ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

IL FUTURO DELLA TERRA, CAMBIAMENTI CLIMATICI E SFIDE SOCIALI

Ciclo XXXVI

Settore Concorsuale: 04/A4 - GEOFISICA

Settore Scientifico Disciplinare: GEO/10 – GEOFISICA DELLA TERRA SOLIDA

Neural Nets: applications to two relevant seismological problems

Presentata da: ***MIRIANA DI DONATO***

Coordinatore Dottorato **Supervisore**

SILVANA DI SABATINO **SILVIA CASTELLARO**

Esame finale anno 2024

Abstract

Italy is among the Mediterranean countries with a high seismic hazard, and risk. Over the past century, it experienced several earthquakes that caused human and economic losses. There is not yet a way to predict earthquakes, but through appropriate monitoring techniques, seismic studies, territorial planning and anti-seismic engineering design, we can enhance our level of prevention against this natural event.

In this thesis we aim at providing a contribution to two subsets of the site effect assessment branch of research, and specifically to some aspects related to the Horizontal to Vertical Spectral Ratio (HVSR) technique and Ground Motion Modelling (GMM). To this aim, we employ machine learning techniques, here applied in a not-yet-experimented way.

In conclusion, in this thesis we explored a number of different neural network-based approaches, such as artificial neural network (ANN), convolutional neural network (CNN) which includes also U-Net, to different classification or prediction problems in seismology. This was important to learn the advantages and disadvantages of the different approaches and constitutes a first basis to build more robust studies in the future and to expand the applications studied so far.

In the first case, we employed Artificial Neural Networks (ANN) and a pre-trained SqueezeNet to separate stratigraphic from non-stratigraphic H/V peaks. This is important to capture the features of H/V curves that are really linked to the soil properties and not to external interference. Implementing neural networks for HVSR analysis could complement traditional methods (e.g. SESAME criteria), simplifying peak recognition and supporting operators in data interpretation.

In the second case, we used a U-Net neural net capable of generating region-specific, fully non-ergodic, and data-driven GMM, an approach not previously explored in the Italian context. Afterwards, we compared the outcomes with traditional ground motion prediction equations. Potential future developments of this methodology could offer shaking maps in areas currently lacking seismic data, benefiting regions with limited network distribution or for real-time mapping of ground shaking during an earthquake.

Contents

Introduction

Earthquakes, one of the most devastating natural phenomena, continue to pose an unpredictable threat to global communities, causing loss of lives, structural damage, and significant economic losses. At present, earthquake prediction remains an elusive goal while prevention is still the only way to mitigate the impacts of such events.

Seismic prevention has many possible declinations, including: i) studies on the characteristic earthquakes expected at a site of interest and their average occurrence rate of time-occurrence (probabilistic and deterministic seismic hazard assessments), ii) studies aimed at assessing the ground motion features expected at sites of interest (site effect studies), and of course iii) seismic engineering studies to design and build structures capable to withstand the expected seismic actions.

The research that we are going to present in this thesis falls in the second item (ii) listed above, that is in the realm of site effect assessment and provides a contribution to two specific subjects within this main topic, that are:

a) geophysical techniques to image the subsoil with the aim of seismic hazard assessment. In particular, it focuses on one aspect of the H/V or HVSR (Horizontal-to-Vertical Spectral Ratio), technique, which is a widely diffuse approach to measure subsoil resonances and map the main seismo-stratigraphic reflectors in the subsoils,
b) ground motion prediction 'equations' or ground motion modelling (GMM).

The element of novelty of this thesis will be the use of different classes of neural networks to approach the problems just described.

In the first application, we will explore the potential of artificial neural networks and convolutional neural networks, known as SqueezeNet, for the automatic detection of H/V peaks of stratigraphic vs non-stratigraphic origin. Particularly, but not only, in urban areas, H/V curves are often disturbed by artefactual peaks of often not-immediately clear origin (typical due to machineries or to the self-resonances of large nearby structures that propagate to the surrounding soil). Artefacts have been reported even at several hundred meters or even kilometers from their sources, particularly at low frequencies. Due to the widespread use of the H/V curve, training a network capable to discriminate between a stratigraphic and an artefactual peak is an interesting task and it also forces us to reason on what makes a peak appear stratigraphic or not in the seismic signal.

In the second application, we will attempt to produce ground motion maps (that is maps of the expected ground motion at any site of interest, starting from given earthquakes). We will do this by using fully data-

driven procedures, that is based solely on data, which surpass the traditional empirical approaches resulting into Ground Motion Prediction Equations (GMPE). The latter requires the assumption of a functional model between some input parameters and the desired output (the ground motion prediction) while we aim at reaching the same prediction with no assumptions about the functional model.
To this aim, we will still apply neural networks: in this case a U-Net algorithm.

The thesis is organized as it follows:
Chapter 1 provides a brief history on the artificial intelligence evolution over time and a general overview of the main existent neural network approaches.

Chapter 2 introduces the first seismological application of neural networks explored in this thesis, that is the automatic (neural network-based) stratigraphic vs non-stratigraphic H/V peak detection.

Chapter 3 deals with the second seismological application of neural networks explored in this thesis, that is the fully data-driven generation of ground motion maps.

Chapter 4 is dedicated to a final discussion of the main results, open questions, and future developments.

Chapter 1. Neural networks

This chapter provides an overview on machine learning and deep learning, with a particular focus on neural networks. We explore the history of these networks, inspired by the functioning of the human brain, starting from the single neuron (perceptron), and progressing to multilayer networks (multilayer perceptron). Subsequently, we explore the progression of neural networks, including the definition of convolutional neural networks from their origin to current configurations. To conclude, we present a general insight of the extensive use of these machine learning techniques in the field of seismology.

1.1 Artificial Intelligence and its subfield

Artificial intelligence, machine learning, deep learning, and neural networks are terms that the public often mistakenly uses interchangeably. Machine learning, deep learning, and neural networks are all specific fields within Artificial Intelligence (AI). However, the relationship between these fields is hierarchical (Figure 1). Neural networks are a subset of machine learning, and deep learning is a subset of neural networks.

The term Artificial Intelligence was first coined in the 1950s by John McCarthy and refers to computational machines capable to simulate human cognitive abilities, such as reasoning, planning, and creativity.

Machine learning (ML) is defined as the capability of a machine to learn from data and experience in order to make predictions with minimal human intervention.

Deep learning (DL), on the other hand, is one of the various approaches within machine learning. While in ML statistical methods are often used to analyse data and make predictions, DL utilizes artificial neural networks to learn complex patterns from data and perform specific tasks such as image recognition or natural language processing.

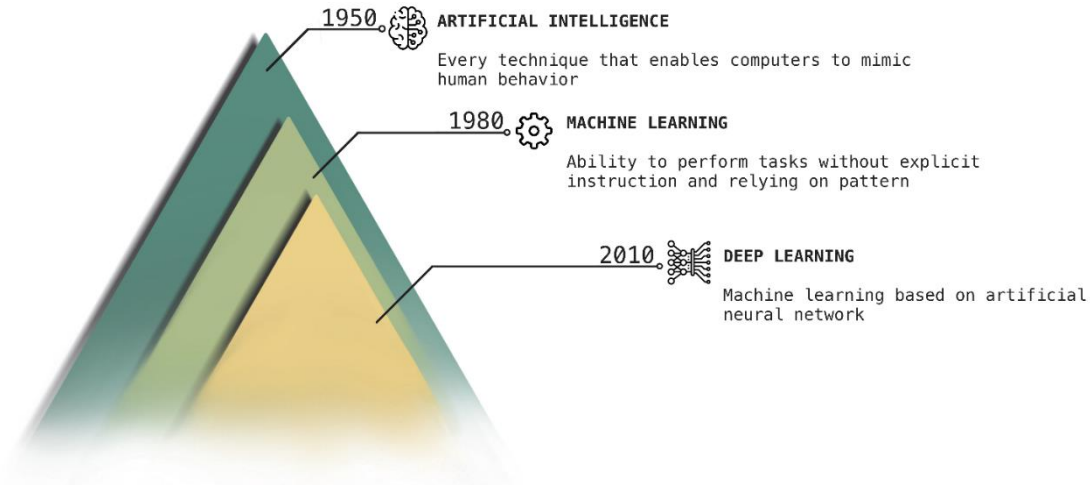

Figure 1. Scheme illustrating the evolution and nesting of artificial intelligence (AI) technologies.

1.2 Machine learning

ML is based on the idea that machines can learn from data and get better over time without being explicitly programmed (Mitchell, 1997). There are three subcategories of machine learning approaches, each with distinct and interesting applications (Figure 2).

Supervised machine learning models are trained on a set of labelled input data, where the desired output is already known. This approach is used for *classification*, where the model categorizes inputs into predefined classes, and *regression*, where it predicts numerical values. These models are widely used in areas like voice recognition, text classification, and image recognition.

Unsupervised machine learning models are trained on a set of unlabelled input data. *Clustering* and *dimensionality reduction* are common unsupervised techniques. These approaches are useful when there are no labels to guide the training, allowing the model to discover interesting patterns or trends by itself that people aren't explicitly looking for. It is useful for anomaly detection and data clustering.

Reinforcement machine learning trains model using trial and error approaches. The goal is to maximize cumulative rewards over time. This method is often used in gaming, robotics, optimization problems, or to train self-driving vehicles and resource management.

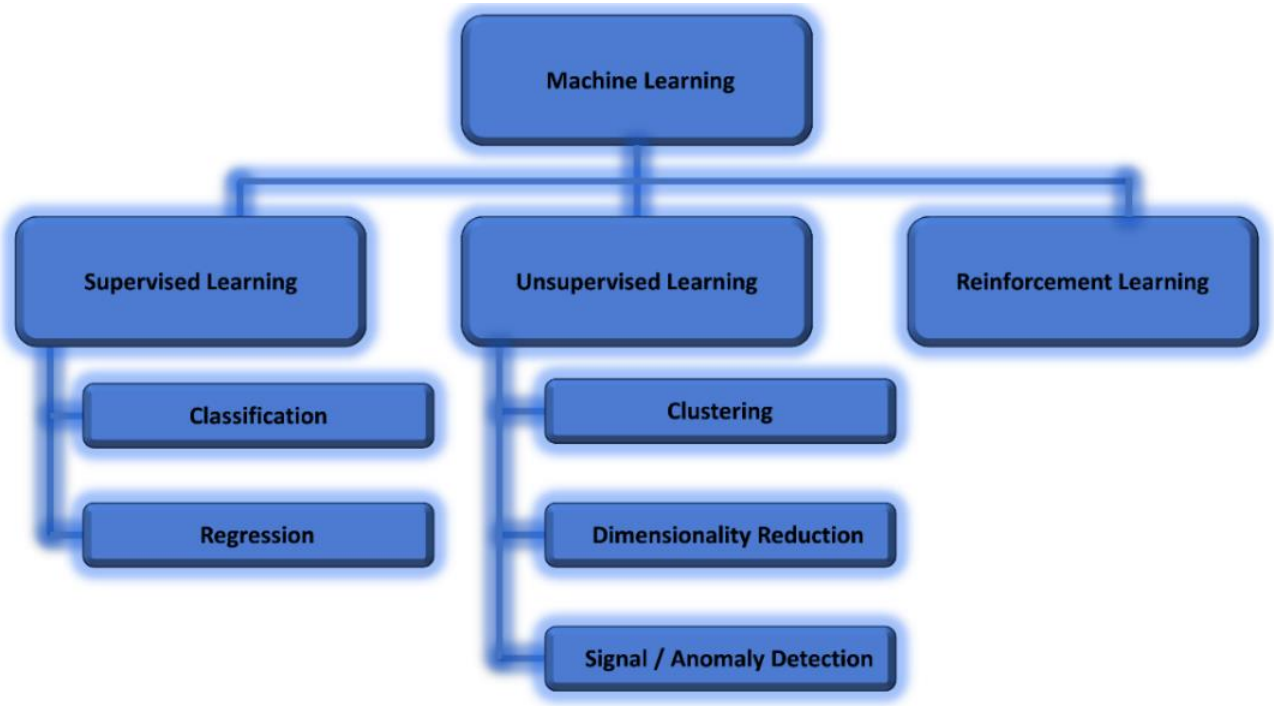

Figure 2. Diagram of machine learning paradigms: Supervised (with classification and regression), Unsupervised (including clustering, dimensionality reduction and signal/anomaly detection), and Reinforcement Learning.

1.3 Deep Learning

Deep Learning is a subset of Machine Learning that utilizes neural networks for automatic learning. An artificial neural network consists of various layers, and, in this section, we will examine how these layers work together to let a net to learn from data.

1.3.1 Definition and History of Artificial Neural Networks

Artificial Neural Networks (ANNs) are models inspired by the human brain and are capable in autonomous learning from data. They are characterized by artificial neurons, like those in our brain.

In real neurons, there are components called dendrites, axons, and synapses (Figure 3 a, c). Dendrites collect electrical signals, while axons send signals to terminal axons. At the end of axons, synapses transform the activity into electrical signals that influence other neurons. Learning occurs when signals are transmitted between neurons.

The functioning of artificial neural networks is very similar to the brain's one. We can compare synapses to the connections between artificial neurons. Each connection is represented by a specific weight. Information transmission and processing occur through the activation of neurons, which multiply the input signal by the connection weight and sum the results to produce the output. Therefore, neural networks are composed of groups of artificial neurons organized into layers connected to each other. Typically, there is an input layer, an output layer, and one or more intermediate or hidden layers. Each layer contains one or more neurons (Figure 3 b, d).

McCulloch and Pitts in 1943 (McCulloch & Pitts, 1943) proposed the first model of a neuron known as the "McCulloch-Pitts neuron", which later became the basis for the development of the first artificial neurons in neural networks. The model was very simple and could perform logical functions using binary inputs and outputs. Later on, this model was improved by Rosenblatt, who proposed the *Perceptron* in 1958 (Rosenblatt, 1958). This new model had superior learning capabilities compared to the previously proposed one (Figure 3 b). However, the limitations of the *Perceptron* became evident when trying to apply it to more complex and non-linearly separable problems.

In the 1940s, the first artificial neurons were created. However, a significant revolution occurred in the 1980s with the introduction of the error backpropagation algorithm by Rumelhart, Hinton, and Williams (Rumelhart et al., 1986). This algorithm facilitated the creation of more complex neural network models, surpassing the limitations of earlier models such as the *Perceptron* (Figure 3 d). Today, this technique is widely used in various areas of artificial intelligence, such as image recognition and natural language processing. This is

made possible thanks to the use of increasingly deep neural networks (i.e. with more layers), the growth of available data, and advancements in technology.

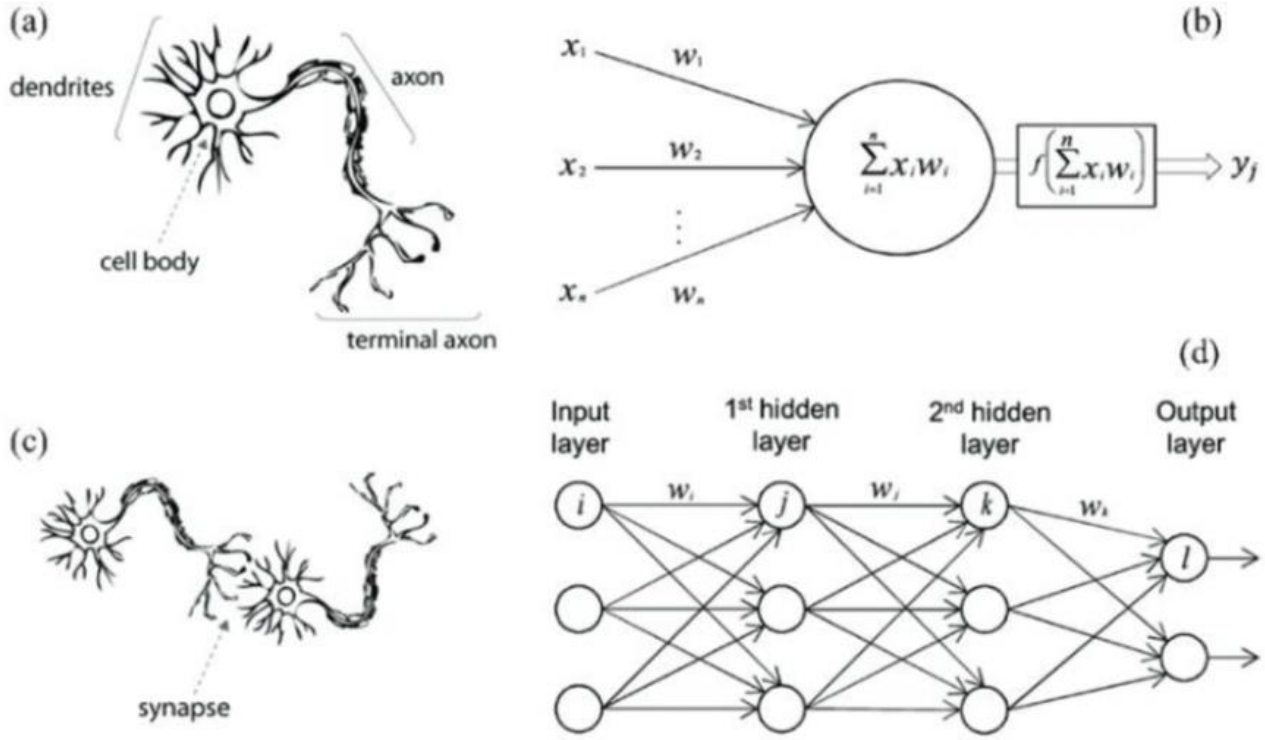

Figure 3. Image depicting a biological neuron in the human brain (a) compared with an artificial neuron (b). Illustration of interconnected biological neurons through synapses (c) compared with an artificial neural network (d) using artificial neurons and connections similar to biological synapses (Sumari et al., 2021).

1.3.2 Artificial Neural Networks Architecture

In Neural Networks, individual neurons work all together and are connected to each other through weighted connections. In general, they are composed of groups of artificial neurons organized into layers. Typically, the input layer receives information, then hidden layers identify complex patterns from input data. Each neuron in a hidden layer is connected to all neurons in the following layer, and the learning process takes place through the updating of weights associated with these connections.

The output layer typically contains a number of neurons equal to the number of possible classes. Each neuron in the output layer will be a number in the [0, 1] range, representing the confidence of the attribution of the examined example to a specific output class. In regression problems the output layer can also have a single neuron and produce a single numerical value.

Therefore, the architecture of the network includes input layer, a number of hidden layers, and output layer, the connections and activation functions.

1.3.2.1 Neurons

In the context of neural networks, the fundamental building blocks are neurons. Neurons are individual units that process information and contribute to the overall functionality of the network. Figure 4 illustrates a fundamental concept in neural networks, known as "nodes" or "neurons". Each node or neuron receives inputs, denoted as a_1, a_2, a_3, …, a_n, each associated with a "weight", denoted as w_1, w_2, w_3, …, w_n. These weights indicate the importance of each input for the neuron.

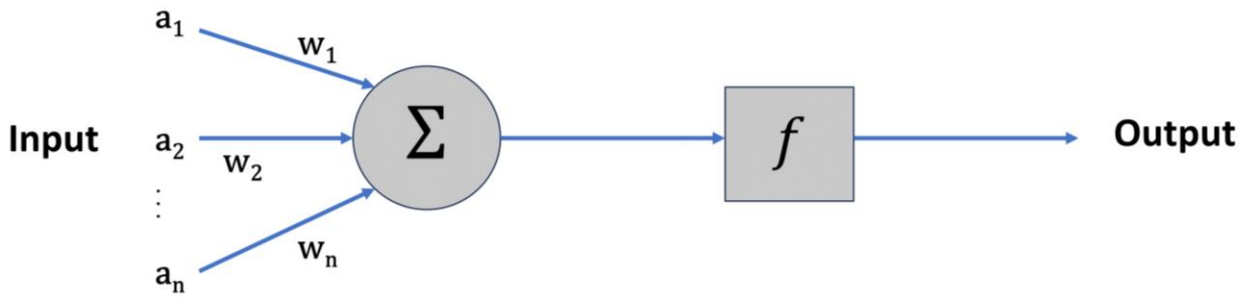

Figure 4. Diagram illustrating the operation of a single neuron: a_1, a_2, ..., a_n represent the inputs, w_1, w_2, ..., w_n are the corresponding weights, the weighted sum of inputs is performed, followed by the activation function (f) application, resulting in the output.

The neuron processes these inputs by multiplying them with their respective weights and summing the results (Eq. 1). Therefore, the output of net can be described as:

$$Output = f\left(\sum_{i=0}^{n} w_i a_i\right) = f(w_1 a_1 + w_2 a_2 + \ w_3 a_3 + \cdots + \ w_n a_n) \quad (1)$$

where f is an activation function that determines the final output of the node. So, weights influence how much each input contributes to the output of neuron, allowing the neural network to learn and adapt to patterns in the input data during the training process.

1.3.2.3 Activation function

The weighted sum of inputs, added to the bias, is then processed through an activation function (f in Eq. 1). This function determines if the neuron should activate or not in response to the input. This process repeats for each neuron in every layer of the network. Among the most common, we find:

- ***Threshold function***: the function returns 1 if the weighted sum of input signals is greater than or equal to specific threshold, and 0 otherwise.

- ***Sigmoid function*** (LeCun et al., 1998)***:*** the function spans between 0 and 1 in a continuous interval (Figure 5), and have typical S-shape. Could be used to predict the probability as an output of model.

- ***ReLU function*** (Nair & Hinton, 2010)***:*** is calculated as max (0, x). That is, if the input value (x) is negative, it returns 0; otherwise, it returns the value x (Figure 5). When using the ReLU function, it is useful to normalize the input values between zero and one.

- ***LeakyReLU function*** (Maas et al., 2013)***:*** the ReLU function can deactivate some neurons in each iteration, for example in case where the value is negative. This is known as the dying ReLU problem. Unlike ReLU, which returns 0 for negative input values, LeakyReLU returns a small negative quantity for these values (Figure 5).

- ***Softmax function*** (Bishop, 2006): this function returns values ranging from 0 to 1. So, if initially, we have vectors with values greater than 1 or less than 0, applying the softmax function normalizes these values to fall within the desired range of 0 to 1. Additionally, the sum of the values obtained after applying the function will always be equal to 1, making its output interpretable as a probability distribution.

- ***Hyperbolic tangent function (tanh)***: it is similar to the Sigmoid function, but its range goes from -1 to +1 (Figure 5).

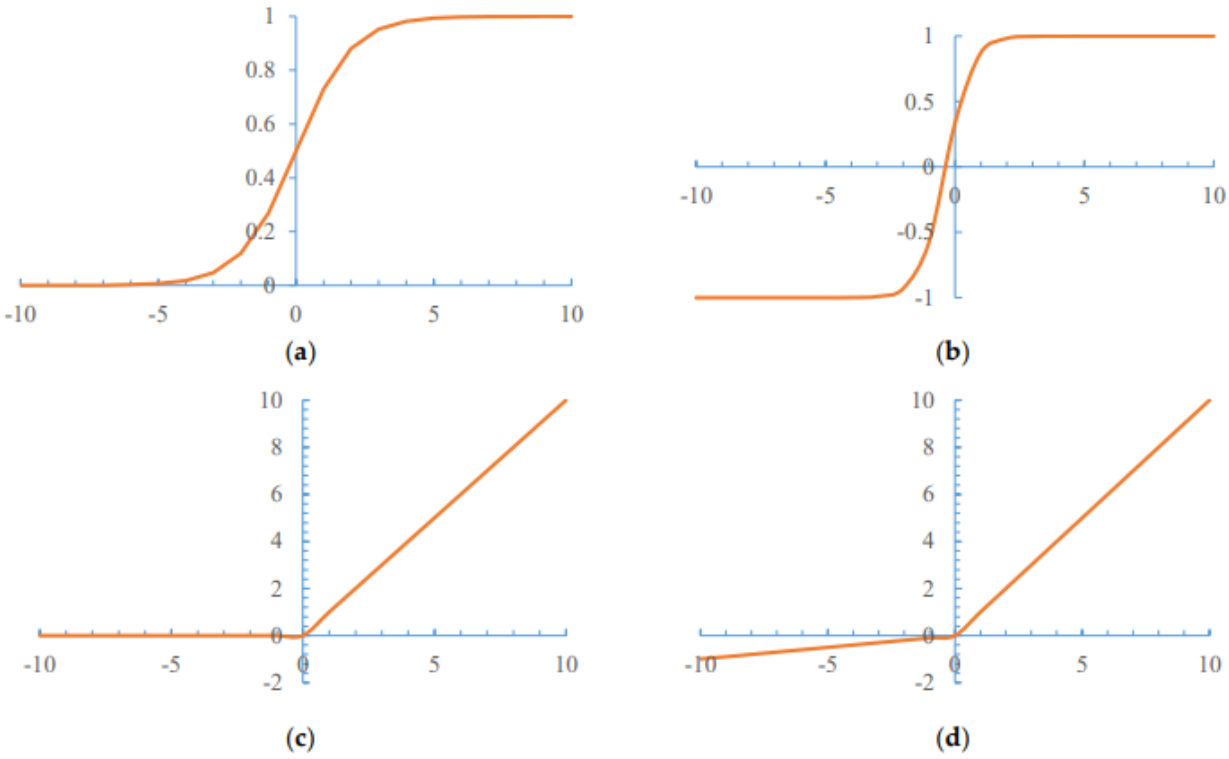

Figure 5. Nonlinear function: a) Sigmoid function; b) Tanh function; c) ReLU function; d) Leaky ReLU function (Yang & Yang, 2018).

1.3.2.4 Exploring Neural Network Architectures: Feedforward vs Recurrent Networks

There exist two main architectures in neural networks: the feedforward network and recurrent network (Figure 6). These architectures indicate different ways in which neurons in different layers of the network are interconnected.

- ***Feedforward Network:*** each input and relative output move forward in one direction. Connections backward or within the same layer are not permitted. This category includes Multilayer Perceptron Neural Networks (MLP). It is by far the most widely used type of network. In this work, only feedforward networks will be considered.

- ***Recurrent Network*****:** this is similar to feedforward net, but feedback connections are allowed (typically to neurons in the same layer, but also backward). This category includes specific net know as Long Short-Term Memory Network (LSTM).

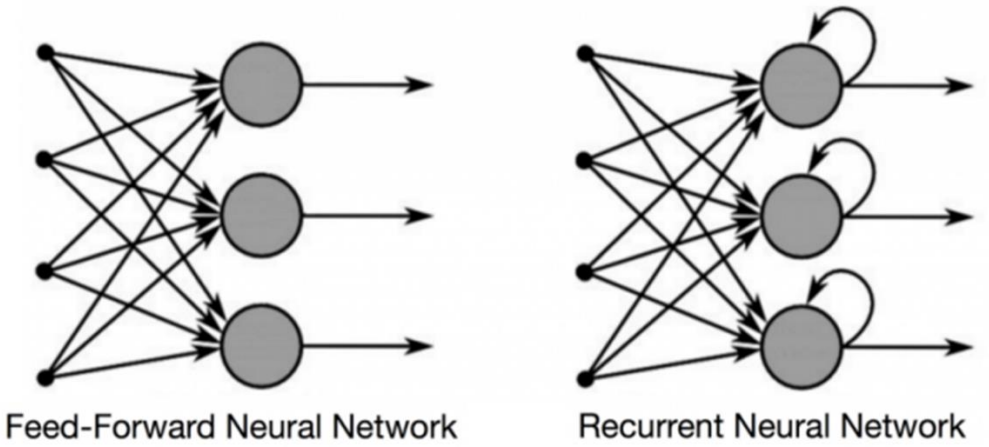

Figure 6. Feed-Forward Neural Network (FFNN) vs Recurrent (RNN). In FFNN, data move in a single direction, while in RNN, there is a loop allowing for a cyclical flow of information (Eliasy & Przychodzen, 2020).

1.3.3 Training of a Neural Network

Training a neural network is the process by which the model learns from input data during the training phase. Here is a simplified overview of how neural network learns:

1) **Weight initialization**: initially, in the neural network the weights of connections between nodes are set randomly.
2) ***Forward propagation***: during this phase, each input moves forward through the network, layer by layer. The signal propagation through the network is simulated by this process, resulting in a final predicted output.

3) ***Error calculation or loss function***: the predicted output is compared with the actual output (labelled in the training data). The difference between them represents the network classification error. The goal is to minimize the error as much as possible (section 1.3.3.1).
4) ***Backpropagation*:** once the error is quantified, it is propagated backward through the network, calculating how much each connection contributed to the error. This process allows to the network to adjust its weights reducing the error.
5) ***Optimization*:** optimization algorithms, such as gradient descent, are used to gradually update the weights of connections. The goal is to reduce the overall error, thus improving the network performance, in the fastest way. Steps 2-4 are iterated many times during the training phase, in order to achieve an adequate accuracy or stability in predictions (Section 1.3.3.3)

1.3.3.1 Loss function

Loss functions quantify the difference between the predicted output of the network and the expected output (labelled or targeted). The assessment of a loss function is fundamental in the learning process of any network. During the training process, it is important to minimize the loss function so that the network can generate increasingly accurate prediction.

Among the most frequently used loss functions there is the Mean Squared Error (MSE), which consists in arithmetic average of the squared differences of prediction errors:

$$MSE = \frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2 \qquad (2)$$

In other words, it is the sum, over all points i in the dataset N, of the squared difference between the expected value y_i (target) and the predicted model value $\hat{y}_i$.

1.3.3.2 Training and Validation Datasets

The division of the dataset into training and validation holds fundamental importance in the context of machine learning. Initially, the training set, typically comprising 70-80% of the data, is employed to instruct the model. During this phase, the model learns and adapts to complex patterns internal to the data present in the training set.

The validation set is equal in size to the remaining 20-30 % of data and serves as an evaluation set, measuring how well the model performs on data that were not previously seen. This process allows to evaluate the ability of the model to generalize to new data.

Furthermore, this division permits the prevention of overfitting, which is a frequent problem when the model is too similar to the training dataset and loses its ability to adapt to new data. Additionally, the split facilitates the optimization of the hyperparameters of the model (see Cap 1.3.3.4), allowing for a more precise configuration without excessively impacting the training set.

1.3.3.3 Optimization algorithms

Gradient descent is an optimization algorithm used in machine learning to minimize the loss function. The process involves different steps in the training phase, which are:

1) **Gradient calculation**: the gradient of the loss function with respect the model parameters is calculated, which can be visualized as the slope of a hill.

2) **Weight update**: the goal is to move along the hill to reach the lowest point (local minimum). To achieve this, the model adjusts its weights in the opposite direction of the gradient aiming at minimizing the loss function.

3) **Repeated iterations**: these steps (1-2) are repeated until the lowest point of the hill, where the error is minimized or reaches an acceptable level, is reached.

It is worth noting that while traditional gradient descent is one optimization method, there are many optimizers available, each with its own variations. Examples include Adaptive Moment Estimation (Adam) (Kingma & Ba, 2017) and Levenberg-Marquardt (Marquardt, 1963; Hagan & Menhaj, 1994).

1.3.3.4 Hyperparameters

In the context of machine learning, Sebastian Raschka and Vahid Mirjalili in their 2019 book "Python Machine Learning" helped clarifying the distinction between parameters and hyperparameters, providing a more specific definition of both concepts (Raschka & Mirjalili, 2019).

Parameters: are internal variables in the neural network that are not manually adjusted but, instead, are modified by optimization algorithms. An example, the weights of connections between neurons which are automatically adjusted to improve the network performance during training.

Hyperparameters: are external variables to the neural network that are manually chosen before the learning process begins, such as the number of epochs, batch size, and learning rate. Subsequently, based on the model performance, they are adjusted until an optimal configuration is achieved.

The **learning rate** determines how large the weight updates should be during the model training process. Essentially, it determines how quickly the neural network learns from training data. If learning rate is too high, weight updates became large, causing divergent behavior, and as a result, it does not find a stable solution. Conversely, if learning rate is too low, weight updates are very small, so it takes a long time to reach the minimum (Figure 7).

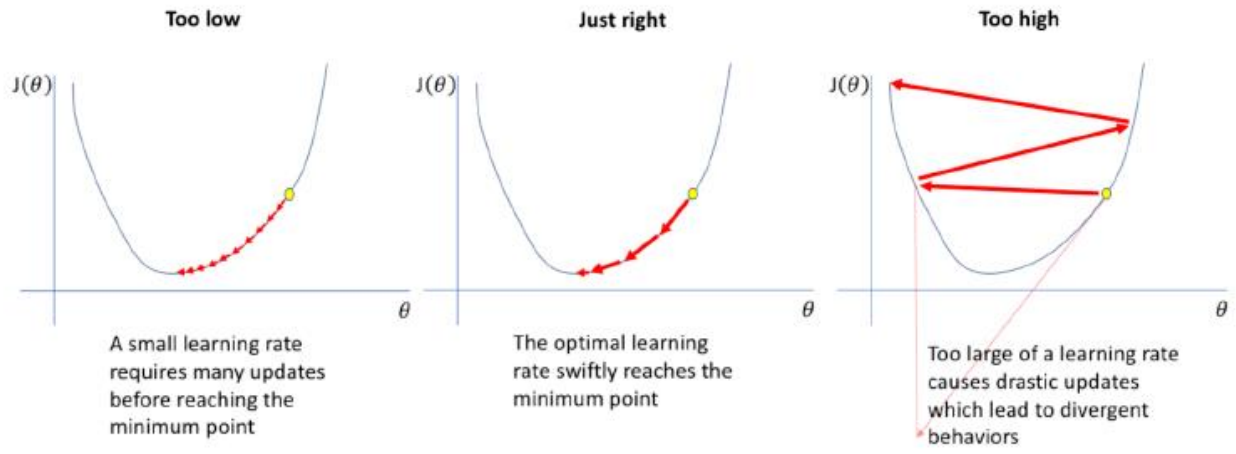

Figure 7. Comparison of learning rates in optimizing an objective function. From left to right: too low learning rate (slow convergence), correct learning rate (efficient convergence), too high learning rate (risk of oscillations and challenges in converging to the global minimum), (Jeremy Jordan website).

During the training of the network, analysing the entire training dataset can be challenging due to the large number of examples. To simplify the analysis, the dataset is divided into subsets known as batches, with each batch containing a specific number of examples referred to as the **batch size**. For instance, if we have 1000 examples and set the batch size to 500, it means that the training set will be divided into two batches, each comprising 500 examples.

With each passage through a batch of training examples, the weights of the network are updated. Once the process is completed for all batches, one epoch concludes. Therefore, the **number of epochs** represents how many times the entire training dataset has been analysed during the model training process.

Choosing the right number of epochs is important to prevent two main problems: underfitting and overfitting (Figure 8). Opting for too few epochs may prevent the network from reaching optimal performance, hence the model is not learning enough. While an excessively high number of epochs can result in the model learns the training data too well, compromising its ability to generalize on new and unseen data.

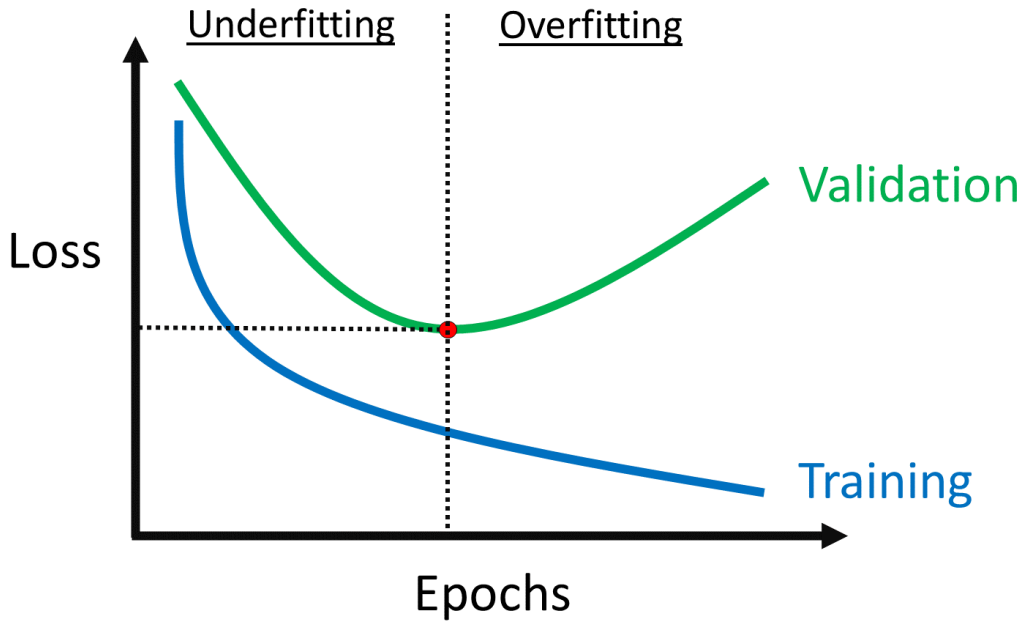

Figure 8. Analysis of the loss function during training and validation across epochs. The trend highlights that before reaching the best performance (red dot), there is underfitting, while immediately after, there is excessive overfitting problems.

1.3.4 Convolutional Neural Networks (CNN)

The roots of Convolutional neural network (CNN or ConvNet) can be attributed to the idea of "Gabor neurons" proposed by David Hubel and Torsten Wiesel, neuroscientists who studied the visual cortex in the brain (Hubel & Wiesel, 1968). This concept inspired Yann LeCun, one of the pioneers of CNNs, who developed LeNet-5, one of the earliest CNNs, to recognize handwritten digits in postal codes (LeCun et al. 1998).

In the last ten years, there has been a rapid development of CNNs, and currently, these networks are extensively utilized in the field of deep learning (Krizhevsky et al., 2017; Zhou, 2020; G. Li et al., 2021). The primary advantage is its ability to automatically extract relevant features directly from data. In this sense, CNN have shown to be superior compare to other artificial neural networks (Gu et al., 2017) especially in some areas such as object detection, image classification, computer vision (Fang et al., 2020), speech processing (Palaz et al., 2019), face recognition (H.C. Li et al., 2020) and segmentation. Also, CNNs have found extensive practical application across diverse fields, such as autonomous vehicles, security camera system and other.

