American Fisheries Society Symposium 48:xx–xx, 2006

Hindcasting Reference Conditions in Streams

Bruce W. Kilgour*

Stantec Consulting Ltd., 1505 Laperriere Avenue
Ottawa, Ontario, K1Z 7T1, Canada

Les W. Stanfield

Ontario Ministry of Natural Resources
Lake Ontario Salmonid Unit
Picton, Ontario K0K 2T0, Canada

Abstract.—Assessments of stream fish or benthos assemblages normally involve a contrast of conditions at "test" sites to conditions represented by "regional" reference sites that are either minimally or least disturbed. Identification of reference sites is difficult and normally involves a variety of subjective criteria. The development of reference models for stream fish and benthos in the Canadian tributaries of Lake Ontario is particularly challenging because there are few undeveloped areas and there is no consensus on criteria for a least-disturbed condition. Rather than identify sites as representing a least-disturbed condition, we developed a series of models that relate the existing biophysical condition of streams (i.e., the fish, benthos, and instream habitat) to landscape (i.e., slope, geology, catchment area) and land use/land cover (percent impervious cover [PIC]). Relationships between indices of biophysical condition and PIC can be used to "hindcast" or estimate the expected biophysical condition at a variety of land cover scenarios. The models cannot be used to predict conditions outside the calibration data range, but this approach does allow us to make use of a disturbance gradient and make predictions with a minimal number of least-disturbed sites. The difference between the hindcast reference and present day conditions is an estimate of present-day impacts. Results from this exercise provided an estimate of the magnitude of impairment of streams in the Canadian portion of the Lake Ontario region.

INTRODUCTION

Ecological monitoring is required in order to understand if human-related stressors have undue influence on environmental resources. In aquatic systems, fish and benthic macroinvertebrate assemblages are often used as monitoring endpoints (Karr and Chu 1999). Data from reference sites are typically used to judge the degree of impairment of conditions in "test" sites that are physically or chemically disturbed as a result of human activity. Where there are large differences in biophysical conditions between reference and test sites, test sites are deemed impaired (Environment Canada 1998; Bailey et al. 2003).

Impact assessments historically involved the comparison of conditions at one or a few reference sites, against conditions at the test site (Green 1979; Environment Canada 1998). No two locations are perfectly alike, and there can be large natural differences in biophysical conditions that are confounded with human disturbances and thus make an assessment of condition difficult. The reference condition approach (RCA) is a generalized sampling design (Hughes et al. 1986; Hughes 1994; Reynoldson et al. 1997; Bailey et al. 1998) in which multiple regional reference locations are sampled. Data from the reference sites serve two purposes. First, the data

*Corresponding author: bkilgour@stantec.com

can be used to model an expected biophysical condition. Second, the data can be used to better characterize the background variability in biophysical conditions. When effects are shown to exceed the natural variation observed in regional reference locations, there is greater cause for concern than if effects exceed the variation at a single reference location. Reference-condition-approach models have been developed in Canada to conduct assessments of benthos assemblages in Great Lakes bays (Reynoldson and Day 1998), the Fraser River basin (Reynoldson et al. 1997; Reynoldson et al. 2001), and the Yukon (Bailey et al. 1998). However, the RCA can also be used to model the expected fish assemblage and physico-chemical attributes, so long as the variables used to predict the expected condition do not vary with anthropogenic disturbance (Oberdorff et al. 2002; Pont et al. 2005).

The RCA design relies on a relatively large number (10–100) of relevant reference locations. Reference sites are variously defined but are typically considered least- or minimally disturbed locations. Maude and Di Maio (1996) were able to develop an RCA model for benthos assemblages found in headwater tributaries of the Oak Ridges Moraine, Ontario. However, within the larger ecoregion that the moraine influences, there are 6–7 million people and the area has been developed for more than 200 years. Current land use is dominated by urbanization and agriculture. Reforested lands are limited and mainly restricted to headwater areas and riparian corridors. While many stream segments are in reasonably good condition, there are few streams that could truly be classified as unimpaired or minimally disturbed. There are very few minimally disturbed reference headwater and lower reaches. It is, therefore, difficult to use the conventional RCA design in the Ontario tributaries of Lake Ontario because there are too few sites of any one size or type to construct predictive models.

Variables used to predict expected conditions are termed primary or normative (Imhof et al. 1996) variables, which are not easily altered by humans, measured at the landscape scale, and presumed to relate to fish, benthos, and physical features of streams. For example, underlying surficial geology is a reasonably good predictor of the kinds of fish found in a stream. Brook trout and other coldwater fishes are found in streams where well-drained soils or karst topographies predominate (Ricker 1932; Seelbach et al. 1997). The effects that surficial geology has on the receiving environment quality may, however, be altered by modifying land cover. Other obvious primary variables include stream size (Fausch et al. 1984) and slope (Hughes and Gammon 1987; Kilgour and Barton 1999).

Our objective is to describe an alternative approach to defining reference conditions for areas like the Lake Ontario region. The approach uses relationships between biophysical response variables and landscape features including indicators of human development, such as percent impervious cover (PIC). Relationships with PIC are then used to back-calculate or hindcast the condition that is expected to have occurred prior to development. Residual noise in the response variable, unaccounted for by the gradient in PIC or other natural features is considered to be a measure of the background variability. That residual variation is then used to standardize deviations from expected conditions, such that the degree of impairment is re-expressed in terms of standard deviations. Re-expressing effects as standard deviations puts all variables on a common scale, facilitating comparison among indices of biophysical condition (Kilgour et al. 1998). The general approach provides a means of quantifying the degree of impairment from pre- to postdevelopment for the Lake Ontario region. Others have used environmental gradients to develop tolerance indices for fishes and invertebrates (e.g., Hilsenhoff 1988; Fore et al. 1996; Whittier and Hughes 1998), but hindcasting to reference conditions is a new variation on that theme.

METHODS

Fish, benthos, temperature, and instream habitat conditions were characterized at stream sites located on the north shore of Lake Ontario and draining parts of the Oak Ridges Moraine (Figure 1). Data were collected between 1995 and 2002 using methods described in the Ontario stream assessment protocol (Stanfield et al. 1997; Stanfield and Kilgour 2005, this volume). Randomly selected sites were a minimum of 40 m long, with boundaries at crossovers (i.e., where the thalweg is through the middle of the stream). About half of the total sites sampled during this period were used to develop the hindcasting models, while the remainder were used to validate them in a separate exercise (Stanfield and Kilgour 2005).

Fish assemblage data (721 sites) were collected by single-pass electrofishing and standardized as biomass (g/m^2) and richness (number of species). Benthic macroinvertebrates were sampled from crossovers using a modified version of the U.S. Environmental Protection Agency rapid bioassessment protocol, involving a stationary kick from an area of about 1 m^2, using a screen with 500-μm mesh to collect the animals. Benthic macroinvertebrates were live sorted and identified to major groups (families and orders; Plafkin et al. 1989). Benthos assemblage data (583 sites) were used to estimate the Hilsenhoff biotic index (HBI; Hilsenhoff 1988) with a modification as described in Stanfield and Kilgour (2005) and richness (number of major groups). Additional multivariate metrics of the fish and benthos assemblages were derived using correspondence analysis (CA; Rohlf 1993). Correspondence analysis is an ordination method that simultaneously orders sites and taxa in biplots. Sites close together in the biplots (i.e., similar site

Figure 1. Study area and sampling sites.

scores) have similar assemblages, while taxa close together covary. Sites are assigned scores along each of two axes (in this case), and those scores were used as multivariate metrics of fish and macroinvertebrate assemblages.

