Università degli Studi della Basilicata

International Ph.D. in
"Applied Biology and Environmental Safeguard"

Fluid injections in the subsurface: a multidisciplinary approach for better understanding their implications on induced seismicity and the environment.

Coordinator   Ph.D. candidate

Prof. ssa Patrizia Falabella   Serena Panebianco

Supervisors

Prof. Michele Paternoster
Dott. Tony Alfredo Stabile

Cycle XXXIV

Abstract

Fluid injections in the subsurface: a multidisciplinary approach for better understanding their implications on induced seismicity and the environment.

Fluid injections in the subsurface are common operations in underground industrial activities such as oil and gas exploitation, geothermal energy development, and carbon capture and storage (CCS). In recent years, it became a focal point as new drilling technologies (e.g., hydraulic fracturing) enable the extraction of oil and gas in unconventional reservoirs and the development of CCS injection techniques became a key research topic in the context of the low-carbon energy transition. Fluid injections have drawn the attention also in the general public because of their main potential implications such as the induced seismicity phenomenon (Rubinstein and Mahani, 2015) and the environmental pollution (Burton et al., 2016, Pitchel et al., 2016). Considering the strong socioeconomic impact of fluid injection operations (National Research Council, 2013; Ellsworth, 2013; Grigoli et al., 2017) the current research in this field needs the integration of multidisciplinary studies, involving knowledge on geology, seismology, source physics, hydrogeology, fluid geochemistry, rocks geomechanics for a complete understanding of the phenomenon and to set-up the most effective and "best practice" protocols for the monitoring of areas where injection operation are performed.

On this basis, this work applied a multidisciplinary approach integrating seismological methods, geochemical studies, and machine learning techniques. Two key-study areas characterized by high fluid-rock interaction and fluid-injection in the subsurface were analyzed: i) the High Agri Valley (hereinafter HAV), hosting the largest onshore oil field in West Europe, in which wastewater disposal operations have been carried out since 2006 at the Costa Molina 2 injection well and where both natural and induced seismicity clusters were recognized; ii) the Mefite d'Ansanto, the largest natural emission of CO_2-rich gases with mantle-derived fluids (from non-volcanic environment) ever measured on the Earth (Carcausi et al., 2013; Caracausi and Paternoster, 2015; Chiodini et al, 2010).

Regarding the HAV study area, we reconstructed the preliminary catalogue of seismicity through accurate absolute locations in a 3D-velocity model (Serlenga and Stabile, 2019) of earthquakes detected from the local seismic INSIEME network managed by the CNR-IMAA. A total of 852 between local tectonic and induced earthquakes occurred in the HAV between September 2016 and March 2019. We tested the potential of the unsupervised machine-learning approach as an automated tool to make faster dataset exploratory analysis, founding the density-based approach (DBSCAN algorithm-Density-Based Spatial Clustering of Applications with Noise, Ester et al., 1996) particularly suitable for the fast identification of clusters in the catalogue resulting from both injection-induced events and tectonic local earthquake swarms. Moreover, we proposed a semi-automated workflow for earthquake detection and location with the aim to improve the current standard procedures, quite time-consuming and strictly related to human operators. The workflow, integrating manual, semi-automatic and automatic detection and location methods enabled us to characterize a low magnitude natural seismic sequence occurred in August 2020 in the southwestern area of the HAV (Castelsaraceno sequence) in a relatively short time with respect to the application of standard techniques, thus representing a starting point for the improvement of the efficiency of seismic monitoring techniques of both anthropogenic and natural seismicity in the HAV. Our multidisciplinary approach involved the geochemical study of the HAV groundwaters with the aim to: (1) determine the geochemical processes controlling the chemical composition; (2) define a geochemical conceptual model regarding fluid origin (deep vs shallow) and mixing processes by means isotopic data; (3) establish a geochemical baseline for the long-term environmental monitoring of the area. A total of 39 water samples were collected from springs and wells located at the main hydro-structures bordering the valley to determine chemical (major, minor and trace elements) and isotopic composition (e.g., dD, $d^{18}O$, d13C-TDIC and noble gas). All investigated water samples have a meteoric origin, although some springs show long and deep flow than the other ones, and a bicarbonate alkaline-earth

composition, thus suggesting the carbonate hydrolysis as the main water-rock interaction process. Our results demonstrated that HAV groundwater is chemically suitable for drinking use showing no criticalities for potentially toxic metals reported by the Italian and European legislation guidelines. Particular attention was given on thermal water of Tramutola well, built by Agip S.p.a. for oil & gas exploration, with the occurrence of bubbling gases. The geochemical study highlighted a substantial difference of these CH_4-dominated thermal fluids with the rest of the dataset. Helium isotope ($^3He/^4He$) indicate a prevalent radiogenic component with a contribution of mantle-derived helium (~20%) and the average $\delta^{13}C$-CO_2 value is of – 4.6 ‰ VPDB, consistent with a mantle origin. Methane isotope composition indicates a likely microbial isotopic signature ($\delta^{13}C$-CH_4 =−63.1‰, −62.4‰, δD-CH_4=−196‰, −212‰), probably due to biodegradation processes of thermogenic hydrocarbons. The methane output at the well, evaluated by means of anemometric measurement of the volume flow (m^3/h) is of ~156 t/y, that represent about 1.5% of total national anthropogenic sources related to fossil fuel industry (Etiope et al., 2007). Our work highlighted that Tramutola well may represent a key natural laboratory to better understand the complex coupling effects between mechanical and fluid-dynamic processes in earthquake generation. Moreover, the integration of seismic and geochemical data in this work allowed us to identify the most suitable locations for the future installation of multiparametric stations for the long-term monitoring of the area and development of integrated research in the HAV.

Regarding the Mefite d'Ansanto, we analyzed the background seismicity in the emission area recorded by a dense temporary seismic network deployed at the site between 30-10-2019 and 02-11-2019. First, we implemented and tested an automated detection algorithm based on non-parametric statistics of the recorded amplitudes at each station, collecting a total dataset of 8561 events. Then, both unsupervised (DBSCAN) and supervised (KNN-k-nearest neighbors classification, Fix & Hodges, 1951) machine learning techniques were applied, based on specific parameters (duration, RMS-amplitude and arrival slope) of the detected events. DBSCAN algorithm allowed to determine characteristic bivariate correlations among tremors parameters: a high linear correlation (r~0.6-0.7) between duration and RMS-amplitude and a lower one (r~0.5-0.6) between amplitude and arrival slope (first arrival parametrization). These relationships let us to define training samples for the KNN algorithm, which allowed to classify tremor signals at each station and to automatically discriminate between tremors and accidentally detected anthropogenic noise. Results allowed to extract new information on seismic tremor at Mefite d'Ansanto, previously poorly quantitively analyzed, and its discrimination, thus providing a starting workflow for monitoring the non-volcanic emission. Isotopic geochemistry ($^3He/^4He$, $^4He/^{20}Ne$, $\delta^{13}C_{CO2}$) indicated a mixing of mantle (30%-40%) and crust-derived fluids. The source location of the emission related tremor would represent a step forward in its characterization, and for setting up more advanced automated detection and machine learning classification techniques to exploit the information provided by seismic tremor for an improved automatic monitoring of non-volcanic, CO_2-gas emissions.

List of Figures

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 3

Chapter 4

List of Acronyms and Abbreviations

ANN	Artificial Neural Networks
AP	Apulia Platform
CCS	Carbon Capture and Storage
CLP	Campano-Lucana Platform
CM2	Costa Molina 2
COVA	Centro Oli Val d'Agri
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DD	Double-Difference earthquake location method
DNN	Deep Neural Network
EAFS	Eastern Agri Fault System to the northeast
EC	Electric conductivity
EDT	Equal Differential Time
EMMWL	Eastern Mediterranean Meteoric Water Line
EMSC	European Mediterranean Seismological Center
FDSN	International Federation of Digital Seismograph Networks
FM	First motions
GFZ	Research Center for Geoscience
GMWL	Global Meteoric Water Line
GSN	Global Seismographic Network
GWB	Geochemist's Workbench
HAV	High Agri Valley
HAVO	High Agri Valley Geophysical Observatory
HDPE	High-Density Polyethylene
HVSR	Horizontal-to-Vertical Spectral Ratio technique
II	Injection induced earthquakes
INGV	Istituto Nazionale di Geofisica e Vulcanologia
INSIEME	INduced Seismicity in Italy: Estimation, Monitoring, and sEismic risk mitigation project
IPGP	Institut de Physique du Globe des Paris
IRIS	Incorporated Research Institutions for Seismology
ISNet	Irpinia Seismic Network
KNN	K-Nearest Neighbors
LE	Local natural earthquakes
LMWL	Local Meteoric Water Line
LU	Lagonegrese Units
MAC	Maximum Admissible Concentrations
ML	Machine Learning
MMFS	Monti della Maddalena extensional Fault System
MORB	Mid Ocean Ridge Basalt
MT	Magnetotelluric
ORP	Redox potential
PDF	Density Probability Function

PSD	Power Spectral Density
QD	Quaternary deposits
RI	Reservoir-Induced seismic events
RMS	Root-Mean-Square misfit
RSNC	Rete Sismica Nazionale Centralizzata
SCOLV	SeisComP3 Origin Locator Viewer
SI	Saturation indices
SIMWL	Southern Italy Water Line
SLCM	Subcontinental Litospheric Mantle
SSP	Seismic Source Parameters
STDR	Station Distribution Ratio
TDIC	Total Dissolved Inorganic Carbon
TDS	Total Dissolved Solids
USGS	U.S. Geological Survey
UTC	Coordinated Universal Time
VSMOW	Vienna Standard Mean Ocean Water standard

Chapter 1 – Introduction

Fluid injections in the subsurface are common operations in underground industrial activities such as oil and gas exploitation, geothermal energy, and carbon capture and storage (CCS). In the recent years, it became a focal point in energy development as new drilling technologies (e.g., hydraulic fracturing) enable the extraction of oil and gas in unconventional reservoirs and deep injection wells are drilled to dispose wastewater generated by oil and gas production. Moreover, the increasing manifestations of an already started climate change made as urgent as ever an energy transition from high-carbon centered to low-carbon, favoring the research and development of Carbon Capture and Storage (CCS) techniques of CO_2 sequestration in subsurface reservoirs to reduce anthropogenic CO_2 emissions to the atmosphere. Fluid injections became a topic of interest also in the general public because of their main potential implications such as the induced seismicity phenomenon (Rubinstein and Mahani, 2015) and the environmental pollution.

From the environmental point of view the main concerns ascribed to fluid injections, are related to the contamination of both soils and sediments (Burton et al., 2016; Vengosh et al., 2014) and water (Vengosh et al., 2014,Fontenot et al., 2013; Jackson et al., 2013), possibly associated to spills of fluids during drilling and fracturing processes or to failure of wastewater pipelines or of the well casings and corrosion of pipes and tanks (Pichtel, 2016). Regarding induced seismicity, it has been documented since the 1920s in oil and gas exploitation (Goose Creek oil field ,USA) (Pratt and Johnson, 1926). Today it is commonly accepted that the term "induced seismicity" is synonymous with anthropogenic seismicity that is, the one directly related to human activity. However, several studies (i.e., McGarr et al., 2002; Shapiro, 2015) discriminate between induced (*stricto sensu*) and triggered seismicity. The first term refers to earthquakes which nucleation and rupture processes are driven by stress changes principally caused by human activities; the latter, defines seismic events in which human activity could act as a trigger, while tectonic stress plays a primary role (Dahm et al., 2015). In this work, all the previous terms are considered interchangeable. Most induced seismic events are classified as microseisms (M≤2). However, potential exists to produce earthquakes that can be felt by people and cause damage to buildings and infrastructures. For this reason, in the recent years the phenomenon has drawn attention due to the increasing number of reported cases occurred near industrial facilities carrying out fluid injection operations and increased the public concern about the development of these industrial activities (National Research Council, 2013; Ellsworth, 2013; Grigoli et al., 2017). The increase of induced seismicity occurred in Basel (Switzerland), led to the definitive suspension of its Enhanced Geothermal System (EGS) project in 2009, due to the raised concern on the topic. Although seismic events occurring close to industrial facilities often raise concerns among the local communities, discriminating between anthropogenic and natural seismicity is not an easy task. An example is the May 2012 Emilia (Italy) seismic sequence that culminated with a magnitude 5.9 (Ml) event on 20 May 2012 and a magnitude 5.8 (Ml) event 9 days later (Galli et al., 2012). It caused intense political debate after the ICHESE (International Commission on Hydrocarbon Exploration and Seismicity in the Emilia Region) scientific commission stated a possible correlation with the hydrocarbon production at the Cavone oil field located just to the northwest of the 29 May

aftershock epicentral area (Grigoli et al., 2017). These results were later disproved from other research groups (Astiz et al., 2014; Dahm et al., 2015; Juanes et al., 2016) which excluded pressure changes associated to the hydrocarbon exploitation at the Cavone field in earthquakes triggering and concluded that the 2012 Emilia earthquake sequence was neither triggered nor induced by field depletion, but very likely of tectonic origin (Dahm et al., 2015). Despite these scientific findings, the anthropogenic origin of the Emilia earthquakes remains a topic of public debate in Italy and raised the attention on the induced/triggered seismicity of local population. This case highlights the difficult issue of discriminating between natural and anthropogenic seismic earthquakes. Considering the strong socioeconomic impact of fluid injection, environmental issues and its seismic implications (National Research Council, 2013; Ellsworth, 2013; Grigoli et al., 2017) the current research in this field needs the integration of multidisciplinary studies, involving knowledge on geology, geodesy, seismology, source physics, hydrogeology, fluid geochemistry, rocks geomechanics for a complete understanding of the phenomenon and to set-up the most effective and "best practice" protocols for the monitoring of areas where injection operation are performed.

The multidisciplinary research in this field aims to better understand the complex interactions among rocks, faults, and fluids involved in earthquake generation, and how perturbations of the stress field, at either small or regional scale, may affect the stability of faults over time. This is crucial to understand both natural and induced earthquakes nucleation processes. Between other physical and geomechanical factors, fluid pore pressure perturbation has been widely recognized as major factor in induced seismicity occurrence (Shapiro et al., 2006; Stabile et al., 2014; Zoback and Harjes, 1997) due to its influence in reducing faults strength. On the other hand, over-pressurized crustal fluids may have been one of the main favoring factors in large ruptures associated with the most damaging seismic sequences (e.g, the 2016-2017 Amatrice-Norcia seismic sequence, (Chiarabba et al., 2018). It follows that, better understanding the complex interactions among rock, faults, and crustal fluids as a unique, complex system may be a key point not only for comprehend the fluid-injection induced seismicity, but also to assess the relationships between fluids migration in the subsurface and earthquakes generations in tectonically active areas. Moreover, in seismic regions, fault systems and fractures act as preferential conduits for natural fluid migration in the subsurface, in which pristine chemical and isotopic composition can be modified by both mixing processes of fluids coming from different sources and secondary chemical processes at shallow levels. Modeling the origin and mixing of fluids and the water–rock–gas interactions in aquifer systems can represent an improvement in the comprehension of crustal deformation processes. This would allow also to identify the most suitable sites for long-term geochemical monitoring of some chemical-physical parameters to highlight any relationship between seismicity and fluid transfer and/or variations in water-gas interactions. The identification of pure geochemical precursors of seismic activity are a still-open issue in scientific research. However, geochemical changes or anomalies prior to seismicity have been historically recorded. For instance, Italiano et al. (2001) reported seismically induced variations in gas composition and flow of a deep source CH_4-dominated fluid monitored at the Tramutola well in coincidence with the 1996 ML 4.9 Irpinia earthquake (Southern Apennines, Italy). It follows that the geochemical characterization of the High Agri Valley groundwaters, along with the assessment of the of origin and mixing of fluids of different sources (mantle, crust, atmosphere) can represent an improvement in the comprehension

of its crustal deformation processes and for the better understanding of the possible seismic-related geochemical processes.

From the social, economic, and environmental point of view, the multidisciplinary characterization of areas with high-fluid-rock interaction it is useful for the development of specific protocols for the management of fluid injection operations, the improvement of monitoring strategies of areas in which industrial operation are carried out and the definition of the associated hazard and risk. Improve seismic monitoring, intended as to identify, and localize the seismicity in a volume surrounding the area where human activities take place (MiSE, 2014), is one of the main purposes of the seismological research. It must allow the reconstruction of the spatio-temporal and magnitude evolution of the seismicity for the improvement of traffic light systems (Guidelines for Monitoring Seismicity, Ground Deformation and Pore Pressure in Subsurface Industrial Activities, MiSE, 2014) for risk mitigation and management of fluid injection activities, along with the development of methodologies to discriminate natural and induced seismicity. Since the environmental risk related to fluid injection operations, their monitoring strategies must involve the evaluation of groundwater resources through the assessment of their geochemical features, the discrimination of the geogenic and anthropogenic influences on groundwater quality and the definition of a geochemical baseline for the long-term monitoring of industrialized areas.

On this basis, this PhD project applied a multidisciplinary approach, which results would contribute to a step-forward in the development of monitoring strategies of industrial areas where fluid-injection operations are carried out. In doing this, two key-study areas characterized by high fluid-rock interaction and fluid-injection in the subsurface were analyzed:

(1) the High Agri Valley, a tectonically active area in southern Italy which hosts the largest onshore oil field in Europe with a current production rate of about 80.000 barrels/day (Buccione et al., 2021). Beyond its natural seismicity, two different clusters of anthropogenic microseismicity were recognized in the area related to: (a) the wastewater disposal operations at the Costa Molina 2 injection well (Hager et al., 2021, Stabile et al., 2021); (b) the water loading and unloading operations in the Pertusillo artificial reservoir (Telesca et al., 2015). Therefore, it is an ideal study area to deepen the understanding of driving processes of both natural and anthropogenic earthquakes, and to improve the current methodologies for seismic and geochemical monitoring in areas where industrial operations are carried out. Aim of this work is to provide an upgraded reconstruction of spatio-temporal evolution of the High Agri Valley seismicity through non-linear absolute earthquake locations in a 3D-velocity model of the area. Moreover, we show the results of an advanced workflow, involving manual, semi-automated and automated techniques for earthquake location, which would represent a useful tool for seismic monitoring for a faster and efficient characterization of both natural and induced-seismicity clusters. This research contributes also to define a model of fluids origin and the water–rock–gas interaction processes in aquifer systems of the High Agri Valley and to provide a geochemical baseline of the area. Moreover, the work allowed to identify the most suitable locations for the installation of multiparametric geochemical stations for the long-term, continuous, and high-resolution geochemical monitoring of the area.

(2) the Mefite d'Ansanto (Rocca San Felice, southern Italy), the largest natural emission of CO_2-rich gases with mantle-derived fluids (from non-volcanic environment) ever measured on the Earth (Chiodini et al., 2010). Given its massive emission and the geological setting in which it arises, the Mefite d'Ansanto site offered a unique opportunity to develop monitoring techniques of the effects of a large gas leakage from a CO_2 buried gas reservoir of a CCS injection project (Chiodini et al., 2010, Pruess, 2008) and for the study of deep-fluid migration on the surface in high seismic hazard areas. In this work, we developed an automated detection algorithm and applied a machine-learning approach on the gas-emission related seismic tremor, which enabled its automated classification and discrimination, thus providing a starting workflow to monitor the non-volcanic emission either to gain information on its evolution for seismic hazard purposes or for management of both natural and CCS leakage CO_2 gas emissions. From the geochemical point of view, a model of fluids origin was obtained from the analysis Helium and carbon isotopic signatures of the emitted free gases at Mefite d'Ansanto. The model, in accordance with previous observations (Caracausi and Paternoster, 2015; Italiano et al., 2000) confirmed the deep source of these fluids mainly derived from the mixing between crust and mantle endmembers, with a negligible influence of the atmosphere.

Chapter 2 – Methodologies

2.1 Introduction

This chapter is a review of the basic principles of both seismological and geochemical methods applied in this work, to act as a reference in the following chapters. A multidisciplinary approach was addressed with seismological, integrated machine learning techniques and geochemical studies applied on the two key study areas (Figure 2.1).

For the characterization of induced seismicity in the High Agri Valley, advanced seismological techniques (earthquake detection and location, magnitude and source parameters estimation) were applied (a) to better reconstruct the spatio-temporal evolution of both natural and induced seismic events; (b) to test an automated event detection and location approach consisting of a single-station template matching technique followed by an automatic picking of P- and S-wave arrival times with a deep-neural-network algorithm (PhaseNet, Zhu and Berosa, 2018), which was applied to a natural seismic sequence occurred in August 2020 in the southeastern area of the High Agri Valley, close the municipality of Castelsaraceno (PZ) (c) to investigate source-based discrimination criteria between natural and induced earthquakes through the estimation of earthquake source parameters (seismic moment Mo, corner frequency fc, high frequency spectral decay factor γ, stress drop $\Delta\sigma$) on both tectonic and induced events of the study area. Geochemical analyses were integrated for the evaluation of the High Agri Valley groundwater origin, to track fluids migration patterns in it and to define a geochemical baseline for the environmental monitoring of the area. These methods involved the chemical and isotopic analysis of both the liquid and gaseous matrices in the selected springs and piezometers: (a) water stable isotopes (^{18}O and ^{2}H) composition for the reconstruction of fluid circulation models; (b) noble gases (He) and carbon (^{13}C) chemical and isotopic composition to assess fluid origin and the water-gas rock interactions effects.

With the purpose to detect and characterize tremor signals recorded at the Mefite d'Ansanto gas emission, a new automated and machine learning-based method has been developed in this work. An automated detection algorithm through Python-Obspy (Krischer et al., 2015) seismic signal processing tools have been implemented (see Chapter 4) to collect information regarding their time-frequency parameters (duration, amplitude, frequency content, polarization). Then, both machine-learning unsupervised (DBSCAN- Density-Based Spatial Clustering of Applications with Noise) and supervised (KNN – k-nearest neighbors classification) algorithms were applied to the collected tremor parameters for automated classification of tremor signals at each station and automated discrimination between tremors and anthropogenic noise.
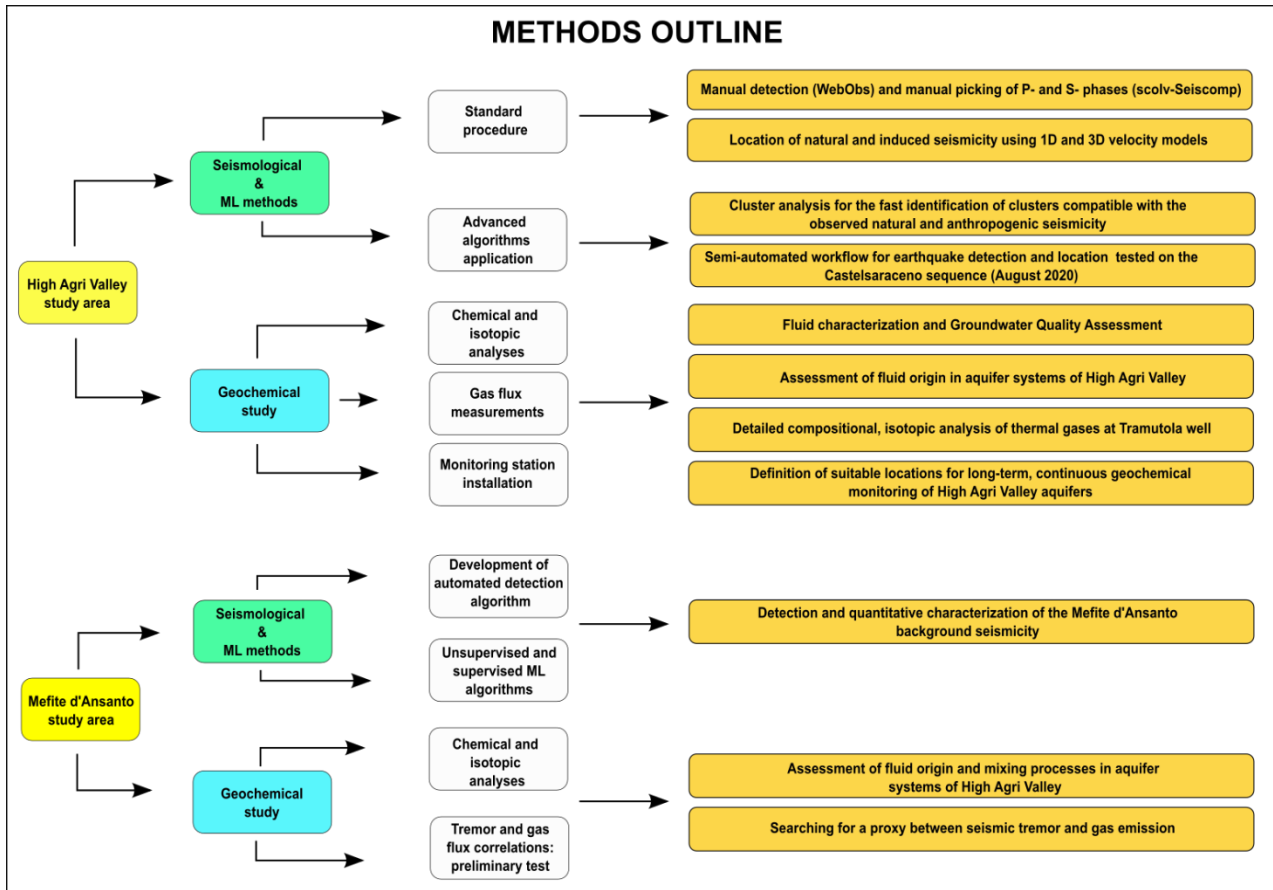
Figure 2. 1 – Schematic outline of the multidisciplinary methods addressed in this work. Seismological methods, integrated machine learning techniques and geochemical studies were applied in the two keys study areas. Main aims and obtained results from each method are listed in the orange boxes.

2.2 Seismological Methods

2.2.1 Seismic networks

Most seismic experiments require the acquisition of contemporary recordings from multiple seismometers deployed in the study area thus constituting a seismic network: sets of seismic stations which recordings – possibly in real-time - are collected by the same data center. Depending on the network inter-station average distances, seismic networks can be classified as global, national, regional, and local. Setting up a seismic network is inherently dependent to its scientific purpose such as studying earth structure and tectonics, monitoring of seismic and volcanic hazardous areas, reconstruct spatio-temporal evolution of seismogenic structures, identifying explosions from nuclear tests. Global networks for geodynamic studies are characterized by an average distance of thousands of kilometers between stations, whereas in local, small-scale monitoring networks, stations are deployed within few kilometers distance. Site accessibility and morphology are also key points when setting the number of network stations and their spatial distribution, particularly for earthquake location methods which precision is higher for homogenous distribution (wide azimuthal coverage). The Global Seismographic Network (GSN) is a 152 station, globally distributed, digital seismic network that provides real-time and open access seismological data to the international scientific community. The GSN is jointly managed by IRIS (Incorporated Research Institutions for Seismology) and the USGS (U.S. Geological

6

Survey) and collects stations from networks installed from research centers and universities all over the world (e.g., USGS, UCSD/IDA GEOFON, GEOSCOPE). The Italian National Seismic Network (Rete Sismica Nazionale Centralizzata, RSNC) currently consist of 90 seismic stations managed from the Istituto Nazionale di Geofisica e Vulcanologia (INGV) which data are collected at the seismic data center in Rome. The national network provides to the scientific community and the public near real-time data for seismic and volcanic hazard monitoring purposes, and for location and magnitude estimation of earthquakes occurring in the national territory. Regional networks, with intermediate size between national networks and local seismic arrays, are usually constructed to monitor particularly tectonically active areas. An example is the ISNet (Irpinia Seismic Network) a dense regional seismic network operated in the southern Apennines (Italy) and managed by the Physics Department of the University of Naples Federico II. The network is composed of a total of 30 stations equipped by accelerometers and short period / broadband seismic sensors, installed along complex and active fault systems which are capable of producing up to M 6.9 magnitude earthquakes (Irpinia earthquake, 23 November 1980). ISNet allowed the implementation of new methodologies of real-time data analysis and management and to set a prototype for an earthquake early warning (Iannaccone et al., 2010; Satriano et al., 2008) system in the Campania Region. To develop detailed studies regarding seismicity nucleation processes and/or to obtain, high-resolution datasets and monitor specific seismogenic and volcanic areas, dense-local seismic network should be deployed. The latter are characterized by close stations with average distances within few kilometers. A particular kind of local seismic network is represented by seismic arrays, consisting of a set of sensors deployed in regular geometric patterns (linear, circles, cross, etc.). Seismic arrays have been introduced since 1960's (Keen et al., 1965) to improve signal-to-noise ratio and increase resolution in detecting earthquakes and nuclear explosions. Seismic network and seismic arrays mainly differ from data processing techniques (e.g beamforming): most array processing techniques require high signal coherency between array stations, highly affected from the array geometry, spatial extent, and data quality and station relative timing.

In this work, data from both a local seismic network (CNR-INSIEME seismic network, now HAVO seismic network – http://www.fdsn.org/networks/detail/VD) and seismic array (GEOFON YZ - Mefite Array) have been analyzed and will be described in detail in the next chapters. The INSIEME seismic network (Chapter 3) (Stabile et al., 2020) installed in the High Agri Valley (HAV) (Basilicata Region, Southern Italy) and managed by CNR – Italian National Research Council, was designed for (a) the study the seismic processes related to the occurrence of anthropogenic induced seismicity and (b) to collect open-access high-quality seismic data for studying such phenomena to be shared with the scientific community. The Mefite d'Ansanto (GEOFON YZ) (Chapter 4) is a dense temporary array, installed between 28-10-2019 and 05-11-2019 by the German Research Center for Geoscience (GFZ) as part of a temporary experiment (East Pollino Experiment, Southern Italy, network DOI doi:10.14470/LM582526, Passarelli et al., 2017) with the aim to study the background seismicity related to the massive gas emission in the site.

2.2.2 Earthquake detection

The dataset of local tectonic and induced earthquakes in the High Agri Valley study area has been updated to develop a reference database for seismological analysis. With this aim, the detection of events recorded by local (the CNR "INSIEME") and national (the INGV- Rete Sismica Nazionale) seismic networks from September 2016 up to March 2019 have been carried out both manually and automatically. Manual detection enabled us to identify a new natural seismic sequence occurred in August 2020 in the southeastern part of the HAV (Castelsaraceno, PZ). For manual detections, a dedicated intranet system (WebObs – Beauducel et al., 2020) developed by the Institut de Physique du Globe des Paris (IPGP) has been used. In it, a near real-time multi-record strip-chart (SefraN) of the configured stations seismograms from the local INSIEME (INS1, INS2, INS3, INS4, INS5, INS6, INS7, INS8, MARCO) and national RSNC (SLNC, MCEL, MTSN, SIRI, SCHR) networks, allowed earthquakes detection (Figure 2.2). Each declared event is then saved in the WebObs system and collected in a Daybook - whether it can be located or not - which stores the estimated event information (start time, end time, t_S - t_P delay, overall duration, first pick station). These will be automatically sent to the SeisComP3 database, for event location through SeisComP3 Origin Locator Viewer (SCOLV).
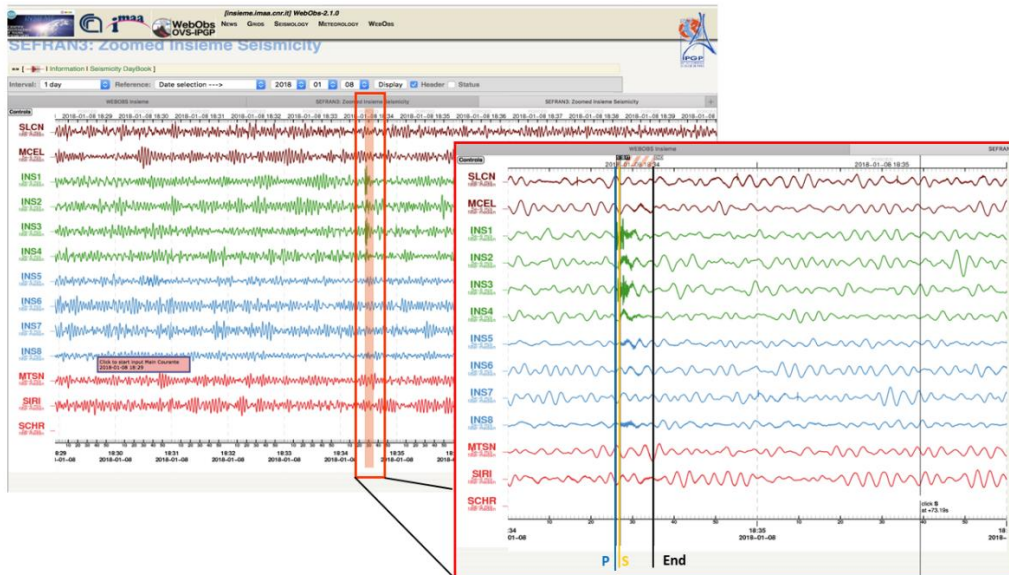
Figure 2. 2 - Multi-record strip-chart (SefraN) of the configured stations on the WebObs intranet system. It allowed the manual detection of earthquakes and their collection in the database (Daybook).

2.2.2.1 Template Matching Algorithm

To increase the dataset completeness magnitude, weak earthquakes have been detected by cross-correlating the continuous data streams with the waveforms of selected master events through a single-station Template Matching Algorithm (Gibbons and Ringdal, 2006; Roberts et al., 1989; Stabile et al., 2021) available at the CNR-IMAA. The method is based on the analysis of three-component seismic data from a single station - which generally is the closest to the analyzed seismicity cluster - and allows to efficiently detect small earthquakes even in noisy time-series by measuring their waveform similarity with one (or more) waveform template. Waveform similarity is estimated through cross-correlating the template waveform with successive

time segments of the recorded data (Gibbons and Ringdal, 2006) . The method is based on the observation that two waveforms recorded from the same sensor would show high similarity if: (a) the inter-event distance is small; (b) these were produced by similar source mechanisms. Uchida (2019) estimated that a set of earthquakes with source distance not greater than a quarter of the dominant wavelength show almost identical waveforms when filtered in a sufficiently low-frequency band. Hence, the method is highly suitable to detect as many microearthquakes as possible which are likely to have occurred close to the site of interest (e.g., belonging to the same cluster or seismic sequence) and which escape manual detection due to low signal-to-noise ratio. Let be $w_{N,\Delta t}(t_0)$ the vector of N consecutive samples of a non-zero time series (waveform) w(t) with Δt sampling interval at time of the first sample t_0:

$$w_{N,\Delta t}(t_0) \;=\; [w(t_0), w(t_0 + \Delta t), \dots, w(t_0 + (N-1)\Delta t]^T \qquad \text{Eq. 2.1}$$

The inner product between two waveforms $v_{N,\Delta t}(t_v)$ and $w_{N,\Delta t}(t_w)$ is defined by:

$$
\begin{aligned}
\langle v(t_v), w(t_w) \rangle_{N,\Delta t} \;&=\; \big\langle v(t_v)_{N,\Delta t}, w(t_w)_{N,\Delta t} \big\rangle \qquad \text{Eq. 2.2}\\
&=\; \sum_{i=0}^{N-1} v(t_v + i\Delta t) w(t_w + i\Delta t)
\end{aligned}
$$

And their normalized cross-correlation coefficient (XC) will be given by:

$$XC[v(t_v), w(t_w)]_{N,\Delta t} \;=\; \frac{\big\langle v(t_v)_{N,\Delta t}, w(t_w)_{N,\Delta t} \big\rangle}{\sqrt{\langle v(t_v), v(t_v) \rangle_{N,\Delta t}\ \langle w(t_w), w(t_w) \rangle_{N,\Delta t}}} \qquad \text{Eq. 2.3}$$

Being normalized, theorical XC will always lie in the interval [−1, 1] with the extreme values occurring when one of the time-series is an exact multiple of the other, through α with the same sign of XC:

$$v_{N,\Delta t}(t_v) \;=\; \alpha w_{N,\Delta t}(t_w) \qquad \text{Eq. 2.4}$$

For real waveforms the unity will almost never be achieved but high value of *XC* will indicates high degree of waveform similarity. If w(t) is the recorded signal at a single seismometer channel at the starting time (t_M) of a data window containing the master template waveform, the cross-correlation function which measures the similarity between the waveform in the immediately following time (t) and the template waveform will be:

$$XC_w(t)_{N,\Delta t} \;=\; XC[w(t), w(t_M)]_{N,\Delta t} \qquad \text{Eq. 2.5}$$

It must be noted that if $XC_w(t)_{N,\Delta t}$ is a detector of waveform similarity, but it does no constitutes a detection (Gibbons & Ringdal, 2006). Indeed, high values of XC can be related to template waveform length and complexity: if w(t) is bandpass filtered in a narrow frequency band and the template signal is short and monochromatic, then it is likely that XC_w(t) will frequently reach a high value so. To facilitate the automatic detection of significant values of the correlation coefficient, a scaled correlation coefficient can be calculated (*XC'*) (Gibbons & Ringdal, 2006) which measures the correlation at a given time t in relation to its RMS values in a time interval (I) in the vicinity of t. Given the I interval and XC_{RMS} given by, respectively:

$$I = [t - b, t - a] \cap [t + a, t + b]$$ Eq. 2.6

$$XC_{RMS}(t) = \sqrt{\frac{\sum_{tk \in I}\left[XC_w(t_k)_{N,\Delta t}\right]^2}{\sum_{tk \in I} 1}}$$ Eq. 2.7

With $0 < a < b$ empirically determined such that: (1) the a value would be sufficiently large to produce a time interval I which does not include significant peaks in the autocorrelation function; (2) the b values will vary according to the frequency content of the signal since it is set to produce a data window long enough to provide a relatively stable estimate of the variation anticipated in the correlation function in the absence of close matches with the waveform template (Gibbons & Ringdal, 2006). From the (2.6) and (2.7) the scaled correlation parameter (*XC'*) will be estimated as:

$$XC'_w(t)_{N,\Delta t} = \frac{XC_w(t)_{N,\Delta t}}{XC_{RMS}(t)}$$ Eq. 2.8

In this work a single-station template matching algorithm, based on the previously described technique and implemented by Stabile et al. (2021), was applied both on natural earthquakes of the Castelsaraceno seismic sequence and on induced seismicity cluster of the CM2 injection well. Master templates were selected at the nearest station of the analyzed seismicity of the INSIEME seismic network to automatically detect weaker micro-earthquakes with high similarity of waveforms. All the three-component templates and the continuous data stream of the analyzed station were band-pass filtered in the same range and then differentiated to obtain their ground acceleration displacement. Finally, each master template component was separately correlated with its correspondent component of the continuous data stream. For an unknown event to be automatically detected two conditions would be contemporary fulfilled (Figure 2.3): (1) the obtained cross-correlation parameter (*XC'*) > 0.5; (2) The first condition would be satisfied on the vertical component and at least on one of the 2 horizontal components.
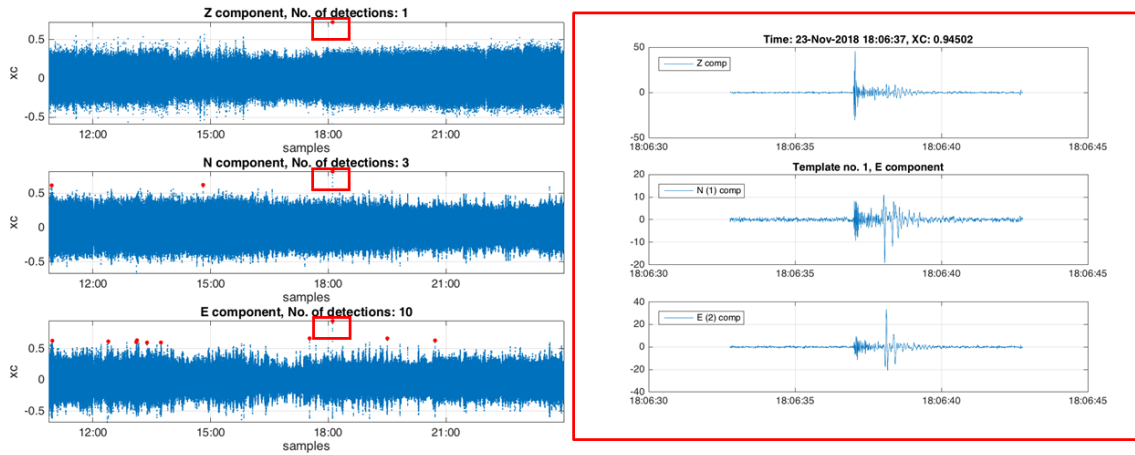
Figure 2. 3 – Automatedly detected event (right red square) through the Master Template Matching algorithm applied on High Agri Valley microseismicity. The image shows (on the left) the cross-correlation coefficient (xc) variations at 3-component waveforms of the same station. The event was detected because of its xc values > 0.6 on the vertical (Z) component and in both the horizontal ones (N and E).

2.2.3 Earthquake Location

Earthquake location consists in determine the source of the radiating seismic energy by defining its hypocenter – the physical location of the starting point of the rupture process – geographical coordinates (longitude x_0, latitude y_0, depth z_0) and origin time (t_0) intended as the start time of the source rupture. The location may also refer to the epicenter (the projection on the earth's surface of the hypocenter), or the centroid (center of earthquake's energy release). Earthquakes can be located by using different techniques ranging from linearized methods to non-linear probabilistic techniques, from absolute locations to relative locations. An absolute earthquake location is estimated within a fixed, geographic and time system (e.g., Coordinated Universal Time (UTC)); while a relative earthquake location is determined with respect to another spatiotemporal object (e.g., another earthquake) (Hardebeck and Husen, 2010). Earthquake location is inherently a non-linear inverse problem. Indeed, because seismic events take place along faults deep in the Earth, their source locations must be inferred from observations at surface: the recorded arrival times of seismic phases at a seismic network. Then a model of seismic waves propagation (velocity) must be assumed "a priori" to predict the source location parameters (x_0, y_0, z_0, t_0) (Zollo and Emolo, 2011). Seismic velocity structure of the Earth in location problems is also unknown, so that, each single earthquake location would require determining the velocity model as well (coupled hypocenter-velocity problem, Thurber, 1992). This is not possible and, consequently, the most representative propagation model for the studied area - where available - is fixed for an earthquake location. For a given model of slowness u(x) (inverse of seismic velocity) and source location $x_{source} = (x_0, y_0, z_0)$ the observed arrival time (t_i') at the ith-station with coordinates $x_{stat} = (x_i, y_i, z_i)$ can be expressed as:

$$t'_i = t_0 + \int_{x_{source}}^{x_{stat}} u(x)\, ds \qquad \text{Eq. 2.9}$$

where x is the vector of spatial coordinates sampling the ray path from the source to the station and t_0 is the origin time. Equation (2.9) is nonlinear since a change in the source location changes the ray path over which the integral is computed. Consequently, source location which depends on the observed arrival times, is a non-linear problem. To overcome the non-linearity of the location problem many methods have been developed, usually based on the iterative, linearized scheme of Geiger (1912) (e.g., Hypo71, Lee and Lahr, 1975). Increasing computing power allowed the introduction of non-linear global search methods. Among the latter methods, in this work we used NonLinLoc (Lomax et al., 2000) which is based on the stochastic exploration of the solution probability distributions of the earthquake location problem.

2.2.3.1 Linearized location methods

Earthquake location expressed from the Equation (2.9) represent a non-linear inverse problem which searches for the hypocentral parameters (x_{source}) better reproducing the observed data (t'_i). An inverse problem, can be summarized by the following matrices expression:

$$d' = g(m) \qquad \text{Eq. 2. 10}$$

where d' and m, are the observed data and the model parameters vectors respectively:

$$d' = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}$$

$$m = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}$$

If an "a priori" velocity model is assumed, d' is the vector of the observed arrival times at each station (t'_i) while g is a non-linear model dependent from the $m = (x_0, y_0, z_0, t_0)$ spatio-temporal hypocentral parameters. To find m-values which better reproduce the observed data (t'_i), linearized location methods set some prior estimate of it (m^0). Given the propagation model, from this 'a priori' start solution, theorical arrival times (d^0) can be estimated to all the considered stations:

$$d^0 = g(m^0) \qquad \text{Eq. 2. 11}$$

Then, the linearized method needs to be iterated for j-steps by perturbing theoretical model parameters (Δm_j^0) using the result of the previous step (m_j^0) as prior estimate for the next iteration:

$$m_j = m_j^0 + \Delta m_j^0 \qquad \text{Eq. 2. 12}$$

The iteration continues until some pre-defined convergence criteria (e.g least square error minimization) between theoretical and observed arrival times is met. Following these principles, the Geiger linearization approach implemented in most of the linearized location methods (e.g., Hypo 71) can be illustrated, considering an homogeneous velocity model, with a constant velocity value (c). The equation (2.9) of the recorded arrival times (t'_i) at the i-th station can be written as the non-linear relation:

$$d'_i = t'_i = t_0 + \frac{1}{c}\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2} \qquad \text{Eq. 2. 13}$$

where t'_i is a i-th element of the observed data vector d'. The equation 2.13 can be linearized through the first two terms of the Taylor series:

$$d'_i = t'_i \approx t_i^0 + \sum_{j=1}^{4} \frac{\partial t'_i}{\partial m_j}\bigg|_{m^0} \Delta m_j^0 \qquad \text{Eq. 2. 14}$$

If we define the difference between the observed and the theorical arrival times, from the previous equation we obtain:

$$\Delta d_i^0 = \Delta t_i^0 = t'_i - t_i^0 \approx \sum_{j=1}^{4} \frac{\partial t'_i}{\partial m_j}\bigg|_{m^0} \Delta m_j^0 \qquad \text{Eq. 2. 15}$$

Where:

$$G_{ij} = \sum_{j=1}^{4} \frac{\partial t'_i}{\partial m_j}\bigg|_{m^0} \qquad \text{Eq. 2. 16}$$

So that, the equation 2.13 can be written as:

$$\Delta d_i = \sum_{j=1}^{4} G_{ij} \, \Delta m_j \qquad \text{Eq. 2. 17}$$

A compact form for the linearized problem (2.17) is given by the matrix equation:

$$\Delta d = G \Delta m \qquad \text{Eq. 2. 18}$$

Which represent a system of linearized equation, which for n observed arrival times:

$$\begin{pmatrix} \Delta d_1 \\ \vdots \\ \Delta d_n \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ \vdots & \vdots & \vdots & \vdots \\ G_{n1} & G_{n2} & G_{n3} & G_{n4} \end{pmatrix} \begin{pmatrix} \Delta m_1 \\ \vdots \\ \Delta m_4 \end{pmatrix} \qquad \text{Eq. 2. 19}$$

To solve the system of equations for each observed data a prediction error is estimated:

$$e_i = d^{obs} - d^{teo} \qquad \text{Eq. 2. 20}$$

Generally, for the solution of the inverse problem we search for the hypocentral parameters which minimize the prediction errors. In the Geiger approach the minimization of the least square of the prediction errors is searched:

$$e = \sum_{i=1}^{N} |e_i|^2 \qquad \text{Eq. 2. 21}$$

where $e_i{}^2$ is the square of the Euclidean distance (L2 norm) between the observed and theorical values. For all the observed arrival time, the system of equations for the prediction minimization can be represented from the matrix expression:

$$e = d - Gm \qquad \text{Eq. 2. 22}$$

Where:

$$e = \sum_{i=1}^{N} |e_i|^2 = e^T e = (d - Gm)^T (d - Gm) \qquad \text{Eq. 2. 23}$$

To minimize it, the partial derivative with respect to m_i parameters will be calculated and posed equal to zero, obtaining the following system of equations:

$$G^T Gm - G^T d = 0 \qquad \text{Eq. 2. 24}$$

Supposing the existence of the invers matrix $(G^T G)^{-1}$, the least squared solution will be:

$$m^{est} = (G^T G)^{-1} G^T d = G^{-g} d \qquad \text{Eq. 2. 25}$$

Where G^{-g} generalized inverse matrix. Solving the (2.25) we can estimate the prior m^0 and determine the best perturbation of the model (Δm_0) which allow to minimize the least square between theorical and observed data. The new model can be estimated given by:

$$m^1 = m^0 + \Delta m^0 \qquad \text{Eq. 2. 26}$$

m^1 will enable to predict new, better fitting theorical data:

$$d^1 = g(m^1) \qquad \text{Eq. 2. 27}$$

The process is repeated iteratively until the prediction error will reach the same order of the mean error of the dataset and/or the obtained model perturbations are not highly significatively with respect to the fixed error threshold. With this method, a single, best-fit hypocenter and origin time location is obtained and its associated, linearly estimated uncertainties (e.g., normally distributed confidence ellipsoid centered on the best-fit hypocenter and origin time). (Husen & Hardebeck, 2010). Iterative, linearized methods have the advantage of being computationally fast, providing reliable solutions for well-constrained earthquake locations. These methods disadvantage is that their solutions are depends on the quality of the initial guess (m^0). In this work, preliminary locations of the High Agri Valley seismicity were performed for earthquakes recorded between September 2016 and March 2019 through the Hypo71 algorithm implementing the 1-D velocity model developed for the High Agri Valley by (Improta et al., 2017) (Figure 2.4). The first arrival times of seismic P- and S-waves were manually picked and located through SeisComP3 Origin Locator Viewer (SCOLV). It is

and interactive tool which enable to analyze detected earthquakes waveforms and review origin information such as picks, location, depth, time, local magnitudes, and event association. For each located earthquake, a preliminary, values of local magnitudes (ML) have been also estimated through SCOLV tools. The location dataset obtained yield to a reference catalogue of tectonic local and induced earthquakes in the study area.
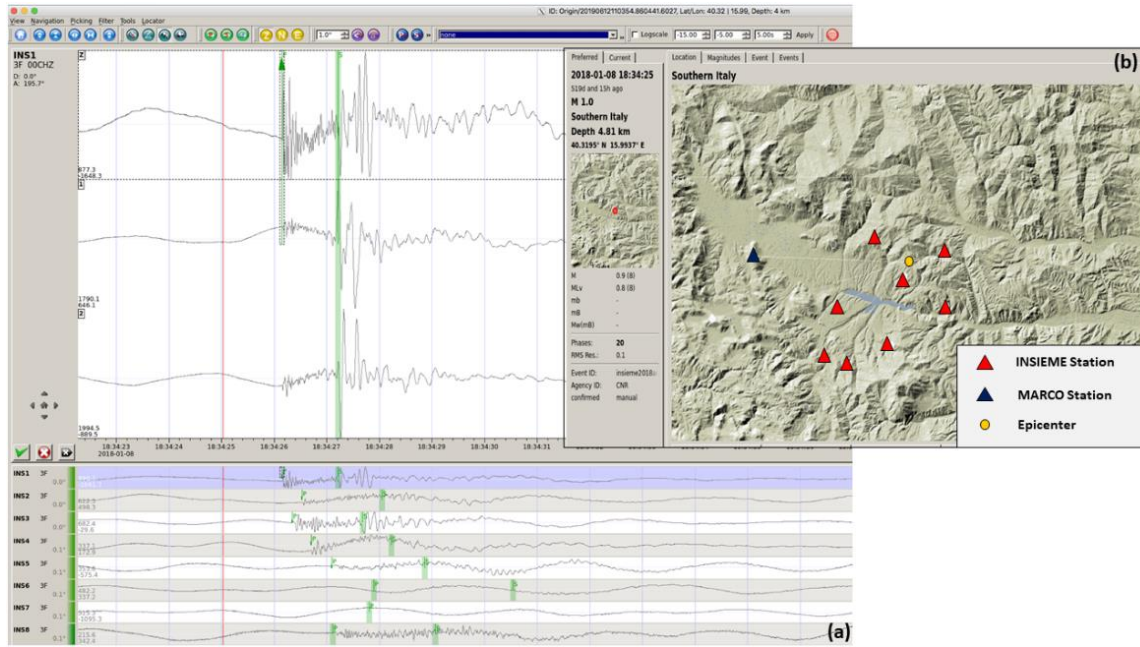
Figure 2. 4 – Interface of the SeisComP3 Origin Locator Viewer (SCOLV) for phase picking (a) and the obtained preliminary location of the 2018-01-08 earthquake (b) through the linearized, iterative, Hypo71 algorithm and the 1-D velocity model developed for the High Agri Valley by Improta et al. (2017)

2.2.3.1 Global search methods: the NonLinLoc software

Nonlinear location, grid search methods do not require the computation of partial derivatives to find the best-fit solution between observed and theorical data. Instead, they are based on deterministic or stochastic searches of the probability distribution of the hypocentral parameters (x_0, y_0, z_0, t_0) in a defined model volume. Once defined, the study area is suitably discretized in a 3D-grid. Then, supposing each grid-node the earthquake source, calculates theorical travel times for the analyzed stations. To find the solution, the times residuals (Eq. 2.20) are mapped through a density probability function (PDF). The solution of the location problem is then represented by the maximum value of the PDF. Methods to explore residual functions can range from the simpler root-mean-square (RMS) misfit (Husen and Kissling, 2001; Nelson and Vidale, 1990) to more complex likelihood functions (Lomax, 2008). When these searches gather information globally throughout the prior probability density function, they are defined global search methods (Lomax et al. 2000, Moser et al., 1992; Tarantola and Valette, 1982). The search methods used in nonlinear locations can be divided in: (a) regular, (b) deterministic, (c) directed and (d) importance sampling (Lomax et al. 2008). The first, include grid-searches, nested grid-searches, and stochastic, "crude" Monte-Carlo searches (e.g., (Sambridge and Mosegaard, 2002) (Husen & Hardebeck, 2010). These can be defined global search methods since their global

sampling of the model space. The disadvantage is their high computation demand. Directed, stochastic methods include evolutionary, adaptive global searches such as the genetic algorithm (e.g.,Sambridge and Drijkoningen, 1992). Importance sampling methods allow an efficient sampling of the global PDF function searching for the sampling density which follows the target function as closely as possible. The oct-tree is an importance sampling algorithm, efficiently applied on earthquake location. The NonLinLoc Software package was applied to obtain more accurate absolute location for both natural and induced seismicity of the High Agri Valley. The software is based on Tarantola & Valette (1982) formulation of probability density function. Given a certain model parameter (m) the probability that m will assume a value between M and $M+\Delta M$ mapped by the probability density function $p(M \leq m \leq M + \Delta M)$ will be:

$$p(M \leq m \leq M + \Delta M) \ = \ \int_{M}^{M+\Delta M} f(m)dm$$
Eq. 2. 28

Assuming an a priori uniform information on the origin time t_0 and both the theoretical model and the observed data characterized by a gaussian uncertainty with C_g and C_d covariance matrices, respectively, the marginal probability distribution $\sigma_p(x_H)$ for the hypocentral location can be obtained as:

$$\sigma_p(x_H) \ = \ K\rho_p(x_H)e^{-\frac{1}{2}L(x_H)}$$
Eq. 2. 29

Where K normalization factor, $\rho_p(x_H)$ probability density function representing the a priori information for the hypocentral location (e.g., limiting the hypocentral location in a determined volume $V_0 = \Delta x \Delta y \Delta z$), $L(x_H)$ a L2 Norm misfit function between theorical and observed data given by:

$$L(x_H) \ = \ \left[\hat{d} - \hat{g}(x_H)\right]^T (C_g + C_d)^{-1}\left[\hat{d} - \hat{g}(x_H)\right]$$
Eq. 2. 30

Where $\hat{d}$ is the vector of the difference between observed data d' and their weighted average, while $\hat{g}(x_H)$ represent the vector of the difference between the theorical arrival times and their weighed average. In both cases weights are given by:

$$w_i = \sum_j w_{ij}$$
Eq. 2. 31

With $w_{ij}= \left[(C_g + C_d)^{-1}\right]_{ij}$. The so calculated probability density function in Eq. 2.28 will represent the solution probabilistic solution of the location problem. The $\sigma_p(x_H)$ does not consider the origin time t_0, which can be calculated from the estimated hypocentral coordinates $x_H = (x_H, y_H, z_H)$ following the principle of maximum likelihood, given by the weighted average of the observed origin time at different stations (Moser et al., 1992):

$$T_0^{ML}(x_H) = \frac{\sum_i \sum_j w_{ij} [d_i - h_i(x_H)]}{\sum_i \sum_j w_{ij}}$$

<div align="right">Eq. 2. 32</div>

Where $h_i(x_H)$ represent the theorical travel time at the ith-station. The NonLinLoc software enable to search the best fit location through the estimation of the $\sigma_p(x_H)$ with different search algorithm (i.e., grid search, Metropolis-Gibbs algorithm, Oct-tree). For the absolute location of the High Agri Valley seismicity the Oct-tree importance sampling method was applied (Figure 2.5). It uses recursive subdivision of rectangular cells in a 3-D space in a cascade structure of sampled cells, such that their spatial density follows the target PDF values. The algorithm is initialized with a global sampling in a wide 3D-grid. The $\sigma_p(x_H)$ value is calculated at each cell of the 3D-grid. Once the cell corresponding to the maximum value of the PDF is estimated, it is further subdivided in a narrow grid. The process is iterated so that this oct-tree structure will have a larger number of smaller cells in areas of higher probability (lower misfit) PDF's relative to areas of lower PDF values. The Oct-tree have been proven to be a fast convergence algorithm (Zollo & Emolo, 2011) which allow an efficient and sampling of the complete probabilistic solution of earthquake location problem (Husen & Hardebeck, 2010, Lomax et al. 2008, Lomax and Curtis, 2001). Finally. direct-search methods are absolute methods which enable to solve the earthquake location problem by sampling either the entire or only parts of the solution space. Even if they can be computationally more expensive, their advantage is to not rely on the quality of an initial guess – as the linearized methods - and to obtain a more complete solution with defined uncertainties.
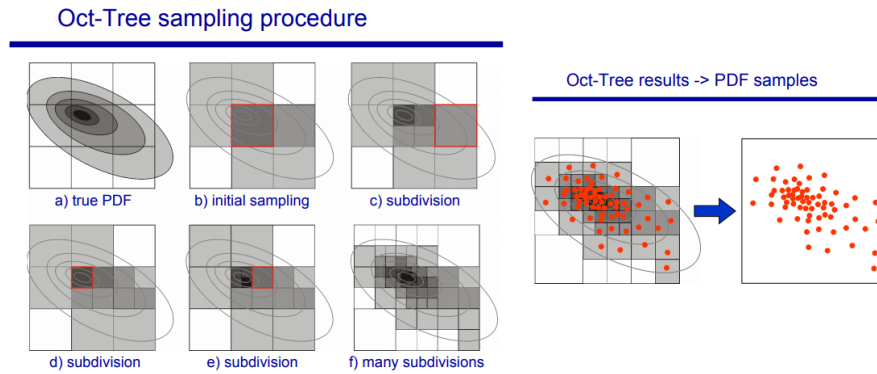
Figure 2. 5 – Oct-Tree Algorithm optimization scheme (from Zollo and Emolo, 2011).

2.2.3.1 Double-differences method

The accuracy of absolute hypocenter locations described in the previous paragraphs is controlled by many factors such as network geometry, recognizable phases and arrival-time reading accuracy, the knowledge of the crustal structure and its velocity model (Gomberg et al., 1990; Pavlis, 1986). The latter, has a key role in the location accuracy because of its influence in the source-stations travel paths and, consequently, in the seismic-phases travel times modeling to solve the inverse problem. Moreover, apply a 1-D velocity model to locate earthquakes can be a good preliminary approach, but it needs to be refined in more accurate 3-D velocity models since three-dimensional velocity variations in the propagation medium can introduce systematic biases

into the estimated travel times. The effects of errors related to unmodelled anomalies in the velocity model can be effectively minimized by using relative earthquake location methods (Frémont and Malone, 1987; Got et al., 1994; Poupinet et al., 1984). In this work double-difference (DD) earthquake location method have been applied (e.g., Waldhauser and Ellsworth, 2000) (Figure 2.6). As a relative earthquake location, it consists in locating a set of seismic events relatively to each other. Particularly, residuals between the observed and theoretical travel-time differences (double-differences) are minimized for pairs of earthquakes at each station while linking together all observed event-station pairs. The method is linearized since its high computational demand due to the high dimensionality of the problem and iteratively search for the least-squares solution adjusting the vector difference between hypocentral model pairs. DD method is based on the principle for which, if two earthquake sources (hypocenters) are close such that their distance is small compared to the event-station distance, then the ray path between the source and a common station are similar along almost the entire travel path. Consequently, analyzing two earthquakes at one station, their difference in travel times can be attributed to the spatial distance between it. The high location accuracy in this case, is due to the common origin of the absolute errors, which differs only in the small region where the ray-path differ at the source. The DD location method incorporates ordinary absolute travel-time measurements and cross-correlation P-and S-wave differential travel-time measurements (Waldhauser & Ellsworth, 2000). Indeed, location precision can be enhanced by improving the relative arrival-time readings accuracy through waveform cross-correlation methods (see paragraph 2.2.2.1). A useful rule of thumb for maximum event separation for waveform similarity is given by (Geller and Mueller, 1980) of about 1/4 wavelength of the highest frequency of importance in the seismogram. The Geiger (1912) truncated series is generally used to linearize the arrival time equation (2.9) at the k station. The resulting problem, for each observation at the k-th station, is to find an earthquake location solution evaluating the travel-time residuals (r_k^i) for the event i, which are linearly related to perturbations (Δm) of the hypocentral parameters:

$$\frac{\partial t_k^i}{\partial m} \Delta m^i = r_k^i \qquad\qquad \text{Eq. 2. 33}$$

Where residuals are the difference between theoretical and observed arrival times $r_k^i = (t^{obs} - t^{cal})_k^i$ and $\Delta m^i = (\Delta x^i, \Delta y^i, \Delta z^i, \Delta t_0^i)$.
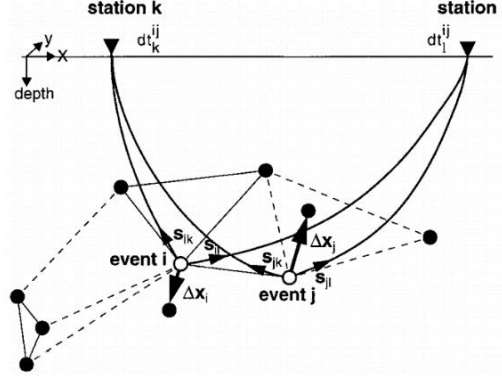
Figure 2. 6 – Double Difference relocation algorithm scheme (Waldhauser and Ellsworth 2000) for the two events i and j mutually recorded at station k and l. Open circles are the initial locations, while solid circles represent the trial hyocenters that are linked to neighboring events by cross-correlation (solid lines) or catalog (dashed lines) data.

DD methods uses waveform cross-correlation to measure travel-time differences between events i and j observed at station k, $(t_k^i - t_k^j)^{obs}$ so that the equation (2.33) cannot be used directly. The equation for relative hypocentral parameters between two events (i and j) is then given by the difference (Fréchet ,1985):

$$\frac{\partial t_k^{ij}}{\partial m} \Delta m^{ij} = dr_k^{ij}$$

Eq. 2. 34

Where $\Delta m^{ij} = (\Delta x^{ij}, \Delta y^{ij}, \Delta z^{ij}, \Delta t_0^{ij})$ is the change in the relative hypocentral parameters between the two events; while $\frac{\partial t_k^{ij}}{\partial m}$ are the components of the slowness vector of the ray connecting the source and receiver, measured at the source. In the obtained equation 2.34, the source is represented by the centroid of the two hypocenters of the events, assuming a constant slowness vector (valid for sufficiently close events); finally, dr_k^{ij} is the so-called double difference: the residual between observed and calculated differential travel time between the two earthquake i and j, given by:

$$dr_k^{ij} = (t_k^i - t_k^j)^{obs} - (t_k^i - t_k^j)^{calc} = dt_k^{ij(obs)} - dt_k^{ij(calc)}$$

Eq. 2. 35

Referring to equation 2.34 the equation 2.35 can be written as:

$$dr_k^{ij} = \frac{\partial t_k^i}{\partial m} \Delta m^i - \frac{\partial t_k^j}{\partial m} \Delta m^j$$

Eq. 2. 36

Using the appropriate slowness (or velocity) vector at the source and the origin time for each event, the DD algorithm combines equation 2.36 from all hypocentral pairs for the *k*-th station, and for all stations, to form a system of linear equations:

$$WGm = Wd$$

Eq. 2. 37

where *G* defines a matrix of size *M* x 4*N* with *M*, number of double difference observations and N, number of events containing the partial derivatives; d is the data vector containing the double difference; *m* is a 4*N*-length

19

vector containing the changes in hypocentral parameters (Δm); W is a diagonal matrix to weight each equation. The matrix G is highly sparse since each equation links together only two events (i.e., of the $4N$ columns in each of the M rows of G, only 8 have nonzero elements). If one event is poorly linked to all other events, matrix G becomes ill conditioned, and the solution to equation (2.38) may become numerically unstable. This can be improved by using only events that are well linked to other events considering only event pairs which have a minimal number of observations (Waldhauser and Ellsworth 2000). This threshold also depends on the geometrical distribution of the stations observing the two events and is usually established using trial and error approach (it should be at least 8 since the number of unknowns for one pair of events is eight) (Husen & Hardenbeck, 2010). The methods find the earthquake location iteratively: first, the initial solution to equation (2.36) is obtained from the starting locations and the a priori quality weights; then, the process is iterated updating the locations, the residuals and the partial derivatives in G. The algorithm starts by using the a priori weights until a stable solution is obtained. Then, it further iterates by re-weighting the data multiplying the a priori weights with new weights related to the misfit of the data from the previous iteration and on distance between the analyzed earthquakes. DD earthquake location is more efficient if events are densely distributed and observed at many stations. The strength of DD earthquake location lies in the fact a high precision of travel-time differences estimation through waveform cross-correlation methods (e.g., Rowe et al., 2002; Schaff and Beroza, 2004). Thus, DD earthquake location can resolve, to a certain extent, absolute earthquake locations (Husen & Hardebeck, 2010). In this work, to relocate the analyzed seismicity in the High Agri Valley, we used the HypoDD Fortran package in which is implemented the the double-difference algorithm of Waldhauser & Ellsworth (2000) described in this paragraph.

2.2.4 Source Parameters

Along with the hypocentral location, the study of seismic source parameters (SSP) is a key point to characterize earthquakes. These enable us to gather insights on the origin of seismicity, retrieving information about the physical and mechanical processes involved in the earthquake nucleation and the energy released. In active tectonic areas where also underground industrial activities are conducted (e.g., the High Agri Valley), such parameters may help to discriminate between induced and natural earthquakes (Grigoli et al., 2017); on the other hand, SSP have been widely used to determine scaling laws for seismic hazard assessments (e.g., (Abercrombie, 2015; Garcia-Aristizabal et al., 2016). SSP are generally estimated in the frequency domain, considering kinematic (Brune, 1970) or dynamic (Madariaga, 1976) source models of the body-wave displacement spectrum. Particularly, the seismic source model proposed by Brune (1970) is the most frequently used for weak to moderate earthquakes (M < 6) from the S-waves displacement spectrum. The model approximates the seismic source to a circular rupture area, of finite radius (r), along which an impulsive variation of shear stress is instantaneously applied. The corresponding modelled spectrum is a ω^{-2} source model (Aki, 1967) typically characterized by (Figure 2.7): (1) an almost flat curve reflecting the approximately constant spectral amplitude at low frequencies; (2) a curve decay from the constant level at a certain corner angular frequency (ωc); (3) a ω^{-2} spectrum decay at high frequencies ($\omega > \omega_c$).
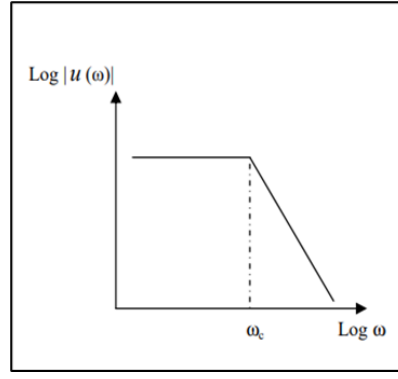
Figure 2. 7 – Theorical ω^{-2} source displacement spectrum.
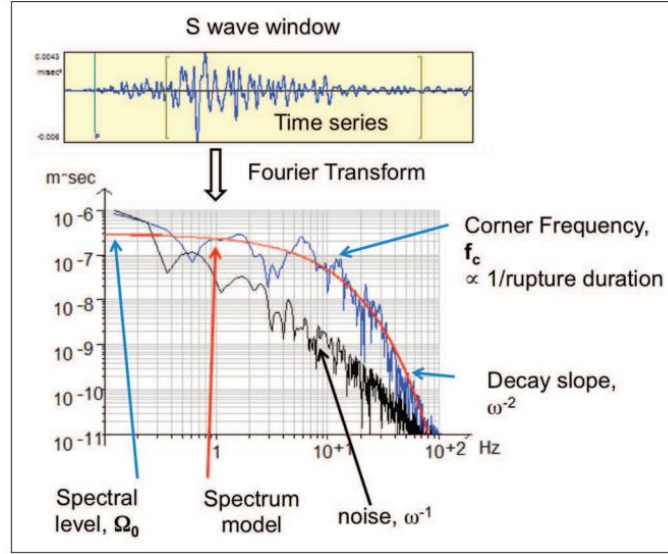
Figure 2. 8 – Example of from ω^{-2} source displacement spectrum (blue spectrum) compared with a high frequency decay ω^{-1} noise spectrum (black spectrum) (from Baig et. al., 2012). From the source spectrum both the corner frequency (f_c) and low-frequency spectral level (Ω_0)(related to M_0) are directly estimated.

The adoption of the ω^{-2} source displacement spectrum allows the estimation of (Figure 2.8): (a) the ω_c, which can also be denoted by the corner frequency $f_c = \omega_c/2\pi$. It is inversely proportional to rupture duration, so to the size of the earthquake source: corner frequency will be higher for smaller earthquakes (with shorter rupture duration). (b) The scalar seismic moment (M_0) computed from the low frequencies flat curve of the spectra. It is a measure of the irreversible inelastic deformation produced in the earthquake generation, given by the "work" accomplished by the seismic shear source in a double-couple seismic source mechanism (Bormann and Wielandt, 2013):

$$M_0 = \mu \bar{u} A \qquad\qquad \text{Eq. 2. 38}$$

Where, μ is the rigidity or shear modulus of the source medium, $\bar{u}$ is the average final displacement after the rupture and A is the surface area of the rupture. From ω_c and M_0 other source parameters can be derived: the seismic moment is used in turn to calculate the moment magnitude (M_W), while the corner frequency (f_c) combined with M_0 allows to obtain the source radius (r) and the associated stress drop ($\Delta\sigma$). The analysis of the source parameters is often complicated by their spectral properties at high frequencies, where path and site

effects are not easily distinguished from the source characteristics. This is particularly true for small microearthquakes, associated to smaller source dimensions and therefore are characterized by higher frequencies (e.g., (Abercrombie, 1995). With aim to estimate earthquake source parameters in the High Agri Valley, from the inversion of seismic signal spectra of the available datasets we adopted the already implemented SourceSpec program (Satriano, 2021). It is a collection of Python programs to determine earthquake source parameters (seismic moment M_0, corner frequency f_c) and the anelastic attenuation term (t^*), directly from the modeling of S-waves displacement spectra. The other source parameters, such as the source radius (r) and the stress drop ($\Delta\sigma$) are computed from the inverted ones, while the attenuation quality factor Q is estimated from the t^*. For each analyzed earthquake, SourceSpec enabled to estimate the Moment Magnitude (M_w) and, when required, also the local magnitude can be obtained (ML). It relies on a non-linear parametric approach (Levenberg-Marquardt algorithm) to obtain the SSP for S-wave source spectra inversion.

2.2.4.1 From the Source Model to the Source Parameters

The earthquake displacement spectrum U(ω) can be expressed by the relationship (Zollo et al., 2014):

$$U(\omega) = S_0(\omega)G(\omega)R(\omega)I(\omega) \qquad \text{Eq. 2. 39}$$

Where $S_0(\omega)$ is the source spectrum; G(ω) is the path attenuation model which takes into account the geometrical spreading, the waves radiation pattern, and anelastic attenuation of the seismic waves along the travel path; R(ω) is the site transfer function related to site (amplification/attenuation) effects related to the geology at the station; I(ω) is the instrument response function, which includes the data logger and sensor transfer functions. The I(ω) is well known so is usually corrected by instrument response deconvolution such that I(ω)=1 in the source modelling. In the far field approximation (R >> λ) – where R is the source-to-receiver distance and λ the observed wavelength - the Brune's source spectrum from which SSP are calculated, can be expressed as:

$$S_0(\omega) = C'_s \frac{\Omega_0}{1 + \left(\frac{\omega}{\omega_c}\right)^\gamma} \qquad \text{Eq. 2. 40}$$

Where Ω_0 is low-frequency spectral level (for $\omega << \omega_c$) (related to M_0), ω_c the corner frequency, γ is a constant that controls the shape of the spectrum curvature around the corner frequency (for the Brune's "omega square", $\gamma = 2$) and C'_s is a constant which accounts for the frequency-independent radiation pattern and the geometrical spreading factor in a vertically varying velocity structure (Aki and Richards,1980):

$$C'_s = \frac{R^c_{\theta\varphi}F_s}{4\pi\rho_h^{1/2}\rho_0^{1/2}c_h^{5/2}c_0^{1/2}R'} \qquad \text{Eq. 2. 41}$$

where the subscripts h and o are for density (ρ) and for P and S waves velocity (c) values at the hypocenter (h) and receiver (o), respectively; $R^c_{\theta\varphi}$ is the radiation pattern coefficient, and F_s is the free-surface coefficient (=2); while, R' is the geometrical spreading factor estimated for a linear variation of velocity with depth (Ben-Menahem and Singh, 1981) which, given R hypocentral distance, is expressed by:

$$R' = \sqrt{\frac{\rho_0 c_0}{\rho_h c_h}} R$$

Eq. 2. 42

Estimating Ω_0 and ω_c from the inversion of the amplitude spectra, the seismic moment (M_0) and the source radius (r) source parameters are estimated as:

$$M_0 = C'_s \Omega_0$$

Eq. 2. 43

$$r = k_s \frac{c}{f_c}$$

Eq. 2. 44

where c is the S-wave velocity and coefficient k_s depends on the adopted circular rupture model (for Brune model, $k_s = 0.37$). M_0 and r are used to estimate the static stress drop (Keleis-Borok, 1959), which represent the total shear stress decrease along the fault plane which nucleated the earthquake:

$$\Delta\sigma = \mu \frac{7\pi}{16} \frac{\bar{u}}{r} = \frac{7}{16} \frac{M_0}{r^3}$$

Eq. 2. 45

Considering the area of the circular Brune's source ($S=\pi r^2$) and considering the logarithms, we obtain the relationship from the Eq. (2.46):

$$\log M_0 = \frac{3}{2} \log S + \log\left(\frac{16\Delta\sigma}{7\pi^{3/2}}\right)$$

Eq. 2. 46

Which shows that, for constant values of stress drop, the seismic source size ($\log S$) is proportional to $2/3 \log M_0$. Has been recognized (Kanamori and Anderson, 1975) that it is valid for a wide range of moderate to large earthquakes (M>4) which showed that $\Delta\sigma$ takes constant values in the range 1-10 MPa independently from earthquake magnitudes. From estimates of seismic moment M_0 (Eq. 2.44) (measured in N m) the moment magnitude M_w is calculated through the relationship (Hanks and Kanamori, 1979):

$$M_w = \frac{2}{3}(\log M_0 - 9.1)$$

Eq. 2. 47

M_w is a measure of an earthquake's magnitude size directly related, through the seismic moment, to the physical properties of the source. It is preferred to other magnitude scales because it does not saturate, underestimating magnitudes as other scales do in certain conditions (see paragraph 2.2.4.1.1 for details).

2.2.4.1.1 Brief remarks on earthquake magnitudes

Magnitude scales were developed to measure the size of an earthquake. Apart from the already defined moment magnitude Mw - related to the physical properties of the source - all magnitude scales introduced before it, were based on measurements of the recorded waveform amplitudes. The original definition of magnitude, introduced by (Richter, 1935) lead to the development of a local magnitude scale (Ml), defined specifically for earthquakes occurring in California and measured with the Wood-Anderson (WA) seismometer characterized by a natural period TS = 0.8 s, a damping factor DS = 0.8 and static magnification Vmax = 2800. The first local magnitude introduced by Richter was given by the logarithm of the maximum trace amplitude (expressed in microns), with which the WA seismometer would register that shock at an epicentral distance (Δ) of 100 km:

$$Ml = \log A_{max} \qquad \text{Eq. 2. 48}$$

This local magnitude was later given the symbol ML (Gutenberg and Richter, 1956). To calculate ML for other epicentral distances further attenuation correction parameters were provided (Gutenberg and Richter, 1942; Hutton and Boore, 1987, Richter, 1935) so that, a general relation for local magnitude can be expressed as (Gutenberg and Richter, 1956):

$$ML = \log_{10}(A_{max}) + 3\log_{10}(\Delta) - 3.37 \qquad \text{Eq. 2. 49}$$

For its intrinsic limitation on the seismic sensor and epicentral distances, a new, body-waves magnitude scale (m_b) were introduced for larger Δ and based on first cycles of body-waves in the recorded waveforms (Gutenberg & Richter, 1956):

$$m_b = \log\left(\frac{A}{T}\right) + Q(h,\Delta) \qquad \text{Eq. 2. 50}$$

Where, T is the period of the body wave (e.g., P-wave or S-wave) used for the amplitude measure (generally, T=1 second) and $Q(h,\Delta)$ is a correction factor accounting for the epicentral distance (Δ) and hypocentral depth (h) experimentally determined. For Δ > 600 km, seismograms are mainly dominated from surface waves, the surface waves magnitude (M_s) were introduced, expressed by the relationship (Gutenberg & Richter, 1956):

$$M_s = \log A_{20} + 1.66\log\Delta + 2.0 \qquad \text{Eq. 2. 51}$$

With A_{20} amplitude (in micron) of the recorded surface wave with a T= 20 seconds period, corrected for the instrument response. For Ms estimation, no depth correction is needed since deep earthquake does not generate high amplitude surface waves due to their different attenuation properties with respect to body waves. Typically, the vertical component of Rayleigh waves is used for M_S calculation. As pointed out in the previous paragraph, especially for large earthquakes magnitude definition, Mw is preferred with respect to other magnitude scales because it relies on physical properties of the source. Conversely, being defined on waveform amplitudes - which in turn depend on the frequency content of seismic waveforms - ML, M_S and m_b, can

saturate, underestimating large earthquake size: the M_s magnitude saturates for $M_S > 8$; the ML saturates for ML > 6.5; while m_b completely saturates for $m_b > 6.5$.

2.2.4.2 Path attenuation model

G(ω) is the anelastic attenuation model which must be evaluated to obtain information on the region under study and obtain the source spectrum. It is related to the anelastic body-wave attenuation along the source-to-receiver path:

$$G(\omega) = e^{-\omega t^*}$$ Eq. 2. 52

Where $t^* = T/Q$ is the attenuation parameter depending on the travel time (T) and the quality factor (Q). In its general formulation, t^* is frequency dependent through a frequency dependent Q:

$$t_c^*(\omega) = \frac{T_c}{Q_0 \omega^n}$$ Eq. 2. 53

Where n is a positive real number, Q_0 is the quality factor evaluated at a reference frequency which often is fixed to 1 Hz value (e.g., Morozov, 2008). The quality factor Q is defined by the relation:

$$\frac{1}{Q(\omega)} = -\frac{\Delta E}{2\pi E}$$ Eq. 2. 54

With the second member representing the fraction of energy dissipated in a cycle by a body wave that propagates in an anelastic medium. Small values of Q are typical of highly attenuating media, while high values of Q characterize poorly attenuating media. After its definition, the Q factor is frequency dependent to satisfy the causality requirements (Aki-Richards, 1980). Conversely, Kjartansson (1979) pointed out that a constant Q model, considering a finite frequency band, also satisfies the causality principle. Therefore, the Q factor can be treated as a constant or frequency dependent. Several theoretical and experimental studies have focused on the problem of the dependence of Q on frequency, since for Q frequency-dependent (e.g., through a power law), the seismic analysis of body waves velocities should consider a strong dependence on frequency (velocity dispersion), and this is rarely inferred for P-waves analysis. Azimi et al. (1968) proposed a Q model which depends on frequency but is constant in the seismic frequency range. For this reason, several studies of source parameters and Q have been carried out by assuming a constant Q (e.g., de Lorenzo et al., 2001, 2010; (Zollo et al., 2014) or frequency-dependent Q (Zollo et al.,2014;Morozov et al., 2008). Correctly estimating the anelastic attenuation properties is one of the main problems for the accurate determination of seismic source parameters, due to their intrinsic correlation with the quality factor Q through the attenuation parameter t^*. This makes difficult to discriminate the intrinsic trade-off within Q and both corner frequency and high-frequency spectral fall-off (γ) (e.g., Sonley and Abercrombie, 2006): both γ and Q control the high frequencies spectrum decay (for $\omega > \omega_c$), and the parameter t^* (along with Q), possibly correlates with corner frequency

and seismic moment. For this reason, robust strategies (e.g., Zollo et al.2014) as well as methods based on spectral ratios (Teng, 1968) have been developed to reduce the correlation effect between the attenuation parameters ($t*$or Q) and source parameters (γ, M_0 or f_c).

