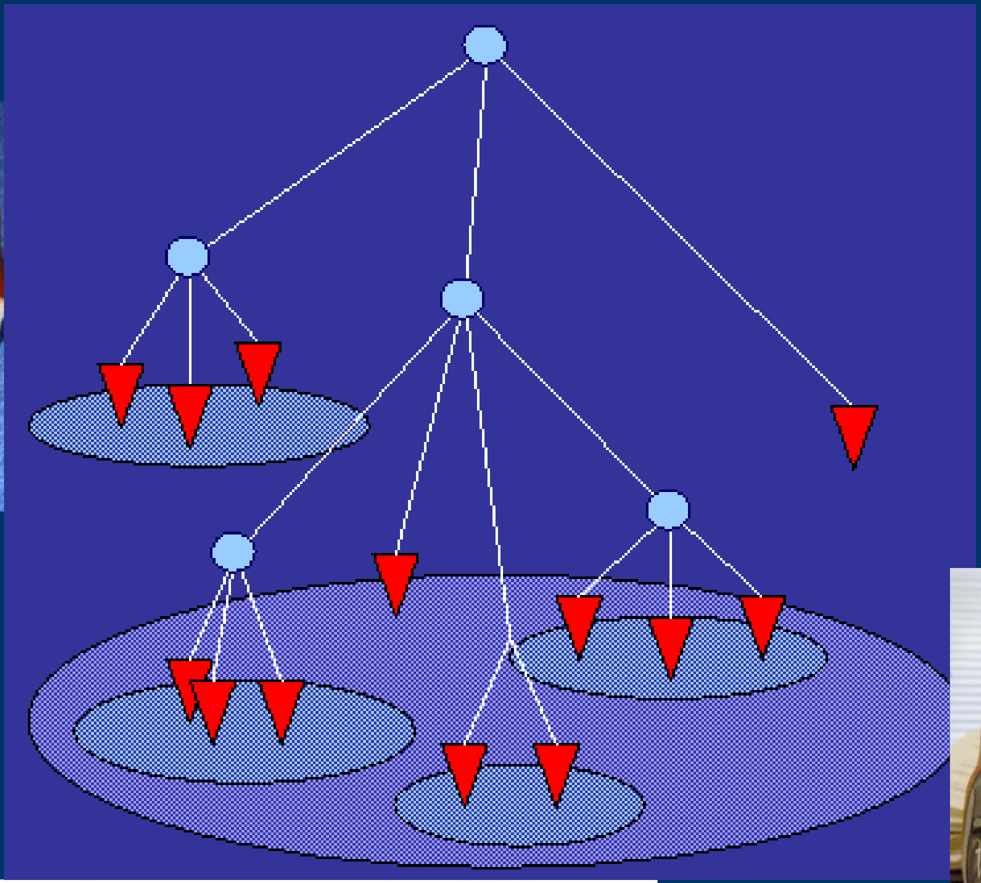
Multilevel modelling of fish abundance in streams

Marco A. Rodríguez

Université du Québec à Trois-Rivières

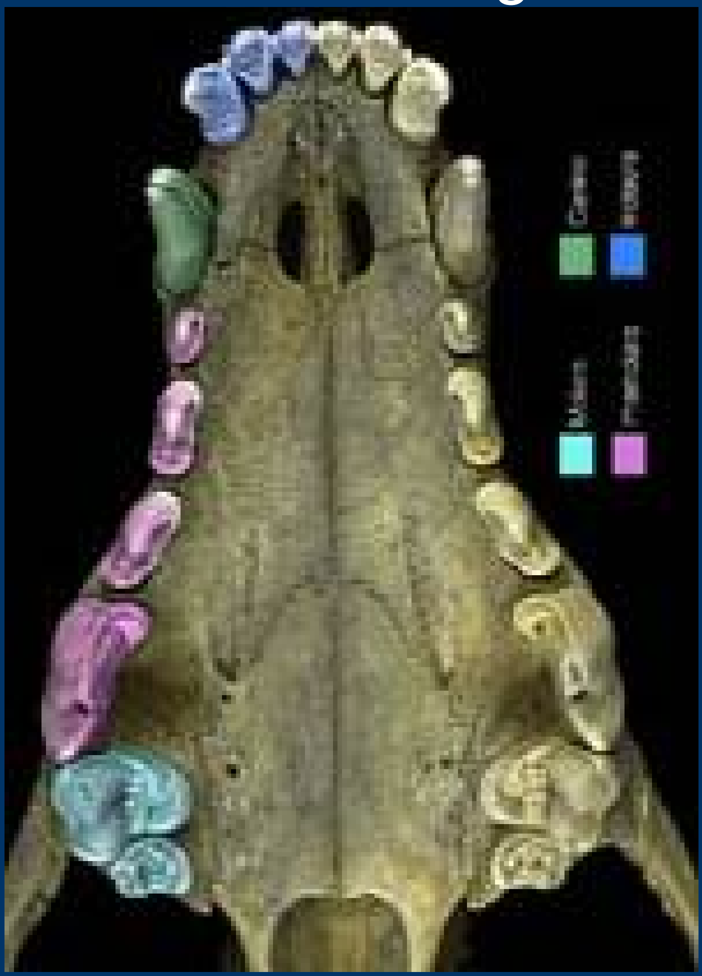
UQTR

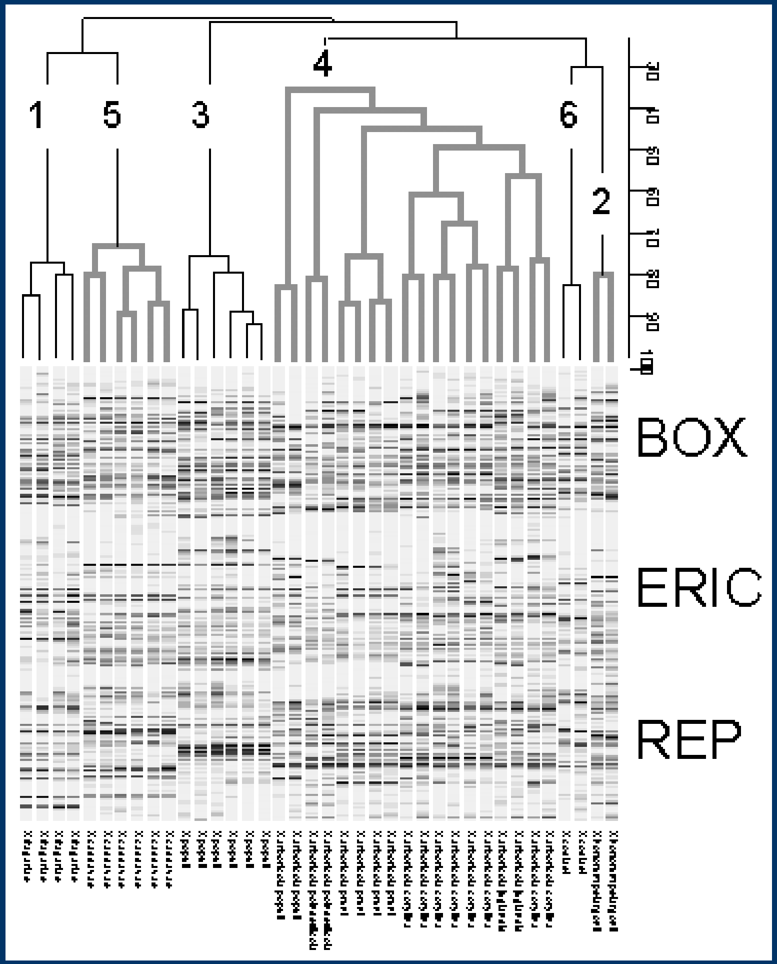
Hierarchies and nestedness are common in nature



