Massachusetts Institute of Technology

UNAVCO

Large continuous network processing and analysis

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

Massachusetts Institute of Technology

GPS Data Processing and Analysis with GAMIT/GLOBK/TRACK
UNAVCO Headquarters, Boulder, Colorado
10–14 August 2015

Material from T. A. Herring, R. W. King, M. A. Floyd (MIT) and S. C. McClusky (now ANU)

Content

- Generating large GAMIT solutions (>50 sites)
  - Regional networks: All sites to be processed
  - Global networks: Make global networks of certain size given list of available sites
- Strategies for large network processing in GLOBK
  - Prototyping tools: Run globk command setup on time series files using tscon and glist. tsfit is used to fit and assess time series

Strategies for Large-network Processing

- Since GAMIT is limited by parameter definitions to 99 sites, with large networks, we divide the processing into sub-nets, each of 30-50 sites (processing is proportional to the cube of the number of parameters, so it's better to have more smaller sub-nets than a few large ones)
- sh_gamit can use the -netext parameter to define multiple day directories (e.g. [DDD]n1, [DDD]n2, ….)
- GLOBK is used to combine the networks for each day
- You can run htoglb to generate binary h-files (.glx) for each subnet, then use sh_glred with the LB and -net options to select the h-files to be combined
- Prototyping programs (tscon, tssum, tsfit) can be used to identify breaks and outliers before running a (time-consuming) velocity solution

Large regional networks

- Program netsel : Subnetting program for regional GPS networks

Usage:
netsel <options>
Options are
-f <file> -- List of rinex files generated with ls -s <rinex files>
-v <file> -- Globk velocity file with site coordinates
-n <number> -- number of sites per network (additional sites added for ties)
-t <number> -- Number of tie sites per network
-s <file> -- Name of station.info file to use (default ../tables/station.info)
-c <code> -- Specifies network code (2-characters). Default ne so that
networks will be ne01, ne02 .... neNN
NEW: 150512
-rw <file> <maxuse> -- sh_gen_stats .rw random walk file name and maximum horizontal
random walk value to be used. Output will be GLOBK use_site commands.
Default for <maxuse> 2 mm^2/yr

Output is nominally written to the screen but is usually redirected to a file. The -rw option is used to sub-net globk solutions

netsel output

NETSEL:
FTPLOG: PBO_2011026.rx
VELFILE: PBO_all.pos
Number of sites per net: 40
NETSEL: PBO_all.pos contains 1358 sites
NETSEL: PBO_2011026.rx contains 1234 sites
Site Range Long   122.1406 310.1850 Latitude   10.2680   82.4940 deg
NETSEL: For 1234 sites, with nominal   40 sites per network, final selection is:
NETSEL: Fin   39 sites in   32 networks with   25 sites in one network
NETSEL: Number of tie sites   1
#NETWORK Number 001 with 39 sites
NN   #   Long   Lat   Name RK
001   1   242.10350   34.12600 AZU1 13
…. List of networks

netsel output and tie

- Algorithm selects sites from highest density regions progressively working to lower density regions.
- Final network ties “centroid” sites of each network together (for case shown here only one tie site (-t option should always be >0)
- Output sites.default.yyyy.ddd to be used in gamit processing.
- -expt code and –netext are normally set to neXX where XX is network number.
- Script file with sh_gamit calls are then passed to sh_PBS_gamit when running on a cluster using Portable Batch System (PBS) and SLURM (normally needs changes for specific installation).

Global Network Selection

- Script sh_network_sel used with program global_sel to make sites.defaults.yyyy.ddd files
- This scripts ftp's lists of available data on a given day and build global networks from this list.
- The core list are 4-char codes of sites to be included if they are available
- Reference list are the initial sites in each network (next slide).
- Each network shares ties sites with each other network. Algorithm in based on keeping sites widely separated.

Reference sites

Reference site lists set initial sites in each network and the number of networks to use. (Default is ref_net.sites, selected with -f option in sh_network_sel).

