

Une méthode d'analyse de séries irrégulières.

Emmanuel Le Tavernier et Patrice Simard

UMR CNRS HEUDIASYC 6599,
Université de Technologie de Compiègne
Centre de Recherches de Royallieu
B.P. 20529, 60205 Compiègne Cedex, France

RÉSUMÉ

Nous proposons une approche destinée à l'analyse de séries irrégulières. Cette approche conduit à une représentation régularité-“résolution” de la série. Elle analyse aussi bien les séries régulières qu'irrégulières et permet surtout de révéler la présence et la localisation de certaines propriétés de cette série comme l'auto-affinité ainsi que la présence de différents modes.

ABSTRACT

This paper suggests a time series analysis approach. It allows to lead to a regularity-scale representation of the series. It can be used for regular and irregular time series and shows up some properties of the series such as self-affinity, and “multi-affinity” in scale.

1 Introduction.

De nombreuses publications ont fourni ces dernières années des estimateurs de régularités liés à la dimension fractale d'une série temporelle [1, 5]. Ces estimateurs sont pour la plupart dérivés d'un modèle de fonctions ou de processus particuliers, dont ils exploitent une propriété. En général, ils demandent d'évaluer une régression linéaire sur un ensemble de points dont l'utilisateur doit au préalable retirer quelques échantillons aux extrémités.

Nous proposons une approche simple pour l'étude et l'analyse d'une série temporelle ne demandant aucune propriété fractale. La méthode qui en découle peut s'appliquer à tout type de série aussi bien régulière qu'irrégulière, sous réserve que la trajectoire du processus étudié soit continue et höldérienne. Elle permet d'apprécier la structure de la série en fonction de sa résolution et d'en déduire les propriétés de régularité et d'auto-affinité du processus générateur.

2 Principe d'analyse.

Soit X un processus stochastique auto-affine de paramètre α et à accroissements stationnaires. Il vérifie donc la propriété suivante : $\forall t, r > 0$,

$$X(rt) \stackrel{d}{=} r^\alpha X(t) \quad (1)$$

où $\stackrel{d}{=}$ désigne l'égalité au sens des distributions.

$$\begin{aligned} (1) \Rightarrow X(t+\varepsilon) - X(t) &\stackrel{d}{=} \varepsilon^\alpha \left[X\left(\frac{t}{\varepsilon} + 1\right) - X\left(\frac{t}{\varepsilon}\right) \right] \\ \Rightarrow |X(t+\varepsilon) - X(t)| &\stackrel{d}{=} \varepsilon^\alpha |X\left(\frac{t}{\varepsilon} + 1\right) - X\left(\frac{t}{\varepsilon}\right)| \\ \Rightarrow E[|X(t+\varepsilon) - X(t)|] &= \varepsilon^\alpha E\left[|X\left(\frac{t}{\varepsilon} + 1\right) - X\left(\frac{t}{\varepsilon}\right)|\right] \end{aligned}$$

$$\Rightarrow E[|X(t+\varepsilon) - X(t)|] \propto \varepsilon^\alpha. \quad (2)$$

Soit maintenant une fonction x continue et höldérienne de paramètre β sur son support, et que l'on suppose provenir d'un processus Y inconnu. Il existe donc une constante C tel que $\forall t, \forall \varepsilon$,

$$|x(t+\varepsilon) - x(t)| \leq C|\varepsilon|^\beta. \quad (3)$$

Soit $\overline{|x(t+\varepsilon) - x(t)|}$ la moyenne sur les instants t des modules des accroissements de largeurs ε de la fonction x , alors

$$\begin{aligned} (3) \Rightarrow \overline{|x(t+\varepsilon) - x(t)|} &\leq C|\varepsilon|^\beta \\ \Rightarrow \overline{|x(t+\varepsilon) - x(t)|} &\propto |\varepsilon|^{\alpha(\varepsilon)}. \end{aligned} \quad (4)$$

où $\alpha(\varepsilon)$ est une fonction telle que $\forall \varepsilon$

$$|\varepsilon|^{\alpha(\varepsilon)} \leq |\varepsilon|^\beta.$$

Si le processus Y est auto-affine de paramètre α . Alors, sous réserve d'ergodicité, les équations (2) et (4) impliquent

$$\overline{|x(t+\varepsilon) - x(t)|} \propto |\varepsilon|^\alpha$$

et donc que $\forall \varepsilon, \alpha(\varepsilon) = \alpha$.

