Geophysical Journal International

Geophys. J. Int. (2025) **244,** 1–19
Research Paper

https://doi.org/10.1093/gji/ggaf402

A Markov chain Monte Carlo method to identify correlations in multidisciplinary time-series: application to the TABOO Near Fault Observatory

G. Poggiali [1,2] N. Piana Agostinetti,[3,4] A. Piersanti,[4] A. Caracausi,[4,5] M. Camarda,[4] E. Serpelloni,[4] E. Tondi[1] and L. Chiaraluce[4]

[1] *Università degli Studi di Camerino, School of Science and Technologies,* 62032*, Camerino, Italy. E-mail: giulio.poggiali@uniroma1.it*
[2] *La Sapienza Università di Roma, Dipartimento di Scienze della Terra,* 00185*, Roma, Italy*
[3] *Università degli Studi di Milano Bicocca, Dipartimento di Scienze dell'Ambiente e della Terra,* 20122*, Milano, Italy*
[4] *Istituto Nazionale di Geofisica e Vulcanologia,* 00143*, Rome, Italy*
[5] *Departamento de Geología, Universidad de Salamanca,* 37008*, Salamanca, Spain*

Accepted 2025 October 7. Received 2025 October 1; in original form 2025 March 11

SUMMARY

We present a workflow for the analysis of multiple time-series of geo-observables contemporaneously recorded in a tectonically active area located in a sector of the Northern Apennines of Italy monitored by the Alto Tiberina Near Fault Observatories (TABOO-NFO), an infrastructure providing high-resolution data sets with high spatial density of networks. With the final aim of unveiling interactions among complex geophysical processes related to the earthquake generation, multidisciplinary data are usually analysed with approaches lacking a unified data-processing framework and relying on qualitative post-processing interpretations preventing a more quantitative and statistical analysis of the results. Alternatively, the joint analysis of different data sets produced by multidisciplinary networks requires a methodology to extract and compare information from different sources in a coherent framework. The workflow is based on Bayesian inference and relies on a reversible jump Markov chain Monte Carlo algorithm allowing us to independently model multiple time-series and extract transients (e.g. rapid temporal changes in the signal) under the form of change points distributions to be used as a common comparable output. We modulate our workflow for the variety of characteristics found in the data sets consisting of V_P/V_S, global navigation satellite system and geochemical time-series, with two strategies: a pre-processing step, which filters out or isolates signals that are not directly related to tectonic activity (e.g. rainfall), and a statistical inference step, where we make use of a versatile parametrization that allow our algorithm to deal with complex time-series exhibiting correlated samples and outliers. For each data set we discuss the advantages of our approach as well as areas of future improvements. The outputs of the analysis are compared with seismicity rate and main earthquakes recorded in the study area considering different ways of gathering the results: single station outputs, multistation and finally a multidisciplinary comparison of change points distributions in time.

Key words: Time-series analysis; Joint inversion; Earthquake interaction, forecasting, and prediction; Continental tectonics: extensional.

1 INTRODUCTION

Understanding seismic activity and its preparatory phase up to the (slow and/or fast) release of energy and deformation is an ambitious goal that involves unravelling mechanical, physical and chemical properties evolving in space and time within complex natural systems. Investigating the relationships between earthquake activity and various measurable parameters is not new: efforts have been made for at least fifty years to try to better understand earthquake generation through the analysis of a wide variety of measurements, mostly in search of precursors (Scholz 1973; Thomas 1988; Cicerone *et al.* 2009, among others). Despite the relevance and popularity of the topic, the problem has rarely been tackled with a true multidisciplinary approach, intended as the

analysis of multiple observables derived from different disciplines and collected simultaneously for the same area.

Our study is based on the idea that a coherent and self-consistent analysis of diverse geo-observables can shed light on the characteristics of seismic activity and, more in general, of complex tectonic processes which comprise multiple phenomena. Consequently, the aim of this work is to develop a novel and unique workflow for the analysis of multiple time-series recorded in the same study area, adopting a state-of-the-art statistical approach.

The workflow is 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 separated in time at so-called *ChangePoints* (CP), and extract variations (transients) which can be, in a final stage, represented and analysed as CP distributions. By transients we mean relatively rapid temporal changes in the signal (e.g. in the mean or slope of the time-series). A CP is the time of occurrence for one of these transients as detected by the algorithm. The ability of Bayesian methods to deal with nonlinear and non-unique problems (Tarantola & Valette 1982; Mosegaard & Tarantola 1995), made these methods appealing in many geophysical fields and the trans-dimensional implementation (Green 1995) has proven successful in a wide variety of applications (Dettmer *et al.* 2010; Bodin *et al.* 2012a, b; Hawkins & Sambridge 2015; Piana Agostinetti *et al.* 2015). Trans-dimensional algorithms, the class of algorithms in which falls the rj-McMC method, 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 number of CPs and, thus, on the model complexity. Given its parsimonious nature (Malinverno 2002), these algorithms make use of a limited number of piecewise functions for each data set, that is, including only piecewise functions directly supported by the data. The occurrence of a discontinuity between two piecewise functions, a CP, indicates the presence of a discontinuity in the time-series. Collecting and analysing all discontinuities between all simulated data sets allows us to relate variations detected at multiple sites and potentially involving different parameters. Examples of applications on various kinds of time-series can be found in literature: borehole temperature data (Hopcroft *et al.* 2007, 2009), geochemical data (Gallagher *et al.* 2011), thermochronology (Gallagher 2012).

Here we present a novel workflow that we apply to the multidisciplinary time-series recorded by the the Alto Tiberina Near Fault Observatory (Chiaraluce *et al.* 2014) (described in Section 1.1), comprising seismological (V_P/V_S), geodetic (Global Navigation Satellite System, GNSS) and geochemical (CO_2 and radon emissions) time-series.

Our workflow is based on a Bayesian framework due to the aforementioned strengths in handling CP detection problems: its parsimonious nature ensures that the model complexity is driven by the data; statistically sound way to separate prior information from the data itself; the ability to estimate the number of CPs as a free parameter is a significant advantage, as it eliminates the need for subjective assumptions. The final output, a common distribution of change points, enables a consistent and comparable analysis across multiple time-series. Other techniques commonly applied to geophysical time-series often rely on simple moving windows or filtering, which may require subjective assumptions on the number of changepoints (for a comparison, see Bernier 1994; for an example on V_p/V_s data, see Lucente *et al.* 2010). For a thorough comparison of methods for offset detection in GNSS data (which can be viewed as a specific CP problem), see Gazeaux *et al.* (2013). While Machine Learning is rapidly changing the landscape of time-series analysis, providing state-of-the-art techniques (e.g. Gao *et al.* 2024), these methods can still require substantial training data or may lack the full probabilistic framework of a Bayesian approach.

The workflow is composed of two main phases. The first phase consists in a pre-processing phase which is necessary for some of the data sets 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).

The second phase, following data set preparation, involves applying the rj-McMC algorithm to each individual time-series and conducting a synoptic analysis of the results. We opted to apply the algorithm to individual time-series, rather than, for example, performing a joint inversion (as in Licciardi *et al.* 2020) of multiple time-series to produce a common CPs distribution. This approach was chosen because we did not want to impose a synchronous CPs structure, given that a variation in one data set could precede or follow a variation in another, or the observed seismicity rate. The application of the rj-McMC algorithm to the single-data time-series determines a set of CPs that describe discontinuities in the time-series itself. We thus obtain a set of CP per data set. After, we examine the collections of CPs obtained for each kind of data (but at different location/sensors/stations) to assess how well the rj-McMC method has worked for that kind of data, and whether there is useful information revealed from the resulting CP distributions. Finally, we compare the CP distributions obtained from the different kind of data, to see if they correspond. Correspondence might suggest a common underlying geological process. Additionally, in case of the availability of an excellent seismic catalogue for the study region, we can compare the time-series and CPs to the seismicity rate.

The proposed method extracts quantitatively comparable information from all the different sources in a unique, coherent framework, straightforward contributes to innovative and all around, multiparameter and multipurpose warning systems for monitoring tectonically active regions and, thus, assists and facilitates expert opinion in relating and comparing the different signals recorded by the increasingly widespread multidisciplinary networks. This kind of approach, based on a purely statistical approach, can also broaden our knowledge about the existence, or not-existence, of relations between earthquakes and various measurable parameters with potential significance as a proxy for the local tectonic process, involving both observables directly linked with seismicity, like geodetic and seismic signals, and others for which the links are indirect and debated, like radon and CO_2 soil flux.

1.1 The Alto Tiberina Near Fault Observatory

In the last decade the development of European research infrastructures such as Near Fault Observatories (NFO) allowed to conduct multidisciplinary experiments in 'natural laboratories', created with the aim of understanding the physical–chemical processes related to earthquakes, tectonic activity and evolution of fault systems (Chiaraluce *et al.* 2022). One NFO has been implemented in central Italy, the Alto Tiberina Near Fault Observatory (TABOO; Chiaraluce *et al.* 2014) (Fig. 1) TABOO 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 allowing complementary investigation

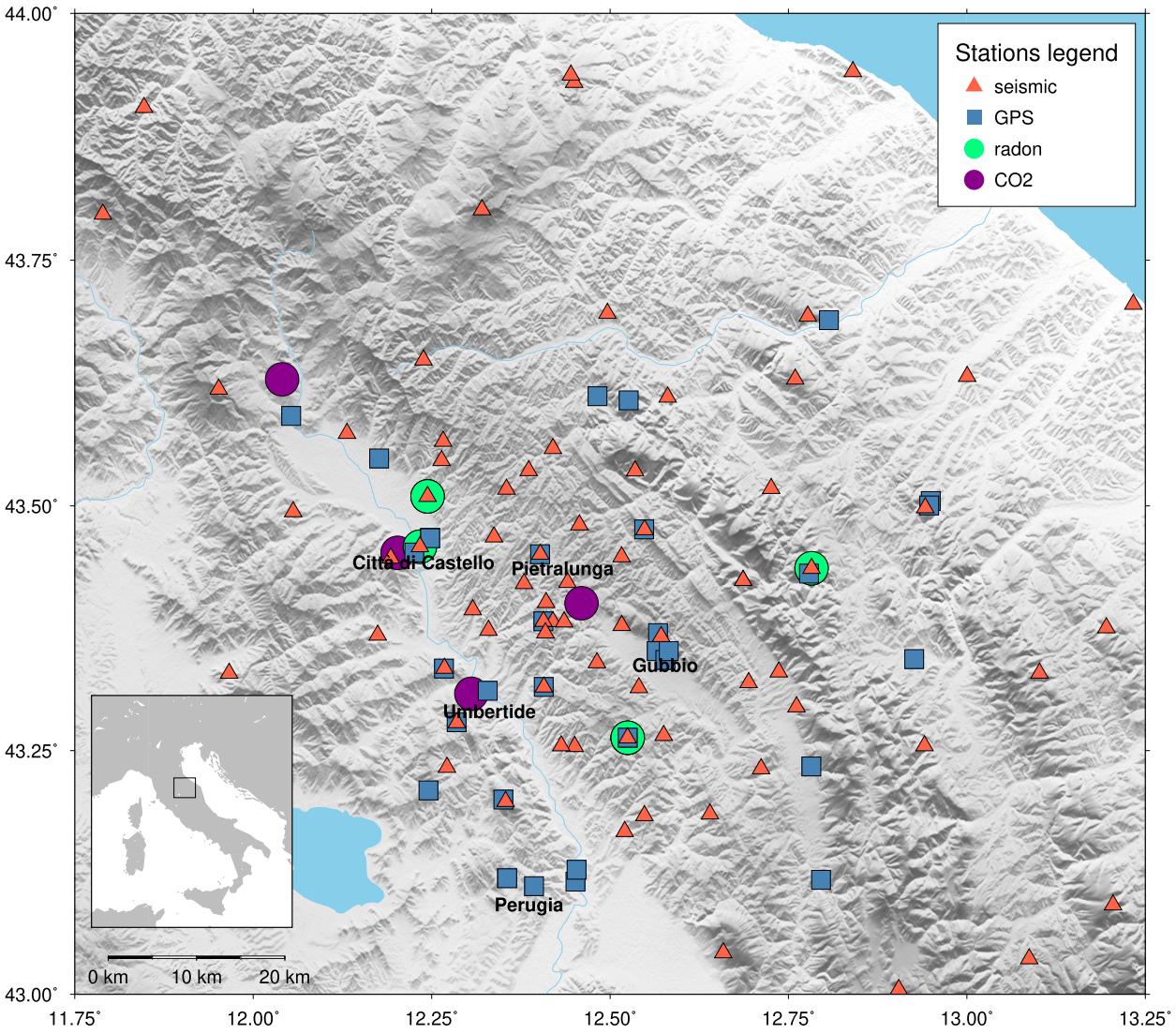

Figure 1. Map of the study area. Only stations used in this work are reported.

