RAPID COMMUNICATION / COMMUNICATION RAPIDE

Determination of optimal riparian forest buffer dimensions for stream biota–landscape association models using multimetric and multivariate responses

Emmanuel A. Frimpong, Trent M. Sutton, Kyoung J. Lim, Peter J. Hrodey, Bernard A. Engel, Thomas P. Simon, John G. Lee, and Dennis C. Le Master

Abstract: The dimensions of riparian buffers selected for stream biota–landscape association models determine correlation strength and subsequent model interpretation. Efforts have been made to optimize buffer dimensions incorporated into models, but none has explicitly determined a single optimum based on both longitudinal and lateral buffer dimensions. We applied partial correlation and multivariate linear regression on functional fish community response attributes and the index of biotic integrity using stream samples (N = 107) from the Eastern Corn Belt Plain Ecoregion of Indiana, USA. Land-cover data in digital format were processed in geographic information systems for an area covering 300 m on either side of selected streams and within 2000 m longitudinally. The optimal buffer dimension for the study area was 30 m laterally and 600 m longitudinally, with a partial correlation of 0.29 (P = 0.002), and there was agreement in the partial correlation and multiple regression models. The longitudinal dimension was more conclusively determined, but the lateral dimension was optimum only with respect to the resolution of the land-use data used. Based on these results, we propose the use of this approach to optimize the riparian buffer parameter in landscape models.

Résumé : La taille des bandes de protection retenue dans les modèles d'association des organismes et des paysages dans les cours d'eau détermine la force de la corrélation et l'interprétation subséquente du modèle. On a déployé beaucoup d'efforts afin d'optimiser la taille des bandes de protection utilisée dans les modèles, mais on n'a pas déterminé de façon explicite un optimum unique basé à la fois sur les dimensions longitudinales et latérales de la bande. Nous avons analysé à l'aide de corrélations partielles et de régressions linéaires multiples les caractéristiques des réponses fonctionnelles des communautés de poissons et les indices d'intégrité biotique dans des échantillons (N = 107) récoltés dans des cours d'eau de l'écorégion de la plaine de l'est de la ceinture de maïs en Indiana, États-Unis. Nous avons traité les données en format digital de couverture végétale à l'aide de systèmes d'information géographique sur une largeur de 300 m sur chaque rive des cours d'eau sélectionnés et sur une distance longitudinale de 2000 m. La taille optimale de la zone de protection dans la zone d'étude est de 30 m de largeur et de 600 m de longueur; la corrélation partielle est de 0,29 (P = 0,002) et les modèles de corrélation partielle et de régression multiple sont en accord. La dimension longitudinale est fixée de façon plus certaine, alors que la dimension latérale est optimale seulement compte tenu de la résolution des données d'utilisation des terres que nous avons employée. D'après ces résultats, notre recommandation est d'utiliser notre méthodologie pour déterminer les caractéristiques optimales des bandes de protection dans les modèles de paysages.

[Traduit par la Rédaction]

Received 23 September 2004. Accepted 24 January 2005. Published on the NRC Research Press Web site at http://cjfas.nrc.ca on 11 February 2005.
J18316

E.A. Frimpong, T.M. Sutton,[1] P.J. Hrodey, and D.C. Le Master. Department of Forestry and Natural Resources, Purdue University, 715 W. State Street, West Lafayette, IN 47906, USA.
K.J. Lim and B.A. Engel. Department of Agricultural and Biological Engineering, Purdue University, 225 S. University Street, West Lafayette, IN 47907, USA.
T.P. Simon. US Fish and Wildlife Service, Bloomington Field Office, 620 S. Walker Street, Bloomington, IN 47403, USA.
J.G. Lee. Department of Agricultural Economics, Purdue University, 403 W. State Street, West Lafayette, IN 47907, USA.

[1]Corresponding author (e-mail: tsutton@purdue.edu).

 doi: 10.1139/F05-020

Introduction

The increased recognition of landscape ecological concepts, in particular, the effects of scale on observed ecological phenomena (Wiens 1989), has resulted in a different view for stream ecology research. A number of investigations over the past two decades involving stream biota–habitat associations have extended beyond the stream channel to present a correlation of land cover in the watershed with either physical habitat or biotic variables measured at sample locations (e.g., Steedman 1988; Townsend et al. 2003; Wang et al. 2003). To demonstrate the influence of scale, many studies have compared the relative amount of variation in reach biotic or physical characteristics explained by land cover in the entire watershed with land cover somewhere in the vicinity of the sample location. While watershed land cover can be accurately quantified, the definition of an area within a watershed often is subject to ambiguity.