Biological data often have clustered or nested structure, in which observations are made on units grouped at different hierarchical levels

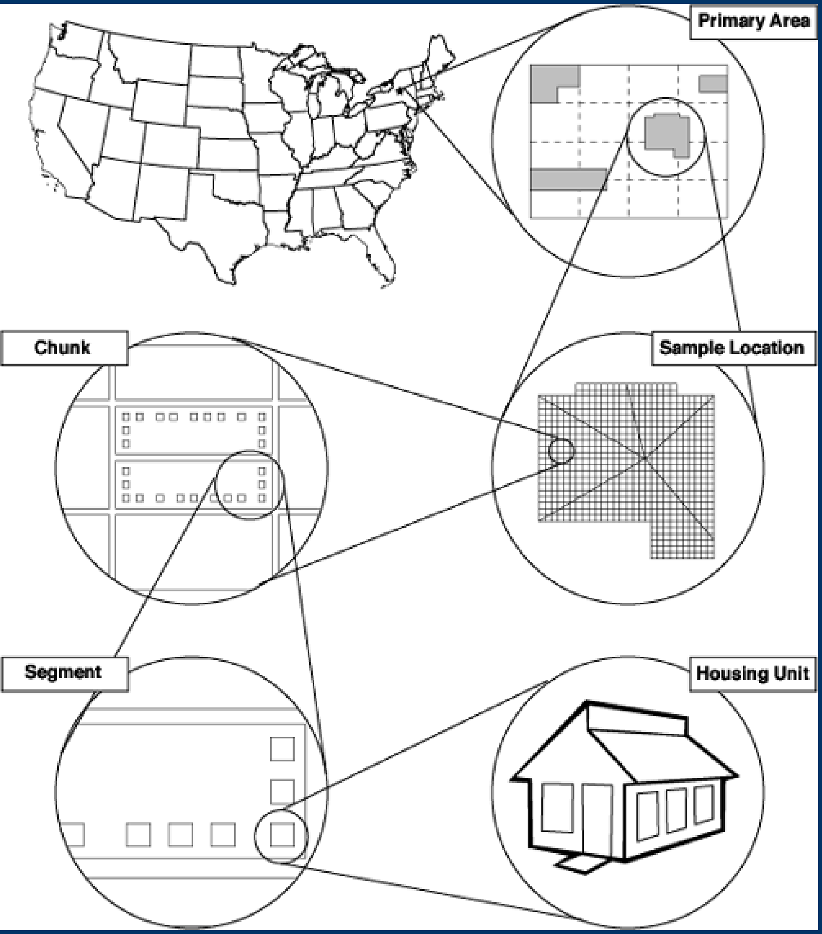
Nestedness in biological data arises both from natural structure…

