Spatial linear models in R
2015



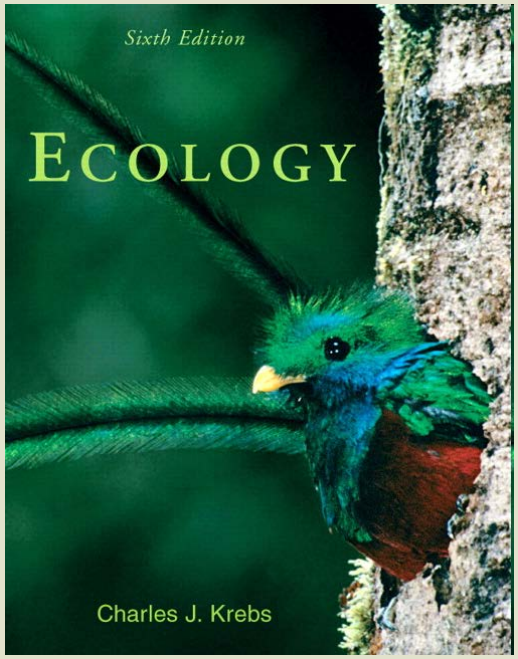
Simoneta Negrete Yankelevich

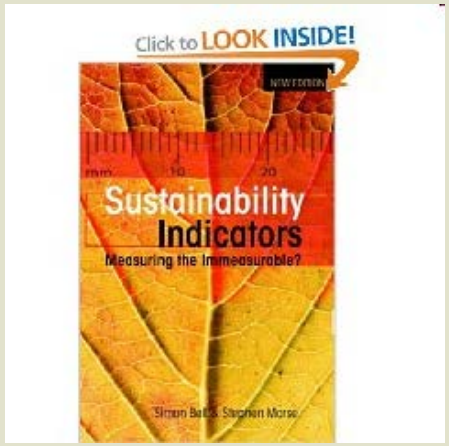
Today

- ¿Why should we care about space?
- The ABC of space
- How to recongnise and measure spatial patterns
- GLS (Generalised least squares): One of the many forms to consider spatial patterns in linear models.

¿Why should we care about space?

¿What is Ecology?





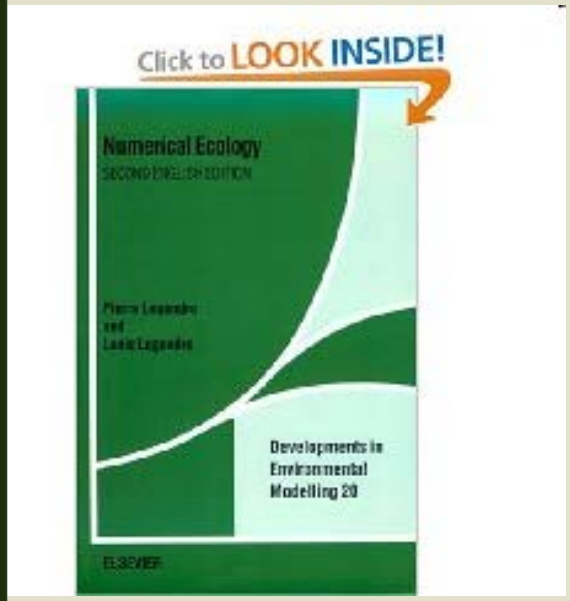
- "We are interested in *where* organisms are, *when* they occur *there* and *why*".

(Krebs, (4th ed) p.3)

- "It can only be considered sustanable if the temporal and spatial scales relevant to the phenomenon are selected"

(Costanza y Patten, 1995). "Defining and predicting sustainability." *Ecological Economics* **15 (3)**: 193–196.