Instream habitat data (578 sites) were collected using a point-transect survey design as described in Stanfield et al. (1997). Additional measures of stream width were obtained from sites where only one transect was surveyed as part of discharge measures (622 sites). Water temperatures were recorded between 1600 and 1700 hours, during low-flow conditions (mid-July to mid-September) when the daily air temperature exceeded 24°C for three consecutive days. Observed water temperatures were standardized to an air temperature of 30°C using known relationships between air and water temperatures (Stanfield and Kilgour 2005).

Digital mapping was used to estimate (1) the upstream catchment area (AREA), (2) the slope of a site (100 m upstream to 100 m downstream; SLOPE), and base flow index (BFI, Piggott et al. 2002; Stanfield and Kilgour 2005). A percent impervious cover (PIC) rating of the catchment was estimated based on the percent cover of a catchment as water (0 PIC), natural/forest (1 PIC), pasture (5 PIC), intensive agriculture (10 PIC), and urban (20 PIC). These PIC ratings were based on a review of observed imperviousness values for different land classifications (see Stanfield and Kilgour 2005), and provide an approximation to the percentage of the surface cover that is impermeable to infiltration. The PIC rating of urban areas, for example, varies between 10 (low intensity housing) and 90 (high intensity industrial areas) (Stanfield and Kilgour 2005). Here, we have assumed that all urban areas have a PIC rating of 20, reflecting a predominance of lower intensity development. We acknowledge that a 20 PIC rating does not adequately reflect the true PIC of all urban areas in the study area.

Hindcast Modeling

Backward-stepwise multiple regression was used to construct multiple regression models that related attributes of fish and benthos assemblages and instream physical habitat characteristics to the landscape predictors (Table 1) (see Stanfield and Kilgour 2005; for details). Not all sites had

Table 1. Regression models relating indices of fish and benthos assemblages, and stream temperature and width, to percent impervious cover (PIC rating) and landscape variables. Values provided for predictors are model coefficients. The mean squared error (MSE) and percent of variance explained (R^2) are also provided. Predictors are defined in the text.

	Fish assemblage			Benthos assemblage			Physical		
Predictor	Log_{10} biomass	Rich	CA Axis 1	HBI	Rich	CA Axis 1	CA Axis 2	Temp	Width
Constant	–3.841	–6.582	–1.625	8.68	10.07	16.26	–0.28	43.84	–0.455
Area	1.866			–0.42		–3.40		–7.977	
Area^2	–0.125	0.189				0.19		0.635	0.030
Slope			–0.243	–0.17		–0.20		–1.813	–0.066
Slope^2			0.027					0.181	
BFI			–0.016	–0.017		–0.077	–0.011		–0.028
BFI^2	<–0.001					0.0007		–0.001	<0.001
PIC	–0.066	0.619	0.476	0.092		0.042	0.102	0.885	
PIC^2		–0.035	–0.016		–0.008			–0.003	0.001
MSE	0.215	5.895	0.804	0.653	4.842	0.760	0.827	10.177	0.491
R^2	0.185	0.372	0.394	0.306	0.08	0.255	0.180	0.299	0.607
n	361	361	361	332	332	332	332	385	373

Notes: BFI = base flow index, MSE = mean squared error, *n* = number of sites, Rich = richness or number of taxa, HBI = Hilsenhoff biotic index, Temp = temperature, CA = Correspondence Analysis, and PIC = percent impervious cover rating.

all biophysical data collected, so each model was constructed for a unique set of sites. Only those biological variables for which PIC was a significant predictor were used for hindcasting expected historical conditions (Figure 2). Models were developed for the $\log_{10}$ of fish assemblage biomass, fish assemblage richness, site scores for the first axis from a CA of the fish assemblage data (fish CA Axis 1), Hilsenhoff biotic index, benthos assemblage richness, site scores for the first two axes from CA of the benthos assemblage (benthos CA Axes 1 and 2), water temperature, and average stream width. Predictors included catchment area, slope, base flow index and PIC. The squared term for each predictor was also included in the models in an attempt to explain curvilinear relationships. Predictors were retained in models if they explained a significant amount of variation in the response variables (at $p < 0.05$).

Model relationships (Table 1) were used to estimate the value of the various biophysical indices for expected reference conditions assuming an overall PIC of 1 (i.e., 100% forest cover; Table 2; Figure 2). As per Table 2, model coefficients for significant predictors were multiplied by the individual site conditions. The sum of the products of model coefficients and site conditions provided the expected condition (i.e., 3.10 in Table 2). The difference between present day values of biophysical indices and the hindcast condition (0.76 in Table 2), was used to assess the level of disturbance. Deviations from expected can be expressed in terms of the original units of measurement (e.g., g/100 m^2 or number of taxa), but results between different indices can be difficult to compare. Alternatively, expressing deviations relative to background variability puts indices on a common scale (i.e., standard deviations, Kilgour et al. 1998; Figure 2). With each of the constructed models, unexplained residual variation (i.e., the mean squared error or MSE) included measurement error and unexplained noise or natural variation among sites that is unrelated to landscape features (Figure 2). The square root of the MSE term (0.215 in Table 2) is an estimate of the among-sites standard deviation. That is, after taking into account the primary features (area, slope, BFI) and PIC rating, the MSE provides an estimate of the among-sites variation. The MSE can, therefore, be used to re-express deviations in terms of the background variability (Figure 2) and facilitate comparison of deviations among different

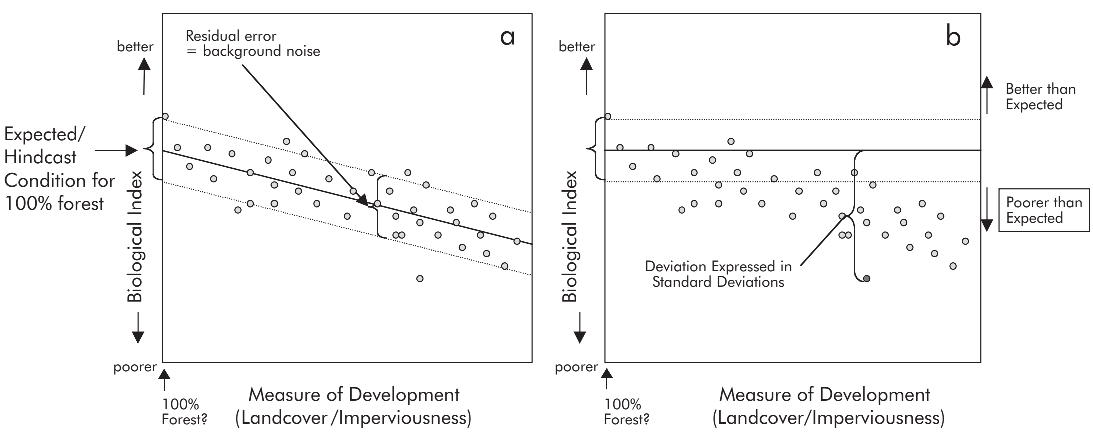

Figure 2. Conceptual illustration of (a) the hindcasting modeling approach, and (b) how the biophysical condition of sites is assessed relative to a hindcast or predicted historical condition. Based on the model in (a), the expected range of values is as shown in (a). Differences between what is predicted and what was observed is expressed relative to the unexplained variability (i.e., as standard deviations in (b)).

