# Minimal requirements for monitoring abundance and distribution of fish communities

*Eric A. Parkinson[1], David Tesch[2] and Art Tautz[3]*

## Abstract

Most methods of monitoring fish populations focus on a localized geographic scale, often individual populations. Strategic planning for biodiversity conservation requires information for multiple species over their entire range but monitoring programming are rarely designed to target such large geographic areas. We utilized existing occurrence data to predict the probability of occurrence for all BC fish species in 18000 watersheds >3rd order at a 1:50000 map scale. These predictions incorporate expert opinions on range, habitat information and watershed connections to observed occurrences to predict occurrence with increasing levels of certainty. This information can used to identify locations where presence/absence data on particular species would provide large increases in prediction quality with minimal amounts of field data collection. Critical information gaps can be used to optimize field data collection activities.

## Introduction

Conservation planning depends heavily on monitoring programs which provides data on species distributions and trends, which are typically inferred from statistical models because of incomplete data coverage. The estimation process involves a series of steps: assembling both the response (fish occurrence) and predictor (habitat) data sets, quality assurance of the data (false reports of presence from species misidentification or accidental occurrences, false reports of absence), estimating a statistical relationship between habitat variables and species presence, testing this model and moving from a habitat suitability model to estimates of probability of occurrence. In this paper, we use examples to illustrate the process by which we are modelling the distribution of all 81 of BCs freshwater fish species.

1 Senior Scientist, Ecosystems Branch, BC Ministry of Environment, 2202 Main Mall, University of BC, Vancouver BC, V6T 1Z4
2 1st Floor - 395 Waterfront Crescent, Victoria BC, V8T 5K7
3 BC Ministry of Environment, 2202 Main Mall, University of British Columbia, Vancouver BC, V6T 1Z4

**Citation:** Parkinson, E.A., D. Tesch, and A. Tautz. 2007. Minimal requirements for monitoring abundance and distribution of fish communities. Paper presented at the "Monitoring the Effectiveness of Biological Conservation" conference, 2-4 November 2004, Richmond, BC. Available at: http://www.forrex.org/events/mebc/papers.html

**Notes:** This paper has been published "as presented" at the Conference, and has not been peer reviewed prior to posting. © Copyright 2007 by the authors.

The use of habitat suitability models to infer species distribution and trends has received a lot of attention recently because of the many successes in using geographic information systems to build large data bases that include both habitat information and observations of species occurrences (e.g. James and McCulloch 2002). The fish occurrence data set that we used is typical in that it has been assembled from a variety of sources with varying degrees of quality assurance and spatial coverage. Both location and species identification are common issues in these types of databases (Altaba 1997, Daan 2001), which means that ensuring consistent data quality is a critical component of data assembly.

A variety of statistical methods are available for quantifying species-habitat associations. However, the differences between the leading contenders are often small (Manel et al 1999, Segurado and Araújo 2004) and no single methodology has a clear advantage in all cases, especially if alternative methods of assessing prediction errors are considered (Fielding and Bell 1997). Statistical relationships may identify locations with suitable habitat but are not a definitive indicator of species presence. Species may be excluded by competitive interactions, unsuitable values of an unmeasured habitat variable or an inability to reach isolated patches. This linkage is especially problematic in freshwater fish which cannot move across watershed boundaries or up over falls and other velocity barriers.

## Methods

Species distributions were mapped at the 1:50,000, 3$^{rd}$ order watershed scale for 19,249 watershed-based polygons in British Columbia, Canada. Each polygon is either a watershed, a coastal area with one or more streams with order <3 or an island. Watersheds of order >3 contain hierarchically arranged parent and daughter polygons. The boundaries of smaller polygons within a watershed coincide with changes in stream order and boundaries of watershed groups. Polygons without mapped surface water were assumed not to support fish, leaving 17,815 polygons with a total area of 944,000 km$^2$. These polygons accounted for > 99% of the total area of BC.

