Describing the habitat of Atlantic salmon parr (*Salmo salar*) using the common spatial dependence between fish abundance and environmental variables: validation and application of a new multi-scale method

Guillaume Guénard – guillaume.guenard@umontreal.ca

Pierre Legendre

Daniel Boisclair

Département de sciences biologiques

Université de Montréal

Atlantic salmon

- Anadromous migratory
- Juveniles (parr) are stream dwelling
- Most wild stocks are at risk in North America

Atlantic salmon

- Anadromous migratory
- Juveniles (parr) are stream dwelling
- Most wild stocks are at risk in North America

Deterioration of habitat (spawning and feeding)

2 Objectives

1. Develop a method to integrate spatial information about environmental descriptors to improve habitat models

2. Build models that describe the habitat used by Atlantic salmon parr

Modeling fish abundance using multiple linear regression

$$\text{Abundance} = \text{Descriptors} \cdot b$$

Observations:
Sampling in space

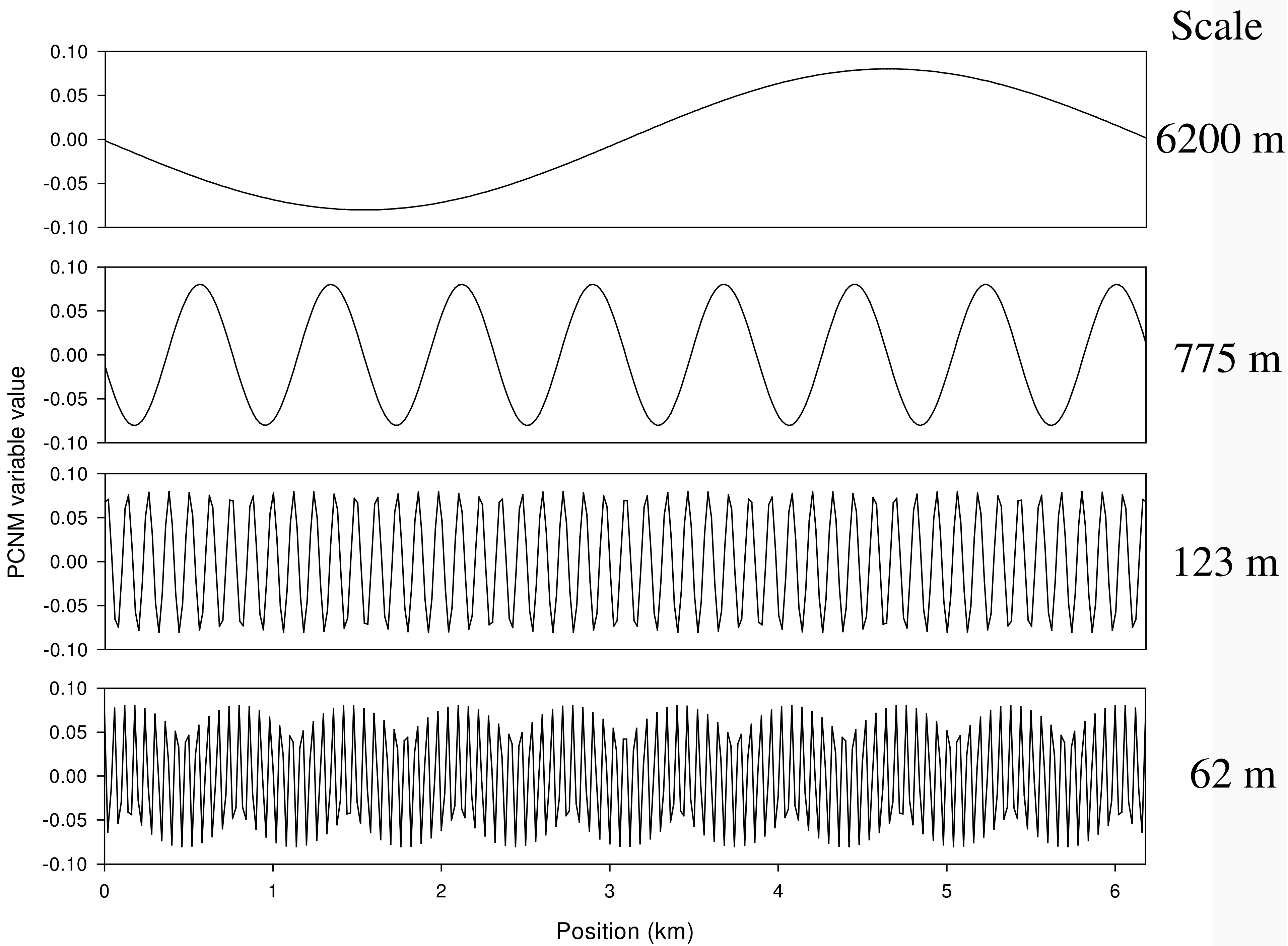
Availability of information: GPS location systems, etc.

What about using this information for modeling purposes?

PCNM variables

1. Truncated matrix of distances among sites
2. Correct the matrix (Euclidean)
3. Number of PCNM variables = $n - 1$
4. Sine-shaped spatial variables
5. Orthogonal to one another
6. Each PCNM variable has a mean of 0

$$\lambda_i \approx \frac{2Ln}{(i+1)\cdot(n-1)}$$



Scale
6200 m
775 m
123 m
62 m
PCNM variable value
0.10
0.05
0.00
-0.05
-0.10
0
1
2
3
4
5
6
Position (km)

Basic idea

If both the fish abundance and a habitat descriptor show high correlation to a given PCNM variable:

1. They are likely to have a common spatial dependence at the scale described by this PCNM variable.

2. This descriptor is suitable for predicting fish abundance **at this** spatial scale.

Method

- Least-square codependence analysis with respect to PCNM variables (matrix **W**)

$$\mathbf{C}_{y,x,\mathbf{W}} = \frac{\mathbf{W}'[y_i - \bar{y}] \circ \mathbf{W}'[x_i - \bar{x}]}{\sqrt{[y_i - \bar{y}]'[y_i - \bar{y}] \cdot [x_i - \bar{x}]'[x_i - \bar{x}]}}$$

$\mathbf{C}_{y,x,\mathbf{W}}$ is the vector of scale-dependant correlations between the fish abundances (*y*) and the habitat descriptor (*x*).