techniques. This section of the Apennine has unique potential for multidisciplinary studies not only because of the dense instrumentation and high-seismicity rate, but also for the occurrence of aseismic deformation (Gualandi *et al.* 2017) and deep fluids circulation (Chiodini *et al.* 2004).

From a seismological point of view, many studies showed a strong microseismic activity occurring along a low-angle normal fault, the Altotiberina fault (ATF) with the majority (∼90 per cent) of earthquakes originating in the hanging wall of the 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 obtain a detailed picture of the subsurface from high-quality seismic catalogues (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 traveltimes of seismic phases to detect variations which are indications of the processes taking place in the subsurface (Li *et al.* 1998; Schaff David & Beroza Gregory 2004; Lucente *et al.* 2010). 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 feature 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 & Bennett 2009; Anderlini *et al.* 2016). The high density of geodetic stations allowed also to detect an aseismic transient deformation episode (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 per cent of lithostatic load has been encountered (Chiaraluce *et al.* 2007), and in very high CO_2 flux emissions at the surface (up to 5800 $t\,yr^{-1}$), values comparable to active volcanic regions (Rogie *et al.* 2000; Marchesini *et al.* 2022).

Geochemical instrumentation installed in the area also includes monitoring of radon emissions (Piersanti *et al.* 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 & Brady 1981; Mollo *et al.* 2011).

This work introduces a novel methodology to leverage the multidisciplinary data set recorded by the NFO. By analysing various time-series within a single, quantitative framework, we aim to enhance the understanding of interactions among fluids, crustal deformation, and seismic activity in the area.

2 METHODOLOGY: THE rj-McMC SAMPLING ALGORITHM

Here our goal is to produce comparable outputs between data derived from different disciplines in order to identify transient signals in the analysed data sets. The workflow we propose relies on Bayesian inferences driven by a rj-McMC algorithm as an analysis tool common to all data sets. We implemented a parametrization that allows us to overcome or at least mitigate some of the data set-specific issues 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 analysed, 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.

2.1 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 eq. 1). All terms, PPD and Likelihood, must be expressed as probability distributions and the relation between them is commonly expressed with the following formula:

$$p(m|d) \propto p(m)p(d|m). \tag{1}$$

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 noise statistics (Mosegaard & Tarantola 1995). The most used technique to obtain a numerical approximation of the PPD is Markov chain Monte Carlo (McMC) sampling (Tarantola & Valette 1982; Mosegaard & Tarantola 1995; Gelman & Rubin 1996): 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, that is, the importance sampling follows the target distribution.

In this study, we adopt the McMC approach of Mosegaard & Tarantola (1995) to sample the PPD. In such approach, a candidate model $\mathbf{m}_{\rm cand}$ is created as a perturbation of the current model $\mathbf{m}_{\rm curr}$ based on the prior distributions, and is accepted or rejected following an acceptance probability:

$$\alpha = min\left[1, \frac{L(\mathbf{m}_{\rm cand})}{L(\mathbf{m}_{\rm curr})}\right], \tag{2}$$

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 & Tarantola 1995). The Likelihood function, given its key-importance for our implementation, is described in details in Section 2.3. The expression used for the acceptance probability is a special case of the general acceptance criterion

$$\alpha = \min\left[1, \frac{q(\mathbf{m}_{\rm curr}|\mathbf{m}_{\rm cand})}{q(\mathbf{m}_{\rm cand}|\mathbf{m}_{\rm curr})}\frac{p(\mathbf{m}_{\rm cand})}{p(\mathbf{m}_{\rm curr})}\frac{L(\mathbf{m}_{\rm cand})}{L(\mathbf{m}_{\rm curr})}|\mathbf{J}|\right] \tag{3}$$

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 vice versa, and a Jacobian term $\mathbf{J}$ of the transformation from $\mathbf{m}_{\rm curr}$ to $\mathbf{m}_{\rm cand}$ which considers the (potential) different number of dimensions in the two models. This last term is introduced to handle the so-called trans-dimensional 'reversible-jump' algorithm 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 modelling, 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 Piana Agostinetti & Bodin (2018), the sampling strategy of Mosegaard & Tarantola (1995) guarantees that the prior distribution equals the proposal distribution and, by adopting specific transformations (see Appendices in Agostinetti & Malinverno 2010), the Jacobian term is unity and can be conveniently ignored.

2.2 Model parametrization

The diverse characteristics of the data sets, discussed in Section 3, result in the need of a versatile parametrization that can adequately adapt the model complexity to each observable with the most suitable set of basis functions. Moreover, due to the use of a variety of geo-observables with different noise statistics, we also add, to the physical model, the so-called 'hyperparameters' (also called *nuisance* parameters) that help in configuring the error models for specific cases (for example, in presence of correlated or uncorrelated noise between data points, or for modulating a poorly known data uncertainty), to avoid introducing bias in the final results given by the unknown error statistics.

Here, for modelling the geo-observables, we adopt a piecewise-modelling. 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 two additional parameters, (a_{k+1} and b_{k+a}, for the data between c_k and $t_{\rm end}$ (end of the observing period), which can be thought of as half-space values. So, the piecewise model vector can be written as:

$$\mathbf{m_{piecewise}} = (k, \mathbf{c}, \mathbf{a}, \mathbf{b}) \ , \tag{4}$$

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

To deal with other properties of the data sets, 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 & 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 is directly related to the level of data noise. As exemplified in Agostinetti & Malinverno (2010), the noise level is inversely proportional to the complexity of the solution. Imposing a user-defined noise level is equivalent to obtaining a solution with a specific complexity: a higher value would determine a less complex solution and vice versa (see example in Fig. S1, Supplementary Material). 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. In this way, we fully implement the 'parsimoniosity' characteristic of the reversible-jump technology.

The second noise parameter r is introduced to treat the serial correlation that some data set exhibit (mainly geochemicals). This property results in a covariance matrix which can no more be assumed as diagonal. We follow the definition of Bodin *et al.* (2012bb) 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 D1 of Bodin *et al.* (2012bb), 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:

$$\mathbf{m_{error}} = (\omega, r) \ . \tag{5}$$

2.3 Likelihood function

Our implementation of the Likelihood function builds on the one described in Poggiali *et al.* (2019) and uses the sum of absolute values of the residuals (L_1-norm). As explained in Mosegaard & 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 2.2), can be written as follows. The covariance matrix can be decomposed as $\mathbf{C_e} = \mathbf{SRS}$ (Malinverno & 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 $\mathbf{e} = \left|d_i^{\rm obs} - d_i^{\rm sim}\right|^{1/2}$ (with $d^{\rm obs}$ and $d^{\rm sim}$ representing observed and simulated data points, respectively). In this case, following Malinverno & Briggs (2004), the likelihood function

$$L(\mathbf{d}|\mathbf{m}) = \frac{1}{2^N \left|\mathbf{C_e}\right|} \exp\left(-\mathbf{e}^T \mathbf{C_e}^{-1} \mathbf{e}\right) \tag{6}$$

simplifies to

$$L(\mathbf{d}|\mathbf{m}) = \frac{1}{(1-r^2)^{n-1} 2 \prod_{i=1}^{n} 10^{\omega} \sigma_i} \exp\left(\frac{-\phi}{1-r^2}\right) \tag{7}$$

with

$$\phi = \phi_1 + \sum_{i=2}^{n-1} \phi_i + \phi_n \tag{8}$$

where:

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

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

and

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

The case of uncorrelated 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.

2.4 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 & 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 & 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 note that a uniform prior is also used to sample the model dimension. This means that every potential number of CPs, namely between 1 and 100, has the same probability of being selected *a priori*. A bit counter-intuitively, 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.

2.5 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 eq. (2) presented before. To produce a candidate model from the current one, we randomly pick one of the following moves (grouped into two main categories):

(1) **moves that affect piecewise parameters:**

(i) (proposed with probability 0.20) perturb intercept value a of a CP;
(ii) (0.20) perturb slope value b of a CP;
(iii) (0.20) perturb the position of a CP in time c;
(iv) (0.1) create a new CP (birth move);
(v) (0.1) delete an existing CP (death move);

(2) **moves that affect noise parameters**:

(i) (0.1) perturb noise scale value ω;
(ii) (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 & Malinverno (2010).

We here briefly depict the sampling workflow, summarizing the main steps, which are repeated for a user defined number of iterations. All saved models are post-processed at the end of the McMC sampling to obtain the approximation of the PPD (from which all figures in this manuscript are drawn):

(i) pick a candidate model $\mathbf{m}_{\text{cand}}$ by perturbing the properties of $\mathbf{m}_{\text{curr}}$ with one of the moves illustrated above;
(ii) compute the likelihood of $\mathbf{m}_{\text{cand}}$;
(iii) accept or reject the candidate model following eq. (2): if $\mathbf{m}_{\text{cand}}$ is accepted then it replaces $\mathbf{m}_{\text{curr}}$, otherwise $\mathbf{m}_{\text{curr}}$ is retained;
(iv) save $\mathbf{m}_{\text{curr}}$ and restart from (A).

We used the versatility of our parametrization to adapt the modelling to the different scenarios that each data set represents:

(i) for V_P/V_S we used a simple parametrization made only by a piecewise constant function (slope parameter fixed to 0);
(ii) GNSS observations are modelled with piecewise linear, noise scale value but no noise correlation parameter;
(iii) Geochemical data sets (radon and CO_2) are modelled using piecewise linear, and both noise parameters.

Each move is tuned, as common practice in Bayesian inversion (Agostinetti & 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. It is worth noting that, as far as the recipe and the moves are statistically correct, the results of the algorithm do not depend from the recipe details. Recipe definition only impacts the sampling efficiency (i.e. how many models need to be sampled to have a stable numerical reconstruction of the PPD).

Synthetic tests of the algorithm are reported in Figs S2 and S3 (Supplementary Material).

