UNIVERSITÀ DEGLI STUDI DI CAMERINO

School of Advanced Studies

DOCTORAL COURSE IN

SCIENCE AND TECHNOLOGY - PHYSICAL AND CHEMICAL PROCESSES IN EARTH SYSTEMS

XXXIV cycle

Multidisciplinary analysis for unravelling physical and chemical signatures of tectonic deformation processes

PhD Student
Giulio Poggiali

Supervisors
Dr. Lauro Chiaraluce
Prof. Emanuele Tondi

Contents

1. Introduction

The study of earthquakes is, by its nature, an interdisciplinary matter. In the last years, our knowledge on rupture nucleation, evolution as well as earthquakes triggering and seismicity pattern evolution, albeit far from being complete, has gone through major steps thanks to technological advances providing a positive combination of high resolution observational data and computational processing capabilities.

From a general standpoint the earthquake generation, encompassing stress accumulation on a fault plane to the final abrupt energy release, is quite well understood, measured, and explained by classic elastic theory and frictional models (see Kanamori and Brodsky, (2004) and references therein). At the same time statistical approaches have been successfully applied to the characterization of seismicity pattern evolution, with consistent observations around the globe and on a wide range of scales. Widely known and adopted examples are the Gutenberg–Richter relation (Gutenberg and Richter, 1941), which describes the expected number of earthquakes (e.g., aftershocks) for each event with a certain magnitude in each area, or the Omori Law (Omori, 1895), describing the frequency of aftershocks decaying with the reciprocal of time after the mainshock occurrence. Yet, we are aware that the mainshock-aftershock classification represents just one of the possible models defining seismicity evolution (Mogi, 1963; Kagan and Knopoff, 1981). However, multiple theories have been successfully developed to model the stress distribution in the crust (Anderson, 1905), the rupture process (e.g. Dieterich, (1978)) and the

nucleation of earthquakes (e.g. Ellsworth and Beroza, (1995)); although we are still far from having a comprehensive view and physical models contributing to the specific processes.

The more closely we look at a particular tectonic area or fault system, the more we become aware of the increasing variability and complexity of the system, with diverse elements interacting with each other, like, among others, stress values (Stein, 1999) and their 3D distribution, mechanical properties of the different lithologies (Trippetta et al., 2010), fluids/rocks interaction processes (Miller, 2013). Details of fault zones are incredibly complex, as is their response to the loading rate, non-uniform in both space and time (Hardebeck and Hauksson, 2001). Then, the occurrence of an earthquake changes the stress and strength of nearby fault segments (Stein, 2003; Belardinelli et al., 2003; Gomberg et al., 2001). Within this context the presence of fluids in the crust can work for example to reduce the friction along the fault plane promoting earthquakes (Hubbert and Rubey, 1959; Sibson, 1973). All this corroborates the idea that crustal faults are highly heterogeneous natural systems, with physical and chemical properties varying in time and space.

Given the complexity and multidisciplinary nature of the problem, the approach of the work has been based on the observation of a (rather) large variety of data derived by the combination of multiple disciplines.

The road of integration An opportunity to study earthquakes with unprecedented detail and with a broad range of multi-parametric instrumentation is given by the recent implementation of Near Fault Observatories (NFO, `www.epos-eu.org/tcs/near-fault-observatories`). NFO are permanent monitoring infrastructures composed of very dense, state of the art networks including not only seismic but also deformation, strain, geochemical and electromagnetic equipment.

These advanced research infrastructures are built with the objective of providing novel and diverse data to better understand the multi-scale physical/chemical processes possibly associated with earthquakes and faulting. They can be considered as on-field laboratories that illuminate underlying active faults by recording and

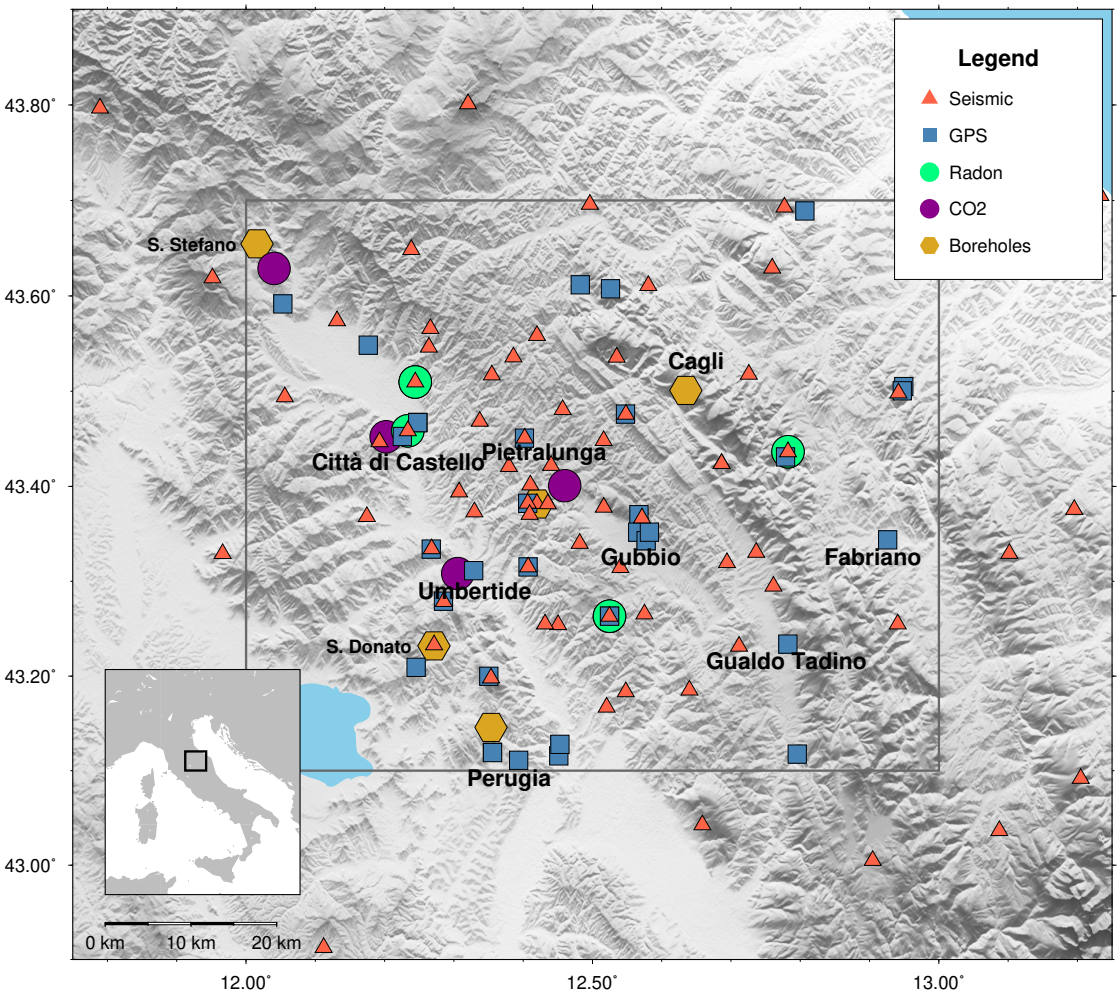

Figure 1.1: Map of the stations and boreholes. Seismic, GNSS, geochemical (CO_2 and radon) stations are reported along with existing boreholes. Rectangle corresponds to the TABOO area.

analyzing multi-disciplinary signals related to the processes that occur in the subsoil, down to very small scales. The high density of seismic networks integrated in NFOs (interstation distances as low as 5 km) allows the detection and location of micro-seismicity down to negative magnitudes. In this framework, small events (with M ∼3) can be considered as local mainshocks, considerably shortening the seismic cycle commonly involving long time scales for moderate-large earthquakes. This provides the opportunity to test models and hypotheses on a more robust statistical basis and to follow the pre-, co- and post-seismic phase of main faulting episode, even if relatively small (e.g., slipping area from tens to hundreds of squared meters). Thus, small sized earthquake sequences potentially provide much more information, contributing to the understanding of the earthquakes preparatory phase and related seismic hazard assessment.

Among the European NFOs operating on different tectonic regimes and areas over Europe, one is installed and developed in central Italy: The Alto Tiberina Near Fault Observatory Chiaraluce et al., (2014)).

Seismotectonic setting TABOO is located at the Tuscany–Umbria–Marche regions boundary within the northern Apennines (Figure 1.1), a NE-verging thrust-fold belt undergoing NE-trending extension at a rate of about 3 mm/yr (Serpelloni et al., 2005). The complex tectonic architecture of the crust is the result of two phases of eastward migrating deformation: an early compression with eastward directed thrusting and a later phase of extension (Elter, 1975; Pauselli et al., 2006). The latter started in the upper Pliocene and is still active in the inner part of Apennines (e.g. Elter, (1975) and Barchi, (2010), and reference therein). During the first compressional phase, the sedimentary cover was deformed in east verging folds and regional-scale thrusts, including a crustal doubling confirmed by borehole data and observed in seismic reflection profiles (Mirabella et al., 2011).

Several studies (e.g. Barchi, (1998) and Boncio et al., (2000)) have documented the presence of a 60 km long, ENE dipping low-angle normal fault (LANF), the Altotiberina Fault (ATF). The ATF is currently the subject of several geological and geophysical studies, being recognized as one of the rare examples of seismically

active low-angle normal faults (Chiaraluce et al., 2007; Hreinsdottir and Bennett, 2009; Vadacca et al., 2016; Valoroso et al., 2017). The main reasons for interest and scientific debate on LANFs is that these faults are characterized by very low dip angles (<30°) and according to classical fault mechanics (i.e., faults in an elastic crust obeying Coulomb friction) these structures should not be formed and/or develop (e.g., being active) in extensional environments characterized by an Andersonian stress field (vertical maximum principal stress).

The ATF has potential for M7 earthquake in case of a rupture of the whole fault length, but its hazard, similarly for the other LANFs worldwide, is still controversial, because there is no evidence of well documented occurrence of moderate-large magnitude events worldwide occurring on such misoriented faults. On the same line there is the lack of evidence of moderate-large earthquakes in the Italian historical catalog in the past 1000 years (Rovida et al., 2011). It is also worth noting that the microseismicity nucleating on the ATF is not able to explain the amount of deformation associated with the short- and long-term slip rate inferred by geological (Collettini and Barchi, 2004) and geodetic studies and data (D'Agostino et al., 2009).

On the contrary, the active role of the ATF in the ongoing extensional deformation across the Apennines has been clearly showed by GNSS observations Hreinsdottir and Bennett, (2009) and Vadacca et al., (2016), suggesting creeping below 5 km of depth as main driving mechanism. This hypothesis is so far coherent with the continuous occurrence of micro-seismicity along the ATF plane and it is also supported by laboratory experiments performed on fault rock samples of the Zuccale low-angle normal fault, considered the (older) exhumed analogue of the ATF (Collettini et al., 2009b). Unfortunately, all these lines of evidence cannot rule out the occurrence of a large earthquake along the ATF potentially nucleating along one of the steeper portions of the fault where along dip staircase trajectories are observed (Valoroso et al., 2017; Mirabella et al., 2011).

Above the ATF, synthetic and antithetic deformations acting on the tectonic pile dissected the main geological units, which, following Mirabella et al., (2011), can be summarized in: (1) the top of the basement, mainly clastic and metased-

imentary rocks of late Paleozoic-middle Triassic age; (2) Triassic Evaporites; (3) a carbonate multilayer platform, lower Jurassic-Oligocene; and (4) early Miocene foredeep turbidites. The area has been object of extensive studies with a number of complementary techniques, thus contributing to a strong background knowledge of the geologic and tectonic setting: e.g. high quality seismic catalogs (Valoroso et al., 2017), seismic tomography (Piana Agostinetti et al., 2017; Moretti et al., 2009), geological (Mirabella et al., 2011) and deterministic velocity models (Latorre et al., 2016) built on the integration of field surveys, seismic reflection profiles and boreholes data.

Tectonic extensional systems and fluids The extension is currently active in a $\sim$ 30 km wide longitudinal zone along the chain where most of recent and historical seismicity is located Figure 1.2. To the west of the active area, extension was active for enough time (>3 Ma) to change the properties of the lithosphere, producing a widespread heat flow anomaly 90 mW/m^2 (Mongelli and Zito, 1991) and a thin crust (20–25 km) (Ponziani et al., 1995). The region characterized by tectonic extension is also affected by widespread and vigorous CO_2 degassing (Chiodini et al., 2004), whose isotopic composition suggests a deep origin due to metasomatization of mantle rocks. CO_2 presence has in fact been measured in San Donato and Santo Stefano boreholes (reported in Figure 1.1). Here a high fluid overpressure at 85% of the lithostatic load have been reported (Chiodini et al., 2004) at depths of 4.8 km and 3.7 km, both within the same geological formation, i.e. the Triassic Evaporites, suggesting that in this area fluids can accumulate in shallow crustal traps, feed the high-CO_2 flux observed at the surface and possibly trigger seismicity.

The main mechanism relating fluids and earthquakes is that high pressure fluids trapped in the lithosphere can escape primarily by large-scale and transient changes in permeability accompanying local fracture and crustal scale earthquakes. If earthquakes provide a trigger for fluid escape, then the pathway taken by the fluids will influence aftershock sequences because the high fluid pressure reduces the effective normal stress while fluid pressure gradients introduce pore-elastic stresses

on the system (see Miller, (2013), where the argument is extensively treated along with the whole geodynamical cycle of fluids in the lithosphere). The hypothesis that high pressure fluids act as an important tectonic driver suggests that lithospheric deformation, and the earthquake cycle, is substantially controlled by the mechanical, chemical, and time-varying hydraulic behavior of high pressure fluids trapped at depth.

In the context of the Apennines, several authors suggest that fluids play a major role in seismogenesis (Chiodini et al., 2004) and in the spatiotemporal evolution of seismicity (Miller et al., 2004). On this line of research there is an increasing evidence, reported by multiple authors, of fluids involvement in recent seismic sequences that occurred in the northern-central Apennines: Umbria-Marche 1997 (Miller et al., 2004; Antonioli, 2005; Lombardi et al., 2010), L'Aquila 2009 (Terakawa et al., 2010; Lucente et al., 2010; Di Luccio et al., 2010; Malagnini et al., 2012), Central Italy 2016 (Chiarabba et al., 2018; Malagnini et al., 2022).

Given the aforementioned large flux of CO_2 from deep source and the overpressure measured in boreholes, several authors suggest that fluids can affect seismic activity in the TABOO area (Valoroso et al., 2017; Marzorati et al., 2014; Piana Agostinetti et al., 2017). This specific area presents indeed a very favorable lithologic and tectonic setting for channeling and trapping CO_2 rich crustal fluids: the presence of the Triassic Evaporites, composed of dolostones and anhydrite layers, suggest a perfect combination of reservoir (dolostones) and sealing (anhydrites) horizons for CO_2 crustal fluids (Trippetta et al., 2013). The primary role of this geological unit in hosting fluids traps is corroborated by laboratory tests, which confirm the very low permeability of the anhydrides, even during deformation (De Paola et al., 2009; Collettini et al., 2009a; Trippetta et al., 2010).

ATF and seismic activity The vast majority of seismic activity is located in the ATF hanging wall, nucleating on minor synthetic and antithetic normal faults (4–5 km along-strike length) and giving rise to seismic sequences with multiple Mw3+ mainshocks (Valoroso et al., 2017), while the ATF footwall is almost aseismic (Chiaraluce et al., 2007).

From 2010 up to now the seismicity occurred in the area (Figure 1.2) can be summarized in three seismic sequences: 1) the Pietralunga sequence, acting with bursts of seismicity over multiple years (2010, 2013 and 2014); 2) the Città di Castello sequence (April and May 2013); 3) Gubbio sequence, active mainly in 2013-2014, but reactivating in the the following years, up to the most recent activity occurred in May 2021. Previous seismic activity recorded in this sector of the Apennines include the 1997 Colfiorito M_W 6.0 earthquake, a few kilometers SE of the study area, 1998 Gualdo Tadino M_W 5.1 and 1984 Gubbio M_W 5.1 events. All these earthquakes activated SW-dipping normal fault systems, thus antithetic to the ATF. The historical earthquakes catalog of the area (Rovida et al., 2011) reports 8 $M_W \geq 5.5$, three of which had $M_W \geq 6.0$ (see Figure 1.2).

1.1 Scientific questions and motivation

The Upper Tiber Valley area presents a compelling seismotectonic setting, with many open questions to be investigated, along with an optimal scientific background and monitoring environment to study such topics with a broad perspective and a high level of detail. The complex seismotectonic setting of the Northern Apennines offers a unique and large combination of ingredients, all potentially showing key signatures of the earthquake's nucleation and preparatory phases. Here there is a constant and high rate of microseismic release, evidence of fluids overpressure at depth, well- and mis-oriented normal faults (respect to the regional extensional stress field), marks of seismic and aseismic activity along them and a state-of-the-art multidisciplinary network monitoring all these components at high resolution in an area whose lithological distribution in 3D is very well constrained. Thus, here there is the potential to investigate a wide range of key topics in modern seismology and fault mechanics, by the integration of seismic with non-seismic data.

That is why this thesis is based on the idea that the joint analysis of all this information can contribute to shed light on earthquake initiation and the mechanisms driving the recurrence and evolution of the seismic sequences. The main questions that I intend to investigate concern signals of various nature and

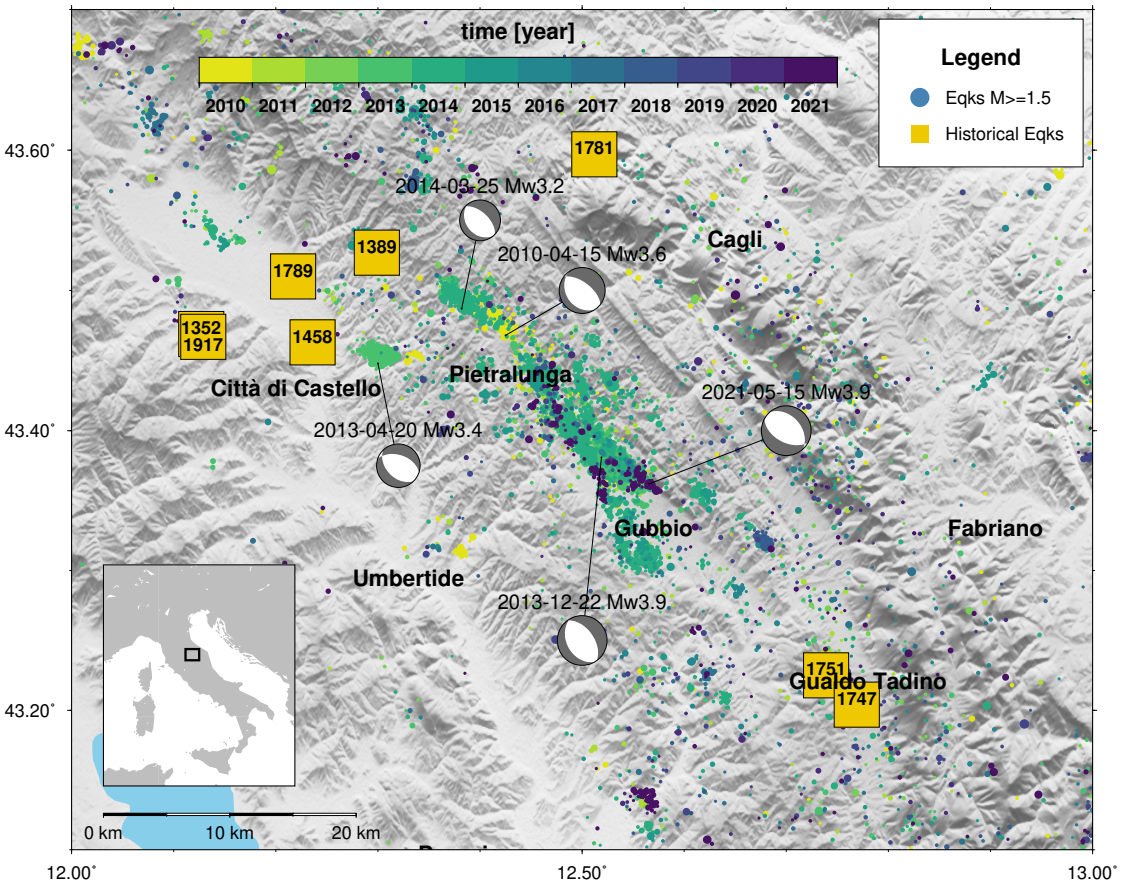

Figure 1.2: Historical and recent instrumental seismicity of the TABOO area. Epicentral location of the earthquakes occurred between 2010-2022 (from INGV catalog `http://iside.rm.ingv.it`) and largest events occurred in the past 1000 years (magnitudes 5.78 to 6.44 from Rovida et al., (2011)).

form that can possibly be detected and traced back to tectonic and seismic processes, giving hints on the driving mechanisms. Are there detectable signals/anomalies linkable to seismic activity in the heterogeneous time series recorded by the multi-parametric instrumentation present in NFOs? If so, what are the relations with local earthquakes and between different datasets? How to compare the heterogeneous time series? And more: what is the role of crustal fluids in the seismic sequences of the area? Are there any identifiable seismic patterns that suggest such interactions? Do local earthquakes nucleate independently of lithology or are some layers more prone to trigger seismicity? These questions are addressed in the main chapters of the thesis, the structure of which is reported in the following.

1.2 Thesis outline

The performed work is presented based on two scientific papers (Chapter 2 and 3) to be submitted to peer reviewed journals. The recurring theme is the multidisciplinary approach revolving around earthquakes and innovative tools to refine and broaden the range of inferences and connections between a variety of data. I address the topic with two complementary paths: the investigation of signals related to active tectonic processes, in seismic and non-seismic data, and a detailed analysis of seismicity to highlight if/how fluids may have interacted with tectonic processes. Additional tests and analyses I performed, not included in publications, are reported in the Appendixes. Finally, I have drawn a general Conclusion of the work (in Chapter 4), summarizing all the work done, the inferences but also and especially the ideas for future work, which is so much but also very promising.

Chapter 2 presents the framework developed for the analysis of multidisciplinary time series applied to the data recorded by TABOO's network. The joint analysis of different datasets is a complex task requiring a methodology to extract and compare useful information from all the different sources in a coherent framework. Here a novel approach is presented, based on Bayesian inference, that relies on a reversible jump Markov chain Monte Carlo algorithm (rj-McMC) to independently model time

series datasets and extract variations/transients under the form of changepoints distributions as a common comparable output. The proposed workflow is applied to V_P/V_S, GNSS and geochemical (Rn and CO_2) time series. For each dataset the output of the analysis is discussed in comparison with seismicity rate and main earthquakes recorded in the study area, looking for a direct connection, (if one can say type of cause and effect), between these data. The analysis of the results is done for single station outputs, multi-station and finally with multidisciplinary comparison of all analyzed dataset from the different disciplines.

Chapter 3 presents an in depth analysis of spatiotemporal patterns of small seismic sequences occurred in the study area, to test the hypothesis of fluid involvement in the process. The earthquakes catalog for Città di Castello and Pietralunga sequences have been built starting directly from raw seismic waveforms to high-resolution hypocentral locations, exploiting the dense TABOO seismic network via the most advanced deep learning phase picking techniques, to detect micro-seismicity with unprecedented resolution. The seismicity patterns are modeled as diffusive processes to investigate the role of fluids in the evolution and triggering of seismicity. The earthquake catalog produced is compared with a detailed seimostratigraphic three-dimensional model of the area, to frame the seismic activity in the geologic and tectonic context allowing to discuss the implications of the hypothesized fluid driven mechanisms.

Chapter 4 synthesizes the results with a discussion about the implications of the presented studies to provide an outlook on future research directions building upon the results of this thesis.

2. Looking for signatures of tectonic processes in multidisciplinary time series by means of Markov chain Monte Carlo approach: TABOO Near Fault Observatory case study

2.1 Introduction

Understanding earthquakes, from the preparatory phase to the (slow or fast) energy release, is an ambitious task that has to deal with the complexity of natural systems, whose mechanical, physical and chemical properties evolve over time. Investigating the relations between earthquakes and various measurable parameters is not new: for at least fifty years much effort have been put in trying to better understand earthquake generation through the analysis of a wide variety of measurements, mostly in search of precursors (Scholz, 1973; Cicerone et al., 2009; Thomas, 1988). Despite the popularity of the problem, it has been rarely faced with a real multidisciplinary approach involving the analysis of multiple different observables at the same time for the same area.

In the last decade the development of research infrastructures such as Near Fault Observatories (NFO) allowed to conduct multidisciplinary experiments in "natural laboratories", created expressly with the aim of understanding the physical-chemical processes related to earthquakes, tectonic activity and evolution of fault systems. One NFO has been installed/developed in central Italy, the Alto Tiberina Near Fault Observatory (TABOO) (Chiaraluce et al., 2014) (Figure 2.1) This NFO is located in the upper Tiber Valley (northern Apennines) and it is equipped with a wide variety of instrumentation, providing continuous acquisition of long time series from high resolution networks of seismic, geodetic and geochemicals. This area has unique potential for multidisciplinary studies not only because of the dense instrumentation and high seismic rate, but also because of the peculiar seismotectonic framework involving aseismic deformation (Gualandi et al., 2017) and deep fluids circulation (Chiodini et al., 2004) in this section of the Apennines. The area has been investigated with a number of complementary techniques: seismic imaging and monitoring, geodetic and geochemicals.

From a seismological point of view, many studies confirmed that the vast majority (~90%) of seismic activity originates in the hanging wall of a low angle normal fault, the Altotiberina fault (ATF) (Chiaraluce et al., 2007), where a complex system of synthetic and antithetic faults give rise to seismic sequences

with multiple Mw3+ mainshocks (Valoroso et al., 2017). This extensive seismic activity and dense network allows to have a detailed picture of the subsurface from high quality seismic catalogs (Valoroso et al., 2017), tomographic (Moretti et al., 2009; Piana Agostinetti et al., 2017) and deterministic (Latorre et al., 2016) velocity models.

Recorded earthquakes can also be exploited to "monitor" elastic properties using travel times of seismic phases to detect variations which are indications of the processes taking place in the subsurface (Lucente et al., 2010; Li et al., 1998; Schaff David P. and Beroza Gregory C., 2004). An example of this approach on the TABOO area is described in Poggiali et al., (2019) using V_P/V_S time series: here most of the variations detected are clustered in space and time near the mainshock of the Gubbio sequence (Dec. 2013).

Another interesting aspect of this area involves geodetic data, which are widely used to monitor crustal deformation and understand seismic and aseismic slip (Kanamori, 1977; Avouac, 2015). Many studies suggest an active role for the ATF in accommodating tectonic extension in this sector of the Apennines, which is supported by GNSS data with numerical models (Vadacca et al., 2016) and it is compatible with creeping behaviour (Hreinsdottir and Bennett, 2009; Anderlini et al., 2016). The high density of geodetic stations allowed also to detect an aseismic transient associated with slow deformation (Gualandi et al., 2017) during the occurrence of a shallow seismic sequence (Gubbio 2013, mainshock $M_W = 3.8$).

Recent studies focused on the geochemical aspect of the fluids that circulate at the main fault depth-level. Interactions of deep fluids of mantle origin (CO_2) with tectonic activity are suggested by several authors as a triggering mechanism (Chiodini et al., 2004), a factor controlling the spatial and temporal evolution of seismicity (Miller et al., 2004; Antonioli, 2005), and a proxy for crustal stress (Camarda et al., 2016). The existence of fluid circulation in the area is documented both in deep boreholes, where fluid overpressure (CO_2) at about 85% of lithostatic load has been encountered (Chiaraluce et al., 2007), and in very high CO_2 flux emissions at the surface (up to 5800 t/yr), values comparable to active volcanic regions.

Geochemical instrumentation installed in the area also includes monitoring of radon emissions (Piersanti, 2015). This radioactive gas have been investigated for decades in relation to earthquake generation (see Cicerone et al., (2009) for a review), and, although relations with seismogenetic processes are still open issues, laboratory experiments show increased radon emissions as a consequence of deformation in rocks (Holub and Brady, 1981; Mollo et al., 2011).

This study is based on the idea that the joint analysis of all these information could shed a light on the mechanism of earthquake nucleation at an intermediate scale (between laboratory and disctructive earthequakes). However, comparing and relating all the different datasets produced by this multidisciplinary networks with tectonic processes has to deal with two levels of complexity: the development of a methodology to extract useful information from all the different sources in a unique and coherent framework; and the need of extracting quantitatively comparable information.

Here we propose a workflow, based on Bayesian inference, that relies on a reversible jump Markov chain Monte Carlo algorithm (rj-McMC) (Green, 1995) as a framework to independently model all the time series with piecewise functions and extract variations/transients which can be finally represented and analyzed as changepoint (CP) distributions. The ability of Bayesian methods to deal with non-linear and non-unique problems (Mosegaard and Tarantola, 1995; Tarantola and Valette, 1982), made these methods appealing in many geophysical fields and the trans-dimensional implementation (Green, 1995) has proven successful in a wide variety of applications (Bodin et al., 2012b; Bodin et al., 2012a; Dettmer et al., 2010; Hawkins and Sambridge, 2015; Piana Agostinetti et al., 2015).

Trans-dimensional algorithms are particularly well suited to deal with CP problems, because they allow the treatment of the number of CPs as a parameter itself to be estimated, without requiring subjective a priori assumptions on the model complexity. Examples of applications on various kinds of time series can be found in literature: borehole temperature data (Hopcroft et al., 2007; Hopcroft et al., 2009), geochemical data (Gallagher et al., 2011), thermochronology (Gallagher 2012).

rj-McMC algorithms, like the one used in this study, are naturally parsimonious (Malinverno, 2002): this assures that only CPs supported by the data are retrieved and that simpler models are naturally preferred over complex models. Model complexity is also directly linked to the level of data noise and the correlation of data points (Piana Agostinetti and Malinverno, 2018). To limit the effect of wrong noise estimations we adopted the "hierarchical Bayes" approach (Malinverno and Briggs, 2004) and let the algorithm estimate a scale factor for errors on the measurements. Furthermore we added a correlation parameter for data noise that can be estimated in the inversion, following the treatment of Bodin et al., (2012a). The latter addition is made to handle the case of correlated measurements that, if assumed uncorrelated, may also result in an unnecessary increase of the model complexity.

The aim of the work is to develop and apply a rj-McMC algorithm for the detection of CPs hidden in different geophysical measurements acquired in the study area by means of a Bayesian algorithm, and to produce an output which can easily be used for integrate multidisciplinary datasets, in a synthesis effort. All analyzed data-sets will be represented in the form of time series, where piecewise functions will be selected ad-hoc for simulating the observed data. Given the parsimonious nature of our approach (Malinverno, 2002), our algorithm will make use of a limited number of piecewise functions for each data-set, i.e. including only functions directly supported by the data. The occurrence of a discontinuity between two functions, a CP, will indicate the presence of a discontinuity in the time series. Collecting and analyzing all discontinuities between all simulated datasets will allow to relate variations detected at multiple sites and potentially involving different parameters. The proposed method could therefore contribute to innovative and heterogeneous warning systems for monitoring tectonically active regions and aid expert opinion in relating and comparing the different signals recorded by the increasingly widespread multidisciplinary networks. This kind of approach can also broaden our knowledge about the existence of relations between earthquakes and various measurable parameters with potential significance as a proxy for the local tectonic process. In fact this study involves both parameters for which links with

seismicity are known, like geodetic and seismic, and parameters for which the links are debated, like radon, or supposed, like CO_2, especially for this area.

The first step in our workflow consists in a pre-processing phase which is necessary for some of the datasets before the application of the rj-McMC algorithm. Different data-sets require different levels of data pre-processing procedures, going from almost nothing (e.g. time series of V_P/V_S ratio), to complex pre-processing workflows including external data/constraints (e.g. geochemical data). For V_P/V_S this first step consisted mainly in building the time series from raw datasets: we followed the same cluster-station logic as described in Poggiali et al., (2019) to accomplish this task. We recognize that geodetic measurements exhibit the presence of seasonal signals which can severely influence the results if not taken into account (Serpelloni et al., 2006): seasonal signals (annual and semiannual) are simulated and the parameters are estimated following a Bayesian approach, before the subsequent rj-McMC modeling. For geochemical measurements, already in a suitable time series format, the main pre-processing step consisted in mitigating the effect of environmental factors known to affect the measurements. In order to remove the most common sources of short-term anomalies not related to tectonic activity, namely rainfall (see, among others, Granieri et al., (2003) for CO_2 and Piersanti et al., (2016) for radon), we implemented a simple but effective filter based on rainfall threshold in a moving window. Moreover raw datasets with hourly sampling (geochemicals) are resampled on a daily basis to remove eventual daily and sub-daily oscillations. After the preparation of all datasets, the McMC algorithm is applied independently on each time series. The collection of CPs obtained for each data is analyzed *per se* and relative to the other time series. Additionally, the time series are compared to the seismicity rate and other signal of the tectonic process.

2.2 Data and time series pre-processing

The multidisciplinary time series analyzed in this study are derived from both raw data and scientific products gathered by the TABOO research infrastructure.