Si par contre le processus n'est pas un processus auto-affine pur, l'étude de la fonction $\alpha(\varepsilon)$ peut permettre de donner des informations sur la structure du processus. Par exemple, si Y est la somme d'un processus auto-affine et d'un bruit blanc, il existe un voisinage de 0 où la fonction $\alpha(\varepsilon)$ tend vers 0.

Remarque. La limite de $\alpha(\varepsilon)$ quand ε tend vers 0 est un indice de régularité moyen de la série. En effet,

$$(4) \Rightarrow \frac{\overline{|x(t+\varepsilon) - x(t)|}}{|\varepsilon|} \propto |\varepsilon|^{\alpha(\varepsilon)-1}$$

Si x est une fonction régulière, alors

$$\lim_{\varepsilon \rightarrow 0} \frac{\overline{|x(t+\varepsilon) - x(t)|}}{|\varepsilon|} = A$$

où A est une constante.

$$\Rightarrow \lim_{\varepsilon \rightarrow 0} \alpha(\varepsilon) - 1 = 0 \Rightarrow \lim_{\varepsilon \rightarrow 0} \alpha(\varepsilon) = 1.$$

Si au contraire, x est irrégulière et donc non dérivable, alors

$$\lim_{\varepsilon \rightarrow 0} \frac{|x(t + \varepsilon) - x(t)|}{|\varepsilon|} = +\infty$$

$$\Rightarrow \lim_{\varepsilon \rightarrow 0} \alpha(\varepsilon) - 1 < 0 \Rightarrow \lim_{\varepsilon \rightarrow 0} \alpha(\varepsilon) < 1.$$

La continuité de x implique $\lim_{\varepsilon \rightarrow 0} \alpha(\varepsilon) > 0$.
 Donc si x est régulière, alors $\lim_{\varepsilon \rightarrow 0} \alpha(\varepsilon) = 1$, alors que si x est irrégulière, $\lim_{\varepsilon \rightarrow 0} \alpha(\varepsilon) \in]0, 1[$.

Dans le cas d'un processus auto-affine pur, les paramètres d'auto-affinité et de régularité sont égaux.

Nous proposons de faire une étude de la fonction $\alpha(\varepsilon)$.

3 Etude de la fonction $\alpha(\varepsilon)$ pour une série.

Soit une série Z de N points, échantillonnée à la période T_e . On suppose que, Z est la représentation échantillonnée d'une fonction continue x tel que $Z[i] = x(i.T_e)$.

On pose $L_k(Z) = (1/(N - k)) \sum_{i=0}^{N-k-1} |Z[i + k] - Z[i]|$.

Nous allons étudier la fonction $\alpha(k)$ tel que

$$L_k(Z) \propto k^{\alpha(k)}. \tag{5}$$

L'équation (5) correspond à peu de choses près à celle introduite par Higuchi dans [2], mais ce dernier suppose une auto-affinité pure du processus générateur et donc pose la fonction $\alpha(\varepsilon)$ constante.

Pour tout couple (k, k') tel que $k < k'$ et dont on connaît les quantités $L_k(Z)$ et $L_{k'}(Z)$, nous calculons le paramètre $\alpha_{k,k'}$ tel que

$$\alpha_{k,k'} = \frac{\log L_k(Z) - \log L_{k'}(Z)}{\log k - \log k'}.$$

Remarque. Si l'on considère Z comme étant de résolution 1, on définit Z_p , Z à la résolution p , la sous série de Z constituée des échantillons $Z_p[i] = Z[i.p]$.

Les coefficients $\alpha_{p.k,p.k'}$ permettent une étude de la série à la résolution p .

3.1 Représentation des $\alpha_{k,k'}$.

Pour visualiser de façon simple la distribution des coefficients $\alpha_{k,k'}$ de la série, nous proposons de représenter sur un même graphe les nuages de points $(\alpha_{k,k'}, k)$ et $(\alpha_{k,k'}, k')$ pour toutes valeurs de k et de k' . Dans certains cas et notamment pour l'étude de processus stochastiques, nous représenterons la densité locale normalisée de ces nuages de points, c'est à dire pour chaque valeur de k , l'histogramme des $\alpha_{k,k'}$ avec le maximum égal à 1.

4 Applications.

Nous testons notre méthode sur différentes séries et nous comparons les résultats avec ceux obtenus par la méthode de Higuchi.

4.1 Signaux réguliers.

Soit Z la série telle que $Z[i] = e^{i.T_e}$ avec $i.T_e \in [0, 1]$. Cette série est l'échantillonnage d'une fonction régulière. La figure 1 représente la distribution des $\alpha_{k,k'}$ en fonction de k et de k' . La droite verticale représente la valeur estimée par la méthode de Higuchi, quant à la courbe, c'est la moyenne sur tous les k' des $\alpha_{k,k'}$.

