Overview

This class gave me ideas at a much faster rate than I was able to process them. I chose to attempt several mini-projects that are not necessarily related. I completed a small project using passive seismic observations from Disney World, FL, then moved to more familiar topics in glaciology, all related to iceberg calving.

Disney World Seismic Data

The concept of a national seismic noise baseline would help seismologists develop instrument vaults, and improve processing for small events where noise causes significant deterioration when processing solutions, like earthquake locations and focal mechanisms. Data from the U.S. National Seismograph Network (operated by USGS) was obtained for station DWPF, which is located in a deep borehole near Disney World, FL. The data consists of the number of ‘hits’ for each of 151 power bins for each of 82 periods, which cover the instrument response band. ‘Hits’ represent Fast Fourier Transforms of complete hour-long data segments. For each FFT the power spectrum is binned into discrete period amplitude bins, such that a ‘hit’ marks the power bin with the most seismic energy for that hour, for each period. Each period has N hits, where N is the number of data segments transformed. A total of 17412 hours of data have been transformed. Thus, the data represent background noise plus small amounts of earthquake energy.

If hits for each period and power are divided by N, the conditional probability of having seismic noise energy given a period is obtained. A histogram for 1s period is shown in Figure 1. Many periods exhibit bimodal and/or highly skewed distributions, because amplitude (power) has a tangible minimum level (0) and no upper bound. Parametric pdf’s cannot accurately describe many of the power density distributions. However, kernel smoothing density estimators can capture such non-parametric properties in data. Non-parametric pdf’s were fit to the data using kernel smoothing techniques (following the handout in class, but I don’t know the reference) using Matlab software. A joint pdf is not obtainable since the probabilities are conditional, but one dimensional pdf’s were fit and assembled on a regular grid (Figure 2). This surface does not represent a pdf, but rather an assembly of pdf’s. The integral of the surface is equal to the number of discrete periods included in the grid. Each individual pdf integrates to one. Slices of this data show interesting properties of seismic noise. Four slices are shown in Figure 3 where a) uses the optimal bandwidth \( h \) for a normal smoothing kernel:

\[
h = \left(\frac{4}{3n}\right)^{\frac{1}{5}} \sigma
\]

where \( n \) is the number of data, and \( \sigma \) is the standard deviation of the data. \( h \) was varied from b) 0.6\( h \) to c) 2\( h \) to study the effects of changing the bandwidth. This figure shows the probability of occurrence at a power for periods of \( \sim 3 \) seconds, which is the characteristic period of storm waves. The sub peaks probably represent hurricanes, which were prevalent in Florida last year. Centered at 1s period is earthquake energy.
One major problem in seismology is removing anthropogenic noise from the earthquake energy as seen by the overlap of power spectra between these two signals. The human signal is centered on 0.4 s and exhibits a strong bimodality from sleep/wake (machines) cycles characteristic of our species. The station has a very narrow band of high probability, low power long period noise, taken from ~70 s period. This is due to the fact that the sensor is vaulted in a ~300 m deep borehole.

The mode is a very important property to extract from this signal (Figure 4). This curve gives the power of maximum probability at each period. In the near future, we will calculate these pdf’s for the entire USNSN and box plot extracted modes. The statistic can be used to identify outliers (noisy stations) and incorporate properties of quiet stations into future vault design.

Columbia Glacier Force Balance Analysis

Alaskan tidewater glaciers exhibit intriguing behavior. They show little relationship with climate, i.e. the positions of their termini do not depend on global climate as terrestrial glaciers do. Rather they fluctuate through cycles of rapid, often catastrophic retreat (km/yr), and slow millennial time scale advances. The mechanism for rapid mass loss (much more rapid than melting could account for) is iceberg calving. This process has played an important role in our planet’s past (Heinrich events) and is important today in Alaska, Greenland, and Antarctica (ice shelf buttressing may turn out to be way more important than we thought!!!) when considering potential changes in global sea level and the spatial distribution of the planet’s population.

