Annual Report to Southern California Earthquake Center:

Simulating SCIGN: Idealized, Realistic Models for Testing Detection Methods for Crustal Deformation Data

3/01/2009 — 2/29/2010

Duncan Carr Agnew (dagnew@ucsd.edu)

Institute of Geophysics and Planetary Physics
Scripps Institution of Oceanography
University of California, San Diego
La Jolla, CA 92093-0225

1. Introduction

A major motivation for measuring crustal deformation has always been the search for possible temporal fluctuations; this is, for example, one of the scientific goals of the Plate Boundary Observatory (PBO). Such fluctuations are of intrinsic interest, and there has always been the possibility that, if detected, they might serve as possible signs of higher periods of earthquake risk. But many claimed changes have been close to the noise levels of the methods available, with better data showing that the changes actually were just noise.

The high precision of GPS has made it possible to detect phenomena never seen before: postseismic deformations have been observed well enough to support inferences about crustal rheology; and transient deformations have been observed at many subduction zones. At the same time, the low cost of GPS has greatly expanded the amount of data: in southern California alone about 450,000 daily displacement numbers are produced each year.

In light of this, the SCEC geodesy group has promoted the development of methods to examine these data for possible changes in deformation rate. Such methods should detect and localize any temporal variations; in the absence of any detections, they could bound temporal variations in strain accumulation.

In August 2008 SCEC supported a small workshop on this problem, which concluded that SCEC geodesy could usefully follow a procedure used by other elements of SCEC: encouraging systematic comparisons between different methods. Such comparisons, when applied to codes for modeling seismic rupture, have clarified discrepancies and resulted in much more reliable modeling codes. The workshop also concluded that such a comparison should begin on synthetic data sets, as similar as possible to real data, but with the noise and any signal known (at least to the person producing the ‘‘data’’), and that the synthetic data should be created by someone not pursuing the detection problem, and provided as a blind test.

I received funding to provide such synthetic datasets; as described below, this was done by developing a software package (`FAKENET`) for simulating the random behavior of geodetic data and combining this with the motions expected from different sources. An early version of this package was used to produce datasets used in the Phase I tests in early 2009, and a later version, described below, produced data for the Phase II tests (summer 2009 and ongoing).

2. Data Models

The basic data model is that the time series at the i-th site, on the j-th day is

$$\mathbf{d}_{ij} = \mathbf{v}_i(t_j - t_0) + \mathbf{c}_{ij} + \mathbf{n}_{ij} + \sum_{n=1}^{N} \mathbf{G}_{in} \cdot \mathbf{s}_{nj}$$

where $\mathbf{d}_{ij}$ is the vector displacement, decomposed into (1) a station velocity $\mathbf{v}_i$; (2) a noise part $\mathbf{c}_{ij}$ which is common to more than one station; (3) a second noise part $\mathbf{n}_{ij}$ peculiar to that station; and (4) a signal part, computed as a sum over slipping fault patches, where $\mathbf{s}_{nj}$ is the (vector) slip on the n-th patch. $\mathbf{G}_{in}$ is the Green function relating slip to displacement: for this we use the usual expressions for a dislocation in a half-space.

We then need statistical models for the noise **c** and **n**; ‘‘statistical’’ meaning that the mathematical model uses random variables, and the simulations use pseudorandom deviates with the same probability behavior. Developing such models for GPS time series has been recognized as important since Johnson and Agnew (1995) pointed out the importance of random-walk motion in estimating uncertainties for station velocities. Subsequent work (Langbein and Johnson 1997, Zhang *et al.* 1997, Wdowinski *et al.* 1997, Mao *et al.* 1999, Williams 2003, Williams *et al.* 2004, Beavan 2005, Langbein 2008, Bos *et al.* 2008, Williams 2008) has shown that it is useful to decompose the noise into several elements.

