

Une méthode d'estimation adaptative de deux dimensions fractales de réseaux de talwegs de ravines

Travail joint avec J.S. Bailly, B. Kaiser, C. Puech et N. Thommeret

Jean-Marc Bardet
bardet@univ-paris1.fr

SAMM, Université Paris 1 (Panthéon-Sorbonne), France

4 juin 2012, MASHS 2012 Paris

Outline

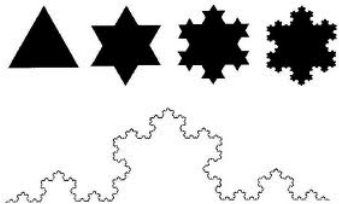
- 1 Introduction
- 2 Deux dimensions fractales "théoriques"
 - Dimension de Hausdorff
 - Dimension de Kolmogorov (ou de Minkowski-Bouligand)
- 3 Deux dimensions fractales "empiriques"
 - Dimension de comptage de boîtes
 - Dimension de corrélation
 - Un problème commun: le choix de l'intervalle de fractalité
- 4 Estimation adaptative de l'intervalle de fractalité
 - Cadre de l'estimation de $\widehat{\dim}_K(A)$
 - Cadre de l'estimation de $\widehat{\dim}_C(A)$
 - Résultats numériques

Plan

- 1 Introduction
- 2 Deux dimensions fractales "théoriques"
 - Dimension de Hausdorff
 - Dimension de Kolmogorov (ou de Minkowski-Bouligand)
- 3 Deux dimensions fractales "empiriques"
 - Dimension de comptage de boîtes
 - Dimension de corrélation
 - Un problème commun: le choix de l'intervalle de fractalité
- 4 Estimation adaptative de l'intervalle de fractalité
 - Cadre de l'estimation de $\widehat{\dim}_K(A)$
 - Cadre de l'estimation de $\widehat{\dim}_C(A)$
 - Résultats numériques

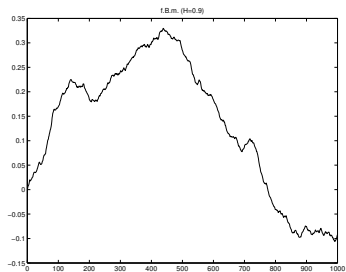
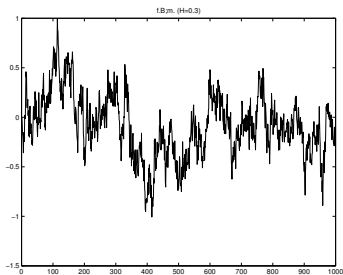
Fractales...

Figure: Deux célèbres exemples de fractales



Fractales...

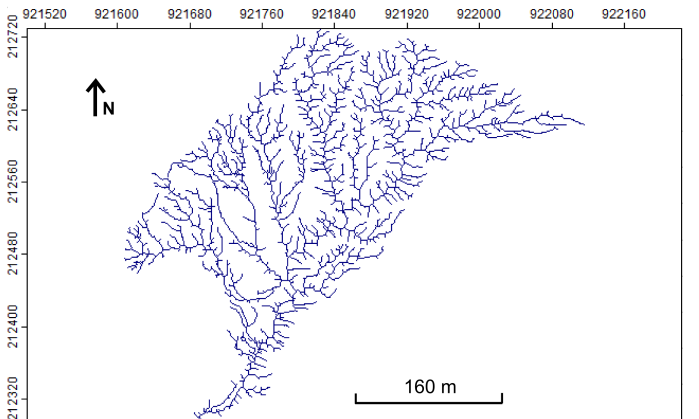
Figure: Autre vision d'une fractale



⇒ Fractale = dimension non entière

Un exemple géographique

Figure: Réseau du Moulin extrait à partir d'un MNT à 1 m (bassins expérimentaux de Draix, système de coordonnées Lambert 3)



Quelques noms associés à la notion de fractale

- Georg Cantor (1845-1918)
- Gaston Julia (1893-1978), puis Adrien Douady (1935-2006),...
- Benoit Mandelbrot (1924-2010)

Objectifs de l'exposé

Problèmes: plusieurs dimensions fractales + difficultés de calcul!

