# Time series and error analysis

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)

# Issues in GNSS error analysis

- What are the sources of the errors ?
- How much of the error can we remove by better modeling ?
- Do we have enough information to infer the uncertainties from the data ?
- What mathematical tools can we use to represent the errors and uncertainties ?

# Determining the uncertainties of GNSS parameter estimates

- Rigorous estimate of uncertainties requires full knowledge of the error spectrum, both temporal and spatial correlations (never possible)
- Sufficient approximations are often available by examining time series (phase and/or position) and reweighting data
- Whatever the assumed error model and tools used to implement it, external validation is important

# Tools for error analysis in GAMIT/GLOBK

GAMIT

- "AUTCLN reweight = Y" (default in sestbl.) uses phase rms from postfit edit to reweight data with constant + elevation-dependent terms

GLOBK

- Rename (eq_file) to "_XPS" or "_XCL" to remove outliers
- "sig_neu" adds white noise by station and span
    - Best way to "rescale" the random noise component
    - A large value can also substitute for "_XPS"/"_XCL" renames for removing outliers
- "mar_neu" adds random-walk noise
    - Principal method for controlling velocity uncertainties
- In the .gdl-files, rescale variances of an entire h-file
    - Useful when combining solutions from with different sampling rates or from different programs (Bernese, GIPSY)

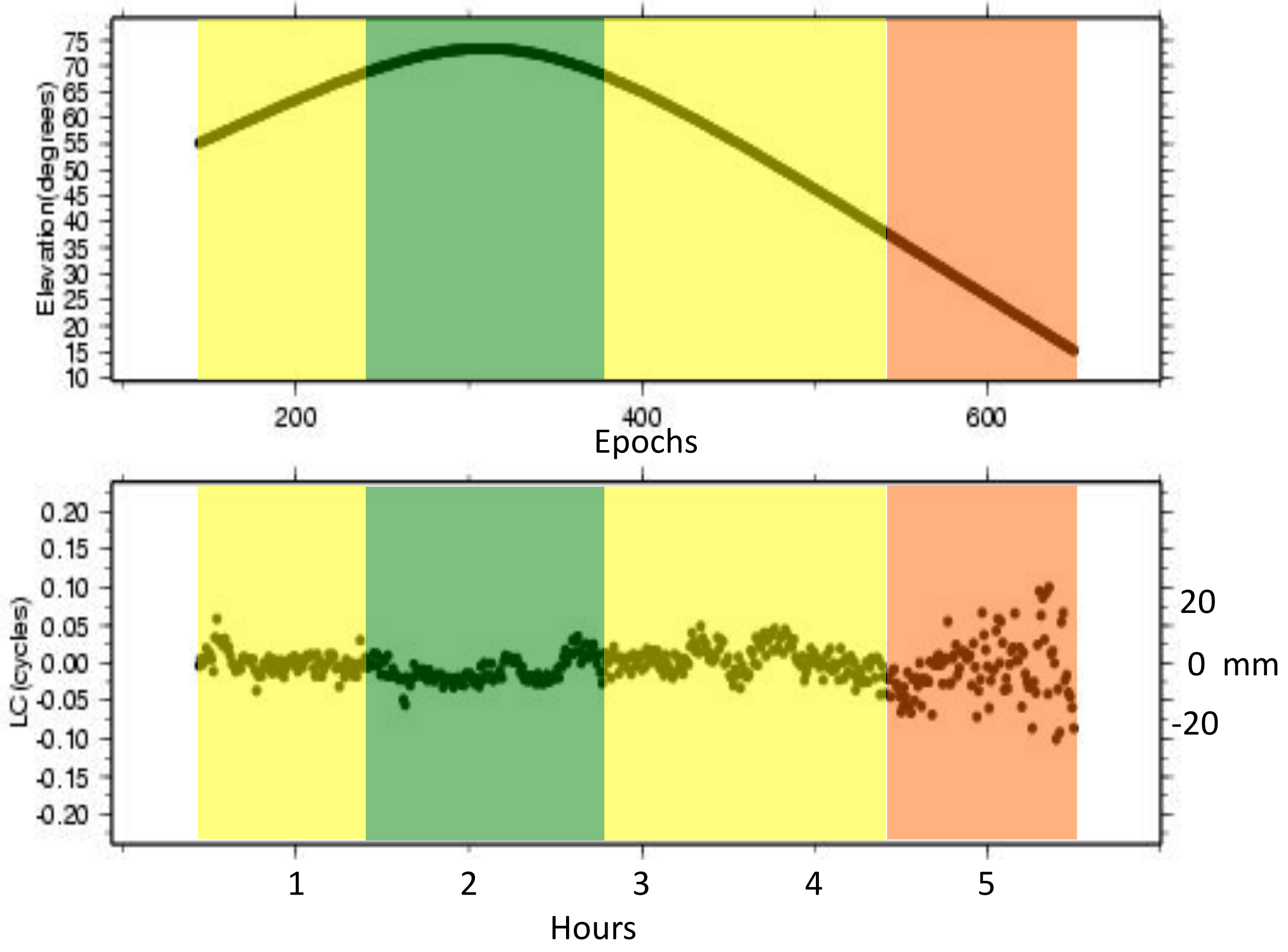
Utilities

- tsview and `tsfit` can generate "_XPS" commands graphically or automatically
- `grw` and `vrw` can generate "sig_neu" commands with a few key strokes
- FOGMEx ("realistic sigma") algorithm implemented in tsview (MATLAB) and `tsfit/ensum`
    - `sh_gen_stats` generates "mar_neu" commands for `globk` based on the noise estimates
- `sh_plotvel` (GMT) allows setting of confidence level of error ellipses
- `sh_tshist` and `sh_velhist` (GMT) can be used to generate histograms of time series and velocities

# Sources of error

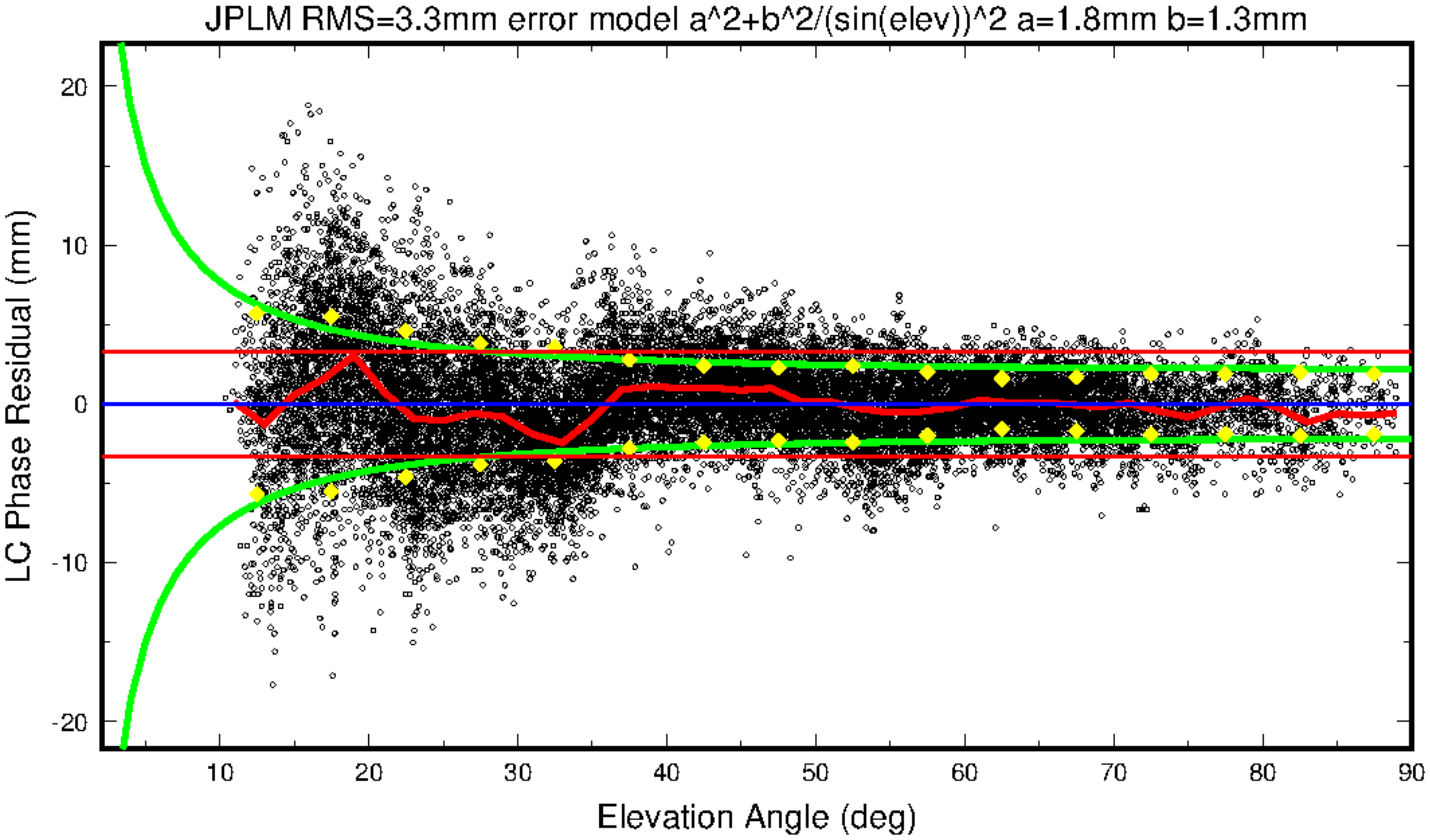
- Signal propagation effects
  - Receiver noise
  - Ionospheric effects
  - Signal scattering (antenna phase center / multipath)
  - Atmospheric delay (mainly water vapor)
- Unmodeled motions of the station
  - Monument instability
  - Loading of the crust by atmosphere, oceans, and surface water
- Unmodeled motions of the satellites

# Characterizing phase noise




Elevation angle and phase residuals for single satellite

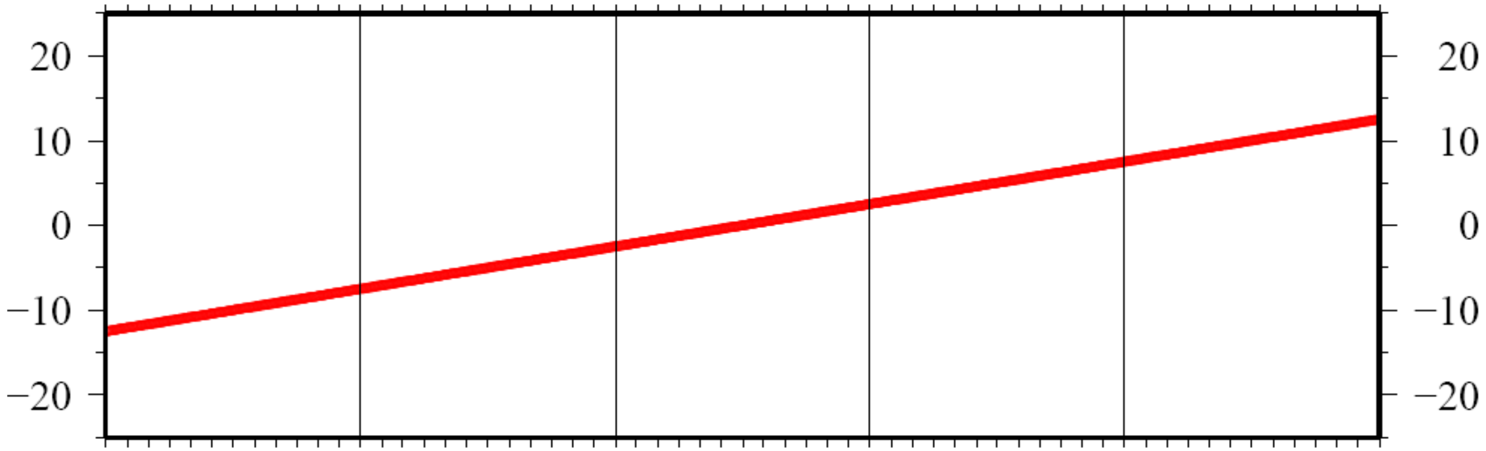
# Characterizing phase noise



# Time series characteristics

# Time series components

$$\underbrace{x^i}_{\text{observed position}} = \underbrace{x_0^i}_{\text{initial position}} + \overbrace{v^i(t - t_0)}^{\text{(linear) velocity term}}$$