In defining the watershed area within the vicinity of a sampled reach, there are two potential approaches. The first approach considers land cover in the watershed within a certain radius of the sample location. However, a variation of this approach does not use a fixed radius but considers land cover in the area above a sample location and below the next sample location for sites nested in a larger watershed (e.g., Steedman 1988). The second approach considers land cover within a certain fixed distance from the stream channel, often called the riparian scale, which may be examined for a certain number of distances (e.g., Wang et al. 2003). Despite the recognition of the importance of both lateral and longitudinal dimensions of riparian areas, optimum dimensions have not been determined in a combined lateral and longitudinal approach. In reported studies that have looked at correlation of varying forested buffer areas with biotic attributes, one dimension of the buffer is usually fixed while the other is varied without explicitly stating the reason for choosing the fixed dimension. Changing the riparian area dimensions in a model may lead to increased or decreased correlation between forested riparian area and biotic attributes, which could alter model interpretation.

Among landscape ecologists who study streams, it is an unresolved question whether whole watershed or riparian land cover has a greater influence on streams and their biota (Rabeni and Sowa 2002; Wang et al. 2003; Allan 2004). To attempt to answer this question, the dimensions of riparian area used in analysis cannot be chosen arbitrarily. Modelers intending to use forested buffer as an explanatory variable for variation in stream biotic attributes can begin with a heuristic search using plausible subwatershed areas to determine the optimum dimension for the geographic area where the model is to be applied. This is a way to define riparian area if the relative influence of riparian and watershed land cover, for example, are to be compared. Because watersheds differ in size, a logical approach to finding a common riparian dimension is to remove the influence of size from the response variable. Alternatively, an interaction of watershed size with riparian forest influence may be assumed to exist and be incorporated in statistical tests. The objectives of this study were (*i*) to determine the lateral and longitudinal dimensions of forest buffer that is most correlated with the index of biotic integrity and its metrics in the Eastern Corn Belt Plain Ecoregion of Indiana, and (*ii*) to demonstrate how the methods used can be generalized to other landscapes and similar modeling efforts involving stream biota – land cover relationships.

Methods

Study area and sampling

We analyzed the association between percent of natural vegetation (called riparian forest or buffer hereafter) and fish assemblage characteristics for 107 stream locations within the approximately 44 000 km^2 Eastern Corn Belt Plain (ECBP) Ecoregion of Indiana. Land use in the ECBP Ecoregion is greater than 75% row-crop agriculture in a predominantly corn – soy bean rotation. Fifty-seven fish assemblage samples were collected by one-pass backpack electrofishing from 1990 through 1994, and 50 additional samples were collected in 2002 and 2003 from reaches that ranged in length from 15 to 20 mean stream widths from June to September. Sampled streams were mostly low-gradient, first- to fifth-order streams, with 0.06%–1.92% main channel slopes (0.23% average) and watershed size ranging from 3.1 to 269.5 km^2. Fish collected were counted, identified to species, and classified into trophic, reproductive, and sensitivity guilds following Simon and Dufour (1998). The multimetric index of biotic integrity (IBI) scores were computed for sites using metric expectations developed by Simon and Dufour (1998). Nine IBI metrics that were correlated with various dimensions of forested buffer were selected to be included in a multivariate model. The selected metrics were number of sensitive species, number of sucker (family Catostomidae) species, number of headwater species, number of darter (Etheostomatini), madtom (*Noturus*), and sculpin (*Cottus*) species, percent abundance of tolerant individuals, proportion of pioneer species, proportion of omnivores, proportion of simple lithophilic spawners, and catch-per-unit-effort (CPUE). The remaining IBI metrics were species richness, number of darter (Etheostomatini) species, number of sunfish species, proportion of insectivorous individuals, number of minnow species, proportion of carnivorous individuals, and proportion of individuals with deformities, eroded fins, lesions, and tumors.