2.2.4.3 Site effects response $R(\omega)$

Spectral properties at high frequencies, are also influenced by site effects $R(\omega)$ which spectral contribution to the earthquake displacement spectrum (Eq. 2.38) needs to be discriminated to enhance the accuracy of the SSP estimation. The $R(\omega)$ is generally unknown and accounts for both the near-site attenuation $K(\omega)$ and the local site amplification $A(\omega)$ (e.g., Scherbaum, 1990, De Lorenzo et al., 2010):

$$R(\omega) \; = \; A(\omega)K(\omega) \qquad\qquad \text{Eq. 2. 55}$$

The $A(\omega)$ depends on geomechanical and elastic properties of the geological substratum near the recording site and is not described by any specific expression; the $K(\omega)$ term is usually described above a frequency threshold (Hanks, 1982) in dependence of the kj high-frequency attenuation parameter (Anderson and Hough, 1984; De Lorenzo et al., 2010, Ktenidou et al., 2014):

$$K(\omega) \; = \; e^{(-\pi\omega kj)} \qquad\qquad \text{Eq. 2. 56}$$

The k_j describes the deviation at high frequencies between the observed source spectra calculated from the seismogram at the j-th station and the modelled ω^{-2} one (e.g., the Brune model).Anderson and Hough (1984) pointed out that k_j includes components related to both the local geology and the regional anelastic structure and can be expressed as (Ktenidou et al., 2014):

$$k_j \; = \; k_0 + k_s + \tilde{k}(R_e) \qquad\qquad \text{Eq. 2. 57}$$

Where k_0 is the site component (at epicentral distance $R_e = 0$), k_s is introduced source contribution term, and $\tilde{k}(R_e)$ is the epicentral distance-dependent term. The k_j influence on the high-frequency seismic spectra are a known problem for source parameter studies, particularly for microearthquakes, since they can alter the shape of the seismic spectra at frequencies comparable to the corner frequency (so that its proper estimation can be crucial (Moratto et al., 2019). Moreover, attenuation site effects are also involved in the so-called corner frequency saturation (Hanks,1982), which is an apparent departure of small-magnitude events from the Aki (1967) scaling law: fc stops to increase for decreasing magnitudes.

The observations on site effects, along with the one related to the path attenuation responses (Paragraph 2.2.4.2) underly the importance of discriminate source, attenuation, and site terms to obtain an accurate estimation of source parameters.

2.3 Geochemical methods

2.3.1 Fluid geochemistry

Fluid geochemistry involves the study of all the physical-chemical processes between rocks, minerals, and fluids (water and gas) in the subsurface. In the natural environment, these mutually interact modifying their pristine compositions; moreover, external anthropogenic activities (industrial processes, farming activities) can contribute to modify it such that geochemistry also contributes to identify the source of pollutants in the environment and their dynamics in the subsurface. Geochemical study of natural fluids, including both waters and gas matrices, represents a broad interdisciplinary science, which can rely on the use of major, minor and trace elements compositions and of isotope ratios for the reconstruction of fluid origin, the water-rock interaction processes, and their pathways towards the surface (e.g., mineral re-equilibrium or mixing).

Water geochemistry accounts on the study of the chemical processes affecting the distribution and circulation of dissolved chemical compounds in natural waters. Indeed the chemical content of waters results mainly from three factors (Bakalowicz, 1994): (1) the type of rock (e.g., mineralogy and grain size) encountered by the flowing water; (2) the climate and environmental conditions, which determine in turn the water flow rate, temperature, pressure, redox conditions, etc.; and (3) the flow conditions, determining especially the presence of a gaseous phase and the time of contact between water and rocks (i.e., residence time of water in the system). The first factor impacts the main geochemical characteristics of waters such as major dissolved species amount and TDS (total dissolved solids) as well as rock pH and electrical conductivity. The (2) and (3) factors are particularly responsible for the solution rate of rocks, depending on flow rate, and the ion content, depending on residence time and chemical content. One of the main purposes of water geochemistry is the definition of water/rocks interactions through the definition of water hydrofacies, based on the concentrations of major constituents - cations (Ca^{2+}, Mg^{2+}, Na^+, K^+) and anions (HCO_3^-, CO_3^{2-}, SO_4^{2-}, Cl^-) - in it. Four main water hydrofacies are defined: (1) Ca-Mg-SO_4-Cl (Sulphate-Chloride-Alkaline-earth), typical of water reaction with gypsum deposits or sulfur oxidation processes related to rocks / water interaction; (2) Na-SO_4-Cl (Sulphate-Chloride-Alkaline) (e.g., saline sea waters); (3) Na-HCO_3 (Alkaline-bicarbonate); (4) Ca-Mg-HCO_3 (Alkaline-earth bicarbonate) typical of carbonate or crystalline rocks/water interaction. Another purpose of water geochemistry is to reconstruct and monitor the groundwater source conditions (e.g., redox condition). These are related to amount of oxygen dissolved, which in turn depends on (a) the mixing time needed to fully saturate the water, (b) the salt content, (c) the water temperature and air pressure, (d) the oxygen consuming microbial activity. However, water redox conditions could be extremely variable across short distances due to the irregular distribution of organic-rich layers and the eventual reduced minerals filling rock fractures in the subsurface. Redox conditions are good indicators of the groundwater age: young, recently recharged groundwaters are likely to be characterized an oxidizing environment; in addition, are important environmental indicators since it strongly affects the mobility and/or persistence of many contaminants in groundwater (e.g., concentrations of arsenic and manganese are more likely to be higher, at levels that exceed human-health thresholds, in reducing-environment waters).

Fluid geochemistry involves the determination of free and dissolved gases (gas geochemistry) origin as well. Indeed, the gas content of natural waters is strictly related to its pathway in the subsurface and to the water

source environment. In addition, chemical and isotopic studies of the gaseous matrix is an important tool to reconstruct waters equilibria with the atmosphere and eventual mixing processes between different reservoirs (atmospheric, crustal and mantle). The latter is frequently investigated in petroleum systems to obtain information on its interaction with the hydrogeological regime or in monitoring deep-derived fluids typically upwelling in volcanic and seismically active areas (Caracausi et al., 2015; Caracausi and Paternoster, 2015; Chiodini et al., 2004) to seek for potential relationships between fluids degassing and volcanic activity or seismogenic processes. Indeed structural-geological features (e.g., faults and the tectonic pattern distribution) affect fluid chemistry, as they allow and facilitate the mixing of groundwater of different origin. In natural waters H_2 and O_2, N_2, CO and CO_2, CH_4, along with noble gases (particularly He) are the main gaseous indicators analyzed both in chemical and isotopic methods.

2.3.1.1 Stable water isotopes and local meteoric water line

Isotopic analysis on water stable isotopes (atoms of the same element with different number of neutrons and mass) are important tools in geochemistry as traces of water origin, recharge areas, circulation paths, mixing or exchange processes, as well as for the study of the global and local water cycle (e.g., Hoffmann et al., 2000). The hydrogen (D) and oxygen (^{18}O) water stable isotopes have a larger number of neutrons lends and a larger atomic weight with their corresponding most abundant isotopes (^1H and ^{16}O, respectively) which induce measurable physical and chemical effects. During phase changes, such as evaporation and condensation, stable isotopes become enriched in one phase and depleted in the other (isotopic fractionation). Fractionation processes leading to a reservoir isotopic enrichment or depletion are quantified through isotope ratios expressed with respect to a standard. For general, atmospheric applications, the usual standard is the Vienna Standard Mean Ocean Water (V-SMOW). To quantify the stable isotope relative ratio between the analyzed samples and the standard, the delta (δ) notation ($\delta^{18}O$ in the Eq. num equation):

$$\delta^{18}O = \left(\frac{R_{sample} - R_{std}}{R_{std}}\right) x\ 1000 = \left(\frac{R_{sample}}{R_{std}} - 1\right) x\ 1000\ (‰) \qquad \text{Eq. 2. 58}$$

where R is the isotopic ratio of concentrations between the rare and abundant molecules for each element for the oxygen R = ^{18}O / ^{16}O and for the hydrogen R = D / ^1H respectively. For more specific studies on recharge areas, circulation path and about the water cycle, the standard starting water, is meteoric water. For this reason, when studying water stable isotopes, it is important to identify the linear relation between δD and $\delta^{18}O$ meteoric water line, first introduced by (Craig, 1961) for global meteoric waters in the form and defined as Global Meteoric Water Line (GMWL):

$$\delta D = 8\delta^{18}O + 10 \qquad \text{Eq. 2.59}$$

With the y-intercept 10 representing the deuterium excess (d-excess). The d-excess may vary locally because it is strictly related to the evaporation conditions. For this reason, regional meteoric water lines have been introduced (e.g., the Mediterranean Water Line (MWL): δD = 8δ^{18}O +22, Gat and Carmi, 1987). Given the

high small-scale variability of the evaporation condition and consequently, of the reference meteoric water lines, to increase the accuracy of detailed hydrogeochemical studies, local meteoric water lines (LMWL) are estimated. For the High Agri Valley, regional reference water lines are: (a) the Southern Italy Water Line (SIMWL: $\delta D = 6.7\delta^{18}O +5.7$, Longinelli and Selmo, 2003); (b) with the Northern Calabria Meteoric Water Line (NCMWL: $\delta D = 6.22\delta^{18}O +5.21$, Vespasiano et al., 2015).

2.3.1.2 Stable Carbon isotopic signatures

Carbon stable isotopes (^{12}C and ^{13}C), conversely to the unstable ones (e.g., ^{14}C), are not subjected to radioactive decay. Indeed, their abundance in natural fluids, is controlled by physical and biological processes such as equilibrium reactions and plants photosynthesis (Laughrey and Baldassare, 2003).Stable carbon isotope geochemistry aims to identify the carbon origin in natural fluids. Indeed, carbon occurs in a wide variety of compounds and is a widespread element in nature in both oxidized environments, such as sedimentary carbonate rocks or gaseous CO_2 to organic substances produced in highly reduced conditions in the biosphere. Given the differences in isotopic mass, different isotopes assume different behavior during natural processes such that, sedimentary carbonates and organic matter possess distinct stable carbon isotope compositions, related to two main reaction mechanisms (Hoefs, 1987): (1) isotope equilibrium exchange reactions in the inorganic carbon cycle (atmospheric CO2-dissolved bicarbonate-solid carbonate) leading to an enrichment of ^{13}C in carbonate rocks; (2) kinetic isotope effects during plants photosynthesis make ^{12}C the most abundant carbon isotope in organic matter. Similarly, to eq (delta notation water stable isotopes), the stable carbon isotopic composition is expressed by the delta notation:

$$\delta^{13}C = \left(\frac{R_{sample}}{R_{std}} - 1\right) x \ 1000 \ (‰) \qquad \text{Eq. 2.60}$$

With the carbon isotopic ratio (R) for both a given sample and the standard is given by R = $^{13}C/^{12}C$. For carbon isotopic composition, the international standard is the Peedee Belemnite (PDB) (Hoefs, 1987) for which have a $\delta^{13}C = 0$ ‰ have been estimated. From Eq. 2.61 follows that: (a) negative $\delta^{13}C$ values are typical of samples enriched in the lighter isotope (^{12}C) relative to the standard and are mainly ascribable to organic matter origin. Positive $\delta^{13}C$ values, represent samples enriched in the heavier isotope (^{13}C) when compared with a carbon origin mainly attributable to the inorganic carbon cycle. In carbon isotope geochemistry, a mainly used parameter is the $\delta^{13}C_{TDIC}$ of Total Dissolved Inorganic Carbon (TDIC). TDIC represents the average of the isotopic composition of the carbon species in solution, weighted on the respective contents of the inorganic carbon compounds: dissolved carbon dioxide (CO_{2aq}), bicarbonate ($HCO3^-$) and carbonate ions (CO_3^{2-}). Their relative concentrations are strictly pH-dependent and reflect the interactions between water, carbonate rocks and CO_2 from different sources. The $\delta^{13}C_{TDIC}$ represents an isotopic balance of the dissolved carbon species and is given by (Grassa et al., 2006):

$$\delta^{13}C_{TDIC} = \delta^{13}C_{CO2aq} * \chi_{CO2aq} + \delta^{13}C_{HCO3} * \chi_{HCO3} + \delta^{13}C_{CO3} * \chi_{CO3} \qquad \text{Eq. 2.61}$$

Where, χ is the molar fraction of each inorganic species in the analyzed sample. Carbon stable isotope geochemistry provides a powerful tool to discriminate natural gases from different sources (Schoell, 1980) and is widely used for determining microbial, thermogenic, and inorganic sources of hydrocarbon gases and CO_2 (Schoell, 1980; Hoefs, 1987). Carbon dioxide is one of the most abundant components in natural gases and earth degassing of deep CO2-rich fluids, commonly reported worldwide in seismically and tectonically active regions (e.g., Caracausi and Sulli, 2019; Chiodini et al., 2004; Italiano et al., 2009; Tamburello et al., 2018). Thus, carbon isotope methods are a key tool to gain information on the escape of natural fluids in tectonically active regions, to identify their sources and their migration patterns through the crust and reconstruct their possible relationships with volcanism and large earthquakes generation.

2.3.1.3 Noble gases geochemistry

The noble gases are one of the most powerful geochemical indicators of natural processes since their key role as a proxy of the behavior and transport of volatiles to the surface from deep levels within the Earth. Noble gases are helium (He), neon (Ne), argon (Ar), krypton (Kr), xenon (Xe), and radon (Rn) and are mutually characterized by an unreactive character due to their chemical structure with the outer electron orbitals filled. Isotopes of all 6 noble gases are produced to some degree by nuclear processes (e.g., ^{4}He is, produced by alpha decay from U and Th; ^{40}Ar is produced by electron capture decay of K). In this work, we will focus on noble gas geochemistry in crustal settings, where the main sources of noble gases are: (a) the atmosphere, as they are introduced into the crust dissolved in groundwater; (b) the mantle, during episodes of magmatic activity; (c) processes of radioactive decay in the crust. ^{36}Ar isotope is mainly atmosphere-derived and introduced in the subsurface trough the flowing groundwater, while mainly radiogenic Argon (^{40}Ar) is produced in the crust by radioactive decay of ^{40}K at a current rate estimated by (Ballentine and Burnard, 2002) of:

$$^{40}\text{Ar atoms g}^{-1}\text{ yr}^{-1} = 102.2 \text{ [K]} \qquad \text{Eq. 2.62}$$

With [K] concentration of the ^{40}K, in parts-per-million (ppm). The ^{3}He isotope has a mainly primordial origin related to Earth degassing of deep, mantle-derived fluids and the ^{4}He, which is produced in the crust by radioactive α-decay of ^{235}U,^{238}U and ^{232}Th with an estimated rated given by (Ballentine & Burnard 2002):

$$^4\text{He atoms g}^{-1}\text{ yr}^{-1} = (3.115 \text{ x } 10^6 + 1.272 \text{ x } 10^5) \text{ [U]} + 7.710 \text{ x } 10^5 \text{ [Th]} \qquad \text{Eq. 2.63}$$

Where [U] and [Th] are concentration in ppm of the ^{235,238}U and ^{232}Th, respectively. Another example are the Neon isotopes (^{21,22}Ne) which are principally produced by crust radioactive decay of radioelements ^{17,18}O, ^{19}F,^{22,23}Na, ^{24,25}Mg (Yatsevich and Honda, 1997), while the ^{20}Ne is mainly atmospheric-derived. Other significant contribution on the noble gases amount in the crust can be given by interplanetary dust particles (IDPs), cosmic ray interaction with the crust and anthropogenic sources. Noble gases geochemistry is a useful tool for the reconstruction of fluid behavior in different geological environments and to discriminate and quantify the contribution of different sources in it due to their inherent characteristics, such as: (a) chemical

inertness and low reactivity, (b) high mobility, (c) relatively low abundances in the solid Earth, and (d) specific isotopic distribution between different terrestrial reservoirs (e.g., mantle, crust, and atmosphere). The latter characteristic is related to Early Earth's processes, such as differentiation into mantle and continental crust, degassing and atmosphere loss, which have developed different reservoirs with a distinct signature of noble gases abundances and isotopic compositions. Therefore, a full understanding of the processes that control the chemical signature of the noble gases in different crustal environments is strongly required for the development of accurate geochemical models. Additionally, noble gas isotope ratios are highly sensitive to trace chemical fractionation processes (i.e., the relative partitioning of the isotopes between two natural systems) so that in multiple-sourced natural fluids, fractionation patterns can provide useful constrains on the timing of geochemical processes occurred before and/or after the fluids mixing. The $^3He/^4He$ isotopic ratio is considered one of the most efficient geochemical tracers. Its variations are related to magmatic/crustal dynamics and therefore is of primary importance in tectonically active areas for volcanic forecasting or for monitoring deep-fluids role in seismic nucleation. Helium isotope geochemistry is a reliable tool to discriminate the fluid source within the Earth due to the specific origin of its two isotopes (the mantle-derived ^3He, the crustal radiogenic derived ^4He). The $^3He/^4He$ production ratio in the crust is in the range of 0.4 to 1 x 10^{-8} (Ballentine and Burnard, 2002), while the atmospheric $^3He/^4He$ ratio, which reflects continued release of primordial ^3He from the Earth, has typically much higher values estimated in the order of 1.384 x 10^{-6}. Indeed, its mass is sufficiently low that it can be accelerated in atmosphere and, since ^3He has much lower mass of the ^4He, it is lost even more easily than ^{4}He. The uniform atmospheric isotope ratio (R/Ra) provides a reference value for He isotope ratios in relation to atmospheric values and is expressed as:

Eq. 2.64

$$\frac{R}{R_A} = \frac{\left(\frac{^3He}{^4He}\right)_{sample}}{\left(\frac{^3He}{^4He}\right)_{atmosphere}}$$

R/R$_A$ ratios crustal values typically range about 0.1 (Ballentine et al., 2002), while in regions of continental extension and magmatism R/RA is usually above 1 indicative of higher He concentrations in crustal fluids (e.g., groundwater, natural gas) due to a flux of He from the mantle.

The summarized methods were applied in this work for the assessment of High Agri Valley (Chapter 3) and Mefite d'Ansanto (Chapter 4) fluids. Both chemical and isotopic (^{18}O and ^2H, $^{13}\delta_{TDIC}$, $^3He/^4He$) analysis were applied on liquid and gaseous matrix of 39 selected samples between natural springs, wells, and piezometers collected in the HAV area. While, 3 free gas samples were chemically and isotopically analyzed at the Mefite d'Ansanto study area.

2.3.2 Water and gas sampling

All water samples were collected in high-density polyethylene (HDPE) bottles in different aliquots (50, 100, 1000 ml) with watertight caps, after priming each bottle at least twice with sampled water. Untreated,

unacidified samples were collected for anions, δD and δ^{18}O analysis. Stabilized samples for cations and trace elements analysis were filtered through a 0.45μm membrane (MF-Millipore filter) and acidified to pH of ~ 2 with suprapure HNO_3 (67-69%). All collected sampled were stored in a refrigerated container (with a temperature of ~ 4°C) during transportation and kept cool until analysis. Dissolved gas samples were collected in 250 ml glass bottles provided with steel caps with rubber seals. A plastic funnel was used to collect free gas samples a the Tramutola sampling site, with a of three-components sampling system made by a funnel, a glass sampler, and a syringe, mutually connected by silicone tubes through a three-way pyrex valve (Figure 2.9). Indeed, employing a funnel allowed to channel the free gas towards a water-filled container used to check the effective functioning of the sampling system through the bubbling produced by the gas. After following the cleaning procedure to purge the glass bottle of air – consisting of slowly withdrawn the gas by the syringe and pushed it hardly into the glass sampler for several times producing bubbling in the water container - the gas was conveyed along the silicon tube into the glass sampler and carefully stored.
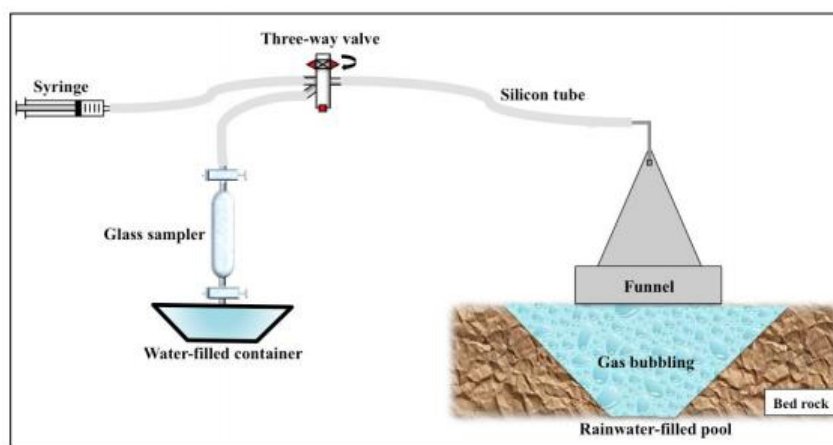
Figure 2. 9 – Scheme of the three-components sampling system (funnel, glass sampler, syringe) for free gases sampling through (from Bordenca,2019).

2.3.3 Physico – chemical parameters

Water physico-chemical parameters (temperature, pH, Eh, conductivity, and dissolved oxygen) were analyzed in situ through an high-resolution multiparametric control unit (WTW Multi 3630 IDS) (Figure 2.10). The total alkalinity was also determined in situ by titration with HCl (0.1 - 0.10 N) using a methyl-orange indicator.

2.3.3.1 Temperature

The temperature parameter has a marked influence on physico-chemical and organoleptic characteristics of waters such as: species solubility, pH variation, decrease of O_2 solubility and consequent CO_2 loss. An increase in temperature causes an increase in electrical conductivity or in the growth of microorganisms due to faster formation of organ halogen compounds in chlorine-treated waters. Temperature measures were carried out through an electric probe with a ± 0.1 °C resolution thermometer.

Figure 2. 10 - High-resolution multiparametric control unit (WTW Multi 3630 IDS) for in situ water physico-chemical parameters (pH, Eh, conductivity, and dissolved oxygen) meaurements.

2.3.3.2 pH

The pH is a logarithmic scale which measures the acidity of aqueous solutions. It is given by the negative of the base-ten logarithm of the molar concentrations of hydrogen ions in the solution:

$$pH = -log_{10}[H^+] \qquad \text{Eq. 2.65}$$

pH value in natural waters is a key indicator, due to its influence in many chemical and biological processes such as macronutrients (P, N, C), micronutrients e heavy metals (Pb, Cu, Cd, etc.) solubility. The pH of the analyzed waters was determined with a ±0.001 pH unit resolution WTW-IDS SenTix® 940 ph-meter through a potentiometric method: the instrument measures the difference in potential between the measuring electrode and a reference electrode with a known potential. The probe is a glass electrode made by a spherical glass membrane, permeable to H^+ ions which contains the reference electrode immerse hydrochloric acid solution buffered at a known pH using a saline bridge. The measured difference in potential is a function of the difference between the concentration of H^+ in the internal solution and in the external one. Following the sampling protocols, two-point calibration analysis (with pH 4 and pH 7 buffers) were performed on the instrument before and during each sampling campaign.

2.3.3.3 Electric Conductivity

Electric conductivity (EC) is a measure of the water capacity to conduct electric alternating current. Natural waters conductivity is mainly related to charged ions dissolved in it so that EC is proportional to electrolytes concentration in the solution. EC is particularly sensitive to temperature variations so that we generally report the EC (measured in μS/cm) referenced to a 25 °C temperatures. EC in situ measurements were addressed through a ±0.1 resolution μS/cm WTW IDS TetraCon® 925 conductivity CELL. We set the instrument on the reference temperature of 25 ° C, correcting it to cell constant verified trough a standard Crison KCl solution at concentration of 0.0100 ± 0.0005 M with a specific conductivity of 1413 μS/cm. Very pure waters (with low

ionic content) are characterized by low electrical conductivity, while groundwaters EC typically ranges between 100 and 1000 μS/cm.

2.3.3.4 Redox potential and Dissolved Oxygen

The dissolved Oxygen (O_2) and the redox potential (ORP or Eh) - which measure the tendency of a chemical species to acquire or lose electrons - both reflect redox conditions of groundwaters. The ORP measurement is generally performed with inert metal electrodes (e. g. platinum) with low resistance and high electron exchangeability (i. e. easily give up electrons to an oxidant or accept electrons from a reductant). The ORP electrode will continue exchange electrons with the analyzed solution until it develops a potential equal to ORP of the solution. The Eh of water samples was measured through ±1 mV with a WTW-IDS SenTix® ORP-T 900 probe after sensor calibration with the ORP Standard for Redox/ORP Electrodes. Dissolved oxygen concentration was measured through a ±0.01 mg/l resolution WTW Optical IDS dissolved oxygen sensor FDO® 925. An air calibration was performed in air-saturate water to calibrate the probe.

2.3.3.5 Alkalinity

Total water alkalinity is the capacity of water to resist acidification and it is represented by the total molar number of bases in solution that can be converted to uncharged species by a strong acid. In this work, natural waters with typical pH < 8.5 values were analyzed, which alkalinity is mainly represented by total concentration of bicarbonates $[HCO3^-]$ in solution. For pH values > 8.5 in the analysis of the total water alkalinity also carbonate and hydroxides should be considered and, if present, borates, phosphates, silicate, ammonia can contribute for the alkalinity. For natural waters, carbonates and bicarbonates have a key role as alkalinity indicators and their equilibrium in solution along with the carbon dioxide content – related to the formation of the carbonic acid – is highly dependent from water pH (e. g. a pH = 7 contains 20 % carbonic acid and 80 % of bicarbonates, without any carbonate in solution). We determined the total alkalinity of the samples by acid-base carbonates titration, using previously standardized 0.1 – 0.10 N hydrochloric acid with methyl orange pH indicator. We added HCl to the solution dosed through a 50 ml burette with $\pm$ 0.05 ml resolution. The bicarbonate present in the solution reacted with HCl with the for–ation of carbonic acid according to the reaction:

$$HCO_{3\,(aq)}^- \; + \; H_{(aq)}^+ \; \rightarrow \; \mathrm{H_2CO_{3\,(aq)}}$$
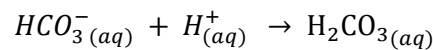
Bicarbonate concentration ($N_{HCO_3^-}$) was estimated from the volume of the HCl needed to reach the titration equivalent point, which corresponds to the complete transformation of bicarbonates in carbonic acid (at pH~4.3) through the following expression:

$$N_{HCO_3^-} \; (eq/L) \; = \; (N_{HCl} \times V_{HCl})/V_{sample} \qquad\qquad \text{Eq. 2.66}$$

*W*here N_{HCl} concentration of the HCl used for the titration analysis, V_{HCl} total volume amount (ml) of dosed HCl to reach titration equivalent point, V_{sample} volume amount (ml) of the analyzed sample.

2.4 Machine learning techniques

2.4.1 Introduction

During the last decades, geosciences were marked by a huge increase of data collection together with the evolution of scientific programming open-source frameworks (e.g. Python) used to process and store large dataset efficiently. With the spreading of monitoring networks and high sampling frequency data fluxes, automatic methods for the analysis of geological and geophysical phenomenon became essential. In this context, Machine Learning (ML) became a powerful tool to explore data patterns and use it to predict future data or to perform other kind of decisions on it (Murphy, 2012). (Karpatne et al., 2017) summarize many geoscience applications for which ML classification and pattern recognition methods are highly suitable such as: (1) characterize objects such as earthquake swarms, foreshocks or aftershocks sequences; (2) estimating geoscience variables difficult to monitor directly (e.g. retrieve surface water availability from ML analysis of remote sensing data); (3) producing long-term forecasts of geological variables through time-series regression ML is becoming increasingly widespread in a data intensive field such as seismology (Kong et al., 2019) where a wide range of automated, machine-learning based methods have been developed for earthquake detection and phase picking (Perol et al., 2018; Zhu and Beroza, 2018), earthquake early warning (Kong et al., 2016) and ground-motion prediction (Alavi and Gandomi, 2011). ML is an interdisciplinary field based on statistics, probability theory, computational complexity theory and other computer science techniques. It is a branch of the broad field of Artificial Intelligence, which aims to make machines simulate intelligent abilities of humans, such as "learning". For ML it is intended in broad sense as an inductive inference (Ratsch, 2004) where machine learns from data (called training samples), which represent incomplete information about some "statistical phenomenon". Specifically, learning is achieved by a set of prediction-based algorithms that extract information from the input data using well-defined optimization rules (Kong et al., 2019). Therefore, the purpose of a machine learning algorithms is to identify (to learn) a function f: $X \rightarrow Y$ that maps the input dataset (X) into an output domain Y (of possible predictions).

Machine learning algorithms are subdivided in two principal classes whether the desired output data have labels (e. g. classification problems) or not (e. g. clustering problems) (Figure 2.11): (1) Unsupervised learning, which typically tries to uncover dataset structures or to detect anomalies in the input unlabeled data. It comprehends clustering and dimensionality reduction algorithms, which group data into categories based on similarity and reduce the input data dimensions, respectively. (2) Supervised machine learning, which comprises predictive modeling algorithms given labeled input data (training data) it learns a function (model) that predicts an appropriate outcome of the given dataset. Supervised algorithms are subdivided into classification and regression algorithms based on whether the expected outputs are categorical (classification) or quantitative (regression).

In this work, both unsupervised DBSCAN algorithm (Density-Based Spatial Clustering of Applications with Noise, Ester et al., 1996.; Schubert et al., 2017) and supervised KNN classification (K-Nearest Neighbors, Cover and Hart, 1967) were implemented through Python-scikit-learn machine learning module (Pedregosa et al., 2011) and applied to time-domain parameters (duration, RMS-amplitude and arrival slope) of the automatedly collected seismic dataset at Mefite d'Ansanto study area. Moreover, a deep-neural-network-based

method (PhaseNet; , Zhu and Beroza, 2018) was applied to automatically pick P- and S-wave arrival times of a natural earthquake sequence occurred close to the municipality of Castelsaraceno (PZ), in the southeastern part of the High Agri Valley, in August 2020. In the next paragraphs, a brief description of the applied machine-learning methods will be given.
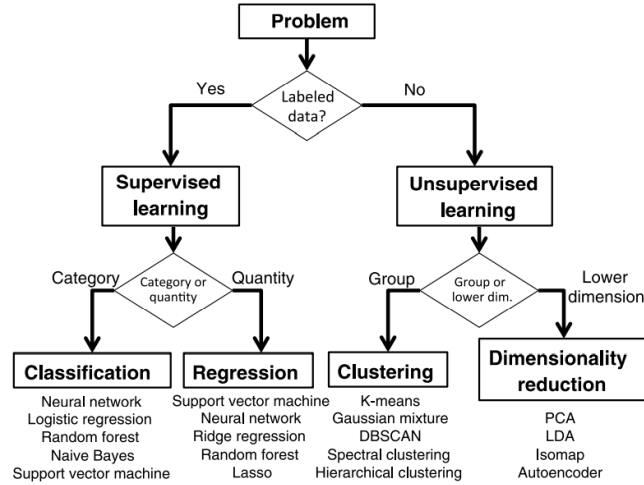
Figure 2. 11- Summary of the main types of "learning" and of most used machine learning algorithms (Kong et al., 2019).

2.4.2 DBSCAN Clustering

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based data clustering algorithm which finds a certain number of clusters in the dataset starting from the estimated density distribution of corresponding nodes. For this reason, DBSCAN can find clusters of any shape. The algorithm requires two hyperparameters - whose value is used to control the learning process- to define a "density-based" cluster: (1) the distance within which to search for close points (ε); (2) the minimum number of samples to create a cluster (n). Indeed, it is based on the notion of "ε-neighborhood" and "density-reachability" (Ansari et al., 2011). Given a set of m data points X = {xi | i = 1 … m}, where each data point is a n-dimensional vector and two data points x_p and x_c both $\in$ X, the ε-neighborhood of x_p is defined as:

$$N_\varepsilon\,(x_p)\;=\;\left\{x_c\;\in\;X\mid d^2(x_p,x_c)\;\leq \varepsilon\right\}$$

Eq. 2.67

re ε the maximum distance which defines the "neighborhood" and d is Euclidean distance between the two data points $x_p \in$ X and $x_c \in$ X is given by:

$$d^2(x_p,x_c)\;=\;\left\|\sum_{k\,=\,1}^{n} x_k^p\,-\,x_k^q\right\|^2$$

Eq. 2.68

n is the minimum number of points required to form a cluster, the point x_c is defined density-reachable from a point x_p if: (1) x_c is part of ε-neighborhood of x_p and (2) if the number of points in the ε-neighborhood of x_p are greater than or equal to n:

$$x_c \in X \qquad\qquad\qquad \text{Eq. 2.69}$$

$$|N_\varepsilon (x_p)| \geq \text{n}$$

DBSCAN would define a cluster with all data points mutually density-connected (named core samples) and, if a data point is density-connected to any data point of the cluster, it would be part of the cluster as well. Two points x_p and x_c are said to be density-connected if there is a point x_0 such that xo and x_p as well as x_0 and x_c are density-reachable. The algorithm starts by randomly selecting a starting data point x_i. If the ε-neighborhood of x_i contains sufficiently many (n) points, a new cluster is set. Conversely, x_i is labeled as noise (outlier). Later this point could be found in the ε-neighborhood of a different data point and would be assigned to a cluster. If a data point is found to be part of a cluster, all the data points in its ε-neighborhood will be added to that cluster and the process continues until the cluster is completely found. Then, a new unlabeled point is selected, and the algorithm will follow the previous steps looking for new cluster or assigning it to the noise. The n hyperparameter primarily controls how tolerant the algorithm is towards noise and should be increase for large and noisy datasets; while the ε parameters controls the local neighborhood of the points so that an appropriate data-driven value should be selected: if a too small ε is chosen most data will not be clustered at all; conversely, if chosen too large, it causes close clusters to be merged.

2.4.3 KNN Classification

The k-Nearest Neighbor (KNN) is a widely used supervised learning algorithm due to its simple and versatile nature. It can be used to solve both data regression and classification problems, but in this work the k-Nearest Neighbor classifier was applied. KNN is a non-parametric algorithm, which means it does not make any assumption on underlying data distribution. It is also defined a lazy learner algorithm because it does not learn from the training set immediately, it stores the dataset and performs the classification only at the final time of decision, instead. The latter characteristics make its application particularly efficient in classification problems where decision boundaries are irregular, as the one encountered in this work (see Chapter 4). As a supervised machine learning algorithm, KNN relies on labeled input data (training data, T) to learn a function (model) that predicts an appropriate classification of the input unlabeled data. KNN classification is based on features similarity, measured as the distance from a number K of labelled points, named as Nearest Neighbors. Different kinds of distance (e. g. Euclidean, Manhattan, Hamming) can be applied as similarity parameter. In its simplest general scheme, the algorithm consist of the following steps: (1) given a training set (T) - which stores labelled training samples (e.g. Class A, Class B in Figure 12a) – and a test object y (the yellow square in Figure 2.12), to be classified with a class label, the algorithm computes the distance between y and all the training objects (Figure 2.12b) to determine its nearest neighbors (k) (Figure 2.12c) . It then classifies y assigning it to the class of the bulk neighboring objects (Class B will be assigned in the example Figure 2.12c) assigning it uniform weights.
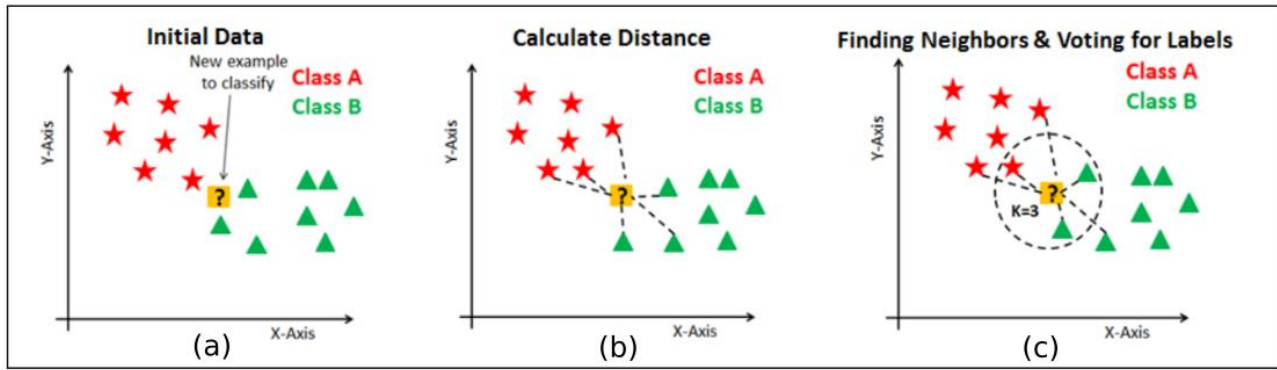
Figure 2. 12 – k-Nearest Neighbor (KNN) Algorithm basic workflow:(a) reads the input unlabeled object (yellow square) which is a vector of attribute values with an unknown class label; (b) computes the distance between the object and all the training labelled (Class A, Class B) objects to determine the requested number of nearest neighbors (k=3 in the example); (c) classifies the object as signing to it the same class of the bulk neighboring objects (modified from: https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn).

K is the core deciding factor (hyperparameter) of the algorithm and its tuning is a key point for the KKN classification performance: if k is too small, noise points have higher influence on the results; if k is too large, then the neighborhood may include too many points from other classes. It must be an odd number for an even number of classes (Kim, 2012). An estimate of the best k can be obtained by cross-validation of the results or by performing a classification model accuracy (the number of correct predictions over the total number of predictions) test for different k values.

2.4.4 Neural networks

Neural networks, also known as artificial neural networks (ANN) are a subset of machine learning algorithms. Their name and structure were inspired by the human nervous system, mimicking the interconnected neurons creating a complex network of signal transmission into the brain. Each cell collects inputs from all other neural cells it is connected to, and if it reaches a certain threshold, it sends the signal to all the interconnected cells. Machine learning neural networks simulate this behavior and its basic unit (neuron) is called "perceptron". The perceptron takes several weighted inputs, summarize them, and if the combined input exceeds a threshold, it will activate and send an output to the surrounding ones. The type of output it sends is determined by an "activation function" (or transfer function). When designing the network, choosing the best activation functions requires some trial and error. There are many kinds of transfer functions, mainly subdivided in Linear and Non-linear Activation Functions. Some of the most common activation functions are sigmoid, hyperbolic tangent, Rectified Linear Unit (ReLU) function and the Softmax activation functions (Pedrycz and Chen, 2020). The equation for a perceptron can be written as:

$$y = \phi \left(\sum_{i=1}^{n} w_i x_i + b \right)$$

Eq. 2.70

Where y is the output signal, φ is the activation function, n is the number of connections to the perceptron, w_i is the weight associated with the ith connection and x_i is the value of the ith connection and b represents the

threshold. A graphical representation can be found in Figure 2.13 where threshold b is a neuron with a constant value of −1. A dynamic threshold can be achieved allowing the network to modify the weight associated to b. In neural networks architectures, perceptrons are organized in layers (node layers) which comprise an input layer, one or more hidden layers, and an output layer (Figure 2.13).
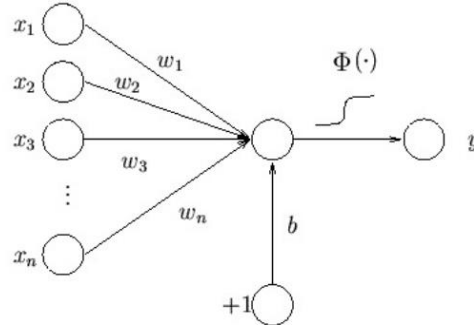
Figure 2. 13 - Graphical representation of a simple perceptron where y is the output signal, Φ is the activation function, n is the number of connections to the perceptron, wi is the weight associated with the ith connection and xi is the value of the ith connection. The b in the figure represents the threshold.

If the output (y) of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network. The learning algorithm of a neural network can either be supervised or unsupervised. Neural networks are often used for classification problems where the algorithm must be able to learn from examples and adapt. In an ANN, this is achieved by updating the weights associated with the connections between the layers. There are several ways of doing this, and one of the most applied is the backpropagation. The method aims to train the network by minimizing the loss as calculated by the cost function. The cost function represents how far off the network is from making accurate predictions based on the input. There are many different types of cost functions which may be used, one of which is the mean squared error:

$$C \ = \ \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Eq. 2.71

where n is the number of data points. The mean squared error represent the average of the square of the difference between the expected output (y_i) and the actual output ($\hat{y}_i$) which was obtained by the network. This process is then used to update the weights, and by repeated use of this process, the network can learn to distinguish between several different classes. One problem with the ANN approach is over-fitting of the data, which happens when the classifier becomes too good at recognizing the training examples, at the expense of not being able to recognize a general input. This can be avoided by cross-validation, where the network is trained on one set of data, and then evaluated on a separate one. When the error starts rising in the validation set, the network might be over-fitted. If previous networks are saved, the network can then be rolled back to the one which gave the smallest error.
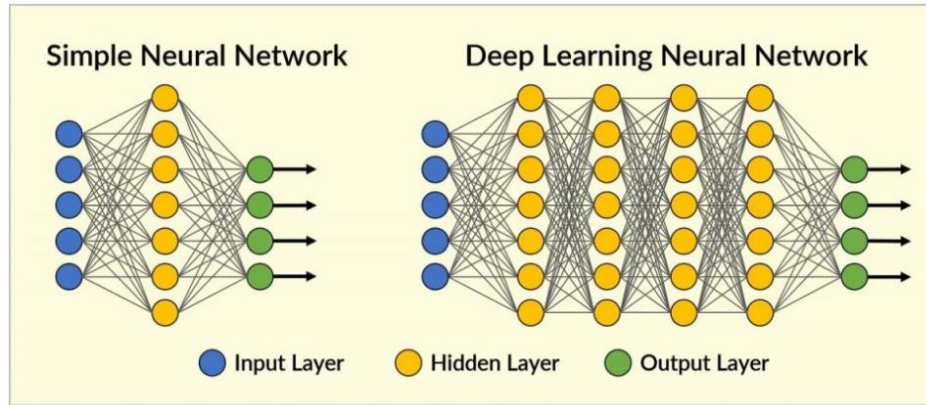
Figure 2. 14 – Schematized structures of a simple Neural Network Algorithm, characterized by one hidden layer and a Deep Learning Neural Network with a more complex structure with at least two hidden layers (from: https://vitolavecchia.altervista.org/differenza-tra-deep-learning-e-rete-neurale/).

2.4.4.1 Deep neural networks

In this work, a deep neural network (DNN) architecture (PhaseNet, Zhu and Beroza, 2018) was applied for seismic phase arrival time automatic picking of the Castelsaraceno natural sequence (Chapter 3). DNN is one type of the broad subclass of Machine Learning algorithms called Deep Learning. Deep learning peculiarity is its higher level of complexity with respect to the previously described algorithm structures. More specifically a deep neural network algorithm differs from a Neural Network (shallow learning algorithm) for multiple layers between the input and output node layers (Figure 2.14). Inspired by deep hierarchical structures of human speech perception and production systems, deep learning techniques received increasing research interests because of their capability of overcoming the drawback of hand-designed features of traditional machine-learning algorithms. Their structures allow the algorithm to understand data in a hierarchical way, with each concept defined in terms of its relation to simpler concepts, making the machine capable to learn from experience, avoiding the need for human operators to formally specify the knowledge (data representation) that the machine needs (Ronneberger et al., 2015, Goodfellow et al., 2016). For this reason, deep learning algorithms are widely used for big data analysis with successful applications to computer vision, recommendation systems, speech recognition, natural language processing, pattern recognition. Deep neural network can be subdivided in three main categories: Convolutional Neural Networks, Pretrained Unsupervised Networks, and Recurrent/Recursive Neural Networks (Pedrycz & Chen, 2020).

The PhaseNet algorithm applied in this work has a Convolutional Network architecture, modified from U-Net (Ronneberger et al. 2015), a deep neural network used in biomedical image processing that seeks to localize properties in an image. The neural network was trained on a large dataset of known earthquake waveforms where P- and S-wave arrival times were picked by experienced analysts. Instead of using manually defined features, it learns the features from labelled data (noise and signal). The algorithm first obtains P-wave, S-wave, and noise probability distributions on the input unfiltered three-component seismic waveforms. Then, predicts P- and S-wave arrival times corresponding to the peaks the P-wave and S-wave probability distributions, respectively. Zhu & Beroza (2018) demonstrated the high accuracy of the PhaseNet deep neural

network as automated phase picking tool with superior performance to standard earthquake monitoring and phase-picking tools (e.g., STA/LTA methods).

Chapter 3 – High Agri Valley study area

3.1 Geological setting and seismicity

The High Agri Valley (HAV hereafter) is located in the northern sector of the Agri Valley (Basilicata Region, Southern Italy). From the geological point of view, it is a NW-SE trending quaternary basin located in the axial zone of the Southern Apennines, an east-verging fold-and-thrust belt developed as an accretionary wedge due to the eastward migration of the compressional tectonics in the Apenninic arc (late Oligocene-early Pleistocene) (Cello et al., 2000; Giano, 2011; Giano et al., 2000; Gueguen et al., 2015). Since the Middle Pleistocene, the onset of NE-SW extensional tectonics in the area controlled the formation and the evolution of the High Agri Valley up to present (Schiattarella et al., 2003). This tectonic regime is still active as witnessed by NW striking high-angle normal and oblique faults bordering the basin, which represent the current main seismogenic structures in the area (Giocoli et al., 2015; Serlenga and Stabile, 2019; Stabile et al., 2014b).

The High Agri Valley geological setting is characterized by quaternary continental deposits overlying a pre-Quaternary substratum. The latter is made by tectonic units overlapped during the middle to late Miocene compressional stage, when the development of major low-angle crustal thrusts accommodated the stacking of the "allochthonous" Apenninic units onto the foreland carbonates sequence of the Apulian Platform. The Appeninic units represent the High Agri Valley outcropping substratum, while the deep Meso-Cenozoic Apulian succession have been only detected in the subsurface during oil exploration in the area. The northeastward propagation of the compressive regime during the formation of the southern Apennines is recorded by a series of thrust-sheet top basins which become progressively more recent toward NE, according to the emplacement age of the Apenninic units (Prosser et al., 2021). The geometrically highest unit is represented by the Liguride complex, consisting of minor fragments of oceanic crust of the Ligurian basin and their related sedimentary cover of Mesozoic to Cenozoic age (Albidona Flysch and Saraceno Formation). This complex derives from NW-dipping subduction of the Ligurian Tethys a late Cretaceous–early Miocene (Knott, 1987). All the underlying tectonic units belong to the paleomargin of the African plate, represented by the Apennine Platform carbonates (Triassic – Cenozoic) tectonically superimposed onto the deep-water deposits of the Lagonegro Basin (Triassic – Miocene). The HAV outcropping substratum comprises the Mesozoic to Cenozoic shallow-water and slope carbonates of the Campania–Lucania platform (Monti della Maddalena unit, D'Argenio et al., 1973) to the SW, thrusted over the pelagic successions (Paleozoic/Triassic to Tertiary Lagonegro units) which largely floor the northeastern part of the valley. Tertiary siliciclastic sediments (Albidona and Gorgoglione flysch Formations) represent the substratum in the SE area of the basin (Carbone et al., 1991). A Quaternary synorogenic succession fills the High Agri Valley basin. It consists of Lower Pleistocene-Holocene continental mainly coarse-grained clastic succession characterized by extreme vertical and lateral variability; this is due to its development in different depocenters originated during the complex extensional-transtensional tectonics (Rizzo et al., 2004; Zembo, 2010). In the study area, the quaternary succession consists of: (i) the lower-middle Pleistocene "Brecce di Galaino" Formation, mainly outcropping in the north-eastern area (Di Niro and Giano, 1995); (ii) the middle-late Pleistocene "Complesso Val d'Agri"

Group (Giano et al., 2000), which largely crops out in the Pertusillo area; (iii) the uppermost Pleistocene - Holocene alluvial and colluvial deposits.

The middle Pleistocene regional extensional tectonics involves the axial zone of the Southern Apennines locally controlling the formation and the evolution of intramountain basins (e.g., Agri Valley). Indeed, the HAV is bounded by two NW-SE striking and oppositely dipping fault systems: the transtensional Eastern Agri Fault System to the northeast (EAFS, Cello et al., 2003) and the Monti della Maddalena extensional Fault System to the southwest (MMFS, (Maschio et al., 2005; Improta et al., 2017) (Figure 3.2). The extensional regime is still active at relatively low rates (2–5 mm/yr; (Ferranti et al., 2014) and it is responsible for most of the present-day seismicity in southern Apennines (Chiarabba et al., 2005). The issue about which of the two extensional fault systems has the main seismogenic potential in the area is still debated. Some authors indicate the MMFS as the main seismogenic structure of the region based on field surveys, paleoseismic trenches, near-surface geophysical prospecting, and GPS analysis (Burrato and Valensise, 2008; Ferranti et al., 2014; Improta et al., 2010; Maschio et al., 2005; Valensise and Pantosti, 2001). Conversely, other authors ascribe most of the natural seismicity nucleation to the EAFS (Benedetti et al., 1998; Cello and Mazzoli, 1998 Borraccini et al., 2002; Cello et al., 2003 ; Barchi et al., 2007). The natural seismicity of the HAV currently consist of very sparse, low-magnitude activity (ML < 3.2) (Serlenga and Stabile, 2019). However, looking at its historical seismicity (e.g., CPTI11 catalogue, Rovida et al., 2011) the area remains one of the most seismic hazardous areas in Italy, with seven historical events with Mw > 4.5 occurred in the last 200 years, including the 1857, Mw 7.0 Basilicata earthquake (Burrato and Valensise, 2008). Along with extension-related natural earthquakes, the present-day seismicity in the HAV is also characterized by induced seismicity from anthropogenic activities. Indeed, two different clusters of induced/triggered micro-earthquakes were recognized in the area to be caused by industrial operations (Figure 3.1):

(1) continued-reservoir-induced seismicity (Ml $\leq$ 2.7) related to the seasonal water level fluctuation of the artificial Pertusillo Lake (Stabile, 2015; Stabile et al., 2014a; Telesca et al., 2015; Valoroso et al., 2009; Vlček et al., 2018);

(2) fluid-induced microseismic swarms located at about 1.3 km SW from the Costa Molina 2 injection well and related to disposal operations of wastewater produced from the exploitation of the biggest onshore oil and gas field in western Europe here located (Improta et al., 2015, 2017; Stabile et al., 2014b; Wcisło et al., 2018).

The twofold nature of the recorded seismicity in the High Agri Valley makes it an ideal study area to deepen the understanding of driving processes of both natural and anthropogenic earthquakes, to improve the current methodologies for the discrimination between natural and induced seismic events and to apply advanced seismological methods to characterize its crustal structure. For this reason, many seismological studies were carried out in the HAV in the last decade starting from the analysis of earthquakes recorded by both national and local seismic networks deployed in the area (Stabile et al., 2020). These allowed to develop local earthquake tomographies (Valoroso et al., 2011; Improta et al., 2017; Serlenga and Stabile, 2019) earthquake relocations, statistical analyses on the spatio-temporal evolution of seismicity (Valoroso et al., 2009; Stabile

et al. 2014a, 2014b; Improta et al. 2015; Stabile et al. 2015; Telesca et al. 2015; Improta et al., 2017; Buttinelli et al., 2016; Stabile et al., 2021) and joint Vp/Vs and attenuation studies (Wcisło et al., 2018).
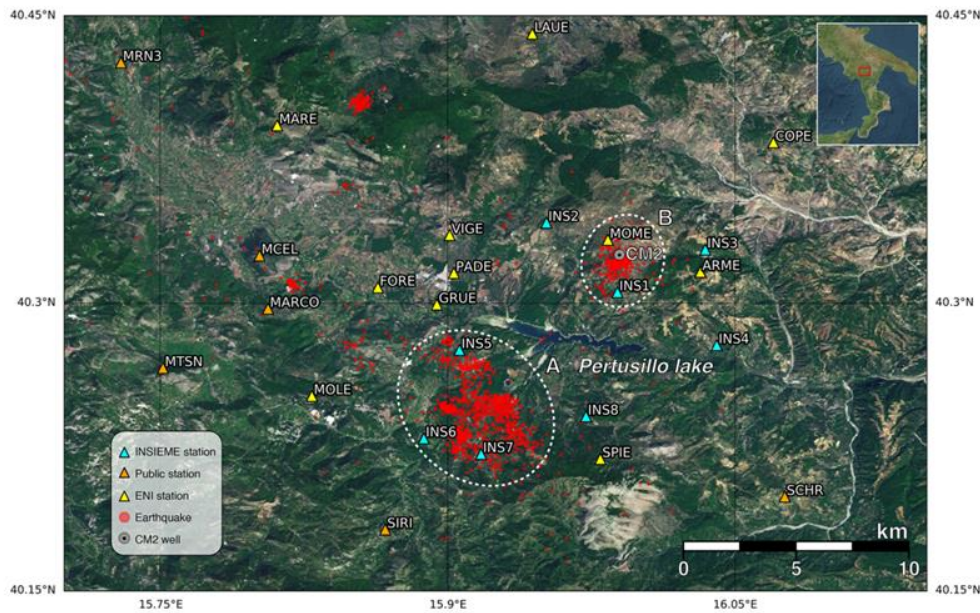
Figure 3. 1 – Induced seismicity clusters in the High Agri Valley and the INSIEME Network deployment (cyan triangles) (from: Stabile et al., 2020)

3.2 Hydrogeological setting

The High Agri Valley is characterized by a high groundwater content. It is favored by the climatic features of the area: according to the Köppen classification, it is a warm-summer Mediterranean type climate area, characterized by cold-humid winter and hot-dry summer. Here, the mean annual precipitation and temperature values, estimated in the years between 2005-2015, were of about 1000 mm/y and 12.5° C respectively with January as the coldest month with an average temperature ranging between 3° C and 4° C. Hot and arid conditions during summer produce periods of drought during July and August. The climate of the area, characterized by high seasonal precipitation rates in autumn and spring and a snowy winter, allows the greatest amount of the aquifer recharge. Furthermore, the fluvial flow in the valley, provide an important amount of surficial water supply to the underground circulation.

Two different kinds of aquifers have been recognized in the HAV (Colella, Lapenna and Rizzo, 2003): (i) fissured and/or karstified aquifers in the Apenninic units belonging to the tectonic substratum; (ii) porous multi-layer aquifers developed in the quaternary succession that fills the basin. Both types present high volumes of groundwater storage and circulation, with the richest amount located in pre-quaternary substratum aquifers. These are characterized by high fracturing indexes and karst permeability and are mainly located in the limestones belonging to the Campano-Lucana Platform and to the underlying Calcari con Selce Formation (Lagonegro Units). Porous aquifers have

been localized in gravelly-sandy permeable deposits belonging to the quaternary succession that fills the basin. These have been described in detail by previous studies (Colella, Colucci e Longhitano, 2003), which mainly recognized three hydrostratighraphic units: (a) a basal aquitard in the lacustrine clays and local gravels largely deposited in the south-central sector of the valley; (b) a multi-layered semi-confined aquifer, localized in the "Complesso dell'Agri" Formation; (c) local aquitards in the overlying Pleistocene-Holocene alluvial deposits. The multi-layered aquifers host the larger groundwater bodies and locally artesian aquifers located at depths of 60-70 m in gravelly-sandy-clayey alluvial-lacustrine deposits. The aquifer shows its maximum thickness between Grumento Nova – Serra Carlea and the Paterno – M. S. Nicola areas and it is dislocated by synsedimentary normal faults which locally contribute to the upwelling of mineralized water. Finally, local groundwater bodies have been localized in recent alluvial deposits at the top of the succession.

The High Agri Valley is also characterized by many natural springs located at both the left and the right banks of the the Agri River, mainly located in correspondence of the geological contact between permeable (limestone-dolomite) pre-quaternary rocks and impermeable clayey and flysch Plio-Pleistocene formations and recent alluvial-lacustrine deposits (Mongelli, Paternoster and Frittella, 2003). Springs are located at different altitudes above the sea level ranging from 1395 m (e.g., "Sorgente Acqua del Pastori", "Aqua del Tasso") to 560 m (e.g., "Sorgente Aggia" and "Fontana Capano"). Springs flow rate follows the seasonal fluctuation of precipitations such that it increases during winter and spring times, and it drops during the summer. In this work, the 35 main natural springs of the High Agri Valley were sampled which will be described in detail in the next paragraphs.
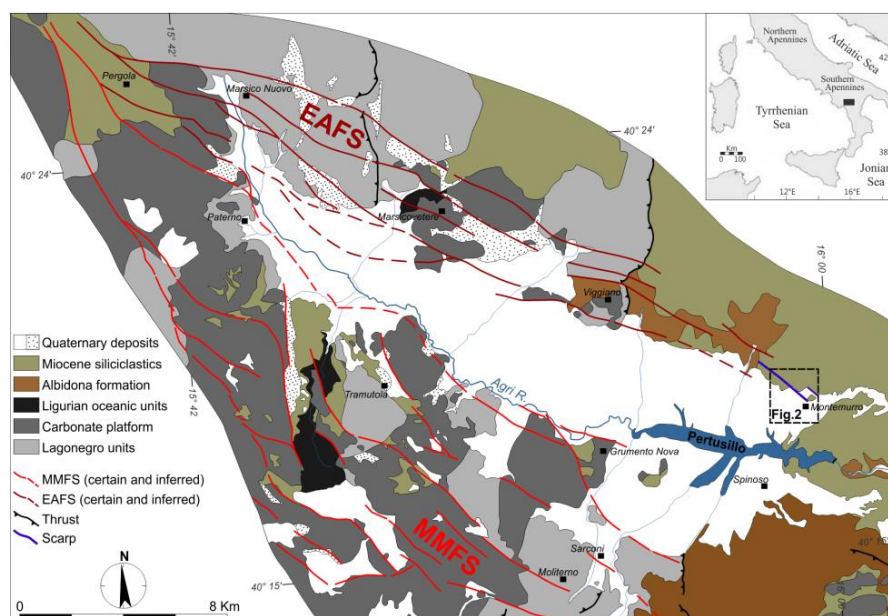
Figure 3. 2 – Geological sketch map of the High Agri Valley (from Giocoli et al., 2015)

3.3 Seismological methods

Induced seismicity has been documented since the 1920s in oil and gas exploitation (Goose Creek oil field (USA), Pratt and Johnson, 1926). In recent years, the phenomenon has drawn attention due to increasing number of reported cases occurred near industrial facilities carrying fluid injection operations (National Research Council, 2013; Ellsworth, 2013; Grigoli et al., 2017). Most of induced seismic events are classified as microseisms (M≤2). However, potential exists to produce earthquakes that can be felt by people and cause damage to buildings and infrastructures. In particular, many studies indicate wastewater disposal as the principal cause of high-magnitude induced seismic events (Ellsworth, 2013; Rubinstein et al., 2014), including the largest fluid-induced event occurred in the world (Mw 5.7 Prague, Oklahoma, USA, 2011) (Keranen et al., 2013). Considering the strong socioeconomic impact of induced seismicity (National Research Council, 2013; Ellsworth, 2013; Grigoli et al., 2017), the current seismological research in this field has two main implications (Stabile et al., 2021): (1) from a scientific point of view, for better understanding the complex interactions among rock, faults, and fluids involved in earthquake generation, and how perturbations of the stress field, even at small scale, may affect the stability of faults over time; (2) from a social and economic point of view, it is useful for the development of specific protocols for the management of fluid injection operations, the improvement of monitoring strategies and the correct implementation of traffic light systems, the definition of the associated hazard and risk, and the discrimination between natural and induced seismicity (Guidelines for Monitoring Seismicity, Ground Deformation and Pore Pressure in Subsurface Industrial Activities, MISE, 2014). To achieve this goals, the set-up of seismic monitoring network is essential to collect high-resolution data: (a) to detect, localize and determine the main parameters of the earthquakes occurring within the study areas; (b) to improve the magnitude completeness of seismicity at the local scale of the monitored areas; (c) to measure with high accuracy the ground accelerations produced at the surface by earthquakes; (d) to build a database of seismic events that can be used as a reference to monitor the spatio-temporal evolution of seismicity and to recognize possible periodic behavior of the induced seismicity in relation to specific operations parameters (e.g., fluid injection volumes and pressure). Efficient and high-quality earthquake detection and location in the area where human activities take place are operations of primary importance in seismic monitoring. The quality of earthquake catalogues depends critically on both the number and accuracy of arrival-time measurements, and to the application of the most suitable location methods, possibly in more reliable, local, 3D-velocity models. As a standard procedure, earthquake detection and arrival-time measurement, or phase picking, is often carried out by network analysts and the quality of both operations is strictly related to its expert judgment and experience. However, with the development of dense monitoring networks and high data fluxes, it is becoming increasingly difficult to manually keep up with the data analysis. For this reason, the development of semi-automated and automated workflows for earthquake detection, location and data processing has become a key challenge for the research to pursue, which would enable seismologists to manage seismic-monitoring operations more efficiently. Algorithms based on waveforms cross-correlation of continuous data streams with the selected master events for the automated detection of weak earthquakes (e.g., Template Matching Algorithm, Roberts et al., 1989; Stabile et al., 2021) have been developed for the improvement of earthquakes catalogues completeness. Moreover, decades of research have

been devoted to automatic phase picking, including methods based the short-term average/long-term average (STA/LTA) method (Allen 1978) and approaches based on higher-order statistic including kurtosis and skewness (Poiata et al., 2016). A recent powerful development was reached through machine-learning approaches, with the implementation of deep-neural-network-based arrival-time picking methods (e.g., PhaseNet, Zhu and Beroza, 2018).

The twofold nature of the recorded seismicity in the High Agri Valley (anthropogenic and natural) makes it an ideal study area to deepen the understanding of driving processes of both natural and anthropogenic earthquakes, and to improve the current methodologies for their discrimination and for seismicity monitoring in areas where industrial operations are carried out. On these grounds, in 2016 the dense INSIEME seismic network (Figure 3.1) was installed in the High Agri Valley (hereinafter HAV), designed, developed, and managed bythe CNR-IMAA for studying induced-seismicity, as well as to collect high-quality seismic data useful for the scientific community. In this work, we first applied a standard detection and location workflow on the collected dataset, which consisted in manual detection of earthquakes and their phase picking, followed by both linearized and probabilistic earthquake locations methods. The approach, despite quite time-consuming, allowed us to provide the High Agri Valley preliminary seismicity catalogue in the time-span September 2016 and March 2019 (https://doi.org/10.5281/zenodo.3632419). Moreover, we tested a semi-automatic workflow for the characterization of low-magnitude tectonic earthquake sequence which occurred on August 2020 in the HAV. Indeed, the study of micro-seismic sequences allowed us to gain insights on fault geometries and properties and it is especially important in high seismic hazard areas. The approach involved the integration of manual and advanced semi-automatic and automatic detection and location methods which enabled us to reveal a 10 km depth anti-apenninic structure, in a relatively short time with respect to the standard techniques, thus representing a starting point for a more efficient monitoring of both anthropogenic and natural seismicity in the High Agri Valley.

3.3.1 The INSIEME network

The INSIEME local seismic network was designed and developed by researchers of the CNR-IMAA in 2016 in the framework of the research project INSIEME (INduced Seismicity in Italy: Estimation, Monitoring, and sEismic risk mitigation, MIUR SIR - Scientific Independence of young Researchers (2015-2019), Stabile et al., 2020). The network has been registered at the International Federation of Digital Seismograph Networks (FDSN) with code 3F until the end of the INSIEME project (23 March 2019); then, it obtained the permanent network code VD (https://doi.org/10.7914/SN/VD, CNR IMAA Consiglio Nazionale delle Ricerche, 2019) as part of the High Agri Valley Geophysical Observatory (HAVO), a multi-parametric network managed by the CNR-IMAA research institute. The network was designed to achieve two main purposes: (a) study the seismic processes related to the occurrence of events belonging to the two induced seismicity clusters in the High Agri Valley and (b) to provide the scientific community with open-access, high-quality seismic data for studying such phenomena and for developing methodologies to discriminate between natural and anthropogenic earthquakes (Stabile et al., 2020). Indeed, the data collected during the INSIEME project (2016-2019) have been already released and open access at the FDSN web services and are available from IRIS DMC

(https://doi.org/10.7914/SN/3F_2016; Stabile and INSIEME Team, 2016). The INSIEME network is composed of eight stations deployed in an area of about 17 km×11 km. Since the main purpose of the network was to detect and to locate the induced microseismicity clusters of the HAV (Ml ≤ 2.7), the seismic stations were organized in two groups of four around each cluster. Serlenga and Stabile (2019) recognized that the anthropogenic clusters in the HAV are characterized by shallow events of about 4–5 km focal depth. Thus, to increase the network resolution, the epicentral distance of the closest station to each cluster have been chosen as less than the focal depth of its events (<4-5 km), and the average distance between stations does not exceed twice events depth (Havskov et al., 2012): the minimum distance between stations (cyan triangles in Figure 3.1) ranges between 2.7 km (INS6 and INS7 stations) and 5.4 km (INS1 and INS4). Each station was equipped with triaxial weak-motion broadband sensors: six 0.05–100 Hz and two 0.0083–100 Hz (INS1, INS7) Trillium Compact Posthole (TCPH) seismometers. To obtain high-quality, low signal-to-noise ratio seismic data, all the sensors were installed in non-toxic PVC (polyvinyl chloride) casings at different depths down to 50 m. The only exception is the INS8 station, which is located in the national park Val d'Agri – Lagonegrese area, subjected to environmental safeguard restrictions, and therefore installed on the surface. Continuous, seismic recordings are acquired at a 250 Hz sampling rate through Centaur digital dataloggers (Nanometrics Inc.) with a dynamic range of 140 dB. The choice of such high sampling frequency allows data acquisition with a Nyquist frequency of 125 Hz, greater than the upper frequency bound of the broadband sensors (100 Hz), thus avoiding the application of temporal anti-aliasing filters to the acquired signals and to record the full spectra content of small earthquakes. Table 3.1 summarizes the INSIEME network station's location, installation depth and sensor characteristics. Moreover, it specifies stations ID's before (network code 3F) and after (network code VD) the end of the INSIEME project. The latter will be used from now on in the rest of the chapter. The final network configuration is reported in Figure 3.1. More detailed information on the INSIEME Seismic Network is reported in Stabile et al., 2020.

Station name (3F)	Station name (VD)	Latitude (°N)	Longitude (°E)	Elevation (m a.s.l.)	Installation date	Installation location	Sensor	Sensor depth (m)
INS1	MONCM	40.305790	15.988603	802	2016-10-12	Borehole	120 s–100 Hz TCP	50
INS2	VIGGN	40.342090	15.951559	1043	2017-03-22	Borehole	20 s–100 Hz TCP	6
INS3	ARMEN	40.328033	16.034446	880	2017-03-22	Borehole	20 s–100 Hz TCP	6
INS4	MONTM	40.278168	16.040405	652	2017-03-22	Borehole	20 s–100 Hz TCP	6
INS5	GRUME	40.275704	15.906211	602	2016-10-13	Borehole	20 s–100 Hz TCP	6
INS6	SARSB	40.229581	15.887608	745	2017-03-22	Borehole	20 s–100 Hz TCP	6
INS7	SARCL	40.221487	15.917465	881	2017-03-23	Borehole	120 s–100 Hz TCP	14
INS8	SPINS	40.241083	15.972221	882	2017-03-02	Surface	20 s–100 Hz TCP	-

Table 3. 1 – Summary of the INSIEME network stations temporary (FDSN code: 3F) and permanent (FDSN code: VD) IDs, location, installation depth and sensor characteristics.

3.3.2 Standard procedure

The continuous data acquisition by the INSIEME seismic network allowed the manual detection and location of a total number of 852 local tectonic and induced earthquakes occurred in the High Agri Valley between September 2016 and March 2019. For the collected dataset, we applied a standard detection and location approach which involved (Figure 3.3): (1) the manual detection of seismic events through the available dedicated intranet system (WebObs, (Beauducel et al., 2020) (see Chapter 2.2.2), which detection information is automatically sent to the SeisComP3 database; (2) the manual pick of the P- and S- waves arrival times through SeisComP3 Origin Locator Viewer (scolv) Software; (3) the earthquake preliminary locations through the linearized absolute location method Hypo71 (Lee and Lahr, 1975) and using the 1-D velocity model developed for the High Agri Valley by Improta et al. (2017); (4) the earthquakes relocation with NonLinLoc code (Lomax et al., 2000) using the Oct-Tree grid search location method and the 3-D velocity model of the study area provided by Serlenga and Stabile (2019), which enabled more accurate locations. To obtain better locations of tectonic earthquakes occurring in the HAV outside the INSIEME network, a virtual seismic network made by a total of 27 seismic station was built, adding recordings of external stations: (a) 11 seismic stations of the Italian National Seismic Network (RSN) (FDSN codes: IV, https://doi.org/10.13127/SD/X0FXnH7QfY, INGV Seismological Data Centre, 2006; MN, https://doi.org/10.13127/SD/fBBBtDtd6q, MedNet Project Partner Institutions, 1990) and managed by Italian National Institute of Geophysics and Volcanology (INGV); (b) 7 stations belonging to the Irpinia Seismic Network (Stabile et al., 2013; Weber et al., 2007; FDSN code: IX), and (c) the MARCO station belonging to the GEOFON network (FDSN code: GE, https://doi.org/10.14470/TR560404, GEOFON Data Centre, 1993), installed south of Tramutola town. All the added stations of the virtual network (Table 3.2) are located within about 60 km distance from the center of the INSIEME network.

3.3.2.1 Preliminary locations

The results of the earthquake preliminary locations through the SeisComP3 Origin Locator Viewer (scolv) are open-access and available on the Zenodo repository as the "Preliminary catalogue of High Agri Valley seismicity (southern Italy) recorded by the temporary INSIEME network" CSV file (https://doi.org/10.5281/zenodo.3632419; Stabile et al., 2020). The catalogue consists of a total of 852 seismic events occurred in the HAV between September 2016 and March 2019. In it, we distinguished three main categories of earthquakes in the High Agri Valley, which were manually picked and located:

EARTHQUAKE LOCATION STANDARD WORKFLOW
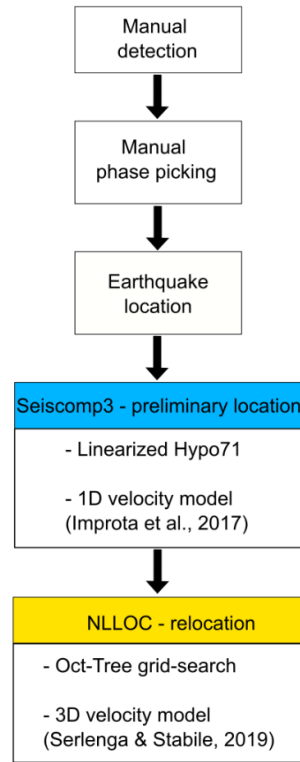
Figure 3. 3 – Standard workflow for earthquake detection, phase picking and location workflow applied for HAV seismicity location.

Station name	Network FDSN code	Latitude °N	Longitude °E	Elevation (m a.s.l)
CDRU	IV	40.49	15.305	1057
CRAC	IV	40.365	16.435	384
CUC	MN	39.994	15.816	637
MATE	IV	40.649	16.704	494
MCEL	IV	40.325	15.802	960
MGR	IV	40.138	15.554	288
MIGL	IV	40.604	16.441	440
MMN	IV	39.891	15.99	921
MRVN	IV	41.061	16.196	610
MTSN	IV	40.266	15.752	1056
ORI	IV	40.051	16.45	375
PTRP	IV	40.522	16.061	1077
PZUN	IV	40.646	15.807	820
SCHR	IV	40.199	16.076	968
SIRI	IV	40.182	15.868	1063
SLCN	IV	40.39	15.633	986
MARCO	GE	40.297	15.806	890
AND3	IX	40.93	15.333	905
CGG3	IX	40.542	15.523	1067
MNT3	IX	40.837	15.007	866
MRN3	IX	40.426	15.73	772
PGN3	IX	40.572	15.797	882
RSF3	IX	40.964	15.176	865
SCL3	IX	40.695	15.511	744
SRN3	IX	40.486	15.458	1067
STN3	IX	40.53	15.652	832
VDP3	IX	40.605	15.572	959

Table 3. 2 – External stations belonging to the virtual network used for HAV seismicity location.

1) A total of 43 Injection induced (II) seismic events, which epicenters concentrate in the seismicity cluster located NE to the Pertusillo lake and close to the Costa Molina 2 well, recognized to be related to fluid-injection operations at the CM2 well (cluster B in Figure 3.1) (Stabile et al., 2021; Hager et al., 2021). The first recording station of such events is MONCM (the INS1 station until 2019, see Table 3.1), the closest receiver, where the signals are characterized by a difference between the arrival times of S- and P-waves of about 1 s and usually a duration of less than 7 s. The average depth for injection induced events is about 4.5 km. Most detected injection-induced events have a magnitude lower than or equal to 1; only two of them have local magnitudes of Ml = 1.1 and Ml = 1.4, respectively. The largest induced event of the catalogue occurred on 29 January 2018 at 15:23:10 (UTC), located at about 1.4 km epicentral distance from MONCM station with a focal depth of about 3.0 km (Ml = 1.4; latitude 40.3182˚N, longitude 15.9842˚E; depth 3.50 km; from https://doi.org/10.5281/zenodo.3632419, Stabile et al., 2020). Depending on the earthquake magnitude, the number of stations that recorded the seismic signals changes from a minimum of three, for the lowest magnitude event, up to 16 stations including those of the virtual network. The highest peak ground velocity amplitude (PGV) measured at MONCM station is about 0.04 mm/s for the strongest event of this category in the catalogue.

2) A total of 117 reservoir-induced seismic events (RI), belonging to the cluster A (Figure 3.1) located SW of the Pertusillo lake. The P-wave arrivals are usually first detected at the three different seismic stations (GRUME, SARSB, or SARCL) indeed, such induced events belong to a wider cluster and are more broadly distributed in the southwestern part of the seismic network. Their average depth is about 4.5 km and in their estimated local magnitudes range $0 \leq Ml \leq 1.8$. The event with maximum recorded local magnitude (Ml = 1.8) occurred on 2 March 2017 at 21:39:41 UTC, located at about 1.9 km epicentral distance from the GRUME station (Ml = 1.8; lat 40.2723˚ N, long 15.8840˚ E; depth = 4.51 km; from https://doi.org/10.5281/zenodo.3632419, Stabile et al., 2020). RI earthquakes are characterized by a difference between S- and P-phases arrival times of about 1 s at the closest station to the epicenter, and by short duration (less than 8 s). The recorded seismic event with the lowest magnitude was detected by seven stations, whereas the strongest earthquake was recorded by 12 stations. Finally, the highest peak ground velocity amplitude recorded up to now for this earthquake category is about 0.08 mm/s.

3) A total of 692 local natural earthquakes (LE). Their average hypocentral depth of about 10 km is ascribable to the typical Apennines crustal earthquakes depth. Most of the located tectonic earthquakes are characterized by a local magnitude < 2. Only 39 seismic events have magnitude greater than 2, and only four earthquakes with a magnitude greater than 3, included in a radius of about 40 km from the center of the INSIEME seismic network, have been recorded. The strongest event close to the INSIEME network (epicentral distance of 16 km from GRUME and SARSB stations) is a Mw = 3.8 earthquake. The highest peak ground velocity amplitude of more than 3 mm/s was recorded at MTSN station, managed by INGV, which was the closest station located at 5 km epicentral distance. The

earthquake was recorded by the whole INSIEME seismic network, as well as by all the stations of the virtual network that were in operation that day (event Ml = 4.0; lat 40.3040° N, long 15.7200° E; depth = 12.10 km reported in https://doi.org/10.5281/zenodo.3632419; Stabile et al., 2020). The main difference with respect to the II and RI events is that the recognized tectonic local earthquakes (LE) are not clustered, but randomly spread in the investigated area. Exceptions are represented by two main local seismic swarms and sequence (Figure 3.7): (a) the Pignola sequence, located to the North of the HAV recorded between the 6 and 7 October 2016, which mainshock corresponds to the earthquake with highest magnitude in the catalogue (Ml = 4.2; lat 40.5373 °N; long 15.7862 °E; depth 15.73 km reported in https://doi.org/10.5281/zenodo.3632419; Stabile et al., 2020); (b) a 0.5> Ml >1.5 seismic swarm made by 13 events recognized NW to the Tramutola town (Lat: 40.33; Long: 15.75) occurred between 26 and 27 October 2016. It is worth noting a slight swarm, which events occurred sparsely over the investigated time span, and located just to the north of Marsicovetere town (Lat: 40.36; Long:15.8), which is assimilable to the previously recognized Marsicovetere natural seismic sequence occurred between September-October 2004 (Improta et al., 2017)
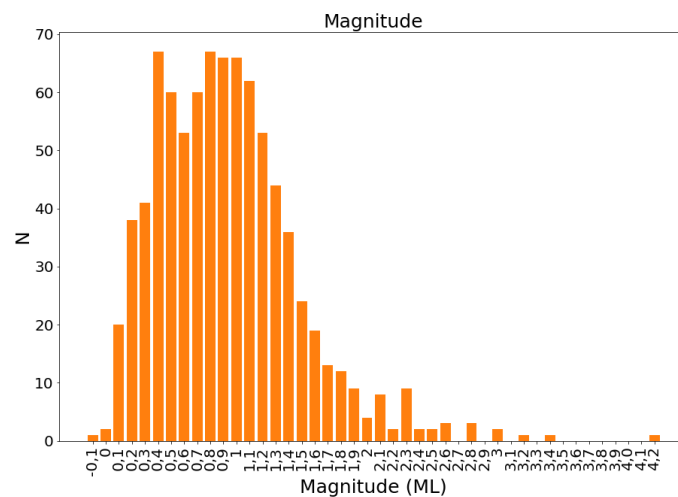
Figure 3. 4 – Local magnitude distribution of the High Agri Valley seismicity preliminary catalogue.

Preliminary local magnitudes (Ml) of earthquakes were estimated through the magnitude calculator (scmag) implemented in SeisComp3. Here, the Ml amplitude calculation is similar to the original definition given by Richter (1935), such that waveforms from both horizontal components are time-windowed and restituted to the Wood-Anderson seismograph. The obtained preliminary magnitude distribution of the catalogue is shown in Figure 3.4. In it, the lower local magnitude values (Ml<2) were estimated for the anthropogenic earthquakes: RI local magnitudes range between Ml = 0.0 – 1.8, while for the II it ranges between Ml = -0.1 – 1.4. The earthquake with the higher magnitude belongs to the LE earthquakes, which show a wider magnitude range (Ml = 0.1 - 4.3). The estimated magnitudes show a bimodal distribution (Figure 3.4), with two most frequent magnitude values at Ml = 0.4 and Ml = 0.8, thus not allowing the identification of a unique magnitude of completeness (Mc) – defined as the minimum magnitude value above which all earthquakes within a certain region are reliably recorded – for the catalogue, at this stage. To check the observed magnitude distribution

behavior, we tested the Seiscomp3 Ml calculation: the scmag tool does not apply any filter on earthquake waveforms before magnitude calculation. Therefore, the magnitude estimation is affected by low-frequency noise amplitudes in the signal which are considered in Ml calculation. This would particularly affect microearthquakes for which the Seiscomp3 magnitude computation would rather consider noise maximum amplitude instead of the event one. We thus interpret the two peaks in the bimodal magnitude distribution as two Mc values: (a) Mc=0.8 representative of the all the events of the catalogue located from the entire virtual network; (b) Mc=0.4 representative of earthquakes located inside the local INSIEME network - much denser and closer to local microseismicity - less affected by the noise in local magnitude estimation. The analysis highlighted the necessity to apply magnitude estimation techniques more suitable for the dataset, particularly for small events (Ml < 0.4). For this purpose, the local magnitude computation through the local scale proposed for southern Italy by (Bobbio et al., 2009) was implemented in this work. The method was tested for the Castelsaraceno sequence (see paragraph 3.3.3.2) and will be further extended to HAV seismicity in the future. Moreover, a significant future development for a better characterization of the HAV seismicity catalogue, will be the computation of the dataset moment magnitude (Mw) from the estimates of seismic moment Mo (Hanks and Kanamori, 1979) and directly related to the source radiated energy rather than to wave propagation effects.

