Glaciated Valley Profiles: An Application of Nonlinear Regression

Mark C. Greenwood
Department of Statistics, University of Wyoming, Laramie, WY 82071

Neil Humphrey
Department of Geology and Geophysics, University of Wyoming, Laramie, WY 82071

Abstract:

The cross valley profile of bedrock valleys that have been extensively glaciated are generally U-shaped; questions of interest include estimation of the actual shape from the data, ability to discriminate between shapes and comparison of these profiles across valleys. Data for such profiles are obtained from digital elevation maps via sampling transects across valleys. Current methodology as in Pattyn & Van Heule (1998) employs "curve fitting" without the use of any diagnostic measures. We consider improvements that include the use of random sampling for selection of profiles and the use and comparison of various nonlinear regression models; illustrated with an example from the Himalayas.

Key Words: valley profiles, nonlinear regression, model selection.

1. Introduction

In geomorphology, there is interest in distinguishing between valleys whose primary force of formation has been the motion of a glacier and those formed by other processes such as rivers. Glaciated valleys tend to have a shape across the valley that is often described as U-shaped, whereas river-cut valleys tend to be V-shaped. The literature in this area has tended towards naïve curve-fitting, with associated limitations on the conclusions available that are imposed by these methods. We describe improvements in the methodology of obtaining and analyzing the valley elevation cross-profiles. First, we present a description of the problem and a brief history of attempts to solve it, then some details of the example used to illustrate the techniques, followed by a discussion of improvements and an application of those methods to a Himalayan valley.

2. Description of the problem

The observed shape of any valley is the culmination of the activities of a number of different erosive and sedimentary processes, as well as the type of bedrock of the valley. A valley whose formation was dominated by the activities of a glacier tends to have profiles that are U-shaped throughout the length of the valley. A valley primarily formed by flowing water tends to a V-shape (McGee, 1894). Subsequently, there has been frequent interest in the geomorphology literature to describe the curvature of observed cross-sections of valleys and use that information to describe the amount of glaciation in a valley.

Describing valleys as either U or V-shaped does not characterize the curvature of a valley, and does not lead directly to a statistical model to employ. A theoretical discussion found in Hirano & Aniya (1988) leads to a suggestion that the ideal shape of a glaciated valley is a catenary curve. (A catenary curve is the curve defined by a chain that is supported only at the ends.) Prior to our work no one has actually fit a catenary curve to valley profiles. Other models that are suggested in the geomorphology literature are power law regressions ($y=\alpha x^{\beta}$) and parabolic regressions (Doornkamp & King, 1971, Wheeler, 1984 and James, 1996, among others). The use of quadratic regression is also suggested as an approximation to the catenary curve that can be fit using conventional least squares techniques (Hirano & Aniya, 1988). The power law regression contains information on curvature in the value of the exponent (β), with values close to 1 indicating a linear valley side, and values closer to 2 indicating a parabolic shaped fit (Graf, 1970, Hirano & Aniya, 1988, James, 1996 and Li, Lui & Cui, 2001(1) among others). Conclusions are also drawn about the amount of glaciation based on the estimated value of this exponent. Drawbacks to the conventional power law model are that it must be fit to each side of the valley separately, and that this often results in fitted values that do not meet at the center of the valley. The observations obtained are locations in space, and the location of (0,0) in the coordinate system is called the datum. The datum choice affects the values of the coefficients, even the coefficient in the exponent supposedly just describing the curvature of the valley profile (Harbor & Wheeler, 1992 and Pattyn & Van Heule, 1998).

A more recent generalization of the power law regression model is the generalized power law (GPL) model ($y-y_0=\alpha|x-x_0|^{\beta}$), where the datum is estimated as part of the model as (x_0, y_0) (Pattyn & Van Heule, 1998). β still contains information on the curvature and this model is invariant to changes in coordinate locations. However, it does not accommodate asymmetric valleys, which do exist. Li, Lui and Cui (2001(2)) attempt to solve this problem of asymmetric profiles by working with the width and depth of the valley, instead of the valley profile. This does not lead to a model that is directly comparable to the others previously described, and is not considered further here.

