Reporting on the condition of stream fish communities in the Canadian tributaries of Lake Ontario, at various spatial scales

L. W. Stanfield
Ontario Ministry of Natural Resources,
Southern Science and Information Section,
Glenora Fisheries Station.

A report to the Southern Ontario Modeling Team,

December 21, 2010

Citation: Stanfield, L.W. 2010. Reporting on the condition of stream fish communities in the Canadian tributaries of Lake Ontario, at various spatial scales. Technical report of the Science and Information Section, Ministry of Natural Resources, Peterborough, On.
Abstract

“Scaling up” field data to provide meaningful summaries for managers that incorporate both the disparate distribution of sampling sites and the confidence in a classification continues to challenge ecologists. In this study, spatial distributions of unimpaired; likely impaired; or impaired conditions of 721 sites were evaluated using different spatial reporting units as a means of illustrating how scaling effects influence reporting perspectives. We test an approach for communicating field results that incorporates both scale and confidence ratings on management decisions. Segment condition, based on the median value of site observations, and a confidence ranking for the classification is determined for each segment. A rank sum approach based on the length of stream sampled and simple scoring criteria (e.g., scores < 50 equal a fail) combine to demonstrate the effect of mapping criteria on the interpretation of results. We demonstrate how patterns in the classifications at each scale can identify priority areas for restoration and how scaling effects influence overall ranking of condition. Heterogeneity in stream condition results in a pattern of larger catchments tending to be in more degraded conditions. As a result, reporting area scores tended to decrease as reporting area size increased. Providing the base data and confidence rankings on maps ensures that decision makers and the public can consider these scaling effects. We suggest that the approach provides a clear and objective means of summarizing condition across multiple scales and believe that its application on a broader basis is feasible.

Keywords: environmental reporting, streams, scaling effects, report card, restoration prioritization, biodiversity conservation, developed landscapes

1 Introduction

To be effective, restoration ecology should; target priority areas; address processes; include a multidisciplinary and hierarchical approach; and ensure that questions are addressed at the appropriate scale [33]. Policies that provide the context and foundation for restoration cover a variety of spatial scales, depending on the target ecosystem such as watersheds; [40, 63] ecoregions [15, 8]; or political jurisdictions [34]. While guidelines for construction and mitigation typically provide direction at the local scale, it is widely recognized that for consistent success, practitioners must consider the broader context of effects to the site (e.g., [33]). Understanding and communicating the interrelationships of factors that influence the hierarchical nature of ecosystems and how reporting area boundaries influences ecosystem indicators has been a major focus of ecology since its origins as a science [28, 31, 26, 70, 67, 62, 69, 1] and is key to ecosystem management and restoration. Effective linkages between ecosystem restoration strategies and ecosystem understanding require studies that use response metrics that are important to the decision makers; are sensitive to the predictor variables and ecological processes under examination and that capture scaling effects [36, 68, 1]. Providing this context to decision makers has been
hampered by the complexity of issues relating to the statistical properties of metrics and the implications from the first law of geography, “Everything is related to everything else, but near things are more related than distant things” sensu Tobler [56], that is the foundation of the laws of spatial autocorrelation [67].

Ecologists are beginning to understand how patterns in ecosystem response to human induced disturbances break down as the scale of measurement changes. Reasons for the break down include, but are not limited to: statistical properties associated with spatial patterns [11]; smoothing effects in attribute properties that occur as the grain size changes [46, 19, 69]; and differences in the variance properties of an attribute that occur over time and as grain size changes [6, 11]. Wu [66] suggests that for landscape attributes that possess hierarchical structure, be it structural or functional, they may form multiple hierarchies in the same landscape that may not correspond to each other precisely in terms of spatial and temporal scales [67]. Terrestrial ecologists are making tremendous progress in quantifying and thereby understanding how scaling effects (proximity - grain - study extent etc.) and hierarchical structure influence the processes that affect terrestrial habitats, [24, 11, 69]. The result is that studies now routinely identify the appropriate scale at which to report terrestrial processes (e.g., 19, 10, 66, 35, 67). Terrestrial ecologists are embracing a new paradigm, “hierarchical patch dynamics” that offers a framework for explicitly incorporating heterogeneity and scale, and for integrating equilibrium, multiple equilibrium and nonequilibrium perspectives [70]. Since rivers are a mirror of their watersheds, [45] the challenge to riverine ecologists is to sort out the dual effects of terrestrial spatial and scaling influences and scaling/autorcorrelation influences that result from the hierarchical influences from the river network (longitudinal) and the residual horizontal influences of the remaining terrestrial habitats.