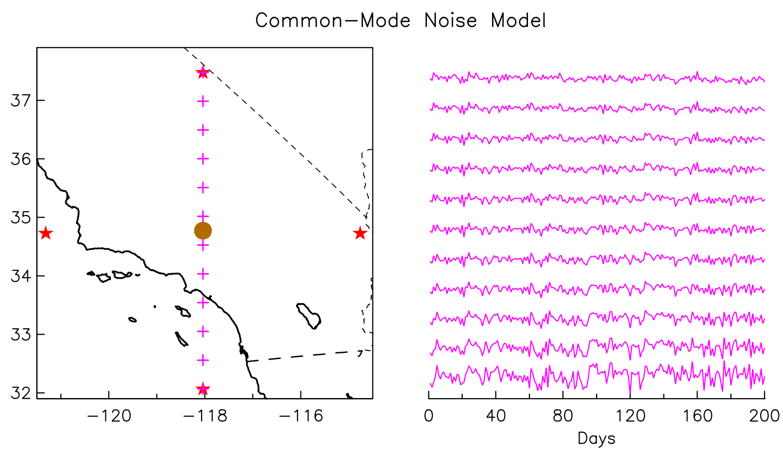

Figure 1

First of all, there is substantial spatial correlation in the displacements between nearby sites, though this diminishes with station separation. We can represent this correlated noise as a blend of several time series:

$$\mathbf{c} = \sum_{k=1}^{K} W_k(\theta, \phi)\mathbf{c}_k(t)$$

where the W’s are K functions of position, designed to sum to 1.0 at all locations but have maxima in different places; each is associated with different a vector noise function of time, $\mathbf{c}_k(t)$. **Figure 1** shows how this works: five series **c** are generated at the points shown by the brown circle and red stars, and the time series at any point are weighted sums of these. The right-hand side shows how this noise varies along a NS profile, mimicking the behavior seen in actual GPS data. The actual form of **c** usually needs to include both random and seasonal signals; seasonal signals also are present at many of the individual stations, and need to be included in the individual series **n**.

At most GPS stations the station noise **n** is temporally correlated out to very long times, which is what statisticians call a long-memory process, and is also represented by power spectral densities $P(f)$ that rise with decreasing frequency. Most analyses suggest that combinations of white $P \propto f^0$, flicker $P \propto f^{-1}$, and random-walk $P \propto f^{-2}$ noises are a good first approximation to the data, though better fits are obtained with general power-law noises ($P \propto f^{-\alpha}$).

