

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

Geofisica

Ciclo XXVII

Settore Concorsuale di afferenza:  
Settore Scientifico disciplinare:

04/A4  
GEO/10

---

Spatial and temporal characterisation  
of ground deformation recorded by  
geodetic techniques

---

Presentata da:

Adriano Gualandi

Coordinatore Dottorato  
Prof. Michele Dragoni

Relatore  
Prof.ssa Maria Elina Belardinelli

Esame finale anno 2015



# Contents

<b>Acknowledgement</b>	<b>3</b>
<b>Introduction</b>	<b>6</b>
<b>1 PCA and the 2009 L'Aquila earthquake</b>	<b>8</b>
1.1 Principal Component Analysis . . . . .	9
1.1.1 Mathematical interpretation . . . . .	9
1.1.2 Geometrical interpretation . . . . .	11
1.1.3 PCA with missing data . . . . .	13
1.2 PCA on GPS time series: the 2009 L'Aquila earthquake . . . .	14
1.2.1 Tectonic framework . . . . .	15
1.2.2 Data . . . . .	17
1.2.3 Results . . . . .	19
1.2.4 Discussion, problems, and limitations . . . . .	25
<b>2 Independent Component Analysis and variational Bayesian approach</b>	<b>29</b>
2.1 Statement of the problem . . . . .	30
2.2 What is independence and how to measure it . . . . .	32
2.3 The variational bayesian ICA . . . . .	35
2.3.1 Variational approximation . . . . .	36
2.3.2 Variational approximation for the ICA problem . . . .	38
2.3.3 How to select the number of components . . . . .	42
2.4 Summary . . . . .	43
<b>3 Test on synthetic data</b>	<b>44</b>
3.1 Data generation . . . . .	44
3.1.1 Description of the sources . . . . .	45
3.1.2 Description of the simulated network of sensors . . . .	51
3.1.3 Logic tree of the synthetic data . . . . .	54
3.2 Results . . . . .	55

3.2.1	Low tectonic rate: $\dot{s}^{lin} = 2 \frac{mm}{yr}$ . . . . .	60
3.2.2	High tectonic rate: $60 \frac{mm}{yr}$ . . . . .	76
3.2.3	Volcanic source . . . . .	78
3.3	Conclusive remarks . . . . .	79
<b>4</b>	<b>Real data applications</b>	<b>81</b>
4.1	The L'Aquila earthquake (2009), central Italy . . . . .	81
4.1.1	Results . . . . .	82
4.1.2	Discussion . . . . .	86
4.2	The Emilia seismic sequence (2012), northern Italy . . . . .	88
4.2.1	Introduction . . . . .	88
4.2.2	Data . . . . .	89
4.2.3	Results . . . . .	91
4.2.4	Discussion . . . . .	98
4.2.5	Conclusive remarks . . . . .	108
4.3	The 2006 Slow Slip Event in Mexico . . . . .	110
4.3.1	Introduction . . . . .	110
4.3.2	Data . . . . .	112
4.3.3	Results . . . . .	113
4.3.4	Discussion . . . . .	116
4.3.5	Conclusive remarks . . . . .	120
	<b>Conclusions</b>	<b>122</b>
	<b>Bibliography</b>	<b>124</b>
	<b>A Credibility intervals for the post-seismic decay constant</b>	<b>138</b>
	<b>B vbICA and FastICA</b>	<b>140</b>
	<b>C Time series</b>	<b>141</b>
C.1	Aquila . . . . .	141
C.1.1	Seasonal signals + co-seismic offsets + post-seismic decay	141
C.1.2	Co-seismic offsets + post-seismic decay . . . . .	141
C.2	Emilia . . . . .	146
C.2.1	Long-term trend + seasonal signals + co-seismic offsets + post-seismic decay . . . . .	146
C.2.2	Seasonal signals + co-seismic offsets + post-seismic decay	146
C.2.3	Co-seismic offsets + post-seismic decay . . . . .	146
C.3	Mexico . . . . .	156
	<b>D Seismic moment estimation from geodetic displacements</b>	<b>158</b>

# Acknowledgement

I took the first steps into the research world when I met for the first time Enrico, in 2009. Since that day he has effectively been an advisor for me, and I want to thank him for all the energy that he spent and everyday spends for making possible the analysis of GPS data in Bologna.

I want to thank my actual advisor Maria Elina. She has always been willing and helpful, and she always left me enough freedom in pursuing the best path for me.

Thank you also to Jean-Philippe, for having given me the possibility to visit for a second time the Tectonic Observatory labs. He encouraged me in my research and he introduced me to wonderful people at Caltech, from both a scientific and a human point of view. In particular, thanks for the good time to Bruno, Christoph, Mortaza, François, Giovanni, Annamaria, Frederic, and Heather.

A heartfelt thanks also to Robert, who received me in his house in Pasadena as a true friend.

Another person that I want to thank is Hugo. I really appreciate the work he has done to maintain the PCAIM code, and the way in which he welcomed me with open arms in Grenoble and in his island, Corsica. This is priceless.

Thanks also to Nathalie, for her openness and for having given me the possibility to work on the Guerrero data.

Thanks to all the people at INGV Bologna who shared ideas and meals with me: Irene, Letizia, Brunella, Pablo, Paola, Barbara, Julie, Simone, Adriano. Thanks also to Anna for the work she did and does in order to simplify the bureaucratic life to the PhD students.

I want to thank also my kung-fu master Daniele, for having helped me in training the balance of both the body and the mind.

A sincere thanks to all my friends who supported me during the difficult times and shared with me the good ones. Thanks to Pic, Riky, Ghera, Poz, Bos, and Caps for the imaginary adventures lived together, and thanks to Laura for all the times she allowed us to occupy her house. Thanks also to

Davide for his honest friendship.

Last but not least, the most important thanks go to my family, from my aunts and uncles to my cousins. To my brother Mario for having taught me to appreciate the differences in people. To my mom for having taught me to avoid envy and be happy for other people's happiness. To my dad for having taught me with his example how to discern what is the right thing to do.

# Introduction

With the improvements and new developments of space geodetic techniques, the number of data for the study of geophysical processes grew a lot in the last decades. In particular, three fundamental aspects improved: the spatial coverage, the temporal coverage, and the accuracy of the measurements. For example, the continuous Global Navigation Satellite System (GNSS) networks as well as the Synthetic Aperture Radar (SAR) images are now standard techniques for the study of crustal deformation, and the increase in the number of available GNSS data in the recent years (about from 2005) has been very rapid, particularly in the Euro-Mediterranean area. Most of the time, in Earth science we try to achieve more knowledge about a process just collecting more data about it. This redundant quantity of information can be very useful, but there is the risk to end up with a huge messy dataset to be untangled. All these data measure the evolution of a certain observable with respect to time. This means that the data are recorded in time series format, and it becomes of fundamental importance to properly use some multivariate statistical technique in order to study them. Much effort has been profused in reducing the dimensionality of vast dataset, and the most widely used technique is probably the Principal Component Analysis (PCA). In the first Chapter of this thesis I show how this method can be used to study the co- and post-seismic crustal deformation related to a mid-size earthquake (the  $M_W$  6.3, 2009 L'Aquila earthquake), following the approach proposed by Kositsky and Avouac (2010).

Nevertheless, one problem remains open. Often, more than one process contribute to the generation of the recorded observable. In the signal processing environment, each process is referred to as a *source signal*, and the observed time series are the *sensor signals*. This means that the sensor outputs consist of a mix of the source signals. One possible approach to explain the data is to assume some source model, and find out the parameters of the model that best fit the data. This forward approach is constraining the sources to have a particular shape, a priori assigned. The PCA technique is excellent in extracting the information contained in the data, and in concen-

trating it in a few number of principal components (PCs). By the way, the PCs do not have any physical meaning, and might represent only a combination of the true physical sources beneath the phenomenon observed. To make possible the interpretation of the components, it becomes of fundamental interest to find out what are the original source signals, making the fewest (and the most reasonable) number of assumptions as possible on the sources. This is the goal of the Blind Source Separation (BSS) problem. One of the most popular approaches to solve the BSS is the Independent Component Analysis (ICA), that I will describe in Chapter 2.

The goal of my PhD thesis is thus to detect signals in geodetic time series, with the aim of allowing an easy interpretation of the multivariate analysis results. Before applying the ICA technique on real data, I perform several tests in order to validate the method and verify its applicability to the geodetic available data. In Chapter 3 I compare the PCA and ICA performance on synthetic data that simulate a continuous GNSS network recording a long-term velocity (tectonic loading), an annual sinusoidal variation (seasonal signal), a step function at a certain epoch (co-seismic displacement) and a logarithmic decay (post-seismic displacement). For one particular case, I also simulate the position time series generated by an inflating and deflating magma chamber via a Mogi source acting at the same time of an earthquake. After this corroboration, in Chapter 4 I apply the ICA to real case studies. In particular, I use continuous GPS position time series in order to characterise the spatio-temporal evolution of the surface displacement relative to the 2009 L'Aquila earthquake ( $M_W$  6.3, central Italy), the 2012 Emilia seismic sequence ( $M_W$  6.1, northern Italy), and the 2006 Guerrero Slow Slip Event (Mexico).

# Chapter 1

## PCA and the 2009 L'Aquila earthquake

*It's just a matter of point of view.*

Let us imagine that we want to study a physical system, and unearth its inner properties. Depending on what are the properties that (a priori) we decide to investigate, we set up some instruments to record the value of the property under exam (or some other observable from which we can deduce the one of interest). Let us suppose to have  $M$  sensors that record at time  $t$  this observable. Let us organize this set of observations in a data vector, that we call  $\mathbf{x}^t$ . This means that, at time  $t$ , we have  $M$  variables, and, if we record the data at  $T$  different epochs, we end up with  $T$  points, each of dimension  $M$ . If  $M$  is greater than 3, we are not able to visualize the data. This means that it is also very difficult (if not impossible) for our brain to perform its favourite task: classify and discern patterns.

It follows that, if we want to give an interpretation to our data, it is necessary to *extract* the *relevant* information from them. This is exactly what PCA does. It performs a transformation such that very few components (this is the extraction part) explain most of the variance of the dataset (i.e., the relevant part). In doing so, the underlying assumption is that the variance is the appropriate measure to describe the carried information. In the following of the Chapter, firstly I describe the mathematical theory behind PCA. After this introduction, I move on applying it to the analysis of a real case, the GPS data collected during the 2009 L'Aquila (central Italy) earthquake.

# 1.1 Principal Component Analysis

## 1.1.1 Mathematical interpretation

From a mathematical point of view, the data can be organised in a matrix, that we call  $\mathbf{X}$ . The first step is to individualize the disposition of the variables, that might be by columns or by rows. From now on, I suppose that each row corresponds to the recorded values of the observable from a certain sensor. This means that  $\mathbf{X}$  is a  $M \times T$  matrix, where  $M$  is the number of variables. All the cases treated in this thesis consider real valued data, and finite dimensional spaces.

Let us consider each row  $\mathbf{x}_j$ ,  $j = 1, \dots, M$ , as the realization of a stochastic process, and let us suppose that it is a non-degenerate Gaussian random vector of dimension  $T$  with independent and identically distributed (i.i.d.) elements. Consequently, given the recorded data (i.e., the sample), we can calculate the sample moments of the distribution of the variables. In other words, we have  $M$  samples that describe the  $M$  variables, and we can describe the data matrix as a vector:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_M \end{pmatrix}$$

Then, we can compute the covariance matrix of this vector, which tells us what is the correlation among the different variables:

$$\begin{aligned} \mathbf{C}_\mathbf{X} &= E [(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T] & (1.1) \\ &= \begin{pmatrix} E[(\mathbf{x}_1 - E[\mathbf{x}_1])(\mathbf{x}_1 - E[\mathbf{x}_1])^T] & \cdots & E[(\mathbf{x}_1 - E[\mathbf{x}_1])(\mathbf{x}_M - E[\mathbf{x}_M])^T] \\ \vdots & \ddots & \vdots \\ E[(\mathbf{x}_M - E[\mathbf{x}_M])(\mathbf{x}_1 - E[\mathbf{x}_1])^T] & \cdots & E[(\mathbf{x}_M - E[\mathbf{x}_M])(\mathbf{x}_M - E[\mathbf{x}_M])^T] \end{pmatrix} \\ &= \frac{1}{T}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T \end{aligned}$$

where  $\cdot^T$  indicates the transpose, and  $\boldsymbol{\mu}$  is the expected value of the vector  $\mathbf{X}$ , i.e.  $\mu_j = E[\mathbf{x}_j]$ , with  $j = 1, \dots, M$ .

Given the samples for the variables  $\mathbf{x}_j$ , it is possible to compute the sample covariance matrix, where the generic  $C^{ij}$  element is:

$$E[(\mathbf{x}_i - \mu_i)(\mathbf{x}_j - \mu_j)^T] = \frac{1}{T} \sum_{t=1}^T (x_{it} - \mu_i)(x_{jt} - \mu_j)^T \quad (1.2)$$

To simplify the notation, let us assume that the data matrix has already been centered, that is  $\mathbf{X} \rightarrow \mathbf{X} - \boldsymbol{\mu}$ .

Being a covariance matrix,  $\mathbf{C}_{\mathbf{X}}$  is positive-semidefinite and symmetric. For the symmetry property, the spectral theorem assures that there exists an orthonormal basis of  $\mathbb{R}^M$  consisting of eigenvectors of  $\mathbf{C}_{\mathbf{X}}$ , that is:

$$\frac{1}{T}\mathbf{X}\mathbf{X}^T = \mathbf{C}_{\mathbf{X}} = \mathbf{U}\mathbf{D}\mathbf{U}^T \quad (1.3)$$

where  $\mathbf{D}$  is diagonal and  $\mathbf{U}$  is an orthogonal matrix. Moreover, the positive-semidefinite property ensures that the diagonal values of  $\mathbf{D}$  are all non-negative. We can thus write:

$$\frac{1}{T}\mathbf{X}\mathbf{X}^T = \mathbf{U}\left(\frac{1}{T}\boldsymbol{\Sigma}^2\right)\mathbf{U}^T = \mathbf{U}\boldsymbol{\Delta}^2\mathbf{U}^T \quad (1.4)$$

where  $\frac{1}{T}\boldsymbol{\Sigma}^2 = \boldsymbol{\Delta}^2$  is a diagonal matrix containing the eigenvalues of  $\mathbf{C}_{\mathbf{X}}$ , and the columns  $\mathbf{u}^j$  of matrix  $\mathbf{U}$ , with  $j = 1, \dots, M$ , are one-by-one orthogonal and are the eigenvectors of the covariance matrix  $\mathbf{C}_{\mathbf{X}}$ , i.e.  $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ .

We can finally write:

$$\mathbf{X} = \mathbf{U}\mathbf{Y} \quad (1.5)$$

and, from the orthogonality of  $\mathbf{U}$ :

$$\mathbf{U}^T\mathbf{X} = \mathbf{U}^T\mathbf{U}\mathbf{Y} = \mathbf{Y} \quad (1.6)$$

The eigenvectors of  $\mathbf{C}_{\mathbf{X}}$ , i.e. the columns of  $\mathbf{U}$  or, equivalently, the rows of  $\mathbf{U}^T$ , transform the data  $\mathbf{x}_j$ , into the Principal Components (PCs)  $\mathbf{y}_j$ , with  $j = 1, \dots, M$ . Thus,

**Definition 1.1:** A PCA transformation of a data matrix  $\mathbf{X}$  is defined by the pair  $\{\mathbf{U}, \boldsymbol{\Delta}^2\}$ , and defines an equivalence class of decompositions, for which the following properties hold:

- 1) the covariance factorizes into  $\mathbf{C}_{\mathbf{X}} = \mathbf{U}\boldsymbol{\Delta}^2\mathbf{U}^T$  where  $\boldsymbol{\Delta}^2$  is diagonal real positive and  $\mathbf{U}$  is full column rank  $M$ ;
- 2)  $\mathbf{U}$  is a matrix whose columns are orthogonal to each other (i.e.,  $\mathbf{U}^T\mathbf{U}$  is diagonal).

The representation of the data via the PCs is given by equation 1.6. It follows that, if we want to have a unique PCA decomposition, we have to impose two constraints. The following two are usually adopted, as well as in this thesis:

- 1) the columns of  $\mathbf{U}$  are unit norm;
- 2) the entries of  $\boldsymbol{\Delta}^2$  are sorted in decreasing order.

## 1.1.2 Geometrical interpretation

From constraint 1) of Section 1.1.1, we argue that the columns of  $\mathbf{U}$ , i.e. the eigenvectors of  $\mathbf{C}_{\mathbf{X}}$ , are an orthogonal basis of a new reference system. The projected data are expressed in the new basis via the following equation:

$$\mathbf{Y} = \mathbf{U}^{-1}\mathbf{X} = \mathbf{U}^T\mathbf{X} \quad (1.7)$$

and the covariance matrix of the data expressed in this system is:

$$\mathbf{C}_{\mathbf{Y}} = \frac{1}{T}\mathbf{Y}\mathbf{Y}^T = \frac{1}{T}\mathbf{U}^T\mathbf{X}(\mathbf{U}^T\mathbf{X})^T = \frac{1}{T}\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U} = \mathbf{U}^T\mathbf{C}_{\mathbf{X}}\mathbf{U} = \mathbf{U}^T\mathbf{U}\Delta^2\mathbf{U}^T\mathbf{U} = \Delta^2 \quad (1.8)$$

From this last equation, we see that the new variables  $\mathbf{y}_j$ , with  $j = 1, \dots, M$ , are uncorrelated. Adding to this the constraint number 2) of Section 1.1.1, we argue that the most of the variance of the data is explained if we project the dataset onto the direction of the first PC, i.e. the direction indicated by the first column of  $\mathbf{U}$ . Continuing the process, the second PC is the one that explains most of the variance of the dataset after all correlation with the first principal component has been subtracted from the points; and so on, until the last PC. In other words, the dataset is progressively better reproduced adding more and more PCs, and the last components have a lower weight in explaining the data if compared to the very first ones.

If we do not care about the dimensionality problem, and we use all the PCs, the orthogonal linear transformation described by a PCA is equivalent to a combination of a translation and a rotation. Indeed, we can summarize it as follows:

- 1) a translation  $-\boldsymbol{\mu}$  to center the variables;
- 2) a rotation  $\mathbf{U}^T$  of the axis.

To visualize these transformations, let us assume to have a small number of sensors ( $M \leq 3$ ), and look at Figure 1.1. In this example, the original data are given by the observation through a camera of the position of an ideal harmonic oscillator with no friction at  $T = 1000$  different epochs. In other words, the data are the projection of the position of the mass of the harmonic oscillator onto the plane perpendicular to the line of sight of the camera. This means that the camera is recording two different variables. Since the camera is not optimally oriented, the records at different times do not belong to the same axis. Moreover, the addition of some Gaussian noise does not allow to the data to be perfectly aligned. The resulting dataset is a composition of  $M = 2$  time series of length  $T = 1000$ .

To perform a PCA, firstly we have to center the dataset. This means that we have to set the origin of the reference system in the centre of the cloud of points in the data space. Figure 1.1a shows the original values recorded (red dots) and the values of the centered data (blue dots). The black arrows,  $\mathbf{u}^1$  and  $\mathbf{u}^2$ , show the two eigenvectors of the data covariance matrix  $\mathbf{C}_{\mathbf{X}}$ . Figure 1.1b illustrates the original data points projected onto the new reference frame.

A further step that is usually performed is the whitening of the data. This is a scale transformation, by a factor  $\sigma_j^{-1}$ , with  $j = 1, \dots, M$ , that is done in order to have on each of them the same variance (see Figure 1.1c and the next Subsection **Singular Value Decomposition**).

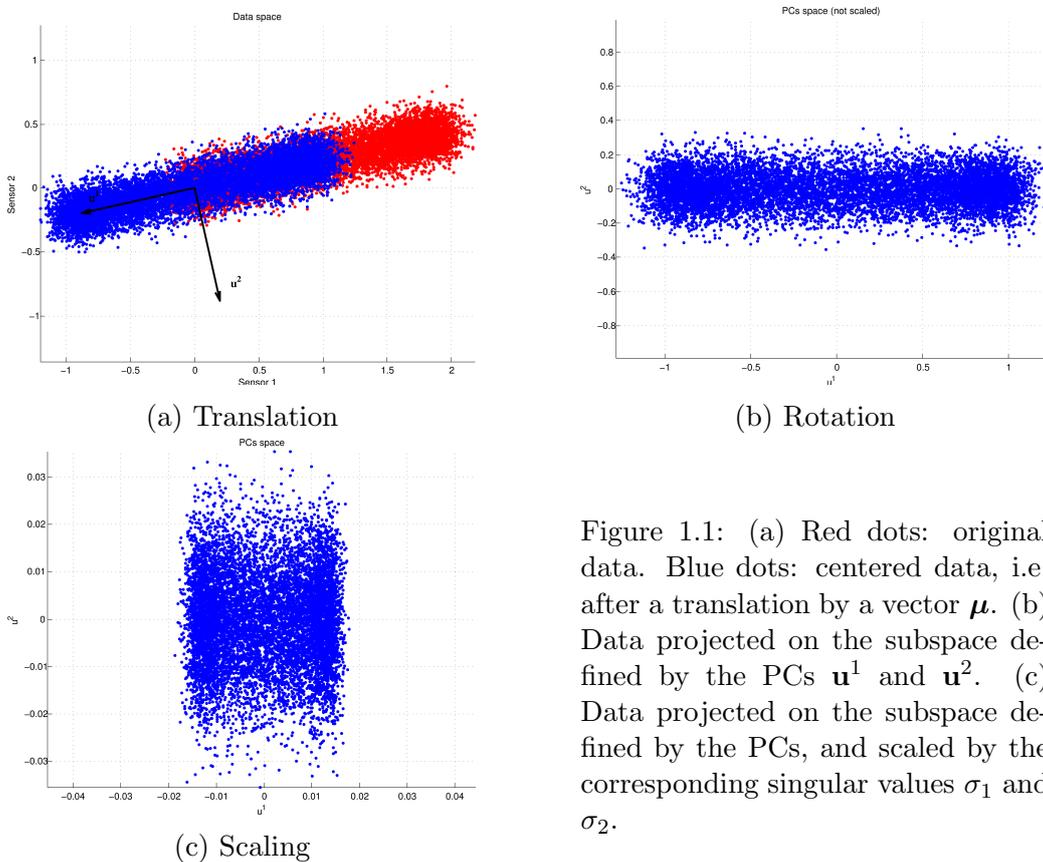


Figure 1.1: (a) Red dots: original data. Blue dots: centered data, i.e. after a translation by a vector  $\boldsymbol{\mu}$ . (b) Data projected on the subspace defined by the PCs  $\mathbf{u}^1$  and  $\mathbf{u}^2$ . (c) Data projected on the subspace defined by the PCs, and scaled by the corresponding singular values  $\sigma_1$  and  $\sigma_2$ .

## Singular Value Decomposition

A very useful property is valid for any finite dimensional  $M \times T$  matrix. This property states that a matrix  $\mathbf{X}$  can be written as the product of an orthogonal matrix ( $\mathbf{U}$ ), a diagonal matrix ( $\boldsymbol{\Sigma}$ ), and another orthogonal matrix

(V):

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T \quad (1.9)$$

This decomposition is called Singular Value Decomposition (SVD). The diagonal values of  $\Sigma$  are the singular values of  $\mathbf{X}$ , while the columns of  $\mathbf{U}$  and  $\mathbf{V}$  are the eigenvectors of  $\mathbf{X}\mathbf{X}^T$  and  $\mathbf{X}^T\mathbf{X}$ , respectively. If we impose the same constraint of unit norm columns also on matrix  $\mathbf{V}$ , we can interpret the eigenvectors of  $\mathbf{X}^T\mathbf{X}$  as a new orthonormal basis for the variables  $\mathbf{x}^t$ , with  $t = 1, \dots, T$ . Note that, in doing so we invert the interpretation of “variables” and “observed sample”. This different point of view implies two different approaches: the T-mode and the S-mode PCA. In the rest of the thesis, I am coherent with the interpretation given in the introduction of this Chapter: the  $\mathbf{X}$  data matrix is composed by  $M$  variables, each described by a sample of  $T$  observations.

### 1.1.3 PCA with missing data

Before performing a PCA, it is necessary to address one additional, and important, issue: how to handle missing data. Indeed, the kind of data that I am going to analyse are not complete, i.e. I do not have exactly  $M \times T$  data, but some of them are not recorded. The fail in recording data can be due to various causes (e.g., from a malfunctioning instrumentation to a theft of the instrumentation itself) that I am not investigating. What I am interested in is to properly take into account the missing data in the analysis. The most studied case occurs when the data are missing completely at random (MCAR). Instead, GPS position time series often present data gaps of variable size, and such a gap may start at every time. There are several techniques to take into account the missing data. In this first Chapter I use the algorithm developed by Srebro and Jaakkola [2003], already implemented in the PCA-based Inversion Method (PCA-IM) code [Kositsky and Avouac, 2010]. This algorithm performs a low-rank approximation weighting the data, and giving a zero weight to the missing ones. In particular, the centering and decomposition steps are repeated iteratively until convergence. One important thing to point out is that the Srebro and Jaakkola [2003] approach can not be properly interpreted as a PCA, since the eigenvectors are not strictly orthogonal. Moreover, the centering stage can be very time consuming when the number of missing data increases. For this reason, I also apply the algorithm proposed by Bailey [2012], based on an Expectation-Maximization routine (EMPCA). This procedure takes into account the uncertainties of the data, as well as the Srebro and Jaakkola [2003] algorithm, but it also forces the orthogonality of the eigenvectors. Moreover,

the centering step is not performed iteratively, and this consistently speeds up the process.

## 1.2 PCA on GPS time series: the 2009 L'Aquila earthquake

The PCA technique has been widely used for the study of geophysical data in general [e.g. Jolliffe, 2002], and for geodetic data in particular [e.g. Dong et al., 2006, Kositsky and Avouac, 2010, Ji and Herring, 2013]. Here I use the PCAIM code [Kositsky and Avouac, 2010] in order to study the GPS position time-series relative to the  $M_W$  6.3, 2009 April 6 L'Aquila earthquake, and, following Perfettini et al. [2010], I show how the PCA approach allows to get some important insights on the mechanical properties of fault zones.

The approach can be synthetically described as follows. Given  $N_{stn}$  GPS stations, to each of them there are three time series associated: one describing the motion along the East-West direction, one along the North-South, and one along the vertical. The position time series can be represented as a  $M \times T$  matrix, where  $M = 3 \times N_{stn}$  and  $T$  is the number of recorded epochs. The application of the PCA technique to this matrix requires two steps: 1) centering the matrix, 2) decomposing the centered matrix. The PCAIM approach uses the Srebro and Jaakkola [2003] decomposition, ensuring the capability to handle missing data. The result of the decomposition can be written as the product of the three matrices  $U$ ,  $\Sigma$ , and  $V$  (see Section 1.1.2). I will refer to  $V$  as the matrix of the PCs or temporal functions. Under the hypothesis that only  $L$  components are used to decompose the data matrix, then  $V$  is a matrix of size  $T \times L$ , where each column  $V_i$  describes the temporal evolution of the relative  $i$ -th component. The  $U\Sigma$  matrix contains the spatial information relative to the dataset, i.e. the weights associated to each temporal function in order to reconstruct the original time series via a linear combination. In order to have a unique PCA decomposition,  $U$  and  $V$  are unitary matrix, and  $\Sigma$  contains the actual weights of the different PCs. Thus, the PCAIM approach is based on the following assumptions [Kositsky, 2011]:

- i) the dataset can be linearly decomposed (linearity)
- ii) a reduced number of components ( $L < M$ ) is necessary to explain the data (low-rank approximation)
- iii) each spatial component of surface displacements (i.e., each column of  $U$ ) can be inverted for a given slip component (invertibility)

- iv) the slip components can be linearly combined to get the fault slip distribution (linearity)

The inversion step is performed using elastic dislocation theory and a least-squares procedure to invert for the slip distribution.

Differently from previous studies on the L'Aquila sequence, where an exponential temporal function has been used to fit the data, the PCAIM approach, under the hypothesis that afterslip is the main cause of the observed post-seismic deformation field, allows me to retrieve the temporal evolution of the slip on the fault with no *a priori* constraints. The multivariate statistical approach allows me also to account for the initial post-seismic deformation in estimating the co-seismic displacements, in a consistent manner for the whole GPS network. In the next Section I will describe the tectonic framework relative to the L'Aquila region. Then, I will describe the adopted processing to get the position time series analysed with PCAIM. After that, I will present the results of the analysis. Finally, I will discuss the results, and I will show some limitations of the PCA approach for the study of geodetic data. Most of the material presented in this Section refers to the continuation of my Master thesis and to the work published on *Geophysical Journal International* [Gualandi et al., 2014], to which the interested reader is referred for more details.

### 1.2.1 Tectonic framework

The  $M_W$  6.3 mainshock of the 2009 central Italy earthquake sequence occurred at 01:32 UTC of April 6 near the city of L'Aquila, causing 308 casualties and damages in a wide area of the Abruzzi region of central Italy. The 2009 seismic sequence included five events with  $5 < M_W < 6$  and activated a NW-SE trending seismogenic volume expanding for nearly 50 km along the central Apennines (see Figure 1.2). Here, both geological and geodetic observations indicate a regional NE-SW oriented extension, occurring at background rates of  $\sim 2\text{-}3$  mm/yr [Serpelloni et al., 2006, Faure-Walker et al., 2010, D'Agostino et al., 2011, Galvani et al., 2012]. This extension is accommodated across a complex belt of subparallel NW-SE oriented normal faults, with lengths ranging from few to 15-20 km, mostly steeply dipping towards the SW [Barchi et al., 2000, Galadini and Galli, 2000, Valensise and Pantosti, 2001, Boncio et al., 2004]. Earthquake fault plane solutions [Pondrelli et al., 2010, Scognamiglio et al., 2010, Herrmann et al., 2011], space geodetic observations [Anzidei et al., 2009, Atzori et al., 2009, Walters et al., 2009, Cheloni et al., 2010, Devoti et al., 2012, Serpelloni et al., 2012] and aftershock distribution [Chiarabba et al., 2009, Chiaraluce et al., 2011] have

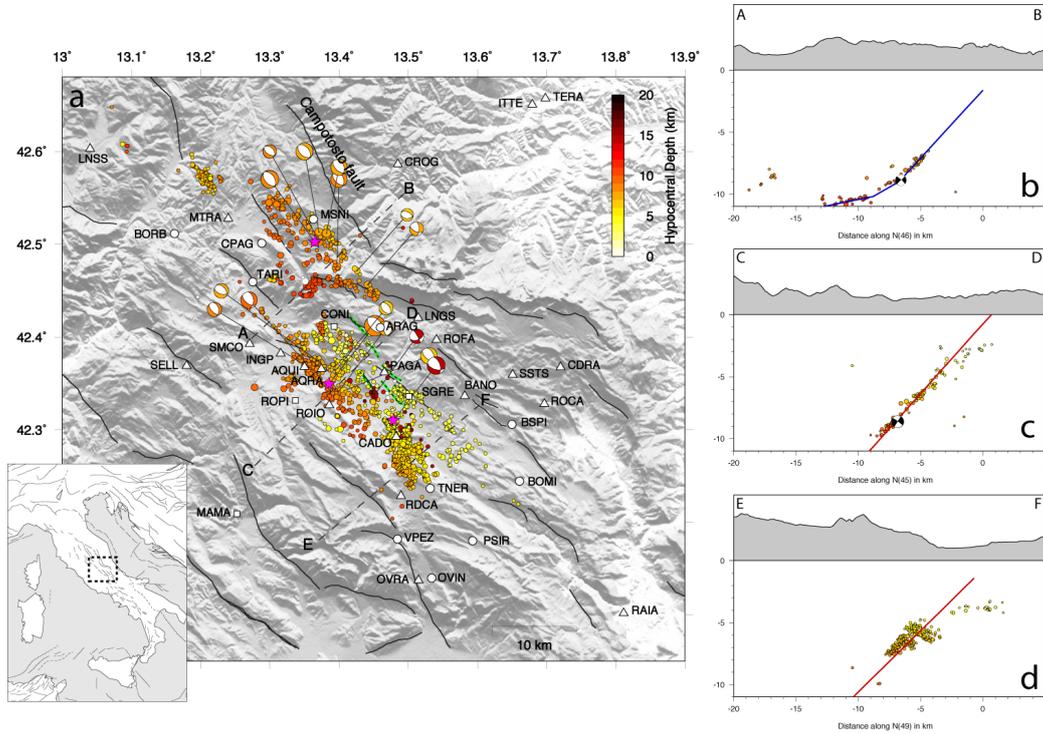


Figure 1.2: (a) Map of the region involved in the seismic sequence. Coloured circles are the aftershocks from Chiaraluce et al. [2011], beach balls are the focal mechanisms from Pondrelli et al. [2010] and purple stars indicate the mainshock (April 6, lon 13.39°, lat 42.35°, depth 8.6 km) and the two most relevant aftershocks (April 7, lon 13.48°, lat 42.31°, depth 14.2 km; April 9, lon 13.36°, lat 42.50°, depth 8.9 km). GPS stations are also shown. White circles/squares indicate those stations that have been used only for the co-seismic/post-seismic analysis. White triangles indicate the stations used in both co-seismic and post-seismic study. Black lines are the active faults from Boncio et al. [2004] and Galli et al. [2008]. Green dashed lines are the observed surface ruptures after the mainshock from Boncio et al. [2010]. Black dashed lines are the sections reported in panels (b), (c) and (d). (b) Section AB of panel (a). The focal mechanism corresponds to the April 9 aftershock. The blue line is the section of the geometry C2 (see table 1 in Gualandi et al. [2014]). (c)(d) Sections CD and EF of panel (a). The focal mechanism (c) represents the mainshock. Red lines are the sections of geometry P4 (see table 1 in Gualandi et al. [2014]).

consistently shown that the largest shocks of the sequence ruptured a set of shallow, SW-dipping, normal faults, running roughly parallel to the NW-SE trending Apennines belt, which are intermediate between the eastern and western fault systems described in Galadini and Galli [2000]. The April 6  $M_W$  6.3 mainshock was followed, on April 7, by a  $M_W$  5.4 aftershock, located

about 15 km SE from the mainshock and at a hypocentral depth of  $\sim 15$  km, and, on April 9, by a  $M_W$  5.2 aftershock (at hypocentral depth of 9 km), occurred on a secondary fault segment, the Campotosto fault, located NW of the mainshock [Chiaraluce et al., 2011]. The Campotosto fault was also later struck, on June 22, by a  $M_W$  4.4 aftershock, at a hypocentral depth  $>10$  km. Precise relocations of aftershocks [Chiaraluce et al., 2011, Valoroso et al., 2013] provide evidence that the L’Aquila sequence evolved on a complex fault system, characterized by major and secondary fault planes, where the two major faults are the Paganica fault and, northwest of it, the Campotosto fault, forming a NW-trending en echelon system 50 km long (see Figure 1.2).

Space geodetic observations (GPS and InSAR) recorded significant post-seismic deformation over a wide area, extending  $>30$  km from the epicentre, the months after the mainshock [Lanari et al., 2010, D’Agostino et al., 2012, Devoti et al., 2012]. Cheloni et al. [2010] and D’Agostino et al. [2012] modelled the post-seismic surface displacements as due to afterslip on the Paganica fault plane, assuming that the post-seismic deformation of GPS observations evolves with time following an exponential decay function, as in Lanari et al. [2010] and Devoti et al. [2012]. Cheloni et al. [2010] and D’Agostino et al. [2012] found that afterslip has not fully propagated to the surface but mostly occurred at a shallow depth, between the main co-seismic patch and the surface, with the post-seismic slip that appears to encircle the main co-seismically slipping portion of the Paganica fault.

### 1.2.2 Data

I use daily position time series obtained from the analysis of the GPS data collected at both continuous (cGPS) and survey-mode (sGPS) networks (see Devoti et al. [2012] and Serpelloni et al. [2012] for a more comprehensive description of the networks and quality of data used). Table 1.1 shows the occupation history of the GPS stations used in this work.

The raw GPS phase data have been processed following the same three-steps procedure described in Serpelloni et al. [2012], to which I refer for a more detailed description of the data analysis strategy used to derive the position time-series. It is known that GPS time-series contain various systematic errors (from network common to site-dependent) and random errors [e.g. Dong et al., 2006, Langbein, 2008, Williams, 2003]. In regional network analysis, the so-called common mode error (CME) is one of the major spatially correlated error sources, which is mitigated through a technique commonly referred to as regional filtering. The “stacking” approach [Wdowinski et al., 1997], commonly used in regional analysis, assumes that the CME is spatially

uniform, which is a good approximation for small networks, but as the spatial extent increases, the assumption is no longer valid [e.g. Márquez-Azúa and DeMets, 2003]. Here I use a PCA-based approach [Dong et al., 2006], which decomposes the network time-series into a set of temporally varying modes (or PCs). Each mode consists of a common temporal function and related different spatial response, providing a mathematical framework to perform spatio-temporal filtering. This approach allows me to remove the assumption of spatially uniform distribution, letting the displacement data reveal the spatial distribution of the CME. The PCA method is applied on the residual time-series (de-trended and cleaned for instrumental offsets, seasonal signals and outliers) of 650 cGPS stations distributed over a wider region, encompassing the Euro-Mediterranean area [Serpelloni et al., 2013], while excluding stations in central Italy, likely affected by post-seismic deformation. I assume the first and second PCs to be representative of the CME, accounting for more than 50% of the total variance. By filtering the time-series from the estimated CME I obtain 35% and 30% reduction of the Weighted Root Mean Squares (WRMS) values in the horizontal and vertical components, respectively, with a significant gain in the signal-to-noise ratio (SNR) of the post-seismic displacements transient. This is particularly important for studying smaller amplitude crustal deformation related to moderate magnitude earthquakes.

The matrix given as input for the analysis with PCAIM contains the CME filtered, de-trended time-series, which are obtained after removing a constant velocity term, a seasonal (annual+semi-annual) term and any non-tectonic offset due to changes in the stations equipment configurations. For those stations for which no reliable pre-earthquake inter-seismic velocities can be determined (e.g. because only post-earthquake data are available or only a short time-interval before the earthquake is available) I constrained their velocities to a priori values obtained from a least-squares interpolation of the 3D velocity field of central Italy [Galvani et al., 2012, Serpelloni et al., 2013]. The time spanned for the co-seismic analysis consists of the period from 2009.0 to 11 days after the mainshock (i.e., 2009.2908), in order to include as much GPS stations as possible. I use the time-series of 67 stations, belonging to both cGPS and sGPS networks. For those sGPS stations that have no data at least in the 5 days before the mainshock we assume a position around 0 and estimate the uncertainty propagating the errors on the velocity and the position at the reference epoch. To study the post-seismic deformation I use the de-trended and filtered time-series of 27 GPS stations that recorded continuously during the first 306 d after the mainshock.

### 1.2.3 Results

#### Slip distribution models

The first step of the PCAIM software consists in decomposing the dataset in its spatial and temporal components. The addition of a second component in the co-seismic case does not improve the offset determination, since the improvement is gained in the noisy part of the time series (see figure S2 of the Supplementary material in Gualandi et al. [2014]). For the post-seismic stage, a  $F$ -test says that a second component improves significantly (at a 95% confidence level) the fit to the data, while a third component is not justified. Figures 1.3 and 1.4 show the temporal functions used for the co- and post-seismic analyses, respectively, i.e. the columns of matrix  $V^{co}$  and  $V^{post}$  (see equation 1.9).

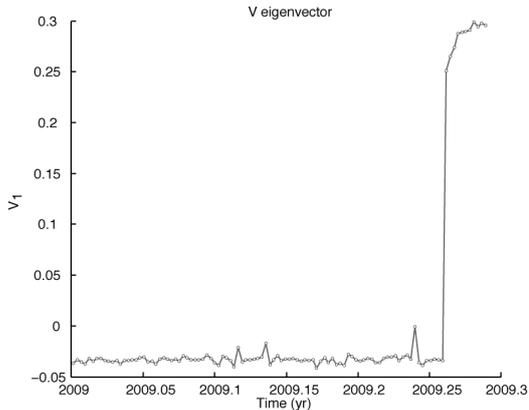


Figure 1.3: Temporal eigenvector  $V^{co}$  for the co-seismic decomposition. To make a realistic estimation of the co-seismic offset considering also stations that missed data immediately after the mainshock it is necessary to retrieve by decomposition the early post-seismic displacement (the last 11 d shown in the figure).

The spatial matrices  $U^*$  (see equation 1.9) are then inverted on a predetermined fault geometry, and the corresponding PCs of slip on the fault are linearly combined using the weights in  $\Sigma^*$  and the temporal functions in  $V^*$ , where  $*$  stands for the co- or post-seismic analysis, respectively. I exploited the relocated aftershock catalogue of Chiaraluce et al. [2011] to define a suite of fault model geometries for the Paganica and Campotosto faults to be used in the co- and post-seismic slip inversions. I tested the sensitivity of GPS observations to the use of increasingly complex fault geometries, adopting as a benchmark for the Paganica fault its geodetic solution (i.e. from Serpelloni et al. [2012]). The fault geometries are obtained with the following procedure: (1) I identify spatial clusters from the relocated aftershocks by applying a  $k$ -means algorithm, therefore, each cluster is defined by a cloud of points; (2) I determine the plane that best approximate the cloud of points through a SVD. This technique allows me to estimate the strike and dip angles of the planes, provided that the number of clusters is already identified, mini-

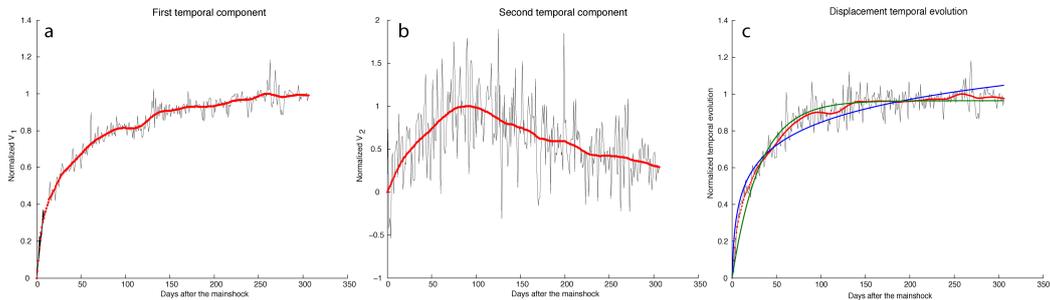


Figure 1.4: (a)-(b) First and second temporal eigenvectors  $V^{post}$  for the post-seismic decomposition. Red lines are obtained filtering high-frequency signal. (c) The red line is the displacement temporal evolution for the post-seismic decomposition, evaluated as the weighted sum of the eigenvectors, where the weights used are the correspondent eigenvalues (the elements of matrix  $\Sigma_{post}$  of the PCA decomposition, see equation 1.1.2). The blue line is the best logarithmic fit and the green line is the best exponential fit. The best parameters estimated with an unconstrained non-linear minimization of the sum of squared residuals (SSR) are:  $A_{log} \sim 0.18$ ,  $\tau_{log} \sim 0.94$  d,  $A_{exp} \sim 0.96$ ,  $\tau_{exp} \sim 31.2$  d.

mizing the degree of subjectivity in tracing fault geometries from aftershock catalogues. For more details, see the Supplementary material of Gualandi et al. [2014]. In order to constrain the absolute position of the fault planes, I forced the shallowest segments to cut the Earth’s surface in correspondence of traces of active faults recognized in the field [e.g. Boncio et al., 2010]. The results of the tests indicated that increasing the complexity of fault geometries, as defined from the aftershock distribution of Chiaraluce et al. [2011], is not justified by the GPS observations available. Accordingly, I adopted the simplest geometry for the Paganica fault, defined by one rectangular plane derived from the northern Paganica seismic cluster (figure S1 of the Supplementary material of Gualandi et al. [2014]).

I have also tested several assumptions about the rake of the inverted slip, and evaluated how these assumptions affect the slip distributions. I performed the inversions assuming three different configurations for the rake: (i) completely variable, (ii) completely fixed along the dip direction ( $-90^\circ$ ) and (iii) slightly variable around the dip direction. The application of these three different constraints is achieved by adopting different weights to the along-strike and along-dip components of the Laplacian operator, whose ratio is indicated by  $R$ . When this ratio is equal to 1, the rake can assume any value, thus corresponding to a completely variable rake configuration, with only the positivity constraint imposed to avoid thrust slip. Increasing the value of  $R$  forces the solution to have rake values closer to  $-90^\circ$ . Here I show the

results corresponding to the best model, obtained for an intermediate value of  $R$  ( $R = 15$ ). For further details, see Section S2.5 of the Supplementary material of Gualandi et al. [2014].

The co-seismic slip distribution model is shown in Figure 1.5. It presents a maximum slip value of  $\sim 130$  cm, confined between 4 and 6 km of depth downdip. Assuming a rigidity modulus of 30 GPa, it is equivalent to a seismic moment  $M_0 = 3.1 \times 10^{18}$  N · m, which is in agreement with previous studies [e.g. Cheloni et al., 2010, D’Agostino et al., 2012, Serpelloni et al., 2012], and corresponds to an earthquake of moment magnitude  $M_W = 6.3$ . The co-seismic slip distribution is available in the format of the Finite-Source Rupture Model Database, and can be downloaded at <http://equake-rc.info/SRCMOD/>.

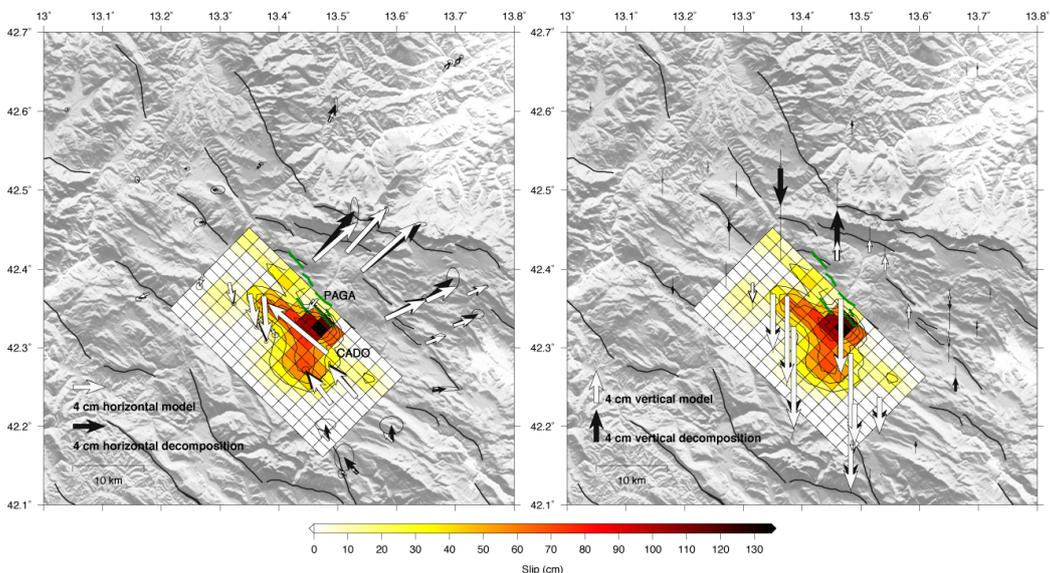


Figure 1.5: Same as figure 4 in Gualandi et al. [2014]. Co-seismic slip model. Black and green lines are the same of Figure 1.2. Contouring lines indicate co-seismic constant slip regions starting from 20 cm to the maximum value, stepping every 20 cm. Left and right panels show horizontal and vertical displacements, respectively, for the stations used in the inversion. The slip and the displacements are computed as the value of the day after the mainshock less the value of the day before.

Figure 1.6 shows the post-seismic slip distribution model for the inversion on the two faults. Comparing the  $\chi^2$  statistics inverting for afterslip only on the Paganica fault and on both the Paganica and Campotosto faults, and using an  $F$ -test, I find a significant (at 95 per cent confidence) improvement in the fit of the post-seismic time series. In particular, the most relevant im-

provement in the fit is recorded at GPS stations near the Campotosto fault (figures 8 and 9 in Gualandi et al. [2014]). The afterslip distribution on the Paganica fault is affected by the introduction of the Campotosto segment, showing a variation in the afterslip module of more than 50 per cent (i.e. up to 8 cm) in the shallowest part of the northern Paganica fault segment, that is the closest to the Campotosto fault. In all the post-seismic inversions, afterslip occurs on two main areas of the Paganica fault plane, located SE and NW of the main co-seismic slip patch. I refer to them as region A and B, respectively (Figure 1.6). In particular, the afterslip patches around region A are stably located on the same position of the fault plane in inversions obtained with different rake constraints. On the contrary, the location of region B varies adopting different rake constraints. I interpret this result

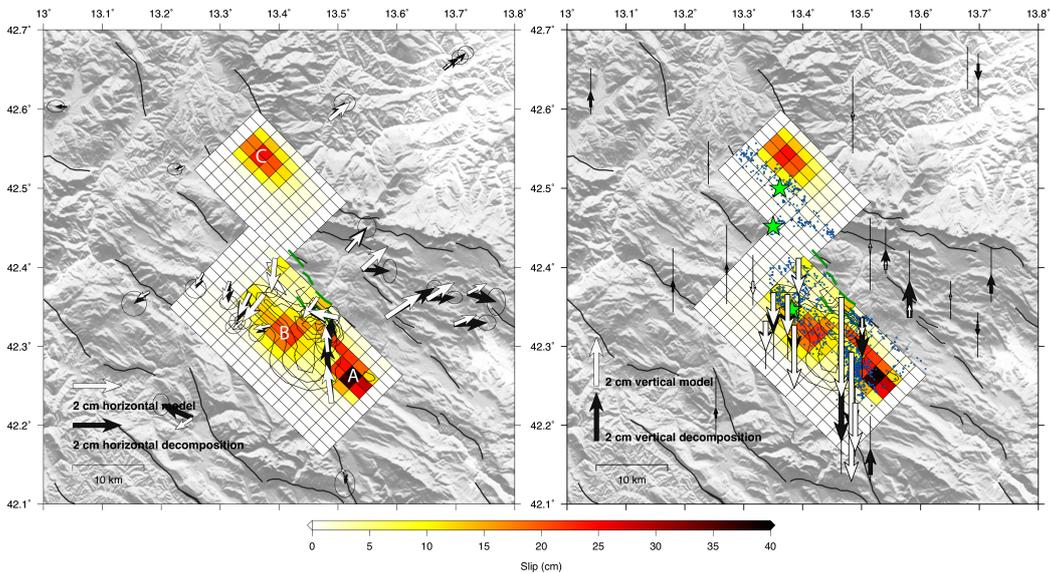


Figure 1.6: Same as figure 9 in Gualandi et al. [2014]. Post-seismic slip model allowing slip on both Paganica and Campotosto faults. Contouring lines as in Figure 1.5. Left and right show horizontal and vertical cumulative displacements, respectively, for the near-field stations. The  $R$  value is set to 15 for both faults. The slip and the displacements are computed as the value of the last day of our analysis (namely, the 306 day after the mainshock) less the value of the first day after the mainshock. Left: Capital letters A, B and C indicate the activated afterslip regions. Right: Green stars show the location of the mainshock (on the Paganica fault) and the two aftershocks of April 9 and June 22. Blue dots are the aftershocks from Chiaraluce et al. [2011].

considering the afterslip region A as a robust feature of the inversion, and a necessary condition to explain the GPS observations. Introducing the Cam-

potosto fault, the fault plane region that undergoes afterslip is the same in all the inversions and is named region C (Figure 1.6). For a rigidity modulus of 30 GPa, the total equivalent seismic moment released by afterslip, in the first 306 d after the mainshock, is  $\sim 1.0 \times 10^{18}$  N · m, corresponding to an earthquake of moment magnitude  $M_W \sim 5.9$ . Instead, afterslip on the Campotosto fault released a total amount of equivalent seismic moment of  $3.5 \times 10^{17}$  N · m, and it is shallower and about complementary to the distribution of aftershocks (Figure 1.6). Afterslip on the Campotosto fault is likely to have loaded these aftershocks. Indeed, introducing the Campotosto fault and taking into account the afterslip of the first 3 d after the mainshock, the Coulomb stress increase ( $\Delta CFF$ ) in the region of the April 9 aftershock doubles, passing from 0.03 to 0.06 MPa. In particular, the afterslip on the Campotosto fault alone induced more than the 98% of the 0.03 MPa increase occurred during the post-seismic stage. The  $M_W$  4.4 aftershock of June 22 (77 d after the mainshock, see Figure 1.6), on the Campotosto fault, is located below the region undergoing afterslip. This event occurred in a fault region that was affected by a negligible co-seismic stress change due to the L’Aquila mainshock ( $\Delta CFF < 0.01$  MPa). Instead, the same fault region experienced a Coulomb stress increase of 0.06 MPa due to afterslip on the Campotosto fault, and it was slightly unloaded by afterslip on the Paganica fault, resulting in a net increase of stress after 77 d of post-seismic stage of 0.05 MPa. These considerations enforce the finding that, following the April 6 mainshock, the Campotosto fault has been activated not only seismically (i.e. aftershocks, [e.g. Chiaraluce et al., 2011]), but also aseismically.

### Fault friction properties

The co-seismic slip distribution and the temporal evolution of afterslip can be used to constrain the parameters of a frictional model of afterslip, based on the rate- and state-dependent constitutive laws. From laboratory experiments [e.g. Dieterich, 1979, Ruina, 1983], it is well known that the friction coefficient ( $\mu$ ) between two blocks of rock varies with the relative velocity between them ( $V$ ) and one or more state parameters ( $\theta_1, \theta_2, \dots$ ), which are representative of the state of the sliding surface. The Dieterich-Ruina rate- and state-dependent friction formulation is:

$$\begin{aligned} \mu(V, \theta) &= \mu^* + a \ln \frac{V}{V^*} + b \ln \frac{V^* \theta}{D_c} \\ \dot{\theta} &= 1 - \frac{V \theta}{D_c} \end{aligned} \quad (1.10)$$

where  $\mu^*$  is the nominal coefficient of friction at the steady reference velocity,  $V^*$  is a reference velocity,  $D_c$  is the so-called critical slip distance,  $a$  and  $b$  are

both positive constants. If  $b > a$  ( $a > b$ ), the constitutive equations predict long-term velocity-weakening (velocity-strengthening) effects on friction.

