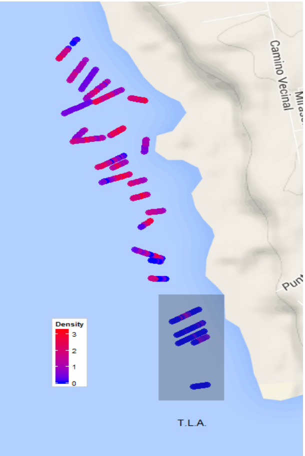
Day 4

6. About Effective Sample Size

Density of Lessonia trabeculata

- 26 transects were considered.
- The distance between two transects was 100 meters.
- The length of the transects was 100 or 200 meters.
- The observations were collected every 10 meters within each transect.
- The observations were collected by divers who follow red marks located on ropes underwater.



A Simple Example

- Cressie (1993, p.16) defined the effective sample size (ESS) as the equivalent number of iid observations associated with a spatial sample of size n.

- Consider a random field $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^2\}$ so that
$$\boldsymbol{Y}(\boldsymbol{s}) = (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n))^\top \sim \mathcal{N}(\mu\mathbf{1}, \boldsymbol{\Sigma}),$$
where $\boldsymbol{\Sigma}_{ij} = \sigma^2\rho^{|i-j|}, i, j = 1, \ldots, n,\ 0 < \rho < 1$. Then
$$\begin{aligned}\text{var}(\overline{Y}) &= \frac{\sigma^2}{n}\left(1 + \frac{2\rho}{(1-\rho)}(1 - 1/n) - 2(\rho/(1-\rho))^2(1-\rho^{n-1})/n\right)\\ &= \frac{\sigma^2}{\text{ESS}}\end{aligned}$$
where
$$\text{ESS} = n/[1 + \frac{2\rho}{(1-\rho)}(1 - 1/n) - 2(\rho/(1-\rho))^2(1-\rho^{n-1})/n]$$

Extension for a Regular Grid in $\mathbb{Z}^2$

A similar formula can be obtained for a spatial process $\{Y(\boldsymbol{s}) : s \in D\}$ where $D = \{(i,j) \in \mathbb{Z}^2 : 1 \leq i \leq m, 1 \leq j \leq n\}$, with the covariance function

$$\mathrm{cov}[Y(\boldsymbol{s}), Y(\boldsymbol{t})] = \sigma^2 \rho^{||\boldsymbol{s}-\boldsymbol{t}||}, \boldsymbol{s}, \boldsymbol{t} \in D.$$

Defining

$$\sum_{\boldsymbol{s},\boldsymbol{t} \in D} \mathrm{cov}[Y(\boldsymbol{s}), Y(\boldsymbol{t})] = \sigma^2 \sum_{\boldsymbol{s},\boldsymbol{t} \in D} \rho^{||\boldsymbol{s}-\boldsymbol{t}||}$$

$$= (m+1)\frac{\sigma^2 \rho}{1-\rho}\left(n - \rho\frac{1-\rho^n}{1-\rho}\right) + (n+1)\frac{\sigma^2 \rho}{1-\rho}\left(m - \rho\frac{1-\rho^m}{1-\rho}\right)$$

$$+ \sigma^2 2 \sum_{i=1}^{n}\sum_{j=1}^{m}(m+1-i)(n+1-j)\rho^{\sqrt{i^2+j^2}} = \sigma^2\ S(\rho, m, n),$$

we obtained

$$\mathrm{var}[\overline{Y}] = \frac{\sigma^2}{(m+1)(n+1)}\left[1 + \frac{2S(\rho,m,n)}{(m+1)(n+1)}\right].$$

Extension for a Regular Grid in $\mathbb{Z}^2$

- Thus the effective sample size is given by[1]

$$\text{ESS} = \frac{(m+1)(n+1)}{\left[1 + \frac{2S(\rho,m,n)}{(m+1)(n+1)}\right]}.$$

- Griffith (2005)[2] suggested the model

$$\boldsymbol{Y} = \mu\mathbf{1} + \boldsymbol{e} = \mu + \boldsymbol{\Sigma}^{-1}\boldsymbol{e}^*,$$

where $\boldsymbol{e}$ and $\boldsymbol{e}^*$, denote the $n \times 1$ error vectors and $\boldsymbol{\Sigma}$ is a covariance matrix.

$$\text{ESS} = \frac{\text{tr}(\boldsymbol{\Sigma}^{-1})}{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\mathbf{1}}\ n,$$

[1] Vallejos, R., Osorio, F., (2014). Effective sample size for spatial process models. Spatial Statistics 9, 66-92.