A CNN is designed in such a way that the "neurons" within a specific layer share the same "weights" and "biases". This implies that each neuron is specialized in detecting a particular feature, such as an edge or

spot, regardless of their exact position in the image (source Matlab). The processing of input data in a CNN is divided into two important steps, as show in Figure 9:

• In the first one (feature learning) the convolutional and pooling layers are responsible for extracting relevant features from the input image.

• The second process (classification) consists of a fully connected traditional neural network layers to classify the image optimally based on the input of extracted features received.

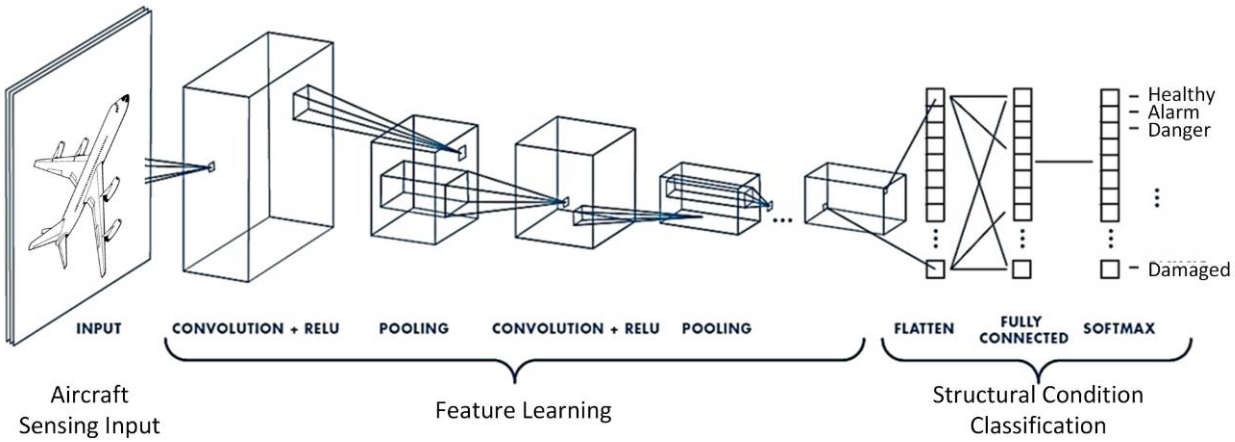

Figure 9. Example of a convolutional neural network with multiple convolutional layers, activation function (ReLU), and pooling used to extract features. Finally, a fully connected layer is employed to classify the image (Tabian et al., 2019).

Below, we describe the three main types of layers in CNN architecture and how they work:

Convolutional layer

In CNNs, structured data such as images is represented by matrices. During the convolution process, a weight matrix called a filter (kernel) moves along the image, multiplying values corresponding to pixels. The step of this filter can be adjusted through the "stride" function. Therefore, convolution is used to process the input through filters to detect specific features in the image.

However, performing this convolution operation may lead to information loss, especially at the image borders. To address this, values greater than or equal to zero are added, a practice known as "padding" (Figure 10). This helps maintain image dimensions and reduces information loss.

After each convolution operation, an activation function is applied (see section 1.3.2.3), allowing the network to capture non-linear relationships in the data.

It is important to note that during the convolution process, there is a risk of overfitting if the feature maps become overly complex (Hawkins, 2004). To mitigate this issue, the pooling layer is applied, helping to simplify the extracted features reducing the number of parameters (Taye, 2023) and increasing the generalization capability of the model and consequently reducing the risk of overfitting (Zafar et al., 2022).

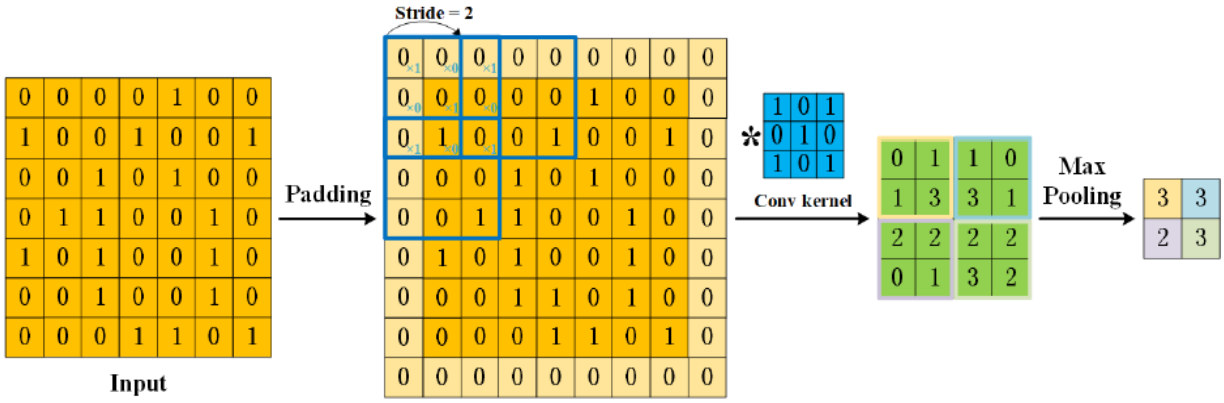

Figure 10. Diagram illustrating the CNN process: the input matrix is exposed to zero padding and convolution with a kernel filter [3x3], utilizing a specific stride (=2), followed by max pooling (Li et al., 2022).

Pooling layer

There are two pooling techniques (Figure 11):

- *Max pooling*: as the filter moves across the input, it selects the pixel with the maximum value to send to the output array. This approach is more commonly used than average pooling.
- *Average pooling*: as the filter moves across the input, it calculates the average of the pixel values.

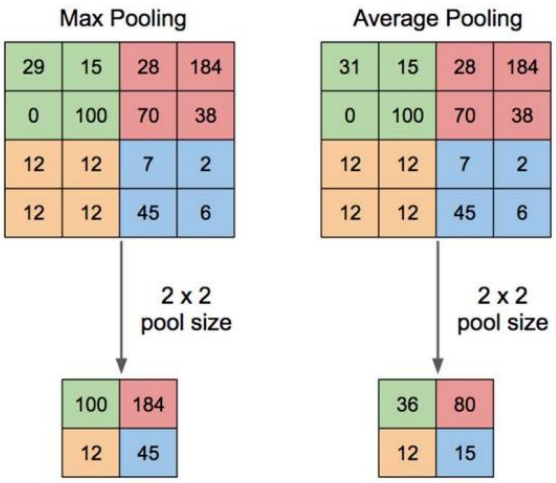

Figure 11. Comparison of max pooling and average pooling operations on a 4x4 input matrix. Using a 2x2 pixel filter, max pooling captures the maximum value within each filter, creating a new output matrix. Conversely, average pooling computes the mean value for each 2x2 filter, resulting in a distinct output matrix on the right (Yani et al., 2019).

Fully connected layer

This layer, positioned at the final stage of CNNs, functions to flatten the feature maps (matrices) derived from the convolution and pooling layers into a one-dimensional vector. The dimensions of this vector align with the number of classes in the classification problem. Probabilities linked to each class are usually derived by

applying an activation function, such as Softmax. The final predicted value corresponds to the class with the highest probability score.

1.3.4.1 Different CNN architecture

In the past decade, as computational power increased and more data became available, CNNs made a strong comeback. In 2012, AlexNet's CNN won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), demonstrating the success of CNNs in large-scale image recognition. Since then, different versions and architectures have kept performing well in later ImageNet competitions. Models like GoogLeNet (Szegedy et al., 2014), VGGNet (Simonyan & Zisserman, 2015) and ResNet (He et al., 2016) have led to significant advances in the field of image recognition. Below are reported some architectures of CNNs developed over time (Li et al., 2022):

1998: LaNet5
2012: AlexNet (Krizhevsky et al., 2017)
2013: ZFNet (Zeiler & Fergus, 2013)
2014: VGG-NET or Visual Geometry Group, GoogLeNet
2015: Inception V2 (Ioffe & Szegedy, 2015) and Inception V3 (Szegedy et al., 2015) , ResNet or Residual Network, U-Net (Ronneberger et al., 2015)
2016/2017: SqueezeNet (Iandola et al., 2016)
2017: ResNext (Xie et al., 2017), Xception (Chollet, 2017), DenseNet (Huang et al., 2017)
2018: MobileNetV2 (Sandler et al., 2018, 2019), ShuffleNet (Zhang et al., 2018)
2019: MobileNetV3 (Howard et al., 2019)
2020: GhostNet (Han et al., 2020)

1.3.4.2 The SqueezeNet architecture

SqueezeNet, introduced in 2016 (Iandola et al., 2016), is a convolutional neural network and its primary innovation lies in the "fire modules", consisting of squeeze and expand layers. These modules use 1x1 and 3x3 convolutions to reduce the number of parameters (Figure 12 top) maintaining high accuracy. The SqueezeNet architecture consists of an initial convolutional layer, followed by 8 fire modules (each containing a squeeze and expand layer), then another convolutional layer, and finally, a softmax layer (Figure 12 bottom). One challenge with using deep neural networks, especially those with lots of layers like AlexNet and VGGNet, is that they need a lot of computational power. However, SqueezeNet has found a balance between accuracy and using memory and processing power more efficiently (Iandola et al., 2016).

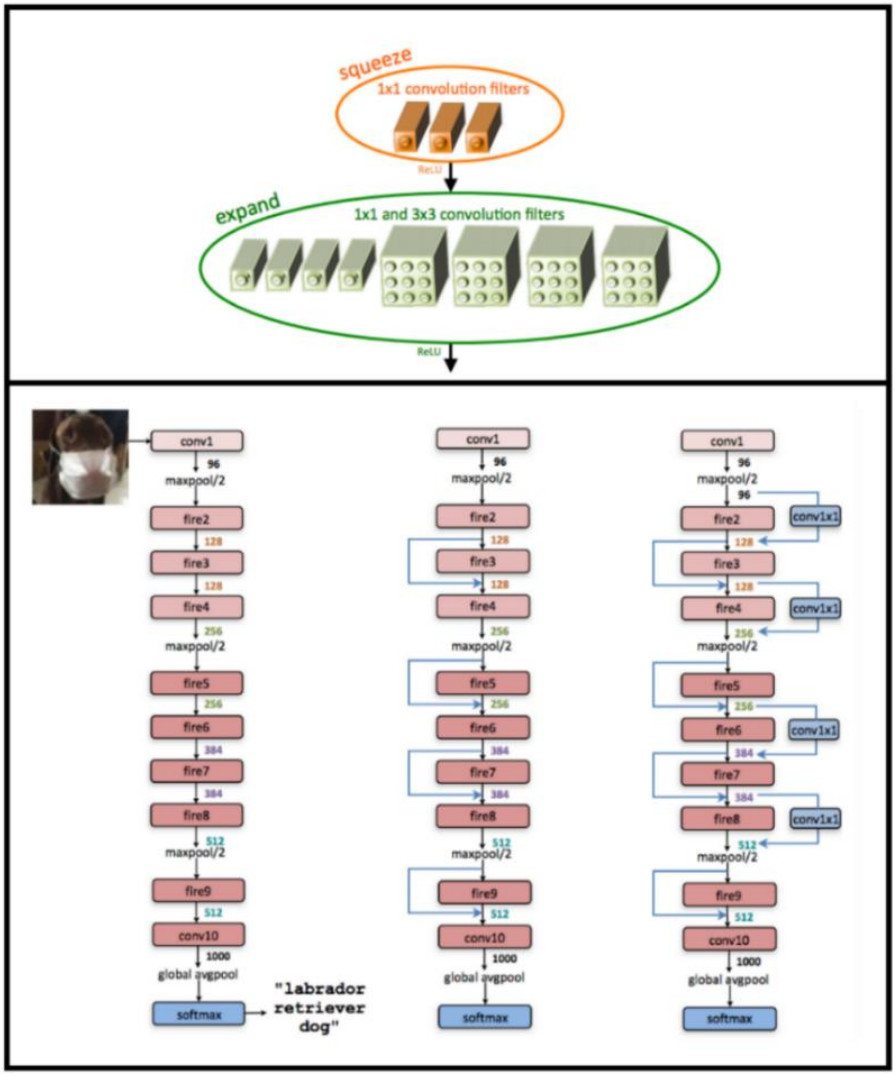

Figure 12. (Top) Fire modules divided into two parts: the 'squeeze' layer consists of a 1x1 convolutional layer, while the 'expand' layer comprises 1x1 and 3x3 convolutional layers. (Bottom) SqueezeNet architecture (Iandola et al., 2016), featuring 'conv' layers (convolutional), 'fire modules,' as well as maxpool and globalavgpool used for dimension reduction. In max pooling, the maximum value in a window is selected, while in global average pooling, the average of the output is computed, respectively.

1.3.4.3 The U-NET architecture

U-Net, proposed in 2015, is a deep neural network architecture primarily used for the detection and segmentation of biomedical images (Ronneberger et al., 2015). The architecture is characterized by a U-shaped structure (Figure 13) where in the descending path of the network, known as the encoder, dimensions are gradually reduced through convolution and pooling layers, capturing the most relevant features. In the ascending path, called the decoder, the expansion of the extracted features generates the final output, restoring the original dimensions. Typically, U-Net uses activation functions such as ReLU, or in our case, leakyReLU, in the internal layers of the network, including the final layer which usually employs a sigmoid function.

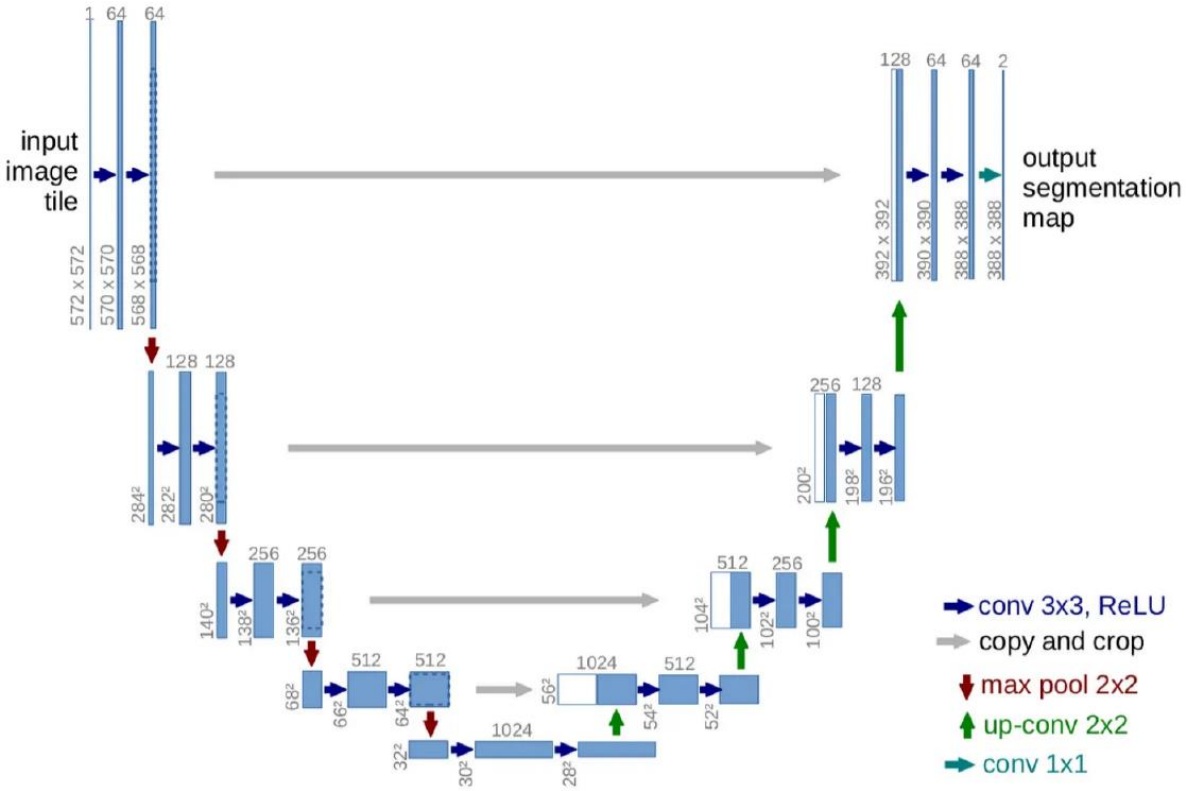

Figure 13. In the diagram, each blue box on the left corresponds to a multi-channel feature map. The x-y dimensions are indicated in the bottom-left corner of these boxes. These feature maps are then processed through various operations, including convolution and other transformation layers, as indicated by the arrows. On the right, white boxes represent the replicated feature maps, and these features are reassembled through decoding layers (Ronneberger et al., 2015).

1.4 Machine learning in Seismology

Artificial neural networks, introduced for the first time in 1943 by Warren McCulloch and Walter Pitts, landed in applied geophysics in the late 1980s. Recent applications of artificial neural networks in geophysics include, e.g., the prediction of site amplification response (Bergamo et al., 2021; Boudghene Stambouli et al., 2017; Salameh et al., 2017) and seismic site classification related issues (Zhao, 2006; Ghasemi et al., 2009; Yaghmaei-Sabegh & Tsang, 2011; Ji et al., 2022; Díaz et al., 2022; Vantassel et al., 2023).

Although machine learning has had a presence in seismology for decades (e.g., Turhan Taner et al., 1988; Dowla et al., 1990), its application has experienced rapid growth in recent years. This is particularly evident in the utilization of ML for primary processing tasks aimed at extracting information about earthquakes and related phenomena. The detailed overview of machine learning applications in seismic contexts presented here is a valuable extract from the contributions of Mousavi & Beroza (2023). In their work, the authors examine the broad spectrum of innovative uses of machine learning in the field of seismology. Mousavi & Beroza (2023) highlight how machine learning is revolutionizing multiple aspects of seismic monitoring and understanding seismic phenomena. From event detection to forecasting and ground motion

characterization, this summary provides a snapshot of the main areas of study and application of this advanced technology in seismology, building on the foundation provided by the research of the two authors. Below is a summary of the main applications that have emerged from their research:

Seismic Event Monitoring:

- Classification of seismic events into explosions versus earthquakes.
- Identification of volcano-seismic events.
- Differentiation of earthquakes caused by mining activities.

Seismic Signal Detection:

- Utilization of machine learning algorithms for the automatic detection of seismic signals.

Seismic Phase Picking:

- Implementation of machine learning for the automatic selection of seismic phases in the data.

Polarity Determination:

- Use of machine learning algorithms to determine the polarity of seismic signals.

Phase Association:

- Machine learning-based approaches for phase association, encompassing both travel time-based and waveform-based methods.

Seismic Source Parameterization:

- Machine learning-based methods for seismic source parameterization, including single-station, multi-station, dynamic multi-station, physics-informed methods, and earthquake source mechanism.

Seismogram Simulation:

- Application of machine learning algorithms for the simulation of seismograms.

Ground Motion Characterization:

- Implementation of machine learning applications for the characterization of ground motion during seismic events.

Earthquake Forecasting:

- Machine learning approaches for earthquake forecasting.

Machine Learning for Exploratory Research:

- Direct investigation of seismic waveform patterns using machine learning.
- Utilization of machine learning as a post-processing tool.
- Applications of machine learning as a tool for exploratory research.

Neural networks have been a useful tool in automatic seismic data processing since the early '80s, primarily utilized in supervised methods, as highlighted by Mousavi & Beroza (2023). While supervised learning remains prevalent, recent trends indicate a growing interest in inverse problems within neural network applications (see Fig. 14 a, b). The field of machine learning offers avenues for data-driven discovery, imaging, and pattern interpretation in seismic data within a high-dimensional space (Mousavi & Beroza, 2023). Notably, the application of unsupervised approaches and Generative Adversarial Networks (GANs), is achieving prominence, particularly in the exploratory analysis of high-dimensional seismic data (refer to Fig. 14 c). Multilayer perceptron and CNNs stand out as the primary neural network types applied in seismological applications. In the bar chart, seismic categories are represented on the y-axis, and the number of published papers is on the x-axis. The different bars are coloured based on the specific machine learning approaches employed, highlighting our significant contribution in these two specific areas of seismology.

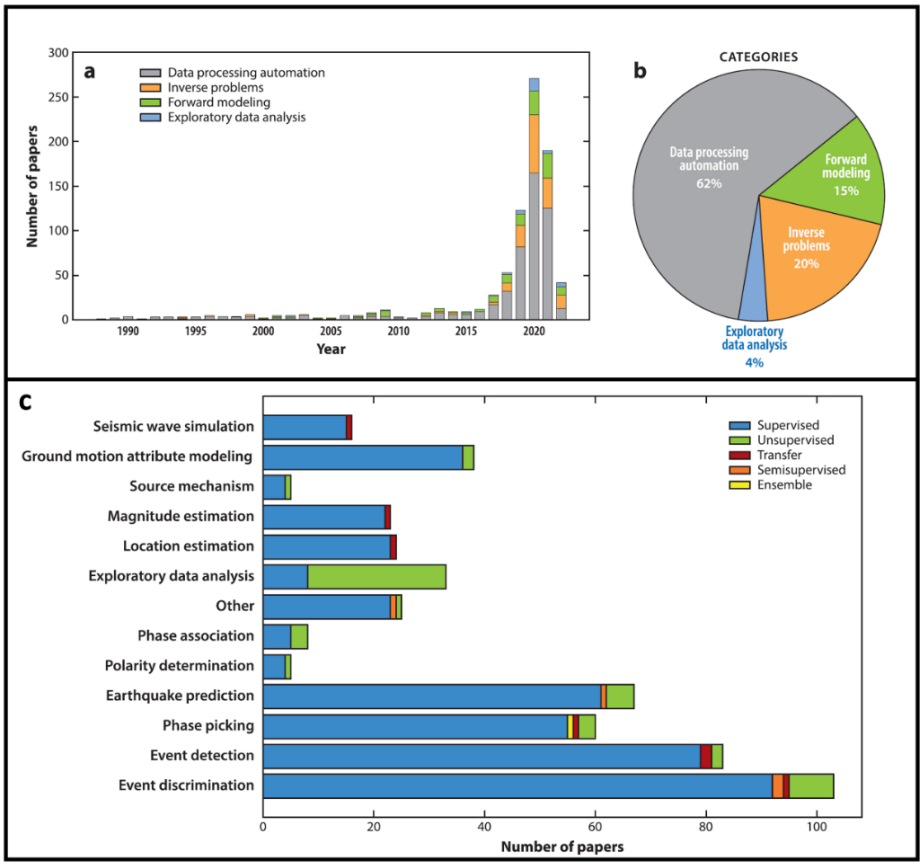

Figure 14. Overview of machine learning (ML) applications in seismology. (a) Bar chart depicts the number of journal publications for ML methods from January 1988 to May 2022, color-coded by task category. (b) Pie chart represents the percentage distribution of machine learning applications across various seismological categories. (c) Distribution of machine learning approaches across different seismological tasks, based on a review of scientific papers (Mousavi & Beroza, 2023).

1.4.1 Transfer Learning

In the context of our study, we adopted an advanced strategy to address the challenges associated with having a relatively limited dataset. To maximize the potential of our approach, we employed a pre-trained neural network, specifically the pre-trained SqueezeNet. This net, originally trained on more than a million images from the ImageNet database (Russakovsky et al., 2015), it is an ideal candidate for transfer learning, especially when there's limited training data available.

Our approach follows some phases. Initially, we froze the weights of the already trained layers of SqueezeNet. This step allowed us to preserve the knowledge acquired during the initial training on ImageNet. Subsequently, we extended the network by adding new layers, including a final layer specifically designed for recognizing distinctive nature of HVSR peaks (see Chap. 2). Additionally, in this phase, we also need to set hyperparameters and number of classes to match our new dataset.

Our goal was to fine-tune the network, training it on a limited and specific dataset. This approach helped us use the existing knowledge in SqueezeNet, making training faster and improving the network ability to recognize the nature of the HVSR peak (see Chap. 2). Therefore, we could summarize as follows:

1) Take the layers from the pre-trained original model.
2) Freeze the layers to preserve the information they contain during future training phases.
3) Add some new trainable layers on bottom of the frozen layers.
4) Train the new layers with the dataset for the specific problem.

1.4.2 Machine learning programming languages

In the vast field of machine learning there are different programming languages but for our project, we mainly used Python and MATLAB®.

We chose to use MATLAB® in our work to create and train two type of neural networks, artificial neural network and pre-trained SqueezeNet. The choice of MATLAB® is driven by fact that offers an interactive and user-friendly interface. We utilized integrated libraries for deep learning, such as the Deep Learning Toolbox, which includes pre-trained models like SqueezeNet. This allowed us to save time in the development phase, but we also customized the solution by training a neural network without using pre-existing models with 7 hidden layers to discriminate the nature of the HVSR peaks (see Chap. 2).

On the other hand, Python offers numerous specialized libraries and frameworks, including Scikit-learn, Tensorflow and PyTorch. In our work, we used Tensorflow, a framework developed by Google in 2014 that

simplifies the creation of deep learning models. Tensorflow provides a rich set of features, utilizing Keras in particular for the definition and training of neural networks through its Application Programming Interfaces (APIs), which serve as the interface between the user and the application. Through powerful libraries such as TensorFlow, we have successfully implemented U-Net neural networks, to create ground motion model (see Chap. 3).

In this context, we aim at providing a contribution to two subsets of the site effect assessment branch of research, and specifically to some aspects related to the HVSR (Horizontal to Vertical Spectral Ratio) technique and GMM (Ground Motion Modelling).

In the first case, we employed Artificial Neural Networks (ANN) and a pre-trained SqueezeNet to separate stratigraphic from non-stratigraphic H/V peaks. This is important to capture the features of H/V curves that are really linked to the soil properties and not to external interference (see Chapter 2).

In the second case, we used a U-Net neural net capable of generating region-specific, fully non-ergodic, and data-driven GMM, an approach not previously explored in the Italian context (See Chapter 3).

Chapter 2. First seismological application of NN: the Horizontal to vertical spectral ratio (HVSR)

In this chapter, we introduce the single-station and passive seismic technique known as H/V or HVSR (Horizontal-to-Vertical Spectral Ratio). This geophysical survey, based on measuring ambient seismic noise and using the spectral ratio between horizontal and vertical components, provides soil resonance and the mapping of seismo-stratigraphic reflectors in the subsurface. It should be noted that H/V curves often exhibit artefactual peaks of unclear origin. Currently, there are no defined criteria for discriminating between the stratigraphic and non-stratigraphic (anthropogenic or artefactual) nature of them. To address this challenge, we have employed two different types of neural networks: artificial neural network and convolutional neural networks.

2.1 Ambient seismic noise

The term "seismic noise" refers to ground vibrations generated by various sources, containing both natural and anthropogenic origins. In various studies, Gutenberg (1911, 1931, 1936, 1958), acknowledged that the origin of seismic tremors below 1 Hz is predominantly natural. In addition to seismic noise caused by oceanic wave activity, which generally involves frequencies much lower than 1 Hz natural sources of seismic tremors also include atmospheric events. Examples of these events are atmospheric pressure variations, wind (wind friction on rough terrain, trees, or constructed objects oscillating or vibrating due to the wind), rushing waters (waterfalls or rapids in rivers and streams), and so on. Subsequent relevant studies further explored the origin of the tremors such as Haubrich et al. (1963) and Longuet-Higgins (1950).

The spectral amplitude both at large and low frequency is ruled by the source intensity: severe weather will determine large-amplitude spectra at low frequency; intense human activity (typical at daytime in urban centers) will determine large amplitude spectra at high frequency.

2.2 Origin of HVSR technique

In the 1970s, the Japanese researchers Nogoshi and Igarashi (Nogoshi & Igarashi, 1970) observed that while the spectra of individual motion components were affected by the seismic source, their ratio maintained temporal stability with site-specific characteristics. This observation gave origin to the H/V technique. Subsequently, Nakamura (1989) contributed to its popularity, interpreting it as an experimental method to assess the resonance frequencies of the subsurface. It involves determining the spectral amplitude ratios between the horizontal components (H) and the vertical component (V) of ambient vibrations measured at a specific point using a three-component seismometer. The spectral ratio serves as an effective normalization tool, reducing source variability and accentuating the path response (i.e., the site effect). The outcome of

such a test is an experimental curve (Figure 15) that shows the average ratio between the horizontal and vertical spectra of ambient noise. The peaks in the H/V curve are generally interpreted as the resonance frequencies of the ground beneath the measurement point (SESAME, 2004).

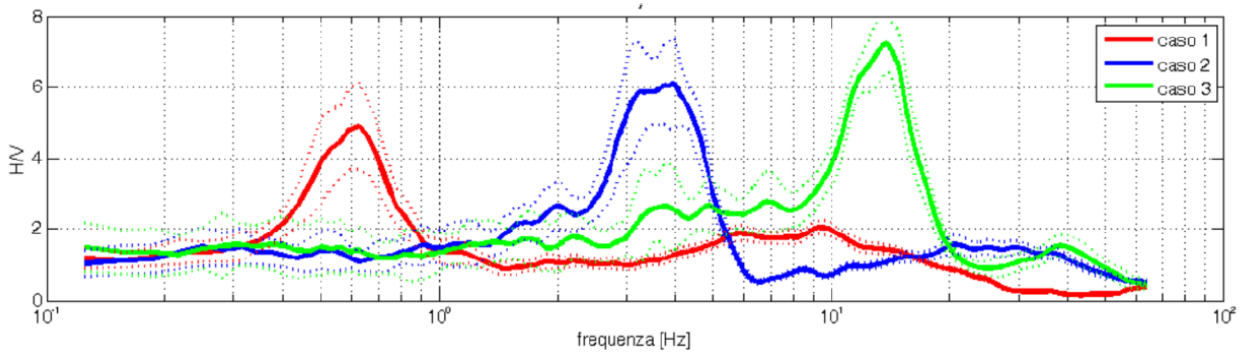

Figure 15. Example of three H/V curve. The peaks indicate seismic bedrock at various depths, generating resonances at different frequencies (Albarello & Castellaro, 2011)

H/V curves play a crucial role in investigating subsurface structure, providing information about the presence of rigid layers through peak amplitudes and relative depth through peak frequency. Moreover, spatial variations in the curves typically indicate lateral subsurface heterogeneities. In summary, the H/V test serves as a valuable method for studying seismic noise for geotechnical purposes, characterized by its simplicity and quick application, though requiring particular caution in interpretation.

This method has been employed by various researchers for stratigraphic imaging, both at shallow depths (Parolai, 2002; Castellaro & Mulargia, 2009a; Gosar & Lenart, 2010; Mantovani et al., 2018) and at greater depths (e.g., Ibs-von Seht & Wohlenberg, 1999; Zor et al., 2010; Scheib et al., 2016 Sgattoni and Castellaro, 2020). Over time, it has been observed that an accurate interpretation of the H/V curve requires the ability to distinguish between H/V peaks of stratigraphic origin and those of anthropogenic origin (Castellaro & Mulargia, 2010; Castellaro, 2016; Molnar et al., 2018). Before investigating into the interpretation of H/V curves, two aspects need to be considered:

- The H/V curve should exhibit limited standard deviation in amplitude and frequency, requiring a cleaning of the trace by removing windows where the ratio is distorted.
- The H/V curve should never be observed alone but always together with the individual spectral components from which it derives. As we will see, this is what permits to differentiate between peaks of stratigraphic and non-stratigraphic nature.

[*The following part of the chapter is rewritten following Di Donato, Castellaro (2024). Performance of different ANNs in microtremor H/V peak classification, Seism. Res. Lett., under publication*]

2.3 Introduction

Here we focus on microtremor H/V curves. The identification of significant peaks therein is traditionally assessed by means of the SESAME (2004) criteria, which, however, cannot discriminate between the stratigraphic and non-stratigraphic (anthropogenic or artefactual) nature of them. SESAME (2004), states that ''*if the microtremor spectra components exhibit a sharp peak (often on the 3 components together), at a particular frequency, there is a 95% chance that this is anthropic "forced" ambient vibration and it should not be considered in the interpretation*''. However, the SESAME (2004) criteria do not provide specific rules to reject these artefacts.
Recently, Wang et al. (2023) proposed an identification protocol for H/V peaks based on a regression tree algorithm. This algorithm, however, does not even consider that H/V peaks can have stratigraphic or non-stratigraphic origins and that their shape is very different in the two cases.

Non-stratigraphic spectral peaks are typically caused by machineries acting in the proximity of the recording station (e.g. wind-mills, industrial machines; Dreossi et al., 2018) or result as an effect of vibrations induced from large nearby structures (tall buildings and bridges or viaducts; Castellaro & Mulargia, 2010; Castellaro et al., 2022). Low frequency (i.e., large wavelength) artefacts can perturb the noise wavefield up to very large distances from the source. They have been documented up to even a few km distance from their source (Bokelmann & Baisch, 1999; Cornou et al., 2004, Dreossi et al., 2018).
The frequency of anthropogenic peaks can also potentially coincide with the frequency of stratigraphic peaks.

H/V curves are not infrequently affected by artefactual peaks, and it is mandatory to discriminate them, before attempting any interpretation of the H/V curves themselves. This is a relevant issue if one considers the worldwide extended use of the H/V technique in seismic microzonation, in seismic soil classification (where it is commonly combined with surface wave multichannel methods to get Vs profiles) and in the classification of soils around seismic stations, which influence the signal recorded by them.
The H/V curves produced by automatic processing of 3C-microtremor recordings from permanent seismic stations would also require automatic algorithms for real-time or off-line identification of non-stratigraphic peaks.

Given the limited capability of standard statistical techniques in this type of signal classification, due to the strong variability of the geometries of non-stratigraphic H/V peaks, we decided to explore the performance

of different artificial neural networks in discriminating between stratigraphic and non-stratigraphic features in microtremor signals.
In the following sections, we first present the features distinguishing stratigraphic from non-stratigraphic H/V peaks, then we illustrate the two types of neural networks trained to this aim, the performance of both and their limitations.

2.4 Stratigraphic vs. non-stratigraphic H/V peaks

Microtremor H/V peaks can be due to the presence of local maxima (peaks) in the horizontal spectral components of motion, to the presence of local minima (troughs) in the vertical component of motion or to the occurrence of both conditions, with different amplitudes.

Peaks in the horizontal components of motion (Figure 16a, b) are typically associated with body wave resonances (e.g., Nakamura, 1989; Van Der Baan, 2009; Castellaro, 2016; Oubaiche et al., 2016; Molnar et al., 2018; Sgattoni & Castellaro, 2020; Molnar et al., 2022; Castellaro & Musinu, 2023). Troughs in the vertical spectral component motion (Figure 16c, d), occurring at or close to the resonance frequency of SH waves, f_0, are instead typical of surface wave trapping phenomena (Fäh et al., 2001; Bonnefoy-Claudet et al., 2008; Castellaro & Mulargia, 2009a; Endrun, 2011; Tuan et al., 2011; Castellaro, 2016; Molnar et al., 2022).
A local maximum in the vertical spectral component follows at about twice the frequency of the trough (Konno & Ohmachi, 1998; Fäh et al., 2001; Tuan et al., 2011; Castellaro, 2016).
Thus, the H/V curves typically exhibit a peak at the frequency of the vertical trough f_0 and the H/V ratio returns to its non-amplified value (H/V $\sim$ 1) at about $2f_0$ (Figure 16, Tuan et al., 2011; Castellaro, 2016).

Another distinctive feature of H/V curves acquired in 1D conditions is that the horizontal components are statistically identical through the whole frequency spectrum as clear in Figure 16. Outside the H/V peak frequency interval, the horizontal and vertical spectra are again statistically identical in amplitude (thus, H/V$\sim$1).
For the meaning of H/V < 1 in wide frequency intervals and for more examples, see, e.g., Castellaro & Mulargia (2009b). This pattern is interpreter as signature of “velocity inversion” and it can be encountered naturally (e.g. gravel overlying silt/clay etc.,) or with presence of anthropic structures (e.g., paving, asphalting above silt/clays etc.,).

H/V peaks with marked troughs in the vertical component dominate in 1D conditions and in low aspect ratio basins, while H/V dominated by local maxima in the horizontal components of motion are typical in markedly 2-3D conditions, like large aspect ratio basins (Bard & Bouchon, 1980a; Roten et al., 2006; Le Roux et al., 2012; Ermert et al., 2014; Sgattoni & Castellaro, 2020; Lattanzi et al., 2023; Sgattoni et al., 2023).

