Transactions of the American Fisheries Society 126:323–329, 1997

Incorporating Stream Level Variability into Analyses of Site Level Fish Habitat Relationships: Some Cautionary Examples

JASON B. DUNHAM[1] AND GARY L. VINYARD

Department of Biology and Program in Ecology, Evolution, and Conservation Biology
University of Nevada, Reno, Nevada 89557, USA

Abstract.—Spatial variation in stream fish populations and habitats can be partitioned into many hierarchical levels (e.g., among drainage basins, streams within basins, and sites within streams). Studies of site level habitat relationships in more than one stream are common in fisheries research. Analyses of such data typically involve multiple regression to relate site level habitat features and fish population characteristics (e.g., biomass). Because sites within streams may not be independent, multiple-regression models should also include qualitative stream effects. As we show here with hypothetical and real examples, ignoring stream effects can lead to erroneous conclusions about the significance of site level habitat variables. Site and stream level effects may function independently or interactively in relation to fish populations. Alternatively, site and stream level effects may be confounded. An example with data on trout populations revealed that highly significant site level effects were only marginally significant after adding qualitative stream effects to the regression model. Further examination of the data revealed that consideration of variation among streams added much insight and complexity to understanding how site level effects may be related to trout populations. Inclusion of stream (or other large-scale) effects in regression models of site level habitat relationships may be a valuable method to more fully understand the spatial scale of habitat variability fish are responding to.

Answers to ecological questions are often scale-dependent (Wiens et al. 1986; May 1994; Root and Schneider 1995; Wu and Loucks 1995). In this regard, studies of relationships between stream fishes and habitat variables (Fausch et al. 1988) are no exception. Stream fish habitat relationships have been examined at a variety of spatial and temporal scales (e.g., Frissell et al. 1986; Baltz 1990; Bozek and Hubert 1992; Strange et al. 1992; Hawkins et al. 1993; Matthews et al. 1994), but the impracticality of considering all relevant spatial and temporal scales is an unavoidable constraint. Here, we briefly consider one aspect of this problem: variation within streams versus among streams in studies of relationships between habitat variation and fish populations.

Spatial variation in fish populations and habitats can be hierarchically partitioned into variability among drainage basins, among streams within drainages, within streams, and so on (see Frissell et al. 1986; Hawkins et al. 1993). Surveys of fish populations at sites within several streams are, therefore, measuring variability due to site level and stream level effects. This particular survey design is common in fisheries research (e.g., Fausch et al. 1988). Investigators typically use multiple-regression models to relate site level habitat variables to fish population characteristics (Fausch et al. 1988). Stream level effects are often ignored in the analysis of this type of data. Perhaps data on stream level variables are lacking, but specific information is not necessary to statistically test for stream effects. Below, we show how ignoring stream effects can bias results of analyses of site level habitat relationships, and we offer simple advice on how to better design and analyze such studies.

Methods

Rather than criticize the work of others, we will consider a hypothetical example and then use original data to show how both stream and site level variation are relevant. Because fish habitat studies commonly use multiple linear regression (Fausch et al. 1988), we will emphasize this statistical method (see Manly 1994 for alternatives). The uses of regression analysis fall into several categories (Myers 1990) including prediction of a response (e.g., prediction of fish population characteristics in relation to environmental variation), variable screening (e.g., to determine which environmental variables are most strongly related to fish populations), model specification (e.g., to determine the form of the relationship between fish populations and environmental variation), and parameter estimation (e.g., to determine the slopes or intercepts of the relationships between environmental variables and fish population characteristics). In the examples that follow, we will emphasize variable screening as our general objective. Our point here is *not* to define specific causal

---

[1] Present address: 1052 SW Western Boulevard, Corvallis, Oregon 97333, USA.