3.3.2.2 Earthquake relocation

The events of the collected seismic catalogue were relocated using the equal differential time (EDT) method implemented in a non-linear global approach algorithm (NonLinLoc code, Lomax et al., 2000) and considering a 3-D velocity model of the area (Serlenga and Stabile, 2019). After a first step, a second iteration of the location algorithm was carried out taking into account station correction obtained in the previous step. This allowed to obtain more accurate event locations with respect to the preliminary locations obtained from the previous linearized location in a 1D velocity model, and to better define the three earthquake categories, thus optimizing the location results. Indeed, earthquake relocations are characterized by low values of the overall arrival-time residuals (RMS) (Figure 3.5c), with a median value of 0.1 s. Their distribution is gaussian, thus reflecting a slight random accuracy error related to the manual picking of the P- and S-phase arrival-times. RMS of P- and S-phases calculated for all the stations of the virtual network used for the earthquake relocations are displayed on map (Figure 3.5a) and also as a plot of the RMS values for each station (Figure 3.5b). These are generally higher for the S-waves arrival times since their manual picking is more challenging due to contamination of the direct S-wave arrival by the P-waves coda and converted phases. Nevertheless, the highest RMS absolute value of about 0.6 is related to P-waves pickings at two stations of the virtual network: ORI, at which the P-waves picking persistently anticipates the phase arrival (RMS = -0.6) and VGG3 station, at which a delay for both P-waves (RMS = 0.57) and S-waves (RMS = 0.51) picks is observed. These anomalously high errors in picking accuracy with respect to the other stations of the virtual network would be imputable to errors in the timing of seismic waveforms belonging to these external stations. The station RMS map view (Figure 3.5a) also confirms the improvement of absolute earthquake location accuracy related to both the local INSIEME network geometry and the availability of a three-dimensional velocity model with the

lowest RMS values recognizable at the inner stations of the 3D tomographic grid (Serlenga and Stabile, 2019). The azimuthal gap distribution of the relocated earthquakes (Figure 3.6) shows a bimodal trend with the highest values (>250°) mainly related to events located outside the study area. Conversely, the earthquakes occurred inside the virtual network (40.20 °N < Latitude <40.50 °N; 15.66 °E < Longitude < 16.19 °E) show a good azimuthal gap (median value of 160.5°) which fulfills the requirements of an azimuthal coverage better than 180° required for high-quality recordings in seismic networks. Earthquake relocation, confirmed a bimodal distribution also for hypocentral depths (Figure 3.6), pointing out its role as discriminating parameter between the anthropogenic seismicity clusters (II and RI) - characterized by depths < 5 km - and tectonic local events with an average hypocentral depth typical of Apennines crustal earthquakes of 10 km.
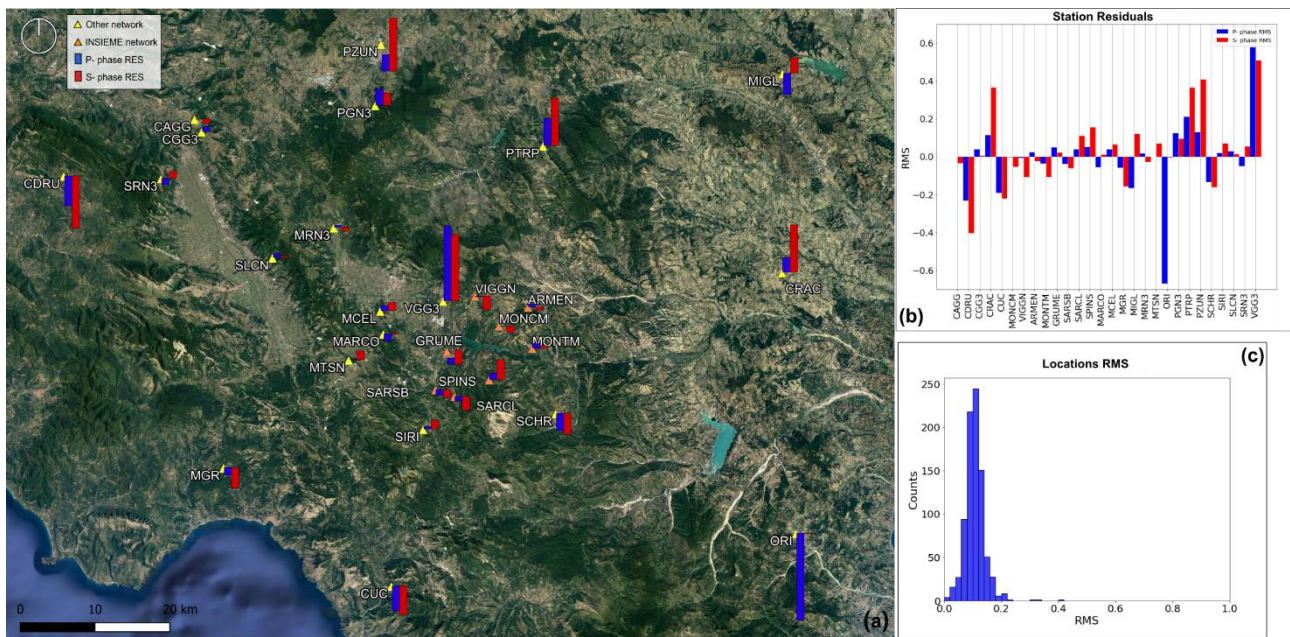
Figure 3. 5 – Resulting total locations RMS (c) and stations residuals for both P- and S-phases reported both in spatial distribution (a) and bar plots (b) for earthquakes absolute locations in a 3D velocity model.
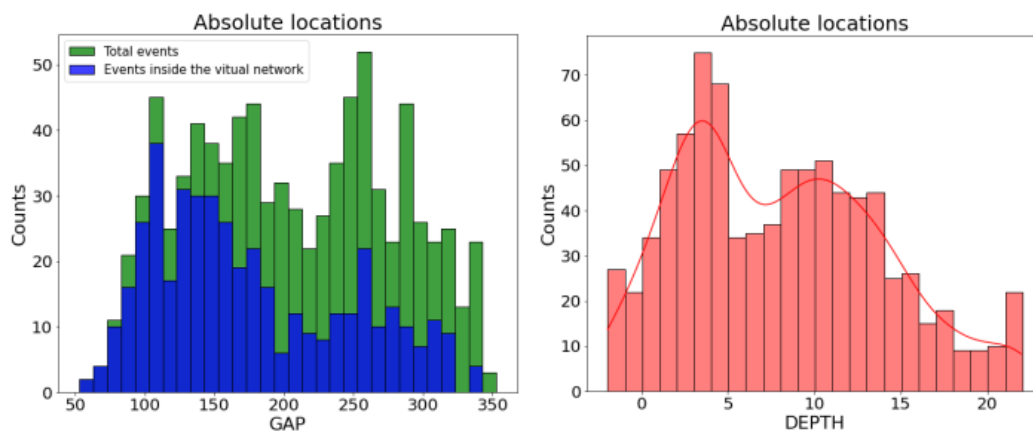
Figure 3. 6 – Azimutal gap (left) and hypocentral depths (right) distributions obtained from the non-linear global locations (NonLinLoc code) performed in in 3D velocity model of the High Agri Valley.

The comparison of the two applied location methods is shown in Figure 3.7. Earthquake relocation with a three-dimensional model, results in a better discrimination of the defined earthquakes categories, with more

defined clusters for both injection-induced events and tectonic local earthquake swarms (e.g., Pignola sequence). Moreover, the NLLOC relocation better highlights the presence of a seismic gap between the High Agri Valley seismicity (to the southwest) and the deeper (mean focal depth 15-20 km) Apennines crustal earthquake to the northeast. The gap occurs in the area where the hydrocarbon production of the Val d'Agri Oil field is currently carried out and it is in accordance with the stabilization effect of the extraction operations shown by Hager et al. (2021) in the area. These authors explained the low seismicity rate in the production area as geomechanically imputable to the decrease in pressure in most of the reservoir and the subsurface aquifers due to production. Indeed, oil extraction would produce a reduction of the principal vertical stress (σ_1) and the subsequent equilibration of the stress state on extensional fault planes within and surrounding the reservoir.
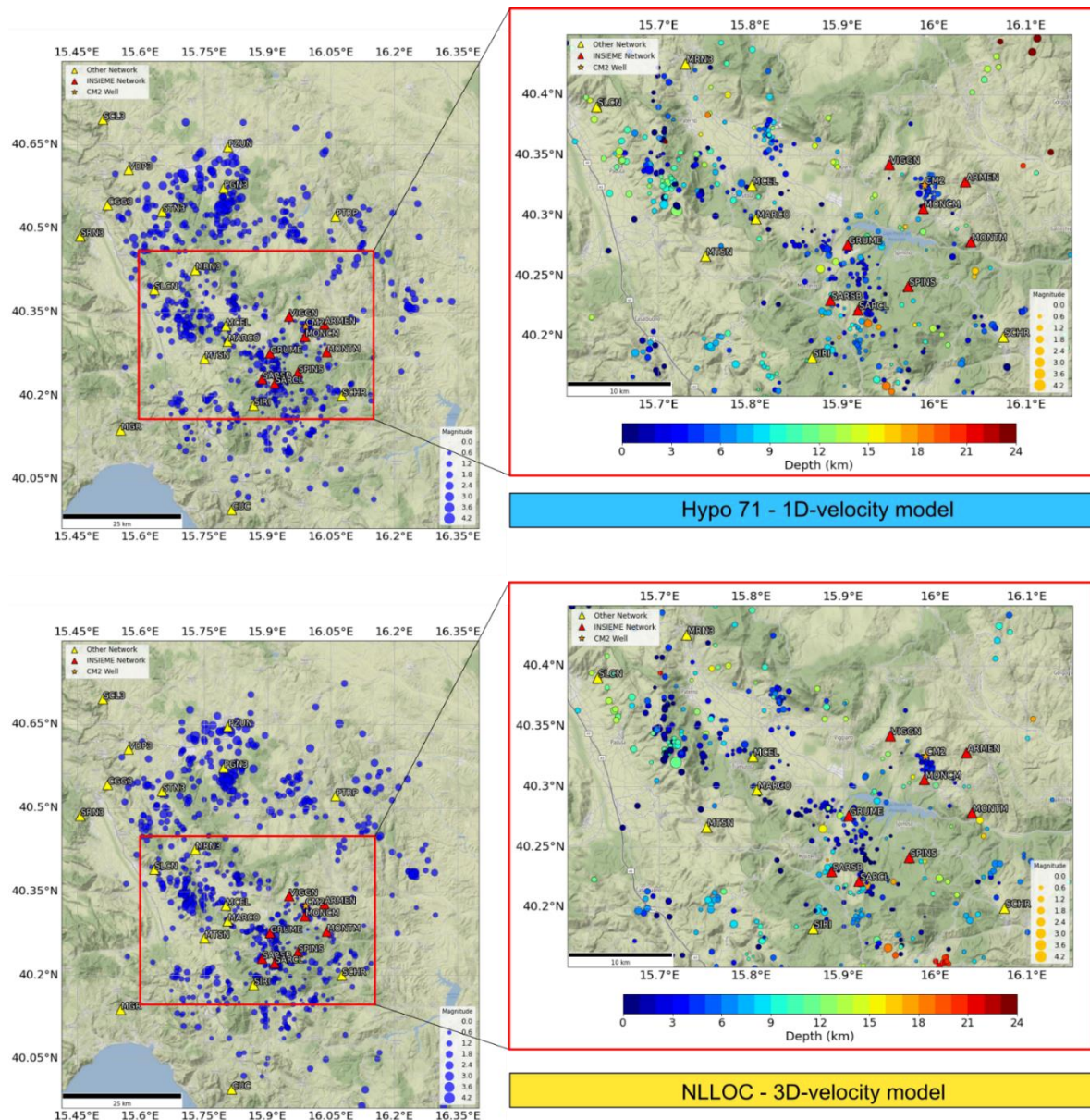
Figure 3. 7 – Comparison of absolute locations spatial distribution in HAV, between the preliminary linearized (Hypo71) method in a 1D-velocity model and the grid-search EDT (NonLinLoc) method in a 3D-velocity model.

3.3.2.3 Seismicity cluster analysis

Cluster analysis represents a useful exploratory technique to verify the existence of similar behavior between observations. Indeed, the purpose of clustering algorithms is to recognize, within a certain dataset, groups or clusters characterized by internal homogeneity with respect to the explored parameters. To verify the cluster behavior of the discriminated earthquakes categories (e.g., II, IE, LE) and better interpret the spatial distribution of the located events in the High Agri Valley, we performed a cluster analysis based on their collected parameters (latitude, longitude, hypocentral depth, Ml). Two kinds of cluster algorithms were applied: the K-Means and the DBSCAN.

The K-Means is commonly used as a first approach algorithm, in exploratory analysis with the general purpose to highlight the main distribution patterns of the analyzed parameters and to detect outliers in the dataset. The algorithm, searches for clusters defined within a distance to their centroid: the detected cluster (C) is defined by the mean of its samples (μ_j) (named centroid), and choosing the centroid that minimizes the inertia, or within-cluster sum-of-squares criterion, given by:

$$\sum_{i=0}^{n} \min_{\mu_i \in C}(\|x_i - \mu_j\|)$$

Eq.3. 1

The inertia is intended as a measure of how internally coherent clusters are. The algorithm was implemented through the open-source scikit-learn machine learning Python module (Pedregosa et al., 2011) on previously normalized parameters to avoid any problem related to their different scale. First, we performed the multivariate cluster analysis exploring the hypocentral coordinates obtained from the earthquake relocations (latitude, longitude, depth) (Figure 3.8) searching only for two clusters to explore the dataset and eventually detect outliers in it. The analysis mainly divided the HAV seismicity based on its hypocentral depth, thus confirming its role as discriminating parameter of the seismicity in the area: the cluster 0 (the yellow one in Figure 3.8) is characterized by hypocentral depths lower than 10 km, typical value estimated for the induced earthquakes; the cluster 1 (the purple one in Figure 3.8) contains deeper tectonic earthquakes (mainly > 10 km). This general analysis also confirmed the observed spatial – depth distribution (Figure 3.7) with the deepest earthquakes (cluster 1) mainly located at highest latitudes (Lat: 40.1 - 40.4 °N) with respect to the surficial ones (Lat: 40.0 - 40.4 °N). To further discriminate the dataset, we performed a four-dimensional K-Means analysis adding the magnitude to the hypocentral parameters. We searched for three clusters at this stage. Again, the main cluster discrimination parameter is the earthquake hypocentral depth, which enabled to better recognize surficial seismicity (cluster 0, Figure 3.9), centered at typical depths of the anthropogenic events (about 5 km) mainly located to the SW. Moreover, it divides a deeper seismicity (>10km) of the cluster 1 (purple one, Figure 3.9) concentrated to the highest latitudes, with a more surficial seismicity (cluster 2, Figure 3.9) characterized by a median hypocentral depth of 10 km, which is randomly spread in the study area. The magnitude was not crucial for the clustering since the discriminated clusters widely superimpose for Ml < 2.0. Nevertheless, they show an increasing magnitude trend from the surficial (cluster 0) to the deeper (cluster 2) one, as was expected. The K-Means analysis allowed us to generally explore the located earthquake parameters correlations and confirm their spatial distribution in the study area. Conversely, it was not highly

effective in the discrimination of anthropogenic and tectonic earthquakes clusters. Indeed, a drawback in the application of the K-Means is the assumption that clusters are convex, isotropic, and even sized, while it poorly responds to clusters with irregular shapes as the ones we were searching for: small clusters of earthquakes related to anthropogenic activities or seismic swarms, within a widespread tectonic seismicity.
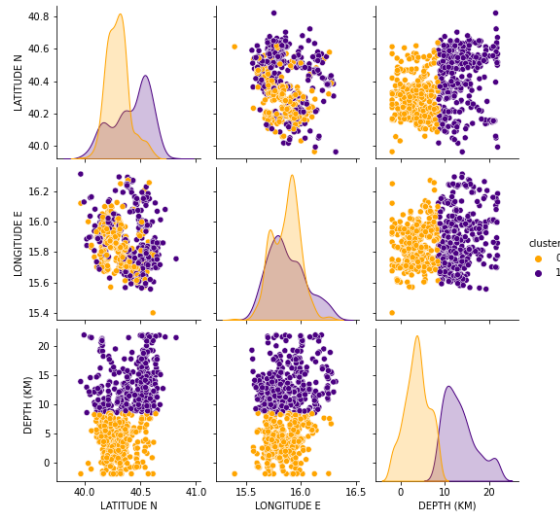
Figure 3. 8 – Resulting pairplot from K-Means clustering performed though earthquakes hypocentral parameters (latitude, longitude, depth) and requiring 2 clusters.
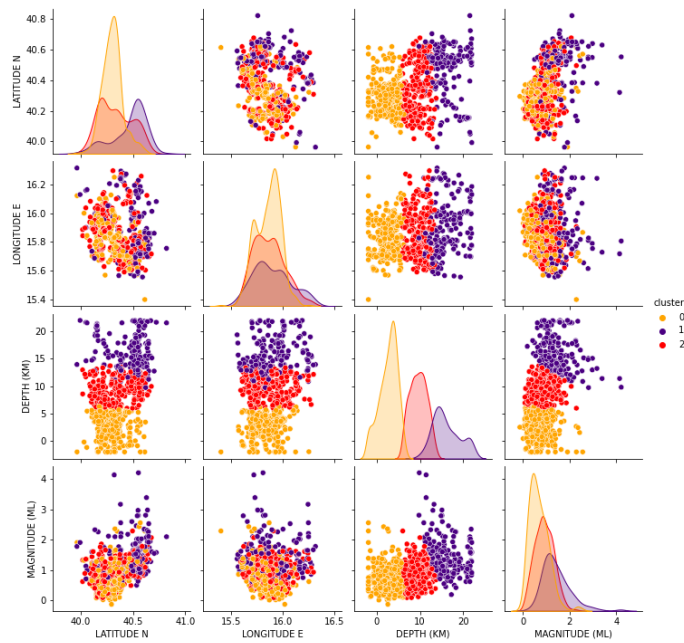
Figure 3. 9 - Resulting pairplot from K-Means clustering performed though earthquakes hypocentral parameters (latitude, longitude, depth) and local magnitude, requiring 3 clusters.

With this aim, the DBSCAN clustering algorithm was applied (see Chapter 2.4.2). It is a density-based algorithm which considers only the density of the features (parameters) in the observation space, thus separating clusters (high density area), from outliers (low density areas). For this reason, clusters found by DBSCAN can be any shape, as opposed to K-Means. Moreover, it does not need any a-priori assumption on

the searched number of clusters in the dataset, but it requires two hyperparameters used to define a cluster: (1) the distance within which search for close points (ε); (2) the minimum number of samples to create a cluster (n). We developed a grid search technique to explore different hyperparameters combinations to obtain the one which maximize the clustering performance for our dataset (ε = 0.6, n = 20) (explained in detail in Chapter 4.4.3.1) and performed the analysis on the located earthquakes parameters (latitude, longitude, depth, magnitude), previously normalized. The results are shown in the pair plot in Figure 3.10.
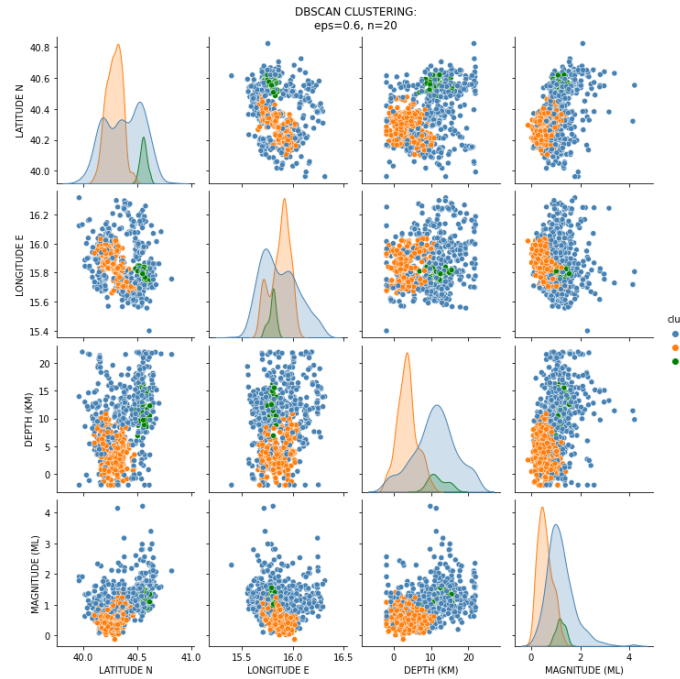
Figure 3. 10 – Resulting pairplot from DBSCAN clustering performed through hypocentral parameters (latitude, longitude, depth) and local magnitude (ε=0.6, n=20).

The DBSCAN cluster analysis identified two main clusters in the dataset and their outliers (cluster -1, Figure 3.10), with the following characteristics:

(a) the cluster 0 (yellow one, Figure 3.10) with typical hypocentral depths lower than 10 km, a magnitude range Ml = -0.1 - 1.8 and located to the SW of the study area (Lat: 40.2 - 40.4 °N; Long: 15.7 e 16.0 °E).

(b) the cluster 1 (green one, Figure 3.10), to which belong earthquakes with hypocentral depths of 10-15 km, with Ml < 2 and located to the North of the study area (Lat: 40.6 °N; Long: 15.75 °E).

(c) the outliers (blue one, Figure3.10) show wider and no-preferential spatial distribution, covering all the latitude (Lat: 40.0-40.7 °N) and longitude (Long: 15.5 – 16.3 °E) ranges of the study area. Their hypocentral depth and magnitude distributions may be assimilated to gaussian (diagonal density plots, figure 3.10), centered to Ml=1.2 and depth of about 12 km.
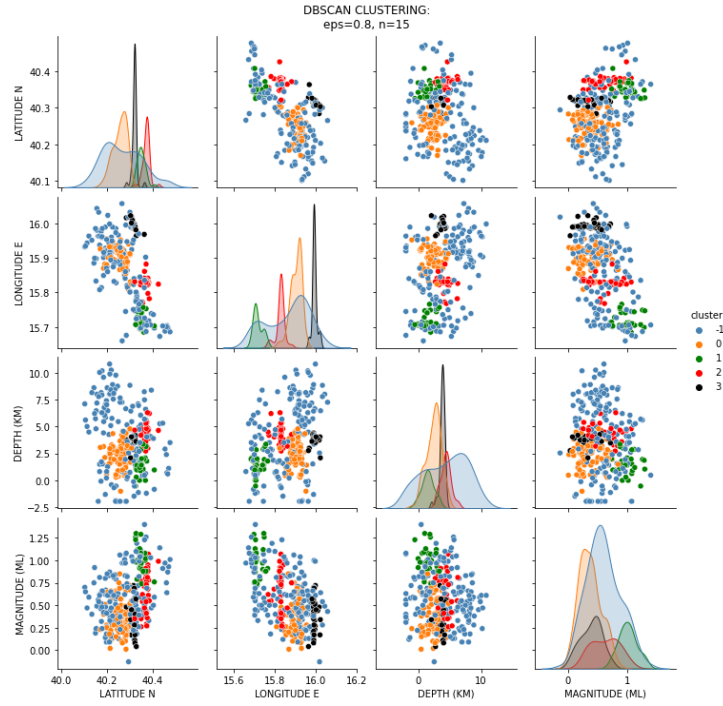
Figure 3. 11 - Resulting pairplot from DBSCAN clustering performed through hypocentral parameters (latitude, longitude,depth) and local magnitude (ε=0.6, n=20).

The analysis outlined that the density-based clustering is more efficient to verify the presence of seismicity clusters within the catalogue. Indeed, it well identified the Pignola sequence located to the North of the study area and which location parameters and magnitude values are compatible with the cluster 1. The latter was well distinguished from the widespread, random seismicity (outliers) which characteristics are ascribable to the background, sparse, Apenninic seismicity of the area. However, the analysis with the available standard earthquake parameters (latitude, longitude, depth, Ml) has not allowed to further differentiate the identified anthropogenic clusters (II, IE) and the HAV local tectonic seismicity, which have been categorized as a unique cluster (cluster 0). For this purpose, we further applied the DBSCAN analysis only to the previous subset (cluster 0), representative of the total HAV seismicity. Its dataset was concentrated in a more restricted area, within a narrow range of latitudes (Lat: 40.10 - 40.5 °N) and longitudes (Long: 15.6 - 16.0 °E), such that a higher value of ε (0.8) was chosen; moreover, we required a smaller minimum number of data to identify a cluster (n=15), due to the small size of the earthquake clusters we want to verify. Since the analyzed subset was characterized by hypocentral depths mainly lower than 10 km, the cluster analysis let us to define only surficial grouped events in both anthropogenic and tectonic local seismicity (depth < 5km), thus identifying as outliers the deeper local background seismicity in the HAV (Figure 3.11). The analysis defined:

1. A small cluster (cluster 3, Figure 3.11) of thirty-seven events, located in the northeastern area (Lat: 40.3, Long 16.0), characterized by hypocentral depth centered around 3-5 km (median 3.85) and earthquake median magnitude of 0.4, which characteristics and the number of earthquakes are ascribable to the injection-induced earthquakes (II).

2. A wider cluster of 88 events (cluster 0, Figure 3.11), located SW (Lat: 40.2 - 40.3 °N, long: 15.8-15.9 °E) of the HAV area, to which belong surficial earthquakes (median 2.6 km) characterized by Ml < 1.0 (median 0.4), which location and size are comparable to the reservoir-induced seismicity (RI).

59

3. Two minor clusters (cluster 1 and cluster 2, Figure 3.11) both located in the northwestern area of the HAV. The location of cluster one (Lat: 40.3 °N; Long: 15.7 °E) suggests a similarity with the recognized Tramutola swarm which median magnitude is equal 1.0. The cluster two is located to the east (Lat: 40.35; Long: 15.81) with respect to the previous one, around Marsicovetere town, thus suggesting its correlation with the observed swarm seismicity cluster in that area.

The results of the performed cluster analysis on the relocated earthquake catalogue underlined the high potential of this unsupervised machine-learning approach as a useful automated tool to make faster dataset exploratory analysis (K-Means) and to identify clusters in the located seismicity (DBSCAN). The density-based approach was particularly suitable to verify the presence of clusters in the located HAV seismicity related to both anthropogenic (II, RI) and local seismic swarms and sequences (e.g., Pignola sequence, Tramutola swarm). The analysis was performed with the available standard earthquake parameters (latitude, longitude, hypocentral depth, Ml) which ranges were highly comparable within the analyzed dataset, except for the hypocentral depth. Future development of the analysis will involve the collection of more discriminating information about the High Agri Valley seismicity (e.g., earthquakes source parameters) thus obtaining a more complex, multivariate analysis of the dataset for a more efficient recognition of clusters within it, and with the aim to find discrimination criteria between induced and tectonic local seismicity.

3.3.3 Advanced algorithms application

The standard workflow described in the previous section, allowed to reconstruct the HAV seismicity catalogue through high-quality earthquake locations, in a 3D-velocity model. However, it was characterized by both manual earthquake detection and phase-picking, quite time-consuming. In this section, we show the results of a faster, semi-automatic workflow which was tested on the low-magnitude tectonic earthquake sequence occurred on August 2020 in the southeastern sector of the HAV, close to the municipality of Castelsaraceno. It was characterized by a 2.9 Ml (magnitude estimation from INGV) mainshock which took place on 7th August. Seismic data, mainly recorded by the eight stations of the HAVO (formerly INSIEME) network, located at a maximum epicentral distance of ~20 km from the sequence cluster, allowed investigating the main properties of this seismic sequence. The advanced algorithm application involved the integration of manual, semi-automatic and automatic techniques in the following stages of the analysis (Figure 3.12): (1) we manually detected microearthquakes in the period between 7 and 10 August, when most of the seismicity occurred; (2) we applied a single-station template matching technique (Roberts et al., 1989; Stabile et al., 2021) looking for microearthquakes belonging to the sequence and possibly escaped to the manual procedure due principally to their low signal-to-noise ratio; (3) the P- and S-wave arrival times were automatically picked by a deep-neural-network algorithm (PhaseNet, Zhu and Beroza, 2018); (4) we provided absolute and relative locations of detected events. Moreover, the source characterization of the mainshock was performed through the estimation of its focal mechanisms (FPFIT, (Reasenberg and Oppenheimer, 1985). The analyses revealed the alignment of the seismic sequence at an average hypocentral depth of about 10 km, along a normal fault with a strike-

slip component. Finally, the mainshock seismic data were inverted for source parameters estimation (SourceSpec, Satriano, 2021).

ADVANCED METHODOLOGIES WORKFLOW

Manual		*Automatic*
detection and picking	Manual detection (WebObs, Beaducel et al.,2020) Single Station Template Matching Detection Algorithm (Stabile et al., 2021)	Deep Neural P- and S-waves arrival time algorithm (PhaseNet, Zhu and Beroza, 2019)
earthquake location	Double difference earthquake locations by means of HypoDD (Waldhauser & Ellsworth, 2000) Magnitude computation through a local magnitude scale for Southern Italy (Bobbio et al., 2009)	Iterative four-step absolute earthquake location by means of NLLoc (Lomax et al., 2000)

Figure 3. 12 – Advanced methodologies for earthquake detection, phase picking and location workflow tested for the study of the Castelsaraceno sequence.

3.3.3.1 Earthquake detection and phase-picking

The detection of the earthquakes belonging to the sequence, was performed in two steps: first, we manually detected microearthquakes in the period between 7 and 10 August, when most of the seismicity occurred; then, we applied a single-station template matching technique for the two nearest stations to the cluster (SARCL, SPINS). The manual detection, performed through the dedicated intranet system (WebObs, Beauducel et al., 2020) (see Chapter 2.2.2) let us define a total number of 60 earthquakes between the 7 and 10 August, well recorded by the eight broadband stations of the INSIEME network. Among them, we identified the main $Ml=1.9$ foreshock (origin time: 2020-08-07, 08:52:34 UTC), followed few hours later by the sequence $Ml = 2.9$ mainshock (origin time 2020-08-07, 13:34:39 UTC). The last manually recognized earthquake of the sequence was detected the 10th August (origin time 19:15:21 UTC). For the detection of small events possibly missed by the conventional manual approach, we applied the single-station template matching algorithm (Stabile et al., 2021), based on the cross-correlation technique developed by Roberts et al., 1989 (described in detail in paragraph 2.2.2.1). We applied the analysis on three-component continuous data streams recorded in the period between July and October 2020 at the two nearest stations to the cluster SARCL (2.9 km epicentral distance from the mainshock) and SPINS (5.7 km epicentral distance from the mainshock) of the HAVO network. The analysis was performed for each single station using 4 master templates for SARCL station (Figure 3.13) and 6 for SPINS station (Figure 3.14). The selected templates are summarized in Table 3.3: three common master events were used, such that a total number of seven templates were selected for the analysis. The high number of templates was selected to detect as many new events as possible. For each template, a cross-correlation window of 1.2 seconds was manually selected around the first P-wave arrival on the vertical component (CHZ) and around the S-wave arrival on the horizontal components: CH1, CH2 for SARCL and CHE, CHN for SPINS. We chose a wide time window in order to limit the number of automated false detection,

since the shorter the length of the cross-correlation window, the higher is the probability to obtain high cross-correlation (high *XC'* values) with the anthropogenic noise.

SARCL STATION		
Master template	**Date**	**Origin time**
SR1	2020-08-07	09:04:10
SR2	2020-08-07	09:09:03
SR3	2020-08-07	13:34:39
SR4	2020-08-07	13:37:16

SPINS STATION		
Master template	**Date**	**Origin time**
SP1	2020-08-07	09:04:10
SP2	2020-08-07	09:09:03
SP3	2020-08-07	13:34:39
SP4	2020-08-07	19:05:44
SP5	2020-08-08	01:35:58
SP6	2020-08-08	07:10:12

Table 3. 3 – Template summary for SARCL station (green table) and SPINS station (orange table). The grey lines represent common earthquakes templates used for both stations.
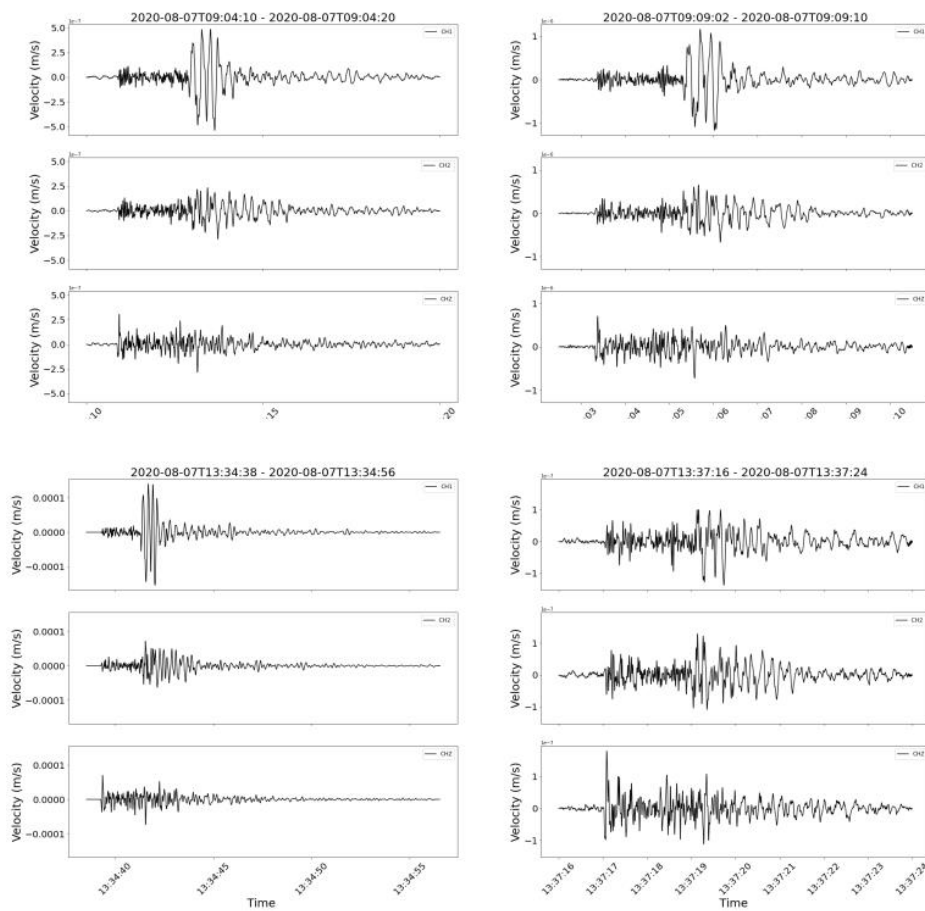
Figure 3. 13 – Master templates selected for SARCL station.
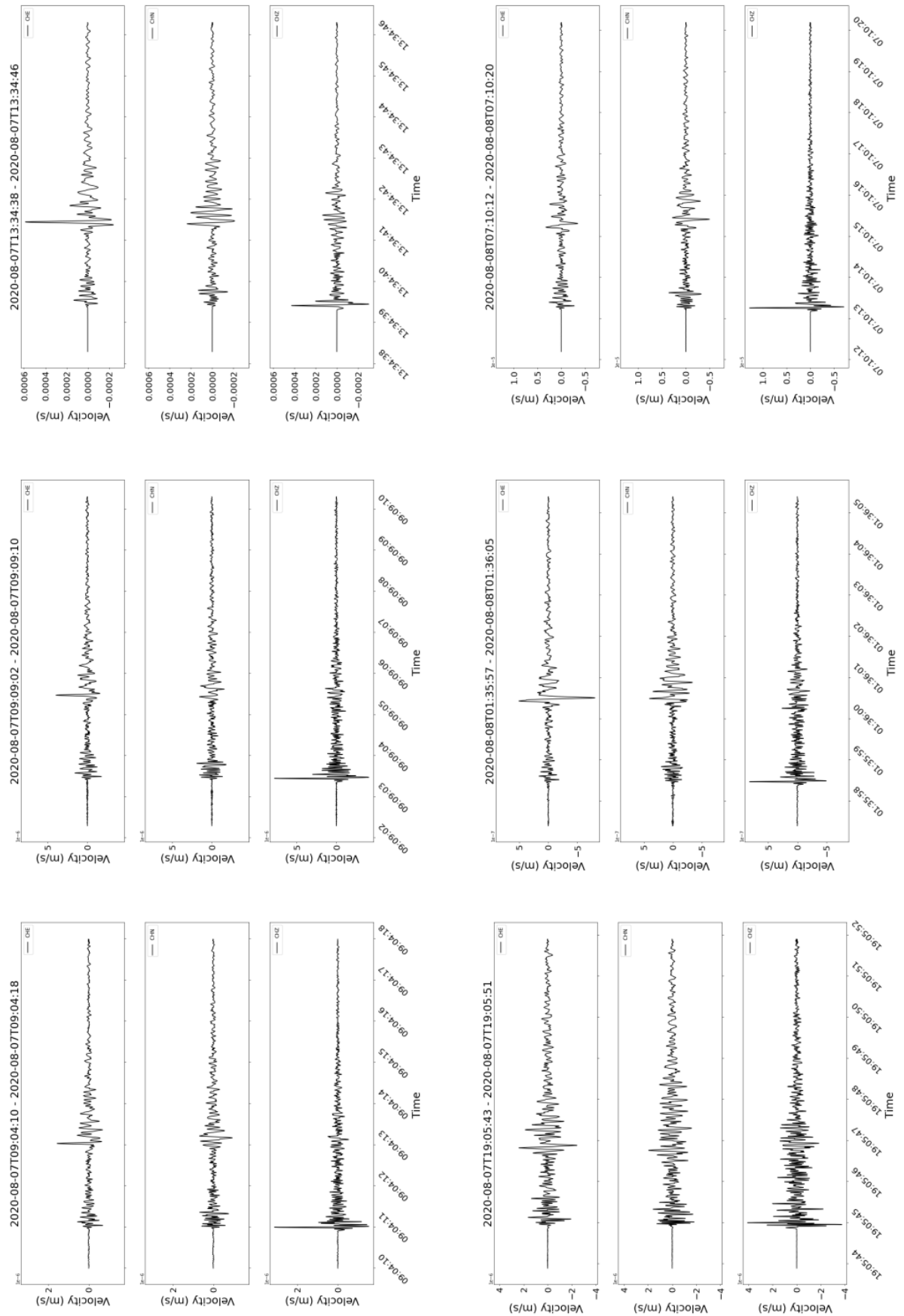
Figure 3. 14 - Master templates selected for SPINS station.

Following a similar procedure of Stabile et al. (2021), all the three-component templates and the continuous data streams ground velocities (m/s) of each analyzed station were first band-pass filtered in the range 1-35 Hz, and then differentiated to obtain their ground acceleration displacement to amplify the peak frequency of the signals. Finally, there were automatically detected only the signals which contemporary fulfilled the two main conditions (Stabile et al., 2021): (1) the obtained cross-correlation parameter (*XC'*) between the master template and the station component data stream was greater than 0.5; (2) the previous condition was satisfied on the vertical component and at least on one of the two horizontal components (Figure 3.15). We set a relatively low value of the detection threshold (*XC'* = 0.5) as we wanted to maximize the number of detected events. However, these led to a high number of detected false events which were manually discriminated, by visual inspection, from the true automatically detected earthquakes. A total of 107 further earthquakes was automatically detected from the analysis at both stations, enabling us to increase the total number of detected earthquakes to 176.
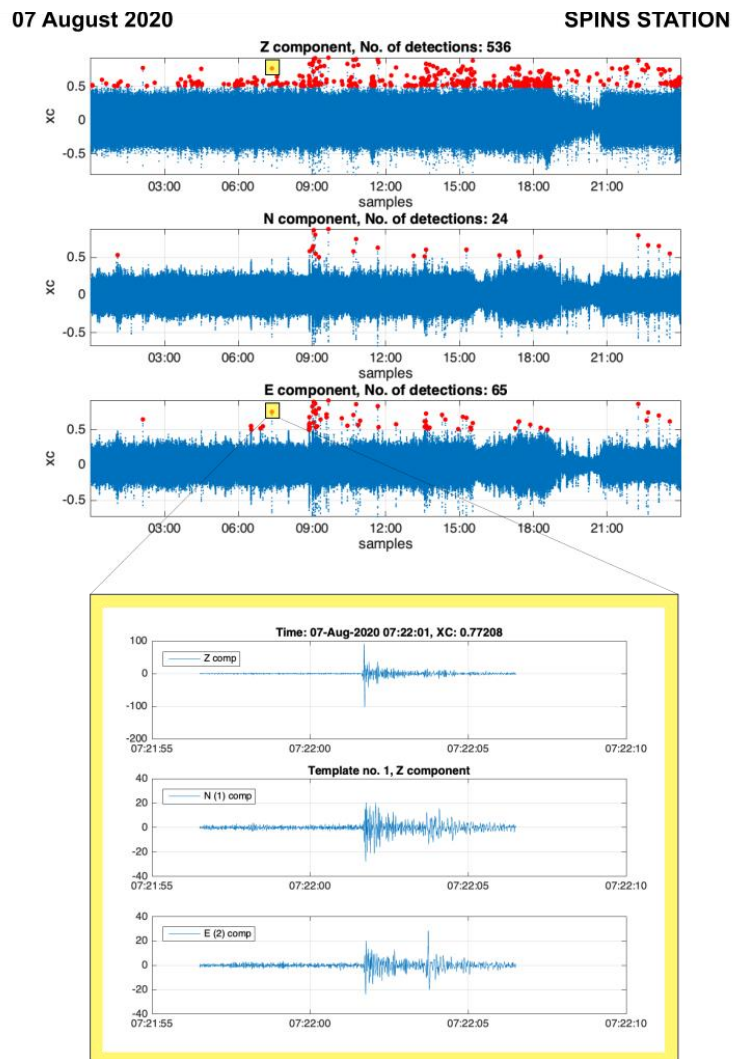
Figure 3. 15 – Example of the single-station template matching algorithm applied to SPINS station on 7 August 2020. The yellow square shows the waveform of an automatedly detected event which fulfilled the described algorithm requirement at the Z and E components.

For the P- and S- waves arrival time picking, we tested a completely automated approach through the application of the PhaseNet algorithm (Zhu and Beroza, 2018). It is a deep neural network algorithm, which allows to perform seismic phase picking without any hand-designed input parameter. Indeed, thanks to their multi-layer neural networks structures, deep-learning algorithm are capable to "understand" data in a hierarchical way, avoiding the need for human operators to formally specify the training feature that they need to work efficiently (paragraph 2.4.4.1 and reference therein). Avoiding any particular pre-processing on the dataset, we applied the algorithm on three-component seismic waveforms, bandpass filtered in the range 2-40 Hz. Therefore, we performed the automated P- and S-wave pickings on all the manually and semi-automatically detected events recorded by all the stations of the INSIEME network (except for MONCM station which was not working in the analyzed timespan) and two of the external stations of the virtual network (MARCO and SIRI stations). For each detected earthquake, a three minutes-long waveform sampled at 250 Hz was given in input to PhaseNet, starting 60 seconds before its detection time (obtained from the single-station template matching algorithm), and ending 120 seconds later; a 30-seconds sliding window over the entire record was chosen for the analysis. The P-wave, S-wave and noise probability distributions are calculated in output from the deep-learning algorithm, which defines the corresponding P- and S- arrival times where the resulting probability distributions peak values go above a certain threshold (Duarte, 2015). A threshold of 0.3 was chosen for our dataset (Figure 3.16). The application of PhaseNet algorithm enabled us to perform the earthquake phases picking in a shorter amount of time compared to a standard manual picking procedure.
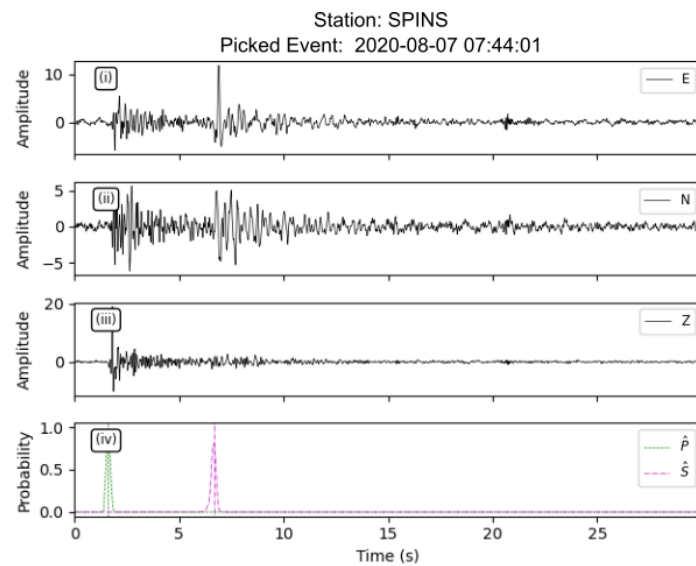
Figure 3. 16 – Example of automatically picking through a deep-learning neural network (PhaseNet, Zhu and Beroza, 2018).

3.3.3.2 Earthquake location

The location of total detected events involved either a manual or an automated approach, indeed it was performed in two main steps. First, we used the P- and S-wave automated arrival picks obtained for all the detected events from the output of the PhaseNet algorithm to perform their absolute locations. As in the previous absolute locations of the HAV, the differential time (EDT) method implemented in the non-linear

global approach (NonLinLoc) was applied and considering the 3D-velocity model (Serlenga and Stabile, 2019) of the High Agri Valley. This time, we implemented an automated, iterative four-step absolute earthquake location: at each step the NonLinLoc location is performed by taking into account station corrections obtained from the previous one. Since we already verified from previous locations in the HAV the high coherence of the used 3D velocity model in the study area, the procedure aims to progressively discard, at each iteration, the high residuals seismic phases (both P- and S- phases residuals), possibly related to wrong and/or multiple automated arrival times picks. The described absolute location enabled us to collect a total catalogue of 130 events belonging to the Castelsaraceno sequence in the investigate period. Most of the located events fall inside a seismicity cluster SW of the Pertusillo Lake close to the Castelsaraceno town (Figure 3.18). The median hypocentral depth of about 11 km is coherent with the typical values for Apenninic seismogenic faults (Figure 3.17). For each located event, the local magnitude (Ml_i) has been computed through the local magnitude scale for southern Italy (Bobbio et al., 2009) as:

$$Ml_i = \psi(Ml_i) = \psi(\log A_{ij} + 1.79 \log R_{ij} - 0.58) \qquad \text{Eq.3. 2}$$

Where, Ml_i is local magnitude of the i-th located event, A_{ij} is the peak displacement (mm) measured from the signal of the i-th event recorded by the j-th station convolved for the response function of the wood Anderson seismograph, R_{ij} the hypocentral distance (km) of the j-th station from the i-th event and ψ is the Huber mean estimator applied to the set of single-station magnitude values Ml_{ij} (Huber, 1964). For the mainshock a Ml 2.9 and a hypocentral depth of 11.45 km were estimated; while, most of the located events are characterized by negative Ml values (median = -0.3) (Figure 3.17), thus revealing the efficiency of the automated approach in detecting and locating small events which would have been missed withthe manual approach.
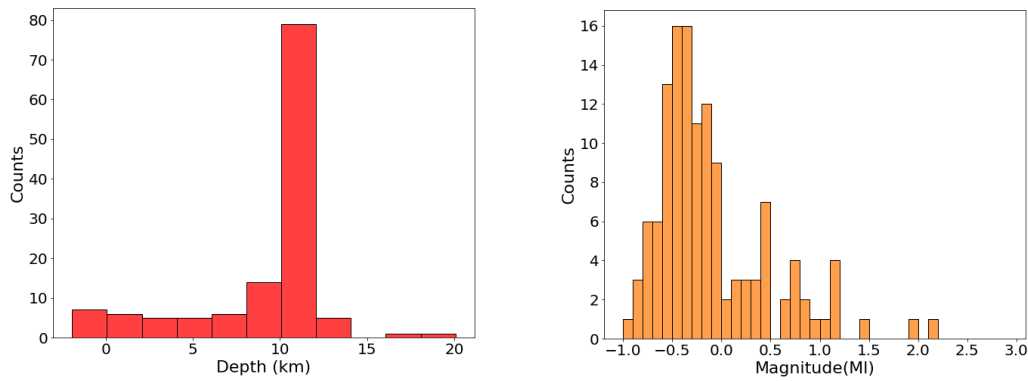
Figure 3. 17 – Hypocentral depths (left) and local magnitudes (right) estimated for the Castelsaraceno sequence.

Subsequently, the locations were refined by applying the double-difference method (hypoDD code, see Chapter 2.2.3.1), starting from the hypocentral parameters determined from the absolute locations and solving the double-difference equations in the same 3D P- and S-wave velocity model used for absolute locations. Only events with a maximum hypocentral separation of 7 km (MAXSEP parameter) and with a minimum number of 6 P- and S-wave linked differential time observations (MINLINK parameter) were relocated. Relative locations are performed by linking each event to maximum 20 other events of the cluster (MAXNGH parameter) thus obtaining a total number of 14382 differential times of P- and S-waves. The conjugate gradient

method (LSQR, Paige and Saunders, 1982) implemented in hypoDD has been used for the minimization of the double-difference residuals for pair of earthquakes at each station (Waldhauser and Ellsworth, 2000). A total dataset of 105 relative located event was obtained. Similarly to the absolute location results, the events fall inside a cluster which appears better defined after relative locations(Figure 3.18). We found that the 75% of the total number of events belonging to the cluster occurred between 7 and 10 August, thus revealing the typical behavior of a mainshock-aftershock sequence (Appendix 3.1). The projection of the relocated hypocenters in a SE-NW section (A'B' in Figures 3.18 and 3.19) revealed a NW dipping anti-apenninic fault that extends over from about 10 to about 12 km at depth below sea level not currently mapped in literature.
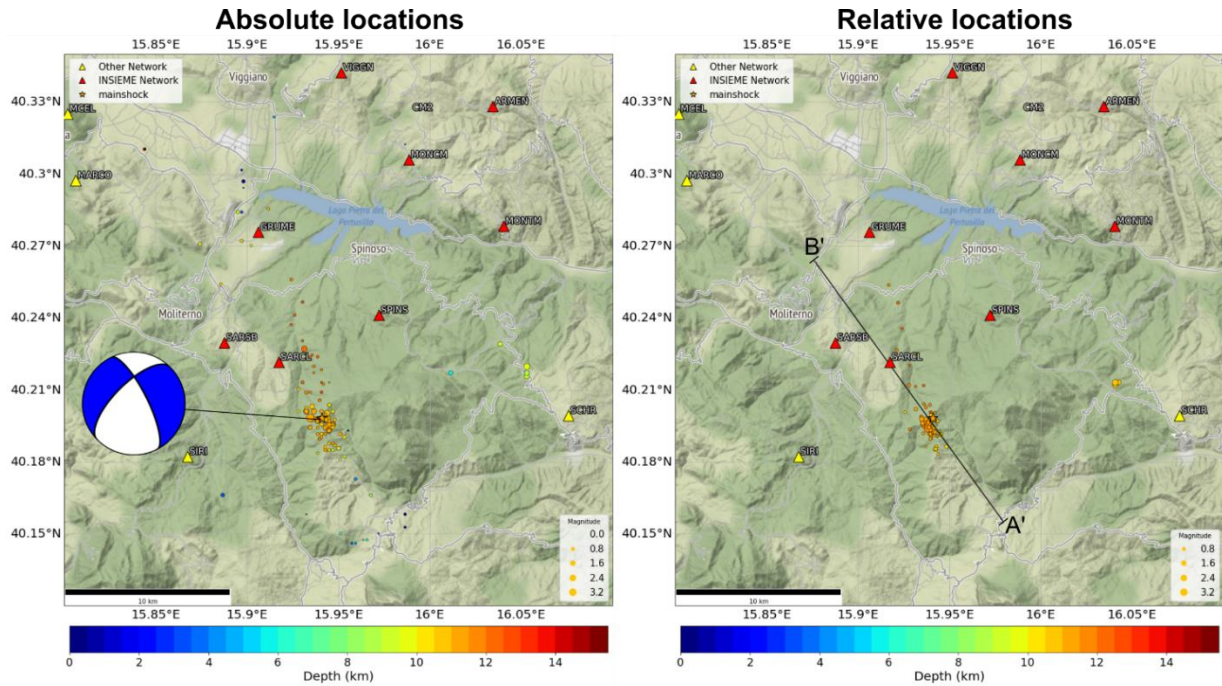
Figure 3. 18 – Comparison between the results of the absolute locations (NonLinLoc) (left) and relative locations (HypoDD) (right). The left panel also show the estiamated focal mechanism for the mainshock. The right panel shows the trace of the section in Figure 3.19.
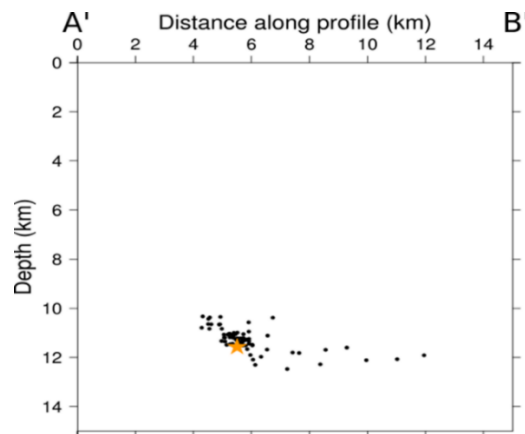
Figure 3. 19 – Cross-section A'B' which shows earthquakes hypocentral distributions obtained from HypoDD relative relocations. Its trace is reported in Figure 3.18

3.3.3.3 Mainshock source characterization

Along with the hypocentral location, the characterization of seismic source parameters enables to gather insights on the physical and mechanical processes involved in the earthquake nucleation and the energy release. We performed the source characterization of the mainshock (Ml=2.9), for which we computed both the focal mechanism and the source parameters (M_0, f_c, Δ_σ) assuming its source as representative of the sequence, at this stage. Further source analysis would be extended to the other events of the sequence in the future. We determined the focal mechanism to retrieve information on the mainshock fault plane orientation and the stress field in which it occurs. The source of small earthquakes is typically approximated to a double couple point source represented by a focal mechanism, or fault plane solution. Knowing the azimuth and take-off angle at which the ray leaves the source - computed for an assumed location and seismic-velocity model - the most used approach for its calculation is the observation of the P-wave first-motion polarities at different stations, and their projection over a reference sphere (focal sphere) around the source. The focal mechanism or "beachball" (Figure 3.20) is the one which best fits the first-motion observations. It results in the focal sphere divided into four quadrants, two in which the first motion should be away from the source (dilatation quadrants), and two in which it should be toward the source (compression quadrants). The   code (Reasemberg and Oppenheimer, 1985) was applied for the fault plane solution computation of the largest event of the sequence (Figure 3.20). Sixteen first motions (FM) were available for the inversion. We obtained a high goodness of the fit with a low resulting RMS equal to 0.09 and a high value of the station distribution ratio (STDR) equal to 0.70. The latter, measures the robustness of the fit related to the number of data lying close to the nodal planes in the solution: the higher STDR, the lower is the number of close data and the higher is the robustness of the fit (Hardebeck and Shearer, 2002). The results (Figure 3.20) showed that the sequence mainshock has normal/lateral left strike-slip faulting kinematic, with one of the two fault plane solutions (strike = 255°, dip = 55°, and rake = -30°) showing the same geometry as the microearthquakes alignment in section (Figure 3.19). Along the surface projection of this seismically inferred fault there are not previously mapped faults. The source parameters (M_0, f_c, $\Delta\sigma$) of the mainshock were retrieved through the S-wave displacement spectra inversion performed using SourceSpec algorithm (Satriano, 2021). The code estimates single events source parameters from station recordings through a ω^{-2} Brune's source spectrum inversion model. A total of 17 stations of the virtual network were used for the inversion (Table 3.1): seven stations from the HAVO Network (sampled at 250 Hz); MARCO, CAGG, SIRI, MGR, SLCN, PTRP (data sampled at 100 Hz) and MRN3, CGG3, PGN3, SRN3 (data sampled at 125 Hz). The excluded stations were automatically skipped from the algorithm since we required the inversion only for stations with a minimum spectral signal-to-ratio (spectral_sn_min) equal to 5.0 and to ignore the ones lacking in manual picks. The inversion was performed through the Reflective Trust Region optimization algorithm (Coleman and Li, 1996) particularly suitable for solving large-scale bound-constrained minimization problems and already implemented in SourceSpec.
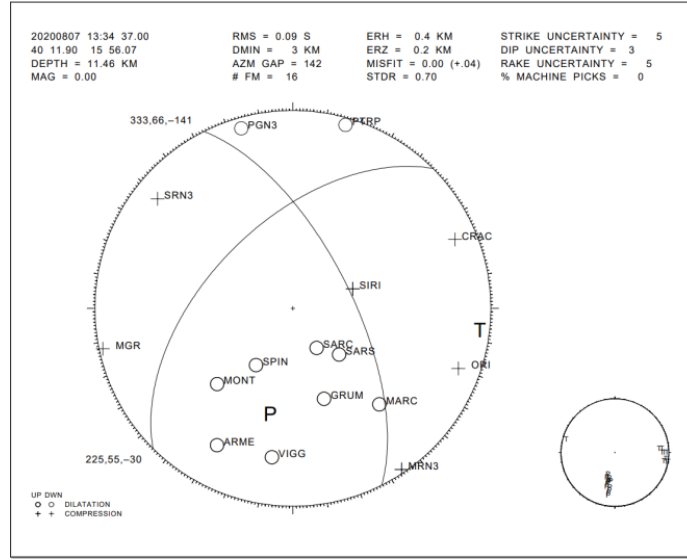
Figure 3. 20 - Fault plane solution of the sequence mainshock obtained from FPFIT code.

We assumed a crustal density of 2700 (kg/m³), a P-wave velocity (Vp) of 6.5 km/s and a S-wave velocity (Vs) of 3.4 km/s. The initial t_0^* value for the inversion (t_star_0) equal to 0.045 was calculated as:

$$t_0^* = \frac{x}{V_s Q}$$

Eq.3. 3

Where x is the mean epicentral distance calculated as the average epicentral distance of all the stations from the mainshock, *Vs* is the S-waves velocity, *Q*=200 is the mean quality factor estimated for the area (Amoroso et al., 2017) With a similar approach the *t** bounds for the inversion algorithm (t_{min}^*, t_{max}^*) where estimated equal to 0.01 and 0.085, respectively, given by:

$$t_{min}^* = \frac{x_{min}}{V_s Q_{max}}$$

Eq.3. 4

$$t_{max}^* = \frac{x_{max}}{V_s Q_{min}}$$

Eq.3. 5

Where x_{min} and x_{max} are the minimum (12 km, SARCL) and maximum (44 km, PGN3) station epicentral distance from the earthquake, respectively; while Q_{min} and Q_{max} are the minimum and maximum quality factor estimated for the area (Amoroso et al., 2017) equal to 150 and 300, respectively. The source parameters inversion was performed at each station on a 2.0 seconds-wide time window (win_length) for both S-waves and noise, choosing a start time of the noise window, respect to the P arrival time (pre_p_time), of 3.0 seconds and a start time of the S-wave window (pre_s_time), respect to the S arrival time, of 0.4 seconds (Figure 3.22). Signal waveforms were bandpass filtered at each station before the analysis. Since stations of different networks consisted of different sensors (e.g., accelerometer, short period, broadband) and were characterized by distinctive sampling frequencies, different frequency ranges for the filtering were chosen. The same minimum value of 0.1 Hz was selected, while the maximum filtering value for each station was chosen equal to the Nyquist frequency (except for the HAVO stations for which a max frequency of 110 Hz was chosen).
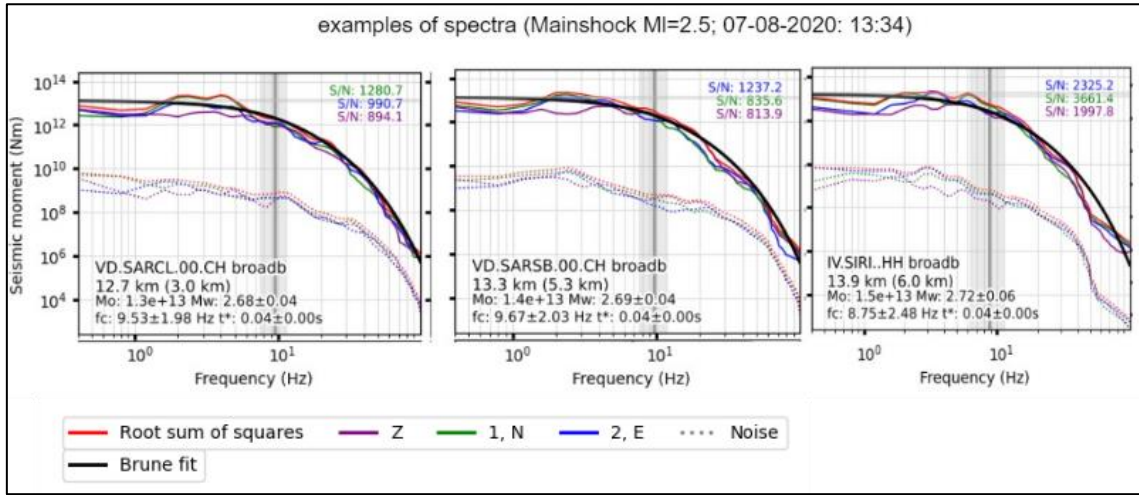
Figure 3. 21 – Example of unwindowed displacement spectra for both the noise (dashed lines) and the signals (solid lines) at three different stations (SARCL, SARSB, SIRI). Two anomalies were observed: (1) a spectral bump at 2 Hz, with a decreasing trend of the S-waves spectrum at lower frequencies; (2) an abrupt-fall of both the noise and the signal amplitude at about 40-60 Hz.

Looking at the performed inversion on unwindowed S-waves and noise displacement spectra, we observed two anomalies at many stations (Figure 3.21): (1) a spectral bump at 2 Hz, with a decreasing trend of the S-waves spectrum at lower frequencies; (2) an abrupt-fall of both the noise and the signal amplitude at about 40-60 Hz. The latter effect was investigated after checking that no mistakes were made in the signal pre-processing. Looking at stations power spectral densities (PSD) calculated for a long-time span of about one year (from 01-07-2020 to 30-05-2021 at http://services.iris.edu/mustang/) we observed its constant trend between 40-100 Hz (Figure 3.23). The anomaly, consistent with the observed abrupt fall of the spectra at those frequencies, would be related to the near-site attenuation $K(\omega)$ (see section 2.2.4.3) likely due also to the decoupling between the sensor and the soil. Since the spectra anomalies previously observed, we performed the source parameters spectral inversion in the frequency range 2-40 Hz (Figure 3.24) which provided average values of Mw, fc and Brune $\Delta\sigma$ equal to 2.6 ± 0.23, 8.82 Hz and 1.40 MPa, respectively. Referring to the relationship between Mw and Ml proposed by (Zollo et al., 2014) for the Irpinia area (Southern Apennines) for which:

$$M_w = 0.74 \, (\pm \, 0.01)Ml + 0.66 \, (\pm \, 0.02) \qquad \text{Eq.3. 6}$$

the Mw obtained for the mainshock is coherent, within the error, with the estimated local magnitude from INGV (Ml=2.9). Moreover, the obtained stress drop $\Delta\sigma$ 1.40 MPa is consistent with the stress drop estimations obtained from other authors (Festa et al., 2021; Stabile et al., 2012) for natural events in Southern Italy. For a more complete characterization of the Castelsaraceno sequence, the source parameter analysis will be extended in the future to other all the collected events.
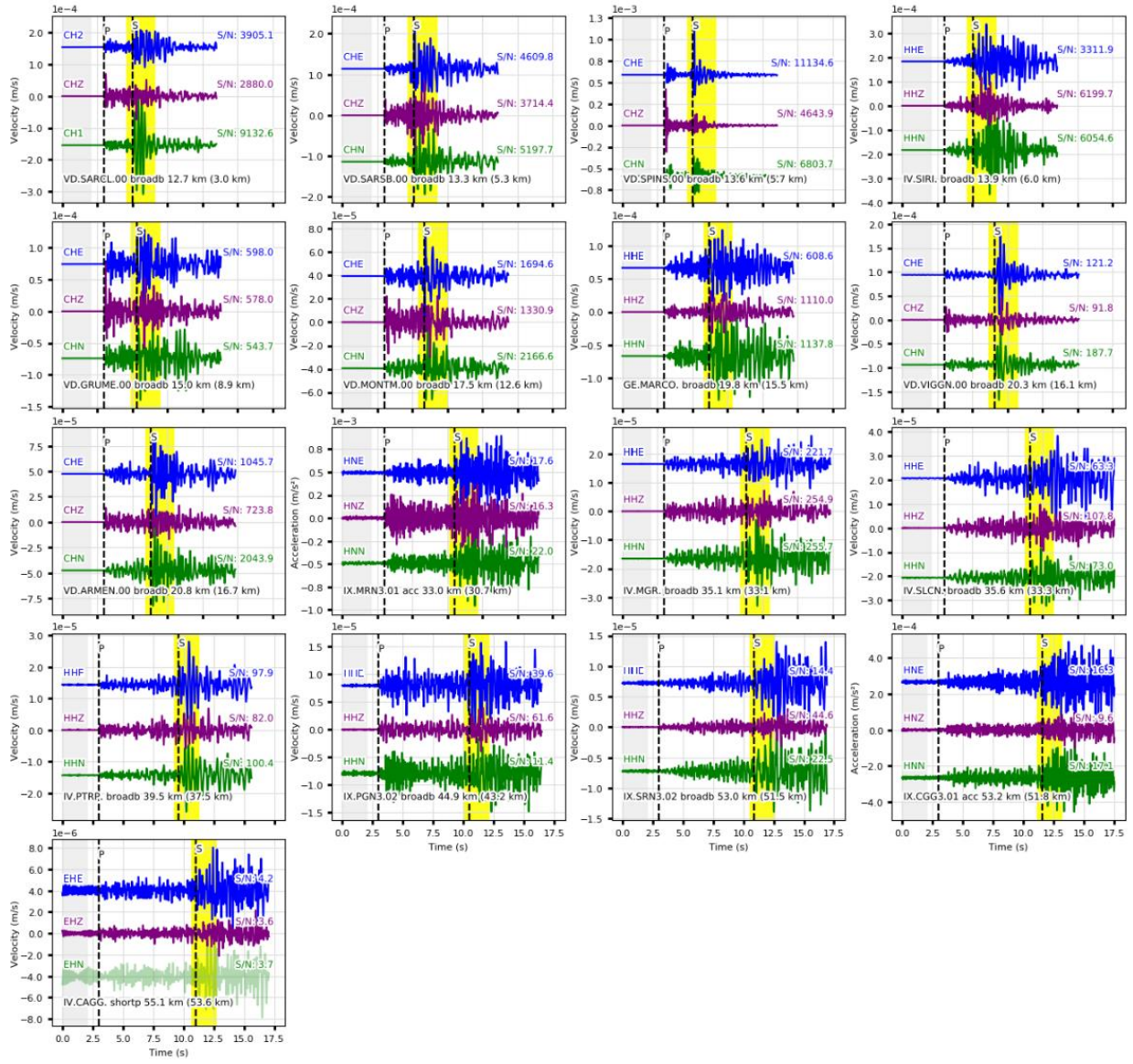
Figure 3. 22 – Mainshock waveforms and S-wave signal window (yellow band) used for source parameters estimation at different stations (SourceSpec, Satriano, 2021)
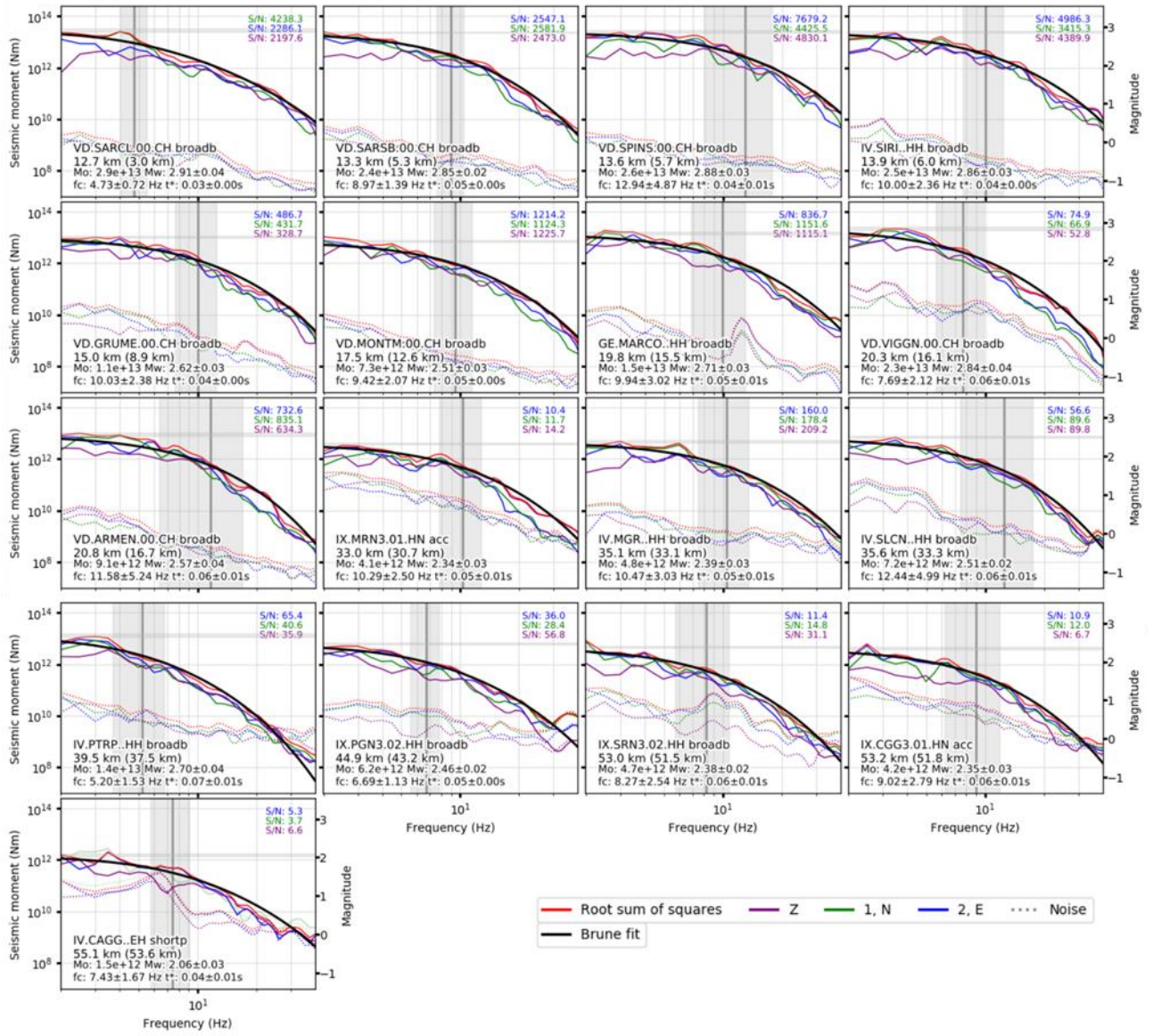
Figure 3. 23 – Results of the S-wave displacement spectra inversions performed at different stations through SourceSpec algorithm (Satriano, 2021). For each station the source parameters (M0, fc, Δσ) of the mainshock were retrieved.
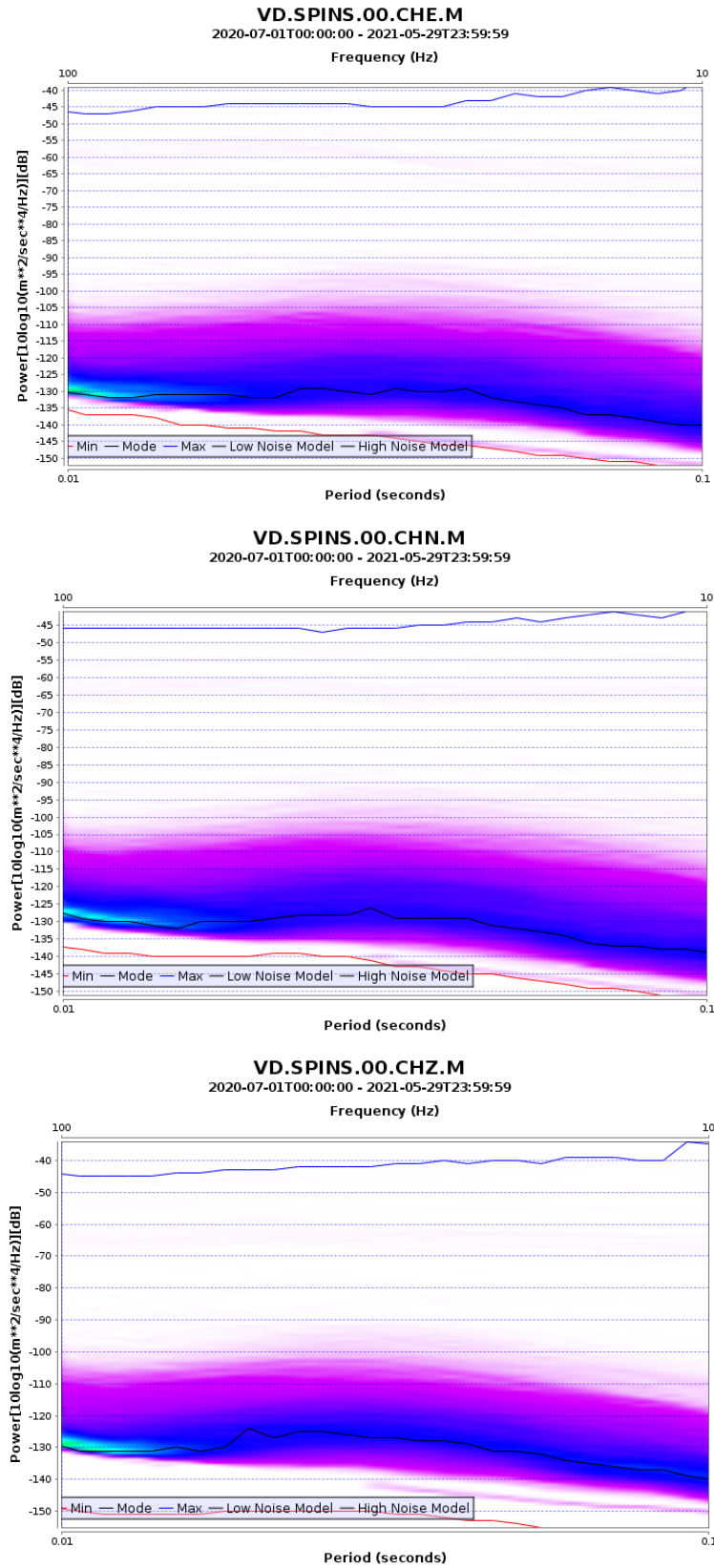
Figure 3. 24 - Power spectral densities (PSD) calculated between 01-07-2020 and 30-05-2021 (http://services.iris.edu/mustang/) for both vertical (CHZ) and horizontal (CHE, CHN) components at SPINS station.

3.4 Geochemical study

Groundwater quality depends on both natural processes (e.g., water-rock interaction, climatic conditions) and anthropogenic activities. The first comprehends mineral precipitation or dissolution, ion-exchange, redox condition, residence time, and mixing between different water types and may have a great impact on groundwater quality (Wang et al., 2018). In water-rock interaction processes, the release of elements from rocks is strictly related either to physical parameters (temperature, residence time, flow rate) or to chemical factors (e.g., weathering of rock-forming minerals, precipitation of sparingly soluble secondary minerals, and presence of reactive gas species such as CO_2) (Apollaro et al., 2020, Paternoster et al., 2017). Between natural processes, environmental redox conditions can have an important role in determining the partitioning between aqueous solution and solid phases since some elements are present in nature in different oxidation states. On the other hand, anthropogenic activities, such as rapid urbanization, industrial operations and intensive agricultural activities have caused a deterioration in water quality worldwide (Melki et al., 2020, Vespasiano et al., 2016). The main concerns on groundwater quality are ascribed to industrial activities and the related fluid injections operations, due to the possible contamination of both soil-sediment (Vengosh et al., 2014, Burton et al., 2014) and water (Vengosh et al., 2014, Jackson et al., 2013, Fontenot et al.,2013), possibly associated to spills of fluids during drilling and injection processes and to the failure of wastewater pipelines or well casings or corrosion of pipes and tanks (Pichtel, 2015). It follows that, the evaluation and management of groundwater resources require an understanding of hydrogeochemical features of the aquifers for (a) the evaluation of the groundwater origin and migration patterns; (b) establish a geochemical baseline for the long-term environmental monitoring, particularly in areas characterized by intense anthropogenic activities. In the HAV, groundwater represents an essential freshwater resource, for both drinking water supply and for agricultural and industrial purposes. Here, together with an ancient rural economy based on agriculture and breeding, an ever-increasing development of industrial activities have been undertaken related to hydrocarbon extraction, of the largest onshore oil field in Western Europe (Paternoster et al., 2021). The first comprehensive characterization and monitoring of groundwaters in HAV was developed in the context of the Agri Fluid project (Colella et al., 2003, Mongelli et al., 2003, Paternoster et al., 2005). A recent update in the hydrogeochemical characterization of the area was developed by Paternoster et al. (2021) who have investigated springs, located mainly on the right northeastern bank of the Agri river, and wells intercepting shallow porous aquifers in the Quaternary alluvial-lacustrine deposits filling the basin. Purpose of this work, is to upgrade the recent hydrogeochemical and groundwater quality observations on the HAV aquifers for a comprehensive hydrogeochemical assessment of the area, better understand the geochemical evolution of groundwater in the study area and define a sustainable groundwater management strategy, considering the growing industrialization in the HAV (Paternoster et al., 2021).

In tectonically active areas (e.g., HAV), fault systems may act as preferential conduits for natural fluids uprising at shallow levels, which may cause modifications in their pristine chemical and isotopic composition due to secondary chemical processes (Caracausi et al. 2003, 2005). Buttitta et al. (2020) demonstrated that microfracturation in stress field due to tectonic may increase the release of He from the rocks sustaining the excess of He in the natural reservoirs with respect to the steady-state diffusive He outputs. It follows that the

geochemical characterization of the HAV groundwaters, along with the assessment of the origin and mixing of fluids of different sources (mantle, crust, atmosphere) can represent an improvement in the comprehension of its crustal deformation processes. It is widely accepted that, in non-volcanic regions, thermal springs and gas vents reflect a relatively deep circulation and surface-emitted fluids can be useful indicators of deep geological processes such as seismic, magmatic or geothermal activities (Italiano et al., 2001). Moreover, groundwater level responses to earthquakes generation have been widely observed as related to co-seismic or post-seismic stress and permeability changes due to seismic activity (Nespoli et al., 2015, 2018). Therefore, in tectonically active regions, the characterization of thermal fluids with a marked deep component, can represent an important tool in the comprehension of seismological phenomenon in the context of the complex relationships among rocks, faults, and fluids involved in earthquake generation (Thomas, 1988; Toutain and Baubron, 1999). After the 1857 Val d'Agri earthquake (MW = 7.0, Boschi et al., 2000), an unusual natural emission of oily fluids took place close to Tramutola (Balderer and Martinelli, 1995, and references therein) and strong anomalies in the fluids flow rate and chemical gas composition were recorded at the Tramutola well in coincidence with the occurrence of 1996 ML 4.9 Irpinia earthquake. Italiano et al. (2001) explained the recorded variations to an earthquake-related transient variation of the permeability in the shallow crust involved in the seismogenic process. However, no such effects were recorded in conjunction with a later ML 5.5 earthquake occurred in 1998 at a comparable epicentral distance from the well. Tramutola well in the HAV represents a key natural laboratory to better understand the complex coupling effects between mechanical and fluid-dynamic processes in earthquake generation processes and to investigate the relationship between geochemical signatures in the fluids and crustal deformations. For this reason, a particular attention was given to the Tramutola site. Moreover, a multi-parametric device for the continuous geochemical monitoring purposes was installed at the well during the last few months of the PhD project. In this work, preliminary observations are reported.

In summary, the objectives of the applied geochemical approach on the High Agri Valley study area are: (1) to determine the geochemical processes controlling the groundwater chemical composition; (2) to define a geochemical conceptual model regarding to fluid origin (deep vs shallow) and mixing processes by means isotopic data; (3) to evaluate the springs of potential interest (i.e. those least affected by shallow and/or deep circulation of fluids), that can be equipped with high-frequency multiparametric monitoring and on data cross analysis with the aim of verifying the dynamic and time evolution of natural processes in relation to the seismicity of the area; (4) to define the geogenic and anthropogenic influences on groundwater quality.

3.4.1 Sampling and analytical procedures

In this work, a total of 39 sample sites were identified for the hydrogeochemical characterization of the High Agri Valley (Figure 3.25). A total of 35 spring water samples was collected, homogeneously located at both the left and the right hydrological sides of the HAV. It occurs at the border hydro-structures of the valley, arising at the geological contacts between the Pre-Quaternary carbonate rocks of the Piattaforma Appenninica and Lagonegro Units (mainly outcropping on the left bank) and the Mesozoic-Cenozoic Ligurian units and their sedimentary cover (e.g., Albidona Formation) (mainly located on right bank of the valley). The

investigated springs are generally characterized by high flow rates, up to 300 l/sec ("Sorgente Aggia"). Most of springs are managed by the Acquedotto Lucano public agency for drinking purposes. In addition, water from 4 piezometers (PZ11, PZ5, PZ6, PZ7), located in the southeastern area of the HAV (Figure 3.25) between Centro Oli Val d'Agri (COVA) (where oil & gas extraction activity is carried out by the Eni S.p.a.) and the Costa Molina 2 (CM2) well (where the related wastewater injection for disposal purposes is operated) were sampled. All the piezometers are managed by ENI S.p.a. for environmental monitoring purposes. The location and water depth from the ground level of the analyzed piezometers are summarized in Table 3.4.
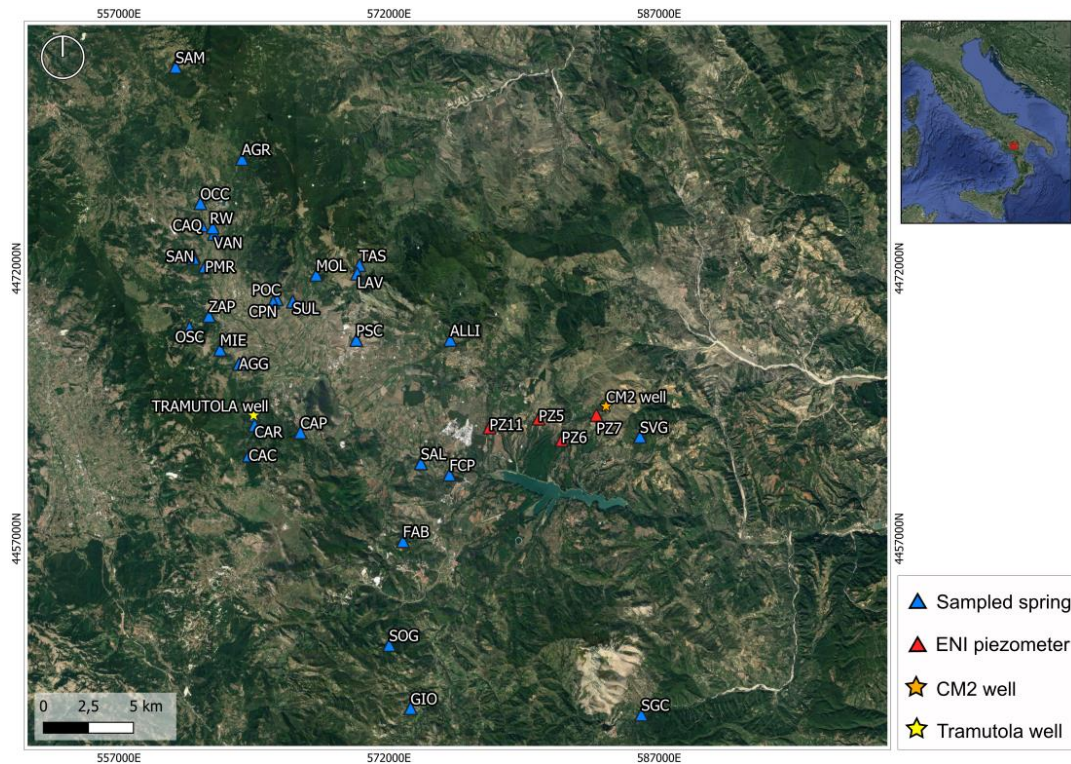
Figure 3. 25 – Sampling site locations in the High Agri Valley study area.