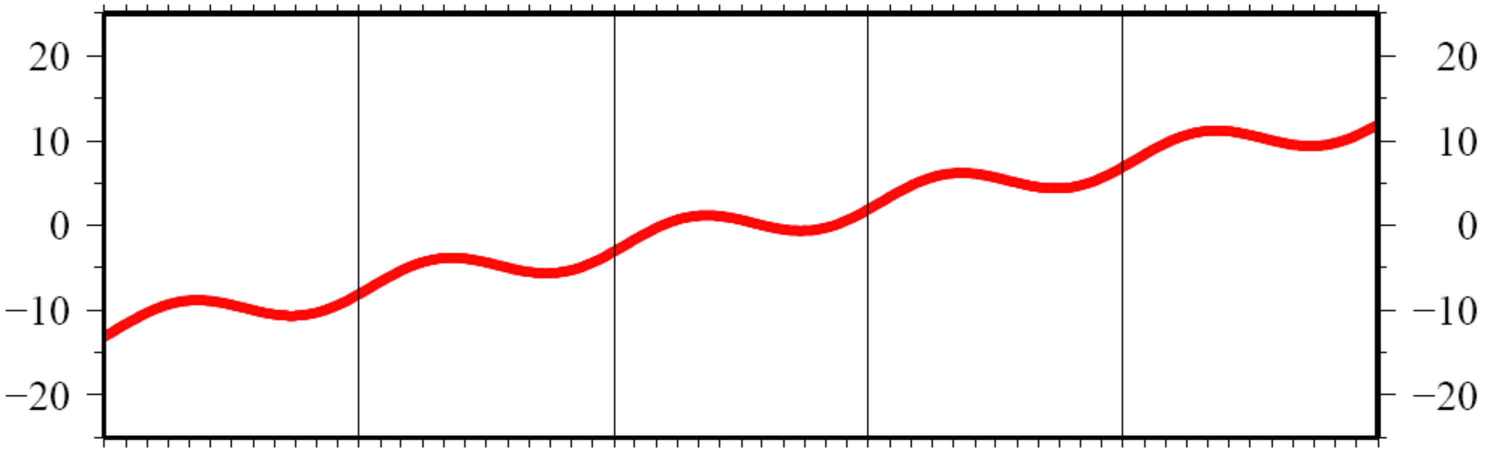


# Time series components

observed position

$$x^i = x_0^i + \overbrace{v^i(t - t_0)}^{\text{(linear) velocity term}} + \underbrace{A_0^i \cos\left(\frac{2\pi(t - t_0)}{T_0} - \tau_0\right)}_{\text{annual period sinusoid}}$$

initial position

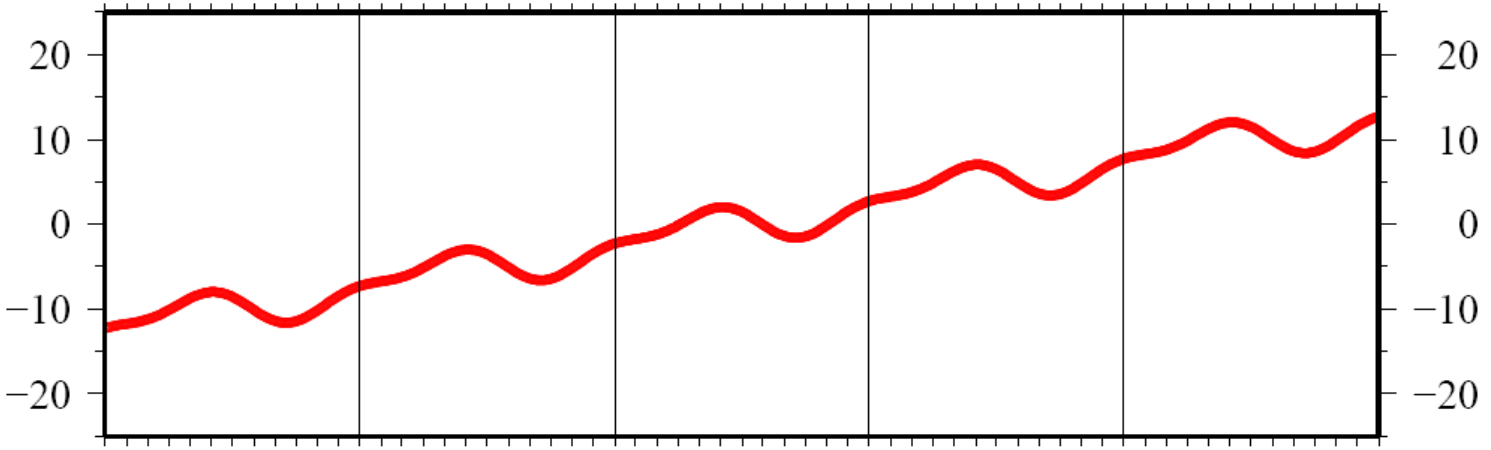


# Time series components

observed position

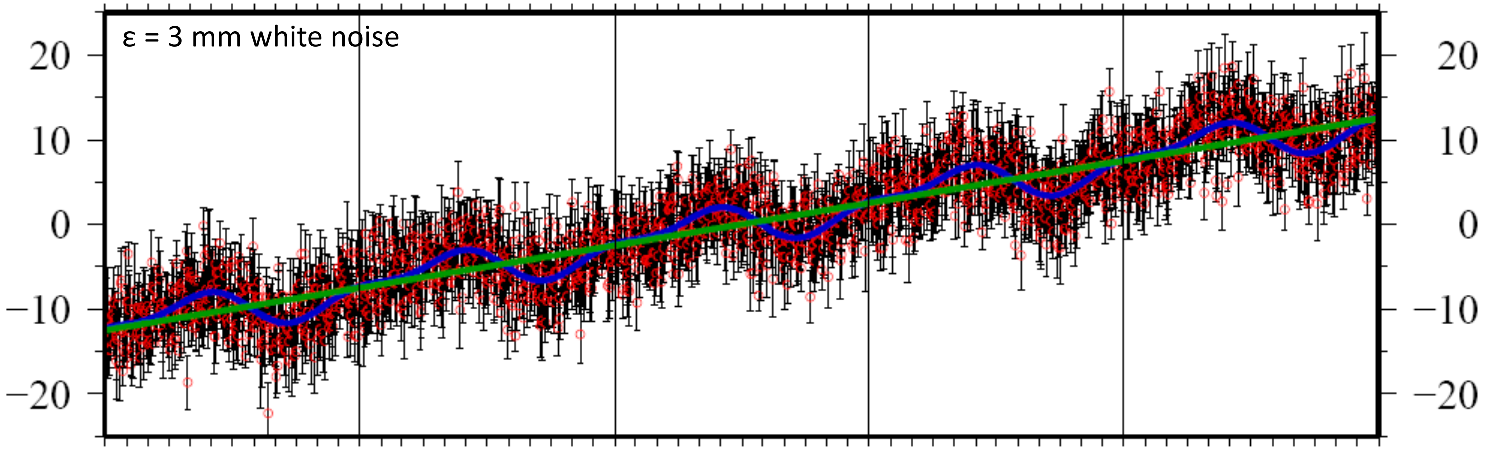
$$x^i = x_0^i + \overbrace{v^i(t-t_0)}^{\text{(linear) velocity term}} + \underbrace{\underbrace{A_0^i \cos\left(\frac{2\pi(t-t_0)}{T_0} - \tau_0\right)}_{\text{annual period sinusoid}} + \underbrace{A_1^i \cos\left(\frac{2\pi(t-t_0)}{T_1} - \tau_1\right)}_{\text{semi-annual period sinusoid}}}_{\text{seasonal term}}$$

initial position

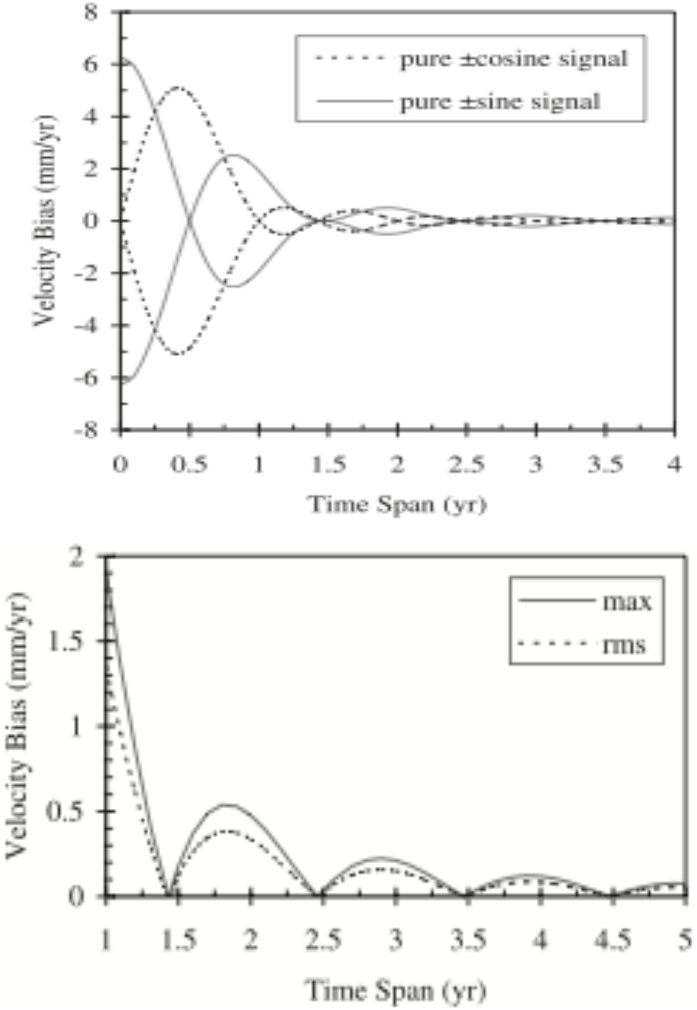


# Time series components

$$\overset{\substack{\text{observed}\\\text{position}\\\downarrow}}{x^i} = \underset{\substack{\uparrow\\\text{initial}\\\text{position}}}{x_0^i} + \overbrace{v^i(t-t_0)}^{\substack{\text{(linear)}\\\text{velocity term}}} + \underbrace{\underbrace{A_0^i \cos\left(\frac{2\pi(t-t_0)}{T_0} - \tau_0\right)}_{\substack{\text{annual period}\\\text{sinusoid}}} + \underbrace{A_1^i \cos\left(\frac{2\pi(t-t_0)}{T_1} - \tau_1\right)}_{\substack{\text{semi-annual}\\\text{period sinusoid}}}}_{\text{seasonal term}} + \varepsilon$$