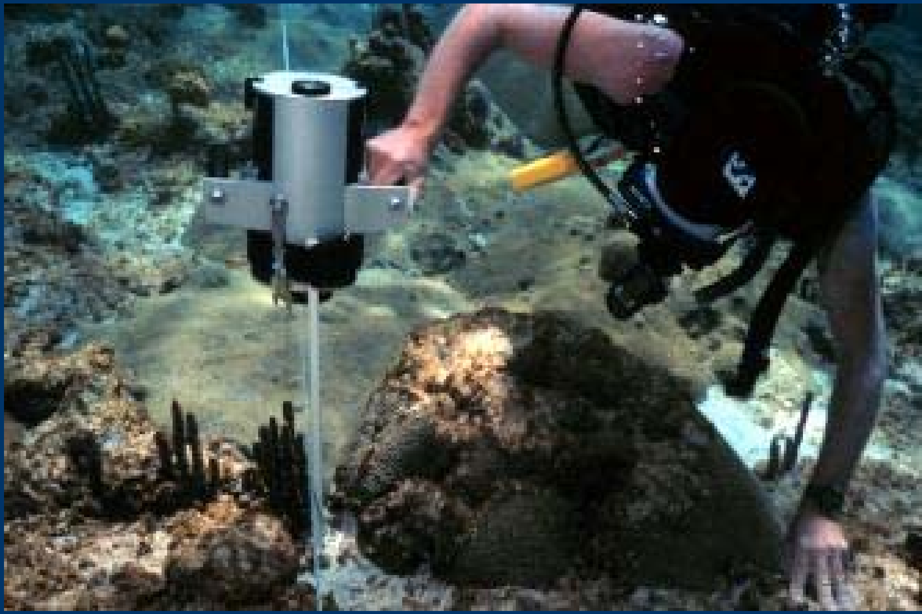


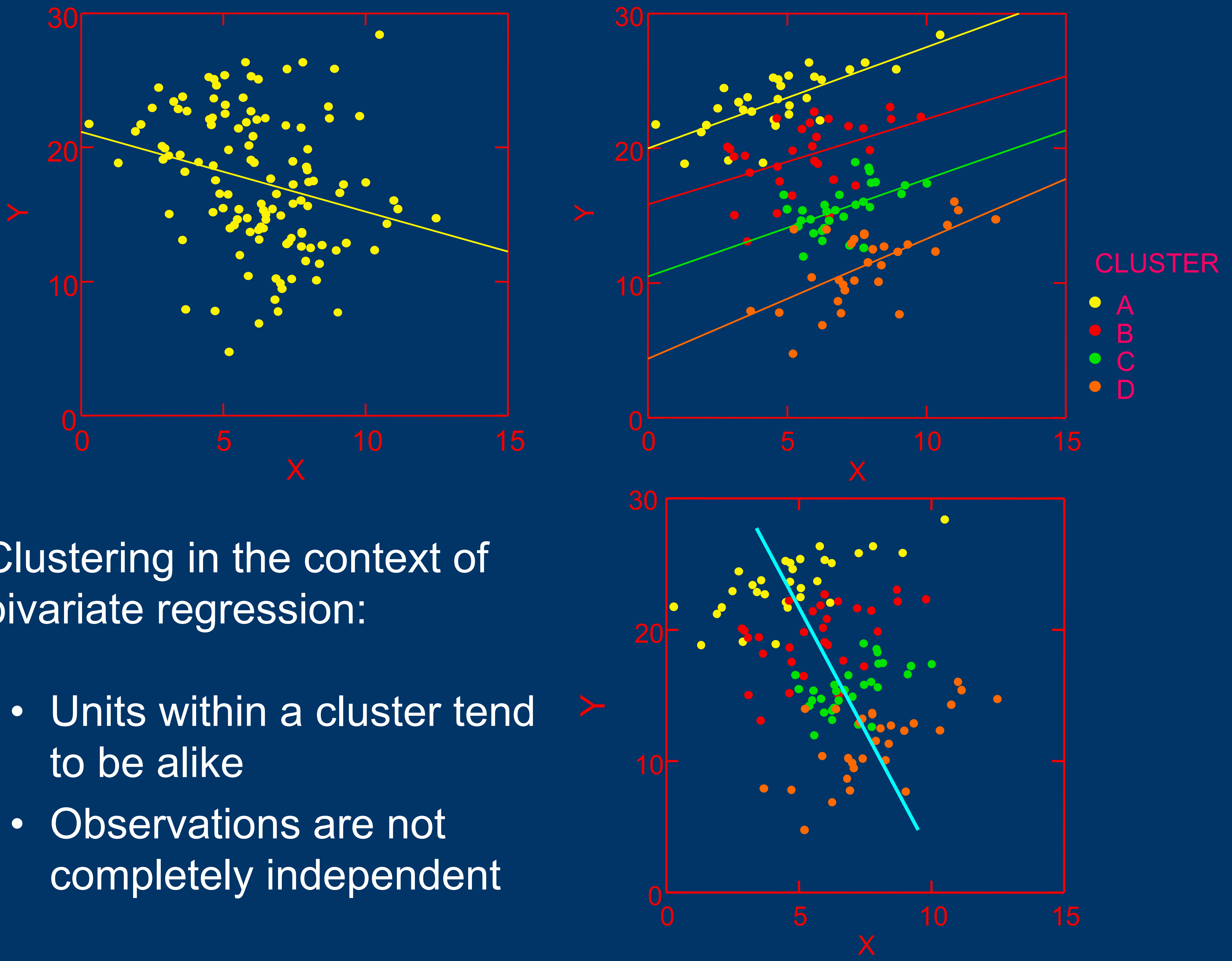


e.g., teeth, or genes, nested within individuals

…and for methodological reasons (e.g. cluster sampling driven by logistical constraints)







Clustering in the context of bivariate regression:

- Units within a cluster tend to be alike
- Observations are not completely independent

Hierarchical linear (HL) models explicitly account for intra-group correlations and allow for modelling of variation at lower levels as a function of higher-level effects

- Random coefficient model
- Variance component model
- Multilevel regression model

Linear mixed-effects model (Laird and Ware 1982):

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, M$$

$$\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$$

- Basic two-level HL model with a single level-1 predictor:

  $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$,   i: sample index, j: group index

  Level-1 coefficients (βs) can be directly related to level-2 predictors (Z_j):

  $\beta_{0j} = \alpha_{00} + \alpha_{01}Z_j + u_{0j}$

  $\beta_{1j} = \alpha_{10} + \alpha_{11}Z_j + u_{1j}$

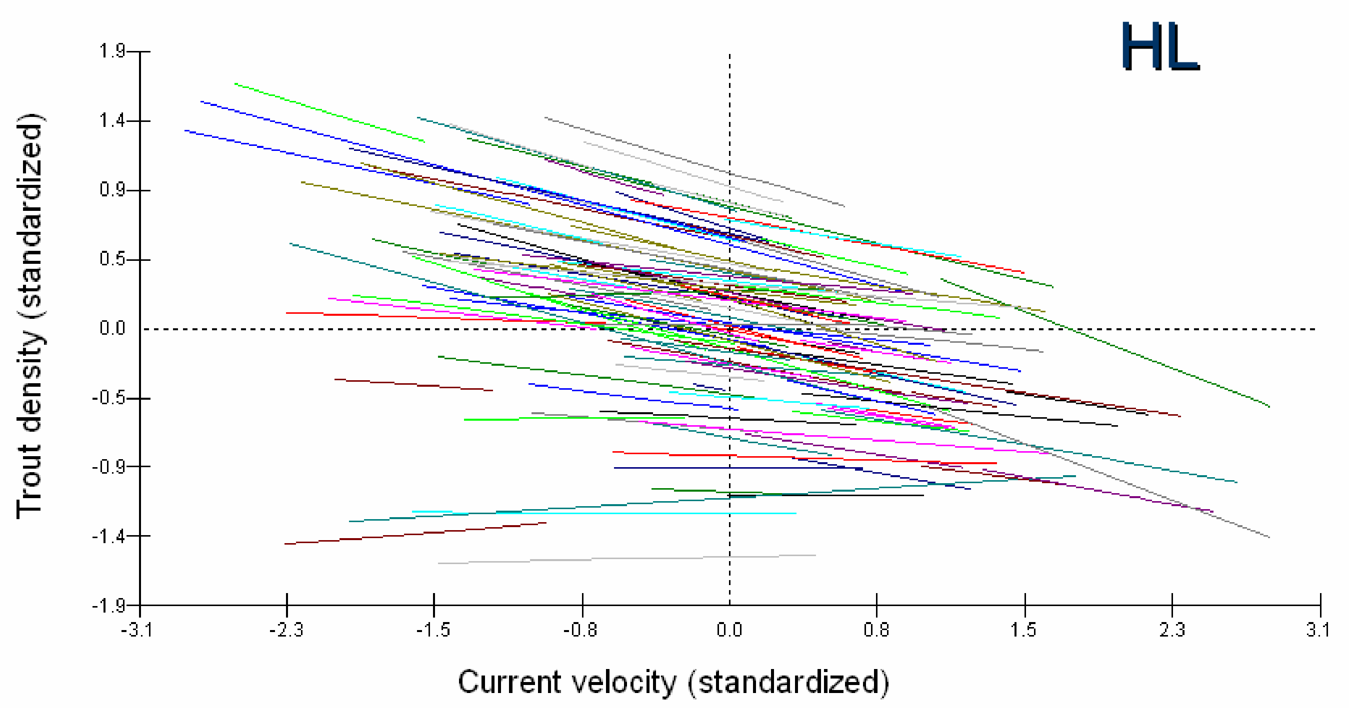
  $Y_{ij} = \alpha_{00} + \alpha_{10}X_{ij} + \alpha_{01}Z_j + \alpha_{11}X_{ij}Z_j + u_{1j}X_{ij} + u_{0j} + e_{ij}$

  $\boldsymbol{u}_j \sim N(\mathbf{0},\boldsymbol{\Omega}_\mathbf{u})$, $e_{ij} \sim N(\mathbf{0},\boldsymbol{\Omega}_\mathbf{e})$

