** Example of processing a complex set of multiyear observations **

Last edited by rwk 150808

Data set: Three years for surveys in multiple parts of
            the eastern-Mediterranean/Caucasus

Strategy:
  - combine daily sGPS GAMIT processing with cGPS h-files
  - aggregate the processing by survey length up to about 30 days
  - use survey-combined H-files to get long-term repeatabilities
    and velocities

Assumed directory structure:
 /EMed
    /2004 /2006 /2008 /tables /vsoln /emed_rinex

Under each year (experiment) directory is the usual structure for sh_gamit and sh_glred (e.g. /tables, /[day-number], /rinex, /gsoln, etc.. However, we have stored all of the RINEX files, in a separately directory elsewhere on the system. We'll call this directory /emed_rinex, and beneath this level will be sub-directories organized by country and year, e.g., /azerbaijan/2004. (This scheme is optional.)

Step 1: Determine the temporal and spatial distribution of the data

For example, if the data are stored in directories by country and year, within each RINEX directory you can use sh_get_times to obtain a list, e.g. in emed_rinex/azerbaijan/2004, type 'sh_get_times -sort -f *.04o > times.azer04.sorted', where the file times.azer04.sorted will have

```
shik2672.04o   2004   267   13:31   269   13:42  48:11  5784 3
khid2672.04o   2004   267   13:39   269   13:41  48:2   5766 3
kurd2731.04o   2004   273   05:06   275   05:13  48:7  11553 3
yevl2735.04o   2004   273   07:58   275   08:01  48:3  11536 3
ayaq2730.04o   2004   273   09:44   275   08:38  46:54  5630 3
kebe2821.04o   2004   282   09:36   284   07:59  46:23 11137 3
qurd2821.04o   2004   282   11:32   284   08:02  44:30 10681 3
shek2841.04o   2004   284   10:13   286   07:05  44:52 10771 3
kate2840.04o   2004   284   13:23   286   11:32  46:9  11046 3
saly2960.04o   2004   296   07:51   298   07:59  48:8  11554 3
bile2960.04o   2004   296   09:13   298   09:16  48:3  11535 3
lenk2960.04o   2004   296   11:57   298   06:02  42:5   5051 3
```

and in emedrinex/georgia/2004/rinex,
type 'sh_get_times -sort -f *.04o > times.georgia.sorted'

```
zinv2730.04o   2004   273   13:15   274   13:14  23:59  2878 2
zinv2740.04o   2004   274   14:23   275   14:19  23:56  2873 2
pasn2750.04o   2004   275   18:01   276   14:24  20:23  2447 2
pasn2760.04o   2004   276   14:25   277   14:19  23:54  2869 2
```

```
pasn2770.04o   2004   277   14:51   277   19:01  4:10    502 1
kazb2780.04o   2004   278   06:31   279   06:29  23:58  2876 2
kazb2790.04o   2004   279   06:30   279   18:04  11:34  1389 1
kazb2791.04o   2004   279   18:05   280   06:50  12:45  1531 2
kudi2810.04o   2004   281   10:11   282   06:29  20:18  2436 2
kudi2820.04o   2004   282   06:30   282   18:04  11:34  1389 1
kudi2830.04o   2004   282   18:05   283   10:22  16:17  1955 2
ingu2840.04o   2004   284   14:46   285   14:24  23:38  2836 2
ingu2850.04o   2004   285   14:25   286   14:19  23:54  2869 2
shua2870.04o   2004   287   10:31   288   10:14  23:43  5692 2
shua2880.04o   2004   288   10:15   289   06:29  20:14  2429 2
shua2881.04o   2004   289   06:30   289   10:21   3:51   464  1
```

So we have survey-mode data for days 267-269, 273-289. Note that all of the Azeri data span parts of three days, so the default setting 'rx_doy_minus 1' in process.defaults will not be sufficient to pick up the data two days earlier. We can change this in process.defaults or make it explicit in the sh_gamit command line.

Step 2: Set up the control files for the first year

Run sh_setup to put templates or actual files you'll need into EMed/2004/tables.

From the /2004 level type 'sh_setup -yr 2004'

Edit process.defaults and sites.defaults for this year.

