

## Chapter 5

# An automated time-window selection algorithm for seismic tomography

### Note

This chapter contains excerpts from a published paper entitled “An automated time-window selection algorithm for seismic tomography,” by Alessia Maggi, Carl Tape, Min Chen, Daniel Chao, and Jeroen Tromp. A. Maggi devised the algorithm and wrote the code, and I was involved in testing and refining some parts of the code. A. Maggi, M. Chen, and I tested the code using three different data sets: global earthquakes (A. Maggi), regional earthquakes from the Japan subduction zone (M. Chen), and crustal earthquakes in southern California (C. Tape). D. Chao (Caltech) did a Summer Undergraduate Research Experience project using and refining the algorithm. His work provided a good starting point for determining a set of user parameters, required by the algorithm, for the southern California data set.

The open-source algorithm FLEXWIN is available for download from the webpage of the Computational Infrastructure for Geodynamics (CIG): [www.geodynamics.org](http://www.geodynamics.org).

## Summary

We present FLEXWIN, an open-source algorithm for the automated selection of time windows on pairs of observed and synthetic seismograms. The algorithm was designed specifically to accommodate synthetic seismograms produced from 3D wavefield simulations, which capture complex phases that do not necessarily exist in 1D simulations or traditional traveltime curves. Relying on signal processing tools and several user-tuned parameters, the algorithm is able to include these new phases and to maximize the number of measurements made on each seismic record, while avoiding seismic noise. Our motivation is to use the algorithm for iterative tomographic inversions, in which the synthetic seismograms change from one iteration to the next. Hence, automation is needed to handle the volume of measurements and to allow for an increasing number of windows at each model iteration. The algorithm is sufficiently flexible to be adapted to many tomographic applications and seismological scenarios, including those based on synthetics generated from 1D models. We illustrate the algorithm using datasets from three distinct regions: the entire globe, the Japan subduction zone, and southern California.

### 5.1 The selection algorithm

Our open-source algorithm, called FLEXWIN to reflect its FLEXibility in picking time WINDows for measurement, operates on pairs of observed and synthetic single component seismograms. The window selection process has five stages, each of which is discussed in Maggi *et al.* (2009): *Stage A*: preprocessing; *Stage B*: definition of preliminary measurement windows; *Stage C*: rejection of preliminary windows based on the content of the synthetic seismogram alone; *Stage D*: rejection of preliminary windows based on the differences between observed and synthetic seismograms; *Stage E*: resolution of preliminary window overlaps. The parameters that permit tuning of the window selection toward a specific tomographic scenario are all contained in a simple parameter file (see Table 5.1). More complexity and finer tuning can be obtained by making some of these parameters time-dependent via user-defined functions that can depend on the source parameters (e.g., event location or depth).

An example of a synthetic seismogram and its corresponding envelope and STA:LTA timeseries  $E(t)$  is shown in Figure 5.1. The  $E(t)$  timeseries starts at its value for a constant

signal, then rises gradually due to the tapered low level numerical noise on the synthetic. At each seismic arrival,  $E(t)$  rises to a local maximum. We can see from Figure 5.1 that these local maxima correspond both in position and in width to the seismic phases in the synthetic, and that the local minima in  $E(t)$  correspond to the transitions between one phase and the next. In the following sections we shall explain how we use these correspondences to define time windows.

Figure 5.2 shows the reduction of candidate windows for the seismogram in Figure 5.1.

## 5.2 Windowing Examples

We present a set of examples showing the results of the FLEXWIN algorithm applied to real data. These examples illustrate the robustness and flexibility of the algorithm. We have applied the algorithm to three tomographic scenarios, with very different geographical extents and distinct period ranges: long-period global tomography (50–150 s), regional tomography of the Japan subduction zone, down to 700 km (6–120 s), and regional tomography of southern California, down to 60 km (2–30 s). For each of these scenarios, we compare observed seismograms to spectral-element synthetics, using our algorithm to select time windows on the pairs of timeseries.

The windowing algorithm itself has little prior knowledge of seismology, other than in the most general terms: it considers a seismogram to be a succession of seismic phases indicated by changes in amplitude and frequency of the signal with time; it is based upon the idea that the short-term to long-term average ratio STA:LTA is a good indicator of the arrival of such phases; it has a notion of the characteristics of an optimal set of data windows. All other prior information — the frequency range to be considered, the portions of the seismogram to be excluded, the acceptable signal-to-noise ratios, the tolerance of dissimilarity between the observed and synthetic seismogram — varies greatly between any two seismological studies. In order to ensure maximum flexibility of our windowing algorithm, all such scenario-dependent information is encapsulated in the tuning parameters of Table 5.1.

We tuned the windowing algorithm separately for each of the three scenarios we present here, and we present examples based on the events listed in Table 5.3. Tuning parameter values for each scenario can be found in Table 5.2, while the functional forms of the time-

dependent parameters can be found in Section 5.3.1. Once tuned for a given scenario, the algorithm is applied to all its events without further modification.

## Local tomography in Southern California

Our last scenario is a local tomographic study of southern California. We apply the windowing algorithm to a set of 140 events within southern California, for which we have computed synthetic seismograms using the spectral-element method and a regional 3D crustal and upper mantle model (Komatitsch *et al.*, 2004). This model contains three discontinuities: the surface topography (included in the mesh), the basement layer that separates the sedimentary basins from the bedrock, and the Moho, separating the lower crust from the upper mantle. The model includes several sedimentary basins, such as the Ventura basin, the Los Angeles basin, and the Salton trough (Komatitsch *et al.*, 2004; Lovely *et al.*, 2006). The smooth 3D background velocity model used in Komatitsch *et al.* (2004) was determined by Hauksson (2000); we use an updated version provided by Lin *et al.* (2007b). The physical domain of the model is approximately 600 km by 500 km at the surface, and extends to a depth of 60 km. Our simulations of seismic waves are numerically accurate down to a period of 2 s.