The selection of numerical methods

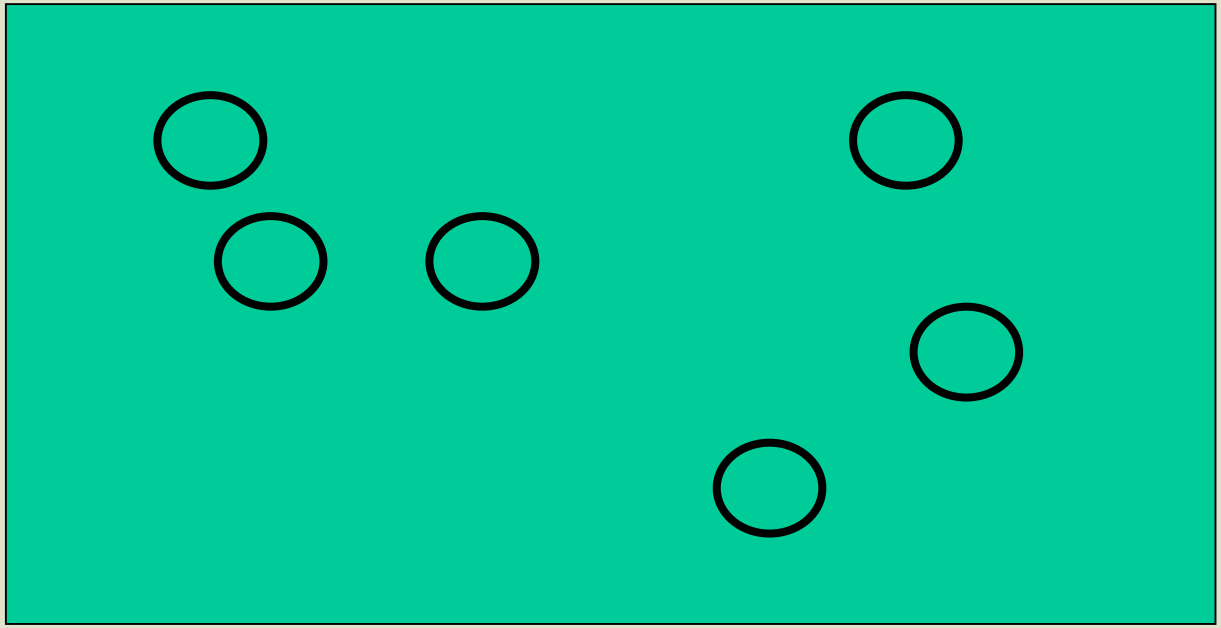


Numerical Ecology (Legendre y Legendre 1998)

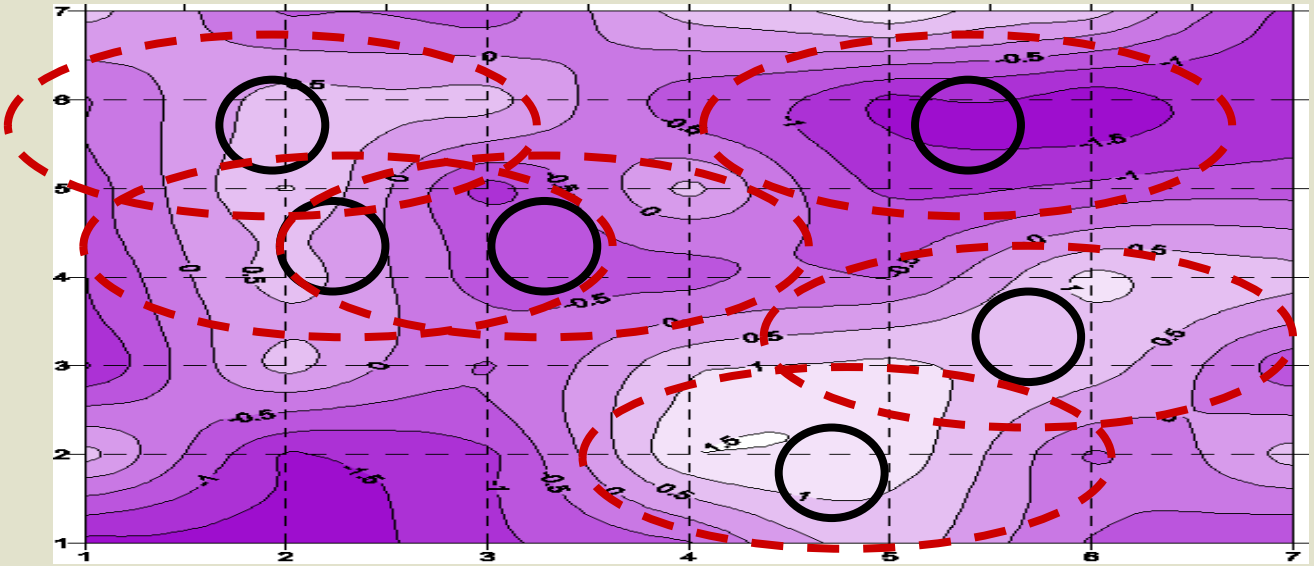
- The statistical methods most commonly used assume independence of observations: distribution in space is uniform or random.

- In nature the rule are patches and gradients.

