

# Sur différentes mesures de dépendances causales entre signaux aléatoires

Pierre-Olivier AMBLARD<sup>1</sup>, Olivier MICHEL<sup>1</sup>

<sup>1</sup>GIPSA-lab, Dépt. Images et Signal (UMR CNRS 5083)

ENSE3-BP 46 38402 Saint Martin d'Hères Cedex

bidou.amblard@gipsa-lab.inpg.fr, olivier.michel@gipsa-lab.inpg.fr

**Résumé** – Nous nous intéressons à des mesures de dépendances causales entre des séries temporelles. Nous montrons qu'une approche fondée sur la prédiction et largement développée par des économétriciens peut être lue dans le cadre de la théorie de l'information dirigée. L'utilisation pratique de ces approches est illustrée par un exemple simple qui met en lumière les difficultés d'estimation des grandeurs d'information, et qui met en avant la relativité des deux analyses à l'observation faite.

**Abstract** – We study two different class of methods to infer causal dependencies between time series. We show that the well-known Granger causality as developed by Geweke and others and based on prediction is included in the theory of directed information. The practical use of these methods is illustrated by a simple example. It reveals the difficulty to correctly estimate and use the directed information. Furthermore, it illustrates the fundamental problem of relativity of the analysis with respect to the observed times series.

## 1 Motivations

Le cadre dans lequel nous nous plaçons ici est celui des réseaux de systèmes en interaction. Une des difficultés importantes rencontrées dans ces études est relative à la nécessité de définir la notion de flot d'information. Un objectif important consiste à comprendre comment l'information circule entre les différents systèmes. Nous supposons disposer pour chaque nœud du réseau des mesures temporelles issues du système porté par le nœud. L'enjeu de ces réflexions est d'identifier à partir de l'ensemble de ces mesures le sens de l'interaction entre deux systèmes, chacun porté par un nœud. Plus particulièrement, nous voulons révéler l'existence de relations dirigées de type maître-esclave entre les systèmes étudiés. Le cadre de ce travail est donc celui de la modélisation de signaux multidimensionnels dont la structure de dépendance est régie par un graphe directionnel. L'objectif est l'estimation de ce graphe à partir des signaux observés.

Ces questions ont été abordées dans diverses disciplines, sous divers angles et dans diverses applications. Elles relèvent par exemple de la notion de causalité développée par Granger, Pearl et d'autres [1, 2], mais également de la théorie de l'information des réseaux [3, 4] ou encore de la théorie des systèmes dynamiques et de leurs couplages [5, 6]. Dans cette étude, nous considérons deux grandes classes de méthodes qui permettent de proposer des réponses aux interrogations posées. Le but du travail est de dégager les points communs entre ces approches, de tenter d'en expliciter les avantages et inconvénients, et enfin de discuter les problèmes d'estimation des grandeurs caractéristiques de ces approches.

– *Approches fondées sur la prédiction (du type « Wiener-*

*Granger* »). Ces méthodes reposent sur une assertion de Wiener selon laquelle un signal  $x_t$  est la cause d'un autre  $y_t$  si la prédiction de  $y_t$  est améliorée par la prise en compte de  $x_t$ . Cette remarque a été largement développée, étendue et discutée par Granger. Ce dernier a mis en évidence ses limitations et a développé des approches opérationnelles pour tester la causalité selon ce principe. Des extensions ont été proposées ensuite par Geweke, puis par Rissanen&Wax [9, 10].

– *Approches fondées sur la théorie de l'information.* Les liens de dépendances causales entre séries temporelles ne peuvent être quantifiés par les grandeurs usuelles de la théorie de l'information, telle l'information mutuelle. Cette dernière mesure, symétrique entre les signaux ou systèmes considérés, ne répond qu'à la question pertinente de dépendance statistique entre deux signaux. La prise en compte d'un sens de circulation de l'information impose d'avoir recours à la théorie de l'information bidirectionnelle, introduite par Marko [7] dans les années 70 et étendue par Massey [4] puis Kramers [8] dans les années 90. Cette théorie repose sur une analyse fine d'un transfert d'information et de la prise en compte de l'influence de la présence possible de « feedback » entre deux séries temporelles. Il en résulte la définition de la notion peu usitée actuellement d'information dirigée.