The 140 events have  $M_w$  magnitudes between 3.5 and 5.5 and were recorded between 1999 and 2007. The locations and origin times are primarily from Lin *et al.* (2007a), and the focal mechanisms are from Clinton *et al.* (2006), Hardebeck and Shearer (2003), or Tan (2006).

We test the windowing code using three period ranges: 6–30 s, 3–30 s, and 2–30 s. The parameters we use for the windowing code are listed in Table 5.2. Figures 5.3 and 5.4 show examples of the output from the windowing algorithm for event 9818433 listed in Table 5.3 recorded at two different stations, while Figure 5.5 shows a summary plot for event 9983429 in the 6–30 s period range.

The windowing algorithm tends to identify five windows on each set of three-component 6–30 s seismograms (Figures 5.3 and 5.5): on the vertical and radial components the first window corresponds to the body-wave arrival and the second to the Rayleigh wave, while windows on the transverse component capture the Love wave. The 2–30 s synthetic seismograms do not agree well with the observed seismograms, especially in the later part of the signal, leading to fewer picked windows. In Figure 5.3c, only three windows are selected by

the algorithm: the P arrival recorded on the radial component, the S arrival on the transverse component, and the Love-wave arrival on the transverse component. The P arrival (PmP or Pn) in fact appears on all three components on both data and synthetics. On the vertical component it is rejected because the cross-correlation value within the time window did not exceed the specified minimum value of 0.85 (Table 5.2). On the transverse component it does not have a large enough signal-to-noise ratio to be picked, but it is evident as a small peak at 36 s in the STA:LTA curve, and more conspicuous when zooming into the synthetics and data. The presence of the P arrival on the transverse component highlights the possibility of measuring subtle phases that may be present in 3D synthetics.

Figure 5.4 shows results for the same event as Figure 5.3, but for a different station, FMP, situated 52 km from the event and within the Los Angeles basin. Comparison of the two figures highlights the characteristic resonance caused by the thick sediments within the basin. This resonance is beautifully captured by the transverse component synthetics (Figure 5.4b, record T), thanks to the inclusion of the basin in the model (*Komatitsch et al.*, 2004). In order to pick such long time windows with substantial frequency-dependent measurement differences, we are forced to lower the minimum cross-correlation value  $CC_0$  for the entire dataset (0.71 in Table 5.2) and increase  $c_{4b}$  to capture the slow decay in the STA:LTA curves (Figure 5.4b, record T). It is striking that although these arrivals look nothing like the energy packets typical for the global case, the windowing algorithm is still able to determine the proper start and end times for the windows. In Figure 5.4c the windowing algorithm selects three short-period body-wave time windows with superb agreement between data and synthetics.

### 5.3 Appendix A: Tuning considerations

FLEXWIN is not a black-box application, and as such cannot be applied blindly to any given dataset or tomographic scenario. The data windowing required by any given problem will differ depending on the inversion method, the scale of the problem (local, regional, global), the quality of the data set and that of the model and method used to calculate the synthetic seismograms. The user must configure and tune the algorithm for the given problem. In this appendix we shall discuss some general considerations the user should bear in mind during the tuning process. For more detailed information on tuning, and for further

examples of tuning parameter sets, we refer the reader to the user manual that accompanies the source code.

The order in which the parameters in Table 5.1 are discussed in the main text of this paper follows the order in which they are used by the algorithm, but is not necessarily the best order in which to consider them for tuning purposes. We suggest the following as a practical starting sequence (the process may need to be repeated and refined several times before converging on the optimal set of parameters for a given problem and dataset).

$T_{0,1}$  : In setting the corner periods of the bandpass filter, the user is deciding on the frequency content of the information to be used in the tomographic problem. Values of these corner periods should reflect the information content of the data, the quality of the Earth model and the accuracy of the simulation used to generate the synthetic seismogram. The frequency content in the data depends on the spectral characteristics of the source, on the instrument responses, and on the attenuation characteristics of the medium. As  $T_{0,1}$  depend on the source and station characteristics, which may be heterogeneous in any given dataset, these filter periods can be modified dynamically by constructing an appropriate user function (e.g., *if station is in list of stations with instrument X then reset T0 and T1 to new values*).

$r_{P,A}$  : In setting the signal-to-noise ratios for the entire seismogram the user is applying a simple quality control on the data. Note that these criteria are applied after filtering. No windows will be defined on data that fail this quality control.

$w_E(t)$  : The short-term average long-term average ratio  $E(t)$  of a constant signal converges to a constant value when the length of the time-series is greater than the effective averaging length of the long-term average. We suggest the user start with a constant level for  $w_E(t)$  equal to this convergence value. The time dependence of  $w_E(t)$  should then be adjusted to exclude those portions of the waveform the user is not interested in, by raising  $w_E(t)$  (e.g., to exclude the fundamental mode surface-wave: *if  $t >$  fundamental mode surface-wave arrival time then set  $w_E(t) = 1$* ). We suggest finer adjustments to  $w_E(t)$  be made after  $r_0(t)$ ,  $CC_0(t)$ ,  $\Delta T_0(t)$  and  $\Delta \ln A_0(t)$  have been configured.