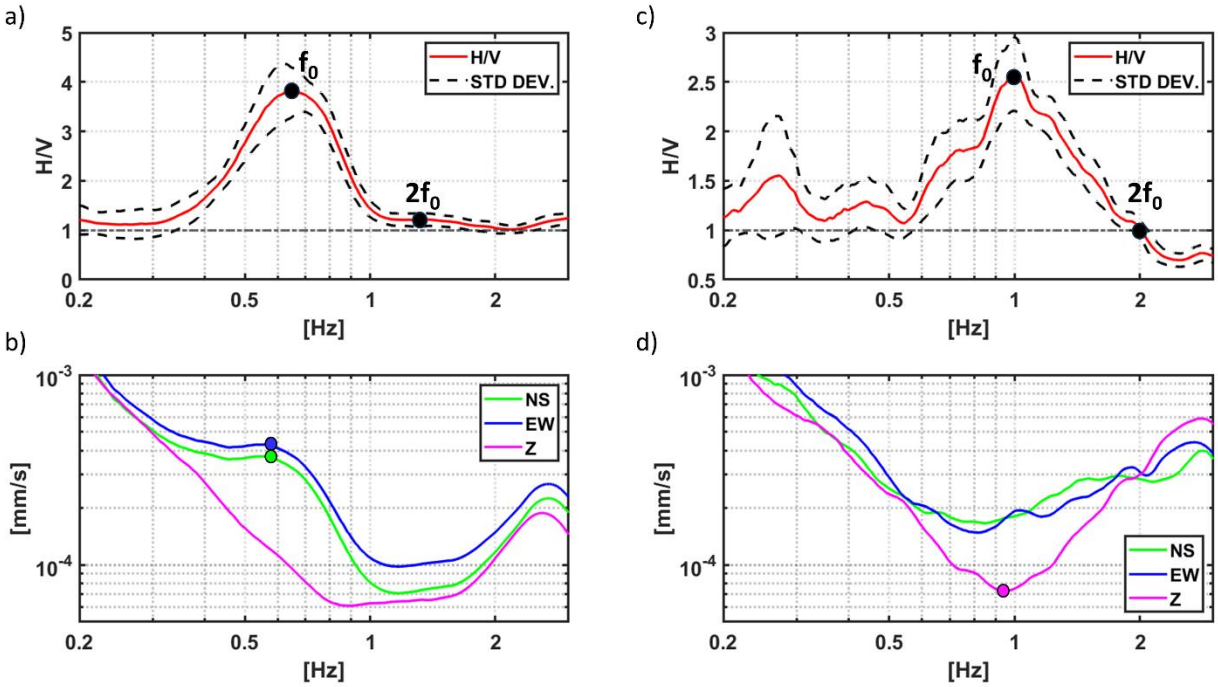

Figure 16. H/V peaks of stratigraphic origin: a, c) H/V curves and b, d) corresponding individual spectra of motion in velocity. Stratigraphic peaks occurring at f_0 typically 'close' around $2f_0$ (with a variability depending on the Poisson's ratio, Tuan et al., 2011). Here the H/V peaks are observed at $f_0 = 0.65\ Hz$ (a) and $f_0 = 1\ Hz$ (b) and the H/V curve turns back to H/V ~ 1 at $\sim 2f_0$ = 1.3 Hz and 2 Hz, respectively. The stratigraphic origin is marked by a local minimum in the vertical spectral component (more evident in panel d) and/or local maxima in the horizontal spectral component (more evident in panel b).

The SESAME (2004) guidelines introduced quantitative criteria to assess the reliability and clarity of peaks in H/V curves. These criteria were mostly based on the statistical analysis of the frequency and amplitude peak stability over time and on the expected geometrical shape for H/V peaks. No other rules were present in the SESAME criteria to discriminate between stratigraphic and non-stratigraphic peaks, beyond the general comment presented in the introduction.

Non-stratigraphic H/V peaks are typically generated by narrow spikes (almost Dirac's deltas, when non-smoothed) visible, with different amplitudes, on all spectral components (Figure 17). Comparing the smoothed (Figure 17b) and non-smoothed (Figure 17c) spectra makes the acknowledgement of artefactual peaks generally straightforward (Castellaro & Mulargia, 2010).

A non-stratigraphic H/V peak at frequency f_0 does not return to H/V ~ 1 at $2f_0$ but well before this value.

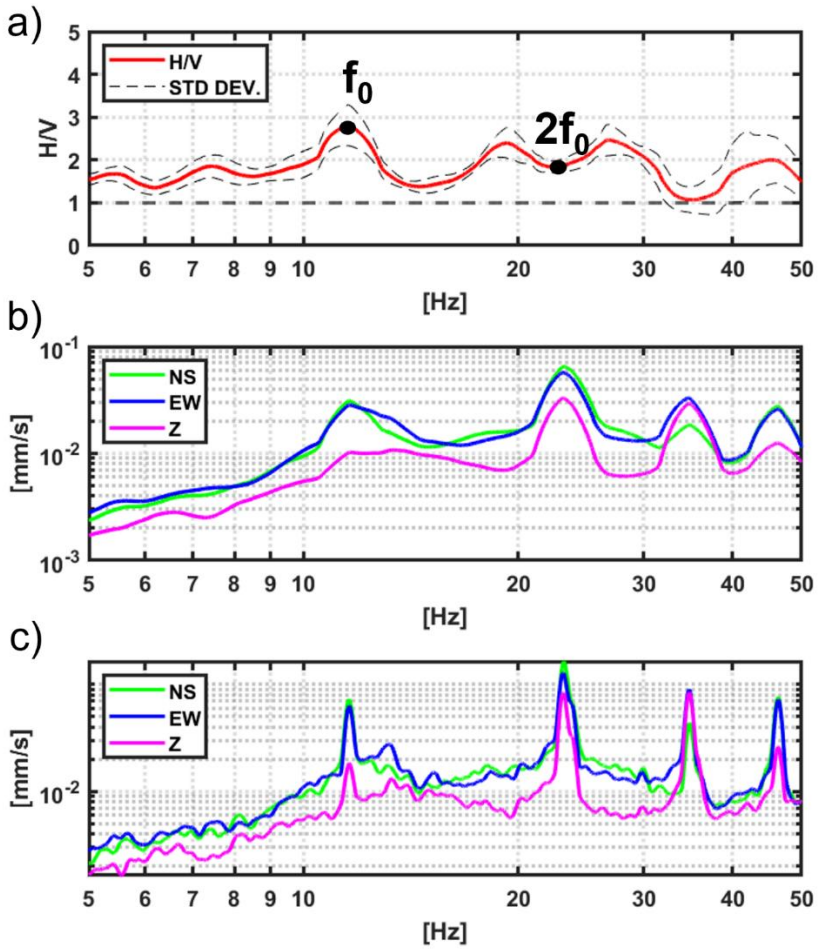

Figure 17. Non-stratigraphic H/V peaks induced by industrial machineries at a few hundred meters distance: a) H/V curve and b) corresponding amplitude spectra of the three components of motion with a 10% (of the central frequency) smoothing. c) is as b) but non-smoothed. The spectral peaks are very sharp in this case and have different amplitudes in the different components. This produces artefactual peaks and troughs in the H/V curve, where the distance between the H/V peaks and the return to H/V ~1 is much smaller than $[f_0, 2f_0]$ *in all cases.*

Other types of non-stratigraphic H/V peaks are those induced by the vibration of large nearby structures or infrastructures (e.g., bridges). They have the same features already discussed for Figure 17, but they typically occur at different frequencies along different components, depending on the shape of the vibrating structures. These can be detected also by means of azimuthal processing (see e.g., Matsushima et al., 2014; Cheng et al., 2020). An example is given in Figure 18.

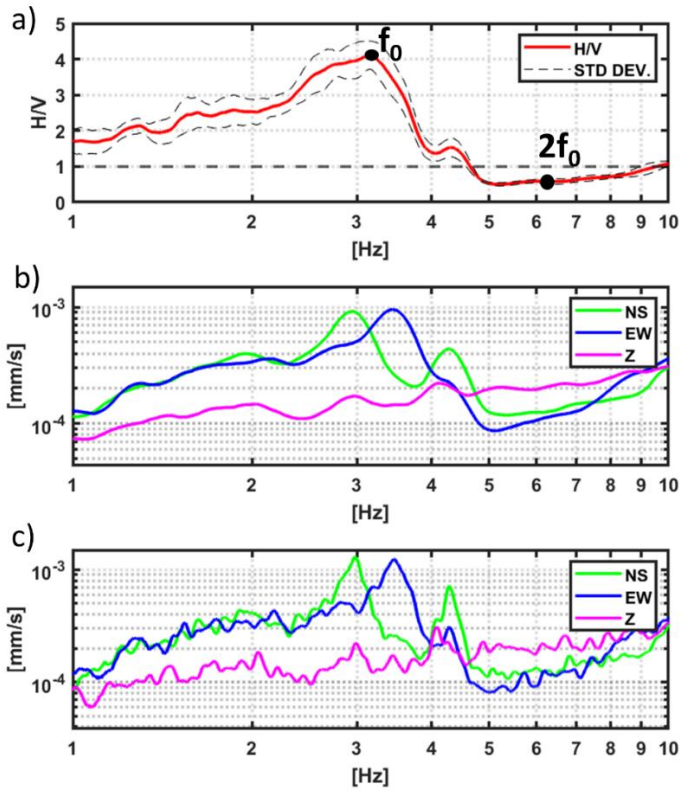

Figure 18. Non-stratigraphic H/V peaks induced by a viaduct with 30 m heigh pillars at approximately 80 m distance: a) H/V curve and b) corresponding amplitude spectra of the three components of motion with a 10% (of the central frequency) smoothing. c) is as b) but non-smoothed. The narrow spikes, occurring at different frequencies and with different amplitudes along different directions, have a clear non-stratigraphic origin.

In Table 1 we summarize the main features distinguishing stratigraphic from non-stratigraphic H/V peaks.

Table 1. Main differences between stratigraphic and non-stratigraphic H/V peaks.

STRATRIGRAPHIC H/V PEAKS (1D CONDITION)	**STRATRIGRAPHIC H/V PEAKS (non-1D CONDITION)**	**NON- STRATIGRAPHIC H/V PEAKS**
- Statistically coincident horizontal spectral components - H/V peak at frequency f_0, H/V trough at approx. $2f_0$ - the horizontal components can be flat at f_0 or show a local maximum - local minimum in the vertical component at f_0, local maximum at approx. $2f_0$ - H/V peaks are thus broad - H/V $\sim$ 1 beyond the peak frequency range	- Horizontal components show peaks at different frequencies - As a consequence, the H/V peaks appear even broader than in the 1D case	- Narrow peaks with different spectral amplitudes and/or frequencies on different spectral components - better detectable in non-smoothed spectra or by using azimuthal processing - the corresponding H/V peaks (or troughs) appear much narrower than in the stratigraphic cases

In some other cases the spectral pattern behind H/V peaks is not fully clear: sometimes due to installation instabilities of the sensor. An example is given in Figure 19, where the H/V peak passes the SESAME criteria, the average NS component is larger than the others in the whole frequency domain and the other spectral components show patterns that cannot be fully reconducted to the cases described in Table 1. We will classify these cases as “other”.

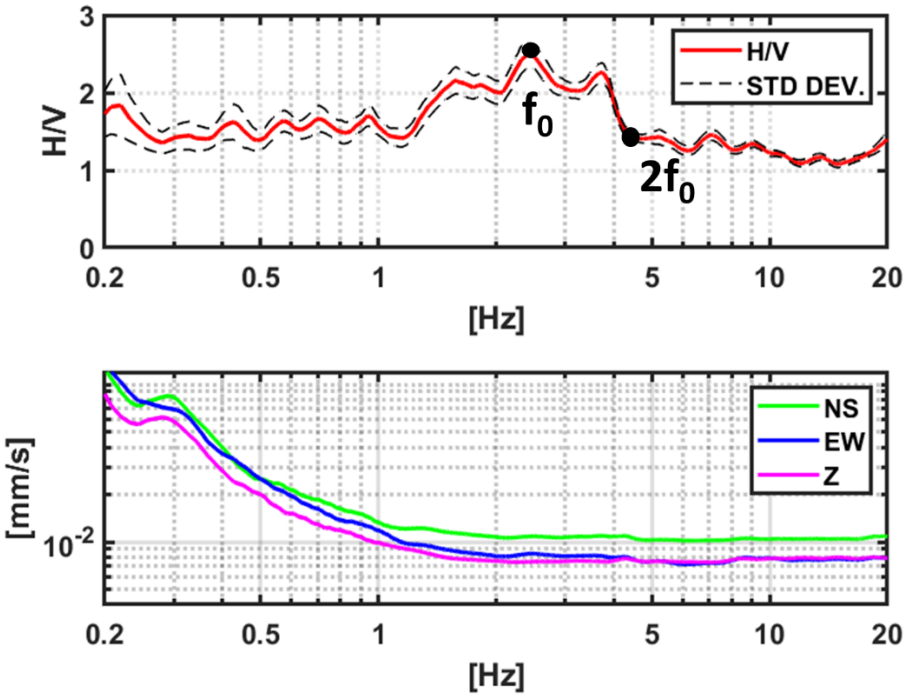

Figure 19. H/V peaks of unclear origin. The NS spectral component is amplified compared to the EW component in a wide frequency interval and no clear spectral pattern as those summarized in Table 1 can be acknowledged. The H/V curve was obtained by averaging the H/V obtained from several 30 s windows; the NS_{avg}, EW_{avg}, Z_{avg} spectra were obtained by averaging the spectra from the same 30 s windows. The H/V curve in the upper panel is therefore not the result of the division of the $\frac{\sqrt{NS_{avg} * EW_{avg}}}{Z_{avg}}$ spectra shown below.

2.5 ANN and CNN approaches

Distinguishing stratigraphic from non-stratigraphic H/V peaks for a human eye is often, though not always, a feasible task, following the indication listed in Table 1. This makes it possible for a human expert to prepare a dataset of stratigraphic, non-stratigraphic and ambiguous cases and provide such dataset to a supervised neural net for training.

To this aim, we trained both an Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN).

ANN are inspired by the structure of the human brain and consist of interconnected layers of ‘neurons’ (nodes). Each node of a layer is connected to all the nodes of the successive layer and these connections are represented by a weight and biases (Bishop, 2006). In order to achieve the capability to perform automatic classification, a supervised neural network must be trained. During the training phase, a number of examples

are presented to the net, together with the desired output and the weights associated to each link change to achieve the desired output. Once the network has been trained, the weights do not change anymore, and new input can be presented to the net to be classified accordingly.

The CNN architecture, originally inspired by the functioning of the animal visual cortex (Hubel & Wiesel, 1968; Fukushima, 1980; Yann, 1990), proved to be suitable to image classification tasks. Typically, a CNN consists of three main types of layers: convolutional, pooling, and fully connected layers. The convolutional and pooling layers are responsible for extracting relevant features from the input image, while the fully connected layer maps these features to the final output, that provides the classification.

A crucial step in the training of both these networks is thus the selection of the significant parameters to be extracted to form the training dataset. These are described in the following paragraphs.

2.5.1 ANN and CNN dataset preparation

The extraction of the parameters from the training database that will constitute the input for the neural network has to be based on the features that make the H/V peak classification possible. From Table 1 and Figure 16 to 18 it is clear that the individual spectral components are more informative about the nature of the H/V peaks, compared to the H/V curves themselves. For this reason, we will use as input the individual spectra and not their ratio. H/V is used in practice to compensate for the amplitude variations of the individual spectra (i.e., to compensate for the source effect) but here we will use a different form of compensation, not to lose information about significant differences between the horizontal components.
Our training database consists of 261 H/V peaks – passing the SESAME criteria – extracted from 184 microtremor recordings, all of duration larger than 20 min, collected at several different sites in Italy and classified by us as stratigraphic, artefactual or unknown/other. The H/V peak frequency distribution for the selected examples is illustrated in Figure 20. This provides some information about the bedrock-like layer depth at the selected sites. We see that the most part of the selected examples was characterized by bedrocks at mid-to-large- depths (frequencies below 6 Hz).

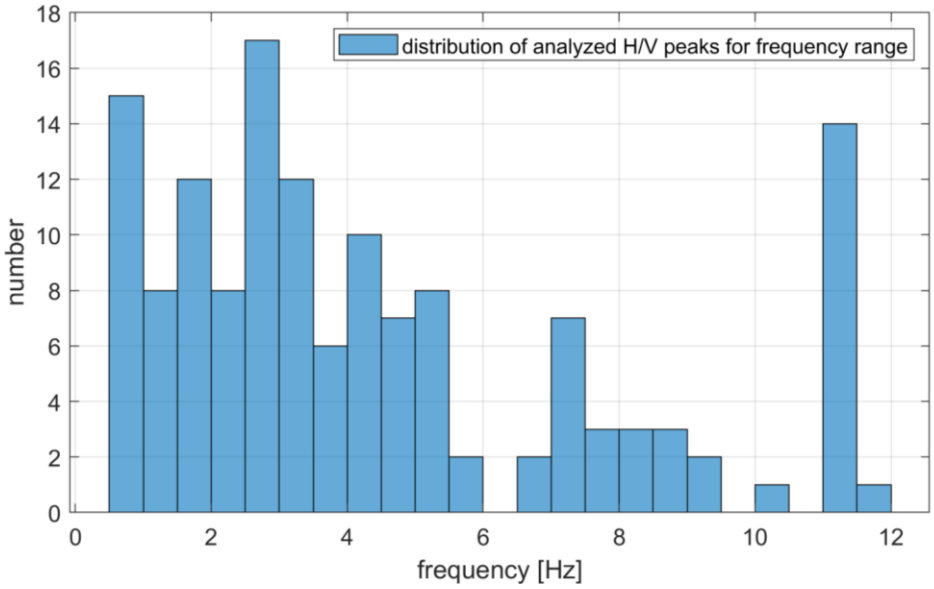

Figure 20. Distribution of the H/V peak frequencies used for the training and validation datasets.

Each of the 184 recordings available was divided into 70 s non-overlapping windows, detrended, tapered, fast Fourier–transformed, and smoothed with a triangular window with width equal to 10% of the central frequency. We tested different smoothing functions (triangular and cosinusoidal with variable width from 1 to 10% the central frequency) because non-stratigraphic peaks are more easily identifiable on poorly smoothed spectra (Figure 17 and Figure 18). However, we observed that the performance of the neural nets was not significantly affected by the smoothing percentage. Therefore, we used 10% smoothing as a final choice. Outlier windows showing spectra beyond two standard deviations from the mean were removed from the analysis in order to analyze only average H/V curves with narrow standard deviation. This is an important processing step, as we discuss in the Appendix.

At the end of the processing, we thus had 3 spectra (NS, EW, Z) and a H/V curve, where the H/V ratio is given by the geometric mean of the horizontal components divided by the vertical one $\frac{\sqrt{NS * EW}}{Z}$.

From each H/V curve that we wanted to classify, we extracted the main peaks according to their prominence. Form each peak, a virtual horizontal line was traced to the left and right, until it crossed the H/V curve because of the presence of another peak (or because the H/V curve ended). The minimum of the signal was searched in the two frequency intervals just defined to the right and to the left of the H/V peak. This point will either be a trough or the curve endpoint. The larger of the two minima interval becomes the reference level and the height above this level is here called 'prominence' (Figure 21). For our case, a peak is considered valid only if its prominence is significantly higher than the surrounding peaks and reaches at least 70 % of the maximum prominence of the signal. We classified the H/V peaks according to their prominence and selected for classification up to a maximum of 3 peaks for each H/V curve.

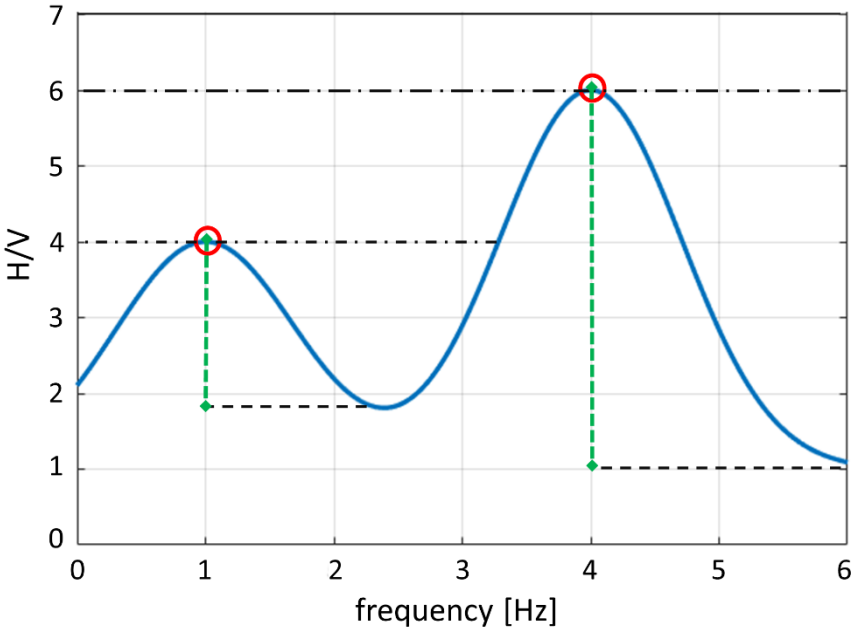

Figure 21. Form each peak, trace a virtual horizontal line to the left and right, until it crossed the curve because of the presence of another peak (or because the curve ends). Search for the minimum of the signal in the two frequency intervals just defined to the right and to the left of the peak. This point will either be a trough or the curve endpoint. The larger of the two interval minima becomes the reference level and the height above this level is here called 'prominence'. For our case, a peak is considered valid only if its prominence is significantly higher than the surrounding peaks and reaches at least 70 % of the maximum prominence of the signal.

Once the prominent peaks have been identified, occurring at frequency f_0, a window centred at f_0 and with a total aperture of $\Delta f = \left[\frac{f_0}{2}, 2f_0\right]$ is considered (Figure 22). Within this window, we resampled each individual spectrum with 30 points and this set of 30 values for each NS, EW, Z component (90 values) was the input provided to the neural net. We tested different number of samples (from 15 to 90) per training example and found that 30 was the smallest number capable to not reduce the neural network performance significantly. The aperture $\Delta f = \left[\frac{f_0}{2}, 2f_0\right]$ for each training example appeared sufficient to include the $[f_0, 2f_0]$ interval, diagnostic of the nature of the H/V peaks (Table 2, Figure 22, Figure 23).

Since a huge variability in terms of spectral amplitude can exist in this $\Delta f = \left[\frac{f_0}{2}, 2f_0\right]$ range, depending on the local ambient noise level, on whether the data are in counts of physical units, etc., in order to simplify the artificial neural net training, we scaled the spectra by dividing each of them by the minimum value of the vertical component. The spectra within the red frame, scaled dividing them by their minimum vertical value, and resampled with 30 samples per component (Figure 22c) constitute one input example for the ANN. Scaling the amplitude is important to compensate for the variability of the source and make the ANN work simpler.

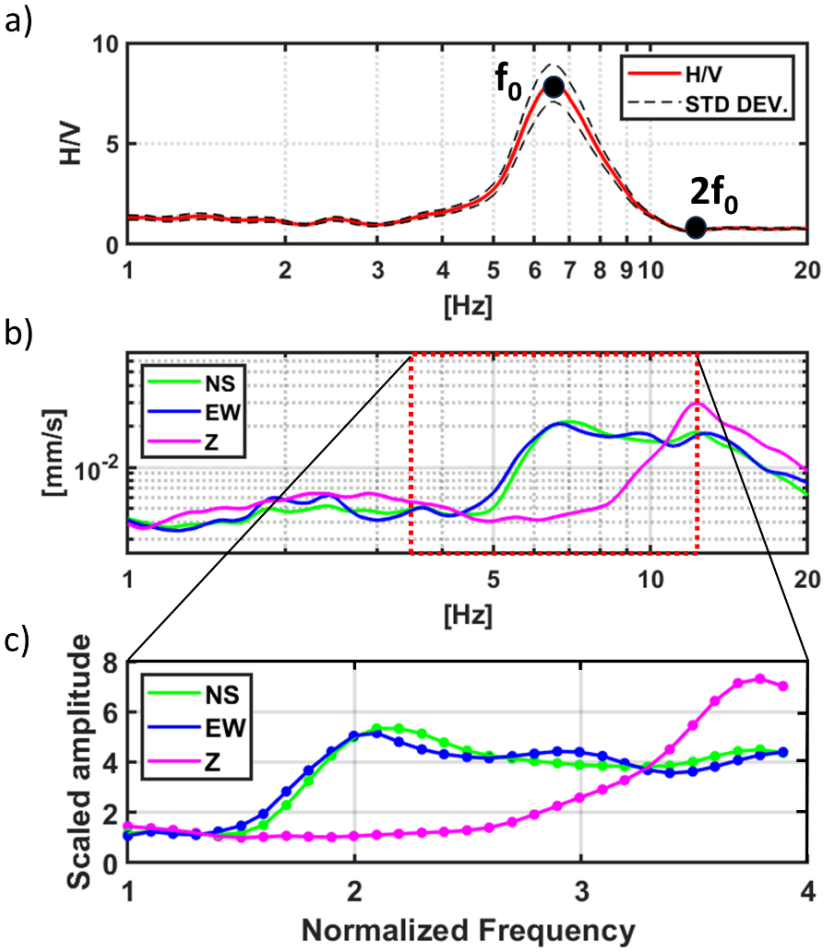

Figure 22. a) H/V curve and b) individual amplitude spectra (in velocity). The red frame centered around the H/V peak at f_0 = 6.5 Hz has a width $\Delta f = \left[\frac{f_0}{2}, 2f_0\right]$ and spans the [3.25, 13] Hz frequency interval. The spectra in this window range from amplitude 4 mm/s to 20 mm/s. The spectra within the red frame, scaled dividing them by their minimum vertical value, and resampled with 30 samples per component (panel c) constitute one input example for the ANN. Scaling the amplitude is important to compensate for the variability of the source and make the ANN work simpler. Frequencies are also normalized (1 stands for $\frac{f_0}{2}$; 2 stands for f_0; 4 stands for $2f_0$). The spectral shape is thus preserved, and the input values remain restricted into homogeneous intervals.

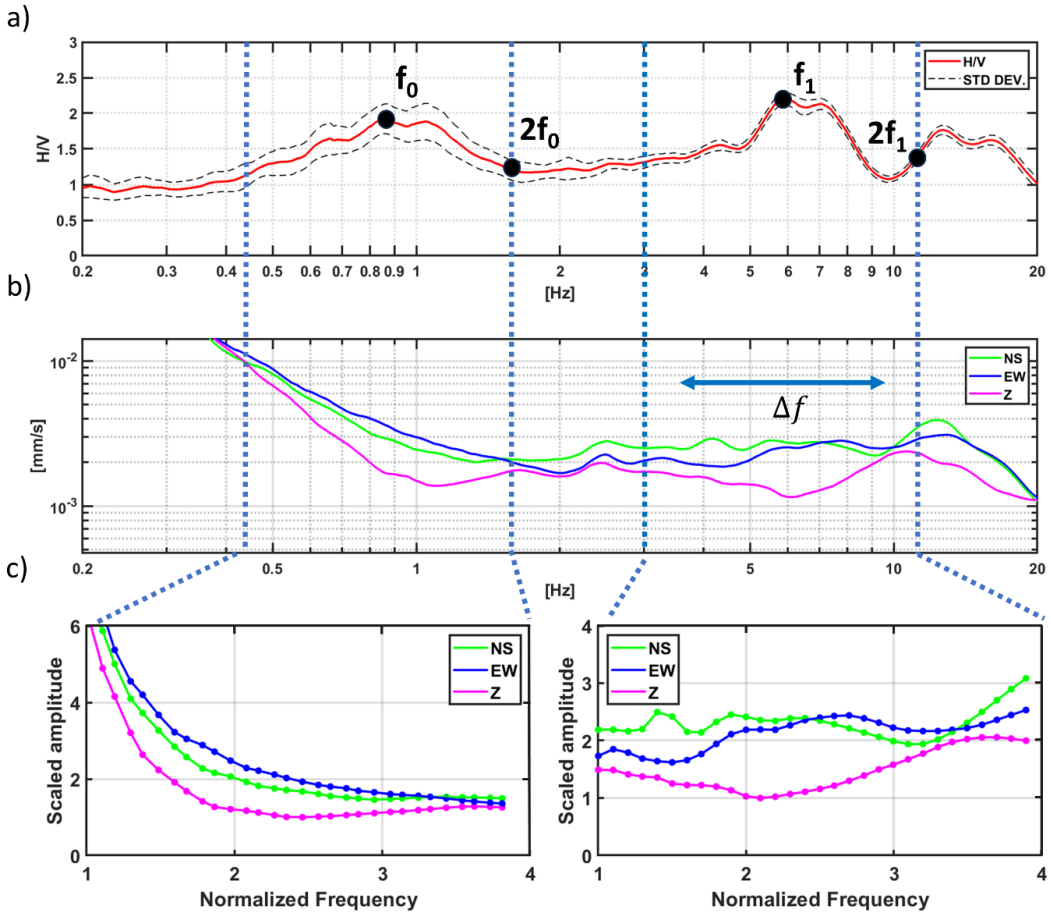

Figure 23. a) In this example two prominent peaks are identified at f_0 = 0.9 and f_1 = 6 Hz in the H/V curve. b, c) The corresponding spectra to be used as input for the ANN and CNN are thus extracted in the Δf_0 = [0.45, 1.8] Hz and Δf_1 = [3, 12] Hz intervals. The normalization and resampling procedures are the same described for Figure 22.

2.5.2 ANN architecture

Once the input examples and data have been defined, the ANN can be trained, which consists in adjusting the weights of the neuron connections to achieve the maximum output value at the expected class (stratigraphic, artefactual, other).

The input dataset was thus split into a training (70%, 183 peaks) and a validation (30%, 78 peaks) set.

The performance of the ANN in terms of mean square error (MSE) between predicted and expected output was then evaluated. Before reaching the final configuration, we tested different network architectures (i.e., different number of hidden layers (from 1 to 7) and of neurons (from 5 to 90), different activation functions (sigmoid, linear), different training algorithms (scaled conjugate gradient backpropagation, conjugate gradient backpropagation with Polak-Ribiére updates) and different training/validation subsets. We obtained the best performance (MSE = 0.0032) with an architecture of 7 hidden layers, each with [5, 10, 10, 10, 10, 10, 5] neurons respectively, activated by a hyperbolic tangent activation function and softmax (Bishop, 2006)

activation function applied to the output layer only. The best performing training algorithm for this type of classification was the Levenberg-Marquardt backpropagation (Marquardt, 1963; Hagan & Menhaj, 1994). The performance of the finally selected classification model is shown in the confusion matrix of Table 2. The precision rate is calculated as the ratio between the number of correct predictions to the total number of predictions. The recall rate measures the sensitivity of the model, and it is the ratio of correct predictions for a specific target to the total number of cases in which it occurs. The overall accuracy represents the total number of correct predictions among all the examples presented for the classification and was about 92% in the validation dataset. The elements outside the diagonal represent the misclassifications.

Table 2. Performance of the trained ANN. Precision is the number of correct predictions to the total number of predictions. Recall rate is the number of correct predictions for a specific target to the total number of cases where the target occurs. The overall accuracy is the total number of correct predictions among all the examples presented for the classification.

	Real target			
Predicted target	Stratigraphic	Artefactual	Other	Precision
Stratigraphic	27	1	0	96.4%
Artefactual	1	24	3	85.7%
Other	1	0	21	95.4%
Recall	93.1%	96.0%	87.5%	
Overall Accuracy				92.3%

2.5.3 CNN

The input of convolutional neural nets are typically images. This does not require extracting points from the spectra but requires some scaling in the input images, so that they are all presented in the same way and the CNN goals can be easier to achieve.

We used a pretrained supervised SqueezeNet model (Iandola et al., 2016) and the same set of 261 H/V peaks used for the ANN, for the training phase. The SqueezeNet we used was trained on more than one million images from the ImageNet database (Russakovsky et al., 2015) and is made available by Matlab® (Mathworks Inc., 2022). The entire network is composed of 68 layers and 75 connections. The requested image input size is 227 x 227 pixels and can be a RGB image. We note that such a matrix of values is much larger than the 90 input values used in the ANN and we could reasonably achieve the same results with smaller matrices. However, processing H/V images is certainly an option (for example when one does not have the data but only the H/V + spectra figure) and we believed that exploring the performance of a network pre-trained on image pattern recognition was interesting, too.

In a pre-trained network like SqueezeNet, when a new convolutional layer is added at the end of the network, the previously learned feature maps is still used. These feature maps were learned during the pre-training process on a large dataset and were already optimized for identifying general patterns in the input data. By

utilizing the learned feature maps, a neural network can quickly adapt to a new task by fine-tuning the weights of a newly added layer on a smaller, task-specific dataset. This process is known as transfer learning (Pratt, 1993) and has shown to be effective, particularly when the new dataset is relatively small.

However, when fine-tuning the network, it is necessary to adjust a range of parameters that control the model, as well as hyperparameters that control the training process, to let the network learn the relevant features of the new dataset.

It was thus necessary to modify the last layers (last convolution 2D Layer and classification output layer) without resetting the weights obtained for the first layers and retrain the network on our database. The feature of this neural network is that, unlike some other NNs, its final layer is a convolutional layer instead of a fully connected layer. In fully connected layers, each neuron is connected to all neurons in the previous layer. This is typical in ANN. Differently, in this CNN, the final layer is another convolutional layer, where neurons perform convolution operations on the input data. This allows to preserve the spatial information of the image and perform a more precise classification. The function used to train the network is a stochastic gradient descent with momentum and the loss function is the cross-entropy loss.

The hyperparameters of the neural network were tuned as follows: the initial learning rate was set to 0.0001, which controls the rate of weight updates during the gradient descent algorithm used to minimize the network loss function. The maximum number of epochs was set to 8, which limits the duration of the training process and prevents overfitting. The batch size was set to 11, which is the number of training examples used in each iteration of the training process. Each epoch involved iterating through all the batches of the dataset to update the weights of the network. This allowed the network to learn from the entire dataset without requiring all the data to be loaded into memory at once.

As per the ANN, the dataset was divided into a training and a validation subset, with 70% and 30% examples respectively, extracted from the prepared database. We obtained an overall accuracy (percentage of examples classified correctly out of the total number of examples) on the validation set of about 95%, as shown in Table 3. The final ANN and CNN architectures are summarized in Figure 24.

Table 3.Performance of the trained CNN. Precision is the number of correct predictions to the total number of predictions. Recall rate is the number of correct predictions for a specific target to the total number of cases where the target occurs. The overall accuracy is the total number of correct predictions among all the presented examples to classify.

	Real target			
Prediction target	Stratigraphic	Artefactual	Other	Precision
Stratigraphic	29	2	1	90.6%
Artefactual	0	22	1	95.6%
Other	0	0	23	100.0%
Recall	100%	91.7%	92.0 %	
Overall Accuracy				94.8%

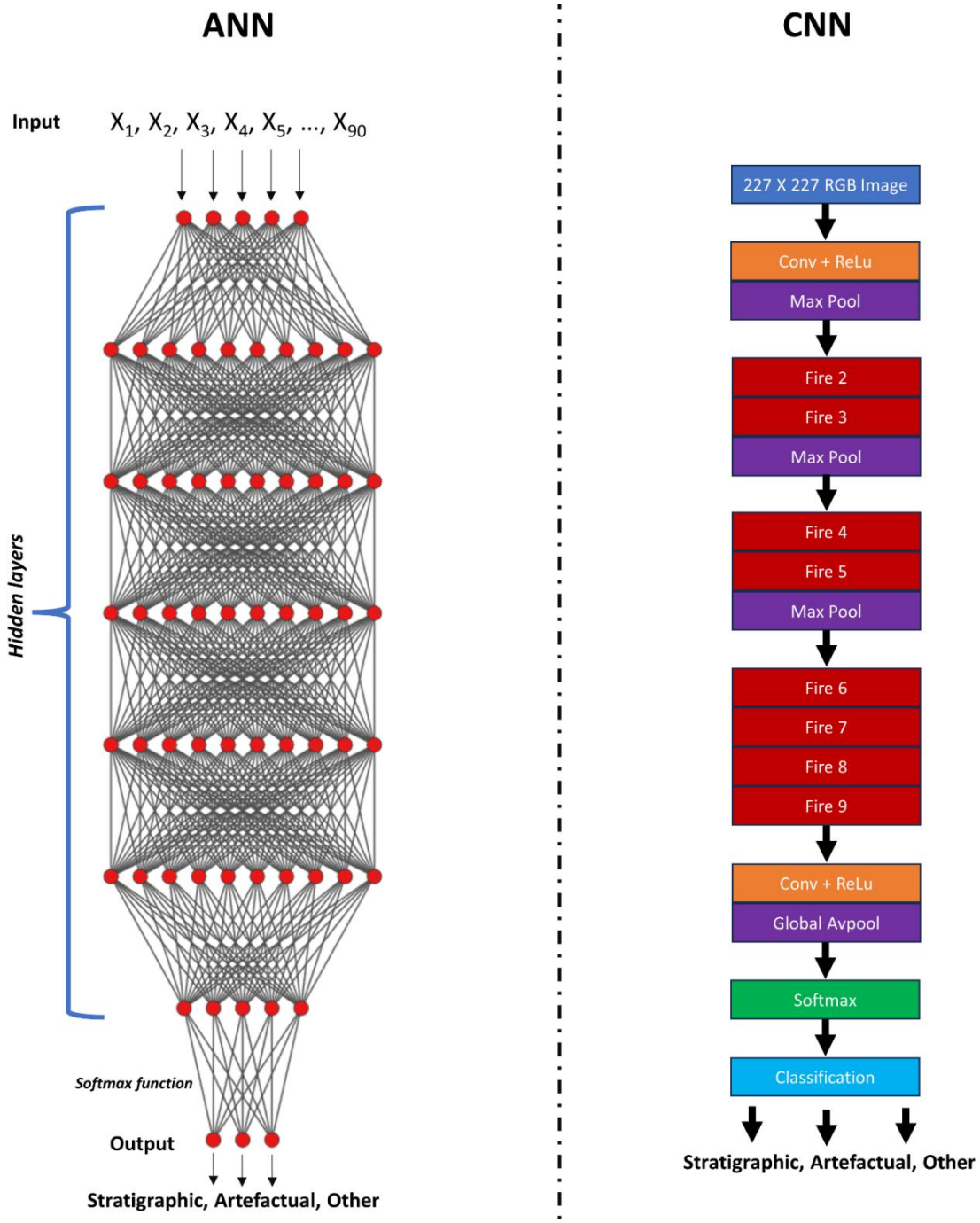

Figure 24. Schematic architecture of the Artificial Neural Network (left) and of the Convolutional Neural Network (right) used in this study. The input of the ANN is a matrix of 30 (scaled spectral amplitude, normalized frequency) values per component (90 values per example); the input of the CNN is a 227 x 227 RGB pixel image of the same normalized spectra. In detail, 'Conv' represents a convolutional layer with a corresponding ReLU (activation function). 'Fire' is a convolutional block with 1x1 and 3x3 convolutions designed to extract information. 'Max Pooling' is employed for dimension reduction by selecting the maximum value in a window. 'Global Average Pooling' is utilized for dimension reduction by calculating the average of the output. Lastly, 'Softmax' assigns probabilities to classes (for more detail see Iandola et al., 2016).

