Geophysical Research Letters

Supporting Information for

"Long-term monitoring of long-period seismicity and space-based SO_2 observations at African lava lake volcanoes Nyiragongo and Nyamulagira (DR Congo)"

Julien BARRIERE[1,2,*], Adrien OTH[2], Nicolas THEYS[3], Nicolas d'OREYE[1,2], and François KERVYN[4]

[1] National Museum of Natural History (NMHN), Walferdange, Luxembourg
[2] European Center for Geodynamics and Seismology (ECGS), Walferdange, Luxembourg
[3] Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Uccle, Belgium
[4] Royal Museum for Central Africa (RMCA), Tervuren, Belgium

Contents of this file

Uploaded separately

Datasets S1-S10

Introduction

This supporting information provides additional text and illustrations about the data and methods. 10 additional data sets used for generating figures in the main manuscript are provided.

S1- Power spectral density analysis and computation of NCFs

Power spectral density (PSD) analysis revealed quite low noise levels at all KivuSNet stations in the microseismic band [0.1-0.3] Hz, probably reflecting the fact that the region is located well in Africa's continental interior [*Oth et al*., 2017]. More interestingly, in the frequency band [0.3-0.9] Hz, a systematic increase of PSD can be observed at all stations (example spectrograms in Figure S2), which may originate from a continuous tremor activity at one or both volcanoes. Following these observations, we compute the NCFs in the frequency band 0.1-0.9 Hz (Figure S3), typically used in ambient seismic noise studies at volcanoes for the retrieval of Green's functions [e.g., *Brenguier et al*., 2016]. Indeed, in order to fulfill (at least partially) some important requirements of the noise wave field properties (e.g., diffusive) to retrieve inter-station Green's functions, the frequency band is typically chosen below 1 Hz, where oceanic microseism noise sources tend to dominate the wavefield (e.g., *Brenguier et al.,* 2008*).* The upper frequency limit may be extended to 1.5 or 2 Hz but a lower limit of 0.9 Hz prevents the potentially contaminating influence of anthropogenic noise sources that are significant, even between 1 to 5 Hz, at most of the KivuSNet stations [*Oth et al*., 2017].

The pre-processing of the raw data prior to the NCFs calculation step is based on the data processing scheme of *Bensen et al.* (2007). The chosen processing criteria are nearly identical to the ones employed by *Ballmer et al.* (2013) for locating tremor sources in Hawaii. 24-hours raw continuous records are synchronized between all stations, detrended, band-pass filtered between 0.05 to 0.95 Hz and downsampled from 50 to 10 Hz. Temporal normalization is then applied using running-absolute-mean normalization with 10 s long data intervals and spectral equalization is finally performed between 0.1 and 0.9 Hz (Figure S3). For two months of high quality continuous seismic records from up to 11 stations (February-March 2016), the NCFs displayed in Figure S3 reveal two dominant frequency patterns. Strong asymmetric arrivals in the microseism band [0.1-0.3] Hz indicate a significant contribution from Lake Kivu, one of the African Great Lakes, as source in the south of the Virunga Volcanic Province (VVP). The tremor-like signals are clearly observed in the interval 0.3-0.9 Hz, as previously inferred from the PSD analysis (see also main text).

S2- Daily tremor location (network response)

We apply a source scanning procedure in the frequency band 0.3-0.9 Hz similarly to *Ballmer et al*. (2013) and *Droznin et al*. (2015) (see also main text), in order to get a final location map of tremor, the so-called 2-D network response R (Figure 1), which can be expressed as follows [*Droznin et al.*, 2015]:

$$R(r)= \sum_{i=1}^{N} \sum_{j=i+1}^{N} S^{i,j} \; [t^i(r) - t^j(r)],$$

where N is the total number of stations, $S^{i,j}$ is the NCF envelope for the station pair i,j and $t^i(r)$ respectively $t^j(r)$ the predicted travel-times between the tested source position r and stations i and j. The network response R(r) is then scaled into the interval [0-1] by applying a min-max normalization. A moving average of +/- 10s was applied to the NCFs prior to the location procedure, allowing to take into account uncertainties associated with the simplified travel-time model and scattering effects [*Ballmer et al*., 2013].

