Assessing relationships between ecological variables at multiple spatial scales under the linear model of coregionalization. Part 2: Testing aspects

Pierre Dutilleul and Bernard Pelletier
McGill University
Ste-Anne-de-Bellevue, Québec, Canada
pierre.dutilleul@mcgill.ca
bernard.pelletier@mcgill.ca

Outline

Reminders

. About the linear model of coregionalization

. About the statistical problem of autocorrelation in correlation analyses with spatial data

Objectives

"Data of the problem"

New results

. Tests for simple and multiple correlations per structure

. Simulation results

. Illustration with real data

Conclusions

Reminder about the Linear Model of Coregionalization (LMC)

A number of spatially sampled variables $Z_j(\mathbf{u})$ are decomposed into

. a deterministic, mean component $m_j(\mathbf{u})$ (representing the large-scale variation) plus

. a random, zero-mean component $R_j(\mathbf{u})$ (representing the variations at other scales):

$$Z_j(\mathbf{u}) = m_j(\mathbf{u}) + R_j(\mathbf{u}).$$

Reminder about the Linear Model of Coregionalization (LMC) (continued)

Under the LMC, each $R_j(\mathbf{u})$ is decomposed into:

. a nugget effect $R_{j1}(\mathbf{u})$ (representing the non-spatial and micro-scale variations);

. a short-range autocorrelated, spatial structure $R_{j2}(\mathbf{u})$ (i.e., the small-scale variation)

. a long-range autocorrelated, spatial structure $R_{j3}(\mathbf{u})$ (i.e., the variation at intermediate scale):

$$R_j(\mathbf{u}) = \Sigma_{s=1\ldots3} R_{js}(\mathbf{u}).$$

Objectives

. Testing the significance of the simple correlation between variable j and variable j' at structure s (s = 1, 2, 3) – parameters of interest: $Corr(R_{js}, R_{j's})$

. Testing the significance of the multiple correlation between variable j and q other variables (j', j'', etc.) at structure s

“Data of the problem”

. R_{js} (j = 1, …, q+1) are not observable.

Note: Only $R_j(\mathbf{u}) = \Sigma_{s=1\ldots3} R_{js}(\mathbf{u})$ is observable if $m_j(\mathbf{u})$ is known or $m_j(\mathbf{u}) = 0$ for all **u**.

. If $m_j(\mathbf{u})$ is unknown , $R_j(\mathbf{u}) \approx$ estimated $R_j(\mathbf{u}) = Z_j(\mathbf{u})$ – estimated $m_j(\mathbf{u})$ (see Part 1: Estimation aspects); same for $R_{j'}(\mathbf{u})$.

. Structural correlations are evaluated from sill matrix estimates (see Part 1).

. R_{j2}, $R_{j'2}$, $R_{j''2}$, etc. and R_{j3}, $R_{j'3}$, $R_{j''3}$, etc. are autocorrelated.

Reminder about the statistical problem of autocorrelation in correlation analyses with spatial data

. In the global, simple linear correlation analysis between spatially sampled variables Z_1 and Z_2, the parameter of interest is Corr(Z_1, Z_2).

. Let r denote Pearson's estimator (commonly called 'sample correlation coefficient') evaluated from samples of size N collected for Z_1 and Z_2.

. In general terms, the presence of spatial autocorrelation in the sample data of Z_1 and Z_2 induces a bias in Var(r).

Reminder about the statistical problem of autocorrelation in correlation analyses with spatial data (continued)

- It follows that the classical t test with N – 2 df built on r/sqrt{(1 – r^2)/(N – 2)} is not valid – it rejects the null hypothesis more often than expected theoretically.
- A solution to this problem is provided by the modified t test with M – 2 df built on the same correlation estimator but with N replaced by

$$M = 1 + Var(r)^{-1},$$

which is the effective sample size that takes the spatial autocorrelation of Z_1 and Z_2 into account (Dutilleul, 1993, *Biometrics*).

New results

- From the variance of the simple structural correlation estimator r_s (to be evaluated from the corresponding GLS sill matrix estimate; see Part 1), it is possible to calculate the effective sample size

$$M_s = 1 + Var(r_s)^{-1} \quad (s = 1, 2, 3),$$

which incorporates the spatial autocorrelation at structure s when present and the uncertainty associated with the GLS sill matrix estimate. This provides modified t tests per structure with $M_s - 2$ df.