2.6 Comparison of the ANN and CNN performances on a fully independent dataset

We tested the trained ANN and CNN on an independent dataset from the ARRA-funded (American Recovery and Reinvestment Act, 2009) site characterization project. Within this project, 184 Californian strong-motion stations sites were characterized by means of geophysical surveys (including microtremor H/V), as well as 4 stations in the Central-Eastern United Sates (CEUS) (Yong et al., 2013). All data were collected for one hour at 200 samples per second and three to six recordings were collected at each investigated site.

All these data were reprocessed and resampled by us, in the same way used to prepare the training set.

From a total amount of about 600, we retained 168 ambient noise recordings and a total amount of 220 peaks to be classified. Those excluded did not contain peaks (seismic stations are often installed on outcropping rock) or had a too large standard deviation, which made them statistically unreliable.

Each prominent peak was both classified by a human operator (us) and classified from the ANN and CNN into the three available classes. The human operator classified the possible output with a binary code [0,0,1] for 'stratigraphic'; [0,1,0] for 'artefactual' and [1,0,0] for 'other' peak. The NN output is a number in the range [0 1] that indicates the probability assigned to each class. From Table 4 we see that the test dataset contains, according to the human operator, 89% stratigraphic, 5.5% artefactual and 5.5% other peaks. We now discuss the performance of our nets in the classification of stratigraphic and non-stratigraphic (anthropic, other) peaks from this dataset.

Table 4. Preliminary confusion matrix for CNN (left) and ANN (right). The sample test contained 220 examples (196 classified by human operators as stratigraphic, 12 as artefactual and 12 as other).

	CNN			ANN		
	Actual target (Human classification)			Actual target (Human classification)		
Predicted target	Other	Artefactual	Stratigraphic	Other	Artefactual	Stratigraphic
Other	8	0	4	10	5	21
Artefactual	1	12	11	2	7	17
Stratigraphic	3	0	181	0	0	158
Recall	66.7%	100.0%	92.3%	83.3%	58.3%	80.6%
Overall Accuracy			91.4%			79.5%

From the preliminary confusion matrix, we see that peaks classified as stratigraphic by humans, are also classified as such from CNN and ANN in 92.3% and 80.6% of cases, respectively.

The overall accuracy (percentage of examples classified correctly out of the total number of examples) on this fully independent dataset is 91.4% and 79.5% for ANN and CNN, respectively.

We note that a different classification between the nets and human output does not necessarily imply an error on either side. When the network suggests a different class compared to human experience, a closer examination of the corresponding H/V curve and spectra should be performed, as they may reveal patterns not immediately understood by the human eye or *vice-versa*.

To present a more visual comparison, in Figure 25, we show the probability of attribution to the stratigraphic class for each input example, where the blue bars indicate when the network attribution coincides with the human classification, and the orange bars indicate when it diverges. The height of each bar represents the confidence of the attribution of the specific example to the corresponding class in the range [0, 1] given by the neural nets.

For the 'stratigraphic class', CNN provides a different classification compared to human classification in 15 cases out of 196, while ANN in 38 cases. Figure 26 Figure 27Figure 27 show the same results for the non-stratigraphic and other class, respectively. The CNN achieved 100% precision (12 correctly classified cases out of 12) for the 'artefactual class', while the ANN gave 5 apparently incorrect cases out of 12. On the other hand, for the 'other class', the CNN achieved a precision of 66% with 4 out of 12 cases apparently misclassified, while the ANN achieved 83% precision with only 2 apparently misclassified cases.

At this point it is important and interesting to study the reasons for these classification differences. We will thus analyse the cases where either one or both networks disagree with human classification.

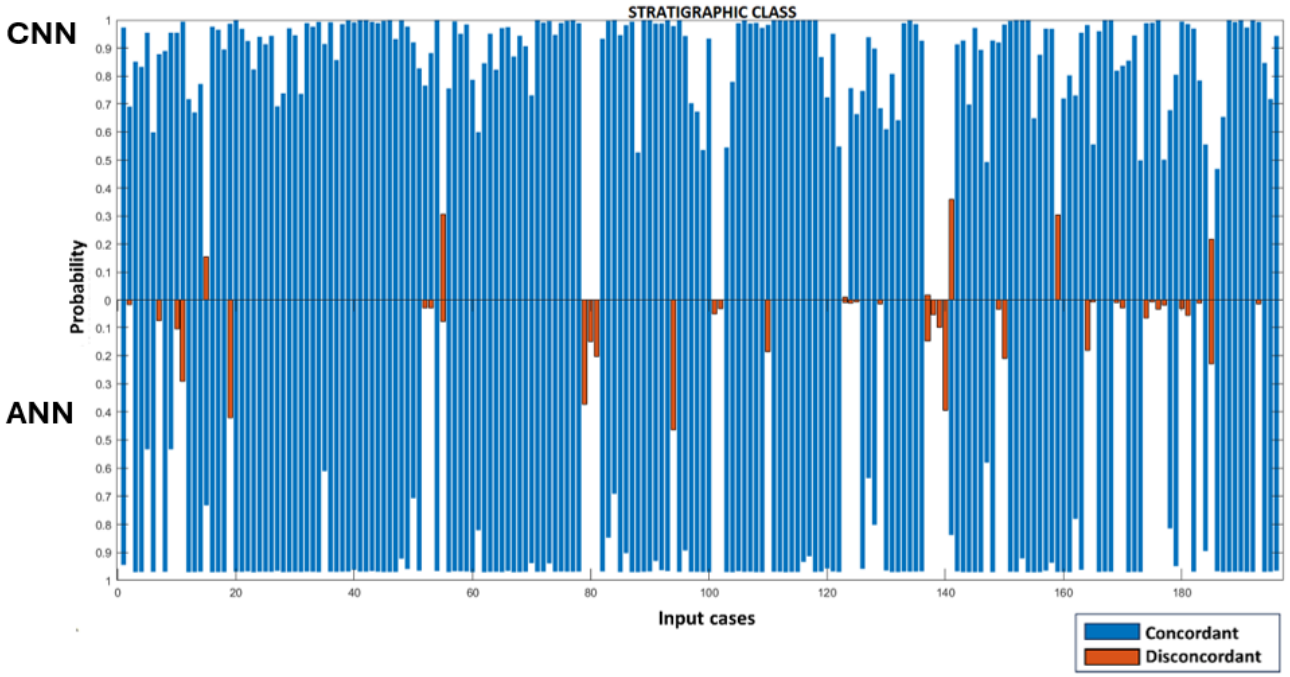

Figure 25. Confidence of attribution of each example to the stratigraphic class according to the CNN (top half) and ANN (bottom half) networks. For these examples the expected output is the 'stratigraphic' class. Blue bars mark the examples for which also the neural nets provided a classification as stratigraphic. Orange bars mark the examples classified into a different class by the nets.

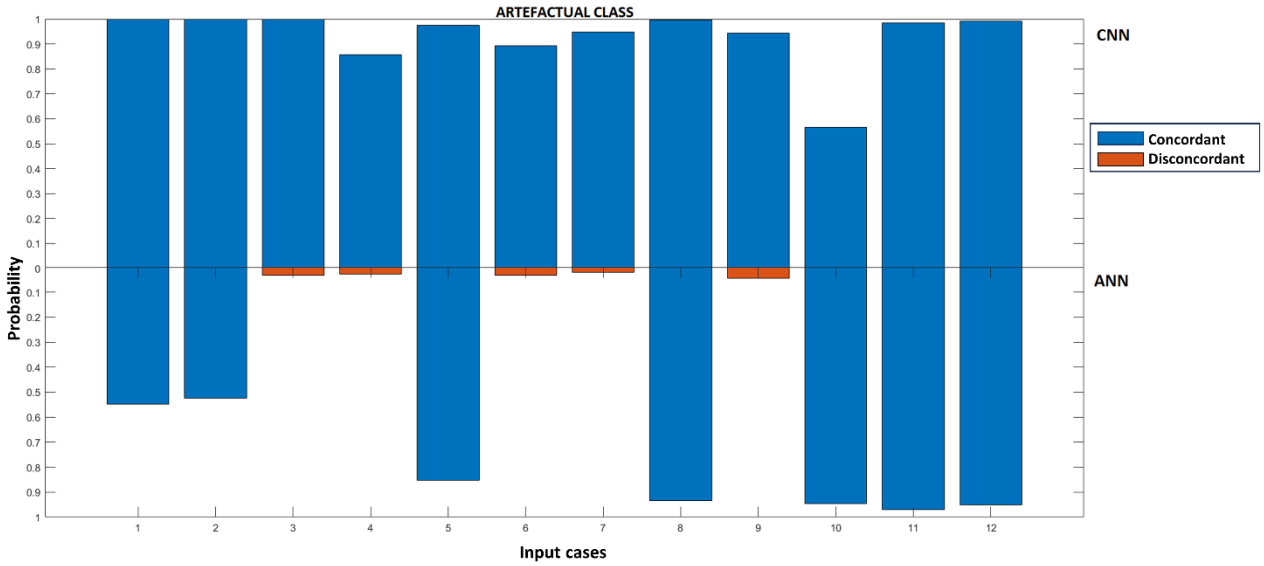

Figure 26. Confidence of attribution of each example to the artefactual class according to the CNN (top half) and ANN (bottom half) networks. For these examples the expected output is the 'artefactual' class. Blue bars mark the example for which also the neural nets provided a classification as artefactual. Orange bars mark the examples classified into a different class by the nets.

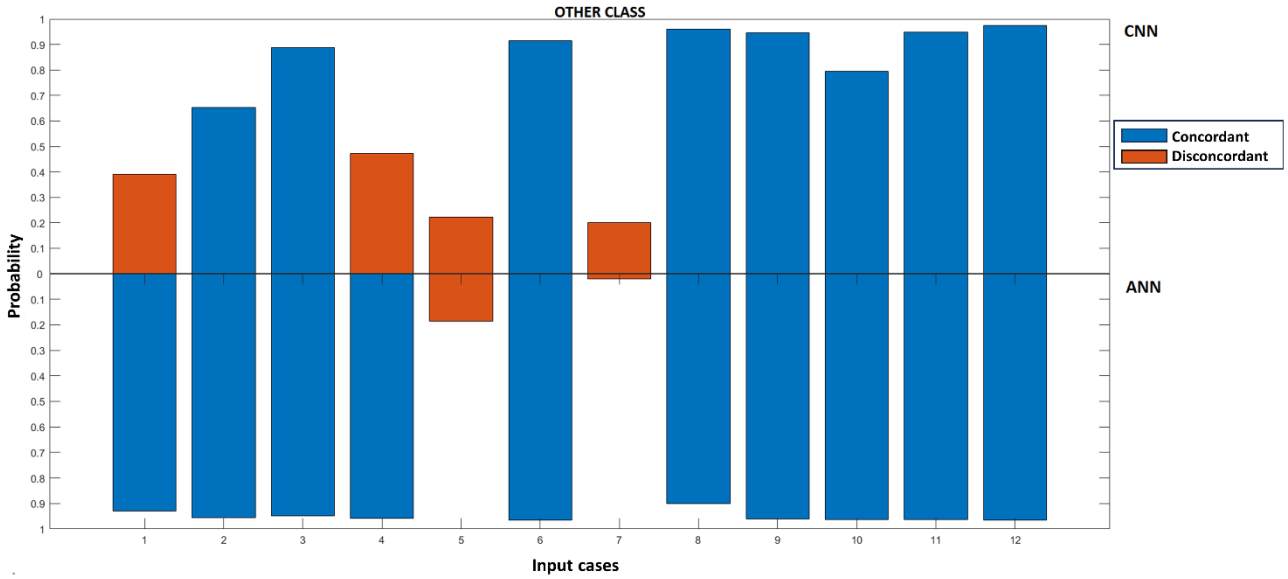

Figure 27. Confidence of attribution of each example to the other class according to the CNN (top half) and ANN (bottom half) networks. For these examples the expected output is the 'other' class. Blue bars mark the example for with also the neural nets provided a classification as 'other'. Orange bars mark the examples classified into a different class by the nets.

2.7 Discussion

The cases in which the networks disagree with human classification can be categorized, with reference to the stratigraphic class, into two types: i) presence of stratigraphic H/V peaks with superimposed artefacts; ii) multiple H/V peaks at close distance.

As an example of the first case, let us consider Figure 28. This example shows a clear stratigraphic H/V peak at 1.5 Hz, disturbed by a number of artefacts, the most prominent of which occur at 1.4 and 1.8 Hz (panel a,

b). When these are removed by filtering the recording, the stratigraphic nature of the H/V peak is even more evident (panel c, d). This example was classified as stratigraphic by the human operators because in training the net it was not possible to provide two possible output for the same input. However, CNN and ANN respectively assign a probability of 30% and 4% to the stratigraphic class; of 70% and 93% to the artefactual class, correctly acknowledging the simultaneous presence of artefacts. These cases cannot therefore be considered as misclassified. This pattern occurred 3 times in our dataset and the the preliminary confusion matrix (Table 4) changes to what is shown in Table 5. The network rate of success increases to 92.7% and 80.9%, respectively.

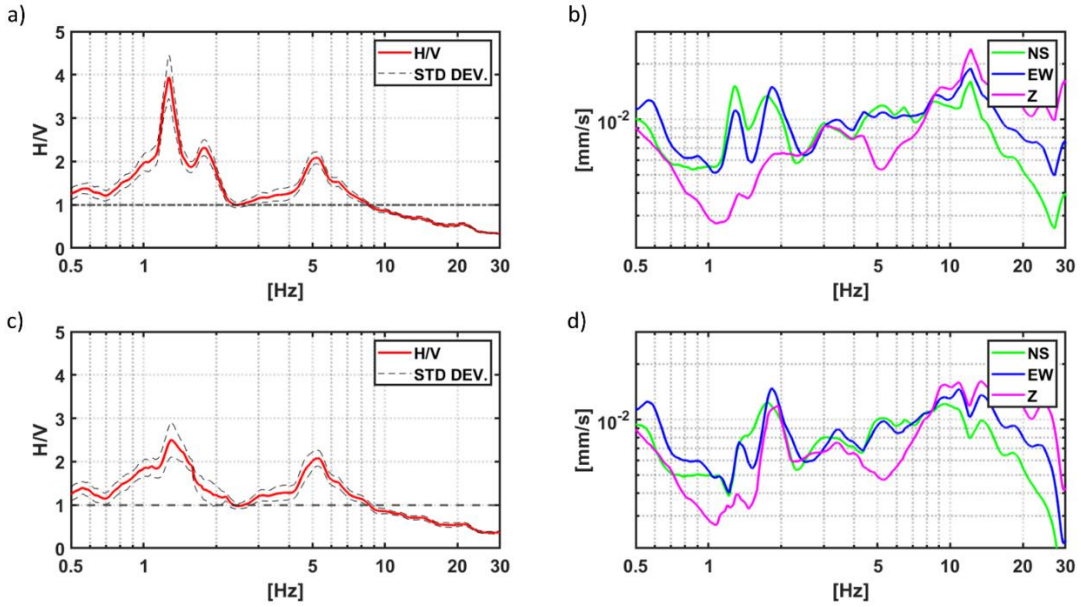

Figure 28. Example of H/V curve with stratigraphic peaks at 1.5 and 5 Hz, with superimposed artefacts: (a, b) original H/V curve and corresponding amplitude spectra of the three components of motion; (c, d) H/V curve and corresponding spectra after a notch-filter application to reduce the main artefactual disturbances. The stratigraphic origin of the 1.5 and 5 Hz peaks is clearly suggested by the local minima in the vertical spectral component of motion.

Table 5. Final confusion matrix for CNN (left) and ANN (right). The sample test contained 220 examples (196 classified by human operators as stratigraphic, 12 as artefactual and 12 as other).

	CNN			ANN		
	Actual target			Actual target		
Predicted target	Other	Artefactual	Stratigraphic	Other	Artefactual	Stratigraphic
Other	8	0	4	10	5	21
Artefactual	1	12	8	2	7	14
Stratigraphic	3	0	184	0	0	161
Recall	66.7%	100.0%	93.8%	83.3%	58.3%	82.1%
Overall Accuracy			92.7%			80.9%

The second group of misclassifications arises when in the $\Delta f = \left[\frac{f_0}{2}, 2f_0\right]$ window selected around the prominent H/V peak, a second H/V peak (or part of it) is present. In other words, when two H/V peaks occur in a narrow frequency interval. An example is provided in Figure 29, where the $\Delta f = \left[\frac{f_0}{2}, 2f_0\right]$ window around the prominent peak includes part of another apparent peak on the left side. In this case, both networks classify (with 99% and 96% confidence, respectively) the 3 Hz peak as an artifact. This pattern occurred 5 time in our dataset.

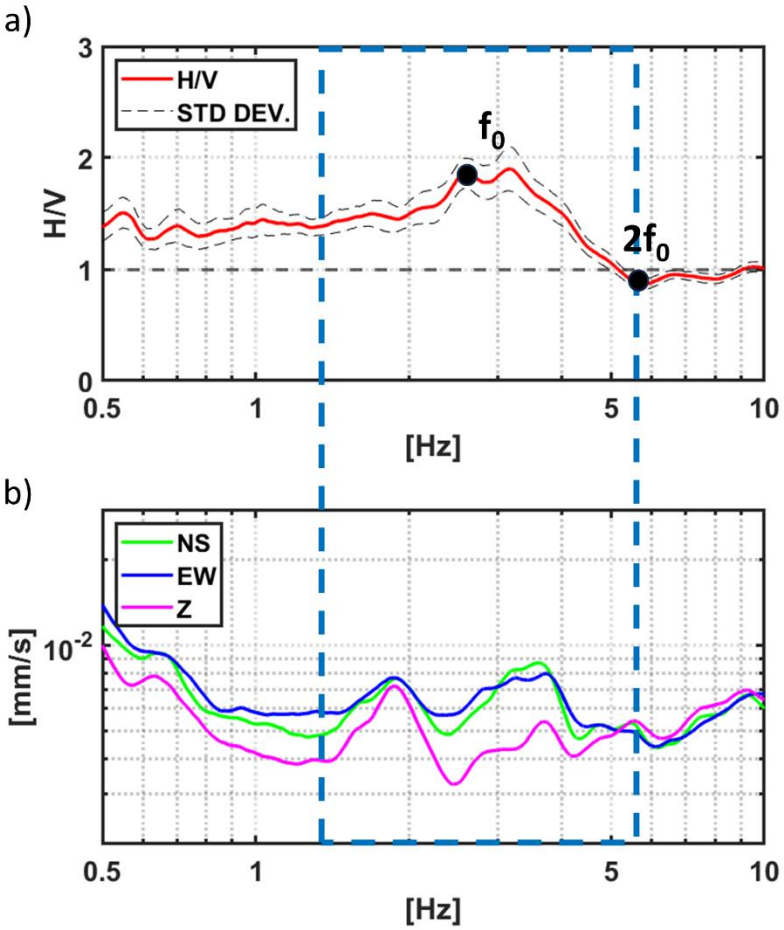

Figure 29. a) H/V curve of stratigraphic peak at 3.1 Hz and b) corresponding individual spectra of motion in velocity. The dashed blue lines indicate $\Delta f = \left[\frac{f_0}{2}, 2f_0\right]$ window used to identify the prominent peak and it is evident that the window includes, on the left side, part of another peak.

So far, we have discussed the divergence between the NNs and human classification, however, there are also cases where the CNN and ANN classifications diverge on the same examples. CNN seems to have more problems in identifying as stratigraphic peaks spectra those with weak local maxima in the horizontal components and less pronounced local minima in the vertical component compared to ANN. In these cases the CNN finds it hard to produce a unique answer and assigns large probabilities to all classes (e.g., 46% artefactual, 36% stratigraphic, 18% other for the 4.8 Hz peak in Figure 30).

In general, by also observing Figure 25 and Figure 26, ANN seems to be less precise in this task compared to the CNN. This discrepancy can be attributed, beyond the specific differences between the two types of nets and input size, also to criteria that we had to apply to extract the input parameters from the spectra to train the ANN and that we did not have to make explicit for CNN.

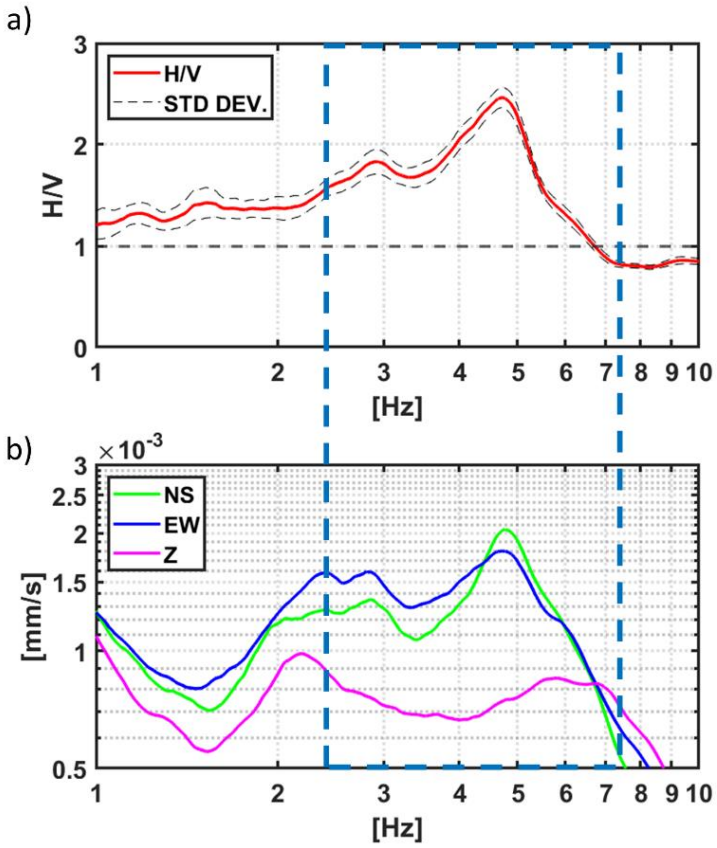

Figure 30. a) H/V curve with a stratigraphic peak at 4.8 Hz and b) corresponding individual spectra of motion in velocity. The dashed blue frame indicates the $\Delta f = \left[\frac{f_0}{2}, 2f_0\right]$ *window used to identify the prominent peak. This example was classified by CNN as 46% artefactual, 36% stratigraphic, 18% other and this is understandable for several reasons: the horizontal components are not statistically identical and there is more than one local maxima in each horizontal component within the selected frequency interval.*

2.8 Conclusions

Seismic networks, seismic microzonation surveys, and seismic exploration of the subsoil provide thousands and thousands of microtremor recordings all around the globe. When processed in terms of H/V spectral ratios, these can unveil useful stratigraphic features about the subsoil. However, it is mandatory to distinguish stratigraphic from non-stratigraphic features before any interpretation. In this paper we first review what makes a H/V peak stratigraphic or not. Then, we train two types of supervised neural nets to classify H/V peaks automatically, or at least to support human decisions by means of an extended knowledge base.

We trained an ANN that uses as input a set of sorted values extracted from the individual microtremor spectra. Then, we trained a CNN working directly on images of the microtremor spectra. In both cases the data were scaled to span a common range of size and values, to make the nets work easier. The nets were trained on an Italian dataset of H/V curves and tested on an independent dataset of Californian curves, collected by different instruments but processed in the same way. On a fully independent database compared to the training one, the two nets showed the 81% and 93% accuracy, respectively in classifying the presented input, compared to the human choices.

In this specific task, using a CNN is more convenient than ANN. The former does not require the extraction of specific input values from the individual spectra (and all the problems connected to the choice of the most suitable number of input data, for example) but allows to achieve a classification directly from an image. It must be noted, however, that this is the case because we used a pre-trained CNN where the optimization of the parameter extraction from the images was already implemented. However, the CNN still seems to have a slightly larger generalization capability in this case.

Some of the independent examples presented to the networks did not fall in single output class but showed both artefacts and stratigraphic peaks in the same example. Our training dataset did not include such phenomenology, but the nets still managed to classify them as belonging with equal probability to two classes.

These results were obtained on relatively small training databases, composed of approximately 200 examples and with a large unbalance in number of examples for the three output categories. Such an unbalance somehow reflects the real-world distribution but can be a limitation in the NN training.

Discriminating artefactual from stratigraphic peaks in microtremor recordings is compulsory before any use of the data but this is a quite neglected step. It is quite impressive that the H/V curves are still very rarely analysed together with the individual components of motion, which is where the origin of the peaks can better be assessed. The proposed approach can support the operators and the automatic processing systems in their decisions, as well as complement the SESAME criteria.

Chapter 3. Second seismological application of NN: the ground motion prediction equations (GMPEs)

In this chapter, we will explore same aspects of Ground Motion Prediction Equations (GMPE) or Ground Motion Models (GMM). These models constitute one of the key components of probabilistic seismic hazard studies and are the basis of as well as for seismic engineering and territorial/urban planning.

GMPEs are typically computed by means of regression of some parameters of interest. Typically, if the parameter of interest is peak ground acceleration (PGA), this is expressed as a function of earthquake magnitude and distance.
Here we apply neural nets, specifically a U-NET architecture, to produce region-specific, fully non-ergodic, and data-driven GMM to the Italian context and compare the results with traditional approaches.

3.1 Seismicity

Seismicity, or seismic activity, is distributed worldwide following the continuous movement of tectonic plates. These plates are in constant interaction, and their collisions, separations, or mutual sliding generate phenomena such as earthquakes and volcanoes.

Every year, the world is stroke by earthquakes that, with their destructive force, cause significant losses in terms of human lives and economic impact. The most recent earthquake, centred in Turkey in February 2023 with a magnitude of 7.2, led to a tragedy of immense proportions, with over 60,000 lives lost. However, this is just the latest in a series of seismic disasters, reminiscent of events like the earthquake in Indonesia in 2004, which resulted in the loss of 184,332 lives, and the one in Haiti in 2010, with 222,587 deaths. Looking at economic losses as well, the earthquake in Japan in 2011 marked an enormous cost of 360 million dollars, while the one in Turkey resulted in an economic loss of 105 million dollars.

The analysis conducted by Masi & Nicodemo in 2022, based on data extracted from the Emergency Events Database for the periods 1980-2000, 2001-2020, and 1890-2020, reveals interesting trends in mortality causes related to catastrophic events (Table 6). In the period 1908-2020, earthquakes accounted for 33% of the total global deaths from disasters.

Specifically, during the 1980-2000 period, disasters with the largest number of fatalities were predominantly linked to droughts, constituting about 42% of the overall human losses. However, between 2001 and 2020, earthquakes took on a predominant role, causing approximately 54% of the total deaths.

It is essential to note that this estimate does not include data from 2021 forward, and therefore, it does not reflect more recent events. In particular, the earthquakes in Turkey and Morocco (2023), causing over 60,000

and 2,900 deaths, respectively, serve as an example of significant impact that occurred after the period considered in Masi's analysis.

Table 6 Percentage distribution of casualties in natural disasters across three distinct periods (From Masi and Nicodemo 2022).

	1980-2000		2001-2020		1980-2020	
Disaster type	**Deaths**	**% of total**	**Deaths**	**% of total**	**Deaths**	**% of total**
Earthquake	163,574	12%	721,094	54%	884,668	33%
Landslide	15,790	1%	17,328	1%	33,118	1%
Wildfire	1,302	0%	1,533	0%	2,835	0%
Flood	153,065	11%	104,773	8%	257,838	10%
Storm	258,572	19%	200,142	15%	458,714	17%
Volcanic activity	24,976	2%	1,601	0%	26,577	1%
Drought	560,461	42%	21,232	2%	581,693	22%
Extreme temperature	15,592	1%	171,332	13%	186,924	7%
Epidemic	144,820	11%	102,322	8%	247,142	9%
Total	1,338,152		1,341,357		2,679,509	

The seismic context in Italy has been marked by events that, while presenting an apparently lower magnitude compared to certain global catastrophes, have however left an indelible mark in terms of human and economic losses. Let us begin with the earthquake in Reggio and Messina in 1783, which marked a tragedy estimated at around 50,000 deaths, highlighting the vulnerability of the communities of that area. Moving forward in time, the earthquake in Basilicata in 1857, with its magnitude of 7.1, led to the loss of 11,000 lives, leaving a trail of destruction. The Messina earthquake in 1908, with a magnitude of 7.1, was one of the most devastating events in Italian history, resulting in the loss of half the population of the Sicilian city and a third of the Calabrian population, totalling approximately 100,000 deaths. In 1915, the Marsica earthquake, with a magnitude of 7, marked another dark chapter, causing the death of over 30,000 people and testing the resilience of the affected communities (Figure 31 a).

Moving forward, the Friuli earthquake in 1976, with a magnitude of 6.4, struck forcefully, destroying 19,000 homes and causing the loss of 990 human lives. In 1980, the Irpinia earthquake, measuring 6.9 in magnitude, left an indelible mark, destroying 362,000 homes and resulting in the death of 2,735 people, with 9,000 injured. Then we come to the L'Aquila earthquake in 2009, with a magnitude of 6.1, which led to the loss of 309 lives and estimated economic damages of 16 million dollars (Figure 31 b). Subsequently, the seismic sequence in Amatrice in 2016 inflicted further suffering, with the loss of 300 lives and estimated economic damages of 10 million dollars (Figure 31 c). These data underscore the urgency of addressing a seismic risk analysis plan and implementing effective readiness and mitigation strategies.

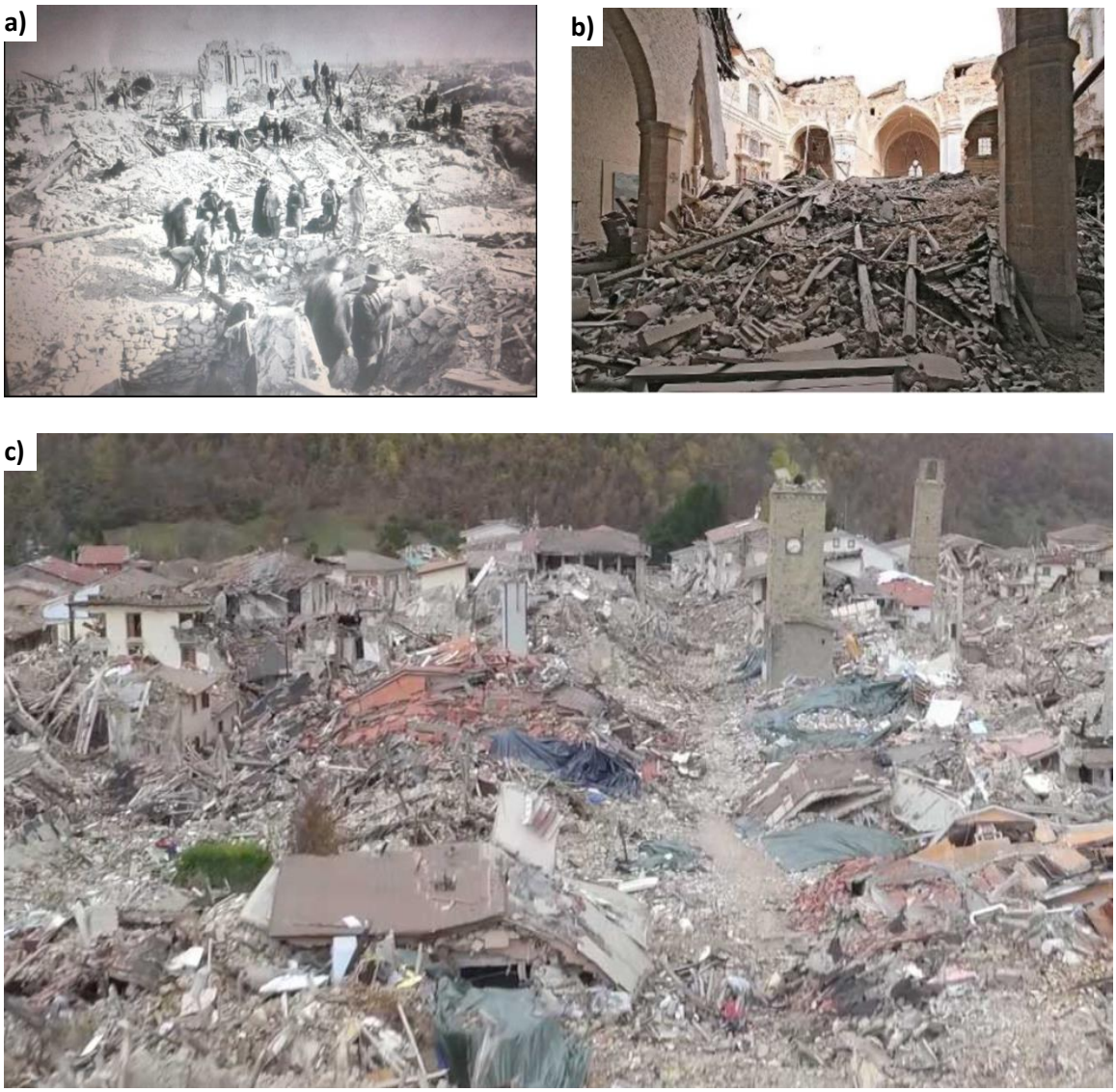

Figure 31. Depicting three catastrophic events: a) the Marsica earthquake, b) the L'Aquila earthquake featuring the representation of the destroyed Collefiorito Basilica, and c) the Amatrice earthquake showcasing an aerial view of the devastated city.

3.1.1 Seismicity in Italy

Italy is exposed to a high seismic hazard due to its geographical position in the convergence zone between the African and Eurasian plates. This positioning causes strong compressive forces, leading to significant seismic activity. Over 2,500 years, Italy has experienced over 30,000 earthquakes of medium and high intensity, with a higher concentration in the central-southern part of the peninsula, Calabria, Sicily, and some northern areas. Seven earthquakes in the 20th century exceeded a magnitude of 6.5 (from Civil Protection).

In Italy, seismic activity resulted in significant economic damages estimated by the Civil Protection at around 135 billion euros over the last forty years, leaving a deep impact on the historical, artistic, and monumental heritage of the country. The seismic vulnerability is particularly high due to population density and the

fragility of building structures, leading to proportionally higher damages compared to similar regions such as California or Japan (Figure 32 a, b).

The fragility of Central and Northen Italy was clearly evident in the seismic sequences of 2009, 2012, and 2016-2017, highlighting an issue that unfortunately exists in other regions of Italy as well, such as Campania, Basilicata, Calabria, and eastern Sicily, all characterized by significant seismic potential. Moreover, according to seismic hazard maps, there exist the probability of approximately 20% of a destructive earthquake occurring within a year (SIGEA, 2018).

The fragility of structures, precarious infrastructure, and population density contribute to Italian high vulnerability, with 70% of buildings constructed before the 1980s, that is before present anti-seismic regulations. (Figure 32 c, ENEA, 2015), Despite the inability to predict seismic events, it is fundamental to mitigate risks, reduce vulnerability, and improve territorial resilience to address the constant seismic threat. Current efforts focus on monitoring the territory through advanced analysis techniques, both remotely using satellites and sensors, and on-site with seismic microzonation studies providing essential insights for territorial planning.

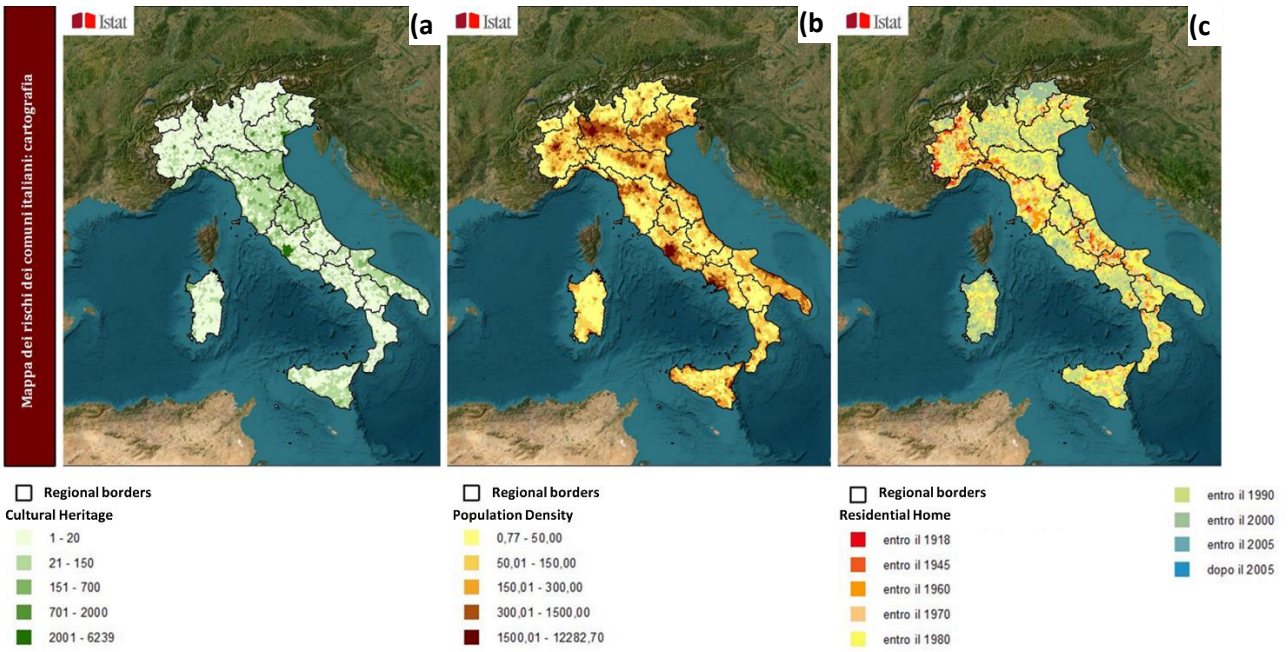

Figure 32. Maps illustrating a) the distribution of cultural assets, b) population density, and c) construction year of residential infrastructure.

