Tectonic Deformation Estimation using Stream Gradients: Nonparametric Function Estimation from Difference Data

Tim C. Hesterberg
T_Hesterberg@FandM.edu
Department of Mathematics
Franklin & Marshall College

Dorothy J. Merritts
D_Merritts@FandM.edu
Department of Geosciences
Lancaster, PA 17604-3003

Abstract

In 1811-1812 three great (8.0+) earthquakes occurred near New Madrid, Missouri. We estimate coseismic deformation in this area using stream elevation data from topographic maps. Streams have a natural profile, the gradient of which depends on the resistance of underlying sediment and the volume of stream flow. If tectonic processes elevate the upstream end of a segment a different amount than the downstream end, the stream will attempt to return to its natural gradient by incising, aggrading, or altering its sinuousity. This adjustment takes time, so deviations from the natural gradient may indicate geologically recent deformation. We use penalized regression splines to estimate the natural stream profile and the deformation of the ground surface. Estimation of the natural profile and deformation is based on nonparametric regression of the form $y_2 - y_1 = f(x_2) - f(x_1)$.

Keywords: B-splines, Geomorphology, Regression splines, SL index, Tectonics, Tensor product, Ridge Regression.

1 Introduction

A number of different features of the landscape have been used to estimate coseismic deformation that has occurred prior to recent geodetic surveys, including fault-bounded mountain fronts, offset stream channels, or uplifted coastlines along active plate margins. Use of such features has been unsuccessful in continental interiors because the terrain in such areas typically has little relief and faults rarely rupture the surface. For example, our study area is within the New Madrid Seismic Zone (NMSZ), in an area mantled by more than 1 km of alluvial sediment that obscures crustal displacement.

Our purpose in this article is to describe a method of using topographic information for stream networks to identify crustal deformation. Stream channels have natural gradients that primarily depend on resistance to flow (dependent on bedrock or sediment type and properties)

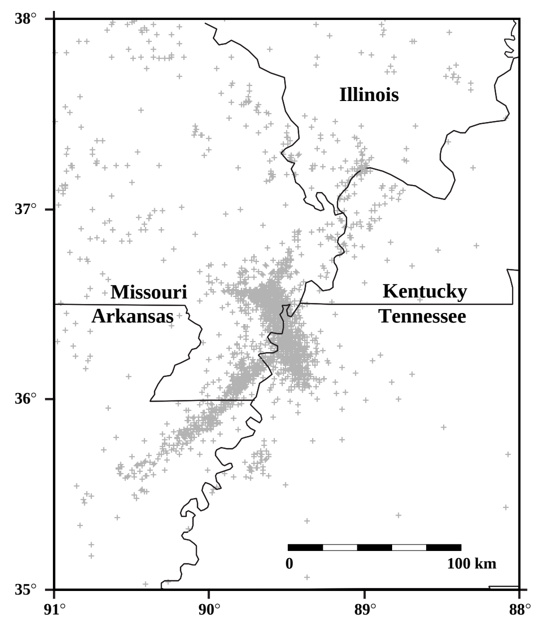

Figure 1: Earthquake epicenters of New Madrid region from 1974 to 1991.

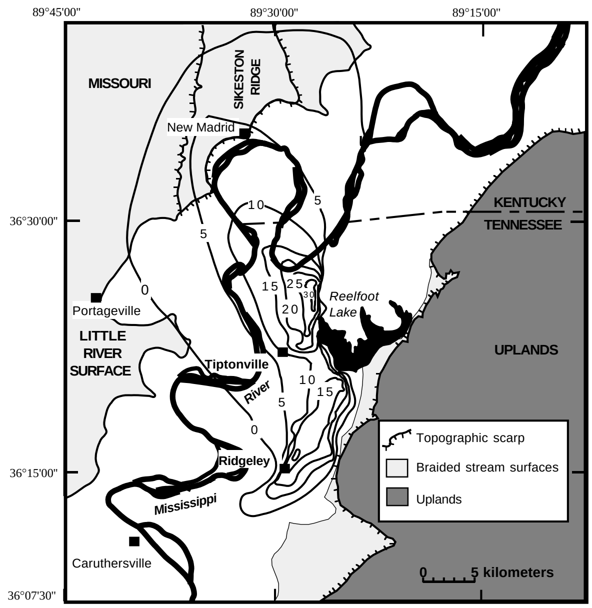

Figure 2: Surface contours (feet) in the Lake County Uplift Region, from Russ (1982).