Piezometer	Municipality	Lat UTM-WGS84	Long UTM-WGS84	UTM Zone UTM-WGS84	Quote m	Water table m
PZ11	Montemurro	4462817	577590	33T	582	15.3
PZ6	Viggiano	4462110	581597	33T	637.27	20
PZ5	Viggiano	4463309	580324	33T	708.31	32.9
PZ7	Viggiano	4463506	583526	33T	951.17	-

Table 3. 4 – ID, location and water table depth of the analyzed piezometers managed by Eni S.p.a.

Particular attention was given to the Tramutola site (TRA). It is the AGIP N2 deep well drilled for oil and gas exploration in the HAV in the period between 1936-1939. The well has a depth of about 404 m, it crosses clays, silicic clays and silicic limestones belonging to the Lagonegro units (ENI, 1972; Barbier et al., 1998, Italiano et al., 2001) and it encountered an artesian aquifer, such that it is characterized by the continuous emission of water and bubbling gas. The thermal water has a constant temperature of about 28°C with a Na-

HCO$_3$ hydrofacies (Italiano et al., 2001, Colangelo et al., 2007); while the emitted gases are estimated to be CH$_4$-dominated (more of the 80%) with high concentrations of nitrogen (e.g., Caracausi and Paternoster, 2015). In this work, a total output of the emitted gases from the Tramutola well was estimated through anemometric measurements performed through a *testo 400* control unit provided of a vane anemometer (Ø=100mm) for flow velocity measurement (±0.01 m/s resolution). Moreover, a multi-parametric device (OTT ecoLog 800 3G) was installed to measure Temperature, pH and water level with high sampling period: (a) 1-minute for water level and (b) 5-minutes for both electrical conductivity and Temperature. Water level data are collected with a ± 0.01 m resolution pressure probe equipped with a relative pressure probe cell, through hydrostatic pressure measurements of the water column. The specific electrical conductivity with a ± 0.001 µS/cm resolution using a 4-graphite electrode conductivity sensor integrated with a thermometer (resolution ± 0.001 °C). The preliminary results of the continuous monitoring will be discussed in detail in the next paragraphs. All the investigated samples were collected during the winter season, between October 2020 and February 2021.

3.4.1.1 Sampling, field measurements, and laboratory analyses

For the sampling, high-density polyethylene (HDPE) bottles of different aliquots (50, 100, 1000 ml) with watertight caps were used. A plastic funnel was used to collect free gases samples a the Tramutola sampling site. Temperature, pH, electric conductivity (EC, measured at 25° C), and redox potential (Eh) were measured in situ using a high-resolution multiparametric probes (WTW Multi 3630 IDS), after calibration in the laboratory by means WTW traceable standard solutions. Total alkalinity was determined in situ by titrating unfiltered samples with 0.01 N HCl with methyl-orange indicator. Major anions (Cl$^-$, NO$_3^-$, NO$_2^-$, SO$_4^{2-}$, F$^-$ and Br$^-$) and cations (Na$^+$, K$^+$, Mg^{2+}, Ca^{2+} and NH$_4^+$) were determined by ionic chromatography through a Dionex Aquion RFIC Ion Chromatographer at Environmental Geochemistry Laboratory (University of Basilicata, Potenza) (Figure 3.26). For the evaluation of the analytic error in the ion concentration measurements, the electroneutrality was estimated as follows:

$$\text{Electroneutrality (e)} = \frac{[CAT - AN]}{CAT + AN} \times 100 \ (\%)$$

where CAT and AN represent the sum of the measured cations and anions, respectively. Values above ±5% indicate an analytical error in the measurements (Matthess, 1982). This threshold was never exceeded in the analyses performed in this study.
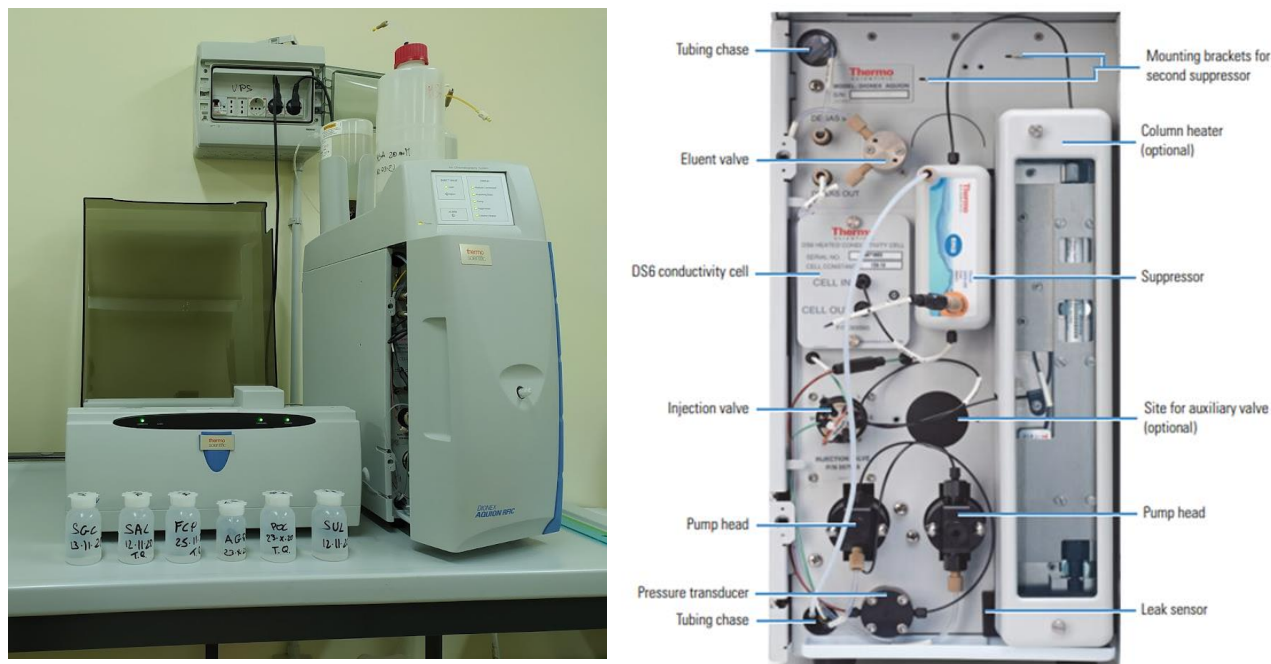
Figure 3. 26 - Dionex Aquion RFIC Ion Chromatographer for major elements analysis of the collected water samples

Minor and trace elements (Li, Si, Al, B, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, As, Se, Rb, Sr, Zr, Cd, Sn, Sb, Ba, Cs, U, Hg, Pb) (Appendix 3.3) were determined by inductively Coupled Plasma Mass Spectrometer (ICP-MS) through a Perkin Elmer Sciex ELAN 9000 ICP/MS instrument at Activation Laboratories Ltd (Actlabs, Ancaster, Ontario, Canada). δD and $\delta^{18}O$ values were performed by laser spectroscopic analysis (LGR DLT-100) at the laboratories of INGV (Istituto Nazionale di Geofisica e Vulcanologia), Sezione di Palermo. Analytical precision is ±1‰ for δD and ± 0.1‰ for $\delta^{18}O$. The chemical composition of dissolved gases was determined using the methodology proposed by (Capasso and Inguaggiato, 1998) based on the partitioning equilibrium of gaseous species between the liquid and the gas phase. The analytical determinations of dissolved gases were carried out using an Agilent GC 7890B gas chromatography system with argon as carrier gas and equipped with a double detector (TCDFID). The uncertainties of measurements are within 5%.

The isotopic composition of total dissolved inorganic carbon (TDIC) was determined using the methodology proposed by Capasso et al. (2005), and it is based on the chemical and physical stripping of CO_2 by the addiction of 150–200 μl of 100% H_3PO_4. The isotopic values were measured using a Thermo Delta V Plus isotope ratio mass spectrometer coupled to a GasBenchII and the results are reported in δ ‰ vs. V-PDB standard. The standard deviation of $^{13}C/^{12}C$ ratio is ±0.15‰. The carbon isotopic composition of CO_2 ($\delta^{13}C_{CO2}$) was determined using a Thermo Delta XP Isotope Ratio Mass Spectrometer coupled with a Thermo Scientific™ TRACE™ Ultra Gas Chromatograph. Separation prior to analysis was done through a 30-m Q-plot column (i.e., of 0.32 mm). The resulting $\delta^{13}C_{CO2}$ values are expressed in per mil notation with respect to the international Vienna Pee Dee Belemnite (VPDB) standard and analytical uncertainties of ±0.15‰. The carbon and deuterium isotopic composition of CH_4 ($\delta^{13}C_{CH4}$ and δD_{CH4}) was determined using a Thermo TRACE GC interfaced to a Delta Plus XP gas source mass spectrometer and equipped with a Thermo GC/C III (for Carbon) and with GC/TC peripherals (for Hydrogen). The $^{13}C/^{12}C$ ratios are reported as $\delta^{13}C_{CH4}$ values with respect to the VPDB standard, and $^2H/^1H$ ratios are reported here as δD_{CH4} values with respect to the

Vienna Standard Mean Ocean Water (VSMOW) standard. The analytical uncertainty of the measurements was 0.1‰.

^{3}He, ^{4}He and ^{20}Ne, and the ^{4}He/^{20}Ne ratios were determined by separately injecting He and Ne into a split flight tube mass spectrometer (GVI-Helix SFT, for He analysis) and then into a multicollector mass spectrometer (Thermo-Helix MC plus, for Ne analysis), after standard purification procedures (Correale et al., 2012). The analytical error was generally less than 1%. Measured He isotopes values are reported as R/RA, where R = ^{3}He/^{4}He, measured in the sample, and RA = ^{3}He/^{4}He of air (1.4×10^{-6}). Helium concentrations in the analyzed HAV samples (TRA 27-06-2020; TRA 28-02-2021; SUF) range from 1.9 to 508.10 ppm. ^{4}He/^{20}Ne ratio is 120–1,690 times higher than that measured in air (^{4}He/^{20}Ne = 0.318), confirming very low air contamination, and validating the accuracy of the results. The R/RA values were corrected for atmospheric contamination based on the ^{4}He/^{20}Ne ratio (Sano & Wakita, 1985) as follows:

$$\frac{R}{R^a} = \frac{\left(\frac{R_m}{R_a}\right)\cdot\left(\frac{^4He}{^{20}Ne}\right)_m - \left(\frac{^4He}{^{20}Ne}\right)_a}{\left(\frac{^4He}{^{20}Ne}\right)_m - \left(\frac{^4He}{^{20}Ne}\right)_a}$$

Eq.3. 7

where subscripts m and a refer to measured and atmosphere theoretical values, respectively; while, (^{4}He/^{20}Ne)$_a$ = 0.318) (Ozima & Podosek, 2002). The correction on the ^{3}He/^{4}He ratio is small or negligible when (^{4}He/^{20}Ne)$_m$ >> (^{4}He/^{20}Ne)$_a$. The Ar elemental and isotopic compositions (^{36}Ar, ^{38}Ar, and ^{40}Ar) were quantified in a multicollector mass spectrometer (Helix MC-GVI). The analytical uncertainty (1σ) for single ^{40}Ar/^{36}Ar measurements was <0.1%. ^{40}Ar was corrected for air contamination (^{40}Ar*) in samples showing ^{40}Ar/^{36}Ar > 315, assuming that the ^{36}Ar present derived from atmosphere, as follows:

$$^{40}Ar^* = {}^{40}Ar_{\text{sample}} - {}^{36}Ar_{\text{sample}} \cdot \left(\frac{^{40}Ar}{^{36}Ar}\right)_{\text{air}}$$

Eq.3. 8

All the gas samples were analyzed at the laboratories of INGV (Istituto Nazionale di Geofisica e Vulcanologia), Sezione di Palermo, for their chemistry and for the isotopic compositions of noble gases (He, Ne, and Ar), C of CO_2, and C and H of CH_4. Analyses are reported in Tables 3.8, 3.9, 3.10.

3.4.1.2 Geochemical Modeling

The Geochemist's Workbench (GWB 8.0, Bethke et al., 2010) and PHREEQC (Parkhurst, 1995) software were used to calculate saturation indices (SI) for the main mineral phases involved in the rock-water interaction processes of the aquifer, the TDIC (Total Dissolved inorganic Carbon), the electron activity (pe) estimation. The GWB is an integrated geochemical modelling package used for balancing chemical reactions, calculating stability diagrams and the equilibrium states of natural waters, tracing reaction processes, modelling reactive transport, plotting the results of these calculations, and storing the related data. Saturation indices (SI) are defined as:

$$SI = \log\left(\frac{IAP}{K_t}\right)$$

Eq.3. 9

Where, IAP is the ion activity product of the mineral–water reaction and Kt is the thermodynamic equilibrium constant at the measured temperature. Thus, SI = 0 indicates a thermodynamic equilibrium state, and values > 0 denote oversaturation and values > 0 denote oversaturation. The electron activity (pe) indicates the tendency of a solution to donate or accept a proton and is calculated as:

$$pe = -\log a_{e^-}$$

Eq.3. 10

Where, a_{e^-} is the electron activity. Low values of pe reflect strong tendency for the solution to donate protons, thus that the solution is reducing.

3.4.2 Results

3.4.2.1 Water geochemistry

Temperature, pH, EC, Eh, dissolved oxygen, the Total Dissolved Solids (hereafter TDS) and major elements of the analyzed waters are listed in Appendix 3.2, together with the geographic coordinates (UTM-WGS84) of the sampling points. The TDS was calculated as the sum (mg/l) of the major cations (NH_4^+, Li^+, Na^+, K^+, Mg^{2+}, Ca^{2+}) and anions (F^-, Cl^-, Br^-, NO_3^-, SO_4^{2-}, NO_2^-, HCO_3^-) in the samples. Physico-chemical parameters distributions of the 39 collected samples are shown Figure 3.27. The investigated springs are characterized by cold waters, with temperature values between 7.4°C and 16.8°C, except for the Tramutola samples characterized by high temperature (T=28°C). Most of the samples show an alkaline pH (median 7.64) with a maximum value of 8.81 (OCC) and a minimum value of 6.81 (PZ7). The TDS (median value of 371.2 mg/L) and EC values reflect slightly saline waters with EC values ranging between 298 - 1178 μS/cm (median EC=458 μS/cm). Three samples showed anomalous values of EC and TDS, thus showing a higher saline content: PZ7 (EC=1675 μS/cm) and the Tramutola ones (EC=2590 for TRA 13.11.2020; EC=2510 μS/cm for TRA 22.02.2021). Most of the analyzed samples exhibit mainly an oxidizing environment with Eh values predominantly positive (Eh median value equal to 242.8 mV) except for the SUF sample for which a negative value (Eh=-113,5 mV) was measured, thus suggesting a reducing circulation environment. A predominantly oxidizing environment for the analyzed waters is indicated also from their dissolved oxygen concentrations, which median value is equal to 9.310 mg/l. However, the O_2 shows the highest variability in the dataset, as testified from the skewness of its distribution (Figure 3.27), which show a group of outliers (PZ11, PZ6, PZ5, PZ7, SUL, SUF, TRA samples) with low values ranging between 1.0 and 4.4 mg/l. Physico-chemical parameters distributions highlighted a substantial difference of the Tramutola well samples compared to the rest of the dataset. This was confirmed also by their hydrofacies in the Piper Diagram (Figure 3.28. It shows a predominant bicarbonate alkaline-earth composition for all the investigated samples, except for the Tramutola ones (red dots in Figure 3.28) belonging to sodium-bicarbonate hydrofacies. The bicarbonate alkaline-earth waters show a slight increasing trend for the Mg/Ca ratios ranging between 0.03 (CAQ) and 0.43 (AMO) which remain lower than the typical value of pure dolomite (0.6) (Celico, 1988). This agrees with the main outcropping lithological units in the area consisting of dominant Ca-rich host rocks (Limestone) and,

for a lesser extent, of Ca-Mg rich host rocks (Dolomitic limestone and Dolostone) mainly outcropping on the southwestern side of the High Agri Valley.
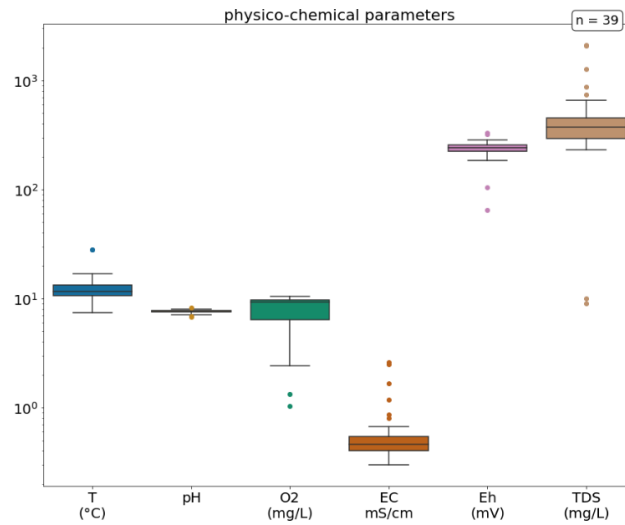
*Figure 3. 27 – Phisico-chemical parameters (from the left: Temperature, pH, Dissolved oxygen, Electrical conductivity, Eh, Total Dissolved Solids) distributions boxplots of the 39 samples. For each distribution, the 25th percentile is assumed as the first quartile (Q1) and the 75th percentile as the third quartile (Q3) and defined the interquartile range (IQR=Q3-Q1); while the minimum and maximum values are found at the end of each box 'whisker'. Outlier data points are located outside the whiskers of the box plot. These are observations numerically distant from the rest of the data which value is outside 1.5 times the interquartile range above the upper quartile and below the lower quartile (Q1 - 1.5 * IQR or Q3 + 1.5 * IQR).*
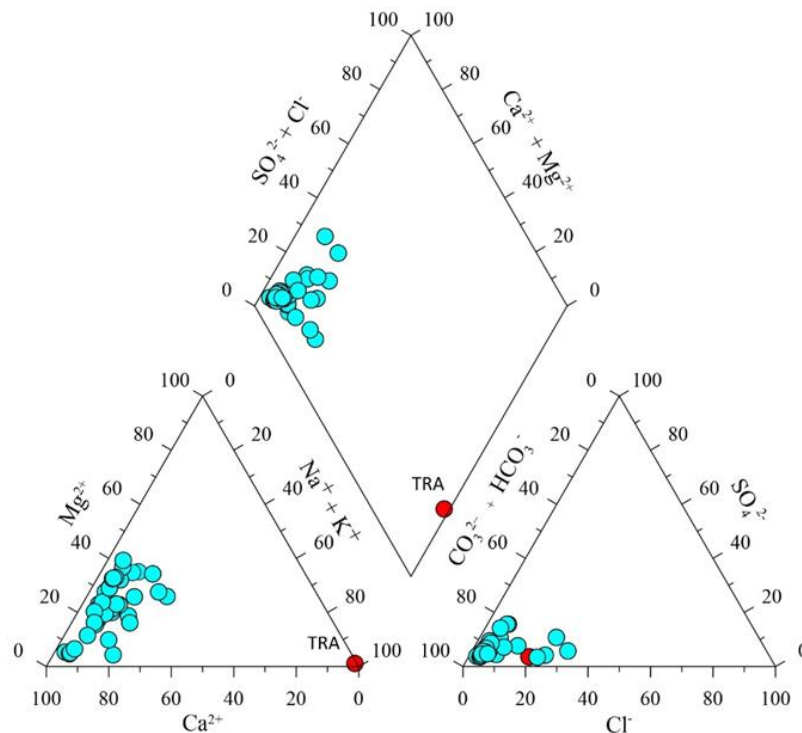
Figure 3. 28 - Piper Diagram for the reconstruction of the samples hydrofacies.
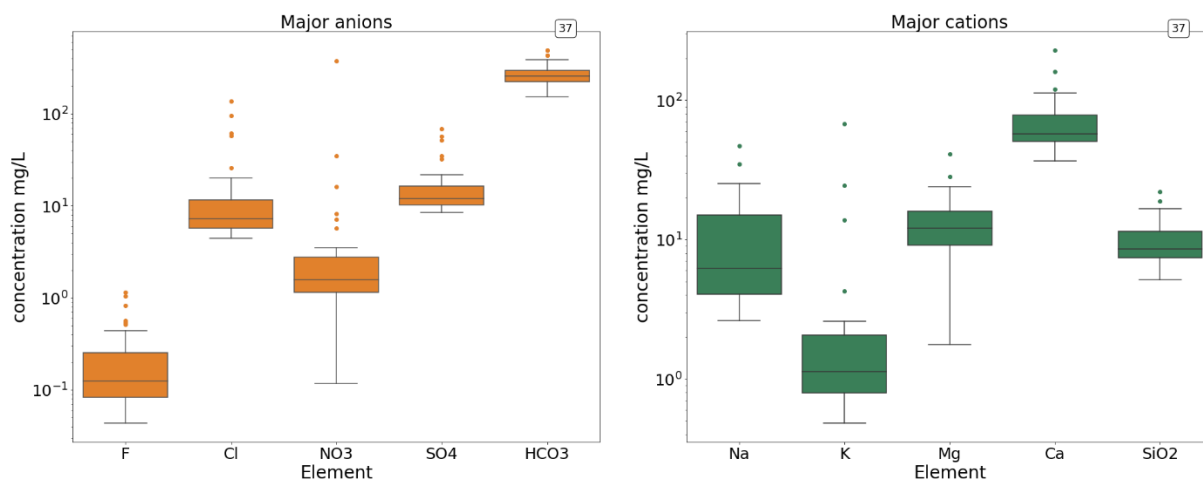
Figure 3. 29 – Major anions (on the left) and major cations (on the right) distribution for 37 HAV samples. TRA outlier samples were excluded from the distributions.

Box-and-whisker plots of the major constituents (cations and anions) are showed in Figure 3.29. TRA samples were excluded from the distributions and will be separately analyzed later in the work. Nitrate and magnesium count the greatest concentration variations. Among cations, Ca^{2+} is the most abundant (median 57.4 mg/l), followed by Mg^{2+} (median 12 mg/l), SiO_2 (median 8.5 mg/l), and Na^+ (median 6.2 mg/l). K^+ values are generally low (median 1.1 mg/l) except for PMR (median 24.4 mg/l) and FAB sample having the highest value (median 67.7 mg/l). Bicarbonate is the main anion in solution with concentrations ranging between 153 and 488 mg/l, with an average median value of 256.2 mg/l. Sulphate contents are not highly variable (median 12.0 mg/l) except for higher concentration values recorded at PZ6 (median 68.8 mg/l), PZ7 (median 56.9 mg/l) and PZ11 (median 51.2 mg/l). F^- is the minor constituent of the analyzed waters with the lowest concentrations (median 0.1 mg/l). Cl^- show mainly low concentrations with a median value of 7.2 mg/l. However, high values of Cl^- were recorded at PZ5 (median 136.8 mg/l), PZ7(median 95.4 mg/l), FAB (median 60.9 mg/l) and LAV (median 57,4 mg/l) samples. Most of the investigated water samples have low NO_3^- contents ranging from below instrumental detection limit (1.9 mg/l) to 16.2 mg/l. Only two water samples (PZ5, PZ7) are outliers, with NO_3^- values of 34.9 mg/l and 373.9 mg/l, respectively. All the analyzed trace and ultratrace elements are reported in Appendix 3.3. $\delta^{18}O$ and δD showed narrow ranges of variability between -6.4 and 8.9 ‰ and between -44 and -55 ‰, respectively. $\delta^{13}C$-TDIC was determined for 19 samples (SGC, POC, OCC, CAQ, PMR, VAN, MOL, PSC, OSC, AMO, AGG, MIE, FAB, SOG, SAM, CAP, ALLI, SUL, SUF). It ranges between -8.3 and -15.30 ‰ PBD. Isotopic data along with the saturation indices of calcite, dolomite, and albite are provided in Appendix 3.4.

3.4.2.2 Gas geochemistry

Table 3.5 lists the analytical results of the gas-chromatographic analyses performed on 10 samples (SAM, OSC, AGG, POC, MOL, PSC, ALLI, SUF, PZ7, PZ11). The compositions of the dissolved gases are reported in cm^3 STP/L of water. These were calculated starting from the gas-chromatographic analyses and considering the solubility coefficients (Bunsen coefficient "β", cc_{gas}/mL_{water} STP) of each gas species, the volume of gas

extracted (cm^3) and the volume of the water sample, following the equation (Martorelli et al., 2016, Italiano et al., 2013):

$$G_c = \{[G_{gc}] * V_{\gamma_e} + (G_{gc} * \beta_G * VW)\} VW^{-1} * V_{\gamma_e} * V_{\gamma_i^{-1}}/100 \qquad \text{Eq.3. 11}$$

where G_C is the concentration of the selected gas species, G_{gc} is its concentration measured by gas chromatography (volume %), and $V\gamma e$ and $V\gamma i$ represent the extracted and the introduced gas volumes, respectively, while VW is the volume of the analyzed water sample. All volumes were carefully measured at the equilibration temperature. In addition, two free gas samples were collected at the Tramutola well (TRA 27.06.2020 and TRA 28.02.2021). We calculated the dissolved gases compositions as percentage of gas volume in equilibrium with water (% vol). The CO_2 concentration in the collected samples ranges from 0.3 to 5.0% vol. Besides CO_2, CH_4 (8.07×10^{-5} to 10.4×10^1 %vol), He (1.69×10^{-2} to 7.84×10^{-2} %vol), H_2 (1.44×10^{-2} to 7.22×10^{-1} %vol), N_2 (76.4 to 90.6 %vol), and O_2 (3.87×10^{-1} to 2.34×10^{-1} %vol) are present in the gas samples. O_2 amounts are much lower in the TRA free gas (0.06-0.07%). The average free gas composition in Air is reported in Table 3.5 for comparison. The ternary plot (Figure 3.30) shows that all samples show an O_2/N_2 ratio lower than the typical ASW values. Oxygen depletion is particularly marked for PZ7, SUF, TRA and PZ11 samples. The degradation of organic matter in oxic environments, which consumes oxygen and produces carbon dioxide and water through the following reaction:

$$CH_2O + O_2 \rightarrow CO_2 + H_2O$$
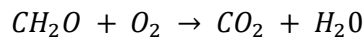
can probably justify these data. However, process ascribable to redox reactions between the gases and the surrounding rocks cannot be excluded. Between the analyzed samples, SUF shows the highest enrichment in N_2 with respect to O_2, which is probably due to a deeper and longer hydrogeological circuit. TRA sample cluster close to the CO_2 vertex, indicate a CO_2 marked contribution, probably linked to external source (i.e., deep).

The CH_4–N_2–CO_2 ternary diagram is shown in Figure 3.31. Most of the HAV groundwater samples show low concentrations of methane, except for SUF and TRA that rare CH4-rich. The variable N_2/CO_2 ratios measured in other samples is due to oxygen consumption related to organic activity in the surficial environment, as previously reported.
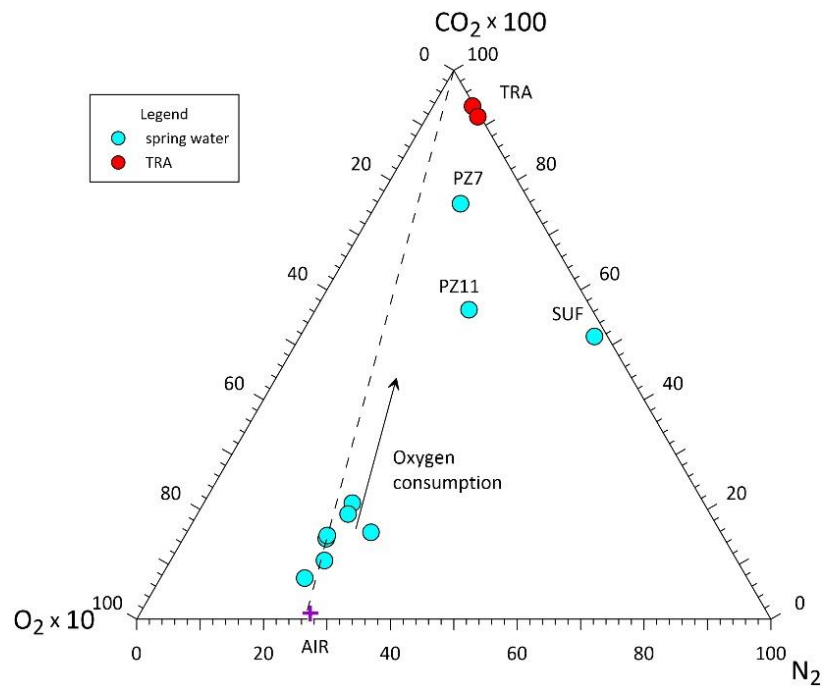
Figure 3. 30 - O$_2$–N$_2$–CO$_2$ triangular plots for dissolved gases (blue dots) and free gases (red dots). The dissolved gas composition in Air-Saturated Water (Ozima and Podosek, 2002) is reported for comparison (purple cross). The dashed line represents the O$_2$/N$_2$ ratio in atmospheric air.
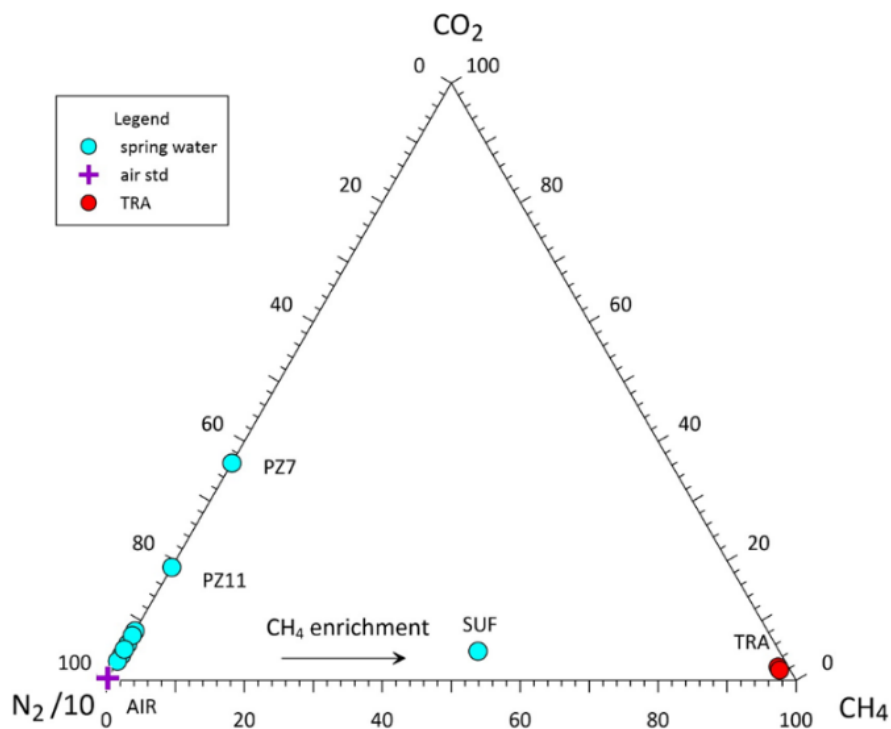
Figure 3. 31 – CO$_2$ –N$_2$ –CH$_4$ triangular plot for dissolved gases (blue dots) and free gases (red dots).

Sample	Sampling data	He ccSTP	He %	H$_2$ ccSTP	H2 %	O$_2$ ccSTP	O2 %	N$_2$ ccSTP	N2 %	CO ccSTP	CO %	CH$_4$ ccSTP	CH4 %	CO$_2$ ccSTP
SAM	22/10/2020	b.d.l.	b.d.l.	2.02E-03	0.13	5.65	21.65	10.38	77.83	b.d.l.	b.d.l.	7.02E-05	2.36E-04	3.72
OSC	22/10/2020	b.d.l.	b.d.l.	4.09E-04	0.02	5.96	18.99	12.89	80.28	1.26E-04	5.33E-04	4.88E-05	1.37E-04	6.33
AGG	22/10/2020	b.d.l.	b.d.l.	2.73E-04	0.01	6.72	21.04	12.88	78.61	1.05E-04	4.38E-04	3.29E-03	9.11E-03	3.04
POC	22/10/2020	b.d.l.	b.d.l.	1.22E-02	0.72	5.40	19.22	11.40	80.06	3.63E-04	1.71E-03	6.89E-05	2.17E-04	4.95
MOL	22/10/2020	b.d.l.	b.d.l.	6.01E-04	0.04	6.59	23.38	11.09	76.36	6.20E-04	2.91E-03	2.92E-05	9.23E-05	1.90
PSC	22/10/2020	3.19E-04	0.04	1.29E-03	0.07	6.49	21.62	12.06	77.84	b.d.l.	b.d.l.	b.d.l.	0.00E+00	4.25
ALLI	22/10/2020	b.d.l.	b.d.l.	4.30E-04	0.02	4.55	15.90	12.38	83.65	2.39E-04	1.10E-03	4.68E-05	1.46E-04	3.46
SUF	05/03/2021	5.27E-04	0.08	1.57E-03	0.11	0.09	0.39	10.97	88.23	b.d.l.	b.d.l.	2.89E+00	1.04E+01	6.78
PZ7	05/03/2021	1.62E-04	0.02	b.d.l.	0.00	2.55	7.33	15.55	87.66	b.d.l.	b.d.l.	1.63E-03	4.11E-03	49.50
PZ11	05/03/2021	b.d.l.	b.d.l.	1.18E-03	0.06	2.51	7.27	15.95	90.61	b.d.l.	b.d.l.	3.18E-05	8.07E-05	20.69
AIR							20.95		78.08					0.36

Table 3. 5 – Dissolved gas He, H$_2$,0$_2$, N$_2$, CO, CO$_2$, CH$_4$, composition (ccSTP and %vol) from gas chromatography analysis. The free gas composition in Air (Ozima and Podosek, 2002) is reported for comparison.

3.4.3 Discussion

3.4.3.1 Water-rock interaction processes

The investigated waters showed that HCO_3^- is the main anion dissolved and it is higly correlated with TDS values (Figure 3.32, R=0.956, p-value < 0.0001). Exception are PZ5 and PZ7 samples, which show an enrichment in TDS values, probably due to high NO_3^- concentrations. The interpretation of the anomalous NO_3^- content in these samples will be further discussed in the next paragraphs.
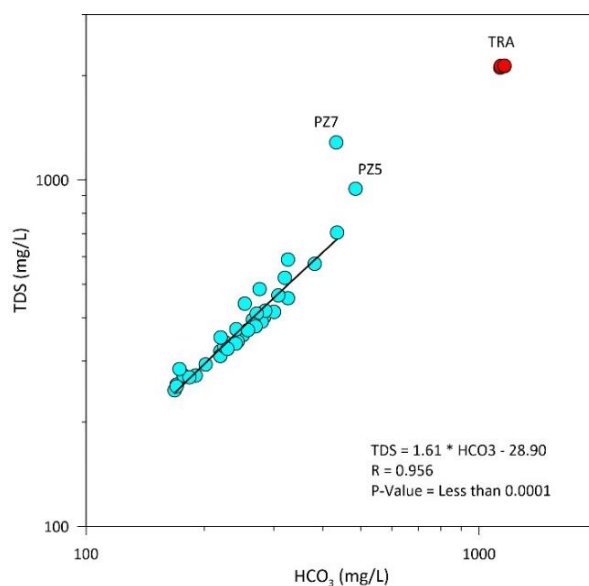
Figure 3. 32 – Relationship between HCO3- and TDS in HAV samples and Tramutola thermal waters (red dots).
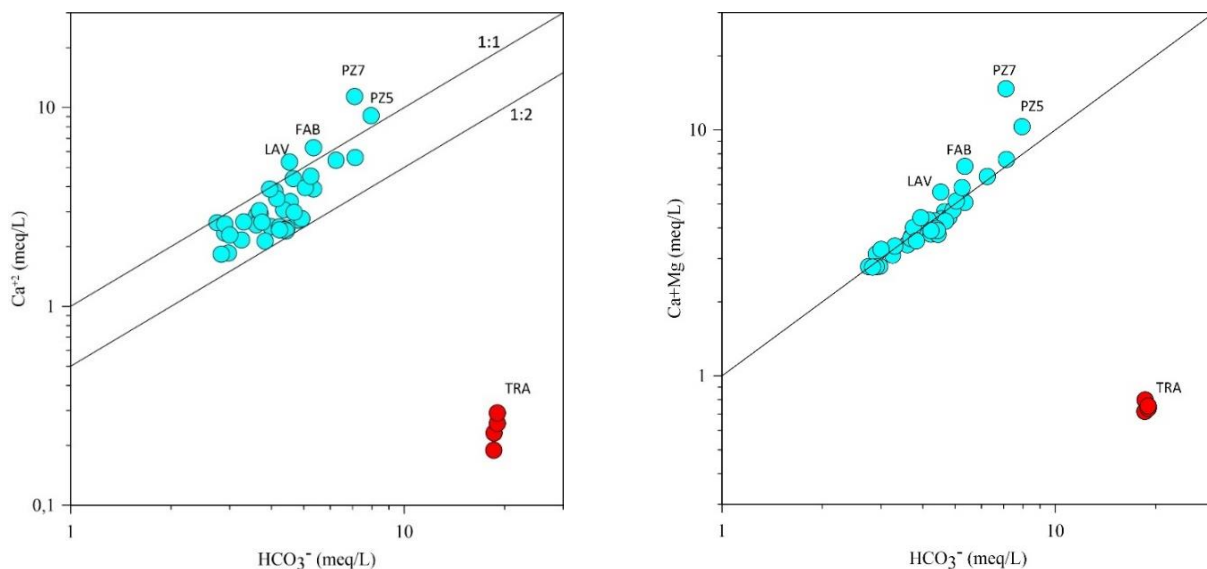
Figure 3. 33 – Binary plot of Ca^{2+} vs. HCO_3^- concentrations (on the left) and of $(Ca^{2+} + Mg^{2+})$ vs. HCO_3^- concentration (on the right).

The HAV geological substratum is principally made by carbonate rocks (limestones and dolostones). The dissolution of carbonates in karst aquifers generally provide Ca^{2+}, Mg^{2+} and HCO_3^- in groundwaters and can be expressed as follows:

$$CaCO_3 + H_2O + CO_2 \rightarrow Ca^{2+} + 2HCO_3^-$$

$$CaMg(CO_3)_2 + 2H_2O + 2CO_2 \rightarrow Ca^{2+} + Mg^{2+} + 4HCO_3^-$$
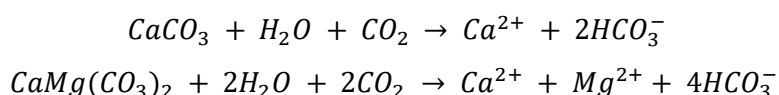
According to the previous reactions, the dissolution of calcite and dolomite would release equally amounts of (a) Ca^{2+} and HCO_3^-, and (b) $(Ca^{2+} + Mg^{2+})$ and HCO_3^-, with equivalent charge ratios equal to 1:1. If the Ca^{2+}, Mg^{2+}, and HCO_3^- contents were derived from the dissolution of dolomite, all of the groundwater samples would be observed along the 1:1 line of $(Ca^{2+} + Mg^{2+})$ and HCO_3^-. Similarly, if the dissolution of calcite was the sole source of Ca^{2+} and HCO_3^- in groundwater, the ratio of Ca^{2+} and HCO_3^- (expressed in meq/L) should be close to 1:1 (Figure 3.33). If the ratio of Ca^{2+} and HCO_3^- are in the range of 1:1 to 1:2, the dissolution of the two carbonate minerals may have occurred simultaneously. Accordingly, most the HAV groundwater samples were distributed between the 1:1 and 1:2 lines of Ca^{2+} and HCO_3^- (Figure 3.33) meaning that the dissolution of calcite and dolomite is the main process influencing the chemical composition of the investigated water. This is confirmed by calcite and dolomite saturation indexes. In fact, the most of waters is over-saturated with respect to dolomite and in equilibrium with respect to calcite (Figure 3.34). PZ7, PZ5, FAB and LAV samples were plotted above the 1:1 and 1:2 lines of Ca^{2+} and HCO_3^- suggesting another source accountable for the extra Ca^{2+}, such as the dissolution of gypsum minerals and silicate rocks or cation exchange between Ca^{2+} and Na^+. Conversely, the Tramutola samples (red dots, Figure 3.33) display a high deficit of (Ca^{2+}) with respect to the Ca^{2+}/HCO_3^- 1:1 ratio, thus invoking another kind of process.
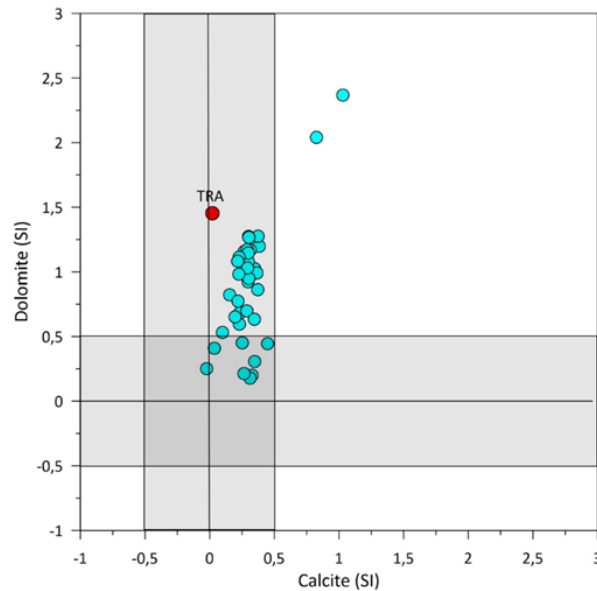
Figure 3. 34 - Diagram of the saturation indexes (SI) of dolomite and calcite for the HAV samples.
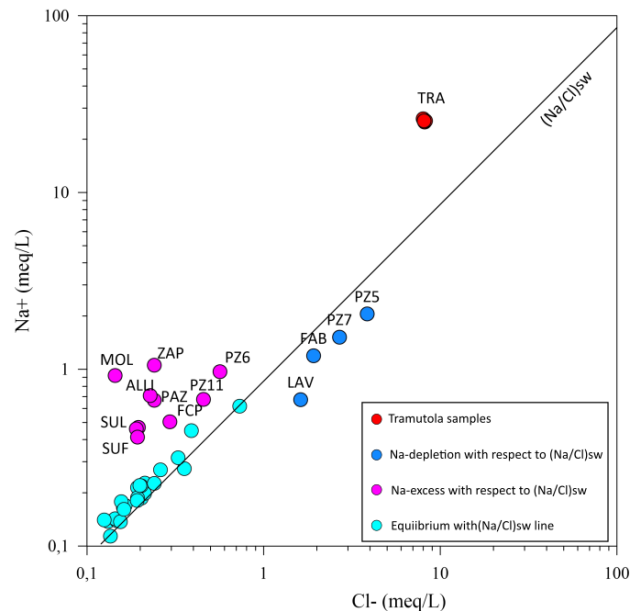
Figure 3. 35 – Na$^+$ and Cl$^-$ samples concentrations with respect to the typical Na+/Cl- ratio for sea water (sw). Figure highlights 4 groups of samples: (a) springs falling on the (Na/Cl)$_{sw}$ line (cyan dots); (b) samples with a slight Na-depletion with respect to (Na/Cl)$_{sw}$ line (blue dots); (c) samples with an enrichment in sodium with respect to the (Na/Cl)$_{sw}$ line (purple dots); (d) Tramutola samples characterized by a marked sodium enrichment (red dots).

Given the geological setting of the HAV, the presence of gypsum minerals is to be excluded. To verify the occurrence of other processes, Na and Cl concentrations of the investigated waters are plotted together with Na$^+$/Cl$^-$ ratio in sea water (Figure 3.35). The springs with the lowest concentrations, show a strong positive correlation and fall on the (Na/Cl)$_{sw}$ line, suggesting that these samples are influenced by spray marine contribution in rainwaters and therefore directly linked to the recharging process. LAV, FAB, PZ7 and PZ5 samples (blue dots in Figure 3.35) show a slight Na-depletion with respect to (Na/Cl)$_{sw}$ line, while a group of 8 samples (MOL, ZAP, PAZ, ALLI, SUL, SUF, PZ5, PZ11) (purple dots, Figure 3.35) were plotted above the (Na/Cl)$_{sw}$ line with an enrichment of sodium. Finally, Tramutola samples are characterized by a marked sodium excess. Cation exchange is an important hydrochemical process that has significant influence on the evolution of hydrochemical characteristic of groundwater (Charfi et al., 2013).

To verify the process in the HAV samples we used the relationship of $[(Ca^{2+}+Mg^{2+}) - (HCO_3^-)]$ vs. (Na^+-Cl^-). In this graph (Figure 3.36a), $(Ca^{2+}+Mg^{2+})$ concentration is corrected by HCO_3^-, to exclude the contribution of ions from carbonates. While Na^+ concentration is corrected by Cl^- to exclude Na^+ input from atmos deposition (Biswas et al., 2012; Jalali, 2009). In aquifers affected by processes of ion exchange, samples with an excess of Na^+ with respect to Cl^- - or $(Ca^{2+}+Mg^{2+})$ with respect to (HCO_3^-) – are balanced by the deficit of $(Ca^{2+}+Mg^{2+})$ or Na^+, respectively. Conversely, all data would group at the origin of the coordinate system. Analyzing Figure 3.36a, we found that most of investigated waters fall at the origin of the coordinate system thus excluding for these springs the occurrence of ion exchange processes. Conversely, Tramutola waters (red dots, Figure 3.36a) show a high enrichment in sodium that is balanced by a similar deficit of $(Ca^{2+}+Mg^{2+})$. This means that cation exchange with Na release and (Ca + Mg) uptake is one of the main hydrogeochemical processes controlling groundwater composition for the Tramutola thermal waters. The proposed process was also suggested by Italiano et al. (2001) and may be suitable for Tramutola waters since they rise across clays and carbonate rocks. On the other hand, an inverse cation exchange with a deficit of Na and excess of (Ca + Mg) can be proposed for PZ7, PZ5, FAB and LAV samples (blue dots in Figure 3.36a). These waters react with carbonate rocks dissolving Ca, Mg and HCO_3 ions, then Na enrichment occurs through an inverse cation exchange with Na uptake on clay-mineral surfaces. Again, PZ7 sample shows a marked (Ca + Mg) excess with respect to other samples, possibly related to anthropogenic activities, as highlighted by an anomalously NO_3^- value (370 mg/L). Cation exchange processes does not explain the higher $(Na/Cl)_{sw}$ ratio of the group of 8 samples (purple dots in Figure 3.36a). For it, Na-Feldspar hydrolysis processes are rather evidenced by their strongly positive correlation (R=0.82, p-values=0.0068) between sodium enrichment (Na-Cl) and the albite (Na-feldspar) saturation index (Figure 3.36b). This would agree with the siliceous substratum ("Scisti Silicei" formation, Scandone, 1967) across which these samples rise.
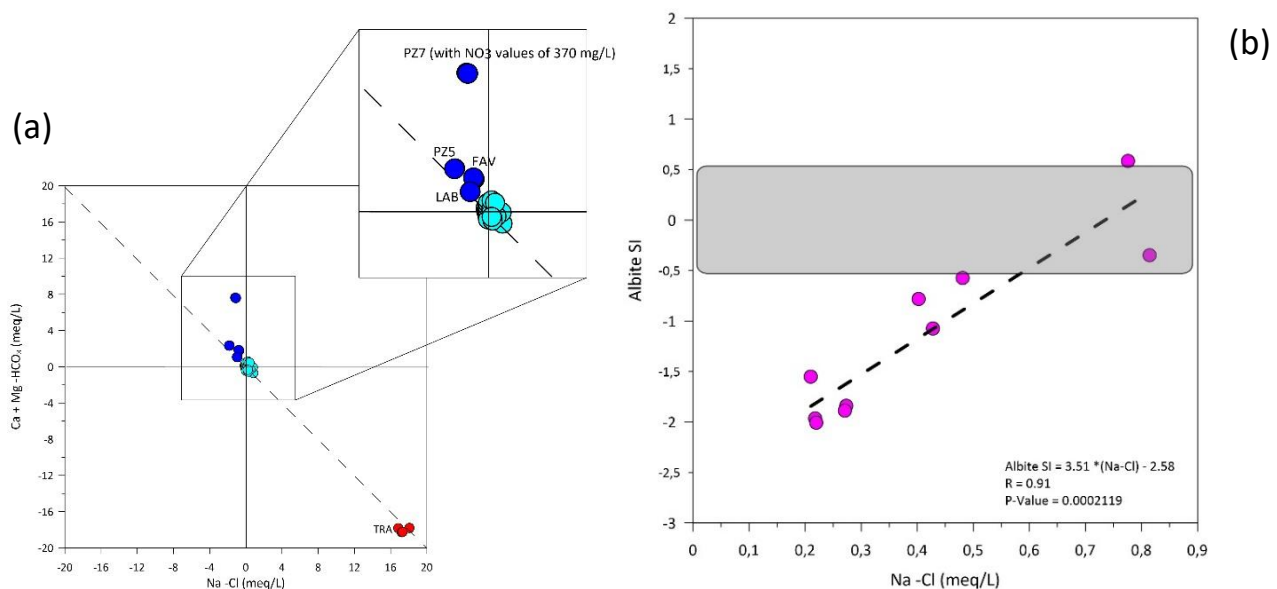
Figure 3. 36 – Comparison of the Na-enrichment with respect to seawater [Na-(Na/Cl)sw x Cl] versus (a) carbonate content (Ca+Mg-HCO3) (blues dots: samples with slight Na-depletion (FAB, LAV, PZ7,PZ5); red dots: TRA samples); (b) the albite SI (saturation index) for MOL, ZAP, PAZ, ALLI, SUL, SUF, PZ5, PZ11 samples (purple ds)

3.4.3.2 Isotopic constraints

Water stable isotopes are used to define water origin and the main processes occurring during water circulation (i.e., recharge areas, circulation paths, and mixing or exchange processes) (Clayton et al., 1966, Clarke and Fritz, 1997, Paternoster et al., 2008). The δD and $\delta^{18}O$ values of the collected samples are reported in Appendix 3.4 and displayed in Figure 3.37, together with the Global Meteoric Water Line (GMWL: $\delta D = 8.17*$ $\delta^{18}O+10.35$, Schemmel et al., 2013), the regional Eastern Mediterranean meteoric water line (EMMWL: $\delta D = 8.0* \delta^{18}O+22$, Gat and Carmi, 1971) and Southern Italy Water Line (SIMWL: $\delta D = 6.17 * \delta^{18}O+5.2$, Longinelli and Selmo, 2003). Moreover, the Local Meteoric Water Line (LMWL: $\delta D = 7.59 * \delta^{18}O+12.7$) estimated for the HAV area from Colella et al. (2003) is also reported (red dashed line, Figure 3.36). The collected δD and $\delta^{18}O$ isotopic values are similar to those measured by Paternoster et al. (2021). Considering all the analyzed samples the following linear regression equation is obtained (black solid line in Figure 3.37):

$$\delta D = 4.79* \delta^{18}O - 10.41 \qquad\qquad \text{Eq. 3. 12}$$

The best fit line (R=0.76, p-value < 0.0001) lies between the two local reference lines of SIMWL and LMWL, thus indicating that springs are fed by rainwater and have meteoric origin. An exception is represented by Tramutola well (TRA) water samples which show an enrichment in $\delta^{18}O$ with respect to the other groundwater samples of the area and fall out from the range defined by the LMWL.
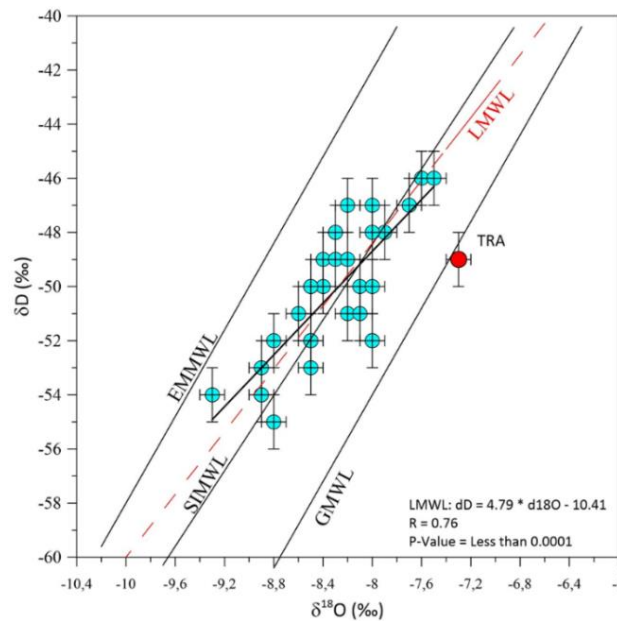
Figure 3. 37 - δ18O and δD of the groundwater samples, as compared with meteoric water lines (Northern Calabria Meteoric Water Line (NCMWL); Southern Italy Meteoric Water Line (SIMWL); Global Meteoric Water Line (GMWL) and Local Meteoric Water Line (LMWL). The uncertainties 1σ (±0.1% for δ18O and ±1% for δD) were reported.

For TRA thermal waters the water-rock interaction processes may have determined an enrichment in heavier oxygen isotopes. In order to estimate the average infiltration elevation of the meteoric recharge we use equation proposed for southern Italy by Vespasiano et al. (2015):

$$\delta\ 18O = 0.00194 \times Hi\ 5.91 \qquad \text{Eq. 3. 13}$$

Where Hi is the expected infiltration elevation. We found that water samples located on the hydrological left of the HAV basin (AMO, PEN, SGC, SAL, PMR, GIO, OSC, AGG, SAN, MIE, SOG, CAC, CAR, CAP, FCP, PAZ, ZAP, FAB) have recharge (infiltration) elevations between 923 m and 1387 m (median 1195 m) a.s.l. with an estimated mean difference between the recharge altitude and the respective spring of about 550 m. Samples located on the right of the hydrological basin (AGR, POC, OCC, CAQ, VAN, PSC, TAS, SVG, CPN, SAM, MOL, ALLI, PZ11, PZ6, SUL , SUL ,SUF, PZ5, PZ7, LAV) showed recharge altitudes in the range 820 - 1540 m (1250 m) a.s.l., with a minor difference between the recharge altitudes and the springs (about 450 m). The infiltration elevation is compatible with the hydrogeological basin of the Agri River for both sides. The estimated mean recharge altitude-spring is 100 meters higher for springs located on the hydrological left of the valley, suggesting their longer migration paths with respect to spring waters located on the hydrological right side of the HAV.
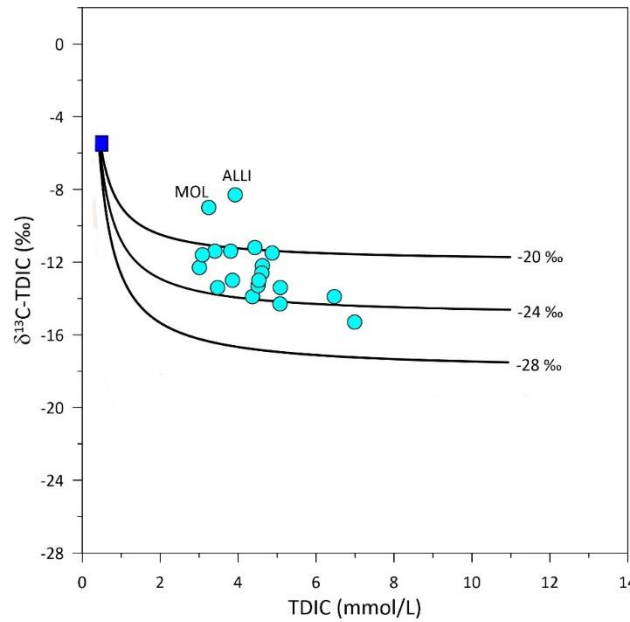
Figure 3. 38 - TDIC versus $\delta^{13}C$-TDIC diagram. The compositions of HAV springs (blue circles) are compared with the theoretical curves representing the evolution of groundwater infiltrating through dissolution of soil CO2 deriving from oxidation of organic matter and root respiration (red curves).

The $\delta^{13}C$ -TDIC values range from -15.3‰ and -8.3‰ vs-PBD (Appendix 3.4). In Figure 3.38, the TDIC and $\delta^{13}C$-TDIC values of sampled waters are compared to the theoretical curves representing dissolution of soil CO_2 and equilibrium with calcite (black theoretical curves in Figure3.38). No deeply derived CO_2 input was assumed for the calculation of the model theoretical evolutions. Each model was computed by means of the PHREEQC code (Parkhurst and Appelo, 1999) starting from TDIC and $\delta^{13}C$-TDIC values of 2.5 x 10^{-5} mol/kg and -5.5‰, respectively. These values are representative of the average composition of rainwater for the latitudes of the study area, in the assumption of isotopic equilibrium between water and atmospheric CO_2 (Chiodini et al., 2000). The theoretical model was computed through the addition in 200 steps of 0.01 mol of

soil CO_2 to 1 kg of the infiltrating rainwater, considering three possible values for $\delta^{13}C_{soil}$ (20‰, 24‰ and 28‰), intended as the biogenically derived carbon compound active in the soil of the recharge areas during the infiltration of meteoric waters (Chiodini et al., 2000). At each step, calcite (Ccarb) was dissolved at equilibrium condition. The isotopic composition of calcite ($\delta^{13}C_{carb}$) was set at 2.21 ± 0.66‰ (mean and standard deviation of 567 samples of Apennine carbonate rocks, Chiodini et al., 2004).

In TDIC vs. $\delta^{13}C$ diagram the most of the investigated HAV samples are characterised by TDIC < 0.008 mol kg^{-1} and are generally positioned along the theoretical curves compatible with a closed-system equilibration between carbonate minerals and dissolution of soil CO_2. Exceptions are MOL and ALLI samples which show a $\delta^{13}C$-TDIC > -10‰ (-9.0‰ and -8.3‰, respectively), thus suggesting a possible slight contribution for deep carbon source in TDIC composition. In order to eliminate the effect of carbonate dissolution the variables TDIC and $\delta^{13}C$-TDIC were calculated as Cext and $\delta^{13}C$-ext.

It is known that the total dissolved inorganic carbon (TDIC) transported by groundwaters circulating in carbonate aquifers can be considered as the sum of carbon deriving from carbonate dissolution within the aquifer and the contribution from sources external to it, according to the following simple equation:

Eq.3. 14

$$TDIC = C_{carb} + C_{ext}$$

where C_{ext} represents the amount of carbon deriving from external sources and C_{carb} states for the concentration of carbon deriving from carbonate mineral dissolution. Assuming that in the investigated area (HAV), Ca and Mg amount in water samples entirely derive from carbonate mineral dissolution, we estimated the carbon share from carbonate mineral dissolution as given by (Chiodini et al., 2004):

$$C_{carb} = Ca + Mg \qquad \text{Eq.3. 15}$$

Where concentrations are expressed in mol/l. From Eq. (1) and Eq. (2), first the C_{ext} for each sample was calculated, then corresponding $\delta^{13}C_{ext}$ isotopic composition was estimated as (Chiodini et al., 2004):

$$\delta^{13}C_{ext} = (\delta^{13}C\text{-TDIC} \times \text{TDIC} - \delta^{13}C_{carb} \times C_{carb}) / C_{ext} \qquad \text{Eq.3. 16}$$

The results are shown in Figure 3.39. Most of the investigated waters have $\delta^{13}C_{ext}$ values ranging between -31.2‰ and -20.8‰, compatible with carbon deriving from a biological source active in the soil during the infiltration of the rainwater. These water samples overlap on the theoretical curves representing the dissolution of soil CO_2 deriving from oxidation of organic matter and root respiration. Only two water samples (ALLI and MOL) are compatible with slightly input of deeply derived CO_2. Alternatively, these values could be related to a different organic carbon source with a more positive isotopic signature. Results at ALLI and MOL springs, let us to identify it as possibly suitable locations for the future installation of multi-parametric geochemical stations for the continuous and long-term geochemical monitoring of the area.
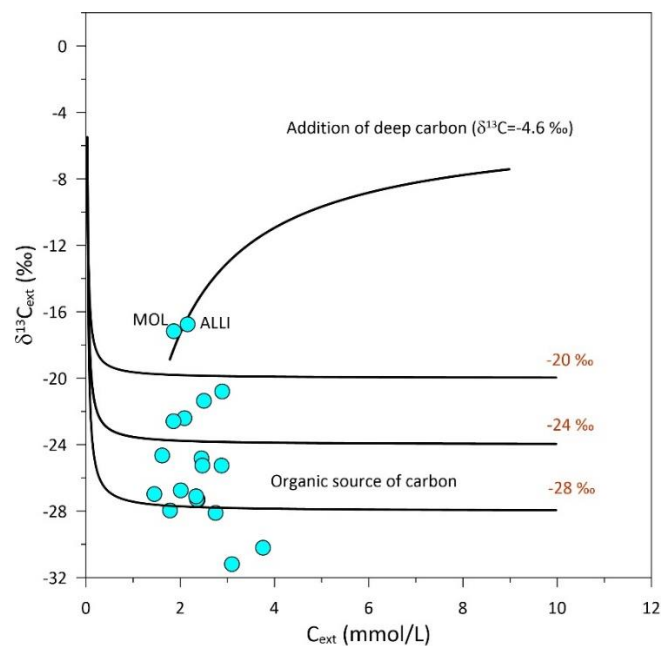
Figure 3. 39 - Cext versus 13δCext diagrams. The compositions of HAV springs (blue circles) are compared with the theoretical curves representing different amount (mmol/L) of Cinf.

3.4.4 Groundwater quality

The groundwater quality is a fundamental issue for both environmental and human health purposes (Mongelli et al., 2019). The quality of water supplies for drinking use is regulated in Italy by the D. Lgs 31/2001, acceptance of the 98/83/EU directive (R. Italiana, "Decreto Legislativo 2 febbraio 2001, n. 31 Attuazione della direttiva 98/83/CE relativa alla qualità delle acque destinate al 92onsume umano"), which provides the reference guidelines for the measurement of water indicator parameters (e.g., odor, color, taste, pH, and hardness) and chemical and microbiological compositional limits for water quality assessment. In this study, the assessment of groundwater quality was performed considering the maximum admissible concentrations (hereafter MAC) of both trace and major elements imposed as chemical and indicator by the Italian legislation for drinking waters. Both the regulated trace (Figure 3.42, 3.43) and major elements (Figure 3.41) amounts in HAV samples were compared with MAC and are summarized in Table 3.6 and Table 3.7, respectively. Between the dosed minor trace elements (Appendix 3.4), the environmental indicators of the Italian legislation (As, B, Cu, Cd, Fe, Hg, Mn, Ni, Pb, Se, Sb, V) were analyzed. Their distributions and variability ranges are represented in Figure 3.40. Here, outlier data points (observations numerically distant from the rest of the data) were calculated as the one which value is outside 1.5 times the interquartile range above the upper quartile (Q3=75th percentile of the distribution) and below the lower quartile (Q1=25th percentile of the distribution). B and Fe recorded the highest concentrations in the range between 10^1 and 10^2 µg/l. Outlier values of boron were recorded at PAZ and ZAP samples (90 and 157 µg/l, respectively), while SUF showed particularly high Fe content (60 µg/l) with respect to other samples. As, Cu, Mn, Ni, Pb, Se and V showed minor amounts between 10^{-1} and 10 µg/l, followed by Sb ($10^{-2} – 10^{-1}$ µg/l) (Figure 3.40). A particularly high value of As was estimated at VAN sample (3.47 µg/l); outlier values of Mn were recorded at SUL, SUF, AGG, LAV samples (37.3 µg/l, 29.2 µg/l, 17.0 µg/l, 12.0 µg/l, respectively), while an anomalous Cu content (> 1.0 µg/l) was

detected at PZ6, CAR, AMO, PZ7, POC, PEN, and SUF (2.9 μg/l, 2.7 μg/l, 2.5 μg/l, 2.2 μg/l, 1.8 μg/l, 1.7 μg/l, 1.6 μg/l, respectively). Higher values of Pb with respect to other samples were observed at PEN (1.35 μg/l), AMO (0.84 μg/l) and MOL (0.57 μg/l). Particularly high content of Ni was recognized at PZ7 (3.4 μg/l), AMO (1.8 μg/l) and PZ6 (0.9 μg/l), while PZ5 and PZ11 showed outlier values of Sb (0.14 μg/l and 0.17 μg/l, respectively). Cd and Hg were characterized by very low content, below the detection limit (b.d.l.: 0.01 and 0.2 μg/l, respectively).
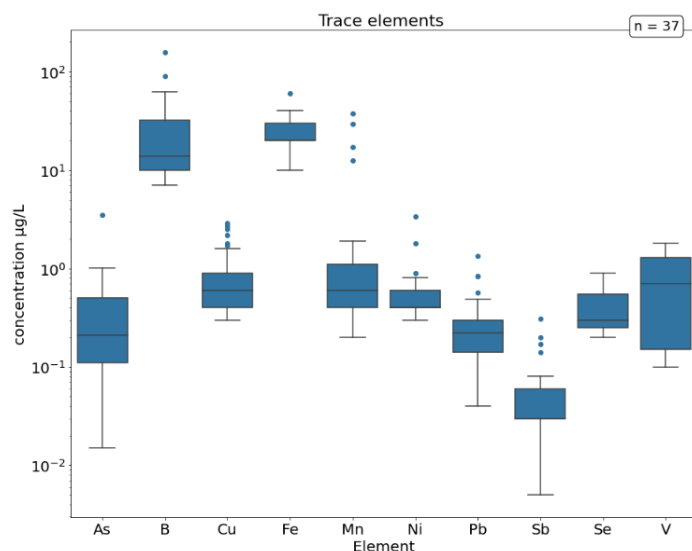
Figure 3. 40 – Distributions of traces elements used as chemical indicators for groundwater quality analysis. For each distribution, the 25th percentile is assumed as the first quartile (Q1) and the 75th percentile as the third quartile (Q3) and defined the interquartile range (IQR=Q3-Q1); while the minimum and maximum values are found at the end of each box 'whisker'. Outlier data points are located outside the whiskers of the box plot.

Among the environmental indicators B, Cu, As, Mn, Pb and Sb were detected in all the samples and analyzed in this work (Figure 3.43, Figure 3.44). Regarding the major elements regulated from the Italian legislation for drinking waters NH$_4$, Cl, Fe, Mn, SO$_4$, Na, F, NO$_2$, and NO$_3$ amounts were analyzed and compared with their MAC. Between the analyzed samples NO$_2$ was characterized by concentrations below the detection limit (0.2 mg/l), while most of the major elements were above the prescribed limits for drinking use. The analysis of both trace and major elements, highlighted the high quality of the HAV groundwater for drinking use after the Italian and European legislation, thus confirming previous groundwater quality assessment performed by Paternoster et al. (2021). The only exception was represented by NO$_3$ at PZ7 which anomalous higher amount (370 mg/l) is above the prescribed limit (MAC for NO3: 50 mg/l) and linked to presence of manure. Indeed, the piezometer is in a scarcely inhabited area where animal grazing represents the main anthropic activity.

Element	Maximum amount (μg/l)	Minimum amount (μg/l)	Mean (μg/l)	Standard deviation (μg/l)	D. Lgs. 31/2001 Maximum admissible concentrations (MAC) (μg/l)
NO_2	b.d.l	b.d.l	b.d.l	b.d.l	50
B	157.00	7.00	26.22	29.21	1000
V	1.80	0.10	0.56	0.62	50
Ni	3.40	0.30	0.37	0.63	20
Cu	2.90	0.30	0.90	0.72	1000
As	3.47	0.02	0.38	0.58	10
Se	0.90	0.20	0.21	0.26	0.01
Cd	b.d.l	b.d.l	b.d.l	b.d.l	5
Sb	0.31	0.01	0.06	0.06	5
Hg	b.d.l	b.d.l	b.d.l	b.d.l	1
Pb	1.35	0.04	0.29	0.26	10
Al	25.00	3.00	5.49	5.16	200

Table 3. 6 – HAV minor and traces amounts samples were compared with their D. Lgs 31/2001 maximum admissible concentrations (MAC).

Element	Maximum amount (mg/l)	Minimum amount (mg/l)	Mean (mg/l)	Standard devation (mg/l)	D. Lgs. 31/2001 Maximum admissible concentrations (MAC) (mg/l)
NH_4	0.39	0.05	0.17	0.13	0.5
Cl	136.80	4.44	17.09	27.45	250
Fe	60.00	0.01	8.65	14.37	0.2
Mn	37.30	0.20	3.20	8.05	0.05
SO4	68.82	8.45	17.22	14.03	250
Na	47.22	2.62	10.25	10.14	200
F	1.14	0.04	0.25	0.27	1.5
NO_3	373.89	0.12	13.21	61.25	200

Table 3. 7 – HAV major elements amounts samples were compared with their D. Lgs 31/2001 maximum admissible concentrations (MAC).
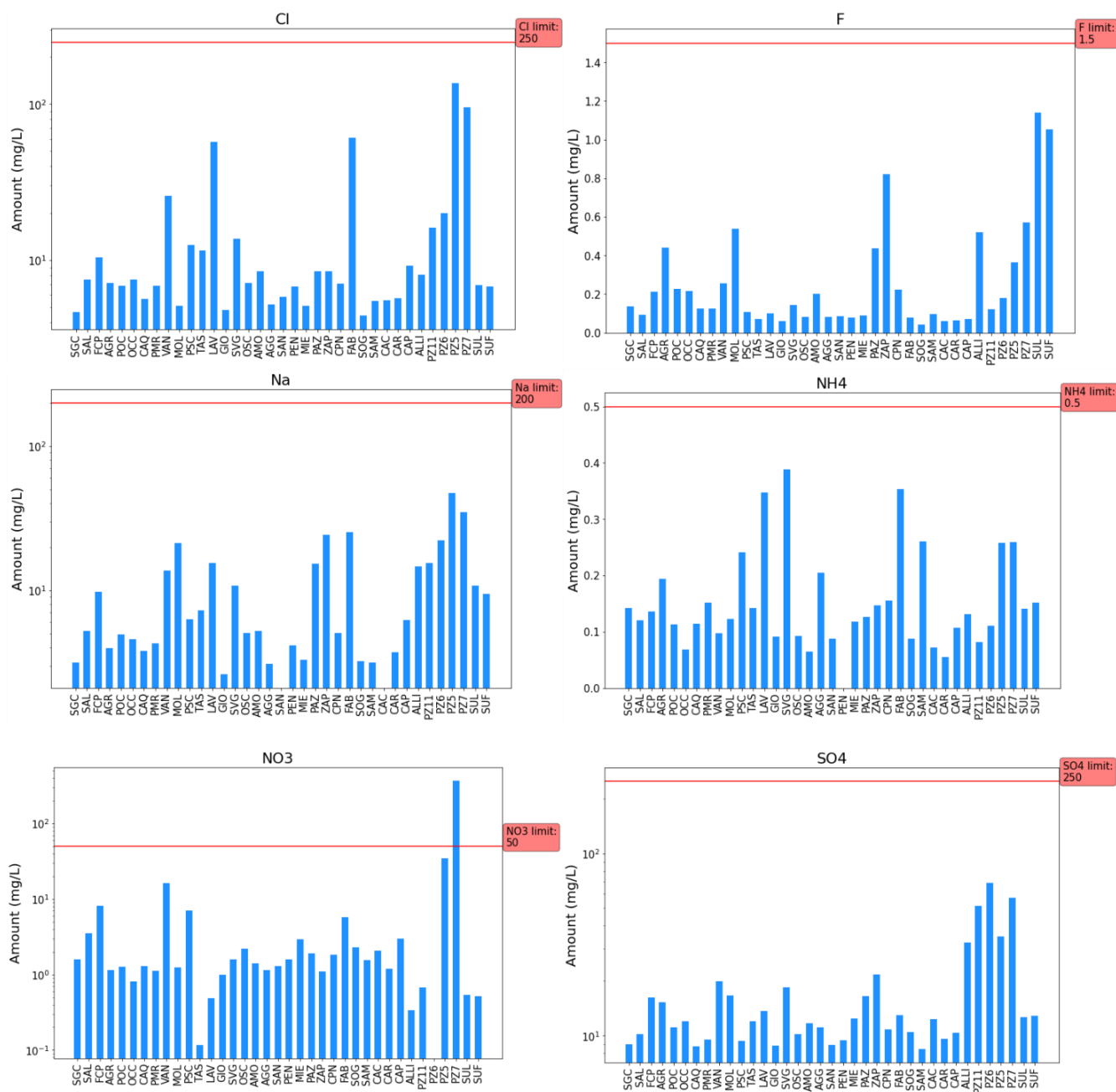
Figure 3. 41 – HAV major elements amount for each sample compared with their D. Lgs 31/2001 maximum admissible concentration (MAC) limit (red line).
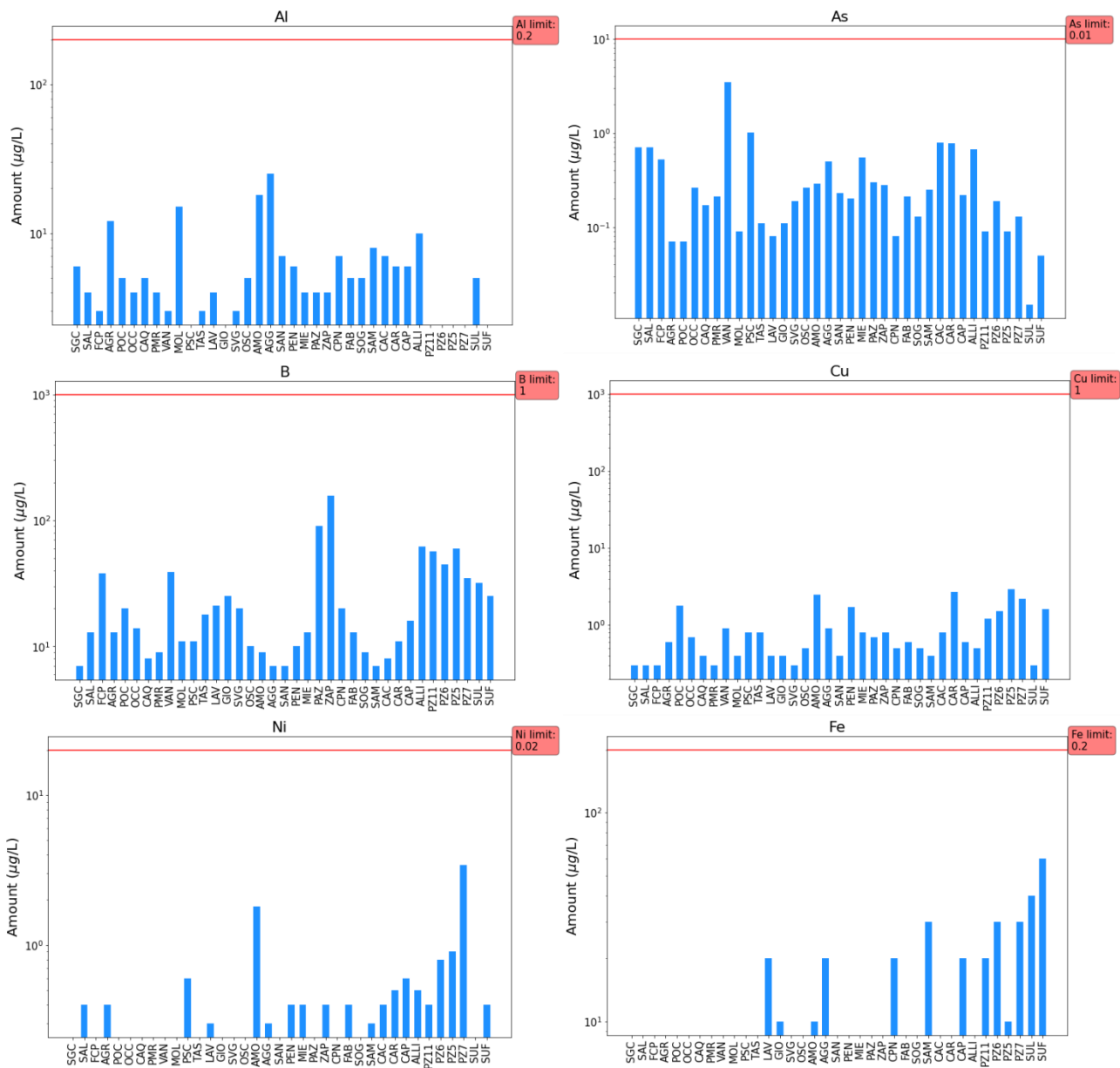
Figure 3. 42 – HAV traces elements amount for each sample compared with their D. Lgs 31/2001 maximum admissible concentration (MAC) limit (red line).
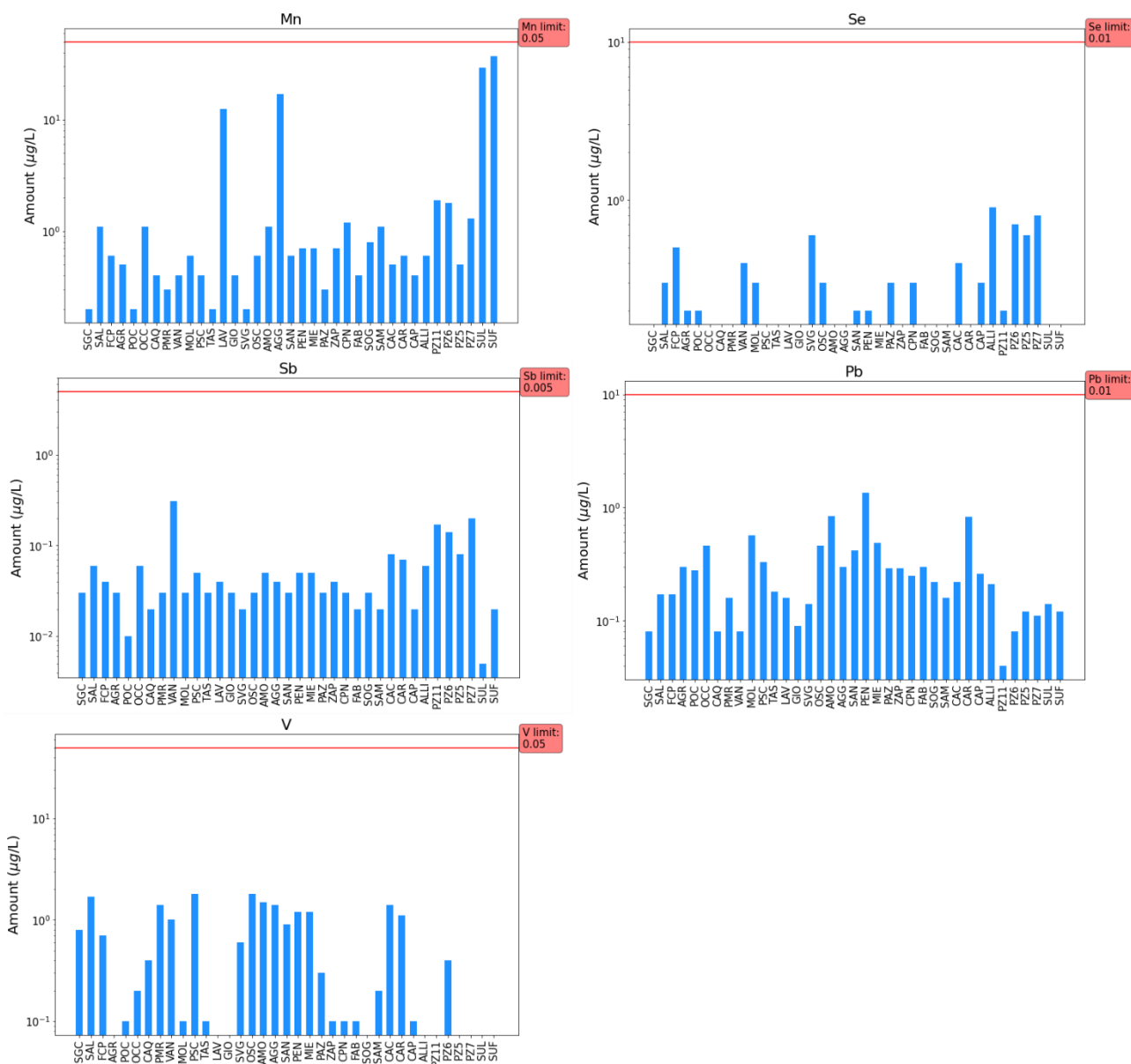
Figure 3. 43 – HAV traces elements amount for each sample compared with their D. Lgs 31/2001 maximum admissible concentration (MAC) limit (red line).