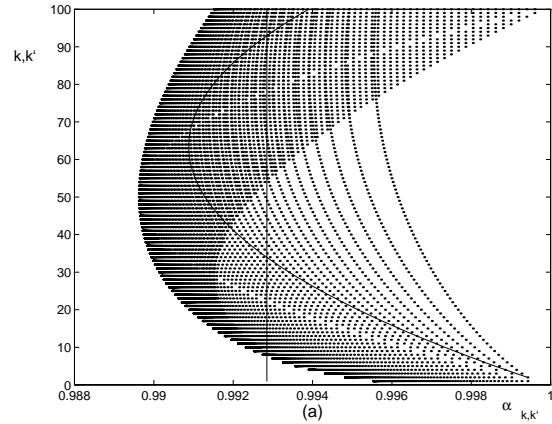


FIG. 1 — Application sur la série $Z[i] = e^{i.T_e}$

Sur une telle série, la méthode de Higuchi donne une valeur du paramètre d'auto-affinité d'environ 0.993 et donc tend à faire croire que la série est fractale. La représentation des $\alpha_{k,k'}$ permet de visualiser plusieurs informations sur la série étudiée. On y voit qu'elle n'est pas fractale et donc le processus non auto-affine. En effet, les $\alpha_{k,k'}$ ne se stabilisent pas dans le voisinage d'une valeur de α particulière. Au sujet de la régularité de la série, il apparaît clairement que lorsque l'on tend vers la résolution 0 de la série, donc quand les valeurs de k et de k' tendent vers 0, les coefficients $\alpha_{k,k'}$ convergent vers 1 ce qui signifie que la série est régulière.

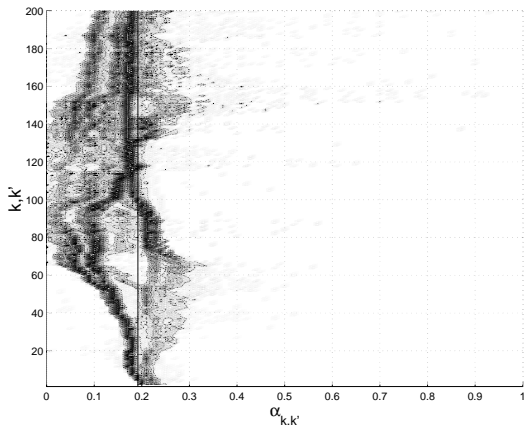
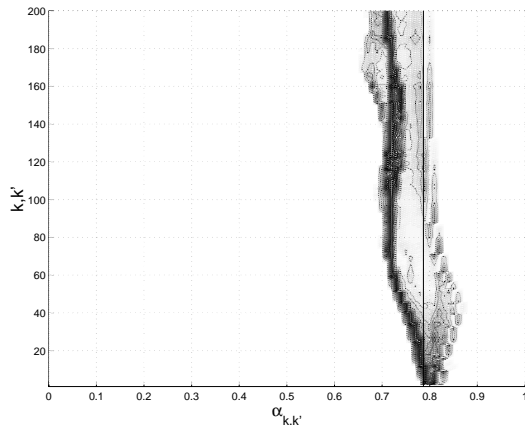
4.2 Signaux fractals.

Les applications suivantes vont être faites sur des modèles de mouvements Brownien fractionnaires [4], les ARIMA(0,d,0) [3]. Les mouvements Brownien fractionnaires ont un paramètre d'auto-affinité H qui est lié au paramètre d des ARIMA par la formule $H = d + 1/2$.

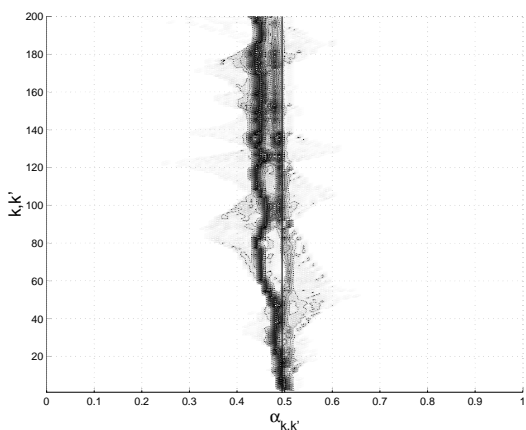
La figure 2 représente la densité locale des points $(\alpha_{k,k'}, k)$ et $(\alpha_{k,k'}, k')$ pour un ARIMA de paramètre $H = 0.2$. La droite verticale indique la valeur estimée par la méthode de Higuchi. Les mêmes résultats mais obtenus sur la réalisation d'un ARIMA, dont $H = 0.5$, sont représentés sur la figure 3 et ceux pour un ARIMA de paramètre $H = 0.8$ sont sur la figure 4.