3.1.2 Seismic Risk

Seismic risk is determined by a combination of hazard, vulnerability, and exposure. It measures the potential damages, based on the type of seismic activity, building resilience, and human factors (nature, quality, and

quantity of exposed assets), expected within a specific time frame. In a more formal probabilistic context, seismic risk is measured by both the severity of a specific loss metric and the annual frequency or probability of surpassing that level of loss (Bommer, 2022). Therefore, the seismic risk is represented by the combination of the three parameters described below and quantifies the expected economic and social loss in a specific area over a certain time interval and can be summarize in this formula:

$$Risk = Hazard * Exposure * Vulnerability \quad (3)$$

Where:

- *Hazard* refers to the probability that an earthquake of a given intensity occurs in a specific interval time in a specific geographical area. In other words, it measures how likely an earthquake event is to happen in a particular region over time. This concept is often assessed through the analysis of historical seismicity and the recurrence of earthquakes in each area. In Italian national territory we use the probabilistic seismic hazard map that expresses the maximum values of horizontal acceleration on the ground and refers to earthquake levels with a probability of exceeding 10% in fifty years (source INGV).
- *Exposure* refers to the presence and vulnerability of critical elements, such as buildings, infrastructure, and the population, in a seismically active area. Exposure involves how susceptible elements that could be damaged during an earthquake are present in a particular zone. It is determined by the combination of population density with the potential presence of heritage and critical infrastructure. For example, high exposure might indicate a significant density of vulnerable buildings or infrastructure in a high seismic risk area (SIGEA, 2018).
- *Vulnerability* refers to the susceptibility of buildings, infrastructure, and other structures to suffer damage or collapse during an earthquake. Vulnerability depends on construction quality, the types of materials used, and the design of structures. An assessment of vulnerability takes into account how elements exposed to a potential earthquake may be compromised or damaged based on the intrinsic characteristics of the structures themselves.

In the context of seismic risk, hazard refers to the intrinsic characteristics of natural events, such as earthquakes, while risk takes into account human presence and infrastructure exposed to these hazards.

3.2 Ground motion

Seismic waves, generated when faults release accumulated tension during earthquake, are a direct manifestation of the energy released. These waves propagate through the Earth's crust, causing oscillatory movements of the ground with variations in intensity, duration, and direction.

The ground motion, felt as tremors or shaking, represents one of the direct impacts of earthquakes. The intensity of this movement is influenced by various factors, including the magnitude of earthquake, the depth and the distance from hypocenter.

During ground shaking, structures and the ground experience dynamic forces that can result in significant damage to buildings, bridges, infrastructure, and other civil works. Additionally, seismically induced phenomena, as summarized in Figure 33, are mentioned.

During a seismic event, there can be direct effects such as ground deformation along the fault that caused the earthquake (surface faulting). Additionally, there can be secondary consequences not directly related to fault movement but induced by the propagation of seismic waves from the hypocentral region. In this case, the waves shaking the water-saturated or partially water-saturated sandy soil can lead to a sudden increase in water pressure with the loss of soil resistance, resulting in liquefaction phenomena. Alternatively, it may trigger landslides, especially in the presence of topographic relief.

For example, in Italy, the CEDIT catalogue (Martino et al., 2020), in its updated version, has compiled data on 173 earthquakes distributed across more than 1,575 Italian locations. It observes that out of a total of 3,989 earthquake-induced effects recorded, there are 2,222 landslides (56%), 903 ground fractures (23%), 486 liquefaction phenomena (12%), 183 surface faulting (4%), and 195 instances of permanent changes in the ground plane (5%). Another earthquake-induced effect is the tsunami, which can also be generated by eruptions or submarine landslides triggered by an earthquake.

Considering that earthquake effects are also influenced by local geological and morphological characteristics, we can introduce the phenomenon of amplification. This is generally higher in areas with soft soil cover, such as alluvial or lacustrine soils, and there may also be topographic amplification at morphological elements like hills, slopes, or canyons.

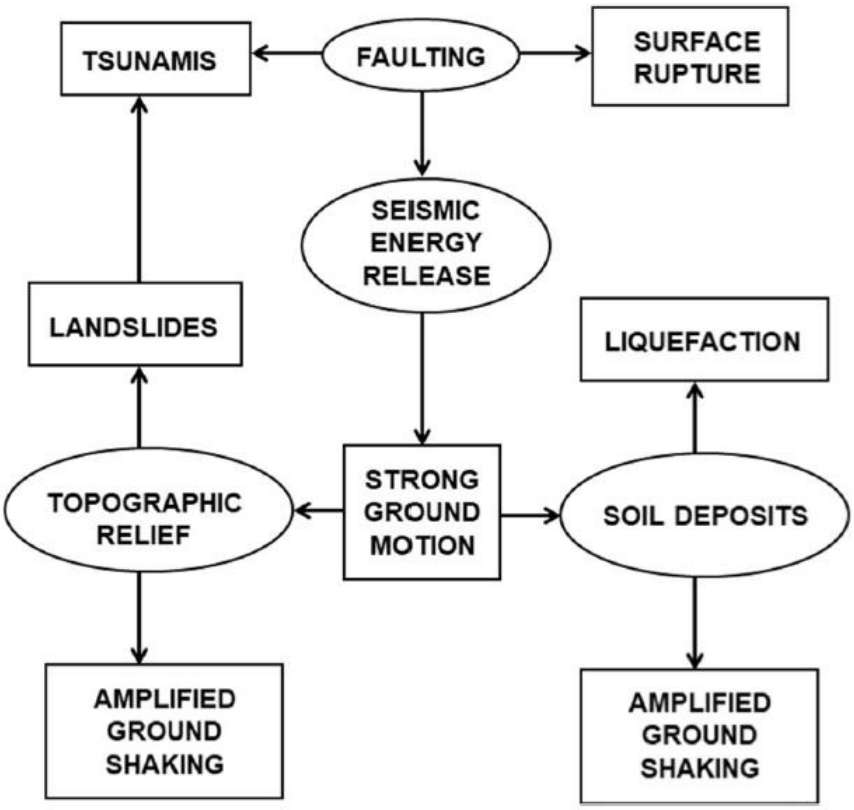

Figure 33. Interactions between ground motion (box) and the natural environment (ellipses) can release a diverse range of seismic hazard effects, both primary and secondary phenomena (rectangle), (Bommer 2022).

The perceived intensity of shaking in a given area during an earthquake is influenced by the amplitude and frequency of seismic waves. Ground motion intensity is commonly measured using among the other parameters like peak ground acceleration (PGA), peak ground velocity (PGV), or peak ground displacement (PGD), along with spectrum acceleration (SA) expressed in terms of gravity (g) or centimetres per second squared (cm/s^2). These parameters provide valuable information about the intensity and features of the earthquake in a particular area.

To understand these parameters, let us take a step back and explain the difference between an accelerogram, a Fourier spectrum, and a response spectrum in acceleration.

The accelerogram is a graphical representation of the ground acceleration recorded during a seismic event. The recording is done through an accelerometer that have three axes: north-south (NS), east-west (EW), and vertical (Z) and represents changes in terms of acceleration, velocity, and displacement of soil along each of these axes in the time domain.

To obtain the Fourier spectrum from an accelerogram, it is necessary to transform the signal from the time domain to the frequency domain. This is because a complex seismic signal can be viewed as the sum of different harmonics and through the Fourier transform allows us to identify different frequencies that

compose the signal. From this analysis, it is possible to obtain the PGA, PGV, and PGD, represented by the maximum value of the respective variable.

Now, we consider different single-degree-of-freedom systems (SDOF), each characterized by distinct periods and subjected to seismic excitation at their base. The Fourier spectrum is obtained from the response of each SDOF system, and the response spectrum (Sa) is formed by enveloping the maximum acceleration values corresponding to specific periods (Figure 34) and provides a representation of structural behaviour.

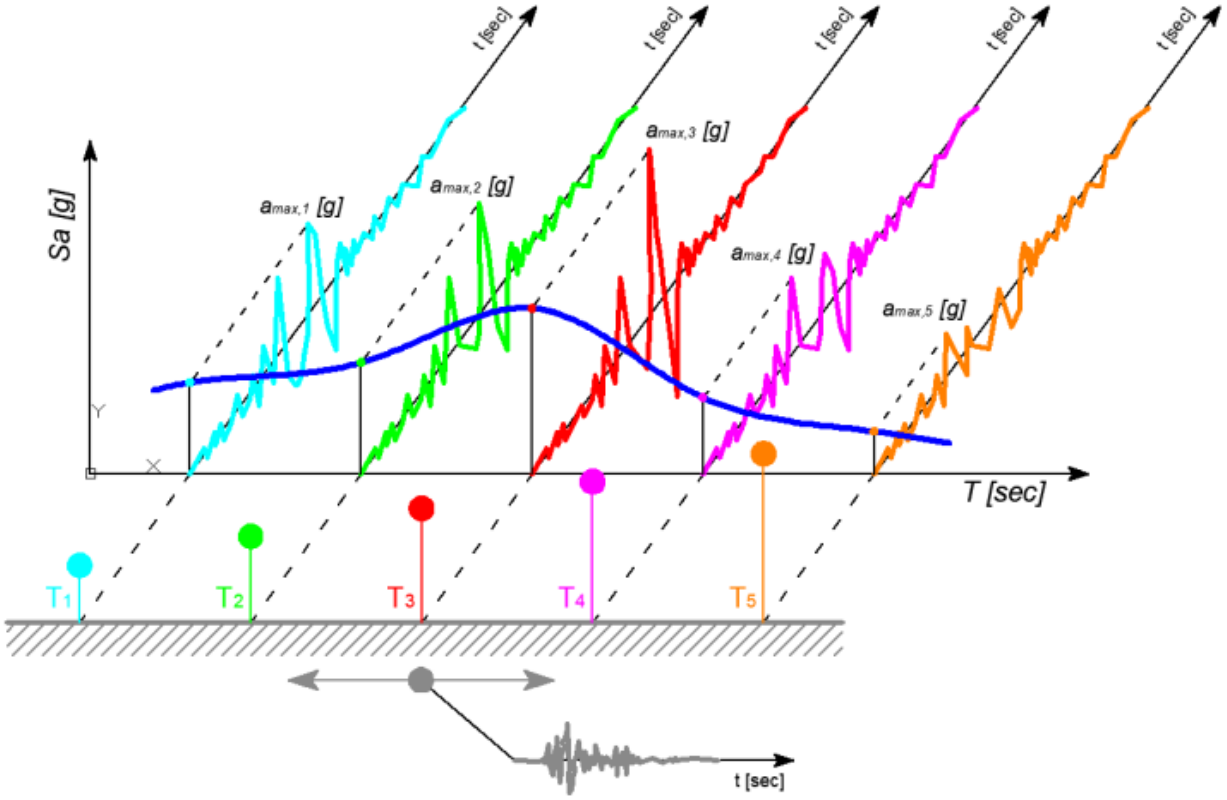

Figure 34. The bars on the bottom indicate buildings of increasing eigen-periods (T1, T2, …) put in motion by the earthquake represented on the bottom. The response of each building to the earthquake is given by the accelerogram in colour. The envelope of the maximum acceleration (PGA) of each spectrum, given as a function of the structural eigen-period, is called response spectrum.

3.2.1 Seismic hazard analysis (SHA)

The seismic hazard analysis (SHA) of a region is important for community safety and infrastructure design. This analysis provides detailed information on the probability and intensity of seismic events, forming the basis for seismic regulations and the design of structures resilient to seismic forces. Results from SHA are commonly applied in the seismic design of buildings. Additionally, planning critical infrastructure such as dams, nuclear power plants, pipelines, and offshore platforms requires specific studies on geotechnical

hazards, including SHA. Furthermore, a thorough assessment of seismic hazard is essential for the development of emergency plans and alertness measures, serving as critical elements for safeguarding human lives and reducing damage in seismic situations.

SHA can be conducted using both probabilistic and deterministic approaches, integrating different perspectives in the assessment of seismic risk. In the context of Deterministic Seismic Hazard Analysis (DSHA) specific-magnitude earthquakes and geographic locations are selected and considered (Kramer, 1996). These deterministic seismic scenarios represent hypothetical events that could occur in certain areas.

On the other hand, Probabilistic Seismic Hazard Analysis (PSHA) is frequently employed to denote a methodology for evaluating the ground motion level expected with varying probabilities at sites characterized by specific geological conditions over a specific period of time. This level of shaking is often quantified through a characteristic parameter, typically expressed as the maximum horizontal acceleration on rigid ground, known as PGA. This probabilistic approach considers various seismic sources, their characteristics, and the uncertainties associated with seismic parameters and with the specific modelling strategy.

In Italy the official seismic hazard map, on which the national building codes are based, is computed by means of PSHA approaches.

The most common probabilistic approach, introduced by Cornell (1968), is based on knowledge of earthquake source zones, earthquake catalogues, and ground motion attenuation relationships. The Cornell method is articulated into four steps (Figure 35):

1) Identification and characterization of seismogenic source.

2) Characterization of seismicity in each seismogenic source through a recurrence relationship of events, i.e. the annual occurrence rate of earthquakes with different magnitude, expressed in terms of Gutenberg & Richter (1944) or Youngs & Coppersmith (1985) laws.

3) Use of attenuation relationship or Ground Motion Prediction Equation (GMPEs) to predict seismic motion at any distance from the seismogenic source (see paragraph 4.5).

4) Evaluation of the effect of uncertainties in earthquake location, size and ground motion prediction. These uncertainties are combined typically according to ‘logic tree’ approaches, resulting in an assessment of the probability of exceeding a threshold value of the seismic motion parameter of interest within a specific period of time.

5) A number of assumptions underlying the Cornell-based PSHA approach exist, and specifically:

   i. earthquakes are assumed to be time-independent (Poissonian distribution),

   ii. the number of events per magnitude class is expected to follow an exponential distribution (Gutenberg-Richter law),

   iii. earthquakes are assumed to be uniformly distributed within seismogenic zones.

In earlier years the outcomes of probabilistic analysis were associated with what is termed "hard rock", meaning they did not account for the effect of shallow geological layers and local topographic conditions on the amplitude, duration, and frequency content of ground motion. In recent years, several attempts to include also these site effects in GMPE have been made. However, GMPE uncertainties remain very large, particularly at mid-to-high-frequencies.

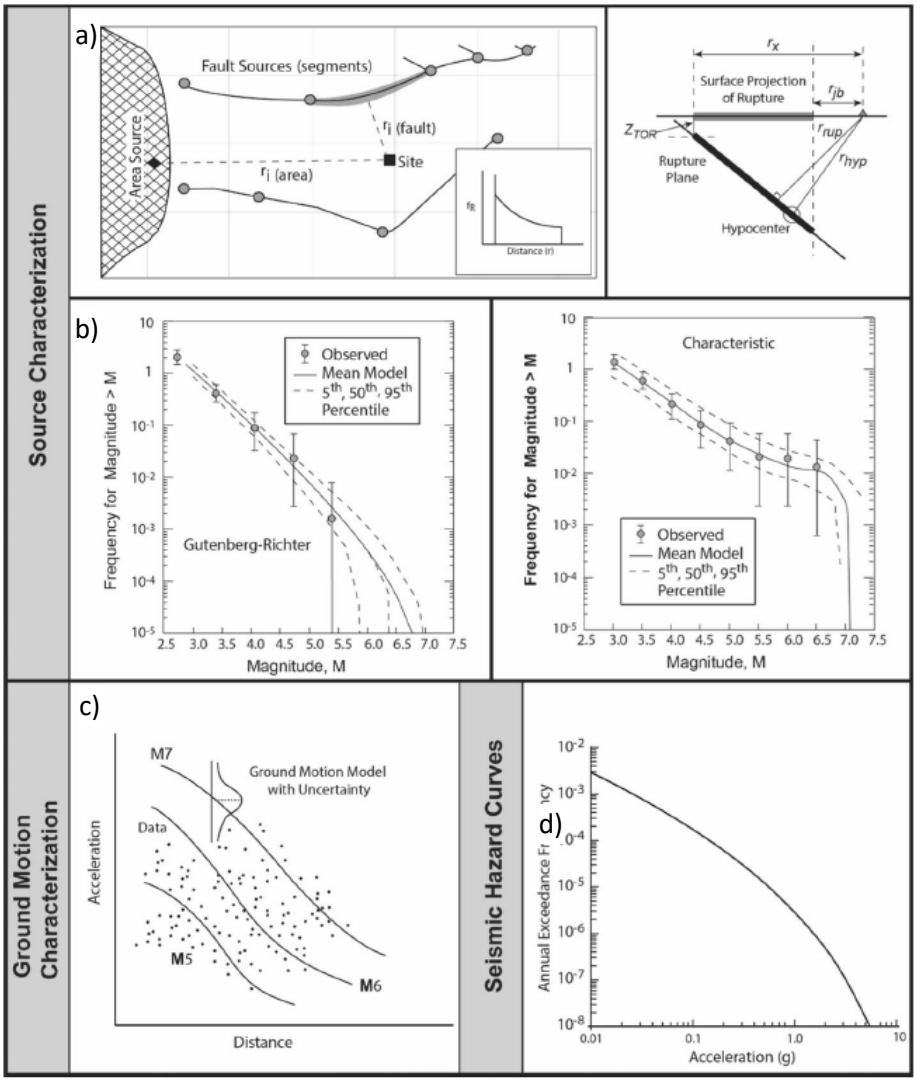

Figure 35. The different phases of PSHA (Probabilistic Seismic Hazard Assessment). a) Identification and characterization of seismogenic sources, (e.g. fault, or areal source). The variability in source to site distance can be describes as probability density function (Fr). b) Description of the seismicity of a seismogenic source in a given period of time. c) Generation of ground motion model and related uncertainty. d) Integration of the whole set of information to generate seismic hazard curves (Bommer 2022).

3.3 Ground Motion Prediction Equation (GMPE)

Ground Motion Prediction Equations are typically empirical models based on observed data which describe the median and standard deviation of ground motion (e.g., PGA or PSA) as a function of the earthquake magnitude and distance. The term "attenuation relationship" is often linked with GMPE, suggesting that ground motion amplitude decreases with distance from the seismic source, regardless of earthquake magnitude.

However, such relation is contaminated by a number of other effects (e.g., the soil condition, the specific range of frequency, and therefore wavelengths, that one considers, the focal mechanism of the considered events, etc.), the GMPE became more and more complex over the years to include also these effects. Various reviews of attenuation relationships have been conducted by several authors, especially by Douglas (2022) who reviewed GMPEs published from 1964 to 2022, summarizing 485 and 316 empirical GMMs developed in different parts of the world to predict PGA and elastic PSA ordinates, respectively. The selection of appropriate equations in areas where national regulations do not provide adequate PSHA studies can be a dauting task.

In PSHA, one typically uses GMPEs computed from data collected in the specific region of interest. However, if the tectonic setting, geological conditions, and seismic activity are similar, GMPEs do not follow to national borders.

In Italy, the most commonly applied GMPEs were described by Sabetta & Pugliese (1996), Akkar & Bommer (2010), Bindi et al. (2011), Cauzzi et al. (2015) and Lanzano et al. (2019b). These attenuation laws have been modified several times with the increase in available seismic data.

According to Al Atik et al. (2010), an empirical GMPE can be describe as a function:

$$Y = f(X_{es}, \theta) + \Delta \tag{4}$$

Where Y is the natural logarithm of the measure ground-motion (such as PGA, PGV or Sa at given period, T, etc.) observed at site *s* during earthquake *e*, $f(X_{es}, \theta)$ is the mean ground motion model (in log terms), X_{es} is the set of parameters that refers to earthquake *source* (i.e. magnitude, focal depth, style of faulting), the *path* (R, distance source to site), *site* effect (e.g. the average shear-wave velocity in the upper 30m,Vs_{30}, site response). The symbol θ is the vector of model coefficients, and Δ is the misfit of data (residuals). Starting from the general equation (4), we can derive a more specific expression by incorporating the previously mentioned parameters as follow:

$$\log(Y) = f_{source}(M) + f_{path}(M, R) + f_{site}(Vs_{30}) + \Delta \tag{5}$$

The term Δ can be decomposed in *between-event* residuals, δB_e, and *within-event* residuals, δW_{es} (Figure 36). The *between-event* residual describes the randomness in the source process such as stress drop and variation of slip in space and time (Al Atik et al., 2010). The *within-event* residual is the combination between ground motion variability coming from record-specific factor (Al Atik et al., 2010, Villani & Abrahamson 2015).

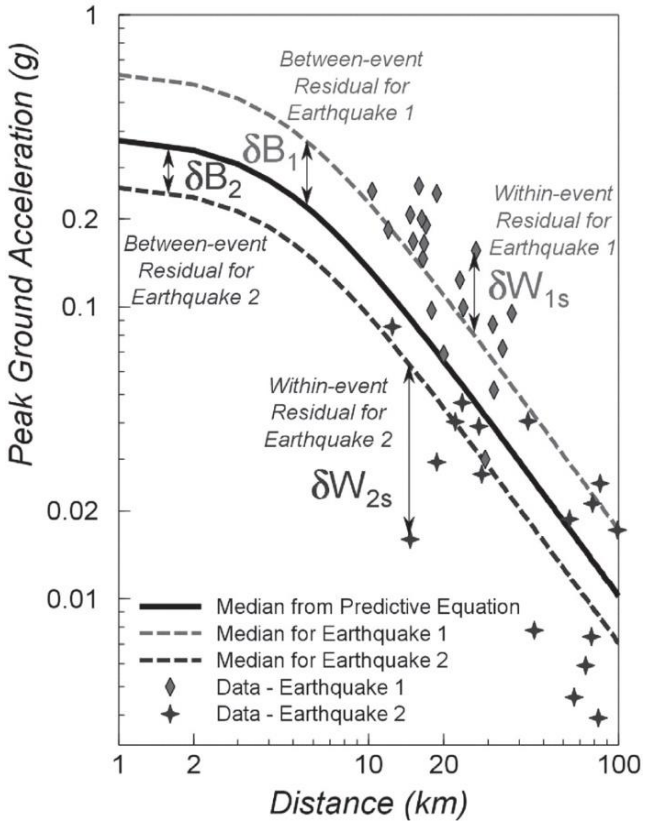

Figure 36. Example of median PGA as a function of distance from the epicenter for two earthquakes with the same magnitude. The variability of ground motion is examined through the concepts of Between-event and Within-event variability. The dispersion of Between-event residuals is described by the Between-event variability, δB_e. Similarly, within-event variability reflects the dispersion of within-event residuals, δW_{es} (Al Atik et al., 2010).

These residuals are normally distributed with zero mean and standard error, τ and φ respectively, therefore the total standard deviation of the ground-motion model is:

$$\sigma_{tot} = \sqrt{t^2 + \varphi^2} \tag{6}$$

The σ_{tot} can be divided into two components: aleatory variability, which is conceptually linked to true randomness in a process, and epistemic uncertainty, resulting from a lack of knowledge. In the context of Ground Motion Models (GMMs), the observed aleatory variability is a result of unmodeled effects, denoted as δB_e and δW_{es}. This observed variability arises because a simplified model is used for the complex earthquake process and wave propagation. Consequently, the value of σ_{tot} significantly affects the results of probabilistic seismic hazard analysis. This influence is extensively discussed in studies such as Douglas (2003),

Restrepo-Vélez & Bommer, (2003), Bommer & Abrahamson (2006), Strasser et al. (2009) , and Figure 37 is reproduced from these discussions.

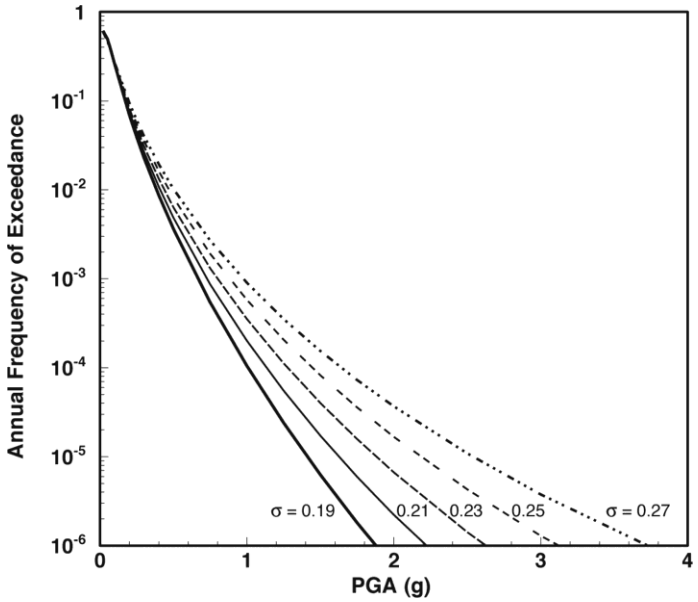

Figure 37 Impact of the aleatory variability of ground motion (σ) on seismic hazard results (Strasser et al., 2009).

If "*the ground-motion uncertainty computed from a global dataset (i.e., including various sites and various sources for multiple events) is assumed to be the same as the variability at a single site"* (Anderson & Brune, 1999, Rodriguez-Marek et al. 2013) we introduce ergodic assumption and as consequences we tend to significantly overestimate the aleatory standard deviation (Villani & Abrahamson, 2015). Hence, studies by Rodriguez-Marek et al. (2013) and Villani & Abrahamson (2015) emphasize the need to separate uncertainty in GMPEs and then remove the ergodic assumption to successfully reduce the total standard deviation.

Nowadays, with the growing volume of available seismic data, there has been a shift away from the ergodic assumption towards non-ergodic approaches. In these approaches, the repeatable terms of variability associated with source, path, and site can be better modelled (see Eq. 5). By doing so, these terms are separated from aleatory variability and incorporated into the quantification of epistemic uncertainty (e.g., Rodriguez-Marek et al., 2013; Villani & Abrahamson, 2015; Baltay et al., 2017; Lanzano et al., 2017). Instances of fully non-ergodic models can be identified in Lin et al. (2011), Baltay et al. (2017), Lanzano et al. (2017), Kuehn & Abrahamson (2020), Sgobba et al. (2019, 2021), Parker et al. (2020).

In recent years, empirical laws describing ground motion have been modified to also incorporate nonlinearity, thus becoming even more complex. Due to these complications, machine learning techniques are increasingly used to generate GMM (Mohammadi et al., 2023). Derras et al. (2012) used feed-forward neural network on KiK-net seismic database composed by 3,891 records from 398 sites and 335 earthquakes to predict PGA. Derras et al. (2014) developed GMM for Europe applying the quasi Newton back propagation method to estimate PGA, PGV and Sa using 1,088 recordings from 320 earthquakes. Pozos-Estrada et al.

(2014) used of ANN to predict the PGA and Sa for Mexican inslab and intraplate earthquakes. Dhanya & Raghukanth (2018) used 13,552 ground motion records from 288 earthquakes to generate a GMM for active shallow crustal regions applying ANN. Lilienkamp et al. (2022) used U-NET on KiK-net seismic database composed by 46,191 records from 65 seismic station and 2,864 earthquake to predict Sa at period T=1 s to generate nonergodic fully-data driven model for Kanto basin (Japan).

3.3.1 Selected empirical GMPE for Italy

We describe here three empirical GMPEs that we will use as a comparison term for the NN-GMM model that we are going to develop. Specifically, our comparison will be performed against the Akkar & Bommer (2010), Bindi et al. (2011), and Kotha et al. (2020) models.

- Akkar & Bommer (2010) developed a model calibrated on events that occurred in Europe, the Middle East, and the Mediterranean. The dataset employed in their study is comparable to the one utilized in Akkar & Bommer (2007).
- Bindi et al. (2011) developed a model calibrated on events that occurred in Italy. It is an upgrade of the ITA08 model (Bindi et al., 2010), enriched with recordings relevant to the 2009 L'Aquila seismic sequence.
- Kotha et al. (2020) upgrade dataset RESORCE (Akkar et al., 2014) based region- and site-specific GMM for shallow crustal earthquakes using the new European Strong-Motion (ESM) dataset (Bindi et al., 2018; Lanzano et al., 2019a).
- The main features behind these three models are listed in Table 7.

These models derive directly from Ground-Shaking Intensity Model Service (eGSIM), operating as a web service implemented on OpenQuake (Pagani et al., 2014). The eGSIM is an integral part of the European Facilities for Earthquake Hazard and Risk (EFEHR), a non-profit network composed of organizations and community resources dedicated to advancing earthquake hazard and risk assessment in the Euro-Mediterranean area.

Table 7. Key elements of the selected GMPE for Italy.

GMPE features	Akkar & Bommer (2010)	Bindi et al. (2011)	Kotha et al. (2020)
Number of earthquakes	131	218	242
Number of records	532	1213	18222
Horizontal component	GM	GM	RotD50
Minimum response period	0.05	0.04	0.01
Maximum response period	3	2	8
Magnitude scale	Mw	Mw	Mw
Minimum magnitude	5	4	3
Maximum magnitude	8.6	6.9	7.4
Ground motion parameter	PGA, PGV and Sa	PGA, PGV and Sa	PGA, PGV and Sa

3.4 Non-ergodic and fully data-drive GMM for Italy

Technological advancements and the availability of a large number of seismic stations in areas where they were previously lacking, has been a significant improvement in the availability and quality of seismic data. This progress has allowed us to move from traditional ground motion models based on ergodic assumptions, towards non-ergodic models that better capture repeatable terms such as source, site, and path and consequently reducing aleatory variability. This progress has also led to the necessity of generating seismic shaking maps, particularly in areas lacking observed seismic data, which has become increasingly relevant tool for earthquake engineering and territorial planning.

Recent studies, such as the one conducted by Sgobba et al. (2019, 2020) have developed an empirical non-ergodic Ground Motion Model for the northern and central regions of Italy. These models consider the spatial correlation of systematic residuals terms related to source, site, and path by employing geostatistical methods such as fitting a semivariogram and kriging interpolation to generate ground shaking maps.

In our study, we adopted an end-to-end approach utilizing a neural network, specifically the U-net. This will work as a filter trying to detect the relationships within the input data and providing a data-driven spatial interpolation with no need to define the type and parameters of an interpolation function, as shown in a prior study by Lilienkamp et al. (2022) on the Kanto Basin in Japan. Additionally, this approach should be designed to effectively reflect residual terms related to site, path, and source. This underscores the model capability to generate a non-ergodic GMM.

In our case, this method represents an innovative application to the Italian context, where this application to generate region-specific fully non-ergodic GMM had not been previously explored.

3.4.1 Ground motion dataset

The case study is focused on the northern to central parts of Italy and includes the epicenters of the earthquakes belonging to recent seismic sequences occurred in recent years. Data used in this study sourced from Italian strong motion data. From the ITACA database (ITalian ACcelerometric Archive; Brunelli et al., 2022), we extracted 22,392 records from various seismic stations, referring to 1,196 earthquakes occurred between 1990 and 2021 (Figure 38). This extensive collection provides a complete overview of seismicity in the considered region.

To ensure a targeted focus and optimize the effectiveness of the U-NET model, we selected stations with a dataset of at least 10 records each. Consequently, out of a total of 678 available stations, we retained and analysed 376, as highlighted in Figure 39. While the distribution offers comprehensive coverage of seismic

features, there is a noticeable concentration of stations along the Apennine Mountain range, where recent seismic sequences have occurred, as evident from the Figure 39.

The magnitudes of the events include in our dataset span the [3-6.3] interval, while the epicentral distances cover the [10-500] km range. The [10-80] km distance range is particularly well sample as highlighted in Figure 40.

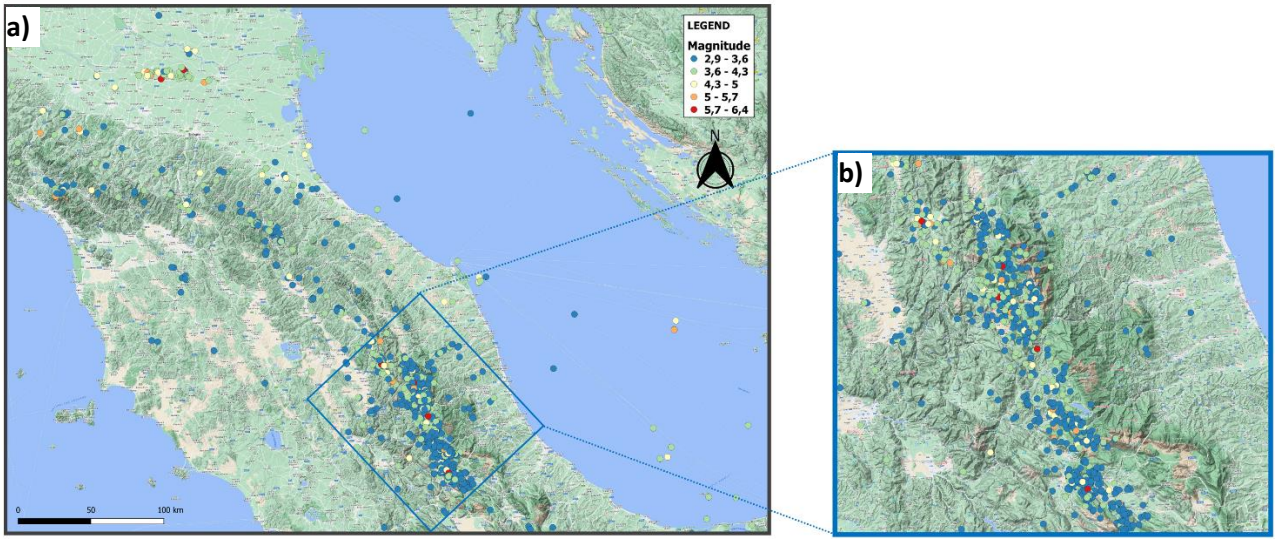

Figure 38. a) Distribution of seismic events used as input to train and test the U-Net built to predict ground motion decay from a seismic source. b) Zoom on central Italy zones, where the majority of recent seismic sequences occurred.

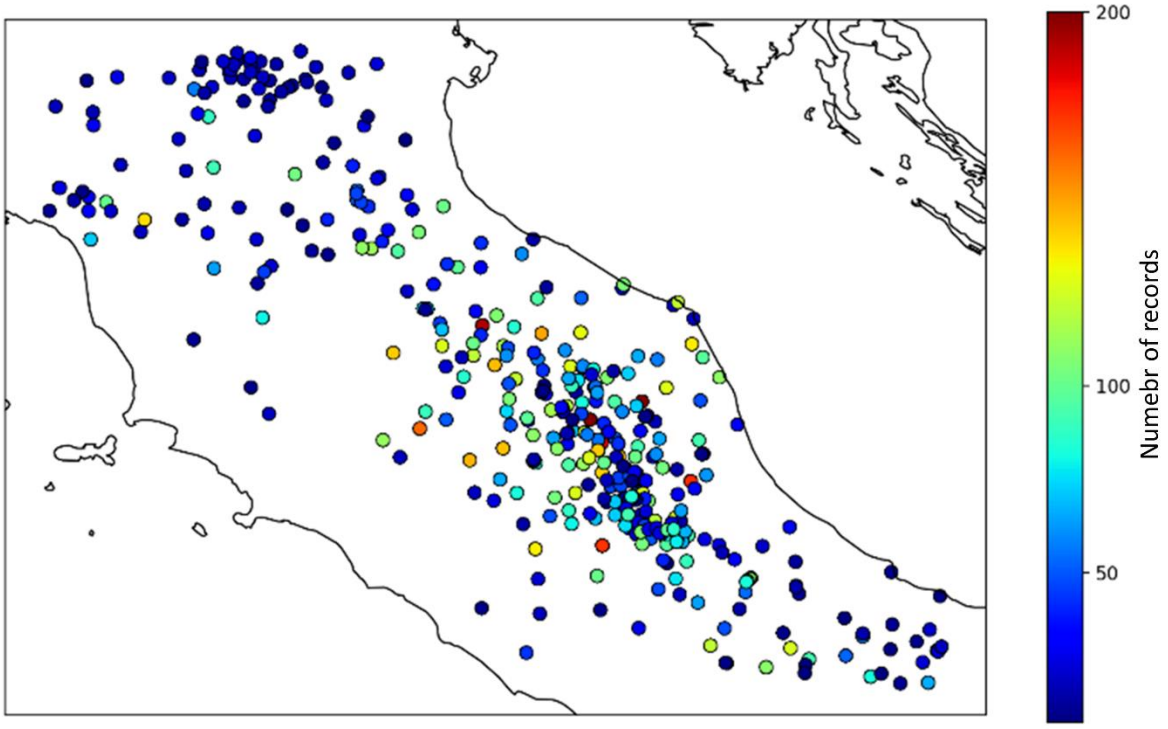

Figure 39. Distribution of seismic station (circle) used to train and test U-NET. Colours are proportional to the number of records available for each station, according to the colour scale on the right.