3.4.5 Gas geochemistry of the Tramutola borehole for hydrocarbon exploitation

The HAV is one of the areas of Italy with the highest seismogenic potential as demonstrated by the historical Mw 7.0 strong earthquake occurred in 1857 (Mallet, 1862). The HAV is characterized by seismogenic fault systems capable of producing large events, which are alternatively associated to the Eastern Agri Fault System (EAFS), a NW-trending, SW-dipping normal-fault system bounding the eastern side of the basin (Benedetti et al., 1998; Barchi et al., 2007), or to the Monti della Maddalena Fault System (MMFS), a NW-trending, NE-dipping normal-fault system cross-cutting the mountain range to the west (Improta et al., 2010). The HAV hosts also the largest onshore Western European oil field. The Tramutola borehole (N2 Agip) is located on the western side of the basin and it was drilled from AGIP S.p.a. for oil & gas exploitation in the period between 1936-1939. Right after its drilling, the well enabled methane average production at a rate of 1300 m^3/day. The

well is currently out of production. Differently from HAV groundwaters, the Tramutola gas samples (TRA 27-06-2020, TRA 05-04-2021) are CH_4-dominated (mean 82.63%), with minor amounts of N_2 (mean 12.98%), CO_2 (mean 1.72%), O_2 (mean 0.06%) and C_2H_6 (mean 0.32%). These data are similar to those measured from other authors (Etiope et al., 2007, Italiano et al., 2000,2001, Minissale et al., 2019) for Tramutola gas composition. Methane was absent in most of the analyzed HAV gas samples, except for SUF which showed a particularly marked oxygen depletion and a significative amount of CH_4 (10.4%) (CH_4–N_2–CO_2 ternary diagram, Figure 3.30). SUF gas composition suggested the possible deeper and longer hydrogeological circuit for this groundwater. For this reason, SUF gas composition and isotopic signatures were investigated in this section and compared with the one of the Tramutola thermal waters.

3.4.5.1 Helium and carbon isotopes

An important geochemical tool to detect deep fluids and their possible origin is represented by noble gases. Given their chemical inertness, the distinct isotopic signature, and the negligible isotopic fractionation in gas-water-rock interactions, these are considered excellent natural tracers to discriminate their contribution from the mantle and crustal sources with respect to the atmosphere (e.g., Ballentine and Burnard, 2002; Caracausi and Paternoster, 2015; Caracausi and Sulli, 2019). Between noble gases, Helium is widely used for this purpose (Ozima & Podosek, 2002). Indeed, mantle, crustal, and atmospheric helium sources are characterized by well-defined He isotopic compositions: ^3He has a primordial origin and is usually degassed from the mantle; ^4He is produced by U and Th decay, therefore the typical He isotopic ratio in crustal fluids is about 0.01–0.03Ra (Ozima & Podosek, 2002).^3He and ^4He concentrations range from 4.81E-04 ppm (TRA) to 4.07E-05 ppm (SUF), and from 285.16 ppm (TRA) to 21.04 ppm (SUF), respectively. In the TRA sample the $^4He/^{20}Ne$ ratio is 366.89, three orders of magnitude higher the ratio in air (0.318), therefore atmospheric contribution is negligible. Conversely, SUF sample is affected by air contamination (about 85%) as the value of the $^4He/^{20}Ne$ ratio (1.9) is slightly higher than recorded in air-saturated water (ASW, $^4He/^{20}Ne$ = 0.2672; Holocher et al., 2002). The measured $^3He/^4He$ ratios of samples range from 1.38 Ra (SUF) to 1.20 Ra (TRA), where Ra denotes the atmospheric $^3He/4He = 1.38 \times 10^{-6}$, Ozima & Podosek, 2002). The latter are in good agreement with previously reported helium isotope data (Italiano et. al., 2001, Caracausi and Paternoster, 2015b). The carbon isotopic composition of CO_2 ($\delta^{13}C_{CO2}$) for TRA sample is of -3.5‰ VPDB, and this value is in the range of the mantle values (−4 ± 1‰, Deines, 2002; Sano and Marty, 1995). This measured value results to be similar to that measured from Italiano et al. (2001), while is enriched in heavier carbon isotopes respect with data ($\delta^{13}C_{CO2}$= -16.1‰) from Tassi et al. (2012). We report $^3He/^4He$, $CO_2/^3He$, C-isotopes in CO_2 and CH_4, and He abundance results (water phase only) in Table 3.9 with the major gas chemistry in Table 3.8

Sample ID	data	He (%)	O_2 (%)	N_2 (%)	CO_2 (%)	CH_4 (%)	C_2H_6 (%)	C_3H_8 (%)
TRA	27/06/2020	0.026	0.0584	12.78	1.92	81.94	0.31	b.d.l.
TRA	05/03/2021	0.0276	0.0748	13.17	3.3	83.31	0.32	b.d.l.
TRA	11/02/2022	0.0283	0.0507	12.97	1.47	85.96	0.33	b.d.l.

Table 3. 8 – Tramutola free gases samples: major gas chemistry.

Sample ID	Sampling date	Type	^{3}He ppm	^{4}He ppm	^{40}Ar ppm	^{38}Ar ppm	^{36}Ar ppm	^{3}He/^{4}He	R/Ra	R/Rac	^{4}He/^{20}Ne	CO$_2$/^{3}He
SUF	05/03/2021	dissolved gas	4.07E-05	21.04	8129.8	5.18	27.35	1.93E-06	1.38	1.45	1.87	
TRA	27/06/2020	free gas	5.20E-04	294.74	909.4	1.118E-02	0.06	1.76E-06	1.26	1.26	508.10	3.69E+07
TRA	28/02/2021	free gas	4.42E-04	275.57	1102.3	1.381E-02	0.07	1.60E-06	1.15	1.15	225.68	
TRA	11/02/2022	free gas	4.34E-04	267.21	1054.3	5.874E-01	3.183	1.62E-06	1.16	1.26	372.16	3.39E+07

Table 3. 9 – Isotopic compositions of noble gases (He, Ar), He isotopic ratios and CO$_2$/^{3}He in SUF (dissolved gas) and TRA (free gas) samples.

^{3}He/^{4}He ratios can be explained in terms of He of different origin (atmosphere, mantle, and crust) that mixes during the migration through the crust. With the aim to estimate the He percentage of each endmember in the HAV fluids, we applied the approach proposed by Sano et al. (1997) that is based on mixing between fluids sources from three endmembers (air, crust, and mantle), which He isotopes and ^{4}He/^{20}Ne ratios are strongly different each other. The following equations (Sano and Wakita, 1985) have been used:

$$\left(\frac{^3He}{^4He}\right) = \left(\frac{^3He}{^4He}\right)_M \times M + \left(\frac{^3He}{^4He}\right)_C \times C + \left(\frac{^3He}{^4He}\right)_A \times A \qquad \text{Eq.3. 17}$$

$$1/\left(\frac{^3He}{^{20}Ne}\right) = M/\left(\frac{^3He}{^{20}Ne}\right)_M + C/\left(\frac{^3He}{^{20}Ne}\right)_C + A/\left(\frac{^3He}{^{20}Ne}\right)_A \qquad \text{Eq.3. 18}$$

$$M + C + A = 1 \qquad \text{Eq.3. 19}$$

where M, C and A represent the fractions of mantle, continental crust, and atmospheric helium, respectively. A SCLM (Subcontinental Lithospheric Mantle) mantle with a He-isotopic ratio of 6.1 ± 0.9 Ra was chosen as mantle-endmember. Indeed, is widely recognized that such value is highly representative of mantle He composition in the European and Mediterranean area (Dunai and Baur, 1995; Gautheron et al., 2005; Torfstein et al., 2013) with estimated values ranging from 6.2 to 6.5 Ra. TRA samples (TRA 27/06/2021, TRA 28/02/2021, TRA 11/02/2022) fall along a mixing line between air and a prevailing crustal origin endmember, with a significant contribution of mantle-derived helium (about 20%) (red dots in Figure 3.45). He isotopic data are in accordance with those collected from previous papers (Italiano et. al., 2001, Caracausi and Paternoster., 2015b), which range between 1.11-1.17 (empty dots in Figure 3.45). Conversely, SUF sample show a large atmospheric input (about 85 %) (blu dot in Figure 3.45), this is probably due to the huge meteoric recharge and relatively rapid circulation in this area, probably masking the deep contribution, if any.
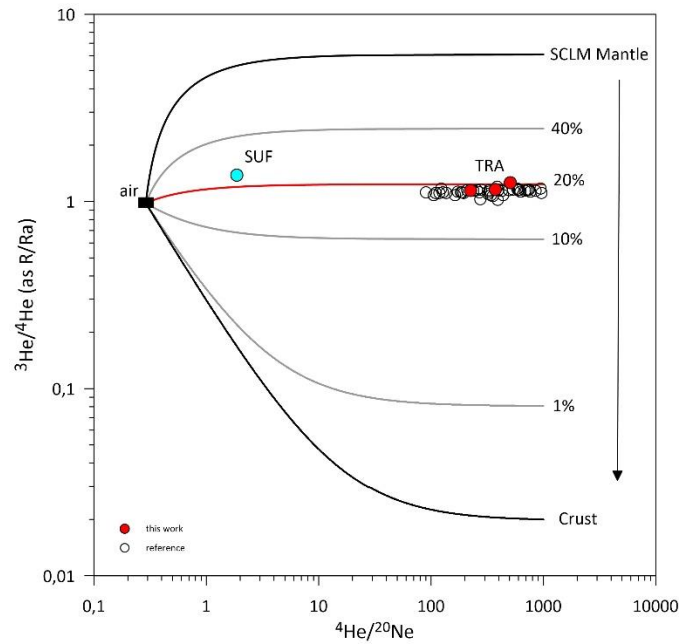
Figure 3. 44 – Helium isotopic ratios (as R/Ra values) and He/Ne ratio values for the TRA and SUF and comparison with literature reference data (Italiano et. al., 2001, Caracausi and Paternoster., 2015b). The theoretical lines (black lines) represent binary mixing trends for mantle-originated, crustal and atmospheric helium, were calculated using the following endmembers: air (^{3}He/^{4}He = 1.39 × 10^{-6}, ^{4}He/^{20}Ne = 0.318; Ozima et al.2002); SCLM Mantle (^{3}He/^{4}He = 6.1 ± 0.9 Ra, ^{4}He/^{20}Ne = 1000; Gauteron et al.2002.; Buttitta et al. 2020); crust (^{3}He/^{4}He = 0.02 Ra, ^{4}He/^{20}Ne = 1000; Ballentine et al., 2002). Further mixing curves (orange lines) between with 1%, 10%, 20% and 40% mantle contribution have also been added.

3.4.5.2 Migration of the Mantle-Derived Fluids

It is known that the presence of a significant mantle He signature can be linked to: i) mantle-derived melt intrusions into the crust (e.g., Ballentine and Burnard, 2002; Sano and Wakita, 1985; Allard et al.,1997; Italiano et al., 2000), which in turn are generally associated to high heat flux; ii) mantle-derived fluids uprising through lithospheric faults (e.g., Burnard et al., 2012; Caracausi and Sulli, 2019). The HAV area is not characterized by high heat flow (average value of about 40mWm^{-2}, Megna et al., 2014). making less probable the occurrence of magmatism at depth. Conversely, it known that brittle tectonics have strongly controlled the formation and evolution of the HAV basin. The studied area is characterized by active extensional tectonics and is bordered by NW striking high-angle normal and oblique fault systems (i.e., the MMMF fault system to the southeast and the EASF fault system to the northwest) which represent the current main seismogenic structures in the area. Earthquake locations data, along with other geophysical investigations (e.g., magnetotelluric model of the area, Balasco et al., 2015, 2021; Vp/Vs velocity models, Valoroso et al., 2011, Improta et al., 2017) showed that these faults systems involve either the quaternary and the Apennines allochthonous units with surficial faults, nor the Apulian Platform tectonic basement with deeper (12-15 km) regional faults (Maschio et al., 2005, Menardi Noguera and Rea, 2000). Thus, the hypothesis of the deep-rooted regional faults through which mantle fluids migrated to the shallower crust is more suitable for the uprising of mantle-derived gas at the Tramutola well). A Magnetotelluric (MT) model was developed from Balasco et al. (2015) in the Tramuola area (Figure3.46). It consisted of a ~ 15 km long 2-D resistivity model with an investigation depth of ~ 10 km. Electrical resistivity is dominated by geological units' porosity and the contained fluids. The model interpretation obtained on the basis of the surface geology and subsurface data across the HAV (e.g., Mazzoli

100

et al., 2001; Borraccini et al., 2002; Shiner et al., 2004; Valoroso et al., 2011) highlighted the superposition of three stacked, lateral varying layers with different thickness and resistivity values: a surficial low–medium resistivity layer (10–250 Ωm) associated with the Quaternary deposits (QD) and to the allochthonous units (LU, CLP); and a deeper high resistivity layer (> 90 Ωm) related to the Apulia Platform (AP), separated by a thin intermediate domain of medium resistivity values (40– 110 Ωm), associated prevalently with the mélange zone and to the Pliocene terrigenous marine deposits (M-TM) (Balasco et al., 2015). In it, sharp lateral resistivity variations due to circulating fluids are interpreted as faults that, based on accurate focal mechanism computations, display normal-faulting kinematics (Figure 3.46). At depth, the major feature identified from the resistivity model (F1) is interpreted as a SW-dipping fault that cuts the top of AP just below the Tramutola well in southwest part of the valley. The lateral resistivity variations of this MT model are spatially coincident and compatible with high Vp/Vs values highlighted from Valoroso et al. (2011) in the same area (Figure 3.47), thus confirming their relationship with fluid circulation along fault systems. The deep source of Tramutola thermal fluids, and their rising at surface, would testify the connection of the observed deep-rooted faults (e.g., F1, Balasco et al., 2015) with surficial crustal layers, and their role as preferential conduits for fluid circulation in the crust below the well. The considered geophysical surveys show high-resolution information on the structural setting up to about 15 km below the Tramutola well. However, crustal cross-section of the Basilicata oil field of the HAV - constructed by the integration of outcrop geology and available subsurface datasets and reported in literature (Figure 3.48) - show the existence of steep rooting, crustal thrust ramps offsetting the Moho discontinuity and linking to the surficial extensional structures, which would represent suitable conduits for the mantle-source gas uprising to the shallower crust above the Tramutola well. Alternatively, the hypothesis the existence of crustal traps storing volatiles from old magmatic activities at depth (e.g., fossil magmatic gases; Ballentine and Burnard 2002) should be taken into account to explain the presence of a significant mantle He signature in the Tramutola thermal gases.
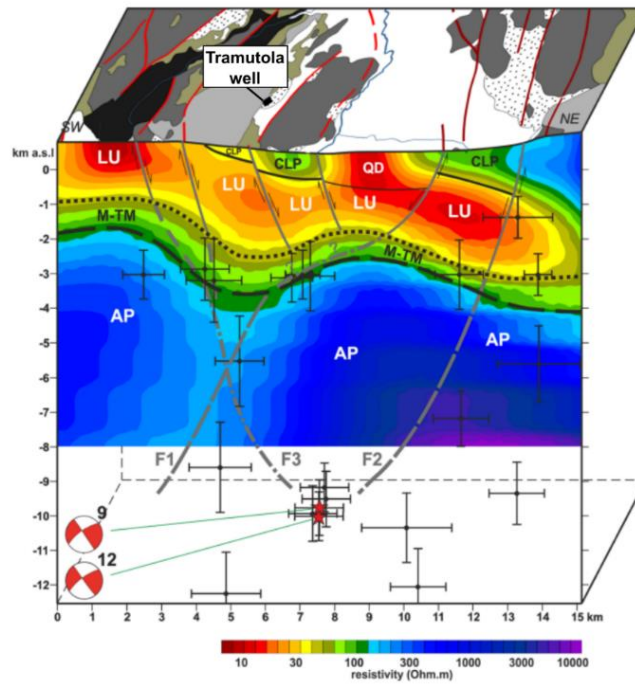
Figure 3. 45 – MT model compared with morphostructural map of the Agri Valley and projection along the MT model of seismicity. The focal mechanisms of the two events represented with red stars display normal-faulting kinematics on two possible NW–SE trending fault plane solutions (modified from Balasco et al., 2015).
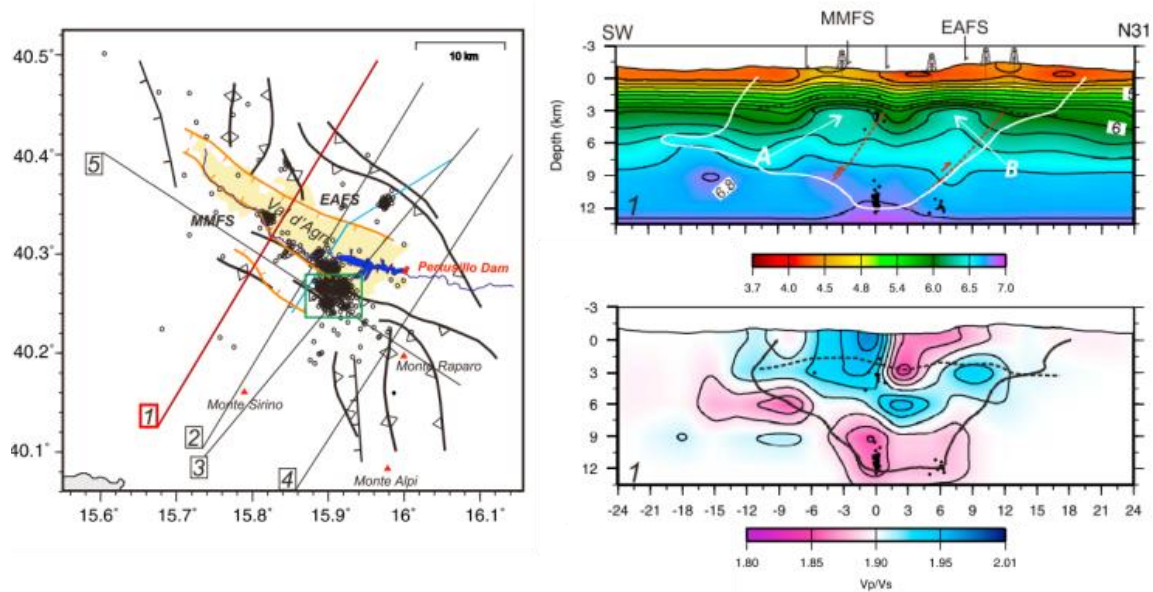
Figure 3. 46 – Map location (on the left) of section 1 (red line) of the Vp and Vp/Vs vertical sections (on the right) obtained in the Tramutola area from Valoroso et al., 2011. Relocated microearthquakes within ±1.5 km from each section are plotted. Map shows the Quaternary basin (yellow area), the Pertusillo dam, the schematic surface traces of the EAFS and MMFS (orange lines) together with main thrusts and anticlines (black lines) (modified from Valoroso et al., 2011).
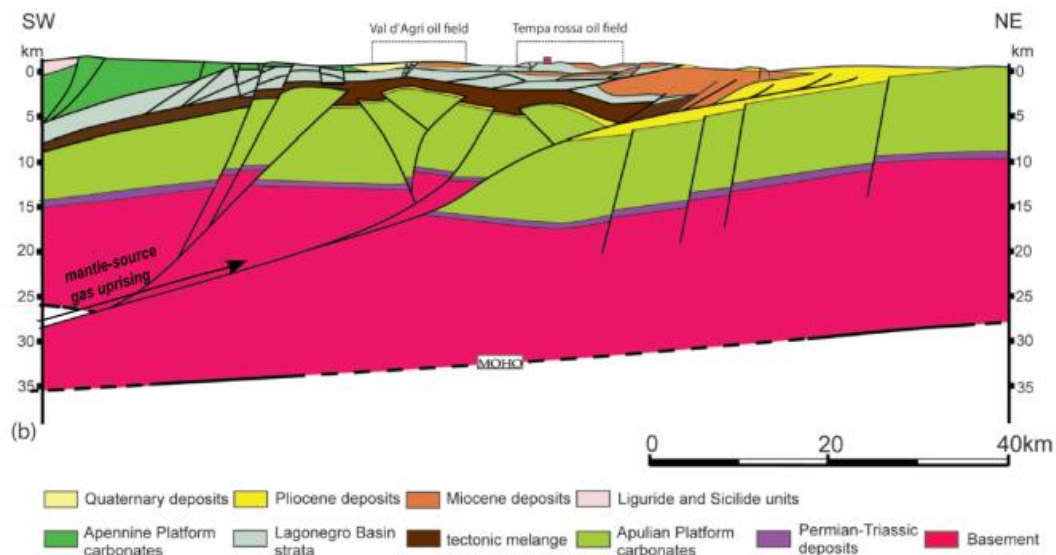
Figure 3. 47 - Geological cross-section the southern Apennines (MAZZOLI et alii, 2013, modified), showing location of site of geotherm calculation.

3.4.5.3 Carbon and hydrocarbon origin

In sedimentary environments the CH_4-dominated gases are principally produced by metabolic and biosynthetic activity (biogenesis) or decomposition of organic matter buried in sediments at T>150 °C (thermogenesis) (Schoell, 1980, 1988; Whiticar, 1999 and references therein). Despite its origin, the CH_4 isotopic composition could be modified by different post-generation processes (i.e., mixing, migration, biodegradation, thermochemical sulfate reduction, and oxidation, Milkov & Etiope, 2018). To assess gas origin differences among seep typologies and from reservoir gas the $\delta^{13}C_{CH4}$ vs C1/ (C2 + C3), and $\delta^{13}C_{CH4}$ versus $\delta^{13}C_{CO2}$ diagrams are used (Milkov & Etiope, 2018). Usually, thermogenic CH_4 has $\delta^{13}C_{CH4}$ values between −50 and−30‰V-PDB, whereas microbial CH_4 shows $\delta^{13}C_{CH4}$ values <−50‰ V-PDB (e.g., McCollom and Seewald, 2007). However, these values may vary depending on processes occurring during gas migration, such as isotopic fractionation by diffusion (Prinzhofer and Battani, 2003) or secondary methanogenesis and anaerobic biodegradation (Etiope et al., 2009). Thermogenic gases are produced by the decay of organic matter at T>150 °C (gases) and are commonly characterized by $CH_4/(C_2H_6+C_3H_8)$ concentration ratios (the so-called "Bernard parameter") minor than 100; whereas higher ratios (>1000) are expected when hydrocarbon production derives exclusively from microbial activity (Jenden et al., 1993). In this work, only for Tramutola site, we measured a $\delta^{13}C_{CH4}$ value of -63.1‰ (TRA 27-06-2021) and -62.4‰ (TRA 11-02-2022) VPDB, and $CH_4/(C_2H_6+C_3H_8)$ concentration ratio of about 260 (Table 3.10). These measured data are different respect with those reported by Etiope et al. (2007) $\delta^{13}C_{CH4}$ =-42.12‰) and Tassi et al. (2012) ($\delta^{13}C_{CH4}$=−47.1‰). In the $\delta^{13}C_{CH4}$ vs C1/(C2+C3) (Bernard et al.,1978) (Figure 3.48) the data fall in a mixing area between biogenic and thermogenic field, suggesting probably a modification of pristine carbon isotope signature due to secondary post-genetic processes. About that we propose that hydrocarbon origin in Tramutola fluids could be linked to either (1) biodegradation processes of thermogenic hydrocarbons, earlier generated by thermocracking process, or (2) ongoing microbial methanogenesis in the shallow organic-rich clays that interbed the porous media hosting the gas. Further analysis would be needed in the future to constrain

hydrocarbons origin in Tramutola site and eventually verify the occurrence of these secondary post-genetic processes.
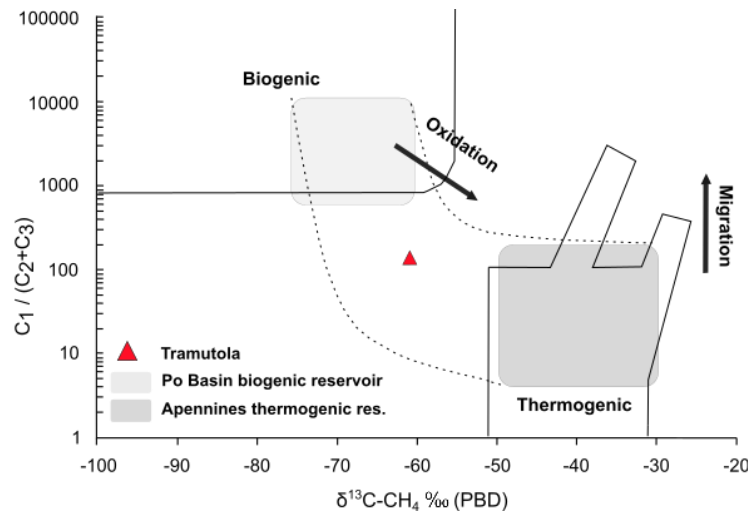
Figure 3. 48 – Carbon isotope vs. alkanes ratio (Bernard et al., 1978). Two mixing lines are plotted by assuming microbial Po Plain end-members (d13C = -76 to -60‰, C1/(C2 + C3) = 600 to 10000), and Apenninic thermogenic end-members (d13C = -50 to 30‰, C1/(C2 + C3) = 4 to 200) (modified from Etiope et al., 2007).

Sample ID	Sampling date	$\delta^{13}C_{CO2}$ ‰ V-PDB	$\delta^{13}C_{CH4}$ ‰V-PDB	$\delta^{13}D_{CH4}$ ‰V-PDB	$CH_4/(C_2H_6+C_3H_8)$
TRA	27/06/2020	-3.4	-63.1	-196	264
TRA	11/02/2022	-6.0	-62.4	-212	260

Table 3. 10 – Tramutola free gas samples measured C of CO_2, and C and H of CH_4 and $CH_4/(CH_2CH_6 + C_3H_8)$.

The CO_2 carbon origin is determined by means $\delta^{13}C$ values, which are diagnostic of the original geochemical environment. These, allow to discriminate between a magmatic source ($8‰ < \delta^{13}C < -4‰$; Sano & Marty, 1995), the contribution subducted marine limestone with $\delta^{13}C = 0‰$, and matter of organic origin with much lighter $\delta^{13}C = -25‰$ (Hoefs, 2015). As known from other studies in hydrothermal gases (Capasso et al., 2005), the $\delta^{13}C$ isotopic signature can be potentially modified by gas-water interaction in which CO_2 dissolves preferentially with respect to the other species. Thus, to constrain the origin of CO_2 in the fluids emitted at Tramutola well, we used the relationship between the elemental ratio $CO_2/^3He$ and the isotopic signature $\delta^{13}C_{CO2}$ (Sano & Marty, 1995; and references therein). The $CO_2/^3He$ systematic is useful to calculate the proportions of mantle, carbonate rocks (CAR), and organic sediments (ORG) in the total carbon composition. Volatile systematics for a hydrothermal degassing system is affected by the balance between volatile input from the subsurface (e.g., magma reservoir, deep fluid circulation) and volatile output to the surface as hydrothermal phase (e.g., degassing). In addition, values of $CO_2/^3He$ ratios depend on the lower solubility in aqueous fluids of He relative to CO_2. It follows that, the residual phase of a hydrothermal system would be characterized by high $CO_2/^3He$ ratio if the volatile output overbalances the volatile input and continuous hydrothermal degassing would also result in lowering of the He contents of the residual fluid. Moreover, secondary processes such as the loss of CO_2 in response to calcite precipitation can promote $CO_2/^3He$ and $\delta^{13}C$ values fractionation (Halldorsson et al., 2013). Figure 3.49 plots two mixing curves modelled considering both

an organic and a limestone endmember, assuming a starting mantle-like $\delta^{13}C$ composition of $-4‰$ and $CO_2/^3He$ $= 2 \times 10^9$ (Marty and Jambon, 1987) or $\delta^{13}C=-3.5‰$ and $CO_2/^3He = 7x10^9$ (Brauer et al., 2016). For the calculation of the mixing lines the following equations were used (Zhang et al., 2021):

$$\left(\frac{^{13}C}{^{12}C}\right)_{Obs} = f_M \times \left(\frac{^{13}C}{^{12}C}\right)_M + f_L \times \left(\frac{^{13}C}{^{12}C}\right)_L + f_S \times \left(\frac{^{13}C}{^{12}C}\right)_S \qquad \text{Eq.3. 20}$$

$$1/\left(\frac{^{12}C}{^3He}\right)_{Obs} = f_M/\left(\frac{^{12}C}{^3He}\right)_M + f_L/\left(\frac{^{12}C}{^3He}\right)_L + f_S/\left(\frac{^{12}C}{^3He}\right)_S \qquad \text{Eq.3. 21}$$

$$f_M + f_L + f_S = 1 \qquad \text{Eq.3. 22}$$

where Obs, M, L, and S represent the observed values for samples, mantle, CAR and ORG endmembers respectively. For CAR, and ORG endmembers, those defined in literature (Sano and Marty ,1995) where adopted: $\delta^{13}C=0\pm2‰$ and $CO_2/^3He = 10^{13}$ for CAR and $\delta^{13}C = 30 \pm 10 ‰$ and $CO_2/^3He = 10^{13}$ for ORG. A SLCM mantle endmember ($\delta^{13}C=-3.5‰$ and $CO_2/^3He = 7x10^9$, Brauer et al., 2016) was considered as the source of mantle carbon. TRA free gas samples (Figure 3.49) show low $CO_2/^3He$ ratios with respect to the Carbon endmembers, thus suggesting the influence in the phase partition of additional processes such as calcite precipitation (Hahm et al., 2008, Halldorsson et al., 2013) that can result in CO_2 loss and the consequent lower $CO_2/^3He$ ratios. The proposed interaction may be suitable for Tramutola waters since they rise across carbonatic rocks. An alternative explanation can be recalled if we consider that $CO_2/^3He$, He/CO_2, CH_4/CO_2 ratios, and $\delta^{13}C$ isotopic signature can be potentially modified by gas-water interaction in which CO^2 dissolves preferentially with respect to He and CH_4 gas, thus causing a decrease in the $CO_2 /^3He$ ratio.
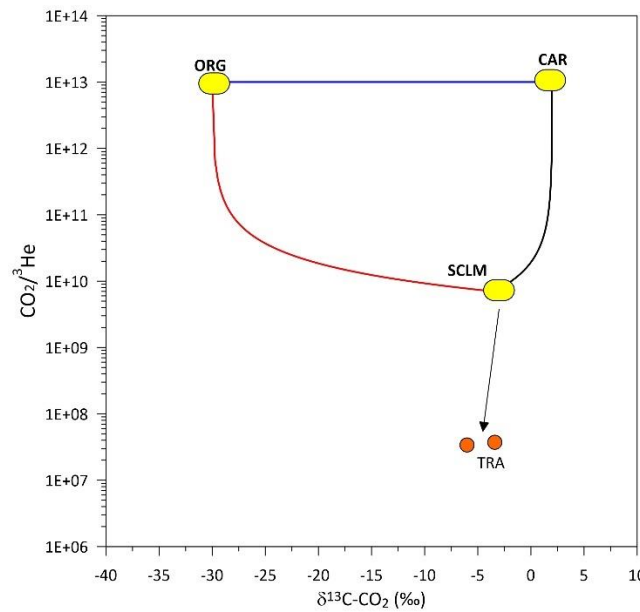
Figure 3. 49 - Plot of $CO_2/^3He$ versus $\delta^{13}C$-CO_2 (‰). Abbreviations: SCLM, sub-continental lithospheric mantle; CAR,carbonate; ORG, organic matter.

3.4.5.5 Methane and Carbon dioxide flux estimations

The methane output emitted from the Tramutola well was evaluated by means of anemometric measurement of the volume flow (m^3/h) and flow velocity (m/s). The analysis was performed through a *testo 400* control unit provided of a vane anemometer (Ø=100mm) for flow velocity measurement with ±0.01 m/s resolution and with a funnel of 200 mm diameter (testovent 417) for volume flow measurements (Figure 3.50). It was positioned directly over well, after a metal structure with a diameter equal to the well has been mounted on the well, and then covered by a plastic sheet assured to the flow funnel by means of a perforated band in pre-galvanized steel and stainless-steel hose clamps (Figure 3.51). This enabled us to efficiently convey most of the analyzed gas flux through the measurement funnel. Flow velocity and volume flow data were simultaneously collected in a one-hour long time window (Time UTC: 2021-04-21 from 13:22:23 to 14:22:23) and with the same sampling period of 1 second. The resulting measurement time-series (Figure 3.52) shows slight variations in the analyzed timespan for the two parameters with the volume flow ranging between a minimum of 21.6 and maximum peak of 30.1 m3/h (mean value of 26.6 m3/h) which reflect velocity flow variations in the range between 0.92 and 1.25 m/s, with an average flux velocity of 1.11 m/s.

Figure 3. 50 - testo 400 control unit used for anemometric measuremets at Tramutola well, vane anemometer (Ø=100mm) probe for flow velocity measurement and the provided funnel set (testovent 417) of 200 mm diameter (on the right) and 300 mm diameter (in the left).

Figure 3. 51 – Anemometric measurement of free gas output setting at Tramtuola well. The measurement system was covered by a plastic sheet assured to the flow funnel by means of a perforated band in pre-galvanized steel and stainless-steel hose clamps to better isolate the emitted free gases from the atmospheric ones.

Figure 3. 52 – Resulting time-series of anemometric measurement of the volume flow (m3/h) (green line) and flow velocity (m/s) (yellow line)

Known CH_4 (82.625%) and CO_2 (1.715%) values, the volume flow measurements were used for the estimation of their total output from. Since the measurement system was not completely isolated, we considered a volume flow 10% error in the measurements, thus considering a respectively higher flow with respect to recorded mean and equal to Q=29.3 m³/h (corresponding to 8.13 L/s). The total output from bubbling gases ($G_{free}(L/s)$) of CH_4 and CO_2 was calculated through the following general formula:

$$G_{free}(L/s) \ = \ Q \ (L/s) \ x \ \frac{G \ (\%)}{100}$$

Eq.3.25

Where Q is the mean volume flow expressed in l/s and G is the gas concentration (%) measured in the bubbling gas samples. From Eq.3.25 a total CH4 gas flux ($CH_{4,free}$) of 6.72 l/s and a total CO2 gas flux ($CO_{2,free}$) of 0.139 L/s were estimated. The corresponding values expressed in liter-per-year are shown in Table 3.11. Considering the gas molar volume value (22.414 L/mol) in STP conditions (T=0°C, P=1 atm), we estimated the total output (mol/s) as:

$$G_{free}(mol/y) = \frac{G_{free}(L/y)}{22.414 \left(\frac{L}{mol}\right)}$$

Eq.3. 26

A total gas flux of 9.45 x 10^6 and 1.96 x 10^5 mol/y were measured for CH$_4$ and CO$_2$, respectively. From the calculated $G_{free}(mol/y)$ in Eq. 3.26, the total tons/year CO$_2$ and CH$_4$ outputs were estimated as:

$$G_{free}(t/y) = \frac{G_{free}(mol/y) \times m_G}{10^6}$$

Eq.3. 27

Where m_G is the molecular mass of the gas ($m_{CH4} = 16.04$ g/mol; $m_{CO2}= 44.01$ g/mol). We obtained a total CH$_4$ output at the Tramutola well ($CH_{4,free}$) is equal to about 152 t/y and a total gas flux $CO_{2,free}$ of 8.6 t/y (Table 3.11). In addition, we considered the dissolved methane aliquot in water, applied the following:

$$CH_{4,diss}(cm^3/s) = CH_{4,diss}(cm^3 STP/L) \times Q$$

Eq.3. 28

Where $CH_{4,diss}$ is the dissolved methane amount, measured in $cm^3 STP/L$ from gas chromatography analysis; Q is the mean water flow (10.02 L/s) estimated at the Tramutola well. A $CH_{4,diss}= 232.9$ cm^3/s. Following a similar procedure as the one used for free gases, the total dissolved CH$_4$ flux (mol/s) was estimated considering the equivalent molar volume of an ideal gas in a 1 cm^3 volume:

$$CH_{4,diss}(mol/s) = \frac{CH_{4,diss}(cm^3/s)}{22.4 \times 10^3}$$

Eq.3. 29

A total $CH_{4,diss}= 0.010$ mol/s was obtained. Using the corresponding value of 3.26 x 10^5 mol/year, from Eq.3.26 the total dissolved methane output from Tramutola well was estimated as equal to 5.23 t/y (Table 3.11). The previous calculations of bubbling ($CH_{4,free}$) and dissolved ($CH_{4,diss}$) component of methane output allowed us to estimate the total methane output flux at Tramutola well as the sum:

$$CH_{4,tot} = CH_{4,free} + CH_{4,diss}$$

Eq.3. 30

A total flux amount of $CH_{4,tot}= 156.41$ t/y was measured, less than the one estimated at Tramutola well from other authors of about 2400 t/y (Etiope et al., 2007, Italiano et al., 2000).

We compared the results with other non-volcanic CH$_4$-dominated gas manifestations (i.e., mud volcanoes, dry-seeps, bubbling water wells, Etiope et al., 2007) recognized in Italy (Etiope et al., 2002, 2007; Italiano et al., 2000) (Figure 3.53). The biggest mud volcano in Italy (Maccalube, Sicily) emits about 400 t/y (Etiope et al., 2002), about twice the total methane output estimated for the Tramutola well. However, the comparison shows that methane output at Tramutola is one order of magnitude higher of most of the analyzed emissions, and one

of the highest CH4 outputs in Italy, such that it significantly contributes to the total CH4 emission into the atmosphere. The measured $CH_{4,tot}$ has a order of magnitude equal to 1.5% of the total national anthropogenic methane sources related tofossil fuel industry (Etiope et al., 2007). With a similar approach, we compared the estimated total budget of mantle-derived CO_2 estimated from the analyzed free gas samples at Tramutola with the main volcanic and non-volcanic (e.g., Mefite, Contursi) CO_2-dominated gas emissions in Italy (Caracausi et al., 2015, and reference therein) (Figure 3.54). Results show that the estimated CO_2 output at Tramutola is much lower with respect to the analyzed emissions: at least six orders of magnitude with respect to the highest volcanic emission (Etna volcano, 2×10^{11} mol/y) and four order of magnitude with respect to the largest non-volcanic emission (Mefite d'Ansanto 7×10^9 mol/y). The latter is the largest non-volcanic, mantle-derived, CO_2 emission ever measured in the Earth (Chiodini et al., 2010) with a total budget comparable with the one of volcanoes (Figure 3.54) and will be analyzed with more detail in the next chapter.
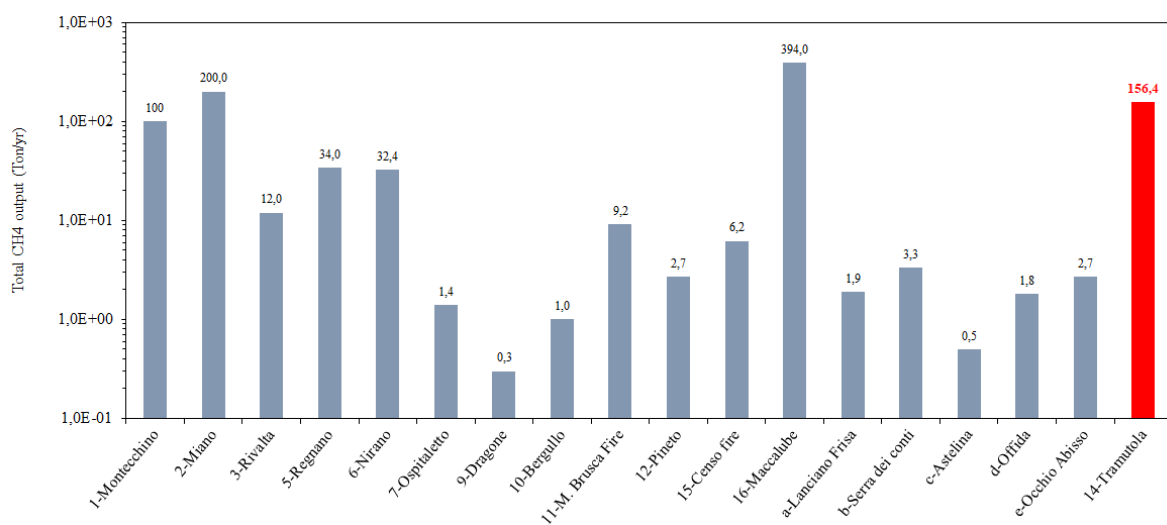
Figure 3. 53 - Comparison of the estimated CH4 total budget (t/y) at Tramutola well (red bar) with other non-volcanic CH4-dominated gas manifestations (Etiope et al., 2002, 2007, Italiano et al., 2000) recognized in Italy.
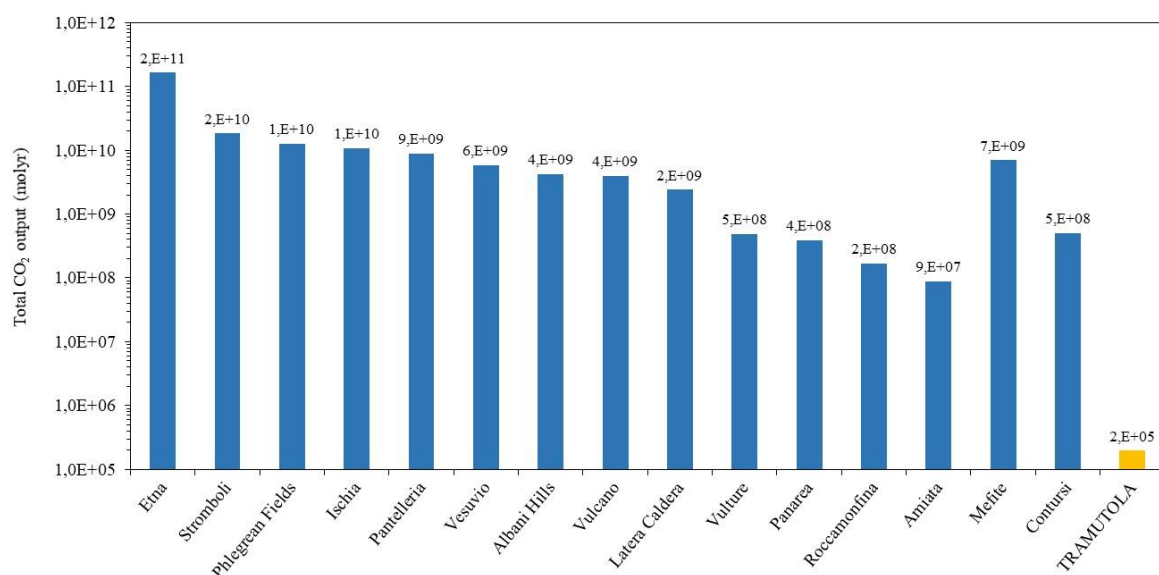
Figure 3. 54 - Comparison of the estimated CO2 total budget (mol/y) at Tramutola well (yellow bar) with the main volcanic and non-volcanic mantle-derived CO2 emissions (Caracausi et al., 2015) in Italy.

	Gas	Mean Volume flow Q L s^{-1}	G %	Gas total output L/y^{-1}	Gas total output mol/y^{-1}	Gas total output t/y^{-1}
free gases	CH$_4$	8.13	82.625	2.12E+08	9.45E+06	151.18
	CO$_2$	8.13	1.715	4.40E+06	1.96E+05	3.14

	Gas	Mean Water flow Q L s^{-1}	G cm^3 L^{-1} (STP)	G cm^3 s^{-1}	Gas total output (mol/y^{-1})	Gas total output (t/y^{-1})
dissolved gases	CH$_4$	10.02	23.2	232.39	3.27E+05	5.23

Table 3. 11 – CH$_4$ and CO$_2$ output estimations for Tramutola well free gases samples (yellow table) and CH4 output estimation from dissolved gases (orange table) obtained from TRA 26.06.2020 sample.

3.4.5.6 A conceptual model for the gas uprising in the old Tramutola borehole

The diagram in Figure 3.55 summarizes the proposed fluid mixing paths reconstructed for Tramutola thermal waters and the locations of the three potential gas sources and their possible mixing processes:

(a) the surficial one would be located at depth between 30-150 m, in the silicic limestones belonging to the Lagonegrese Units. Its composition (N$_2$ + O$_2$) would be mainly related to atmospheric gas and rainwater entering the system through water infiltration (N$_2$ + O$_2$). A small additional CO2 amount would be related to dissolution processes of biological CO$_2$.

(b) the crustal CH$_4$-dominant gas source, was identified in correspondence of the hydrocarbon reservoir within the limestones belonging to the Apulian Platform Units which was identified at 2000-2200 m in the Tramutola area (Balasco et al., 2021). Here CH$_4$ would be related to thermogenic hydrocarbon origin, which subsequently underwent biodegradation processes. Additional CO$_2$ in this reservoir could be probably related to the anaerobic hydrocarbon oxidation produced by biodegradation processes. Helium signatures in the Tramutola fluids are related to a prevailing crustal component (^{4}He).

(c) The deeper, mantle-gas source is testified by the measured He isotopic ratio. According to the literature, the Moho discontinuity runs parallel to the bottom of the Apulian Platform and reaches a depth of about 25-35 km in the study area (i.e., Adriatic Moho, Menardi and Noguera Rea, 2000, Megna et al., 2014). The significant mantle He signature at Tramutola would be linked to mantle-derived fluids uprising through lithospheric normal faults as demonstrated in other Apennine's areas (e.g., Burnard et al., 2012; Caracausi and Sulli, 2019). These tectonic structures would affect both the deeper Apulian Platform Limestones and overlying Liguride Complex featured at depth by listric geometry as suggested by reflection seismic profiles (Menardi and Noguera Rea, 2000). Fluid circulation in such deep structures below the well are compatible with high Vp/Vs values highlighted from Valoroso et al. (2011) in the area. An additional mantle-derived CO$_2$ component is hypothesized as suggested by $\delta^{13}C_{CO2}$ measured values (-3.5 ‰). However, since the already

supposed crustal and surficial sources of CO_2, further investigation would be needed in the future on the CO_2 mass balance from the different reservoirs to justify the measured value of $\delta^{13}C_{CO2}$ at surface.
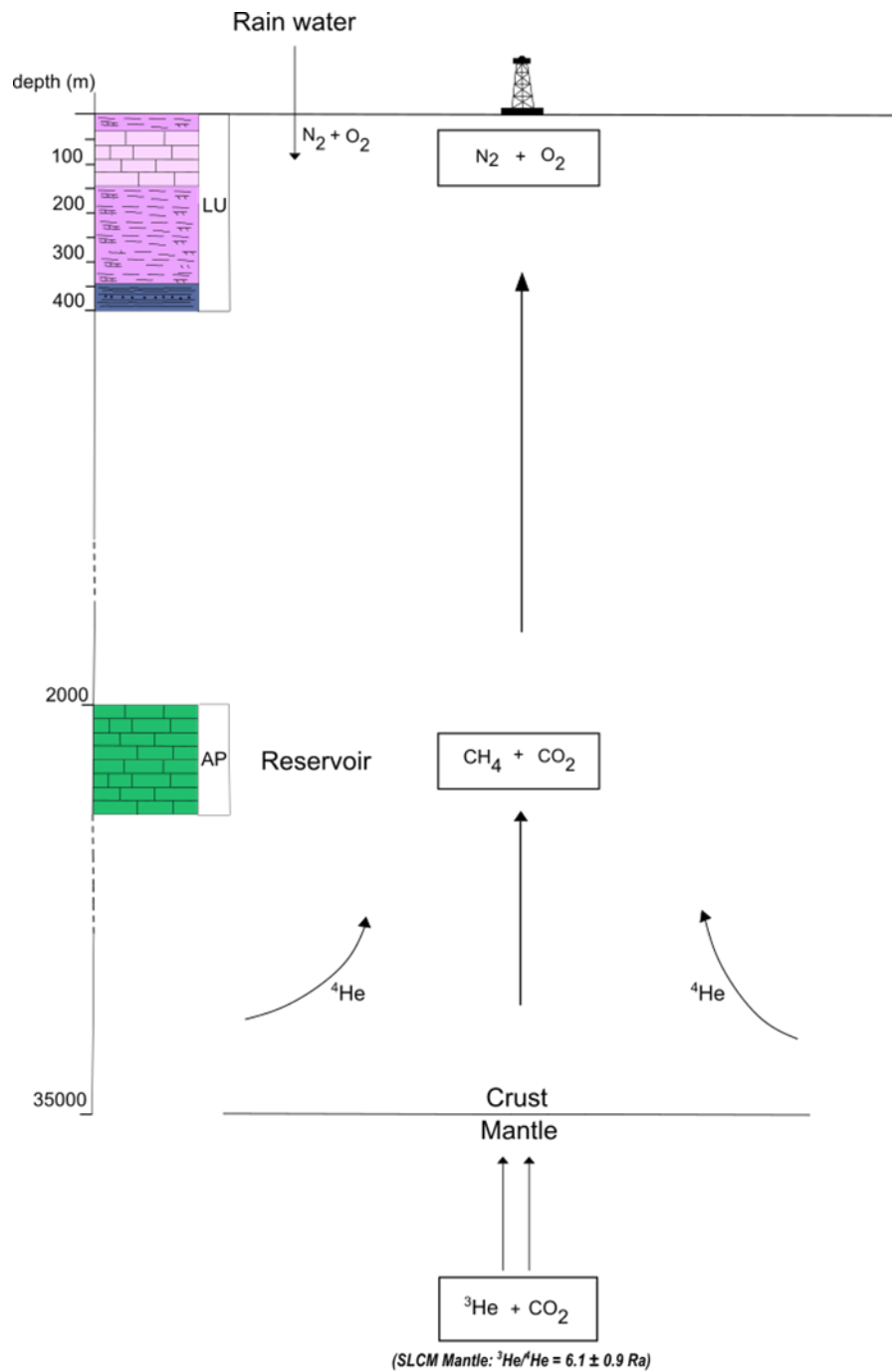
Figure 3. 55 – Scheme of the proposed fluid mixing paths reconstructed for Tramutola thermal waters and the locations of three potential gas sources and their possible mixing processes (depth axis is not in scale).

3.4.5.7 Preliminary observations on the continuous well monitoring

In this work, preliminary observations on the Tramutola well continuous monitoring are reported. An OTT ecoLog 800 (3G) groundwater datalogger was installed into the well during the last few months of the PhD project, on 24 April 2021 (Figure 3.56). The multi-parametric geochemical station (station name: 001 TRAMUTOLA) is provided with three different probes to collect high-resolution, high-sampling frequency data for the continuous monitoring of the groundwater level and physico-chemical parameters at the well, such as temperature and specific electrical conductivity. The OTT ecoLog 800 (3G) measured values are transmitted in near real-time through a GSM connection to the Data Center located in the Geochemical Environmental Laboratory at the University of Basilicata (Department of Science, Potenza, Italy).

Figure 3. 56 - OTT ecoLog 800 (3G) groundwater datalogger installation at Tramutola well.

Water level data are collected with a ± 0.01 m resolution pressure probe equipped with a relative pressure probe cell, through hydrostatic pressure measurements of the water column. The OTT ecoLog 800 (3G) measures the specific electrical conductivity with a ± 0.001 μS/cm resolution using a 4-graphite electrode conductivity sensor. The electrical conductivity meter is integrated with a temperature sensor (resolution of ± 0.001 °C). A high sampling frequency was chosen to obtain a high-resolution monitoring of the collected parameters: (a) 1-minute sampling period for water level and (b) 5-minutes sampling period for both electrical conductivity and temperature. Figure 3.57 (for temperature and conductivity) and Figure 3.58 (for water level) show preliminary results of the near-continuous monitoring of the Tramutola well in measured parameters time-series between 03 August and 03 November 2021. Indeed, experimental data collected during the first 3-months test period were excluded since contaminated from external effects related to well cleaning issues. Water temperature variations in the analyzed period are showed in Figure 3.58a. In agreement with the collected TRA samples, temperature of the thermal waters at Tramutola well recorded slight variations in the analyzed timespan with a median value equal to 27.86 °C. The high sampling frequency allowed to record daily water temperature fluctuations of about 0.03 °C (Figure 3.57a) and to observe a general decreasing seasonal trend: the higher temperatures were recorded in August (with a maximum value of 27.95 °C on 13th August 2021); the minimum temperature of 27.76 °C was recorded on 15th October 2020 coinciding with the occurrence of anomalously cold weather conditions. The average electric conductivity in the analyzed period

is of 2630 μS/cm. Figure 3.57b shows high variability of the measured conductivity values with a non-stationary alternation of maximum lobes and minimum peaks which are not correlated to the nearly constant temperature variations. Such a high-frequency conductivity fluctuations may be rather explained as the effect of water bubbling at the Tramutola well. Further observations in a longer timespan would be needed in the future to better comprehend the observed conductivity variation pattern. Negligible water level variation was measured in the period under analysis with a median value equal to 0.208 m (Figure 3.58). The maximum fluctuation between the highest measure (0.214 m) and the minimum one (0.196) of about 2 centimeters (0.18 m) was recorded just after the well cleaning occurred on 03/08/2021, followed by water level increase and subsequent stabilization.
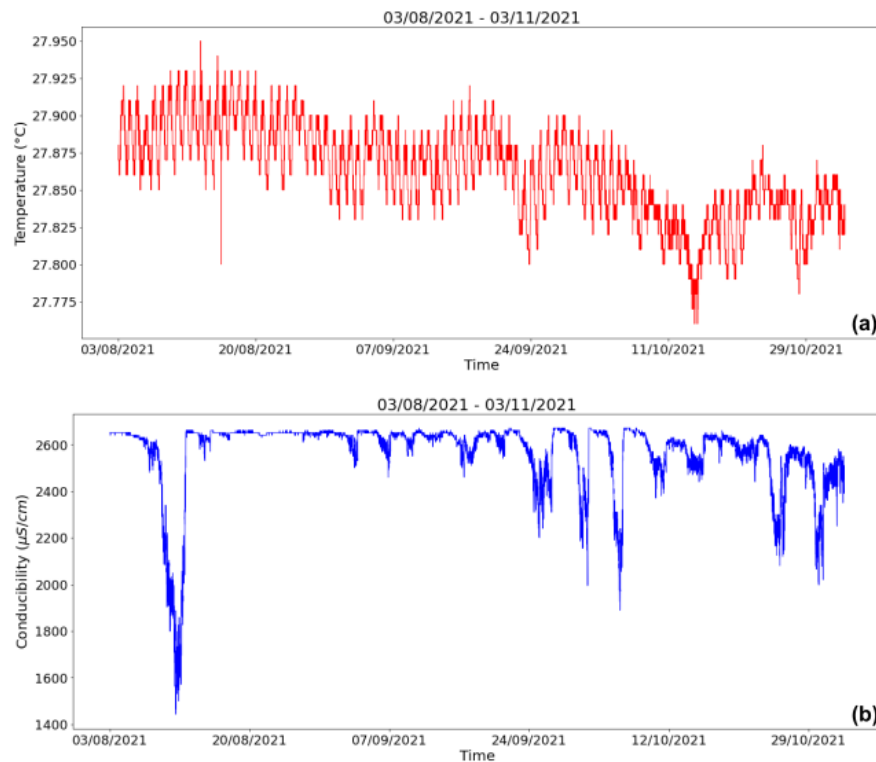
Figure 3. 57 – Tramutola well measured parameters time-series: temperature (a) and electrical conductivity (b).

We compared the observed groundwater level time-series with the total daily precipitation data (mm) recorded from a weather station located in Tramutola (Station ID: ITRAMUTO2; Latitude: 40.353°N; Longitude: 15.81°E; https://www.wunderground.com/dashboard/pws/ITRAMUTO2) (Figure 3.58). Mainly absent and scarce precipitation were recorded in the area in the analyzed timespan, with two main rainfall events occurred on 05th August and 10th October and characterized by a cumulative precipitation of 22.86 mm and 24.13 mm, respectively. No correlation between precipitation and groundwater level can be observed in the analyzed period. Moreover, the measured groundwater level was preliminarily compared with the time series of 14 Mw≥6.5 teleseisms occurred worldwide in the analyzed timespan (source: European-Mediterranean Seismological Center, EMSC) (Figure 3.58, Table 3.12). Indeed, Barberio et al. (2020) recently identified co-seismic and post-seismic groundwater level fluctuations in the Central Apennines related to the ground motion produced by Rayleigh seismic waves deriving from distant, high magnitude (Mw≥6.5) earthquakes. No clear groundwater level perturbation can be observed at this scale of the investigation, especially for earthquakes

with Mw ≤ 7.0. However, the two largest earthquakes - the 12/08/2021 18:35:20 (UTC) Mw=8.1 event (South Sandwich Islands region) and the 02/10/2021 06:28:18 (UTC) Mw=7.3 earthquake (Vanuatu region) – seems to coincide in time with negative water level peaks, followed by a water level increase (Figure 3.58). Further detailed investigations would be needed in the future to verify and better understand eventual anomalous impulsive change of the groundwater level at Tramutola well produced from both teleseisms (Barberio et al., 2020) and local earthquakes, which location data are currently not available. However, our preliminary analysis highlighted how the continuous monitoring of groundwater level at Tramutola well and the better definition of a model for its thermal fluids uprise would be a key tool for future investigation on the scientific still open-issue of co-seismic and post-seismic groundwater-level responses to earthquakes and their possible role as geochemical indicator for seismic hazard and monitoring purposes.
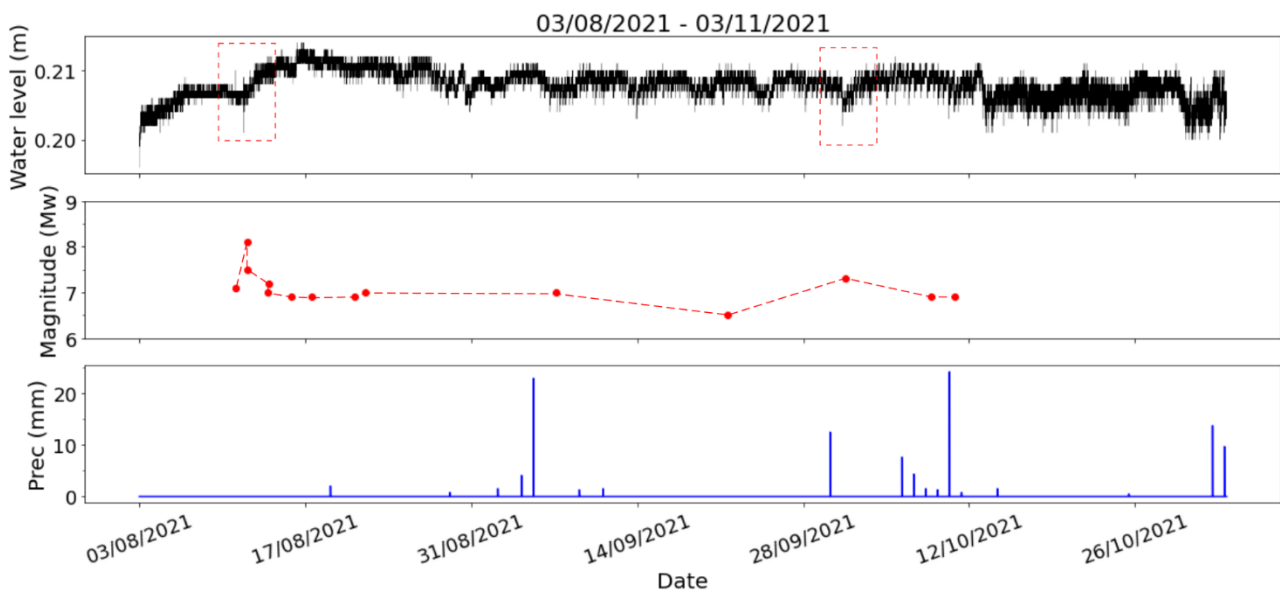
Figure 3. 58 - Comparison of the measured groundwater level at Tramutola well (black line) with Mw≥ 6.5 earthquakes (red dashed line) and the total daily precipitation data (mm) (blu line) occurred in the analyzed timespan.

Date	Origin Time (UTC)	Latitude (DD)	Longitude (DD)	Mw
11/08/2021	17:46:13	6.54	126.73	7.1
12/08/2021	18:32:55	-57.57	-25.46	7.5
12/08/2021	18:35:20	-58.52	-25.26	8.1
14/08/2021	11:57:42	55.32	-157.51	7.0
14/08/2021	12:29:09	18.36	-73.48	7.2
16/08/2021	11:10:36	-58.44	-24.45	6.9
18/08/2021	10:10:05	-14.87	167.16	6.9
22/08/2021	00:45:08	-60.17	-24.43	6.9
22/08/2021	21:33:20	-60.32	-24.99	7.0
08/09/2021	01:47:46	16.98	-99.77	7.0
22/09/2021	09:57:08	12.16	-87.78	6.5
02/10/2021	06:29:18	-21.10	174.91	7.3
09/10/2021	10:58:28	-21.11	174.55	6.9
11/10/2021	09:10:24	56.41	-156.67	6.9

Table 3. 12 - Mw≥6.5 teleseisms origin time (UTC), latitude (decimal degrees), longitude (decimal degrees) and Magnitude (Mw)occurred worldwide between 03 August and 03 November 2021 (source: European-Mediterranean Seismological Center, EMS

Chapter 4 – Mefite d'Ansanto study area

4.1 Introduction

Natural, non-volcanic CO_2 gas emissions are often located in tectonically active areas, where deep fluids migration and ascent on the surface is enabled by fault systems and highly fractured rocks (i.e., Brune et al., 2017; Caracausi and Paternoster, 2015; Chiodini et al., 2004; Italiano et al., 2009; Tamburello et al., 2018). It is well documented that tectonic discontinuities could act as preferential conduits for deep fluids circulation within the crust, guiding their migration patterns in the subsurface (e.g., Caracausi et al., 2015, 2013; Di Luccio et al., 2018). Moreover, the study of the relationships between non-volcanic CO_2 gas emissions and large earthquakes generation have been receiving ever greater interest since (a) the observed close spatial correlation between seismically active and deep fluid discharge areas (Chiodini et al., 2020, 2004) and (b) the potential fluids role in earthquake nucleation processes (Miller, 2013; Tamburello et al., 2018). Chiodini et al. (2004, 2020) showed an intriguing spatial relationship between deep-source CO_2 gas reservoirs and earthquake locations in the Apennine chain (Italy), where widespread non-volcanic, natural CO_2-rich gas-emissions occur (Ascione et al., 2018). These authors observed a remarkably drop in CO_2 degassing along the chain axial zone, currently undergoing extensional tectonics and where the seismicity is concentrated, thus suggesting a direct relationship between Apennine's earthquakes nucleation and over-pressurized deep CO_2 gas reservoirs. Co-seismic release of crustal-trapped over-pressurized fluids have been identified as one of the main triggering mechanisms of aftershocks sequences related to large earthquakes (Miller et al., 2004); while, fluid pore pressure role in reactivation of pre-existing faults may have been one of the main favoring factors of large ruptures associated with the most damaging seismic sequences occurred in Italy in the last decades such as the 2009 L'Aquila (Di Luccio et al., 2010) and 2016-2017 Amatrice-Norcia (Chiarabba et al. ,2018) ones. On the other hand, fluid pore-pressure role in reducing faults strength and consequently favoring earthquakes nucleation, have been widely recognized as a triggering factor of anthropogenic induced / triggered seismicity related to fluid injections in the subsurface (Zoback & Harjes 1997, Shapiro et al. 2006) or related to water level load fluctuations in artificial lakes (Stabile et al., 2014). However, constraining the fluid role in tectonic earthquakes triggering mechanisms is difficult since it is not easy to discriminate their effects from those of other driving forces such as aseismic slip (e. g. (Lohman and McGuire, 2007; Vidale and Shearer, 2006). It follows that monitoring non-volcanic gas emissions is a key point to assess the complex relationships between fluids migration in the subsurface, Earth degassing and large earthquakes generations. This could be crucial in areas characterized by high seismic hazard, such the Apennine chain, currently considered one of the highest hazard areas in Europe. Here is located the Mefite d'Ansanto study area: it is the largest natural, non-volcanic, CO_2-dominant gas manifestations recognized along the chain (Ciotoli et al., 2014, 2007; Caracausi and Paternoster, 2015) and it occurs at the northern end of the faults system which generated one the most damaging earthquakes occurred in Italy in the last decades, the Mw= 6.9 1980 Irpinia earthquake. Therefore, the better understanding of the complex evolution of the Mefite d'Ansanto and monitoring it spatiotemporal evolution could be a key tool to address the role of crustal fluids in the generation of large earthquakes in Irpinia.

In the last decades ever increasing attention have been given by the scientific community to CO_2 gas emissions from the environmental point of view, since their substantial contribution to the present-day global carbon output in the atmosphere and their influence on climate (Chiodini et al., 2004,Frondini et al., 2019; Hunt et al., 2017; Lee et al., 2016). Indeed, the main driver of climate change has been recognized the global surface temperature rise, directly correlated with the CO_2 emissions to the atmosphere. The increasing manifestations of an already started climate change made necessary an energy transition from high- to low-carbon centered. For this reason, the development of Carbon Capture and Storage (CCS) techniques for both CO_2 utilization (e.g concrete, aggregates, polymers, algae agriculture, and algae fuels) and sequestration in subsurface reservoirs, became an increasingly important challenge for the scientific, the industrial and the political community given the urgent need to reduce CO_2 emissions to reverse the trend of the global climate warming (Lau et al., 2021). Given its massive natural CO_2 emission ($\approx$ 2000 tons/day) and the geological setting in which it arises, the Mefite d'Ansanto offers a unique opportunity for measuring, modeling and develop monitoring techniques of the effects of a large gas leakage from a CO_2 buried gas reservoir of a CCS injection project (Chiodini et al., 2010; Pruess, 2008).

So far, background seismicity at Mefite site have been analyzed only in terms of ground motion polarization from ambient noise measurements (Pischiutta et al., 2013). In this work, we upgraded the current knowledge about the emission-related seismic tremor at Mefite d'Ansanto, currently poorly quantitatively analyzed. Seismic tremor in volcanic areas is an extensively studied phenomenon, due to its recognized association with the increase of volcanic activity and its role in forecasting incipient eruptions as well as in understanding the interaction between magmatic fluids and the surrounding rocks. Recent works (Chiodini et al., 2017; Giudicepietro et al., 2019) demonstrated that seismic tremor characterization in volcanic gas emissions areas (e.g., fumarolic vents) is a powerful tool to monitor volcanic unrest related to hydrothermal activity and for risk mitigation purposes. Conversely, in non-volcanic areas, seismic tremors from gas emissions have been hardly analyzed (Pischiutta et al., 2013). Nevertheless, its quantitative characterization has great potential in shedding light on the geodynamics of tectonic complex degassing regions as well as in monitoring the close relationship between seismic activity and deep fluid discharge. We developed an automated detection algorithm and applied a machine-learning based classification approach which enabled the automated Mefite tremor characterization and discrimination, thus providing a starting workflow to monitor the non-volcanic emission either to gain information on its evolution for seismic hazard purposes or providing an early step-forward in the automated management techniques of both natural and CCS leakage CO_2 gas-emissions.

4.2 The Mefite d'Ansanto

Known since the V century BC (Lisio et al., 2014), the Mefite d'Ansanto is the largest natural emission on Earth of non-volcanic, CO_2-rich gases with an estimated total gas flux of about 2000 tons per day (Chiodini et al., 2010). The emitted gas composition consists of CO_2 (98 vol%) with minor amounts of N_2 (1.3 vol%), H_2S (0.33 vol%) CH_4 (0.23 vol%) (Chiodini et al., 2010). It occurs in the Irpinia area (Campania Region, Italy) in Southern Apennines, one of the highest seismic hazard areas in Europe. The East-verging thrust belt was

originated from the late Oligocene-early Pleistocene west-dipping subduction of the Adriatic plate (Cello et al., 2000; Doglioni et al., 1999; Gueguen et al., 2015) and it has been undergoing a NE-SW extensional tectonic since Middle-Late Pleistocene which currently control the moderate to strong seismicity ofthe area. Extensional processes mainly involve the axial zone of the chain and developed up to 15 km deep NW–SE striking fault systems (Chiodini et al., 2010; Cinque et al., 1993) which can generate up to M 7 earthquakes (e.g., Mw 6.9, 1980 Irpinia Earthquake, Del Pezzo et al., 1983) and, along with, the gas leakage is favored. Mefite d'Ansanto developed at the northern end of the fault system activated by the Irpinia earthquake in the well-known thermal anomaly area of the Mt. Forcuso (Festa et al., 2021; Improta et al., 2014, 2003). It is the largest of natural, non-volcanic, CO_2-dominant gas manifestations extensively recognized along the Southern Apennines active fault systems (Chiodini et al., 2004; Ciotoli et al., 2014, 2007, Caracausi and Paternoster, 2015b), which deep origin is suggested by their isotopic compositions (i.e., ^{13}C and $^{3}He/^{4}He$)(Ascione et al., 2018; Caracausi et al., 2013; Chiodini et al., 1999; Italiano et al., 2009) and is attributed to both upper mantle degassing and/or crustal metamorphism of the Apulia limestones belonging to the Adria subducting plate.

Geologically, the Mefite area is characterized by calcareous-siliciclastic flysch deposits (Upper Trias – Middle Cretaceous) and deep-water sediments belonging to the Lagonegro basin and the Argille Scagliose units (Sicilide complex) (Mostardini and Merlini, 1986, Pischiutta et al., 2012). The region surrounding the emission area is geometrically characterized by alternating and NW-striking synform and antiform structures: the Paternopoli (PS) and the Trevico synforms (TS), separated by the broad Frigento antiform (FA). Improta et al. (2000) interpreted these structures as an effect of the irregular morphology of the top of the Apulia Platform Carbonates related to thrust ramp-flat structures. (Patacca and Scandone, 2001) proposed a different interpretation for this structural pattern, suggesting that Frigento antiform derives from a duplex structure within the Apulia limestones. The Mefite d'Ansanto is located above the structural high point of the NW – SE-trending Frigento Antiform. Which show outcropping, siliceous claystones with limestones (Unità del Sannio and Unità del Torrente Calaggio, Scrocca et al., 2007). The analysis of the geological map produced by the ''SISCAM'' project (Pischiutta et al., 2012; Vilardo et al., 2009) revealed the presence of a fault that changes its strike from NW–SE (in the northernmost sector) to NNW–SSE (in the southernmost sector) of the Frigento area. This change occurs close to the Mefite d'Ansanto emission, but this fault segment is not mentioned in literature. Its nature and kinematic are still debated, also due to the absence of fault zone outcrops and to currently poor constraining data from field investigations in the area due to the bad rheology and the high alteration of the outcropping geological units at Mefite.

The Mefite d'Ansanto emission is fed by a 1128 to 1600 m-deep gas reservoir which was encountered by the 1850 m deep Monte Forcuso 001 well (https://www.videpi.com/videpi/pozzi/dettaglio.asp?cod=3920). It was drilled at about 2 km east from the emission and allowed to recognize a reservoir gas pocket which sits at the top of permeable limestones ("Piattaforma Apula Interna" Formation, Mesozoic) covered by a clayey cap formation ("Unità Lagonegresi", Miocene*)* (Chiodini et., 2010). At Mefite the continuous gas manifestation has been probably active since the beginning of the extensional tectonic in Southern Apennines, due to the combination of the constant leakage by the buried reservoir and recurrent faulting processes. At surface the emission composition extends in an area of about 4.000 m^2 (Chiodini et al., 2010) and occurs mainly as diffuse

soil degassing with local bubbling pools and pressurized vents, with the gas flowing toward SW along a tectonically driven narrow river as evidenced also by the lack of vegetation in the area (Figure 4.1).
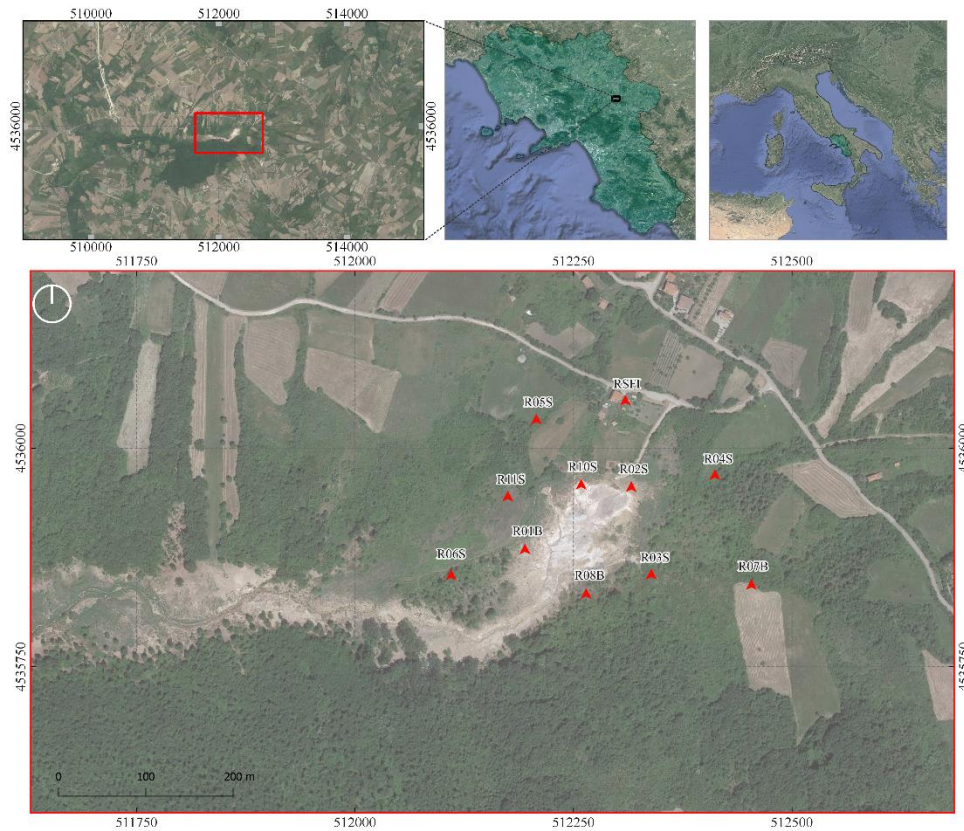
Figure 4. 1 – The Mefite d'Ansanto gas emission location in the Irpinia area (Campania Region, Italy). The lower image shows a detail of the emission site, recognizable from the lack of vegetation, and the location of the 11 temporary seismic station of the Mefite d'Ansanto array.

4.3 The Mefite d'Ansanto array

With the improvement of sensor technology and the decrease of its costs, the deployment of dense temporary arrays became a popular tool in seismology. These seismic networks are spatially denser than the permanent networks thus furnishing a high-resolution imaging and spatio-temporal monitoring of faults zones and volcanic areas. The improved resolution provided by closely spaced stations is used to record smaller earthquakes and other sources of seismic radiation (e.g., tremor and anthropogenic sources) which cannot be detected from regional networks (e.g., Inbal et al., 2016; Cara et al.; 2019, Gradon et al., 2019). The background seismicity at Mefite d'Ansanto was collected by a temporary dense seismic network (code: GEOFON YZ, doi:10.14470/LM582526) installed at the emission between 28-10-2019 and 05-11-2019 by the German Research Center for Geoscience (GFZ) as part of a temporary experiment (East Pollino Experiment, Southern Italy, Passarelli et al., 2017), with the aim to monitor the interaction between crustal fluids and earthquake occurrence. The dense seismic array was installed around the emission site and consisted of 11 surface seismic stations (Figure 4.1 and Table 4.1). The maximum distance between stations was of 340 m with a maximum slope of 44 m. Two kinds of triaxial seismometers were installed: (a) 4 Broadband sensors (Trillium-Compact

120s) installed at R01B, R07B, R08B and RSFI; (b) 7 Short-period sensors (SM-6/g=28) at stations R02S, R03S, R04S, R05S, R06S, R10S, R11S. Continuous waveforms recordings were acquired by the YZ network with a CUBE Datalogger at 200 samples per second. Table 4.1 shows the summary of station locations, characteristics, and installation times.

Station Code	Network	Longitude (DD)	Latitude (DD)	Altitude (m)	Sampling Rate (Hz)	Seismometer	Installed	Uninstalled
R01B	YZ	15.1448	40.9740	709.80	200	Trillium-Compact-120/GIPP/g=75	28/10/2019 00:00	04/11/2019 00:00
R02S	YZ	15.1463	40.9747	727.35	200	SM-6/g=28	28/10/2019 00:00	05/11/2019 00:00
R03S	YZ	15.1466	40.9738	714.44	200	SM-6/g=28	29/10/2019 00:00	05/11/2019 00:00
R04S	YZ	15.1474	40.9748	725.86	200	SM-6/g=28	29/10/2019 00:00	05/11/2019 00:00
R05S	YZ	15.1450	40.9754	727.27	200	SM-6/g=28	29/10/2019 00:00	05/11/2019 00:00
R06S	YZ	15.1438	40.9738	703.04	200	SM-6/g=28	29/10/2019 00:00	05/11/2019 00:00
R07B	YZ	15.1479	40.9737	722.37	200	Trillium-Compact-120/GIPP/g=75	29/10/2019 00:00	05/11/2019 00:00
R08B	YZ	15.1457	40.9736	704.83	200	Trillium-Compact-120/GIPP/g=75	28/10/2019 00:00	05/11/2019 00:00
R10S	YZ	15.1456	40.9747	715.45	200	SM-6/g=28	29/10/2019 00:00	05/11/2019 00:00
R11S	YZ	15.1446	40.9746	718.6	200	SM-6/g=28	29/10/2019 00:00	05/11/2019 00:00
RSFI	YZ	15.1462	40.9756	748.54	200	Trillium-Compact-120/GIPP/g=75	28/10/2019 18:01	05/11/2019 16:30

Table 4. 1 – Summary of Mefite array station locations, characteristics, and installation times.