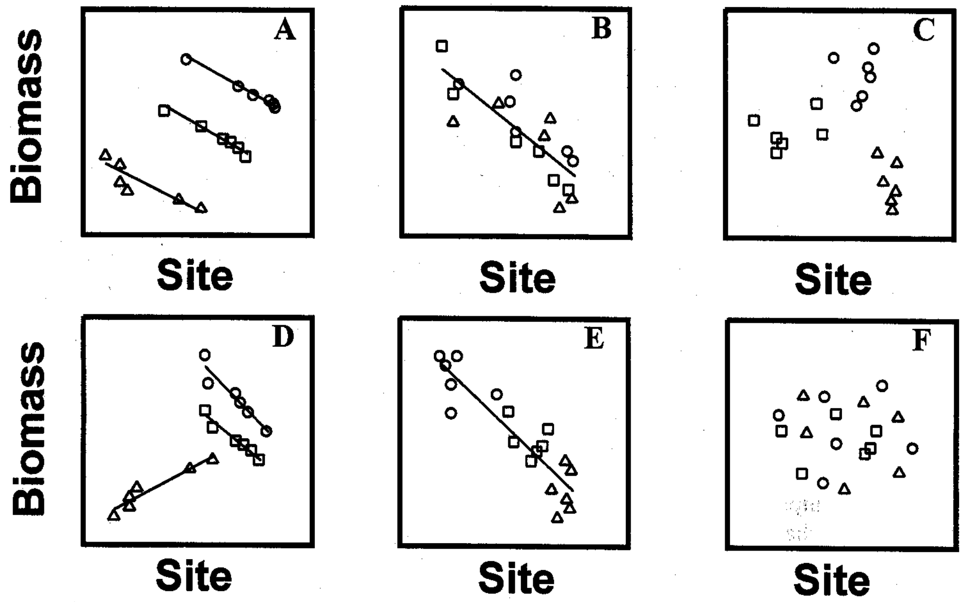

FIGURE 1.—Six possible relationships between fish biomass, qualitative stream effects (△, □, and ○ = stream categories 1, 2, and 3, respectively), and quantitative site level variation: (**A**) both stream and site level effects are significant (constant slope, different intercepts); (**B**) site level effects only (slopes and intercepts equal); (**C**) stream effects only (slopes equal to zero, different intercepts); (**D**) stream·site interaction (unequal slopes and intercepts); (**E**) stream and site level effects confounded (cannot separate stream and site effects); (**F**) no relationship with stream or site effects.

mechanisms that affect fish populations but rather to look for important patterns of variation in an exploratory analysis.

Regression models with a categorical (also called class, indicator, dummy, treatment, design, or qualitative) variable, such as "stream" and one quantitative regressor variable take the following form:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{1}$$

for category 1 and

$$y_i = (\beta_0 + \beta_2) + \beta_1 x_i + \epsilon_i \tag{2}$$

for category 2 (and so on, for each category), where y_i is the measured response variable for the ith observation, β_0 is the y intercept, β_0 is the intercept difference between category 1 and category 2, and β_1, x_i, and ϵ_i are the slope, measured predictor variable, and random error, respectively (see Myers 1990 for extensions of this simple model). In the examples that follow, β_0 represents stream effects and β_i is the slope relating site level variation (x_i) to fish biomass (y_i).

Regression analyses with qualitative and quantitative regressors belong to a general class of linear statistical models that includes analysis of variance (ANOVA) and analysis of covariance (ANCOVA; Neter et al. 1985; Myers 1990; Littell et al. 1991). In other words, regression models with one or more quantitative regressors and one or more categorical variables are statistically identical to ANCOVA models wherein the quantitative regressors are viewed as covariates. The term ANCOVA implies the quantitative regressors are not of direct interest to the study but are used to adjust the treatment means (mean values for each stream, in this case) in relation to the covariate(s) (site level variables in this case; see Kuehl 1994). An important assumption of ANCOVA is that covariate and treatment effects do not interact (e.g., no stream·site level variable interactions). The statistical models in equations (1) and (2) do not include interaction terms, but these should always be tested first (Neter et al. 1985).

Hypothetical Examples

We initially consider hypothetical data with three stream categories and six sites per stream to illustrate six possible types of relationships between fish populations (e.g., biomass) and stream level and site level variation (Figure 1). Similar examples can be found in texts by Neter et al.

(1985) and Myers (1990). First, it is possible that both stream and site level effects are significant but independent (i.e., equal slopes with different intercepts; Figure 1A). Alternatively, only site level (intercepts and slopes equal; Figure 1B) or stream level (different intercepts, slopes not significant; Figure 1C) effects may be important. Stream level and site level effects may also be interactive (i.e., slope depends on stream; Figure 1D) or confounded (stream and site effects cannot be distinguished; Figure 1E). The sixth possibility is the null case in which fish populations and habitat are not related at any level (Figure 1F). The point here is that four of six possibilities involve stream level effects. These simple examples illustrate that stream level effects may be important in analyses of relationships among site level variation and fish population characteristics.

