Enhancing reliability of soil sealing indicators by use of geostatistical modeling

Patrick Sillard

Observation and Statistics Service - French Ministry of Environment, Energy and Marine affairs

Nov. 2016

Context of the study

Public policies aiming at

- limiting the impact of human being on nature
- limiting soil consumption (artificialisation, sealing) which has a negative impact on biodiversity and water flow

Role of NSI's

- Computing indicators that makes it possible to follow the impact of public policies in this field
- Being able to qualify statistics computed on various geostatistical sources

Motivation of the paper

Many different Sources : CORINE Land Cover (Geog./HRL), Teruti,. . .

And apparent inconsistencies between published statistics

Sources : CORINE Land Cover (Geog. HRL), Teruti-Lucas survey,*[national geographic databases, land tax databases]*. . .

Source	artificialisation	imperviousness
CORINE-Geog.	5.8%	.
CORINE-HRL		2.8%
Teruti (LUCAS)	9.3%	4.6%

Goal of the study

$\Longrightarrow$ to develop a statistical model that makes all these statistics more consistent : Are there biases or larger standard deviations than expected ?

Teruti-Lucas [TL]

A "classical" statistical survey (European)

- one observation point per km^2 : 3×10^5 points
- the collector observes land occupation and use within a 3m-extent circle (according to a specific classification)
- Systematic sample
- Confidence intervals (CI) are computed for national rates. For example, the national rate for artificial soils is 9.3% and the 99% CI is [9.1% , 9.5%].

RÉPUBLIQUE FRANÇAISE
Ministère de l'Environnement, de l'Énergie et de la Mer

CORINE Land Cover

2 products : CLC-G & CLC-HRL

- A geographic database [CLC-G] made of polygons coherent with respect to land cover, produced at a medium-sized map
- A raster database [CLC-HRL] for imperviousness made of 20m-cells giving the local degree of imperviousness (between 0 and 1) : 1.3×10^9 points

A "classical" geographic database (European)

- the production scale of CLC-G is small, then the contours are rather generalized
- the raster layer describes a 20m-scale phenomenon, not a 2m-scale phenomenon as TL
- national rates are computable, but no CI is published

A geospatial process

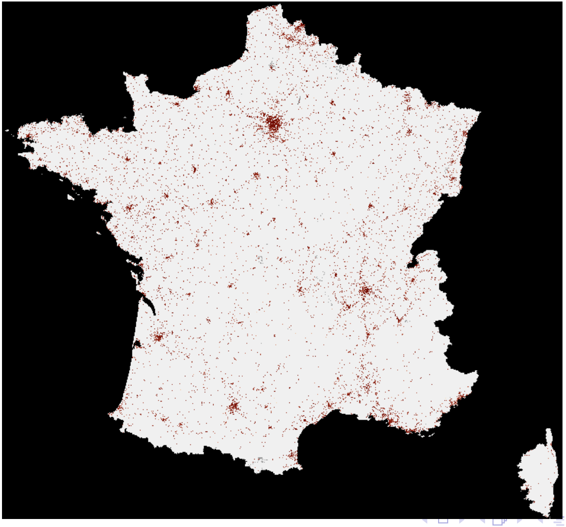
Imperviousness and Artificial lands are local phenomenons

- Highly concentrated, very rare, omnidirectional
- Developing at a very detailed scale : the transition zone between sealing and unsealing soils has an extension smaller than 1m (a few tenth of a meter)

From a probabilistic perspective

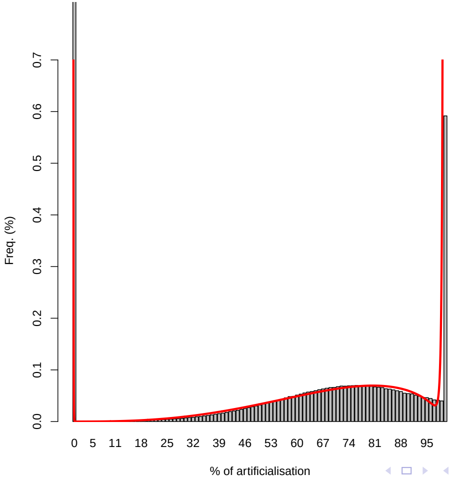
- A stochastic process with a 2-D continuous geographic support
- Rare. Possibly, at a certain scale, be considered as a binary variable. The (local) rate must be a continuous RV over $[0, 1]$
- Strong auto-correlation, isotropic
- Probably stationary