Although a number of attempts have been made to address the question of the shape of a glaciated valley in contrast to other types of valleys, there is still no definitive method of addressing this issue. This is demonstrated in that the most recent papers are still attempting very different methods of obtaining solutions. Nearly all previous work in this area has suffered from a lack of an objective method of obtaining information, "curve-fitting" techniques employed as opposed to statistical modeling, a lack of assessment of the adequacy of those models that are considered, and a more thorough comparison of models.

3. Example

To illustrate the methodology, an example from the Dona Khola Drainage in the Himalayas is used. This valley is known to have been glaciated; it is intended to illustrate the techniques in a valley that should be relatively U-shaped. All data used here come from a Stereo Optic Satellite image that was converted to a Digital Elevation Map (DEM) with a 100m x 100m pixel size; the mean elevation is available for each pixel. Figure 1 shows an image of the DEM with the drainage of interest in the box (top) and a satellite image of the area can be found in the bottom of Figure 1. Note that this image shows the Tulagi glacier at the beginning (right side) of the drainage.

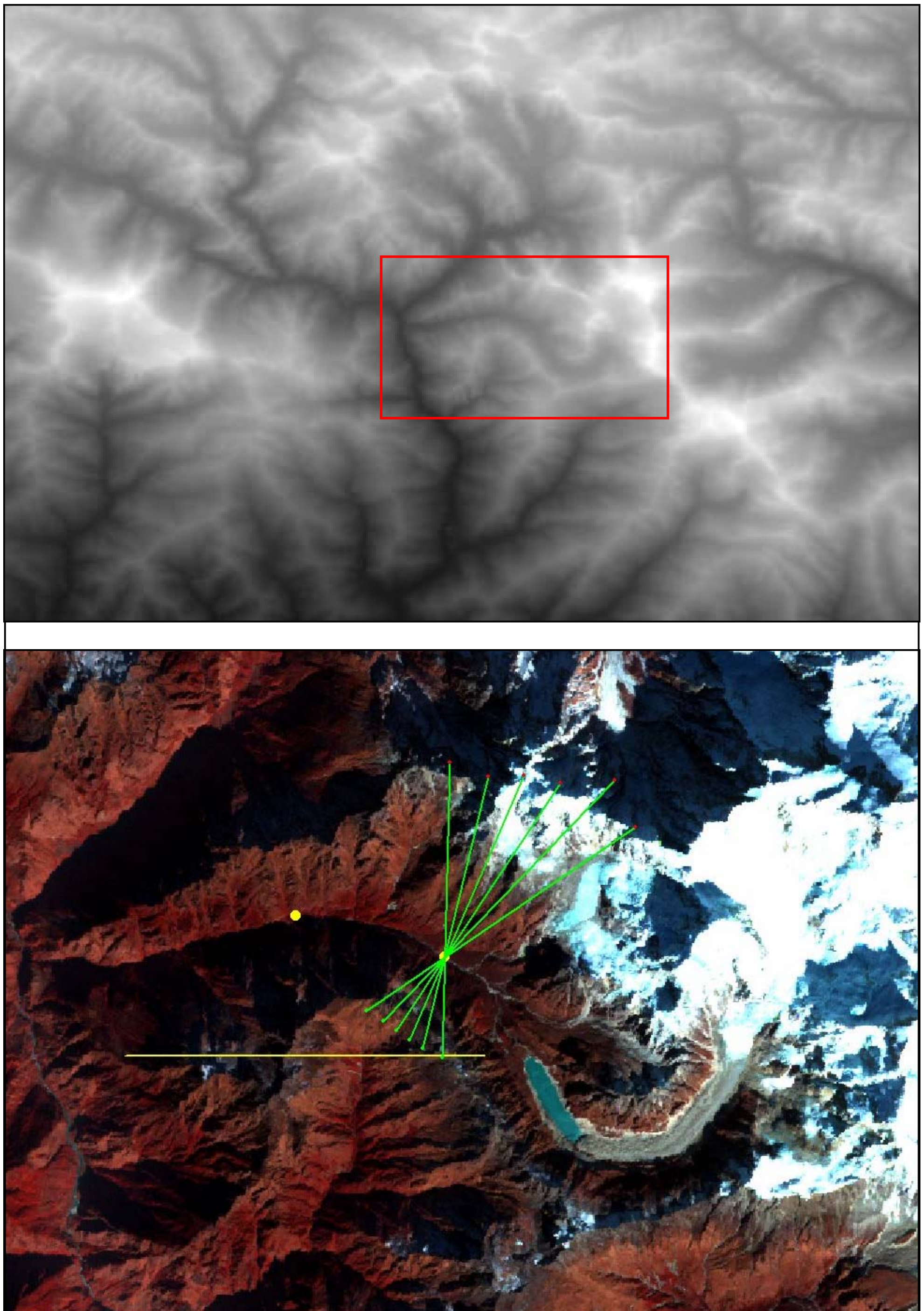


Figure 1. DEM (top) and satellite image (bottom) of the study area, with detail of the Dhona Khola drainage. Yellow dots in bottom image are sampled locations. Yellow line was used to generate the sample and the green lines are sampled transects from Location 2.

4. Methodology

In looking at current methods used to address this problem, there are some noticeable limitations. The first is the choice of the location in the valley to sample. There is no discussion of an objective sampling methodology, and a subjective choice could be a source of considerable bias. It is also not trivial to determine the transect that is perpendicular to the run of the valley, especially when valleys are curved. Additionally, for a given transect, the furthest that you could sample is to the top of the neighboring ridge, which biases profiles to V-shapes. Since glaciers usually do not occupy the entire valley and only fill the bottom few hundreds of meters of the valleys, the inclusion of the upper portion of the wall can cause this bias. An objective method of determining this 'wall height' for a profile is discussed. Once a profile is selected, it is unusual in the geomorphology literature for more than two models to be compared, model adequacy is rarely considered, and point estimates are interpreted without regard to the precision of the estimates. Improvements are discussed for all of these problems.