For a velocity-strengthening region experiencing a positive Coulomb stress change  $\Delta CFF$  due to the co-seismic rupture, the following relation is valid [Perfettini and Ampuero, 2008]:

$$\Delta CFF = (a - b)\sigma \ln \frac{\beta}{V_{pl}} \quad \text{for } \Delta CFF > 0 \quad (1.11)$$

where the effective normal stress on the patch  $\sigma$ , and the loading velocity  $V_{pl}$ , can be estimated from field measurements (I refer to Gualandi et al. [2014] for the discussion of the estimation of the values used for these two parameters).  $\beta$  is the starting sliding velocity of the region at the beginning of the post-seismic stage. Knowing the co-seismic slip distribution allows me to calculate the Coulomb stress variation on a given patch<sup>i</sup>. Thus, the frictional parameter  $a - b$  can be expressed as a function of the starting velocity  $\beta$ . If I model the afterslip  $\delta(t)$  on a velocity-strengthening region with a spring-slider system [Marone et al., 1991], the following relation is valid (appendix A of Gualandi et al. [2014]):

$$\delta(t) - \delta(t_1) \approx \alpha \ln \left[ \frac{\alpha + \beta t}{\alpha + \beta t_1} \right] \quad \text{for } t_1 \leq t \ll t_d = \frac{\alpha}{V_{pl}} \quad (1.12)$$

where  $\alpha = (a - b)\sigma/k = V_{pl}t_d$  is a characteristic slip,  $k$  is the stiffness of the spring in the fault analogue model,  $t_d$  is a characteristic decay time, and  $t_1$  is the epoch corresponding to the first recorded data. Since afterslip patches have a stationary location (i.e. creep propagation is not evident) it is not necessary to model the temporal evolution of slip taking into account elastic interaction between patches [e.g. Wesson, 1980]. Then, as in Perfettini et al. [2010], the model is applied to single patches, neglecting the mutual elastic interaction between them during the post-seismic stage. The values for  $\delta(t)$  are deduced directly from the data, thus I can estimate the parameters  $\alpha$  and  $\beta$ . The values obtained ensure the validity of relation 1.12 for all the patches belonging to regions A and C (Figure 1.6), for which I have estimated the  $a - b$  frictional parameter. The estimate of  $a - b$  for a given patch is slightly affected by the current uncertainty of the long-term fault slip rate ( $V_{pl}$ ), which is assumed to vary in the range 0.2-3 mm/yr (see equation 1.11). I found  $a - b$  values in the range  $10^{-4} - 10^{-3}$ , with the most frequent value of the order of  $10^{-3}$ .

---

<sup>i</sup>The computation of the  $\Delta CFF$  has been performed using the Coulomb 3.3 software [Toda et al., 2011].

#### 1.2.4 Discussion, problems, and limitations

The density of the GPS network operating during the post-seismic stage in the epicentral region motivated me in investigating implications on the mechanics of faulting during the L'Aquila sequence. A better understanding of the physical processes that control the relative amount and location of seismic and aseismic sliding is a key goal in the study of fault mechanics. Rate- and state-dependent friction allows a fault to develop both seismic instabilities and aseismic sliding, according to the value of the constitutive and mechanical parameters [e.g. Rice and Gu, 1983]. Geodetic data recording co-seismic and post-seismic deformation allow us to infer information about the frictional properties of fault planes involved in a seismic sequence [e.g. Hsu et al., 2006, Fukuda et al., 2009, Perfettini et al., 2010]. The  $a - b$  values found for region A on the Paganica fault are in agreement with studies of fault rocks typical of these regions at elevated temperatures and under fluid-saturated conditions [Scuderi et al., 2013]. Small  $a - b$  values, such as  $10^{-3}$  [Marone et al., 1991], characterize fault regions where transitions between velocity weakening ( $a - b < 0$ ) and velocity strengthening ( $a - b > 0$ ) occur. These regions may undergo both afterslip and aftershocks during the post-seismic phase [Boatwright and Cocco, 1996]. Figure 1.7(a) shows the distribution of afterslip on the Paganica fault plane superimposed to the aftershock projections on the same plane. The afterslip region A is also characterized by a high concentration of aftershocks, in agreement with the estimates of the  $a - b$  parameter. Figure 1.7(b) shows the slip temporal evolution and the fit of a spring-slider model for the main afterslip patch of region A considered in the computation of  $a - b$ .

All these considerations are made possible because of the assumption that afterslip was the dominant mechanism of post-seismic deformation. Moreover, before performing the PCA decomposition, I have corrected the GPS position time series for a long-term linear trend and two seasonal signals, with constant amplitude and periods of 0.5 and 1 yr, respectively. It is not guaranteed that these assumptions are correct. If we use the *raw* position time series, i.e. the time series not corrected for any modelled deformation, and we directly apply the PCA algorithm on them, the result is more difficult to interpret (Figure 1.8).

In general, it is also possible that deformation mechanisms different from afterslip are active *at the same time* during the sequence, as well that the assumptions of constant long-term velocity, or constant seasonal amplitude and period, are wrong. A problem arises when we do not have a reliable model describing a deformation signal that we are looking for in the time series. Such signals are often referred to as *transient* signals. A lot of effort

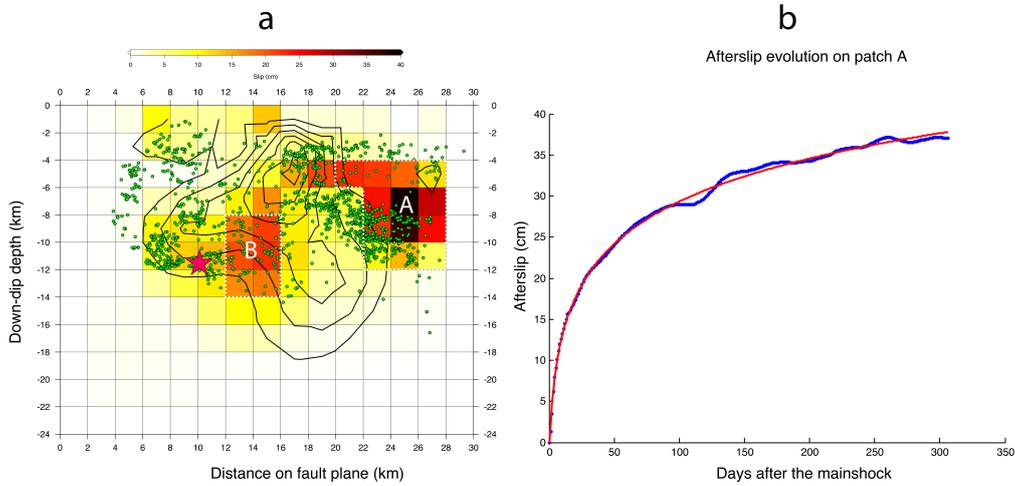


Figure 1.7: (a) Afterslip distribution on the Paganica fault plane when both Paganica and Campotosto faults are taken into account in the inversion. Green dots are the projected aftershocks used to deduce the fault and derived from Chiaraluce et al. [2011]. Contour lines represent the co-seismic slip, as in previous figures. The purple star localizes the mainshock event. Capital letters A and B indicate the main afterslip regions. Region A is used for the calculation of the frictional parameter  $a - b$ . (b) Afterslip history of the main afterslip patch experiencing co-seismic stress increase. The blue dots represent the afterslip deduced from the inversion of GPS data. The red line represents the frictional model expressed by the formula 1.12. The parameter values are:  $\alpha \sim 7.7$  cm,  $\beta \sim 4.3$  cm/d and have been obtained through an unconstrained non-linear minimization of the sum of squared residuals.

has been put in the development of algorithms for the detection of transient events causing displacements of the Earth's surface [e.g. Lohman and Murray, 2013, and references therein]. The PCA has also been used for this goal [e.g. Ji and Herring, 2013], but it is not the most suitable technique for tackling this problem if multiple sources of displacement are acting simultaneously. For this reason I introduce in the next Chapter the Independent Component Analysis (ICA) technique.

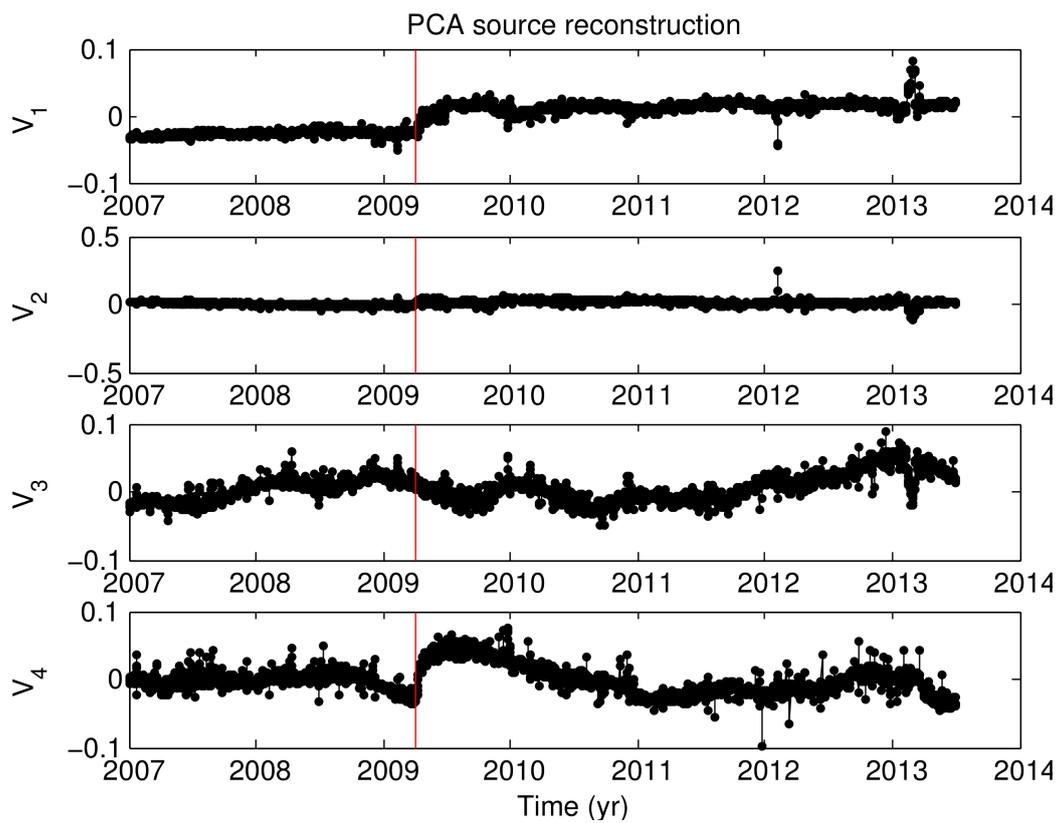
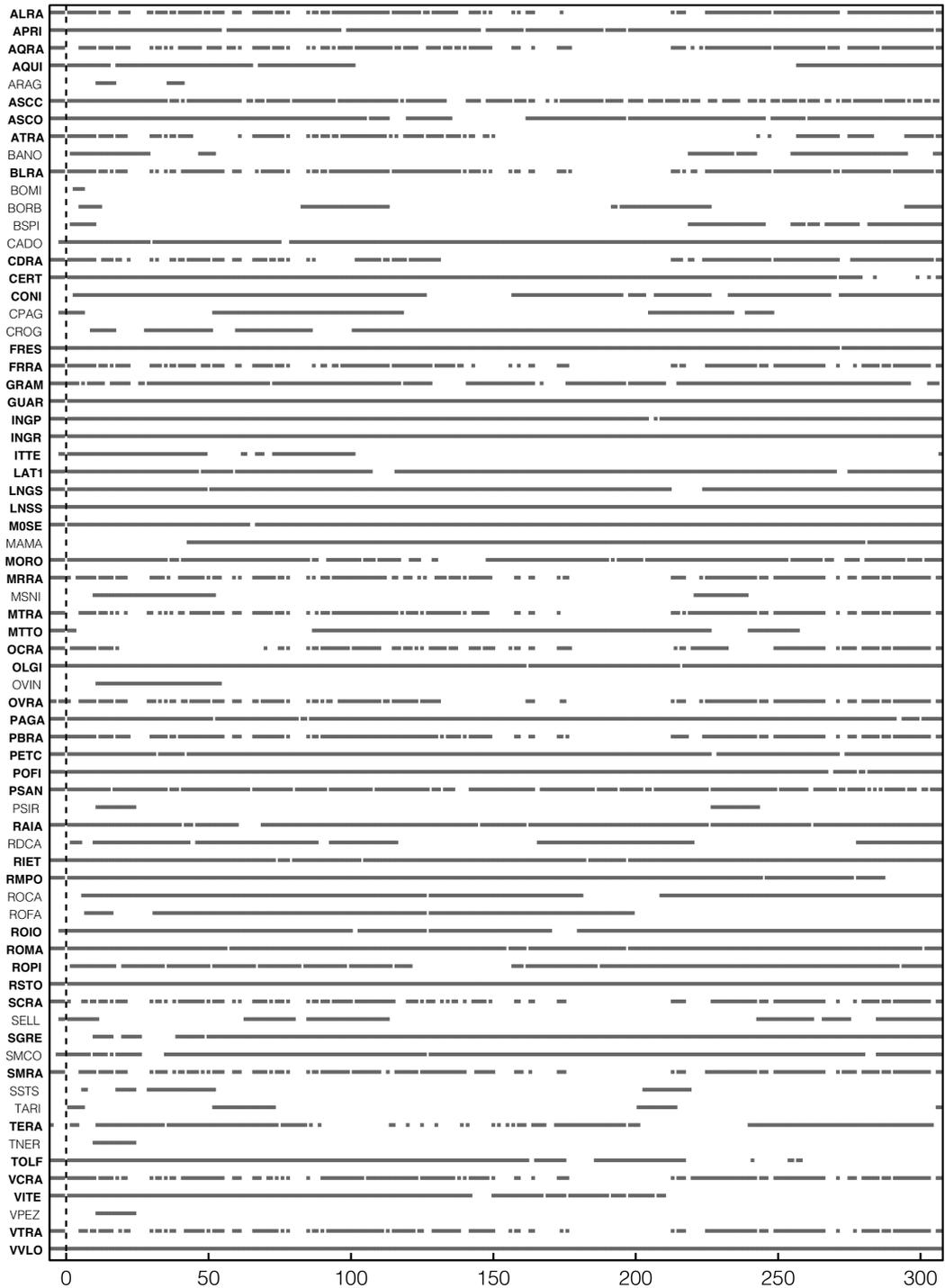


Figure 1.8: PCA decomposition with 4 PCs performed on the time series relative to 12 cGPS stations that recorded continuously during both the pre- and post-seismic stages. The time series used as input for PCA have been corrected only for the CME.

Table 1.1: Occupation history of the GPS stations used in Gualandi et al. [2014]. Horizontal axis: days after the mainshock (dashed line). Vertical axis: stations used. Names in black indicate continuous GPS stations, while names in grey indicate survey-mode GPS stations.



## Chapter 2

# Independent Component Analysis and variational Bayesian approach

*Experience with real-world data, however, soon convinces one that both stationarity and Gaussianity are fairy tales invented for the amusement of undergraduates.*

Thomson, 1994

For what seen in the previous Chapter, the PCA might be not the proper technique to use for investigating the physical phenomena under study. The key point is that nature is fairly complex, and what we do observe is the result of a mix of processes. We are interested in separating all these processes, and learn how they do work individually. Then, in order to join all them together, it is necessary to know how they do interact one with the other, and how to sum up their effects to give the final result, i.e. the observation. Most of the times, we automatically simplify the world, and make the following hypothesis by tacit agreement:

- 1) we think that these processes can be studied separately;
- 2) we just sum up the effects of the separated processes to end up with the resulting observed effect.

The first assumption can be reformulated as the belief that there is no mutual interaction, or this is negligible, between the phenomena under study, i.e. they are reciprocally *independent*. The second hypothesis states that the effect of the sum of the processes is equivalent to the *linear* sum of the

effect of each process individually considered. This assumption allows us to remain in the arena of linear decomposition techniques. PCA is one of these techniques, but instead of assuming independence between the components, it only requires them to be uncorrelated. The independence requirement is much more stringent, and it needs some more advanced work. Therefore, what we are looking for is an *analysis* technique, that is able to extract some *components* mutually *independent* one from the other. In other words, we want to perform an Independent Component Analysis (ICA), that is exactly the subject of this Chapter.

## 2.1 Statement of the problem

As defined by Priemer [1990], in this work I refer to a signal as to “a function that conveys information about the behavior of a system or attributes of some phenomenon”. For all the cases studied in this thesis, the domain of the signals is time. The codomain may vary with the problem under study. For example, if I study the seismic cycle, the codomain of the signal is the slip on the fault; this source signal generates the recorded one, that is the position variation of the sensors located at the Earth’s surface. In general, two different signals might be two functions having the same domain, but two different codomains. The essential is that even if the codomains are different, the resulting effects of the two signals influence in some way the observable recorded by the sensors. For example, if we measure the volume variation with time of an iron bar, the volume variation will be the observable recorded. The causes of this variation might be different, e.g. a temperature variation or a stress applied to the bar. The codomain of the first signal is the temperature, while the codomain of the second one is the stress. Both these signals generate an effect recorded by the volume variation. It is the work of the researcher to give the proper interpretation to the proper causes.

This is true also for geodetic measurements. The effects that we observe at the Earth’s surface consist of position variations. Our goal is to identify the signals that compose the observations, and give them the proper interpretation. As I already pointed out in Chapter 1, when we have to handle datasets belonging to a  $M$ -dimensional space, with  $M > 3$ , it becomes tough to see what is going on. In order to enlight the inner structure of the data, we search for a preferential position for looking to them. Nevertheless, this change of point of view might not be enough. It might be that we have to stop thinking in an Euclidean world. Indeed, up to now we attempted to faithfully reproduce the dataset from an Euclidean point of view. The PCA

minimizes the following quantity:

$$\chi^2 = \sum_{j=1}^M \sum_{t=1}^T [\mathbf{X} - \mathbf{U}\Sigma\mathbf{V}^T]_{jt}^2 \quad (2.1)$$

But what if this assumption were wrong? In other words, what if this minimization is not giving us the *best*<sup>i</sup> point of view? Maintaining the Euclidean approach could introduce some misinterpretation, and it is necessary to modify some of our assumptions if we want to address this *Blind Source Separation* (BSS) problem. Considering the two assumptions made at the beginning of this Chapter, we can write the following equation:

$$\mathbf{X}_{M \times T} = \mathbf{A}_{M \times L} \mathbf{S}_{L \times T} + \mathbf{N}_{M \times T} \quad (2.2)$$

where  $\mathbf{X}$  is the data matrix,  $M$  is the number of observed time series,  $T$  is the number of recorded epochs; each row of  $\mathbf{S}$  contains a source signal, for a total of  $L$ ;  $\mathbf{A}$  is the so-called mixing matrix; and  $\mathbf{N}$  is some additive noise. Therefore, the problem is to attempt to find all the elements on the right side of equation 2.2, knowing only the left side, i.e. the data.

Often, the noise term is incorporated as one of the sources, and equation 2.2 is written as follows:

$$\mathbf{X}_{M \times T} = \mathbf{A}_{M \times L} \mathbf{S}_{L \times T} \quad (2.3)$$

Herauld and Jutten [1986] solved this problem in signal processing using a recursive artificial neural network, assuming a number of sources equal to the number of sensors (square problem), an instantaneous and linear mix, and imposing the mutual *independence* of the underlying signals. It was the first ICA ever done, and from that date this field grew a lot. Comon [1994] gave the first mathematical framework for the study of ICA, introducing a formalism similar to the PCA. Indeed, from equation 2.3, we can easily see the similarity with equation 1.5, and we can give a definition for the ICA similar to that used for the PCA [Comon, 1994]:

**Definition 2.1:** The ICA of a random vector  $\mathbf{x}^t$  of size  $M$  with finite covariance  $\mathbf{C}_{\mathbf{X}}$  is a pair  $\{\mathbf{A}, \Delta^2\}$  of matrices such that

- 1) the covariance factorizes into  $\mathbf{C}_{\mathbf{X}} = \mathbf{A}\Delta^2\mathbf{A}^T$  where  $\Delta^2$  is diagonal real positive and  $\mathbf{A}$  is full column rank  $M$ ;

---

<sup>i</sup>The meaning of *best* varies with the goal of the analysis. PCA is the best decomposition in minimizing the  $\chi^2$ , but it is not the best if we are interested in recovering the sources from a mixed recorded signal.

- 2) the observation can be written as  $\mathbf{x}^t = \mathbf{A}\mathbf{s}^t$ , where  $\mathbf{s}^t$  is a  $M \times 1$  random vector with covariance  $\Delta^2$  and whose components are “the most independent possible”, in the sense of the maximization of a given “contrast function” that will be subsequently defined.

Comparing this definition with the definition 1.1 given in Chapter 1, we see that the first property is the same. What is now required is the independence of the data projected onto the new reference frame. This means that we are still looking for a function  $\Upsilon$  that relates the data space  $\mathbb{R}^M$  to an independent source space  $\mathbb{R}^L$ , or vice-versa. As specified by assumption 2) made at the beginning of this Chapter, we seek a linear  $\Upsilon$ . In the PCA case,  $\Upsilon$  is the rotation matrix  $\mathbf{U}^T$ . In the ICA literature [e.g. Hyvärinen and Oja, 2000, and references therein] it is usually indicated by  $\mathbf{W}$ , and it is a matrix such that  $\mathbf{W}\mathbf{X} = \mathbf{S}$ , and the orthogonality constraint is relaxed. Moreover, the contrast function is not the one defined by equation 2.1, but by a function  $\varphi$  that maximizes the independence among the components. This function is not unique, and we have to choose one among a set of possible functions. In order to make this choice properly, it is necessary to know what *independence* means.

## 2.2 What is independence and how to measure it

Let me recall what said about the data rows in Section 1.1.1, and let me relax one hypothesis assumed there: the Gaussianity. Let us consider each row  $\mathbf{x}_j$ ,  $j = 1, \dots, M$ , as the realization of a stochastic process, and let us suppose that it is a non-degenerate random vector of dimension  $T$  with i.i.d. elements. Neglecting moments of order higher than the second, and searching for a decorrelation of the covariance matrix, the PCA implicitly assumes a Gaussian distribution for the data<sup>ii</sup>. On the contrary, now the first two moments might be, in general, not enough to describe the variable  $\mathbf{x}_j$ . This means that stopping our analysis to the second-order moments, disregarding the higher ones, is no more a good strategy to pursue. To understand how to proceed, we need the following definition of independence. For all the following definitions, I assume that the random variables are continuous, since the sources that I study may have any value in  $\mathbb{R}$ .

**Definition 2.2:** Given two (scalar-valued) random variables  $s_1$  and  $s_2$ , with probability density functions  $p_1(s_1)$  and  $p_2(s_2)$ , then we say that  $s_1$  and  $s_2$  are

---

<sup>ii</sup>To be more general, PCA assumes a distribution described using only two moments. Usually, it is used a Gaussian distribution.

independent iff the combined random variable  $(s_1, s_2)$  has a joint probability density function given by:

$$p(s_1, s_2) = p_1(s_1)p_2(s_2) \quad (2.4)$$

In general, a set of  $L$  random variables  $\mathbf{s} = \{s_1, \dots, s_L\}$  is statistically independent iff their joint probability density function factorises such that

$$p(\mathbf{s}) = \prod_{i=1}^L p_i(s_i) \quad (2.5)$$

Equation 2.5 states a property that we want for our set of statistically independent signal sources. Whatever will be our choice for the contrast function  $\varphi$ , it must take into account this definition. Comon [1994] argued that a natural way to check whether  $\mathbf{s}$  has independent components is to measure how different are the two sides of equation 2.5. To measure the difference between two probability density functions (pdfs), Basseville [1989] indicated the class of  $f$ -divergences as a very important one. For instance, the *Kullback-Leibler (KL)-divergence* is a widely used measure of statistical dependence between two random variables [Kullback and Leibler, 1951]. It is a non-negative quantity, and in the case it is applied to equation 2.5 it takes also the name of *mutual information (MI)* [Cover and Thomas, 1991], and it is defined as follows:

$$MI[\mathbf{s}] = KL \left[ p(\mathbf{s}) \parallel \prod_{i=1}^L p_i(s_i) \right] \doteq \int p(\mathbf{s}) \ln \frac{p(\mathbf{s})}{\prod_{i=1}^L p_i(s_i)} d\mathbf{s} \geq 0 \quad (2.6)$$

where  $d\mathbf{s} = \prod_i ds_i$ . The minimization of the  $MI$  is probably the most intuitive method for finding independent sources. Indeed,  $MI(\mathbf{s}) = 0$  iff the variables  $s_i$  are mutually independent, and is strictly positive otherwise. Nevertheless, the actual  $p(\mathbf{s})$  is unknown, and the quantity  $MI$  can not be directly used as a contrast function. For practical purposes, the contrast function is approximated by the Edgeworth expansion of the  $MI$ , and consists of a combination of third- and fourth-order marginal cumulants. Comon [1994] found that a particular contrast function of interest is the sum of squares of marginal cumulants of order 4. The fourth-order cumulant (i.e., the kurtosis) is the classical measure for non-gaussianity. It has been shown [e.g. Hyvärinen and Oja, 2000] that the minimization of the  $MI$  is equivalent to the maximization of the non-gaussianity of the ICs. Another way to look at the problem is through another quantity: the *negentropy*. The negentropy is based on the information-theoretic quantity of (differential) entropy, that

is the basic concept of information theory. Entropy was defined by Shannon [1948] for the discrete case, and the differential entropy is its counterpart in the continuous case. Given a random vector  $\mathbf{s}$  with pdf  $p(\mathbf{s})$ , the entropy of  $\mathbf{s}$  is defined as [Cover and Thomas, 1991, Papoulis, 1991]:

$$H(p(\mathbf{s})) = H[\mathbf{s}] \doteq - \int p(\mathbf{s}) \ln p(\mathbf{s}) d\mathbf{s} \quad (2.7)$$

A Gaussian variable has the largest entropy among all random variables of equal variance (for a proof, see e.g. Cover and Thomas [1991] or Papoulis [1991]).

The negentropy is defined as:

$$J(p(\mathbf{s})) = J[\mathbf{s}] \doteq H_{\mathcal{N}}[\mathbf{s}] - H[\mathbf{s}] \quad (2.8)$$

where  $H_{\mathcal{N}}[\mathbf{s}]$  is the entropy of a Gaussian distribution with the same covariance as  $\mathbf{s}$ .

Comon [1994] shown that the negentropy and the *MI* are related, and that maximising the non-Gaussianity of the source signals is equivalent to minimising the *MI* between them (if the data are pre-whitened).

The idea of using deviation from Gaussianity to measure independence is based on the Central Limit Theorem (CLT). Indeed, the CLT expresses the concept that the distribution of a sum of independent random variables tends toward a Gaussian distribution. Thus, a sum of two independent random variables usually has a distribution that is closer to Gaussian than any of the two original random variables [Hyvärinen and Oja, 2000]. This is also the reason why two Gaussian sources are indistinguishable for an ICA algorithm: not more than one Gaussian source is allowed into the mix. Indeed, the joint pdf of two Gaussian, uncorrelated, and unit variance random variables mixed via an orthogonal matrix is equal to the product of the two pdfs. This can be easily seen graphically (see, for example, figure 7 of Hyvärinen and Oja [2000]): the orthogonal mixing matrix is a rotation, and the joint pdf is a Gaussian in two dimensions, with the same variance in each direction (unit variance), and not aligned along any particular direction (the variables are uncorrelated). Under any rotation, we end up with the same pdf, and it is not possible to find a preferential direction for projecting the data.

This last example makes clear that the ICA is strictly linked to the PCA. The approach sketched up to now is also called the *mapping* approach. Indeed, it is conceptually the same of the one described for performing a PCA: we seek a function that projects the data from the data space ( $\mathbb{R}^M$ ) to the source space ( $\mathbb{R}^L$ ).

A general problem with this approach for the determination of the sources in an ICA is that the estimation of the contrast function is difficult, and approximations are needed. This means that when we minimize (or maximize) the approximated contrast function, we make some mistakes due to the non-optimal choice of the approximation. Another possible approach is the one that uses generative models. Roberts and Everson [2001] have shown that minimising the *MI* corresponds to maximising a generative log-likelihood to within an additive constant. This means that there exists a generative model for all the possible mapping procedures. Under a generative model approach, instead of looking for the best mapping function  $\Upsilon : \mathbb{R}^M \mapsto \mathbb{R}^L$ , a parametric model is built in order to imitate the forward mixing process of  $L$  independent source signals:  $\Upsilon : \mathbb{R}^L \mapsto \mathbb{R}^M$ . The contrast function is the likelihood of the model generating the observed data, or, in a Bayesian framework, it is the *a posteriori* pdf of the model parameters. In a generative model different parameters need to be estimated, and this significantly slows down the process if compared to the traditional mapping approach. Nevertheless, it brings a very valuable advantage: it allows to treat non-square and noisy mixing problems with no particular effort. Here I adopt the algorithm developed by Choudrey [2002], called variational Bayesian ICA (vbICA)<sup>iii</sup>. The variational approach allows to speed up the estimation of the parameters using the calculus of variations, a concept formerly developed in mechanical statistics. In the next Section I present the basic concepts of the algorithm following Choudrey [2002].

## 2.3 The variational bayesian ICA

A generative model  $\mathcal{M}$  is characterised by some observed variables ( $\mathbf{X}$ ), some latent variables ( $\mathbf{H}$ ), some hidden parameters ( $\Theta$ ), and the mutual relationships between all these quantities. The observations and the hidden variables are quantities identified with the “real world” whose distributions are being modelled using the parameters. Parameters exist to make the model “work” and their distributions (if any) are modelled using further parameters (called *hyper-parameters*). Both the parameters and the hidden variables are unknown, and they are collected under the same name of weights, indicated as  $\mathbf{W} = \{\mathbf{H}, \Theta\}$ . The particular model  $\mathcal{M}$  represents all the assumptions and beliefs about the model, like the structure and the values of the hyper-parameters governing the prior pdf of the weights  $p(\mathbf{W}|\mathcal{M})$ .

The goal of a generative model is to find the *best* weights in order to

---

<sup>iii</sup>The original code is downloadable at <http://www.robots.ox.ac.uk/~parg/projects/ica/riz/code.html>. Last access 8 March 2015

explain the observations *and* match the *a priori* knowledge. In a Bayesian framework, given a model  $\mathcal{M}$  and the observed data  $\mathbf{X}$ , maximising the posterior pdf over weights  $\mathbf{W}$  given the data  $\mathbf{X}$  is the best choice for  $\mathbf{W}$ :

$$p(\mathbf{W}|\mathbf{X}, \mathcal{M}) = \frac{p(\mathbf{X}|\mathbf{W}, \mathcal{M})p(\mathbf{W}|\mathcal{M})}{p(\mathbf{X}|\mathcal{M})} \quad (2.9)$$

where the denominator is called the evidence for  $\mathcal{M}$  and is expressed by:

$$p(\mathbf{X}|\mathcal{M}) = \int p(\mathbf{X}|\mathbf{W}, \mathcal{M})p(\mathbf{W}|\mathcal{M})d\mathbf{W} \quad (2.10)$$

For brevity, I will drop the dependency on the model  $\mathcal{M}$  in the next equations.

In practice, the computation of the integral in equation 2.10 is intractable in most of the cases, since it has to be calculated in the whole weight space, and also in this case approximations are needed in order to evaluate it. The approximation used by Choudrey [2002] consists in a variational approximation. Let me first briefly describe the general concept behind it, and then I will report how it has been applied to the ICA problem.

### 2.3.1 Variational approximation

Variational approximations is a body of deterministic techniques for making approximate inference for parameters in complex statistical models. Summaries of contemporary variational approximations can be found in the works of Jordan et al. [1999], Jordan [2004], Titterington [2004], and Bishop [2006].

The most common variant of variational approximation is the so-called density transform approach, which involves approximation of posterior densities by other densities for which inference is more tractable. Let  $p'$  be an arbitrary density over the weight space. Then, the logarithm of the marginal likelihood for data  $\mathbf{X}$  can be written as:

$$\begin{aligned} \ln(p(\mathbf{X})) &= \ln(p(\mathbf{X})) \int p'(\mathbf{W})d\mathbf{W} \\ &= \int p'(\mathbf{W})\ln(p(\mathbf{X}))d\mathbf{W} \\ &= \int p'(\mathbf{W})\ln\left(\frac{p(\mathbf{X}, \mathbf{W})}{p(\mathbf{W}|\mathbf{X})}\frac{p'(\mathbf{W})}{p'(\mathbf{W})}\right)d(\mathbf{W}) \\ &= \int p'(\mathbf{W})\ln\frac{p(\mathbf{X}, \mathbf{W})}{p'(\mathbf{W})}d\mathbf{W} + \int p'(\mathbf{W})\ln\frac{p'(\mathbf{W})}{p(\mathbf{W}|\mathbf{X})}d\mathbf{W} \end{aligned} \quad (2.11)$$

Remembering equation 2.6, the second term of the r.h.s. is the *KL*-divergence between the two pdfs  $p'(\mathbf{W})$  and  $p(\mathbf{W}|\mathbf{X})$ , i.e. it measures how

far the arbitrary approximating distribution  $p'(\mathbf{W})$  is from the posterior  $p(\mathbf{W}|\mathbf{X})$ . Instead, the first term of the r.h.s. is known as the Negative variational Free Energy (NFE), and, recalling the definition of differential entropy (see equation 2.7), can be written as:

$$\begin{aligned} NFE[\mathbf{X}] &= \int p'(\mathbf{W}) \ln \frac{p(\mathbf{X}, \mathbf{W})}{p'(\mathbf{W})} d\mathbf{W} \\ &= \int p'(\mathbf{W}) \ln(p(\mathbf{X}, \mathbf{W})) d\mathbf{W} - \int p'(\mathbf{W}) \ln(p'(\mathbf{W})) d\mathbf{W} \quad (2.12) \\ &= \langle \ln(p(\mathbf{X}, \mathbf{W})) \rangle_{p'(\mathbf{W})} + H[\mathbf{W}] \end{aligned}$$

where  $\langle \cdot \rangle_{p'}$  indicates the expected value given the pdf  $p'$ .

Because of the non-negativity of the  $KL$ -divergence, the following inequality holds:

$$\ln(p(\mathbf{X})) \geq NFE[\mathbf{X}] \quad (2.13)$$

with equality iff the approximating  $p'(\mathbf{W})$  equals the true (unknown) posterior  $p(\mathbf{W}|\mathbf{X})$ . Noting that the log-evidence is not dependent on the weights  $\mathbf{W}$ , maximising the NFE w.r.t. the approximating posterior  $p'(\mathbf{W})$  necessarily minimise the  $KL$ -divergence between the approximating and true posterior. The essence of the density transform variational approach is approximation of the posterior density  $p(\mathbf{W}|\mathbf{X})$  by a  $p'(\mathbf{W})$  for which  $p(\mathbf{X}; p') = \exp(NFE[\mathbf{X}])$  is more tractable than  $p(\mathbf{X})$ . Tractability is achieved by restricting  $p'$  to a more manageable class of densities, and then maximising  $p(\mathbf{X}; p')$  over that class [Ormerod and Wand, 2010]. The most common restriction for the  $p'$  density is that  $p'(\mathbf{W})$  factorizes into  $\prod_{i=1}^N p'(\mathbf{w}_i)$  for some partition  $\{\mathbf{w}_1, \dots, \mathbf{w}_N\}$  of  $\mathbf{W}$ . This restriction is also known as *mean field* approximation and derives from statistical mechanics [e.g. Callen, 1985, Parisi, 1988]. Restriction of  $p'$  to a subclass of product densities gives rise to explicit solutions for each product component in terms of the others. These, in turn, lead to an iterative scheme for obtaining the simultaneous solutions [e.g., Algorithm 1 of Ormerod and Wand, 2010]. It is thus necessary to define a partition of the weights  $\mathbf{W}$ , decide for each of them which density has to be used for their description, and enforce a particular factorisation. Then, the NFE is maximised for each factorising posterior iteratively until convergence. In the next Section I will briefly describe the model proposed for solving the ICA problem in particular. The exposition follows the nomenclature of Choudrey [2002], but the same concepts have been published almost at the same time by Chan et al. [2003].

### 2.3.2 Variational approximation for the ICA problem

Recalling equation 2.2, I indicate the observed data with  $\mathbf{X}$ , while the weights  $\mathbf{W}$  will be given (at least) by  $\{\mathbf{A}, \mathbf{S}, \mathbf{N}\}$ . MacKay [1995] has shown that there is no need to specify functional forms for the posteriors if conjugate forms for the densities are chosen. This means that once we have defined the prior distributions, the posteriors will come from the maximisation of the NFE. The prior density over the mixing matrix is given by a product of Gaussians:

$$p(\mathbf{A}) = \prod_{i=1}^L \prod_{j=1}^M \mathcal{N}(A_{ji}|0, \alpha_{ji}) \quad (2.14)$$

The noise is assumed to be Gaussian with zero mean and diagonal precision  $\mathbf{\Lambda}$ , i.e.:

$$p(\mathbf{n}^t|\mathbf{\Lambda}) = \mathcal{N}(\mathbf{n}^t; \mathbf{0}, \mathbf{\Lambda}) \quad (2.15)$$

where  $\mathbf{n}^t$  is the noise vector at a given time  $t$ . In particular, the prior over the sensor noise precision is assumed to be a product of Gamma distributions:

$$p(\mathbf{\Lambda}) = \prod_{j=1}^M \mathcal{G}(\Lambda_j; b_{\Lambda_j}, c_{\Lambda_j}) \quad (2.16)$$

The sources instead are modelled similarly to the Independent Factor Analysis of Attias [1999], where each independent component  $i = 1, \dots, L$  is modelled by a mixture of  $m_i$  Gaussians (MoG). In this way, the source model should be capable of encompassing multimodal distributions, and move beyond traditional ICA. The importance of being able to describe also multimodal distributions in geophysics, and particularly in geodesy, is shown by Figure 2.1, where some typical signals related to GPS observations are presented. In order to describe  $\mathbf{S}$  we need the parameters governing the Gaussians of the mix  $(\mu_{i,q_i}, \beta_{i,q_i})$  that describe the source  $i$ , as well as the probability that a given Gaussian of the mix is actually used to explain the source  $i$  ( $\pi_{i,q_i}$ ). This means that the prior distribution over  $\mathbf{S}$  at time  $t$  is:

$$p(\mathbf{s}^t|\boldsymbol{\theta}) = \prod_{i=1}^L \sum_{q_i=1}^{m_i} \pi_{i,q_i} \mathcal{N}(s_i^t; \mu_{i,q_i}, \beta_{i,q_i}) \quad (2.17)$$

where the variable  $q_i$  is an indicator variable signifying which Gaussian component of the  $i$ -th source is chosen for generating  $s_i^t$  and takes on values of  $\{q_i = 1, q_i = 2, \dots, q_i = m_i\}$ .  $\mu_{i,q_i}$  and  $\beta_{i,q_i}$  are the mean and the precision of the  $q_i$  Gaussian for the  $i$ -th source, respectively. The probabilities

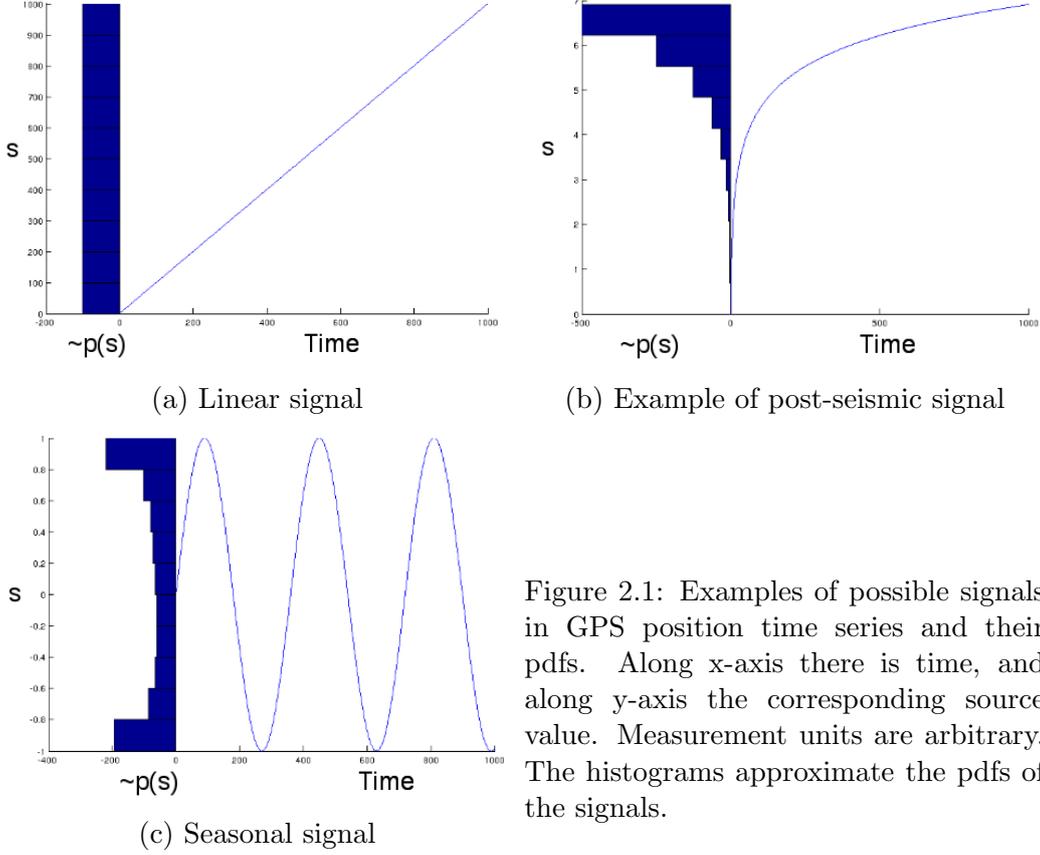


Figure 2.1: Examples of possible signals in GPS position time series and their pdfs. Along x-axis there is time, and along y-axis the corresponding source value. Measurement units are arbitrary. The histograms approximate the pdfs of the signals.

$\pi_{i,q_i} = p(q_i^t = q_i | \boldsymbol{\pi}_i)$  are the mixing proportions, and their probability density depends on the actual Gaussian of the mix  $q_i$  considered among the  $m_i$ . Thus, the parameters of source  $i$  are  $\boldsymbol{\theta}_i = \{\boldsymbol{\pi}_i, \boldsymbol{\mu}_i, \boldsymbol{\beta}_i\}$  and the complete parameter set of the source model is  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_L\}$ . Instead, the complete collection of possible source states is denoted  $\mathbf{q} = \{q_1, q_2, \dots, q_L\}$  and runs over all  $\mathbf{m} = \prod_i m_i$  possible combinations of source states. Then, the prior over the states and the sources is given by:

$$p(\mathbf{q} | \boldsymbol{\pi}) = \prod_{t=1}^T \prod_{i=1}^L \pi_{i,q_i} \quad (2.18)$$

$$p(\mathbf{S} | \mathbf{q}^t, \boldsymbol{\theta}) = \prod_{t=1}^T \prod_{i=1}^L \mathcal{N}(s_i^t; \mu_{i,q_i}, \beta_{i,q_i}) \quad (2.19)$$

and  $p(\boldsymbol{\theta}) = p(\boldsymbol{\pi})p(\boldsymbol{\mu})p(\boldsymbol{\beta})$ . The prior over the mixture proportion is assumed to be a product of symmetric Dirichlet distributions:

$$p(\boldsymbol{\pi}) = \prod_{i=1}^L \mathcal{D}(\boldsymbol{\pi}_i; \lambda_{i0}) \quad (2.20)$$

the prior over the means is a product of Gaussians:

$$p(\boldsymbol{\mu}) = \prod_{i=1}^L \prod_{q_i=1}^{m_i} \mathcal{N}(\mu_{i,q_i}; m_{i0}, \tau_{i0}) \quad (2.21)$$

and the prior over the precisions is a product of Gammas:

$$p(\boldsymbol{\beta}) = \prod_{i=1}^L \prod_{q_i=1}^{m_i} \mathcal{G}(\beta_{i,q_i}; b_{i0}, c_{i0}) \quad (2.22)$$

Thus, the whole set of hidden variables and parameters used in the model can be expressed as the collection  $\mathbf{W} = \{\mathbf{A}, \mathbf{S}, \boldsymbol{\Lambda}, \mathbf{q}, \boldsymbol{\theta}\}$ . Choudrey [2002] tested two different factorisation for  $p'(\mathbf{W})$ :

- i)  $p'(\mathbf{W}) = p'(\boldsymbol{\Lambda})p'(\mathbf{A})p'(\mathbf{S})p'(\mathbf{q})p'(\boldsymbol{\theta})$
- ii)  $p'(\mathbf{W}) = p'(\boldsymbol{\Lambda})p'(\mathbf{A})p'(\mathbf{S}|\mathbf{q})p'(\mathbf{q})p'(\boldsymbol{\theta})$

showing that the second outperforms the first. In this thesis, I use only the second factorisation.

Once established all the hyper-parameters that determine the prior distributions, a set of learning equations allows us to compute the parameters that govern the distributions of the posterior densities under a variational approximation. I do not discuss here this set of equations, and I refer to Choudrey [2002] (in particular, chapter 5 and appendix C). What is important to point out is that I have modified this set of equations in order to take into account missing data in the learning of the parameters of the posterior distributions. In particular, I have used the modifications proposed by Chan et al. [2003] where it has been added an indicator variable  $\mathbf{o}^t$  for observed entries in  $\mathbf{x}^t$ , such that  $o_j^t = 1$  if  $x_j^t$  is observed, and  $o_j^t = 0$  if  $x_j^t$  is missing, with  $j = 1, \dots, M$  and  $t = 1, \dots, T$ .

In order to visualize the described model, it is possible to make use of the direct graph shown in Figure 2.2. The circles indicate random variables, while the squares indicate deterministic variables.

From what said, the role of priors is particularly relevant. In all the case studies I use general values of  $m_{i0} = 0, \tau_{i0} = 1, b_{i0} = 10^3, c_{i0} = 10^{-3}$ .

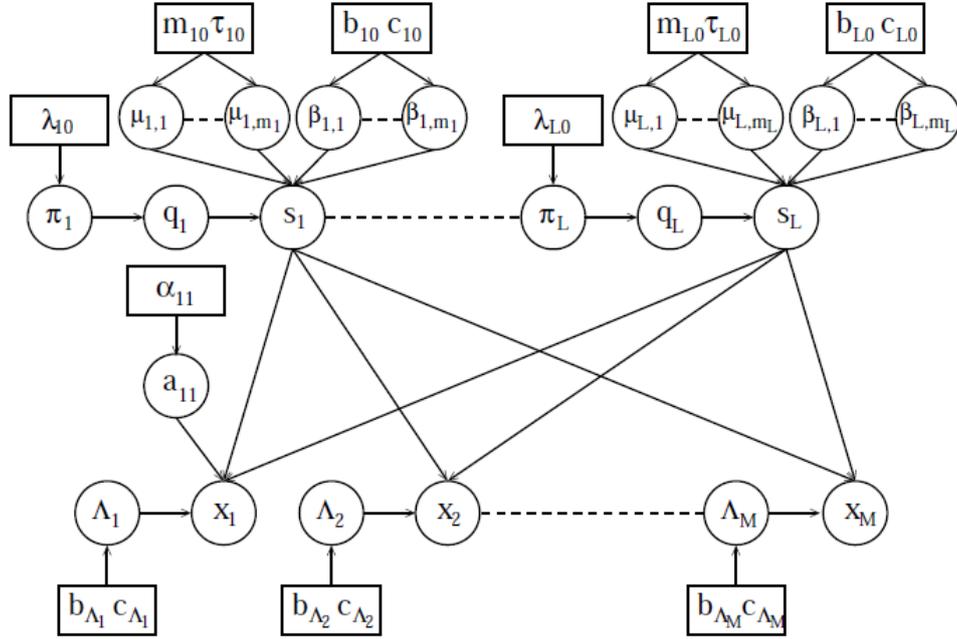


Figure 2.2: Same as figure 5.1 of Choudrey [2002]. Bayesian Independent Component Analysis as a Graphical Model. Circles represent random variables and rectangles represent hyper-parameters. The meaning of each symbol is indicated in the main text. The mixing matrix parameters are all summarized in one symbol for brevity.

This set of values is not particularly stringent, and let the mix of Gaussians enough freedom to reproduce the data. Choudrey [2002] suggest to use values  $1 \leq b \leq 10^3$  with  $bc \approx 1$  for Gamma distributions. Also the number of Gaussians to be used in modelling each source must be chosen. Choudrey [2002] indicates that, for most datasets, 3-5 components have been found to be enough. Here, in order to describe the sources related to the GPS observations, I use 4 components for each MoG.

Another important factor that controls the final result of the vbICA algorithm is the choice of the initialization point. Two possible initialization solutions are adopted: a first one that uses the PCA decomposition, and a second one completely random. If the PCA initialization is set, then the  $\mathbf{U}\Sigma$  and the  $\mathbf{V}$  values are used to set the starting point for the mixing matrix and the sources, respectively. In particular, since the sources are described by a MoG, the actual  $\mu_{i,q_i}$  values are set using a  $k$ -means algorithm on the  $i$ -th column of  $\mathbf{V}$ , and the mixing probabilities are all equals.

### 2.3.3 How to select the number of components

In dimensionality reduction problems the selection of the number of components to use in order to best describe the observations is always a key point. An ICA does not organize into a diagonal matrix the variances of the dataset, and thus we can not use a threshold of explained variance as usually done for a PCA. It is still possible to use a reduced  $\chi^2$  test, or a  $F$ -test (e.g., see Kositsky and Avouac [2010] for an application to a PCA). Anyway, the vbICA algorithm offers another powerful tool, called Automatic Relevance Determination (ARD). The capability of setting the strengthness or weakness of the priors is given by the hyperparameters that control their distributions. A strong confidence in the starting model implies a strong similarity between the posteriors and the priors. In particular, under weak priors the data play the most important role in guiding the learning of the posterior parameters, and an increasing number of components is used to fit (or over-fit) the data. The ARD exploits the fact that each of the  $L$  columns of the mixing matrix is associated with one of the  $L$  sources. Instead of assigning a different variance to each element of the mixing matrix, let us associate only one variance value  $\alpha_i$ , with  $i = 1, \dots, L$ , to each column. This makes the mixing matrix depending on the set of parameters  $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_L\}$ , which define the precision associated to each column of  $\mathbf{A}$  and define how strong is the assumption that the mean value of the columns of the mixing matrix is zero. In other words, a certain  $\alpha_i$  defines how relevant is the source  $i$  for the explanation of the data. Also these parameters  $\boldsymbol{\alpha}$  follow a distribution, and their priors is supposed to be a product of Gamma distributions:

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^L \mathcal{G}(\alpha_i; b_{\alpha_i}, c_{\alpha_i}) \quad (2.23)$$

and the factorisation of the weights is rewritten as:

$$p'(\mathbf{W}) = p'(\boldsymbol{\Lambda})p'(\mathbf{A})p'(\boldsymbol{\alpha})p'(\mathbf{S}|\mathbf{q})p'(\mathbf{q})p'(\boldsymbol{\theta}) \quad (2.24)$$

In this way, the posterior precision over the mixing matrix elements depends on a prior precision that is adapted during the learning process, embodied in  $\alpha_i$ . Large  $\alpha_i$  corresponds to a posterior over mixing matrix column  $i$  dominated by the prior density, effectively setting the elements of column  $i$  to zero. This will result in heavy suppression of the  $i$ -th source signal. By monitoring the variance of each source signal - or, equivalently, the values of  $\alpha_i$  - the most likely number of sources supported by the observation data can be ascertained [Choudrey, 2002]. Here, to determine the suggested number of components to retain, I directly compare the variances associated to the

posterior of each column  $i$ , and if the maximum variance is more than 10 times bigger than the smallest one, I consider the source associated with the latter as noise, and discard the last component. As suggested by Choudrey [2002], a prudent strategy would be to use ARD to narrow down the search, then use individual model selection via the NFE over the model orders of interest. This is exactly the approach that I take in the rest of the thesis.

## 2.4 Summary

Most of the interesting phenomena in nature are described by non-Gaussian distributions. This makes the point of moving beyond the PCA approach, and explore the ICA technique. In geophysics, ICA algorithms have been applied both to atmospheric data [Unkel et al., 2010, 2011] and solid Earth data [Bottiglieri et al., 2007, 2010, Forootan and Kusche, 2012]. All these applications have used different kinds of mapping approaches. The limitation of the mapping approach is intrinsic in the approximation needed to make the contrast function tractable. In this Chapter I have briefly exposed the concepts and the model behind a variational bayesian approach to the ICA. Some limitations are still present, like the assumption of linear mixing matrix or not moving sources. Anyway, this generative model approach allows to naturally study non-square problems, i.e. problems where the number of sensors exceeds the number of sources. Moreover, the approximation of the sources via a MoG enables to treat multi-modal distribution of the sources. And finally, the modification proposed by Chan et al. [2003] allows to handle missing data. These three characteristics (non-square problems, multi-modal source distributions, missing data) likely affect geodetic sources and measurements.

# Chapter 3

## Test on synthetic data

*Γνώθι σεαυτόν*

*Nosce te ipsum*

### 3.1 Data generation

To validate the modified algorithm described in Chapter 2, I perform tests on synthetic time series, realized in order to simulate continuous GPS (cGPS) data and evaluate the ability of the algorithm to retrieve the known sources. To reproduce cGPS data, first I create some source signals ( $\mathbf{S}$ ), then I sum up a specific linear combination of them (i.e., I use a predetermined mixing matrix  $\mathbf{A}$ ), and finally I add some noise ( $\mathbf{N}$ ).

