Assessing Relationships Between Ecological Variables at Multiple Spatial Scales under the Linear Model of Coregionalization

Part 1: Estimation Aspects

Bernard Pelletier, Pierre Dutilleul, Guillaume Larocque and James W. Fyles
McGill University, Canada

90th Annual Meeting of the Ecological Society of America
Montréal, Québec
August 9, 2005

Conceptual Framework

Our method is based on the geostatistical model:

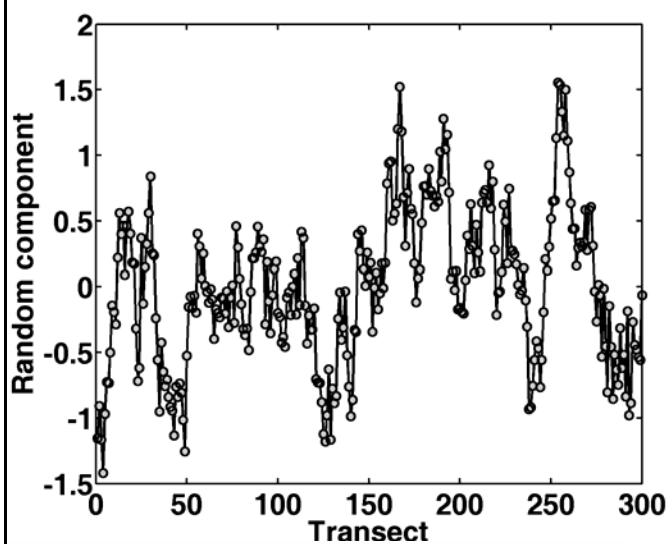
$$\mathbf{Z}(\mathbf{u}) = \mathbf{m}(\mathbf{u}) + \mathbf{R}(\mathbf{u}),$$

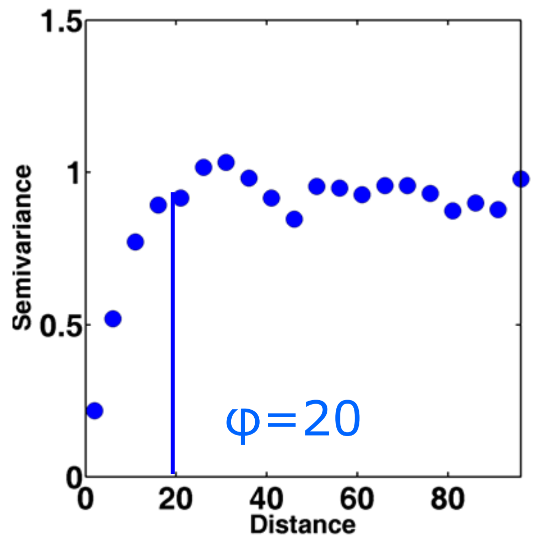
$\mathbf{Z}(\mathbf{u})$ the spatial variables $\mathbf{Z}$ at sampling locations $\mathbf{u}$

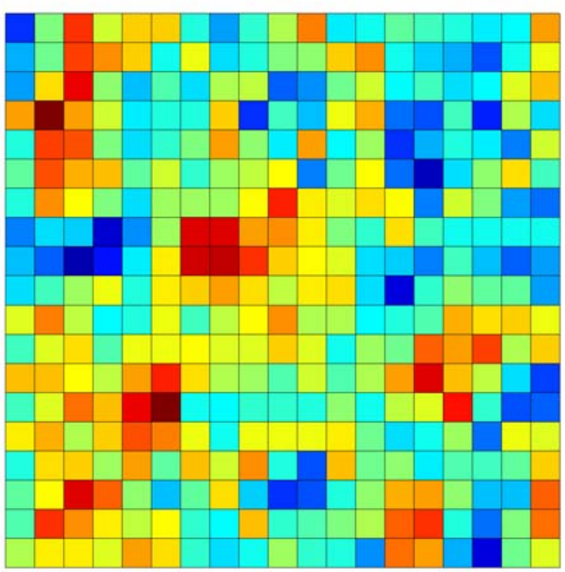
$\mathbf{m}(\mathbf{u})$ is the large-scale component (**deterministic**)

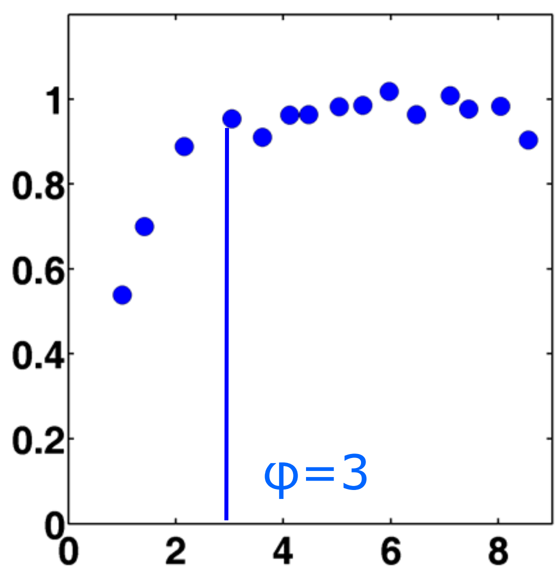
$\mathbf{R}(\mathbf{u})$ is the “small-scale” component (**random**)