On remarque, sur chacune de ces trois figures, que la régularité théorique du signal étudié est atteinte lorsque l'on tend vers la résolution 0. Il apparaît aussi clairement des dominantes fractales.

La figure 2 révèle une grande dispersion des coefficients $\alpha_{k,k'}$ comparée à celles des deux autres figures. On peut tout de

FIG. 2 — Application sur un ARIMA avec $H = 0.2$.FIG. 4 — Application sur un ARIMA avec $H = 0.8$.

même remarquer sur cette figure une dominante d'auto-affinité légèrement inférieure à 0.2 lorsque k est supérieur à 120 ainsi que pour k proche de 20.

FIG. 3 — Application sur un ARIMA avec $H = 0.5$.

Sur la figure 3, résultats obtenus à partir d'un ARIMA de régularité 0.5, on distingue nettement une auto-affinité de 0.45 pour les valeurs de k supérieures à 60, mais aussi une deuxième zone auto-affine de 0.5 pour les k compris entre 0 et 20. La série générée par cet ARIMA possède donc deux modes bien distincts qui dépendent de la résolution à laquelle on étudie la série.

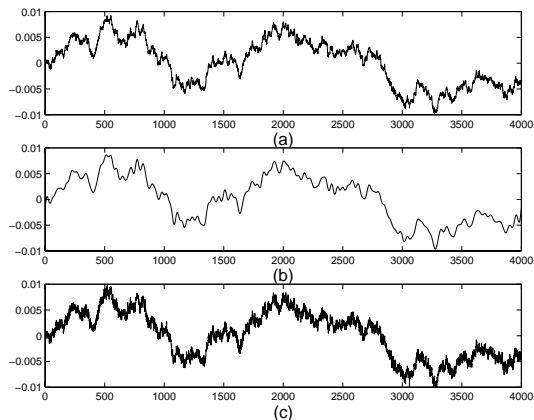
La figure 4, quant à elle met en évidence une localisation des $\alpha_{k,k'}$ autour de 0.7 pour les valeurs de k supérieures à 60 puis les $\alpha_{k,k'}$ tendent vers 0.8 lorsque k tend vers 0. La distribution des $\alpha_{k,k'}$ est particulièrement régulière.

Les résultats obtenus avec la méthode de Higuchi pour les trois séries étudiées peuvent être considérés comme valables puisqu'ils sont proches des valeurs théoriques. Calculés à partir de la régression linéaire de la représentation log-log des coefficients $L_k(Z)$, les résultats obtenus avec la méthode de Higuchi tiennent comptes de toutes les valeurs de $L_k(Z)$ y compris les possibles valeurs aberrantes. La régression faite sur une échelle log-log "pondère" chacune des valeurs de

$L_k(Z)$ en favorisant les termes de faibles valeurs de k . Notre méthode ne met en évidence que les zones où la majorité des coefficients $\alpha_{k,k'}$ sont localisée. Les $\alpha_{k,k'}$ provenant des valeurs aberrantes de $L_k(Z)$ ne sont donc pas visibles. Les informations contenues sur ces trois figures apportent des éléments d'analyses beaucoup plus pertinents et permettent de mieux localiser les propriétés d'auto-affinités du signal.

4.3 Signaux lissés et bruités.

Nous allons étudier les représentations que l'on obtient avec notre méthode sur des signaux lissés puis sur des signaux bruités.

FIG. 5 — (a) Trajectoire d'un ARIMA $H=0.5$, (b) même trajectoire lissée, (c) même trajectoire bruitée.

Nous avons repris le signal ARIMA de paramètre $H = 0.5$ étudié précédemment et représenté sur la figure 5(a). Dans un premier temps, nous avons fait subir à ce signal un lissage. Le signal obtenu est représenté sur la figure 5(b). La figure 6 représente la densité locale des $\alpha_{k,k'}$ obtenus à partir de ce signal.