REF_NET NET1 ONSA|ALGO|KOUR|S071|WDC1|WDC3

REF_NET NET2 AMC2|MATE|KHAJ|KOKB

REF_NET NET3 NYAL|CHUR|CRO1|TWTF

REF_NET NET4 GOL2|NIST|PIE1|WSRT

REF_NET NET5 BREW|STJO|IENG|NOT1

REF_NET NET6 WAB2|BRUS|NLIB|HOB2

Prototyping tools

- There are two programs that are used for prototyping solutions are:
  - tscon which converts a variety of data formats into the PBO .pos format while allowing a new reference frame realization using techniques similar to GLORG stabilization. Stabilization can used to test selection of reference sites.
  - tsfit which fits time series with a variety of models some of which can be specified in a GLOBK .eq file format. tsfit also output a globk apriori coordinate files. Use of realistic sigma option here and sh_gen_stats allows process noise to be set for globk (site dependent random walk variances)
- The program, tssum can be used to extract and append pbo time series files from globk and glred output files (normally .org files). Output of PBO format line is now default.

Prototyping concept

- The general idea of the solution prototyping is to generate an earthquake file and a list of stabilization sites that can be used in both velocity and time series analysis in GLOBK and GLRED runs. Tsfit can also be used to generate apriori coordinate files for use in tscon and globk/glred.
- GLIST can be used with eq_files and use_site type commands to get full list of sites that will be in the solution. Model summary is also now included.
- Both tscon and tsfit can read standard globk earthquake and apriori coordinate files (include EXTENDED entries). The programs do not manipulate covariance matrices and so it assumed that an initial time-series solution exists with stabilized coordinates (i.e., the output of a glred run with stabilization).

Process

- Basic processing ordering:
  - First run glred to generate time series with the pbo output option set. This solution might for example use ITRF08 sites for stabilization, or for more regionally focused networks, globk might be used for a velocity solution and the good sites from this analysis used as the stabilization sites in the glred run.
  - (There is a "catch-22" here in that knowing which sites are well behaved requires generating time series first and so these approaches tend to be iterative with the list of good sites being determined from their behavior in different analyses.)
  - Once the initial time-series are generated, tscon can be used to generate new time-series with different stabilization sites and with different apriori coordinate models than those used in the original run.
  - Analyses of these time series can be carried out using tsfit to estimate new apriori coordinate models and additional parameters associated with seasonal variations, earthquake post-seismic deformations and jumps in the time series due to antenna and the instrument changes and earthquakes.

Basic Processing (cont.)

- The statistics of the fits to the time series are generated by tsfit and these can be used to judge the quality of the analyses. The summary file output by tsfit can be used in the version of sh_gen_stats with the –ts option.
- Removal of outlier data using an n-sigma condition can also be preformed by tfsit with the output in standard eq-file format.
- The new coordinate apriori files from tsfit can be used in a new reference frame realization using tscon. The newly generated time series can be used to refine the analysis more using tsfit. Iterating the reference frame in this manner could lead to some systematic behaviors and it is ideally best to generate the reference frame with a globk solution.

Prototyping output

- At the completion of the tscon/tsfit process, there should be available an earthquake file that contains earthquakes, renames for offsets and for time series editing (renames to _XPS names), and an apriori coordinate file with optional EXTENDED entries that should provide a good match to the behavior of the time series.
- A refined list of reference frame sites and process noise models may also have been generated (sh_gen_stats).
- The earthquake and apriori file and other information can be used in an updated globk velocity solution or in glred repeatability time series run. These final globk and glred analyses should run with no major problems and would be used to generate final results.

tsfit

- tsfit is a program to fit PBO-formatted times series using a globk eathquake file input and other optional parameters (such as periodic signals). PBO format time series are generated program tssum to extract the time series. tssum allows incremental updates of time series rather the full re-generation used by ensum and multibase.
- For the prototyping role, the most important commands are eq_file (input) and out_aprf and rep_edits (outputs).
- The command line for tsfit is:

  tsfit <command file> <summary file> <list of files/file containing list>

tsfit commands

- EQ_FILE <File Name>
  - Name of standard globk earthquake file. Command may used multiple times as in the lastest version of globk.
- OUT_APRF <file name>
  - Specifies name of a globk apriori coordinate file to be generated from the fits. This file contains EXTENDED entries if needed and can be used directly in globk or tscon.
- REP_EDITS <rename file>
  - Set to report edits to file <rename file>. Edit lines start with R. The rename file if given will contain globk rename to _XPS lines.
- REAL_SIGMA
  - Apply the tsview/ensum realistic sigma algorithm to generate sigmas that account for temporal correlations in the data. This option is needed to use sh_gen_stats. Now called the FOGMEX algorithm.

