User’s Manual

Full-Wave Ambient Noise Tomography

Yang Shen, Wei Zhang, Zhigang Zhang, …

Graduate School of Oceanography

University of Rhode Island

Version 1.0

Draft, 03.2011

Contents

1. Introduction

Seismic records at pairs of stations are weakly correlated due to ambient seismic waves that propagate between the stations. Empirical Green's Functions (EGFs) can be extracted from ambient noise by cross-correlating records at the stations. Numerous studies have shown that EGFs provide useful signals at a wide frequency range and can be used to image the earth structure (e.g., Shapiro et al., 2005; Nishida et al., 2009). Because EGFs do not depend on earthquake sources and the locations of these "virtual sources" are known precisely, ambient noise tomography has become a powerful tool in seismic imaging, particularly in aseismic regions.

This User's Manual provides a step-by-step explanation of the procedure and scripts used in full-wave ambient noise tomography (Shen and Zhang, 2010). The method is based on the scattering-integral approach (Zhao et al., 2005; Zhang et al., 2007; Chen et al., 2007a,b; Zhang and Shen, 2008). The users are assumed to have basic training in seismology and be familiar with shell programming. Figure 1 shows the flow chart of full-wave ambient noise tomography. Station Strain Green Tensors (SGTs) are constructed from a 3D reference model by finite-difference simulation of the response to orthogonal unit impulsive point forces acting at stations. Travel time, amplitude, or waveform anomalies are measured from observed and synthetic waveforms at stations. The station SGTs are used to calculate finite-frequency sensitivities to perturbations in Vp, Vs (or bulk and shear moduli), density, and attenuation. The measurements and structural sensitivity kernels are used to invert for the earth structure. The tomographic inversion results and additional constraints (e.g., receiver function solutions, well logs) are added to the 3D reference model. This process can be repeated to progressively improve the resolution.

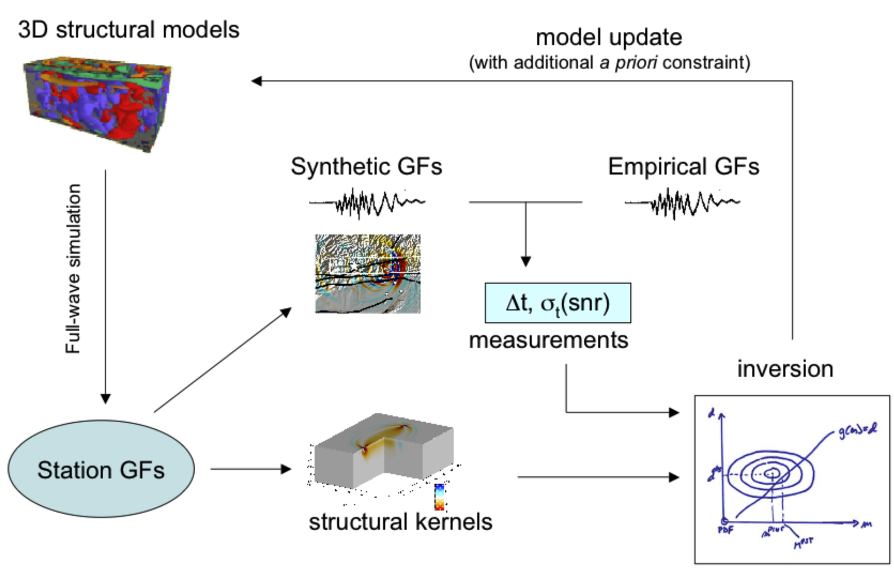

Figure 1. Flow chart of full-wave ambient noise tomography.

Depending on the dimension of the study area, the forward wave propagation simulation may be carried out in a (local) 3D Cartesian coordinate (Zhang et al., 2008; Zhang and Chen, 2011) or a spherical coordinate (Zhang et al., 2011). Examples in this User's Manual are based on wave simulation in the spherical coordinate for a regional-scale study.

We use (c/bash) shell scripts to drive nearly all data processing and computation. Modifications are needed for different users, study areas, data sets, etc. We suggest that users follow a similar directory and file structure as in the examples until they are familiar with the entire procedure.