- Extension 1: Modified F tests for assessing the significance of multiple structural correlations (i.e., in a regionalized multiple regression with q random explanatory variables)
- Extension 2: Modified F tests per structure in a regionalized RDA
- Repetition of tests: Bonferroni correction

Simulation results

. Descriptive statistics on effective sample sizes

. Validity and power analyses for modified F tests per structure

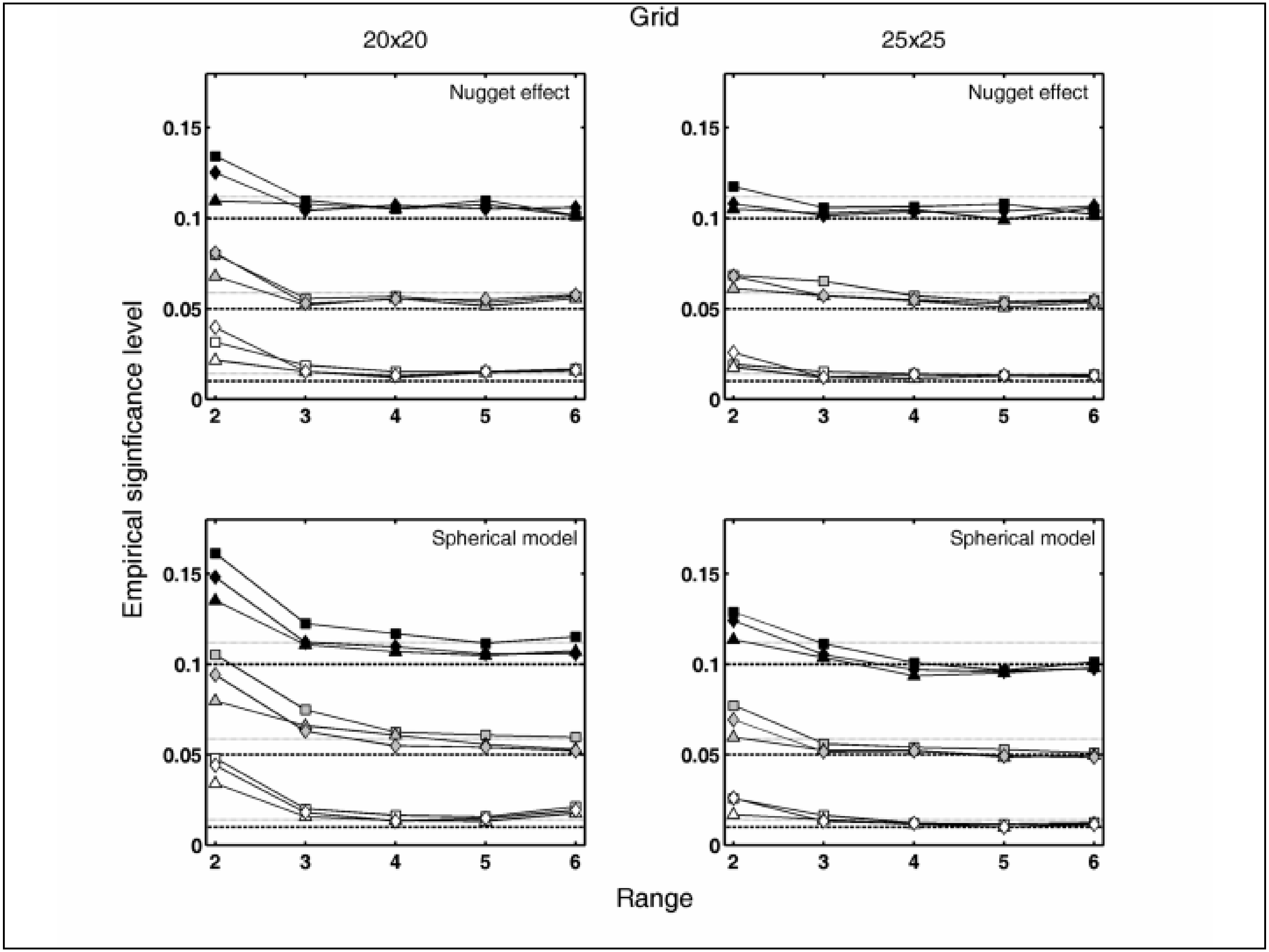
. Simulation parameter values: grid sizes of 20x20 and 25x25, two structures (i.e., s = 1, 2), and values of 2, 3, 4, 5, 6 for the range of the spherical structure.

