

MSNoise, a Python Package for Monitoring Seismic Velocity Changes Using Ambient Seismic Noise

by Thomas Lecocq, Corentin Caudron, and Florent Brenguier

INTRODUCTION

The usage of seismic ambient noise has recently proved its efficiency in different contexts, from imaging to monitoring. The impulse response (or Green's function [GF]) between two sensors can be reconstructed from the correlation of seismic noise recorded (Campillo and Paul, 2003). This method has provided excellent results in imaging the Earth's interior, from global to regional or local scales. More recently, the method was extended to study time-dependent variations in those GF. A change in the delay times might originate from a change in the medium velocity or from a dramatic change in the position of the source or of one/many scatterers. Several studies using seismic ambient noise have shown that small perturbations within a volcanic edifice can be detected as changes in seismic-wave properties (Sens-Schönfelder and Wegler, 2006; Brenguier, Shapiro, *et al.*, 2008; Duputel *et al.*, 2009; Mordret *et al.*, 2010; Brenguier *et al.*, 2011; Anggono *et al.*, 2012). Contrary to the use of active sources or earthquake coda waves, the technique takes advantage of the continuous sampling of the medium using around-the-clock records from seismic stations. The method has proven its ability to evidence temporal physical changes in fault zones (Wegler and Sens-Schönfelder, 2007; Brenguier, Campillo, *et al.*, 2008), the lunar environment (Sens-Schönfelder and Wegler, 2011), or to detect instrumental errors (Stehly *et al.*, 2007; Sens-Schönfelder, 2008).

Some codes have already been presented to compute cross correlations of seismic noise, for example, within Seismic Analysis Code (SAC) (Goldstein *et al.*, 2003) or within Computer Programs in Seismology (CPS) 3.3 (Herrmann, 2002). To our knowledge, no integrated solution has been published to go from the raw waveforms to the travel-time variations, automatically detecting changes in the data archive and computing only what is necessary on a scheduled basis (hourly, daily) and which is also usable as a research tool to process archives. Automating the detection is important, as, for example, data streams coming from real-time telemetered stations contain gaps that could be filled up later on, and provide important data to be analyzed. Such a tool must be able to work with any common seismic format, be reasonably fast, optimized whenever

possible. It should interact with a data archive and a database with high-level helper functions in order to be pluggable and extensible. Finally, it must produce exportable data, either in waveform format, tabular text files, or high-quality figures. This is the purpose of Monitoring using Seismic Noise (i.e., MSNoise).

We do not want to provide another black box to the users so MSNoise is an integrated solution, its code is open, commented, and documented. The processing workflow is separated in steps and one can easily replace one step by another code, providing that the inputs and outputs are respected.

In this paper, the software steps are presented both from the IT and scientific/methodological points of view. Example graphical outputs are shown and described. We finally validate MSNoise using archive data from the Piton de la Fournaise volcano (La Réunion, France), where precursory seismic velocity changes have been observed prior to two eruptions in October and December 2010, during the UnderVolc project (Brenguier *et al.*, 2012). UnderVolc is an Agence Nationale de la Recherche (ANR) project and stands for "UNDERstanding VOLcanic Processes: Towards Eruption Prediction and Risk Mitigation, an application to Piton de La Fournaise volcano, La Réunion (2009–2013)."

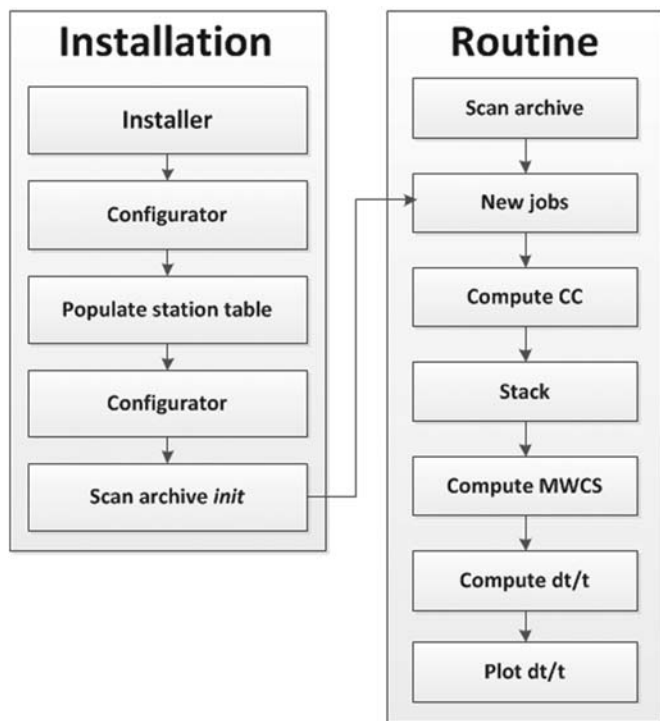
MSNoise is written in Python to be fully cross platform. It is open source and free, for noncommercial educational and research usage. MSNoise is licensed under the European Union Public License (EUPL v1.1). Commercial services related to MSNoise, for example, providing consultancy or support for its implementation or maintenance, are prohibited without signed agreement with the original authors. MSNoise is available on <http://www.msnoise.org/> (last accessed March 2014).

SOFTWARE FUNCTIONALITY

The workflow of MSNoise is pretty straightforward (Fig. 1) and is composed of simple steps that are described extensively on the companion website.

Installation and Configuration

MSNoise requires Python and only a few extra packages for running on any operating systems (OS). The waveform archive must be either a known format, for example SeisComP Data Structure (SDS) or Buffer of Uniform Data structure (BUD), or one has to define its own. MSNoise uses a database for storing configuration bits, waveform metadata, and jobs. This database can be either MySQL or sqlite, but we recommend using MySQL for performance reasons. The package comes with an installer script



▲ **Figure 1.** Schematic view of the MSNoise workflow, the one-time installation part enters the routine workflow before the job definition step.

that will initialize the connection to the database and store the user name and password locally to allow the Configurator and all processes to interact with the database. This Configurator GUI (built using Enthought Tools Suite [Enthought, 2008]) shows three different panels: the general MSNoise configuration, the stations configuration, and the filters configuration. All configuration elements are described in the online documentation.

Automatic Job Definition

Data Discovery

To automatically run every night on the data acquired during the previous day, MSNoise needs to check the data archive for new or modified files. Those files could have been acquired during the last day, but could also be the data from a previously offline station and contain useful information for, say, a month ago. The time to search for is defined in the *config*. MSNoise uses the *find* command (*gnufind* on Windows) with the *-mtime* argument to locate new or modified files. Once located and read (using *obspy* [Beyreuther et al., 2010; Megies et al., 2011]) they are inserted (if new) or updated (if modified) in the *data availability* table. This table can then be turned into a *data availability plot* (Fig. 2). Upon first run, the script must be called with the *init* argument in order to avoid using the *mtime* argument for the *find* command and to insert all discovered files in the *data availability* table.

Job Definition

Once all new/modified files have been identified, MSNoise will define which days need to be processed and for which station

pair(s). For a set of N stations, there will be M pairs to compute (equation 1a). If one wants auto-correlation, M will be defined by equation (1b):

$$M = N \times (N - 1) / 2 \quad (1a)$$

and

$$M = N \times (N - 1) / 2 + N. \quad (1b)$$

Outputs, Extensibility, and Plugability

The original goal of MSNoise is to provide $\delta v/v$ plots over time. These data can be represented as several plots, one per station pair, one average for all station pairs, etc. MSNoise comes with several example plots that one can duplicate or hack to meet its needs. MSNoise uses *pandas* for data analysis (McKinney, 2012) as it provides great helper functions to analyse time series data (rolling average, resampling, detrending time series containing NaNs...) and *matplotlib* (Hunter, 2007) for static graphical outputs.

MSNoise has been designed with plugability and extensibility in mind. All the communications with the database, with the data archive or with the cross-correlation files go through helper functions. None of the processing steps makes direct connections to the database or to the data files. Extension makers are advised to look at the documentation of the functions to build their extensions. A good starting point is to have a look at the example output plots provided with the package. We want to emphasize that although we currently prove MSNoise is working as expected (see the Validation section below), one might be interested in modifying or using other computation routines. This can be easily done by replacing steps in the workflow.