Over estimation of degrees of freedom

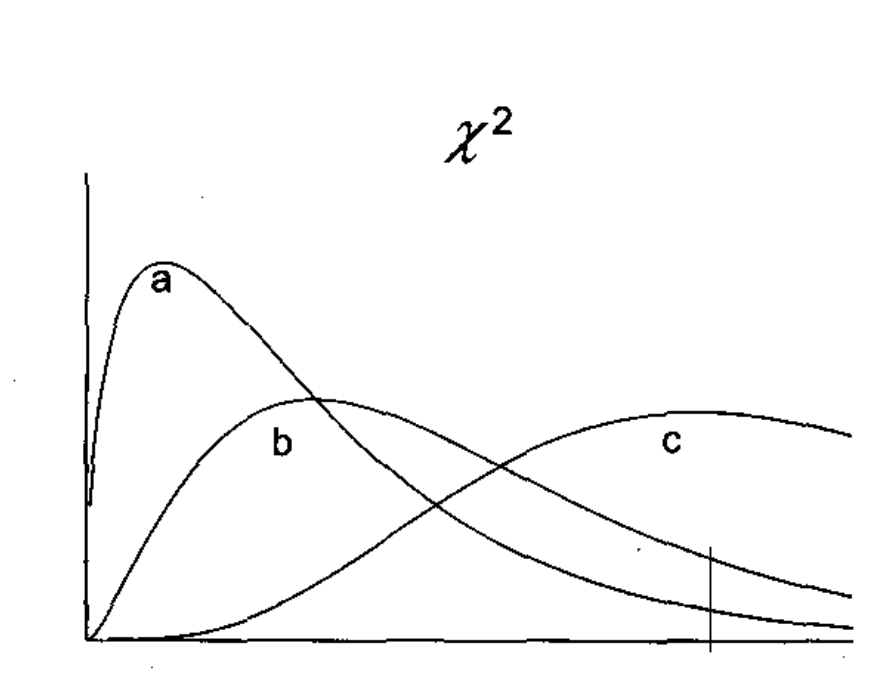


d.f.= (6-1)



d.f. = (*4.5*-1)

Comparison with test statistic (X^2)



$$X^2 = 35.5$$

Assuming Independence in space when there isn't one, increases the probability of false positives in hypothesis testing.

Spatial structuring is functional

In homogenous or random ecosystems

•Consumers would not me close to resources
•It would be impossible for an organism to predict where to find a mating partner.
•Soil conditions around a plant would not be particularly suitable or unsuitable for its offspring.

•Spatial structuring of ecosystems is functional and that means it should be central focus of attention in Ecological research.

The ABC of space

Types of spatial data

- Point pattern. The process determines the position of individuals in space.
- Geostatistical. The process affects the value of the variable in different points of space.
- Areal. The process generates qualitative or quantitative changes in contiguous surfaces, mosaics or grids. The information of which cells are neighbours is available.

The ABC of space

- Patterns

  Patch: area “with delimited borders” characterized by a higher concentration than expected at random.

  Gradient: Gradual increase in a single direction.

  uniform (disperse or overdisperse):More homogeneous than expected at random.

Pattern is used as the opposite of random.

BEFORE DOING ANYTHING!!!!

- ¿Does your hypothesis predict who, when and WHERE?
- Define the temporal and spatial extension in which your phenomenon expresses itself.
- Determine if the temporal and spatial resolution of the study design will capture the process.
- Make sure that the sampling design will produce the type of data appropriate for the statistical strategy that you are planning to use

How to detect and measure spatial pattern?

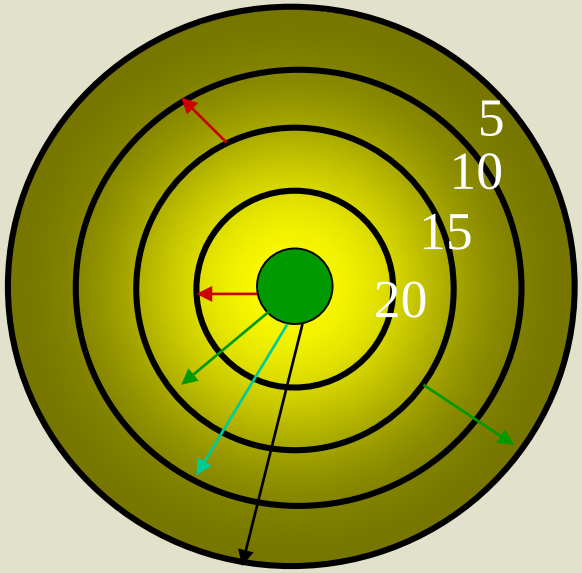
Spatial dependence

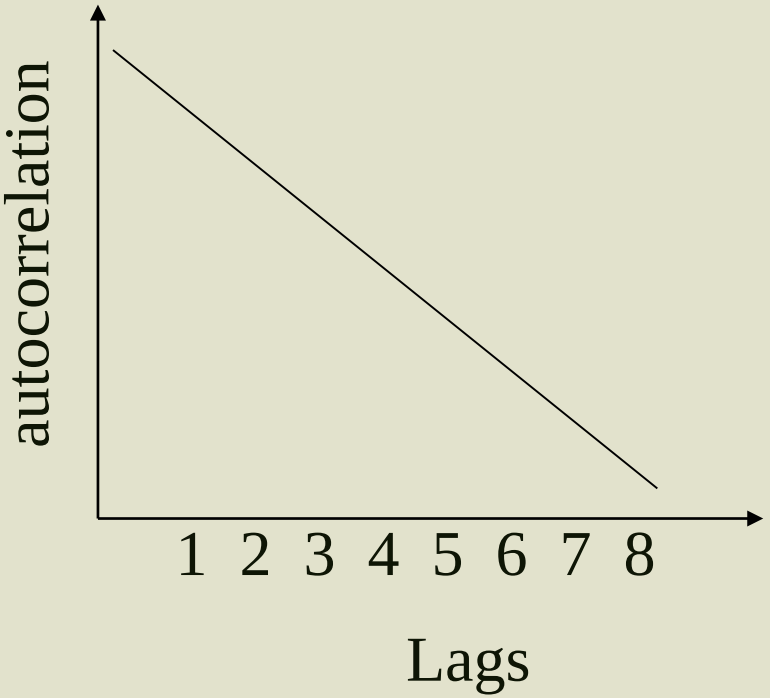
- Spatial dependence or autocorrelation is the correlation among the values of a variable given their position in space.

