

Towards generalised reference condition models for environmental assessment: a case study on rivers in Atlantic Canada

D. G. Armanini · W. A. Monk · L. Carter · D. Cote ·
D. J. Baird

Received: 7 March 2012 / Accepted: 21 November 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Evaluation of the ecological status of river sites in Canada is supported by building models using the reference condition approach. However, geography, data scarcity and inter-operability constraints have frustrated attempts to monitor national-scale status and trends. This issue is particularly true in Atlantic Canada, where no ecological assessment system is currently available. Here, we present a reference condition model based on the River Invertebrate Prediction and Classification System approach with regional-scale applicability. To achieve this, we used biological monitoring data collected from wadeable streams across Atlantic Canada together with freely available, nationally consistent geographic information system (GIS)

environmental data layers. For the first time, we demonstrated that it is possible to use data generated from different studies, even when collected using different sampling methods, to generate a robust predictive model. This model was successfully generated and tested using GIS-based rather than local habitat variables and showed improved performance when compared to a null model. In addition, ecological quality ratio data derived from the model responded to observed stressors in a test dataset. Implications for future large-scale implementation of river biomonitoring using a standardised approach with global application are presented.

Keywords Macroinvertebrate · Bioassessment · Geographic information system · RIVPACS · Monitoring

D. G. Armanini · W. A. Monk · D. J. Baird
Environment Canada @ Canadian Rivers Institute,
Department of Biology, University of New Brunswick,
10 Bailey Drive, PO Box 4400, Fredericton,
NB E3B 5A3, Canada

D. G. Armanini (✉)
Prothea srl,
Milan, Italy
e-mail: d.armanini@protheagroup.com

L. Carter
Environment Canada/Atlantic Water Quality Monitoring,
45 Alderney Drive,
Dartmouth, NS B2Y 2N6, Canada

D. Cote
Parks Canada, Terra Nova National Park,
Glovertown, NL, Canada

Introduction

Our need to monitor the ecological condition of river ecosystems has created two schools of ecological assessment: multivariate prediction and multimetric description. The former has been widely applied in Canada (e.g. Reynoldson et al. 1995), the UK (e.g. Clarke et al. 1996), Australia (e.g. Davies 2000) and some USA states (Bonada et al. 2006), whilst the latter has been adopted mainly by continental Europe (e.g. Hering et al. 2004; Buffagni et al. 2009) and also by some US states (e.g. Barbour et al. 1999). Each approach is based on a comparison of observed conditions

in natural or nearly natural reference sites and test sites of regulatory concern. The reference condition approach (RCA) in its simplest form is a comparison of an observed condition (O) with an expected condition (E) (Reynoldson et al. 2001). In Canada, a multivariate prediction model, Benthic Assessment of Sediment (BEAST), was developed by Reynoldson et al. (1995) to assess the ecological quality of streams and rivers.

Recent efforts have led to the development of a national-scale biomonitoring program managed by Environment Canada, known as the Canadian Aquatic Biomonitoring Network (CABIN; <http://cabin.cciw.ca>). The establishment of the CABIN program has catalyzed a process of data and metadata integration, which focuses on Canadian freshwater ecosystems. Adopting a partner-network model (i.e. a model where several groups and institutions contribute to the implementation of the program), the program has focused on the sampling of benthic macroinvertebrates, the most commonly used biological indicator group for biomonitoring of rivers (e.g. Resh 2008). Predictive models have been developed for selected regions of the country (e.g. the Fraser River, British Columbia—BEAST, Reynoldson et al. 2001) but not yet for Atlantic Canada (comprising the provinces of New Brunswick, Prince Edward Island (PEI), Nova Scotia and Newfoundland and Labrador). However, these models require data to be collected in a standardised format, with field-observed local habitat variables required for model construction. This places a significant constraint on the modelling process, particularly in data-sparse remote regions such as those found in Canada, where biological data may exist in a compatible format, yet standardised field-observed variables are either inconsistently measured, or absent.

Geospatial data are an effective surrogate for local field-collected variables in providing the information needed to develop multivariate prediction models (Hawkins et al. 2000; Hargett et al. 2007; Poquet et al. 2009). Geology and climate, parameters widely mapped using geographic information system (GIS), have been recognized as large-scale drivers of macroinvertebrate community composition in the literature (Omernik 1987; Snelder et al. 2004). One of the advantages of using national-scale geospatial data is the ability to directly compare habitat data consistently amongst locations, thereby reducing or eliminating the challenge of data interoperability (i.e. the ability to integrate data collected with different protocols or different operators for ecological assessment), which can arise in bottom-

up, partner-based networks, such as EC's CABIN program.

This paper describes the development of a RCA model based on the River Invertebrate Prediction and Classification System (RIVPACS) approach using landscape variables extracted from standardised geospatial data paired with benthic macroinvertebrate data generated from samples of wadeable streams across New Brunswick, Nova Scotia, Prince Edward Island and Newfoundland. The analyses also aim to validate the developed model against independent datasets and against a null model. A null model is formulated on an assumption that the occurrence probabilities of taxa in reference sites are not driven by natural-gradient variables (Van Sickle et al. 2005), and thus, a null model does not employ the procedure of clustering reference sites into groups as per the standard RIVPACS approach. Finally, one local case study with known environmental impairments was used to test the performance of the method.