Objectifs:

- 1 Présenter des dimensions fractales "théoriques" et "empiriques"
- 2 Mettre en place une méthode **adaptative** de calcul des dimensions empiriques
- 3 Mesurer la **robustesse** de l'estimation
- 4 Appliquer sur des **fractales simulées** et sur des **réseaux de talwegs de ravines**

Plan

- 1 Introduction
- 2 Deux dimensions fractales "théoriques"
 - Dimension de Hausdorff
 - Dimension de Kolmogorov (ou de Minkowski-Bouligand)
- 3 Deux dimensions fractales "empiriques"
 - Dimension de comptage de boîtes
 - Dimension de corrélation
 - Un problème commun: le choix de l'intervalle de fractalité
- 4 Estimation adaptative de l'intervalle de fractalité
 - Cadre de l'estimation de $\widehat{\dim}_K(A)$
 - Cadre de l'estimation de $\widehat{\dim}_C(A)$
 - Résultats numériques

Définition de la dimension de Hausdorff

Soit A un ensemble de points.

Definition

Pour $r > 0$, et $\sigma = (\sigma_n)_{n \in \mathbb{N}}$ un recouvrement dénombrable de A tel que $\sup_{n \in \mathbb{N}} \text{diam}(\sigma_n) \leq r$, pour tout $\alpha \geq 0$, on appelle définit:

$$m^\alpha(A) = \lim_{r \rightarrow 0} m_r^\alpha(A) \text{ avec } m_r^\alpha(A) = \inf_{\sigma} \sum_n \text{diam}(\sigma_n)^\alpha.$$

La **dimension de Hausdorff** de A , $\dim_H(A)$, vérifie:

$$\dim_H(A) = \sup\{\alpha \in \mathbb{R} \text{ tel que } m_r^\alpha(A) > 0\} = \inf\{\alpha \in \mathbb{R} \text{ tel que } m_r^\alpha(A) = 0\}$$

Quelques propriétés

Propriété

- $\dim(A)$ est un réel positif qui existe pour tout ensemble A
- A fini $\implies \dim(A) = 0$
- Pour une C^1 -sous-variété F de \mathbb{R}^n de dimension $m \leq n$, alors $\dim_H(F) = m$
- Détermination théorique de $\dim(A)$ uniquement possible dans des cas simples (invariance par changement d'échelle, voir Falconer, 1990)

Définition et propriétés

Soit $N(r, A)$ le nombre minimum de boules fermées de rayon r nécessaire pour recouvrir A (on pourrait aussi considérer des pavés de longueur r). Alors:

Definition

La **dimension de Kolmogorov** de A , $\dim_K(A)$ est définie par:

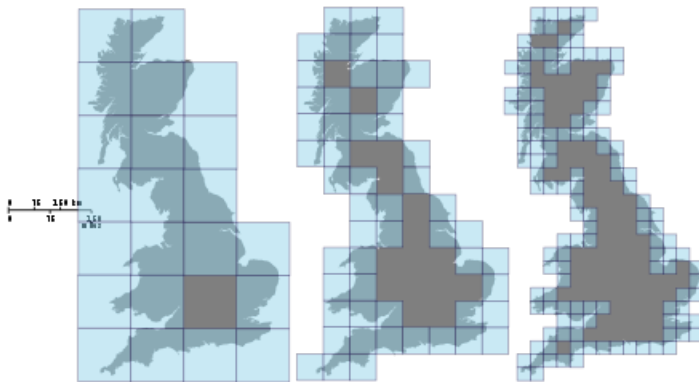
$$\dim_K(A) = \lim_{r \rightarrow 0} \frac{\log[N(r, A)]}{\log(1/r)} \quad (\text{si cette limite existe}).$$

Proposition

- *cette limite n'existe pas toujours \implies limite supérieure et inférieure.*
- *pour tout $F \subset \mathbb{R}^n$, $\dim_H(F) \leq \diminf_K(F) \leq \limsup_K(F)$.*