Common sources of valley profiles for this type of work are from field surveying, 'contour crossing' on topographic maps and using satellite imagery. The second two methods may be the least effective in obtaining information for this application, as the number and quality of observations are usually not in the most interesting sections of the profiles (James, 1996). However, field surveying methods are time consuming and costly. Sampling from satellite imagery leads to extremely efficient data collection compared to field surveying, with the 'contour crossing' method also saving time in comparison to field surveying. Sampling from satellite imagery can easily be accomplished using a computer package such as ArcView. This allows for increases in the volume of information available for analysis, but possibly a limited knowledge of the valley under consideration. With field sampling, knowledge of the valley would be high, but the amount of data collected would likely be limited. Another issue is that it is difficult, even with field surveying, to know the extent of glaciation on the valley wall.

One of the simplest improvements to current methods that we suggest is to employ random sampling. We are interested in describing the shape of a valley, and if we define a relatively homogeneous section of the valley as the population of interest, we can randomly sample from this population. Along this section of the valley, a line parallel to the run of the section of the valley is defined (yellow line in Figure 1(bottom)), and a location on this line is randomly selected using a uniform distribution. From this selected point, the middle or center of the valley is determined (2 locations were sampled, yellow dots in Figure 1). A number of different transects are selected that go through this point, to aid in determining the transect that is perpendicular to the run of the valley (green lines in Figure 1 for Location 2). This process could be repeated to select additional locations in the valley, and future research will look at using a multivariate analysis in this situation.

James (1996) discusses the effects of different choices of wall heights in a particular valley. Ideally, the wall height would be chosen as the maximum height in the valley that has been glaciated. This is not possible since it is difficult or impossible to determine this height especially from satellite information. To standardize the choice of wall height without biasing the results to favor U or V shapes a priori, an objective method of making this choice is needed. On each side of a valley, there is a ridge or top of the valley wall. This implies that the side of the valley is going from horizontal at the bottom of the valley to horizontal at the top of the ridge. Between the two locations, there must be at least one, and usually only one, inflection point. We choose the minimum of those two inflection points for the wall height for that profile. The inflection points are determined by fitting a cubic polynomial to each side of the profile. The profile that is perpendicular to the run of the valley would have the shortest distance across the valley. To choose this optimal profile, we use the minimum inflection point criterion on each observed profile in a location and select the profile that has the minimum distance between inflection points across the valley.