Ces différentes approches (moins pour la deuxième qui est nouvelle dans ce contexte) ont été largement étudiées dans diverses littératures, appliquées avec plus ou moins de bonheur dans des disciplines aussi variées que neurosciences, économie ou physique. Dans le paragraphe 2, des liens théoriques entre ces approches sont mis en évidence. Le paragraphe 3 est dédié à une discussion sur la nature et l'utilisation de ces méthodes

dans un contexte pratique. Dans la suite, nous nous placerons sous des hypothèses simples de signaux aléatoires conjointement gaussiens.

## 2 Liens théoriques

**Approches du type « Wiener-Granger »**. Une modélisation paramétrique est adoptée pour obtenir une approche opérationnelle :

$$\begin{cases} x_t = \sum_{s=1}^{\infty} A_{i,s}x_{t-s} + \sum_{s=0}^{\infty} B_{i,s}y_{t-s} + u_{i,t} \\ y_t = \sum_{s=0}^{\infty} C_{i,s}x_{t-s} + \sum_{s=1}^{\infty} D_{i,s}y_{t-s} + v_{i,t} \end{cases} \quad (1)$$

où  $i$  paramètre le type de liens entre les deux séries  $x_t$  et  $y_t$ . Par convention  $i = 1$  désigne le cas dans lequel les séries sont non couplées, de sorte que  $B_{1,s} = 0, C_{1,s} = 0, \forall s$  et les résidus  $u_{1,t}$  et  $v_{1,t}$  sont blancs et indépendants. Pour  $i = 2$ , les séries sont couplées uniquement dynamiquement au sens où  $B_{1,0} = 0, C_{1,0} = 0$ . Les résidus  $u_{2,t}$  et  $v_{2,t}$  sont blancs (principe d'orthogonalité statistique pour l'estimation en moyenne quadratique). Enfin dans le cas  $i = 3$ , les séries sont entièrement couplées,  $B_{3,s} \neq 0, C_{3,s} \neq 0, \forall s$ , les résidus  $u_{3,t}$  et  $v_{3,t}$  sont blancs, mais ne sont éventuellement plus indépendants l'un de l'autre. On se place dans l'hypothèse restrictive selon laquelle les signaux sont conjointement gaussiens. Par conséquent, les seules statistiques d'intérêt sont les statistiques d'ordre 2. On notera  $\Gamma_{i,XY}$  les covariances ou matrices de covariance entre les grandeurs (ou vecteurs)  $X$  et  $Y$  sous le modèle  $i$ .

Notons que dans le cas  $i = 2$ , les résidus  $u_{2,t}$  et  $v_{2,t}$  peuvent être corrélés *instantanément*. Dans ce cas, le modèle 2 peut être transformé en un modèle de type 3 en appliquant la transformation

$$\begin{pmatrix} x'_t \\ y'_t \end{pmatrix} = \begin{pmatrix} I & -\Gamma_{2,uv}\Gamma_{2,vv}^{-1} \\ -\Gamma_{2,vu}\Gamma_{2,uu}^{-1} & I \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix}$$

$I$  étant l'identité. Cette transformation donne en effet des coefficients  $B'_{2,0}$  et  $C'_{2,0}$  non nuls, et des résidus  $u'_{2,t}$  et  $v'_{2,t}$  décorrelés (propriétés de l'estimation linéaire en moyenne quadratique) et donc indépendants sous l'hypothèse gaussienne adoptée.

