Ge167 Homework 3
Due in class, Thursday, May 7

The goal of this homework is to get used to the character of typical time series from continuous GPS observations. You will learn more about where these kinds of time series come from in the lectures from Susan Owen. These particular time series are from 4 CGPS sites in southern California downloaded as processed time series from `www.sopac.ucsd.edu` a couple years ago. More recently processed time series may have improved a bit. These sites were all installed prior to the October 16, 1999 M_w 7.1 Hector Mine, CA earthquake. For each time series, estimate a constant offset, a secular (constant velocity) rate, annual and semi-annual periodic terms, the coseismic step, and a logarithmic decay. The general form for each component of the displacement, u, should be something like:

$$\begin{aligned} u^i(t) &= u_0^i + v^i t \\ &+ \sum_p \left(s_p^i \sin(2\pi t/T_p) + c_p^i \cos(2\pi t/T_p)\right) \\ &+ \sum_j h_j^i \mathcal{H}(t - t_j) \\ &+ \sum_k r_k^i \mathcal{H}(t - t_k) \log(1 + (t - t_k)/\tau) \end{aligned}$$

where $i =$ (north, east, up), and $\mathcal{H}$ is a Heaviside step function. You should be able to set up the estimation of the relevant constants as a standard linear over-determined problem, $Gm = d$, where $m^i = [u_o^i, v^i, s_{p=1..P}^i, c_{p=1..P}^i, h_{j=1..J}^j, r_{k=1..K}^i]^T$. You can have multiple periodic signals (I recommend periods, T_p, of 365.25 days and perhaps half that). Above, we explicitly include the possibility of having jumps both due to earthquakes (with transients) and jumps without transients (hardware issues or small earthquakes). In other words, you expect the number of jumps to be greater than or equal to the number of transients you assume ($J \geq K$). Note that the date of the jumps are assumed known a priori, otherwise the problem is no longer linear. Please consider all 3 components from all 4 sites. Similarly, to keep the solution linear, you will have to fix the characteristic time constant, τ, and only solve for the amplitude. You can iterate on τ manually or use a more sophisticated approach if you like - the values should be of order a few days. The time series will have data gaps, with different gaps for each site. They also do not all start on the same date. For numerical reasons, you may find it best to shift your time axis in the estimation (not in plotting) such that you start at zero, not at some arbitrary large time (like matlab's internal dates).

- Take a look at the formal model covariance matrix from your estimation (C_m not C_d) - are there significant tradeoffs between your model parameters for a given time series? You may find it more useful to convert the model covariance to model correlation.
- Tabulate the rates and jumps along with estimated errors.
- Make a map showing the estimate coseismic steps. Do they make sense?
- Plot the original time series with your parametric curve fit superimposed (remove the constant offset, u_o^i, from both data and model to facilitate intercomparison).
- Plot the time series with the offset, the linear rate, and the earthquake jumps removed from both data and model.

- Now plot the total residual time series with all model terms removed.

- What is the typical RMS of the total residul for the 3 different components? (Note we are assuming random Gaussian noise, which is probably wrong). Is there behavior in the residual suggesting that your parameterization is insufficient? Do the residuals stay approximately constant with time? Things to contemplate: Are seasonal periodic signals necessarily sinusoidal? Depending on the source of these, do we expect the amplitude to be the same year after year?

- Is there any commonality to the residuals from *a given component* of the 4 different time series? If you add the residuals from all the sites, this gives you a crude estimate of what is called Common Mode Error (CME) which is partially due to problems estimating the reference frame in each day's solution. What is the RMS of the CME? Does it change amplitude with time? You can remove this estimate of CME from the original time series, does it change your model parameter estimates?

- What challenges can you imagine would occur with estimating CME?

The file `site_coordinates` gives you the latitude and longitude coordinates for plotting purposes. I have given you several useful Matlab files to get you started:

- `load_gps_data.m` - reads in gps data file for a given site and also calculates the date vector in matlabs internal data manipulation format. It provides data, the three components and their 1 sigma variances. You should include these in your estimation.

- `data_prep.m` - a code that calls `load_gps_data.m` and makes a sample plot. This code could serve as the basis for reading all the files.

- `yearday.m` - converts from year/dayofyear to Matlab's internal date format

Useful built-in matlab commands for dealing with dates include `datestr, datenum, and datetick`. On all your plots, please make sure to choose the y-axis limits appropriately in order to both show the dynamic range of a given time series while at the same time facilitating the intercomparison of similar quantities from different time series.