Ecology, 79(2), 1998, pp. 448–460

INFERRING ECOLOGICAL RELATIONSHIPS FROM THE EDGES OF SCATTER DIAGRAMS: COMPARISON OF REGRESSION TECHNIQUES

FREDERICK S. SCHARF,[1,4] FRANCIS JUANES,[1,2] AND MICHAEL SUTHERLAND[3]

[1]*Department of Forestry and Wildlife Management, University of Massachusetts, Amherst, Massachusetts 01003 USA*
[2]*Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts, Amherst, Massachusetts 01003 USA*
[3]*Statistical Consulting Center, University of Massachusetts, Amherst, Massachusetts 01003 USA*

Abstract. Scatter diagrams have historically proved useful in the study of associative relationships in ecology. Several important ecological questions involve correlations between variables resulting in polygonal shapes. Two examples that have received considerable attention are patterns between prey size and predator size in animal populations and the relationship between animal abundance and body size. Each is typically illustrated using scatter diagrams with upper and lower boundaries of response variables often changing at different rates with changes in the independent variables. Despite recent statistical contributions that have stimulated an interest in characterizing the limits of a variable, a consensus on an appropriate methodology to quantify the boundaries of scatter diagrams has not yet been achieved. We tested regression techniques based on least squares and least absolute values models using several independent data sets on prey length and predator length for piscivorous fishes and compared estimated slopes for consistency. Our results indicated that least squares regression techniques were particularly sensitive to outlying *y* values and irregularities in the distribution of observations, and that they frequently produced inconsistent estimates of slope for upper and lower bounds. In contrast, quantile regression techniques based on least absolute values models appeared robust to outlying *y* values and sparseness within data sets, while providing consistent estimates of upper and lower bound slopes. Moreover, the use of quantile regression eliminated the need for an excess of arbitrary decision-making on the part of the investigator. We recommend quantile regression as an improvement to currently available techniques used to examine potential ecological relationships dependent upon quantitative information on the boundaries of polygonal relationships.

Key words: animal abundance–size relationships; prey–predator size relationships; quantile regression; regression techniques; scatter diagrams.

INTRODUCTION

The ability to make quantitative predictions based on the interdependence of two variables is a central theme in ecology. When examining bivariate data, scatter diagrams illustrate associative relationships and provide a graphical representation of variation and have historically proven useful for exploring a diversity of ecological phenomena spanning several disciplines (Beverton 1962, Begon and Mortimer 1986, Lawton 1989). However, potentially useful information obtainable from the edges of scatter diagrams is often neglected. The frequent disregard for determining quantitatively the magnitude of the boundaries of the relationship between two variables stems from a lack of consensus on an appropriate statistical method and the tendency for ecologists to focus primarily on average relationships. Recently, however, there has been a growing interest among ecologists in addressing biological questions involving the limits or extremes in a variable. For example, Gaines and Denny (1993) present an alternative body of statistics to test hypotheses regarding extreme values of probability distributions. In spite of recent statistical contributions and an awareness of the potential ecological utility contained in knowledge on the limits of a variable, techniques to evaluate the edges of scatter diagrams and understand more completely the information they provide have received little attention (but see Maller et al. 1983, Blackburn et al. 1992).

Many important ecological questions deal specifically with two interdependent variables that exhibit polygonal relationships. Two examples have been especially prominent in the literature: patterns in body size between prey and predators and the relationship between animal abundance and body size. Both relationships center around large scale patterns in animal assemblages and can incorporate broad taxonomic ranges. Typically, least squares regression applications

Manuscript received 26 June 1996; revised 18 February 1997; accepted 11 March 1997.

[4] Present address: Texas Parks and Wildlife Department, Seabrook Marine Laboratory, P.O. Box 8, Seabrook, Texas 77586 USA.

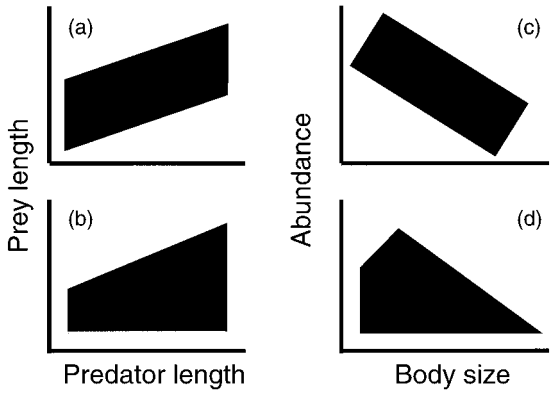

FIG. 1. Examples of typically observed patterns illustrating the relationship between (a and b) predator size and prey size, and (c and d) abundance and body size for animal

have been used to provide quantitative estimates of central tendency for various data sets collected to examine these associations, whereas the boundaries of the relationships are predominately estimated by eye (but see Blackburn et al. 1992).

When examining predator–prey interactions, scatter diagrams have been used to link prey and predator body size and to illustrate ontogenetic patterns in prey size use for a variety of predator taxa (Schoener and Gorman 1968, Paine 1976, Vezina 1985, Cohen et al. 1993). For animal taxa, regression analyses normally indicate an increase in mean prey size with predator size. However, mean prey size does not usually increase proportionately with predator size and substantial variability exists in most cases (Cohen et al. 1993). Disproportionate increases in prey size with increasing predator size can lead to a variety of polygonal patterns of prey size and predator size (Fig. 1a and 1b). Similarly, numerous studies of various taxa and size ranges have used scatter diagrams to characterize the relationship between animal abundance and body size (Peters and Wassenberg 1983, Juanes 1986, Nee et al. 1991, Blackburn and Gaston 1994, Ebenman et al. 1995). Typically, animal abundance declines with increasing body size, although investigations to date reveal several asymmetric distribution patterns resulting in polygonal shapes (Lawton 1989) (Fig. 1c and d).

The observed patterns between prey and predator body sizes and animal abundance and body size suggest that the upper and lower limits of these relationships may often change at different rates. The existence of various ecological patterns with disparate edges warrants separate examination of upper and lower bounds; however, a refined technique to accomplish these goals is not in widespread use. Simply defining the boundaries using the observed extremes in the data is unsatisfactory because there are commonly a limited number of observations near the edges, often making them indiscernible by eye alone. Moreover, the presence of outlying values may result in considerable variation around the true edge. Based on the considerable interest directed at understanding these and other ecological relationships and the fact that hypothesis testing requires measures of variance coupled with parameter estimates, it is evident that a consensus on an appropriate statistical technique used to estimate the edges of scatter diagrams is needed.

In this paper we extend a least squares (LS) regression technique originally described by Blackburn et al. (1992) to estimate slopes of lower as well as upper bounds of scatter diagrams. We apply several versions of this LS regression technique and compare results for consistency. We also describe and apply a quantile regression technique based on a least absolute values model. Comparisons are made between the two regression techniques for validity of slope estimates. Prey size–predator size data for piscivorous fishes are used to test the appropriateness of each regression technique. It will be shown, however, that the described techniques are widely applicable to other types of data and may prove valuable when examining a variety of ecological questions.

