below this threshold are set to zero in

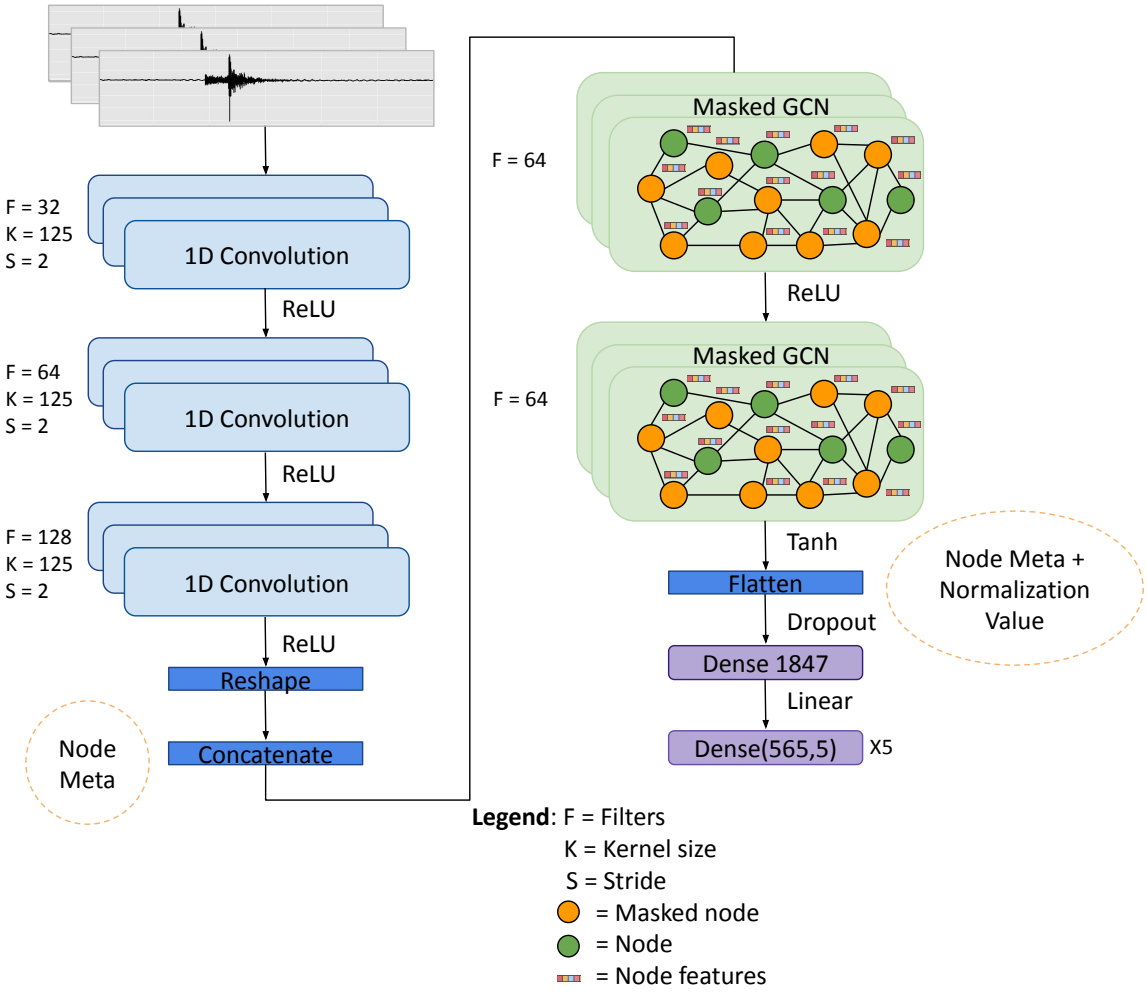


Figure 3 The architecture of the proposed model for IM prediction. Boxes shaded grey represent inputs, shaded light blue represent convolutional layers, shaded blue represent operators, shaded green represent graph convolutional layers and shaded violet represent fully connected layers.

the adjacency matrix, directly affecting the final graph topology. The resulting adjacency matrix is fixed for our network configuration, ensuring a consistent spatial structure while allowing event-specific features to vary at the node level, as explained in detail in Section 3.2. The output of the GCN layers is passed through a "tanh" activation function to introduce non-linearity and bound the values within [-1,1]. After the GCN layers, the network output is flattened and combined with additional information—including the maximum amplitude value—in a fully connected layer. We show that adding this knowledge in the neural network improves the results (see Section 4.1), consistent with the findings of Jozinović et al. (2022). In this work, we consider the maximum amplitude value from the vertical component waveforms recorded across all stations within the time window. In parallel, we verified that similar results were obtained when using the maximum amplitude across all three components (vertical, north–south, and east–west), which is the solution to be adopted in the operational phase. Finally, the last fully connected layer produces arrays of size (565,5), where 565 is the number of stations, and 5 is the number of predicted IMs per station (PGA, PGV, SA(0.3), SA(1.0) and SA(3.0)). The base-10 logarithm has been applied to all the IMs (i.e. $\log_{10} IM$).

The dataset comprises 975 seismic events, each recorded across up to 565 stations (but 21 stations on average), resulting in a total of 35,078 waveform instances. The data were randomly split into a training set (80%) and a test set (20%). Model training was conducted using a 5-fold cross-validation approach, where 1/5 of the training set was used for validation in rotation. The model was trained for 100 epochs. ReduceLROnPlateau learning rate scheduler, starting at 0.0001, was employed to dynamically adjust the learning rate during training. The reduction is applied by a factor of 0.5 if the validation loss does not improve for three consecutive epochs. The mean squared error (MSE) loss function was used for model optimization, with the RMSProp optimizer. To mitigate overfitting, an additional dropout layer with a rate of 0.4 was applied before concatenating the maximum amplitude. Training was performed on an NVIDIA RTX 4090 GPU, utilizing approximately 9.2 GB of memory, and required a total of 1 hour and 12 minutes to complete across all 5 folds. In an operational setting, during testing, the model predicts each feature in 1 second, leading to a total of 5 seconds for all

5 features across 195 traces, excluding dataset loading time. Additionally, in a real-time deployment, the system would need to integrate a model such as PhaseNet (Zhu and Beroza, 2019) or STA/LTA method for detecting the first P-wave arrival. Then, our pipeline incurs a cost of 1.06 seconds for data preprocessing and 0.7 seconds for the prediction for a single event, resulting in a total computation time of approximately 1.76 seconds per event.

3.2 Mask Method

The Masked GCN model utilizes a dynamic masking mechanism to determine which stations contribute to the model's predictions, enabling flexible input data configurations and enhancing real-time adaptability for seismic monitoring. The mask is implemented as a tensor, where each value indicates whether the input data from a given station should be included (1) or excluded (0) (Yang et al., 2019). This mask is applied to the waveform data prior to their input into the model and during loss computation. This mask is designed to include only waveforms where the P-wave arrives within the first 10 seconds starting from the P arrival at the first recording station. During the training phase, we used the manually picked P-wave arrival times from the INSTANCE metadata for each station. This approach minimizes the inclusion of noisy or irrelevant signals, allowing the model to focus on meaningful seismic data. Conversely, in the testing phase, we computed the travel times assuming a P-wave velocity of 6.5 kms^{-1} based on the P arrival at the closest station provided in the INSTANCE metadata for each earthquake. A more detailed explanation of the mask's functionality is provided in Figure S1. By restricting the input to this 10-second window, the mask increases the likelihood of capturing the P-wave arrival, and often both the P- and S-waves of the stations close to the hypocenter, at each station within this timeframe. This targeted selection strategy not only enhances the quality of input data by prioritizing critical early-wave information but also optimizes the model's responsiveness in real-time scenarios. In practical terms, this procedure would be based on the station with the first trigger which should be a good approximation of the station closest to the epicenter (and the epicenter itself in the regions with moderate-to-high network density), especially for earthquakes with M$\geq$ 3 (used in this study).