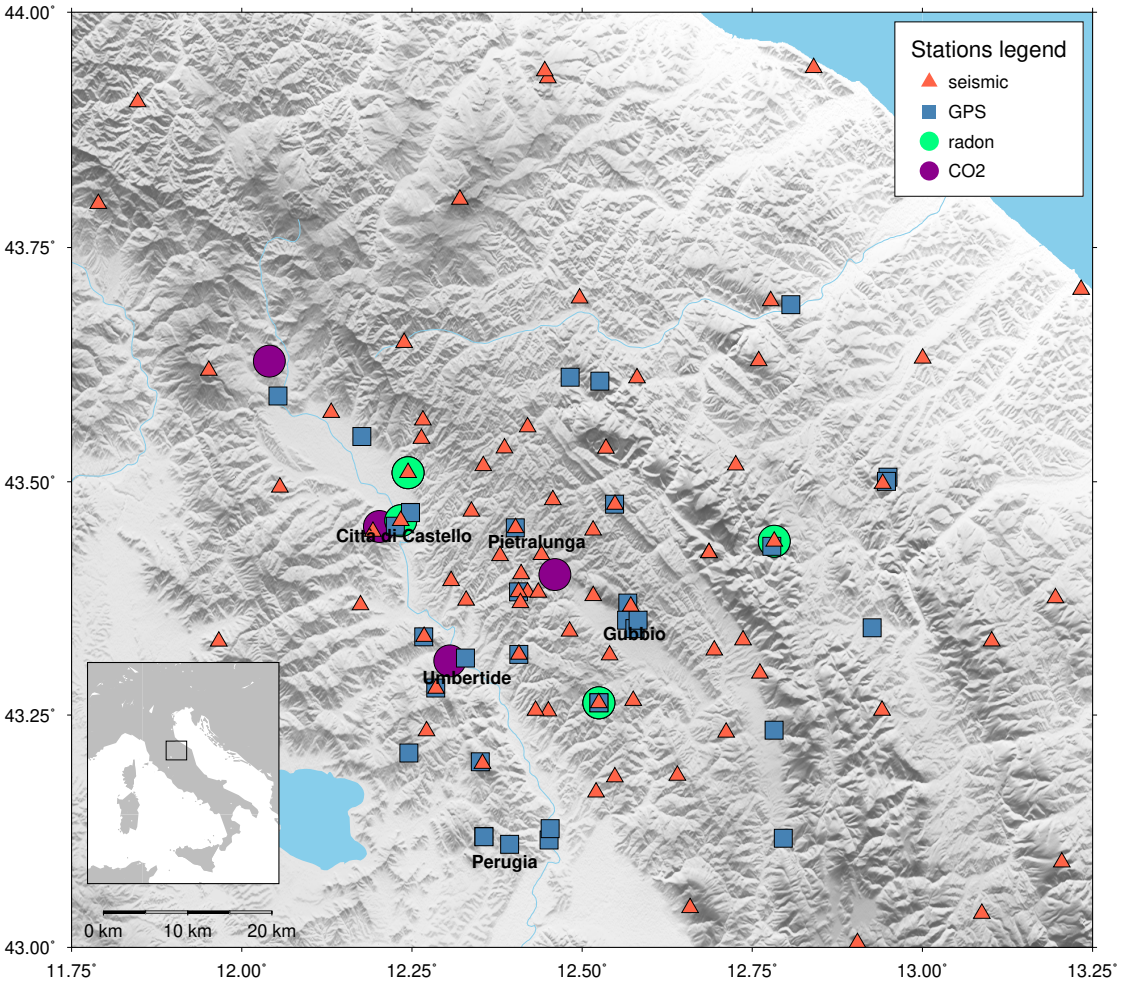

Figure 2.1: Map of the study area. Only stations used in this work are reported.

Datasets include time series of geodetic, geochemical (CO_2 flux and radon emissions) and V_P/V_S ratios for a maximum timespan of 10 years (January 2010 to December 2019). A crucial step in searching for variations related to tectonic processes is defining what kind of signals we are interested in. This approach would require us to know the actual relations between every observable we analyze and the tectonic processes taking place. Which is unfortunately unknown. To overcome this limitation we chose the more viable option of defining which are the signals that we are *not* interested in before the CP detection phase. With this strategy we aim to remove or isolate the contributions which are known to affect the measurements but are not related to seismic activity. This includes mainly seasonal signals (present in geodetic and geochemical data) and short-term effects due to meteorological parameters such as rainfall (observed in geochemical data).

Datasets have very different properties from each other and have required a variable amount of preparation/preprocessing in order to be analysed with the McMC algorithm: V_P/V_S data needed to be spatially organized in time series, GNSS data required the removal of annual and semiannual periodicity (Serpelloni et al., 2006) (without considering the creation of positional time series, not treated here); geochemical data have been filtered and resampled. In the following, we explain in details all the preprocessing steps for each data-set.

2.2.1 Global Navigation Satellite System data

We analysed horizontal components of continuous GNSS time series from stations located inside the TABOO area (Figure 2.1). These time series represent the evolution of each site position in time relative to Eurasian plate over a maximum of 10 years from January 2010 to December 2019. For a detailed description of the procedure adopted to produce GNSS positional time series used in this work we refer to Serpelloni et al., (2006). In a first moment, the instrumental offsets are removed. Then, a standard McMC search (Sambridge and Mosegaard, 2002) is operated to define the most probable values of the annual and semiannual periodicity which generally affect GNSS time series (Serpelloni et al., 2006). In details, we consider the model:

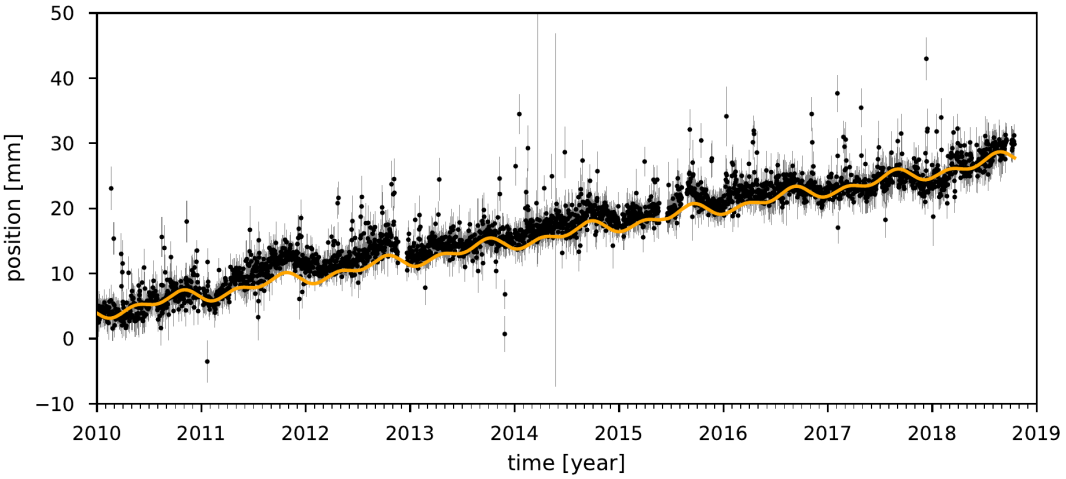

Figure 2.2: Example of data from the E-W component of GNSS station ARCE (black points). Estimated sinusoid (annual plus semiannual components) superimposed in orange (arbitrary intercept and slope).

$$p_{gps}(\mathbf{m}) = \alpha_1 sin(2\pi\gamma_1 t + \beta_1) + \alpha_2 sin(\pi\gamma_2 t + \beta_2)\,, \tag{2.1}$$

where: $p_{gps}(\mathbf{m})$ are the periodic contributions in the GNSS series that we wish to infer, $\alpha_{1,2}$, $\beta_{1,2}$ and $\gamma_{1,2}$ are respectively the parameters of the annual and semiannual terms (t is measured in years). $\gamma_{1,2}$ have values between 0.9 and 1.1 and are used to adjust annual and semiannual periods, so that they are not strictly annual and semiannual. The mean posterior values of $\alpha_{1,2}$, $\beta_{1,2}$ and $\gamma_{1,2}$ are used to estimate the periodic signals. After the seasonal estimation, time series are analyzed with the rj-McMC algorithm to search for sudden changes. For each GNSS station our data consist of position time series and associated measurement error of the two horizontal components (E-W and N-S), sampled at 1 sample per day (example of E-W component in Figure 2.2). Time series shorter than 2 years are not considered. We used a total of 26 GNSS stations in the area (Figure 2.1).

2.2.2 Geochemical data

Although environmental / meteorological parameters (e.g. temperature, pressure, rainfall, etc.) are known "exogenous" factors in the modulation of geochemical signals (see, among others, Pinault and Baubron, (1996) and Camarda et al., (2019) and references therein, respectively for radon and CO_2), the specific role and contribution of each one is still an open issue, often further complicated by site effects. Nonetheless it is possible to focus on a subset of factors, widely recognized in literature for having a major influence in the modulation of geochemical emissions, and implement strategies in the workflow in order to mitigate at least the effect of these known sources of anomalies. Furthermore the pre-processing we used on geochemical datasets is functional only to the subsequent application of the rj-McMC algorithm, that is to remove the main sources of CPs not accountable to tectonic activity. Thus, we are not interested in identifying, for example, the cause of a seasonal oscillation, but only to account for it in our model.

CO_2

CO_2 dataset analyzed in this study consists of CO_2 flux measures from 4 stations: Fungaia, Migianella, Nogna, Uppiano (Figure 2.3). The stations are positioned in sites characterized by strong CO_2 flux, up to 5800 t/year (Camarda et al., 2019), evidenced in some sites by the presence of bubbling water at the surface. The timespan considered in this study goes from June 2015 to December 2019, with almost 3 years covered by all four stations. In each station the data are sampled hourly. We first proceeded by removing measurements related to known instrument malfunctions (i.e. Migianella data from the end of May to the end of July 2016). Because of the very wide range of values, spanning 4 orders of magnitude, we worked on the log transformed data (base 10 logarithm). This simplifies the McMC sampling by "compressing" the data range. Many studies investigated the relations between meteorological parameters and CO_2 emissions, (from our knowledge these studies are all performed in volcanic areas) highlighting a complex behavior, often site specific, but always characterized by long term seasonal oscillations and short

term effects related to rainfall and soil water content (Granieri et al., 2003; Camarda et al., 2019; Viveiros et al., 2008).

In order to remove anomalous signals related to strong precipitation episodes we implemented a simple filter based on a threshold over the cumulative rainfall calculated in a moving window. We used a threshold of 1 mm of rainfall and a window length of two days. These values were found with empirical criteria: 1) keeping most of the data points (more than 75% of considered days are retained with such parameters, counted from first to last valid data) 2) removing the clear spikes/drops associated with strong rainfalls (see Figure 2.4 for an example of how this filter works on a dataset). Being the cumulative sum for each sample calculated over the previous 2 days, the moving window has the side effect of accounting for some time for soil humidity to recover from wet conditions. For our purposes we preferred this way of treating the effect of rainfall over, for example, subtracting a regression function modeled on soil humidity, both for simplicity reasons and to manipulate the datasets as little as possible. Moreover removing data points does not pose any problem for the McMC algorithm because these methods can deal with data gaps by simply tending to the prior in such no-data zones, without any need of interpolation.

After the removal of data points attributable to rainy events, we resample each dataset at 1 sample per day, taking the mean and the standard deviation respectively as our final measurements and errors to be analyzed with the McMC algorithm. Periods with no meteorological data available were not considered: e.g. Nogna meteorological station unfortunately reported malfunctions until November 2017, so a large fraction of this dataset was discarded. The effect of daily and/or sub daily cycles, frequently documented in literature for CO_2 time series (Granieri et al., 2003; Rinaldi et al., 2012), is bypassed here because of the daily resampling that removes these high frequency oscillations.

As seen in geodetic time series, seasonal phenomena can generate periodic fluctuation in the geochemical time series. Even in this case, we can estimate the parameters related to a periodic function (see Eq. 2.1 during the pre-processing steps to isolate such contribution before the rj-McMC step. In particular, we adopt

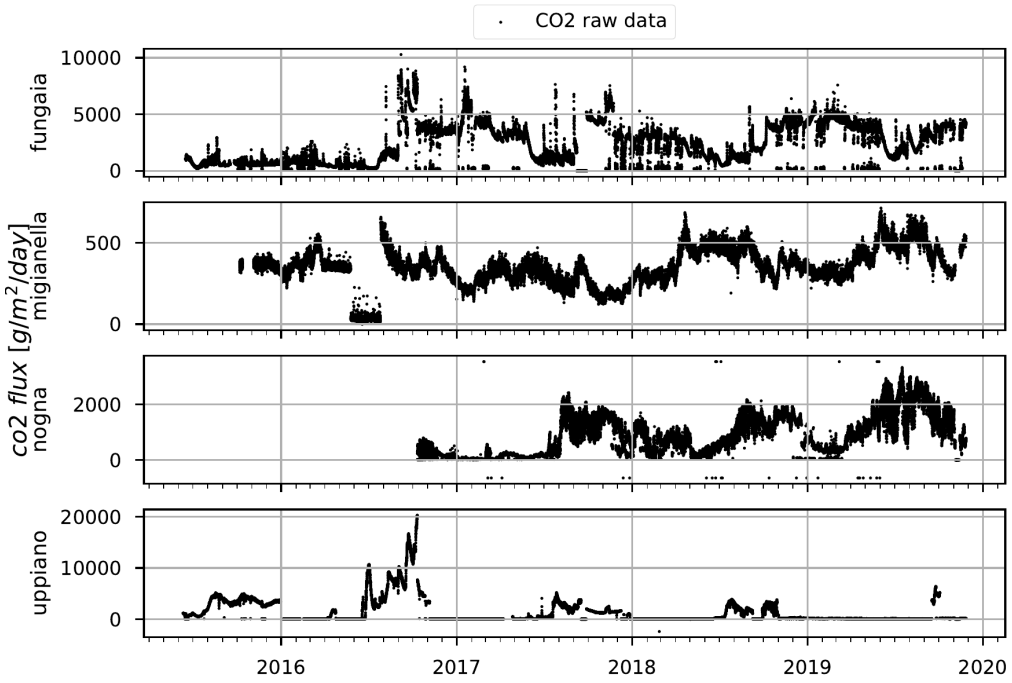

Figure 2.3: CO_2 raw dataset.

the same scheme used in Section 2.2.1 and we consider a periodic model with annual and semiannual periodicities. Parameters related to such model are estimated using a standard McMC approach. In the following rj-McMC application, mean posterior values of such parameters are considered. An example of the application of this procedure to one CO_2 station (Fungaia)is reported in Figure 2.5.

Finally we expect some degree of serial correlation between the measurements. In the assumption of uncorrelated samples, this would lead to an unnecessarily high complexity of the resulting models as a consequence of the wrong assumption. We deal with this property of the dataset by introducing a correlation parameter in the parametrization, as described in the method section.

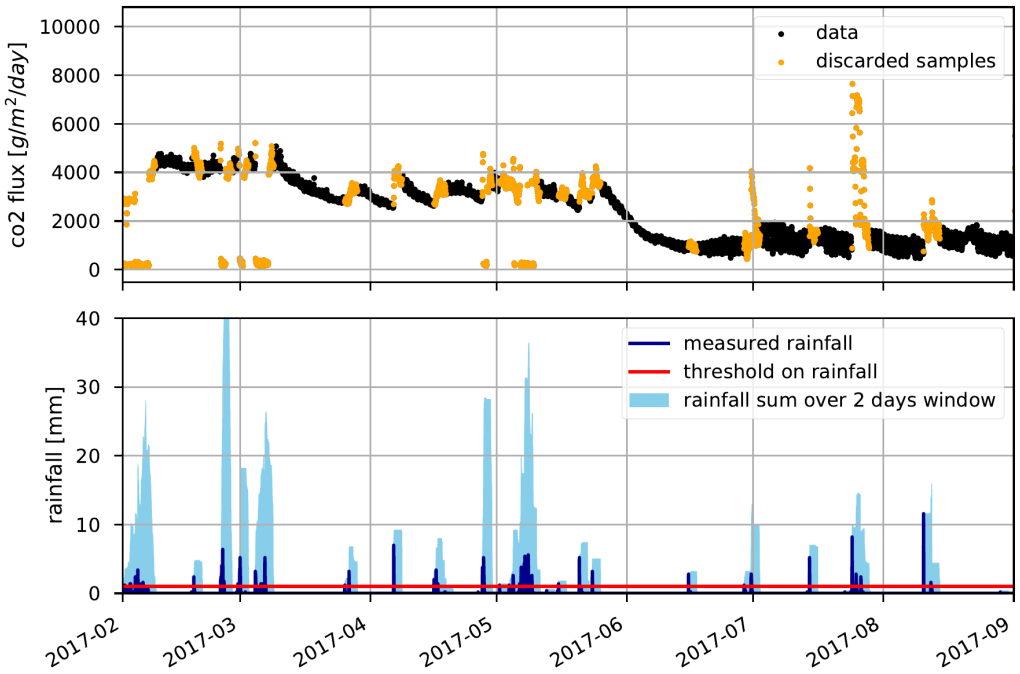

Figure 2.4: A zoom on Fungaia CO_2 station: here the relation between rainfall (bottom) and CO_2 measurements (top) is evidenced. The implemented filtering, based on a simple rainfall threshold in a moving window, is effective in identifying anomalous data points (marked in orange in top panel) that will not be considered in the following steps. We also note that the effect of rain is not always the same but results in peaks and drops of the measured CO_2 flux.

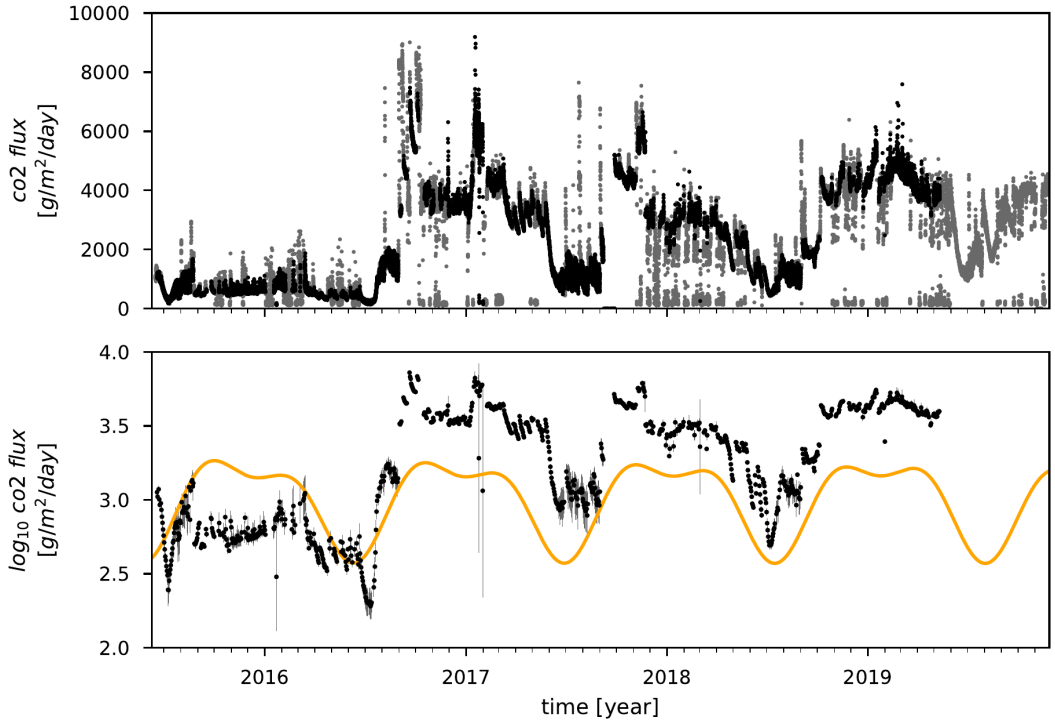

Figure 2.5: Example of the pre-processing steps for a CO_2 station (Fungaia). Top: raw data in grey and filtered data in black. The grey points are discarded measurements due to rainfall events or lack of meteorologic recordings (e.g. at the end of the time series). Bottom: Log transformed data, daily resampled with errorbars calculated as daily std. The sinusoid estimated with annual plus semi-annual components is superimposed in orange (arbitrary intercept and slope).

Radon

Radon emissions that we used as raw data are continuous measures of concentrations recorded by 4 stations located inside the TABOO area: BADI, CDCA, MURB, SSFR (Figure 2.6). The stations are part of the italian radon monitoring network (IRON); details about station sites, installations types and instrumentation can be found in Cannelli et al., (2018). The instruments used in each station use an acquisition window of 2 hours to obtain a radon concentration reading, so every time series is sampled at 12 samples per day. The timespan we analyzed goes from the end of 2013 to the beginning of 2019, with a good coverage for most of the considered time window and a maximum extension of more than 4 years of measurements.

As with CO_2 time series we first deleted the few data points related to instrument malfunctions and log transformed the values. In fact, the stations exhibit a wide range of variability: from stations with lower and more stable measurements (like CDCA and BADI) to the much higher mean and variance of SSFR, which exhibit also spikes with values over 5000 Bq/m^3 (Figure 2.6). Radon measurements are the effect of very complex interactions between several factors, and the influence of meteorological parameters is widely documented (Pinault and Baubron, (1996) and Siino et al., (2019) and references therein). Both short (daily and sub-daily) and long (annual and semiannual) oscillations have been highlighted (Siino et al., 2019; İnan et al., 2012) and a strong effect of rainfall on the measurements is constantly evidenced (Piersanti et al., 2016; Cannelli et al., 2018; Tommasone Pascale et al., 2015), frequently marked by peaks and/or drops in radon time series data. Moreover the correlations between meteorological parameters and radon emissions exhibit a site-specific behaviour (Piersanti, 2015; Piersanti et al., 2016).

We applied the same filtering that we used for CO_2 data to remove samples affected by rainfall episodes. The moving window has been kept 2 days long and the threshold has been increased to 10 mm with an analogue procedure based on keeping most of the data and removing anomalous signals observed in conjunction with heavy rain events. Some examples of peaks related to rain events detected in CDCA and BADI stations are further explored in Cannelli et al., (2018). After

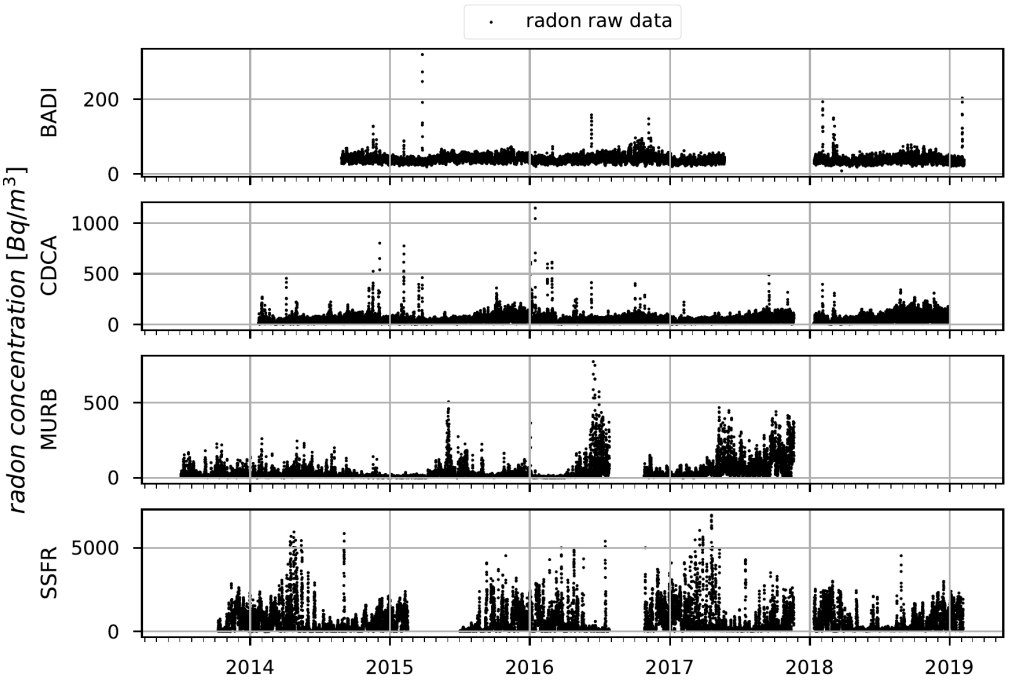

Figure 2.6: Radon raw dataset.

daily resampling of the remaining data, ~90% of days are retained with respect to the original datasets (counted from first to last valid data). As with the previous CO_2 case, we take the daily mean and standard deviation as the final values and associated errors that will be the input of the McMC algorithm. The resampling procedure removes daily (and sub-daily) periodicity evidenced in literature, which are generally attributed to temperature and pressure cycles.

The seasonal parameters are estimated for Radon time series following the same approach presented for GNSS and CO_2. Again, this implies a preliminary application of a standard McMC scheme to a given model. For consistency, we use the same model as in GNSS and CO_2 (annual and semiannual periodicities only). An example of the application of this procedure to one radon station (CDCA)is reported in Figure 2.7.

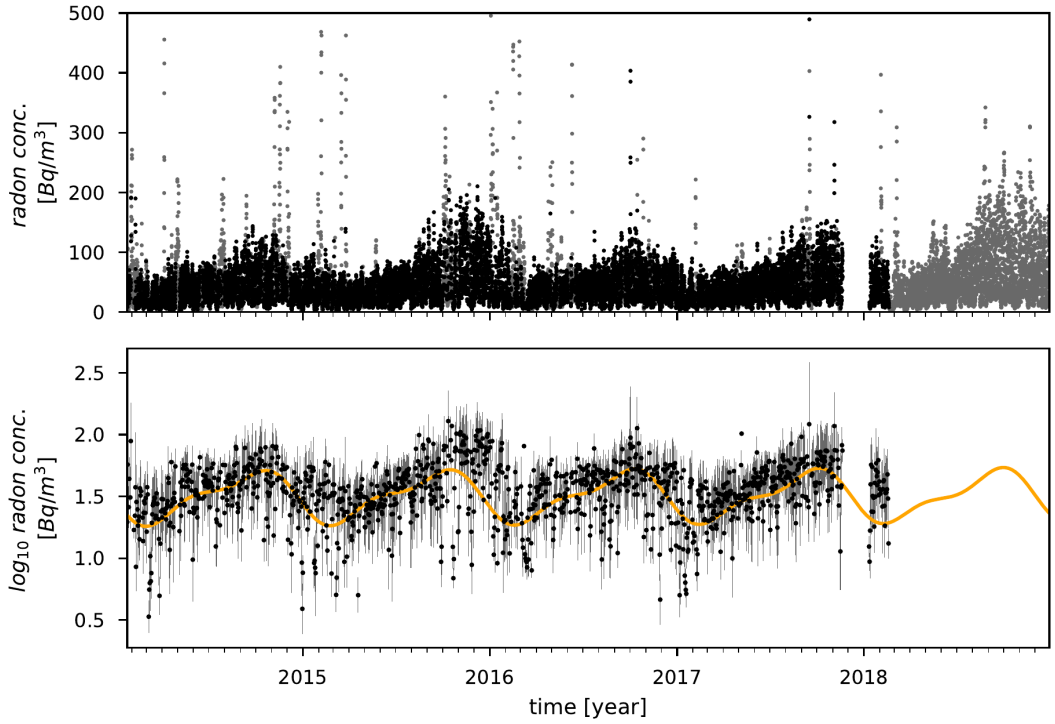

Figure 2.7: Example of the pre-processing steps for a radon station (CDCA). Top: raw data in grey and filtered data in black. The grey points are discarded measurements due to rainfall events or lack of meteorologic recordings (e.g. at the end of the time series). Bottom: Log transformed data, daily resampled with errorbars calculated as daily std. The sinusoid estimated with annual plus semi-annual components is superimposed in orange (arbitrary intercept and slope).

2.2.3 V_P/V_S

The average V_P/V_S ratio for a give ray-path can be computed from the P- and S- waves travel-times (Wadati and Oki, 1933) using the formula: $\widetilde{V_P/V_S} = (t_S - t_P)/t_P + 1$, where $\widetilde{V_P/V_S}$ is the average value along a the ray-path connecting an earthquake and a seismic station, t_P and t_S are the computed travel-times (i.e. P- and S- waves arrival times minus the event origin time). The V_P/V_S ratio values and errors used in this work are derived from a database of 82962 events recorded in the TABOO area from April 2010 to December 2019. The procedure to create time series from the database is the same as in Poggiali et al., (2019), here we report a brief summary.

Each time series is constructed with the goal of monitoring V_P/V_S values in time for a specific rock volume to detect V_P/V_S ratio variations. In order to sample the same rock volume the ray paths have to be similar, so the time series are constructed on a hypocenter-based selection:

1. the study area is gridded with 500 m spacing in three dimensions;

2. each node of the grid is the center of a 1 km^3 sphere, called *cluster* hereinafter;

3. a cluster-station time series is computed for a given station and all events within the cluster;

4. a cluster-station time series is retained for the next rj-McMC algorithm, if it contains at least 100 V_P/V_S ratio values.

After this procedure we obtained 10560 time series. More than half of the stations do not have enough data to originate a time series, so only 38 stations from the starting 90 are actually used in the following analysis. To avoid subjective bias, we do not remove outliers values, which are present evidently due to mislocated events in the catalog (i.e. mislocated events display erroneous origin-time, which, in turn, affect P- and S- wave travel times computation and, finally, V_P/V_S ratio estimates).

2.3 Methodology: the rj-McMC sampling algorithm

Our goal is to produce comparable outputs between data derived from different disciplines in order to identify significant signals/transients in multiple datasets. The workflow we propose relies on Bayesian inferences driven by a rj-McMC algorithm as an analysis tool common to all datasets. We implemented a parameterization that allows us to overcome or at least mitigate some of the dataset-specific issues seen in the previous section such as outliers and serial correlation. In the proposed workflow we model each time series with piecewise linear functions separated by CPs. Together with the distributions of each parameter analyzed, we produce distributions of CPs over time for each time series, which is a suitable kind of output to search for common changes between different disciplines and in relation to seismic activity. In the following we describe the workflow we used and the details of our algorithm.

2.3.1 Bayesian inference

Given a physical or mathematical model and a pool of data, the aim of Bayesian inference (Bayes, 1763) is to estimate a probability distribution for the model parameters (so called *posterior probability distribution, PPD*) from prior information on the model and a likelihood function representing the information contained in the data (the PPD calculation includes also an "evidence" term, which is not a function of the model and is often neglected, giving rise to the proportionality symbol in Equation 2.2). All components (PPD, prior and likelihood) must be expressed as probability distributions and the relation between them is commonly expressed with the following formula:

$$p(m|d) \propto p(m)p(d|m) \tag{2.2}$$

Here the posterior pdf $p(m|d)$ of the model m given data d is linked to our prior knowledge on the model, represented by the prior $p(m)$, and a term that quantifies

the probability of observing the measured data given that particular model $p(d|m)$, which is the likelihood function. The likelihood function quantifies how well the set of parameters composing the model can fit the observed data and can be expressed in different forms depending on the assumptions made on the statistical character of the noise (Mosegaard and Tarantola, 1995). The most used technique to obtain a numerical approximation of the PPD is Markov chain Monte Carlo (McMC) sampling (Gelman and Rubin, 1996; Mosegaard and Tarantola, 1995; Tarantola and Valette, 1982): a sequence of models is generated with a random walk constructed to have the PPD as its equilibrium distribution. Different sets of parameters, called states, are visited along the chain and the next state depends only on the current one. The first part of the chain (burn-in) where the random walk moves towards the high probability region is discarded, after that the random walk is assumed to be stationary i.e. the importance sampling follows the target distribution.

In this study we adopt the McMC approach of Mosegaard and Tarantola, (1995) to sample the PPD, which involves two steps. The first step consist in drawing a candidate model $\mathbf{m}_{cand}$ from the prior distribution, for example replacing one of the parameters in the current model $\mathbf{m}_{curr}$ with a value extracted from the prior probability distribution for such parameter. In the second step the candidate is accepted or rejected following an acceptance probability:

$$\alpha = min\left[1, \frac{L(\mathbf{m}_{cand})}{L(\mathbf{m}_{curr})}\right] \tag{2.3}$$

where L represents the likelihood function. It can be proven that if the candidate models are generated by sampling the prior distribution, this acceptance probability results in a random walk that samples the PPD (see Mosegaard and Tarantola, 1995). The expression used for the acceptance probability is a special case of the general acceptance criterion

$$\alpha = min\left[1, \frac{q(\mathbf{m}_{curr}|\mathbf{m}_{cand})}{q(\mathbf{m}_{cand}|\mathbf{m}_{curr})}\frac{p(\mathbf{m}_{cand})}{p(\mathbf{m}_{curr})}\frac{L(\mathbf{m}_{cand})}{L(\mathbf{m}_{curr})}|\mathbf{J}|\right] \tag{2.4}$$

that involves, in addition to the likelihood ratio, also the prior ratio p, a proposal ratio q to move from a candidate model to the current model and viceversa, and a Jacobian term $\mathbf{J}$ of the transformation from $\mathbf{m}_{curr}$ to $\mathbf{m}_{cand}$ which considers

the (potential) different number of dimensions in the two models. The McMC implementation used here adopts the transdimensional "reversible-jump" formulation proposed by Green, (1995), where the model dimensionality can vary along the chain (i.e. different number of parameters). The most important feature of rj-McMC algorithms is that model complexity is determined by the data: this characteristic is of primary importance in time series analysis and CPs modeling, where the number of partitions can be inferred without any subjective assumptions. In our algorithm model dimensionality can increase or decrease only by one at each iteration, falling in the category of "birth-death" algorithms (e.g., Malinverno, 2002; Hopcroft et al., 2007). As explained in Agostinetti and Malinverno, (2010), on which our algorithm is derived, the sampling strategy of Mosegaard and Tarantola, (1995) guarantees that the prior distribution equals the proposal distribution and, by adopting specific transformations, the jacobian term is unity and can be conveniently ignored.

2.3.2 Model parametrization

The diverse characteristics of the datasets, discussed in the previous section (section 2.2), result in the need of a versatile parameterization that can adequately adapt the model complexity to each observable with the most suitable set of basis functions. Here we adopt a piecewise-modeling. In other words, we consider a model that operates linear data interpolation in a given number of time-windows. Due to our trans-D approach, the number of time-windows is not constant, but it can vary from one model to the next model during the McMC sampling. Basically, the piecewise parameters fit the linear trends between two CPs, so these are the parameters that control local properties of the time series and its variations.

