Batch processing with `sh_gamit`

T. A. Herring M. A. Floyd
Massachusetts Institute of Technology, Cambridge, MA, USA

GNSS Data Processing and Analysis with GAMIT/GLOBK and `track`

Virtual (online only) hosted by EarthScope
22–26 July 2024

http://geoweb.mit.edu/gg/courses/202407_EarthScope/

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

Outline

- Setup, operation and options for GAMIT processing with `sh_gamit`
  - Directory structures
  - Main functions in GAMIT
    - Programs called that run the GAMIT processing
  - Files that are important in processing
  - Summary files
  - Residual plots
  - Problems that can happen and suggestions

Overview of `sh_gamit`: Getting started

- `sh_setup` will create tables/ directory, then local specifics can be set
  - In tables/, process.defaults and sites.defaults are the two main files that need to be edited; sittbl. may also need editing to ensure some constrained stations in the network to be processed; sestbl. is edited if non-standard processing.
  - In tables/, a priori coordinate (.apr) file created (name in process.defaults); additional coordinates are put into tables/lfile.
- Create rinex/ directory (e.g. `mkdir rinex`) in which to store local RINEX files to be processed; RINEX data in public archives will automatically be downloaded

- `sh_gamit -expt <expt-name> -s <yr> <start-doy> <stop-doy>`
  - Common options are: `-orbit`, `-eops`, `-gnss`, `-netext`

Directory structure

- Top level: global tables and survey directories
- Within each top-level directory:
  brdc/ gfiles/ glbf/ gsoln/ igs/ rinex/ tables/
  001/ 002/ … (these directories are created as needed)
  (day directories with network extent (-netext option useful for multi-GNSS and subnetwork processing. -yrext useful for keeping years separate).
- Generally 50-60 sites is the largest network processed in GAMIT; networks larger than 99 stations require sub-netting of sites (see `netsel`, `global_sel` and `sh_network_sel`)
- Tables are linked from day directories to experiment tables/ and then to ~/gg/tables/
- GAMIT processing occurs in the day (<DDD>) directories
- GLOBK processing occurs in gsoln/

Files provided or created automatically

- Satellite orbits
  - IGS SP3-files (tabular) and/or g-files (initial conditions (ICs) for GAMIT)
  - `arc` integrates to get t-files (tabular)
- Earth orientation parameters (EOPs: ut1.; wob.): downloaded if needed for current day
- Leap second file; linked to gg/tables/ (update approximately yearly or with leap second announcements*)
- Satellite clock (j-) files; from RINEX navigation (brdc) or SP3 file
- Receiver and antenna characteristics (rcvant.dat, hi.dat); linked to gg/tables
- Differential code biases (dcb.dat); updated monthly
- Antenna phase center models (antmod.dat); linked to gg/tables/ (also needs to be updated when new antennas added)
- Lunar, solar and planetary ephemerides (nbody); linked to gg/tables/
  - Formerly tables (soltab., luntab., nutabl.) linked to gg/tables/, which needed to be updated yearly; no longer necessary
- Ocean tide grid (optional); linked to gg/tables/ from gg/GRIDS/
- Atmospheric loading grid (optional); linked to gg/tables/ from gg/GRIDS/; need to update yearly
- Mapping function grid (optional); linked to gg/tables/ from gg/GRIDS/; need to update yearly

* New Leap seconds may be eliminated soon

Files you need to worry about

- RINEX files: local plus list in sites.defaults
- Control files
  - process.defaults : minor edits for each survey
  - sestbl. : experiment and models setup (unchanged for most processing)
  - sites.defaults : sites to include or omit and source of metadata (can have year and doy extent e.g., sites.defaults.2024.193; form checked first)
  - sittbl. : sites constrained for ambiguity resolution
  - globk.cmd : `use_site`, `apr_neu`, `apr_svs`, `apr_wob`, `apr_ut1`, `sig_neu` and `mar_neu` commands
  - glorg.cmd : `apr_file`, `pos_org`, `stab_site` commands
- a priori coordinates (.apr-file, l-file)
- Metadata (station.info)
- Differential code biases (dcb.dat); download current values once per month
- Satellite characteristics (svnav.dat); download current with each new launch
  - Now linked to igs_metadata.snx, which is maintained by the IGS