○ : Hadmard product

Method

- Calculate a pivotal statistic to test the significance of each coefficient of vector **C** using permutations

$$t^2_{y,x_k,\mathbf{w}_s} = (n-p-1)\frac{\mathbf{w}_s'[y_i-\bar{y}]\circ\mathbf{w}_s'[x_{i,k}-\bar{x}_k]}{\sqrt{\sum_i\left([y_i-\bar{y}]-\mathbf{w}_s'[y_i-\bar{y}]\right)^2\cdot\sum_i\left([x_{i,k}-\bar{x}_k]-\mathbf{w}_s'[x_{i,k}-\bar{x}_k]\right)^2}}$$

Permutation testing

Two permutation approaches are possible

- Permute *y* and *x* together (pairwise) with respect to **W**.

H_0: variables *y* and *x* are correlated but their correlation is not spatially dependent.

- Permute y and x independently.

H_0: variables *y* and *x* are neither correlated nor spatially dependent.

Corrections for multiple testing

1. Non-sequential Šidák correction of *p*-values

$$p' = 1 - (1 - p)^k$$

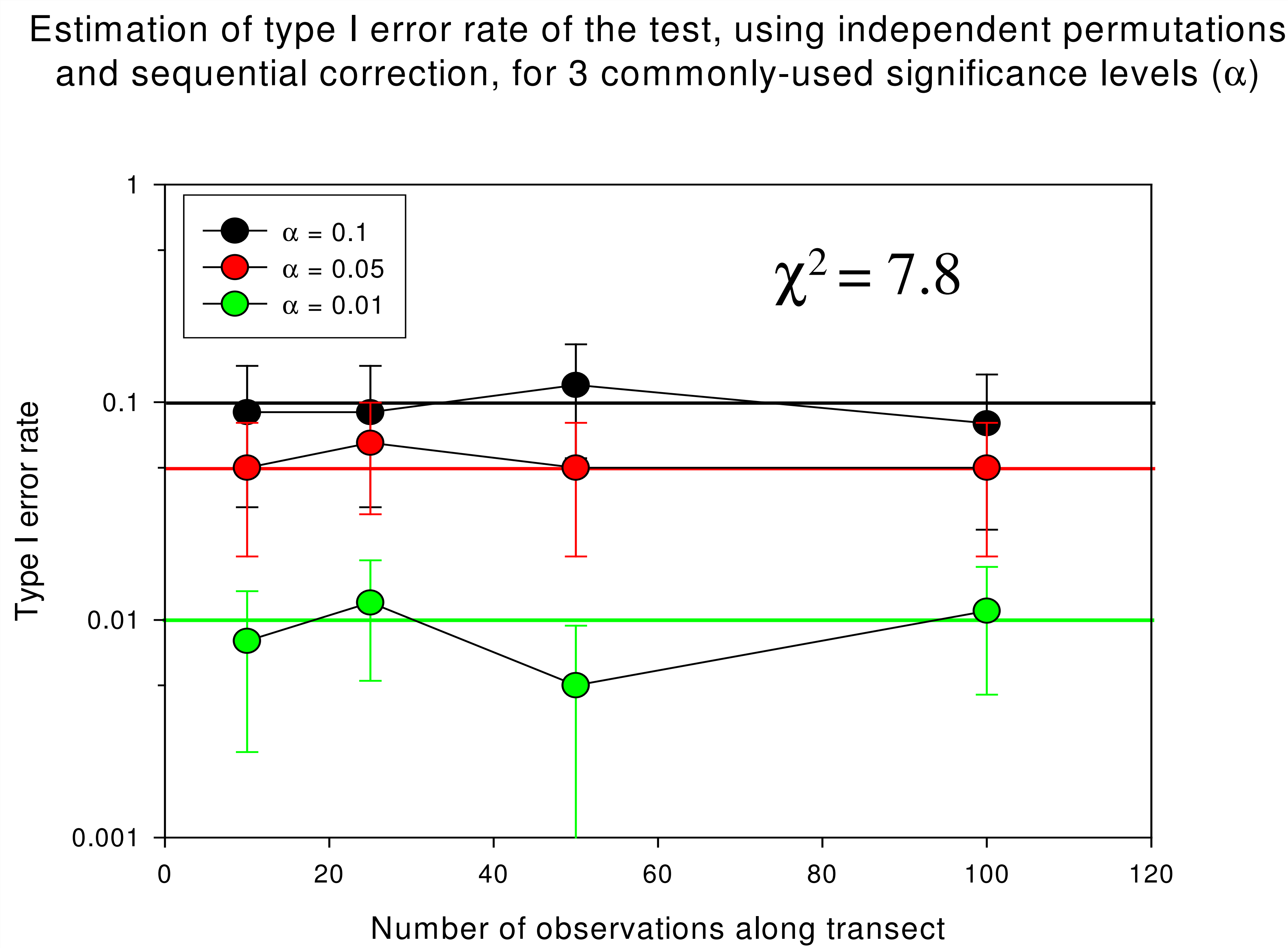
2. Sequential Šidák correction of *p*-values

$$p' = 1 - (1 - p)^{k - step - 1}$$

Stepwise testing procedure

1. Compute all $\mathbf{C}_{y,x,\mathbf{W}}$ coefficients.
2. Sort them in decreasing order of absolute values.
3. Test the first or subsequent coefficient(s) for significance.
4. Use a correction for multiple testing in the test of subsequent coefficients.
5. Repeat from step 3 as long as coefficients are significant. Stop when a non-significant coefficient is found.

Simulation results


Estimation of type I error rate of the test, using independent permutations and sequential correction, for 3 commonly-used significance levels (α)
α = 0.1
α = 0.05
α = 0.01
$\chi^2 = 7.8$
Type I error rate
1
0.1
0.01
0.001
0
20
40
60
80
100
120
Number of observations along transect

Results of all simulations using independant pairs of random variables

Permutation	Correction	χ^2_{11}
Pairwise	Non-sequential	17.7
Pairwise	Sequential	14.4
Independent	Non-sequential	16.9
Independent	Sequential	7.8

Results

Simulating type II error rate, using pairs of variables with know common spatial dependence…