Another improvement is to consider more than two models. Some of these models are described in the geomorphology literature and others are included for their ability to take on U or V-shapes or as generalizations of previously used models. The simplest model is the quadratic regression model ($y=\beta_0+\beta_1x+\beta_2x^2+\varepsilon$), which is restricted to relatively U-shaped profiles and uses conventional least squares, so has been discussed frequently. The power law regression model ($y= \alpha x^{\beta}+\varepsilon$) is the most discussed in this situation, but has major limitations. The model is usually fit using a log transformation of both sides of the equation, to allow use of conventional least squares. It must be fit to each side of the profile separately, so does not allow direct comparison with other models through model selection techniques, thus will not be considered in model selection results. Another drawback to this model is that the datum choice (choice of (0,0) in the coordinate system) affects the coefficients and fits for this model. It does, however, provide some interesting information on curvature in β.

A modification of the power law model that solves some of these problems is the GPL regression model as suggested by Pattyn & Van Heule (1998), which is $y-y_0= \alpha|x-x_0|^{\beta}+\varepsilon$. α, β, y_0 and x_0 are estimated in this model. This solves the datum choice problem, and provides a model that can accommodate either V or U-shapes. It does not accommodate asymmetric valleys, but it does provide a very useful generalization of the power law model.

The catenary curve regression model ($y=\beta_0+\beta_1\cosh(x/\beta_1+\beta_2)+\varepsilon$) can be approximated by the quadratic model, but describes the shape that Hirano & Aniya (1988) argue is the ideal shape for a glaciated valley. It is restricted to fitting a U-shape only.

The quartic regression model ($y= \beta_0+\beta_1x+\beta_2x^2+\beta_3x^3+\beta_4x^4+\varepsilon$) is a higher order version of the quadratic model and considered because it can have steeper sides than a similar quadratic model and also adds two more possible inflection points. Another, even more flexible model from Ratkowski (1990) is a ratio of polynomials regression model ($y= (\beta_0+\beta_1x+\beta_2x^2)/(1+\beta_3x+\beta_4x^2)+\varepsilon$), which can take on a variety of shapes.

The GPL model, discussed above, works well for symmetric profiles. For asymmetric profiles, modification of the GPL model can allow each side its own shape or curvature parameters. The 6 parameter, modified GPL is $y= \alpha_1|x-x_0|^{\beta 1}+y_0+\varepsilon$ if $x<x_0$ where x_0 estimates the center location and $y= \alpha_2|x-x_0|^{\beta 2}+y_0+\varepsilon$ if $x\geq x_0$. A slightly more restricted version of this model, where the α coefficient is shared between both sides is the 5 parameter, modified GPL(5). This model is $y= \alpha|x-x_0|^{\beta 1}+y_0+\varepsilon$ if $x<x_0$ and $y= \alpha|x-x_0|^{\beta 2}+y_0+\varepsilon$ if $x\geq x_0$.

Model estimates are sensitive to parameter estimates; models were estimated using Matlab and also S-Plus. This sensitivity is especially noticeable for the GPL models, where the estimation of x_0 is very sensitive to initialization.

We use two methods to describe the shape of an observed profile, using model selection criteria to assess which model is "best" and looking at the estimated coefficients from a model that has a flexible shape. The first method can provide information on curvature based on the type of model that is favored. If a U-shaped model such as the catenary curve is favored, the profile could be described as U-shaped. However, if a more flexible model such as one of the GPL models or the ratio of polynomials model are favored, this suggests that the profile is not U-shaped. The second method of describing curvature uses the GPL model, as it can take on a variety of behaviors, including both U and V shapes. The curvature is then described by the estimate of β and the associated confidence interval. Both methods are illustrated in the results in section 5.

