

Application of adaptive cluster sampling to low-density populations of freshwater mussels

DAVID R. SMITH, RITA F. VILLELLA and
DAVID P. LEMARIÉ

U.S. Geological Survey, Leetown Science Center, Kearneysville, West Virginia 25430

E-mail: david_r_smith@usgs.gov

E-mail: rita_villella@usgs.gov

E-mail: david_lemarie@usgs.gov

Received August 1999; Revised January 2002

Freshwater mussels appear to be promising candidates for adaptive cluster sampling because they are benthic macroinvertebrates that cluster spatially and are frequently found at low densities. We applied adaptive cluster sampling to estimate density of freshwater mussels at 24 sites along the Cacapon River, WV, where a preliminary timed search indicated that mussels were present at low density. Adaptive cluster sampling increased yield of individual mussels and detection of uncommon species; however, it did not improve precision of density estimates. Because finding uncommon species, collecting individuals of those species, and estimating their densities are important conservation activities, additional research is warranted on application of adaptive cluster sampling to freshwater mussels. However, at this time we do not recommend routine application of adaptive cluster sampling to freshwater mussel populations. The ultimate, and currently unanswered, question is how to tell when adaptive cluster sampling should be used, i.e., when is a population sufficiently rare and clustered for adaptive cluster sampling to be efficient and practical? A cost-effective procedure needs to be developed to identify biological populations for which adaptive cluster sampling is appropriate.

Keywords: adaptive cluster sampling, biological surveys, freshwater mussels, sampling rare populations, unionids

1352-8505 © 2003  Kluwer Academic Publishers

1. Introduction

Adaptive cluster sampling (Thompson and Seber, 1996) is an intuitively appealing approach to sampling clustered populations when cluster membership is not known prior to sampling. Perhaps most appealing, from a biologist's perspective, is that adaptive cluster sampling resembles the manner a biologist collects biological organisms, especially rare ones. Adaptive cluster sampling, which results in design-based estimates of population density, allows a biologist to survey for an organism of interest, and once the organism is found, the biologist can continue to search and collect organisms nearby.

1352-8505 © 2003  Kluwer Academic Publishers

Increasing effort in response to finding what you are looking for is the natural practice of those who search for rare and clustered objects, e.g., biologists, anglers, hunters, birders, bargain shoppers, and even some game players (consider, for example, the game of Battleship by Milton-Bradley[®]).

From a statistical perspective, adaptive cluster sampling is appealing because it is efficient for rare and clustered populations (Thompson and Seber, 1996; Christman, 1997). Density of rare and clustered populations is difficult to estimate precisely using conventional sampling designs. The two factors that interact to determine efficiency of adaptive cluster sampling are within network variance relative to population variance and final sample fraction relative to initial sample fraction (Thompson, 1990; Smith *et al.*, 1995). Adaptive cluster sampling, as with conventional cluster sampling, is efficient when within network (or within cluster) variance is close to population variance, which occurs when the population is clustered. Simultaneously, adaptive cluster sampling is efficient when final sample size is close to initial sample size, which occurs when the population is rare.

Two issues temper the appeal of adaptive cluster sampling: (1) efficiency depends on degree of rarity and clustering, which is unknown prior to sampling, and (2) the final sample size is random, which adds uncertainty to survey planning. Misapplication of adaptive cluster sampling can result in failure to achieve desired precision or in a final sample size that exceeds a survey's budget. Excessive final sample size can result when the biological population is not as rare as or is more widely distributed than anticipated; practitioners of adaptive cluster sampling refer to this problem colloquially as encountering "the cluster from hell". Recent work has focused on limiting the final sample size (Brown and Manly, 1998; Thompson and Seber, 1996; Salehi and Seber, 1997). However, if and when a population is sufficiently rare and clustered for adaptive cluster sampling to be efficient is, at this time, an open question.