process.defaults

- Controls:
  - Data and processing directory structure
  - Some session parameters (e.g. start time, length and data interval, and apr-file name)
  - Peripheral book-keeping (e.g. files to compress, archive or delete, and email address for summary)

sites.defaults

- Controls sites to be in included in experiment of given name
  - Whether or not these sites should be downloaded from a public data server (use “ftprnx” flag)
  - Whether or not these sites should have their metadata updated from the RINEX file header (not recommended, particularly for continuous sites; use “xstinfo” flag)
- May use one sites.defaults file with multiple experiment names and use “`-expt`” option in `sh_gamit` to process only certain sites
- May use a different experiment setup for different days
  - `sh_gamit` will look first for “sites.defaults.YYYY.DDD”, then a generic “sites.defaults” file

autcln.cmd

- Controls all parts of the phase cleaning algorithm
- Defaults generally work well for all experiments
  - May occasionally wish to change:
    - elevation mask
    - criteria to keep more data from sites with bad a priori co-ordinates
    - Changes are often needed for short-duration sessions (e.g., 3-hr session lengths) because autcln defaults eliminate data if durations are short because it concludes that the data are "corrupt."

.apr-file

- Controls a priori (input) coordinates of sites
- Convergence of (non-linear) processing is about 1:1000, i.e. 10 m accuracy for a priori co-ordinate will result in final coordinate accurate to about 10 mm
  - Important to have good a priori coordinates; post year 2000, RINEX header coordinates should be fine unless the receiver has been moving. (See use_rxc option process.defaults)
- Utilities include: `sh_rx2apr`
- .The experiment l-file is initialized each day with the coordinates in the .apr-file specified in process.defaults (while retaining any entries added during prior processing for sites not in the .apr_file

station.info

- Controls site occupation metadata
  - Site name
  - Start and stop times of occupation
  - Receiver and antenna information (types, serial numbers, firmware, heights)
  - With IGS20 standards, antenna azimuth is also needed (~/gg/tables/station.info.dAZ)
- Utilities include `sh_upd_stnfo` which invokes program `mstinf`
- Options for metadata include
  - Pre-prepared station.info (`sh_upd_stnfo`, `make_stnfo`)
    - Must set “xstinfo” in sites.defaults
  - RINEX headers (`sh_gamit` default but may change soon)
    - Update station.info unless an entry already exists for the day being processed or “stinf_unique” is set to “-u” in process.defaults and entry has not changed
    - Can be used with non-standard receiver and antenna names specified in guess_rcvant.dat (ideally your RINEX files have the IGS official receiver and antenna names. It is critical that this information is correct
- THIS IS A VERY IMPORTANT FILE!
  - If you do not get this file correct (and verified) before processing, you may lose a lot of time reprocessing phase data at the GAMIT (slowest) stage

sestbl. (“session table”)

- Controls processing setup
  - Observables to use (e.g. LC, L1+L2, etc.)
  - Experiment (orbits and EOPs) type
    - “BASELINE” solves for site coordinates only using fixed orbital parameters [default]
    - “ORBIT” solves for orbital parameters only using fixed site coordinates (from .apr-file)
    - “RELAX.” solves for both site and orbital parameters
  - Models used
  - Specification of the lower frequency to use is done with the -lfreq option in sh_gamit (defaults are automatically set for different GNSS systems. For L5 processing with GPS, use -lfreq 5 (but note many ground antennas don't have L5 calibrations; L2 used in these cases)

sittbl. (“sites table”)

- Controls:
  - Site-specific information for processing
    - Constraint (accuracy) of a priori coordinates in .apr-file
    - The distribution version is set for the ITRF2020 reference frame sites.