The masking process is dynamic and operator-adjustable, allowing the system to adapt to network changes, such as the addition or deactivation of stations. If a station goes offline or becomes unreliable, its mask value can be set to [0], effectively excluding it from the model's predictions without requiring structural modifications or retraining. This flexible masking strategy overcomes a common limitation of graph convolutional networks, which typically struggle to handle changes in node structures. By selectively masking inactive nodes, the model maintains a consistent graph structure, ensuring seamless adaptation to variations in the operational network while preserving the reliability and relevance of its predictions.

4 Results

4.1 Evaluating the Impact of Additional Knowledge

In the first part of this study, we explored the additional knowledge needed to make reliable predictions of the maximum values of the IMs. To assess the effectiveness of our model, we compared the training performed using the (i) coordinates of the first station that records the P wave, (ii) maximum amplitude information and (iii) interstation distances (i.e., the distance matrix) as additional metadata inserted into the last fully connected layer. These three options were added in varying combinations to create eight different configurations, allowing us to evaluate the individual and combined impact of each metadata type on model performance. More specifically, in the first configuration, neither the coordinates of the first P-wave station nor the interstation distances were incorporated, while the normalized amplitude value was included. In the second configuration, all three additional features were included. In the third configuration, all metadata except the normalized amplitude value were integrated. The same approach was applied to the remaining configurations, ensuring a full exploration of all eight possible combinations. A complete representation of the eight configurations and the results of all the experiments is shown in Table S1.

For each experiment, the Masked GCN model results from all five-fold cross-validation test sets are averaged. The best results were achieved when all information, except the coordinates of the first P-wave station, was included (i.e., Experiment 6 for SA(0.3) and SA(1.0)) or when only the maximum amplitude information was integrated in the model (i.e., Experiment 1 for PGA, PGV and SA(3.0)). The mean squared errors (MSE) were: 0.188 for PGA, 0.199 for PGV, 0.181 for SA(0.3), 0.229 for SA(1.0) and 0.267 for SA(3.0). Remarkably, when the third feature is also added, model performance deteriorates (Table S1) compared to that observed in Experiments 1 and 6. This result might originate from the redundancy of information inserted in the model. That is, the knowledge conveyed by the coordinates of the first station that records the P wave is likely captured already by the adjacency matrix and the interstation distances. Finally, we evaluated the performance of the model's best configuration with and without applying the aforementioned mask, which focuses on stations where the P-wave arrives within the first 10 seconds. The results indicate that the MSE of the Masked GCN model for the 5 IMs is lower than that of the model without the mask (i.e., 0.203 for PGA, 0.210 for PGV, 0.194 for SA(0.3), 0.243 for SA(1.0) and 0.278 for SA(3.0)). This outcome is expected, as the selection of good quality data recorded near the epicenter improves the accuracy of our predictions. It is worthwhile to stress here that the methodology does not avail of any explicit hypocenter location and magnitude input information. Overall, Experiment 1 was found to be the most effective configuration for all five IMs and was adopted for our model. Furthermore, we conducted an additional experiment using PhaseNet as the pretrained encoder, selected for its greater ca-

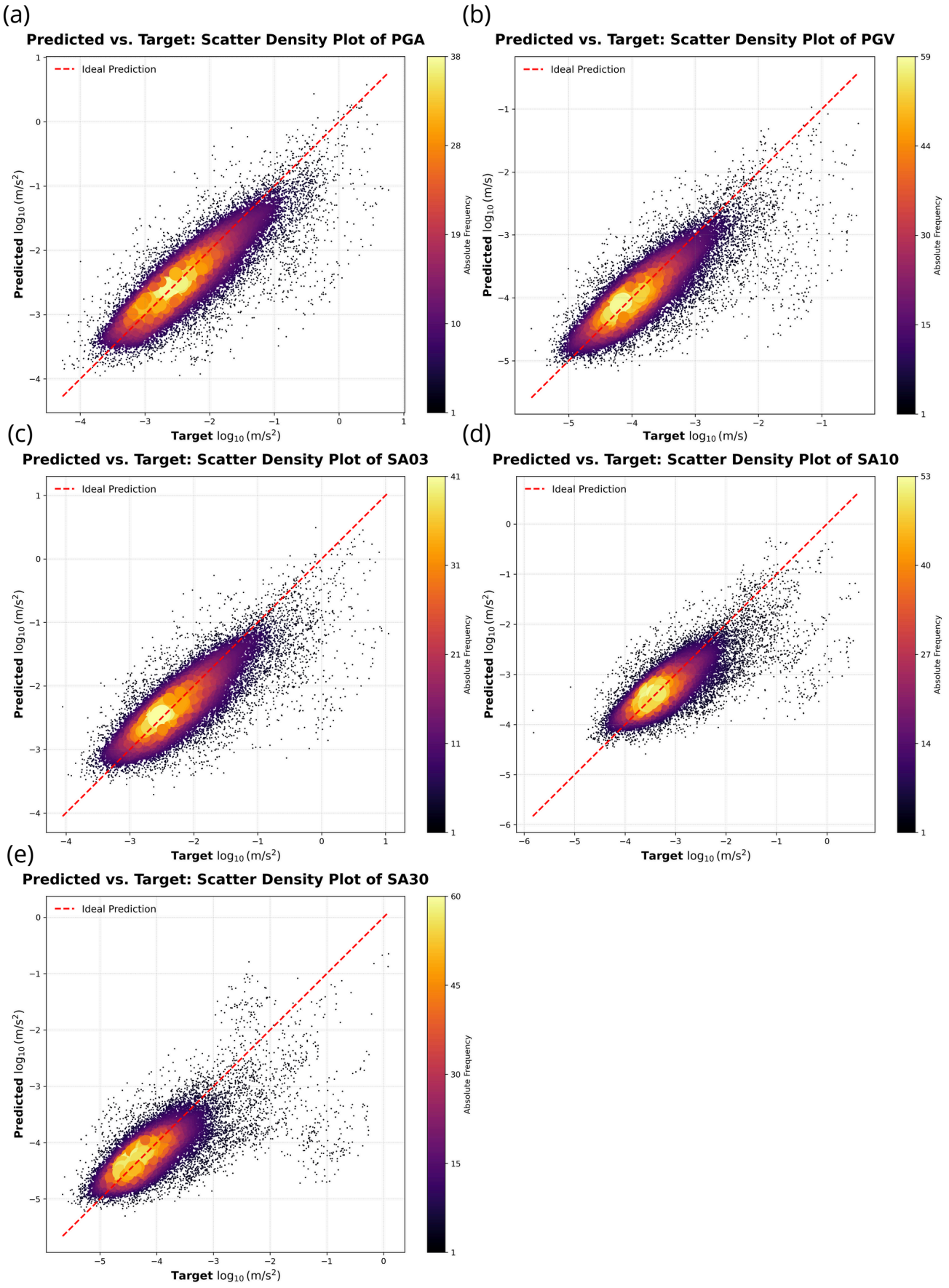


Figure 4 Observed versus predicted scatterplot of the testing dataset for: (a) PGA, (b) PGV, (c) SA(0.3), (d) SA(1.0) and (e) SA(3.0). Scatter points are both color- and size-coded, with black indicating low-density areas and yellow representing high-density areas. The colorbar shows the total number of data points in each area.

pacity and training on the STEAD dataset via SeisBench. Initializing the encoder with these weights improved performance by reducing MSE by 2.7% for PGA, 5.5% for PGV, and 4.1% for SA(0.3) compared with Experiment 1, and lowered GPU memory usage from 9.2 GB to 2.58 GB. We will further investigate this approach and evaluate alternative pretrained temporal encoders and techniques in future work.