For V_P/V_S time-series the algorithm was run with 10 independent chains for 10M iterations each; the first half of each chain is discarded as burn-in. As introduced before, the data sets that require seasonal components (GNSS and geochemical) are modelled 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 was first run 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.

3 TIME-SERIES PRE-PROCESSING

The multidisciplinary time-series analysed in this study are derived from both raw data and scientific products gathered by the TABOO research infrastructure. Data sets include time-series of geodetic, geochemical (soil CO_2 flux and radon emissions) and V_P/V_S ratios for a maximum time span of 10 yr (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 analyse and the tectonic processes taking place which is instead 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 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). After this process, we will be able to perform the search for the variations (e.g. change points) and observe the outcomes.

Data sets have very different properties from each other and have required a variable amount of preparation/pre-processing in order to be analysed with the McMC algorithm. For V_P/V_S, this first step consisted mainly in building the time-series from raw data sets: we followed the same cluster-station logic as described in Poggiali *et al.* (2019) to accomplish this task. On the other hand, 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 (mainly annual and semi-annual) are simulated and the parameters are estimated following an additional Bayesian approach, before the subsequent rj-McMC modelling. 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

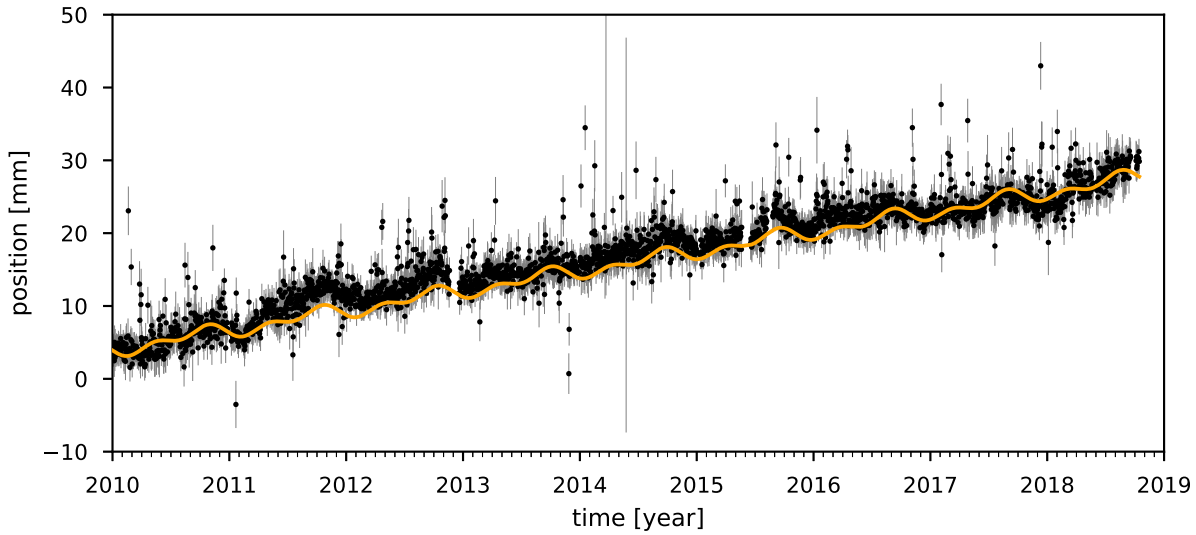

Figure 2. Example of data from the E–W component of GNSS station ARCE (black points). Estimated sinusoid (annual plus semi-annual components) superimposed in orange (arbitrary intercept and slope).

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 data sets with hourly sampling (geochemicals) are resampled on a daily basis to remove eventual daily and subdaily oscillations.

3.1 Global navigation satellite system (GNSS) data

We analysed horizontal components of continuous GNSS displacement time-series from stations located inside the TABOO area (Fig. 1) and contributing to the INGV Rete Integrata GNSS Nazionale (http://ring.gm.ingv.it). These time-series represent the daily evolution of each site position in time relative to the Eurasian plate over a maximum of 10 yr 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 step, the instrumental offsets, due to changes in the station's equipment, are removed. Then, a standard parameter search (Sambridge & Mosegaard 2002) is operated to define the most probable values of the annual and semi-annual periodicity which generally affect GNSS time-series (Serpelloni *et al.* 2006). In details, we consider the model:

$$p_{\mathrm{gps}}(\mathbf{m}) = \alpha_1 sin(2\pi\gamma_1 t + \beta_1) + \alpha_2 sin(\pi\gamma_2 t + \beta_2)\,, \qquad (9)$$

where: $p_{\mathrm{gps}}(\mathbf{m})$ 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 semi-annual terms (t is measured in years). $\gamma_{1,2}$ have values between 0.9 and 1.1 and are used to adjust annual and semi-annual periods, so that they are not strictly annual and semi-annual. The optimal 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 analysed 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 Fig. 2). Time-series shorter than 2 yr are not considered. We used a total of 26 GNSS stations in the area (Fig. 1).

3.2 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 & Baubron 1996; 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 data sets 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.

3.2.1 CO_2 data

CO_2 data set analysed in this study consists of time-series of soil CO_2 flux (CO_2 hereafter) acquired at four different TABOO sites: Fungaia, Migianella, Nogna and Uppiano (Fig. 3). The stations are positioned close to sites characterized by strong CO_2 flux, up to 5800 t yr^{-1} (Camarda *et al.* 2019), evidenced in some cases by the presence of bubbling water at the surface. The time span considered in this study goes from June 2015 to December 2019, with almost 3 yr 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 behaviour, often site specific, but always characterized by long-term seasonal oscillations and short-term effects related to rainfall and soil water

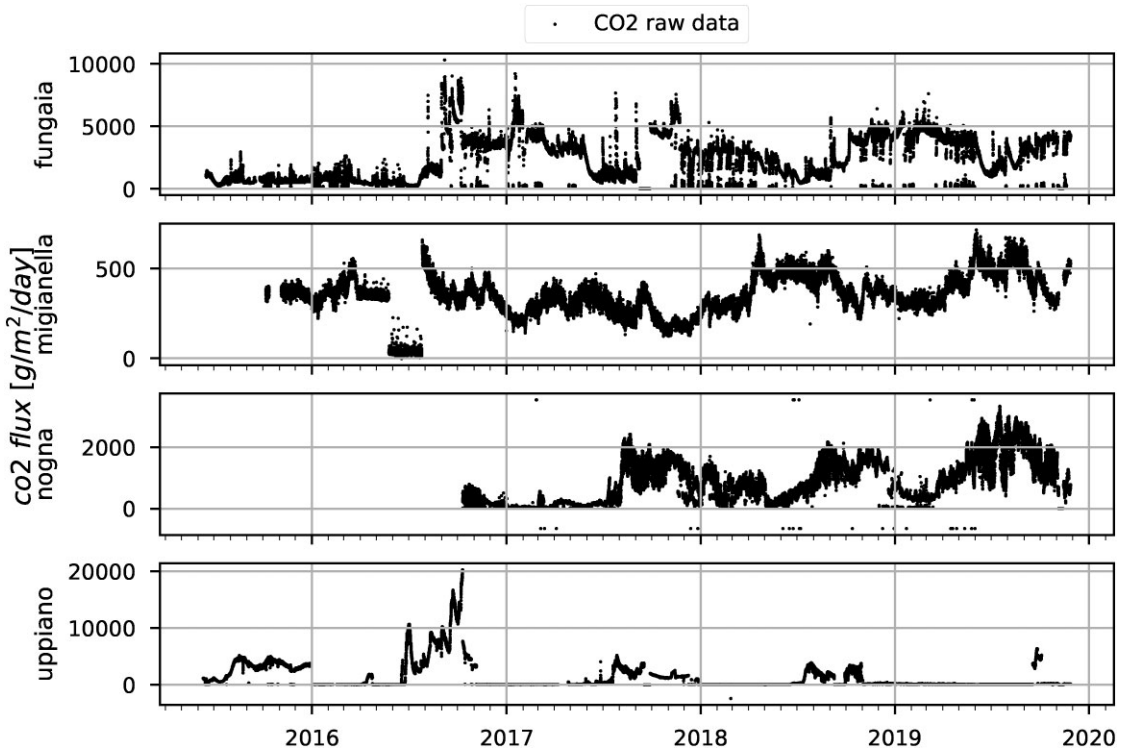

Figure 3. CO_2 raw data set.

content (Granieri *et al.* 2003; Viveiros *et al.* 2008; Camarda *et al.* 2019).

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 per cent 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 Fig. 4 for an example of how this filter works on a data set). Being the cumulative sum for each sample calculated over the previous 2 d, 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 modelled on soil humidity, both for simplicity reasons and to manipulate the data sets 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. This means that in that cases the posterior will equal the prior, gaining no additional information from the likelihood.

After the removal of data points attributable to rainy events, we resample each data set at 1 sample per day, taking the mean and the standard deviation, respectively, as our final measurements and errors to be analysed with the McMC algorithm. Periods with no meteorological data available were not considered: for example, Nogna meteorological station unfortunately reported malfunctions until November 2017, so a large fraction of this data set was discarded. The effect of daily and/or subdaily 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. 9) during the pre-processing steps to isolate such contribution before the rj-McMC step. In particular, we adopt the same scheme used in Section 3.1 and we consider a periodic model with annual and semi-annual periodicities. Parameters related to such a model are estimated using a standard parameter search. 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 Fig. 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 data set by introducing a correlation parameter in the parametrization, as above.

3.2.2 Radon

Radon emissions that we used as raw data are continuous measures of concentrations recorded by four stations located inside the TABOO area: BADI, CDCA, MURB, SSFR (Fig. 6). The stations are part of the Italian Radon mOnitoring Network; 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 hr to obtain a radon concentration reading, so every time-series is sampled at 12 samples per day. The time span we analysed 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 yr of measurements.

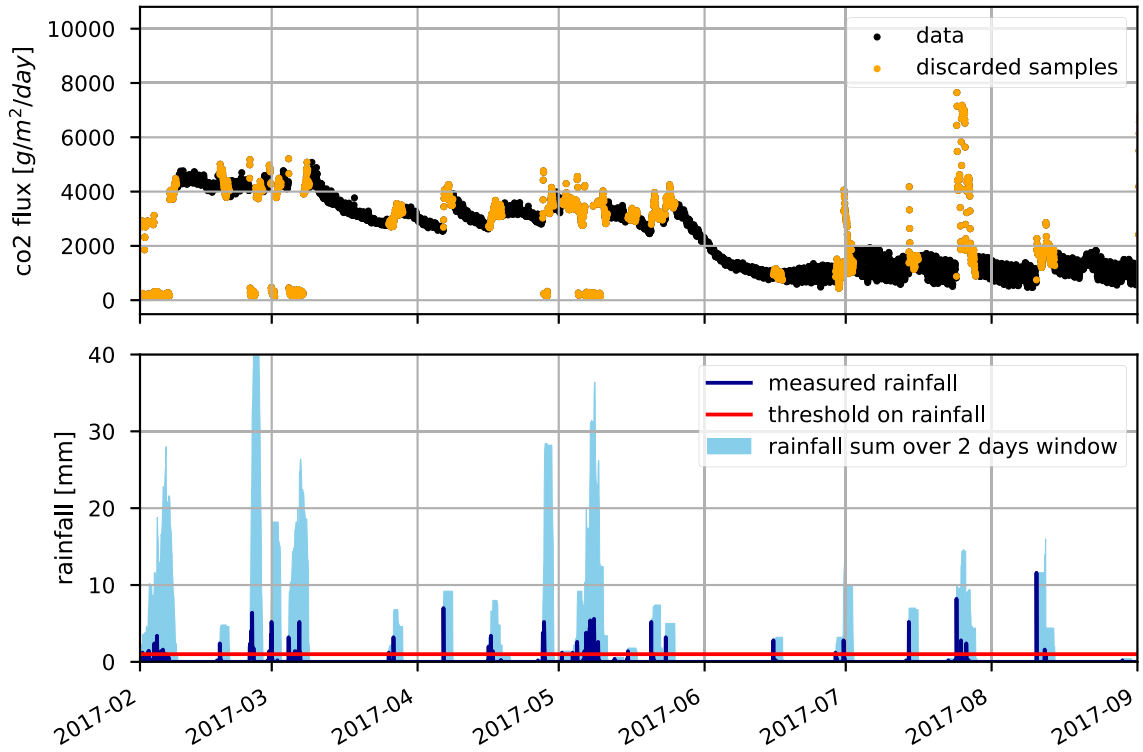

Figure 4. A zoom on Fungaia CO_2 station: here the relation between rainfall (bottom) and CO_2 measurements (top) is evidenced. The implemented filtering, based on a simple rainfall threshold in a moving window, is effective in identifying anomalous data points (marked in orange in top panel) that will 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.