METHODS

Data sets

To identify the strengths and weaknesses of two regression techniques in estimating the slopes of the upper and lower bounds of scatter diagrams, we selected six independent data sets from a larger body of data illustrating the relationship between the body sizes of predators and prey. Observations within each data set consisted of individual piscivorous fish predators and individual prey and their respective body lengths for a specific predator population. A diverse group of fish predators was chosen with individual species representing freshwater habitats (tigerfish *Hydrocynus brevis*), marine pelagic environments (bluefish *Pomatomus saltatrix*), as well as benthic (winter skate *Raja ocellata*) and demersal (red hake *Urophycis chuss*; white hake *Urophycis tenuis*; and European hake *Merluccius merluccius*) marine habitats. The six data sets used in these analyses were selected to provide sufficient variation among species assemblages as to the numbers of observations contained in the data set (e.g., red hake $n = 59$, winter skate $n = 701$), the presence or absence of extreme values in the y direction (e.g., winter skate and the three hakes), and the irregularity or sparseness present in the data set (e.g., bluefish and European hake). We postulated that variation in the characteristics present across data sets would be necessary to provide an appropriate setting in which to compare results of regression techniques for consistency.

Each data set was initially tested for a significant relationship between predator length and prey length using both ordinary least squares regression analysis and a quantile regression technique estimating the me-

TABLE 1. Description of the four versions of the least squares regression technique.

Observations used to fit least squares regressions	Methods of data partitioning: Equal numbers of observations	Methods of data partitioning: Equal increments of the independent variable (predator length in mm)
Minimum or maximum y value paired with corresponding x value	$N_{(x,y)}$	$Inc_{(x,y)}$
Minimum or maximum y value paired with median or midpoint x value	$N_{(med\text{-}x,y)}$	$Inc_{(mid\text{-}x,y)}$

dian (i.e., 50th quantile). Each data set exhibited a highly significant positive slope for each regression ($P < 0.0001$), with the exception of the red hake data set (LS $P = 0.135$; quantile regression $P = 0.155$). To confirm the existence of polygonal relationships between the body sizes of predators and prey examined in this study, each data set was then tested for heteroscedastic distribution of errors using two tests based on methods used by Terrell et al. (1996). In each test, residuals generated from a LS regression through all observations were used as the dependent variable with predator length as the independent variable in an ordinary LS regression (Glejser 1969, Harvey 1976). For each test, a significant, positive estimate of the slope would be evidence of heteroscedastic error distribution. For the six data sets, eleven of twelve tests (two tests per data set) produced significant, positive estimates of slope (exception was test 1 for red hake; $P = 0.20$). These results indicate significant heteroscedastic distribution of errors within each data set (with the exception of red hake), with variance increasing as a function of predator size. The distribution patterns contained in these data sets are commonly observed when examining prey size–predator size relationships of piscivorous fishes, with maximum prey size increasing with predator size while minimum prey size remains relatively constant. Therefore, we expected lower bound slope estimates to be modest or statistically insignificant for each data set.

Least squares regression technique

Blackburn et al. (1992) introduced a LS regression technique to estimate the slopes of the upper bounds of animal abundance vs. body size scatter plots. Observations were grouped into a number of equal (n) size classes and the uppermost points within each size class were used to fit a LS regression model estimating the slope of the upper bound. The authors concluded that this approach was adequate for estimating the slopes of the boundaries of polygonal relationships.

Four versions of the LS regression technique were applied (Table 1). Versions differed in the methods applied to partition the data into a number of equal size classes and in the choice of x and y values used to estimate the regression slope. Data were either separated into size classes containing an equal number of observations or divided into size classes representing equal increments of the independent variable (predator length measured in millimeters). LS regression slopes were then estimated by pairing the minimum or maximum prey length (y) with either its corresponding predator length (x) or with the median or midpoint predator length (x) within each size class. Regressions were weighted by the number of observations within each size class for versions with size classes representing equal increments of the independent variable. The four versions of the LS regression technique were applied to the six data sets. Estimates of upper and lower bound slope were tested for homogeneity (Sokal and Rohlf 1995) across different versions for each predator–prey assemblage.

The question of how many size classes to use during application of the LS regression technique has been previously addressed (Blackburn et al. 1992). Based on repeated analysis of simulated data, Blackburn et al. (1992) found that estimates of upper bound slope converged toward the mean regression slope with increasing number of size classes and became highly variable when too few size classes were used. They suggested that between six and fifteen size classes was adequate. We chose to divide data into eight size classes for all analyses, regardless of the methods of data partitioning. To examine the effect of size class number on estimates of LS regression slope, we partitioned data from four predator–prey assemblages into size classes containing an equal number of observations and estimated regression slopes using the minimum y value paired with its corresponding x value within each size class (as in version $N_{(x,y)}$). Estimates of upper and lower bound slope were generated using a range of size class numbers between 10 and 50 and observed trends examined. To increase the overall diversity of predators examined in this study, the data sets used to analyze the influence of size class number on slope estimates consisted of predator–prey body length information for four species of piscivorous fish separate from the six species used in the slope comparison analyses. The predator group is diverse and includes a freshwater piscivore (walleye *Stizostedion vitreum*), a predatory reef species (coral trout *Plectropomus leopardus*), a

marine, demersal piscivore (Atlantic cod *Gadus morhua*) and an elasmobranch (spiny dogfish *Squalus acanthias*).

Quantile regression technique

A quantile regression technique (StataCorp. 1995) based on least absolute values regression was applied to the same six predator–prey data sets. In least absolute values regression, the regression model is fit by minimizing the sum of the absolute values of the residuals (Harris 1950, Bloomfield and Steiger 1980, Rousseeuw and Leroy 1987) rather than the sum of the squares of the residuals as in ordinary LS regression models. Regression models based on least absolute values criteria are resistant to extreme outlying values in the y direction (Narula and Wellington 1982, Rousseeuw and Leroy 1987) to the extent that model fit is unaffected by changes in y values so long as the signs of the residuals are maintained (Bloomfield and Steiger 1983). Moreover, regression models that minimize the sum of absolute deviations are particularly appropriate for estimating nonparametric measures of location (e.g., medians, quantiles) (Koenker and Bassett 1978).

The quantile-based regression technique used in this study estimates quantiles of the dependent variable ranging from 0 to 100, conditional on the values of the independent variable (StataCorp. 1995). Quantile regression estimates are obtained through minimization of the following quantity:

$$\sum_i \left| y_i - \sum_j \beta_j x_{ij} \right| h_i$$

where h_i is a multiplier equal to the chosen quantile value (i.e., 0.50 for the median) if the residual between the absolute value symbols is positive or one minus the chosen quantile value if otherwise. A method of Koenker and Bassett (1982) and Rogers (1993) is generally used to estimate the variance–covariance matrix of the coefficients and generate estimates of regression coefficient standard errors. However, this method tends to underestimate standard errors for data sets with heteroscedastic error distributions (Rogers 1992). Therefore, for this study, a bootstrap resampling procedure provided by Gould (1992) was employed to estimate standard errors of quantile regression coefficients. When the bootstrap resampling procedure is used, only estimates of standard error and significance levels are affected, with estimates of quantile regression coefficients remaining unchanged. Twenty replications were used for each application of the quantile regression technique, as this has been shown to be sufficient for cases of heteroscedastic error distributions (StataCorp. 1995).

Unlike the LS regression technique, no data partitioning is necessary when applying the quantile regression technique; quantile regression models are fit to the original unmodified data. Therefore choices regarding number of size classes, methods of data partitioning, and appropriate x and y values used to fit the regression model are eliminated. However, the decision as to which quantiles best represent the boundaries of the data is an arbitrary one and must be made by the investigator. For this reason, we examined upper and lower bound slope estimates generated by the quantile regression technique using three different pairs of quantiles to represent the edges of the data; the 10th and 90th quantiles, the 5th and 95th quantiles, and the 1st and 99th quantiles. Estimates of lower bound (1st, 5th, and 10th quantiles) and upper bound (90th, 95th, and 99th quantiles) slope were then tested for homogeneity for each predator–prey assemblage. Estimates of slopes generated by the quantile regression technique were then tested for homogeneity against estimates of slopes generated by the LS regression technique.