4.2 Model performance

Considering the results of the Masked GCN model on each individual earthquake and each IM metric, Figure 4 displays scatter density plots comparing predicted versus target values of the test set. The analysis also includes ShakeMap estimates. In Figure 4 the x-axis represents the target values in on a base-10 logarithmic scale, while the y-axis shows the predicted values, also on a base-10 logarithmic scale and in the same units. A red dashed line indicates the ideal prediction line, where the predicted values match the target values. The scatter points are coded by color and size to represent data density, with smaller dots indicating low-density areas and larger dots highlighting high-density areas. The colorbar, ranging from dark to light colors, represents the number of data points. The points cluster around the ideal prediction line, indicating a general agreement between predicted and target values. However, the presence of a few outlier events in the test set can be attributed to the unbalanced dataset, which contains limited training examples for larger magnitudes, making it challenging for the model to generalize effectively at higher magnitudes. Furthermore, rather poor predictive performance is also observed in a few borderline earthquake cases located in areas with a sparse distribution of stations (e.g., offshore or near the coast). As a result, these events are characterized by only a limited number of stations providing earthquake P-wave signals (or even S-wave signals for the nearest ones) within the 10-second window.

4.3 Examples of event data prediction

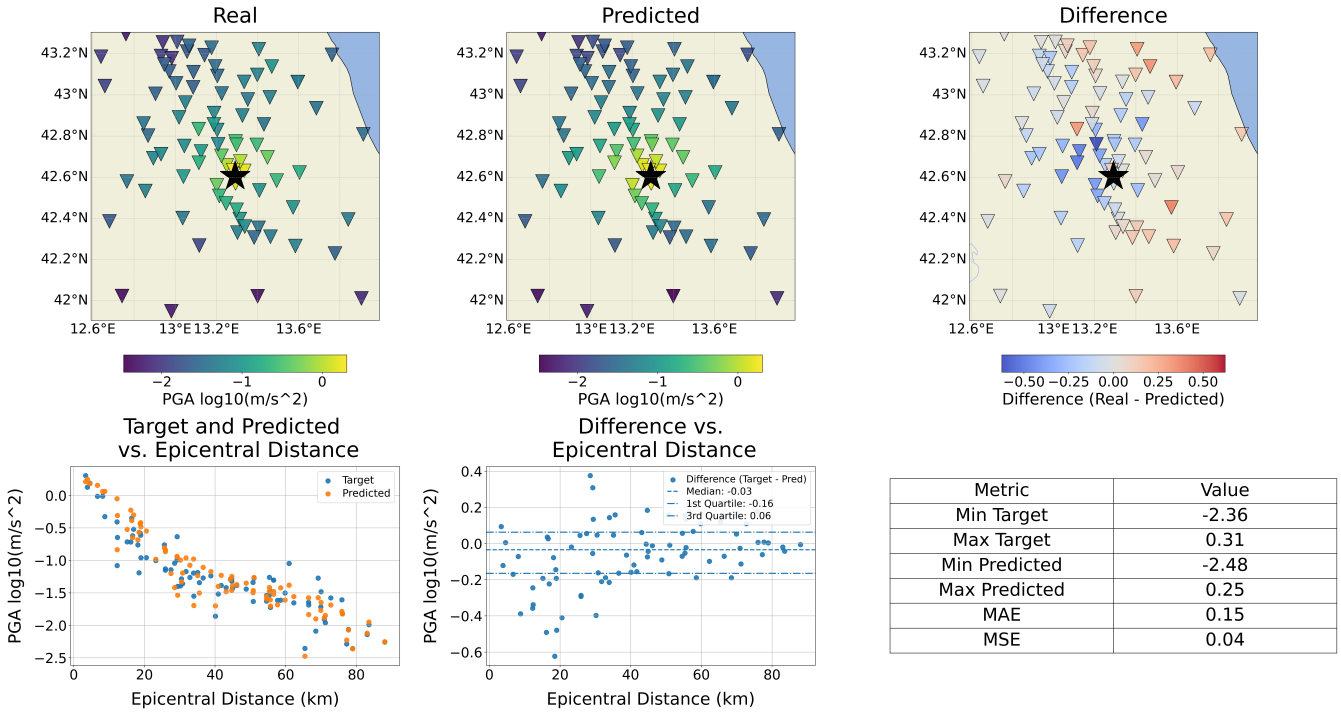
The panels in Figures 5, 6 and 7 display six sample events from the test set, presenting a comparison of real and predicted PGA values both spatially and as a function of distance. The maps and scatter plots in Figures 5, 6 and 7 clearly illustrate the performance of the predictive model, while the summary table provides a quantitative evaluation of the model's accuracy. Specifically, the top row illustrates the spatial distribution of observed PGA (left), predicted PGA (center), and their differences (right). The maps highlight the station locations, with color gradients representing the magnitude of the values: from purple (low) to yellow (high) for observed and predicted values, and from light blue (underpredictions) to orange (overpredictions) for the differences. The epicenter is marked by a black star. At the bottom, the left scatter plot shows observed (light blue) and predicted (orange) PGA values as a function of epicentral distance. The center scatter plot illustrates the residuals (observed - predicted) against epicentral distance, with dashed lines indicating the median, the first and the third quartile residuals, respectively. The table (bottom right) summarizes key metrics, including the range of observed and predicted PGA values, the mean absolute error (MAE) and the mean squared error (MSE).

The six events shown include the 26 August 2016 M 4.8, the 15 September 2016 M 3.7, the 26 October 2016 M 3.1, the 14 December 2016 M 3.8, the 6 April 2009 M 6.1 and the 13 June 2013 M 3.8. They have been selected because they are representative of the overall results. The relatively low MAE and MSE values for the first four earthquakes (Figs 5 and 6) demonstrate a strong agreement between observed and predicted values, representing the accuracy achieved in 96.4% of the cases within the test set. In contrast, the last two earthquakes are representative of the remaining 3.6% (the outliers) of the testing events, where predictive performance is significantly reduced. This lower accuracy can be primarily attributed to: (i) the limited availability of earthquake waveform data for larger magnitudes (Fig 7a) and (ii) poor station coverage—especially the uneven distribution in different directions surrounding the epicenter—which leads to incomplete or unbalanced waveform data and, consequently, to inaccurate ground motion predictions, potentially confusing the model during inference (Fig 7b).