Similar to terrestrial habitats a variety of primary variables (climate - geology, glacial history, etc.) operate in interconnected ways to influence stream habitats in a hierarchical nature [17, 7, 60, 13, 45, 18, 64, 48, 9]. These factors operate in similar ways for both terrestrial and aquatic habitats and as such, classification schemes such as those produced by Maxwell et al [30] have been widely applied by both terrestrial and aquatic ecologists (see [21]). Unfortunately, because of their nested hierarchy and inherent variability, classification of riverine systems has required another level of complexity to help understand how scaling effects influence the processes and communities of these ecosystems [60, 13, 48]. In summary, geomorphic and climatic processes influence stream habitat units (grain sizes) differentially in a hierarchy from ecoregion to watershed, to reach, to site, to a microhabitat unit, such as an individual pool that create a flowing dynamic equilibrium that is intimately connected to the landscape and the river network. Natural (weather and geomorphology) and human disturbances also interact with these grain sizes differentially, depending on both the nature of the disturbance to habitats; the sensitivity of the grain size to each disturbance (see [17]) for point source impacts, and Dunne and Leopold [7] for landscape
impacts); and the potential residual hierarchical influences of upland habitats. Given this complexity, it is not surprising that at present, there is a poor understanding of how both the spatial and hierarchical scaling factors combine to influence the expression of stream metrics, yet this is critical to implementation of stream restorations. Wu [67] refers to this problem as spatial transmutation or, the tendency for the inappropriate extrapolation of statistical relationships from one scale to another. This is of concern to restoration ecologists because, if metrics used to study patterns respond differently depending on underlying scaling factors then efforts to direct restoration to specific habitats based on this analysis could lead to project failures.

Detecting patterns in ecosystems is affected by observed variance in measured attributes at the level in the hierarchy of interest and unfortunately, these values are known to change properties as grain size increases [67, 11, but the causes of these changes are the result of a variety of factors. While ecologists have long known of the importance of standardization of sampling methods (see [14]), a major impediment to developing the understanding of ecological responses to altered landscapes is that as grain size increases the availability of datasets collected using comparable methods decreases, and therefore variance resulting from sampling methods increases. As such, analysis of ecological condition at larger grain sizes is often restricted to coarser metrics that are less sensitive to changes in ecological condition (e.g., species presence verses abundance of all species) [1]. This can generate a mismatch between managerial needs and scientific understanding. Fortunately, patterns in data tend to manifest themselves at distinctive operational scales despite the heterogenous nature of the landscape [67] and the hierarchical structure of river networks [13].

For researchers working in flowing waters there has been increasing interest and availability of datasets (see [47]) collected using standard methods [49, 2] that cover large geographic areas that are useful for addressing a number of landscape and scaling questions (see papers in [16]). These large datasets are in general, not evenly distributed on the landscape and tend to be spatially autocorrelated [62]. How grain size and the distribution of data affect patterns in stream metrics on the landscape and how reporting data at inappropriate scales might influence policy/restoration decisions has yet to be demonstrated.

The objective of this paper is to illustrate how reporting data at varying reporting scales can influence perspective on the overall state of the rivers and to illustrate an approach that can help address these effects and contribute to more informed decisions about management priorities.

The intent is to demonstrate how perspective changes when spatially extensive datasets are summarized at various reporting scales and to demonstrate how GIS mapping approaches might facilitate a consistent means of interpreting/reporting data at multiple scales and that demonstrates the inherent limitations in the data. A second objective of this paper is to evaluate the accuracy of stream segment
classification based on models, as this is integral to understand and describe areas where field data are not available. Wang et al [62] identified this need as an important step to validating the segment as the primary reporting unit for streams.

2 Methods

The study covers streams located on the north shore of Lake Ontario (Figure 1). Two physiographic features, the Oak Ridges Moraine and the Niagara Escarpment, dominate this area and are influential in generating both natural and development spatial patterns in the area. Most forested areas are located on top of the moraine and escarpment, along the river valleys and remnant wetlands generating a strong north-south gradient [50]. Low intensity agriculture dominates on the tablelands and urban areas are predominant in proximity to Lake Ontario and in the west-central part of the study area, generating both a south to north and a west to east gradient of development [50]. These two features provide an abundance of groundwater discharge ensuring that northern headwater segments are cold and that water temperatures gradually increase towards the river mouths.

Streams in this region now have a relatively diverse fish assemblage that can range from cold-water intolerant communities of brook trout (*Salvelinus fontinalis*) and slimy sculpin (*Cottus cognatus*) to tolerant and warm water assemblages that contain species such as common shiner (*Luxinus cornutus*) and fathead minnows (*Pimephales promelas*). Additionally, these streams now contain a variety of naturalized non-native salmonids: rainbow trout (*Onchorhynchus mykiss*), chinook salmon (*Oncorhynchus tshawytscha*), coho salmon (*Oncorhynchus kisutch*) and brown trout (*Salmo trutta*).

2.1 Developing a Fish Assemblage Metric:

This study builds on the results and utilizes the same base field data, of an earlier study by Stanfield and Kilgour [52] and Kilgour and Stanfield [23]. Because the modeling is the foundation of the reporting process that follows, we provide an extensive description of how the models were developed. For more details and the background tables, refer to the earlier publications [52, 23].
The analysis utilizes fish assemblage data from 1995 and 2002 collected using single pass electrofishing (standardized to g/100m2). Several species were inconsistently identified; mottled sculpin (Cottus bairdi) and slimy sculpin (C. cognatus); American brook lamprey (Lampetra appendix) and sea lamprey (Petromyzon marinus) and are therefore combined for this analysis. Species found at < 5% of sites are not included in the analysis. Log densities of each species are summarized to provide an objective measure of the contrast in fish assemblages between sites.

