# Generating time series with `glred`

M. A. Floyd  T. A. Herring
*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/

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

# Post-processing: GLOBK

- Convert ASCII h-files to binary h-files (`htoglb` in glbf/)
- Generate and chronological list of binary h-files (`glist` in gsoln/)
- At this point, diverge in approach depending on solution sought
  - More details about `glred`, `globk` and `glorg` in advanced course lectures
- Similarly to `sh_gamit`, the batch script `sh_glred` will run all of the above command steps (and more, introduced in next slides)
  - User may just need to edit `globk` and/or `glorg` command files to achieve most desired types of solution

# sh_glred

- glred is just a way of invoking globk to process one day at a time; sh_glred is a script to invoke glred easily for a sequence of days
- Once you've run sh_gamit for a sequence of days, you will have on each day an h-file of loosely constrained parameter estimates and covariances. If you have appropriately constructed command files for globk (globk.cmd) and glorg (glorg.cmd) in [expt]/gsoln/, you can obtain time series using, e.g.

```
sh_glred -expt [expt] -s [start yr] [start_doy] [stop yr] [stop doy] -opt H G T
```

  which will translate the GAMIT plain text h-files into GLOBK binary h-files (H), run GLOBK (G) and run sh_plot_pos (T)
- Self-guided templates will guide you in constructing the command files (~/gg/tables/globk.cmd and ~/gg/tables/glorg.cmd)

# Manual sequence

- `htoglb` (i.e. `sh_glred -opt H`)
  - Converts plain text h-files output from GAMIT to binary h-files (in glbf/) for input to GLOBK
- `ls`
  - Create list of binary h-files to process (in gsoln/)
- `glist`
  - Create chronological list of h-files to process and associated information
- `glred` (i.e. `sh_glred -opt G`)
  - Create ".org"-file(s) with individual solutions
- `sh_plot_pos` (i.e. `sh_glred -opt T`)
  - Create ".pos" (time series) file(s) and time series plots
- `globk`
  - Create combined (or velocity) solution
- `glorg`
  - Additional `glorg` runs for different reference frames

# htoglb

- Creates binary h-files for input to GLOBK
  - All metadata etc. carried forward from GAMIT
- Not restricted to plain text h-files from GAMIT
  - May also use SINEX (Software INdependent EXchange format), GIPSY's "stacov" files, etc.
  - But beware of constraints implicit in solutions from other software/processing runs!
- For example, from glbf/

```
htoglb . '' ../[0-3][0-9][0-9]/h*a.*
```

- (Use pair of quotes '' if no satellite data is to be saved. This is normal case).

# GLOBK checks

- List files to be processed by GLOBK, e.g. from gsoln/
  `ls ../glbf/h*.glx > expt.glx.gdl`
- Run pre-processing checks using `glist`

```
glist expt.glx.gdl 201407_NSFBay.sum +1 ~/gg/tables/itrf08_comb.eq:A 201407_NSFBay.gdl
```

  - This will also calculate if any overlapping h-files should be combined with `glred` (e.g. multiple networks on the same day)
- Inspect any errors (e.g. site name clashes)

# Create time series

- `glred` simply runs the main program, `globk`, once per interval (e.g. daily) to combine data over that interval into one solution and one effective time series point
  `glred 6 glred_20150811.prt glred_20150811.log 201407_NSFBay.gdl globk.cmd`
  - Assess solution by looking at "POS STATISTICS" lines
- Old example (Example 2 on this course's web page) using `sh_glred` with "`-opt E`" creates:
  - "mb"-files (time series) with `multibase`
  - "psbase"-files (PostScript) with `sh_baseline`
- Updated, preferred method is `sh_glred` with "`-opt T`":
  - `tssum` to create ".pos"-files (time series) from ".org"-file output from `glred`
  - `sh_plot_pos` to create PostScript plots
    - ".org"-file may be input to `sh_plot_pos`, which will run `tssum` for you, e.g.
      `sh_plot_pos -f glred_YYYYMMDD.org -d _YYYYMMDD` ...

# Time series solution files

**Old scheme**

- “.org”-file
- `ensum`
  - “VAL”-file (time series values)
  - “SUM”-file (statistics)
- `multibase`
  - “mb”-files
- `sh_baseline`
  - Time series plots

`sh_plotcrd`

**Current scheme**

- “.org”-file
- `tssum`
  - “.pos”-files
  - `tsfit`
    - “.res”-files
- `sh_plot_pos`
  - Time series plots