[2] Griffith, D. (2005). Effective Geographic Sample Size in the Presence of Spatial Autocorrelation. Annals of the Association of American Geographers 95, 740-760.

ESS for General Spatial Regression Processes

- The definitions of Griffith (2005) and Vallejos and Osorio (2014) only considered regression processes with a constant mean.
- Let $\{Y(\boldsymbol{s} : \boldsymbol{s} \in \mathbb{R}^d)\}$ be a random field. Assume that $Y(\cdot)$ has been measured at $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \in \mathbb{R}^d$, and the vector of observations is $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^\top$. Consider the spatial regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

  where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}(\boldsymbol{\theta}))$.
- How to compute the ESS for $\boldsymbol{Y}$?

ESS for General Spatial Regression Processes

- The Fisher information quantity for $\boldsymbol{\beta}$ is

$$I(\boldsymbol{\beta}) = \boldsymbol{X}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \boldsymbol{X}.$$

- One way to generalize a formula consist of defining a weighted ESS of the form

$$\text{ESS} = \sum_{i=1}^{p} v_i \boldsymbol{x}_i^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \boldsymbol{x}_i,$$

where v_i are suitable weights to be determined.

- v_i are related to the number of covariates included in $\boldsymbol{X}$.

- Our goal is to have a measure satisfying $1 \leq \text{ESS} \leq n$.

ESS for General Spatial Regression Processes

Definition

Consider a spatial regression model as $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}(\boldsymbol{\theta}))$ with $\boldsymbol{X} = [\mathbf{1}, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p]$, $\boldsymbol{x}_j = [x_j(\boldsymbol{s}_1), x_j(\boldsymbol{s}_2), \ldots, x_j(\boldsymbol{s}_n)]^\top$, and $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n \in D$, such that $\boldsymbol{x}_j^\top \boldsymbol{x}_j = n$, for all $j = 2, \ldots, p$. The effective sample size of $\boldsymbol{Y}$ is defined as[3]

$$\text{ESS} = \frac{\operatorname{tr}[\boldsymbol{X}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \boldsymbol{X}]}{p} = \frac{1}{p} \sum_{j=1}^{p} \boldsymbol{x}_j^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \boldsymbol{x}_j. \tag{2}$$

If $\boldsymbol{X} = \mathbf{1}$ in (2) , ESS $= \mathbf{1}^\top \boldsymbol{R}(\boldsymbol{\theta})^{-1} \mathbf{1}$. If $\boldsymbol{R}(\boldsymbol{\theta}) = \boldsymbol{I}$, then ESS $= n$. That is, we recover the definition given by Vallejos and Osorio (2014).

[3] Acosta, J., Vallejos, R. (2018). Effective sample size for spatial regression processes. Electronic Journal of Statistics 12, 3147-3180.

An Application

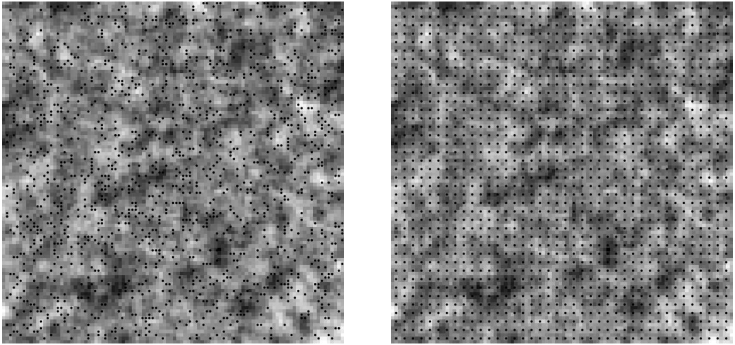


Figure: (a) Original image X generated from a SAR process, and the 1542 points sampled using a random sampling scheme; (b) Transformed image Y generated form image X, and the 1600 points sampled using a systematic sampling scheme.