Table 2. Example calculations for hindcasting using the model for $\log_{10}$ of fish biomass (g/100 m^2) at a site in Wilmot Creek. BFI = base flow index; PIC = percent impervious cover rating.

Model parameters	Model coefficients	Site conditions	Result
Constant	–3.84	1	–3.84
Log_{10} area	1.87	7.41	13.86
Log_{10} area^2	–0.125	54.8	–6.85
Slope	0	0.49	0
Slope^2	0	0.24	0
BFI	–0.0001	51.8	–0.0052
BFI^2	0	2684	0
PIC	–0.066	1	–0.066
PIC^2	0	0	0
Hindcast (estimated) $\log_{10}$ fish biomass (*a*)		3.10	
Observed (present-day) $\log_{10}$ fish biomass (*b*)	2.34		
Difference (*c*)		(*a*) – (*b*)	0.76
MSE			0.215
Difference re-expressed in units of standardized deviation (SDs)		$\frac{(c)}{\sqrt{MSE}}$	1.64

biological indicators. In the example, the difference (0.76 g/100 m^2) between present day (2.34 g/100 m^2) and hindcast (3.10 g/100 m^2) fish biomass, re-expressed relative to the unexplained variation (MSE = 0.215) was 1.64 standard deviations (Table 2). These reexpressed deviations are termed effect sizes (e.g., Kilgour et al. 1998).

Deviations from the expected hindcast reference condition were expressed relative to the estimated standard deviation for the nine variables for which PIC rating was a significant predictor (Table 1). Fish and benthos assemblages were classified as being in (1) unimpaired when effects were within the expected range of hindcast conditions (i.e., within ± 2 SDs of the predicted mean value), (2) likely impaired when effects were between 2 and 3 SDs from the predicted mean and when the effect was in a "poorer" direction (e.g., reduction in richness or biomass), (3) impaired when effects exceeded 3 SDs and the effect was in a poorer direction, and (4) unimpaired when effects exceeded 2 SDs and were in a "better" direction (e.g., increase in richness or biomass). Sites were classified separately for each fish and benthos assemblage index, and measures of stream temperature and width.

RESULTS

The streams used in this study represented a broad assortment of typical wadeable streams in southern Ontario. Catchment areas varied from 2 to 90,000 ha (average of 4,500 ha), and in slope from 0% to 25% (average of 2%). Catchment land cover varied from 0 to 98% forest, from 0% to 100% urban, and from 0% to 100% agriculture. Fish and benthos assemblages included those representative of both high and low habitat quality.

Fish Assemblages

Present-day biomass of the fish assemblage varied from 1 to 7,000 g/100 m^2 (average of 540 g/100 m^2), while the number of species varied between 1 and 15 (average of 6) per site. Correspondence Analysis (CA) Axis 1 separated coldwater salmonid assemblages (low Axis 1 scores) from warmwater cyprinid and centrarchid assemblages (high Axis 1 scores; Figure 3A). The hindcasting regression models (Table 1) were used to predict Axis 1 scores, biomass and richness of the fish assemblage assuming 100% forest cover. The expected fish

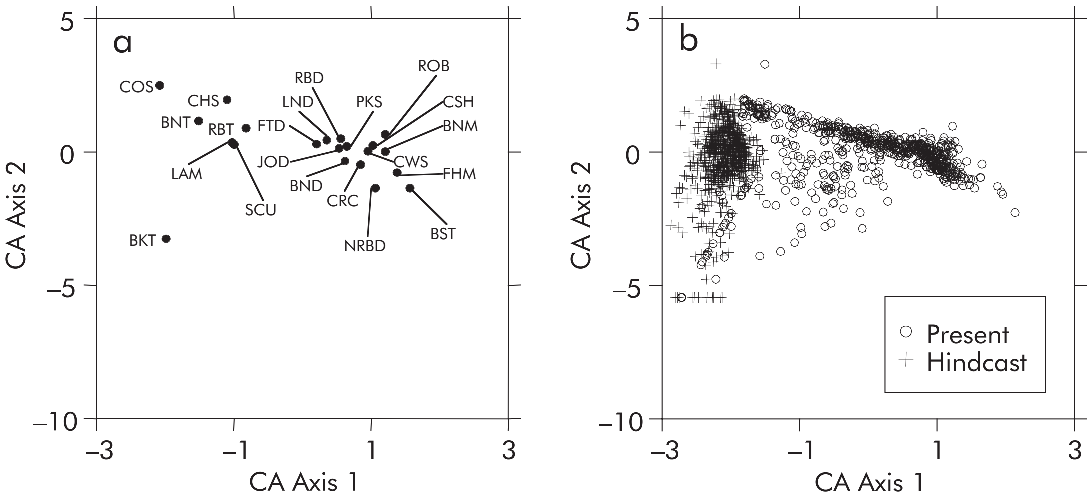

Figure 3. Correspondence analysis (CA) taxa scores (a) for the current fish assemblage, and site scores observed currently and hindcast (b). BST = brook stickleback *Culaea inconstans*; CHS = Chinook salmon *Oncorhynchus tshawytscha*; FTD = fantail darter *Etheostoma flabelare*; COS = coho salmon *O. kisutch*; LAM = lamprey family Petromyzontidae; NRBD = northern redbelly dace *Phoxinus eos*; ROB = rock bass *Ambloplites rupestris*; RBD = rainbow darter *Etheostoma caeruleum*; PKS = pumpkinseed *Lepomis gibbosus*; BNM = bluntnose minnow *Pimephales notatus*; CSH = common shiner *Luxilus cornutus*; FHM = fathead minnow *P. promelas*; BKT = brook trout *Salvelinus fontinalis*; JOD = Johnny darter *Etheostoma nigrum*; BNT = brown trout *Salmo trutta*; LND = longnose dace *Rhinichthys cataractae*; RBT = rainbow trout *O. mykiss*; SCU = sculpin family Cottidae; CRC = creek chub *Semotilus atromaculatus*; WS = white sucker *Catostomus commersonii*; BND = eastern blacknose dace *Rhynichthys atratulus*.

assemblage for most sites in the dataset was cold-/coolwater, consisting principally of salmonids and sculpins (Figure 3B). Hindcast biomass of fish assemblages varied between 6 and 1,000 g/100 m^2 (average of 600) that was generally higher than was observed with the present-day data (Figure 4A). Hindcasting models predicted between 0 and 9 fish species (average of 4) per site, which was lower than the present-day condition of most sites (Figure 4B). Effects on the fish

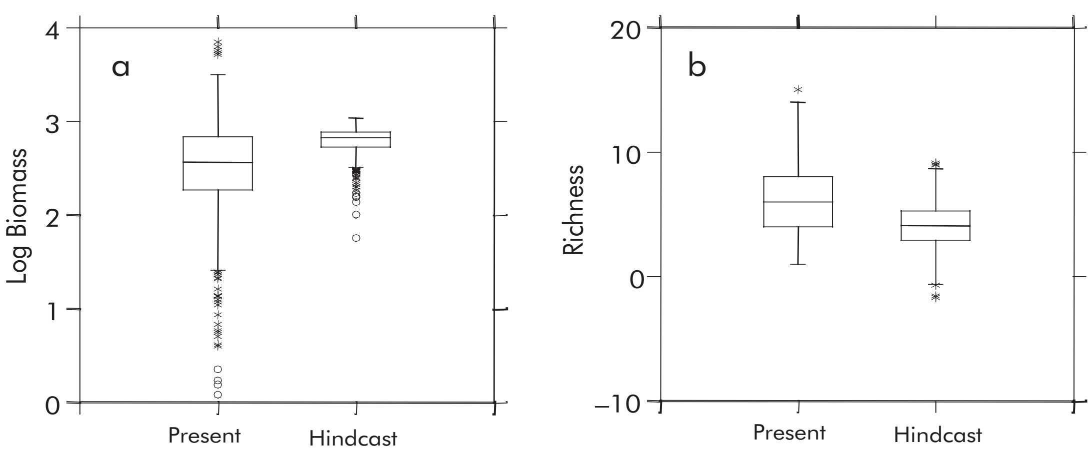

Figure 4. Present and hindcast conditions for the log of fish assemblage biomass (a) and number of taxa (b). In the box plots, the center horizontal line marks the median of observations; the box shows the range within which the central 50% of the values fall, while the whiskers illustrate the data range.