Correspondence analysis (CA) is used to generate a synthetic score for each site that reflected the contrast in fish assemblage composition for the study area. The CA ordination calculates a set of “synthetic” variables (axes) that best explain variations in taxa abundances across sites. Calculation of site and taxa scores on the first ordination axis is done by iteratively estimating the weighted-average sample scores and the weighted average taxa scores. Sample scores in CA are usually scaled to a mean of zero and standard deviation of 1 [55]. There was a gradient in the CA axis scores for fish. Cold-water sensitive taxa were located on the extreme left side of the correlation (high negative scores) and sites with warm water tolerant species were found on the extreme right side (high positive numbers) (Figure 2). For the remainder of the paper we refer to the CA axis 1 fish assemblage values as the shortened “fish assemblage metric”.

![Figure 1. Sites used in this study and major geographic features.](image-url)
Catchment boundaries and areas for each site are determined using Arc-hydro (v1.1 for ArcGIS 9.0), [29]. Surficial geology is attributed for each site and converted to a baseflow index (BFI) using a summed rank percentile approach [43]. Slope for each site is determined over a 200 m length of stream, centered at each site. Kilgour and Stanfield [23] created a landscape disturbance index, (LDI) to provide a single metric of the overall land cover in a catchment that reflected forested, agriculture and urban lands. The LDI is calculated by using a summed ranking of the land cover data (land cover 28, [38]), with the following rankings: urban (0.2), intensive agriculture (0.1), rangeland (0.05), forest (0.01) and water (0). These ratings loosely reflect ratings used to measure percent impervious cover (PIC), the term that Stanfield and Kilgour [23] used to describe this index. Each of the landscape variables represents predictor variables for subsequent model development designed to evaluate the deviation from a “reference” condition.

The input data to a general linear model were drainage area, slope, BFI, LDI, and their squared terms (to account for potential curvilinear relationships), along with the fish assemblage metric values for each site. The model predicts a threshold response at LDI ratings of 8-10 and a linear response below this threshold (Table 1) (Figure 3). Because issues of autocorrelation can confound these types of models, Stanfield and Kilgour [52] confirmed the significance of the models by evaluating the LDI effect on the residuals of the model after accounting for the primary landscape variables (area, BFI and slope). The absence of catchment
area in the fish assemblage model was likely due to the strong correlation of slope and catchment area in this dataset. Finally, the pattern of the relationship between the observed and predicted fish assemblage metric was similar for both

![Graphs showing relationships between landscape and fish assemblage composition as determined through Canonical Correspondence Analysis (CCA).](image)

Figure 2. Relationship between landscape and fish assemblage composition as determined through Canonical Correspondence Analysis (CCA). Reprinted with permission and species acronyms are defined in Table 6 [52]. Por=porosity of soils; str=stream; PIC is analogous to LDI; BFI=baseflow index.

the calibration and validation data sets, thereby providing further support that the model was robust.

2.2 Evaluating the state of the fish assemblage metric:

The power of the fish assemblage metric relationship with landscape variables is relatively strong for this type of model (adjusted $r^2=0.39$) and the mean sum of the error term is low, thereby enabling the possibility of quantifying sites that significantly differ from the predicted condition, or the degree of impairment at each site. Kilgour and Stanfield [23] proposed the use of hindcasting to provide a benchmark reference state for landscapes where truly reference conditions did not exist and applied this approach to this study area and dataset. This approach hindcasts the fish assemblage metric values for each site, by applying the sum of the products of model coefficients and site landscape conditions, after setting forest cover to 100% (LDI = 1).

For each site, the difference from hindcasted condition was calculated and compared to the standard deviations (the square root of the MSE, 0.8 in Table 1) observed from the model. Sites within 2 SD in either direction of the mean (i.e., <95%) are considered unimpaired; sites within 2-3 SD (> 95% but <99%) are likely impaired and sites >3 SD are considered impaired. Sites with current
conditions that are better than predicted in the reference state by >2 SD are classified as “better than expected” [22]. In this manner, regardless of catchment primary conditions, a standardized measure of site condition is calculated. For example, a site with a fish assemblage value of 4.7 and hindcasted community of 1.9 deviates by a score of 2.8. This site would be classified as impaired because the difference is greater than 3 SD (2.7) from the mean relationship, for sites with similar BFI and slope.

2.3 Translating site condition to a landscape perspective:

GIS is used to define and attribute a variety of spatial reporting units that are typically used by managers to report on the state of land/riverscapes (see Figure 4).

An automated GIS demarcated and attributed valley segments, the first reporting level in the hierarchy. Segment boundaries occurred at changes in Strahler order, hydraulic conductivity (as determined from surficial geology) and intersections.

Table 1: Mean sum of the error (MSE) terms and model coefficients for the fish assemblage metric (CA axis 1), (from [52]). Numbers in brackets indicate the thresholds for 2 and 3 standard deviations used to classify sites.