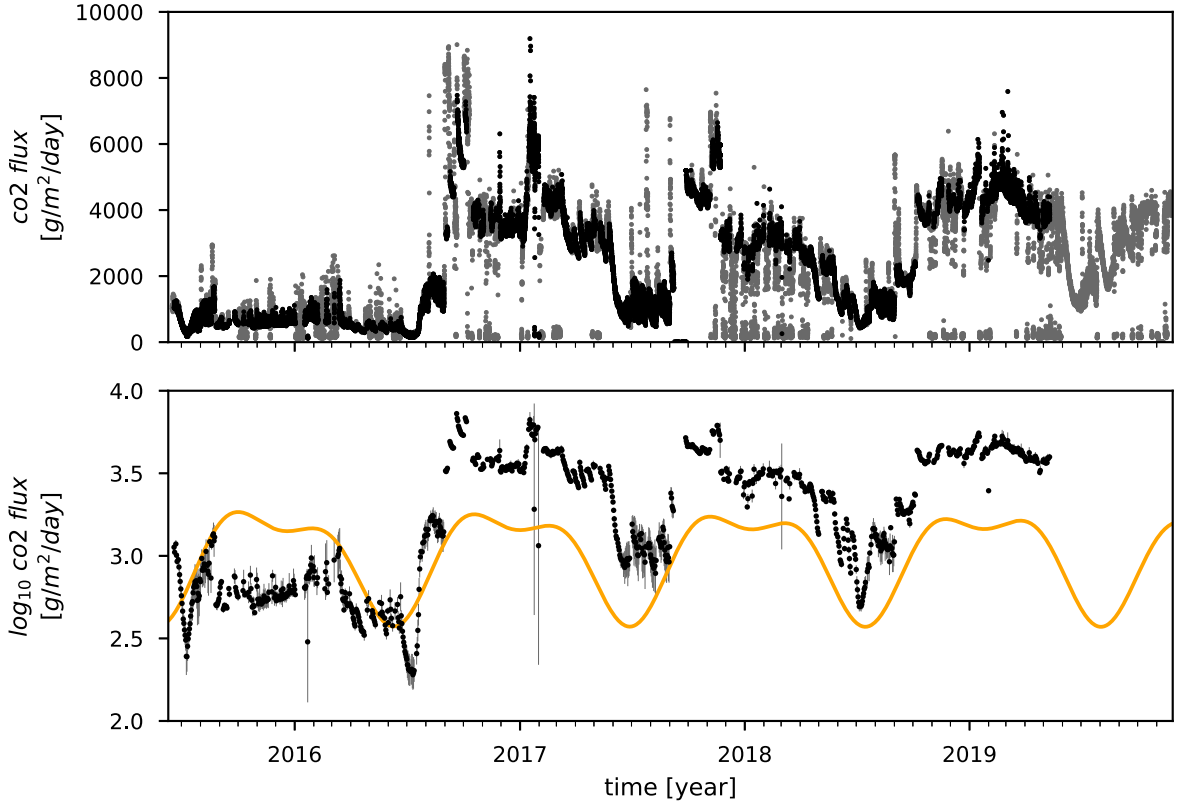

Figure 5. Example of the pre-processing steps for a CO_2 time-series (Fungaia). Top: raw data in grey and filtered data in black. The grey points are discarded measurements due to rainfall events or lack of meteorologic recordings (e.g. at the end of the time-series). Bottom: log transformed data, daily resampled with error bars calculated as daily std. The sinusoid estimated with annual plus semi-annual components is superimposed in orange (arbitrary intercept and slope).

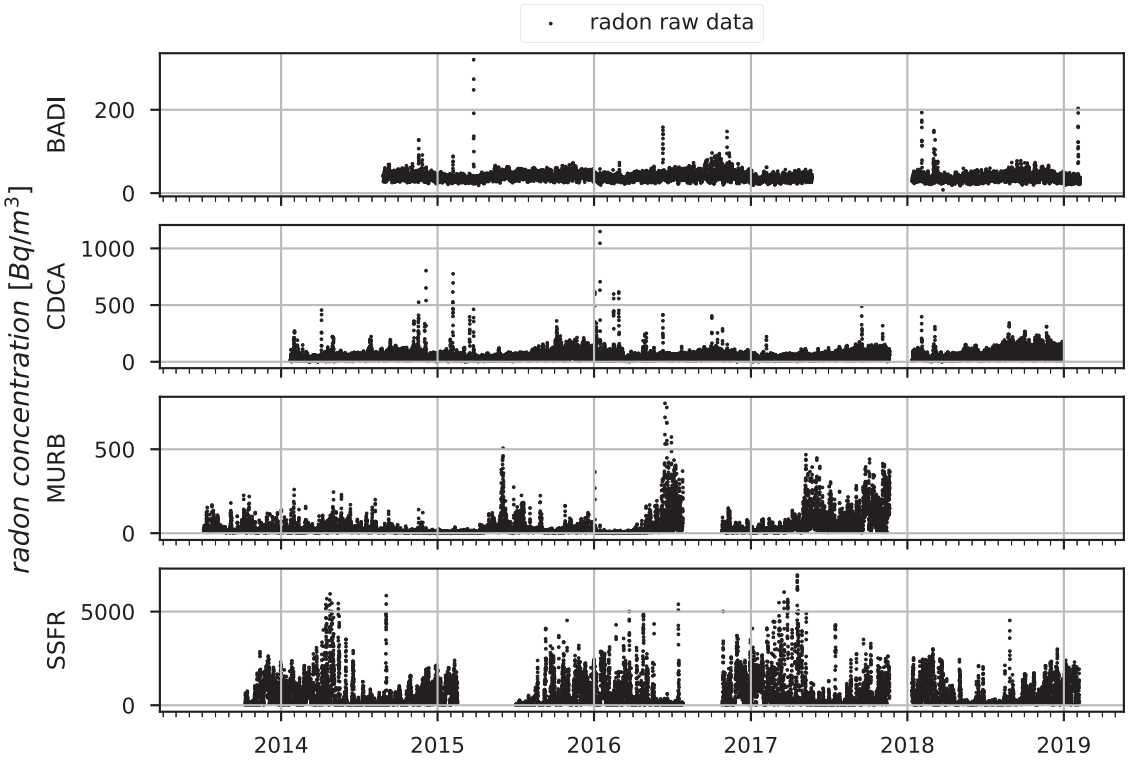

Figure 6. Radon raw data set.

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 (Fig. 6). Radon measurements are the effect of very complex interactions between several factors, and the influence of meteorological parameters is widely documented (Pinault & Baubron 1996; Siino *et al.* 2019 and references therein). Both short (daily and subdaily) and long (annual and semi-annual) oscillations have been highlighted (İnan *et al.* 2012; Siino *et al.* 2019) and a strong effect of rainfall on the measurements is constantly evidenced (Tommasone Pascale *et al.* 2015; Piersanti *et al.* 2016; Cannelli *et al.* 2018), frequently marked by peaks and/or drops in radon time-series data. Moreover the correlation between meteorological parameters and radon emissions exhibits a site-specific behaviour (Piersanti *et al.* 2015, 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 d 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 daily resampling of the remaining data, ∼90 per cent of days are retained with respect to the original data sets (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 subdaily) 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 parameter search to a given model. For consistency, we use the same model as in GNSS and CO_2 (annual and semi-annual periodicities only). An example of the application of this procedure to one radon station (CDCA) is reported in Fig. 7.

3.3 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 traveltimes (Wadati & Oki 1933) using the formula: $V_P\tilde{/}V_S = (t_s - t_p)/t_p + 1$, where $V_P\tilde{/}V_S$ is the average value along a the ray path connecting an earthquake and a seismic station, t_p and t_s are the computed traveltimes (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 82 962 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 hypocentre-based selection:

(i) the study area is gridded with 500 m spacing in three dimensions;

(ii) each node of the grid is the center of a 1 km^3 sphere, called *cluster* hereinafter;

(iii) a cluster-station time-series is computed for a given station and all events within the cluster;

(iv) 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.

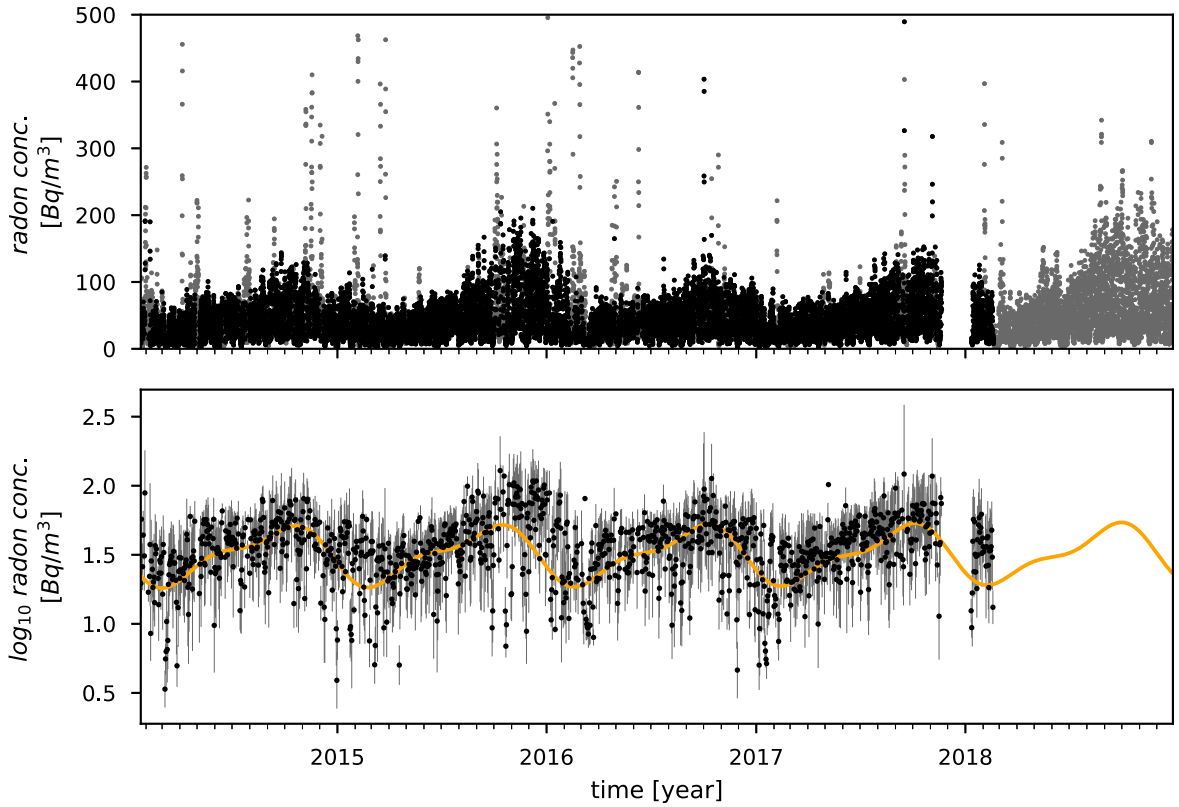

Figure 7. Example of the pre-processing steps for a radon station (CDCA). Top: raw data in grey and filtered data in black. The grey points are discarded measurements due to rainfall events or lack of meteorologic recordings (e.g. at the end of the time-series). Bottom: log transformed data, daily resampled with error bars calculated as daily std. The sinusoid estimated with annual plus semi-annual components is superimposed in orange (arbitrary intercept and slope).