assemblage were most evident with the multivariate descriptor (CA Axis 1), with effect sizes at most sites in the study area more than 2 SDs from the expected hindcast condition (Figure 5). Present day biomass and richness of the fish assemblages were generally within ± 2SDs of the hindcast conditions (Figure 5). Fish assemblages within the immediate vicinity of Toronto (Humber, Don, Credit) and immediately east (Lynde, Oshawa) were mostly impaired, while those further to the east and south in the more agrarian catchments were generally in good condition (Figure 6C).

Benthos Assemblage

Benthos taxa richness varied between 1 and 17 (average of 10) per site, while the HBI varied between 3 and 8 (average 5.4). Low Axis 1 and 2 scores were coincident with benthos assemblages comprised of more sensitive groups such as mayflies (Ephemeroptera), caddisflies (Trichoptera), and stoneflies (Plecoptera), while higher Axis 1 and 2 scores were coincident with benthos assemblages more typically associated with degraded conditions (i.e., Oligochaeta, Chironomidae, Isopoda; Figure 7A). Based on the hindcasting models, benthos assemblages have changed from sensitive to tolerant assemblages (Figure 7B). Those changes were reflected in

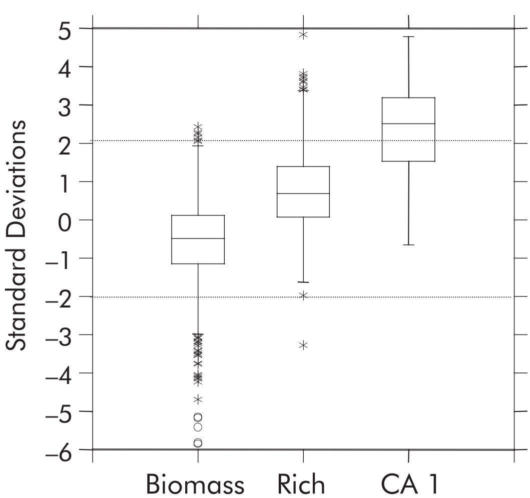

Figure 5. Standardized effect sizes for log of fish assemblage biomass, number of taxa, and CA Axis 1 scores. In the box plots, the center horizontal line marks the median of observations, the box shows the range within which the central 50% of the values fall, while the whiskers illustrate the data range.

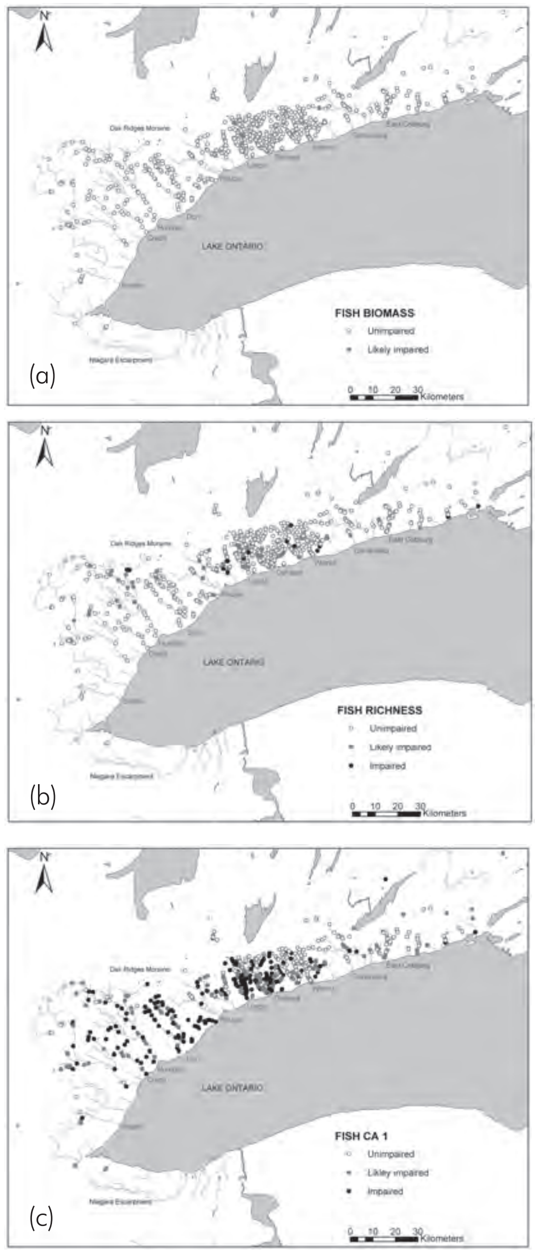

Figure 6. Spatial distribution of southern Ontario sampling sites and the estimated condition for log of fish biomass (a), number of fish species (b), and CA Axis 1 scores.

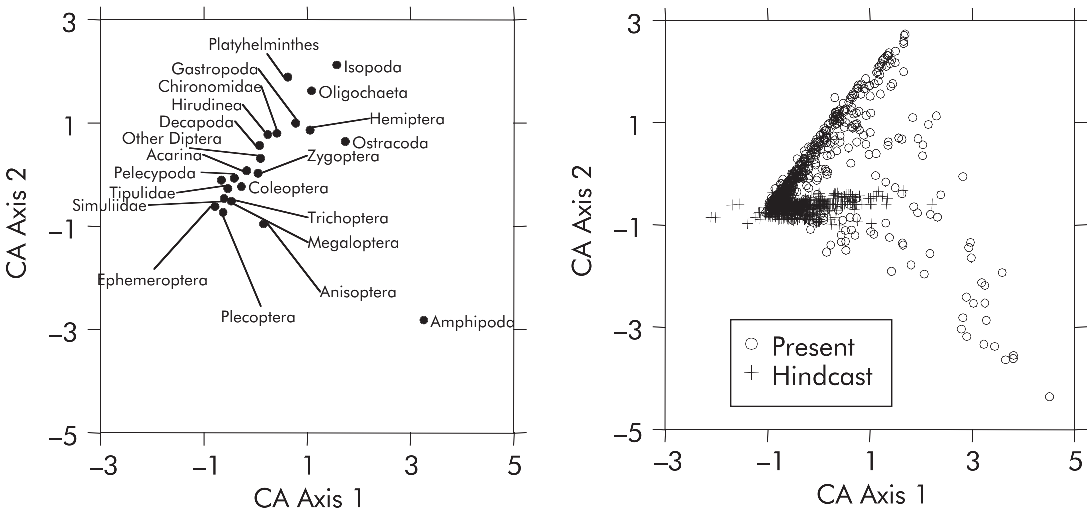

Figure 7. Correspondence analysis (CA) taxa scores (a) for the current benthos assemblage and site scores observed currently and hindcast (b).

significant shifts in the HBI and greater variation in richness of the benthos assemblage (Figure 8). The hindcast condition for the HBI was predicted to range between 3.2 and 5.9 with an average of 4.8, which is about half a unit lower than the present-day condition (Figure 8). Effects were also more evident with the multivariate metrics of the benthos assemblage (i.e., CA Axes 1 and 2) and the HBI than they were with richness (Figure 9). Effects were not as evident with the benthos assemblage as they were with the fish assemblage, and streams in the Toronto area were not shown to be as degraded using benthos as they were with fish (Figure 10).

