

Using GPS data in teaching data analysis

Seth Stein

Department of Earth and Planetary Sciences
Northwestern University, Evanston IL 60208

Although GPS data are most naturally used in teaching classes about tectonics, they are also very useful in teaching data analysis. This handout shows two homework problems used in a data analysis course for advanced undergraduates and beginning graduate students.

These use the fact that despite their complex derivation, the final GPS data are positions that are easy to visualize, as are uncertainties in positions and velocities.

One problem uses GPS measurements of two sites, one with clear sky visibility and one surrounded by buildings. Collecting a large number of position measurements illustrates how Gaussian distributions arise from large sets of data, and how the standard deviation can be interpreted physically. The second uses a large dataset of site velocities to illustrate how the uncertainty in a site velocity as a function of measurement interval can be derived from the linear propagation of errors.

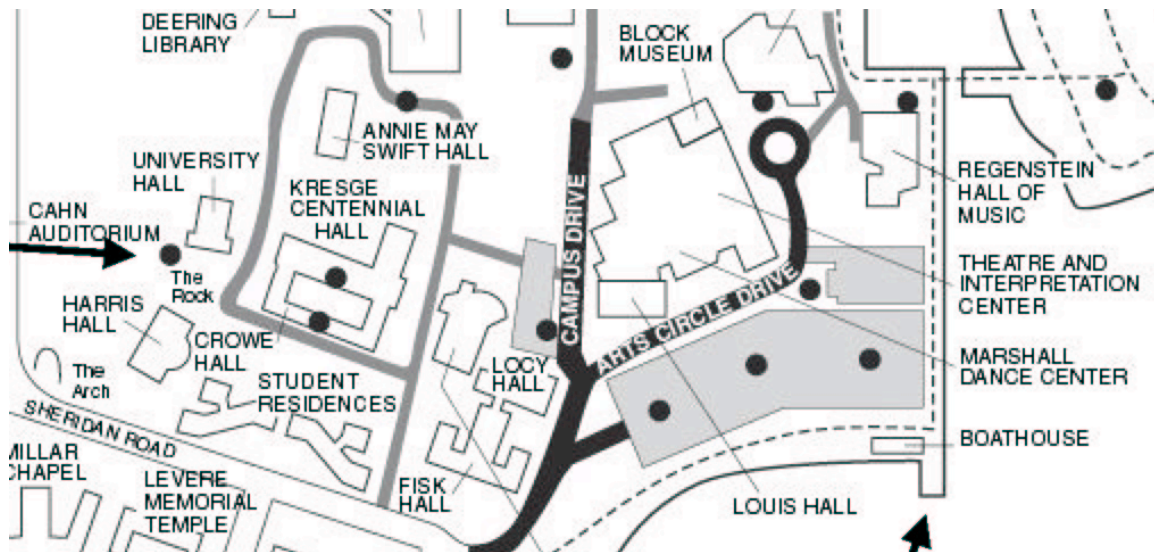
The course handouts and problem sets are at
<http://www.earth.northwestern.edu/people/seth/326>

The class also uses examples from other branches of the earth sciences and also finance, for which there is a large popular literature and natural student interest:

“Most disasters come from the fact that individual scientists do not have an innate understanding of standard error or a clue about critical thinking. As a practitioner of uncertainty I have seen more than my share of snake-oil salesmen dressed in the garb of scientists. The greatest fools of randomness will be found among these.”

Fooled by Randomness: The Hidden Role of Chance in Markets and Life
Nassim Nicholas Taleb

We will collect a set of GPS data giving the positions of two points on campus: “the point” on the lakefill (SE corner) and “the rock” (N side)



The point (SE corner)

Take 7 measurements, at least an hour apart. Put the results into the list on the next page, and the spreadsheet in

<http://www.earth.northwestern.edu/people/seth/326/gpshw.xls>

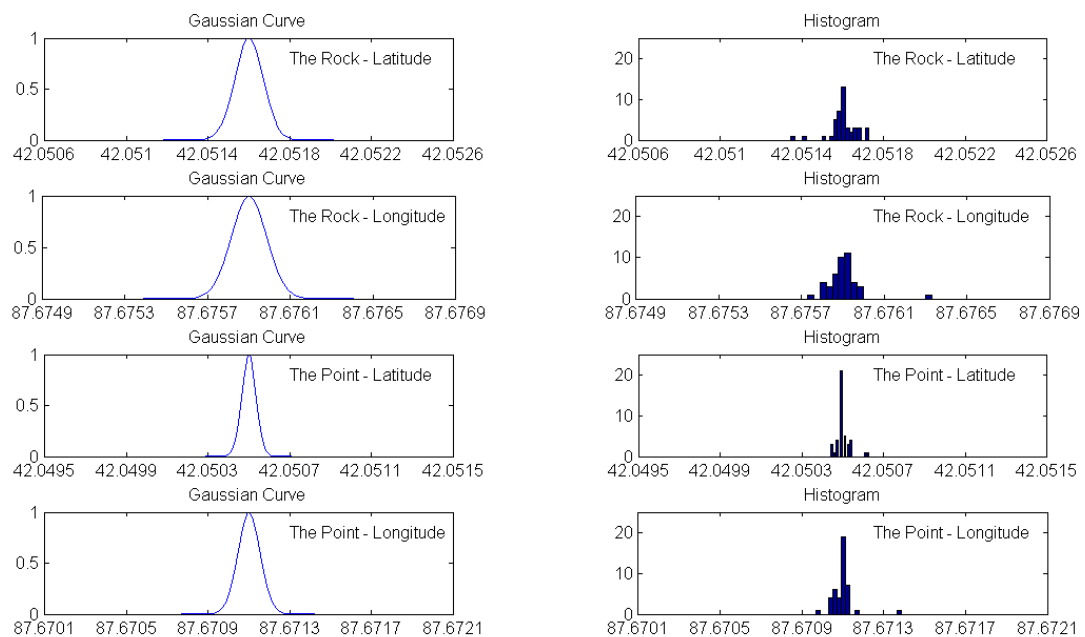
Then:

- 1) Combine the measurements into one spreadsheet (or dataset)
- 2) Find the mean and standard deviation of the latitude and longitude of each of the two sites.
- 3) Assuming 1 degree equals 111 km, give the standard deviations in meters.
- 4) Plot histograms of the positions and a Gaussian curve that describes each.

Mean and standard deviation:

	The Rock		The Point	
	Latitude	Longitude	Latitude	Longitude
Mean (degree)	42.05157	87.67593	42.0505	87.67115
SD (deg)	0.00007	0.00009	0.00004	0.00005
SD (meter)	8	9	4	6

(Assuming 1 degree equals 111 km, SD in meters.)

Plots:

Measurements within bins are plotted in the histograms. The Gaussian curves are plotted by using the mean and standard deviation of the measurements. Comparing shape of the two categories of figures, same characteristics can be found.

Measurements at the Rock have larger standard deviations than those at the Point. The reason is that the sky view is better near the Rock, which means more GPS satellites are available and the measurements are more accurate.

Measurement of latitude tends to have better accuracy. I think this is related to the azimuth coverage of the satellites. Maybe there are more satellites along the longitudinal direction.

Propagation of errors

We often use the *propagation of errors*, a general method for finding the relation between the uncertainty in a function and the uncertainty in the variables that it depends on. If z is a function of multiple variables, then

$$z = f(u, v, \dots),$$

and we have N measurements of (u, v, \dots) . The mean value of the function is its value for the mean of the arguments,

$$\bar{z} = f(\bar{u}, \bar{v}, \dots),$$

and its variance is

$$\sigma_z^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (z_i - \bar{z})^2.$$

We can show (Stein & Wyssession, 6.5.1) that

$$\sigma_z^2 = \sigma_u^2 \left(\frac{\partial z}{\partial u} \right)^2 + \sigma_v^2 \left(\frac{\partial z}{\partial v} \right)^2$$

This relation, called the propagation of errors equation, illustrates that the extent to which the uncertainty in each variable contributes to the uncertainty in a function depends on the partial derivative of the function with respect to that variable. Here we assume that the variations in the different variables are uncorrelated (which is not always the case).

1) We can use this to show that as we observe over longer times, estimates of geodetic velocities improve. Consider measuring the rate v of motion of a monument that started at position x_1 and reaches x_2 in time T . If the position uncertainty is given by its standard deviation σ , use the propagation of errors equation

$$v = (x_1 - x_2)/T \quad \text{to show that} \quad \sigma_v = 2^{1/2} \sigma / T \quad (8)$$

where σ_v is the uncertainty of the inferred rate. Thus the longer we wait, the smaller the velocity uncertainty becomes, even if the data do not become more precise.

2) Fit a curve of the expected uncertainty versus time to the data in the plot and estimate the uncertainty in position.

GPS velocity estimate uncertainty vs measurement timespan

- 40

CALAIS ET AL.: DEFORMATION OF NORTH AMERICAN PLATE

Expect
 $\sigma \propto$
 $1/\text{Timespan}$

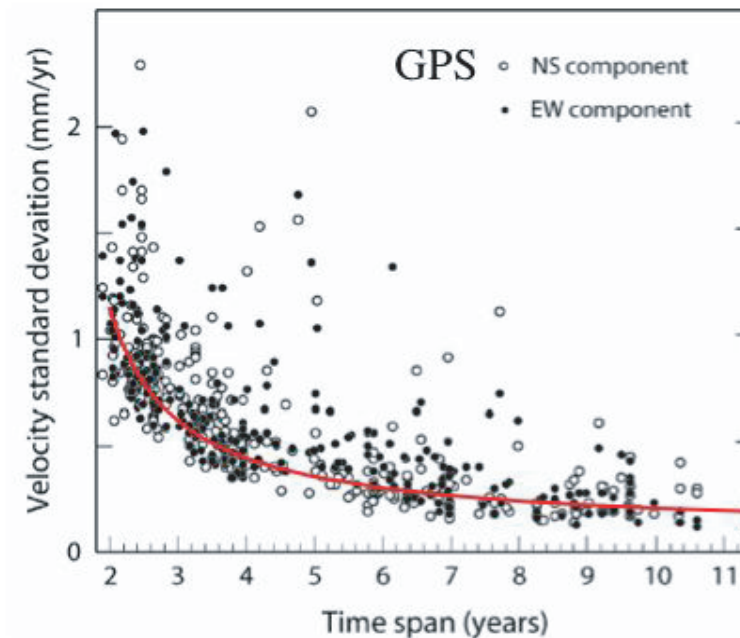


Figure 4. Velocity standard deviation as a function of measurement time span.

Calais et al, 2006

30