`sh_gamit` internal operation

- The following programs are run by the script:
- `makexp` and `makex` prepare the data
- `fixdrv` prepares the batch control files
- `arc` integrates GNSS satellite orbits
- `model` calculates theoretical (modeled) phase and partial derivatives of phase with respect to parameters
- `autcln` repairs cycle slips, removes phase outliers, and resolves the wide-lane ambiguities
- `solve` estimates parameters via least-squares, resolving the narrow-lane ambiguities and creating an h-file for GLOBK (user constraints are removed in the h-file to allow reference frame definition)

Steps in the standard GAMIT batch sequence

- `arc`, `model`, `autcln`, `solve` for initial solution
  - 5-minute sampling, no ambiguity resolution (GCR only)
  - update lfile. for coordinates adjusted > 30 cm
  - look at: autcln.prefit.sum; q<expt>p.ddd
- `model`, `autcln`, `solve` for final solution
  - 2-minute sampling, ambiguity resolution, data noise model based on residuals from prefit solution.
  - Look at: autcln.post.sum, q<expt>a.ddd
- Final solution repeated if NRMS reduced by > 30% from initial solution, to assure good editing and linear adjustment of parameters (original final solution files overwritten)

What `solve` produces

- Print output is the q-file, which records

  in detail

  - A constrained solution without ambiguities resolved (“GCR”)
  - A constrained solution with ambiguities resolved (“GCX”)
  - These are the solutions you should examine, along with the `autcln` summary files, to assess the quality of the solution

  and in summary only

  - A loose solution without ambiguities resolved (“GLR”)
  - A loose solution with ambiguities resolved (“GLX”)
- Updated l-file for successive iterations or days
- Useful output for GLOBK is the h-file (analogous to the IGS standard SINEX file), which contains the parameters estimates and full covariance matrix.
- (There is also an o-file, which is just the q-file but in more machine-readable form, and is seldom used; and, if orbits adjusted, an updated g-file)

A priori coordinates (`sh_gamit`)

- Create l-file in day directory by merging existing lfile. and apr_file from ../tables/ (apr_file has priority)
- If site not found in l-file
  - Use RINEX header coordinates ("`set use_rxc = Y`" in process.defaults, good for modern (post selective availability, in 2000) data

  or
  - Use pseudorange data in RINEX file to estimate point position or differential position relative to a site in sites.defaults ("`set use_rxc = N`", default)
- During the `sh_gamit` run, the coordinates are updated (and copied to ../tables/lfile.) if they are in error by > 30 cm

Ambiguity resolution

- (L2 – L1) integers resolved by `autcln` and passed to `solve` in the n-file ("LC_AUTCLN" option in sestbl.)
  - Weak dependence on geometry
  - Need current differential code bias file dcb.dat
  - Use "LC_HELP" for codeless data (before ~ 1995) or if problems (default max distance is 500 km)
- Narrow-lane (L1) resolved by solve
  - Strong dependence on phase noise and models
  - 5–10 cm constraints on a priori coordinates usually sufficient

sh_gamit_<DDD>.summary (also email)

- Preamble

```
Input options -d 2002 30 31 32 33 -expt ncar -pres ELEV -yrext -netext a
Processing 2002 031 GPS week 1151 4 Raw 2
/data51/tah/SENH02/glob02/suomi/2002_031a
Disk Usage:  12678.4  Free  76447.4 Mbyte. Used 15%
```

- Summary Statistics (from autcln)

```
Number of stations used 4 Total xfiles 4
Postfit RMS rms, to and by satellite
RMS  IT Site   All  01  02  03  04  05  06  07  08  09 ...
RMS  20 ALL    4.8   4   5   6   5   5   4   5   4   5 ...
Best and Worst two sites:
RMS  20 TMGO   3.2   3   3   4   4   4   3   3   3   4 ...
RMS  20 SA09   4.6   4   4   5   4   5   4   4   4   5 ...
RMS  20 PLTC   5.4   4   5   5   6   5   4   5   5   6 ...
RMS  20 SA13   5.5   5   5   6   5   5   5   5   5   6 ...
```

sh_gamit_<DDD>.summary (also email)