The signals are created with the goal of simulating daily data, as usually recorded by continuous GPS networks, around an active fault and a volcanic region. The time spanned for the analysis is 5 years, and the number of cGPS stations is 20, for a data matrix  $X_{M \times T} = X_{60 \times 1827}$  (considering also leap years). As reference epoch I assume the starting time  $T_{start} = 0$ . This means that all the data of the epochs that follow  $T_{start}$  are measurements of the displacement along east, north, and vertical directions relative to the first epoch. The unit of measurement considered for the position time series is mm. The simulation of the seismic cycle is performed considering only one planar fault as the source of the deformation, while the volcanic activity is simulated using an inflating and deflating Mogi source. The volcanic source can be generally imaged as a source of transient deformation superimposed on the tectonic deformation due to fault activity. The geometry of the fault is maintained constant, and what varies is the extension of the fault plane that undergoes slip as well as the value of slip.

The tests are performed modifying 2 different conditions:

- 1) source intensities (signal-to-noise ratio, SNR, and signal-to-signal ratio, SSR).
- 2) network quality (geometry, w.r.t. the fault plane, and percentage of missing data);

Point 1) depends only on the sources, while point 2) relies only on the sensors. Of course, the worst case scenario is realized when I have a low source signal intensity (e.g., a small earthquake) and a bad network quality (unfavourable geometry and high percentage of missing data).

### 3.1.1 Description of the sources

#### Seismic cycle sources

In order to reproduce the seismic cycle, I use three different source signals:

- 1) a linear function;
- 2) a Heaviside function;
- 3) a logarithmic function.

With these three signals I want to simulate the 1) inter-seismic, 2) co-seismic, and 3) post-seismic stages, respectively. The usage of these three functions is overall justified by the observation of real world data. GPS data have been widely used to study the seismic cycle, and interseismic velocities are now well constrained by geodetic measurements [e.g., Grenerczy et al., 2005, Galvani et al., 2012]. Thus, it is common to explain the secular linear trend observed in GPS time series as due to inter-seismic motion of crustal blocks, while the differences in trends among sites, e.g. across a fault, as due to inter-seismic deformation, due to long-term slip of the fault during the inter-seismic phase in the deep creeping portion of the fault plane, and locking in the shallow brittle portion. It is obvious the usage of a Heaviside function to model a co-seismic deformation. Indeed, sampling the data daily, it is not possible to see the evolution of the co-seismic rupture, and what is recorded is just a jump from the position before the earthquake and after it. Finally, it is less obvious the usage of a logarithmic function to represent a post-seismic decay. The post-seismic deformation, in fact, can be driven by different phenomena, each of them following a characteristic evolution with time [e.g. Barbot et al., 2008, and references therein]. Afterslip has been modelled using a logarithmic function [e.g., Marone et al., 1991], and I decide to describe the post-seismic source with this function. It is important

to underline that this choice is not critical for the goal of this Chapter. The main interest of this Chapter is to test the ability of vbICA to recover a transient signal, where, following the definition given by Riel et al. [2014], a transient deformation signal is a nonperiodic, nonsecular accumulation of strain in the crust. I am focusing only on a logarithmic function, but the vbICA method does not take advantage of this since it does not impose any constraint on the shape of the recovered source functions.

Let me write down more explicitly the mathematical expressions used to simulate the three seismic stages:

$$s^{lin}(x, y, z, t) = q^{lin}(x, y, z) + m^{lin}(x, y, z)t \quad (3.1a)$$

$$s^{co}(x, y, z, t, t^{co}) = A^{co}(x, y, z)H(t - t^{co}) \quad (3.1b)$$

$$s^{post}(x, y, z, t, t^{co}, \tau) = A^{post}(x, y, z)ln(1 + \frac{t - t^{co}}{\tau}) \quad (3.1c)$$

where  $t$  is time, and  $(x, y, z)$  is a point in the space.

The parameters that define the sources are 6, and they are separately defined for each patch  $p$  (i.e., the fault plane is discretized into several smaller rectangular patches), located at point  $(x, y, z)$ :

- 1)  $q^{lin}$ : arbitrary starting point of the creep value at time  $T_{start}$ ;
- 2)  $m^{lin} = \dot{s}^{lin}$ : slip rate;
- 3)  $A^{co}$ : co-seismic slip;
- 4)  $t^{co}$ : epoch of the earthquake;
- 5)  $A^{post}$ : post-seismic amplitude;
- 6)  $\tau$ : post-seismic decay time.

Varying these 6 parameters I can vary the SSR between the different sources related to the seismic cycle. In particular, I maintain the same value for the following three parameters:  $q^{lin} = 0$  mm,  $t^{co} = 1096$  d (= 3 yr) and  $\tau = 1$  d. I limit the cases under study allowing the remaining three parameters to assume the values shown in Table 3.1.

The co-seismic and post-seismic amplitudes are intended to vary in order to generate events with different energies. Assuming that the fault under study is embedded in a homogenous and elastic half-space, then the seismic moment associated to an earthquake (or the equivalent seismic moment associated to afterslip) can be calculated using the formula:

$$M_0 = \mu \mathcal{A} \delta \quad (3.2)$$

Table 3.1: Source parameters case study

Parameter	Values			Unit of measurement
$m^{lin}$	2	12	60	mm/yr
$A^{co}$	400		200	mm
$A^{post}$	30.3	15.2	9.1	mm

where  $\mu$  is the rigidity modulus and  $\mathcal{A}$  is the area that slips an amount  $\delta$  during the co-seismic period (or during the post-seismic period). Equation 3.2 shows that the energy released by an earthquake depends on three parameters. In all my simulations, I keep constant the rigidity modulus  $\mu = 30$  GPa (a typical value for the crust, e.g. Kanamori and Brodsky [2004]), and I vary the remaining two parameters. For the purposes of this Chapter, it is sufficient to use uniform slip distributions, so the co-seismic and post-seismic amplitudes take only the values shown in Table 3.1. The different scenarios generated comprise three options for the energy released by the earthquake: for the first scenario I consider an earthquake of moment magnitude greater than 6.5; for the second in the range [6.0, 6.5]; and for the third lower than 6.0. Since the moment magnitude is related to the seismic moment (in the International System of units) by the Hanks and Kanamori [1979] formula

$$M_W = \frac{2}{3} \text{Log}(M_0) - 6 \quad (3.3)$$

then I can vary the extension of the slipping portion to vary the moment magnitude of the generated events. Figure 3.1 shows the three different scenarios proposed. I use a fixed planar fault geometry, described by the 7 parameters of Table 3.2. The tectonic regime is set to simulate a thrust fault (rake  $-90^\circ$ , black arrows in Figure 3.1). The associated  $M_W^{co}$  are 6.85, 6.29, 5.94, respectively, and the corresponding equivalent  $M_W^{post}$  after 2 years are 6.57, 6.09 and 5.80.

Table 3.2: Fault plane geometry parameters

$x_{\text{top centre}}$	0	km
$y_{\text{top centre}}$	0	km
Length	46	km
$z_{\text{top}}$	-2	km
$z_{\text{bottom}}$	-26	km
Strike	45	$^\circ$
Dip	40	$^\circ$

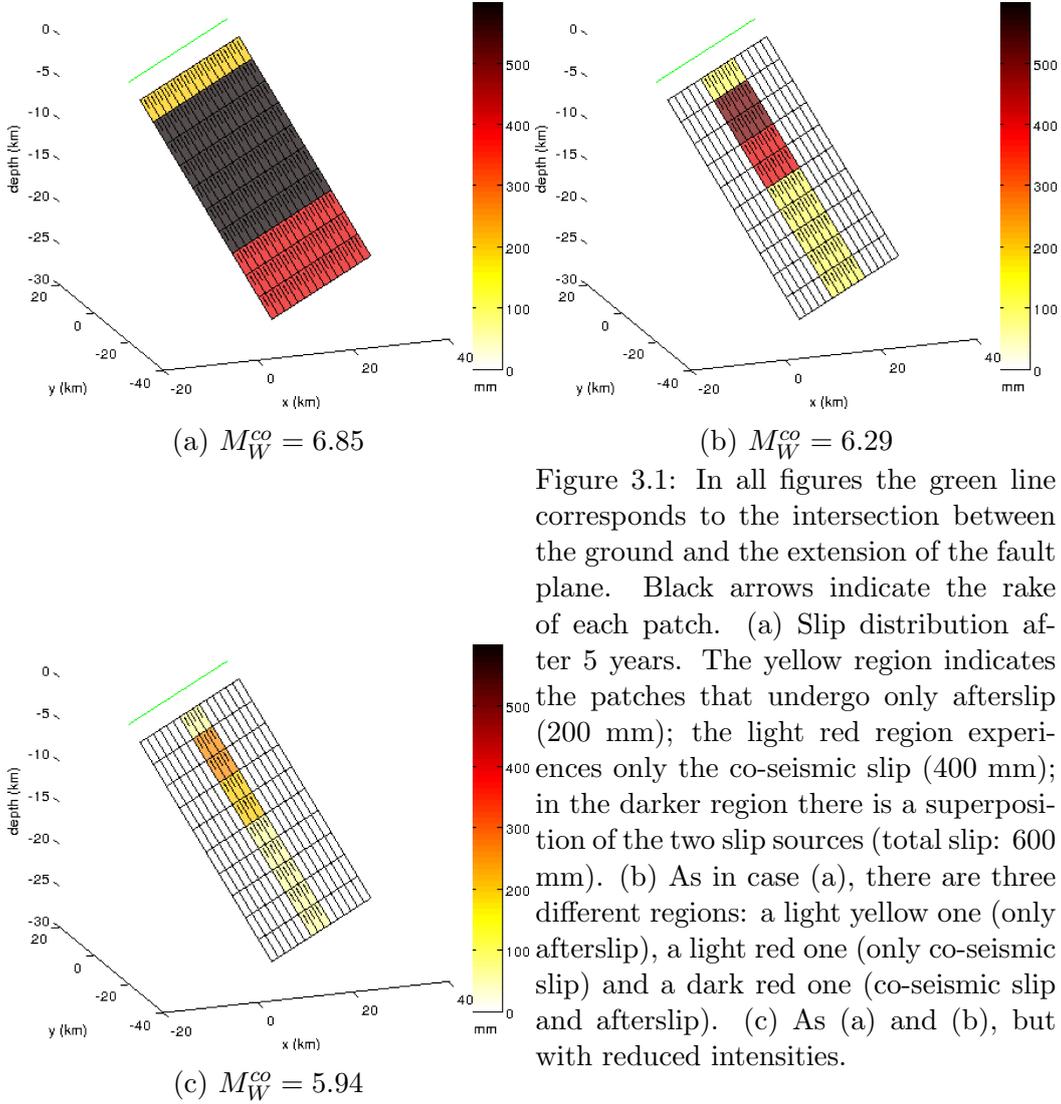


Figure 3.1: In all figures the green line corresponds to the intersection between the ground and the extension of the fault plane. Black arrows indicate the rake of each patch. (a) Slip distribution after 5 years. The yellow region indicates the patches that undergo only afterslip (200 mm); the light red region experiences only the co-seismic slip (400 mm); in the darker region there is a superposition of the two slip sources (total slip: 600 mm). (b) As in case (a), there are three different regions: a light yellow one (only afterslip), a light red one (only co-seismic slip) and a dark red one (co-seismic slip and afterslip). (c) As (a) and (b), but with reduced intensities.

The  $z_{bottom}$  parameter (see Table 3.2) is taken as representative of the locking depth, and the creeping portion of the fault (not shown in Figure 3.1), located at depths greater than  $z_{bottom}$ , is slipping at a constant rate  $m^{lin}$ . Its extension along strike is much bigger (100 times) than the locking depth, and it is 1000 km depth. It is worth considering that, although it is not the most sophisticated representation of the earthquake cycle, it is sufficient for my purposes to generate a secular trend in the data as I have already described.

Now that I have generated the slip evolution on each patch of the fault, I propagate it up to the surface to generate the displacements at specified locations (see Section 3.1.2). To perform this propagation I use the Green's

functions by Okada [1985], represented by the matrix  $\mathbf{G}$ , whose element  $G_{pj}$  indicates the displacement at the station  $j$  due to a 1 meter slip at patch  $p$ . Indicating with  $\mathbf{S}^{slip}$  the sources that generate slip, then the displacement at the given stations is given by:

$$\mathbf{X}_{M \times T}^{slip} = \mathbf{G}_{M \times P} \mathbf{S}_{P \times T}^{slip} \quad (3.4)$$

where  $M$  and  $T$  are defined in Section 3.1, and  $P$  is the total number of patches.

### Seasonal source

Another typical signal that characterizes GNSS time series data is the so-called seasonal signal. It is usually modelled by sinusoidal functions with different periods and amplitudes. For the purpose of this Chapter, I introduce only one oscillatory source signal, having two different amplitudes for the horizontal and vertical components. The general form of the observed signal is given by the following equation:

$$x^{sin}(x, y, t) = A^{sin}(x, y) \sin(\nu^{sin} t + \varphi^{sin}) \quad (3.5)$$

where  $A^{sin}(x, y)$  is the amplitude of the signal which depends on the position at the surface  $(x, y)$ ,  $\nu^{sin}$  is the frequency of the signal and  $\varphi^{sin}$  is the phase of the signal. I assume fixed values for  $\nu^{sin} = 2\pi$  1/yr and  $\varphi^{sin} = 0$ . Furthermore, I assume that the amplitude parameter depends on  $(x, y)$  as a bidimensional Gaussian having its centre in  $(x^{sin}, y^{sin}) = (-100, -100)$  km, having a maximum value of 1.5 and 3 mm for the horizontal and vertical components, respectively:

$$x^{sin}(x, y, t) = A_{max}^{sin} \frac{1}{\sqrt{2\pi\sigma_{sin}^2}} e^{-\frac{(x-x^{sin})^2+(y-y^{sin})^2}{2\sigma_{sin}^2}} \sin(2\pi t) \quad (3.6)$$

where  $\sigma_{sin}$  is a parameter that controls the decay of the amplitude, and it is maintained fixed during all the simulations.

From equation 3.6 it is possible to see that the spatial and temporal dependencies are separated, and that the role of the source signal is played by the function  $\sin(2\pi t)$ , while the corresponding column of the mixing matrix is given by the modulation of the amplitude:

$$\mathbf{X}_{M \times T}^{sin} = \mathbf{A}_{M \times 1}^{sin} \mathbf{S}_{1 \times T}^{sin} = \left\{ \left[ \begin{array}{c} A_E^{sin}(x_j, y_j) s^{sin}(t) \\ A_N^{sin}(x_j, y_j) s^{sin}(t) \\ A_V^{sin}(x_j, y_j) s^{sin}(t) \end{array} \right] \right\} \quad (3.7)$$

where E, N and V stand for east, north and vertical,  $j = 1, \dots, \frac{M}{3}$ , and  $t = 1, \dots, T$ .

## Volcanic source

In order to simulate the volcanic activity I use a Mogi source located at 5 km of depth, NW w.r.t. the fault used to mimic the seismic cycle (40 km to the W and 60 km to the N w.r.t. the projection to the surface of the centre of the top of the fault used). I have used a linear combination of arctangent functions to reproduce the change of volume  $\mathcal{V}$  with time  $t$  associated with the inflation and deflation of the magma chamber. In particular, the formula used is the following:

$$\mathcal{V}(t; a_{infl}, t_{infl}, a_{defl}, t_{defl}) \propto \frac{\text{atan}[a_{infl}(t - t_{infl})]}{\max\{\text{atan}[a_{infl}(t - t_{infl})]\}} - \frac{\text{atan}[a_{defl}(t - t_{defl})]}{\max\{\text{atan}[a_{defl}(t - t_{defl})]\}} \quad (3.8)$$

The values used for the 4 parameters are:  $a_{infl} = 0.01$  1/yr,  $t_{infl} = 4$  yr,  $a_{defl} = 0.05$  1/yr,  $t_{defl} = 4.5$  yr. The maximum volume variation corresponds to  $\sim 4.9 \times 10^4$  mm<sup>3</sup>, and occurs at  $\sim 4.3$  yr from the beginning of the analysis. The displacements at the surface have been calculated using the code of François Beauducel (<http://www.ipgp.fr/~beaudu/matlab.html#Mogi>, last access 10 March 2015).

---

All the described signals simulated at the station positions have a spatial and a temporal signature at the Earth's surface. In general, an observed signal is described using a function of both position of the sensor,  $(x, y, z)$ , and time,  $t$ . For all the sources used, I have assumed that it is possible to separate the spatial and temporal dependencies, i.e. I can write:

$$s_i(x, y, z, t, \boldsymbol{\beta}_i) = s_i^s(x, y, z, \boldsymbol{\beta}_i^s) s_i^t(t, \boldsymbol{\beta}_i^t) \quad i = 1, \dots, N_{sources} \quad (3.9)$$

where  $\boldsymbol{\beta}_i = \{\boldsymbol{\beta}_i^s, \boldsymbol{\beta}_i^t\}$  is the parameter vector containing the spatial and temporal parameters relative to source  $i$ , and  $N_{source}$  is the total number of sources used.

This assumption is forced by the nature of the static ICA algorithm that I am testing. Solving the BSS problem with a linear combination of sources does not allow me to determine the proper intensity of the sources. Indeed, any multiplication by a factor  $\alpha$  of the mixing matrix can be adjusted dividing by the same factor  $\alpha$  the source matrix. It follows that any (finite) intensity for the sources is admissible, and what I am going to recover is only the temporal evolution of the sources, i.e.  $s_i^t(t, \boldsymbol{\beta}_i^t)$ ,  $i = 1, \dots, N_{sources}$ .

From now on, if not differently specified, I will refer to the source signals as to the temporal part of the sources  $s_i(x, y, z, t, \boldsymbol{\beta}_i)$ , i.e.  $s_i^t(t, \boldsymbol{\beta}_i^t)$ , with  $i = 1, \dots, N_{sources}$ . The temporal parameters used are  $t^{co}$ ,  $\tau$ ,  $t_{infl}$ , and  $t_{defl}$ , and all of them are fixed to a predetermined value, as specified previously in this Section.

Accordingly, I can represent the observed signals  $\mathbf{X}(x, y, t)$  using two different plots: one that shows the spatial pattern at the surface (i.e., the mixing matrix values  $\mathbf{A}(x, y)$ ), and one that shows the evolution with time (i.e., the source signals, or their temporal evolution with time,  $\mathbf{S}(t)$ ). The plot relative to the spatial extent of the observed signals can be interpreted as the spatial distribution of the relative contribution to each station given by a certain source. I will often refer to  $\mathbf{A}$  as to the spatial pattern of the source, instead of the spatial pattern of the observed signal. The actual spatial distribution of the source might be obtained solving the corresponding inverse problem related to these new data, i.e. the columns of the mixing matrix:

$$\bar{A}(x, y) = g(\bar{S}^s(x, y, z)) \quad (3.10)$$

where  $\bar{\cdot}$  indicates the vectorization of the corresponding matrix, and  $g$  is the function that relates the model to recover (the spatial distribution of the sources  $\bar{S}^s(x, y, z)$ ) and the observed data (the spatial distribution corresponding to the mixing matrix  $\bar{A}(x, y)$ ).

Figures 3.2, 3.3, 3.4, 3.5 and 3.6 show the plots of the actual  $\mathbf{A}$  and  $\mathbf{S}$  described previously in this Section in a region of  $200 \times 200$  km<sup>2</sup> around the projection onto the surface of the mid-point of the planar fault's top edge. From now on, in this Chapter I use the  $\hat{\cdot}$  symbol to specify the quantities obtained from the vbICA decomposition:  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{S}}$ , and  $\hat{\mathbf{N}}$ .

### 3.1.2 Description of the simulated network of sensors

In previous Section I have described the sources used to generate the data. Now I am going to describe both the sensor location and the station quality (specified in terms of % of missing data, rather than noise).

Real world data offer a wide casuistry, and I attempt to reproduce only a few of the possible conditions in terms of network geometry and data quality. In particular, I test:

- 1) two different geometries,
- 2) three different data missing configurations.

Moreover, it has been shown [Langbein, 2008] that GPS data include also white and coloured noise, in particular pink (or flicker) and red (or

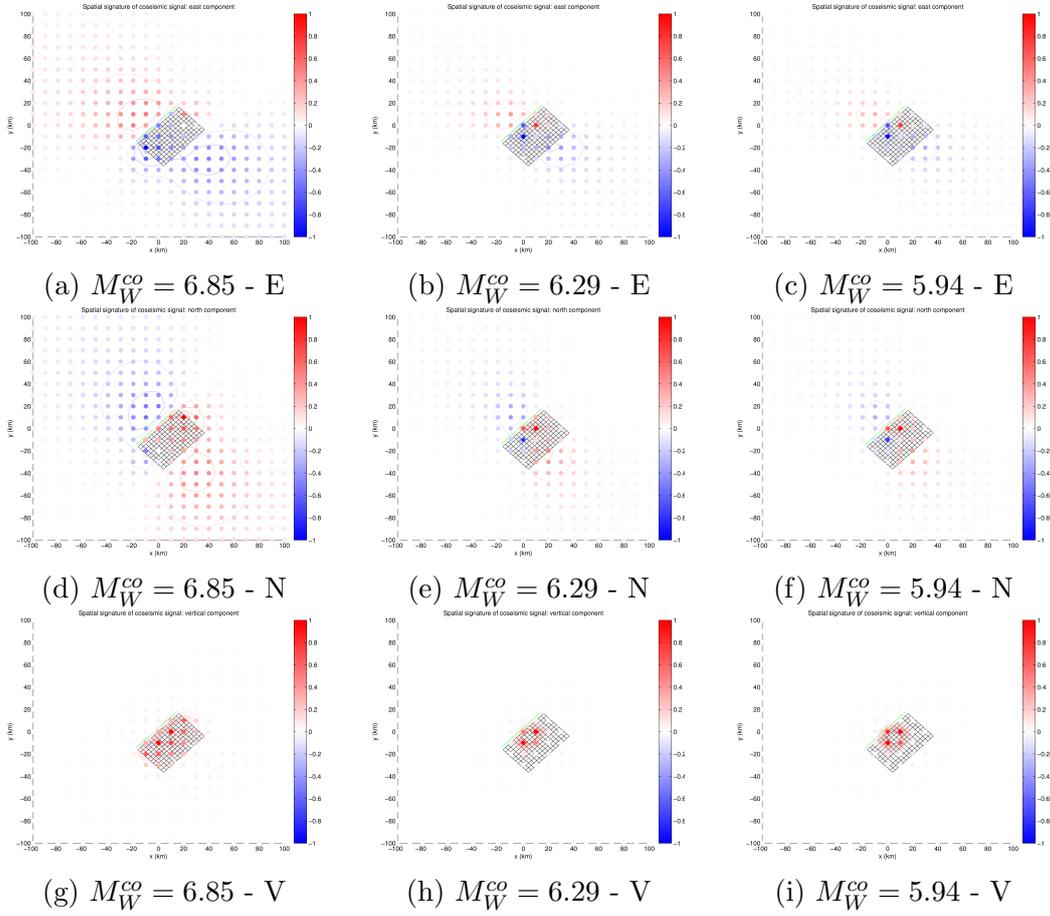


Figure 3.2: Spatial response of the co-seismic source at the surface, for all the study cases. E, N and V stand for east, north and vertical.

Brownian or random-walk). I simulate these three noises generating them one time, and fixing them to a determined value. In order to vary the SNR I prefer to control the signal power through the source intensities, since the noise generation is performed randomly and it is out of control.

The two different geometries tested are shown in Figure 3.7, and I refer to them as Network 1 (N1) and Network 2 (N2).

The possibility of a GPS network to record a particular source signal depends on the geometry of the network w.r.t. the position of the source and the source intensity. Here, since I am simulating deformation at the Earth's surface mainly due to earthquakes related processes, I refer to the inter-seismic locking depth as a reference distance to define a geometric quality of a GPS network. In particular, the first geometry tested has half of the total number of stations (i.e., 10) located in a square box of side equal to twice the

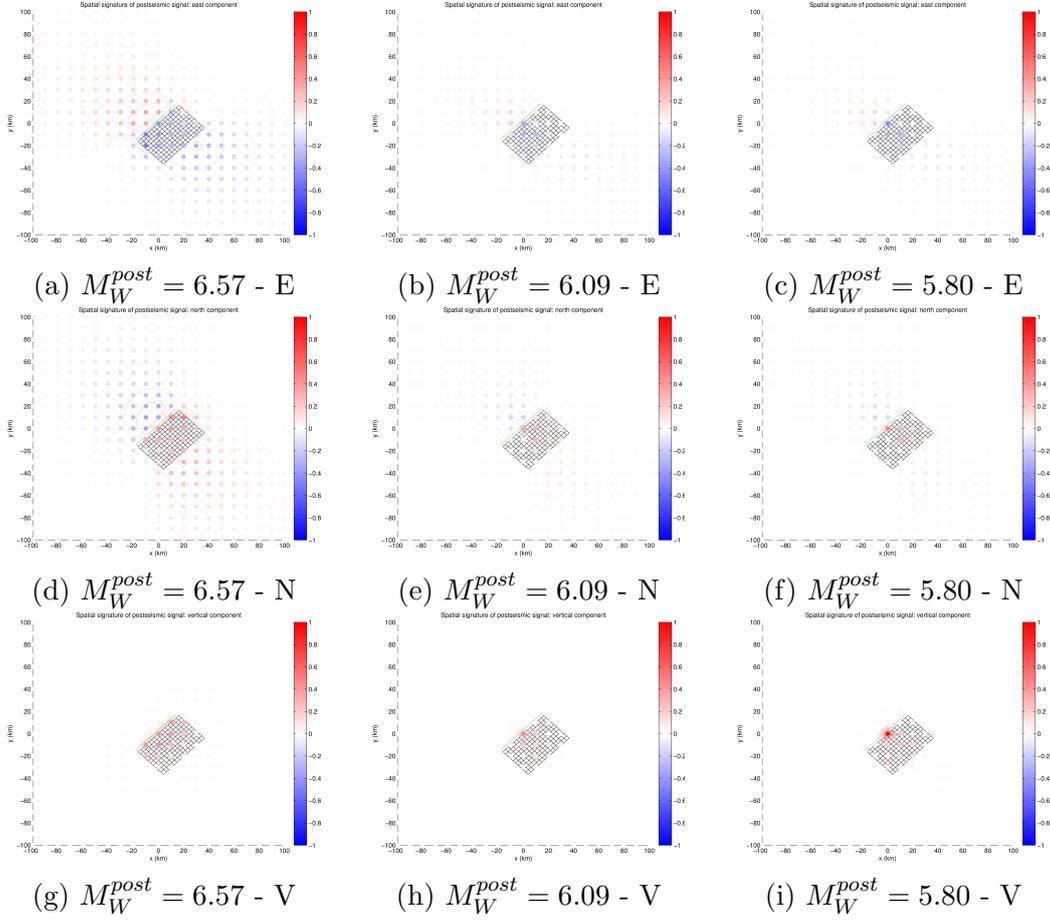


Figure 3.3: Spatial response of the post-seismic source at the surface, for all the study cases. E, N and V stand for east, north and vertical.

locking depth, and centered at the surface projection of the mid-point of the upper bound of the fault plane. The second network geometry consists in 20 stations, but this time with only  $\frac{1}{10}$  of the total number of stations located in the box described above.

The generation of gaps (missing data) in the simulated position time series is performed knowing that in real world data most of the times GPS time series present broad empty gaps, mainly due to malfunctioning or theft of instrumentation, rather than MCAR data. For this reason, I randomly create a bunch of gaps with different lengths, and delete the data associated with these gaps. After a first random generation to create a 5% and a 25% missing data dataset, I hold these two masks and maintain them for the creation of the further datasets obtained modifying the source intensities. I do this in order to avoid to continuously generate random gaps, that would

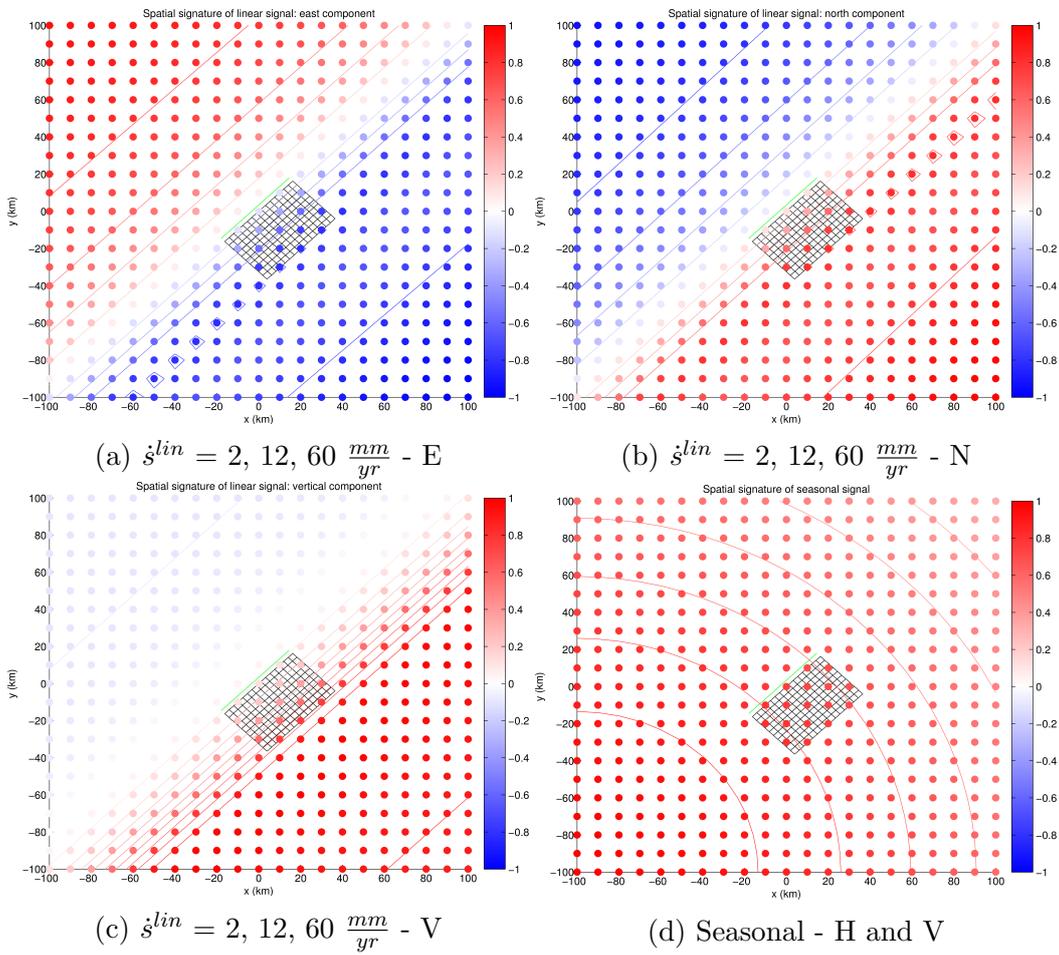


Figure 3.4: Interseismic and seasonal spatial responses for all the study cases.

be again a factor out of control during the analysis.

### 3.1.3 Logic tree of the synthetic data

In order to show more clearly the logic behind all the tests, Figure 3.8 illustrates all the case studies taken into account. In addition to the cases shown in Figure 3.8, I perform a simulation taking into account the co- and post-seismic signals, the seasonal signal, and the volcanic signal. For this particular situation I have used only the values relative to the network N2,  $M_W^{co} = 6.85$ ,  $M_W^{post} = 6.57$ , and 5% of missing data.

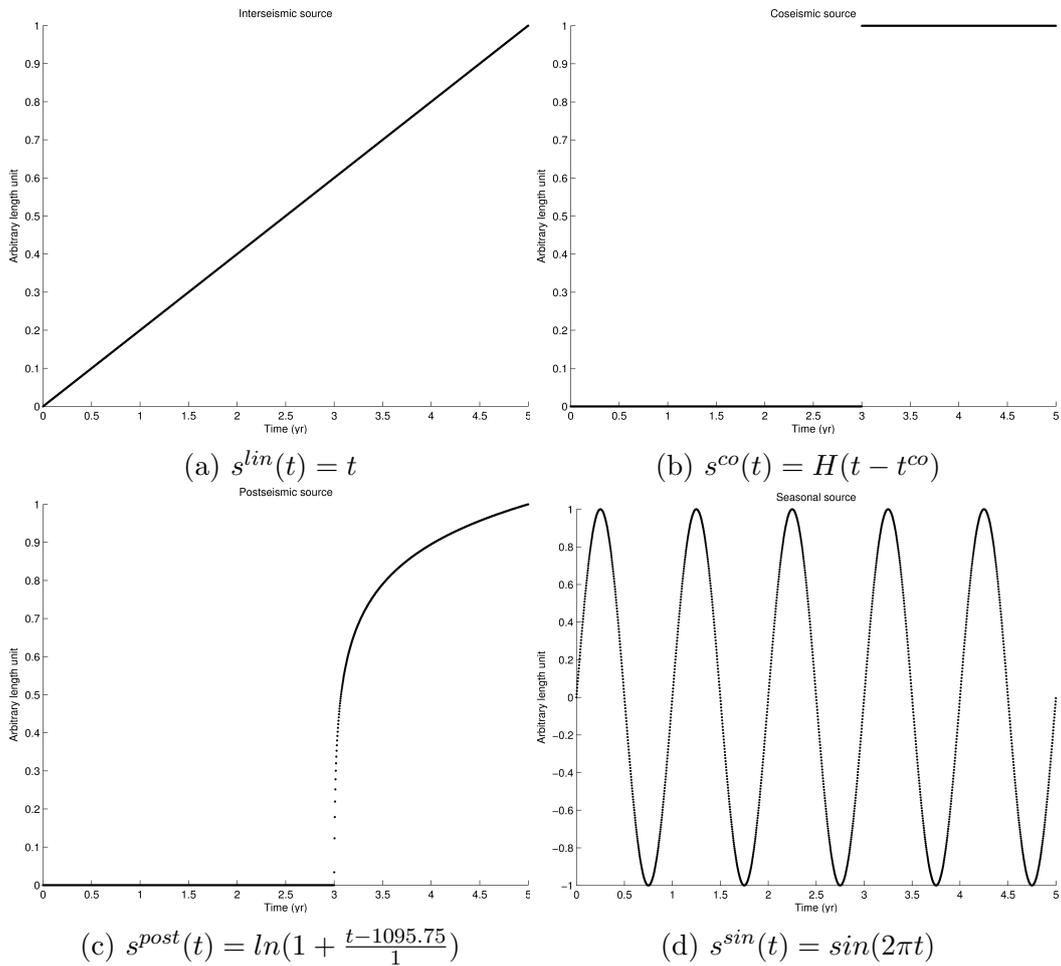


Figure 3.5: Temporal signatures of the displacement at the surface, that are equivalent to the sources used to generate the observed signal.

## 3.2 Results

Before proceeding with the exposition of the results, it is necessary to point out that two of the sources described in Section 3.1.1 are not independent. These two sources are the co-seismic and the post-seismic ones. Indeed, from the temporal evolution of the second we can deduce something of the temporal evolution of the first, and vice-versa. Let us imagine that we know the value of the post-seismic source at a certain epoch. If it is equal to 0 (or the reference point before the earthquake), then we immediately know the value of the co-seismic source at the same time; if it is different from the pre-earthquake value, then again we know the value of the co-seismic source at that epoch. It follows that the probabilities of the two sources do not

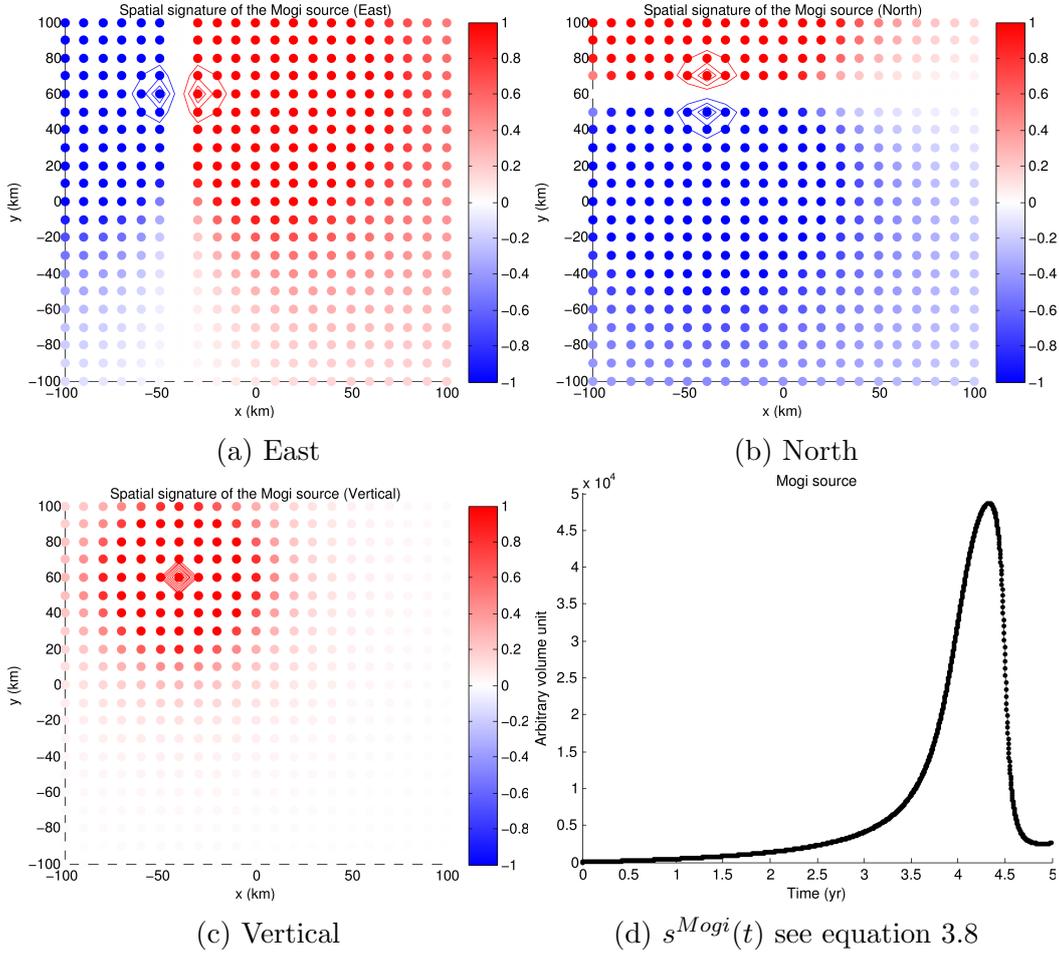


Figure 3.6: Mogi source spatial and temporal responses. The maximum displacement associated to the volcanic source recorded at one of the stations is  $\sim 13$  *mm*.

factorize:  $p(s^{co}, s^{post}) \neq p(s^{co})p(s^{post})$ . This is a problem for the application of the ICA algorithm, since it is based on this factorization. Accordingly, I decide to pre-process the time series correcting them for the co-seismic jump recovered using a PCA with the same number of components used for the subsequent ICA. In so doing, I assume that the PCA reproduces satisfactorily well the observed (simulated) time series data using few components, and I minimize the effect of the noise in the determination of such a signal. This, unfortunately, is not always the case, since sometimes I handle cases having a very low SNR (see Tables 3.3, 3.4, 3.5 and 3.6 for the SNR values of each source and each configuration studied). Moreover, also the geometry configuration of the network affects the SNR recorded. Indeed, for the three

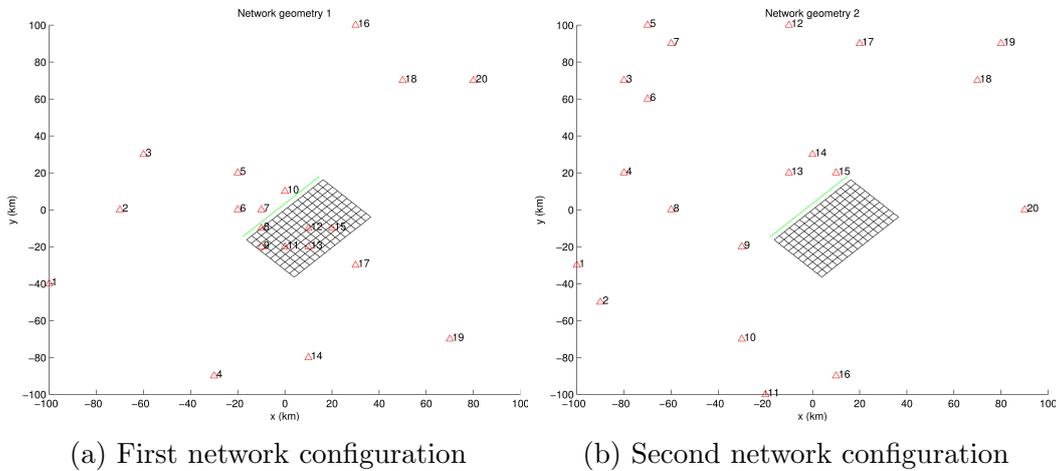


Figure 3.7: Network configurations.

co-seismic sources tested, the percentage of time series having a  $\text{SNR} > 1$  is 70%, 45% and 23% for the network configuration N1, but these values drop to 56%, 10% and 3% in case of network N2. The same argumentation is valid for the post-seismic signal, passing from values of 63%, 28% and 8% to values of 28%, 3% and 0%. The linear and seasonal signal are less affected by the network configuration since they are less localized than the signals related to the seismic event. Indeed, for the linear signals tested this percentage is about 5%, 60% and 95% for the three creep rates adopted, while for the seasonal signal it is 28-30% for all the configurations. Finally, for the case with also a Mogi source, the SNR related to the volcanic signal is greater than 1 for 6% of the stations. In critical situations (i.e., less than 5% of the stations show a  $\text{SNR} > 1$  and those few stations do not have a relevant SSR compared to other signals) it is a good idea to use ad hoc techniques, in place of multivariate statistical ones. Unfortunately, in real data applications we do not know the actual SNR value and we have to figure out what are the relevant signals recorded by the data.

Bearing this in mind, the total number of cases studied is  $55 = 2$  network configurations  $\times$  3 missing data masks  $\times$  3 earthquake energy release  $\times$  3 inter-seismic slip rates + 1 with a Mogi source. Before moving on, it is important to notice that for high values of the inter-seismic slip rate the linear signal becomes dominant, and the entire dataset is strongly correlated. In other words, the cloud of points in the data space is strongly aligned, and this makes the search of the IC directions difficult. Choudrey [2002] pointed out that “as a particular cluster becomes more and more correlated (or, equivalently, its independent directions become less orthogonal), the second

	$\dot{s}^{lin} = 2 \frac{mm}{yr}$			$\dot{s}^{lin} = 12 \frac{mm}{yr}$			$\dot{s}^{lin} = 60 \frac{mm}{yr}$			
	MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%	
$M_W^{co} = 6.85$	N1	0.26	0.26	0.24	9.45	9.50	8.58	236.32	237.54	214.40
	N2	0.22	0.21	0.19	7.97	7.73	6.84	199.17	193.29	170.91
$M_W^{co} = 6.85$	N1	0.26	0.26	0.24	9.45	9.50	8.58	236.32	237.54	214.40
	N2	0.22	0.21	0.19	7.97	7.73	6.84	199.17	193.29	170.91
$M_W^{co} = 6.85$	N1	0.26	0.26	0.24	9.45	9.50	8.58	236.32	237.54	214.40
	N2	0.22	0.21	0.19	7.97	7.73	6.84	199.17	193.29	170.91

Table 3.3: Linear SNR values

	$\dot{s}^{lin} = 2 \frac{mm}{yr}$			$\dot{s}^{lin} = 12 \frac{mm}{yr}$			$\dot{s}^{lin} = 60 \frac{mm}{yr}$			
	MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%	
$M_W^{co} = 6.85$	N1	339.58	324.46	328.63	339.58	324.46	328.63	339.58	324.46	328.63
	N2	14.97	14.20	13.16	14.97	14.20	13.16	14.97	14.20	13.16
$M_W^{co} = 6.85$	N1	12.34	12.08	11.54	12.34	12.08	11.54	12.34	12.08	11.54
	N2	0.85	0.83	0.70	0.85	0.83	0.70	0.85	0.83	0.70
$M_W^{co} = 6.85$	N1	1.30	1.28	1.22	1.30	1.28	1.22	1.30	1.28	1.22
	N2	0.08	0.08	0.06	0.08	0.08	0.06	0.08	0.08	0.06

Table 3.4: Co-seismic SNR values

		$\dot{s}^{lin} = 2 \frac{mm}{yr}$			$\dot{s}^{lin} = 12 \frac{mm}{yr}$			$\dot{s}^{lin} = 60 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	64.30	63.35	63.04	64.30	63.35	63.04	64.30	63.35	63.04
	N2	4.43	4.19	3.37	4.43	4.19	3.37	4.43	4.19	3.37
$M_W^{co} = 6.85$	N1	2.22	2.11	1.84	2.22	2.11	1.84	2.22	2.11	1.84
	N2	0.10	0.09	0.07	0.10	0.09	0.07	0.10	0.09	0.07
$M_W^{co} = 6.85$	N1	0.31	0.29	0.26	0.31	0.29	0.26	0.31	0.29	0.26
	N2	0.013	0.012	0.009	0.013	0.012	0.009	0.013	0.012	0.009

Table 3.5: Post-seismic SNR values

		$\dot{s}^{lin} = 2 \frac{mm}{yr}$			$\dot{s}^{lin} = 12 \frac{mm}{yr}$			$\dot{s}^{lin} = 60 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	0.98	0.98	0.97	0.98	0.98	0.97	0.98	0.98	0.97
	N2	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
$M_W^{co} = 6.85$	N1	0.98	0.98	0.97	0.98	0.98	0.97	0.98	0.98	0.97
	N2	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
$M_W^{co} = 6.85$	N1	0.98	0.98	0.97	0.98	0.98	0.97	0.98	0.98	0.97
	N2	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11

Table 3.6: Seasonal SNR values

moment starts to dominate so the cluster becomes easier to describe by a Gaussian density. [...] The limit of correlation that vbICA can successfully resolve seems to be about 0.67.” He suggested two possible solutions to work around this problem:

- a) to use a random initialization point instead of the SVD decomposition;
- b) to decorrelate or “whiten” (also “sphere”) the observation data.

I try both these solutions for the highly correlated cases ( $\dot{s}^{lin} = 12$  and  $60 \frac{mm}{yr}$ ). Using a random initialization helps only in few cases, and it is computationally intensive, since it is necessary to perform a number of decompositions sufficient to allow the algorithm to find the proper path to follow. As a first step, I start analysing only the  $\dot{s}^{lin} = 2 \frac{mm}{yr}$  case. After that, I move on considering the possibility to reduce the correlation of the other two scenarios. Finally, I show the results concerning the case with also a Mogi source.

### 3.2.1 Low tectonic rate: $\dot{s}^{lin} = 2 \frac{mm}{yr}$

The first thing to do is to determine the number of ICs that I am going to use to analyse the data. I compare three different methods: 1) a  $\chi^2$  test, 2) a  $F$ -test, and 3) the ARD criterion described in Chapter 2.

The  $\chi^2$  test and the  $F$ -test are based on the definition of the following measure:

$$\chi^2 = \sum_{j=1}^M \sum_{t=1}^T w_{jt} (x_{jt} - x_{jt}^{dec})^2 \quad (3.11)$$

where  $w_{jt}$  is the weight associated to the measure  $x_{jt}$ , i.e.  $w_{jt} = \frac{1}{\sigma_{jt}^2}$ , where  $\sigma_{jt}$  is the error (or standard deviation) of the  $j$ -th time series at epoch  $t$ . As expected, the  $\chi^2$  assumes the lowest value for the PCA reconstruction, since it is exactly the quantity that PCA minimises. The loss in  $\chi^2$ , that is the lower accuracy in reconstructing the data using a  $L_2$ -norm, is the reckoning for the higher accuracy in the separation of the original sources. Anyway, this is not always the case if we consider datasets with missing data. In 8 over 12 cases with missing data in the time series, the vbICA algorithm reconstructs better than the PCA the data gaps, as testified by the  $\chi^2$  values computed using the unmasked data and the PCA/ICA decompositions obtained masking the data (see Table 3.7 and 3.8, second and third columns relative to the  $\dot{s}^{lin} = 2$  mm/yr case).

A  $\chi^2$  test states if a model is explaining satisfactorily well the data, in the light of the error variance. In practice, if the reduced  $\chi^2$ , defined as

		$\dot{s}^{lin.} = 2 \frac{mm}{yr}$			$\dot{s}^{lin.} = 12 \frac{mm}{yr}$			$\dot{s}^{lin.} = 60 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	4.8%	3.4%	5.2%	5.7%	4.1%	6.2%	5.6%	4.0%	6.2%
	N2	1.9%	1.0%	6.1%	3.6%	3.6%	5.5%	2.2%	2.4%	7.6%
$M_W^{co} = 6.85$	N1	2.3%	2.6%	5.5%	2.3%	2.5%	NaN	2.3%	2.4%	4.3%
	N2	1.6%	1.9%	2.9%	9.3%	9.5%	3.8%	4.6%	4.6%	3.8%
$M_W^{co} = 6.85$	N1	2.4%	2.6%	7.7%	3.6%	3.8%	NaN	4.9%	4.9%	5.8
	N2	2.1%	2.2%	7.6%	2.2%	2.2%	2.7%	2.0%	2.1%	2.6%

Table 3.7:  $\frac{\chi_{ICA}^2 - \chi_{PCA}^2}{\chi_{PCA}^2}$

		$\dot{s}^{lin.} = 2 \frac{mm}{yr}$			$\dot{s}^{lin.} = 12 \frac{mm}{yr}$			$\dot{s}^{lin.} = 60 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%	MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	4.8%	5.7%	38.1%	5.7%	6.4%	41.9%	5.6%	6.1%	44.7%
	N2	1.9%	1.0%	4.8%	3.6%	3.2%	20.1%	2.2%	1.8%	-98.6%
$M_W^{co} = 6.85$	N1	2.3%	2.4%	-12.4%	2.3%	2.2%	NaN	2.3%	2.2%	9.3%
	N2	1.6%	1.9%	0.9%	9.3%	10.9%	10.9%	4.6%	5.1%	5.0%
$M_W^{co} = 6.85$	N1	2.4%	2.5%	-95.6%	3.6%	3.9%	NaN	4.9%	5.2%	7.0%
	N2	2.1%	2.1%	-94.2%	2.2%	2.2%	0.8%	2.0%	2.0%	0.6%

Table 3.8:  $\frac{\chi_{ICA}^2 - \chi_{PCA}^2}{\chi_{PCA}^2}$  filling the gaps

$\chi_{red}^2 = \frac{\chi^2}{dof}$ , where  $dof$  are the degrees of freedom of the model under investigation, is  $\leq 1$ , than the model is complex enough, and it is not necessary to add further complications (i.e., we do not have to add more components to explain the data).

The  $F$ -test in regression problems says if, given two models one of which (model 1) nested in the other (model 2), the complexity of the more complex model is justified by the better explanation of the data. In other words, it tests if we have to add one component to our decomposition or not. The  $F$ -value derived from the data is defined as follows:

$$F = \frac{\frac{\chi_1^2 - \chi_2^2}{(N-p_1) - (N-p_2)}}{\frac{\chi_2^2}{N-p_2}} = \frac{\chi_1^2 - \chi_2^2}{\chi_2^2} \frac{N - p_2}{p_2 - p_1} \quad (3.12)$$

where  $\cdot_{1,2}$  indicates the quantity relative to model 1 or 2,  $N$  is the total number of data ( $N = MT - N_{missingdata}$ ),  $p_{1,2}$  is the number of parameters of the corresponding model. It follows that  $(N - p_1) - (N - p_2) = p_2 - p_1$  and  $N - p_2$  are the degrees of freedom of the numerator and the denominator, respectively. Once the  $F$ -value is calculated from the data, it is necessary to compare it with the  $F$  statistics at a certain confidence level, that is usually defined at 95% level.

Performing these tests for both the PCA and ICA decompositions, I find that in all cases the reduced  $\chi^2$  test shows that one component is sufficient to explain the data. This choice is clearly incorrect, since I have used more than one source to create the observed time series, and would prevent the detection and separation of the actual sources that generated the data. Consequently, it is necessary to use a different test. The  $F$ -test for the PCA decomposition shows that 3 components are not enough, and demands for more complex models. This means that it introduces also some noise into the reconstruction. This can be due to the fact that the signals under investigation have a  $SNR < 1$  in most of the cases, thus the noise pops up as a relevant signal, even if it does not show a common temporal evolution. Instead, the  $F$ -test performed on the vbICA decomposition says that 2 components are sufficient in all cases but the one with 0% of data missing, the network geometry configuration N1, and a co-seismic source corresponding to a  $M_W=6.85$  earthquake. In such a case, the  $F$ -test points out that the 3 components decomposition is the most appropriate. Finally, the ARD selection criterion picks the number of components illustrated in Table 3.9.