4.3.1 Calibration analysis

The precision of seismic amplitude measurements at stations is critical for estimating seismic wave spatial gradients used for inferring seismic propagation, strains, stresses, and rotations. Recorded seismic wave amplitudes (measured in counts) depend on the instrument transfer function, the orientation of the instrument, the physical characteristics of installation, the structure of the Earth under the site, the vicinity to the seismic waves source. For these reasons, the recorded seismic waves at different stations, could have different amplitudes ranges even if recorded at close stations of a dense array. Nevertheless, they share some common ranges in which signals can be recorded simultaneously (Clinton and Heaton, 2002) (e.g., signals of the same earthquake). In these common ranges, if we consider the same kind of sensor (e.g., broadband, or short period), nearly identical seismograms are expected after the removal of instrument response. Instrument and metadata errors in one of the sensors or eventual amplification/de-amplification site effects will produce inconsistent waveforms between the same kind of sensor. The Mefite array was made by both, broadband and short-period seismometers. These instruments have different dynamic ranges and are sensitive to different frequencies, so that we developed the amplitude calibration analysis separately for the two types of sensors (Figure 4.3, Figure 4.4, respectively). During the installation period, the array recorded two regional earthquakes: (a) Mw 4.7 Albania earthquake (origin time (UTC): 2019-11-01 05:26:50, source: European Mediterranean Seismological Center - EMSC) (Figure 4.2); (b) m_b 4.7 Bosnia-Herzegovina earthquake (origin time (UTC): 2019-11-02 14:12:30, source: European Mediterranean Seismological Center - EMSC). These were recorded almost simultaneously from all the station of the network thus they were used for the network calibration and orientation analysis. For stations amplitude calibration, the Albania earthquake was analyzed.
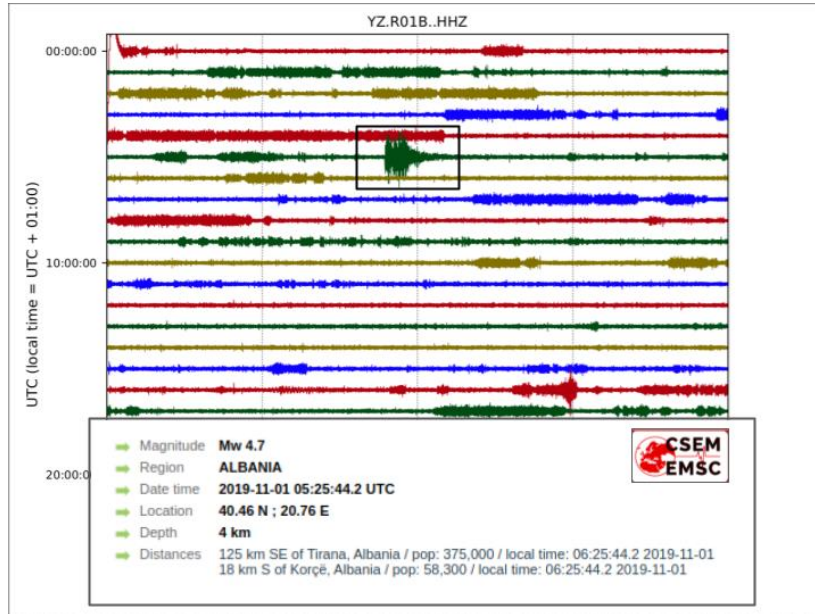
Figure 4. 2 - Albania earthquake recorded from the station R01B of the Mefite array (origin time: 2019-11-01 05:26:50 UTC) (source EMSC)

For all the stations, the recorded velocity displacement amplitudes (m/s), corrected for the instrument response, were analyzed in the same 5 minutes continuous record (05:25:00 – 05:30:00 UTC) around the earthquake waveform. The traces were tapered with a Hann window function and then bandpass filtered in a sufficiently narrow and low-frequency range – 0.08-0.2 Hz for broadband stations and 0.2-2 Hz for short period stations – to exclude any local spectral content effect at each station. The results for broadband stations and short period sensors are shown in Figure 4.3 and Figure 4.4, respectively.The good calibration of the broadband stations is testified by the very similar amplitudes recorded in either for the horizontal (E and N) or for the vertical (Z) components. Most of the short period sensors showed high amplitude coherency for all the sensors components. The only exception was represented by a de-amplification effect recognized at the horizontal channels of station R05S (N-S component) and, particularly, for station R06S (E-W component). To better evaluate the anomalous amplitudes at these stations, we compared the absolute amplitudes recorded at R05S and R06S with a stable station (R02S), representative for all the other short-period stations of the array. The absolute amplitudes were calculated in 30 s time window within the previous one (05:26:50 – 05:27:20 UTC), applying the same bandpass filter, as the vectorial compositions of the recorded amplitude at each sensor component (A_E, A_N, A_Z) as:

$$A_{abs} = \sqrt{A_E{}^2 + A_N{}^2 + A_z{}^2}$$

Eq. 4. 1

Both the component and absolute amplitude analysis showed that, while the recorded amplitudes or R05S and R02S are comparable with the others, the E-component of R06S station showed a de-amplification (Figure 4.5). The latter can be interpreted as due to a local site effect or to a station orientation issue.
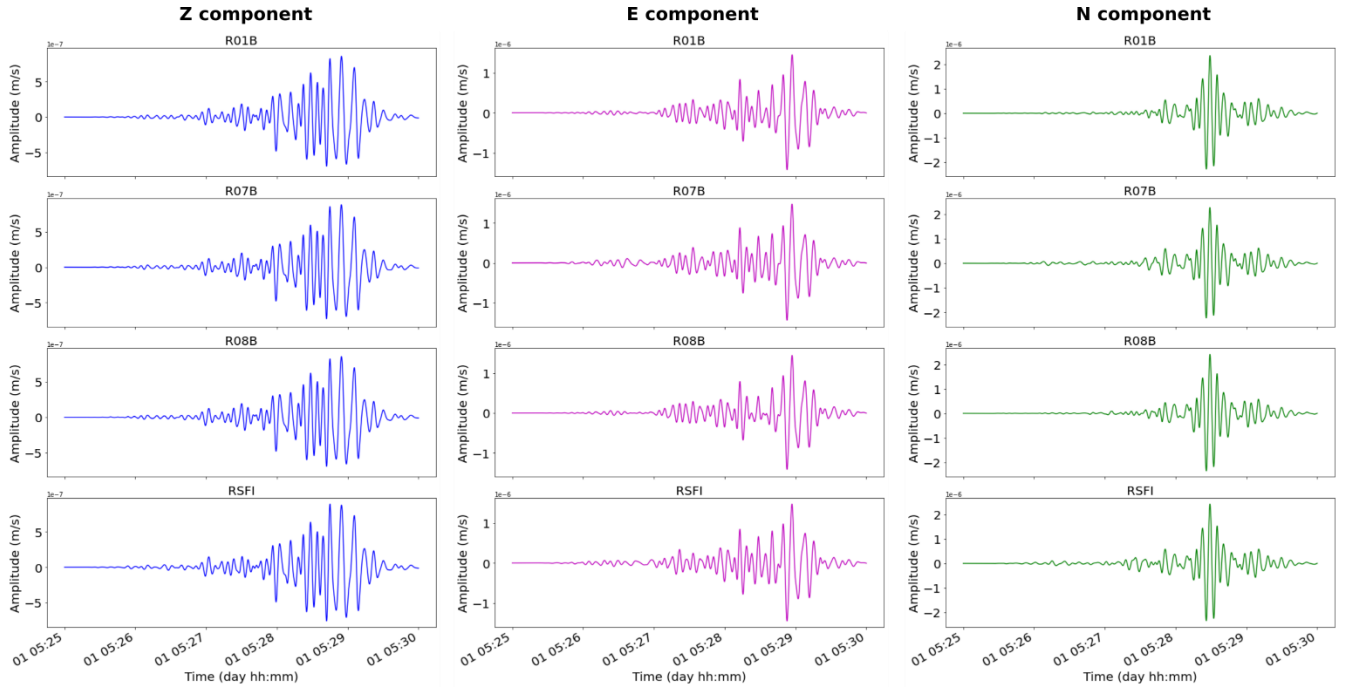
Figure 4. 3 – Amplitude comparison of 5 minutes continuous records (05:25:00 – 05:30:00 UTC) of the Albania earthquake at broadband stations. The signals are bandpass filtered in the range 0.08-0.2 Hz.
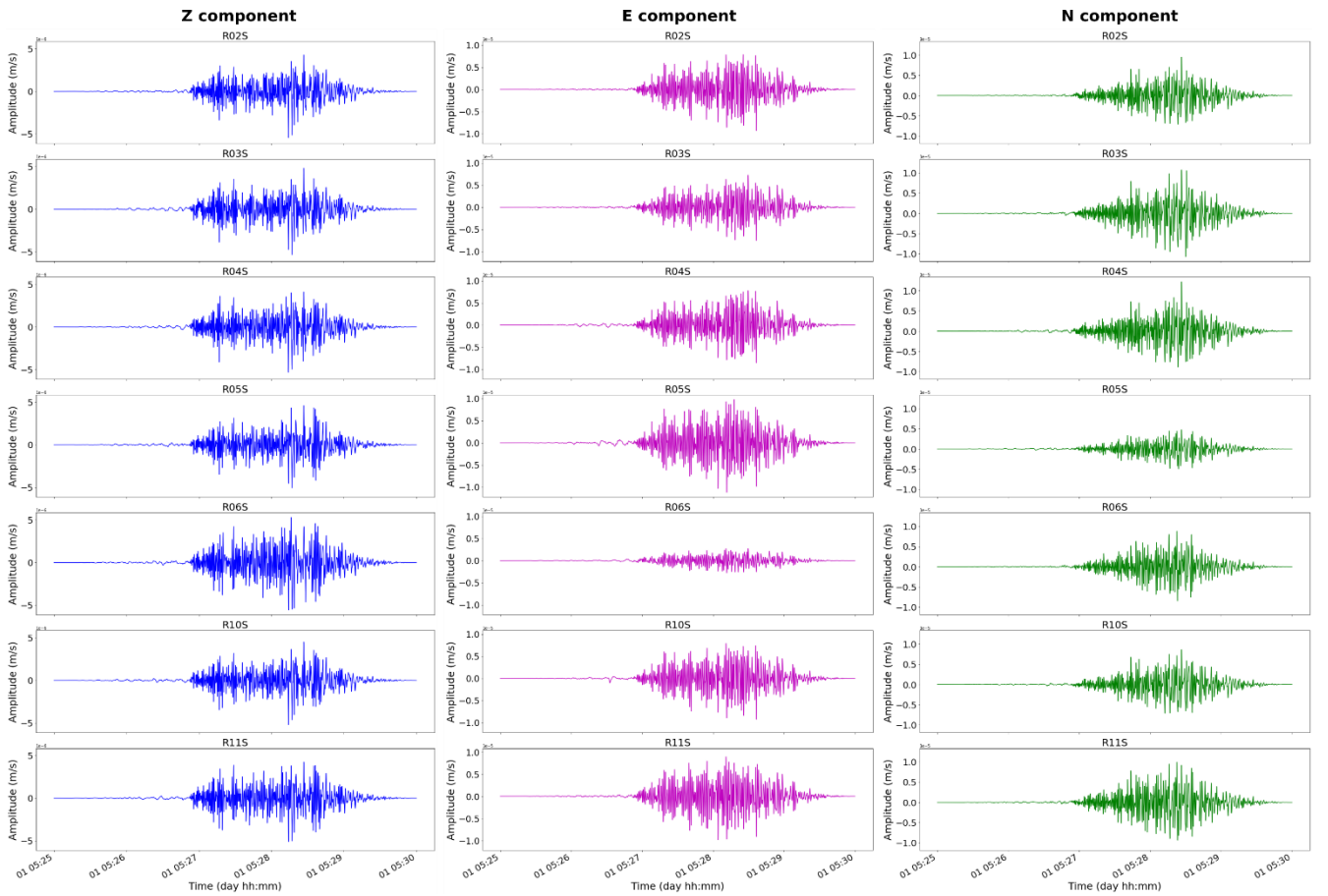
Figure 4. 3 - Amplitude comparison of 5 minutes continuous records (05:25:00 – 05:30:00 UTC) of the Albania earthquake at short period stations. The signals are bandpass filtered in the range 0.2-2 Hz.

121
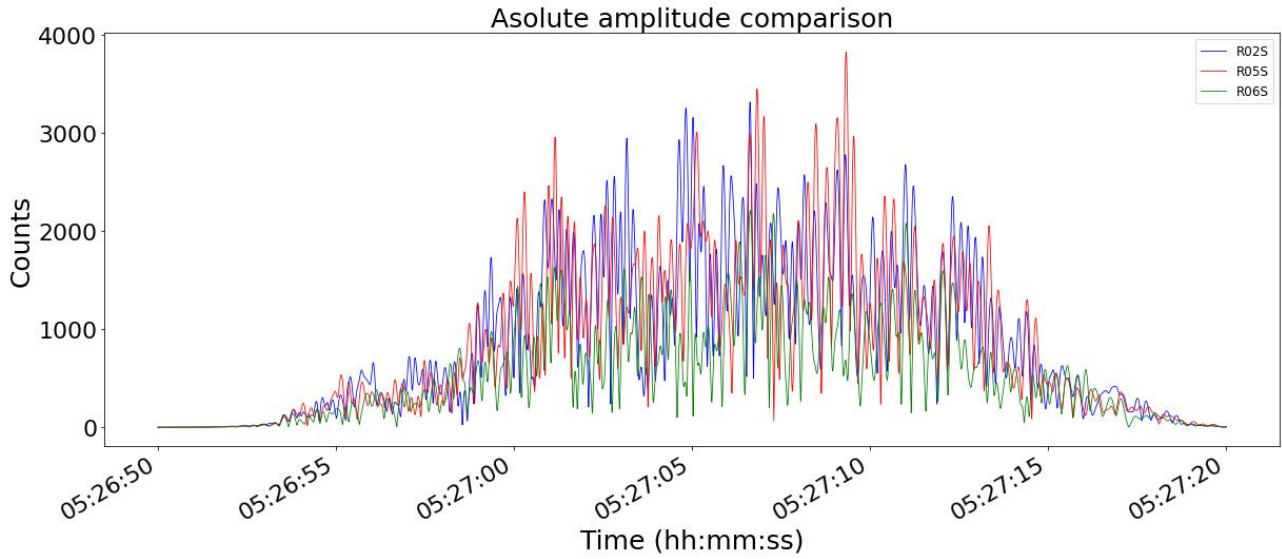
Figure 4. 4 – Absolute amplitudes comparison between the stable station (R02S) and the ones which showed de-amplification effects (R05S, R06S)

4.3.2 Network orientation analysis

Following the previous observation on stations amplitudes, the array stations orientation analysis has been carried out testing two different approaches: (a) polarization analysis of the recorded waves at different stations of the array; (b) signals cross-correlation. Both methods were applied on the wavefield related to the regional $m_b = 4.7$ Bosnia-Herzegovina earthquake, recorded from the Mefite array on 2 November 2019 (origin time 14:12:30 UTC).

The polarization analysis was performed on 3-component, bandpass filtered (0.1 - 0.5 Hz) signals applying the Particle motion method (Jurkevics, 1988). For each station the mean polarization azimuth was estimated for the selected time window. The method was carried out on: (1) 20 s long P-waves time window (14:12:34 - 14:12:54); (2) 56 s long window (4:13:47 - 14:14:15) containing both S and Love seismic phases. Then, the calculated signal azimuths were compared with the Earthquake waves propagation direction, calculated as the clockwise angle from the N of the linear direction joining the estimated event source coordinates (44.32 N, 17.61 E, source: EMSC) with the array centroid (Figure 4.6). We assumed the condition of plane wave approximation which is satisfied if the distance (d) between sensors is much less than the dominant wavelength (λ) of the recorded signal ($d \ll \lambda$). The observed high coherency of the array stations azimuths, for both the P, S and Love phases, with respect to the earthquake direction, confirmed the good orientation alignment of the network stations. Again, R05S and R06S stations represent an exception for the analysis being slightly out of orientation. The result suggest that the polarization analysis is influenced by the recorded amplitudes and from the directionality of amplification site effects which were observed in the previous amplitude calibrations analysis.
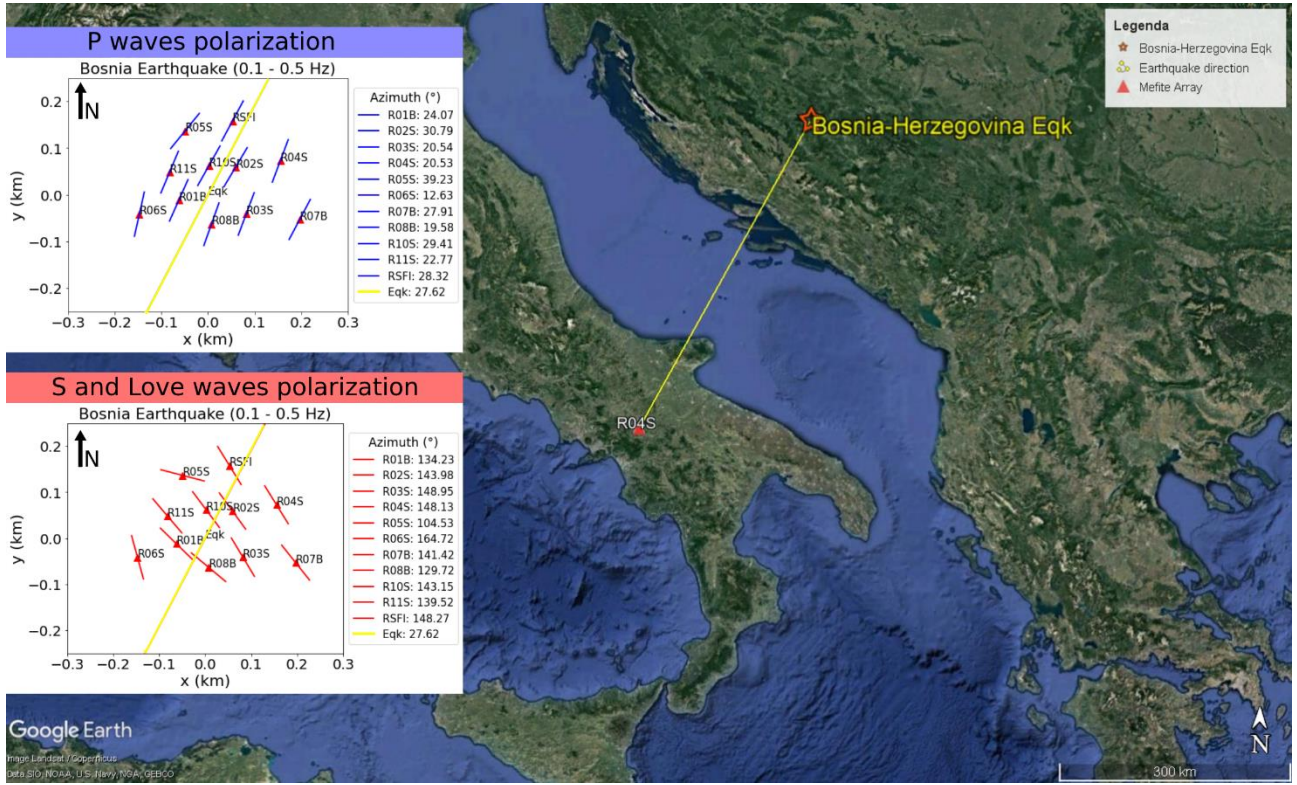
Figure 4. 5 - Polarization analysis for station alignment assessment under the condition of condition of plane wave approximation (d <<< λ). The mean station azimuth were compared with the earthquake plane direction with respect to the Mefite array.

Consequently, a different approach was tested, to obtain array station alignment errors uncorrelated to eventual site effects on the recorded amplitudes. The estimates of station azimuthal orientations with respect to a reference station were obtained by applying a methodology based on the maximization of the cross-correlation among the horizontal traces of adjacent sensor pairs (Stabile et al., 2020; Zeng and McMechan, 2006). Even in this case, for each pair of adjacent sensors we assumed the condition of plane wave approximation (d << λ). For each angle θ ranging from 0 to 360° with a step size of 0.5°, we computed the normalized cross-correlation between the north component of the signal recorded by the reference station (SrN) and the first horizontal component of the signal recorded by sensor with unknown orientation and rotated anticlockwise by the angle θ (Su_1^{θ}). In addition, we computed the normalized cross-correlation between the east component of the signal recorded by the reference station (SrE) and the second horizontal component of the signal recorded by sensor with unknown orientation and rotated anticlockwise by the angle θ (Su_2^{θ}). For each angle θ, the maximum values of the cross-correlation between SrN and Su_1^{θ} (A^{θ}) and between SrE and Su_2^{θ} (A^{θ}) were retrieved. Then, the sensor orientation with respect to the reference sensor was given by the following equation:

$$\theta^{BEST} = \theta : \max_{0° \leq \theta \leq 360°}(A^{\theta} B^{\theta})$$

Eq. 4. 2

where θ^{BEST} is the angle for which the product between A^{θ} and B^{θ} was maximum. R01B was chosen as reference station. By applying Eq. (4.1) over the Bosnia earthquake recordings, we obtained the estimates of θ^{BEST} for each station of the array (Table 4.2). The analysis highlighted high signals cross-correlation for all

stations, with a cross-correlation coefficient (CC) greater than 0.92. Moreover, it confirmed that all the stations of the array are well aligned with a maximum rotation error from the reference station of 5.2°.

Station	θ^{BEST}	CC
R02S	5.0	0.96
R03S	1.2	0.98
R04S	2.0	0.93
R05S	3.2	0.95
R06S	2.6	0.97
R07B	2.0	0.96
R08B	4.2	0.99
R10S	5.2	0.97
R11S	1.8	0.99
RSFI	4.2	0.93

Table 4. 2 – Results of the cross-correlation method for the station alignment analysis.

Station	θ^{BEST}	PM error (P wave)
R02S	5.0	6.7
R03S	1.2	3.5
R04S	2.0	3.5
R05S	3.2	15.2
R06S	2.6	11.4
R07B	2.0	3.8
R08B	4.2	4.5
R10S	5.2	5.3
R11S	1.8	1.3
RSFI	4.2	4.3

Table 4. 3 – Comparison of station orientation angle with respect to the reference station R01B calculated with the cross-correlation method (θ^{BEST}) and the Particle Motion method.

Table 4.3 shows the comparison between the results obtained from the application of the above-described methods. The particle motion azimuth error (PM error) - defined as the difference between the estimated P-wave azimuths with respect to R01B – and θ^{BEST} estimated from the cross-correlation analysis are consistent for most of the stations, thus indicating the good effectiveness of both methods in the study of network alignment analysis. Nevertheless, the not diverging values at R05S and R06S stations orientation confirm that, unlike the polarization analysis, the cross-correlation method is not influenced by the directional amplification due to site effects. We thus concluded that:

(a) The polarization method represents a good preliminary method to determine array stations orientation errors but is influenced by amplitude directionality related to site amplification effects (e.g., R05S and R06S stations). It can be useful to choose the best-oriented station for the subsequent application of the cross-correlation method.

(b) The signal cross-correlation method must be used to obtain more precise array station orientation errors and a more reliable estimation of the network stations alignment.

(c) The analysis confirmed that all the stations of the array are well aligned.

4.3.3 Horizontal-to-vertical spectral ratio analysis

With the purpose to investigate the supposed local ground effects at the network stations, the horizontal-to-vertical spectral ratio technique (HVSR; Nakamura, 1989) was addressed. The HVSR technique allows to retrieve information about the shallow-subsoil seismic properties by single-station measurements (Lunedei and Malischewsky, 2015) and assumes that the effects of the source in the recorded waveform are supposed to be minimized normalizing the horizontal ground-motion spectral component by the vertical one. From the experimental point of view, this technique requires a three-component ground-motion acquisition and is performed through the ratio between its horizontal and vertical Fourier spectra. This ratio, which is a function of the frequency, is called the H/V (or HVSR), is usually computed by using ground-motion velocity spectra and can be expressed as (Lunedei & Malischewsky, 2015):

$$HVSR\,(\omega)\,=\,\frac{A_{H,surface}^{F}\,(\omega)}{A_{V,surface}^{FW}\,(\omega)} \qquad \text{Eq. 4. 3}$$

where $A_{H,surface}^{FW}\,(\omega)$ and $A_{V,surface}^{FW}\,(\omega)$ are the total full-wavefield (FW) spectral amplitudes of the horizontal and vertical ground motion at Earth's surface, respectively, while ω is the angular frequency. In this paragraph, preliminary results are described, since a larger and more robust earthquake dataset would be needed for a more detailed characterization of the local ground fundamental frequencies. Indeed, the HVSR method was applied only on the two regional earthquakes (Albania and Bosnia-Herzegovina) recorded from the temporary seismic network, since no other regional or local seismic event was recorded in the installation period.

The analysis was performed through the Geopsy software (Wathelet et al., 2020) on 150 seconds time window for both the Albania (2019-11-01 time: 05:26:50- 05:29:20 UTC) and Bosnia-Erzegovina (2019-11-02 - 14:12:30- 14:50:00 UTC) earthquakes. Before computing the fast Fourier transform, the mean and the trend were removed from the time series and the analyzed 3-component signals were tapered by applying a Hann window with 5% bandwidth. Then, each frequency spectrum was computes and smoothed through the Konno–Ohmachi function (Konno and Ohmachi, 1998), with a smoothing coefficient (b-parameter) equal to 40. A 150 s time window was selected for the analysis which includes the whole earthquake waveform. Then, HVSR was retrieved separately calculating the spectral ratio of each horizontal amplitude spectral component over the vertical amplitude spectral component (EW/Z and NS/Z, respectively). The HVSR analysis (Figure 4.7) highlighted that the site effects are not uniform at all the stations of the array, thus suggesting possible local directionality effects (e.g., R05S station). Moreover, some H/V peaks for all the stations were identified at frequencies below 10 Hz (Figure 4.7), except for R06S station which shows a de-amplification effect at about 1 Hz. For the other stations, the first H/V main peak is recognizable at 0.7-0.8 Hz with higher amplitudes at the inner stations of the array (R10S, R02S, R03S); the second H/V peak is variable from station to station in the range 2-7 Hz and predominant at the external stations (R04S, R05S, R07B, R11S, RSFI). The two peaks would suggest the presence of two main local discontinuities in the emission area: a surficial and locally irregular one – to which the 2-7 Hz peaks would be related – and a deepest one, testified from the presence of the 0.7 Hz peak. Further analysis would be needed in the future for the geological-structural interpretation of

the observed H/V spectral ratios and for a better constrained H/V analysis possibly on a larger waveform dataset.

The results of the array amplitudes calibration, the station orientations, and the preliminary H/V analysis, displayed a major stability of amplitudes recorded at the vertical components with respect to the horizontal ones, which can be influenced from site amplification/de-amplification effects. In the next paragraphs, the developed automated tremor detection algorithm will be described. It is based on the measured tremor amplitudes at each station, such that it was applied only to the more stable vertical component of each station.
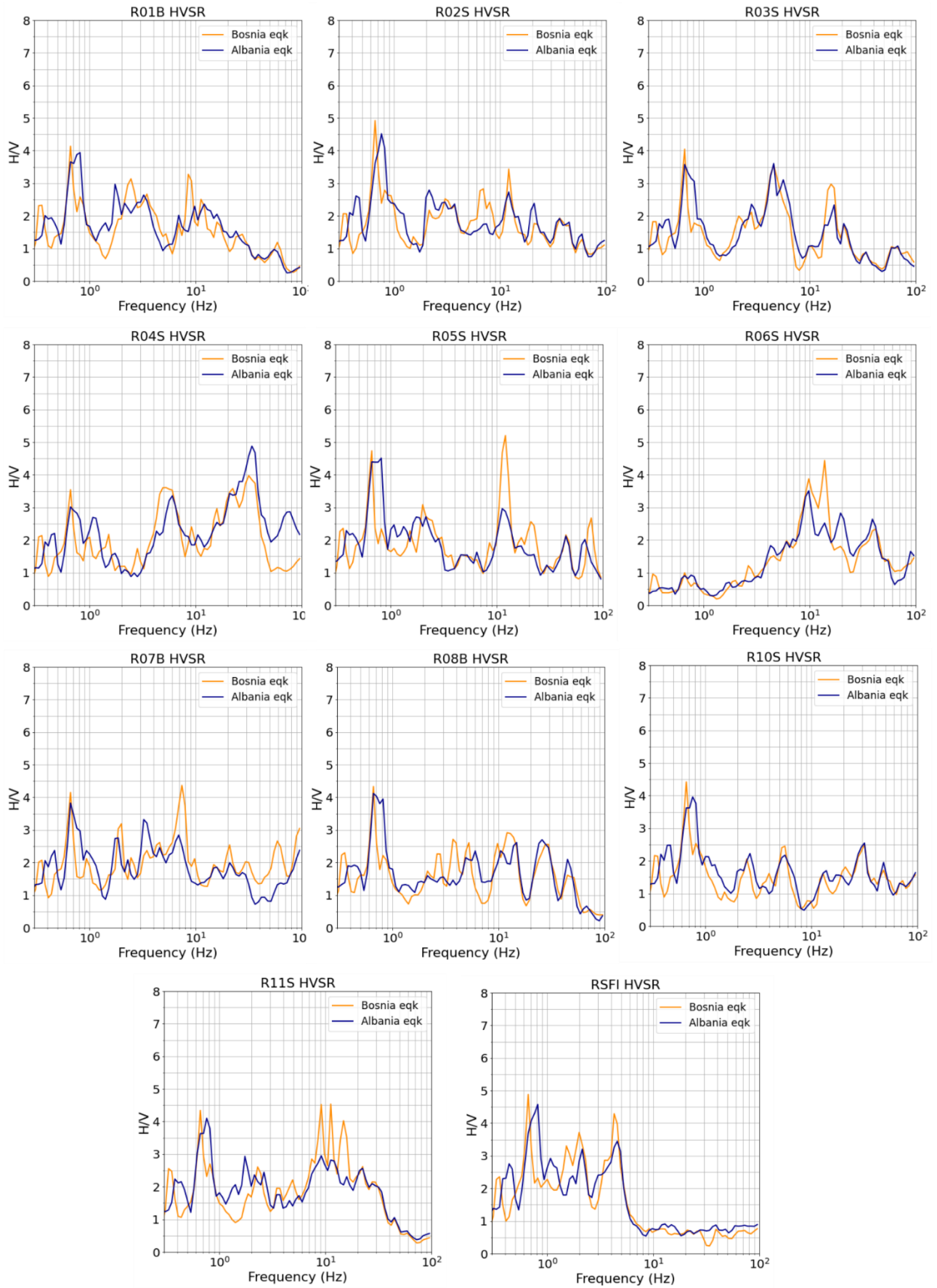
Figure 4. 6 – Comparison of HVSR analysis performed with the recorded regional earthquakes at each station of the array.

4.4 Tremor detection and classification: methods e results

4.4.1 Time-frequency observations

We analyzed continuous data recorded from both broadband (R01B, R07B, R08B and RSFI) and short-period (R02S, R03S, R04S, R05S, R06S, R10S, R11S) stations in the period between 2019-10-30 and 2019-11-02 during which the 11 stations worked jointly. The time-frequency observation of the background seismicity recorded at Mefite enabled us to observe a non-stationary signal characterized by the emergent onset of seismic tremors marked by (Figure 4.8): (a) higher amplitudes with respect to the recorded background noise at all the stations of the array; (b) variable duration (from tens of second to tens of minutes); (c) a prominent spectral peak in the frequency range between 6 and 15 Hz.
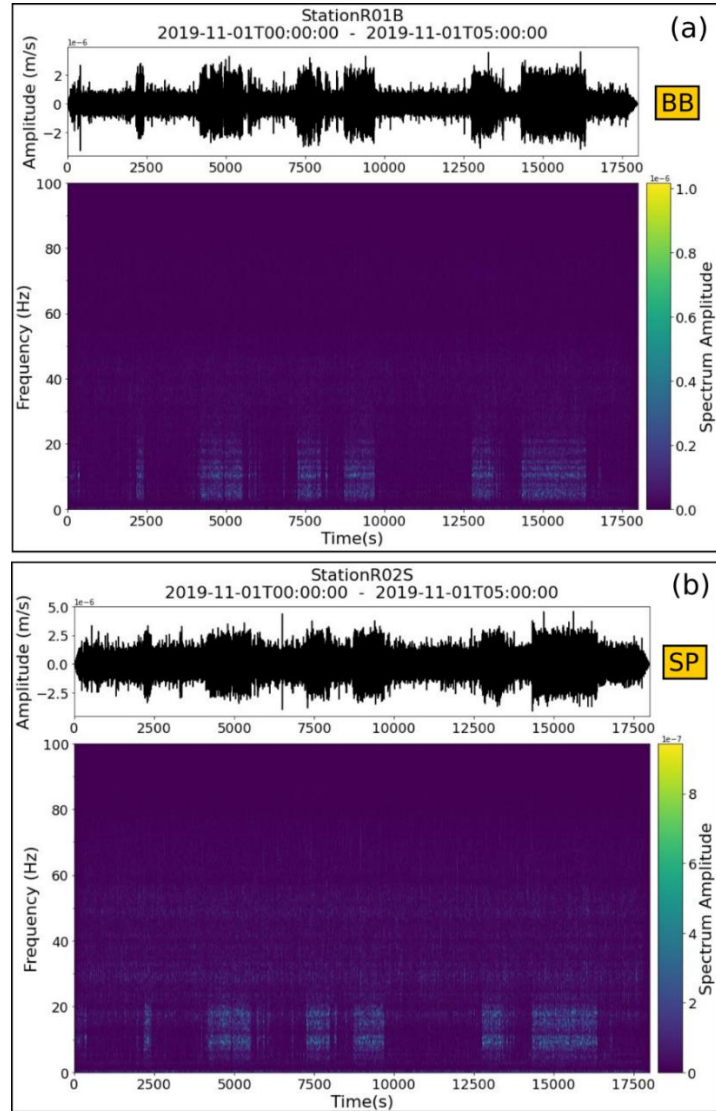
Figure 4. 7 - Five-hour long vertical component seismograms and their relative spectrograms recorded at the (a) R01B broadband and (b) R02S short-period stations.

4.4.2 Automated detection algorithm

Exploiting continuous recordings of the Mefite array we implemented an automated detection algorithm through Python-Obspy (Krischer et al., 2015) seismic signal processing tools (Figure 4.10, 4.11). The algorithm is based on non-parametric statistics of the recorded amplitudes at each station and was applied only to their vertical Z-components to avoid problems related to local site amplifications. Detection is performed through the following steps (Figure 4.10): (a) the trace is bandpass filtered in the observed peak-frequency range of 6 - 15 Hz to enhance tremors onset over the record; (b) the smoothed envelope of the signal is calculated by applying a centered moving average function of 3000 points (15 seconds windows); (c) tremor start- and end-times are measured as the times when the envelope goes above and below, respectively, a given threshold. Consecutive detected signals closer than 15 seconds are joined into a single detection, while events shorter than 15 s are excluded from detected signals. Amplitude thresholds for detection were estimated from a statistical analysis of the recorded amplitudes. For a more robust statistical choice of detection thresholds, we made no assumption on the amplitude's statistical distributions, analyzing it only in terms of percentiles. Considering a background noise mean amplitude value at the 16th and 84th percentiles –which would correspond to the 68.3% of the population of a normal distribution – we estimated the detection threshold amplitude as the absolute value of its 90th percentile. This enabled to estimate data-driven thresholds for each station and for each analyzed day (Figure 4.9). Both the threshold estimation and the tremor detection methods were applied on raw amplitudes (counts). Detection algorithm was applied to all the stations of the array from 2019-10-30 to 2019-11-02, collecting a total dataset of 8561 signals.
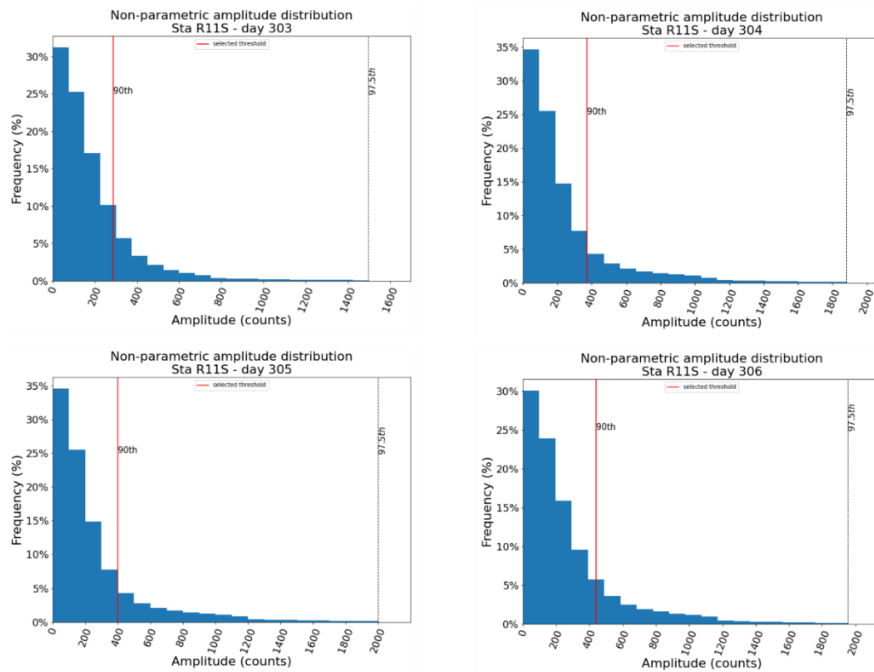
Figure 4. 8 – Non-parametric distribution of the absolute amplitude values (counts) recorded at station R11S from 30 October 2019 (303) to 02 November 2019 (306). The images show the distribution until the 97.5th percentile (vertical dashed line) and the selected detection thresholds at the 90th percentiles (vertical red line).

With the aim of characterizing the collected dataset, the developed detection algorithm calculates and stores in output, for each event, the following parameters: (1) Start time; (2) End time; (3) Duration (s); (4) RMS ground velocity amplitude (m/s) calculated from the instrument-corrected signal (m/s):

$$A_{RMS \ (\frac{m}{S})} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} Ai^2}$$

Eq. 4. 4

(5) Arrival slope, calculated as the mean derivative of the signal envelope 10 seconds before the detection trigger: the higher the arrival slope, the more impulsive the signal onset would be.
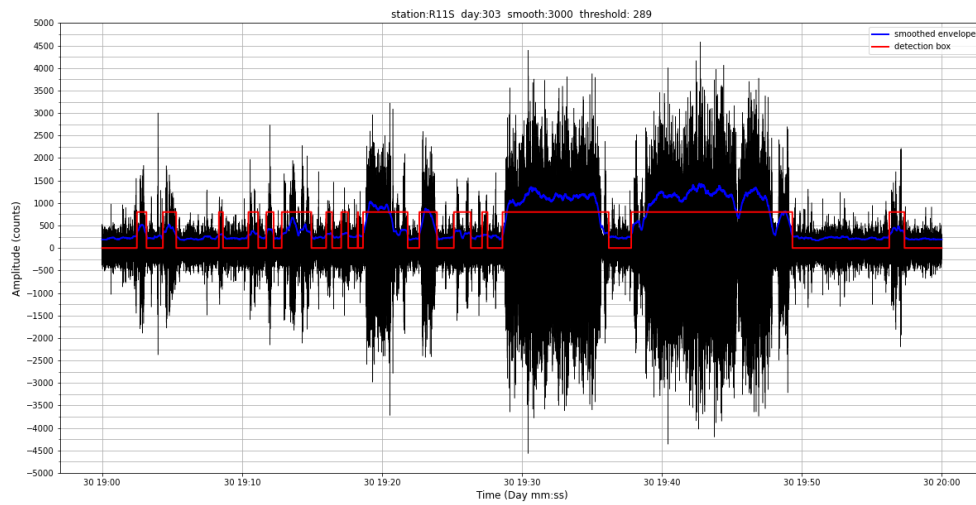
Figure 4. 9 - Detection algorithm workflow applied to a 1-hour signal (19:00:00 - 20:00:00 UTC time) recorded on 30 October 2019 (Julian day 303) at station R11S. First, the smoothed envelope of the 6-15 Hz filtered waveform (blue curve) is calculated; then, event detection is performed through a box function (red curve) which assumes non-zero values for envelope amplitudes greater than the threshold.
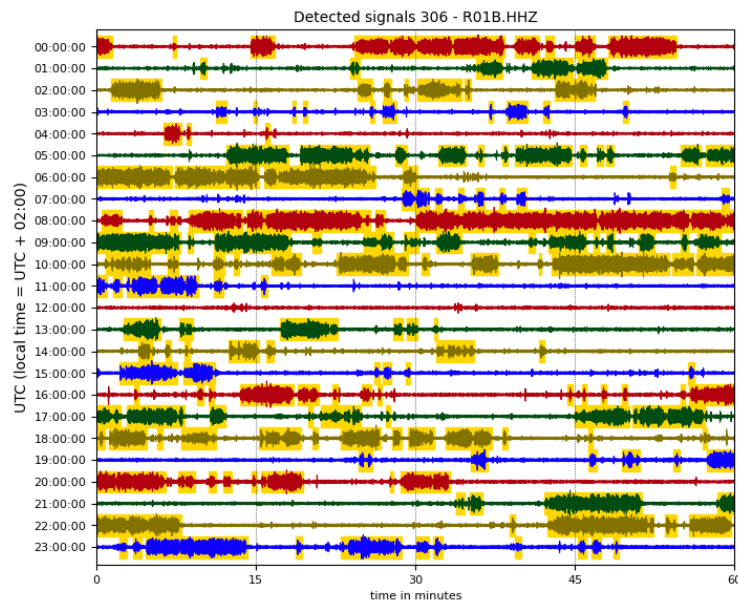
Figure 4. 10 - Helicorder display of the detection results at station R01B on 02 November 2019. Each yellow bar on the plot corresponds to the signal automatically detected from the developed algorithm.

4.4.3 Spectral analysis

The developed detection algorithm enabled us to automatically extract a large tremor dataset for each station of the array and to collect information regarding their time-frequency parameters (duration, RMS amplitude, arrival slope, frequency content, polarization). The spectral analysis of the detected signals at each station of the array evidenced two main characteristics of their frequency content (Figure 4.12): (1) amplitude spectra are characterized by three to four peaks in the range between 7-20 Hz and (2) these peaks vary from station to station, with different central frequency and/or different relative amplitude. Moreover, comparing amplitude spectra of tremor events and background noise (Figure 4.12) we observed the same peaks pattern, which eventually can be amplified at the occurrence of a tremor event. This would suggest that tremors do not have a specific frequency characteristic with respect to noise; there is only a higher spectral amplification in the frequency range between 6 and 15 Hz without a significative change of spectral shape (Figure 4.12). Thus, the signal frequency content has been excluded from the machine learning analysis at this stage.
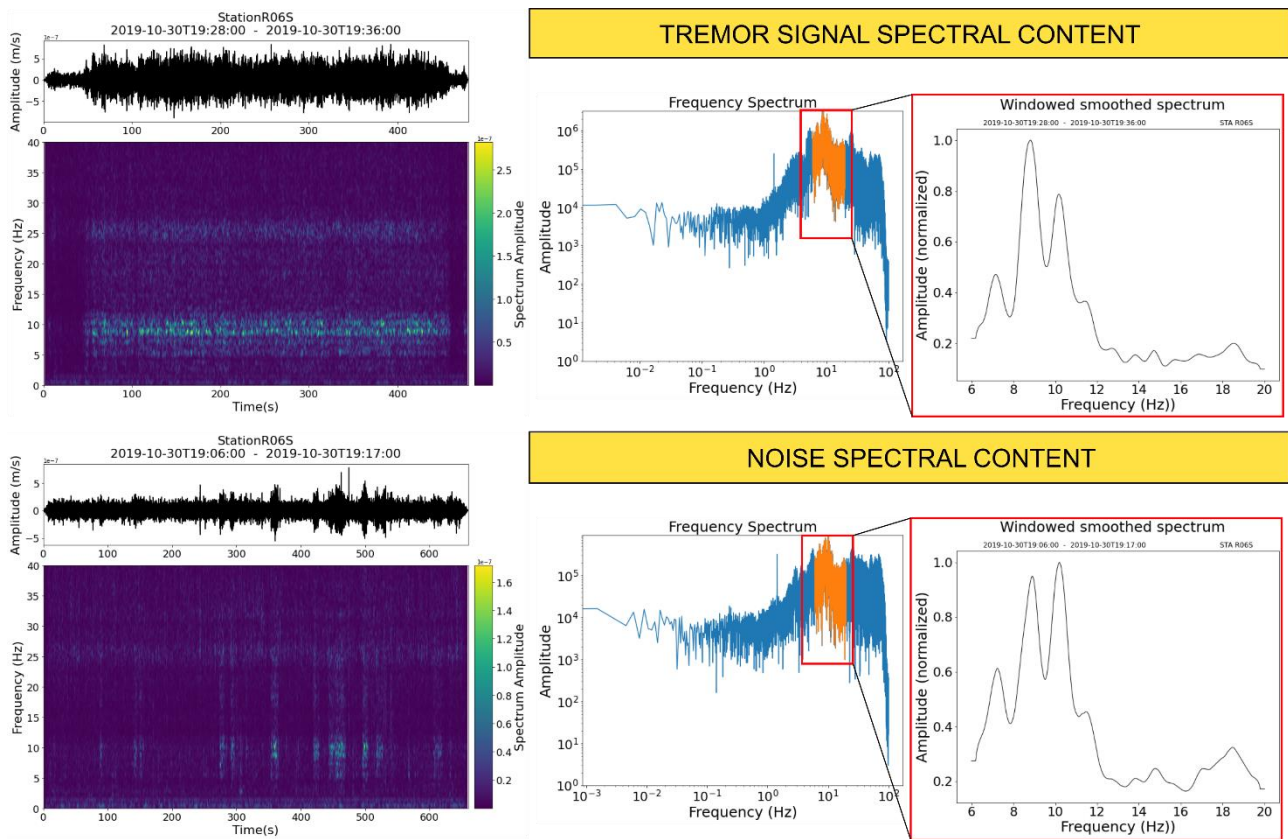
Figure 4. 11 – Comparison of spectral content comparison between 11 minutes-long noise and a tremor event recorded at station R06S the same hour. For both signals the image shows the spectrogram, the frequency spectrum and its windowed and smoothed portion (6-20) which highlights the main spectral peaks.

4.4.3 Tremor signals automated machine-learning classification

With the aim of characterizing typical tremor parameters distributions and classifying non-volcanic tremor signals at each station – discriminating between tremors and accidentally detected anthropogenic noise –, we applied an automated machine-learning approach. In this section, we show the results of both unsupervised DBSCAN algorithm (Density-Based Spatial Clustering of Applications with Noise – (Ester et al., 1996; Schubert et al., 2017) and supervised KNN classification (k-nearest neighbors –Fix & Hodges, 1951, Cover & Hart, 1967). The algorithms were implemented through Python-scikit-learn machine learning module (Pedregosa et al., 2011) and applied to time-domain tremor parameters (duration, RMS-amplitude and arrival slope). DBSCAN was first used to determine bivariate correlations among the analyzed parameters and let us define training samples for the subsequent application of the supervised KNN classification.

4.4.3.1 Signal clustering

The first step in the automated tremor signal characterization was to apply a cluster analysis at each station. Clustering is a powerful machine-learning tool to explore dataset structures and was applied to standardized values of tremor duration, RMS amplitude and arrival slope to avoid misleading clustering results related to different parameters scale. Since no previous information about tremor parameters distribution was available, the DBSCAN clustering algorithm was applied to avoid any a-priori assumptions on the number of clusters we were looking for, considering only the density of the features (tremor parameters) in the observation space (unsupervised model). DBSCAN is a density-based data clustering algorithm which finds an optimal number of clusters in the dataset starting from the estimated density distribution of corresponding nodes. The algorithm requires two hyperparameters - whose value is used to control the learning process- to define a "density-based" cluster: (1) the distance within which search for close points (ε); (2) the minimum number of samples to create a cluster (n). We applied a grid search to different hyperparameters combinations to obtain the one which maximize the clustering performance for each station dataset, calculating the following metrics: (a) number of clusters found; (b) mean noise points distance from the closest 6 points (6-N); (c) number of identified noise points (outliers) (Figure 4.13). The method allowed us to choose for each station cluster analysis the values for ε and n which jointly minimize the number of outliers and maximize the 6-N distance for the observed signals. For every station dataset the cluster analysis identified a single cluster and its outliers (noise points) (Appendix 4.2). The identified cluster mainly included impulsive events characterized by sequences of short tremors of few seconds (Figure 4.14a). The recognized impulsive signals showed a typical shorter duration (median < 100 s) and smaller amplitudes with respect to the identified outliers. The latter included mainly two kinds of signals: high-amplitudes and longer (>100 s) tremor signals and anthropic spikes (Figure 4.15), thus indicating a low effectiveness of the unsupervised algorithm for the automated discrimination between tremor and anthropic signals. Conversely, the cluster analysis allowed the recognition of characteristic bivariate correlations among the parameters of the detected signals (Figure 4.14b): (a) a good linear correlation (Pearson's r ≈ 0.6 - 0.7) between signal duration and amplitude; (b) a lower linear correlation (Pearson's r ≈

0.4 - 0.6) between amplitude and arrival slope; (c) no correlation between signal duration and their impulsive onset (arrival slope).
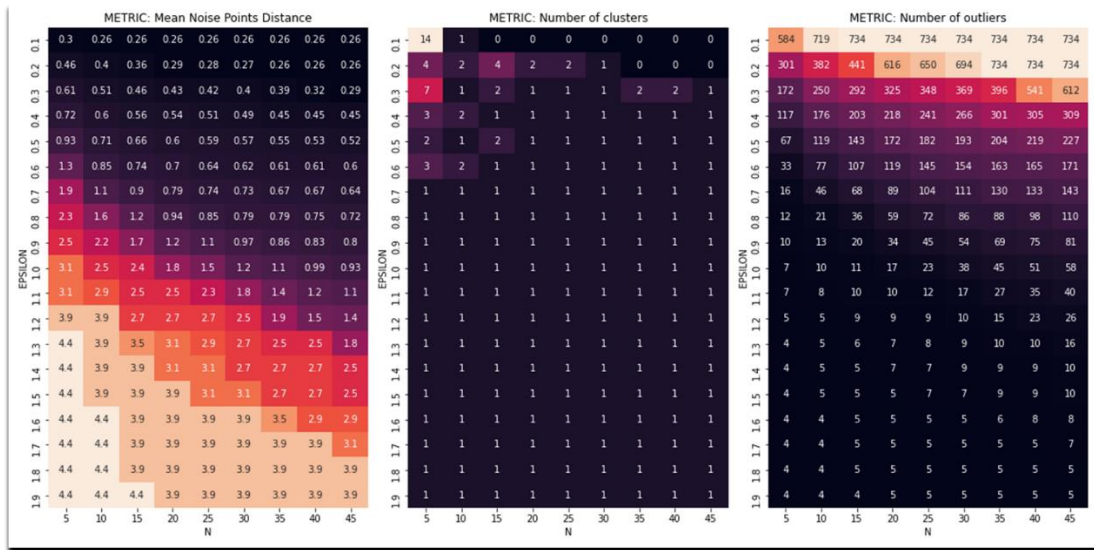
Figure 4. 12 –The developed grid search metrics for the best hyperparameters combination choice.
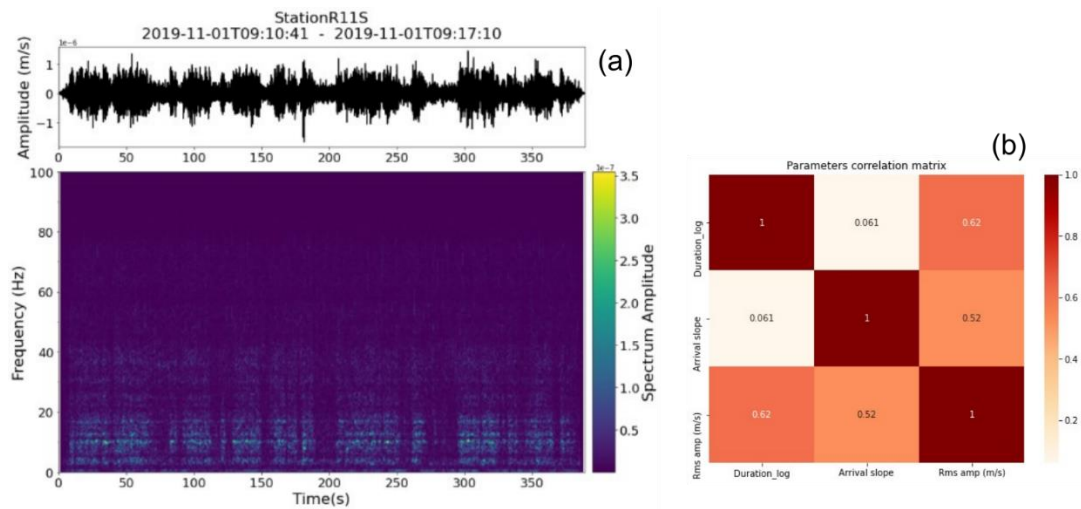
Figure 4. 13 - Seismogram and spectrogram of impulsive tremor (a) and cluster parameters linear correlation matrix (b) obtained through DBSCAN cluster analysis
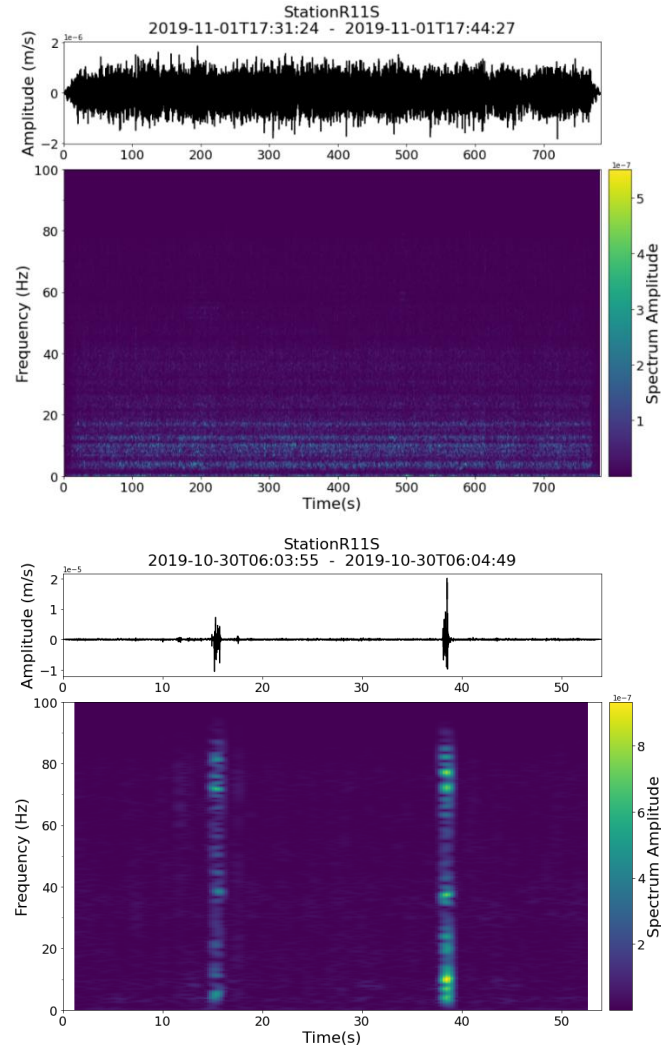
Figure 4. 14 - Seismogram and spectrogram of DBSCAN clustering outliers: (a) long-high amplitude recorded tremors; (b) anthropic spikes.

4.4.3.2 Signal classification

To overcome the low effectiveness of the DBSCAN clustering in the automated discrimination between tremor and anthropic signals and to better characterize the recorded tremor at each station, we decided to classify the collected datasets of tremors through the K-Nearest Neighbor (KNN) Classification algorithm. As a supervised machine learning algorithm, KNN relies on labeled input data (training data) to learn a function (model) that predicts an appropriate classification of the input unlabeled data. KNN is a widely used classifier thanks to its non-parametric nature and its efficiency in classification problems where decision boundaries are irregular, as in the case of the long-high amplitude tremors and the station-to-station different anthropic signals (jointly classified as noise-points in the DBSCAN clustering analysis). KNN classification is based on features similarity, measured as the distance from a number K of labelled points (Nearest Neighbors). We chose the Euclidean distance as similarity parameter. The number K is the core deciding factor (hyperparameter) of the algorithm: each data point will get the most recurring label amongst its closest K points. With the aim of evaluating the best K-value for the classification analysis, we implemented a model accuracy test: for any given training dataset, we generated different models for different K-values knowing that it must be odd

134

numbers for an even number of classes (Kim et al., 2012) (here: tremor signal, non-tremor signal) and selected the K-value which would produce the maximum classification model accuracy. It is defined as the number of correct predictions over the total number of predictions and was estimated as the Jaccard similarity index. For each station, a training data file was created with two classes of data samples selected by both visual inspection and time-frequency analysis. Tremor signals were assigned to the class 0, while anthropic spikes were labelled as class 1. Then, the KNN Classifier was trained through the sample signals parameters. Based on the previously observed bivariate correlation between tremor parameters, a first classification through the two uncorrelated parameters (Duration and Arrival slope) of the training dataset was attempted. However, adding signal amplitudes improved the discrimination of high-amplitude anthropic spikes so that a 3 parameters classification was addressed. The automated classification enabled to discriminate the accidentally detected anthropic signals from the analyzed seismic tremor and to confirm typical tremor parameters (duration, amplitude, arrival slope) (Appendix 4.1). The classified tremor signals evidenced an even stronger positive correlation between the logarithm of tremor duration and amplitude (Pearson's r of 0.7 - 0.8) then the one observed in the clustering analysis, while the previously evidenced correlation between amplitude and arrival slope was confirmed. Tremor duration can be highly variable (from 20 seconds to 30-40 minutes) as testified by the high skewness of its distribution at all stations (Figure 4.19c and 4.16) but has also a consistent median value < 100 s (in the range of 52 - 72 seconds). The collected RMS amplitudes are generally of the order of 10^{-7} m/s and show mainly symmetric distributions (Figure 4.19a). Particularly, the highest amplitudes of tremors have been recorded at station R10S (median value 6.24×10^{-7} m/s) while the lowest at station R06S (median value of 8.81×10^{-8} m/s). Amplitudes time series (Figure 4.17) show a good correlation for all stations except for slightly different trends in R05S, R06S, R11S, RSFI due to few outliers (Figure 4.19a) not well discriminated from the automated classification. Moreover, the RMS-amplitude time series display a slight double trend with smaller amplitudes recorded during the first day (30 October 2019) which increased during the following analyzed days. The same trend, even more pronounced, can be observed in the arrival slope time series (Figure 4.18) confirming that it is an indicator of the signals first arrival. This is also evinced by looking at their box plots (Figure 4.19a, b) which show similar trends at all the stations except for RSFI. The latter is the noisier station of the array, close to the main road and installed in an inhabited area so that can comprise anthropic noise with similar characteristics to the tremor which have not been well discriminated by the KNN classification.
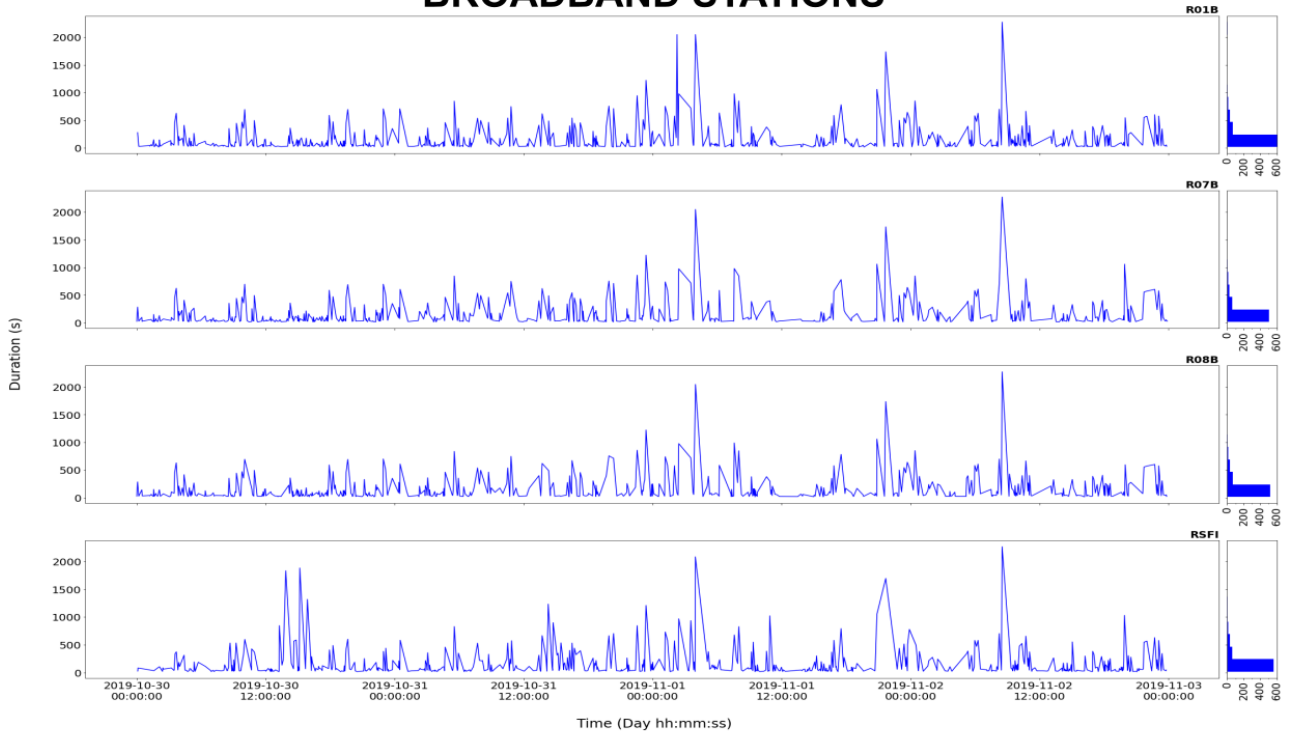
Figure 4. 15 – Duration time-series of the automatedly classified tremors signals at both broadband and short-period stations

BROADBAND STATIONS
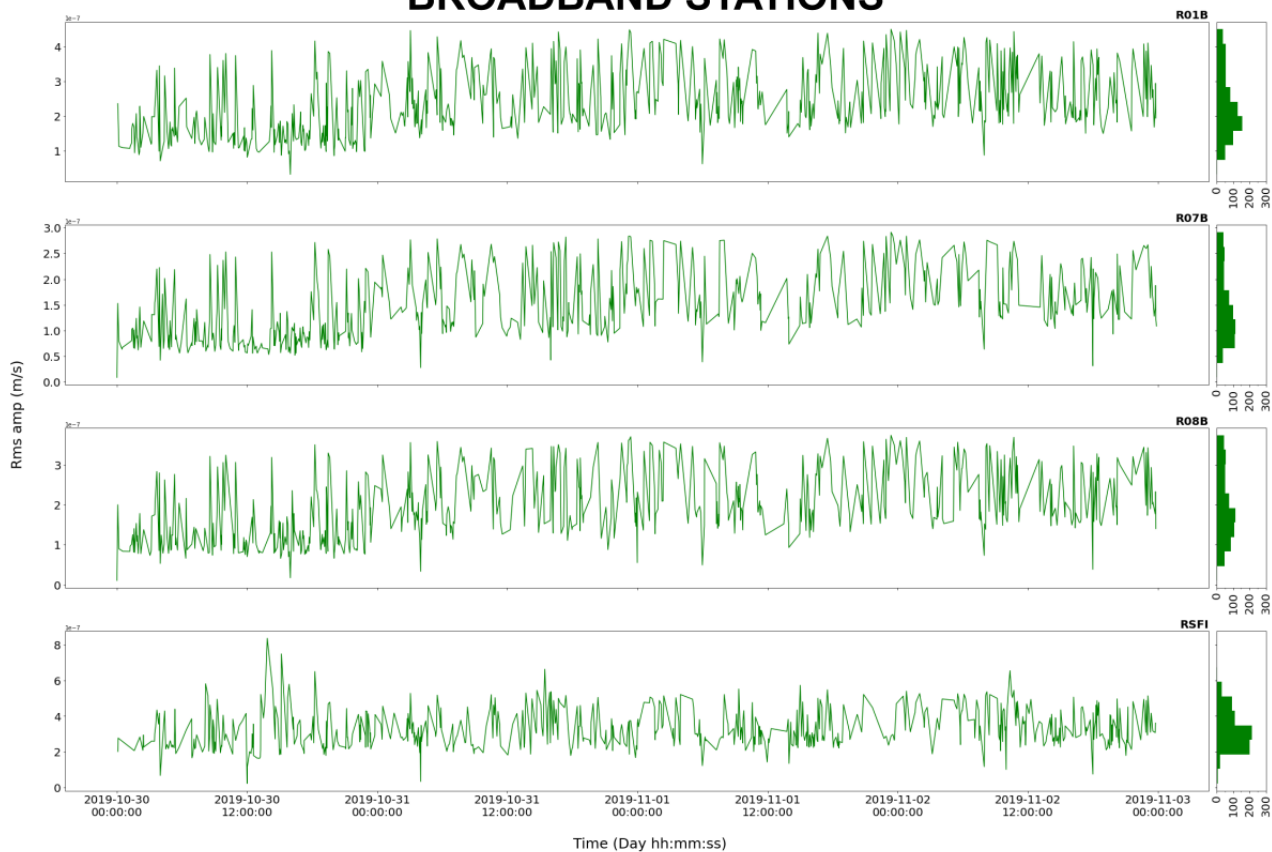
SHORT PERIOD STATIONS
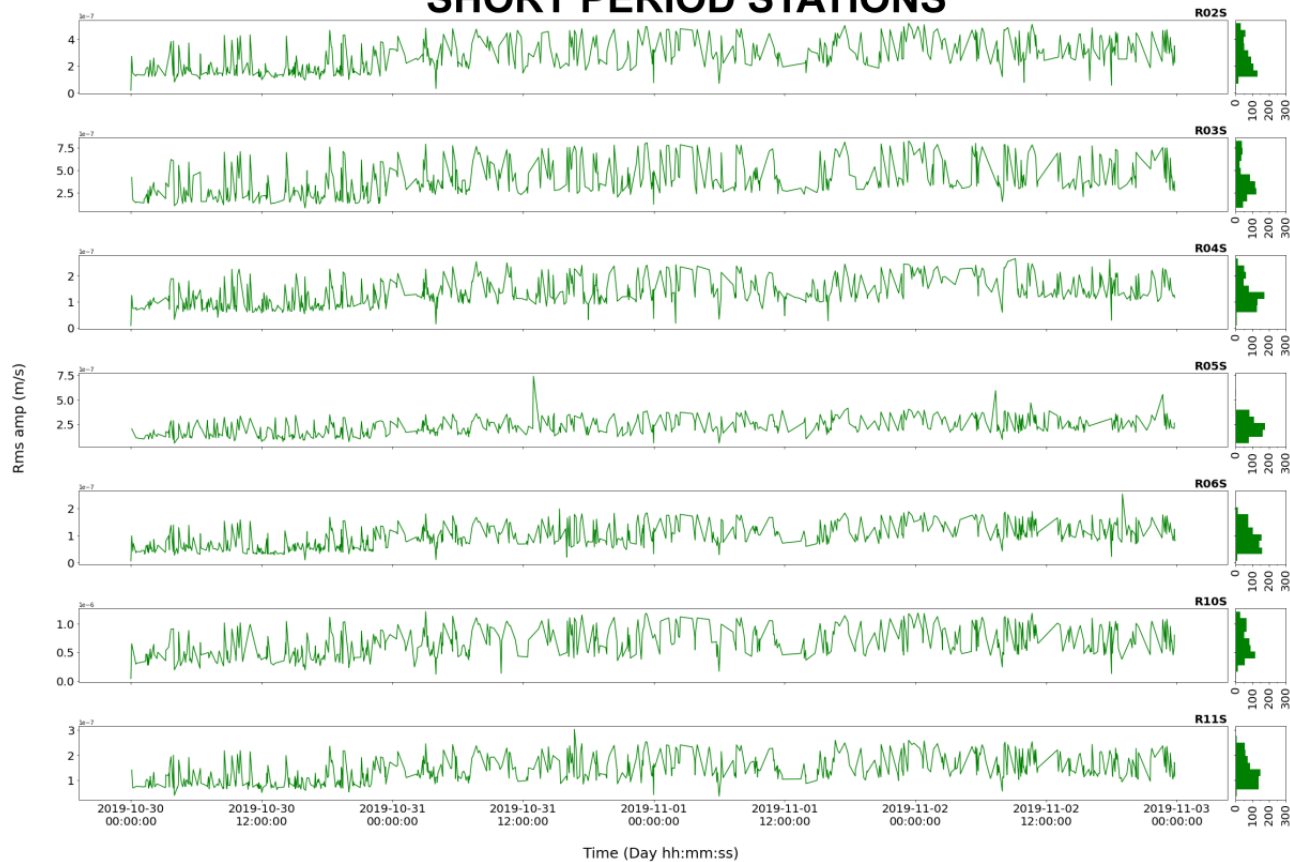
Figure 4. 16 – RMS-amplitude time-series of the automatedly classified tremors signals at both broadband and short-period stations.
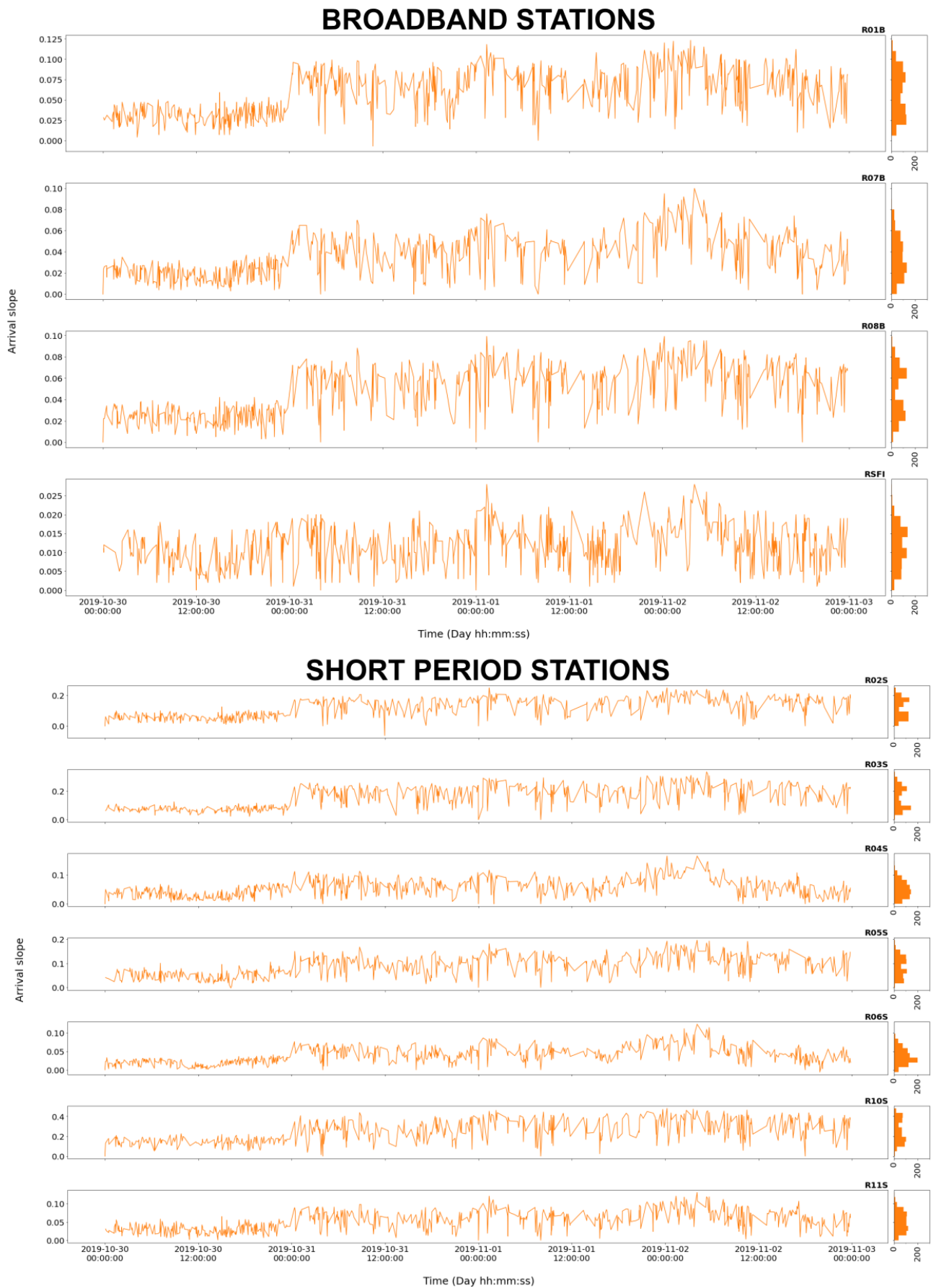
Figure 4. 17 – Arrival slope time-series of the automatedly classified tremors signals at both broadband and short-period stations.
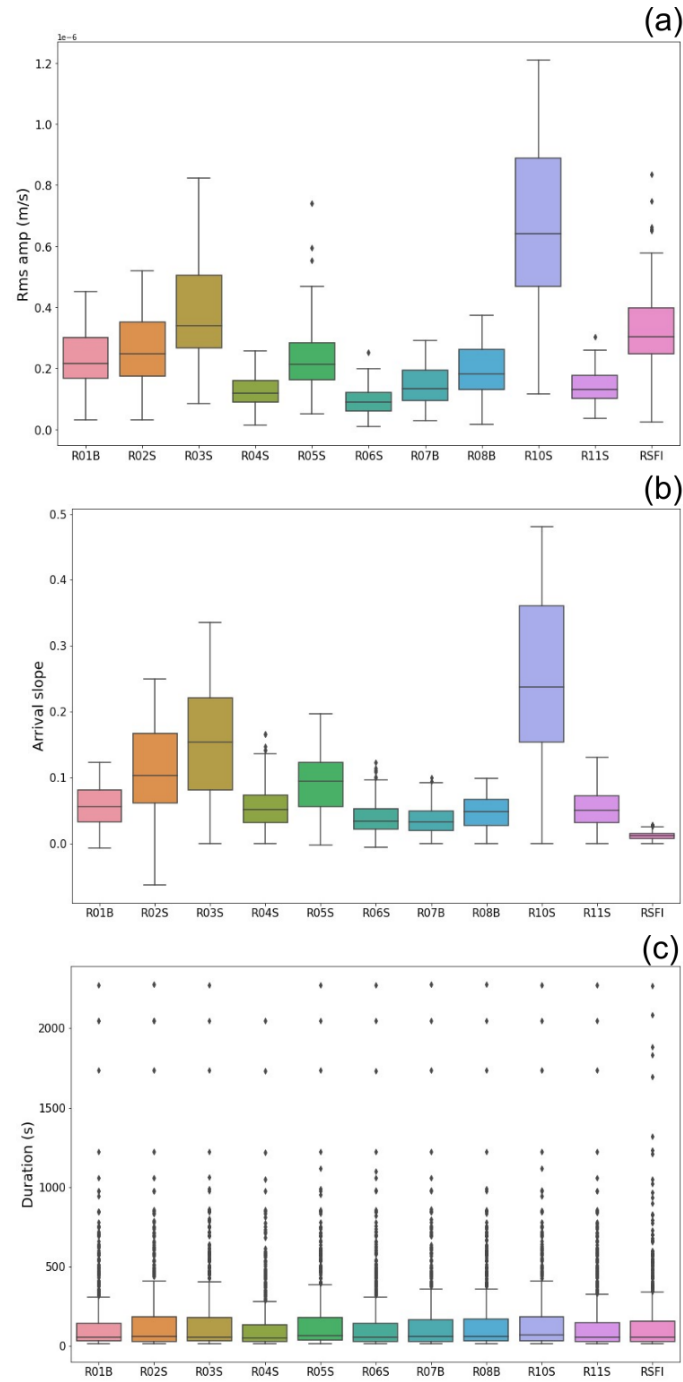
Figure 4. 18 - Box Plot representation of the automatedly classified tremor parameters distributions at each station of the array: (a) RMS-amplitude (m/s), (b) arrival slope, (c) duration (s).

4.4.3.2 Polarization analysis

To test whether the measured arrival times were related to seismic waves propagating from a common source, we developed a signal binding algorithm. It analyzes events start times (detection triggers) at all the stations of the network and associates them to the same tremor based on two conditions: (1) given a constant, sliding time window of 10 seconds, it gathers only signals which start times results within it; (2) if the resulting number of associated stations is larger than 5, an event is declared. A total of 673 events were selected. The polarization analysis of the selected events was carried out at all stations of the array by applying the particle motion method (Jurkevics,1988) on three-component, 6-15 Hz bandpass filtered signals. For each tremor there was estimated

139

the average azimuth over its total duration. Results are displayed in the rose diagrams with 10° bins (Figure 4.20). We obtained extremely steady tremor azimuths at each station suggesting at each station their propagation from a common source. Signals show three principal polarization directions: (a) NNE-SSW for the internal stations (R01B, R02S, R03S, R08B, R10S) of the network; (b) near E-W for most of the external stations (R05S, R06S, R07B); (c) NE-SW for R04S and RSFI stations.
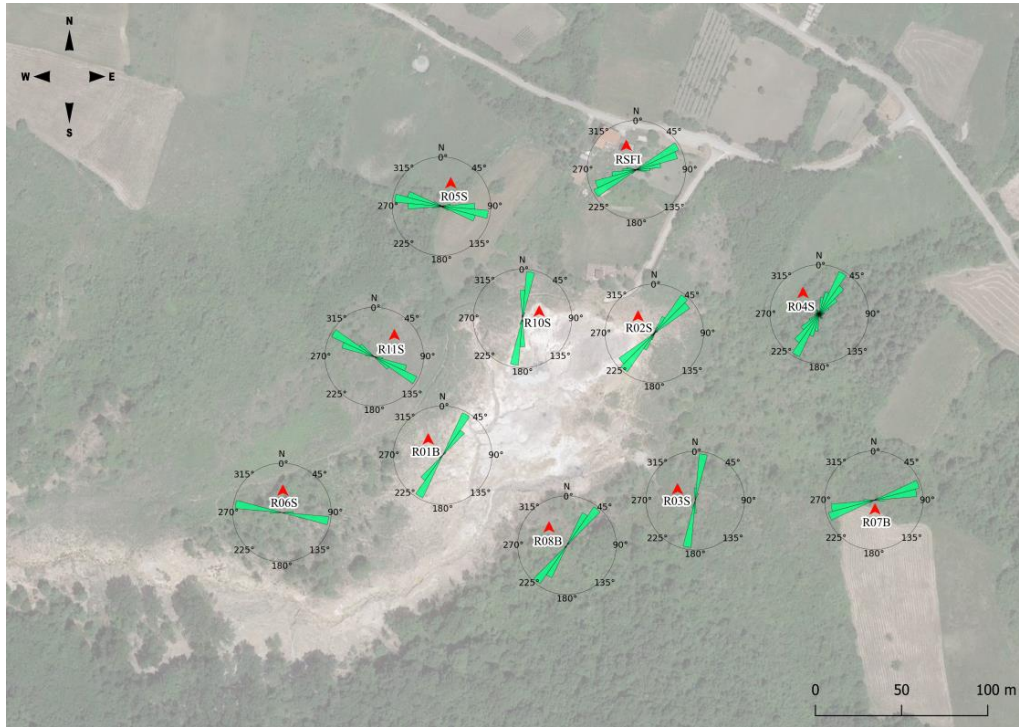
Figure 4. 19 - The map shows the rose diagrams of the mean azimuths calculated for each classified tremor signal at each station on 10° wide bins rose diagrams.

4.5 Tremor detection and classification: discussion

4.5.1 Tremor parameters distributions and correlations

We extracted new information on the previously poorly quantitatively analyzed seismic tremor at Mefite d'Ansanto through an automated and machine-learning approach which was addressed for tremor detection, classification, and discrimination, thus providing a starting workflow for monitoring the non-volcanic emission. The machine-learning discrimination between tremor and anthropic signals allowed us to test the effectiveness of the automated detection algorithm developed in this work. For each analyzed station at least the 75% of the detected signals have been classified from the KNN algorithm as tremors events (Figure 4.21).
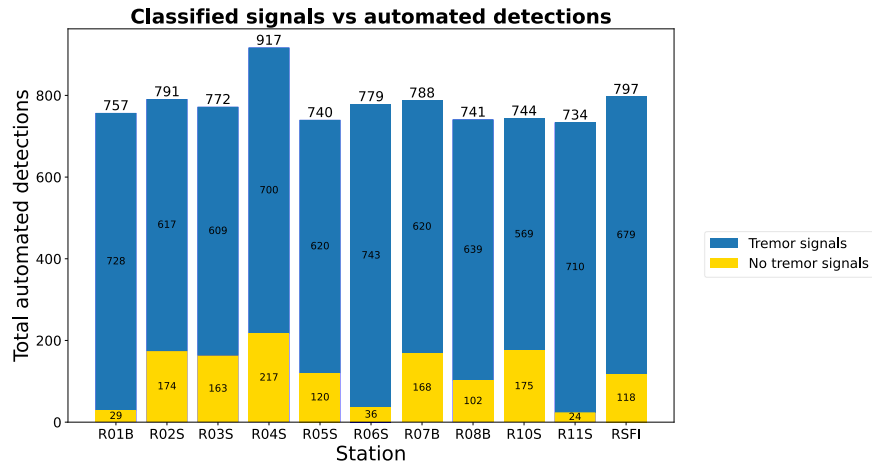
Figure 4. 20 – Discrimination between the KNN-classified tremor (blue) and non-tremor (yellow) over the total automatedly detected signals at each station.

The machine learning analysis on the detected signals was performed at individual stations using the collected tremor parameters (duration, arrival slope, RMS velocity amplitudes) as learning features. First, the unsupervised (DBSCAN) clustering algorithm enabled the pattern recognition of the unknown tremor parameters distributions, while the supervised classification (KNN) algorithm, allowed the tremor discrimination from anthropic signals. We identified typical tremors parameters correlations from the cluster analysis, then confirmed from the subsequent supervised classification. Their intrinsic nature was confirmed by the high consistency at all the stations (Appendix 4.1) with respect to the inhomogeneous ones for the automatically discriminated no-tremor signals. Classified tremors showed a linear correlation (r ≈ 0.4 - 0.6) between the first arrival parametrization (arrival slope) and the RMS amplitudes (Appendix 4.1), thus confirming the effectiveness of the introduced parameter in the estimation of the emergent onset of tremors, testified also by the similar distributions and time-series variability of these parameters (Figure 4.16, 4.17, 4.18). Tremor amplitudes showed similar order of magnitudes at all stations given the dense network setting with respect to the emission. On the other hand, higher amplitude values were recorded by the inner stations of the array (R10S, R03S R02S, R01B), closest to the emission area. An exception is represented by RSFI and R05S outer stations which higher amplitudes would be related to the influence of a smaller, isolated gas emission just to the NW of the main emission area that was observed during a recent survey on the field. The highest amplitudes were recorded at the R10S station, the closest to the bubbling pools and located in the NW area of the emission; while the smallest at R06S, an outer station located at SW of the Mefite. These amplitude observations would reflect the spatial distribution of different emission sources (e.g., bubbling water, pressurized vents) in the Mefite d'Ansanto area. A better source(s) location and assessment would be needed to further investigate the observed tremor amplitude distribution.

4.5.1.1 Tremor amplitude-duration distribution analysis

The high linear correlation between the logarithm of durations and the RMS amplitudes (r ≈ 0.6 - 0.7) of the investigated tremors would suggest an exponential scaling law between the two parameters, similar to the one modelled by (Aki and Koyanagi, 1981) for volcanic tremors. The same distribution has been tested from

(Benoit et al., 2003) at 8 different volcanoes and observed for non-volcanic deep low-frequency (DLF) tremors recorded in subductions areas during interseismic periods (Watanabe et al., 2007). The model is described by the equation:

$$d(D_R) = d_t e^{-\lambda D_r} \qquad \text{Eq. 4. 5}$$

where, d is the duration of tremor greater than or equal to the reduced displacement D_R, d_t is the total duration of tremor, and λ is the slope of the linearized exponential regression or scaling parameter. The reduced displacements (D_R) in the model, defined as the root-mean square (RMS) ground displacement corrected for geometrical spreading, has units of distance * amplitude (e.g., m^2) and is thus a measure of the intensity of the tremor source, similarly to the magnitude for earthquakes (Benoit et al., 2003). The D_R accounts for the instrument magnification (*M*) and the distance to the source (*r*) and is related to the type of waves composing the tremor. The reduced displacement was first defined for body waves (Aki and Koyanagi, 1981) as:

$$D_R = \frac{A}{2\sqrt{2}} * \frac{r}{M} \qquad \text{Eq. 4. 6}$$

Then, Fehler (1983) defined the surface wave reduced displacement as:

$$D_R = \frac{A}{2\sqrt{2}} * \frac{\sqrt{r\lambda}}{M} \qquad \text{Eq. 4. 7}$$

In contrast with earthquake source processes, self-similar (scale-invariant) and well-described by a power law (e.g., frequency-size distribution for earthquakes, Gutenberg & Richter, 1954), the tremor exponential distribution describes scale-bound source processes (e.g., fixed source dimension with variable excess pressures of fluids) (Watanabe et al., 2017). It follows that, the mean of an exponential distribution (characteristic size) which completely describes it, would be related to source processes. For tremor amplitude-duration model, the characteristic amplitude is represented by the inverse of the scaling parameter (λ^{-1}) and is physically interpreted as proportional to a geometric dimension of the source (e.g., length fluid conduits) (Benoit et al., 2003).
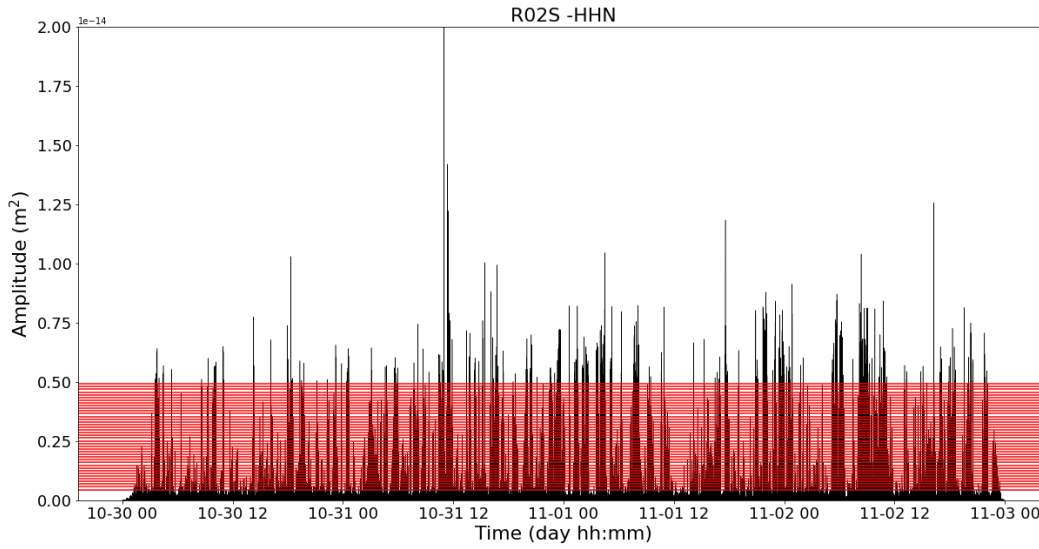
Figure 4. 21 – Example of the selected 40 evenly spaced amplitudes (Ai) used for the fit at the N horizontal component of station R02S. These were calculated starting from the chosen minimum amplitude (A_{min}) with a pace (ΔA) equal to one quarter of it ($\Delta A = \frac{A_{min}}{4}$) such that, the maximum sample value (A_{max}) would be an order of magnitude greater than A_{min}.