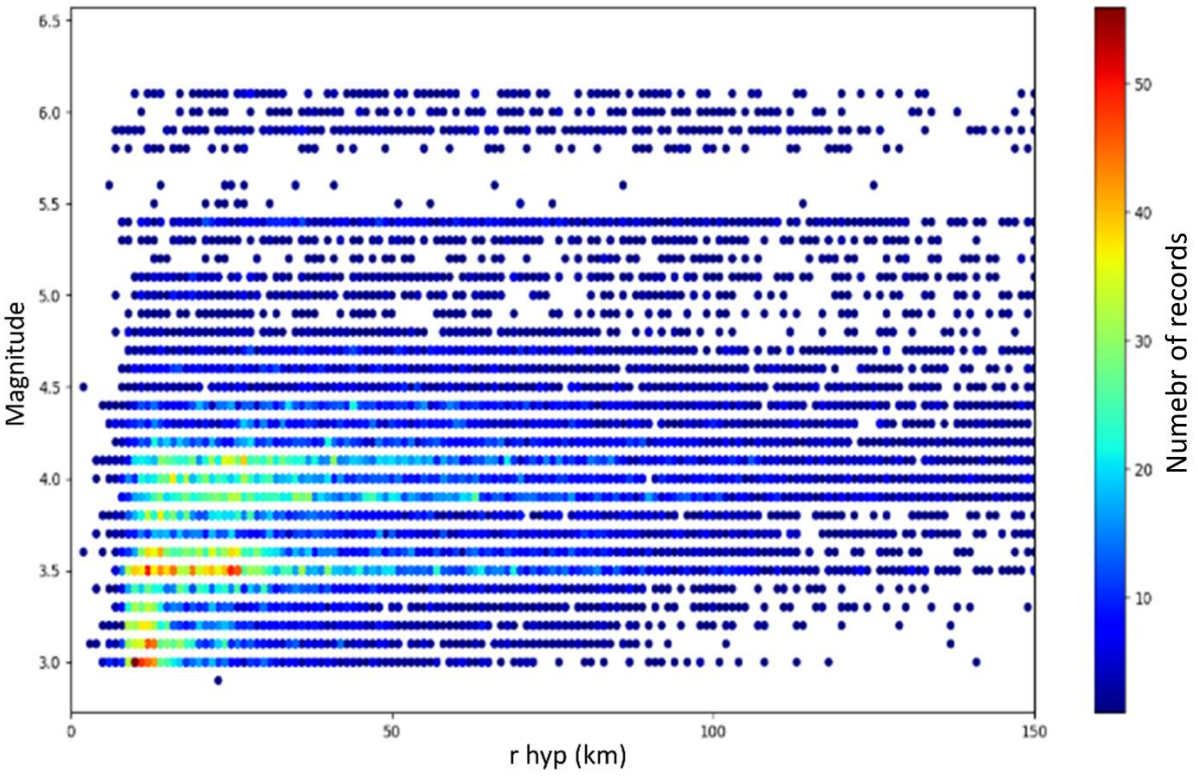

Figure 40. Distribution of all the seismic records that constitute the training and test dataset for U-NET as a function of their magnitude and hypocentral distance.

In our analysis, we designated the natural logarithm of the geometric mean of the two horizontal components at period of 1 second (ln (Sa(T=1s))) as the intensity measure (IM) of ground motion to predict on a spatial grid of points, starting from a given earthquake. The input and output grid that covers the selected portion of central Italy consists of 768 x 400 pixels (Figure 41 a). Each cell of the grid measures 700 x 700 m (Figure 41 b). We chose a period of 1 second due to its widespread application in probabilistic seismic hazard analysis.

Shorter periods (larger frequencies) would turn into short wavelengths and would be very sensitive to the details of local geology and geomorphology. Longer periods (smaller frequencies) would represent very long waves, possibly not even generated by small magnitude events.

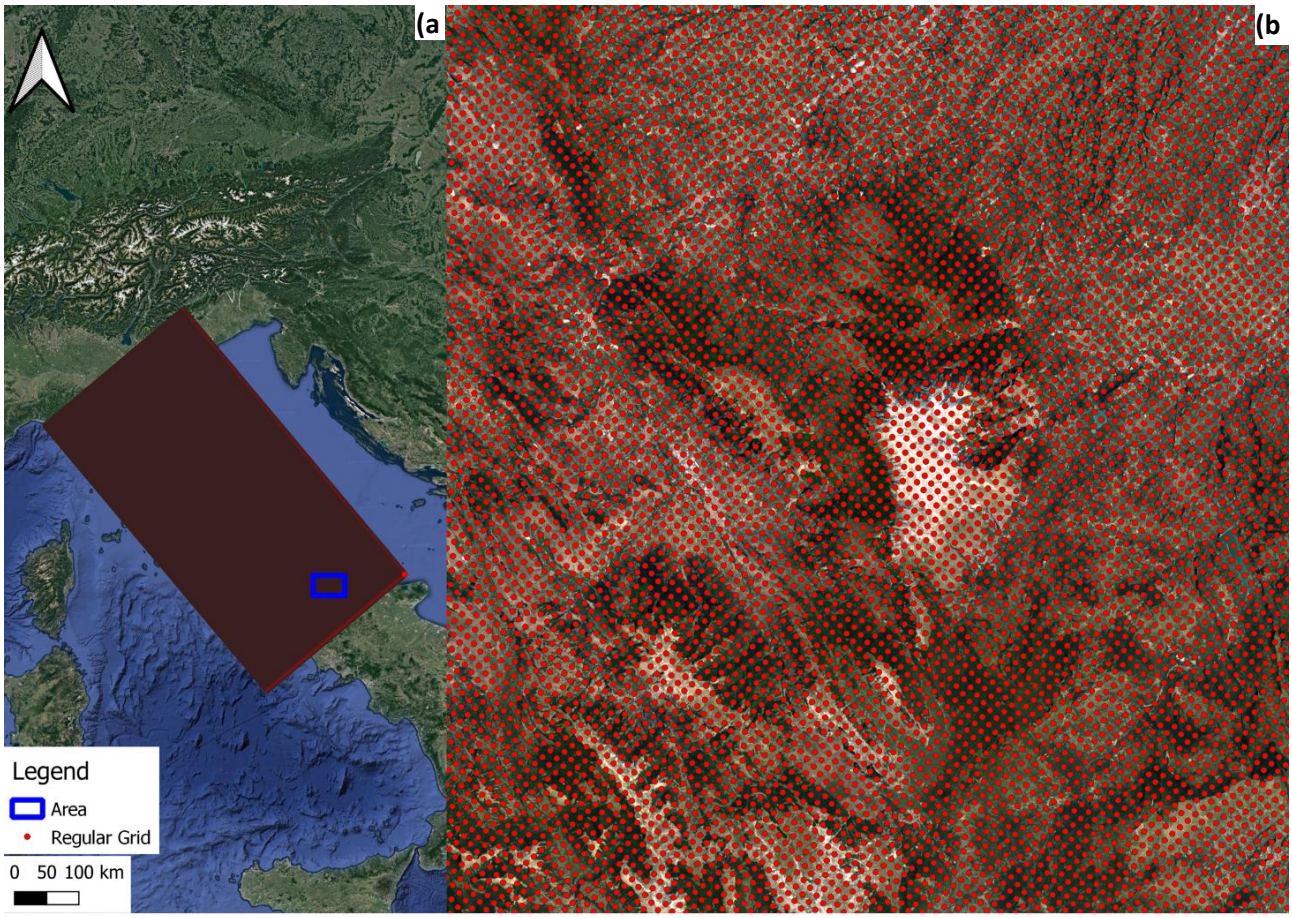

Figure 41 a) Regular grid (768 x 400 points) covering the study area, depicting the distribution of points (in red) from the northern to central regions of Italy. b) Zooming into a specific area (blue box) shows the distribution and regular spacing of points or nodes in the grid, with each point positioned at equidistant intervals of 700 x 700 m.

3.4.2 Input data for U-NET

Seismological parameters

In order to develop a neural net capable to predict ground motion at different distances from a given earthquake, we used as parameters for the training set all or some combinations among:

a. the 5% damping response spectrum ordinate at specific periods (e.g., 0.5, 1, 2 s)
b. local magnitude
c. geographic coordinates of the event
d. epicentral depth of event
e. hypocentral distance of the earthquake from each output evaluation point
f. soil category of the evaluation point
g. elevation and slope of the evaluation point

for each of the 22,392 selected records. These input data were referred to the position of the selected seismic stations (Figure 39) and the goal of the network is to predict their values at the nodes of a regular grid (Fig. 41).

Parameters a) to d) were extracted from ITACA.

For each seismic event, we identified the seismic stations that recorded the seismic signal. Subsequently, we linked the parameter a (i.e. Sa), which is our target for the network, to these specific grid points/nodes (seismic station), without extending the coverage to the entire grid.

The hypocentral distance (parameter d) was determined by first converting the geographic coordinates and epicentral depth (ϕ, λ, h) into cartesian coordinates using the ECEF (Earth-Centered, Earth-Fixed) formula:

$$X = (N(\phi) + h)\, cos\phi\, \cos\lambda$$

$$Y = (N(\phi) + h)\, cos\phi\, \sin\lambda$$

$$Z = \left((1 - e^2)N(\phi) + h\right) sin\phi$$

where:

$$N(\phi) = \frac{a}{\sqrt{1 - e^2 sin^2\phi}}$$

(7)

Where ϕ, λ, h, are latitude, longitude (in radians) and depth in meter, a = 6378137 is the equatorial radius (semi-major axis) in meters and represent World Geodetic System 1984 (WGS84) ellipsoid constant radius, and e = 8.181919084262 e^{-2} is eccentricity for WGS84 ellipsoid.
Once the conversion was completed, the hypocentral distance from each point of a regular grid was computed as:

$$Dist = \sqrt{(X_{rg} - X_{ev})^2 + (Y_{rg} - Y_{ev})^2 + (Z_{rg} - Z_{ev})^2} \quad (8)$$

Where X_{rg}, Y_{rg}, Z_{rg} and X_{ev}, Y_{ev}, Z_{ev} are cartesian coordinate of regular grid (rg) and epicentral depth of event (*ev*), respectively.

The parameter b, magnitude, generally expressed as moment magnitude (Mw) was calculated from local magnitude (ML) estimates by using regression equations of the type:

$$M_w = bM_L + a \quad (9)$$

where, following Castellaro et al. (2006), b = 0.890 and a = 0.283.

From surface geology to geophysical data

To obtain shear wave velocities that will be used to define the soil category (input in the list above) we used a shapefile derived from the European Seismic Risk Model (ESMRS20 project, from Weatherill et al., 2023). This is a collector of lithological and stratigraphic information about Europe, from three distinct sources, integrated in a unified map (Figure 42 a). After identifying the most common lithologies in Italy (Figure 42 b), we grouped them into three main categories (Figure 42 c)**:**

i) **loose sediment** (CLAY+SAND+ALLU),

ii) **soft rocks** (SSNT+MARL+LMST) and,

iii) **hard rock** (VOLC),

characterized by upper-interval velocity of 300, 1600, 2400 m/s, respectively.

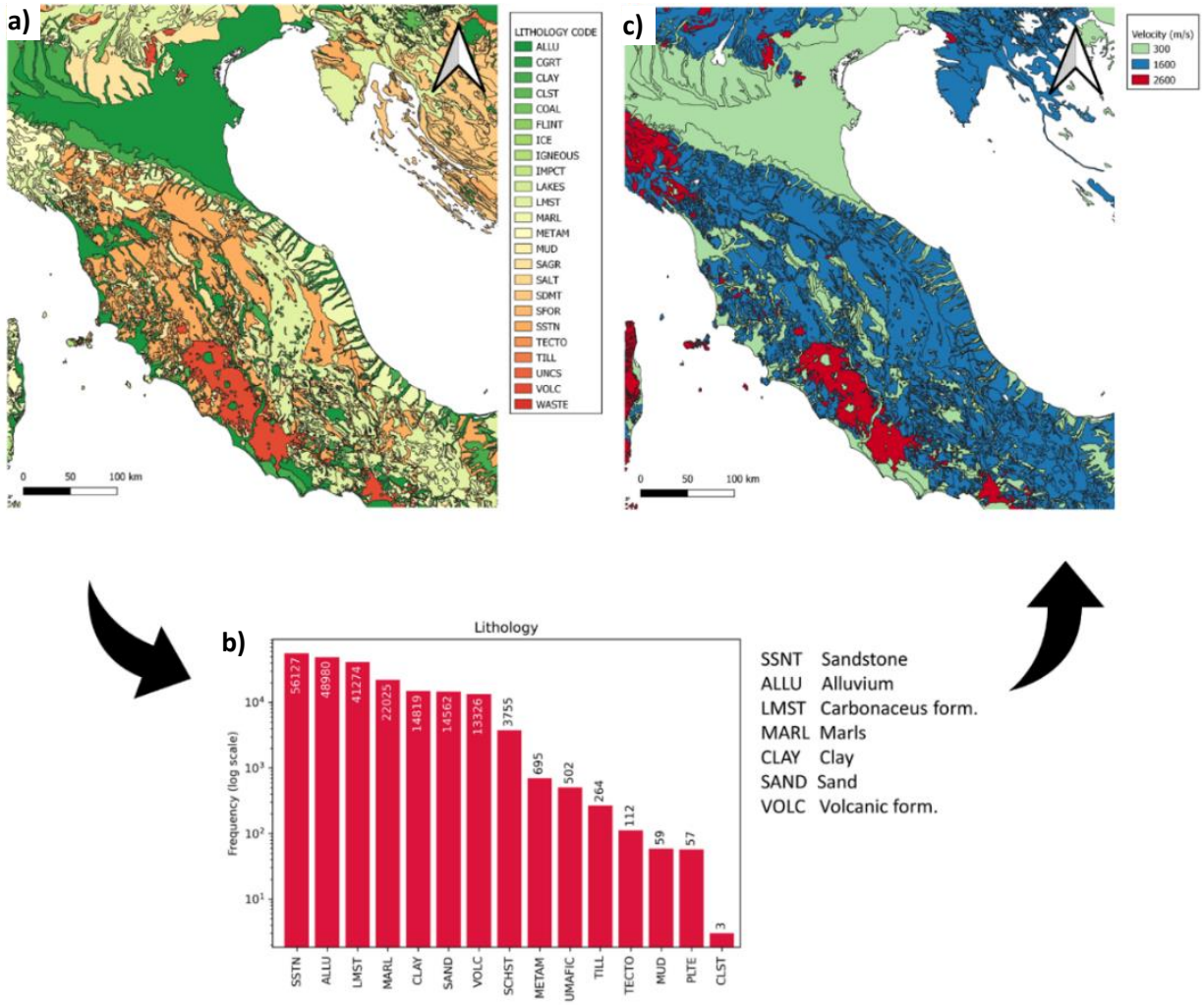

Figure 42 a) Stratigraphic information for the area of interest according to the categories of the ESMRS20 project, b) gathering of the lithological formations of panel a) into three main groups, based on their Vs_{30} values as loose sediment [0 300] m/s, soft rock [300, 1600] m/s, and hard rock > 1600 m/s.

Morphological data

The morphological information (topography and slope) of the studied area were acquired using the TINITALY (Tarquini et al., 2023) digital elevation model (DEM), with resolution of 10-meter cell size. The geographical projection used is the Universal Transverse Mercator (UTM), under the WGS 84.

To extract elevation points from the DEM and link them to each point in the regular grid (Figure 41), we utilized a *Point Sampling tool* available in QGIS. In Figure 43, we can observe the distribution of the selected topographical data. Furthermore, in addition to topography, we also calculated slope as an indicator of terrain steepness or inclination.

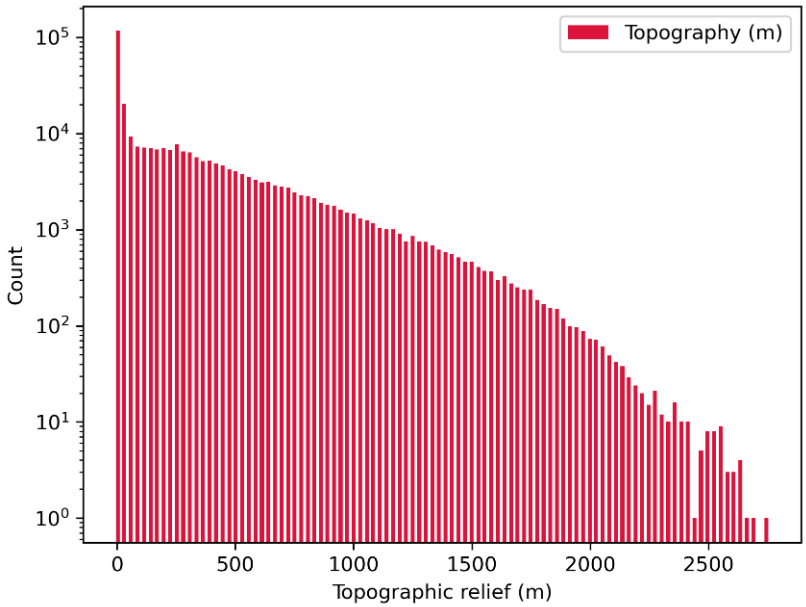

Figure 43. Distribution of the elevation above the mean sea level of the points in the input/output grid used to train and test the U-NET.

3.5 Architecture and functioning of U-Net

The type of neural network chosen to estimate ground motion at the nodes of the grid (Figure 41) starting from all or a combination of parameters listed on page 65 is the U-Net, developed by Ronneberger et al. (2015). The input parameters are presented to the U-Net for training in the form of a stack of maps covering the selected geographic area. Each layer in this stack maps moment magnitude (Mw), hypocentral distance ($\ln(r_{hypo})$), depth, coordinates (longitude and latitude), velocity, and topographic information, as detailed in the previous sections. The U-Net processes this input stack through its various layers, encompassing contraction, maximum depth, and expansion phases (see paragraph 1.3.4.3). Throughout this process, the

network learns the relationships and patterns present in the data. Ultimately, the U-Net produces two output maps, representing the predicted intensity measure (IM) of the ground motion at each node of the selected grid, in terms of mean and variance.

3.5.1 Training phases

In its initial state, when the U-Net is still untrained, it functions as a random outputs map generator. The aim of our process is to obtain accurate estimates for the mean ($\hat{y}$) and variance ($\hat{\sigma}^2$) of the target IM. To achieve this goal, the U-Net clearly has to be trained. During the training process, the U-Net minimizes the loss L_e, indicating the discrepancy between the U-Net outputs and the observations y_e for an event e with N_{obs} observations. This minimization follows a negative log-likelihood approach based on a normal distribution (Lilienkamp et al., 2022):

$$L_e = \sum_{n=1}^{N_{obs}} ln\hat{\sigma}_{e,n}^2 + \frac{(y_{e,n}-\hat{y}_{e,n})^2}{\hat{\sigma}_{e,n}^2} \quad (10)$$

The loss is iteratively reduced using the gradient descent method, specifically the Adam optimizer (Kingma & Ba, 2015). The gradients are calculated through backpropagation. The training process basically corresponds to a nonlinear least-squares regression.

After each training epoch, the U-Net is tested on new seismic events. These belong to a validation dataset, which includes events that weren't part of the initial training.

The ability of the U-Net to apply its learned knowledge to new events is checked to identify the moment when the improvement in predicting the validation dataset slows down. This indicates that the model has reached a point where it cannot significantly improve its performance in accurately predicting new, previously unseen events. Upon reaching this point, the training is considered completed, and the U-Net can thereafter be utilized as a GMM.

We highlight that, when representing observations (y) as point-wise measurements, the U-Net provides predictions (mean and variance) in the format of maps indicating that the U-Net has the capability to automatically interpolate the acquired attenuation relationship throughout the entire output area. This interpolation occurs even if the loss function is computed solely at points corresponding to actual observations.

Before proceeding with the actual training of the network, we dedicated significant effort to a phase of preliminary experiments. The goal of this phase was to explore various key parameters to optimize the overall performance of the neural network. Additionally, as a common practice in neural network applications, we normalized the data (input and target) using the formula:

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (11)$$

Where x represents the value of the variable to be normalized, in our case M, r_{hypo}, longitude and latitude, depth, topography and velocity and natural logarithm of the geometric mean of the two horizontal components at period of 1 s.

This scaling process aimed at ensuring that the normalized values fall within the range of 0 to 1 with uniform distribution of data, facilitating a more effective convergence during the training phase.

During the initial training phase, we encountered the challenge of "edge effects", which not only obscured interpolation but also masked the mean values of ground motion intensity, resulting in a concentration of higher values at the map edges (Figure 44 a). These effects specifically occurred from our initial use of "zero padding". This technique, involving the addition of zeros around the input matrix, caused issues during convolution. Pixels at the matrix edges, surrounded by zeros, were only partially influenced by the kernel window, risking the loss of important information at the map edges.

Subsequently, to address this issue, we used "symmetric padding". In this approach, instead of adding zeros, we symmetrically extended the original data beyond the matrix edges. This modification proved effective in preserving data coherence and continuity, eliminating edge effects, and enabling the network to handle the spatial distribution of motion intensity values more effectively during training (Figure 44 b).

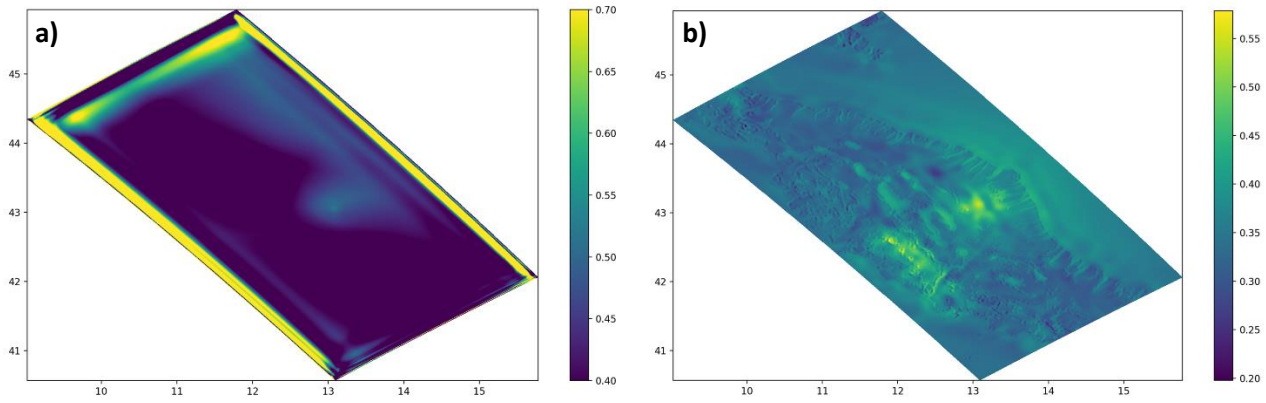

Figure 44 a) Example of original output map of the U-Net where border effects are clearly visible. b) Same output map after border effect removal achieved by adding symmetrical padding. The color scale represents normalized values of Sa at period T = 1 s based on equation (11).

Furthermore, we explored the sensitivity of the network to different activation functions (Tanh, LeakReLU, ReLU, PReLU), testing various number of epochs and learning rates (see cap. 1.3.2.3).

We set the number of epochs to 200 during the model training to evaluate its performance. This was part of an initial exploration phase aimed at understanding where the model showed the best convergence. In case of underfitting that indicating insufficient capacity of the model to capture complexities in the data, we

gradually increased the number of epochs, reaching a maximum of 5,000, to identify the optimal convergence point, which ultimately turned out to be 700.

Simultaneously, we varied the learning rate during these experiments, exploring values ranging from 0.1 to 0.0001. Experimenting with different learning rate settings, we aimed at identifying an optimal configuration that ensured effective convergence without falling into a situation where the model overly adapts to the training data, losing its generalization ability. The optimal learning rate was found to be 0.0001.

Once these parameters were selected, we adopted a gradual approach to model optimization, beginning with the simplest configuration that comprised only two inputs: magnitude and distance. Subsequently, we systematically increased complexity by incorporating additional variables, such as event coordinates, depth, velocity, topography, and slope (Figure 45). The decision to develop the neural networks with an expanding number of inputs is guided by the necessity to assess whether the inclusion of new parameters increase the predictive capability of the model. When there was no improvement, this suggested that the added parameter did not exert a significant influence on the output, or that its representation was not accurate enough to exert an influence on the output.

Another factor to consider during training is the variable we are attempting to estimate, for example, Sa at different periods (e.g. 0.1, 0.5, 1, or 2 s). We know that the wavelength is proportional to the period. Hence, at shorter periods, the wavelength is shorter, whereas at longer periods, the wavelength is longer. Therefore, when training a neural network to predict the response spectrum amplitude, we need to consider that at shorter periods (e.g., 0.1 s) the network may require greater detail in the input information (e.g. surface geology). In other words, we expect a network trained to predict the Sa at period of 2 s to perform better than one trained for 0.1 s, due to the input data resolution.

To provide a quantitative measure of the performance of models on a validation database, we opted for a trial-and-error approach, based on the analysis of error metrics such as Root means square error ($RMSE$) and coefficient of determination (R^2).

$RMSE$ can be expressed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y_i - \hat{y}_i\|^2}{N}} \tag{12}$$

where N is the number of data points, y_i is the i-th observation, and $\hat{y}_i$ is its corresponding prediction.

R^2 is calculated as:

$$R^2 = 1 - \frac{\sum(\hat{y}_i - y_i)^2}{\sum(\bar{y} - y_i)^2} \tag{13}$$

where, in the numerator, there is the sum of squared errors generated by our model, and in the denominator, the sum of squared errors generated by the model that uses the mean ($\bar{y} =$ mean of observation).

This process guided us toward the optimal input configuration, identifying which variables significantly contribute to enhance the model performance and which might have limited impact. Specifically, we selected the configuration with a high R^2 value, indicative of our model ability to explain variability in the validation dataset, and a low $RMSE$ value, measuring the accuracy of predictions ($\hat{Y}$) compared to actual observations (y).

Table 8. Different sets of data used to train the U-Net n. The first column tells the configuration name (Config). A X in the other columns indicate whether magnitude (M), natural logarithm of hypocentral distance (ln (r_{hypo})), longitude (Lon), latitude (Lat), epicenter depth, shear wave velocity (vel), elevation (Topo), slope were used as input in each training configuration. The last two columns indicate the root mean square error (RMSE), and coefficient of determination (R^2) of the predicted vs expected output on validation dataset.

Config	M	ln (r_{hypo})	Lon	Lat	Depth	Vel	Topo	Slope	RMSE	R^2
2 c	X	X							0.71	0.79
4 c	X	X	X	X					0.75	0.77
5 c	X	X	X	X	X				0.71	0.79
6 c	X	X	X	X	X	X			0.70	0.80
7 c	X	X	X	X	X	X	X		0.63	0.83
8 c	X	X	X	X	X	X	X	X	0.69	0.80
5 c bis*	X	X			X	X	X		0.68	0.81
7 c bis*	X	X	X	X	X	X		X	0.66	0.82

After a thorough analysis of the performance of different model configurations, we identified as optimal configuration the one with seven inputs (configuration 7 c in table 8, magnitude, radius, depth, lon, lat, velocity, topography), characterized by a RMSE of 0.63 and a R^2 of 0.83. The RMSE value can be consider in the range compared to the typical ranges of GMPE found in the last 40 years (Strasser et al., 2009).

After the identification of the most appropriate configuration, we reviewed the analysis to assess the impact of the different parameters on the model performance. For instance, in the testing with configuration 2 c (using only magnitude and distance), the addition of longitude and latitude led to an increase in RMSE (configuration 4 c). To explore this effect, we created configuration 5 c bis*, excluding longitude and latitude from configuration 7c. The results indicated that, while configuration 5 c bis* surpassed configuration 5 c, it still didn't reach the same level of the configuration 7 c (Table 8).

Subsequently, we explored the possibility of improving model performance by replacing topography with slope (7c bis*). However, the results obtained did not show a substantial improvement. This could be because

the topographic distribution might already include slope implicitly. Additionally, calculating slope based on 700-meter side cells may not be realistic and thus not meaningful.

These choices are guided by the search for the most effective configuration, balancing model performance with the number of inputs considered. Configuration 7c has proven to be the optimal choice in terms of predictive accuracy.

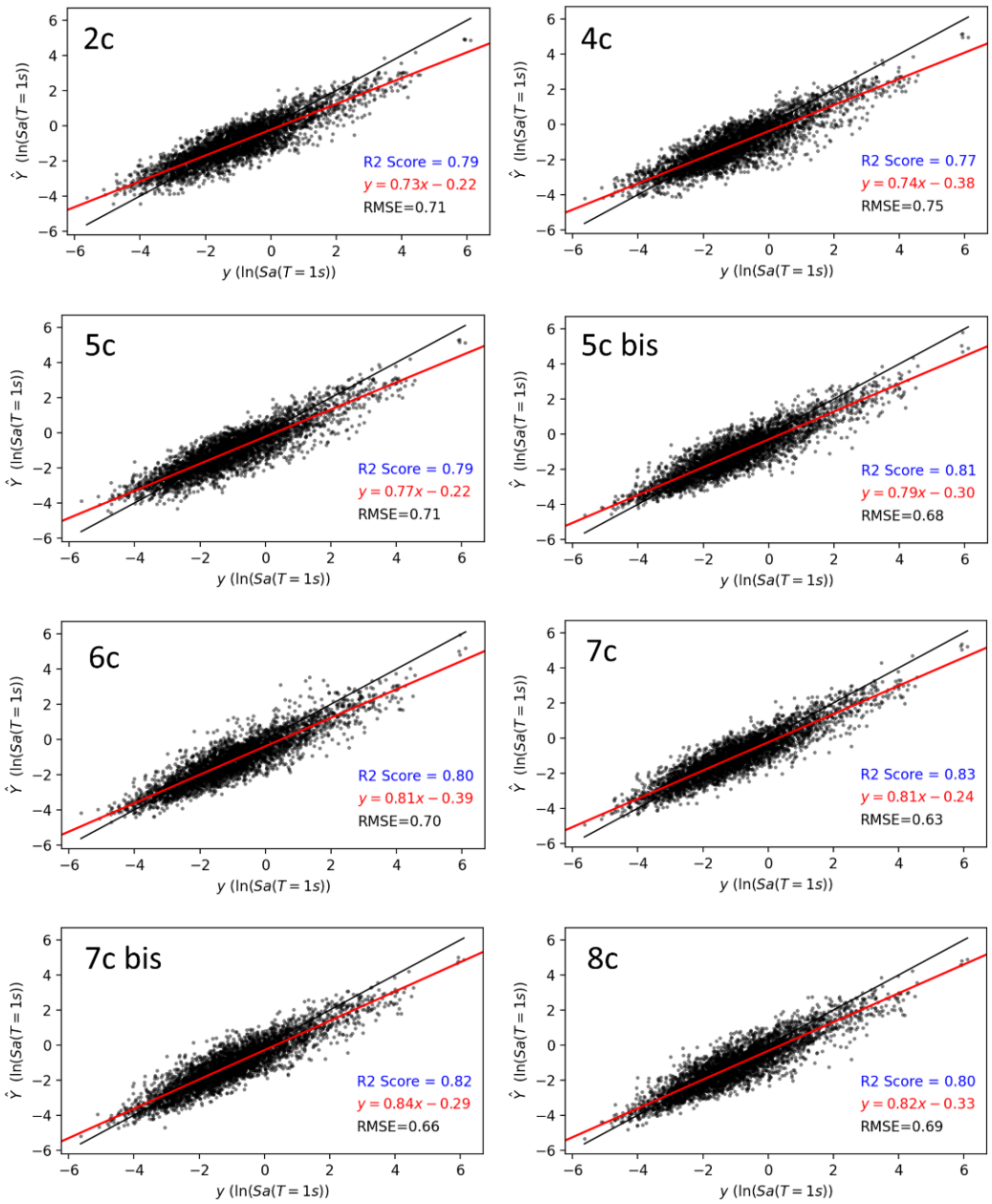

Figure 45. Predicted (y-axis) vs observed (x-axis) response spectra amplitudes at period T = 1 s for different training configurations, listed in Table 8, used for the U-Net. The black line indicates the ideal case of total coincidence between the predicted and observed values; the red line is the least-squared regression.

Starting from the configuration with seven inputs as a starting point, we explored the potential impact of varying reference periods (Figure 46). Initially, we had adopted Sa at 1 second as the main target.

Subsequently, we conducted a series of tests exploring different periods (i.e., Sa at 0.15, 0.20, 0.25, 0.35, 0.5, 0.6, 0.7, 0.8, 0.9, 1.2 and 2 s).

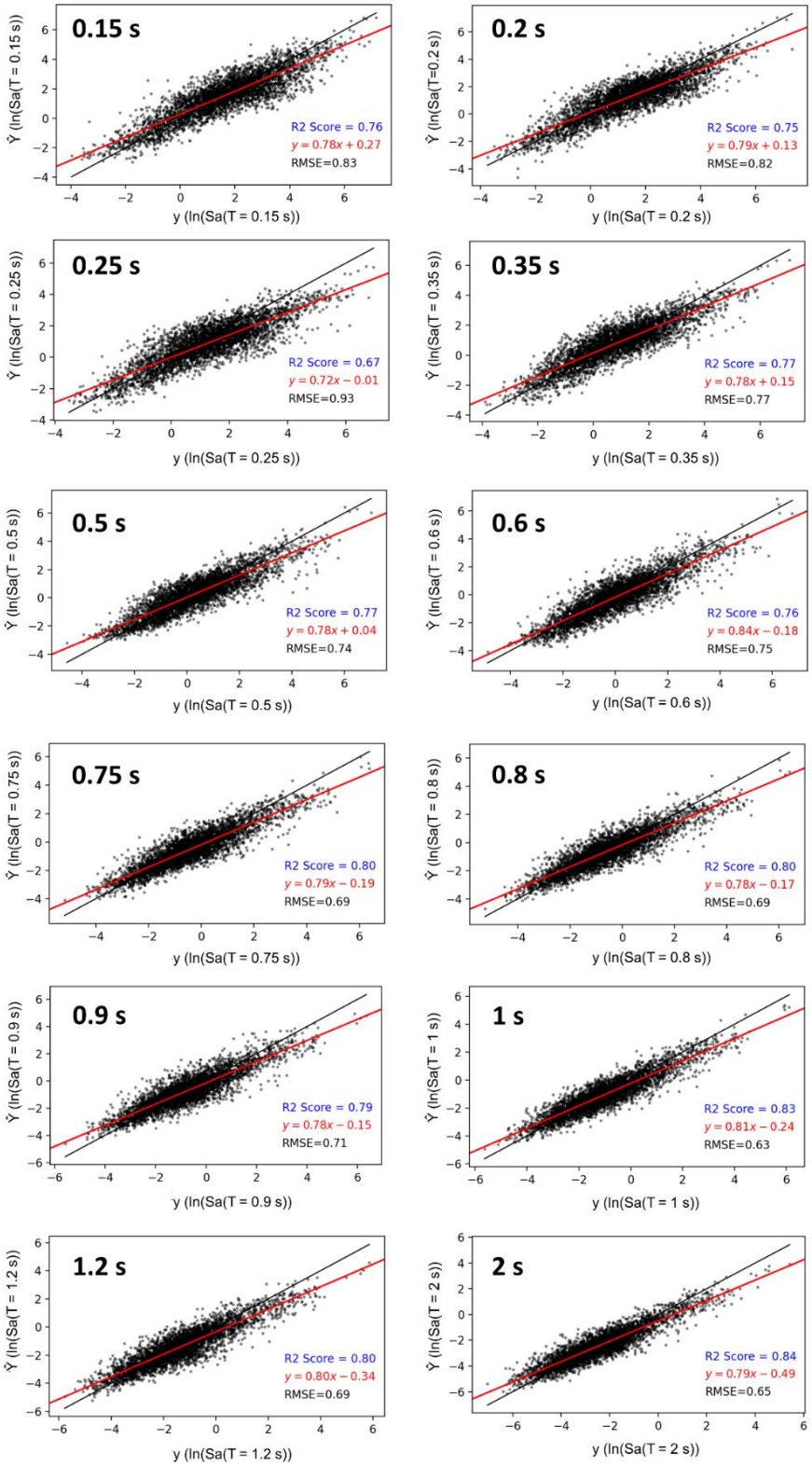

Figure 46.Predicted (y-axis) vs observed (x-axis) response spectra amplitudes at different periods (indicated in each panel) for the configuration 7c. The black lines indicate the ideal case of total coincidence between the predicted and observed values; the red line is the least-squared regression.

As expected, the R^2 of the model increases with increasing periods while the RMSE, on the opposite, decreases (Figure 47). This suggests that an accurate description of input related to surface geology may be crucial, especially at lower periods associated with higher frequencies. Conversely, at higher periods, topography and surface geology (our case velocity) appear to have limited or no influence, based on the observed trend in the data.

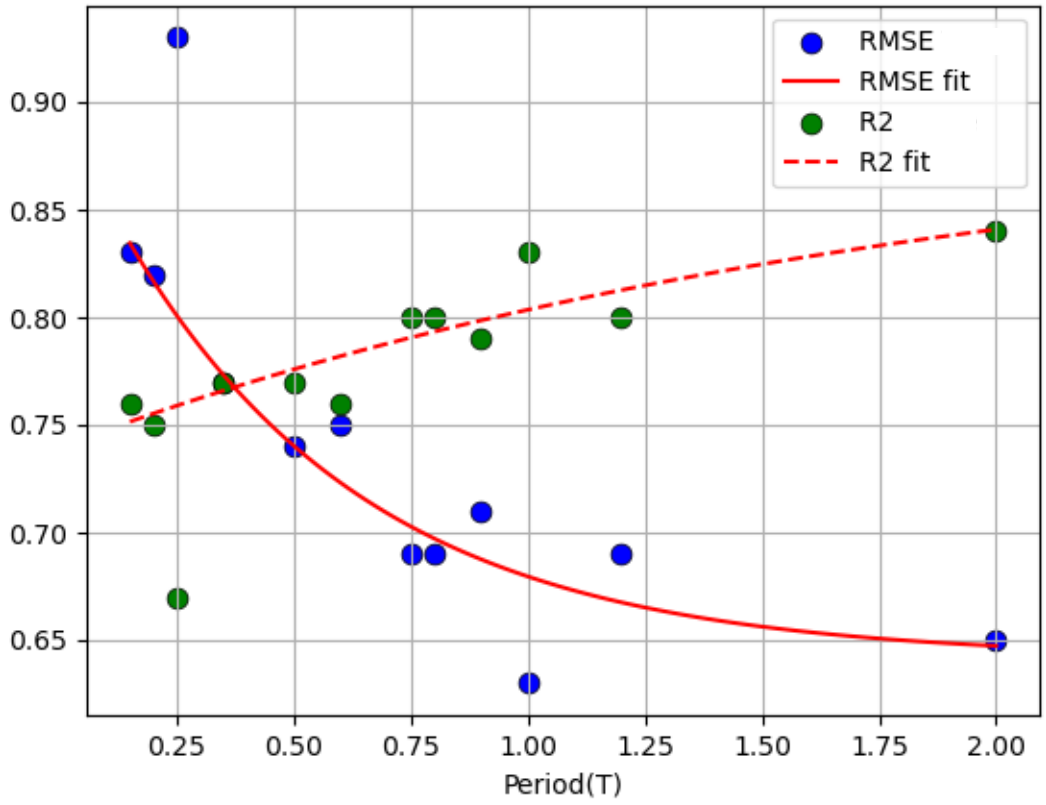

Figure 47. RMSE and R^2 values for U-Net predicting response spectra amplitudes at different periods (T). Dashed red lines represent the R^2 fit, while the solid red line represents the RMSE fit.