Random component **R(u)**









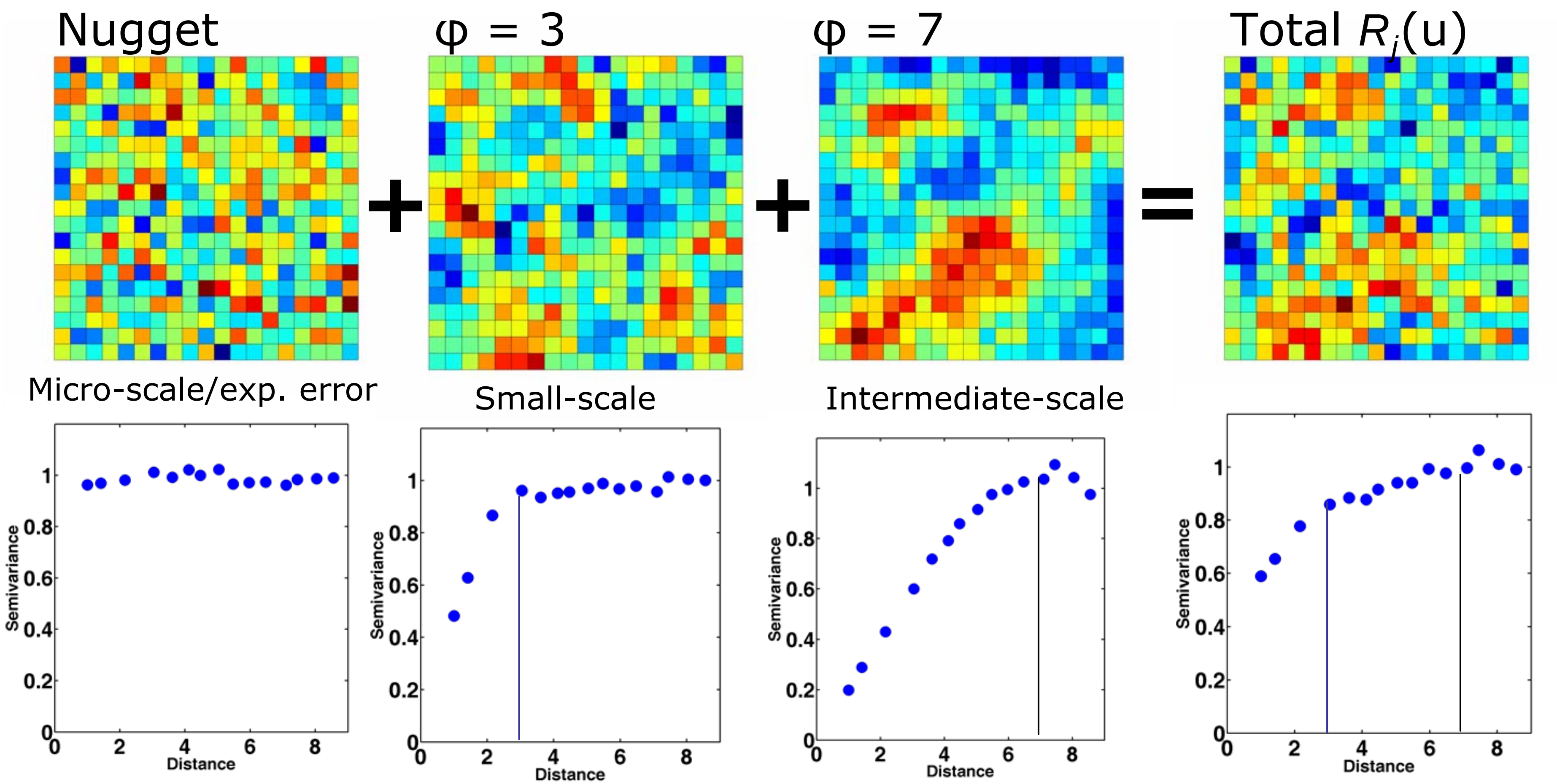
Probabilistic Approach: "Small-scale" patterns $R_j(\mathbf{u})$ viewed as the outcome of a spatial process with a given range of autocorrelation (φ)

