


# Selecting time windows of seismic phases and noise for engineering seismology applications: a versatile methodology and algorithm

Vincent Perron<sup>1,2,3</sup>  · Aurore Laurendeau<sup>4,7</sup> ·  
Fabrice Hollender<sup>1,3</sup> · Pierre-Yves Bard<sup>3</sup> · Céline Gélis<sup>2</sup> ·  
Paola Traversa<sup>5</sup> · Stéphane Drouet<sup>6</sup>

Received: 14 October 2016 / Accepted: 28 March 2017 / Published online: 17 April 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Seismic signal windowing is the preliminary step for many analysis procedures in engineering seismology (standard spectral ratio, quality factor, general inversion techniques, etc.). Moreover a noise window is often necessary for the data quality control through the signal-to-noise verification. Selecting the noise window can be challenging when large heterogeneous datasets are considered, especially when they include short pre-event noise signals. This study proposes a fully automatic and configurable (i.e., with default parameters that can also be user-defined) algorithm to windowing the noise and the P, S, coda and full signal once the P-wave ( $T_P$ ) and S-wave ( $T_S$ ) first arrivals are known. An application example is given on a KiK-net dataset. A Matlab language implementation of this algorithm is proposed as an online resource.

**Keywords** Seismic phase windowing · Noise windowing · Signal-to-noise ratio · Matlab algorithm

---

**Electronic supplementary material** The online version of this article (doi:[10.1007/s10518-017-0131-9](https://doi.org/10.1007/s10518-017-0131-9)) contains supplementary material, which is available to authorized users.

---

✉ Vincent Perron  
vincent.perron.mail@gmail.com

<sup>1</sup> French Alternative Energies and Atomic Energy Commission (CEA), Cadarache, 13108 Saint-Paul-lez-Durance Cedex, France

<sup>2</sup> IRSN, PRP-DGE/SCAN/BERSSIN, BP 17, 92262 Fontenay-aux-Roses, France

<sup>3</sup> Univ. Grenoble Alpes, IFSTTAR, ISTerre, CS 40700, 38058 Grenoble, France

<sup>4</sup> CEA, DAM, DIF, 91297 Arpajon, France

<sup>5</sup> EDF, CEIDRE/TEGG/SGG, 905 av. du Camp de Menthe, 13097 Aix-en-Provence, France

<sup>6</sup> GEOTER, Fugro Group, Auriol, France

<sup>7</sup> Present Address: Instituto Geofisico Escuela Politecnica Nacional, Quito, Ecuador

## 1 Introduction

Selecting specific signal phases (i.e., P, S, or coda waves) is required for diverse applications in seismology. For instance, the early part of shear-wave phase is often used for site effects assessment (e.g., Borchardt 1970; Jongmans and Campillo 1993; Horike et al. 2001; Satoh et al. 2001) and is the basis of the evaluation of the kappa parameter (Anderson and Hough 1984; Ktenidou et al. 2014), while the quality factor (Q) related to the attenuation is regularly estimated from the later coda arrivals (e.g., Aki and Chouet 1975; Mayor et al. 2016). Moreover, to estimate the quality of a signal and its frequency range of validity, the signal-to-noise ratio (SNR) is computed from the ratio between the Fourier amplitude spectrum (FAS) estimated on parts of the signal and the FAS generally evaluated for a noise window of the same duration that is often selected before the event.

While numerous studies have proposed automatic picking algorithms to determine the P-wave ( $T_P$ ) and/or S-wave ( $T_S$ ) first arrivals (e.g., Baer and Kradolfer 1987; Sleeman and van Eck 1999; Zhao and Takano 1999; Zhang et al. 2003; Stefano et al. 2006; Wong et al. 2009; Küperkoch et al. 2010; Tan and He 2016), there have only been a few that have offered a solution for windowing the different phases of the earthquake signal. Most studies have considered a constant window duration for every event, without taking into account the earthquake rupture duration or the expansion of the signal duration as waves are being propagated to larger distances (Phillips and Aki 1986; Bonilla et al. 1997; Drouet et al. 2010; Douglas et al. 2010). Recently, some studies have proposed more complex approaches based on the signal analysis (e.g., Maggi et al. 2009) or based on a model using the information extracted from seismic bulletins (e.g., Kishida et al. 2016).

When working with a large and heterogeneous dataset, once the  $T_P$  and  $T_S$  first arrivals have been picked, defining a specific window can be complicated. This complexity increases when a noise window has to be assessed for SNR computation. Indeed, time-series extracted from triggered instruments and/or generated automatically from regional or national networks, often present short and variable pre-event noise durations. When a large dataset has to be processed, as for generalized inversion techniques (GIT—e.g., Drouet et al. 2008) or ground motion prediction equations (e.g., Laurendeau et al. 2013), then a complex automatic procedure has to be used to avoid the introduction of poor quality data into the processing and to minimize the number of rejected data due to difficult window selection.

The motivation behind the present study was to provide a suitable windowing tool for spectral estimation on different phases with due account to signal-to-noise ratio issues. A method to select the phase windows of any dataset for which the  $T_P$  and  $T_S$  first arrivals are known is proposed, and a suitable solution to estimate the noise window from heterogeneous datasets with variable noise level and duration is provided. An automatic Matlab algorithm was developed and tested on a KiK-net Japanese dataset that is composed of more than 2000 manually picked events with short and variable durations of pre-event noise (Laurendeau et al. 2013). The records are accelerograms from local to regional events that are used between 0.25 and 30 Hz mostly. This study was initially developed for the application of GIT to a specific KiK-net subset (Foundotos et al. 2016—same issue), and for the correction of the KiK-net surface records for local site effects for prediction of hard-rock reference ground motion prediction (Laurendeau et al. 2016—same issue): more details on the dataset can be found in these papers.

After a short reminder on the relationships between window duration and the associated minimum frequency valid for the FAS, ( $f_{min}$ ), we present the model proposed for the

seismic-phase windowing in a first step, and the methodology used for the noise window selection in a second step. In a final step, some examples are given to discuss the windowing obtained for local, regional and teleseismic events, as well as for complex examples including after- or fore-shocks.

## 2 Spectral resolution and sensitivity to time window duration

Many studies have tested the sensitivity of their data to signal window duration (e.g., Satoh et al. 2001; Ktenidou 2010; Douglas et al. 2010). These have mostly reported only limited dependence, provided that the same seismic phase is considered. These observations have sometimes been used to justify the choice of a constant window duration for every event.