COMPUTATION

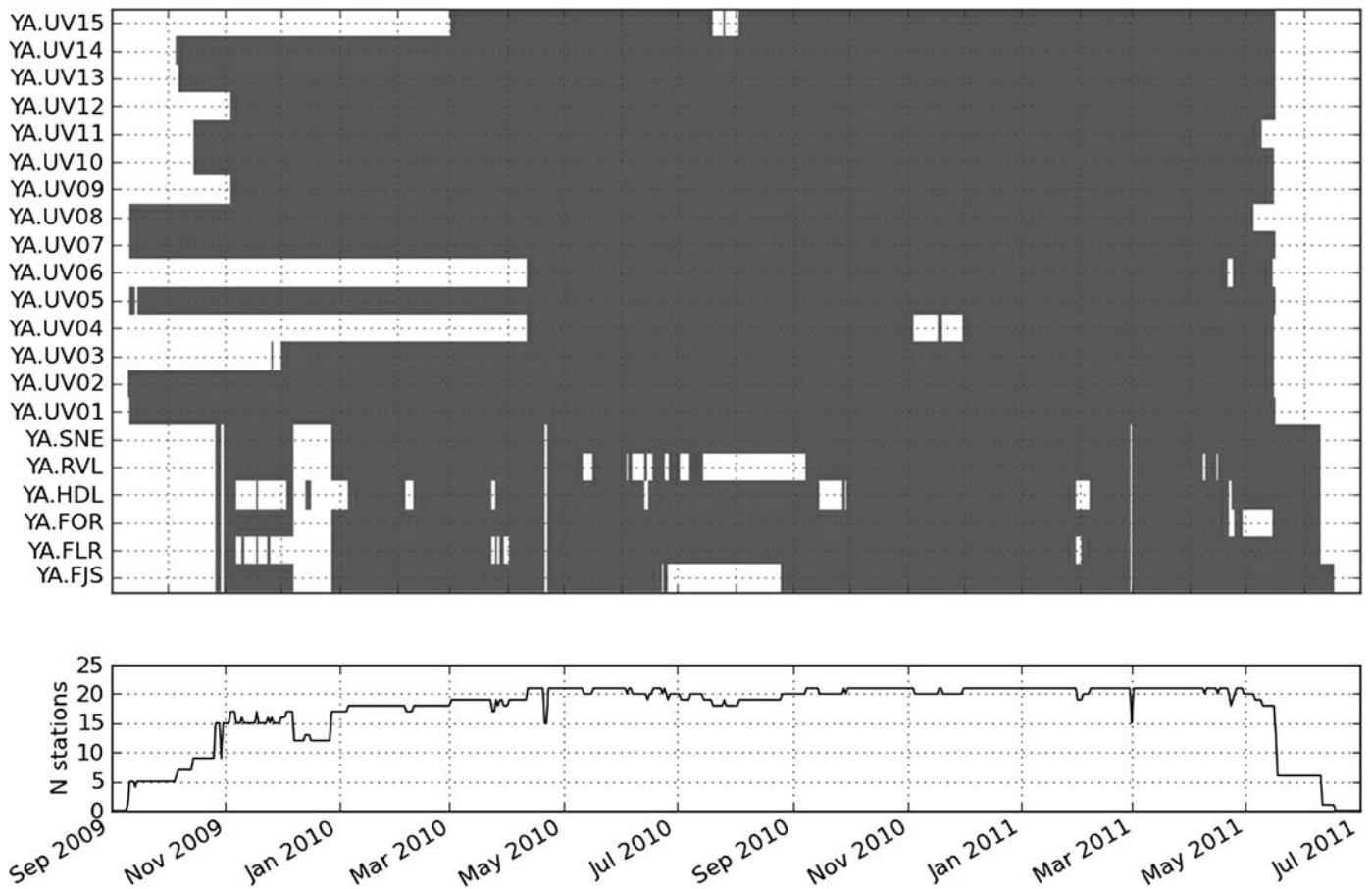
Our approach for continuously measuring seismic velocity changes of the Earth's Interiors relies on three steps. (1) The first step is computing cross-correlation functions (CCFs) of ambient seismic noise time series at different dates for individual pairs of seismic sensors; (2) measuring travel-time delays of different arrivals (direct or coda waves) between these individual CCFs and a defined reference CCF, and (3) averaging these travel-time delays for different correlation lag times over different sensor pairs and interpreting these travel-time delays using a simple model of uniform relative velocity change within the studied area ($\delta v/v = \text{constant}$).

Computation of Cross-Correlation Functions

For each cross-correlation job, the waveforms are preprocessed and then analyzed. Jobs are independent, so several computations can be run in parallel, usually depending on ones' CPU-RAM-Disk speed triplet.

Waveform Preprocessing

For each station, all files that potentially contain data for the day are opened. The traces are then merged and split, to obtain



▲ **Figure 2.** Data availability for the UnderVolc project: days with at least some data are plotted in gray and the N number of available stations for each date is summarized in the lower part. This plot is made automatically using the `plot_data_availability.py` script.

the most continuous chunks possible. The different chunks are then demeaned, tapered, and merged again to a one day long trace. If shorter than one day, the trace is padded with zeroes. If longer, it is cut to match the start/end of the day. Subsequently, each one day long trace is low passed, high passed then decimated, or downsampled. Decimation is faster, but only allows decimation by integer factors, whereas the downsampling supports any factor which allows the usage of heterogeneous station configurations. Decimation/Downsampling are configurable, and users are advised to test both. The resampling method used in MSNoise comes from the audio world and provides excellent quality results (de Castro Lopo, 2013). Low-pass and down-sampling values must be defined together properly.

Processing

Once all waveforms are loaded in memory, the computation is done by iteration on the station pairs, on the different components to compute and then on the different filters for each defined window (30 minutes slices by default, configurable).

$$\begin{aligned} R &= N \cos(Az) + E \sin(Az) \\ T &= N \sin(Az) - E \cos(Az). \end{aligned} \quad (2)$$

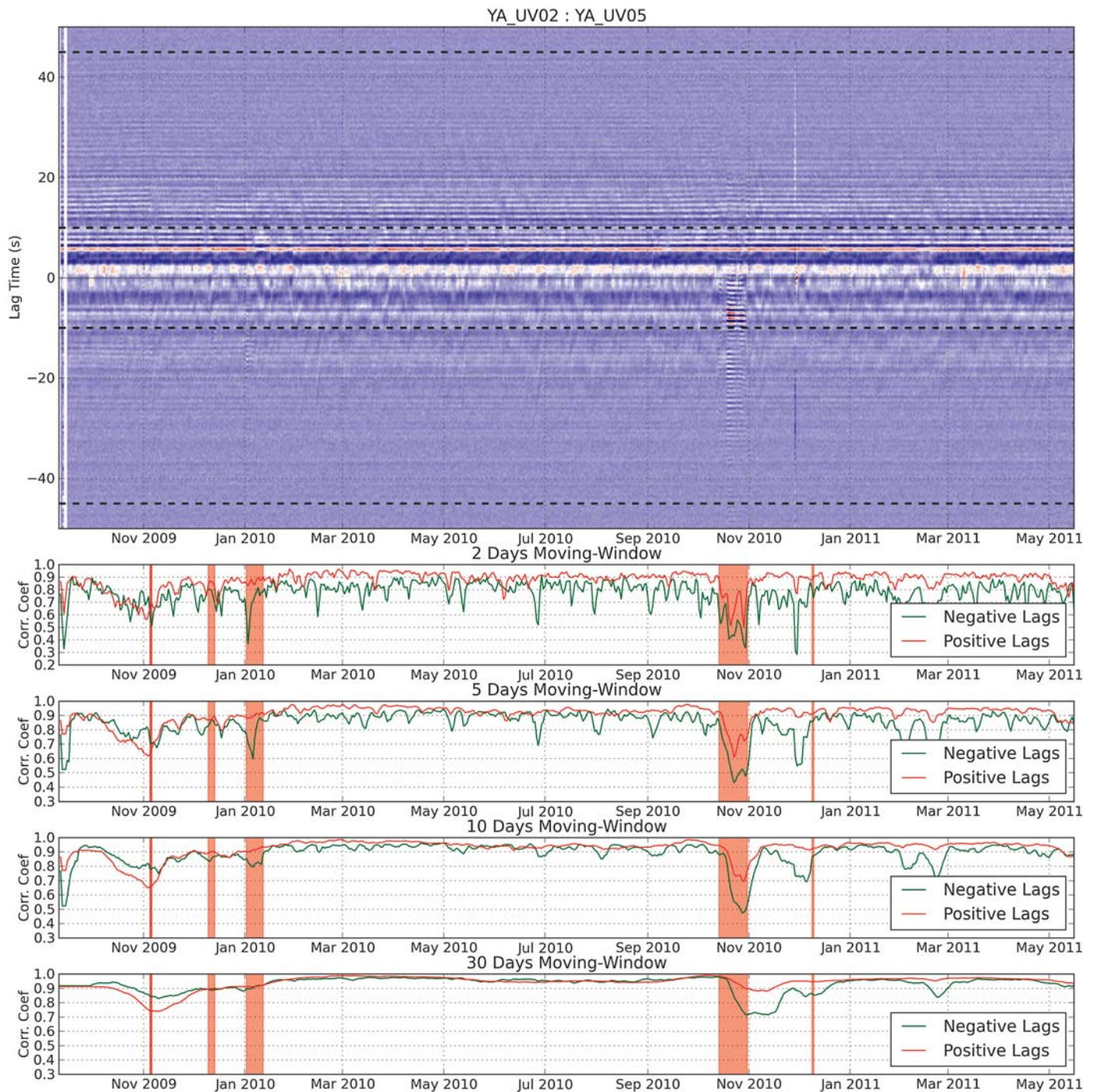
The radial (R) and transverse (T) components are calculated from the rotation of the east (E) and north (N) components. The rotation angle is defined as the azimuth (Az) between the two sensors.