Variogram is used to represent that spatial process

Deterministic Approach: Interested in the explicit description of patterns

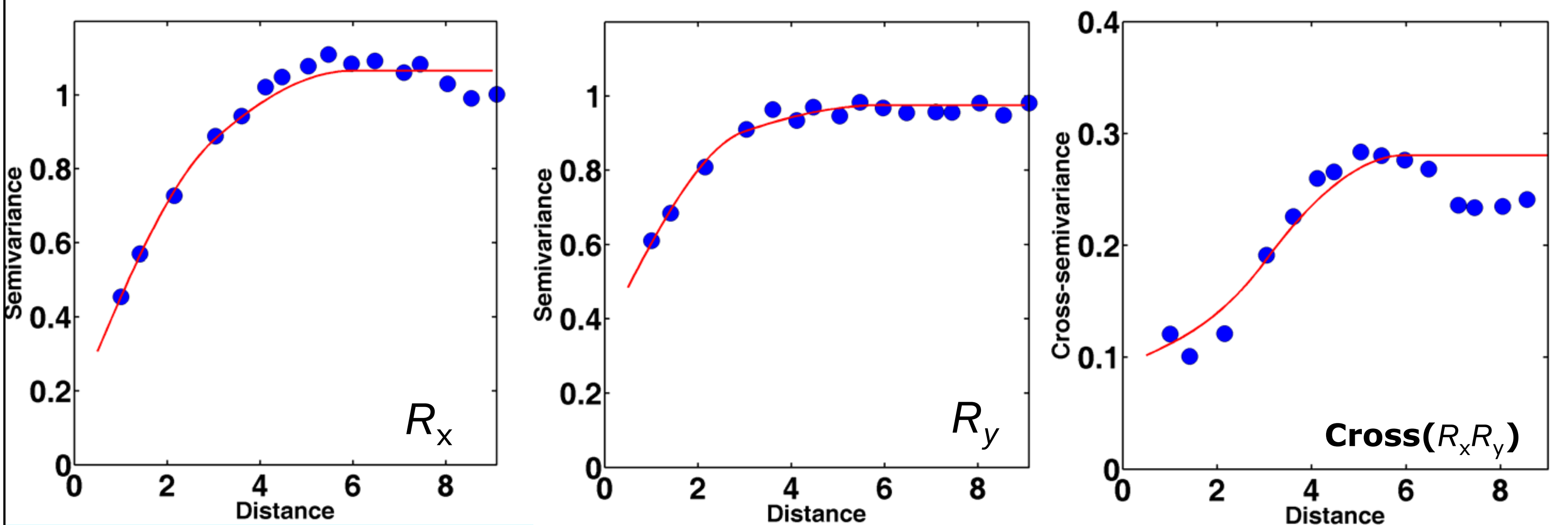
Superposition of random components

$R_j(\mathbf{u})$ ➔ superposition of spatial components with different range of autocorrelation



Linear Model of Coregionalization (LMC)

In the LMC, the random component $R_j(\mathbf{u})$ of each variable $Z_j(\mathbf{u})$ is viewed as the outcome of the <u>same</u> combination of underlying spatial processes.



$\gamma_{Rx}(\mathbf{h})$ = **0.16** nugget + **0.31** sph(3) + **0.60** sph(6)
$\gamma_{Ry}(\mathbf{h})$ = **0.36** nugget + **0.39** sph(3) + **0.23** sph(6)
$\gamma_{Rx\text{-}Ry}(\mathbf{h})$ = **0.09** nugget + **-0.12** sph(3) + **0.31** sph(6)

Journel & Huijsbregt, 1978; Goovaerts, 1997

Coregionalization analysis

$$\gamma_{Rx}(\mathbf{h}) = \boxed{\mathbf{0.16}}\ \text{nugget} + \boxed{\mathbf{0.31}}\ \text{sph}(3) + \boxed{\mathbf{0.60}}\ \text{sph}(6)$$
$$\gamma_{Ry}(\mathbf{h}) = \boxed{\mathbf{0.36}}\ \text{nugget} + \boxed{\mathbf{0.39}}\ \text{sph}(3) + \boxed{\mathbf{0.23}}\ \text{sph}(6)$$
$$\gamma_{Rx\ Ry}(\mathbf{h}) = \boxed{\mathbf{0.09}}\ \text{nugget} + \boxed{\mathbf{-0.12}}\ \text{sph}(3) + \boxed{\mathbf{0.31}}\ \text{sph}(6)$$

Matrices of sill estimates

$$\begin{bmatrix} 0.16 & 0.09 \\ 0.09 & 0.36 \end{bmatrix} \quad \begin{bmatrix} 0.31 & -0.12 \\ -0.12 & 0.39 \end{bmatrix} \quad \begin{bmatrix} 0.60 & 0.31 \\ 0.31 & 0.23 \end{bmatrix}$$

$r_{\mathbf{nug}} = \mathbf{0.32}$ $\quad r_{\mathbf{sph(3)}} = \mathbf{-0.31}$ $\quad r_{\mathbf{sph(6)}} = \mathbf{0.82}$

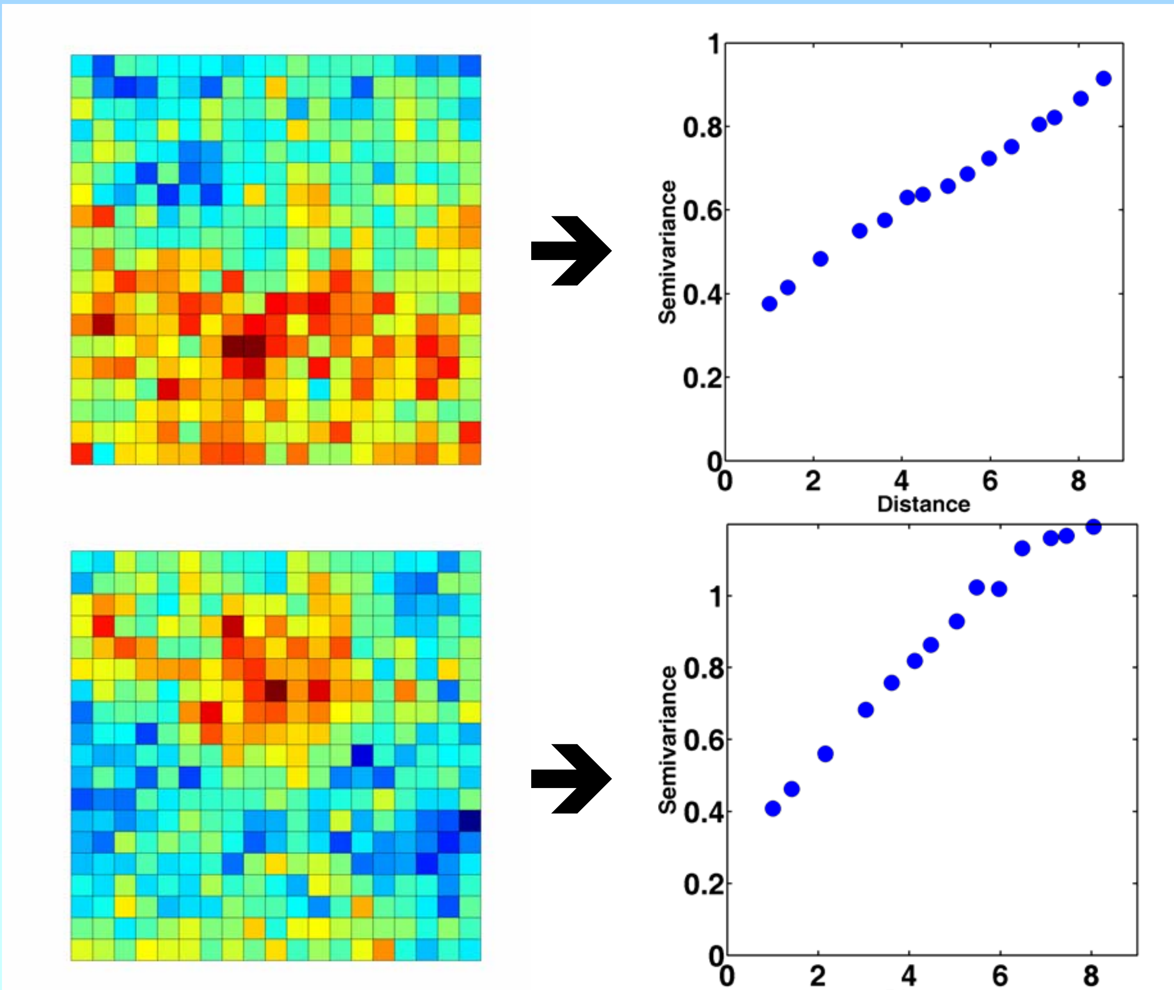
In the multivariate case, sill estimate matrices can be used in regionalized versions of PCA or RDA (Wackernagel et al., 1989)