Spatial data computations

United States Census Bureau streams were imbedded into the US Geological Survey (USGS) 30-m digital elevation model (DEM) and used to delineate watersheds for all sample locations and generate stream networks using the Watershed Delineator extension in ArcView 3.3 (ESRI, Redlands, California). We used a threshold area of approximately 0.5 km^2 to define a stream. To determine land cover in watersheds near the time of sampling, the following two sources of 30-m resolution digital land-cover data were used: the National Land Cover Data (sites sampled from 1990–1994) and the National Agricultural Statistical Service crop-use data (sites sampled in 2002 and 2003). Land-cover grids were reclassified as forest, agricultural, or developed and converted to vector data. Around each sample location, 10 concentric buffer polygons established 200 m apart were created, and an additional 10 buffer polygons were created 30 m apart around the entire stream network. These two sets of polygons were intersected together with the vectorized land-cover data and watershed boundary, thereby enabling the computation of the cumulative percent of forest in each of 100 combinations of buffer dimensions within each watershed. The buffer dimensions were up to 2000 m in steps of 200 m longitudinally upstream and up to 300 m in steps of 30 m laterally on either side of the stream. It was believed that this range covered the optimum buffer dimension based on studies such as Barton et al. (1985) and Steedman (1988). Because of the large amount of data extractions involved, most of the overlay and computations were automated in ArcView using customized Avenue scripts.

Statistical models

Fish assemblage – riparian buffer associations were investigated by two underlying statistical models, with all analyses performed in SAS® 8.02 (SAS Institute Inc., Cary, North Carolina). The first model was a first-order partial correlation of IBI and percent forested buffer evaluated by partialing out the effect watershed area. The partial correlation model is described by Neter et al. (1996) and the analysis was available in SAS through the PARTIAL statement of the CORR procedure. In the application of this model to three variables, Y_1 was IBI score at the sample location, Y_2 was $\log_{10}$(1 + percent forested buffer of a given lateral and longitudinal dimension), and Y_3 was $\log_{10}$(watershed size). The dimensions of the riparian buffer used to compute Y_2 were varied to find the combination that maximized the partial correlation coefficient.

The second model was a multivariate linear regression in which simultaneous significance tests on the relationship between explanatory variables and more than one response variable may be obtained. The simultaneous test is only as good as the common trait of the system being modeled by the response variables together (Finn 1974; Johnson and Wichern 2002). For this model, consider m responses Z_1, $Z_2,\ldots,Z_m$ and a single set of predictor variables Y_1, $Y_2,\ldots,Y_r$. The multivariate linear regression model is

$$(1)\qquad \underset{(n\times m)}{\mathbf{Z}} = \underset{(n\times(r+1))}{\mathbf{Y}}\ \underset{((r+1)\times m)}{\boldsymbol{\beta}} + \underset{(n\times m)}{\boldsymbol{\epsilon}}$$

where $\mathbf{Z}$ is a matrix of response variables, $\mathbf{Y}$ is the matrix of predictors, $\boldsymbol{\beta}$ is the matrix of parameters, and $\boldsymbol{\epsilon}$ is a matrix of error terms. Each Z_i follows the linear regression model. Finn (1974) provides the theory of the test of simultaneous hypotheses on rows of $\boldsymbol{\beta}$. These tests were available in SAS through the MTEST statement in the REG procedure. The model was applied in this study using nine response variables for each of 107 sites. For the $\mathbf{Z}$ matrix, the variable Z_1 was the number of sensitive species, Z_2, the number of sucker (family Catostomidae) species, Z_3, the number of headwater species, Z_4, the number of darter (Etheostomatini), madtom (*Noturus*), and sculpin (*Cottus*) species, Z_5, the percent abundance of tolerant individuals, Z_6, the proportion of pioneer species, Z_7, the proportion of omnivores, Z_8, the proportion of simple lithophilic spawners, and Z_9, CPUE. The $\mathbf{Y}$ matrix had the following three variables: Y_2 and Y_3 (defined in the previous paragraph) and the multiplicative term Y_2Y_3 representing the interaction between forested buffer and watershed area. The hypotheses that the parameters for the predictors Y_2 and Y_2Y_3 are zero were tested simultaneously. The interpretation of the hypotheses together is that neither riparian forest nor its interaction with watershed area has a significant relationship with any of the response variables Z_1 to Z_9. The dimensions of the riparian buffer were varied to find the region of minimum probability of type-I error for rejecting the simultaneous hypothesis.

Response visualization

Because the riparian buffer was two-dimensional, the response of fish community attributes to varying buffer dimensions was best visualized as a surface. No detailed theory was applied in fitting the response surface. Simple linear interpolations of the partial correlation coefficient (for the first model) and the probability of a type-I error (for the second model) were performed. Procedures described by Cohen et al. (1998), and available through SAS/INSIGHT® (SAS Institute Inc.), provided quick contour plots to visually determine the approximate optimal buffer dimensions for the study area. The graphing software SigmaPlot® (Systat Software Inc., Point Richmond, California) provided more editing functionality and was used to create the plots included in this report.