# Velocity errors due to seasonal signals in continuous time series



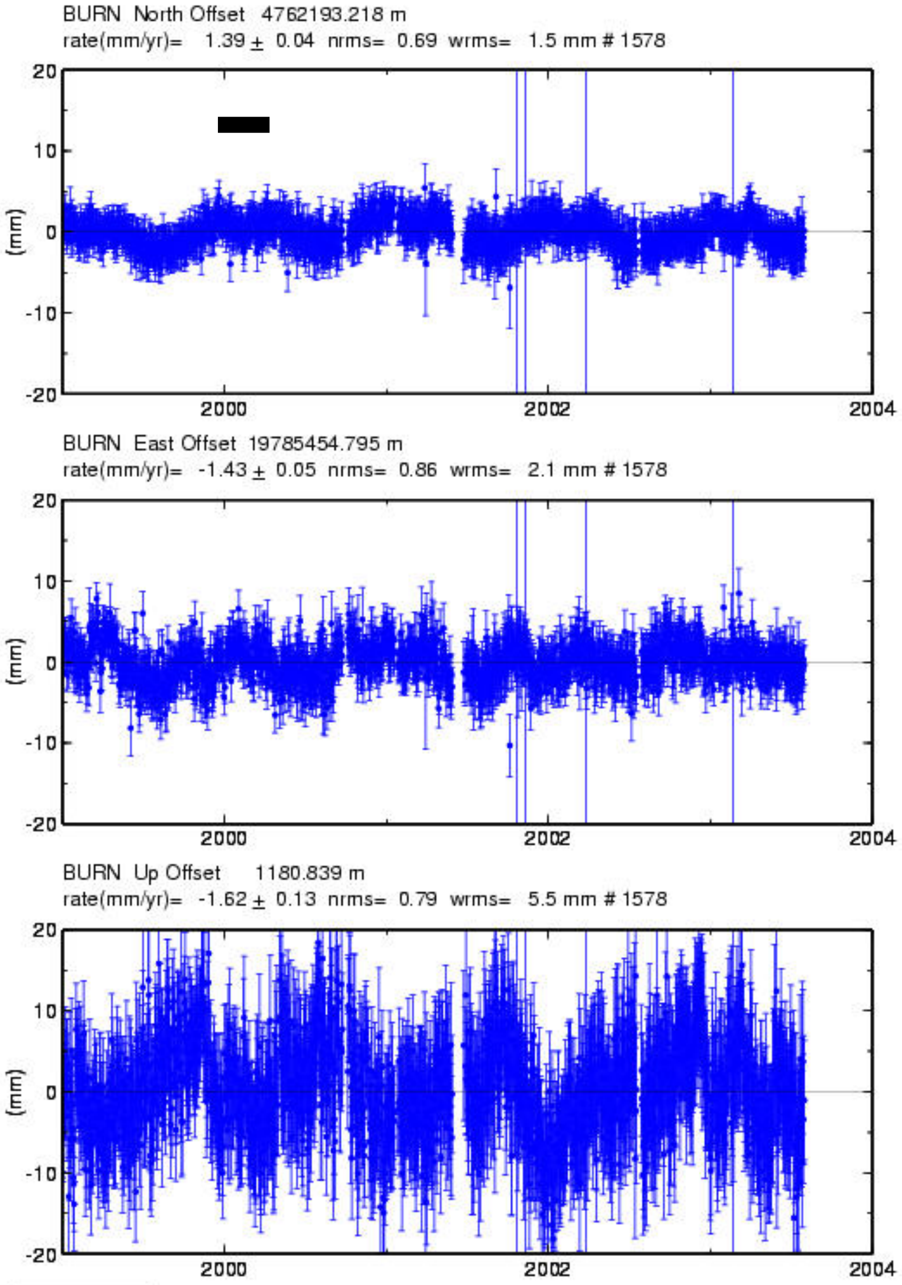

- Theoretical analysis of a continuous time series by Blewitt and Lavallee (2002,2003)

- 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

# Characterizing the noise in daily position estimates

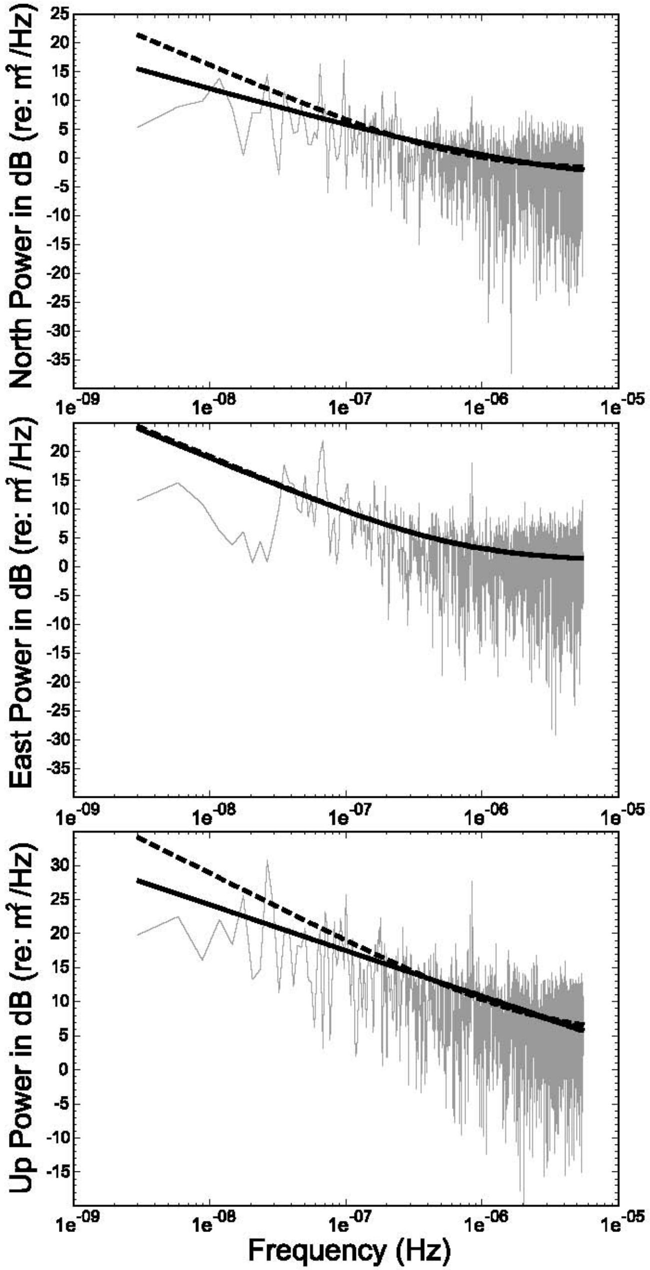
Note temporal correlations of 60-200 days and seasonal terms



# Spectral analysis of the time series to estimate an error model

Figure 5 from Williams et al. (2004): Power spectrum for common-mode error in the SOPAC regional SCIGN analysis. Lines are best-fit white noise plus flicker noise (solid = mean amplitude; dashed = maximum likelihood estimation)

Note lack of taper and misfit for periods > 1 yr (frequencies < $\pi \times 10^{-8}$)

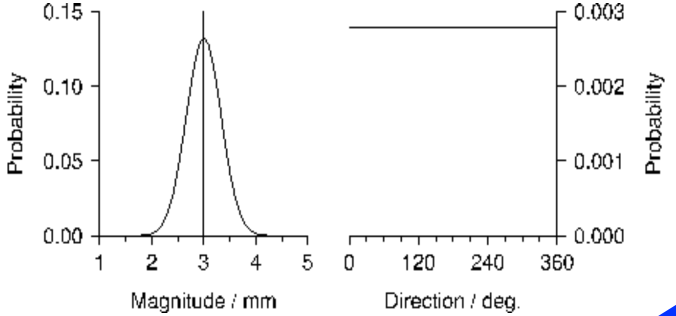


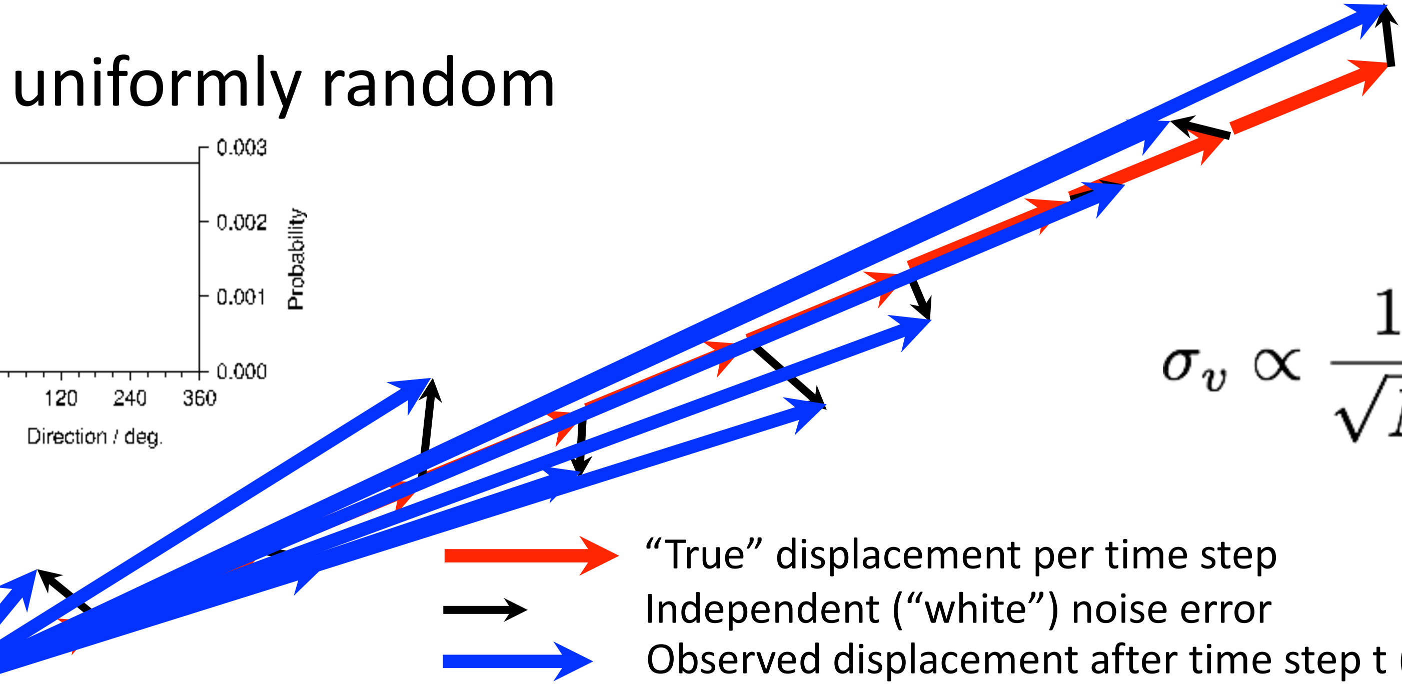
# Summary of spectral analysis approach

- Power law: slope of line fit to spectrum
  - 0 = white noise
  - −1 = flicker noise
  - −2 = random walk
- Non-integer spectral index (e.g. “fraction white noise” → $1 > k > -1$ )
- Good discussion in Williams (2003)
- Problems:
  - Computationally intensive
  - No model captures reliably the lowest-frequency part of the spectrum