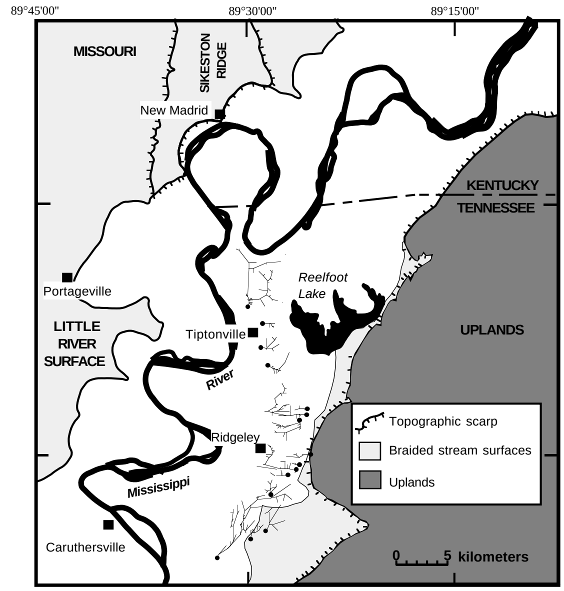

Figure 3: Watersheds analyzed in this study.

and volume of water and sediment flowing in the stream (Schumm, 1977, 1983). The gradient is steeper near the headwaters, and less steep downstream where the flow volume is greater. If tectonic processes change the gradient of an alluvial stream segment by elevating the upstream end of a segment more or less than the downstream end, the stream will attempt to return to its natural gradient, by incising, aggrading, or altering its sinuosity (Ouchi 1985). However, that process takes time, so that deviations of streams from their natural gradients could indicate recent tectonic deformation. Merritts (1987) and Merritts & Vincent (1989) found that stream networks are sensitive to slow rates of surface uplift and tilt, even where faults have not ruptured the surface. The focus of this article is on statistical methodology; further discussion of our results is in Merritts & Hesterberg (1994).

We begin with a description of our study area in Section 2. In Section 3 we describe a procedure for estimating the natural profile of streams in a study area, regardless of tectonic disturbances. In Section 4 we describe how to estimate the natural profile and surface deformation simultaneously. This regression procedure can extract a subtle signal from noisy data.

The procedures in Section 3 and 4 are based on using nonparametric regression from difference data, in which the response variable cannot be observed directly; instead the difference in values of the response variable is observed for pairs of values of the explanatory variable(s).

2 Lake County Uplift Area

In 1811 and 1812 the three largest historic earthquakes (magnitude 8.1-8.3) in the North American continental plate occurred in the NMSZ. Figure 1 shows the location of recent small earthquakes in this area, recorded by St. Louis University and the Center for Earthquake Research and Information; these may be aftershocks of the great earthquakes. Our study area is is the Lake County Uplift Area (LCU), a region within the NMSZ where surface deformation is known to have occurred. Russ (1982) used warped stream levies and other evidence to estimate the uplift pattern shown in Figure 2.

We digitized topographic data for 16 small watersheds that drain the LCU, shown in Figure 3, obtaining map coordinates (x and y), elevation z, and arclength of each segment l. We worked from 1:24,000 scale topographic maps obtained from the U.S. Geological Survey, which have a contour interval of five feet, using custom software *DigiStream* which is available from the authors.