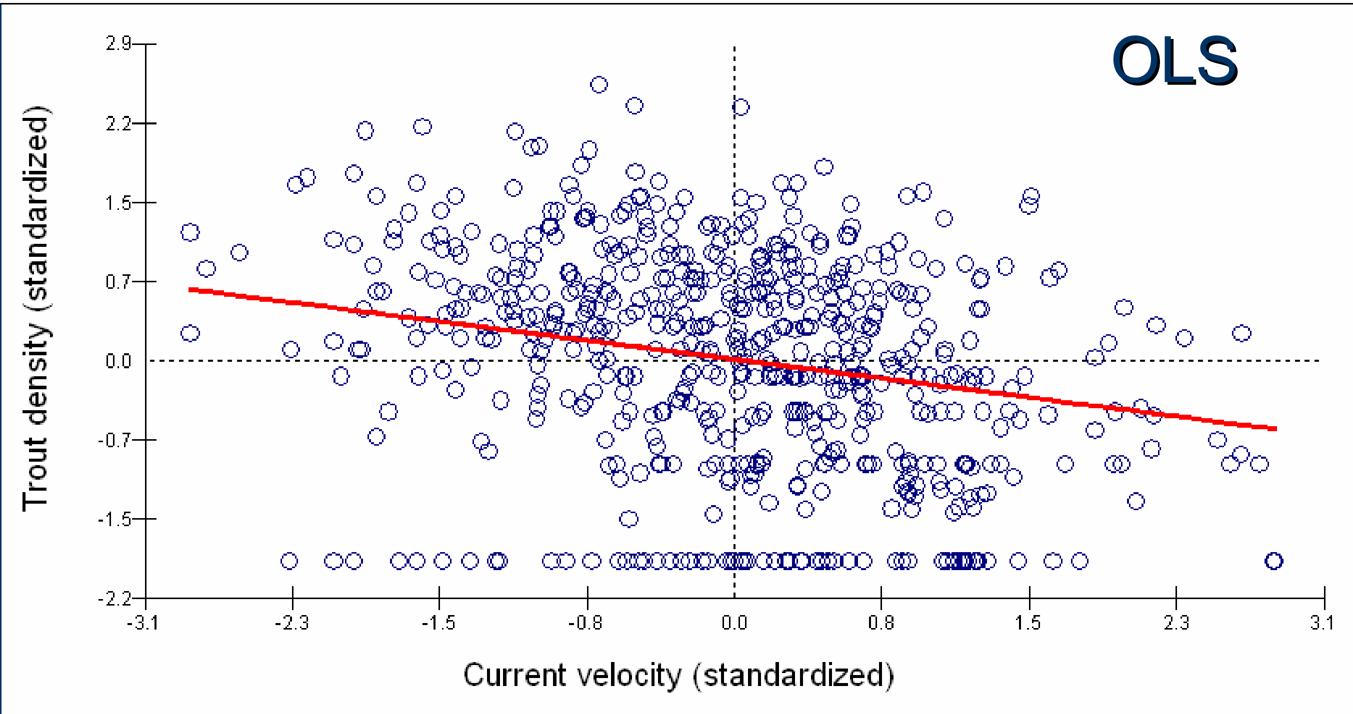
- A key difference between HL and ordinary least squares (OLS) regression lies in the random parts: $u_{1j}X_{ij} + u_{0j}$

- Hierarchical linear model:
  - Units within groups can be dependent
  - Groups are considered as randomly sampled from a larger population of groups
  - Parameters that vary across groups are composed of fixed and random portions
- Ordinary least squares regression:
  - All units are considered independent regardless of group affiliation
  - Parameters are fixed and do not vary across groups

- Group-specific intercept and slope
- Empirical Bayes shrinkage estimators ("borrowing-of-strength")



- Common intercept
- Common slope



Submodels of the general multilevel (HL) model			
Model	Parameter noise	Level-1 predictors	Level-2 predictors
General multilevel	Yes	Yes	Yes
Random coefficients	Yes	Yes	No
Means(slopes)-as-outcomes	Yes	No	Yes
Random effects ANOVA	Yes	No	No
Interactions	No	Yes	Yes

modified from Steenbergen and Jones (2002)

Case study examining the relationship between population density of brook trout and habitat features nested at different hierarchical levels in a drainage basin

Objectives

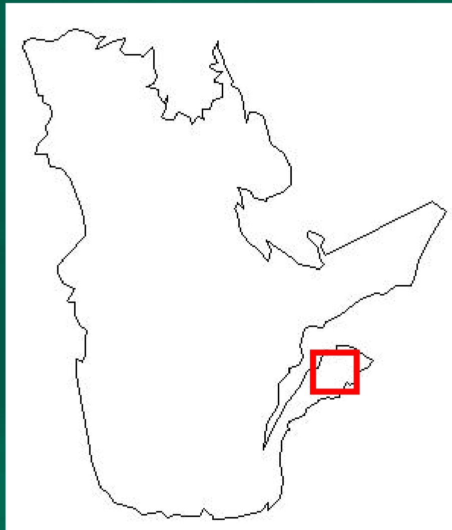
Questions about units at different levels

- How does brook trout density vary with local habitat (level-1) characteristics such as current velocity?
- How does brook trout density vary with larger-scale reach (level-2) characteristics such as basin area?