Data and methods

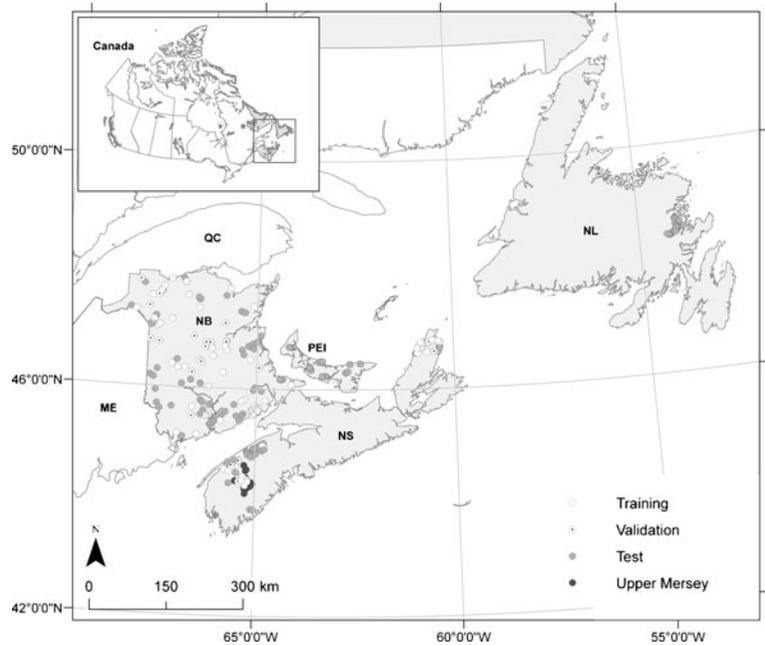
Study area

Figure 1 shows locations of the 582 benthic macroinvertebrate samples collected between 2002 and 2008 in the Atlantic Maritime ecozone. All samples were collected in unique locations for reference sites (see below for definitions), whilst repeated samples were collected at some test sites. Forests in the Atlantic Maritime ecozone constitute 90 % of the total land cover and are referred to as the New England–Acadian forests, comprising temperate broadleaf and mixed forest (Benke and Cushing 2005). Newfoundland is located in the north central boreal forest subregion (Meades and Moores 1994), and it shows, in broader terms, a cool summer subtype of a humid continental climate, a hilly topography and forests that comprise about 50 % of the area (Gauthier, Poulin, Theriault, Ltd. 1977).

Benthic macroinvertebrate data employed in our study were drawn from the following sources (see also Table 1 for a summary):

- (a) Environment Canada CABIN dataset (kick-net)
Riverine benthic macroinvertebrate data were extracted from Environment Canada's online

Fig. 1 The distribution of biomonitoring sites employed for reference condition approach model development collected in New Brunswick, Nova Scotia, Prince Edward Island and Newfoundland. Training and validation datasets only include sites classified as ‘reference’ (see text for further explanation) whilst potentially impacted samples are indicated as ‘test’



CABIN database (<http://cabin.cciw.ca>, consulted December 2009; data usage permission was obtained wherever needed) for all samples located in New Brunswick, Nova Scotia, PEI and Newfoundland. All samples were previously collected using a standardized travelling kick-net method, which disturbs the stream substrate using a triangular net of 400- μ m mesh size, whilst walking backwards upstream. The collector zig-zags across the river from bank to bank in an

upstream direction for 3 min (Reynoldson et al. 2007). Physical habitat and water chemistry data were collected following standardised procedures outlined in detail in the CABIN protocol (Reynoldson et al. 2007).

Two additional data sets were used: one from Environment Canada’s National Agri-Environmental Standards Initiative project (<http://tinyurl.com/EC-NAESI>) and one from a National Defense project based at Canadian

Table 1 List of data sources considered in the present paper with a summary of the main data set features

Dataset	CABIN database			New Brunswick ^a
	Standard CABIN	NAESI	National Defence	
Sampling Years	2002–2008	2006	2008	2004, 2006, 2007, 2008
Provinces	NB, NS, PEI, NL	NB	NB	NB
Sampling net		Kick-net		U-net
Sampling mesh size		400 μ m		(a) 250 μ m and >1 mm but only latter retained and (b) 400
Sampling method	Single 3-min sample		Composite of 3 \times 1-min samples	3 rep samples each of 3 \times 1-min collections
Habitat			Composite	
Season			Fall	
Taxonomic resolution			Mixed level, adjusted	

^a Two datasets combined after specific data analysis (Brua et al. 2010)

Forces Base Gagetown. Data from both projects were collected according to the standard CABIN approach with the only difference being that in the former study, samples were a composite of three independent 1-min samples, instead of a standard single 3-min sample.

(b) Province of New Brunswick dataset (U-net)

An analysis by Brua et al. (2010) demonstrated that benthic macroinvertebrate samples from two datasets, one from the Canadian Rivers Institute (CRI) and one from the New Brunswick Department of Environment (NBDENV), were statistically comparable and were merged accordingly.

Samples were collected using a U-net (250- μ m mesh size for CRI data and 400- μ m mesh size for NBDENV data). Large rocks within the U-net sampling area were rubbed by hand inside a U-net to dislodge any clinging macroinvertebrates into the net. Bottom sediments were disturbed by hand for 1 min to a depth of approximately 2 cm. At each site, three replicate samples were taken, and within each replicate, three 1-min U-net collections were pooled. NBDENV samples were processed using a 400- μ m sieve whilst CRI samples were split between a fine (250 μ m to <1 mm) and coarse (>1 mm) fraction. Only the coarse fraction of the macroinvertebrate community data from the CRI samples was retained for analysis to improve data comparability (Monk and Curry 2007).

Temporal variability

In order to reduce the effects of temporal variability on the analysis, all the samples included in the analysis were collected in the fall, thus limiting the potential impact of seasonal difference. Assessments of site-level inter-annual variability are currently limited because repeat sampled data are not available for the majority of sites. For the analyses in this paper, we have included reference site data collected from 2002 to 2008, which may account for some sources of inter-annual variability in the model development.

Taxonomic adjustment

Site data were collated in a taxon abundance matrix, with individual taxa adjusted mostly to family level, with some exceptions at higher levels (i.e. Acarina, Collembola, Cyclopoida, Gastropoda, Hydracarina,

Nemata, Oligochaeta, Oribatei, Ostracoda, Platyhelminthes, Prostigmata). After removing taxa occurring in less than 5 % of the samples (McCune and Grace 2002), 60 taxa remained as the focus of the data analysis.

The data were expressed as relative abundance, i.e. the sum of the abundance of all the taxa in a sample is equal to 1. Relative abundance was chosen to reduce noise from natural fluctuations in raw abundance values.

