

Analyse des séries temporelles

Didier Delignières

Septembre 2007

1. Définitions

Les statistiques classiques opèrent le plus souvent par comparaison de moyennes (notamment par Analyse de variance) ou par l'étude des relations entre variables (par l'analyse des corrélations). Le concept d'échantillon est central dans cette approche. Un échantillon est un ensemble de données collectées sur des « sujets » différents, mais supposés issus d'une même population parente. Ceci signifie qu'au regard de la variable mesurée, les sujets ne diffèrent pas au-delà « d'erreurs » aléatoires (liées par exemple à l'histoire des sujets, à divers facteurs non maîtrisés dans la standardisation expérimentales, à de l'erreur de mesure, ect.).

En d'autres termes, chaque valeur x_i de l'échantillon considéré peut être exprimé ainsi :

$$x_i = \mu + \varepsilon_i \quad (1)$$

μ représentant le « valeur vraie » de la variable mesurée, et ε_i une variable aléatoire centrée distribuée normalement. Dans la mesure où la moyenne de ε_i est nulle, on conçoit que la moyenne théorique de l'échantillon est μ (c'est à dire que la moyenne empirique tend à se rapprocher asymptotiquement de μ lorsque l'effectif de l'échantillon augmente). On le voit, les statistiques classiques reposent sur des hypothèses fortes, et la vérification de la normalité de la distribution des échantillons n'est pas superflue. A noter que dans un échantillon, l'ordre dans lequel se présentent les données n'a aucun effet sur les résumés statistiques (moyenne ou écart-type notamment).

Les statistiques classiques visent donc à approcher une valeur vraie, supposée stable et partagée par un ensemble de « sujets », en multipliant les mesures supposées indépendantes. Les statistiques en séries temporelles vont plutôt tenter d'analyser l'évolution dans le temps d'une variable donnée. Il s'agira donc de mesurer de manière répétée un phénomène. Si les mesures sont réalisées de manière régulière, on parlera de séries temporelles. Chaque valeur est alors distante de la précédente d'un intervalle de temps constant.

Fréquemment cependant les mesures sont réalisées non à l'échéance d'un intervalle temporel fixé à l'avance, mais à l'occurrence d'un événement donné. C'est le cas par exemple dans les expériences de tapping, où l'on collecte des séries d'intervalles inter-tappes : les valeurs ainsi collectées demeurent ordonnées, mais les valeurs successive ne correspondent plus à des dates absolues définies par une fréquence d'acquisition.

Séries temporelles et séries événementielles sont le plus souvent traitées indifféremment dans la littérature. Ce qui ne va pas sans poser problème, par exemple dans l'application des analyses spectrales (que signifie la fréquence dans une série événementielle ?). Certains travaux ont montré cependant que la construction de séries événementielles apportait des informations plus riches sur le système considéré que l'utilisation de séries temporelles (voir Delignières, Torre & Lemoine, 2005 ; Torre, Delignières & Lemoine, 2007).

Il existe deux catégories de modèles pour rendre compte d'une série temporelle. Les premiers considèrent que les données sont une fonction du temps

$$y_t = f(t) \quad (2)$$

On suppose alors un déterminisme temporel strict. On retrouve ce type de modèle dans les équations différentielles ordinaires, par exemple :

$$y_t = y_0 + vt \quad (3)$$

Une seconde catégorie de modèles cherche à déterminer chaque valeur de la série en fonction des valeurs qui la précède :

$$y_t = f(y_{t-1}, y_{t-2}, \dots) \quad (4)$$

Ce type de modèle exploite une fonction itérative. C'est notamment le cas des modèles ARIMA qui seront détaillés dans ce cours. A noter que ces deux types de modèles sont invertibles. L'équation différentielle (3) peut être exprimée sous forme itérative :

$$y_t = y_{t-1} + vdt \quad (5)$$

A noter que dans les deux cas ces modèles peuvent inclure des termes stochastiques, soit :

$$y_t = f(t) + \varepsilon_t \quad \text{ou} \quad y_t = f(y_{t-1}) + \varepsilon_t \quad (6)$$

ε_t représentant un bruit gaussien de moyenne nulle.

L'objectif premier des statistiques en séries temporelles est d'analyser la nature des dépendances temporelles dans les séries, c'est-à-dire la manière dont une valeur est liée aux valeurs précédemment observées. Un des postulats de base de ces approches est donc l'existence de telles dépendances, qu'il conviendra de caractériser. Une série temporelle composée de valeurs successives non corrélées (c'est-à-dire un bruit blanc) est considéré comme un cas limite.

On peut noter cependant que souvent, on a recours à l'analyse de variance pour traiter les séries de mesures répétées. Or une des assumptions majeures de l'ANOVA est que les résidus des différentes mesures ne sont pas auto-corrélés. Ce n'est évidemment pas le cas si la performance à l'essai t est liée à la performance réalisée à l'essai t-1.

2. Rappels statistiques

Les statistiques en séries temporelles empruntent aux statistiques classiques un certain nombre de procédures dont on peut rappeler les principes essentiels.

2.1. Covariance et corrélation

Si X_1 et X_2 sont fortement liées par un lien positif, on pourrait penser définir la covariance en développant l'idée suivante : Quand X_1 prend une valeur positive, alors X_2 prend vraisemblablement aussi une valeur positive, et vice-versa. Cette idée ne convient pas, car la Covariance doit rester inchangée quand les distributions de probabilité des variables sont translatées par des quantités arbitraires. Au lieu de mesurer X_1 et X_2 à partir de 0, nous allons donc les mesurer à partir de références qui se translatent en même temps que les distributions, par exemple leurs moyennes μ_1 et μ_2 , soit :

Quand $(X_1 - \mu_1)$ prend une valeur positive, alors $(X_2 - \mu_2)$ prend vraisemblablement aussi une valeur positive, et vice-versa. Ainsi, si X_1 et X_2 sont fortement (positivement)

liées, les quantités $(X_1 - \mu_1)$ et $(X_2 - \mu_2)$ seront le plus souvent simultanément positives, ou simultanément négatives.

Le produit $(X_1 - \mu_1)(X_2 - \mu_2)$ sera donc alors le plus souvent positif, soit parce que les deux quantités sont positives, soit parce que les deux quantités sont négatives. La **Covariance** de X_1 et X_2 . est définie comme l'**espérance** du produit $(X_1 - \mu_1)(X_2 - \mu_2)$

$$Cov(X_1, X_2) = E[(X_1 - \mu_1).(X_2 - \mu_2)] \quad (7)$$

Une faiblesse de la Covariance est qu'elle n'est pas invariante dans un changement d'unités utilisées pour exprimer les valeurs des deux variables X_1 et X_2 . Par exemple, la valeur de la Covariance de "Taille" et "Poids" d'une population change si les tailles sont exprimées en pouces au lieu de centimètres, ou les poids sont exprimés en "pounds" au lieu de kilogrammes, alors que la force du lien entre ces deux grandeurs ne dépend évidemment pas des unités utilisées pour les exprimer. C'est pourquoi on utilise plus fréquemment, et à fin de comparaison entre situations, le **coefficient de corrélation**.

Supposons que l'unité utilisée pour mesurer X_1 soit divisée par 2 (et donc que les valeurs de X_1 soient multipliées par 2). Alors la covariance $Cov(X_1, X_2)$ est également multipliée par 2. Mais l'écart-type (racine carrée de la variance) de X_1 est également multiplié par 2, et donc le rapport :

$$Cov(X_1, X_2) / (Var(X_1))^{1/2} \quad (8)$$

reste inchangé. Le même argument s'applique à X_2 , et plus généralement, à tout changement d'unités pour la mesure de X_1 et de X_2 . Donc, en toute généralité, le nombre :

$$\rho = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1) \cdot Var(X_2)}} \quad (9)$$

ne dépend pas des unités dans lesquelles X_1 et X_2 sont exprimées. ρ varie de +1 (corrélacion maximale positive) à -1 (corrélacion maximale négative).

On utilise couramment une formule plus pratique, dérivée de la précédente:

$$r = \frac{N \sum xy - \sum x \sum y}{\sqrt{(N \sum x^2 - (\sum x)^2)(N \sum y^2 - (\sum y)^2)}} \quad (10)$$

2.2. Corrélation partielle

La corrélation partielle permet parfois d'éviter de se laisser abuser par certaines corrélations artefactuelles. Par exemple, si l'on calcule les corrélations entre (1) l'intelligence, (2) le poids, et (3) l'âge. On trouve une corrélation $r_{12}=.60$ entre intelligence et poids. Cette corrélation doit être rapportée aux deux autres, soit $r_{13}=.69$, et $r_{23}=.88$. La corrélation partielle exprime la liaison entre deux variables, en contrôlant l'influence d'une troisième variable: dans ce cas on calcule la liaison entre intelligence et poids, à âge constant.

La procédure de corrélation partielle consiste à calculer l'équation de régression de (1) vers (3). Puis on calcule la corrélation entre les résidus de cette régression et (2).

$$\log(y) = a(\log(x)) + b \quad \text{donc} \quad y = e^{a(\log(x)) + b}$$

$$y = e^{a(\log(x))} * e^b$$

$$y = e^b * x^a$$

A = a et B = e^b

D'autres fonctions plus complexes peuvent être obtenues par la même méthode. Le choix du meilleur ajustement se fait ensuite par comparaison des coefficients de corrélation de chaque équation.

2.4. Régression multiple

La régression permet de résumer la relation entre deux variables, et donc de prédire une variables Y en fonction d'une variable X. Mais la prédiction d'une variable donnée peut être plus fine si l'on prend en compte plus de variables prédictives. La régression multiple permet de calculer une équation additive de forme:

$$V_p = A_1 V_1 + A_2 V_2 + A_3 \quad (17)$$

prédisant une variable V_p à partir de deux (ou plus) variables mesurées. L'équation de régression multiple est caractérisée par un coefficient de régression multiple, exprimant la précision de la prédiction. A noter que les valeurs des coefficients A₁ et A₂ dépendent évidemment des unités de mesure utilisées pour mesurer V₁ et V₂. On utilise plus usuellement l'équation de régression normalisée :

$$Z_{V_p} = \beta_1 Z_{V_1} + \beta_2 Z_{V_2} \quad (18)$$

Dans cette équation Z_{V_p}, Z_{V₁} et Z_{V₂} sont les valeurs centrées réduites.

