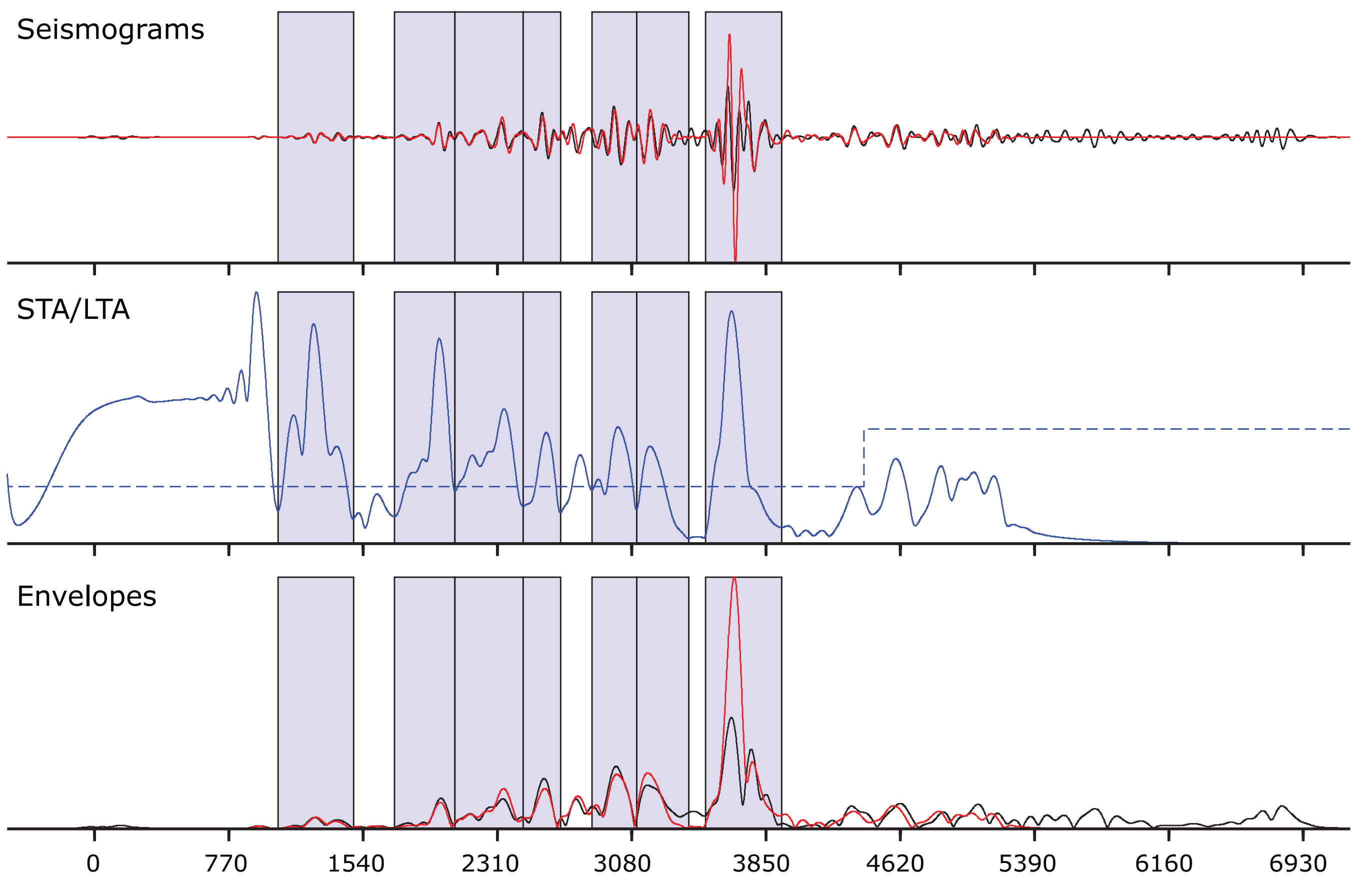
FLEXWIN User's Manual

Alessia Maggi



Contents

Chapter 1

Introduction

The FLEXWIN software package automates the time-window selection problem for seismologists. It operates on pairs of observed and synthetic single component seismograms, defining windows that cover as much of a given seismogram as possible, while avoiding portions of the waveform that are dominated by noise.

FLEXWIN selects time windows on the synthetic seismogram within which the waveform contains a distinct energy arrival, then requires an adequate correspondence between observed and synthetic waveforms within these windows.

There is no restriction on the type of simulation used to generate the synthetics. Realistic Earth models and more complete wave propagation theories yield waveforms that are more similar to the observed seismograms, and thereby promote the selection of measurement windows covering more of the available data. The input seismograms can be measures of displacement, velocity or acceleration. There is no requirement for horizontal signals to be rotated into radial and transverse directions.

FLEXWIN is a configurable data selection process that can be adapted to different tomographic scenarios by tuning a handful of parameters. Although the algorithm was designed for use in 3D-3D adjoint tomography, its inherent flexibility should make it useful in many data-selection applications.

For a detailed introduction to FLEXWIN as applied to seismic tomography, please consult Maggi et al. [2009]. If you use FLEXWIN for your own research, please cite Maggi et al. [2009].

Chapter 2

Getting started

Here is where you find basic information for obtaining and installing the FLEXWIN package. For details of how to tune the algorithm to your seismograms, see chapter 3.