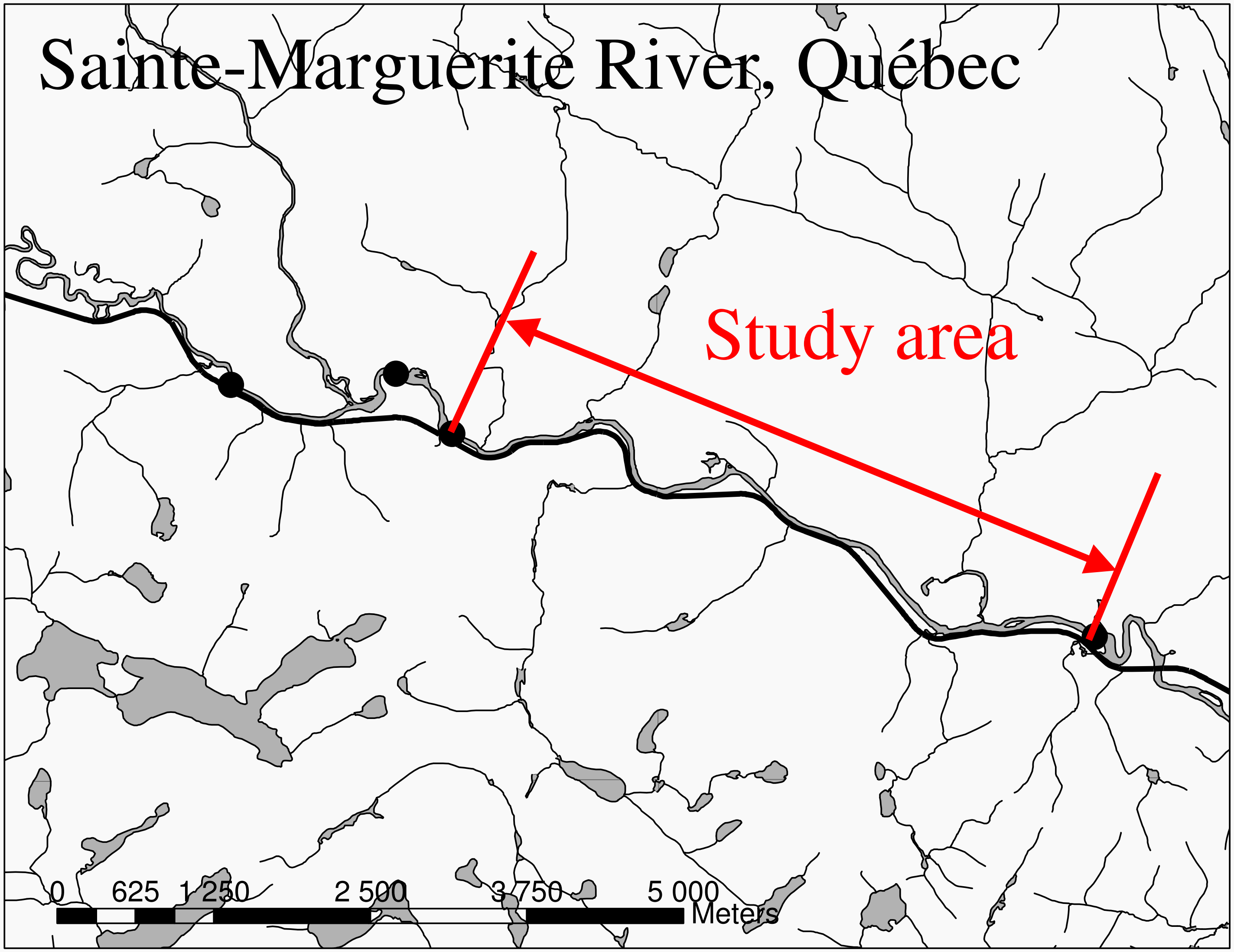
…remains to be done!

Predicting Salmon parr relative abundance

$$\hat{\mathbf{y}}_{s,k} = \mathbf{w}_s \left[\frac{\sqrt{[y_i - \bar{y}]'[y_i - \bar{y}] \cdot [x_{i,k} - \bar{x}_k]'[x_{i,k} - \bar{x}_k]} \cdot \mathbf{C}_{y, x_k, \mathbf{w}_s}}{\mathbf{w}_s '[x_{i,k} - \bar{x}_k]} \right]$$

$$\hat{\mathbf{y}}_k = \sum_s \hat{\mathbf{y}}_{s,k} \qquad \hat{\mathbf{y}} = \sum_k \hat{\mathbf{y}}_k$$