Our species occurrence data set was extracted from the Fisheries Information Summary System (FISS), a database that is maintained jointly by the Governments of BC and Canada (http://www.env.gov.bc.ca/fish/fiss/update.htm), which contains species occurrence records from a variety of data sources. These records are linked to the watershed polygons using watershed codes and geographic coordinates. In 2003, 117,000 occurrence records for 81 species of freshwater fish in BC were extracted from FISS. Although the source of each record can be traced, some duplication occurred because some sources relied on earlier records for some species. Much of this duplication was removed by using location and date information but some duplicated records are probably still present, especially for common species. The error checking process ensured that none of these records were high leverage points.

Currently accepted ranges (McPhail and Carveth 1993) were used as a baseline for defining more exact distributions. The terms “range” and “distribution” are intended to represent “Extent of Occurrence” and “Area of Occupancy” as defined by IUCN (2001). Out-of-range records were all individually checked. Decisions on range extensions were based the amount of sampling, the prevalence of records in the extension and the presence of plausible alternative species. The number of reliable records required to justify a range extension ranged from one (rare species) to three (common migratory species).

Habitat variables were derived from three major sources: the polygons themselves, intersection with other maps, and the Watersheds BC FISS database (Table 1). Since watersheds with MaxOrder > 3 contain both tributary and headwater polygons, some variables can take on the values of either the polygon or the watershed. For example, area can be the polygon area or the area of the entire watershed including all the tributary polygons.

We used multiple logistic regression to model the distribution (presence/absence) of each species as a function of habitat characteristics. The twenty-six habitat descriptors (Table 2) were reduced to principle component scores, which described vectors that captured habitat variation among watersheds within the range of a species. This procedure removed the correlation among variables and produced axes that correspond to factors, such as watershed size, temperature, lake influence, gradient, dryness and glacial influence, which are important in driving distributions of stream fishes. Only data from watersheds within a species range, that had visited at least once, were used in the logistic regression. The dependent variables were PCAs describing variation among watersheds within the range and the independent variable was whether or not a given species had ever been observed in each sampled watershed. The output of this analysis is a Logistic Regression Habitat Score (LRHS), which can be used to rank watersheds in terms of their suitability for each species.

## Results

### Data Checks

Out-of-Range Occurrences

Two percent (2233) species occurrence records fell outside the published ranges. Most (1939) involved probable species identification errors for records bull trout (*Salvelinus confluentus*) and Dolly Varden (*S. malma*), which have only recently been recognized as separate species and are difficult to distinguish morphologically (Hass 1993). Species identification error was also judged to be the cause of 11 out-of-range records involving two sculpin species pairs, which are also difficult to distinguish morphologically. Range extensions via introductions were recognized in 3 species (carp *Cyprinus carpio*, rainbow trout *Oncorhynchus mykiss*, pumpkinseed sunfish *Lepomis gibbosus*). A range contraction in one species (chinook salmon, *O. tshawytscha*) due to hydroelectric development was confirmed. A range extension was not made for White Sturgeon (*Acipenser transmontanus*) based on 11 the out-of-range occurrences from a single source. Other sources have reported repeated absences of this species in the same area (Upper Columbia).

