# Statistical tools for evaluating habitat use in ecological studies: two applications

Marco A. Rodríguez
Université du Québec à Trois-Rivières
marco.rodriguez@uqtr.ca

# Predicting microhabitat selection in juvenile Atlantic salmon by the use of logistic regression and classification trees

Katrine Turgeon

# Overview

- Habitat models traditionally consider only behaviour of active fish
- Daytime sheltering in summer can be a key factor affecting production in juvenile Atlantic salmon
- Models including both activity and sheltering behaviours may provide:
  - better understanding of stream salmonid biology
  - more accurate predictions of spatial distribution

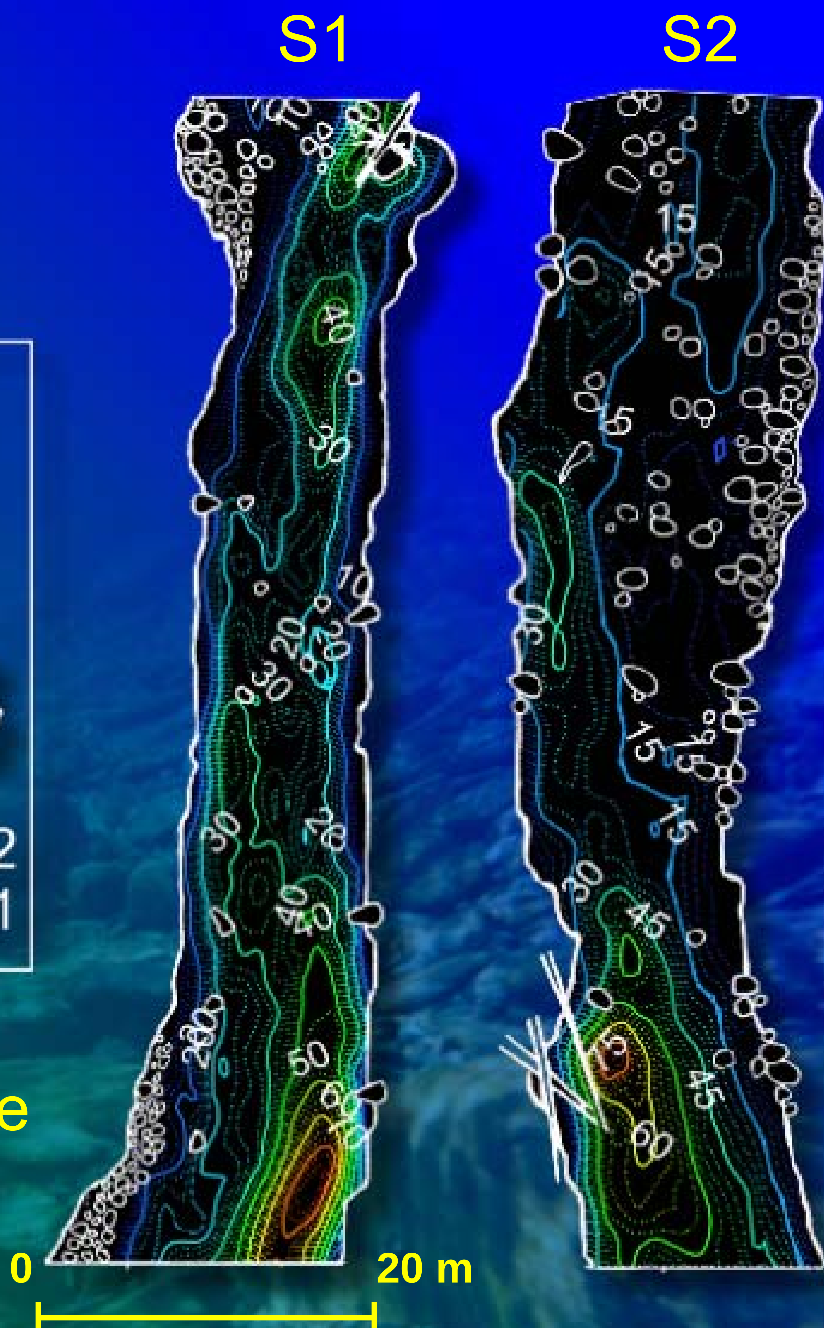
- An ideal habitat model should be:
  - Accurate and general
  - Parsimonious (ease of application and interpretation; reasonable demands in data collection and computation)
- The modelling process is incomplete without validation and assessment of model performance
- We compared the ability of logistic regression (LR) and classification tree (CT) models to predict habitat use in Atlantic salmon

# Methods

## Study Area


Québec

Big Jonathan Brook drainage



- Study period: 28 June - 29 August 2002
- Snorkelling observations of fish at focal (“presence”) points
- Unoccupied (“absence”) points selected at random
- Environmental measurements at each point:
  - Water depth
  - Water velocity (at 15% and 40% depth from bottom)
  - Substratum size
  - Instream and overhead cover
  - Distance to river bank

## Logistic regression:

logit(probability of occurrence) modelled as a linear combination of habitat predictors

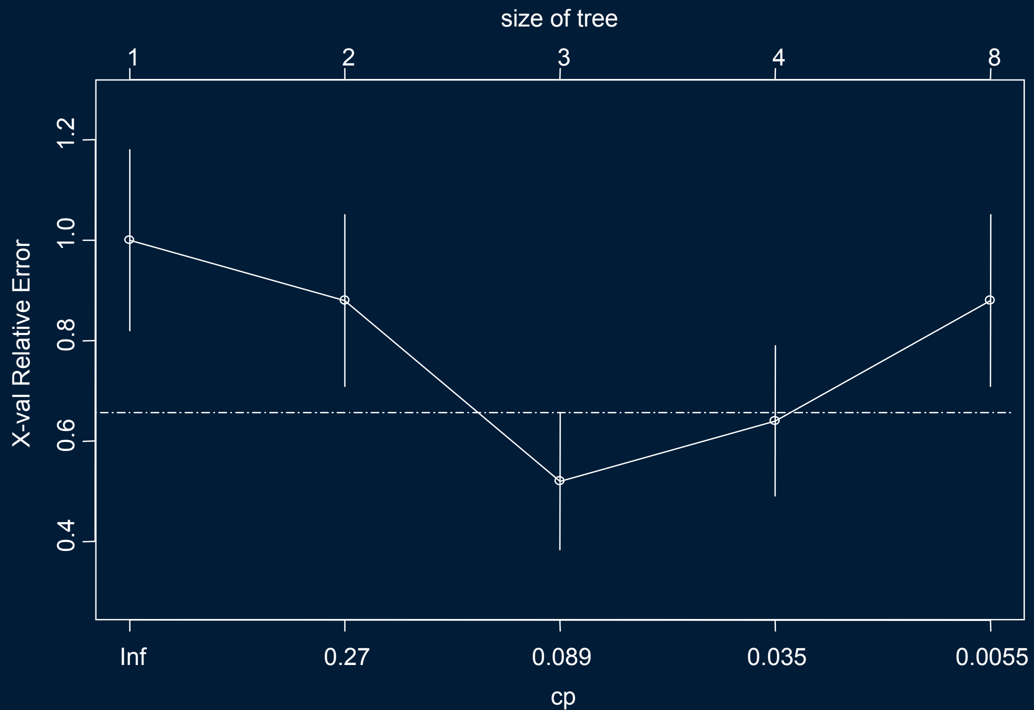
$$p = \frac{e^{\beta_0 + \sum_{i=1}^{k} \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^{k} \beta_i x_i}}$$

Binary predictions based on optimal selection threshold (ODT), or on midpoint theshold with $p = 0.5$

# Classification trees:

Optimal splitting values for habitat predictors are chosen by recursive partitioning to allocate cases to relatively homogenous groups (Gini coefficient)

Pruning followed by 10-fold cross-validation



## Model development

- Models were calibrated separately for each section and behaviour

## Model validation and evaluation

- Crossover field tests were used to validate models and assess transferability
- Model performance was assessed in terms of:
  - Correct classification rate, sensitivity, and specificity
  - Chance-adjusted measures: Cohen’s Kappa, Matthew’s correlation, normalized mutual information, and log odds-ratio

## Confusion matrix

| | | Predicted | |
|---|---|---|---|
| | | Absence | Presence |
| Observed | Absence | *a* | *b* |
| | Presence | *c* | *d* |

Correct classification rate (CCR): proportion of all cases correctly predicted
Specificity: proportion of true absences correctly predicted
Sensitivity: proportion of true presences correctly predicted

Cohen's kappa (κ, proportion of specific agreement; range: -1 to 1), Matthews correlation (MC; range: -1 to 1), normalized mutual information (NMI; range: 0 to 1), and log odds-ratio (LOR; range: -∞ to ∞):

$$\kappa = \frac{(a+d)-((a+c)(a+b)+(b+d)(c+d))/N}{N-((a+c)(a+b)+(b+d)(c+d))/N}$$

$$MC = \frac{ad-cb}{\sqrt{(a+c)(a+b)(b+d)(c+d)}}$$

,

$$NMI = 1-\frac{-a\ln(a)-b\ln(b)-c\ln(c)-d\ln(d)+(a+b)\ln(a+b)+(c+d)\ln(c+d)}{N\ln N-(a+c)\ln(a+c)-(b+d)\ln(b+d)}$$

$$LOR = \ln\left(\frac{ad}{cb}\right)$$

*a* = true presences, *b* = false presences, *c* = false absences, and *d* = true absences;