After this procedure we obtained 10 560 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 catalogue (i.e. mislocated events display erroneous origin-time, which, in turn, affect *P*- and *S*-wave traveltimes computation and, finally, V_P/V_S ratio estimates).

4 RESULTS

First, we analyse some examples of the application of the Bayesian algorithm to show the general behaviour with different data sets 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 analyse 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 kinds of data sets (grey histograms in right panels of Fig. 8). Typically V_P/V_S time-series can be modelled 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 data set modelled with periodic signals shows some time windows of considerable length (months to almost a year in the examples of Fig. 8) where seasonal oscillations are enough to properly model the data sets. These portions of the time-series are identified by absence of CPs.

The effect of L_1-norm likelihood implemented in the algorithm can be clearly seen in panel (a) of Fig. 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 down weighted with the L_1-norm and do not give rise to un-necessary CPs. Exceptions are observed, as expected, in case of isolated data points with low error or clusters of data points far from neighbouring values, which cannot be considered isolated outliers. These situations are observed more frequently with radon stations, of which examples are visible in panel (d) of Fig. 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 Fig. 8).

A feature observed in the models obtained from the GNSS time-series, as the one in panel (a) of Fig. 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 be seen in Fig. A1. 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

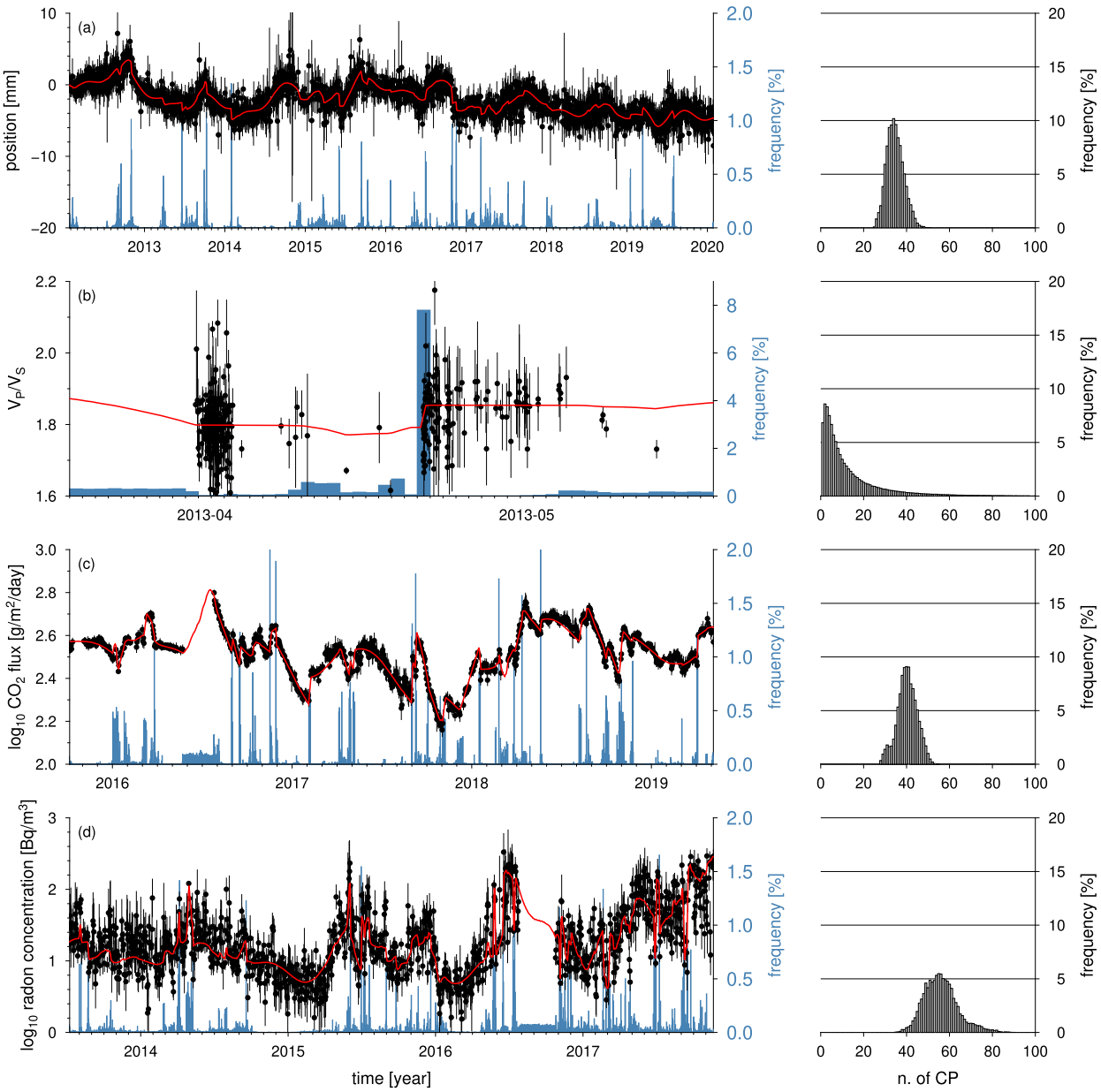

Figure 8. Examples of outputs from different data sets. In all panels, data points are depicted in black, mean model from PPD in red, CP distribution in blue, number of CP on side panels. Histograms represent relative frequencies as percentages. The examples, from top to bottom panel, are respectively: GNSS, V_P/V_S, CO_2 flux, radon.

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 behaviour is expected and it is a consequence of the automatic procedure for time-series creation, without any fine tuning on specific space-time windows, that is, not tailored to seismic sequences. CPs used to fit sparse data are plateau-shaped, as high as 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 Fig. 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 does not show any CP, meaning that a high density of measurements does not translate 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.

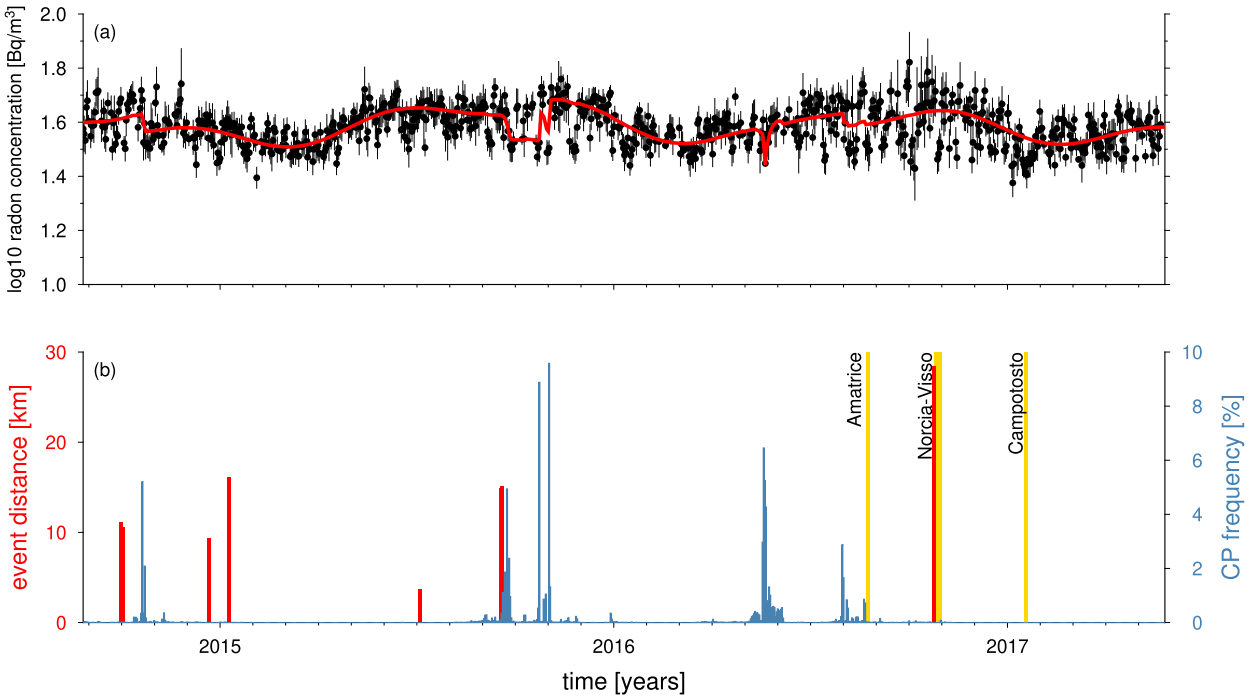

Figure 9. Radon station BADI. Top: data (black) and mean model from PPD (red). Bottom: CP distribution (blue), local earthquakes (red, $M \geq 3$) and regional earthquakes (yellow, $M \geq 5$).

A commonly observed feature of geochemical time-series is the relatively high number of CPs used for the modelling. Considering that geochemical time-series are less than 5 yr long, the CPs density over time is much higher compared to GNSS time-series, which are modelled with an average ∼30 CPs and are almost always longer than 5 yr (except two stations). The reason why such complex models are sampled by the algorithm probably arises from specific characteristics of CO_2 flux and radon data sets. Radon data sets exhibit frequent peaks and drops visible in multiple stations: CDCA and MURB have approximately the same number of CPs (∼60) in the same time span (4 yr). 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 multiday scale, a site specific feature already observed in radon measurements (Siino *et al.* 2019). In all mentioned cases, the data points cannot be considered outliers by the algorithm as a consequence of data set properties and estimated noise level. This happens despite using an L_1-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 behaviour, somehow similar to overfitting, is nonetheless determined by the data.

In the end, for both CO_2 flux 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 a case in which the direct comparison with local and regional seismicity is actually possible. In Fig. 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 is not 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. Performing further analysis taking into account for example the stations hosting lithologies, may be considered.

Finally, to look for common features possibly hidden in the complex output of multiobservables results, we gathered all CPs for each data set and represented them as percentage distributions on a common time axis (Figs 10 and 11, which is a zoom on a time span when all time-series are available).