Le calcul des équations de régression multiple est rapidement réalisé par les logiciels de statistique. Il peut être utile cependant de maîtriser les algorithmes de calcul, pour construire ses propres routines. Soit une régression multiple comprenant trois prédicteurs :

$$Z_{V_p} = \beta_1 Z_{V_1} + \beta_2 Z_{V_2} + \beta_3 Z_{V_3}$$

On suppose connues les coefficients de corrélation entre les quatre variables. Il s'agit de résoudre par substitution le système d'équations suivant :

$$r(1,p) = \beta_1 + r(1,2)\beta_2 + r(1,3)\beta_3$$

$$r(2,p) = r(2,1)\beta_1 + \beta_2 + r(2,3)\beta_3$$

$$r(3,p) = r(3,1)\beta_1 + r(3,2)\beta_2 + \beta_3$$

On obtient les solutions suivantes :

$$\beta_1 = r(1,p) - r(1,2)\beta_2 - r(1,3)\beta_3$$

$$\beta_2 = (-r(2,p) + (r(2,1)r(1,p)) + \beta_3(r(2,3) - (r(2,1)r(1,3)))) / ((r(2,1)r(1,2)) - 1)$$

$$\beta_3 = ((-r(3,1)r(1,p)) - (((-r(3,1)r(1,2)) + r(3,2)) * (-r(2,p) + (r(2,1)r(1,p)))) / ((r(2,1)r(1,2)) - 1) + r(3,p)) / (((r(2,3) - (r(2,1)r(1,3)))) * ((-r(3,1)r(1,2)) + r(3,2)) / ((r(2,1)r(1,2)) - 1) + ((-r(3,1)r(1,3)) + 1))$$

Les coefficients pour variables brutes sont calculées ainsi :

$$A_1 = \beta_1(\sigma_{V_p}/\sigma_{V_1})$$

$$A_2 = \beta_2(\sigma_{V_p}/\sigma_{V_2})$$

$$A3 = \beta3(\sigma Vp/\sigma V3)$$

Et l'ordonnée à l'origine :

$$A4 = m_{Vp} - A1m_{V1} - A2m_{V2} - A3m_{V3}$$

Pour quatre prédicteurs, on obtient mes solutions suivantes :

$$Z_{Vp} = \beta1Z_{V1} + \beta2Z_{V2} + \beta3Z_{V3} + \beta4Z_{V4}$$

$$\beta1 := r(1,p)-r(1,2)*\beta2-r(1,3)*\beta3-r(1,4)*\beta4$$

$$\beta2 = ((r(2,p)-r(2,1)*r(1,p))+r(2,1)*r(1,3)-r(2,3))*\beta3+(r(2,1)*r(1,4)-r(2,4))*\beta4/(1-r(2,1)*r(1,2))$$

$$\beta3 = ((r(3,p)-r(3,1)*r(1,p))+r(3,1)*r(1,2)-r(3,2))*r(2,p)-r(2,1)*r(1,p))/(1-r(2,1)*r(1,2))+((r(3,1)*r(1,2)-r(3,2))*r(2,1)*r(1,4)-r(2,4))/(1-r(2,1)*r(1,2))+r(3,1)*r(1,4)-r(3,4))*\beta4/((1-r(3,1)*r(1,3))-r(3,1)*r(1,2)-r(3,2))*r(2,1)*r(1,3)-r(2,3))/(1-r(2,1)*r(1,2))$$

$$\beta4 = ((r(4,p)-r(4,1)*r(1,p))+r(4,1)*r(1,2)-r(4,2))*r(2,p)-r(2,1)*r(1,p))/(1-r(2,1)*r(1,2))+((r(4,1)*r(1,2)-r(4,2))*r(2,1)*r(1,3)-r(2,3))/(1-r(2,1)*r(1,2))+r(4,1)*r(1,3)-r(4,3))*((r(3,p)-r(3,1)*r(1,p))+r(3,1)*r(1,2)-r(3,2))*r(2,p)-r(2,1)*r(1,p))/(1-r(2,1)*r(1,2))/((1-r(3,1)*r(1,3))-r(3,1)*r(1,2)-r(3,2))*r(2,1)*r(1,3)-r(2,3))/(1-r(2,1)*r(1,2)))/((1-r(4,1)*r(1,2)-r(4,2))*r(2,1)*r(1,4)-r(2,4))/(1-r(2,1)*r(1,2))-(((r(2,1)*r(1,3)-r(2,3))*r(4,1)*r(1,2)-r(4,2))/(1-r(2,1)*r(1,2))+r(4,1)*r(1,3)-r(4,3))*((r(3,1)*r(1,2)-r(3,2))*r(2,1)*r(1,4)-r(2,4))/(1-r(2,1)*r(1,2))+r(3,1)*r(1,4)-r(3,4))/((1-r(3,1)*r(1,3))-r(3,1)*r(1,2)-r(3,2))*r(2,1)*r(1,3)-r(2,3))/(1-r(2,1)*r(1,2))))$$

Ces équations ne sont fournies ici que pour ceux qui voudraient les utiliser pour intégration dans des routines de programmation.

3. Fonction d'auto-corrélation et fonction d'auto-corrélation partielle

L'auto-corrélation consiste à corrélérer la série avec elle-même, en introduisant un décalage entre les deux échantillons. L'autocorrélation de décalage 0 est égale à 1, puisque la série est corrélée avec elle-même. Puis la corrélation diminue au fur et à mesure que le décalage s'accroît. On appelle fonction d'auto-corrélation $C(\tau)$ la fonction faisant correspondre au décalage τ l'auto-corrélation correspondante. Si le signal présente une périodicité, on retrouve un maximum d'auto-corrélation pour chaque période significative du signal. Vereijken (1991) utilise ainsi l'auto-corrélation pour déterminer la fréquence d'oscillation du simulateur de ski. Une série temporelle de bruit blanc présente la particularité de ne présenter que des auto-corrélations nulles, quel que soit le décalage ($C(\tau)=0$).

On peut également calculer une fonction d'auto-corrélation partielle. Dans ce cas on calcule pour chaque décalage la corrélation entre la série et la série décalée, en contrôlant l'influence de tous les décalages intermédiaires.

Les logiciels de statistiques fournissent généralement les fonctions d'auto-corrélation et d'auto-corrélation partielle ; Il peut être utile cependant de connaître les algorithmes de calcul.

Soit $r(k)$ l'auto-corrélation de décalage k

Soit $rp(k)$ l'auto-corrélation partielle de décalage k (donc contrôlant tous les décalages intermédiaires)

$$rp(1) = r(1)$$

$$rp(2) = (r(2) - (r(1))^2) / (1 - r(1)^2)$$

$$\text{phi}(1, 1) = rp(1)$$

$$\text{phi}(2, 2) = rp(2)$$

Pour les décalages k supérieurs à 2, et pour des valeurs de j variant de 1 à k-2:

$$\text{phi}(k - 1, j) = \text{phi}(k - 2, j) - (\text{phi}(k - 1, k - 1) * \text{phi}(k - 2, k - 1 - j))$$

Pour les valeurs de k supérieures à 3

$$rp(k) = (r(k) - \sum_{j=1}^{k-1} \text{phi}(k - 1, j) * r(k - j)) / (1 - \sum_{j=1}^{k-1} \text{phi}(k - 1, j) * r(j))$$

4. Corrélations à court terme

Les modèles ARIMA, développés par Box et Jenkins (1976) permettent de modéliser les dépendances à court terme dans les séries. On entend par dépendance à court terme le fait que la valeur actuelle ne soit déterminée que par un ensemble limité de valeurs précédentes.

4.1. Processus d'intégration

Une différenciation d'ordre 1 suppose que la différence entre deux valeurs successives de y est constante.

$$y_t - y_{t-1} = \mu + \varepsilon_t \quad (19)$$

μ est la constante du modèle, et représente la différence moyenne en y. Un tel modèle peut être représenté comme un accroissement linéaire en fonction du temps. Si μ est égal à 0, la série est stationnaire.

Les modèles d'ordre 2 travaillent non plus sur les différences brutes, mais sur les différences de différence. La seconde différence de y au moment t est égale à $(y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$, c'est-à-dire à $y_t - 2y_{t-1} + y_{t-2}$. Un tel modèle obéira à l'équation de prédiction suivante :

$$y_t - 2y_{t-1} + y_{t-2} = \mu + \varepsilon_t \quad (20)$$

ou encore:

$$y_t = \mu + 2y_{t-1} - y_{t-2} + \varepsilon_t \quad (21)$$

4.2. Processus auto-régressifs

Les modèles auto-régressifs supposent que y_t est une fonction linéaire des valeurs précédentes.

$$y_t = \mu + \phi_1 y_{(t-1)} + \phi_2 y_{(t-2)} + \phi_3 y_{(t-3)} + \varepsilon_t \quad (22)$$

Littérairement, chaque observation est constituée d'une composante aléatoire (choc aléatoire, ε) et d'une combinaison linéaire des observations précédentes. ϕ_1 , ϕ_2 et ϕ_3 dans cette équation sont les coefficients d'auto-régression

A noter que cette équation porte soit sur les données brutes, soit sur les données différenciées. L'association d'un processus auto-régressif et d'une différenciation donne:

$$y_t - y_{t-1} = \mu + \phi(y_{t-1} - y_{t-2}) + \varepsilon_t \quad (23)$$

Ce qui peut également être écrit:

$$y_t = \mu + y_{t-1} + \phi(y_{t-1} - y_{t-2}) + \varepsilon_t \quad (24)$$

Notez qu'un processus auto-régressif ne sera stable que si les paramètres sont compris dans un certain intervalle ; par exemple, s'il n'y a qu'un paramètre auto-régressif, il doit se trouver dans l'intervalle $-1 < \phi_1 < +1$. Dans les autres cas, les effets passés s'accumuleraient et les valeurs successives des y_t se déplaceraient infiniment vers l'avant, ce qui signifie que la série ne serait pas stationnaire.

4.3. Processus de moyenne mobile

Les modèles à moyenne mobile suggèrent que la série présente des fluctuations autour d'une valeur moyenne. On considère alors que la meilleure estimation est représentée par la moyenne pondérée d'un certain nombre de valeurs antérieures (ce qui est le principe des procédures de moyennes mobiles utilisées pour le lissage des données). Ceci revient en fait à considérer que l'estimation est égale à la moyenne vraie, auquel on ajoute une somme pondérée des erreurs ayant entaché les valeurs précédentes :

$$y_t = \mu - \theta_1 \varepsilon_{(t-1)} - \theta_2 \varepsilon_{(t-2)} - \theta_3 \varepsilon_{(t-3)} + \varepsilon_t \quad (25)$$

Littérairement, chaque observation est composée d'une composante d'erreur aléatoire (choc aléatoire, ε) et d'une combinaison linéaire des erreurs aléatoires passées. θ_1 , θ_2 et θ_3 sont les coefficients de moyenne mobile du modèle.