Other tsfit commands

- PERIODIC <Period (days)>
  - Estimates Cosine and Sine terms with Period. This command may be issued multiple times to estimate signals with different periods.
- DETROOT <det_root>
  - String to be used at the start of the site dependent parameter estimate files. Each site generates its own file. Default is ts_. NONE generates no files
- VELFILE <vel file name>
  - Name of the output file containing velocity estimates in the standard globk velocity file format.
- NSIGMA <nsigma limit>
  - Edit time series based on a n-sigma condition.
- File names in tsfit can use the @ wild card to replace strings based on the summary file name (same as globk)

Other tsfit commands

- MAX_SIGMA <Sig N> <Sig E> <Sig U> meters
  - Allows limit to be set on sigma of data included in the solutions.
  - Default values are 0.1 meters in all three coordinates.
- TIME_RANGE <Start Date> <End Date>
  - Allows time range of data to be processed to be specified. Dates are Year Mon Day Hr Min. End date is optional.
- OUT_EQROOT <root for Earthquake files> <out days>
  - Specifies the root part of the name for earthquake estimates outputs. The outputs are in globk .vel file format and so can be used with sh_plotvel and velview. The outputs are coseismic offset and log and exponential coefficient estimates. If the <out days> argument is included the total post-seismic motion is computed that many days after each of the earthquakes. If exponential and log terms are estimated for the same event (same eq_def code) then they are summed and correlations accounted for in computing the sigmas of the total motion. Output file format is .vel file format.

tscon

- The program tscon converts timeseries from Reason/JPL/SIO XYZ files and SCEC CSV format to PBO time series format and optionally re-realizes the reference frame used to generate the time series for the format above and standard PBO time series files generated with tssum.
- The program assumes that the position time series are reported at a regular 1-day interval. This is the normal timing used in gamit for 24-hr sessions of data.
- The command line for tscon is:

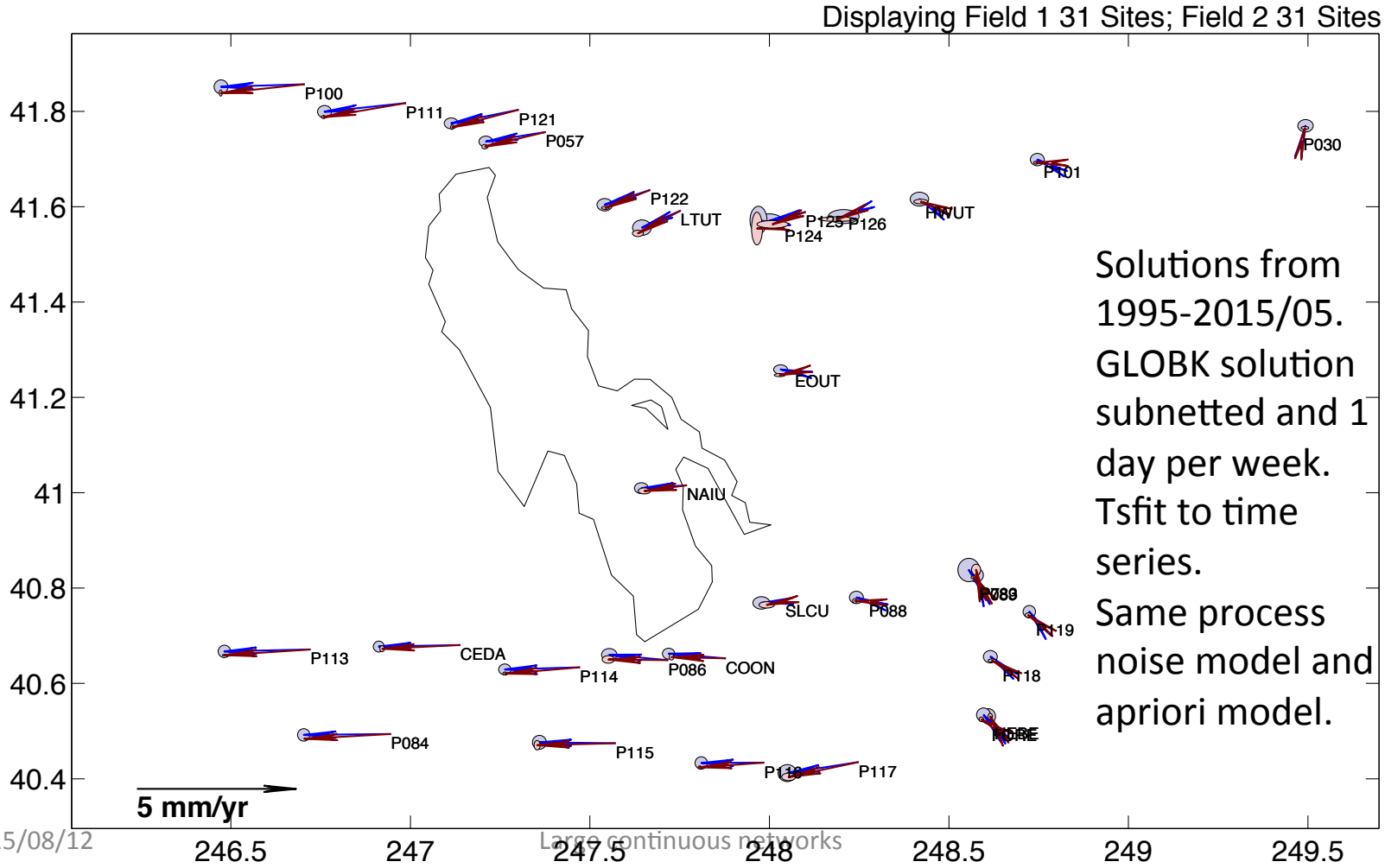
  tscon <dir> <prod_id> <cmd file> <XYZ/PBO files/file with list>