We did not make any selections, apart from removing V_P/V_S CPs with less than 10 per cent of data in the time-series before or after (one of the criteria used also in Poggiali *et al.* 2019). This is done in order to eliminate CPs at the edges of time-series, not supported by 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

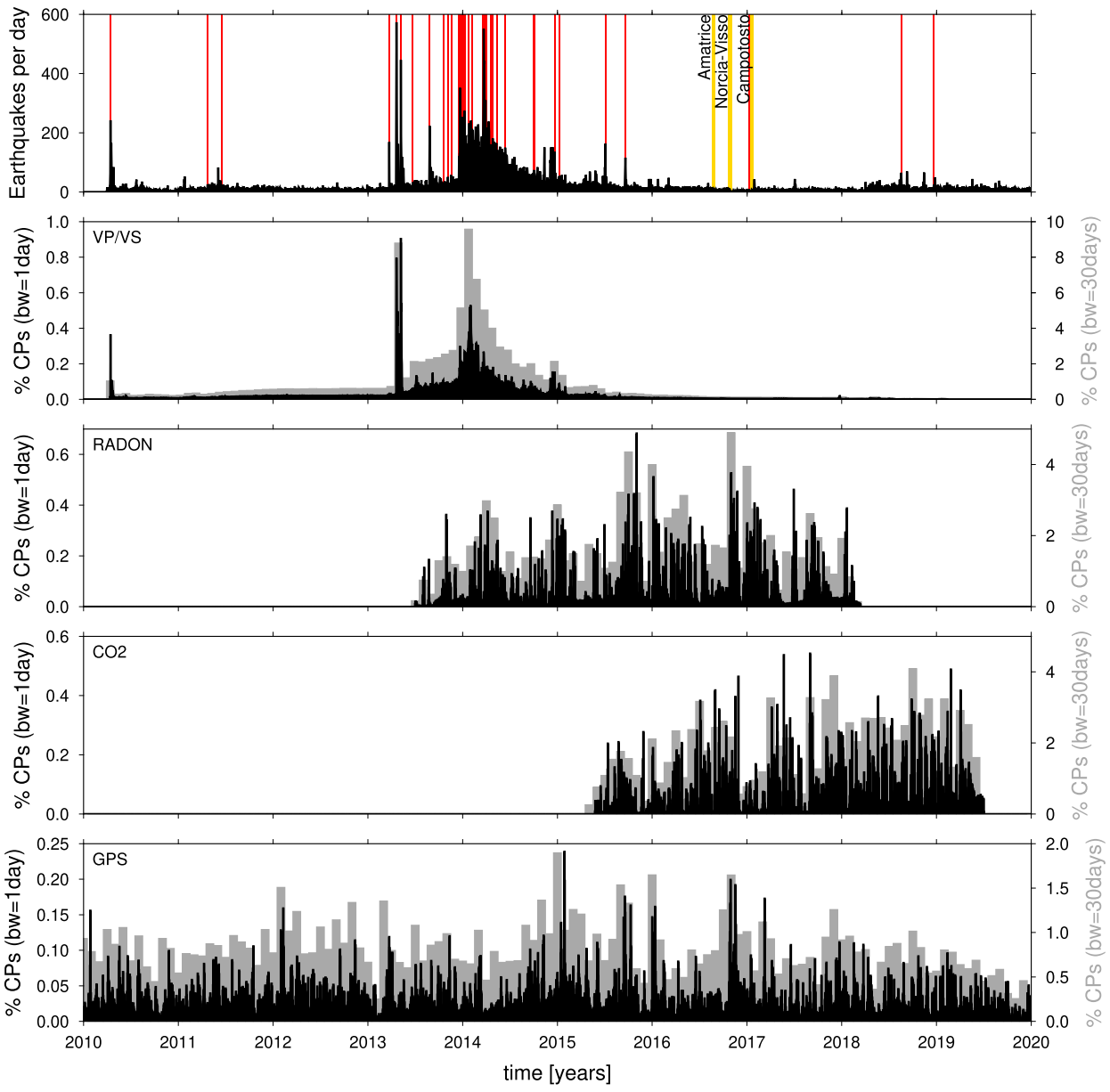

Figure 10. CPs distributions of the multidisciplinary time-series and earthquakes distribution (top). Black histograms use a binwidth of 1 d and grey histograms use a binwidth of 30 d. Local earthquakes ($M \geq 3$) in red and regional earthquakes ($M \geq 5$) in yellow.

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 the appendix. Other features of GNSS models are better evidenced with alternative representations shown in the appendix.

Unfortunately most geochemical data are available only after main seismic sequences of the study area, so with current time-series we cannot 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 (Fig. 9) because of the simple resulting model (few CPs).

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

5 DISCUSSION AND CONCLUSIONS

We generated a novel workflow to analyse different kind of time-series looking for hidden correlations between multidisciplinary time-series provided by TABOO NFO, a monitoring infrastructure composed by dense networks of seismic, geodetic and geochemical (radon and CO2 flux) measuring stations monitoring an area of the northern Apennines characterized by a high rate of (micro-)seismic release (Chiaraluce *et al.* 2014).

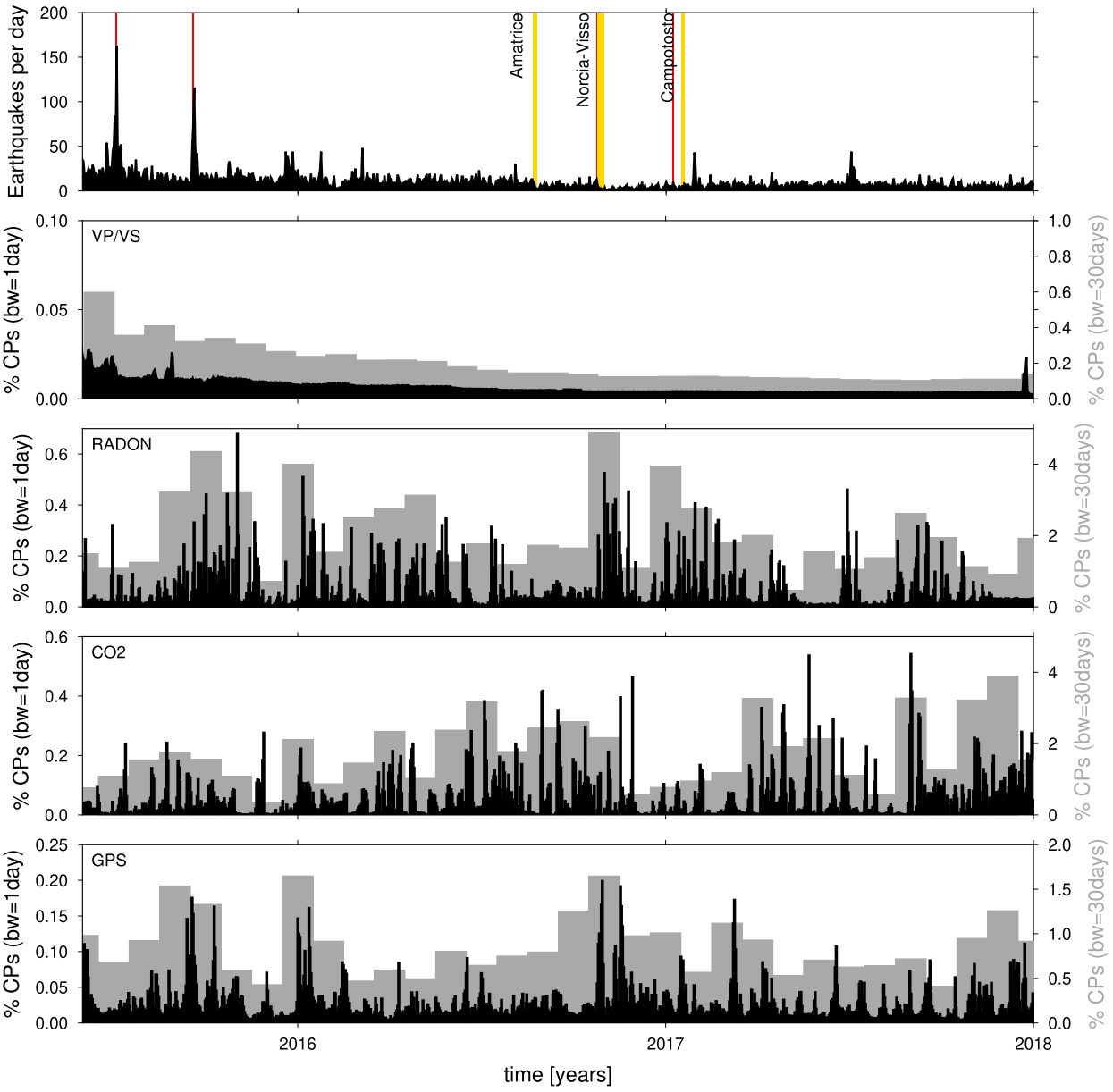

Figure 11. CPs distributions of the multidisciplinary time-series and earthquakes distribution (top). Zoom on the time window where all data sets are available. Black histograms use a binwidth of 1 d and grey histograms use a binwidth of 30 d. Local earthquakes ($M \geq 3$) in red and regional earthquakes ($M \geq 5$) in yellow.

The proposed workflow, based on a Bayesian framework, is aimed to independently analyse different time-series and produce a standardized and comparable output in a synthesis effort. In the building process we have dealt with the challenges related to the different characteristics of the data sets in two main ways: (1) a pre-processing aimed at removing known signals of non-tectonic origin affecting some of the observables and (2) the implementation of an algorithm concerning the likelihood function, covariance matrix and model parametrization allowing the proposed the procedure to be quite 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 allowing to deal with complex data sets including outliers, unknown noise scale and correlation between the measurements. All these intricated factors are treated as unknowns and can be estimated by our approach.

From the modelling point of view, the algorithm is forced to be effectively data-driven and weakly influenced by operator decisions. The Bayesian inversion controls the number and positioning in time of the change points, while the inversions are performed independently so the resulting distribution are a consequence of data and errors properties. Nonetheless, the basic functioning of the CP detecting algorithm, operating on piecewise linear estimates is intrinsically suitable to work on observables that display simple (multiple) linear trends, more than on massively environmentally driven observables like geochemical ones.

However, a successful application can depend on the assumptions made in the pre-processing phase. Especially for some data sets, this step may strongly influence the time-series to be analysed with the Bayesian algorithm and finally the complexity and interpretability of the results. To tackle these issues our approach was keeping 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 semi-annual seasonal components for all data sets 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 of 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 overcomplicated and non-meaningful model.

With these pre-requisite geochemical data are probably the most complicated data to analyse, not only because of the data themselves (high variance, correlation), but also because of unknown factors influencing the measurements. Considering the annual plus semi-annual periodicity, rain effect and daily resampling (to remove subdaily cycles) has proven necessary but not sufficient to produce easily interpretable CPs distribution, in terms of an eventual link to tectonic processes. 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 together with the application of ad-hoc pre-processing strategies are crucial for the evaluation of non-obvious periodicities and other signals of various origin that can affect measurements in a systematic way. Unfortunately, the temporal coverage of the actual geochemical time-series is limited in relation to the major seismic sequences of the area, therefore we did not have the opportunity to fully compare a potentially more evident signal or the lack thereof.