One of the smaller watersheds is shown in Figure 4. There are seven segments in this watershed, several of which have subsegments, where the segment crosses a contour line. Segment 7 is the outlet of the watershed.

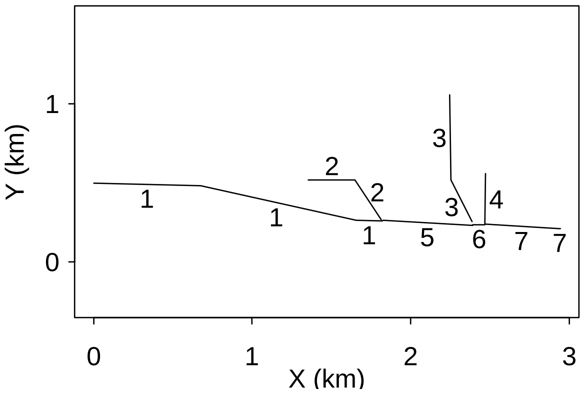

Figure 4: Example watershed with seven segments.

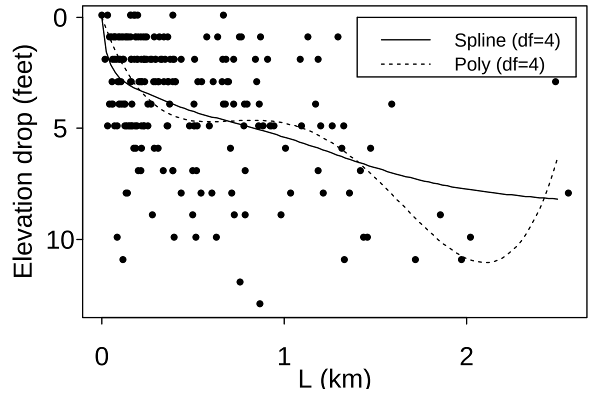

Figure 5: Elevation drops for feeder segments, with two estimates of the natural profile. L is the length of the segment from the head to the end of the current subsegment.

3 Natural Profile Estimation

We describe in this section a procedure for estimating the natural profile of streams in a study area, using nonparametric regression. We assume that the flow at any point on a stream network is a smooth function of (but not necessarily proportional to) the length of all segments upstream of the point. For comparison, the SL index (Hack 1973), the standard procedure for analyzing stream gradients, assumes that flow is a function of the length of the single longest sequence of segments upstream of a point. The SL index also assumes that the profile is given by a parametric curve $z = a - b\log(L)$.

Figure 5 shows a scatterplot of the elevation drop of feeder segments (segments with nothing flowing into them, e.g. segments 1-4 in Figure 4) vs the length from the head of the segment. Each subsegment of a feeder segment is represented, by plotting the elevation drop from the beginning of the segment to the end of the subsegment, vs the length of the stream from the beginning of the segment to the end of the subsegment. It is evident that there is considerable variability in the elevation drops, but nonparametric regression, or scatterplot smoothing, may be used to obtain an estimate of the profile. Two such estimates are shown on the plot. A polynomial of degree 4, fitted by least-squares with no intercept (so the curve passes through the origin), produces an unacceptable estimate of the profile. A much better curve is obtained by a (cubic) smoothing spline with the equivalent of four degrees of freedom; see Hastie & Tibshirani (1990) for a discussion of this and other smoothing procedures, and of equivalent degrees of freedom. See also de Boor (1978) for a more complete discussion of splines. We also perform a square-root transformation on L prior to the smooth, and use an artificial point at the origin with weight 1000 to ensure that the curve passes near the origin.