Results and discussion

The IBI scores ranged from 16 to 56 with an average of 38 (maximum score is 60), and the percent of forested area in buffers ranged from 0% to 100% in all combinations of lateral and longitudinal distances. The partial correlation between IBI scores and the percent of forested riparian buffer was maximized at 600 m in the longitudinal (upstream) direction and peaked laterally at closest to the stream with a partial correlation of 0.29 (P = 0.002) (Fig. 1). The multivariate test also had a minimum probability of type-I error centered on 600 m longitudinally and below 60 m (Fig. 2). If a 30 m or 60 m × 600 m buffer dimension is selected, the probability of a type-I error is less than 1% if we reject the hypothesis that neither percent forested buffer nor its interaction with watershed area has a significant relationship with any of the nine fish assemblage attributes. The lateral pattern of partial correlation between 30 m and 300 m shows that at any fixed longitudinal dimension, the lateral dimension nearest the stream has the strongest correlation. However, the change in correlation is not linear and decreases steeply away from the stream, leveling off at 150 m. At a fixed interval, the statistical significance of the difference in correlation between two lateral dimensions was improbable the greater the distance they are from the stream.

Considering the behavior of the partial correlation and the probability of a type-I error around the optimum point, it is unlikely that there are other optima within the watershed. Barton et al. (1985) similarly found that the optimal stream bank distance upstream needed to explain weekly mean temperature and the distribution of trout species in southern Ontario streams was approximately 1 km. This trend is also consistent with Steedman's (1988) observation that partial basin land use provided better predictive power of regression models for IBI, particularly with increasing mean watershed size. It appears that for agricultural areas, the influence of forest on fish communities is localized. The optimum buffer dimension based on our data resolution was approximately

Fig. 1. Response of the partial correlation coefficient for the index of biotic integrity (IBI) with percent forested riparian buffer for a given lateral and longitudinal buffer dimensions upstream of sample locations. The correlation was evaluated by partialing out the effect of watershed size on IBI for 107 first- to fifth-order streams of the Eastern Corn Belt Plain Ecoregion of Indiana, USA.

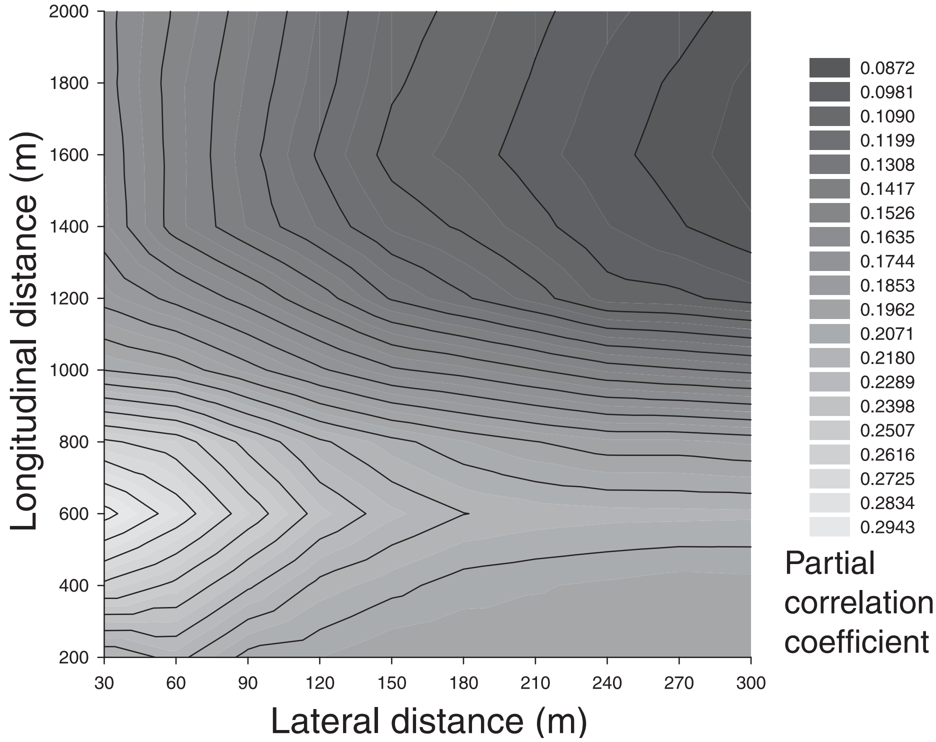
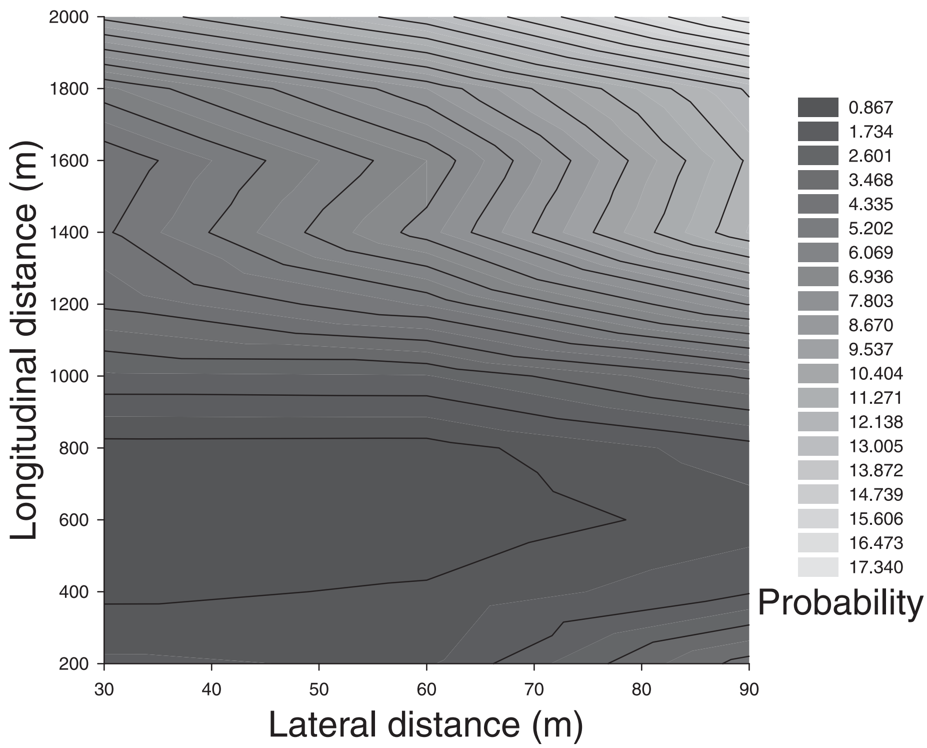