Although there have been many simulation studies of adaptive sampling, there have been few documented field trials on real biological populations. Our objective was to apply adaptive cluster sampling to biological populations, which tend to be spatially clustered and found at low density, and evaluate its performance. Freshwater mussels are benthic macroinvertebrates that cluster spatially and are frequently found at low densities (Kovalak *et al.*, 1986; Downing and Downing, 1992; Strayer and Ralley, 1993). Thus, freshwater mussels seem to be ideal candidates for adaptive cluster sampling (Dorazio, 1999). Strayer *et al.* (1996) applied adaptive cluster sampling to freshwater mussel populations, but did not evaluate the application. We applied adaptive cluster sampling at 24 sites in the Cacapon River, WV to estimate density of freshwater mussels. In Section 2, we outline the survey design. In Section 3, we evaluate our application of adaptive cluster sampling in terms of efficiency, yield of individuals, and detection of uncommon species. We offer conclusions and suggestions in Section 4.

2. Application to a survey of freshwater mussels

Conservation of North American freshwater mussels (Bivalvia: Unionidae) is at a critical juncture with their extinction rates rivaling those of tropical species (Ricciardi and Rasmussen, 1999). Freshwater mussel diversity is threatened by anthropogenic activities

such as dam building, land use induced changes in water quality, spread of exotic species (e.g., zebra mussels), and over-harvest of some large river species (Williams *et al.*, 1993).

Identifying locations of populations and estimating population parameters, such as population density, are integral to research, conservation, and management of freshwater mussels. There is an important, unmet need for sampling techniques that will yield precise estimates of low-density (sparse) populations of freshwater mussels (Strayer *et al.*, 1996). Spatial aggregation of freshwater mussels, a characteristic that adds to the challenge of precise estimation, is likely due to the role of aggregation in reproduction and the association between mussel occurrence and patchily distributed habitat. Downing *et al.* (1993) found that fertilization success was strongly correlated with spatial aggregation. Strayer and Ralley (1993) found that current speed, which exhibits within-site variation, was a useful predictor of mussel occurrence.

2.1 Sampling methods

We applied adaptive cluster sampling at sites along the Cacapon River, WV, which is a 3rd order stream in the upper Potomac River drainage. Sampling was restricted to the riffle habitat type because it is shallow and does not require SCUBA diving to collect mussels. The extent of the site was determined by the extent of the habitat.

To avoid over-sampling (which could occur when the final sample size is many multiples of the initial sample size), we used double sampling for stratification (Cochran, 1977) to limit the application of adaptive cluster sampling only to low-density sites. First, we selected 30 sites throughout the Cacapon River in a constrained random fashion by selecting one site at random in each 4 km stretch of river. Then at each site, we assessed density level of all species qualitatively by searching in a haphazard fashion throughout the site and counting mussels using glass-bottom buckets for 1 hr (e.g., two biologists for 30 min or four biologists for 15 min). Based on the qualitative assessment of density, we classified sites into two strata (i.e., low and high density). From previous experience sampling mussels in this river, we decided *a priori* that 30 mussels counted per hour would be the cut off between low (< 1 mussel per m^2) and high-density (≥ 1 mussel per m^2) strata. Site area (including islands) ranged from 337 to 2177 m^2 in low-density sites, and 199 to 3020 m^2 in high-density sites. Finally, we used adaptive and conventional sampling to subsample the low and high-density sites, respectively. Here, we report only on the results from adaptive cluster sampling; results from other sites will be reported elsewhere.

We used 0.25 m^2 quadrats, a standard sampling unit size for freshwater mussels, to sample within a site. The initial sample of 0.25 m^2 quadrats was placed systematically throughout the site to achieve good spatial coverage. Bank-to-bank transects were located at 5 m intervals with a random start; then quadrats were placed along transects at 3 m intervals with a random start on each transect. The initial sample size depended on the wetted area of a site (excluding islands). Adaptive sampling was triggered by detection of ≥ 1 mussel in a quadrat. The neighborhood included the quadrat that met the criterion and the four adjacent quadrats.

The manner of placing initial quadrats resulted in joint inclusion probabilities equal to zero for some units. Thus, variance could not be estimated without further assumptions. We proceeded by treating the initial sample as a simple random sample, and we expected that variances would be overestimated as a result (Thompson, 1992).