The model selection criterion used here is the predictive discrepancy criterion (PDC) (Davies & Cavanaugh, 2002). This model selection criterion has fewer assumptions than other model selection criteria such as AIC or AIC_C. The AIC_C is known to perform well even with small samples, but requires continuous second derivatives with respect to the parameters (Hurvich & Tsai, 1989). This assumption is violated in the GPL models. The PDC does not require this assumption, and performs as well as or better than AIC_c in many situations. Compared to AIC_c, the PDC does tend to include the GPL models more frequently in the application considered.

The PDC is similar to AIC or AIC_c in that it is intended to estimate the expected discrepancy, a measure of the separation between the true model and a fitted, approximating model. The PDC uses cross validation to find an estimator of the expected discrepancy. Under the assumption of normal, independent errors, the PDC is $\sum_{i=1}^{n} \ln \hat{\sigma}_{-i}^{2} + \sum_{i=1}^{n} (y_i - \hat{y}_{i,-i})^2 / \hat{\sigma}_{-i}^{2}$, where $\hat{y}_{i,-i}$ is the fitted value for y_i and $\hat{\sigma}_{-i}^{2}$ is the MLE for the variance, both based on the data set with y_i removed (Davies & Cavanaugh, 2002). One drawback to this criterion is that the models must be fit repeatedly. In models where convergence was problematic, application of this criterion can be complicated.

5. Results

Figure 2 shows the observed profiles that correspond to the green lines in Figure 1. There are noticeable differences between the different profiles, which suggests that the choice of angle in the sample is important. The inflection point criterion is applied and the optimum profile that is selected for this location is displayed in Figure 3. Model

Figure 2. Plot of Location 2 Profiles

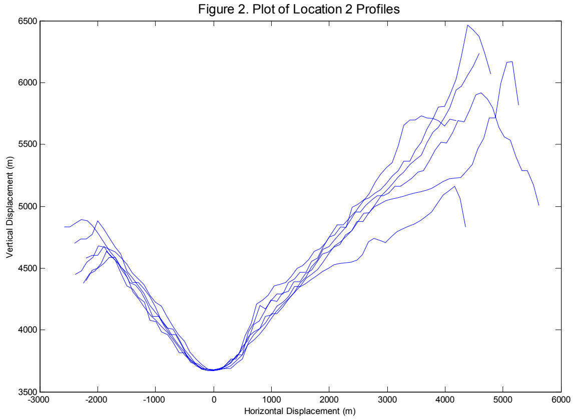
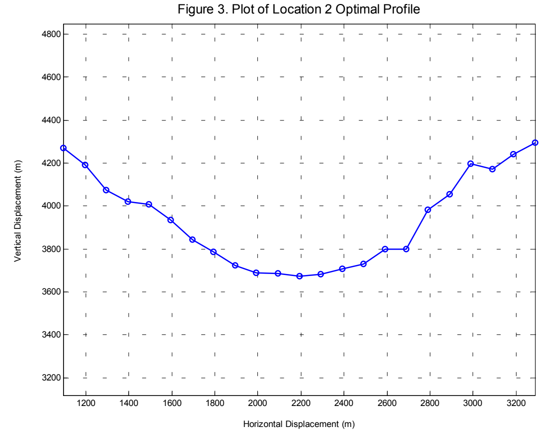


Figure 3. Plot of Location 2 Optimal Profile



selection is performed on each of the 6 profiles from this location and the model that is favored for each profile using the PDC is displayed in Table 1.

Table 1. Model Selection, Location 2

Profile	PDC Favored Model
1*	Quartic
2	Mod GPL(5)
3	Ratio
4	Mod GPL(6)
5	Ratio
6	Ratio

* Profile in Figure 3.

These results show differences in the favored models depending on the angle of the profile selected. The selection of the quartic model for the optimum profile indicates that the shape of this profile is not similar to a catenary curve, and that this profile may have a more complicated shape than simply U or V. (As a note, when this optimal profile is analyzed using conventional power law methods it would be described as being closer to U-shaped than V-shaped.)