Les mesures proposées par Geweke [9] et étendues par Rissanen&Wax [10] utilisent les covariances des résidus  $u_{i,t}$  et  $v_{i,t}$  dans (1). Notons que  $\Gamma_{i,uu}$  représente la variance de l'erreur d'estimation de  $x_t$  à partir de son passé si  $i = 1$ , et du passé de  $y_t$  si  $i = 2$ , et de la valeur instantanée  $y_t$  si  $i = 3$ . Si  $x_{1:t}$  représente la collection des valeurs du signal  $x$  de l'instant initial 1 à l'instant  $t$ ,  $\varepsilon_{\infty}^2(x_t|x_{1:t-1}, y_{1:t}) = \lim_{t \rightarrow +\infty} \varepsilon^2(x_t|x_{1:t-1}, y_{1:t})$  est la variance asymptotique de l'erreur d'estimation de  $x_t$  à partir de  $x_{1:t-1}, y_{1:t}$ . Selon la méthodologie introduite par Geweke, nous considérons les indices suivants

$$F_{y \rightarrow x} = \log \frac{\varepsilon_{\infty}(x_t|x_{1:t-1})}{\varepsilon_{\infty}(x_t|x_{1:t-1}, y_{1:t-1})}$$

$$F_{x \rightarrow y} = \log \frac{\varepsilon_{\infty}(y_t|y_{1:t-1})}{\varepsilon_{\infty}(y_t|x_{1:t-1}, y_{1:t-1})}$$

$$F_{x,y} = \log \frac{\varepsilon_{\infty}(x_t|x_{1:t-1}, y_{1:t-1})}{\varepsilon_{\infty}(x_t|x_{1:t-1}, y_{1:t})}$$

$F_{x,y}$  est symétrique en  $x$  et  $y$  alors que les deux autres ne le sont pas et impliquent une direction de couplage.  $F_{x \rightarrow y}$  est appelée par Geweke mesure de « feedback » linéaire de  $x$  vers  $y$  et  $F_{x,y}$  est dénommée mesure de dépendance causale instantanée. Dans [10], Rissanen&Wax généralisent cette approche en utilisant la théorie du codage prédictif. La mesure envisagée n'est plus une variance d'erreur de prédiction, mais une quantité d'information (exprimée en bits) nécessaire à la prédiction. Dans le cas gaussien et sous l'hypothèse d'un ordre de modèle connu, cette approche permet de retrouver les indices de Geweke-Granger. Notons que la quantité d'information relative au choix de l'ordre du modèle fait partie intégrante du calcul de la mesure de Rissanen&Wax. En pratique en effet, l'approche de Geweke-Granger repose sur l'identification de modèles dont les ordres sont supposés connus *a priori*, ou préalablement estimés.

**Information dirigée.** Dans [4], Massey propose qu'en présence de feedback, la mesure correcte d'information entre deux signaux n'est plus l'information mutuelle, mais l'information dirigée, définie par

$$I(x_{1:t} \rightarrow y_{1:t}) = \sum_{k=1}^t I(x_{1:k}; y_k | y_{1:k-1})$$

où  $I(x_{1:k}; y_k | y_{1:k-1})$  est l'information mutuelle conditionnelle usuelle [3]. Lorsque les signaux sont stationnaires, l'extensivité de l'information dirigée permet de définir le taux d'information dirigée  $I_{\infty}(x \rightarrow y) = \lim_{t \rightarrow +\infty} (1/t)I(x_{1:t} \rightarrow y_{1:t})$ . D'autre part dans [8], Kramers établit le résultat  $I_{\infty}(x \rightarrow y) = \lim_{k \rightarrow +\infty} I(x_{1:t}; y_t | y_{1:t-1})$ , valide sous hypothèse de stationnarité.