2.2 Evaluation of the design

We evaluated adaptive versus conventional sampling by fixing cost and comparing precision, as measured by coefficient of variation (CV). For each site, CV was calculated for adaptive cluster sampling (CV_{ACS}), simple random sampling given the initial sample size ($CV_{SRS|n_1}$), and simple random sampling given the final sample size ($CV_{SRS|\nu}$). The final sample size (ν) was a count of all network quadrats plus edge units. To predict CV for conventional sampling, we estimated the population variance from the initial sample ($\hat{\sigma}^2$) then calculated the variance as $(1 - f_n)\hat{\sigma}^2/n$, where f_n was the sampling fraction for sample size of n , and then we set n to the value of the initial sample size for $CV_{SRS|n_1}$ and the final sample size for $CV_{SRS|\nu}$. This prediction is commonly done when planning a survey based on data from a pilot survey. The estimate of population variance derived from a pilot survey is used commonly to calculate sample size or to predict estimator precision for a given sample size. Here we use an estimate of population variance based on the initial sample to predict estimator CV for the initial and final sample sizes.

We used logistic regression to compute the probability of sampling species other than *Elliptio complanta* (eastern elliptio) for initial and adaptive quadrats. The Cacapon River mussel fauna included six species. *E. complanata*, which tends to be the most common species in Atlantic slope rivers, was the most abundant species in the Cacapon River. Other species were comparatively uncommon. These species were *Lampsilis cariosa* (yellow lampmussel), *Elliptio fisheriana* (northern lance), *Alasmidonta varicosa* (brook floater), *Lasmigona subviridis* (green floater), and *Strophitus undulatus* (squawfoot).

3. Results

Adaptive cluster sampling did not improve relative precision over simple random sampling for fixed sample size (Table 1). Efficiency tended to decrease as density increased, and the greatest inefficiency was at the site with the highest density ($0.71/\text{m}^2$) and variance-to-mean ratio (1.60). If the cost to locate sampling units is high then precision for fixed cost might be more favorable for adaptive cluster sampling, which tends to arrange units close together. However, cost to place a quadrat was minimal. Therefore, we do not expect that the comparison of precision would change substantially, if we fixed on total cost rather than sample size.

Yield, in terms of collecting individual mussels and detecting uncommon species, was higher in adaptively sampled quadrats. For reaches where density exceeded 0.15 mussels per m^2 , mussels per quadrat tended to be higher for adaptive quadrats (Fig. 1). Based on the logistic regression model in Table 2, the odds ratio for sampling uncommon species in adaptive relative to initial quadrats was 27.7 (95% CI: 1.1, 678.5). We show predicted probabilities of sampling uncommon species in Fig. 2.

The final sample size was at most 3.8 times the initial sample size (Table 1). At approximately 72% of the adaptively sampled reaches, final sample size was ≤ 1.5 times the initial sample size; at approximately 88% of the reaches, final sample size was ≤ 2.0 times the initial sample size.

Table 1. Results from 18 sites on the Cacapon River, WV, where freshwater mussels were sampled and adaptively cluster sampling was applied. There were 24 sites where adaptive cluster sampling was applied; however, at six of those sites freshwater mussels were not detected. Coefficient of variation (CV) was calculated for adaptive cluster sampling (CV_{ACS}), simple random sampling given the initial sample size ($CV_{SRS|n=n1}$), and simple random sampling given the final sample size ($CV_{SRS|n=\nu}$).