The curves in Figure 5 are estimated using only feeder

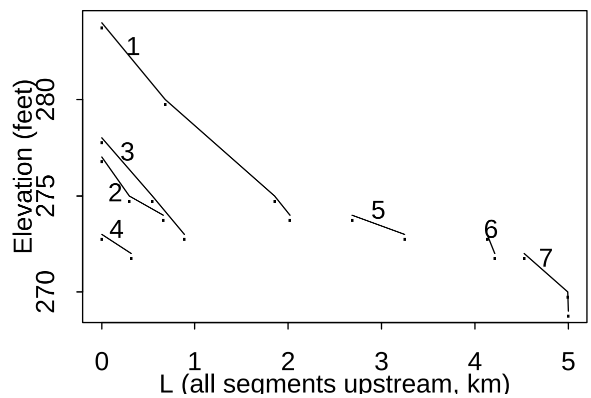

Figure 6: Longitudinal view of all segments in example watershed.

segments. Adding downstream segments results in longitudinal profiles like the one shown in Figure 6. The beginning elevation of segment 5 matches the end elevation of segments 1 and 2, and the value of L at the upstream end of segment 5 matches the sum of lengths of segments 1 and 2. We can move the individual segments, or even subsegments, up or down, without changing the gradients. In Figure 7 the subsegments of the sample watershed are translated vertically so that they begin on the estimated profile from Figure 5. For this particular watershed the early subsegments are less steep, and the later subsegments steeper, than the natural profile estimated using feeder segments from all watersheds.

In order to incorporate information from downstream segments in estimating a natural profile we must abandon simple scatterplot smoothing, because downstream segments consist of pairs of observations which may be translated vertically arbitrarily. The natural profile should reproduce the observed elevation drops as closely as possible, rather than the elevations. We estimate the

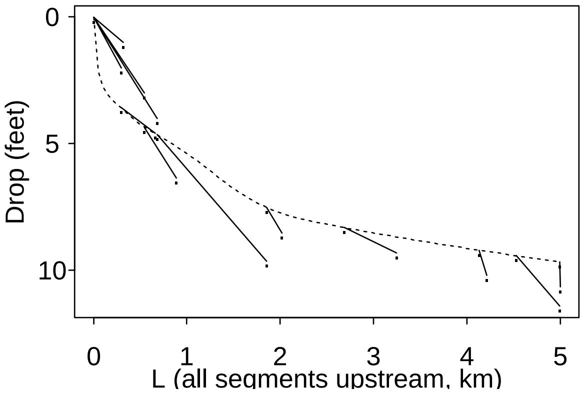

Figure 7: Longitudinal profile of example watershed, curve estimated from feeder segments from all watersheds.

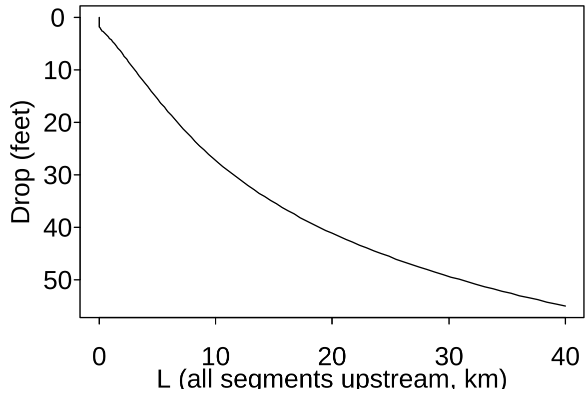

Figure 8: Estimated natural profile, using all segments and nonparametric estimation on difference data.

profile f using a model of the form

$$\mathrm{drop} = z_1 - z_2 = f(L_1) - f(L_2) + \epsilon \qquad (1)$$

where f is a smooth nonparametric function (with $f(0) = 0$), L_1 and L_2 are the values of L at the upstream and downstream ends of the subsegment, and z_1 and z_2 are the elevations at the ends of the subsegment.

