Generating velocity solutions with `globk`

T. A. Herring   M. A. Floyd   R. W. King

Massachusetts Institute of Technology, Cambridge, MA, USA

UNAVCO Headquarters, Boulder, Colorado, USA
19–23 June 2017

http://web.mit.edu/mfloyd/www/courses/gg/201706_UNAVCO/

Overview

- Basics of "velocity" solutions
  - Invoked with "apr_neu all xx xx xx <NEU velocity sigmas>"
- Strategies for setting up solutions (they can take a long time to run)
- Strategies for speeding up solutions.
- Methods for "cleaning up" potential problems
- Different reference frame realizations
- Some examples.
- *These solutions involve making decisions about how to treat data and the type of solution to be created – lots of decisions*

GLOBK velocity solutions

- The aim of these solutions is to combine many years of data to generate position, velocity, offset, and postseismic parameter estimates. Not uncommon to have 10000 parameters in these solutions.

- Input requirements for these solutions:
  - a priori coordinate and velocity file. Used as a check on positions in daily solutions (for editing of bad solutions) and adjustments are a priori values (a priori sigmas are for these values)
  - Earthquake file which specifies when earthquakes, discontinuities, and mis-named stations affect solution. Critical that this file correctly describe data.
  - Process noise parameters for each station. Critical for generating realistic standard deviations for the velocity estimates (`sh_gen_stats`).

Velocity solution strategies

- In general careful setup (i.e., correct apriori coordinate, earthquake file and process noise files) is needed since each run that corrects a problem can take several days. Incorrect solutions may not complete correctly and results may be subtly wrong.

- General strategy for iteratively generating velocity solution:
  - Define a core-set of sites (usually 20-200 sites) where the solution runs quickly. Test files on this solutions and use the coordinate/velocity estimates to form the reference frame for time series generation.
  - Time series using these reference frame sites and then test (RMS scatter, discontinuity tests) to form a more complete earthquake and apriori coordinate/velocity files.
  - Steps above are repeated, usually increasing number of stations until solution is complete. As new stations are added missed discontinuities and bad process noise models can cause problems.

- Aim here is make sure that when a large solution is run (maybe several days of CPU time) that the run completes successfully.

# General methods for increasing speed and to allow for parallel runs

- Approaches to increase speed:
  - Pre-combine daily solutions into weekly to monthly solutions and use these combined solutions in the velocity solutions. There are many advantages to this approach:
    - Runs are much faster. Each processing step takes about the same time with the monthly as a daily file but there are 30 fewer files so 30 times faster.
    - Numerical rounding errors are much better when monthlies are used
    - "MIDP" output option refers the solutions to the middle of the month. (Earlier versions used last day of month as reference time, natural time for a sequential Kalman filter.
    - Random walk process noise models correct when velocity NOT estimated in combinations.
    - Care needed here when "eq_log" is used for solutions far away in time from the earthquake.
  - Run decimated solutions (e.g., one day per week). Works fine and changing start day does not have large effect due to correlated noise models. Care needed when different start day results are combined to avoid white noise sigma reduction.
  - Sub-netting in GLOBK to generate each solution with smaller number of stations. Sub-net velocity solutions are combined with GLOBK. Use `netsel` with `—rw` option to make GLOBK "use_site" list (Current PBO approach)

Before velocity runs

- Surveys may be combined into one solution per survey

- No need to re-run `glred` again to see long-term time series

- Multiple ".org"-files may be read by `tssum` or `sh_plot_pos`
  - `tssum ts_pos mit.final_igb08 -R survey1_comb.org survey2_comb.org ...`
    - `ts_pos is the name of a directory for the .pos files. (. can be used)`
  - `sh_plot_pos -f survey1_comb.org survey2_comb.org -k ...`

