Users' manual for DSurfTomo

Hongjian Fang (fanghj@mail.ustc.edu.cn) and Huajian Yao (hjyao@ustc.edu.cn)

School of Earth and Space Sciences, University of Science and Technology of China
Last modified: Sep 7, 2015

1, Description

DSurfTomo is the surface wave inversion program that can directly invert surface wave dispersion data to 3D shear wave speed without the intermediate step of constructing the phase or group velocity maps. A fast march method (FMM) (Rawlinson et al., 2004) is used to compute, at each period, surface wave travel times and ray paths between sources and receivers. This avoids the assumption of great-circle propagation that is used in most surface wave tomographic studies, but which is not appropriate in complex media. Please refer to Fang et al.(2015 , GJI) for the detailed description of the method.

2, Installation

This program has been tested successfully on Linux(CentOS, Debian, Ubuntu etc.)/Unix platform with gfortran compiler. You may need to change the Makefile a little bit if you are using other compiler.
Firstly, make a directory where you want to put the DSurfTomo code and example in
mkdir DSurfTomo
Then, cope the source code to DSurfTomo
cd DSurfTomo
cp ~/pathwhereyouputsrc/DSurfTomo.tgz .
At last, install
tar –xzvf DSurfTomo.tgz
./configure
Then you can find the executable file in the 'bin' directory.
You may need to add this path into PATH variable
echo 'export PATH = $PATH:/yourpath/SurfTomo/bin' >> ~/.bashrc

3, Data preparation

You need to prepare all the dispersion measurements as followed:

```
 #    25.148500 121.511100 1 2 0
   25.158529 121.476890 0.7990
   25.133539 121.499190 1.0420
 #    25.158529 121.476890 1 2 0
   25.119850 121.473190 0.6460
 #    25.128920 121.417420 1 2 0
   25.119850 121.473190 0.9430
 #    25.119850 121.473190 1 2 0
   25.090361 121.462250 0.8280
   25.083694 121.435220 1.0870
   25.133539 121.499190 1.3910
   25.102119 121.515930 1.2950
 #    25.090361 121.462250 1 2 0
```

25.083694 121.435220 1.3240
25.061472 121.431140 0.7490
25.083694 121.435220 1 2 0
25.065750 121.477750 0.6220
25.061472 121.431140 0.8190
……

Lines begin with '#' represent the sources, followed by source latitude, source longitude, period index (integer), wave type and velocity type (explain latter). Then followed by the receiver data: the first two columns are the latitude and longitude of the receivers, the third column is phase or group velocity (surface wave dispersion measurements).

Period Index (integer): index of the period vector that is listed in the parameter file DSurfTomo.in

Wave type: **2 for Rayleigh wave and 1 for Love wave**

Velocity type: **0 for phase velocity and 1 for group velocity**.

Note that you can put four kinds of data and **jointly invert** them to get the 3D shear velocity model. You can give the file a name, e.g. surfdata.dat

(You can use "ExtractData.py" in the script directory to extract the data from the file after you get the phase of group velocity using Huajian Yao's program. After changing some parameters (stations' name etc) in the script file, run 'python ExtractData.py > surfdataTB.dat')

3, the initial model

The file name of the initial model must be 'MOD', the content looks like:

```
  0.0  0.2  0.4  0.6  0.8  1.1  1.4  1.8  2.5
  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900
0.900  0.900  0.900  0.900
  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900
0.900  0.900  0.900  0.900
  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900
0.900  0.900  0.900  0.900
  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900  0.900
0.900  0.900  0.900  0.900
```

……

The first line is depth grid point.

Then followed by shear velocity value, one line represents shear velocity value at different latitudes (ascending order or descending order?) at a single longitude and a certain depth, then followed by next longitude, then different depth. A simple python script (GenerateIniMOD.py) in 'scripts' directory can be used to output an initial model. After changing some parameters according to your case, run 'python GenerateIniMOD.py').

4, the input file for the inversion

the file name must be 'DSurfTomo.in' and it contains:

```
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c INPUT PARAMETERS
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
surfdataTB.dat                  c: data file
18 18 9                         c: nx ny nz (grid number in lat lon and depth direction)
25.2  121.35                    c: goxd gozd (upper left point,[lat,lon])
0.015 0.017                     c: dvxd dvzd (grid interval in lat and lon direction)
20                              c: less than nrc*numf
4.0  0.0                        c: weight damp
0.1                             c: minthk (about 1/3 of grid interval in vertical direction)
0.5 2.8                         c: minimum velocity, maximum velocity (priori information)
10                              c: maxiter (iteration number)
0.2                             c: sparsity fraction
26                              c: kmaxRc (followed by periods)
0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9 3.0
0                               c: kmaxRg
```

```
0                              c: kmaxLc
0                              c: kmaxLg
0                              c: synthetic flag(0:real data,1:synthetic)
0.02                           c: noiselevel
2.5                            c: threshold(second)
```

The file is kinda self-explanatory.
One thing to be noted is nx and ny had better to be 2^n+2, nz had better to be 2^n+1 to avoid the boundary problem in wavelet domain, where n is an integer, for example n can be 3,4,5 or even 6 if the data constraint is good enough. Remember there's no such requirement in spatial domain.
The sixth line is the start point in latitude and longitude, note the direction of Latitude is from North to South, and from West to East for Longitude.