Being k the unknown number of CPs and c_i the times of occurrence of each CP, we model the data in the time-window between c_i and c_{i-1} with a slope a_i and an intercept b_i ($c_0 = t_{start}$, representing the starting date). Because we used a left-continuous CP definition, we need an additional set of parameters for the data between c_k and t_{end} (end of the observing period), which can be thought of as half-space values. So, the piecewise model vector can be written as:

$$\mathbf{m_{piecewise}} = (k, \mathbf{c}, \mathbf{a}, \mathbf{b}) \ , \tag{2.5}$$

where $\mathbf{c}$ is a k-vector and $\mathbf{a}, \mathbf{b}$ are $(k+1)$-vectors.

To deal with other properties of the datasets, such as outliers presence and serial correlation, we included two parameters related to the treatment of data noise (often called *hyperparameters*) in our rj-McMC implementation. The first one is the noise scale parameter ω which is adopted following the "hierarchical Bayes" approach (Malinverno and Briggs, 2004) and is defined in the same way as in Poggiali et al., (2019) where the original error values are multiplied by a factor 10^{ω} . In Bayesian inference the role of measurement errors is of fundamental importance and it is even more so in trans-dimensional implementations, where the complexity of the solution is a variable and it's directly related to the level of data noise. As exemplified in Agostinetti and Malinverno, (2010) the noise level is inversely proportional to the complexity of the solution. Imposing an a-priori noise level is equivalent to obtaining a solution with a specific complexity: a higher value would determine a less complex solution and vice versa. By adopting the hierarchical Bayes approach and sampling ω as a parameter we let the algorithm "modulate" the level of data noise and this ultimately controls the complexity of the solution in a totally data-driven way. As a consequence, measurement errors retain only their relative importance to each other.

The second noise parameter r is introduced to treat the serial correlation that some dataset exhibit (geochemicals). This property results in a covariance matrix which can no more be assumed as diagonal. We follow the definition of Bodin et al., (2012a) where the correlation between samples is described in the covariance matrix either as an exponentially decaying function or with a Gaussian correlation law. The first type of noise correlation, of which the details can be found in appendix D2 of Bodin et al., (2012a), is the one adopted in our implementation. This expression of noise correlation is very convenient from a computational point of view because both the determinant and the inverse of the covariance matrix have an analytic formulation that can be exploited to perturb r or ω directly. Thus, our model is completed by:

$$\mathbf{m_{error}} = (\omega, r) \ . \tag{2.6}$$

2.3.3 Likelihood function

Our implementation builds on the one described in Poggiali et al., (2019) and uses the same sum of absolute differences in the likelihood function (l1 norm). As explained in Mosegaard and Tarantola, (1995) this is equivalent to expressing experimental uncertainties with a Laplacian function instead of the more common Gaussian function. This has the advantage of being a more robust estimator, which is suitable in the presence of outliers that otherwise would have determined a greater complexity (more CPs) of the models in order to "fit" outlier values. The expression for the likelihood, combining the choice of a Laplacian function and exponential correlation (see Section 2.3.2), can be written as follows. The covariance matrix can be decomposed as $\mathbf{C_e} = \mathbf{SRS}$ Malinverno and Briggs, 2004, where: $\mathbf{S}$ is the diagonal $n \times n$ matrix with the square root of the data uncertainties along the diagonal multiplied by the square root of the noise scale parameter 10^{ω}, and $\mathbf{R}$ is the $n \times n$ exponential correlation matrix (n is the number of data-points). The square root of the absolute value of the residuals vector is $\mathbf{e} = \left| d_i^{obs} - d_i^{sim} \right|^{1/2}$. In this case, following Malinverno and Briggs, (2004), the likelihood function

$$L(\mathbf{d}|\mathbf{m}) = \frac{1}{2\left|\mathbf{C_e}\right|} exp\left(-\mathbf{e}^T {\mathbf{C_e}}^{-1} \mathbf{e}\right) \tag{2.7}$$

simplifies to

$$L(\mathbf{d}|\mathbf{m}) = \frac{1}{(1-r^2)^{n-1} 2 \prod_{i=1}^{n} 10^{\omega} \sigma_i} exp\left(\frac{-\phi}{1-r^2}\right) \tag{2.8}$$

with

$$\phi = \phi_1 + \sum_{i=2}^{n-1} \phi_i + \phi_n \tag{2.9}$$

where:

$$\phi_1 = \frac{\left| d_1^{obs} - d_1^{sim} \right|}{10^{\omega}\sigma_1} - r \frac{\left| d_1^{obs} - d_1^{sim} \right|^{1/2} \left| d_2^{obs} - d_2^{sim} \right|^{1/2}}{10^{\omega} \sigma_1^{1/2} \sigma_2^{1/2}} \ ,$$

$$\phi_n = \frac{\left| d_n^{obs} - d_n^{sim} \right|}{10^{\omega}\sigma_n} - r \frac{\left| d_n^{obs} - d_n^{sim} \right|^{1/2} \left| d_{n-1}^{obs} - d_{n-1}^{sim} \right|^{1/2}}{10^{\omega} \sigma_n^{1/2} \sigma_{n-1}^{1/2}} \ ,$$

and

$$\phi_i = -r\frac{\left|d_{i-1}^{obs} - d_{i-1}^{sim}\right|^{1/2}\left|d_i^{obs} - d_i^{sim}\right|^{1/2}}{10^{\omega}\sigma_{i-1}^{1/2}\sigma_i^{1/2}} + (1+r^2)\frac{\left|d_i^{obs} - d_i^{sim}\right|}{10^{\omega}\sigma_i} - r\frac{\left|d_{i+1}^{obs} - d_{i+1}^{sim}\right|^{1/2}\left|d_i^{obs} - d_i^{sim}\right|^{1/2}}{10^{\omega}\sigma_{i+1}^{1/2}\sigma_i^{1/2}} \quad .$$

The case of independent samples is handled as a special case of the previous expression for the likelihood with $r = 0$. Instead, imposing $\omega = 0$ is equivalent to using original errors without making inference on the noise scale factor.

2.3.4 Prior probability distributions

In Bayesian inference every prior knowledge we have is expressed by means of probability distributions, which, combined with the likelihood, will produce the PPD. Therefore the choice of the prior distributions to be adopted must be made carefully (Roy and Romanowicz, 2017) in order to balance an efficient sampling (defining not too wide bounds) and an unconstrained solution. One of the criticisms that is in fact brought against the Bayesian approach is that the prior can be tuned to lead the sampling towards preferred solutions (Scales and Snieder, 1997). Our approach is to use uniform prior distributions for all parameters to avoid any preference over the solution, and setting sufficiently wide bounds. It is important to notice that a uniform prior is also used to sample the model dimension. This means that model with a different number of CPs, namely between 1 and 100, are equally probable a priori. A bit counter-intiuitively, the consequence is that having a small amount of information in the data does not automatically translate in sampling low dimensional models. If few or no information are contained in the data, both low and high dimensional models will be accepted along the McMC sampling.

2.3.5 Sampling recipe

After initializing the model parameters in the current model by drawing values from the prior distributions, the sampling of the PPD is done by proposing a new candidate model and accepting or rejecting it according to Equation 2.3 presented

before. To produce a candidate model form the current one, we randomly pick one of the following moves (grouped into two main categories):

1. moves that affect piecewise parameters:

   1.1 (proposed with probability 0.20) perturb intercept value a of a CP;

   1.2 (0.20) perturb slope value b of a CP;

   1.3 (0.20) perturb the position of a CP in time c;

   1.4 (0.1) create a new CP (birth move);

   1.5 (0.1) delete an existing CP (death move);

2. moves that affect noise parameters:

   2.1 (0.1) perturb noise scale value ω;

   2.2 (0.1) perturb noise correlation parameter r.

The uniform priors associated to the model parameters are sampled following the strategy proposed in Appendix A of Agostinetti and Malinverno, (2010).

We here briefly depict the sampling workflow, summarizing the main steps. All saved model are post-processed at the end of the McMC sampling to obtain the approximation of the PPD (from which all figures in the present manuscript are drawn):

A. pick a candidate model $\mathbf{m}_{cand}$ by perturbing the properties of $\mathbf{m}_{curr}$ with one of the moves illustrated above;

B. compute the likelihood of $\mathbf{m}_{cand}$;

C. accept or reject the candidate model following Equation 2.3: if $\mathbf{m}_{cand}$ is accepted then it replaces $\mathbf{m}_{curr}$, otherwise $\mathbf{m}_{curr}$ is retained;

D. save $\mathbf{m}_{curr}$ and restart from (A).

We used the versatility of our parametrization to adapt the modeling to the different scenarios that each dataset represents:

- for V_P/V_S we used a simple parameterization made only by a piecewise constant function (slope parameter fixed to 0);

- GNSS data is modeled with piecewise linear, noise scale value but no noise correlation parameter;

- Geochemical datasets (radon and CO_2) are modeled using piecewise linear, and both noise parameters.

Each move is tuned, as common practice in Bayesian inversion (Agostinetti and Malinverno, 2010), by keeping the value of the acceptance ratio (accepted over proposed models) between 0.25 and 0.5 to balance between the exploration of the parameter space and sampling efficiency.

For V_P/V_S time series the algorithm ran on 10 independent chains for 10M iterations each; the first half of each chain is discarded as burn-in. As introduced before, the datasets that require seasonal components (GNSS and geochemical) are modeled with a two step procedure separating the sampling of periodic parameters from piecewise parameters. We use the first part of the chains to infer only periodic parameters: the obtained ensemble of models is used to compute the average values of each periodic parameter. In the second part of the chains we fix seasonal parameters to these average values and we start sampling the piecewise parameters. So, for GNSS and geochemical time series, the algorithm is first ran with 20 chains for 500k iterations, of which the first 400k are discarded as burn-in. The last 100k of this run are used to infer periodic parameters. Periodic parameters are then kept fixed and the 20 chains are iterated for 1M models of which 900k are discarded as burn-in. In each case (V_P/V_S, GNSS, geochemicals) we retain only 1 model every 100.

2.4 Results

First we analyze some examples of the application of the Bayesian algorithm to show the general behavior with different datasets and to highlight positive and negative characteristics of our approach in relation to data properties. We then

show a case where the inverted model is simple enough to be visually compared with main earthquakes near the measuring station. Finally, we gather the whole ensemble of retrieved CPs distributions from multidisciplinary time series and analyze it in relation with seismic activity recorded inside the study area.

A first observation is that the complexity of the solutions, represented by the number of CPs used in the modelling, varies considerably across different kind of datasets (grey histograms in right panels of Figure 2.8). Typically V_P/V_S time series can be modeled with few CPs, GNSS time series result in a moderate number of CPs (generally below 50), and geochemical time series commonly need a high number of CPs. The number of CPs for geochemical data is generally between 40 and 80, with few notable exceptions: radon station BADI with less than 10 CPs and SSFR with 100 CPs, which is the maximum value of the prior range on CPs number (i.e. 0 to 100).

Each dataset modeled with periodic signals shows some time windows of considerable length (months to almost a year in the examples of Figure 2.8) where seasonal oscillations are enough to properly model the datasets. These portions of the time series are identified by absence of CPs.

The effect of l1-norm likelihood implemented in the algorithm can be clearly seen in panel (a) of Figure 2.8: in this GNSS time series the presence of some outliers is evident, but it is also evident that they are not affecting the resulting model. Outliers are indeed properly downweighted with the l1-norm and do not give rise to unneeded CPs. Exceptions are observed, as expected, in case of isolated data points with low error or clusters of data points far from neighboring values. These situations are observed more frequently with radon stations, of which examples are visible in panel (d) of Figure 2.8 (MURB station). The effect of outliers on inverted models is recognizable by the spike-like peaks or drops in the average models from PPD (red lines in Figure 2.8).

A feature observed in the models obtained from the GNSS time series, as the one in panel (a) of Figure 2.8, is that most CPs do not mark large offsets. One of the few exceptions is visible in this example: a clear offset in October 2016 related to the Central Italy seismic sequence. This offset is detected in multiple stations, as can

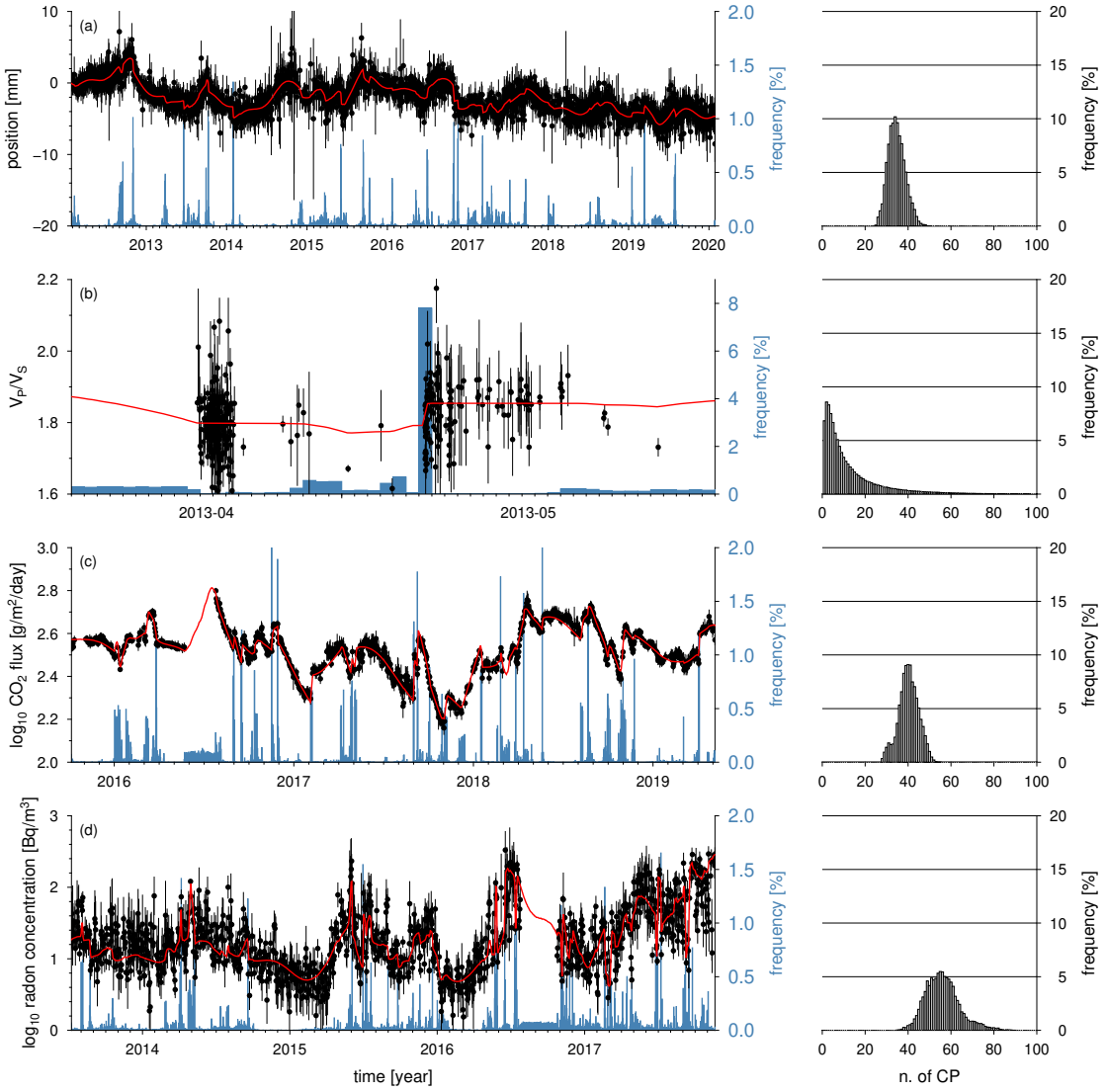

Figure 2.8: Examples of outputs from different datasets. In all panels data points are depicted in black, mean model from PPD in red, CP distribution in blue, number of CP on side panels. The examples, from top to bottom panel, are respectively: GNSS, V_P/V_S, CO_2, radon.

be seen in Figure 2.12. More details on this and other common features observed in multiple GNSS stations can be found in the appendix. Most variations are much less evident and probably due to local effects or adaptations of the piecewise-linear model to different amplitude modulations of the seasonal components. The high sensitivity of the algorithm to such small changes is in accordance with the estimated scalar factor for the noise level in the data uncertainties, about $\omega = -0.33$, which means that previously estimated noise values are practically halved.

Many V_P/V_S time series are sparse and with large gaps between data. In these ranges of low data-density, CPs can be placed to model few data points distant from each other. This behavior is expected and it is a consequence of the automatic procedure for time series creation, without any fine tuning on specific space-time windows, i.e. not tailored to seismic sequences. CPs used to fit sparse data are plateau-shaped, not much higher than the prior level, and very different from the spike-shaped ones marking abrupt changes occurring in zones with high data density. In panel (b) of Figure 2.8 both types are visible: a well constrained V_P/V_S variation is marked with a clear peak in CPs distribution in May 2013; while between April and May a plateau-type CPs are visible. The latter case represents a situation where a CP is supported by the data, but the position in time is not well constrained. It is also interesting to note that the first "cluster" of data points doesn't show any CP, meaning that a high density of measurements does not translates to increased CPs occurrence if the data does not support it. The no data zone at the beginning of the time series is characterized by a flat CPs distribution tending to the prior level. An analysis of CPs distribution of all time series is discussed later.

A commonly observed feature of geochemical time series, is the relatively high number of CPs used for the modeling. Considering that geochemical time series are less than than 5 years long, the CPs density over time is much higher compared to GNSS time series, which are modeled with an average ~30 CPs and are almost always longer than 5 years (except two stations). The reason why such complex models are sampled by the algorithm probably arises from specific characteristics of CO_2 and radon datasets. Radon datasets exhibit frequent peaks and drops visible in

multiple stations: CDCA and MURB have approximately the same number of CPs (~60) in the same timespan (4 years). Marked drops can be seen in CDCA, while in MURB both peaks and drops are visible. SSFR is an extreme case: together with a very high variance, this station is also probably affected by strong peaks on a multi-day scale, a site specific feature already observed in radon measurements (Siino et al., 2019)). In all mentioned cases the data points can not be considered outliers by the algorithm as a consequence of dataset properties and estimated noise level. This happens despite using an l1-norm and considering a non diagonal covariance matrix which accounts for correlated measurements. CO_2 data, on the other hand, are characterized by very stable measurements on the daily scale. With such low noise data, even little variations are considered significant by the algorithm and this results in the high number of CPs observed. This behavior, somehow similar to overfitting, is nonetheless determined by the data.

In the end, for both CO_2 and radon data, models complexity has an obvious effect on the interpretability of the results: with an high number of CPs the importance of each one in terms of “anomalous signal” is lowered and it becomes increasingly more difficult to detect "significant" variations and eventually relate each one with potential causes.

BADI is the only geochemical station with a low number of CPs (<10), so it is the only case in which a direct comparison with local and regional seismicity is actually possible. In (Figure 2.9) we show: (a) data and mean model from PPD (the piecewise model is superimposed on the seasonal signal); (b) CPs distribution and earthquakes ($M \geq 3$) within 20 km distance from BADI station. At least in this case there isn't a consistent relation between seismic activity and variations in radon emissions retrieved with our methodology. We may try not to consider the second half of the time series (from 2016 onwards) and speculate on a link between radon and seismicity in 2014 and 2015. Even with these premises the only cases in which a connection can be hypothesized are October 2014 and September 2015 events: in both the variation would be post-seismic. Other events, even closer to the station, would not produce similar effects.

To highlight common features possibly hidden in the complex output of multi-

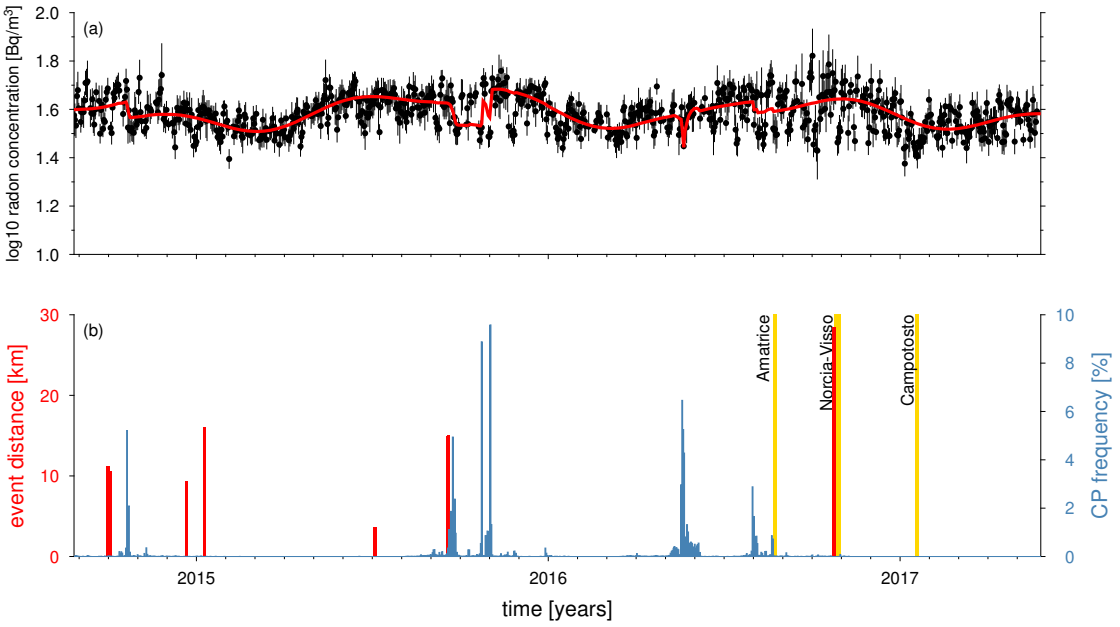

Figure 2.9: Radon station BADI. Top: data (black) and mean model from PPD (red). Bottom: CP distribution (blue), local earthquakes (red, $M \geq 3$) and regional earthquakes (yellow, $M \geq 5$).

observables results, we gathered all CPs for each dataset and represented as percentage distributions on a common time axis (Figure 2.10).

We didn't make any selections, apart from removing V_P/V_S CPs with less than 10% of data in the time series before or after (one of the criteria used also in Poggiali et al., (2019)). This is done and in order to eliminate CPs at the edges of time series, not supported but enough points, probably due to the sparsity of V_P/V_S time series mentioned before. V_P/V_S distribution shows clear peaks: the most evident related to Gubbio seismic sequence of 2013-2014 (as seen in Poggiali et al., (2019)), but other peaks are evident in 2010 (Pietralunga sequence) and 2013 (Città di Castello). Detected CP are not "validated" in any way: some could be related to non-meaningful variations (too little variation with respect to uncertainties in posterior models) or originate from outliers presence. This level of detail is nonetheless sufficient to highlight main features.

GNSS CPs distribution shows no obvious common features, apart from two notable peaks: the one near the end of 2016 related to central Italy earthquakes and the one at the beginning of 2015, which is discussed in appendix. Other features of GNSS models are better evidenced with alternative representations shown in appendix.

Unfortunately most geochemical data are available only after main seismic sequences of the study area, so with current time series we can't make an ad-hoc comparison of geochemical response in relation to local seismic sequences. A possible attempt for a comparison with seismicity has been done with BADI radon station (Figure 2.9) because of the simple resulting model (few CPs).

No common pattern of CPs are observed, indicating that, at this scale and with this processing, no common response is detected (both inter- and intra- discipline). If present, it remains buried by site-specific effects.

2.5 Discussion and Conclusions

In this study we analyzed different kind of time series with a Bayesian approach in search for common variations and/or tectonic related signals. The opportunity

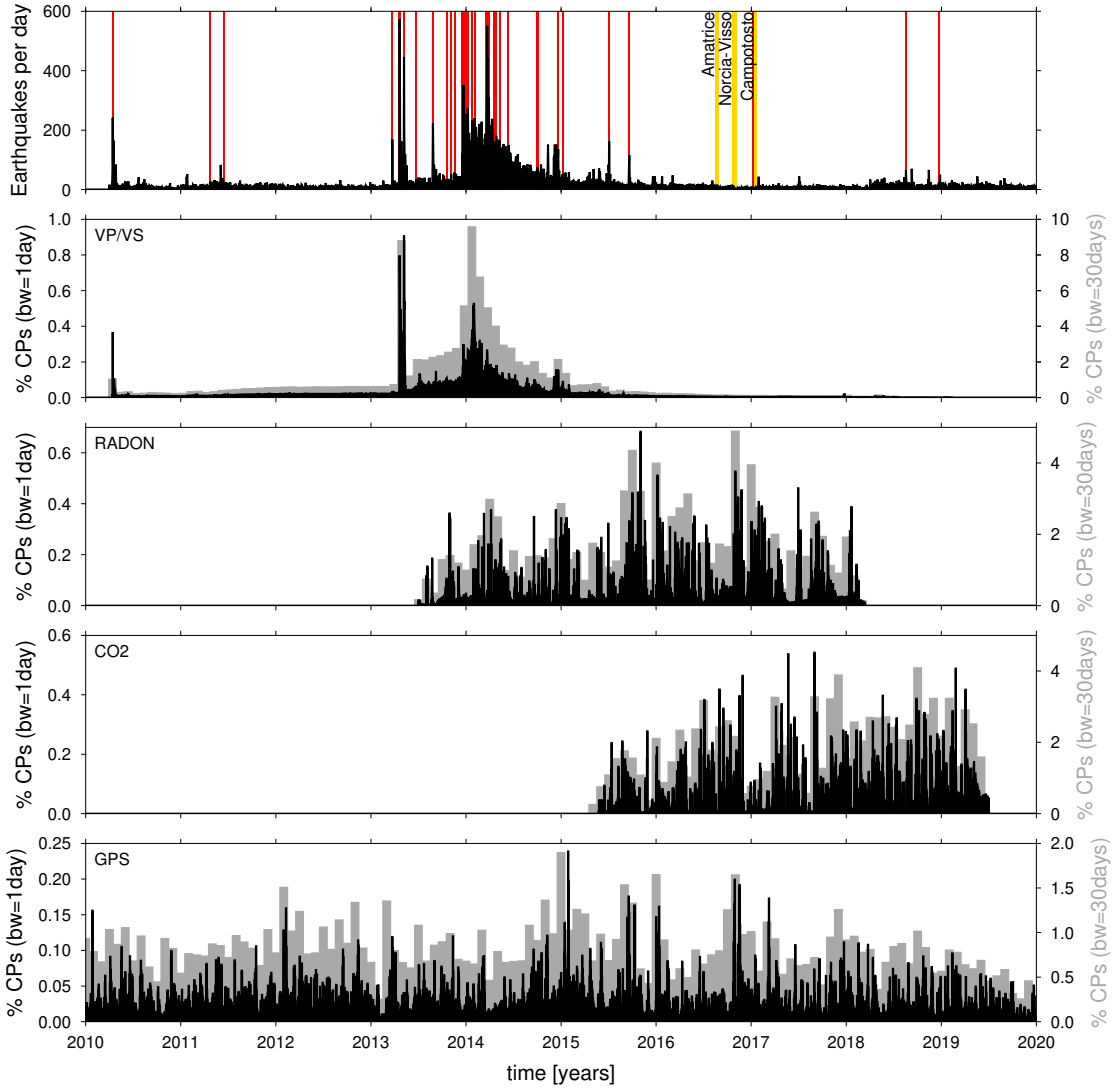

Figure 2.10: CPs distributions of the multidisciplinary time series and earthquakes distribution (top). Black histograms use a binwidth of 1 day, grey histograms use a binwidth of 30 days. Local earthquakes ($M \geq 3$) in red and regional earthquakes ($M \geq 5$) in yellow.

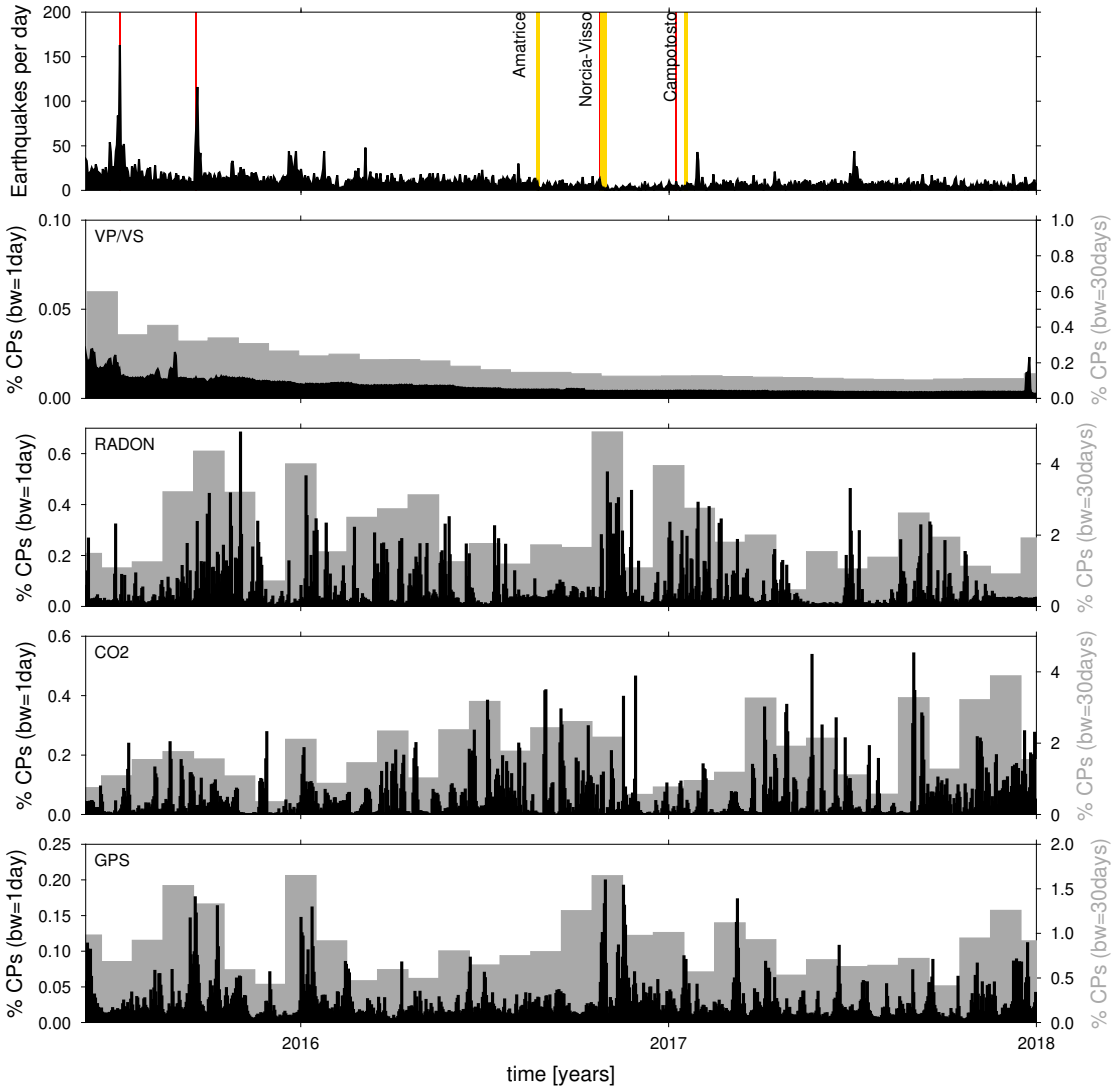

Figure 2.11: CPs distributions of the multidisciplinary time series and earthquakes distribution (top). Zoom on the time window where all datasets are available. Black histograms use a binwidth of 1 day, grey histograms use a binwidth of 30 days. Local earthquakes ($M \geq 3$) in red and regional earthquakes ($M \geq 5$) in yellow.

for this multidisciplinary experiment is given by the TABOO NFO infrastructure, composed by multiple seismic, geodetic and geochemical (radon and CO_2) measuring stations, monitoring an area of the northern Apennines characterized by a high seismicity rate. The proposed workflow, based on a Bayesian framework, is aimed to independently analyze the different time series and finally produce a standardized and comparable output, in a synthesis effort. We have dealt with the challenges related to the different characteristics of the datasets in two ways: 1) a pre-processing aimed at removing known signals of non-tectonic origin affecting some of the observables and 2) algorithm implementations regarding the likelihood function, covariance matrix, and model parametrization.

The proposed implementations allowed the algorithm to be very versatile. The choices we made for the base function (piecewise-linear), likelihood (l1 norm) and covariance matrix (correlation parameter) made the algorithm adaptable to a wide variety of situations and able to deal with complex datasets including outliers, unknown noise scale and correlation between the measurements, which are both treated as unknowns and can be estimated. From the modeling point of view, the algorithm is data-driven and weakly influenced by subjective decisions: Bayesian inversion controls CPs number and positioning in time, inversions are performed independently and resulting distribution are a consequence of data and errors properties.

Nonetheless a successful application depends also on the assumptions made in the pre-processing phase. Especially for some datasets, this step strongly influences the time series to be analyzed with the Byesian algorithm and finally the complexity and interpretability of the results. Our approach on these topics was to keep the pre-processing step as simple as possible in order to produce suitable input data for the McMC inversion. More specifically, we considered annual and semiannual seasonal components for all datasets apart from V_P/V_S, and for geochemical data we proposed a filter based on rainfall thresholds. Aspects regarding periodicities and "exogenous signals", as described in the data section, are very complex and still debated, therefore we cannot aim to produce time series completely free from non-tectonic signals, but only to mitigate them in a consistent way. Considering these

limitations, we acknowledge the potential incompleteness of the pre-processing due to unknown factors and the expected effects on the results would be an increased model complexity. This is true even in the worst case scenario of a wrong estimation of the seasonal components: the piecewise-model would still "fit" the data, but with an over-complicated and non meaningful model.

Geochemical data are probably the most complicated data to analyze, not only because of the data (high variance, correlation), but also because of unknown factors influencing the measurements. Considering annual plus semiannual periodicity, rain effect and daily resampling (to remove sub-daily cycles) has proven necessary but often not sufficient to produce interpretable CPs distribution. Seismic related variations can still be detected with our approach, but are difficult to separate from variations due to other causes. In this context longer time series and improvements of the network are crucial for the evaluation of non obvious periodicities and other signals of various origin that can affect measurements in a systematic way. Also the temporal coverage of the actual geochemical time series is limited in relation to the major seismic sequences of the area, therefore we didn't have the opportunity to compare a potentially more evident signal or the lack thereof. An alternative way to analyze geochemical results can involve a quantitative estimation of variations on the modeled time series. The time distribution of CPs can be paired with, for example, the cumulative variation on the PPD mean on a moving window.