tscon commands

- Summary of commands are:
  - eq_file <file name> (maybe issued mutliple times)
  - apr_file <apriori coordinate file> (may be issued multiple times)
  - stab_site <list of stablization sites> (multiple times)
  - pos_org <xtran> <ytran> <ztran> <xrot> <yrot> <zrot> <scale>
  - stab_ite [# iterations] [Site Relative weight] [n-sigma]
  - stab_min [dHsig min pos] [dNEsig min pos]
  - cnd_hgtv [Height variance] [Sigma ratio]
  - time_range [Start YY,MM,DD,HR,MIN] [End YY,MM,DD,HR,MIN]
- These commands mimic the glorg equivalent commands and operate is very similar way. There are some small differences because tscon starts with frame realized time series.

Example: Zoom of PBO field

- Sample comparison of GLOBK and time-series analysis. Field 1 is GLOBK, Field 2 is time series analysis with tsfit



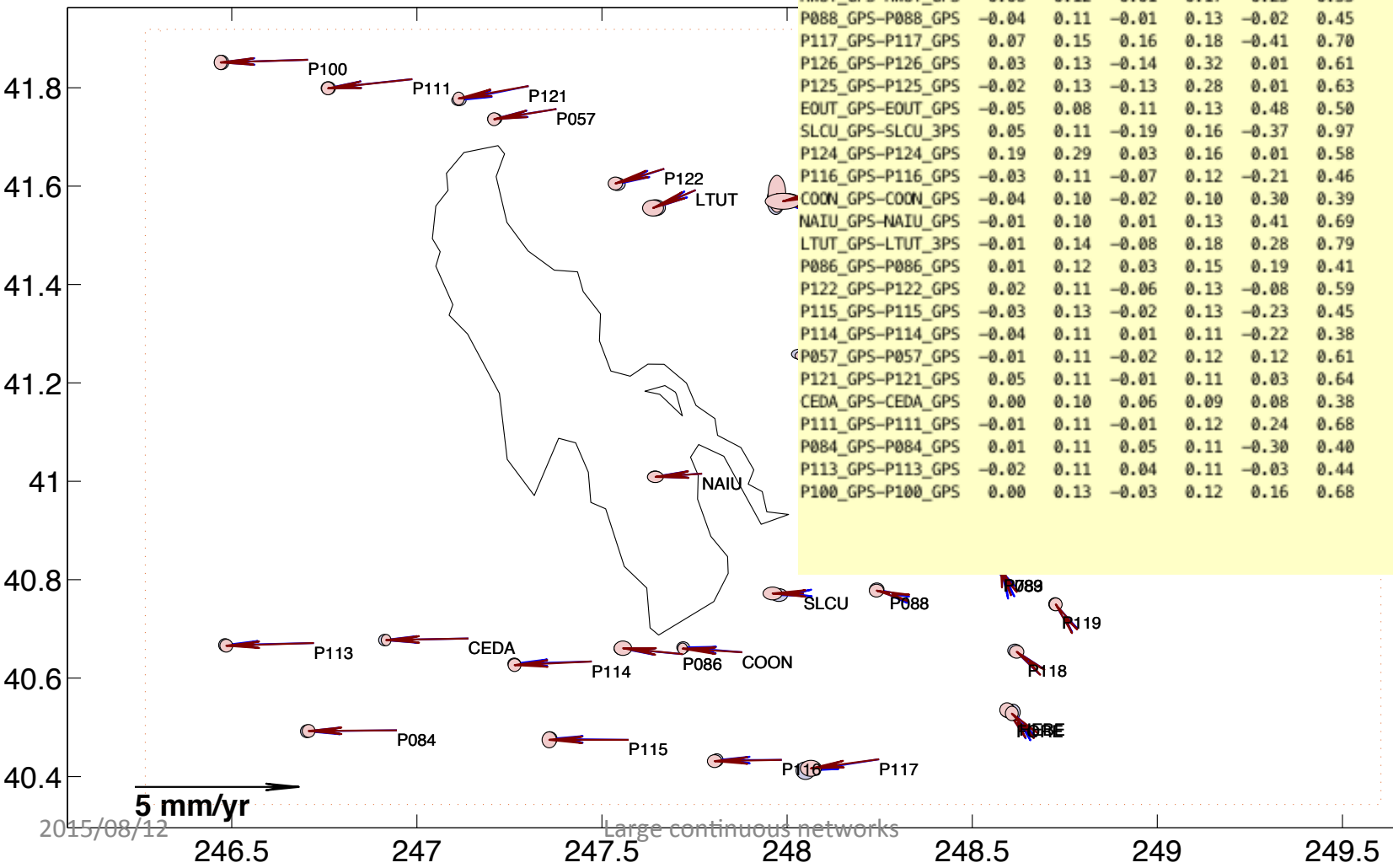
Comparison

Alignment of two fields: tsfit Kalman filter solution

Alignment Results from 31 sites and 3 param (click to remove)

Param	Est	+-	C	WRMS (mm/yr)	NRMS
dN mm/yr	-0.11	0.01	N	0.04	0.36
dE mm/yr	-0.00	0.01	E	0.06	0.49
dU mm/yr	0.63	0.04	U	0.25	0.51

Sites	NRes	+-	ERes	+-	URes	+-
P030_GPS-P030_GPS	-0.02	0.11	-0.03	0.13	0.23	0.59
P101_GPS-P101_GPS	-0.01	0.11	-0.01	0.12	0.15	0.59
P119_GPS-P119_GPS	-0.01	0.11	0.02	0.11	-0.10	0.36
P118_GPS-P118_GPS	-0.04	0.11	0.05	0.13	-0.18	0.36