Effect of spatial drift on variogram

In the presence of spatial drift $m_j(\mathbf{u})$, the variogram of $Z_j(\mathbf{u})$ is biased (not bounded) and cannot be used in coregionalization analysis.

Nugget + Sph (4) + Linear gradient

Nugget + Sph (4) + Large Patches



We need to work with detrended observations

$$\hat{R}_j(\mathbf{u}) = Z_j(\mathbf{u}) - \hat{m}_j(\mathbf{u})$$

But also interested in **large-scale** “variation” associated with spatial drifts

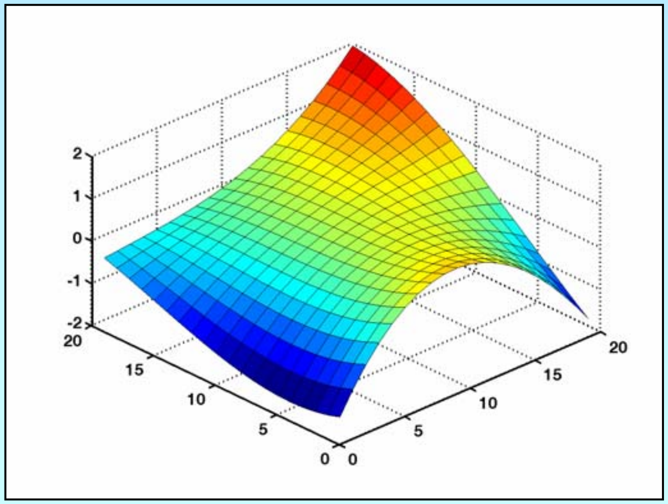
Estimation of spatial drift (1)

1- Estimation of the drift is performed by Generalized Least Squares (GLS)

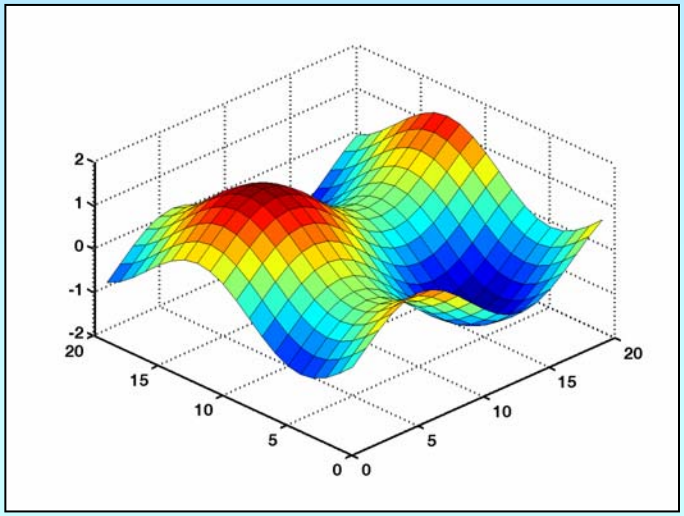
- takes spatial autocorrelation into account
- provides "drift estimates" with higher precision.

2- Global drift estimation: Parametric

Estimate of $m_j(\mathbf{u})$ = the value of a function of spatial coordinates expressed as:



polynomial of a given degree



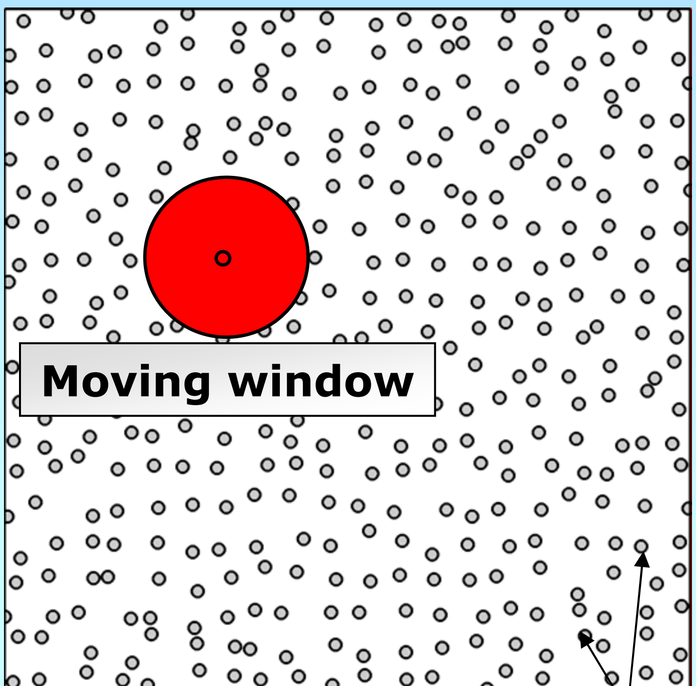
sum of cosine and sine waves

Estimation of spatial drift (2)