V_P/V_S data confirm to be well suited to be analyzed with a Bayesian CPs detection algorithm because of favorable features of data distribution and noise, and also because of the simple piecewise-constant model that can be adopted. The piecewise-constant parametrization used for V_P/V_S time series also makes CPs distributions immediately interpretable because each CP identifies a variation in the time series (e.g. variation in the elastic properties of the sampled volume). In other kind of datasets this connection is not as obvious, because a potentially interesting variation could lie between CPs (e.g linear increase or decrease). Peaks in CPs distributions are observed in conjunction with multiple seismic sequences recorded in the area: most of the CPs gather around the Gubbio seismic sequence at the end of 2013, but other peaks are observable in relation to Pietralunga

(2010) and Città di Castello (2013) sequences Figure 2.10. These peaks in CPs distributions are not further investigated because the purpose of this study was to verify multidisciplinary relations.

The proposed method has proved to be applicable successfully also on GNSS time series. Seismic related offsets are observed in multiple stations, but not in relation to local earthquakes. A long period (~4 years) pattern emerge from the analysis of the ensemble of resulting models (Figure 2.12). A common (unknown) signal is also detected at the beginning of 2015 (Figure 2.13). The techniques here applied can be used as an offsets detection tool (not only tectonic, but also related to instrument failures) for GNSS time series.

This approach offers promising insights as a workflow for multidisciplinary time series analysis. The results will become more interesting from an interpretative point of view with improvements to the pre-processing step, which in this experiment has been intentionally left simplified in order to manipulate the data as little as possible. In this context the estimation of the periodic compontents could benefit from a trans-dimensional implementation on the number of sinusoids, in order to include additional periodicities, when needed, in a data-driven way. A refinement in the estimate and removal of "exogenous" contributions would allow applications of the approach proposed here to multi-parametric monitoring of active seismic areas as a tool aiding expert opinion.

2.A Observations on GNSS models ensemble

In Figure 2.12 we show the median (black line) of mean posterior models for E-W (top) and N-S (bottom) components. Models are detrended after inversions and the estimated annual and semiannual signals are not considered. Gathering all piecewise models from GNSS inversions, some common features emerge.

A long period oscillation is visible on the E-W component with a period of approximately 4 years. This kind of low frequency signals have been observed in GNSS time series and have been related to hydrologic processes involving groundwater recharge/discharge. An in-depth analysis can be found in Serpelloni et al., (2018) for the Alpine region, and Silverii et al., (2016) and Mandler et al., (2021) for the Apennines. An hydrological origin could be an explanation of the long period oscillation.

Another feature observed on many resulting models is the V-shaped signal near beginning of 2015, visible in E-W component of Figure 2.12. This feature is also evident in Figure 2.13, where the azimuth distribution of GNSS stations is represented as histograms in time. Clear maxima are oserved near 2015, indicating that a noticeable number of stations are moving in the same directions.

Apart from the seismic offset related to 2016 Central Italy earthquakes, no obvious coseismic signals are observed in relation to earthquakes occurred inside the study area.

Acknowledging the complexity of positional time series (Chanard et al., 2020), it is worth noting that the features here described arise from an ensemble of independently analyzed time series. Looking at the multi-station features instead of single station models, partially avoids the risk of misinterpreting site-dependent signals as tectonic motion.

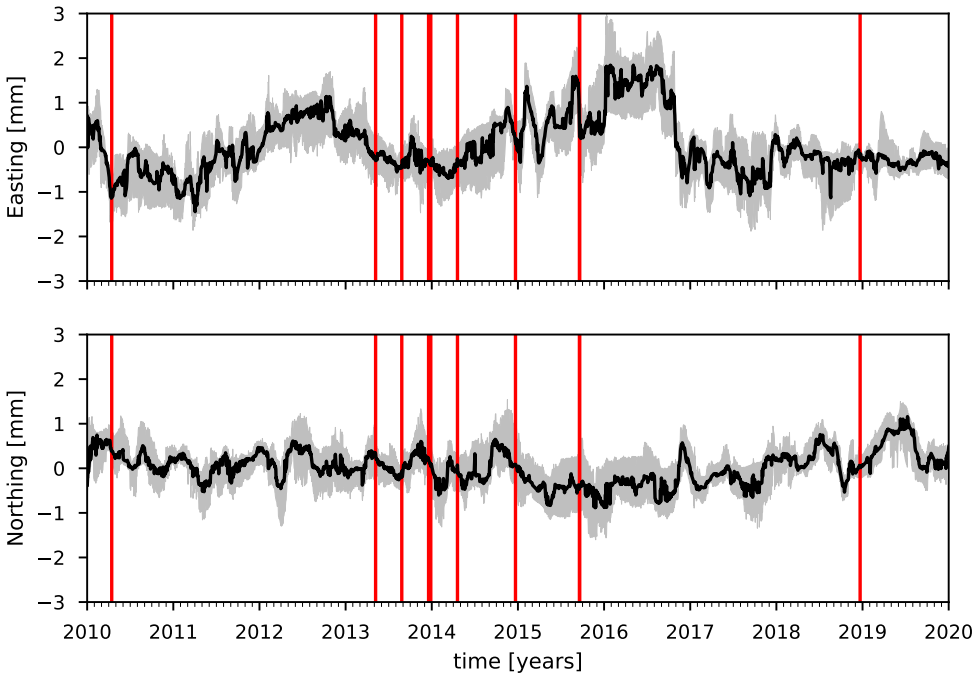

Figure 2.12: Median of the posterior mean models from all GNSS stations. Models are detrended post-processing and don't include the estimated seasonal contributions. Grey shade represents interquartile range. Top:E-W component. Bottom:N-S component. Red lines represent the occurrence of earthquakes $M \geq 3.5$ inside the study area.

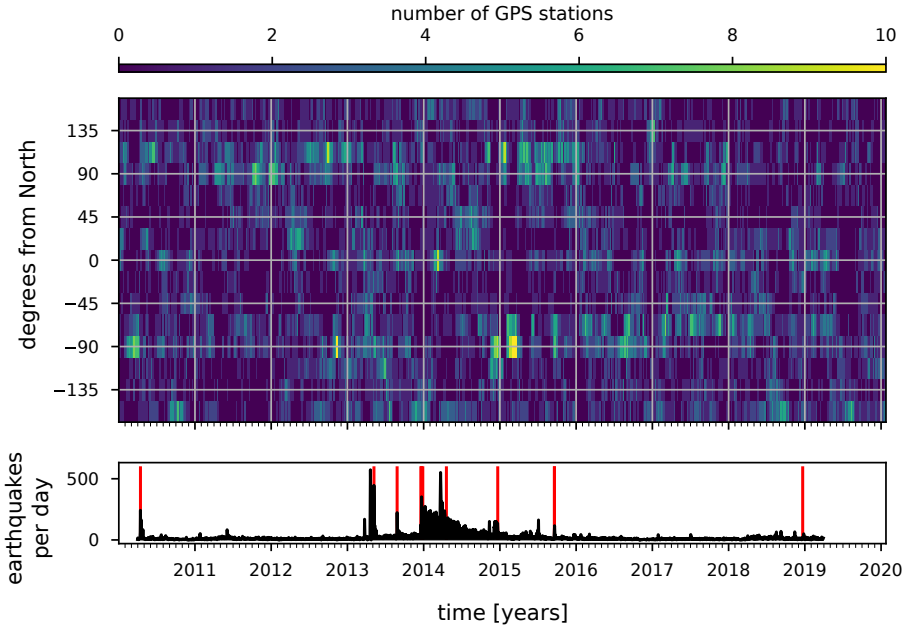

Figure 2.13: Top: distribution of the direction of movement from posterior models as histograms in time; all GNSS stations inside the study area are considered. It's a representation of how many stations are moving in the same direction at the same time. Bottom: seismicity inside the study area. Red lines represent the occurrence of earthquakes $M \geq 3.5$.

2.B Periodic signals and trans-dimensional implementation

Time series are influenced, at a variable level, by meteorologic parameters. Some of these parameters (e.g temperature) translate to seasonal oscillations in measured time series, others in short-term anomalies (e.g. rain), as described in section 2.2.

In a first phase of data exploration, geochemical time series have been analyzed in search for relations with meteorological parameters to be used in the modeling of seasonal signals. As already pointed out, a complex aspect of these datasets is the site specific effect observed in CO_2 and radon with respect to meteorological parameters. As shown in Figure 2.14 and Figure 2.15, the correlation between available parameters (temperature, humidity, atmospheric pressure) and geochemical measurements is highly variable, with both positive and negative correlation values for the same parameter at different sites.

The determination of cause-effects relations between meteo and geochemical measurements is a very complex topic, which is beyond the scope of this work. Even if the specific cause of the fluctuations are not identified, seasonal oscillations are indeed present, as shown in Figure 2.16 and Figure 2.17. The frequency spectra are estimated using a least-squares spectral analysis method (Lomb-Scargle), which allows to analyze series with gaps and irregular sampling, unlike Fourier analysis. A strong daily signal is observed in radon time series, along with an annual oscillation and a minor semi-annual oscillation. For CO_2 the long period signals are still present, but daily oscillations are not so prominent.

Apart from the annual and semi-annual signals present in both GNSS (Serpelloni et al., 2006) and geochemical data (Siino et al., 2019), other oscillating contribution could be present in CO_2 and radon time series, as suggested by Figure 2.16 and Figure 2.17. Since the number of sinusoids required to model the time series is not exactly known, the idea to approach this problem has been the integration of this unknown periodic contributions in the rj-McMC algorithm. The number of sinusoids should not increase indefinitely, due to the parsimonious characteristics of such algorithms (Malinverno, 2002), guaranteeing against "overfitting" the data

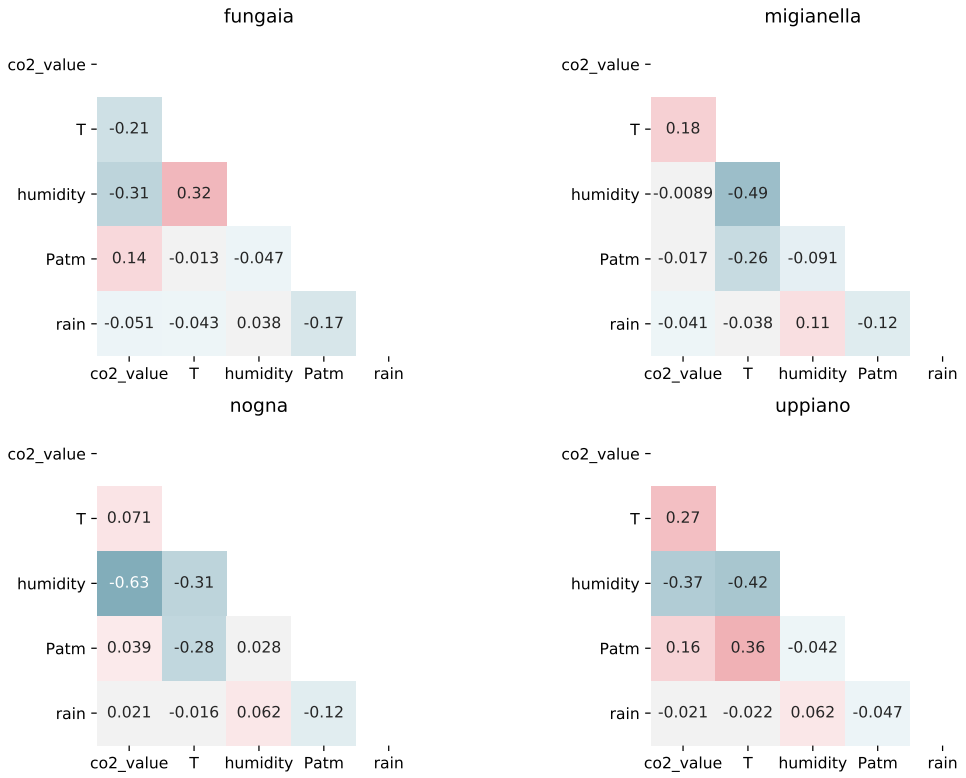

Figure 2.14: Correlation coefficient between CO_2 time series and meteorological parameters.

Figure 2.15: Correlation coefficient between radon time series and meteorological parameters.

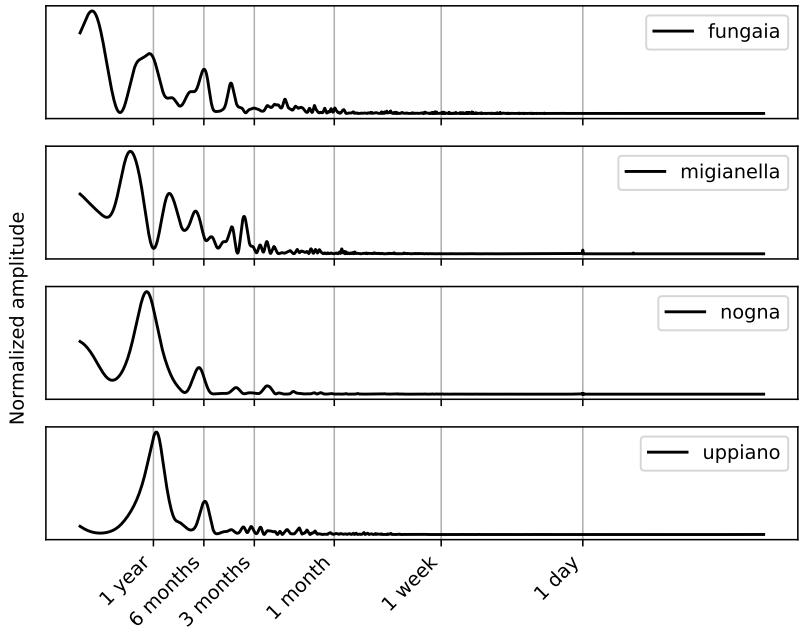

Figure 2.16: Lomb-Scargle periodogram of CO_2 time series.

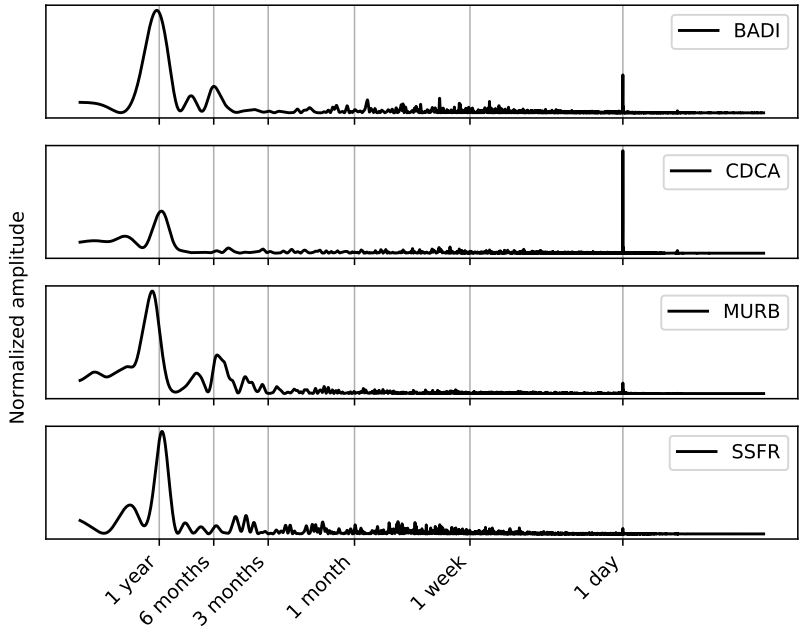

Figure 2.17: Lomb-Scargle periodogram of radon time series.

with an unnecessary high number of sinusoids. Thus, we explored this possibility by implementing a "double trans-dimensional" version of the code, meaning that the algorithm would be trans-dimensional in both the number of CPs and the number of sinusoids, the latter treated in this version as an additional unknown parameter. The equation reported for the modeling of annual and semi-annual contributions:

$$p(\mathbf{m}) = \alpha_1 sin(2\pi\gamma_1 t + \beta_1) + \alpha_2 sin(\pi\gamma_2 t + \beta_2) \tag{2.10}$$

becomes a more general

$$p(\mathbf{m}) = \sum_{i=1}^{n} \alpha_i sin(2\pi\gamma_i t + \beta_i). \tag{2.11}$$

The parametrization includes a periodic model vector $\mathbf{m_{periodic}} = (n, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ with $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ vectors of length n describing amplitude, frequency and phase of each of the n estimated sinusoids. The complete parametrization include this periodic part and the piecewise model vector described in the method section: $\mathbf{m_{piecewise}} = (k, \mathbf{c}, \mathbf{a}, \mathbf{b})$.

The moves have been updated accordingly to sample the additional periodic parameters. The following moves have been added:

1. perturb amplitude value α of a sinusoid;
2. perturb phase value β of a sinusoid;
3. perturb frequency γ of a sinusoid;
4. create a new sinusoid (birth move);
5. delete an existing sinusoid (death move).

This approach have been extensively tested on synthetic and real data, resulting in a complex coexistence of periodic and piecewise parametrization. On the simpler synthetic cases, like the one shown in Figure 2.18. Here the dataset is correctly modeled with 2 CPs and 2 sinuoids.

A slightly complex synthetic model (Figure 2.19) highlights a non-optimal modeling, with neither the number of CPs nor sinuoids converging to the right

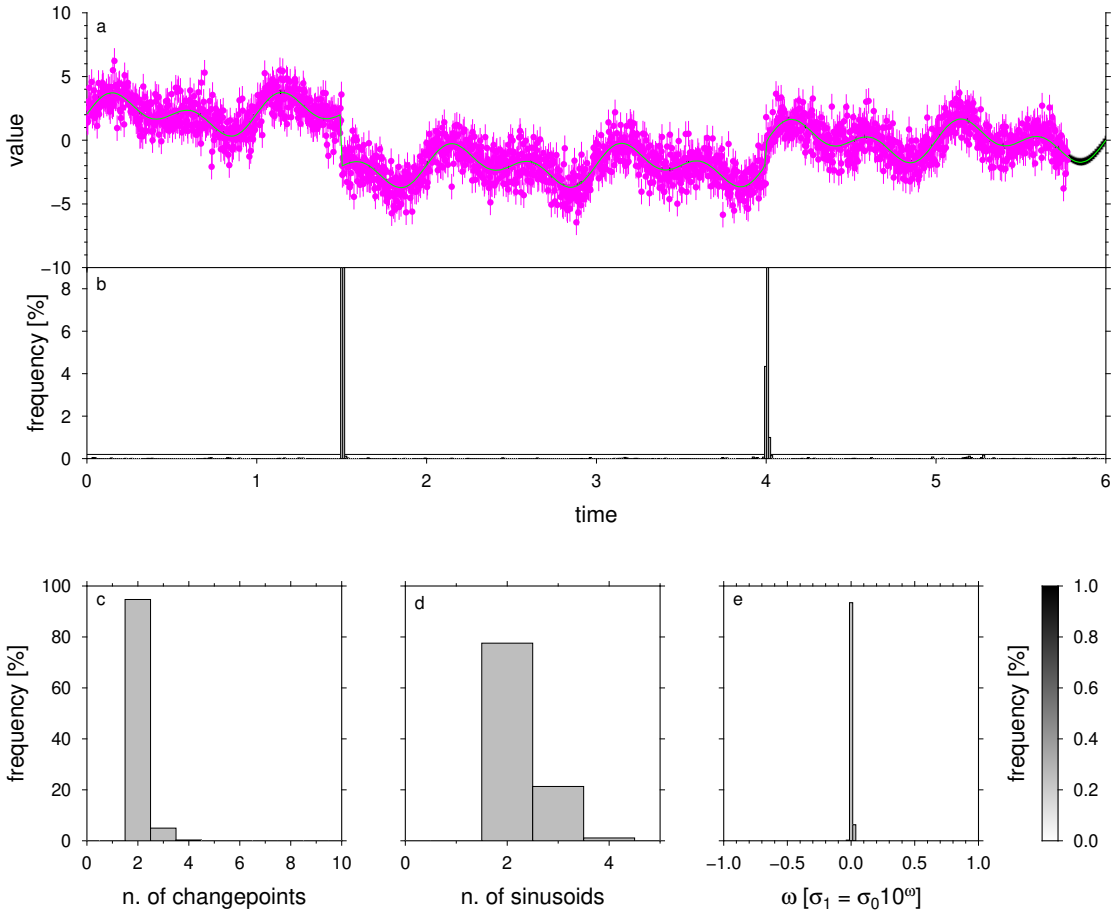

Figure 2.18: Synthetic test made with 2 sinusoids and 2 CPs. Data (magenta) and mean model from PPD (green) in panel (a); CP distribution in panel (b); number of CPs in panel (c), number of sinusoids in panel (d), noise scale in panel (e).

value (respectively 6 and 2). Moreover the distributions appear bimodal, indicating the (obvious) strong link between the piecewise and periodic parameterizations influencing each other. The issues encountered here and in numerous other test are due at least to the following factors:

- the two parametrization are "competing" to fit the datasets, with the piecewise part able to adapt to any sinusoidal model, being it correct or incorrect;
- the sampling along one or more chains can get stuck in a high likelihood area due to an incorrect sinusoidal and piecewise combination;
- the noise scale hyperparameter can contribute to this situation by lowering its value and making it impossible to change the parametrization towards correct values, especially because changing the sinusoidal model affects the whole time series at once.

Despite being interesting from a methodological point of view, this approach have proven not to be reliably applicable, needing further improvements and testing. A successful application may require a more sophisticated sampling, involving techniques like Simulated Annealing (Laarhoven and Aarts, 1987) to increase the exploration of the solution space.

To analyze the real datasets, the simplified procedure involving separate sampling of periodic and piecewise parameters have been ultimately adopted, as described in Chapter 2. Adopting the strategy of sampling the two parameterizations separately, the number of sinusoids had to be fixed to avoid the possibility of CPs being modeled with a high number of sinusoids.

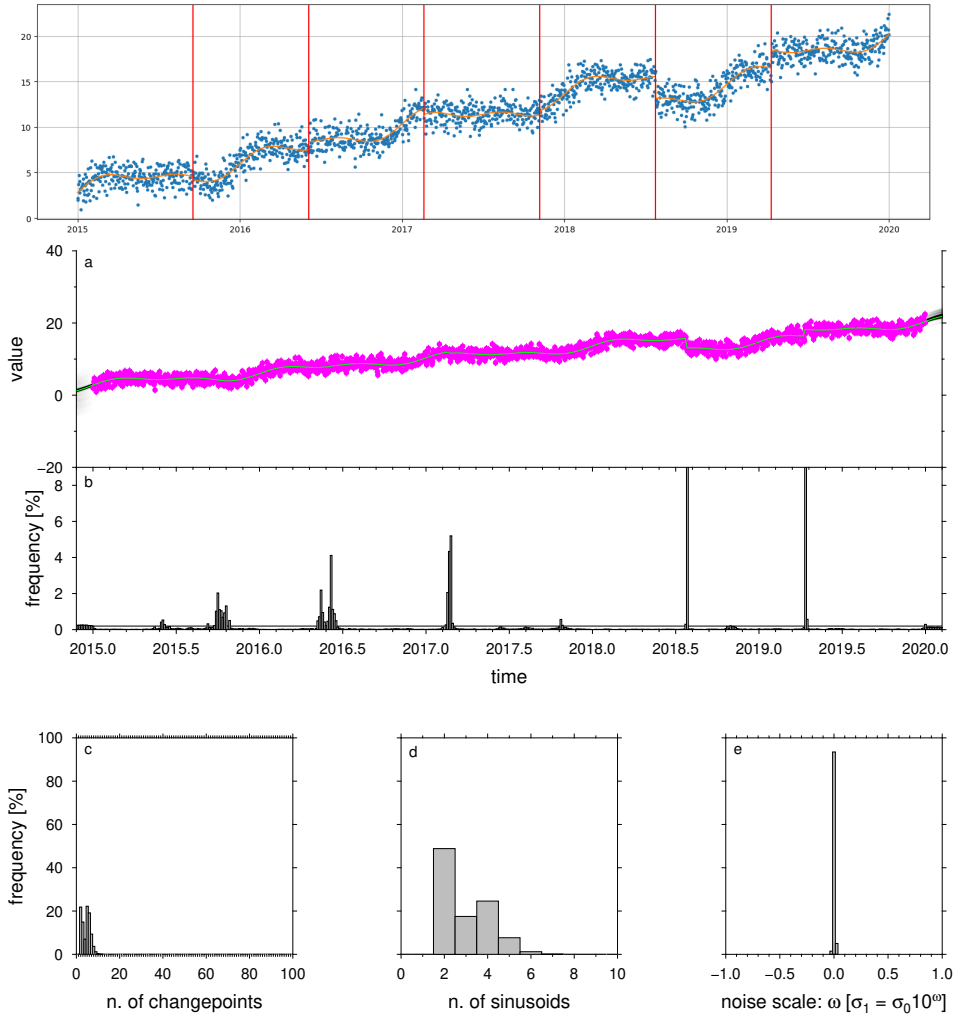

Figure 2.19: Synthetic test. Top panel shows the real model (orange) and the synthetic dataset (blue) with CPs highlighted in red. Panels are the same as in Figure 2.18.

3. Diffusion processes in minor normal faulting seismic sequences monitored by the Alto Tiberina Near Fault Observatory (Northern Apennines, Italy)

3.1 Introduction

The role of fluids in seismogenesis have been extensively studied and diverse mechanisms have been invoked to explain their involvement in earthquakes generation processes and seismic sequences evolution: dilatancy related to changes in shear stress and/or mean stress, localized post-seismic fluid redistribution around rupture, and post-seismic discharge of fluids from over-pressured portions of the crust (Sibson, (1994) and Sibson, (2000) and references therein). The effect of pore-fluid flow in triggering seismicity has long been suggested (Nur and Booker, 1972), and specific characteristics and patterns related to possibly fluid-driven seismic activity have been highlighted both in aftershocks (Noir et al., 1997; Antonioli, 2005), swarm like sequences (Saccorotti et al., 2002; Hainzl, 2004; Ross and Cochran, 2021) as well as in induced seismicity episodes (Talwani, 1984; Shapiro et al., 1997; Shapiro et al., 1999).

A typical signature referred to a possible ongoing diffusive process is commonly visible in the spatiotemporal migration of hypocenters, where the seismic cloud expands over time. This feature has been initially used by Shapiro et al., (1997) to model a pore-pressure perturbation caused by fluid injections into a borehole. They solved the diffusion equation for the case of a point pore-pressure source in a homogeneous isotropic saturated poroelastic medium and the distance of the propagating pore-pressure front from the source is estimated as

$$r = \sqrt{4\pi D t}, \tag{3.1}$$

where t is time from the injection of fluid, D is hydraulic diffusivity ($m^2 s^{-1}$) and r is distance (radius, in a homogeneous isotropic medium). The indication is that the triggering of an earthquake may occur at any time t at location r after the pore-pressure front reaches point r. So, in a $(r-t)$ plot, the fluid triggered seismicity should lie below the parabola defined by Shapiro et al., (1997) as "triggering front" (see Equation 3.1). Whilst, if earthquake triggering occurs immediately after the pore pressure perturbation, in a $(r-t)$ plot we should observe a relatively narrow parabolic cluster of seismicity along that parabola. Both these approaches measure the distance between hypocenters, thus considering that pore-pressure relaxation

begins at the origin time and hypocenter of the causative shock. This method has then been widely used to diagnose diffusive processes in natural earthquakes (Duverger et al., (2015), Ruhl et al., (2016), De Barros et al., (2019), De Barros et al., (2020), and Ross and Cochran, (2021) among others).

In the last twenty years in Italy, and specifically for extensional seismic sequences occurred along the Apennines, there is increasing indication of interaction between fluids and earthquakes. Miller et al., (2004) shown that the 1997 Umbria-Marche seismic sequence may have been driven by a fluid pressure pulse generated from the coseismic release of trapped high-pressure fluids by comparing aftershocks hypocenters to a modeled high-pressure diffusion front. Antonioli, (2005) observed that the spatio-temporal migration of seismicity is consistent with fluid flow and Lombardi et al., (2010) showed that the temporal increase of background seismicity rate can be explained by transient perturbation caused by poroelastic relaxation. The 1997 sequence was anticipated by foreshocks activity and Ripepe, (2000), following a dilatancy model, suggested that the relatively low V_P/V_S ratio observed before the main shocks could indicate the presence of fluid in the focal volume.

Fluids played a key role also in 2009 L'Aquila seismic sequence. Malagnini et al., (2012) shown that the Campotosto fault experienced a strong reduction in shear strength due to the increasing pore fluid pressure triggered by the previous main event. Di Luccio et al., (2010) observed that pore pressure diffusion controls the space-time evolution of aftershocks and detected a related increment in V_P/V_S values. V_P/V_S is also analyzed by Lucente et al., (2010) supporting a dilatancy phenomenon before the mainshock. Basing on the analysis of focal mechanisms Terakawa et al., (2010) proposed for L'Aquila a scenario involving high pressure fluids at hypocentral depths, similar to what described by Miller et al., (2004) for the Umbria-Marche sequence.

More recently fluid involvement have also been hypothesized for the 2016-2017 Central Italy seismic sequence by Chiarabba et al., (2018) highlighting V_P/V_S anomalies imaged by tomographic models and by Malagnini et al., (2022) analyzing seismic attenuation and space-time distributions of events.

The case of a fluid-driven seismicity is supported by the presence of a known

source of fluids in the inner sector of the Apennines. Chiodini et al., (2004) reported in fact about large CO_2 degassing sources, whose isotopic composition suggests a deep origin due to metasomatization of mantle rocks. Despite the still debated origin of the CO_2, its flux sharply decreases in a narrow band where most of the chain seismicity concentrates, suggesting a contributing role for fluids overpressure in seismogenesis.

In the northern Apennines specifically, fluid overpressure has been measured in San Donato and Santo Stefano boreholes at 85% of lithostatic load (Chiodini et al., 2004) at a depth of 4.8 km and 3.7 km (boreholes reporeted in Figure 3.1). This evidence is framed in the context of a E-NE-trending active extension at about 3 mm/yr (Serpelloni et al., 2005), ongoing since the Quaternary and following an upper Miocene-lower Pliocene compressional phase always eastward trending (Chiaraluce et al., 2007). The main geological units composing the resulting thrust and fold belt are summarized in (Mirabella et al., 2011) as: (1) the top of the basement (late Paleozoic-middle Triassic); (2) Triassic Evaporites; (3) a carbonate multilayer platform (lower Jurassic-Oligocene); and (4) foredeep turbidites (early Miocene). The overpressure observed in the deep boreholes are probably due to the accumulation of gas in crustal traps favored by the tectonic setting together with the presence of the Evaporitic layer (Trippetta et al., 2010; Trippetta et al., 2013).

Thus, given the marked presence of fluids in the subsurface of this sector of the Northern Apennines, their claimed involvement in the generation of the seismic activity of the area (Valoroso et al., 2017; Marzorati et al., 2014; Piana Agostinetti et al., 2017), and not least the presence of a very dense seismic network pertaining to the Alto Tiberina Near Fault Observatory (TABOO) (Chiaraluce et al., 2014), we have chosen to investigate the connection between fluids and seismogenesis by analyzing of the seismicity pattern of two small seismic sequences.

We report in Figure 3.1 the 2010-2015 seismicity of the study area as reconstructed by (Valoroso et al., 2017); the vast majority of the events occur in the hanging wall of a low angle normal fault, the Altotiberina fault (ATF) (Chiaraluce et al., 2007), within a complex system of synthetic and antithetic faults dissecting the stratigraphic succession and giving rise to seismic sequences with multiple

Mw3+ mainshocks.

The two small but prolific sequences we analyzed occurred nearby Città di Castello (CdC) and Pietralunga (see boxes in Figure 3.1). The first is active in April-May 2013 and the second starts in 2010 and reactivates in 2013 and 2014 (Figure 3.2). Following the approach proposed by Shapiro et al., (1997) we tested the hypothesis of the presence of a *triggering front* due to an evolving pore-pressure diffusion process, by checking for the parabolic signature in the spatiotemporal distribution of earthquakes.

Since the approach requires the largest as possible number of events and robust hypocentral locations (Shapiro et al., 2003), we generated a new earthquakes catalog for both sequences at higher resolution, by applying a deep neural network algorithm directly starting from the continuous waveforms provided by the very dense TABOO network consisting of a dense array of seismometers installed both at surface and in boreholes, allowing a very low detection threshold. The high detection capability of TABOO network and the superior performance of artificial intelligence-based algorithms, allow us to treat local small earthquakes ($3 < M_L < 4$) as a mainshock leading to a proper aftershock sequence. These kind of algorithms have in fact rapidly spread for the phase detection task because of a superior performance with respect to standard picking algorithms, on speed, recall and accuracy (Ross et al., 2018; Zhu and Beroza, 2019; Mousavi et al., 2020).For our application in a low-magnitude and high-seismicity context, another advantage to consider is the increased capability of these methods in detecting small magnitude earthquakes (Mousavi et al., 2020; Park et al., 2020), which can enhance not only the number of detected events, but also the robustness of the features that we aim to investigate.