Example: Long-term time series for survey sites

Reasonable repeatability
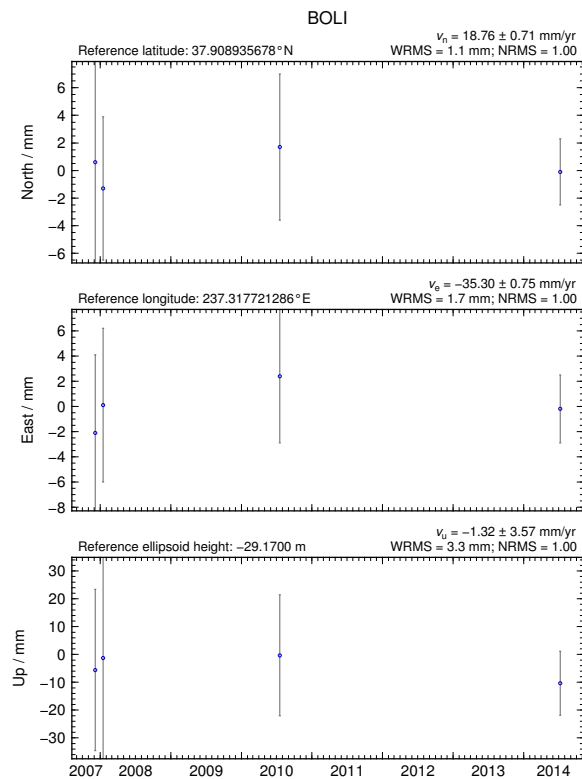
Outlier in vertical
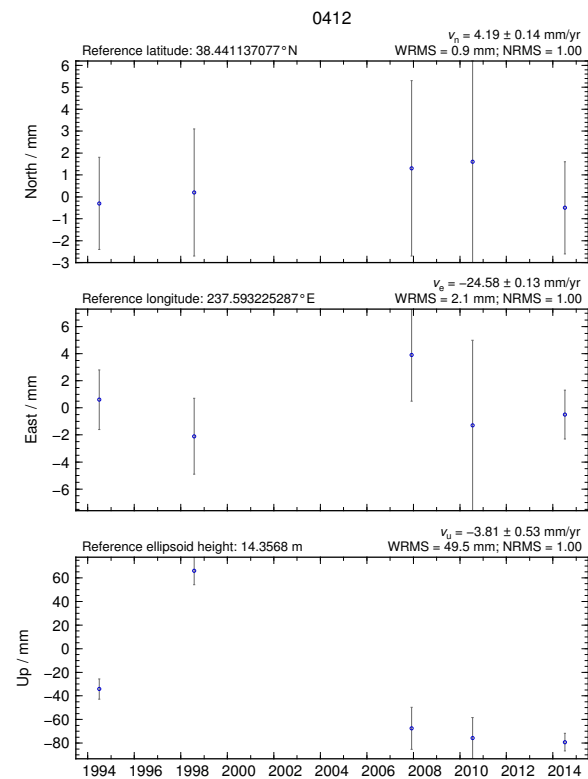
Excluding outliers or segments of data

- Create "rename" file records and add to `globk` command file's "eq_file" option, e.g.

  ```
  rename PTRB      PTRB_XPS h1407080610_nb4a
  rename PTRB      PTRB_XPS 2014 07 07 18 00 2014 07 08 18 30
  rename ABCD      ABCD_XCL 2013 07 08 00 00
  ```

- "XPS" will not exclude data from `glred` (so still visible in time series) but will exclude data from `globk` (combination or velocity solution)

- "XCL" will exclude data from all `glred` or `globk` runs

Run `globk`

- Create new ".gdl"-file with *combined* binary h-files, e.g. from vsoln/, assuming standard directory hierarchy
  - `ls ../*/gsoln/*.GLX > vsoln.glx.gdl`
- Optionally run `glist` to see size of solution
  - Recommended to prevent problems during long `globk` run
  - `glist` can read earthquake file and `globk` use site type commands. (Useful if a `globk` solution seems to be missing or has extra sites.)
- Run `globk`
  - This may take many hours for very large/long velocity solutions
  - Use `tsfit` with earthquake file to generate a priori site coordinates. Be careful if ~/gg/tables/itrf08_xxx.apr files also used because some site names permutations may have inconsistent coordinates (use `unify_apr` to be safe)

`glorg` for different reference frames

- No need to re-run `globk` every time you want
- `glorg` is usually called from `globk` command file ("org_cmd" option) but `glorg` may be run separately
- globk 6 globk_vel.prt globk_vel.log globk_vel.gdl globk_vel.cmd
- glorg globk_vel_noam.org ERAS:… glorg_vel.cmd vel.com
- Must have saved the ".com"-file!
  - e.g. "com_file @.com"
  - Do not use "del_scra yes" in `globk` command file
  - "apr_neu" must be loosely constrained ("apr_rot" and "apr_tran" will also need to be used for sestbl. "BASELINE" experiment solutions.

Use of equates

- With earthquakes and discontinuities, there can be many site names for the same physically location:
  - Equate commands in `glorg` allow the velocity adjustments at these sites to be made the same (or constrained to be the same within a specified sigma)
  - "eq_dist" allows site separate by distance to equated (and constrained in latest `glorg`).
  - "eq_4char" equates sites with same 4-character name (useful to stop equates at sites that share antennas).
  - Chi-squared increments of equates allows assessment of equates (use "un_equate" for large chi-squared values)
  - Use "FIXA" option to make a priori the same for equated sites (better to use consistent a priori file).

Uses of `sh_gen_stats`

- Velocity solutions are often iterative:
  - Generate time series using some reference frame sites (IGb08 sites initially for example).
  - Fit to the time series (`tsfit`) to:
    - Find outliers, nature of earthquakes (log needed?), discontinuities
    - Self consistent a priori file.
    - Used FOGMEx model (realistic sigma) to get process noise model and list of low-correlated noise reference frame sites). Use "stabrad" option for dense networks
  - Run `globk` velocity solution to refine reference frame site coordinates and velocities
  - Re-generate time series and repeat.