GIS data extraction

A 3-arcsec (approximately 90 m) continuous Shuttle Radar Topography Mission Digital Terrain Model (SRTM-DTM) was used for watershed delineation (Jarvis et al. 2006; and see <http://srtm.csi.cgiar.org>). The SRTM-DTM was processed at a 30-m resolution to remove all depressions through a combination of filling and breaching. The stream and lake network and associated metadata were created from the SRTM-DTM using the Burn function. This simple method allows the location of known mapped water features to be embedded into the SRTM-DTM using a method first introduced by Maidment (1996). Upstream catchments for each of the benthic macroinvertebrate sample sites were delineated using the Spatial Analyst Hydrology toolset in ArcGIS® version 9.2 (ESRI, St. Paul, MN, USA). A series of geospatial variables were then extracted for each of the delineated catchments (Table 2).

Training, validation and test dataset

The analysis in this paper uses 582 samples collected in Atlantic Canada. Samples were categorised according to the CABIN protocol (Reynoldson et al. 2007): (1) *reference*: no observed modifying influences within the vicinity of the reach at the time of sampling, this being confirmed by later, more detailed examination of surrounding land use; (2) *potential reference*: no observed modifying influences within the vicinity of the reach at the time of sampling or (3) *test*: one or more modifying influences present within the vicinity of the reach at the time of sampling (<http://cabin.cciw.ca>; Armanini et al. 2011). For the present paper, both ‘reference’ and ‘potential reference’ samples were merged into a single ‘reference sample’ category. Two subsets of reference samples were created

Table 2 List of the environmental variables considered in the present paper and related acronyms with details on the GIS data layers and sources

Group	Variables	Data layer	Data source
Stream morphology	Catchment area (km ²) Average slope (%)	Digital elevation model	NASA Shuttle Radar Topography Mission
Climate	Long-term average precipitation (mm) Long-term average temperature range (°C)	Climate (precipitation and temperature)	Environment Canada—Meteorological Service of Canada
Geology	Sedimentary and volcanic rocks (%) Intrusive rocks (%) Sedimentary rocks (%) Volcanic rocks (%)	Geological map of Canada, major rock categories	Geological Survey of Canada, Natural Resources Canada

following the approach outlined by Hargett et al. (2007): (a) a ‘training’ dataset used for model development and (b) a ‘validation’ dataset used to measure overall model performance (*sensu* Van Sickle et al. 2006). Thus, samples in the combined reference dataset were randomly assigned to two subsets: 75 % to the training dataset and 25 % to the validation set. This procedure produced 128 training samples and 42 validation samples, making a total of 170 samples, each collected at a unique location (Fig. 1).

A third dataset included sites potentially impacted accordingly to the CABIN protocol (Reynoldson et al. 2007) or provincial classification (Province of New Brunswick, personal communication): 412 test samples were collected from the four provinces studied. According to CABIN, sites are flagged as test sites for a range of reasons, and they could thus include both impacted and unimpacted sites.

A subset of potentially impacted sites with detailed physical and chemical variables from the Upper Mersey CABIN studies was selected to explore the ability of the predictive model to reflect potential departure of sites from the regional reference state. The Upper Mersey Study was originally designed as a comprehensive suite of reference and test sites to monitor aquatic health and assess the ecological effects of forestry management activities on benthic macroinvertebrate communities in the upper Mersey watershed (<http://cabin.cciw.ca>). Samples and the environmental variables were collected using the standardised CABIN protocol (Table 1; Reynoldson et al. 2007). The majority of samples collected on PEI and Newfoundland were categorized as test sites,

and we recommend that further data collection in the province should be oriented towards the collection of reference data, if possible.

Data analysis

Biological data interoperability comparison

A permutational ANOVA (PERMANOVA; Anderson 2001), computed in the *Vegan* package (Oksanen et al. 2009) using 999 permutations, was used to assess biological data compatibility attributable to data source. Based on the information summarized in Table 1, a three-factor PERMANOVA was performed considering Years (8 years, 2002–2009), Province (three provinces, NB, NL and NS) and Sampling Method (CABIN Kick-net Single 3-min sample, CABIN Kick-net Composite of 3×1-min samples and Provincial U-net). Differences between the two datasets collected in New Brunswick, described in the “Study area” section (b), have not been tested here as they have already been demonstrated as interoperable (Brua et al. 2010). Both the calibration and validation datasets were used in the PERMANOVA, representing a total of 170 reference samples.

Reference condition model development

We developed a predictive model based on the RIVPACS approach, based on procedures described in recent literature (Clarke et al. 1996, 2003; Moss et al. 1999; Hawkins et al. 2000; Van Sickle et al. 2006) and a paired null predictive model following the

approach of Van Sickle et al. (2006). To date, RCA models have been traditionally developed in Canada following the BEAST approach (Reynoldson et al. 1995). However, for this model we have decided to diverge from this approach and instead to develop a RIVPACS-type approach, for two reasons: (1) the RIVPACS approach is backed by a larger body of international scientific literature and its features have been widely discussed and studied in a variety of ecological settings (see citations at the beginning of the paragraph), and (2) RIVPACS-type models necessarily limit operator choices during model development, particularly in the biological clustering phase. This latter point supports our desire to reduce the level of subjectivity in model construction, which creates inconsistencies in model performance amongst studies and promotes model subjectivity.

Biological classification

To assess the biological similarity amongst samples of community assemblages (*sensu* Faith et al. 1987), the Bray–Curtis dissimilarity measure was selected as a robust indicator of differences amongst benthic macroinvertebrate samples (Reynoldson et al. 2001). The dissimilarity matrix was built using relative abundance data and considering only the calibration dataset. As recommended by Reynoldson et al. (2001), benthic macroinvertebrate sample data were not transformed, as dominance of taxa is an important property of biological communities.

Agglomerative hierarchical cluster analysis (Kaufman and Rousseeuw 1990) was applied using the *Agnes* function in the R package *Cluster* (Maechler et al. 2005) using an unweighted pair group method with arithmetic averages. The agglomeration coefficient was computed as it provides a measure of the average height of the mergers in a dendrogram. To select the number of clusters to be retained, an internal validation approach was selected (Handl et al. 2005). The development of the RIVPACS-style weighting approach limits the need to optimize the clustering phase and identify the ‘best’ model (Moss et al. 1999).