In addition to the potential differences in the input seismic signal delimited by different windows, the resolution of the FAS differs as well. Indeed, the longer the selected time window, the higher the number of wavelengths considered for the FAS at each frequency. For instance, a 10-s-long window contains only one wavelength of a 10-s period, but 10 wavelengths of a 1-s period, and 100 wavelengths of a 0.1-s period, while a 1-s-long window contains one tenth of the wavelengths at each of these periods. If  $N$  is the number of wavelengths necessary to insure a good spectral resolution, then the minimal reliable frequency for the FAS is given by:

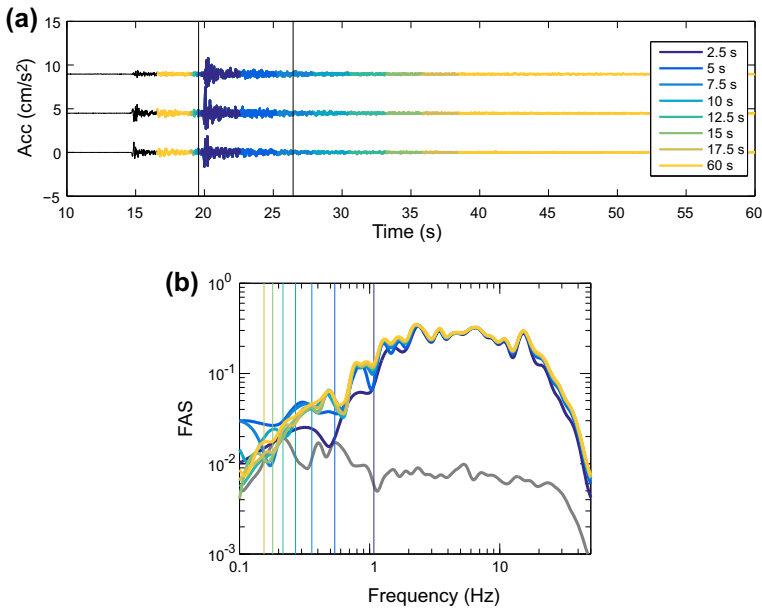
$$f_{min} = \frac{N}{D}, \quad (1)$$

where  $D$  is the duration of the window considered. The higher is  $N$ , the better is the spectral resolution. Based on our experience, taking  $N = 10$  is enough to give the assurance of good spectral resolution. However, taking such a high  $N$  number is not always possible, as this depends on the seismic phase or noise duration available, and on the minimal frequency necessary for the application. When it is required, the  $N$ -value can be optimized by tested it values for different signals. Figure 1 shows the sensitivity of the FAS to the S-wave window duration ( $D_S$ ). The colors from blue to yellow show the results for window lengths from 2.5 to 60 s. In this example and for various other KiK-net signals tested (not represented here), the minimal  $N$ -value for this dataset is around three. Indeed, clear discrepancies appear at low frequency for the shortest window, generally just below the  $f_{min}$  criteria (vertical lines) obtained with  $N = 3$ . Because the KiK-net dataset present very limited duration of noise for the analyses and tests with  $N = 3$  appear satisfying, we keep this  $N$ -value in the following examples.

In agreement with the literature, we find a good stability of the FAS over the frequency range where the  $N$ -value criterion is satisfied (Fig. 1). Thus, the FAS seems weakly dependent on the duration of the signal considered, provided the most energetic part of the signal is common to every window. It means that small variations on the duration of the selected phase window would lead to negligible change on the FAS.

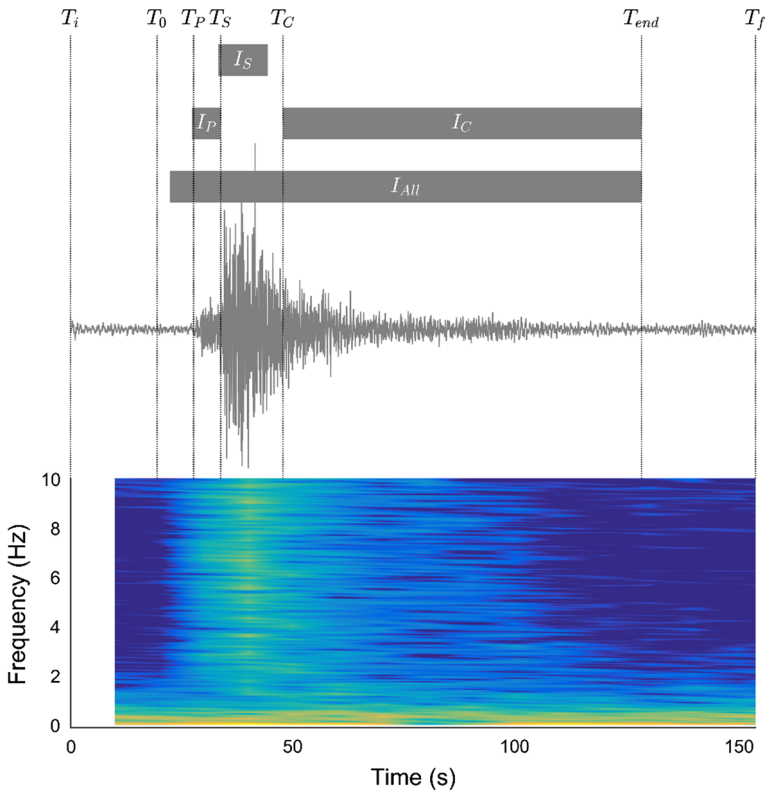
## 3 Seismic-phase windowing

Phase windowing consists of using the first arrival time of P waves and S waves to automatically select different windows: for P, S, coda, and full signals. The nomenclature for the phase intervals and the different times considered is given in Fig. 2. In addition to



**Fig. 1** S-wave window duration impact on the computation of the Fourier acceleration spectrum (FAS). **a** The time histories of the three components and the selected window are shown. The window is a little larger because the cosine taper is applied outside the limit interval of the defined S-wave window. **b** The corresponding FAS are shown. FAS in gray correspond to the noise spectrum. The vertical lines indicate the minimum frequency associated with the S-wave window and allowed to have at least three wavelengths ( $N = 3$ ). In this example, it is necessary to have at least 10 s of signal to have a reliable spectrum at 0.3 Hz

$T_P$  and  $T_S$ , the ending signal time ( $T_{end}$ ) can be defined as well. This time is used as an upper limit for the duration of the S phase, the coda phase, and the full signal. We recommend selecting  $T_{end}$  directly from the spectrogram with precautions in the chosen color scale, to be able to detect the end of the coda waves at low frequency, as well as eventual strong noise or aftershocks at each frequency. If the  $T_{end}$  value is not provided, then it is automatically taken as the time corresponding to 95% of the cumulated energy evaluated on the three components between  $T_P$  and the end of the record. It is particularly useful to pick  $T_{end}$  for low SNR records and in the case of close aftershocks or strong transient noise, which can be included by the cumulated energy approach. Moreover, the cumulated energy approach presents the drawback that it depends on the duration and level of noise included in the record around the signal, especially when the latter is weak. The time of the initial sample of the record is denoted  $T_i$ , while the final sample is given by  $T_f$ . The time for the earthquake occurrence is noted as  $T_0$ . The method and its associated algorithm have been developed for FAS processing purposes. The window's edge must be tapered to satisfy the infinite signal assumption made for the Fourier transform on a finite window. Thus, the rate of tapering ( $\alpha$ ) can be specified in the window selection process, to enlarge the windows and apply the tapering outside the accurate delimitation of the phase windows.