Le lissage affectant les hautes fréquences, il reste les composantes basses fréquences de l'ARIMA. On retrouve donc sur la figure 6 une zone d'auto-affinité de 0.5 pour les valeurs de

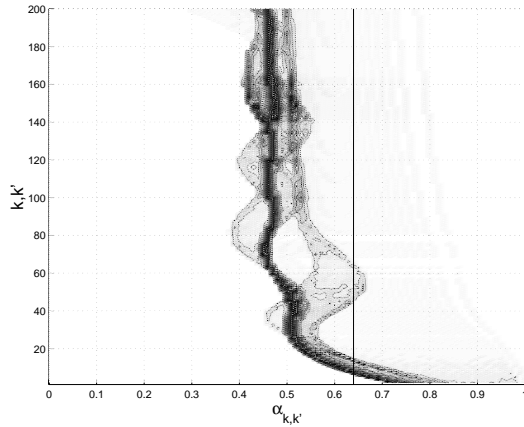


FIG. 6 — Localisation des $\alpha_{k,k'}$ pour l'ARIMA lissé.

k supérieures à 60. Sur le signal d'origine cette zone se situait aux environs de 0.45.

Le lissage a surtout eu des répercussions sur la régularité du signal. Ceux sont les coefficients $\alpha_{k,k'}$ des basses résolutions qui ont subi les modifications. En effet, il y a une déviation vers 1 des $\alpha_{k,k'}$ lorsque la résolution tend vers 0. Le signal tend à être régulier ce qui est caractéristique d'un signal lissé.

Dans un deuxième temps, nous avons bruité la même réalisation de l'ARIMA 0.5, le signal obtenu est représenté sur la figure 5(c). La figure 7 représente la localisation des $\alpha_{k,k'}$ pour ce signal bruité.

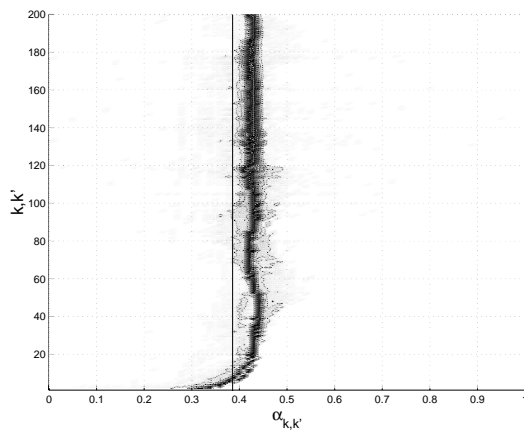


FIG. 7 — Localisation des $\alpha_{k,k'}$ pour l'ARIMA bruité.

On retrouve une zone d'auto-affinité pour les valeurs de k supérieures à 40 ainsi qu'une nouvelle déviation des $\alpha_{k,k'}$ pour les basses résolutions. Mais cette fois ci, la déviation tend vers 0 lorsque les résolutions diminuent. Le bruitage ajoute de l'irrégularité sur les hautes fréquences du signal. Une telle déviation signifie qu'un processus de régularité inférieure à 0.4 s'ajoute au processus auto-affine dont le paramètre est supérieur à 0.4.

Sur de tels signaux la méthode de Higuchi ne permet pas de mettre en évidence la tendance fractale du signal et la tendance

de lissage ou de bruitage. Elle évalue des tendances "moyennes" qui ne correspondent à aucune des caractéristiques du signal. Bien entendu, si l'on se restreint à une des zones d'auto-affinité du signal, la méthode de Higuchi donnera la valeur d'auto-affinité de cette zone. Mais sans la connaissance de la présence et de la localisation de ces zones, il peut être hasardeux de se fier aux résultats obtenus par la méthode de Higuchi.

5 Conclusion

La méthode que nous proposons a comme intérêt de visualiser de manière différente le signal étudié. On observe et localise les plages de résolutions du signal où apparaissent des propriétés d'auto-affinité, ainsi que les tendances de régularité de la série. On a pu voir avec les deux derniers exemples, que la fractalité et l'auto-affinité peuvent n'être visibles qu'à certains niveaux de résolutions du signal et donc une représentation telle que nous la proposons est tout à fait utile pour une bonne analyse des séries temporelles.

Références

- [1] K.J. Falconer. *Fractal Geometry : mathematical foundations and applications*. Wiley, 90.
- [2] T. Higuchi. "Approach to an irregular time series on the basis of the fractal theory". *Physica D*, Vol.31 :277–283, 88.
- [3] J. Kasdin. "Discrete simulation of colored noise and stochastic processes and $1/f^\alpha$ power law noise generation". *Proc. IEEE*, Vol.83(5) :802–827, 95.
- [4] B.B. Mandelbrot. *The fractal geometry of nature*. W.H. Freeman, 82.
- [5] C. Tricot. *Courbes et dimension fractale*. Springer-Verlag, 93.