S3- Location procedure for LP events

The location procedure for LP events is close to the one employed by *Poiata et al.* (2016) based on station-pair time-delay estimate (TDE) functions. The TDE-based method consists in projecting and stacking the cross-correlation functions onto each 3D differential travel-time grid calculated once for each station-pair. Using a simple 1D velocity model (Figure S5) with essentially one interface at depths of interest [obtained from a receiver function analysis at two earlier broadband stations in the VVP, *Mavonga et al*., 2010], 3D travel-time grids are calculated using the Grid2Time routine of the NonLinLoc program [*Lomax et al*., 2009], based on the Eikonal finite-difference scheme of *Podvin and Lecomte* [1991].

First a traditional STA/LTA detection algorithm is applied to the continuous data. Considering that the vast majority of events in the area will be of low SNR, we set the minimum threshold to the rather low value of 3. A potential event to be located is declared when this trigger ratio is exceeded at least at three stations within a 10 s time interval. The location procedure is carried out in several frequency bands in order to accurately locate the LP events. In our case, the most reliable frequency band was determined to be 0.5-2 Hz. While the P phase is undetectable for most events, we highlight the S phase by computing envelope functions of filtered signals at each available station. The steps of this signal processing procedure are the following:

1. Normalization of traces E_1 and N_1: The East (E) and North (N) components are filtered in the desired frequency band and normalized.
2. Calculation of envelopes $Ef_1^{E,N}$: A centered running maximum (+/- 0.5s) is applied to squared Hilbert envelopes of traces E_1 and N_1 followed by a smoothing operator (centered moving average of +/- n seconds, here n=2).

3. Calculation of $Ef_2^{E,N}$: In order to make the envelopes more impulsive around the S arrival, the envelopes $Ef_1^{E,N}$ are multiplied by their respective leading gradient as follows:

$$Ef_2^{E,N}(t) = Ef_1^{E,N}(t) * [Ef_1^{E,N}(t+n) - Ef_1^{E,N}(t)].$$

4. The final envelope Ef_3 is obtained by summing Ef_2^E and Ef_2^N and setting negative values to zero:

$$Ef_3(t) = a * [Ef_2^E(t) + Ef_2^N],$$

$$where \begin{cases} a = 0, [Ef_2^E(t) + Ef_2^N] < 0. \\ a = 1, [Ef_2^E(t) + Ef_2^N] > 0. \end{cases}$$

Ef_3 represents a bell-shaped envelope mostly time-limited between the S-phase onset and the first maximum (Figure S4). While more standard envelopes such as $Ef_1=[Ef_1^E+Ef_1^N]$ are strongly affected by the wave train characteristics, the shape of Ef_3 is a suitable property to impulsive cross-correlation functions between station pairs, i.e. more precise time-lag estimates. Depending on the SNR and the sharpness of the S phase, secondary lobes may be generated and will result in a smaller contribution to the final stacked cross-correlation functions, thus acting as a damping factor for less accurate S phase envelopes. If no distinct dominant peak can be retrieved, which means that the S phase is mostly indistinguishable from the noise, the considered station is disregarded. Source scanning is performed for all available station-pairs and a coherent source region is found where the stacked TDE functions add constructively (see Figure S9).

S4- Template matching of LP events and tremor

The template consists in 20s long Hilbert envelopes for vertical and horizontal components (Figure S7). Since the records at RGB exhibit generally a broader frequency content than the more distant stations for such LP event, we enlarge the frequency band up to 4 Hz, the lower frequency bound remaining at 0.5 Hz. To ensure a reliable detection without being too restrictive, the minimum level of correlation between the 3-components template and the continuous data is set to 1.8 (i.e., 0.6 on average per channel).

At this stage it is important to point out that the discrimination between shallow volcano-tectonic earthquakes (VT) and LP events, which are originating from two distinct source processes, can be ambiguous, notably if they are recorded at relatively distant stations (approx. 20 km at best for RGB, the closest to Nyamulagira). We therefore performed a standard frequency-domain analysis of the master event (Figure S9A) at several stations in order to verify that the dominant low-frequency content of the signal is more

suggestive of a source process rather than path effects attenuating the high-frequencies of a VT event (Figure S8).

Finally, we tracked the variability of the tremor signature in the VVP between April 2014 and February 2017 via the correlation of the daily NCFs with a 6-months stacked reference NCF at the station pair GOM-RGB. This 6-months period has been chosen between 26 May and 25 November 2016 because it is associated with constant daily locations of tremor at Nyiragongo and the absence of any surface activity at Nyamulagira, thus being representative of a dominant and persistent tremor source at Nyiragongo (see section 3. *Results and Discussion*).