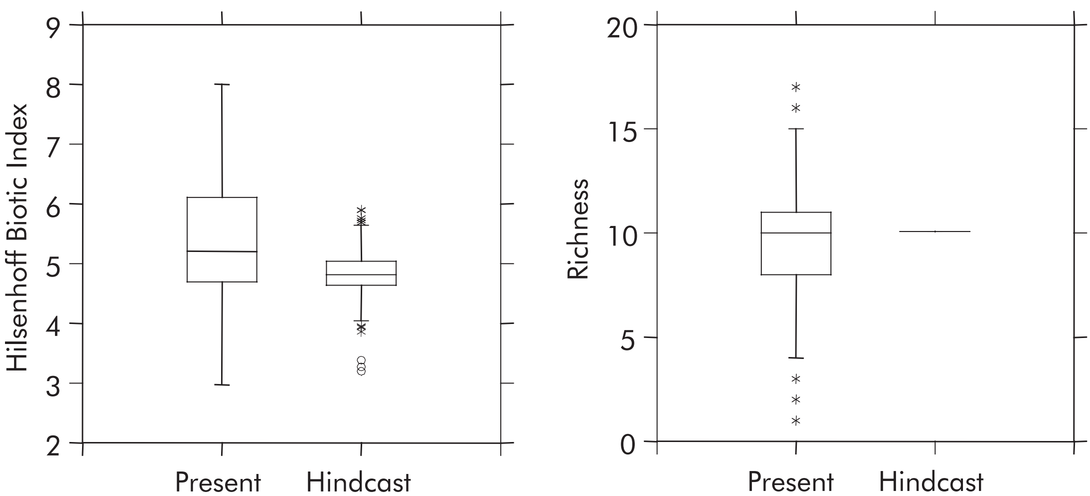

Figure 8. Present and hindcast conditions for the Hilsenhoff biotic index (a) and number of taxa (b). In the box plots, the center horizontal line marks the median of observations; the box shows the range within which the central 50% of the values fall, while the whiskers illustrate the data range.

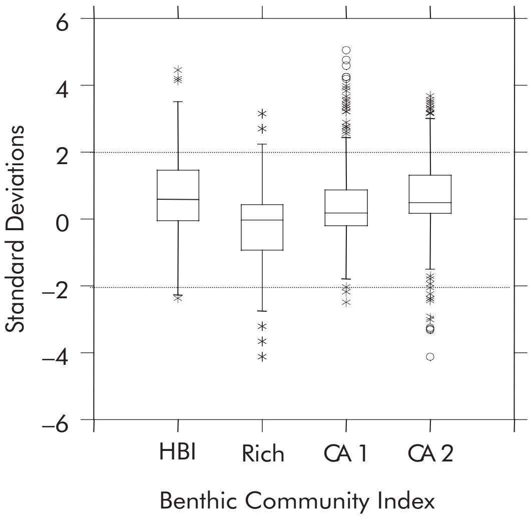

Figure 9. Standardized effect sizes for the Hilsenhoff biotic index, number of taxa, and CA Axis 1 and 2 scores. In the box plots, the center horizontal line marks the median of observations; the box shows the range within which the central 50% of the values fall, while the whiskers illustrate the data range.

Stream Temperature and Width

Present-day standardized stream temperature varied between 9°C and 33°C (average of 21°C), while historical temperatures were predicted to be lower and varied between 12°C and 25°C (average 17°C; Figure 11). Average stream widths varied between 0.4 and 11.5 m (average 4.3 m), while historically, streams were predicted to be narrower with widths varying between 0.1 and 9 m (average 1.1 m; Figure 11). The estimated changes in stream temperatures and widths were not large when expressed relative to the background noise in each of the two variables, with few effects exceeding 2 SDs from the hindcast condition (Figure 12). As with the benthos assemblage data, spatial trends in effects on width and temperature were not evident (Figure 13).

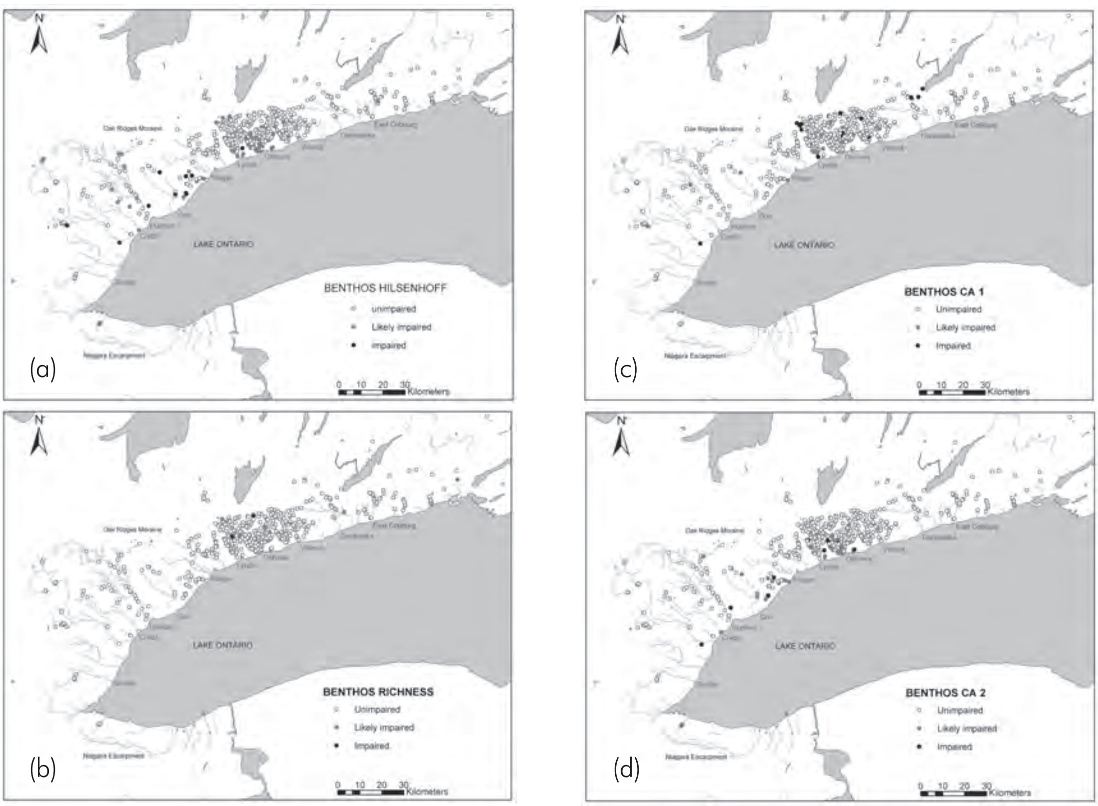

Figure 10. Spatial distribution of southern Ontario sampling sites and the estimated condition for the Hilsenhoff biotic index (a), number of taxa (b), and CA Axis 1 and 2 scores (c, d).