*N* = *a* + *b* + *c* + *d* = total number of cases

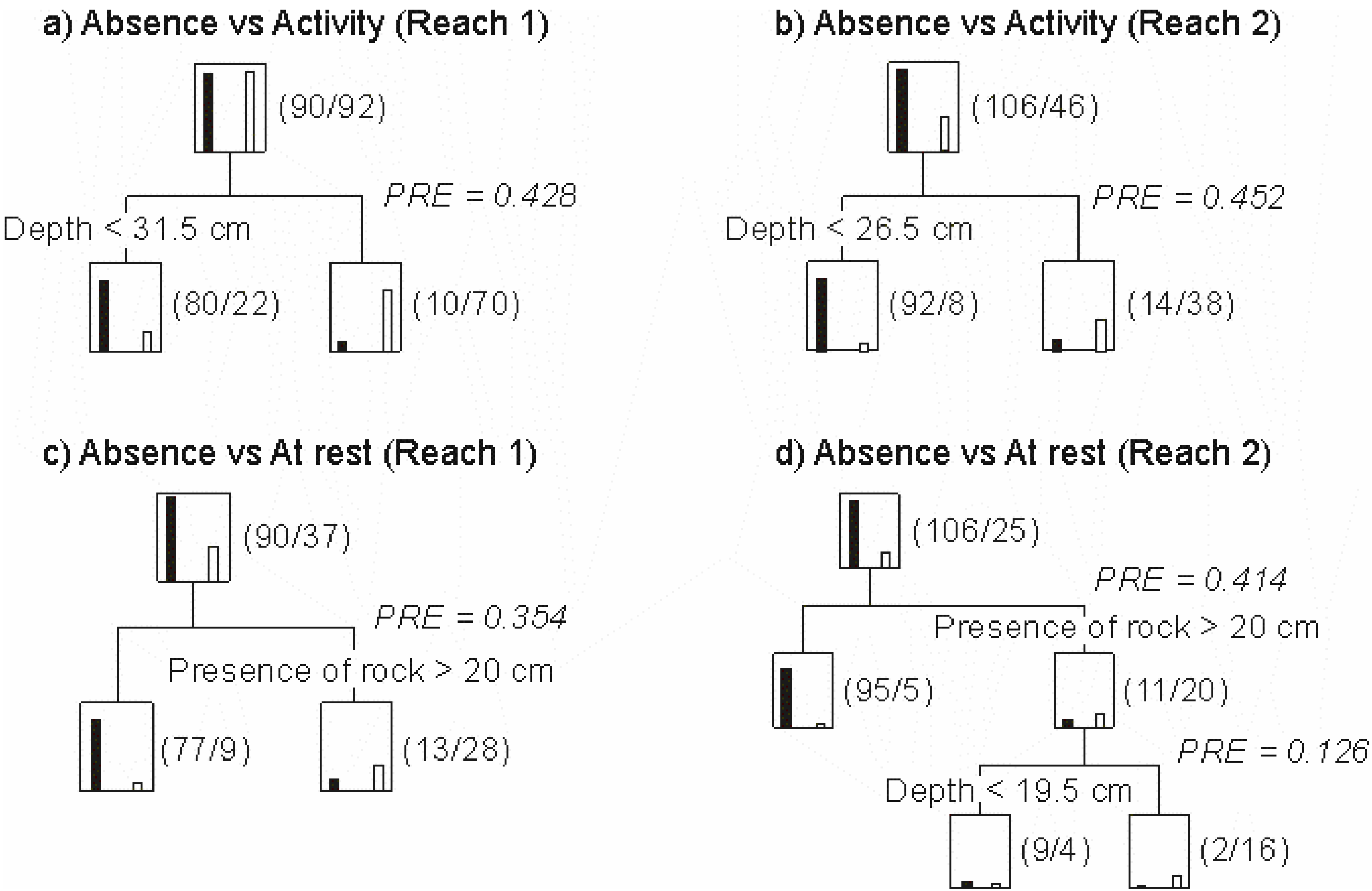
## Prediction Maps

- Instream habitat features were quantified at fixed points on uniform XY grids (1 x 1 m cells)
- LR and CT models were used to predict the spatial distribution of fish based on instream characteristics
- Prediction maps were then compared with observed fish distributions

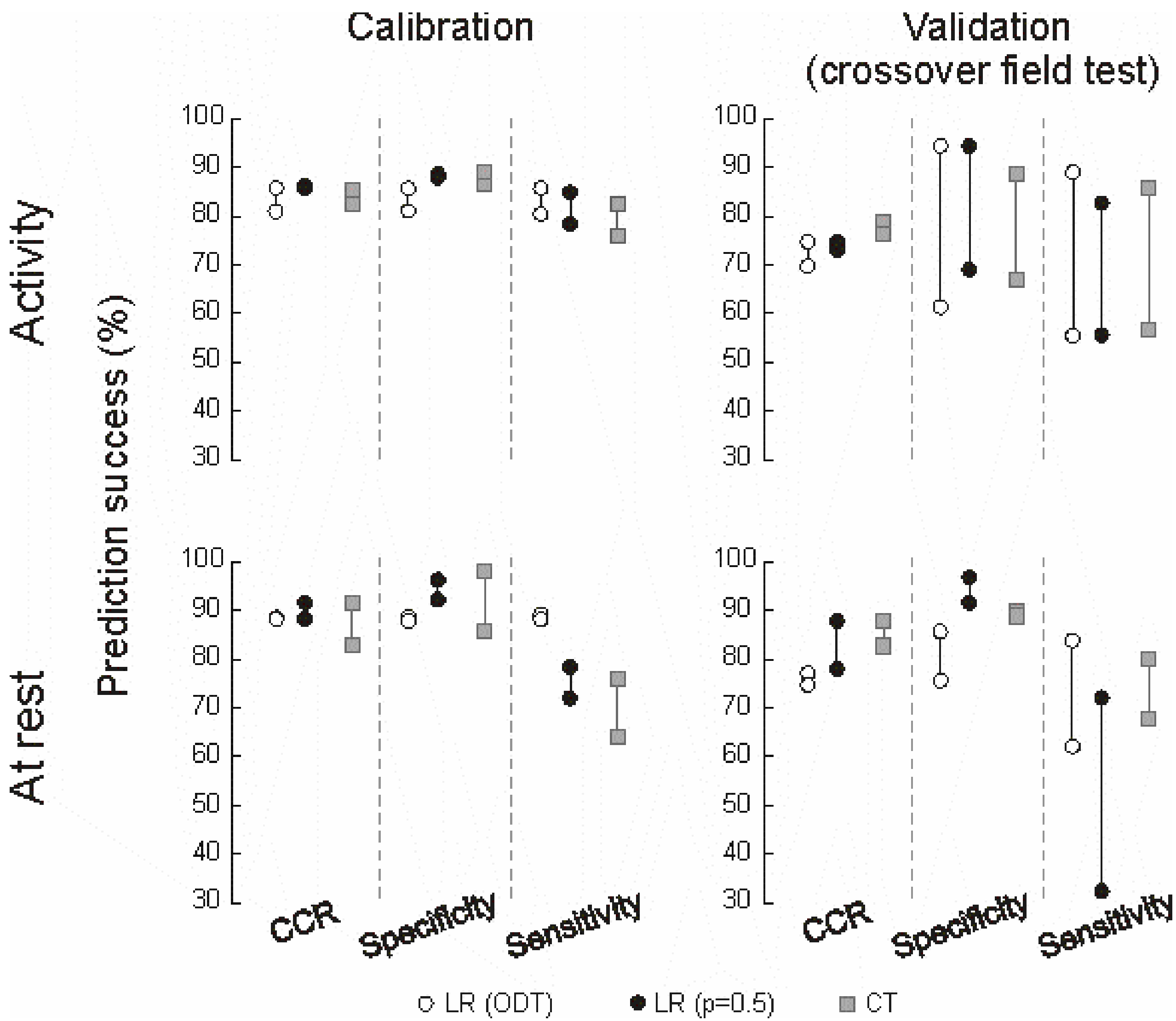
# Results

Coefficients of logistic regression models for activity and resting behaviours, by reach. Coefficients are given only for terms retained by the stepwise selection procedure ($p < 0.05$). All models were globally significant at $p < 0.0001$. McFadden's $\rho^2$ is reported for each model also