`sh_plot_pos`

# Recommended strategy for stabilization

- In the template files, globk.cmd and glorg.cmd:
  - default apr-file is ~/gg/tables/igb14_comb.apr
  - default eq-file is ~/gg/tables/igb14_comb.eq
  - default stab-file is ~/gg/tables/igb14_comb.stab_site
- igb14_comb.apr is a combined .apr-file, using many publicly available coordinate sources, all aligned to ITRF2014
- igb14_comb.eq is the associated .eq-file with defined discontinuities
  - equipment changes
  - earthquakes
  - etc.
- igb14_hierarchy.stab_site uses the established IGS core network hierarchy to choose stabilizing sites, e.g.
  - `stab_site DRAO/BREW/NANO/ALBH/HOLB`
    means use DRAO if available in the solution (e.g. h-files), otherwise use BREW if available, otherwise use NANO, etc.
- Equivalent files for previous realizations of ITRF2014 (“itrf2014…”, “igs14…”) and ITRF2008 (“itrf08…”, “igb08…”) still available

# Inspect consistency of stabilization statistically

- It is a good idea to have thought about your reference frame stabilization when setting up your experiment, e.g. sites.defaults, before running `sh_gamit`
- Desire as many well-defined (e.g. IGS) sites as possible for redundancy
  - Recommended to use some of the sites (preferring the first column) in ~/gg/tables/igb14_hierarchy.stab_site when selecting your processing network, e.g. additional sites listed in your sites.defaults
  - But remember trade-off with processing time, e.g. processing time scales proportionally to $n^3$

```
grep '^POS S' glred_20150811.org
POS STATISTICS: For   51 RefSites WRMS ENU   2.15   2.55   6.19  mm    NRMS ENU   0.71   0.84   0.63 L0104260000_tg1a.glx
POS STATISTICS: For   54 RefSites WRMS ENU   2.17   2.42   6.03  mm    NRMS ENU   0.74   0.80   0.63 L0104270000_tg1a.glx
POS STATISTICS: For   50 RefSites WRMS ENU   2.12   2.25   6.34  mm    NRMS ENU   0.71   0.75   0.67 L0104280000_tg1a.glx
POS STATISTICS: For   54 RefSites WRMS ENU   2.19   2.31   5.23  mm    NRMS ENU   0.80   0.81   0.58 L0104300000_tg1a.glx
POS STATISTICS: For   54 RefSites WRMS ENU   1.83   2.17   6.34  mm    NRMS ENU   0.64   0.75   0.68 L0105010000_tg1a.glx
POS STATISTICS: For   54 RefSites WRMS ENU   2.09   2.63   6.47  mm    NRMS ENU   0.80   0.98   0.75 L0105020000_tg1a.glx
```

# “.pos”-files

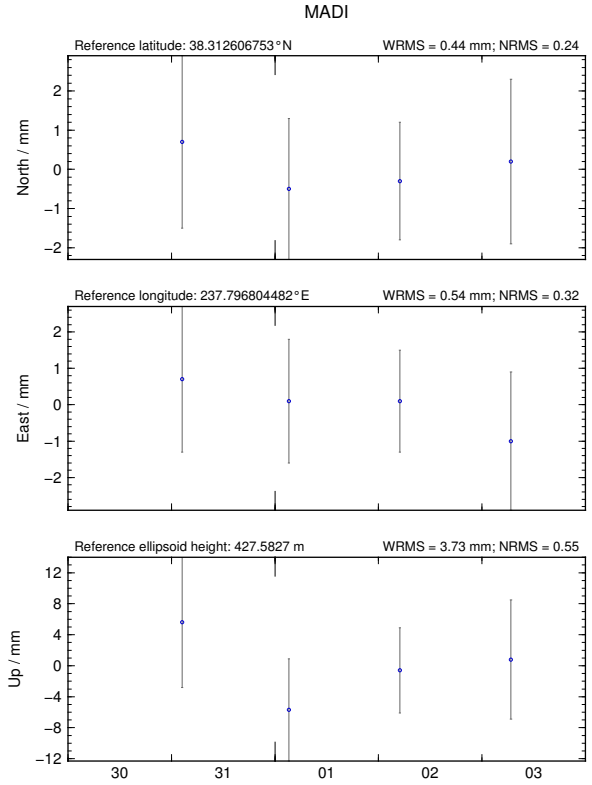
- These contain your time series solution
- Long format in various coordinate systems
  - Geocentric (X, Y, Z)
  - Geodetic (lon., lat., height)
  - Local (east, north, up)
- Can be input to `tsfit` (interactive version of GGMatlab tool “tsview”), `sh_cats` (requires CATS) and `sh_hector` (requires Hector)
- Both “.pos”-files and “.res”-files can be plotted with `sh_plot_pos`

