

Sur l'estimation de l'entropie de transfert dans la détection de connectivité cérébrale

JIE ZHU^{1,2}, JEAN-JACQUES BELLANGER^{1,2}, REGINE LE BOUQUIN JEANNES^{1,2}

¹ INSERM, U 1099, Rennes, F-35000, France

² Université de Rennes 1, LTSI, F-35000, France

Campus de Beaulieu, Bât 22, Université de Rennes 1, 35042, Rennes Cedex, France

^{1,2}{regine.le-bouquin-jeannes, jean-jacques.bellanger}@univ-rennes1.fr

Résumé - Ce travail s'inscrit dans le contexte d'une meilleure compréhension de l'activité épileptique et plus précisément de l'organisation des réseaux épileptogènes responsables de la génération et/ou de la propagation des crises et d'extraction de leurs caractéristiques, par application de méthodes statistiques d'analyse de connectivité. Plus particulièrement, il vise à préciser des difficultés liées à l'estimation de l'entropie de transfert, usuellement présentée comme pertinente dans la recherche de causalités au sein d'un ensemble d'enregistrements électroencéphalographiques. Nous mettons en évidence, au moyen de signaux simulés suivant des modèles autorégressifs vectoriels et un modèle physiologiquement réaliste, que les méthodes généralement proposées sont quelque peu en difficulté pour estimer cette entropie sur une durée d'observation n'excédant pas trois ou quatre secondes.

Abstract - This study comes within the scope of better understanding causal relationships between structures of the brain involved in the onset and propagation of epileptic seizures. Our objective concerns effective connectivity to discover whether neural activity in a given substructure influences activity in another area of the brain. This study is devoted to analyze transfer entropy, which is often regarded as an efficient nonlinear, nonparametric approach to overcome linear methods limitations. Now, building efficient transfer entropy estimators and testing them still ask open questions. The problem addressed in this contribution is the influence of the length of the observation when estimating this entropy using different estimators on short length signals, not exceeding a few seconds.

1 Introduction

Notre objectif est de contribuer aux méthodologies de mesure de connectivité cérébrale à des fins de diagnostic de l'épilepsie qui renvoient aux notions fondamentales de connectivité structurelle pour l'organisation anatomique des liaisons corticales, de connectivité fonctionnelle qui concerne les dépendances statistiques entre activités, et de connectivité effective qui cible les interactions causales entre différentes structures cérébrales. Pour détecter des relations causales, deux approches sont principalement envisagées. La première se base sur un modèle, qu'il soit de type boîte noire comme pour l'index de causalité de Granger [1] qui s'appuie sur une représentation AR, ou basé sur un fondement physiologique comme dans le cas de la modélisation causale dynamique [2]. La seconde approche, non paramétrique et non linéaire, est intéressante en absence d'*a priori* et en cas de dynamiques sous-jacentes non linéaires. Elle s'illustre particulièrement par l'entropie de transfert $TE_{x \rightarrow y}$ d'un signal X vers un signal Y . Il est utile ici de distinguer deux problèmes : le premier est l'estimation de l'entropie de transfert théorique et le deuxième est l'utilisation d'estimations $\widehat{TE}_{x \rightarrow y}$ et $\widehat{TE}_{y \rightarrow x}$ pour décider si X influe ou non sur Y et réciproquement. C'est la première question qui est principalement considérée ici.

2 L'entropie de transfert

En posant $U_n^m = (U_n, \dots, U_{n-m+1})$, l'entropie de transfert (TE) est définie sur une paire ordonnée $(X, Y)_{n \in \mathbb{Z}}$ de signaux conjointement stationnaires par la distance de Kullback-Leibler entre les densités de probabilité prédictives

$$P_{Y_{n+1}/Y_n^k=y^k, X_n^l=x^l}(y), y \in \mathbb{R}, (y^k, x^l) \in \mathbb{R}^{k+l}$$

et

$$P_{Y_{n+1}/Y_n^k=y^k}(y), y \in \mathbb{R}, y^k \in \mathbb{R}^k$$

pour mesurer la quantité d'information transmise du processus X vers le processus Y , et notée $TE_{x \rightarrow y}$. La première est définie conditionnellement à $X_{n'}$ et $Y_{n'}$, $n' < n$, à des instants antérieurs à n , et la seconde est seulement définie conditionnellement au passé de Y . On a donc :