In order to attenuate the parasitic signals from seismic events (being local, regional, or global), we apply a Windsorizing (Tukey, 1962) of three times RMS to the traces as a first-processing step. A second step consists of whitening the amplitude of the signal between two configured frequencies.

The cross correlation is conducted in the frequency domain. Let $x(t)$ and $y(t)$ be two time series and $X(f)$ and $Y(f)$ their Fourier transforms. The correlation operation is then defined as

$$C(f) = X^*(f) \times Y(f), \quad (3)$$

with $X^*(f)$ being the complex conjugate of $X(f)$. If $x(t) = y(t)$, then the operation is called auto correlation. The CCF $c(t)$ is the inverse Fourier transform of $C(f)$. If configured (not by default), the $c(t)$ for each 30 minutes slice can be stored to an output directory on the computer. Again, if configured (default), the day stack of the non-zero, non-infinite, and non-NaN 48×30 minutes $c(t)$ are stored.



▲ **Figure 3.** Cross-correlation functions (CCFs) through time for two random stations (YA.UV02 and YA.UV05) of the UnderVolc project (above). The horizontal dashed lines represent the limits in the negative and positive time lags defining two domains that are compared to the reference (here, a stack of all available data). Correlation coefficients are determined by comparing different moving-window stacks to the reference (lower four graphs). Eruptions of the Piton de la Fournaise are shown as red vertical bands.

Stacking Strategy and Export

MSNoise is capable of using a reference function (REF) defined by absolute or relative dates span. For example, an absolute range could be from 1 January 2010 to 31 December 2011 and a relative range could be the last 200 days. To define this reference, we propose to plot all CCF already computed and check for their

consistency. Once a stable enough period has been identified, it can be defined as the reference function (REF) for further analysis.

The correlation coefficient of each daily CCF with a given REF can be evaluated and is already a good indicator that something happened under, around or to ones' sensors (Fig. 3).

Signals in the negative and positive time lags can sometimes be different (Mordret *et al.*, 2010), as shown on Figure 3. A change in the velocity should bring smooth changes in the correlation, whereas changes in the origin of the seismic noise or in the position of some of the scatterers should dramatically change the correlation (Wegler and Sens-Schönfelder, 2007).

Computing velocity variations by comparing daily CCF with the REF can bring strong variations because of noisy CCF. Cross correlating an increasing number of days stacked together with the REF helps deciding a reasonable number of days for which the correlation coefficient reaches 0.96. A drawback of taking a too long moving-window stack is an immediate decrease in resolution. Liu *et al.* (2010) have demonstrated that the choice of the window size might highlight or hide features in the data, so we made MSNoise capable of exporting multiple moving-window stacks, for example, 2, 5, 10, and 30 days stacks. Once the user has defined the reference and a (some) reasonable moving average(s), both need to be exported before computing the relative travel-time variations. Similarly, the definition of REF requires attention. If one takes the whole archive as reference, but it actually contains strong-velocity variations, the quality of the reference will be greatly diminished (Duputel *et al.*, 2009; Mordret *et al.*, 2010). The investigation of stable periods compared with the total data length is highly recommended.

Only data for new/modified dates need to be exported. If any cross-correlation (CC) job has been marked “Done” within the last day, the stacks will be calculated, and a new travel-time variation (DTT) job will be inserted in the database. Note, the subsequent $\delta t/t$ calculations will only be done for dates for which new/modified CC is available, or for which a new REF has been determined. If one defines a moving-window REF, then the REF stack will need to be calculated every time, and the travel-time variations will need to be completely recalculated too. With a fixed REF, only the $\delta t/t$ for the previous day should be calculated, which makes the whole process faster.

As described in Hadziioannou *et al.* (2009), the accuracy of seismic velocity change measurements relies on the reconstruction of stable in time CCF. As a first step, one must thus ensure that sufficiently long noise time series have been correlated in order to reach, for every sensor pair, a coherence between individual and the reference CCFs of a certain threshold. An obvious trade-off is the longer the time series, the higher the coherence, but the lower the time resolution.

Relative Travel-Time Variations

Time Delay Measurements

Two techniques can be used to estimate the time delays: the moving-window cross spectrum analysis (MWCS; first introduced by Ratdomopurbo and Poupinet, 1995) and passive interferometry (Sens-Schönfelder and Wegler, 2006). Although two studies highlighted the stability of the second technique over the first (Duputel *et al.*, 2009; Hadziioannou *et al.*, 2009), the MWCS technique has the advantage of operating in the frequency domain, in which the bandwidth of coherent signal in the correlation function can be clearly defined (Clarke *et al.*,

2011). Furthermore, the recent development of the technique by Clarke *et al.* (2011) has led to some improvement of MWCS technique, particularly the estimation of statistical parameters which allows assessment of the variability above which perturbations can be reliably observed. In MSNoise, time delays are computed following the methodology of Clarke *et al.* (2011), rewritten in Python for consistency. In the Validation section of this paper, we compare the results obtained by MSNoise with the ones obtained using Daniel Clarke’s measured FORTRAN program (personal comm., 2011).

The current CCF is compared with the reference. Both time series are sliced in several overlapping windows. Each slice is mean adjusted and cosine tapered (1% taper at both ends) before Fourier transformation to the frequency domain. $F_{\text{cur}}(\nu)$ and $F_{\text{ref}}(\nu)$ are the first halves of the Hermitian symmetric Fourier-transformed segments that were padded with zeroes to a length equal to the next power of 2 of their length. The cross spectrum $X(\nu)$ is defined as

$$X(\nu) = F_{\text{ref}}(\nu) \times F_{\text{cur}}^*(\nu), \quad (4)$$

in which asterisk denotes the complex conjugation. $X(\nu)$ is then smoothed by convolution with a Hanning window. The similarity of the two time series is assessed using the cross coherence between energy densities in the frequency domain:

$$C(\nu) = \frac{|\overline{X(\nu)}|}{\sqrt{|F_{\text{ref}}(\nu)|^2 |F_{\text{cur}}(\nu)|^2}}, \quad (5)$$

in which the overline here represents the smoothing of the energy spectra for F_{ref} and F_{cur} and of the spectrum of X . The mean coherence for the segment is defined as the mean of $C(\nu)$ in the frequency range of interest. The time delay between the two cross correlations is found in the unwrapped phase, $\phi(\nu)$, of the cross spectrum and is linearly proportional to frequency:

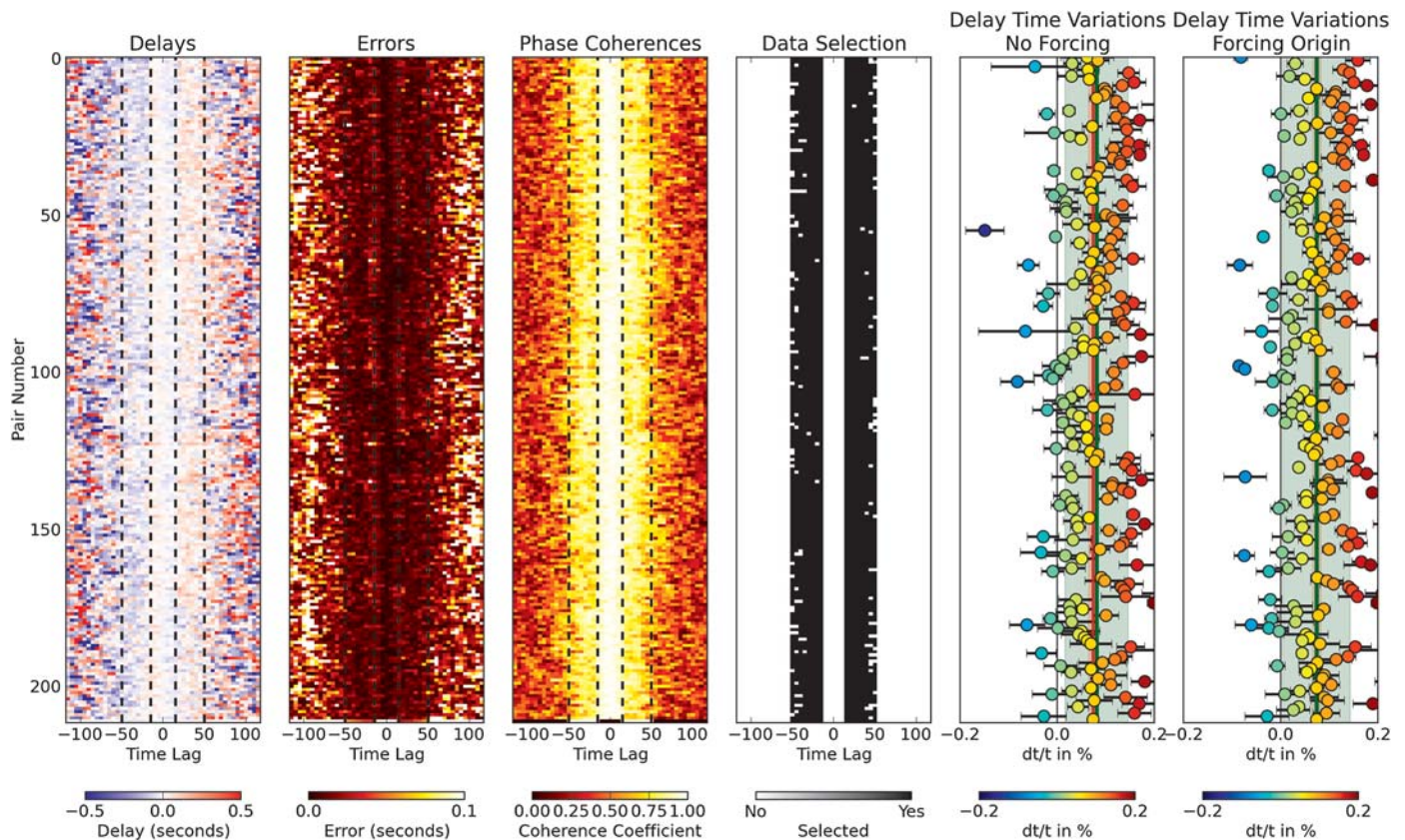
$$\phi_j = m\nu_j, \quad m = 2\pi\delta t. \quad (6)$$

The time shift for each window between two signals is the slope m of a weighted linear regression of the samples within the frequency band of interest. The weights are those introduced by Clarke *et al.* (2011), which incorporate both the cross-spectral amplitude and cross coherence, unlike Poupinet *et al.* (1984). The errors are estimated using the weights (thus the coherence) and the squared misfit to the modeled slope:

$$e_m = \sqrt{\sum_j \left(\frac{w_j \nu_j}{\sum_i w_i \nu_i^2} \right)^2 \sigma_\phi^2}, \quad (7)$$

in which w are weights, ν are cross coherences, and σ_ϕ^2 is the squared misfit of the data to the modeled slope and is calculated as

$$\sigma_\phi^2 = \frac{\sum_j (\phi_j - m\nu_j)^2}{N - 1}. \quad (8)$$



▲ **Figure 4.** Example delay, error, phase coherence, data selection, and delay time variation matrices, here for all station pairs available on 12 October 2010 during the UnderVolc project. Each line of the matrices corresponds to a station pair, whereas each column of the four first matrices is the central time lag of the moving windows in the MWCS process. Each matrix has its own color scale. The bottom line of the matrices is the filtered column weighted average (see text for details) and is called “ALL” within MSNoise. The “ALL” line is only calculated between $-\text{left_maxlag}:\text{left_minlag}$ and $\text{right_minlag}:\text{right_maxlag}$ time lags (dashed lines). Each $\delta t/t$ value determined by the weighted least-square regression (forced to cross the origin or not) of the selected data is represented as a scatter point with error bars. The red vertical line is the $\delta t/t$ (no forcing: $0.074\% \pm 0.009\%$ and with forcing: $0.074\% \pm 0.008\%$) determined for the “ALL” line and the green line is the weighted average (without forcing: $0.081\% \pm 0.065\%$ and with forcing: $0.074\% \pm 0.069\%$) of all $\delta t/t$. Forced and not-forced $\delta t/t$ values are very similar, but the error on the latter is an order of magnitude larger than on the former.

The output of this process is a table containing, for each moving window: the central time lag, the measured delay, its error, and the mean coherence of the segment. Tables are saved as text files (comma-separated values [csv]) to the disk. For any given day, MSNoise computes M sets (1) of delay time versus lag time and saves the result in text files.

Velocity Variations Computation

If one assumes a relative velocity variation $\delta v/v$ homogeneous in space and a relative time shift $\delta t/t$ between the reference and the current CF, it has been demonstrated (Ratdomopurbo and Poupinet, 1995) that

$$\delta v/v = -\delta t/t. \quad (9)$$

In the following sections, we will mainly speak about $\delta t/t$ because this is the raw information stored by MSNoise, and one can easily convert it to $\delta v/v$ when needed.

The MWCS delay-time tables saved at the previous step can then be visualized as a delay matrix for each day (Fig. 4),

evidencing the slight delay differences between pairs as a function of the time lag. MSNoise is able to compute $\delta t/t$ for each station pair, but also for a filtered weighted average of the pairs. An extra line is added at the bottom of the matrix once all station pairs are loaded. This line, called the “ALL” line within MSNoise, will contain a filtered column weighted average of the delay times. For each column (lag time), delay times must satisfy given rules to be included in the weighted mean calculation. Example rules are presented for the UnderVolc data in the Validation section.

The selection of the minimum and maximum lag times for the time shift computations is of much importance. As scattered waves traveling along longer paths accumulate larger time delays, most of the studies performed in volcanic environment were using the late arrivals of the CCF (e.g., Duputel *et al.*, 2009, 5–20 s or Mordret *et al.*, 2010, 10–30 s). An exception arises from the recent study of Anggono *et al.* (2012), in which ± 10 to ± 1 s of lag times are used. Because of poor waveform similarity in the coda part (> 10 s), they neither used the

MWCS nor the stretching method, but compared the maximum of the daily CCF with the one from a reference CCF. We advise the user to perform preliminary tests in different windows and to avoid working around zero time lag. Moreover, the CCF is merely symmetric on volcanoes especially at short lag times (i.e., the ballistic waves described by Duputel *et al.*, 2009). This might evidence an inhomogeneous distribution of sources, a preferred location for the scatterers which act as secondary sources (Paul, 2005; Stehly *et al.*, 2006) or reflect clock problems (Sens-Schönfelder, 2008). Mordret *et al.* (2010) averaged the causal and anticausal parts of the CCFs to achieve a better coherency and stability of the reconstructed wavefield. However, one can also decide to solely work in the causal or acausal part of the retrieved CCFs. Ambient noise sources change in time (change in spectrum, in location and amplitude), and thus seismic velocity changes may be biased by these changes. These changes in the noise source properties mostly affect the direct reconstructed waves. In order to estimate this bias, one may thus, in a preliminary step, measure seismic velocity changes taking into account only measured travel-time shifts of the direct or early arrivals of the CCFs.