# sh_plot_pos

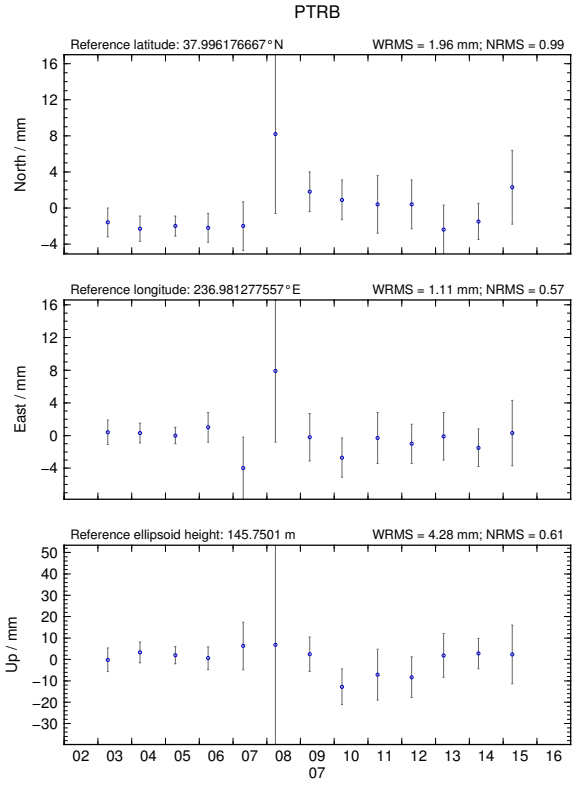
- Uses GMT and has many features including options to:
  - Read in ".org"-files, ".pos"-files (output of `tssum`) and ".res"-files (output of `tsfit`) [`-f` option]
  - Run `tsfit` (GLOBK's curve-fitting module) on input ".pos"-files [`-t` option]
  - Calculate basic statistics (e.g. WRMS, NRMS)
  - Add vertical lines at epochs specified by renames, earthquakes or user [`-b`, `-e` and `-l` options, respectively]
  - Specify fixed start and end times of time series [`-t1`, `-t2` options]
  - etc.

# Inspect consistency of time series

## Good repeatability




## Outlier



# 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` and `globk` runs

# Iterating your solution

- First time series may only be stabilized by previously well-defined sites, e.g. ITRF sites
- Once a high-quality position (and velocity) estimate for a previously unknown or new site is available, we can use all sites to stabilize
- This approach may be used with both time series (e.g. `glred`) and velocity (e.g. `globk`) solutions

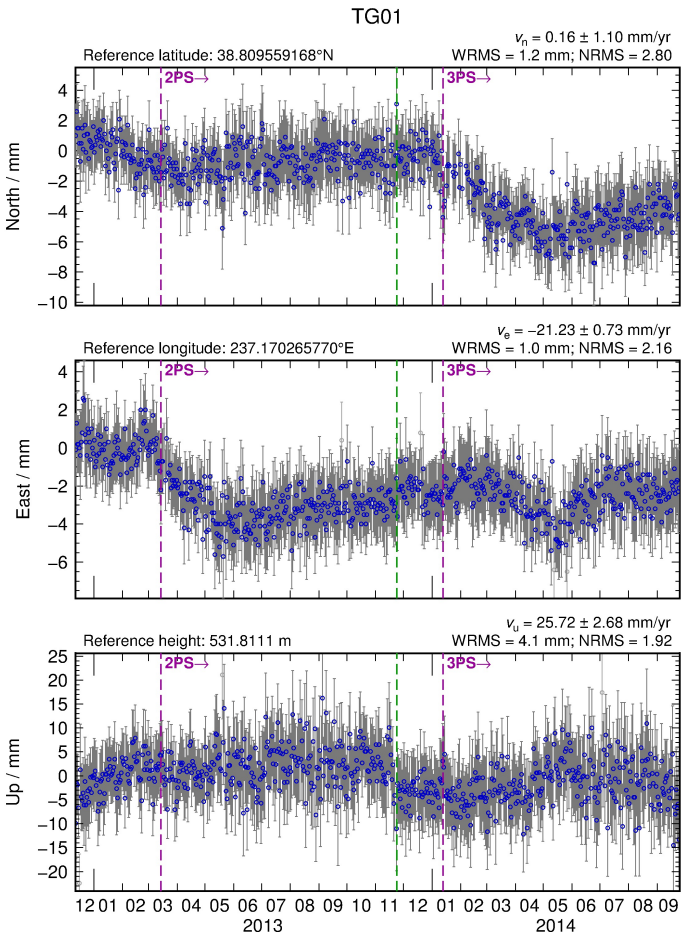
# Changes to equipment

- Antenna is main concern, although receiver may also affect continuity of calculated position
  - Even antennas of the same model type may not be manufactured within acceptable geodetic tolerance
- `stinf_to_rename` useful for creating automatic list