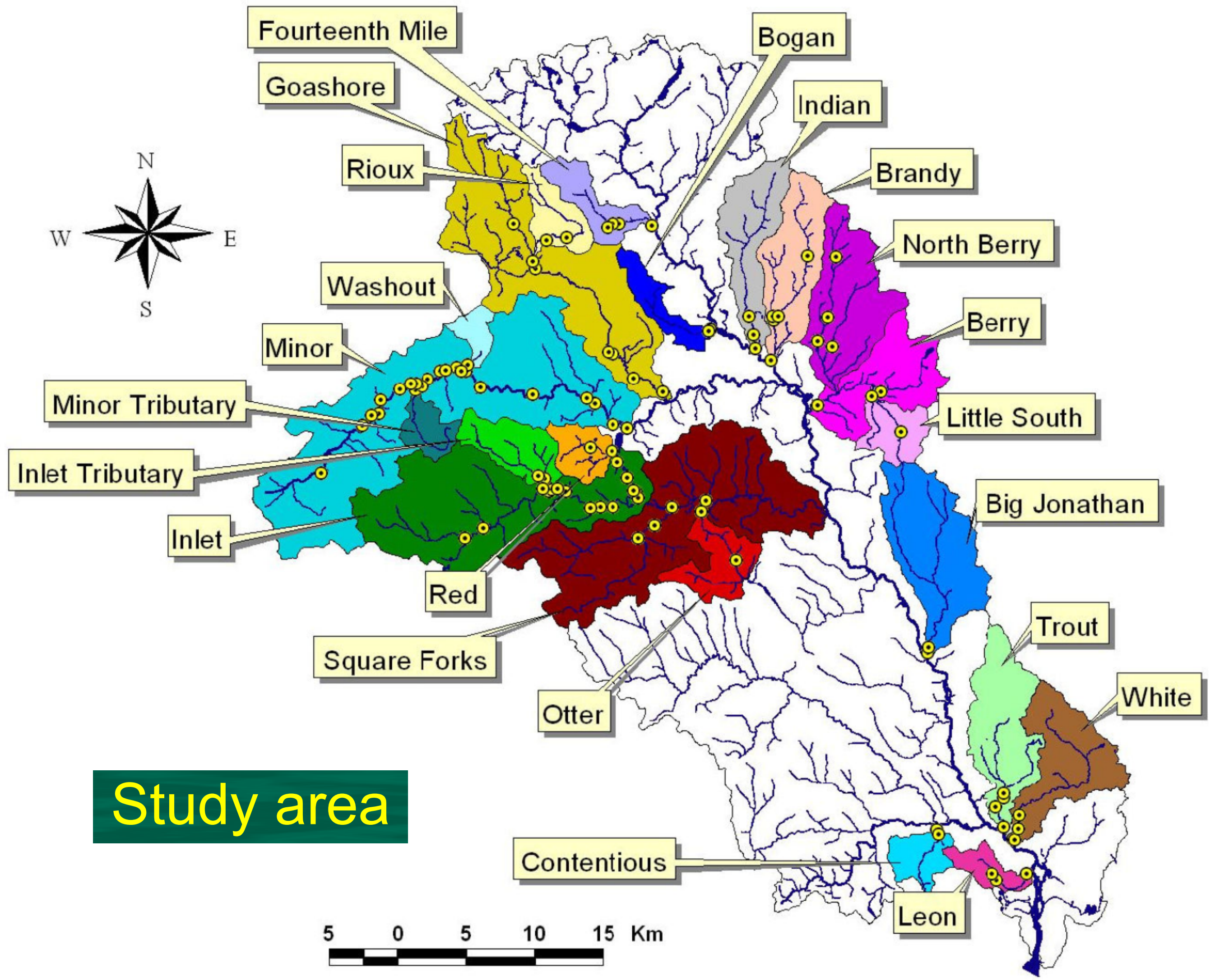
Examination of cross-level interactions

- Contextual effects, e.g., does the relationship between brook trout density and current velocity (level 1) depend on basin area (level 2)?