$r_0(t)$ ,  $CC_0(t)$ ,  $\Delta\tau_{\text{ref}}$ ,  $\Delta\tau_0(t)$ ,  $\Delta \ln A_{\text{ref}}$  and  $\Delta \ln A_0(t)$  : These parameters — window signal-to-noise ratio, normalized cross-correlation value between observed and synthetic seismograms, cross-correlation time lag, and amplitude ratio — control the degree of well-behavedness of the data within accepted windows (Stage D). The user first sets constant

values for these four parameters, then adds a time dependence if required. Considerations that should be taken into account include the quality of the Earth model used to calculate the synthetic seismograms, the frequency range, the dispersed nature of certain arrivals (e.g., for  $t$  corresponding to the group velocities of surface-waves, reduce  $CC_0(t)$ ), and *a priori* preferences for picking certain small-amplitude seismic phases (e.g., for  $t$  close to the expected arrival for  $P_{\text{diff}}$ , reduce  $r_0(t)$ ).  $\Delta\tau_{\text{ref}}$  and  $\Delta \ln A_{\text{ref}}$  should be set to zero at first, and only reset if the synthetics contain a systematic bias in traveltimes or amplitudes.

$c_{0-4}$  : These parameters are active in Stage C of the algorithm, the stage in which the suite of all possible data windows is pared down using criteria on the shape of the STA:LTA  $E(t)$  waveform alone. We suggest the user start by setting these values to those used in our global example (see Table 5.2). Subsequent minimal tuning should be performed by running the algorithm on a subset of the data and closely examining the lists of windows rejected at each stage to make sure the user agrees with the choices made by the algorithm.

$w_{\text{CC}}$ ,  $w_{\text{len}}$  and  $w_{\text{nwin}}$  : These parameters control the overlap resolution stage of the algorithm (Stage E). Values of  $w_{\text{CC}} = w_{\text{len}} = w_{\text{nwin}} = 1$  should be reasonable for most applications.

The objective of the tuning process summarily described here should be to maximize the selection of windows around desirable features in the seismogram, while minimizing the selection of undesirable features, bearing in mind that the desirability or undesirability of a given feature is subjective, and depends on how the user subsequently intends to use the information contained within the data windows.

### 5.3.1 Examples of user functions for southern California

As concrete examples of how the time dependence of the tuning parameters can be exploited, we present here the functional forms of the time dependencies used for the southern California tomographic scenario described in the text (Section 5.2). We use predicted arrival times derived from 1D Earth models to help modulate certain parameters. Note, however, that the actual selection of individual windows is based on the details of the waveforms, and not on information from 1D Earth models.

In the following,  $t_P$  and  $t_S$  denote the start of the time windows for the crustal P wave and the crustal S wave, computed from a 1D layered model appropriate to Southern California (*Wald et al.*, 1995). The start and end times for the surface-wave time window,

$t_{R0}$  and  $t_{R1}$ , as well as the criteria for the time shifts  $\Delta\tau_0(t)$ , are derived from formulas in *Komatitsch et al.* (2004). The source-receiver distance (in km) is denoted by  $\Delta$ .

For the 6–30 s and 3–30 s data, we use constant values of  $r_0(t) = r_0$ ,  $CC_0(t) = CC_0$ ,  $\Delta\tau_0(t) = \Delta\tau_0$ , and  $\Delta \ln A_0(t) = \Delta \ln A_0$ . We exclude any arrivals before the P wave and after the Rayleigh wave. This is achieved by the box-car function for  $w_E(t)$ :

$$w_E(t) = \begin{cases} 10w_E & t < t_P, \\ w_E & t_P \leq t \leq t_{R1}, \\ 10w_E & t > t_{R1}. \end{cases} \quad (5.1)$$

For the 2–30 s data, we avoid selecting surface-wave arrivals as the 3D model used to calculate the synthetics cannot produce the required complexity. The water-level criteria then becomes:

$$w_E(t) = \begin{cases} 10w_E & t < t_P, \\ w_E & t_P \leq t \leq t_S, \\ 10w_E & t > t_S. \end{cases} \quad (5.2)$$

Table 5.1: Overview of standard tuning parameters, and of fine tuning parameters. Values are defined in a parameter file, and the time dependence of those that depend on time is described by user-defined functions.

Standard tuning parameters:	
$T_{0,1}$	bandpass filter corner periods
$r_{P,A}$	signal-to-noise ratios for whole waveform
$r_0(t)$	signal-to-noise ratios single windows
$w_E(t)$	water-level on short-term:long-term ratio
$CC_0(t)$	acceptance level for normalized cross-correlation
$\Delta\tau_0(t)$	acceptance level for time lag
$\Delta \ln A_0(t)$	acceptance level for amplitude ratio
$\Delta\tau_{\text{ref}}$	reference time lag
$\Delta \ln A_{\text{ref}}$	reference amplitude ratio
Fine tuning parameters:	
$c_0$	for rejection of internal minima
$c_1$	for rejection of short windows
$c_2$	for rejection of unprominent windows
$c_{3a,b}$	for rejection of multiple distinct arrivals
$c_{4a,b}$	for curtailing of windows with emergent starts and/or codas
$w_{CC}$ $w_{\text{len}}$ $w_{\text{nwin}}$	for selection of best non-overlapping window combination

Table 5.2: Values of standard and fine-tuning parameters for the three seismological scenarios discussed in *Maggi et al.* (2009).