# Changes in multipath environment





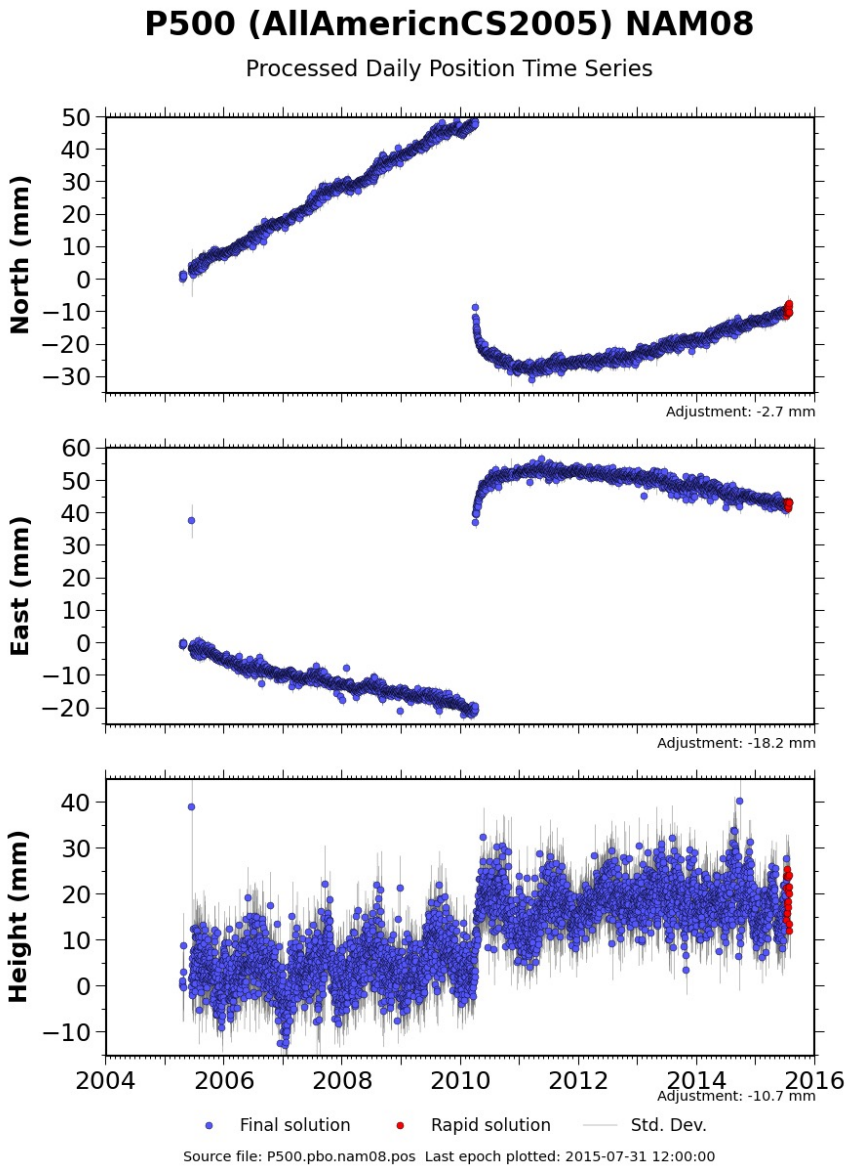


# Rename scheme

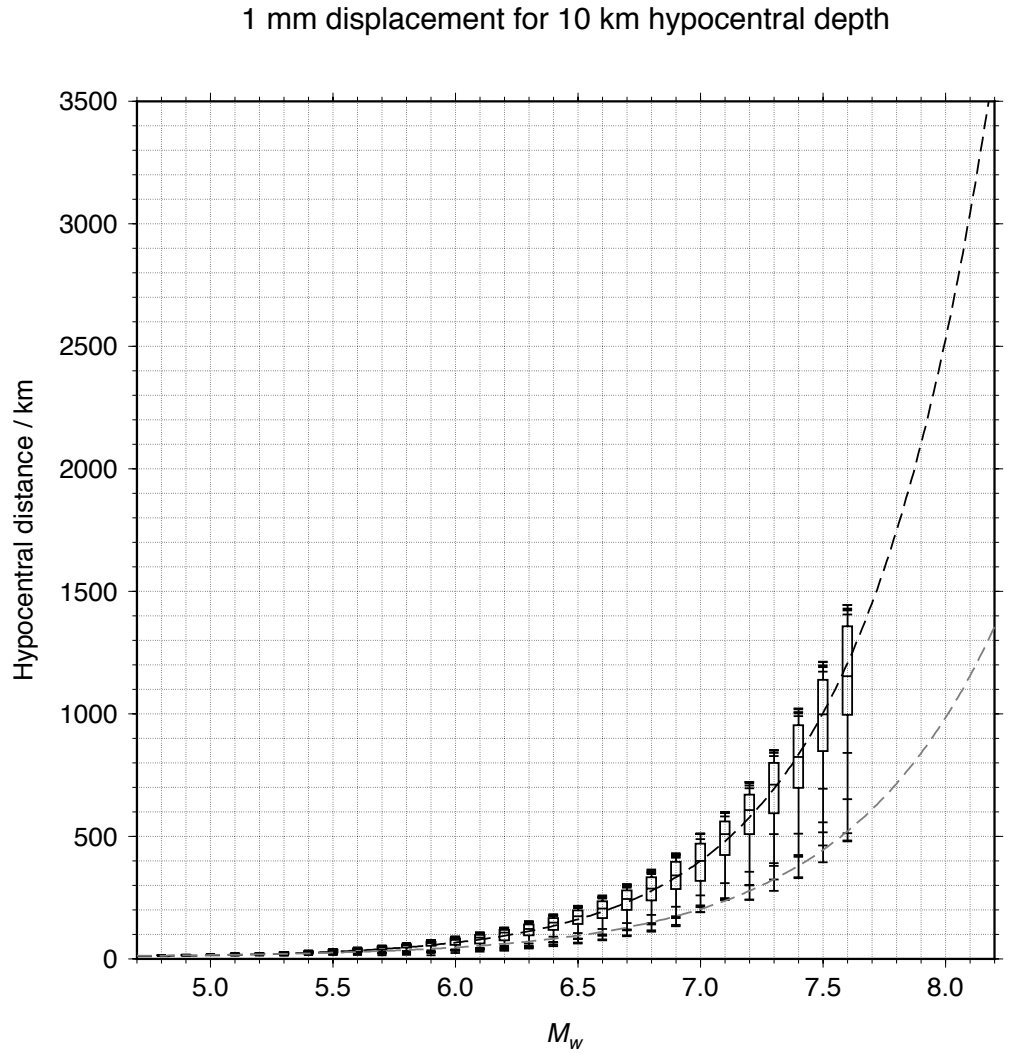
- Rename scheme uses first character of suffix
  - XXXX_GPS → XXXX_2PS → XXXX_3PS → … etc.
- Beyond “9PS”, convert to 10th letter of alphabet (“JPS”) and beyond
  - … → XXXX_9PS → XXXX_JPS → XXXX_KPS → … etc.
  - But remember “XPS” is a special case for excluding sites from `globk` runs
- For manual renames (e.g. due to manual inspection rather than known events), use 8 available letters in alphabet before “J”
  - XXXX_APS, XXXX_BPS, etc.
  - But remember “GPS” is a special case (the default)