RESULTS

Least squares regression technique

Results of application of the LS regression procedure were highly variable across versions within each data set (Table 2). Significant slope estimates for the boundaries of the data often were not detected for data sets with significant mean relationships. Moreover, versions that produced significant estimates of lower bound slope did not always generate a significant estimate of upper bound slope, although the mean relationships within the data sets were significantly positive.

All four versions yielded nonsignificant slope estimates for minimum prey length consumed by tigerfish, whereas each version produced significant estimates of upper bound slope that did not differ from one another ($F_{(3, 113)} = 0.050$; $P > 0.75$) (Fig. 2a). No significant regressions were generated by any version for either minimum or maximum prey length consumed by winter skate (Fig. 2b). The relationship between minimum prey size and bluefish size was significant for all but version $N_{(\text{med-}x,y)}$, whereas slope estimates for maximum prey size consumed were significant for all but version $\text{Inc}_{(x,y)}$ (Fig. 2c). Significant estimates of lower bound slope for bluefish were indistinguishable from one another ($F_{(2, 295)}$) = 0.039; $P > 0.75$), however, significant differences were detected among estimates of upper bound slope ($F_{(2, 295)} = 5.416$; $P < 0.005$), with estimates generated by versions $N_{(\text{med-}x,y)}$ and $\text{Inc}_{(\text{mid-}x,y)}$ being different (post hoc slope comparisons). Two of the four versions produced significant estimates of lower bound slope for European hake that were statistically similar ($F_{(1, 611)} = 0.003$; $P > 0.75$), whereas only one version produced a significant estimate of upper bound slope (Fig. 2d). No significant estimates of upper bound slope and two significant estimates of lower bound slope that were not statistically different ($F_{(1,57)} = 0.015$; $P > 0.75$) were generated for red hake (Fig. 2e). Similarly, for white hake, no significant estimates of upper bound

TABLE 2. Upper and lower bound slope estimates ($\pm s_b$) for the six piscivores described in the text. The four versions are based on modifications made by the authors to the central philosophy of the least squares regression technique described by Blackburn et al. (1992).

Species	Version				Data source
	$N_{(x,y)}$	$N_{(\text{med-}x,y)}$	$\text{Inc}_{(x,y)}$	$\text{Inc}_{(\text{mid-}x,y)}$	
Tigerfish (n = 115)					
Lower bound	NS	NS	NS	NS	Lewis (1974)
Upper bound	0.230 ± 0.054**	0.232 ± 0.050**	0.253 ± 0.059**	0.251 ± 0.055**	
Winter skate (n = 701)					
Lower bound	NS	NS	NS	NS	NMFS (NEFSC)
Upper bound	NS	NS	NS	NS	
Bluefish (n = 297)					
Lower bound	0.106 ± 0.036*	NS	0.097 ± 0.026*	0.108 ± 0.026**	Juanes and Conover (1995)
Upper bound	0.596 ± 0.044***	0.843 ± 0.119***	NS	0.351 ± 0.132*	
European hake (n = 613)					
Lower bound	NS	NS	0.101 ± 0.040*	0.098 ± 0.039*	Guichet (1995)
Upper bound	NS	0.646 ± 0.250*	NS	NS	
Red hake (n = 59)					
Lower bound	NS	NS	0.251 ± 0.084*	0.236 ± 0.087*	NMFS (NEFSC)
Upper bound	NS	NS	NS	NS	
White hake (n = 101)					
Lower bound	NS	NS	NS	0.111 ± 0.045*	NMFS (NEFSC)
Upper bound	NS	NS	NS	NS	

Notes: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS indicates $P > 0.05$. NMFS (NEFSC) = National Marine Fisheries Service (Northeast Fisheries Science Center).

slope and only one significant estimate of lower bound slope were produced (Fig. 2f).

LS regression estimates of upper and lower bound slope did not significantly increase or decrease with increasing number of size classes for four predator–prey assemblages with the exception of a significant increase in estimates of lower bound slope for Atlantic cod ($F_{(3, 134)} = 3.453$; $P < 0.025$); however, obvious trends were evident for each species (Fig. 3). Estimates of upper bound slope showed a decreasing pattern, whereas estimates of lower bound slope tended to increase with increasing numbers of size classes, with each set of slopes showing convergence toward the mean regression slope. Further, probability values of the regression coefficients decreased with increasing number of size classes across species, especially for estimates of lower bound slope (Table 3).

Quantile regression technique

Estimates of slope produced by the quantile regression technique using the two less extreme sets of quantiles (10th and 90th quantiles; and 5th and 95th quantiles) corresponded closely with only two of twelve comparisons yielding statistically significant differences. Differences were detected, however, between estimates of slope for the 10th/90th and 5th/95th quantiles and estimates of slope for the 1st/99th quantiles (i.e., the most extreme quantiles) (Table 4).

For tigerfish, estimates of slope for each of the three quantiles used to represent lower bounds were not significant, whereas each of the three quantiles representing upper bounds produced significant estimates of slope that were statistically indistinguishable ($F_{(2, 113)} = 0.587$; $P > 0.50$) (Fig. 4a). Significant estimates of slope were produced only for the two less extreme sets of quantiles for winter skate (Fig. 4b). Slope estimates were not statistically different between lower bound quantiles ($F_{(2, 699)} = 1.111$; $P > 0.25$) or upper bound quantiles ($F_{(2, 699)} = 0.057$; $P > 0.75$). Bluefish results for estimates of lower bound slope were similar to results for LS versions, with each quantile yielding significant estimates that did not differ from one another ($F_{(2, 245)} = 0.555$; $P > 0.50$) (Fig. 4c). Estimates of upper bound slope for bluefish generated by each quantile were significant and were not statistically different from each other ($F_{(2, 295)} = 0.189$; $P > 0.75$) (Fig. 4c). Results of quantile regression applications to the European hake data set yielded a significant estimate of lower bound slope only for the 10th quantile, whereas estimates of upper bound slope were significant for each quantile with no statistically detectable differences ($F_{(2, 611)} = 1.922$; $P > 0.10$) (Fig. 4d). Only one significant slope estimate was generated for the red hake data set (90th quantile) (Fig. 4e). For the white hake data set, estimates of upper bound slope were significant for each quantile with statistically detectable differences between them ($F_{(2,99)} = 7.654$; $P < 0.001$; post hoc slope comparisons indicate differences between the slope estimate for the 99th quantile and the other two estimates) (Fig. 4f). No significant estimates of lower bound slope were generated for white hake.