$$TE_{x \rightarrow y} = \int_{\mathbb{R}^{k+l+1}} \log \left(\frac{P_{Y_{n+1}/Y_n^k=y^k, X_n^l=x^l}(y)}{P_{Y_{n+1}/Y_n^k=y^k}(y)} \right) d(y, y^k, x^l) \quad (1)$$

où $(y, y^k, x^l) \in \mathbb{R}^{1+k+l}$. En théorie, $TE_{x \rightarrow y}$ est non négative et égale à zéro ssi les densités prédictives sont pp égales, ce qui correspond alors à l'absence de causalité de X vers Y . Une difficulté pratique relève

du choix des paramètres k et l définissant les ordres des prédicteurs. Seul le choix $k = l = \infty$ correspondrait à une définition complètement non paramétrique. Dans [3], Schreiber relie ces ordres à un caractère markovien postulé conjointement pour $(X, Y)_{n \in \mathbb{Z}}$, *i.e.* $\forall (k' > k, l' > l)$:

$$P_{Y_{n+1}/Y_n^k, X_n^l}(\cdot) = P_{Y_{n+1}/Y_n^{k'}, X_n^{l'}}(\cdot)$$

ou marginalement, *i.e.* $\forall (k' > k, l > 0)$:

$$P_{Y_{n+1}/Y_n^{k'}, X_n^l}(\cdot) = P_{Y_{n+1}/Y_n^k}(\cdot).$$

Ceci équivaut à introduire deux hypothèses H_1 (influence causale de X vers Y) et H_0 (Y indépendant de X). Toute estimation $\widehat{TE}_{x \rightarrow y}$ de $TE_{x \rightarrow y}$, fonction d'une observation $(X, Y)_{n=1, \dots, N}$, peut alors être assimilée à une statistique utile pour discriminer entre ces hypothèses. Ainsi, une utilisation classique de l'entropie de transfert consiste à proposer un estimateur \widehat{TE} , à obtenir empiriquement sa distribution $p_{\widehat{TE}}^0$ sous H_0 (par exemple au moyen de données de substitution) et à calculer une p-value pour rejeter ou accepter H_0 . Disposer d'un bon estimateur permettra donc une discrimination performante. La question adressée se limite ici à une analyse comparative de performances pour différents estimateurs, en fonction de la durée d'observation.

2.1 Estimation

Si les densités de probabilité sont connues, l'estimateur de Monte-Carlo naturel de $TE_{x \rightarrow y}$, à partir des données (y_{n+1}, y_n^k, x_n^l) , $n = n_0, \dots, N-1$, avec $n_0 = 1 + \max(k, l)$, s'écrit

$$\widehat{TE}_{x \rightarrow y} = \frac{1}{N - n_0} \sum_{n=n_0}^{N-1} \log \left(\frac{P_{Y_{n+1}, Y_n^k, X_n^l}(y_{n+1}, y_n^k, x_n^l) P_{Y_n^k}(y_n^k)}{P_{Y_{n+1}, Y_n^k}(y_{n+1}, y_n^k) P_{Y_n^k, X_n^l}(y_n^k, x_n^l)} \right) \quad (2)$$

Les densités étant inconnues, deux approches sont possibles : on peut leur substituer des estimations non paramétriques à noyau où adopter une méthodologie comme celle proposée dans [4] et inspirée de [5], où l'information mutuelle est estimée en utilisant une approche du type K -plus proches voisins.

2.2 Estimateur basé sur les noyaux

Dans cette approche, notée KDE par la suite, l'estimation d'une distribution basée sur un noyau Ker s'écrit :

$$\hat{p}_U(u) = \frac{1}{N} \sum_{n=1}^N \frac{1}{h} Ker \left(\frac{u - u_n}{h} \right) = \frac{1}{N} \sum_{n=1}^N Ker_h(u - u_n) \quad (3)$$

soit encore, pour la densité de probabilité conjointe estimée en un point (y_{n+1}, y_n^k, x_n^l) , (*i.e.* une réalisation dans la loi de Y, Y^k, X^l) :