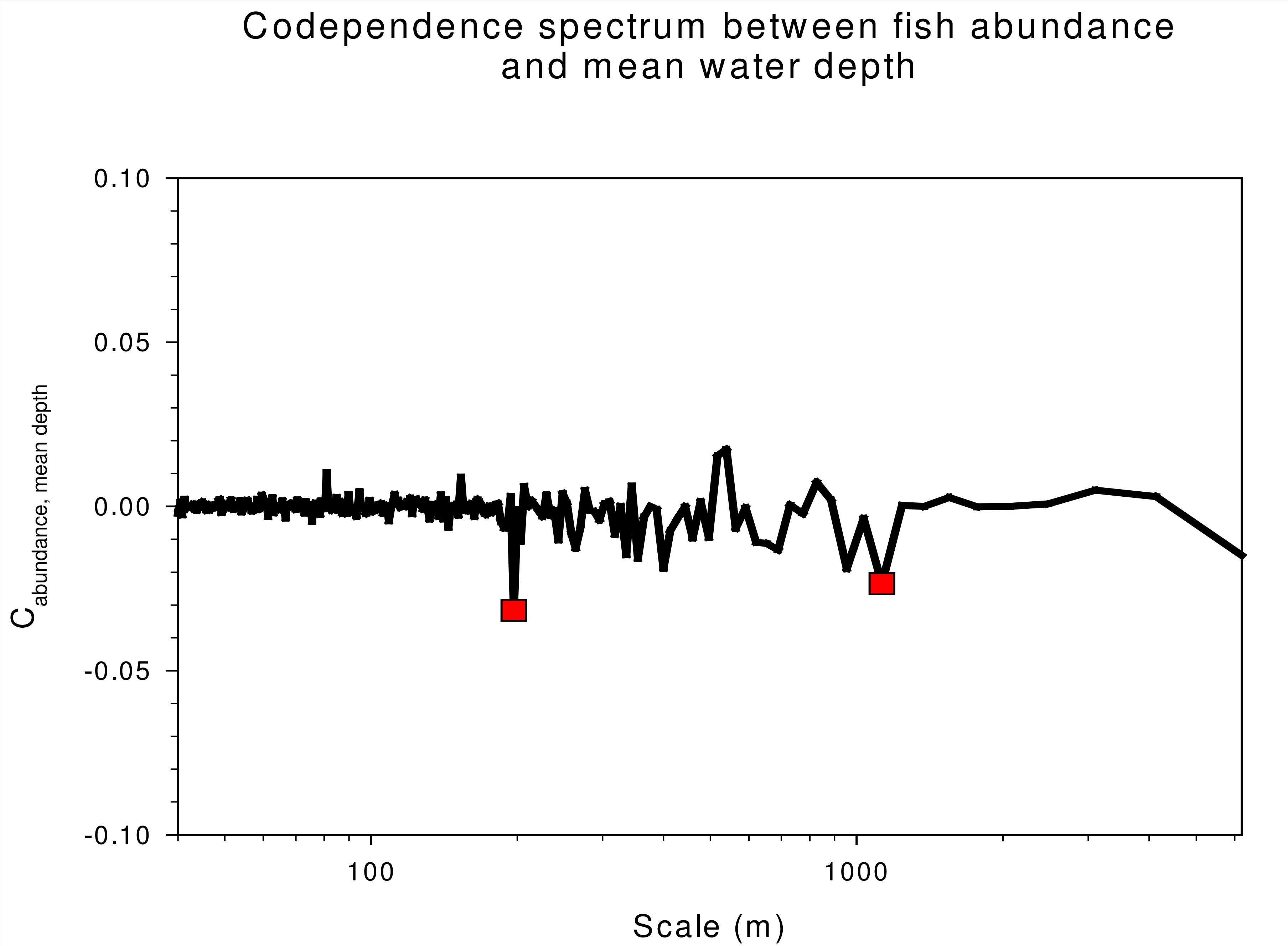
Application to habitat modeling

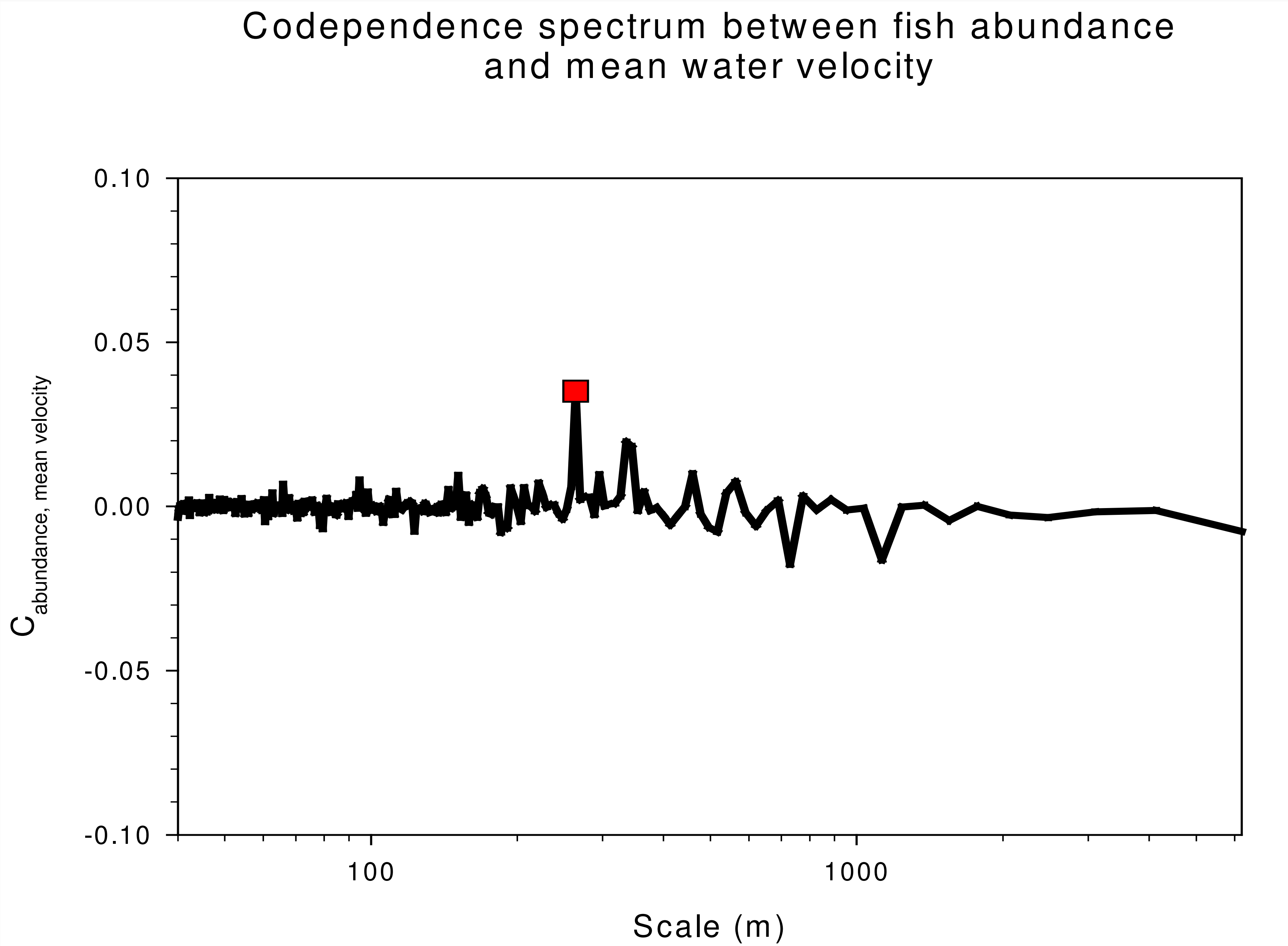


Sampling

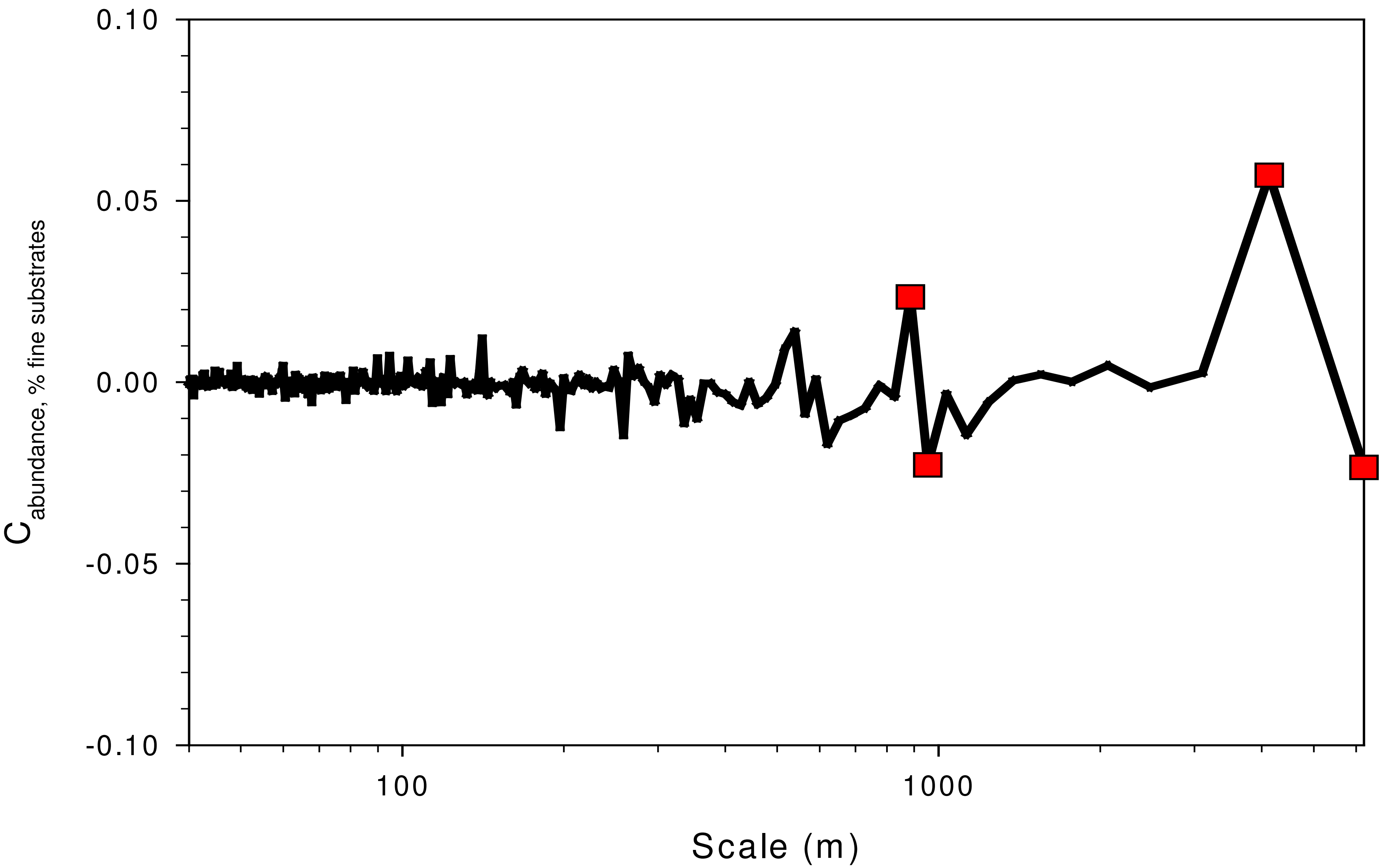
1. Salmon parr counted by snorkling
2. Mean water velocity and depth