# “White” noise

- Time-independent (uncorrelated)
- Magnitude has continuous probability function, e.g. Gaussian distribution
- Direction is uniformly random



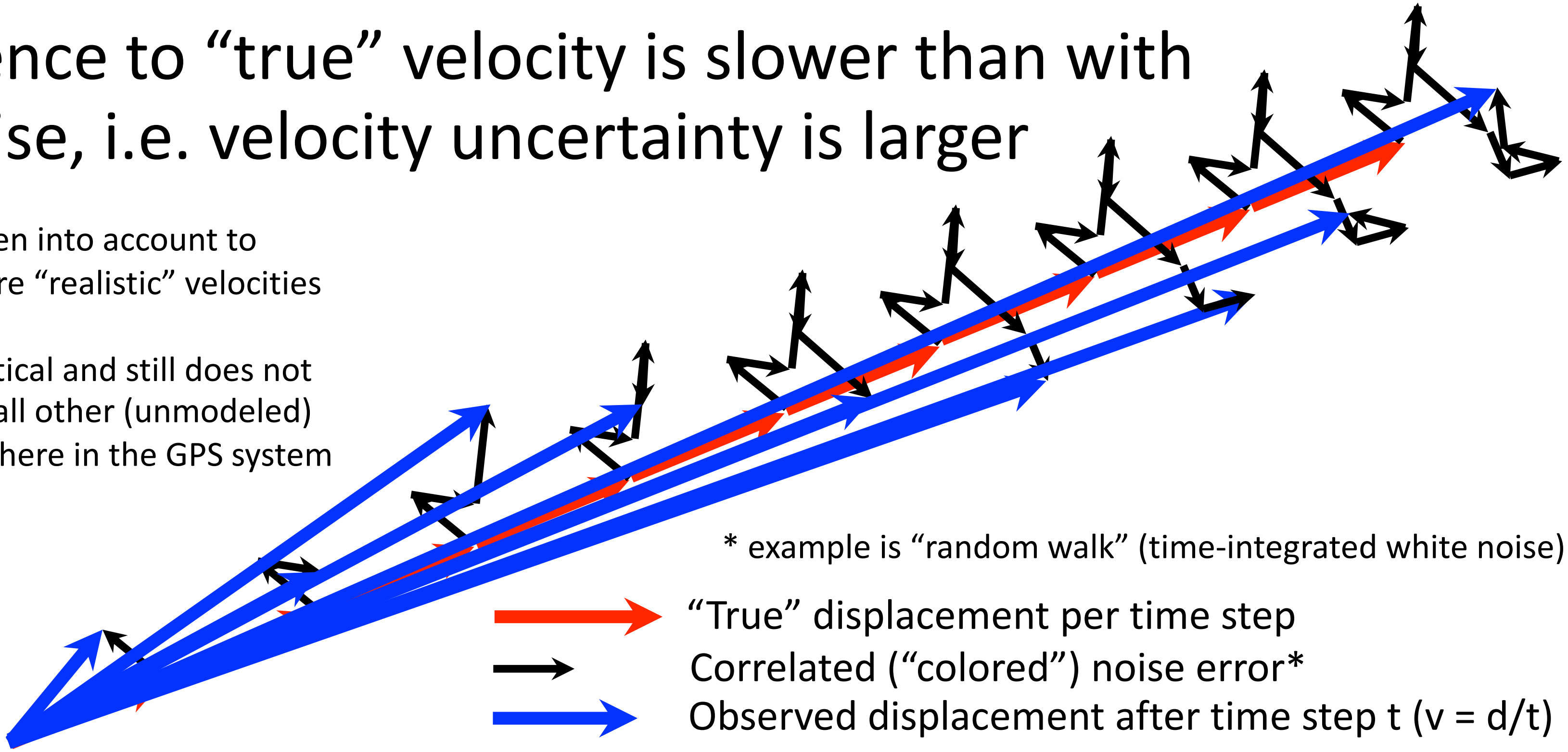

$$\sigma_v \propto \frac{1}{\sqrt{N}}$$

# “Color” noise

- Time-dependent (correlated): power-law, first-order Gauss-Markov, etc.
- Convergence to “true” velocity is slower than with white noise, i.e. velocity uncertainty is larger

Must be taken into account to produce more “realistic” velocities

This is statistical and still does not account for all other (unmodeled) errors elsewhere in the GPS system



# CATS (Williams, 2008)

- Create and Analyze Time Series
- Maximum likelihood estimator for chosen model solves for
  - Initial position and velocity
  - Seasonal cycles (sum of periodic terms) [optional]
  - Exponent of power law noise model
- Requires some linear algebra libraries (BLAS and LAPACK) to be installed on computer (common nowadays, but check!)
  - Information on M. Floyd’s experience of compiling CATS at http://web.mit.edu/mfloyd/www/computing/cats/
- Formerly at http://www.pol.ac.uk/home/staff/?user=WillSimoCats
  - However, above web page and source code no longer seem to available
  - Possibly a sign that CATS is superseded by Hector

# Hector (Bos et al., 2013)

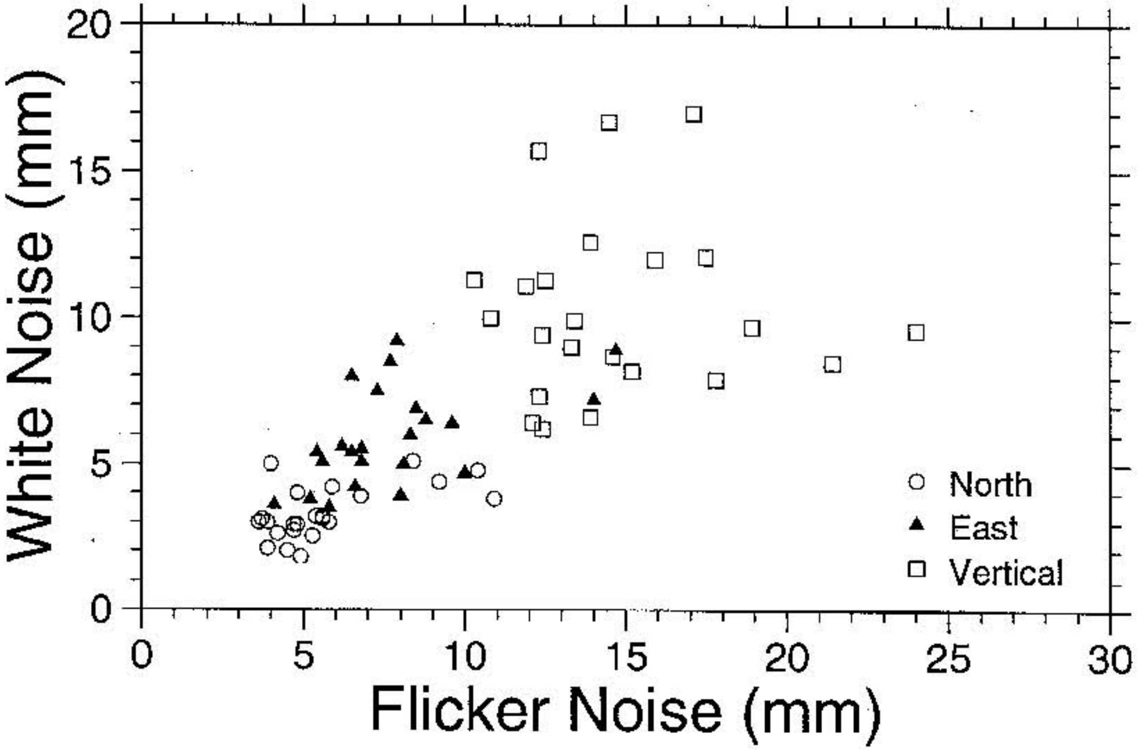
- Much the same as CATS but faster algorithm
- Maximum likelihood estimator for chosen model solves for
  - Initial position and velocity
  - Seasonal cycles (sum of periodic terms) [optional]
  - Exponent of power law noise model
  - Also, as of Hector version 1.6:
  - Changes in linear velocity
  - Non-linear motions (logarithmic and/or exponential decays)
- Requires ATLAS linear algebra libraries to be installed on computer
- Pre-complilled executables available, but tricky to install from source due to ATLAS requirement
- http://segal.ubi.pt/hector/

# 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

# Approximations (Mao et al., 1999)

Use white noise statistics (wrms) to predict the flicker noise




White noise vs flicker noise from Mao et al. (1999) spectral analysis of 23 global stations

# 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

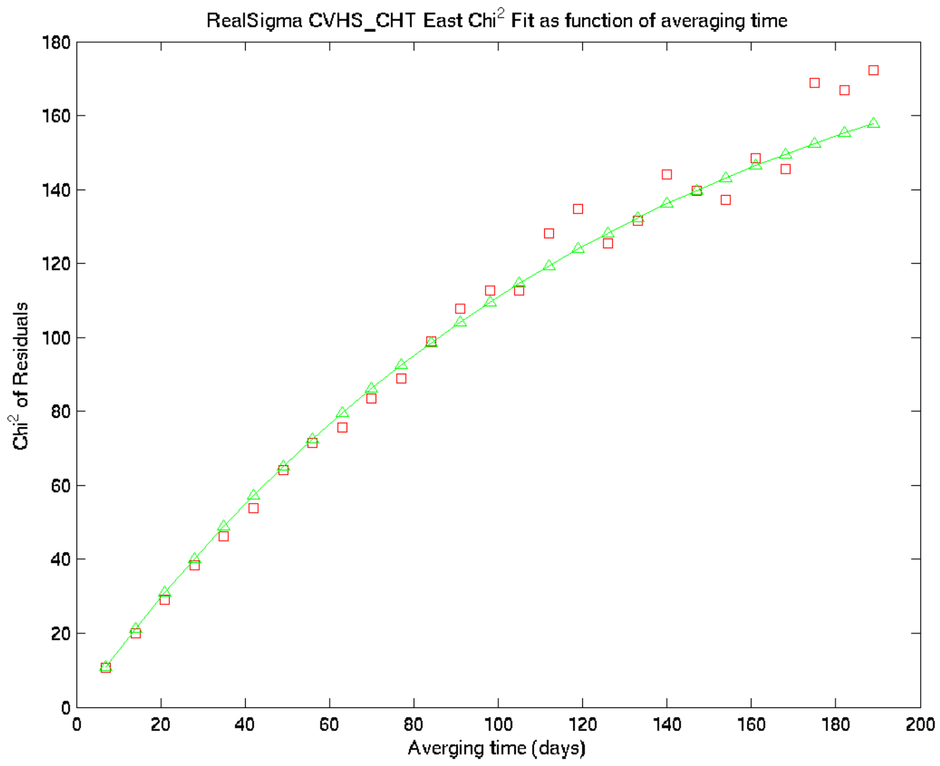
- The departure from a white-noise ($\sqrt{N}$) reduction in noise with averaging provides a measure of correlated noise.