4.4 Benchmarking the Model Against Baselines

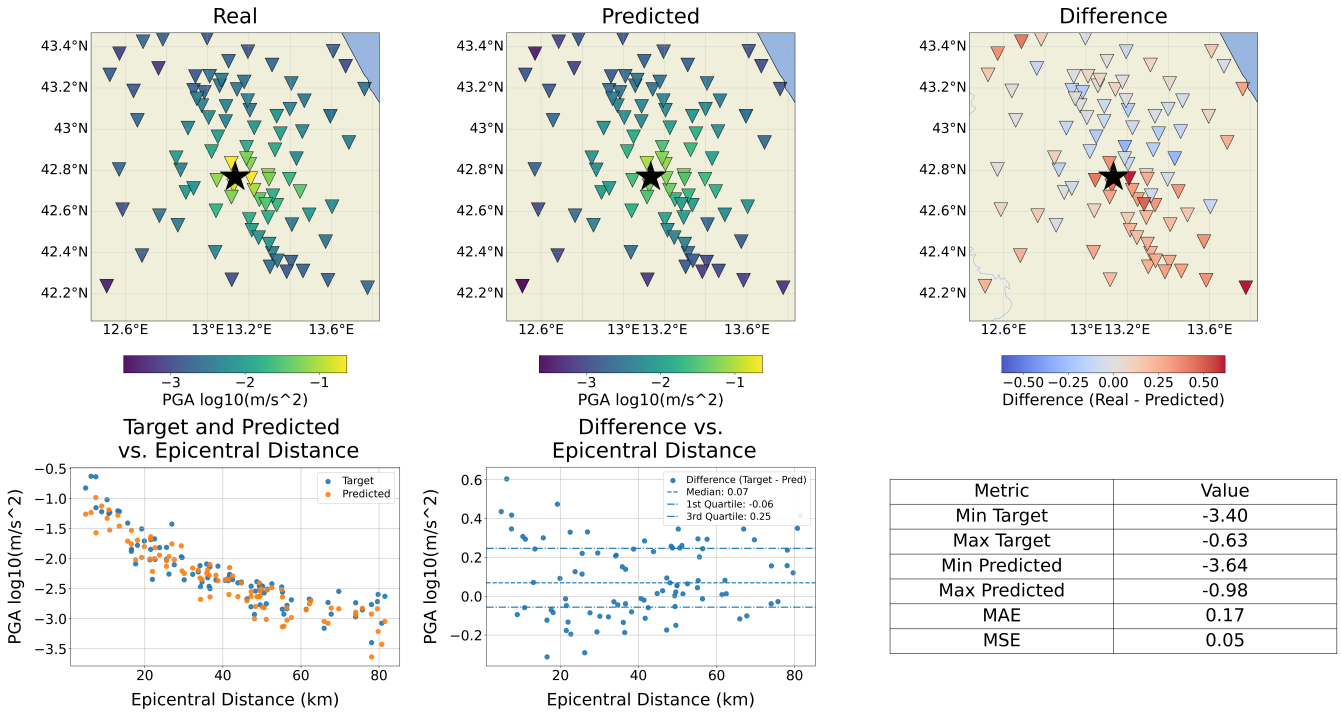
To evaluate the performance of the Masked GCN model against a baseline case, we compared our test set results with predictions obtained using the GMM by Bindi et al. (2011), calibrated specifically for Italy. For this comparison, we excluded the ShakeMap predictions used when no observed data are available (as it relies on the Bindi et al. (2011) GMM itself). For the GMM predictions, we used the INSTANCE metadata (Michelini et al., 2021) and we applied the necessary corrections based on the EC8 site classes of the stations. No between-event correction (Atik et al., 2010), however, has been applied when predicting the ground motion using the GMM by Bindi et al. (2011). The residuals between the base-10 logarithms of observed and predicted values (i.e. $\log_{10} \frac{IMobs}{IMpred}$) have been calculated together with their mean, median and standard deviation (SD) for the results obtained with (i) the Masked GCN model and (ii) the GMM by Bindi et al. (2011) (Table S2 and Fig 8). Large residual values [$\log_{10} \frac{IM_{obs}}{IM_{pred}} > |1|$] were removed resulting in 94.21% of the data retained for the GCN predictions and 97.56% for the GMM estimates. Our analysis shows that the median values of $\log_{10} \frac{IM_{obs}}{IM_{pred}}$, which reflect model bias, are significantly reduced in the Masked GCN model, particularly for PGV, SA(0.3) and SA(1.0). In terms of standard deviation, the values are similar across both models and are consistent with those reported by Bindi et al. (2011), who observed a standard deviation between 0.34 and 0.38 for $\log_{10} \frac{IM_{obs}}{IM_{pred}}$.

a) Comparison of Real and Predicted — Event: 7224451 — Magnitude 4.8



Metric	Value
Min Target	-2.36
Max Target	0.31
Min Predicted	-2.48
Max Predicted	0.25
MAE	0.15
MSE	0.04

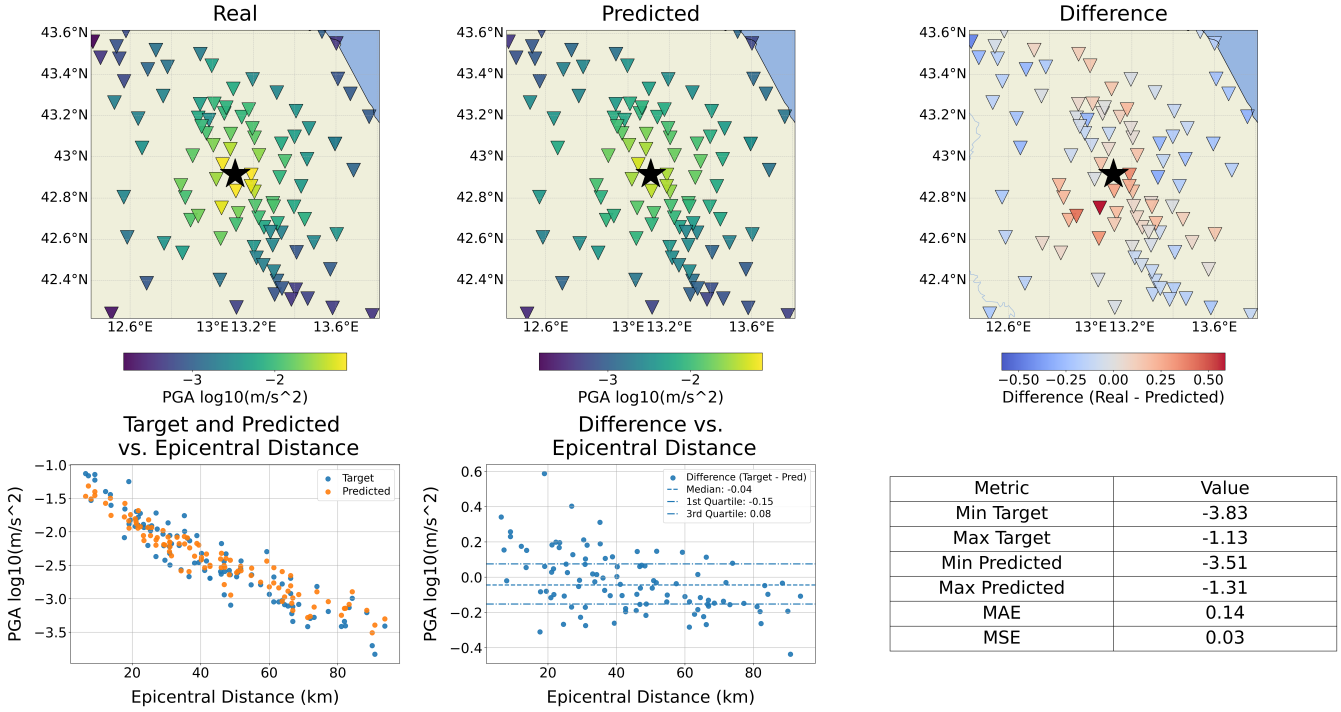
b) Comparison of Real and Predicted — Event: 8034701 — Magnitude 3.7



Metric	Value
Min Target	-3.40
Max Target	-0.63
Min Predicted	-3.64
Max Predicted	-0.98
MAE	0.17
MSE	0.05

