2000

Sample Size Requirements for In Situ Vegetation and Substrate Classifications in Shallow, Natural Nebraska Lakes

Craig P. Paukert
South Dakota State University

David W. Willis
South Dakota State University

Richard S. Holland
Nebraska Game and Parks Commission

Paukert, Craig P.; Willis, David W.; and Holland, Richard S., "Sample Size Requirements for In Situ Vegetation and Substrate Classifications in Shallow, Natural Nebraska Lakes" (2000). *Nebraska Game and Parks Commission -- Staff Research Publications*. 48.
https://digitalcommons.unl.edu/nebgamestaff/48

North American Journal of Fisheries Management 22:1329–1333, 2002

Sample Size Requirements for In Situ Vegetation and Substrate Classifications in Shallow, Natural Nebraska Lakes

CRAIG P. PAUKERT*[1] AND DAVID W. WILLIS

Department of Wildlife and Fisheries Sciences, South Dakota State University,
Brookings, South Dakota 57007, USA

RICHARD S. HOLLAND

Nebraska Game and Parks Commission,
Post Office Box 30370, Lincoln, Nebraska 68701, USA

Abstract.—We assessed the precision of visual estimates of vegetation and substrate along transects in 15 shallow, natural Nebraska lakes. Vegetation type (submergent or emergent), vegetation density (sparse, moderate, or dense), and substrate composition (percentage sand, muck, and clay; to the nearest 10%) were estimated at 25–70 sampling sites per lake by two independent observers. Observer agreement for vegetation type was 92%. Agreement ranged from 62.5% to 90.1% for substrate composition. Agreement was also high (72%) for vegetation density estimates. The relatively high agreement between estimates was likely attributable to the homogeneity of the lake habitats. Nearly 90% of the substrate sites were classified as 0% clay, and over 68% as either 0% or 100% sand. When habitats were homogeneous, less than 40 sampling sites per lake were required for 95% confidence that habitat composition was within 10% of the true mean, and over 100 sites were required when habitats were heterogeneous. Our results suggest that relatively high precision is attainable for vegetation and substrate mapping in shallow, natural lakes.

Habitat classification is commonly used to assess differences over large scales (e.g., among lakes or streams). In addition, fish population characteristics, such as abundance and growth, can often be attributed to these large-scale differences. For example, bluegill *Lepomis macrochirus* growth was reduced at high abundances of submergent macrophytes in a Wisconsin lake (Trebitz et al. 1997), and largemouth bass *Micropterus salmoides* abundance increased with increasing aquatic vegetation abundance in Texas and South Dakota impoundments (Durocher et al. 1984; Guy and Willis 1991). Nebraska sandhill lakes with higher levels of emergent vegetation typically contained higher-quality bluegill populations (Paukert et al. 2002). Therefore, vegetation assessment is evidently needed for effective management of fisheries in lentic systems.

Habitat assessment techniques are often designed to optimize time rather than accuracy (Platts et al. 1983) and therefore commonly employ subjective visual estimates. Previous work on visual estimation of stream habitat attributes has indicated that observer experience, habitat complexity, and habitat classification scale all affect precision (Platts et al. 1983; Roper and Scarnecchia 1995; Wang et al. 1996). However, research on precision of habitat classification has focused on lotic habitats and, to our knowledge, has not been assessed in lentic habitats.

Our objective was to determine the precision of visually-estimated vegetation and substrate classifications in shallow, natural lakes. We focused on the classification of vegetation and substrate to allow assessment of fish community relationships. Therefore, we were more concerned with large-scale precision (e.g., the coverage of dense emergent vegetation within a lake) than with microhabitat-scale precision. We wanted to determine whether our method was sufficiently precise to warrant its use as a rapid and effective vegetation and substrate assessment that fisheries and watershed managers could complete with limited time and experience.

Methods

Fifteen lakes in the sandhills of north-central Nebraska were selected for this study. All lakes were shallow (mean depth range = 1.2–3.0 m), had regular shorelines (shoreline development index range = 1.2–2.6), and were relatively clear (Secchi disk transparency mean = 1.2 m; range = 0.4–2.5 m) (Paukert and Willis 2000). Most lakes were small, with a mean lake area of 149 ha); five lakes were smaller than 50 ha, and the largest lake was 907 ha (Paukert and Willis 2000).

---

* Corresponding author: cpaukert@usgs.gov

[1] Present address: Southwest Biological Center, U.S. Geological Survey, 2255 North Gemini Drive, Flagstaff, Arizona 86001, USA.