To highlight potential taxa indicators of the different clusters, the indicator value (IndVal) (Dufrene and Legendre 1997) method was run using the *duleg* function in R. Only calibration samples were included in the analysis. Such analysis should be seen as qualitative only, not for the purpose of describing the different

clusters, as consideration of indicator value significance would be inappropriate, due to circularity (i.e. the clusters were observed within the same set of biological observations).

Environmental drivers and development of discriminant models

For the development of a RIVPACS-based RCA model, the selection of one or more environmental variables that can correctly discriminate amongst the identified biological groups was performed using the best-subsets multiple discriminant function (DF) modelling procedure developed by Van Sickle et al. (2006). Group size was used as a prior probability in predicting group membership probabilities from the DF model (Clarke et al. 2003). The procedure by Van Sickle et al. (2006) uses Wilk’s lambda values to measure group separation and to rank the different models obtained by the subset procedures. Van Sickle et al. (2006) suggested that the root mean square error (RMSE) of O/E can be used to gain information on the bias and variability of prediction errors. Over-fitting was visually checked in addition to looking at the RMSE of O/E at increasing model orders (following Van Sickle et al. 2006).

Computation of O/E measure

Following the recommendations of Hawkins et al. (2000), Clarke & Murphy (2006) and Van Sickle et al. (2007), we have included only the relatively common taxa, to increase the power of detecting deviation from reference conditions using our model. Hence and to this end, we have selected a cutoff level of 0.5 (i.e. when probability of occurrence of a taxon exceeds 50 %) to include taxa in the predictive model, following Van Sickle et al. (2007). Taxa richness O/E was computed based on the algorithm of Van Sickle et al. (2006; 2007) for both DF-based and null predictive models.

Prediction of habitat group membership

Following selection of the best DF model, estimates of the probabilities that a new site belonged to each of the clusters were calculated. The probability of a new sample belonging to a given group was based on the

Mahalanobis distance of the sample site from the centre of each group (Clarke et al. 2003). The model was further assessed by evaluating if predictor variables at a site were within the range of predictor variables at reference sites (Moss et al. 1987; Clarke et al. 1996; Hawkins et al. 2000). To this end, a chi-squared test based on the multivariate distance between the set of predictor values observed in a test samples and those in reference samples was used to assess outliers. Sites were flagged as outliers where $\alpha < 0.01$, according to the procedure of Hawkins et al. (2000).

Null model development

Recently, benthic scientists have explored application of the concept of null distribution of biota in different fields of environmental sciences including bioassessment (Van Sickle et al. 2005), although the full implications of such a theoretical model shift have not been fully explored. Such ‘null models’ are formulated based on an assumption that the occurrence probabilities of taxa in reference sites are not driven by regionally discrete natural gradient variables (Van Sickle et al. 2005), and as a result, there is no support for the clustering of reference sites into groups. For example, Van Sickle and Hughes (2000) have highlighted the need to contrast the many predictive models developed in the Oregon (USA) for biomonitoring purposes against a simple assumption of null distribution to validate the need for clustering of the communities. Therefore, a null model has been derived for this analysis assuming that all reference samples belong to a single group.

Comparison of null model and predictive models

The O/E of taxa richness was used to compare the DF-based and null models using three approaches: (1) the mean of the O/E measure as, theoretically, the mean O/E values of a reference site should be as close to 1 as possible in order to avoid over- or underestimation of ecological quality at test sites; (2) the standard deviation of the O/E measure, as a measure of the ability of a model to explain the “natural” (non-anthropogenic) sources of assemblage variation amongst sites. Since the null model makes no attempt to account for such variation, it has the highest SD(O/E) and hence serves as a baseline for DF-based model performance; and (3) the ability of the O/E measure to reflect changes in

environmental stressors and to detect potential stress condition of the biological community. The modelling procedure used to develop the predictive model has been compiled as a set of function scripts for the R language (R Development Core Team 2009), partly based on scripts of Van Sickle et al. (2006), and is freely available from the senior author upon request.

Results

Biological data interoperability comparison

A three-factor PERMANOVA analysis was performed to compare taxa composition based on Years, Province and Sampling Method (Table 1). As expected, a portion of the variance observed was due to inter-annual variability (partial $R^2=0.13$, $P=0.001$). The inter-annual component interacted significantly both with Province and Sampling methods (partial $R^2=0.001$ and 0.002 , respectively), but the portion of variance explained for Province and Sampling methods was very small (partial $R^2 \sim 2\%$ for both variables). A significant effect with a limited variance explained was observed for both Sampling Methods (partial $R^2=6\%$, $P=0.001$) and Province (partial $R^2=3\%$, $P=0.001$). No significant interactions were observed between Province and Sampling Method. Overall, the low amount of variance explained by Sampling Method, Province and by the relative Year interactions indicates considerable overlap in taxa composition for the components comprising the separate factors. This confirmed data interoperability and justified merging of those datasets.

Biological community classification for the RIVPACS-based model

Bray–Curtis dissimilarities had a median of 0.59 (average value of 0.60) with an inter-quartile range between 0.49 and 0.71. Visual assessment of dendrogram and breaks in slope of the agglomeration function indicated the selection of three groups as an optimal clustering.

The majority of group 1 ($n=31$) sites were located in southern and central New Brunswick and central Nova Scotia. Sites were characterised by an intermediate to large watershed area with low slope and elevation relative to the other groups. Group 2 ($n=39$) sites were strongly affiliated with southern New

Brunswick, particularly Fundy National Park, and Cape Breton with a single site in Newfoundland. These sites were associated with intermediate to high slopes within the catchment and low to intermediate elevations relative to the other groups. Group 3 ($n=58$) sites were located across the study region with no clear geographical affiliation. Due to the large number of sites within group 3, there was a wide range in elevation with sites characterised by a small to intermediate catchment area relative to the other groups. Differences in stream-type distribution between test and reference sites were visually assessed, and the range of variation of variables, such as elevation and catchment size, were fully overlapping across the three groups reducing the risk of stream-type differences between test and reference sites.