Given the low SNR of the linear signal ( $< 0.3$ ), it does not surprise that in most of the cases the linear signal is not recognised. Moreover, it is important to point out that the ARD method chooses 3 components when 25% of the data is missing. In all these cases, one of the 3 ICs contains

		$\dot{s}^{lin} = 2 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	2	2	2
	N2	2	2	3
$M_W^{co} = 6.29$	N1	2	2	3
	N2	2	2	2
$M_W^{co} = 5.94$	N1	2	2	3
	N2	2	2	3

Table 3.9: ARD - Number of components selection. It is indicated the number of components selected using the ARD criterion, for all the interseismic case  $\dot{s}^{lin} = 2 \frac{mm}{yr}$ , the 3 co-seismic and post-seismic cases (different values of  $M_W^{co}$ ), the 3 data missing masks (different values of MD), and the 2 network geometries (N1 and N2).

both the linear signal and a residual co-seismic offset. This means that the PCA reconstruction of the offset, and its consequent subtraction from the dataset, is not good enough, and it leaves in the processed data a residual offset. The SNR of this residual signal is stronger than, or comparable to, the SNR of the linear signal, and they are captured together. A similar problem affects also the two cases with  $M_W=5.94$ , where the post-seismic signal is detected but the number of ICs is 2 (namely, when the N1 configuration and the 0% and 5% of data missing masks are used, see Figures 3.13b and 3.13d). Moreover, looking at Figures 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14, we see that the smaller the post-seismic source, the higher the influence of the linear signal in the corresponding IC. This is due to the fact that only two ICs are used to decompose the data. This presence of two sources in one IC might introduce some error in the computation of the time decay constant of the post-seismic signal, as discussed later in Subsection **Post-seismic decay constant**.

## Comparing PCA and ICA

The strength of the ICA method consists in giving to the researcher the possibility to easily visualize the main ongoing processes affecting the data. Nevertheless, it is necessary to determine if a particular IC is properly representing the actual source or not. Indeed, looking at Figures 3.9-3.14, we can qualitatively say that the ICA decomposition is able to reproduce and isolate the post-seismic decay and the seasonal signal, but we need a more objective way to quantify how the ICA is able to recover the signals.

In order to compare the PCA and ICA algorithms, it is necessary to

compare the PCA eigenvectors and the ICA sources assuming a common normalization. As it can be seen by Definitions 1.1.1 and 2.1, a PCA and a ICA decomposition are equivalence classes of decompositions, and it is necessary to impose some constraints in order to have a unique PCA and ICA. These constraints have been specified for the PCA decomposition (see Section 1.1.1), and, since it is always possible to decompose a matrix in a unit column norm matrix and a diagonal matrix, in order to compare the  $\mathbf{X}_{ICA}^{dec}$  matrix to the SVD decomposition, I rewrite it as follows<sup>1</sup>:

$$\mathbf{X}_{ICA}^{dec} = \hat{\mathbf{A}}\hat{\mathbf{S}} = \mathbf{U}_{ICA}\boldsymbol{\Sigma}_{A_{ICA}}\boldsymbol{\Sigma}_{S_{ICA}}\mathbf{V}_{ICA}^T = \mathbf{U}_{ICA}\boldsymbol{\Sigma}_{ICA}\mathbf{V}_{ICA}^T \quad (3.13)$$

where  $\hat{\mathbf{A}} = \mathbf{U}_{ICA}\boldsymbol{\Sigma}_{A_{ICA}}$ ,  $\hat{\mathbf{S}} = \boldsymbol{\Sigma}_{S_{ICA}}\mathbf{V}_{ICA}^T$ ;  $\mathbf{U}_{ICA}$  and  $\mathbf{V}_{ICA}^T$  are two unit column norm matrices;  $\boldsymbol{\Sigma}_{A_{ICA}}$  and  $\boldsymbol{\Sigma}_{S_{ICA}}$  are two diagonal matrices, as well as  $\boldsymbol{\Sigma}_{ICA}$ . Adopting this decomposition, I can directly compare the  $\mathbf{U}_{PCA/ICA}$  and  $\mathbf{V}_{PCA/ICA}$  matrices of the respective decompositions with the normalised spatial and temporal parts of the actual sources,  $\mathbf{U}$  and  $\mathbf{V}$ .

The reconstruction error is quantified by the mean-square error (MSE) between the reconstructions and the original sources:

$$\text{MSE} = \frac{1}{L} \sum_{i=1}^L \text{MSE}_i = \frac{1}{LT} \sum_{i=1}^L \sum_{t=1}^T (\hat{v}_{it} - v_{it})^2 \quad (3.14)$$

where  $L$  is the number of sources,  $T$  the number of epochs,  $\mathbf{V}_{T \times L}$  is the normalised actual source matrix and  $\hat{\mathbf{V}}_{T \times L}$  is the reconstructed source matrix. This quantity measures how far (in a Euclidean sense) the PCs and the ICs are from the original sources. A smaller MSE value correspond to a better reconstruction of the sources. We can also check how good is the reconstruction of a particular source using the  $\text{MSE}_i$ , defined as the MSE relative to the  $i$ -th component:  $\frac{1}{T} \sum_{t=1}^T (\hat{v}_{it} - v_{it})^2$ .

The residual cross-talk ( $Xtalk$ ), defined as:

$$Xtalk = \frac{1}{T(L^2 - L)} \sum_{i=1}^L \sum_{j \neq i=1}^L \sum_{t=1}^T |\hat{v}_i v_j - v_i v_j| \quad (3.15)$$

measures the ability of the decomposition method to separate the sources. The correlation of the original sources signals is deducted to ensure their  $Xtalk$  is zero [Choudrey, 2002].

<sup>1</sup>Let us suppose that  $\mathbf{A}$  is a generic matrix of size  $M \times N$ . Let us consider the diagonal matrix  $\mathbf{D}$  of size  $N \times N$ , having as elements the norm of the columns of  $\mathbf{A}$ . If we divide each element of  $\mathbf{A}$  for the norm of the corresponding column, and we call this matrix  $\mathbf{B}$ , then the original matrix can be written as  $\mathbf{A} = \mathbf{B}\mathbf{D}$ .

Adopting the ARD criterion, the decomposition does not extract the linear trend source for 14 cases over 18 (see Table 3.9), not selecting the proper number of components. In such cases, i.e. when  $L = 2$ , in order to compute the MSE and  $Xtalk$  statistics I use only the two sources having the highest linear correlation with the corresponding ICs. In this way, I can objectively state that the linear source is the one among the three sources that has the lowest correlation with the ICs, and for this reason it is to be considered the one that has not been recovered.

Tables 3.10, 3.11, 3.12, show the MSE values obtained for all the synthetic tests, and the relative percentage gains obtained analysing the data with the ICA technique instead of the PCA one. We can also refer to only the post-seismic and seasonal signals. Tables 3.13 and 3.14 show the MSE relative to the individual ICs.

		$\dot{s}^{lin} = 2 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	0.0347	0.0366	0.0534
	N2	0.0598	0.0557	1.3585
$M_W^{co} = 6.29$	N1	0.0825	0.0864	1.3176
	N2	0.1602	0.1423	0.3092
$M_W^{co} = 5.94$	N1	0.7565	0.6588	2.9953
	N2	0.1897	0.1731	1.3235

Table 3.10:  $MSE_{PCA}$

		$\dot{s}^{lin} = 2 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	0.0241	0.0245	0.0310
	N2	0.0187	0.0242	1.0220
$M_W^{co} = 6.29$	N1	0.0254	0.0261	1.0129
	N2	0.0910	0.0742	0.2755
$M_W^{co} = 5.94$	N1	0.0405	0.0430	2.0322
	N2	0.1523	0.1281	2.1488

Table 3.11:  $MSE_{ICA}$

In all cases but one, the vbICA algorithm reduces the MSE, as expected. The only scenario where the MSE is better for a PCA is the worst possible case studied, i.e. the one with 25% of missing data, a small post-seismic source and an unfavourable GPS network. For this configuration, the SNR of the post-seismic source is equal to 0.009, resulting in a very noisy component.

		$\dot{s}^{ln} = 2 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	44%	49%	72%
	N2	221%	130%	33%
$M_W^{co} = 6.29$	N1	226%	230%	30%
	N2	76%	92%	12%
$M_W^{co} = 5.94$	N1	1767%	1430%	47%
	N2	25%	35%	-38%

Table 3.12:  $\frac{MSE_{PCA} - MSE_{ICA}}{MSE_{ICA}}$

		$\dot{s}^{ln} = 2 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	0.0019	0.0016	0.0019
	N2	0.0019	0.0064	0.0115
$M_W^{co} = 6.29$	N1	0.0039	0.0043	0.0053
	N2	0.0745	0.0571	NI
$M_W^{co} = 5.94$	N1	0.0176	0.0195	0.0796
	N2	NI	NI	1.9633

Table 3.13:  $MSE_{ICA}$  post-seismic signal. NI stands for Not Identified. See Sub-section **Post-seismic decay constant**.

		$\dot{s}^{ln} = 2 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	0.0222	0.0229	0.0291
	N2	0.0167	0.0178	0.0250
$M_W^{co} = 6.29$	N1	0.0214	0.0219	0.0275
	N2	0.0165	0.0171	0.0243
$M_W^{co} = 5.94$	N1	0.0229	0.0235	0.0289
	N2	0.0169	0.0188	0.0239

Table 3.14:  $MSE_{ICA}$  seasonal signal.

### Post-seismic decay constant

In order to quantitatively assess the proper values for the parameters of a deterministic temporal evolution function, I need to estimate the credibility interval of such parameters given a particular model. For the case under study, I want to verify if, given a logarithmic model to fit the IC associated with the post-seismic decay, I can properly deduce the value of the time decay

constant. For what said in Chapter 2, the ICs are not uniquely determined, and we can multiply them for an arbitrary factor. This means that the amplitude parameter is not constrained, and I can not compare it with the original value used to generate the data. Moreover, I removed the co-seismic signal during a preprocessing step, and it is possible that part of the co-seismic jump is still present in the recovered IC. For this reason, I use the following fitting model:

$$g(m_1, m_2, m_3, t) = m_1 + m_2 \ln\left(1 + \frac{t}{m_3}\right) \quad (3.16)$$

where the model space has 3 dimensions, and  $m_1$ ,  $m_2$  and  $m_3$  are the model parameters.

A great advantage of the vbICA w.r.t. other decomposition approaches is that it gives an approximation of the pdf of the ICs, i.e. of the sources. From this approximated pdf I can extract not only the value of the component at a particular epoch, but I can also calculate the corresponding variance. This means that we have an estimation of the uncertainty related to each epoch of the IC. This considerably helps in determining if a particular model  $f(t)$  is properly fitting the component. Indeed, let us consider the IC related to the post-seismic signal ( $IC_{post}$ ). For the structure of the vbICA algorithm, I assume that the component is represented by  $T$  random variables independent and identically distributed (iid). In other words, I am not assuming any temporal dependency between the value of the IC at time  $t$  and the value at time  $t + \delta t$ <sup>ii</sup>. A common simplification adopted in describing GPS time series is to assume that each observation has an associated error that is the square root of the variance of the measure, and that such a random variable is described by a Gaussian distribution. I assume the same for each of the values of the  $IC_{post}$ . To solve this fitting problem I use a Bayesian approach that gives me the credible volume for the parameter vector  $\mathbf{m} = [m_1, m_2, m_3]$  (see Appendix A). I explore the ranges  $[-0.03, 0]$ ,  $[\varepsilon, 0.04]$  and  $[\varepsilon, 50]$  for the three parameters  $m_1$  (offset constant),  $m_2$  (amplitude of the post-seismic source, or  $A$ ) and  $m_3$  (time decay of the post-seismic source, or  $\tau$ ), where  $\varepsilon \simeq 2.2 \times 10^{-16}$ . These ranges are chosen around the values found using an unconstrained non-linear minimization of the sum of squared residuals, and ensure me to explore all the regions of the model space where the likelihood function is significantly greater than 0<sup>iii</sup>. If a relevant portion of the model

---

<sup>ii</sup>An attempt to consider this dependency has been done by Choudrey [2002]. He developed a Dynamic ICA, based on Hidden Markov Models (HMMvbICA). It might be a good idea for the future to follow the same approach also for the analysis of geodetic data.

<sup>iii</sup>I say that the likelihood is significantly greater than 0 if even considering a  $m_3$  range 100 times wider, the sum of all the new added points investigated would not contribute

space having a likelihood function significantly different from 0 is not taken into account, than the uncertainty on the time decay parameter is too high, and such a parameter is not resolved. The grid step adopted for the three parameters is 0.005, 0.0001 and 0.5 respectively. The actual  $m_3$  value is 1 d (see Section 3.1.1), and the credible intervals at 68.27% and 99.99% are listed in Tables 3.15 and 3.16.

For 6 over 18 cases the  $m_3$  parameter is not resolved, and all the values spanned are acceptable or the post-seismic source has not been identified (NI). These 6 cases correspond to the N2 geometry scenarios for the intermediate and small post-seismic source intensities. This proves that it is necessary to have a good quality network if we want to study crustal displacements of the order of few mm with multivariate statistical techniques such as vbICA.

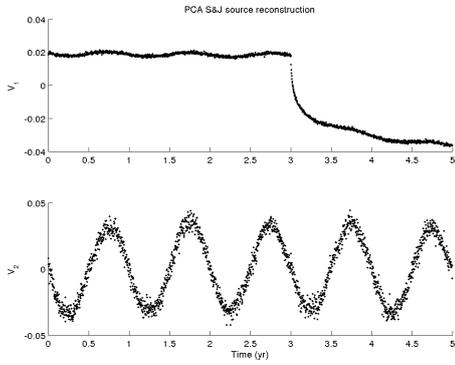
In more than one case I recover more than one credible interval. This is the direct consequence of the non linearity of the forward model. Indeed, the likelihood function has few local maxima, and I can not exclude the possibility that the proper model does not belong to the global maximum. Moreover, as every inverse problem, the solution is not unique, and to give only the maximum likelihood model could be misleading.

In all cases where the number of ICs used is 2, it is probably more correct to use a forward model (i.e., equation 3.16) that takes into account also the linear signal. Nevertheless, the number of parameters used in equation 3.16 seems to be already enough to correctly guess the decay constant at a 99.99% level.

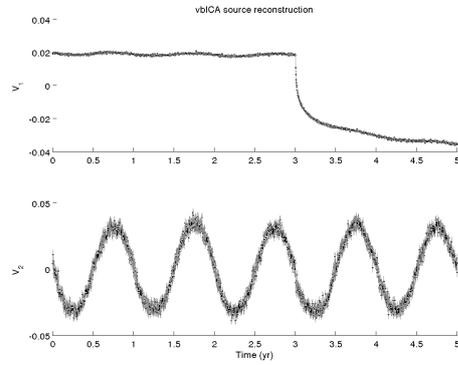
---

more than  $\frac{1}{10}$  of the contribution to the volume given by the point of maximum likelihood.

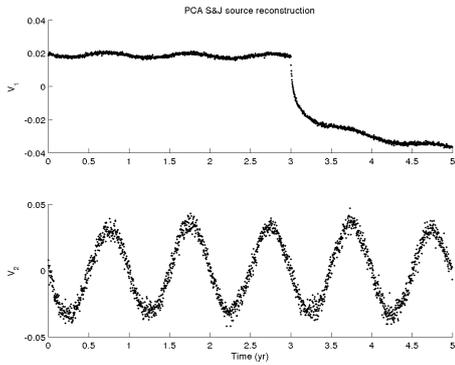




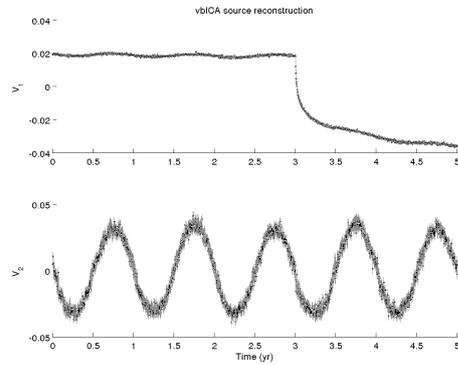
(a) PCA - MD 0%



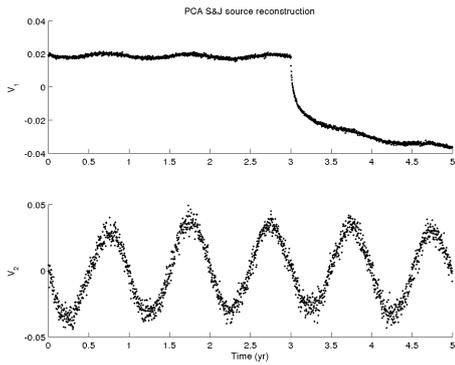
(b) vbICA - MD 0%



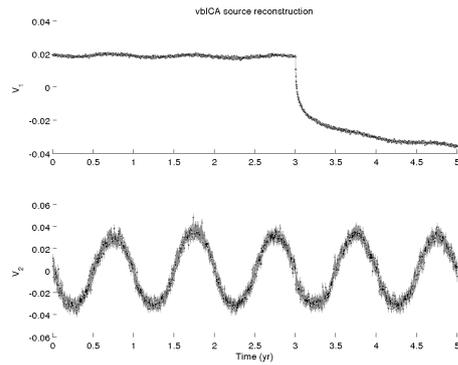
(c) PCA - MD 5%



(d) vbICA - MD 5%

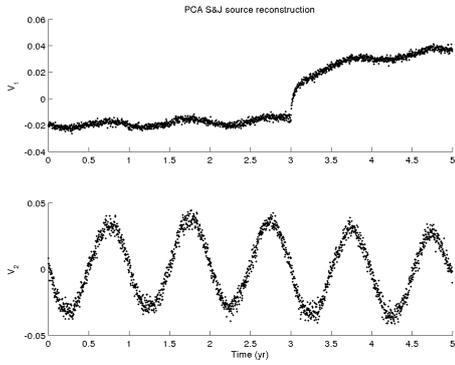


(e) PCA - MD 25%

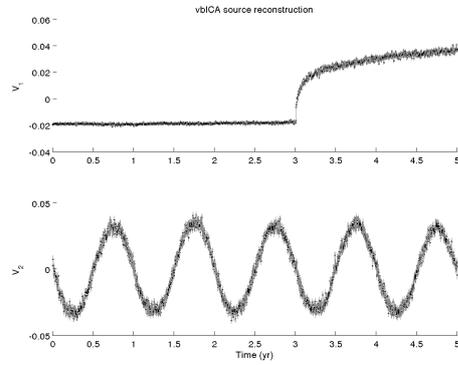


(f) vbICA - MD 25%

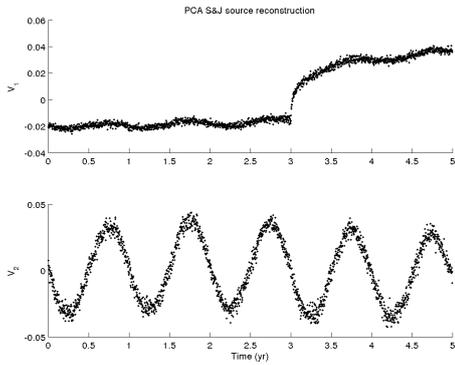
Figure 3.9:  $N1 - M_W^{post} = 6.57$



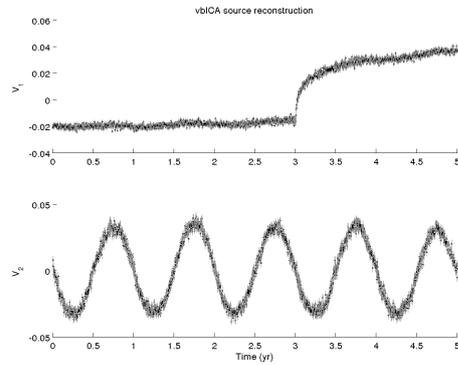
(a) PCA - MD 0%



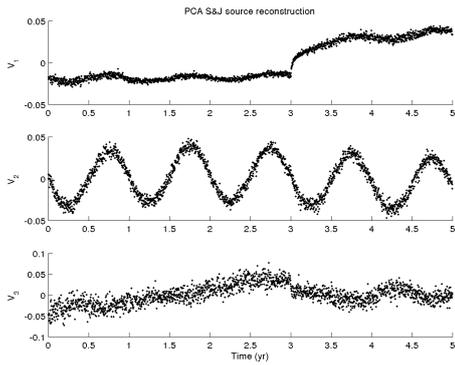
(b) vbICA - MD 0%



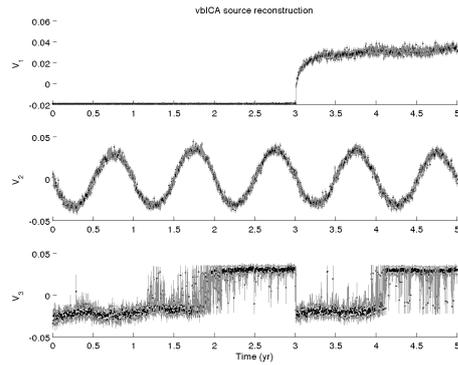
(c) PCA - MD 5%



(d) vbICA - MD 5%

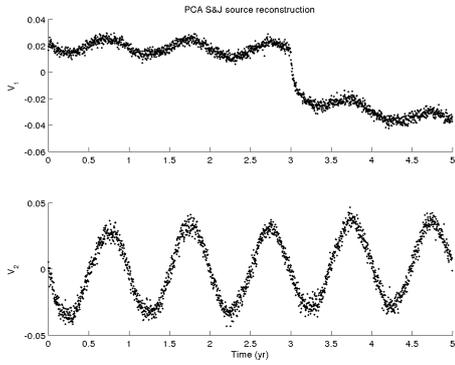


(e) PCA - MD 25%

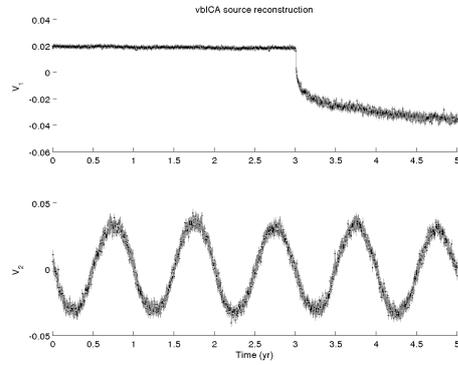


(f) vbICA - MD 25%

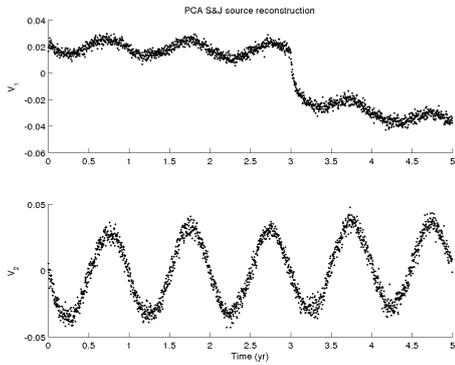
Figure 3.10:  $N_2 - M_W^{post} = 6.57$



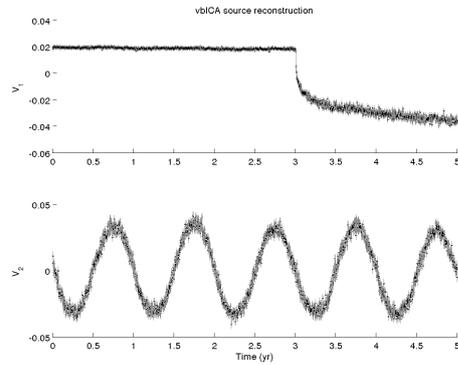
(a) PCA - MD 0%



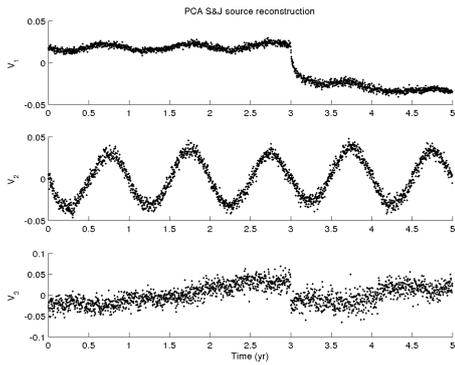
(b) vbICA - MD 0%



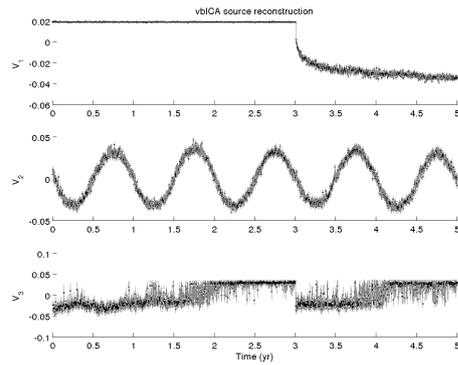
(c) PCA - MD 5%



(d) vbICA - MD 5%

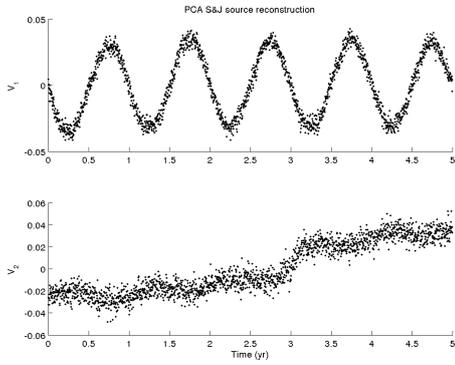


(e) PCA - MD 25%

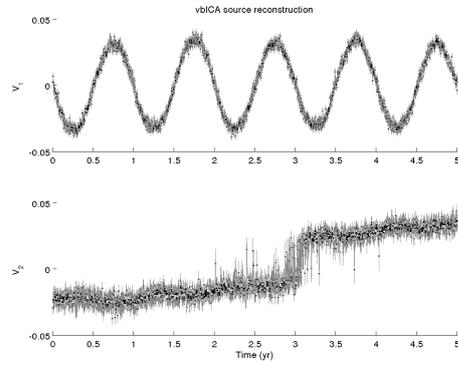


(f) vbICA - MD 25%

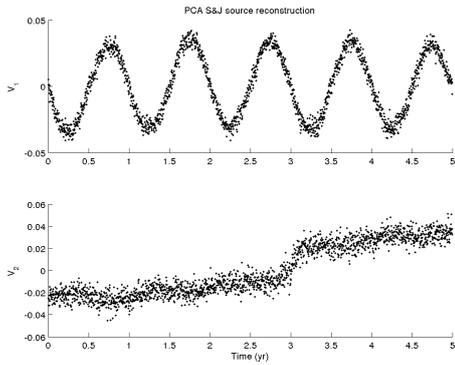
Figure 3.11:  $N1 - M_W^{post} = 6.09$



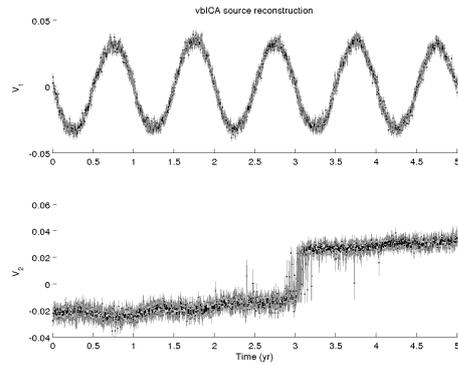
(a) PCA - MD 0%



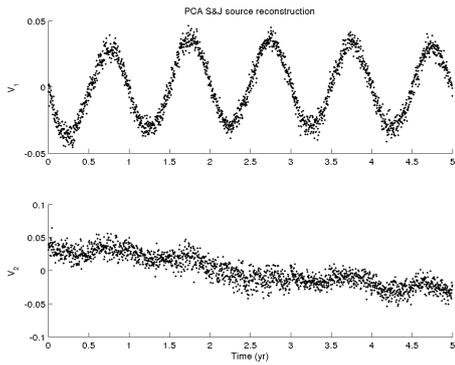
(b) vbICA - MD 0%



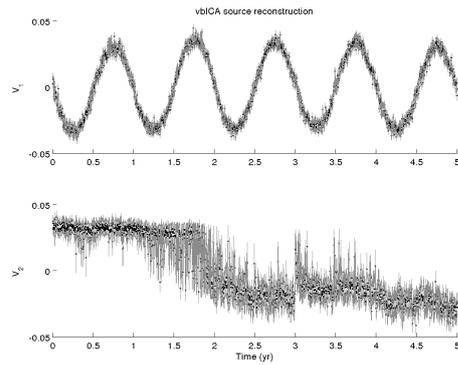
(c) PCA - MD 5%



(d) vbICA - MD 5%

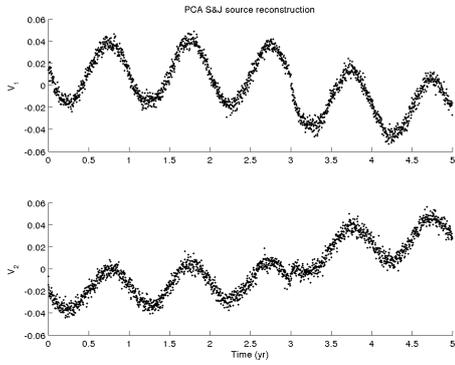


(e) PCA - MD 25%

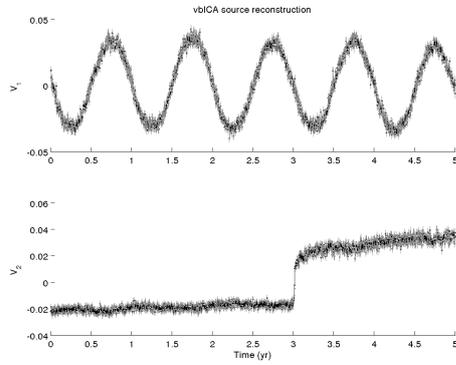


(f) vbICA - MD 25%

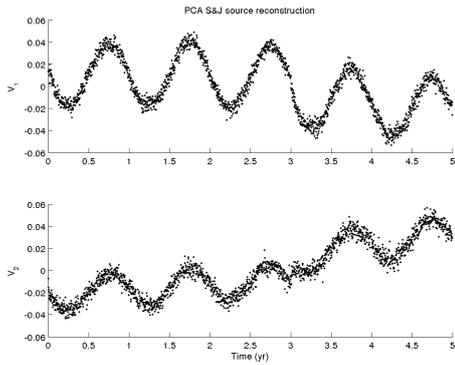
Figure 3.12:  $N_2 - M_W^{post} = 6.09$



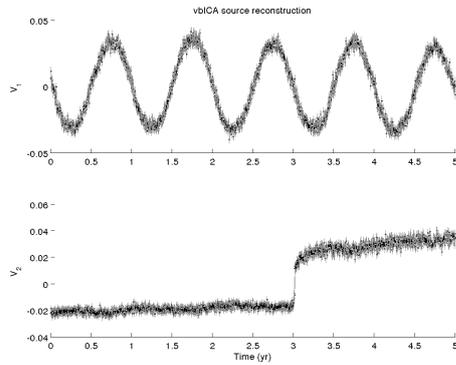
(a) PCA - MD 0%



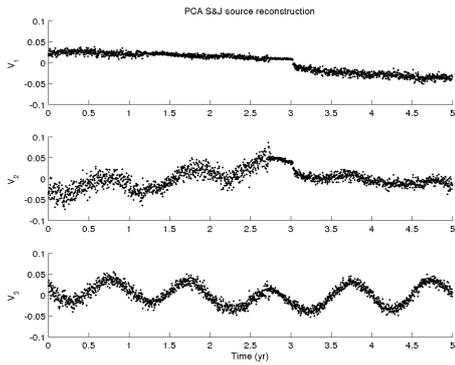
(b) vbICA - MD 0%



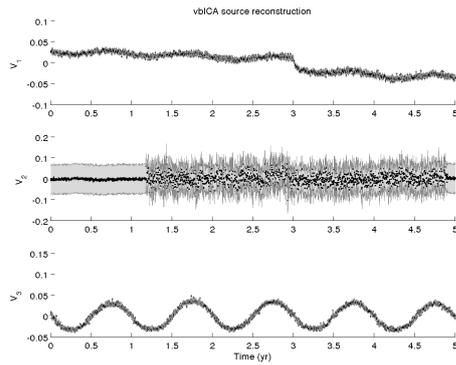
(c) PCA - MD 5%



(d) vbICA - MD 5%

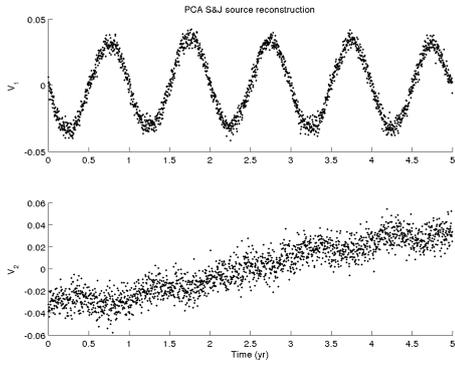


(e) PCA - MD 25%

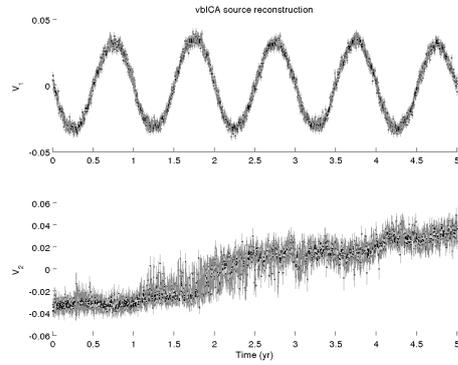


(f) vbICA - MD 25%

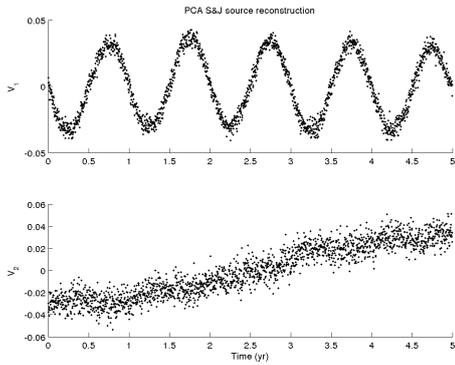
Figure 3.13:  $N1 - M_W^{post} = 5.80$



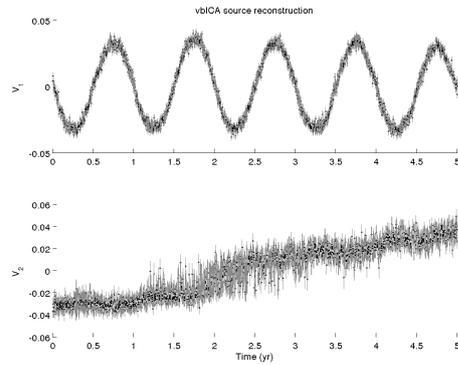
(a) PCA - MD 0%



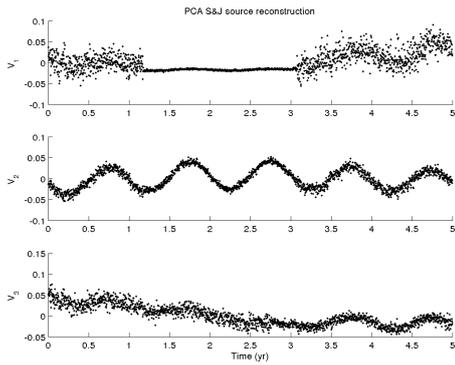
(b) vbICA - MD 0%



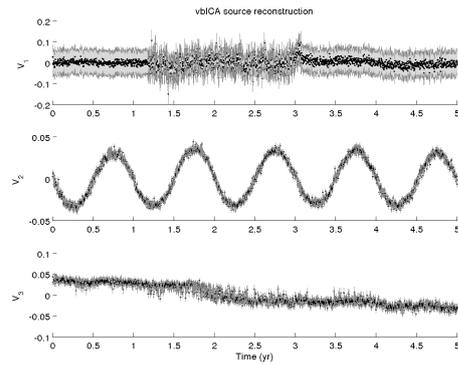
(c) PCA - MD 5%



(d) vbICA - MD 5%



(e) PCA - MD 25%



(f) vbICA - MD 25%

Figure 3.14:  $N_2 - M_W^{post} = 5.80$

		$\dot{s}^{lin} = 2 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	[1.0]	[1.0]	[1.0]
	N2	[2.0,2.5]	[1.5]	[0.5]
$M_W^{co} = 6.29$	N1	[1.5],[4.0]	[1.0]	[1.0]
	N2	[15.5,37.5]	[5.0],[6.0,9.5],[10.5],[38.0,50.0]	[NI]
$M_W^{co} = 5.94$	N1	[11.5,15.5]	[10.5,15.5]	[0.5,1.0]
	N2	[NI]	[NI]	[ $\varepsilon$ ,50.0]

Table 3.15:  $m_3$  ranges in d: 68.27% credible intervals

		$\dot{s}^{lin} = 2 \frac{mm}{yr}$		
		MD 0%	MD 5%	MD 25%
$M_W^{co} = 6.85$	N1	[1.0]	[1.0]	[1.0]
	N2	[1.0],[2.0,2.5],[4.5,5.5]	[1.5]	[0.5,1.0]
$M_W^{co} = 6.29$	N1	[1.5,2.0],[3.0,5.0]	[0.5,1.0],[2.0,3.0]	[0.5,1.0],[2.0]
	N2	[0.5,50.0]	[0.5,50.0]	[NI]
$M_W^{co} = 5.94$	N1	[0.5,1.5],[2.5,5.5],[8.0,21.5]	[0.5,5.5],[7.0,22.5]	[0.5,1.5]
	N2	[NI]	[NI]	[ $\varepsilon$ ,50.0]

Table 3.16:  $m_3$  ranges in d: 99.99% credible intervals

### 3.2.2 High tectonic rate: $60 \frac{mm}{yr}$

As previously said (see Section 3.2), to study the cases with high ( $> 10 \frac{mm}{yr}$ ) creep rates, it is necessary to reduce the correlation of the dataset. In order to do it I try two different solutions. The first consists in changing the reference frame, subtracting to all the stations the corresponding horizontal, and eventually also the vertical, velocities derived from a given reference station. The second consists in detrending the data. Of course, in both cases the estimation of the linear trend can be performed only if I have at my disposal enough consecutive data for which the linear signal is the only non-stationary one. Indeed, in this way the trend estimation is not significantly affected by the other superimposed signals.

For the synthetic data under study, the first solution is helping only for the network configuration N2. This is due to the fact that I use as a reference station the station number 1 (see Figure 3.7). The stations in configuration N1 are more spread than in configuration N2, and consequently the correlation remains high ( $> 0.67$ ). Instead, in the N2 case most of the stations are

located North-West of the source, close to the reference station. It follows that the linear trend is largely reduced, and the correlation of the dataset is reduced below the 0.67 threshold. Among all the N2 cases, I have already proven that the post-seismic signal is not recovered for the  $M_W^{post} = 6.09, 5.80$  scenarios because of the very small SNR. Therefore, I search for the vbICA decomposition only for the  $M_W^{post} = 6.57$  case.

For the MD 0% case the vbICA decomposition reduces the global MSE value, but a strong cross-talk between the ICs representing the linear and the post-seismic sources is still present. This is due to the fact that the vbICA algorithm is trapped near to the local PCA maximum for the NFE (see Figures 3.15a and 3.15b). To escape from it, I try different random initializations. This brings to different results, and Figure 3.15c shows the decomposition associated with the highest NFE value. The credible interval at 68% for the post-seismic decay constant is given by [0.5], and at 99.99% is [0.5,1.0]. This result confirms the ability of vbICA to properly reconstruct the temporal evolution of the post-seismic source used to generate the data.

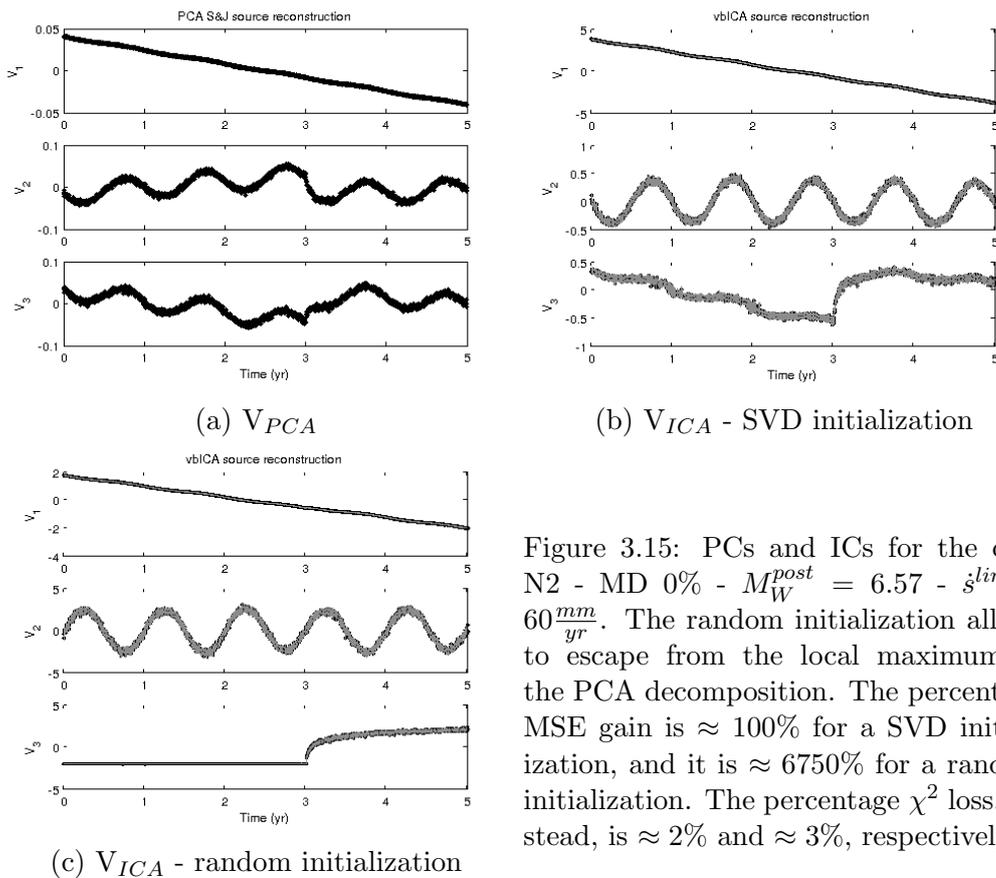


Figure 3.15: PCs and ICs for the case N2 - MD 0% -  $M_W^{post} = 6.57$  -  $\dot{s}^{lin} = 60 \frac{mm}{yr}$ . The random initialization allows to escape from the local maximum of the PCA decomposition. The percentage MSE gain is  $\approx 100\%$  for a SVD initialization, and it is  $\approx 6750\%$  for a random initialization. The percentage  $\chi^2$  loss, instead, is  $\approx 2\%$  and  $\approx 3\%$ , respectively.

Considering the second possible pre-processing step of the data, i.e. the trend subtraction from all the time series, it considerably helps in the synthetic cases under study, reducing the correlation of the dataset below the 0.67 threshold. Nevertheless, this solution must be taken as *extrema ratio*. Indeed, it affects the physical assumption made to explain the data. Removing different velocities from different stations implies to consider the linear trend an actual signal. This might not be the case, and care must be taken in the subsequent analysis of signals having a curvature different from zero (e.g., acceleration of the ground due to isostatic adjustment).

### 3.2.3 Volcanic source

After having set the number of components to 3, as suggested by the ARD criterion, I have adjusted the parameters of the a priori distributions to maximise the NFE. The results of a PCA and a vbICA are shown in Figure 3.16. The loss in the  $\chi^2$  passing from the PCA to the vbICA is  $\sim 3\%$ , while the gain in the MSE is greater than 500%. The post-seismic decay constant is correctly inferred from the first IC. Despite the low percentage of stations with  $\text{SNR} > 1$ , the signal of the volcanic source is also correctly recovered.

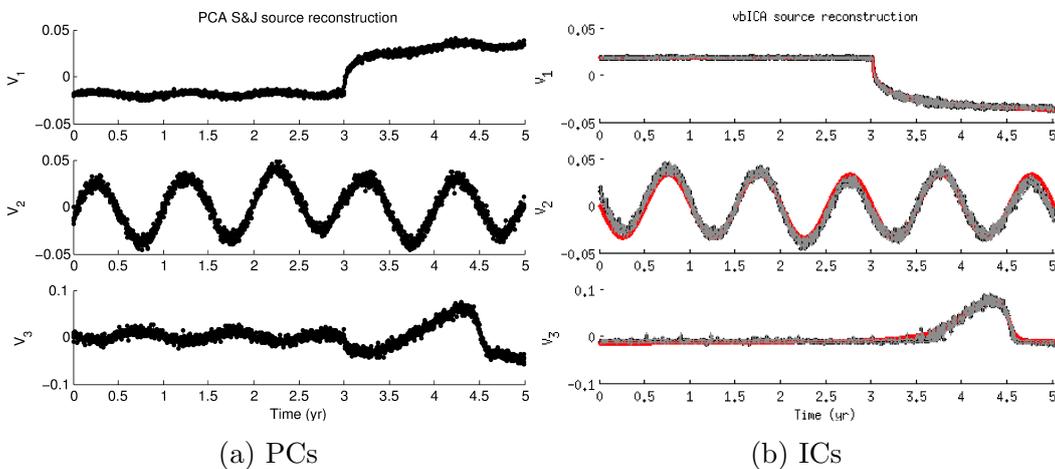


Figure 3.16: PCs and ICs relative to the case with co- and post-seismic signals mixed with a seasonal signal and a Mogi source signal. I have used only one network configuration (N2, see Section 3.1.2) and one seismic scenario to test the capability of the ICA algorithm to separate the temporal evolution of the post-seismic and the Mogi sources. Red lines in (b) indicate the actual sources.

### 3.3 Conclusive remarks

In this Chapter I have tested the ability of the modified vbICA algorithm in recovering the original sources that produce a synthetic displacement time series dataset. It has already been proven [Choudrey, 2002] the superiority of the vbICA algorithm compared to other standard ICA techniques (e.g., InfoMax or IMAX [Bell and Sejnowski, 1995], FastICA [Hyvärinen and Oja, 1997], JADE [Cardoso, 1999]) in solving the BSS problem. As an example, in Appendix B, I show the ICs and the corresponding MSE resulting from a FastICA analysis of the data relative to the case including also a Mogi source. The vbICA method adopted in this thesis is performing better.

Nevertheless, it is necessary to have a good geodetic network geometry configuration if we want to detect signals with small intensities. The definition of “good” is related to the signal we are interested in. In this Chapter I have shown that having 10 continuous GPS stations located in an area not wider than a box of side two times the locking depth of a fault where an earthquake of moment magnitude  $\sim 6$  occurred allows me to recover the post-seismic time decay constant of the source that generated the data within a credible interval of the 99.99%.

To conclude, the steps to best approach a dataset derived from the real world are:

- 1) Center the dataset, taking into account the missing data
- 2) Check the correlation of the dataset
  - 2a) if the correlation is greter than 0.67, go to point 3)
  - 2b) if the correlation is lower than 0.67 go to point 4)
- 3) Detrend the data
- 4) Correct for the offsets
- 5) Perform a vbICA

The vbICA is performed starting from a PCA initialization. If the ICs are very close to the PCs, this might mean two different things: or the ICs are orthogonal and very similar to the PCs, or the NFE local maximum corresponding to the PCA solution is very close to the global maximum. If the latter is the actual situation, then it is a good idea to try with some random initializations. The selection of the number of components is done using the ARD criterion and adopting prior parameters not very constraining (in particular, I use the values specified in Section 2.3.2). After having determined

the proper number of components, it is necessary to tune the prior parameters. This tuning can be performed evaluating the negative free energy. This is the approach that I use in the next Chapter, where I analyse real data.

# Chapter 4

## Real data applications

*I can live with doubt, and uncertainty, and not knowing.  
I think it's much more interesting to live not knowing than have  
answers which might be wrong.*

R.P. Feynman, 1981

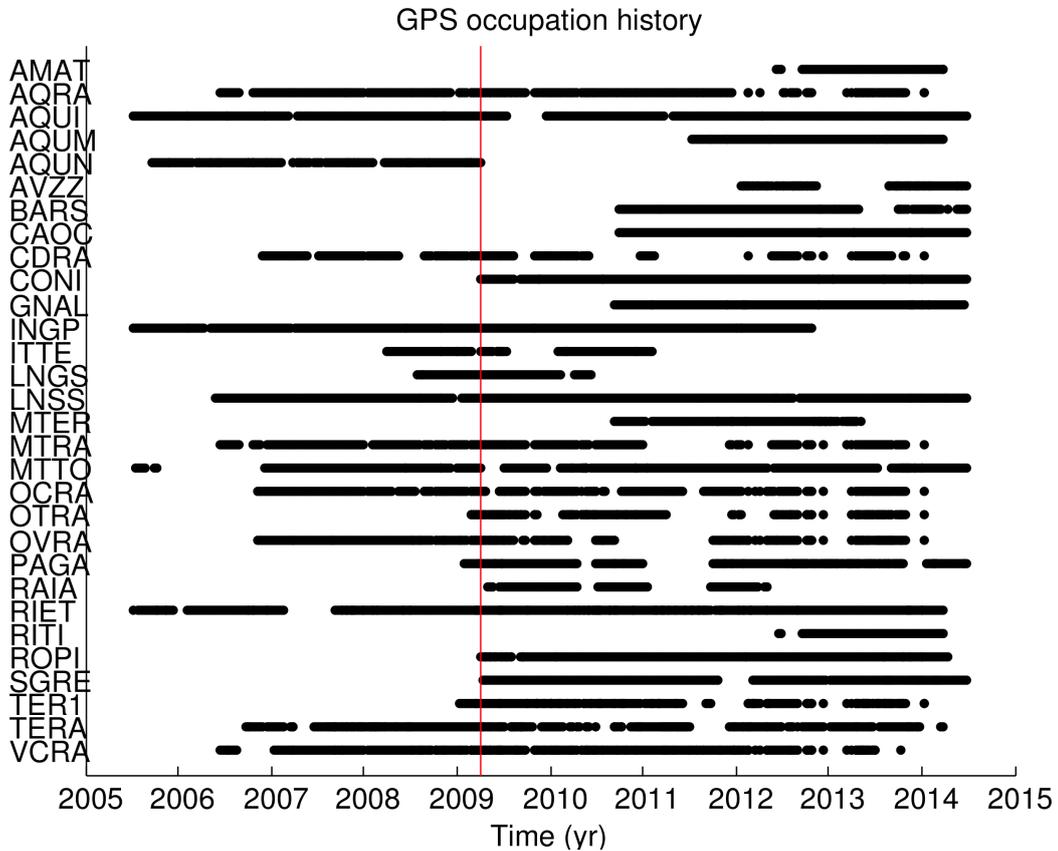
In Chapter 3 I have validated the modified vbICA algorithm on synthetic data. The tests performed were focused on the detection of the displacement signals measured at the surface, at virtual GPS stations, due to the inter-, co-, and post-seismic phases of the seismic cycle. In this Chapter I apply the previously described algorithm to real data. In particular, I focus on ground deformation using GPS data. It is worth considering that the algorithm is applicable to different kinds of dataset, provided that they are time series and can be organised in matrix fashion. The Chapter is structured as follows. First, I study a subset of cGPS stations relative to the  $M_W$  6.3 L'Aquila earthquake of 2009 (central Italy). I use only stations that recorded continuously before and after the mainshock. Then, I consider the Emilia seismic sequence of 2012 (northern Italy), where two mainshocks of comparable magnitudes activated two different structures 9 days apart. Finally, I will apply the vbICA algorithm to the Mexico Slow Slip Event (SSE) of 2006 recorded by 15 GPS stations.

### 4.1 The L'Aquila earthquake (2009), central Italy

The tectonic setting of the area struck by the L'Aquila earthquake has been described in Section 1.2. I have also discussed the limits related to a PCA approach for the study of geodetic time series. Here I directly present

the results of the application of a PCA and a vbICA onto 12 cGPS stations that have recorded continuously during both the pre-seismic and post-seismic stages. I consider the time interval that ranges from 2007.0 to 2013.5. For the epochs preceeding 2007.0 only 4 stations were continuously operative (see Table 4.1). The list of the 12 stations used is: AQRA, AQUI, CDRA, INGP, LNSS, MTRA, MTTO, OCRA, OVRA, RIET, TERA, VCRA. Despite the selection of only continuous stations, the dataset still presents an elevated amount of missing data ( $> 25\%$ ).

Table 4.1: Occupation history of the cGPS stations. The red line indicates the April 6 2009 mainshock.



#### 4.1.1 Results

In removing the co-seismic offset I use a PCA at 3 components for the time span  $[2009.0, 2009.5]$ , as suggested by a reduced  $\chi^2$  test. For the application of the ICA I use the ARD criterion for the selection of the number

of components. I start the analysis using the time series filtered only for the CME (see Section 1.2.2 for more details) and rotated in a fixed Eurasian reference frame. The PCA and ICA decompositions on the corrected time series are very similar (Figure 4.1). The first and second ICs are probably

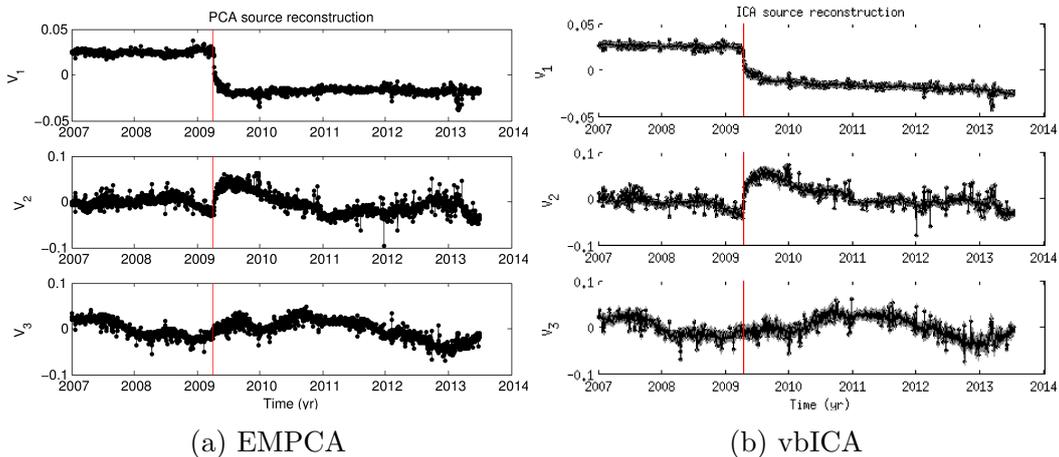


Figure 4.1: PCs (a) and ICs (b) on the data corrected for the CME and the co-seismic offset.

affected by a high cross-talk, and the linear trend is not properly separated from the post-seismic decay. Randomly changing the initialization point of the algorithm does not lead to significant improvement. For this reason I decide to detrend the time series using the data before the occurrence of the mainshock. The results on the detrended time series are shown in Figure 4.2 (for the displacements time series, see Appendix C.1). An ARD criterion points to 3 as the most appropriate number of components. The first IC can be explained with afterslip. The logarithmic approximation used also in Section 1.2 is no longer valid after  $\sim 150$  d after the mainshock, as shown in Figure 4.3. Using the model expressed by equation 3.16, the blue and red lines indicate the best fits for the first 306 d and 150 d after the mainshock, respectively. The 99% credible interval (see Appendix A) for the time decay constant deduced using only the first 150 d after the mainshock is  $[0.3, 0.6]$  d (Figure 4.3).