A weighted least-square regression (WLS) is subsequently computed for each $M + 1$ line of the matrix, through the satisfying points (in black in the data selection matrix on Fig. 4). The WLS is computed two times, first by allowing for a constant (equation 10a) and second by forcing the regression to cross the origin ([0,0]) (equation 10b). Errors on each parameter are computed, ea and em for a and m , respectively, in equation (10a) and em_0 for m_0 in equation (10b). The presence of

a significantly non-null constant a can reflect a possible instrument drift. The results, for a single day, can be plotted as a scatter plot of $\delta t/t$ versus pair number (Fig. 4, right two plots) or as a matrix (Fig. 5). The former also shows the slopes determined for the “ALL” line (red vertical line) and the weighted average value of the slopes determined for all pairs (green line). In this example, results are highly similar, but, even visually, some station pairs seem to behave differently. The matrix (Fig. 5) helps in identifying station groups, which could be analyzed together. As an example in the case of the UnderVolc project, one could define two groups: “crater stations” and “large interdistance stations”, for example,

$$\delta t = a + mt \quad (10a)$$

and

$$\delta t_0 = m_0 t, \quad (10b)$$

in which the subscript 0 is added to stress the [0,0] forcing.

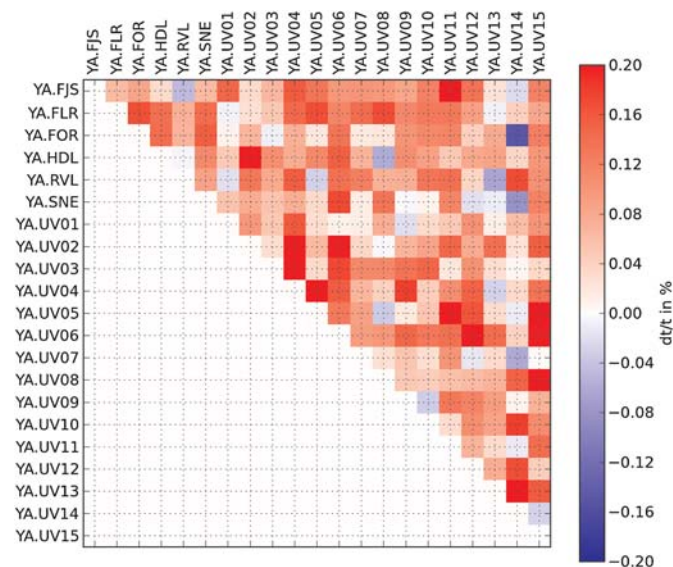
The six values (the slopes m and m_0 , the constant a , and their uncertainties) are saved to the disk, for each pair and for the “ALL” line, in order to be easily readable by plotting or exporting routines.

VALIDATION

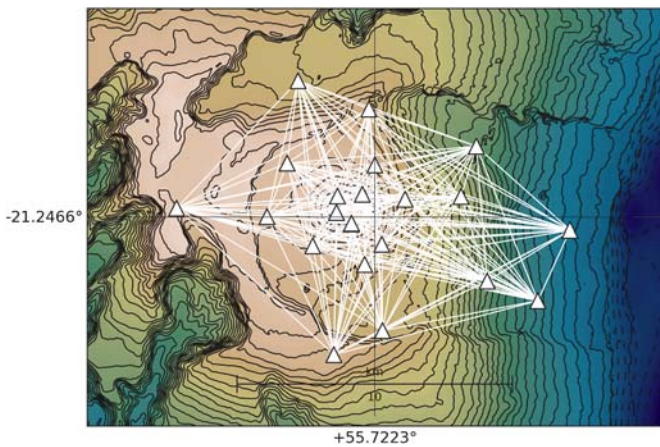
To validate our processing work flow, we tested it against the same data set that was used to successfully identify precursory velocity changes under the Piton de la Fournaise volcano for eruptions in October and December 2010 (Breuquier *et al.*, 2012) during the UnderVolc project.

Data and Parameters

The data set consists of 21 broadband three-component seismic stations placed on the Piton de la Fournaise volcano at various interdistances ranging from <1 to >10 km (Fig. 6). Most of the stations were installed by the beginning of November 2009 and were removed at the end of June 2011 (Fig. 2). In fact, most of the station locations were taken over by Observatoire Volcanologique du Piton de la Fournaise (OVPF) network, using new station codes. A total of 11,393 Z-component files have been processed in this test. For computing the CCF, the 100 Hz waveforms have been bandpassed between 0.01 and 8.0 Hz, then decimated to 20 Hz (factor 5). Only ZZ components have been calculated for different stations (no auto correlation). Each day was cut in 30 minute slices, and the resulting CCFs were saved to disk, together with the daily stack. Stacks have been calculated for 2, 5, 10, and 30 moving-window days. The MWCS frequency band is 0.2–0.85 Hz and is computed for 10 second sliding windows overlapping by 50%. For $\delta t/t$ calculations, both causal and acausal parts are used, lag times between ± 5 and ± 50 s are selected. Remaining selection parameters are 0.5 minimum coherence, 0.1 s maximum error, and 0.5 s maximum dt. Example timing on a 4 vCPU virtual machine and data volumes are given in Tables 1 and 2.



▲ **Figure 5.** Example matrix for the $\delta t/t$ values presented as a scatter plot on Figure 4. Each row or column is representing a unique station, so each cell a station pair. The color of the cell depends on the calculated $\delta t/t$ (the slope) for each pair. Here, the diagonal is empty because auto correlation was not computed. Thanks to the reciprocity principle, CCFs are only computed once per station pair.



▲ **Figure 6.** Station map of the UnderVolc project. Stations are the white triangles and paths between each station pair are drawn in white overlaid on a digital elevation model (DEM) of the Piton de la Fournaise volcano. The station map without the DEM is created automatically using the `plot_station_map.py` script.

Results

The $\delta v/v$ variations of the volcano during the UnderVolc project as obtained by applying the same methodology on the delay times calculated with the FORTRAN code of Clarke *et al.* (2011) and MSNoise show very similar results (Fig. 7). Slight differences below 0.02% are noted, whereas the signal of the eruptions is stronger than 0.05%. A general trend is present and denotes an apparent slight change in the velocity of 0.2% per year. This long-term component (LTV) *sensu* Brenguier, Shapiro, *et al.* (2008), can be removed to highlight shorter time-scale variations. Analyzing both codes in depth, we assume that the little differences might come from differences either in the way the cosine-smoothing function is constructed, or in the phase calculation code. We choose to use well-tested numerical routines included in the *numpy*, *scipy* (Jones *et al.*, 2001; Oliphant, 2006), and *statsmodels* (Seabold and Perktold, 2010) packages instead of rewriting Clarke's FORTRAN functions.