An Application

- We consider an original image (X) of size 100×100 from a SAR process.
- Another image (Y) was generated transforming original one, using

$$Y = 0.5X + X^2 + X^3$$

- For each image, we compute the effective sample size $\widehat{\mathsf{ESS}}_X$ and $\widehat{\mathsf{ESS}}_Y$ and the corresponding variances $\mathrm{var}[\widehat{\mathsf{ESS}}_X]$ and $\mathrm{var}[\widehat{\mathsf{ESS}}_Y]$.
- $\widehat{\mathsf{ESS}}_{XY} \approx \dfrac{\widehat{\mathsf{ESS}}_X + \widehat{\mathsf{ESS}}_Y}{2}$ and

  $\mathsf{var}[\widehat{\mathsf{ESS}}_{XY}] \approx \dfrac{\mathsf{var}[\widehat{\mathsf{ESS}}_X] + \mathrm{var}[\widehat{\mathsf{ESS}}_Y]}{4}$.
- In this case $\widehat{\mathsf{ESS}}_{XY} \approx 1542$.

An Application

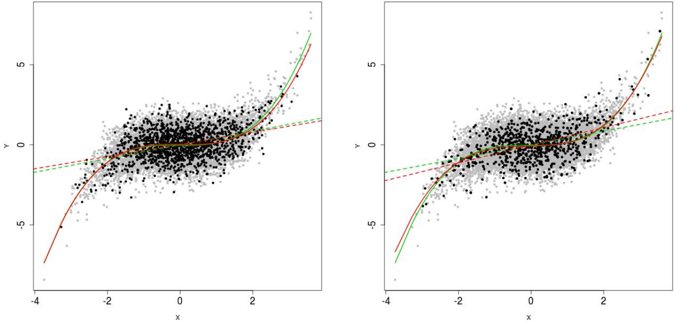


Figure: Spatially simplified scaterplots. (Left) Scatterplot generated with the 1542 sampled points from a random sampling scheme; (Right) Scatterplot generated with the 1600 sampled points from a systematic sampling sheme. The red lines represent the fitted lines using the sampled points while the green lines represent the fitted lines using the whole information.

Introduction to Codispersion

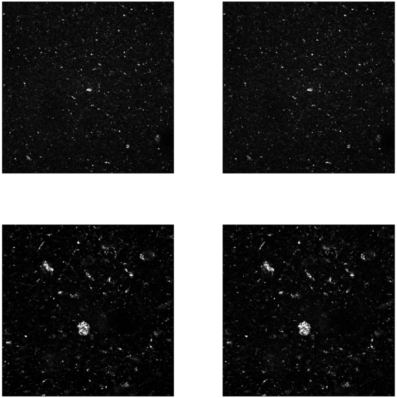


Figure: Dispersion of nanotubes. Images taken at NIST, USA. The images of the top were taken at the same distance. The images of the bottom were taken at the same distance but closer.

7. The Codispersion Coefficient

- Consider two intrinsically stationary processes $\{X(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ and $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ with variograms $2\gamma_X(\cdot)$ and $2\gamma_Y(\cdot)$.

- suppose that the cross-variogram associated with processes $X(\cdot)$ and $Y(\cdot)$ is denoted as $\gamma_{XY}(\cdot)$.

- The codispersion coefficient is a normalized version of the cross-variogram given by

$$\rho_{XY}(\boldsymbol{h}) = \frac{\gamma_{XY}(\boldsymbol{h})}{\sqrt{\gamma_X(\boldsymbol{h})\gamma_Y(\boldsymbol{h})}}, \tag{3}$$

  where $\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{h} \in D$, $\gamma_X(\boldsymbol{h}) = \mathbb{E}[X(\boldsymbol{s}+\boldsymbol{h}) - X(\boldsymbol{s})]^2$ and similarly for $\gamma_Y(\cdot)$.