If you use full-wave ambient noise tomography in your research, please consider citing the following reference (to be updated):

Shen, Y. and W. Zhang (2010), Full-wave ambient noise tomography of the northern Cascadia, SSA meeting (abstract), Seismological Research Letters, 81, 300.

2. Directory and file structure

We process continuous seismic data to extract EGFs on a workstation and run wave propagation simulation, kernel calculation and inversion on a Linux cluster.

On the workstation, the directory structure for data processing looks like following:

$proj_hm_dir (directory for the project, usually in user's home directory)

- csh (subdirectory containing shell scripts)
- data_reqs (subdirectory containing data request files)
- matlab (matlab scripts)
- STinfo (subdirectory containing station files)

In addition to $proj_hm_dir, the user needs a large internal or external disk ($disk) to store the data downloaded from the IRIS DMC and intermediate data generated in processing. When EGFs are obtained, they are copied to the "data" directory on the cluster (see below).

3. Data Processing

Data processing consists of following steps. Each involves one or more c-shell scripts listed in Appendix. Starting from (e), the scripts can be daisy-chained as in '**./csh/rerun.csh**', though power outage, matlab server/license problem, and most likely errors in the scripts may cause a breakdown of the chain, resulting in incomplete or erroneous results. We recommend that users test individual scripts with a small data sets (a few stations), before run them in a daisy chain.

A. Prepare a station list for the study area

First, find all available stations in the study area from the IRIS DMC (http://www.iris.washington.edu/SeismiQuery/station.htm). You will need the following station information in the station list: network, station, elevation, start time, end time, site like, latitude, and longitude.

Depending on the wave periods to be used, it may be unnecessary to download broadband records with a high sampling rate (e.g., BHZ at 40 sps). Some stations have multiple data streams (e.g., BHZ and LHZ). It may be sufficient to use LHZ records with a sampling rate of 1 sps. We suggest at least 7 samples per minimum period for adequate signal fidelity.

Use "gmap" at the IRIS DMC to find the stations with the right channel and time, (e.g., http://www.iris.edu/gmap/?chan=BHZ&minlat=45&maxlat=55&minlon=5&maxlon=155&timewindow=1995-2010). Copy and save the station names at the right side of the "gmap" window. Run '**./csh/sort_stn_lst.csh**' to select the wanted ones from the list of available stations.

B. Make data request files

Run '**./csh/mk_yearly_req.csh**' to general BREQ_FAST files.

For LH channels, we ask for one full-year record of a single channel at a single station in each request. For BH channels, we request 6-month or monthly record to limit the size of the resulting SEED files (large files tend to hang up during ftp download).

Note: The horizontals of borehole stations are different (LH1 and LH2).

C. Send data request files to the IRIS DMC

Run '**./send_allreq.csh**'

D. Download SEED files from the IRIS DMC

After the data requests are submitted to the IRIS DMC, you will receive automated emails about the status of your request. You may also check the IRIS web site to see if the data are ready to be downloaded (http://www.iris.washington.edu/data/, view request status and shipment). Follow the instructions in the email from IRIS to download the files, if the files are small and can be downloaded within a few hours. Otherwise, go to your directory at the IRIS ftp site and make a list of the SEED files and save it in the directory $disk/seed_data.

Run '**./csh/fetch_seed.csh**' to fetch the SEED files.

This script compares the SEED files in the local directory against the list and determines which SEED files are yet to be downloaded from IRIS. You may run several ftp streams using the script. Give each ftp stream a different "tag" to uniquely identify the SEED files to be downloaded. Sometimes ftp hangs up and you have to terminate the ftp job. The ftp site has a maximum connection time, so you may run out of time before the files are all downloaded. You can restart '**./fetch_seed.csh**" to download the remaining files. Use '**./csh/find_incomplete_seed.csh**' to remove incomplete SEED files in your local directory, and then do a final sweep with '**./fetch_seed.csh**'.

E. Extract sac files and remove instrument response

Run '**./rdseed2sac_yr.csh**' to extract daily sac files from SEED and remove instrument response. The output is ground displacement. The script checks instrument reversal and gain. For password protected data, use '**./openssl.csh**' first to remove encryption.

