



# Large continuous network processing and analysis

T. A. Herring   M. A. Floyd

*Massachusetts Institute of Technology, Cambridge, MA, USA*

GNSS Data Processing and Analysis with GAMIT/GLOBK and track

UNAVCO Headquarters, Boulder, Colorado, USA

24–28 August 2020

[http://geoweb.mit.edu/~floyd/courses/gg/202008\\_UNAVCO/](http://geoweb.mit.edu/~floyd/courses/gg/202008_UNAVCO/)

Material from R. W. King, T. A. Herring, M. A. Floyd (MIT) and S. C. McClusky (now at 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-networks, each of 30–50 sites
  - Processing time is proportional to the cube of the number of parameters, so it's faster to have more smaller sub-nets than a few large ones
- `sh_gamit` can use the “`-netext`” option 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 “`-opts LB`” and “`-net`” options to select the h-files to be combined. If running GAMIT in baseline mode, use the `-a` option in `htoglb`.
- 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: Sub-netting 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> -- Desired number of sites in each network (will be adjusted to make all
networks about the same size)
-t <number> -- Number of tie sites per network (must be >0)
-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

# netssel 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
```

## netse1 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.defaults.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) (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 may be used for prototyping solutions
  - `tscon` 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 be used to test selection of reference sites.
  - `tsfit` fits time series with a variety of models some of which can be specified in a GLOBK `.eq` file format. `tsfit` also outputs GLOBK a priori 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 a priori coordinate files for use in `tscon` and `globk/glred`.
- `glist` can be used with “`eq_file`” 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 a priori coordinate files (including “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).

# Basic processing

- First run `glred` to generate time series. This solution might for example use ITRF2014 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 a priori coordinate models than those used in the original run.
- Analyses of these time series can be carried out using `tsfit` to estimate new a priori 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 `tsfit` with the output in standard eq-file format.