- Two observations that are close together have similar values, as their distance increases, the relationship fades-off until it becomes undetectable. We then tentatively conclude that the observations become spatially independent.

Spatial dependence

Seeds that fall from a tree

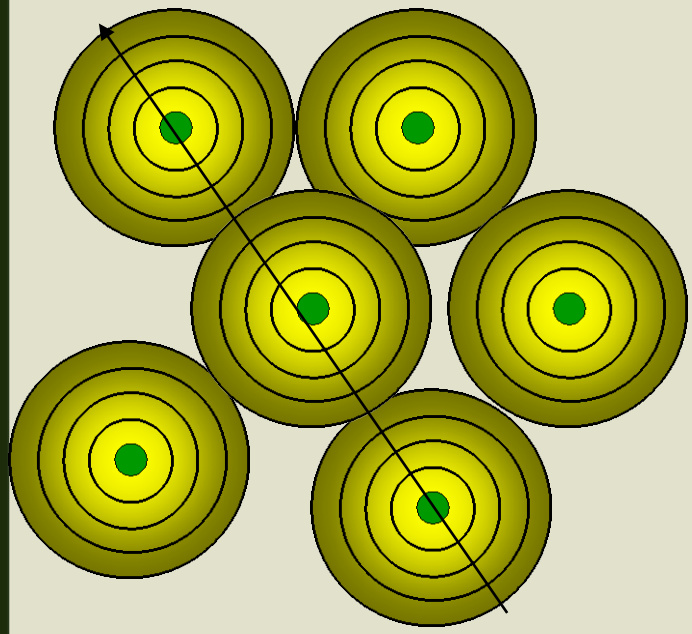


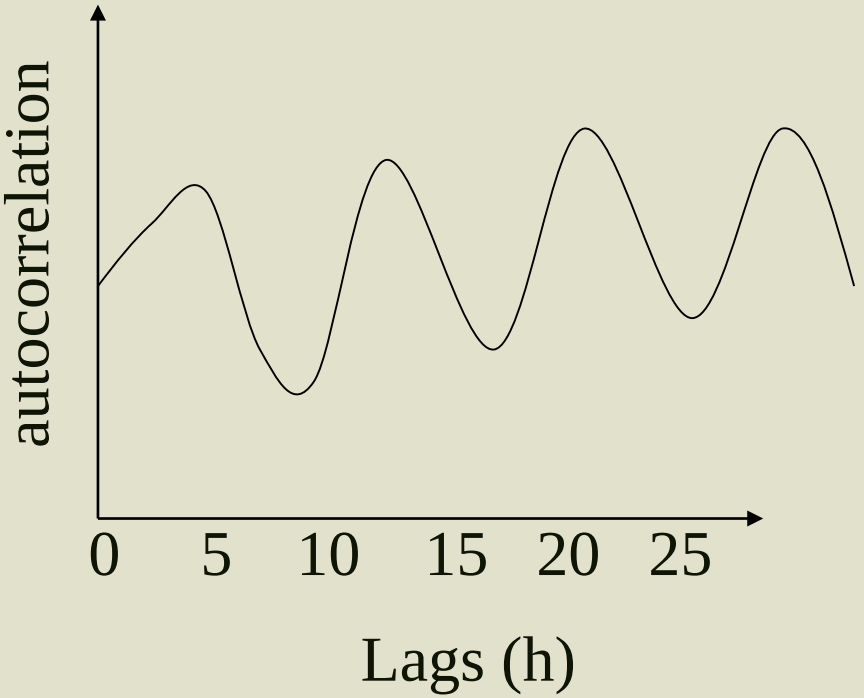


It is measured through the comparison of lags or distance intervals

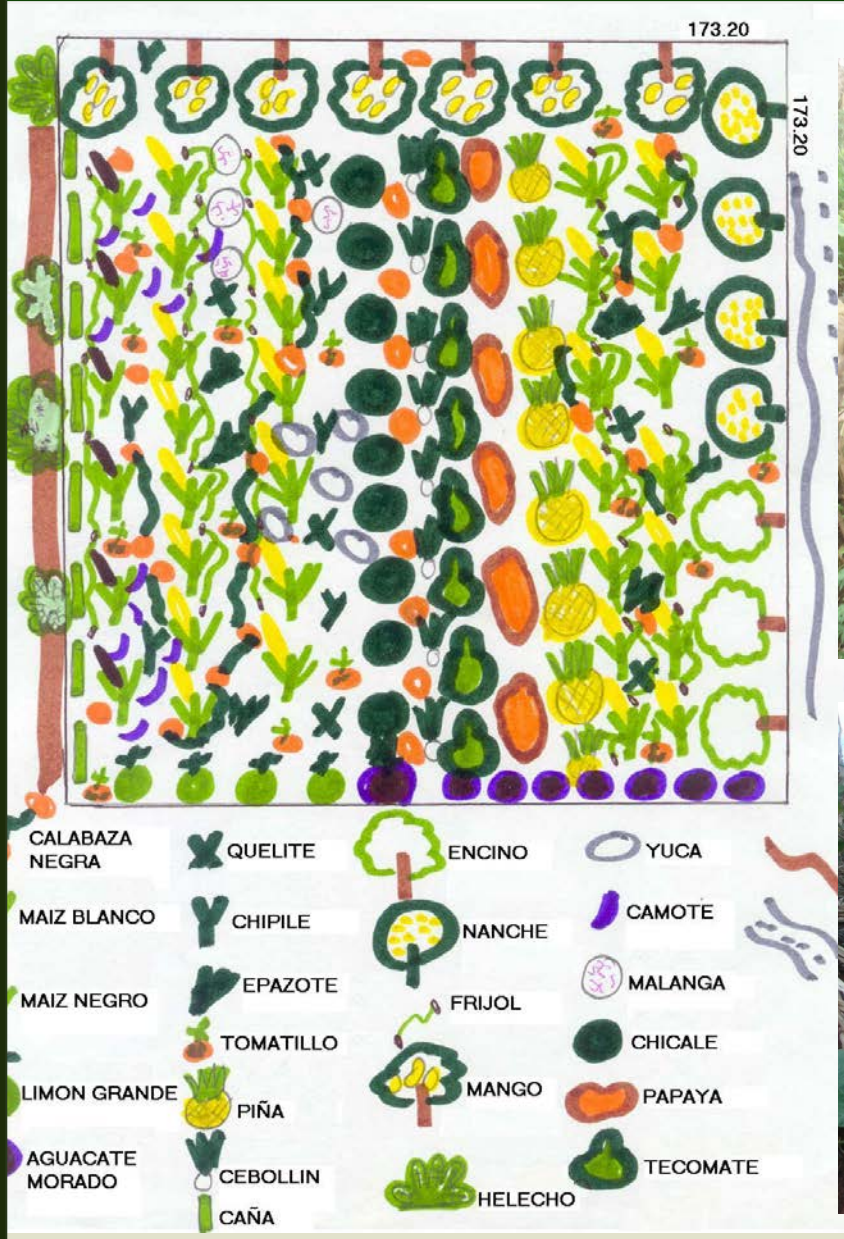
Spatial dependence

Seed abundance in a plantation





Popoluca Maize polycultures, Los Tuxtlas, Mexico