Make appropriate changes to the template process.defaults:

```
# Directory path for RINEX archives (search all levels)
 set rnxfnd = "/emed_rinex"
# H-files archive
 set hfnd = /net/everest/raid6/ftp/pub/MIT_GLL/H04
# Set compress (copts), delete (dopts) and archive (aopts) options.
 set copts = ( o q mkx )
 set dopts = ( D ao b c g i j k m p ps t x y )
```

The rnxfnd option will cause sh_gamit to make links for all RINEX files within the requested span from from the 2004/rinex directory to the /emed_rinex/azerbaijan/2004/rinex and /emed_rinex/georgia/2004/rinex directories. The copt/dopt settings shown here represent a space-saving approach, retaining only the files you are likely to view unless something goes wrong. We rarely use the option to put some files in an archive directory, preferring to save only the minimal files in the day directories themselves.

Make appropriate changes to the template sites.defaults to automatically download and include in the processing (by being present in 2004/rinex) the ITRF08 sites you wish to use to tie the local h-files to the MIT global files (at present the site name entries must be lowercase):

```
# Tie stations for MIT global h-files, used also for GAMIT
constraints
#  North
 bucu_gps emed ftprnx
#  West
 ankr_gps emed ftprnx
 # Central
 nssp_gps emed ftprnx
#  South
 tela_gps emed ftprnx
#  East
 kit3_gps emed ftprnx
```

Make the appropriate changes to the template sittbl. to provide moderate ITRF08 constraints in the GAMIT processing to help in resolving ambiguities:

```
SITE              FIX    --COORD.CONSTR.--
      << default for regional stations >>
ALL               NNN    100.  100.   100.
      << IGS core stations >>
BUCU BUCU_GPS     NNN    0.050 0.050  0.05
TELA TELA GPS     NNN    0.050 0.050  0.05
KIT3 KIT3_GPS     NNN    0.050 0.050  0.05
```

Step 3: Create station.info

In 2004/tables, which will contain the MIT version of station.info, type

```
 sh_upd_stnfo -ref station.info -w station.info.mit -l sd
```

which will result in the output file having only the entries for sites listed with ftprnx in sites.defaults.

Using the log sheets as a guide, in emed_rinex/azerbaijan/2004/rinex and emedrinex/georgia/2004/rinex, create survey-specific station.info entries manually (editor or using 'make_stnfo') or using 'sh_upd_stnfo -f *.04o' to create entries from the RINEX headers and then editing any entries that need to be changed. The starting file must be named station.info since it is continually concatenated with the information from each new RINEX file;
and the starting file must have a legitimate set of headers though these do not need to match those in files you wish to merge such as the MIT station.info file. Rename these survey-specific files to , e.g. station.info.azer04, and copy them to 2004/tables. Then merge the files:

```
sh_upd_stnfo -ref station.info.mit -merge station.info.azer04 station.info.georgia04
```

which will create a merged file, station.info.new, which you should rename to station.info for processing.

Step 4: Do the GAMIT processing

```
sh_gamit -expt emed -c -s 2004 267 269 -pres ELEV -rx_doy_minus 2 >&! sh_gamit.log
sh_gamit -expt emed -c -s 2004 273 289 -pres ELEV -rx_doy_minus 2 >&! sh_gamit.log
```

The -c option prevents the compression of the RINEX files at the end of processing each day. Since the "files" in 2004/rinex are actually links to/emed_rinex directories, if the compression takes place, you'll get a (compressed) copy of the RINEX files in EMed/2004/rinex, unnecessarily duplicating the files in /emed_rinex.

After each day is run, you'll get an email copy of the sh_gamit_[ddd].summary file for that day. Check it for sites included, norms, ambiguities resolved, and failed editing (autcln rms 0.). Fix any problems and rerun those days.

Step 5: Generate the daily time series for editing

If not there already, edit /2004/gsoln/globk_comb.cmd to add a "contingent/option" line for saving an output h-file:

```
COMB out_glb H------_EMED.GLX
```

This command will be executed only if you include '-globk_cmd_prefit COMB' in the sh_glred command line.