It is interesting to notice that now a seasonal signal appears in the third IC and second PC. In a PCA it is mixed with the post-seismic decay, while the ICA clearly separate the two sources. A frequency analysis reveals that the third IC has a peak at frequencies of 1 cycle/yr, corresponding to a period of 1 yr, while there is no evidence of a signal with a period of 0.5 yr, as usually assumed in the analysis of GPS time series. Considering the  $1\sigma$  nominal uncertainty given by the GAMIT analysis as representative of the

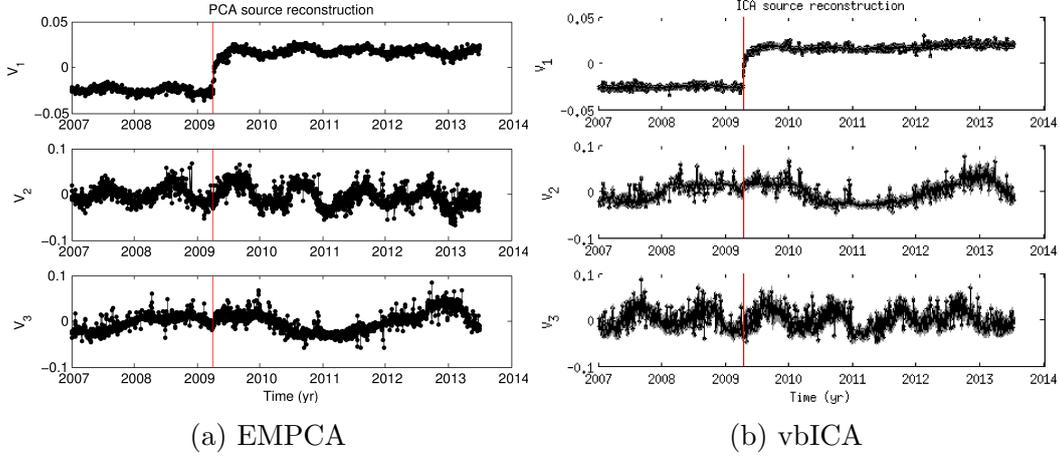


Figure 4.2: PCs (a) and ICs (b) on the data corrected for the CME, the co-seismic offset, and a linear trend.

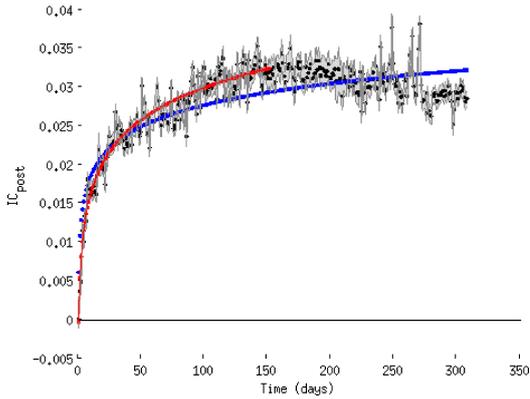


Figure 4.3: First IC - Fit with a logarithmic model (equation 3.16). Blue line: best fit over the first 306 d. Red line: best fit over the first 150 d. The time decay constant for the credible interval at 99% varies in the range [0.3, 0.6] d.

error in the determination of the station position, I consider as representative for the noise on a time series the average value of the uncertainties associated to every single measurement. Thus I estimate the SNR value for each time series as the square of the ratio between the signal maximum displacement and the noise amplitude. The SNR of the third IC is greater than 1 only for the vertical component. Its spatial response is homogeneous over the region, indicating that all the stations are in phase. Instead, the second IC shows a SNR greater than 1 only for the horizontal components. Assuming that the annual signal can be properly explained using a sinusoid with a constant amplitude and period, I proceed removing also this modellization.

The PCA and ICA on the time series filtered for the CME, the co-seismic offset, a linear trend, and an annual signal are shown in Figure 4.4 (for the displacements time series, see Appendix C.1). Again, the first IC can be associated with afterslip, and the post-seismic decay extinguishes after

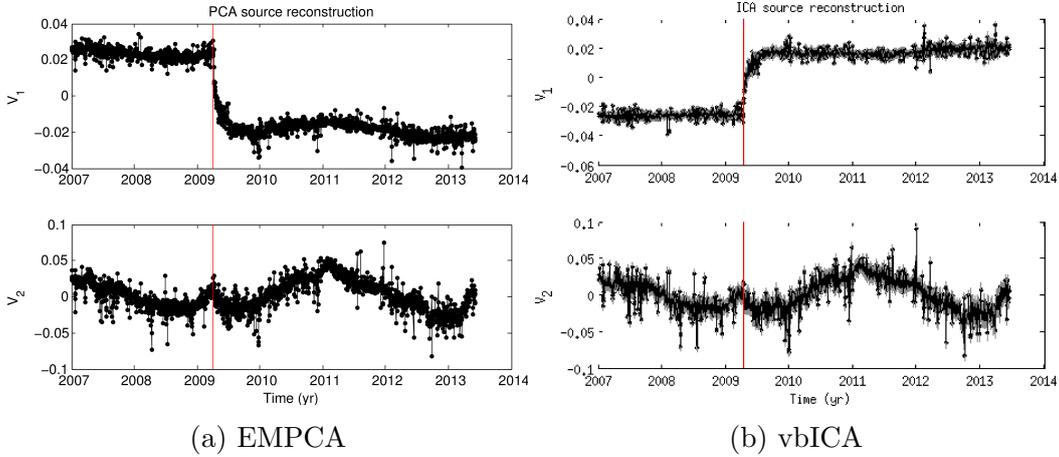


Figure 4.4: PCs (a) and ICs (b) on the data corrected for the CME, the co-seismic offset, a linear trend, and a sinusoid with annual period.

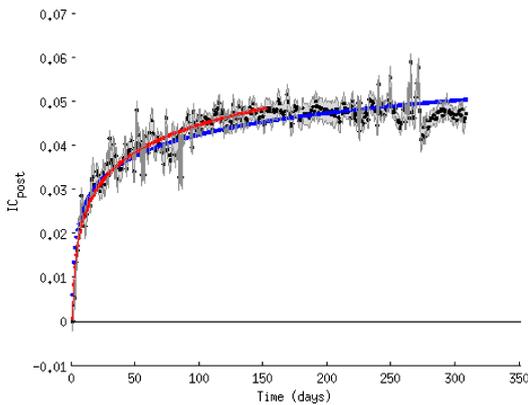


Figure 4.5: First IC - Fit with a logarithmic model (equation 3.16). Blue line: best fit over the first 306 d. Red line: best fit over the first 150 d. The time decay constant for the credible interval at 99% varies in the range  $[0.4, 0.5]$  d.

$\sim 150$  d (see Figure 4.5).

The second IC is not a common signal in GPS time series. It shows a multi-annual period of about 4 yr, but the time spanned is too short to fully characterize this periodicity. A peculiar deviation from a multi-annual sinusoid is observed at times around the mainshock. In order to better study this particular anomaly and reduce the influence of the multi-annual signal, I decide to focus the analysis to a time span sufficiently small around the time of the mainshock. I redo the analysis of the 12 cGPS time series in the time range from 2008.0 to 2010.0. The ARD criterion suggests that only one component is enough to explain the data. In Figure 4.6a I show a decomposition at 2 components. The second IC has a SNR lower than 1 for all the three directions (east, north, and vertical). The station with the highest displacement associated to the second IC is MTTO. Removing this station from the analysis makes the signal related to the second IC vanish,

as can be seen from Figure 4.6b.

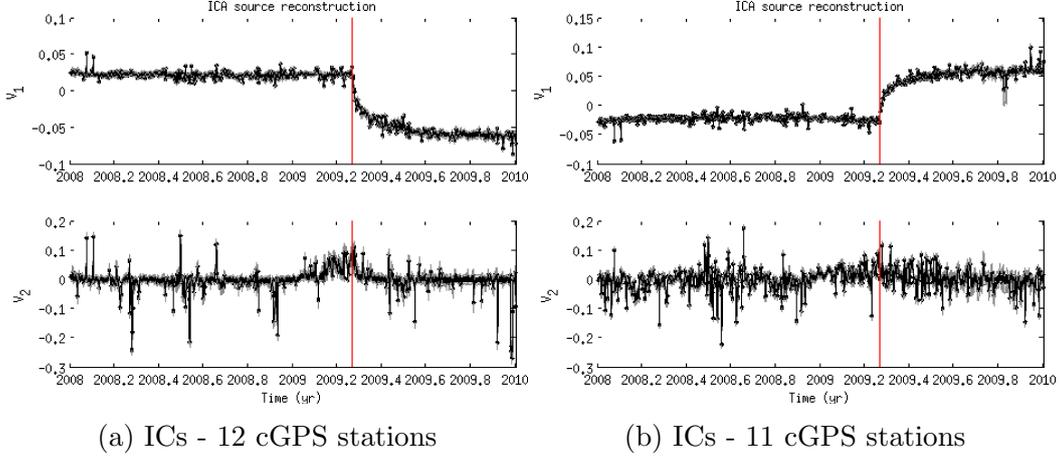


Figure 4.6: ICs on the data corrected for the CME, the co-seismic offset, a linear trend, and a sinusoid with annual period.

### 4.1.2 Discussion

The analysis presented in this Section refers to a subset of the cGPS stations that recorded data during the pre- and post-seismic stages relative to the 2009 L’Aquila earthquake. The dataset used for this study is different w.r.t. to the one used in Chapter 1, and the inversion for afterslip would be less accurate since the number of stations that I am using for the analysis of the pre- and post-seismic time series is lower. The goal of the application of an ICA technique is to look for other possible sources of deformation acting at the same time of the afterslip source. The results show that the post-seismic deformation is mainly driven by a process that decays as equation 3.16 during the first 150 d after the mainshock. Such a temporal behaviour is related to afterslip mechanism [Marone et al., 1991]. Thus, the application of the vbICA technique allows me to validate the assumption that afterslip is the main source of deformation, made in Section 1.2 in order to use the PCA technique for the analysis of the post-seismic GPS time series relative to the L’Aquila 2009 earthquake.

A second component has been detected and separated by the vbICA algorithm. Silverii et al. [2014] also found a cyclic signal with a multi-annual period applying a low-pass filter to the GPS time series. They correlated this signal to rainfalls associated to the North Atlantic Current, and suggested the possibility of a link between the observed deformation and the hydrological loading. Nevertheless, the observed displacement is acting mostly on the

horizontal components. The spatial response associated with the second IC is shown in Figure 4.7. Red coloured stations exhibit a displacement that

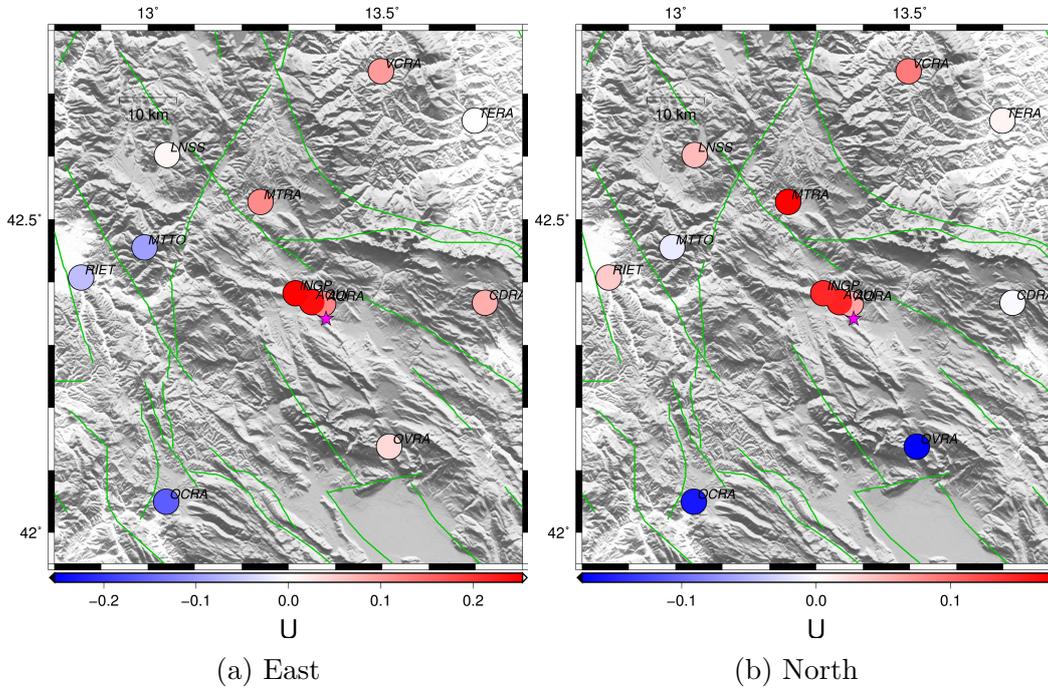


Figure 4.7: Spatial response of the second IC (see Figure 4.4b). The magenta star correspond to the mainshock. The green lines correspond to the fault traces from the structural model of Italy.

is toward East or North if the function described by  $V_2$  in Figure 4.4b is increasing. The same stations move toward West or South if  $V_2$  is decreasing. Blue coloured stations act in the opposite way. The pattern shown in Figure 4.7 indicates the presence of an alternate state of contraction and extension of the region over long periods. In particular, just before the mainshock the deviation from the long-period sinusoid indicates an extension of the hangingwall in the SW-NE direction. This extension started about the beginning of 2009 (see Figures 4.4b and 4.6a). At the same time a strong foreshock sequence activated the deepest portion of the fault until a week before the main shock, when the largest foreshock ( $M_W$  4.0) triggered a minor antithetic segment. Seismicity jumped back to the main fault plane a few hours before the mainshock [Chiaraluce et al., 2011]. This could indicate that the extension experienced by the entire region is the same that initiated the foreshock activity. Anyway, a more detailed analysis that focuses only on the period from 2008.0 to 2010.0 has shown that such a signal is mainly related to the presence of the MTTO station. Furthermore, I can not exclude

the possibility that the presence of this deviation from a sinusoidal behaviour is caused by a mismodel of the co-seismic offset, removed from the analysis before the application of the vbICA algorithm. In order to get more reliable results and to understand better the origin of the observed multi-annual signal it is necessary to maintain a higher number of stations operative in continuous mode.

## 4.2 The Emilia seismic sequence (2012), northern Italy

### 4.2.1 Introduction

In May-June 2012, a seismic sequence struck the Emilia region (northern Italy). The mainshock, of  $M_L$  5.9, occurred on May 20, 2012, at 02:03 UTC. In addition, on May 29, 2012, a  $M_L$  5.8 earthquake occurred in the morning (07:00 UTC), followed by two events within just 5 min of each other, one at 10:55 UTC ( $M_L$  5.3) and the second at 11:00 UTC ( $M_L$  5.2). All of the focal mechanisms were thrust, in some cases with different, but small, percentages of strike-slip components, consistently showing compressional kinematics with E-W oriented reverse nodal planes [Pondrelli et al., 2012]. Seismological studies do not agree on the moment magnitude associated to each event. Pondrelli et al. [2012] determined a  $M_W$  of 6.1 for the first event, and a  $M_W$  of 5.9 for the second. Scognamiglio et al. [2012] found  $M_W$  values of 5.86 and 5.66, respectively. Instead, Malagnini et al. [2012] report a  $M_W$  of 5.63 and 5.44.

From geological studies it has been shown that this region is affected by slow deformation rates due to the northwards motion of the northern Apennines fold-and-thrust belt, which is buried beneath the sedimentary cover of the Po Plain [e.g., Picotti and Pazzaglia, 2008, Toscani et al., 2009]. More recently, the deformation rates have been monitored using GPS measurements. Different geodetic studies [e.g., Serpelloni et al., 2006, Devoti et al., 2011, Bennett et al., 2012] agree that less than  $2 \frac{\text{mm}}{\text{yr}}$  of SW-NE shortening are accommodated across this sector of the Po Plain.

Various GPS networks, mainly managed by private and regional institutions for cadastral and topographic applications, were operative during the seismic sequence, and recorded the surface position at a daily sample rate. Serpelloni et al. [2012] performed a uniform slip inversion for the co-seismic displacement of the two events, finding a  $M_W$  of 6.1 for the May 20<sup>th</sup> event, and 5.8 for the May 29<sup>th</sup> one, on two North vergin thrust faults. They also inverted for the fault plane geometry directly from the GPS data, finding a

good agreement between the seismicity distribution derived from the INGV catalog (<http://iside.rm.ingv.it>) and the geodetically derived fault plane relative to the 20<sup>th</sup> of May. For the 29<sup>th</sup> of May event the modelled fault plane imaged by GPS is shifted by ca. 5 km to the SW with respect to the crustal volume affected by the aftershocks. The fault planes described in their work have strike of  $\sim 109^\circ$  and  $\sim 102^\circ$  for the May 20 and 29 events respectively, and they are dipping towards South.

Pezzo et al. [2013] inverted for both GPS and InSAR displacement data, inverting for a non-uniform slip distribution. In their work they neglected the contribution of the underground fluid pressure. Extensive liquefaction phenomena were observed over an area of  $\sim 1200 \text{ km}^2$  following the 20 and 29 May mainshocks. Because of the high susceptibility to liquefaction of the alluvial plain, this seismic sequence produced the most prominent extensive liquefaction phenomena of the last century in Italy [Emergeo, 2013]. Here I show that the vbICA algorithm is able to identify the transients of surface displacement that are likely related to both the afterslip and the poroelastic response of the medium. This suggests that for the Emilia seismic sequence fluids played a significant role which should not be neglected if we want to better understand crustal deformation associated to this seismic event.

## 4.2.2 Data

I select the GPS stations in an area including a radius of 50 km from the average longitude and latitude of the two mainshocks:  $lon = 11.165^\circ$ ,  $lat = 44.869^\circ$  (red dot in Figure 4.8). This selection points out to 28 GPS stations included in the area, 20 of which are classified as continuous GPS (cGPS) stations, and the remaining 8 as survey GPS (sGPS). Considering as a time span the one that goes from the first to the last epoch available for the 28 stations, the dataset shows  $>50\%$  of missing data. Among these 28 stations I select only those having continuous data during the post-seismic period and having at least 2.5 years of data before the first mainshock. This last criterion is necessary in order to reduce the uncertainties in the modelling of the seasonal cycles [Langbein, 2008]. This selection reduces the number of stations available to 17. Finally, I exclude the stations named SGIP and BOLG. The former is located nearly at the same longitude and latitude of the station PERS, and its quality is much lower than the one of its neighbour because a strong multipath component affects the solution for the position of the station. The latter is close to the stations BOLO and BLGN, but it is affected by a local signal due to a monumentation problem (Bacchetti, personal communication). Consequently, the results shown in this Section refer to the analysis of 15 stations, and their location is shown in Figure 4.8.

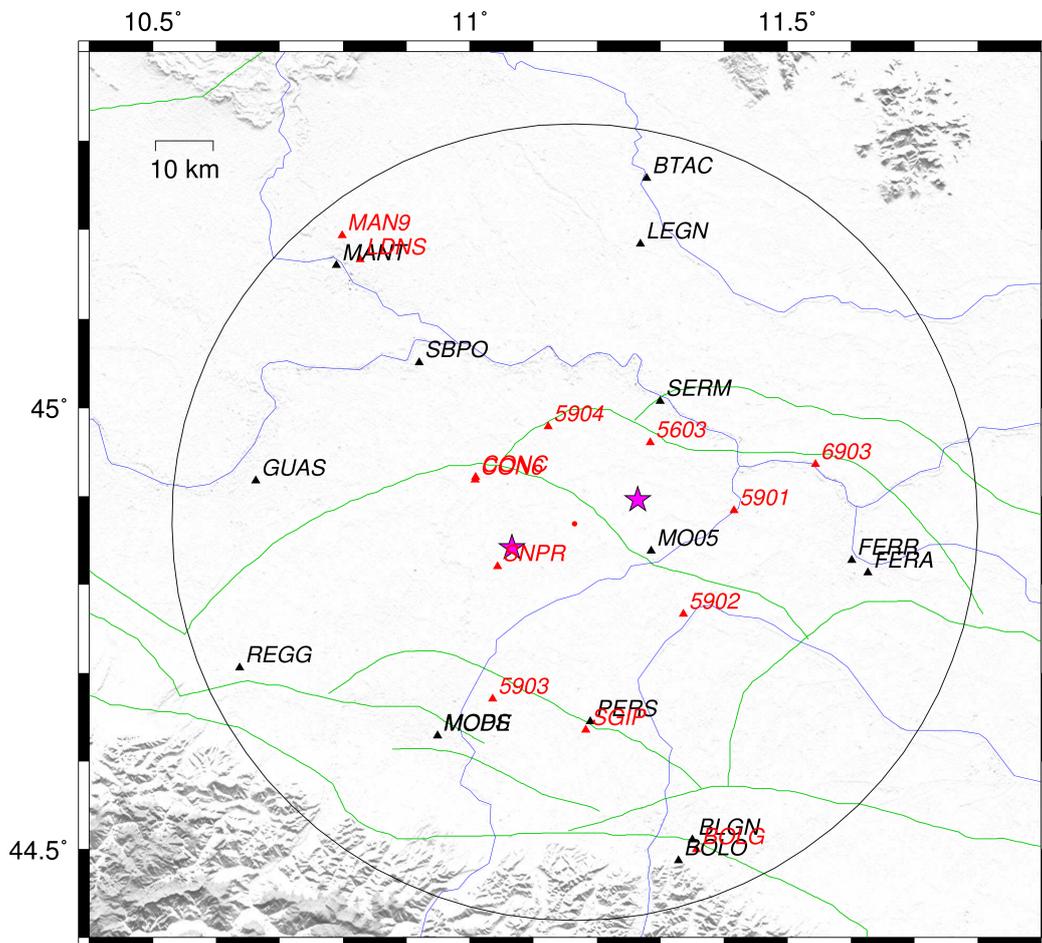
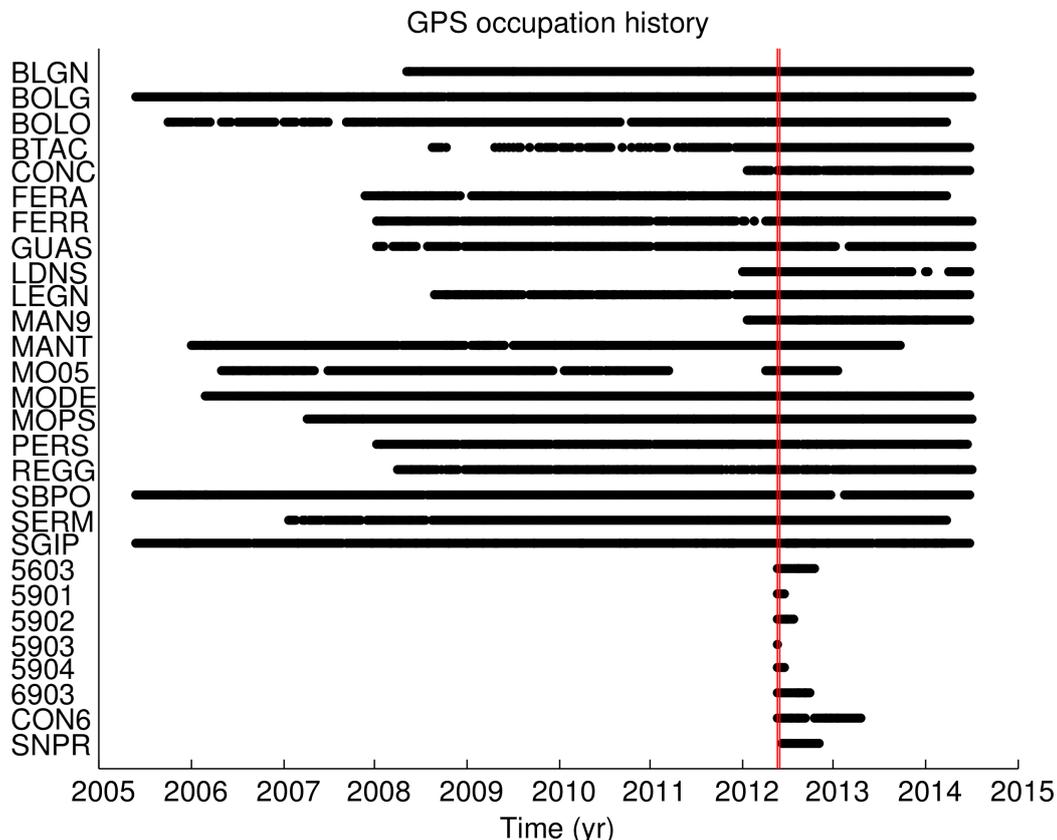


Figure 4.8: Region under study. The black circle indicates the 50 km radius area taken into account. The two magenta stars correspond to the 20<sup>th</sup> and 29<sup>th</sup> mainshocks from the ISIDE catalogue (<http://iside.rm.ingv.it/>). The red dot is the average epicentre of the two mainshocks. The green lines correspond to the fault traces from the structural model of Italy. The black triangles correspond to the cGPS stations used in this Chapter. The red triangles correspond to the GPS stations discarded from the analysis.

The data have been processed with the GAMIT/GLOBK software (<http://www-gpsg.mit.edu/~simon/gtgk>), and they are daily sampled, with the value assigned to the position of the station corresponding to the average value of all the data of the day. For all the calculations that follow I assume that a point in a time series is representative of the position of the station at noon of the corresponding day. The only exception is for the days of the two mainshocks. In this case the data have been processed considering only the data after the earthquake and belonging to the same day. The first and

Table 4.2: Occupation history of the GPS stations. The red lines indicate the two mainshocks.



last data available correspond to May the 22<sup>nd</sup> of 2005, and July the 1<sup>st</sup> of 2014. Table 4.2 shows the occupation history of the 28 stations. Most of them have been activated for the first time around the beginning of 2008. To help the stability of the EMPCA and vbICA algorithms, I prefer to start the analysis from January the 1<sup>st</sup> of 2008. The percentage of missing data for the 15 stations considered and for this time span is  $\sim 17\%$ .

I have filtered the time series for the Common Mode Error (CME) using the approach proposed by Dong et al. [2006] (see Section 1.2.2).

### 4.2.3 Results

In order to apply the vbICA algorithm for the detection of post-seismic transients I have to correct the data for the co-seismic offsets. This correction is necessary because of the non-independence between the co- and post-seismic signals (see Section 3.2). I have extracted the co-seismic offsets

applying a 10 components PCA spanning the time that ranges from 2012.0 to 2012.6, i.e. focusing on a time span close to the events. The choice of 10 components is performed applying a F-test to the PCA. Instead, the choice of the time span 2012.0-2012.6 is arbitrary, and it is aimed at explaining the data around the co-seismic jumps. Indeed, if I focus on a longer time span the risk is to give little importance to the variance due to the steps, and I do not reproduce them adequately. The high number of PCs used and the restricted time window analysed enable me to well recover the co-seismic steps, allowing me to remove two Heaviside functions corresponding to the time of the earthquakes.

The estimated co-seismic displacements are shown in Table 4.3 and in Figures 4.9.

Table 4.3: Co-seismic offset estimation: E = East, N = North, V = Vertical

Stn name	May 20 <sup>th</sup> (mm)			May 29 <sup>th</sup> (mm)		
	E	N	V	E	N	V
BLGN	0.4	5.8	5.6	-2.3	2.6	2.2
BOLO	0.1	6.0	4.8	-1.4	1.4	5.2
BTAC	-0.4	-2.5	-6.0	-1.3	-1.1	8.1
FERA	0.6	-0.2	1.4	-0.1	-0.9	2.9
FERR	0.2	-0.6	1.3	-0.7	0.1	2.9
GUAS	0.0	-0.3	0.4	-1.6	0.0	1.2
LEGN	0.2	-2.6	0.3	-1.1	-2.9	4.3
MANT	-1.2	-1.4	5.5	-0.3	-0.9	1.7
MO05	-5.5	-32.8	83.1	4.3	-0.7	-5.5
MODE	4.2	5.6	0.6	3.3	8.5	0.4
MOPS	3.1	3.4	1.8	4.8	10.8	2.1
PERS	5.5	19.4	1.3	-2.8	8.1	4.5
REGG	1.0	0.2	-0.2	0.9	1.8	4.9
SBPO	0.4	-1.4	9.2	0.5	-4.1	2.0
SERM	-2.3	-15.1	-3.2	-3.4	-5.2	5.5

Considering the  $1\sigma$  nominal uncertainty given by the GAMIT analysis as representative of the error in the determination of the station position, I consider as representative for the noise on a time series the average value of the uncertainties associated to every single measurement. Thus I estimate the SNR value for each time series as the square of the ratio between the signal amplitude (see Table 4.3 for the co-seismic signals estimated) and the noise amplitude. The mean SNR of the May 20<sup>th</sup> co-seismic signal for the

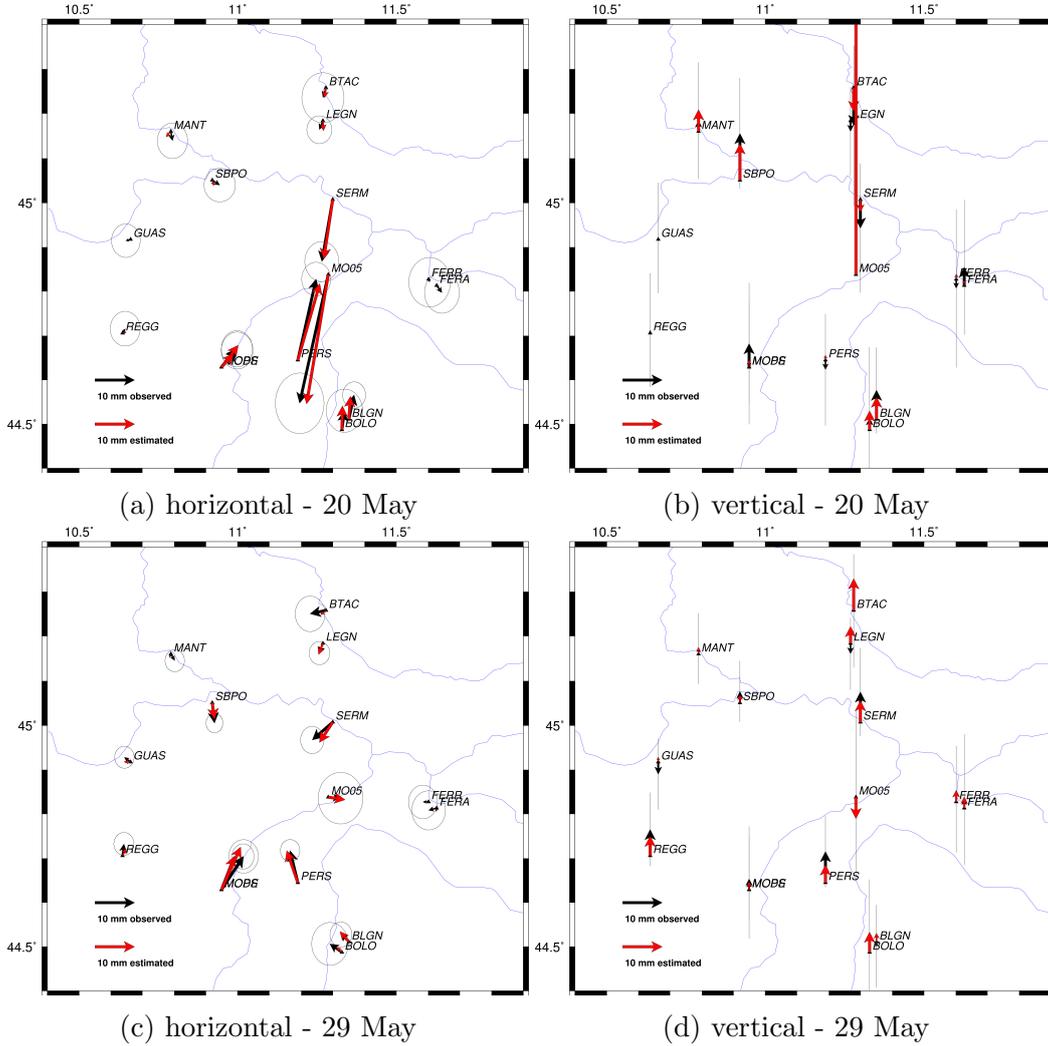


Figure 4.9: Co-seismic displacements. Black arrows indicate the offsets computed using the first day available after the earthquake and the last before it. The ellipsoids in 4.9a and 4.9c, and the vertical bars in 4.9b and 4.9d correspond to the  $2\sigma$  uncertainties. The red arrows are the offset estimated with a PCA, and used to correct the time series.

east, north, and vertical directions is  $\sim 2.7$ ,  $\sim 31.7$ , and  $\sim 7.2$  respectively. For the May 29<sup>th</sup> event it is  $\sim 2.2$ ,  $\sim 8.3$ , and  $\sim 0.4$ . As expected for a thrust fault striking  $\sim 90^\circ$ , the highest value is in the north direction. The only mean SNR value below 1 is found for the vertical signal generated by the second mainshock. This explains the reason why Serpelloni et al. [2012] did not consider the vertical component for such event.

After the removal of the co-seismic estimated offsets, I proceed analysing

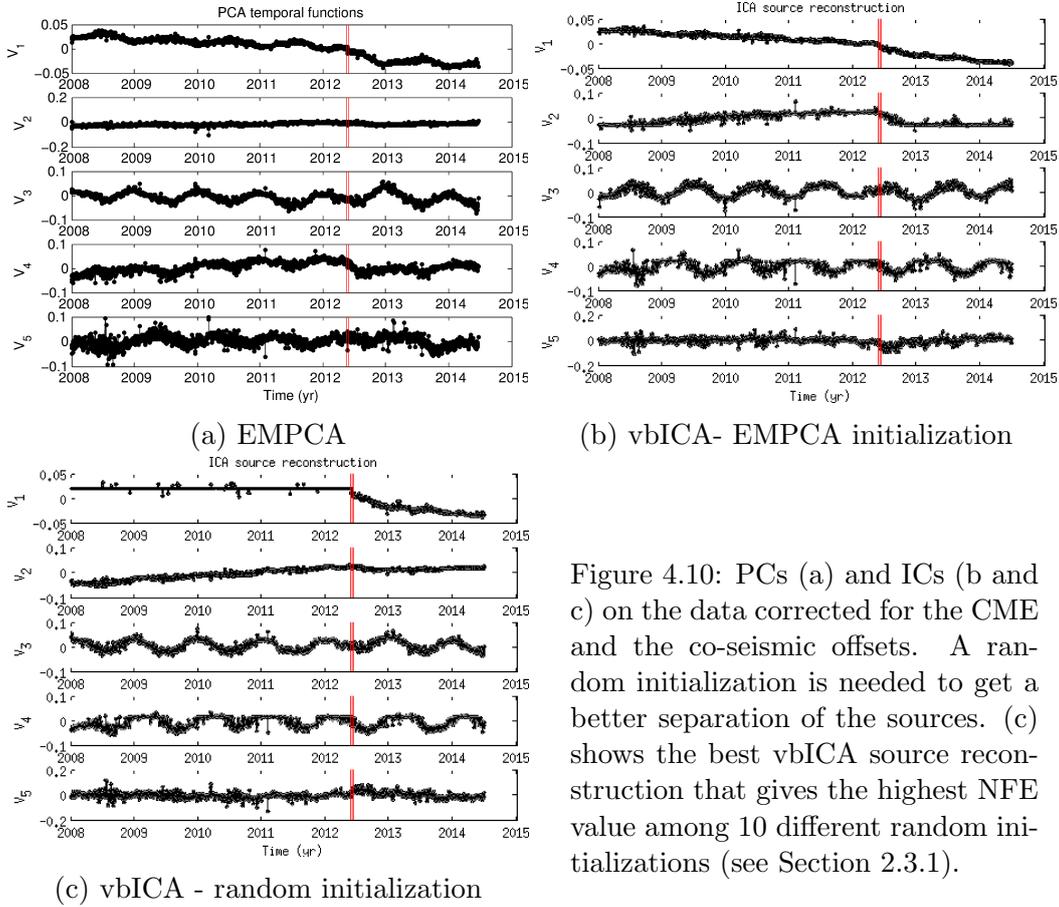


Figure 4.10: PCs (a) and ICs (b and c) on the data corrected for the CME and the co-seismic offsets. A random initialization is needed to get a better separation of the sources. (c) shows the best vbICA source reconstruction that gives the highest NFE value among 10 different random initializations (see Section 2.3.1).

the corrected time series. Figure 4.10a shows the recovered temporal functions using the EMPCA algorithm (from now on, PCA). At least 4 components over 5 show a seasonal signal with roughly an annual period. Moreover, it persists also a mix between a linear trend and a post-seismic decay. The second PC shows a slight slope, and the eigenvalue associated is the second more relevant one. The choice of 5 components is based on an ARD criterion method applied on the vbICA decomposition (see Chapter 2). In particular, the vbICA sources obtained from a PCA initialization are shown in Figure 4.10b. Two seasonal signals with an annual period and roughly  $90^\circ$  out of phase are caught by ICs number 3 and 4. The first two ICs still show a cross-talk between a linear trend and a post-seismic decay. This can be due to the non optimal choice of the initialization point, as already shown on synthetic data (see Section 3.2.2). To escape from this local maximum defined by the PCA solution, I run 10 vbICA decompositions starting from a random initialization point. Figure 4.10c shows the decomposition with the highest NFE value among 10 random initializations.

I can not quantify the reduction in the cross-talk, because I do not know the original sources; nevertheless, the improvement in extracting the independent sources of deformation seems evident comparing Figures 4.10a and 4.10c. It is interesting to notice that a fifth IC is suggested by the ARD criterion. Such IC does not seem to be affected by the different initialization point (see  $V_5$  in Figures 4.10b and 4.10c). It clearly shows a decay during the post-seismic stage, but it is not monotonic. Such a behaviour is not typical of afterslip processes, and in order to better investigate it I decide to remove a linear trend from the data. In fact, for the study of such a transient signal the long-term linear trend is acting as noise, and the high power of the latter reduces the SSR for the proper detection of the former. This detrending procedure is commonly adopted in GPS studies [e.g. Wdowinski et al., 1997], and it can help in the analysis not only of the temporal evolution of the sources, but also (and perhaps more importantly) of the spatial pattern associated to each of them.

The long-term velocity is computed using only the data before the occurrence of the first main earthquake, namely the  $M_W6.1$  of the 20<sup>th</sup> of May. The time series corrected for the inter- and co-seismic signals are shown in Appendix C.2. The ARD criterion indicates 4 ICs as the proper number of components. A comparison between the PCA temporal functions and the vbICA sources is shown in Figure 4.11. From this analysis it is clear how the seasonal signals are not separated by a PCA, and it offers a glaring example of the advantage of the ICA over a PCA.

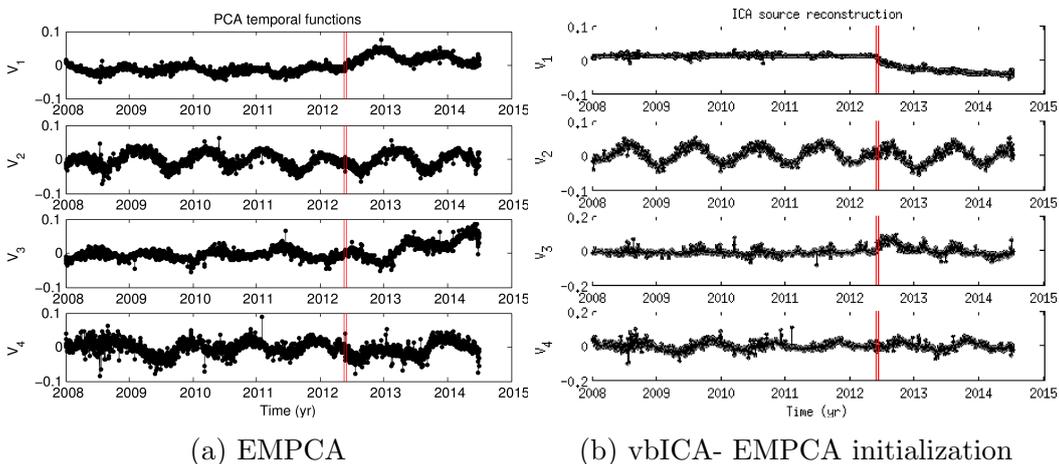


Figure 4.11: PCs (a) and ICs (b) on the data corrected for the CME, the co-seismic offsets and a linear trend.

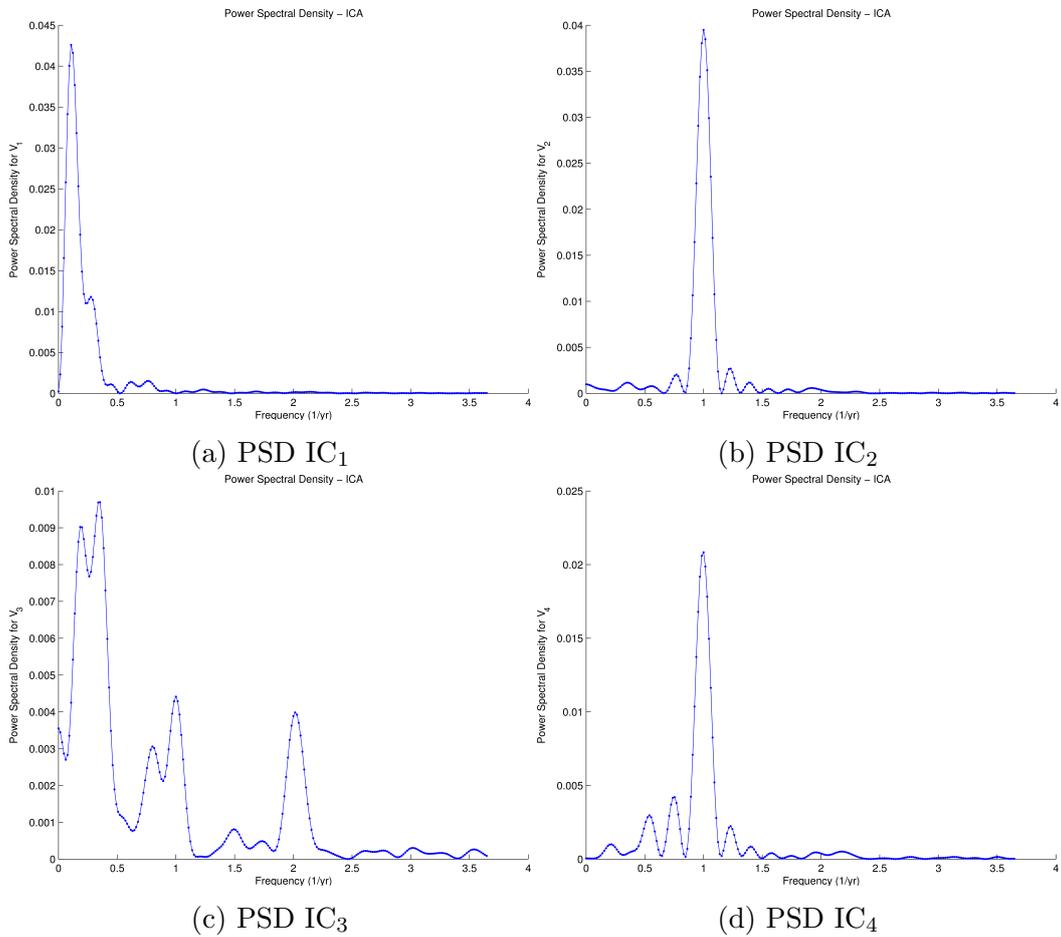


Figure 4.12: Power Spectral Densities (PSDs) of the four ICs of Figure 4.11b.

Performing a frequency analysis on the ICs I find that there are peaks corresponding to signals having annual and semi-annual periods (see Figure 4.12). It is interesting to notice that the semi-annual peak is present only in the third IC, while the annual one is present in more than one component. The separation of the annual signals in the second and fourth ICs appears to be very clear. These components show a sinusoidal signal  $\sim 90^\circ$  out of phase. In particular, the second IC has a peak roughly corresponding to the solstices, and the fourth IC to the equinoxes. This is an evidence that more than one annual source contributes to the ground displacement recorded by the GPS stations.

An interesting result is that the decay characterized by the fifth IC of the raw time series analysis (see Figure 4.10c) persists also in the detrended time series (see  $V_3$  in 4.11b). In order to better characterize such a transient signal I remove two sine functions with constant amplitude and period. Removing

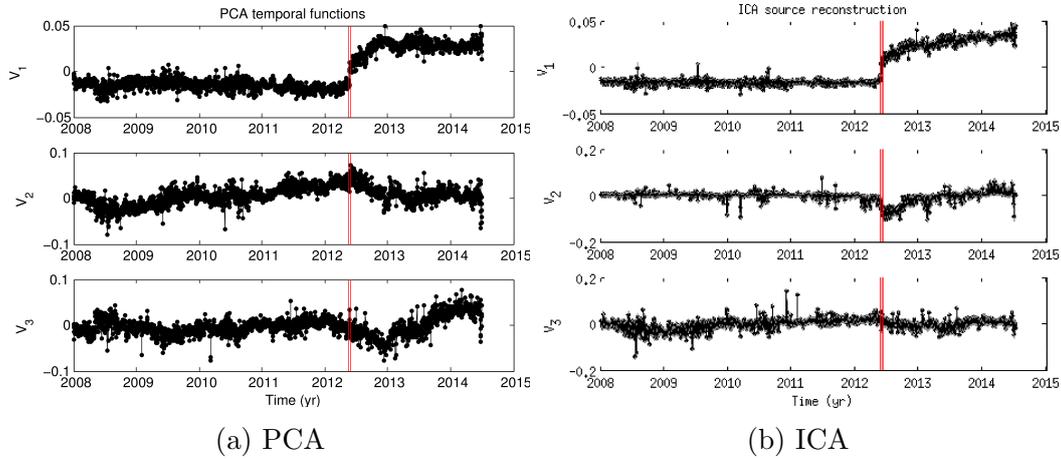


Figure 4.13: PCA and ICA - linear trend, annual and semi-annual signals removed

this signals can help in improving the SSR for the study of small transients related to the seismic sequence.

The results of the PCA and ICA decomposition on the cleaned data are shown in Figure 4.13. This time an ARD criterion suggests to use 3 components. Looking at the PCs (Figure 4.13a) it seems that three different processes are mixed together:

- 1) a monotonic post-seismic decay
- 2) a non-monotonic post-seismic transient signal
- 3) a cyclic signal with long period

A frequency analysis performed on the PCs reveals that the second and third PCs share a peak at 0.5 cycles/yr, corresponding to a signal with a period of 2 yr. Also signals with longer periods in the range between  $\sim 3$  and 5 yr appear in the Power Spectral Densities (PSDs). The same frequency analysis performed on the ICs reveal that the 2 yr cycle is now confined in the third IC. All the three ICs show long period signals, however the first two ICs have a flat behaviour for most of the time. The third IC instead has a peak for periods of  $\sim 4$ -5 yr. The time span studied consists of 6.5 yr, limiting my possibility to accurately detect periodic signals with such low frequencies. In order to better understand the origin of the third IC it would be necessary to expand the timeline back in time, or continue collecting data in the future.

For the moment, I have presented the results of a more and more refined analysis of the displacement time series related to the Emilia seismic

sequence. I have motivated the assumptions behind the conceptual model consisting of a linear trend and two seasonal signals of annual and semi-annual period. Finally I have detected two transient signals occurring during the post-seismic stage. Such signals are statistically independent in the time domain, if I consider both the pre- and post-seismic stages. If I take a look to the source space, or the independent components space, I can recognize that there is a correlation between the first and second IC for a subsample of times. In particular, this correlation holds for the data belonging to the post-seismic stage (see Figure 4.14). This means that an ICA performed on the entire dataset, considering both pre- and post-seismic epochs, does not find the independent sources in the post-seismic stage only. In order to prop-

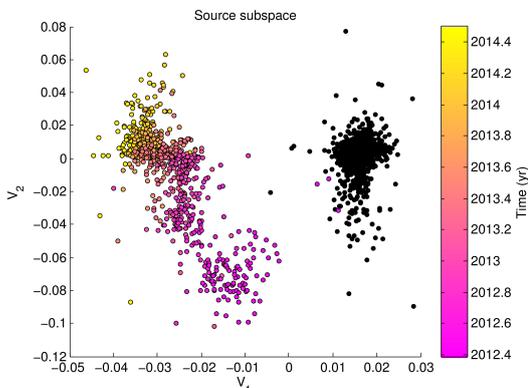


Figure 4.14: Source subspace V1-V2. Black dots represent points belonging to the pre-seismic stage. Coloured dots belong to the post-seismic stage.

erly separate the independent components that generate the observed data during the post-seismic phase, I perform an ICA reducing the time span and considering only the epochs after the May 20<sup>th</sup> mainshock. The results of the temporal sources, obtained both from a PCA and an ICA, are shown in Figure 4.15. An ARD criterion indicates three sources as the proper number of components for the analysis. The ICA decomposition separate two different decay processes, and a cyclic signal (Figure 4.15b, 4.15d, and 4.15f). Conversely, once again the PCA shows a mix of processes (Figures 4.15a, 4.15c, and 4.15e). In order to get more insights into the physical processes that play a role during the post-seismic stage, it is necessary to characterize also the spatial patterns of the ICs. I am going to discuss them, as well as the possible origin of such signals, in the next Section.

#### 4.2.4 Discussion

Figures 4.16 and 4.17 summarize the ICA decomposition of the cleaned post-seismic time series (i.e., after having removed the CME, a linear long-term trend, and two sinusoid functions with constant amplitude, and annual

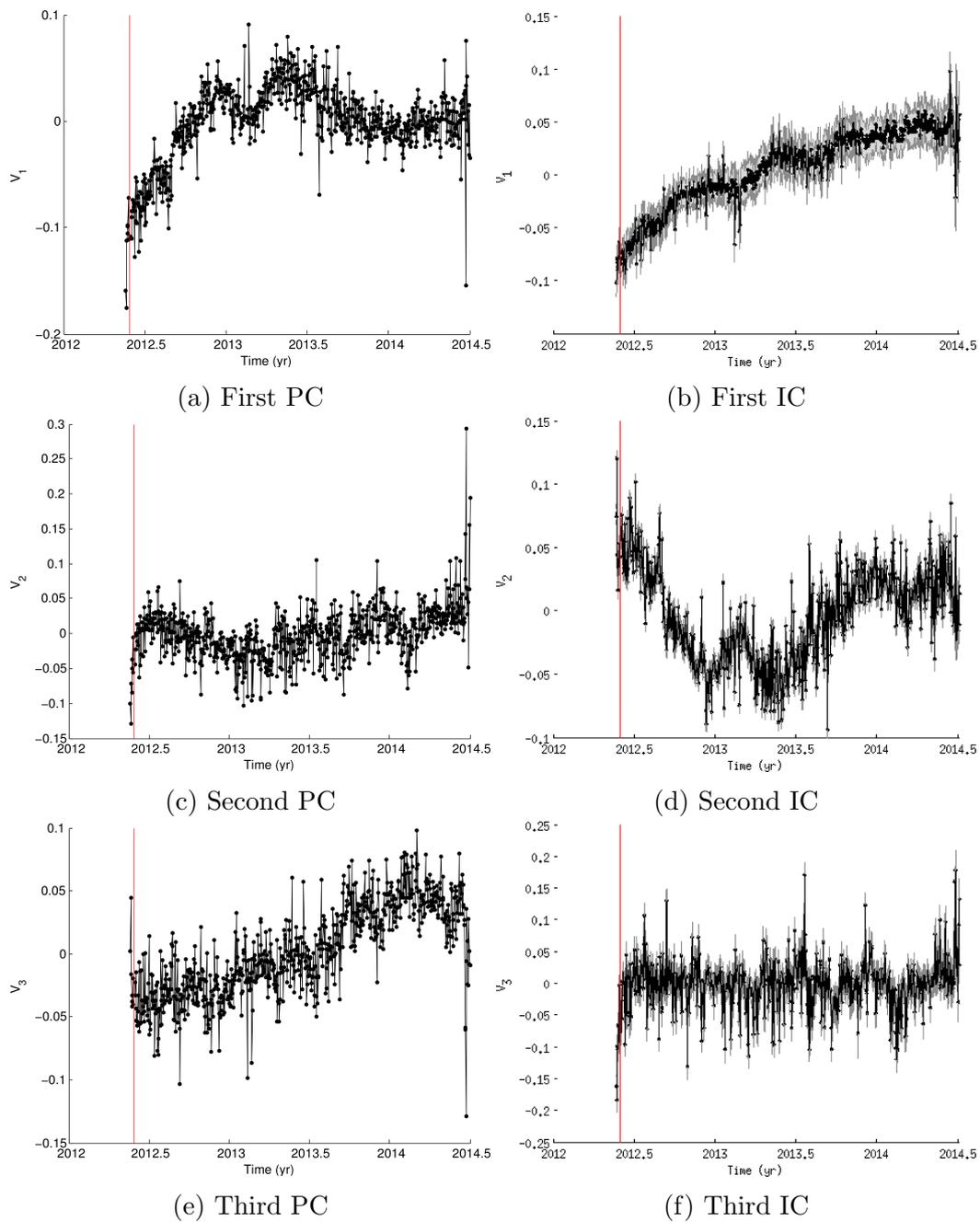


Figure 4.15: PCA and ICA - linear trend, annual and semi-annual signals removed - only post-seismic analysis

and semi-annual period). The temporal evolution  $V$  of each component is plotted at the top of each Figure. The spatial response  $U$  is plotted with

vectors for the horizontal components, and with a colour scale for the vertical one. Both  $U$  and  $V$  are non-dimensional, since the unit of measurement is carried by the weight  $S$  (top-right corner of each Figure), as it was for the SVD formalism (see Section 1.1.2). This means that, given a vector for the horizontal components and a scalar for the vertical component associated with a particular station  $j$ , when the  $V$  temporal function increases (decreases), then the position of the station  $j$  moves toward east-north-up (west-south-down) if the vector is pointing toward east-north and the scalar associated to the vertical component of that station is positive, i.e. red coloured. In order to correctly interpret this kind of plot it is important to stress that the vectors and the scalars associated to each station do not show the direction of the motion of the station itself. This is the case only if the corresponding temporal source is a monotone increasing function. The result shown in Figures 4.16 and 4.17 corresponds to state that the observed post-seismic displacements can be explained as the linear combination of two decaying processes, and a periodic contribution. Let me analyse them one by one, in both time and space domain.