Dans le cas de processus conjointement gaussiens, le lien entre information dirigée et approche de Geweke-Granger est obtenu *via* la proposition suivante :

$$I_{\infty}(x \rightarrow y) = F_{x \rightarrow y} + F_{x,y}$$

En effet, la règle de chainage pour les informations mutuelles conditionnelles permet de décomposer  $I(x_{1:t}; y_t | y_{1:t-1})$  en  $I(x_{1:t-1}; y_t | y_{1:t-1}) + I(x_t; y_t | x_{1:t-1}, y_{1:t-1})$ , [3]. D'autre part, dans le cas de trois vecteurs aléatoires  $x, y, z$ , l'information mutuelle conditionnelle s'écrit  $I(x; y | z) = I(x; y, z) - I(x; y)$ . Si  $\varepsilon^2(x|y)$  représente la matrice de covariance conditionnelle, alors dans le cas de vecteurs conjointement gaussiens  $I(x; y) = 1/2 \log |\varepsilon^2(x)| / |\varepsilon^2(x|y)|$ , où  $|\cdot|$  dénote le déterminant. Ainsi,  $I(x; y | z) = 1/2 \log |\varepsilon^2(x|y)| / |\varepsilon^2(x|y, z)|$ . Dans le cas qui nous intéresse, nous obtenons

$$I(x_{1:t-1}; y_t | y_{1:t-1}) = \frac{1}{2} \log \frac{\varepsilon_{\infty}^2(y_t | y_{1:t-1})}{\varepsilon_{\infty}^2(y_t | x_{1:t-1}, y_{1:t-1})}$$

$$I(x_t; y_t | x_{1:t-1}, y_{1:t-1}) = \frac{1}{2} \log \frac{\varepsilon_{\infty}^2(y_t | x_{1:t-1}, y_{1:t-1})}{\varepsilon_{\infty}^2(y_t | x_{1:t}, y_{1:t-1})}$$

Donc, en prenant la limite de  $I(x_{1:t}; y_t | y_{1:t-1})$  quand  $t$  tend vers  $\infty$ , le résultat  $I_{\infty}(x \rightarrow y) = F_{x \rightarrow y} + F_{x,y}$  est obtenu. ■

Ainsi sous l'hypothèse que les signaux en jeu sont conjointement gaussiens, les deux approches se rejoignent asymptotiquement. Toutefois, leur nature et leur utilisation pratique diffèrent.

### 3 Sur l'utilisation en pratique

Bien que les deux approches se rejoignent théoriquement, elles sont de nature très différentes et nécessitent des méthodes d'implantation différentes. Dans le cadre des méthodes fondées sur la prédiction et sous des hypothèses gaussiennes, les rapports de variances suivent des lois connues [11]. Les mesures permettent à l'aide de tests d'hypothèse de trancher entre l'existence ou l'absence de liens de dépendance causale, au sens de la prédiction. Rappelons que dans ce contexte les apports de l'approche de Rissanen&Wax dans [10] qui prend en compte dans la construction de la mesure le problème de l'estimation des ordres des modèles.

Par nature, l'approche fondée sur la théorie de l'information bidirectionnelle répond mieux à nos problématiques : suivre les flots d'information dans les réseaux de systèmes en interaction. L'information dirigée ne répond pas directement à la question de dépendance causale entre deux séries temporelles mais quantifie l'information circulant entre les deux en tenant en compte de la direction. Cette mesure amène donc des réponses sur la nature maître-esclave des liens entre systèmes.

L'inconvénient de l'information dirigée réside dans son estimation et quand cet aspect pratique est abordé, les principes fondés sur la prédiction se révèlent facilement opérationnels. En effet, sous les hypothèses de résidus gaussiens, les tests fondés sur la prédiction peuvent être qualifiés complètement, au moins asymptotiquement. Lorsque l'on considère l'information dirigée, les problèmes de grandes dimensions, de corrélation temporelle des séries et de finitude du nombre des mesures conduisent à de grandes difficultés d'estimation. Pour illustrer ces différents points nous examinons maintenant un problème jouet simple dont nous donnons trois lectures.

**Première lecture.** Considérons deux signaux  $x_t, y_t$  dont une portion est représentée sur la figure (1). Les fonctions de corrélation  $E[x_t y_{t-\tau}]$  et d'information mutuelle  $I(x_t; y_{t-\tau})$  estimées sont également tracées sur cette figure. Ces graphes montrent à l'évidence que  $y$  est corrélé au passé de  $x$  (pics en  $\tau = -2, -3$ ), mais également anticorrélé à son futur (pics en  $\tau = 2, 3$ ). Ces évidences permettent de conclure à un probable « feedback » entre ces deux séries temporelles.