Edit glorg_comb.cmd to specify the stabilization you want for the time series (apr_file, stab_site list, parameters to be estimated). Then type

```
sh_glred -s 2004 267 2004 289 -expt emed -local -opt H G T  >&! sh_glred.log
```

Setting '-local' instructs sh_glred to use MIT global h-files only for days on which a day directory exists (i.e. days for which you have survey-mode data). The output of this command will be a gdl file, log file, and org file for each day and the usual output of sh_plot_pos: a tsfit sum file, postscript plots of time series for each station and histograms of nrms and wrms. (Using E instead of T for –opt will invoke ensum and sh_plotcrd rather than tsfit and sh_plot_pos; we plan to phase out this plotting scheme in 2016.)

Check the stabilization by typing in /gsoln

```
  grep 'POS STAT' *.org
```

The number of sites in the stabilization should be what you expected (4 or more for translation-only, 8 or more well-distributed sites for translation +rotation), the wrms values should be 1-3 mm, and the nrms values less than ~2.

View all the plots ('gs *.ps') and note any outliers or noisy segments. (With a data set that has multiple surveys within a year, you might create sub-directories below /gsoln, e.g. /gsoln/plots_267-289, in which to store the plots.)

Notes:

```
ardt 267-269 downweight (DW) by adding 6 mm; check long-term plots
       later
arga 282 U-30, DW 100
atht 269 N-5 E-10  308 E+5
boyl 270-272 DW 4
bucu 286 N+3 287 U-10
cagl 265-268 DW 5 291 E+7 293 U+15
madr 282-283 E+5 290 E+5
ozdm 267-269 N/E slope DW 15 or DW 267 15  LT E favors 267 N favors
       268-269
reun 257 U+20
torl 273-275 DW 5
triz 269 U+40
zeck 286 N+5 287 U-15
```

Since experience has shown that with the standard data weighting in GAMIT using the autcln rms produces short-term nrms with a median value of about 0.7 (see pshist_nrms.SUM.emed), I treat any single value or nrms that is greater than 1.5 (twice the median) as a two-sigma (95% level) outlier and apply a downweight by adding white noise with sig_neu, either enough to make the nrms about 1.0 or to increase the error bar to be significantly greater than the value of the outlier. In the case of an outlier, adding only enough white noise to equal the residual is not as objective as adding enough noise to make the outlier of negligible weight, so you should pay attention to how much the reweighted outlier is biasing the position estimate. If your primary interest is horizontal deformation, you might be very generous in the noise added to vertical outliers so as to prevent their having any effect on the horizontals.

Generate sig_neu reweights using 'grw' and add these to globk_comb.cmd using, e.g.'source ../../tables/emed_comb.reweights'.

Step 6: Generate the combined h-file for the survey

```
sh_glred -s 2004 267 2004 289 -ncomb 32 -globk_cmd_prefix COMB -expt emed -opt G
```

which will create H040923_EMED.GLX and globk_emed_04267-289.org, the latter to be checked for appropriate chi2 increments (~1.0) in the stacking.

*** Steps 1-6 to be repeated for each survey. ***

Step 7: Generate the multi-year time series

Working in a directory parallel to the survey/year directories, e.g. EMed/vsoln, create the multi-year gdl file by, e.g.

```
 ls ../????/gsoln/H*GLX > emed.gdl
```

Copy into /vsoln the template files globk_long.cmd and glorg_long.cmd from ~gg/tables and edit them for your analysis, generally needing to change only the use_site list (or 'sourced' file) and a list of edits

in the globk command file, and the stab_list list in the glorg command file.

As a check on severe conflicts due to misnamed sites, and on the application of renames for instrumental changes and earthquakes, it's a good idea at this point to run 'glist', which reads the h-files, applies renames, and compares the coordinates with those in the apr_file:

```
glist emed.gdl emed.glist +1 ../tables/itrf08_comb.eq:../tables/emed_renames.eq
               ' ' ../tables/itrf08_comb.apr  >! glist.out
```

The glist.out file will echo coordinate differences, and the emed.glist file will list all of the sites and their epochs. If you want to generate an explicit use_site list at this point, based on proximity and adequate epochs and time span, you can run 'glist2cmd' (type to see help).