From a biological standpoint, IndVal analysis identified several indicator taxa for each group (Table 3). Most notably, group 1 was characterized by Chironomidae (Diptera) and Naididae (Haplotaxida); group 2 by Baetidae (Ephemeroptera) and Ephemerellidae (Ephemeroptera), Rhyacophilidae (Trichoptera), Chloroperlidae (Plecoptera) and Perlodidae (Plecoptera) and group 3 by Hydropsychidae (Trichoptera) and Brachycentridae (Trichoptera), Oligochaeta and Platyhelminthes. The biological community structure appears to reflect changes in stream habitat with Ephemeroptera, Plecoptera and Trichoptera taxa, known for their preference for well-oxygenated water, characterizing the cluster with higher slope and thus more turbulent flows (i.e. group 2), whilst more potamal taxa dominated the sites with low slope and elevation (i.e. group 1). The group 3 was characterized by generalist taxa able to colonize the range of environmental conditions experienced in the sites of group 3.

Identification of environmental drivers and development of a DF models using best-subset procedure

More than 250 possible configurations were computed using the selected predictors in the screening or best-subset procedure, and five best models were retained for each model order (defined as the number of predictors included in the DF model) for an overall output of 36 models. For the training sites, the overall precision and accuracy increased with the order of the best DF models, i.e. the RMSE (O/E), decreased (Fig. 2). However, models with an order greater than 4 did not

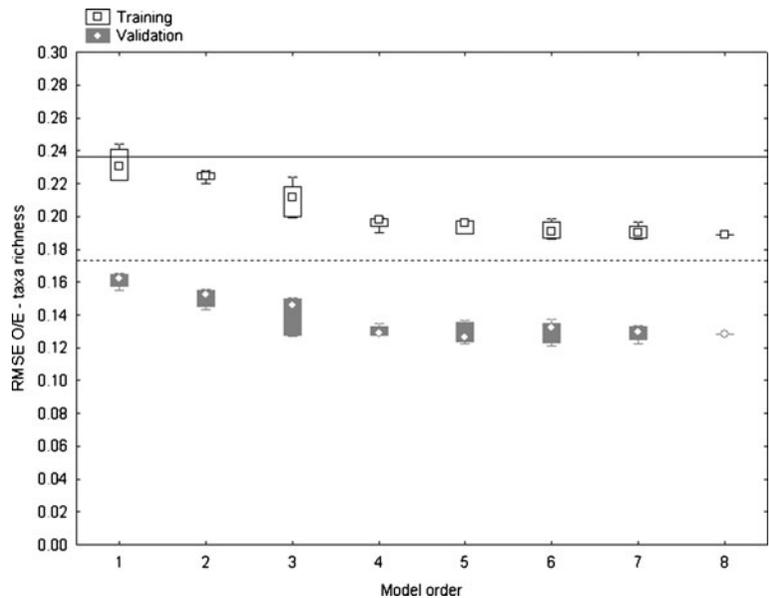
Table 3 Indicator value (IndVal) for the three biological groups selected for the development of the RIVPACS-based model

Class	Taxa	IndVal (%)	
1	Chironomidae	62	
	Naididae	52	
	Oribatei	49	
	Polycentropodidae	49	
	Hydroptilidae	43	
	Empididae	42	
	Corydalidae	38	
	Enchytraeidae	36	
	Leuctridae	35	
	2	Baetidae	89
Rhyacophilidae		69	
Chloroperlidae		57	
Perlodidae		57	
Ephemerellidae		51	
Heptageniidae		46	
Lepidostomatidae		44	
3		Hydropsychidae	66
		Brachycentridae	60
		Oligochaeta	56
	Platyhelminthes	50	
	Elmidae	49	
	Perlidae	47	
	Philopotamidae	46	
	Tipulidae	46	
	Leptophlebiidae	42	
	Hydracarina	42	
Gastropoda	38		
Glossosomatidae	37		

Only values higher than 33 % are presented

appear to provide an improvement in both precision and accuracy. Therefore, a fourth-order model was selected, and by examining the best model in terms of RMSE(O/E) for each index, two DF solutions were presented as best possible models: (1) long-term annual temperature range (in degree Celsius), intrusive rocks (in percent), sedimentary rocks (in percent) and sedimentary and volcanic rocks (in percent) and (2) long-term annual temperature range (in degree Celsius), sedimentary rocks (in percent), sedimentary and volcanic rocks, and average slope (in percent). For the training dataset, the two models showed equal Wilks' lambda values (0.426), but the first model showed higher cross-validated classification accuracy

Fig. 2 Root mean-square errors of O/E taxa richness from predictive models are based on 36 best discriminant function models from training (*squares*) and validation (*circles*). The RMSE of the null model is depicted by a *solid line* for the training dataset and a *dashed line* for the validation dataset. Median, 25–75 %, and minimum-maximum range are represented by *square*, *box* and *whisker*, respectively



(73 as compared with 68 %). The first model was thus selected as best model. The DF model was then used to classify the entire dataset available (training, validation and test dataset).

Only one sample out of 170 reference samples and only one sample of the 412 test sites were flagged as outliers ($\alpha < 0.05$) based on the chi-squared test. A comparison of the RMSE of O/E measures between DF-based predictive and null models in training and validation datasets showed a slight improvement in accuracy for the DF-based predictive models (Table 4). Similar results were obtained for the two models when looking at standard deviation of the O/E ratio (Table 4). Box plots of O/E values for reference and selected test datasets, namely the Upper Mersey and PEI National Park dataset as far as the test dataset are concerned, indicated no overlap between the datasets (Fig. 3). This result was evident for both null and DF models. In the PEI National Park dataset, the two models showed comparable results, whilst in the Upper Mersey the null model showed a higher deviance from the reference sites in the null model. The O/E measures both for the DF-based and the null-based models responded to changes in the availability of dissolved oxygen (Table 5), whilst no response was observed for both total Nitrogen and Phosphorous concentrations. Nevertheless, the number of cases available was limited to 17, due to lack of environmental data for site interpretation.