- The new coordinate a priori 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 a priori 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 a priori 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 earthquake 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` (no longer recommended).
- 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 be used multiple times as in the latest version of GLOBK.
- OUT\_APRF <file name>
  - Specifies name of a GLOBK a priori 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 tsfit/ensum/tsview 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 (Floyd and Herring, 2020).

# 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 time series from Reason/JPL/SIO XYZ files, SCEC CSV, and UNR formats 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 `ts sum`. (Program use file extension to determine type)
- 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.
- If no `<cmd file>` is used i.e., `''` used as argument, program simply converts the file format.
- 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 multiple 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 in a very similar way. There are some small differences because tscon starts with frame realized time series.

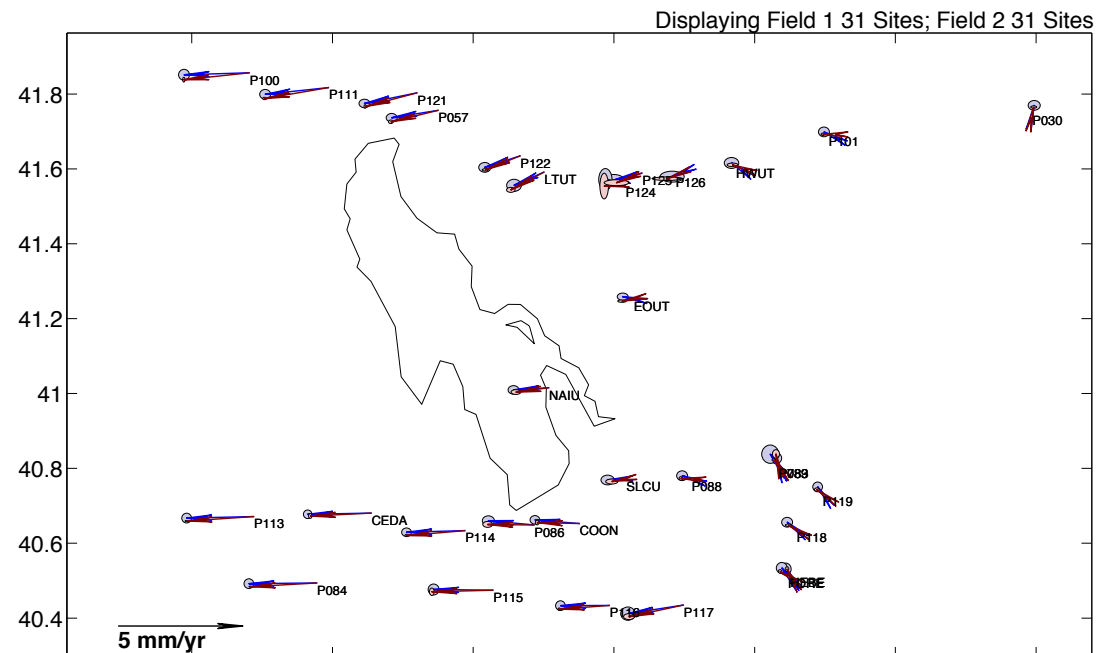
## Example: Small region of PBO field

Sample comparison of `globk` and time series analysis. Field 1 is `globk`, Field 2 is time series analysis with `tsfit`

Solutions from 1995–2015/05

- `globk` solution sub-netted and 1 day per week
- `tsfit` to time series.

Same process noise model and a priori model.



# Comparison

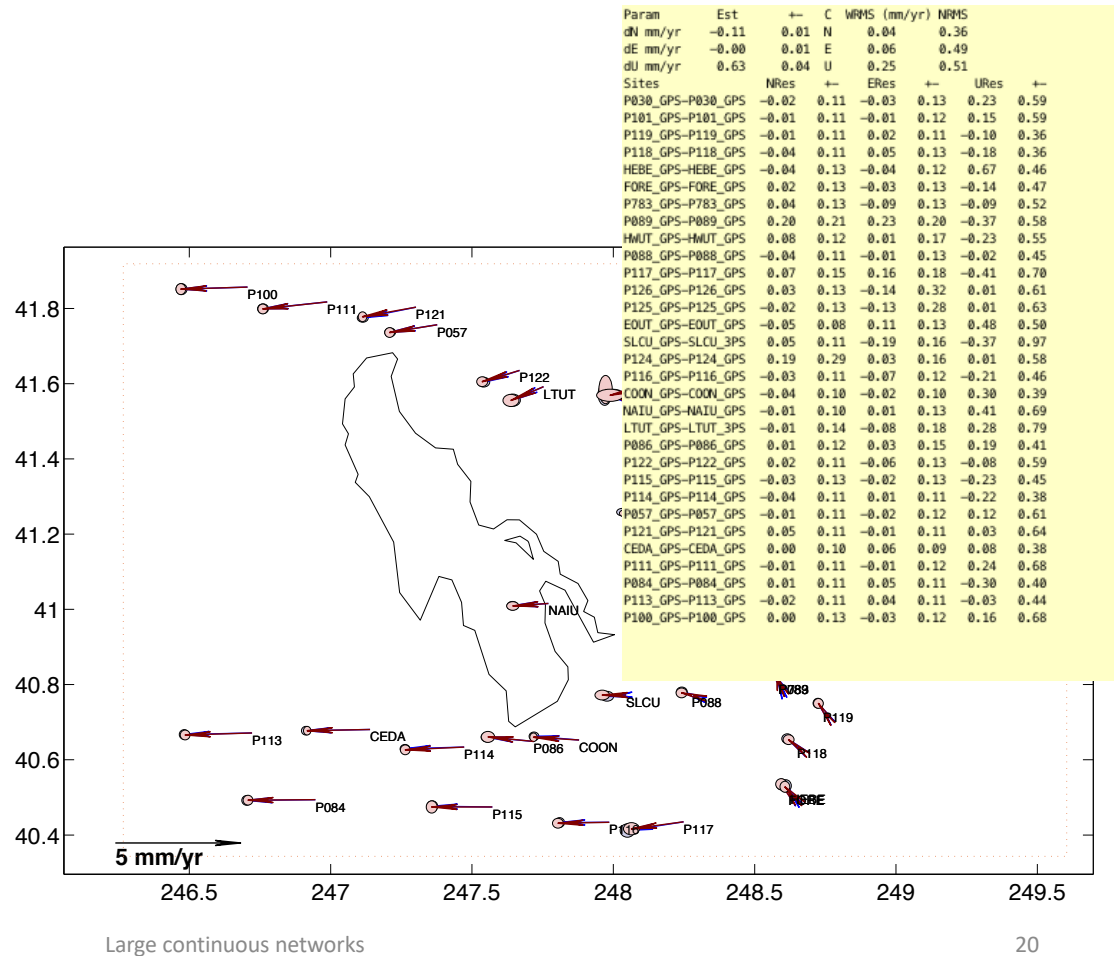
Alignment of two fields: `tsfit`  
Kalman filter solution

This gives offset and RMS differences:

NE < 0.1 mm/yr

U Mean 0.6 mm/yr/RMS 0.25 mm/yr

All NRMS values less than 0.5 meaning differences are about 0.5 sigma.



# 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:
  - 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 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 a priori 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 a priori 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
    - “MIDP” output option refers the solutions to the middles of the month. (Default is 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)
  - `net_sel` 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.