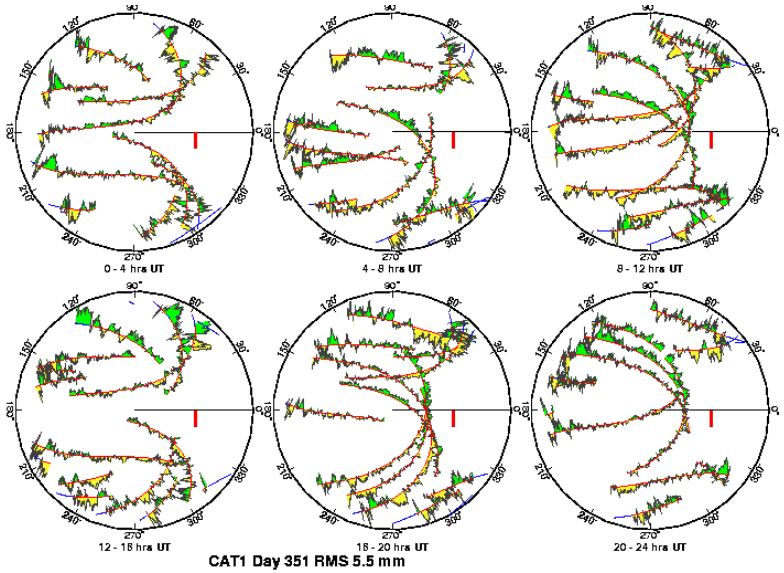
- Solution statistics from solve

```
Double difference statistics
 Prefit nrms:  0.31280E+03    Postfit nrms: 0.21324E+00     (Constrained free)
 Prefit nrms:  0.31185E+03    Postfit nrms: 0.21818E+00     (Constrained fixed)
 Prefit nrms:  0.31272E+03    Postfit nrms: 0.20470E+00     (Loose free)
 Prefit nrms:  0.31185E+03    Postfit nrms: 0.20756E+00     (Loose fixed)
 Number of double differences:  12447
 Numbers of WL and NL biases 120   Percent fixed 95% WL  85% NL
```

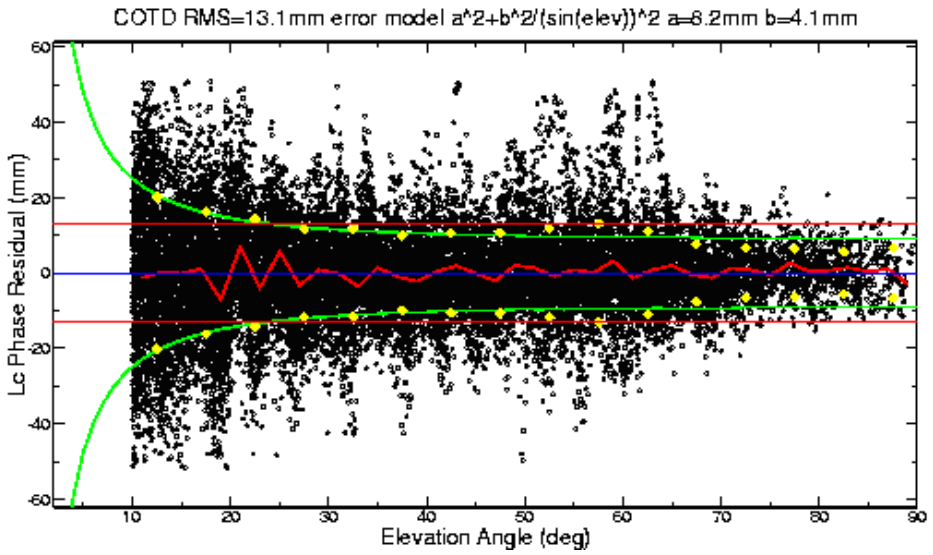
- Any large adjustments to positions (> 0.3 m) ...
- Things to note:
  - Number of stations matches expectation
  - Site postfit RMS values 3–10 mm
  - No stations with RMS = 0 (implies no data retained by autcln)
  - Postfit nrms from solve ~ 0.2 for constrained and loose solutions
  - “Most” ambiguities resolved (70–85% for noisy days, > 90% for best)