3. Substrate composition: 6 grain size classes
4. Size of sampling unit in the river: 20 m

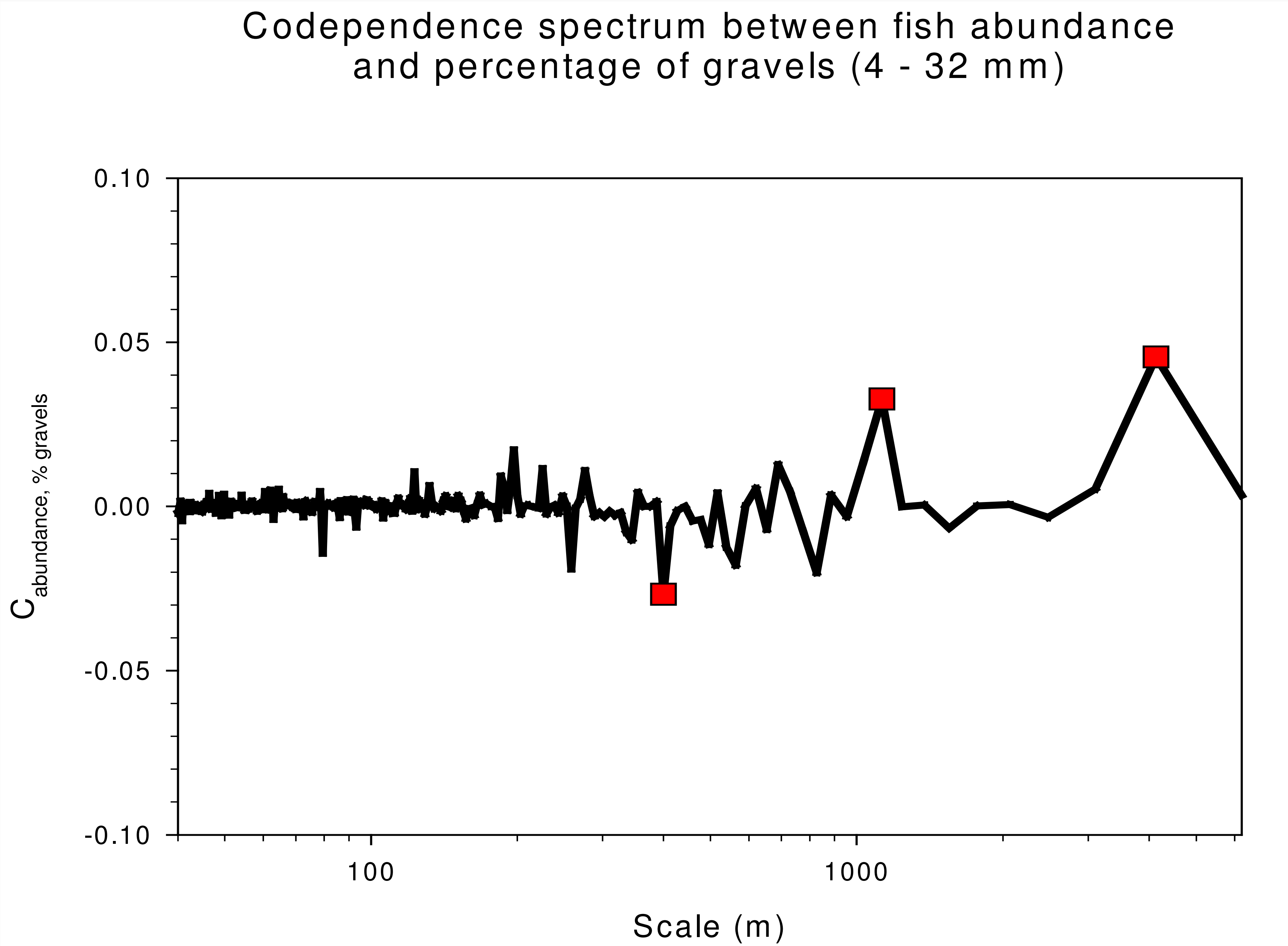
Results


Codependence spectrum between fish abundance
and mean water depth
0.10
0.05
0.00
-0.05
-0.10
C abundance, mean depth
100
1000
Scale (m)