Implementation

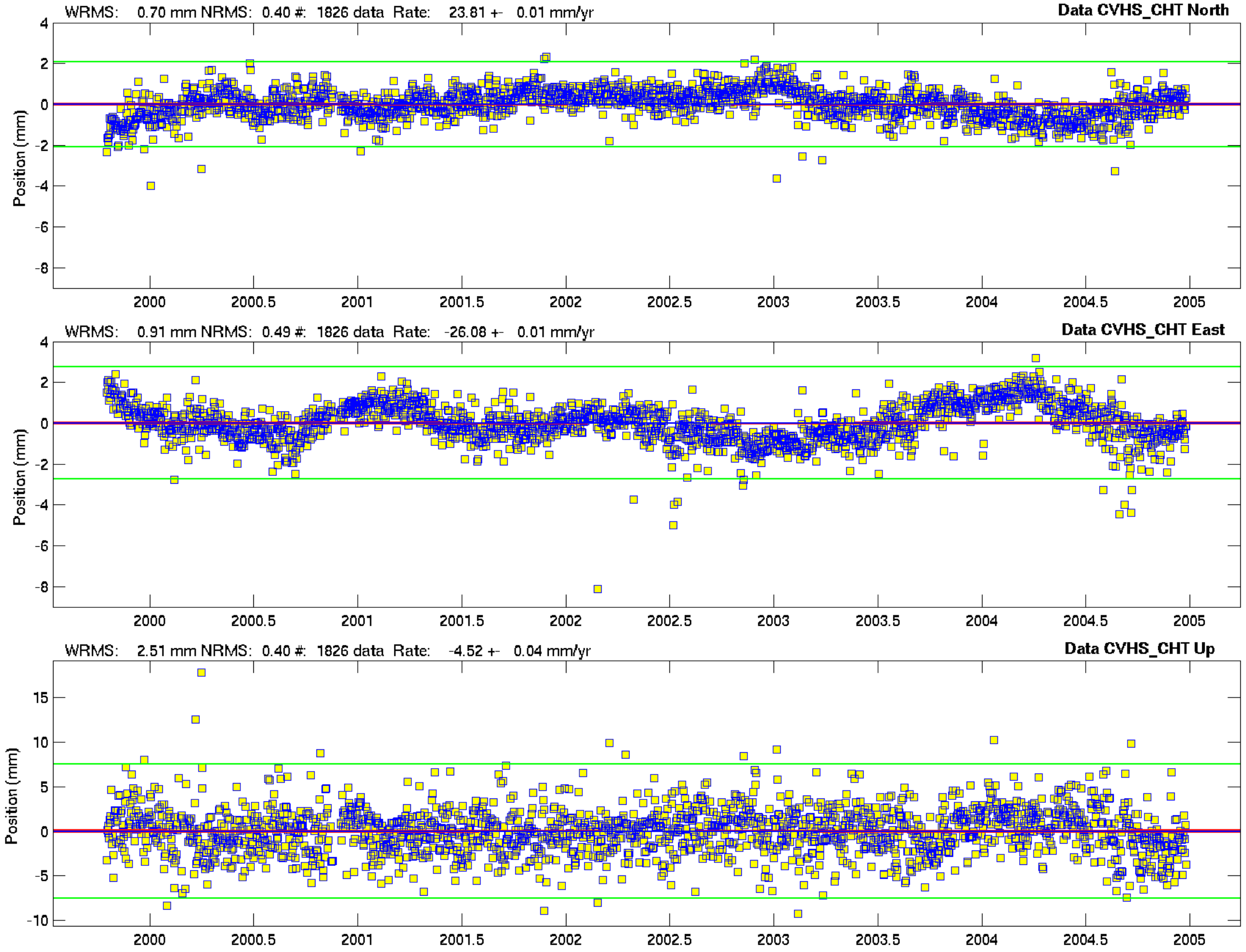
- Fit the values of $\chi^2$ versus averaging time to the exponential function expected for a first-order Gauss-Markov (FOGM) process (amplitude, correlation time)
- Use the $\chi^2$ value for infinite averaging time predicted from this model to scale the white noise sigma estimates from the original (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)

# Extrapolated variance (FOGMEx)

- For independent noise, variance ∝ $1/\sqrt{N_{\text{data}}}$
- For temporally correlated noise, variance (or $\chi^2$/d.o.f.) of data increases with increasing window size
- Extrapolation to "infinite time" can be achieved by fitting an asymptotic function to RMS as a function of time window
  - $\chi^2$/d.o.f. ∝ $e^{-\sigma\tau}$
- Asymptotic value is good estimate of long-term variance factor
- Use "real_sigma" option in `tsfit`