Un exemple géographique célèbre

Figure: Calcul de la dimension de Kolmogorov de la côte de la Grande-Bretagne



Plan

- 1 Introduction
- 2 Deux dimensions fractales "théoriques"
 - Dimension de Hausdorff
 - Dimension de Kolmogorov (ou de Minkowski-Bouligand)
- 3 Deux dimensions fractales "empiriques"
 - Dimension de comptage de boîtes
 - Dimension de corrélation
 - Un problème commun: le choix de l'intervalle de fractalité
- 4 Estimation adaptative de l'intervalle de fractalité
 - Cadre de l'estimation de $\widehat{\dim}_K(A)$
 - Cadre de l'estimation de $\widehat{\dim}_C(A)$
 - Résultats numériques

Définition de la dimension de comptage de boîtes

A : ensemble dont on observe un **nombre fini** de points ou de segments (arbre)

Definition

Pour $0 < r_1 < r_2 < \dots < r_p$, la **dimension de comptage de boîtes** (**box-counting dimension**) de A , $\widehat{\dim}_K(A)$ est l'estimateur par moindres carrés ordinaires \hat{a} du modèle linéaire:

$$\log[N(r_i, A)] = a * \log(1/r_i) + b \quad \text{pour } i = 1, \dots, p$$

Remarque: si A infini et $r_p \rightarrow 0$, $\widehat{\dim}_K(A) \rightarrow \dim_K(A)$ (si existence).

Définition de la dimension de corrélation

Proposée par Grassberger et Proccacia (1983), version probabiliste (Renyi, 1970)

On suppose $A = \{x_1, \dots, x_N\}$.

Definition

Pour $0 < r_1 < r_2 < \dots < r_p$, la **dimension de corrélation** de A , $\widehat{\dim}_C(A)$ est l'estimateur par moindres carrés ordinaires \hat{a} du modèle linéaire:

$$\log[C(r_i, A)] = a * \log(1/r_i) + b \quad \text{pour } i = 1, \dots, p$$

$$\text{où } C(r, A) = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} \mathbb{1}_{\|x_i - x_j\| \leq r}$$

Remarque: estimation de $\widehat{\dim}_C(A)$ (si existence) où $\frac{\log(P(r, A))}{\log(1/r)} \rightarrow \widehat{\dim}_C(A)$ ($r \rightarrow 0$) avec $P(r, A)$ proportion de couples de A dont interdistance $\leq r$.

Relations entre dimensions

Proposition

Pour A ensemble, $\dim_H(A) \leq \dim_C(A) \leq \dim_K(A)$ (si existence).

Remarque: On aura donc souvent $\widehat{\dim}_C(A) \leq \widehat{\dim}_K(A)$, mais pas toujours...

Comment calculer $\widehat{\dim}_C(A)$ et $\widehat{\dim}_K(A)$

Problème: Comment choisir (r_1, \dots, r_p) ?

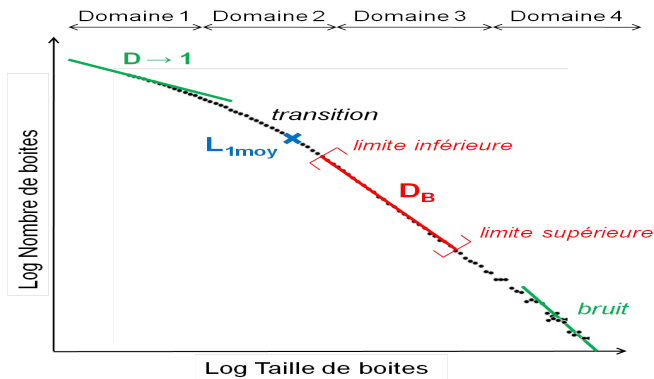
Comme A est fini:

- Si les r_i sont trop "petits", $\widehat{\dim}_K(A) \simeq 1$ et $C(r_i, A) = 0$
- Si les r_i sont trop "grands", $N(r_i, A) \simeq 1$ et $C(r_i, A) \rightarrow 1$ (sautent par paliers)

Objectif: Trouver la "bonne zone" pour (r_1, \dots, r_p) : [intervalle de fractalité](#)

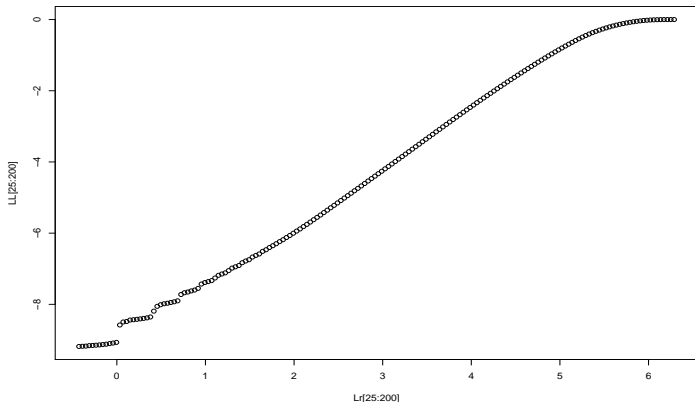
Exemple sur le réseau de talwegs pour $\widehat{\dim}_K(A)$

Figure: Nuage de points $(\log(1/r_i), \log(N(r_i, A)))$



Exemple sur le réseau de talwegs pour $\widehat{\dim}_C(A)$

Figure: Nuage de points $(\log(1/r_i), \log(C(r_i, A)))$



Plan

- 1 Introduction
- 2 Deux dimensions fractales "théoriques"
 - Dimension de Hausdorff
 - Dimension de Kolmogorov (ou de Minkowski-Bouligand)
- 3 Deux dimensions fractales "empiriques"
 - Dimension de comptage de boîtes
 - Dimension de corrélation
 - Un problème commun: le choix de l'intervalle de fractalité
- 4 Estimation adaptative de l'intervalle de fractalité
 - Cadre de l'estimation de $\widehat{\dim}_K(A)$
 - Cadre de l'estimation de $\widehat{\dim}_C(A)$
 - Résultats numériques

Trois zones de linéarités différentes

Trois zones se distinguent assez nettement:

- Une zone linéaire avec $a \simeq 1$
- Une zone linéaire avec $a = \widehat{\dim}_K(A)$
- Une zone linéaire avec $a > 0$

Détection de ruptures dans modèles linéaires

On utilise le contraste suivant:

Pour $t_1 < t_2$ des entiers positifs:

$$\begin{aligned}(\widehat{t}_1, \widehat{t}_2) = \operatorname{Argmin}_{1 < t_1 < t_2 < p} & \left(\sum_{t=1}^{t_1} (\log(N(r_t)) - \widehat{a}_0 + \log(r_t))^2 \right. \\ & \left. + \sum_{t=t_1+1}^{t_2} (\log(N(r_t)) - \widehat{b}_0 - \widehat{b}_1 \log(r_t))^2 + \sum_{t=t_2+1}^p (\log(N(r_t)) - \widehat{c}_0 - \widehat{c}_1 \log(r_t))^2 \right)\end{aligned}$$

où \widehat{a}_0 (resp. $(\widehat{b}_0, \widehat{b}_1)$ et $(\widehat{c}_0, \widehat{c}_1)$) est l'estimateur MCO sur $\{1, \dots, t_1\}$ (resp. $\{t_1 + 1, \dots, t_2\}$ et $\{t_2 + 1, \dots, p\}$)

Remarque théorique

Proposition

D'après Bai et Perron (1998) ou Lavielle et Moulines (2000), sous certaines hypothèses assez faibles,

$$\|(\widehat{t}_1, \widehat{t}_2) - (t_1^*, t_2^*)\| = O_P(1).$$

Une "grande zone linéaire"

On observe:

- Une "grande zone linéaire" centrale avec $a = \widehat{\dim}_C(A)$
- Pour r petit, la même linéarité plus bruitée
- Pour r grand, une rupture "douce" (non abrupte)

⇒ Difficultés pour utiliser la même détection que pour $\widehat{\dim}_K(A)$

Technique de type sélection de modèles

On suppose le modèle:

$$Y_t = f(t) + \varepsilon_t \quad t \in \{1, \dots, p\}$$

avec $f(t) = at + b$ pour $t \in \{t_1^* + 1, \dots, t_2^*\}$.