This area is therefore a perfect place to investigate the evolution of minor extensional seismic sequences and the effects of fluids on seismogenesis because of 1) the presence of a known source of CO_2 at depth; 2) the geologic and tectonic setting favoring fluid overpressure and 3) the presence of a high-resolution data. Thus, once relocated the new catalogs of the CdC and Pietralunga sequences generated by means of a state of art detection, association, and location procedure, we verify

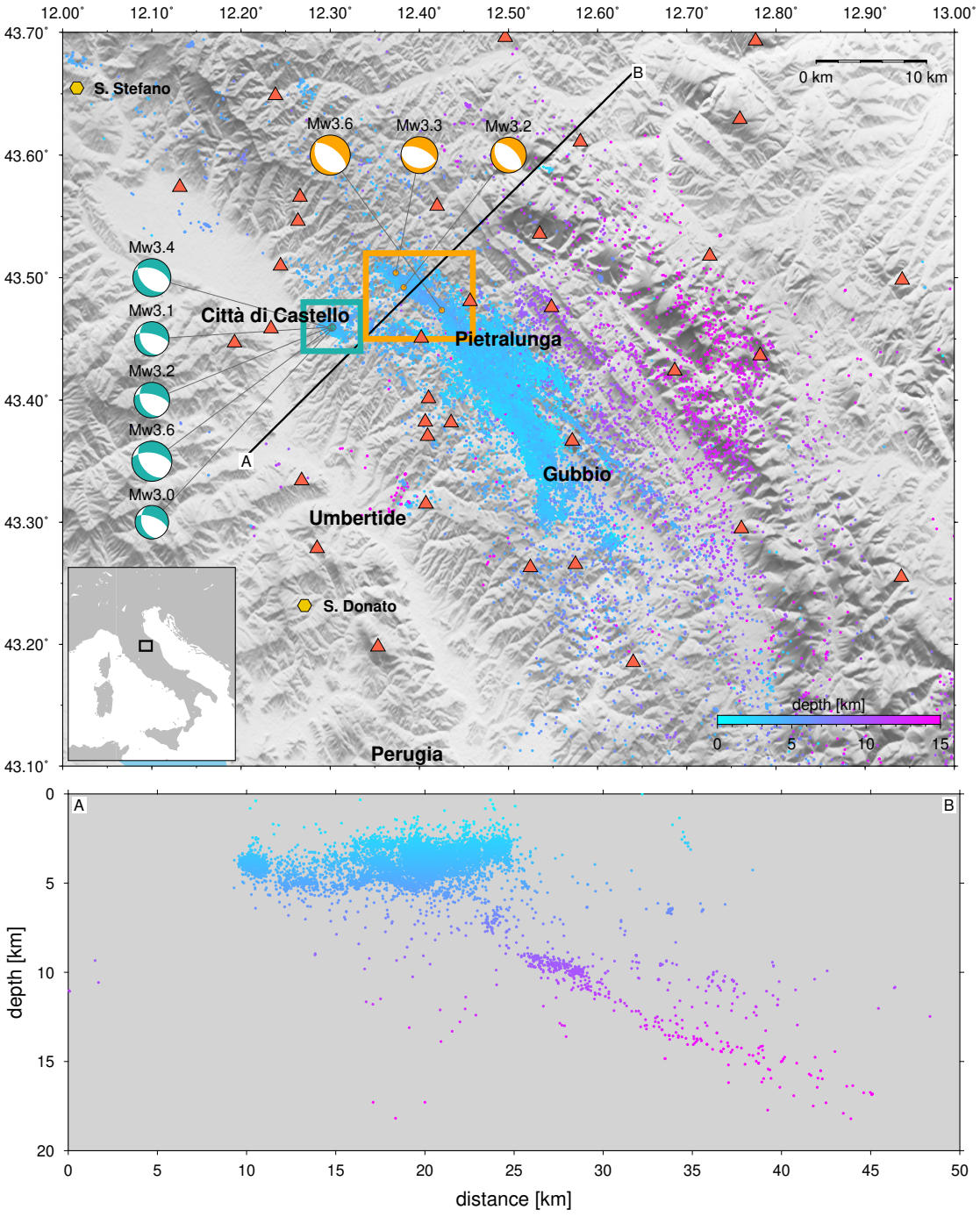

Figure 3.1: Top: map of the study area (events from Valoroso et al., (2017) catalog). Città di Castello and Pietralunga areas are marked respectively by green and orange rectangles. Red triangles represent seismic stations. Locations of Santo stefano and San Donato boreholes are marked with yellow hexagons. Focal mechanisms from (`http://cnt.rm.ingv.it/tdmt`. Bottom: cross section (width=20 km) highlighting the high seismic activity at shallow depths (<7 km) and the ATF visible at higher depths, dipping toward NE.

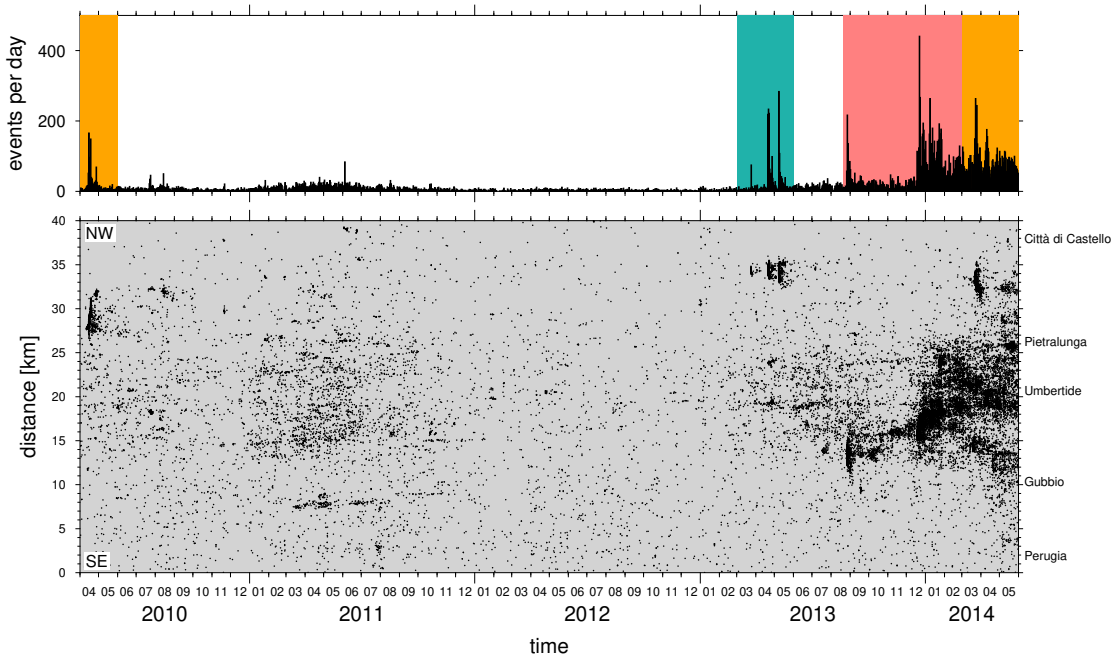

Figure 3.2: Events per day (top) and space-time distribution of seismicity (bottom) in the TABOO area. Città di Castello and Pietralunga sequences are highlighted in green and orange, as in boxes of Figure 3.1. Gubbio sequence, not treated here, is highlighted in red.

if the seismicity distribution is consistent with a diffusive pattern. Finally, we compare the located earthquakes with a deterministic seismostratigraphic model of the upper crust (from Latorre et al., (2016)), to investigate the possible contribution of the lithology in controlling seismic activity evolution.

3.2 Data analysis

3.2.1 Event detection, picking and association

Our workflow consisted of five main steps (see Figure 3.3): (1) continuous waveforms retrieval; (2) application of PhaseNet (Zhu and Beroza, 2019), a deep-neural network algorithm for phase detection and picking; (3) phase association and preliminary location using REAL (Zhang et al., 2019); (4) absolute earthquake location with NonLinLoc (NLL), a probabilistic non-linear hypocenter inversion algorithm (Lomax et al., 2000; Lomax et al., 2014); (5) and a final relocation of a subset of the events using hypoDD. We started by collecting raw seismic data (e.g., continuous waveforms) recorded by TABOO network consisting of about 55 stations (whose distribution is showed in Figure 3.1) including very- and short-periods plus broad band three-component sensors. All the recordings, including the DH channels, have been re-sampled at 100 Hz. We selected three specific time windows corresponding to the CdC (from March 2013 to May 2013) and Pietralunga (from April 2010 to May 2010 and then from March 2014 to May 2014) seismic sequences; see Figure 3.2). The seismicity peaks observed from the second half of 2013 to the first months of 2014 pertain to the Gubbio sequence (red highlight in Figure 3.2). For this swarm-like sequence, not treated here, seismic and geodetic data reveal aseismic contribution in the deformation process (Gualandi et al., 2017).

After preparing the waveforms, we have applied a deep-neural-network-based seismic arrival-time picking algorithm (PhaseNet, Zhu and Beroza, (2019)), allowing us to detect P- and S-waves arrival time in the seismic waveforms. We ran PhaseNet on the three-component waveforms high-pass filtered at 1 Hz. The high-pass

Figure 3.3: Scheme of the adopted workflow.

filtering further increases the detection capabilities of PhaseNet, by reducing the low frequency noise with respect to the unfiltered datasets used in the training phases (Zhu, p.c.). In other words the algorithm has learned the features of P and S arrivals in more complicated cases, thus its work with filtered waveforms is facilitated. Traces have been processed in chunks of 30 s length with 50% overlap in order not to cut earthquake signals. PhaseNet outputs probability distributions of P arrivals, S arrivals and noise and is designed such that peaks in probability mark the arrival time of the phases. To define a P or S phase we used a threshold value for peaks probability of 0.5, as in the original work from (Zhu and Beroza, 2019) and in other PhaseNet applications (Park et al., (2020); Liu et al., (2020)). Duplicated picks from overlapping windows are identified on the basis of their timestamps and the one with higher probability is retained. The total number of picked P- and S- phases amounts respectively to 3611057 and 3286647. Picks probability distributions for the three considered time windows highlighted in Figure 3.2 are reported in Figure 3.4.

The output P and S picks from PhaseNet are used as input for REAL, which associates and initially locates earthquakes primarily through counting the number of P and S picks and secondarily from travel time residuals by grid searching (Zhang et al., 2019). We kept most of REAL parameters to default ones: the grid step is 0.02° in horizontal and 2 km in depth reaching 30 km; the searched area is 0.25° centered at the station recording the initiating P phase. Minimum time distance between 2 events is kept at 5 seconds. Thresholds for P phases and total phases

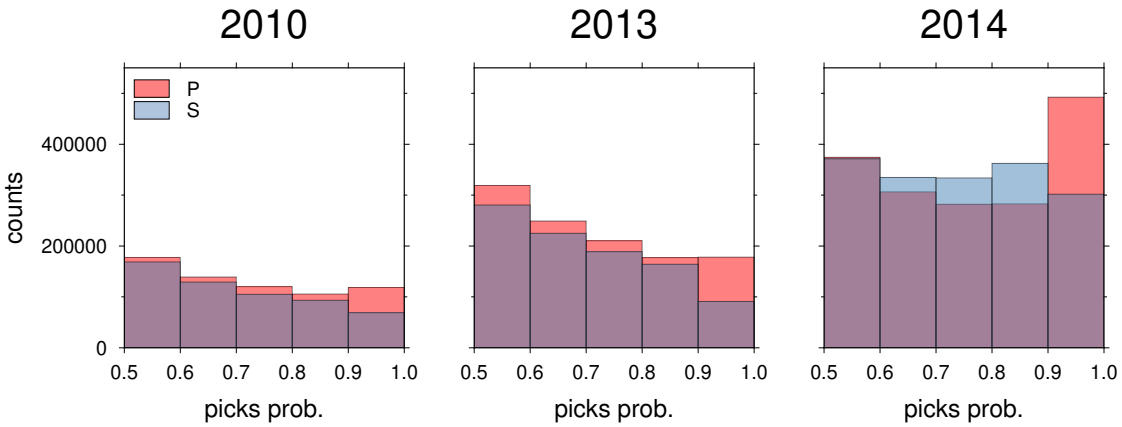

Figure 3.4: Picks probabilities for each of the three considered time windows.

are kept respectively at 3 and 5. These thresholds have been intentionally left to low values because of two reasons: the first one is that we are mostly working with microseismicity and in this context a low number of phases is expected for very small local earthquakes; the second one is since the catalog will be refined in the next steps anyway. So, our goal for this preliminary location is to keep as many events as possible. Acknowledging that we are probably carrying more false events, we aim to filter them out in the next steps, after assessing their quality not only on the number of phases but considering more constraints from location parameters and uncertainties. As in Ross and Cochran, (2021) no magnitude estimation have been made because it is not needed for our purposes and, dealing with very small events, it is an inherently difficult task. We retrieved magnitudes for our final catalog by merging the events with the corresponding events in the catalog of Valoroso et al., (2017).

3.2.2 Events location

Firstly, we located all the retrieved events in absolute terms using the NonLinLoc (NLL) (Lomax et al., 2000). To compute the hypocentral locations we initially established the more suitable NLL code settings and tuning parameters for the TABOO area and network. The 1-D velocity model chosen for the locations is the one proposed by Carannante et al., (2013) because it generates the best residual

Pick probability (p)	Weight	Error [s]
$p \geq 0.85$	0	0.05
$0.7 \leq p < 0.85$	1	0.10
$0.6 \leq p < 0.7$	2	0.25
$0.5 \leq p < 0.6$	3	0.50

Table 3.1: Picks weighting scheme.

distribution and formal errors. Then, to partially overtake the oversimplified assumption of using a 1D velocity model in the generation of the starting locations, we used mean static station corrections available for the whole TABOO network (from Bagh p.c.).

NLL (Lomax et al., 2000) is based on a probabilistic, nonlinear formulation of the earthquake location problem, thus it does not provide a single "best" solution to the hypocenter location but an estimation of the posterior probability distribution for the hypocentral coordinates and origin times, allowing us to retrieve a complete description of location uncertainties. To this regard, we must point out that as with most machine learning picking algorithms, PhaseNet does not provide an estimate of the time uncertainty on the picks, but only a probability value (going from 0 to 1; the last picking has the largest probability to be correct). Thus, to perform the probabilistic hypocenter location with NLL we weighted the picks according to their associated PhaseNet output probability (see Table 3.1), as in e.g. Tan et al., (2021) and Liu et al., (2020).

The NLL output catalog counts more than 160k events, located employing $\sim$ 2.4M phases (P- and S-), thus using about one third of the total detected phases.

To classify the quality of our absolute locations, we adopted the procedure proposed by Michele et al., (2019) to the resulting catalog. These authors proposed a criterion to assess the location quality, consisting of the combination of the uncertainty estimates, properly normalized, provided by the NLL code. The procedure quantifies the location quality in terms of a unique numeric normalized value, named quality factor (qf), which varies between 0 (best) and 1 (worst). Then, the obtained location is assigned to a quality class depending on the qf parameter

according to the scheme: A-class ($0 < \text{qf} \leq 0.25$), B-class ($0.25 < \text{qf} \leq 0.50$), C-class ($0.50 < \text{qf} \leq 0.75$) and D-class ($0.75 < \text{qf} < 1.00$). We report in Figure 3.5 the statistical distribution of the location parameters of the NLL output catalog, grouped by quality classes.

Even if we are mainly interested in the relative distribution of the seismic events, after ranking all the earthquakes locations we selected only class A and B as input for the relocation process. Quality classes for the whole TABOO area and for CdC and Pietralunga areas (boxes in Figure 3.1) are shown in Figure 3.6. The selected subsets have been relocated with HypoDD (Waldhauser, 2000) using travel times of P and S waves. DD algorithm assumes that, if the hypocentral separation between two earthquakes is small compared to the event-station distance, the ray paths are similar along almost the entire ray path. Therefore, the difference in travel times for two events observed at one station can be attributed to the spatial offset between the events with high accuracy. In this way, the ray paths between the source region and a common station can be considered similar along almost the entire ray path. Therefore, the difference in travel times for two events observed at one station can be attributed only to the spatial offset between the events, and the latter can be computed with high accuracy by differencing Geiger's equation for earthquake location (see Waldhauser, (2000) for a comprehensive review). HypoDD solves the minimization problem by weighted least squares (LSQR, Paige, (1982)) using the conjugate gradients method. Improved hypocenters are found by iteratively adjusting the vector difference between hypocentral pairs. Hypocentral parameters and partial derivatives are updated after each iteration. We estimated hypocentral errors by using the singular value decomposition (SVD) (see Waldhauser, (2001)) for a subset of 200 earthquakes for each sequence, because error estimates obtained with LSQR are overly optimistic. Thus, from the initial set of earthquakes, we end up with 22k relocated events for Pietralunga and 9k for CdC with most of the locations possessing formal horizontal and vertical errors, very similar for the two sets, respectively 50 m and 100 m (See histograms in Figure 3.7). Consequently, structural and patterns details with 500 m of dimension can be interpreted.

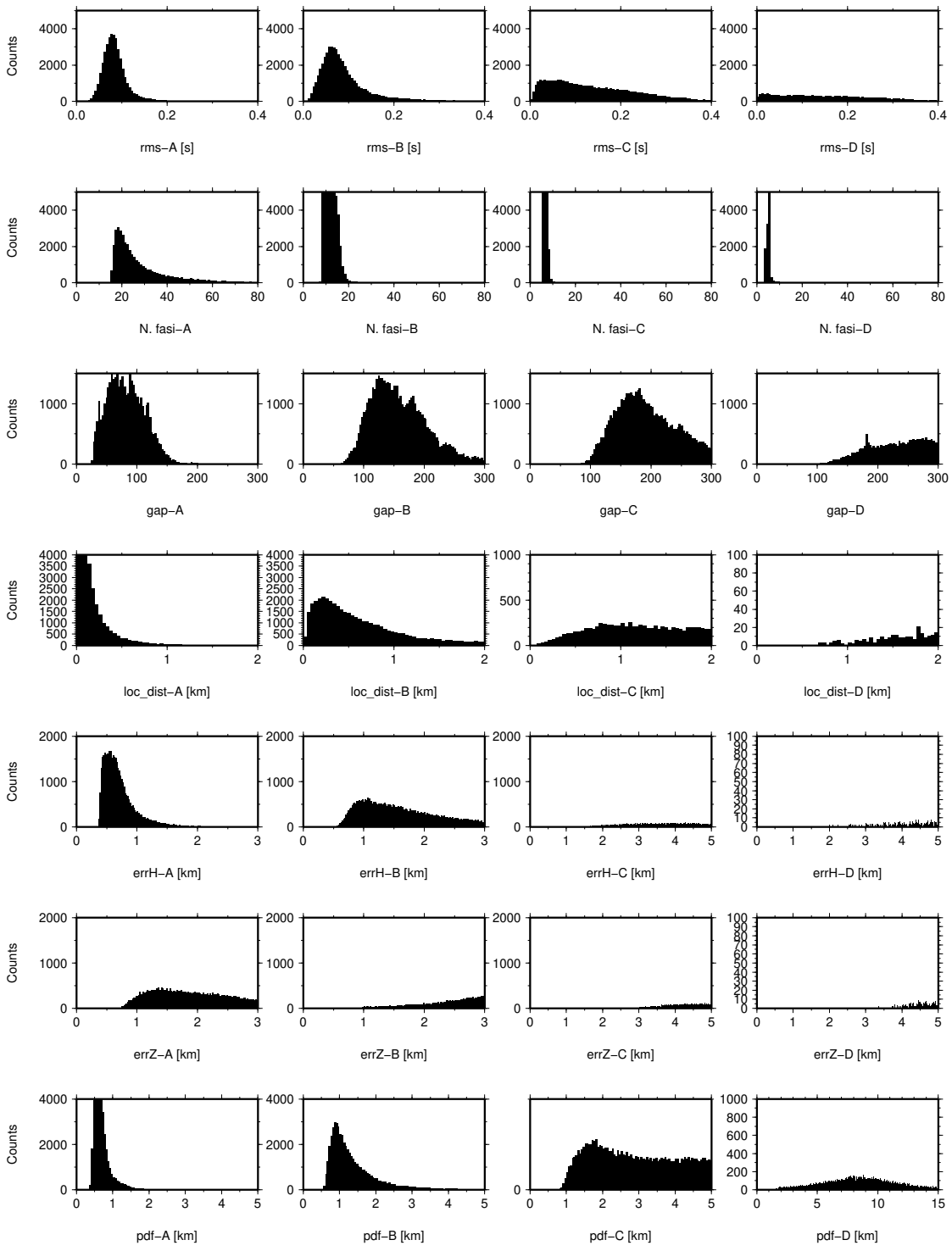

Figure 3.5: Location parameters grouped by quality classes (A-D).

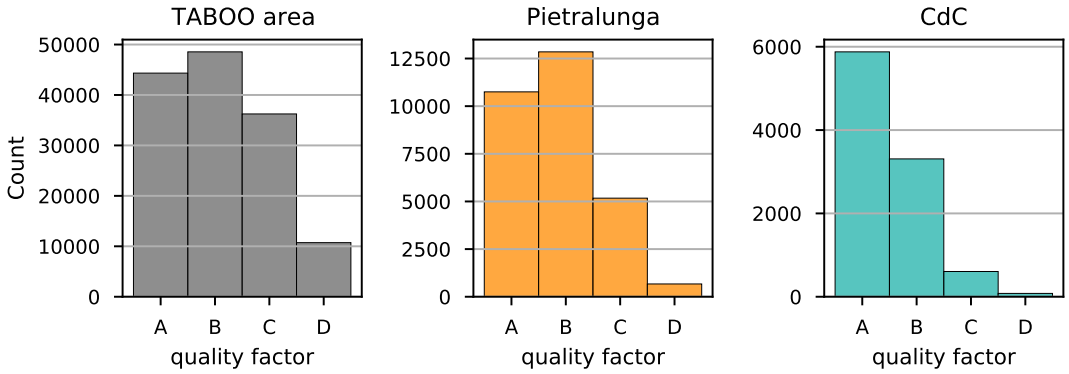

Figure 3.6: Events count for each quality class in the whole TABOO area (left), Pietralunga (center) and CdC (right).

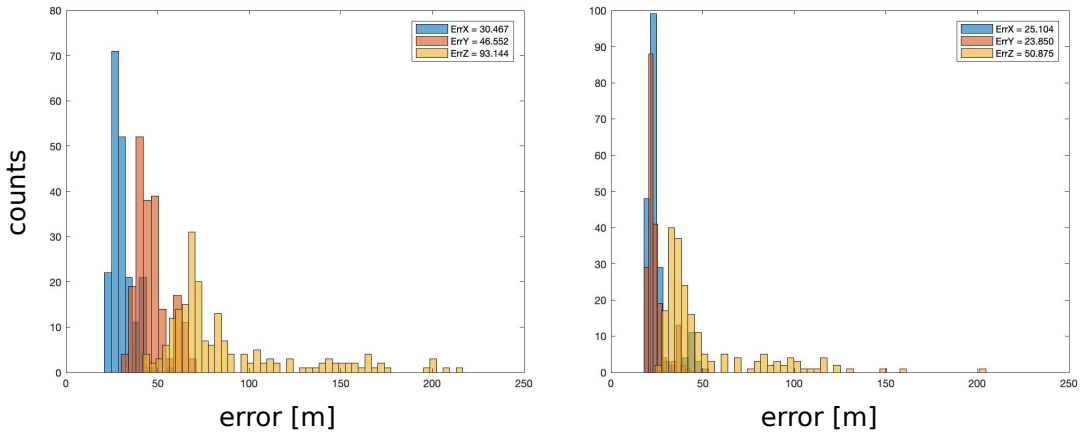

Figure 3.7: Horizontal and vertical errors for the CdC (left) and Pietralunga (right) relocated catalogs.

3.3 Results

The final relocated catalog of CdC and Pietralunga sequences counts 6 times the number of events detected in previous catalog (Valoroso et al., 2017): the events number increased from 1.5k to 9k events for CdC and from 3.1k to 20k events for Pietralunga Figure 3.8. For both sequences the spatial distribution of events is similar to the catalog of Valoroso et al., (2017) but with a much higher density of events, proving the reliability of the new detections. In Figure 3.9 we show the cumulative number of events vs time. The curves have similar shape, with corresponding times of sudden increases in both catalogs, but with a marked difference in the absolute number of detected events.

In the following we analyze the seismicity patterns of both sequences and we discuss relations with hosting lithologies.

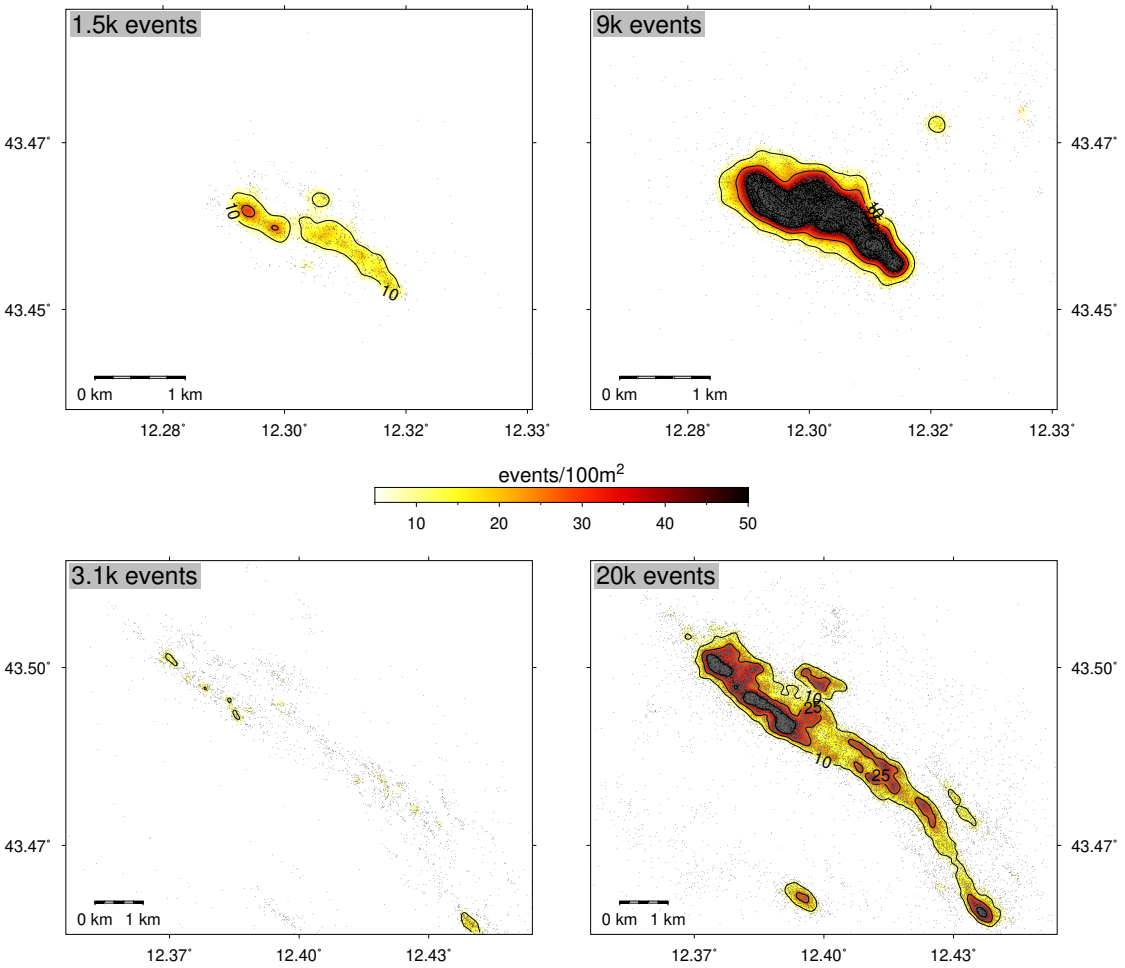

Figure 3.8: Comparison of events detected in this study (right column) vs previous catalog (left column) for Città di Castello (top row) and Pietralunga (bottom row).

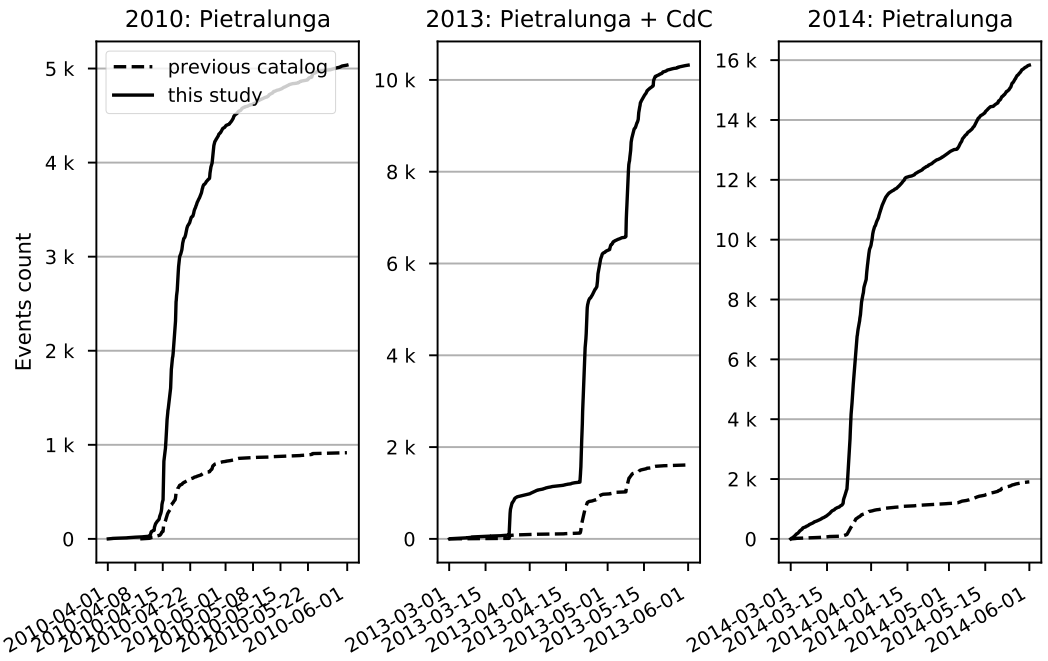

Figure 3.9: Cumulative number of relocated events in comparison with previous catalog for the considered time windows and areas.

3.3.1 Pietralunga sequence

The Pietralunga sequence involves a fault system of $\sim$ 10 km of length (Pietralunga fault, PIEF), NW-SE oriented, with segments activated in different periods from 2010 to 2014. In Figure 3.10 we show the seismicity distribution in map view (a), cross sections orthogonal to the PIEF (b) and space time diagrams grounded on along strike sections (c).

The sequence starts in April 2010 with five days of foreshocks (maximum magnitude $< M_L 1.7$, Valoroso et al., (2017)) leading to the M_W 3.6 earthquake occurred on April 15 2010. From this event onwards, an along-strike migration begins, which is represented with red dots in Figure 3.10, and specifically in the along strike plot of panel (c). The NW trending migration velocity is compatible with the value of 0.4 km/day from Marzorati et al., (2014). The 2010 sequence activated the south-eastern portion of the PIEF:a 6 km long segment well imaged in the cross sections of Figure 3.10(b) with seismicity dipping NE at 60°, compatible with time domain moment tensor focal mechanism solution of the mainshock (`http://cnt.rm.ingv.it/tdmt`). A shorter antithetic fault is also activated in the same period, well resolved here thanks to the high number of detected events. The mainshock is located at the deeper termination of the two faults (Figure 3.10(b), section 3). A high seismic activity persist for about 3 weeks in total.

The $r - t$ plot we report in Figure 3.11 shows the foreshocks moving toward the M_W 3.6 mainshock, whose occurrence coincides with the beginning of a clear diffusive pattern, with an estimated hydraulic diffusivity value of about $2m^2s^{-1}$. The earthquakes with magnitude $\geq M_L 2.5$ are all positioned near the triggering front. After less than 3 days from the mainshock the seismic cloud moves away from the source and the rock volume near the mainshock becomes characterized by a very low density of events. This pattern is compatible with a backfront signature (Parotidis et al., 2004; Segall and Lu, 2015) observed in induced seismicity but also in natural earthquakes (Ross and Cochran, 2021), and it is a characteristic related to the end of the fluid injection transient.

After 3 years of quiescence, seismic activity starts again in March 2013 (yellow dots in Figure 3.10) with a M_W 3.3 taking place in the northern section of the

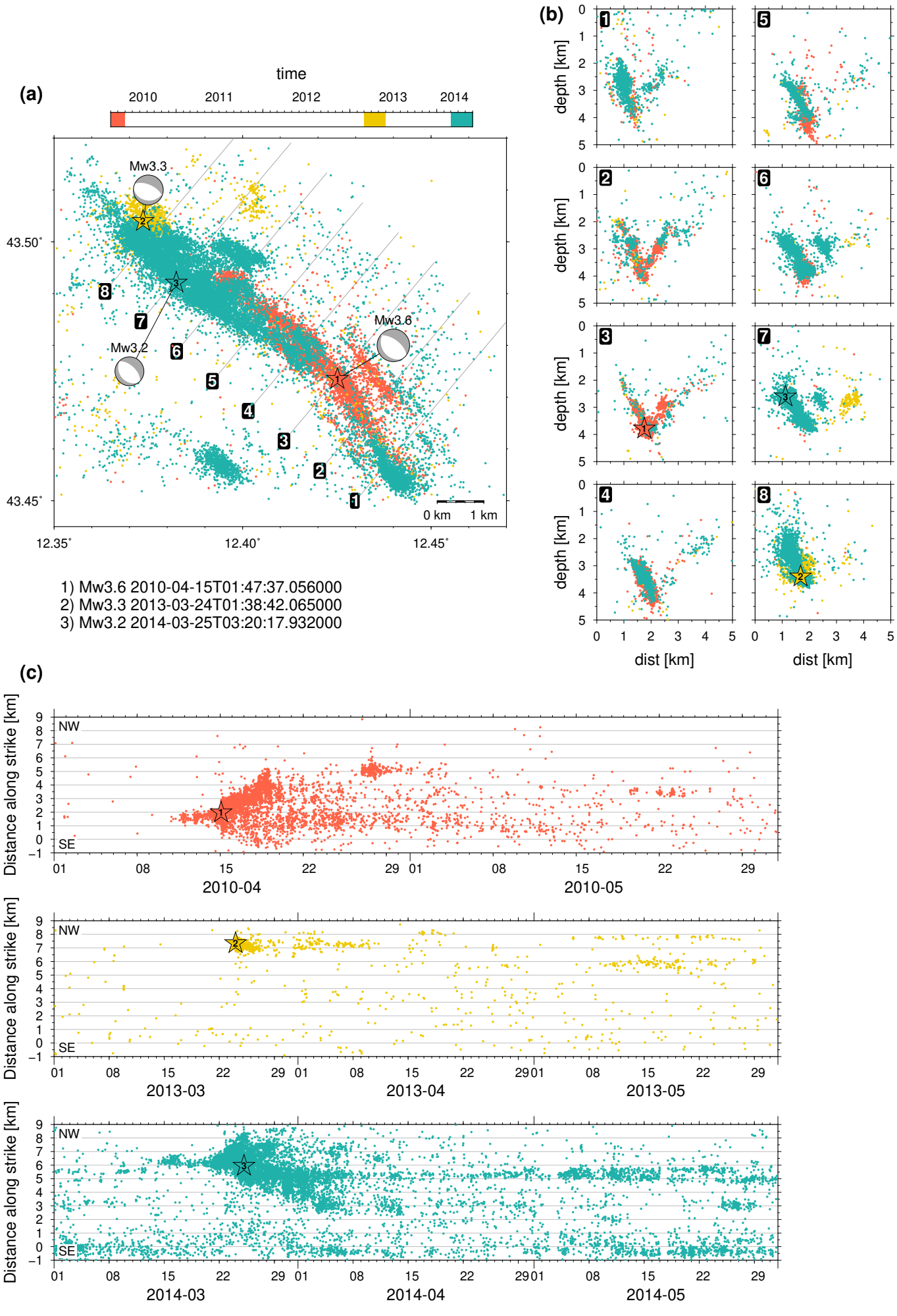

Figure 3.10: Pietralunga sequence: map (a), cross sections (b) and along-strike space-time diagram (c). Distance = 0 in along-strike plots corresponds to section 1.