Across technique slope comparisons

Statistical differences were not detected between estimates of upper bound slope for tigerfish across both regression techniques ($F_{(6, 113)} = 0.425$; $P > 0.75$). Similarly, estimates of lower bound slope for bluefish were

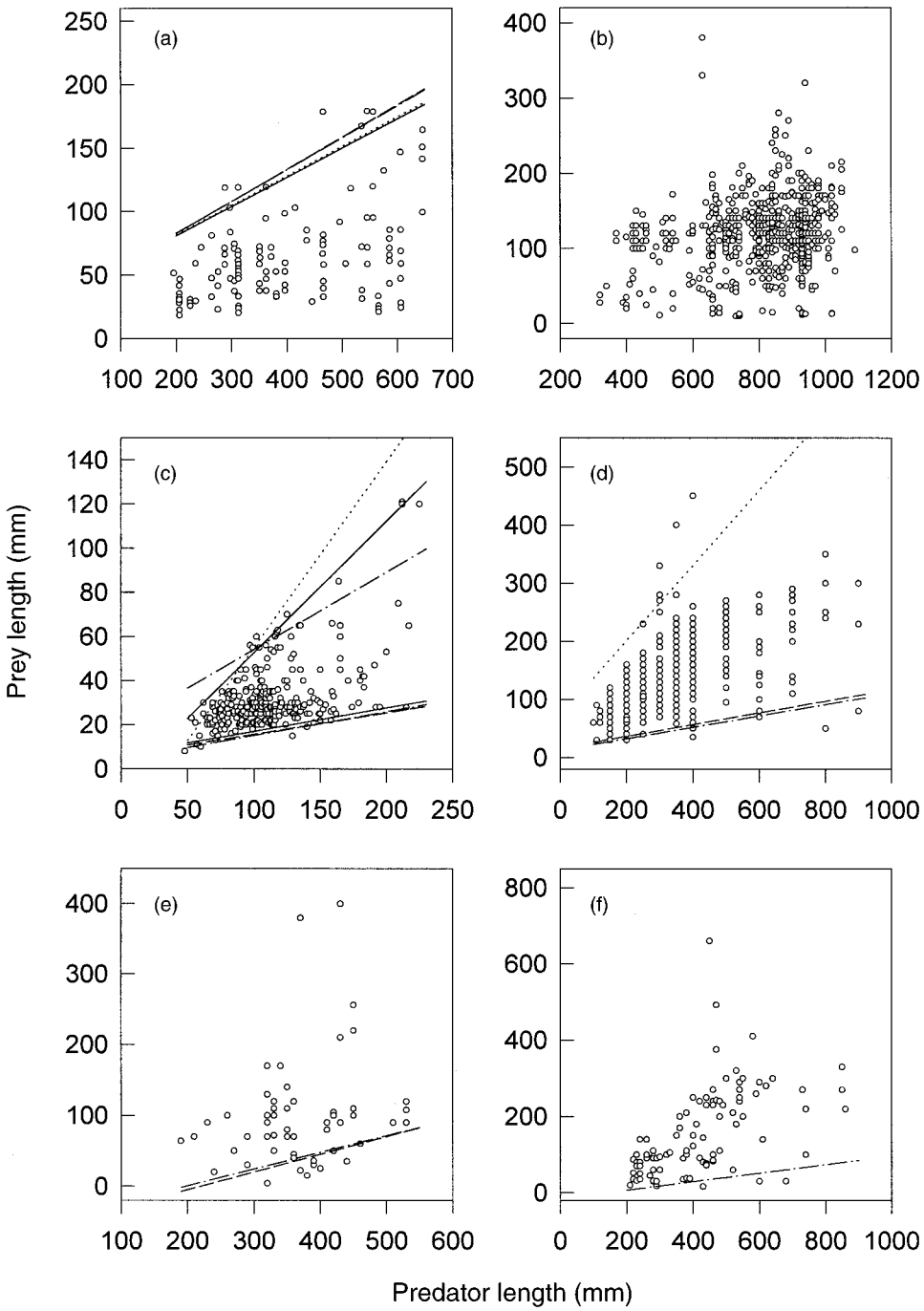

FIG. 2. Plots of piscine prey length vs. predator length for (a) tigerfish, (b) winter skate, (c) bluefish, (d) European hake, (e) red hake, and (f) white hake, illustrating estimates of upper and lower bound slope generated by the LS regression technique. All significant slope estimates are shown for each predator–prey assemblage (e.g., panel b shows no lines because no significant LS slope estimates were generated for the boundaries of the winter skate data set). Each version of the LS regression technique is illustrated using different regression line configurations: solid line, $N_{(x,y)}$; dotted line, $N_{(\text{med-}x,y)}$; dashed line, $\text{Inc}_{(x,y)}$; dashed/dotted line, $\text{Inc}_{(\text{mid-}x,y)}$.

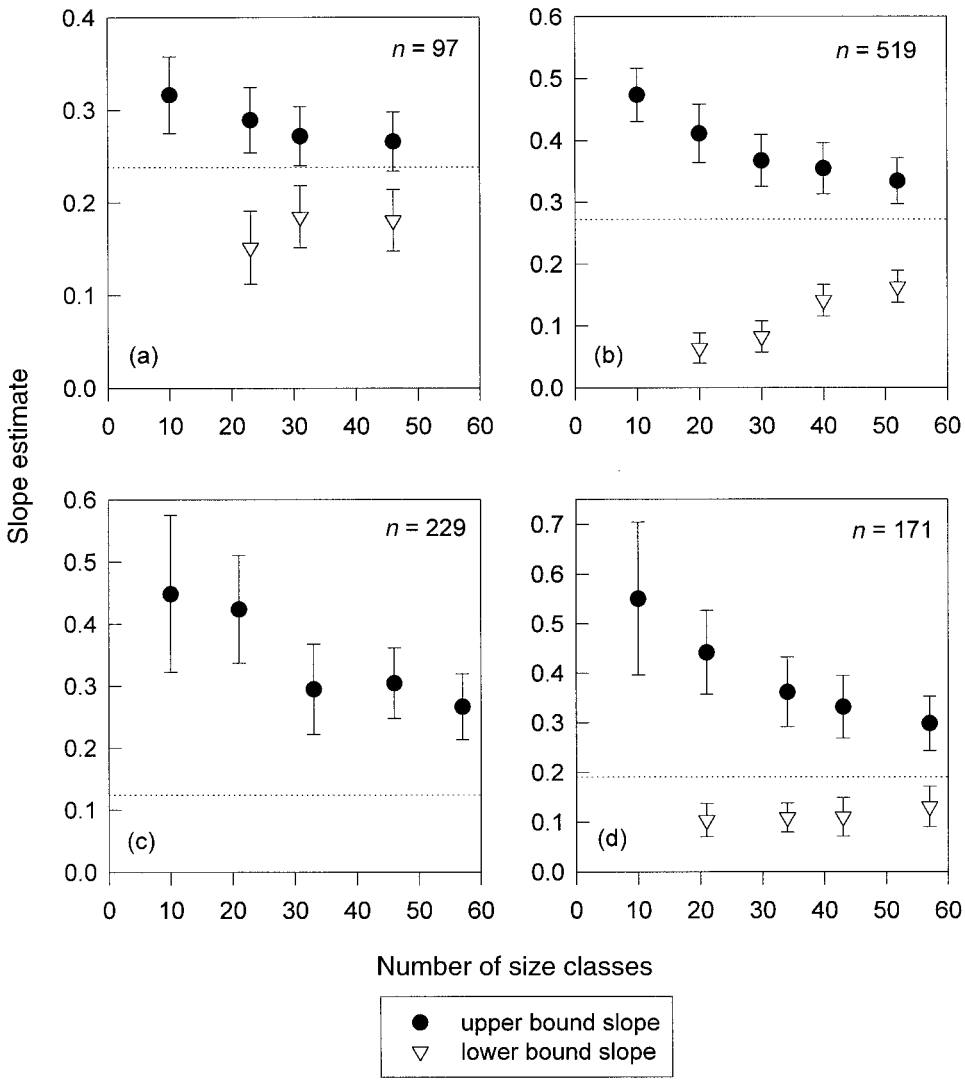

FIG. 3. Plots of upper and lower bound slope estimates ($\pm s_b$) vs. the number of size classes for (a) walleye, (b) Atlantic cod, (c) spiny dogfish, and (d) coral trout. Significant heteroscedasticity of error distributions was detected within each data set. Symbols represent significant slope estimates only. Dashed lines represent slope estimates from mean LS regressions.