2.1 System requirements

In order to install and run, FLEXWIN requires:

- UNIX operating system (Linux, Solaris, MacOS . . .)
- a fortran compiler (gfortran, ifort, etc...)
- other packages : SAC (Seismic Analysis Code, available from IRIS); GMT (Generic Mapping Tools) for the plotting scripts

FLEXWIN requires the following libraries external to the package in order to compile and run: `libsacio.a` and `libsac.a`. Both libraries are distributed by IRIS as part of the SAC package (version 101.2 and above). The IRIS download site (as of 30-March-2009) is here: `http://www.iris.edu/software/sac/sac.request.htm`. (To check your version, type sac.)

2.2 Obtaining the code

The code is available as a gzipped tarball from CIG (Computational Infrastructure for Geodynamics, `http://www.geodynamics.org`). The tarball is unpacked by typing `tar xvzf flexwin.tgz`.

The package contains the flexwin code and documentation, as well as a set of test data, examples of user files for different scenarios, and a set of utility scripts that may be useful for running flexwin on large datasets.

2.3 Compilation

If your compiler of choice is gfortran, then you should be able to use the `make_gfortran` makefiles with only minor modifications (notably you may need to change the search path for the `libsacio.a` library). If you prefer another compiler, you should modify the OPT and FC lines in the makefiles accordingly. We tested the code using gfortran version 4.1.2 (To check your version, type `gfortran --version`.)

Important note: All the code is compiled with the -m32 option, which makes 32bit binaries. This option is currently required to enable compatibility with SAC. Future versions of the SAC distribution may no longer require this compatibility flag.

Steps to compile the flexwin package:

1. Compile `libtau.a` and create `iasp91.hed` and `iasp91.tbl`. In the `flexwin/ttimes_mod` directory type: `make -f make_gfortran`. This will compile `libtau.a`, and two programs, `remodl` and `setbrn`. The makefile will also run `remodl` and `setbrn` to create the `iasp91.hed`and `iasp91.tbl` files. You should then type `make -f make_gfortran install` to install the iasp91 files.
2. Compile flexwin. Edit the `make_gfortran` file in the flexwin root directory to ensure the `SACLIBDIR` environment variable points to the location of your SAC libraries (by default `$SACHOME/lib`). Then type `make -f make_gfortran`.

You should end up with the `flexwin` executable. The program requires the `iasp91.hed` and iasp91.tbl files (or symbolic links to them) to be present in the directory from which the code is launched.

2.4 Running the Test case

You should test your compiled code on the test_data dataset provided. In the flexwin/test_data directory, type ./flexwin < input.test. The results of your run will be found in the MEASURE subdirectory, and should match those found in the MEASURE.orig subdirectory.

You can also test the basic plotting script by running ./plot_seismos_gmt.sh MEASURE/ABKT.II.LHZ, whose output will be MEASURE/ABKT.II.LHZ.seis.eps. file. Your result should be identical to what is shown in Figure 2.1.

2.5 Running FLEXWIN

In general, flexwin is run as follows: ./flexwin < input where the input file is formatted as follows:

```
327
RAW_DATA/9627721.CI.ADO.BHR.sac.d.fil
SYNTH/ADO.CI.BHR.new.fil
MEASURE/ADO.CI.BHR
RAW_DATA/9627721.CI.ADO.BHT.sac.d.fil
SYNTH/ADO.CI.BHT.new.fil
MEASURE/ADO.CI.BHT
RAW_DATA/9627721.CI.ADO.BHZ.sac.d.fil
SYNTH/ADO.CI.BHZ.new.fil
MEASURE/ADO.CI.BHZ
....
```

i.e. the number of traces to be measured, followed by (in order) the path to the raw data sac file, the path to the synthetic sac file and the path and basename for the (many!) output files for that trace.