$$\hat{p}_{YYX} \left(y_{n+1}, y_n^k, x_n^l \right) = \frac{1}{|\Delta|} \sum_{m \in \Delta} \frac{1}{h^{k+l}} Ker_h(y_{n+1} - y_{m+1}) \prod_{j=1}^k Ker_h \left(y_n^k(j) - y_{m-j+1} \right) \prod_{j=1}^l Ker_h \left(x_n^l(j) - x_{m-j+1} \right) \quad (4)$$

où $\Delta = \{m / \max(k, l) < m \leq N-1, |m-n| > \tau_c\}$, et τ_c , le temps de décorrélation, sont introduits dans [6] pour éviter une sur-représentation des voisins temporels sur le support d'un noyau pouvant advenir du fait de la non indépendance stochastique des observations (on ne se trouve pas dans le cas d'observations i.i.d.). Le paramètre h à fixer impose la taille commune des noyaux gaussiens symétriques utilisée pour les signaux X et Y , préalablement normalisés. Dans l'équation (4), $u(j)$ représente la $j^{\text{ème}}$ composante du vecteur u . A partir de cette approche, et en nous appuyant sur des travaux en estimation non paramétrique de densités de probabilité [7], nous avons développé une version améliorée à noyau adaptatif, notée AKDE (Adaptive Kernel Density Estimator), comme suit :

Soit une largeur de bande $h_0 \in]0; 3]$ qui maximise (2).

Comme proposé dans [7], on calcule

$$h_n = h_0 \left(\hat{p}_n^{h_0} / g \right)^{-r} \quad (5)$$

où g est la moyenne géométrique de $(\hat{p}_n^{h_0}(x_m))_m$ et

$\hat{p}_n^{h_0} = \hat{p}_{YYX}^{h_0}(y_{n+1}, y_n^k, x_n^l)$. Le paramètre de sensibilité r satisfait généralement $0 < r < 1$. Comme il apparaît difficile de choisir une valeur de r pour les 4 estimations de densités de probabilité, nous avons choisi de ne calculer que $\hat{p}_{YYX}(y_{n+1}, y_n^k, x_n^l)$ selon (2) en laissant les autres densités inchangées. L'algorithme qui sélectionne une paire (r_s, h_s) s'écrit donc comme suit :

$V = \emptyset$

pour $h_0' \in \{v_1, v_2, \dots, v_{n_s}\} \subset]0; 3]$

calculer $F : r \in \{r_1, r_2, \dots, r_{n_r}\} \subset]0; 1[\rightarrow \widehat{TE}(r, \hat{p}^{h_0'})$

si F est unimodale, alors $V = V \cup \{h_0'\}$

fin pour

$$(r_s, h_s) = \arg \max_{r, h_0' \in V} \left[\widehat{TE}(r, \hat{p}^{h_0'}) \right]$$

$$\widehat{TE} = \max_{r, h_0' \in V} \left[\widehat{TE}(r, \hat{p}^{h_0'}) \right]$$

La valeur sélectionnée h_s est souvent proche de la valeur initiale h_0 , ce qui permet de limiter l'espace de recherche.

2.3 Estimateur basé sur la fonction digamma

Cet estimateur, noté KNN par la suite, a la forme :

$\widehat{TE}_{x \rightarrow y} = \psi(K) + \left\langle \psi(p_{y_n^k} + 1) - \psi(p_{y_{n+1} y_n^k} + 1) - \psi(p_{y_n^k x_n^l} + 1) \right\rangle$
où $\psi(\bullet)$ est la fonction digamma, $\langle \bullet \rangle$ traduit la

moyenne sur les échantillons, et p_c désigne le nombre de voisins dans les espaces correspondants, la dimension maximale étant $k+l+1$. Le nombre de voisins dans l'espace de plus grande dimension est pris égal à K .

Pour les 2 estimateurs proposés, les ordres de prédiction $k \leq k_{\max}$ et $l \leq l_{\max}$ sont déterminés en utilisant le critère BIC (Bayesian Information Criterion) généralisé [8], k_{\max} et l_{\max} étant fixés par l'utilisateur.