	Global	Japan		S. California		
$T_{0,1}$	50, 150	24, 120	6, 30	6, 30	3, 30	2, 30
$r_{P,A}$	3.5, 3.0	3.5, 3.0	3.5, 3.0	3.0, 2.5	2.5, 3.5	2.5, 3.5
$r_0$	2.5	1.5	3.0	3.0	4.0	4.0
$w_E$	0.08	0.10	0.12	0.18	0.11	0.07
$CC_0$	0.85	0.70	0.73	0.71	0.80	0.85
$\Delta\tau_0$	15	12.0	3.0	8.0	4.0	3.0
$\Delta \ln A_0$	1.0	1.0	1.5	1.5	1.0	1.0
$\Delta\tau_{\text{ref}}$	0.0	0.0	0.0	4.0	2.0	1.0
$\Delta \ln A_{\text{ref}}$	0.0	0.0	0.0	0.0	0.0	0.0
$c_0$	0.7	0.7	0.7	0.7	1.3	1.0
$c_1$	4.0	3.0	3.0	2.0	4.0	5.0
$c_2$	0.3	0.0	0.6	0.0	0.0	0.0
$c_{3a,b}$	1.0, 2.0	1.0, 2.0	1.0, 2.0	3.0, 2.0	4.0, 2.5	4.0, 2.5
$c_{4a,b}$	3.0, 10.0	3.0, 25.0	3.0, 12.0	2.5, 12.0	2.0, 6.0	2.0, 6.0
$w_{CC}, w_{\text{len}}, w_{\text{nwin}}$	1, 1, 1	1, 1, 1	1, 1, 1	0.5,1.0,0.7	0.70,0.25,0.05	1,1,1

Table 5.3: Example events used in *Maggi et al.* (2009). The identifier refers to the CMT catalog for global events and Japan events, and refers to the Southern California Earthquake Data Center catalog for southern California events.

Identifier	Latitude	Longitude	Depth, km	Moment, N m	$M_w$	Location
Global						
101895B	28.06	130.18	18.5	5.68e19	7.1	Ryukyu Islands
200808270646A	-10.49	41.44	24.0	4.68e17	5.7	Comoros Region
050295B	-3.77	-77.07	112.8	1.27e19	6.7	Northern Peru
060994A	-13.82	-67.25	647.1	2.63e21	8.2	Northern Bolivia
Japan						
051502B	24.66	121.66	22.4	1.91e18	6.1	Taiwan
200511211536A	30.97	130.31	155.0	2.13e18	6.2	Kyushu, Japan
091502B	44.77	130.04	589.4	4.24e18	6.4	Northeastern China
Southern California						
9983429	35.01	-119.14	13.5	9.19e15	4.6	Wheeler Ridge
9818433	33.91	-117.78	9.4	3.89e15	4.3	Yorba Linda