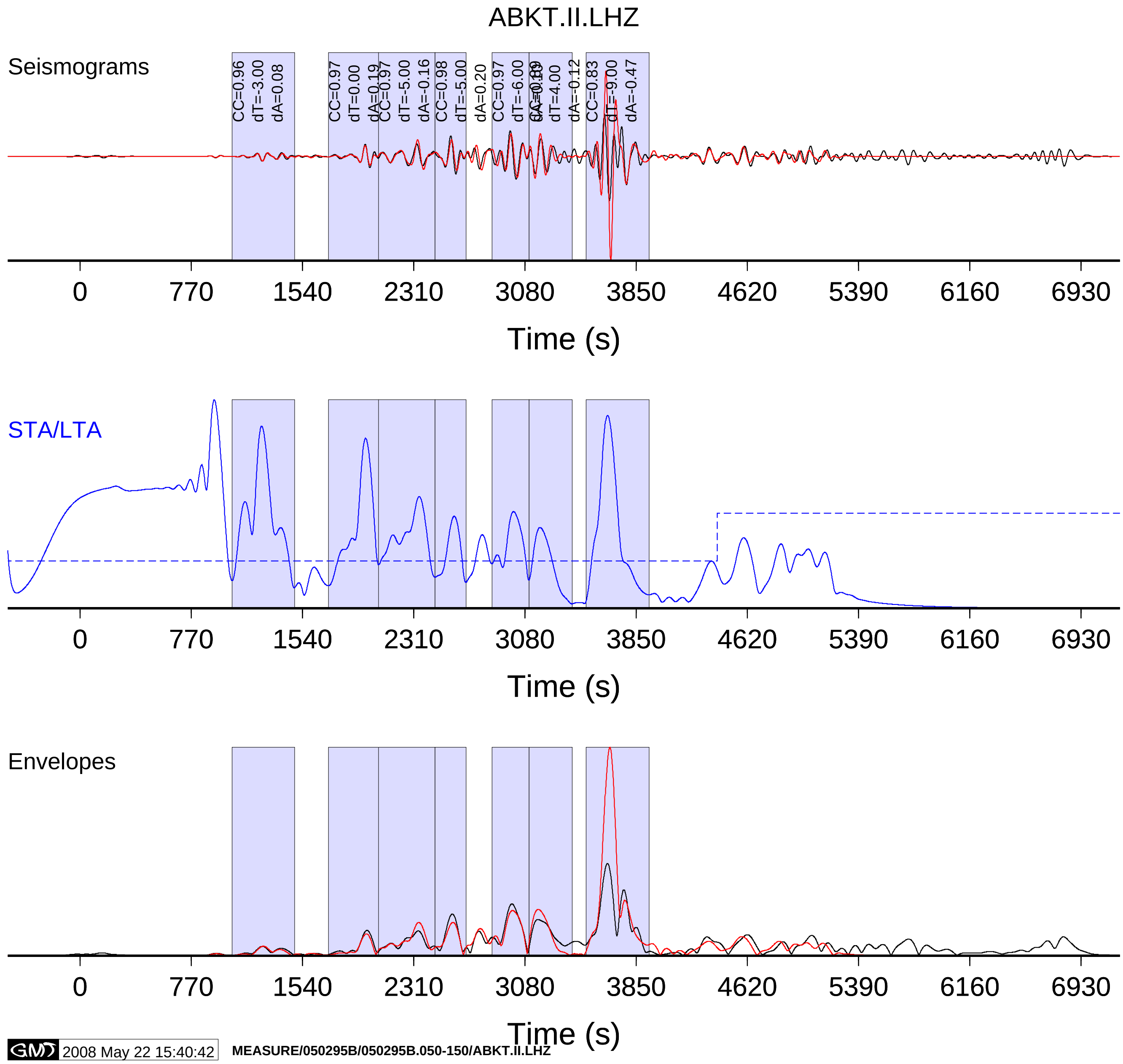


Figure 2.1: Windowing results for the test data set, plotted using the `./plot_seismos_gmt.sh` script.

2.6 Output files

Most output files are in ascii. All file names start with the basename given in the input file for that trace:

basename.obs ascii observed seismogram (filtered)

basename.syn ascii synthetic seismogram (filtered)

basename.obs_lp.sac sac observed seismogram (filtered)

basename.syn_lp.sac sac synthetic seismogram (filtered)

basename.env.obs ascii envelope of observed seismogram (filtered)

basename.env.syn ascii envelope of synthetic seismogram (filtered)

basename.win list of windows with theoretical phase arrival times

basename.win.qual list of windows with Tshift,CC,dlnA values

basename.phases theoretical arrival times of phases

basename.stalta STA:LTA timeseries used to select the windows, and the time-dependent values of the STA:LTA water level, the cross-correlation limit CC_0, the time-lag limit $\Delta\tau_0$, amplitude ratio limit $\Delta \ln A_0$ and the window signal to noise limit r_0.

basename.info information on the path and some statistics

For more details about the file formats, read the `write_` subroutines in `io_subs.f90`.

2.7 Scripts

Several plotting routines (`plot_*.sh`) are provided in the `scripts` subdirectory as examples for plotting seismograms, measurements and adjoint sources. All plotting is done using GMT (Generic Mapping Tools). These scripts will need to be modified to suit your particular plotting needs.

The script `extract_event_windowing_stats.sh` extracts statistical information on the window selection process, on the measurements. Again, you can use use this script as a template for your own information extraction needs.

2.8 Pre-processing suggestions

Pre-processing is a subtle procedure that can affect the selection of windows. Here we list some suggestions for pre-processing:

1. Interpolate raw data and “raw” synthetics using the same time-step.
2. Cut the data and synthetic seismograms based on the data record. If the data record starts before the synthetic record (which is common for local earthquakes), then consider padding zeros before the synthetic record; information prior to the origin time (and P arrival) is useful in assessing the signal-to-noise ratio in the observed records.
3. Bandpass both the data and synthetics over the desired period range, rather than doing this within FLEXWIN, thereby limiting the number of filtering operations. However, for initial determination of the bandpass range of interest, it is simpler to experiment using the FLEXWIN parameters `WIN_MIN_PERIOD` and `WIN_MAX_PERIOD`.

Chapter 3

Tuning FLEXWIN for your seismograms

FLEXWIN is adapted to your specific problem by modifying the values of the parameters in Table 3.1, and the functional form of those parameters that are time-dependent. We consider the algorithm to be correctly adapted when false positives (windows around undesirable features of the seismogram) are minimized, and true positives (window around desirable features) are maximized. The choice of what makes an adequate set of windows remains subjective, as it depends strongly on the quality of the input model, the quality of the data, and the region of the Earth that the tomographic inversion aims to constrain.

The base values of the various parameters are set in the `PAR_FILE`, which is read at run time. Examples of base parameter values for the three tomographic scenarios discussed by Maggi et al. [2009] can be found in Table 3.2. The functional forms of the time dependent parameters may be adjusted by modifying `user_parameters.f90` (see next section), and re-compiling the code.

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
$\mathrm{CC}_0(t)$	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
$\Delta\tau_{\mathrm{ref}}$	reference time lag
$\Delta \ln A_{\mathrm{ref}}$	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 un-prominent 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_{len} w_{nwin}	for selection of best non-overlapping window combination

Table 3.1: 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.

3.1 User parameters

The main user parameters in the `PAR_FILE` are:

`WIN_MIN_PERIOD` Corresponds to T_0 in Table 3.1, and is the short wavelength cut-off for the band-pass filter applied to the raw synthetic and observed seismograms.