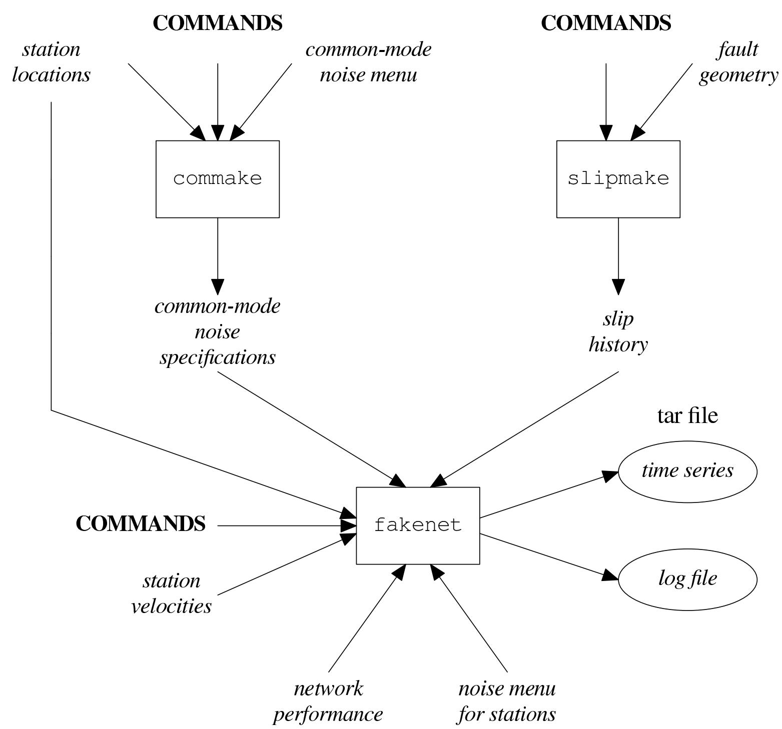

Figure 2

3. Program Package

In its current form, the `FAKENET` package consists of Fortran-77 programs, a detailed user manual, and data files. Each program reads in a set of commands which cause each program to read from input files, and write to one or more output files. The displacement timeseries produced include (1) velocities, randomly perturbed by an amount set by the user; (2) gaps, as specified by an input file; (3) randomly-assigned noise levels for each time series, with noise types including white, flicker, and random-walk; (4) common-mode noise (including sinusoidal variations) that decorrelates with increasing distance between stations; and (5) signals from slip on faults, including propagating slip, with time constants and time variations set by the user.

The package produces a set of files, each containing the timeseries for a particular site; these are combined into a single compressed tarfile, which, like the individual files, is assigned a unique name depending on when the package was run. The program also generates a *log file*, which documents the actual settings of the inputs, so that the input files can be recreated and the program rerun; this log file serves as an ‘‘audit trail’’ for what was done.

For ease of use (in Southern California, at any rate), the current distribution includes, from the SCEC Community Fault Model, the rectilinear approximations to faults, adjusted to make the top and bottom edges horizontal. The distribution also includes files of station coordinates, network performance, and station velocities, all derived from the SOPAC time series distribution and metadata; and also a file of noise parameters (provided by John Langbein) from which samples are randomly drawn to set the station noise parameters.

Figure 2 shows how simulated time series are created. The main program (`fakenet`) reads from several *input files*; these describe the station locations and velocities, the slip on faults, the parameters for common-mode noise, and so on. The program also reads instructions given in a **command file**, which contains multiple lines of commands with parameters. Some of the input files for `fakenet` are generated by other programs: `slipmake` produces the file giving fault slip time histories, and `commake` creates the file for common-mode noise; making these programs separate has simplified development and debugging. These programs also read both command and input files. For `commake`, the inputs are files of station locations and noise parameters; for `slipmake`, a file of fault geometry.

A simulation program needs a ‘‘good’’ pseudorandom number generator; this package uses the ‘‘Mersenne twister’’ of Matsumoto and Nishimura (1998), which passes all the tests of L’Ecuyer and Simard (2007). Noise models now implemented include random-walk, flicker, and white noise, using algorithms from Thomas *et al.* (2007), and R. Voss (in Gardner 1978).