S5- LP detection capability

Because tremor and LP events overlap in their dominant frequency content, the detection of LP events during periods of high tremor activity could in principle be hampered and thus lead to a misrepresentation of the occurrence rate of LP events. In order to assess whether this might be the case, we compare two distributions of peak LP event amplitudes averaged over the three components (Fig. S10A,B) during high and low tremor activity determined on the basis of the RSAM level (Fig. S10C). The relationship "count of events – peak amplitude" (Fig. S10A,B) allows to determine an approximate amplitude threshold below which the template matching procedure is most likely not able to detect all LP events due to a poor SNR. A threshold $A_T = 30x10^{-9}$ m is determined for the high tremor activity period while it is slightly smaller during low tremor activity (approx. $17x10^{-9}$ m). While both periods include a similar number of days (414 and 367 for high and low activity, respectively), 1,717 more events with peak amplitudes below A_T are detected during low tremor activity as a potential result of the better SNR.

In order to assess if some variations in the occurrence rates of LP events (see Figure 2) could be due to a detection capability difference, we plot the cumulative sum of events once only considering events with peak amplitudes above the determined threshold A_T and once without applying this selection for the period April 2015-September 2016 (Fig. S10D). Between April 2014 and April 2016, a two-years period including high and low tremor activity, the removal of small amplitude events (approx. 19% of all events) does not affect the relative features of the occurrence rate of LP events (same shape of cumulative curve). However, significantly more weak amplitude events (approx. 40%) are detected since the disappearance of surface activity at Nyamulagira in April 2016, coinciding with the end of lava lake activity. This difference is therefore more likely due to real physical source properties rather than a simple effect of detection failure. These observations imply that the detection capability of LP events is, if at all, only weakly dependent on the level of tremor activity.

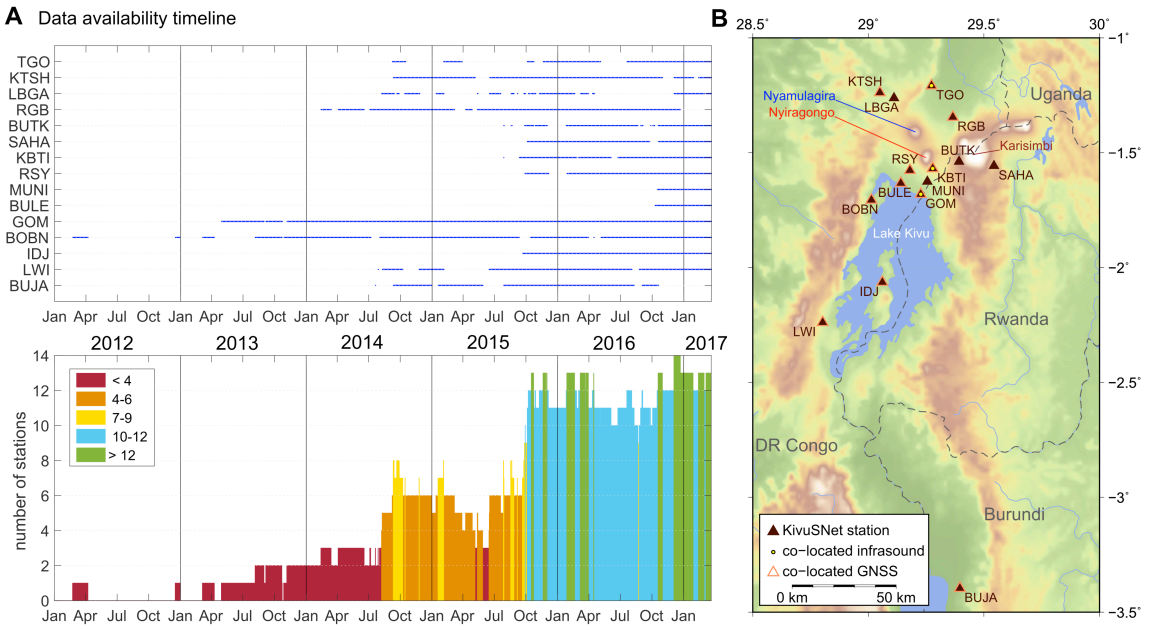

Figure S1: **A)** KivuSNet data availability timeline since the installation of AfricaArray station BOBN (2012). Real-time transmission based on cellular network was installed in October 2015, allowing to continuously keep operational a network of 7 to 11 stations with near-real time capabilities since then. **B)** Map of the VVP-North Kivu region including the KivuSNet seismic stations (black triangles). The station GOM is located in the city of Goma (DR Congo), separated from the neighboring town of Rubavu (Rwanda) by the boundary line.