Avec $C > 2$, on pose le contraste pénalisé suivant:

$$J(t_1, t_2) = \frac{1}{t_2 - t_1 - 2} \sum_{t=t_1+1}^{t_2} \widehat{\varepsilon}_t^2 \left(1 + \frac{C}{\sqrt{t_2 - t_1}}\right)$$

$$\implies (\widehat{t}_1, \widehat{t}_2) = \operatorname{Argmin}_{1 < t_1 < t_2 < p} J(t_1, t_2)$$

Remarque: En pratique, nous avons choisi $C = \sqrt{p}$ (on suppose $t_1 = \lceil \tau_1 n \rceil$ et $t_2 = \lceil \tau_2 n \rceil$).

Résultats sur le réseau de talwegs de ravines

	$\widehat{\dim}_K(A)$	Intervalle de fractalité	R^2	$\widehat{\dim}_C(A)$	Intervalle de fractalité
Moulin 1 m	1.70	[5.3 m, 44.3 m]	0.9998	1.81	[10.6 m, 32.2 m]
Moulin 2 m	1.57	[8.7 m, 40.5 m]	0.9997	1.81	[15.3 m, 41.3 m]

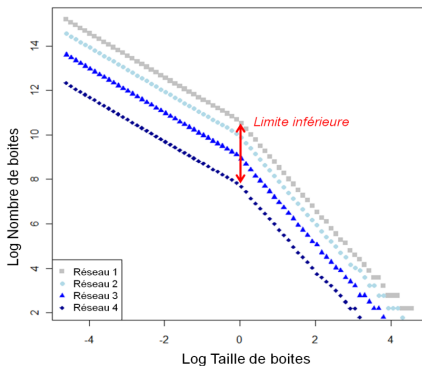
Résultats pour des arbres simulés

Pour des arbres binaires bruités avec angle de 60° et rapport $1/\sqrt{3} \pm 0.03$:

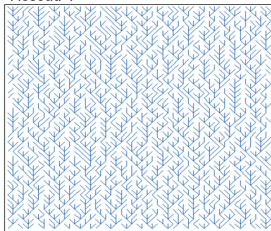
Ordre	$\widehat{\dim}_C(A)$	IC $\widehat{\dim}_C(A)$	$\widehat{\dim}_K(A)$	IC $\widehat{\dim}_K(A)$
5	1.27	[1.14-1.40]	1.02	[0.97-1.07]
6	1.31	[1.24-1.41]	1.01	[1.08-1.29]
7	1.38	[1.28-1.45]	1.14	[1.05-1.22]
8	1.40	[1.35-1.43]	1.33	[1.29-1.37]
9	1.41	[1.38-1.43]	1.34	[1.31-1.37]
10	1.41	[1.39-1.43]	1.42	[1.35-1.49]

Sur des arbres simulés: réseau de Scheidegger

Figure: Nuage de points $(\log(1/r_i), \log(N(r_i, A)))$



Réseau 4



Résultats

Ordre	$\widehat{\dim}_C(A)$	IC $\widehat{\dim}_C(A)$	$\widehat{\dim}_K(A)$	IC $\widehat{\dim}_K(A)$
3	1.26	[0.97-1.53]	2.28	[1.60-2.96]
4	1.39	[1.10-1.74]	1.68	[1.50-1.86]
5	1.47	[1.24-1.71]	1.77	[1.73-1.81]
6	1.67	[1.49-1.77]	1.78	[1.73-1.83]
7	1.80	[1.79-1.81]	1.85	[1.85-1.86]