`WIN_MAX_PERIOD` Corresponds to T_1 in Table 3.1, and is the long wavelength cut-off for the band-pass filter applied to the raw synthetic and observed seismograms.

`SNR_INTEGRATE_BASE` Corresponds to r_P in Table 3.1, and is the minimum signal to noise ratio on the power of the observed seismogram for windowing to continue.

`SNR_MAX_BASE` Corresponds to r_A in Table 3.1, and is the minimum signal to noise ratio on the modulus of the observed seismogram for windowing to continue.

`WINDOW_S2N_BASE` Corresponds to r_0 in Table 3.1, and is the minimum signal to noise ratio for a window on the observed seismogram to be acceptable.

`STALTA_BASE` Corresponds to w_E in Table 3.1, and is the water level to be applied to the synthetic short-term/long-term average waveform in order to generate candidate time windows. See Figure 3.1a.

`CC_BASE` Corresponds to CC_0 in Table 3.1, and is the minimum normalized cross-correlation value between synthetic and observed seismogram for a window to be acceptable.

`TSHIFT_BASE` Corresponds to $\Delta\tau_0$ in Table 3.1, and is the maximum cross-correlation lag (in seconds) between synthetic and observed seismogram for a window to be acceptable.

`DLNA_BASE` Corresponds to $\Delta \ln A_0$ in Table 3.1, and is the maximum amplitude ratio ($\Delta \ln A$ or $\Delta A/A$) between synthetic and observed seismogram for a window to be acceptable.

`TSHIFT_REFERENCE` Corresponds to $\Delta\tau_{\mathrm{ref}}$ in Table 3.1, and allows for a systematic traveltime bias in the synthetics.

`TSHIFT_REFERENCE` Corresponds to $\Delta \ln A_{\mathrm{ref}}$ in Table 3.1, and allows for a systematic amplitude bias in the synthetics.

`C_0` Corresponds to C_0 in Table 3.1, and is expressed as a multiple of w_E. No window may contain a local minimum in its STA:LTA waveform that falls below the local value of $C_0 w_E$. See Figure 3.1b.

`C_1` Corresponds to C_1 in Table 3.1, and is expressed as a multiple of T_0. No window may be shorter than $C_1 T_0$.

`C_2` Corresponds to C_2 in Table 3.1, and is expressed as a multiple of w_E. A window whose seed maximum on the STA:LTA waveform rises less than $C_2 w_E$ above either of its adjacent minima is rejected. See Figure 3.1c.

`C_3a` Corresponds to C_{3a} in Table 3.1, and is expressed as a fraction. It regulates the acceptable height ratio between local maxima in a given window on the STA:LTA waveform. See Figure 3.2.

`C_3b` Corresponds to C_{3b} in Table 3.1, and is expressed as a multiple of T_0. It regulates the acceptable time separation between local maxima in a given window on the STA:LTA waveform. See Figure 3.2.

`C_4a` Corresponds to C_{4a} in Table 3.1, and is expressed as a multiple of T_0. It limits the length of a window before its first local maximum in STA:LTA.

`C_4b` Corresponds to C_{4b} in Table 3.1, and is expressed as a multiple of T_0. It limits the length of a window beyond its last local maximum in STA:LTA. See Figure 3.1d.

`WEIGHT_AVERAGE_CC` Corresponds to w_{CC} in Table 3.1, and is the weight given to the average cross-correlation value in the process of resolving window overlaps. See Figure 3.3.

`WEIGHT_SPACE_COVERAGE` Corresponds to w_{len} in Table 3.1, and is the weight given to the time span covered by windows in the process of resolving window overlaps.

	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	0.11	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
$\Delta\tau_{\mathrm{ref}}$	0.0	0.0	0.0	4.0	2.0	1.0
$\Delta \ln A_{\mathrm{ref}}$	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
$w_{\mathrm{CC}}, w_{\mathrm{len}}, w_{\mathrm{nwin}}$	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 3.2: Values of standard and fine-tuning parameters for the three seismological scenarios discussed Maggi et al. [2009]. This table is identical to Table 3 of that study.

`WEIGHT_N_WINDOWS` Corresponds to w_{nwin} in Table 3.1, and is the weight given to the total number of windows in the process of resolving window overlaps.

3.2 Time dependence of user parameters

A subset of the FLEXWIN parameters from Table 3.1 are time-dependent (where time is measured along the seismogram). This feature enables the user to exercise fine control of the windowing algorithm. The user can modulate the time-dependence of these parameters by editing the `set_up_criteria_arrays` subroutine in the `user_functions.f90` file. This subroutine is called after the seismograms have been read in, and the following variables have been set:

`npts, dt, b` Number of points, time step, time of first point with respect to the reference time of both seismograms. The observed and synthetic seismograms should have identical values of these three quantities.

`evla, evlo, evdp, stla, stlo` Event latitude, event longitude, event depth (km), station latitude, station longitude, read from the observed seismogram.