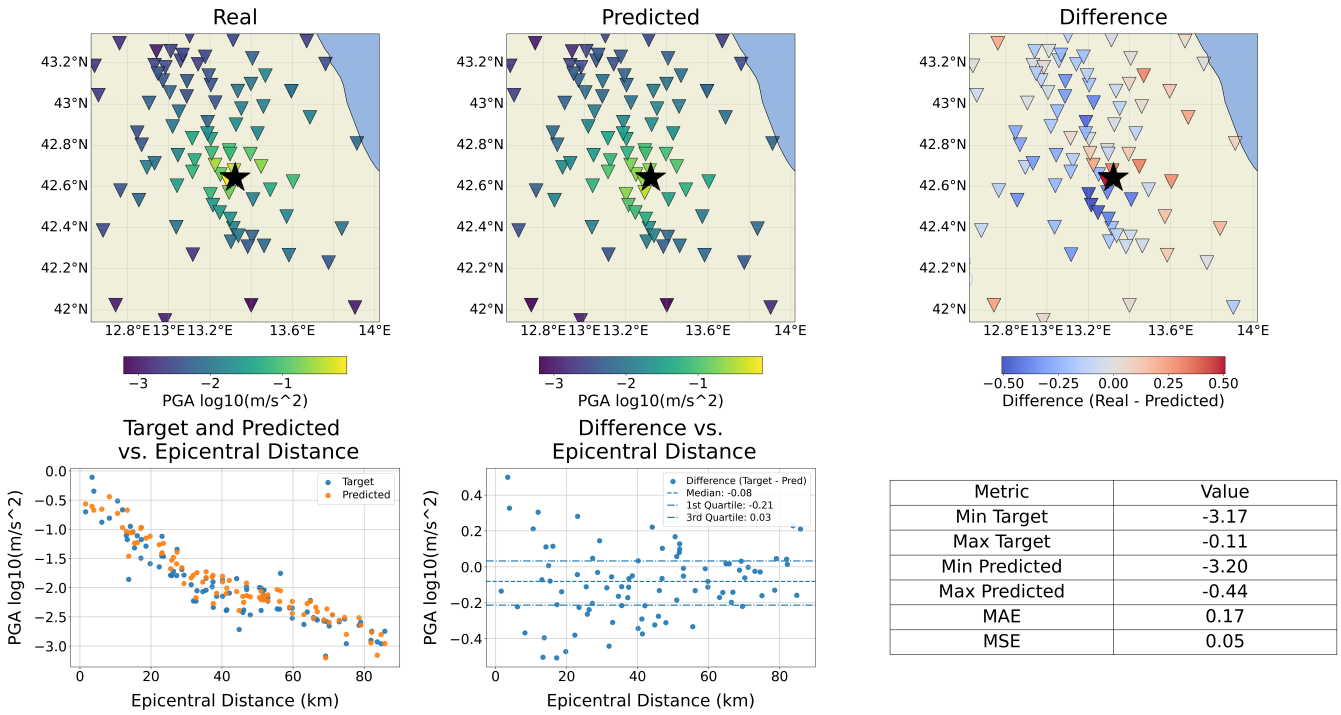
Figure 5 Our model results for (a) the 26 August 2016 M 4.8 and (b) the 15 September 2016 M 3.7. The top row presents spatial distributions of observed PGA (left), predicted PGA (center), and their differences (right) across the stations, with color gradients representing the respective values. The bottom row includes scatter plots of PGA values as a function of epicentral distance: observed vs. predicted values (left) and their differences (right). Key performance metrics, such as MAE, MSE, and the range of observed and predicted PGA values, are summarized in the accompanying table (bottom right).

a) Comparison of Real and Predicted — Event: 8673931 — Magnitude 3.1



Metric	Value
Min Target	-3.83
Max Target	-1.13
Min Predicted	-3.51
Max Predicted	-1.31
MAE	0.14
MSE	0.03

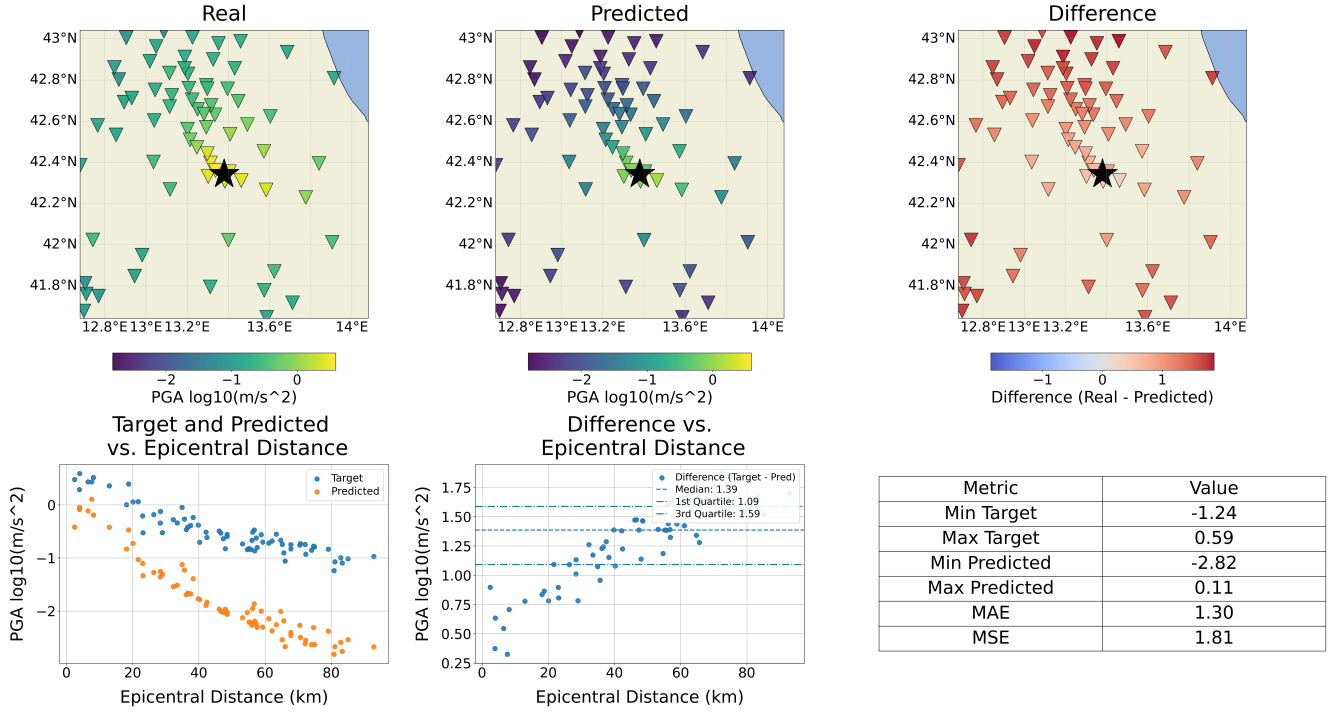
b) Comparison of Real and Predicted — Event: 11510961 — Magnitude 3.8



Metric	Value
Min Target	-3.17
Max Target	-0.11
Min Predicted	-3.20
Max Predicted	-0.44
MAE	0.17
MSE	0.05