F. Find and remove incomplete sac files

Some sac files are incomplete (missing data in the daily files). If the number of incomplete files is relatively small, then removing incomplete sac files simplifies subsequent processing.

Run '**./csh/find_and_rm_incomplete_sac.csh**'

G. Delete stations with insufficient data

Run '**./csh/count_sac_files.csh**'

H. Check station orientation

Some stations, even the GSN stations, are/were not correctly oriented. This is perhaps not an issue if instrument reversal and gain are checked in “rdseed”.

To find out if the orientations are correct, run '**./csh/find_orientation.csh**'

To rotate the horizontals to N and E, run '**./csh/rotate_NE2NE.csh**'

To rotate LH1 and LH2 (borehole instrument) to N and E, run '**./csh/rotate_LH12toLHNE.csh**'

I. Re-sample

As a rule of thumb, you need ~7 points per minimum period to obtain good signal fidelity. For example, if the minimum period to be used is 15 s, it is safe to use a 0.5 sample-per-second sampling rate to reduce the cost of subsequent computation.

Run '**./csh/interp2halfsps.csh**'

J. Frequency Time Normalization (FTN)

Run '**./csh/cal_FTN_sac.csh**'

We use the signal normalization method of Ekstrom et al. (2009), instead of the commonly used one-bit normalization. This is a relatively time-consuming calculation. If necessary, you may run it on multiple nodes/cpu cores. There is no point to set a high frequency limit above what you will be able to use. The highest useable frequency depends on the available computational resources. We set the lowest frequency and the frequency interval to half the lowest frequency to be used (For example, if we want to use up to 100 s period waves, we set the low frequency limit and frequency interval in the FTN calculation to 0.5*0.01 = 0.005 Hz).

K. Sort the FTN files by day

Run '**./csh/sort_FTN_by_day.csh**'

L. Remove days that have less then 2 stations

Run '**./csh/count_sac_by_day.csh**'

M. Synchronize files

SAC files extracted from SEED start at slightly different time within the sampling interval (e.g., less than +/- 1 s for LH channels). They need to be synchronized.

Run '**./csh/sync_sac.csh**'

N. Delete time segments with large (M>5.5) earthquakes

Run '**./csh/del_EQs.csh**'

(http://www.iris.edu/SeismiQuery/sq-events.htm)

O. Calculate cross correlation between records at pairs of stations

Run '**./csh/cal_xcorr.csh**'

P. Find and remove corrupted cross correlations

Matlab crash and perhaps some other problems may cause cross-correlation to be “nan” or “0.000000e+00”. A single “nan” file spoils the stacked result for the station pair. Use '/**find_nan_files.csh**', to find and remove those corrupted files.

Q. Select station pairs

Count the number of cross correlation files in each station pairs and remove those with insufficient (e.g., < 300 days) day files '**./count_xcorr_files.csh**'

R. Stack cross correlations and calculate EGFs

'**./stack_xcorr_bothsides.csh**' or '**./stack_xcorr.csh**' (folded about the origin)

Use ‘./matlab/plt_egfs_by_dist.m’ to plot EGFs as a function of distances between station pairs.