The seventh line is the grid interval in Latitude and Longitude.
Need to ensure that all the sources and receivers are in the region of goxd-(nx-3)*dvxd~goxd and gozd~gozd+(ny-3)*dvzd, otherwise the program will complain and stop.

The eighth line is the maximum number of data for a certain receiver. We output this number when we reformat the data file. You can also set it to nrc*numf, which is the number of stations times the total number of periods for dispersion measurements.

"weight" is the balancing parameter between data fitting term and smoothing (or sparsity in wavelet domain) regularization term, sometimes it is tricky to choose an appropriate one since L-curve does not always work. From my perspective, an appropriate weight will lead to reasonable velocity change at each iteration. At the first few iterations, velocity change can reach about 0.4. If it is too large, it means the weight you choose is too small and vice versa. Generally, a few trials will be more than enough to choose a good weight. The velocity change will decrease in the last a few iterations as the inversion converges, or the program will complain by outputting a lot of garbage information.

"damp" is the input parameter that lsqr needs, it controls the amplitude of the inverted parameter. We typically set it to zero.

"minthk" is the parameter we use to transform from the grid point model to the layered model when computing sensitivity kernel. It should be about 1/3 of the minimum grid interval in the depth direction. minimum and maximum velocity is some a priori information you have about the study area, which can be roughly estimated from dispersion data if you are familiar with surface wave theory. You can set a large interval if you are not sure, and it doesn't affect the final result too much.

"maxiter" is the iteration number of the inversion. It usually converges after a few iterations.

Sparsity fraction parameter means how sparsity the sensitivity matrix is, 5 percent will be enough in the spatial domain, it will be more (e.g., 20 percent) in the wavelet domain).

KmaxRg, kmaxRc, kmaxLc, kmaxLg means the number of periods for Rayleigh wave phase velocity, Rayleigh wave group velocity, Love wave phase velocity and Love wave group velocity, respectively. Note if the number of periods is zero, you need not write the periods following the number, e.g. the kmaxRg.

"synthetic flag" set to 0 means inversion using real data, 1 for synthetic data; it needs a file named "MOD.true" when the flag is set to 1, otherwise the program will stop. You can also generate

‘MOD.true’ using ‘GenerateTrueMOD.py’ in the ‘scripts’ directory after changing some parameters. The command is ‘python GenerateTrueMOD.py’

“noise level” means how much Gaussian noise do you want to add to your synthetic data. For instance, 0.02 means the random Gaussian noise added has zero mean and a standard deviation of 2%*(observed travel time).

The last line represents the threshold, misfit greater than the threshold will be reweighted with a small weight in order to prevent outliers affect the final result too much.

There are many scripts to generate checkerboard and initial model, and make quick plots of the results. We put them in the 'scripts' directory, you can find and change them according to your own case.
Syntax to run the program:
DSurfTomo DSurfTomo.in
or just
DSurfTomo
then type the input file name from the keyboard.

Output files
The final velocity model (SurfTomo.inMeasure.dat):
First column : longitude
Second column : latitude
Third column : depth
Fourth column : shear velocity
You can quickly plot some slices using “plotslice.gmt” in the “scripts” directory by the following command:
csh plotslice.gmt SurfTomo.inMeasure.dat depth1 depth2 depth3 depth4.

The raypath distribution of the final model (raypath.out)
number of ray path segments
latitude longitude
…
Velocity model at each iteration (IterVel.out)
Similar to initial model

Residual of first and last iteration (residual*.dat)
Distance ForwardT ObserveT weightedForwardT weightedObserveT weight
…

Feel free to contact us if you have any questions.

References
Fang, H., Yao, H., Zhang, H., Huang, Y. C., & van der Hilst, R. D. (2015). Direct inversion of surface wave dispersion for three-dimensional shallow crustal structure based on ray tracing: methodology and application. Geophysical Journal International, 201(3), 1251-1263.
Rawlinson, N. & Sambridge, M., 2004. Wave front evolution in strongly heterogeneous layered media using the fast marching method, *Geophys. J. Int.,* **156**(3), 631–647