# Earthquakes

- Earthquakes occur at known times
- May exhibit more than just a discontinuity
- Take care when earthquake occurs in the middle of processing day
  - Some data will fall before and some after ground displacement
  - Time series point on day of earthquake may appear between pre-earthquake and post-earthquake position



# Radius of influence




- "eq_def" line in eq-file contains earthquake ID (two characters), location, etc.
  - ID is used to substitute last two characters of 8-character site name, e.g. XXXX_GPS → XXXX_GSN
- `sh_makeeqdef` will search archives (ANSS ComCat or ISC) to generate "eq_def" records

# Recommendations

- Know your goals
  - Only fit “nuisance” terms
  - It is usually best not to try to fit signals that you are interested in, e.g. seasonal terms if you are studying these.
- Depending on your goal (e.g. linear tectonic velocities), sometimes you just have to abandon data as it is likely to do more harm than good (rename to xxxx_XPS or xxxx_XCL)
  - Adding large process noise in `globk` is one approach but be careful not to make too large
  - GLOBK “sig_neu” command can be used for small duration “bad” events

# Short- vs long-term time series

- Exactly the same procedure is used for short (e.g. survey) and long (e.g. years of continuous data) time series
- The only difference may be the number and type of input h-files, e.g.
  - Daily survey h-files (short-term time series)
  - Combine into one solution (short-term position combination)
  - Several combined survey files over years (long-term time series)
  - Several combined survey files over years (long-term velocity combination)

# `tsfit` and tsview

- `tsfit` is the command-line tool for fitting time series and generating statistics
  - Input ".pos"-files, optionally .eq-files
  - Fits linear rate and choice of common parameters
    - Periodic terms
    - Discontinuities and earthquakes
    - Post-seismic decays
  - Outputs
    - statistics of fit
    - standard (position and velocity) .apr-files
    - extended (periodic, logarithmic decay, etc.) .apr-files
    - Residuals to fit (".res"-files)
- tsview is an alternative that, via a MATLAB interface, allows interaction