Phase residual plots

- Set with "`-pres elev`" in `sh_gamit` command (requires GMT)
- Postscript files in day directory, by default converted to PNG in figs/ directory and then erased (requires GMT's `psconvert` or `ps2raster`)
- Use to assess multipath, water vapor, and antenna phase center model

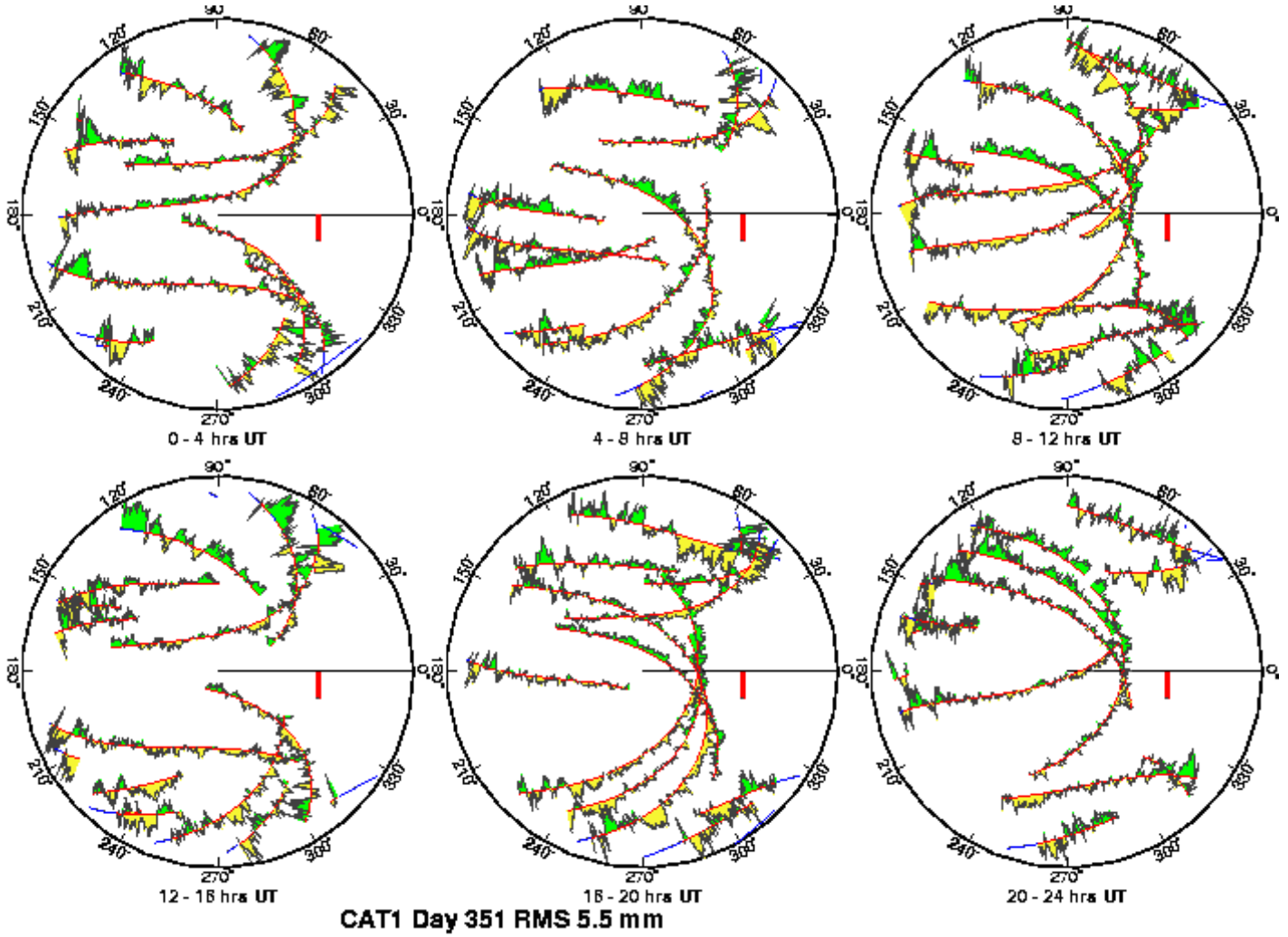


"Sky plot"



Phase vs elevation angle

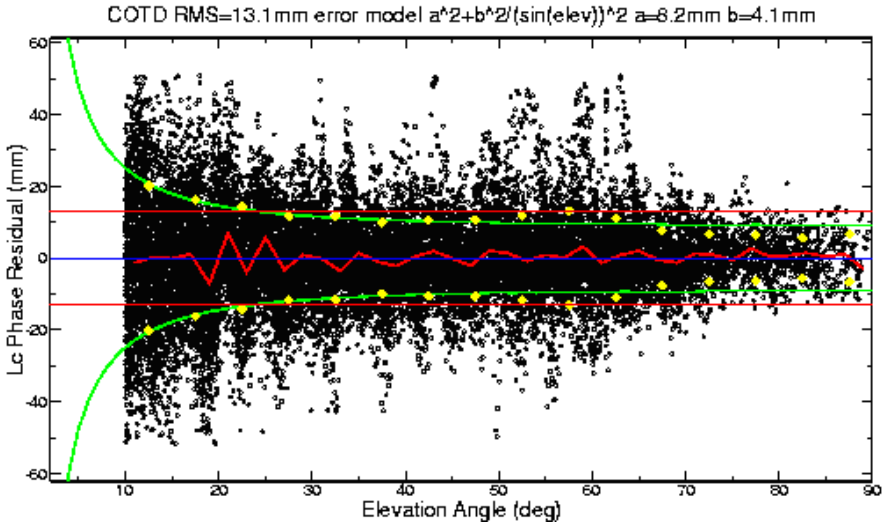