3 Résultats

Nous avons expérimenté les deux estimateurs précédents et les avons comparés à deux techniques de référence. La première est l'indice de causalité de Granger divisé par deux (noté GC/2, équivalent à l'entropie de transfert dans le cas gaussien, utilisant k et l donnés par BIC). La deuxième est celle disponible dans la boîte à outils Trentool (notée TRL) [4], méthode qui utilise une technique de plongement avec reconstruction d'une trajectoire d'état [9]. Deux modèles autorégressifs (AR) vectoriels et un modèle physiologique ont été considérés. Pour simuler des situations réalistes, ces modèles génèrent des séries temporelles stationnaires présentant des densités spectrales de puissance à bande étroite et, de ce fait, des fonctions de corrélation temporelles à décroissance lente. Le premier modèle AR est unidirectionnel (modèle 1) et les signaux générés sont de la forme :

$$\begin{cases} x_n = 0.95\sqrt{2}x_{n-1} - 0.9025x_{n-2} + u_n \\ y_n = 0.5x_{n-3} - 0.4y_{n-2} + v_n \end{cases} \quad (6)$$

où u_n et v_n sont des bruits blancs gaussiens centrés, normés et indépendants. Sur la Figure 1 sont données les évolutions des différents estimateurs en fonction de la longueur des signaux pour 100 essais. Pour les méthodes KNN et TRL, le nombre de voisins était fixé à 4 et, pour AKDE, $\tau_c = 20$. Les estimateurs s'avèrent tous présenter un biais important et un écart-type non négligeable pour des durées courtes ($n=8$, soit 256 points correspondant à 1 s d'estimation). Si l'estimateur de l'indice de Granger est le moins biaisé, sa variance est, en revanche, très élevée sur des horizons courts, à l'inverse de l'estimateur KDE, dont les fluctuations sont plus réduites mais qui présente un biais important avec un comportement asymptotique difficile à apprécier. L'estimateur KNN proposé est un compromis intéressant sur l'ensemble des estimateurs retenus, témoignant de défauts comparables mais plus limités. Notons que la valeur exacte de l'entropie de transfert de X vers Y vaut ici 0,41 (obtenue avec GC/2 et N grand). A titre d'illustration, nous avons également représenté sur la Figure 2 un exemple de résultat de la méthode AKDE pour différentes valeurs de h_0 et pour une longueur de signal $N = 2^{11}$. La valeur initiale à retenir ici est $h_0 = 0,42$ conduisant à une valeur de l'entropie de transfert effectivement proche de la valeur exacte (0,41).

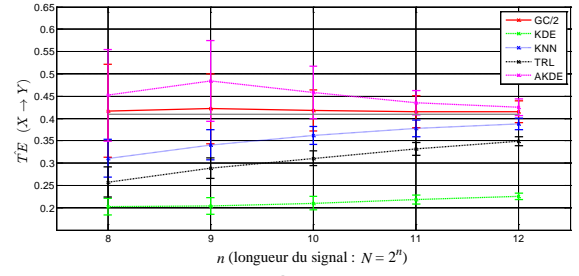


Figure 1. $\widehat{TE}_{x \rightarrow y}$ (modèle 1)

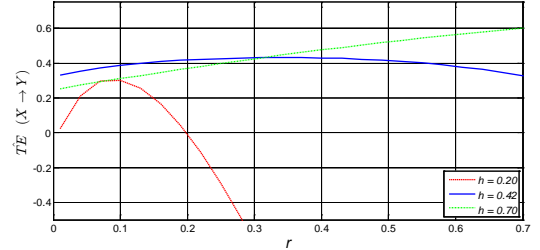


Figure 2. $\widehat{TE}_{x \rightarrow y}$ ($N = 2^{11}$, modèle 1)

Le second modèle AR est bidirectionnel (modèle 2) :

$$\begin{cases} x_n = 0.5x_{n-1} + 0.3y_{n-2} + u_n \\ y_n = 0.5x_{n-3} - 0.4y_{n-1} + v_n \end{cases} \quad (7)$$

où u_n et v_n sont des bruits tels que ceux définis à l'équation (6). Les conclusions issues de l'analyse précédente s'étendent ici comme illustré Figures 3.a et 3.b, les valeurs exactes de l'entropie de transfert étant alors respectivement de 0,15 et 0,06 pour $TE_{x \rightarrow y}$ et $TE_{y \rightarrow x}$.