Note: I ignore step P (find_nan_files.csh). After stacking,

(1) find_nan_EGFs.m

(2) find_nan_files.csh

(3) stack_xcorr_bothsides.csh

(4) split_sgfs.csh

(5) cnv_egf_ax.csh

(6) find_gcp.csh

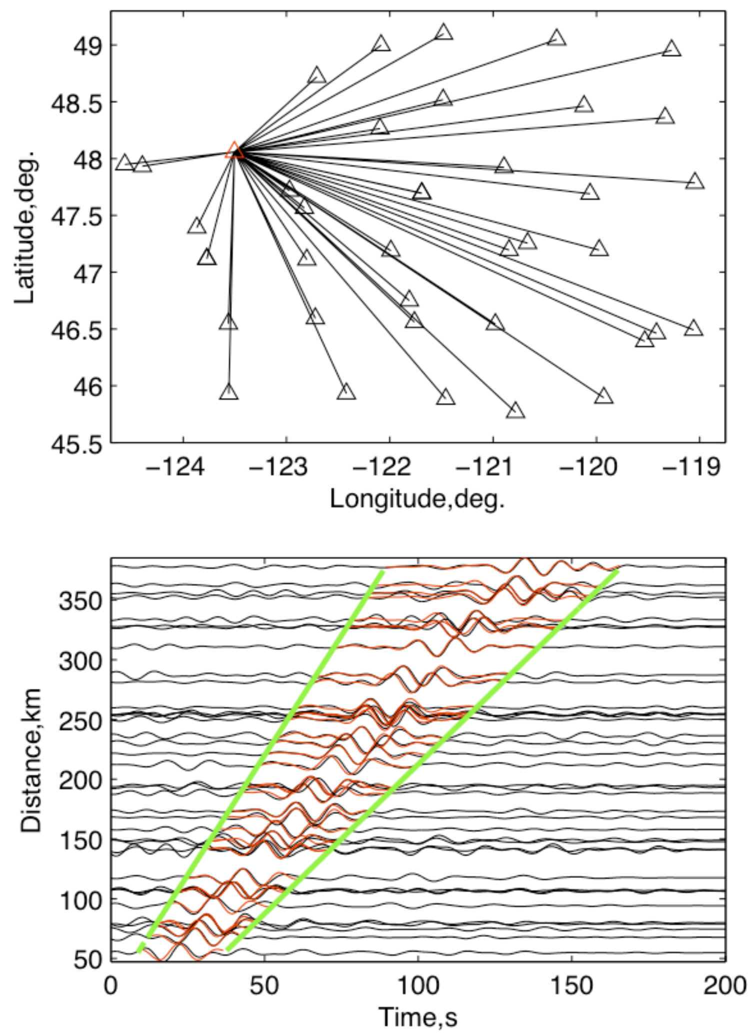

Figure 2. (Top) Lines connect the "virtual source" (red triangle) and receivers. (Bottom) Empirical Green's Functions (black) are plotted as a function of station distances. The red lines are the synthetic Green's functions for the 4th-iteration model of northern Cascadia (Shen and Zhang, 2010). The green lines mark the arrival times for waves at the speeds of 2 and 4 km/s. The waveforms have been filtered between 10-20 s period.

4. Wave Propagation Simulation and Inversion

On the cluster, the directory structure looks like following:

$proj_hm_dir (directory for the project, usually on the I/O node as it takes a large amount of disk space)

codes (subdirectory containing codes for wave simulation, kernel calculation, & inversion)

csh (additional shell scripts to prepare EGFs for measurements)

data (subdirectory containing EGFs and derivatives)

ite_0? (?=1,2,3,.., model iteration)

matlab (matlab scripts)

mfiles (special matlab functions)

misc (miscellaneous files, including topography, coastlines, etc)

model_updates (subdirectory to update the model after each inversion)

STinfo

4.1 Select a model and simulation parameters

It pays to take time to select a model and optimum simulation parameters by running test simulations. See FD3Dspher User's Manual (version 1.0) for instructions on how to set up a model and run simulations. Here we summarize the main steps briefly. Define,

$sim_test = $FD3Dspher/run/$your_study_area

You may use a composite or volumetric model as your initial reference model. To define the study area, edit define_latlon.m in ./$sim_test/config.

To set up a composite model (e.g., AK135+CRUST2.0), run conf_media_composite.m

Output: SeisMedia.composite.$area.crust2.d1800.nc (copy it to $sim_test)

To generate grids in the latitude, longitude, and radial directions, run ./config/grid/creat_grid_xy.m and creat_grid_z.m

Outputs: grid[x,y,z].dat

Edit $sim_test/SeisGrid.conf using the grids in grid?.dat

Edit other configuration files (SeisFD3D.conf, SeisSource.conf, SeisStation.conf), then

run ./bin/seis3d_grid

qsub pbs_media_mpi.sh

Make sure the maximum time step is OK by checking output spher.media.*

Change the grids or the size of the time step in SeisFD3D.conf if necessary & inspect the discretized media file (using ./draw_media_surf_all.m).

run ./bin/seis3d_metric

run ./bin/seis3d_source

run ./bin/seis3d_station

qsub pbs_wave_mpi.sh

Inspect the snapshots and waveform (draw_snap_surf_spher.m and draw_seismo_single.m). If the waveforms are OK, you are ready to go to the next step.