Now to generate the time series, run glred in /vsoln (sh_glred works only for daily processing):

```
rm globk_rep.prt globk_rep.org globk_rep.log
glred 6 globk_rep.prt globk_rep.log emed.gdl  globk_long.cmd >&! glred.out
mv globk_rep.org globk_rep_130708a.org
```

where I've used the current date and a letter extent to identify the solution. The 'rm' step is to remove any leftover files since the new ones will be concatenated onto the end of any old ones of the same name.

Again check the stabilization quality with

```
  grep 'POS STAT' globk_rep_130708a.org
```

Then, perhaps in a sub-directory (e.g. /plots),

```
sh_plot_pos –f ../globk_rep_130708a.org –r –t1 1993-001 –t2 2013-060-30 –h - -k -v
```

View the plots (‘gs *.ps’) and make note of any problems:

```
bgnt 04  E+8
blot 04(1) N-50 E-35 typo?
bucu 04 systematic cf later, DW 5
cnca N nrms 1.6  DW N only, 5
garb 06 N-10
gr02 N nrms 1.8 DW 5 both
hoss DW N 4
```

As noted in the lecture, the small number of epochs (even for the cGPS, after aggregation) means that you have to be careful to remove or down-weight any outliers or noisy segments. As with the daily time series, the typical long-term nrms median is about 0.7, so I use the same "1.5 threshold" to indicate a 2-sigma outlier.

Step 8: Perform a velocity solution

As with continuous data, we must consider the effect of temporally correlated noise in the estimates and uncertainties of the velocities. So although we cannot determine objectively the character of correlated noise for sites with only a few epochs, to obtain a realistic uncertainty, we must nevertheless include some in the solution. For GLOBK, this is usually accomplished by applying a random walk using the mar_neu command. The value to use could be determined from prior experience in other regions, from the median of the values estimated for cGPS stations in your region, or,iteratively from a validation of the velocities against a model. A reasonable value to start would be 1 mm/sqrt(yr) for the horizontals and 3 mm/sqrt(yr) for the verticals, e.g.

```
mar_neu all .000001 .000001 .000009 0 0 0
```

Run the velocity solution:

```
rm globk_vel.prt globk_vel.org globk_vel.log
globk 6 globk_vel.prt globk_vel.log emed.gdl  globk_long.cmd  VEL >&! glred.out
mv globk_vel.org globk_vel_130708a.org
```

where the VEL option refers to the VEL keywords in globk_long.cmd and glorg_long.cmd which turn on velocity estimation. This scheme allows you to use the same command files for repeatabilities as for velocities without any explicit editing. With the org_opt setting in the template (CMDS GLDF PSUM VSUM FIXA RNRP), all of the information you need to evaluate both the stacking (globk) and the stabilized solution (glorg) will appear in the glorg print file, globk_vel.org. Things to check are the chi2 increments from the stacking (all should be of order ~1.0), the stabilization (number of sites retained in the iteration, and the N E U rms and nrms values), and the uncertainties of the velocity estimates (1-2 mm/yr or less horizontal for a 3-year span of data using at least 12 hrs of data each year and a random-walk of 1 mm/sqrt(yr).

Step 9: Generate the times series for the velocity solution

In principle, the statistics relevant to the velocity solution should flow from time series generated with a stabilization that includes all of the well-determined sites in the solution. If your solution has a strong IGS cGPS component over the full span, then the improvement using the regional stations will be minor. On the other hand, if some or all of your epochs are weak in IGS coverage, using the regional stations in the stabilization is quite important to evaluate the actual scatter of each station's epoch-by-epoch positions without confusion added by a weak stabilization at some epochs. To get the new times series, create an apr file from your velocity solution,

```
sh_exglk -f globk_vel_130708a.org -apr vel_130708a.apr
```

Next create a site_site list with all of the sites that have reasonably good position and velocity uncertainties. A utility to help with this is 'vel2stab'.

Then repeat Step 7 with the new apr-file (only) and the new stabilization list. If this time series provides new information about the editing/weighting of the coordinate estimates, then you should repeat the velocity solution (Step 8) and generation of the time series (Step 9).