#R1

#C1

Autocorrelation function: Semi-variance

Traditionally d is h in variograms

$$\hat{\gamma}(h) = \frac{1}{2W(h)} \sum_{i=1}^{W(h)} [y_i - y_{i+h}]^2$$

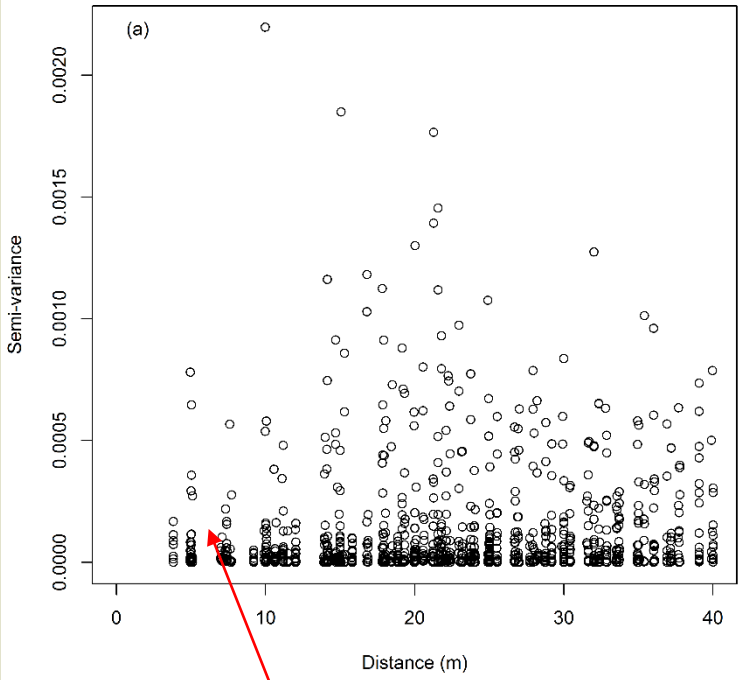
Weighted mean

Squared differences

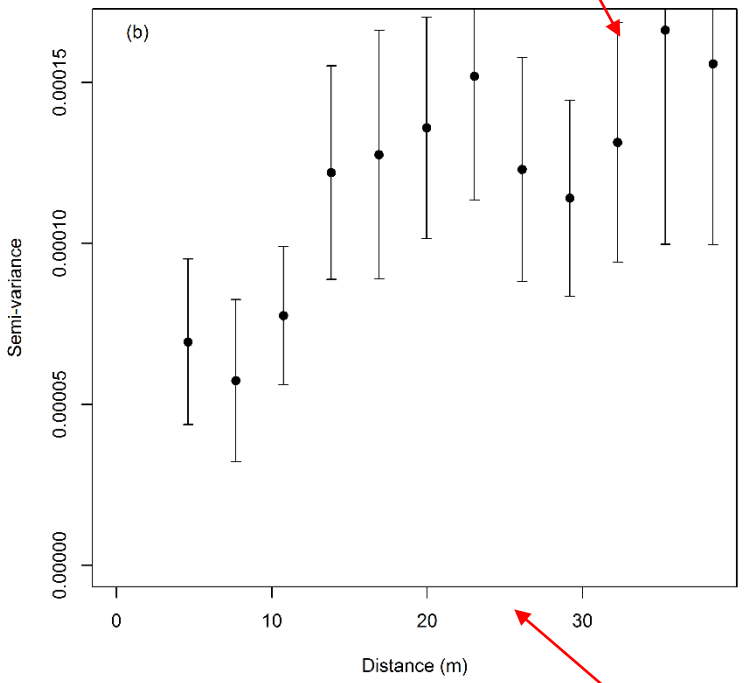
Autocorrelation functions: Variograms

CI increase with distance, confidence in estimations diminishes

Cloud Variogram



Variograma



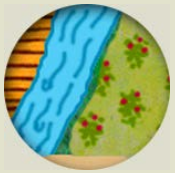
Raw squared differences (each point is a pair of observations)