The directory ./ite_0? should have the following subdirectories:

./sim_input (netCDF files for the coordinate and media, copied from $sim_test/input)

./sim.station (station SGTs and velocity on the free surface)

./measure (phase measurements)

./sim.kernel (sensitivity kernels)

./syn.seismograms (synthetic seismograms from each virtual source)

./inv.structure.kerVpVs (inversion)

4.2 Calculate SGTs and Seismograms (directory: sim.station)

In the subdirectory skel/fx under sim.station

Make necessary links

ln -s $sim_test/bin bin

ln -s ../../../sim.input input

Edit the configuration files

mkdir checkpoint

mkdir output

mkdir input.src

In sim.station, run ./skel2station.sh to generate station directories

Run ./subjob.csh and ./subjob_timed.csh

The scripts use a list, work.lst, to keep track which stations have been calculated. The stations that have been calculated have "#!" before their names.

Note: to clean the node memory, /home/$user/bin, edit list_node_ipc to list nodes need to be cleaned and then run clean_ips_priv.sh

4.3 Calculate synthetic seismograms at stations (directory: syn.seismograms)

Run syn.seismograms/snap2sac_spher.m

The script reads outputs from sim.station and writes seismograms at receivers in sac.

4.4 Measure phase delays (directory: measure)

First, prepare EGFs for travel time measurement:

Split EGFs into positive and negative time lags,

../../csh/split_egfs.csh

Convolve EGFs with source time function used in the calculation of SGTs,

../../csh/cnv_egf.csh

run ../../csh/find_gcp.csh to exclude station pairs outside or too close to the model boundaries;

Measure delay by cross-correlating observed and synthetic EGFs,

./measure_phase_delay.csh

Assemble the measurements into a list,

./assemble_measure.csh

Output: measure_result.dat

Plot the ray paths,

./plt_GCPs.csh

Plot delay vesus distance,

./plt_data_dist.csh

Plot EGFs vs. Synthetics,

../../matlab/plt_data_vs_syn_gao.m

4.5 Calculate sensitivity kernels (directory: sim.kernel)

Link the executables,

ln -s ../sim.input input

ln -s $FD3Dspher/code/bin bin

Generate conf files for the kernel calculation,

./kern_measure2kernelconf.sh

Creat directories for the kernel files,

./kern_conf2skel.sh

Generate synthetics in ascii in the receiver directories, .

/kern_synthetic4kernel.sh

Generate matlab fileter file,

./kern_createfilterfile.m

Edit the configuration file, TomoKernel.conf

Submit job, qsub pbs_kernel_mpi.sh

(Monitor the memory use in the nodes. If memory swap occurs, then you need to reduce the block_size to calculate the kernel in blocks).

Plot sensitivity kernels,

draw_kernel_spher.m and draw_kernel_spher_multi.m

4.6 Inversion (directory: inv.structure.kerVpVs)

Generate inversion block dimensions and indexes,

inv_make_block_stride_xygridcent.m

Generate smoothness constraints,

inv_make_smooth.m

1th: first derivative (flatness); 2th: 2nd derivative (see Menke, p53)

After we have the inversion block dimension, we can assemble the kernels

./sim.kernel/run.kernel.assm.sh

(assembled results are saved in ./inv.structure.kerVpVs/G_spool)

Note: pay attention to the threshold kernel value, below which the sensitivity is ignored

Creat a model directory (e.g., Freq3456.Z.model)

Generate a list of the assembled kernels,

./list_G_raw.sh

Generate station and event ("virtual source") lists,

./mk_st_list.csh

Edit the following lists: inv_black_list, inv_st_list, inv_ev_list, inv_cmp_list, inv_freq_list, inv_twin_list, inv_phase_list (frequencies have to be in the same format as in ./measure/measure_result.dat)

Generate the final list used in inversion,

./inv_Gd_list.sh

Run ./Freq3456.Z.model/run.solver.1th.sh, with various damping and smoothness constraints

Use plt_slice.m too view the resulting model

Find the tradeoff between the solutions and dampling/smooothness paramfeters,
plt_tradeoff.csh

Select a model with the optimum variance reduction and model norm.