Discussion

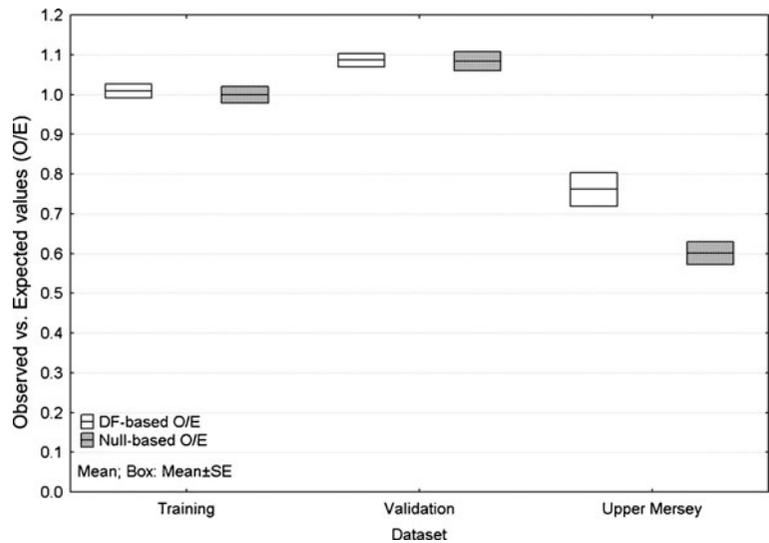
Models used to support sound and sustainable environmental management require the availability of consistent data with spatiotemporal coverage appropriate for the management questions being posed. Efforts have been made to establish large-scale and long-term hydrological, physicochemical and biological monitoring programs to collect and store information about the status of the environment. However, these programs are difficult to sustain over large land areas such as Canada, as they require careful spatiotemporal coordination. In reality, annual data are often collected sequentially at non-matching stations and/or at different times of year (Armanini et al. 2011). In such situations, maximizing inter-operability of the available data is key to reducing the need for further data

Table 4 Comparison of the distribution of selected biological metrics O/E ratio in the training and validation datasets (see method for details) for the DF-based and the null-based models

Taxa richness	Training		Validation			
	Mean	SD	RMSE	Mean	SD	RMSE
DF-based model	1.009	0.191	0.190	1.087	0.105	0.135
Null-based model	1.000	0.237	0.236	1.084	0.152	0.172

SD standard deviation of the O/E metric, RMSE root mean square error of the O/E metric

Fig. 3 Box plots of O/E values for ‘reference’, ‘test’ and Upper Mersey datasets for DF-based (a) and null (b) predictive models. Mean and plus/minus standard error (\pm SE) are represented by line and box, respectively



collection and makes the best use of resources already invested in a monitoring network. Standard models that can support different data sources and rely on easily accessible environmental information are needed to build large-scale models that support simple and rigorous adaptability and replicability. However, potential issues due to differences in sampling methods, seasonal differences and other sources of variability highlight the need to assess dataset properties before merging datasets. Here, such differences have been minimised, with no relevant differences observed, allowing the merging of different data sources, broadening the applicability of the derived RCA model.

Assuring that bioassessment models incorporate all sources of natural variability and are capable of highlighting anthropogenic change are key goals of bioassessment. However, financial and resource constraints pose a significant challenge for the full implementation of bioassessment programs.

Inter-annual variability is the most commonly neglected feature of assessment systems with a few

notable exceptions (e.g. Humphrey et al. 2000). Unfortunately, in Atlantic Canada, a multi-year dataset for all reference sites is not available to account for such variability at the site level. Due to the lack of available data, inter-annual variability was assessed by incorporating reference samples collected across multiple years within the model. Future versions of the model should focus on more fully exploring variation at the individual site level where data can be obtained. Moreover, changes to stream ecosystems arising from anthropogenic climate change simply underline the importance of establishing systematic long-term monitoring across a subset of reference sites—a practice never fully adopted within the CABIN program to date.

We have successfully created a RCA model based solely on GIS information with an accuracy level comparable to other approaches undertaken using locally derived field variables (e.g. Hawkins et al. 2000). Using GIS data can greatly reduce the cost of developing RCA models (Hargett et al. 2007) as it allows

Table 5 Coefficients of determination r^2 (reported in italic if p value is above 0.05) for selected environmental variables and O/E measures in the Upper Mersey datasets (see “Data and methods” for details) for the DF-based and the null-based models

Variable	Mean	SD	<i>N</i>	DF-based O/E		Null-based O/E	
				r^2	<i>p</i> value	r^2	<i>p</i> value
Dissolved oxygen (mg/L)	6.80	1.73	17	0.237	0.047	0.313	0.020
Nitrogen—total (mg/L)	0.20	0.07	17	0.061	0.338	0.199	0.072
Phosphorus—total (mg/L)	0.03	0.02	17	0.001	0.916	0.124	0.165

SD standard deviation of the environmental variable, *N* number of valid cases

the re-use of existing data and promotes data interoperability amongst scarce biological studies including those with missing or divergent field information. In this study, GIS-derived variables which served as the main predictors of community clustering were related to geology and climate. Geology (e.g. rock type) and climate (e.g. temperature range) are major influences of river habitat structure and also of benthic community composition (Omernik 1987; Snelder et al. 2004). Moreover, such GIS-based broad scale variables seem to be less influenced by human activities than field collected information, e.g. water chemistry or channel substrate (Poquet et al. 2009). However, caution should be exercised when using geospatial data, since metadata for each layer may be inadequate to support their use, as without ground-truthing and data quality assessment, data merging can be problematic. For this reason, we would urge a wider adoption of metadata standards for each GIS layer to support replicability and quality assurance to facilitate their widespread use for model development.