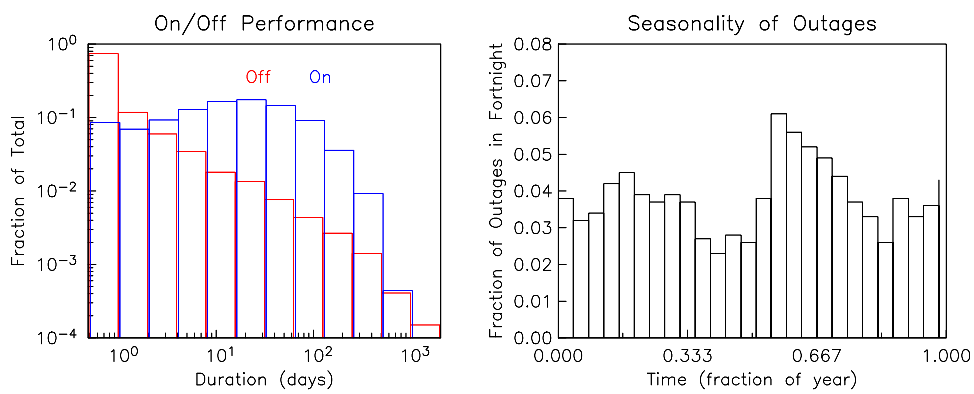

Figure 3

The existing package uses a fixed set of on-off times, taken from the SCIGN analysis to represent data gaps, and the start and end of a particular station. **Figure 3** shows several features of the SCIGN timeseries. The left plot in **Figure 3** is a histogram of durations of intervals of data being available, or not; the ‘‘on’’ times are roughly hyperbolically distributed up to a cutoff time, a much heavier-tailed distribution than would be expected from a Poisson process (fortunately); the durations of ‘‘off’’ periods are much more tightly clustered around small values. The right-hand side of **Figure 3** shows that outages in southern California GPS stations also have a definite variation with time of year.

As noted above, if the random components are set to zero, `FAKENET` can be used for conventional forward modeling. By providing standard noise models, it should also aid in the development of analysis tools for GPS data; to circle back to the start of this report, if we are going to look for signals we had better understand the noise.

The `FAKENET` package is freely available.

References

Beavan, J. (2005). Noise properties of continuous GPS data from concrete pillar geodetic monuments in New Zealand and comparison with data from U.S. deep drilled braced monuments, *J. Geophys. Res.*, **110**, 1-13, doi:doi:10.1029/2005JB003642

Bos, M. S., R. M. S. Fernandes, S. D. P. Williams, and L. Bastos (2008). Fast error analysis of continuous GPS observations, *J. Geodesy*, **82**, 157-166, doi:10.1007/s00190-007-0165-x

Gardner, M. (1978). Mathematical games: white and brown music, fractal curves and 1/f fluctuations, *Sci. Amer.*, **239(4)**, 16-32.

Johnson, H., and D. Agnew (1995). Monument motion and measurements of crustal velocities, *Geophys. Res. Lett.*, **22**, 2905-2908.

L'Ecuyer, P., and R. Simard (2007). TestU01: A C library for empirical testing of random number generators, *ACM Trans. Math. Softw.*, **33**, 22, doi:10.1145/1268776.1268777

Langbein, J. (2008). Noise in GPS displacement measurements from Southern California and Southern Nevada, *J. Geophys. Res.*, **113**, B05405, doi:10.1029/2007JB005247.

Langbein, J. O., and H. Johnson (1997). Correlated errors in geodetic time series: Implications for time-dependent deformation, *J. Geophys. Res.*, **102**, 591-603.

Mao, A. L., C. G. A. Harrison, and T. H. Dixon (1999). Noise in GPS coordinate time series, *J. Geophys. Res.*, **104**, 2797-2816.

Matsumoto, M., and T. Nishimura (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Trans. Modeling Comp. Simul.*, **8**, 3-30 .

Thomas, D. B., W. Luk, P. H. W. Leong, and J. D. Villasenor (2007). Gaussian random number generators, *ACM Computing Surv.*, **39**, 11.

Wdowinski, S., Y. Bock, J. Zhang, and P. Fang (1997). Southern California permanent GPS geodetic array: spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 Landers earthquake, *J. Geophys. Res.*, **102**, 18057-18070.

Williams, S. D. P. (2003). The effect of coloured noise on the uncertainties of rates estimated from geodetic time series, *J. Geod.*, **76**, 483-494.

Williams, S. D. P., Y. Bock, P. Fang, P. Jamason, R. M. Nikolaidis, L. Prawirodirdjo, M. Miller, and D. J. Johnson (2004). Error analysis of continuous GPS position time series, *J. Geophys. Res.*, **109**, B03412, doi:10.1029/2003JB002741

Williams, S. D. P. (2008). CATS: GPS coordinate time series analysis software, *GPS Solutions*, **12**, 147-153, doi:10.1007/s10291-007-0086-4

Zhang, J., Y. Bock, H. Johnson, P. Fang, J. Genrich, S. Williams, S. Wdowinski, and J. Behr (1997). Southern California Permanent GPS Geodetic Array: Error analysis of daily position estimates and site velocities, *J. Geophys. Res.*, **102**, 18035-18055.