Fig. 2. Probability of type-I error for rejection of the null hypothesis that neither the percent forested riparian buffer nor its interaction with watershed size has a relationship with nine fish community attributes for a range of lateral and longitudinal buffer dimensions upstream of sample locations. The 107 streams of the Eastern Corn Belt Plain Ecoregion of Indiana, USA, varied from first to fifth order.



30 m × 600 m. Below this optimum, insufficient information has been included, and beyond the optimum, too much information has been included, with both leading to weaker correlations or higher probability of error in the choice of buffer dimensions. There is a caveat in our choice of the lateral optimum in that we do not know what the trend would

be laterally at dimensions less than 30 m. It is tempting to extrapolate the lateral trend to the stream and suggest that correlations should increase to the stream bank. Aerial photos would permit such fine-resolution investigations. In this study, we used land cover at the time of sampling; however, similar analyses could be performed with historical land cover if that is deemed more relevant to the model at stake. We obtained similar results using both multimetric and multivariate response variables, but the multivariate approach is considerably more complex. In real-world applications, the multivariate approach will be more useful when more than one biotic variable or combinations of variables (for example, IBI and an invertebrate community index or a mixture of metrics from these two indices) are under consideration for a single model.

It should be pointed out that while this method leads to identification of the zone of highest correlation between some landscape variable and stream biotic attributes, the variation accounted for by a given landscape variable may be small and require that additional or alternative landscape attributes be investigated. For instance in this study, the 30 m × 600 m forested buffer accounted for only about 10% of the variation in IBI at the ecoregional scale. In a related study, another landscape variable, a combined lengths and steepness of slopes factor, within a 300 m × 1000 m zone accounted for more variation in IBI than forested buffer in the ECBP Ecoregion. The 30 m × 600 m forested buffer assumed greater importance in smaller watersheds within the ecoregion, accounting for 50%–60% of the variation in IBI in watersheds one-third to one-quarter of the size of the ECBP Ecoregion, where surficial geology was more homogenous. We observed that as the spatial extent of a model was expanded, less biotic variation was accounted for by forested buffer. We attribute this to increased heterogeneity of surficial geology over a large area.