Sky plots



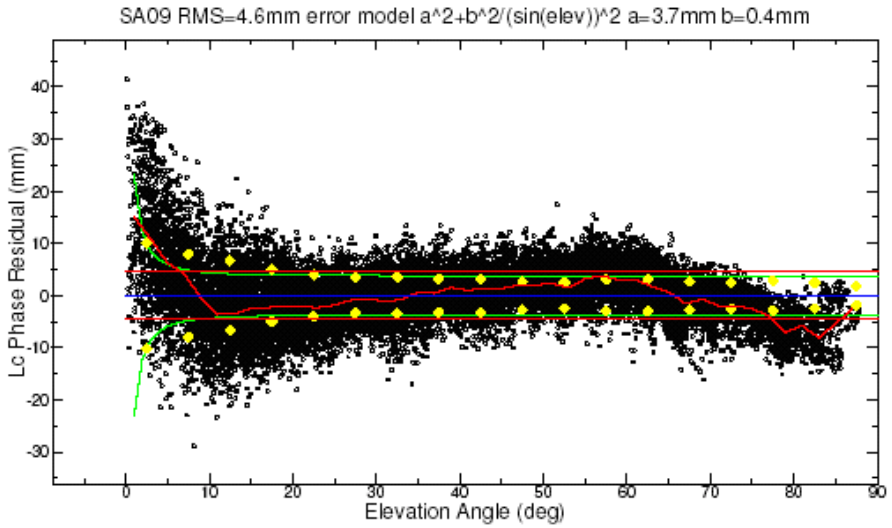
High residuals in the same place at different times suggest multipath

High residuals appearing in a given place only at one time suggest water vapor

Phase vs elevation angle



Normal pattern: bands are high-frequency multipath; red is smoothing of individual values, showing no strong systematics. Mid-elevation angle noise could be atmospheric delay errors?