Nous appliquons la méthodologie présentée précédemment. Sous des hypothèses de signaux conjointement gaussiens, les distributions des indices de Geweke sont connues [9], et peuvent être utilisées pour définir une stratégie de test optimal. Toutefois, l'approche de Geweke et l'approche fondée sur l'information dirigée peuvent être utilisées sous des hypothèses différentes. Si l'hypothèse gaussienne est absente, les distributions des indices de Geweke et de l'information dirigée estimée sont inconnues. La détermination d'un seuil de décision ne peut donc être traitée analytiquement. Nous proposons d'avoir recours à une randomisation des données pour simuler l'hypothèse nulle selon laquelle il n'y a pas de lien entre les signaux  $x_t$  et  $y_t$ . Pratiquement, les signaux étant de taille finie, nous générons  $n_p$  permutations aléatoires  $\sigma(t)$  des indices temporels pour créer  $n_p$  signaux  $y_{\sigma(t)}$ . L'effet de la permutation

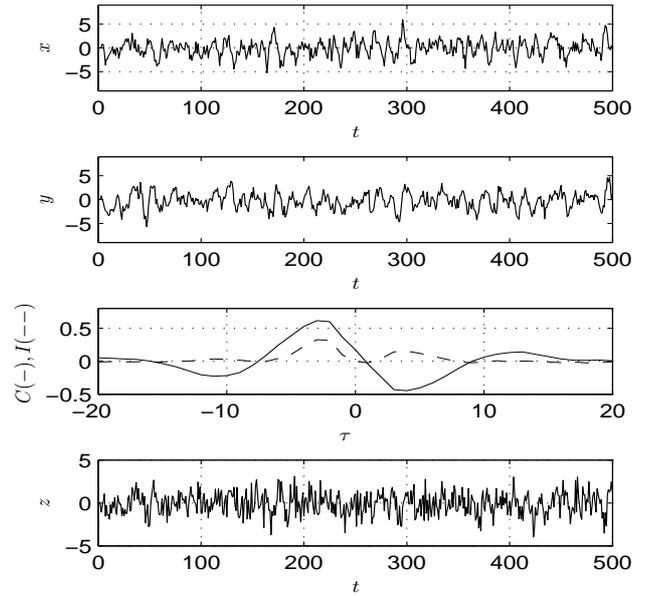


FIG. 1 – Signaux  $x_t$  et  $y_t$  observés et dont on cherche des liens de dépendance. Fonctions de corrélation et d'information mutuelle entre ces signaux. Signal  $z_t$  supplémentaire.

aléatoire des dates est de briser l'éventuelle dépendance entre  $x_t$  et  $y_t$ . Nous calculons les indices de dépendances pour les  $n_p$  signaux, évaluons l'histogramme des indices, et fixons le seuil de test de sorte à ce que le taux de fausses alarmes soit fixé (nous choisissons ici 5%).

Précisons ici que les difficultés d'estimation ne sont pas les mêmes entre l'approche de Geweke et l'utilisation de l'information dirigée. Ici, les indices de Geweke sont directement obtenus à partir de la puissance des erreurs de prédiction estimées classiquement. Le premier point concerne l'horizon de calcul de l'information dirigée. Ce point délicat existe aussi pour les indices de Geweke mais est moins crucial et peut être levé avec l'approche de Rissanen&Wax. Dans le cas de l'information dirigée, les problèmes de dimension sont par contre primordiaux. Un horizon fini est utilisé et la mesure s'écrit

$$I(x \rightarrow y) = \frac{1}{p} \sum_{k=1}^p I(x_{t-k:t}; y_t | y_{t-k:t-1})$$