**Fig. 2** East–West component record from a  $M_L$  2.2 earthquake at 56 km epicentral distance and occurring at  $T_0$ . The P, S, coda and all phases are represented by the gray bands indicating  $I_P$ ,  $I_S$ ,  $I_C$  and  $I_{All}$ , respectively, with their corresponding durations ( $D_P$ ,  $D_S$ ,  $D_C$ ,  $D_{All}$ ) and considering a rate of tapering of 5% ( $tx = 0.05$ ). The first arrival times ( $T_P$ ,  $T_S$ ,  $T_C$ ), phases ending ( $T_{end}$ ) and beginning ( $T_i$ ) and ending ( $T_f$ ) of the record are also shown. *Bottom* spectrogram for the seismic energy as a function of the time and frequency

### 3.1 P-wave windowing

The P-wave window is the easiest phase to delimit as it starts at  $T_P$  and ends at  $T_S$ . The duration of the P-wave phase ( $D_P$ ) can be written as:

$$D_P = \frac{T_S - T_P}{1 - tx}, \tag{2}$$

where  $tx$  takes into account the single edge enlargement in the noise before the P-wave onset, to apply the tapering on this pre-event noise and thus to avoid losing some parts of the P-wave phase in the tapering process for the FAS evaluation. Finally, the P-wave window interval is defined as:

$$I_P = [T_P - D_P \times tx; T_S]. \tag{3}$$

### 3.2 S-wave windowing

The S-wave window duration is given according to a source term through the inverse of the corner frequency ( $f_c$ —Brune 1970), and a propagation term taking into account the difference time between the P-wave and S-wave first arrivals ( $T_S-T_P$ ):

$$D_S = \max \left\{ \frac{1}{f_c} + (T_S - T_P) \right\} / (1 - 2tx). \tag{4}$$

Here, the window is being enlarged by a factor  $tx$  for both edges. Thus, the enlargement of the first edge includes a small portion of P waves in the S window, although P waves are already included in the S-wave phase anyway. A minimal duration ( $D_{smin}$ ) can be defined for the question of spectral resolution at low frequency.  $f_c$  is estimated directly by the Brune (1970) relationship [Eq. (4)] from the seismic moment ( $M_0$ ), considering a stress drop ( $\Delta\sigma$ ) of 10 bars and a mean shear-wave velocity ( $\beta_S$ ) for the crust of 3500 m/s:

$$f_c = 0.37\beta_S \left[ \frac{16 \times \Delta\sigma \times 10^5}{7 \times M_0} \right]^{\frac{1}{3}}. \tag{5}$$

$\Delta\sigma$  and  $\beta_S$  can be, however, easily adapt to the target region if needed. The seismic moment can be deduced from the moment magnitude ( $M_W$ ) according to Eq. (6) (Hanks and Kanamori 1979):

$$M_0 = 10^{1.5M_w+9.1}. \tag{6}$$

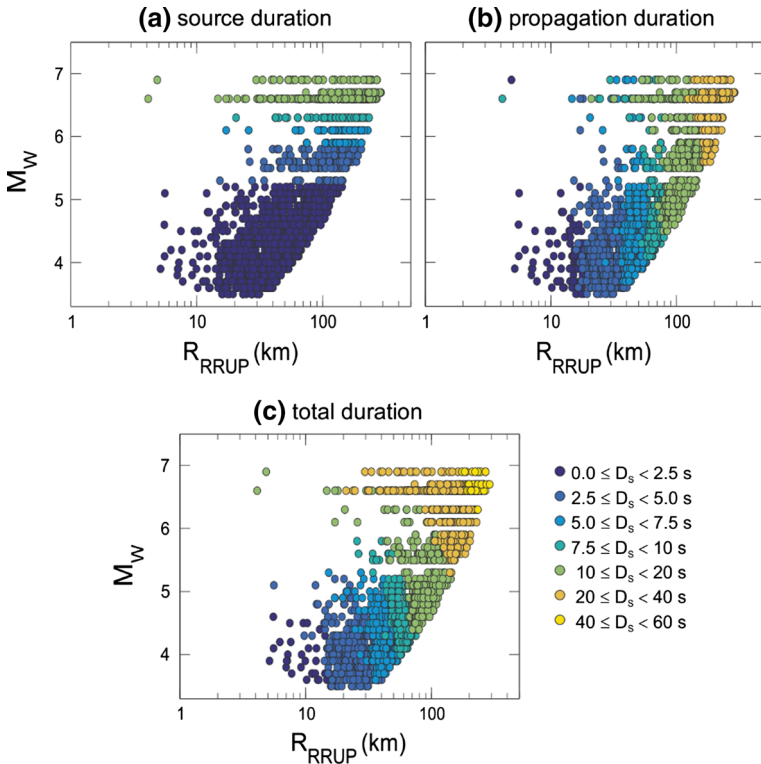
If the moment magnitude is not available,  $M_W$  can be approximated by the local magnitude ( $M_L$ ) extracted from the seismic catalog. This estimation of the source duration is anyway approximate, but is supported by the observed stability in the spectrum evaluated from windows with different  $D_S$  as shown in Fig. 1. Kishida et al. (2016) proposed a similar formulation for  $D_S$ , with a part related to the source with different durations defined empirically according to the magnitude, and a part related to the propagation defined as one tenth of the hypocentral distance ( $0.1R_h$ ).

First, for the source term,  $f_c$  is high for small and moderate earthquakes ( $M < 5$ ), and thus  $D_S$  is only controlled by the propagation term. Then, the source duration can be neglected for  $M < 5$ , making Eq. (4) usable without the need for parameters other than  $T_S$  and  $T_P$ . The use of Eq. (4) for earthquakes with  $M > 7.5$  can lead to very large source durations (see Kishida et al. 2016). A maximal S-wave window duration ( $D_{smax}$ ) can be chosen in this context. Secondly, for the propagation term, ( $T_S-T_P$ ) is widely accepted to be close to a 1/8 of the hypocentral distance given in kilometers, making both expressions similar. However, the formulation in Eq. (3) has the advantage of being independent of uncertainties in the source localization (especially the depth) given by the seismic catalog.