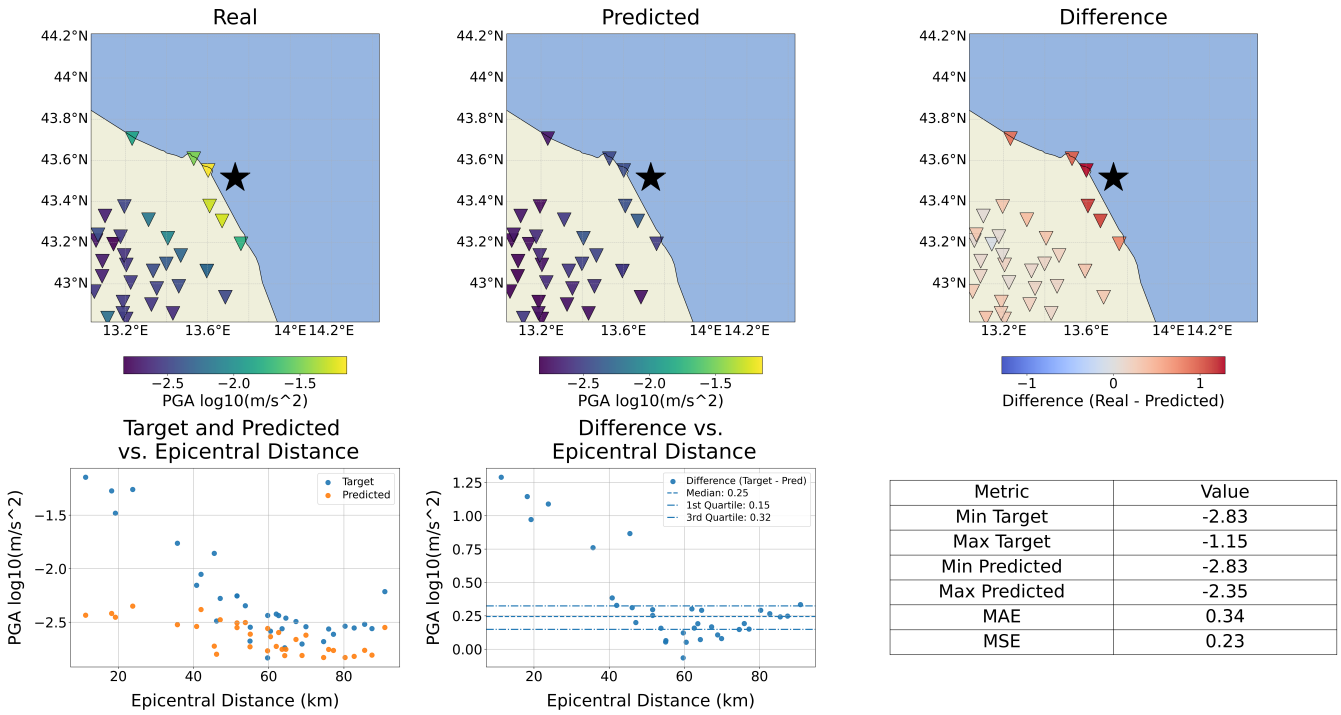
Figure 6 Our model results for (a) the 26 October 2016 M 3.1 and (b) the 14 December 2016 M 3.8. Refer to the caption of Figure 5 for further details.

a) Comparison of Real and Predicted — Event: 1895389 — Magnitude 6.1



Metric	Value
Min Target	-1.24
Max Target	0.59
Min Predicted	-2.82
Max Predicted	0.11
MAE	1.30
MSE	1.81

b) Comparison of Real and Predicted — Event: 2131571 — Magnitude 3.8



Metric	Value
Min Target	-2.83
Max Target	-1.15
Min Predicted	-2.83
Max Predicted	-2.35
MAE	0.34
MSE	0.23

Figure 7 Our model results for (a) the 6 April 2009 M 6.1 and (b) the 13 June 2013 M 3.8. Refer to the caption of Figure 5 for further details.

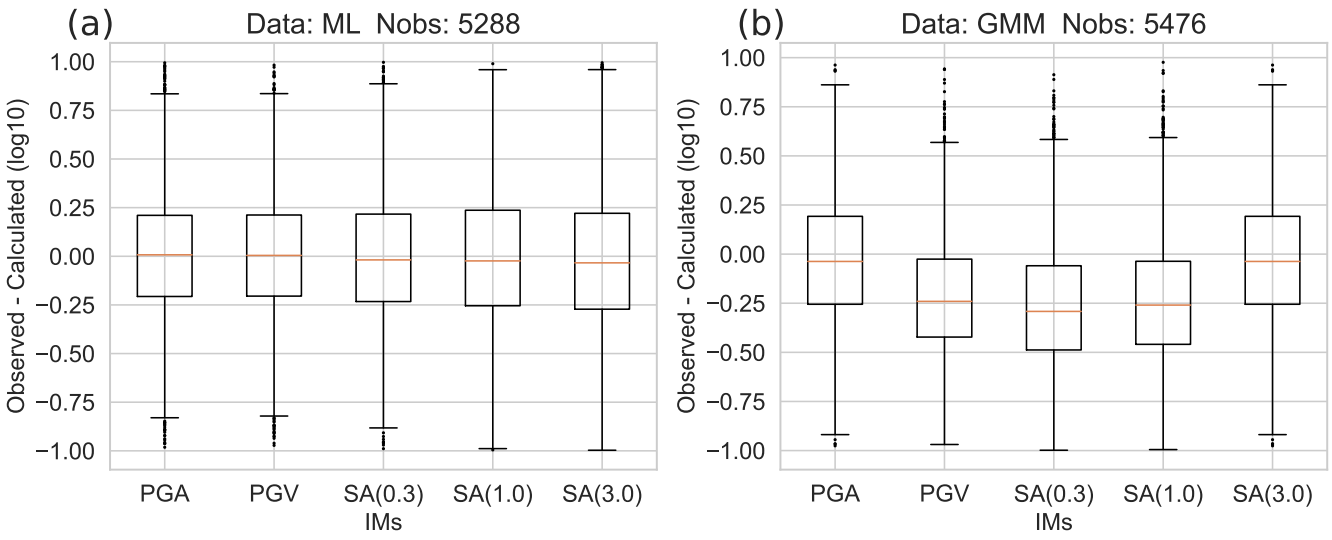


Figure 8 Boxplots for the residuals $[\log_{10} \frac{IM_{obs}}{IM_{pred}}]$ of the test set (195 events): (a) masked GCN model results for the observed IMs and (b) GMM results.

We also compared our work with the models proposed by Jozinović et al. (2020) and Bloemheuvel et al. (2023) (Table S3). These baseline models were not given the exact same input data as our proposed Masked TISER-GCN; instead, they were provided with different time series per station for different earthquakes in the same area. Overall, when evaluating the individual performance of the models, our Masked TISER-GCN performs comparably to the baselines. In terms of MSE, our model shows an improvement of 6% on PGA, 5.5% on PGV, and 5% on SA(0.3) compared to the TISER-GCN model proposed by Bloemheuvel et al. (2023). When compared to the CNN model from Jozinović et al. (2020), the improvement is even more pronounced, achieving 17% on PGA, 31% on PGV, and 33% on SA(0.3).

4.5 Effect of Window length on Model Performance

Sections 4.1, 4.2, 4.3 and 4.4 present results based on a 10-second input window length. In the final part of this study, we investigated how our model performs when the window length is reduced, as shorter windows could lead to earlier predictions—a critical aspect for integrating this methodology into EEW systems. However, using a smaller input window introduces greater challenges, as fewer stations will have received sufficient information to predict IMs at more distant locations. To mitigate this, we allowed data from as many stations as possible to be considered. Specifically, we included all stations that recorded the earthquake within 5 seconds, regardless of their distance from the first recording station. Our findings indicate that the input window can be reduced by half while still keeping comparable performance (Fig 9). Specifically, For PGA and PGV, the MSE of the residuals between the base-10 logarithms of observed and predicted values $[\log_{10} \frac{IM_{obs}}{IM_{pred}}]$ is 0.205. For SA(0.3), the MSE increased to 0.198, while for SA(1.0) and SA(3.0), the misfit increased slightly to 0.236 and 0.269, respectively. Overall, these results emphasize the effectiveness of the GCN layers and our implementation. Given that GCN layers are designed to facilitate node feature sharing during the convolution process, they enable the transfer of a substantial amount of information between nodes, even when the input is reduced by half, supporting the feasibility of our approach for real-time applications such as EEW.

5 Discussion

In this study, we demonstrated that our Masked GCN model can accurately predict earthquake IMs at recording stations using only waveforms captured within a 10-second time window starting from the earthquake's first P-wave arrival times. This approach enables accurate prediction of IMs at distant stations, even before they record the earthquake signal or its maximum values, based solely on the initial recordings at stations near the epicenter. To achieve this, we used 3C station waveform data and the spatial pattern of the waveforms across the recording network, without requiring any prior knowledge of the earthquake's location or magnitude. Our approach is conceptually similar to GMMs, but instead of relying on earthquake source (location and magnitude) and site parameters, it uses direct ground motion observations from the station network, specifically the waveform ground motion patterns.