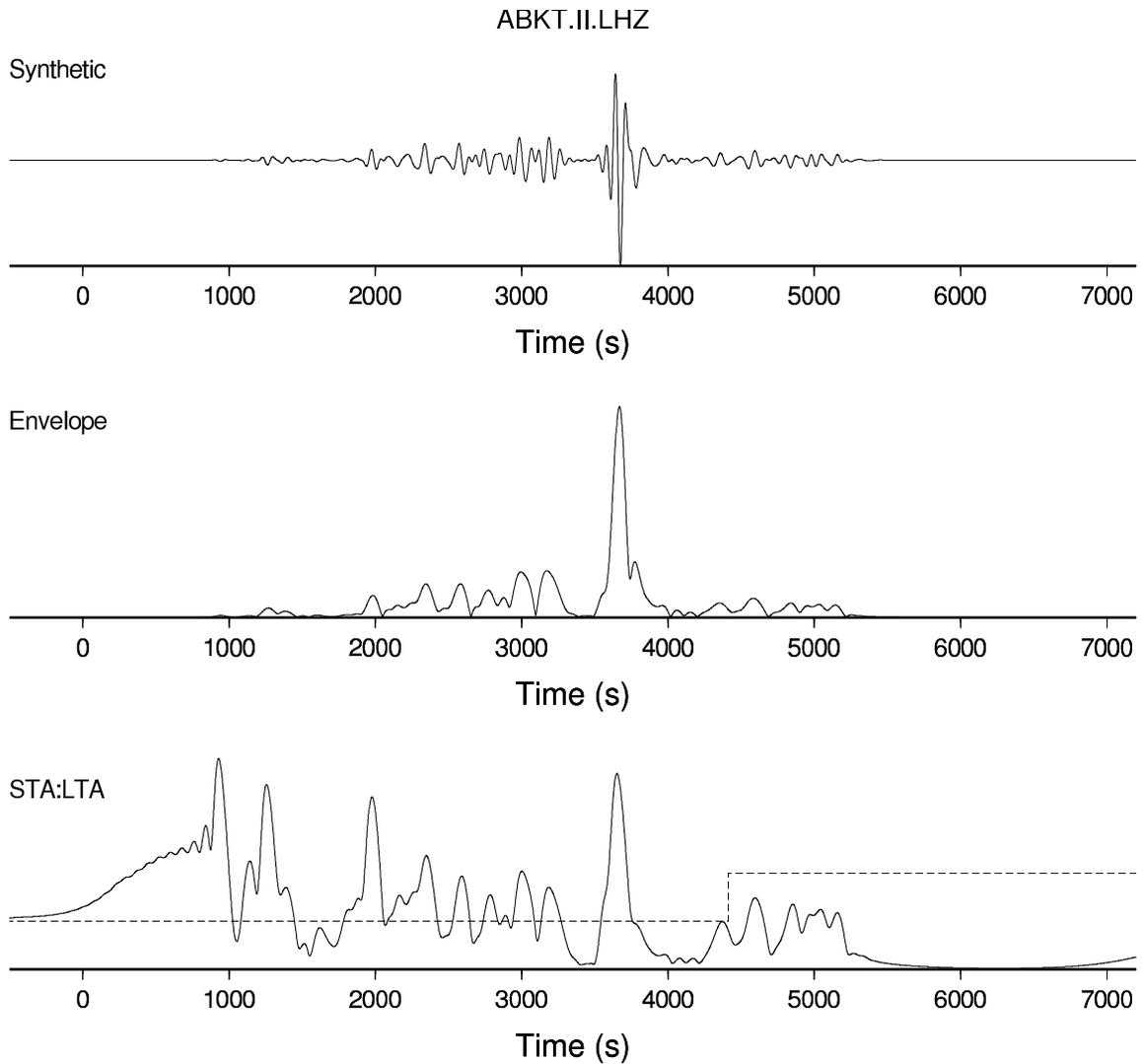


Figure 5.1: Synthetic seismogram and its corresponding envelope and STA:LTA timeseries (Maggi *et al.*, 2009, Figure 1). The seismogram was calculated using SPEC-FEM3D and the Earth model S20RTS (Ritsema *et al.*, 1999) for the CMT catalog event 050295B, whose details can be found in Table 5.3. The station, ABKT, is at an epicentral distance of 14100 km and at an azimuth of 44 degrees from the event. The top panel shows the vertical component synthetic seismogram, filtered between periods of 50 and 150 seconds. The center panel shows its envelope, and the bottom panel shows the corresponding STA:LTA waveform. The dashed line overlaid on the STA:LTA waveform is the water-level  $w_E(t)$ .

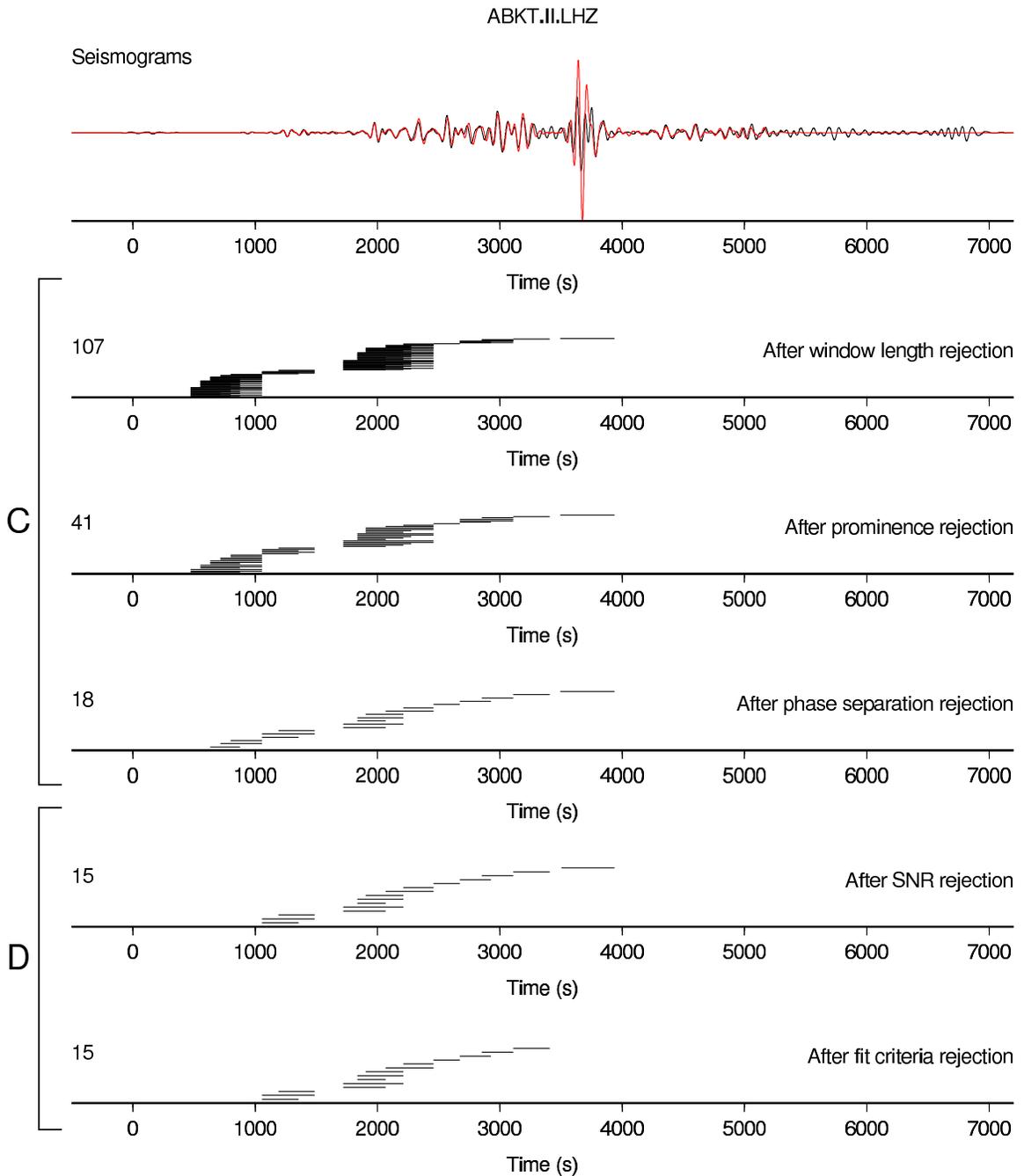


Figure 5.2: Window rejection applied to real data (*Maggi et al.*, 2009, Figure 4). Top panel: observed (black) and synthetic (red) seismograms for the 050295B event recorded at ABKT (see Figure 5.1). Subsequent panels: candidate windows at different stages, separated into Stage C (shape based rejection) and Stage D (fit based rejection). Each candidate window is indicated by a black segment. The number of windows at each stage is shown to the left of the panel.