Methods

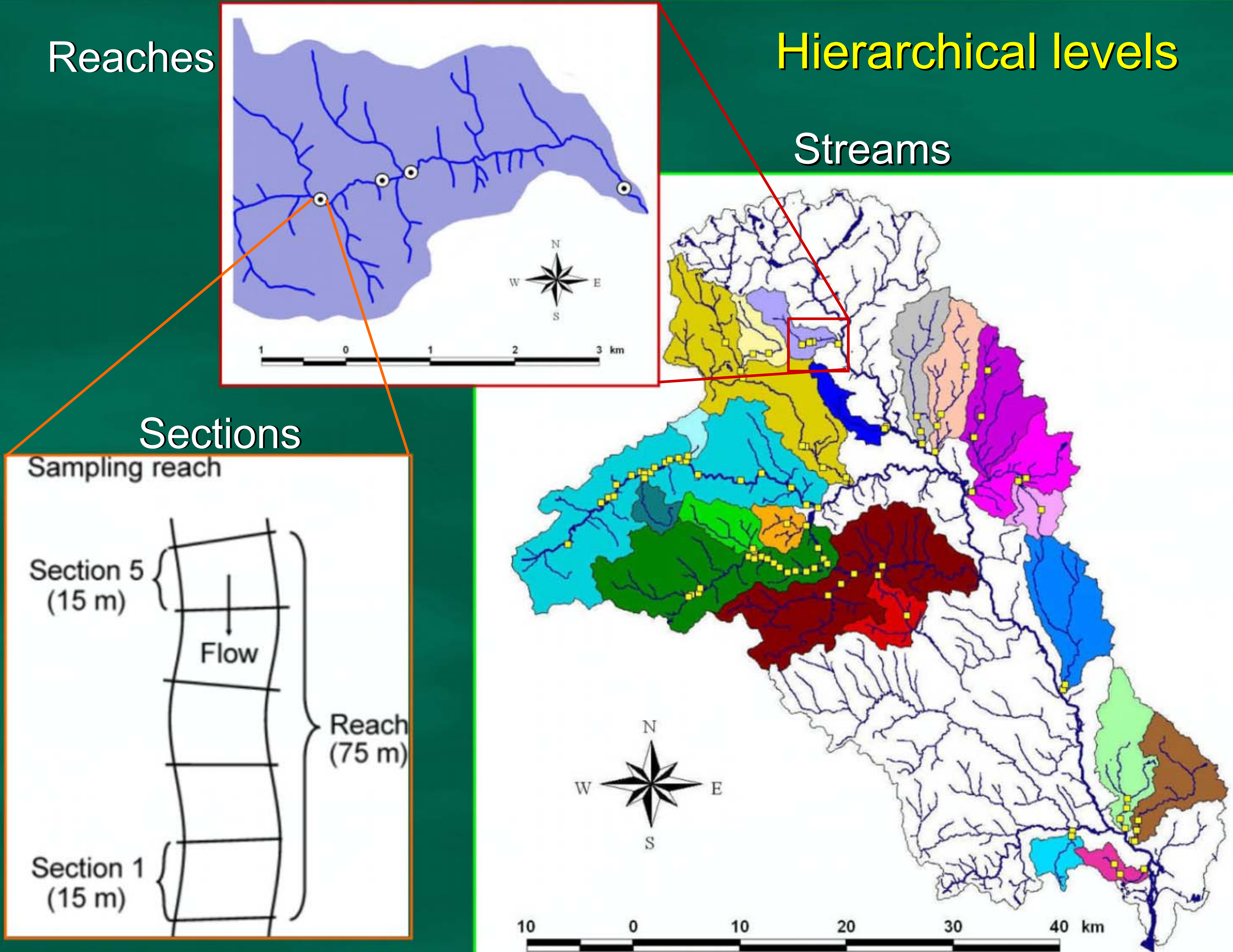


Julie Deschênes



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

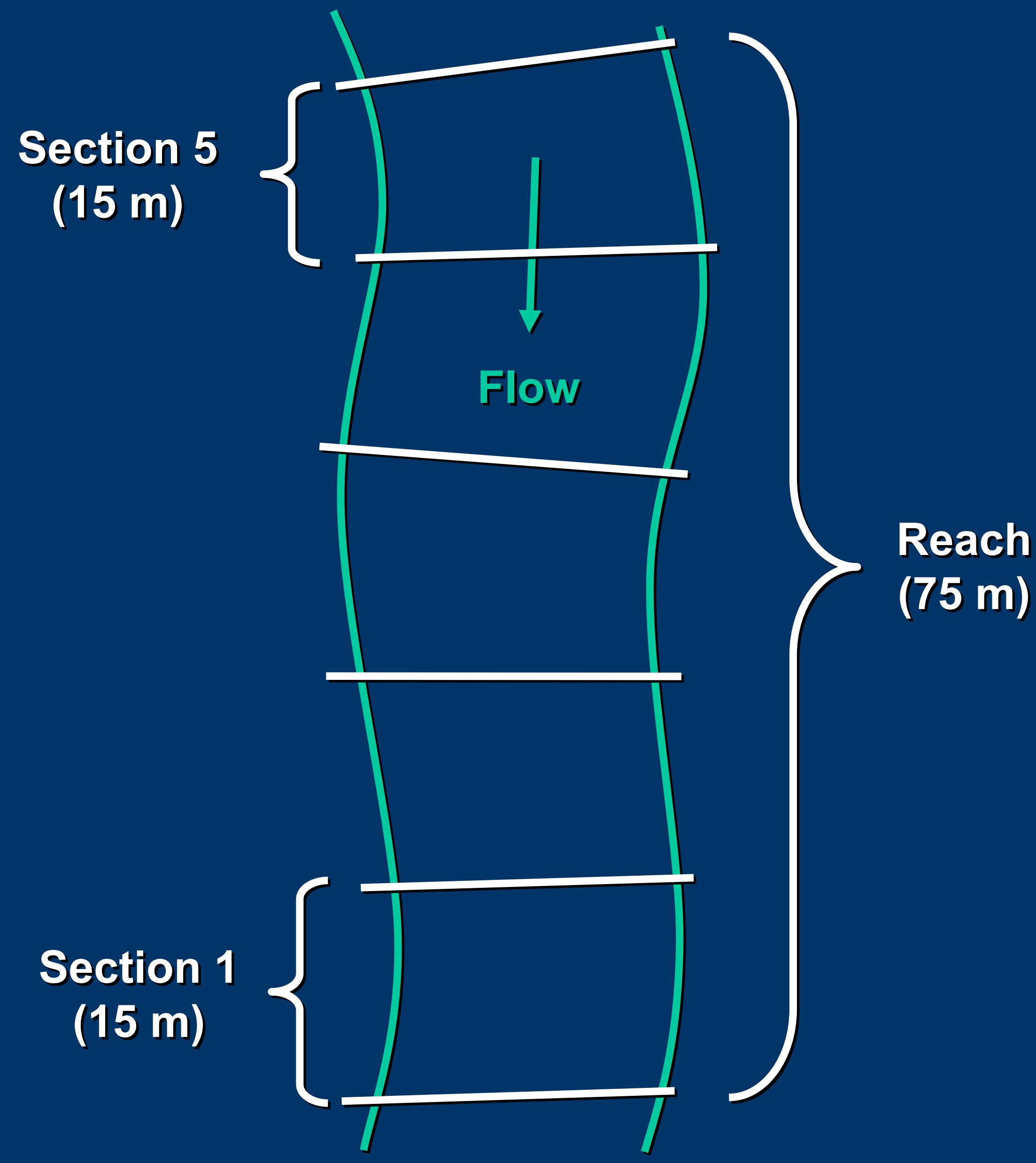
Hierarchical levels



Sampling and measurement

At each section:

- Brook trout density
  - 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

- HL model (multilevel regression) comprising three levels:
    1. Sections (i)
    2. Reaches (j)
    3. Streams (k)

  - Autocorrelated errors (AR1) for residuals at level 1 to account for spatial proximity (dependence) of sections

  - Estimation by full ML (IGLS) because interest focused on fixed effects

  - Deviance tests for selection of environmental predictors

- Ordinary least squares multiple regression using the same predictors as the hierarchical linear model

Results

- Initial step: decomposition of the total variance by use of an intercept-only (variance components) model

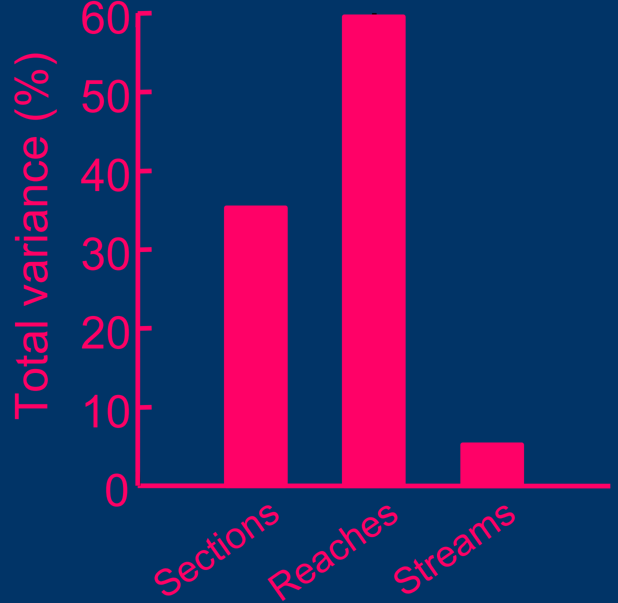
$$Y_{ijk} = \beta_0 + u_{0jk} + v_{0k} + e_{ijk}$$

β_0 = overall mean of Y

e_{ijk} = variance of Y at section level: 35.3%

u_{0jk} = variance of Y at reach level: 59.5%

v_{0k} = variance of Y at stream level: 5.2%



- Trout density did not vary significantly at the highest level, across streams
  - The model was therefore simplified to two levels:
    - sections (level 1) within reaches (level 2)

HL model

$$\text{trout density}_{ij} \sim N(XB, \Omega)$$

$$\text{trout density}_{ij} = \beta_{0ij} + \beta_{1j}\text{current velocity}_{ij} + \beta_2\text{woody debris}_{ij} + \beta_3\text{cover}_{ij} + \beta_4\text{sub-basin area}_j + \beta_5\text{height at flood}_j + \beta_6\text{current.sub-basin}_{ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0,\ \Omega_u) \ : \ \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u10} & \sigma^2_{u1} \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0,\ \Omega_e) \ : \ \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

OLS model

$$\text{trout density}_{ij} \sim N(XB, \Omega)$$

$$\text{trout density}_{ij} = \beta_{0i} + \beta_1\text{current velocity}_{ij} + \beta_2\text{woody debris}_{ij} + \beta_3\text{cover}_{ij} + \beta_4\text{sub-basin area}_j + \beta_5\text{height at flood}_j + \beta_6\text{current.sub-basin}_{ij}$$

$$\beta_{0i} = \beta_0 + e_{0ij}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0,\ \Omega_e) \ : \ \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

HL model

$$\text{trout density}_{ij} \sim N(XB, \Omega)$$

$$\text{trout density}_{ij} = \beta_{0ij} + \beta_{1j}\text{current velocity}_{ij} + 0.155(0.034)\text{woody debris}_{ij} + 0.109(0.042)\text{cover}_{ij} + -0.212(0.064)\text{sub-basin area}_{j} + -0.210(0.070)\text{height at flood}_{j} + -0.095(0.042)\text{current.sub-basin}_{ij}$$

$$\beta_{0ij} = 0.014(0.064) + u_{0j} + e_{0ij}$$

$$\beta_{1j} = -0.224(0.043) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.372(0.068) & \\ -0.053(0.028) & 0.045(0.025) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.334(0.048) \end{bmatrix}$$