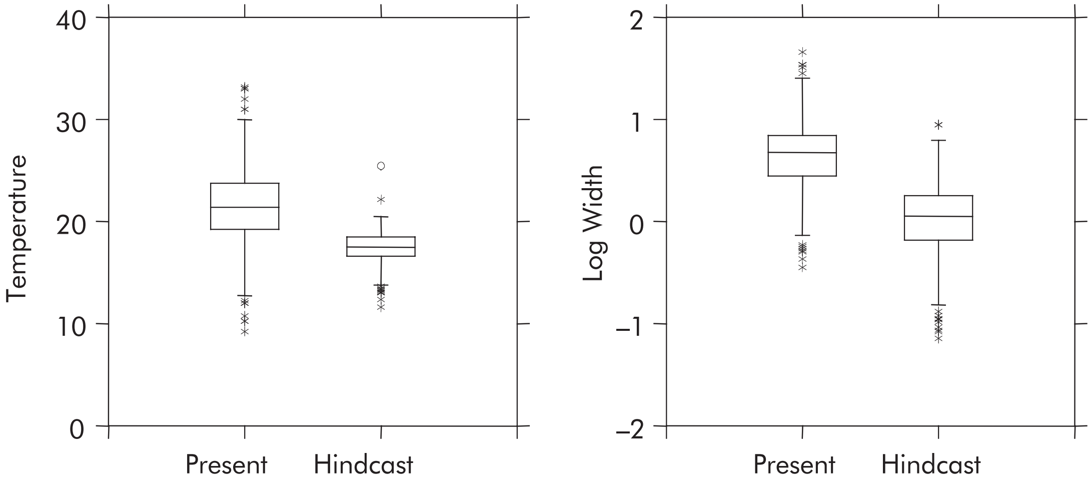

Figure 11. Present and hindcast conditions for maximum stream temperature standardized to an air temperature of 30°C (a) and average stream width (m) (b). In the box plots, the center horizontal line marks the median of observations; the box shows the range within which the central 50% of the values fall, while the whiskers illustrate the data range.

DISCUSSION

We have demonstrated how to hindcast historical expected conditions based on models that relate instream biophysical conditions to land use/land cover with the PIC rating, and then how to assess present-day conditions. We demonstrated significant shifts in biophysical conditions in the study area, with fish and benthos assemblages shifting from sensitive coldwater taxa toward tolerant warmwater taxa. The assessments based on this hindcasting approach are conservative in potentially underestimating the true degree of impairment, and they reflect our understanding of how the streams in the study area have changed over time. Others (e.g., Martin 1984; Steedman 1988; Wichert 1994) have demonstrated impairment in fish assemblages in the greater Toronto area and have related indices of composition to measures of urbanization and agriculture. The hindcasting approach, however, provides an objective measure of the degree or magnitude of changes from presumed historical undeveloped periods. No previous studies have attempted to conduct a hindcast assessment of this study area, although Van Sickle et al. (2004) used hindcasting models to demonstrate fundamental historical changes in Willamette Valley, Oregon streams.

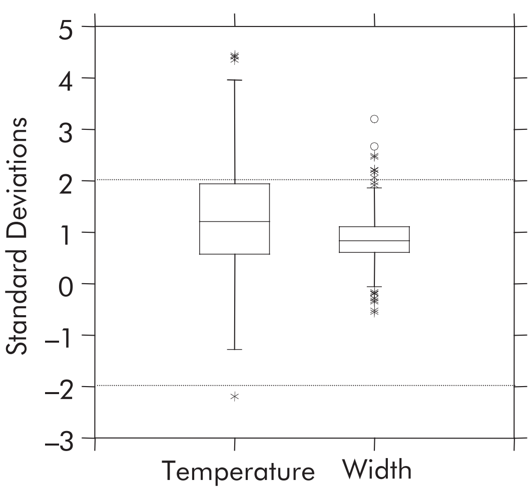

Figure 12. Standardized effect sizes for maximum stream temperature standardized to an air temperature of 30°C and average stream width. In the box plots, the center horizontal line marks the median of observations; the box shows the range within which the central 50% of the values fall, while the whiskers illustrate the data range.

There are several caveats when using hindcast models in assessments. First, we assumed that the reference condition was represented or

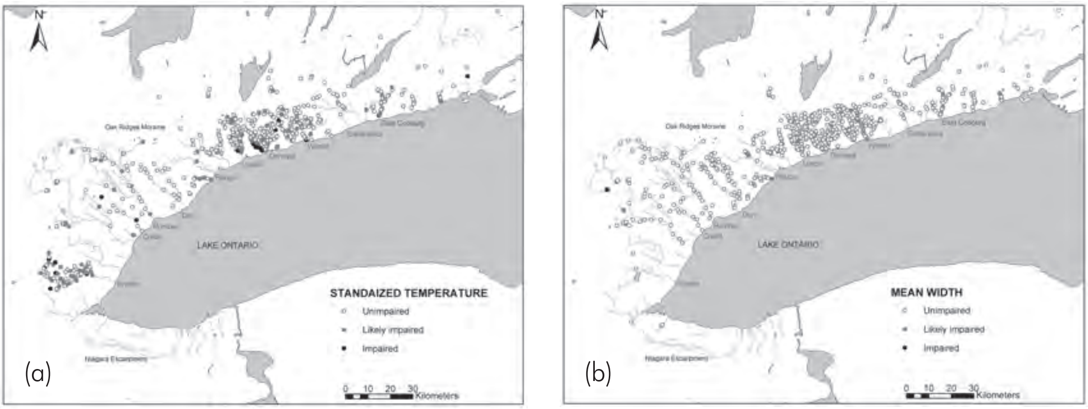

Figure 13. Spatial distribution of southern Ontario sampling sites and the estimated condition for maximum stream temperature standardized to an air temperature of 30°C (a) and average stream width (b).

characterized by 100% forest cover (1 PIC rating). Forest cover was set to 100% purposely to demonstrate the full extent of potential impairment that could have occurred at each of the sites in the dataset. Assessments based on other land cover targets are obviously possible.

Second, caution is necessary in hindcasting outside the calibration range. For example, if there are no minimally disturbed reference sites for larger reaches, then hindcasting to a pristine condition for large reaches is not recommended. In this study, there were at least a few sites in headwater catchments with 100% forest cover, which technically allowed us to hindcast to that level of forest cover. We did not observe 100% forest cover in larger streams near Lake Ontario, so hindcasting to 100% forest cover for larger reaches should only be done with caution.

Third, although we demonstrated that it is possible to hindcast an historical condition, we have not validated the predictions. It would be useful to demonstrate this approach using sites for which there are both present-day and historical reference data. That, however, is not possible for this study area for the following reasons. First there are no quantitative preEuropean settlement data. The entire landscape was clear-cut soon after European settlement in the area, and Atlantic salmon S*almo salar* was locally extirpated partially as a result. Loss of such a keystone species ensures that current fish assemblages are fundamentally different from historic ones (Gresh et al. 2000; Stanfield and Jones 2003).

Fourth, the chronological sequence of land use changes on the landscape was not unidirectional. That is, while much of the study area was clear-cut in the 1800 s and early 1900s, some previously agricultural areas have recovered to a forested condition. Some of the unexplained variation in the models may be due to the differential rates and direction of land use changes, as well as ghost effects from past land uses (Harding et al. 1998). Incorporation of the chronology of events (if it can be determined) could improve the predictive power of relationships.