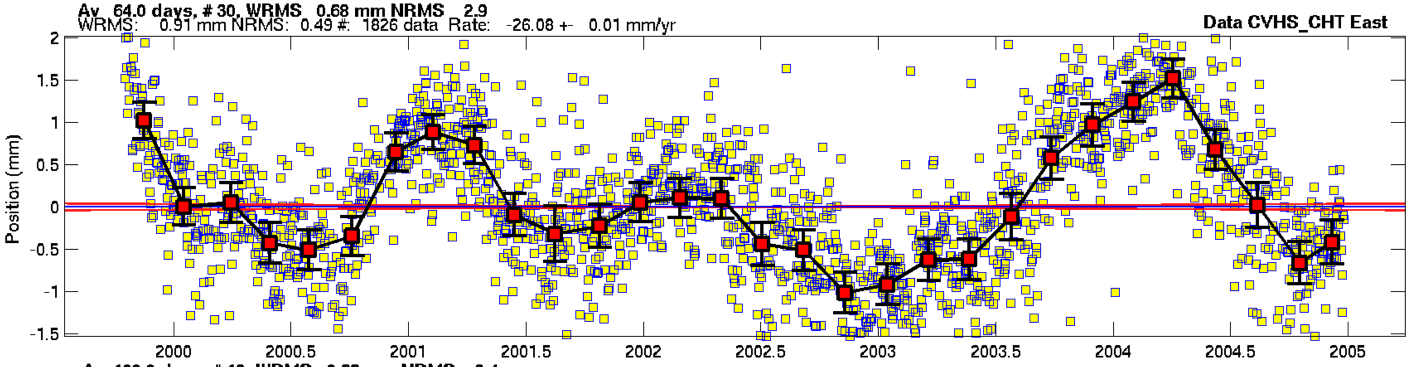


# Understanding the FOGMEx algorithm: Effect of averaging on time-series noise

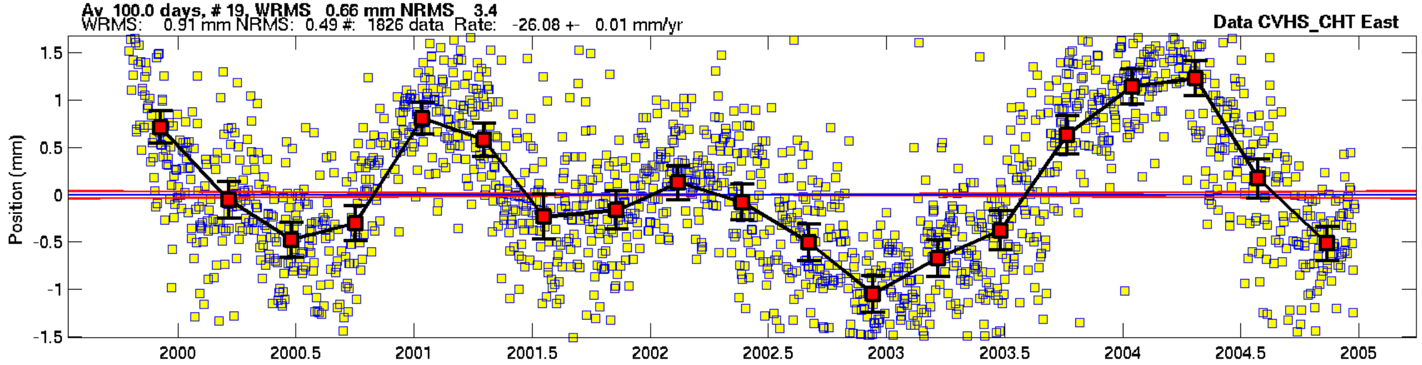


# Same site, east component (daily wrms 0.9 mm, nrms 0.5)

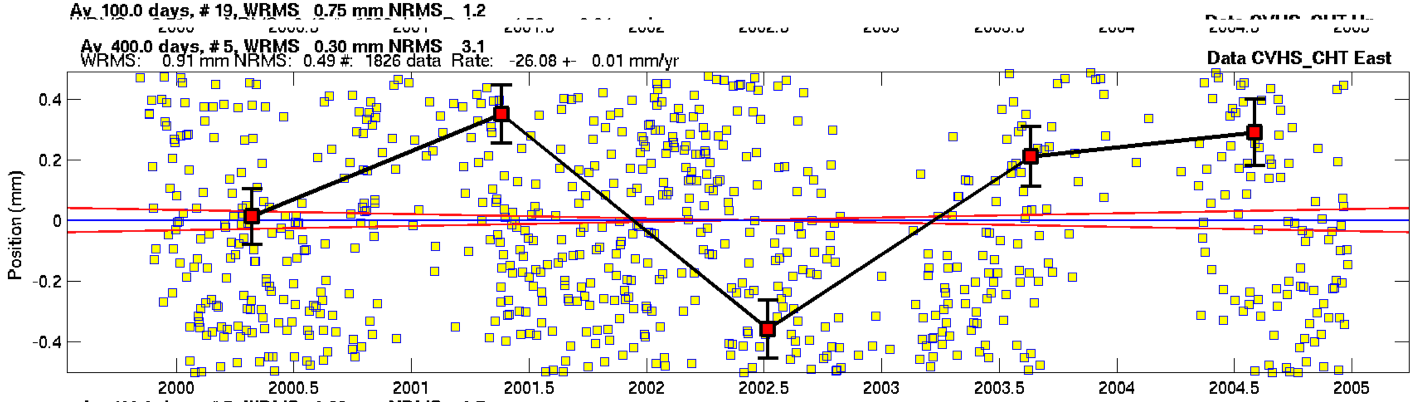
64-d avg
wrms 0.7 mm
nrms 2.0




100-d avg
wrms 0.6 mm
nrms 3.4




400-d avg
wrms 0.3 mm
nrms 3.1



# Using tsview to compute and display the FOGMEx (“realistic-sigma”) results

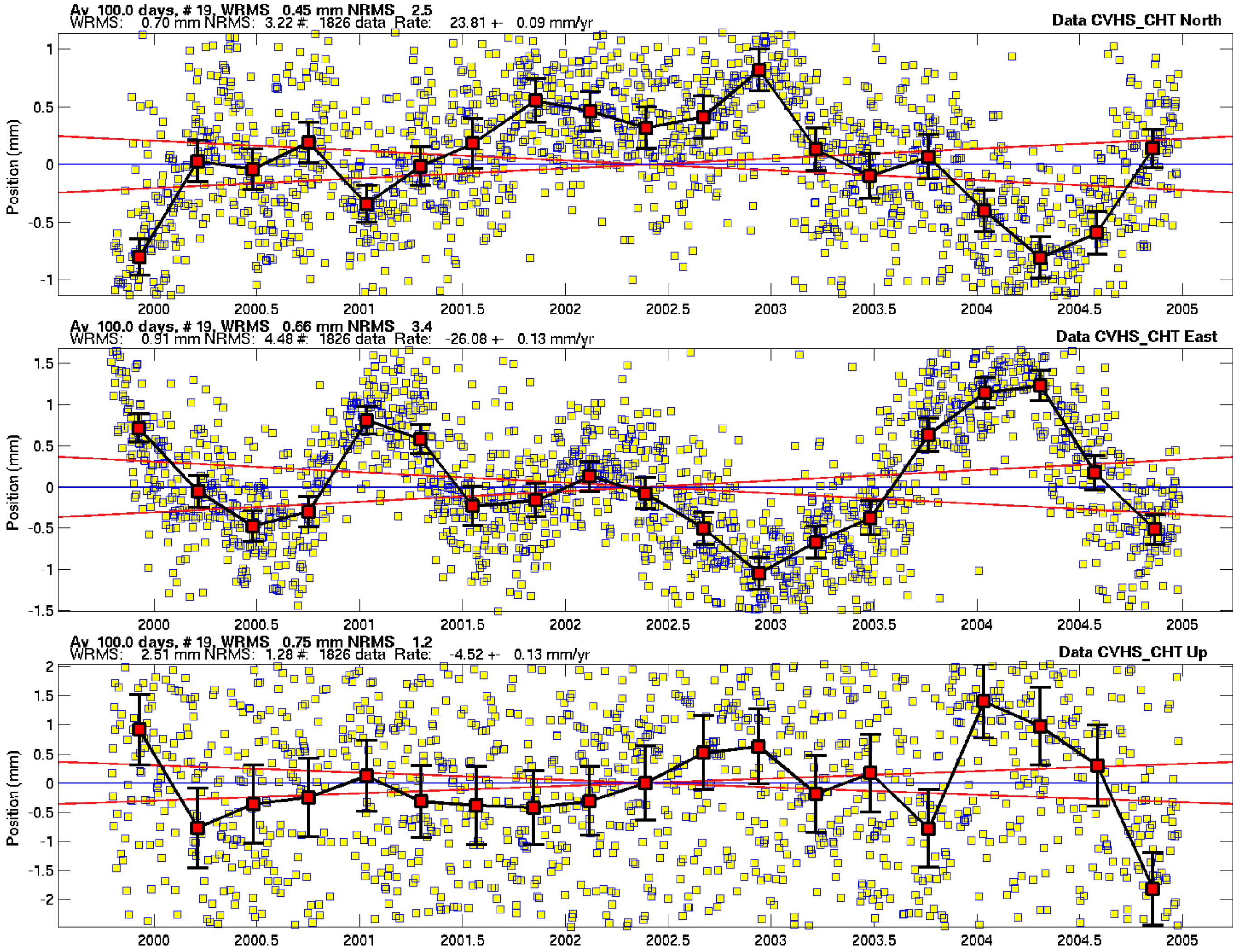
Note rate uncertainties with the “realistic-sigma” algorithm :

0.09  mm/yr N

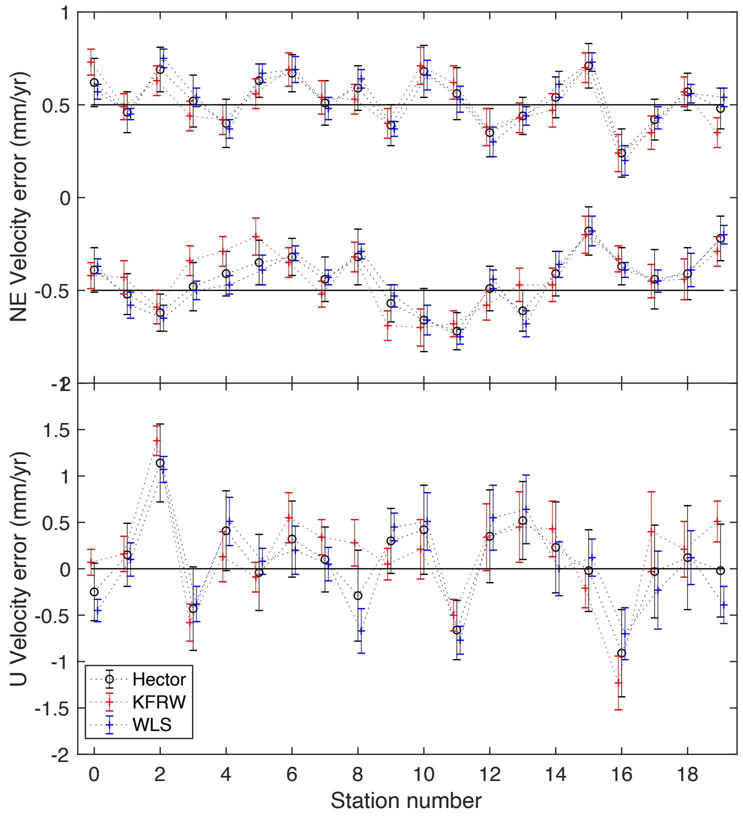
0.13 mm/yr E

0.13 mm/yr U

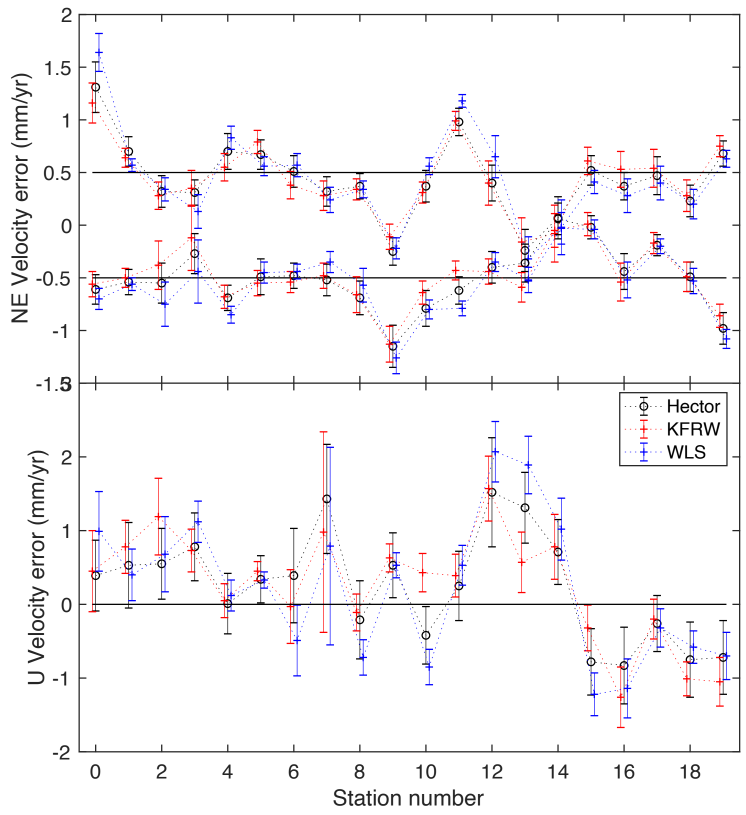
Red lines show the 68% probability bounds of the velocity based on the results of applying the algorithm.



# Comparison of estimated velocity uncertainties using spectral analysis (Hector) and Gauss-Markov fitting of averages (FOGMEx)




Floyd and Herring (2020), Figure 7




Floyd and Herring (2020), Figure 8

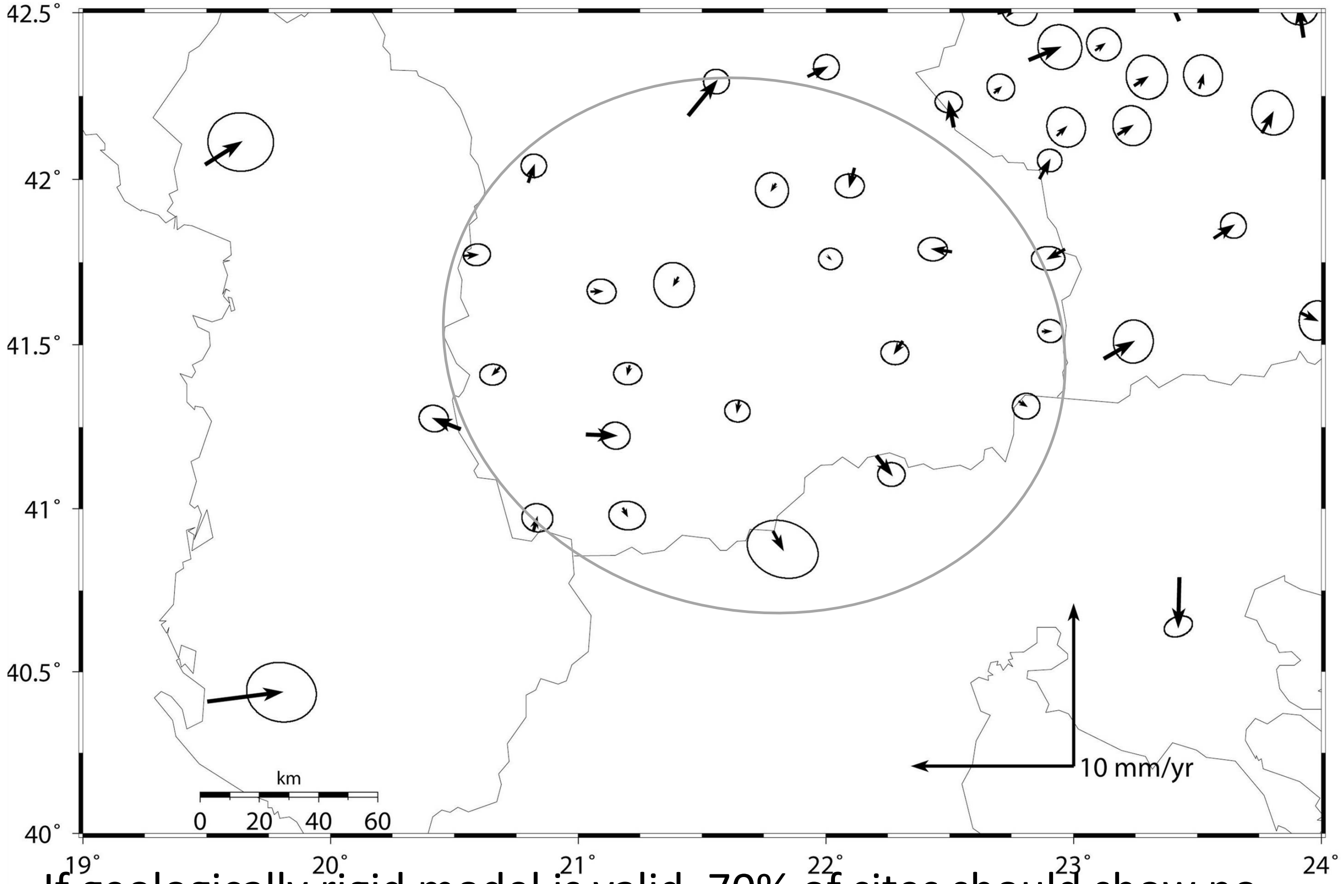
# Summary of practical approaches

- White noise + flicker noise (+ random walk) to model the spectrum (Williams et al., 2004)
- White noise as a proxy for flicker noise (Mao et al., 1999)
- Random walk to model to model an exponential spectrum (Herring "FOGMEx" algorithm for velocities)
- "Eyeball" white noise + random walk for non-continuous data
- All approaches require common sense and verification