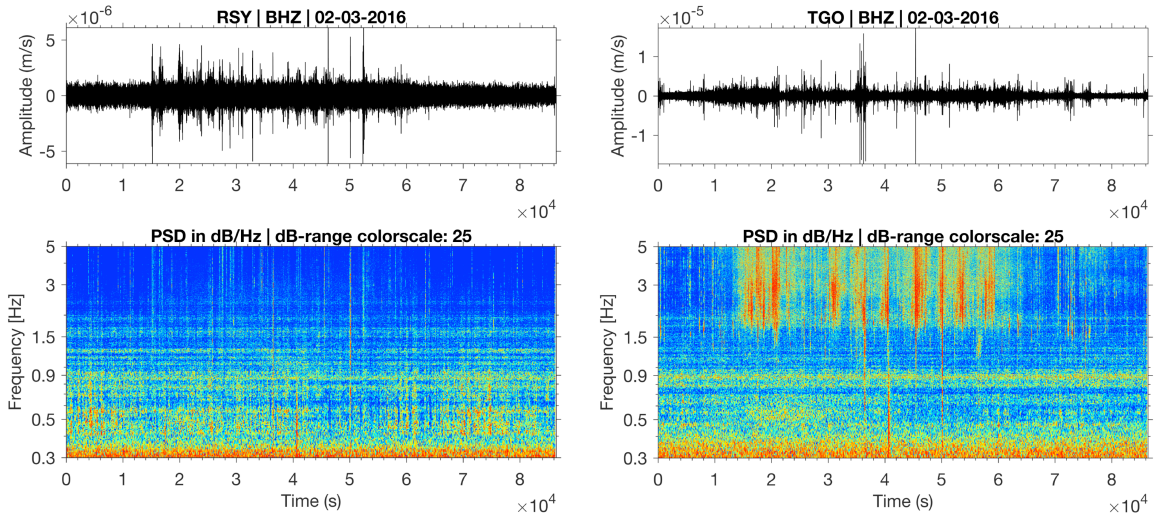

Figure S2: Raw seismograms and power spectral densities (PSD) for the same day (3 Feb. 2016) exemplified at stations RSY and TGO between 0.3 and 5 Hz. Continuous tremor patterns are clearly visible between 0.3 to 0.9 Hz. High amplitudes of PSD near 0.3 Hz correspond to the edge of the microseism band.

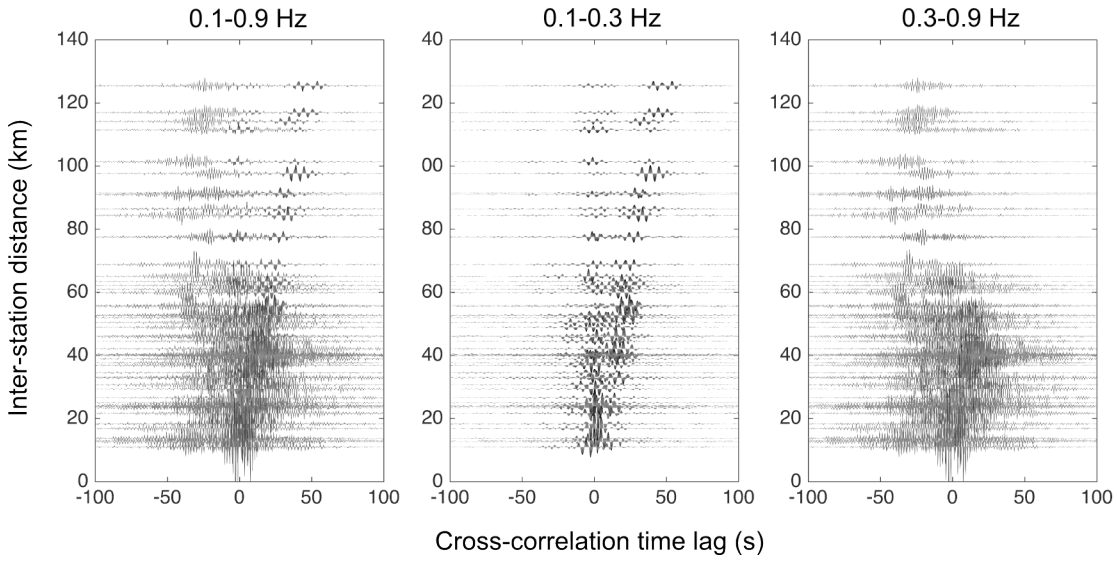

Figure S3: Daily NCFs stacked for a period of two months with excellent data availability (February-March 2016, 9 to 11 stations) sorted by inter-station distance for three different frequency bands: 0.1-0.9 Hz**,** 0.1-0.3 Hz and 0.3-0.9 Hz. The most southern station of each station pair is always taken as reference for the cross-correlation calculation (noise coming from the south would contribute to the positive components).