# Time series characteristics

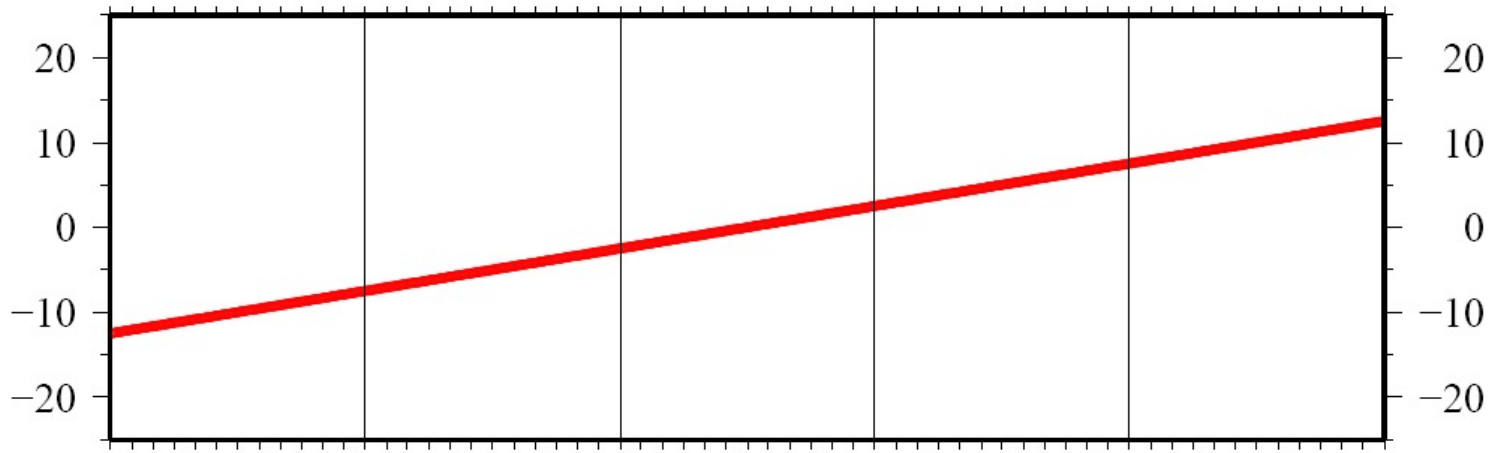
# Time series components

observed position

(linear) velocity term

$$x^i = x_0^i + v^i(t - t_0)$$

initial position



# Time series components

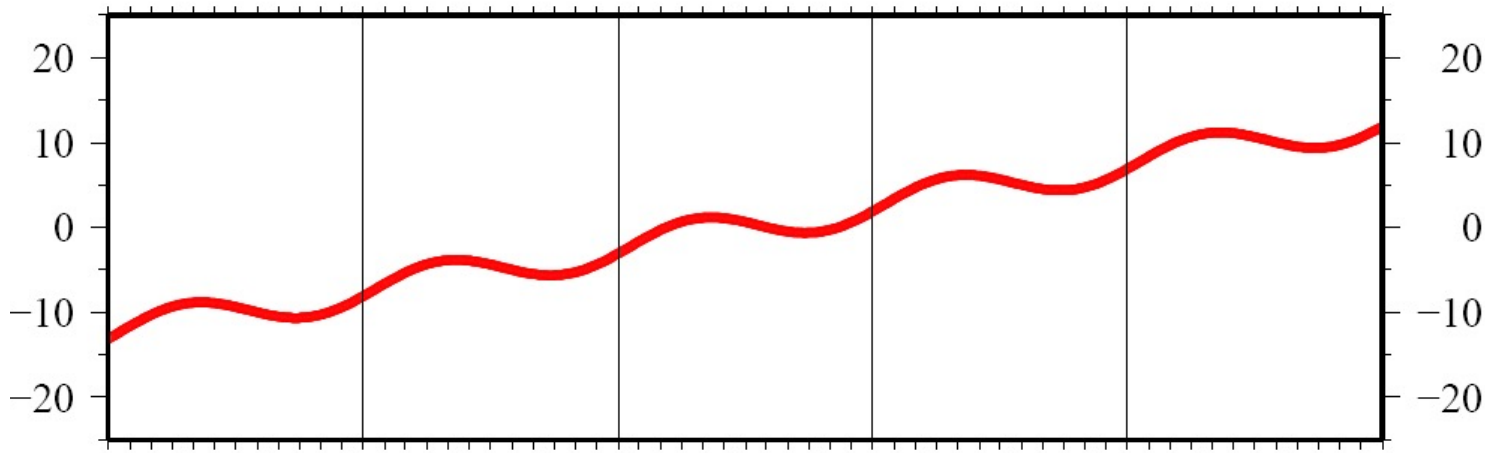
observed position

(linear) velocity term

$$x^i = x_0^i + v^i(t - t_0) + A_0^i \cos\left(\frac{2\pi(t - t_0)}{T_0} - \tau_0\right)$$