TABLE 3. Probability values of regression coefficients generated by the LS regression technique using an increasing number of size classes for four piscivorous fish predators.

No. size classes	Walleye		Atlantic cod		Spiny dogfish		Coral trout	
	lower	upper	lower	upper	lower	upper	lower	upper
10	0.188	0	0.197	0	0.601	0.008	0.100	0.007
20	0.001	0	0.017	0	0.303	0	0.006	0
30	0	0	0.003	0	0.133	0	0.001	0
40	0	0	0	0	0.078	0	0.007	0
50	NA	NA	0	0	0.062	0	0.002	0

Note: A zero value indicates $P < 0.001$. Numbers of size classes listed are approximate. Variation from these numbers was necessary to ensure equal numbers of observations within each size class. NA = insufficient data available to estimate slopes. Walleye data from Nielsen (1980); Atlantic cod and spiny dogfish data from NMFS (NEFSC); coral trout data from Kingsford (1992).

TABLE 4. Slope estimates ($\pm s_b$) generated by the quantile regression technique using three different sets of quantiles to represent upper and lower bounds for the same six species listed in Table 2.

Species	Quantile		
	10th/90th	5th/95th	1st/99th
Tigerfish ($n = 115$)			
Lower bound	NS	NS	NS
Upper bound	0.231 ± 0.070**	0.292 ± 0.084**	0.337 ± 0.050***
Winter skate ($n = 701$)			
Lower bound	0.078 ± 0.017***	0.047 ± 0.024*	NS
Upper bound	0.109 ± 0.015***	0.117 ± 0.030***	NS
Bluefish ($n = 297$)			
Lower bound	0.088 ± 0.018***	0.101 ± 0.020***	0.120 ± 0.026***
Upper bound	0.538 ± 0.035***	0.589 ± 0.039***	0.584 ± 0.099***
European hake ($n = 613$)			
Lower bound	0.121 ± 0.024***	NS	NS
Upper bound	0.350 ± 0.026***	0.400 ± 0.028***	0.947 ± 0.412*
Red hake ($n = 59$)			
Lower bound	NS	NS	NS
Upper bound	0.600 ± 0.256*	NS	NS
White hake ($n = 101$)			
Lower bound	NS	NS	NS
Upper bound	0.727 ± 0.103***	0.844 ± 0.281**	2.476 ± 0.534***

Notes: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS indicates $P > 0.05$.

statistically similar across regression techniques ($F_{(5, 295)} = 0.174$; $P > 0.75$). Significant differences were detected between estimates of upper bound slope for bluefish across regression techniques ($F_{(5, 295)} = 3.221$; $P < 0.01$); however, post hoc tests indicated that these were due to differences stated earlier between versions $N_{(\text{med-}x,y)}$ and $\text{Inc}_{(\text{mid-}x,y)}$ of the LS regression technique. Estimates of lower and upper bound slope for European hake were not significantly different across regression techniques (lower bound $F_{(4, 611)} = 0.127$; $P > 0.75$) (upper bound $F_{(3, 611)} = 1.279$; $P > 0.25$).

DISCUSSION

Ontogenetic increases in prey size range have been observed for several predator taxa (Sabelis 1992, Arnold 1993, Bremigan and Stein 1994). The inclusion of larger prey sizes combined with the retention of small prey items in the diet lead to polygonal predator size–prey size relationships with notable scaling differences between lower and upper bounds. Polygonal patterns of a similar nature are frequently observed when examining the relationship between animal abundance and body size (Damuth 1981, Lawton 1989, Currie 1993). Lawton (1989) implies that questions regarding the overall distribution shape may potentially provide more information than average relationships, whereas Maurer and Brown (1988) suggest that defining the constraints of upper and lower bounds of polygonal relationships should be a focus of future macroecological research. However, because an uncomplicated statistical technique providing unbiased parameter estimates and a measure of error has not been widely accepted or implemented, hypotheses concerned with the potential relationships found in the edges of scatter diagrams have often been ignored. Thomson et al. (1996) propose several potential statistical methodologies to analyze asymmetric distributions, but suggest that more research be directed at the testing and development of new or previously established techniques.

Least squares regression technique

In this study, within species comparisons made across applications of different versions of the LS regression technique yielded a high degree of variability among slope estimates of upper and lower bounds (Table 2). Although only one of five comparisons between significant LS slope estimates produced statistically detectable differences, results were generally inconsistent. One or more LS regression versions often produced significant estimates of upper or lower bound slope, whereas the remaining versions failed to detect a significant relationship between minimum or maximum prey length and predator length. Moreover, LS regression versions that produced significant estimates of lower bound slope frequently failed to produce significant estimates of upper bound slope for the same data set, which is in direct contrast to our expectations. As stated above, we expected lower bound slopes to be modest or not significant with upper bound slopes increasing at a much faster rate based on the positive mean relationships within each data set and the patterns typically observed when examining the feeding ecology of piscivorous fishes.

We observed a declining trend in estimates of upper bound slope as the number of size classes increased. A slight increase in estimates of lower bound slope was