A Real-World Example

To further illustrate these possibilities, we used data from surveys of Lahontan cutthroat trout (*Oncorhynchus clarki henshawi*) abundance and several site level habitat variables at 49 sites in seven headwater streams in the upper Humboldt River drainage, northeast Nevada (Dunham and Vinyard, unpublished data). Three streams (West Marys River, East Marys River, and Marys River Basin Creek) are perennially interconnected tributaries of the upper Marys River, whereas T, Foreman, Gance, and Frazer creeks are isolated stream habitats. Catchment basin areas and average elevations of these streams range from 13 to 27 km^2 and from 1,975 to 2,670 m, respectively.

All streams were sampled within a very short time period (6–8 weeks) during summer base flow to minimize temporal differences (Austen et al. 1994). In each stream we surveyed seven sections (approximately 25 m each) with multiple-pass electrofishing (van Deventer and Platts 1989). Two stream sites were dropped from the analysis because cutthroat trout were absent for reasons unrelated to habitat conditions (e.g., physical migration barriers). Habitat variables measured at each site included average water depth, maximum water depth, wetted channel width, numbers of large woody debris pieces, canopy density, undercut bank volume, instream macrophyte and woody cover, conductivity, and pH. Details of this work will be published in another paper.

Multiple-regression analyses were conducted with SAS Institute software (Freund and Littell 1991; Littell et al. 1991) to examine relationships between $\log_{10}$ cutthroat trout biomass (g fish/m) and site level habitat variables. Logarithmic transformation of the response was needed to stabilize variance heterogeneity (tested with ANOVA on the absolute residuals; Kuehl 1994) and to linearize relationships between biomass and the regressor variables (determined by examination of partial regression plots; Myers 1990). We selected the best linear regression model by using "all subsets" regression (SAS PROC REG RSQUARE option; Littell et al. 1991). Model selection criteria included mean square error, Mallows' C_p, R^2, and adjusted R^2, which adjusts for number of variables in the model (Myers 1990; Freund and Littell 1991). Multicollinearity, influential observations, variance heterogeneity, and other diagnostics were evaluated to ensure model assumptions were satisfied (Phillipi 1994). After variable selection, we reanalyzed the reduced model with stream effects included in the model as a categorical variable. Stream·site interactions were tested as well.

Results and Discussion

Model selection criteria identified instream macrophyte cover and wetted stream width as the two best site level predictor variables. With these two site level variables, we investigated four models corresponding to the possible fish habitat relationships shown in Figure 1A–D (Table 1). To look at site effects only (Figure 1B), we conducted a separate multiple regression with stream width and instream macrophyte cover (Table 1A). Regression results indicated only width effects were significant ($P < 0.0001$), so instream macrophyte cover was dropped from subsequent analyses. Analysis with a model including both width and stream effects (Table 1B, compare to Figure 1A) showed strong stream effects ($P < 0.0005$), but stream width was only weakly ($P = 0.045$) related to cutthroat trout biomass. Stream·width interactions were not significant (Table 1C), but stream effects alone explained more than 75% of the variation in cutthroat trout biomass. Adding stream width to the model only increased R^2 by 3% (from 0.75 to 0.78, Table 1B, D). If stream effects were ignored in this analysis, wetted width and cutthroat trout biomass would appear to be strongly related (Figure 2), but our analysis with stream effects indicates biomass variation among streams is much greater than variation among sites within streams (width effects).

A closer examination of the plot of stream width versus fish biomass per meter (Figure 2) shows our real-world example is not as clear-cut as the hypothetical examples (Figure 1). The finding that

TABLE 1.—Analysis of variance tables for different models of relationships between Lahontan cutthroat trout biomass and stream and site level habitat variation for (A) multiple regression with two site level predictor variables ($R^2 = 0.62$); (B) a model with qualitative stream effects and stream width ($R^2 = 0.78$); (C) same model as (B), but with a stream*width interaction in the model ($R^2 = 0.80$); (D) a model with stream effects only ($R^2 = 0.75$). Mean square (MS) values are from partial sums of squares (SAS type III) except for case (C), wherein MS values are from SAS Type I sums of squares (see Myers 1990; Littell et al. 1991).