Moreover, the wide range of applicability of the model was confirmed by the limited number of reference and test sites flagged as outliers (chi-squared test). Only one out of more than 400 test samples not used in model development was found to be out of the range of applicability of the model in Atlantic Canada. The Upper Mersey test dataset, selected due to the presence of catchment disturbance, showed divergent O/E values with respect to the calibration and validation reference dataset, confirming the ability of the model to detect impairment signals in the system. Moreover, the O/E metrics responded positively to changes in the availability of Oxygen in a dataset subject to silviculture practices. Nevertheless, no response to concentration of nitrogen (N) and phosphorous (P) was evident and this might be linked to the low average concentrations observed at the test sites for both parameters. According to the values suggested by Chambers et al. (2012), both average values were well below the suggested N and P thresholds for preserving high water quality.

From a standpoint of parsimony, the performance of complex RCA models should be demonstrably superior to a null model (Van Sickle et al. 2006; Carlisle et al. 2008), as there is the need to demonstrate that the DF model has succeeded in explaining observed patterns in assemblage variation across the region of applicability (Van Sickle et al. 2005). In the

data analysed here, an improvement in precision was observed when using a RCA model as compared with a null model. Standard deviation of O/E measures showed improved values both in the training and validation datasets when compared to the null models. The DF-based predictive model from this analysis outperformed the null model overall. However, the performance of the null model is still noteworthy. If adequate (null) performance can be similarly demonstrated elsewhere, it could have significant implications for the development of a cost-effective, nationwide assessment system, an important consideration in a geographically diverse country such as Canada, with a significant portion of remote river habitat.

Innovation in the field of bioassessment can be achieved by combining multivariate prediction and multi-metric methods, i.e. computing biological based diagnostic metrics with an O/E approach to enhance the diagnostic power of reference condition models. With the increasing availability of stressor-specific metrics (e.g. Von der Ohe and Liess 2004; Armanini et al. 2011), managers increasingly have an option to employ tools that combine the strong predictive character of a multivariate prediction system with the diagnostic potential of biological metrics. Two examples where such an approach has experienced some limited success are the BMWP metric in the UK (Clarke et al. 1996) and the SIGNAL approach in Australia (Davies 2000). Nevertheless, to integrate biological metrics into RCA models, additional research is required both for model development (e.g. Van Sickle et al. 2006) and for O/E interpretation.

In conclusion, we have demonstrated an approach to regional ecological bioassessment model development based on the predictive model and reference condition approaches. Future efforts should focus on the application of similar RCA models at regional scale to other remote areas such as northern Canada, with a view to implementing innovative biological diagnostics using O/E measures for improved management of freshwater ecosystems, including the incorporation of the temporal component of natural variability.

Acknowledgments We would like to acknowledge partners who have made this paper possible by sharing their data: (1) New Brunswick Department of Environment (New Brunswick DENV); (2) Department of National Defence Canada, Base Gagetown; (3) Environment Canada's National Agri-Environmental Standards Initiative (NAESI); (4) Canadian Rivers Institute (CRI); (5) the Canadian Aquatic Biomonitoring

Network (CABIN) network and (6) Parks Canada. We would also like to thank Dr. John Van Sickle (USEPA) for generously sharing R scripts and for reviewing an earlier version of the manuscript and Pierre Martel (Parks Canada) for providing access to the Upper Mersey dataset. This research was supported through program funding from Environment Canada and through a Canadian National Science and Engineering Research Council Discovery Grant to DJB.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Anderson, M. J. (2001). A new method for a non-parametric multivariate analysis of variance. *Austral Ecology*, *26*, 32–46.
- Armanini, D. G., Horrigan, N., Monk, W. A., Peters, D. L., & Baird, D. J. (2011). Development of a benthic macroinvertebrate flow sensitivity index for Canadian rivers. *River Research and Application*, *27*, 723–737. doi:10.1002/rra.1389.
- Barbour, M. T., Gerritsen, J., Snyder, G. E., & Stribling, J. B. (1999). *Rapid bioassessment protocols for use in wadable streams and rivers: periphyton, benthic macroinvertebrates and fish*. EPA 841-b-99-002 (2nd ed.). Washington, D.C.: USEPA, Office of Water. 153 pp.
- Benke, A. C., & Cushing, C. E. (Eds.). (2005). *Rivers of North America*. Amsterdam: Elsevier. 1144 pp.
- Bonada, N., Prat, N., Resh, V. H., & Statzner, B. (2006). Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. *Annual Review of Entomology*, *51*, 495–523.
- Brua, R. B., Culp, J. M., & Benoy, G. A. (2010). Comparison of benthic macroinvertebrate communities by two methods: kick- and U-net sampling. *Hydrobiologia*, *658*(1), 293–302. doi:10.1007/s10750-010-0499-x.
- Buffagni, A., Armanini, D. G., & Erba, S. (2009). Does the lentic–lotic character of rivers affect invertebrate metrics used in the assessment of ecological quality? *Journal of Limnology*, *68*(1), 92–105.
- Carlisle, D. M., Hawkins, C. P., Meador, M. R., Potapova, M., & Falcone, J. (2008). Biological assessments of Appalachian streams derived from predictive models for fish, macroinvertebrate, and diatom assemblages. *Journal of the North American Benthological Society*, *27*, 16–37.
- Chambers, P. A., McGoldrick, D. J., Brua, R. B., Vis, C., Culp, J. M., & Benoy, G. A. (2012). Development of environmental thresholds for nitrogen and phosphorus in streams. *Journal of Environmental Quality*, *41*(1), 7–20.
- Clarke, R. T., & Murphy, J. F. (2006). Effects of locally rare taxa on the precision and sensitivity of RIVPACS bioassessment of freshwaters. *Freshwater Biology*, *51*(10), 1924–1940.
- Clarke, R. T., Furse, M. T., Wright, J. F., & Moss, D. (1996). Derivation of a biological quality index for river sites: comparison of the observed with the expected fauna. *Journal of Applied Statistics*, *23*, 311–332.
- Clarke, R. T., Wright, J. F., & Furse, M. T. (2003). RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling*, *160*, 219–233.
- Davies, P. E. (2000). Development of a national river bioassessment system (AusRivAS) in Australia. In J. F. Wright, D. W. Sutcliffe, & M. T. Furse (Eds.), *Assessing the biological quality of freshwaters: RIVPACS and other techniques* (pp. 113–124). Ambleside: Freshwater Biological Association.
- Dufrene, M., & Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs*, *67*, 345–366.
- Faith, D. P., Minchin, P. R., & Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, *69*, 56–68.
- Gauthier, Poulin, Theriault, Ltd. (1977). *A biophysical classification of Terra Nova National Park. A report for Parks Canada, Atlantic Region*. Halifax: Historic Properties.
- Handl, J., Knowles, K., & Kell, D. (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics*, *21*(15), 3201–3212.
- Hargett, E., ZumBerge, J., Hawkins, C. P., & Olson, J. (2007). Development of a RIVPACS-type predictive model for bioassessment of wadeable streams in Wyoming. *Ecological Indicators*, *7*, 807–826.
- Hawkins, C. P., Norris, R. H., Hogue, J. N., & Feminella, J. W. (2000). Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications*, *10*, 1456–1477.
- Hering, D., Moog, O., Sandin, L., & Verdonschot, P. F. M. (2004). Overview and application of the AQEM assessment system. *Hydrobiologia*, *516*, 1–20.
- Humphrey, C. L., Storey, A. W., & Thurtell, L. (2000). AUSRI-VAS: operator sample processing errors and temporal variability—implications for model sensitivity. In J. F. Wright, D. W. Sutcliffe, & M. T. Furse (Eds.), *Assessing the biological quality of fresh waters. RIVPACS and other techniques* (pp. 143–163). Ambleside: Freshwater Biological Association.
- Jarvis, A., Reuter, H. I., Nelson, A., Guevara, E. (2006). Hole-filled SRTM for the globe, Version 3, available from the CGIAR-CSI SRTM 90m Database: <http://srtm.csi.cgiar.org>.
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: an introduction to cluster analysis*. New York: Wiley.
- Maechler, M., Rousseeuw, P., Struyf, A., & Hubert, M. (2005). Cluster analysis basics and extensions. <http://cran.r-project.org/web/packages/cluster/>.
- Maidment, D. R. (1996). GIS and hydrologic modelling—an assessment of progress. In: International NCGIA Conference, 3/Workshop on Integrating GIS and Environmental Modelling 1996, Santa Fé, New México. Proceedings (pp. 21–25). Santa Fé: NCGIA.
- McCune, B., & Grace, J. B. (2002). *Analysis of ecological communities*. Glendon Beach: MjM Software. 304 pp.
- Meades, W. J., & Moores, L. (1994). *Forest site classification manual: a field guide to the forest types of Newfoundland*. Forestry Canada and Newfoundland Dept of Forestry and Agriculture, FRDA Rep. No. 003. St. John's: FRDA.