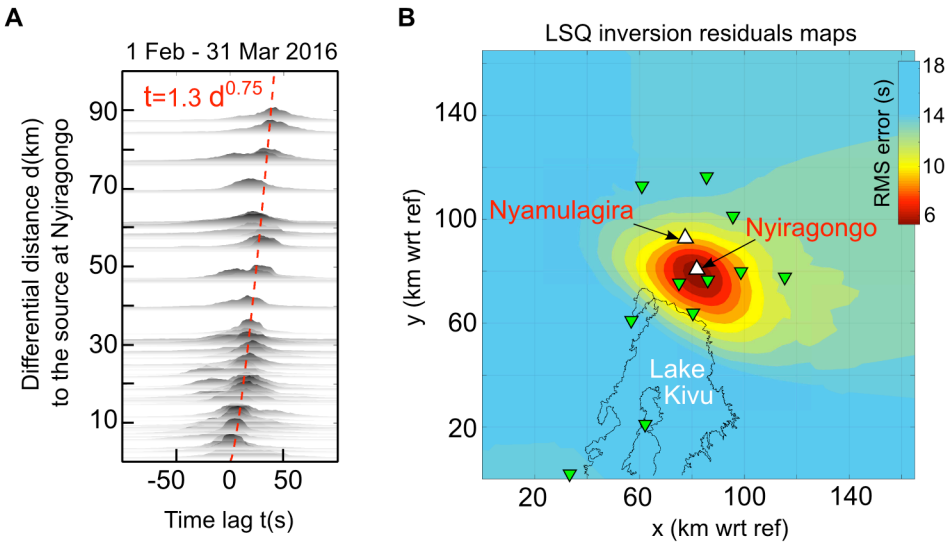

Figure S4: (**A**) Best-fitting travel-time curve using stacked NCFs for a period of two months with excellent data availability and with NCFs having the same characteristics (1 February to 31 March 2016, 9 to 11 stations) and (**B**) corresponding least-square inversion residuals map. Delay times are picked at times corresponding to the absolute maxima of NCF envelopes, which are smoothed by a +/- 10s centered moving average window.

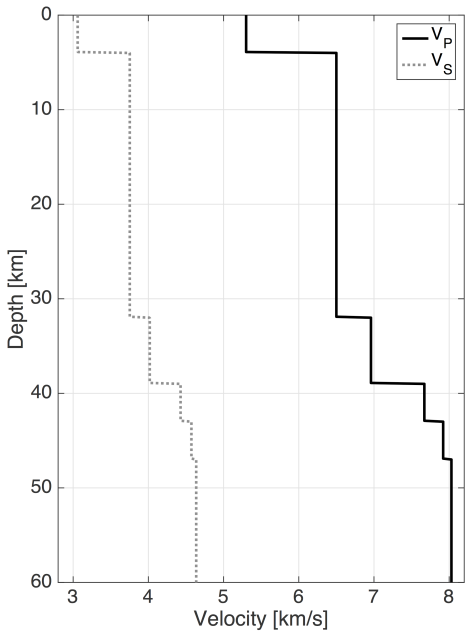

Figure S5: 1D velocity model for the VVP region [Mavonga et al., 2010] used for the location of LP event (see Figure S9).