The temporal evolution of the first IC resembles very closely a typical post-seismic decay usually associated with afterslip on the fault plane. Such a phenomenon can be explained by rate-and-state friction laws, and consists in aseismic slip occurring in (presumably) stable regions of the fault, namely velocity strengthening regions, experiencing a positive Coulomb Failure Function variation during the earthquake rupture [Marone et al., 1991]. A fit with a logarithmic function of the post-seismic decay represented by the first IC is shown in Figure 4.18a. In particular, I use a three parameters model as the one described by equation 3.16. The misfit between the best model and the first IC is shown in Figure 4.18b. Such a model has a reduced  $\chi^2$  value of  $\sim 2.2$ . A value greater than 1 for the reduced  $\chi^2$  indicates that the fit has not fully captured the data, or that the error variance has been underestimated. Both the possibilities can be valid in this case. Firstly, the IC number 1 may contain the information not only of the afterslip related to the first event, but it may summarize both the afterslip processes, occurred on two different structures. It follows that a simple logarithmic function may be inappropriate to fully explain the observation. Secondly, the pdf of the IC number 1 is unimodal but asymmetric. Each point of the temporal source is obtained as the sum of a mix of Gaussians (see Chapter 2), and the final variance on such points is calculated considering each of them normally distributed. This may be a wrong assumption, and may be the reason of an underestimation of the variance, and hence the error bar associated to each point. Another interesting feature to notice in the misfit plot is that a semi-annual signal persists. This is the main reason for the reduced  $\chi^2$  value

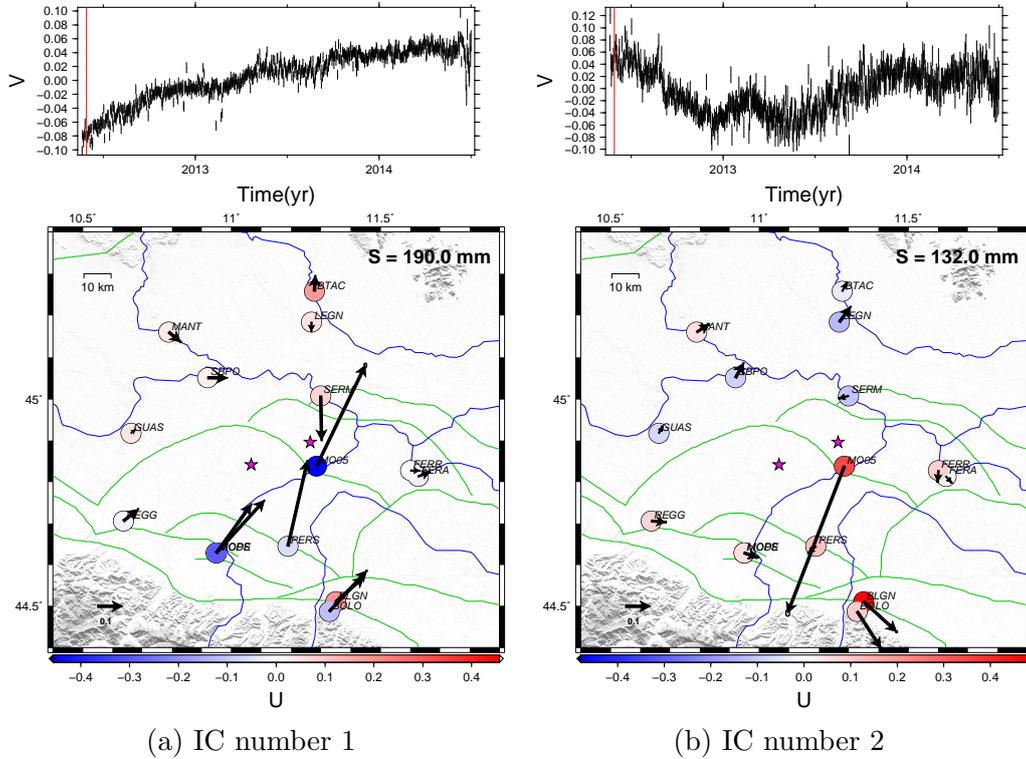
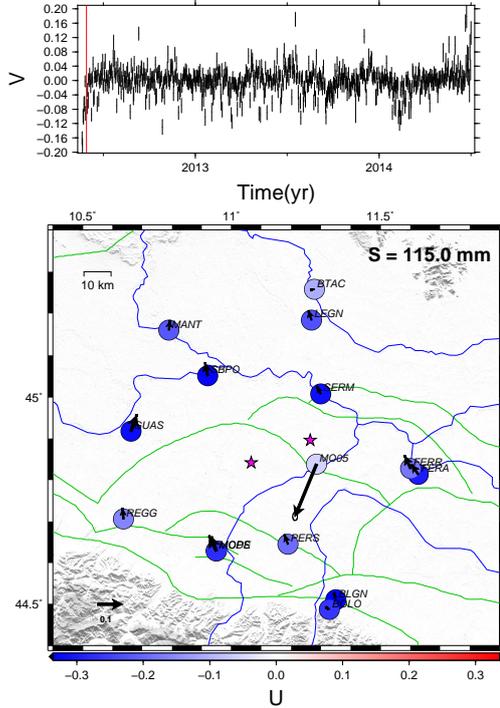


Figure 4.16: Spatial pattern (U), weight (S), and temporal evolution (V) associated with the first and second ICs relative to the post-seismic time series after having removed a linear trend and two seasonal signals, with annual and semi-annual period.

greater than 1. A 2.15 cycles/yr peak is observed in the PSD of the misfit time series (Figure 4.18c), and indicates that the sinusoidal function with constant amplitude and period used to remove the semi-annual signal is not working properly. An interesting feature is that the oscillations mitigate over time, and tend to vanish about 2 yr after the mainshock. It could be that a secondary signal, generated by the co-seismic rupture, is superimposed to the afterlip process, but its power is too tiny to be properly detected, and further considerations would not be reliable.

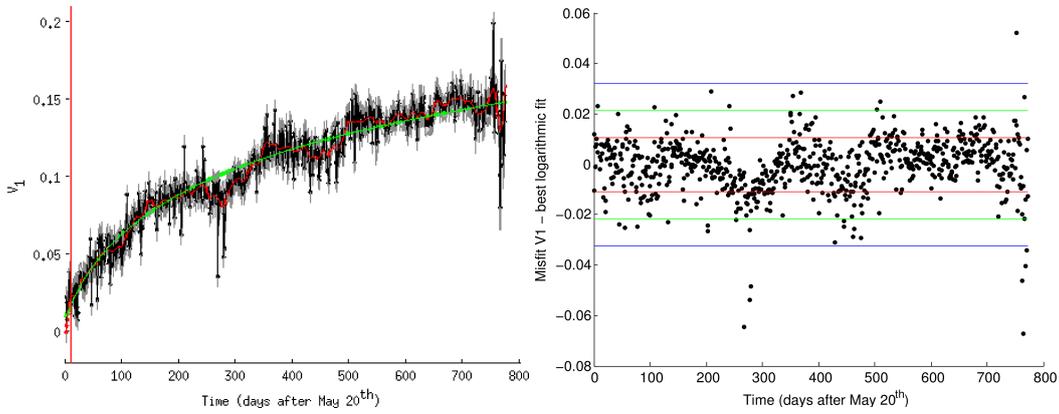
The mean east, north, and vertical displacements associated to the first IC are  $\sim 2.1$ ,  $3.4$ , and  $3.0$  mm. The highest displacements in the east, north, and vertical directions are  $\sim 6$ ,  $12$ , and  $13$  mm, respectively, and have been recorded at station MO05. The generated displacements are similar in magnitude, but the uncertainty on the vertical position is usually about three times larger than on the horizontal. It follows that the vertical motion is



(a) IC number 3

Figure 4.17: Spatial pattern (U), weight (S), and temporal evolution (V) associated with the third IC relative to the post-seismic time series after having removed a linear trend and two seasonal signals, with annual and semi-annual period.

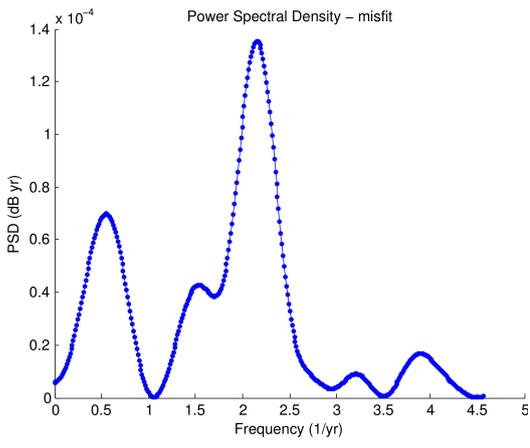
mostly buried in the noise. Indeed, the average SNR for the east, north, and vertical displacements produced by the first IC is  $\sim 2.8$ ,  $7.2$ , and  $0.4$ , respectively. This means that the most reliable interpretation can be deduced from the analysis of the horizontal components. The only two stations consistently showing a  $\text{SNR} > 1$  for the vertical component are MO05 and MOPS. Both of them subside during the post-seismic stage (see Appendix C.2). Moreover, performing an ICA on the vertical time series alone, there are no ICs showing a long-term decay like the one of Figure 4.18a. For this reason, I consider only the horizontal pattern for the interpretation of the first IC, and only the MO05 and MOPS stations as representative for the vertical pattern. The motion of the stations along the NS direction shows a sharp separation between the hangingwall (moving northward) and the footwall (moving southward). The motion along the EW direction is much smaller if compared to the NS one, as expected if we associate the first IC to afterslip on the fault plane. All the stations show an eastward motion, except the three stations belonging to the footwall and nearly orthogonal to the strike (namely, BTAC, LEGN, and SERM). In order to verify if this IC can be explained by afterslip, an inversion of the superficial displacements is needed [Kositsky and Avouac, 2010]. Such an inversion can give some hints on the frictional properties of



(a)

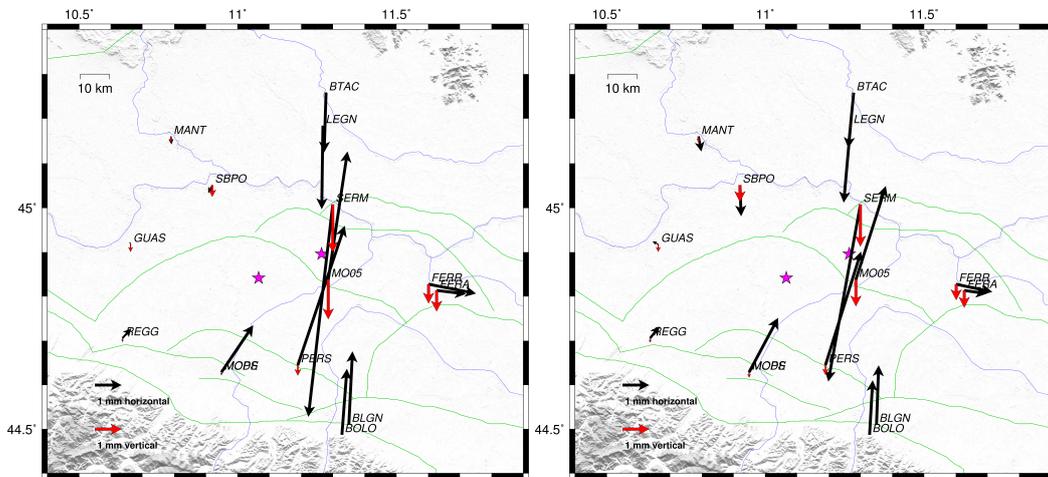
(b)

Figure 4.18: (a) Logarithmic fit to the post-seismic stage of the first IC (shown in Figure 4.16a). The green line indicates the best logarithmic model, obtained for the parameters  $m_1 = 0.01$ ,  $m_2 = 0.05$ , and  $m_3 = 52$  d. The red line is a 10 d window filtered version of the IC. The vertical red line corresponds to the May 29 mainshock, and the time 0 d corresponds to the May 20 mainshock. (b) Misfit between the post-seismic  $V_1$  and the best model of Figure (a). The red, green, and blue lines indicate the 1, 2, and 3 times RMSE thresholds. (c) Power Spectral Density of the misfit of Figure (b).



(c)

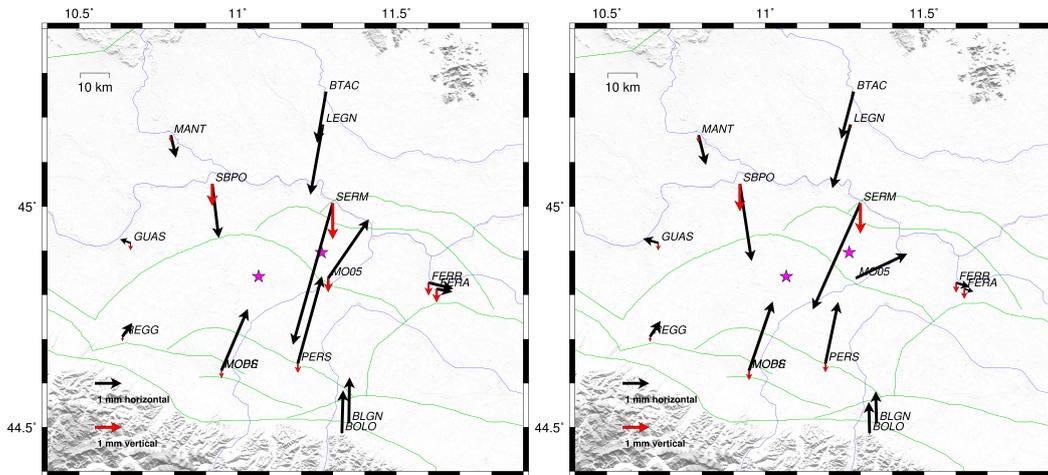
the fault [e.g. Johnson et al., 2006, Perfettini et al., 2010, Gualandi et al., 2014]. Unfortunately, for this study the number of GPS stations available is limited, and likely not sufficient to obtain a reliable non-uniform slip distribution from the inversion of the displacements at the surface. Anyway, here I provide some hints on the position of the patches likely activated by afterslip on the two fault planes, performing some forward models. I can qualitatively explain the observed post-seismic patterns of displacement by forward modelling using uniform slip rectangular dislocations constrained to lie on the assumed fault planes. For the latter ones I use a geometry derived



(a) 100% fault 1 - 0% fault 2

(b) 75% fault 1 - 25% fault 2

Figure 4.19: Direct models - horizontal (black) and vertical (red) predictions. Fault 1 refers to the seismogenic source of the May 20<sup>th</sup> event. Fault 2 refers to the seismogenic source of the May 29<sup>th</sup> event.

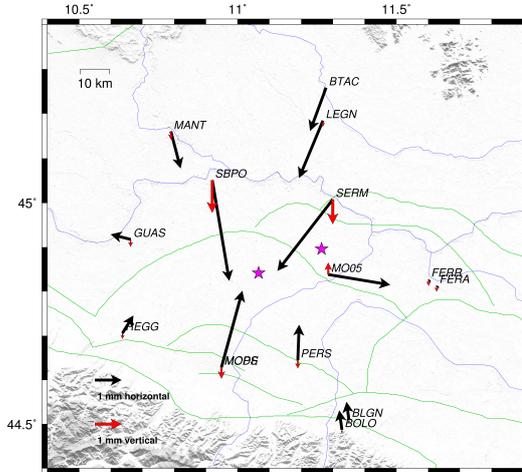


(a) 50% fault 1 - 50% fault 2

(b) 25% fault 1 - 75% fault 2

Figure 4.20: Direct models - horizontal (black) and vertical (red) predictions. Fault 1 refers to the seismogenic source of the May 20<sup>th</sup> event. Fault 2 refers to the seismogenic source of the May 29<sup>th</sup> event.

from the relocated seismicity catalogue of Chiarabba et al. [2014]. The technique adopted is the same described in Gualandi et al. [2014]. For the length



(a) 0% fault 1 - 100% fault 2

Figure 4.21: Direct models - horizontal (black) and vertical (red) predictions. Fault 1 refers to the seismogenic source of the May 20<sup>th</sup> event. Fault 2 refers to the seismogenic source of the May 29<sup>th</sup> event.

of the dislocation I use 20 km (same length as the mainshock fault [Cesca et al., 2013]), and for the width I use 2 km. Moreover, I assume a uniform slip such that the equivalent seismic moment released during the first 2 yr after the mainshock is equal to an earthquake of moment magnitude 6.0. This estimation of the seismic moment associated to the post-seismic stage is deduced in Appendix D. With this simple model I can qualitatively explain the displacements associated to the first IC and interpret it as the result of afterslip on the rectangular dislocation. By changing the depth of the rectangular dislocation, the only models qualitatively in agreement with the observation confine the afterslip on the first fault plane to a range of depth between 3 and 6 km, and on the second fault plane between 5 and 11 km. A deeper afterslip for both the faults would produce uplift for the MO05 and MOPS stations. Instead, a shallower afterslip on the fault of the second event would produce uplift for the MOPS station. I can not exclude that also deep/shallow regions around the co-seismic rupture of the first/second event experienced afterslip, but I can state that the greater amount of afterslip occurred in the range of depth reported above. The same exercise can be done to fit the co-seismic data, finding results in agreement with the uniform inversion performed by Serpelloni et al. [2012]. From table 1 of Serpelloni et al. [2012], I can deduce that the majority of the afterslip relative to the first mainshock occurred at a shallower depth w.r.t. the co-seismic rupture. On the contrary, the main afterslip on the fault of the May 29<sup>th</sup> mainshock is located at a greater depth w.r.t. the co-seismic slip distribution. Figures 4.19, 4.20, and 4.21 show different configurations, progressively increasing the percentage of afterslip on the second structure. The station closest to the fault system is MO05, and

its post-seismic deformation is clearly not explained if I consider afterslip occurring only on the second structure (see Figures 4.16a and 4.21). Instead, considering afterslip only on the first fault, the displacements of westernmost stations (GUAS, MANT, REGG, and SBPO) are underestimated if compared to the displacements of the other stations of the network (see Figures 4.16a and 4.19a). From this simple direct model I deduce that the first IC represent the post-seismic response related to the afterslip occurred on both the faults activated during the seismic sequence. A good repartition of the afterlip seems to be the one of Figure 4.19b, where the afterslip on the fault of the May 20 event is 3 times larger than the one occurred on the May 29 structure. Interestingly, this ratio corresponds to the ratio between the seismic moments of the co-seismic events found by Serpelloni et al. [2012].

As regard the second IC (see Figure 4.16b), its temporal evolution shows an oscillation with a period longer than 1 yr. A frequency analysis reveals a peak at about 0.5 cycles/yr, corresponding to a period of about 2 yr. The time spanned in the post-seismic stage is  $\sim 2.1$  yr. Consequently, a longer time span is necessary to properly understand and characterize this signal. The only stations showing a SNR greater than 1 are BLGN (east and vertical directions), and MO05 (east and north directions). In particular, the data at my disposal for MO05 end at epoch  $\sim 2013.0$ , making impossible to recognize such a fluctuation for this station. This means that this signal could be related to some local effects that affect the BLGN station.

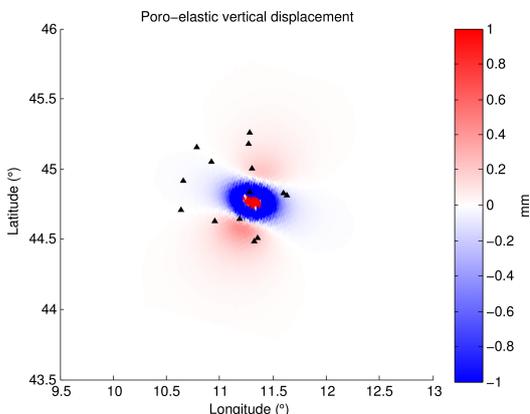


Figure 4.22: Poro-elastic displacement at the surface on an elastic half-space computed using an undrained and drained Poisson's ratio of 0.3 and 0.12, respectively, if only the first 2 km depth layer reaches the drained condition, while the deeper half-space keeps undrained. The uniform co-seismic slip distribution of Serpelloni et al. [2012] is adopted (Belardinelli, personal communication). Surface displacement were computed by means of Edgrn-Edcmp [Wang et al., 2013]. Black triangles indicate the cGPS stations.

The temporal evolution of the third component (see Figure 4.17) shows a very fast decay that starts immediately after the first mainshock, and lasts about 2 weeks. The average SNR related to this third component is  $\sim 0.1$ , 0.5, and 1.0 for the east, north, and vertical directions, respectively. This implies that the source of displacement affects mainly the vertical component. A process inducing post-seismic subsidence is due to the large amount of extruded sand that produced important emptying and compaction in the liquefied beds. This phenomenon produced at the surface a localised subsidence, sometimes accompanied by ponding [Emergeo, 2013]. The region interested by such a deformation mechanism cover an area of about 1200 km<sup>2</sup>. Nevertheless, the liquefaction may be associated to relevant but localised subsidence or bulging related to sediments extrusion. The observed spatial pattern of the third IC is instead nearly homogeneous, and it affects the whole network. Post-seismic deformation in the upper crust has been explained also advocating the occurrence of pore-pressure transients [e.g. Peltzer et al., 1996, Jönsson et al., 2003]. Changes in fluid pressure induce strains, and conversely, changes in stress or strain induce changes in pore pressure [Segall, 2010]. Readjustment of pore fluid pressure (poro-elastic rebound) in the crust can cause volume changes following earthquakes, resulting in post-seismic subsidence over areas of co-seismic pressure increase and vice-versa [Fielding et al., 2009]. Unfortunately, generally the equilibrium and flow equations for a porous elastic medium are fully coupled, and there are only some special situations in which the governing equations uncouple. Instead of using analytical solutions of the poro-elastic problem, it is of common use to determine the cumulative post-seismic response as the one where the crust relaxes from the completely undrained state (immediately after the earthquake) to the fully drained (or infinite time) response [Segall, 2010]. In order to calculate the cumulative poro-elastic deformation, I can use a co-seismic slip distribution to compute the displacement field for a dislocation in a homogeneous elastic half-space using both undrained and drained Poisson's ratios. Then, the poro-elastic deformation corresponds to the difference between the displacements in the drained and undrained conditions. For all the calculations that follow, the uniform co-seismic slip distribution of Serpelloni et al. [2012] has been used. For a thrust fault, a cumulative poro-elastic response in a homogeneous half-space predicts uplift for the entire region orthogonal to the strike, and subsidence along the strike and outside the fault. This model is in contrast with the observation consisting in a general subsidence of the network. A possibile mechanism to explain this discrepancy consists in considering a layered half-space, where the drained condition is reached only by the shallower sandstone layer. Indeed, the thermal gradient of the region is 40 K/km [Nespoli et al., in prep.], sufficiently high to have gaseous

state fluids already at few km of depth. This implies that the pore pressure effects on the solid matrix can be neglected, and I can consider the same Poisson's ratio for the undrained and drained condition below the first 2 km. From this assumption, and using a 2 km thick sandstone layer, it follows that the predicted poro-elastic vertical displacement consists of a restricted region around the co-seismic rupture that experiences uplift, surrounded by a wider region experiencing subsidence. This subsidence region extends along strike, while in the direction orthogonal to the strike two uplifting lobes appear (see Figure 4.22). The boundary of the described two regions are known with a poor level of accuracy, and the extension of the subsidence region is hardly computable using a uniform co-seismic slip distribution and an a priori depth of the sandstone drained layer. I have shown this simple conceptual model in order to point out that the fast decay associated to the third IC may be related to poro-elastic rebound, but the spatial sparsity of the data does not allow to define a more refined model.

#### 4.2.5 Conclusive remarks

I have presented the results from the analysis of time series from 15 cGPS stations located at a distance  $< 50$  km from the fault system that generated the seismic sequence occurred in May 2012 in the Emilia region (northern Italy). The proposed scenario is interesting because of the high number of different sources of displacement acting together at the same time, with associated deformations and displacements that are toward their lower bounds of the technique resolution. From the analysis of the time series rotated in a fixed Eurasian reference frame, the modified vbICA algorithm is able to detect 5 different signals: one mostly linear trend, two nearly annual oscillations, and two signals related to post-seismic processes (see Figure 4.10c). Because of the non-stationarity and the high power of the linear signal, in order to facilitate the separation of the remaining signals, I have detrended the time series and redone the analysis. Two annual signals  $\sim 90^\circ$  out of phase persist. The correct interpretation of the annual signals in the displacement time series is still not fully understood. Dong et al. [2002] have studied the annual variations in GPS time series at a global scale, and found that  $\sim 40\%$  of the power of the observed annual vertical variations in site positions can be explained by the joint contribution from pole tide effects, ocean tide loading, atmospheric loading, nontidal oceanic mass, and groundwater loading. Moreover, the coordinate time series used in this study have been derived from cGPS data processed as 24 hour batch. For this reason, residual semi-diurnal and diurnal crustal tide signatures are under-sampled, resulting in aliased periodic signals in the coordinate time series. A secondary aliasing

effect arises from the repeat period of the satellite orbits being longer than the Nyquist period of the semi-diurnal and diurnal tidal signatures. Penna and Stewart [2003] have shown that the beating of the two aliased signals invariably results in spurious signatures in the time series with semi-annual and annual period. After having verified the presence of annual and semi-annual periods in the oscillatory signals (Figure 4.12), in order to better focus on the transient signals related to the seismic cycle I have decided to filter out two sinusoidal functions with a constant amplitude and period from the time series. This analysis has led me to the individuation of two post-seismic decay signals, very similar to those detected by the first analysis on the raw time series. Considering a time span that takes into account both the pre- and post-seismic epochs, I find that one of the two detected ICs shows a non-monotonic behaviour. In particular, such IC would correspond to a displacement source that activates and bounces back to its previous condition. In a non-volcanic scenario such the Emilia one, I do not know any post-seismic process that can act in a similar way. After having noticed the strong correlation sussisting between the post-seismic data of the ICs related to the two decays (Figure 4.14), I have decided to redo the analysis only on the data after the May 20 mainshock. Differently from a PCA approach, using a vbICA algorithm I detect two different monotonic decays: a first one that lasts hundreds of days, and a second one completed in few days. This finding reveals the existence of more than one deformation process involved during the post-seismic phase of the Emilia seismic sequence, and demonstrates that the selection of the time span to be analysed is crucial in determining the ICA results and their interpretation in term of sources. I have shown how the slow decay ( $IC_1$ ) can be associated with afterslip occurred on the two fault planes activated during the sequence. In particular, I suggest that shallow afterslip is associated to the fault of the May 20 mainshock, while on the fault of the May 29 mainshock afterslip mainly occurred at a depth greater than the one of the co-seismic slip distribution. The vbICA approach is not able to separate the two afterslip processes on the two different seismic structures. It may be the case that the two pdfs related to the two processes are too similar to be distinguished by an ICA approach at all. This finding does not allow me to determine if the fault of the second mainshock experienced aseismic slip after the 20<sup>th</sup> of May and before the 29<sup>th</sup> of May, as proposed by Pezzo et al. [2013]. I have also shown that the fast decay ( $IC_3$ ) is mainly associated with the vertical component, and I have proposed a simple dislocation model in a stratified poro-elastic half-space in order to qualitatively explain the observations. The results suggest that only a shallow layer of sandstone has been drained during the post-seismic stage, generating the subsidence of the network associated to the rapid post-seismic decay captured by the third

IC (Figure 4.17). A denser spatial coverage of the area would increase the detail of information about the processes involved, but unfortunately, the available InSAR interferograms do not allow to capture such a fast transient. In conclusion, it is possible that the fluids played some role in the seismic sequence, especially during the time between the two mainshocks. For this reason, a realistic model should take their effect into account.

## 4.3 The 2006 Slow Slip Event in Mexico

### 4.3.1 Introduction

In Chapter 3 I have applied the vbICA algorithm to synthetic data relative to typical seismic cycle phases, and in Sections 4.1 and 4.2 of this Chapter I have applied it to co- and post-seismic displacements associated to two moderate earthquakes recorded by GPS measurements. In all the case studies the main signal to be detected was a post-seismic one. In order to show the capability of the vbICA technique to capture the spatio-temporal characteristics of different processes, and its independence from the particular temporal evolution of the source under study, here I show the results relative to the analysis of position time series from 9 continuous GPS stations that recorded a Slow Slip Event (SSE) in the Guerrero region (Mexico). The Guerrero subduction segment is part of the Middle American Trench (MAT). It extends from  $261^{\circ}\text{E}$  to  $258.4^{\circ}\text{E}$  and encompasses the NW Guerrero gap where no large subduction earthquake occurred since 1911 ( $M_S$  7.8) [Ortiz et al., 2000] and the SE Guerrero gap which overlaps the rupture lengths of the 1907 ( $M_S$  7.9), 1957 ( $M_W$  7.8) and 1962 ( $M_W$  7.1 and  $M_W$  7.0) earthquakes [Ortiz et al., 2000, Anderson et al., 1994] (4.23).

In January 1997, a continuous GPS receiver was installed in Cayaco, Guerrero region (Mexico). Since then, the GPS network grew up to more than 20 continuous stations, and it has been widely exploited for the detection of transient displacements at the surface. In particular, since 1998 one large SSE has been observed in Guerrero approximately every 3-5 yr: in 1998 [e.g. Lowry et al., 2001, Larson et al., 2004, Vergnolle et al., 2010], 2001/2002 [e.g. Kostoglodov et al., 2003, Iglesias et al., 2004, Vergnolle et al., 2010], 2006 [e.g. Larson et al., 2007, Correa-Mora et al., 2009, Vergnolle et al., 2010, Radiguet et al., 2011, Hooper et al., 2012, Cavalié et al., 2013], 2009/2010 [e.g. Walpersdorf et al., 2011], and 2014 (N. Cotte, personal communication).

Most of the geodetically instrumented subduction zones (e.g., Alaska, Cascadia, Costa Rica, Japan, Mexico, New Zealand) have revealed the existence of SSEs [e.g., Schwartz and Rokosky, 2007, and references therein].

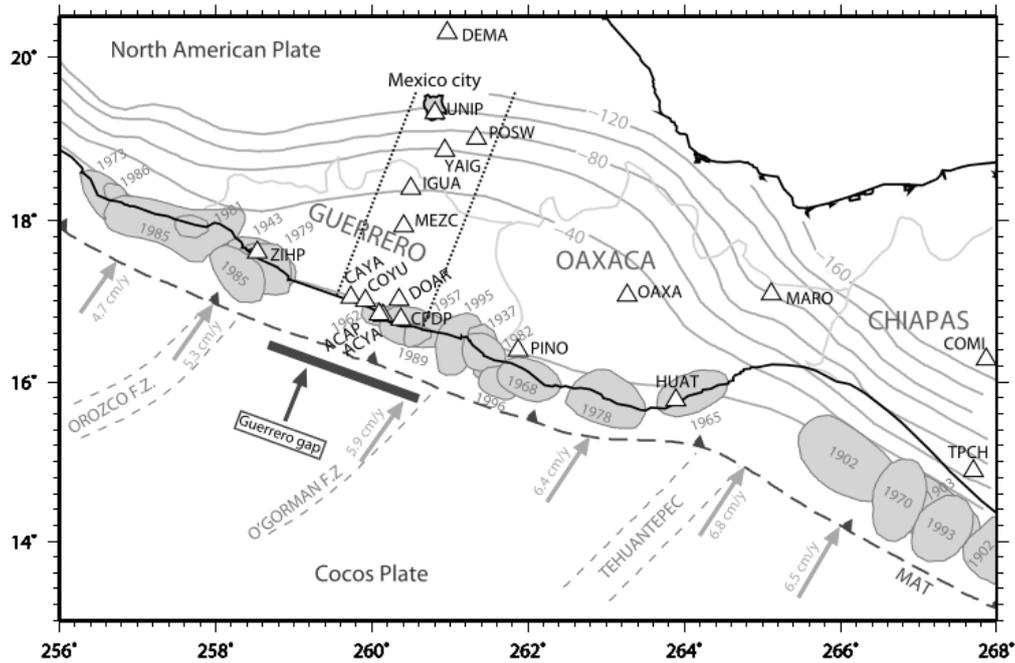


Figure 4.23: Seismotectonic map (after Franco et al. [2005]) and GPS station locations (white triangles). The arrows indicate the direction and magnitude of NUVEL1A relative plate motion between the Cocos and North American plates [DeMets et al., 1994]. The gray patches represent the major earthquake rupture zones. The gray lines with negative values (in km) indicate the slab interface isodepths.

The discovery of SSEs has revolutionized the understanding of how faults accommodate relative plate motions. Faults were previously thought to relieve stress either through continuous aseismic sliding, or as earthquakes resulting from instantaneous failure of locked faults. In contrast, SSEs proceed so slowly that slip is limited and only low-frequency (or no) seismic waves radiate [Peng and Gomberg, 2010]. In some occasions SSEs have been detected before the occurrence of large earthquakes, and it has been proposed that the aseismic slip associated to the SSE triggered the subsequent mainshock [e.g. Ito et al., 2013, Ruiz et al., 2014]. This observation elects them valuable seismic precursors to investigate.

The Guerrero 2006 SSE is one of the world's largest observed SSE, and it offers a unique opportunity to study this kind of phenomena. Usually, SSEs studies only model the cumulative displacements, giving a static image of the slip evolution. For the first time the slip propagation on the subduction interface during the 2006 SSE was analysed by Radiguet et al. [2011]. In

particular, they performed an inversion in two steps. First, they inverted the cumulative GPS displacements in order to retrieve the total slip amplitude. Second, they inverted for the initiation time and duration of the slip, using a linearized least-squares inversion procedure and assuming a functional form for the slip function. Here, I use the vbICA approach in order to retrieve the original sources that generated the observed displacements, with the goal of studying their evolution through space and time. In so doing I will characterize only the displacement at the surface, but still this enables me to extract information on the SSE source without performing any inversion for the slip on the subduction interface. The goal of this study is to show that the spatio-temporal evolution of the 2006 SSE in Mexico can be explained using a limited number of ICs, which bring interesting information about the propagation of the slip source.

### 4.3.2 Data

The Guerrero 2006 SSE was recorded at 15 cGPS stations of the Mexican permanent GPS network “SSN-Sismologia-UNAM” (<http://usuarios.geofisica.unam.mx/vladimir/gpsred/gpsred.html>). The stations are located along the coast and along a linear transect perpendicular to the trench, between Acapulco and the north of Mexico City (Figure 4.23). I use the detrended GPS position time series processed by Vergnolle et al. [2010], which show a noise reduction of 50 per cent w.r.t. previous studies of the Guerrero region (see Appendix C.3). Differently from Vergnolle et al. [2010] and Radiguet et al. [2011], who filled the gaps with a linear interpolation of a low-pass filtered version of the original time series, I directly use the daily sampled position time series. Radiguet et al. [2011] discarded the stations DEMA, DOAR, HUAT, OAXA, PINO, and ZIHP from the analysis of the temporal evolution of the SSE, because of the important gaps in the data. Considering all the 15 cGPS stations at my disposal, the percentage of missing data in the time interval between the first and the last epoch available (2005.0014 and 2008.2391, respectively) is  $\sim 34\%$ , while removing the stations previously listed the percentage of missing data drops to  $\sim 28\%$ . The qualitative interpretation of the final results is unaffected by the removal of the stations with missing data, but in order to constrain better the temporal evolution of the SSE, here I show the results of the analysis performed on the subset of 9 over 15 cGPS stations.

### 4.3.3 Results

An ARD criterion suggests to keep two ICs, but I prefer to maintain also the third IC in order to get a clearer result on the first two ICs, and to separate a cyclic contribution. Figures 4.24 and 4.25 show the temporal functions associated with the three PCs (left) and ICs (right). I have initialized the vbICA algorithm with two different PCA approaches: the one described by Bailey [2012] (EMPCA), and the one described by Srebro and Jaakkola [2003]. The former maintains the PCs orthogonal one to the other, while

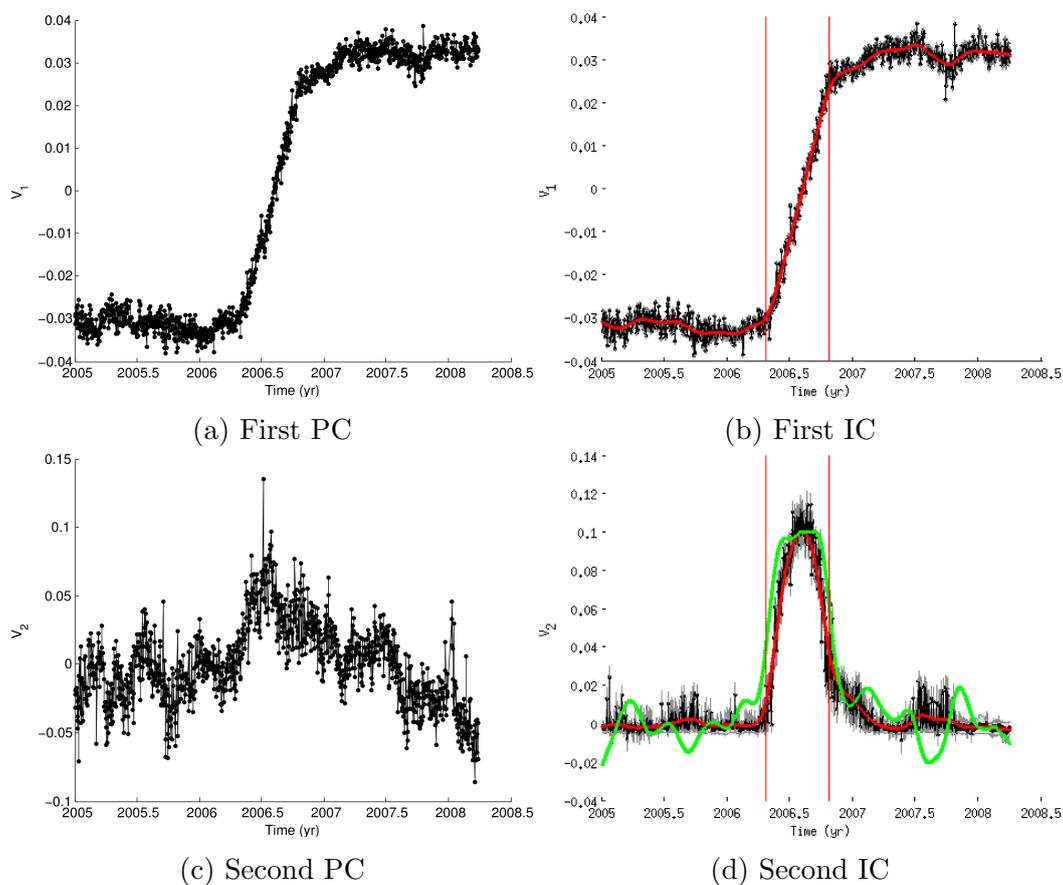


Figure 4.24: PCA and ICA - linear trend removed - first and second components. (b) The vertical red lines correspond to the delimitation of different stages (see main text). The red line is a zero-phase digital filtered version of the IC, obtained using a 6 week window for the numerator coefficients (see the Matlab function `filtfilt`). This filtering allows to discard the high-frequency daily scatter. (d) Same as (b); the green line corresponds to a filtered version of the discrete time derivative of the red line of (b).

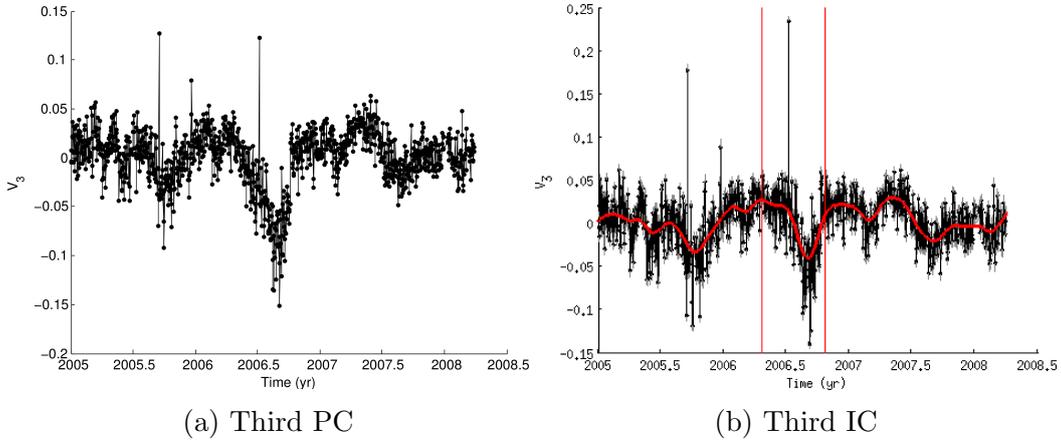


Figure 4.25: PCA and ICA - linear trend removed - third component. Colours as in Figure 4.24.

the latter does not have this constraints in performing a low-rank approximation. For this reason the second approach (here shown in Figures 4.24a, 4.24c, and 4.25a) is closer to an ICA decomposition. The ICs are not significantly affected by the initialization point.

In order to describe the temporal evolution of the SSE, I divide the time spanned into three stages:

- I) from 2005.0 to 2006.3: time before the onset of the SSE
- II) from 2006.3 to 2006.8: rise time of the SSE
- III) from 2006.8 to 2008.2: time after the end of the SSE

During stage I), the first IC is nearly flat. At time 2006.3 the deformation associated to the first source starts, and rapidly increases until epoch 2006.8. The time span covered by this deformation episode is 0.5 yr, i.e. 182.5 d, and, following Radiguet et al. [2011], I refer to it as the rise time  $t_r$  of the source. Finally, the third stage is nearly flat, and shows some oscillations around a constant value, probably due to the presence of some seasonal signal, not originally removed from the time series.

The bell-shaped second IC is atypical for slip events. It can still be splitted into 3 stages: a first flat one, a second bell-shaped, and a third nearly flat, reaching a value equal to the one of the first stage. It is interesting to notice that this IC resembles very closely the time derivative of the first IC (green line in Figure 4.24d). It is unlikely that the second IC may be associated with slip on the subduction fault. This pattern, in fact, implies that a region of the fault plane first move in one direction, and than goes back

to the same starting position. Such a behaviour is however not physically reasonable, maintaining during the slip event the same tectonic regime.

Finally, the third IC shows periodic variations. A frequency analysis highlights a main peak at about 0.8 cycles/yr, and a secondary peak at 1.75 cycles/yr. These frequencies do not correspond to the typical annual and semi-annual periods, and can thus potentially be related to signals not linked with seasonal variations.

Figure 4.26 shows both the temporal and the spatial response of the three ICs. The notation is the same of Figures 4.16 and 4.17.

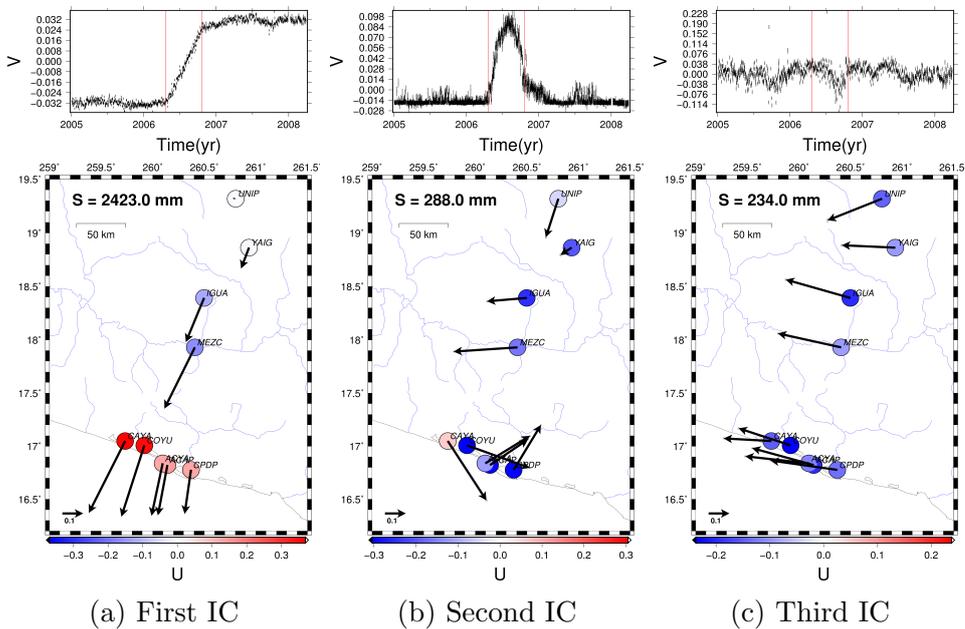


Figure 4.26: Temporal and spatial patterns for the ICs.

The interpretation of the first IC seems straightforward. Slip occurred on the subduction interface and has generated a displacement at the surface, mainly affecting the North direction, as expected by the tectonic setting. All the stations consistently move southward, in a direction nearly orthogonal to the coast. The vertical motion is sharply geographically separated: the stations along the coast show uplift, while the inland stations show subsidence. This corresponds to the case if the source is located in a region between the coastal and the inland stations.

The interpretation of the second IC is more tricky. Its weight is  $\sim 8$  times smaller than the one relative to the first IC. Because of its shape in time, it acts as a secondary source generating a null cumulative displacement. Indeed, during stages I) and III) all the stations are nearly at the same position and

do not move. Instead, during the first part of stage II), i.e. when the temporal source (first IC) increases, the stations in the North move mainly toward West and subside, while the stations along the coast move mainly toward East and subside. The opposite motion is generated during the second part of stage II), i.e. when the temporal source decreases. Such a behaviour can not be explained with slip on the subduction interface. In the next Section I propose an explanation for this observation interpreting the temporal evolution of the first IC as the true shape of the source that have generated the data, and hypothesising the possibility for this source to move in the space.

Finally, also the third IC shows a peculiarity feature. All the stations coherently move in the same direction, i.e. the spatial response is homogeneous for the network, and the SNR is greater than 1 only for the east direction. In particular, the orientation of the horizontal responses for the first and third ICs is nearly orthogonal. Thus, for the third IC the horizontal motion occurs in a direction nearly parallel to the coast, i.e. orthogonal to the long term loading velocity (see Figure 4.23), and the direction of motion during the SSE, as from the first IC. This observation, together with the fact that the cycles in the temporal function are not strictly related to seasons, indicates that this IC may refer to a common cyclic source, neither of annual or semi-annual periods, as commonly found in GPS time series.

#### 4.3.4 Discussion

In general, any technique that aims at solving a BSS problem should be able to treat the case of non instantaneous mix and moving sources, but because of its mathematical complexity it is common to assume an instantaneous mix of sources fixed in space. These two assumptions allow to uncouple the spatial and temporal dependencies, and explain the observed data as a linear combination of the temporal evolution of the sources. In particular, the linearly combined quantities correspond to the temporal information related to the sources (source matrix), while the weights are collected in a matrix that contains the information about the relative position between the sources and the receivers (mixing matrix) (for more details, see Chapter 2).

The observation that the second IC resembles the time derivative of the first IC for the Guerrero 2006 SSE requires some additional investigation. Let me now consider only the first two ICs to explain the GPS position time series relative to the 2006 Guerrero SSE. This means that the position time series at the surface can be explained as a linear combination of two static sources (i.e., fixed in space), where one is the time derivative of the other:

$$\mathbf{X}(\mathbf{x}, t) = \mathbf{a}_1(\mathbf{x})s(t) + \mathbf{a}_2(\mathbf{x})\dot{s}(t) \quad (4.1)$$

where  $\mathbf{X}$  is a 3-dim vector corresponding to the position recorded at time  $t$  by the GPS station located at  $\mathbf{x}$ ,  $\mathbf{a}_1$  and  $\mathbf{a}_2$  are two 3-dim vectors containing in each component the corresponding mixing value constant with time, and  $s$  is the time-dependent source. I indicate the second source, i.e. the time derivative of the first, with  $\dot{s}$ . It follows that the dimension of  $\mathbf{a}_2$  is equal to the dimension of  $\mathbf{a}_1$  multiplied by time.

Radiguet et al. [2011] have shown that for the 2006 Guerrero SSE the assumption of not moving sources is wrong. They have also found that the rise time (i.e. the local duration of slip) does not show large spatial variations and is equal to about 185 d. This means that there is a common evolution in time that characterizes the source, and that this source is moving. The rise time of the first IC (see Figure 4.24b) is about 0.5 yr, in agreement with their finding. This observation makes the IC<sub>1</sub> a good candidate for the signal associated to a moving source. So, let me assume that only one source  $s_1(t)$  is generating the observed displacements. The observations say that a second component is needed, so the general formulation for explaining the data with a linear combination of two sources is:

$$X(\mathbf{x}, t) = \mathbf{b}_1(\mathbf{x})s_1(t) + \mathbf{b}_2(\mathbf{x})s_2(t) \quad (4.2)$$

In order to consider the movement of the source  $s_1$ , let me hypothesise that the second source is a copy of the first, but delayed in time of an amount  $\Delta t$ . This is equivalent to say that the source  $s_1$  migrated, so that its contribution to the observed displacements is changed from  $\mathbf{b}_1$  to  $\mathbf{b}_2$ . In this situation the ICA algorithm can not separate the sources  $s_1$  and  $s_2$  because their pdfs are exactly the same. The hypothesis made about  $s_2$  allows me to write the following relation:

$$s_2(t + \Delta t) = s_1(t) \quad (4.3)$$

or, equivalently:

$$s_2(t) = s_1(t - \Delta t) \quad (4.4)$$

In general, in order to move from the value  $s_1(t - \Delta t)$  to the value  $s_1(t)$  it is necessary to add a quantity  $\Delta s_1$  that depends on  $t$  and  $\Delta t$ . I can thus rewrite the general linear combination 4.2 as:

$$\begin{aligned} X(\mathbf{x}, t) &= \mathbf{b}_1(\mathbf{x})s_1(t) + \mathbf{b}_2(\mathbf{x})s_1(t - \Delta t) \\ &= \mathbf{b}_1(\mathbf{x})s_1(t) + \mathbf{b}_2(\mathbf{x})[s_1(t) - \Delta s_1(t, \Delta t)] \\ &= [\mathbf{b}_1(\mathbf{x}) + \mathbf{b}_2(\mathbf{x})]s_1(t) - \mathbf{b}_2(\mathbf{x})\Delta s_1(t, \Delta t) \end{aligned} \quad (4.5)$$

For a time delay  $\Delta t \neq 0$ , this equation can be rewritten using the difference quotient notation:

$$X(\mathbf{x}, t) = [\mathbf{b}_1(\mathbf{x}) + \mathbf{b}_2(\mathbf{x})]s_1(t) - \mathbf{b}_2(\mathbf{x})\Delta t \frac{\Delta s_1(t, \Delta t)}{\Delta t} \quad (4.6)$$

Equation 4.6 can now be compared to equation 4.1. In general, it is not guaranteed that equation 4.6 is valid in the limit  $\Delta t \rightarrow 0$ . Nonetheless, the above considerations induce me to interpret the  $\mathbf{a}_1$  mixing factor as the contribution of all the mixing coefficients relative to the regions activated by the source  $s_1$  during the event, and  $\mathbf{a}_2$  as a quantity directly proportional to the delay time of the source, i.e.:

$$\mathbf{a}_1(\mathbf{x}) = \mathbf{b}_1(\mathbf{x}) + \mathbf{b}_2(\mathbf{x}) \quad (4.7)$$

$$\mathbf{a}_2(\mathbf{x}) = -\mathbf{b}_2(\mathbf{x})\Delta t \quad (4.8)$$

The quantity  $\tau_{ij} = -\frac{a_{2i}(\mathbf{x})}{a_{1j}(\mathbf{x})} = -\frac{b_{2i}(\mathbf{x})}{b_{1j}(\mathbf{x})+b_{2j}(\mathbf{x})}\Delta t$ , with  $i, j = 1, 2, 3$ , is a  $3 \times 3$  matrix, and in particular it is a tensor:

$$\tau(\mathbf{x}) = \begin{pmatrix} -\frac{a_{2x}}{a_{1x}} & -\frac{a_{2x}}{a_{1y}} & -\frac{a_{2x}}{a_{1z}} \\ -\frac{a_{2y}}{a_{1x}} & -\frac{a_{2y}}{a_{1y}} & -\frac{a_{2y}}{a_{1z}} \\ -\frac{a_{2z}}{a_{1x}} & -\frac{a_{2z}}{a_{1y}} & -\frac{a_{2z}}{a_{1z}} \end{pmatrix} \quad (4.9)$$

where I have dropped the dependency of  $\mathbf{a}_1$  and  $\mathbf{a}_2$  from  $\mathbf{x}$  for brevity.