3.6 Evaluation of U-Net predictions

The U-Net residuals, representing the difference between the observed values at seismic stations and the predicted values from the neural network at the corresponding locations, are shown in Figure 48. The dependence of the residuals with the hypocentral distance is not centered around zero, suggesting potential underestimation of values predicted by the model compared to observed values. Additionally, we note higher uncertainties at shorter distances, gradually decreasing with distance as expected.

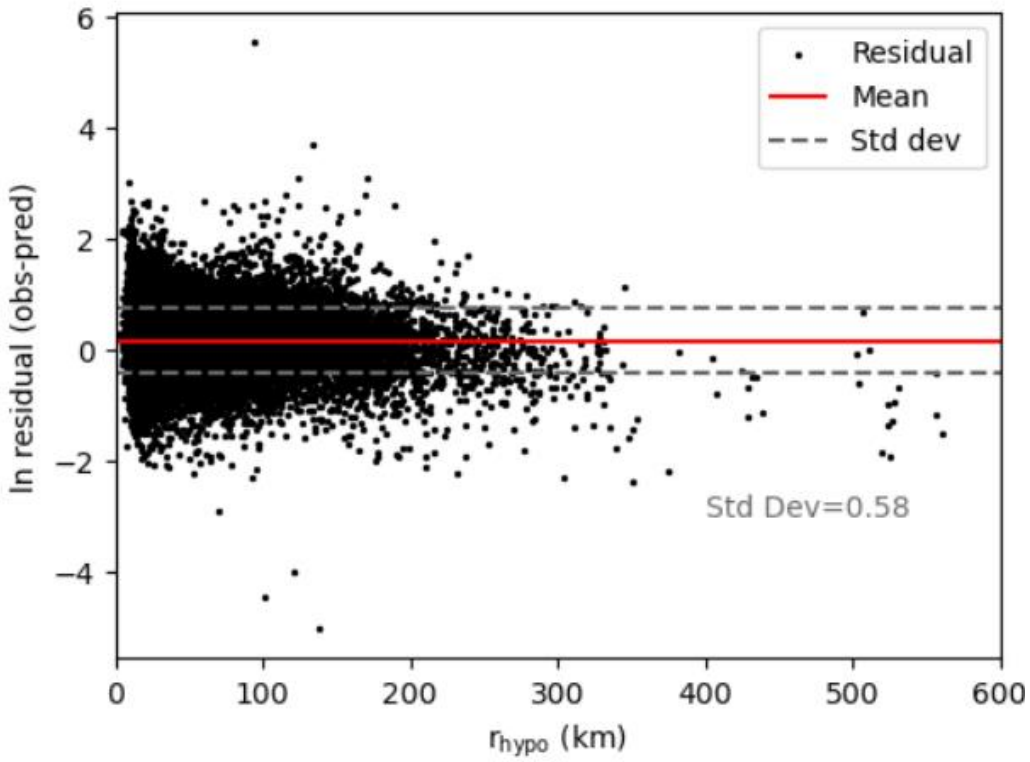

Figure 48. Distribution of ln(SA(T=1s) residuals, the red solid line shows the mean and dashed line the standard deviation.

We then standardize the residuals ($\tilde{\Delta}_n$) by dividing each individual residual by its predicted standard deviation as follows:

$$\tilde{\Delta}_n = \frac{y_n - \hat{Y}_n}{\hat{\Sigma}_n} n \in \{1, \dots, N\} \quad (14)$$

in which N denotes the number of observations from events.

The standardized residuals and the distributions of $\tilde{\Delta}_n$ for the training and validation sets, as well as the combination of both, appeared similar and approximately normal (Figure 49).

These findings suggest that the model effectively captures variability in the data, and its key assumptions, such as the uniform distribution of residual variance (homoscedasticity) and normality of residuals, seem to be reasonably satisfied.

Despite the overall good shape of the distribution of $\tilde{\Delta}_n$, we observed that the mean is not centered around zero, indicating a tendency of the model to underestimate the observed values. However, no clear patterns or significant asymmetries emerged, and the distribution shows uniform dispersion around the mean. The standard deviation of $\tilde{\Delta}_n$ for the entire dataset is 1.04.

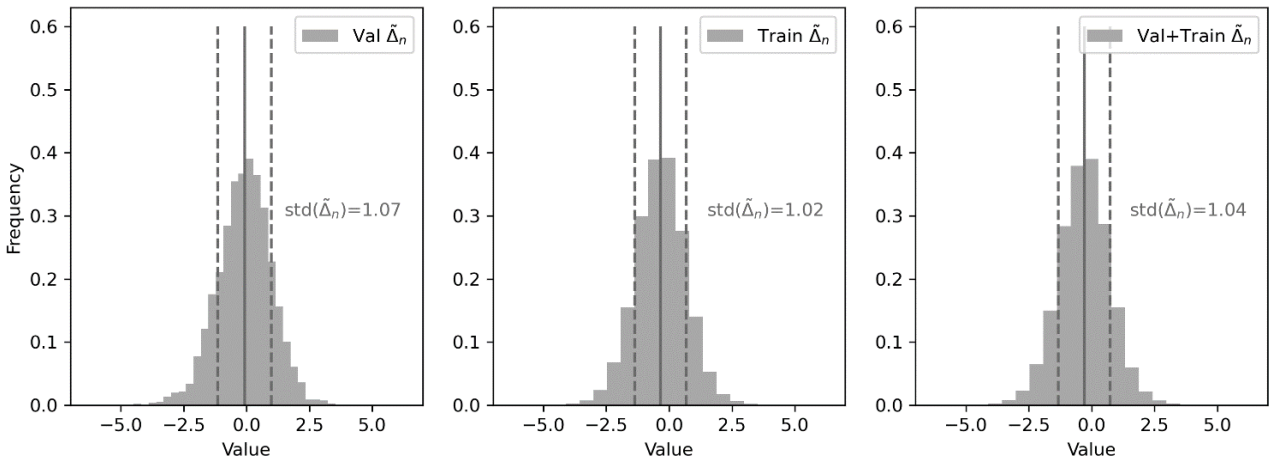

Figure 49. Standard normally distribution of standardized residuals ($\tilde{\Delta}_n$) for validation, training, and entire dataset. Solid and dashed vertical lines indicate the mean and standard deviation (std) of standardized residuals($\tilde{\Delta}_n$) respectively.

In Figure 50a, we present an example of a prediction map derived from our validation dataset, focusing on the 2009 L'Aquila earthquake with magnitude of 6.1.The obtained output from U-Net seems to somehow replicate the complex surface geology (Figure 50 b). Specifically, we expect that in areas with loose sediments, the predicted values will be significantly higher, reflecting the expected seismic amplification in such contexts. Conversely, in regions characterized by soft and hard rocks, we expected lower predicted values, consistent with the expected lower amplification during a seismic event.

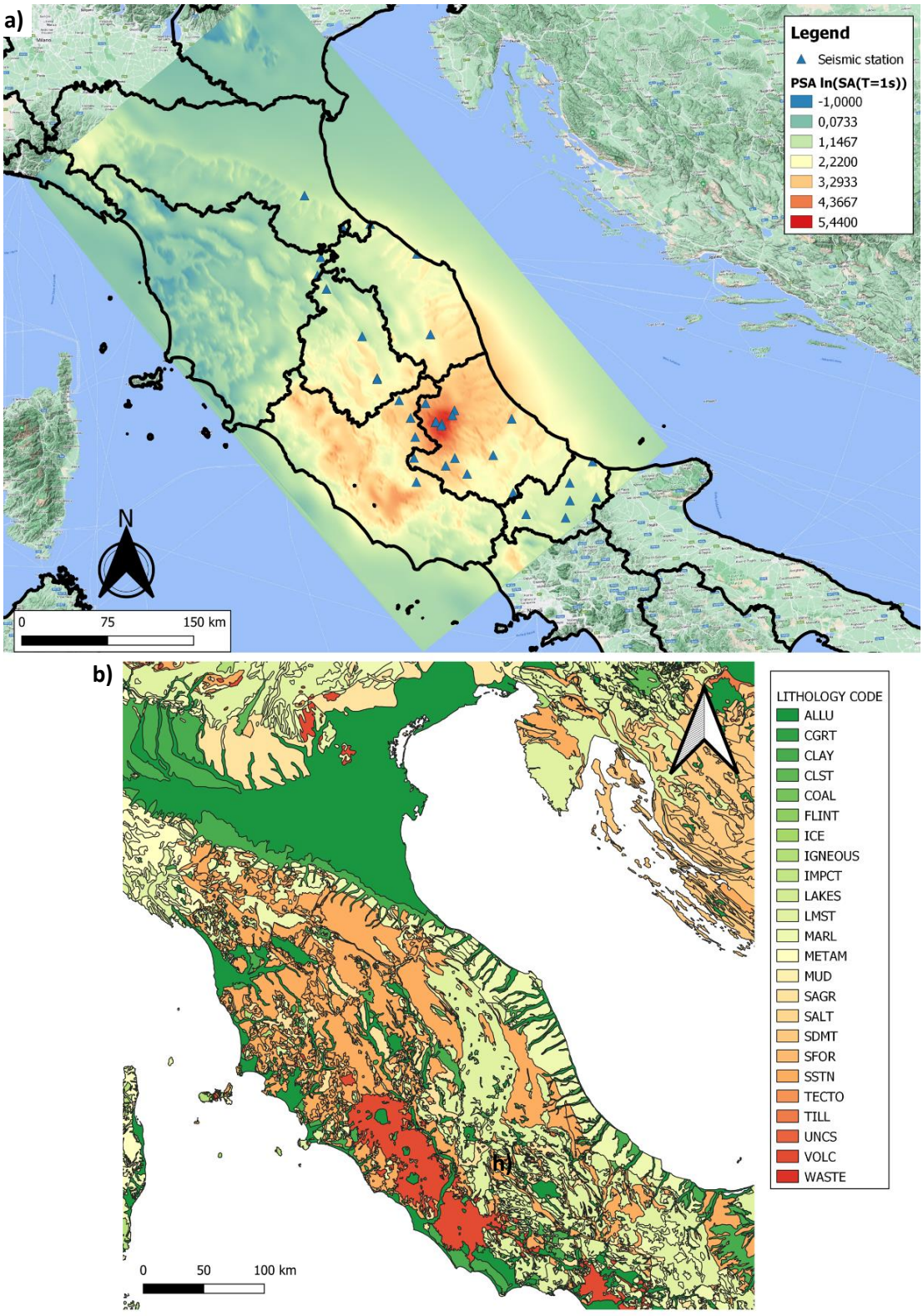

Figure 50. Top: Predicted Sa values obtained from the best-performing U-net (input configuration 7c in table 8). The distribution of seismic stations used as validation points is show by the blue triangles. Results refer to the April 6th 2009 L'Aquila M 6.1 earthquake. Bottom: lithology map for the same area.

We conducted a more in-depth analysis by comparing the mean prediction generated by U-Net with the observed data (Figure 51 a) at seismic station. We see that the uncertainty is quantified around 0.815 (expressed in cm/s^2).

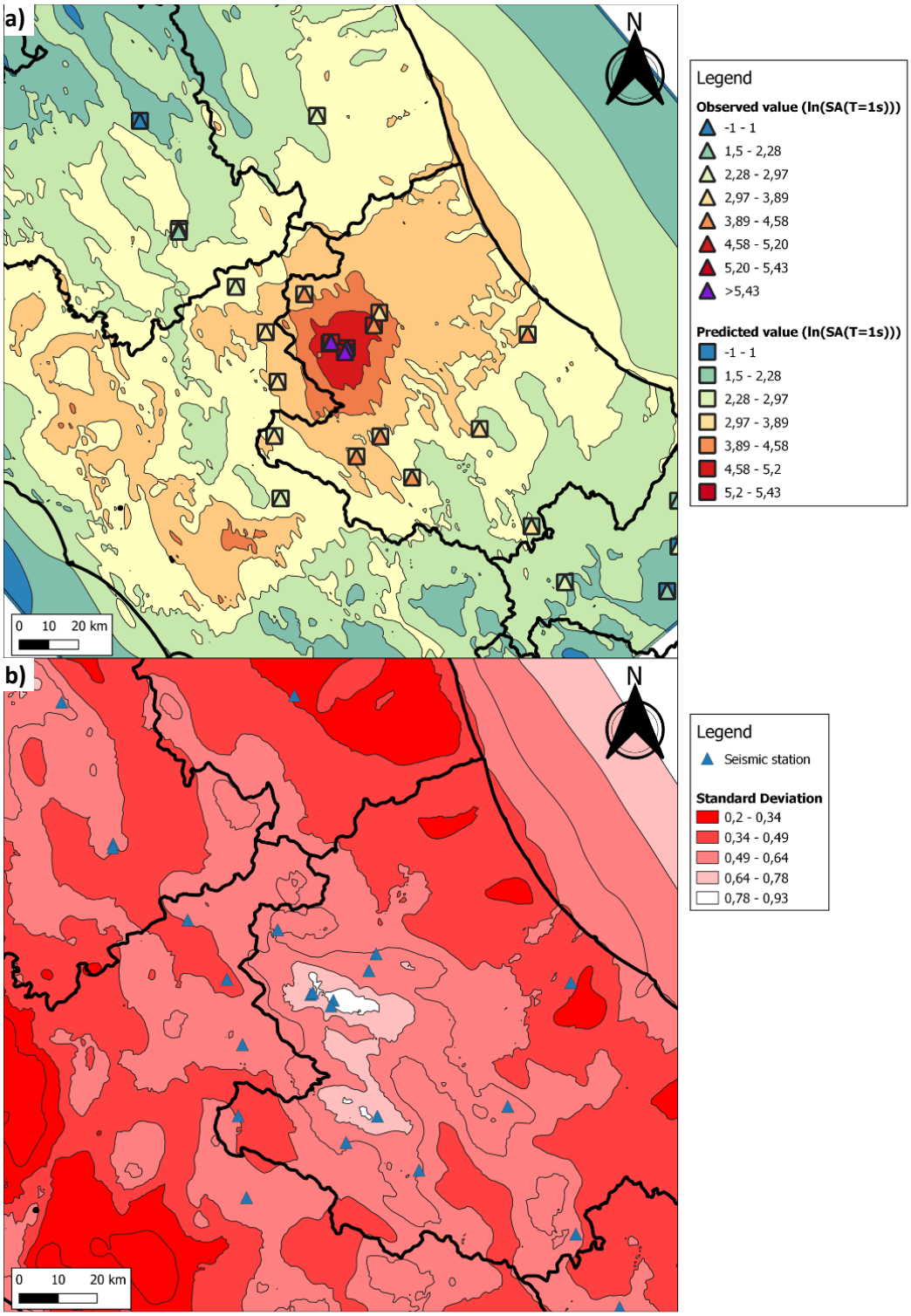

Figure 51. Response spectra values at 1 s predicted by our U-Net for the 6.1 magnitude L'Aquila earthquake (Apr.6th 2009). In (a), the squares on the plotted raster data represent the mean prediction at locations with available observations (triangles). Panel (b) displays the predicted standard deviation. Blue triangles indicate station locations, including those without records for this specific event.

In the analysis of the discrepancy between predicted and observed data, we plotted Sa (T=1s) values against hypocentral distance for the same event (Figure 52). We observe that differences are more pronounced at

short hypocentral distances but significantly reduced at greater distances, indicating increased consistency between the model predictions and observed spectral acceleration values for this validated event.

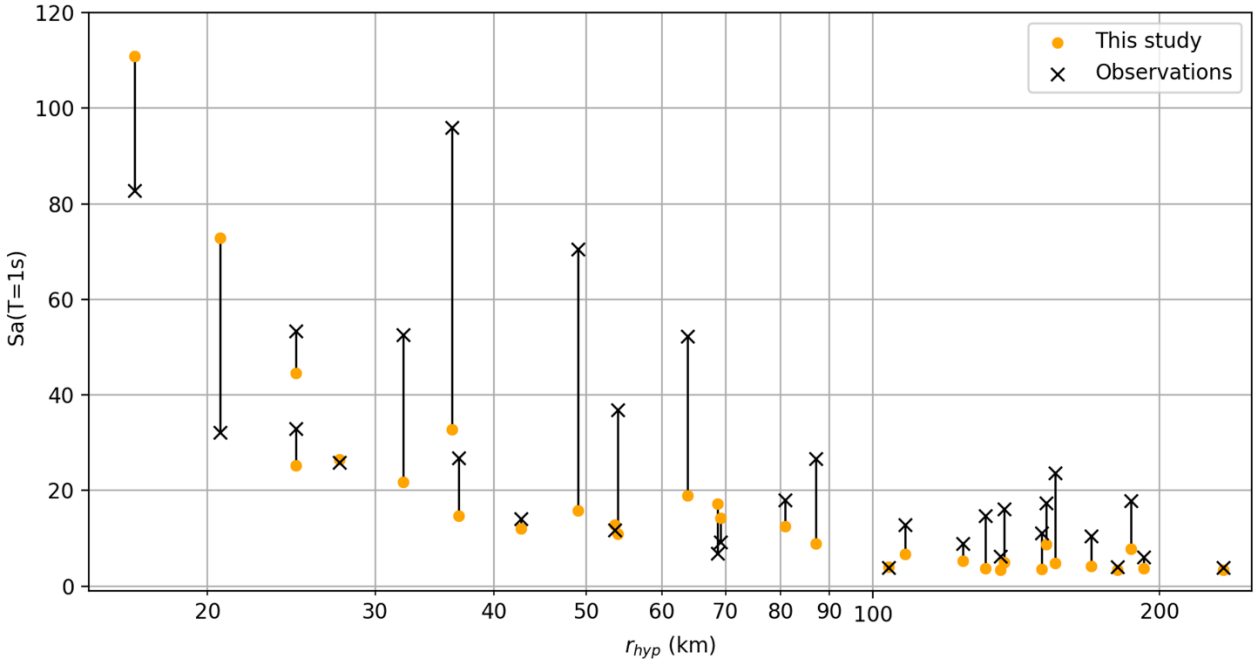

Figure 52. The comparison of our predictions (orange points), expressed in cm/s², with observations (black crosses) from the validation event of L'Aquila. The root mean square error (RMSE) is 0.81 in natural log units.

3.7 Discussion

GMPE are typically obtained by quantifying the coefficients of a given function (ergodic approach) from available data. While using the U-Net approach, on the opposite, we do not need to assume any specific regression model and the procedure is said to be entirely data-driven.

We now discuss these two (ergodic vs data-driven) approaches. The comparison between the U-Net predictions for the example just discussed and the ergodic GMMs developed for Italy (Bindi et al., 2011) and pan – Europe (Kotha et al. 2020, Akkar & Bommer 2010) is given in Figure 53. We want to highlight that these comparisons are not intended to determine the superiority of one model over the other. Rather, our goal is to show that a U-Net approach comes to results very similar to well-established and mature ergodic GMM approaches, basically by doing no assumptions.

In Figure 53, we present a direct comparison between the predicted data for the selected event from the validation dataset (Figure 51) and traditional GMMs. We not only compare the predicted values at site observations (black dots) but also the distribution of all predicted values from the model (gray dots). We note

that the results obtained from the U-NET are all well concentrated within standard deviation of the models by Akkar & Bommer (2010), Bindi et al. (2011) and Kotha et al (2020).

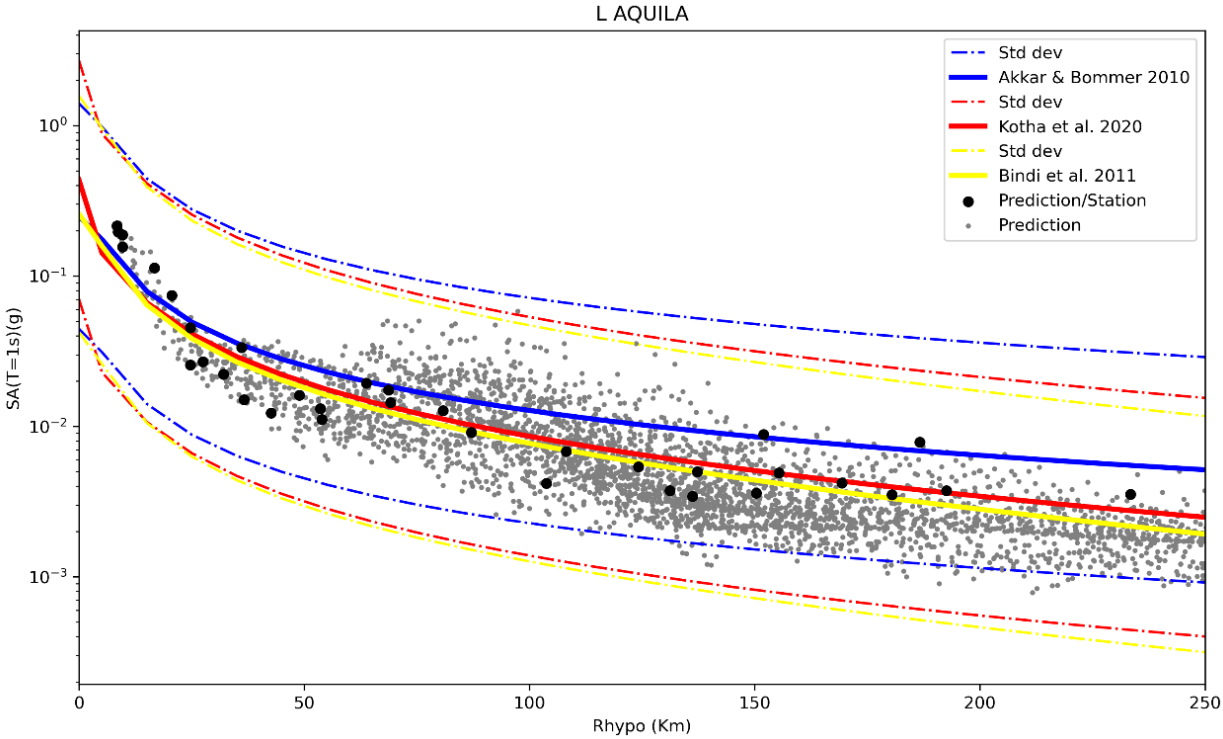

Figure 53. Direct comparison between the predictions of our model (black dots) at observed sites and all predicted values (cloud of gray dots) for the event introduced in Figure 50. Additionally, empirical GMMs are represented with different colours using solid and dashed lines. The solid line represents the mean, and the dashed line represents the standard deviation.

During further analyses, we directly compared our model with the ShakeMap (Oliveti et al., 2022) for the reference event (L'Aquila, 2009, mainshock) provided by the INGV website (Figure 54). The INGV site offers contours for various levels of 1-second Sa expressed as a percentage of gravity (% g). Overlaying these contours onto our map, we observed that the distribution of U-Net predicted Sa values approximately matches with the distribution of values indicated by the ShakeMap, especially in areas with high ground motion intensity.

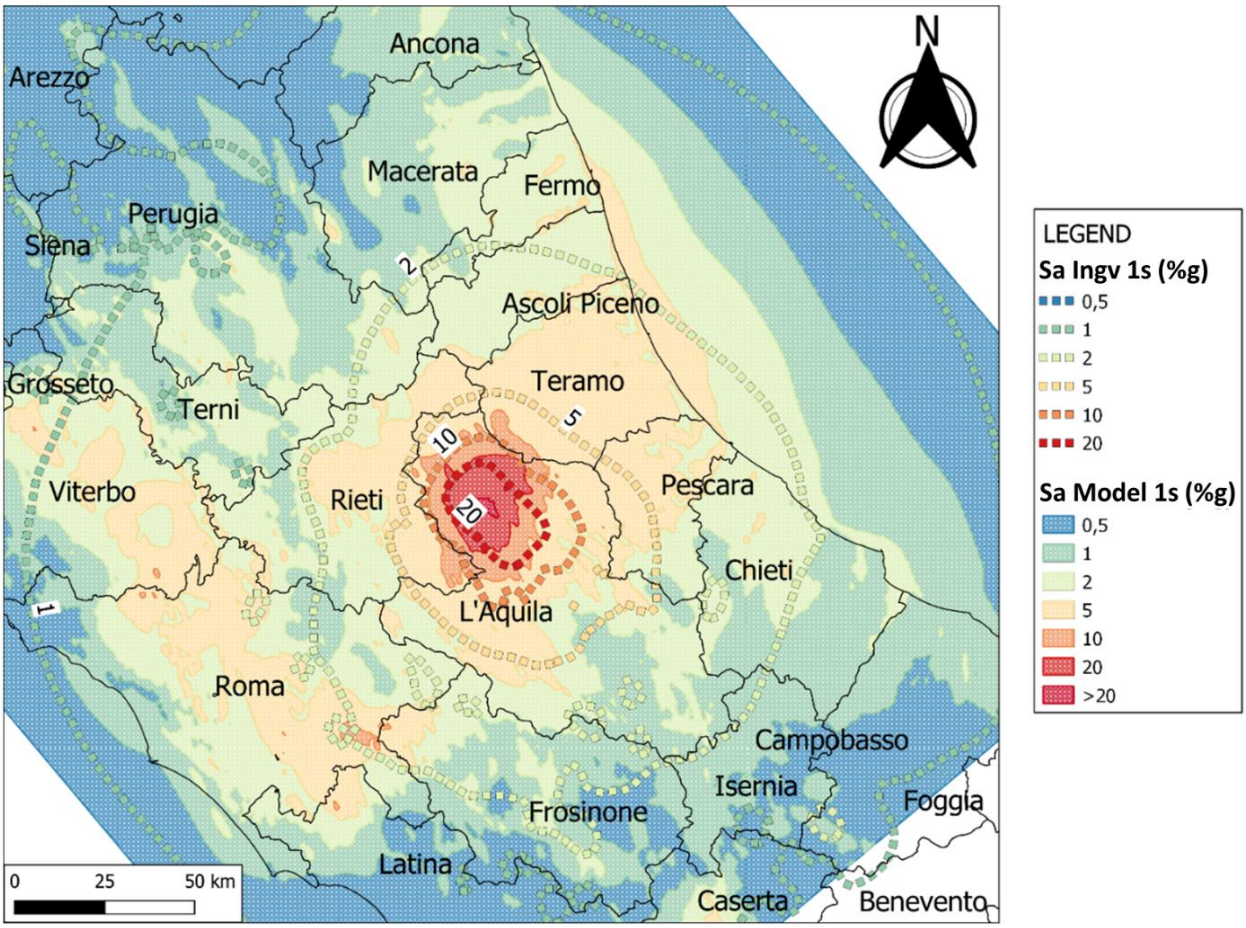

Figure 54. Comparison between our model predictions and the INGV ShakeMap for the reference event. The Sa (T=1s) values are expressed in percentage of gravity (% g).

Until now, we have focused on comparing predictions for single event (L'Aquila, 2009, mainshock) with conventional GMPEs. Subsequently, we grouped all predicted values at site observation within entire database, using magnitude intervals. For instance, we considered events with magnitudes ranging from 4 to 4.4, 4.5 to 4.9, 5 to 5.4, 5.5 to 5.9, and 6 to 6.4. Following that, we calculated the average of these grouped data, also considering hypocentral distance intervals. This process was also applied to all observed values.

The outcome was then compared to ground motion models, as illustrated in Figure 55, showing that the average predictions tend to underestimate the mean observation for each magnitude range. Nevertheless, these predictions reflect the trends of the respective attenuation laws. It is significant (and definitely expected) that disparities between predictions and observations remain substantial at shorter distances, a phenomenon attributed to the limited presence of seismic stations in close proximity to earthquake events and also linked to near-field effects such as directivity, non-planarity of seismic wavefronts, but that can also be attributed to biased recordings at some seismic station as described in Castellaro et al. (2022).

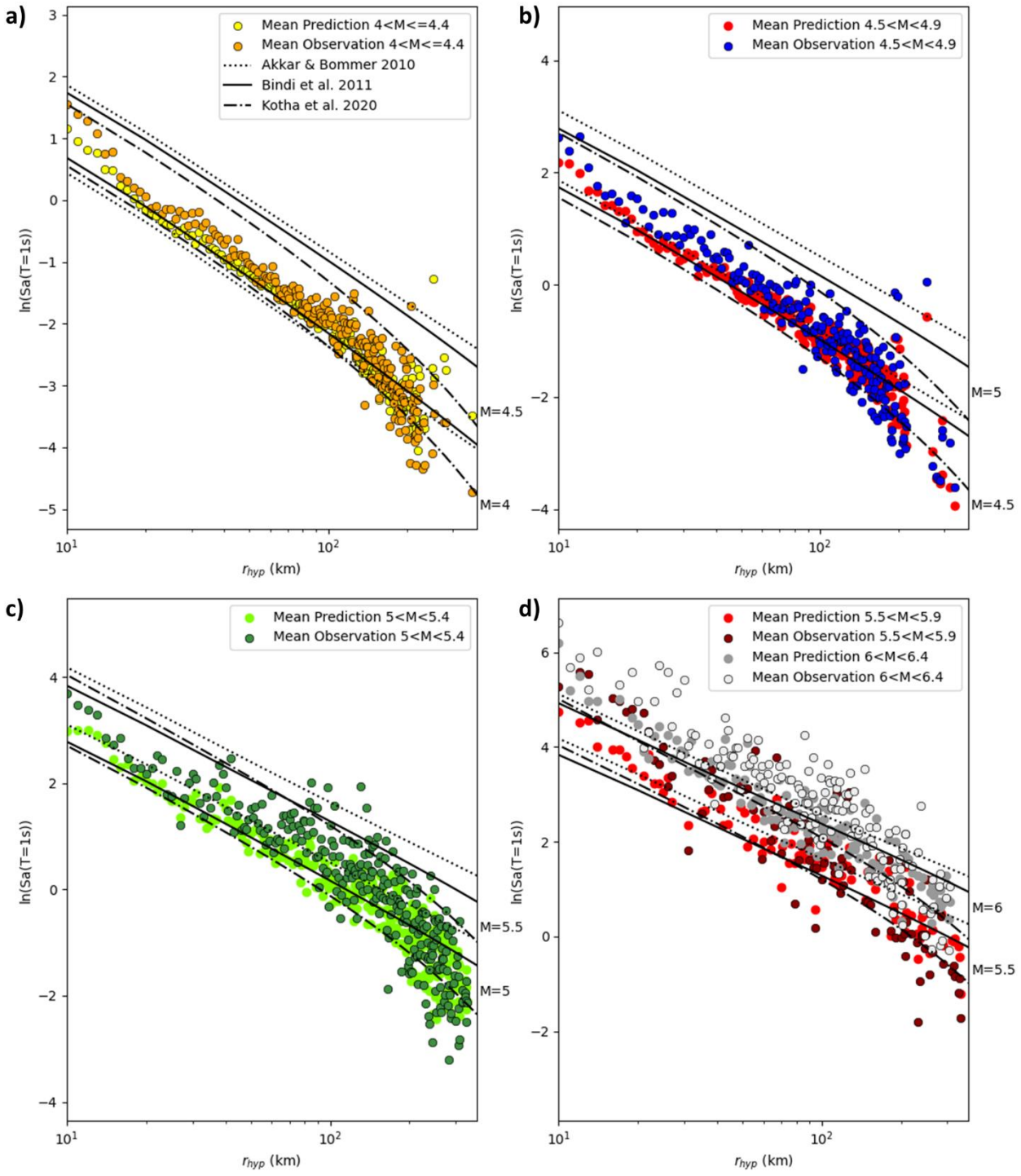

Figure 55. Comparison between the U-Net GMM developed in this study, observation and empirical GMMs (lines), for (a) events between 4 and 4.4 magnitude, (b) from 4.5 to 4.9, (c) from 5 to 5.4 and (d) from 5.4 to 6.3. The colored dots are derived from averaging suitable records and prediction of ln (Sa(T=1s)) along M and rhypo bins.

3.8 Conclusion

Ground Motion Models play an important role in estimating the intensity of ground shaking (e.g., PGA, PGV, Sa) and are widely used in civil and earthquake engineering studies, deterministic or probabilistic seismic hazard analyses, the creation of seismic hazard maps for building codes, and seismic hazard assessments for infrastructure design.

The increasing number of stations and data availability is now facilitating the generation of non-ergodic GMM in order to try to reduce aleatory uncertainties by improving the modelling of relevant terms such as path, site, and source characteristics. Despite developments towards non-ergodic models, aleatory uncertainty for GMM remains stable and large (Strasser et al., 2009). This is because the ground motion intensity depends on many factors such as magnitude, fault style, source-to-site distance, and site effects, with a sensitivity increasing with increasing frequency range of interest. Our simplification of such a complex earthquake phenomenon, coupled with a lack of comprehensive information, prevents the accurate capture of both linear and nonlinear behaviours inherent in earthquake motions. Hence, a possible approach to address this issue is to use alternative machine learning algorithms that offer greater flexibility in capturing the nonlinear and complex patterns within the ground shaking.

In this work, we introduced a fully data-driven based on neural network, unconstrained approach to assess ground motion in Italian context. The approach is based on the training of a U-Net and this methodology stands out for its advanced data-driven spatial interpolation, distinguishing itself from the models currently proposed by Sgobba et al (2019, 2020).

When applying this machine learning technique, the important aspect to consider is the accurate selection of inputs and targets. Therefore, we conducted several tests, starting from the simplest configuration and gradually moving towards the more complex ones, to understand how much these inputs effected the model performance. A similar approach was taken in evaluating the target, showing that for shorter periods, network may require greater detail in the input information.

Finally, the results of this study show that the proposed U-Net-based GMM can capture the behaviour of ground motion using minimum seismological data and basically no assumptions. This is clear from its overlapping performance compared to conventional GMMs. As well as traditional GMMs, the neural network results also show a limited applicability in predicting ground motion near the source. This definitely requires more quality and density in the input data. Nonetheless, this method present opportunities for future research, and we are hopeful that U-Net-based GMMs could become a valuable element in hazard and risk studies, especially in data-rich regions like California and Japan.

Chapter 4. General conclusions

One of the natural phenomena that causes large numbers of casualties and economic losses is earthquake. While from an engineering perspective prevention means designing and building structures capable to withstand seismic motion; from a seismological perspective, the tools available for prevention mainly rely on seismic hazard and risk studies.

In this thesis, we explored the application of machine learning techniques (specifically, neural networks) to improve the assessment of the expected ground motion features at sites of interest. In detail, we focused on two specific subjects within this main topic: the HVSR (Horizontal to Vertical Spectral Ratio) technique and GMM (Ground Motion Modelling).

In the first case, we employed Artificial Neural Networks (ANN) and a pre-trained SqueezeNet to separate stratigraphic peaks from non-stratigraphic ones. This is important to capture the features of H/V curves that are really linked to the soil properties and not to external interference. The results showed that the CNN achieved an accuracy of 93%, surpassing the performance of the ANN, which was approx. 80%. The first approach relies on a much larger number of input compared to the second one and this can be an explanation for the different performance.

In the second case, we used a U-Net neural net capable of generating region-specific GMMs, an approach not previously explored in the Italian context. The results showed that the proposed U-Net-based GMM can capture the behaviour of ground motion using minimal seismological data, performing in the same way compared to conventional GMMs. However, a limitation emerged in its applicability to predict ground motion near the seismic source.

Despite promising results, it is important to carefully consider these limitations to ensure the reliability and practical applicability of the proposed models. Firstly, both nets were trained on imbalanced classes, which could impact predictive ability in real-world scenarios. In the case of HVSR curves, results were obtained on relatively small training databases, with a significant imbalance in the number of examples for the three output categories. It has to be noted, however, that such an imbalance also reflects the distribution of real cases, where non-stratigraphic H/V peaks are luckily not the majority of H/V peaks. Additionally, for the U-Net model, the choice of the number and type of inputs is essential. The limited availability of data representing high magnitudes, greater than 4.5, and the scarcity of seismic stations near events pose further challenges.

4.1 Future Developments of the Two Methods

Regarding the limited availability of data for moderate-high magnitude earthquakes, we recognize that such events are insufficient in our context. In situations where collecting real data is impractical, one can consider using simulated data. Simulating earthquakes with higher magnitudes can provide an artificial dataset, allowing for model training without the need for real seismic recordings. Additionally, using pre-trained networks on larger datasets or from similar contexts can reduce the dependence on a high number of recordings. This, however, may lead to a feedback loop as the use of modelled data, generated through Ground Motion Prediction Equations, would result in the network identifying and reproducing a pattern that is already known.

For HVSR case, increasing the diversity and quantity of data will result in more reliable and generalizable models for HVSR curve classification.

Discussing potential future developments of GMM, they could provide some information on ground motion even in areas currently lacking data and thus be of utility also in areas with non-advanced seismic network distributions. Additionally, generating earthquake-induced shaking maps has various applications on:

- Real-time processing: real-time mapping during an earthquake, continuously updating them with new data allowing for a more accurate and timely assessment of seismic situations.
- Future scenarios: reproducing potential scenarios of future earthquakes.
- Retrospective analysis: improving the comprehension of past seismic effects by generating retrospective shaking maps.
- PSHA Integration: integral to probabilistic seismic hazard analysis, potentially reducing uncertainties within the analysis through comprehensive shaking maps.