We can also look to the results of model fits from the GPL model to describe the curvature of the profile, assuming it is symmetric. The interpretation of the exponent (β) in the GPL model is of interest, but the approximate 95% confidence interval around the parameter provides more interesting conclusions. These confidence intervals are meaningful to interpret since we collected this information using random sampling. In looking at those intervals, note that if β were 1, the valley sides would be considered linear and if it were 2, the sides would be described as parabolic. The interpretation of the amount of glaciation present is also possible by looking at the range of possible values for β. In looking at the results in Table 2, profiles 4 and 5 have confidence intervals that contain 1, all others exclude 1 and none contain 2. This corresponds to the description of this location as being somewhere between V and U-shaped, but that the shape is not distinctly U or V-shaped. Note that this model was not selected by the model selection criterion, but that there is interesting information contained in its results. Throughout the modeling process, the residuals of each model are considered, and the best fitting models' residuals were always adequate.

Table 2. GPL: 95% CI for β, $y=\alpha|x-x_0|^{\beta}+y_0+\varepsilon$

Profile	Lower Limit	Upper Limit
1	1.0701	1.7136
2	1.0494	1.3538
3	1.0095	1.4744
4	0.9344	1.2168
5	0.9205	1.3864
6	1.0769	1.4612

Figure 4. Plot of Location 2, Optimal Profile, Different Wall Heights

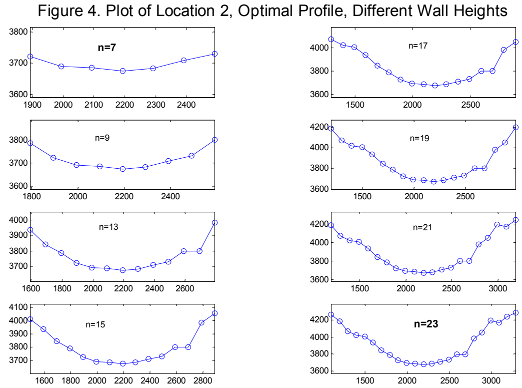
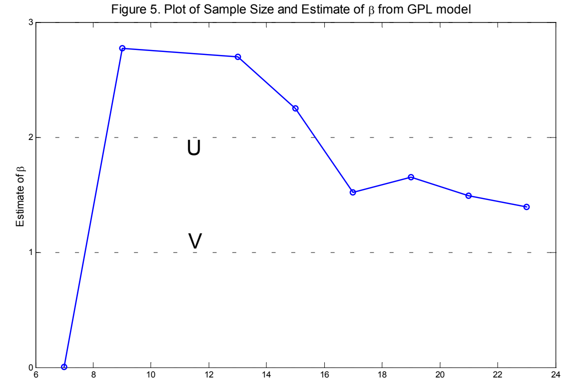


Figure 5. Plot of Sample Size and Estimate of β from GPL model



The other aspect to choosing a profile is the selection of the wall height for a given profile. To illustrate the importance of this choice, for the optimum profile discussed above, alternate wall heights are considered, starting from the inflection point criteria and reducing the wall height. The different profiles considered are displayed in Figure 4. As the sample size is reduced, the profiles seem to become more U shaped. This is also illustrated in Figure 5, which displays the estimates of β for the GPL model for the different sample sizes. It indicates that the profiles are U-shaped or more curved initially, then they become more V-shaped as the sample size increases. Model selection results also support the previous discussion and are displayed in Table 3. As the sample size increases, the shapes go from distinctly U-shaped (quadratic and catenary) to more flexible shapes (quartic).

Table 3. Model Selection Results, Different Wall Heights.

# Observations	PDC favored
7	Quadratic
9	Catenary
13	Quadratic
15	Catenary
17	Quartic
19	Mod GPL(6)
21	Quartic
23	Quartic

The model selection results for the other randomly selected location are displayed in Table 4 and mirror those for the first location, demonstrating that the angle of the selected profile makes a difference in the final results. The ratio of polynomials model is favored, which indicates that the shape is not distinctly U or V-shaped.