We tested the scaling relationships between amplitude and duration of the seismic tremors recorded at Mefite, following a similar procedure to the one explained by Benoit et al. (2003), which was slightly modified, since no source location information were available for the calculation of the D_R (Eq. 4.6, Eq. 4.7). The analysis was performed with the following steps: (a) first, the 4 days-long trace (between 30-10-2019 and 02-11-2019) was filtered in the tremor's peak frequency range (6-15 Hz); (b) the ground displacement (m) was obtained from the instrument response deconvolution; (c) it was squared to obtain ground displacement in D_R units (m^2); (d) evenly spaced sample amplitudes (A_i) were chosen for the fit. These were selected as 40 evenly spaced amplitude values, calculated starting from the chosen minimum amplitude (A_{min}) with a space (ΔA) equal to one quarter of it ($\Delta A = \frac{A_{min}}{4}$) such that, the maximum sample value (A_{max}) would be an order of magnitude greater than A_{min} (Figure 4.22). We selected the A_{min} as the 97th percentile of each station recorded amplitudes distribution to confidently exclude from the analysis most of the noise-related amplitudes; (e) we calculated the cumulative duration (d_{cum}) counting the number of trace samples ($n_{samples}$) with amplitude major or equal to each sample amplitude A_i:

$$d_{cum} = n_{samples} * f_s$$

<div align="right">Eq. 4. 8</div>

where, f_s is the sampling frequency (200 Hz) of traces. The amplitude selection criterion described at point (d) was applied after observing the exponential fit obtained considering all the recorded amplitudes (Figure 4.23). It highlighted a decrease of the overall exponential fit due to the contamination of the noise at small amplitudes and non-tremor spikes (e.g., anthropic noise) at high amplitudes (Figure 4.23). Thus, we decided to explore a narrower amplitude range (A_{min}, A_{max}) more representative of tremor's distribution.
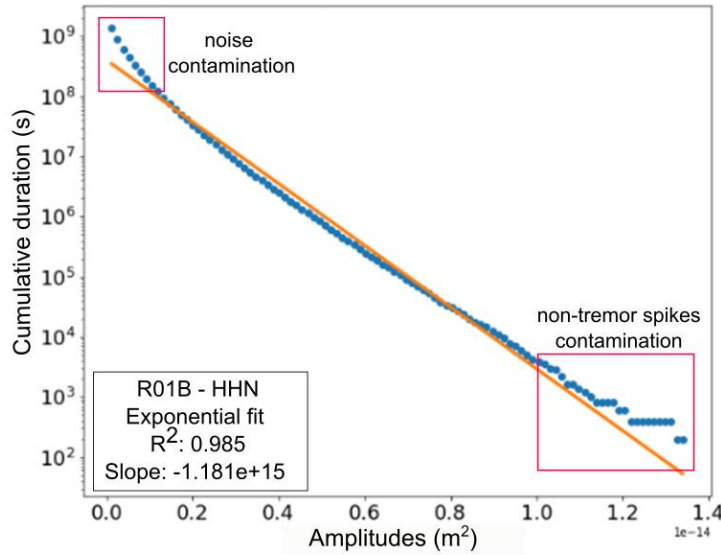
Figure 4. 22 - Exponential fit obtained for the N component of station R01B exploring all the recorded amplitude at the station. It highlighted the contamination of the noise at small amplitudes and non-tremor spikes (e.g., anthropogenic noise) at high amplitudes

Moreover, we estimated that, the contamination due to eventual high-amplitude spikes in the distribution, started at values equal or greater than seven times the selected A_{min}. To better constrain the analysis to tremor-related amplitudes, we fitted only amplitudes (A_{fit}) equal to:

$$A_{fit} = A_i \leq 7A_{min} \qquad \text{Eq. 4. 9}$$

We performed the analysis on both vertical (Figure 4.26) and horizontal components (Figure 4.24, Figure 4.25). The results are summarized in Table 4.4. The goodness-of-the-fit was estimated through the coefficient of determination (R^2). It is the is measure of variance of the modeled distribution, and its best score possible is 1.0. We found a high correlation coefficient of the exponential distributions for the tremor recorded at most of the stations, with higher fit values at the horizontal components (mean $R^2 = 0.97$), for most of which the exponential model is representative for the 99% of the distribution ($R^2=0.99$). The major contamination of high-amplitude spikes recorded at vertical components would explain their dataset lower fit (mean $R^2 = 0.96$). The lowest R^2 were estimated at RSFI station, the noisier station of the array. Assuming an average $\lambda = 3.05 \pm 1.16 \times 10^{15}$, from the highest exponential model fit at the EW components, it is much larger than the one estimated from Benoit et. (2003) for volcanic tremor. Consequently, the characteristic tremor amplitude ($\lambda^{-1} = 3.75 \pm 1.38 \times 10^{-16}$ m^2) is much smaller than the one determined in volcanic areas. This would agree with the scale-bound source processes described by tremor signals: the analyzed tremor at Mefite d'Ansanto, would be related to source processes more local and much less energetic than the ones recorded in volcanic areas. However, since the strict dependence of the exponential model from the source-receiver distance corrected amplitude (D_R), a detailed tremor source location would be needed to improve the analysis and to better assess the observed exponential amplitude-duration scaling law of gas-emission related tremors at the Mefite d'Ansanto.

STA	R^2 (EW)	λ (EW)	λ^{-1} (m^2) (EW)	R^2 (NS)	(λ) (NS)	λ^{-1} (m^2) (NS)	R^2 (Z)	(λ) (Z)	λ^{-1} (m^2) (Z)
R01B	0.99	2.25E+15	4.45E-16	0.99	1.49E+15	6.70E-16	0.99	9.09E+15	1.10E-16
R02S	0.99	1.79E+15	5.59E-16	0.99	1.72E+15	5.80E-16	0.97	6.87E+15	1.46E-16
R03S	0.97	4.04E+15	2.47E-16	0.98	1.52E+15	6.56E-16	0.98	2.19E+15	4.56E-16
R04S	0.98	4.22E+15	2.37E-16	0.98	3.56E+15	2.81E-16	0.92	1.97E+15	5.08E-16
R05S	0.97	1.84E+15	5.45E-16	0.97	2.62E+15	3.81E-16	0.95	9.40E+15	1.06E-16
R06S	0.97	4.69E+15	2.13E-16	0.96	1.55E+15	6.44E-16	0.95	4.09E+15	2.44E-16
R07B	0.97	3.10E+15	3.23E-16	0.96	7.01E+15	1.43E-16	0.92	1.89E+15	5.30E-16
R08B	0.99	4.70E+15	2.13E-16	0.99	2.83E+15	3.54E-16	0.99	1.03E+15	9.67E-16
R10S	0.99	1.81E+15	5.54E-16	0.99	5.59E+15	1.79E-16	0.99	1.52E+15	6.58E-16
R11S	0.99	2.62E+15	3.82E-16	0.99	5.41E+15	1.85E-16	0.98	2.72E+15	3.68E-16
RSFI	0.90	2.48E+15	4.04E-16	0.87	8.88E+15	1.13E-16	0.90	5.80E+15	1.72E-16
mean	0.97	3.05E+15	3.75E-16	0.97	3.83E+15	3.80E-16	0.96	4.23E+15	3.88E-16
stdev	0.03	1.16E+15	1.38208E-16	0.04	2.53E+15	2.20555E-16	0.03	3.07E+15	2.70322E-16

Table 4. 4 – The table summarize the exponential fit results (R^2, λ, λ^{-1}) for all the stations EW (blu panel), NS (green panel) and Z (orange panel) components and their mean values and standard deviations.
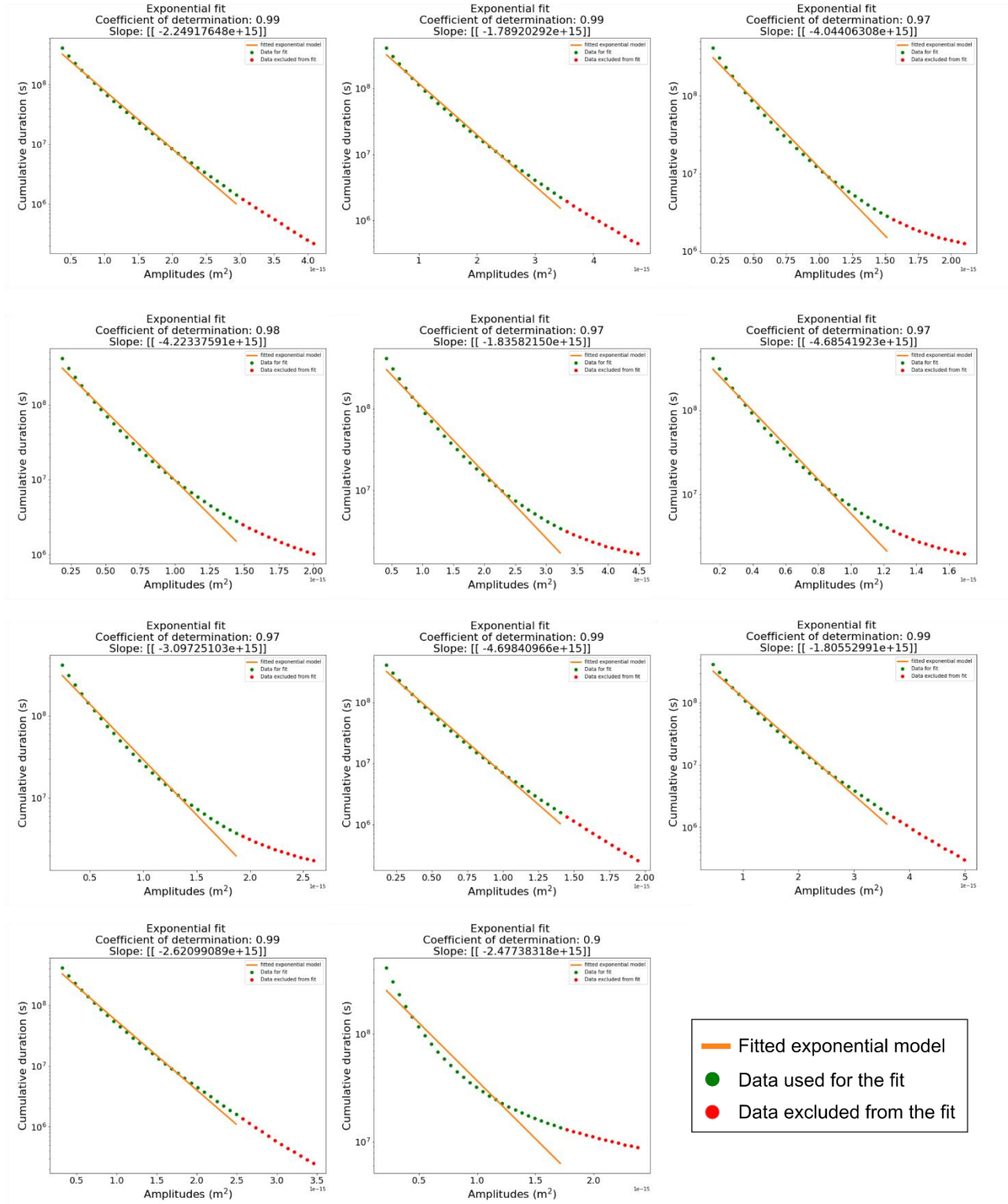
Figure 4. 23 - Exponential fit results for tremor amplitudes recorded at the EW component for each station of the array.
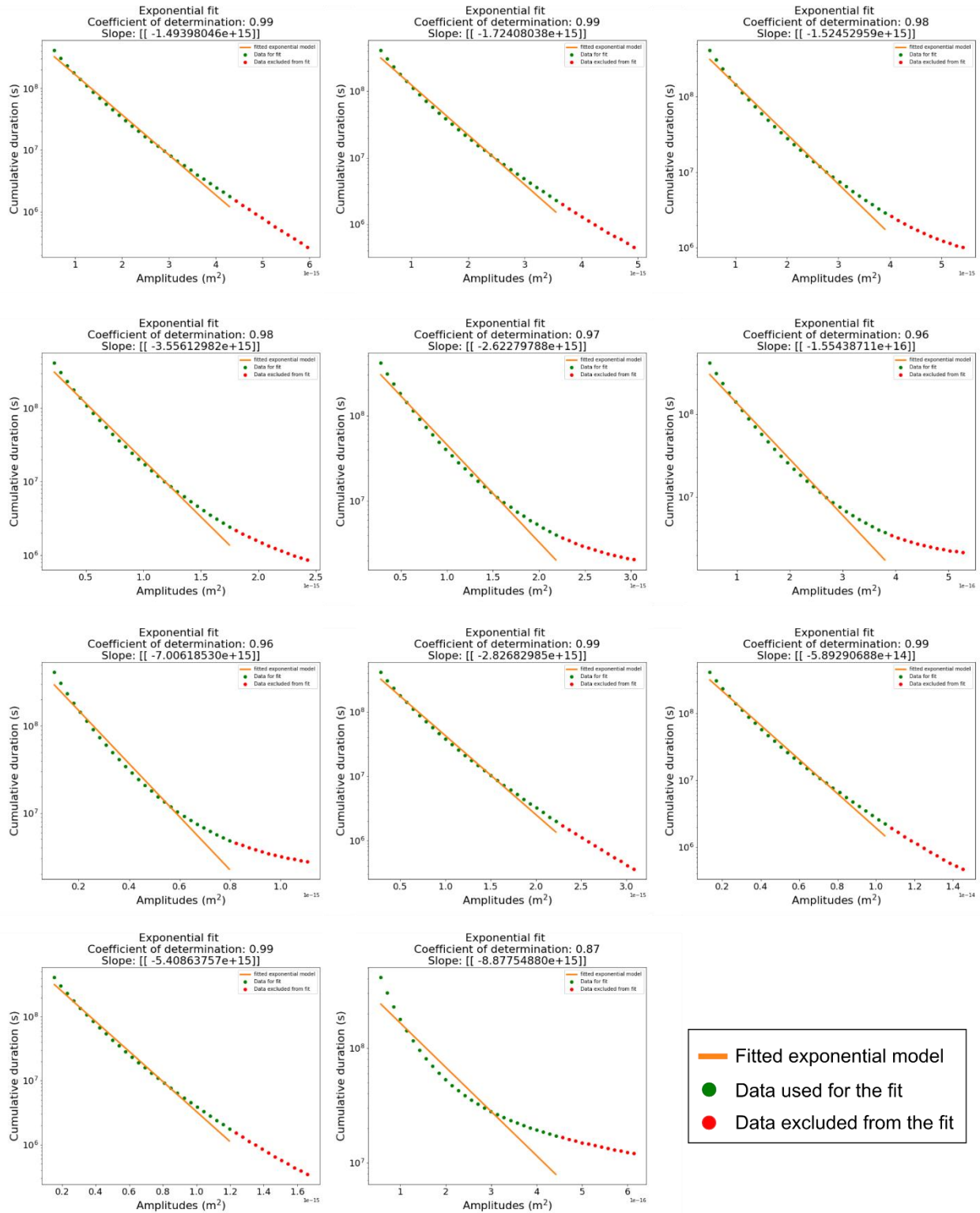
Figure 4. 24 - Exponential fit results for tremor amplitudes recorded at the NS component for each station of the array.
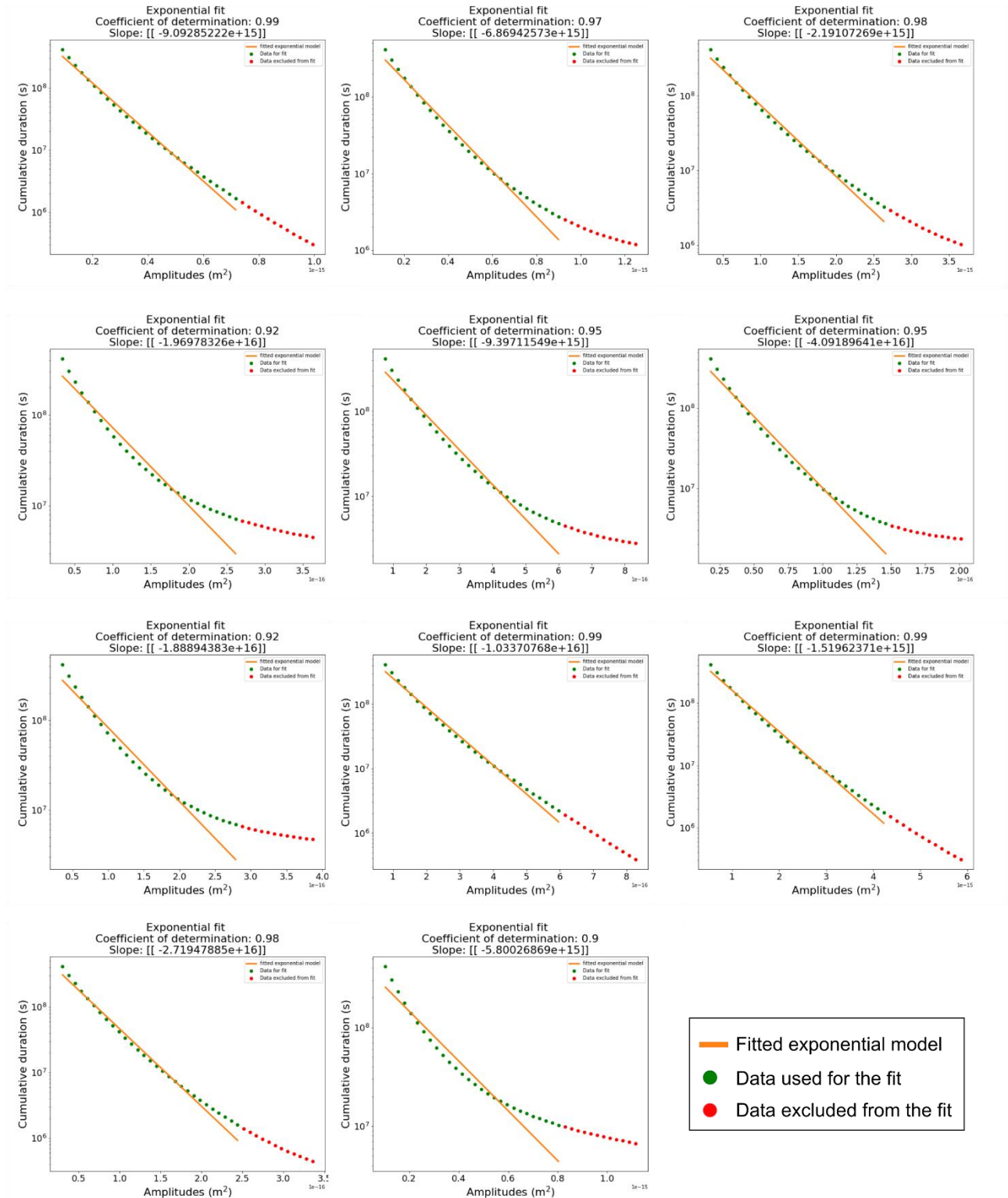
Figure 4. 25 - Exponential fit results for tremor amplitudes recorded at the Z component for each station of the array.

4.5.2 Tremor characterization

The classified tremor in the gas emission area recorded variable velocity displacement amplitudes at different stations and is not stationary with estimated durations ranging between tens of seconds to tens of minutes. Two main kind of tremor signals were observed: (a) long duration (>100 s) and high amplitudes tremors; (b) pulsating tremor made by sequences of closely spaces short tremors of few seconds. Both types of signals occur with no specific patterns. The alternation of impulsive and long tremor recalls the ones described by (Hsu et al., 2013) related to intermittently emitting gases by active mud volcanoes and to the so-called spasmodic and pulsating tremors associated with ash and gas emissions of intermediate energy (Arámbula-Mendoza et al., 2016). The signals spectral content has shown similarities with fumarolic tremors (Giudicepietro et al., 2017, Chiodini et al., 2010) with a spectral peak in a narrow frequency band around 10 Hz. However, tremor at Mefite showed more than one frequency peak in the range between 5 - 20 Hz. It would recall harmonic tremor characterized by multiple peaks at unrelated frequency observed in volcanic settings (e.g., Kilauea volcano, Hawaii) and explained as related to several resonant chambers in the shallow surface layers of active volcanoes (Hurst and Sherburn, 1993) which often resonate at frequencies of a few hertz, in the same frequency range as volcanic tremor (Gordeev, 1992). The main tremors frequency peaks at Mefite can vary from station to station for the same signal. This, would suggest different source-related effects in the gas emission area, characterized by multiple and diffuse potential tremor sources (gas vents, water bubbling) which, together with the local morphology, would cause peculiar spectral patterns at each station.

The polarization analysis (Jurkevics, 1988; Montalbetti and Kanasewich, 1970) showed consistent tremor azimuth at each station for all the analyzed days and suggests that the seismic wave field generated from tremors at Mefite d'Ansanto is local (Figure 4.20). Classified tremor average azimuths calculated at each station show divergent directions: the outer stations (R05S, R06S, R07B, R04S, RSFI) show mainly E-W polarization direction almost radial with respect to emission area for most of the stations, while the internal ones (R01B, R02S, R03S, R08B, R10S) are more variable with a main polarization direction of NNE-SSW. The extremely uneven topography effects of the Mefite d'Ansanto site could explain this polarity difference at the stations from signals generated by a local, internal source with respect to the array. This is also suggested looking at the results of the polarization analysis performed on the external wavefield related to the regional earthquake (mb = 4.7, Bosnia-Herzegovina) recorded from the array (see paragraph 4.3.2). The analysis was performed with the same method but on an external wavefield and, conversely to the observations on tremor's polarization, we observed high coherency between the estimated signal azimuth at all stations with respect to the regional earthquake direction.

4.6 Gas composition and flow measurements: preliminary data and future perspectives

Mefite d'Ansanto is the largest of the deep-source, CO_2-dominant, gas manifestations extensively recognized along the Southern Apennines active fault systems (Chiodini et al., 2004; Ciotoli et al., 2007; Ciotoli et al., 2014; Caracausi and Paternoster, 2015). In this work we analyzed the chemical and isotopic composition of

free gases) collected at the Mefite d'Ansato area (MEPF, MEP, MEF) (Table 4.5, 4.6). Chemical compositions of the investigated gases agree with the one previously defined in literature (Italiano et al., 2000, Chiodini et al.,2010, Caracausi et al., 2013). It mainly consists of CO_2 (mean value of 96.95 vol%) and minor amounts of non-atmospheric N_2 (mean 1.58 vol%), CH_4 (mean 0.21 vol%) and H_2S (mean 0.11 vol%). Also, O_2 concentrations was very low (mean 0.04%), thus indicating a low atmospheric contribution. This is supported by the high $^4He/^{20}Ne$ ratios, ranging from 40.8 to 45.3, higher than atmosphere one (0.318, Ozima and Posodek, 2002). The $^3He/^4He$ values (expressed as R/Ra) are in the range 2.63 - 2.90, similar to those reported in Caracausi and Paternoster (2015). To define the helium isotope signature, R/Ra versus He/Ne diagram (Figure 4.27) was used. As previously highlighted by Caracausi and Paternoster (2015), in the Mefite area, located in the southern Apennine geodynamic scenario, the SCLM (Sub-lithospheric Continental Mantle) with $^3He/^4He$ = 6.1 ± 0.9 Ra and $^4He/^{20}Ne$ ratio = 1000 (Gautheron and Moreira, 2002) is highly representative of the mantle He composition in the area (Dunai and Baur, 1995; Gautheron et al., 2005; Torfstein et al., 2013). Taking into account the SLCM mantle endmember, the He isotope signature at Mefite area is mainly derived from the mantle with a value from 30 - 40%.

It is known that $\delta^{13}C$ values measured in free gases are diagnostic of the carbon origin, because the magmatic source ($-8‰ < \delta^{13}C < -4‰$; Sano & Marty, 1995) has values different than contribution from subducted marine limestone ($\delta13C = 0‰$) and matter of organic origin ($\delta^{13}C = -30±10‰$; Hoefs, 2015). In addition, the He-CO_2 systematic allows to get more detailed information about the carbon source. Thus, we correlated the variation of the $CO_2/^3He$ versus $\delta^{13}C$ ratio based on the approach proposed by Sano and Marty (1995 and references therein). Figure 4.28 plots two mixing curves modeled considering both an organic and a limestone endmember, in which the mantle corresponds to $CO_2/^3He = 7.0 \times 10^9$ and $\delta^{13}C = -4.0‰$ (Brauer et al., 2016). For both organic and limestone endmembers, a value of $CO_2/^3He = 1.0 \times 10^{13}$ is assumed, whereas for organic and limestone $\delta^{13}C$ endmember is assumed $\delta^{13}C = -30‰$ and $\delta^{13}C = 0‰$, respectively (Sano & Marty, 1995). In the Mefite gases a slightly variability of the $CO_2/^3He$ ratio is observed. The data overlaps with SCLM endmembers with a low contribution from limestone. Therefore, a mixing of mantle and crust-derived fluids is mainly responsible for the Helium and carbon isotopic signature of the Mefite gases.

Sample ID	Sample Name	Sampling date	Latitude UTM-WGS84	Longitude UTM-WGS84	Zone UTM-WGS84	CH₄ (%)	O₂ (%)	N₂ (%)	CO₂ (%)	H₂S (%)
MEF	Mefite Pozzo fango	26/06/2020	4536037	512679	33T	0.20	0.03	1.46	96.83	0.13
MEPF	Mefite Polla Acqua	26/06/2020	4536043	512687	33T	0.23	0.05	1.74	97.09	0.14
MEP	Mefite	26/06/2020	4535918	512252	33T	0.20	0.05	1.55	96.95	0.06

Table 4. 5 – Samples IDs, names, sampling date, location, and analyzed free gas composition (% vol).

Sample ID	⁴He/²⁰Ne	⁴He ppm	²⁰Ne ppm	R/Ra	R/Rac	δ¹³C_CO2	δ¹³C_CH4
MEF	40.83	12.63	0.309	2.90	2.92	0.1	-29.0
MEPF	45.35	14.77	0.326	2.89	2.90	0.1	-29.0
MEP	42.09	12.70	0.302	2.63	2.64	0.1	-28.4

Table 4. 6 – Samples analyzed isotopic composition.
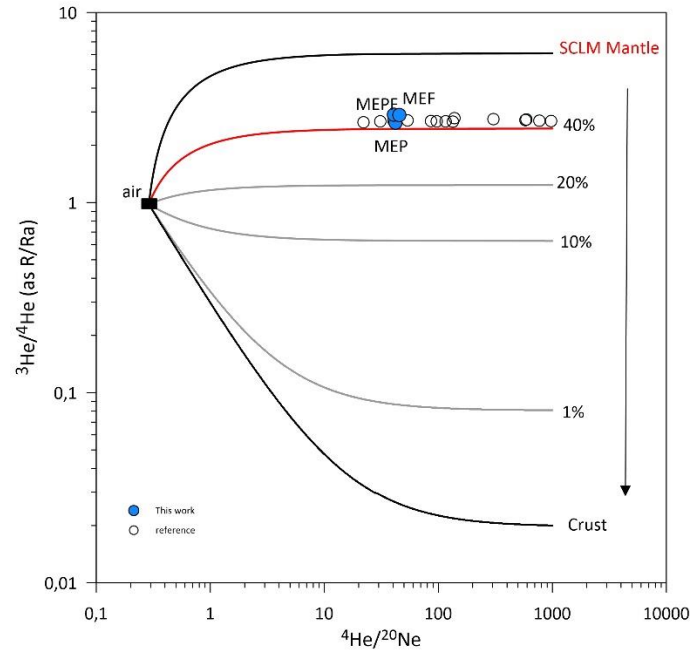
Figure 4. 26 - Helium isotopic ratios (as R/Ra values) and He/Ne relationships for the Mefite samples and comparison with literature reference data (Italiano et. al., 2001, Caracausi and Paternoster., 2015b). The theoretical lines (black lines) represent binary mixing trends for mantle-originated, crustal and atmospheric helium, were calculated using the following endmembers: air ($^3He/^4He = 1.39 \times 10^{-6}$, $^4He/^{20}Ne = 0.318$; Ozima et al.2002); SCLM Mantle (red solid line) ($^3He/^4He = 6.1 \pm 0.9$ Ra, $^4He/^{20}Ne = 1000$; Gauteron et al.2002.; Buttitta et al. 2020); crust ($^3He/^4He = 0.02$ Ra, $^4He/^{20}Ne = 1000$; Ballentine et al., 2002). Western side of central-southern Apennine (WA) mantle endembers (red dashed lines) between air and deep fluids with He-mantle contribution typical of the Phlegrean Fields ($^3He/^4He = 2.8$ Ra) and the Vesuvius volcano ($^3He/^4He = 3.6$ Ra) were also added.
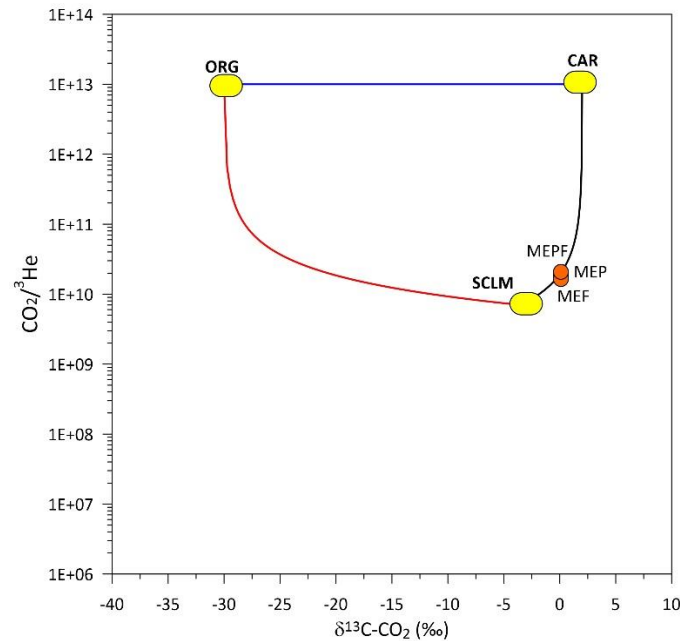
Figure 4. 27 - Plot of CO2/ 3He versus δ13C-CO2 (‰) for TRA and Mefite samples (see Chapter 4). Abbreviations: SCLM, sub-continental lithospheric mantle; CAR, carbonate; ORG, organic matter.

4.6.1 Tremor and gas flux correlation: preliminary test

Tremor signals at Mefite d'Ansanto can contain key (and ignored) information on fluid-rock interactions at depth. This is particularly true if it is demonstrated that the analyzed seismic tremors are a proxy of the gas flux, which can be possible since the widely observed relationships in volcanic environments between infrasound, tremor, and gas flux at open-vent volcanoes and fumaroles (Chiodini et al., 2017; Delle Donne et

al., 2016; Matoza et al., 2013; McKee et al., 2017). Chiodini et al. (2010) performed gas flow rate estimation at Mefite d'Ansanto with the aim to model the gravity driven flow of CO_2 flowing toward southwest over the valley and identified the most dangerous areas, where air CO_2 concentrations can be higher than the safety threshold for human health. However, no previous studies are currently reported in literature which establish a correlation between seismic tremor observation and gas flux measurements at Mefite gas emission. In this work, a first attempt of multiparametric experiment was performed by means of contemporary seismic and anemometric measurement of the volume flow (m^3/h) at one pressurized vent located at the southeastern side of the emission area (Figure 4.29). The analyzed vent was chosen since a PVC tube for monitoring purposes was already installed in the area. The experiment was carried out jointly acquiring continuous seismic and anemometric recordings for 19 hours - from 2021-07-01 18:00:00 (UTC) to 2021-0-02 13:00:00 (UTC). The anemometric measurements were performed through a *testo 400* control unit provided of a vane anemometer (Ø=100mm) for flow velocity measurement and with a 200 mm funnel already described in Chapter 3 (Paragraph 3.8.2) installed directly on the emission vent (Lat: 40.9738°N, Lon: 15.1458°E) (Figure 4.30b). The testo 400 control unit was set in polystyrene thermal box, covered by an aluminum sheet to avoid instrument overheating (Figure 4.30a). The seismic station provided with a 120 s broadband sensor was installed to the northwest (Lat: 40.9748°N, Lon: 15.1455°E, Figure4.28 and Figure 4.30c). Seismic data were acquired with a 250 Hz sampling frequency, while a 1 second sampling period was applied for gas volume flow measurements. The linear distance between the two station was of 113 m. Figure 4.31 shows the total anemometric time series obtained in the analyzed 19 hours-long time-window. The chosen vent was characterized by a high gas volume flow, almost constant during the analyzed timespan, with slight recorded variations between 418.5 and 440.8 m^3/s and an estimated median value of 430 m^3/h. No clear correlations with the recorded seismic tremor were observed in this preliminary test to constrain the relationship between gas volume flow and the recorded seismic tremor. However, it highlighted that the anemometer volume flow measurement precision ($\pm$0.1 m^3/h) does not allow a resolution comparable to the one of the seismic methods for the chosen gas vent, characterized by a high-volume flow. Further improvement of the experiment would be the choice of less energetic emission vents at Mefite to obtain a higher resolution in the gas flux measurements. Moreover, field observations at the emission area showed that it is characterized by different sources (e.g., bubbling site, pressurized vents) in many cases producing emission-related sounds such that, a suitable improvement of the experiment, would be the addition of hydrophone records looking for a proxy between seismic tremor and gas flow at Mefite d'Ansanto.

Figure 4. 28 – Location of seismic (red triangle) and anemometric (orange triangle) stations deployed for the joint measurements preliminary test at Mefite d'Ansanto. Figure also shows the chosen pressurized emission vent for gas flow measurements.

Figure 4. 29 – Preliminary test measurement installation: (a) testo 400 control unit deployed in a polystyrene thermal box and (b) covered by an aluminum sheet to avoid instrument overheating; (c) seismometer installation.
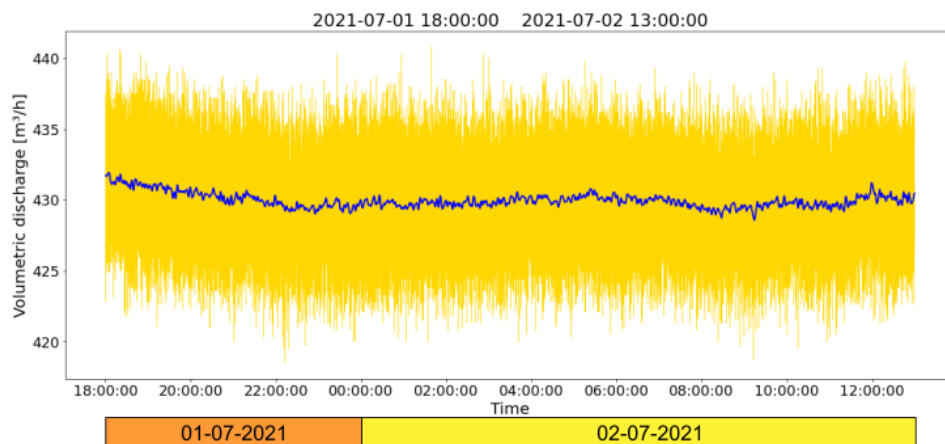
Figure 4. 30 – Volume flow (m^3/h) time-series in the analyzed 19 hours- preliminary long time-window test: raw data (yellow line) along with smoothed data (1D-gaussian filter, $\sigma=50$) (blue line) are reported.

Chapter 5 – Conclusion and future perspectives

Considering the strong socioeconomic impact of fluid injection environmental issues (Vengosh et al., 2014, Burton et al., 2014, Pichtel, 2015) and its seismic implications (National Research Council, 2013; Ellsworth, 2013; Grigoli et al., 2017) the current research in this field needs the development of comprehensive studies involving the integration of different research fields (e.g., geology, seismology, hydrogeology, rocks geomechanics) for a more complete understanding of the complex relationships between rocks, faults, and fluids as a complex system. This may be crucial to constrain source processes involved in both induced and natural earthquake generation and for the development of effective seismic and environmental monitoring strategies (e.g., traffic-light systems, long-term geochemical monitoring) of areas where fluids injections are carried out. On this basis, this PhD project applied seismic, geochemical and machine learning methods with the aim to contribute in a step-forward in the development more effective, semi-automated and long-term monitoring strategies of two key study areas: (a) the High Agri Valley (Chapter 3), the largest onshore oil field in Europe located in the tectonically active southern Apennines where fluid-related clusters of anthropogenic microseismicity have been recognized (Huger et al., 2021, Stabile et al., 2021, Telesca et al., 2015); (b) the Mefite d'Ansanto, the largest natural emission of CO_2-rich gases with mantle-derived fluids (from non-volcanic environment) ever measured on the Earth (Caracausi et al., 2015, Chiodini et al, 2010, Italiano et al., 2001,) which arise in one of the high seismic hazard areas of southern Apennines (Irpinia area) and which offers the opportunity to study the complex interaction between fluid migration in the subsurface, Earth degassing and earthquake generation and develop monitoring techniques of large gas leakage from Carbon Capture and Storage (CCS) injection projects (Chapter 4).

5.1 High Agri Valley study area

In this work, both standard and advanced methodologies were applied for the characterization of the High Agri Valley seismicity.

The standard detection and location workflow, consisted in manual detection of earthquakes and their phase picking, followed by both linearized and probabilistic earthquake locations methods. The approach, despite quite time-consuming, allowed us to provide the High Agri Valley preliminary seismicity catalogue in the time-span between September 2016 and March 2019 (available at: https://doi.org/10.5281/zenodo.3632419, Stabile et al., 2020) obtained from linearized (HYPO71, Lee and Lahr, 1975) locations in a 1D-velocity model of the study area (Improta et al.,2017). The results displayed the microseismic nature of anthropogenic earthquakes in the area (Ml<2). However, the analysis of the preliminary local magnitude distribution highlighted the necessity to apply magnitude estimation techniques more suitable for the dataset in the future, particularly for low magnitude events (Ml < 0.4). We improved the preliminary catalogue through earthquakes relocation by means of the equal differential time (EDT) method implemented in a non-linear global approach algorithm (NonLinLoc code, Lomax et al., 2000) and considering a 3-D velocity model recently developed for the High Agri Valley from Serlenga and Stabile (2019). The application resulted in a more accurate event locations (overall median RMS = 0.1) and in a better discrimination of the defined earthquakes categories and

more defined clusters in the HAV seismicity resulting from both injection-induced events and tectonic local earthquake swarms (e.g., Pignola sequence). Moreover, the NLLOC relocation better highlighted the presence of a seismic gap between the High Agri Valley seismicity (to the southwest) and the deeper (mean focal depth 15-20 km) Apennines crustal earthquake to the northeast and occurring in the area where the hydrocarbon production of the Val d'Agri Oil field is currently carried out. Its location is in accordance with the stabilization effect of the extraction operations suggested by Hager et al., 2021. The performed cluster analysis on the relocated High Agri Valley seismicity catalogue underlined the high potential of this unsupervised machine-learning approach as a useful automated tool to make faster dataset exploratory analysis (K-Means) and to identify clusters in the located seismicity (DBSCAN). The density-based approach (DBSCAN) was particularly suitable to verify the presence of clusters in the located HAV seismicity related to both anthropogenic and local seismic swarms and sequences and to discriminate if from background, sparse, Apenninic seismicity. The analysis was performed with the available standard earthquake parameters (latitude, longitude, hypocentral depth, Ml) which ranges were highly comparable within the analyzed dataset, except for the hypocentral depth. Future development of the analysis will involve the collection of more discriminating information about the High Agri Valley seismicity (e.g., earthquakes source parameters) for a more efficient recognition of clusters within it, and with the aim to find discrimination criteria between induced and tectonic local seismicity.

A semi-automated workflow for earthquake detection and location was tested in this work with the aim to improve the current standard procedures, quite time-consuming and strictly related to human operators for which is becoming increasingly difficult to manually keep up with the data analysis, with the ever-growing development of dense monitoring networks and high data fluxes. The approach involved the integration of manual (WebObs, Beauducel et al., 2020) and advanced semi-automatic (single-station template matching algorithm, Stabile et al., 2021) and automatic (e.g., PhaseNet deep-learning algorithm, Zhu and Beroza, 2018) detection and location methodologies and was tested for the characterization of a low-magnitude tectonic earthquake sequence which occurred in August 2020 in the southeastern sector of the High Agri Valley, close to the municipality of Castelsaraceno and characterized by a 2.9 Ml mainshock on 7th Augsut.

An automated, iterative four-step absolute earthquake (EDT) linear global approach (NonLinLoc) location method in the HAV 3D-velocity model (Serlenga and Stabile, 2019) was implemented to progressively discard, at each iteration, the high residuals seismic phases (both P- and S- phases residuals), possibly related to wrong and/or multiple automated arrival times picks. We improved the local magnitude estimation through the local magnitude scale proposed Bobbio et al., 2009 for southern Italy. The method will be extended to all the HAV seismicity in the future, with the aim to overcome the observed limitation in the default estimated local magnitudes. Absolute locations were refined by applying the double-difference method (hypoDD code, Waldhauser and Ellsworth, 2000) in the same 3D P- and S-wave velocity model (Serlenga ann Stabile, 2019). The 75% of the total number of events occurred between 7 and 10 August, thus revealing the typical behavior of a mainshock-aftershock sequence and unfolded a 10 km depth anti-apenninic structure. The mainshock focal mechanism (FPFIT code, Reasemberg and Oppenheimer, 1985) highlighted a (strike = 255°, dip = 55°, and rake = -30°) normal/lateral left strike-slip faulting kinematic; while, its source parameters (M_0, f_c, $\Delta\sigma$) retrieved

through the S-wave displacement spectra inversion (SourceSpec algorithm, Satriano, 2021) showed a stress drop $\Delta\sigma$=1.40 Mpa consistent with the one obtained from other authors (Festa et al., 2021; Stabile et al., 2012) for natural events in Southern Italy.

The proposed semi-automatic detection and location method enabled us to characterize the Castelsaraceno sequence in a relatively short time with respect to the standard techniques. thus, representing a starting point for the improvement of the efficiency of seismic monitoring techniques of both anthropogenic and natural seismicity in the High Agri Valley.

The geochemical study of the HAV groundwaters aimed to assess the geochemical processes controlling its groundwater chemistry and to define a model of fluids origin and the water–rock–gas interaction processes in aquifer systems of the High Agri Valley. Between all analyzed samples, all the examined groundwater has a meteoric origin although springs located on the hydrological left of the valley, show longer migration paths in the subsurface and some spring (SUF, TRA) showed a deep flow than the other ones. Samples water and gas chemistry highlighted a substantial difference of the Tramutola well thermal fluids compared to the rest of the dataset. Most of the HAV samples hydrofacies (except for the Tramutola ones) show a predominant bicarbonate alkaline-earth composition reflecting the main geochemical process affecting water chemistry is the calcite (and subordinately dolomite) hydrolysis. Minor silicate dissolution processes were recognized in some samples, related to different processes: (a) inverse cation exchange with depletion of Na for Ca ions at water-clay mineral surface interaction; (b) Na-Feldspar hydrolysis processes. Differently from most HAV springs water samples, the observed Na-HCO$_3$-Cl water type of Tramutola well suggested more complex processes of carbonate dissolution and subsequent direct cation exchange due their contact, at shallow levels in the well, to clay rocks. The proposed interaction was also suggested by Italiano et al. (2001) and may be suitable for Tramutola waters since they rise across clays and carbonatic rocks (ENI, 1946). The total dissolved carbon isotopic constraints (δ^{13}C-TDIC) allowed to model the evolution of infiltrating waters and let us identify the oxidation and dissolution of soil organic matter as the main source of dissolved carbon during the infiltration of the meteoric waters from the recharge areas. Only for two water samples (ALLI and MOL) the total amount of carbon deriving from external sources with respect to carbonate dissolution (C$_{ext}$) suggested a deep source input, comparable with the isotopic signature of the CO$_2$ of deep provenance measured at the Tramutola well samples. Groundwater quality and its suitability for drinking use were evaluated for all the analyzed samples - except for the Tramutola thermal water - by comparing different chemical parameters and with those reported by the Italian (D. Lgs 31/2001) and European (98/83/CE) legislation guidelines. Results confirmed previous groundwater quality assessment performed by Paternoster et al. (2021) to both spring and well samples in the HAV and demonstrated that most of HAV groundwater is chemically suitable for drinking use with respect to the analyzed traces, minor and major inorganic chemical compounds. The only exception was represented by PZ7 sample which anomalous amount of NO$_3$ amount (370 mg/l) above the prescribed limit (50 mg/l) is imputable to presence of manure due to animal grazing main anthropic activity.

In this work, a detailed compositional and isotopic analysis of thermal fluids at Tramutola well was gaseous species was performed to discriminate its natural sources (crustal, mantle and atmospheric) and to define CH$_4$

and CO_2 origin and output. Unlike HAV groundwater, TRA gases are CH_4-dominated (82%), CO_2 is up to 1.92%, O_2 and N_2 up to 0.06% and 12.8 % respectively. The $^4He/^{20}Ne$ ratio in the collected gases is three orders of magnitude higher than air-one (0.318), highlighting that atmospheric contribution is negligible. Helium isotope ($^3He/^4He$, expressed as R/Ra) indicate a prevalent radiogenic component with a contribution of mantle-derived helium (~20%). An additional mantle-derived CO_2 component is hypothesized as suggested by $\delta^{13}C_{CO2}$ measured values (-3.5 ‰). TRA gases have low $CO_2/^3He$ ratios compared to mantle carbon endmember, probably due to secondary processes such as calcite precipitation. Since also crustal and surficial sources of CO_2 were hypothesized, further investigation would be needed in the future on the CO_2 mass balance from the different reservoirs to justify the measured value of $\delta^{13}C_{CO2}$ at surface. Methane isotope composition indicates a likely microbial isotopic signature ($\delta^{13}C$-CH_4 =−63.1‰, δD-CH_4=−196‰), probably due to either (1) biodegradation processes of thermogenic hydrocarbons, earlier generated by thermocracking process or (2) ongoing microbial methanogenesis in the shallow organic-rich clays hosting the gas. The methane output, evaluated by means of anemometric measurement of the volume flow (m3/h) and flow velocity (m/s), is of ~156 t/y, that represent about 1.5% of total national anthropogenic sources related to fossil fuel industry (Etiope et al., 2007. The measured CO_2 output (3.2 ton/y) is from 2 to 5 orders of magnitude lower than budget of mantle-derived CO_2 from Italian volcanoes.

Earthquake locations data, along with other geophysical investigations (e.g., magnetotelluric model of the area, Balasco et al., 2015, 2021; Vp/Vs velocity models, Valoroso et al., 2011, Improta et al., 2017, Serlenga and Stabile, 2019) showed the presence of deep-rooted faults in the crust below the Tramutola well. The demonstrated deep source of Tramutola thermal fluids, and their rising at surface, would testify the connection of these structures with surficial crustal layers and their role as preferential conduits for fluid circulation in the crust below the well. Our study highlighted as the Tramutola well site may represent a key natural laboratory in the HAV to better understand the complex coupling effects between mechanical and fluid-dynamic processes in earthquake generation. A multi-parametric geochemical station for the continuous geochemical monitoring purposes was installed at the well during the last few months of the PhD project. In this work, preliminary observations on three-month long period on the measured temperature, electrical conductivity, and water level are reported. However, continuous monitoring of groundwater level at Tramutola well would be a key tool for future investigation on the still scientific open-issue of co-seismic and post-seismic groundwater-level responses to earthquakes and their possible role as geochemical indicator for seismic hazard and monitoring purposes.

5.2 Mefite d'Ansanto study area

In this work, we developed an automated detection algorithm and applied a machine-learning approach, which enabled the automated Mefite tremor's classification and discrimination, thus providing a starting workflow to monitor the non-volcanic emission either to gain information on its evolution for seismic hazard purposes or providing an early step-forward in the automated management techniques of both natural and CCS leakage CO_2 gas-emissions. Indeed, similarly to the extensively application of the phenomenon in volcanic settings -

for volcanic activity monitoring and forecasting - the tremor recognized at Mefite could have a key role in the assessment of the complex evolution of this massive non-volcanic gas emission.

The extraction of new information on the previously poorly quantitatively analyzed Mefite seismic tremor was enabled by the installation of a very dense seismic network around the emission, which allowed to record high-resolution data of the local and emission-related background seismicity. The developed automated detection algorithm allowed to collect a large dataset of tremors at Mefite d' Ansanto to be analyze with machine-learning approach at individual stations. The latter enabled tremor classification, and its discrimination from accidentally detected anthropic noise, thus providing a starting workflow for monitoring the non-volcanic emission. The automatically classified tremor signals are characterized by typical parameters (duration, RMS amplitude, arrival slope) correlations which intrinsic nature was confirmed by the high consistency of at all the stations of the array. Particularly, the linear correlation between the logarithm of durations and the RMS amplitudes would be explained from the observed exponential scaling law between the two parameters, similar to the one modelled by Aki and Koyanagi (1981) for volcanic tremor. The classified tremor in the gas emission area is not stationary and consist of two main kind of tremor signal with no specific daily distribution over the records: (a) long duration and high amplitudes tremors; (b) pulsating tremor made by sequences of closely spaces short tremors of few seconds. The signals showed common characteristics at all the stations, such as: (1) an emergent onset almost simultaneous with higher amplitudes with respect to the recorded background noise; (2) highly variable durations ranging between 20 seconds and 30-40 minutes; (3) tremor spectral content with a spectral peak in a narrow frequency band around 10 Hz, with multiple peaks at frequencies in the range between 5 - 20 Hz which amplitude locally changes at different stations. The polarization analysis performed to automated classified tremors showed a consistent tremor azimuth at each station thus confirming their common source area and suggests that the seismic wave field generated from tremors at Mefite d'Ansanto is local. Future developments of tremor source location methods would be needed - involving non-conventional techniques (e.g., based on polarization analysis) - to discriminate eventual multiple tremors sources as suggested by tremor spectral content and RMS amplitudes at different stations of the dense array, which would reflect the spatial distribution of different local emission sources (gas vents, water bubbling). The source location of the emission related tremor at Mefite d'Ansanto would represent a step forward in its characterization to further understand its source processes, and for setting up more advanced automated detection and machine learning classification techniques to exploit the information provided by seismic tremor for an improved automatic monitoring of non-volcanic, CO_2 gas emissions. The free gases samples collected at Mefite showed a composition similar the one defined in literature (Italiano et al., 2000, Chiodini et al.,2010, Caracausi et al., 2013), mainly consisting of CO_2 (mean 96.95 vol%) and minor amounts of non-atmospheric N_2 (mean 1.58 vol%), CH_4 (mean 0.21 vol%) and H_2S (mean 0.11 vol%). Gas chemical and isotopic analyses ($^3He/^4He$, R/Rac, $^4He/^{20}Ne$, $\delta^{13}C_{CO2}$) indicate a trivial atmospheric source contribute to the gas composition. Considering the position of Mefite in the geological context of the area, the western Apennines mantle (Caracausi and Paternoster, 2015b) a SLCM mantle endmember is highly representative of the mantle derived He in the Mefite d'Ansanto samples. We found that the He isotope signature at Mefite area is mainly derived from the mixing of mantle (30%-40%) and crust-derived fluids. In this work, a first attempt of multiparametric

experiment was performed by means of contemporary seismic and anemometric measurement of the volume flow (m^3/h) at one pressurized vent located at the southeastern side of the emission area looking for a proxy between seismic tremor observation and gas flux measurements at Mefite gas emission. No clear correlations with the recorded seismic tremor were observed in this preliminary test. However, it highlighted that (a) a further improvement of the experiment would be the choice of less energetic emission vents at Mefite to obtain a higher resolution in the gas flux measurements; (b) the addition of hydrophone records of the emission-related sounds would be a further improvement of the experiment.

References

Abercrombie, R.E., 1995. Earthquake source scaling relationships from −1 to 5 ML using seismograms recorded at 2.5-km depth. Journal of Geophysical Research: Solid Earth 100, 24015–24036. https://doi.org/10.1029/95JB02397

Abercrombie, R.E., 2015. Investigating uncertainties in empirical Green's function analysis of earthquake source parameters: uncertainties in EGF analysis. J. Geophys. Res. Solid Earth 120, 4263–4277. https://doi.org/10.1002/2015JB011984

Aki K. & Richards P. G. 1980. Quantitative Seismology, Theory and Methods. Volume I: 557 pp., 169 illustrations. Volume II: 373 pp., 116 illustrations. San Francisco: Freeman. Price: Volume I, U.S. 35.00. ISBN 0 7167 1058 7 (Vol. I), 0 7167 1059 5 (Vol. II). Geological Magazine 118, 208–208. https://doi.org/10.1017/S0016756800034439

Aki K., 1967. Scaling law of seismic spectrum. J. Geophys. Res. 72, 1217–1231. https://doi.org/10.1029/JZ072i004p01217

Aki K., Koyanagi R., 1981. Deep volcanic tremor and magma ascent mechanism under Kilauea, Hawaii. Journal of Geophysical Research: Solid Earth 86, 7095–7109. https://doi.org/10.1029/JB086iB08p07095

Alavi, A.H., Gandomi, A.H., 2011. Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. Computers & Structures 89, 2176–2194. https://doi.org/10.1016/j.compstruc.2011.08.019

Amoroso, O., Russo, G., De Landro, G., Zollo, A., Garambois, S., Mazzoli, S., Parente, M., Virieux, J., 2017. From velocity and attenuation tomography to rock physical modeling: Inferences on fluid-driven earthquake processes at the Irpinia fault system in southern Italy: From Seismic Tomography to Rock Modeling. Geophys. Res. Lett. 44, 6752–6760. https://doi.org/10.1002/2016GL072346

Anderson, J.G., Hough, S.E., 1984. A model for the shape of the fourier amplitude spectrum of acceleration at high frequencies. Bulletin of the Seismological Society of America 74, 1969–1993. https://doi.org/10.1785/BSSA0740051969

Ansari, Z., Azeem, M.F., Ahmed, W., Babu, A.V., 2011. Quantitative Evaluation of Performance and Validity Indices for Clustering the Web Navigational Sessions 10.

Apollaro, C., Caracausi, A., Paternoster, M., Randazzo, P., Aiuppa, A., De Rosa, R., Fuoco, I., Mongelli, G., Muto, F., Vanni, E., Vespasiano, G., 2020. Fluid geochemistry in a low-enthalpy geothermal field along a sector of southern Apennine's chain (Italy). Journal of Geochemical Exploration 219, 106618. https://doi.org/10.1016/j.gexplo.2020.106618

Arámbula-Mendoza, R., Valdés-González, C., Varley, N., Reyes-Pimentel, T.A., Juárez-García, B., 2016. Tremor and its duration-amplitude distribution at Popocatépetl volcano, Mexico: TREMOR AT POPOCATÉPETL. Geophys. Res. Lett. 43, 8994–9001. https://doi.org/10.1002/2016GL070227

Ascione, A., Ciotoli, G., Bigi, S., Buscher, J., Mazzoli, S., Ruggiero, L., Sciarra, A., Tartarello, M.C., Valente, E., 2018. Assessing mantle versus crustal sources for non-volcanic degassing along fault zones in the actively extending southern Apennines Mountain belt (Italy). GSA Bulletin 130, 1697–1722. https://doi.org/10.1130/B31869.1

Astiz, L., Dieterich, J.H., Frolich, C., Hager, B.H., Juanes, R., Shaw, J.H., 2014. On the potential for induced seismicity at the Cavone oilfield: analysis of geological and geophysical data, and geomechanical modeling. Technical Report. Report for the Laboratorio di Monitoraggio Cavone. http://labcavone.it/documenti/32/allegatrapporto-studiogiacimento.pdf

Azimi S.A.Kalinin A.V. Kalinin V.V and Pivovarov B.L.1968. Impulse and transient characteristics of media with linear and quadratic absorption laws. Izvestiya - Physics of the Solid Earth 2. p.88-93

Bakalowicz, M., 1994. Water Geochemistry: Water Quality and Dynamics, in: Groundwater Ecology. Elsevier, pp. 97–127. https://doi.org/10.1016/B978-0-08-050762-0.50011-5

Balasco, M., Cavalcante, F., Romano, G., Serlenga, V., Siniscalchi, A., Stabile, T.A., Lapenna, V., 2021. New insights into the High Agri Valley deep structure revealed by magnetotelluric imaging and seismic tomography (Southern Apennine, Italy). Tectonophysics 808, 228817. https://doi.org/10.1016/j.tecto.2021.228817

Balasco, M., Giocoli, A., Piscitelli, S., Romano, G., Siniscalchi, A., Stabile, T.A., Tripaldi, S., 2015. Magnetotelluric investigation in the High Agri Valley (southern Apennine, Italy). Natural Hazards and Earth System Sciences 15, 843–852. https://doi.org/10.5194/nhess-15-843-2015

Balderer, W., Martinelli, G., Aeschbach, W., Kipfer, R., Kahr, G., Nuesch, R., and Wolf, M. 1996. Isotopic and chemical investigations of water/rock interaction processes in fluids and gases occurring in a seismically active area of the Irpinia-Basilicata Apennine region in southern Italy. IAEA.

Ballentine, C.J., Burgess, R., Marty, B., 2002. Tracing Fluid Origin, Transport and Interaction in the Crust. Reviews in Mineralogy and Geochemistry 47, 539–614. https://doi.org/10.2138/rmg.2002.47.13

Ballentine, C.J., Burnard, P.G., 2002. Production, Release and Transport of Noble Gases in the Continental Crust. Reviews in Mineralogy and Geochemistry 47, 481–538. https://doi.org/10.2138/rmg.2002.47.12

Barberio, M.D., Gori, F., Barbieri, M., Billi, A., Caracausi, A., De Luca, G., Franchini, S., Petitta, M., Doglioni, C., 2020. New observations in Central Italy of groundwater responses to the worldwide seismicity. Sci Rep 10, 17850. https://doi.org/10.1038/s41598-020-74991-0

Barchi, M., Amato, A., Cippitelli, G., Merlini, S., Montone, P., 2007. Extensional tectonics and seismicity in the axial zone of the Southern Apennines. Boll. Soc. Geol. It.(Italian Journal of Geosciences).

Beauducel, F., Lafon, D., Béguin, X., Saurel, J.-M., Bosson, A., Mallarino, D., Boissier, P., Brunet, C., Lemarchand, A., Anténor-Habazac, C., Nercessian, A., Fahmi, A.A., 2020. WebObs: The Volcano Observatories Missing Link Between Research and Real-Time Monitoring. Front. Earth Sci. 8, 48. https://doi.org/10.3389/feart.2020.00048

Benedetti, L., Tapponnier, P., King, G.C.P., Piccardi, L., 1998. Surface Rupture of the 1857 Southern Italian Earthquake? Terra Nova 10, 206–210. https://doi.org/10.1046/j.1365-3121.1998.00189.x

Ben-Menahem, A., Singh, S.J., 1981. Atmospheric and Water Waves and Companion Seismic Phenomena, in: Ben-Menahem, A., Singh, S.J. (Eds.), Seismic Waves and Sources. Springer, New York, NY, pp. 768–839. https://doi.org/10.1007/978-1-4612-5856-8_9

Benoit, J.P., McNutt, S.R., Barboza, V., 2003. Duration-amplitude distribution of volcanic tremor. J. Geophys. Res. 108. https://doi.org/10.1029/2001JB001520

Bernard, B.B., Brooks, J.M., Sackett, W.M., 1978. Light hydrocarbons in recent Texas continental shelf and slope sediments. Journal of Geophysical Research: Oceans 83, 4053–4061.

Biswas, A., Nath, B., Bhattacharya, P., Halder, D., Kundu, A.K., Mandal, U., Mukherjee, A., Chatterjee, D., Mörth, C.-M., Jacks, G., 2012. Hydrogeochemical contrast between brown and grey sand aquifers in shallow depth of Bengal Basin: Consequences for sustainable drinking water supply. Science of The Total Environment 431, 402–412. https://doi.org/10.1016/j.scitotenv.2012.05.031

Bobbio, A., Vassallo, M., Festa, G., 2009. A Local Magnitude Scale for Southern Italy. Bulletin of the Seismological Society of America 99, 2461–2470. https://doi.org/10.1785/0120080364

Bormann, P., Wielandt, E., 2013. Seismic Signals and Noise. New Manual of Seismological Observatory Practice 2 (NMSOP2) 7 mb, 62 pages. https://doi.org/10.2312/GFZ.NMSOP-2_CH4

Borok, V.K., 1959. On estimation of the displacement in an earthquake source and of source dimensions. Annals of Geophysics 12, 205–214. https://doi.org/10.4401/ag-5718

Bräuer, K., Geissler, W.H., Kämpf, H., Niedermannn, S., Rman, N., 2016. Helium and carbon isotope signatures of gas exhalations in the westernmost part of the Pannonian Basin (SE Austria/NE Slovenia): Evidence for active lithospheric mantle degassing. Chemical Geology 422, 60–70. https://doi.org/10.1016/j.chemgeo.2015.12.016

Brune, J.N., 1970. Tectonic stress and the spectra of seismic shear waves from earthquakes. Journal of Geophysical Research (1896-1977) 75, 4997–5009. https://doi.org/10.1029/JB075i026p04997

Brune, S., Heine, C., Clift, P.D., Pérez-Gussinyé, M., 2017. Rifted margin architecture and crustal rheology: Reviewing Iberia-Newfoundland, Central South Atlantic, and South China Sea. Marine and Petroleum Geology 79, 257–281. https://doi.org/10.1016/j.marpetgeo.2016.10.018

Buccione, R., Fortunato, E., Paternoster, M., Rizzo, G., Sinisi, R., Summa, V., Mongelli, G., 2021. Mineralogy and heavy metal assessment of the Pietra del Pertusillo reservoir sediments (Southern Italy). Environ Sci Pollut Res 28, 4857–4878. https://doi.org/10.1007/s11356-020-10829-6

Burnard, P., Bourlange, S., Henry, P., Geli, L., Tryon, M.D., Natal'in, B., Sengör, A.M.C., Özeren, M.S., Çagatay, M.N., 2012. Constraints on fluid origins and migration velocities along the Marmara Main Fault (Sea of Marmara, Turkey) using helium isotopes. Earth and Planetary Science Letters 341–344, 68–78. https://doi.org/10.1016/j.epsl.2012.05.042

Burrato, P., Valensise, G., 2008. Rise and Fall of a Hypothesized Seismic Gap: Source Complexity in the Mw 7.0 16 December 1857 Southern Italy Earthquake. Bulletin of the Seismological Society of America 98, 139–148. https://doi.org/10.1785/0120070094

Burton, T.G., Rifai, H.S., Hildenbrand, Z.L., Carlton, D.D., Fontenot, B.E., Schug, K.A., 2016. Elucidating hydraulic fracturing impacts on groundwater quality using a regional geospatial statistical modeling approach. Science of The Total Environment 545–546, 114–126. https://doi.org/10.1016/j.scitotenv.2015.12.084

Buttinelli, M., Improta, L., Bagh, S., Chiarabba, C., 2016. Inversion of inherited thrusts by wastewater injection induced seismicity at the Val d'Agri oilfield (Italy). Sci Rep 6, 37165. https://doi.org/10.1038/srep37165

Buttitta, D., Caracausi, A., Chiaraluce, L., Favara, R., Gasparo Morticelli, M., Sulli, A., 2020. Continental degassing of helium in an active tectonic setting (northern Italy): the role of seismicity. Sci Rep 10, 162. https://doi.org/10.1038/s41598-019-55678-7

Capasso, A., Migliaccio, M., 2005. Evoluzione del settore termale. Milano: FrancoAngeli.

Capasso, G., Inguaggiato, S., 1998. A simple method for the determination of dissolved gases in natural waters. An application to thermal waters from Vulcano Island. Applied Geochemistry 13, 631–642. https://doi.org/10.1016/S0883-2927(97)00109-1

Caracausi, A., Italiano, F., Martinelli, G., Paonita, A., Rizzo, A., 2005. Long-term geochemical monitoring and extensive/compressive phenomena: case study of the Umbria Region (Central Apennines, Italy). Annals of Geophysics 48. https://doi.org/10.4401/ag-3178

Caracausi, A., Italiano, F., Paonita, A., Rizzo, A., Nuccio, P.M., 2003. Evidence of deep magma degassing and ascent by geochemistry of peripheral gas emissions at Mount Etna (Italy): Assessment of the magmatic reservoir pressure. Journal of Geophysical Research: Solid Earth 108.

Caracausi, A., Martelli, M., Nuccio, P.M., Paternoster, M., Stuart, F.M., 2013. Active degassing of mantle-derived fluid: A geochemical study along the Vulture line, southern Apennines (Italy). Journal of Volcanology and Geothermal Research 253, 65–74. https://doi.org/10.1016/j.jvolgeores.2012.12.005

Caracausi, A., Paternoster, M., 2015. Radiogenic helium degassing and rock fracturing: A case study of the southern Apennines active tectonic region. J. Geophys. Res. Solid Earth 120, 2200–2211. https://doi.org/10.1002/2014JB011462

Caracausi, A., Paternoster, M., Nuccio, P.M., 2015. Mantle CO2 degassing at Mt. Vulture volcano (Italy): Relationship between CO2 outgassing of volcanoes and the time of their last eruption. Earth and Planetary Science Letters 411, 268–280. https://doi.org/10.1016/j.epsl.2014.11.049

Caracausi, A., Sulli, A., 2019. Outgassing of Mantle Volatiles in Compressional Tectonic Regime Away From Volcanism: The Role of Continental Delamination. Geochem. Geophys. Geosyst. 20, 2007–2020. https://doi.org/10.1029/2018GC008046

Celico, P., 1988. Prospezioni idrogeologiche. Liguori Editore, Napoli.

Cello, G., Gambini, R., Mazzoli, S., Read, A., Tondi, E., Zucconi, V., 2000. Fault zone characteristics and scaling properties of the Val d'Agri Fault System (Southern Apennines, Italy). Journal of Geodynamics 29, 293–307. https://doi.org/10.1016/S0264-3707(99)00043-5

Cello, G., Mazzoli, S., 1998. Apennine tectonics in southern Italy: a review. Journal of Geodynamics 27, 191–211. https://doi.org/10.1016/S0264-3707(97)00072-0

Cello, G., Tondi, E., Micarelli, L., Mattioni, L., 2003. Active tectonics and earthquake sources in the epicentral area of the 1857 Basilicata earthquake (southern Italy). Journal of Geodynamics 36, 37–50. https://doi.org/10.1016/S0264-3707(03)00037-1

Cerling, T.E., Solomon, D.K., Quade, J., Bowman, J.R., 1991. On the isotopic composition of carbon in soil carbon dioxide. Geochimica et Cosmochimica Acta 55, 3403–3405. https://doi.org/10.1016/0016-7037(91)90498-T

Charfi, S., Zouari, K., Feki, S., Mami, E., 2013. Study of variation in groundwater quality in a coastal aquifer in north-eastern Tunisia using multivariate factor analysis. Quaternary International, Terrestrial archives of the Mediterranean 302, 199–209. https://doi.org/10.1016/j.quaint.2012.11.002

Chiarabba, C., De Gori, P., Cattaneo, M., Spallarossa, D., Segou, M., 2018. Faults Geometry and the Role of Fluids in the 2016–2017 Central Italy Seismic Sequence. Geophysical Research Letters 45, 6963–6971. https://doi.org/10.1029/2018GL077485

Chiarabba, C., Jovane, L., DiStefano, R., 2005. A new view of Italian seismicity using 20 years of instrumental recordings. Tectonophysics 395, 251–268. https://doi.org/10.1016/j.tecto.2004.09.013

Chiodini, G., Cardellini, C., Amato, A., Boschi, E., Caliro, S., Frondini, F., Ventura, G., 2004. Carbon dioxide Earth degassing and seismogenesis in central and southern Italy. Geophysical Research Letters 31. https://doi.org/10.1029/2004GL019480

Chiodini, G., Cardellini, C., Di Luccio, F., Selva, J., Frondini, F., Caliro, S., Rosiello, A., Beddini, G., Ventura, G., 2020. Correlation between tectonic CO_2 Earth degassing and seismicity is revealed by a 10-year record in the Apennines, Italy. Sci. Adv. 6, eabc2938. https://doi.org/10.1126/sciadv.abc2938

Chiodini, G., Frondini, F., Cardellini, C., Parello, F., Peruzzi, L., 2000. Rate of diffuse carbon dioxide Earth degassing estimated from carbon balance of regional aquifers: The case of central Apennine, Italy. J. Geophys. Res. 105, 8423–8434. https://doi.org/10.1029/1999JB900355

Chiodini, G., Frondini, F., Kerrick, D.M., Rogie, J., Parello, F., Peruzzi, L., Zanzari, A.R., 1999. Quantification of deep CO_2 fluxes from Central Italy. Examples of carbon balance for regional aquifers and of soil diffuse degassing. Chemical Geology 159, 205–222. https://doi.org/10.1016/S0009-2541(99)00030-3

Chiodini, G., Giudicepietro, F., Vandemeulebrouck, J., Aiuppa, A., Caliro, S., De Cesare, W., Tamburello, G., Avino, R., Orazi, M., D'Auria, L., 2017. Fumarolic tremor and geochemical signals during a volcanic unrest. Geology 45, 1131–1134. https://doi.org/10.1130/G39447.1

Chiodini, G., Granieri, D., Avino, R., Caliro, S., Costa, A., Minopoli, C., Vilardo, G., 2010. Non-volcanic CO_2 Earth degassing: Case of Mefite d'Ansanto (southern Apennines), Italy. Geophys. Res. Lett. 37, n/a-n/a. https://doi.org/10.1029/2010GL042858

Cinque, A., Patacca, E., Scandone, P., & Tozzi, M. (1993). Quaternary kinematic evolution of the Southern Appennines. Relationships between surface geological features and deep lithospheric structures. Annals of Geophysics, 36(2).

Ciotoli, G., Bigi, S., Tartarello, C., Sacco, P., Lombardi, S., Ascione, A., Mazzoli, S., 2014. Soil gas distribution in the main coseismic surface rupture zone of the 1980, M s = 6.9, Irpinia earthquake (southern Italy). J. Geophys. Res. Solid Earth 119, 2440–2461. https://doi.org/10.1002/2013JB010508

Ciotoli, G., Lombardi, S., Annunziatellis, A., 2007. Geostatistical analysis of soil gas data in a high seismic intermontane basin: Fucino Plain, central Italy. J. Geophys. Res. 112, B05407. https://doi.org/10.1029/2005JB004044

Clark, I and Fritz, P. (1997) Environmental Isotopes in Hydrogeology. CRC Press, New York, 328 p.

Clayton, R.N., Friedman, I., Graf, D.L., Mayeda, T.K., Meents, W.F., Shimp, N.F., 1966. The origin of saline formation waters: 1. Isotopic composition. Journal of Geophysical Research (1896-1977) 71, 3869–3882. https://doi.org/10.1029/JZ071i016p03869

Clinton, J.F., Heaton, T.H., 2002. Potential Advantages of a Strong-motion Velocity Meter over a Strong-motion Accelerometer. Seismological Research Letters 73, 332–342. https://doi.org/10.1785/gssrl.73.3.332

Colella, A., Lapenna, V., and Rizzo, E., 2003. La struttura sepolta del bacino dell'Alta Val d'Agri (Pleistocene, Basilicata). Le risorse idriche sotterranee dell'Alta Val d'Agri. Ed. Collana Editoriale di Studi e Ricerche dell'Autorità Interregionale di Bacino della Basilicata. n. 3. Potenza

Coleman, T.F., Li, Y., 1996. An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds. SIAM J. Optim. 6, 418–445. https://doi.org/10.1137/0806023

Correale, A., Martelli, M., Paonita, A., Rizzo, A., Brusca, L., Scribano, V., 2012. New evidence of mantle heterogeneity beneath the Hyblean Plateau (southeast Sicily, Italy) as inferred from noble gases and geochemistry of ultramafic xenoliths. Lithos 132–133, 70–81. https://doi.org/10.1016/j.lithos.2011.11.007

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inform. Theory 13, 21–27. https://doi.org/10.1109/TIT.1967.1053964

Craig, H., 1961. Isotopic Variations in Meteoric Waters. Science 133, 1702–1703. https://doi.org/10.1126/science.133.3465.1702

Dahm, T., Cesca, S., Hainzl, S., Braun, T., Krüger, F., 2015. Discrimination between induced, triggered, and natural earthquakes close to hydrocarbon reservoirs: A probabilistic approach based on the modeling of depletion-induced stress changes and seismological source parameters. J. Geophys. Res. Solid Earth 120, 2491–2509. https://doi.org/10.1002/2014JB011778

de Lorenzo, S., Zollo, A., and Mongelli, F.,2001. Source parameters and three-dimensional attenuation structure from the inversion of microearthquake pulse width data: Qp imaging and inferences on the thermal state of the Campi Flegrei caldera (southern Italy), J. Geophys. Res., 106 (B8), 16265– 16286, doi: 10.1029/2000JB900462.

de Lorenzo, S., Zollo, A., and Zito, G., 2010. Source, attenuation, and site parameters of the 1997 Umbria-Marche seismic sequence from the inversion of P wave spectra: A comparison between constant QP and frequency-dependent QP models, J. Geophys. Res., 115, B09306, doi:10.1029/2009JB007004.

Deines, P., 2002. The carbon isotope geochemistry of mantle xenoliths. Earth-Science Reviews 58, 247–278. https://doi.org/10.1016/S0012-8252(02)00064-8

Deines, P., Langmuir, D., Harmon, R.S., 1974. Stable carbon isotope ratios and the existence of a gas phase in the evolution of carbonate ground waters. Geochimica et Cosmochimica Acta 38, 1147–1164. https://doi.org/10.1016/0016-7037(74)90010-6

Del Pezzo, E., Iannaccone, G., Martini, M., Scarpa, R., 1983. The 23 November 1980 southern Italy earthquake. Bulletin of the Seismological Society of America 73, 187–200. https://doi.org/10.1785/BSSA0730010187

Delle Donne, D., Ripepe, M., Lacanna, G., Tamburello, G., Bitetto, M., Aiuppa, A., 2016. Gas mass derived by infrasound and UV cameras: Implications for mass flow rate. Journal of Volcanology and Geothermal Research 325, 169–178. https://doi.org/10.1016/j.jvolgeores.2016.06.015

Di Luccio, F., Chiodini, G., Caliro, S., Cardellini, C., Convertito, V., Pino, N.A., Tolomei, C., Ventura, G., 2018. Seismic signature of active intrusions in mountain chains. Sci. Adv. 4, e1701825. https://doi.org/10.1126/sciadv.1701825

Di Luccio, F., Ventura, G., Di Giovambattista, R., Piscini, A., Cinti, F.R., 2010. Normal faults and thrusts reactivated by deep fluids: The 6 April 2009 Mw 6.3 L'Aquila earthquake, central Italy. Journal of Geophysical Research: Solid Earth 115. https://doi.org/10.1029/2009JB007190

Doglioni, C., Harabaglia, P., Merlini, S., Mongelli, F., Peccerillo, A., Piromallo, C., 1999. Orogens and slabs vs. their direction of subduction. Earth-Science Reviews 45, 167–208. https://doi.org/10.1016/S0012-8252(98)00045-2

Dunai, T.J., Baur, H., 1995. Helium, neon, and argon systematics of the European subcontinental mantle: Implications for its geochemical evolution. Geochimica et Cosmochimica Acta 59, 2767–2783. https://doi.org/10.1016/0016-7037(95)00172-V

Ellsworth, W.L., 2013. Injection-Induced Earthquakes. Science 341, 1225942–1225942. https://doi.org/10.1126/science.1225942

ENI, 1972. Acque Dolci Sotterranee. C. Colombo, ed. Milano.

Ester, M., Kriegel, H.-P., Xu, X., 1966. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise 6.

Etiope, G., Feyzullayev, A., Milkov, A.V., Waseda, A., Mizobe, K., Sun, C.H., 2009. Evidence of subsurface anaerobic biodegradation of hydrocarbons and potential secondary methanogenesis in terrestrial mud volcanoes. Marine and Petroleum Geology 26, 1692–1703. https://doi.org/10.1016/j.marpetgeo.2008.12.002

Etiope, G., Klusman, R.W., 2002. Geologic emissions of methane to the atmosphere. Chemosphere 49, 777–789. https://doi.org/10.1016/S0045-6535(02)00380-6

Etiope, G., Martinelli, G., Caracausi, A., Italiano, F., 2007. Methane seeps and mud volcanoes in Italy: Gas origin, fractionation and emission to the atmosphere. Geophys. Res. Lett. 34, L14303. https://doi.org/10.1029/2007GL030341

Ferranti, L., Burrato, P., Pepe, F., Santoro, E., Mazzella, M.E., Morelli, D., Passaro, S., Vannucci, G., 2014. An active oblique-contractional belt at the transition between the Southern Apennines and Calabrian Arc: The Amendolara Ridge, Ionian Sea, Italy: Active contraction Amendolara Ridge. Tectonics 33, 2169–2194. https://doi.org/10.1002/2014TC003624

Festa, G., Adinolfi, G.M., Caruso, A., Colombelli, S., De Landro, G., Elia, L., Emolo, A., Picozzi, M., Scala, A., Carotenuto, F., Gammaldi, S., Iaccarino, A.G., Nazeri, S., Riccio, R., Russo, G., Tarantino, S., Zollo, A., 2021. Insights into Mechanical Properties of the 1980 Irpinia Fault System from the Analysis of a Seismic Sequence. Geosciences 11, 28. https://doi.org/10.3390/geosciences11010028

Fischer, H., Wagenbach, D., Kipfstuhl, J., 1998. Sulfate and nitrate firn concentrations on the Greenland ice sheet: 2. Temporal anthropogenic deposition changes. Journal of Geophysical Research: Atmospheres 103, 21935–21942. https://doi.org/10.1029/98JD01886

Fix, E. and Hodges, J.L. (1951) Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field.

Fontenot, B.E., Hunt, L.R., Hildenbrand, Z.L., Carlton Jr., D.D., Oka, H., Walton, J.L., Hopkins, D., Osorio, A., Bjorndal, B., Hu, Q.H., Schug, K.A., 2013. An Evaluation of Water Quality in Private Drinking Water Wells Near Natural Gas Extraction Sites in the Barnett Shale Formation. Environ. Sci. Technol. 47, 10032–10040. https://doi.org/10.1021/es4011724

Frémont, M.-J., Malone, S.D., 1987. High precision relative locations of earthquakes at Mount St. Helens, Washington. Journal of Geophysical Research: Solid Earth 92, 10223–10236. https://doi.org/10.1029/JB092iB10p10223

Frondini, F., Cardellini, C., Caliro, S., Beddini, G., Rosiello, A., Chiodini, G., 2019. Measuring and interpreting CO_2 fluxes at regional scale: the case of the Apennines, Italy. Journal of the Geological Society 176, 408–416. https://doi.org/10.1144/jgs2017-169

Garcia-Aristizabal, A., Caciagli, M., Selva, J., 2016. Considering uncertainties in the determination of earthquake source parameters from seismic spectra. Geophys. J. Int. 207, 691–701. https://doi.org/10.1093/gji/ggw303

Gat, J., Carmi, I., 1987. Effect of climate changes on the precipitation patterns and isotopic composition of water in a climate transition zone: Case of the Eastern Mediterranean Sea area. Proceedings of the Vancouver Symposium 168.

Gautheron, C., Moreira, M., 2002. Helium signature of the subcontinental lithospheric mantle. Earth and Planetary Science Letters 199, 39–47. https://doi.org/10.1016/S0012-821X(02)00563-0

Gautheron, C., Moreira, M., Allègre, C., 2005. He, Ne and Ar composition of the European lithospheric mantle. Chemical Geology 217, 97–112. https://doi.org/10.1016/j.chemgeo.2004.12.009

Geiger, L., 1912. Probability method for the determination of earthquake epicenters from the arrival time only.Bull. St. Louis. Univ8, 60–71.