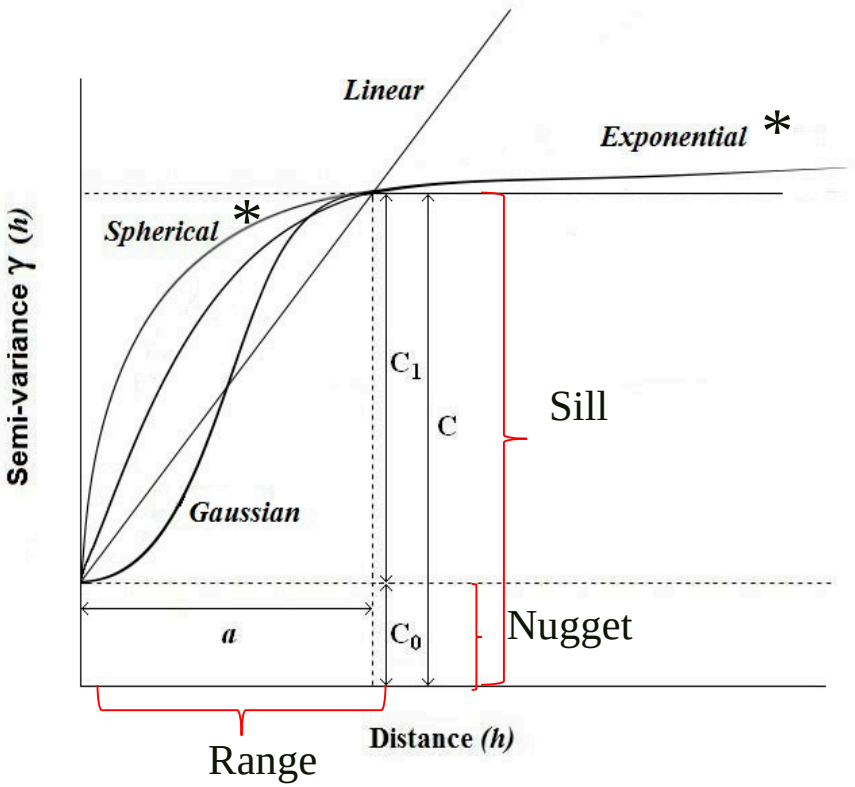
The weighted mean square with exploratory CI.

#R2

#C2

Variography: Theoretical Variograms





Nugget effect: $\gamma(h) = C_0$

Exponential: $\gamma(h) = C_0 + C_1\left[1 - \exp\left(-3\frac{h}{a}\right)\right]$

Spherical : $\gamma(h) = C_0 + C_1\left[1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right]$

if $h \le a;\ \gamma(h) = C\ if\ h > a$

Linear: $\gamma(h) = C_0 + bh$

Gaussian: $\gamma(h) = C_0 + C_1\left[1 - \exp\left(-3\frac{h^2}{a^2}\right)\right]$

* Asymptotic models: when a reaches 90% of the asymptote.