V_P/V_S data confirm to be well suited to be analysed with a Bayesian CPs detection algorithm (as previously demonstrated by Poggiali *et al.* 2019) because of favourable features of data distribution, noise and 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 data set, this connection is not as obvious due to a potentially interesting variation which 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 (Fig. 10).

The proposed method has also proved to be applicable successfully on GNSS time-series. Seismic related offsets are observed in multiple stations, but not in relation to local earthquakes. A long-period pattern emerges (4 yr, if the signal is assumed to be periodic) from the analysis of the ensemble of resulting models (Fig. A1). A common (unknown) signal is also detected at the beginning of 2015 (Fig. A2). Our methodology can be used as an offset detection tool (not only tectonic but also related to instrument failures) for GNSS time-series.

Thus, the proposed approach offers promising insights as a workflow for multidisciplinary time-series analysis. In this first experiment, the pre-processing has been intentionally left oversimplified in order to manipulate the data as little as possible. The estimation of the periodic components could for example benefit from a trans-dimensional implementation on the number of sinusoids, in order to include additional periodicities, when needed, in a data-driven way. Also, a refinement in the estimate and removal of exogenous (e.g. non-tectonic) contributions would allow applications of the approach proposed here to multiparametric monitoring of active seismic areas as a tool aiding expert opinion. Finally, together with the improvements of the pre-processing step, we believe that the investigation may become more interesting from an interpretative point of view with the availability of longer time-series covering the same time interval and collected in many areas with different tectonic characteristics in order to have an increasingly comprehensive and meaningful statistic over time.

ACKNOWLEDGMENTS

GP did most of this work during his doctoral thesis at the University of Camerino, and then finalized it together with LC during his Postdoc at La Sapienza University of Rome supported by the ERC Adv. Grant 'The physics of earthquake faulting: learning from laboratory earthquake prediction to improve forecasts of the spectrum of tectonic failure modes'—TECTONIC—under the H2020 programme (https://cordis.europa.eu/project/id/835012).

SUPPORTING INFORMATION

Supplementary data are available at *GJIRAS* online.

Please note: Oxford University Press are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

DATA AVAILABILITY

All the (time-series) data are available at the following open access portals:

(i)Seismic and geochemical data—https://fridge.ingv.it;
(ii)GNSS data—https://gnssdata-epos.oca.eu/

The code used in this study is available from the authors upon request.

REFERENCES

Agostinetti, N.P. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **181,** 858–872.

Anderlini, L., Serpelloni, E. & Belardinelli, M.E., 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, *Geophys. Res. Lett.,* **43**(9), 4321–4329.

Antonioli, A., 2005. Fluid flow and seismicity pattern: evidence from the 1997 Umbria-Marche (central Italy) seismic sequence, *Geophys. Res. Lett.,* **32**(10), 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, *Annu. Rev. Earth Planet. Sci.,* **43**(1), 233–271.

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, *Phil. Trans.,* **53,** 370–418.

Bernier, J., 1994. Statistical detection of changes in geophysical series, *Engineering Risk in Natural Resources Management: with Special References to Hydrosystems under Changes of Physical or Climatic Environment,* pp. 159–176, Springer.

Bodin, T., Sambridge, M., Rawlinson, N. & Arroucau, P., 2012a. Transdimensional tomography with unknown data noise: transdimensional tomography, *Geophys. J. Int.,* **189**(3), 1536–1556.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012b. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117**(B2), doi:10.1029/2011JB008560.

Camarda, M., De Gregorio, S., Di Martino, R.M.R. & Favara, R., 2016. Temporal and spatial correlations between soil CO_2 flux and crustal stress: soil CO_2 flux and crustal stress, *J. Geophys. Res.: Solid Earth,* **121**(10), 7071–7085.

Camarda, M., De Gregorio, S., Capasso, G., Di Martino, R.M.R., Gurrieri, S. & Prano, V., 2019. The monitoring of natural soil CO_2 emissions: issues and perspectives, *Earth Sci. Rev.,* **198,** 102 928, doi:10.1016/j.earscirev.2019.102928.

Cannelli, V., Piersanti, A., Galli, G. & Melini, D., 2018. Italian Radon mOnitoring Network (IRON): a permanent network for near real-time monitoring of soil radon emission in Italy, *Ann. Geophys.,* **61**(2018), 1, doi:10.4401/ag-7604.

Chanard, K., Métois, M., Rebischung, P. & Avouac, J.P., 2020. A warning against over-interpretation of seasonal signals measured by the Global Navigation Satellite System, *Nat. Commun.,* **11**(1), 1375, doi:10.1038/s41467-020-15100-7.

Chiaraluce, L., Chiarabba, C., Collettini, C., Piccinini, D. & Cocco, M., 2007. Architecture and mechanics of an active low-angle normal fault: Alto Tiberina Fault, northern Apennines, Italy, *J. geophys. Res.,* **112**(B10), doi:10.1029/2007JB005015.

Chiaraluce, L. *et al.*, 2014. The Alto Tiberina Near Fault Observatory (northern Apennines, Italy), *Ann. Geophys.,* **57**(3).

Chiaraluce, L. *et al.*, 2022. The near fault observatory community in europe: a new resource for faulting and hazard studies, *Ann. Geophys.,* **65**(3), DM316, doi:10.4401/ag-8778.

Chiodini, G., Cardellini, C., Amato, A., Boschi, E., Caliro, S., Frondini, F. & Ventura, G., 2004. Carbon dioxide Earth degassing and seismogenesis in central and southern Italy: carbon dioxide earth degassing and seismogenesis, *Geophys. Res. Lett.,* **31**(7), n/a–n/a.

Cicerone, R.D., Ebel, J.E. & Britton, J., 2009. A systematic compilation of earthquake precursors, *Tectonophysics,* **476**(3–4), 371–396.

Dettmer, J., Dosso, S.E. & Holland, C.W., 2010. Trans-dimensional geoacoustic inversion, *J. acoust. Soc. Am.,* **128**(6), 3393–3405.

Gallagher, K., 2012. Transdimensional inverse thermal history modeling for quantitative thermochronology, *J. geophys. Res.,* **117**(B2), doi:10.1029/2011JB008825.

Gallagher, K., Bodin, T., Sambridge, M., Weiss, D., Kylander, M. & Large, D., 2011. Inference of abrupt changes in noisy geochemical records using transdimensional changepoint models, *Earth planet. Sci. Lett.,* **311**(1–2), 182–194.

Gao, S., Koker, T., Queen, O., Hartvigsen, T., Tsiligkaridis, T. & Zitnik, M., 2024. Units: a unified multi-task time series model, *Adv. Neural Inf. Process. Syst.,* **37,** 140 589–140 631.

Gazeaux, J. *et al.*, 2013. Detecting offsets in GPS time series: first results from the detection of offsets in GPS experiment, *J. Geophys. Res.: Solid Earth,* **118**(5), 2397–2407.

Gelman, A. & Rubin, D.B., 1996. Markov chain Monte Carlo methods in biostatistics, *Stat. Methods Med. Res.,* **5**(4), 339–355.

Granieri, D., Chiodini, G., Marzocchi, W. & Avino, R., 2003. Continuous monitoring of CO_2 soil diffuse degassing at Phlegraean Fields (Italy): influence of environmental and volcanic parameters, *Earth planet. Sci. Lett.,* **212**(1–2), 167–179.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika,* **82**(4), 711–732.

Gualandi, A., Nichele, C., Serpelloni, E., Chiaraluce, L., Anderlini, L., Latorre, D., Belardinelli, M.E. & Avouac, J.P., 2017. Aseismic deformation associated with an earthquake swarm in the northern Apennines (Italy): aseismic swarm related deformation, *Geophys. Res. Lett.,* **44**(15), 7706–7714.

Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using trans-dimensional trees, *Geophys. J. Int.,* **203**(2), 972–1000.

Holub, R.F. & Brady, B.T., 1981. The effect of stress on radon emanation from rock, *J. Geophys. Res.: Solid Earth,* **86**(B3), 1776–1784.

Hopcroft, P.O., Gallagher, K. & Pain, C.C., 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, *Geophys. J. Int.,* **171**(3), 1430–1439.

Hopcroft, P.O., Gallagher, K. & Pain, C.C., 2009. A Bayesian partition modelling approach to resolve spatial variability in climate records from borehole temperature inversion, *Geophys. J. Int.,* **178**(2), 651–666.

Hreinsdottir, S. & Bennett, R.A., 2009. Active aseismic creep on the Alto Tiberina low-angle normal fault, Italy, *Geology,* **37**(8), 683–686.

İnan, S. *et al.*, 2012. Seasonal variations in soil radon emanation: long-term continuous monitoring in light of seismicity, *Nat. Hazards,* **62**(2), 575–591.

Kanamori, H., 1977. Seismic and aseismic slip along subduction zones and their tectonic implications, *Island Arcs, Deep Sea Trenches and Back-Arc Basins,* pp. 163–174, American Geophysical Union (AGU).

Latorre, D., Mirabella, F., Chiaraluce, L., Trippetta, F. & Lomax, A., 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, *J. Geophys. Res.: Solid Earth,* **121**(11), 8113–8135.

Li, Y.G., Vidale, J.E., Aki, K., Xu, F. & Burdette, T., 1998. Evidence of shallow fault zone strengthening after the 1992 M7.5 Landers, California, Earthquake, *Science,* **279**(5348), 217–219.

Licciardi, A., Gallagher, K. & Clark, S.A., 2020. A Bayesian approach for thermal history reconstruction in basin modeling, *J. Geophys. Res.: Solid Earth,* **125**(7), e2020JB019384, doi:10.1029/2020JB019384.

Lucente, F.P., De Gori, P., Margheriti, L., Piccinini, D., Di Bona, M., Chiarabba, C. & Piana Agostinetti, N., 2010. Temporal variation of seismic velocity and anisotropy before the 2009 M_w 6.3 L'Aquila earthquake, Italy, *Geology,* **38**(11), 1015–1018.

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.,* **151**(3), 675–688.

Malinverno, A. & Briggs, V.A., 2004. Expanded uncertainty quantification in inverse problems: hierarchical Bayes and empirical Bayes, *Geophysics,* **69**(4), 1005–1016.

Mandler, E., Pintori, F., Gualandi, A., Anderlini, L., Serpelloni, E. & Belardinelli, M.E., 2021. Post-seismic deformation related to the 2016 Central Italy seismic sequence from GPS displacement time-series, *J. Geophys. Res.: Solid Earth,* **126**(9), e2021JB022200, doi:10.1029/2021JB022200.

Marchesini, B., Carminati, E., Aldega, L., Mirabella, F., Petrelli, M., Caracausi, A. & Barchi, M.R., 2022. Chemical interaction driven by deep fluids in the damage zone of a seismogenic carbonate fault, *J. Struct. Geol.,* **161,** 104 668, doi:10.1016/j.jsg.2022.104668.

Miller, S.A., Collettini, C., Chiaraluce, L., Cocco, M., Barchi, M. & Kaus, B.J.P., 2004. Aftershocks driven by a high-pressure CO_2 source at depth, *Nature,* **427**(6976), 724–727.

Mollo, S., Tuccimei, P., Heap, M.J., Vinciguerra, S., Soligo, M., Castelluccio, M., Scarlato, P. & Dingwell, D.B., 2011. Increase in radon emission due to rock failure: an experimental study: RADON EMISSIONS FROM ROCKS, *Geophys. Res. Lett.,* **38**(14), n/a–n/a.