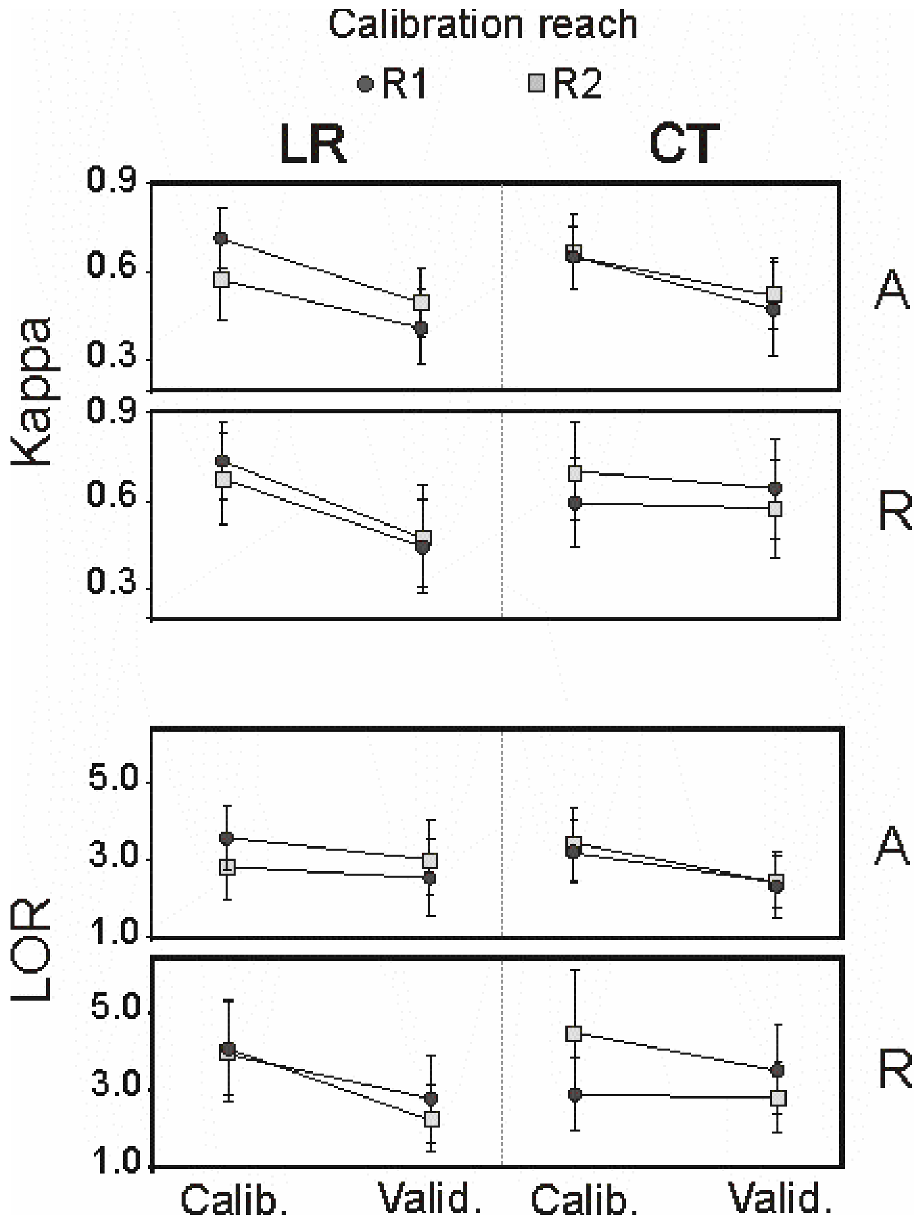
| Model term | Activity | | At rest | |
|---|---|---|---|---|
| | Reach 1 (N=182) | Reach 2 (N=152) | Reach 1 (N=127) | Reach 2 (N=131) |
| Constant | 0.807 | -0.673 | -1.656 | -1.701 |
| Depth | 3.544 | 3.959 | 0.638 | - |
| Velocity at 40% | -0.075 | - | - | 1.155 |
| Distance to bank | 0.854 | - | 0.673 | - |
| Substratum size | - | - | 0.622 | 1.008 |
| Rock > 20 cm | 0.610 | - | 1.816 | 1.721 |
| $Depth^2$ | - | -1.756 | - | - |
| $(Velocity\ at\ 40\%)^2$ | -0.497 | - | - | -0.862 |
| Substratum • Depth | - | - | -0.722 | - |
| McFadden's $\rho^2$ | 0.52 | 0.45 | 0.47 | 0.50 |


a) Absence vs Activity (Reach 1)
(90/92)
PRE = 0.428
Depth < 31.5 cm
(80/22)
(10/70)
b) Absence vs Activity (Reach 2)
(106/46)
PRE = 0.452
Depth < 26.5 cm
(92/8)
(14/38)
c) Absence vs At rest (Reach 1)
(90/37)
PRE = 0.354
Presence of rock > 20 cm
(77/9)
(13/28)
d) Absence vs At rest (Reach 2)
(106/25)
PRE = 0.414
Presence of rock > 20 cm
(95/5)
(11/20)
PRE = 0.126
Depth < 19.5 cm
(9/4)
(2/16)