In this thesis we explored a few different neural network-based approaches (ANN, CNN, U-Net) to different classification or prediction problems in seismology. It was important to learn the advantages and disadvantages of the different approaches and it constitutes a first basis to build more robust studies in the future and to expand the applications studied so far.

In our opinion, larger space for improvement can be found in the automatic study of HVSR features, where establishing deterministic equations to identify stratigraphic or non-stratigraphic peaks (or similar issues) is more difficult than letting a network find its path the solution of the problem.

Discussing potential future developments, refining techniques related to HVSR curves could complement traditional methods (e.g., SESAME criteria), simplifying peak recognition and supporting operators in data interpretation.

In the case of GMM, on the opposite, we found that NN reached the same conclusions as traditional GMPE and did not provide a real improvement in the knowledge of the phenomenon. The main advantage in this case lies in the fact that NNs do not requires any explicit assumption in the regression functions to use.

Acknowledgments

We warmly thank the two reviewers, Veronica Pazzi and Maria Rosaria Gallipoli, for their careful revision and valuable suggestions. We would like to thank the German Research Centre for Geosciences (GFZ, Potsdam) for the opportunity to conduct our research, especially Prof. Fabrice Cotton and his research group, who allowed us to study and test the U-Net network for generating a non-ergodic and fully data-driven ground motion model (GMM) for the Italian context.

References

Akkar, S., & Bommer, J. J. (2007). Prediction of elastic displacement response spectra in Europe and the Middle East. *Earthquake Engineering & Structural Dynamics*, *36*(10), 1275–1301. https://doi.org/10.1002/eqe.679

Akkar, S., & Bommer, J. J. (2010). Empirical Equations for the Prediction of PGA, PGV, and Spectral Accelerations in Europe, the Mediterranean Region, and the Middle East. *Seismological Research Letters*, *81*(2), 195–206. https://doi.org/10.1785/gssrl.81.2.195

Akkar, S., Sandıkkaya, M. A., Şenyurt, M., Azari Sisi, A., Ay, B. Ö., Traversa, P., Douglas, J., Cotton, F., Luzi, L., Hernandez, B., & Godey, S. (2014). Reference database for seismic ground-motion in Europe (RESORCE). *Bulletin of Earthquake Engineering*, *12*(1), 311–339. https://doi.org/10.1007/s10518-013-9506-8

Al Atik, L. ., Abrahamson, N., Bommer, J. J., Scherbaum, F., Cotton, F., & Kuehn, N. (2010). The Variability of Ground-Motion Prediction Models and Its Components. *Seismological Research Letters*, *81*(5), 794–801. https://doi.org/10.1785/gssrl.81.5.794

Albarello, D., & Castellaro, S. (2011). Tecniche sismiche passive: Indagini a stazione singola. *Supplemento alla rivista Ingegneria Sismica Anno XXVIII*, *2*, 32–49.

Anderson, J. G., & Brune, J. N. (1999). Probabilistic seismic hazard assessment without the ergodic assumption. *Seismol. Res. Lett*, *70*, 19–28.

Baltay, A. S., Hanks, T. C., & Abrahamson, N. A. (2017). Uncertainty, Variability, and Earthquake Physics in Ground-Motion Prediction Equations. *Bulletin of the Seismological Society of America*, ssabull;0120160164v1. https://doi.org/10.1785/0120160164

Bard, P.-Y., & Bouchon, M. (1980a). The seismic responce of sediment-filled valleys. Part 1. The case of incident SH waves. *Bulletin of the Seismological Society of America*, *70*(4), 1263–1286.

Bergamo, P., Hammer, C., & Fäh, D. (2021). On the Relation between Empirical Amplification and Proxies Measured at Swiss and Japanese Stations: Systematic Regression Analysis and Neural Network Prediction of Amplification. *Bulletin of the Seismological Society of America*, *111*(1), 101–120. https://doi.org/10.1785/0120200228

Bindi, D., Kotha, S.-R., Weatherill, G., Lanzano, G., Luzi, L., & Cotton, F. (2018). The pan-European engineering strong motion (ESM) flatfile: Consistency check via residual analysis. *Bulletin of Earthquake Engineering*, *17*(2), 583–602. https://doi.org/10.1007/s10518-018-0466-x

Bindi, D., Luzi, L., Massa, M., & Pacor, F. (2010). Horizontal and vertical ground motion prediction equations derived from the Italian Accelerometric Archive (ITACA). *Bulletin of Earthquake Engineering*, *8*(5), 1209–1230. https://doi.org/10.1007/s10518-009-9130-9

Bindi, D., Pacor, F., Luzi, L., Puglia, R., Massa, M., Ameri, G., & Paolucci, R. (2011). Ground motion prediction equations derived from the Italian strong motion database. *Bulletin of Earthquake Engineering*, *9*(6), 1899–1920. https://doi.org/10.1007/s10518-011-9313-z

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bokelmann, G. H. R., & Baisch, S. (1999). Nature of narrow-band signals at 2.083 Hz. *Bulletin of the Seismological Society of America*, *89*(1), 156–164. https://doi.org/10.1785/BSSA0890010156

Bommer, J. J. (2022). Earthquake hazard and risk analysis for natural and induced seismicity: Towards objective assessments in the face of uncertainty. *Bulletin of Earthquake Engineering*, *20*(6), 2825–3069. https://doi.org/10.1007/s10518-022-01357-4

Bommer, J. J., & Abrahamson, N. A. (2006). Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates? *Bulletin of the Seismological Society of America*, *96*(6), 1.967-1.977.

Bonnefoy-Claudet, S., Kohler, A., Cornou, C., Wathelet, M., & Bard, P.-Y. (2008). Effects of Love Waves on Microtremor H/V Ratio. *Bulletin of the Seismological Society of America*, *98*(1), 288–300. https://doi.org/10.1785/0120070063

Boudghene Stambouli, A., Zendagui, D., Bard, P.-Y., & Derras, B. (2017). Deriving amplification factors from simple site parameters using generalized regression neural networks: Implications for relevant site proxies. *Earth, Planets and Space*, *69*(1), 99. https://doi.org/10.1186/s40623-017-0686-3

Brunelli, G., Lanzano, G., D'Amico, M. C., Felicetta, C., Luzi, L., Mascandola, C., Pacor, F., Russo, E., & Sgobba, S. (2022). *ITACAext flatfile: Parametric table of metadata and strong motion intensity measures (Version 1.0) [Data set]. Istituto Nazionale di Geofisica e Vulcanologia (INGV).ITACAext flatfile [Data set]. Istituto Nazionale di Geofisica e Vulcanologia (INGV).* https://doi.org/DOI: 10.13127/itaca32/itacaext_flatfile.1.0

Castellaro, S. (2016). The complementarity of H/V and dispersion curves. *GEOPHYSICS*, *81*(6), T323–T338. https://doi.org/10.1190/geo2015-0399.1

Castellaro, S., Alessandrini, G., & Musinu, G. (2022). Seismic Station Installations and Their Impact on the Recorded Signals and Derived Quantities. *Seismological Research Letters*, *93*(6), 3348–3362. https://doi.org/10.1785/0220220029

Castellaro, S., & Mulargia, F. (2009). VS30 Estimates Using Constrained H/V Measurements. *Bulletin of the Seismological Society of America*, *99*(2A), 761–773. https://doi.org/10.1785/0120080179

Castellaro, S., & Mulargia, F. (2010). How Far from a Building Does the Ground-Motion Free-Field Start? The Cases of Three Famous Towers and a Modern Building. *Bulletin of the Seismological Society of America*, *100*(5A), 2080–2094. https://doi.org/10.1785/0120090188

Castellaro, S., & Mulargia, F. (2009b). The Effect of Velocity Inversions on H/V. *Pure and Applied Geophysics*, *166*(4), 567–592. https://doi.org/10.1007/s00024-009-0474-5

Castellaro, S., & Mulargia, F. (2009a). VS30 Estimates Using Constrained H/V Measurements. *Bulletin of the Seismological Society of America*, *99*(2A), 761–773. https://doi.org/10.1785/0120080179

Castellaro, S., Mulargia, F., & Kagan, Y. Y. (2006). Regression problems for magnitudes. *Geophysical Journal International*, *165*(3), 913–930. https://doi.org/10.1111/j.1365-246X.2006.02955.x

Castellaro, S., & Musinu, G. (2023). Resonance versus Shape of Sedimentary Basins. *Bulletin of the Seismological Society of America*, *113*(2), 745–761. https://doi.org/10.1785/0120210277

Cauzzi, C., Faccioli, E., Vanini, M., & Bianchini, A. (2015). Updated predictive equations for broadband (0.01–10 s) horizontal response spectra and peak ground motions, based on a global dataset of digital acceleration records. *Bulletin of Earthquake Engineering*, *13*(6), 1587–1612. https://doi.org/10.1007/s10518-014-9685-y

Cheng, T., Cox, B. R., Vantassel, J. P., & Manuel, L. (2020). A statistical approach to account for azimuthal variability in single-station HVSR measurements. *Geophysical Journal International*, *223*(2), 1040–1053. https://doi.org/10.1093/gji/ggaa342

Chollet, F. (2017). *Xception: Deep Learning with Depthwise Separable Convolutions* (arXiv:1610.02357). arXiv. http://arxiv.org/abs/1610.02357

Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the seismological society of America*, *58*(5), 583–1606.

Cornou, C., Gu﷿guen, P., Bard, P.-Y., & Haghshenas, E. (2004). Ambient noise energy bursts observation and modeling: Trapping of harmonic structure-soil induced?waves in a topmost sedimentary layer. *Journal of Seismology*, *8*(4), 507–524. https://doi.org/10.1007/s10950-004-1980-7

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

Derras, B., Bard, P.-Y., Cotton, F., & Bekkouche, A. (2012). Adapting the Neural Network Approach to PGA Prediction: An Example Based on the KiK-net Data. *Bulletin of the Seismological Society of America*, *102*(4), 1446–1461. https://doi.org/10.1785/0120110088

Dhanya, J., & Raghukanth, S. T. G. (2018). Ground Motion Prediction Model Using Artificial Neural Network. *Pure and Applied Geophysics*, *175*(3), 1035–1064. https://doi.org/10.1007/s00024-017-1751-3

Díaz, J. P., Sáez, E., Monsalve, M., Candia, G., Aron, F., & González, G. (2022). Machine learning techniques for estimating seismic site amplification in the Santiago basin, Chile. *Engineering Geology*, *306*, 106764. https://doi.org/10.1016/j.enggeo.2022.106764

Douglas, J. (2003). Earthquake ground motion estimation using strongmotion records: A review of equations for the estimation of peak ground acceleration and response spectral ordinates. *Earth Sci. Rev.*, *61*, 43.

Douglas, J. (2022). *Ground motion prediction equations*.

Dowla, F. U., Taylor, S. R., & Anderson, R. W. (1990). Seismic discimination ith artificial neural networks:preliminary results ith regional spectral data. *Bulletin of the Seismological Society of America*, *80*(5), 1346–1373.

Dreossi, I.,Barnaba C., Castellaro S. (2018). Book of Abstracts: 36th General Assembly of the European Seismological Commission, 2-7 September 2018, Valletta- Malta. (2018). Mistral Service.

Eliasy, A., & Przychodzen, J. (2020). The role of AI in capital structure to enhance corporate funding strategies. *Array*, *6*, 100017. https://doi.org/10.1016/j.array.2020.100017

Endrun, B. (2011). Love wave contribution to the ambient vibration H/V amplitude peak observed with array measurements. *Journal of Seismology*, *15*(3), 443–472. https://doi.org/10.1007/s10950-010-9191-x

ENEA. (2015). *Avezzano 1915-2015: Cento anni di ingegneria sismica*.

Fäh, D., Kind, F., & Giardini, D. (2001). A theoretical investigation of average *H/V* ratios. *Geophysical Journal International*, *145*(2), 535–549. https://doi.org/10.1046/j.0956-540x.2001.01406.x

Fang, W., Love, P. E. D., Luo, H., & Ding, L. (2020). Computer vision for behaviour-based safety in construction: A review and future directions. *Advanced Engineering Informatics*, *43*, 100980. https://doi.org/10.1016/j.aei.2019.100980

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/BF00344251

Ghasemi, H., Zare, M., Fukushima, Y., & Sinaeian, F. (2009). Applying empirical methods in site classification, using response spectral ratio (H/V): A case study on Iranian strong motion network (ISMN). *Soil Dynamics and Earthquake Engineering*, *29*(1), 121–132. https://doi.org/10.1016/j.soildyn.2008.01.007

Gosar, A., & Lenart, A. (2010). Mapping the thickness of sediments in the Ljubljana Moor basin (Slovenia) using microtremors. *Bulletin of Earthquake Engineering*, *8*(3), 501–518. https://doi.org/10.1007/s10518-009-9115-8

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, L., Wang, G., Cai, J., & Chen, T. (2017). *Recent Advances in Convolutional Neural Networks* (arXiv:1512.07108). arXiv. http://arxiv.org/abs/1512.07108

Gutenberg, B. (1911). *Die seismische Bodenunruhe* [Ph.D. thesis]. University of Gottigen,Germany.

Gutenberg, B. (1931). Microseisms in North America. *Bull. Seismol. Soc. Am.*, *21*, 1–24.

Gutenberg, B. (1936). On microseisms. *Bull. Seismol. Soc. Am.*, *26*, 111–117.

Gutenberg, B. (1958). Microseisms. *Advances in Geophysics*, *5*, 53–92.

Gutenberg, B., & Richter, C. F. (2019b). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, *34*(4), 185–188.

Hagan, M. T., & Menhaj, M. (1994). Training feed-forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, *5*(6), 989–993.

Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). GhostNet: More Features From Cheap Operations. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1577–1586. https://doi.org/10.1109/CVPR42600.2020.00165

Haubrich, A. R., Munk, W. H., & Snodgrass, F. E. (1963). Comparative spectra of microseisms and swell. *Bull. Seism. Soc. Am.*, *53*, 27–33.

Hawkins, D. M. (2004). The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*, *44*, 1–12.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., & Le, Q. (2019). Searching for MobileNetV3. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 1314–1324. https://doi.org/10.1109/ICCV.2019.00140

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269. https://doi.org/10.1109/CVPR.2017.243

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). *SQUEEZENET: ALEXNET-LEVEL ACCURACY WITH 50X FEWER PARAMETERS AND <0.5MB MODEL SIZE*. 14.

Ibs-von Seht, M., & Wohlenberg, . (1999). Microtremor measurements used to map thickness of soft sediments. *Bulletin of the Seismological Society of America*, *89*, 250–279.

Ioffe, S., & Szegedy, C. (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift* (arXiv:1502.03167). arXiv. http://arxiv.org/abs/1502.03167

Ji, K., Ren, Y., Wen, R., Zhu, C., Liu, Y., & Zhou, B. (2022). HVSR-based Site Classification Approach Using General Regression Neural Network (GRNN): Case Study for China Strong Motion Stations. *Journal of Earthquake Engineering*, *26*(16), 8423–8445. https://doi.org/10.1080/13632469.2021.1991520

Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization* (arXiv:1412.6980). arXiv. http://arxiv.org/abs/1412.6980

Konno, K., & Ohmachi, T. (1998). Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. *Bulletin of the Seismological Society of America*, *88*(1), 228–241. https://doi.org/10.1785/BSSA0880010228

Kotha, S. R., Weatherill, G., Bindi, D., & Cotton, F. (2020). A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe. *Bulletin of Earthquake Engineering*, *18*(9), 4091–4125. https://doi.org/10.1007/s10518-020-00869-1

Kramer, S. L. (1996). Geotechnical earthquake engineering. *Pearson Education India*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kuehn, N. M., & Abrahamson, N. A. (2020). Spatial correlations of ground motion for non-ergodic seismic hazard analysis. *Earthq. Eng. Struct. Dyn.*, *49*, 4–23.

Lanzano, G., Luzi, L., Pacor, F., Felicetta, C., Puglia, R., Sgobba, S., & D'Amico, M. C. (2019b). A Revised Ground-Motion Prediction Model for Shallow Crustal Earthquakes in Italy. *Bulletin of the Seismological Society of America*, *109*(2), 525–540.

Lanzano, G., Pacor, F., Luzi, L., D'Amico, M., Puglia, R., & Felicetta, C. (2017). Systematic source, path and site effects on ground motion variability: The case study of Northern Italy. *Bulletin of Earthquake Engineering*, *15*(11), 4563–4583. https://doi.org/10.1007/s10518-017-0170-2

Lanzano, G., Sgobba, S., Luzi, L., Puglia, R., Pacor, F., Felicetta, C., D'Amico, M., Cotton, F., & Bindi, D. (2019a). The pan-European Engineering Strong Motion (ESM) flatfile: Compilation criteria and data statistics. *Bulletin of Earthquake Engineering*, *17*(2), 561–582. https://doi.org/10.1007/s10518-018-0480-z

LeCurn, Y., Bottou, L., Orr, G. B., & Müller, K. R. (1998). Efficient BackProp. *In: Orr, G.B., Müller, KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*, *1524*(Springer, Berlin, Heidelberg).

Li, G., Zhang, M., Li, J., Lv, F., & Tong, G. (2021). Efficient densely connected convolutional neural networks. *Pattern Recognition*, *109*, 107610. https://doi.org/10.1016/j.patcog.2020.107610

Li, H.-C., Deng, Z.-Y., & Chiang, H.-H. (2020). Lightweight and Resource-Constrained Learning Network for Face Recognition with Performance Optimization. *Sensors*, *20*(21), 6114. https://doi.org/10.3390/s20216114

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2022). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(12), 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827

Lilienkamp, H., Von Specht, S., Weatherill, G., Caire, G., & Cotton, F. (2022). Ground-Motion Modeling as an Image Processing Task: Introducing a Neural Network Based, Fully Data-Driven, and Nonergodic Approach. *Bulletin of the Seismological Society of America*, *112*(3), 1565–1582. https://doi.org/10.1785/0120220008

Lin, P.-S., Chiou, B., Abrahamson, N., Walling, M., Lee, C.-T., & Cheng, C.-T. (2011). Repeatable Source, Site, and Path Effects on the Standard Deviation for Empirical Ground-Motion Prediction Models. *Bulletin of the Seismological Society of America*, *101*(5), 2281–2295. https://doi.org/10.1785/0120090312

Longuet-Higgins, M. D. (1950). A theory on the origin of microseisms. *Philos. Trans. R. Soc. Lond.*, *243*, 1–35.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proceedings of the 30 Th International Conference on Machine Learning*, *28*.

Mantovani, A., Valkaniotis, S., Rapti, D., & Caputo, R. (2018). Mapping the Palaeo-Piniada Valley, Central Greece, Based on Systematic Microtremor Analyses. *Pure and Applied Geophysics*, *175*(3), 865–881. https://doi.org/10.1007/s00024-017-1731-7

Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal on Applied Mathematics*, *11*, 431–441.

Martino, S., Caprara, P., Fiorucci, M., & Marmoni, G. M. (2020). Il Catalogo CEDIT: dall'inventario degli effetti sismoindotti all'analisi di scenario. *Mem. Descr. Carta Geol. d'It.*, *107*, 441–450.

Masi, A., & Nicodemo, G. (2022). *Reconstruction, recovery and socio-economic development of the Basilicata region, southern Italy: Lessons and experience after the 1980 earthquake*. https://doi.org/10.4430/bgo00405

Matsushima, S., Hirokawa, T., De Martin, F., Kawase, H., & Sanchez-Sesma, F. J. (2014). The Effect of Lateral Heterogeneity on Horizontal-to-Vertical Spectral Ratio of Microtremors Inferred from Observation and Synthetics. *Bulletin of the Seismological Society of America*, *104*(1), 381–393. https://doi.org/10.1785/0120120321

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill International Editions Computer Science Series.

Mohammadi, A., Karimzadeh, S., Banimahd, S. A., Ozsarac, V., & Lourenço, P. B. (2023). The potential of region-specific machine-learning-based ground motion models: Application to Turkey. *Soil Dynamics and Earthquake Engineering*, *172*, 108008. https://doi.org/10.1016/j.soildyn.2023.108008

Molnar, S., Cassidy, J. F., Castellaro, S., Cornou, C., Crow, H., Hunter, J. A., Matsushima, S., Sánchez-Sesma, F. J., & Yong, A. (2018). Application of Microtremor Horizontal-to-Vertical Spectral Ratio (MHVSR) Analysis for Site Characterization: State of the Art. *Surveys in Geophysics*, *39*(4), 613–631. https://doi.org/10.1007/s10712-018-9464-4

Molnar, S., Sirohey, A., Assaf, J., Bard, P.-Y., Castellaro, S., Cornou, C., Cox, B., Guillier, B., Hassani, B., Kawase, H., Matsushima, S., Sánchez-Sesma, F. J., & Yong, A. (2022). A review of the microtremor horizontal-to-vertical spectral ratio (MHVSR) method. *Journal of Seismology*, *26*(4), 653–685. https://doi.org/10.1007/s10950-021-10062-9

Morasca, P., D'Amico, M., Sgobba, S., Lanzano, G., Colavitti, L., Pacor, F., & Spallarossa, D. (2022). Empirical correlations between an FAS non-ergodic ground motion model and a

GIT derived model for Central Italy. *Geophysical Journal International*, *233*(1), 51–68. https://doi.org/10.1093/gji/ggac445

Mousavi, S. M., & Beroza, G. C. (2023). Machine Learning in Earthquake Seismology. *Annual Review of Earth and Planetary Sciences*, *51*(1), 105–129. https://doi.org/10.1146/annurev-earth-071822-100323

Nair, V., & Hinton, G. E. (2010). *Rectified Linear Units Improve Restricted Boltzmann Machines*.

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. *Quarterly Report of Railway Technical Research*, *3*(1), 25–33.

Nogoshi, M., & Igarashi, T. (1970). On the propagation characteristics of microtremors. *J. Seism. Soc. Japan*, *23*, 264–20.

Oliveti, I., Faenza, L., & Michelini, A. (2022). New reversible relationships between ground motion parameters and macroseismic intensity for Italy and their application in ShakeMap. *Geophysical Journal International*, *231*(2), 1117–1137. https://doi.org/10.1093/gji/ggac245

Oubaiche, E. H., Chatelain, J., Hellel, M., Wathelet, M., Machane, D., Bensalem, R., & Bouguern, A. (2016). The Relationship between Ambient Vibration H/V and SH Transfer Function: Some Experimental Results. *Seismological Research Letters*, *87*(5), 1112–1119. https://doi.org/10.1785/0220160113

Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi, M., Panzeri, L., Simionato, M., & Vigano, D. (2014). OpenQuake Engine: An Open Hazard (and Risk) Software for the Global Earthquake Model. *Seismological Research Letters*, *85*(3), 692–702. https://doi.org/10.1785/0220130087

Palaz, D., Magimai-Doss, M., & Collobert, R. (2019). End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition. *Speech Communication*, *108*, 15–32. https://doi.org/10.1016/j.specom.2019.01.004

Parolai, S. (2002). New Relationships between Vs, Thickness of Sediments, and Resonance Frequency Calculated by the H/V Ratio of Seismic Noise for the Cologne Area (Germany). *Bulletin of the Seismological Society of America*, *92*(6), 2521–2527. https://doi.org/10.1785/0120010248

Pozos-Estrada, A., Gómez, R., & Hong, H. P. (2014). Use of neural network to predict the peak ground accelerations and pseudo spectral accelerations for Mexican Inslab and Interplate Earthquake. *Geofis. Int*, *53*, 39–57.

Pratt, L. (1993). Discriminability-based transfer between neural networks. *NIPS Conference: Advances in Neural Information Processing Systems*, 204–211.

Raschka, S., & Mirjalili, V. (2019). *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2* (3rd ed.). Packt Publishing.

Restrepo-Vélez, L. F., & Bommer, J. J. (2003). An exploration of the nature of the scatter in ground-motion prediction equations and the implications for seismic hazard assessment. *Journal of Earthquake Engineering*, *7*(S1), 171–199.

Rodriguez-Marek, A., Cotton, F., Abrahamson, N. A., Akkar, S., Al Atik, L., Edwards, B., Montalva, G. A., & Dawood, H. M. (2013). A Model for Single-Station Standard Deviation Using Data from Various Tectonic Regions. *Bulletin of the Seismological Society of America*, *103*(6), 3149–3163. https://doi.org/10.1785/0120130030

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation* (arXiv:1505.04597). arXiv. http://arxiv.org/abs/1505.04597

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65(6)*, 386–408.

Rumelhart, D. E., Hintont, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors*.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sabetta, F., & Pugliese, A. (1996). Estimation of response spectra and simulation of nonstationary earthquake ground motions. *Bulletin of the Seismological Society of America*, *86*(2), 337–352.

Salameh, C., Bard, P.-Y., Guillier, B., Harb, J., Cornou, C., Gérard, J., & Almakari, M. (2017). Using ambient vibration measurements for risk assessment at an urban scale: From numerical proof of concept to Beirut case study (Lebanon). *Earth, Planets and Space*, *69*(1), 60. https://doi.org/10.1186/s40623-017-0641-3

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2019). *MobileNetV2: Inverted Residuals and Linear Bottlenecks* (arXiv:1801.04381). arXiv. http://arxiv.org/abs/1801.04381

Scheib, A., Morris, P., Murdie, R., & Delle Piane, P. (2016). A passive seismic approach to estimating the thickness of sedimentary cover on the Nullarbor Plain, Western Australia. *Australian Journal of Earth Sciences*, *63*(5), 583–598. https://doi.org/10.1080/08120099.2016.1233455

Sgattoni, G., & Castellaro, S. (2020). Detecting 1-D and 2-D ground resonances with a single-station approach. *Geophysical Journal International*, *223*(1), 471–487. https://doi.org/10.1093/gji/ggaa325

Sgattoni, G., Lattanzi, G., & Castellaro, S. (2023). An experimental approach to unravel 2D ground resonances: Application to an alluvial-sedimentary basin. *Earth, Planets and Space*, *75*(1), 74. https://doi.org/10.1186/s40623-023-01825-4

Sgobba, S., Lanzano, G., & Pacor, F. (2021). Empirical nonergodic shaking scenarios based on spatial correlation models: An application to central Italy. *Earthquake Engineering & Structural Dynamics*, *50*(1), 60–80. https://doi.org/10.1002/eqe.3362

Sgobba, S., Lanzano, G., Pacor, F., Puglia, R., D'Amico, M., Felicetta, C., & Luzi, L. (2019). Spatial Correlation Model of Systematic Site and Path Effects for Ground-Motion Fields in Northern Italy. *Bulletin of the Seismological Society of America*, *109*(4), 1419–1434. https://doi.org/10.1785/0120180209

SIGEA. (2018). *Geologia dell'Ambiente: Rischio sismico in Italia: Analisi e prospettive per una prevenzione efficace in un Paese fragile*.

Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv:1409.1556). arXiv. http://arxiv.org/abs/1409.1556

Strasser, F. O., Abrahamson, N. A., & Bommer, J. J. (2009). Sigma: Issues, Insights, and Challenges. *Seismological Research Letters*, *80*(1), 40–56. https://doi.org/10.1785/gssrl.80.1.40

Sumari, P., Saqib Jamal, S., & Abualigah, L. (2021). A Novel Deep Learning Pipeline Architecture based on CNN to Detect Covid-19 in Chest X-ray Images. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(6), 2001–2011. https://doi.org/10.17762/turcomat.v12i6.4804

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). *Going Deeper with Convolutions* (arXiv:1409.4842). arXiv. http://arxiv.org/abs/1409.4842

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). *Rethinking the Inception Architecture for Computer Vision* (arXiv:1512.00567). arXiv. http://arxiv.org/abs/1512.00567

Tabian, I., Fu, H., & Khodaei, Z. S. (2019). A Convolutional Neural Network for Impact Detection and Characterization of Complex Composite Structures. *Sensors*, *19*(22), 4933. https://doi.org/10.3390/s19224933

Tarquini, S., Isola, I., Favalli, M., Battistini, A., & Dotta, G. (2023). TINITALY, a digital elevation model of Italy with a 10 meters cell size (Version 1.1). *Istituto Nazionale di Geofisica e Vulcanologia (INGV).*

Taye, M. M. (2023). Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions. *Computation*, *11*(3), 52. https://doi.org/10.3390/computation11030052

Tuan, T. T., Scherbaum, F., & Malischewsky, P. G. (2011). On the relationship of peaks and troughs of the ellipticity (H/V) of Rayleigh waves and the transmission response of single layer over half-space models: Relationship of peaks and troughs of H/V-ratio. *Geophysical Journal International*, *184*(2), 793–800. https://doi.org/10.1111/j.1365-246X.2010.04863.x

Turhan Taner, M., Lu, L., & Baysal, E. (1988). Unified method for 2-D and 3-D refraction statics with first break picking by supervised learning. *SEG Tech. Program Expand.*, *Abstr. 1988:172–74*.

Van Der Baan, M. (2009). The origin of *SH* -wave resonance frequencies in sedimentary layers. *Geophysical Journal International*, *178*(3), 1587–1596. https://doi.org/10.1111/j.1365-246X.2009.04245.x

Vantassel, J. P., Stolte, A. C., Wotherspoon, L. M., & Cox, B. R. (2023). AutoHVSR: A machine-learning-supported algorithm for the fully-automated processing of horizontal-to-vertical

spectral ratio measurements. *Soil Dynamics and Earthquake Engineering*, *173*, 108153. https://doi.org/10.1016/j.soildyn.2023.108153

Villani, M., & Abrahamson, N. A. (2015). Repeatable Site and Path Effects on the Ground-Motion Sigma Based on Empirical Data from Southern California and Simulated Waveforms from the CyberShake Platform. *Bulletin of the Seismological Society of America*, *105*(5), 2681–2695. https://doi.org/10.1785/0120140359

Wang, P., Zimmaro, P., Ahdi, S. K., Yong, A., & Stewart, J. P. (2023). Identification Protocols for Horizontal-to-Vertical Spectral Ratio Peaks. *Bulletin of the Seismological Society of America*, *113*(2), 782–803. https://doi.org/10.1785/0120210304

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). *Aggregated Residual Transformations for Deep Neural Networks* (arXiv:1611.05431). arXiv. http://arxiv.org/abs/1611.05431

Yaghmaei-Sabegh, S., & Tsang, H.-H. (2011). A new site classification approach based on neural networks. *Soil Dynamics and Earthquake Engineering*, *31*(7), 974–981. https://doi.org/10.1016/j.soildyn.2011.03.004

Yang, J., & Yang, G. (2018). Modified Convolutional Neural Network Based on Dropout and the Stochastic Gradient Descent Optimizer. *Algorithms*, *11*(3), 28. https://doi.org/10.3390/a11030028

Yani, M., Budhi Irawan, S., Si. ,. M. T., & Casi Setiningsih, S. T., M. T. (2019). Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry's Nail. *Journal of Physics: Conference Series*, *1201*(1), 012052. https://doi.org/10.1088/1742-6596/1201/1/012052

Yong, A., Martin, A., Stokoe, K., & Diehl, J. (2013). *ARRA-funded VS30 measurements using multi-technique approach at strong-motion stations in California and central-eastern United States*. U.S. Geological Survey.

Youngs, R. R., & Coppersmith, K. J. (1985). Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates. *Bulletin of the Seismological society of America*, *75*(4), 939–964.

Zafar, A., Aamir, M., Mohd Nawi, N., Arshad, A., Riaz, S., Alruban, A., Dutta, A. K., & Almotairi, S. (2022). A Comparison of Pooling Methods for Convolutional Neural Networks. *Applied Sciences*, *12*(17), 8643. https://doi.org/10.3390/app12178643

Zeiler, M. D., & Fergus, R. (2013). Visualizing and Understanding Convolutional Networks. *CoRR*. https://doi.org/1311.2901v3

Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6848–6856. https://doi.org/10.1109/CVPR.2018.00716

Zhao, J. X. (2006). An Empirical Site-Classification Method for Strong-Motion Stations in Japan Using H/V Response Spectral Ratio. *Bulletin of the Seismological Society of America*, *96*(3), 914–925. https://doi.org/10.1785/0120050124

Zhou, D.-X. (2020). Theory of deep convolutional neural networks: Downsampling. *Neural Networks*, *124*, 319–327. https://doi.org/10.1016/j.neunet.2020.01.018

Zor, E., Özalaybey, S., Karaaslan, A., Tapırdamaz, M. C., Özalaybey, S. Ç., Tarancıoğlu, A., & Erkan, B. (2010). Shear wave velocity structure of the İzmit Bay area (Turkey) estimated from active-passive array surface wave and single-station microtremor methods: Shear velocity structure of the İzmit area. *Geophysical Journal International*, *182*(3), 1603–1618. https://doi.org/10.1111/j.1365-246X.2010.04710.x

Appendix

The SESAME (2004) guidelines recommend ensuring the statistical significance of the H/V peaks before interpreting them. The H/V curves are generated by splitting the recorded signals into windows and by computing the spectral ratio on each. The final H/V curve is their average. Sometimes, isolated disturbed windows (typically those where the operators stand too close to the seismometer) significantly alter the spectral ratios. Let us consider the average H/V curve and the corresponding spectra obtained by splitting a microtremor recording into 70 s windows (Figure A1). The spectrogram (Figure A1e) clearly shows that a few of these windows significantly differ from the others. Once these transients have been removed (see Castellaro and Mulargia, 2009 for the procedure followed to remove them), the standard deviation around the individual spectra and the H/V ratio becomes acceptable and the spectral pattern turns back to its typical feature (eye-shape around f_0) described in Figure 1.

If we present to the networks the uncleaned data (Figure A1a, c) these would be wrongly classified as stratigraphic while presenting to the networks the cleaned data (Figure A1b, d) turns into a correct classification.

If we look at the positive side, misclassifications from the neural nets could sometimes mark poorly cleaned recordings (i.e. recordings where transients were not sufficiently removed).

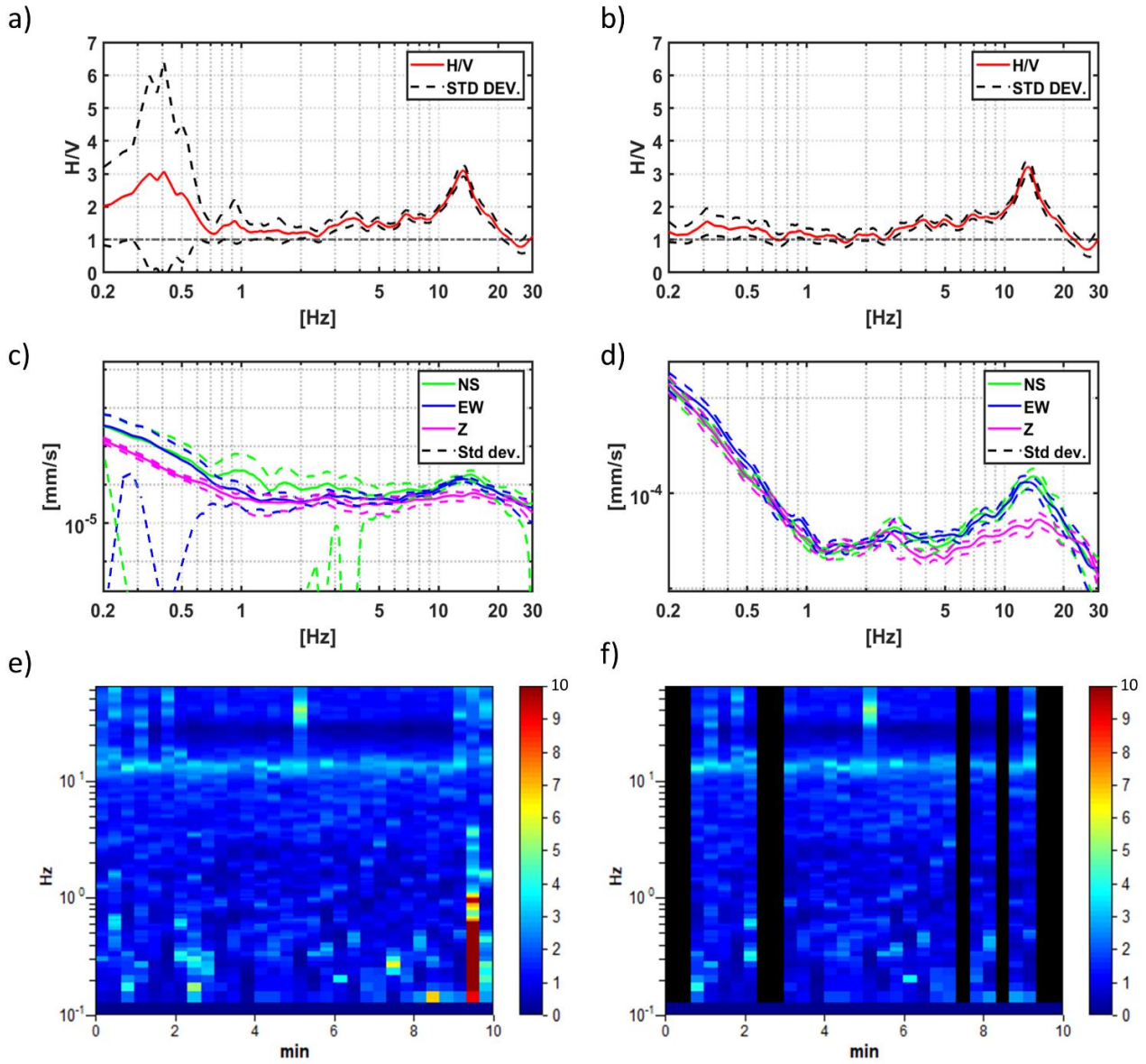

Figure A1. Average H/V curves and their standard deviations recorded at the same site before (a) and after (b) transient removal. Corresponding individual spectra of motion in velocity and standard deviation for individual components before (c) and after (d) transient removal. e) Time stability of the H/V curve during the measurement (H/V amplitude in color, time on the x-axis, frequency on the y-axis) and d) remaining time windows after transient removal.