Table 1
Timing of the Different Steps of the Workflow

Step	Timing
Installation	10 minutes
Configuration	5 minutes
Scanning data archive	30 minutes
Identifying 102,626 new jobs	30 minutes
Processing cross-correlation jobs (1 filter)	21 hours
Stacking REFs for 210 pairs	33 minutes
Stacking 4 moving windows for all pairs	6 hours
Computing MWCS	8 hours
Computing $\delta t/t$	1 hour

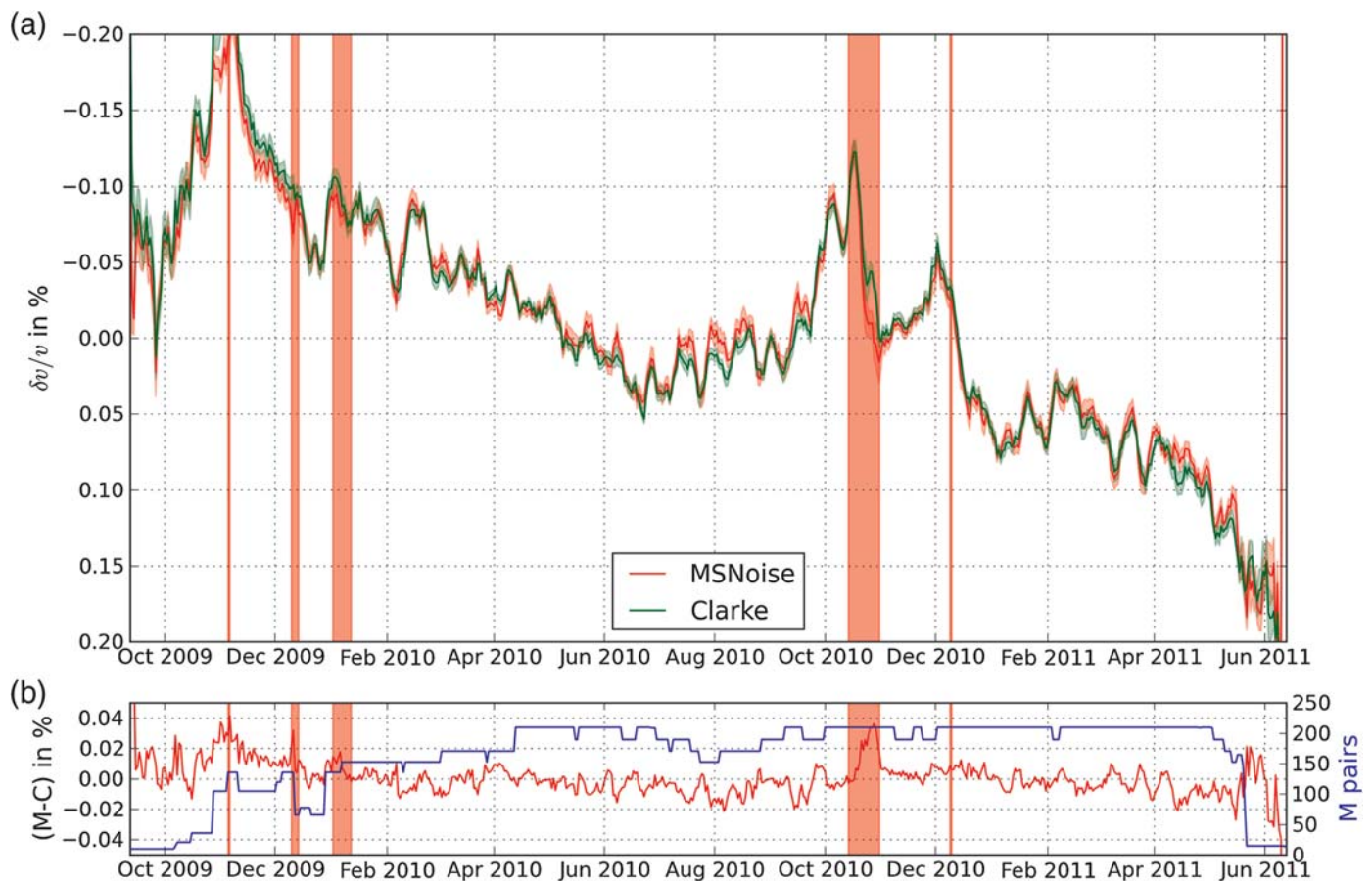
Table 2
Volume of the Archive and the Outputs of MSNoise

Element	Volume
UnderVolc Data archive	382 GB
1-day stack for 210 pairs, 1 filter	5.9 GB
REFs for 210 pairs	13 MB
Each moving-window stack for 210 pairs, 1 filter	6.0 GB

Eruptions of November 2009, January 2010, October 2011, and December 2011 are preceded by a decrease of the $\delta v/v$ (going up on Fig. 8). There seem to be no precursory signs of October 2009 eruption. We evidence the November 2009, October 2010, and December 2010 eruptions with a precursory time of about 15 days. For the last two eruptions, an increase of the $\delta v/v$ is noticed during the day before the actual eruption. This could be explained by a velocity increase at depth, by the stronger seismic activity during the hours preceding the eruption or by the presence of eruptive tremor in the minutes before. The last two are provoking a change in the noise source provenance and are disturbing the system. It is worth noting the decorrelation of the moving-stack function relative to the REF, five days before the eruption (Fig. 3). As Liu *et al.* (2010) have shown, a change in the stacking duration might highlight or erase signal in the evolution of the $\delta v/v$ curve. We show on Figure 3 the different decorrelation patterns when 2, 5, 10, or 30 days stacks are compared with the same REF. For example, precursory changes of $\delta v/v$ before the November 2009 eruption are visible for stack durations up to 10 days, but are completely erased when using the 30 day stack (Fig. 8). Short period (of less or equal to a few days) variations are visible on the 1, 2, and 5 day stacks (Fig. 8) and could have different origins, for example, linked with “local” causes like changes in the barometric pressure, rainfall, or seismic activity, or more “regional” ones like changes in the noise source positions. Although possible, the short period of a few days makes the “regional” causes less likely than the “local” ones.

CONCLUSIONS AND PERSPECTIVES

We present the first complete software package for computing and monitoring relative velocity variations using ambient seismic noise. MSNoise is a fully integrated solution that automatically scans data archives and determines which jobs need to be done whenever the scheduled task is executed. The whole package is written in Python, open source and freely available on <http://www.msnoise.org> (last accessed March 2014). The processing steps have been tested against codes used to successfully identify precursory velocity variations (Clarke *et al.*, 2011) under the Piton de La Fournaise volcano (La Réunion Island, France) during the UnderVolc project (Brenguier *et al.*, 2012). Using the same data set, results of both processing codes are highly similar. MSNoise is pluggable and extensible and is *de facto* a useful tool for researchers as they can only focus on



▲ **Figure 7.** (a) Relative seismic velocity variations in the Piton de la Fournaise volcano during the UnderVolc project, using D. Clarke’s FORTRAN code for MWCS (green) and using our implementation (red), eruptions are marked as red vertical lines. (b) The absolute difference in percent between the two results (red) and the number of station pairs available for each day of the project (blue).

implementing their processing code, while benefiting from its robust framework.