Received December 4, 2001; accepted April 1, 2002

To estimate vegetation and substrate classification, we established 5–10 equally-spaced transects perpendicular to the longest axis within each of the 15 lakes in July 1999. We used five transects in lakes with areas of 50 ha or less, with sampling stations 50-m apart along each transect. For 51–100-ha lakes, sampling stations were spaced at 100-m intervals along seven transects, and for 101-ha and larger lakes, sampling was conducted at 100-m intervals along 10 transects. The first sampling location along each transect was randomly selected as 10, 20, 30, 40, or 50 m from shore. The distance between points was measured with a tape, and a compass was used to navigate along the transects. At each station, the boat was stabilized either by anchors or, when wind conditions permitted, by use of the outboard motor. Two independent observers conducted the habitat sampling, and worked from the same side of the boat at each site; however, the sampled side varied between sites.

Habitat at each sampling site was categorized as either vegetated or open water prior to sampling. Therefore, we assumed 100% agreement between observers for the initial categorization of vegetated or open water. Substrate was not sampled at sites where vegetation was sampled and vice versa. Consequently, we only estimated substrate composition in open-water areas and not within vegetation beds. When the site was classified as open water, substrate was collected with an Ekman dredge (McMahon et al. 1996). The percentages of muck (i.e., organic matter), sand, clay, and other material (e.g., detritus, cobble, etc.) in the substrate sample were visually identified to the nearest 10%. For vegetated sites, we classified the vegetation as emergent (foliage extending above water surface; i.e., cattail *Typha* spp., reed *Phragmites* spp., and bulrush *Scirpus* spp.) or submergent (anchored to bottom by roots or rhizomes, and foliage primarily submerged; i.e., pondweed *Potamogeton* spp., water milfoil *Myriophyllum verticillatum*, and coontail *Ceratophyllum demersum*) (Larson 1993). If a site contained mixed beds of submergent and emergent vegetation, the dominant vegetation type was used to classify that site. Only three lakes had floating vegetation (i.e., yellow water lily *Nuphar variegata*); therefore, this vegetation type was excluded from further analysis. In addition, vegetation density was visually estimated as either dense (plant stems $<$ 10 cm apart, on average), moderate (stems $\geq$ 10 cm and $<$ 30 cm apart), or sparse (stems $>$ 30 cm apart). All measurements were made in a 1-m^2 area surrounding the transect point.

The two observers independently collected all measurements. Both observers worked in the same boat, recorded their own data sheets, and observed the same vegetation and substrate samples. Prior to the study, the observers discussed vegetation and substrate categories and therefore had similar training (e.g., Roper and Scarnecchia 1995). However, neither observer had any specific training on vegetation and substrate measurements prior to the study.

The percentage of vegetation coverage for each lake was estimated by dividing the total number of sites with the dominant vegetation category (e.g., sparse emergent, dense submergent, etc.) by the total number of sites sampled in the lake and multiplying this value by 100. We also determined the percentage of observations that were in agreement between the two observers for each vegetation type or substrate site sampled. Estimates of substrate agreement were calculated as the absolute value of the difference between the first observer's estimate and the second observer's estimate. Because vegetation density was classified as sparse, moderate, or dense for the dominant vegetation type, we calculated the observer agreement rate for the individual vegetation density categories.

For each vegetation type and density category, Pearson's product-moment correlations were used to determine the association between the vegetation coverage estimates made by the two observers for all 15 lakes. In addition, correlations were used to determine the relation between the estimates made by the two observers for each substrate category for all sites where substrate was sampled. The sample size required for 95% confidence that vegetation coverage fell within 10% of the true mean value was estimated for each lake with sample size equations for proportions (Krebs 1989).

Results

A total of 644 sites were sampled among the 15 lakes, including 312 substrate sites and 332 vegetated sites. The dominant vegetation type was similarly classified by both observers 92% of the time. On only 26 occasions (8%) did the observers differ in their estimates of dominant vegetation type. With regard to vegetation density, estimates by the two observers were in 71% agreement, whereas 27% of the estimates disagreed only by one category level (e.g., sparse versus moderate).

For the substrate descriptions of sites, the observers had relatively high agreement for sand (74.7%), muck (62.5%), and clay (90.1%). Ob-

TABLE 1.—Pearson's product-moment correlations (r = coefficient) between estimates of vegetation coverage, substrate type and vegetation density, determined by two independent observers in Nebraska Sandhill lakes in 1999. For vegetation, 332 sites were sampled; fifteen lakes were sampled for vegetation estimates; 312 sites within these lakes were sampled for substrate estimates.

