Geofísica Internacional (2014) 53-1: 77-85

Original paper

Reducing wind noise in seismic data using Non-negative Matrix Factorization: an application to Villarrica volcano, Chile

Giuseppe Cabras*, Roberto Carniel, Joshua P. Jones and Minoru Takeo

Received: January 17, 2013; accepted: April 05, 2013; published on line: December 11, 2013

Resumen

La separación de distintas fuentes existentes en una sola señal sísmica es un problema interesante y al mismo tiempo difícil. En este trabajo presentamos un método semi-ciego para la separación de un solo canal sísmico para mejorar la parte de señal de origen volcánico. En este método, el esquema de descomposición de las fuentes se basa en una factorización en matrices dispersas y non-negativas (Non-negative Matrix Factorization, NMF) de la representación tiempo-frecuencia del único canal sísmico vertical. Como caso de estudio se presenta una aplicación a partir de datos sísmicos registrados en el volcán Villarrica, Chile, uno de los más activos de los Andes meridionales. Los datos analizados están fuertemente contaminados por el ruido del viento y el procedimiento propuesto se utiliza para separar un componente de origen volcánico de otro de origen meteorológico.

Palabras clave: Tremor volcánico, NMF, Villarrica, reducción del ruido del viento.

Abstract

Single channel source separation of seismic signals is an appealing but difficult problem. In this paper, we introduce a semi-blind single-channel seismic source separation method to enhance the components of volcanic origin. In this method, the source decomposition scheme is addressed as a Sparse Non-negative Matrix Factorization (NMF) of the time-frequency representation of the single vertical seismic channel. As a case study we present an application using seismic data recorded at Villarrica volcano, Chile, one of the most active in the southern Andes. The analysed dataset is strongly contaminated by wind noise and the procedure is used to separate a component of volcanic origin from another of meteorological origin.

Key words: volcanic tremor, NMF, Villarrica, wind noise reduction.

G. Cabras*
Dipartimento di Chimica, Fisica e Ambiente
Universitá degli Studi di Udine
Via delle Scienze, 206
Udine, Italy
**Corresponding authors:* giuseppe.cabras@uniud.it

R. Carniel
Laboratorio di misure e trattamento dei segnali
Dipartimento di Ingegneria Civile e Architettura
Universitá degli Studi di Udine
Via delle Scienze, 206
Udine, Italy

Earthquake Research Institute
University of Tokyo
Tokyo, Japan
E-mail: roberto.carniel@uniud.it

Geognosis Proj., Int. Syst. Lab.
National Research University
of Information Technologies
Mechanics and Optics
Kronverksky Prospect, 49
St. Petersburg 197101, Russia

J.P. Jones
Department. of Physics - CCIS
University of Alberta,
Edmonton AB T6G 2E1
Canada
E-mail: jpjones@ualberta.ca

M. Takeo
Earthquake Research Institute
University of Tokyo
Tokyo, Japan
E-mail: takeo@eri.u-tokyo.ac.jp

Introduction

Villarrica volcano (Chile), located in the southern Andes (39.42º S, 71.93º W) has more than 60 important historical eruptions (Casertano, 1963). As other volcanoes of basic composition (Behncke *et al*., 2003; Behncke, 2009), Villarrica can not only produce effusive and moderate explosivity activity, but also pyroclastic flows that represent the most dangerous scenario (Moreno, 1998). It is currently characterized by the presence of a small (30-40 m wide) summit lava lake which produces a persistent volcanic tremor and discrete events associated to strombolian activity (Ortiz *et al*., 2003; Tárraga *et al*., 2008). In November 2004 we installed an L22 three-component geophone (f_0 = 2.0 Hz) approximately 800 m from the summit crater. The seismometer recorded continuously for a period of 10 days, between November 8, 2004 and November 17, 2004.