One way to fit this model would be to include each subsegment in the regression as two observations, one for the upstream and one for the downstream end, with a single dummy variable to allow for vertical translation, giving a model of the form

$$z_{(i,j)} = f(L_{(i,j)}) + \sum_{k=1}^{n} \beta_k D_k + \epsilon \qquad (2)$$

where $j = 1, 2$ for the upstream and downstream ends of a subsegment, and D_i is a dummy variable which is one for observations $(i, 1)$ and $(i, 2)$ and zero otherwise, $i = 1, \ldots, n$. In a dataset with n subsegments, this would require a regression on $2n$ observations using n dummy variables. This would be computationally expensive to fit using a standard package, but might be feasible using a backfitting procedure (Hastie & Tibshirani 1990).

Our approach is to use regression splines, in particular B-splines. We use cubic B-splines s_j, which form a basis for the set of cubic splines, i.e. any cubic spline can be represented in the form

$$f(x) = \beta_0 + \sum_{j=1}^{J} \beta_j s_j(x). \qquad (3)$$

Regression splines are described in Hastie & Tibshirani (1992), and implemented in *S-Plus* (Chambers & Hastie 1992) as well as in other software such as *Matlab*. We choose the degrees of freedom J and the knots in advance, and transform L prior to computing the splines, e.g. by a square root transformation. Then (1) reduces to a regression of the form

$$z_1 - z_2 = \sum_{j=1}^{J} \beta_j (s_j(\sqrt{L_1}) - s_j(\sqrt{L_2})) + \epsilon \qquad (4)$$

which can easily be estimated by ordinary linear regression; the (i, j) element of the design matrix is the difference $s_j(\sqrt{L_{(i,1)}}) - s_j(\sqrt{L_{(i,2)}})$. There is no intercept in the regression. For comparison, in fitting a polynomial $f(x) = \sum_{j=0}^{J} \beta_j x^j$ to difference data, the regression would be of the form

$$\Delta z \equiv z_1 - z_2 = \sum_{j=1}^{J} \beta_j (x_2^j - x_1^j) + \epsilon \qquad (5)$$

We also use a dummy variable, which is 1 for feeder subsegments and 0 for other subsegments. This allows the estimated curve to drop sharply near $L = 0$. The resulting curve is shown in Figure 8. The gradient is very steep at first, then gradually levels off.

4 Deformation Estimation

Deformation of the earths surface during the 1811-1812 earthquakes caused the upstream ends of some segments to be elevated a different amount than the downstream ends, so those segments became more or less steep than their natural profile. We assume that streams have not had time to adjust their channels to return to their natural profile; this is more likely to be true for smaller streams, such as we analyze, than for larger streams. Then the actual drop of any segment should consist of the drop determined by the natural profile, plus the differential elevation gain from tectonic deformation. Our model for this is

$$\Delta z = f(L_1) - f(L_2) + g(x_1, y_1) - g(x_2, y_2) + \epsilon \qquad (6)$$

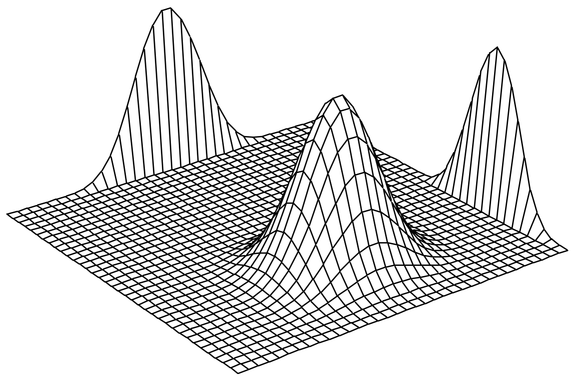

Figure 9: Tensor product B-splines. The product spline is $g_{m,k}(x,y) = s_m(x)S_k(y)$.