# Model performance: Correct classification rate, specificity, and sensitivity

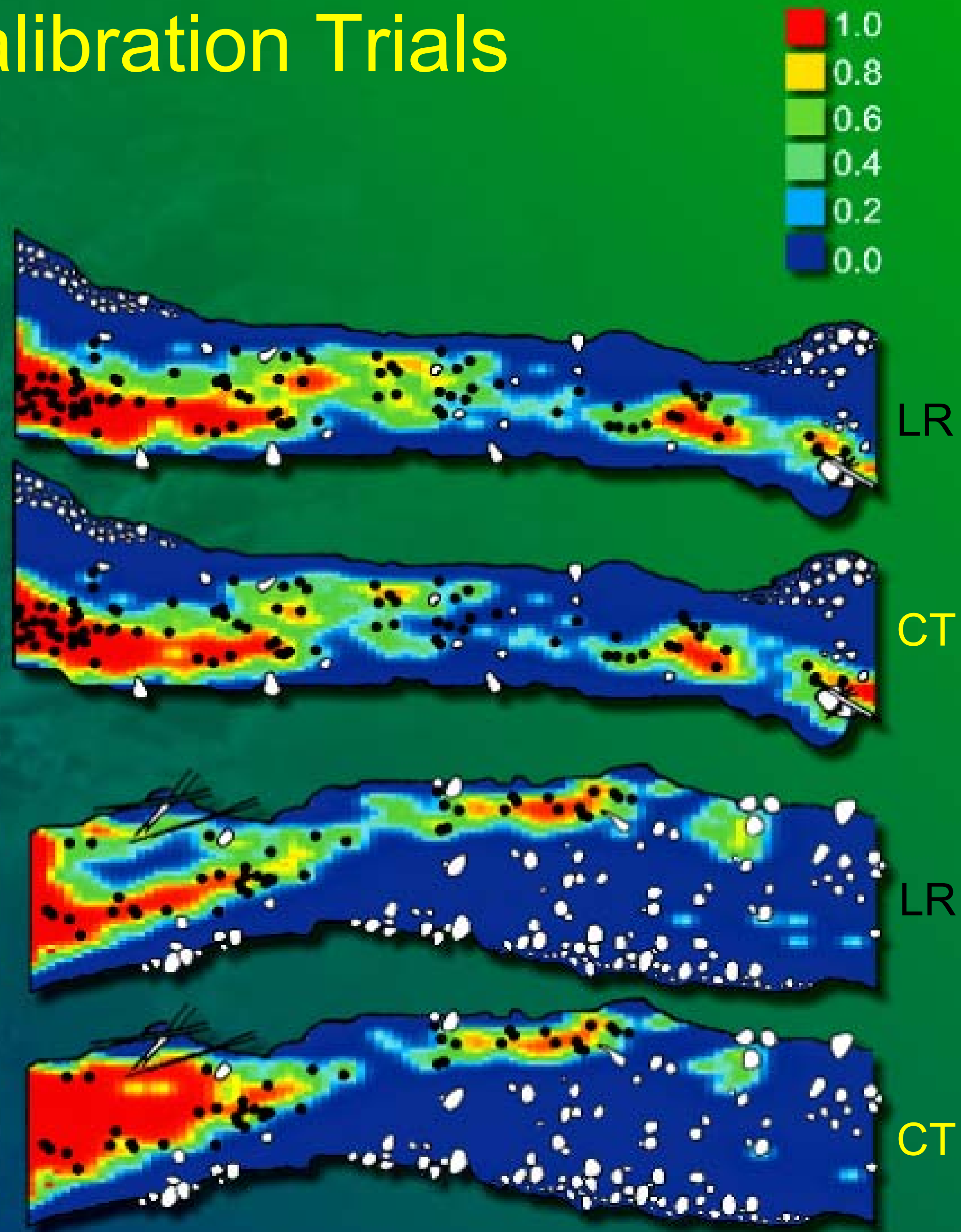


# Model performance: Chance-corrected measures



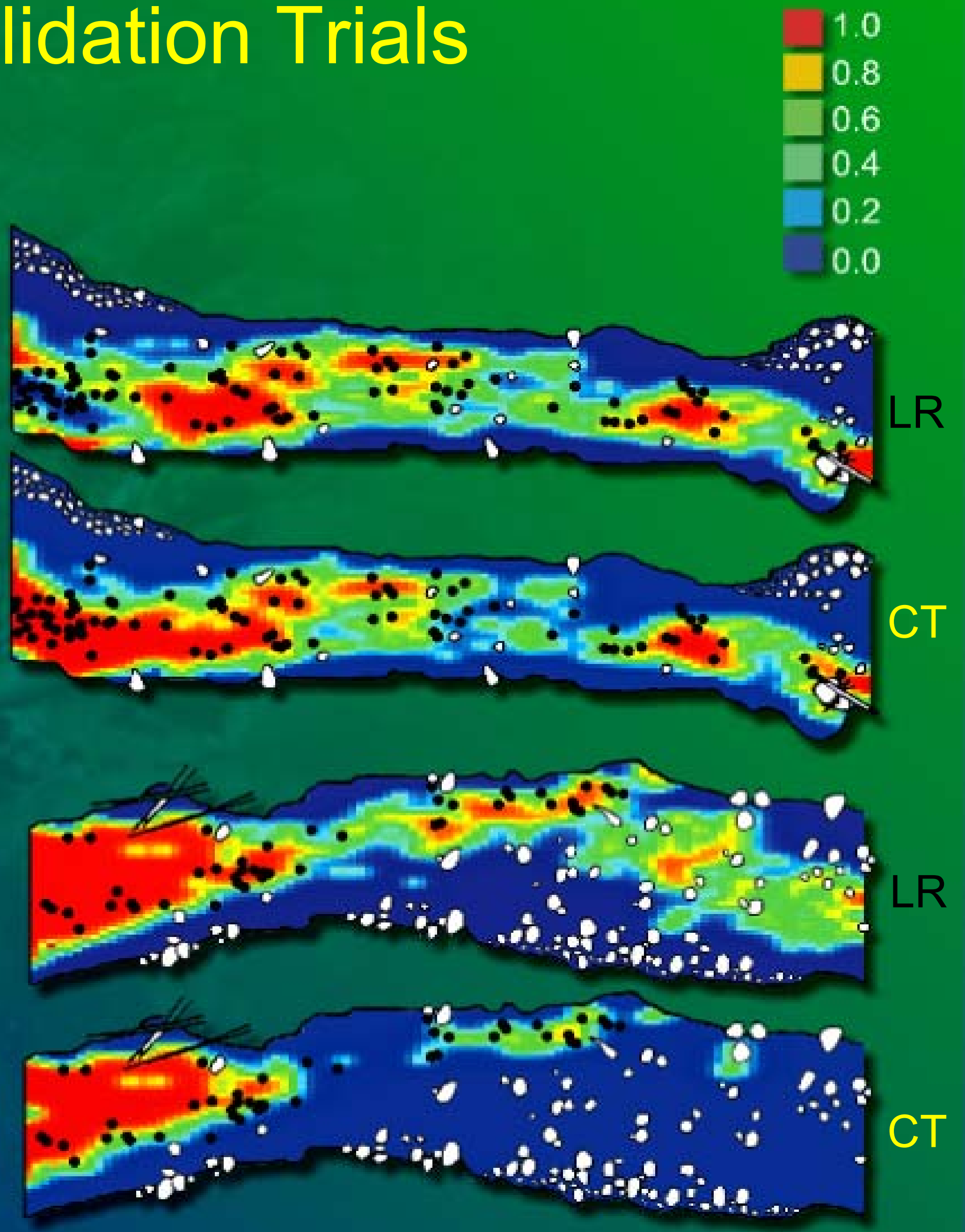
# Activity: Calibration Trials

Prediction Maps



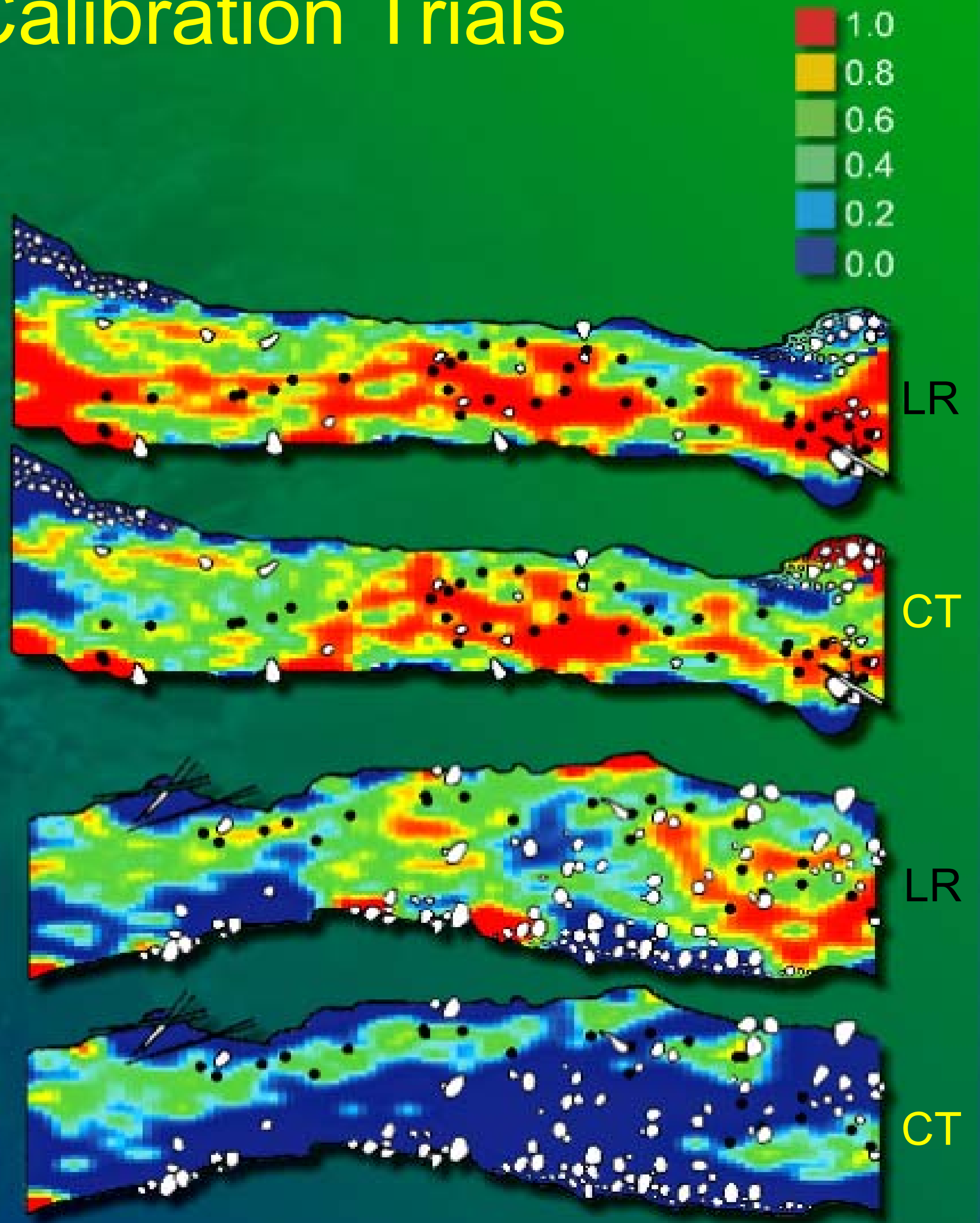
# Activity: Validation Trials

Prediction Maps



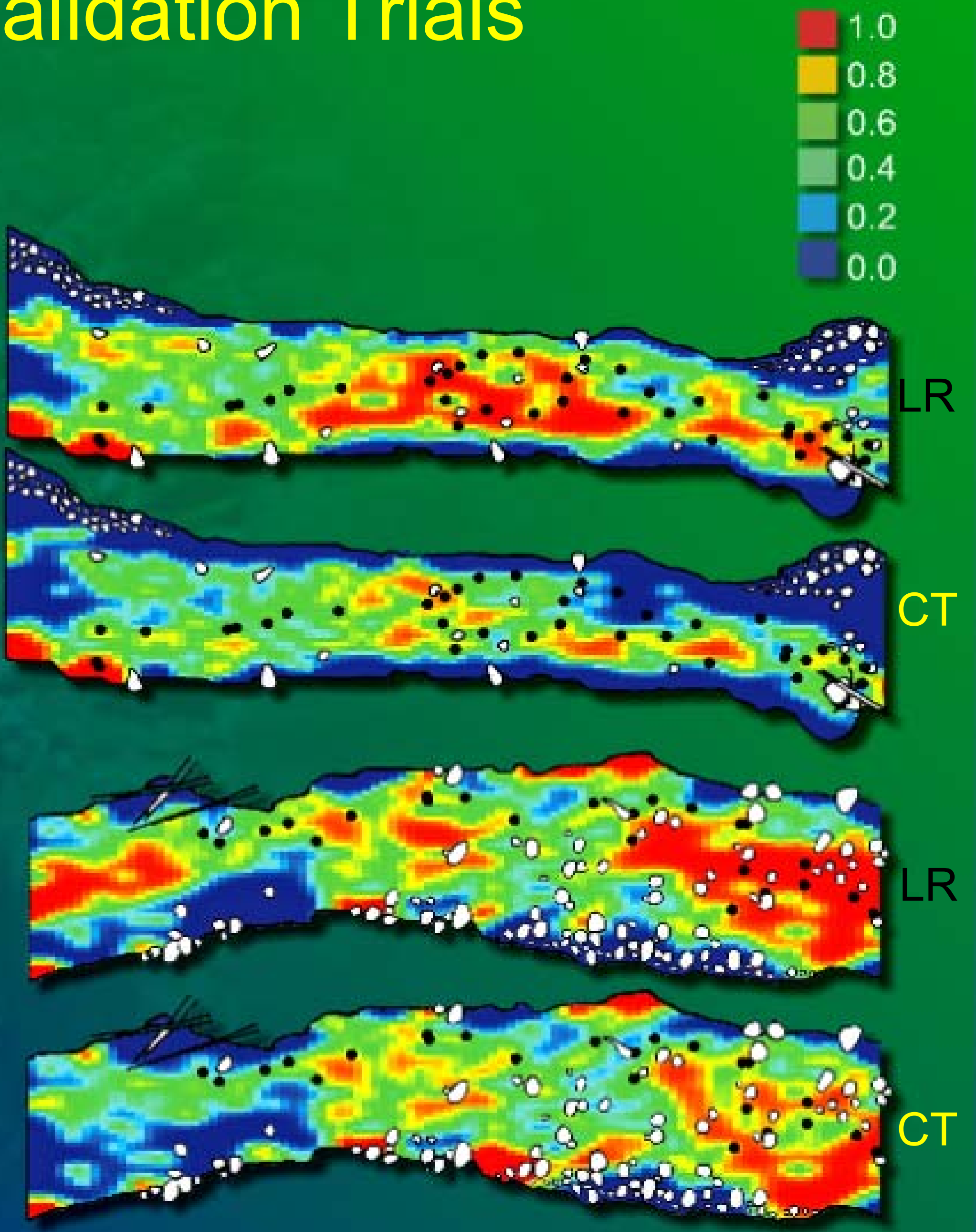
# Sheltering: Calibration Trials

Prediction Maps



# Sheltering: Validation Trials

Prediction Maps



# Conclusions

## Habitat selection and behaviour

**Activity**
Selection mostly a function of water depth

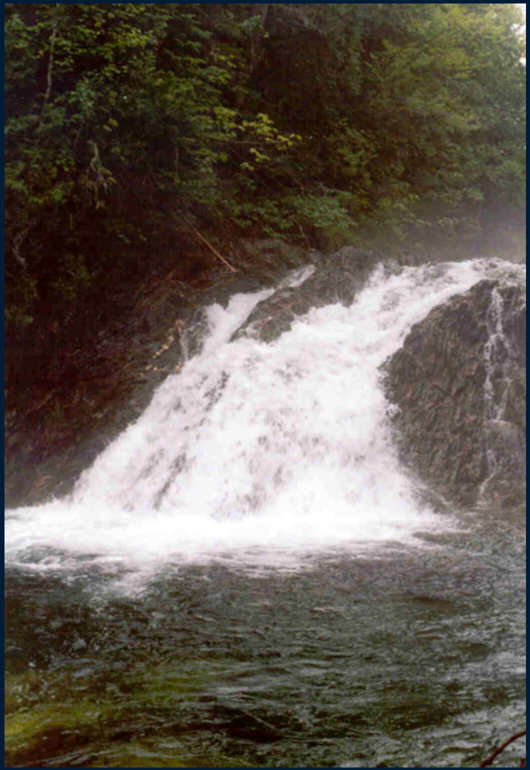
**Sheltering**
Selection for an unembedded rock > 20 cm

- LR and CT models had high:
  - Accuracy in calibration trials
  - Transferability in field validation trials