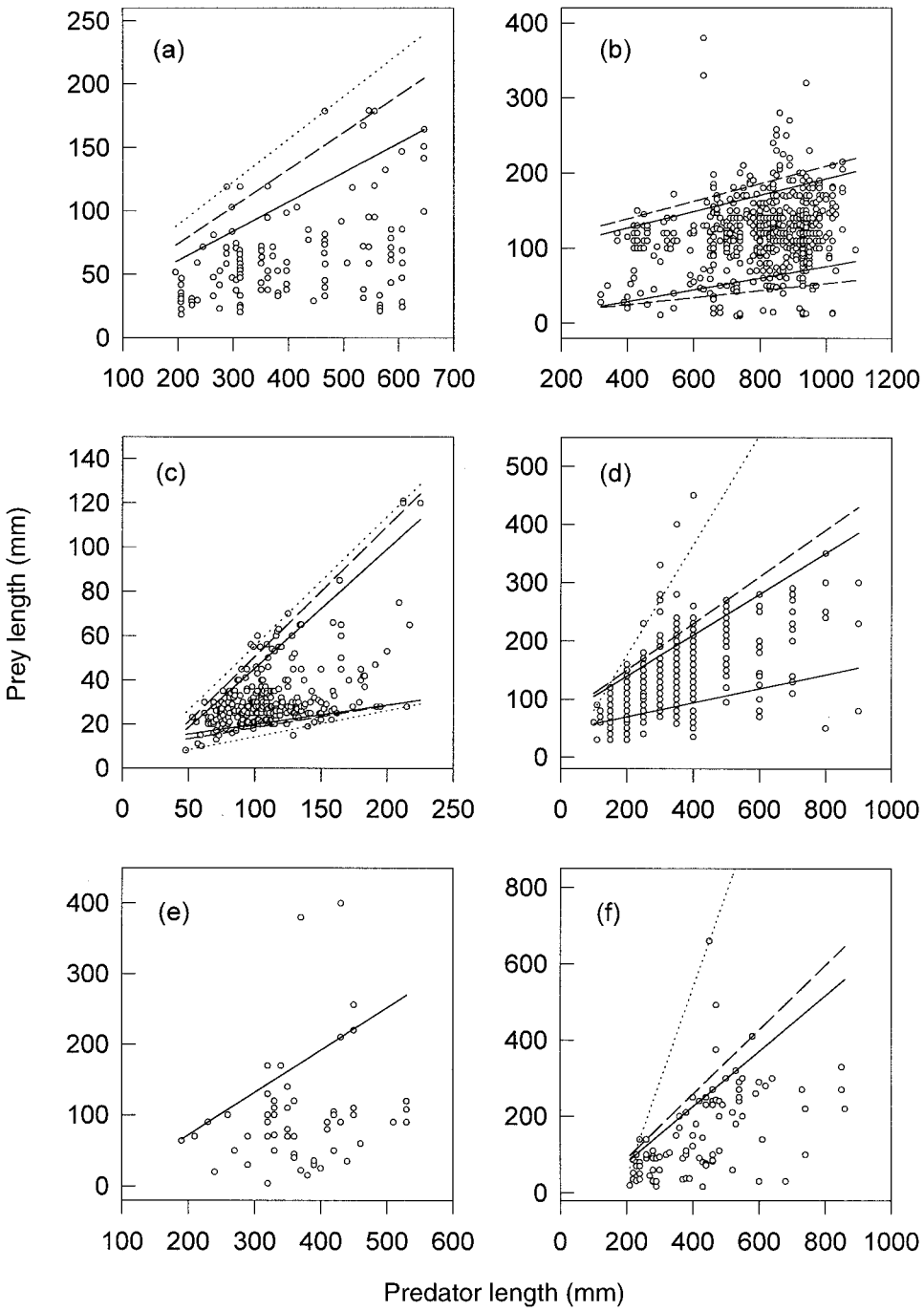

FIG. 4. Plots of piscine prey length vs. predator length for (a) tigerfish, (b) winter skate, (c) bluefish, (d) European hake, (e) red hake, and (f) white hake, illustrating estimates of upper and lower bound slope generated by the quantile regression technique. All significant slope estimates are shown for each predator–prey assemblage. Each quantile estimated by the quantile regression technique is illustrated using different regression line configurations: solid line, 10th and 90th quantiles; dashed line, 5th and 95th quantiles; dotted line, 1st and 99th quantiles.

also evident across data sets. Our results indicate that estimates of both upper and lower bound slope converge toward the mean regression slope as the number of size classes increases, which is in agreement with the findings of Blackburn et al. (1992). In addition, probability values of significance tests tended to decrease as the number of size classes increased. Based on regression principles (Draper and Smith 1981), these results would be expected. When applying the LS technique, more observations are used to estimate the

slopes of upper and lower bounds as size class number increases (Blackburn et al. 1992). The number of size classes is ultimately limited to the number of observations in the data set and any regression generated with this maximum number of size classes is simply a regression through all observations, or the mean regression. Hence, the estimate of a regression slope generated from data partitioned into size classes will approach the estimate of the mean regression slope as size class number increases. Further, as the number of observations used to estimate the regression increases, the amount of variation explained by the regression model should increase. Therefore, increasing sample size would be expected to generate lower probability values for significance tests, as seen here.

The results of this study indicate that estimates of LS regression slopes may depend upon four factors which are subject to arbitrary decision-making by the investigator. The first is the method of data partitioning, which because of the uneven distribution of observations in some data sets (i.e., bluefish, European hake, and white hake) affected both regression significance levels and values of slope estimates. Second is the decision of which *x*-*y* pairs to use in fitting the regression model. This decision may have only a minor influence on results as slope estimates generated by N versions of the LS regression technique were more similar to each other than those generated by Inc versions and vice versa. The third factor involves the number of size classes used. Our results suggest that as the number of size classes increases, estimates of upper and lower bound slope will converge toward the mean regression slope. Hence, decisions regarding size class number may considerably affect regression coefficient values and consequently, conclusions drawn from those values. The fourth factor involves the diagnosis of outliers (Barnett and Lewis 1994) (especially in the *y* direction) and the decision to include or remove them from the regression model. The extreme *y* values contained in the winter skate data set and the three hake data sets appeared to have considerable effects on significance levels and values of slope estimates generated by each version of the LS regression technique. Therefore, decisions concerning outliers could prove critical in estimation of upper and lower bounds of scatter diagrams when employing this technique. When estimating the edges of a scatter diagram with no significant mean or median relationship (i.e., the red hake data set), the LS regression technique was consistent across different versions. No significant estimates of upper bound slope were produced, whereas two versions did generate significant estimates of lower bound slope. Bivariate scatter diagrams should not, however, be discarded if mean or median relationships are not evident, as their edges may still provide information on the limitations imposed by one variable on another (Thomson et al. 1996). In the case of the red hake data set, the edges do not appear to contain any additional information and no clear patterns are present in the existing data.

Quantile regression technique

Regression techniques involving quantiles enjoy a long statistical history within the larger body of regression techniques using absolute deviations which have been successfully employed in various areas of research outside of ecology (Mudrov et al. 1968, Blattberg and Sargent 1971, Taylor 1974). They are a generalization of median regression, a technique of fitting best medians using minimization of sums of absolute deviations first suggested by Boscovich in 1757 and later developed by Laplace (Stuart and Ord 1991). Bloomfield and Steiger (1980) and Koenker and D'Orey (1987) detail the problem of minimizing a sum of absolute deviations as a linear programming problem, which makes application of quantile regression techniques possible on modern personal computers.

The quantile regression technique applied here produced results that were markedly more consistent than those produced by the LS technique (Table 4). There were no appreciable differences between slope estimates generated when using either the 10th and 90th quantiles or the 5th and 95th quantiles, despite the fact that they are different quantiles estimating lines through different regions of the data, so that one would expect slight differences in equations. There were, however, significant differences between results produced when using these two sets of quantiles and results generated when using the 1st and 99th quantiles. Slope estimates for 99th quantile upper bounds were generally either not significant statistically or were much larger than estimates produced for less extreme quantiles (i.e., 90th and 95th) coupled with large error terms. When using the 90th or 95th quantiles to represent upper bounds, the quantile regression technique was insensitive to data set characteristics such as the presence of outlying *y* values (e.g., winter skate, European hake, and white hake data sets) or uneven distribution of observations (e.g., bluefish, European hake, and white hake data sets). However, the presence of outlying *y* values appeared to have a significant effect on estimates of slope for 99th quantile upper bounds. As with the LS regression technique, quantile regression did not detect significant slopes for the edges of a scatter diagram with no underlying mean or median relationship (i.e., the scatter diagram for red hake).

The only decision exposed to the subjectivity of the investigator when applying the quantile regression technique is the choice of quantiles used to represent upper and lower bounds. Recent efforts directed at the estimation of quantile regression coefficient standard errors suggest that extreme quantiles may not be dependable in many situations, especially when data are limited. Rogers (1992) suggests that $n > 5/q$ or $n > 5/(1 - q)$ to ensure reliable estimates of variance for quantile regression coefficients. Our results indicate

that a more conservative approach may be appropriate, such as $n > 10/q$ or $n > 10/(1 - q)$. This paper was focused on the estimation of slopes of the boundaries of scatter diagrams; however, the quantile regression technique will estimate quantiles with values between 0 and 100. For the purposes of examining the internal structure of a scatter plot, Thomson et al. (1996) present a partitioned regression technique that consecutively subdivides the point cloud using major-axis regression. The potential drawback of the partitioned regression technique is that a sufficient number of observations is needed to allow for an adequate number of divisions. With the exception of extreme quantiles, the quantile regression technique is not constrained by data requirements of this sort and may be better suited for illustrating scatter diagram structure. In general, quantile choice will be dependent upon available ecological information, the amount of data collected, and the nature of the research question.