Table 1. Habitat variables associated with each watershed-based polygon

| Major Theme | Variable Name | Description |
|---|---|---|
| Size | Order | The maximum stream order (REF) |
| Size | Magnitude | The maximum stream magnitude |
| Size | Area | Area (ha) of the watershed |
| Size | BL Length | Length of blue lines (streams lakes etc.) in the watershed |
| Size | QBank | Bankful discharge estimated empirically, corrected empirically at small stream sizes. $\mathbf{Q}_{\mathbf{bank}} = k * DrainageArea^{0.68}$ |
| Size | Width | Bankful width estimated from the hydraulic equation. $\mathbf{W}_{\mathbf{bank}} = a\ \mathbf{Q}_{\mathbf{bank}}\ b$ |
| Temperature | MinTemp | Minimum summer temperature (Porter ) |
| Temperature | MeanTemp | Mean summer temperature (Porter ) |
| Temperature | Max Temp | Maximum summer temperature (Porter ) |
| Temperature | MeanElev | Mean elevation of the watershed |
| Temperature | MinElev | Minimum elevation of the watershed |
| Lake Influence | WetP | Proportion of watershed area occupied by wetlands |
| Lake Influence | LakeP | Proportion of watershed area occupied by wetlands |
| Lake Influence | BLStreamP | Proportion of blue line in watershed that are streams (i.e. not lakes or wetlands) |
| Gradient | ElevSDPoly | Standard deviation of elevation within the polygon |
| Gradient | ElevSDWshd | Standard deviation of elevation within the watershed |
| Gradient | GradTerrPoly | Average terrain gradient of the polygon |
| Gradient | GradTerrWshd | Average terrain gradient of the watershed |
| Gradient | GradBLPoly | Average gradient of blue lines in the polygon |
| Gradient | GradBLWshd | Average gradient of blue lines in the watershed |
| Gradient | GradBLMStem | Average gradient of polygons mainstem blue line |
| Dryness | KFactor | A measure of wateryield |
| Dryness | BLDensity | Density of stream bluelines |
| Glacial | IceP | Proportion of watershed area that is permanent ice |
| Glacial | ElevMax | Maximum elevation of the watershed |
| Glacial | AlpP | Proportion of the watershed area that is alpine |

Table 2. Distribution of WP occurrences in watersheds classified using a Logistic Regression Habitat Score. Lines 1, 2 and 3 correspond to scores that include 60%, 90% and 100% of known WP occurrences.

| | Sampled Watersheds | | | | |
|---|---|---|---|---|---|
| Logistic Regression Habitat Score | All Sites | WP Not Observed | WP Observed | % with known WP Observations | Unsampled Watersheds |
| > 0.37 | 27 | 8 | 19 | 70% | 4 |
| > 0.083 | 56 | 28 | 28 | 50% | 28 |
| > 0.024 | 115 | 84 | 31 | 27% | 93 |
| < 0.024 | 502 | 502 | 0 | 0% | 1454 |
| Total Sites | 617 | 586 | 31 | 5% | 1547 |

Identification of Habitat Outliers: e.g. White Sturgeon
The two highest leverage points in the initial logistic regression of white sturgeon presence/absence as a function of habitat PCAs were both headwater streams, which typically support only salmonids and sculpins. Dewdney Creek is an above-barrier tributary of the Coquihalla River which may be a miscoding of Dewdney Slough, where sturgeon are common. The species list for the Hammling Creek site includes 11 species, most of which are not typically found in headwater streams. Both were eliminated from the final habitat model.

**Species Distributions**

There were 5 general methods for mapping distributions within ranges. Restricted range species (19 species or populations) are known to be in <10 watersheds. Expert opinion of habitat must be used for poorly sampled species (31 species) have too few records (<10) to generate a habitat model. Invasive exotics (8 species) have distributions that are expanding through a combination of natural and anthropogenic dispersal. Their distribution maps include current known occurrences and potential distribution based on habitat associations in their native range. Statistical habitat models are evaluated for 37 species that include habitat specialists and habitat generalists. Habitat specialists have an informative habitat model because occurrences are restricted to well-defined habitat types. For habitat generalists, the habitat model is uninformative because the species is found in a variety of habitats throughout their range.

Informative Habitat Models: e.g. Walleye (Stizostedion vitreum, WP)
The WP logistic regression model accounted for 55% of the variance in occurrence. The LRH Scores predict that WP are more likely to be present in watersheds that are lower in elevation and gradient, larger in size and with a higher proportion of wetlands. This assessment is consistent with a distribution that stretches across the prairies, the Canadian Shield and into the American mid west. A habitat preference for lakes and large, slow moving rivers is well documented (Scott and Crossman 1973). Within the BC native range of WP, 617 watersheds have had at least one sampling event. WP have been observed in 31 (5%) of these 617 watersheds (Figure 1). WP were not

observed in watersheds with LRHS < 0.024. Of the 115 watersheds with an LRHS >.024, WP were observed in 31 (27%). Within the native range of WP, 1547 watersheds had not been sampled. Only 93 of unsampled watersheds have LRHS values that indicate suitable habitat for WP. WP are also an exotic species that has been introduced into the Columbia River where it is found in the mainstem Columbia below Arrow lakes and in the Kettle River.

