

L'estimation des taux d'artificialisation et d'imperméabilisation en France

Patrick Sillard

Insee-DMCSI

Juin 2018

Plan

- 1 Introduction
- 2 Les sources de données
- 3 Que mesure-t-on ?
- 4 Modèle statistique & simulations
- 5 Résultats
- 6 Conclusion

Contexte

Des politiques publiques qui visent à

- limiter l'impact de l'homme sur la nature
- limiter la consommation d'espace (artificialisation, imperméabilisation) qui ont un impact négatif sur la biodiversité et l'écoulement des eaux

Expertiser les statistiques de couverture

- Utilité de disposer d'indicateurs qui permettent de suivre l'impact des comportements et des politiques mises en œuvre dans ce champ
- Nécessité de qualifier les statistiques calculées sur différentes sources

But de l'étude

Sources variées : CORINE Land Cover (Geog./HR), Teruti,...

et des incohérences apparentes entre les statistiques produites

Sources : CORINE Land Cover (Geog. HR), enquête Teruti-Lucas, [base géographiques nationales, fichiers fonciers]...

Source	artificialisation	imperméabilisation
CORINE-Geog.	5,8%	.
CORINE-HR		2,8%
Teruti (LUCAS)	9,3%	4,6%

Objectif

⇒ développer un modèle statistique cohérent qui explique le niveau des écarts constatés par la présence de biais ou d'aléas (incertitude)

Plan

- 1 Introduction
- 2 Les sources de données**
- 3 Que mesure-t-on ?
- 4 Modèle statistique & simulations
- 5 Résultats
- 6 Conclusion

Teruti-Lucas [TL]

Une enquête "classique" (cadre européen)

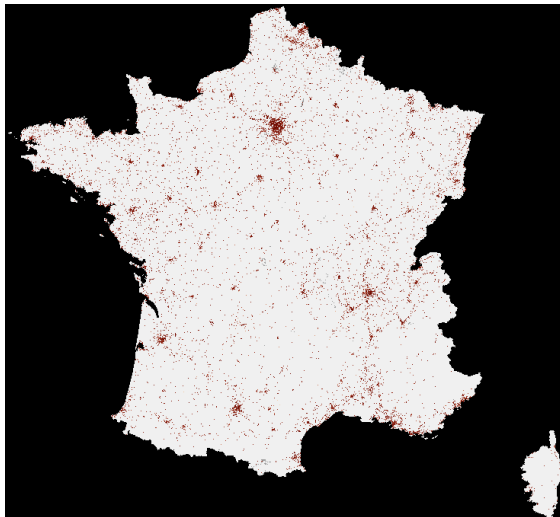
- une observation par km^2 : 309 000 points
- l'enquêteur relève, sur le terrain, l'occupation et l'utilisation du sol dans un cercle de 3m de diamètre (double nomenclature spécifique)
- sondage systématique
- intervalles de confiance (IC) publiés pour les taux nationaux de couverture. Par exemple, le taux national pour les sols imperméabilisés est 4,6% ; $\sigma_{ech} = 0,18pts$.

CORINE Land Cover-HR

Un raster d'1 milliard de cellules

- une base de données raster [CLC-HR] pour l'imperméabilisation constituée de cellules de 20m de côté, donnant le degré d'imperméabilisation (entre 0 et 1) :
 1.3×10^9 cellules
- la couche raster décrit un phénomène d'extension typique de 20m, a priori différent de celui relevé dans Teruti (extension 3m)
- les taux nationaux sont calculables mais aucun IC n'est publié à ce jour

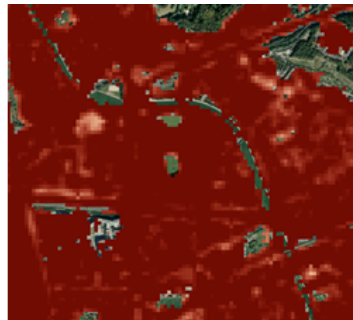
Carte de la couche raster d'imperméabilisation (CLC-HR)



Plan

- 1 Introduction
- 2 Les sources de données
- 3 Que mesure-t-on ?**
- 4 Modèle statistique & simulations
- 5 Résultats
- 6 Conclusion

L'exemple du Stade de France



Rouhaud, 2016

Un autre exemple en milieu rural



Un même phénomène observé à des échelles variées

L'imperméabilisation et l'artificialisation sont des phénomènes localisés

- Très concentrés, rares, isotropes
- Se développent à des échelles très fines : la zone de transition entre des sols imperméables et non imperméables est de quelques cm

De quoi a-t-on besoin ?