# External validation of velocity uncertainties by comparing with a geophysical model

Simple case: assume no strain within a geologically rigid region

GMT plot at 70% confidence

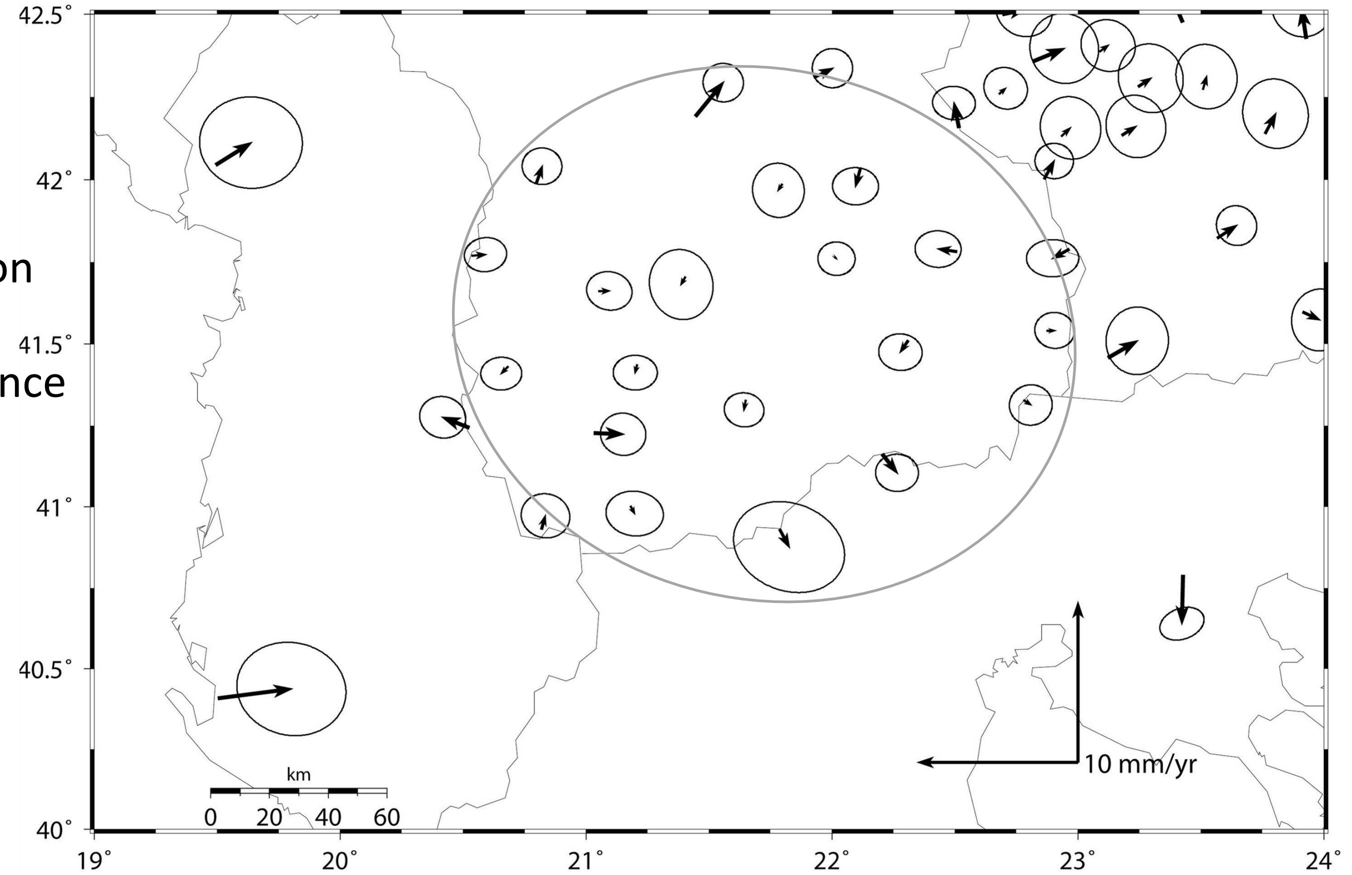



17 sites in central Macedonia: 4–5 velocities pierce error ellipses

If geologically rigid model is valid, 70% of sites should show no statistically significant motion, i.e. velocity lies within error ellipse

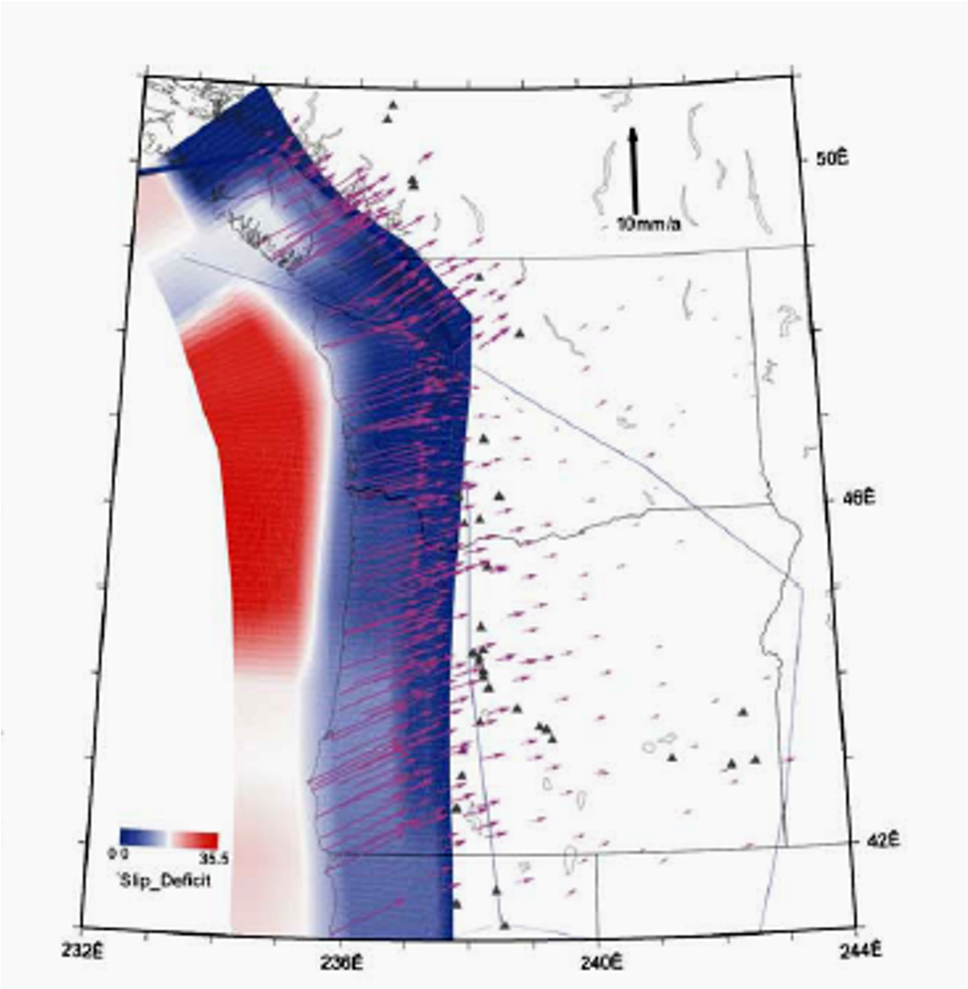
# External validation of velocity uncertainties by comparing with a geophysical model