The target outputs—PGA, PGV and SA at 0.3, 1, and 3-second periods—are predicted as a regression problem for all 565 stations using the Masked GCN model. This marks a significant advancement over the original TISER-GCN (Bloemheuvel et al., 2023), which was constrained by a fixed network geometry with only 39 stations. The model generally performs consistently well across all IM types. As expected, its performance improves when additional knowledge is integrated into the model and when stations receiving the P-wave within the first 10 seconds are prioritized.

Due to the Gutenberg-Richter magnitude distribu-

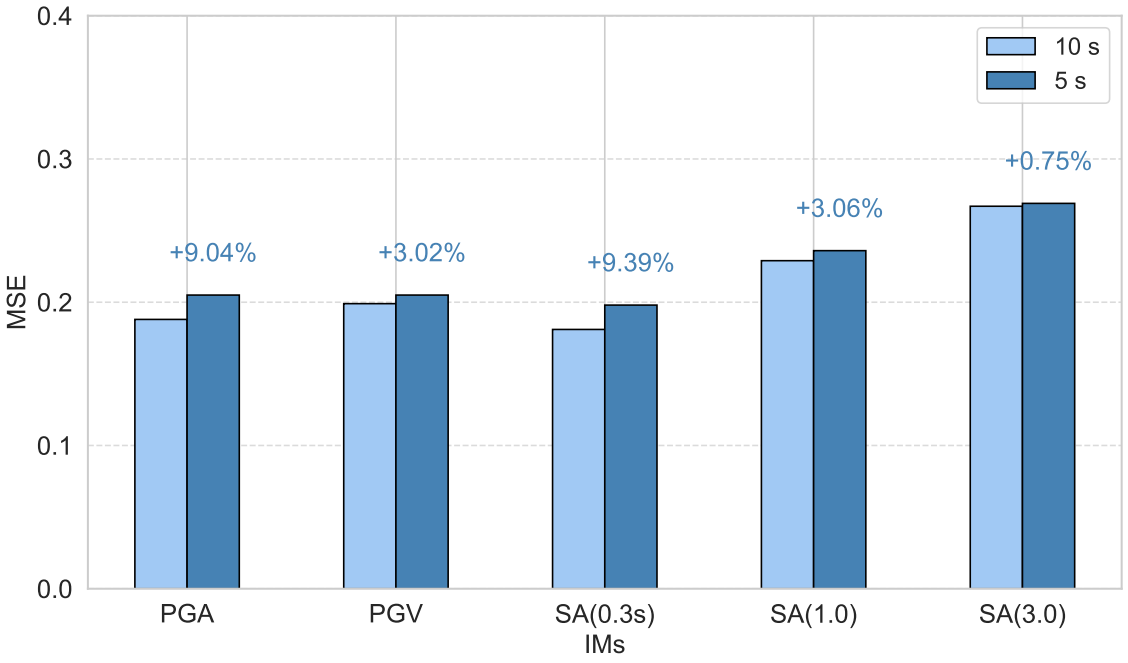


Figure 9 MSE at 10s and 5s input window lengths for each IM. Our model achieves approximately the same MSE values even with half the input, demonstrating the effectiveness of GCNs in processing spatial information.

tion, our dataset is significantly imbalanced, with a much higher frequency of smaller-magnitude events compared to larger ones. This imbalance presents challenges in predicting the IMs for larger events, as illustrated in Figure 7a. One potential solution involves applying data augmentation techniques, such as those proposed by Tang et al. (2021), by introducing additional training data for larger magnitudes through random transformations of existing events. Alternatively, synthetic seismograms could be generated for hypothetical larger events at the same stations. Another approach to address limited data is transfer learning (Tan et al., 2018), where a pre-trained model from a larger dataset is fine-tuned on a smaller, domain-specific dataset. As shown by (Jozinović et al., 2022), a pre-trained network designed for other seismological applications can be effectively adapted for tasks such as IM prediction or magnitude estimation. Similarly, the network developed in this study could be utilized as a pre-trained model for IM prediction in other regions.

To validate the methodology, we used processed waveforms with preliminary data cleaning to exclude stations affected by missing or erroneous data due to equipment failures or transmission issues. Following Jozinović et al. (2020), missing data were replaced with zeros, and ShakeMap-predicted target values were adopted to ensure the availability of target data during training. Despite these limitations, the Masked GCN model demonstrated robust accuracy in predicting IMs, suggesting its ability to learn meaningful spatial relationships between stations, conditioned on the input data. This implies that the model can exploit the spatial configuration of the seismic network and adapt its predictions based on the available waveform signals, even if it does not explicitly model the physical process of seismic wave propagation.

To evaluate the predictive performance of the Masked GCN model, we compared it against GMMs proposed by Bindi et al. (2011), which are employed in ShakeMap configurations for Italy (Michelini et al., 2020). As shown in Figure 8, the Masked GCN model outperforms the GMMs in terms of IM residuals. This is likely due to (i) the independence of the Masked GCN model from earthquake magnitude uncertainties (eliminating the need for between-event correction terms), (ii) the relatively uniform source-receiver geometry and wave paths within the study area, and (iii) the model's ability to capture and adjust for local site effects.

Notably, the earthquakes in our dataset are spatially concentrated, reducing variability at individual stations caused by differing wave paths. This likely contributed to the strong performance of our model. Nevertheless, we are confident that by incorporating a larger dataset with a broader spatial distribution of epicenters, the Masked GCN model can learn site- and path-specific characteristics for each station and zone, provided there is sufficient station coverage in these areas. Future studies will address these aspects by expanding the geographic scope and dataset size. In this context, a more accurate selection of P-wave velocity could be necessary to refine the travel-time estimates, considering a larger area that may require the adoption of different velocity models, depending on the depth and specific geology.

Our results (Figs 4, 5 and 6) confirm that the proposed methodology can predict ground motion effectively. Combined with the translational invariance properties of neural networks, this suggests that the approach could be implemented in real-time systems. In such a setup, continuously streamed data could be used to predict IMs while leveraging the dynamic masking mechanism to flexibly select which stations contribute to the

model's predictions, allowing for adaptable input configurations.

6 Conclusions

A Masked GCN model was employed to predict IMs, including PGA, PGV, and SA at 0.3 s, 1 s, and 3 s, using multistation 3C waveforms within a 10-second time window starting at the first P-wave arrival. The dataset consists of waveforms recorded by 565 stations from 975 earthquakes ($M \geq 3.0$, depth < 30 km) in central Italy.

The IM predictions were achieved without prior knowledge of earthquake location, distance, and magnitude. The model performance improved when additional information—such as the coordinates of the first-arrival station and the maximum amplitude value—was incorporated. To optimize the performance further, a dynamic masking mechanism was introduced to manage which stations contributed to the predictions, enabling flexible input configurations.

Our findings show consistent model performance across all IM types. A comparison between the Masked GCN model and the Bindi et al. (2011) GMM, calibrated for earthquakes in Italy, revealed that the Masked GCN exhibited no prediction bias while maintaining a similar residual variance.

Model performance degraded only in two scenarios: (i) for large-magnitude events due to the limited number of such earthquakes available for training, and (ii) in regions with sparse station coverage (e.g., offshore or near the coast). Our model achieves comparable performance while using only half of the input window length (5 s).

Overall, the results suggest that the proposed approach could be integrated into EEW systems, leveraging the Masked GCN model's ability to prioritize critical early-wave information through dynamic masking. Although this work focused on a geographically concentrated dataset, the methodology could be extended to larger areas with more widespread seismicity, provided sufficient training data is available. Overall, we showed that the proposed Masked GCN could achieve ground motion predictions as reliably as those in Bloemheuvel et al. (2023) without being restricted to a fixed station geometry.

Acknowledgements

The research has been funded by The European Union project "A Digital Twin for Geophysical Extremes" (DT-GEO) (No: 101058129). The constructive and insightful criticism of Giuseppe Costantino and another anonymous reviewer is greatly acknowledged.