Comme précédemment cette équation porte soit sur les données brutes, soit sur les données différenciées si une différenciation a été nécessaire. Pour l'association d'une différenciation et d'un terme de moyenne mobile on aura :

$$y_t - y_{t-1} = \mu - \theta \varepsilon_{t-1} + \varepsilon_t \quad (26)$$

Ce qui peut également être écrit:

$$y_t = \mu + y_{t-1} - \theta \varepsilon_{t-1} + \varepsilon_t \quad (27)$$

Un modèle de moyenne mobile correspond à des séries exhibant des fluctuations aléatoires autour d'une moyenne variant lentement. Plutôt que de prendre comme précédemment la valeur précédente comme prédicteur, on utilise une moyenne de quelques observations précédentes, de manière à éliminer le bruit, et estimer plus précisément la moyenne locale.

4.4. Les modèles ARIMA

L'objectif essentiel des modèles ARIMA est de permettre une prédiction de l'évolution future d'un phénomène. Son développement dans le domaine de l'économétrie est basé sur ce principe. Un autre intérêt, peut-être plus essentiel en ce qui concerne la recherche scientifique, est de comprendre la signification théorique de ces différents processus. Il est clair cependant que cette interprétation dépend de la nature du phénomène étudié, et des modèles dont le chercheur dispose pour en rendre compte.

Un modèle ARIMA est noté (p,d,q) , p correspondant au nombre de termes auto-régressifs, d au nombre de différenciations, et q au nombre de termes de moyenne mobile. De nombreuses combinaisons sont envisageables. On peut ici décrire trois modèles particulièrement courants :

- le modèle ARIMA(0,0,0) décrit un processus non différencié à bruit blanc, suggérant des fluctuations aléatoires autour d'une valeur de référence stable. Cette valeur de référence peut

être considérée comme une caractéristique stable du système étudié (trait de personnalité, mémoire, capacité stabilisée, etc..)

- le modèle ARIMA (0,1,1), sans constante significative, obéit à l'équation suivante :

$$y_t = y_{t-1} - \theta \varepsilon_{t-1} + \varepsilon_t \quad (29)$$

Ce modèle suggère que la valeur de référence évolue d'une mesure à l'autre. Plus précisément, la valeur de référence est fonction de la valeur de référence précédente et de l'erreur ayant entaché la mesure précédente. Le processus tend simultanément à préserver la valeur de référence et à garder la mémoire des perturbations successives les erreurs successives. La valeur du coefficient de moyenne mobile détermine l'équilibre entre ces deux tendances opposées. Ce modèle a été notamment utilisé pour décrire la dynamique de l'estime de soi (Fortes, Delignières & Ninot, 2004).

- le modèle ARIMA (1,0,0) : un tel processus auto-régressif suggère que le phénomène étudié possède une dynamique de relaxation par rapport à une valeur attractive. Lorsque le processus s'écarte de cette valeur, il tend à la rejoindre au bout d'un certain temps. La valeur du coefficient d'auto-régression détermine la force de l'attracteur. Cette dynamique a été retrouvée dans la dynamique de mémorisation d'une séquence morphocinétique (Delcor, Cadopi, Delignières & Mesure, 2003).

4.5. Identification des modèles ARIMA

4.5.1. Différenciation

L'estimation des modèles ARIMA suppose que l'on travaille sur une série stationnaire. Ceci signifie que la moyenne de la série est constante dans le temps, ainsi que la variance. La meilleure méthode pour éliminer toute tendance est de différencier, c'est-à-dire de remplacer la série originale par la série des différences adjacentes. Une série temporelle qui a besoin d'être différenciée pour atteindre la stationnarité est considérée comme une version intégrée d'une série stationnaire (d'où le terme *Integrated*).

La correction d'une non-stationnarité en termes de variance peut être réalisée par des transformations de type logarithmique (si la variance croît avec le temps) ou à l'inverse exponentielle. Ces transformations doivent être réalisées avant la différenciation.

Une série stationnaire fluctue autour d'une valeur moyenne et sa fonction d'autocorrélation décline rapidement vers zéro. Si une série présente des auto-corrélations positives pour un grand nombre de décalages (par exemple 10 ou plus), alors elle nécessite d'être différenciée.

L'ordre optimal de différenciation est souvent celui pour lequel l'écart-type est minimal. Un accroissement de l'écart-type doit donc être considéré comme un symptôme de sur-différenciation. Un troisième symptôme de sur-différenciation est un changement systématique de signe d'une observation à l'autre.

Un modèle sans différenciation suppose que la série originale est stationnaire. Un modèle avec une différenciation d'ordre 1 suppose que la série originale présente une tendance constante. Un modèle avec une différenciation d'ordre 2 suppose que la série originale présente une tendance variant dans le temps.

4.5.2. Termes auto-régressifs et de moyenne mobile.

Après que la série ait été stationnarisée, l'étape suivante consiste à identifier les termes AR et MA nécessaires pour corriger les auto-corrélations résiduelles. Cette analyse est basée sur l'examen des fonctions d'auto-corrélation et d'auto-corrélation partielle. Rappelons que

l'auto-corrélation est la corrélation d'une série avec elle-même, selon un décalage (*lag*) défini. L'auto-corrélation de décalage 0 est par définition égale à 1. La fonction d'auto-corrélation fait correspondre à chaque décalage l'auto-corrélation correspondante.

D'une manière générale, une corrélation partielle entre deux variables est la quantité de corrélations qui n'est pas expliquée par les relations de ces variables avec un ensemble spécifié d'autres variables. Dans le cas des séries temporelles, la corrélation partielle de décalage k est la corrélation entre y_t et y_{t-k} , contrôlant l'influence des $k-1$ valeurs interposées.

L'autocorrélation de décalage 1 est la corrélation entre y_t et y_{t-1} . On suppose que c'est également la corrélation entre y_{t-1} et y_{t-2} . Si y_t et y_{t-1} sont corrélés, et que y_{t-1} et y_{t-2} le sont également, on peut supposer qu'une corrélation sera présente entre y_t et y_{t-2} . C'est-à-dire que la corrélation de décalage 1 se propage au décalage 2 et sans doute aux décalages d'ordre supérieurs. Plus précisément, la corrélation attendue au décalage 2 est la carré de la corrélation observée au décalage 1. L'auto-corrélation *partielle* de décalage 2 est donc la différence entre l'auto-corrélation de décalage 2 et la corrélation attendue due à la propagation de la corrélation de décalage 1.

Les outils principaux utilisés lors de la phase d'identification sont donc les tracés de la série, les corrélogrammes d'autocorrélation (FAC), et d'autocorrélation partielle (FACP). La décision n'est pas simple et les cas les plus atypiques requièrent, outre l'expérience, de nombreuses expérimentations avec des modèles différents (avec divers paramètres ARIMA). Toutefois, les composantes des séries chronologiques empiriques peuvent souvent être assez bien approchées en utilisant l'un des 5 modèles de base suivants, identifiables par la forme de l'autocorrélogramme (FAC) et de l'autocorrélogramme partiel (FACP). Puisque le nombre de paramètres (à estimer) de chaque type dépasse rarement 2, il est souvent judicieux d'essayer des modèles alternatifs sur les mêmes données.

(1) Un paramètre autorégressif (p) : FAC - décomposition exponentielle ; FACP - pic à la période 1, pas de corrélation pour les autres périodes.

(2) Deux paramètres autorégressifs (p) : FAC - une composante de forme sinusoïdale ou un ensemble de décompositions exponentielles ; FACP - pics aux périodes 1 et 2, aucune corrélation pour les autres périodes.

(3) Un paramètre de moyenne mobile (q) : FAC - pic à la période 1, aucune corrélation pour les autres périodes ; FACP - exponentielle amortie.

(4) Deux paramètres de moyenne mobile (q) : FAC - pics aux périodes 1 et 2, aucune corrélation pour les autres périodes ; FACP - une composante de forme sinusoïdale ou un ensemble de décompositions exponentielles.

(5) Un paramètre autorégressif (p) et un de moyenne mobile (q) : FAC - décomposition exponentielle commençant à la période 1 ; FACP - décomposition exponentielle commençant à la période 1.

5. Corrélations croisées

5.1. Cross-corrélation et cross-corrélation partielle

Le calcul du coefficient de cross-corrélation permet de rendre compte du degré d'association entre deux séries temporelles X et Y : il s'agit simplement d'une corrélation de Bravais-Pearson, prenant en considération les couples successifs de valeurs synchrones (x_t et y_t) dans les deux séries. Le coefficient de cross-corrélation peut également être calculé en introduisant un décalage temporel entre les deux séries : le calcul est alors réalisé en prenant

en compte des couples de valeurs séparées par un décalage constant (x_t et y_{t+k}). Ce calcul peut être réalisé pour un ensemble de décalages, positifs et négatifs, débouchant sur une fonction de cross-corrélation, mettant en relation le décalage et le coefficient de cross-corrélation correspondant. Ces fonctions de cross-corrélation permettent en particulier la détermination du décalage correspondant à l'association maximale entre les deux variables, suggérant que l'influence d'une série sur l'autre s'exerce selon un certain délai temporel. Le signe de ce décalage permet également d'identifier le sens de cette influence : si les évolutions de la série X précèdent des évolutions similaires de la série Y, on peut en conclure une influence de la première variable sur la seconde.

La figure 1 donne un exemple d'utilisation de la crosscorrélation. Les données sont tirées d'une expérience d'oscillation de l'avant-bras en synchronisation avec un métronome. Les carrés blancs représentent la fonction de cross-corrélation entre l'erreur au métronome et la période d'oscillation. On voit que les périodes futures sont négativement corrélées à l'erreur présente (justifiant l'hypothèse d'une correction). Inversement, l'erreur est positivement corrélée aux périodes précédentes. Les deux autres courbes représentent les fonctions de cross-corrélation entre l'erreur au métronome et les demi-périodes aller et retour des oscillations. On remarque que les corrélations négatives précédemment relevées entre l'erreur et les périodes futures sont principalement expliquées par un effet sur la seconde demi-période du cycle. A l'inverse, l'effet de la période sur l'erreur à venir est principalement expliquée par la première partie du cycle.