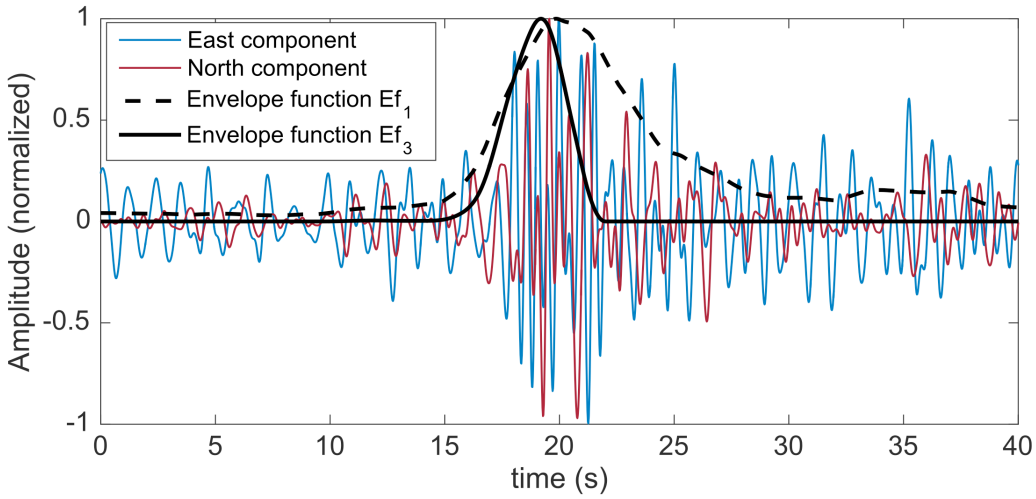

Figure S6: Example of horizontal seismic traces for one LP event recorded at station RGB. The solid black line represents the final envelope Ef_3 used as input into the cross-correlation location procedure. Its shape better captures the S-phase onset than the more standard envelope $Ef_1=[Ef_1^E+Ef_1^N]$ (dashed black line).

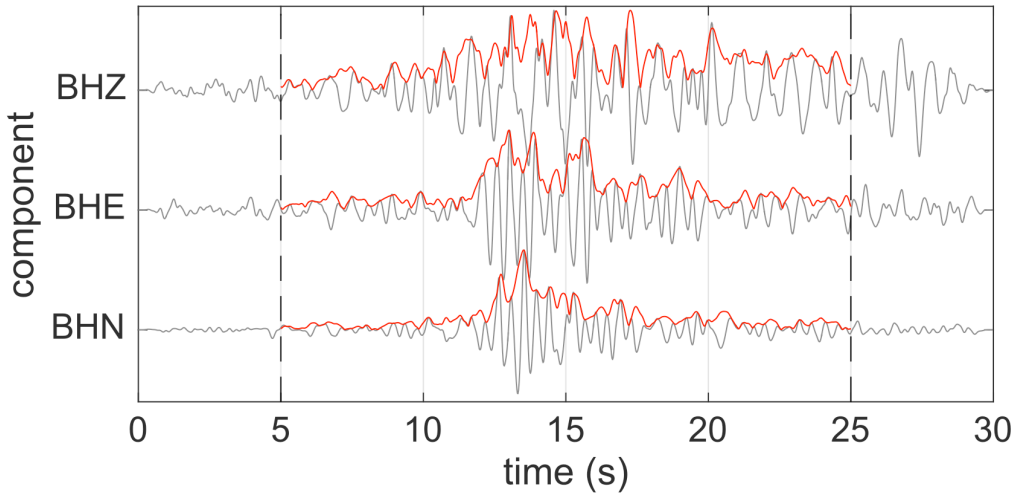

Figure S7: 3-component template for station RGB (Hilbert envelopes in red) associated with the master event extracted from the cluster of LP events at Nyamulagira (see Figure S9A).