Fifth, hindcasting is only as good as the variables on which the hindcasting is based. In this demonstration it was assumed that PIC rating was the principle driving measure of effects in the study area. There is good reason to make that assumption considering the number of studies demonstrating effects related to PIC (Stanfield and Kilgour 2005). In the event that other variables override the effects of PIC (e.g., point source discharges, migration barriers), the hindcast models and associated predictions of historical conditions and present-day assessments will not apply. One concern with using

PIC rating in this study was the tendency for PIC to be naturally correlated with underlying surficial geology, and thus the base flow index. In this study area, urban development tended to occur on clay-till plains close to Lake Ontario, while agriculture and forest cover tended to be more prevalent on higher sand/gravel morainal deposits (Figure 1). Stanfield and Kilgour (2005) and Stanfield et al. (2005, this volume) demonstrated that PIC rating was able to account for significant amounts of variability in biophysical responses, through a process of partial regressions. The modeling conducted here has assumed that PIC rating effects overrode effects related to variations in surficial geology.

The strength of the relationship between the hindcasting and response variables is of little concern. Either the predictor variable explains significant amounts of variation in the response variable or not. When the predictor (PIC rating) explains a large amount of variation in the predictor, then the background range of natural variability will be a smaller range than when the predictors explain less variability in the response variable. Comparison of test sites to the normal range of variability follows the same process, whether the range is considered small or large. Through exploratory analyses, we found that the percent of variance explained in response variables was increased if models were developed for smaller portions of the overall study area, and therefore recommend the construction of models for smaller study areas when the data are available. The models constructed for the larger study area, however, provide information useful for screening assessments. That is, where indices of composition or habitat features exceed the normal range of hindcast conditions based on large-scale models, there is good certainty that effects are significant and deserve consideration. In contrast, not exceeding normal ranges based on large-scale models does not imply effects are not large or important. Also Riseng et al. (2005, this volume) found that large-scale data sets offered more sensitive models than smaller scale data sets because the former included more minimally disturbed sites.

Finally, despite concerns over confounding, the observed effects developed by the hindcasting models were modest compared to what was anticipated. Streams in the Toronto area are considered highly degraded (Wichert 1994), and larger effects were anticipated. The models used here were derived from a large area and thus represented regional relationships. Subsequent modeling and analyses may demonstrate that subsets of models for smaller study areas further reduces unexplained variability in biophysical conditions and thus increases our ability to detect effects relative to unexplained variability. Riseng et al. (2005) found that smaller scale models were better able to incorporate local effects from dams and point source discharges. Our regional models provide a set of numeric biophysical criteria that can be used to assess the level of degradation. More subtle effects might be discernible with more local models, but effects documented using our regional models deserve management consideration. The noise associated with these regional models may partially explain why standardized effects on stream temperatures and widths were small, even though effects in terms of the original units of measurement were quite significant. Stream temperatures, for example, increased from an expected average of 17°C to a present-day average of 21°C. Differences of 4°C can substantially alter production of salmonids and invertebrate assemblages (Bisson and Davis 1976; Hughes and Davis 1986). When re-expressed in units of standard deviations, that difference (4°C) was just greater than 1, which is not a large statistical difference. Further stratification of the database to account for regional differences might make the assessment of effects on temperature, width, and other response variables more sensitive.

Benthos assemblage data and measures of stream temperature and width were relatively insensitive. Our method of sampling the benthos assemblage, however, was a fairly coarse tool, principally recommended for use in rapid screenings of problem areas (Plafkin et al. 1989). Had the benthos assemblage been sampled more thoroughly and had more individuals been counted to lower taxonomic levels, greater effects would likely have been observed (Furse et al. 1984; Wright et al. 1995; Cao et al. 2002). The protocols for measuring stream width and temperature were fairly standard (and rigorous), so the lack of effects was probably a function of local factors such as groundwater inputs, shading, and adjacent land use being important. Gregory et al. (1991) emphasized that riparian trees are critical for shading small streams and for providing habitat structure.

In this analysis, we did not evaluate the statistical significance of differences from hindcast to present-day conditions. Rather, we visually examined the differences between expected hindcast and present-day conditions, expressed relative to the unexplained noise in the response variables. Effects expressed in terms of standard deviations are becoming a popular means of articulating the potential ecological significance of an effect (Lowell 1997; Kilgour et al. 1998). In the reference-condition approach, the comparison of one sample from one site to a set of reference samples from a number of sites can be analyzed as a typical two-sample contrast or *t*-test, though some (e.g., Kilgour et al. 1998) have argued it should be a one-sample contrast based on comparison to noncentral *t* or *F*-distributions. Sample sizes were very high in this study (500+) resulting in very high statistical power. Any site with an effect exceeding 2 SDs would very likely differ significantly from the hindcast reference condition. For the purpose of this manuscript, therefore, specific testing of statistical significance was not considered critical. Quantifying the magnitude of effects was, however, considered more informative.

Our analysis confirms that the biophysical condition of tributaries on the north shore of Lake Ontario varies with the amount of development in the region. Here, a measure of human development (percent impervious cover, PIC rating) was used to hindcast to expected reference conditions. By expressing deviations from the expected hindcast condition relative to the unexplained residual variability (as in a standard deviation), deviations in all biophysical variables are expressed on a common scale and can thus be easily compared.

ACKNOWLEDGMENTS

This study required the contributions of several people, many of whom contributed to the primary paper by Stanfield and Kilgour (2005). For this paper, we thank Robert McGlaughlin for the concept of hindcasting, Keith Somers for reviewing the finalized approach, Gen Carr for helping with statistical and editorial issues, two anonymous reviewers for providing comments that improved the paper, and Robert Hughes for an excellent editorial review. Brent Harlow and Jason Borwick generated the maps used in this paper. Funding from the City of Toronto and in-kind contributions from the Ontario Ministry of Natural Resources, Jacques Whitford Limited, and Stantec Consulting Limited are appreciated.

REFERENCES

Bailey, R. C., M. G. Kennedy, M. Z. Dervish, and R. M. Taylor. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. Freshwater Biology 39:765–774.

Bailey, R. C., R. H. Norris, and T. B. Reynoldson. 2003. Bioassessment of freshwater ecosystems: the reference condition approach. Kluwer, Boston.

Bisson, P. A., and G. E. Davis. 1976. Production of juvenile chinook salmon, *Onchorhynchus tshawytscha*, in a heated model stream. Fishery Bulletin 74:763–774.

Cao, Y., D. P. Larsen, R. M. Hughes, P. L. Angermeier, and T. M. Patton. 2002. Sampling effort affects multivariate comparisons of stream assemblages. Journal of the North American Benthological Society 21:701–714.

Environment Canada. 1998. Pulp and paper technical guidance for aquatic environmental effects monitoring. Environment Canada, EEM/1998/1, Hull, Quebec.

Fausch, K. D., J. R. Karr, and P. R. Yant. 1984. Regional application of an index of biotic integrity based on stream-fish communities. Transactions of the American Fisheries Society 113:39–55.

Fore, L. S., J. R. Karr, and R. W. Wisseman. 1996. Assessing invertebrate responses to human activities: evaluating alternative approaches. Journal of the North American Benthological Society 15:212–231.