Density (no./m ²)	Standard Error	Variance to Mean Ratio	Initial Sample Size	Final Sample Size	Ratio of Final to Initial Sample Size	Coefficient of variation			Ratio of $CV_{SRS n=\nu}$ to CV_{ACS}
						CV_{ACS}	$CV_{SRS n=n1}$	$CV_{SRS n=\nu}$	
0.03	0.0327	1	121	124	1.02	0.99	0.99	0.88	0.89
0.06	0.0566	1	70	74	1.06	0.99	0.99	0.96	0.97
0.08	0.0825	1	48	52	1.08	0.99	0.99	0.95	0.96
0.11	0.0523	0.92	149	164	1.10	0.49	0.49	0.47	0.96
0.12	0.0806	0.99	69	77	1.12	0.70	0.70	0.66	0.94
0.18	0.0877	0.97	119	151	1.27	0.49	0.49	0.48	0.98
0.18	0.0867	0.97	90	111	1.23	0.49	0.49	0.44	0.90
0.19	0.0884	1.3	125	145	1.16	0.46	0.46	0.43	0.93
0.19	0.1086	0.97	62	73	1.18	0.56	0.56	0.52	0.93
0.20	0.1085	0.97	62	76	1.23	0.56	0.56	0.50	0.89
0.33	0.1404	0.94	67	107	1.60	0.42	0.43	0.34	0.81
0.33	0.1141	1.52	112	157	1.40	0.35	0.39	0.33	0.94
0.34	0.1013	0.92	120	178	1.48	0.30	0.30	0.25	0.83
0.48	0.1991	0.90	42	67	1.60	0.41	0.42	0.33	0.80
0.57	0.2226	0.95	38	146	3.84	0.39	0.55	0.27	0.69
0.59	0.1453	1.49	137	264	1.93	0.25	0.26	0.18	0.72
0.63	0.285	1.21	38	71	1.87	0.46	0.42	0.30	0.65
0.71	0.2031	1.60	95	315	3.32	0.29	0.31	0.17	0.59

Table 2. Logistic model coefficients where the outcome is encountering a rare species of mussel. Independent variables included number of quadrats and quadrat type (initial or adaptive). Global null hypothesis (all coefficients = 0) was rejected ($X^2 = 29.74$, 2 df, $P = 0.0001$). Interaction between number of quadrats and quadrat type was not significant ($X^2 = 0.42$, 1 df, $P = 0.51$).

Variable	Estimated Coefficient	Standard Error	Coeff./SE	Odds Ratio
Intercept	-6.6971	2.0395	3.2837	.
Number of quadrats	0.0697	0.0208	3.3510	1.072
Quadrat type	3.3223	1.6314	1.9756	27.723

4. Discussion

We found that application of adaptive cluster sampling increased yield of individual mussels and detection of uncommon species; however, it did not improve the precision of density estimates. In surveys of freshwater mussels, the goal of high yields of individuals and detection of uncommon species often accompanies the goal of estimating population density or abundance. In multiple-species surveys, emphasis is often on the uncommon species. In such cases, malacologists might prefer sampling designs that increase the likelihood of encountering uncommon species or collecting more individuals of those species.

At sites where we applied adaptive cluster sampling, density was ≤ 0.71 per m² and