Recently we submitted a paper on a time series analysis of force balance at Columbia Glacier AK, using an extensive record (but I need more data) of positions and speeds on this glacier both prior to, and during its rapid retreat. This paper is available as a pdf at http://instaar.colorado.edu/shad. Force balance analysis (Van der Veen and Whillans, 1989) outputs four terms at each spatial point describing stresses on the ice at a point. These include an action force or driving force arising from gravity ($\tau_d$), plus retarding forces including friction at the sides and bed of an ice column ($\tau_f$ and $\tau_b$), and an along flow coupling term ($F_{lon}$), such that ice a point ‘feels’ the dynamics of ice upstream and downstream of itself.

Many aspects of this class were incorporated into the paper while dealing with reviewer comments. The neatest thing was to transform plots showing a low quality statistical analysis into a more robust plot. The statistics are simple, but the idea greatly strengthened the paper. Figure 5 shows the plot before updates, where the mean, max, and minimum centerline value was plotted a function of time to represent the glacier at each time, which isn’t a great assumption. This plot was transformed to a time series of box plots where the top panel shows the temporal distribution of samples, which is highly irregular. This was poorly shown prior to update and has important implications for interpretation. Now instead of a single centerline value, all centerline values are shown which does more justice to the dynamics of the glacier. Ratios of force balance terms to the driving stress show the relative importance of each term in the retreat, and are boxplotted in Figure 6. An example of a feature visible in these new updated plots is the change in skew for $\tau_b/\tau_d$, which describes the importance of basal friction. During the retreat (begins 1980, still underway) the skew of this ratio changes from negative to
positive as the domain size shrank and became dominated by a region of steeply sloped and thick ice.

Another major improvement comes from a statistical analysis between force balance terms and surface speed. Figure 7 shows the driving stress (often thought of as main control on ice motion), basal drag (fuzzily important when sliding over the bed dominates the flow, like in this case) and the ice speed along the glacier centerline, $\xi$ (low $\xi$ = upstream, high $\xi$ = downstream; all terms are normalized by the maximum value). A change in the relationship occurs as the retreat begins. Ice speed shows dramatic changes, while the stresses show minor or no changes. Before review, this paper had just mentioned this, without any quantification. I thought it was obvious. First, boxplots were drawn, which graphically shows the large change in speed in both the up- and downstream regions, strongest in the downstream region. Changes in driving stress appear small, as do changes in basal drag. I used a Rank sum test (Helsel and Hirsch) to quantify these changes. At the 95% confidence level there was indeed a significant change in velocity, but none in driving stress or basal drag in the upstream region. There was a change in basal drag at the 95% confidence level with a p value of 0.62 of a change in basal drag near the terminus. These statistics support the hypothesis that at the onset of retreat there was a shift in the controlling mechanism for ice motion from driving stress to basal drag. The relationship between basal drag and ice speed would be something like

$$u = k \frac{\tau_b^n}{\tau_b^m}$$

where $k$ is a dimensional constant and $u$ is the sliding speed. Basal drag can be thought of as a proxy for effective pressure, and the concept of an effective pressure sliding rule is growing stronger in glaciology (replace $\tau_b$ above, with $p_{eff}$, which is the overburden pressure – basal water pressure. To show this proxy is valid at Columbia Glacier I show the relation between effective pressure (calculated via assumption of basal water connection to sea level pressure) and velocity, which is similar (and stronger) than the relation between velocity and basal drag (they are indeed different and both hard or impossible to measure at tidewater glaciers due to extensive surface fracturing). The relationship is obvious, but I removed linear and polynomial trends (1985 speed) from the data and computed Pearson’s r to show the strengthened correlation with initiation of retreat. $R$ changed from –0.13 to -0.9, a significant change, and the first ever quantification of this relationship from data (Figure 8).

I also wrote a routine to calculate Kendall’s Tau, but this didn’t turn out to be that useful. The gridding routine I used in this paper is basically a kernel smoothing filter, but I didn’t know that when I wrote it up, its nice to formally learn about this! I made a circular shaped kernel and gridded the data by assigning a linear weight as a function of radius from each grid point but within the circle. Its like a cone shaped smoothing kernel. Next time I will use the normal kernel for this. I could easily use the 2-D kernel smoothing density estimator I wrote for the seismology data, but didn’t end up using. This was a result of misunderstanding the data. At first, I thought I was given joint probability and so developed a 2-D routine since Matlab only has a packaged smoothing estimator for 1-D. I’ve compared mine with R and the results are very similar.