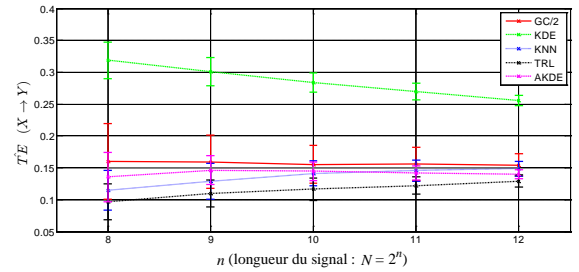


Figure 3.a. $\widehat{TE}_{x \rightarrow y}$ (modèle 2)

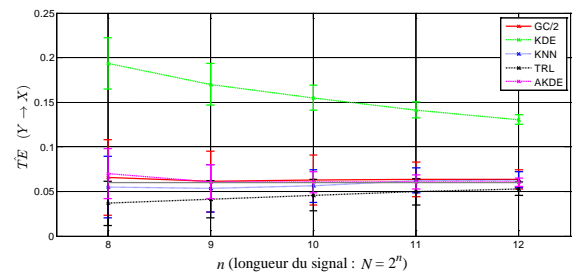


Figure 3.b. $\widehat{TE}_{y \rightarrow x}$ (modèle 2)

Le modèle physiologique de populations neuronales retenu est celui décrit dans [10] et comporte deux populations X et Y . C'est un modèle du type agrégation de paramètres (lumped parameters model) qui, pour chaque population, a la forme d'un système d'équations différentielles stochastique, l'une des équations incluant deux variables d'entrée. La première représente les incréments d'un processus brownien, s'interprétant comme l'influence de l'activité neuronale de fond dans l'ensemble des neurones connectés à la

population d'intérêt, la seconde étant fonction de l'état dynamique de l'autre population dont on modélise l'influence de manière privilégiée. Ces équations intègrent des opérateurs non linéaires modélisant des effets de seuillage et de saturation. Dans la situation modélisée, seule la population X influence la population Y . Sans vérité terrain sur les valeurs de TE, nous donnons sur les Figures 4.a et 4.b des résultats moyennés sur 50 réalisations, sans recouvrement. Les paramètres retenus étaient $k_{\max} = l_{\max} = 2$ pour GC/2, KDE AKDE, et KNN, $\tau_c = 50$ pour AKDE, et le nombre de voisins égal à 4 pour KNN et TRL. Les résultats montrent un biais et une variance qui se stabilisent à partir de 2^{10} points (soit 4 s).

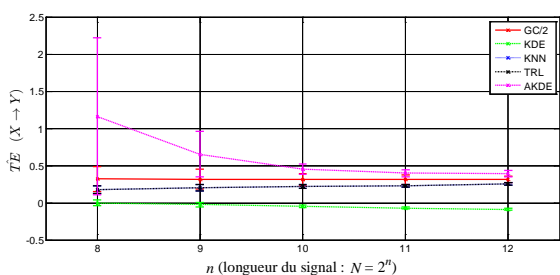


Figure 4.a. $\widehat{TE}_{x \rightarrow y}$ (modèle physiologique)

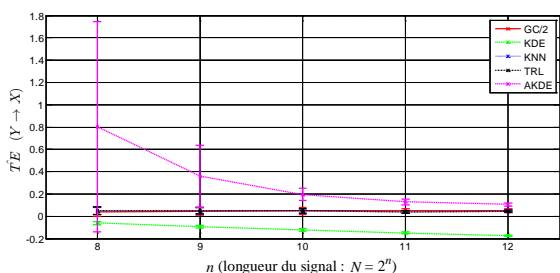


Figure 4.b. $\widehat{TE}_{y \rightarrow x}$ (modèle physiologique)

Ces résultats sont résumés en calculant l'indice de performance (critère de contraste) Ind , donné par

$$Ind = \frac{2(\widehat{TE}_{x \rightarrow y} - \widehat{TE}_{y \rightarrow x})}{\sigma_{\widehat{TE}_{x \rightarrow y}} + \sigma_{\widehat{TE}_{y \rightarrow x}}}, \quad (8)$$

(où σ représente l'écart-type) dans le Tableau 1, à l'exception de ceux obtenus avec l'approche KDE, qui présente un biais excessif.