Furse, M. T., D. Moss, J. F. Wright, and P. D. Armitage. 1984. The influence of seasonal and taxonomic factors on the ordination and classification of running-water sites in Great Britain and on the prediction of their macro-invertebrate communities. Freshwater Biology 14:257–280.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley, New York.

Gregory, S. V., F. J. Swanson, W. A. McKee, and K. W. Cummins. 1991. An ecosystem perspective of riparian zones: focus on links between land and water. BioScience 41:540–551.

Gresh, T., J. Lichatowich, and P. Schoonmaker. 2000. An estimation of historic and current levels of salmon production in the Northeast Pacific ecosystem: evidence of a nutrient deficit in the freshwater systems of the Pacific Northwest. Fisheries 25(1):15–21.

Harding, J. S., E. F. Benfield, P. V. Bolstad, G. S. Helfman, and E. B. D. Jones, III. 1998. Stream biodiversity: the ghost of land use past. Proceedings of the National Academy of Sciences 95:14843–14847.

Hilsenhoff, W. L. 1988. Rapid field assessment of organic pollution with a family-level biotic index. Journal of the North American Benthological Society 7:65–68.

Hughes, R. M. 1994. Defining acceptable biological status by comparing with reference conditions. Pages 31–47 *in* W. S. Davis and T. P. Simon, editors. Biological assessment and criteria: tools for water resource planning and decision making. Lewis, Boca Raton, Florida.

Hughes, R. M., and G. E. Davis. 1986. Production of coexisting juvenile coho salmon and steelhead trout in heated model stream communities. Pages 322–337 *in* J. Cairns, Jr., editor. Community toxicity testing. American Society for Testing and Materials, Special Technical Publication 920, Philadelphia.

Hughes, R. M., and J. R. Gammon. 1987. Longitudinal changes in fish assemblages and water quality in the Willamette River, Oregon. Transactions of the American Fisheries Society 116:196–209.

Hughes, R. M., D. P. Larsen, and J. M. Omernik. 1986. Regional reference sites: a method for assessing stream potential. Environmental Management 10:629–635.

Imhof, J. G., J. Fitzgibbon, and W. K. Annable. 1996. A hierarchical evaluation system for characterizing watershed ecosystems for fish habitat. Canadian Journal of Fisheries and Aquatic Sciences 53(Supplement 1):312–326.

Karr, J. R., and E. W. Chu. 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington, D.C.

Kilgour, B. W., and D. R. Barton. 1999. Associations between stream fish and benthos across environmental gradients in southern Ontario, Canada. Freshwater Biology 41:553–566.

Kilgour, B. W., K. M. Somers, and D. E. Matthews. 1998. Using the normal range as a criterion for ecological significance in environmental monitoring and assessment. Écoscience 5:542–550.

Lowell, R. B. 1997. Discussion paper on critical effect size guidelines for EEM using benthic invertebrate studies. Report to the Environmental Effects Monitoring Program. Environment Canada, Hull, Quebec.

Martin, D. K. 1984. The fishes of the Credit River: cultural effects in recent decades. M.Sc. thesis. University of Toronto, Toronto.

Maude, S. H., and J. Di Maio. 1996. Benthic macroinvertebrate communities and water quality of headwater streams of the Oak Ridges Moraine: reference conditions. Ontario Ministry of the Environment and Energy, Etobicoke, Ontario.

Oberdorff, T., D. Pont, B. Hugueny, and J-P. Porchers. 2002. Development and validation of a fish-based index for the assessment of 'river health' in France. Freshwater Biology 47:1720–1734.

Piggott, A., D. Brown, and S. Moin. 2002. Calculating a groundwater legend for existing geological mapping data. Environment Canada, Burlington, Ontario.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross, and R. M. Hughes. 1989. Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish. U.S. Environmental Protection Agency, EPA/440/4–89/001, Washington, D.C.

Pont, D., B. Hugueny, and T. Oberdorff. 2005. Modelling habitat requirement of European fishes: do species have similar responses to local and regional environmental constraints? Canadian Journal of Fisheries and Aquatic Sciences 62:163–173.

Reynoldson, T. B., and K. E. Day. 1998. Biological sediment guidelines for the Great Lakes. Environment Canada, Burlington, Ontario.

Reynoldson, T. B., D. M. Rosenberg, and V. H. Resh. 2001. Comparison of models predicting invertebrate assemblages for biomonitoring in the Fraser River catchment, British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 58: 1395–1410.

Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of the North American Benthological Society 16: 833–852.

Ricker, W. E. 1932. Studies of the speckled trout (*Salvelinus fontinalis*) in Ontario. Publications of the Ontario Fisheries Research Laboratory 44:69–110.

Riseng, C. M., M. J. Wiley, R. J. Stevenson, T. Zorn, and P. W. Seelbach. 2005. Comparison of coarse versus fine scale sampling on statistical modeling of landscape effects and assessment of fish assemblages of the Muskegon River, Michigan. Pages xx–xx *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influences of landscape on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Rohlf, F. J. 1993. NTSYS-pc, numerical taxonomy and multivariate analysis system, Version 1.80. Exeter Software, Setauket, New York.

Seelbach, P. W., M. J. Wiley, J. C. Kotanchik, and M. E. Baker. 1997. A landscaped-based ecological classification system for river valley segments in lower Michigan (MI-VSEC Version 1.0). Michigan Department of Natural Resources, Fisheries Research Report Number 2036, Ann Arbor.

Stanfield, L. W., and M. L. Jones. 2003. Factors influencing rearing success of Atlantic salmon stocked as fry and parr in Lake Ontario tributaries. North American Journal of Fisheries Management 23:1175–1183.

Stanfield, L., M. Jones, M. Stoneman, B. Kilgour, J. Parish, and G. Wichert. 1997 Stream assessment protocol for southern Ontario. V 3.1 training manual. Ontario Ministry of Natural Resources, Peterborough.

Stanfield, L. W., and B. W. Kilgour. 2005. Effects of percent impervious cover on fish and benthos assemblages and instream habitats in Lake Ontario tributaries. Pages xxx–xxx *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influences of landscape on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Stanfield, L. W., S. F. Gibson, and J. A. Borwick. 2005. Using a landscape approach to identify the distribution and density patterns of salmonids in Lake Ontario tributaries. Pages xxx–xxx *in* R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Influences of landscape on stream habitats and biological assemblages. American Fisheries Society, Symposium 48, Bethesda, Maryland.

Steedman, R. J. 1988. Modification and assessment of an index of biotic integrity to quantify stream quality in southern Ontario. Canadian Journal of Fisheries and Aquatic Sciences 45:492–501.

Van Sickle, J., J. Baker, A. Herlihy, P. Bayley, S. Gregory, P. Haggerty, L. Ashkenas, and J. Li. 2004. Projecting the biological condition of streams under alternative scenarios of human land use. Ecological Applications 14:368–380.

Whittier, T. R., and R. M. Hughes. 1998. Evaluations of fish species tolerances to environmental stressors in lakes in the northeastern United States. North American Journal of Fisheries Management 18:236–252.

Wichert, G. 1994. Fish as indicators of ecological sustainability: historical sequences in Toronto area streams. Water Pollution Research Journal of Canada 29:599–617.

Wright, I. A., B. C. Chessman, P. G. Fairweather, and L. J. Benson. 1995. Measuring the impact of sewage effluent on the macroinvertebrate community of an upland stream: the effects of different levels of taxonomic resolution and quantification. Australian Journal of Ecology 20:142–149.