The S-wave interval is finally defined as:

$$I_S = \left[ T_S - D_S \times tx; \min \left\{ \begin{matrix} T_S + D_S(1 - tx) \\ T_{end} \end{matrix} \right\} \right]. \tag{7}$$

Figure 3 shows the variation of source and path component duration of  $D_S$  as defined by Eq. (4) with the magnitude and rupture distance for the KiK-net dataset example. The maximum duration due to the source is around 17 s, and that due to the propagation is



**Fig. 3** Duration of the S-wave windows (c) with respect to the term duration of the source (a;  $M_W$  effect) and the path (b;  $R_{RRUP}$  effect)

around 38 s. The total duration has a minimum of 1.2 s and a maximum of 55 s. However, in the following, we consider a target minimal frequency of 0.3 Hz and a minimum of three wavelengths included inside the window ( $N = 3$ ). Equation (1) finally gives a minimal duration  $D_{Smin} = 10$  s for applications on the KiK-net dataset.

### 3.3 Coda-wave windowing

Aki (1969) defined the beginning of the coda phase as twice the S-wave travel time ( $2(T_S - T_0)$ ) after earthquake occurrence  $T_0$ . To be independent of the information extracted from the seismic bulletin, we propose a formulation equivalent to the Aki (1969) definition, but based only on the  $T_P$  and  $T_S$  parameters. Using the approximation commonly accepted that  $R_h \propto 8(T_S - T_P)$  and that  $\beta_S \approx 3.5$  km/s and considering  $R_h/\beta_S \approx (T_S - T_0)$ , we easily find  $2(T_S - T_0) \approx 4.6(T_S - T_P)$ . The time of the first ‘arrival’ of the coda phase ( $T_C$ ) is finally  $T_C = 4.6(T_S - T_P) + T_0$ , which is equivalent to  $T_C = 2.3(T_S - T_P) + T_S$ . This formulation is also empirically confirmed through the good coefficient of determination  $R^2 = 0.98$  for the linear regression on the local events ( $R_h < 500$  km) between Aki (1969) and our  $T_C$  expression. This definition of  $T_C$  has the advantage that it is independent of the uncertainty on the time origin of the earthquake. Finally, the coda-wave interval can be written as:

$$I_C = [3.3T_S - 2.3T_P; T_{end}] \tag{8}$$

Its associate duration is simply:

$$D_C = T_{end} - (3.3T_S - 2.3T_P) \tag{9}$$

A minimal coda wave window duration ( $D_{Cmin}$ ) can be defined. If  $D_C < D_{Cmin}$ , then no coda wave window is returned by the algorithm since the signal is too weak and the amplitude of the coda waves falls rapidly below the noise level. A coda signal is generally available only for events with good SNR (i.e. >10) in a large frequency range.

### 3.4 Full-signal windowing

A full-signal interval ( $I_{All}$ ) is also proposed that starts at  $T_P$  and finishes at  $T_{end}$ . This takes into account the enlargement on the first edge for the tapering, and it is defined as:

$$I_{All} = [T_P - D_{All} \times tx; T_{end}] \quad \text{with} \quad D_{All} = \frac{T_{end} - T_P}{1 - tx} \tag{10}$$

This full signal window is particularly useful when no specific phase is mandatory, and to obtain long enough windows ( $D_{All}$ ) to assess spectra with good resolution up to low frequencies.

## 4 Noise window selection

The noise window selection generally consists of taking the duration of the target window and reporting it before the first P-wave arrival. Here, a more complex scheme has been developed, to take into account the availability of data with short windows of pre-event noise. Figure 4 shows the pre-event noise duration that is available for the KiK-net dataset. This shows generally that these durations are short and variable, which makes the noise window selection difficult. Only one noise window of duration ( $D_N$ ) is selected to assess the SNR of several seismic phase windows of different durations. The duration of the target noise window ( $D$ ) has to be at least as long as the longest seismic signal window requested ( $D_P, D_S, D_C$  or  $D_{All}$ ), or long enough to satisfy the  $f_{min}$  criterion. Thus, FAS estimated from these windows of different durations have to be normalized by the square root of the duration to obtain the Fourier amplitude spectrum density ( $FASD = FAS/\sqrt{D}$ ) that is length-independent, for SNR purposes especially.

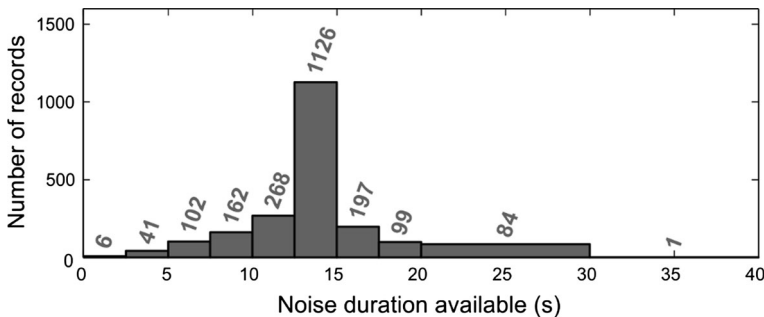


Fig. 4 The number of records versus the pre-event noise duration available



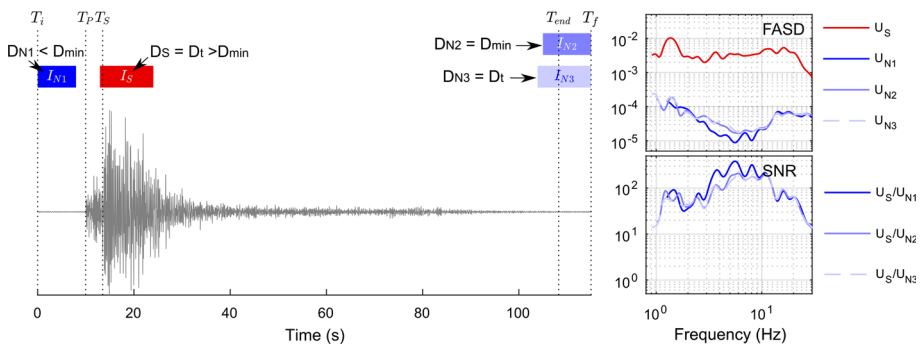
In the present example, we consider a minimal noise duration  $D_{min}$  of 10 s to be consistent with the minimal S-wave window duration taken previously. However, 10 s of noise is not available before the event for all of the records of the KiK-net dataset, as shown in Fig. 4. The dataset contains 311 records without 10 s of pre-event noise. Thus, different noise window definitions are tested, for which the energy was then compared. The idea was to select a noise window with sufficient duration, and also a window with a representative level of energy (without strong seismic signal included). To do this, one pre-event noise window ( $I_{N1}$ ) of duration ( $D_{N1}$ ) is tested as well as two post-events noise windows: a short window  $I_{N2}$  of duration  $D_{N2}$ , and a long window  $I_{N3}$  of duration  $D_{N3}$ .