Source	df	MS	*F*	*P*
A: Site effects only				
Model	2	3.334	35.32	0.0001
Error	44	0.094		
B: Stream and site effects				
Stream	6	0.323	5.25	0.0005
Stream width	1	0.263	4.28	0.0451
Error	39	0.061		
C: Stream, site, and stream*site interaction effects				
Stream	6	1.360	21.22	0.0001
Stream width	1	0.263	4.11	0.0507
Interaction	6	0.047	0.74	0.6231
Error	33	0.064		
D: Stream effects only				
Stream	6	1.360	20.44	0.0001
Error	40	0.067		

stream level effects dominate site level (e.g., width) variation masks potentially more complex patterns in the data. Only two streams (West Marys and East Marys rivers) showed significant ($P < 0.05$) negative width effects (regression slopes) when separate regressions were fitted for each stream. These two streams appear to show within-stream width effects that correspond to the hypothetical example in Figure 1A or, perhaps, Figure 1B depending on intercept differences. Other streams show no such relationship. The pattern suggests a stream·width interaction, but interaction effects were too weak to be detected with a sample size of only six or seven sites per stream.

In other streams shown in Figure 2, stream width appears to be confounded with variation in biomass among streams (Figure 1e). For example, variation in fish biomass within streams *and* stream width in Frazer, Gance, and Foreman creeks appear to vary together as a function of stream. More information is needed to determine if biomass differences among these streams can be attributed to width or to some other stream level characteristic. In conclusion, stream width may truly be an important *correlate* (no causation implied) of fish biomass in the seven streams in this example, but addition of stream effects to the model reveals ad-

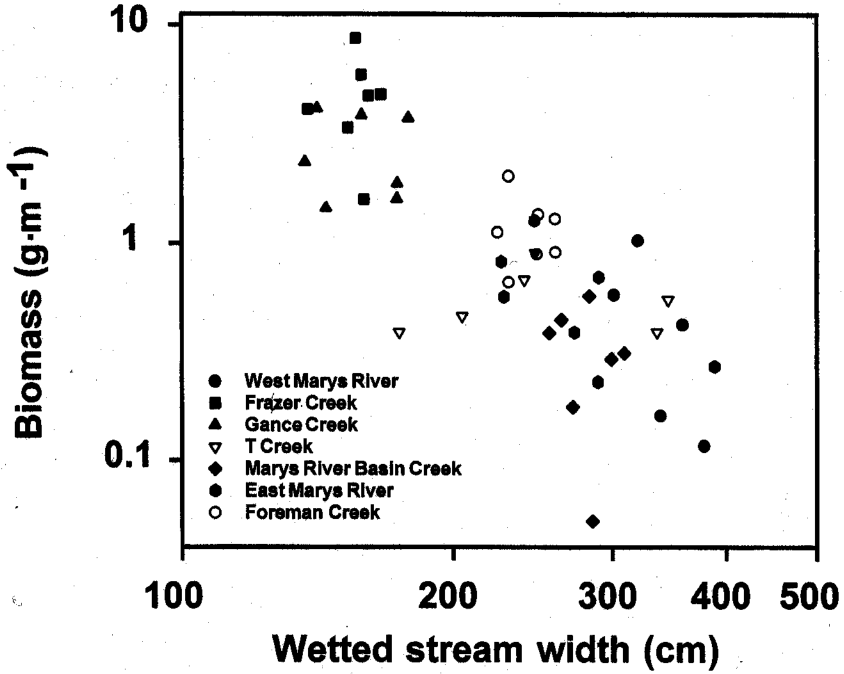

FIGURE 2.—Plot of Lahontan cutthroat trout biomass against wetted stream width (log scale). Compare with possible relationships in Figure 1.

ditional complexity that would not be apparent by an analysis of stream width (site effects) alone. In other words, if the data were analyzed by assuming only width (site) effects (see Figure 1B, Table 1A), this complexity would be missed.

We emphasize the use of multiple linear regression here, but other related statistical methods are available for detecting spatial structure in ecological data sets, such as spatial autocorrelation analysis (Legendre 1993; Hinch et al. 1994; Fortin and Gurevitch 1994). In terms of equations 1 and 2, autocorrelation refers to correlations among the error terms (ϵ_i) among sites within streams. Our example data set with Lahontan cutthroat trout, consisting of only six or seven sites per stream, was too small to test for autocorrelation with available methods (Fortin and Gurevitch 1994).