Some comparisons: Approach

- Use `sh_exglk -f <soln.org> -vel <soln.vel> -rmdup`
to extract velocity estimates (`rmdup` removes equated sites with the same estimates)

- Program `velrot` allows fields to be compared (change frames and merge fields as well). For example:

`velrot solna.vel nam08 solnb.vel IGS08 '' '' '' '' N` compares to solutions directly (use "RT" instead of "N" to allow rotation and translation rates). Use `grep '^S '` to get statistics.

Comparisons: Decimation

Decimation: Different days of week (1996-2015 solution, small subset of sites):

```
Un-aligned fields
compare 1 NMT_vel_150418_day1.vel NMT_vel_150418_day3.vel

S Component North    #    75 WMean  -0.00 WRMS    0.04 mm/yr, NRMS    0.198

S Component East     #    75 WMean  -0.02 WRMS    0.04 mm/yr, NRMS    0.203

S Component Up       #    75 WMean   0.03 WRMS    0.16 mm/yr, NRMS    0.180

S Component Horz     #    75 WMean  -0.01 WRMS    0.04 mm/yr, NRMS    0.200

compare 2 NMT_vel_150418_day1.vel NMT_vel_150418_day5.vel

S Component North    #    74 WMean  -0.01 WRMS    0.04 mm/yr, NRMS    0.207

S Component East     #    74 WMean  -0.02 WRMS    0.05 mm/yr, NRMS    0.225

S Component Up       #    74 WMean   0.04 WRMS    0.19 mm/yr, NRMS    0.212

S Component Horz     #    74 WMean  -0.01 WRMS    0.04 mm/yr, NRMS    0.217

compare 3 NMT_vel_150418_day3.vel NMT_vel_150418_day5.vel

S Component North    #    76 WMean  -0.01 WRMS    0.03 mm/yr, NRMS    0.177

S Component East     #    76 WMean  -0.01 WRMS    0.03 mm/yr, NRMS    0.161

S Component Up       #    76 WMean   0.01 WRMS    0.13 mm/yr, NRMS    0.142

S Component Horz     #    76 WMean  -0.01 WRMS    0.03 mm/yr, NRMS    0.169
```

Comparison: Time series vs GLOBK

- ## PBO Combined analyses:

```
Un-aligned fields (no rotation and translation).
compare 1 PBO_vel_150425.vel PBO_vel_150425KF.vel

S Component North    #  2105 WMean  -0.01 WRMS    0.12 mm/yr, NRMS    0.925
S Component East     #  2105 WMean  -0.00 WRMS    0.13 mm/yr, NRMS    0.934
S Component Up       #  2105 WMean   0.02 WRMS    0.31 mm/yr, NRMS    0.871
S Component Horz     #  2105 WMean  -0.01 WRMS    0.12 mm/yr, NRMS    0.929
compare 4 PBO_vel_150425.vel PBO_vel_150425_NAM08.vel

S Component North    #  1972 WMean   0.03 WRMS    0.13 mm/yr, NRMS    0.965
S Component East     #  1972 WMean   0.02 WRMS    0.15 mm/yr, NRMS    1.049
S Component Up       #  1972 WMean  -0.07 WRMS    0.41 mm/yr, NRMS    0.943
S Component Horz     #  1972 WMean   0.02 WRMS    0.14 mm/yr, NRMS    1.008
compare 7 PBO_vel_150425KF.vel PBO_vel_150425_NAM08.vel

S Component North    #  1969 WMean   0.04 WRMS    0.16 mm/yr, NRMS    0.952
S Component East     #  1969 WMean   0.02 WRMS    0.17 mm/yr, NRMS    0.967
S Component Up       #  1969 WMean  -0.08 WRMS    0.44 mm/yr, NRMS    0.935
S Component Horz     #  1969 WMean   0.03 WRMS    0.16 mm/yr, NRMS    0.959
```

PBO_vel_150425.vel: tsfit solution to time series

PBO_vel_150425KF.vel: tsfit Kalman filter solution to timeseries

PBO_vel_150425_NAM08.vel: GLOBK combined velocity solution (NMT+CWU), decimated 7 days, 28-subnet combination. Reference frame realization to NAM08 frame sites (~600)

See Herring et al., Reviews of Geophysics, 2016 for more detailed comparisons.

Final comments

- Practice large solutions with decimated data sets and small networks (run time increased cubically with number of stations)

- Make sure your a priori coordinates files are consistent (especially with equates)
  - Use the out_aprf command in tsfit to generate an apriori which is consistent with your timeseries estimates.