$$I_{N1} = \left[ \max \left\{ (T_P - 0.1) - \max_{T_i} \left\{ \frac{D_{min}}{D_t} \right\} \right\}; T_P - 0.1 \right]; \tag{11}$$

$$I_{N2} = \left[ \max \left\{ T_f - \max_{T_S + D_S} \left\{ \frac{D_{min}}{D_{N1}} \right\} \right\}; T_f \right]; \tag{12}$$

$$I_{N3} = \left[ \max \left\{ T_f - \max_{T_S + D_S} \left\{ \frac{D_{min}}{D_t} \right\} \right\}; T_f \right]. \tag{13}$$

The pre-event window ( $I_{N1}$ ) is the preferred one, as no earthquake signal can be introduced in the noise evaluation. However, if the pre-event noise that is available is too short, or if there is a fore-shock before the mainshock, then other windows have to be considered. The post-event window  $I_{N3}$  is longer than  $I_{N2}$  only when the target phase duration ( $D_t$ ) is greater than  $D_{min}$  and  $D_{N1}$ . No S wave can be introduced in the post-event window, while the coda wave can be accepted to minimize the number of rejected signals due to a lack of noise available. An example is given in Fig. 5 that illustrates the noise window selecting process for the S-wave duration targeted on a record presenting a limited pre-event noise duration available. In this example,  $I_{N3}$  is preferable to the two other noise windows because it is the only one that has the same duration as the targeted S-wave window while it provides



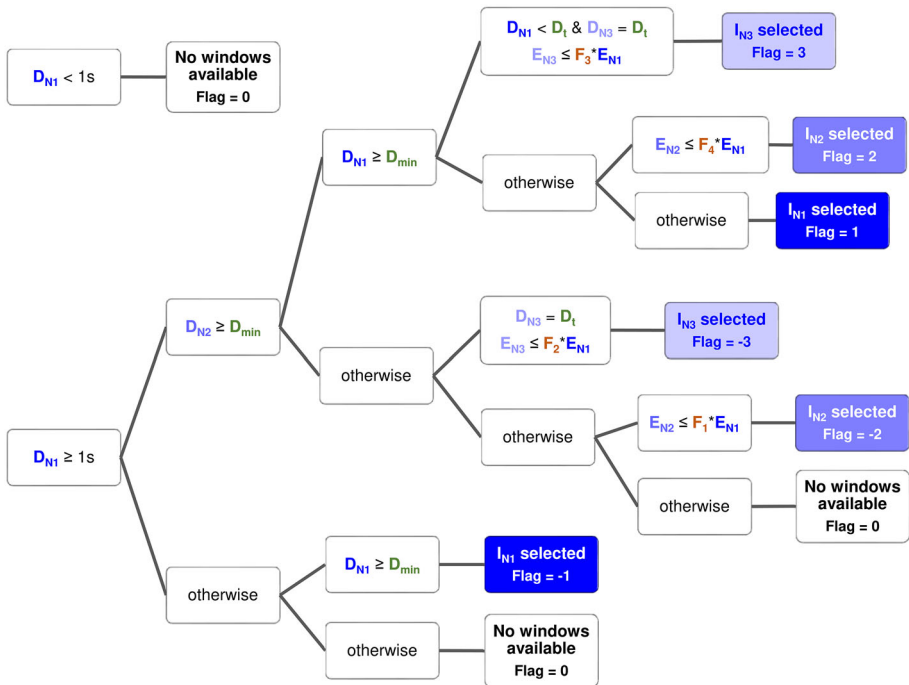
**Fig. 5** Example of the noise window selection process when the S-wave window duration is targeted ( $D_t = D_s$ ) and considering a 10 s minimal window duration ( $D_{min} = 10$  s). The S-wave window ( $I_S$ ), the pre-event noise window ( $I_{N1}$ ) and the short and long post-event noise window ( $I_{N2}$  and  $I_{N3}$ ) are represented as well as their corresponding amplitude spectral density (FASD) and deduce signal-to-noise ratio (SNR)

similar FASD than the two other noise windows even if some coda waves may be included inside  $I_{N3}$ . This latter verification is carried out on the comparison of the mean energy of the different windows. A minimum pre-event noise duration of 1 s is mandatory for these comparisons. To be able to compare just the representativeness of each noise window, the mean energy ( $E$ ) is estimated for the three noise windows ( $E_{N1}$ ,  $E_{N2}$ ,  $E_{N3}$ ) according to the following definition:

$$E = \frac{\sum_{i=N_{min}}^{Nf} FAS_{EW}(f_i)^2 + FAS_{NS}(f_i)^2 + FAS_Z(f_i)^2}{3 \times (Nf - N_{min} + 1)}, \tag{14}$$

where  $N_{min}$  is the index of the minimum frequency ( $f_{min}$ ),  $Nf$  is the number of frequency samples, and  $FAS$  are the Fourier amplitude spectra computed for the three components.

A scheme can take into account the length of each window as well as their mean energy, to determine the best noise window for the noise FASD evaluation, as detailed in Fig. 6. The energy for the three noise window comparisons is weighted by some factors defined by the user ( $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$ ). This allows  $I_{N1}$ ,  $I_{N2}$ , or  $I_{N3}$  to be favored, depending on the duration available for each one, and the number of rejected records to be minimized due to lack of noise for the SNR evaluation. In addition to the noise and the P, S, coda and all signal windows, the algorithm returns a Flag value that indicates which noise window has been selected and how. For the KiK-net application the scheme of the noise window selection process is configured with:  $D_{min} = 10$  s,  $F_1 = 5$ ,  $F_2 = 3$ ,  $F_3 = 2$ , and  $F_4 = 0.67$ .



