

# Estimateurs de relations inter-canaux : précision et intérêt clinique en épilepsie

Abel, KINIE, Jean-Jacques BELLANGER, Fabrice WENDLING

Laboratoire Traitement du Signal et de l'Image, INSERM  
Université de Rennes1 – Bât 22, Campus de Beaulieu – 35042 RENNES Cedex

Abel.Kinie@univ-rennes1.fr, Jean-Jacques.Bellanger@univ-rennes1.fr  
Fabrice.Wendling@univ-rennes1.fr

**Résumé** – Nous considérons dans cet article les problèmes posés, dans le cadre de l'analyse de signaux stéréoelectroencéphalographiques de profondeur en épilepsie, par l'évaluation au cours du temps (très importante pour l'interprétation des crises) du « couplage » statistique entre les signaux recueillis pour une paire arbitraire de capteurs. D'une part le caractère non stationnaire de ce couplage mène à procéder à des estimations sur des fenêtres temporelles relativement courtes et d'autre part durant les crises d'épilepsie les signaux enregistrés présentent un caractère non linéaire affirmé. Ces deux points appauvrissent clairement l'arsenal des méthodes *a priori* disponibles pour évaluer la précision des estimateurs pouvant être calculés dans ce contexte. Pour surmonter cette difficulté nous proposons d'utiliser des méthodes de rééchantillonnage (Bootstrap) puisque ces dernières présentent l'avantage de minimiser les informations probabilistes *a priori* nécessaires. L'absence de valeur de référence pour les statistiques de couplage mesurées sur les signaux réels et le volume disponible limité pour ces derniers rend difficile la validation expérimentale des méthodes de Bootstrap. Nous répondons ici à cette difficulté en introduisant un modèle non linéaire fondé physiologiquement. Des résultats préliminaires sont présentés et discutés.

**Mots clés** : regression nonlinéaire, Bootstrap, précision , EEG, épilepsie

**Abstract** – The general context of this paper is the analysis of stereoelectroencephalographic (depth recording) signals. In this context, we consider the problems linked to the evaluation of the statistic coupling (of major importance for the interpretation of seizures) between signals recorded on a pair of sensors. On the one hand, the non stationary characteristics of this coupling leads to estimate the quantities on sliding windows of relatively short durations. On the other hand, during the seizures, recorded signals exhibit actual non linear characteristics. These two points diminish the number of possible methods able to evaluate the accuracy of the estimators that can be used. To bypass this difficulty we propose to use re-sampling methods (Bootstrap) presenting the advantage of minimizing the *a priori* necessary probabilistic information. The absence of reference values for the coupling statistics measured on real signals, added to the limited amount available data, make the experimental validation of bootstrap methods difficult. Here, we address this difficulty by using a non linear physiologically relevant model. Preliminary results are presented and discussed.

**Key word** : nonlinear regression, bootstrap, accuracy , EEG, epilepsy.

## 1. Introduction

Dans l'analyse clinique des épilepsies, il est essentiel d'appréhender comment se recompose, fonctionnellement, le système d'intercommunications entre différentes régions du cerveau durant les crises. Une information disponible qui traduit, avec une résolution temporelle maximale, l'évolution dynamique de ces mécanismes est celle fournie par les enregistrements stéréoelectroencéphalographiques de profondeur (SEEG) obtenus au moyen de capteurs implantés chirurgicalement au sein de régions du cerveau soupçonnées de participer, en interagissant anormalement entre elles, au démarrage des crises. Ces signaux sont utilisés dans le travail présenté ici. Etudier l'évolution des couplages entre différents groupes neuronaux amène à analyser le comportement statistique conjoint des signaux recueillis au sein des structures concernées. Pour cela ont été utilisées dans la littérature différentes méthodes de traitement du signal visant

à estimer sur une fenêtre glissante pour une paire de signaux scalaires ( $X_1$ ,  $X_2$ ) et sous l'hypothèse de stationnarité locale (une à quelques secondes) une fonctionnelle  $R$  de leur loi de probabilité conjointe.

Dans ce contexte, les questions méthodologiques qui se posent alors, de manière non indépendantes, sont :

- Quelles fonctionnelles et quels estimateurs de ces fonctionnelles retenir ? Quelles sont les précisions de ces estimateurs ?