Treatment of more complex relationships

In this paper, the application of both regression techniques was done with the assumption that the edges of the scatter diagrams used were inherently linear and that predator length (i.e., the independent variable) was measured without error. The quantile regression technique should be proficient in dealing with curvilinear edges of scatter diagrams as nonlinear (e.g., logarithmic) transformations of data do not bias coefficient estimates of quantile regression (Bassett 1992). However, coefficient estimates from LS regression technique applications to log-transformed data do not remain unbiased after back transformation (Beauchamp and Olsen 1973). Although the quantile regression methodology as we present it is free of consideration of measurement errors in the independent variables, this mimics the long rich history of classical least squares regression. During the last decade, however, there has been a burst of interest in errors-in-variables regression for least squares, and recently these results are being generalized to other fitting criteria. However, the limited results available for regression using absolute deviations prevent a thorough application of them here.

Further ecological relationships

In addition to relationships between prey and predator sizes and animal abundance and body size, quantile regression techniques should prove valuable in several other areas of ecological research. Recently, the limitations in standing stock of fishes imposed by various habitat variables were assessed using quantile regression (Terrell et al. 1996). However, there are many ecological relationships for which the boundaries are often ignored. For example, much attention has focused on relationships between abundance and distribution of both plant and animal species (Brown 1984, Gotelli and Simberloff 1987, Gaston 1988, Gaston and Lawton 1988). One potential question may be how the number of inhabited locations limits the maximum abundance within locations or vice versa. Currie (1991) illustrates the relationship between energy and species richness using scatter diagrams. It is evident that minimum and maximum values of species richness may change at different rates over the full range of energy values and that it may be useful to identify each separately (i.e., how does annual potential evapotranspiration constrain minimum and maximum species richness?). Another example is the relationship between the number of predator and prey species able to coexist in a given ecological system (Jeffries and Lawton 1985). In addition to examining average ratios of predators to prey, one might ask how the number of prey species limits the maximum number of predator species in a given area. Other areas of ecology that have utilized scatter diagrams and hence may benefit from a further examination of the boundaries of the association include relationships between productivity and soil characteristics in forest ecosystems (O'Neill and DeAngelis 1981), the relationship between animal production and respiration (Humphreys 1979), the dependence of net primary production in terrestrial vegetation upon the length of the growing season (Lieth 1975), and how nest size may physically constrain reproductive success of egg guarding male fishes (Torricelli et al. 1993).

Conclusions

The LS and quantile regression techniques examined in this paper are each capable of estimating the boundaries of scatter diagrams. However, LS approaches as presented here are burdened with several arbitrary decisions regarding data partitioning and numbers of size classes prior to application of regression techniques. These decisions can significantly affect regression estimates of scatter diagram edges. Moreover, LS approaches define the boundaries using the literal extremes in the data, which often consist of only a limited number of observations. The relative sparseness of data near scatter diagram edges coupled with the potential presence of outlying values may often cause LS regression estimates to differ considerably from the true scatter diagrams edges.

In contrast, quantile regression procedures based on least absolute values models eliminate the need for an excess of arbitrary decisions. The choice of a quantile to best represent the edges of a scatter diagram is subjective and will depend on each particular data set. However, we have provided a rather conservative rule relating the number of observations contained in a data set to an appropriate quantile. The rule should make quantile choice decisions less subjective and help avoid estimation of extreme quantiles with insufficient data.

Our results indicate that quantile regression techniques are robust with respect to extreme outlying y values and sparseness contained within data sets relative to regression procedures based on LS models.

Although limitations exist, they clearly represent an improvement over currently available techniques used to estimate quantitatively the edges of bivariate scatter diagrams. Thomson et al. (1996) note that communication of the results of descriptive ecological studies would be improved by commonly accepted statistical methodologies to deal with asymmetric distributions. The results obtained in this study will hopefully stimulate further research to achieve this goal.

ACKNOWLEDGMENTS

We thank Robert Guichet for providing data and the National Marine Fisheries Service (NEFSC) for providing access to the NEFSC Food Chain Dynamics Investigation data base. Special thanks to Rodney Rountree for guidance and initial suggestions regarding the NEFSC data base. We are grateful to Tim Blackburn, John Boreman, Aaron Ellison, Rodney Rountree, Andrew Taylor, and an anonymous reviewer for comments that greatly improved the manuscript. This work was supported by a research grant through the Cooperative Marine Education and Research Program of the National Oceanic and Atmospheric Administration (to F. Juanes). The U.S. Government is authorized to produce and distribute reprints for governmental purposes notwithstanding any copyright notation that may appear hereon. The views expressed herein are those of the authors and do not necessarily reflect the view of NOAA or any of its sub-agencies.

LITERATURE CITED

Arnold, S. J. 1993. Foraging theory and prey-size-predator-size relations in snakes. Pages 87–115 *in* R. A. Seigel and J. T. Collins, editors. Snakes- ecology and behavior. McGraw-Hill, New York, New York, USA.

Barnett, V., and T. Lewis. 1994. Outliers in statistical data. Third edition. Wiley and Sons, New York, New York, USA.

Bassett, G. 1992. The Gauss Markov property for the median. Pages 23–31 *in* Y. Dodge, editor. L_1 statistical analyses. North-Holland, Amsterdam, The Netherlands.

Beauchamp, J. J., and J. S. Olsen. 1973. Corrections for bias in regression estimates after logarithmic transformation. Ecology **54**:1403–1407.

Begon, M., and M. Mortimer. 1986. Population ecology: a unified study of animals and plants. Blackwell Scientific, London, England.

Beverton, R. J. H. 1962. Long-term dynamics of certain North Sea fish populations. Pages 242–259 *in* E. D. LeCren and M. W. Holdgate, editors. The exploitation of natural animal populations. Blackwell Scientific, Oxford, England.

Blackburn, T. M., J. H. Lawton, and J. N. Perry. 1992. A method of estimating the slope of upper bounds of plots of body size and abundance in natural animal assemblages. Oikos **65**:107–112.

Blackburn, T. M., and K. J. Gaston. 1994. Body size and density: the limits to biomass and energy use. Oikos **69**: 336–339.

Blattberg, R., and T. Sargent. 1971. Regression with non-gaussian stable disturbances: Some sampling results. Econometrica **39**:501–510.

Bloomfield, P., and W. Steiger. 1980. Least absolute deviations curve fitting. SIAM Journal on Scientific and Statistical Computing **1**:290–301.

Bloomfield, P., and W. L. Steiger. 1983. Least absolute deviations: theory, applications, and algorithms. Birkhauser Verlag, Boston, Massachusetts, USA.

Bremigan, M. T., and R. A. Stein. 1994. Gape-dependent larval foraging and zooplankton size: implications for fish recruitment across systems. Canadian Journal of Fisheries and Aquatic Sciences **51**:913–922.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. American Naturalist **124**:255–279.

Cohen, J. E., S. L. Pimm, P. Yodzis, and J. Saldana. 1993. Body sizes of animal predators and animal prey in food webs. Journal of Animal Ecology **62**:67–78.

Currie, D. J. 1991. Energy and large-scale patterns of animal- and plant-species richness. American Naturalist **137**:27–49.