- It is clear that $|\rho_{XY}(\boldsymbol{h})| \leq 1$.

Time Series Examples

For the error vector $(\epsilon_1(t), \epsilon_2(t))^\top$ with mean $\mathbf{0}$ and correlation structure given by

$$\mathrm{cor}(\epsilon_1(t), \epsilon_2(s)) = \begin{cases} \rho & , \text{ if } s = t \\ 0 & , \text{ otherwise,} \end{cases}$$

we proved that

Models	Equations	$\rho(1)$	Constant
AR(1)	$X(t) = \phi_1 X(t-1) + \epsilon_1(t)$ $Y(t) = \phi_2 X(t-1) + \epsilon_2(t)$	$\rho \cdot c_1$	$c_1 = \frac{(2-\phi_1-\phi_2)\sqrt{(1+\phi_1)(1+\phi_2)}}{2(1-\phi_1\phi_2)}$
MA(1)	$X(t) = \epsilon_1(t) + \theta_1 \epsilon_1(t-1)$ $Y(t) = \epsilon_2(t) + \theta_2 \epsilon_2(t-1)$	$\rho \cdot c_2$	$c_2 = \frac{(2-\theta_1-\theta_2+2\theta_1\theta_2)}{2\sqrt{(1-\theta_1+\theta_1^2)(1-\theta_2+\theta_1^2)}}$

Spatial AR Process Example

Consider the two AR-2D process of the form

$$X(i,j) = \phi_1 X(i-1,j) + \phi_2 X(i,j-1) + \phi_3 X(i-1,j-1) + \epsilon_1(i,j),$$
$$Y(i,j) = \psi_1 Y(i-1,j) + \psi_2 Y(i,j-1) + \psi_3 Y(i-1,j-1) + \epsilon_2(i,j),$$

where $\epsilon_1(i,j)$ and $\epsilon_2(i,j)$ are zero-mean i. i. d. random variables with variances σ^2 and τ^2, respectively, and the correlation structure is defined as

$$\mathrm{cor}[\epsilon_1(i,j), \epsilon_2(k,l)] = \begin{cases} \rho, & (i,j) = (k,l), \\ 0, & (i,j) \neq (k,l). \end{cases}$$

Basu and Reinsel (1993) studied the conditions for the parameters to be in the stationaty region.

Spatial AR Process Example

If the parameters of the models are in the stationary region, one can obtain

$$\rho_{XY}(\boldsymbol{h}) = K \cdot \rho,$$

where $K = \frac{K_1}{2R_1R_2}$ and

$$K_1 = C(\phi_1, \phi_2, \phi_3, \psi_1, \psi_2, \psi_3, 0, 0) - [\psi_1^{h_1}\psi_2^{h_2} + \phi_1^{h_1}\phi_2^{h_2}]C(\phi_1, \phi_2, \phi_3, \psi_1, \psi_2, \psi_3, h_1, h_2),$$

$$R_1 = \sqrt{C(\phi_1, \phi_2, \phi_3, \phi_1, \phi_2, \phi_3, 0, 0) - \phi_1^{h_1}\phi_2^{h_2}C(\phi_1, \phi_2, \phi_3, \phi_1, \phi_2, \phi_3, h_1, h_2)},$$

$$R_2 = \sqrt{C(\psi_1, \psi_2, \psi_3, \psi_1, \psi_2, \psi_3, 0, 0) - \psi_1^{h_1}\psi_2^{h_2}C(\psi_1, \psi_2, \psi_3, \psi_1, \psi_2, \psi_3, h_1, h_2)}.$$

$$C(\phi_1, \phi_2, \phi_3, \psi_1, \psi_2, \psi_3, h_1, h_2) = \sum_{k,l,m,r} \frac{(k+l+m)!}{k!l!m!} \frac{(k+l+2m+r+h_1+h_2)!}{r!(k+m-r+h_1)!(l+m-r+h_2)!}$$
$$\times (\phi_1\psi_1)^k(\phi_2\psi_2)^l(\phi_3\psi_1\psi_2)^m\left(\frac{\psi_3}{\psi_1\psi_2}\right)^r.$$