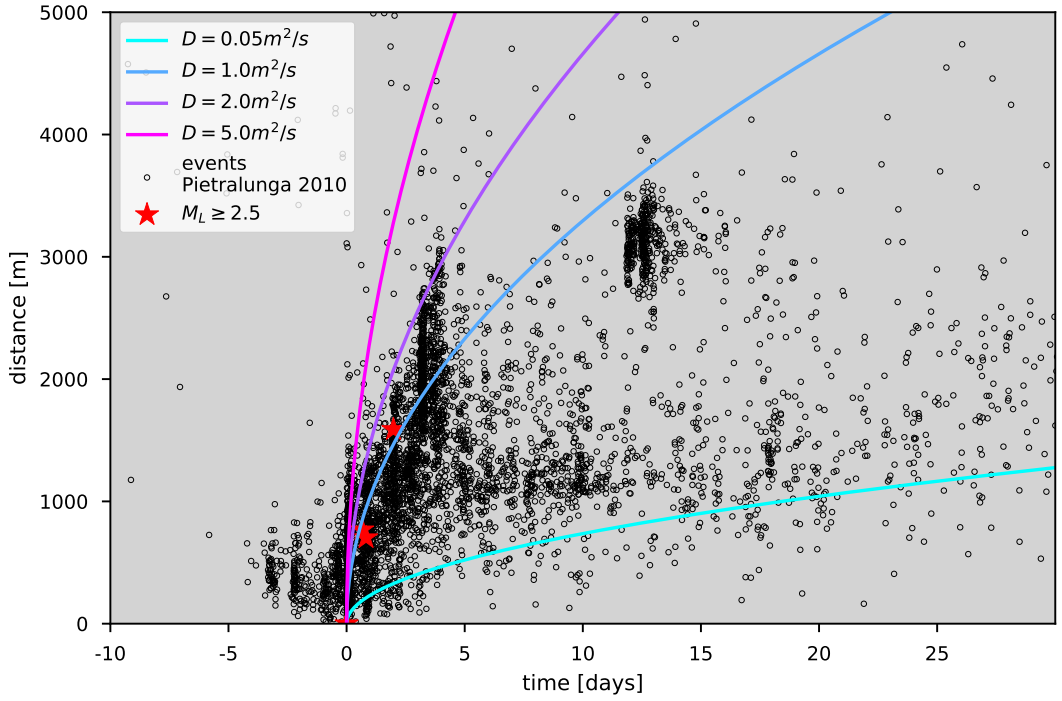

Figure 3.11: r-t plot Pietralunga 2010

PIEF system. This part of the sequence does not show an along strike migration (Figure 3.10 (c)), and appears to be clustered in a short segment of the fault, ~ 1 km long. The seismic activity continues weeks later and seems to occur on a short (< 0.5 km) antithetic fault, eastward to the main fault (Figure 3.10(b), section 7).

Seismicity renews in March 2014; events are represented with blue dots in Figure 3.10. This part of the sequence fills the gap between the previous phases of 2010 and 2013, occurred respectively in the south-eastern and north-western sections of the PIEF. Foreshocks are detected, similarly to the 2010 parts. This time the foreshocks do not lead to a clear mainshock but to multiple (relatively) major earthquakes with $M_L \geq 2.5$ and a $M_W 3.2$ on March 25. The along-strike migration is again visible (Figure 3.10 (c)) and takes place in both NW and, more extensively, SE direction, respectively with velocities of ~ 0.5 and 0.25. his part of the sequence involves a NE dipping segment along with a minor parallel segment ~ 1 km north-east of the main one (Figure 3.10(b), sections 6-7). The SE migration also reactivates parts of the fault that were previously activated in 2010 (Figure 3.10(b), sections 4-5). This feature is not visible in the catalog of Valoroso et al., (2017), evidencing again the outcome of a higher detection capability. A sustained seismic activity in the south-eastern part is observed (Figure 3.10(b), section 1), which appears clearly unrelated to the described migration (blue dots in Figure 3.10(c) at distance ≤ 1 km). This is part of the Gubbio sequence occurring few kilometers SE (Figure 3.1).

Given that the 2014 events do not show a clear mainshock, essentially behaving like a small swarm, we tested multiple events as the origin of $r - t$ plots. In Figure 3.12 and Figure 3.13 we selected two diverse earthquakes with $M_L \geq 2.5$ occurring at the beginning of the migration pattern observed in Figure 3.10 (c), after the foreshocks. Using different earthquakes as origin does not hide the presence of the diffusive pattern but slightly changes the value of estimated hydraulic diffusivity: $< 1 m^2 s^{-1}$ in Figure 3.12 and $> 1 m^2 s^{-1}$ in Figure 3.13. In both cases the number of events near the origin point decrease markedly after 5 days, but activity continues at greater distances, which, similarly to 2010, suggest again the presence of a back front pattern compatible with the cessation of the fluid diffusion stimulation.

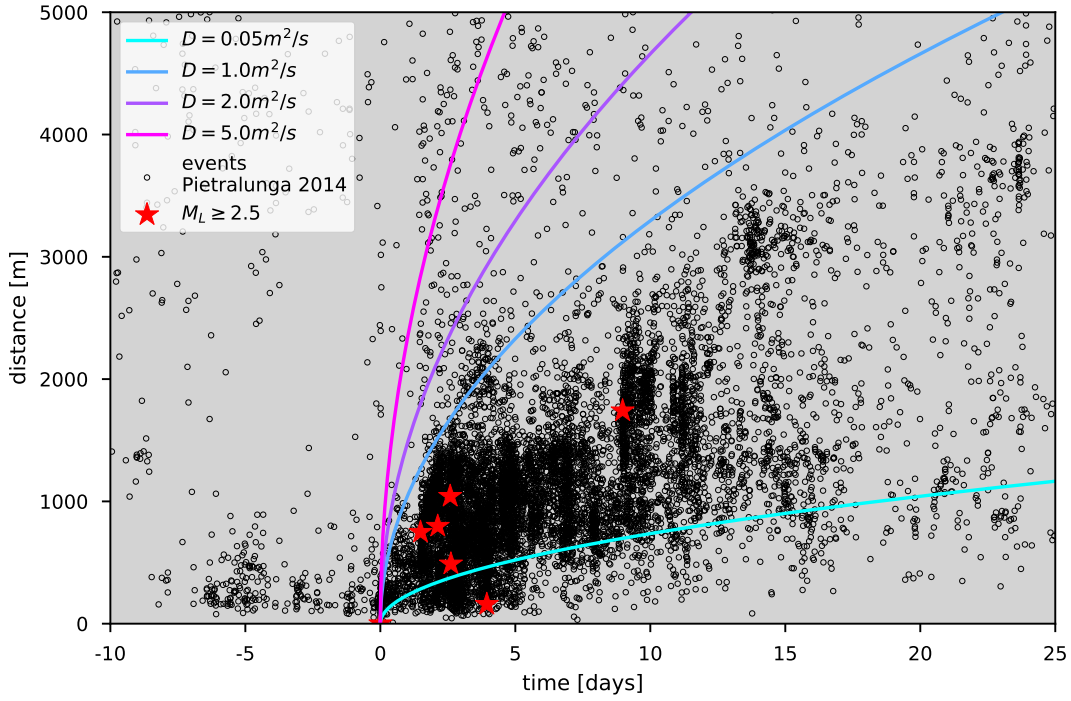

Figure 3.12: r-t plot Pietralunga 2014. Origin: 2014-03-21 04:41:02 M 2.5

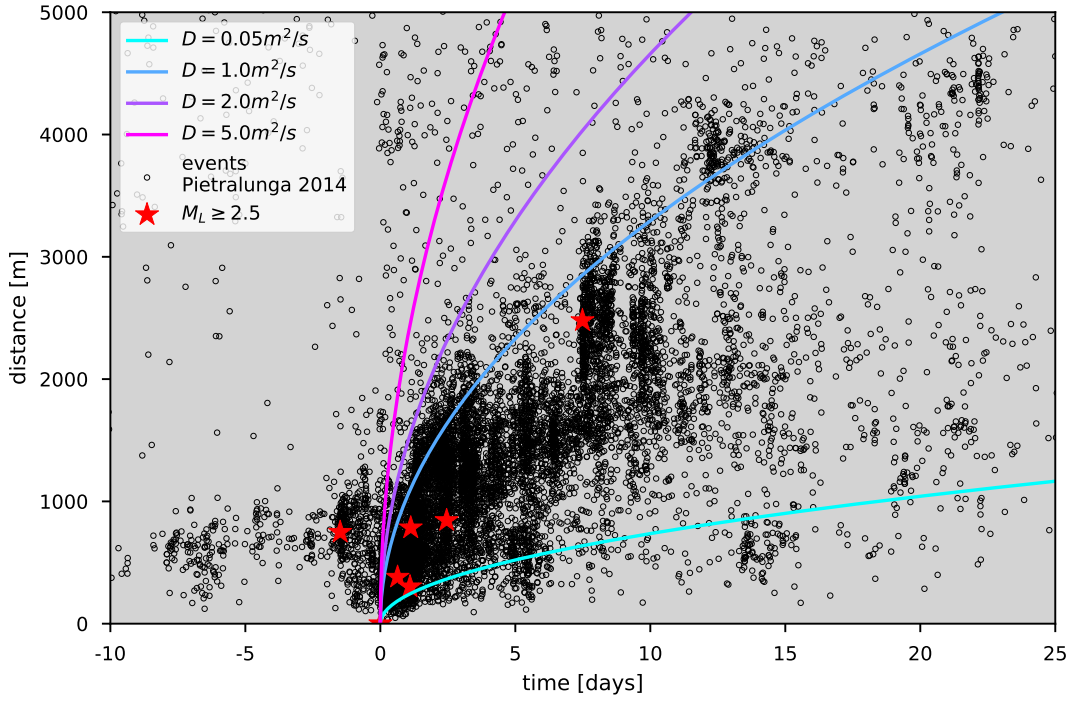

Figure 3.13: r-t plot Pietralunga 2014. Origin: 2014-03-22 16:23:19 M 2.6

3.3.2 Città di Castello sequence

The CdC sequence starts and ends between a brief time window of one month and involves a fault segment with a length of about 2.5 km. It can be temporally divided in two main bursts of activity: the first (red dots in Figure 3.14) starts on April 20 2013 with a M_W 3.4 earthquake, followed the next day by two other earthquakes (M_W 3.1 and M_W 3.2) nucleating very close to each other both in space and time. The migration of seismicity is evident for this first part of the sequence in the along strike direction (Figure 3.14(c)). Firstly, we observe a vertical (i.e., co-seismic) alignment of points in the first hours after the mainshock, probably due to a static stress transfer effect. Then, the distribution of seismicity starts showing a clear migration of $\sim$ 0.5 km/day. No foreshocks have been detected before the M_W 3.4 earthquake of April 20. After almost three days of lower seismic activity, the seismicity increases again (yellow dots in Figure 3.14), insisting on the central segment of the fault.

For the CdC sequence we noticed a scattering of the shallower events, possibly related to a larger depth uncertainty. This can be due to the combination of two factors: 1) the sequence is placed near the western boundary of the TABOO network, thus having a slightly poorer stations coverage; 2) the two seismic stations nearest to the sequence are unfortunately very noisy (CDCA) and with a persistent occurrence of gaps (BADI).

The $r - t$ plot relative to this first part of the sequence is shown in Figure 3.15. In this case the origin (i.e. the source of fluids) is not coincident with the first earthquake, but we moved it about 750 m. This value has been found with a grid search over possible sources near the M_W 3.4, of which the exact location has been checked by manually reviewing the associated P and S phases, to make sure it wasn't a mislocated event. Unlike with borehole fluid injection, in natural earthquake applications we do not know the exact location of the fluid source (Pacchiani and Lyon-Caen, 2010). Usually the source is assumed coincident with the first earthquake in a swarm or with the mainshock in a mainshock-aftershock sequence (as in the other r-t plots of the present work), but the initial propagation of pore pressure could as well be seismically silent (Duverger et al., 2015). Still, we

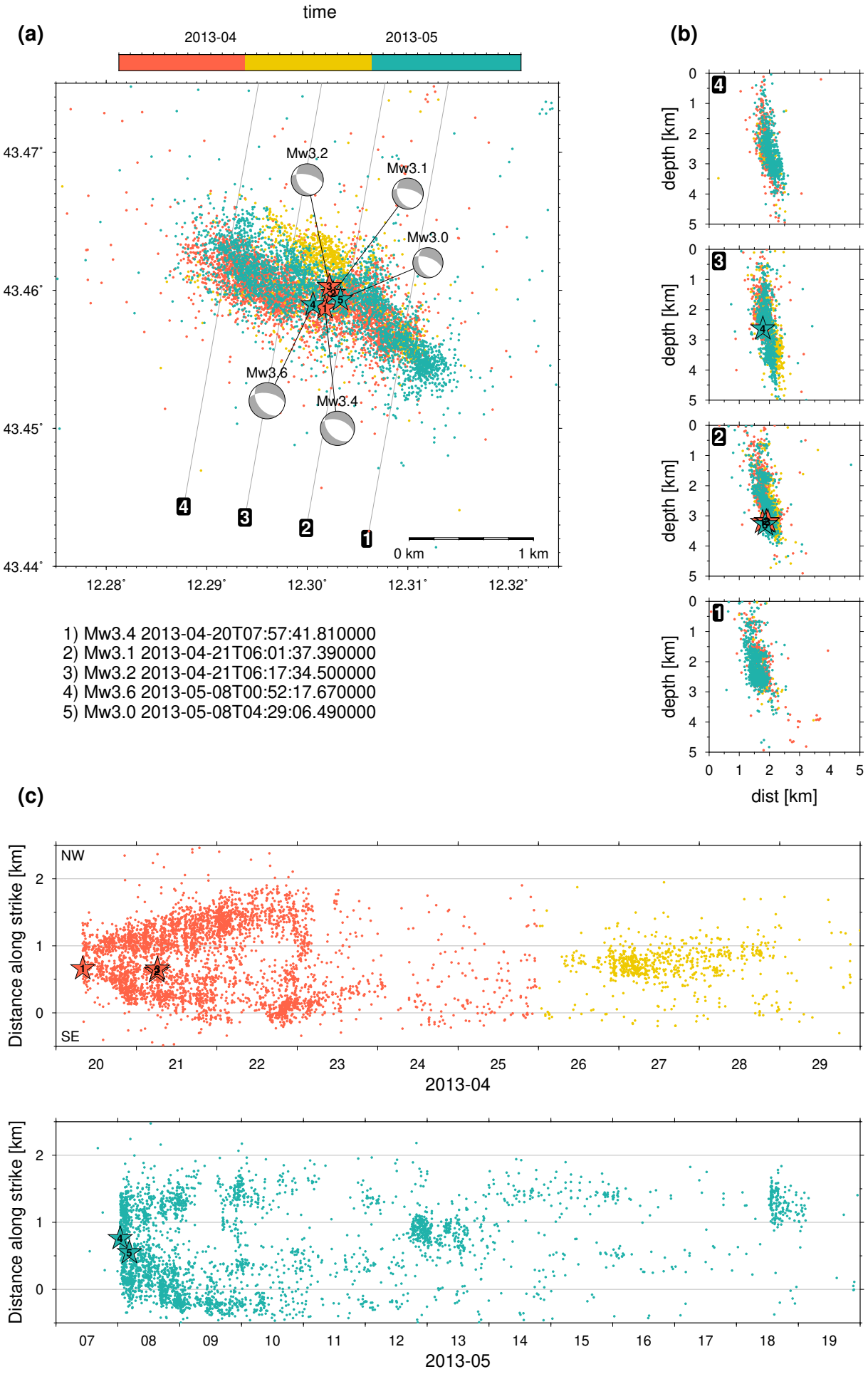

Figure 3.14: Città di Castello sequence: map (a), cross sections (b) and along-strike space-time diagram (c). Distance = 0 in along-strike plots corresponds to section 1.

decided to try using only the events located with arrival times from the two nearest stations (CDCA and BADI) to mitigate the aforementioned issue (Figure 3.15, bottom). Reference curves with different diffusivity values are shown along with the spatiotemporal distribution of hypocenters. A diffusive pattern is evident, with a hydraulic diffusivity value of $\sim 1m^2s^{-1}$.

The second burst of the CdC sequence (light blue dots in Figure 3.14) starts on May 8 with a M_W 3.6, followed less than 4 hours later by a M_W 3.0. The spatiotemporal evolution (Figure 3.14(c)) is similar to the previous one, but with a less defined migration along-strike. Seismicity moves away from the M_W 3.6 earthquake with an expanding front and reactivates in a small cluster in the middle section (12-13 May) after a couple of days of lower activity. The along-strike migration is less clear in this case, probably because of a stronger co-seismic triggering at the beginning. Interestingly, the mainshocks of both the first and second phase of the sequence are almost co-located in map (Figure 3.14 (a)) and in sections (Figure 3.14 (b), section 2).

The $r-t$ plot for the May portion of the CdC sequence is shown in Figure 3.16. The spatiotemporal pattern does not initially follow a diffusive pattern (in the first hours from the M_W 3.6), where the distribution of events is probably dominated by (co-seismically induced) static stress transfer effects. The diffusive pattern becomes evident from ~ 0.5 days, with a hydraulic diffusivity estimated between $1m^2s^{-1}$ and $2m^2s^{-1}$. It is in fact frequently and reasonably assumed that the evolution of a seismic sequence is not driven by a single mechanism, but more realistically by a complex interaction between both hydraulic and mechanical properties (Hainzl, 2004; De Barros et al., 2019; De Barros et al., 2021), the role of which can be primary or secondary at different stages of the sequence. As with Pietralunga sequence, the events move away from the origin point, leaving the rock volume near the source practically aseismic.

For both Pietralunga and CdC sequences our hydraulic diffusivity estimates $(1-2m^2s^{-1})$ lie well within the $0.1-10m^2s^{-1}$ interval, commonly observed for induced and natural earthquakes (Talwani, 1984).

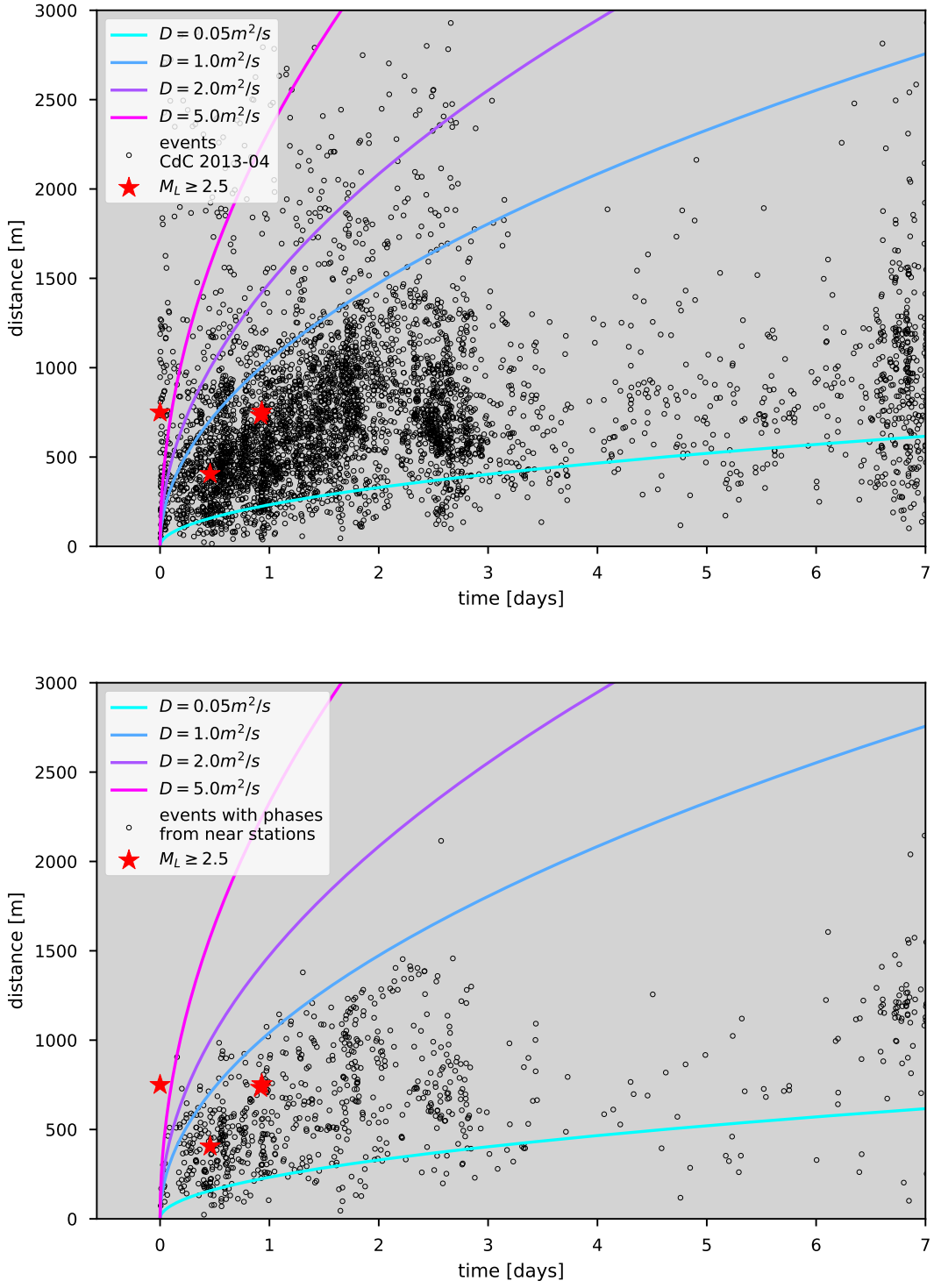

Figure 3.15: $r - t$ plot for the first part of Città di Castello seqeunce (April 2013). The time of the origin of the plot is set on the M_W 3.4 of 2013-04-20 00:52:17, the origin is moved in space to better describe the diffusive pattern. Explanation for events selection and origin shifting can be found in the text. Top: all events. Bottom: only events with phases from nearest stations.

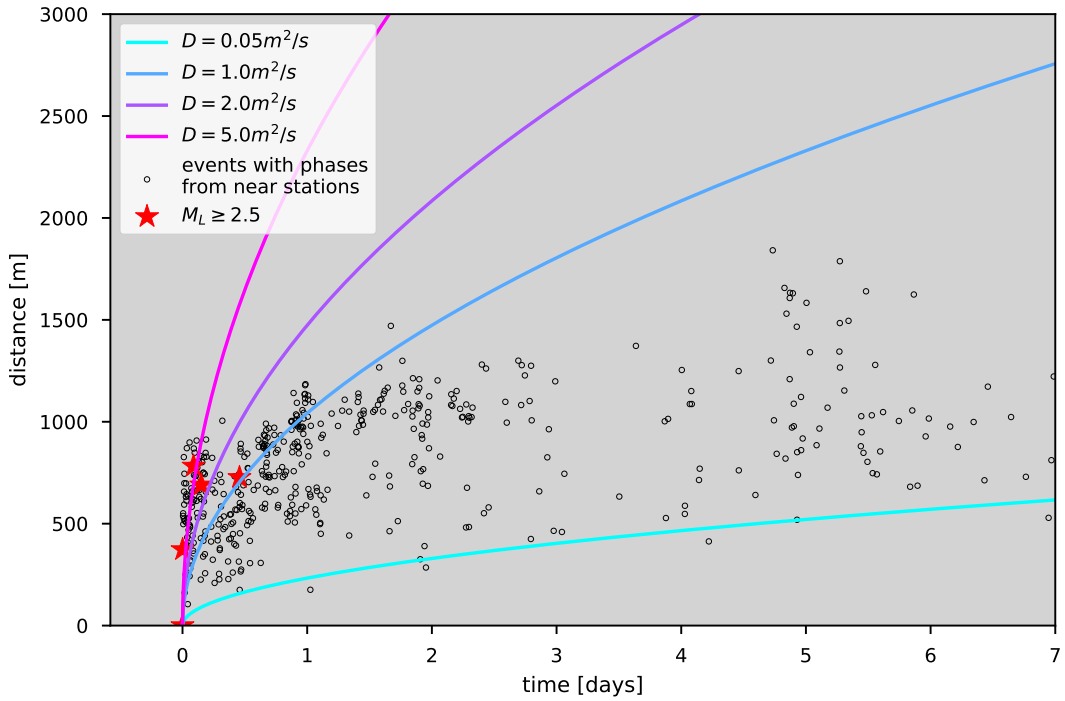

Figure 3.16: $r - t$ plot for the first part of Città di Castello seqeunce (May 2013). Only the events with phases from nearest stations are selected, explanation in the text. The origin of the plot is set on the M_W 3.6 of 2013-05-08 00:52:17.

3.3.3 Lithology and seismicity

We have compared the seismicity distribution retrieved in this study with the actual lithology and geological structures profiting from the availability of a detailed 3D seismostratigraphic model for the TABOO area Latorre et al., (2016). The three-dimensional model has been constructed by combining information from active seismic (derived from the interpretation of 300 km of seismic reflection profiles), data from kilometers deep boreholes, detailed geological surveys and direct measurements of P and S wave velocities performed in situ and in laboratory.

The Pietralunga and CdC sequences are superimposed on the seismostratigraphic units from Latorre et al., (2016) and shown respectively in Figure 3.17 and Figure 3.18. The complete succession of seismostratigraphic units defined in Latorre et al., (2016) include, from bottom to top: (1) the Paleozoic Crystalline Basement; (2) the Acoustic Basement composed of Paleozoic-Triassic clastic and metasedimentary rocks; (3) the Triassic Evaporites; (4) the Jurassic-Oligocene Carbonatic Multilayer; and (5) the Miocene Turbidites.

The most remarkable feature is the constant involvement of the Triassic Evaporites (highlighted in magenta in the figures) in the development of seismic sequences. In the case of Pietralunga most of the events occurred inside the Triassic Evaporites and Carbonatic Multilayer (Figure 3.17). For Pietralunga 2010 we observe that the sequence starts in the evaporitic layer, with the mainshock located at the lower termination of the NE dipping main fault and the antithetic segment. The occurrence of the M_W3.6 earthquake marks the beginning of the migration, with its main directions towards the surface and to NW, as previously described in along-strike plots (Figure 3.10 (c)). The vertical development of the sequence, which spreads from the bottom up from the Triassic Evaporites towards the overlying Carbonatic Multilayer, is in agreement with observations of Marzorati et al., (2014) and Latorre et al., (2016). The upward migration along with the diffusion pattern highlighted in the $r - t$ plot of this portion of the sequence (Figure 3.11) corroborates the fluid-driven hypothesis.

The 2014 seismicity takes place inside the same seismostratigraphic units but, unlike the 2010 part, the migration pattern doesn't seem to have a component

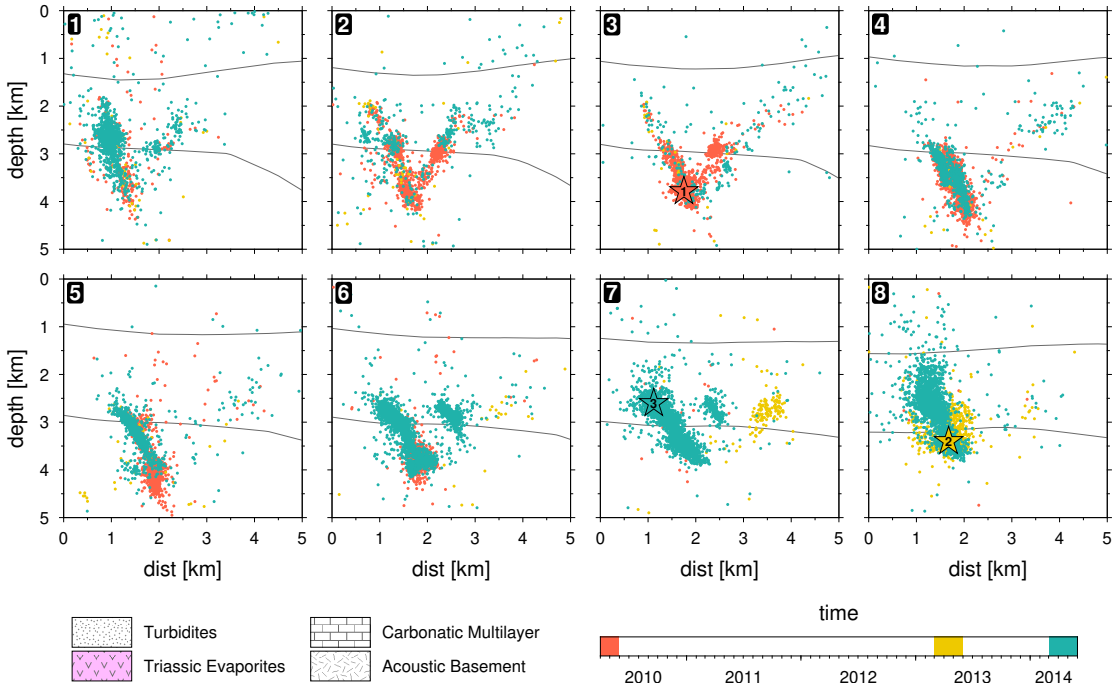

Figure 3.17: Pietralunga sections with seismostratigraphic units from Latorre et al., (2016). Triassic Evaporites highlighted in magenta.

directed towards the surface. The sequence seems dominated by the SE migration observed in along-strike plots (Figure 3.10(c)): starting from the northern segment of the fault activated in the short burst of 2013, seismicity moves towards south-east, ultimately connecting with the 2010 fault segment.

The CdC sequence shows a marked clustering of all the $M_W > 3$ earthquakes both in map (Figure 3.14(a)) and in sections (Figure 3.18, section 2). In Figure 3.18 we show only events with phases from nearest stations, thus mitigating the aforementioned scattering observed in the CdC sequence. Almost all of the $M_W > 3$ earthquakes occur inside the Triassic Evaporites, suggesting a strong lithological control over seismicity. The first burst of seismicity, starting with the M_W3.4 on 2013-04-20, triggers a migration with a primary along-strike component, but also a vertical one, similar to what is observed in 2010 Pietralunga sequence. The occurrence of subvertical clouds of events have been observed in relation with fluids migration (Shapiro et al., 2003; Di Luccio et al., 2010), also in the northern

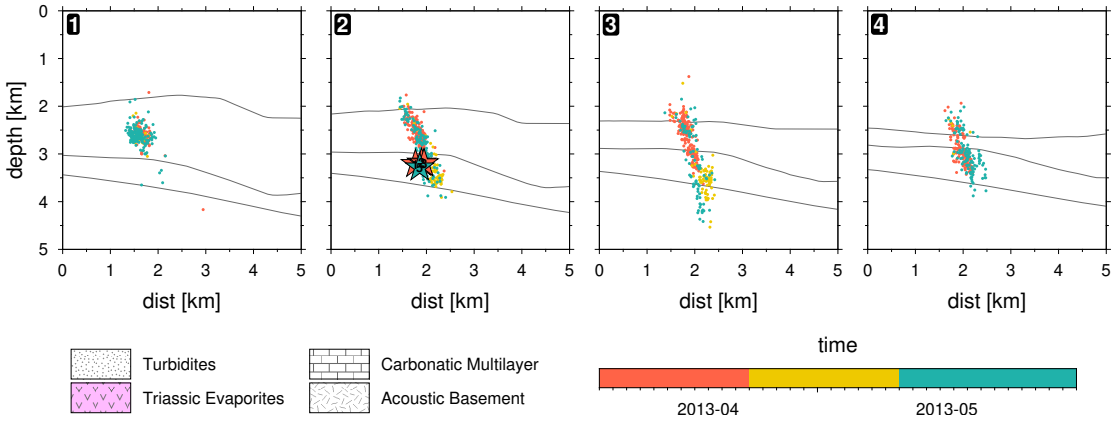

Figure 3.18: CdC sections with seismostratigraphic units from Latorre et al., (2016). Triassic Evaporites highlighted in magenta. Only events with phases from nearest stations are shown.