Regression
It is important to realize we know next to nothing about tidewater glaciers and why they retreat. Conceptual models that describe observations exist, but very little quantified work exists, mostly because of the difficulty in obtaining measurements in such an unstable environment. Columbia glacier moves 30+ m/d at its terminus nad has retreated over 15 km in the past 25 years. This serves as a preface for the rest of this project, where the ideas are there, but the data won't speak, or I just haven’t found it yet. When considering tidewater glacier dynamics, predictive capability seems a ways off, but a mere understanding of what processes are or are not important in governing flow, retreat and/or advance would be a significant contribution. In the early 1980’s, Brown and others (1982) published an empirical calving law from a linear regression between calving rate and water depth, which quickly became accepted by the community because of the high $R^2$ value and for lack of anything better, even though there was no physical basis for their model. Since then, the idea that flotation level or buoyancy is important has gained momentum (Meier and Post, 1987; Van de Veen, 1996, Vieli and others, 2001; O’Neel and others, 2003). The extensive data set at Columbia Glacier offer an opportunity to investigate different parameters and their relationship with calving. Suspect parameters include, but are certainly not limited to: ice speed, strain rate, ice thickness, water depth, ice cliff height, height above buoyancy and surface slope. Calving rate is calculated as the difference between incoming ice speed and rate of length change:

$$U_c = U_i - \frac{dL}{dt}$$

This difference equation is highly dependent on the time scale over which it is calculated (Van der Veen, 1996). At annual and longer time scales there is a nearly perfect correlation between ice speed and calving rate that has no physical meaning. Hence, ice speed is difficult to analyze. An independent quantity for calving flux is available at Columbia Glacier, called ice-to-be-calved. This is a photogrammetrically determined quantity that I will use in the analysis (Krimmel, 2001). I wish to identify potentially important parameters in controlling calving rate through ordinary linear regression, and multiple linear regressions. In the near future these should be computed with nonparametric methods, as the data is mostly non-normal.

The underlying idea for these regressions is to first choose the best of the three possible buoyancy variables. Simultaneous identification of linear relations between the force balance terms and calving rate using multiple linear regression is the next step. An additional multiple regressions between along flow strain rate, surface slope and speed against calving flux will also be performed. The important parameters from this regression and important parameters from the regression of calving and force balance terms would be used as a full model subjected to stepwise regression to possibly determine the main controls on calving, incorporating the best buoyancy variable, the significant force balance variables, and any other geometric and/or motion variables that turn out to be significant. I avoid putting all these into one huge regression model because they are not all significant and I feel lumping them together based on physics might be a stronger method.

This analysis is temporarily halted by data mining complications, but the idea is to identify the important parameters through two stepwise regressions. My feeling is that the data need to be very carefully considered prior to the regressions, as the calving calculations are quite sensitive. I had hoped to complete this, but didn’t come close.
Local regression methods will also prove to be insightful, as it is more likely that this process is highly non-linear, with calving increasing rapidly as some of these parameters approach critical values. This will be a good thing to continue working on next semester.

Buoyancy regression

Three of the culprits often blamed to various degrees for calving are water depth, ice thickness at the terminus, and height above buoyancy, \( H_b \), which captures part of both these parameters.

\[
H_b = H - \frac{\rho_w}{\rho_i} d_w
\]

where \( H \) is the ice thickness at the terminus and \( d_w \) is the water depth. \( \rho_i \) and \( \rho_w \) are the density of ice and water. Perhaps a single parameter linear regression will reveal which quantity is most important for iceberg calving.

Figure 9 shows the results of single linear regression analysis between calving rate (dependent variable) and \( H, D_w \) and \( H_b \). None of the data are normally distributed, which is a first hint that least squares regression is not the best choice. In the future, a local regression should be performed or some transform to normalize the data. Scatter plots do show fairly strong linear relationships between the data, strongest for height above buoyancy. A table of values interpreting the regressions is given below.