Noise often affects records of volcanic tremor or ambient ground-motion recordings used e.g. for HVSR estimates of seismic site amplification. Benson *et al*. (2007) introduce several methods to reduce the effects of local, non-stationary noise sources, earthquakes and instrumental irregularities on ambient noise. Lambert *et al*. (2011) propose four methodologies focused on the removal of the effects of anthropogenic noise. In this work, we apply an innovative wind noise reduction procedure to the tremor recorded at Villarrica, based on Non-negative Matrix Factorization with Sparse Coding (Hoyer, 2002) and on the construction of a wind noise dictionary which is estimated from the available seismic recording itself. The presented procedure can be extended to the filtering of wind noise in other volcanic geophysical time series, such as the ones recorded by infrasonic sensors (Ichihara *et al*., 2012).

A two training step strategy to Single-Sensor Seismic Analysis

In this paper, we extend the Non-negative Matrix Factorization (NMF) and Sparse Coding (SC) procedure introduced in (Cabras *et al*., 2012), where the basic idea is that we can obtain a meaningful part-based factor decomposition (Lee & Seung, 1999) from a single-channel time series imposing only the constraints of non-negativity and sparseness of the data. For a general discussion on NMF and SC, see Cichocki *et al*. (2009). As we will describe in detail below, the main contribution with respect to the original procedure (Cabras *et al*., 2012) lies in the training stage, where we learn two sets of basis components in two steps directly from the available dataset: in the first training step we learn an approximate volcanic-set basis components (or preliminary volcanic-set dictionary), $\tilde{D}_s'$, selecting a non windy data-set for training; in the second training step we learn the wind-set basis components (or noise-set dictionary), D_n, selecting a windy data-set for training. The separation stage remains the same of (Cabras *et al*., 2012), providing a fixed wind-set basis components, D_n to the constrained sparse NMF learning loop.

We can state an NMF problem as follows: given a non-negative data matrix $\mathbf{X} \in \mathbb{R}_+^{F \times T}$ (with $x_{ft} \geq 0$ or equivalently $X \geq 0$) and a reduced rank K ($K \leq min(F,T)$), find two non-negative matrices $\mathbf{D} \in \mathbb{R}_+^{F \times K}$, called dictionary or basis components and $\mathbf{H} \in \mathbb{R}_+^{K \times T}$, called sparse code or weights, which factorize X as well as possible, that is:

$$X = DH + E, \tag{1}$$

where $\mathbf{E} \in \mathbb{R}_+^{F \times T}$ represents the approximation error to minimize. The meaning of dictionary matrix D, sparse code matrix H and rank K depends on the specific application and signal representation. To estimate the parameters of NMF, (i.e. the factor matrices D and H), we need to minimize the measure of similarity (or cost function C) between the data matrix X and the estimated model matrix $\hat{X} = DH$; the simplest and widely used measure is the squared Euclidean distance (or Frobenius norm):

$$argmin_{D,H} C_F(X \parallel DH) = \tfrac{1}{2} \parallel XDH \parallel_F^2 + \lambda \parallel H \parallel_1 \tag{2}$$

where λ ia a non-negative regularization parameter that controls the tradeoff between sparseness and reconstruction error and $\parallel H \parallel_1$ is an ℓ_1 norm regularization function proposed in Hoyer (2002).

Cabras *et al*. (2012) adopted the single channel enhancement model originally developed for processing audio records (Cabras *et al*., 2010) to separate a "high convective", relatively transient, seismic source of interest from a "low convective", relatively continuous, "noise" in a single-sensor seismic time series recorded at Erta 'Ale volcano (Harris *et al*., 2005). Erta 'Ale is characterized by the presence of a permanent lava lake that produces, in a similar way to Villarrica, at least part of a continuous volcanic tremor (Jones *et al*., 2006), which is really characterized by a superposition of different independent sources (Jones *et al*., 2012). In the Erta 'Ale case study, we learned the basis components of the noise $n(t)$, denoted by D_n, in a single step training stage, because in our data-set we have segments of pure "low convective noise" for training.

In the present case study of Villarrica volcano, this strategy is not directly applicable. In fact, our data-set is characterized by a relatively continuous "volcanic tremor", our source of interest, immersed in a relatively transient "wind" noise that we want to suppress. Wind components change rapidly in