- Relatively simple LR and CT models sufficed to generate accurate prediction maps

- However, CT models:
  - Were easier to build and interpret
  - Were more parsimonious
  - Had less variable performance in validation trials

# Community structure of stream fishes: A tale of environment and scale



Julie Deschênes

# Introduction

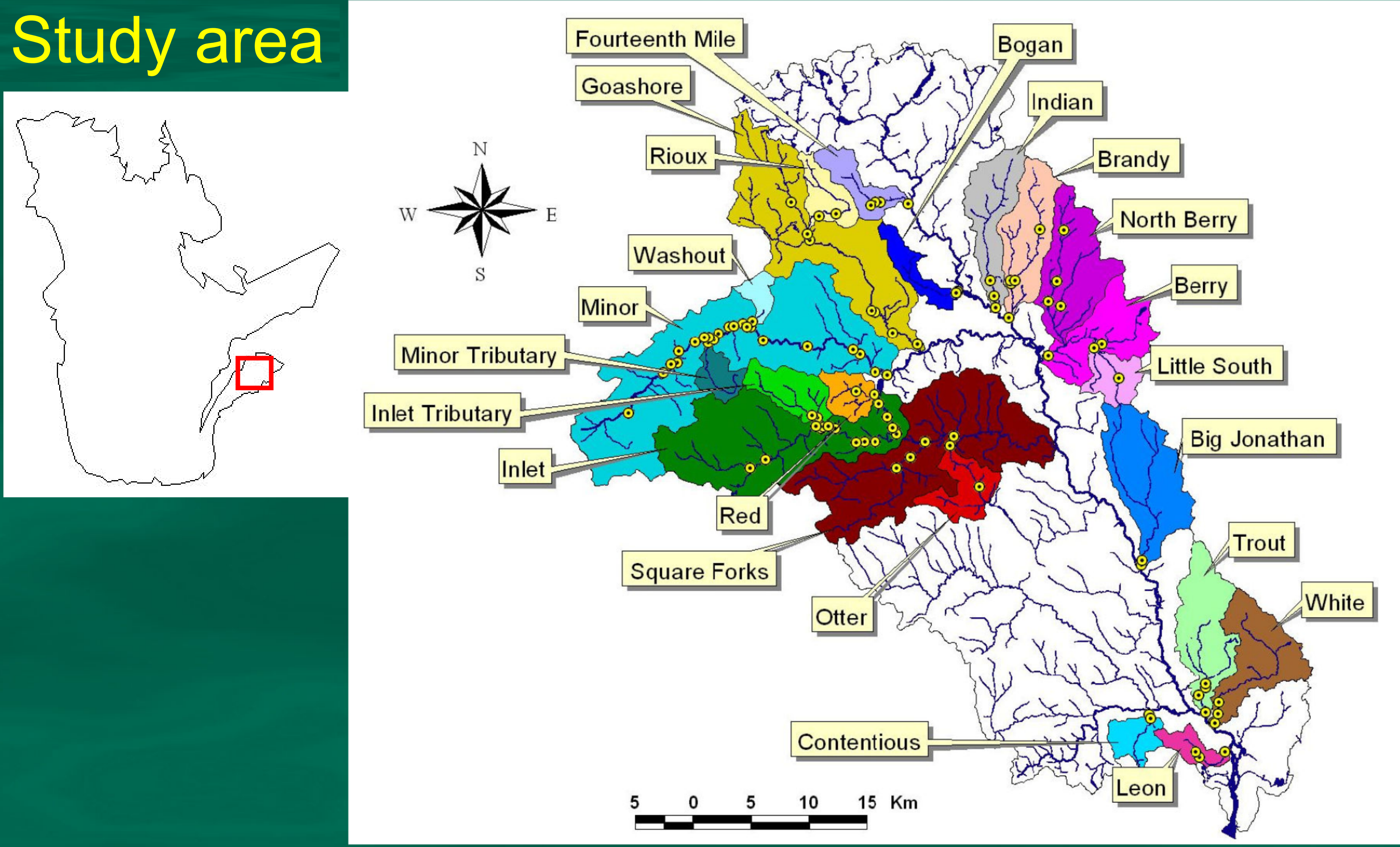
- Spatial scale may influence the relationship between fish distribution and environmental features
  - e.g., scale-dependent effects of:
    - overall cover on abundance of masu salmon
    - woody debris on abundance of golden perch

# Main objectives

- Determine how environmental influences on stream fish communities vary with spatial scale

- Identify the environmental variables most strongly related to community structure at different scales