3- Local drift estimation: Nonparametric

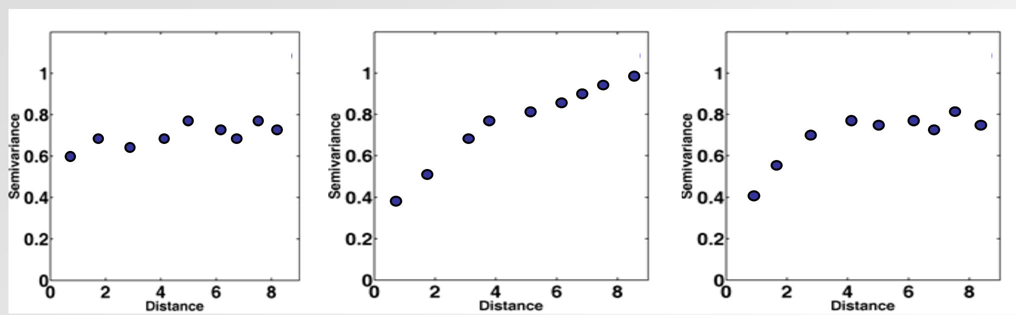
$m_j(\mathbf{u})$ is locally estimated within a window around each sampling location.

Local polynomial of order 0 (i.e., constant), 1 or 2 can be used for the estimation procedure



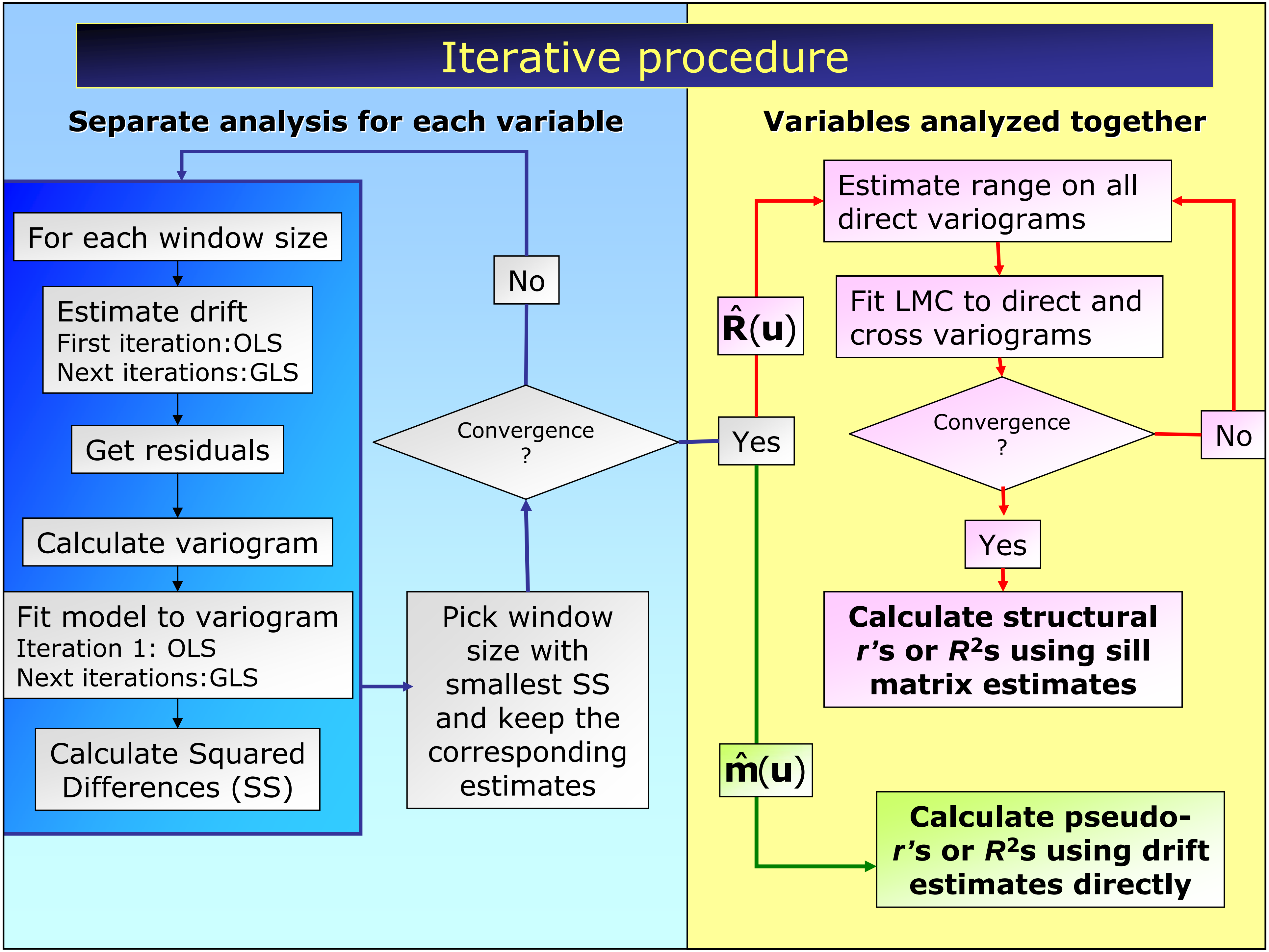
Sampling locations

Choice of window size



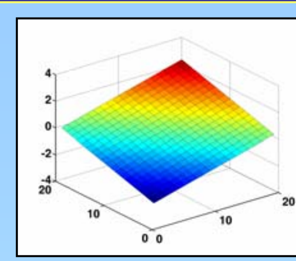
Too small? Too large? OK?