Positive spatial (and temporal) autocorrelation may not bias parameter estimates, but variances may be underestimated, inflating the chance of finding a significant result when in fact one does not exist: a type I statistical error (Zar 1984). In our real-world example, we may expect positive spatial autocorrelation due to movement of individuals among sites within streams or among sites with similar habitat characteristics.

A common remedy for dealing with autocorrelated errors is to add additional regressor variables to remove the autocorrelation (Myers 1990). The fact that site level variability (i.e., stream width) was only marginally significant ($P < 0.045$) after adding stream effects in the regression analysis suggests that autocorrelation may have been present.

Conclusion

Results of the above analyses strongly suggest cutthroat trout biomass is related to stream level and possibly site level habitat variation (at least with respect to site level variables measured in this study). In some cases, the analyses suggest stream and site effects may be confounded. At this point, we know little about specific stream level habitat characteristics, but at least we have a clearer idea of the relevant spatial scales for addressing cutthroat trout habitat relationships (note that here we are purposely ignoring the problem of using population density or biomass alone as an indicator of habitat quality; van Horne 1983).

An important caveat to note is that analyses such as these described here to analyze stream effects must use biologically meaningful definitions of "stream" or other such unit (e.g., "basin," "reach," or "channel unit"; see Frissell et al. 1986). In our example with Lahontan cutthroat trout, streams in the analyses are meaningful units because they are either isolated from other fluvial habitats or tributary to larger mainstem habitats. In the latter case, we consider the tributary streams to be meaningful units because they have very different characteristics than their larger mainstem counterparts (Minshall et al. 1983).

A second important caveat of the approach we use here is that pattern detection methods, such as regression analysis, do not necessarily imply causal relationships. Such analyses may be essential, however, to the formation of causal hypotheses and more experimental, process-oriented study (Ludwig and Reynolds 1988). Our point here was not to discuss all of the potential larger-scale, stream level processes that may actually affect stream fishes but, rather, to show how site and stream level variability may be considered together in a correlation or regression analysis. In our example, lack of specific information on exactly what stream effects (e.g., stream size, elevation, geomorphology, or other characteristics) were related to Lahontan cutthroat trout populations and how these characteristics affected the populations indicates larger-scale experimental and observational studies are needed to better understand this species.

To summarize, studies of site level fish habitat relationships from sites in more than one stream should include consideration of the following: (1) stream effects are potentially very important in developing habitat models to predict fish biomass; (2) interactions among stream and site level variation also may be important; (3) inclusion of effects at different scales is useful for determining the scale of habitat variability (e.g., the "environmental grain," sensu Levins 1968) fish are responding to; (4) specific characteristics of streams need not be known to include stream effects in a habitat model; and (5) sampling designs should include enough sites and streams to detect effects at different scales (Cohen 1988).

Lack of attention to stream (or other large-scale) effects may partly explain why habitat-based models often have little explanatory power (Fausch et al. 1988). Even worse, as evidenced by the examples presented here, failure to consider more than one spatial (or temporal) scale can lead to serious misinterpretations. We do not, however, wholly condemn habitat models based on limited spatial scales. Such information is very useful provided the inherent constraints of the data are acknowledged. Obviously, it is practically impossible to consider all relevant spatial and temporal scales in studies of stream fish habitat relationships. Nonetheless, as we have shown here, ignoring the problem of scale in fisheries can have serious consequences in a scale-dependent world.

Acknowledgments

Major funding for this work was provided by a grant from the U.S. Forest Service, Intermountain Research Station to G. Vinyard (INT-92731-RJVA). Additional support was provided by the Biological Resources Research Center and Biology Department, University of Nevada–Reno. This work would not have been possible without permission of the U.S. Fish and Wildlife Service Region I and Nevada Department of Wildlife (permits S-9039, S-0153, S-12357). We thank P. Coffin, G. Hoelzer, S. Jenkins, G. Johnson, D. Lee, J. McIntyre, R. Nelson, M. Reed, and B. Rieman for their suggestions during the formative stages of this research; S. Breck, S. Dunham, S. Garland, T. Jenni, P. Fischer, M. Olsen, M. Peacock, and M. Rahn for assistance with field data collection; and K. Obermeyer, C. Rabeni, B. Rieman, S. Sowa, and two anonymous reviewers for critical comments on various drafts of the manuscript.