Pour estimer l'information mutuelle conditionnelle, nous utilisons les techniques d'estimation de l'entropie fondées sur les plus proches voisins [12, 13], étendues astucieusement à l'information mutuelle par Kraskov [14] puis à l'information mutuelle conditionnelle par Frenzel&Pompe [15]. Nous ne détaillerons pas ici ces estimateurs mais notons que leur comportement est raisonnable pour des tailles de signaux de plusieurs milliers d'échantillons, à condition de rester dans des dimensions faibles (3 à 4) et que les signaux soient faiblement corrélés temporellement. Ceci explique les valeurs proches de zéros mais négatives des seuils obtenues lors des mesures après permutations (voir le tableau 1). Les estimateurs souffrent d'un biais négatif et n'assure pas au résultat d'être positif (propriété requise pour une information mutuelle). Toutefois, nous

	$F_{x \rightarrow y}$	$F_{y \rightarrow x}$	$F_{x,y}$	$I(x \rightarrow y)$	$I(y \rightarrow x)$
seuils	1.016	1.016	1.002	-0.016	-0.013
tests	1.289	1.112	1.007	0.08	0.02

TAB. 1 – Exponentielle des indices de Geweke et Informations dirigée. Signaux de 2000 échantillons, passé de longueur 20 pour Geweke, de 3 pour les informations. 1000 permutations utilisées pour Geweke, 100 pour les informations.

	$F_{x \rightarrow y z}$	$F_{y \rightarrow x z}$	$F_{x,y z}$	$I(x \rightarrow y z)$	$I(y \rightarrow x z)$
tests	1.218	1.003	1.011	0.02	-0.025

TAB. 2 – Exponentielle des indices de Geweke et Informations dirigée conditionnellement à  $z_t$ . Signaux de 2000 échantillons, passé de longueur 20 pour Geweke, de 3 pour les informations.

avons vérifié la convergence vers zéro pour des tailles de signaux croissantes. De plus, mentionnons que les temps de calcul sont considérablement plus longs que pour des estimations en moyenne quadratique !

Le tableau (1) résume les seuils de test ainsi obtenus pour les différents indices, et les valeurs des indices pour les données à effectivement tester. Chaque indice est supérieur au seuil prédéterminé. La conclusion est donc la présence d'un feedback entre  $x_t$  et  $y_t$ , et la présence d'une causalité instantanée.

**Deuxième lecture.** Outre les signaux  $x_t$  et  $y_t$  utilisés, nous supposons maintenant avoir une autre information à travers la mesure d'un troisième signal  $z_t$  a priori lié aux deux autres. Nous reprenons la même méthodologie que précédemment à la différence près que tous les calculs sont effectués conditionnellement à  $z_{1:t}$ . Autrement dits les prédictions dans l'approche de Geweke utilisent systématiquement les données  $z_t$  conjointement aux autres ; les informations mutuelles conditionnelles en jeu dans le calcul de l'information dirigée sont également conditionnelles à  $z_{1:t}$  (ou à  $z_{t-p:t}$  dans le cas du calcul sur un horizon fini). Le tableau (2) présente les indices de mesure dépendance. Nous avons présenté les résultats pour l'information dirigée, mais ces résultats sont sujet à caution étant donné la petite taille des échantillons et les grandes dimensions considérées. De plus, les seuils ne sont pas ici calculés, car l'ajout de  $z_t$  ne permet pas une randomisation simple des données conduisant à l'hypothèse nulle. Les indices de Geweke, calculés conditionnellement à  $z_t$ , montrent toutefois que le couplage instantané entre  $x$  et  $y$  et que le feedback de  $x_t$  vers  $y_t$  demeurent, mais que le feedback de  $y_t$  vers  $x_t$  n'existe pas conditionnellement à  $z_t$ . Ceci met en évidence que le couplage de  $y$  vers  $x$  est indirect et se fait par  $z$ .