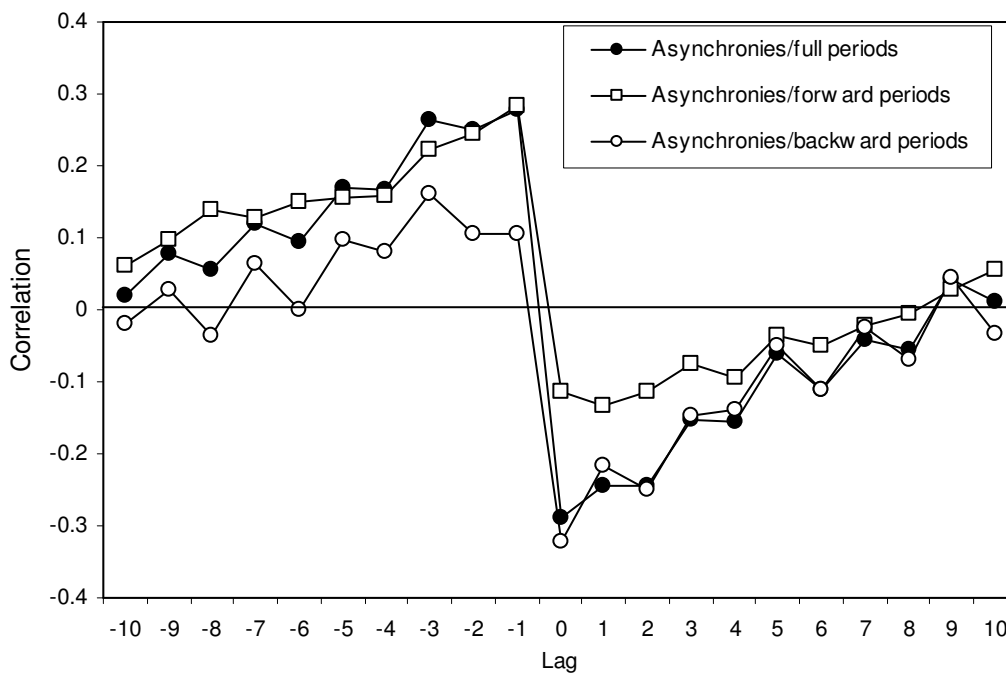


Figure 1 : Fonctions de cross-corrélation (voir détails dans le texte)

On peut également calculer des fonctions de cross-corrélation partielle, contrôlant l'influence des décalages intermédiaires. Notez que dans ce cas on a le choix de contrôler soit la première variable, soit la seconde. La figure 2 illustre sur les données présentées précédemment, l'intérêt de cette analyse. Les ronds blancs reprennent la fonction de cross-corrélation entre erreur et période (seuls les décalages positifs sont représentés). Les deux autres courbes représentent les fonctions de cross-corrélation partielles, contrôlant l'erreur (ronds blancs) ou la période (carrés noirs). On voit que lorsque l'on contrôle l'erreur, la persistance de la crosscorrélation disparaît : il n'y a pas de corrélation significative au-delà de

la première valeur, suggérant que l'erreur affecte simplement la période suivante. La persistance des corrélations dans la fonction de cross-corrélation n'était donc liée qu'à des effets indirects. Lorsque l'on contrôle la période, la cross-corrélation conserve sa persistance, et présente même des valeurs plus fortes (en valeurs absolues). Cet effet semble lié à la présence de corrélations fractales dans les erreurs, un résultats difficile à expliquer dans le cadre de ce cours.

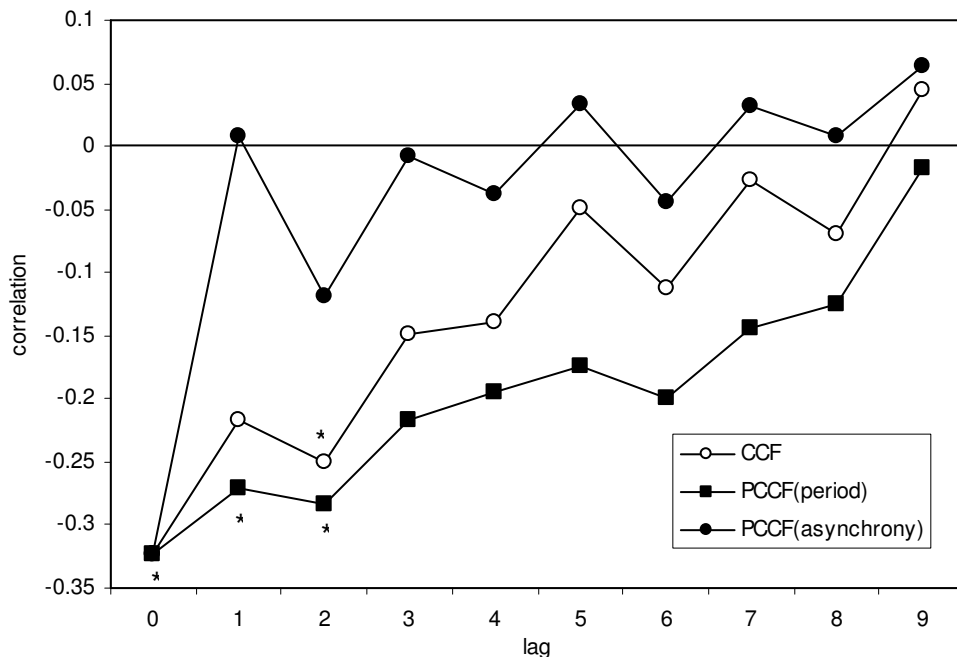


Figure 2 : Fonctions de cross-corrélation partielles (voir détails dans le texte)

5.2. Cross-corrélation fenêtrée

Supposons deux séries, représentant l'évolution de deux variables couplées de manière non-linéaire. En d'autres termes, la direction du processus d'influence entre ces deux séries varie au cours du temps, ainsi que la force de cette influence. Si l'on applique une approche cross-corrélationnelle à ces séries, prises dans leur ensemble, le décalage correspondant au maximum de la fonction de cross-corrélation représentera en quelque sorte la « moyenne » des décalages qui auraient pu être successivement mesurés au cours de l'observation.

Une solution pour analyser ce type de phénomène est de calculer la cross-corrélation sur une fenêtre limitée, et de faire glisse ensuite cette fenêtre tout au long des séries analysées. La figure 3 montre deux séries correspondant à l'angulation de la hanche (en haut) et à celle de la cheville (en bas), lors du maintien de la station de bout chez un patient paraplégique sous stimulation électrique fonctionnelle (Simon, 2006).

La fonction de cross-corrélation fenêtrée indique comment le couplage entre les deux articulations peut varier de la phase (cross-corrélation positive) à l'antiphase (cross-corrélation négative) au cours du même essai.

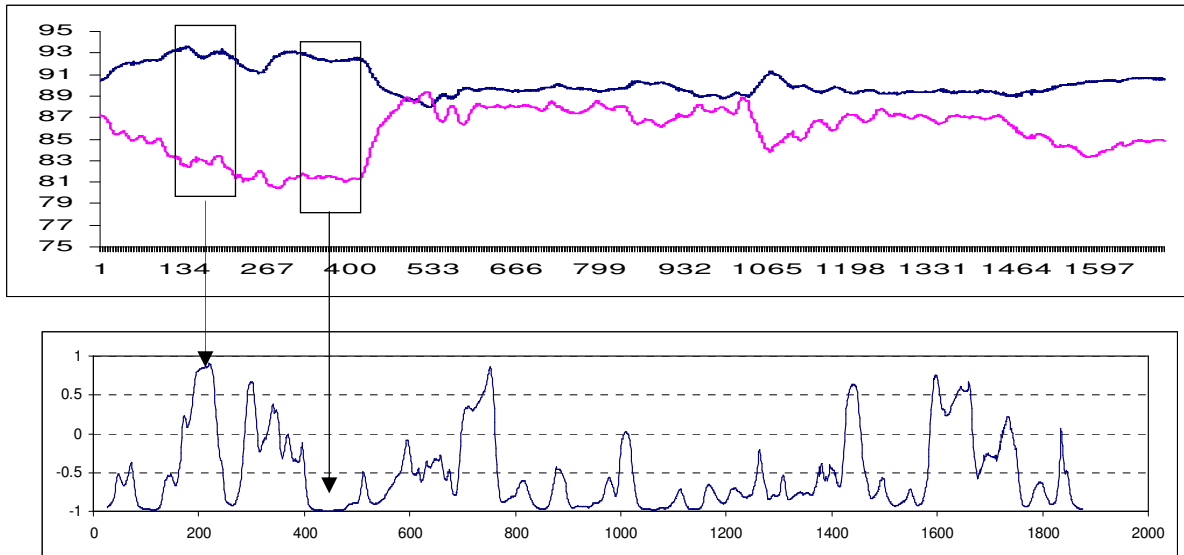


Figure 3. Mise en évidence des processus de couplage par cross-corrélation. D'après Simon (2006).

La figure 4 reproduit les fonctions de cross-corrélation fenêtrées obtenues entre la valeur physique perçue, et ses quatre dimensions constitutives (endurance, compétence sportive, apparence, et force). On voit qu'à certains moments les quatre variables contribuent simultanément à la variable supérieure. Puis les différentes variables prennent successivement l'ascendant, dans un processus d'alternance.

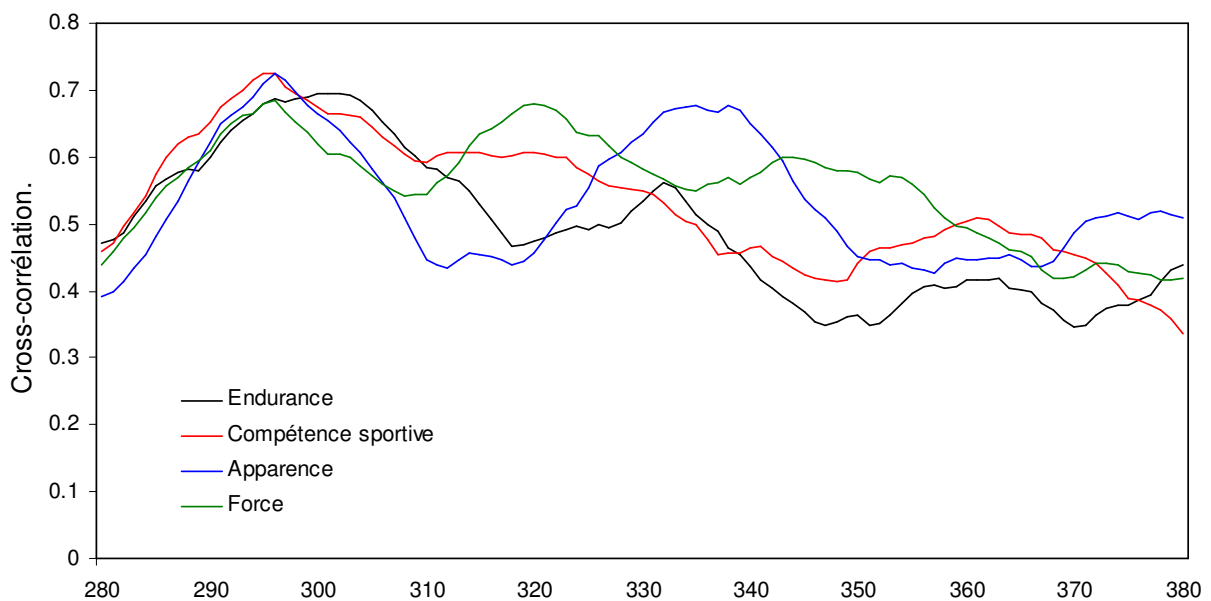


Figure 4: Analyse, par cross-corrélation fenêtrée, de la contribution de l'endurance, la compétence sportive, l'apparence et la force à la valeur physique perçue (Fortes, 2003).