Calculating the characteristic polynomial equation for the tensor  $\tau$  at each position  $\mathbf{x}$ , I find that only the coefficients for the third and second grade of the equation are significantly different from zero. This implies that the eigenvalue  $\lambda = 0$  has algebraic multiplicity equal to 2, and only one real eigenvalue is non-zero. Moreover, for each eigenvalue the geometric and algebraic multiplicity is the same, therefore the matrix is diagonalizable. The trace ( $tr$ ) of a tensor is an invariant for rotation. The value  $\hat{\tau}(\mathbf{x}) = \frac{1}{3}tr(\tau(\mathbf{x}))$  does not depend on the reference system and it is proportional to  $\Delta t$ , the delay time of the effect of  $s_2$  on the receiver w.r.t. those caused by  $s_1$ , via a non-dimensional coefficient  $B(\mathbf{x})$  that depends on the position  $\mathbf{x}$  of the receiver and the position of the sources  $s_1$  and  $s_2$ :  $\hat{\tau}(\mathbf{x}) = B(\mathbf{x})\Delta t$ . I interpret the quantity  $\hat{\tau}(\mathbf{x})$  as the *perceived* time of activation at the station located in  $\mathbf{x}$ . Given two different GPS stations located at  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , if  $\hat{\tau}(\mathbf{x}_1) > \hat{\tau}(\mathbf{x}_2)$  then the station in  $\mathbf{x}_2$  has been activated before the station in  $\mathbf{x}_1$ , and vice-versa. Table 4.4 reports the calculated  $\hat{\tau}$  values for the 9 cGPS stations. The different perceived times of activation directly reflect the movement of the source in the crust, but in order to get the slip on the subduction interface it is necessary to perform an inversion of the surface displacement. I do not have a fault model at my disposal, but looking at Figure 4.27 it is evident a gradient in the scalar field  $\hat{\tau}(\mathbf{x})$ , indicating a migration of the event. In

particular, it started in the North-West region and propagated toward South-East, as also suggested by Radiguet et al. [2011]. The only station showing a  $\hat{\tau}$  value not consistent with the nearby is UNIP. Associating the first IC to the source signal, the station UNIP is the only one that shows a  $\text{SNR} < 1$  in every direction (east, north, and vertical). This can be the reason for the anomalous  $\hat{\tau}$  deduced. For the station YAIG the vertical response is below the threshold  $\text{SNR} = 1$ , and I use instead of  $\hat{\tau}$  the mean of the east and north responses. The first station affected by the SSE (i.e., the one with lowest  $\hat{\tau}$ ) started to show a displacement associated to the event about 54 d before  $t^* = 2006.3$ , i.e. in late February 2006 ( $t_a \sim 2006.1521$ ). The last station involved started its motion in  $t_b \sim 2006.563$  (late July of 2006). The rise time is about 0.5 yr, so the end of the event corresponds to the epoch  $\sim 2007.063$ , i.e. late January 2007. This means that the SSE lasted not less than 11 months. These results are in excellent agreement with the findings of Radiguet et al. [2011].

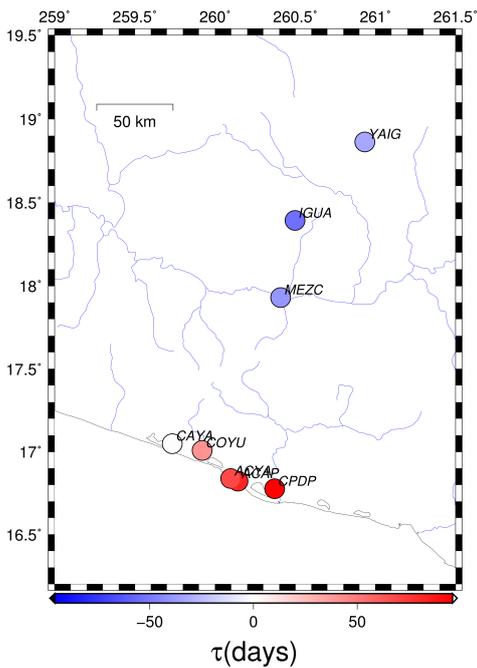


Figure 4.27:  $\hat{\tau}(\mathbf{x})$ . The value for the station UNIP is not shown. For station YAIG  $\hat{\tau}$  is computed using only the East and North direction because they are the only two with a  $\text{SNR} > 1$ , i.e.  $\frac{\tau_{11} + \tau_{22}}{2}$ .

After these considerations, the third component can be looked under a different perspective. If the first component is mainly associated with N-S displacement recorded at the surface, and the second IC gives information about the propagation of the source of the SSE, the third IC mainly affects the E-W motion (see Figure 4.26c). A frequency analysis reveals that the peaks do not correspond to seasonal cycles. This could be an indication that the 2006 Guerrero SSE have had a wavelike behaviour, undetectable by

Table 4.4:  $\hat{\tau}(\mathbf{x})$  (days)

Stn name	$\tau_{11}$	$\tau_{22}$	$\tau_{33}$	$\hat{\tau}$	$\hat{\tau}_{\text{SNR}>1}$
ACAP	150	19	77	82	82
ACYA	150	20	35	68	68
CAYA	40	-31	-6	1	1
COYU	103	-12	32	41	41
CPDP	162	38	87	96	96
IGUA	-82	-3	-76	-54	-54
MEZC	-79	-3	-37	-40	-40
UNIP	-166	3014	-245	868	-
YAIG	-53	-13	-759	-275	-22

classical seismology instrumentation. Further investigations are necessary in order to better understand the spectral characteristics of SSEs. I can not exclude that the third IC might instead be related to a common noise that is affecting the network. In order to address this question it would be necessary to extend the GPS network, as well as study subsequent events and see if they show a similar behaviour.

### 4.3.5 Conclusive remarks

I have studied with a static ICA algorithm (i.e., with a mixing matrix constant w.r.t. time) the daily GPS position time series relative to the 2006 Guerrero SSE. The analysis shows a peculiar behaviour, with one of the ICs assuming the value corresponding to the temporal derivative of the main IC. Moving from the assumptions that: 1) the linear combination of two components is sufficient to explain the data, and 2) only one source is the cause of the observed data, and this source is allowed to move in the space, I have sketched a conceptual framework to explain the observations. Without performing any inversion, I have proposed to interpret the reference system invariant quantity  $\hat{\tau}(\mathbf{x})$  characteristic of each station as the perceived time of activation at the station located in  $\mathbf{x}$ . The gradient of  $\hat{\tau}(\mathbf{x})$  may thus suggest the direction of perceived propagation of the SSE source effects at the surface.

The reason why the observed displacements can be explained using only two components is still unanswered. The algorithm used for the spatio-temporal characterisation of the ground displacements is not aimed at solving the BSS problem relative to moving sources. Nonetheless, it seems that the hypothesis of having only one source that is moving can provide a simpler interpretation of the observed temporal series and useful information on the

spatial propagation of the source, even in case, such as the present one, where the analysis is performed with a static ICA technique. The results presented here are still preliminary, and the interesting scenario concerning only one source, but moving in the space, needs to be further investigated in the future, starting from the application of static ICA to synthetic test time series.

# Conclusions

In this thesis I have presented the application of multivariate statistical techniques for the detection and characterisation of ground displacements recorded by GNSS stations. In particular, I have focused my attention to two linear decomposition techniques: the Principal Component Analysis (PCA) and the Independent Component Analysis (ICA). The first one allows reducing the dimensionality of the data space preserving as much as possible the explained variance. This tool is largely used for compressing data and extract their inner structure. In the geodetic field this approach reveals to be very useful, and in Chapter 1 I have shown how it can be used to get new insights into some important mechanical characteristics of the fault zone relative to a mid-size earthquake. Nonetheless, some problems persisted. Indeed, real world observations are usually the result of a combination of effects, and in order to reveal the inner structure of the data it is more useful to search for an intrinsic coordinate frame where the informative content of the data is maximised. For this reason, in Chapter 2 I have introduced the theory behind an advanced ICA technique based on variational approximation and bayesian inference (vbICA), aimed at solving the Blind Source Separation (BSS) problem. This approach extends the classical mapping approach used to perform ICA and makes it more flexible, allowing me to retrieve sources with multi-modal distributions and to have an estimation of the probability density functions (pdfs) associated to each IC. Following Chan et al. [2003], I have modified the code of Choudrey [2002] in order to take into account missing data, since usually significant gaps affect GNSS position time series. In Chapter 3 I have tested the modified algorithm on synthetic data that simulate different GNSS network configurations located around an active seismic fault and a volcanic source. In particular, I have used a combination of a linear function, a Heaviside function, and a logarithmic function in order to mimic the seismic cycle. I have also added a sinusoidal signal to reproduce seasonal effects, and white and coloured noise. The volcanic source has been reproduced using a Mogi source whose temporal evolution corresponds to a combination of arctangent functions in order to mimic an inflating and

deflating magma chamber. The results have shown that if the network is sufficiently dense (10 stations inside an area corresponding to a square of side twice the locking depth, and centered around the projection onto the Earth's surface of the mid-point of the top edge of the fault) and continuous, then the temporal evolution of the post-seismic source relative also to moderate size earthquakes can be correctly inferred. After this validation, in Chapter 4 I have applied the algorithm to real world data. For all the case studies presented the number of stations operative in continuous mode was limited ( $\leq 15$ ), but with the usage of the proposed technique it has been possible to extract useful informations about the spatial and temporal characteristics of the processes that generated the observations. In particular, the possibility to recover the actual temporal structure of the sources has also enabled me to reconstruct a more faithful mixing matrix w.r.t. the one obtained with different multivariate statistical techniques like PCA. This capability of the ICA approach allows me to visualize in a more clear way the spatial extent of the geophysical ongoing processes.

Thanks to the vbICA approach I have been able to verify the hypothesis that the afterslip deformation mechanism is the main ongoing process during the post-seismic stage relative to the 2009 L'Aquila (central Italy) earthquake. This information corroborates the results found using the PCA-based Inversion Method (PCAIM, Kositsky and Avouac [2010]) for the determination of the frictional properties of a velocity-strengthening region presented in Chapter 1. The multivariate analysis has also allowed me to detect a peculiar multi-annual signal. In order to better understand its origin it is necessary to expand the timeline of the analysis. This will be possible in the next years if a continuous GPS network remains operative in the region. I have also shown that the selection of the time spanned and of the stations used for the analysis are two crucial factors that may affect the final decomposition results. This is due to the fact that the presented multivariate techniques are fundamentally data driven, with all the deriving pros and cons.

The application to the 2012 Emilia (northern Italy) seismic sequence has revealed the presence of two post-seismic decay processes, not properly separated by a PCA technique. I have found that only one of the two decays is associated with afterslip occurred on both the structures activated during the sequence. It is likely that the vbICA has not been able to separate two afterlip processes each one occurring on a single structure because of the similarity of the pdfs of the two signals. Despite the number of GPS stations did not allow me to perform a distributed slip inversion on the fault planes, I have been able to infer from forward modelling the position of the regions that likely have been activated by afterslip. The analysis of the second rapid ( $\sim 2$  weeks) decay signal has induced me to think that it is related to a pro-

elastic deformation mechanism. Further investigations with spatially denser techniques like InSAR could give some more hints about this process. Anyway, the main problem for the detection of such a fast transient resides in the poor temporal resolution of InSAR time series.

Finally, the application to the Guerrero 2006 Slow Slip Event (SSE) (Mexico) has pushed the algorithm to its limits. Indeed, the current version of the vbICA technique is not aimed at treating the BSS problem with moving sources, while Radiguet et al. [2011] have shown that the Guerrero 2006 SSE migrated. I have presented a possible interpretation of the decomposition, suggesting that only one source is active and it is moving in the space. I have introduced the quantity defined by the tensor  $\tau(\mathbf{x})$ , whose trace is an invariant and can be interpreted as the perceived time of activation at a given point  $\mathbf{x}$  at the surface. The problem considered is a simplification of the actual BSS problem with moving sources, since the observed data are the result of a mix of only one source with itself. More investigations are needed in order to better understand the behaviour of the vbICA algorithm in such a simplified case. Some desirable routes to explore in the future should consider the possibility of having a mixing matrix that depends also on time, or the possibility of having correlation across time in the sources.

# Bibliography

- Z. Altamimi, X. Collilieux, and L. Metivier. ITRF2008: an improved solution of the International Terrestrial Reference Frame. *Journal of Geodesy*, 8(85):457–473, 2011. DOI: 10.1007/s00190-011-0444-4.
- J.G. Anderson, J.N. Brune, J. Prince, R. Quaas, S.K. Singh, D. Almora, P. Bodin, M. Oñate, R. Vasquez, and M. Velasco. The Guerrero accelerograph network. *Geofis. Int.*, 33:341–371, 1994.
- M. Anzidei, E. Boschi, V. Cannelli, R. Devoti, A. Esposito, A. Galvani, D. Melini, G. Pietrantonio, F. Riguzzi, V. Sepe, and E. Serpelloni. Co-seismic deformation of the destructive April 6, 2009 L’Aquila earthquake (central Italy) from GPS data. *Geophysical Research Letters*, 36(L17307), 2009.
- H. Attias. Independent Factor Analysis. *Neural Computation*, 11:803–851, 1999.
- S. Atzori, I. Hunstad, M. Chini, S. Salvi, C. Tolomei, C. Bignami, S. Stramondo, E. Trasatti, A. Antonioli, and E. Boschi. Finite fault inversion of DInSAR coseismic displacement of the 2009 L’Aquila earthquake (central Italy). *Geophysical Research Letters*, 36(L15305), 2009.
- S. Bailey. Principal Component Analysis with Noisy and/or Missing Data. *Publications of the Astronomical Society of the Pacific*, 124(919):1015–1023, 2012.
- S. Barbot, Y. Hamiel, and Y. Fialko. Space geodetic investigation of the coseismic and postseismic deformation due to the 2003 Mw 7.2 Altai earthquake: implications for the local lithospheric rheology. *Journal of Geophysical Research*, 113(B03403), 2008. doi:10.1029/2007JB005063.
- M. Barchi, F. Galadini, G. Lavecchia, P. Messina, A.M. Michetti, L. Peruzza, A. Pizzi, and E. Tondi. Sintesi delle conoscenze sulle faglie attive in Italia Centrale: parametrizzazione ai fini della caratterizzazione della pericolosità

- sismica. *CNR-Gruppo Nazionale per la Difesa dai Terremoti, Roma*, 62, 2000.
- M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18(4):349–369, 1989.
- A.J. Bell and T.J. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.
- R.A. Bennett, E. Serpelloni, S. Hreinsdóttir, M.T. Brandon, G. Buble, T. Basic, G. Casale, A. Cavaliere, M. Anzidei, M. Marjonovic, G. Minelli, G. Molli, and A. Montanari. Syn-convergent extension observed using the RETREAT GPS network, northern Apennines, Italy. *Journal of Geophysical Research*, 117(B4), 2012. DOI: 10.1029/2011JB008744.
- C.M. Bishop. *Pattern Recognition and Machine Learning*. New York, Springer, 2006.
- J. Boatwright and M. Cocco. Frictional constraints on crustal faulting. *Journal of Geophysical Research*, 101(B6):13895–13909, 1996.
- P. Boncio, G. Lavecchia, and B. Pace. Defining a model of 3D seismogenic sources for Seismic Hazard Assessment applications: The case of central Apennines (Italy). *Journal of Seismology*, 8:407–425, 2004.
- P. Boncio, A. Pizzi, F. Brozzetti, G. Pomposo, G. Lavecchia, D. Di Naccio, and F. Ferrarini. Coseismic ground deformation of the 6 April L’Aquila earthquake (central Italy,  $M_W$  6.3). *Geophysical Research Letters*, 37 (L06308), 2010. doi:10.1029/2010GL042807.
- M. Bottiglieri, M. Falanga, U. Tammaro, F. Obrizzo, P. De Martino, C. Godano, and F. Pingue. Independent component analysis as a tool for ground deformation analysis. *Geophys. J. Int.*, 168:1305–1310, 2007. doi:10.1111/j.1365-246X.2006.03264.x.
- M. Bottiglieri, M. Falanga, U. Tammaro, P. De Martino, F. Obrizzo, C. Godano, and F. Pingue. Characterization of GPS time series at the Neapolitan volcanic area by statistical analysis. *J. of Geophys. Res.*, 115(B10416), 2010. doi:10.1029/2009JB006594.
- H.B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. John Wilwy and Sons Inc., New York, 2nd edition, 1985.

- J.F. Cardoso. High-order contrasts for Independent Component Analysis. *Neural Computation*, 11:157–192, 1999.
- O. Cavalié, E. Pathier, M. Radgiet, M. Vergnolle, N. Cotte, A. Walpersdorf, V. Kostoglodov, and F. Cotton. Slow slip event in the Mexican subduction zone: Evidence of shallower slip in the Guerrero seismic gap for the 2006 event revealed by the joint inversion of InSAR and GPS data. *Earth and Planetary Science Letters*, 367:52–60, 2013.
- S. Cesca, T. Braun, F. Maccaferri, L. Passarelli, E. Rivalta, and T. Dahm. Source modelling of the M5-6 Emilia-Romagna, Italy, earthquakes (2012 May 20-29). *Geophysical Journal International*, 2013. doi:10.1093/gji/ggt069.
- K. Chan, T.-W. Lee, and T.J. Sejnowski. Variational Bayesian Learning of ICA with Missing Data. *Neural Computation*, 15(8):1991–2011, 2003.
- D. Cheloni, N. DAgostino, E. DAnastasio, A. Avallone, S. Mantenuto, R. Giuliani, M. Mattone, S. Calcaterra, P. Gambino, D. Dominici, F. Radicioni, and G. Fastellini. Coseismic and initial post-seismic slip of the 2009 Mw 6.3 LAquila earthquake, Italy, from GPS measurements. *Geophysical Journal International*, 181:1539–1546, 2010.
- C. Chiarabba, A. Amato, M. Anselmi, P. Baccheschi, I. Bianchi, M. Cattaneo, G. Cecere, L. Chiaraluce, M.G. Ciaccio, P. De Gori, G. De Luca, M. Di Bona, R. Di Stefano, L. Faenza, A. Govoni, L. Improta, F.P. Lucente, A. Marchetti, L. Margheriti, F. Mele, A. Michelini, G. Monachesi, M. Moretti, M. Pastori, N. Piana Agostinetti, D. Piccinini, P. Roselli, D. Seccia, and L. Valoroso. The 2009 L’Aquila (central Italy) MW6.3 earthquake: Main shock and aftershocks. *Geophysical Research Letters*, 36, 2009. doi:10.1029/2009GL039627.
- C. Chiarabba, P. De Gori, L. Improta, F.P. Lucente, M. Moretti, A. Govoni, M. Di Bona, L. Margheriti, A. Marchetti, and A. Nardi. Frontal compression along the Apennines thrust system: The Emilia 2012 example from seismicity to crustal structure. *Journal of Geodynamics*, 82:98–109, 2014. doi:10.1016/j.jog.2014.09.003.
- L. Chiaraluce, L. Valoroso, D. Piccinini, R. Di Stefano, and P. De Gori. The Anatomy of the 2009 L’Aquila Normal Fault System [central Italy] Imaged by High Resolution Foreshock and Aftershock Locations. *Journal of Geophysical Research*, 2011.

- R.A. Choudrey. *Variational Methods for Bayesian Independent Component Analysis*. Pattern analysis and machine learning - robotics research group, University of Oxford, 2002.
- P. Comon. Independent component analysis, A new concept? *Signal Processing*, 36:287–314, 1994.
- F. Correa-Mora, C. DeMets, E. Cabral-Cano, O. Diaz-Molina, and B. Marquez-Azua. Transient deformation in southern Mexico in 2006 and 2007: Evidence for distinct deep-slip patches beneath Guerrero and Oaxaca. *Geochemistry Geophysics Geosystems*, 10(2), 2009. doi:10.1029/2008GC002211.
- T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.
- N. D’Agostino, S. Mantenuto, E. D’Anastasio, R. Giuliani, M. Mattone, S. Calcaterra, P. Gambino, and L. Bonci. Evidence for localized active extension in the central Apennines (Italy) from global positioning system observations. *Geology*, 39(4):291–294, 2011. doi: 10.1130/G31796.1.
- N. D’Agostino, D. Cheloni, G. Fornaro, R. Giuliani, and D. Reale. Space-time distribution of afterslip following the 2009 LAquila earthquake. *Journal of Geophysical Research*, 117(B02402), 2012. doi:10.1029/2011JB008523.
- C. DeMets, R.G. Gordon, D.F. Argus, and S. Stein. Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions. *Geophys. Res. Lett.*, 21:2191–2194, 1994. doi:10.1029/94GL02118.
- R. Devoti, A. Esposito, G. Pietrantonio, A.R. Pisani, and F. Riguzzi. Evidence of large scale deformation patterns from GPS data in the Italian subduction boundary. *Earth and Planetary Science Letters*, 311(3-4):230–241, 2011. doi:10.1016/j.epsl.2011.09.034.
- R. Devoti, L. Anderlini, M. Anzidei, A. Esposito, A. Galvani, G. Pietrantonio, A. Pisani, F. Riguzzi, V. Sepe, and E. Serpelloni. The coseismic and postseismic deformation of the L’Aquila, 2009 earthquake from repeated GPS measurements. *Ital. J. Geosci.*, 131(3):348–358, 2012. doi: 10.3301/IJG.2012.15.
- J.H. Dieterich. Modeling of fault friction: 1. Experimental results and constitutive equations. *Journal of Geophysical Research*, 84(B5):2161–2168, 1979.

- D. Dong, P. Fang, Y. Bock, M.K. Cheng, and S. Miyazaki. Anatomy of apparent seasonal variations from GPS-derived site position time series. *Journal of Geophysical Research Solid Earth*, 107, 2002. doi:10.1029/2001JB000573.
- D. Dong, P. Fang, Y. Bock, F. Webb, L. Prawirodirdjo, S. Kedar, and P. Jamason. Spatiotemporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis. *Journal of Geophysical Research Solid Earth*, B3(111), 2006. DOI: 10.2029/2005JB003806.
- Working Group Emergeo. Liquefaction phenomena associated with the Emilia earthquake sequence of May-June 2012 (Northern Italy). *Nat. Hazards Earth Syst. Sci.*, 13(4):935–947, 2013. DOI: 10.5194/nhess-13-935-2013.
- J.P. Faure-Walker, G.P. Roberts, P.R. Sammonds, and P. Cowie. Comparison of earthquake strains over  $10^2$  and  $10^4$  year timescales: Insights into variability in the seismic cycle in the central Apennines, Italy. *Journal of Geophysical Research*, 115(B10418), 2010. doi:10.1029/2009JB006462.
- E.J. Fielding, P.R. Lundgren, R. Bürgmann, and G.J. Funning. Shallow fault-zone dilatancy recovery after the 2003 Bam earthquake in Iran. *Nature*, 458, 2009. doi:10.1038/nature07817.
- E. Forootan and J. Kusche. Separation of global time-variable gravity signals into maximally independent components. *Journal of Geodesy*, 86(7):477–497, 2012.
- S.I. Franco, V. Kostoglodov, K.M. Larson, V.C. Manea, M. Manea, and J.A. Santiago. Propagation of the 20012002 silent earthquake and intraplate coupling in the Oaxaca subduction zone. *Earth Planets Space*, 57:973–985, 2005.
- J. Fukuda, K.M. Johnson, K.M. Larson, and S. Miyazaki. Fault friction parameters inferred from the early stages of afterslip following the 2003 Tokachi-oki earthquake. *Journal of Geophysical Research*, 114(B4), 2009. doi:10.1029/2008JB006166.
- F. Galadini and P. Galli. Active tectonics in the central Apennines (Italy) - Input data for seismic hazard assessment. *Natural Hazards*, 22:225–270, 2000.

- P. Galli, F. Galadini, and D. Pantosti. Twenty years of paleoseismology in Italy. *Earth Sci. Rev.*, 88(1-2):89–117, 2008. doi:10.1016/j.earscirev.2008.01.001.
- A. Galvani, M. Anzidei, R. Devoti, A. Esposito, G. Pietrantonio, A.R. Pisani, F. Riguzzi, and E. Serpelloni. The interseismic velocity field of the central Apennines from a dense GPS network. *Annals of Geophysics*, 55(5), 2012. DOI: 10.4401/ag-5634.
- G. Grenerczy, G. Sella, S. Stein, and A. Kenyeres. Tectonic implications of the GPS velocity field in the northern Adriatic region. *Geophys. Res. Lett.*, 32(L16311), 2005. i:10.1029/2005GL022947.
- A. Gualandi, E. Serpelloni, and M.E. Belardinelli. Space-time evolution of crustal deformation related to the  $M_W$  6.3, 2009 L’Aquila earthquake (central Italy) from principal component analysis inversion of GPS position time-series. *Geophys. J. Int.*, 197(1):174–191, 2014. doi:10.1093/gji/ggt522.
- T.C. Hanks and H. Kanamori. A Moment Magnitude Scale. *Journal of Geophysical Research*, 84(B5), 1979.
- J. Herault and C. Jutten. Space or time adaptive signal processing by neural models. In *Proceedings AIP Conference: Neural Networks for Computing*, volume 151, pages 206–211. J.S. Denker, Ed. American Institute for Physics, 1986.
- R.B. Herrmann, L. Malagnini, and I. Munafó. Regional moment tensors of the 2009 LAquila earthquake sequence. *Bull. seism. Soc. Am.*, 101:975–993, 2011.
- A. Hooper, D. Bekaert, K. Spaans, and M. Arikan. Recent advances in SAR interferometry time series analysis for measuring crustal deformation. *Tectonophysics*, 514-517:1–13, 2012.
- Y.J. Hsu et al. Friction afterslip following the 2005 Nias-Simeulue earthquake, Sumatra. *Science*, 312:1921–1926, 2006.
- A. Hyvärinen and E. Oja. A fast fixed-point algorithm for Independent Component Analysis. *Neural Computation*, 9(7):1483–1492, 1997.
- A. Hyvärinen and E. Oja. Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4-5):411–430, 2000.

- A. Iglesias, S.K. Singh, A.R. Lowry, M. Santoyo, V. Kostoglodov, K.M. Larson, and S.I. Franco-Sánchez. The silent earthquake of 2002 in the Guerrero seismic gap, Mexico (Mw=7.6): Inversion of slip on the plate interface and some implications. *Geofísica Internacional*, 43(3):309–317, 2004.
- Y. Ito, R. Hino, M. Kido, H. Fujimoto, Y. Osada, D. Inazu, Y. Ohta, T. Iinuma, M. Ohzono, S. Miura, M. Mishina, K. Suzuki, T. Tsuji, and J. Ashi. Episodic slow slip events in the Japan subduction zone before the 2011 Tohoku-Oki earthquake. *Tectonophysics*, 600:14–26, 2013. doi:10.1016/j.tecto.2012.08.022.
- K.H. Ji and T.A. Herring. A method for detecting transient signals in GPS position time-series: smoothing and principal component analysis. *Geophysical Journal International*, 193(1):171–186, 2013. doi:10.1093/gji/ggt003.
- K.J. Johnson, R. Bürgmann, and K. Larson. Frictional properties on the San Andreas fault near Parkfield, California, inferred from models of afterslip following the 2004 earthquake. *Bulletin of Seismological Society of America*, 96(4B):S321–S338, 2006.
- I.T. Jolliffe. *Principal Component Analysis*. Springer, 2nd edition, 2002.
- S. Jónsson, P. Segall, R. Pedersen, and G. Björnsson. Post-earthquake ground movements correlated to pore-pressure transients. *Nature*, 424:179–183, 2003.
- M.I. Jordan. Graphical Models. *Statistical Science*, 19:140–155, 2004.
- M.I. Jordan, Z. Ghahramani, T.S. Jaakkola, and L.K. Saul. An Introduction to Variational Methods for Graphical Models. *Machine Learning*, 37:183–233, 1999.
- H. Kanamori and E. Brodsky. The physics of earthquakes. *Rep. Prog. Phys.*, 67(1429), 2004. doi:10.1088/0034-4885/67/8/R03.
- A.P. Kositsky. *PCAIM User’s Manual*. California Institute of Technology, 2011.
- A.P. Kositsky and J.P. Avouac. Inverting geodetic time series with a principal component analysis-based inversion method. *Journal of Geophysical Research*, 115(B03401), 2010. doi:10.1029/2009JB006535.

- V. Kostoglodov, S.K. Singh, J.A. Santiago, S.I. Franco, K. Larson, A.R. Lowry, and R. Bilham. A large silent earthquake in the Guerrero seismic gap, Mexico. *Journal of Geophysical Research*, 30(15, 1807), 2003. doi:10.1029/2003GL017219.
- S. Kullback and R.A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951. doi:10.1214/aoms/1177729694.
- R. Lanari, P. Berardino, M. Bonano, F. Casu, A. Manconi, M. Manunta, M. Manzo, A. Pepe, S. Pepe, E. Sansosti, G. Solaro, P. Tizzani, and G. Zeni. Surface displacements associated with the L’Aquila 2009  $M_W$  6.3 earthquake (central Italy): New evidence from SBAS-DInSAR time series analysis. *Geophysical Research Letters*, 37(L20309), 2010. doi:10.1029/2010GL044780.
- J. Langbein. Noise in GPS displacement measurements from Southern California and Souther Nevada. *Journal of Geophysical Research*, 113(B05405), 2008. doi:10.1029/2007JB005247.
- K.M. Larson, A.R. Lowry, V. Kostoglodov, W. Hutton, O. Sánchez, K. Hudnet, and G. Suárez. Crustal deformation measurements in Guerrero, Mexico. *Journal of Geophysical Research*, 109(B04409), 2004. doi:10.1029/2003JB002843.
- K.M. Larson, V. Kostoglodov, S. Miyazaki, and J.A. Santiago. The 2006 aseismic slow slip event in Guerrero, Mexico: New results from GPS. *Geophysical Research Letters*, 34(L13309), 2007. doi:10.1029/2007GL029912.
- R.B. Lohman and J.R. Murray. The SCEC Geodetic Transient-Detection Validation Exercise. *Seismological Research Letters*, 84(3):419–425, 2013. doi: 10.1785/0220130041.
- A.R. Lowry, K.M. Larson, V. Kostoglodov, and R. Bilham. Transient fault slip in Guerrero, southern Mexico. *Geophysical Research Letters*, 28(19): 3753–3756, 2001.
- D.J.C. MacKay. Developments in probabilistic modelling with neural networks - ensemble learning. In *Proceedings of the third Annual Symposium on Neural Networks*, pages 191–198, Nijmegen, The Netherlands, 1995. Springer.
- L. Malagnini, R.B. Hermann, I. Munafó, M. Buttinelli, M. Anselmi, A. Akinci, and E. Boschi. The 2012 Ferrara seismic sequence: Regional crustal

- structure, earthquake sources, and seismic hazard. *Geophysical Research Letters*, 39(L19302), 2012. doi:10.1029/2012GL053214.
- C.J. Marone, C.H. Scholz, and R. Bilham. On the Mechanics of Earthquake Afterslip. *Journal of Geophysical Research*, 96(B5):8441–8452, 1991.
- B. Márquez-Azúa and C. DeMets. Crustal velocity field of Mexico from continuous GPS measurements, 1993 to June 2001: implications for the neotectonics of Mexico. *J. Geophys. Res.*, 108(B9), 2003. doi:10.1029/2002JB002241.
- M. Nespoli, M. Todesco, E. Serpelloni, M.E. Belardinelli, M. Bonafede, M. Marcaccio, A.P. Rinaldi, L. Anderlini, and A. Gualandi. Modelling earthquake effects of groundwater levels: evidences from the 2012 Emilia earthquake (Italy). in prep.
- Y. Okada. Surface deformation due to shear and tensile faults in halfspace. *Bulletin of the Seismological Society of America*, 1985.
- M. Ortiz, S.K. Singh, V. Kostoglodov, and J. Pacheco. Constraint on the rupture areas of the Acapulco-San Marcos, Mexico earthquakes of the 1962 (M 7.1; 7.0) and 1957 (M 7.7), based on the analysis of tide records. *Geofis. Int.*, 39:337–348, 2000.
- J.T. Ormerod and M.P. Wand. Explaining Variational Approximations. *The American Statistician*, 64(2), 2010. DOI:10.1198/tast.2010.09058.
- A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 3rd edition, 1991.
- G. Parisi. *Statistical Field Theory*. Addison-Wesley, Redwood City, CA, 1988.
- G. Peltzer, P. Rosen and F. Rogez, and K. Hudnut. Postseismic rebound in fault step-overs caused by pore fluid-flow. *Science*, 273:1202–1204, 1996.
- Z. Peng and J. Gomberg. An integrated perspective of the continuum between earthquakes and slow-slip phenomena. *Nature Geoscience*, 3:599–607, 2010. doi:10.1038/ngeo940.
- N.T. Penna and M.P. Stewart. Aliased tidal signatures in continuous GPS height time series. *Geophys. Res. Lett.*, 30(23), 2003. doi:10.1029/2003GL018828.

- H. Perfettini and J.P. Ampuero. Dynamics of a velocity strengthening region: implications for slow earthquakes and postseismic slip. *Journal of Geophysical Research*, 113(B9), 2008. doi:10.1029/2007JB005398.
- H. Perfettini, J.P. Avouac, H. Tavera, A. Kositsky, J.M. Nocquet, F. Bondoux, M. Chlieh, A. Sladen, L. Audin, D.L. Farber, and P. Soler. Seismic and aseismic slip on the Central Peru megathrust. *Nature*, 465(6 May 2010), 2010. doi:10.1038/nature09062.
- G. Pezzo, J.P.M. Boncori, C. Tolomei, S. Salvi, S. Atzori, A. Antonioli, E. Trasatti, F. Novali, E. Serpelloni, L. Candela, and R. Giuliani. Coseismic Deformation and Source Modeling of the May 2012 Emilia (Northern Italy) Earthquakes. *Seismological Research Letters*, 84(4), 2013. doi:10.1785/0220120171.
- V. Picotti and F.J. Pazzaglia. A new active tectonic model for the construction of the Northern Apennines mountain front near Bologna (Italy). *Journal of Geophysical Research*, 113(B08412), 2008. doi:10.1029/2007JB005307.
- S. Pondrelli, S. Salimbeni, A. Morelli, G. Ekstrom, M. Olivieri, and E. Boschi. Seismic moment tensors of the April 2009, L'Aquila (Central Italy), earthquake sequence. *Geophysical Journal International*, 180:238–242, 2010.
- S. Pondrelli, S. Salimbeni, Paolo Perfetti, and Peter Danecek. Quick regional centroid moment tensor solutions for the Emilia 2012 (northern Italy) seismic sequence. *Annals of Geophysics*, 55(4), 2012.
- R. Priemer. *Introductory Signal Processing*. World Scientific, 1990. p. 1, doi:10.1142/9789814434409\_0001.
- M. Radiguet, F. Cotton, M. Vergnolle, M. Campillo, B. Valette, V. Kostoglodov, and N. Cotte. Spatial and temporal evolution of a long term slow slip event: the 2006 Guerrero Slow Slip Event. *Geophysical Journal International*, 184:816–828, 2011. doi:10.1111/j.1365-246X.2010.04866.x.
- J.R. Rice and J.C. Gu. Earthquake aftereffects and triggered seismic phenomena. *Pure Appl. Geophys.*, 121(2):187–219, 1983.
- B. Riel, M. Simons, P. Agram, and Z. Zhan. Detecting transient signals in geodetic time series using sparse estimation techniques. *Journal of Geophysical Research*, pages 5140–5160, 2014. doi:10.1002/2014JB011077.

- S.J. Roberts and R. Everson. *Independent Component Analysis: Principles and Practice*. Cambridge University Press, 2001.
- A.L. Ruina. Slip instability and state variable friction laws. *Journal of Geophysical Research*, 88:10359–10370, 1983.
- S. Ruiz, M. Metois, A. Fuenzalida, J. Ruiz, F. Leyton, R. Grandin, C. Vigny, R. Madariaga, and J. Campos. Intense foreshocks and a slow slip event preceded the 2014 Iquique  $M_W$  8.1 earthquake. *Science*, 345(6201):1165–1169, 2014. DOI: 10.1126/science.1256074.
- S.Y. Schwartz and J.M. Rokesky. Slow slip events and seismic tremor at circum-Pacific subduction zones. *Rev. Geophys.*, 45(RG3004), 2007. doi:10.1029/2006RG00020.
- L. Scognamiglio, E. Tinti, A. Michelini, D.S. Dreger, A. Cirella, M. Cocco, S. Mazza, and A. Piatanesi. Fast Determination of Moment Tensors and Rupture History: What Has Been Learned from the 6 April 2009 L’Aquila Earthquake Sequence. *Seismological Research Letters*, 81(6):892–906, 2010. doi: 10.1785/gssrl.81.6.892.
- L. Scognamiglio, L. Margheriti, F.M. Mele, E. Tinti, A. Bono, P. De Gori, V. Lauciani, F.P. Lucente, A.G. Mandiello, C. Marcocci, S. Mazza, S. Pintore, and M. Quintiliani. The 2012 Pianura Padana Emiliana seismic sequence: locations, moment tensors and magnitudes. *Annals of Geophysics*, 4(55), 2012. doi: 10.4401/ag-6159.
- M.M. Scuderi, A.R. Niemeijer, C. Collettini, and C. Marone. Frictional properties and slip stability of active faults within carbonate-evaporite sequences. The role of dolomite and anhydrite. *Earth Planet. Sci. Lett.*, 369-370:220–232, 2013.
- Paul Segall. *Earthquake and Volcano Deformation*. Princeton University Press, 2010.
- E. Serpelloni, G. Casula, A. Galvani, M. Anzidei, and P. Baldi. Data analysis of permanent GPS networks in Italy and surrounding regions: application of a distributed processing approach. *Annals of Geophysics*, 49(4/5), 2006.
- E. Serpelloni, L. Anderlini, A. Avallone, V. Cannelli, A. Cavaliere, D. Cheloni, C. D’Ambrosio, E. D’Anastasio, A. Esposito, G. Pietrantonio, A.R. Pisani, M. Anzidei, G. Cecere, N. D’Agostino, S. Del Mese, R. Devoti, A. Galvani, A. Massucci, D. Melini, F. Riguzzi, G. Selvaggi, and V. Sepe. GPS observations of coseismic deformation following the May 20 and 29,

- 2012, Emilia seismic events (northern Italy): data, analysis and preliminary models. *Annals of Geophysics*, 55(4), 2012. doi: 10.4401/ag-6168.
- E. Serpelloni, C. Faccenna, G. Spada, D. Dong, and S.D.P. Williams. Vertical GPS ground motion rates in the EuroMediterranean region: New evidence of velocity gradients at different spatial scales along the NubiaEurasia plate boundary. *Journal of Geophysical Research: Solid Earth*, 118(11):6003–6024, 2013. DOI: 10.1002/2013JB010102.
- C.E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.
- F. Silverii, N. D’Agostino, and M. Metois. Postseismic Viscoelastic Relaxation Following the L’Aquila 2009 Earthquake: Implications for Lithospheric Rheology of the Apennines. AGU Fall Meeting, 2014.
- N. Srebro and T. Jaakkola. Weighted low-rank approximations. *Twentieth International Conference on Machine Learning*, 2003. <http://ttic.uchicago.edu/nati/Publications/SrebroJaakkolaICML03.pdf>.
- A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.
- D.M. Titterton. Bayesian Methods for Neural Networks and Related Models. *Statistical Science*, 19:128–139, 2004.
- S. Toda, R. Stein, V. Sevilgen, and J. Lin. Coulomb 3.3 Graphic-rich deformation and stress-change software for earthquake, tectonic, and volcano research and teaching-user guide. *U.S. Geological Survey Open-File Report*, 1060:63, 2011. Available at: <http://earthquake.usgs.gov/research/modeling/coulomb/download/Coulomb31UserGuide.pdf> (last accessed 1 November 2013).
- G. Toscani, P. Burrato, D. Di Bucci, S. Seno, and G. Valensise. Plio-Quaternary tectonic evolution of the Northern Apennines thrust fronts (Bologna-Ferrara section, Italy): seismotectonic implications. *Italian Journal of Geosciences*, 128(2):605–613, 2009. doi:10.3301/IJG.2009.128.2.605.
- S. Unkel, N.T. Trendafilov, A. Hannachi, and I.T. Joliffe. Independent exploratory factor analysis with application to atmospheric science data. *Journal of Applied Statistics*, 37(11):1847–1862, 2010.

- S. Unkel, A. Hannachi, N.T. Trendafilov, and I.T. Joliffe. Independent Component Analysis for Three-Way Data With an Application From Atmospheric Science. *Journal of Agricultural, Biological, and Environmental Statistics*, 16(3):319–338, 2011.
- G. Valensise and D. Pantosti. The investigation of potential earthquake sources in peninsular Italy: A review. *Journal of Seismology*, 5:287–306, 2001.
- L. Valoroso, L. Chiaraluce, D. Piccinini, R. Di Stefano, D. Schaff, and F. Waldhauser. Radiography of a normal fault system by 64,000 high-precision earthquake locations: the 2009 LAquila (central Italy) case study. *Journal of Geophysical Research*, 118, 2013. doi:10.1002/jgrb.50130.
- M. Vergnolle, A. Walpersdorf, V. Kostoglodov, P. Tregoning, J.A. Santiago, N. Cotte, and S.I. Franco. Slow slip events in Mexico revised from the processing of 11 year GPS observations. *Journal of Geophysical Research*, 115(B08403), 2010. doi:10.1029/2009JB006852.
- A. Walpersdorf, N. Cotte, V. Kostoglodov, M. Vergnolle, M. Radiguet, J.A. Santiago, and M. Campillo. Two successive slow slip events in 2009-2010 by a dense GPS network in Guerrero, Mexico. *Geophysical Research Letters*, 38(L15307), 2011. doi:10.1029/2011GL048124.
- R.J. Walters, J.R. Elliott, N. D’Agostino, P.C. England, I. Hunstad, J.A. Jackson, B. Parsons, R.J. Phillips, and G. Roberts. The 2009 L’Aquila earthquake (central Italy): A source mechanism and implications for seismic hazard. *Geophysical Research Letters*, 36(L17312), 2009.
- R. Wang, F. Lorenzo Martín, and F. Roth. Computation of deformation induced by earthquakes in a multi-layered elastic crust-FORTRAN programs EDGRN/EDCMP. *Computers and Geosciences*, 29(2):195–207, 2013.
- S. Wdowinski, Y. Bock, J. Zhang, and J. Genrich. Southern California permanent GPS geodetic array: Spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 Landers earthquake. *Journal of Geophysical Research*, 102:18057–18070, 1997.
- R.L. Wesson. Letter to Fournier d’Albe. *USGS*, 1980.
- S.D.P. Williams. Offsets in global positioning system time series. *Journal of Geophysical Research*, 108(B6), 2003. doi:10.1029/2002JB002156.

# Appendix A

## Credibility intervals for the post-seismic decay constant

Let us consider a data vector  $\mathbf{d} \in \mathbb{R}^D$ , and let us suppose that we want to explain the data using a parameter vector  $\mathbf{m} \in \mathbb{R}^M$ . Let us also suppose that we know the relationship  $g : \mathbb{R}^M \mapsto \mathbb{R}^D$  between the model and the data vectors:  $\mathbf{d} = g(\mathbf{m})$ . The misfit function is given by the quantity  $S(\mathbf{m}, \mathbf{d}) = \mathbf{d} - g(\mathbf{m})$  and, given the data  $\mathbf{d}$ , depends on the parameter model vector  $\mathbf{m}$ . For brevity I will consider the observed data as a given vector  $\mathbf{d}^{obs}$ , and the misfit function depending only on the choice of the parameters vector. In general,  $g$  can be a non-linear function, and it is not guaranteed that  $S$  is convex everywhere in the model space. In other words, it is not guaranteed that only one minimum exists for the misfit function.

A bayesian approach to the fitting problem consists in assigning to the data  $\mathbf{d}$ , the parameters  $\mathbf{m}$ , and the model function  $g$  some *a priori* probability density function (pdf), and then solve for the inverse problem to find the *a posteriori* pdf of the parameters. In the particular fitting problem treated in this thesis, the observed data vector  $\mathbf{d}^{obs}$  corresponds to the IC related to the post-seismic decay, the model function  $g$  is given by equation 3.16, and the parameters vector  $\mathbf{m}$  is given by a bias  $m_1$ , an amplitude  $m_2$ , a time decay  $m_3$ , and time  $t$ . I consider only times  $t > t_{eq}$ , i.e. epochs after the occurrence of a mainshock. The time is a known parameter, so it can be neglected in the count of the model space dimensionality. I assume that modelization uncertainties are negligible compared to observational uncertainties. Since the data and model spaces are linear, the solution to the inverse problem can be written as (equation 1.93 of Tarantola [2005]):

$$\sigma_M(\mathbf{m}) = \frac{1}{\nu} \rho_M(\mathbf{m}) \rho_D(g(\mathbf{m})) \quad (\text{A.1})$$

where  $\sigma_M$  is the *a posteriori* probability density of the parameters,  $\rho_M$  is the *prior* probability density in the model space,  $\rho_D$  is the probability density describing the result of the measurement, and  $\nu = \int_{\mathbb{R}^M} \rho_M(\mathbf{m})\rho_D(g(\mathbf{m}))d\mathbf{m}$  is a normalization constant. I will refer to  $\rho_D(g(\mathbf{m}))$  also as the likelihood function  $L(\mathbf{m})$ , which gives a measure of how good a model  $\mathbf{m}$  is in explaining the data.

In the case that no *a priori* information is available about the parameters,  $\rho_M(\mathbf{m})$  can be replaced by its homogeneous limit  $\mu_M(\mathbf{m})$ , and I can rewrite equation A.1 as:

$$\sigma_M(\mathbf{m}) = kL(\mathbf{m}) \quad (\text{A.2})$$

Knowing the *a posteriori* pdf of the parameters  $\mathbf{m}$ , I can compute the  $x\%$  credibility volume  $\mathcal{V}_x$  in the  $M$ -dimensional parameter space, that is determined as that volume  $\mathcal{V}$  such that the probability  $P$  to find a parameter vector in it is  $x\%$ , i.e.  $\mathcal{V}_x = \mathcal{V} \subseteq \mathbb{R}^M$  such that  $P(\mathbf{m} \in \mathcal{V}) = x\%$ . From the definition of probability, I can write:

$$x\% = P(\mathbf{m} \in \mathcal{V}) = \frac{\int_{\mathcal{V}} \sigma_M(\mathbf{m})d\mathbf{m}}{\int_{\mathbb{R}^M} \sigma_M(\mathbf{m})d\mathbf{m}} = \frac{\int_{\mathcal{V}} L(\mathbf{m})d\mathbf{m}}{\int_{\mathbb{R}^M} L(\mathbf{m})d\mathbf{m}} \quad (\text{A.3})$$

Obviously, if  $\mathcal{V} = \mathbb{R}^M$ , then the probability to find  $\mathbf{m}$  in  $\mathcal{V}$  is 100%.

Since the problem is not linear, it is not possible to use a direct formula to solve it and I have to sample the model space. Fortunately, for all the cases treated in this thesis the model space dimensionality is low ( $M = 3$ ), and I can sample the posterior distribution using a grid search approach.

It remains to determine the *a priori* pdf for the data. This choice is case dependent. In the cases treated in this thesis, the data consist in the temporal sources obtained from an ICA. A great advantage of the vbICA technique is that it finds an approximation of the sources using a mix of Gaussian distributions, and thus it is possible to calculate the moments of such a distribution. In particular, for each point of a given IC, i.e. for each element of the temporal sources, I can compute (at least) the variance. I assume that all the  $T$  elements that compose an IC are independent and identically distributed (iid)<sup>1</sup>. In particular, I assume that each of them is normally distributed, and the variance is calculated directly from the mix of Gaussians.

---

<sup>1</sup>Practically, I am not assuming any temporal dependency between the value of the IC at time  $t$  and the value at time  $t + \delta t$ . An attempt to consider this dependency has been done by Choudrey [2002]. He developed a Dynamic ICA, based on Hidden Markov Models (HMMvbICA). It might be a good idea for the future to follow the same approach also for the analysis of geodetic data.

# Appendix B

## vbICA and FastICA

The FastICA algorithm [Hyvärinen and Oja, 2000] is one of the most popular algorithms used in order to perform an ICA. Figure B.1 shows the sources detected by applying FastICA on the synthetic data generated using a Mogi source, a seasonal signal, and a co- and post-seismic signal. At a first glance it is already clear the better performance of the vbICA algorithm (see Figure 3.16b). Calculating the MSE values, the vbICA performance is more than 325% better than the one of FastICA.

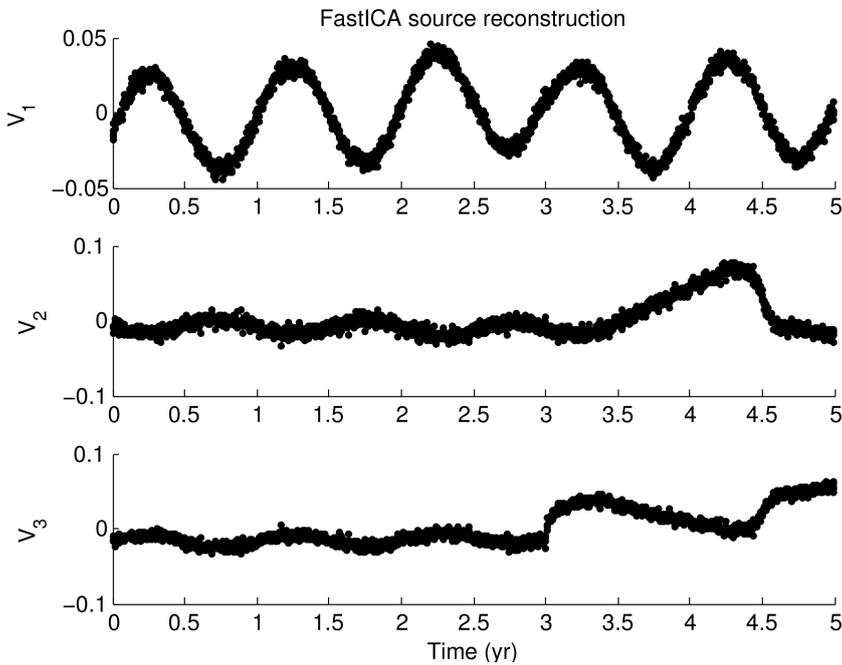


Figure B.1: Sources from the FastICA algorithm

# Appendix C

## Time series

### C.1 Aquila

Here I show the time series of the 9 cGPS stations used for the analysis of the L'Aquila 2009 earthquake (central Italy) (see Section 4.1). In all figures the black dots represent the data, and the grey shadow is the  $1\sigma$  uncertainty on the data. The red dots correspond to the vbICA reconstruction of the data, with its uncertainty shown in light red.

#### C.1.1 Seasonal signals + co-seismic offsets + post-seismic decay

Results on the time series after the removal of a long-term linear trend (Figures C.1, C.2).

#### C.1.2 Co-seismic offsets + post-seismic decay

Results on the time series after the removal of a long-term linear trend and two seasonal signals with annual and semi-annual period (Figures C.3, C.4).

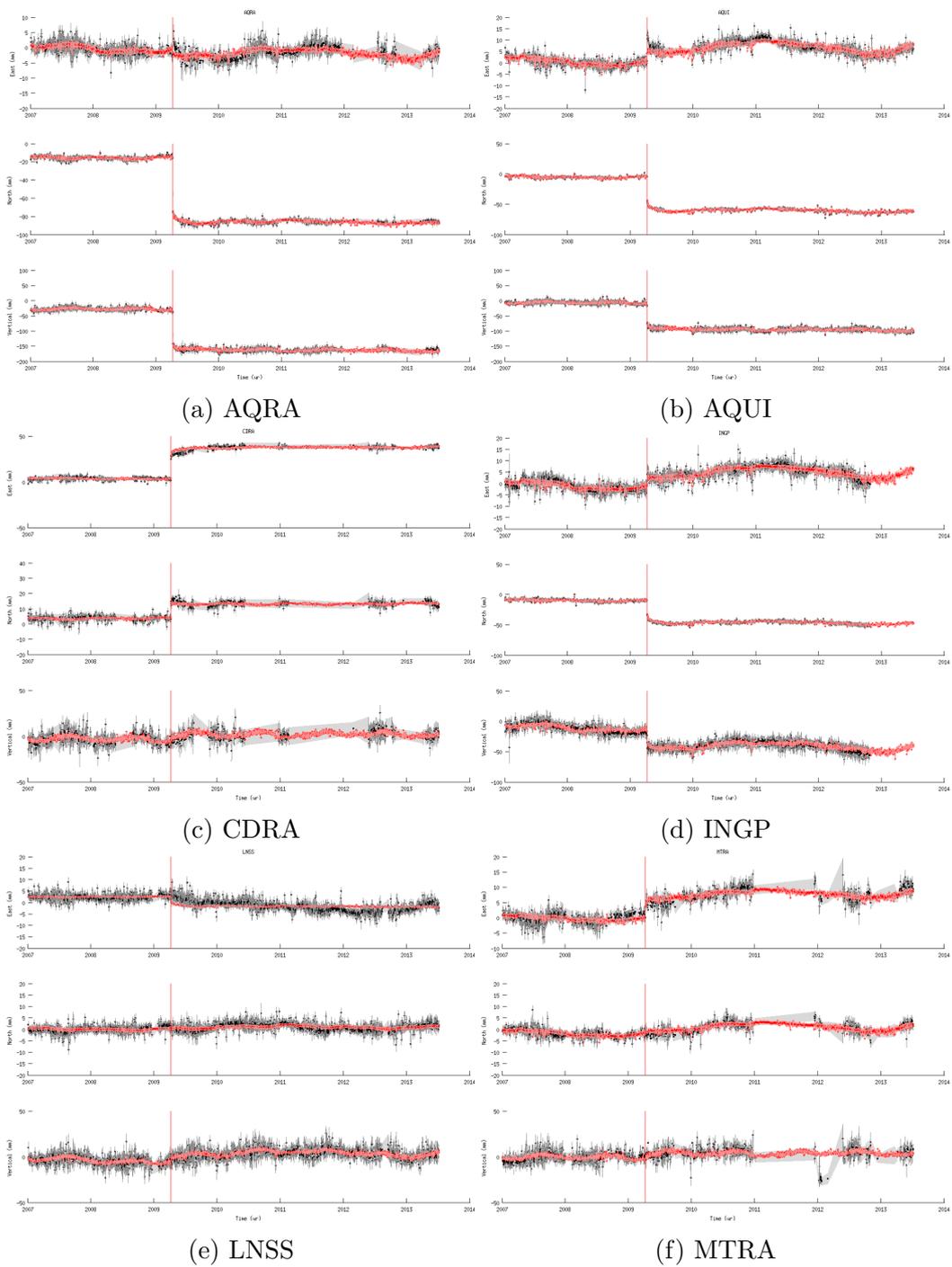


Figure C.1: L'Aquila GPS time series (black dots) and the corresponding vbICA reconstruction (red dots).