initial position

annual period sinusoid



# Time series components

observed position

(linear) velocity term

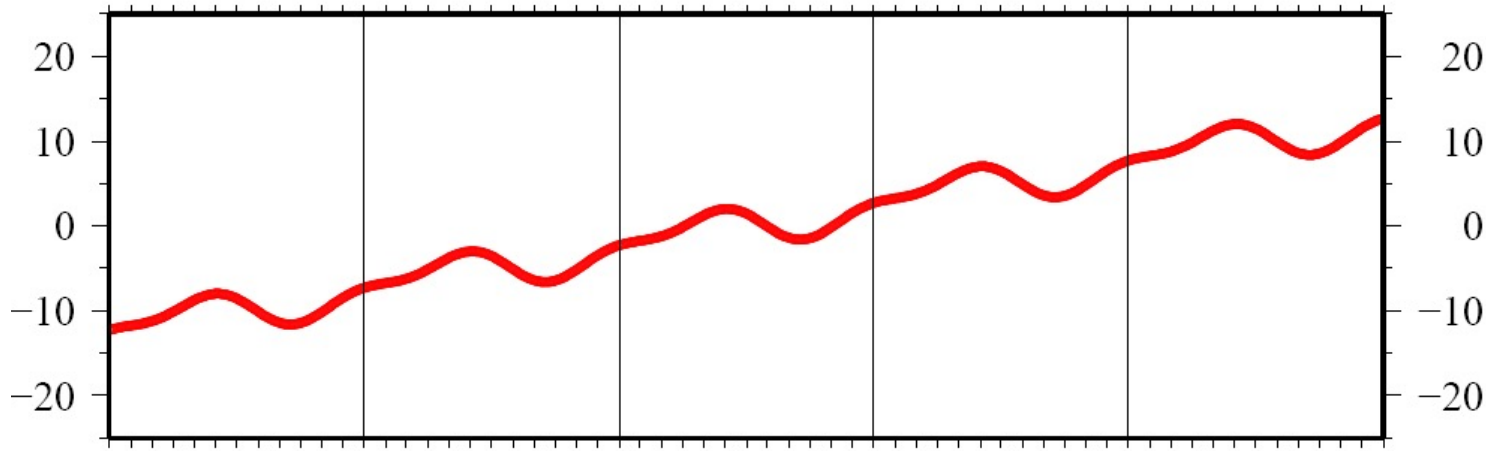
$$x^i = x_0^i + v^i(t - t_0) + A_0^i \cos\left(\frac{2\pi(t - t_0)}{T_0} - \tau_0\right) + A_1^i \cos\left(\frac{2\pi(t - t_0)}{T_1} - \tau_1\right)$$

initial position

annual period sinusoid

semi-annual period sinusoid

seasonal term



# Time series components

observed position

(linear) velocity term

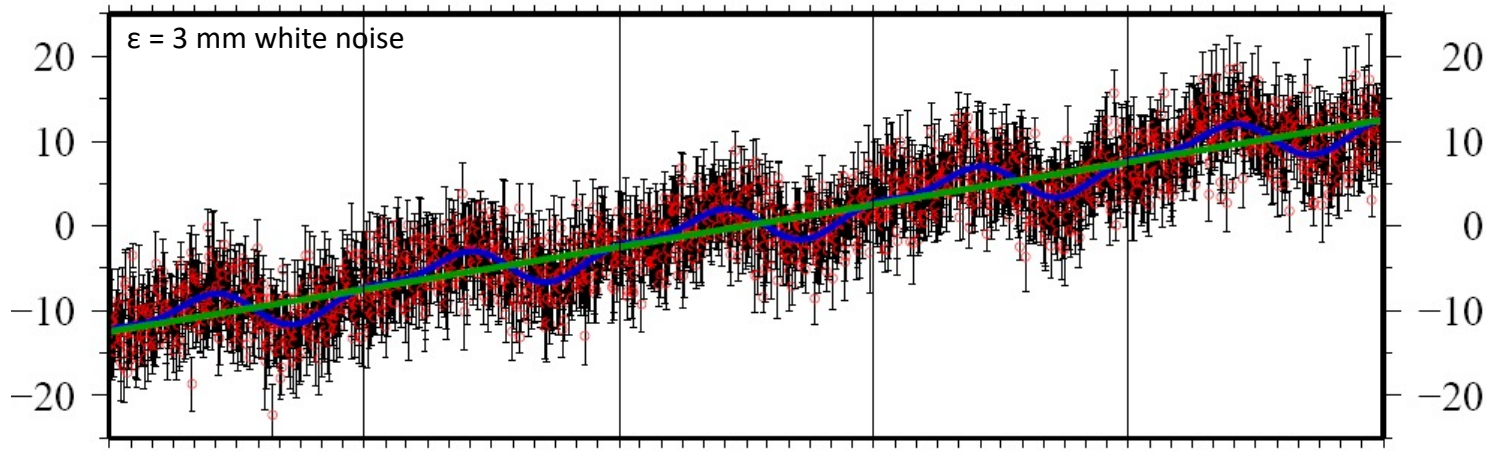
$$x^i = x_0^i + v^i(t - t_0) + A_0^i \cos\left(\frac{2\pi(t - t_0)}{T_0} - \tau_0\right) + A_1^i \cos\left(\frac{2\pi(t - t_0)}{T_1} - \tau_1\right) + \varepsilon$$