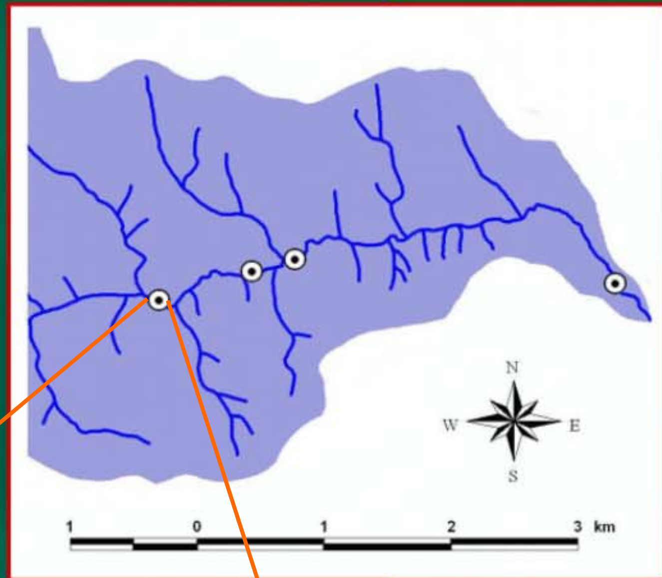
# Study area



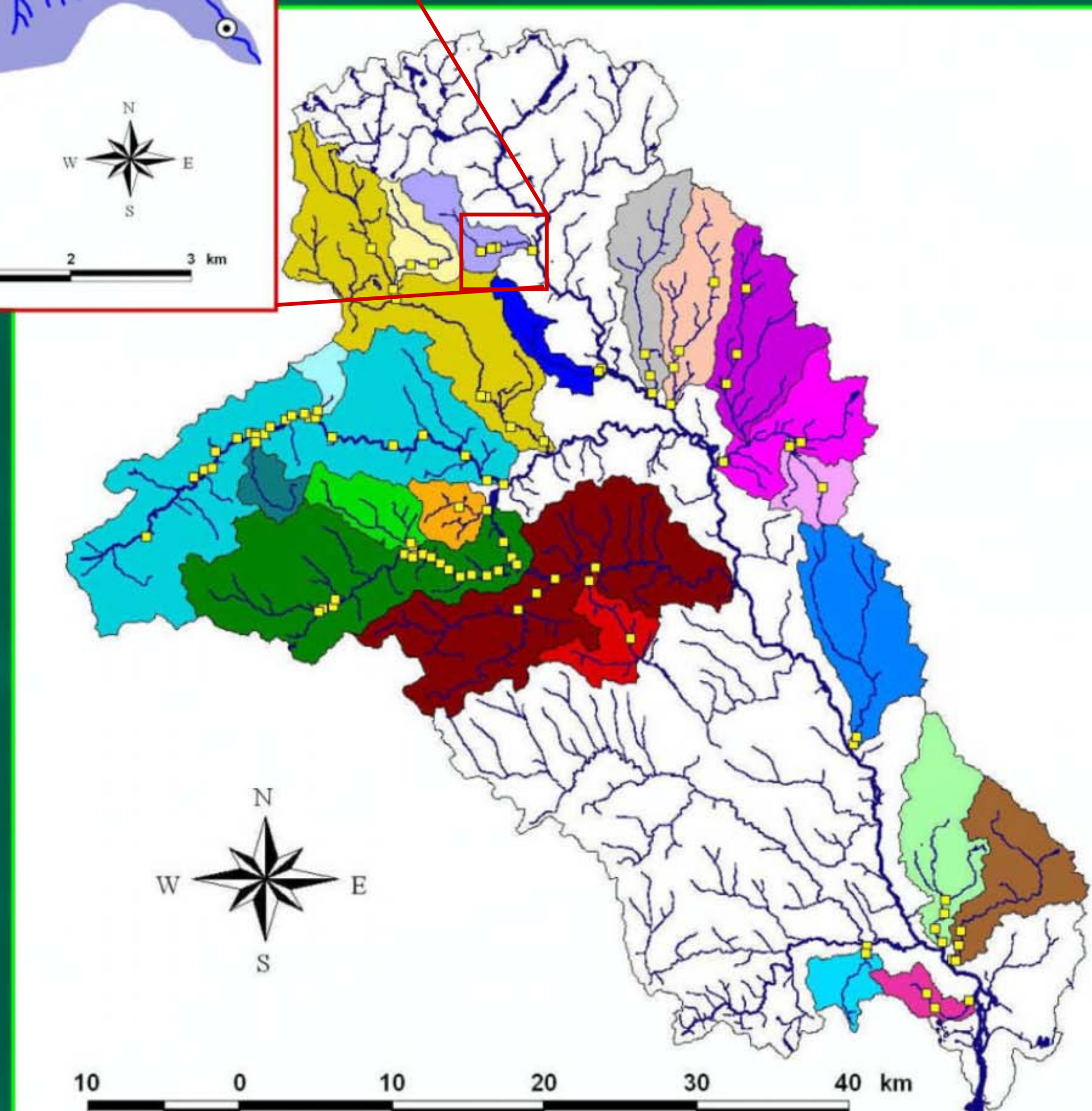

600 sections distributed among 120 reaches and 31 tributary streams of the Cascapedia River, Québec, Canada

# Hierarchical levels

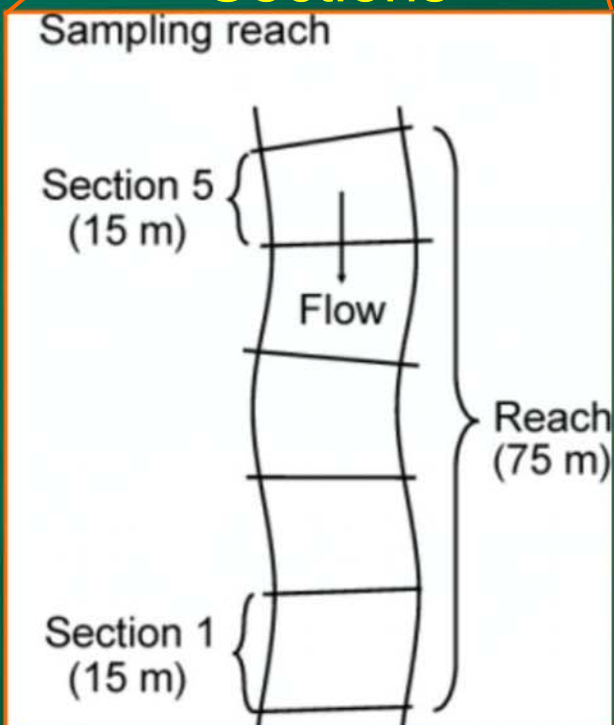
Reaches




Streams




Sections

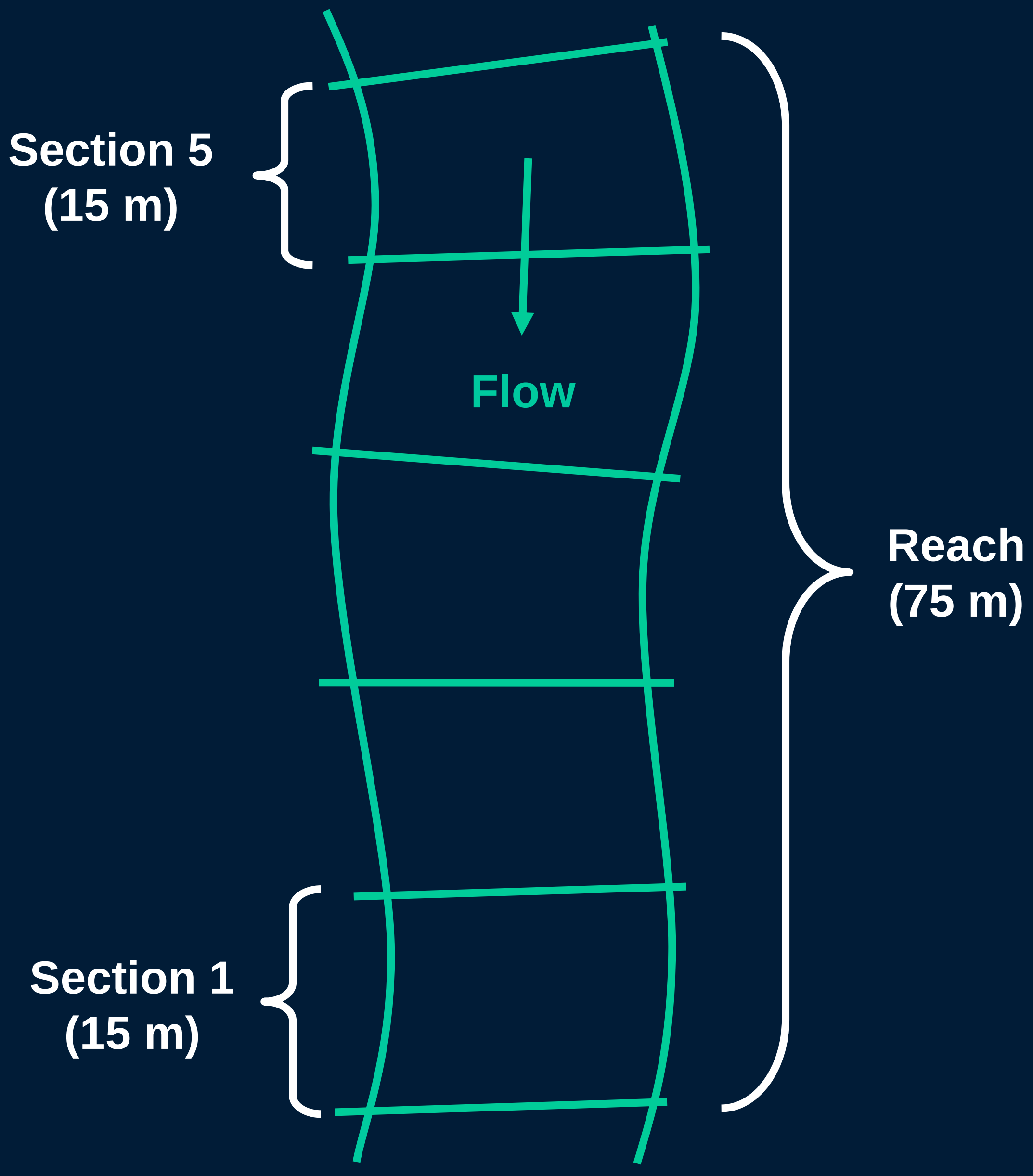


# Sampling and measurement

At each section:

- Fish densities
  - Electrofishing

- Predictor variables
  - 25 environmental features
    - 11 local habitat
    - 11 landscape
    - 3 accessibility



# Environmental predictors

| Variable name | Hierarchical level | Spatial extent |
|---|---|---|
| Mean depth (cm) | Section | Local habitat |
| Mean current velocity ($cm \cdot s^{-1}$) | Section | Local habitat |
| Mean substratum size | Section | Local habitat |
| Plant abundance index | Section | Local habitat |
| Cover index | Section | Local habitat |
| Canopy opening (°) | Section | Local habitat |
| Large woody debris | Section | Local habitat |
| Number of pools | Section | Local habitat |
| Stream slope (°) | Reach | Local habitat |
| Mean wetted width (m) | Reach | Local habitat |
| Temperature (°C) | Reach | Local habitat |
| Terrace width (m) | Reach | Landscape |
| Height at flood (m) | Reach | Landscape |
| Width at flood (m) | Reach | Landscape |
| Entrenchment (%) | Reach | Landscape |
| Altitude (m) | Reach | Landscape |
| Sub-basin area ($km^2$) | Reach | Landscape |
| Total road density ($km \cdot km^{-2}$) | Reach | Landscape |
| Logging 0-4 years old (%) | Reach | Landscape |
| Logging 0-9 years old (%) | Reach | Landscape |
| Logging 0-14 years old (%) | Reach | Landscape |
| Logging 0-19 years old (%) | Reach | Landscape |
| Distance to mainstem | Reach | Accessibility |
| Accessibility index | Reach | Accessibility |
| Distance to mainstem mouth | Stream | Accessibility |