- Est-il possible d'introduire des éléments de modélisation réaliste (non gaussienne et non linéaire) pour les signaux étudiés exploitables pour le point précédent ?

Dans ce papier qui englobe et étend des études que nous avons déjà présentées [1] nous proposons : 1) une méthodologie de simulation réaliste de signaux EEG dans le cadre de l'épilepsie, 2) une méthodologie expérimentale dont l'objet est, au moyen de signaux simulés comme en 1), d'évaluer la précision pour l'estimation de différentes

fonctionnelles. Son organisation est la suivante. Dans les § 2, § 3 et § 4 nous décrirons les fonctionnelles étudiées, leurs estimateurs, et proposons des méthodes de rééchantillonnage (Bootstrap) pour évaluer la précision de ces estimateurs. Dans le § 5 nous présentons un modèle de simulation non linéaire pertinent pour le domaine d'application visé. Dans le § 6 nous présentons quelques résultats préliminaires et nous concluons.

## 2. Fonctionnelles et estimateurs étudiés

### 2.1 Cadre formel de l'étude

Nous considérons une paire de signaux aléatoires réels  $(X_1(t), X_2(t))$  observés pour des valeurs de  $t$  dans un intervalle  $[0, Te]$ . Ici, les valeurs de  $t$  et des différentes variables temporelles seront assimilées à des multiples  $k\theta e$  d'une période d'échantillonnage  $\theta e$  et les intervalles temporels aux ensembles discrets d'échantillons  $y$  étant inclus.

Soit  $P\Delta n$  la mesure de probabilité correspondant à la restriction à l'intervalle  $\Delta n = [nb, L + nb] \in [0, Te]$  de la loi conjointe de  $(X_1, X_2)$  sur  $[0, Te]$ . La suite  $\Delta n, n = 0, 1, \dots, M$  correspond à une suite des fenêtres temporelles de  $1 + L$  échantillons chacune et débutant aux instants  $nb$  avec  $Mb + L \leq Te$ . Si  $(X_1, X_2)$  est stationnaire sur  $[0, Te]$  la loi ne dépend pas de  $n$ . En pratique, pour les signaux qui nous intéressent,  $P\Delta n$  varie avec  $n$  et on désire produire à partir des échantillons sur  $\Delta n$  des indicateurs  $R(P\Delta n)$  de la liaison statistique entre  $X_1$  et  $X_2$  sur  $\Delta n$ . Formellement  $L$  est une fonctionnelle associant à  $P\Delta n$  un nombre entre 0 et 1. On impose  $R(P\Delta n) = 0$  dans le cas où  $X_1$  et  $X_2$  sont indépendants sur  $\Delta n$ ,  $R(P\Delta n)$  pouvant atteindre la valeur maximale de 1 lorsqu'il existe une relation déterministe entre les échantillons de  $X_1$  et de  $X_2$ . Pour mesurer, pour chaque  $n$ , la valeur de  $R(P\Delta n)$  il est nécessaire d'introduire un estimateur  $\hat{R}\Delta n(X_1, X_2)$  calculé à partir des échantillons de  $X_1$  et  $X_2$  sur  $\Delta n$ . On se rend compte facilement que si aucune information *a priori* n'est donnée sur la forme de  $P\Delta n$ , on ne peut pas envisager en général d'estimer correctement  $R(P\Delta n)$  à partir de la seule paire de vecteurs  $(l+1)$  dimensionnels  $X_1$  et  $X_2$ .

Ceci amène à contraindre  $R(P\Delta n)$  à être de la forme  $R(P\delta n)$  où  $P\delta n$  est la restriction de  $P\Delta n$  aux échantillons de  $X_1$  et  $X_2$  correspondants à deux sous ensembles  $\delta_{1n}$  et  $\delta_{2n}$  d'instant d'échantillonnage inclus chacun dans  $\Delta n$ , le premier comportant  $l_1 < L$  échantillons de  $X_1$  et le deuxième  $l_2 < L$  échantillons de  $X_2$ .