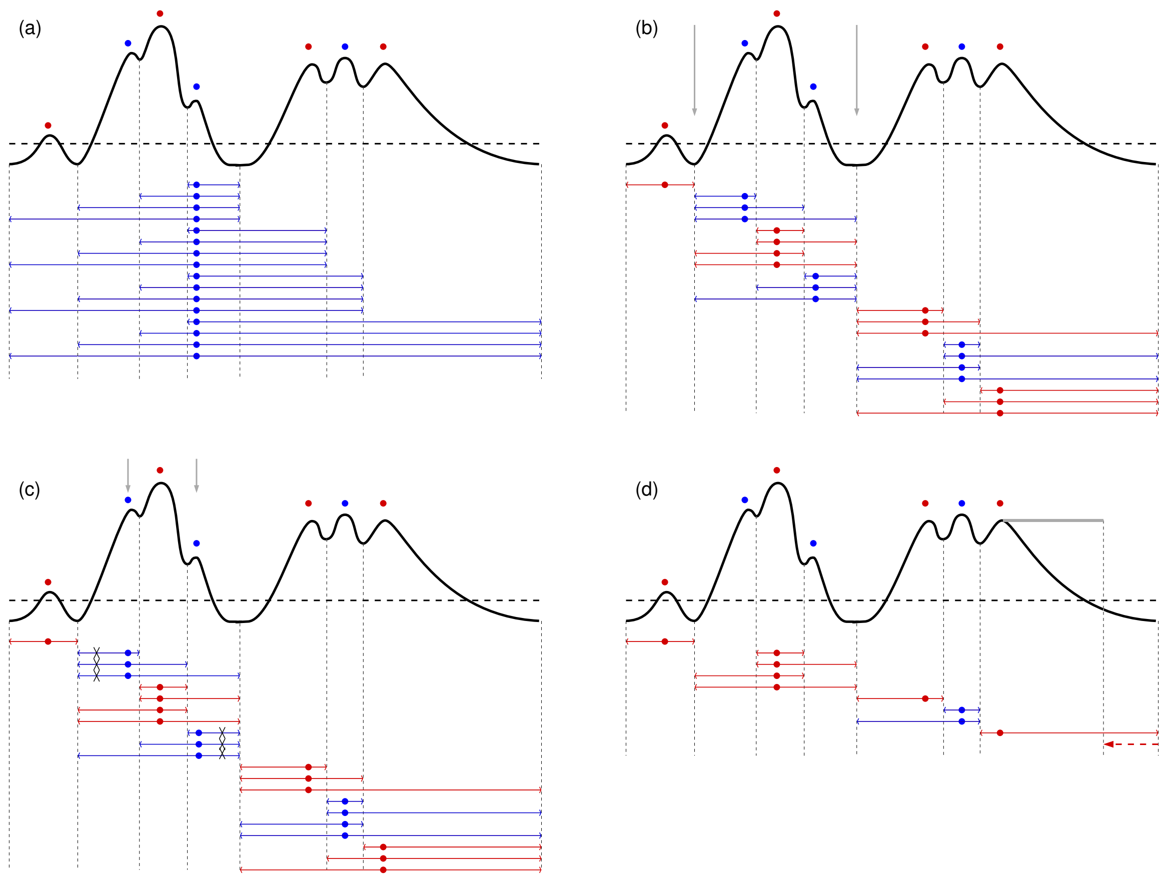


Figure 3.1: (a) Window creation process. The thick black line represents the STA:LTA waveform $E(t)$, and the thick horizontal dashed line its water level $w_E(t)$. Local maxima are indicated by alternating red and blue dots, windows are indicated by two-headed horizontal arrows. The time of the local maximum used as the window seed t_M is denoted by the position of the dot. Only windows for the fourth local maximum are shown. (b) Rejection of candidate windows based on the amplitude of the local minima. The two deep local minima indicated by the grey arrows form virtual barriers. All candidate windows that cross these barriers are rejected. (c) Rejection of candidate windows based on the prominence of the seed maximum. The local maxima indicated by the grey arrows are too low compared to the local minima adjacent to them. All windows that have these local maxima as their seed are rejected (black crosses over the window segments below the time series). (d) Shortening of long coda windows. The grey bar indicates the maximum coda duration $c_{4b}T_0$. Note that after the rejection based on prominence represented in (c) and before shortening of long coda windows represented in (d), the algorithm rejects candidate windows based on the separation of distinct phases, a process that is illustrated in Figure 3.2.