- Un objet-modèle qui rend compte de la finesse des échelles en jeu
- Un objet-modèle qui permet de rendre cohérentes les échelles d'observation : le phénomène est unique (commun), mais il est observé à des échelles différentes

Plan

- 1 Introduction
- 2 Les sources de données
- 3 Que mesure-t-on ?
- 4 Modèle statistique & simulations**
- 5 Résultats
- 6 Conclusion

Un processus spatial

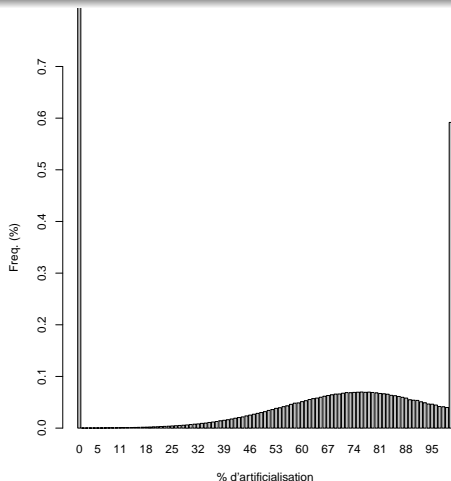
De quoi dispose-t-on pour estimer la structure du processus ? Réponse : de CLC-HR !

- Couverture homogène en termes de contenu et d'incertitude
- Une échelle d'observation relativement fine
- Un volume de données "tractable"

Un modèle probabiliste pour l'imperméabilisation

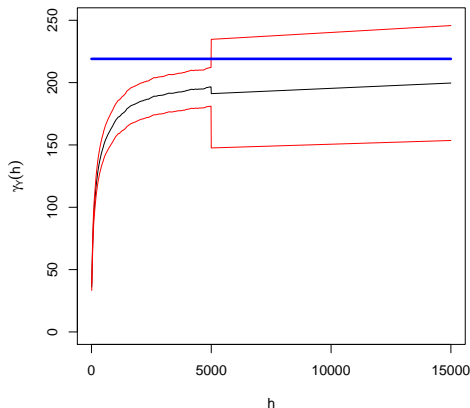
- Processus stochastique à support spatial continu (2D)
- Rare. A certaines échelles, peut être vu comme une variable binaire. Le taux de couverture local est une variable continue sur $[0, 1]$
- Processus stationnaire isotrope (à démontrer) : la corrélation entre deux points ne dépend que de la distance qui les sépare et non de leurs localisations particulières

Histogramme des pixels d'imperméabilisation (CLC-HR)



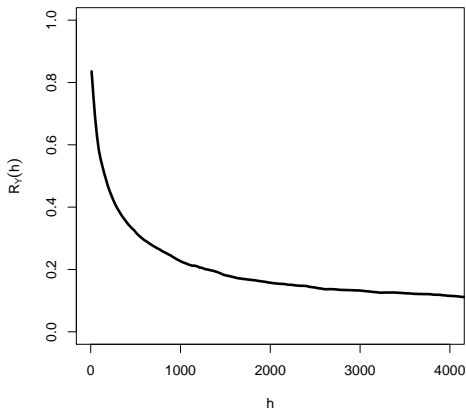
96,2% des pixels comportent un taux d'imperméabilisation nul. La barre de l'histogramme représentée dans la figure en 0 est tronquée pour permettre la représentation des autres modalités d'imperméabilisation qui autrement seraient quasiment confondues avec l'axe des abscisses.

Variogramme empirique des pixels d'imperméabilisation (CLC-HR)



Variogramme estimé empiriquement sur l'ensemble de la France métropolitaine (échelle des abscisses en m). Les courbes en rouges indiquent les intervalles de confiance à deux écarts-types calculés point par point, sans exploiter d'hypothèse de continuité du variogramme.