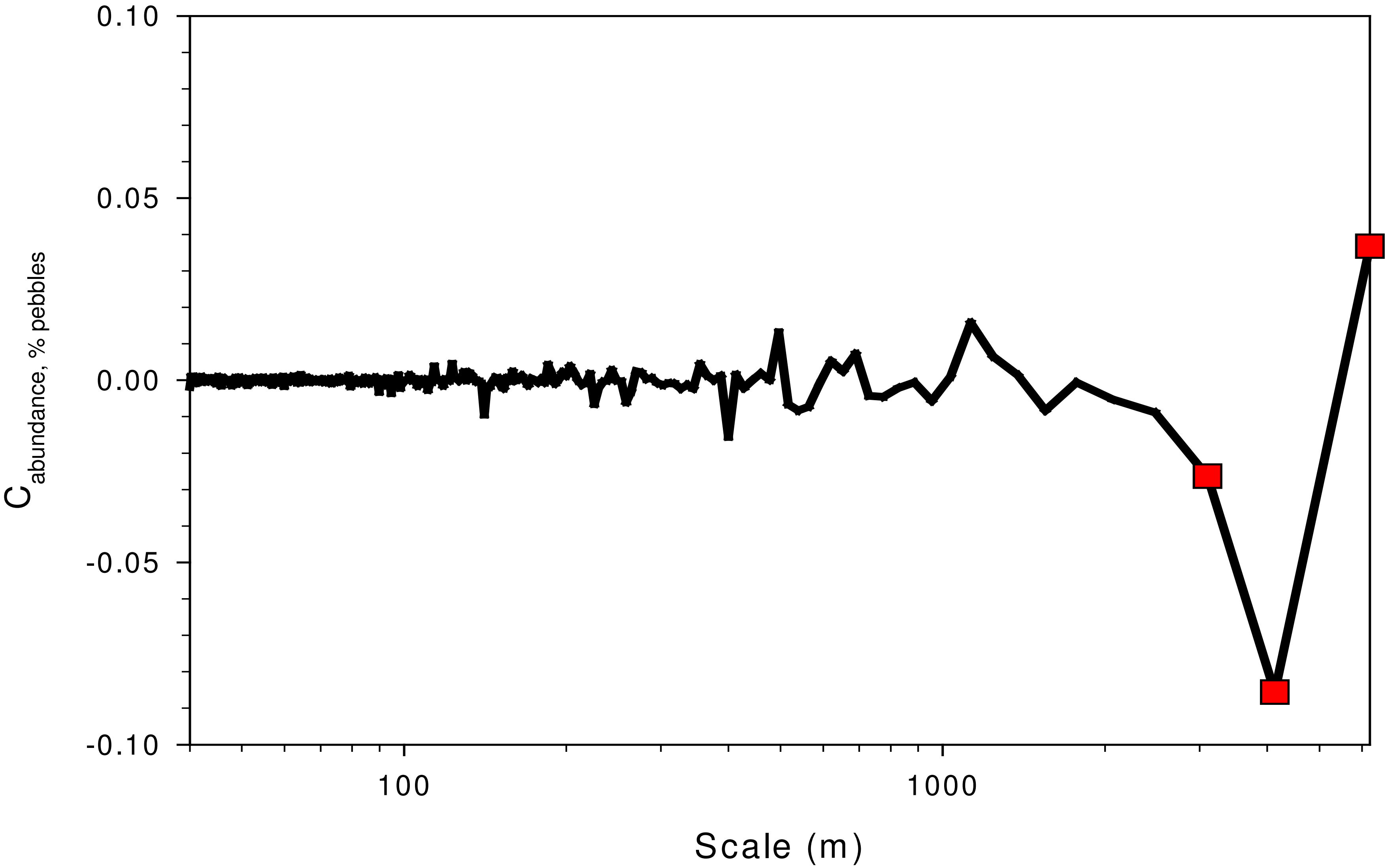

Codependence spectrum between fish abundance
and mean water velocity
0.10
0.05
0.00
-0.05
-0.10
$C_{abundance, mean velocity}$
100
1000
Scale (m)

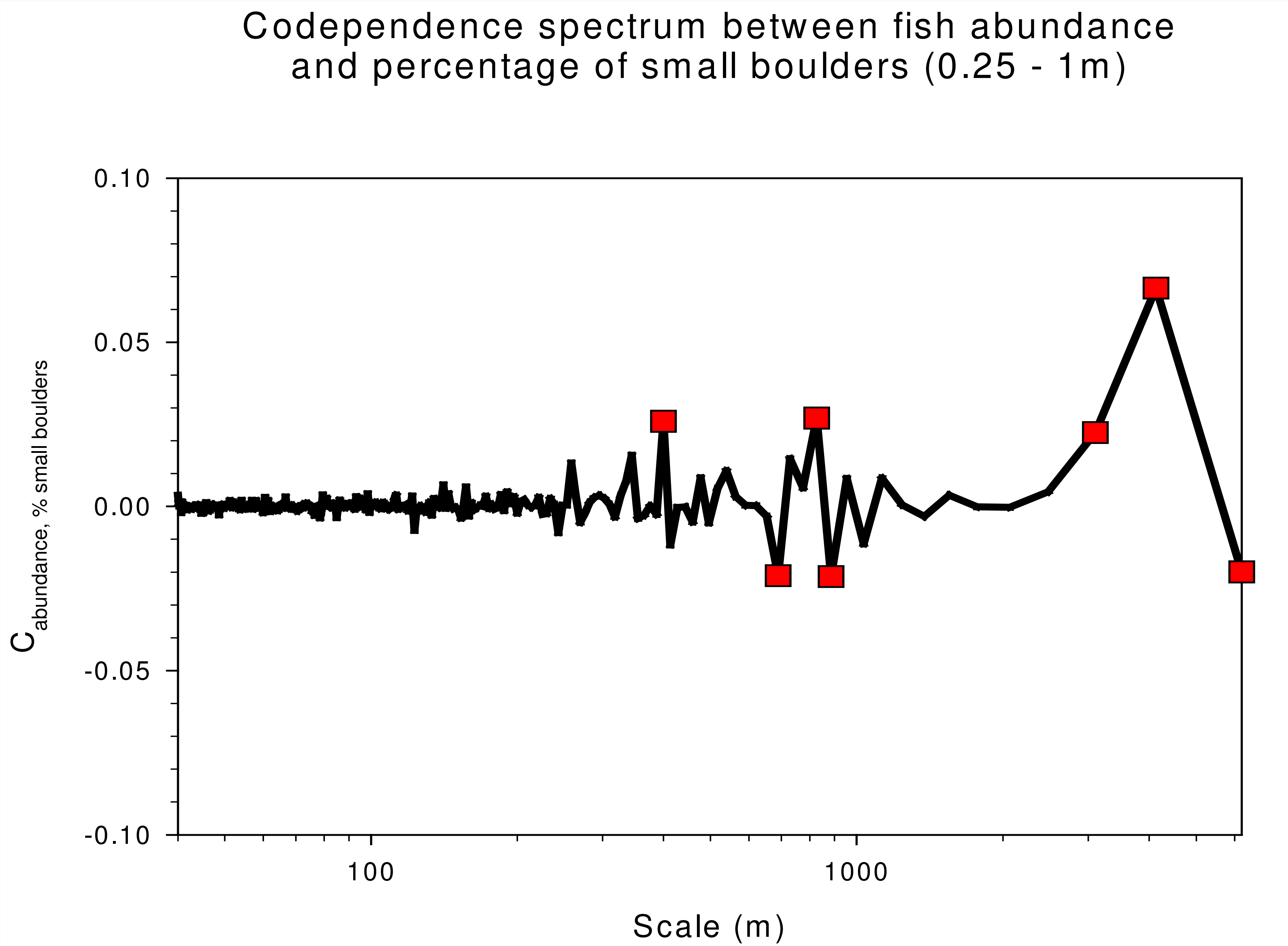
Codependence spectrum between fish abundance
and percentage of fine substrates (0.5 - 4 mm)