<table>
<thead>
<tr>
<th>MSE (2SD/3SD)</th>
<th>Constant</th>
<th>slope</th>
<th>slope²</th>
<th>BFI</th>
<th>LDI</th>
<th>LDI²</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8 (1.8/2.7)</td>
<td>-1.63</td>
<td>-0.24</td>
<td>0.03</td>
<td>-0.02</td>
<td>0.48</td>
<td>-0.02</td>
</tr>
</tbody>
</table>
with water bodies greater than 10 ha in size. Hydraulic conductivity boundaries reflected changes from poorly (1x10-5 to 1x10-2 m/yr), moderately (1x10-2 to 1x101 m/yr) or well drained (1x101 to 1x107 m/yr) soil porosities, as described by the Ontario geological survey [37]. Segments are typically much longer than 1 km in length and contain all minor tributaries (i.e., < 2nd order). The predicted fish assemblage for segment utilize the ALIS datasets [27] which provide comparable measures of geology and land cover for each segment catchment (all upstream areas). Measures of slope differ slightly with this dataset, representing the slope along the entire length of the segment.

Managers from within the study area identified catchments as a second level reporting unit (Figure 4). Catchments represent areas that have specific properties to warrant management plans (e.g., drainage, biodiversity or land use). These moderate sized reporting areas include all the drainage area upstream of an outlet to a mainstem tributary, such that a catchment is always a subcomponent of a watershed. In this study, a watershed is a drainage area upstream of the intersection of the stream with the Lake Ontario shoreline. We use the quaternary and tertiary watershed delineations as our means of reporting conditions across the medium and large scale. Quaternary watersheds typically contain one or more major stream systems and a number of smaller systems, while tertiary watersheds contain several quaternary watersheds that have similar geography and comparable hydrograph conditions, [5]. In this study area, there were 27 quaternary and 4 tertiary watersheds.

2.4 Summarizing data within a segment

Based on the recommendation of Wu and Li [69] we evaluated several approaches for summarizing the condition of each reporting area (e.g., mean, median, mode, best, worst, random, and the mean difference of aggregated scores) during a preliminary analysis phase. This analysis determined that the mean condition was not appropriate for categorical data and all but the median and the modal methods are subject to the effect of outliers. The modal value (most frequently occurring) is not effective with small sample sizes. Therefore, the median deviation from expected is used to describe segment condition. Further, variability in the number of sites available for each reporting area and in site classifications combine to effect the confidence in interpretations of landscape patterns. For each segment, a confidence-rating scheme based on a combination of the congruence of ratings and the number of observations in a segment (Table 2) is applied.

To compare segment predictions to the observed data we applied the fish assemblage model to the landscape conditions summarized from the downstream location of each segment. In effect, this provides a comparison of the reliability
of using landscape data alone to predict segment condition. In conducting the analysis, a number of segments where multiple observations were available demonstrated a longitudinal gradient in condition indicative of conditions where local land use activities might be overwhelming larger scale processes. Therefore, an additional classification identifies segments that had at least three categories of condition is applied to this reporting area scale as a means of identifying potentially high priority areas for exploring local restoration opportunities.

2.5 Summarizing data within larger reporting units

The condition of each larger reporting area is determined using a weighted sum approach. Area weighting is a standard statistical approach when the sampling units (segments) vary in size [4]. Table 3 contains the weighting factors for each category of disturbance. For example, a reporting area with data for only three segments of 1000 m of unimpaired, 1500 m likely impaired, and 2000 m impaired lengths would have an overall rating of 0.29 (e.g., 1000*0.7+1500*0.4+2000*0.0/4500). If for the same example dataset used above, segment confidence ratings were high for the unimpaired and the impaired segments and low for the likely impaired segment, the overall confidence rating for this reporting area would be 0.7 (i.e. 1000*1+1500*0.1+2000*1/4500). Results of this analysis are illustrated geospatially and summarized for each level in the reporting hierarchy.

Table 2. Decision criteria used to classify confidence ratings in each segment

<table>
<thead>
<tr>
<th>Number of Sites/Segment</th>
<th>Site Classifications (summary of site ranking by reporting area)</th>
<th>Proposed Confidence Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>---</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>2 identical</td>
<td>Medium</td>
</tr>
<tr>
<td>2</td>
<td>2 different</td>
<td>Low</td>
</tr>
<tr>
<td>3</td>
<td>3 identical</td>
<td>High</td>
</tr>
<tr>
<td>3</td>
<td>2 identical, one site differs by 1 category</td>
<td>Medium</td>
</tr>
<tr>
<td>3</td>
<td>2 identical, one differs by 2+ categories</td>
<td>Low</td>
</tr>
<tr>
<td>3</td>
<td>all different</td>
<td>Low</td>
</tr>
<tr>
<td>4</td>
<td>4 identical</td>
<td>High</td>
</tr>
<tr>
<td>4</td>
<td>3 identical, one site class differs by 1 category</td>
<td>High</td>
</tr>
<tr>
<td>4</td>
<td>3 identical, one site class differs by 2+ categories</td>
<td>Medium</td>
</tr>
<tr>
<td>4</td>
<td>even split (2 site class categories observed)</td>
<td>Low</td>
</tr>
<tr>
<td>4</td>
<td>3 categories observed</td>
<td>Low</td>
</tr>
<tr>
<td>4</td>
<td>4 categories observed</td>
<td>Low</td>
</tr>
<tr>
<td>5 or more</td>
<td>70-100% of site classes identical</td>
<td>High</td>
</tr>
<tr>
<td>5 or more</td>
<td>50-70% site classes identical</td>
<td>Medium</td>
</tr>
<tr>
<td>5 or more</td>
<td>&lt;50% identical</td>
<td>Low</td>
</tr>
</tbody>
</table>
Note 1: identical infers all sites are classified similarly, all sites compared relative to the hierarchy presented above, such that a site that differs by one category is either one higher or lower in the hierarchy.