**Fig. 6** Scheme of the noise window selection methodology.  $D_{N1}$ ,  $D_{N2}$  and  $D_{N3}$  are the pre-event and the short and the long post-event noise window duration respectively. The corresponding spectral energy of these noise windows are given by  $E_{N1}$ ,  $E_{N2}$  and  $E_{N3}$ . Few other parameters can be easily adapted to each dataset: the minimal and target duration ( $D_{min}$  and  $D_t$ ) and the weighted factors  $F_1$ – $F_4$

**Table 1** Number of records selected both at depth and at the surface from the 2119 events of the KiK-net dataset for each noise window definition

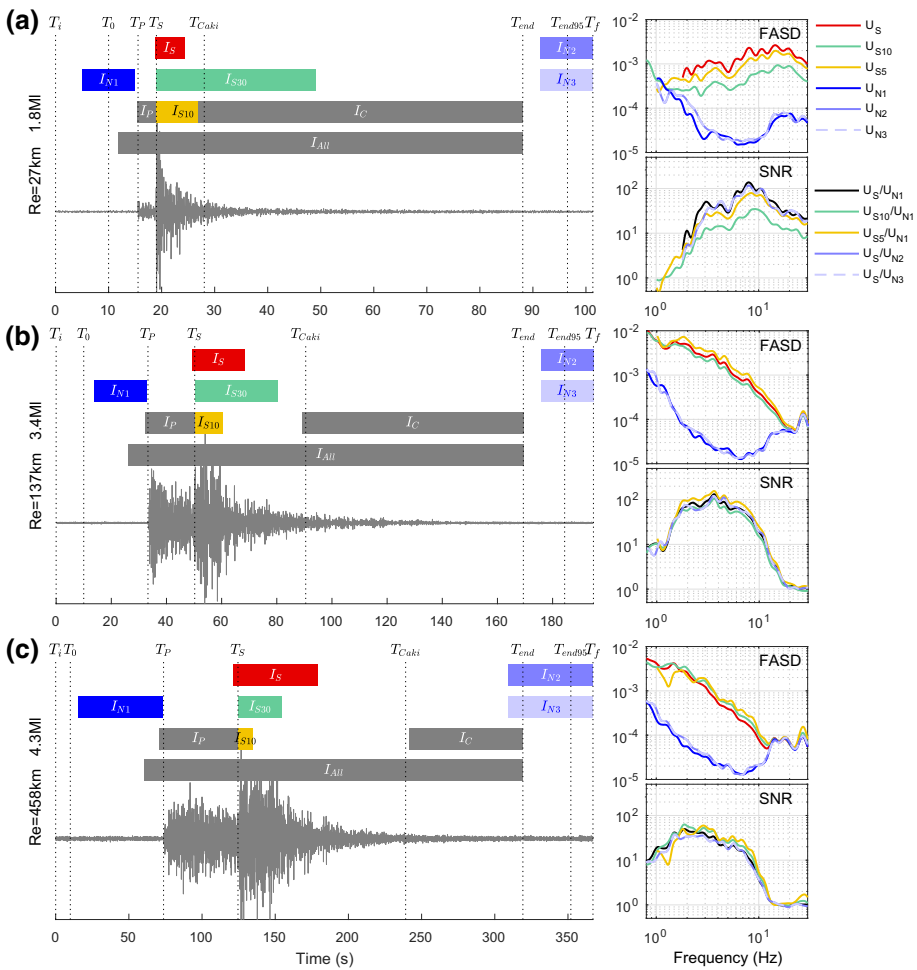
Flag	Definition	Records (n)	
		At depth	At the surface
−3	Long post-event noise window ( $I_{N3}$ ) selected without the possibility to take a pre-event noise window	297	250
−2	Short post-event noise window ( $I_{N2}$ ) selected without possibility to take a pre-event noise window	60	72
−1	Pre-event noise window ( $I_{N1}$ ) selected without possibility to take a post-event noise window	31	33
0	No noise window selected	6	30
1	Pre-event noise window ( $I_{N1}$ ) selected	1538	1547
2	Short post-event noise window ( $I_{N2}$ ) selected	6	25
3	Long post-event noise window ( $I_{N3}$ ) selected	181	162

The main idea is that the pre-event window is favored when both the pre-event and post-event windows are longer than  $D_{min}$  and  $D_t$  (Flag 1). However if the pre-event window is shorter than  $D_t$ , then the longest post-event window is preferred (Flag 3). In the same way, if  $I_{N1}$  is shorter than  $D_{min}$ , then the long post-event is preferred if not too much signal is included inside it (Flag-3), otherwise the short post-event window is chosen (Flag-2).

These Flags for the noise selection are indicated in Fig. 6 and are given in Table 1, with the corresponding number of events that were selected for the KiK-net dataset example. Most of the noise windows selected were pre-event windows. The post-event windows selected were mostly constituted by the long window. Only a few signals are removed from the dataset due to the absence of a noise window for the SNR evaluation. All of the parameters and factors given in this article can be adjusted as an input of the algorithm.

## 5 Application examples and discussion

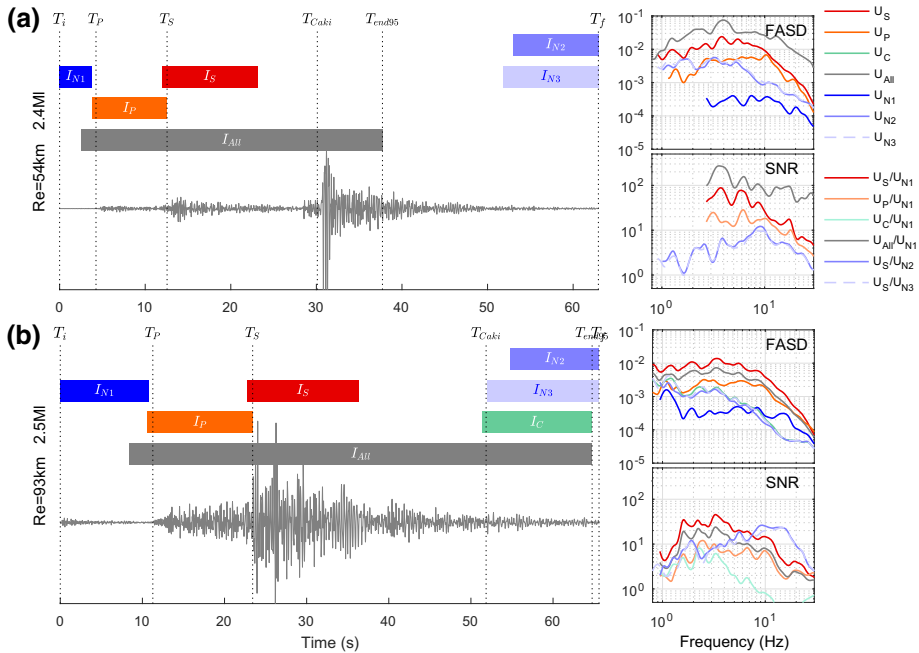
The advantage of our windowing formulation is illustrated in Fig. 7 by the comparison with two simple formulations: a 30- and a 10-s constant S-wave window. Three earthquakes recorded in Provence (Southeastern France) with increasing epicentral distance are presented. The influence of the S-wave window duration is visible on the FASD and the SNR. For the closest earthquake (7a), the 30-s window includes a lot of coda waves and underestimates greatly the FASD obtained with the 10-s window and our formulation. For the intermediate epicentral distance (7b), our formulation is in between the two constant window leading to a FASD that is close to the average between the three window definitions. Only the beginning of the S-wave window is included in the two constant windows for the regional earthquake (7c) while our formulation provides longer duration leading to a slightly lower FASD amplitude even if the three curves are very similar. Thus, a constant window duration is not adapted when working with datasets composed by both local and regional earthquakes while our formulation gives a suitable solution. Concerning the other phases of the signal, we used the classical P wave formulation that seems appropriate, while our coda wave window formulation begins very close to the one predicted by Aki (1969), as expected. The  $T_{end95}$  cumulative energy estimation is always later than the