Codependence spectrum between fish abundance
and percentage of gravels (4 - 32 mm)
0.10
0.05
0.00
-0.05
-0.10
$C_{abundance, \% gravels}$
100
1000
Scale (m)

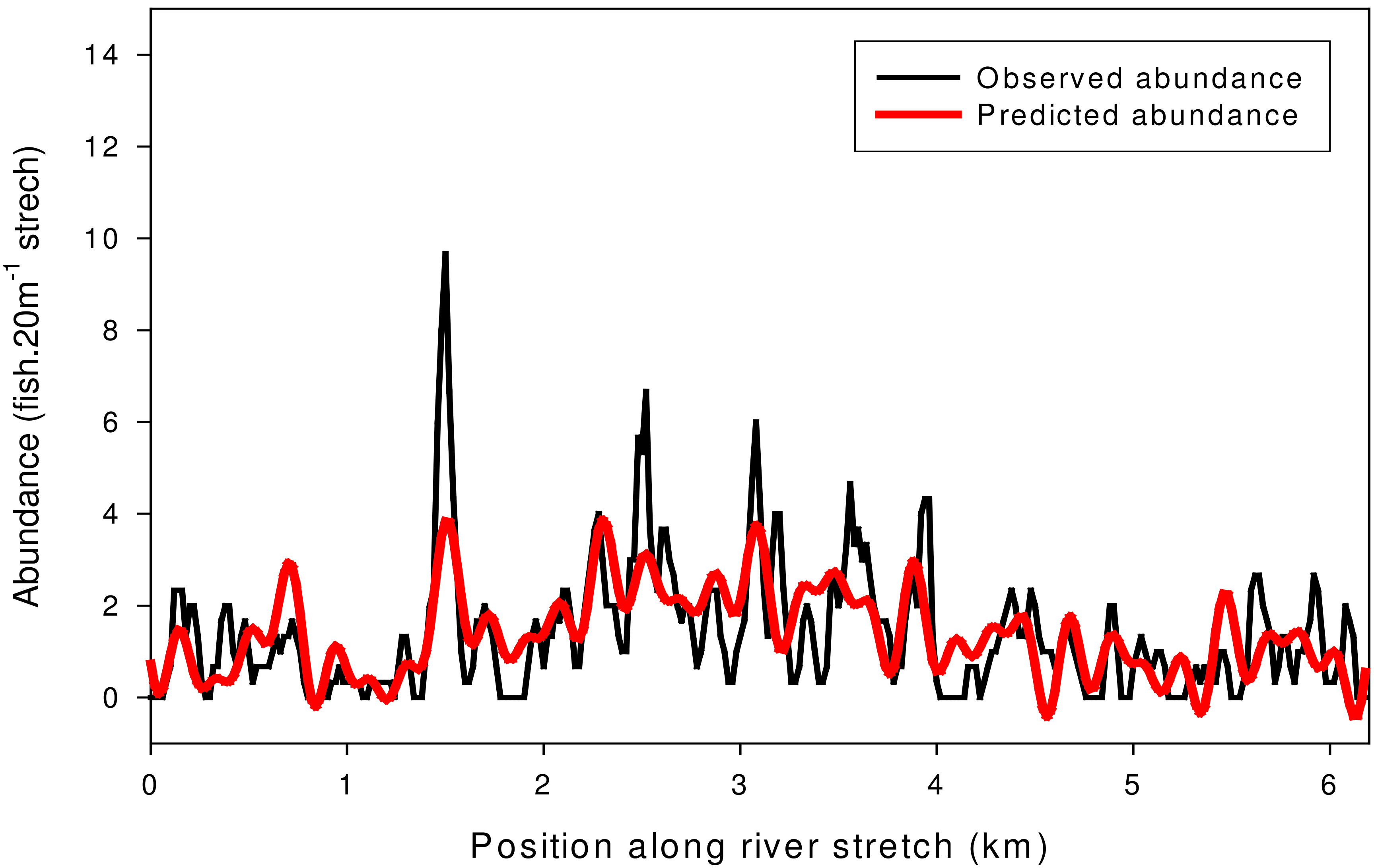
Codependence spectrum between fish abundance
and percentage of pebbles (32 - 64 mm)