With this approach, each reporting level larger than a segment has the potential of scores from 0-100. Illustrating the relative rankings of each reporting area on maps poses a challenge, since for clarity, classification is required and like any classification system, selection of the thresholds can influence interpretation of the results. Since one of our objectives is to develop an approach that could be easily understood by the public in support of watershed report cards, academic report card rules are applied: i.e., A > 80, B 70-80, C 60-70, D 50-60 and F < 50.

3.0 Results:

Table 3. Weightings for summary of segment conditions and confidence ratings.

<table>
<thead>
<tr>
<th>Segment Category</th>
<th>Weighting Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better than expected</td>
<td>1.0</td>
</tr>
<tr>
<td>Unimpaired</td>
<td>0.7</td>
</tr>
<tr>
<td>Likely impaired</td>
<td>0.4</td>
</tr>
<tr>
<td>Impaired</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Confidence Rating</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>1.0</td>
</tr>
<tr>
<td>Medium</td>
<td>0.5</td>
</tr>
<tr>
<td>Low</td>
<td>0.1</td>
</tr>
</tbody>
</table>

3.1 Classified Segments
Classification generated a similar numbers of segments in each of the categories of unimpaired, likely impaired and impaired (Table 4). The rarity of the classification better than expected meant that no segments achieved this rating. Seventeen segments (6%) have three classes of conditions indicating they are likely under the influence of local land use activities. Classified segments in closer proximity to urban lands tended to be impaired and likely impaired, in contrast to unimpaired sites, which tended to be located in more northern or eastern areas (Figure 5). Moving in an easterly direction from the Greater Toronto Area (GTA), there is a zone of heterogeneity in the classifications (see the inset of Figure 5). In these areas, proximal segments contrasted greatly.

Table 4: Summary of the median fish assemblage classification and confidence ratings for each segment with field data.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Confidence in rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Not impaired</td>
<td>20</td>
</tr>
<tr>
<td>Likely impaired</td>
<td>10</td>
</tr>
</tbody>
</table>
Figure 5: Classification of segment condition based on fish assemblage.
The majority (63%) of the confidence ratings for segments are low, however the most common cause (55%) of this classification was due to only one site record being available for the segment (Figure 6). Many of the remaining low confidence ratings were the result of different classifications between mainstem and tributary systems associated with one segment, perhaps artificially deflating the confidence that managers might have in the overall classification system used. Only 16% of segments had high confidence ratings in the classification. As an example, Salem Creek east of Brighton (Figure 7), had the greatest contrast in classifications of any segment, with seven unimpaired, three likely impaired and one impaired site. The median classification for this segment is unimpaired. Salem Creek had a dam wash out the year prior to the field surveys reported here, supporting the contention that indeed local land use activities are likely overwhelming landscape influences on this segment of stream.

3.2 Evaluating the accuracy of the models to the segment level

Regardless of the number of observations per segment, there was good correlation between observed and predicted fish assemblage condition for segments classified as either unimpaired (95%), or impaired (81%). Rarely (2%) was a segment classified as unimpaired, for which field data concluded that it as not in this condition (Table 5). The model was largely unable to predict sites classified as being in the intermediate reporting class of likely impaired.
Figure 7: Site classifications for fish assemblages on Salem Creek

Table 5: Comparison of observed and predicted fish assemblage condition within a segment (bolded numbers indicate agreement)

<table>
<thead>
<tr>
<th>Observed median category</th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>unimpaired</td>
<td><strong>54</strong></td>
<td>0</td>
</tr>
<tr>
<td>likely impaired</td>
<td>58</td>
<td><strong>55</strong></td>
</tr>
<tr>
<td>Impaired</td>
<td>9</td>
<td>4</td>
</tr>
</tbody>
</table>
3.3 Reporting fish assemblage condition to higher-level spatial scales:

Fish assemblage data was sufficient to classify 48% of the length of streams located within the 59 example catchments, although the length of coverage in individual catchments ranged from 3-100% (Figure 8). Since this study area has no segments identified as better than expected, the maximum score attainable within a catchment was 70 (or B). Only 16 of the catchments received this rating and these tended to be located in the eastern part of the study area. In contrast, western catchments tended to receive failing classifications, and this grade was the most common in the study area at 61% of catchments (Figure 8; Table 6). Despite the large dataset, only 20% of catchments are classified with high confidence with the majority of catchments having low (47%) ratings.