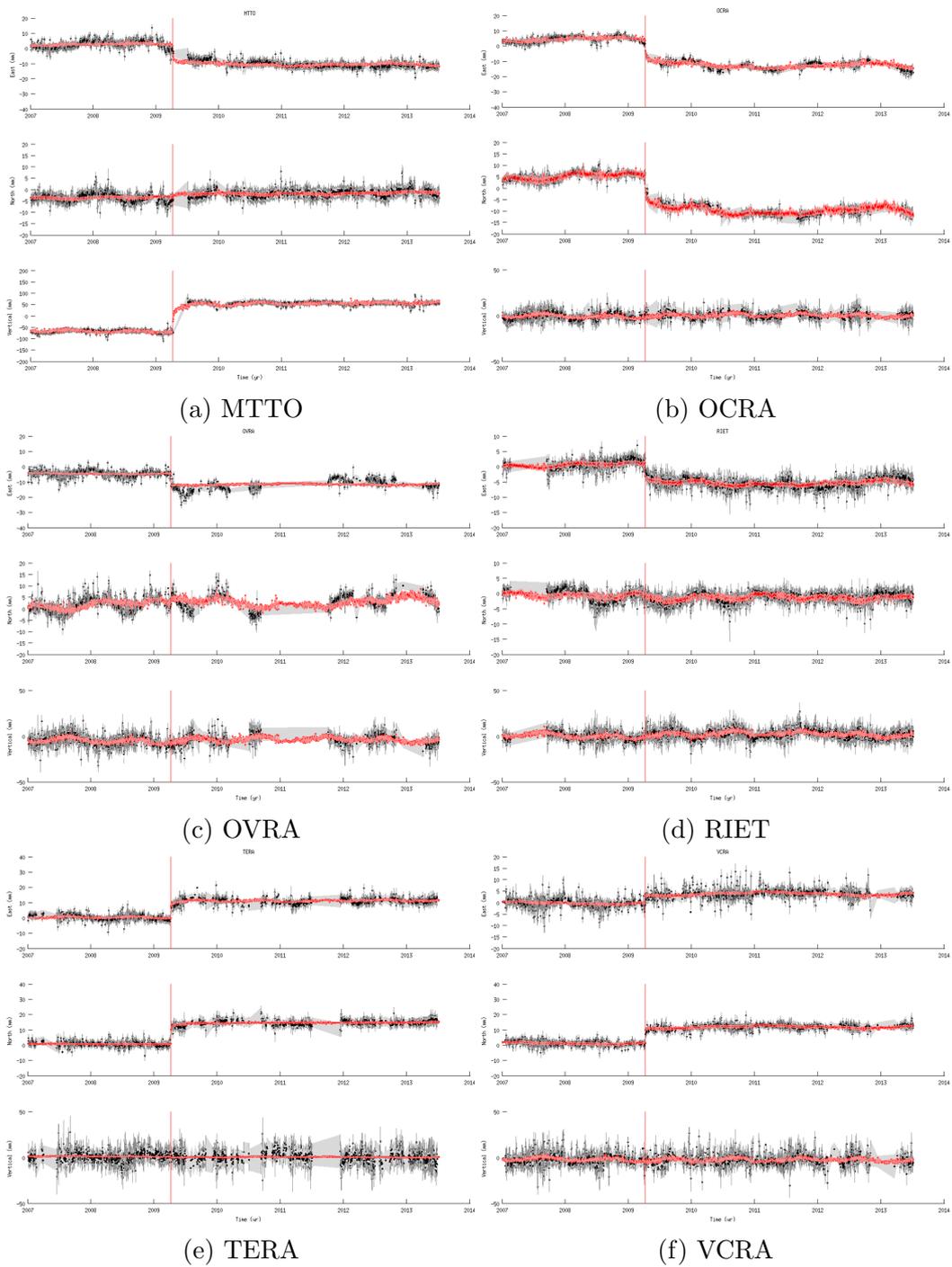


Figure C.2: L'Aquila GPS time series (black dots) and the corresponding vbICA reconstruction (red dots).

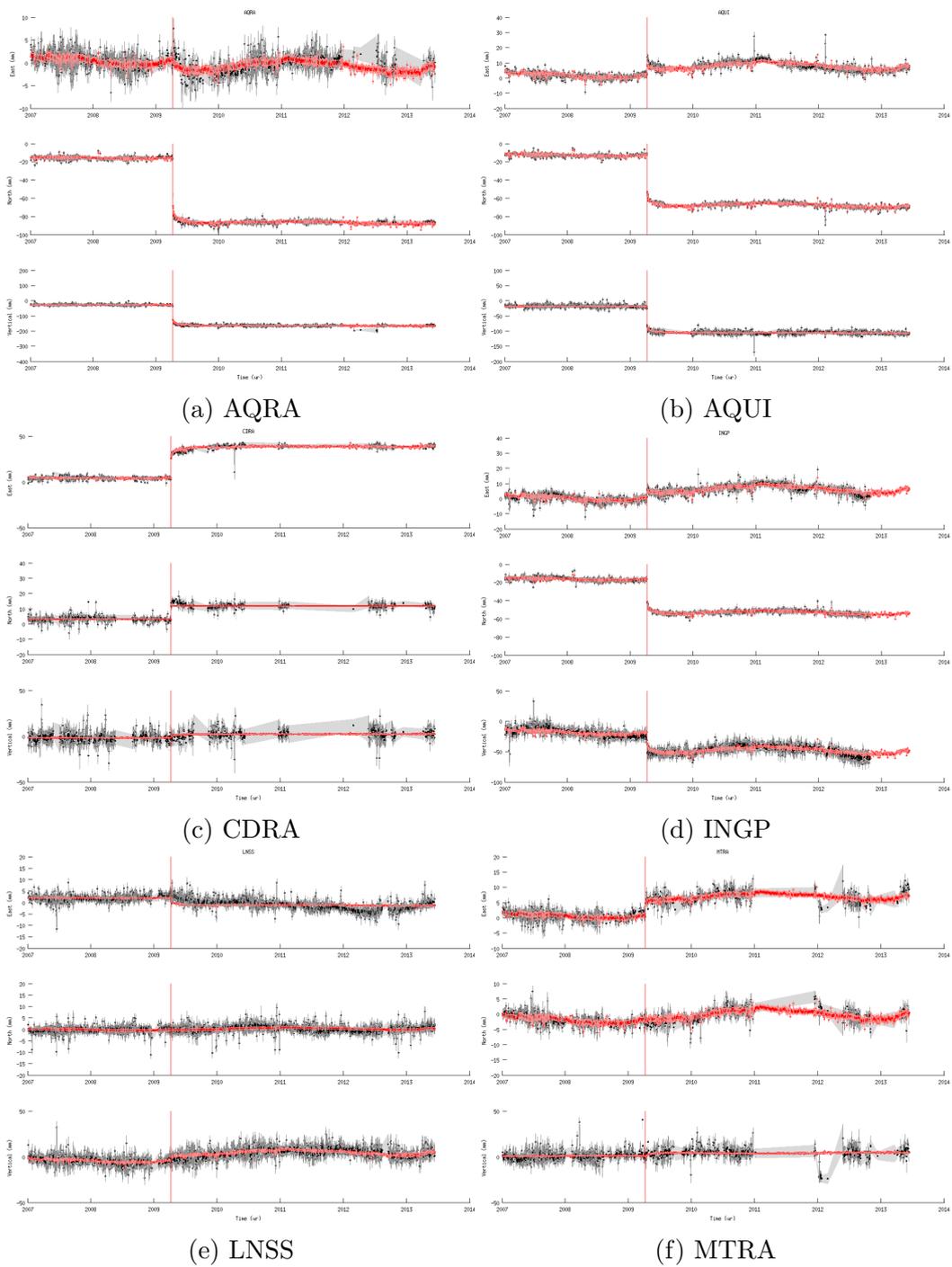


Figure C.3: L'Aquila GPS time series (black dots) and the corresponding vbICA reconstruction (red dots).

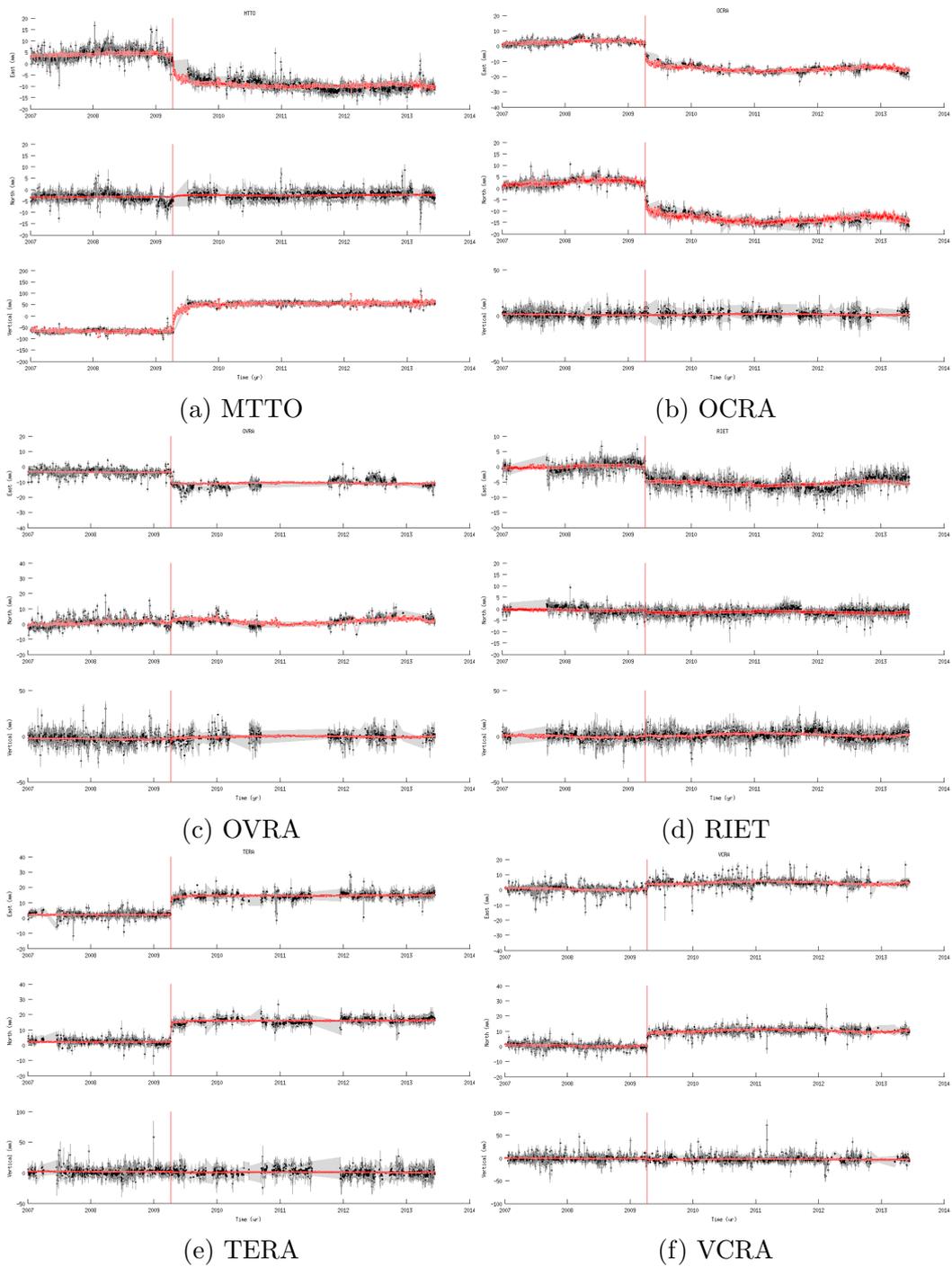


Figure C.4: L'Aquila GPS time series (black dots) and the corresponding vbICA reconstruction (red dots).

## C.2 Emilia

Here I show the time series of the cGPS stations used for the analysis of the Emilia seismic sequence (see Section 4.2). In all figures the black dots represent the data, and the grey shadow is the  $1\sigma$  uncertainty on the data. The red dots correspond to the vbICA reconstruction of the data, with its uncertainty shown in light red.

### C.2.1 Long-term trend + seasonal signals + co-seismic offsets + post-seismic decay

Results on the time series filtered only for the CME (raw time series) (Figures C.5, C.6, C.7). The original time series are generated in the ITRF2008 reference frame [Altamimi et al., 2011]. I correct them for a rotation about the Euler rotation vector specified by the pole of latitude  $55.77^\circ$ , longitude  $-97.82^\circ$ , and magnitude  $0.261^\circ/Myr$  (i.e., Eurasia fixed reference frame).

### C.2.2 Seasonal signals + co-seismic offsets + post-seismic decay

Results on the time series after the removal of a long-term linear trend (Figures C.8, C.9, C.10).

### C.2.3 Co-seismic offsets + post-seismic decay

Results on the time series after the removal of a long-term linear trend and two seasonal signals with annual and semi-annual period (Figures C.11, C.12, C.13).

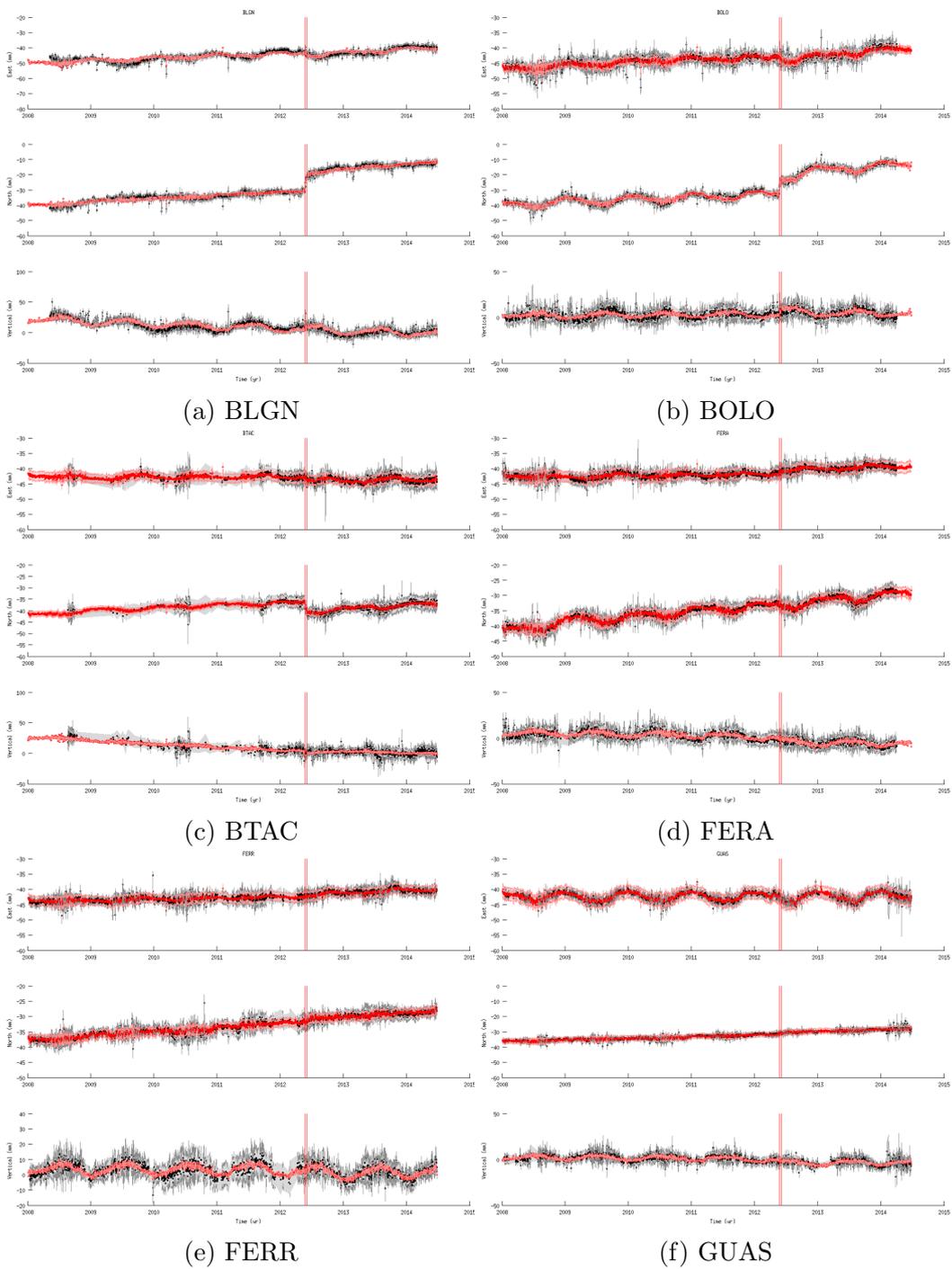


Figure C.5: Raw time series (black dots) and the corresponding vbICA reconstruction (red dots).

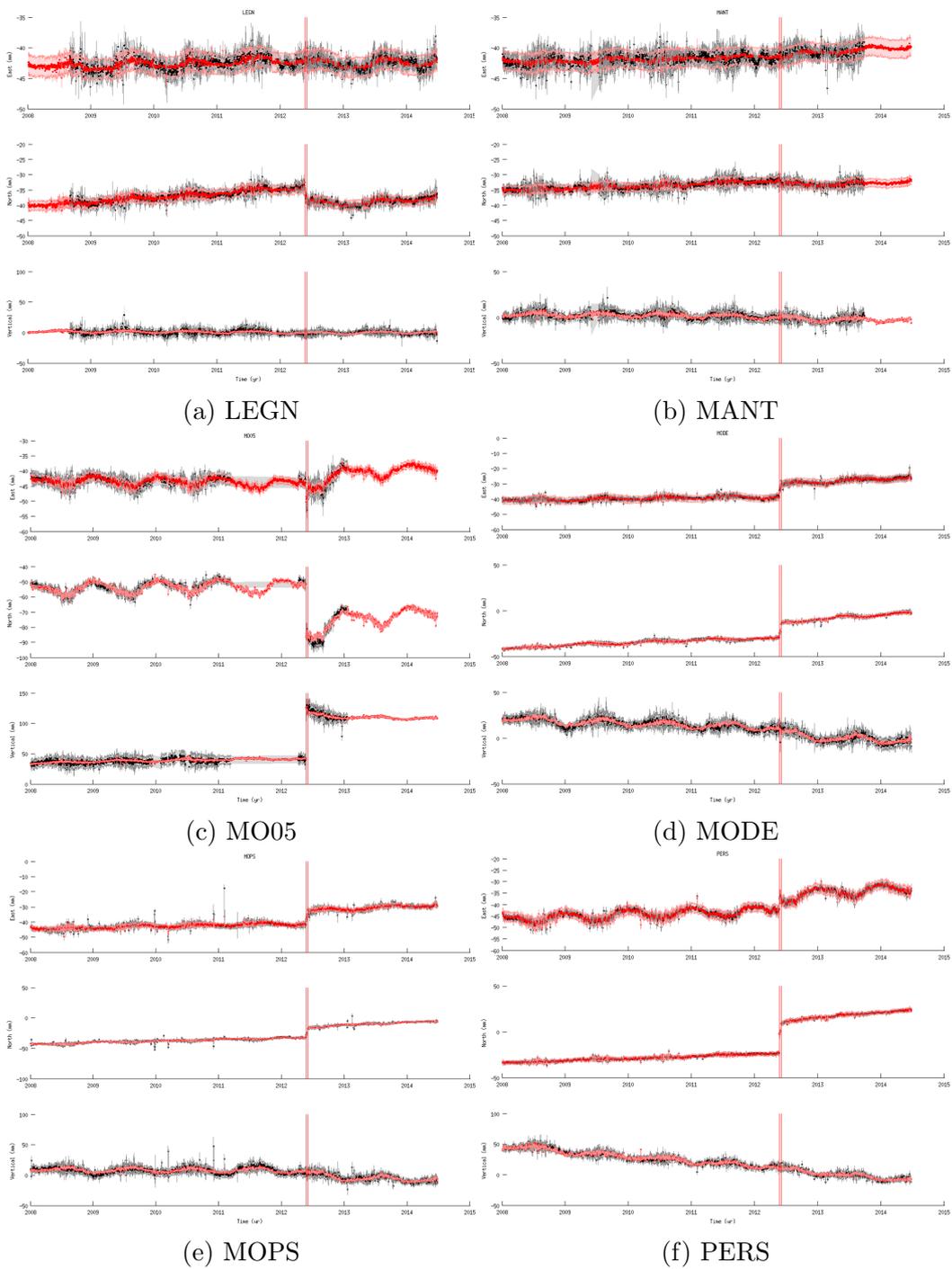
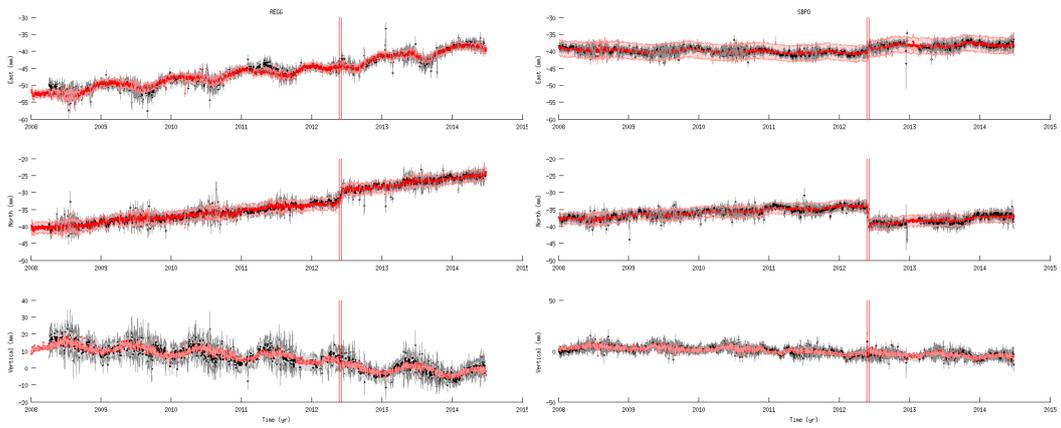
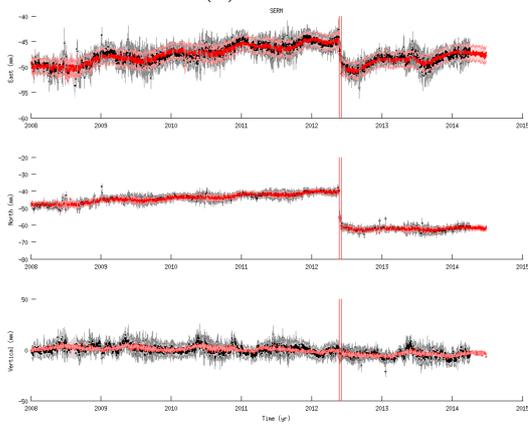


Figure C.6: Raw time series (black dots) and the corresponding vbICA reconstruction (red dots).



(a) REGG

(b) SBPO



(c) SERM

Figure C.7: Raw time series (black dots) and the corresponding vbICA reconstruction (red dots).

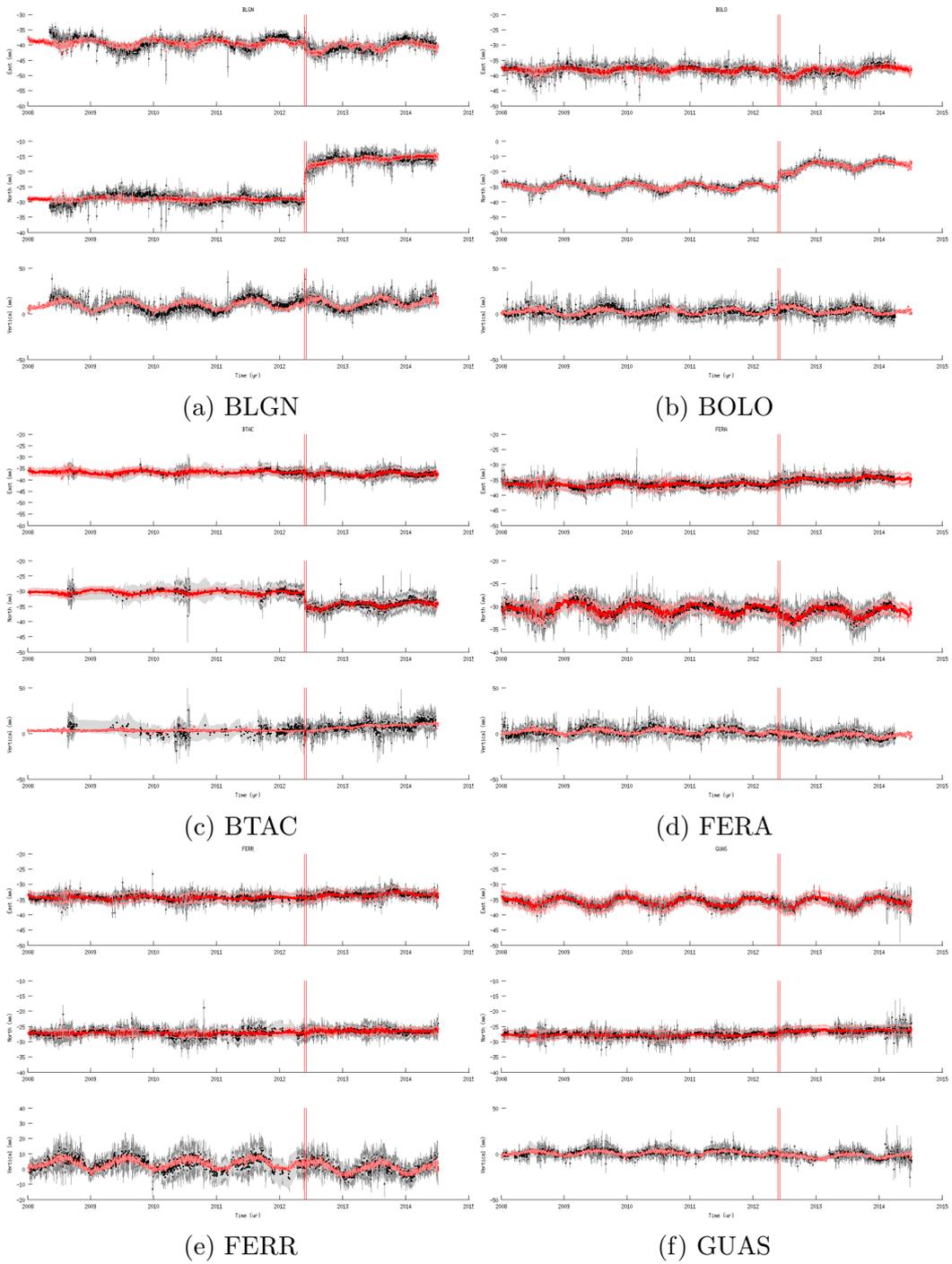


Figure C.8: Detrended time series (black dots) and the corresponding vbICA reconstruction (red dots).

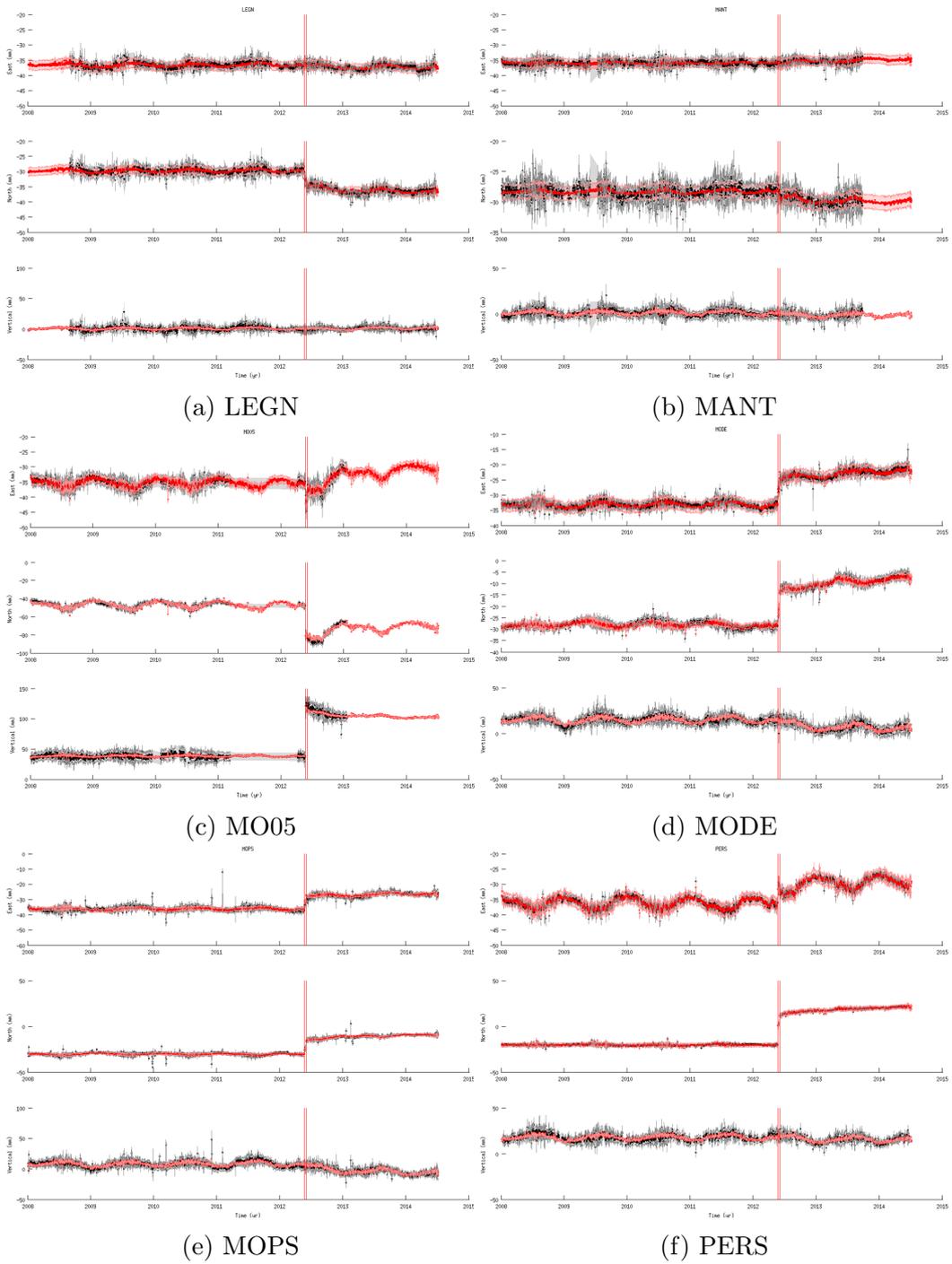
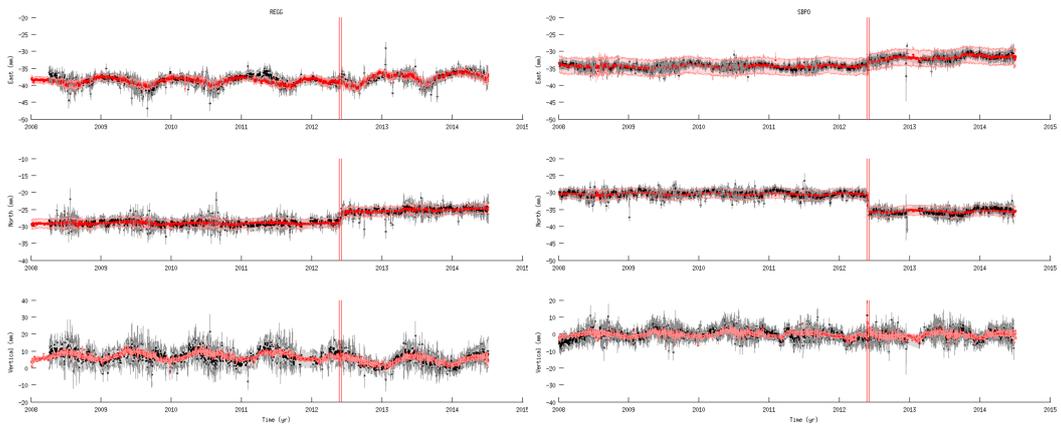
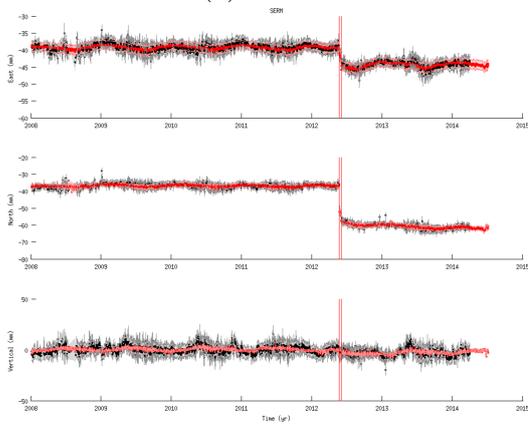


Figure C.9: Detrended time series (black dots) and the corresponding vbICA reconstruction (red dots).



(a) REGG

(b) SBPO



(c) SERM

Figure C.10: Detrended time series (black dots) and the corresponding vbICA reconstruction (red dots).

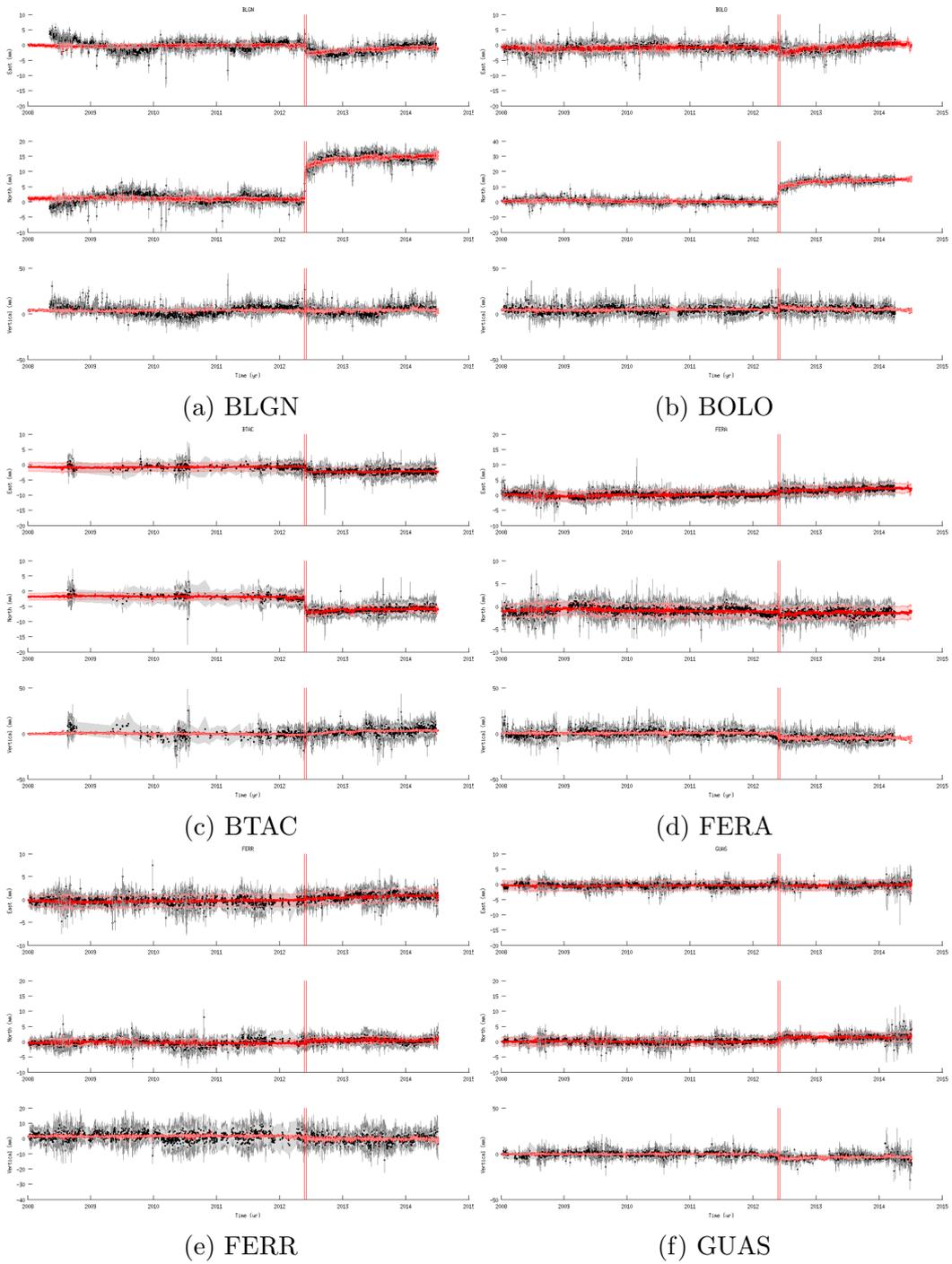


Figure C.11: Detrended time series (black dots) and the corresponding vbICA reconstruction (red dots).

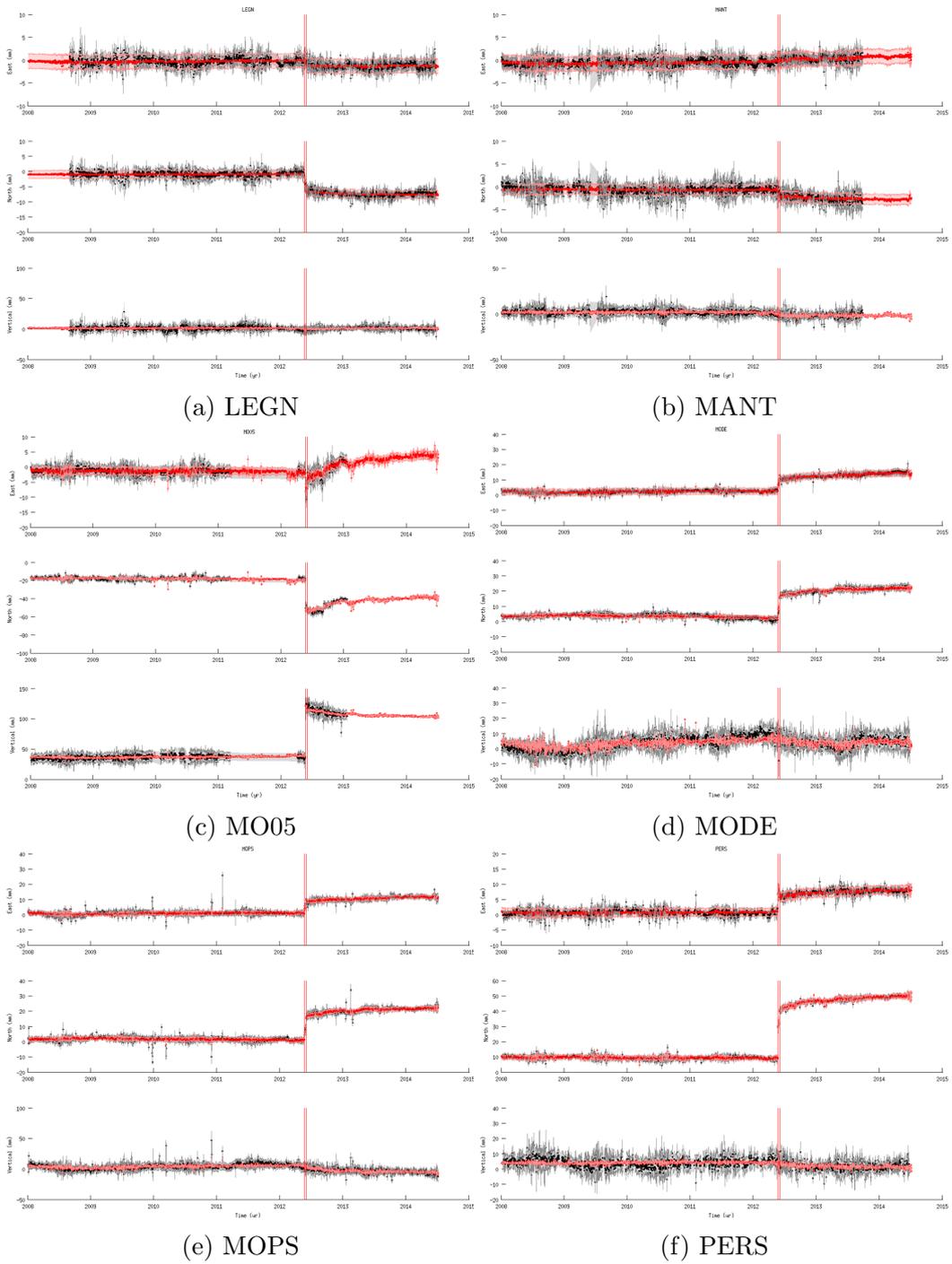
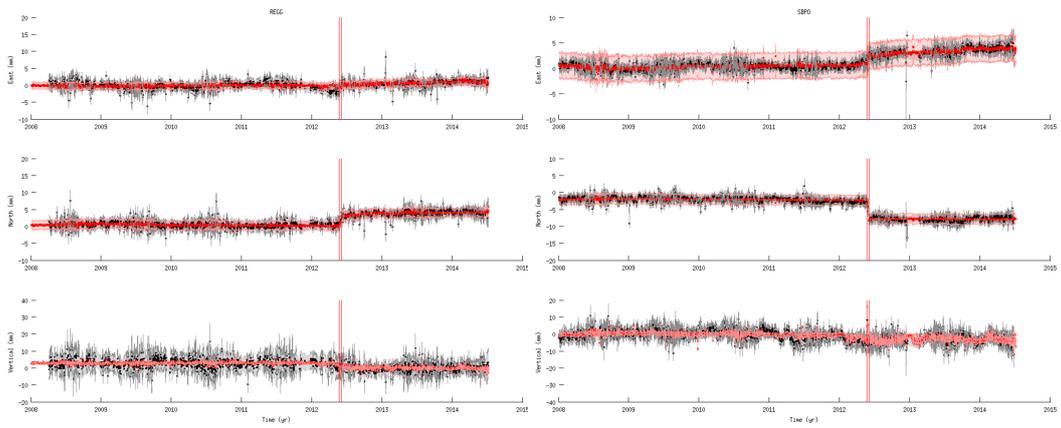
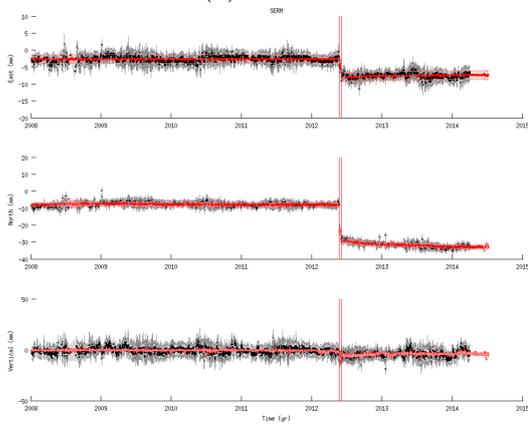


Figure C.12: Detrended time series (black dots) and the corresponding vbICA reconstruction (red dots).



(a) REGG

(b) SBPO



(c) SERM

Figure C.13: Detrended time series (black dots) and the corresponding vbICA reconstruction (red dots).

### C.3 Mexico

Here I show the time series of the cGPS stations used for the analysis of the Guerrero 2006 SSE (Mexico) (see Section 4.3). In all figures the black dots represent the data, and the grey shadow is the  $1\sigma$  uncertainty on the data. The red dots correspond to the vbICA reconstruction of the data, with its uncertainty shown in light red.

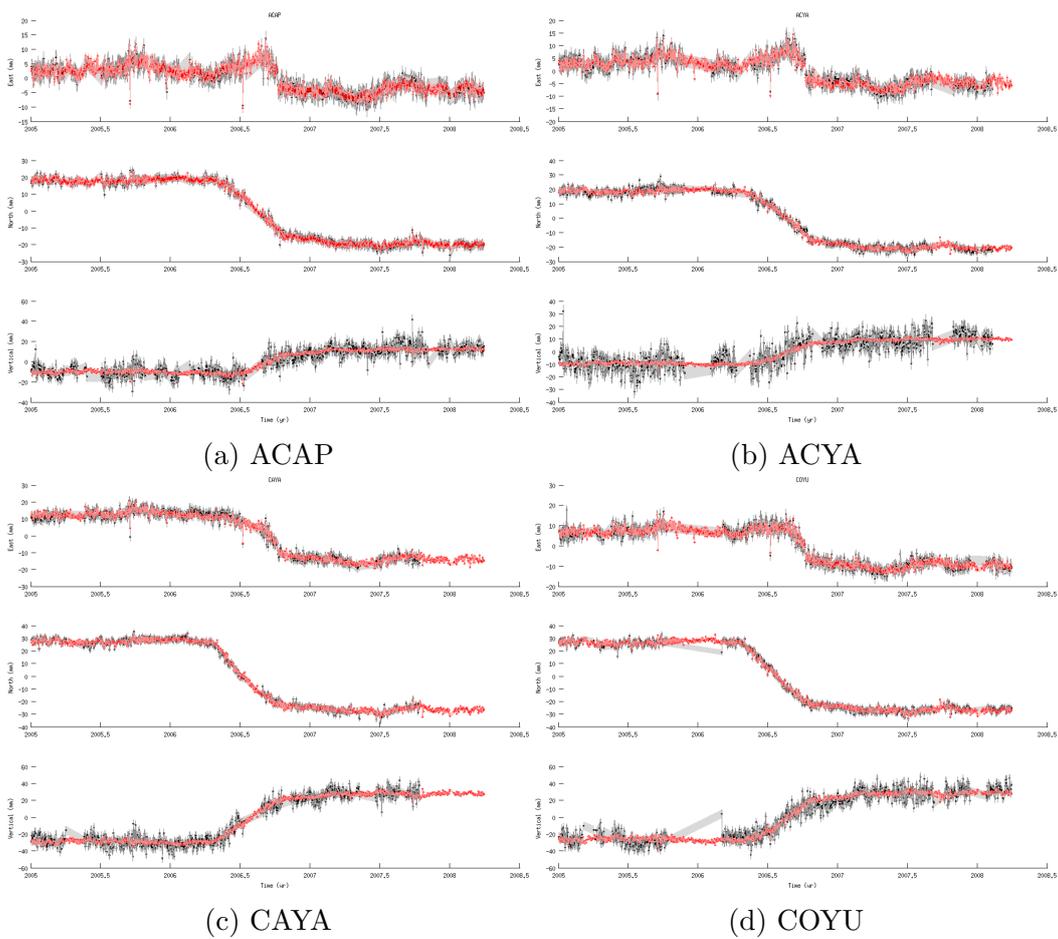


Figure C.14: Mexico GPS time series (black dots) and the corresponding vbICA reconstruction (red dots).

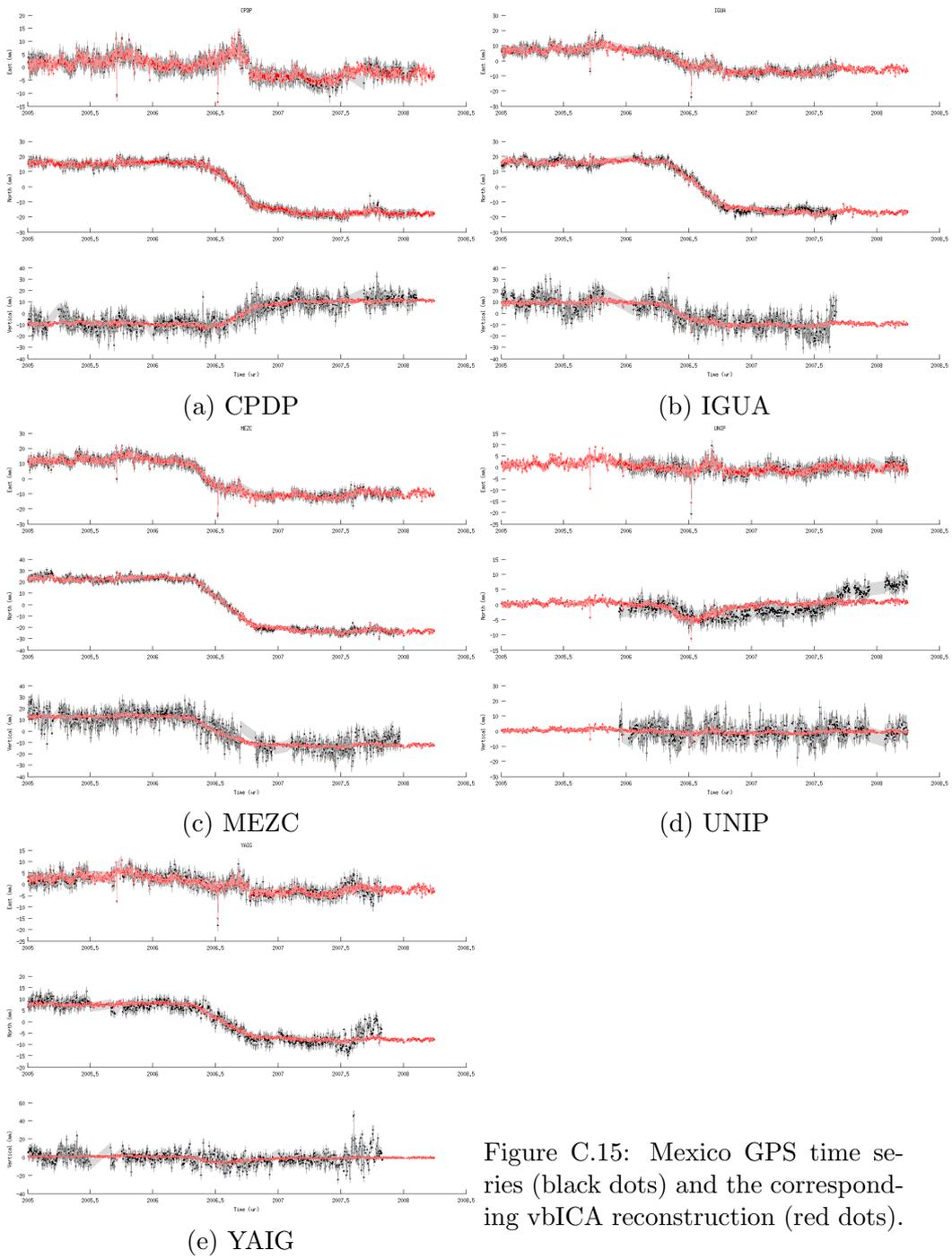


Figure C.15: Mexico GPS time series (black dots) and the corresponding vbICA reconstruction (red dots).

# Appendix D

## Seismic moment estimation from geodetic displacements

The co-seismic displacements estimated at the 15 cGPS stations (see Table 4.3) can be used to determine the energy of the second event, once the energy of the first event is known. Indeed, the displacement recorded at the surface is directly proportional to the work done to move the stations. From the law of conservation of energy, I can relate the work that creates the observed displacement to the energy of the earthquake (the energy released by the earthquake multiplied by seismic efficiency [?]). In other words, the energy of a seismic event is proportional to the amount of displacement recorded at the surface. If we assume point-like sources separated by a distance which is smaller than the minimum source station distance and with the same seismic efficiency, the following proportionality holds:

$$\frac{X_1}{E_1} = \frac{X_2}{E_2} \quad (\text{D.1})$$

where  $X$  is the total displacement recorded at the surface by the entire network of stations, and  $E$  is the energy released by the source that generated the observed displacement. The application of this formula to the Emilia seismic sequence gives the results summarized in Table D.1. The moment

Table D.1: Estimated moment magnitude for the May 29<sup>th</sup> earthquake.

<b>Paper</b>	<b>May 20<sup>th</sup></b>	<b>May 29<sup>th</sup></b>	<b>May 29<sup>th</sup> - geodesy</b>
Pondrelli et al. [2012]	6.1 ± 0.1	5.9 ± 0.1	5.92 ± 0.16
Scognamiglio et al. [2012]	5.86 ± 0.01	5.66 ± 0.01	5.68 ± 0.13
Malagnini et al. [2012]	5.63 ± 0.01	5.44 ± 0.01	5.45 ± 0.13

magnitude is calculated using the formula of Hanks and Kanamori [1979].

The uncertainties are propagated taking into account the logarithmic relationship between the seismic moment and the moment magnitude. Using different reference seismic moments for the May 20<sup>th</sup> earthquake I obtain different seismic moments corresponding to the May 29<sup>th</sup> event, as shown in Table D.1. Anyway, it is remarkable that the internal coherence is preserved. This means that if I have, from seismological observations, a good estimation of the energy of the main event, I can estimate the energy of a near event detected by geodesy. This can be particularly useful for two reasons: 1) In considering equation D.1 I am assuming that the stations are not affected by the spatial extension of the source and then by the non-uniformity of the slip distribution. In other words, it can be applied only if the stations are far enough from the epicenter, therefore their distribution can not resolve adequately the slip distribution on the fault plane. Most of the time, this is the case in geodesy, and this equation allows me to get the information about the amount of energy released without solving for an inverse problem to get the slip distribution and then the seismic moment released. 2) I can estimate the energy released also by aseismic events, such as afterslip induced by the event with known magnitude and closely located Slow Slip Events (SSEs), provided that the equivalent seismic moment of one of them is known.