Bad pattern: systematic signature of smoothed values indicates a poor model of the antenna phase pattern (perhaps a misidentified antenna in station.info)

(Green lines show the elevation-dependent noise model shown at top and used to reweight the data in solve)

What can go wrong?

- Site missing (not listed)
  - No RINEX data within session span: check RINEX file and/or makex.expt.infor
  - Too few data, x-file too small and not used: check RINEX file size, change “minxf” in process.defaults
- Site in solution but no data or adjustment
  - a priori coordinates > 10 m off: check range rms in autcln.prefit.sum,
    - run `sh_rx2apr` differentially for several RINEX files
  - bad receiver: examine RINEX files or initial c-files with `cview`
- q-file nrms > 0.2
  - solution over-constrained: check GCX vs GLX nrms, rerun with only one site constrained in the sittbl.

Problems with a priori coordinates

- Need to be good to < 10 m to get through `autcln`
- Safest source is a previous solution or a pseudorange solution using `svsp3` (`sh_rx2apr` still used svpos/svdiff only rinex 2)
- Range rms and bias flags added from `autcln` summary file are a useful check
- Convergence is 1:100 to 1:1000 (1 m error in .apr-file can lead to 1–10 mm error in adjustment), hence automatic update of l-file for iteration of second GAMIT solution
- Watch for repeated updates in email summary as a sign of bad data