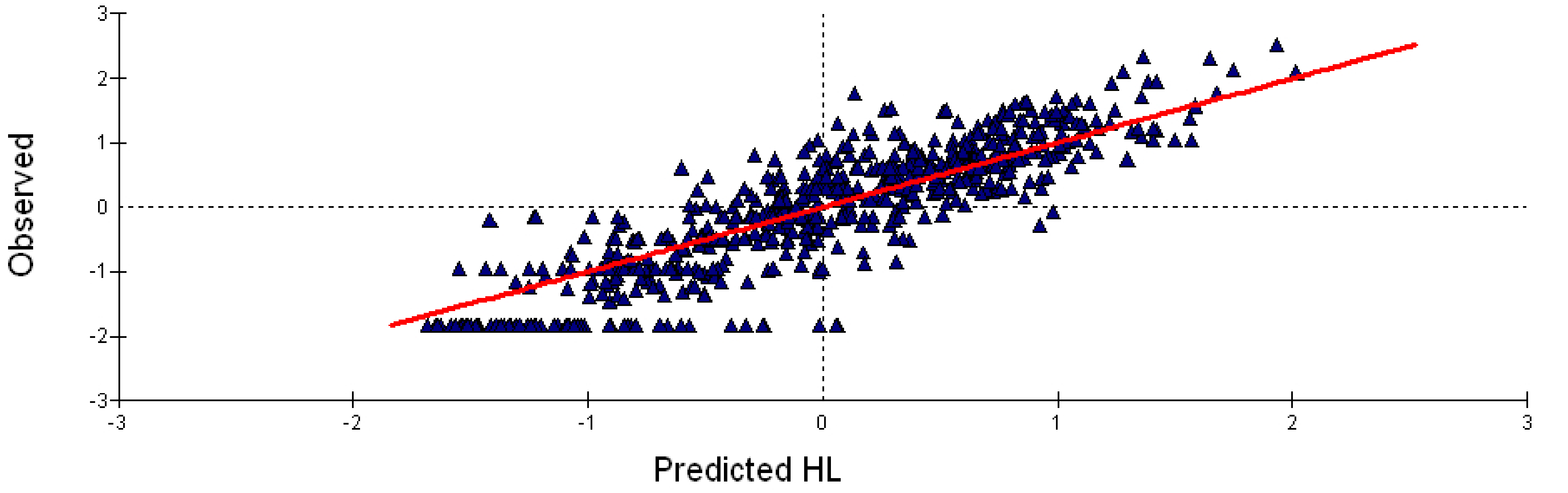
OLS model

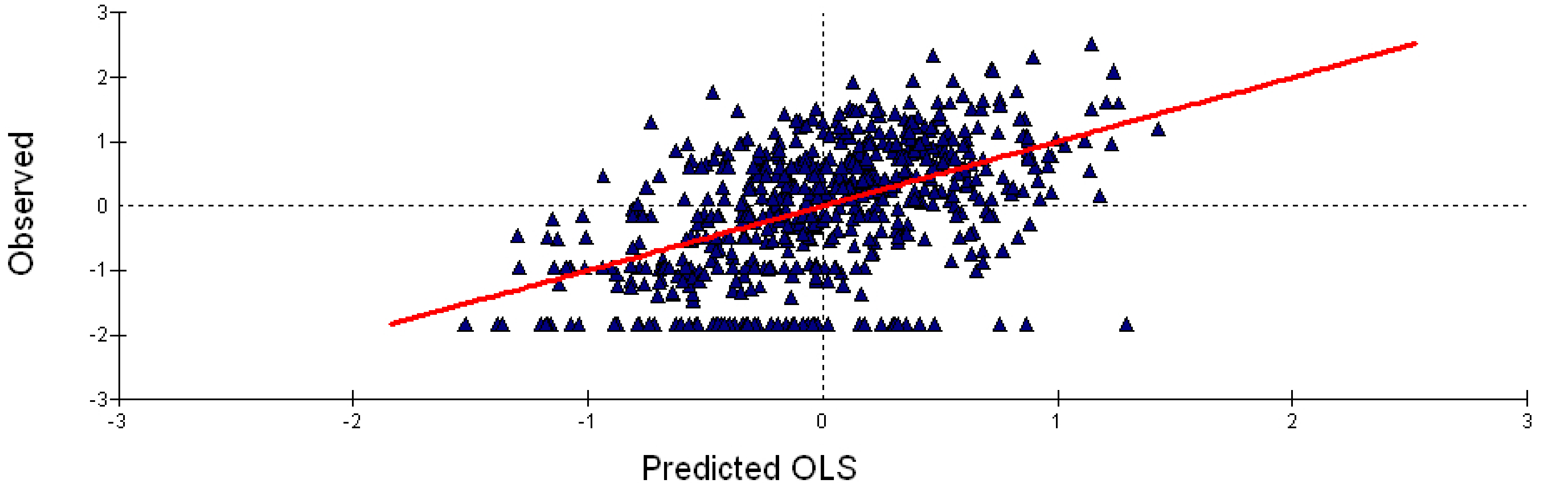
$$\text{trout density}_{ij} \sim N(XB, \Omega)$$

$$\text{trout density}_{ij} = \beta_{0i} + -0.308(0.036)\text{current velocity}_{ij} + 0.224(0.036)\text{woody debris}_{ij} + 0.133(0.036)\text{cover}_{ij} + -0.189(0.037)\text{sub-basin area}_{j} + -0.179(0.040)\text{height at flood}_{j} + -0.137(0.036)\text{current.sub-basin}_{ij}$$

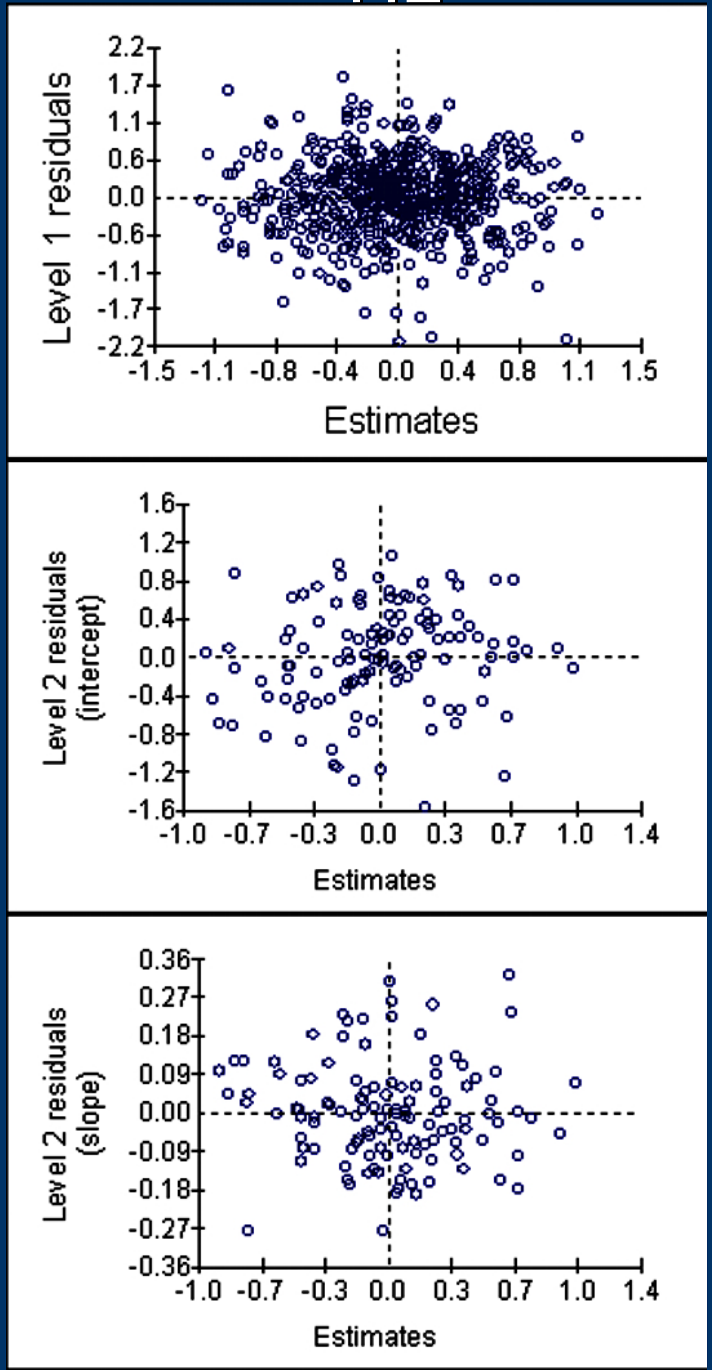
$$\beta_{0i} = 0.015(0.035) + e_{0ij}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.736(0.042) \end{bmatrix}$$