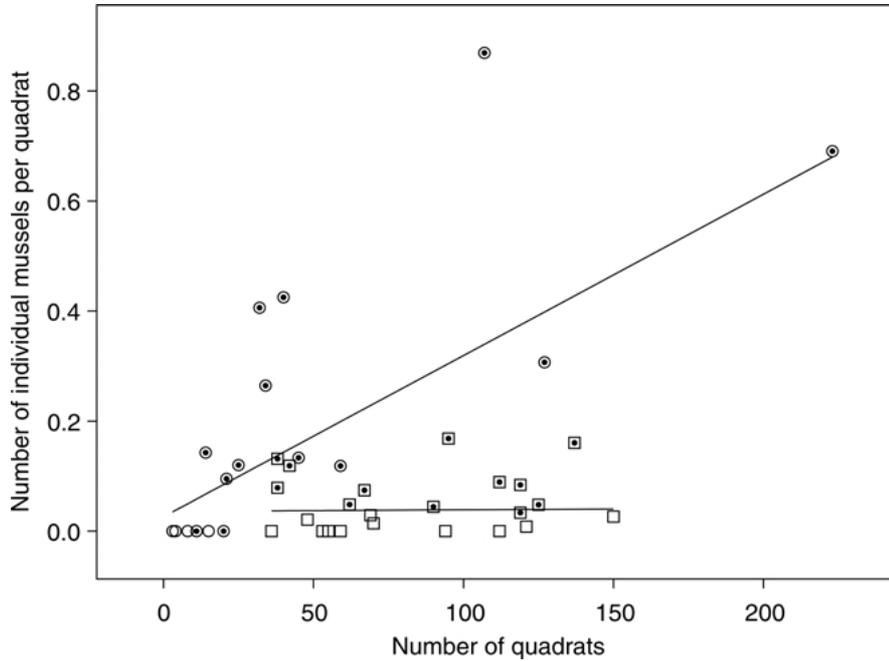


Figure 1. Yield of individual mussels in adaptively (○) and conventionally (□) placed quadrats at sites with low-densities ($< 1/m^2$) of freshwater mussels. Sites with density $\geq 0.15 m^2$ are shown with embedded dots. Robust regression lines are shown.

variance-to-mean ratio was ≤ 1.6 . At five of the sites where only conventional sampling was applied, density was > 1 per m^2 and for three of those, variance-to-mean ratio was ≥ 2.6 (i.e., 2.6, 2.72, and 2.89). It is possible that adaptive cluster sampling would have been efficient at those high-density sites. To guard against excessive sampling effort some measure to control final sample size would have to have been implemented at high-density sites. When applying adaptive cluster sampling to freshwater mussels, Strayer *et al.* (1996) controlled final sample size by varying the criterion used to trigger adaptive sampling, setting it higher at sites where a reconnaissance indicated high-density populations. In our application, we were reluctant to increase the criterion as a way to control final sample size, because the practice has an unpredictable effect on efficiency. Increasing the criterion decreases both the within network variance and the effective sample size. The difference between the within-network and population variances would decrease, because units with values below the criterion would not be included in the network and units within networks would have more uniformly high values. At the same time, the difference between the initial and effective sample sizes would decrease, because maximum network size would be reduced. However, these effects influence efficiency in opposite directions. Reducing within-network variance tends to decrease efficiency, while reducing effective sample size tends to increase efficiency. With so many possible criteria and no guidance on the sensitivity of efficiency to these changes, we do not believe that increasing the criterion is a practical method to control final sample size and, at the same time, enhance or maintain efficiency.

Recent work has identified several other approaches to limit final sample size. Brown

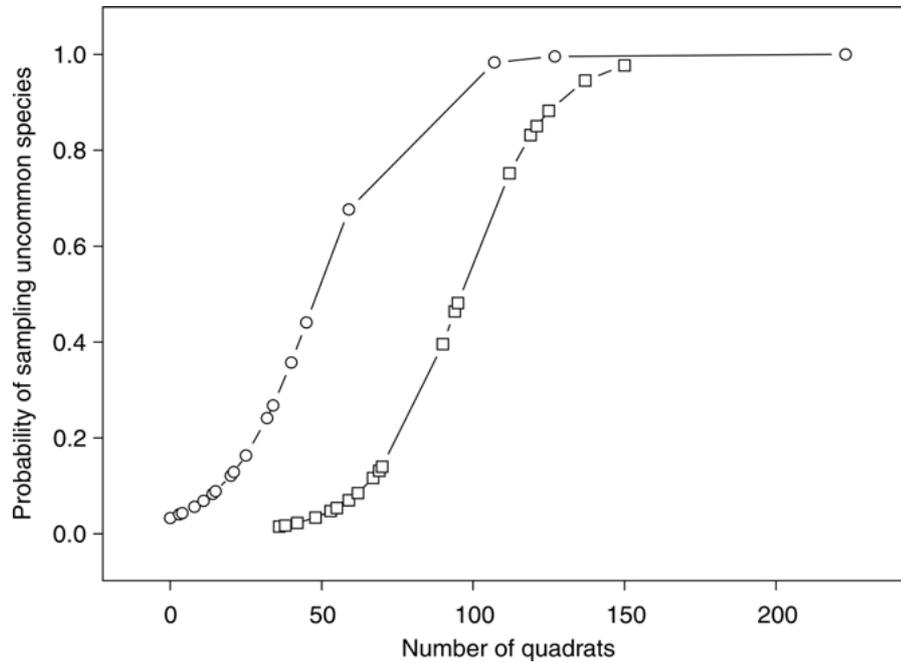


Figure 2. Probability of sampling uncommon species in adaptively (○) and conventionally (□) placed quadrats at sites with low-densities ($< 1/m^2$) of freshwater mussels. Probabilities were from the logistic regression model presented in Table 2.