5.3. Fonction de cross-corrélation fenêtrée

Une méthode de cross-corrélation à fenêtre mobile (*windowed cross-corrélation*) a été récemment proposée par Boker, Xu, Rotondo et King (2002). Cette méthode consiste à

calculer la fonction de cross-corrélation au niveau d'un segment temporel limité, où l'on fait l'hypothèse d'une stationnarité relative de l'association entre les deux séries. On répète ensuite ce traitement en décalant d'une valeur la fenêtre de calcul, et ainsi de suite jusqu'à ce que l'ensemble de la série ait été traité. L'ensemble de ces calculs permet d'obtenir une série de fonctions locales de cross-corrélation. Un algorithme de recherche de pic permet ensuite de déterminer les maxima de ces fonctions, et d'établir la série des décalages successifs correspondant à ces maxima.

Soient deux séries X et Y, de longueur totale N , $X = \{x_1, x_2, \dots, x_N\}$ et $Y = \{y_1, y_2, \dots, y_N\}$. Soient n la longueur de la fenêtre mobile, et $d (>0)$ le décalage maximum soumis à l'analyse. On verra plus loin les contraintes pesant sur la détermination de ces deux paramètres. L'analyse commence avec un premier segment longueur n de la série X, $X_{(a,n)} = \{x_a, x_{a+1}, x_{a+2}, \dots, x_{a+n-1}\}$. Pour l'ensemble des décalages k compris entre $-d$ et $+d$, on calcule le coefficient de cross-corrélation locale entre les deux séries, défini par :

$$r(X_{(a,n)}, Y_{(a+k,n)}) = \frac{1}{n} \sum_{i=a}^{a+n} \frac{(x_i - \bar{X}_{(a,n)})(y_{i+k} - \bar{Y}_{(a+k,n)})}{SD(X_{(a,n)})SD(Y_{(a+k,n)})} \quad (30)$$

où $\bar{X}_{(a,n)}$ et $\bar{Y}_{(a+k,n)}$ représentent les moyennes des deux segments des séries X et Y de longueur n commençant respectivement par x_a et y_{a+k} , et $SD(X_{(a,n)})$ et $SD(Y_{(a+k,n)})$ leurs écart-type. On obtient donc une série de $2d+1$ coefficients, définissant la fonction locale de cross-corrélation, indexée sur la valeur centrale du segment $X_{(a,n)}$.

On peut noter que classiquement, le calcul d'un coefficient de cross-corrélation de décalage k entraîne la perte de $|k|$ données dans chacune des séries, ce qui génère une diminution progressive des degrés de libertés au fur et à mesure que le décalage croît, en valeur absolue. Dans le cas présent, le décalage de la série Y s'opère sans perte de données, puisque la définition de chaque segment $Y_{(a+k,n)}$ s'opère par glissement d'une fenêtre de longueur n sur la série Y. Afin de permettre ce glissement sans perte de données pour les décalages négatifs au début de l'analyse, on commence les calculs pour $a = d$. Les calculs sont ensuite répétés pour des valeurs incrémentées de a ($a = d+1, d+2, \dots, N-d$). La valeur maximale permet le glissement du segment $Y_{(a+k,n)}$ jusqu'au décalage d , sans perte de données. L'ensemble de ces calculs produit une série de $N-2d$ fonctions locales de cross-corrélation. Un algorithme de recherche de pic permet ensuite de déterminer les maxima de ces fonctions, et d'établir la série des décalages successifs correspondant à ces maxima.

Les résultats, représentés sous forme de diagrammes de densité (Figure 5), font clairement apparaître la récurrence de dynamiques locales, d'une série à l'autre, selon des délais temporels qui peuvent être importants. Ce traitement a utilisé une fenêtre de 15 données et un décalage maximum de 50 (soit en tout 101 décalages explorés, de -50 à 50). On peut remarquer que de nombreux îlots de corrélation situent à des décalages de ± 28 ou ± 42 , suggérant la récurrence occasionnelle de dynamiques hebdomadaires dans les séries. On retrouve cependant surtout une zone de cross-corrélations élevées autour du décalage zéro, qui laisse supposer que c'est à ce niveau proximal que les influences directes de série à série se situent.

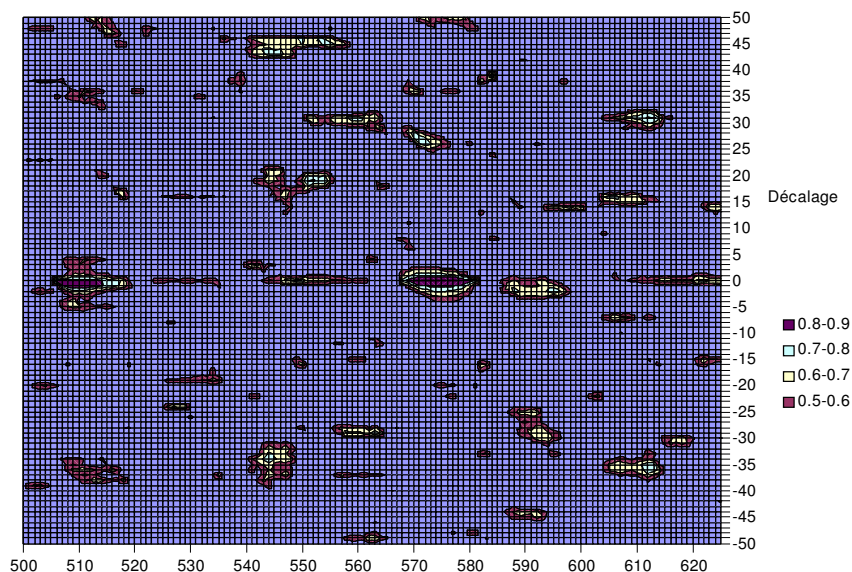


Figure 5 : Représentation des résultats de l'analyse de cross-corrélation fenêtrée. En abscisse, on reporte les indices temporels de la valeur centrale des fenêtres, et en ordonnée les décalages soumis à analyse. Les couleurs indiquent la grandeur du coefficient de cross-corrélation. Seules les corrélations supérieures à 0.5 ont été indiquées. Ce graphique correspond aux rapports entre Estime Globale de soi et Valeur Physique Perçue). Cette figure rend compte d'environ deux mois d'évaluation, à raison de deux évaluations par jour. L'étendue des décalages permet de mettre en évidence des délais de plus ou moins trois semaines (d'après Fortes, 2003).

6. Corrélations à long terme

Les processus ARIMA sont caractérisés par une mémoire à court terme : La valeur actuelle est déterminée la valeur qui précède, et l'importance des valeurs précédentes décroît de manière exponentielle. De nombreux phénomènes naturels semblent posséder des dépendances temporelles à plus long terme : les corrélations dans la série persistent de manière durable. Ces caractéristiques signent la présence d'une structure fractale.

6.1. Auto-similarité et dépendance à long terme

On appelle bruit toute variation imprévisible d'une quantité dans le temps. Le *bruit blanc* est le plus aléatoire des bruits. Dans ce cas, il n'existe aucune corrélation entre les accroissements successifs de la quantité. La fonction d'auto-corrélation est donc nulle ($C(\tau)=0$), et la densité spectrale est équivalente quelle que soit la fréquence (ce qui est une caractéristique de la lumière blanche, d'où le nom attribué à ce type de distribution). Le bruit blanc a une distribution gaussienne, et une moyenne nulle.

Le *mouvement brownien* est l'intégration d'un bruit blanc. On parle également de marche au hasard biaisée (*biased random walk*). En une dimension, il représente par exemple le mouvement d'une particule le long d'une ligne sous l'effet de chocs de direction et d'intensité aléatoires.

Soit $x(t)$ la position de la particule au temps t

$$x(t)=x(t-1) + \Delta x(t),$$

les accroissements $\Delta x(t)$ suivant une distribution à bruit blanc. Einstein (1905) a montré que dans le cas d'un mouvement brownien, la variance des accroissements est proportionnelle à l'intervalle de temps considéré. Les accroissements étant de moyenne nulle, cette variance correspond à l'espérance mathématique du carré des accroissements :

$$\begin{aligned}\text{Var}(\Delta x) &= E((\Delta x - E(\Delta x))^2) \\ &= E(\Delta x^2) - E(\Delta x)^2 \\ &= E(\Delta x^2)\end{aligned}$$

Cette espérance mathématique peut être estimée par la moyenne calculée sur un intervalle considéré, $\langle \Delta x^2 \rangle$. La relation d'Einstein prend la forme suivante :

$$\text{Var}(\Delta x) = \langle \Delta x^2 \rangle = 2D\Delta t \quad (31)$$

où le paramètre D est le coefficient de diffusion. Le coefficient de diffusion est une mesure moyenne de l'activité stochastique, c'est-à-dire qu'il est lié à l'amplitude et à la fréquence des chocs. La relation ci-dessus est aisément extensible aux mouvements brownien bi- ou tridimensionnels.

Le terme de *mouvement brownien fractionnaire* a été introduit par Mandelbrot et van Ness (1968), et constitue une généralisation du mouvement brownien. Pour cette famille de processus, la relation d'Einstein est généralisée sous la loi d'échelle suivante:

$$\text{Var}(\Delta x) = \langle \Delta x^2 \rangle \propto \Delta t^{2H} \quad (32)$$

Ou encore

$$\sigma_{\Delta x} = \propto \Delta t^H \quad (33)$$

dans laquelle H peut être n'importe quel réel compris entre 0 et 1. Pour un mouvement classique brownien, $H=0.5$, et l'on retrouve la relation linéaire d'Einstein.

Le mouvement brownien fractionnaire (fBm) est caractérisé par la présence de corrélations à long terme, lorsque H est différent de 0.5. Un exposant supérieur à 0.5 révèle des phénomènes de persistance, c'est-à-dire que l'évolution de la série tend à suivre des tendances. Si la série a augmenté précédemment, la probabilité est forte qu'il continue à le faire. Les séries persistantes ont une mémoire à long terme, c'est-à-dire qu'il existe une corrélation à long terme entre les événements actuels et les événements futurs. Chaque observation porte la mémoire des événements qui l'ont précédé.

A l'inverse, un exposant inférieur à 0.5 révèle un phénomène d'anti-persistance. Dans ce cas les accroissements successifs tendent à être négativement corrélés. Une augmentation de la variable tend à être suivi d'une diminution, et vice-versa.

Il existe une autre famille de processus, liée à la précédente : les bruits gaussiens fractionnaires (fGn). Un fGn est la différentiation d'un fBm, et les deux séries sont caractérisées par le même exposant H . $H = 0.5$ correspond donc au mouvement brownien dans la famille des fBm, et au bruit blanc dans la famille des fGn. Ce modèle fBm/fGn est fondamental. Sa méconnaissance dans de nombreux articles publiés avant 2000 a donné lieu à de nombreuses interprétations erronées des résultats (voir Eke et al., 2000, Delignières et al., 2006). La figure 6 illustre ce modèles dichotomique.