**Fig. 7** Example of the signal phase and noise windowing for 3 earthquakes located at 27 (a), 137 (b) and 458 km (c) from the recording site. The P-wave window ( $I_P$ ), the S-wave window ( $I_S$ ), the coda wave window ( $I_C$ ) and the full signal window ( $I_{AU}$ ) are represented. To make the comparison between our formulation and simple constant window formulation, 30- and 10-s long windows ( $I_{S30}$  and  $I_{S10}$ ) are also represented as well as their corresponding amplitude spectral density (FASD) and associated signal-to-noise ratio (SNR). The picked phase ending ( $T_{end}$ ) and the one predicted by the 95% fractile of the cumulative energy ( $T_{end95}$ ) are represented as well as the Aki (1969) coda beginning formulation ( $T_{Caki}$ ) that can be compared to our formulation ( $I_C$ ). Identically to Fig. 5, the noise window selection process is also represented by the pre-event noise window ( $I_{N1}$ ) and the short and long post-event noise window ( $I_{N2}$  and  $I_{N3}$ ) and their associated FASD and SNR. Here, the pre-event noise duration is long enough to target  $D_S$  leading to the same duration for the three noise windows ( $D_{N1} = D_{N2} = D_{N3} = D_S$ )

manually picked one, and this duration increase may be accentuated for longer records. The three noise windows have the same duration and present similar FASD. However, long pre-event noise is available and this may not be true for each record of every dataset.

Figure 8 presents two examples recorded during the 2014 Cephalonia seismic crisis in Greece, for which the selection of noise and phase windows is uneasy and might be unreliable. Here, the phase ending time has not been picked and is defined by  $T_{end95}$ . In the



**Fig. 8** Examples of uneasy signal phase and noise windowing when no phase ending ( $T_{end}$ ) has been picked and when either after—(a) or fore-shock (b) are present in the record. Similarly to Fig. 7, all the signal phases ( $I_P$ ,  $I_S$ ,  $I_C$  and  $I_{All}$ ) and noise windows ( $I_{N1}$ ,  $I_{N2}$  and  $I_{N3}$ ) are represented with their corresponding FASD and SNR. Here, the pre-event noise is limited, leading to test different post-event noise windows durations

first case (a), the target earthquake is followed by a stronger one that strongly biases the  $T_{end95}$  estimation and leads to include the P- and S-wave phases of the second earthquake in the coda and the full signal phase of the first event. However, the duration of the coda wave is not long enough here to satisfy the minimal coda wave duration criteria ( $D_C < D_{Cmin} = 10$  s). Moreover, for this particular example, the noise selection leads to the rejection of this record (Flag = 0) since the pre-event noise window is too short and the FASD of the two post-event noise windows are too different from the pre-event one.

In the second example, a small fore-shock is present in the pre-event noise while the record is cut before the actual end of the signal. Here the signal phases are not biased but the noise selection is complex. Indeed,  $I_{N1}$  is enriched at high frequency compared to  $I_{N2}$  and  $I_{N3}$  while it is the opposite for low frequency due to the presence of long-period coda waves in the post-event noise window. When using our parameterization ( $D_{min} = 10$  s,  $F_1 = 5$ ,  $F_2 = 3$ ,  $F_3 = 2$ , and  $F_4 = 0.67$ ), the algorithm selects the pre-event noise window since it exhibits sufficient duration for good spectral resolution. The SNR is, however, significantly lower at higher frequency. To avoid such difficulty, we strongly recommend visualizing and flag such records when picking P- and S-wave first arrivals. Picking  $T_{end}$  or testing more accurate automatic procedures than the cumulative energy one is also required for an accurate coda and full signal phase window ending.

## 6 Conclusions

Seismic signal windowing is the preliminary step for many applications in seismology and for SNR verification. While the vast majority of previous studies have used very simple windowing formulations, such as constant duration, we propose a more complex method that takes into account source and propagation terms. This study provides a suitable solution for heterogeneous datasets where the P-wave and S-wave first arrivals have been picked beforehand. The earthquake signal phases are selected exclusively from the  $T_P$  and  $T_S$  parameters for the majority of the events, which makes the windowing independent of the uncertainties present in the information given by the seismic bulletin. For strong earthquakes ( $M > 6$ ) that have source durations that are not negligible, a source term is estimated through the inverse of the corner frequency evaluated from the magnitude.

Selecting the noise window can be challenging when large heterogeneous datasets are considered, especially if for some events the duration of the available pre-event noise is short. The noise window has to be the most representative of the noise level, and long enough to allow SNR estimation with good resolution at low frequency. To get around this issue, we defined and tested three different windows: one pre-event and two post-event windows. A scheme is proposed for selecting of the best noise window in terms of duration (as long as possible) and mean energy (as low as possible), without including undesirable transients. This approach gave good results on the KiK-net dataset, with a very limited number of rejected signals.

A Matlab algorithm was developed and can be adapted to each dataset through a few parameters to be defined by the user. This algorithm is freely available as electronic online supplementary material of this paper and ready to be used for every windowing application.