This study employed a new approach to modeling stream biota – land use association. The methodology is applicable to modeling in any geographic location and any biotic response. Stream landscape models have been developed based on ecoregions or watersheds as the spatial unit and response variables have been predominantly fish and macroinvertebrate community attributes. Political boundaries are less meaningful as spatial units, and stream physical and chemical habitat variables are not the ultimate response following landscape modification. Riparian landscape alteration causes changes in the abundance and composition of the aquatic biota, mediated by the changes in physical and chemical habitat and energy sources over what appears to be a short distance (about a kilometre or less). Forested reaches typically have cooler temperatures, wider channels, and fewer sediments (Allan 2004). The effectiveness of the ecosystem functions performed by riparian vegetation is affected by landform (slopes, geology, etc.) and location (Rabeni and Sowa 2002; Allan 2004). The question of where and how much of landscape alteration will cause a hypothesized change in fish or macroinvertebrate community attributes can be explored using the methods of this study. Researchers interested in specific functional groups can focus on those groups. It is generally agreed that different stream fauna respond to landscape processes at different scales. Our methodology provides a means for evaluating different species and functional groups in different landscape settings to corroborate this generally held view. To apply this method, all one needs is spatially referenced biological sample locations and digital land-cover and elevations data. Large sample sizes are recommended owing to the inherent variability in biological systems. The choice of response variables should be determined by the management question posed. Specific species – land cover relations would be appropriate in systems dominated by one or a few species but will have limited practical use in multispecies systems. This method will be more useful in landscapes dominated by one of two land-cover types, agricultural or natural vegetation, where physical processes linking landscapes and streams are diffuse and effects of vegetation or agricultural runoff decline with distance from the upstream source. Our specific findings apply to the ECBP Ecoregion and perhaps similar agricultural landscapes where forests are located predominantly in floodplains.

Acknowledgements

We wish to thank the Indiana Department of Environmental Management for providing the 1990–1994 biological dataset and the Purdue University Department of Agricultural and Biological Engineering for providing all spatial data. We also thank Cameron Guenther for his field support with sampling and fish identification. This project, funded through the Purdue University Department of Forestry and Natural Resources, was approved for publication as manuscript 2005-17567 by the Purdue University Agricultural Research Program.

References

Allan, J.D. 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annu. Rev. Ecol. Evol. Syst. **35**: 257–284.

Barton, D.R., Taylor, W.D., and Biette, R.M. 1985. Dimensions of riparian buffer required to maintain trout habitat in southern Ontario streams. N. Am. J. Fish. Manag. **5**: 364–378.

Cohen, M-D., Chen, H., Yuan, Y., and Wicklin, F. 1998. New Features in SAS/INSIGHT® in version 7. Applications Division, SAS Institute Inc., Cary, North Carolina.

Finn, J.D. 1974. A general model for multivariate analysis. Holt, Rinehart, and Winston, New York.

Johnson, R.A., and Wichern, D.W. 2002. Applied multivariate statistical analysis. 5th ed. Prentice Hall, Upper Saddle River, New Jersey.

Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. 1996. Applied linear statistical models. 4th ed. McGraw-Hill, New York.

Rabeni, C.F., and Sowa, S.P. 2002. A landscape approach to managing the biota of streams. *In* Integrating landscape ecology into natural resources management. *Edited by* J. Liu and W.W. Taylor. Cambridge University Press, Cambridge, UK. pp. 114–143.

Simon, T.P., and Dufour, R. 1998. Development of index of biotic integrity expectations for the ecoregions of Indiana: V. Eastern Corn Belt Plain. US Environmental Protection Agency (EPA), Region 5, Chicago, Illinois. EPA 905-R-96-004.

Steedman, R.J. 1988. Modification and assessment of an index of biotic integrity to quantify stream quality in southern Ontario. Can. J. Fish. Aquat. Sci. **45**: 492–501.

Townsend, C.R., Doledec, S., Norris, R., Peacock, K., and Arbuckle, C. 2003. The influence of scale and geography on relationships

between stream community composition and landscape variables: description and prediction. Freshw. Biol. **48**: 768–785.
Wang, L., Lyons, J., Rasmussen, P., Seelbach, P., Simon, T., Wiley, M., Kanehl, P., Baker, E., Niemela, S., and Stewart, P.M. 2003. Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest Ecoregion, U.S.A. Can. J. Fish. Aquat. Sci. **60**: 491–505.
Wiens, J.A. 1989. Spatial scaling in ecology. Funct. Ecol. **3**: 385–397.