Table 4. Model Selection Results, Location 1

Profile	PDC favored
1	Ratio
2	Ratio
3	Mod GPL(6)
4	Mod GPL(6)
5	Quartic or Mod GPL(6)
6	Mod GPL(6)

6. Conclusions

The valley under consideration seems to be somewhere between V and U-shaped. This is based on the results from the optimal profile for location 2. From above, the 95% confidence interval for the GPL model is between 1 and 2 and does not contain 1 or 2, indicating that the shape is between V and U-shaped. Additionally, the model selection results favored the quartic model, which is a more flexible model than the quadratic or catenary, and suggests that this profile is not simply U-shaped.

With the current methods that are used in this area, there are still only limited conclusions available concerning the amount of glaciation in a valley. With the usual sample sizes, it is even difficult to distinctly describe the shape of a single valley profile. Much of this is due to the number of observations per profile, as well as the focus on only one profile to describe the shape of an entire valley. With our improvements in the sampling process, valley profiles are not pre-selected to be U-shaped, so any evidence of an observed U-shape is even stronger evidence in the direction of glaciation. Model selection and inference concerning model parameters do aid in describing the shape of a valley. With increasing precision that is becoming available in satellite information, the ability to describe and distinguish the shape of glaciated valley profiles should improve.

Special thanks to Snehalata Huzurbazar, Ian Fairweather, and Joel Harper for their help in the preparation of both this paper and the corresponding presentation. Also thanks to Joe Cavanaugh for information regarding the PDC.

References:

Davies, S. and Cavanaugh, J. 2002. 'Cross Validation Model Selection Criteria Based on the Kullback-Liebler Discrepancy', working paper.

Doornkamp, J. C. and King, C. A. M. 1971. *Numerical Analysis in Geomorphology*, Edward Arnold, London.

Graf, W.L. 1970. 'The Geomorphology of the Glacial Valley Cross-Section', *Arctic and Alpine Research*, **2**, 303-312.

Harbor, J. M. 1990. 'A Discussion of Hirano and Aniya's (1988, 1989) Explanation of Glacial-Valley Cross Profile Development', *Earth Surface Processes and Landforms*, **15**, 369-377.

Harbor, J. M. 1992. 'On the Mathematical Description of Glaciated Valley Cross Sections', *Earth Surface Processes and Landforms*, **17**, 477-485.

Hirano, M. and Aniya, M. 1988. 'A Rational Explanation of Cross-Profile Morphology for Glacial Valleys and of Glacial Valley Development', *Earth Surface Processes and Landforms*, **13**, 707-716.

Hirano, M. and Aniya, M. 1989. 'A Rational Explanation of Cross-Profile Morphology for Glacial Valleys and of Glacial Valley Development: A Further Note', *Earth Surface Processes and Landforms*, **14**, 173-174.

Hurvich, C. M. and Tsai, C. 1989. 'Regression and Time Series Model Selection in Small Samples', *Biometrika*, **76**(2), 297-307.

James, L. A. 1996. 'Polynomial and Power Functions for Glacial Valley Cross-Section Morphology', *Earth Surface Processes and Landforms*, **21**, 413-432.

Li, Y., Liu, G. and Cui, Z. 2001(1). 'Glacial Valley Cross-Profile Morphology, Tian Shan Mountains, China', *Geomorphology*, **38**, 153-166.

Li, Y, Liu, G. and Cui, Z. 2001(2). 'Longitudinal Variations in Cross-Section Morphology along a Glacial Valley: A Case-Study from the Tien Shan, China', *Journal of Glaciology*, **47**(157), 243-250.

McGee, W. J. 1894. 'Glacial Canons', *Journal of Geology*, **2**, 350-364.

Pattyn, F. and Van Heule, W. 1998. 'Power Law or Power Flaw?', *Earth Surface Processes and Landforms*, **23**, 761-767.

Ratkowski, D. A. 1990. *Handbook of Nonlinear Regression Models*, Marcel Decker, New York.

Wheeler, D. A. 1984. 'Using Parabolas to Describe the Cross-Sections of Glaciated Valleys', *Earth Surface Processes and Landforms*, **9**, 391-394.