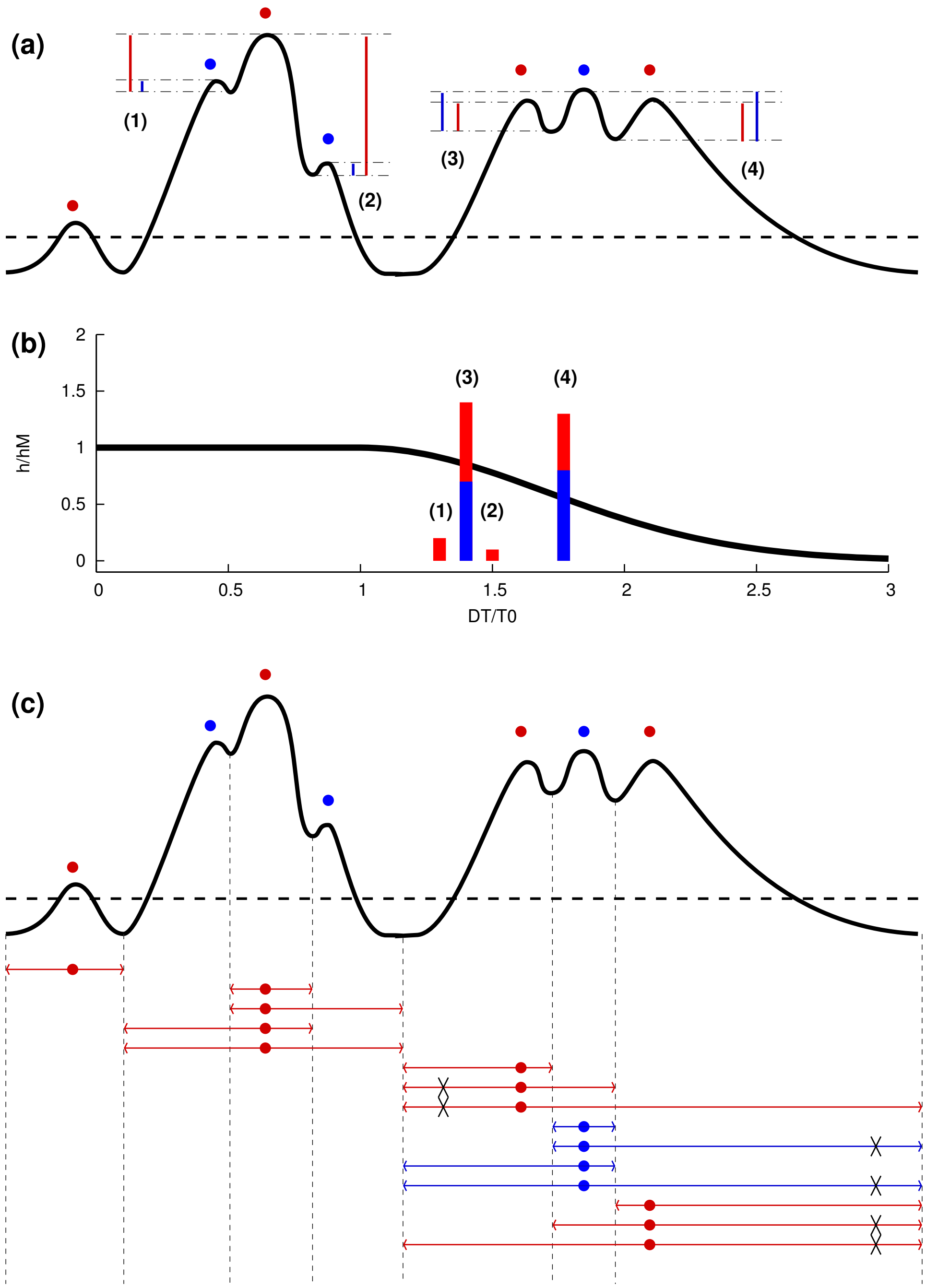


Figure 3.2: Rejection of candidate windows based on the separation of distinct phases. (a) Heights of pairs of local maxima above their intervening minimum. (b) The black line represents the limiting value of h/h_M. Vertical bars represent h/h_M for each pair of maxima. Their position along the horizontal axis is given by the time separation ΔT between the maxima of each pair. The color of the bar is given by the color of the seed maximum corresponding to h_M. Bars whose height exceeds the line represent windows to be rejected. (c) The windows that have been rejected by this criterion are indicated by black crosses.

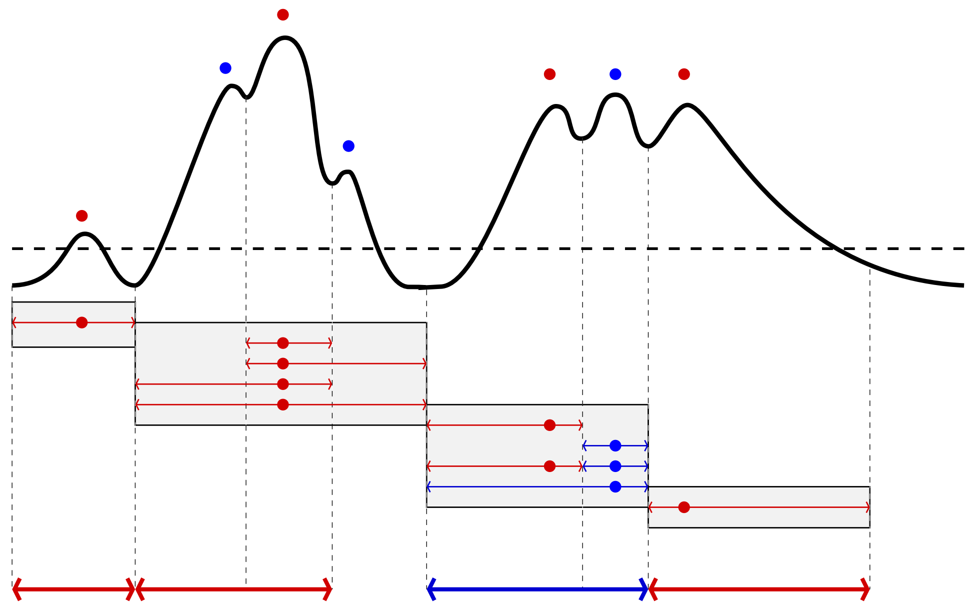


Figure 3.3: The selection of the best non-overlapping window combinations. Each grey box represents a distinct group of windows. Non-overlapping subsets of windows are shown on separate lines. Only one line from within each group will be chosen, the one corresponding to the highest weighted score. The resulting optimal set of data windows is shown by thick arrows.

`azimuth, backazimuth, dist_deg, dist_km` Calculated from the event and station locations above.

`kstnm, knetwk, kcmpnm` Station name, network name, component name, read from the observed seismogram.

`num_phases, ph_names, ph_times` Number, names and arrival times of standard seismic phases calculated through IASPEI91 using the event depth and epicentral distance.