Map of imperviousness raster (CLC-HRL)

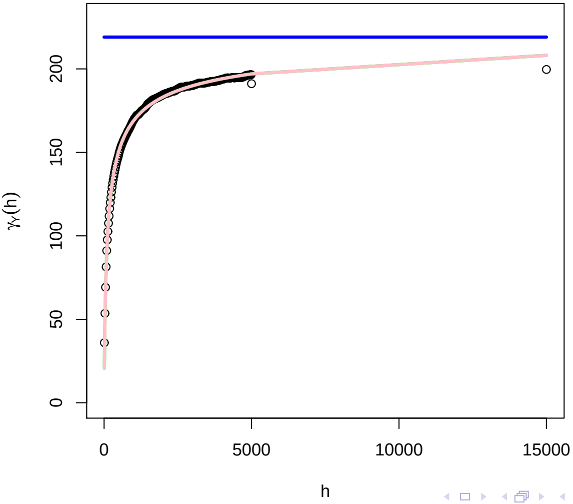


Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Ministère de l'Environnement, de l'Énergie et de la Mer

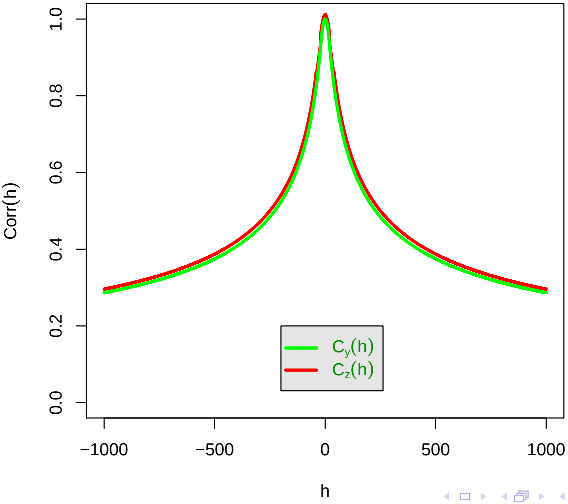
Marginal distribution function of imperviousness raster process (CLC-HRL)



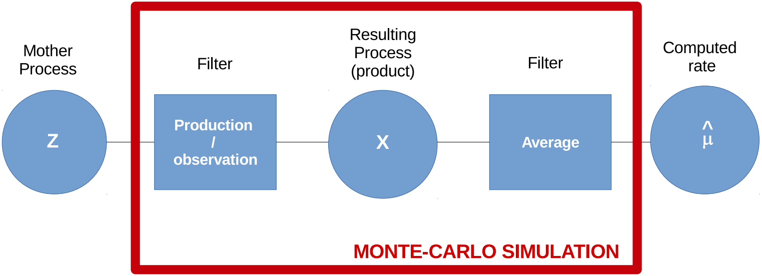
Variogram of imperviousness raster process (CLC-HRL)



Autocorrelation of imperviousness model process



Monte-Carlo simulations based on this statistical model

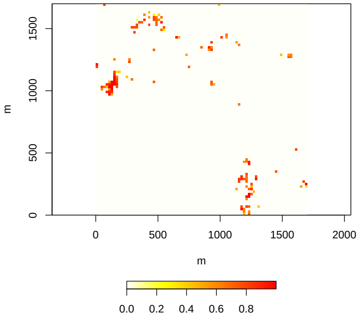


If we are able to draw samples of the process, then :

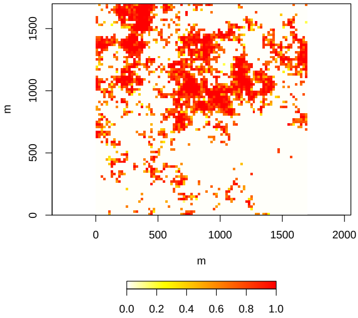
- we can compute the distribution of the estimated rates
- we can compute Confidence Intervals (CI)