———. 1993. What shape is the relationship between body size and population density? Oikos **66**:353–358.

Damuth, J. 1981. Population density and body size in mammals. Nature **290**:699–700.

Draper, N. R., and H. Smith. 1981. Applied regression analysis. Second edition. John Wiley and Sons, New York, New York, USA.

Ebenman, B., A. Hedenstrom, U. Wennergren, B. Ekstam, J. Landin, and T. Tyrberg. 1995. The relationship between population density and body size: the role of extinction and mobility. Oikos **73**:225–230.

Gaines, S. D., and M. W. Denny. 1993. The largest, smallest, highest, lowest, longest, and shortest: extremes in ecology. Ecology **74**:1677–1692.

Gaston, K. J. 1988. Patterns in the local and regional dynamics of moth populations. Oikos **53**:49–57.

Gaston, K. J., and J. H. Lawton. 1988. Patterns in body size, population dynamics, and regional distribution of bracken herbivores. American Naturalist **132**:662–680.

Glejser, H. 1969. A new test for heteroskedasticity. Journal of the American Statistical Association **64**:316–323.

Gotelli, N. J., and D. Simberloff. 1987. The distribution and abundance of tallgrass prairie plants: a test of the core-satellite hypothesis. American Naturalist **130**:18–35.

Gould, W. W. 1992. Quantile regression with bootstrapped standard errors. Stata Technical Bulletin **9**:19–21.

Guichet, R. 1995. The diet of European hake (*Merluccius merluccius*) in the northern part of the Bay of Biscay. ICES Journal of Marine Science **52**:21–31.

Harris, T. 1950. Regression using minimum absolute deviations. The American Statistician **4**:14–15.

Harvey, A. C. 1976. Estimating regression models with multiplicative heteroscedasticity. Econometrica **44**:461–465.

Humphreys, W. F. 1979. Production and respiration in animal populations. Journal of Animal Ecology **48**:427–453.

Jeffries, M. J., and J. H. Lawton. 1985. Predator–prey ratios in communities of freshwater invertebrates: the role of enemy free space. Freshwater Biology **15**:105–112.

Juanes, F. 1986. Population density and body size in birds. American Naturalist **128**:921–929.

Juanes, F., and D. O. Conover. 1995. Size-structured piscivory: advection and the linkage between predator and prey recruitment in young-of-the-year bluefish. Marine Ecology Progress Series **128**:287–304.

Kingsford, M. J. 1992. Spatial and temporal variation in predation on reef fishes by coral trout (*Plectropomus leopardus*, Serranidae). Coral Reefs **11**:193–198.

Koenker, R., and G. Bassett. 1978. Regression quantiles. Econometrica **46**:33–50.

Koenker, R., and G. Bassett. 1982. Robust tests for heteroscedasticity based on regression quantiles. Econometrica **50**:43–61.

Koenker, R., and V. D'Orey. 1987. Computing regression quantiles. Applied Statistics **36**:383–393.

Lawton, J. H. 1989. What is the relationship between population density and body size in animals? Oikos **55**:429–434.

Lewis, D. S. C. 1974. The food and feeding habits of *Hydrocynus forskahlii* Cuvier and *Hydrocynus brevis* Gunther in Lake Kainji, Nigeria. Journal of Fish Biology **6**:349–363.

Lieth, H. 1975. Primary productivity in ecosystems: com-

parative analysis of global patterns. Pages 67–88 *in* W. H. van Dobben and R. H. Lowe-McConnell, editors. Unifying concepts in ecology. Dr. W. Junk Publishers, The Hague, The Netherlands.

Maller, R. A., E. S. de Boer, L. M. Joll, D. A. Anderson, and J. P. Hinde. 1983. Determination of the maximum foregut volume of western rock lobsters (*Panulirus cygnus*) from field data. Biometrics **39**:543–551.

Maurer, B. A., and J. H. Brown. 1988. Distribution of energy use and biomass among species of North American terrestrial birds. Ecology **69**:1923–1932.

Mudrov, V. I., V. L. Kushko, V. I. Mikhailov, and E. M. Osovitskii. 1968. Some experiments on the use of the least-moduli method in processing orbital data. Cosmic Research **6**:421–431.

Narula, S. C., and J. F. Wellington. 1982. The minimum sum of absolute errors regression: A state of the art survey. International Statistical Review **50**:317–326.

Nee, S., A. F. Read, J. J. D. Greenwood, and P. H. Harvey. 1991. The relationship between abundance and body size in British birds. Nature **351**:312–313.

Nielsen, L. A. 1980. Effect of walleye (*Stizostedion vitreum vitreum*) predation on juvenile mortality and recruitment of yellow perch (*Perca flavescens*) in Oneida Lake, New York. Canadian Journal of Fisheries and Aquatic Sciences **37**:11–19.

O'Neill, R. V., and D. L. DeAngelis. 1981. Comparative productivity and biomass relations of forest ecosystems. Pages 411–449 *in* D. E. Reichle, editor. Dynamic properties of forest ecosystems. Cambridge University Press, Cambridge, England.

Paine, R. T. 1976. Size-limited predation: an observational and experimental approach with the *Mytilus–Pisaster* interaction. Ecology **57**:858–873.

Peters, R. H., and K. Wassenberg. 1983. The effect of body size on animal abundance. Oecologia **60**:89–96.

Rogers, W. H. 1992. Quantile regression standard errors. Stata Technical Bulletin **9**:16–19.

———. 1993. Calculation of quantile regression standard errors. Stata Technical Bulletin **13**:18–19.

Rousseeuw, P. J., and A. M. Leroy. 1987. Robust regression and outlier detection. John Wiley and Sons, New York, New York, USA.

Sabelis, M. W. 1992. Predatory arthropods. Pages 225–264 *in* M. J. Crawley, editor. Natural enemies. Blackwell Scientific, London, England.

Schoener, T. W., and G. C. Gorman. 1968. Some niche differences in three lesser Antillean lizards of the genus *Anolis.* Ecology **49**:819–830.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. Third edition. W. H. Freeman, New York, New York, USA.

StataCorp. 1995. Stata Statistical Software: Release 4.0 College Station, TX: Stata Corporation.

Stuart, A., and J. K. Ord. 1991. Kendall's advanced theory of statistics, Volume 2. Fifth edition. Oxford University Press, New York, New York, USA.

Taylor, L. D. 1974. Estimation by minimizing the sum of absolute errors. Pages 169–190 *in* P. Zarembka, editor. Frontiers in econometrics. Academic Press, New York, New York, USA.

Terrell, J. W., B. S. Cade, J. Carpenter, and J. M. Thompson. 1996. Modeling stream fish habitat limitations from wedge-shaped patterns of variation in standing stock. Transactions of the American Fisheries Society **125**:104–117.

Thomson, J. D., G. Weiblen, B. A. Thomson, S. Alfaro, and P. Legendre. 1996. Untangling multiple factors in spatial distributions: lilies, gophers, and rocks. Ecology **77**:1698–1715.

Torricelli, P., M. Lugli, and L. Bobbio. 1993. The importance of male–male competition and sexually selected dimorphic traits for male reproductive success in site-attached fishes with parental care: the case of the freshwater goby *Padogobius martensi*. Pages 231–256 *in* F. A. Huntingford and P. Torricelli, editors. Behavioural ecology of fishes. Harwood Academic, Chur, Switzerland.

Vezina, A. F. 1985. Empirical relationships between predator and prey size among terrestrial vertebrate predators. Oecologia **67**:555–565.