<table>
<thead>
<tr>
<th></th>
<th>H vs UC</th>
<th>Dw vs Uc</th>
<th>Hb vs Uc</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^2 )</td>
<td>0.43</td>
<td>0.54</td>
<td>0.66</td>
</tr>
<tr>
<td>F &amp; p</td>
<td>53; 4E-10</td>
<td>83; 2E-13</td>
<td>134; 0</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.66</td>
<td>-0.74</td>
<td>-0.81</td>
</tr>
<tr>
<td>MI</td>
<td>0.034</td>
<td>0.032</td>
<td>0.113</td>
</tr>
</tbody>
</table>

The regression residuals ice thickness and water depth are skewed, but height above buoyancy residuals are more normally distributed, as shown in histograms and qq plots in Figure 9. This is not that surprising given the non-normal input. Next, I checked for heteroskedascity by plotting the residuals against the predicted values and each independent variable. All three seem to suggest that a non-linear regression would be more appropriate, because the residuals increase in magnitude as both \( X \) and \( y \) do. Acf analysis of the residuals show that there are significant correlations between the residuals at low lags, especially for ice thickness. The lowest correlated residuals are for \( H_b \). This indicates a violation of a fundamental assumption in linear least squares regression, which requires independence for the residuals, and is a common result when time series are studied. This implies that the mean square error may seriously underestimate the true variance and that the null hypothesis is too easily rejected. It does give an indication of the data memory, and can be more robustly studied by doing an ACF analysis using the mutual information statistic. Mutual information statistics show that all three variables are correlated in some sense, that is, all the values are outside of the bootstrapped confidence intervals (Figure 10). All tests point towards the height above buoyancy as the most important parameter for calving.

Calving and Force Balance

One idea is to seek relationships between the above mentioned force balance terms and calving. For force balance terms, I use the most frequently occurring spatial location of each term along the centerline. From this analysis, I obtained a time series of 42 points with highly irregular spacing. Timing of force balance data is independent of
ice-to-be-calved, and there are only 38 data points at simultaneous intervals. Preliminary analysis shows only marginal relationships between each term and calving (Figure 11). Likewise with an all inclusive multiple regression, all beta’s are insignificantly different from zero, even after detrending the data. Bloody outliers! A more complete time series would be helpful. The data mining for this was done very quickly, and not very carefully, and I think with a more careful analysis something may come of it. The correlation between lateral drag and calving has the best visual correlation, and is strongly linked with speed, suggesting speed may indeed be important, although none of the force balance terms showed any significant correlation even in a multiple regression (all beta confidence intervals included zero). This was also true when a weighted least squares regression was applied.
Figure 1. Histogram and kernel smoothing density PDF for a single period of seismic data. The x-axis shows the power in Db and the y-axis is the probability of occurrence.
Figure 2. Assemblage of 1-D nonparametric PDF’s onto a regular grid. Log period is on the x-axis, power (Db) on the y-axis and the probability of occurrence at each period is given on the z-axis and shown in colors corresponding to the color scale on the right.
Figure 3. Slices through Figure 2 at particular periods. The bandwidth changes from $2h$, to $h$, to $0.6h$ where $h$ is the optimal bandwidth as given in the text. Strong bimodality is evident especially in the human band, which has significant overlap with the earthquake band.

Figure 4. The mode from Figure 2.
Figure 5. Figure from O’Neel and others showing improvement in information transmission using box plots.
Figure 6. Box plots showing stress ratios in the force balance for Columbia Glacier, during its rapid retreat. The relative importance of each term is shown in the figure.
Figure 7. Changes in velocity, driving stress and basal drag during the retreat of Columbia Glacier. A rank sum test was used to quantify the changes.

Figure 8. Changes in the relation between velocity and effective pressure during the retreat of Columbia Glacier. Data was detrended and a Kendal Tau correlation coefficient as well as pearsons r were calculated to show the change in correlation at the onset of retreat.
Figure 9. Regression between buoyancy variables and Calving flux, with model diagnostics.
Figure 10. Box plots of Mutual information statistic and confidence intervals on the correlations.
Figure 11. Results of multivariate linear regression between force balance terms and calving flux. Only single variable regression are shown with OLS in blue and WLS in red. Outliers make all correlations insignificant, and better data mining needs to be done before any progress can be made.

References