Uninformative Habitat Models: e.g., Slimy Sculpin (*Cottus cognatus* CCG)
The logistic regression model was poor, accounting for only 13% of the variance in occurrence. The habitat model predicts that CCG are more likely to be present at lower elevations and moderate gradients, in large watersheds with a low prevalence of lakes and wetlands. Both McPhail and Lindsey (1973) and Scott and Crossman (1973) emphasize a preference for cold temperatures and gravely substrate.

The geographic pattern of occurrences (Figure 2) suggests that CCG may be absent from large areas of its range, since CCG is often locally abundant and is among the easiest of sculpins to identify. To quantify this, the number of CCG observations and the number of observations of all species were tallied for each watershed group (WG) within the CCG range. The ratio of these provides an index of the importance of CCG as a component of the WG fish fauna. Sixty-eight (out of 176) WGs where there were 2 or fewer CCG observations and where CCG made up <0.5% of all observations within a WSG were designated as peripheral range (Figure 2). These WGs accounted for only 10/1960 (0.5%) of CCG observations.

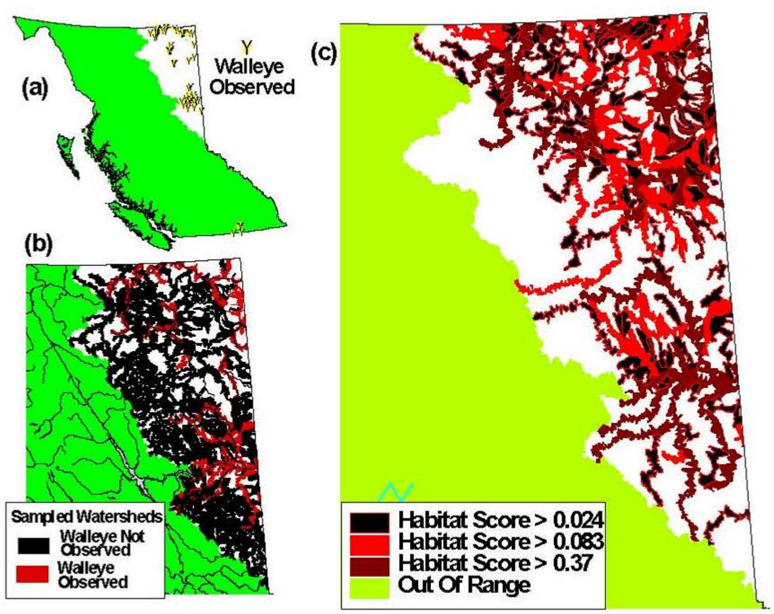


Figure 1. a) the native range and observed locations of walleye (WP) in BC.
(b) Sampled watersheds within the native range with and without WP observed.
(c) Habitat Score from the logistic regression. The breakpoints in the graduated colour are chosen such that 60% of observed occurrences are in watersheds with the darkest colour, 90% of occurrences are in watersheds with the darkest 2 colours, and no occurrences of WP have been observed in the uncoloured watersheds.

## Discussion

Monitoring the distribution of a species can be viewed as an information hierarchy that provides increasing specific information at various spatial and temporal scales (Angermeier et al 2002). At the least-detailed level, ranges define broadly contiguous geographic areas outside of which the species is not found. Within ranges, species occupy suitable habitat at increasingly finer scales such as watershed, reaches or microhabitats. Suitable habitats can be ranked in terms of the maximum density or maximum rate of population growth that they can support. Status can be defined in terms of the current density or growth rate as a fraction of these maxima. Watershed scale variables have been successfully used to predict the distribution of single species (Dunham et al. 2003) and the effects of single variables on multiple species (Poff and Allan 1995) but not with the entire fish fauna over an area as large as BC.