I. Imperviousness simulation (CLC-HRL type)

example 1



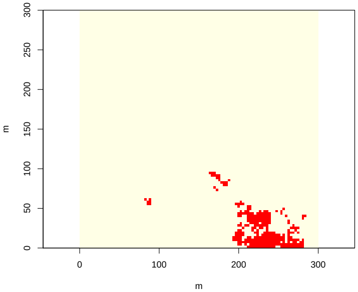
example 2



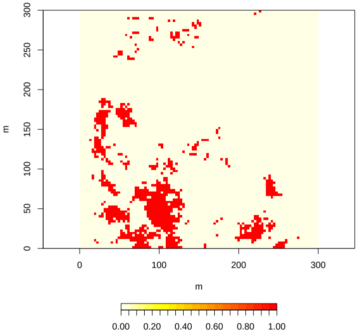
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Ministère
de l'Environnement,
de l'Énergie
et de la Mer

II. Imperviousness simulation (TL type)

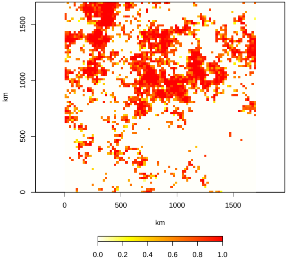
example 1

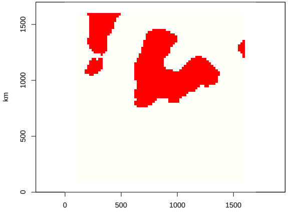


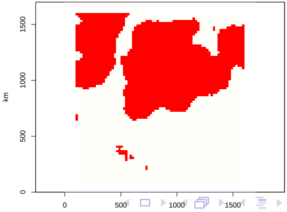
example 2



III. Imperviousness simulation (CLC-G type)







Estimation of rates

- $\mathbf{Y}$: vector of observations of size n
- μ the rate to be estimated
- $\mathrm{var}(\mathbf{Y}) = \sigma^2\Omega$, Ω being the correlation matrix

$$\hat{\mu} = \frac{1}{n}\mathbf{1}^T.\mathbf{Y}$$

$\Omega = I_n$	Any Ω
$\mathrm{var}(\hat{\mu})_0 = \dfrac{\sigma^2}{n}$	$\mathrm{var}(\hat{\mu}) = \dfrac{\sigma^2}{n^2}\sum_{i,j}\Omega_{i,j}$

$$\Rightarrow \mathrm{var}(\hat{\mu}) = \underbrace{\frac{1}{n}\left(\sum_{i,j}\Omega_{i,j}\right)}_{\text{Variance Inflation Factor}} \mathrm{var}(\hat{\mu})_0$$

Summary of rates and Confidence Intervals

Variable	Process		3σ-Confidence Interval	
			Sample mean	
	Mean	St. dev.	(1)	(2)
Imperviousness rate CLC-HRL	2.8%	14.8%	[2.799% ; 2.801%]	[2.1% ; 3.5%]
Imperviousness rate Teruti-Lucas	4.6%	20.1%	[4.9% ; 4.7%]	[3.4% ; 5.8%]
Artificialization rate Teruti-Lucas	9.3%	29.0%	[9.1% ; 9.5%]	[7.6% ; 11.0%]
Artificialization rate CLC-G	5.8%	/	/	/

(1) : without taking into account autocorrelation

(2) : taking into account autocorrelation

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Ministère de l'Environnement, de l'Énergie et de la Mer

Conclusion

Computing statistical indicators over geospatial information

- we must model the mother process and simulate the production process
- Geographical databases look exhaustive but they rely on a transformation of reality and then
  1. The consequences (biases) on the statistics we want to compute must be checked
  2. The distribution of the resulting statistics must be computed
- Classical statistical surveys also have drawbacks. In the case of Teruti-Lucas, two problems arise :
  1. observations points are not independent and the systematic sample design is sub-optimal
  2. Std of rates is underestimated if we don't take into account correlation coming from the underlying process