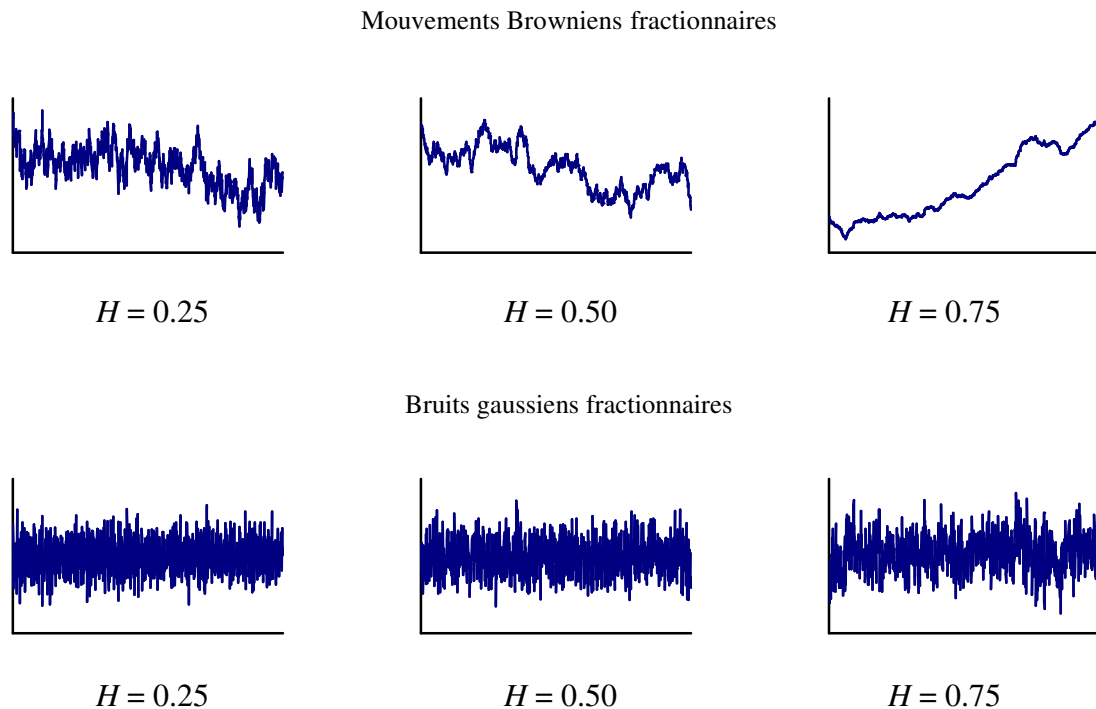


Figure 6: Exemples graphiques de séries fractales. Les graphes du haut représentent des mouvements browniens fractionnaires (fBm) et les graphes du bas les bruits gaussiens fractionnaires (fGn) correspondants, pour trois valeurs typiques de l'exposant d'échelle. Les graphes du centre représentent le mouvement brownien ordinaire en haut ($H=0.5$), et sa série différenciée (bruit blanc) en-dessous. Les colonnes de gauche et de droite représentent respectivement les fBm anti-persistants et persistants, et leurs fGn correspondant.

Le mouvement brownien fractionnaire est également caractérisé par des propriétés d'auto-similarité ou d'invariance d'échelle. L'auto-similarité d'un objet fractal renvoie au fait que l'objet peut être décomposé en sous-unités, puis en sous-sous unités, qui possèdent les mêmes propriétés statistiques que l'objet global. Quelle que soit l'échelle d'observation, on observe donc des caractéristiques identiques. Pour mettre en évidence l'auto-similarité d'une courbe représentée en deux dimensions, on sélectionne un segment de la courbe, et on applique aux deux échelles un facteur identique d'amplification. Les propriétés statistiques de la courbe ré-étalonnée doivent être similaires à celles de la courbe d'origine.

Ce concept de structure fractale peut être étendu aux séries temporelles. L'auto-similarité signifie alors qu'en moyenne, les fluctuations sur une échelle de temps sont similaires aux fluctuations sur d'autres échelles de temps. Cette extension pose cependant problème. Si une série temporelle est classiquement représentée sur un plan en deux dimensions, ces deux dimensions renvoient à des grandeurs physiques différentes (le temps en

abscisse, la variable évoluant dans le temps en ordonnée). On aura donc besoin de deux facteurs d'amplification distincts (un pour le temps, et l'autre pour la variable).

En termes mathématiques, une série temporelle sera auto-similaire si :

$$y(t) \stackrel{d}{=} a^\alpha y\left(\frac{t}{a}\right) \quad (34)$$

$\stackrel{d}{=}$ signifie que les propriétés statistiques des deux côtés de l'équation sont identiques. En d'autres termes, un processus auto-similaire $y(t)$, avec un paramètre α a une probabilité de

distribution identique à un processus ré-étalonné de manière adéquate, $a^\alpha y(t/a)$, c'est-à-dire une série temporelle qui a été ré-étalonnée sur l'axe des x par un facteur a ($t \rightarrow t/a$), et sur l'axe des y par un facteur a^α ($y \rightarrow a^\alpha y$). L'exposant α est appelé paramètre d'auto-similarité.

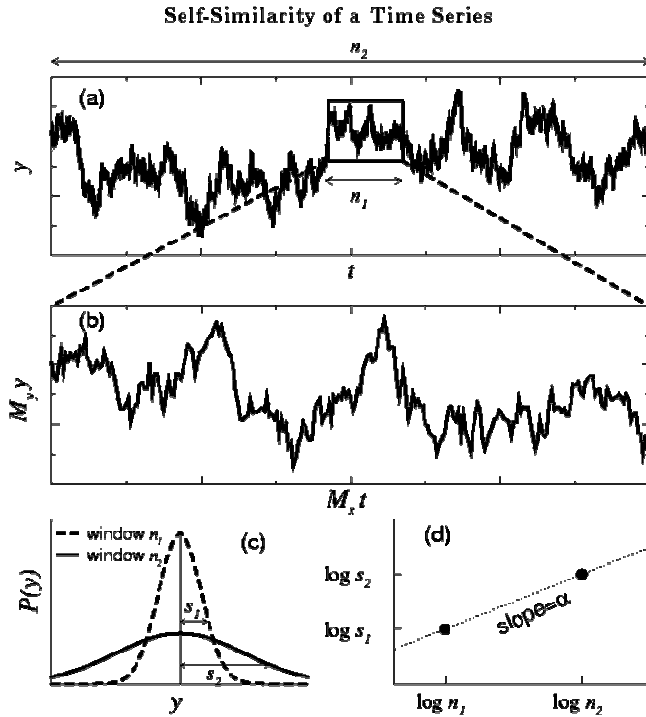


Figure 7 : Phénomène d'auto-similarité dans une série temporelle.

En pratique, il est impossible de déterminer si deux processus sont statistiquement identiques, car ce critère nécessite qu'ils aient des fonctions de distribution identiques (comprenant la moyenne, la variance, mais également tous les moments d'ordres plus élevés). On approche généralement cette équivalence par un critère plus souple en n'examinant que les deux premiers moments (moyenne et variance). La figure 7 illustre cette démarche, dans le cadre d'une série temporelle.

Soit M_x le facteur d'amplification appliqué au temps et M_y le facteur appliqué à la variable. Le paramètre d'auto-similarité peut être calculé par la simple relation :

$$\alpha = \frac{\ln M_y}{\ln M_x} \quad (35)$$

M_x peut être calculé en faisant le rapport des tailles de la série initiale et de la série extraite ($M_x = n_2/n_1$). Une estimation de M_y peut être réalisée par le rapport des écarts types des deux distributions ($M_y = s_2/s_1$). On obtient donc :

$$\alpha = \frac{\ln M_y}{\ln M_x} = \frac{\ln s_2 - \ln s_1}{\ln n_2 - \ln n_1} \quad (36)$$

Cette relation est simplement la pente de la droite joignant les deux points (n_1, s_1) et (n_2, s_2) sur un graphe log-log.

6.2. La fonction d'auto-corrélation

Une série présentant des corrélations à long terme est définie comme une série présentant une fonction d'auto-corrélation possédant une propriété très précise: $C(\tau) \approx \tau^{-\gamma}$, avec $0 < \gamma < 1$. En d'autres termes la fonction d'auto-corrélation suit une loi puissance. Ceci implique que dans le domaine fréquentiel, on a également une loi puissance $1/f^\beta$ ($\beta = 1 - \gamma$).

La figure 8 contraste à titre d'exemple les fonctions d'auto-corrélation d'un processus auto-régressif et celle d'un bruit gaussien fractionnaire. La première présente une extinction exponentielle de l'auto-corrélation, alors que la seconde présente une diminution plus lente, de type puissance. On considère cependant que la fonction d'auto-corrélation ne contient pas suffisamment d'information pour une estimation valide de l'exposant fractal.

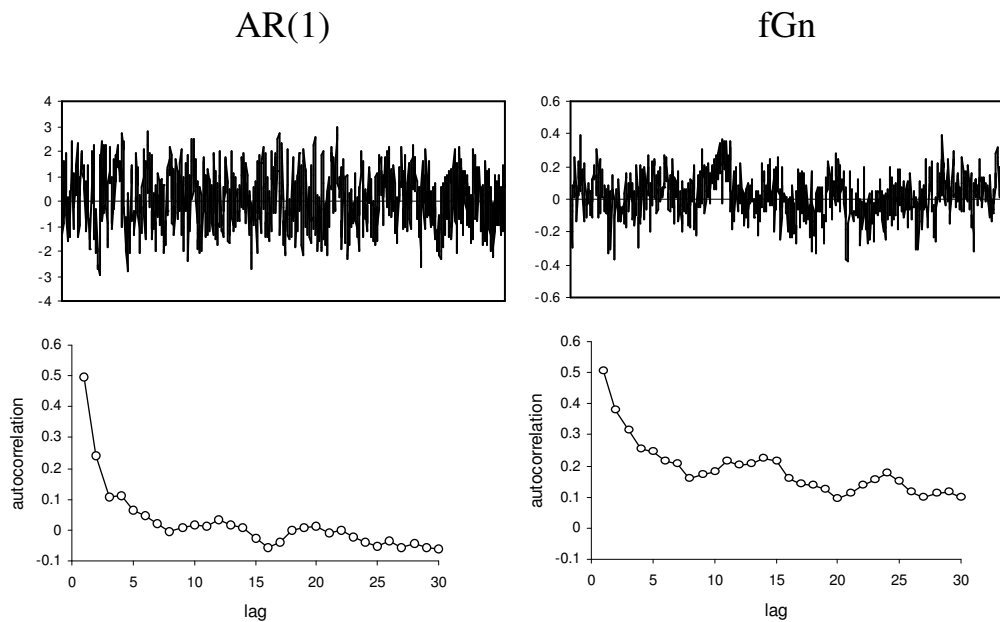


Figure 8 : Fonctions d'auto-corrélation de processus auto-régressifs et de processus fractals.