and Manly (1998) suggested that sampling could stop after cumulative sample size meets or exceeds a preset value. The stopping rule reduces variation in final sample size, but causes estimates to be biased. However, Brown and Manly (1998) found the stopping rule gave reasonable results, in terms of mean square error, for several simulated populations. Salehi and Seber (1997) suggest designing surveys based on pilot survey information from two-stage adaptive cluster sampling. First select a small number of primary units using random sampling, and then select secondary units within the primary sampling units by adaptive cluster sampling. The authors indicate how one can use the pilot survey data to compute the number of primary units required for a desired final sample size or level of precision. Thompson and Seber (1996:160–161) discuss other approaches to control the final sample size that utilize stratification, order statistics, partitioning into blocks (similar to primary units in Salehi and Seber, 1997), and neighborhood definitions.

Unfortunately, we do not know if application of adaptive cluster sampling at high-density sites on the Cacapon River would have been efficient or practical. We do know, however, that at high-density sites precise estimates of density can be achieved by applying conventional quadrat-based sampling without the variability in final sample size that adaptive sampling introduces (Downing and Downing, 1992; Smith *et al.*, 2001a, Smith *et al.*, 2001b). We had hoped that adaptive cluster sampling would have been efficient at low-density sites because large sample sizes are necessary to achieve precise estimates of low-density populations using conventional sampling (Smith *et al.*, 2001a). Because of the importance of finding uncommon species and collecting individuals of those species, additional research is warranted on application of adaptive cluster sampling

to freshwater mussels. However, at this time we cannot recommend routine application of adaptive cluster sampling to low-density freshwater mussel populations.

Our application and evaluation was compromised by use of systematic sampling methods with joint inclusion probabilities equal to zero for some units. We proceeded by applying inclusion probabilities appropriate for simple random sampling and approximating variances. Hence our results are approximate. We continue to favor systematic sampling because of good spatial coverage and ease of implement in the field, which are desirable properties for surveys of biological populations (Christman, 2000). Currently, we implement systematic sampling with multiple random starts so that joint inclusion probabilities exceed zero for all units and variance can be estimated without bias (Smith *et al.*, 2001a).

The ultimate (and currently unanswered) question is how to tell when adaptive cluster sampling should be used, i.e., when is a population sufficiently rare and clustered for adaptive cluster sampling to be efficient and practical? We applied adaptive cluster sampling to populations that we suspected would be rare and clustered. Some of the populations turned out to be rare and clustered, some were just rare, and some were just clustered. However, even for the rare and clustered populations adaptive cluster sampling was not efficient. A cost-effective procedure needs to be developed to identify populations for which adaptive cluster sampling is appropriate. Part of adaptive cluster sampling's appeal is the fact that it can be applied without prior knowledge of cluster location. Adaptive cluster sampling will lose its appeal if effective application demands *a priori*, detailed, and costly information on density and distribution.

Acknowledgments

We thank Mary Christman and anonymous referees for helpful comments on an earlier draft. David Weller, Priscilla Young, and William Bartles assisted in data collection and data management.