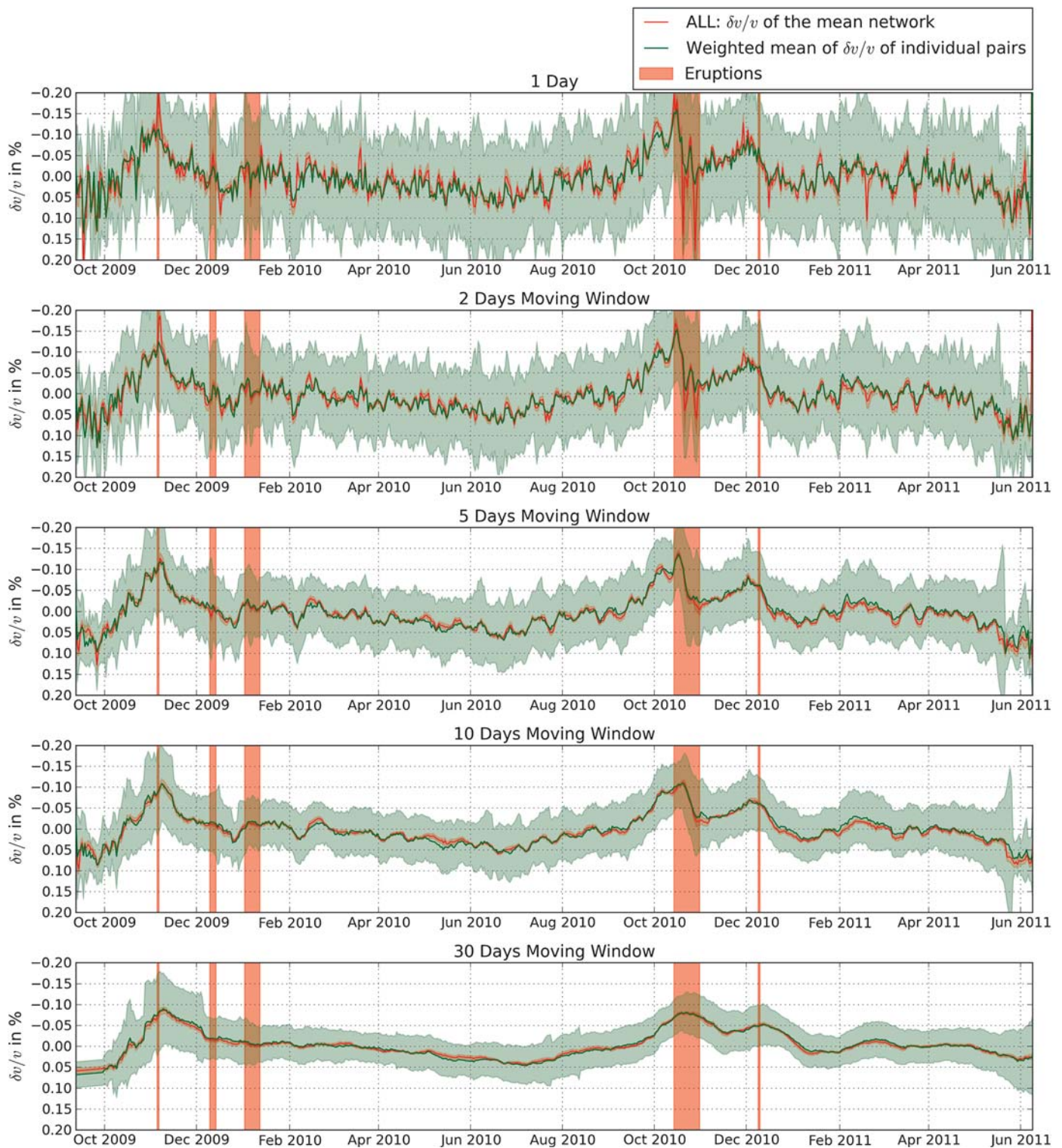
Our approach relies on the assumption that every wavepacket of the coda part of the reconstructed CCFs has experienced a uniform relative velocity change ($\delta v/v = \text{constant}$) along its travel path. This is a strong assumption that may not be valid in all cases. The consequence is that the measured relative seismic velocity change may strongly differ from “real” $\delta v/v$ ’s within the studied medium. A main limitation arises from the fact that the coda wavepackets may travel areas of nonuniform velocity changes. In that case, the measured $\delta v/v$ is an average of the real $\delta v/v$, and one may, for example, misinterpret a change of δv as an increase of $\delta v/v$ whereas it could simply be associated with a spatial extent of a constant value $\delta v/v$.

Finally, although we present the validation on MSNoise in a volcanic environment, its application should not be restricted to this specific case, and we are looking forward to seeing applications applied to different contexts and to receiving comments and contributions from other research groups. Contributions could be, for example, a new correlation algorithm, a new $\delta t/t$ calculation technique, the implementation of attenuation computation (like in Prieto *et al.*, 2011), the usage of the delay

matrices to locate variations of $\delta v/v$, or decorrelation at different time lags (like in Larose *et al.*, 2010; Obermann *et al.*, 2013). Other contributions could include the analysis of much higher-frequency data (higher than 1 kHz such as in Olivier *et al.*, 2012), or the usage of the CCF as input to a frequency-time analysis (FTAN) analysis (as described in Levshin *et al.*, 1989) for computing dispersion and subsequent inversion of 1D surface-wave velocity profiles. We are planning to implement the possibility to define multiple references in the same way as the multiple stacks and the possibility to define several “station groupings” similar to the “ALL” (see [Velocity Variations Computation](#) section). ☒

ACKNOWLEDGMENTS

Corentin Caudron’s Ph.D. work is supported by the Belspo (Action 2 Grant WI/33/J02). Thomas Lecocq’s 1 month stay at the Observatoire Volcanologique du Piton de la Fournaise (OVPF) has been partly supported by the Institut du Physique du Globe (IPGP). The data used for the analysis were collected by the IPGP/OVPF, and the Institut des Sciences de la Terre (ISTerre) within the framework of the ANR_08_RISK_011/UnderVolc project. The sensors are property of the French



▲ **Figure 8.** Comparison of the detrended $\delta v/v$ results obtained using MSNoise for five different moving-window stacks (1, 2, 5, 10, and 30 days) without forcing the slope estimate to cross the origin (equation 10) in the $\delta t/t$ estimation process. Eruptions of the Piton de la Fournaise are shown as red vertical bands. The $\delta v/v$ of the whole network, the “ALL” line (red), matches the weighted mean of all individual pairs (green). Four out of the five eruptions that occurred during the UnderVolc project are preceded by precursory changes in $\delta v/v$ (see text).

transportable seismic network, Sismob (INSU-CNRS). Thierry Camelbeeck (Royal Observatory of Belgium [ROB]) is acknowledged for his continuous support of this project and Daniel Clarke for providing his moving-window cross spectrum (MWCS) codes in order to compare and validate our own. OVPF staff is thanked for their help throughout the UnderVolc project and for their warm welcome during Thomas' visit in 2011. Josef Perktold, of *statsmodels*, is acknowledged for his kind help on the usage of the weighted least-square regression. Christoph Gohlke from the Laboratory for Fluorescence Dynamics, University of California (Irvine) is acknowledged for providing an extensive list of prebuilt Python packages for Windows on his website (<http://www.lfd.uci.edu/~gohlke/pythonlibs/>; last accessed March 2014). Finally thanks to all the alpha/beta users of MSNoise for providing inestimable feedback!

The editor-in-chief, associate editor, and Reviewer Weston Thelen are acknowledged for their constructive comments.