- Monk, W. A., & Curry, A. (2007). *Annual report on developing an interpretive model for water classification of flowing waters in New Brunswick: preliminary report 2006–2007*. NB Environmental Trust Fund, 63pp.
- Moss, D., Furse, M. T., Wright, J. F., & Armitage, P. D. (1987). The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. *Freshwater Biology*, *17*, 41–52.
- Moss D., Wright J. F., Furse M. T., & Clarke R. T. (1999). A comparison of alternative techniques for prediction of the fauna of running-water sites in Great Britain. *Freshwater Biology*, *41*, 167–181.
- Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2009). Vegan: Community Ecology Package. R package version 1.15-3. <http://CRAN.R-project.org/package=vegan>.
- Omernik, J. M. (1987). Ecoregions of the conterminous United States. *AAG/A*, *771*, 118–125.
- Poquet, J. M., Alba-Tercedor, J., Puntí, T., Sánchez-Montoya, M. D., Robles, S., Álvarez, M., et al. (2009). The MEDiterranean Prediction And Classification System MED-PACS: an implementation of the RIVPACS/AUSRIVAS predictive approach for assessing Mediterranean aquatic macroinvertebrate communities. *Hydrobiologia*, *623*, 153–171.
- R Development Core Team, 2009. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Resh, V. H. (2008). Which group is best? Attributes of different biological assemblages used in freshwater biomonitoring programs. *Environmental Monitoring and Assessment*, *138*, 131–138.
- Reynoldson, T. B., Bailey, R. C., Day, K. E., & Norris, R. H. (1995). Biological guidelines for freshwater sediment based on Benthic Assessment of Sediment the BEAST using a multivariate approach for predicting biological state. *Australian Journal of Ecology*, *20*, 198–219.
- Reynoldson, T. B., Rosenberg, D. M., & Resh, V. H. (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., Logan, C., Pascoe, T., & Thompson, S.P. (2007). CABIN Canadian Aquatic Biomonitoring Network invertebrate biomonitoring field and laboratory manual. Canada National Water Research Institute, Environment Canada Retrieved December 18 2009, from http://cabin.cciw.ca/Application/Downloads/cabin_protocol.doc.
- Snelder, T. H., Cattaneo, F., Suren, A. M., & Biggs, B. J. F. (2004). Is the river environment classification an improved landscape-scale classification of rivers? *Journal of the North American Benthological Society*, *23*, 580–598.
- Van Sickle, J., & Hughes, R. M. (2000). Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. *Journal of the North American Benthological Society*, *193*, 370–384.
- Van Sickle, J., Hawkins, C. P., Larsen, D. P., & Herlihy, A. T. (2005). A null model for the expected macroinvertebrate assemblage in streams. *Journal of the North American Benthological Society*, *24*, 178–191.
- Van Sickle, J., Huff, D. D., & Hawkins, C. P. (2006). Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. *Freshwater Biology*, *51*, 359–372.
- Van Sickle, J., Larsen, D. P., & Hawkins, C. P. (2007). Exclusion of rare taxa affects performance of the O/E index in bioassessments. *Journal of the North American Benthological Society*, *26*, 319–331.
- Von der Ohe, P., & Liess, M. (2004). Relative sensitivity distribution of aquatic invertebrates to organic and metal compounds. *Environmental Toxicology and Chemistry*, *23*, 150–156.