Descriptive statistics on effective sample sizes for the modified F tests per structure

	•20x20 Grid						•25x25 Grid				
•Range (spherical)	•2	•3	•4	•5	•6		•2	•3	•4	•5	•6
	•Nugget effect										
•Model-based, •estimated range											
•Mean	•44.6	•95.2	•125.9	•146.5	•163.3		•67.0	•151.8	•201.3	•233.0	•258.2
•Std	•33.6	•24.6	•22.9	•23.4	•24.3		•43.1	•30.4	•28.6	•29.4	•30.0
•Median	•37.4	•93.9	•124.9	•146.0	•163.2		•58.7	•150.5	•200.7	•232.9	•257.9
•Min	•1.0	•13.2	•54.6	•68.6	•76.4		•1.0	•29.5	•119.3	•144.3	•159.7
•Max	•223.3	•201.0	•219.2	•256.3	•276.1		•259.6	•262.2	•312.3	•337.1	•352.1

Descriptive statistics on effective sample sizes for the modified F tests per structure (continued)

	•20x20 Grid						•25x25 Grid				
•Range (spherical)	•2	•3	•4	•5	•6		•2	•3	•4	•5	•6
	•Spherical structure										
•Model-based, •estimated range											
•Mean	•27.5	•36.4	•33.6	•29.8	•26.8		•41.5	•57.3	•52.7	•46.36	•41.5
•Std	•13.1	•6.6	•4.6	•3.8	•3.5		•15.1	•8.0	•5.6	•4.62	•4.2
•Median	•25.5	•36.1	•33.5	•29.8	•26.7		•40.0	•57.1	•52.6	•46.28	•41.4
•Min	•1.1	•15.1	•15.8	•13.2	•11.5		•1.3	•32.2	•34.2	•31.36	•28.1
•Max	•101.7	•68.0	•51.6	•44.0	•41.7		•133.8	•88.5	•70.8	•61.13	•55.5



Grid
20x20
25x25
Nugget effect
Spherical model
Empirical significance level
Range
0
0.05
0.1
0.15
2
3
4
5
6