Moretti, M., De Gori, P. & Chiarabba, C., 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, *Geophys. J. Int.,* **176**(3), 833–846.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.: Solid Earth,* **100**(B7), 12 431–12 447.

Piana Agostinetti, N. & Bodin, T., 2018. Flexible coupling in joint inversions: a Bayesian structure decoupling algorithm, *J. Geophys. Res.: Solid Earth,* **123**(10), 8798–8826.

Piana Agostinetti, N., Giacomuzzi, G. & Malinverno, A., 2015. Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **201**(3), 1598–1617.

Piana Agostinetti, N., Giacomuzzi, G. & Chiarabba, C., 2017. Seismic swarms and diffuse fracturing within Triassic evaporites fed by deep degassing along the low-angle Alto Tiberina normal fault (central Apennines, Italy), *J. Geophys. Res.: Solid Earth,* **122**(1), 308–331.

Piersanti, A., Cannelli, V. & Galli, G., 2015. Long term continuous radon monitoring in a seismically active area, *Ann. Geophys.,* **58,** 16, doi:10.4401/ag-6735.

Piersanti, A., Cannelli, V. & Galli, G., 2016. The Pollino 2012 seismic sequence: clues from continuous radon monitoring, *Solid Earth,* **7**(5), 1303–1316.

Pinault, J.L. & Baubron, J.C., 1996. Signal processing of soil gas radon, atmospheric pressure, moisture, and soil temperature data: a new approach for radon concentration modeling, *J. Geophys. Res.: Solid Earth,* **101**(B2), 3157–3171.

Poggiali, G., Chiaraluce, L., Di Stefano, R. & Piana Agostinetti, N., 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), *Geophys. J. Int.,* **217**(2), 1217–1231.

Rinaldi, A.P., Vandemeulebrouck, J., Todesco, M. & Viveiros, F., 2012. Effects of atmospheric conditions on surface diffuse degassing: atmospheric conditions and diffuse degassing, *J. Geophys. Res.: Solid Earth,* **117**(B11), n/a–n/a.

Rogie, J.D., Kerrick, D.M., Chiodini, G. & Frondini, F., 2000. Flux measurements of nonvolcanic co2 emission from some vents in central italy, *J. Geophys. Res.: Solid Earth,* **105**(B4), 8435–8445.

Roy, C. & Romanowicz, B.A., 2017. On the implications of A Priori constraints in transdimensional Bayesian inversion for continental lithospheric layering: constraints in Bayesian inversion, *J. Geophys. Res.: Solid Earth,* **122**(12), 10 118–10 131.

Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems: Monte Carlo inversion, *Rev. Geophys.,* **40**(3), 3–1-3-29.

Scales, J.A. & Snieder, R., 1997. To Bayes or not to Bayes?, *Geophysics,* **62**(4), 1045–1046.

Schaff David, P. & Beroza Gregory, C., 2004. Coseismic and postseismic velocity changes measured by repeating earthquakes, *J. Geophys. Res.: Solid Earth,* **109**(B10), doi:10.1029/2004JB003011.

Scholz, H.C., 1973. Earthquake prediction: a physical basis, *Science,* **181,** 803–809.

Serpelloni, E., Casula, G., Galvani, A., Anzidei, M. & Baldi, P., 2006. Data analysis of permanent GPS networks in Italy and surrounding regions: application of a distributed processing approach, *Ann. Geophys.,* **49,** 32.

Serpelloni, E., Pintori, F., Gualandi, A., Scoccimarro, E., Cavaliere, A., Anderlini, L., Belardinelli, M.E. & Todesco, M., 2018. Hydrologically induced Karst deformation: insights from GPS measurements in the Adria-Eurasia plate boundary zone, *J. Geophys. Res.: Solid Earth,* **123**(5), 4413–4430.

Siino, M., Scudero, S., Cannelli, V., Piersanti, A. & D'Alessandro, A., 2019. Multiple seasonality in soil radon time series, *Sci. Rep.,* **9**(1), 8610, doi:10.1038/s41598-019-44875-z.

Silverii, F., D'Agostino, N., Métois, M., Fiorillo, F. & Ventafridda, G., 2016. Transient deformation of Karst aquifers due to seasonal and multiyear groundwater variations observed by GPS in southern Apennines (Italy), *J. Geophys. Res.: Solid Earth,* **121**(11), 8315–8337.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.,* **20**(2), 219–232.

Thomas, D., 1988. Geochemical precursors to seismic activity, *Pure appl. Geophys.,* **126**(2–4), 241–266.

Tommasone Pascale, F., Carbone, P., De Francesco, S., Cuoco, E. & Tedesco, D., 2015. Rainstorm-induced soil 222Rn concentration spikes observed in Southern Italy, *Environ. Earth Sci.,* **73**(12), 8177–8187.

Vadacca, L., Casarotti, E., Chiaraluce, L. & Cocco, M., 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), 1537–1549.

Valoroso, L., Chiaraluce, L., Di Stefano, R. & Monachesi, G., 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, *J. Geophys. Res.: Solid Earth,* **122**(12), 10 220–10 240.

Viveiros, F., Ferreira, T., Cabral Vieira, J., Silva, C. & Gaspar, J., 2008. Environmental influences on soil CO_2 degassing at Furnas and Fogo volcanoes (São Miguel Island, Azores archipelago), *J. Volc. Geotherm. Res.,* **177**(4), 883–893.

Wadati, K. & Oki, S., 1933. On the Travel Time of Earthquake Waves. (Part II), *J. Meteorol. Soc. Japan Ser. II,* **11**(1), 14–28.

APPENDIX A: OBSERVATIONS ON GNSS MODELS ENSEMBLE

In Fig. A1, 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 semi-annual 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 yr. 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 the beginning of 2015, visible in the E–W component of Fig. A1. This feature is also evident in Fig. A2, where the azimuth distribution of GNSS stations is represented as histograms in time. Clear maxima are observed in 2015, indicating that a notable 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 analysed time-series. Looking at the multistation features instead of single station models, partially avoids the risk of misinterpreting site-dependent signals as tectonic motion.

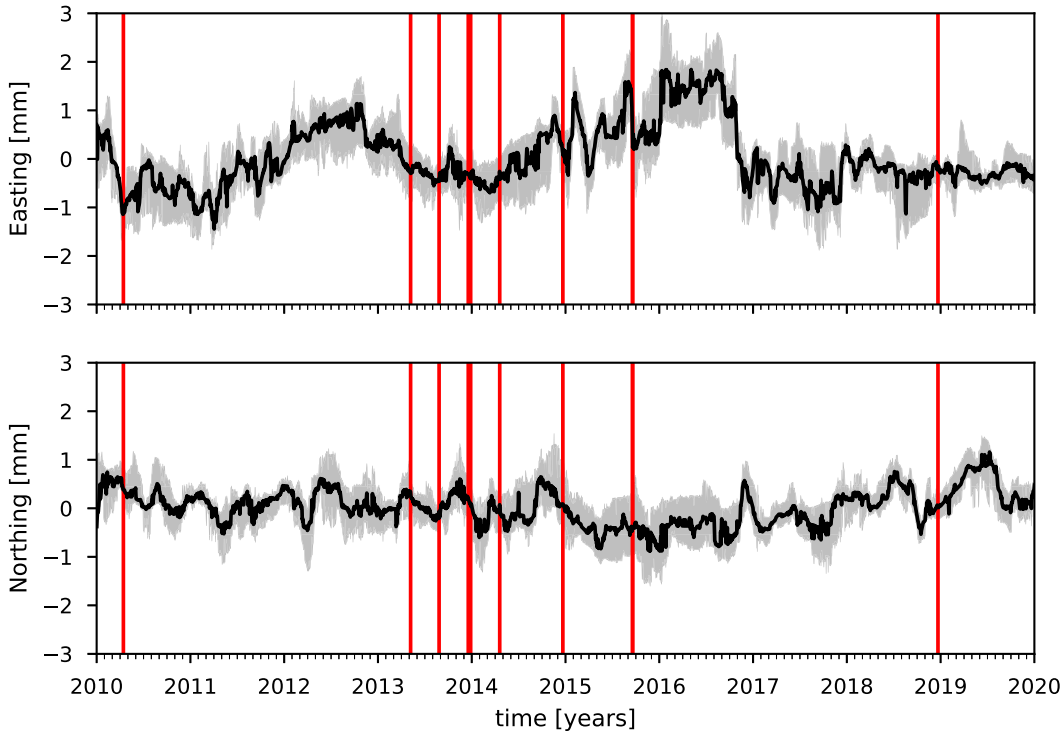

Figure A1. Median of the posterior mean models from all GNSS stations. Models are detrended in post-processing and do not include the estimated seasonal contributions. Grey shade represents interquartile range. Top: E–W component. Bottom: N–S component. Red lines represent the occurrence of earthquakes $M \geq 3.5$ inside the study area.

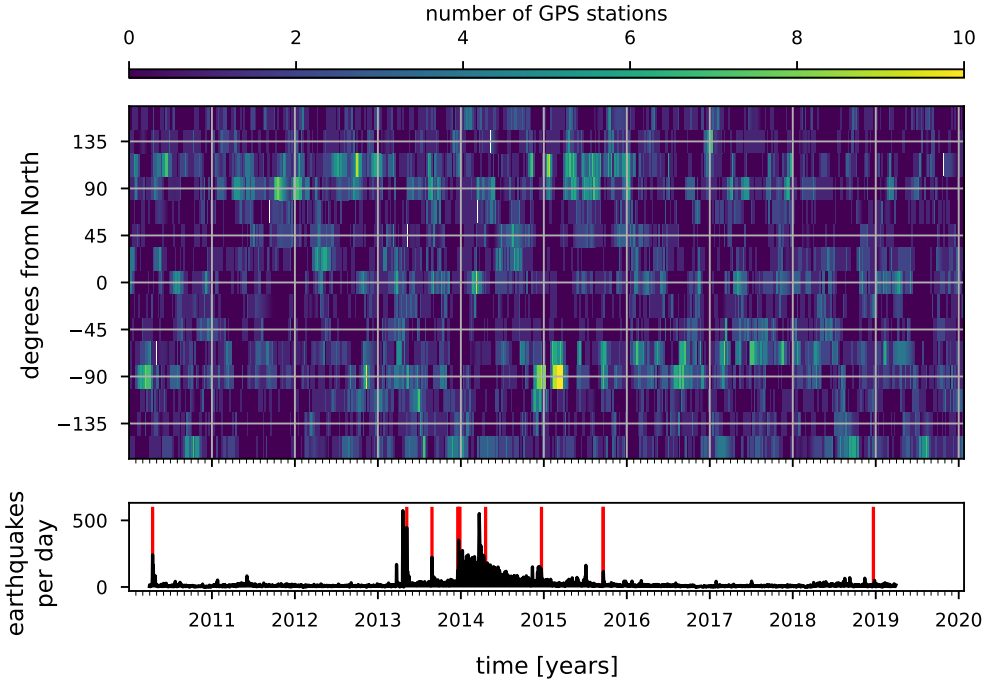

Figure A2. Top: Distribution of the direction of movement from posterior models as histograms in time; all GNSS stations inside the study area are considered. It is a representation of how many stations are moving in the same direction at the same time. Bottom: seismicity inside the study area. Red lines represent the occurrence of earthquakes $M \geq 3.5$.