D'où l'autocorrélation empirique (modèle) des pixels d'imperméabilisation (CLC-HR)



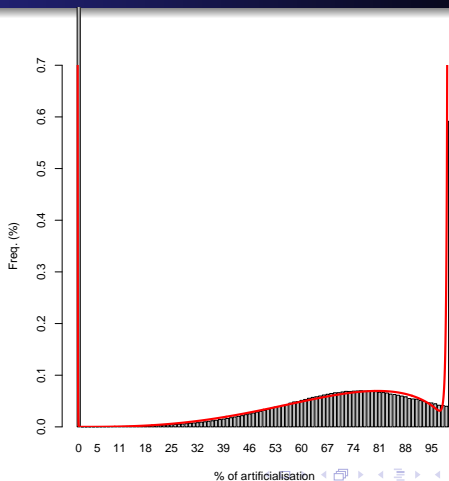
abscisses : distance entre deux points, en m; ordonnées : corrélation (ici positive, entre 0 et 1)

Portée pratique (distance nécessaire pour tomber à 5% de la valeur en 0) = 56km

Modélisation de la distribution empirique par une distribution continue

Composition de trois lois beta :

$f_Y(x) = p_1 \beta_{a_1, b_1}(x) + (1 - p_1 - p_3) \beta_{a_2, b_2}(x) + p_3 \beta_{a_3, b_3}(x)$. Les paramètres p_1 , p_3 , a_2 , b_2 sont estimés par MV.



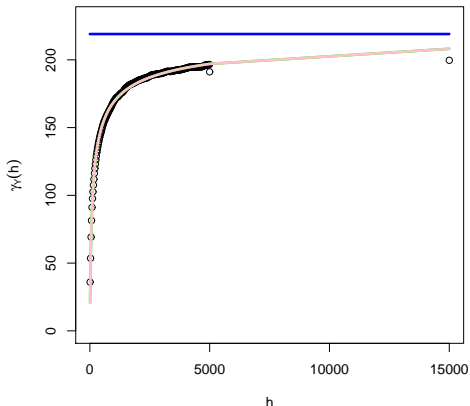
Modélisation de l'autocovariance

Fonction modèle :

$$\hat{C}_Y(h) = A \left(1 + \frac{|h|}{b} \right)^{-\alpha}$$

Estimation de A , b , α par MV :

Paramètre	Valeur
A	227.9 (1.5)
b	38.1 (1.6)
α	0.41 (0.01)



Le processus est **stationnaire à mémoire longue**.

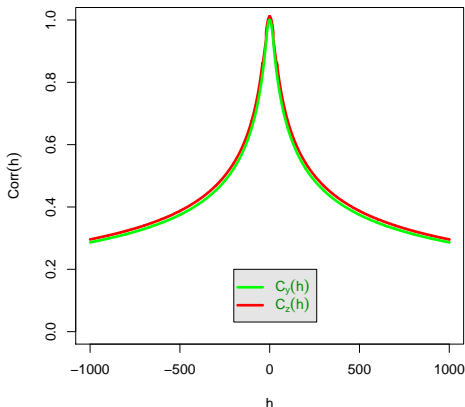
En déduire l'autocovariance du processus métrique

Si Z est le processus métrique (observé par Teruti) et Y est le processus raster moyenné à 20m (i.e. sur $\mathcal{P}(\mathbf{M})$), les deux sont reliés :

$$Y(\mathbf{M}) = \frac{\int_{m \in \mathcal{P}(\mathbf{M})} Z(m) dm}{\int_{m \in \mathcal{P}(\mathbf{M})} dm}$$

et il est possible de déduire l'autovariance de Z connaissant celle de Y .

Notons toutefois qu'on ne connaît pas $C_Z(h)$ pour $h < 20m$.



Principe de simulation

Simuler ?