Criteria for choosing window size➔ Variogram of residuals with the "best" fitted model

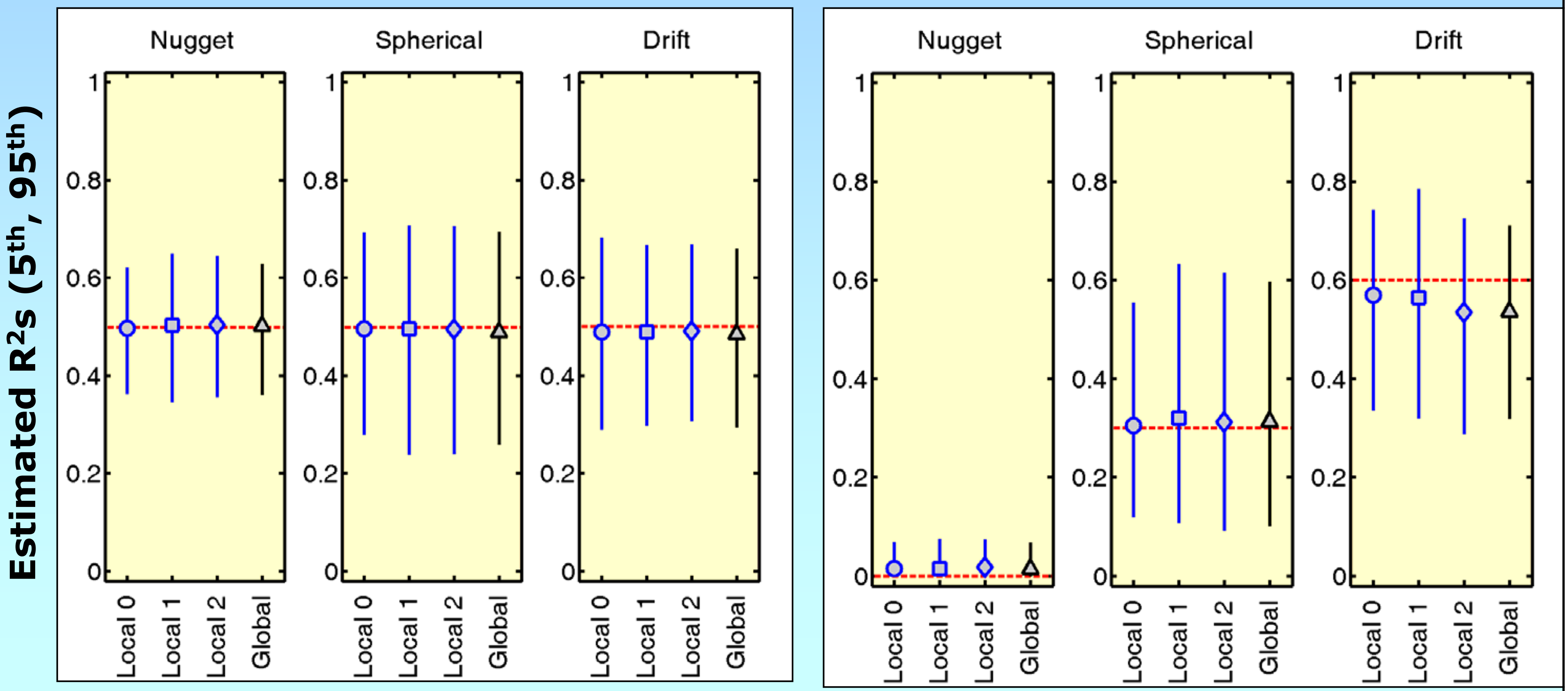

Iterative procedure
Separate analysis for each variable
Variables analyzed together
For each window size
Estimate drift
First iteration:OLS
Next iterations:GLS
Get residuals
Calculate variogram
Fit model to variogram
Iteration 1: OLS
Next iterations:GLS
Calculate Squared Differences (SS)
Pick window size with smallest SS and keep the corresponding estimates
Convergence ?
No
Yes
$\hat{R}(u)$
$\hat{m}(u)$
Estimate range on all direct variograms
Fit LMC to direct and cross variograms
Convergence ?
No
Yes
Calculate structural r's or R^2s using sill matrix estimates
Calculate pseudo-r's or R^2s using drift estimates directly

Simulation results (1)

500 simulations, 2 variables
20x20 grid
Each component with 1/3 of total variation
Nugget + Sph (3) + Linear gradients



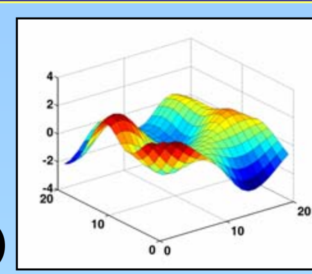
Structural ρ^2s: 0.5, 0.5, 0.5 **Structural ρ^2s: 0.0, 0.3, 0.6**