where (x_1, y_1) and (x_2, y_2) are the coordinates (longitude and latitude) of the upstream and downstream ends of a segment, and g represents the deformation surface. We represent g using tensor product splines — if $s_m(x), m = 1, \ldots, M$ and $S_k(y), k = 1, \ldots, K$ are B-splines in x and y, then the tensor product splines are $g_{m,k}(x,y) = s_m(x)S_k(y)$. One such product spline is shown in Figure 9. In addition, because B-splines are normally used in a regression with an intercept, the terms $g_{m,0}(x,y) = s_m(x)$ and $g_{0,k}(x,y) = S_k(y)$ should also be used in a regression. Then g is of the form

$$g(x,y) = \beta_0 + \sum_{m,k} \beta_{m,k} s_m(x) S_k(y) \qquad (7)$$

where the double sum is for $m = 0, \ldots, M$ and $k = 0, \ldots, K$ with the exception of $m = 0$, $k = 0$. Incorporating this into the regression results in a model of the form

$$\begin{aligned} \Delta z &= \textstyle\sum_{j=1}^{J} \beta_j (s_j(\sqrt{L_1}) - s_j(\sqrt{L_2})) + \eta I(\text{feeder}) \\ &\quad + \textstyle\sum_{m,k} \beta_{m,k} (g_{m,k}(x_1,y_1) - g_{m,k}(x_2,y_2)) + \epsilon \end{aligned} \qquad (8)$$

There are some difficulties with this procedure. If the knots in x and y are placed in a reasonably fine grid then the number of product splines, and degrees of freedom in the regression, can be very large. There may be problems with colinearity or near-colinearity. Indeed, because the streams are not located uniformly on a rectangular grid, some of the product splines may be zero for all observed segment ends, in particular those product splines which have their peak near the corners of the grid. There are two add-hoc remedies and a general remedy for these problems.

The first ad-hoc remedy is that product splines which are zero on observed coordinates may be omitted from the regression. The second is to transform the coordinates so that the segment ends more nearly fill out a rectangle on the transformed coordinate system. Rotations would be useful if a study area has a long axis which lies in other than a north-south or east-west direction. Our streams have a long axis in the north-south direction already, but they do follow a banana shape rather than a rectangular shape. We use a transformation of the form $x^* = x + ay + by^2$, with a and b chosen to eliminate the curve of the banana.

The more general remedy is to use ridge regression. Instead of ordinary least squares, the parameters in (8) can be chosen to minimize

$$\sum_{i=1}^{n} \hat{\epsilon}_i^2 + \lambda \sum_{m,k} (\beta_{m,k})^2, \qquad (9)$$

where λ is a penalty factor. The amount of penalty can be varied, to produce smoother and flatter surfaces at one extreme, or unflattened and unstable surfaces at the other. We have chosen a λ that appears to give reasonably smooth estimates of deformation; the result is shown in a contour plot in Figure 10. Uplift is generally greatest along the middle (from west to east) of the study area, with low areas about one-quarter and one-half the distance from north to south. The pattern exhibits four distinct high areas, three of which generally coincide with areas of uplift determined by Russ (1982). However, fine details might result from the inherent randomness in stream channel elevations or the limitations of working from maps with 5-foot contour intervals in a region of very little relief. The contours are less accurate on the edges of the study area. The overall amount of uplift shown is less than determined by Russ (1982), probably because of the flattening effect of the ridge regression.

We are investigating a more general ridge regression procedure, in which the penalty term would be based on the smoothness of the deformation estimate. This criterion is motivated by the energy criterion that determines thin-plate splines. The penalty can be expressed as a quadratic function function of the regression coefficients, $\sum_u \sum_v c_{u,v} \beta_u \beta_v$, for some constants c, where u and u represent pairs of indices (m,k).

5 Summary and Conclusions

A bootstrap analysis (not described here) shows that the deformation estimate is highly statistically significant. Comparison of the deformation estimate with the topography shows that the deformation estimation procedure is not simply reproducing the contours of the land. Thus it appears that the procedure is finding some real effect in the landscape, and the agreement with the contours of Russ suggests that it is the recent tectonic deformation. We are in the process of testing the method in other study areas.