REFERENCES

- Anggono, T., T. Nishimura, H. Sato, H. Ueda, and M. Ukawa (2012). Spatio-temporal changes in seismic velocity associated with the 2000 activity of Miyakejima volcano as inferred from cross-correlation analyses of ambient noise, *J. Volcanol. Geoth. Res.* **247/248**, 93–107, doi: [10.1016/j.jvolgeores.2012.08.001](https://doi.org/10.1016/j.jvolgeores.2012.08.001).
- Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.* **81**, no. 3, 530–533, doi: [10.1785/gssrl.81.3.530](https://doi.org/10.1785/gssrl.81.3.530).
- Brenguier, F., M. Campillo, C. Hadziioannou, N. M. Shapiro, R. M. Nadeau, and E. Larose (2008). Postseismic relaxation along the San Andreas fault at Parkfield from continuous seismological observations, *Science* **321**, no. 5895, 1478–1481, doi: [10.1126/science.1160943](https://doi.org/10.1126/science.1160943).
- Brenguier, F., D. Clarke, Y. Aoki, N. M. Shapiro, M. Campillo, and V. Ferrazzini (2011). Monitoring volcanoes using seismic noise correlations, *Comptes Rendus Geosci.* **343**, nos. 8/9, 633–638, doi: [10.1016/j.crte.2010.12.010](https://doi.org/10.1016/j.crte.2010.12.010).
- Brenguier, F., P. Kowalski, T. Staudacher, V. Ferrazzini, F. Lauret, P. Boissier, P. Catherine, A. Lemarchand, C. Pequegnat, O. Meric, C. Pardo, A. Peltier, S. Tait, N. M. Shapiro, M. Campillo, and A. Di Muro (2012). First results from the UnderVolc high resolution seismic and GPS network deployed on Piton de la Fournaise Volcano, *Seismol. Res. Lett.* **83**, no. 1, 97–102, doi: [10.1785/gssrl.83.1.97](https://doi.org/10.1785/gssrl.83.1.97).
- Brenguier, F., N. Shapiro, M. Campillo, V. Ferrazzini, Z. Duputel, O. Coutant, and A. Nercessian (2008). Towards forecasting volcanic eruptions using seismic noise, *Nature Geosci.* **1**, no. 2, 126.
- Campillo, M., and A. Paul (2003). Long-range correlations in the diffuse seismic coda, *Science* **299**, no. 5606, 547–549, doi: [10.1126/science.1078551](https://doi.org/10.1126/science.1078551).
- Clarke, D., L. Zaccarelli, N. M. Shapiro, and F. Brenguier (2011). Assessment of resolution and accuracy of the Moving Window Cross Spectral technique for monitoring crustal temporal variations using ambient seismic noise, *Geophys. J. Int.* **186**, no. 2, 867–882, doi: [10.1111/j.1365-246X.2011.05074.x](https://doi.org/10.1111/j.1365-246X.2011.05074.x).
- de Castro Lopo, E. (2013). *Secret Rabbit Code (aka libsamplerate)*, <http://www.mega-nerd.com/SRC/> (last accessed March 2014).
- Duputel, Z., V. Ferrazzini, F. Brenguier, N. Shapiro, M. Campillo, and A. Nercessian (2009). Real time monitoring of relative velocity changes using ambient seismic noise at the Piton de la Fournaise volcano (La Réunion) from January 2006 to June 2007, *J. Volcanol. Geoth. Res.* **184**, no. 1/2, 164–173, doi: [10.1016/j.jvolgeores.2008.11.024](https://doi.org/10.1016/j.jvolgeores.2008.11.024).
- Enthought (2008). *The Enthought Python Distribution*, <http://code.enthought.com> (last accessed March 2014).
- Goldstein, P., J. A. Snoke, M. Firpo, and L. Minner (2003). Signal processing and analysis tools for seismologists and engineers, in *The LASPEI International Handbook of Earthquake and Engineering Seismology*, W. H. K. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger (Editors), Academic press, London, 2000 pp.
- Hadziioannou, C., E. Larose, O. Coutant, P. Roux, and M. Campillo (2009). Stability of monitoring weak changes in multiply scattering media with ambient noise correlation: Laboratory experiments, *J. Acoust. Soc. Am.* **125**, no. 6, 3688–3695.
- Herrmann, R. (2002). *Computer Programs in Seismology 3.3*, <http://www.eas.slu.edu/eqc/eqccps.html> (last accessed March 2014).
- Hunter, J. (2007). Matplotlib: A 2D graphics environment, *Comput. Sci. Eng.* **9**, no. 3, 90–95, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- Jones, E., T. Oliphant, P. Peterson, and SciPy development team (2001). *SciPy: Open source scientific tools for Python*, <http://www.scipy.org/> (last accessed March 2014).
- Larose, E., T. Planes, V. Rossetto, and L. Margerin (2010). Locating a small change in a multiple scattering environment, *Appl. Phys. Lett.* **96**, no. 20, 204101–204101-3.
- Levshin, A. L., T. Yanovskaya, A. Lander, B. Bukshin, M. P. Barmin, L. Ratnikova, and E. Its (1989). *Seismic Surface Waves in a Laterally Inhomogeneous Earth*, V. I. Keilis-Borok (Editor), Kluwer Academic Publishers, Norwell, Massachusetts, 293 pp.
- Liu, Z., J. Huang, and J. Li (2010). Comparison of four techniques for estimating temporal change of seismic velocity with passive image interferometry, *Earthq. Sci.* **23**, no. 5, 511–518, doi: [10.1007/s11589-010-0749-z](https://doi.org/10.1007/s11589-010-0749-z).
- McKinney, W. (2012). *Python for Data Analysis*, O'Reilly Media, Cambridge, 470 pp.
- Megies, T., M. Beyreuther, R. Barsch, L. Krischer, and J. Wassermann (2011). ObsPy—What can it do for data centers and observatories? *Ann. Geophys.* **54**, no. 1, 47–58, doi: [10.4401/ag-4838](https://doi.org/10.4401/ag-4838).
- Mordret, A., A. Jolly, Z. Duputel, and N. Fournier (2010). Monitoring of phreatic eruptions using interferometry on retrieved cross-correlation function from ambient seismic noise: Results from Mt. Ruapehu, New Zealand, *J. Volcanol. Geoth. Res.* **191**, nos. 1/2, 46–59, doi: [10.1016/j.jvolgeores.2010.01.010](https://doi.org/10.1016/j.jvolgeores.2010.01.010).
- Obermann, A., T. Planes, E. Larose, C. Sens-Schonfelder, and M. Campillo (2013). Depth sensitivity of seismic coda waves to velocity perturbations in an elastic heterogeneous medium, *Geophys. J. Int.* **194**, no. 1, 372–382, doi: [10.1093/gji/ggt043](https://doi.org/10.1093/gji/ggt043).
- Oliphant, T. E. (2006). *Guide to NumPy*, Brigham Young University, Provo, Utah, 371 pp.
- Olivier, G., F. Brenguier, M. Campillo, R. Lynch, and N. Shapiro (2012). Monitoring of velocity variations in an underground mine using ambient seismic noise, in *Abstract Book of AGU Fall Meeting 2012*.
- Paul, A. (2005). Empirical synthesis of time-asymmetrical green functions from the correlation of coda waves, *J. Geophys. Res.* **110**, no. B8, doi: [10.1029/2004JB003521](https://doi.org/10.1029/2004JB003521).
- Poupinet, G., W. L. Ellsworth, and J. Frechet (2011). Monitoring velocity variations in the Crust using earthquake doublets: An application to the Calaveras fault, California, *J. Geophys. Res.* **89**, no. B7, 5719–5731.
- Prieto, G. A., M. Denolle, J. F. Lawrence, and G. C. Beroza (2011). On amplitude information carried by the ambient seismic field, *Comptes Rendus Geosci.* **343**, nos. 8/9, 600–614, doi: [10.1016/j.crte.2011.03.006](https://doi.org/10.1016/j.crte.2011.03.006).
- Ratdomopurbo, A., and G. Poupinet (1995). Monitoring a temporal change of seismic velocity in a volcano: Application to the 1992 eruption of Mt. Merapi (Indonesia), *Geophys. Res. Lett.* **22**, no. 7, 775–778.
- Seabold, J., and J. Perktold (2010). Statsmodels: Econometric and statistical modeling with python, in S. van der Walt and J. Millman (Editors), *Proc. of the 9th Python in Science Conference*, Austin, Texas, <http://conference.scipy.org/proceedings/scipy2010/> (last accessed March 2014).

- Sens-Schönfelder, C. (2008). Synchronizing seismic networks with ambient noise, *Geophys. J. Int.* **174**, no. 3, 966–970.
- Sens-Schönfelder, C., and U. Wegler (2006). Passive image interferometry and seasonal variations of seismic velocities at Merapi volcano (Indonesia), *Geophys. Res. Lett.* **33**, L21302, doi: [10.1029/2006GL027797](https://doi.org/10.1029/2006GL027797).
- Sens-Schönfelder, C., and U. Wegler (2011). Passive image interferometry for monitoring crustal changes with ambient seismic noise, *Comptes Rendus Geosci.* **343**, nos. 8/9, 639–651, doi: [10.1016/j.crte.2011.02.005](https://doi.org/10.1016/j.crte.2011.02.005).
- Stehly, L., M. Campillo, and N. M. Shapiro (2006). A study of the seismic noise from its long-range correlation properties, *J. Geophys. Res.* **111**, no. B10, doi: [10.1029/2005JB004237](https://doi.org/10.1029/2005JB004237).
- Stehly, L., M. Campillo, and N. M. Shapiro (2007). Traveltime measurements from noise correlation: Stability and detection of instrumental time-shifts, *Geophys. J. Int.* **171**, no. 1, 223–230, doi: [10.1111/j.1365-246X.2007.03492.x](https://doi.org/10.1111/j.1365-246X.2007.03492.x).
- Tukey, J. W. (1962). The future of data analysis, *Ann. Math. Stat.* **33**, no. 1, 1–67, doi: [10.1214/aoms/1177704711](https://doi.org/10.1214/aoms/1177704711).
- Wegler, U., and C. Sens-Schönfelder (2007). Fault zone monitoring with passive image interferometry, *Geophys. J. Int.* **168**, no. 3, 1029–1033, doi: [10.1111/j.1365-246X.2006.03284.x](https://doi.org/10.1111/j.1365-246X.2006.03284.x).

*Thomas Lecocq
Corentin Caudron¹
Seismology-Gravimetry
Royal Observatory of Belgium
Avenue Circulaire 3
B1180 Brussels, Belgium
thomas.lecocq@seismology.be*

*Florent Brenguier²
ISTerre University of Grenoble
Box 53
38041 Grenoble CEDEX 9
France*

¹ Also at DSTE, Université Libre de Bruxelles, Belgium; now at the Earth Observatory of Singapore.

² Also at Observatoires Volcanologiques et Sismologiques, IPGP, France.