Data are available for 27 watersheds (Figure 9) and 22 quaternary watersheds (Figure 10; Table 6). As reporting area size increased, the scores generally decreased. This occurred because in this study area, it is rare to find large areas of streams in unimpaired condition. All three tertiary watersheds are classified as F, with scores of: 2HB=37; 2HC=19; and 2HD=48.
Figure 10: Classification of Quaternary watershed condition based on fish assemblage.

Table 6: Summary of catchment and watershed scores by confidence ratings. Criteria to summarize data are defined in Table 3.

<table>
<thead>
<tr>
<th>Catchment Scores</th>
<th>Criteria to summarize data are defined in Table 3.</th>
<th>Confidence Ratings</th>
<th>Catchments</th>
<th>Watersheds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low (&lt;33)</td>
<td>Medium (33-66)</td>
<td>High (&gt; 66)</td>
</tr>
<tr>
<td>B (70 – 80)</td>
<td></td>
<td>8</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>C (60-70)</td>
<td></td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D (50-60)</td>
<td></td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>F (&lt; 50)</td>
<td></td>
<td>17</td>
<td>12</td>
<td>7</td>
</tr>
</tbody>
</table>

1: ratings based on the weighted sum (see Table 3) of the length of classified segments in each reporting area.
4.0 Discussion

While there are many studies that summarize stream data across one reporting scale, we are unaware of any studies that compare the effects of summarizing stream condition at multiple levels in the hierarchy, as is done here. This study demonstrates that there is good agreement between observed and predicted stream fish assemblages at the segment scale, especially if the fish assemblage is impaired or unimpaired. While it is clear that mapping hindcasted site data provides an effective means of summarizing fish assemblage data across a broad geographic area. It is also clear that smoothing effects that result from summarizing findings to larger areas and choice of mapping thresholds has a great affect on interpretation of study findings. While showing classified site conditions and confidence rankings on all maps offers a tool for interpreting the findings, it is clear that scaling effects can have tremendous influence on reporting results of stream data. Finally, this represents the first attempt at describing the condition of stream fish assemblages across the entire study area and suggests that there are disturbing spatial trends. An expansion and discussed of these findings and their implications follows.

4.1 Correlation between observed and predicted condition of segments

The strong correlation between site observations and segment conditions support Seelbach et al [48] and Wang et al [62] recommendation that stream segments represents the most appropriate lowest level of reporting unit with reliable predictability from landscape models. These findings give confidence that segments classified as undisturbed and disturbed are likely accurate and that segments classified as likely impaired warrant additional analysis, and more sampling, to truly ascertain stream condition. That there was poor agreement in the classification of likely impaired segments is likely the result of these sites being located on the cusp of the threshold, where the model response has the steepest slope. In these areas, local conditions could easily have a greater affect on fish communities than at sites that are above the threshold (> 10 LDI) or are in least disturbed areas (< 4 LDI). Use of a more liberal classification scheme, for example adding in another category for 1 SD sites, generated greater contrast in classes, but lower agreement (23%) between the predicted and observed segment condition. Scaling up data to larger spatial areas however, poses a series of challenges to understanding scales of influence.

4.2 Scale of reporting influences perception of condition:

That stream condition tended to decline as scaling size increased is indicative of the challenges associated with scaling influences, although it is difficult to disentangle true effects from perception. Scaling up the information on stream condition poses several challenges because of issues associated with the scales of reporting, homogenization or smoothing effects, classification criteria,
aggregation effects [46, 32, 19] and the nature of the river network [60]. This study clearly shows how these combined effects influence the interpretation of condition as a reporting area increases. Smoothing effects can occur from combining reporting areas that differ in aerial sizes. Further, aggregation of reporting areas (e.g., watersheds) that result from the combination of varying numbers and areas of sub-units can result in misleading results when comparing reporting areas. Terrestrial studies have shown that even when aerial units are of the same size, combining data into increasingly larger aerial units may result in different data values and inferences (see [19]). Because watersheds are a primary influence on the hydrologic cycle that governs conditions at a site and these vary in size, it is not feasible to address smoothing effects by standardizing grain size, as is widely used as a means of addressing this challenge in terrestrial studies [65, 58, 35]. From this decision, smoothing effects were inevitable, since catchment sizes varied and in this study area unimpaired areas tended to be located in smaller streams, ensuring their contribution to larger reporting areas was lower. As reporting area increases the differences between reporting areas in the number of contributing sub-areas and the differences in areas between reporting areas likely magnify these scaling effects, making comparisons between the largest reporting areas (e.g., quaternary and tertiary watersheds) less meaningful.