Drift estimation procedures **Drift estimation procedures**

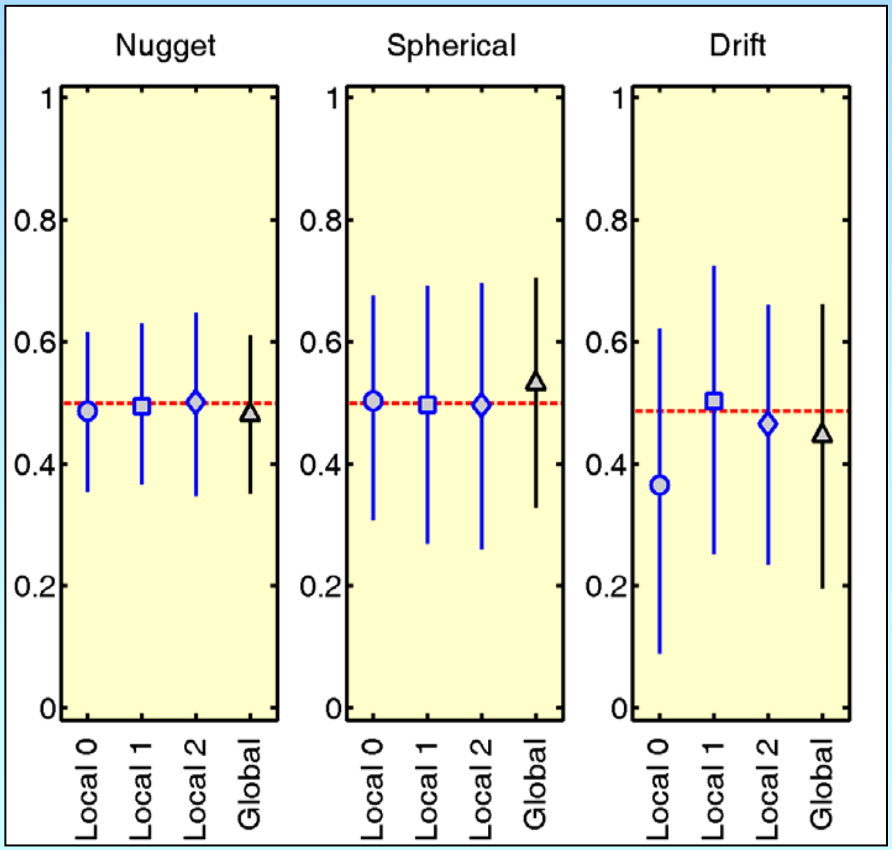
Simulation results (2)

500 simulations, 2 variables
20x20 grid
Each component with 1/3 of total variation
Nugget + Sph (3) + Large patches (Gaussian model)



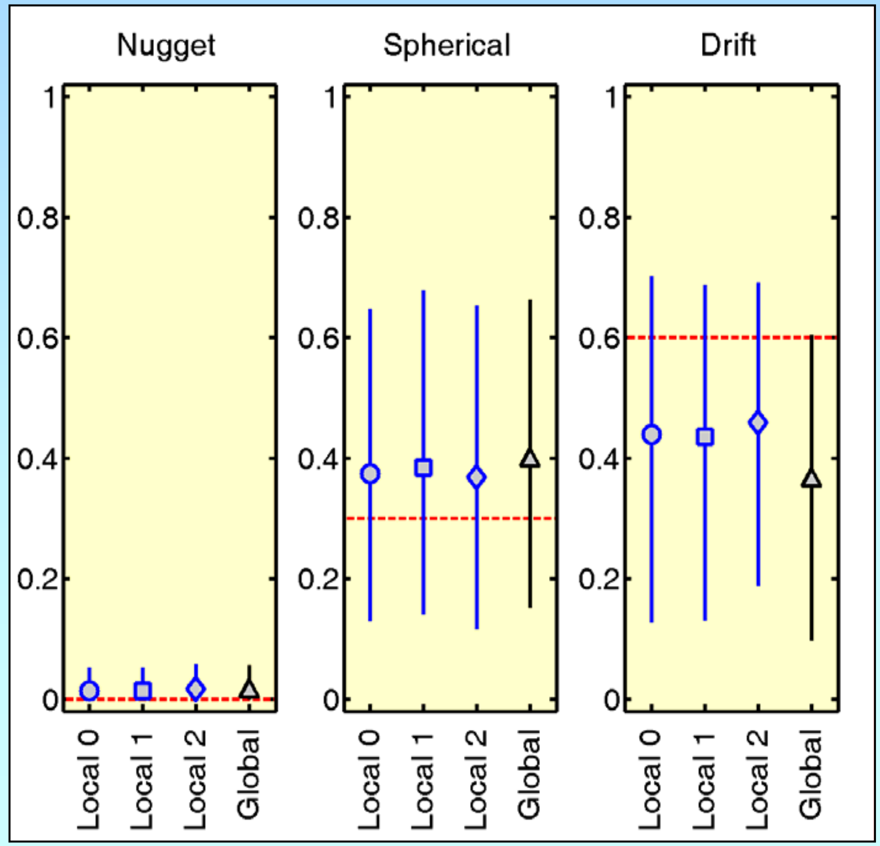
Estimated R^2s (5th, 95th)

Structural ρ^2s: 0.5, 0.5, 0.5



Drift estimation procedures

Structural ρ^2s: 0.0, 0.3, 0.6

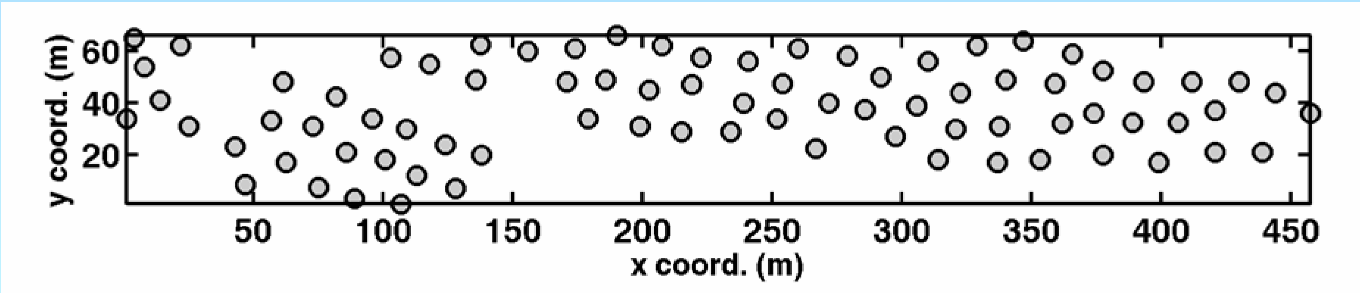


Drift estimation procedures

Forest data set (1)