Apennines (Calderoni et al., 2009) and can be explained by the movements of fluids along with negative pressure gradients, reasonably assumed to be directed towards the surface (Pacchiani and Lyon-Caen, 2010). In this context the lithological units near the Triassic Evaporites are involved as a consequence of the expansion of the diffusion front. Considering the shifting of the origin used in Figure 3.15 to model the diffusive pattern, the vertical migration could actually start near the Evaporites - Carbonates interface instead of starting completely inside the evaporitic layer. This would mean a less upward directed migration and a greater involvement of the Carbonatic layer in the diffusion process. This consideration does not affect the overall diffusive behavior and the evident along-strike migration, observed firstly in Figure 3.14(c).

3.4 Discussion and conclusions

We investigated the potential role of fluids in the development of two small seismic sequences (mainshocks magnitude < 4) occurred in the Northern Apennines: The Città di Castello and the Pietralunga sequence.

The study area has a unique potential to test the hypothesis of fluid involvement

in seismogenesis for multiple reasons. Firstly, a CO_2 source of claimed mantle origin (Chiodini et al., 2004), with well documented degassing phenomena at the surface (Chiaraluce et al., 2014), along with the favorable geologic and tectonic setting for the development of overpressure in crustal traps (Trippetta et al., 2013), as evidenced by measurements in deep (4-5km) boreholes with values comparable to the 85% of the lithostatic load. Furthermore, the presence of a dense monitoring network (Chiaraluce et al., 2014), which allows a very low event detection threshold, unlocking the possibility to analyze local small sequences with a resolution comparable to the one reached for bigger sequences, with the advantage of a much higher frequency of occurrence.

We exploited the potential of the seismic network by running a state-of-the-art deep learning phase picking algorithm (PhaseNet, Zhu and Beroza, (2019)) directly running on the raw waveforms. The retrieved P- and S-waves arrival times are then associated to common events with a recently developed algorithm (REAL, Zhang et al., (2019)). Absolute locations are obtained with a global probabilistic location method (NonLinLoc, Lomax et al., (2000)) and then the seismic events, after undergoing through quality selections criteria, are relocated with the double-difference algorithm (HypoDD, Waldhauser, (2000)).

The final catalog for Pietralunga and CdC sequences consist of roughly 6 times more events than the previously available catalogs for the same areas and time windows (Figure 3.8 and Figure 3.9). The results of our workflow confirms the detection capabilities of recently developed phase picking methods based on artificial intelligence, highlighting the superior performances especially in a microseismicity context.

The dense catalogs are analyzed to test the hypothesis of fluid diffusion processes driving the spatiotemporal migration of seismicity. By modeling the seismic sequences using the solution to the diffusion equation proposed by (Shapiro et al., 1997), we detected patterns compatible to fluid diffusion in both Pietralunga (Figure 3.11, Figure 3.12 or Figure 3.13) and CdC sequences (Figure 3.15, Figure 3.16). The diffusive pattern, sometimes coupled with a backfront signature defining the end of the diffusion effect, is highlighted in space-time diagrams (named $r - t$

plots) together with the characteristic prolonged and along-strike migration of the seismicity.

The estimated values of hydraulic diffusivity are in the order of $1 - 2m^2s^{-1}$, perfectly within ranges commonly observed for induced and natural earthquakes (Talwani, 1984), ranging from 0.1 to $10m^2s^{-1}$. These values are also consistent to the relation proposed by Chiaraluce, (2012) between diffusivity values and earthquake magnitudes in extensional tectonic environments. In this regard our estimated values are representative of M < 4 earthquakes, and rank on the lower end with respect to other diffusivity values estimated for larger earthquakes in the Apennines, with the Umbria-Marche (Antonioli, 2005), L'Aquila (Malagnini et al., 2012) and 2016 Central Italy (Malagnini et al., 2022) sequences showing larger diffusivity values (at least one order of magnitude larger).

CdC and Pietralunga sequences appear to be both driven by pore-pressure diffusion, but in between them they differ in some aspects, testifying the complex and variable behavior observable even at this small scale within a common seismotectonic context. We do not detect foreshocks for the CdC sequence and the early stage of the activation of both April and May segments of this fault system, show a sudden co-seismic migration along strike, occurring immediately after the main rupture episodes (as showed by the clouds of events, vertically aligned in the along-strike plots; Figure 3.14(c)). We interpreted this portion of the sequence as related to triggering effects due to static stress transfer (Stein, 1999). Whilst the Pietralunga sequence shows foreshocks activity and no evident static triggering signatures; more similarly to a swarm-like behavior.

Other interesting features are evident by comparing the seismicity distribution with the three-dimensional distribution of the lithologies at depth. We benefit in fact from the availability of a detailed deterministic seimostratigraphic model of the area (from Latorre et al., (2016)). As shown in Figure 3.17 and Figure 3.18 an involvement of the Triassic Evaporites in the seismic activity is ubiquitous. Pietralunga 2010 sequence, before spreading upwards toward the Carbonatic Multilayer, nucleates within the Triassic Evaporites (Figure 3.17, section 3). The same for the 2014 activity (in agreement with Marzorati et al., (2014), Latorre et al., (2016) and

Valoroso et al., (2017)). While all the CdC mainshocks, despite the slightly lower location resolution, show a consistent pattern and are all clustered inside the Triassic Evaporites.

Thus, we confirm the Triassic Evaporites as a preferential lithology for earthquakes nucleation, indicating a strong mechanical control and corroborating their seismogenic role (Mirabella et al., (2008), Collettini et al., (2009a), and Miller et al., (2004) for the 1997 Umbria-Marche sequence and Marzorati et al., (2014) and Latorre et al., (2016) for the Pietralunga one). This lithology has been suggested to have an important part in the trapping of fluids: composed of a series of anhydritic and dolomitic layers, the Triassic Evaporites may act as barriers to fluid flow due to the exceptionally low permeability of the anhydrite (Trippetta et al., 2010; Trippetta et al., 2013), substantially contributing to the overpressure measured in boreholes. Thus, in the case of a deep origin for the CO_2 (e.g., from mantle degassing, as proposed by (Chiodini et al., 2004)), the flux coming from below may remain trapped in this low permeability layers, locally generating overpressure, as observed in the Santo Stefano and San Donato boreholes, specifically in the dolostone layers that would act as CO_2 reservoirs. Moreover anhydrites can keep a low permeability even during deformation due to their tendency to deform with a ductile behaviour (De Paola et al., 2009). The brittle deformation during the main events of the sequences probably links the different dolostones reservoirs and fluids, triggering the diffusive flow driving the seismic activity.

As proposed by Collettini et al., (2009a), brittle faulting within the Triassic Evaporites could generate the paths for crustal scale fluid flow (Cello et al., 2001). In this context the Pietralunga sequence may be an example of two fluid driven seismic phases migrating along the PATF with opposite directions: towards NW in 2010 and mainly towards SE in 2014. The starting of migration on opposite ends of the fault indicates multiple sources of fluids separated by ~ 5 km. The distribution of events in CdC sequence suggest a single source of fluids as the driving mechanism for the aftershocks of both April and May 2013. In this case the slight increase in hydraulic diffusivity between the first and the second phase could be indicative of the propagation in an already ruptured medium.

Our observations of diffusive patterns, the involvement of the Triassic Evaporites and the nucleation of multiple mainshocks in this lithology, are all supporting the case of fluid driven seismicity with strong lithological control in CdC and Pietralunga sequences. Moreover, our findings confirm that small extensional sequences show similar characteristics with respect to large ones (e.g. Antonioli, (2005); Di Luccio et al., (2010); Malagnini et al., (2022); respectively for Umbria-Marche 1997, L'Aquila 2009 and Central Italy 2016), such as along strike seismicity migration, fluid-driven aftershocks and multi-segments ruptures; underlining a similar behavior on different orders of magnitude.

The analysis of seismic patterns in small sequences represents a fruitful line of research, giving insights to understand preparatory phases and aftershocks evolution with advantages in information availability with respect to major sequences, provided a dense instrumentation and high performance detection techniques.

This highlights the relevance of implementing high resolution networks to observe and model characteristics of small sequences that can be transferred to major ones. The ongoing development of seismic networks and Near Fault Observatories is essential for this kind of applications, in order to be able to resolve fault structural details and seismicity patterns at the hundreds of meters scale.

3.A V_P/V_S analysis on foreshocks

The very high number of events detected with the workflow applied in Chapter 3 allowed to perform a detailed analysis on the foreshock sequence of Pietralunga 2010.

A small cluster (radius=500 m) centered near the hypocenter of 15 April 2010 mainshock and including the foreshocks is used to select the corresponding events and calculate V_P/V_S values, creating time series for each station in Figure 3.19. V_P/V_S time series are analyzed with the rj-McMC algorithm developed in Chapter 2: a piecewise constant modeling is used.

While in ATPI and PIEI stations no variations are detected, ATPC ($\sim$ 750 events detected in April 2010 for this station) shows an increase in V_P/V_S values before the mainshock. Such variation is visible in Figure 3.20. The observed trend is verified also with a simple moving median, computed using different number of samples (see Figure 3.21).

The increase in V_P/V_S preceding the mainshock is similar to what observed by Lucente et al., (2010) for the 2009 L'Aquila sequence, compatible with a dilatancy phenomenon and corroborating the important role played by fluids diffusion.

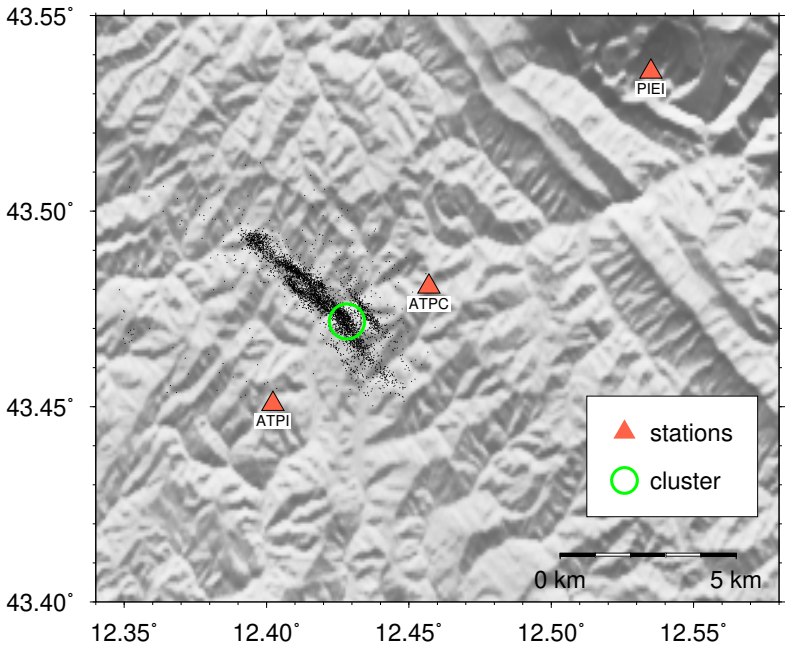

Figure 3.19: Map showing stations and selected cluster for V_P/V_S analysis on foreshocks.

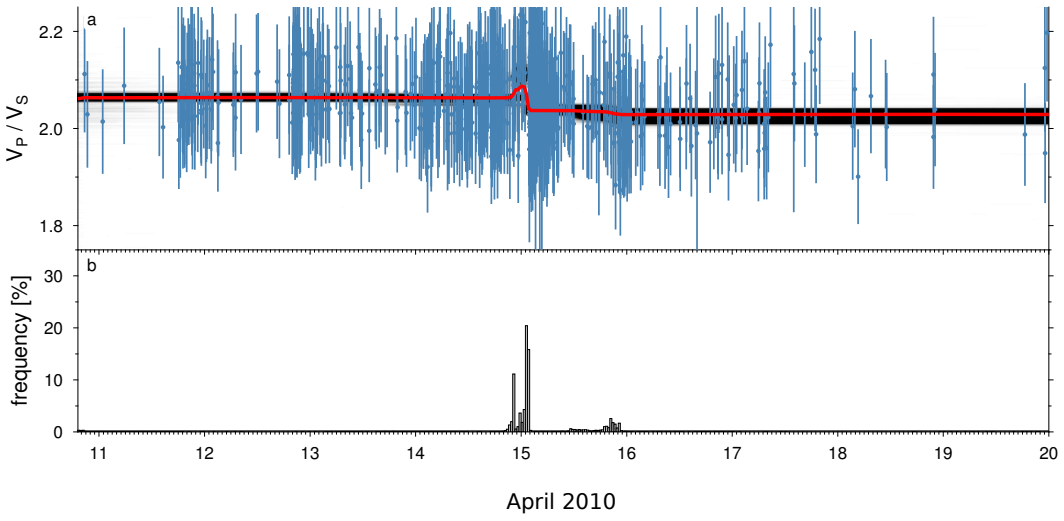

Figure 3.20: ATPC V_P/V_S analysis. Panel (a): data points in blue; red line is mean model from PPD; grayscale background is marginal PPD on V_P/V_S values. Panel (b) CP distribution in time.

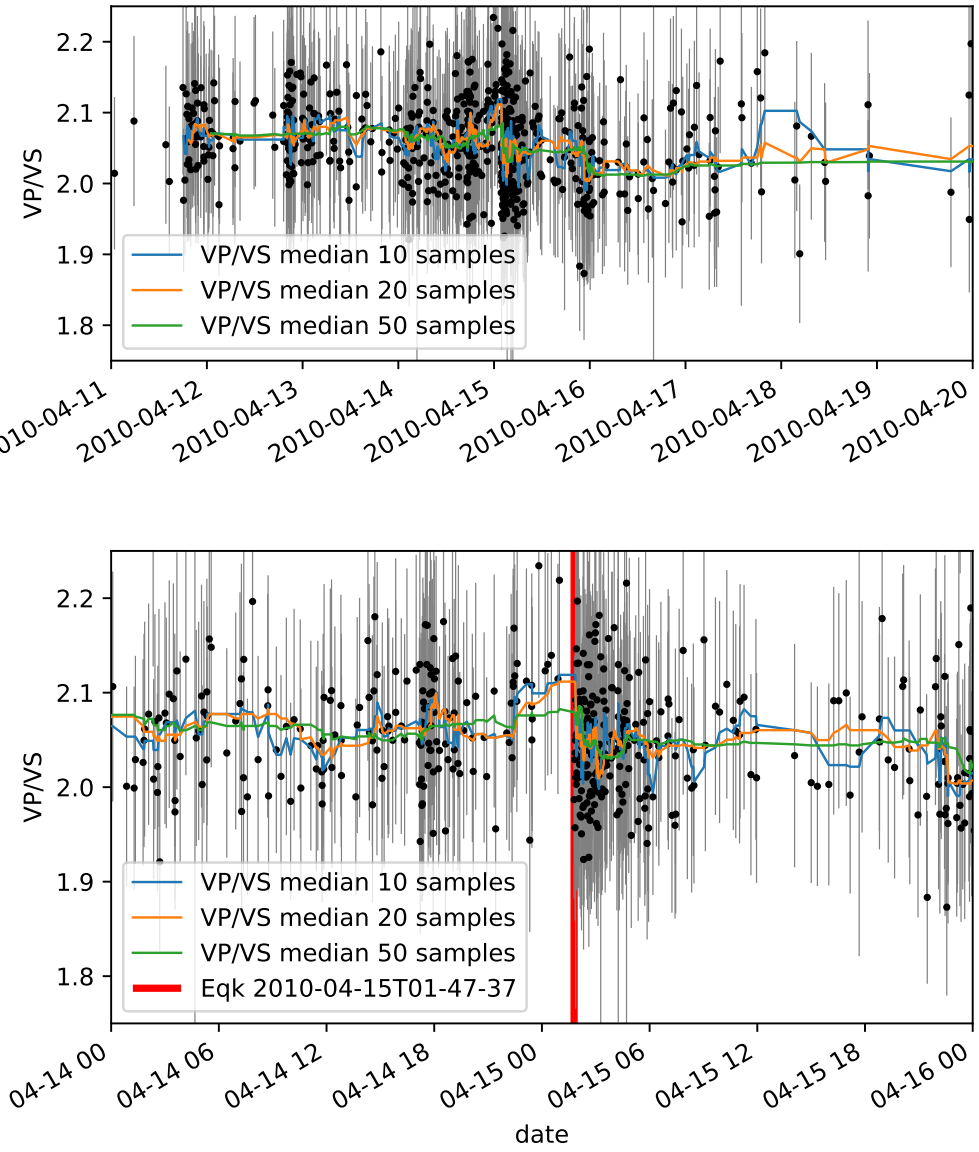

Figure 3.21: Top: same data as in Figure 3.20 modeled with a moving median computed on a different number of samples. Bottom: zoom from 14 April 2010 to 16 April 2010; vertical red line marks the occurrence of the mainshock.

3.B Application of location workflow to an ongoing seismic sequence: Gubbio 2021

On 15 May 2021 at 07:56:01 (UTC time) a M_W 3.9 normal faulting earthquake occurred near Gubbio. The event was well recorded by a large number of INGV stations pertaining to TABOO network, allowing an optimal detection and relocation of the seismic activity. The event gave the onset to a local seismic sequence characterized by the occurrence of 160 detected/relocated events between 15-16 May.

This sequences gave the opportunity to apply the location workflow used in Chapter 3 (without the final relocation) to an ongoing seismic sequence. The relative map and cross sections are shown in Figure 3.22. Grey and black points (e.g., events) represent respectively the background activity from 2010 to 2015 and the Gubbio swarm of 2013-2014 (from Valoroso et al., (2017)). Green points represent the 345 earthquakes occurred between 10 and 14 of May and the red ones are the 326 events of the 15 and 16 of May after the M_W 3.9 whose location is represented by a white star. The recent sequence reactivates a minor antithetic splay of the ATF placed on the Gubbio fault hanging wall (both outlined in cross sections).

Aprat from the larger number of detected events (more than double), the new locations obtained with the workflow based on the deep learning picker (Zhu and Beroza, 2019) includes also some previously undetected foreshocks (see events occurred between 0 and 5 km of distance from Gubbio in space-time diagram of Figure 3.23).

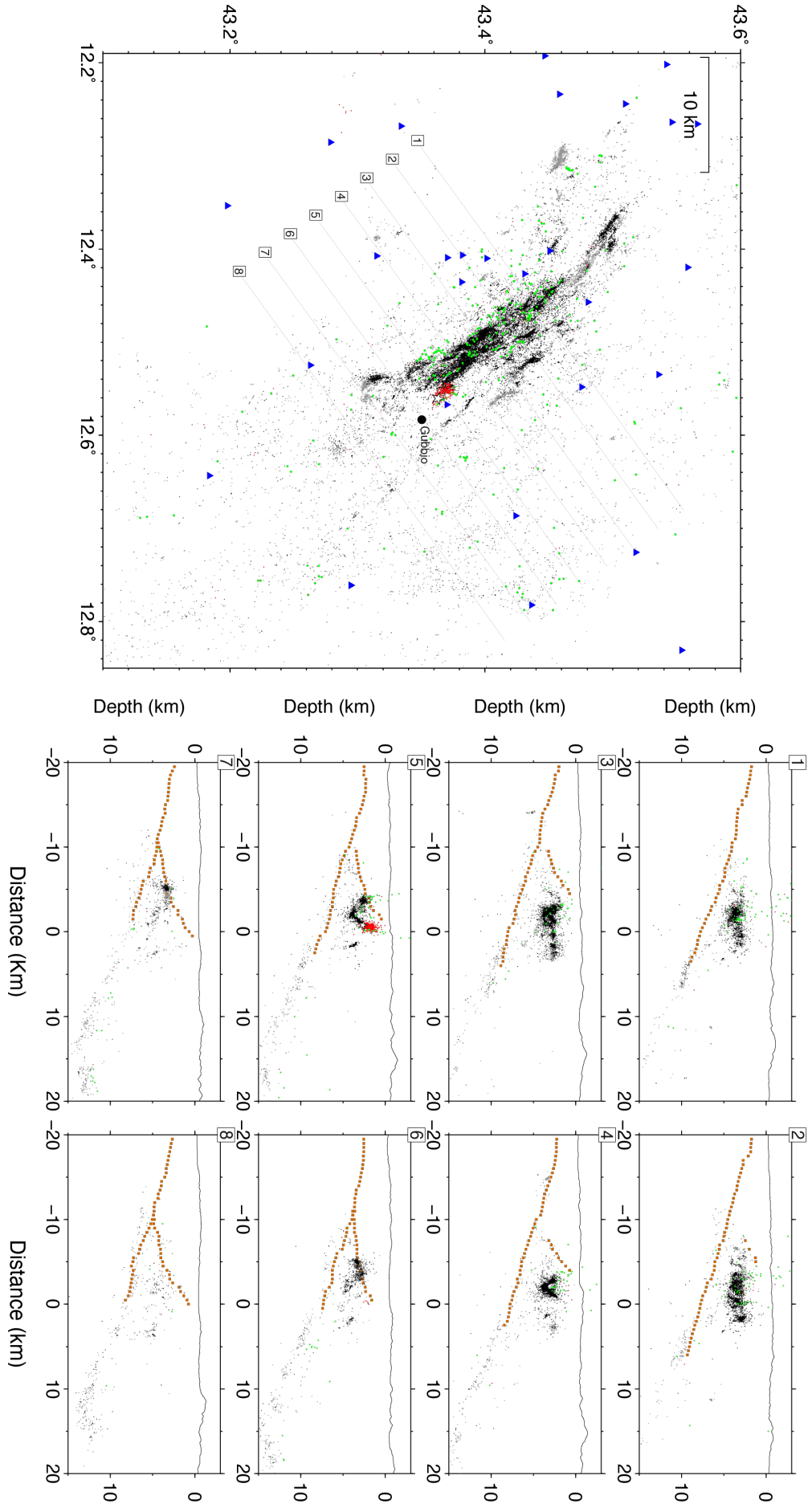

Figure 3.22: Map and sections of events. Grey and black points represent background activity from 2010 to 2015 and Gubbio swarm of 2013-2014. Green points represent the earthquakes occurred between 10 and 14 May 2021 and the red ones are the events of the 15 and 16 May 2021 after the M_W 3.9, whose location is represented by a white star.

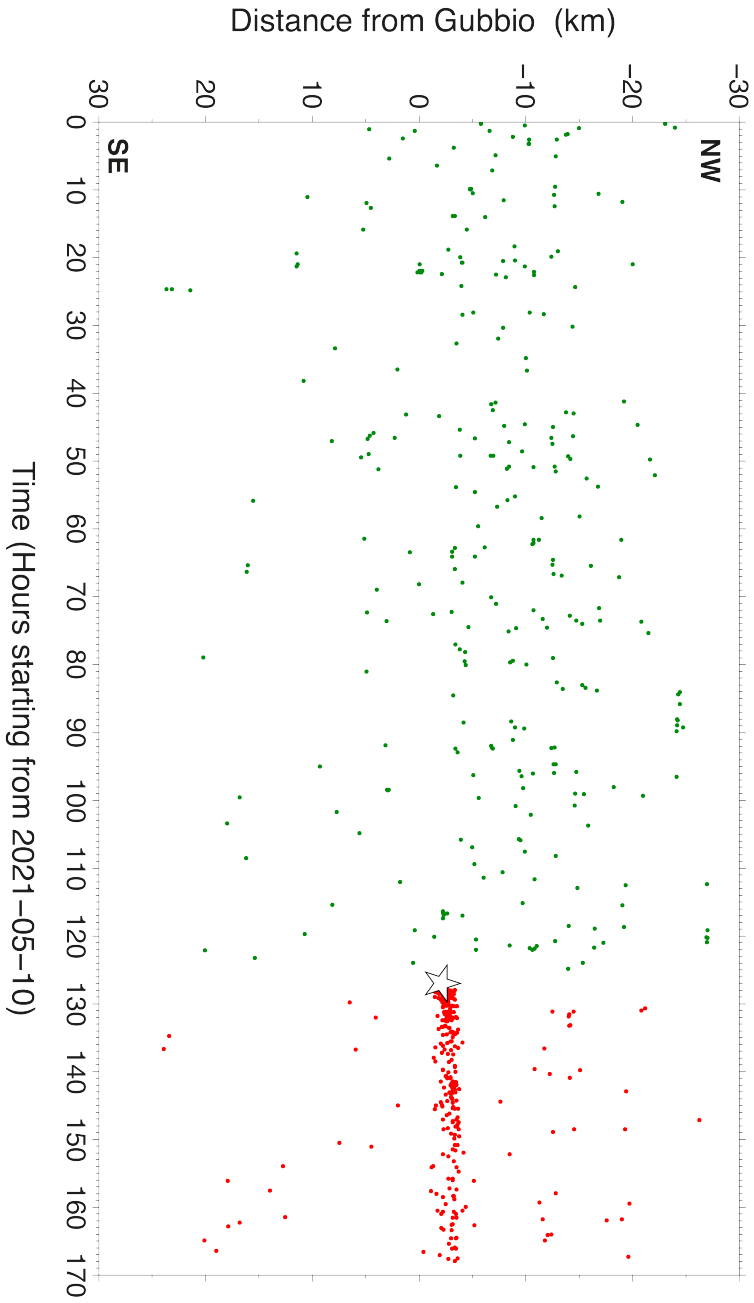

Figure 3.23: Space-time diagram of Gubbio 2021 sequence.

4. Conclusion and outlook

In my research I was prompted by the idea that the contribution from multiple disciplines is essential for the advancement of our knowledge of earthquakes phenomena. I considered questions regarding potentially significant anomalies (e.g., variations) detectable in multiparametric data or the possible role played by crustal fluids in triggering and constraining the seismic activity and its evolution, as some of the key questions in modern earthquakes Science. And I believe that answering these questions requires considering phenomena at the boundaries between different research fields and the availability of long time series of high-resolution data.

In this thesis I faced the analysis of multidisciplinary raw data and derived scientific products potentially related to tectonic processes, with the aim to generate a common environment able to integrate and compare the diverse datasets. One day in the future, the outcome of these studies could provide a contribution to aiding expert opinions pertaining to monitoring agencies. I exploited previous knowledge of the area and together with the usage of state-of-the-art methods to better detect and characterize the seismic activity of the study region, I looked for relationships between complex processes occurring in the crust.

The whole work has been possible due to the recent development of a modern research infrastructure including dense seismological, geodetic and geochemical networks monitoring a small (3000 km2) region of the Northern Apennines characterized by a high rate of micro-seismic release and possessing moderate to large earthquakes potential.

I studied (see Chapter 2) the integration of diverse time series derived from multiparametric sensors and their relation to seismicity activity detected inside the study area. A primary goal was the development of a method that allows the comparison of all the different quantities in a coherent way. In this context the big advantage of the proposed reversible-jump Markov chain Monte Carlo (rj-McMC) approach is the ability to estimate the number of discontinuities, (Change Points; CP), required to model the time series, based only on data and errors. This is a data-driven approach which is particularly valuable for this kind of analysis involving many different time series from multiple stations, where an ad hoc modeling would be not only infeasible, but also lead to a completely subjective result. The output of the application of the algorithm has been analyzed in terms of CP distributions and mean models obtained from the posterior probability distribution (PPD).

Further analysis can include any statistical parameters obtained from the PPD, making the output exploitable in different ways. The multidisciplinary comparison has been made for the whole study area, highlighting no obvious variations occurring in multiple analyzed parameters. Another useful representation of the results could involve a more focused selection of specific stations and seismic events (e.g., pertaining to a specific seismic sequence). Unfortunately, the overlap between the timing of occurrence of the local main seismic sequences and the availability of all datasets was not favorable. Given the results reported in Chapter 3, an interesting proxy for future analysis would be to verify if a variation in the CO_2 flux can be detected at the surface before/during/after the occurrence of fluid driven sequences like the ones I described. This would be a first indication of hydraulic connectivity from (few kilometers of) depth to the free surface.

The availability of long-term time series of multiparametric data is crucial to have the possibility of reliably identifying tectonic signals (being pre- co- or post-seismic). The identification of earthquake related signals is in fact further complicated by the influence of non-tectonic effects (Earth's tides, barometric pressure, temperature, rainfall, and other factors (Hartmann and Levy, 2005)) on many measured parameters. Some of these effects may be corrected by disposing of longer time series, which would allow the identification, for example, of

long-term fluctuations or recurring meteorological effects in addition to the ones considered. This is of paramount importance in the pre-processing phase I adopted (see section 2.2). By considering more of these "exogenous" signals, the resulting CP distribution in geochemical or GNSS time series would be much simpler, thus allowing an easier reading of the data and perhaps a more meaningful comparison.

The complex CP distributions also arise, as explained in Chapter 2, from the lowering of the noise scale parameter. This behavior is not imposed but determined by the data distribution. Nonetheless it results in a high sensitivity of the algorithm, to even small variations. If such variations are deemed as unnecessary or unwanted, the causes should be further investigated to develop strategies that can mitigate the issue acting on error evaluation and/or treatment. The novel approach I proposed has a good potential in multiparametric monitoring of seismically active regions and lays a foundation to a more systematic, data-driven and coherent handling of the ever increasing datasets recorded by modern networks. A fruitful application is nonetheless tied to the level of knowledge that we have on the effects of different sources of anomalies on the measured parameters.

In Chapter 3 the study changes focus from the exploration of tectonic related signals in multidisciplinary datasets to the effects that composite processes happening in the crust may have on the spatiotemporal distribution of the seismicity. I have been working to test the hypothesis of fluid diffusion processes occurring in the study area at the source depth, by identifying compatible signatures in seismicity pattern evolution.

Detailed analysis of such patterns in small seismic sequences (mainshocks magnitude < 4) faces significant observational challenges. Nevertheless, these small earthquake sequences are often the only source of information; large earthquakes are in fact a lot less frequent (even if possibly impending). Thus, the analysis of micro seismicity is the one providing valuable information on preparatory phases, triggering mechanisms and fault mechanics, evidencing similarities or differences at various scales of the problem.

I exploited the detection capability threshold of TABOO's seismic network by means of novel deep learning techniques applied to detection and localization

problems, with the aim to build a detailed picture of the spatiotemporal evolution of seismicity. The goodness of the obtained results in not only expressed by the increased detection capability, but it is also reflected in the enhancement of the quality of the retrieved earthquakes catalog, due to the availability of a significatively larger number of seismic phases (e.g., P- and S-waves arrival times) allowing the adoption of more stringent quality control systems (filtering) on the detected events while still keeping a plentiful catalog. Further improvements in locations robustness would be gained by including an estimate of the time uncertainty in the output of the phase detection algorithms, which is a useful feature, although uncommon for these machine learning techniques (an exception is, for example, the one proposed by Mousavi et al., (2020)). Acknowledging the generalization capabilities of the deep learning algorithms to waveforms from different regions around the world (Mousavi et al., 2020), an interesting test would be the assessment of the eventual level of improvements in precision and recall using transfer learning to update the training of the model with local waveforms. Regarding the precision of the relocation process, especially valuable when dealing with small structures, a substantial increase would be expected performing analysis of waveforms similarities via cross-correlation techniques (Schaff, 2004).

The spatiotemporal distributions of events obtained from the produced catalog strongly suggest the occurrence of diffusive processes during both Città di Castello and Pietralunga seismic sequences. These small sequences show characteristics very similar to the ones characterizing larger extensional sequences (e.g. Colfiorito 1997 L'Aquila 2009, Central Italy 2016), indicating that the fluid-driven mechanism may be widespread in the Apennines, disregarding the magnitude difference. This finding supports the already strong case of deeply trapped CO_2 substantially contributing to seismicity, earthquakes, and aftershocks in the Apennines (Chiarabba et al., (2009), Lucente et al., (2010), Miller et al., (2004), Terakawa et al., (2010), among others).

The results advise also for a strong lithological (e.g., mechanical) control on earthquake nucleation and distribution, with the low permeability Triassic Evaporites seemingly able to promote earthquake nucleation as a response to

development of (CO_2) fluid overpressure. These considerations corroborates the hypothesis advanced in multiple studies performed in the same area (e.g. Trippetta et al., (2013); Mirabella et al., (2008); Marzorati et al., (2014), Latorre et al., (2016); Piana Agostinetti et al., (2017)) and agrees with what observed in laboratory tests involving the same lithology (De Paola et al., (2009); Collettini et al., (2009a), Trippetta et al., (2010)).

Thus, based on my experience, I strongly believe that our improvement of the understanding of natural phenomena (e.g., earthquake) is largely data-driven. The improvement of our observational capabilities allows testing a broader range of models and hypotheses searching for new hints of crustal processes interacting with each other. The integration of high-resolution data coming from different disciplines can help us in obtaining a comprehensive picture to unravel tectonic evolution and fault zone processes. The analysis of small seismicity proves to be a prolific line of research, providing valuable information that can be extended on the behavior of larger earthquakes and narrowing the gap with laboratory experiments, which have made much progress recently (Rouet-Leduc et al., 2017), but with which the integration constitutes a further step to overcome.

In my PhD journey, which unfortunately included very few trips due to the (partially still ongoing) pandemic, I was lucky enough to be able to explore a multitude of topics. I had the chance to explore datasets that I had never seen before, learn about how broad and branched the earthquakes Science is, and deepened many interesting aspects in the study of seismic sequences: from the delicate steps of locating earthquakes to the analysis of seismic catalogs. I also had the opportunity to participate (virtually) in seminars where I was able to share some of my ideas and work, but, more importantly, learn from the others, their research approach and hypothesis. The multidisciplinary approach that I followed has undoubtedly shaped me, making me able to wander on many facets of earthquake Science, from the more seismological aspects to the computational methods and observational skills, without losing my bearing. Due to the social distancing imposed by the pandemic, I managed to appreciate the value of scientific collaboration even more. I enjoyed every occasion in which I had the opportunity

to cooperate with those who share the same topics and issues, receiving (and I hope sometimes even giving), a valuable contribution.

References

Agostinetti, N. P. and A. Malinverno (2010). "Receiver function inversion by trans-dimensional Monte Carlo sampling". *Geophysical Journal International*. DOI: 10.1111/j.1365-246X.2010.04530.x.

Anderlini, L., E. Serpelloni, and M. E. Belardinelli (2016). "Creep and locking of a low-angle normal fault: Insights from the Altotiberina fault in the Northern Apennines (Italy): Creep and Locking of Altotiberina Fault". *Geophysical Research Letters* 43.9, pp. 4321–4329. DOI: 10.1002/2016GL068604.