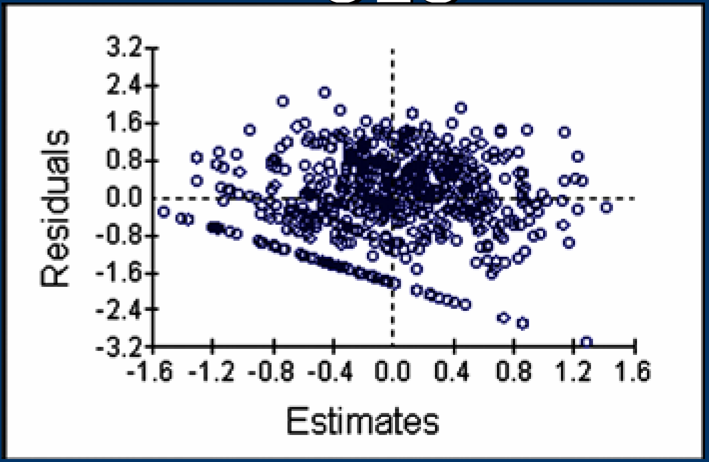

Observed
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Predicted HL


Observed
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Predicted OLS

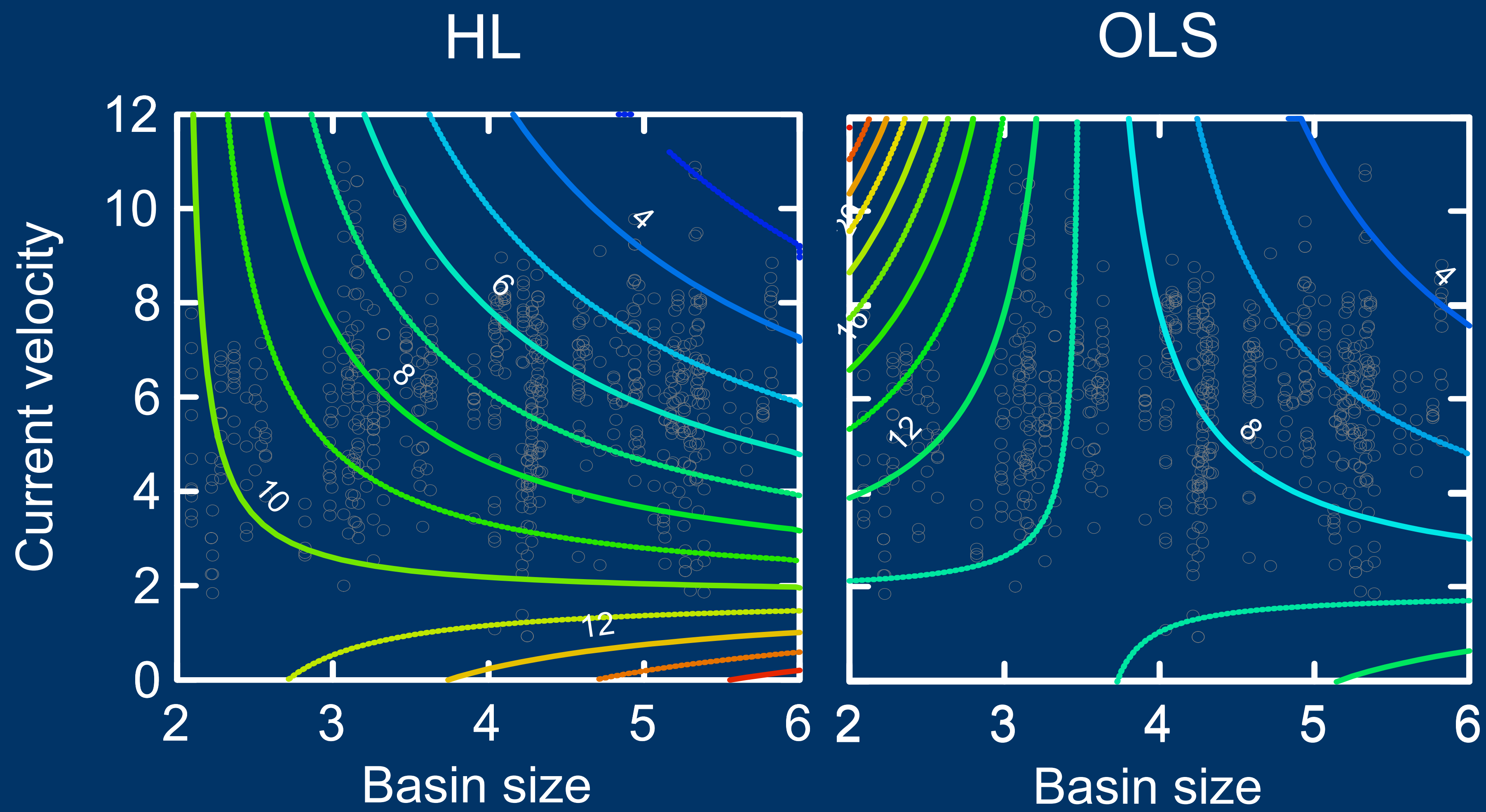
HL


Level 1 residuals
Estimates
Level 2 residuals (intercept)
Estimates
Level 2 residuals (slope)
Estimates

OLS


Residuals
Estimates

Effect of Current x Basin interaction on brook trout density



Conclusions

- Decomposition of variation showed that brook trout density
  - varied more across reaches than across sections within reaches
  - did not vary significantly across streams
- Brook trout density was related to both section- and reach-level features

- One significant cross-level interaction in the HL model:
  - The influence of current velocity on brook trout density depended on basin size (a contextual effect)
    - In larger basins, density declined with increasing current velocity
    - In smaller basins, density varied little with current velocity
- In contrast, the OLS model indicated that brook trout density declined with current velocity in larger basins but increased with current velocity in smaller basins

Ignoring the hierarchical structure of data may result in erroneous statistical and ecological conclusions

Acknowledgments

Fonds de recherche sur la nature et les technologies Québec

NSERC CRSNG

CIRSA
CENTRE INTERUNIVERSITAIRE DE RECHERCHE SUR LE SAUMON ATLANTIQUE

Société Cascapédia, inc.

Julie Deschênes
Katrine Turgeon
Véronique Bérard
Karen Brassard
Claudia Cossette
Marie-Josée Gagnon
Catherine Sauvé
Émilie Racicot