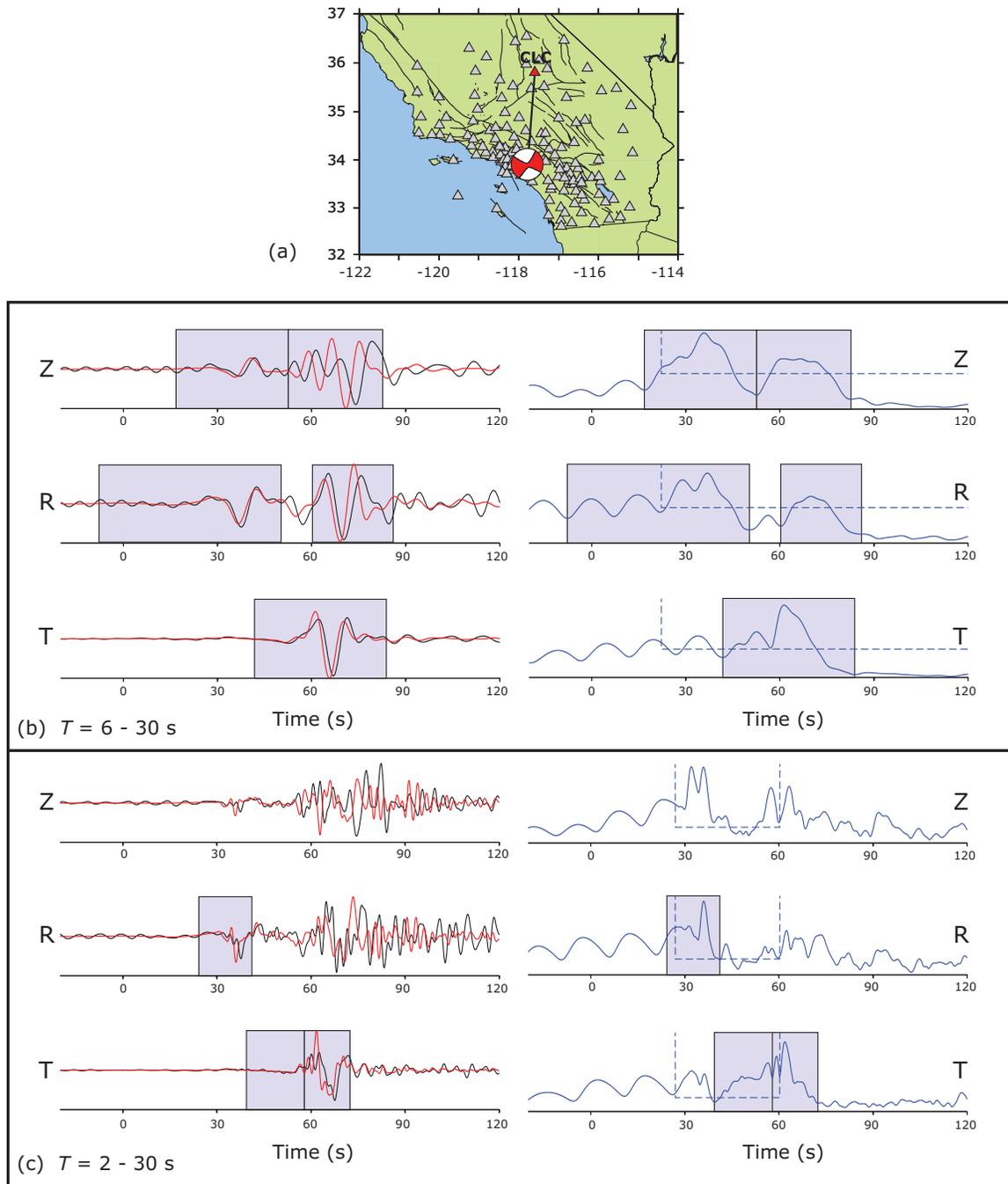


Figure 5.3: Window selection results for event 9818433 from Table 5.3 recorded at station CLC ( $\Delta = 211.7$  km) (*Maggi et al.*, 2009, Figure 15). (a) Map showing all stations with at least one measurement window for the period range 6–30 s for this event. Red triangle denotes station CLC. (b) Results for station CLC for the period range 6–30 s. Vertical (Z), radial (R), and transverse (T) records of data (black, left column) and synthetics (red, left column), as well as the STA:LTA records (right column) used to produce the window picks. (c) Results for station CLC for the period range 2–30 s. Note that corresponding lower-passed filtered versions are shown in (b).