References

- Brown, J.A. and Manly, B.J.F. (1998) Restricted adaptive cluster sampling. *Environmental and Ecological Statistics*, **5**, 49–63.
- Christman, M.C. (1997) Efficiency of some sampling designs for spatially clustered populations. *Environmetrics*, **8**, 145–66.
- Christman, M.C. (2000) A review of quadrat-based sampling of rare, geographically clustered populations. *Journal of Agricultural, Biological, and Environmental Statistics*, **5**, 168–201.
- Cochran, W.G. (1977) *Sampling Techniques* (3rd edition), Wiley, New York.
- Downing, J.A. and Downing, W.L. (1992) Spatial aggregation, precision, and power in surveys of freshwater mussel populations. *Canadian Journal of Fisheries and Aquatic Science*, **49**, 985–91.
- Downing, J.A., Rochon, Y., and Pérusse, M. (1993) Spatial aggregation, body size, and reproductive success in the freshwater mussel. *Elliptio complanata*. *Journal of the North American Benthological Society*, **12**, 148–56.
- Dorazio, R.M. (1999) Design-based and model-based inference in surveys of freshwater mollusks. *Journal of the North American Benthological Society*, **18**, 118–31.

- Kovalak, W.P., Dennis, S.D., and Bates, J.M. (1986) Sampling effort required to find rare species of freshwater mussels. In *Rationale for Sampling and Interpretation of Ecological Data in the Assessment of Freshwater Ecosystems*, B.G. Isom (ed.), ASTM STP 84. American Society for Testing and Materials, Philadelphia, pp. 34–45.
- Ricciardi, A. and Rasmussen, J.B. (1999) Extinction rates of North American freshwater fauna. *Conservation Biology*, **13**, 1220–22.
- Salehi, M. and Seber, G.A.F. (1997) Two-stage adaptive cluster sampling. *Biometrics*, **53**, 959–70.
- Smith, D.R., Conroy, M.J., and Brakhage, D.H. (1995) Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, **51**, 777–88.
- Smith, D.R., Vilella, R.F., and Lemarié, D.P. (2001a) Survey protocol for assessment of endangered freshwater mussels in the Allegheny River. *Journal of North American Benthological Society*, **20**, 118–32.
- Smith, D.R., Vilella, R.F., Lemarié, D.P., and von Oettingen, S. (2001b) How much excavation is needed to monitor freshwater mussels? In *Freshwater Mollusk Symposium Proceedings*, R.A. Tankersley, D.I. Warmolts, G.T. Watters, B.J. Armitage, P.D. Johnson, and R.S. Butler (eds), Ohio Biological Survey, Columbus, Ohio, pp. 203–18.
- Strayer, D.L. and Ralley, J. (1993) Microhabitat use by an assemblage of stream-dwelling unionaceans (Bivalvia), including two rare species of *Alasmidonta*. *Journal of North American Benthological Society*, **12**, 247–58.
- Strayer, D.L., Sprague, S.J., and Claypool, S. (1996) A range-wide assessment of populations of *Alasmidonta heterodon*, and endangered freshwater mussel (Bivalvia:Unionidae). *Journal of the North American Benthological Society*, **15**, 308–17.
- Thompson, S.K. (1990) Adaptive cluster sampling. *Journal of the American Statistical Association*, **85**, 1050–59.
- Thompson, S.K. (1992) *Sampling*, Wiley, New York.
- Thompson, S.K. and Seber, G.A.F. (1996) *Adaptive sampling*, Wiley, New York.
- Williams, J.D., Warren, Jr., M.L., Cummings, K.S., Harris, J.L., and Neves, R.J. (1993) Conservation status of freshwater mussels of the United States and Canada. *Fisheries*, **18**, 6–22.

Biographical sketches

David Smith is a biological statistician, Rita Vilella is an ecologist, and David Lemarié is a fisheries biologist in the Aquatic Ecology Lab of the United States Geological Survey, Leetown Science Center.

Dave Smith's interest in efficient sampling methods for the study of rare and spatially clustered organisms has lead him to evaluate sampling designs for aquatic species, such as freshwater mussels and horseshoe crabs. His other research interests include the interaction between geomorphology and biota on estuarine beaches, tagging models for anadromous fish, and design of studies to assess landscape effects on biodiversity.

Rita Vilella's primary research interests are in unionid life history and population dynamics. Current research activities include evaluating environmental determinants such as stream hydraulics on unionid mussel distribution and abundance.

Dave Lemarié's research interests include the ecology of fish, aquatic insects, freshwater mussels, and salamanders. He has been active in development of survey methods for freshwater mussels, including evaluation of tagging techniques.