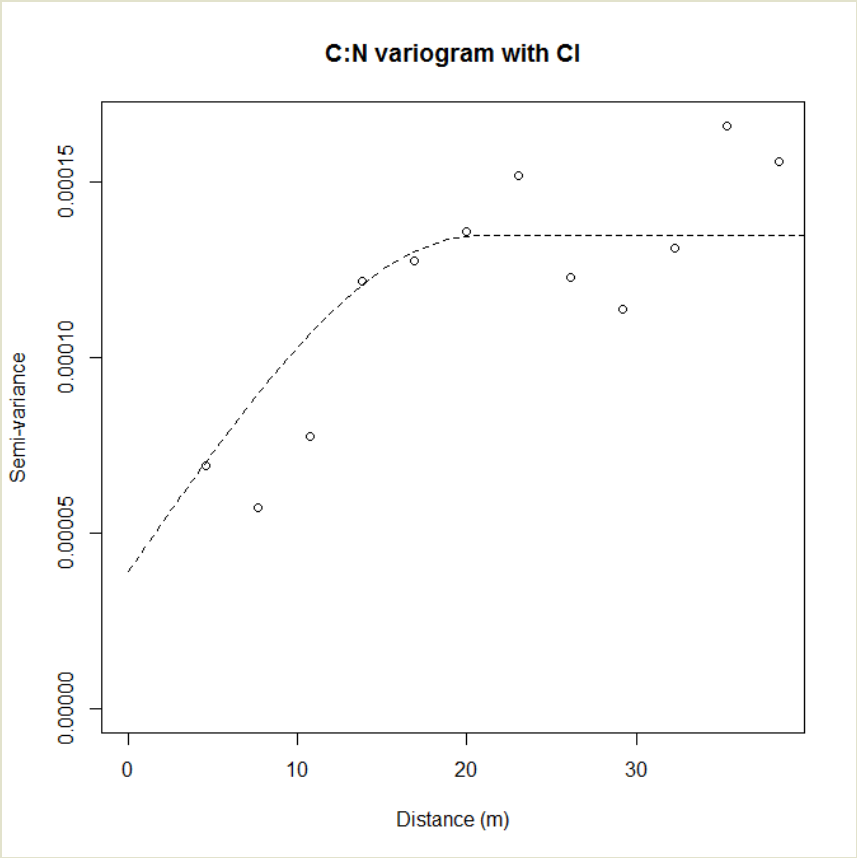
#R3

#C4

Then what?

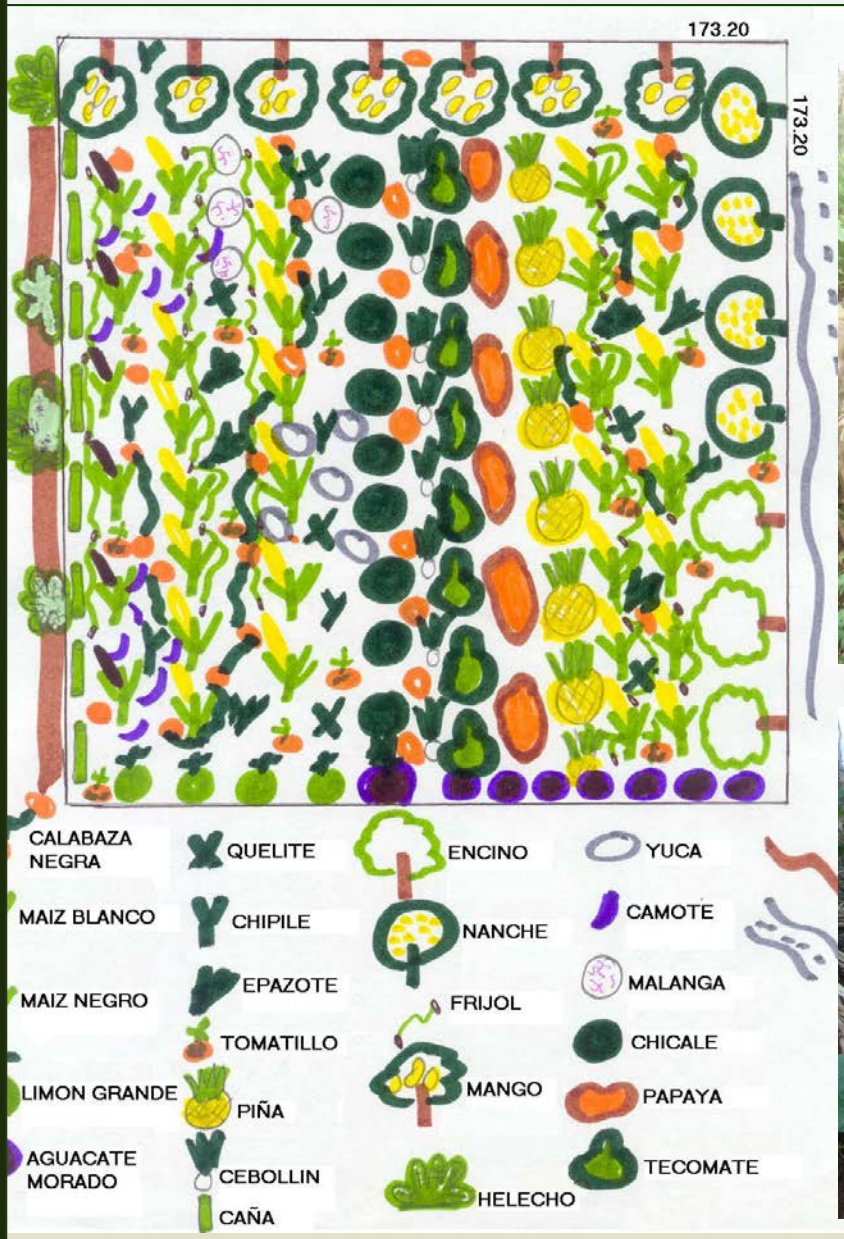
- ML Variogram:

Range: 20.7 m
nugget: $3.9X10^{-5}$
Sill: $1.0\ X10^{-4}$



Conclusion: Subtle patchy pattern (20.7 m) within a gradient.

Popoluca Maize polycultures, Los Tuxtlas, Mexico



Spatial linear models: GLS

Spatial linear models

- When one of the important explanatory variables was not measured and it prints a spatial patter on the dependent variable.

- When the dependent variable has a normal distribution there is GLS (Generalised least squares or kriging regression, a type of mixed model with an R-side effect).

A type of inferential analysis

- Requires significance tests of statistics.
- You have to comply with distributional assumptions once the spatial pattern is accounted for.
- Randomisations of positions (restricted or not) are used to generate the distribution that represents H0.

GLS: a type (strange) of mixed model

There is no random factor!

Sides

G R

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \Lambda \mathrm{i})$$

variance-covariance matrix. can be calculated with the sill, nugget and range parameters of the variogram.

#R4

#C5

What did we learn about the C:N in polycultures?

- There are two spatial structures in its distributions (gradients and patches of 23m).
- When we include the nearest species in the model, the patchy pattern is explained but not the gradient.
- This pattern is mainly associated with a lower C:N close to sweet potatoes (legumes that are symbionts with Bradirizobia a nitrogen fixing bacteria from the forest!).

Autorregresive models: SAR

Exogenous autocorrelation (unknown)

SARerr: $\boldsymbol{Y} = \boldsymbol{\beta X} + \lambda \boldsymbol{W} u + \boldsymbol{\varepsilon}$

Endogenous autocorrelation

SARlag: $\boldsymbol{Y} = \rho \boldsymbol{WY} + \boldsymbol{\beta X} + \boldsymbol{\varepsilon}$

Endogenous and exogenous (known) autocorrelation

SARmix: $\boldsymbol{Y} = \rho \boldsymbol{WY} + \boldsymbol{\beta X} + \theta \boldsymbol{WX} + \boldsymbol{\varepsilon}$

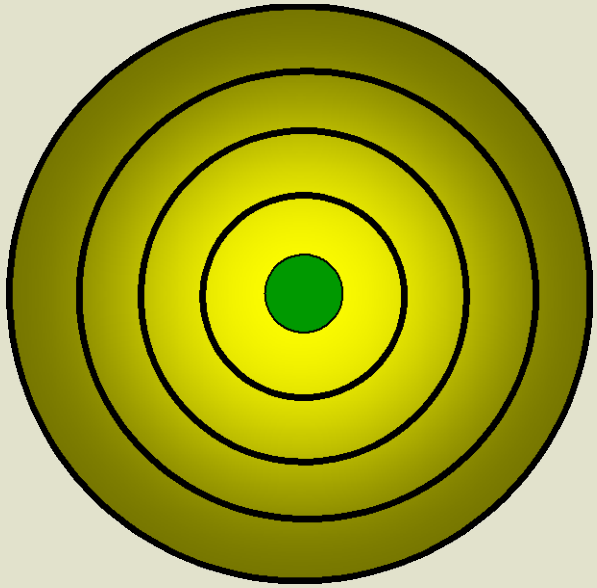
Autorregression coefficients

Weighted Neighbourhood matrix

Spatially dependent error

SARlag and SARmix: direct and indirect impacts

- Indirect impacts (spill overs) are like an expansive wave (seabird nesting and gregarious behaviour).



In summary

- There is a world to be explored! And…it is a spatially explicit one.
- Spatially explicit models are complex. They ARE NOT an easy way to get read of design problems.
- Spatially explicit models ARE an elegant and informative way to understand the role of space in Ecology.
- This is only a starter…have a nice spatial trip!