The `set_up_criteria_arrays` subroutine first sets up non time-modulated versions of the time-dependent parameters using a simple loop over time:

```
! --------------------------------------------------------------------
! This is the basic version of the subroutine - no variation with time
! --------------------------------------------------------------------
  do i = 1, npts
    time=b+(i-1)*dt
    DLNA_LIMIT(i)=DLNA_BASE
    CC_LIMIT(i)=CC_BASE
    TSHIFT_LIMIT(i)=TSHIFT_BASE
```

```
    STALTA_W_LEVEL(i)=STALTA_BASE
    S2N_LIMIT(i)=WINDOW_AMP_BASE
  enddo
```

It is then up to the user to modulate these values for the specific problem at hand. For example, should the user want to discourage the algorithm from picking windows beyond the end of the first surface wave-train $R1$, the following lines should be added to the `user_functions.f90` file to raise the water level on the STA:LTA waveform by a factor of ten:

```
! --------------------------------
! Set approximate end of rayleigh wave arrival
R_vel=3.2
R_time=dist_km/R_vel
! --------------------------------
! modulate criteria in time
do i = 1, npts
  time=b+(i-1)*dt
  if (time.gt.R_time) then
    STALTA_W_LEVEL(i)=STALTA_BASE*10.0
  endif
enddo
```

Should the user want the algorithm to be less stringent in its requirements for the cross-correlation fit for the surface wave portion of the seismograms, and allow a greater travel-time lag for deeper events, the required lines could be:

```
! --------------------------------
! Set approximate end of rayleigh wave arrival
R_vel=3.2
R_time=dist_km/R_vel
! --------------------------------
! Set approximate start of love wave arrival
Q_vel=4.2
Q_time=dist_km/Q_vel
! --------------------------------
! modulate criteria in time
do i = 1, npts
  time=b+(i-1)*dt
  ! if we are in the surface wave times, then make the cross-correlation
  ! criterion less severe
  if (time.gt.Q_time .and. time.lt.R_time) then
```

```
    CC_LIMIT(i)=0.9*CC_LIMIT(i)
  endif
  ! --------------------------------
  ! modulate criteria according to event depth
  !
  ! if an intermediate depth event
  if (evdp.ge.70 .and. evdp.lt.300) then
    TSHIFT_LIMIT(i)=TSHIFT_BASE*1.4
  ! if a deep event
  elseif (evdp.ge.300) then
    TSHIFT_LIMIT(i)=TSHIFT_BASE*1.7
  endif
enddo
```

The above examples illustrate the power of the `user_functions.f90` file. The user can choose to include/exclude any portion of the seismogram, and to make the rejection criteria for windows more or less stringent on any other portion of the seismogram. All the seismogram-dependent variables whose values are known when the `set_up_criteria_arrays` subroutine is executed may be used to inform these choices, leading to an infinite number of windowing possibilities. The careful user will use knowledge of the properties of the observed data set, the limitations of the synthetic waveforms, and the final use to which the selected windows will be put in order to tailor the subroutine to the needs of each study.

For a given set of data and synthetics, the `PAR_FILE` and `user_functions.f90` files uniquely determine the windowing results.

3.2.1 Examples of user functions

Here we present the time dependencies of tuning parameters used for three tomographic scenarios [Maggi et al., 2009]: global, Japan and southern California. In each example 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.

Global scenario

In the following, h indicates earthquake depth, t_Q indicates the approximate start of the Love wave predicted by a group wave speed of 4.2 km s^{-1}, and t_R indicates the approximate end of the Rayleigh wave predicted by a group wave speed of 3.2 km s^{-1}. In order to reduce the number of windows picked beyond R1, and to ensure that those selected beyond R1 are a very good match to the synthetic waveform, we raise the water level on the STA:LTA waveform and impose stricter criteria on the signal-to-noise ratio and the waveform similarity after the approximate end of the surface-wave arrivals. We allow greater flexibility in cross-correlation time lag $\Delta\tau$ for intermediate depth and deep earthquakes. We lower the cross-correlation value criterion for surface-waves in order to retain windows with a slight mismatch in dispersion characteristics.

We therefore use the following time modulations:

$$w_E(t) = \begin{cases} w_E t \leq t_R, \\ 2w_E t > t_R, \end{cases} \tag{3.1}$$

$$r_0(t) = \begin{cases} r_0 & t \leq t_R, \\ 10r_0 & t > t_R, \end{cases} \tag{3.2}$$

$$\mathrm{CC}_0(t) = \begin{cases} \mathrm{CC}_0 & t \leq t_R, \\ 0.9\mathrm{CC}_0 & t_Q < t \leq t_R, \\ 0.95 & t > t_R, \end{cases} \tag{3.3}$$

$$\Delta\tau_0(t) = \begin{cases} \begin{cases} \tau_0 & t \leq t_R, \\ \tau_0/3 & t > t_R, \end{cases} & h \leq 70 \text{ km} \\ 1.4\tau_0 & 70 \text{ km} < h < 300 \text{ km}, \\ 1.7\tau_0 & h \geq 300 \text{ km}, \end{cases} \tag{3.4}$$

$$\Delta \ln A_0(t) = \begin{cases} \Delta \ln A_0 & t \leq t_R, \\ \Delta \ln A_0/3 & t > t_R. \end{cases} \tag{3.5}$$

Japan scenario

In the following, t_P and t_S denote the start of the time windows for P- and S waves, as predicted by the 1-D IASPEI91 model [Kennett and Engdahl, 1991], and t_{R1} indicates the end of the surface-wave time window. For the 24–120 s data, we consider the waveform between the start of the P wave to the end of the surface-wave. We therefore modulate $w_E(t)$ as follows:

$$w_E(t) = \begin{cases} 10w_E & t < t_P, \\ w_E & t_P \leq t \leq t_{R1}, \\ 10w_E & t > t_{R1}. \end{cases} \tag{3.6}$$