Habitat type	Correlation		Range (% of coverage)	
	r	*P*	Observer 1	Observer 2
Emergent vegetation				
Sparse	0.80	<0.001	0.0–19.5	0.0–14.6
Moderate	0.42	0.11	0.0–5.1	0.0–7.1
Dense	0.86	<0.001	0.0–5.2	0.0–5.2
Emergent vegetation combined	0.86	<0.001	0.0–22.0	0.0–21.4
Submergent vegetation				
Sparse	0.84	<0.001	0.0–34.2	0.0–34.2
Moderate	0.96	<0.001	0.0–35.9	0.0–48.7
Dense	0.99	<0.001	0.0–74.5	0.0–64.7
Submergent vegetation combined	0.99	<0.001	0.0–97.1	0.0–97.1
Substrate				
Sand	0.91	<0.001	0–100	0–100
Clay	0.56	<0.001	0–50	0–80
Muck	0.90	<0.001	0–100	0–100

servers differed by over 20% in only 5.4% of estimates for sand, 9.8% for muck, and 4.1% for clay. However, substrate was relatively homogeneous for all the lakes. The first observer classified 68% of all substrate categories as 0% sand, whereas the second observer classified 66% as 0% sand. The first observer also classified 54.5% of all substrate sites as 100% muck, whereas the second observer classified 50.3% of sites as 100% muck. A between-observer mean percentage composition of 0% accounted for 89.8% of clay observations, and 53.7% of muck observations and 68.4 % of sand observations averaged either 0% or 100%. Only 7.3–9.9% of observations (depending on observer) were classified as 30–70% sand, and only 13.8–15.3% of observations were classified as 30–70% muck.

In general, estimates for vegetation and substrate classifications were highly correlated between observers (Table 1). For vegetation density categories of submergent and emergent vegetation types, observer estimates were highly correlated. Moderate emergent vegetation had a limited range for both observers, which likely caused the relatively poor correlation. The maximum observer difference was only 16.0% for submergent vegetation (the between-observer mean was 56.0%) and only 7.7% for emergent vegetation (the among-observer mean was 3.8%). Although substrate categories had a finer scale (i.e., increments of 10%), correlations between observer estimates were high. Clay, which composed less than 50% of the substrate at any site, had the lowest correlation coefficient, but the relation was still highly significant.

The number of sample sites required to produce 95% confidence that vegetation coverage was within 10% of the true mean varied by vegetation type and coverage. The required sample size was highest at intermediate vegetation coverages for submergent vegetation (Figure 1). The required sample size for emergent vegetation increased with increasing vegetation coverage, but maximum coverage was only 17% (Figure 1). When submergent vegetation coverage was less than 20%, 40 sites sampled per lake was sufficient; however, when submergent vegetation coverage approached 50%, about 100 locations per lake was required (Figure 1). For emergent vegetation, close to 40 sample sites were sufficient when emergent vegetation was less than 10%. At the highest level of emergent vegetation, 60 samples were needed for 95% confidence that vegetation coverage was within 10% of the true mean.

Discussion

Between-observer precision of vegetation and substrate classifications was relatively high in these shallow, natural lakes. Observers agreed closely (92% of the time) when evaluating the dominant vegetation type as either emergent or submergent. Even with the finer scale of vegetation density, agreement between observers was still 71%. Precision of habitat classification has been studied in smaller streams (e.g., Wright et al. 1981; Platts et al. 1983; Simonson 1993; Roper and Scar-

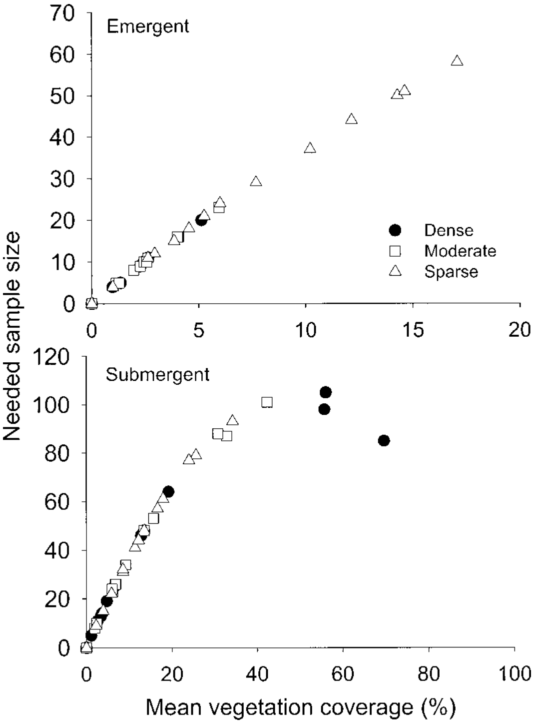

FIGURE 1.—Relation between mean vegetation coverage for three density categories of emergent (top panel) and submergent (bottom panel) vegetation and the number of samples per lake required for 95% confidence that vegetation coverage was within 10% of the true mean.

necchia 1995; Wang et al. 1996; Dare and Hubert 2000), but no such studies have concentrated on lentic environments. Our results suggest that high precision in lake habitat classification is possible with this relatively simple system.