Variance reduction: 1-∑(di-di_pre)^2/∑di^2
Chi square: (∑(di/σi)^2)/N
Note: di is the observed ith misfit between observation and synthetics, dt
di_pre is the predicted ith misfit dt between observation and synthetics, K(dc/c)dv
LSQ: K(dc/c)dv=dt where dc/c is the velocity perturbation and dv volume.

4.7 Update model (directory: $projhmdir/model_updates)

Copy model input files,../ite_01/sim.input, to input_ite_01 and updated_input_ite_01 (media files in updated_input_ite_01 will be rewritten. if the study box is changed in the next iteration, copy input and correspoinding files from the $sim_test directory).

Copy configureation file,
cp ../ite_01/sim.station/skel/fx/SeisFD3D.conf SeisFD3D.conf_ite_01

Update the model by "TriScatteredInterp" the inversion result and adding it to the previous model, update_model_smth.m

Compare current and updated model, './plt_models.m'

Set up the directory and files for the next iteration,
'setup4next_ite.csh

Return to 4.2 and iterate until a satisfactory model is obtained.

Compare results of each iteration with matlab/hist_xcorreff.m.

References

Chen, P., T.H. Jordan, and L. Zhao, (2007a), Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods, Geophys. J. Int., 170, 175-181.

Chen, P., L. Zhao, and T.H. Jordan (2007b), Full 3D tomography for crustal structure of the Los Angeles Region, Bull. Seismol. Soc. Am., 97 (4), 1094-1120, doi: 10.1785/0120060222.

Ekström, G., G.A. Abers, S.C. Webb (2009), Determination of surface-wave phase velocities across USArray from noise and Aki's spectral formulation, Geophys. Res. Lett., 36, L18301, doi:10.1029/2009GL039131.

Nishida, K., J.-P. Montagner, and H. Kawakatsu (2009), Global surface wave tomography using seismic hum, *Science*, **326**, 112.

Shapiro, N.M., M. Campillo, L. Stehly, M. Ritzwoller (2005), High-resolution surface-wave tomography from ambient seismic noise, Science, 307, 1615-1618.

Shen, Y., W. Zhang (2010). Full-wave ambient noise tomography of the northern Cascadia, SSA meeting (abstract), Seismological Research Letters, 81, 300.

Shen, Y., Z. Zhang, and L. Zhao (2008). Component-dependent Frechet sensitivity kernels and utility of three-component seismic records, Bull. Seism. Soc. Am., 98 (5), 2517-2525, doi: 10.1785/0120070283.

Zhang, W. and X.F. Chen (2011), 3D Elastic Wave Numerical Modeling In the Presence of Surface Topography by a Nonstaggered-Grid Finite Difference Method on Curvilinear Grid, submitted to GJI.

Zhang, W., Y. Shen, and X.F. Chen (2008), Numerical simulation of strong ground motion for the Ms 8.0 Wenchuan earthquake of 12 May 2008, Science in China Series D-Earth Sciences, **51(12)**, 1673-1682.

Zhang, W., Y. Shen, and L. Zhao (2011), 3D Seismic Wave Modeling in Spherical Coordinate by a Nonstaggered-grid Finite Difference Method on Nonuniform Grids, manuscript in preparation.

Zhang, Z. and Y. Shen (2008). Cross-dependence of finite-frequency compressional waveforms to shear seismic wave-speeds, Geophys. J. Int., 174, 941-948, doi:10.1111/j.1365-246X.2008.03840.x.

Zhang, Z., Y. Shen, and L. Zhao (2007). Finite-frequency sensitivity kernels for head waves, Geophys. J. Int. 171, 847-856.