En désignant par  $\delta_n + k$  l'ensemble des instants de  $\delta_{1n}$  et  $\delta_{2n}$  décalés simultanément de  $k\theta e$  et en tenant compte de la stationnarité de  $P\Delta n$  sur  $\Delta n$ , il est clair que les lois  $P_{\delta_n+k}$  sont identiques pour tout décalage  $k$  maintenant les instants d'échantillonnage dans la fenêtre  $\Delta n$ . En considérant un

nombre  $K$  de décalages, il est alors possible d'utiliser à  $n$  fixé les  $K$  vecteurs  $y_{n,k}, k = k_1, \dots, k_K$ , de la forme  $y_{n,k} = [X_1(i \cdot \theta e), i \in \delta_{1n} + k, X_2(j \cdot \theta e), j \in \delta_{2n} + k]^T$ , pour estimer  $R(P_{\delta_n})$ . Les valeurs  $k_1, \dots, k_K$  peuvent être choisies de manière à éviter une redondance d'information puisqu'en général  $y_{n,k}$  et  $y_{n,k'}$  ne seront pas indépendants pour  $|k - k'|$  trop faible. Un estimateur  $\hat{R}_{\delta_n}(X_1, X_2)$  de  $R(P_{\delta_n})$  se ramènera alors à une statistique  $T(y_{n,k}, k = k_1, \dots, k_K)$  qu'on notera  $T_n$  dans la suite.

### 2.2 Fonctionnelles introduites dans l'étude

Notre but n'est pas ici de proposer des indicateurs de relation  $R(P_{\delta_n})$  originaux mais de bien situer par rapport à notre présentation générale des indicateurs déjà introduits dans la littérature et de proposer une méthodologie pour les exploiter en connaissance de cause quand à leur précision lorsqu'on les estime sur une suite d'intervalles  $\Delta n$  recouvrant le support d'une crise.

Nous nous sommes ainsi limités aux trois fonctionnelles  $R$  suivantes qui, chacune, n'utilise que des lois marginales bidimensionnelles de  $P_{\delta_n}$  (1 échantillon pour  $X_1$  et 1 échantillon pour  $X_2$ ).

#### 2.2.1 Module de la fonction de cohérence moyennée sur une bande de fréquences

En considérant que la longueur des fenêtres  $\Delta n$  est plus grande que le temps de corrélation (et d'intercorrélation) des fonctions d'auto et d'intercorrélation  $\Gamma_{X_1}, \Gamma_{X_2}, \Gamma_{X_1, X_2}$  les lois marginales d'ordre 2 de  $P\Delta n$  suffisent à leur connaissance et donc à celle des densités spectrales  $\gamma_{X_1}, \gamma_{X_2}, \gamma_{X_1, X_2}$  associées. A  $P\Delta n$  on peut alors associer sans ambiguïté la fonction de cohérence définie sur  $[-1/2, 1/2]$

$$\rho_{1,2}(f) = \frac{\gamma_{X_1 X_2}}{\gamma_{X_1}^{1/2} \gamma_{X_2}^{1/2}} \quad (1)$$

L'indicateur de relation que nous associons à  $\rho_{1,2}$  est :

$$R_1(P_{\delta_n}) = \overline{\rho_{1,2}(F)} = \frac{1}{\#(F)} \sum_{f \in F} |\rho_{1,2}(f)| \quad (2)$$

Il correspond à la moyenne arithmétique des modules de  $\rho_{1,2}$  sur un ensemble de fréquences  $F$  présélectionné par l'expérimentateur. Il prend ses valeurs entre 0 et 1. L'indépendance entre  $X_1$  et  $X_2$  implique la valeur nulle et une dépendance sous forme de filtrage linéaire entre  $X_1$  et  $X_2$  implique la valeur unité si  $F$  est inclus dans le support de la réponse fréquentielle du filtre. L'ensemble  $\delta_n$  peut être constitué dans ce cas de  $\delta_{1n} = \{nb\}$  et  $\delta_{2n} = \{nb, nb+1, \dots, nb+\tau\}$  où  $\tau$  est tel que  $X_1(t)$  et  $X_2(t')$  deviennent indépendants si  $|t - t'| > \tau$