Geller, R.J., Mueller, C.S., 1980. Four similar earthquakes in central California. Geophys. Res. Lett. 7, 821–824. https://doi.org/10.1029/GL007i010p00821

Giano, S., 2011. Quaternary alluvial fan systems of the Agri intermontane basin (southern Italy): tectonic and climatic controls. Geologica Carpathica 62, 65–76. https://doi.org/10.2478/v10096-011-0006-y

Giano, S.I., Maschio, L., Alessio, M., Ferranti, L., Improta, S., Schiattarella, M., 2000. Radiocarbon dating of active faulting in the Agri high valley, southern Italy. Journal of Geodynamics 29, 371–386. https://doi.org/10.1016/S0264-3707(99)00058-7

Gibbons, S.J., Ringdal, F., 2006. The detection of low magnitude seismic events using array-based waveform correlation. Geophysical Journal International 165, 149–166. https://doi.org/10.1111/j.1365-246X.2006.02865.x

Giggenbach, W.F., 1996. Chemical Composition of Volcanic Gases, in: Scarpa, R., Tilling, R.I. (Eds.), Monitoring and Mitigation of Volcano Hazards. Springer, Berlin, Heidelberg, pp. 221–256. https://doi.org/10.1007/978-3-642-80087-0_7

Giggenbach, W.F., Sano, Y., Wakita, H., 1993. Isotopic composition of helium, and CO_2 and CH_4 contents in gases produced along the New Zealand part of a convergent plate boundary. Geochimica et Cosmochimica Acta 57, 3427–3455. https://doi.org/10.1016/0016-7037(93)90549-C

Giocoli, A., Stabile, T.A., Adurno, I., Perrone, A., Gallipoli, M.R., Gueguen, E., Norelli, E., Piscitelli, S., 2015. Geological and geophysical characterization of the southeastern side of the High Agri Valley (southern Apennines, Italy). Nat. Hazards Earth Syst. Sci. 15, 315–323. https://doi.org/10.5194/nhess-15-315-2015

Giudicepietro, F., Chiodini, G., Caliro, S., De Cesare, W., Esposito, A.M., Galluzzo, D., Lo Bascio, D., Macedonio, G., Orazi, M., Ricciolino, P., Vandemeulebrouck, J., 2019. Insight Into Campi Flegrei Caldera Unrest Through Seismic Tremor Measurements at Pisciarelli Fumarolic Field. Geochem. Geophys. Geosyst. 20, 5544–5555. https://doi.org/10.1029/2019GC008610

Gomberg, J.S., Shedlock, K.M., Roecker, S.W., 1990. The effect of S-wave arrival times on the accuracy of hypocenter estimation. Bulletin of the Seismological Society of America 80, 1605–1628. https://doi.org/10.1785/BSSA08006A1605

Gordeev, E.I., 1992. Modelling of volcanic tremor wave fields. Journal of Volcanology and Geothermal Research 51, 145–160. https://doi.org/10.1016/0377-0273(92)90065-L

Got, J.-L., Fréchet, J., Klein, F.W., 1994. Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea. Journal of Geophysical Research: Solid Earth 99, 15375–15386. https://doi.org/10.1029/94JB00577

Graham, D.W., 2002. Noble Gas Isotope Geochemistry of Mid-Ocean Ridge and Ocean Island Basalts: Characterization of Mantle Source Reservoirs. Reviews in Mineralogy and Geochemistry 47, 247–317. https://doi.org/10.2138/rmg.2002.47.8

Grassa, F., Capasso, G., Favara, R., Inguaggiato, S., 2006. Chemical and Isotopic Composition of Waters and Dissolved Gases in Some Thermal Springs of Sicily and Adjacent Volcanic Islands, Italy. Pure appl. geophys. 163, 781–807. https://doi.org/10.1007/s00024-006-0043-0

Grigoli, F., Cesca, S., Priolo, E., Rinaldi, A.P., Clinton, J.F., Stabile, T.A., Dost, B., Fernandez, M.G., Wiemer, S., Dahm, T., 2017. Current challenges in monitoring, discrimination, and management of induced seismicity related to underground industrial activities: A European perspective: CHALLENGES IN INDUCED SEISMICITY. Rev. Geophys. 55, 310–340. https://doi.org/10.1002/2016RG000542

Gueguen, E., Bentivenga, M., Colaiacovo, R., Margiotta, S., Summa, V., Adurno, I., 2015. The Verdesca landslide in the Agri Valley (Basilicata, southern Italy): a new geological and geomorphological framework. Nat. Hazards Earth Syst. Sci. 15, 2585–2595. https://doi.org/10.5194/nhess-15-2585-2015

Gutenberg, B., Richter, C.F., 1956. Earthquake magnitude, intensity, energy, and acceleration: (Second paper). Bulletin of the Seismological Society of America 46, 105–145. https://doi.org/10.1785/BSSA0460020105

Hager, B.H., Dieterich, J., Frohlich, C., Juanes, R., Mantica, S., Shaw, J.H., Bottazzi, F., Caresani, F., Castineira, D., Cominelli, A., Meda, M., Osculati, L., Petroselli, S., Plesch, A., 2021. A process-based approach to understanding and managing triggered seismicity. Nature 595, 684–689. https://doi.org/10.1038/s41586-021-03668-z

Hahm, D., Hilton, D.R., Cho, M., Wei, H., Kim, K.-R., 2008. Geothermal He and CO2 variations at Changbaishan intra-plate volcano (NE China) and the nature of the sub-continental lithospheric mantle. Geophysical Research Letters 35.

Halldórsson, S.A., Hilton, D.R., Troll, V.R., Fischer, T.P., 2013. Resolving volatile sources along the western Sunda arc, Indonesia. Chemical Geology 339, 263–282. https://doi.org/10.1016/j.chemgeo.2012.09.042

Hanks, T.C., 1982. fmax. Bulletin of the Seismological Society of America 72, 1867–1879. https://doi.org/10.1785/BSSA07206A1867

Hanks, T.C., Kanamori, H., 1979. A moment magnitude scale. J. Geophys. Res. 84, 2348. https://doi.org/10.1029/JB084iB05p02348

Hardebeck, J., Husen, S., 2010. Earthquake location accuracy. https://doi.org/10.5078/CORSSA-55815573

Hardebeck, J.L., 2002. A New Method for Determining First-Motion Focal Mechanisms. Bulletin of the Seismological Society of America 92, 2264–2276. https://doi.org/10.1785/0120010200

Hoefs, J., 1987. Variations of Stable Isotope Ratios in Nature, in: Hoefs, J. (Ed.), Stable Isotope Geochemistry, Minerals and Rocks. Springer, Berlin, Heidelberg, pp. 66–197. https://doi.org/10.1007/978-3-662-09998-8_3

Hoefs, J., 2015. Isotope Fractionation Processes of Selected Elements, in: Hoefs, J. (Ed.), Stable Isotope Geochemistry. Springer International Publishing, Cham, pp. 47–190. https://doi.org/10.1007/978-3-319-19716-6_2

Hoffmann, G., Jouzel, J., Masson, V., 2000. Stable water isotopes in atmospheric general circulation models. Hydrological Processes 14, 1385–1406. https://doi.org/10.1002/1099-1085(20000615)14:8<1385::AID-HYP989>3.0.CO;2-1

Holocher, J., Peeters, F., Aeschbach-Hertig, W., Hofer, M., Brennwald, M., Kinzelbach, W., Kipfer, R., 2002. Experimental investigations on the formation of excess air in quasi-saturated porous media. Geochimica et Cosmochimica Acta 66, 4103–4117. https://doi.org/10.1016/S0016-7037(02)00992-4

Hsu, S.-K., Wang, S.-Y., Liao, Y.-C., Yang, T.F., Jan, S., Lin, J.-Y., Chen, S.-C., 2013. Tide-modulated gas emissions and tremors off SW Taiwan. Earth and Planetary Science Letters 369–370, 98–107. https://doi.org/10.1016/j.epsl.2013.03.013

Hunt, J.A., Zafu, A., Mather, T.A., Pyle, D.M., Barry, P.H., 2017. Spatially Variable CO2 Degassing in the Main Ethiopian Rift: Implications for Magma Storage, Volatile Transport, and Rift-Related Emissions. Geochem. Geophys. Geosyst. 18, 3714–3737. https://doi.org/10.1002/2017GC006975

Hurst, A.W., Sherburn, S., 1993. Volcanic tremor at Ruapehu: Characteristics and implications for the resonant source. null 36, 475–485. https://doi.org/10.1080/00288306.1993.9514593

Husen, S., Kissling, E., 2001. Postseismic fluid flow after the large subduction earthquake of Antofagasta, Chile. Geol 29, 847. https://doi.org/10.1130/0091-7613(2001)029<0847:PFFATL>2.0.CO;2

Hutton, L.K., Boore, D.M., 1987. The ML scale in Southern California. Bulletin of the Seismological Society of America 77, 2074–2094. https://doi.org/10.1785/BSSA0770062074

Iannaccone, G., Zollo, A., Elia, L., Convertito, V., Satriano, C., Martino, C., Festa, G., Lancieri, M., Bobbio, A., Stabile, T.A., Vassallo, M., Emolo, A., 2010. A prototype system for earthquake early warning and alert management in southern Italy. Bull Earthquake Eng 8, 1105–1129. https://doi.org/10.1007/s10518-009-9131-8

Improta, L., Bagh, S., De Gori, P., Valoroso, L., Pastori, M., Piccinini, D., Buttinelli, M., 2017. Reservoir structure and wastewater-induced seismicity at the Val d'Agri oilfield (Italy) shown by three-dimensional Vp and Vp/Vs local earthquake tomography. Journal of Geophysical Research: Solid Earth, 122, 9050– 9082. https://doi.org/10.1002/2017JB014725

Improta, L., Bonagura, M., Capuano, P., Iannaccone, G., 2003. An integrated geophysical investigation of the upper crust in the epicentral area of the 1980, Ms=6.9, Irpinia earthquake (Southern Italy). Tectonophysics 361, 139–169. https://doi.org/10.1016/S0040-1951(02)00588-7

Improta, L., De Gori, P., Chiarabba, C., 2014. New insights into crustal structure, Cenozoic magmatism, CO2 degassing, and seismogenesis in the southern Apennines and Irpinia region from local earthquake tomography. Journal of Geophysical Research: Solid Earth 119, 8283–8311. https://doi.org/10.1002/2013JB010890

Improta, L., Ferranti, L., De Martini, P.M., Piscitelli, S., Bruno, P.P., Burrato, P., Civico, R., Giocoli, A., Iorio, M., D'Addezio, G., Maschio, L., 2010. Detecting young, slow-slipping active faults by geologic and multidisciplinary high-resolution geophysical investigations: A case study from the Apennine seismic belt, Italy. J. Geophys. Res. 115, B11307. https://doi.org/10.1029/2010JB000871

Improta, L., Valoroso, L., Piccinini, D., Chiarabba, C., 2015. A detailed analysis of wastewater-induced seismicity in the Val d'Agri oil field (Italy): wastewater induced seismicity in Italy. Geophys. Res. Lett. 42, 2682–2690. https://doi.org/10.1002/2015GL063369

Italiano, F., Martelli, M., Martinelli, G., Nuccio, P.M., 2000. Geochemical evidence of melt intrusions along lithospheric faults of the Southern Apennines, Italy: Geodynamic and seismogenic implications. J. Geophys. Res. 105, 13569–13578. https://doi.org/10.1029/2000JB900047

Italiano, F., Martelli, M., Martinelli, G., Nuccio, P.M., Paternoster, M., 2001. Significance of earthquake-related anomalies in fluids of Val D'Agri (southern Italy). Terra Nova 13, 249–257. https://doi.org/10.1046/j.1365-3121.2001.00346.x

Italiano, F., Martinelli, G., Bonfanti, P., Caracausi, A., 2009. Long-term (1997-2007) geochemical monitoring of gases from the Umbria-Marche region. Tectonophysics 476, 282–296. https://doi.org/10.1016/j.tecto.2009.02.040

Italiano, F., Yuce, G., Uysal, I.T., Gasparon, M., Morelli, G., 2014. Insights into mantle-type volatiles contribution from dissolved gases in artesian waters of the Great Artesian Basin, Australia. Chemical Geology 378–379, 75–88. https://doi.org/10.1016/j.chemgeo.2014.04.013

Jackson, R.B., Vengosh, A., Darrah, T.H., Warner, N.R., Down, A., Poreda, R.J., Osborn, S.G., Zhao, K., Karr, J.D., 2013. Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction. Proc Natl Acad Sci U S A 110, 11250–11255. https://doi.org/10.1073/pnas.1221635110

Jalali, M., 2009. Geochemistry characterization of groundwater in an agricultural area of Razan, Hamadan, Iran. Environ Geol 56, 1479–1488. https://doi.org/10.1007/s00254-008-1245-9

Jenden, P.D., Drazan, D.J., Kaplan, I.R., 1993. Mixing of Thermogenic Natural Gases in Northern Appalachian Basin1. AAPG Bulletin 77, 980–998. https://doi.org/10.1306/BDFF8DBC-1718-11D7-8645000102C1865D

Juanes, R., Jha, B., Hager, B.H., Shaw, J.H., Plesch, A., Astiz, L., Dieterich, J.H., Frohlich, C., 2016. Were the May 2012 Emilia-Romagna earthquakes induced? A coupled flow-geomechanics modeling assessment. Geophys. Res. Lett. 43, 6891–6897. https://doi.org/10.1002/2016GL069284

Jurkevics, A., 1988. Polarization analysis of three-component array data. Bulletin of the Seismological Society of America 78, 1725–1743. https://doi.org/10.1785/BSSA0780051725

Kanamori, H., Anderson, D.L., 1975. Theoretical basis of some empirical relations in seismology. Bulletin of the Seismological Society of America 65, 1073–1095. https://doi.org/10.1785/BSSA0650051073

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V., 2017. Machine Learning for the Geosciences: Challenges and Opportunities. arXiv:1711.04708 [physics].

Keen, C.G., Montgomery, J., Mowat, W.M.H., Mullard, J.E., Platt, D.C., 1965. British seismometer array recording systems. Radio Electron. Eng. UK 30, 297. https://doi.org/10.1049/ree.1965.0106

Keranen, K.M., Savage, H.M., Abers, G.A., Cochran, E.S., 2013. Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence. Geology 41, 699–702. https://doi.org/10.1130/G34045.1

Kim, J., 2012. Department of electrical engineering and computer science. SIGDA Newsl. 20, 91. https://doi.org/10.1145/378886.380416

Kjartansson, E., 1979. Constant Q-wave propagation and attenuation. Journal of Geophysical Research: Solid Earth 84, 4737–4748. https://doi.org/10.1029/JB084iB09p04737

Knott, S.D., 1987. The Liguride Complex of Southern Italy —a Cretaceous to Paleogene accretionary wedge. Tectonophysics 142, 217–226. https://doi.org/10.1016/0040-1951(87)90124-7

Kong, Q., Allen, R.M., Schreier, L., 2016. MyShake: Initial observations from a global smartphone seismic network. Geophys. Res. Lett. 43, 9588–9594. https://doi.org/10.1002/2016GL070955

Kong, Q., Trugman, D.T., Ross, Z.E., Bianco, M.J., Meade, B.J., Gerstoft, P., 2019. Machine Learning in Seismology: Turning Data into Insights. Seismological Research Letters 90, 3–14. https://doi.org/10.1785/0220180259

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., Wassermann, J., 2015. ObsPy: a bridge for seismology into the scientific Python ecosystem. Comput. Sci. Disc. 8, 014003. https://doi.org/10.1088/1749-4699/8/1/014003

Ktenidou, O.-J., Cotton, F., Abrahamson, N.A., Anderson, J.G., 2014. Taxonomy of : A Review of Definitions and Estimation Approaches Targeted to Applications. Seismological Research Letters 85, 135–146. https://doi.org/10.1785/0220130027

Laughrey, C.D., Baldassare, F.J., 2003. Some applications of isotope geochemistry for determining sources of stray carbon dioxide gas. Environ. Geosci. 10, 107–122. https://doi.org/10.1306/eg100303003

Lee, H., Muirhead, J.D., Fischer, T.P., Ebinger, C.J., Kattenhorn, S.A., Sharp, Z.D., Kianji, G., 2016. Massive and prolonged deep carbon emissions associated with continental rifting. Nature Geosci 9, 145–149. https://doi.org/10.1038/ngeo2622

Lee, J.-Y., Marti, K., Severinghaus, J.P., Kawamura, K., Yoo, H.-S., Lee, J.B., Kim, J.S., 2006. A redetermination of the isotopic abundances of atmospheric Ar. Geochimica et Cosmochimica Acta 70, 4507–4512. https://doi.org/10.1016/j.gca.2006.06.1563

Lee, W.H., Lahr, J.C., 1975. HYPO71 (revised; a computer program for determining hypocenter, magnitude, and first motion pattern of local earthquakes (Report No. 75–311), Open-File Report. https://doi.org/10.3133/ofr75311

Lisio, A., Russo, F., Sisto, M., 2014. La Mefite nella Valle d'Ansanto (Irpinia, Campania): il valore paradigmatico di un geoarcheosito. Geologia dell'Ambiente - Italian Magazine of Environmental Geology 2–7.

Lohman, R.B., McGuire, J.J., 2007. Earthquake swarms driven by aseismic creep in the Salton Trough, California: OBSIDIAN BUTTES SWARM. J. Geophys. Res. 112. https://doi.org/10.1029/2006JB004596

Lomax, A., 2008. Location of the Focus and Tectonics of the Focal Region of the California Earthquake of 18 April 1906. Bulletin of the Seismological Society of America 98, 846–860. https://doi.org/10.1785/0120060405

Lomax, A., Curtis, A., 2001. Fast, probabilistic earthquake location in 3D models using oct-tree importance sampling 1.

Lomax, A., Virieux, J., Volant, P., Berge-Thierry, C., 2000. Probabilistic Earthquake Location in 3D and Layered Models, in: Thurber, C.H., Rabinowitz, N. (Eds.), Advances in Seismic Event Location, Modern Approaches in Geophysics. Springer Netherlands, Dordrecht, pp. 101–134. https://doi.org/10.1007/978-94-015-9536-0_5

Longinelli, A., Selmo, E., 2003. Isotopic composition of precipitation in Italy: a first overall map. Journal of Hydrology 270, 75–88. https://doi.org/10.1016/S0022-1694(02)00281-0

Lunedei, E., Malischewsky, P., 2015. A Review and Some New Issues on the Theory of the H/V Technique for Ambient Vibrations, in: Ansal, A. (Ed.), Perspectives on European Earthquake Engineering and Seismology, Geotechnical, Geological and Earthquake Engineering. Springer International Publishing, Cham, pp. 371–394. https://doi.org/10.1007/978-3-319-16964-4_15

Madariaga, R., 1976. Dynamics of an expanding circular fault. Bulletin of the Seismological Society of America 66, 639–666. https://doi.org/10.1785/BSSA0660030639

Mallet, R., 1862. Great Neapolitan Earthquake of 1857: The First Principles of Observational Seismology as Developed in the Report to the Royal Society of London of the Expedition Made by Command of the Society Into the Interior of the Kingdom of Naples, to Investigate the Circumstances of the Great Earthquake of December 1857. Chapman and Hall.

Martelli, M., Nuccio, P.M., Stuart, F.M., Burgess, R., Ellam, R.M., Italiano, F., 2004. Helium–strontium isotope constraints on mantle evolution beneath the Roman Comagmatic Province, Italy. Earth and Planetary Science Letters 224, 295–308. https://doi.org/10.1016/j.epsl.2004.05.025

Martorelli, E., Italiano, F., Ingrassia, M., Macelloni, L., Bosman, A., Conte, A.M., Beaubien, S.E., Graziani, S., Sposato, A., Chiocci, F.L., 2016. Evidence of a shallow water submarine hydrothermal field off Zannone Island from morphological and geochemical characterization: Implications for Tyrrhenian Sea Quaternary volcanism: A New Hydrothermal Field Off Zannone Island. J. Geophys. Res. Solid Earth 121, 8396–8414. https://doi.org/10.1002/2016JB013103

Marty, B. and Dauphas, N. (2003). The nitrogen record of crust-mantle interaction and mantle convection from Archean to present. Earth and Planetary Science Letters 206: 397-410.

Marty, B., Jambon, A., 1987. C3He in volatile fluxes from the solid Earth: implications for carbon geodynamics. Earth and Planetary Science Letters 83, 16–26. https://doi.org/10.1016/0012-821X(87)90047-1

Maschio, L., Ferranti, L., Burrato, P., 2005. Active extension in Val d'Agri area, Southern Apennines, Italy: implications for the geometry of the seismogenic belt. Geophysical Journal International 162, 591–609. https://doi.org/10.1111/j.1365-246X.2005.02597.x

Matoza, R.S., Fee, D., Neilsen, T.B., Gee, K.L., Ogden, D.E., 2013. Aeroacoustics of volcanic jets: Acoustic power estimation and jet velocity dependence. Journal of Geophysical Research: Solid Earth 118, 6269–6284. https://doi.org/10.1002/2013JB010303

Matthess, G. (1982) The Properties of Groundwater. Department of Environmental Science, John Wiley and Sons Inc., New York, 406 p

Mazzoli, S., Barkham, S., Cello, G., Gambini, R., Mattioni, L., Shiner, P., Tondi, E., 2001. Reconstruction of continental margin architecture deformed by the contraction of the Lagonegro Basin, southern Apennines, Italy. Journal of the Geological Society 158, 309–319. https://doi.org/10.1144/jgs.158.2.309

McCollom, T.M., Seewald, J.S., 2007. Abiotic Synthesis of Organic Compounds in Deep-Sea Hydrothermal Environments. Chem. Rev. 107, 382–401. https://doi.org/10.1021/cr0503660

McGarr, A., Simpson, D., Seeber, L., 2002. 40 Case histories of induced and triggered seismicity, in: International Geophysics. Elsevier, pp. 647–661. https://doi.org/10.1016/S0074-6142(02)80243-1

McKee, K., Fee, D., Yokoo, A., Matoza, R.S., Kim, K., 2017. Analysis of gas jetting and fumarole acoustics at Aso Volcano, Japan. Journal of Volcanology and Geothermal Research 340, 16–29. https://doi.org/10.1016/j.jvolgeores.2017.03.029

Melki, S., Asmi, A.M.E., Sy, M.O.B., Gueddari, M., 2020. A geochemical assessment and modeling of industrial groundwater contamination by orthophosphate and fluoride in the Gabes-North aquifer, Tunisia. Environ Earth Sci 79, 135. https://doi.org/10.1007/s12665-020-8857-0

Milkov, A.V., Etiope, G., 2018. Revised genetic diagrams for natural gases based on a global dataset of >20,000 samples. Organic Geochemistry 125, 109–120. https://doi.org/10.1016/j.orggeochem.2018.09.002

Miller, S.A., 2013. Chapter 1 - The Role of Fluids in Tectonic and Earthquake Processes, in: Dmowska, R. (Ed.), Advances in Geophysics, Advances in Geophysics. Elsevier, pp. 1–46. https://doi.org/10.1016/B978-0-12-380940-7.00001-9

Miller, S.A., Collettini, C., Chiaraluce, L., Cocco, M., Barchi, M., Kaus, B.J.P., 2004. Aftershocks driven by a high-pressure CO2 source at depth. Nature 427, 724–727. https://doi.org/10.1038/nature02251

Minissale, A., Donato, A., Procesi, M., Pizzino, L., Giammanco, S., 2019. Systematic review of geochemical data from thermal springs, gas vents and fumaroles of Southern Italy for geothermal favourability mapping. Earth-Science Reviews 188, 514–535. https://doi.org/10.1016/j.earscirev.2018.09.008

Mongelli, G., Argyraki, A., Lorenzo, M.L.G., Shammout, M.W., Paternoster, M., Simeone, V., 2019. Groundwater Quality in the Mediterranean Region. Geofluids 2019, e7269304. https://doi.org/10.1155/2019/7269304

Mongelli, G., Paternoster, M., and Frittella, F., 2003. Caratterizzazione geochimica delle acque di sorgente dell'Alta Val d'Agri. Le risorse idriche sotterranee dell'Alta Val d'Agri. Autorità interregionale di Bacino della Basilicata,

Montalbetti, J.F., Kanasewich, E.R., 1970. Enhancement of Teleseismic Body Phases with a Polarization Filter. Geophysical Journal International 21, 119–129. https://doi.org/10.1111/j.1365-246X.1970.tb01771.x

Moratto, L., Romano, M.A., Laurenzano, G., Colombelli, S., Priolo, E., Zollo, A., Saraò, A., Picozzi, M., 2019. Source parameter analysis of microearthquakes recorded around the underground gas storage in the Montello-Collalto Area (Southeastern Alps, Italy). Tectonophysics 762, 159–168. https://doi.org/10.1016/j.tecto.2019.04.030

Morozov, I.B., Zhang, C., Duenow, J.N., Morozova, E.A., Smithson, S.B., 2008. Frequency Dependence of Coda Q, Part I: Numerical Modeling and Examples from Peaceful Nuclear Explosions. Bulletin of the Seismological Society of America 98, 2615–2628. https://doi.org/10.1785/0120080037

Moser, T.J., Nolet, G., Snieder, R., 1992. Ray bending revisited. Bulletin of the Seismological Society of America 82, 259–288. https://doi.org/10.1785/BSSA0820010259

Mostardini F. & Merlini S. (1986): Appennino centro meridionale sezioni geologiche e proposta di modello strutturale - 73° Congresso della Società Geologica Italiana

Murphy, K.P., 2012. Machine learning: a probabilistic perspective, Adaptive computation and machine learning series. MIT Press, Cambridge, MA.

Nakamura, Y. (1989). A Method for Dynamic Characteristics Estimation of Subsurface using Microtremor on the Ground Surface, Quarterly Report of RTRI, 30:1, 25-33

National Research Council. 2013. Induced Seismicity Potential in Energy Technologies. Washington, DC: The National Academies Press.https://doi.org/10.17226/13355.

Nelson, G.D., Vidale, J.E., 1990. Earthquake locations by 3-D finite-difference travel times. Bulletin of the Seismological Society of America 80, 395–410. https://doi.org/10.1785/BSSA0800020395

Nespoli, M., Belardinelli, M.E., Anderlini, L., Bonafede, M., Pezzo, G., Todesco, M., Rinaldi, A.P., 2017. Effects of layered crust on the coseismic slip inversion and related CFF variations: Hints from the 2012 Emilia Romagna earthquake. Physics of the Earth and Planetary Interiors 273, 23–35. https://doi.org/10.1016/j.pepi.2017.10.011

Noguera, A.M., Rea, G., 2000. Deep structure of the Campanian–Lucanian arc (southern Apennine, Italy). Tectonophysics 324, 239–265.

Ozima M. and Posodek F.A., 2002. Noble Gas Geochemistry, 2nd ed. xiv + 286 pp. Cambridge, New York, Melbourne: Cambridge University Press, ISBN 0 521 80366 7.3. Geological Magazine 140, 616–617. https://doi.org/10.1017/S0016756803258349

Paolo Galli, Sergio Castenetto, Edoardo Peronace, 2012. The MCS macroseismic survey of the Emilia 2012 earthquakes. Annals of Geophysics 55, 19. https://doi.org/10.4401/ag-6163

Parkhurst, D.L. and Appelo, C.A.J. (1999) User's Guide to PHREEQC (Version 2)—A Computer Program for Speciation, Batch-Reaction, One-Dimensional Transport, and Inverse Geochemical Calculations. U.S. Geological Survey, Water Resources Investigations Report 99-4259, Washington DC.

Passarelli, L.; Govoni, A.; Maccaferri, F.; Woith, H.; Strollo, A.; Zieke, T.; Margheriti, L.; Martinelli, G.; Dahm, T., 2017): East Pollino Experiment, Southern Italy. GFZ Data Services. Other/Seismic Network. doi:10.14470/LM582526.

Passarelli, Luigi; Govoni, Aladino; Francesco, Maccaferri; Woith, Heiko; Strollo, Angelo; Zieke, Thomas; Margheriti, Lucia; Martinelli, Giovanni; Dahm, Torsten (2017): East Pollino Experiment, Southern Italy. GFZ Data Services. Other/Seismic Network. doi:10.14470/LM582526

Patacca, E., Scandone, P., 2001. Late thrust propagation and sedimentary response in the thrust-belt—foredeep system of the Southern Apennines (Pliocene-Pleistocene), in: Vai, G.B., Martini, I.P. (Eds.), Anatomy of an Orogen: The Apennines and Adjacent Mediterranean Basins. Springer Netherlands, Dordrecht, pp. 401–440. https://doi.org/10.1007/978-94-015-9829-3_23

Paternoster, M., Buccione, R., Canora, F., Buttitta, D., Panebianco, S., Rizzo, G., Sinisi, R., Summa, V., Mongelli, G., 2021. Hydrogeochemistry and Groundwater Quality Assessment in the High Agri Valley (Southern Italy). Geofluids 2021, 1–15. https://doi.org/10.1155/2021/6664164

Paternoster, M., Liotta, M., Favara, R., 2008. Stable isotope ratios in meteoric recharge and groundwater at Mt. Vulture volcano, southern Italy. Journal of Hydrology 348, 87–97. https://doi.org/10.1016/j.jhydrol.2007.09.038

Paternoster, M., Oggiano, G., Sinisi, R., Caracausi, A., Mongelli, G., 2017. Geochemistry of two contrasting deep fluids in the Sardinia microplate (western Mediterranean): Relationships with tectonics and heat sources. Journal of Volcanology and Geothermal Research 336, 108–117. https://doi.org/10.1016/j.jvolgeores.2017.02.011

Paternoster, M., Scarfiglieri, A., and Mongelli, G., 2005. Groundwater chemistry in the high Agri Valley (southern Apennines, Italy), Geo Acta, vol. 4, pp. 25–41

Pavlis, G.L., 1986. Appraising earthquake hypocenter location errors: A complete, practical approach for single-event locations. Bulletin of the Seismological Society of America 76, 1699–1717. https://doi.org/10.1785/BSSA0760061699

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12 (2011) 2825-2830.

Pedrycz, W., Chen, S.-M. (Eds.), 2020. Deep Learning: Concepts and Architectures, Studies in Computational Intelligence. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-31756-0

Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. Sci. Adv. 4, e1700578. https://doi.org/10.1126/sciadv.1700578

Pichtel, J., 2016. Oil and Gas Production Wastewater: Soil Contamination and Pollution Prevention. Applied and Environmental Soil Science 2016, 1–24. https://doi.org/10.1155/2016/2707989

Pischiutta, M., Anselmi, M., Cianfarra, P., Rovelli, A., Salvini, F., 2013. Directional site effects in a non-volcanic gas emission area (Mefite d'Ansanto, southern Italy): Evidence of a local transfer fault transversal to large NW–SE

extensional faults? Physics and Chemistry of the Earth, Parts A/B/C 63, 116–123. https://doi.org/10.1016/j.pce.2013.03.008

Pischiutta, M., Salvini, F., Fletcher, J., Rovelli, A., Ben-Zion, Y., 2012. Horizontal polarization of ground motion in the Hayward fault zone at Fremont, California: dominant fault-high-angle polarization and fault-induced cracks: Horizontal polarization in Hayward fault zone. Geophysical Journal International 188, 1255–1272. https://doi.org/10.1111/j.1365-246X.2011.05319.x

Poiata, N., Satriano, C., Vilotte, J.-P., Bernard, P., Obara, K., 2016. Multiband array detection and location of seismic sources recorded by dense seismic networks. Geophys. J. Int. 205, 1548–1573. https://doi.org/10.1093/gji/ggw071

Poupinet, G., Ellsworth, W.L., Frechet, J., 1984. Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras Fault, California. Journal of Geophysical Research: Solid Earth 89, 5719–5731. https://doi.org/10.1029/JB089iB07p05719

Pratt, W.E., Johnson, D.W., 1926. Local Subsidence of the Goose Creek Oil Field. The Journal of Geology 34, 577–590. https://doi.org/10.1086/623352

Prinzhofer, A., Battani, A., 2003. Gas Isotopes Tracing: an Important Tool for Hydrocarbons Exploration. Oil & Gas Science and Technology - Rev. IFP 58, 299–311. https://doi.org/10.2516/ogst:2003018

Prosser, G., Palladino, G., Avagliano, D., Coraggio, F., Bolla, E.M., Riva, M., Catellani, D.E., 2021. Stratigraphic and Tectonic Setting of the Liguride Units Cropping Out along the Southeastern Side of the Agri Valley (Southern Apennines, Italy). Geosciences 11, 125. https://doi.org/10.3390/geosciences11030125

Pruess, K., 2008. On CO_2 fluid flow and heat transfer behavior in the subsurface, following leakage from a geologic storage reservoir. Environ Geol 54, 1677–1686. https://doi.org/10.1007/s00254-007-0945-x

Rätsch, G., 2004. A Brief Introduction into Machine Learning.

Reasenberg, P., Oppenheimer, D.H., 1985. FPFIT, FPPLOT and FPPAGE; Fortran computer programs for calculating and displaying earthquake fault-plane solutions (Report No. 85–739), Open-File Report. https://doi.org/10.3133/ofr85739

Richter, C.F., 1935. An instrumental earthquake magnitude scale. Bulletin of the Seismological Society of America 25, 1–32.

Rizzo, E., Colella, A., Lapenna, V., Piscitelli, S., 2004. High-resolution images of the fault-controlled High Agri Valley basin (Southern Italy) with deep and shallow electrical resistivity tomographies. Physics and Chemistry of the Earth, Parts A/B/C 29, 321–327. https://doi.org/10.1016/j.pce.2003.12.002

Roberts, R.G., Christoffersson, A., Cassidy, F., 1989. Real-Time Event Detection, Phase Identification and Source Location Estimation Using Single Station Three-Component Seismic Data. Geophysical Journal International 97, 471–480. https://doi.org/10.1111/j.1365-246X.1989.tb00517.x

Robinson, D., Scrimgeour, C.M., 1995. The contribution of plant C to soil CO_2 measured using $\delta 13C$. Soil Biology and Biochemistry 27, 1653–1656. https://doi.org/10.1016/0038-0717(95)00109-R

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Rowe, C.A., Aster, R.C., Borchers, B., Young, C.J., 2002. An Automatic, Adaptive Algorithm for Refining Phase Picks in Large Seismic Data Sets. Bulletin of the Seismological Society of America 92, 1660–1674. https://doi.org/10.1785/0120010224

Rubinstein, J.L., Ellsworth, W.L., McGarr, A., Benz, H.M., 2014. The 2001-Present Induced Earthquake Sequence in the Raton Basin of Northern New Mexico and Southern Colorado. Bulletin of the Seismological Society of America 104, 2162–2181. https://doi.org/10.1785/0120140009

Rubinstein, J.L., Mahani, A.B., 2015. Myths and Facts on Wastewater Injection, Hydraulic Fracturing, Enhanced Oil Recovery, and Induced Seismicity. Seismological Research Letters 86, 1060–1067. https://doi.org/10.1785/0220150067

Sambridge, M., Drijkoningen, G., 1992. Genetic algorithms in seismic waveform inversion. Geophysical Journal International 109, 323–342. https://doi.org/10.1111/j.1365-246X.1992.tb00100.x

Sambridge, M., Mosegaard, K., 2002. Monte Carlo Methods in Geophysical Inverse Problems. Reviews of Geophysics 40, 3-1-3–29. https://doi.org/10.1029/2000RG000089

Sano, Y., Gamo, T., Williams, S.N., 1997. Secular variations of helium and carbon isotopes at Galeras volcano, Colombia. Journal of Volcanology and Geothermal Research 77, 255–265. https://doi.org/10.1016/S0377-0273(96)00098-4

Sano, Y., Marty, B., 1995. Origin of carbon in fumarolic gas from island arcs. Chemical Geology 119, 265–274. https://doi.org/10.1016/0009-2541(94)00097-R

Sano, Y., Wakita, H., 1985. Geographical distribution of 3He/4He ratios in Japan: Implications for arc tectonics and incipient magmatism. Journal of Geophysical Research: Solid Earth 90, 8729–8741.

Sano, Y., Wakita, H., Italiano, F., Nuccio, M.P., 1989. Helium isotopes and tectonics in southern Italy. Geophysical Research Letters 16, 511–514.

Satriano, C., Zollo, A., Iannaccone, G., Bobbio, A., Cantore, L., Crosta, M.D., Elia, L., Festa, G., Lancieri, M., Martino, C., Romeo, A., Vassallo, M., 2008. The Irpinia Seismic Network (ISNet): a modern facility for earthquake early warning 3.

Schaff, D.P., Beroza, G.C., 2004. Coseismic and postseismic velocity changes measured by repeating earthquakes. Journal of Geophysical Research: Solid Earth 109. https://doi.org/10.1029/2004JB003011

Schiattarella, M., Di Leo, P., Beneduce, P., Ivo Giano, S., 2003. Quaternary uplift vs tectonic loading: a case study from the Lucanian Apennine, southern Italy. Quaternary International 101–102, 239–251. https://doi.org/10.1016/S1040-6182(02)00126-X

Schoell, M., 1980. The hydrogen and carbon isotopic composition of methane from natural gases of various origins. Geochimica et Cosmochimica Acta 44, 649–661. https://doi.org/10.1016/0016-7037(80)90155-6

Schoell, M., 1983. Genetic characterization of natural gases. AAPG Bulletin (American Association of Petroleum Geologists); (United States) 67:12.

Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X., 2017. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. ACM Trans. Database Syst. 42, 1–21. https://doi.org/10.1145/3068335

Scrocca, D., Sciamanna, S., Di Luzio, E., Tozi, M., Nicolai, C., Gambini, R., 2007. Structural setting along the CROP-04 deep seismic profile (Southern Apennines - Italy). Bollettino della Societa Geologica Italiana, Supplemento 7, 283–296.

Serlenga, V., Stabile, T.A., 2019. How do Local Earthquake Tomography and inverted dataset affect earthquake locations? The case study of High Agri Valley (Southern Italy). Geomatics, Natural Hazards and Risk 10, 49–78. https://doi.org/10.1080/19475705.2018.1504124

Shapiro, S.A., 2015. Fluid-Induced Seismicity. Cambridge University Press.

Shapiro, S.A., Kummerow, J., Dinske, C., Asch, G., Rothert, E., Erzinger, J., Kümpel, H.-J., Kind, R., 2006. Fluid induced seismicity guided by a continental fault: Injection experiment of 2004/2005 at the German Deep Drilling Site (KTB). Geophysical Research Letters 33. https://doi.org/10.1029/2005GL024659

Shiner, P., Beccacini, A., Mazzoli, S., 2004. Thin-skinned versus thick-skinned structural models for Apulian carbonate reservoirs: constraints from the Val d'Agri Fields, S Apennines, Italy. Marine and Petroleum Geology 21, 805–827. https://doi.org/10.1016/j.marpetgeo.2003.11.020

Snyder, G., Poreda, R., Fehn, U., Hunt, A., 2003. Sources of nitrogen and methane in Central American geothermal settings: Noble gas and 129I evidence for crustal and magmatic volatile components. Geochemistry, Geophysics, Geosystems 4, 1–28.

Stabile, T.A., 2015. Relationship Between Seismicity and Water Level of the Pertusillo Reservoir (southern Italy). BGTA. https://doi.org/10.4430/bgta0161

Stabile, T.A., Giocoli, A., Lapenna, V., Perrone, A., Piscitelli, S., Telesca, L., 2014a. Evidence of Low-Magnitude Continued Reservoir-Induced Seismicity Associated with the Pertusillo Artificial Lake (Southern Italy). Bulletin of the Seismological Society of America 104, 1820–1828. https://doi.org/10.1785/0120130333

Stabile, T.A., Giocoli, A., Perrone, A., Piscitelli, S., Lapenna, V., 2014b. Fluid injection induced seismicity reveals a NE dipping fault in the southeastern sector of the High Agri Valley (southern Italy). Geophys. Res. Lett. 41, 5847–5854. https://doi.org/10.1002/2014GL060948

Stabile, T.A., Iannaccone, G., Zollo, A., Lomax, A., Ferulano, M.F., Vetri, M.L.V., Barzaghi, L.P., 2013. A comprehensive approach for evaluating network performance in surface and borehole seismic monitoring. Geophysical Journal International 192, 793–806. https://doi.org/10.1093/gji/ggs049

Stabile, T.A., Satriano, C., Orefice, A., Festa, G., Zollo, A., 2012. Anatomy of a microearthquake sequence on an active normal fault. Sci Rep 2, 410. https://doi.org/10.1038/srep00410

Stabile, T.A., Serlenga, V., Satriano, C., Romanelli, M., Gueguen, E., Gallipoli, M.R., Ripepi, E., Saurel, J.-M., Panebianco, S., Bellanova, J., Priolo, E., 2020. The INSIEME seismic network: a research infrastructure for studying induced seismicity in the High Agri Valley (southern Italy) 20.

Stabile, T.A., Vlček, J., Wcisło, M., Serlenga, V., 2021. Analysis of the 2016–2018 fluid-injection induced seismicity in the High Agri Valley (Southern Italy) from improved detections using template matching. Sci Rep 11, 20630. https://doi.org/10.1038/s41598-021-00047-6

Tamburello, G., Pondrelli, S., Chiodini, G., Rouwet, D., 2018. Global-scale control of extensional tectonics on CO2 earth degassing. Nat Commun 9, 4608. https://doi.org/10.1038/s41467-018-07087-z

Tarantola, A., Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion. Rev. Geophys. 20, 219. https://doi.org/10.1029/RG020i002p00219

Tassi, F., Fiebig, J., Vaselli, O., Nocentini, M., 2012. Origins of methane discharging from volcanic-hydrothermal, geothermal and cold emissions in Italy. Chemical Geology 310–311, 36–48. https://doi.org/10.1016/j.chemgeo.2012.03.018

Tedesco, D., Allard, P., Sano, Y., Wakita, H., Pece, R., 1990. Helium-3 in subaerial and submarine fumaroles of Campi Flegrei caldera, Italy. Geochimica et Cosmochimica Acta 54, 1105–1116. https://doi.org/10.1016/0016-7037(90)90442-N

Telesca, L., Giocoli, A., Lapenna, V., Stabile, T.A., 2015. Robust identification of periodic behavior in the time dynamics of short seismic series: the case of seismicity induced by Pertusillo Lake, southern Italy. Stoch Environ Res Risk Assess 29, 1437–1446. https://doi.org/10.1007/s00477-014-0980-6

Thomas, D., 1988. Geochemical precursors to seismic activity. PAGEOPH 126, 241–266. https://doi.org/10.1007/BF00878998

Thurber, C.H., 1992. Hypocenter-velocity structure coupling in local earthquake tomography. Physics of the Earth and Planetary Interiors 75, 55–62. https://doi.org/10.1016/0031-9201(92)90117-E

Torfstein, A., Hammerschmidt, K., Friedrichsen, H., Starinsky, A., Garfunkel, Z., Kolodny, Y., 2013. Helium isotopes in Dead Sea Transform waters. Chemical Geology 352, 188–201. https://doi.org/10.1016/j.chemgeo.2013.06.008

Toutain, J.-P., Baubron, J.-C., 1999. Gas geochemistry and seismotectonics: a review. Tectonophysics 304, 1–27. https://doi.org/10.1016/S0040-1951(98)00295-9

Uchida, N., 2019. Detection of repeating earthquakes and their application in characterizing slow fault slip. Prog Earth Planet Sci 6, 40. https://doi.org/10.1186/s40645-019-0284-z

Valensise, G., Pantosti, D., n.d. The investigation of potential earthquake sources in peninsular Italy: A review 20.

Valoroso, L., Improta, L., Chiaraluce, L., Di Stefano, R., Ferranti, L., Govoni, A., Chiarabba, C., 2009. Active faults and induced seismicity in the Val d'Agri area (Southern Apennines, Italy). Geophysical Journal International 178, 488–502. https://doi.org/10.1111/j.1365-246X.2009.04166.x

Valoroso, L., Improta, L., De Gori, P., Chiarabba, C., 2011. Upper crustal structure, seismicity and pore pressure variations in an extensional seismic belt through 3-D and 4-D V P and V P /V S models: The example of the Val d'Agri area (southern Italy). J. Geophys. Res. 116, B07303. https://doi.org/10.1029/2010JB007661

Vengosh, A., Jackson, R.B., Warner, N., Darrah, T.H., Kondash, A., 2014. A Critical Review of the Risks to Water Resources from Unconventional Shale Gas Development and Hydraulic Fracturing in the United States. Environ. Sci. Technol. 48, 8334–8348. https://doi.org/10.1021/es405118y

Vespasiano, G., Apollaro, C., De Rosa, R., Muto, F., Larosa, S., Fiebig, J., Mulch, A., Marini, L., 2015. The Small Spring Method (SSM) for the definition of stable isotope–elevation relationships in Northern Calabria (Southern Italy). Applied Geochemistry 63, 333–346. https://doi.org/10.1016/j.apgeochem.2015.10.001

Vespasiano, G., Cianflone, G., Cannata, C.B., 2016. Analysis of groundwater pollution in the Sant'Eufemia plain (Calabria – South Italy). Italian Journal of Engineering Geology and Environment 5–15. https://doi.org/10.4408/IJEGE.2016-02.O-01

Vidale, J.E., Shearer, P.M., 2006. A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers. Journal of Geophysical Research: Solid Earth 111.

Vilardo, G., Guido, V., Terranova, C., Matano, F., nardò, S., 2009. Ground deformation due to tectonic, hydrothermal, gravity, hydrogeological, and anthropic processes in the Campania Region (Southern Italy) from Permanent Scatterers Synthetic Aperture Radar Interferometry. Remote Sensing of Environment 113, 197–212. https://doi.org/10.1016/j.rse.2008.09.007

Vlček, J., Eisner, L., Stabile, T.A., Telesca, L., 2018. Temporal Relationship Between Injection Rates and Induced Seismicity. Pure Appl. Geophys. 175, 2821–2835. https://doi.org/10.1007/s00024-017-1622-y

Waldhauser, F., Ellsworth, W.L., 2000. A Double-difference Earthquake location algorithm: Method and application to the Northern Hayward Fault, California. Bulletin of the Seismological Society of America. https://doi.org/10.1785/0120000006

Wang, X., Zhang, L., Zhao, Z., Cai, Y., 2018. Heavy metal pollution in reservoirs in the hilly area of southern China: Distribution, source apportionment and health risk assessment. Science of The Total Environment 634, 158–169. https://doi.org/10.1016/j.scitotenv.2018.03.340

Watanabe, T., Hiramatsu, Y., Obara, K., 2007. Scaling relationship between the duration and the amplitude of non-volcanic deep low-frequency tremors. Geophys. Res. Lett. 34, L07305. https://doi.org/10.1029/2007GL029391

Wathelet, M., Chatelain, J.-L., Cornou, C., Giulio, G.D., Guillier, B., Ohrnberger, M., Savvaidis, A., 2020. Geopsy: A User-Friendly Open-Source Tool Set for Ambient Vibration Processing. Seismological Research Letters 91, 1878–1889. https://doi.org/10.1785/0220190360

Wcisło, M., Stabile, T.A., Telesca, L., Eisner, L., 2018. Variations of attenuation and VP/VS ratio in the vicinity of wastewater injection: A case study of Costa Molina 2 well (High Agri Valley, Italy). GEOPHYSICS 83, B25–B31. https://doi.org/10.1190/geo2017-0123.1

Weber, E., Iannaccone, G., Zollo, A., Bobbio, A., Cantore, L., Corciulo, M., Convertito, V., Di Crosta, M., Elia, L., Emolo, A., Martino, C., Romeo, A., Satriano, C., 2007. Development and Testing of an Advanced Monitoring Infrastructure (ISNet) for Seismic Early-warning Applications in the Campania Region of Southern Italy, in: Gasparini, P., Manfredi, G., Zschau, J. (Eds.), Earthquake Early Warning Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 325–341. https://doi.org/10.1007/978-3-540-72241-0_16

Whiticar, M.J., 1999. Carbon and hydrogen isotope systematics of bacterial formation and oxidation of methane. Chemical Geology 161, 291–314. https://doi.org/10.1016/S0009-2541(99)00092-3

Yatsevich, I., Honda, M., 1997. Production of nucleogenic neon in the Earth from natural radioactive decay. J. Geophys. Res. 102, 10291–10298. https://doi.org/10.1029/97JB00395

Yuce, G., Italiano, F., D'Alessandro, W., Yalcin, T.H., Yasin, D.U., Gulbay, A.H., Ozyurt, N.N., Rojay, B., Karabacak, V., Bellomo, S., Brusca, L., Yang, T., Fu, C.C., Lai, C.W., Ozacar, A., Walia, V., 2014. Origin and interactions of fluids circulating over the Amik Basin (Hatay, Turkey) and relationships with the hydrologic, geologic and tectonic settings. Chemical Geology 388, 23–39. https://doi.org/10.1016/j.chemgeo.2014.09.006

Zembo, I., 2010. Stratigraphic architecture and quaternary evolution of the Val d'Agri intermontane basin (Southern Apennines, Italy). Sedimentary Geology 223, 206–234. https://doi.org/10.1016/j.sedgeo.2009.11.011

Zeng, X., McMechan, G.A., 2006. Two methods for determining geophone orientations from VSP data. GEOPHYSICS 71, V87–V97. https://doi.org/10.1190/1.2208935

Zhang, M., Xu, S., Zhou, X., Caracausi, A., Sano, Y., Guo, Z., Zheng, G., Lang, Y.-C., Liu, C.-Q., 2021. Deciphering a mantle degassing transect related with India-Asia continental convergence from the perspective of volatile origin and outgassing. Geochimica et Cosmochimica Acta 310, 61–78. https://doi.org/10.1016/j.gca.2021.07.010

Zhu, W., Beroza, G.C., 2018. PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method. Geophysical Journal International. https://doi.org/10.1093/gji/ggy423

Zoback, M.D., Harjes, H.-P., 1997. Injection-induced earthquakes and crustal stress at 9 km depth at the KTB deep drilling site, Germany. Journal of Geophysical Research: Solid Earth 102, 18477–18491. https://doi.org/10.1029/96JB02814

Zollo, A., Emolo, A., 2011. Terremoti e onde: introduzione alla sismologia moderna, 1. ed. italiana., Biblioteca scientifica. Liguori, Napoli.

Zollo, A., Orefice, A., Convertito, V., 2014. Source parameter scaling and radiation efficiency of microearthquakes along the Irpinia fault zone in southern Apennines, Italy. J. Geophys. Res. Solid Earth 119, 3256–3275. https://doi.org/10.1002/2013JB010116

Appendix

Appendix 3.1 – Castelsaraceno sequence HypoDD output catalgue

Event ID	Date	Origin Time	Latitude DD	Longitude DD	Depth Km	ML	errlat	errlon	errz
1	07/08/2020	04:29:18	40.2027	15.9371	11.81	-0.3	118.2	108.8	112.4
2	07/08/2020	07:21:59	40.1916	15.939	11.46	-0.1	131.1	111.4	82.1
3	07/08/2020	08:52:32	40.1939	15.939	11.62	2.1	82.5	65.1	52.9
4	07/08/2020	09:00:30	40.1948	15.9386	11.5	1.1	85.9	93.7	71.4
5	07/08/2020	09:03:08	40.1946	15.9385	11.34	0.7	72.5	74.9	52.8
6	07/08/2020	09:04:09	40.1936	15.9382	11.29	0.4	68.6	73.1	53.7
7	07/08/2020	09:07:39	40.1953	15.9375	11.37	-0.1	81.8	80.6	48.2
8	07/08/2020	09:08:34	40.1957	15.937	11.26	0.7	69.4	67.6	53
9	07/08/2020	09:09:01	40.195	15.9363	11.34	0.8	79.4	94.7	72.6
10	07/08/2020	09:12:10	40.1954	15.9375	11.38	0.4	79.4	60.5	50.6
11	07/08/2020	09:16:38	40.1957	15.9356	11.51	0.7	70.3	62.3	49.6
12	07/08/2020	09:17:28	40.2119	15.9278	11.82	-0.2	155.5	201.4	94.3
13	07/08/2020	09:20:33	40.1975	15.9365	11.44	0.3	64.9	70.6	59
14	07/08/2020	09:35:00	40.1918	15.9398	11.04	0.3	77.3	63.3	53.2
15	07/08/2020	09:39:56	40.1945	15.9405	11.06	0.1	75.3	66.6	90.9
16	07/08/2020	10:40:52	40.1941	15.9414	11.2	-0.2	111.6	111.6	64
17	07/08/2020	10:47:28	40.1859	15.943	10.83	-0.2	98	75.8	72.7
18	07/08/2020	11:40:22	40.1961	15.9378	11.5	0.7	70.6	73.4	50.3
19	07/08/2020	13:34:37	40.1939	15.9382	11.57	2.9	67.4	75.2	61.7
20	07/08/2020	13:37:15	40.185	15.9422	10.64	0	108.6	95.2	85.7
21	07/08/2020	13:46:49	40.1988	15.9345	12.3	-0.3	119.5	147.8	87.9
22	07/08/2020	14:02:55	40.2006	15.9256	10.38	-0.5	150.2	178.8	124.5
23	07/08/2020	14:05:19	40.1907	15.9432	11.33	-0.4	104	94.6	64.4
24	07/08/2020	14:25:16	40.1925	15.9409	11.04	0	83.5	54.1	46.8
25	07/08/2020	15:07:58	40.2266	15.9212	11.6	-0.4	436.3	604	230.3
26	07/08/2020	15:28:11	40.2355	15.9239	12.11	-0.5	201.3	293.7	120.6
27	07/08/2020	15:32:13	40.1898	15.9398	11.19	-0.2	97	72.4	66.7
28	07/08/2020	17:23:47	40.1961	15.9377	11.05	0.6	67.3	79.8	56.9
29	07/08/2020	19:05:43	40.2125	16.041	10.83	0.8	82.5	89	89.7
30	07/08/2020	20:14:55	40.2463	15.9219	12.07	-0.4	265.9	272.9	146.7
31	07/08/2020	20:17:07	40.1909	15.94	11.49	-0.5	118.2	88.2	70.3
32	07/08/2020	21:36:57	40.2119	16.0406	10.59	0.6	99.5	102.6	95.8
33	07/08/2020	21:48:54	40.2041	15.9359	11.68	-0.4	115.5	132.4	68.6
34	07/08/2020	22:16:20	40.1935	15.9284	11.9	-0.4	112.9	135.8	122.5
35	07/08/2020	22:36:41	40.2118	15.9359	12.47	-0.5	184.7	310.9	108
36	07/08/2020	22:40:29	40.1854	15.9418	10.37	-0.3	136.9	146.8	128.9
37	07/08/2020	22:56:58	40.2114	15.9314	11.8	-0.6	156	189	88.7
38	07/08/2020	23:06:40	40.1928	15.9413	11.18	-0.1	64.9	81.8	60.3
39	07/08/2020	23:26:19	40.1865	15.9427	10.65	-0.4	154.6	177.4	131.1

40	07/08/2020	23:33:20	40.2007	15.9399	11.51	-0.5	161.6	130.2	92.5
41	08/08/2020	00:08:48	40.2202	15.9281	12.28	-0.6	223	341.1	121.5
42	08/08/2020	00:11:59	40.2006	15.9393	12.09	-0.7	147.8	149.9	110.6
43	08/08/2020	00:13:42	40.1918	15.9357	11.23	-0.4	100.7	83.8	65.5
44	08/08/2020	00:50:07	40.1924	15.942	11.07	-0.2	67.5	73	51.7
45	08/08/2020	01:07:56	40.1954	15.9431	11	-0.8	190.7	202.8	140.5
46	08/08/2020	01:07:56	40.191	15.9435	10.83	-0.8	63.2	77.5	61.4
47	08/08/2020	01:35:56	40.1914	15.9452	10.35	-0.2	124.1	136.6	99.4
48	08/08/2020	02:16:46	40.1927	15.9444	11.35	-0.6	86.2	78.9	56.2
49	08/08/2020	02:24:43	40.1969	15.9435	10.99	-0.3	113.9	128.4	104.7
50	08/08/2020	03:09:45	40.1916	15.9459	10.65	-0.6	82.5	112.8	64.1
51	08/08/2020	03:13:35	40.1971	15.9385	11.3	-0.3	123.4	83.7	62.3
52	08/08/2020	03:20:13	40.1833	15.9436	10.79	-0.5	152.7	150	121.2
53	08/08/2020	03:21:43	40.2005	15.9419	11.24	-0.6	167.9	190.2	129.2
54	08/08/2020	03:22:28	40.1931	15.9398	11.44	-0.4	95.6	81.5	60.2
55	08/08/2020	03:23:09	40.1948	15.9377	11.47	0.4	63.2	75.9	53.1
56	08/08/2020	03:48:21	40.2537	15.9167	11.91	-0.5	290.2	286.8	161.3
57	08/08/2020	03:53:27	40.1968	15.9364	11.66	1.1	117	130.5	97.6
58	08/08/2020	03:54:04	40.1917	15.9395	11.04	-0.3	84.6	68.7	44.9
59	08/08/2020	03:54:38	40.2242	15.9316	11.69	-0.6	277.6	348	109.3
60	08/08/2020	04:00:30	40.2131	16.0427	10.97	0.4	81.9	86.1	89.1
61	08/08/2020	04:13:27	40.191	15.9416	11.2	-0.5	94.4	86.9	60.9
62	08/08/2020	04:53:17	40.2131	16.0407	10.93	1.4	86.1	90.8	98.6
63	08/08/2020	05:47:34	40.1863	15.9483	10.33	-0.2	136.6	128.4	90.2
64	08/08/2020	07:10:11	40.1978	15.941	11.24	0.9	82.7	76.1	71.8
65	08/08/2020	11:59:27	40.1994	15.9401	10.57	-0.1	100.5	110.2	104.4
66	08/08/2020	13:19:41	40.1978	15.937	11.34	0.1	71.7	64.9	51.7
67	08/08/2020	18:46:30	40.1896	15.9395	11.07	-0.3	95.3	87.9	55
68	08/08/2020	21:40:05	40.1958	15.9383	11.52	1	89.3	80.8	71.5
69	08/08/2020	21:40:53	40.1957	15.939	11.22	-0.2	67.5	81.7	55.5
70	08/08/2020	22:22:54	40.1932	15.9397	11.1	-0.2	68.1	79.9	59
71	08/08/2020	22:40:29	40.1944	15.9388	11.22	-0.1	73.3	76.3	60.6
72	08/08/2020	22:55:51	40.1964	15.9384	11.2	-0.1	87.6	89.6	68.4
73	08/08/2020	23:46:12	40.1965	15.94	11.45	-0.3	84	122.9	68.4
74	09/08/2020	01:08:11	40.2037	15.9393	11.97	-0.4	211.4	203.1	138.9
75	09/08/2020	02:52:42	40.1863	15.937	10.66	-0.6	97.1	80.1	86.3
76	09/08/2020	07:25:09	40.2019	15.9315	11.11	-0.1	93.6	100.4	82.6
77	09/08/2020	21:50:08	40.1966	15.935	10.95	-0.2	59.1	61.9	45.4
78	10/08/2020	08:35:18	40.1978	15.9348	11.48	0.4	99.7	124.5	79.8
79	10/08/2020	09:35:50	40.1961	15.9352	11.26	0.1	92.8	76.9	58
80	10/08/2020	10:16:42	40.1853	15.9426	10.43	-0.4	134.1	124.1	101.8

Appendix 3.2 – Location, physico-chemical parameters, and major elements of the analyzed samples

Sample ID	Sampling date	Name	Long UTM(WGS84)	Lat UTM(WGS84)	Zone UTM(WGS84)	Altitude m	Flow rate L/s	O_2 mg/L	pH	EC µS/cm	Eh mV	T °C	TDS mg/L	Na^+ mg/L	K^+ mg/L	Ca^{2+} mg/L	Mg^{2+} mg/L	SiO_2 mg/L	Cl mg/L	NO_3^- mg/L	NO_2^- mg/L	SO_4^{2-} mg/L	HCO_3 mg/L
SGC	13/11/2020	S. Giovanni	586023	4447243	33T	765	12.0	9.3	7.472	509	209.1	12.5	397.7	3.17	0.80	88.16	3.14	6.42	4.66	1.60	b.d.l.	8.93	287
SAL	11/12/2020	Fontana del Salice	573775	4461185	33T	547	10.4	8.7	7.463	531	228.8	13.5	449.6	5.23	1.03	78.13	14.30	9.20	7.53	3.53	-	10.17	329
FCP	25/11/2020	Fontana Carpineta	575367	4460559	33T	554	20.0	6.5	7.429	541	208.1	14.0	454.3	9.80	4.28	79.24	14.37	11.34	10.43	8.20	-	16.20	311
AGR	23/10/2020	Capo d'Agri	563843	4478079	33T	930	30.0	10.6	7.925	298	235.5	8.4	230.1	3.97	0.65	37.13	11.42	8.35	7.19	1.15	-	15.21	153
POC	23/10/2020	Pozzo Capano	565770	4470297	33T	630	30.0	9.5	7.627	369	332.5	12.3	309.6	4.93	1.12	57.42	6.89	10.91	6.84	1.27	-	11.16	220
OCC	12/10/2020	Occhio	561528	4475619	33T	791	4.5	10.3	8.281	432	253.3	11.4	386.9	4.56	1.52	67.52	11.94	10.49	7.50	0.82	-	12.04	281
CAQ	12/10/2020	Capo d'Acqua	561602	4474386	33T	700	10.0	9.8	7.564	401	254.3	12.0	349.3	3.79	0.85	76.20	2.36	7.28	5.68	1.30	-	8.74	250
PMR	12/10/2020	Palmiro	561790	4472099	33T	659	11.9	9.6	7.695	426	265.7	11.4	391.0	4.31	0.85	61.40	14.74	6.42	6.85	1.11	-	9.50	268
VAN	12/10/2020	Fontana S. Giovanni	562285	4473921	33T	689	2.7	8.9	7.503	538	235.6	13.0	426.3	13.77	24.38	69.93	9.98	15.84	25.93	16.22	-	19.82	256
MOL	23/10/2020	Molinara	567963	4471645	33T	890	12.5	10.5	8.056	321	321.5	8.6	289.1	21.17	0.79	36.64	11.41	7.92	5.11	1.25	-	16.66	195
PSC	23/10/2020	Pozzo Peschiera	570187	4468035	33T	702	70.0	9.6	7.736	406	257.1	12.8	332.1	6.30	2.59	58.61	9.26	7.92	12.60	7.08	-	9.36	226
TAS	25/11/2020	Tasso	570367	4472185	33T	1290	12.5	9.7	7.851	351	209.5	7.4	266.0	7.27	1.15	46.87	9.66	6.21	11.61	0.12	-	12.00	177
LAV	25/11/2020	Lato Viggiano	570239	4471716	33T	1121	4.0	8.4	7.529	652	183.5	10.7	490.5	15.47	1.21	106.65	3.37	8.56	57.41	0.49	-	13.61	292
GIO	25/11/2020	Giordano	573207	4447596	33T	1110	12.5	10.4	7.878	368	226.0	8.6	243.4	2.62	0.52	52.86	1.77	6.42	4.80	0.99	-	8.79	171
SVG	13/11/2020	S. Salvatore	585950	4462667	33T	922	-	9.3	7.348	673	252.5	11.0	553.5	10.76	2.23	108.80	12.68	22.04	13.81	1.60	-	18.40	384
OSC	22/10/2020	Oscuriello	560925	4468706	33T	913	23.0	9.3	7.627	463	250.3	9.5	394.0	5.05	0.98	54.38	20.77	8.99	7.17	2.19	-	10.16	293
AMO	22/10/2020	Amoroso	560925	4468706	33T	913	35.0	9.8	7.644	494	261.0	9.7	406.3	5.20	0.85	55.44	23.74	7.49	8.52	1.39	-	11.72	299
AGG	22/10/2020	Aggia	563682	4466705	33T	608	130.0	10.4	7.840	352	253.2	10.6	277.7	3.10	0.66	43.26	11.39	5.35	5.21	1.14	-	11.12	201
SAN	22/10/2020	Monaco Santino	561168	4472667	33T	685	35.0	9.9	7.763	354	234.0	10.6	298.5	3.71	0.78	51.49	10.22	7.92	5.83	1.29	-	8.90	220
PEN	11/11/2020	Peschiera Santino	561136	4472576	33T	680	2.5	9.3	7.754	481	236.3	10.2	336.2	4.17	1.55	50.59	15.90	7.92	6.83	1.59	1.08	9.41	244
MIE	11/12/2020	S. Miele	562625	4467497	33T	730	5.5	9.4	7.666	414	243.6	11.7	288.2	3.29	0.55	53.37	8.61	5.14	5.12	2.95	0.21	12.38	201
PAZ	13/11/2020	Acqua Panzone 1	562010	4469366	33T	635	5.0	6.3	7.566	489	264.3	14.1	406.7	15.37	1.79	59.67	15.48	13.48	8.51	1.91	-	16.46	287
ZAP	13/11/2020	Acqua Panzione 2	562010	4469366	33T	635	2.0	4.6	7.616	465	246.7	16.8	398.6	24.27	2.06	49.38	15.73	16.05	8.53	1.10	-	21.73	275
CPN	12/10/2020	Fontana Capano	565577	4470112	33T	607	4.5	9.5	7.651	377	268.5	12.2	320.2	5.07	1.16	60.79	7.14	11.56	7.06	1.84	-	10.78	226
FAB	11/12/2020	Fabbricata	572786	4456867	33T	729	12.5	9.5	7.318	859	241.7	11.9	656.3	25.31	67.69	120.00	9.14	9.20	60.96	5.70	-	12.97	354
SOG	11/12/2020	Sorgituro	572019	4451104	33T	920	12.0	10.2	7.811	458	256.2	9.8	252.5	3.24	0.48	52.04	2.29	8.13	4.44	2.30	-	10.45	177
SAM	22/10/2020	San Michele	560157	4483184	33T	1116	24.0	10.5	7.910	340	258.0	7.9	261.0	3.16	1.02	45.74	12.03	8.35	5.47	1.57	-	8.45	183
CAC	22/10/2020	Capo Caolo	564222	4461623	33T	725	10.0	9.0	7.640	418	240.4	11.0	343.1	3.53	0.76	50.42	15.33	5.78	5.54	2.07	-	12.35	256
CAR	11/12/2020	Caramma	564548	4463300	33T	651	4.0	9.3	7.693	457	259.2	12.2	334.8	3.70	0.59	53.13	16.50	5.56	5.73	1.20	-	9.57	244
CAP	11/12/2020	Capo d'Acqua	567088	4462899	33T	650	10.0	9.0	7.488	481	256.0	11.5	371.2	6.20	1.62	77.98	6.12	14.77	9.23	2.99	-	10.41	256
ALLI	23/10/2020	Alli	575408	4468034	33T	844	25.0	8.4	7.820	413	283.2	12.1	349.1	14.70	1.13	42.69	17.13	12.20	8.07	0.34	-	32.37	232
PZ11	22/02/2021	Piezometro - PZ11	577590	4462817	33T	582	-	3.8	7.135	616	242.8	10.9	515.9	15.51	1.95	90.57	15.97	8.99	16.18	0.67	-	51.28	323
PZ6	22/02/2021	Piezometro - PZ6	581597	4462110	33T	637.27	-	3.6	7.707	796	233.2	10.9	738.6	22.27	2.45	112.36	23.98	18.83	20.07	-	-	68.82	488
PZ5	22/02/2021	Piezometro - PZ5	580324	4463309	33T	708.31	-	4.4	6.846	1178	223.0	16.4	878.9	47.22	2.27	160.07	28.14	11.34	136.80	34.98	-	34.90	433
PZ7	22/02/2021	Piezometro - PZ7	583526	4463506	33T	951.17	-	3.1	6.818	1675	216.8	14.4	1267.2	34.94	2.52	227.84	40.89	16.69	95.43	373.89	-	56.98	433
TRA 13.11.20	13/11/2020	Pozzo Tramutola	564493	4463843	33T	620	26.0	1.0	7.887	2590	104.5	28.0	2074.0	555.01	7.33	3.80	6.41	17.55	288.60	0.54	-	53.47	1135
TRA 22.02.21	22/02/2021	Pozzo Tramutola	564493	4463843	33T	620	26.0	1.3	7.882	2510	65.0	28.0	2107.2	599.74	8.10	4.62	6.90	17.12	283.56	0.52	-	52.17	1135
SUL	11/12/2020	Acqua Solfata	566667	4470162	33T	636	2.8	3.6	7.573	424	192.2	14.9	10.0	10.79	0.87	48.84	18.69	14.12	6.93	-	-	12.61	281
SUF	22/02/2021	Acqua solfata	566667	4470162	33T	636	10	2.41	7.604	412	-113.5	15.0	9.0	9.49	0.68	47.70	17.88	13.05	6.83	-	-	12.82	256

Appendix 3.3 – Samples trace and ultratrace elements from ICP-MS analysis

Sample ID	Li μG/L	Si μG/L	Al μG/L	B μG/L	Ti μG/L	V μG/L	Cr μG/L	Mn μG/L	Fe μG/L	Ni μG/L	Cu μG/L	Zn μG/L	As μG/L	Se μG/L	Rb μG/L	Sr μG/L	Zr μG/L	Cd μG/L	Sn μG/L	Sb μG/L	Cs μG/L	Ba μG/L	U μG/L	Hg μG/L	Pb μG/L
Detection Limit	1	200	2	3	0.1	0.1	0.5	0.1	10	0.3	0.2	0.5	0.03	0.2	0.005	0.04	0.01	0.01	0.1	0.01	0.001	0.1	0.001	0.2	0.01
SGC	-	3000	6	7	0.5	0.8	-	0.2	-	-	0.3	1.2	0.7	-	0.596	89.3	-	-	0.1	0.03	0.17	23.1	0.661	-	0.08
SAL	1	4300	4	13	0.6	1.7	-	1.1	-	0.4	0.3	1.3	0.7	0.3	1.24	121	-	-	0.1	0.06	0.018	15	0.796	-	0.17
FCP	4	5300	3	38	0.1	0.7	-	0.6	-	-	0.3	2.9	0.52	0.5	1.28	256	-	-	0.4	0.04	0.014	25.5	0.652	-	0.17
AGR	3	3900	12	13	-	-	-	0.5	-	0.4	0.6	4.1	0.07	0.2	0.708	324	0.06	-	-	0.03	0.013	11.6	0.144	-	0.3
POC	3	5100	5	20	-	0.1	-	0.2	-	-	1.8	5.2	0.07	0.2	0.985	288	0.04	-	-	0.01	0.015	12.9	0.225	-	0.28
OCC	1	4900	4	14	0.8	0.2	-	1.1	-	-	0.7	16.5	0.26	-	1.08	113	-	-	0.4	0.06	0.02	18.8	0.34	-	0.46
CAQ	-	3400	5	8	0.6	0.4	-	0.4	-	-	0.4	4.5	0.17	-	1.12	97	-	-	-	0.02	0.002	11.2	0.287	-	0.08
PMR	-	3000	4	9	0.5	1.4	-	0.3	-	-	0.3	3.7	0.21	-	1.25	68.3	-	-	0.2	0.03	0.004	7.7	0.605	-	0.16
VAN	1	7400	3	39	0.4	1	-	0.4	-	-	0.9	15.2	3.47	-	4.05	171	-	-	-	0.31	0.015	30.9	0.334	-	0.08
MOL	3	3700	15	11	-	0.1	-	0.6	-	-	0.4	2.5	0.09	0.4	0.598	374	0.04	-	-	0.03	0.013	9.8	0.146	0.8	0.57
PSC	-	3700	3	11	0.5	1.8	-	0.4	-	0.6	0.8	5.9	1.01	0.3	2.06	87.7	-	-	-	0.05	0.02	12.2	0.38	-	0.33
TAS	1	2900	-	18	0.3	0.1	-	0.2	20	0.3	0.8	1.7	0.11	-	1.17	106	-	-	-	0.03	0.027	13.7	0.109	-	0.18
LAV	2	4000	4	21	0.1	-	-	12.5	20	0.3	0.4	2.7	0.08	-	1.2	165	-	-	-	0.04	0.006	96.7	0.212	-	0.16
GIO	-	3000	-	25	-	-	-	0.4	10	-	0.4	2.5	0.11	-	0.559	133	-	-	-	0.03	0.016	13.3	0.075	-	0.09
SVG	6	10300	3	20	1	0.6	-	0.2	-	-	0.3	1.9	0.19	0.6	0.413	247	0.03	0.01	0.1	0.02	0.003	49.1	0.782	-	0.14
OSC	-	4200	5	10	0.7	1.8	-	0.6	-	-	0.5	2.2	0.26	0.3	1.7	61.7	0.06	-	-	0.03	0.012	8.1	0.698	-	0.46
AMO	9	3500	18	9	0.9	1.5	-	1.1	10	1.8	2.5	12.2	0.29	-	4.87	51.9	0.08	0.03	0.3	0.05	0.014	8.9	0.536	-	0.84
AGG	-	2500	25	7	0.9	1.4	-	17	20	0.3	0.9	2.4	0.5	-	1.61	54.8	0.02	-	-	0.04	0.017	9.8	0.32	-	0.3
SAN	-	3700	7	7	-	0.9	0.5	0.6	-	-	0.4	3	0.23	0.2	1.65	67.5	0.03	-	-	0.03	0.005	9.3	0.37	-	0.42
PEN	-	3700	6	10	0.7	1.2	-	0.7	-	0.4	1.7	12.8	0.2	0.2	2.19	54.1	0.04	0.02	0.2	0.05	0.011	8	0.47	-	1.35
MIE	-	2400	4	13	0.5	1.2	-	0.7	-	0.4	0.8	2.8	0.55	0.3	0.975	52.2	-	0.01	0.2	0.05	0.009	7.7	0.325	-	0.49
PAZ	24	6300	4	90	0.1	0.3	-	0.3	-	-	0.7	3	0.3	-	1.61	231	0.01	0.01	0.2	0.03	0.022	50.2	0.554	-	0.29
ZAP	45	7500	4	157	-	0.1	-	0.7	-	0.4	0.8	3	0.28	0.3	1.78	378	0.01	0.01	0.2	0.04	0.036	81.3	0.449	-	0.29
CPN	3	5400	7	20	-	0.1	-	1.2	20	-	0.5	4.9	0.08	-	1.03	302	0.02	-	-	0.03	0.01	13.6	0.227	-	0.25
FAB	1	4300	5	13	0.8	0.1	-	0.4	-	0.4	0.6	3.6	0.21	-	1.57	214	0.01	-	0.2	0.02	0.02	26.5	0.344	-	0.3
SOG	2	3800	5	9	0.2	-	-	0.8	-	-	0.5	3.2	0.13	-	0.835	98.3	-	-	0.2	0.03	0.019	12	0.084	-	0.22
SAM	-	3900	8	7	0.6	0.2	-	1.1	30	0.3	0.4	1.7	0.25	-	0.956	51	0.17	-	-	0.02	0.007	10.6	0.147	-	0.16
CAC	-	2700	7	8	0.2	1.4	-	0.5	-	0.4	0.8	2.9	0.79	0.4	1.45	74.1	-	-	-	0.08	0.019	10.6	0.918	-	0.22
CAR	-	2600	6	11	0.4	1.1	-	0.6	20	0.5	2.7	11.5	0.77	0.3	1.28	65.8	-	-	0.2	0.07	0.016	11.3	0.46	-	0.83
CAP	3	6900	6	16	0.3	0.1	-	0.4	-	0.6	0.6	7	0.22	0.9	1.54	127	-	-	0.2	0.02	0.03	31	0.32	-	0.26
ALLI	17	5700	10	62	-	-	-	0.6	-	0.5	0.5	2.5	0.67	0.2	0.76	456	0.06	0.02	0.2	0.06	0.026	50.6	0.241	-	0.21
PZI1	10	4200	-	57	-	0.4	-	1.9	20	0.4	1.2	52.3	0.09	0.7	0.289	534	0.01	0.04	0.2	0.17	-	66.6	0.849	-	0.04
PZ6	24	8800	-	45	-	-	-	1.8	30	0.9	1.5	9.7	0.19	0.6	0.796	948	0.01	0.04	0.2	0.14	0.003	222	2.32	-	0.08
PZ5	22	5300	-	60	-	-	-	0.5	10	0.5	2.9	7.7	0.09	0.2	0.508	883	-	0.04	0.3	0.08	0.007	68.4	1.23	-	0.12
PZ7	38	7800	-	35	-	-	-	1.3	30	3.4	2.2	27	0.13	0.8	0.774	1660	-	0.05	0.4	0.2	-	169	3.15	-	0.11
TRA 13.11.20	881	8200	-	5170	-	-	-	36.1	-	-	-	6.9	-	-	10.4	2300	0.33	-	0.8	-	3.37	219	-	-	0.52
TRA 22.02.21	770	8000	-	8820	-	-	-	37.8	90	-	1.3	4.8	-	-	11	2470	0.1	-	0.2	-	3.43	213	0.018	1.6	0.11
SUL	8	6600	5	32	-	-	-	29.2	40	0.4	0.3	3	-	-	0.863	612	0.02	-	0.2	-	0.032	21.3	0.055	-	0.14
SUF	6	6100	-	25	-	-	-	37.3	60	0.4	1.6	5.2	0.05	-	0.801	650	0.01	-	0.2	0.02	0.028	21.7	0.076	-	0.12

Appendix 3.4 – Isotopic data, carbon composition (TDIC, C$_{carb}$, C$_{ext}$) and saturation index (SI)

Sample ID	δD$_{H2O}$	δ^{18}O$_{H2O}$	δ^{13}C-TDIC	TDIC	C$_{carb}$	C$_{ext}$	δ^{13}C$_{ext}$	SI-calcite	SI-dolomite	SI-albite
	‰ vs. V-SMOW	‰ vs. V-SMOW	‰ PDB	mol/L	mmol/L	mmol/L	‰	Log Q/K	Log Q/K	Log Q/K
SGC	-53	-8.9	-14.3	5.08E-03	2.33	2.75	-28.11	0.34	0.29	-2.40
SAL	-47	-7.7	-	-	-	-	-	0.34	1.01	-1.99
FCP	-51	-8.6	-	-	-	-	-	0.29	0.92	-1.55
AGR	-52	-8	-	-	-	-	-	0.20	0.92	-1.18
POC	-51	-8.2	-11.4	3.81E-03	1.72	2.09	-22.40	0.23	0.60	-1.56
OCC	-50	-8.5	-12.2	4.62E-03	2.18	2.45	-24.83	1.03	2.36	-1.66
CAQ	-49	-8.2	-13.9	4.37E-03	2.00	2.37	-27.31	0.32	0.19	-2.18
PMR	-50	-8.5	-12.6	4.60E-03	2.14	2.47	-25.26	0.38	1.20	-2.32
VAN	-52	-8.8	-13.3	4.51E-03	2.16	2.35	-27.30	0.22	0.67	-0.92
MOL	-50	-8.1	-9	3.25E-03	1.38	1.86	-17.16	0.33	1.19	0.58
PSC	-52	-8	-13	3.85E-03	1.84	2.01	-26.75	0.37	1.00	-
TAS	-52	-8.5	-	-	-	-	-	0.22	0.77	-1.84
LAV	-54	-8.9	-	-	-	-	-	0.43	0.40	-1.34
GIO	-47	-8	-	-	-	-	-	0.31	0.17	-
SVG	-55	-8.8	-15.3	6.99E-03	3.24	3.76	-30.21	0.37	0.86	-0.46
OSC	-49	-8.4	-13.4	5.08E-03	2.21	2.87	-25.25	0.25	1.13	-1.53
AMO	-47	-7.3	-	-	-	-	-	0.30	1.27	-1.22
AGG	-49	-8.5	-11.4	3.40E-03	1.55	1.85	-22.59	0.25	0.97	-1.80
SAN	-47	-8	-	-	-	-	-	0.30	0.94	-1.76
PEN	-48	-8.2	-	-	-	-	-	0.31	1.16	-1.76
MIE	-48	-7	-13.4	3.47E-03	1.69	1.78	-27.97	0.20	0.66	-2.76
PAZ	-48	-7.9	-	-	-	-	-	0.30	1.09	-1.07
ZAP	-48	-8.3	-	-	-	-	-	0.29	1.17	-0.35
CPN	-54	-8.9	-	-	-	-	-	0.28	0.69	-1.32
FAB	-45	-6.4	-13.9	6.47E-03	3.37	3.10	-31.19	0.34	0.64	-1.05
SOG	-49	-8.3	-12.3	3.01E-03	1.39	1.61	-24.65	0.25	0.18	-1.86
SAM	-51	-8.1	-11.6	3.08E-03	1.64	1.45	-26.97	0.29	1.03	-1.44
CAC	-45	-6.7	-	-	-	-	-	0.20	0.93	-2.20
CAR	-49	-8.4	-	-	-	-	-	0.27	1.09	-2.48
CAP	-45	-6.4	-13	4.53E-03	2.20	2.34	-27.11	0.23	0.41	-0.93
ALLI	-50	-8	-8.3	3.93E-03	1.77	2.16	-16.76	0.28	1.22	-0.57
PZ11	-48	-8.3	-	-	-	-	-	-0.02	0.26	-1.96
PZ6	-46	-7.6	-	-	-	-	-	0.78	1.96	-0.78
PZ5	-46	-7.5	-	-	-	-	-	0.09	0.51	-
PZ7	-48	-8	-	-	-	-	-	0.10	0.54	-
TRA	-49	-7.3	-	-	-	-	-	0.02	1.46	-
TRA	-	-	-	-	-	-	-	0.10	1.55	-
SUL	-52	-8.5	-8.3	4.87E-03	1.99	2.89	-20.79	0.21	1.07	-1.84
SUF	-53	-8.5	-11.2	4.43E-03	1.93	2.50	-21.36	0.22	1.09	-2.01