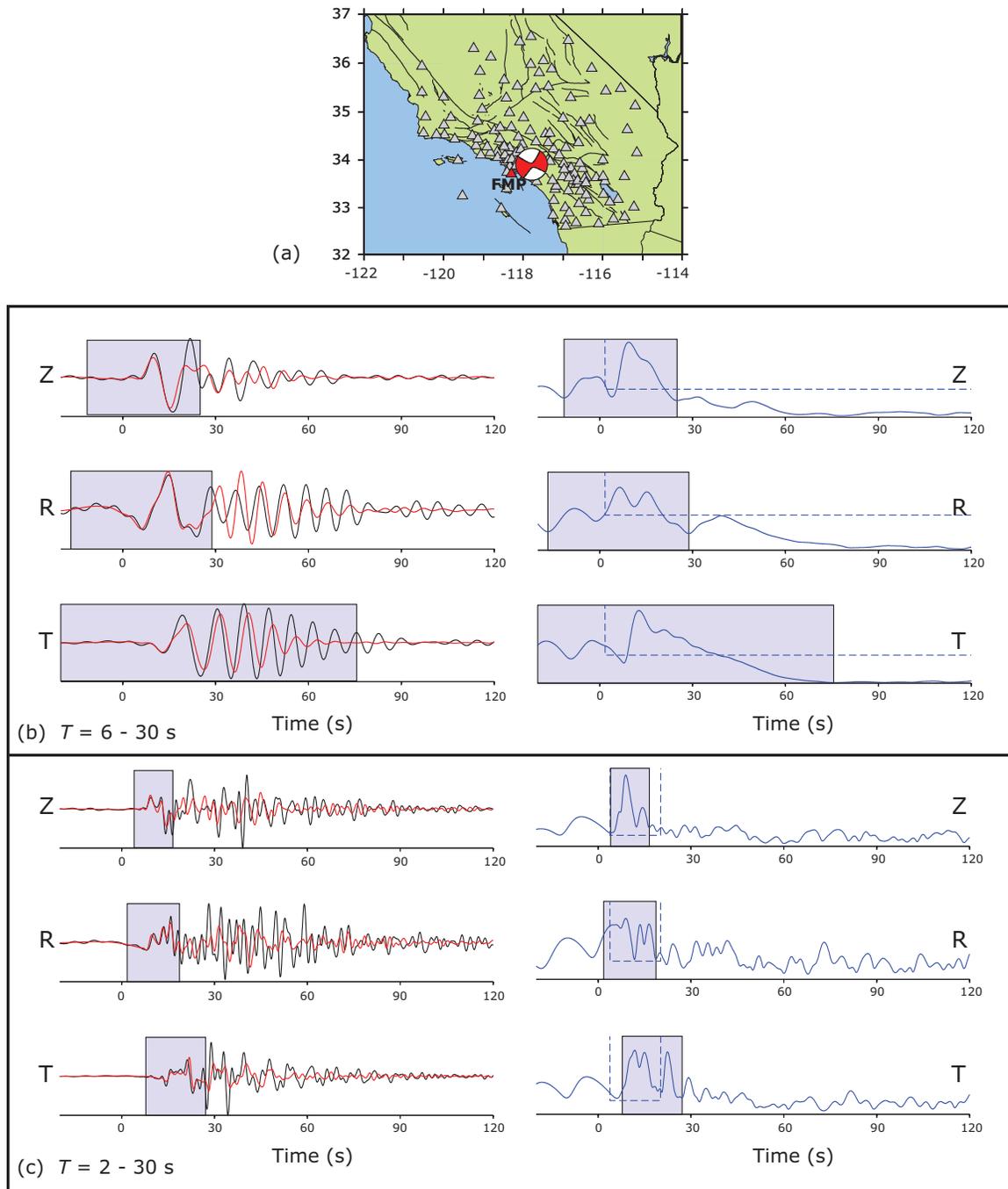


Figure 5.4: Window selection results for event 9818433 from Table 5.3 recorded at station FMP ( $\Delta = 52.2$  km) (Maggi *et al.*, 2009, Figure 16). Same caption as Figure 5.3, but for a different station.