References

Austen, D. J., D. L. Scarnecchia, and E. P. Bergersen. 1994. Usefulness of structural and condition indices in management of high-mountain stream salmonid populations. North American Journal of Fisheries Management 14:681–691.

Baltz, D. M. 1990. Autecology. Pages 585–607 *in* C. B. Schreck, and P. B. Moyle, editors. Methods for fish biology. American Fisheries Society, Bethesda, Maryland.

Bozek, M. A., and W. A. Hubert. 1992. Segregation of resident trout in streams as predicted by three habitat dimensions. Canadian Journal of Zoology 70: 886–890.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences, 2nd edition. Erlbaum Associates, Hillsdale, New Jersey.

Fausch, K. D., C. L. Hawkes, and M. G. Parsons. 1988. Models that predict standing crop of stream fish from habitat variables: 1950–85. U.S. Forest Service General Technical Report PNW-213.

Fortin, M.-J., and J. Gurevitch. 1994. Mantel tests: spatial structure in field experiments. Pages 342–359 *in* S. M. Scheiner, and J. Gurevitch, editors. Design and analysis of ecological experiments. Chapman and Hall, New York.

Freund, R. J., and R. C. Littell. 1991. SAS system for regression. SAS Institute, Cary, North Carolina.

Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. Environmental Management 10:199–214.

Hawkins, C. P., and eleven coauthors. 1993. A hierarchical approach to classifying stream habitat features. Fisheries 18(6):3–12.

Hinch, S. G., K. M. Somers, and N. C. Collins. 1994. Spatial autocorrelation and assessment of habitat–abundance relationships in littoral zone fish. Canadian Journal of Fisheries and Aquatic Sciences 51:702–712.

Kuehl, R. O. 1994. Statistical principles of research design and analysis. Duxbury Press, Belmont, California.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74:1659–1673.

Levins, R. 1968. Evolution in changing environments. Princeton University Press, Princeton, New Jersey.

Littell, R. C., R. J. Freund, and P. C. Spector. 1991. SAS system for linear models. SAS Institute, Cary, North Carolina.

Ludwig, J. A., and J. F. Reynolds. 1988. Statistical ecology, a primer on methods and computing. Wiley, New York.

Manly, B. F. J. 1994. Multivariate statistical methods: a primer, 2nd edition. Chapman and Hall, New York.

Matthews, W. J., B. C. Harvey, and M. E. Power. 1994. Spatial and temporal patterns in the fish assemblages of individual pools in a midwestern stream (U.S.A.). Environmental Biology of Fishes 39:381–397.

May, R. M. 1994. The importance of scale to ecological questions and answers. Pages 1–17 *in* P. J. Edwards, R. M. May, and N. R. Webb, editors. Large scale ecology and conservation biology. Blackwell Scientific Publications, Oxford, UK.

Minshall, G. W., and six coauthors. 1983. Interbiome comparison of stream ecosystem dynamics. Ecological Monographs 53:1–25.

Myers, R. H. 1990. Classical and modern regression with applications, 2nd edition. Duxbury Press, Belmont, California.

Neter, J., W. Wasserman, and M. H. Kutner. 1985. Applied linear statistical models, 2nd edition. Irwin, Homewood, Illinois.

Phillipi, T. E. 1994. Multiple regression: herbivory. Pages 183–210 *in* S. M. Scheiner and J. Gurevitch, editors. Design and analysis of ecological experiments. Chapman and Hall, New York.

Root, T. L., and S. H. Schneider. 1995. Ecology and climate: research strategies and implications. Science 269:334–341.

Strange, E. L., P. B. Moyle, and T. D. Foin. 1992. Interactions between stochastic and deterministic processes in stream fish community assembly. Environmental Biology of Fishes 36:1–15.

van Deventer, J. S., and W. S. Platts. 1989. Microcomputer software system for generating population statistics from electrofishing data—user's guide for MICROFISH 3.0. U.S. Forest Service General Technical Report INT-254.

van Horne, B. 1983. Density as a misleading indicator