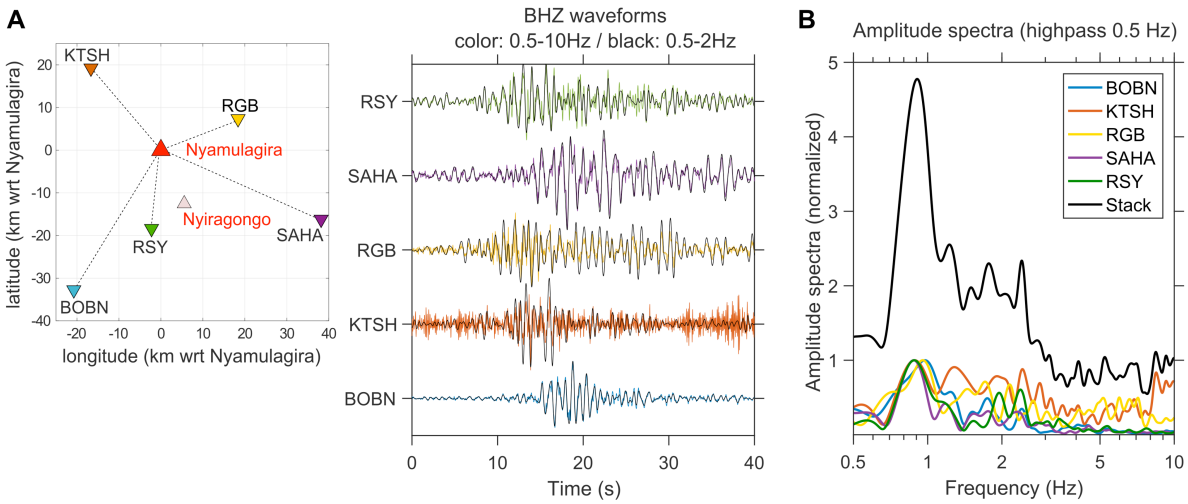

Figure S8: A) Vertical component seismograms for the master LP event, located at Nyamulagira (see Figure S9A), for 5 stations with different azimuths and distance to the source (see map on the left) and **B)** their respective amplitude spectra in the frequency range [0.5 - 10] Hz. A clear peak around 0.9 Hz common to all stations is evidenced by stacking each individual spectrum.

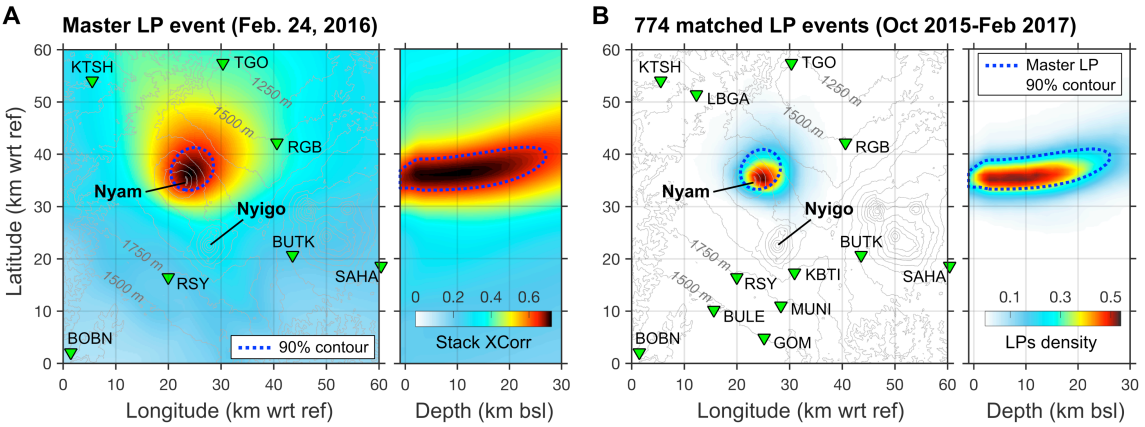

Figure S9: **A)** Maximum-likelihood location (i.e., stack of inter-station time delay estimates obtained by cross-correlation) of the selected master LP event and B) location density maps of 774 events belonging to the master event's family obtained by template matching between October 2015 and February 2017. In **A**), not all stations are displayed since KBTI and LBGA stations were not operational at that time, BULE and MUNI were not installed while records at GOM are too noisy.