Tab.1. Indice de performance (Ind)

	$n = 8$	$n = 9$	$n = 10$	$n = 11$	$n = 12$
GC/2	2.742	3.210	4.641	6.259	9.174
KNN	2.922	4.284	7.582	11.472	15.566
TRL	2.921	4.284	7.582	11.472	15.566
AKDE	0.358	1.010	4.193	7.179	9.317

4 Discussion et conclusion

Même dans le cas simple des modèles AR gaussiens, l'estimateur de Granger présente une variance relativement élevée. Une explication *a priori* plausible pour cette variance élevée était l'effet de la forte corrélation temporelle des séries temporelles étudiées.

Toutefois, des simulations de suites i.i.d. $(Y_{n+1} Y_n^k X_n^l)_n$ de même loi gaussienne que pour le modèle AR unidirectionnel testé n'ont en rien diminué la variance de cet estimateur ni celle de ses concurrents.

De cette analyse, pour les signaux et modèles testés, il ressort par ailleurs les points suivants : (i) la méthode KDE est très biaisée, (ii) l'approche AKDE proposée améliore sensiblement la méthode originale en termes de biais et de variance, (iii) pour des durées inférieures ou égales à 2 s, seule AKDE concurrence l'indice de Granger pour les signaux AR à bande étroite, (iv) AKDE est préférable au logiciel en libre accès Trentool (TRL) ce dernier étant cependant amélioré dans la version KNN proposée qui choisit différemment les ordres et les variables de prédiction, (v) pour le modèle physiologique non linéaire, les méthodes proposées AKDE et KNN détectent correctement le flux de propagation sans gain supplémentaire par rapport à l'indice de Granger ce qui peut s'expliquer par le caractère lisse des non-linéarités intervenant dans ce modèle de populations épileptiques.

Bien que l'estimateur AKDE semble présenter des avantages, en termes de complexité algorithmique, son coût de calcul s'avère le plus élevé, puisque, sous Matlab, pour 2^9 points, il se révèle 25 fois plus important que celui de KNN. Notons également que l'indice de Granger est 2,5 fois plus coûteux que l'estimateur KNN. Pour conclure, outre la proposition d'un nouvel estimateur à noyau de l'entropie de transfert, une comparaison de méthodes linéaire et non linéaires sur différents types de modèles de signaux épileptiques a mis en exergue la difficulté d'élire une méthode globalement optimale.

Références

- [1] C.W.J. Granger, "Investigating causal relations by econometric models and cross-spectral methods", *Econometrica*, vol. 37, pp. 424-438, Aug. 1969.
- [2] K.J. Friston, L. Harrison, and W.D. Penny, "Dynamic Causal Modelling", *NeuroImage*, 19(4):1273-1302, 2003.
- [3] T. Schreiber, "Measuring information transfer", *Physical Review Letters*, vol. 85, pp. 461-464, 2000.
- [4] M. Lindner, R. Vicente, V. Priesemann, M. Wibral, "TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy", *BMC Neuroscience*, 12:119, 2011.
- [5] A. Kraskov, H. Stoegbauer, P. Grassberger, "Estimating mutual information", *Physical Review E*, 69, 066138-1:16, 2004.
- [6] A. Kaiser and T. Schreiber, "Information transfer in continuous processes", *Physica D*, vol. 166, pp. 43-62, 2002.
- [7] J.N. Hwang, "Nonparametric multivariate density estimation: a comparative study", *IEEE Trans. Signal Proc.*, vol. 42, pp. 2795-2810, 1994.
- [8] C. Yang, R. Le Bouquin Jeannès, J.J. Bellanger, H. Shu, "A new strategy for model order identification and its application to transfer entropy for EEG signals analysis", *IEEE Trans. on Biomedical Engineering*, vol. 60, no. 5, pp. 1318-1327, May 2013.
- [9] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series", *Physica A*, vol. 110, pp. 43-50, 1997.
- [10] F. Wendling, A. Hernandez, J.J. Bellanger, P. Chauvel, and F. Bartolomei, "Interictal to ictal transition in human temporal lobe epilepsy: insights from a computational model of intracerebral EEG", *Journal of Clinical Neurophysiology*, vol. 22, pp. 343-356, 2005.