### 2.2.2 Erreur quadratique moyenne normalisée de régression non linéaire de $X_2$ par rapport à $X_1$

Cette fonctionnelle n'utilise que des lois marginales bidimensionnelles. Pour la définir est introduite la fonction du décalage temporel  $\tau$

$$h_2(\tau) = 1 - \frac{\text{VAR}(X_2(t+\tau) / X_1(t))}{\text{VAR}(X_2(t))}$$

où au numérateur VAR désigne une variance conditionnelle. L'indicateur est ici :

$$R_2(P_{\delta n}) = h_m^2 = \max_{|\tau| \leq \tau_{\max}} h^2$$

$\tau_{\max}$ , fixé par l'opérateur, s'interprète ici comme le plus grand décalage au delà duquel  $X_1(t)$  et  $X_2(t+\tau)$  deviennent des variables aléatoires indépendantes et  $\delta n$  correspond à

$$\delta_{1n} = \left\{ nb + \frac{\tau_{\max}}{\theta e} \right\} \text{ et } \delta_{2n} = \left\{ nb, nb+1, \dots, nb + \frac{2\tau_{\max}}{\theta e} \right\}$$

$h_m^2$ , à valeurs dans  $[0,1]$ , est nulle quand  $X_1$  et  $X_2$  sont indépendants. Sa valeur est égale à 1 s'il existe un  $\tau$  et une fonction  $g: \mathfrak{R} \rightarrow \mathfrak{R}$  telle que  $X_2(t+\tau) = g(X_1(t))$ .

### 2.2.3 Entropie mutuelle normalisée (EMMN)

L'indicateur est :

$$R_3(P_{\delta n}) = EMMN = \max_{|\tau| \leq \tau_{\max}} \frac{H_1 + H_2 - H_{1,2}(\tau)}{\min(H_1, H_2)}$$

où  $H_1$ ,  $H_2$  et  $H_{1,2}(\tau)$  correspondent respectivement aux entropies marginales des lois d'ordre 1 de  $X_1$  et  $X_2$  sur  $\delta n$  et à l'entropie conjointe associée à la loi bidimensionnelle de  $(X_1(t), X_2(t+\tau))$  pour  $t$  et  $t+\tau$  « inclus » dans  $\delta n$  qui est ici identique à  $\delta n$  pour  $h_m^2$ . Si  $X_1$  et  $X_2$  sont indépendants  $R_3(P_{\delta n})$  est clairement égale à zéro, la valeur 1 n'étant atteinte, par contre, que si il existe une relation bijective entre  $X_1(t)$  et  $X_2(t+\tau)$  pour une certaine valeur de  $\tau$ .

## 3. Estimation des fonctionnelles 1), 2), 3)

- *Statistique  $T_{c,n}$  pour estimer de la cohérence moyennée  $\bar{\rho}_{1,2(F)}$  sur  $\Delta n$* . Elle est obtenue en remplaçant dans (2)  $\rho_{1,2}(f)$  par son estimation elle même obtenue en substituant aux densités spectrales simples et croisées dans (1) des estimations. Ces dernières sont calculées classiquement, de deux manières. La première non paramétrique utilise la FFT de manière classique en apportant éventuellement une correction du biais [2]. La deuxième utilise un algorithme d'identification AR vectoriel par moindres carrés exacts sous sa forme en treillis. Ces deux approches ont été appliquées à l'analyse en EEG dans la littérature [3] sans avoir été comparées systématiquement du point de vue de leur précision.

- *Statistique  $T_{h_m^2}$  pour estimer de  $h_m^2$  sur  $\Delta n$* . Pour  $h_m^2$  nous avons utilisé une partition du domaine des valeurs de

$X_1(t)$  en classes et un calcul pour chaque valeur de  $\tau$  de la moyenne de  $X_2(t+\tau)$  conditionnellement à chaque classe pour construire une fonction affine approximant par morceaux  $E(X_1(t) / X_2(t+\tau))$ .

- *Statistique  $T_{EMMN,n}$  pour estimer EMMN sur  $\Delta n$* . Pour l'entropie mutuelle normalisée nous avons utilisé pour estimer la loi conjointe et les lois marginales de  $(X_1(t), X_2(t+\tau))$  un histogramme bidimensionnel.