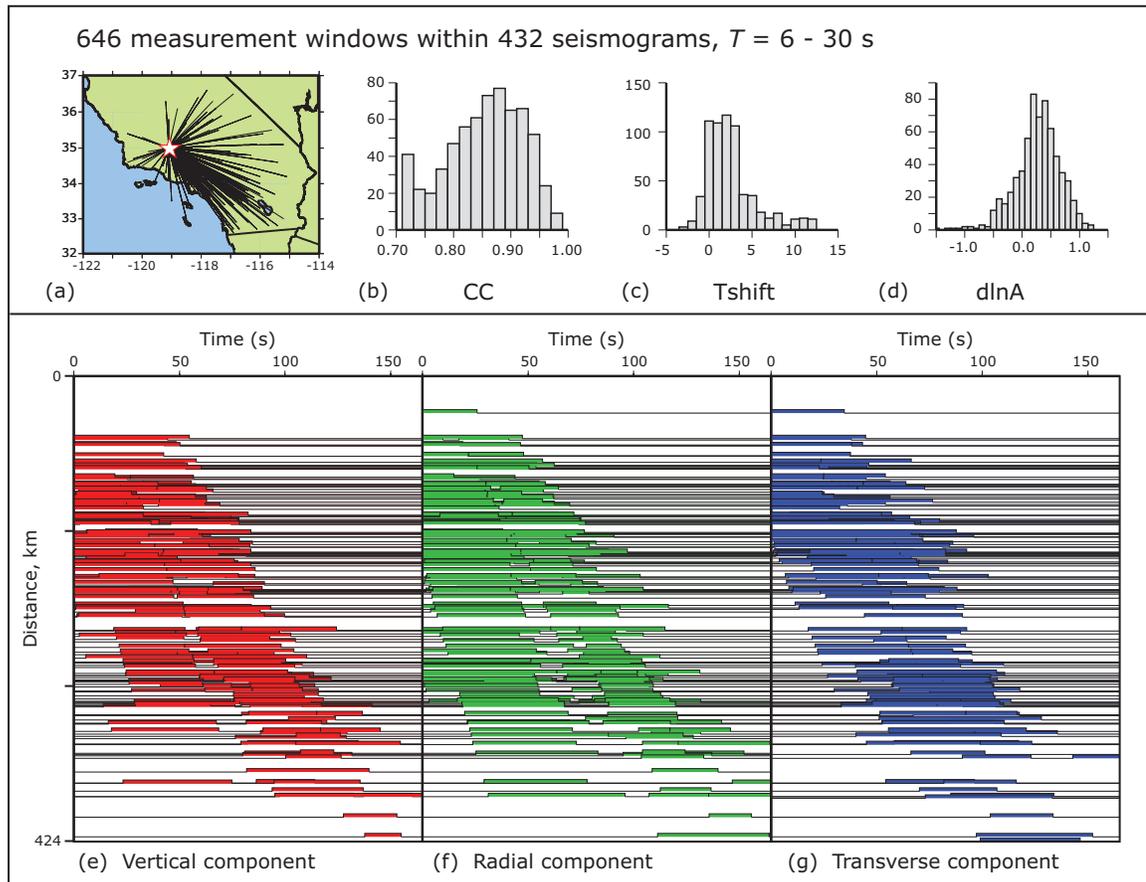


Figure 5.5: Summary plots of windowing results for event 9983429 in Table 5.3, for the period range 6–30 s (*Maggi et al.*, 2009, Figure 17). (a) Map showing paths to each station with at least one measurement window. (b)-(d) Histograms of number of windows as a function of normalized cross-correlation  $CC$ , time lag  $\tau$  and amplitude ratio  $\Delta \ln A$ . (e)-(g) Record sections of selected windows for the vertical, radial and transverse components. The two branches observed on the vertical and radial components correspond to the body-wave arrivals and the Rayleigh wave arrivals.

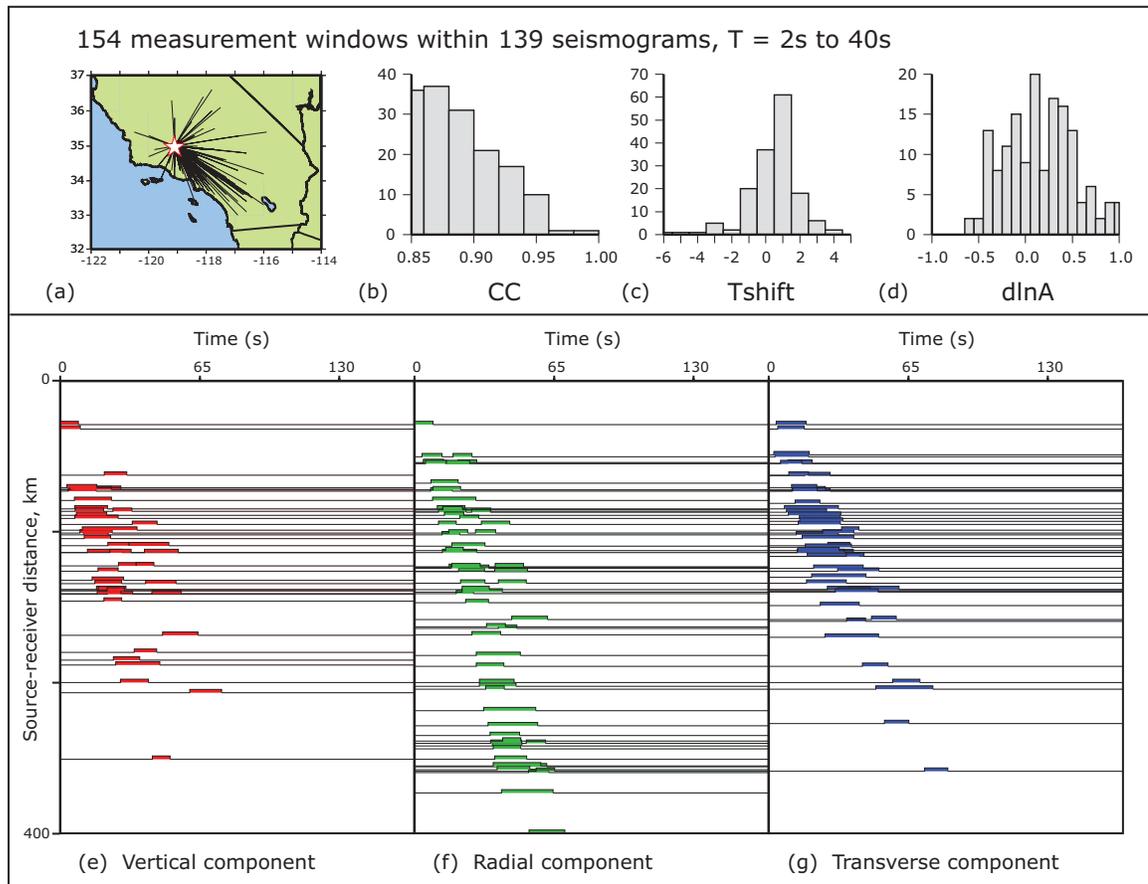


Figure 5.6: Summary plots of windowing results for event 9983429 in Table 5.3, for the period range 2–30 s. Same as Figure 5.5, but the windowing code has been run using a different set of parameters (Table 5.2), so that primarily only the body-wave arrivals are selected.