Same solution plotted with 95% confidence ellipses



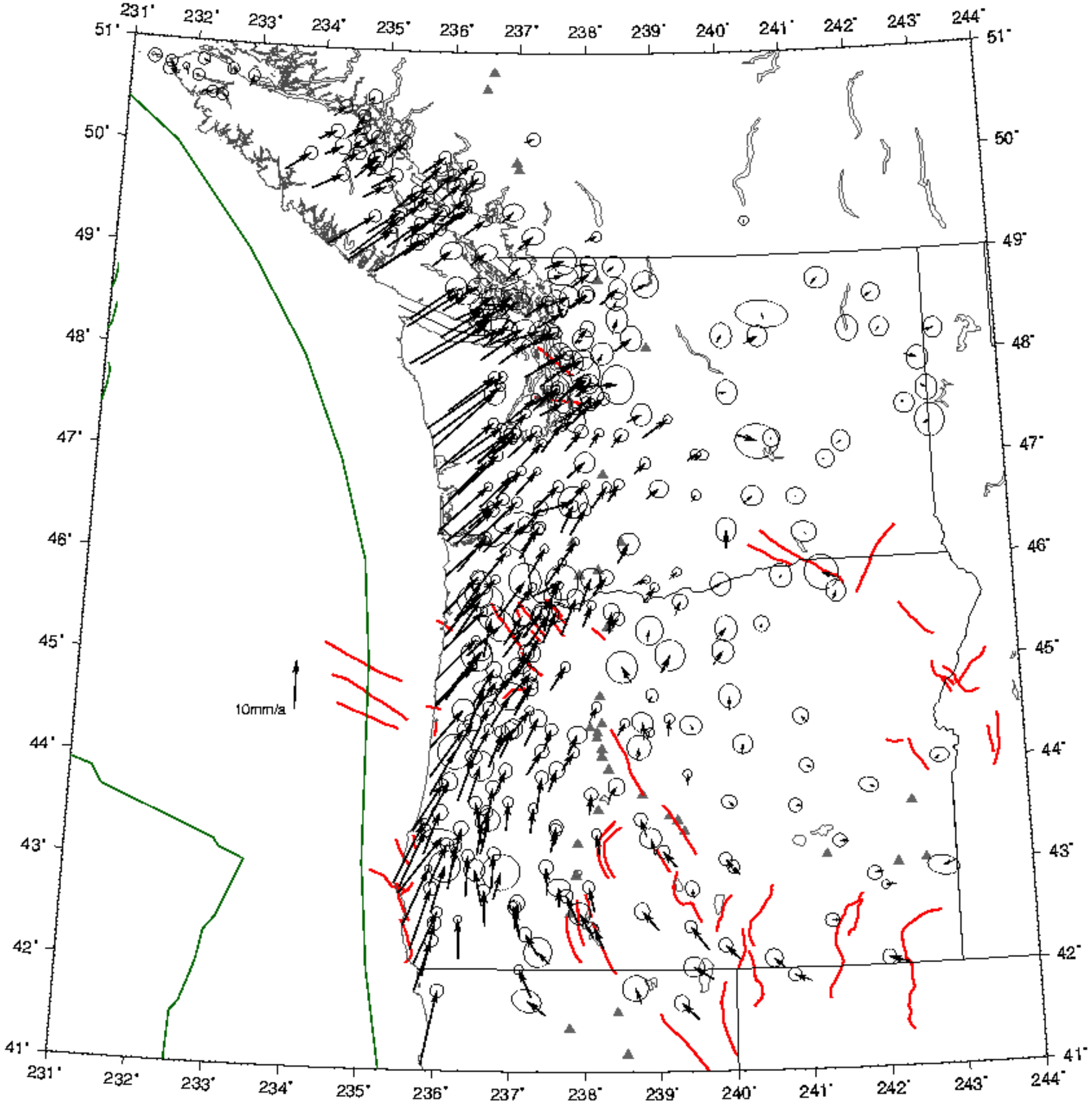

Now 1–2 of 17 velocities pierce error ellipses

# External validation of velocity uncertainties by comparing with a geophysical model



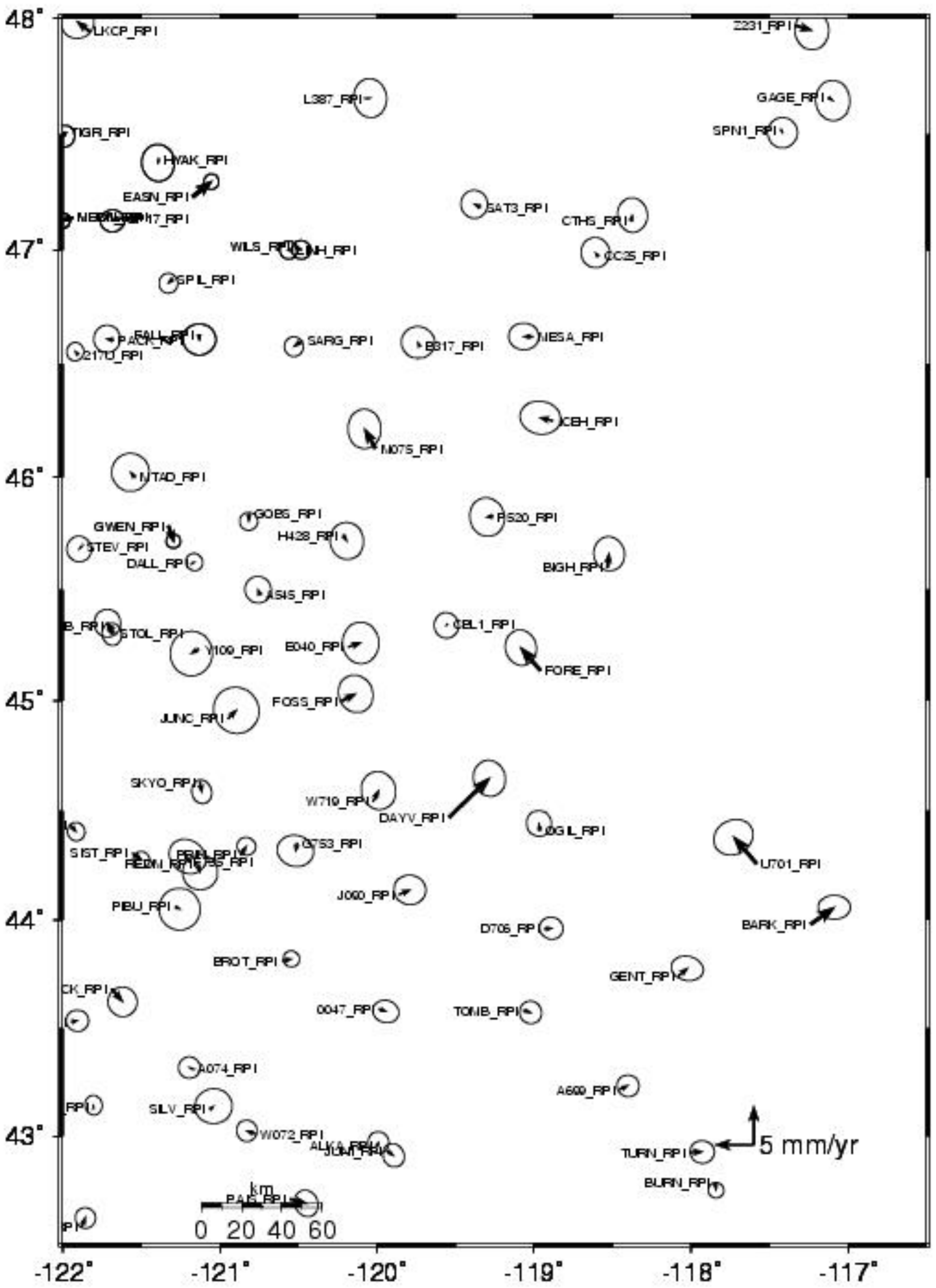

A more complex case of a large network in the Cascadia subduction zone

Colors show slipping and locked portions of the subducting slab where the surface velocities are highly sensitive to the model; area to the east is slowly deforming and insensitive to the details of the model




Velocities and 70% error ellipses for 300 sites observed by continuous and survey-mode GPS 1991-2004

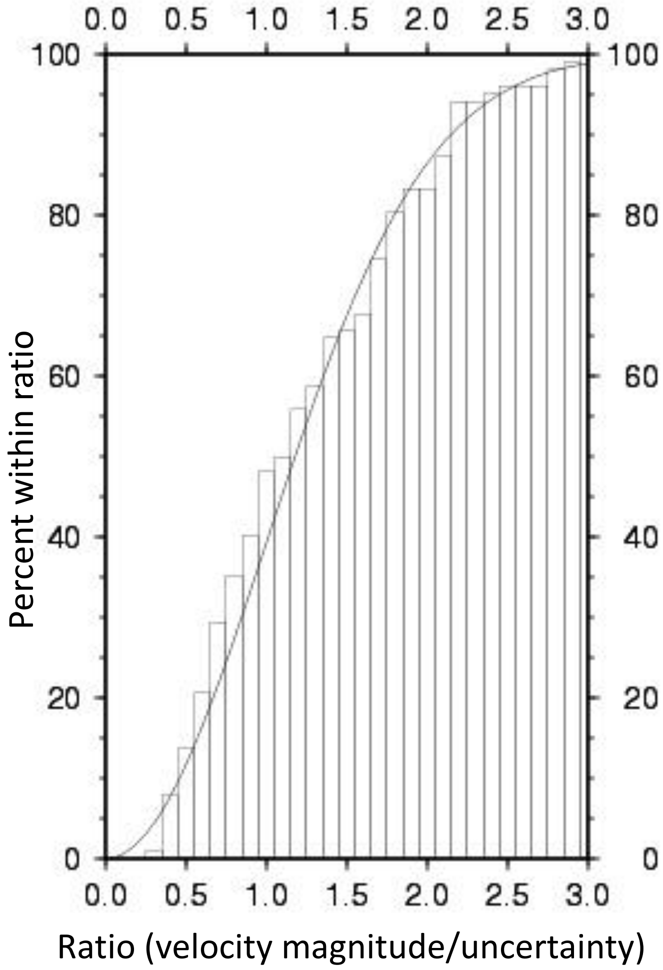
Validation area (next slide) is east of 238°E




Residuals to elastic block model for 73 sites in slowly deforming region

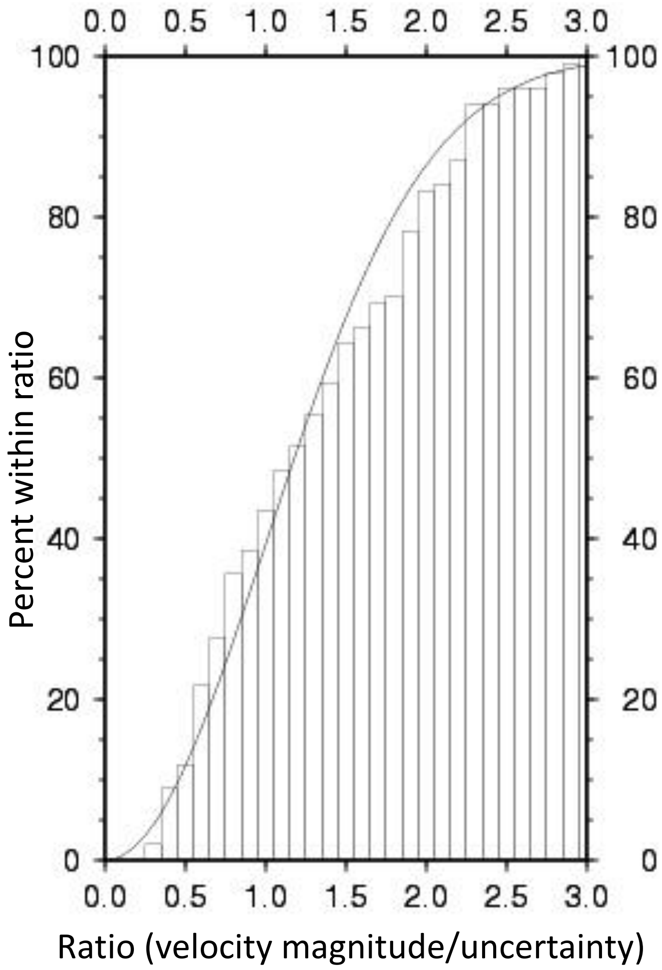
Error ellipses are for 70% confidence:
13-17 velocities pierce their ellipse

# Statistics of velocity residuals



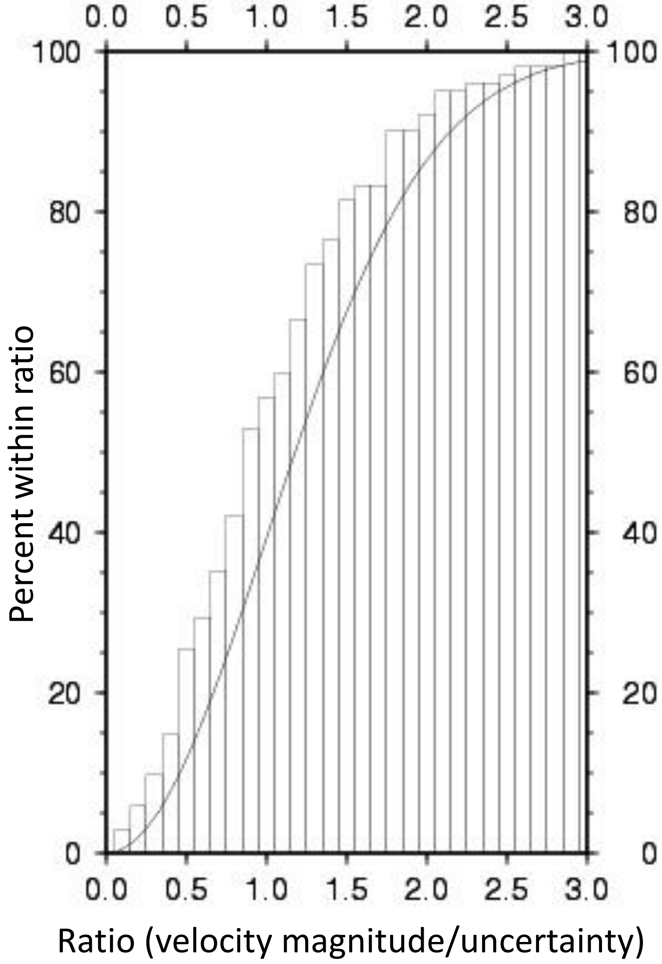

- Cumulative histogram of normalized velocity residuals for eastern Oregon and Washington
  - 70 sites
- Noise added to position for each survey:
  - 0.5 mm random (“sig_neu”)
  - 1.0 mm/sqrt(yr) random walk (“mar_neu”)
- Solid line is theoretical for a χ-distribution

# Statistics of velocity residuals




- Same as last slide but with a smaller random-walk noise added:
  - 0.5 mm random
  - 0.5 mm/yr random walk
  - cf. 1.0 mm/sqrt(yr) RW for "best" noise model
- Note greater number of residuals in range of 1.5–2.0 sigma

# Statistics of velocity residuals




- Same as last slide but with larger random and random-walk noise added :
  - 2.0  mm white noise
  - 1.5 mm/sqrt(yr) random walk
  - cf. 0.5 mm WN and 1.0 mm/sqrt(yr) RW for “best” noise model
- Note smaller number of residuals in all ranges above 0.1-sigma

# Summary

- All algorithms for computing estimates of standard deviations have various problems
  - Fundamentally, rate standard deviations are dependent on low frequency part of noise spectrum, which is poorly determined without very long time series (decades)
- Assumptions of stationarity (constant noise characteristics over time) are often (usually?) not valid
- FOGMEx ("realistic sigma") algorithm is a convenient and reliable approach to getting velocity uncertainties in `globk`
  - We are testing how reliable, in comparison to other methods, given good and bad time series
- Velocity residuals from a physical model, together with their uncertainties, can be used to validate the error model