HEBE_GPS-HEBE_GPS	-0.04	0.13	-0.04	0.12	0.67	0.46
FORE_GPS-FORE_GPS	0.02	0.13	-0.03	0.13	-0.14	0.47
P783_GPS-P783_GPS	0.04	0.13	-0.09	0.13	-0.09	0.52
P089_GPS-P089_GPS	0.20	0.21	0.23	0.20	-0.37	0.58
HWUT_GPS-HWUT_GPS	0.08	0.12	0.01	0.17	-0.23	0.55
P088_GPS-P088_GPS	-0.04	0.11	-0.01	0.13	-0.02	0.45
P117_GPS-P117_GPS	0.07	0.15	0.16	0.18	-0.41	0.70
P126_GPS-P126_GPS	0.03	0.13	-0.14	0.32	0.01	0.61
P125_GPS-P125_GPS	-0.02	0.13	-0.13	0.28	0.01	0.63
EOUT_GPS-EOUT_GPS	-0.05	0.08	0.11	0.13	0.48	0.50
SLCU_GPS-SLCU_3PS	0.05	0.11	-0.19	0.16	-0.37	0.97
P124_GPS-P124_GPS	0.19	0.29	0.03	0.16	0.01	0.58
P116_GPS-P116_GPS	-0.03	0.11	-0.07	0.12	-0.21	0.46
COON_GPS-COON_GPS	-0.04	0.10	-0.02	0.10	0.30	0.39
NAIU_GPS-NAIU_GPS	-0.01	0.10	0.01	0.13	0.41	0.69
LTUT_GPS-LTUT_3PS	-0.01	0.14	-0.08	0.18	0.28	0.79
P086_GPS-P086_GPS	0.01	0.12	0.03	0.15	0.19	0.41
P122_GPS-P122_GPS	0.02	0.11	-0.06	0.13	-0.08	0.59
P115_GPS-P115_GPS	-0.03	0.13	-0.02	0.13	-0.23	0.45
P114_GPS-P114_GPS	-0.04	0.11	0.01	0.11	-0.22	0.38
P057_GPS-P057_GPS	-0.01	0.11	-0.02	0.12	0.12	0.61
P121_GPS-P121_GPS	0.05	0.11	-0.01	0.11	0.03	0.64
CEDA_GPS-CEDA_GPS	0.00	0.10	0.06	0.09	0.08	0.38
P111_GPS-P111_GPS	-0.01	0.11	-0.01	0.12	0.24	0.68
P084_GPS-P084_GPS	0.01	0.11	0.05	0.11	-0.30	0.40
P113_GPS-P113_GPS	-0.02	0.11	0.04	0.11	-0.03	0.44
P100_GPS-P100_GPS	0.00	0.13	-0.03	0.12	0.16	0.68