initial position

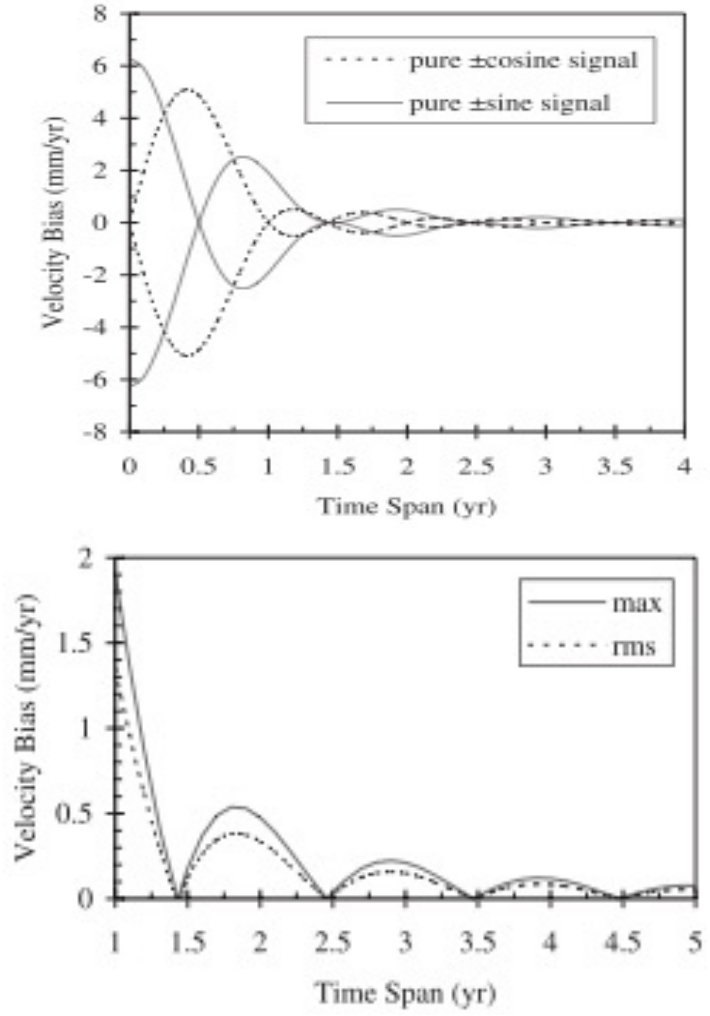
annual period sinusoid

semi-annual period sinusoid

seasonal term



# Velocity errors due to seasonal signals in continuous time series




- Theoretical analysis of a continuous time series by Blewitt and Lavallee (2002,2003)
- Top: Bias in velocity from a 1mm sinusoidal signal in-phase and with a 90-degree lag with respect to the start of the data span
- Bottom: Maximum and rms velocity bias over all phase angles
  - The minimum bias is NOT obtained with continuous data spanning an even number of years
  - The bias becomes small after 3.5 years of observation

# FOGMEx (“realistic sigma”) algorithm for velocity uncertainties

Motivation

- Computational efficiency
- Handle time series with varying lengths and data gaps
- Obtain a model that can be used in `globk`

Concept

- The departure from a white-noise ($\sqrt{N}$) reduction in noise with averaging provides a measure of correlated noise.

Implementation

- Fit the values of $\chi^2$ versus averaging time to the exponential function expected for a first-order Gauss-Markov (FOGM) process (amplitude, correlation time)
- Use the $\chi^2$ value for infinite averaging time predicted from this model to scale the white noise sigma estimates from the original (least-squares) fit
- and/or
- Fit the values to a FOGM with infinite averaging time (i.e., random walk) and use these estimates as input to `globk` (“mar_neu” command)

# `sh_cats/sh_hector`

- Scripts to aid batch processing of time series with CATS or Hector
- Requires CATS and/or Hector to be pre-installed
- Outputs
  - Velocities in ".vel"-file format
  - Equivalent random walk magnitudes in "mar_neu" commands for sourcing in `globk` command file
- Can take a long time!
- Reads GAMIT/GLOBK formats
  - .pos-file(s) as input
  - .eq-file(s) to define discontinuities for estimation of offsets
  - `tsfit` command file containing "eq_file", "max_sigma", "n_sigma" and/or "periodic" options instead of specifying as `sh_cats/sh_hector` options
- Writes files for GLOBK
  - .apr-file(s), including "EXTENDED" terms where periodic and/or non-linear (logarithmic and/or exponential decay) terms have been estimated
  - "mar_neu" commands for equivalent random walk process noise

# Summary

- `sh_glred` is post-processing equivalent to `sh_gamit`
- .pos-file format now standard GLOBK output for time series
- Visual inspection of time series very important for identifying outliers, bad segments of data or other problems like incompatible site IDs
  - `sh_plot_pos` (GMT) and tsview (MATLAB)
  - Populate .eq-file(s) with “rename” commands or use “sig_neu” commands to mitigate impact of poor or incompatible data points during velocity
  - Be aware that some “outliers” may be stabilization issues if they persist across a large part or all of a network at the same time, so check stabilization using “POS STATISTICS” lines in .org-file(s)
    - Numbers of stabilizing sites should be consistent and at least as many as the number of parameters estimated in `glorg` (e.g. three components of rotation and translation)
    - Numbers for wrms should be consistent from day-to-day and small (< 5 mm)
- Batch tools are available for longer, denser, continuous time series where point-by-point visual inspection is unreasonable
  - `tsfit` and tsview