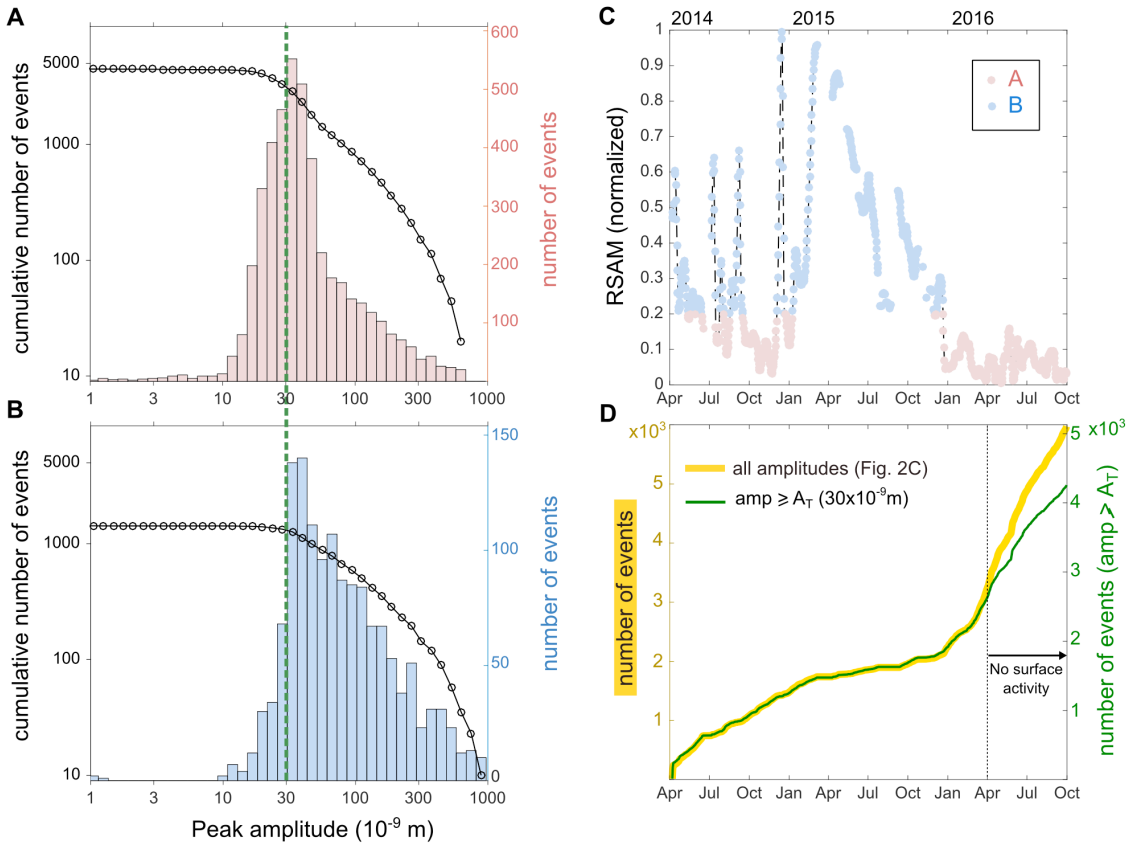

Figure S10: (**A**) and (**B**) Amplitude (colored histograms) and cumulative number of amplitude (black lines and markers) distributions during low and high tremor activity, respectively, determined on the basis of (**C**) normalized RSAM levels. The corresponding time periods for A are displayed in blue, while those for B are shown in red. (**D**) Cumulative number of events, including all events (yellow line, left y-axis) or excluding the smallest events characterized by peak amplitudes below $A_T = 30x10^{-9}$m (green line, right y-axis). The amplitude threshold A_T is indicated as a dashed green line in (A) and (B).

References

Bensen, G. D., M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro and Y. Yang (2007), Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements*, Geophys. J. Int., 169,* 1239–1260*,* doi:10.1111/j.1365-246X.2007.03374.x

Lomax, A., Michelini, A. and Curtis, A. (2009), Earthquake Location, Direct, Global-Search Methods, in *Complexity In Encyclopedia of Complexity and System Science*, pp. 2449-2473, Springer, New York.

Podvin, P. and I. Lecomte (1991), Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools, *Geophys. J. Int., 105*, 271-284, doi:10.1111/j.1365-246X.1991.tb03461.x

Captions for Datasets S1-S10

Dataset S1: Location map for SO2 plume density for August 2016 (matrix 101x101, rows (latitude)=-2:-1° ; columns (longitude) =28.65:29.65° ; step=0.01°).

Dataset S2: Location map for SO2 plume density for February 2016 (matrix 101x101, rows (latitude)=-2:-1° ; columns (longitude) =28.65:29.65° ; step=0.01°).

Dataset S3: Location map for SO2 plume density for January 2015 (matrix 101x101, rows (latitude)=-2:-1° ; columns (longitude) =28.65:29.65° ; step=0.01°).

Dataset S4: Location map for tremor (2D network response) for August 2016 (matrix 101x101, rows (latitude)=-2:-1° ; columns (longitude) =28.65:29.65° ; step=0.01°).

Dataset S5: Location map for tremor (2D network response) for February 2016 (matrix 101x101, rows (latitude)=-2:-1° ; columns (longitude) =28.65:29.65° ; step=0.01°).

Dataset S6: Location map for tremor (2D network response) for January 2015 (matrix 101x101, rows (latitude)=-2:-1° ; columns (longitude) =28.65:29.65° ; step=0.01°).

Dataset S7: Time series (2 columns date and values) of daily LP occurrence rate

Dataset S8: Time series (2 columns date and values) of RSAM at station RGB

Dataset S9: Time series (2 columns date and values) of total mass of SO_2 in the Virunga Volcanic Province (VVP)

Dataset S10: Time series (2 columns date and values) of tremor signature based on NCFs GOM-RGB