## 4. Dispersion des statistiques $T_{c,n}$ , $T_{h_m^2}$ et $T_{EMMN,n}$

Des résultats génériques existent, en dehors du cadre asymptotique, seulement pour  $T_{c,n}$  estimé par la méthode FFT (autopériodogrammes et interpériodogrammes, calculés sur des blocs  $\delta n + k$  présentant un recouvrement partiel bien défini, et moyennés sur différentes valeur de k [2]. Il se trouve que pour des signaux gaussiens la variance et le biais de l'estimation (FFT) de  $|\rho_{1,2}(f)|$  ne dépendent de  $P(\delta n)$  que par l'intermédiaire de  $|\rho_{1,2}(f)|$  elle même. Pour la méthode non paramétrique (AR) seuls les résultats asymptotiques existent dans le cas où l'observation répond effectivement à un modèle AR. Des résultats sur la variance de  $T_{EMMN,n}$  existent [1] pour  $\tau$  fixé dans le cas où les vecteurs  $y_{n,k}$  sont statistiquement indépendants. Ils dépendent de la distribution bidimensionnelle de  $X_1(t)$ ,  $X_2(t+\tau)$ , et ne débouche sur une formule simple que dans le cas gaussien. Pour  $T_{h_m^2}$  une étude de sa distribution empirique a été effectuée dans [1] uniquement sous l'hypothèse d'indépendance sur des segments SEEG réels. Il est cependant difficile d'affirmer que ce résultat puisse s'étendre à d'autres base de données. On dispose donc en général de très peu d'informations pour associer à une statistique  $T_n$  mesurée sur un segment  $\Delta n$  un intervalle de confiance ou toute caractérisation de sa dispersion.

## 5. Evaluation de la précision des $T_n$ par rééchantillonnage

Les difficultés décrites dans le paragraphe précédent sont du même type que celles qui ont amené au développement des méthodes de Bootstrap [4] pour évaluer des distributions d'estimateurs dans un cadre limité de connaissance *a priori* ou lorsque les calculs théoriques sont trop complexes. Ces méthodes initialement développées pour des suites d'échantillons indépendants ont été aménagées pour certaines situations de dépendance comme par exemple, en traitement du signal, pour estimer des paramètres d'un modèle AR [4]. Nous avons initié une étude expérimentale pour apprécier la validité de ce type d'approche dans le cadre de notre problème. La méthodologie est la suivante, pour chaque type de statistique :

- définition d'une technique de rééchantillonnage pour répliquer pour  $\Delta n$  donné,  $n \in \Delta n$  en un nombre  $N_R$  de répliqués ( $X_{1R}, X_{2R}$ )( $n\theta e$ ),  $n \in \Delta n$ .
- calcul de  $T_n$  pour un ensemble de  $\Delta n$  et pour chacun d'entre eux pour les  $N_R$  versions répliquées et construction de sa distribution empirique sur les versions répliquées.
- Comparaison de la distribution obtenue en b) avec une distribution de référence pour  $T_n$ .

Nous ne détaillons pas ici les techniques a) utilisées pour les trois statistiques étudiées. En résumé, nous utilisons deux types de méthodes. Le premier type concerne les trois statistiques étudiées et utilise une généralisation de la technique de répliqués, utilisée pour un processus AR monodimensionnel [4], à deux voies d'observation en introduisant une identification AR vectorielle et en rééchantillonnant aléatoirement dans la suite vectorielle des innovations ainsi obtenues. Le deuxième type concerne  $T_{h_m^2}$  et

$T_{EMMN}$  et s'inspire des techniques de bootstrap en régression non linéaire d'une variable simple par rapport à une autre tout en répliquant  $X_1$  sur  $\Delta n$  par bootstrap AR monovoie. Pour ce qui est de c) qui constitue une étape de validation la difficulté est évidemment de disposer d'une distribution de référence. Comme nous ne disposons pas de signaux réels tels qu'on connaîtrait la valeur des  $T_n$  sur un grand nombre de  $\Delta n$ , la seule solution est de faire appel à des simulations. Notre contribution sur ce plan est de proposer un modèle de génération non linéaire de signaux SEEG argumenté physiologiquement. Nous le présentons brièvement dans le paragraphe suivant.