**Troisième lecture.** Pour terminer la discussion sur ce problème jouet, nous présentons le modèle utilisé :

$$\begin{cases} x_t &= 0.6x_{t-1} - 0.6z_{t-1} + w_{1,t} \\ y_t &= 0.6y_{t-1} + 0.4x_{t-2} + w_{2,t} \\ z_t &= -0.1z_{t-1} + 0.5y_{t-2} + w_{3,t} \end{cases} \quad \Gamma_w = \begin{pmatrix} 1 & 0.1 & 0.1 \\ & 1 & 0.01 \\ & & 1 \end{pmatrix}$$

$\Gamma_w$  représente la matrice de corrélation des suites i.i.d. centrées gaussiennes  $w_{i,t}$ .

**Quelques conclusions.** Nous avons montré ici que les approches d'analyse de causalité fondée sur l'analyse des erreurs de prédictions sont théoriquement incluses dans la théorie de

l'information dirigée. Toutefois, d'un point de vue pratique, la mesure des informations dirigées reste très difficile, surtout lorsque les signaux en jeu sont observés sur des durées courtes et présentent des supports de corrélation larges. Nous avons également montré à travers l'analyse d'un exemple simple un problème fondamental : l'analyse des dépendances (causales ou non) est *relative* aux observations faites. Ce point déjà discuté par Granger se révèle tout aussi crucial quand d'autres types de mesures sont utilisés, en particulier l'information dirigée. La résolution de ce problème est toutefois plus difficile d'un point de vue pratique dans le cas de l'information dirigée puisque son estimation devient très délicate lorsque la dimension des données augmente et donc lorsque le conditionnement est pris en compte.

La théorie de l'information dirigée offre un cadre théorique adéquat pour l'analyse des dépendances entre séries temporelles. Toutefois, le recours à des approches sous-optimales proches est à l'heure actuelle préférable pour des raisons purement pragmatiques.

## Références

- [1] C. W. J. GRANGER. Some recent developments in a concept of causality. *Journal of econometrics*, 39 :199–211, 1988.
- [2] J. PEARL. *Causality : models, reasoning and inference*. Cambridge University Press, 2000.
- [3] J. COVER and B. THOMAS. *Elements of information theory*. Wiley, 1993.
- [4] J.L. MASSEY. Causality, feedback and directed information. In *Proc. Intl. Symp. on Info. th. and its Applications, Waikiki, Hawaii, USA*, nov 1990.
- [5] M. LE VAN QUYEN, J. MARTINERIE, C. ADAM, and F. VARELA. Nonlinear analyses of interictal eeg map the brain interdependences in human focal epilepsy. *Physica D*, 127 :250–266, 1999.
- [6] R. QUIAN QUIROGA, J. ARNHOLD, and P. GRASSERGER. Learning driver-response relationship from synchronisation patterns. *Phys. Rev. E*, 61 :5142–5148, 2000.
- [7] H. MARKO. The bidirectional communication theory—a generalization of information theory. *IEEE Trans. on communications*, 21(12) :1345–1351, December 1973.
- [8] G. KRAMERS. *Directed information for channels with feedback*. PhD thesis, Swiss Federal Institute of Technology Zurich, 1998.
- [9] J. GEWEKE. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77 :304–313, 1982.
- [10] J. RISSANEN and M. WAX. Measures of mutual and causal dependence between two time series. *IEEE Trans. on Information Theory*, 33 :598–601, 1987.
- [11] T. W. ANDERSON. *An introduction to multivariate statistical analysis*. Wiley, 2001.
- [12] L. F. KOZACHENKO and N. N. LEONENKO. Sample estimate of the entropy of a random vector. *Problems Infor. Transmiss.*, 23 :95–101, 1987.
- [13] M. N. GORJA, N. N. LEONENKO, V. V. MERGEL, and P. L. NOVI INVARDI. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *Jour. Nonparam. Stat.*, 17(3) :277–297, 2005.
- [14] A. KRASKOV, H. STOGBAUER, and P. GRASSERGER. Estimating mutual information. *Phys. Rev. E*, 69 :066138, 2004.
- [15] S. FRENZEL and B. POMPE. Partial mutual information for coupling analysis of multivariate time series. *Phys. Rev. Lett.*, 99 :204101, 2007.