Study on tree species influence on forest floor properties

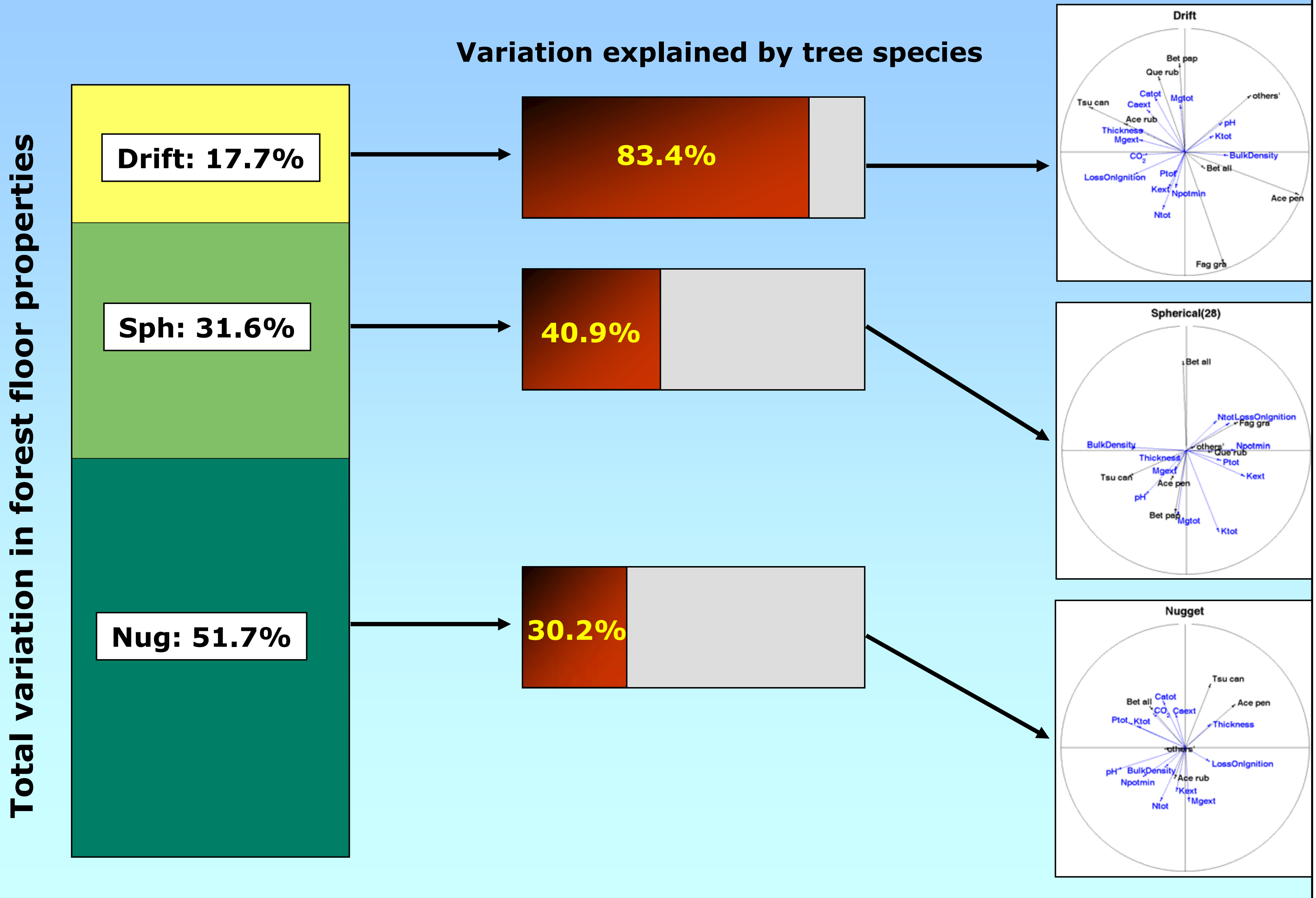
Explanatory variables: Index of influence for 8 tree species
Response variables: 14 forest floor properties
80 observations in mixed-species stand (Beech-Hemlock-Red Maple). *Pelletier et al.(1999), Ecoscience, 6(1):79-91*



Drifts estimated with moving window using a local polynomial of order 1

The LMC was based on a nugget effect and a spherical model with an estimated range of 28m.

Forest data set (2)



Summary

No unique decomposition for $\mathbf{Z}(\mathbf{u}) = \mathbf{m}(\mathbf{u}) + \mathbf{R}(\mathbf{u})$

Drift analysis is conducted jointly with the analysis of structural correlations.

Drift estimation takes into account the spatial autocorrelation.

Only the pseudo-correlations in the drift analysis (i.e., at large scale) result from projections.

The analysis of the random component is done within a probabilistic framework.

Inferences can be made about spatial processes generating the observed patterns.

Structural correlations are modeled in the LMC.

References can be found in Dutilleul's presentation (Part 2)