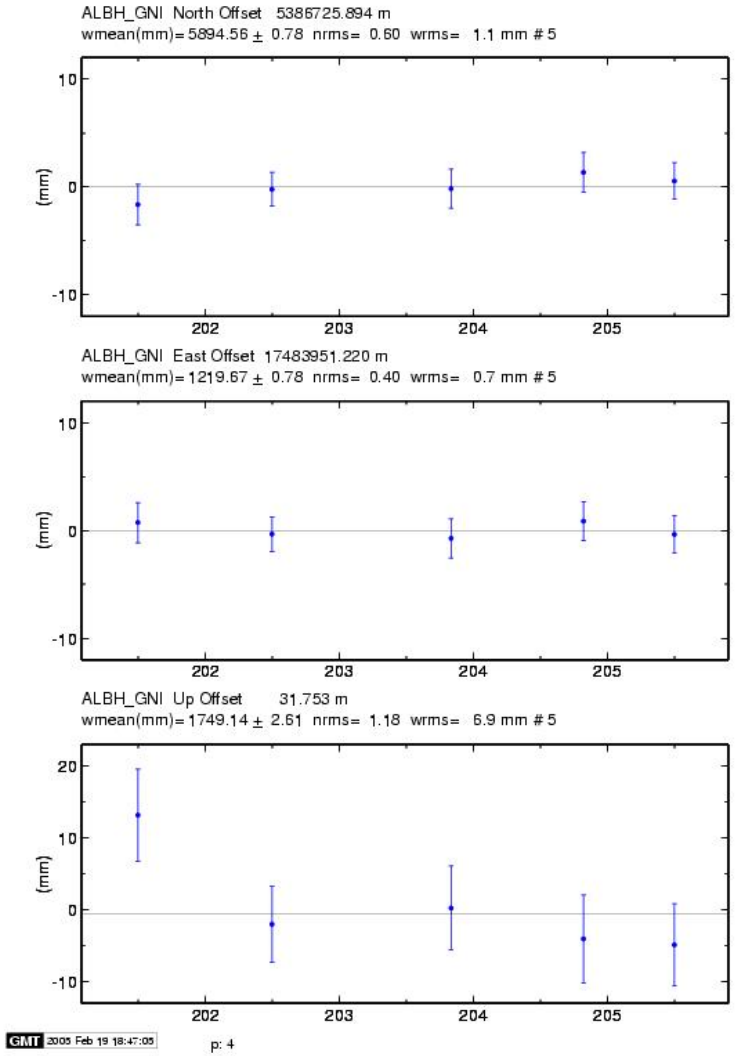
Constraining the GAMIT solution

- Minimal (single-station) constraint is all that's needed for ambiguity resolution, but sittbl. can list several to assure one
- Orbits can be fixed or tightly constrained (.005 ppm) for IGS orbits since at least 1996
  - Use of "BASELINE" mode with "igsf" orbits fixed now recommended for processing regions up to at least 6000 km. Repro3 igsf orbits are now multi-GNSS (automatic selection from CDDIS)
- Look for good (~0.2) loose (GLR/GLX) nrms but elevated constrained nrms (GCR/GCX) as indication of an over-constrained solution

More subtle problems

- Site with high rms in autcln.post.sum
  - High multipath or water vapor: check sky plots of phase
  - Bad receiver: examine RINEX files or initial c-files with `cview`
- Phase vs elevation angle plot large and systematic
  - Misidentified antenna (wrong PCV model)
  - Coupling between antenna and mount
- GAMIT results within normal range but time series shows outlier
  - Survey-mode: antenna not leveled and centered over mark
  - Change in multipath (water, objects) or water vapor
  - Snow on antenna
  - Incorrect ambiguity resolution (east component except for high latitudes)

Example of understanding outliers

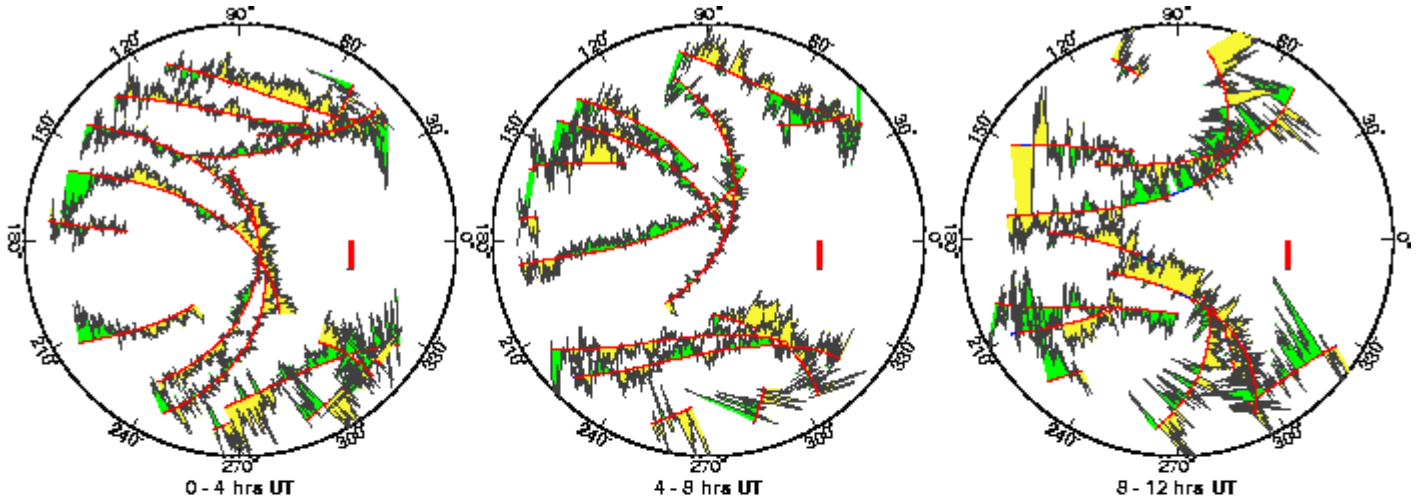


`autcln` **RMS:**

- Day 201 9.6 mm
- Day 202 6.0 mm

Notice height outlier on day 201

ALBH 2003 Day 201



ALBH 2003 Day 202