Example Statistics

GLOBK aligned to weighted least squares (WLS) tsfit.

Param	Est	+-	C	WRMS (mm/yr)	NRMS
dN mm/yr	-0.12	0.01	N	0.04	0.48
dE mm/yr	0.00	0.01	E	0.07	0.67
dU mm/yr	0.53	0.05	U	0.26	0.54

GLOBK aligned to Kalman filter (KF) tsfit.

Param	Est	+-	C	WRMS (mm/yr)	NRMS
dN mm/yr	-0.11	0.01	N	0.04	0.36
dE mm/yr	-0.00	0.01	E	0.06	0.49
dU mm/yr	0.63	0.04	U	0.25	0.51

Comparison of individual sites: Effects of estimation mode and process noise.

```
P122_GPS Ve  -1.43 ± 0.10; Vn  -0.47 ± 0.08; Vu  -0.27 ± 0.56 mm/yr  GLOBK
P122_GPS Ve  -1.49 ± 0.09; Vn  -0.56 ± 0.08; Vu   0.28 ± 0.19 mm/yr  tsfit KF
P122_GPS Ve  -1.41 ± 0.05; Vn  -0.57 ± 0.02; Vu   0.17 ± 0.15 mm/yr  tsfit WLS

P121_GPS Ve  -2.12 ± 0.09; Vn  -0.43 ± 0.07; Vu  -0.12 ± 0.61 mm/yr GLOBK
P121_GPS Ve  -2.13 ± 0.07; Vn  -0.49 ± 0.08; Vu   0.54 ± 0.20 mm/yr tsfit KF
P121_GPS Ve  -2.09 ± 0.02; Vn  -0.55 ± 0.03; Vu   0.55 ± 0.18 mm/yr tsfit WLS
```

Some differences here in the way heights are down weighted in GLOBK frame alignment and minimum process noise values.

GLOBK Velocity Solutions

- The aim of these solutions is to combined 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:
  - Apriori coordinate and velocity file. Used as a check on positions in daily solutions (for editing of bad solutions) and adjustments are apriori values (apriori sigmas are for these values)
  - Earthquake file which specifies when earthquakes, discontinuities, and miss-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.

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. In correct solutions may not complete correctly.
- Previous methods for constructing these solutions:
  - 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.

Velocity strategies

- Other methods that are used in increase speed are:
  - 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
    - New MIDP output option refers the solutions to the middles 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
  - 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.

Summary

- Generating large GAMIT solutions (>50 sites)
  - netsel program: Divides up specific list of stations into sub-networks either for GAMIT or GLOBK processing.
  - sh_network_sel uses global_sel to make global networks of specific size and number based on a large list of available data.
- Strategies for large network processing in GLOBK
  - Prototyping tools: Run globk command setup on time series files using tscon and glist. tsfit is used to fit and assess time series.
- tsview and velview are Matlab interactive programs to assess solutions. velrot also useful for comparing velocity fields.
- Always check the on-line help for these programs because they do evolve with time.