6.3. Méthodes spectrales

Une méthode classique pour analyser ces processus stochastiques consiste à s'intéresser au spectre de puissance de la série. Ce spectre, ou périodogramme, est obtenu par une analyse de Fourier. Le périodogramme présente les fréquences f en abscisse, et le carré des amplitudes correspondantes, notées $S(f)$, en ordonnées. La relation de Mandelbrot et van Ness (1968) peut être exprimée sous la forme:

$$S(f) \propto 1/f^\beta \quad (37)$$

On détermine l'exposant β de la pente de la droite ajustant la relation entre $\text{Log}_{10}f$ et $\text{Log}_{10}S(f)$ (figure 9).

$$\text{Log}_{10}S(f) \propto \text{Log}_{10}f^{-\beta} \quad (38)$$

Pour un processus de bruit blanc, β est égal à 0. Un mouvement brownien est caractérisé par un exposant β égal à 2. Enfin, l'exposant $\beta=1$ correspond au cas particulier où $S(f)$ est proportionnel à $1/f$. Il s'agit du bruit rose, ou bruit $1/f$.

L'exposant β est lié à la fonction d'auto-corrélation, car le spectre de puissance est simplement la transformée de Fourier de la fonction d'auto-corrélation. Si $[C(\tau) \approx \tau^{-\gamma}]$ (cas des mouvements fractionnaires browniens), alors $\beta = 1 - \gamma$. On peut en dériver une relation directe entre β et α : $\beta = 2\alpha - 1$. La DFA est cependant plus adaptée aux séries non stationnaires, car l'analyse de Fourier requiert un minimum de stationnarité.

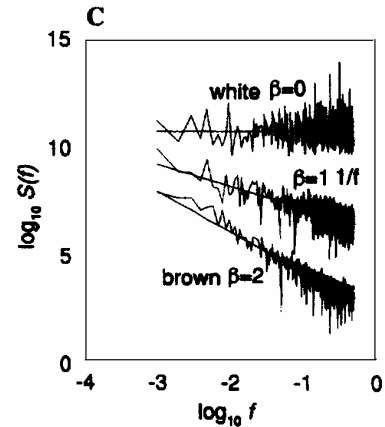


Figure 9 Estimation de l'exposant β (d'après Hausdorff et al., 1996)

6.4. Méthodes temporelles

6.4.1. La Rescaled range Analysis

Hurst (1965) a proposé une méthode appelée analyse des étendues normalisées, ou *Rescaled Range Analysis (R/S Analysis)*. La méthode consiste à estimer, pour un intervalle d'effectif donné, le rapport entre l'étendue R de la série centrée et intégrée, et l'écart-type S de la série originale. En d'autres termes, la série est considérée comme la "cause" générant le déplacement d'une particule (dont rend compte la série centrée et intégrée). On s'intéresse à la différence entre les limites minimale et maximale du déplacement, dans un certain intervalle temporel, normalisé par l'écart-type de la cause ayant généré le déplacement. Dans la logique de l'équation proposée par Einstein, l'exposant de Hurst rend compte de la relation liant R/S à la taille de l'échantillon.

Soit une série temporelle de N nombres, notés $y(t)$. La méthode préconisée par Hurst consiste à prendre en considération une série de subdivisions indépendantes, d'effectif τ . Pratiquement, on part de séries d'effectif 10, puis 11, 12, 13, jusqu'à l'effectif le plus élevé permettant de distinguer deux subdivisions ($N/2$ ou $(N/2)-1$). Pour chaque subdivision considérée, la moyenne des τ données est

$$\langle y \rangle_{\tau} = \frac{1}{\tau} \sum_{t=1}^{\tau} y(t) \quad (39)$$

Pour chaque subdivision, on centralise les données, en leur soustrayant la moyenne locale. Puis on établit une série de valeurs cumulées à l'intérieur de chaque période : on ajoute à chaque donnée la somme des valeurs centrées qui la précèdent:

$$Y(t, \tau) = \sum_{u=1}^t \{y(u) - \langle y \rangle_{\tau}\} \quad (40)$$

Donc pour chaque valeur de t ($1 \leq t \leq \tau$), on a une valeur de $Y(t, \tau)$. $Y(t, \tau)$ est une distribution intégrée. On peut noter que si le signal original est un bruit blanc, $Y(t, \tau)$ est un mouvement brownien. L'étendue R (*range*) est la différence entre le minimum et le maximum de $Y(t, \tau)$

$$R = \max_{1 \leq t \leq \tau} Y(t, \tau) - \min_{1 \leq t \leq \tau} Y(t, \tau) \quad (41)$$

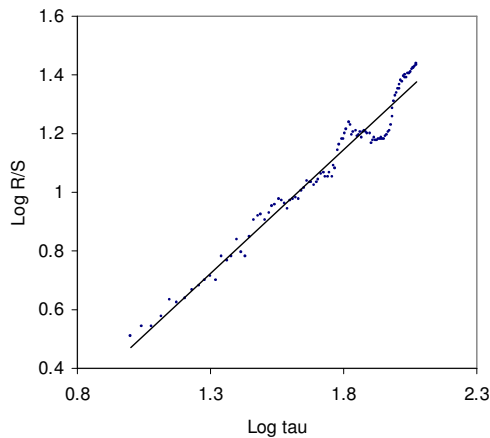


Figure 10: Estimation graphique de l'exposant de Hurst.

Cette étendue est ensuite normalisée au moyen d'une division par l'écart-type local ($S(t, \tau)$). Enfin on procède au moyennage, par niveau d'effectif, des étendues normalisées R/S . Hurst montre que l'étendue normalisée R/S , où S représente l'écart type de la série des $y(t)$ est liée à la grandeur de l'intervalle considéré par la relation suivante:

$$R/S = (a\tau)^H \quad (42)$$

Où a est une constante et H est l'exposant de Hurst. H peut être estimé par la méthode des moindres carrés, ou par l'estimation de la pente de la régression log/log de R/S sur τ . Expérimentalement, R/S doit être calculé pour un ensemble de valeurs de τ . Pour une variable aléatoire (bruit blanc), $H=0.5$,

c'est-à-dire:

$$a\tau = (R/S)^2 \quad (43)$$

L'exposant de Hurst est compris entre 0 et 1. Un exposant supérieur à 0.5 révèle des phénomènes de persistance: le « bruit » n'est pas aléatoire, et la variable est sensible à son histoire. La force du phénomène de persistance croît lorsque H approche de 1. A l'inverse un exposant inférieur à 0.5 révèle un phénomène d'anti-persistance.

L'analyse de Hurst est délicate pour les séries brèves. Des fluctuations importantes peuvent être obtenues avec des séries de moins de 500 données. Ceci explique pourquoi les subdivisions les plus faibles utilisées ne sont pas inférieures à 10. Enfin l'exposant de Hurst est sensible à la taille de l'échantillon. Ainsi, De la Fuente *et al.* (1998) montrent que l'exposant obtenu pour la composante x de l'attracteur de Lorenz est égale à 0.82 pour 5000 données, et 0.62 pour 40000. Par contre, pour un intervalle de temps donné, l'exposant de Hurst ne dépend pas de la fréquence d'échantillonnage (De la Fuente *et al.*, 1998).

6.4.2. La Detrended Fluctuation Analysis

L'analyse des fluctuations redressées (DFA) cherche à mettre en évidence les processus d'auto-similarité dans les séries temporelles. Elle utilise la démarche illustrée dans la figure 7, qui est répétée un grand nombre de fois, sur un ensemble de fenêtres indépendantes et de tailles diverses. Les écarts-types obtenus pour des fenêtres de tailles identiques sont moyennés, et l'exposant α est estimé à partir du graphe log-log de s sur n , en prenant en compte toutes les tailles d'échantillon analysées.

Le traitement de données expérimentales, et notamment physiologiques, pose cependant deux types de problèmes, que la méthode DFA tente de contourner. Le premier est celui du bornage de séries physiologiques dans des limites "normales", le second est celui de la non-stationarité.

Pour un processus auto-similaire avec un $\alpha > 0$, l'écart-type croît avec la taille de la fenêtre selon une loi puissance. De ce fait la série n'est pas bornée. Cependant, la plupart des séries temporelles physiologiques ne possèdent pas cette particularité: l'étendue de la

variable est bornée dans des limites, qui ne changent pas en fonction de la taille de la fenêtre d'observation (Hausdorff, Purdon, Peng, Ladin, Wei & Glodberger, 1996). De ce fait, on n'a pas besoin d'utiliser de facteur d'amplification ($M_y = 1$) pour mettre en évidence l'auto-similarité. Dans ce cas on obtient donc un α égal à 0, ce qui correspond au paramètre caractéristique des processus à bruit blanc. On contourne ce problème en appliquant l'analyse non pas sur la série originale, mais sur la série intégrée.

Un autre problème est que les séries physiologiques sont en général non stationnaires. Une série est stationnaire si la moyenne, l'écart-type et les moments d'ordre supérieur, ainsi que les fonctions de corrélation demeurent invariants malgré les translations temporelles. Les signaux ne remplissant pas ces conditions sont non-stationnaires, et la procédure d'intégration tend à exagérer cette non-stationnarité. La DFA tente de contourner ce problème, en redressant les données à l'intérieur de chaque fenêtre analysée.

Dans une première étape, la série originale $y(i)$ est intégrée. Pour cela on remplace chaque donnée par la somme cumulée des écarts à la moyenne :

$$y(k) = \sum_{i=1}^k [x(i) - \bar{x}] \quad (44)$$

On peut noter que si la série originale constitue un bruit blanc, la série intégrée sera un mouvement brownien. La série intégrée est ensuite divisée en fenêtres indépendantes de longueur équivalente n . Dans chaque fenêtre de longueur n , une droite des moindres carrés est estimée, représentant la tendance de cette fenêtre. On note $y_n(k)$ les coordonnées y sur ces droites. La série intégrée est ensuite redressée en retranchant la tendance locale, dans chaque fenêtre ($y(k) - y_n(k)$). L'analyse porte donc sur les résidus de la régression. Pour une longueur de fenêtre donnée, la grandeur caractéristique des fluctuations pour cette série intégrée et redressée est :

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2} \quad (45)$$

$F(n)$ peut être considéré comme un écart-type, calculé par rapport à la moyenne redressée. Ce calcul est répété pour toutes les tailles de fenêtre pour produire une relation entre n et $F(n)$. De manière typique, $F(n)$ croît avec n . Une relation linéaire, sur un graphe log-log révèle la présence d'auto-similarité (figure 11). La pente de la droite correspond au paramètre α , qui peut être considéré comme une estimation de l'exposant H de l'équation de Mandelbrot et van Ness (1968).