Codispersion versus Pearson's Correlation

Definition

Let $X(\boldsymbol{s})$ and $Y(\boldsymbol{s})$ be two spatial processes defined on the set $D \subset \mathbb{R}^2$. Let $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k$, $k \in \mathbb{N}$ be the locations in space. Then, the empirical estimator of the correlation coefficient ρ is

$$\widehat{\rho} = r = \frac{\displaystyle\sum_{i=1}^{k}(X(\boldsymbol{s}_i) - \overline{X})(Y(\boldsymbol{s}_i) - \overline{Y})}{\sqrt{\displaystyle\sum_{i=1}^{k}(X(\boldsymbol{s}_i) - \overline{X})^2 \sum_{i=1}^{k}(Y(\boldsymbol{s}_i) - \overline{Y})^2}},$$

where $\overline{X} = \frac{1}{k}\sum_{i=1}^{k} X(\boldsymbol{s}_i)$, and similarly for $\overline{Y}$.

Codispersion versus Pearson's Correlation

Definition

Let $X(\boldsymbol{s})$ and $Y(\boldsymbol{s})$ be two spatial processes defined on $D \subset \mathbb{R}^2$. Then, for n sampling sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$, the estimated cross-variogram is defined as

$$\widehat{\gamma}_{XY}(\boldsymbol{h}) = \sum_{s \in N(\boldsymbol{h})} (X(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h}))(Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h})),$$

and the empirical codispersion coefficient is

$$\widehat{\rho}_{XY}(\boldsymbol{h}) = \frac{\displaystyle\sum_{s \in N(\boldsymbol{h})} (X(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h}))(Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h}))}{\sqrt{\displaystyle\sum_{\boldsymbol{s} \in N(\boldsymbol{h})} (X(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h}))^2 \sum_{\boldsymbol{s} \in N(\boldsymbol{h})} (Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h}))^2}}$$

where $N(\boldsymbol{h}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : ||\boldsymbol{s}_i - \boldsymbol{s}_j|| \in T(\boldsymbol{h}), 1 \leq i, j \leq n\}$, $T(\boldsymbol{h})$ is a tolerance region around $\boldsymbol{h}$.

Contaminaton Algorithm

Algorithm 1.2.1: Transformation algorithm in the direction $\boldsymbol{h} = (1, 1)$.

input : An image I of size $n \times m$
output: An image W of size $n \times m$

for $i = 1, 2, \ldots, n$ **do**
  for $j = 1, 2, \ldots, m$ **do**
    if $(i == 1)$ *or* $(j == 1)$ **then**
      $W[i, j] \longleftarrow$ Simulation of a standard normal random variable;
    else
      $W[i, j] \longleftarrow W[i-1, j-1] - I[i-1, j-1] + I[i, j]$
  $W \longleftarrow \frac{W - \min W}{\max W - \min W}$;
  return W;

UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Directional Contamination

Figure: (Left) Original image (Lenna); (Right) Image transformed into the direction $\boldsymbol{h} = (1, 1)$.

Directional Contamination

- The correlation coefficient between the images shown before is $\widehat{\rho} = 0.691357$.
- $\widehat{\rho}_{XY}(1,1) = 1$.
- This is not surprising because the differences needed for the computation of the codispersion coefficient are of the form $W[i,j] - W[i+h_1, j+h_2]$, but $W[i+h_1, j+h_2] = W[i,j] - \mathcal{D}$, where $\mathcal{D}$ is the difference associated with pixel (i,j) in the original image, i.e. $\mathcal{D} = I[i,j] - I[i+h_1, j+h_2]$.
- The codispersion coefficient captures the directional correlation between the two images while the usual correlation does not.

Summary

- The codispersion coefficient is an updated version of the usual correlation
$$\rho_{XY}(\boldsymbol{h}) = K(\boldsymbol{h}, \boldsymbol{\theta}) \cdot \rho.$$

- The codispersion coefficient captures the directional correlation
$$\rho_{XY}(\boldsymbol{h}).$$

Day 5