### 5.1 Principe des simulations utilisant un modèle physiologique de génération de l'EEG

Nous avons utilisé un modèle de génération de l'EEG interprétables physiologiquement, qui se positionne à un niveau (comportement collectifs de population neuronales) raisonnable par rapport aux observations fournies par la SEEG (les capteurs intra-cérébraux utilisés enregistrent l'activité globale de larges populations de neurones et non celle de neurones isolés). Ce type de modèle utilisé par Jansen et al. [5] pour étudier la génération de potentiels évoqués dans le cortex visuel. Dans sa version simple, il modélise l'activité non linéaire d'une macro-colonne corticale: des neurones excitateurs reçoivent un feedback de neurones eux-mêmes excitateurs ou bien inhibiteurs, des effets de saturation (fonction sigmoïde) étant introduits dans les boucles de rétroaction. Dans sa version enrichie, il modélise l'activité de deux macro-colonnes dont le couplage peut être ajusté. C'est cette seconde version qui est utilisée dans le travail présenté ici, adaptée à la génération de certaines activités « épileptiques » et non à celle de potentiels évoqués. Deux paires de signaux ont été générées et étudiées statistiquement avec les différentes méthodes présentées pour une valeur du couplage inter-colonne (paramètre du modèle) donnée. Les signaux générés reproduisent de manière très satisfaisante, qualitativement, des activités épileptogènes observées en SEEG chez des patients. Leur caractère non linéaire (non

gaussien) est patent (apparition des pointes) et se démarque très nettement du modèle AR gaussien.

## 6. Résultats préliminaires et perspectives

Nous ne présentons ici qu'une partie des résultats, explicitée figure 1. Nous jugeons ces résultats encourageants quant à la possibilité de construire des intervalles de confiance, pour les statistiques introduites, sur signaux réels.

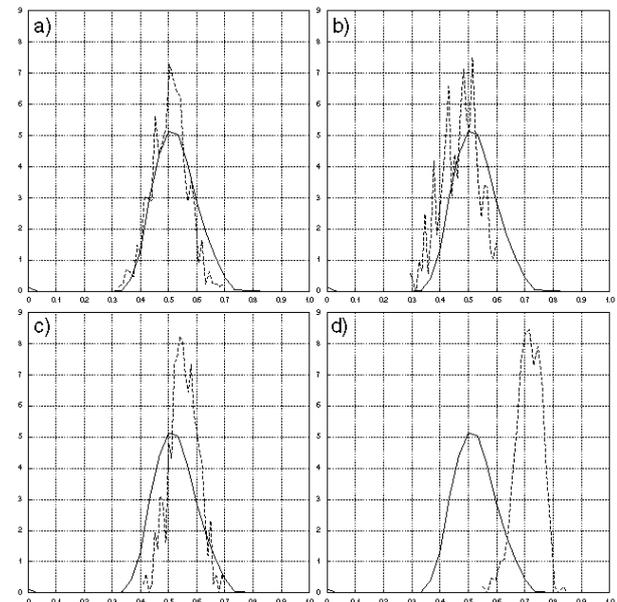


Fig.1 : Procédure de bootstrap AR pour l'estimation de  $T_{h_m^2}$ . Les distributions « vraie » (trait continu) et

« bootstrap » (pointillé) présentent des dispersions du même ordre de grandeur (a à d) pour quatre paires de signaux ayant servi de données initiales dans la procédure de bootstrap. La distribution « vraie » a été obtenue sur deux signaux générés par le modèle non linéaire sur une durée  $T_e = 1500$  s ( $\theta e = 0.005$  s). La distribution de l'estimateur de  $T_{h_m^2}$  a été évaluée sur 3000 intervalles  $\Delta_n$  d'une durée de 6.5 s (1300 échantillons) décalés de 0.5 s. Pour la procédure de bootstrap,  $N_R = 300$ .

## Références

- [1] Kinié A., *Détection de relations de causalité et estimation de temps de retard en épilepsie*, Thèse Université de Rennes 1, 1996.
- [2] Wendling F., LeBouquin-Jeannès R., Bellanger J.J., Faucon G., Chauvel P., *Extraction de signatures épileptiques et mesure de cohérence dans des observations SEEG : étude préliminaire*, Innov. et Technol en Biologie et en Médecine, 19(2):59-75, 1998.
- [3] Gersh W., *Causality or driving in electrophysiological signal analysis*, Mathematical Biosciences, 177-196, 1972.
- [4] Politis D.N., *Computer-intensive methods in Statistical Analysis*, IEEE Signal Processing Magazine, 39-55, January, 1998.
- [5] Jansen B.H., Rit V.G., *Electroencephalogram and visual evoked potential generation in a mathematical model of cortical columns*, Biol. Cybern., 73:357-366, 1995.