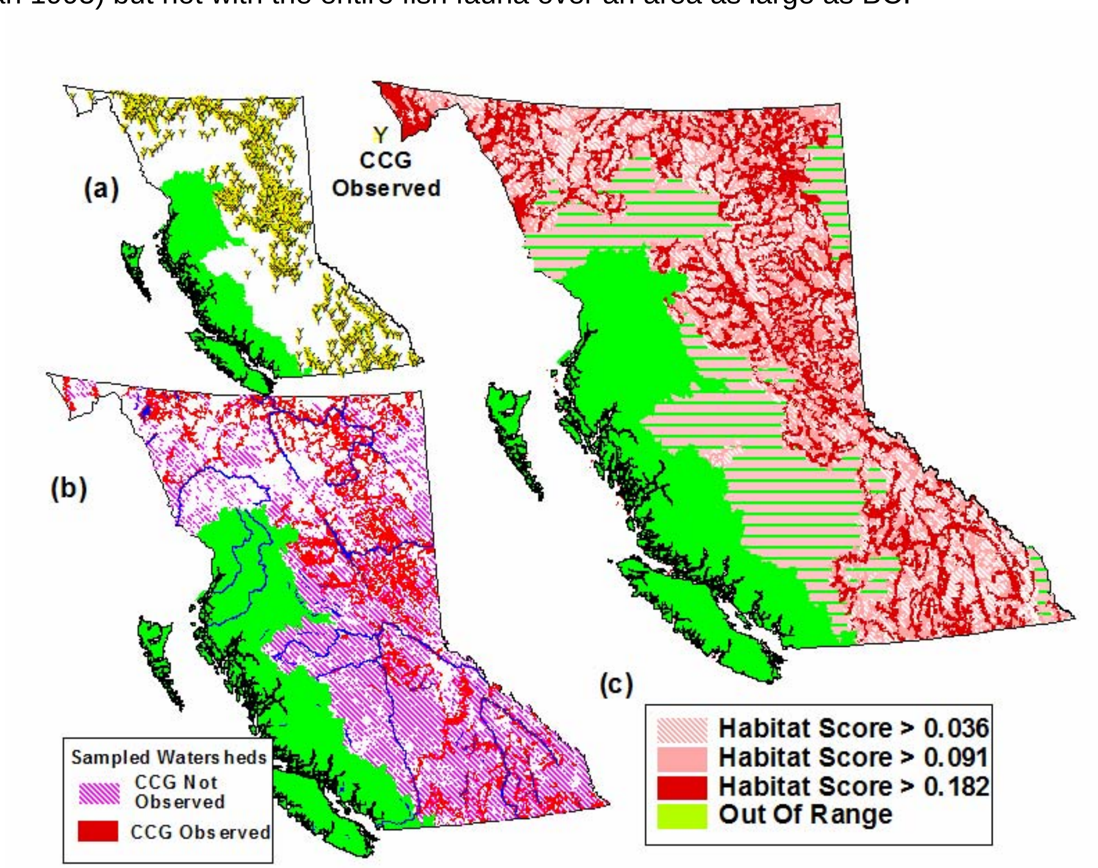


Figure 2. (a) the native range and observed locations of slimy sculpin (CCG) in BC. (b) Sampled watersheds within the native range with and without CCG observed. Uncoloured areas have not been sampled (c) Habitat Score from the logistic regression. The breakpoints in the graduated colour are chosen such that 60% of observed occurrences are in watersheds with the darkest colour, 90% of occurrences are in watersheds with the darkest 2 colours and 1% have been observed in the uncoloured watersheds. Out of range watersheds are green. Areas crosshatched with green represent <1% of 1910 species sampling events.

Range and distribution models should be used more as indicators of where a species is unlikely to be found rather than as indicators of presence. Range boundaries are intended to exclude erroneous occurrences and occasional reports of migratory species outside their normal range (“accidentals” in the bird literature REF) but many species are also absent from large areas of their range. Within ranges, the distribution of suitable habitats often does not reflect species distributions. Distributions can vary temporally with seasons, density or current habitat quality. Suitable habitat may not be occupied because of lack of access or biotic interactions.