**Acknowledgements** This work was conducted within the framework of the CASHIMA project (*Cadarache Seismic Hazard Integrated Multidisciplinary Assessment*), jointly funded by the French Alternative Energies and Atomic Energy Commission (CEA), the Laue-Langevin Institute (ILL) and the International Thermonuclear Experimental Reactor (ITER) organization, and the SIGMA project (*Seismic Ground Motion Assessment*) funded by Electricité De France (EDF), CEA, AREVA and the Ente Nazionale per l'Energia Elettrica (ENEL). We are thankful to Christopher Berrie and to the two anonymous reviewers for their careful suggestions and corrections.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Aki K (1969) Analysis of the seismic coda of local earthquakes as scattered waves. *J Geophys Res* 74:615–631. doi:10.1029/JB074i002p00615
- Aki K, Chouet B (1975) Origin of coda waves: source, attenuation, and scattering effects. *J Geophys Res* 80:3322–3342
- Anderson JG, Hough SE (1984) A model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies. *Bull Seismol Soc Am* 74:1969–1993
- Baer M, Kradolfer U (1987) An automatic phase picker for local and teleseismic events. *Bull Seismol Soc Am* 77:1437–1445
- Bonilla LF, Steidl JH, Lindley GT et al (1997) Site amplification in the San Fernando Valley, California: variability of site-effect estimation using the S-wave, coda, and H/V methods. *Bull Seismol Soc Am* 87:710–730

- Borcherdt RD (1970) Effects of local geology on ground motion near San Francisco Bay. *Bull Seismol Soc Am* 60:29–61
- Brune JN (1970) Tectonic stress and the spectra of seismic shear waves from earthquakes. *J Geophys Res* 75:4997–5009
- Douglas J, Gehl P, Bonilla LF, Gélis C (2010) A  $\kappa$  model for Mainland France. *Pure Appl Geophys* 167:1303–1315. doi:[10.1007/s00024-010-0146-5](https://doi.org/10.1007/s00024-010-0146-5)
- Drouet S, Chevrot S, Cotton F, Souriau A (2008) Simultaneous inversion of source spectra, attenuation parameters, and site responses: application to the data of the French accelerometric network. *Bull Seismol Soc Am* 98:198–219. doi:[10.1785/0120060215](https://doi.org/10.1785/0120060215)
- Drouet S, Cotton F, Guéguen P (2010) Vs30,  $\kappa$ , regional attenuation and Mw from accelerograms: application to magnitude 3–5 French earthquakes: vS30,  $\kappa$ , attenuation and Mw from accelerograms. *Geophys J Int* 182:880–898. doi:[10.1111/j.1365-246X.2010.04626.x](https://doi.org/10.1111/j.1365-246X.2010.04626.x)
- Foundotos L, Hollender F, Laurendeau A et al (2016) Assessing site effects at different scales using different approaches: learnings from generalized inversion of KiK-net data from stiff and rocky sites. *Bull Earthq Eng* (to submitted)
- Hanks TC, Kanamori H (1979) A moment magnitude scale. *J Geophys Res B* 84:2348–2350
- Horike M, Zhao B, Kawase H (2001) Comparison of site response characteristics inferred from micro-tremors and earthquake shear waves. *Bull Seismol Soc Am* 91:1526–1536
- Jongmans D, Campillo M (1993) The response of the Ubaye Valley (France) for incident SH and SV waves: comparison between measurements and modeling. *Bull Seismol Soc Am* 83:907–924
- Kishida T, Ktenidou O-J, Darragh RB, Walter S (2016) Semi-automated procedure for windowing time series and computing Fourier amplitude spectra (FAS) for the NGA-West2 database. Pacific Earthquake Engineering Research Center, Berkeley
- Ktenidou O-J (2010) Experimental and theoretical study of seismic ground motion in the city of Aegion, Greece, focusing on the local site and topographic effects. Doctoral thesis, Aristotle University Thessaloniki
- Ktenidou O-J, Cotton F, Abrahamson NA, Anderson JG (2014) Taxonomy of Kappa: a review of definitions and estimation approaches targeted to applications. *Seismol Res Lett* 85:135–146. doi:[10.1785/0220130027](https://doi.org/10.1785/0220130027)
- Küperkoch L, Meier T, Lee J et al (2010) Automated determination of P-phase arrival times at regional and local distances using higher order statistics. *Geophys J Int* 181:1159–1170. doi:[10.1111/j.1365-246X.2010.04570.x](https://doi.org/10.1111/j.1365-246X.2010.04570.x)
- Laurendeau A, Cotton F, Ktenidou O-J et al (2013) Rock and stiff-soil site amplification: dependency on VS30 and Kappa (K0). *Bull Seismol Soc Am* 103:3131–3148. doi:[10.1785/0120130020](https://doi.org/10.1785/0120130020)
- Laurendeau A, Bard P-Y, Hollender F et al (2016) Derivation of consistent hard rock ( $1000 < V_s < 3000$  m/s) GMPs from surface and down-hole recordings: analysis of KiK-net data. *Bull Earthq Eng* (to submitted)
- Maggi A, Tape C, Chen M et al (2009) An automated time-window selection algorithm for seismic tomography. *Geophys J Int* 178:257–281. doi:[10.1111/j.1365-246X.2009.04099.x](https://doi.org/10.1111/j.1365-246X.2009.04099.x)
- Mayor J, Calvet M, Margerin L et al (2016) Crustal structure of the Alps as seen by attenuation tomography. *Earth Planet Sc Lett* 439:71–80. doi:[10.1016/j.epsl.2016.01.025](https://doi.org/10.1016/j.epsl.2016.01.025)
- Phillips WS, Aki K (1986) Site amplification of coda waves from local earthquakes in central California. *Bull Seismol Soc Am* 76:627–648
- Satoh T, Kawase H, Matsushima S (2001) Differences between site characteristics obtained from micro-tremors, S-waves, P-waves, and codas. *Bull Seismol Soc Am* 91:313–334
- Sleeman R, van Eck T (1999) Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Phys Earth Planet Int* 113:265–275. doi:[10.1016/S0031-9201\(99\)00007-2](https://doi.org/10.1016/S0031-9201(99)00007-2)
- Stefano RD, Aldersons F, Kissling E et al (2006) Automatic seismic phase picking and consistent observation error assessment: application to the Italian seismicity. *Geophys J Int* 165:121–134. doi:[10.1111/j.1365-246X.2005.02799.x](https://doi.org/10.1111/j.1365-246X.2005.02799.x)
- Tan Y, He C (2016) Improved methods for detection and arrival picking of microseismic events with low signal-to-noise ratios. *Geophysics* 81:KS93–KS111. doi:[10.1190/geo2015-0213.1](https://doi.org/10.1190/geo2015-0213.1)
- Wong J, Han L, Bancroft J, Stewart R (2009) Automatic time-picking of first arrivals on noisy microseismic data. *CSEG 0 02 04 06 08 1:1–4*
- Zhang H, Thurber C, Rowe C (2003) Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings. *Bull Seismol Soc Am* 93:1904–1912. doi:[10.1785/0120020241](https://doi.org/10.1785/0120020241)
- Zhao Y, Takano K (1999) An artificial neural network approach for broadband seismic phase picking. *Bull Seismol Soc Am* 89:670–680