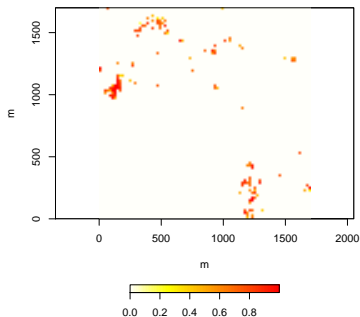
- Z est connu : on dispose de sa densité marginale F et son autocorrélation R

Comment ? → méthode des copules

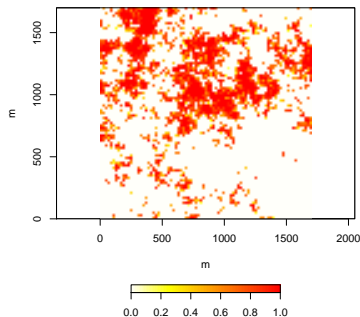
- Soit V , vecteur normal (Ψ) centré-réduit-corrélé (corrélation R). W , de composantes $W_i = F^{-1} \circ \Psi(V_i)$, a pour répartitions marginales F et corrélation R (si F continue).
- Génération de fenêtres de quelques kilomètres de côté

Exemple pour l'imperméabilisation raster (20m – type CLC-HR)

exemple 1



exemple 2



Plan

- 1 Introduction
- 2 Les sources de données
- 3 Que mesure-t-on ?
- 4 Modèle statistique & simulations
- 5 Résultats**
- 6 Conclusion

Variance de quoi ?

Une variable importante

Moyenne du processus Z sur un ensemble (2-D) V :

$$\bar{Z}(V) = \frac{1}{|V|} \int_V Z(\mathbf{x}) d\mathbf{x}$$

La variance de $\bar{Z}(V)$

$$\text{var}(\bar{Z}(V)) = \frac{1}{|V|^2} \int_V \int_V C_Z(\mathbf{u} - \mathbf{v}) d\mathbf{u} d\mathbf{v}$$

A.N. (V - cercle de rayon 417km) :

$$\sqrt{\widehat{\text{var}}(\bar{Z}(V))} = \sqrt{\hat{\phi}_c(T = 417\text{km})} = 2,40\text{pts}$$

Remarque : elle aurait été de $2,3 \times 10^{-5}$ pts en l'absence de corrélation spatiale !

Qu'est-ce que la variance de \bar{Z} ?

- $\text{var}(\bar{Z}(V))$ est la variabilité de la moyenne que l'on obtiendrait si on simulait un grand nombre de fois le processus stochastique générateur de l'imperméabilisation, tel qu'on l'a identifié. C'est la **variance d'estimation** du paramètre $\mathbb{E}Z$.
- On peut aussi considérer la **variance de prédiction** consistant à prédire la réalisation de Z considérée par l'observation d'un échantillon de N points de cette réalisation. Soit $\bar{z}^{(\cdot)} = \frac{1}{N} \sum_{i=1}^N z_i$, la variance de prédiction est :

$$\text{var}(\bar{z}^{(\cdot)} - \bar{Z}(V))$$

La variance d'échantillonnage

Variance de prédiction

Hypothèses simplificatrices : $V =$ cercle de rayon $T = 417km$,
 $N = 309\ 000$ observations élémentaires sur un pavage de cercles de
 $t = 750m$ de rayon \Rightarrow

$$\text{var}(\bar{z}^{(\cdot)} - \bar{Z}(V)) = \frac{1}{N} [\Phi_C(t) - \Phi_C(T)]$$

A.N. :

$$\sqrt{\widehat{\text{var}(\bar{z}^{(\cdot)} - \bar{Z}(V))}} = 0,015\text{pts}$$

\rightarrow Pourrait être utile de simuler le champ pour calculer la variance en tenant compte des vraies formes de V et de celle de la cellule élémentaire d'observation.

Plan

- 1 Introduction
- 2 Les sources de données
- 3 Que mesure-t-on ?
- 4 Modèle statistique & simulations
- 5 Résultats
- 6 Conclusion**

Conclusion

Pour calculer des indicateurs statistiques géospaciaux

- Importance de la modélisation des corrélations pour des processus spatiaux
- Importance de la capacité à simuler
- Différence entre variance d'estimation et variance d'échantillonnage (// notion de superpopulation en sondage)
- Les taux observés pour l'imperméabilisation sont cohérents du point de vue de la variance d'estimation (les processus générateurs sont identiques) mais incohérents du point de vue de la variance de prédiction