At coarse spatial scales, however, (Angermeier et al 2002) suggest that habitat and dispersal ability are the strongest indicators of absence. Occupation of unsuitable habitat is most common at spatial scales that are smaller than typical dispersal distances of individuals. Unsuitable (sink) habitat can be permanently occupied because of continuous emigration of individuals from adjacent suitable (source) habitat (REF from Grouse guy). For fish at the watershed scale, occupation of unsuitable habitat should be less of an issue since barriers to dispersal across watershed boundaries, and homing of migratory species to natal areas should limit distributions to suitable watersheds. Although dispersal into unsuitable habitat is probably not an issue at the watershed scale, barriers which prevent dispersal into suitable habitat are a major obstacle in building accurate distribution models.

Our analyses suggest that currently available data can provide some information on the distribution of all BC fish at the 3$^{rd}$ order watershed scale. About 25% of species are confined to small, well-defined geographic areas and watershed-scale habitat suitability is not an issue. Another 25% of species have too few records to define habitat relationships, but expert opinion, based on similar species and data from other jurisdictions, can be used to rank habitat suitability. We are in the process of evaluating habitat models for the remaining species. For some, such as Walleye, habitat suitability, rather than dispersal appears to be the major factor determining distribution. Core and peripheral habitats can be defined on the basis of reported occurrences. For others, such as slimy sculpin, current and post-glacial barriers to dispersal have limited access to suitable habitat. Access models, based on biogeography, the presence of other species, the presence of known barriers and the presence of upstream observations will have to be an important component of the models for these species.

Relatively small amounts of additional sampling would likely result in large increases in the accuracy of our predictions. Stratified sampling across habitat gradients, such as watershed size or temperature, are necessary in order to validate the models. Standardized methods (RIC REF?) must be used in order to ensure that erroneous observations are minimized and that the probability of absence can be estimated.

## References

Altaba, CR. 1997. Documenting biodiversity: the need for species identifications. Trends in Ecol. Evol. 12:358-359.

Angermeier, PL, Krueger, KL and Dolloff, CA. 2002. Discontinuity in stream-fish distributions: Implications for assessing and predicting species occurrence. P. 519-527. *In:* Predicting species occurrences : issues of accuracy and scale. *Ed by* J. Michael Scott [et al.]. Scott, J. Michael. Island Press. Washington DC.

Daan, N. 2001. The IBTS database: a plea for quality control. ICES 2001 Annual Science Conference CM 2001/T:03.

Dunham, J, Rieman, B and Chandler, G. 2003. Influences of Temperature and Environmental Variables on the Distribution of Bull Trout within Streams at the Southern Margin of Its Range. N. Am. J. Fish. Manage. 23:894-904.

Fielding, AH and Bell, JF 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Cons. 24: 38–49.

IUCN. 2001. IUCN Red List Categories and Criteria: Version 3.1. IUCN Species Survival Commission. IUCN, Gland, Switzerland and Cambridge,UK. ii + 30 pp. available at: http://www.redlist.org/info/categories_criteria2001.html

James F.C. and C.E. McCulloch 2002. Predicting species presence and abundance. p. 461-465 In Predicting Species Occurrences: Issues of Accuracy and Scale. Edited by J. M. Scott, P. J. Heglund, J. B. Haufler, M. L. Morrison, M. G. Raphael, W. B. Wall, and F. Samson. Island Press, Washington.

Manel S, Dias, JM, Buckton, ST and Ormerod, SJ. 1999. Alternative methods for predicting species distribution: An illustration with Himalayan river birds. J. Appl. Ecol. 36:734-747.

MacPhail, JD and Carveth, R. 1993. Field Key to the Freshwater Fishes of British Columbia. Prov. Of BC, Resource Inventory Publications. Available at: http://srmwww.gov.bc.ca/risc/pubs/aquatic/freshfish/index.htm

Poff, NL and Allan, JD. 1995. Functional organization of stream fish assemblages in relation to hydrological variability. Ecology 76:606-627.

Segurado P and Araújo, MB. 2004. An evaluation of methods for modelling species distributions. J Biogeogr. 31:1555-1568.