One criticism of subjective habitat measurements is that observer bias is high and precision and repeatability are low (Poole et al. 1997). Only two observers were involved in our study; however, though neither observer had formal training in habitat classification, the two attained high levels of agreement. Wang et al. (1996) also noted that observer experience had little effect on precision. Our results suggest that untrained observers can attain high precision with this rapid habitat classification technique.

Our vegetation coverages were visually estimated, which typically results in low agreement (Platts et al. 1983; Dare and Hubert 2000). However, the high agreement between observers using visual estimates in our study may be related to scale. We used only three vegetation density categories and only two vegetation types. Roper and Scarnecchia (1995) found that variation among observers in stream habitat classification was due to (1) the level of definition required for classification, (2) the level and uniformity of observer training, and (3) stream channel characteristics. The low number of habitat types and low definition (i.e., three vegetation density categories) may have produced our high precision. Substrate composition was described with more levels (i.e., 10% increments) and exhibited slightly lower agreement. However, at both a coarse scale (vegetation density) and a slightly finer scale (substrate composition), observer agreement was relatively high.

Sample sizes required for high precision were less than 100 sites per lake; 40 sites per lake provided sufficient samples in lakes with homogenous vegetation and substrate. We typically sampled about 30–50 sites in each of two lakes per 8-h day, which included substantial travel time to the lakes. Therefore, up to two lakes per day could be sampled with high precision based on the technique described here. Our results suggest that inexperienced observers can attain relatively high precision in acquiring coarse-scale vegetation and substrate estimates in shallow, natural lakes.

Acknowledgments

We would like to thank P. Chvala and B. Heikes for field assistance. Valentine National Wildlife Refuge allowed access to some study lakes. The project was funded by the Nebraska Game and Parks Commission through Federal Aid in Sport Fish Restoration Project Number F-118-R. The comments received from D. Perleberg and one anonymous reviewer substantially improved the manuscript. This paper was approved for publication by the South Dakota Agricultural Experiment Station as journal series 3283.

References

Dare, M. R., and W. A. Hubert. 2000. Precision and interpretation of data collected using a new measurement technique for microhabitat features at fish locations determined by radio telemetry. Journal of Freshwater Ecology 15:29–38.

Durocher, P. P., W. C. Provine, and J. E. Kraii. 1984. Relationship between abundance of largemouth bass and submergent vegetation in Texas reservoirs. North American Journal of Fisheries Management 4:84–88.

Guy, C. G., and D. W. Willis. 1991. Relationship between environmental variables and density of largemouth bass in South Dakota ponds. Proceedings of the South Dakota Academy of Science 70:109–117.

Krebs, C. J. 1989. Ecological methodology. Harper Collins, New York.

Larson, G. E. 1993. Aquatic and wetland vascular plants in the northern Great Plains. U.S. Forest Service General Technical Report RM-238.

McMahon, T. E., A. V. Zale, and D. J. Orth. 1996. Aquatic habitat measurements. Pages 83–120 *in* B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, Maryland.

Paukert, C. P., and D. W. Willis. 2000. Factors influencing panfish populations in Nebraska sandhill lakes. Nebraska Game and Parks Commission, Federal Aid in Sport Fish Restoration, Project F-118-R, Job 1, Study 1, Completion Report, Lincoln.

Paukert, C. P., D. W. Willis, and J. A. Klammer. 2002. Effects of predation and environment on quality of yellow perch and bluegill populations in Nebraska sandhill lakes. North American Journal of Fisheries Management 22:86–95.

Platts, W. S., W. F. Megahan, and G. W. Minshall. 1983. Methods for evaluating stream, riparian, and biotic conditions. U.S. Forest Service General Technical Report INT-138.

Poole, G. C., C. A. Frissell, and S. C. Ralph. 1997. Instream habitat unit classification: inadequacies for monitoring and some consequences for management. Journal of the American Water Resources Association 33:879–896.

Roper, B. B., and D. L. Scarnecchia. 1995. Observer variability in classifying habitat types in stream surveys. North American Journal of Fisheries Management 15:49–53.

Simonson, T. D. 1993. Correspondence and relative precision of stream habitat features estimated at two spatial scales. Journal of Freshwater Ecology 8: 363–372.

Trebitz, A. S., S. Carpenter, P. Cunningham, B. Johnson, R. Lillie, D. Marshall, T. Martin, R. Narf, T. Pellett, S. Stewart, C. Storlie, and J. Unmuth. 1997. A model of bluegill–largemouth bass interactions in relation to aquatic vegetation and management. Ecological Modelling 94:139–156.

Wang, L., T. D. Simonson, and J. Lyons. 1996. Accuracy and precision of selected stream habitat estimation. North American Journal of Fisheries Management 16:340–347.

Wright, J. F., P. D. Hiley, S. F. Ham, and A. D. Berrie. 1981. Comparison of three mapping procedures developed for river macrophytes. Freshwater Biology 11:369–379.