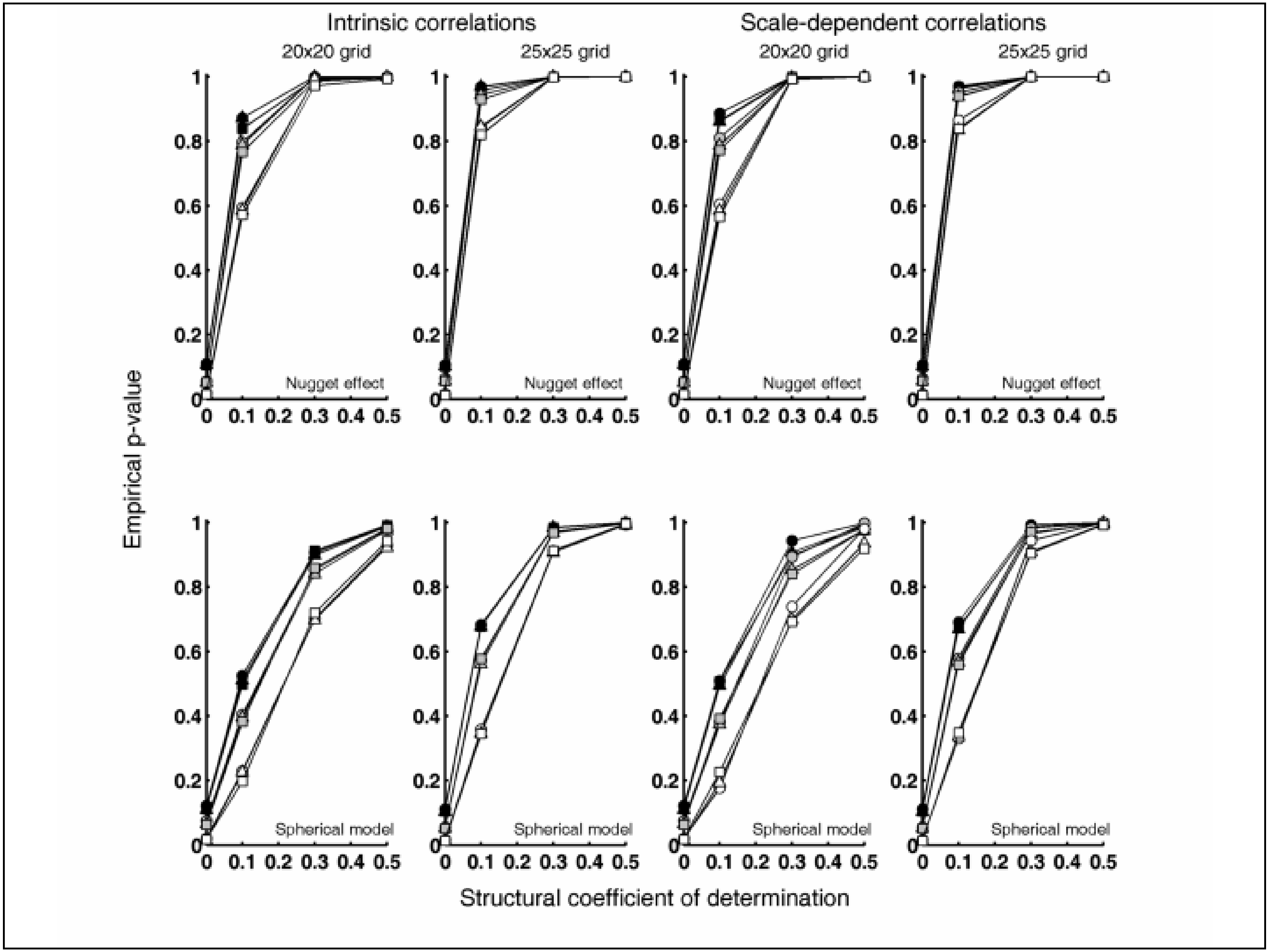

Intrinsic correlations
Scale-dependent correlations
20x20 grid
25x25 grid
20x20 grid
25x25 grid
Nugget effect
Spherical model
Empirical p-value
Structural coefficient of determination
0
0.1
0.2
0.3
0.4
0.5
0.6
0.8
1

Illustration with real data

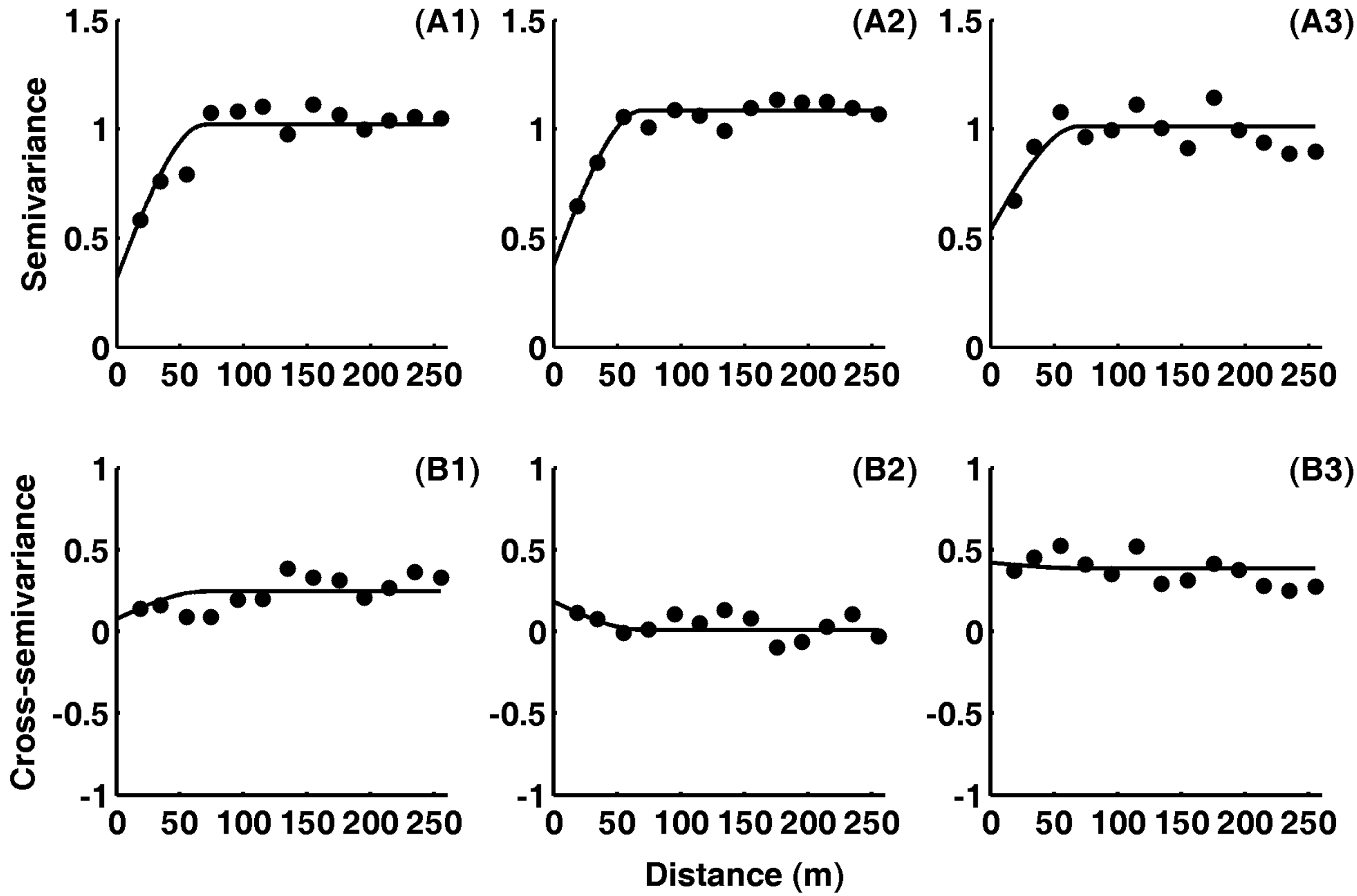
. Example of regionalized multiple regression with two random explanatory variables (soil variables; q = 2) – a plant biomass trait is the dependent variable