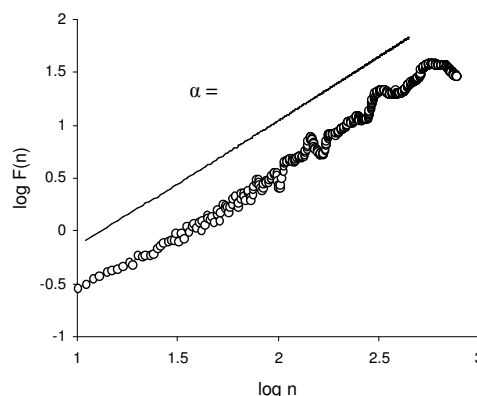


Figure 11 : Graphe de diffusion de la DFA.

Le paramètre α d'une série temporelle intégrée est lié à la fonction d'auto-corrélation $C(\tau)$ de la série originale. Pour un processus à bruit blanc, la valeur à l'instant t est non-corrélée avec les valeurs précédentes. La série intégrée correspond à une marche aléatoire (*random walk*), et donc $\alpha=0.5$. La fonction d'auto-corrélation est égale à 0 pour chaque décalage τ différent de 0.

On peut noter que de nombreux phénomènes naturels sont caractérisés par de fortes auto-corrélations sur une échelle de temps caractéristique τ_0 , et une fonction d'auto-corrélation qui décroît ensuite de manière exponentielle [$C(\tau) \approx \exp(-\tau/\tau_0)$]. La pente initiale de la droite $\log F(n)$ vs $\log n$ peut être différente de 0.5, mais approche 0.5 pour les tailles élevées de fenêtre. Ces processus sont caractéristiques des processus auto-régressifs et de moyenne mobile mis en évidence dans les procédures ARIMA.

Un α supérieur à 0.5 et inférieur ou égal à 1 révèle des auto-corrélations plus persistantes, décroissant selon une loi puissance [$C(\tau) \approx \tau^{-\gamma}$]. La relation entre α et γ est $\gamma = 2 - 2\alpha$. Le cas $\alpha=1$ est spécial et correspond à un bruit 1/f, ou bruit rose (*pink noise*).

Lorsque $\alpha > 1$, les corrélations existent mais ne suivent plus une loi puissance. $\alpha=1.5$ indique que la série originale est un mouvement brownien (une intégration de bruit blanc), appelé encore bruit brun (*brown noise*).

Enfin lorsque $0 < \alpha < 0.5$, des anti-corrélations, décroissant selon une loi puissance sont présentes, indiquant que des valeurs élevées sont suivies de valeurs faibles, et vice-versa.

L'exposant α peut également être vu comme un indicateur de la « rugosité » de la série originale. Plus α est grand, plus la série originale est lisse. Dans ce contexte, le bruit 1/f peut être interprété comme un compromis entre l'imprédictibilité du bruit blanc et le paysage plus lissé du bruit brownien.

6.4.3. La dispersional analysis

Cette méthode a été introduite par Basingthwaighte (1988). La série des $x(t)$ est divisée en intervalles contigus de longueur n . La moyenne de chaque intervalle est calculée, puis l'écart-type (SD) de ces moyennes locales, pour une longueur n donnée. Ces calculs sont répétés pour toutes les longueurs n possibles. SD est liée à n par une relation puissance :

$$SD \propto n^{H-1}$$

La quantité $(H-1)$ est exprimée au travers de la pente de la droite reliant n à SD , en coordonnées logarithmiques (figure 12)^o. Evidemment, comme le nombre de moyennes incluses dans le calcul dépend du nombre d'intervalles disponibles, le SD calculé pour les plus fortes valeurs de n tend à tomber sous la droite de régression et à fausser l'estimation. Caccia et al. (1997) suggèrent d'ignorer les mesures obtenues pour les intervalles les plus longs. Delignières et al. (2006) limitent l'analyse aux écart-types obtenus avec au moins 6 intervalles contigus. Comme la R/S analysis, la Dispersional Analysis ne fonctionne que sur le fGn, et donne des résultats erronés avec les séries de fBm.

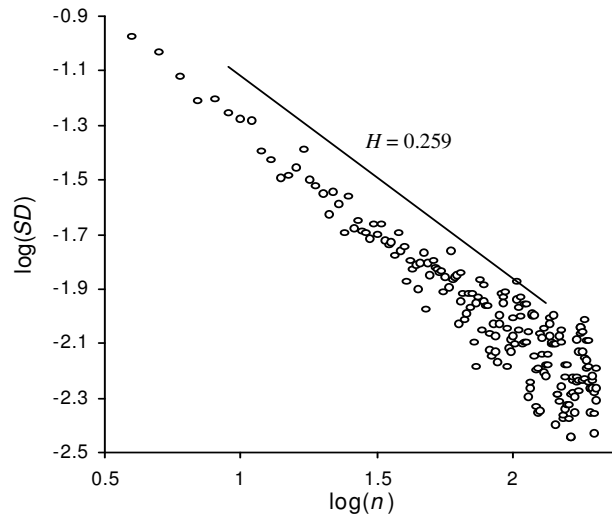


Figure 12 : Dispersional Analysis (séries temporelles d'estime de soi, d'après Delignières, Fortes et Ninot, 2004).

6.5. Modélisation ARMA/ARFIMA

Les méthodes présentées précédemment permettent de mesurer l'intensité des corrélations fractales dans les séries, mais ne peuvent en attester la présence. Ces dernières années, un enjeu a été de mettre au point des tests statistiques permettant de prouver la présence de dépendances à long terme dans les séries. C'est le but de la modélisation ARFIMA/ARMA, proposée par Wagenmakers, Farrell et Ratcliff (2005) et Torre, Delignières et Lemoine (2007).

On a précédemment proposé des formulations additives des modèles ARIMA. On peut également formaliser les modèles ARIMA à l'aide d'un « backshift operator », noté B et défini comme suit:

$$By_t = y_{t-1} \tag{46}$$

Soit par exemple un processus ARIMA (1, 0, 0)

$$\begin{aligned} y_t &= \mu + \phi y_{t-1} + \varepsilon_t \\ y_t - \phi y_{t-1} &= \mu + \varepsilon_t \\ (1 - B\phi)y_t &= \mu + \varepsilon_t \end{aligned}$$

Soit un processus (1, 1, 0)

$$\begin{aligned} y_t - y_{t-1} &= \mu + \phi(y_t - y_{t-1}) + \varepsilon_t \\ (1 - \phi)(y_t - y_{t-1}) &= \mu + \varepsilon_t \\ (1 - \phi)(1 - B)y_t &= \mu + \varepsilon_t \end{aligned}$$

Un modèle ARIMA (p,d,q) peut être généralement formalisé selon l'équation:

$$\phi(B)(1 - B)^d = \theta(B)\varepsilon_t \tag{47}$$

Avec $\phi(B) = 1 - B\phi_1 - B^2\phi_2 - \dots - B^p\phi_p$
 et $\theta(B) = 1 - B\theta_1 - B^2\theta_2 - \dots - B^q\theta_q$

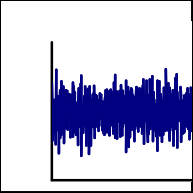
Les modèles ARIMA sont notés (p, d, q) , d étant un entier. Granger (1980) montre que l'on peut rendre compte de corrélations à long terme en permettant à d de prendre des valeurs fractionnaires. Il propose une nouvelle catégorie de modèles, les ARFIMA (auto-regressive fractionally integrated moving average).

La procédure proposée par Wagenmakers, Farrell et Ratcliff (2005) et Torre, Delignières et Lemoine (2007) consiste à comparer 9 modèles ARMA(p,q) et 9 modèles ARFIMA(p,d q), p et q variant systématiquement de 0 à 2. L'évaluation de ces modèles permet de calculer des scores de vraisemblance, et l'on peut sélectionner le meilleur modèle sur la base de critères combinant vraisemblance et parcimonie. Torre et al. (2007) montrent que les meilleures performances sont obtenues avec le Bayes Information Criterion (BIC). A noter également que l'évaluation des modèles ARFIMA teste l'hypothèse nulle $d = 0$, permettant donc d'apporter une preuve statistique de la présence de corrélations à long terme dans les données.

Torre et al. (2007) proposent d'accepter l'hypothèse de corrélation à long terme si (1) pour au moins 90% des séries analysées le meilleur modèle est un ARFIMA, d étant significativement différent de 0.

Références

- Box, G.E.P., & Jenkins, G.M. (1970). *Time Series Analysis, Forecasting and Control.*, San Francisco, CA : Holden-Day.
- Delcor, L., Cadopi, M., Delignières, D., & Mesure, S. (2003). Dynamics of the memorization of a morphokinetic movement sequence. *Neuroscience Letters*, 336, 25-28.
- Delignières, D., Fortes, M., & Ninot, G. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics in Psychology and Life Science*, 8, 479-510.
- Delignières, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M. & Ninot, G. (2006). Fractal analysis for short time series : A reassessment of classical methods. *Journal of Mathematical Psychology*, 50, 525-544.
- Eke, A., Herman, P., Bassingthwaighite, J.B., Raymond, G.M., Percival, D.B., Cannon, M., Balla, I., & Ikrényi, C. (2000). Physiological time series: distinguishing fractal noises from motions. *Pflügers Archives*, 439, 403-415.
- Fortes, M. (2003). *La dynamique de l'estime de soi et du soi physique. Un regard nouveau sur la variabilité et le fonctionnement des modèles hiérarchiques.* Thèse de doctorat STAPS, Université Montpellier I.
- Fortes, M., Delignières, D., & Ninot, G. (2004). The dynamics of self-esteem and physical self: Between preservation and adaptation. *Quality and Quantity*, 38, 735-751.
- Fortes, M., Ninot, G., & Delignières, D. (2004). The hierarchical structure of physical self: Cross-correlational analysis of individual time series. *International Journal of Sport and Exercise Psychology*, 2, 119-132.



Granger, C.W.J., & Joyeux, R. (1980). An introduction to long-memory models and fractional differencing. *Journal of time Series Analysis*, 1, 15-29.

Mandelbrot, B.B., & van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10, 422-437.

Simon, C. (2006). La dynamique des stratégies posturales de patients paraplégiques verticalisés sous stimulation électrique fonctionnelle. Mémoire de Master 2 Recherche, Université Montpellier I

Torre, K, Delignières, D., & Lemoine, L. (2007). Detection of long-range dependence and estimation of fractal exponents through ARFIMA modeling. *British Journal of Mathematical and Statistical Psychology*, 60, 85-106.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of $1/f^\alpha$ noise in human cognition. *Psychonomic Bulletin & Review*, 11, 579–615.