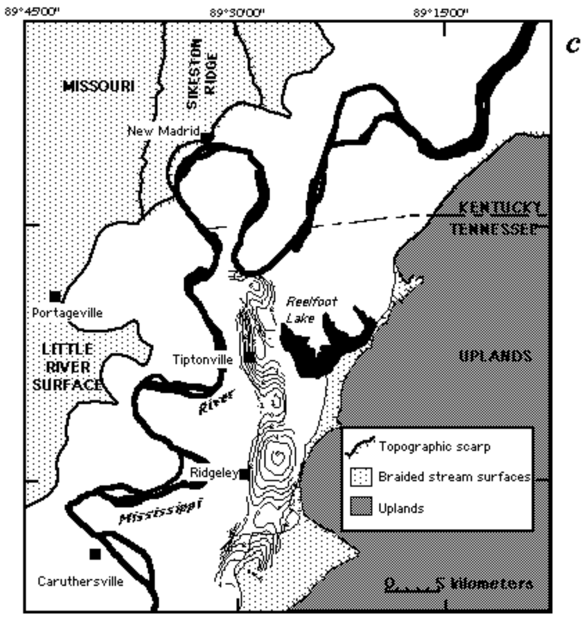

Figure 10: Estimate of surface deformation g (feet).

This procedure can be extended in a number of ways. Differential flow resistance could be estimated by adding a spatial term to the model (6), of the form $(L_2 - L_1)h(x, y)$. Because the procedures described here are based on linear regression they can be made robust in the usual manner, using iterative weighted least squares to reduce the weights on observations with large residuals.

Acknowledgments

We wish to thank Harry Haddon for creating the *DigiStream* software used during the digitization process, and Denise Muriceak, Ed Stoner, Vasilios Lambrinos, and Spiro Gavaris for collecting the data and other assistance, T. Hastie for suggesting regression splines for estimating the natural profile of streams, and several geologists with expertise in the NMSZ for advice, suggestions, and information, including R. Wheeler, E. Schweig, M. Ellis, P. Bodin, and A. Crone. This work was supported by U.S. Geological Survey under award numbers 1408-93-G-2230 and 1434-95-G-2624, and by the National Science Foundation under award number EAR-9418681.

References

R. A. Becker, J. M. Chambers, and A. R. Wilks (1988), *The New S Language*, Pacific Grove, CA: Wadsworth and Brooks/Cole.

Chambers, J. M. and T. J. Hastie (1992), *Statistical Models in S*, Pacific Grove, CA: Wadsworth and Brooks/Cole.

de Boor, C. (1978), *A Practical Guide to Splines,* New York: Springer Verlag.

Hack, J. T. (1978), "Stream-Profile analysis and stream-gradient index," *U.S. Geological Survey Journal of Research" v. 1, p. 421-429.*

Hastie, T. J. and R. J. Tibshirani (1990), Generalized Additive Models *Chapman & Hall, New York, NY.*

Merritts, D. (1987), Geomorphic response to late Quaternary tectonism, coastal northern California, Mendocino triple junction region," Ph.D. Thesis, *Univ. of AZ, Tucson.*

Merritts, D. and K. R. Vincent (1989), "Geomorphic response of coastal streams to low, intermediate, and high rates of uplift, Mendocino triple junction region, northern California," Geol. Soc. Am. Bull. *101, 1373.*

Russ, D. P. (1982), in Investigations of the New Madrid, Missouri, earthquake region , *F. A. McKeown and L. C. Pakiser, Ed., U.S. Geol. Surv. Prof. Pap., 1236,95-113 (1982)*

Schumm, S. A. (1977), The fluvial system *Wiley, New York.*

S. A. Schumm (1983), "Alluvial river response to active tectonics," in Active Tectonics, *National Academy of Sciences Press, Washington, DC, p. 80-94.*

Statistical Sciences Inc. (1991); S-PLUS Reference Manual, *Version 3.0, Seattle: Statistical Sciences, Inc.*