For the 6–30 s data, the fit between the synthetic and observed surface-waves is expected to be poor, as the 3D model used to calculate the synthetics cannot produce the required complexity. We therefore want to concentrate on body-wave arrivals only, and avoid surface-wave windows altogether by modulating $w_E(t)$ as follows:

$$w_E(t) = \begin{cases} 10w_E & t < t_P, \\ w_E & t_P \leq t \leq t_S, \\ 10w_E & t > t_S. \end{cases} \tag{3.7}$$

We use constant values of $r_0(t) = r_0$, $\mathrm{CC}_0(t) = \mathrm{CC}_0$ and $\Delta \ln A_0(t) = \Delta \ln A_0$ for both period ranges. In order to allow greater flexibility in cross-correlation time lag $\Delta\tau$ for intermediate depth and deep earthquakes we use:

$$\Delta\tau_0(t) = \begin{cases} 0.08t_P & h \leq 70 \text{ km}, \\ \max(0.05t_P, 1.4\tau_0) & 70 \text{ km} < h < 300 \text{ km}, \\ \max(0.05t_P, 1.7\tau_0) & h \geq 300 \text{ km}. \end{cases} \tag{3.8}$$

Southern California scenario

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 Δ.

For the 6–30 s and 3–30 s data, we use constant values of $r_0(t) = r_0$, $\mathrm{CC}_0(t) = \mathrm{CC}_0$, $\Delta\tau_0(t) = \Delta\tau_0$, 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 $w_E(t)$:

$$w_E(t) = \begin{cases} 10w_E & t < t_P, \\ w_E & t_P \leq t \leq t_{R1}, \\ 10w_E & t > t_{R1}, \end{cases} \tag{3.9}$$

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:

$$w_E(t) = \begin{cases} 10w_E & t < t_P, \\ w_E & t_P \leq t \leq t_S, \\ 10w_E & t > t_S. \end{cases} \tag{3.10}$$

3.3 Tuning considerations

FLEXWIN is not a black-box application and should not 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, the quality of the model, and the accuracy of the method used to calculate the synthetic seismograms. The user must configure and tune the algorithm for the given problem. Here we discuss general considerations the user should bear in mind.

We suggest the following as a practical starting sequence for tuning the algorithm. Keep in mind that this process may need to be repeated and refined several times before converging on the optimal set of parameters for a given problem and data-set.

$T_{0,1}$: 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 data-set, 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.

$w_E(t)$: For a constant signal the short-term average long-term average ratio, $E(t)$, converges to a constant value when the length of the time-series is greater than the effective averaging length of the long-term average. This value is 0.08 for the short-term average long-term average ratio used in FLEXWIN (it has a small dependence on T_0, which can be ignored in most applications). 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 $r0(t)$, $CC_0(t)$, $\Delta T_0(t)$ and $\Delta \ln A_0(t)$ have been configured.

$r_0(t)$, $\mathrm{CC}_0(t)$, $\Delta\tau_{\mathrm{ref}}$, $\Delta\tau_0(t)$, $\Delta \ln A_{\mathrm{ref}}$ 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. 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_{diff}*, reduce* $r_0(t)$). $\Delta\tau_{\mathrm{ref}}$ and $\Delta \ln A_{\mathrm{ref}}$ 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 control the process by 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 3.2). 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_{CC}, w_{len} and w_{nwin} : These parameters control the overlap resolution stage of the algorithm. Values of $w_{\mathrm{CC}} = w_{\mathrm{len}} = w_{\mathrm{nwin}} = 1$ should be reasonable for most applications.

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

Chapter 4

Miscellaneous

4.1 Bug reports and suggestions for improvements

To report bugs or suggest improvements to the code, please send an email to the CIG Computational Seismology Mailing List (cig-seismo@geodynamics.org) or Alessia Maggi (alessia@sismo.u-strasbg.fr), and/or use our online bug tracking system Roundup (www.geodynamics.org/roundup).

4.2 Notes and Acknowledgments

The main developers of the FLEXWIN source code are Alessia Maggi and Carl Tape. The following individuals (listed in alphabetical order) have also contributed to the development of the source code: Daniel Chao, Min Chen, Vala Hjorleifsdottir, Qinya Liu, Jeroen Tromp. The following individuals (listed in alphabetical order) contributed to this manual: Sue Kientz, Alessia Maggi, Carl Tape.

The FLEXWIN code makes use of filtering and enveloping algorithms that are part of SAC (Seismic Analysis Code, Lawerence Livermore National Laboratory) provided for free to IRIS members. We thank Brian Savage for adding interfaces to these algorithms in recent SAC distributions.

We acknowledge support by the National Science Foundation under grant EAR-0711177. Daniel Chao received additional support from a California Institute of Technology Summer

Undergraduate Reseach Fellowship.

4.3 Copyright

Bibliography

B. L. N. Kennett and E. R. Engdahl. Traveltimes for global earthquake location and phase identification. *Geophys. J. Int.*, 105:429–465, 1991.

D. Komatitsch, Q. Liu, J. Tromp, P. Süss, C. Stidham, and Shaw J.H. Simulations of ground motion in the Los Angeles basin based upon the spectral-element method. *B. Seismol. Soc. Am.*, 94:187–206, 2004.

A. Maggi, C. Tape, M. Chen, D. Chao, and J. Tromp. An automated time-window selection algorithm for seismic tomography. *Geophys. J. Int.*, 2009. (in press).

L. A. Wald, L. K. Hutton, and D. D. Given. The Southern California Network Bulletin: 1990–1993 summary. *Seismol. Res. Lett.*, 66:9–19, 1995.