Data and code availability

The INSTANCE dataset, assembled in HDF5 format, is available at http://doi.org/10.13127/instance (Michelini et al., 2021). Waveforms and metadata have been processed using python libraries. IMs for the stations with no data have been calculated using USGS ShakeMap 4. The Masked GCN model has been developed using the Pytorch library. The code of this paper is available on Mask_GCN GitHub Repository and in Zenodo 10.5281/zenodo.15050355.

References

Atik, L. A., Abrahamson, N., Bommer, J. J., Scherbaum, F., Cotton, F., and Kuehn, N. The variability of ground-motion prediction models and its components. *Seismological Research Letters*, 81 (5):794–801, 2010. doi: 10.1785/gssrl.81.5.794.

Bindi, D., Pacor, F., Luzi, L., Puglia, R., Massa, M., Ameri, G., and Paolucci, R. Ground motion prediction equations derived from the Italian strong motion database. *Bulletin of Earthquake Engineering*, 9:1899–1920, 2011. doi: 10.1007/s10518-011-9313-z.

Bloemheuvel, S., van den Hoogen, J., Jozinović, D., Michelini, A., and Atzmueller, M. Graph neural networks for multivariate time series regression with application to seismic data. *International Journal of Data Science and Analytics*, 16(3):317–332, 2023. doi: 10.1007/s41060-022-00349-6.

Böse, M., Heaton, T., and Hauksson, E. Rapid estimation of earthquake source and ground-motion parameters for earthquake early warning using data from a single three-component broadband or strong-motion sensor. *Bulletin of the Seismological Society of America*, 102(2):738–750, 2012. doi: 10.1785/0120110152.

Danecek, P., Pintore, S., Mazza, S., Mandiello, A., Fares, M., Carluccio, I., Della Bina, E., Franceschi, D., Moretti, M., Lauciani, V., et al. The Italian node of the European integrated data archive. *Seismological Research Letters*, 92(3):1726–1737, 2021. doi: 10.1785/0220200409.

Derras, B., Bard, P. Y., and Cotton, F. Towards fully data driven ground-motion prediction models for Europe. *Bulletin of Earthquake Engineering*, 12(1):495–516, 2014. doi: 10.1007/s10518-013-9481-0.

Fornasari, S. F., Pazzi, V., and Costa, G. A machine-learning approach for the reconstruction of ground-shaking fields in real time. *Bulletin of the Seismological Society of America*, 112(5): 2642–2652, 2022. doi: 10.1785/0120220034.

Gandomi, A. H., Alavi, A. H., Mousavi, M., and Tabatabaei, S. M. A hybrid computational approach to derive new ground-motion prediction equations. *Engineering Applications of Artificial Intelligence*, 24(4):717–732, 2011. doi: 10.1016/j.engappai.2011.01.005.

Hsu, T.-Y., Huang, S.-K., Chang, Y.-W., Kuo, C.-H., Lin, C.-M., Chang, T.-M., Wen, K.-L., and Loh, C.-H. Rapid on-site peak ground acceleration estimation based on support vector regression and P-wave features in Taiwan. *Soil Dynamics and Earthquake Engineering*, 49:210–217, 2013. doi: 10.1016/j.soildyn.2013.03.001.

Hu, J., Ding, Y., Zhang, H., Jin, C., and Wang, Z. A Real-Time Seismic Intensity Prediction Framework Based on Interpretable Ensemble Learning. *Seismological Research Letters*, 94(3):1579–1602, 2023. doi: 10.1785/0220220167.

Iaccarino, A. G., Cristofaro, A., Picozzi, M., Spallarossa, D., and Scafidi, D. Real-time prediction of distance and PGA from P-wave features using Gradient Boosting Regressor for on-site earthquake early warning applications. *Geophysical Journal International*, 236(1):675–687, 2024. doi: 10.1093/gji/ggad443.

Jozinović, D., Lomax, A., Štajduhar, I., and Michelini, A. Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. *Geophysical Journal International*, 222(2):1379–1389, 2020. doi: 10.1093/gji/ggaa233.

Jozinović, D., Lomax, A., Štajduhar, I., and Michelini, A. Transfer

learning: Improving neural network based prediction of earthquake ground shaking for an area with insufficient training data. *Geophysical Journal International*, 229(1):704–718, 2022. doi: 10.1093/gji/ggab488.

Margheriti, L., Nostro, C., Cocina, O., Castellano, M., Moretti, M., Lauciani, V., Quintiliani, M., Bono, A., Mele, F. M., Pintore, S., et al. Seismic surveillance and earthquake monitoring in Italy. *Seismological Research Letters*, 92(3):1659–1671, 2021. doi: 10.1785/0220200380.

Michelini, A., Margheriti, L., Cattaneo, M., Cecere, G., D'Anna, G., Delladio, A., Moretti, M., Pintore, S., Amato, A., Basili, A., et al. The Italian National Seismic Network and the earthquake and tsunami monitoring and surveillance systems. *Advances in Geosciences*, 43:31–38, 2016. doi: 10.5194/adgeo-43-31-2016.

Michelini, A., Faenza, L., Lanzano, G., Lauciani, V., Jozinović, D., Puglia, R., and Luzi, L. The new ShakeMap in Italy: Progress and advances in the last 10 yr. *Seismological Research Letters*, 91(1): 317–333, 2020. doi: 10.1785/0220190130.

Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., and Lauciani, V. INSTANCE–the Italian seismic dataset for machine learning. *Earth System Science Data*, 13(12):5509–5544, 2021. doi: 10.5194/essd-13-5509-2021.

Münchmeyer, J., Bindi, D., Leser, U., and Tilmann, F. The transformer earthquake alerting model: A new versatile approach to earthquake early warning. *Geophysical Journal International*, 225(1):646–656, 2021. doi: 10.1093/gji/ggaa609.

Otake, R., Kurima, J., Goto, H., and Sawada, S. Deep learning model for spatial interpolation of real-time seismic intensity. *Seismological Research Letters*, 91(6):3433–3443, 2020. doi: 10.1785/0220200006.

Spallarossa, D., Kotha, S. R., Picozzi, M., Barani, S., and Bindi, D. On-site earthquake early warning: a partially non-ergodic perspective from the site effects point of view. *Geophysical Journal International*, 216(2):919–934, 2019. doi: 10.1093/gji/ggy470.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*, pages 270–279. Springer, 2018.

Tang, Z., Qiao, Z., Hong, X., Wang, Y., Dharejo, F. A., Zhou, Y., and Du, Y. Data augmentation for graph convolutional network on semi-supervised classification. In *Web and Big Data: 5th International Joint Conference, APWeb-WAIM 2021, Guangzhou, China, August 23–25, 2021, Proceedings, Part II 5*, pages 33–48. Springer, 2021.

Wang, A., Li, S., Lu, J., Zhang, H., Wang, B., and Xie, Z. Prediction of PGA in earthquake early warning using a long short-term memory neural network. *Geophysical Journal International*, 234(1): 12–24, 2023. doi: 10.1093/gji/ggad067.

Wang, C.-Y., Huang, T.-C., and Wu, Y.-M. Using LSTM neural networks for onsite earthquake early warning. *Seismological Research Letters*, 93(2A):814–826, 2022. doi: 10.1785/0220210197.

Yang, L., Wu, F., Wang, Y., Gu, J., and Guo, Y. Masked Graph Convolutional Network. In *IJCAI*, pages 4070–4077, 2019. doi: 10.24963/ijcai.2019/565.

Zhang, H., Melgar, D., Sahakian, V., Searcy, J., and Lin, J.-T. Learning source, path and site effects: CNN-based on-site intensity prediction for earthquake early warning. *Geophysical Journal International*, 231(3):2186–2204, 2022. doi: 10.1093/gji/ggac325.

Zhu, W. and Beroza, G. C. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273, 2019. doi: 10.1093/gji/ggy423.