Anderson, E. M. (1905). "The dynamics of faulting". *Transactions of the Edinburgh Geological Society* 8.3, pp. 387–402. DOI: 10.1144/transed.8.3.387.

Antonioli, A. (2005). "Fluid flow and seismicity pattern: Evidence from the 1997 Umbria-Marche (central Italy) seismic sequence". *Geophysical Research Letters* 32.10, p. L10311. DOI: 10.1029/2004GL022256.

Avouac, J.-P. (2015). "From Geodetic Imaging of Seismic and Aseismic Fault Slip to Dynamic Modeling of the Seismic Cycle". *Annual Review of Earth and Planetary Sciences* 43.1, pp. 233–271. DOI: 10.1146/annurev-earth-060614-105302.

Barchi, M. (1998). "The CROP 03 profile: a synthesis of results on deep structures of the northern Apennines". *Mem. Soc. Geol. It.* 52, pp. 383–400.

Barchi, M. (2010). "The Neogene-Quaternary evolution of the Northern Apennines: crustal structure, style of deformation and seismicity". *Journal of the Virtual Explorer* 36.

Bayes, T. (1763). "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S". *Philosophical Transactions* 53, pp. 370–418. DOI: 10.1098/rstl.1763.0053.

Belardinelli, M. E., A. Bizzarri, and M. Cocco (2003). "Earthquake triggering by static and dynamic stress changes". *Journal of Geophysical Research: Solid Earth* 108.B3. DOI: 10.1029/2002JB001779.

Bodin, T. et al. (2012a). "Transdimensional inversion of receiver functions and surface wave dispersion". *Journal of Geophysical Research* 117.B2. DOI: 10.1029/2011JB008560.

Bodin, T. et al. (2012b). "Transdimensional tomography with unknown data noise: Transdimensional tomography". *Geophysical Journal International* 189.3, pp. 1536–1556. DOI: 10.1111/j.1365-246X.2012.05414.x.

Boncio, P., F. Brozzetti, and G. Lavecchia (2000). "Architecture and seismotectonics of a regional low-angle normal fault zone in central Italy". *Tectonics* 19.6, pp. 1038–1055. DOI: 10.1029/2000TC900023.

Calderoni, G. et al. (2009). "A seismic sequence from Northern Apennines (Italy) provides new insight on the role of fluids in the active tectonics of accretionary wedges". *Earth and Planetary Science Letters* 281.1-2, pp. 99–109. DOI: 10.1016/j.epsl.2009.02.015.

Camarda, M. et al. (2016). "Temporal and spatial correlations between soil CO_2 flux and crustal stress: Soil CO_2 Flux and Crustal Stress". *Journal of Geophysical Research: Solid Earth* 121.10, pp. 7071–7085. DOI: 10.1002/2016JB013297.

Camarda, M. et al. (2019). "The monitoring of natural soil CO2 emissions: Issues and perspectives". *Earth-Science Reviews* 198, p. 102928. DOI: 10.1016/j.earscirev.2019.102928.

Cannelli, V. et al. (2018). "Italian Radon mOnitoring Network (IRON): A permanent network for near real-time monitoring of soil radon emission in Italy". *Annals of Geophysics* 61.Vol 61 (2018), p. 1. DOI: 10.4401/ag-7604.

Carannante, S. et al. (2013). "Deep structure and tectonics of the northern-central Apennines as seen by regional-scale tomography and 3-D located earthquakes:

STRUCTURE AND TECTONICS OF CENTRAL ITALY". *Journal of Geophysical Research: Solid Earth* 118.10, pp. 5391–5403. DOI: 10.1002/jgrb.50371.

Cello, G et al. (2001). "Fault properties and fuid fow patterns from Quaternary faults in the Apennines, Italy", p. 16.

Chanard, K. et al. (2020). "A warning against over-interpretation of seasonal signals measured by the Global Navigation Satellite System". *Nature Communications* 11.1, p. 1375. DOI: 10.1038/s41467-020-15100-7.

Chiarabba, C. et al. (2018). "Faults Geometry and the Role of Fluids in the 2016–2017 Central Italy Seismic Sequence". *Geophysical Research Letters* 45.14, pp. 6963–6971. DOI: 10.1029/2018GL077485.

Chiarabba, C., P. D. Gori, and E. Boschi (2009). "Pore-pressure migration along a normal-fault system resolved by time-repeated seismic tomography". *Geology* 37.1, pp. 67–70. DOI: 10.1130/G25220A.1.

Chiaraluce, L. (2012). "Unravelling the complexity of Apenninic extensional fault systems: A review of the 2009 L'Aquila earthquake (Central Apennines, Italy)". *Journal of Structural Geology* 42, pp. 2–18. DOI: 10.1016/j.jsg.2012.06.007.

Chiaraluce, L. et al. (2007). "Architecture and mechanics of an active low-angle normal fault: Alto Tiberina Fault, northern Apennines, Italy". *Journal of Geophysical Research* 112.B10. DOI: 10.1029/2007JB005015.

Chiaraluce, L. et al. (2014). "The Alto Tiberina Near Fault Observatory (northern Apennines, Italy)". *Annals of Geophysics* 57.3.

Chiodini, G. et al. (2004). "Carbon dioxide Earth degassing and seismogenesis in central and southern Italy: CARBON DIOXIDE EARTH DEGASSING AND SEISMOGENESIS". *Geophysical Research Letters* 31.7, n/a–n/a. DOI: 10.1029/2004GL019480.

Cicerone, R. D., J. E. Ebel, and J. Britton (2009). "A systematic compilation of earthquake precursors". *Tectonophysics* 476.3-4, pp. 371–396. DOI: 10.1016/j.tecto.2009.06.008.

Collettini, C. and M. R. Barchi (2004). "A comparison of structural data and seismic images for low-angle normal faults in the Northern Apennines (Central

Italy): constraints on activity". *Geological Society, London, Special Publications* 224.1, pp. 95–112. DOI: 10.1144/GSL.SP.2004.224.01.07.

Collettini, C., N. De Paola, and D. Faulkner (2009a). "Insights on the geometry and mechanics of the Umbria–Marche earthquakes (Central Italy) from the integration of field and laboratory data". *Tectonophysics* 476.1-2, pp. 99–109. DOI: 10.1016/j.tecto.2008.08.013.

Collettini, C. et al. (2009b). "Development of interconnected talc networks and weakening of continental low-angle normal faults". *Geology* 37.6, pp. 567–570. DOI: 10.1130/G25645A.1.

D'Agostino, N. et al. (2009). "Contemporary crustal extension in the Umbria–Marche Apennines from regional CGPS networks and comparison between geodetic and seismic deformation". *Tectonophysics*. Ten years after the Umbria-Marche earthquake, Central Italy 476.1, pp. 3–12. DOI: 10.1016/j.tecto.2008.09.033.

De Barros, L. et al. (2019). "Fluid-Induced Swarms and Coseismic Stress Transfer: A Dual Process Highlighted in the Aftershock Sequence of the 7 April 2014 Earthquake (Ml 4.8, Ubaye, France)". *Journal of Geophysical Research: Solid Earth* 124.4, pp. 3918–3932. DOI: 10.1029/2018JB017226.

De Barros, L. et al. (2020). "Imbricated Aseismic Slip and Fluid Diffusion Drive a Seismic Swarm in the Corinth Gulf, Greece". *Geophysical Research Letters* 47.9. DOI: 10.1029/2020GL087142.

De Barros, L. et al. (2021). "Migration of fluid-induced seismicity reveals the seismogenic state of faults". *Journal of Geophysical Research: Solid Earth*. DOI: 10.1029/2021JB022767.

De Paola, N., D. R. Faulkner, and C. Collettini (2009). "Brittle versus ductile deformation as the main control on the transport properties of low-porosity anhydrite rocks". *Journal of Geophysical Research: Solid Earth* 114.B6. DOI: 10.1029/2008JB005967.

Dettmer, J., S. E. Dosso, and C. W. Holland (2010). "Trans-dimensional geoacoustic inversion". *The Journal of the Acoustical Society of America* 128.6, pp. 3393–3405. DOI: 10.1121/1.3500674.

Di Luccio, F. et al. (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.B6. DOI: 10.1029/2009JB007190.

Dieterich, J. H. (1978). "Time-Dependent Friction and the Mechanics of Stick-Slip". In: *Rock Friction and Earthquake Prediction*. Ed. by J. D. Byerlee and M. Wyss. Basel: Birkhäuser, pp. 790–806. DOI: 10.1007/978-3-0348-7182-2_15.

Duverger, C. et al. (2015). "The 2003–2004 seismic swarm in the western Corinth rift: Evidence for a multiscale pore pressure diffusion process along a permeable fault system". *Geophysical Research Letters* 42.18, pp. 7374–7382. DOI: 10.1002/2015GL065298.

Ellsworth, W. L. and G. C. Beroza (1995). "Seismic Evidence for an Earthquake Nucleation Phase". *Science* 268.5212, pp. 851–855. DOI: 10.1126/science.268.5212.851.

Elter, P. (1975). "Tensional and compressional areas in the recent Tortonian to present eolution of the Apennines."

Gallagher, K. et al. (2011). "Inference of abrupt changes in noisy geochemical records using transdimensional changepoint models". *Earth and Planetary Science Letters* 311.1-2, pp. 182–194. DOI: 10.1016/j.epsl.2011.09.015.

Gelman, A. and D. B. Rubin (1996). "Markov chain Monte Carlo methods in biostatistics". *Statistical methods in medical research* 5.4, pp. 339–355.

Gomberg, J. et al. (2001). "Earthquake triggering by seismic waves following the Landers and Hector Mine earthquakes". *Nature* 411.6836, pp. 462–466. DOI: 10.1038/35078053.

Granieri, D et al. (2003). "Continuous monitoring of CO 2 soil diffuse degassing at Phlegraean Fields (Italy): influence of environmental and volcanic parameters". *Earth and Planetary Science Letters* 212.1-2, pp. 167–179. DOI: 10.1016/S0012-821X(03)00232-2.

Green, P. J. (1995). "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination". *Biometrika* 82.4, pp. 711–732.

Gualandi, A. et al. (2017). "Aseismic deformation associated with an earthquake swarm in the northern Apennines (Italy): ASEISMIC SWARM RELATED

DEFORMATION". *Geophysical Research Letters* 44.15, pp. 7706–7714. DOI: 10.1002/2017GL073687.

Gutenberg, B. and C. Richter (1941). *Seismicity of the Earth*. Geological Society of America.

Hainzl, S. (2004). "Seismicity patterns of earthquake swarms due to fluid intrusion and stress triggering". *Geophysical Journal International* 159.3, pp. 1090–1096. DOI: 10.1111/j.1365-246X.2004.02463.x.

Hardebeck, J. L. and E. Hauksson (2001). "Crustal stress field in southern California and its implications for fault mechanics". *Journal of Geophysical Research: Solid Earth* 106.B10, pp. 21859–21882. DOI: 10.1029/2001JB000292.

Hartmann, J. and J. K. Levy (2005). "Hydrogeological and Gasgeochemical Earthquake Precursors ? A Review for Application". *Natural Hazards* 34.3, pp. 279–304. DOI: 10.1007/s11069-004-2072-2.

Hawkins, R. and M. Sambridge (2015). "Geophysical imaging using trans-dimensional trees". *Geophysical Journal International* 203.2, pp. 972–1000. DOI: 10.1093/gji/ggv326.

Holub, R. F. and B. T. Brady (1981). "The effect of stress on radon emanation from rock". *Journal of Geophysical Research: Solid Earth* 86.B3, pp. 1776–1784. DOI: 10.1029/JB086iB03p01776.

Hopcroft, P. O., K. Gallagher, and C. C. Pain (2007). "Inference of past climate from borehole temperature data using Bayesian Reversible Jump Markov chain Monte Carlo: Bayesian inversion of borehole data for past climate". *Geophysical Journal International* 171.3, pp. 1430–1439. DOI: 10.1111/j.1365-246X.2007.03596.x.

Hopcroft, P. O., K. Gallagher, and C. C. Pain (2009). "A Bayesian partition modelling approach to resolve spatial variability in climate records from borehole temperature inversion". *Geophysical Journal International* 178.2, pp. 651–666. DOI: 10.1111/j.1365-246X.2009.04192.x.

Hreinsdottir, S. and R. A. Bennett (2009). "Active aseismic creep on the Alto Tiberina low-angle normal fault, Italy". *Geology* 37.8, pp. 683–686. DOI: 10.1130/G30194A.1.

Hubbert, M. K. and W. W. Rubey (1959). "Role of fluid pressure in mechanics of overthrust faulting: I. Mechanics of fluid-filled porous solids and its application to overthrust faulting". *GSA Bulletin* 70.2, pp. 115–166. DOI: 10.1130/0016-7606(1959)70[115:ROFPIM]2.0.CO;2.

Kagan, Y. Y. and L. Knopoff (1981). "Stochastic synthesis of earthquake catalogs". *Journal of Geophysical Research: Solid Earth* 86.B4, pp. 2853–2862. DOI: 10.1029/JB086iB04p02853.

Kanamori, H. (1977). "Seismic and Aseismic Slip Along Subduction Zones and their Tectonic Implications". In: *Island Arcs, Deep Sea Trenches and Back-Arc Basins*. American Geophysical Union (AGU), pp. 163–174. DOI: 10.1029/ME001p0163.

Kanamori, H. and E. E. Brodsky (2004). "The physics of earthquakes". *Reports on Progress in Physics* 67.8, pp. 1429–1496. DOI: 10.1088/0034-4885/67/8/R03.

Laarhoven, P. J. M. v. and E. H. L. Aarts (1987). "Simulated annealing". In: *Simulated Annealing: Theory and Applications*. Springer Netherlands, pp. 7–15.

Latorre, D. et al. (2016). "Assessment of earthquake locations in 3-D deterministic velocity models: A case study from the Altotiberina Near Fault Observatory (Italy): Event Locations in Deterministic Models". *Journal of Geophysical Research: Solid Earth* 121.11, pp. 8113–8135. DOI: 10.1002/2016JB013170.

Li, Y.-G. et al. (1998). "Evidence of Shallow Fault Zone Strengthening After the 1992 M7.5 Landers, California, Earthquake". *Science* 279.5348, pp. 217–219. DOI: 10.1126/science.279.5348.217.

Liu, M. et al. (2020). "Rapid Characterization of the July 2019 Ridgecrest, California, Earthquake Sequence From Raw Seismic Data Using Machine-Learning Phase Picker". *Geophysical Research Letters* 47.4, e2019GL086189. DOI: https://doi.org/10.1029/2019GL086189.

Lomax, A. et al. (2000). "Probabilistic Earthquake Location in 3D and Layered Models". In: *Advances in Seismic Event Location*. Ed. by C. H. Thurber and N. Rabinowitz. Modern Approaches in Geophysics. Dordrecht: Springer Netherlands, pp. 101–134. DOI: 10.1007/978-94-015-9536-0_5.

Lomax, A., A. Michelini, and A. Curtis (2014). "Earthquake Location, Direct, Global-Search Methods". In: *Encyclopedia of Complexity and Systems Science*.

Ed. by R. A. Meyers. New York, NY: Springer New York, pp. 1–33. DOI: 10.1007/978-3-642-27737-5_150-2.

Lombardi, A. M., M. Cocco, and W. Marzocchi (2010). "On the Increase of Background Seismicity Rate during the 1997-1998 Umbria-Marche, Central Italy, Sequence: Apparent Variation or Fluid-Driven Triggering?" *Bulletin of the Seismological Society of America* 100.3, pp. 1138–1152. DOI: 10.1785/0120090077.

Lucente, F. P. et al. (2010). "Temporal variation of seismic velocity and anisotropy before the 2009 M $_{W}$ 6.3 L'Aquila earthquake, Italy". *Geology* 38.11, pp. 1015–1018. DOI: 10.1130/G31463.1.

Malagnini, L. et al. (2012). "Control of pore fluid pressure diffusion on fault failure mode: Insights from the 2009 L'Aquila seismic sequence: DIFFUSION OF FLUIDS AND FAULT FAILURE". *Journal of Geophysical Research: Solid Earth* 117.B5, n/a–n/a. DOI: 10.1029/2011JB008911.

Malagnini, L. et al. (2022). *Crustal Permeability Changes Observed From Seismic Attenuation: Impacts on Multi-Mainshock Sequences*. preprint. Geophysics. DOI: 10.1002/essoar.10510049.1.

Malinverno, A. (2002). "Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem". *Geophysical Journal International* 151.3, pp. 675–688.

Malinverno, A. and V. A. Briggs (2004). "Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes". *GEOPHYSICS* 69.4, pp. 1005–1016. DOI: 10.1190/1.1778243.

Mandler, E. et al. (2021). "Post-Seismic Deformation Related to the 2016 Central Italy Seismic Sequence From GPS Displacement Time-Series". *Journal of Geophysical Research: Solid Earth* 126.9, e2021JB022200. DOI: 10.1029/2021JB022200.

Marzorati, S. et al. (2014). "Very detailed seismic pattern and migration inferred from the April 2010 Pietralunga (northern Italian Apennines) micro-earthquake sequence". *Tectonophysics* 610, pp. 91–109. DOI: 10.1016/j.tecto.2013.10.014.

Michele, M., D. Latorre, and A. Emolo (2019). "An Empirical Formula to Classify the Quality of Earthquake Locations". *Bulletin of the Seismological Society of America* 109.6, pp. 2755–2761. DOI: 10.1785/0120190144.

Miller, S. A. (2013). "The Role of Fluids in Tectonic and Earthquake Processes". In: *Advances in Geophysics*. Vol. 54. Elsevier, pp. 1–46. DOI: 10.1016/B978-0-12-380940-7.00001-9.

Miller, S. A. et al. (2004). "Aftershocks driven by a high-pressure CO2 source at depth". *Nature* 427.6976, pp. 724–727. DOI: 10.1038/nature02251.

Mirabella, F. et al. (2008). "Insights on the seismogenic layer thickness from the upper crust structure of the Umbria-Marche Apennines (central Italy): SEISMOGENIC LAYER AND CRUST STRUCTURE". *Tectonics* 27.1, n/a–n/a. DOI: 10.1029/2007TC002134.

Mirabella, F. et al. (2011). "Tectonic evolution of a low-angle extensional fault system from restored cross-sections in the Northern Apennines (Italy): RESTORATION OF SECTIONS ACROSS A LANF". *Tectonics* 30.6, n/a–n/a. DOI: 10.1029/2011TC002890.

Mogi, K. (1963). "Some discussions on aftershocks, foreshocks and earthquake swarms-the fracture of a semi-infinite body caused by an inner stress origin and its relation the earthquake phenomena". *Bull. Earthq. Res. Inst.* 41, pp. 615–658.

Mollo, S. et al. (2011). "Increase in radon emission due to rock failure: An experimental study: RADON EMISSIONS FROM ROCKS". *Geophysical Research Letters* 38.14, n/a–n/a. DOI: 10.1029/2011GL047962.

Mongelli, F. and G. Zito (1991). "Flusso di calore nella regione Toscana".

Moretti, M., P. De Gori, and C. Chiarabba (2009). "Earthquake relocation and three-dimensional V_p and V_p / V_s models along the low angle Alto Tiberina Fault (Central Italy): evidence for fluid overpressure". *Geophysical Journal International* 176.3, pp. 833–846. DOI: 10.1111/j.1365-246X.2008.03984.x.

Mosegaard, K. and A. Tarantola (1995). "Monte Carlo sampling of solutions to inverse problems". *Journal of Geophysical Research: Solid Earth* 100.B7, pp. 12431–12447. DOI: 10.1029/94JB03097.

Mousavi, S. M. et al. (2020). "Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking". *Nature Communications* 11.1, p. 3952. DOI: 10.1038/s41467-020-17591-w.

Noir, J. et al. (1997). "Fluid flow triggered migration of events in the 1989 Dobi Earthquake sequence of central Afar". *Geophysical Research Letters* 24.18, pp. 2335–2338. DOI: 10.1029/97GL02182.

Nur, A. and J. R. Booker (1972). "Aftershocks Caused by Pore Fluid Flow?" *Science*. DOI: 10.1126/science.175.4024.885.

Omori, F. (1895). "On the After-shocks of Earthquakes". PhD thesis. College of Science, Imperial University.

Pacchiani, F. and H. Lyon-Caen (2010). "Geometry and spatio-temporal evolution of the 2001 Agios Ioanis earthquake swarm (Corinth Rift, Greece)". *Geophysical Journal International* 180.1, pp. 59–72. DOI: 10.1111/j.1365-246X.2009.04409.x.

Paige, C. C. (1982). "LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares", p. 29.

Park, Y. et al. (2020). "Machine-Learning-Based Analysis of the Guy-Greenbrier, Arkansas Earthquakes: A Tale of Two Sequences". *Geophysical Research Letters* 47.6, e2020GL087032. DOI: https://doi.org/10.1029/2020GL087032.

Parotidis, M., S. A. Shapiro, and E. Rothert (2004). "Back front of seismicity induced after termination of borehole fluid injection". *Geophysical Research Letters* 31.2. DOI: 10.1029/2003GL018987.

Pauselli, C. et al. (2006). "The crustal structure of the northern apennines (Central Italy): An insight by the crop03 seismic line". *American Journal of Science* 306.6, pp. 428–450. DOI: 10.2475/06.2006.02.

Piana Agostinetti, N. and A. Malinverno (2018). "Assessing uncertainties in high-resolution, multifrequency receiver-function inversion: A comparison with borehole data". *GEOPHYSICS* 83.3, KS11–KS22. DOI: 10.1190/geo2017-0350.1.

Piana Agostinetti, N., G. Giacomuzzi, and A. Malinverno (2015). "Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sam-

pling". *Geophysical Journal International* 201.3, pp. 1598–1617. DOI: 10.1093/gji/ggv084.

Piana Agostinetti, N., G. Giacomuzzi, and C. Chiarabba (2017). "Seismic swarms and diffuse fracturing within Triassic evaporites fed by deep degassing along the low-angle Alto Tiberina normal fault (central Apennines, Italy)". *Journal of Geophysical Research: Solid Earth* 122.1, pp. 308–331. DOI: 10.1002/2016JB013295.

Piersanti, A. (2015). "Long term continuous radon monitoring in a seismically active area". *Annals of Geophysics* 58, p. 16.

Piersanti, A., V. Cannelli, and G. Galli (2016). "The Pollino 2012 seismic sequence: clues from continuous radon monitoring". *Solid Earth* 7.5, pp. 1303–1316. DOI: 10.5194/se-7-1303-2016.

Pinault, J.-L. and J.-C. Baubron (1996). "Signal processing of soil gas radon, atmospheric pressure, moisture, and soil temperature data: A new approach for radon concentration modeling". *Journal of Geophysical Research: Solid Earth* 101.B2, pp. 3157–3171. DOI: 10.1029/95JB03121.

Poggiali, G. et al. (2019). "Change-point analysis of *VP* / *VS* ratio time-series using a trans-dimensional McMC algorithm: applied to the Alto Tiberina Near Fault Observatory seismic network (Northern Apennines, Italy)". *Geophysical Journal International* 217.2, pp. 1217–1231. DOI: 10.1093/gji/ggz078.

Ponziani, F. et al. (1995). "Crustal shortening and duplication of the Moho in the Northern Apennines: a view from seismic refraction data". *Tectonophysics* 252.1, pp. 391–418. DOI: 10.1016/0040-1951(95)00093-3.

Rinaldi, A. P. et al. (2012). "Effects of atmospheric conditions on surface diffuse degassing: ATMOSPHERIC CONDITIONS AND DIFFUSE DEGASSING". *Journal of Geophysical Research: Solid Earth* 117.B11, n/a–n/a. DOI: 10.1029/2012JB009490.

Ripepe, M. (2000). "Foreshock sequence of September 26th, 1997 Umbria-Marche earthquakes". *Journal of Seismology* 4.4, pp. 387–399. DOI: 10.1023/A:1026508425230.

Ross, Z. E. and E. S. Cochran (2021). "Evidence for Latent Crustal Fluid Injection Transients in Southern California From Long-Duration Earthquake Swarms". *Geophysical Research Letters* 48.12. DOI: 10.1029/2021GL092465.

Ross, Z. E. et al. (2018). "Generalized Seismic Phase Detection with Deep Learning-Short Note". *Bulletin of the Seismological Society of America* 108.5A, pp. 2894–2901. DOI: 10.1785/0120180080.

Rouet-Leduc, B. et al. (2017). "Machine Learning Predicts Laboratory Earthquakes". *Geophysical Research Letters* 44, pp. 9276–9282. DOI: 10.1002/2017GL074677.

Rovida, A. et al. (2011). "Italian Parametric Earthquake Catalogue CPTI15, version 3.0". 3, p. 38.

Roy, C. and B. A. Romanowicz (2017). "On the Implications of A Priori Constraints in Transdimensional Bayesian Inversion for Continental Lithospheric Layering: CONSTRAINTS IN BAYESIAN INVERSION". *Journal of Geophysical Research: Solid Earth* 122.12, pp. 10,118–10,131. DOI: 10.1002/2017JB014968.

Ruhl, C. J. et al. (2016). "Complex spatiotemporal evolution of the 2008 Mw 4.9 Mogul earthquake swarm (Reno, Nevada): Interplay of fluid and faulting". *Journal of Geophysical Research: Solid Earth* 121.11, pp. 8196–8216. DOI: 10.1002/2016JB013399.

Saccorotti, G., G. Ventura, and G. Vilardo (2002). "Seismic swarms related to diffusive processes: The case of Somma-Vesuvius volcano, Italy". *GEOPHYSICS* 67.1, pp. 199–203. DOI: 10.1190/1.1451551.

Sambridge, M. and K. Mosegaard (2002). "MONTE CARLO METHODS IN GEOPHYSICAL INVERSE PROBLEMS: MONTE CARLO INVERSION". *Reviews of Geophysics* 40.3, pp. 3–1–3–29. DOI: 10.1029/2000RG000089.

Scales, J. A. and R. Snieder (1997). "To Bayes or not to Bayes?" *GEOPHYSICS* 62.4, pp. 1045–1046. DOI: 10.1190/1.6241045.1.

Schaff, D. P. (2004). "Optimizing Correlation Techniques for Improved Earthquake Location". *Bulletin of the Seismological Society of America* 94.2, pp. 705–721. DOI: 10.1785/0120020238.

Schaff David P. and Beroza Gregory C. (2004). "Coseismic and postseismic velocity changes measured by repeating earthquakes". *Journal of Geophysical Research: Solid Earth* 109.B10. DOI: 10.1029/2004JB003011.

Scholz, H. C. (1973). "Earthquake prediction : A physical basis". *Science* 181, pp. 803–809.

Segall, P. and S. Lu (2015). "Injection-induced seismicity: Poroelastic and earthquake nucleation effects: INJECTION INDUCED SEISMICITY". *Journal of Geophysical Research: Solid Earth* 120.7, pp. 5082–5103. DOI: 10.1002/2015JB012060.

Serpelloni, E. et al. (2005). "Crustal velocity and strain-rate fields in Italy and surrounding regions: new results from the analysis of permanent and non-permanent GPS networks". *Geophysical Journal International* 161.3, pp. 861–880. DOI: 10.1111/j.1365-246X.2005.02618.x.

Serpelloni, E. et al. (2018). "Hydrologically Induced Karst Deformation: Insights From GPS Measurements in the Adria-Eurasia Plate Boundary Zone". *Journal of Geophysical Research: Solid Earth* 123.5, pp. 4413–4430. DOI: 10.1002/2017JB015252.

Serpelloni, E. et al. (2006). "Data analysis of permanent GPS networks in Italy and surrounding regions: application of a distributed processing approach". *Annals of Geophysics* 49, p. 32.

Shapiro, S. A. et al. (2003). "Triggering of Seismicity by Pore-pressure Perturbations: Permeability-related Signatures of the Phenomenon". In: *Thermo-Hydro-Mechanical Coupling in Fractured Rock*. Ed. by H.-J. Kümpel. Pageoph Topical Volumes. Basel: Birkhäuser, pp. 1051–1066. DOI: 10.1007/978-3-0348-8083-1_16.

Shapiro, S. A., P. Audigane, and J.-J. Royer (1999). "Large-scale *in situ* permeability tensor of rocks from induced microseismicity". *Geophysical Journal International* 137.1, pp. 207–213. DOI: 10.1046/j.1365-246x.1999.00781.x.

Shapiro, S. A., E. Huenges, and G. Borm (1997). "Estimating the crust permeability from fluid-injection-induced seismic emission at the KTB site". *Geophysical*

Journal International 131.2, F15–F18. DOI: 10.1111/j.1365-246X.1997.tb01215.x.

Sibson, R. H. (1973). "Interactions between Temperature and Pore-Fluid Pressure during Earthquake Faulting and a Mechanism for Partial or Total Stress Relief". *Nature Physical Science* 243.126, pp. 66–68. DOI: 10.1038/physci243066a0.

Sibson, R. H. (1994). "Crustal stress, faulting and fluid flow". *Geological Society, London, Special Publications* 78.1, pp. 69–84. DOI: 10.1144/GSL.SP.1994.078.01.07.

Sibson, R. H. (2000). "Fluid involvement in normal faulting". *Journal of Geodynamics* 29.3-5, pp. 469–499. DOI: 10.1016/S0264-3707(99)00042-3.

Siino, M. et al. (2019). "Multiple seasonality in soil radon time series". *Scientific Reports* 9.1, p. 8610. DOI: 10.1038/s41598-019-44875-z.

Silverii, F. et al. (2016). "Transient deformation of karst aquifers due to seasonal and multiyear groundwater variations observed by GPS in southern Apennines (Italy)". *Journal of Geophysical Research: Solid Earth* 121.11, pp. 8315–8337. DOI: 10.1002/2016JB013361.

Stein, R. S. (1999). "The role of stress transfer in earthquake occurrence". *Nature* 402.6762, pp. 605–609. DOI: 10.1038/45144.

Stein, R. S. (2003). "Earthquake Conversations". *Scientific American* 288.1, pp. 72–79.

Talwani, P. (1984). "Pore pressure diffusion and the mechanism of reservoir-induced seismicity". 122, p. 19.

Tan, Y. J. et al. (2021). "Machine-Learning-Based High-Resolution Earthquake Catalog Reveals How Complex Fault Structures Were Activated during the 2016–2017 Central Italy Sequence". *The Seismic Record* 1.1, pp. 11–19. DOI: 10.1785/0320210001.

Tarantola, A. and B. Valette (1982). "Generalized nonlinear inverse problems solved using the least squares criterion". *Reviews of Geophysics* 20.2, pp. 219–232. DOI: 10.1029/RG020i002p00219.

Terakawa, T. et al. (2010). "High-pressure fluid at hypocentral depths in the L'Aquila region inferred from earthquake focal mechanisms". *Geology* 38.11, pp. 995–998. DOI: 10.1130/G31457.1.

Thomas, D. (1988). "Geochemical precursors to seismic activity". *Pure and Applied Geophysics PAGEOPH* 126.2-4, pp. 241–266. DOI: 10.1007/BF00878998.

Tommasone Pascale, F. et al. (2015). "Rainstorm-induced soil 222Rn concentration spikes observed in Southern Italy". *Environmental Earth Sciences* 73.12, pp. 8177–8187. DOI: 10.1007/s12665-014-3976-0.

Trippetta, F. et al. (2010). "Laboratory measurements of the physical properties of Triassic Evaporites from Central Italy and correlation with geophysical data". *Tectonophysics* 492.1, pp. 121–132. DOI: 10.1016/j.tecto.2010.06.001.

Trippetta, F. et al. (2013). "A multidisciplinary study of a natural example of a CO2 geological reservoir in central Italy". *International Journal of Greenhouse Gas Control* 12, pp. 72–83. DOI: 10.1016/j.ijggc.2012.11.010.

Vadacca, L. et al. (2016). "On the mechanical behaviour of a low-angle normal fault: the Alto Tiberina fault (Northern Apennines, Italy) system case study". *Solid Earth* 7.6, pp. 1537–1549. DOI: https://doi.org/10.5194/se-7-1537-2016.

Valoroso, L. et al. (2017). "Mixed-Mode Slip Behavior of the Altotiberina Low-Angle Normal Fault System (Northern Apennines, Italy) through High-Resolution Earthquake Locations and Repeating Events: Seismic activity of low-angle ATF system". *Journal of Geophysical Research: Solid Earth* 122.12, pp. 10,220–10,240. DOI: 10.1002/2017JB014607.

Viveiros, F. et al. (2008). "Environmental influences on soil CO2 degassing at Furnas and Fogo volcanoes (São Miguel Island, Azores archipelago)". *Journal of Volcanology and Geothermal Research* 177.4, pp. 883–893. DOI: 10.1016/j.jvolgeores.2008.07.005.

Wadati, K. and S. Oki (1933). "On the Travel Time of Earthquake Waves. (Part II)". *Journal of the Meteorological Society of Japan. Ser. II* 11.1, pp. 14–28.

Waldhauser, F. (2000). "A Double-Difference Earthquake Location Algorithm: Method and Application to the Northern Hayward Fault, California". *Bulletin*

of the Seismological Society of America 90.6, pp. 1353–1368. DOI: 10.1785/0120000006.

Waldhauser, F. (2001). "hypoDD – A Program to Compute Double-Difference Hypocenter Locations", p. 25.

Zhang, M., W. L. Ellsworth, and G. C. Beroza (2019). "Rapid Earthquake Association and Location". *Seismological Research Letters* 90.6, pp. 2276–2284. DOI: 10.1785/0220190052.

Zhu, W. and G. C. Beroza (2019). "PhaseNet: a deep-neural-network-based seismic arrival-time picking method". *Geophysical Journal International* 216.1, pp. 261–273. DOI: 10.1093/gji/ggy423.

İnan, S. et al. (2012). "Seasonal variations in soil radon emanation: long-term continuous monitoring in light of seismicity". *Natural Hazards* 62.2, pp. 575–591. DOI: 10.1007/s11069-012-0096-6.

List of Figures