In this study, ratings do not consider location of the segment within the stream hierarchy as the intention is to provide an unbiased indicator of system wide condition. This unbiased view and the choice of our indicator as a fish assemblage metric generates a different perspective on stream condition than has been reported in several other studies of these tributaries [20, 54, 51, 53]. In these studies, the focus was on migratory salmonids and in each case, the authors have remarked on the generally high production of salmonids found in many of these systems. For example, Stockwell and Jones [20] identified Wilmot Creek as a premiere trout stream in Ontario and the mainstem supports this classification, being classified in good condition (B), but its tributary catchments are impaired or likely impaired, pulling the overall watershed condition down from a B to a D. This more holistic interpretation highlights the value of the mainstem fishery, while also identifying the potential concerns in the smaller catchments.

It may be that future studies will identify a statistical process to address the impact of scaling and smoothing effects and ways to incorporate segment position in the river network. In the meantime, we agree with Tobler [56] and Jeninski and Wu [19] that for tributaries, where aerial units (catchment sizes) are inherently different, that more emphasis should be placed on mapping approaches that enable comparisons that do no rely on aerial units being of similar size.

4.3 Mapping can provide reliable interpretations of landscape patterns
Patterns are most apparent at the smallest reporting units (e.g., segment and catchment) and inclusion of confidence rankings helped provide context for these interpretations by providing a single “score” that helps users understand the limitations of the data. Hierarchy theory predicts these findings, as higher levels are larger and slower to respond than lower levels, particularly for nested hierarchies [69]. Given the challenges to reporting stream condition at higher levels in the hierarchy, it is recommended that if higher level reporting is carried out, that the base data always be shown on the maps. In this way, users can understand the framework that generates subsequent interpretations of condition.

Using this approach, it is clear that streams on the Canadian side of Lake Ontario show progressive degradation in response to intensification of land use. The patterns observed here agree with our knowledge of the effects of increasing development on fish communities [25, 61, 52] and suggest a mechanism for the effect. This study suggests that impacts occur in phases, with watersheds becoming impaired segment by segment and catchment by catchment, until the entire system is affected. Reversal of this creeping crisis requires acknowledgement of this concept and action in smaller segments/catchments, where it is likely that significant fisheries do not currently exist.

4.4 Aggradation and Scoring Systems matter

This study followed the advice of Fotheringham [12] in the selection of the fish assemblage metric based on the raw field data as being sensitive to changes in land use/land cover and as a metric is not modifiable. The multi-variate metric provides an “unbiased” indicator of the fish assemblage and by using the coefficients for each species enable trend analysis to be conducted on subsequent datasets. In fact, In Ontario, these approaches have begun to resonate with managers as this work has been referenced in two fisheries management plans [59, 39], that will monitor change in stream condition over time using these approaches. Jeninski and Wu [19] suggest that given the multitude of spatial and statistical issues associated with reporting on ecosystem condition that measuring changes in condition on one spatial scale over time has the greatest potential for success in linking cause to effect. Although measuring trends in time and space must still define the reference state from which to measure change.

Pearse et al [43] warned that, our perception of change-in-expected condition “is short sighted at best” and that the inability to evaluate, in an objective way, the degree of change in habitat, is largely due to a lack of knowledge of the reference condition. Xu et al [71] faced similar challenges in their attempt to classify the health of Italian lakes. They used the field data to determine the worst (0) and best (100) score, dividing this range into five equal rankings, which ensured the final classification, would include the full range of scores of sites, but only for the existing state. The approach used in this study assumes the existing state is to some degree degraded and applies the hindcasting approach to determine the
reference condition. This approach provides a stable benchmark from which to measure future change in condition.

The scoring system used here is likely to result in biased interpretations of conditions in comparisons between higher levels in the hierarchy (e.g., quaternary and tertiary watersheds), although the scheme may be appropriate for evaluating the individual condition of lower level reporting areas (e.g., watershed and catchment). The chosen scoring system is somewhat arbitrary and is not as important as the base data that contributes to the rating in each reporting area. This study provides a foundation for rating stream condition based on a standard reference condition that is more important that the actual thresholds for each “grade”. For example, choosing a more conservative rating scheme for a geographic area that is dominated by urban land use (e.g., scoring F as < 40) would result in a better overall grading of stream condition, but provided the same base data are used, would not compromise the scientific credibility of the data. Rather, such a decision might provide a more realistic evaluation of stream condition for an urbanized landscape and enable consistent reporting to the public of watershed health [3].

Finally, providing the base level mapping results regardless of the level in the hierarchy being reported provides an opportunity to identify potential areas requiring management action that may not be obvious otherwise.

4.5 Utility of this approach for guiding management activities

These findings provide guidance to management activities such as where additional monitoring should occur; where land use planning could focus new initiatives; and where stewardship might be more beneficial. As suggested by Wang et al [62], segments identified as unimpaired should be the focus for management action that emphasizes protection. In these areas, instream activities such as cover enhancement or bank stabilization would only be beneficial where local areas are a concern. On the other hand, impaired segments are likely to require large scale planning initiatives at the watershed level to address cumulative effects from land and instream disturbances. The 17 stream-segments classified as likely impaired should be the targets of further investigation in order to confirm the source of the changes and if present to initiate stewardship activities to rectify the local problems (Figure 11). Applying this quantitative approach to prioritize stream and riparian vegetation restoration projects could improve the benefits from these activities at multiple levels. For example, priority areas for vegetation restoration would represent areas where impaired segments along with terrestrial restoration objectives overlap, thereby meeting both terrestrial and aquatic objectives. Resource managers with large
geographic area mandates could incorporate an adaptive monitoring strategy that initially targets all segments with one sample site. Later, additional sampling could target areas classified as likely impaired from the first sample. Such a design takes advantage of the high predictive power of the impaired and unimpaired condition and optimizes the potential to identify locations where local land use is having its greatest effect (see above).