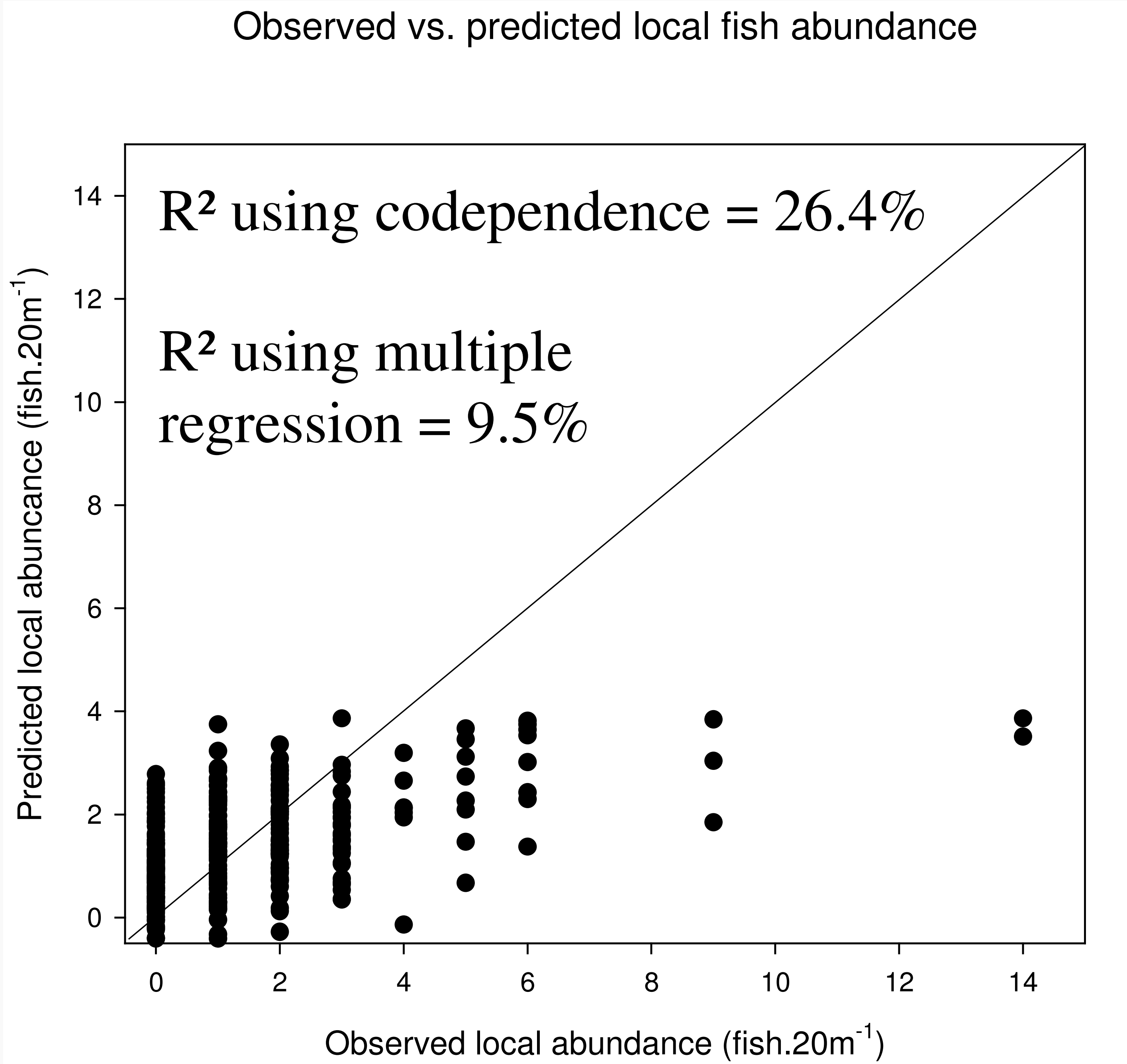




Codependence spectrum between fish abundance
and percentage of small boulders (0.25 - 1m)
0.10
0.05
0.00
-0.05
-0.10
100
1000
Scale (m)
C abundance, % small boulders

Observed and raw predicted fish abundance using mean water velocity, depth, substrate composition




Observed vs. predicted local fish abundance
R² using codependence = 26.4%
R² using multiple regression = 9.5%
Predicted local abuncance (fish.20m-1)
Observed local abundance (fish.20m-1)
0
2
4
6
8
10
12
14

Conclusions

1. The spatial distribution of an environmental descriptor also contains information.
2. This information can be used to improve prediction of Atlantic salmon parr distribution.
3. Further simulations have to be done to determine type II error rate and the statistical power of the method.

The End

Supplementary explanations

$$\mathbf{C}_{y,x,\mathbf{w}_j} = \frac{\sum_{i=1}^{n} \mathbf{w}_{i,j} \cdot (y_i - \bar{y}) \cdot \sum_{i=1}^{n} \mathbf{w}_{i,j} \cdot (x_i - \bar{x})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

$$\hat{\mathbf{y}}_{s,k} = \mathbf{w}_s \left[\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \cdot \sum_{i=1}^{n} (x_{i,k} - \bar{x}_k)^2} \cdot \left(\mathbf{C}_{y,x_k,\mathbf{w}_s} \circ \left(\mathbf{w}_s{}'[x_{i,k} - \bar{x}_k] \right)^{-1} \right) \right]$$

$$\hat{\mathbf{y}} = \sum_{s,k} \hat{\mathbf{y}}_{s,k}$$

Supplementary explanations

$$\mathbf{C}_{y,x,w_j} = \frac{\mathbf{w}_j'[y_i - \bar{y}] \circ \mathbf{w}_j'[x_i - \bar{x}]}{\sqrt{[y_i - \bar{y}]'[y_i - \bar{y}] \cdot [x_i - \bar{x}]'[x_i - \bar{x}]}}$$

$$\mathbf{b}_{s,k} = \mathbf{w}_s'[y_i - \bar{y}] = \frac{\sqrt{[y_i - \bar{y}]'[y_i - \bar{y}] \cdot [x_{i,k} - \bar{x}_k]'[x_{i,k} - \bar{x}_k]} \cdot \mathbf{C}_{y,x_k,\mathbf{w}_s}}{\mathbf{w}_s'[x_{i,k} - \bar{x}_k]}$$

$$\hat{\mathbf{y}}_{s,k} = \mathbf{w}_s \mathbf{b}_{s,k} = \mathbf{w}_s \left[\frac{\sqrt{[y_i - \bar{y}]'[y_i - \bar{y}] \cdot [x_{i,k} - \bar{x}_k]'[x_{i,k} - \bar{x}_k]} \cdot \mathbf{C}_{y,x_k,w_s}}{\mathbf{w}_s'[x_{i,k} - \bar{x}_k]} \right]$$

$$\hat{\mathbf{y}}_k = \sum_s \hat{\mathbf{y}}_{s,k} \qquad \hat{\mathbf{y}} = \sum_k \hat{\mathbf{y}}_k$$