. Two structures (i.e., s = 1, 2) in addition to drifts

. N = 152

. Direct and cross experimental variograms; see figure on the next slide

. Values of structural coefficients of determination: 0.62 (micro-scale, P = 0.02, estimated M_1 = 11.0); 0.14 (small to intermediate scale, P = 0.39, estimated M_2 = 16.1, estimated range: 70)



Semivariance
Cross-semivariance
(A1)
(A2)
(A3)
(B1)
(B2)
(B3)
0
0.5
1
1.5
-1
-0.5
0
50
100
150
200
250
Distance (m)

Conclusions

. Modified t and F tests developed under the LMC are available for assessing structural correlations between spatially sampled variables at micro-, small and intermediate scales.

. The association (pseudo-correlation) between drift estimates at large scale can be tested with the Mantel statistic (Dutilleul et al., 2000, JABES).

. The effective sample sizes M_s (s = 1, 2, 3) can be used to build confidence intervals for structural coefficients of determination, whether unadjusted or adjusted.

Bibliography of the global modified t test

. Dutilleul, P. 1993. Modifying the t test for assessing the correlation between two spatial processes. *Biometrics* 49:305-314.

. Dutilleul, P. and Pinel-Alloul, B. 1996. A doubly multivariate model for statistical analysis of spatio-temporal environmental data. *Environmetrics* 7:551-566.

. Dutilleul, P., Herman, M., and Avella-Shaw, T. 1998. Growth rate effects on correlations among ring width, wood density and mean tracheid length in Norway spruce (*Picea abies* (L.) Karst). *Canadian Journal of Forest Research* 28:56-68.

. Alpargu, G. and Dutilleul, P. 2003. To be or not to be valid in testing the significance of the slope in simple quantitative linear models with autocorrelated errors. *Journal of Statistical Computation and Simulation* 73:165-180.

. Alpargu, G. and Dutilleul, P. Stepwise regression in mixed quantitative linear models with autocorrelated errors. *Communications in Statistics*, in press.

Bibliography of the tests per structure and related work

. Pelletier, B., Dutilleul, P., Larocque, G., and Fyles, J. W. 2004. Fitting the linear model of coregionalization by generalized least squares. *Mathematical Geology* 36:323-343.

. Larocque, G., Dutilleul, P., Pelletier, B., and Fyles, J. W. Conditional Gaussian co-simulation of regionalized components of soil variation. *Geoderma*, in press.

. Larocque, G., Dutilleul, P., Pelletier, B., and Fyles, J. W. Characterization and quantification of uncertainty in coregionalization analysis. In revision.

. Dutilleul, P., Pelletier, B., and Alpargu, G. Modified F tests for assessing the multiple correlation between one spatial process and several others globally and per structure. Submitted.

. Pelletier, B., Dutilleul, P., Larocque, G., and Fyles, J. W. Multi-scale assessment of spatial relationships between ecological variables using coregionalization analysis in the presence of drifts. In preparation.