4.6 Other factors to consider and science needs

A core component and therefore limitation of this study is the use of modeling results based on coarse metrics, of landscape properties based on varying spatial resolution, to determine the condition of the fish community. As indicated earlier, there is much confounding of influences between the landscape variables and spatial autocorrelation not considered in these models. Further, other studies of spatial interaction modelling show that both scale and zoning changes affects model goodness-of-the-fit and parameter estimates (see [19] for a review). Further, this analysis does not consider proximity [61] or spatial heterogeneity of land use typically applied to terrestrial spatial analysis [11]. Therefore, as more refined data layers and new approaches for analyzing the data emerge model predictions are likely to improve. Future analysis of the state of the streams in this area will likely require a reanalysis of the condition of each site, while still
using the same metric of fish assemblage. In this way, trend-in-time analysis will enable comparisons in condition across the landscape.

A conceptual understanding of the effects of hierarchical scaling is well established, (17, 45, 18, 42, 9 for examples) and papers are beginning to emerge that demonstrate the pathways of influence at various spatial scales and at the various levels in the river network (see 10, 61, 71, 53). However, many of the existing solutions to these scaling and autocorrelation challenges routinely applied to terrestrial studies cannot be applied to aquatic studies because of the intertwined issue that watershed areas vary and rivers are organized in a hierarchical network. There continues to be an urgent need for new analytical approaches to tease out these influences on streams (see [66] for an earlier plea). If this approach is not feasible, as suggested by Openshaw et al. [41], also see [12], [19]) then the alternative is to develop a framework around which consistent application of visual interpretations of maps can be applied across the landscape. This study suggests that this approach should include mapping at the segment level, should include site-specific findings and incorporate some measure of confidence in classifications and that caution is necessary if segment data is scaled up to larger spatial areas.

5.0 Conclusions and Recommendations:

This study suggests that the scale at which the stream data is collected (e.g., site or segment) is the best scale at which to present the findings and that as reporting area increases, a variety of statistical and spatial issues combine to reduce the reliability of comparisons in condition between spatial areas. This study demonstrates an approach for summarizing and illustrating site data to higher levels of reporting in the stream hierarchy that both maintain the original data and provide a measure of confidence in the classifications and interpretations. This study provides insight for the pros and cons of various levels of sampling intensity for studies directed at evaluating stream condition. That data from only one site can be sufficient to identify whether a segment is classified as either unimpaired or impaired will be welcome news to managers struggling with the logistics of conducting large scale environmental monitoring. Additionally, if a sample determines a segment to be likely impaired or if a manager suspects local land use is unduly influencing stream condition, this study demonstrates how having at least three sites of data within a stream segment that cover a large longitudinal gradient can help identify potential areas where local restoration activities can be effective. This study provides guidance to the development of Watershed Report Cards or state of the ecosystem reports for each reporting area and provides a point in time assessment of stream conditions that will enable future comparisons. That few watersheds in this study did not demonstrate at least some segments in a degraded condition is evidence that the incremental loss of fish habitat described by Pearce (1985) is still occurring in this area. Incorporating the approach in an adaptive framework optimizes a study design with the goal of classifying the condition of the resources in a jurisdiction and
identifying priority areas for protection (unimpaired segments); land use planning to address large scale issues (impaired segments) or local restoration (segments with a gradient of conditions). Finally, there is an urgent need to better understand the scales of influence of the processes that result in degraded stream conditions, if scaling influence is to be incorporated in watershed restoration activities in time to prevent an even greater impact to streams in this area from the next wave of urban development.

Acknowledgements

This paper is dedicated to the many field technicians and biologists who collected the data that made this analysis possible, specifically, Jeff Anderson, Jon Clayton, Marc Desjardins, and Scott Jarvie, Ian Kelsey and the crews from the Salmonid Ecology Unit. Rebecca Astolfo and Debbie Irvine-Alger provided GIS and mapping support. Scott Gibson helped with the data analysis. Helpful comments on the methods applied were received from Al Dextrase, Nick Jones, Nigel Lester, Bob Metcalfe, and Scott Reid and on an earlier draft of this report from Bob Bergmann, Doug Dodge, Tim Haxton, Bruce Kilgour, Danijela Puric-Mladenovic and Christine Tu. The Ministry of Natural Resources, Water Resources Information Project and by Environment Canada through the Water Use and Supply project provided funding for this project. Cheryl Lewis, Scott Christilaw, Wendy Leger and Silvia Strobl provided administrative support.

References:


[38] Ontario Ministry of Natural Resources (OMNR), *LANDSAT-I MSS: Inventory, Monitoring and Assessment Section*, Peterborough, ON, 1999.