# Three fish species:

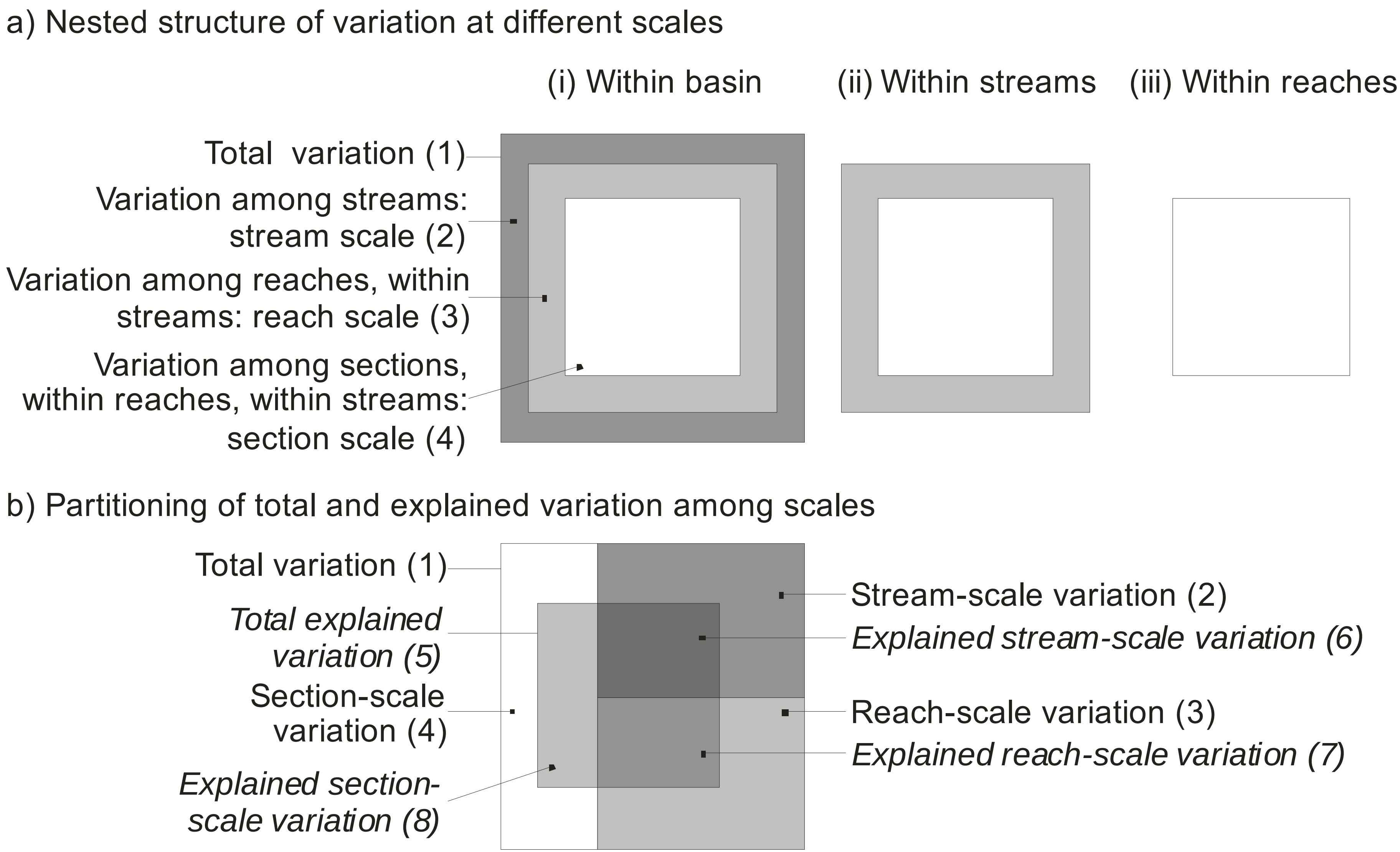

Brook charr
*Salvelinus fontinalis*

Atlantic salmon
*Salmo salar*

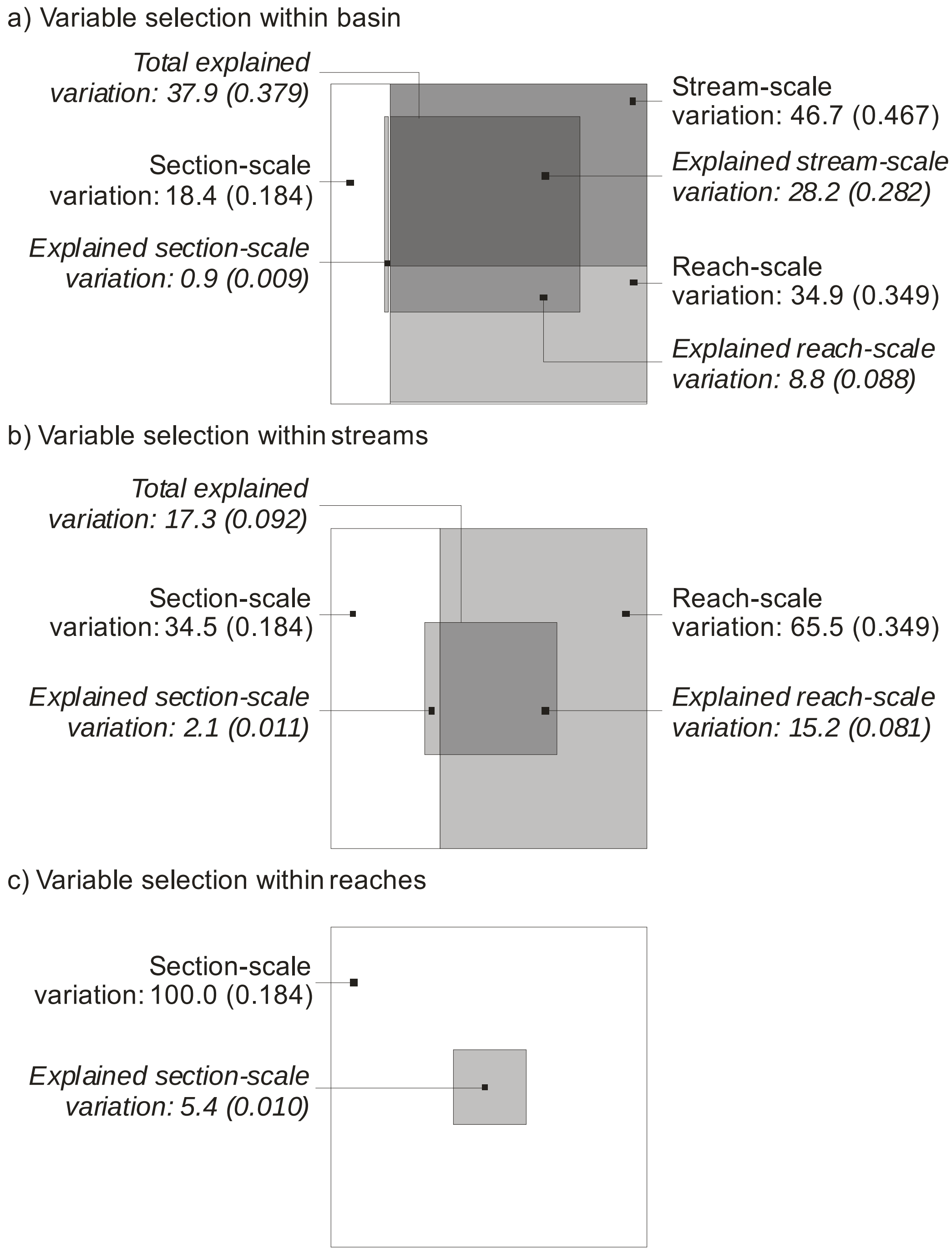
Slimy sculpin
*Cottus cognatus*

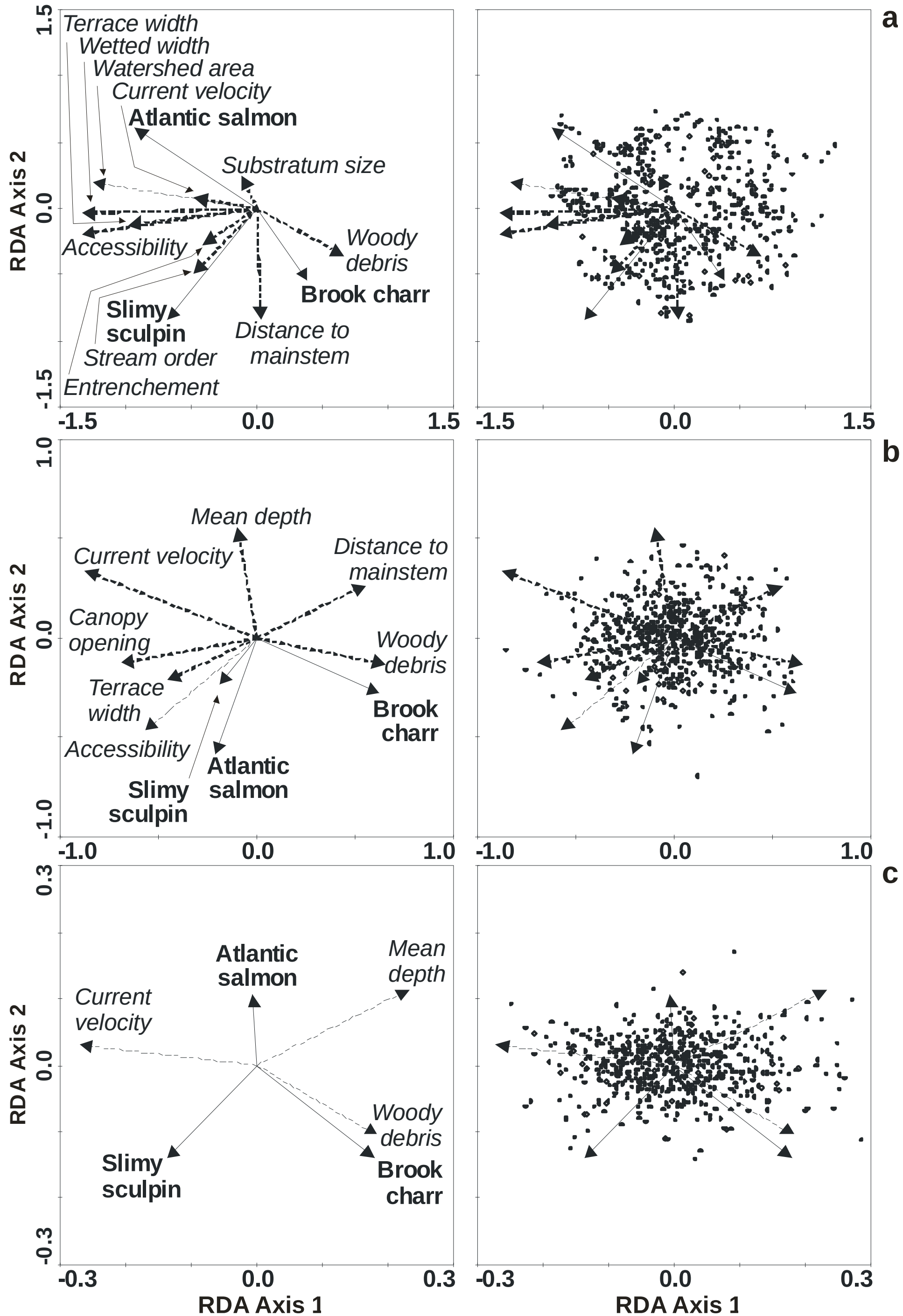
# Statistical analyses

- Redundancy analysis
  - Sampling unit: individual section ($N = 600$)
  - Stepwise variable selection ($p < 0.05$)
  - Restricted permutations (split-plot design)
- Variable selection at each spatial scale
- Hierarchical partitioning of variance
  - Partial redundancy analyses


a) Nested structure of variation at different scales
(i) Within basin
(ii) Within streams
(iii) Within reaches
Total variation (1)
Variation among streams: stream scale (2)
Variation among reaches, within streams: reach scale (3)
Variation among sections, within reaches, within streams: section scale (4)
b) Partitioning of total and explained variation among scales
Total variation (1)
*Total explained variation (5)*
Section-scale variation (4)
*Explained section-scale variation (8)*
Stream-scale variation (2)
*Explained stream-scale variation (6)*
Reach-scale variation (3)
*Explained reach-scale variation (7)*

# Results


a) Variable selection within basin
Total explained variation: 37.9 (0.379)
Stream-scale variation: 46.7 (0.467)
Section-scale variation: 18.4 (0.184)
Explained stream-scale variation: 28.2 (0.282)
Explained section-scale variation: 0.9 (0.009)
Reach-scale variation: 34.9 (0.349)
Explained reach-scale variation: 8.8 (0.088)
b) Variable selection within streams
Total explained variation: 17.3 (0.092)
Section-scale variation: 34.5 (0.184)
Reach-scale variation: 65.5 (0.349)
Explained section-scale variation: 2.1 (0.011)
Explained reach-scale variation: 15.2 (0.081)
c) Variable selection within reaches
Section-scale variation: 100.0 (0.184)
Explained section-scale variation: 5.4 (0.010)


a
b
c
Terrace width
Wetted width
Watershed area
Current velocity
Atlantic salmon
Substratum size
Woody debris
Accessibility
Brook charr
Slimy sculpin
Distance to mainstem
Stream order
Entrenchement
Mean depth
Canopy opening
Terrace width
RDA Axis 1
RDA Axis 2
1.5
0.0
-1.5
1.0
-1.0
0.3
-0.3

Contribution of environmental variables to explained variation in assemblage structure at different spatial scales. Empty cells indicate that a variable had no significant influence at a given scale. Selected environmental variables explained 39.1% of the total variation at the within-basin scale ($E_B$), 17.6% at the within-streams scale ($E_S$), and 5.4% at the within-reaches scale ($E_R$).

| Spatial extent | Environmental variable | Contribution of environmental variable to explained variation at each scale (%) | | |
|---|---|---|---|---|
| | | Within basin ($E_B$) | Within streams ($E_S$) | Within reaches ($E_R$) |
| Accessibility | Accessibility index | 29.6 | 27.6 | |
| Accessibility | Distance to mainstem | 13.0 | 9.6 | |
| Landscape | Stream order | 12.7 | | |
| Landscape | Watershed area | 7.2 | | |
| Landscape | Terrace width | 4.9 | 9.6 | |
| Landscape | Entrenchment | 3.6 | | |
| Local habitat | Mean wetted width | 13.5 | | |
| Local habitat | Mean current velocity | 6.9 | 23.4 | 50.0 |
| Local habitat | Large woody debris | 3.3 | 11.7 | 20.0 |
| Local habitat | Mean substratum size | 2.8 | | |
| Local habitat | Height increment at flood | 2.5 | 6.4 | |
| Local habitat | Mean depth | | 11.7 | 30.0 |
| | | 100.0 | 100.0 | 100.0 |

# Conclusions

- Fish assemblage structure varied more strongly among streams and among reaches than among sections, and the relationships between species assemblage structure and environmental features differed across scales
- The variation explained by environmental variables was highest at the within-basin (among-stream) scale

- Interpretation of environmental effects on fish community structure in the Cascapedia River Basin strongly depended on observational scale
  - large-scale variation in accessibility and stream size explained a major proportion of variation among streams and among reaches
  - effects of water velocity and woody debris were detectable at all scales

- Prediction of fish assemblage structure in streams may be increasingly reliable at coarser spatial scales
- Integration of information across scales is required to fully understand the mechanisms structuring fish assemblages and identify the scales at which they operate most strongly

# Acknowledgments

Katrine Turgeon
Julie Deschênes
Marie-Noëlle Rivard
Myriam Chénier-Soulières
Nicolas Martel
Véronique Bérard

Karen Brassard
Claudia Cossette
Marie-Josée Gagnon
Émilie Racicot
Catherine Sauvé

Fonds de recherche sur la nature et les technologies Québec
NSERC CRSNG
CIRSA