

# DÉTECTION DE L'INDÉPENDANCE LOCALE ENTRE NEURONES

Christine Tuleau-Malot <sup>1</sup> & Patricia Reynaud-Bouret <sup>1</sup> & Vincent Rivoirard <sup>2</sup> & Thomas Bessaïh <sup>3</sup> & Régis Lambert <sup>3</sup> & Nathalie Leresche <sup>3</sup> & Michaël Quiquempoix <sup>3</sup>

<sup>1</sup> *Univ. Nice Sophia Antipolis, CNRS, LJAD, UMR 7351, 06100 Nice, France, malot@unice.fr, reynaudb@unice.fr*

<sup>2</sup> *Univ. Paris Dauphine, CNRS, CEREMADE, UMR 7534, 75775 Paris, France, Vincent.Rivoirard@dauphine.fr*

<sup>3</sup> *Univ. Pierre et Marie Curie, CNRS, NPA, UMR 8246, 75252 Paris, France bessaih.thomas@snu.jussieu.fr, regis.lambert@snu.jussieu.fr, nathalie.leresche@snu.jussieu.fr, michael.quiquempoix@snu.jussieu.fr*

**Résumé.** Un des enjeux actuels des Neurosciences est d'obtenir une meilleure compréhension de la dynamique de l'activité neuronale. En effet, si depuis plusieurs années un phénomène de synchronisation de l'activité neuronale a été mis en évidence d'un point de vue pratique et théorique, il demeure la question de la connectivité fonctionnelle. Si biologiquement, les neurones sont tous connectés, cela ne signifie pas que tous interviennent dans la réponse à une stimulation. La question est donc de pouvoir identifier des assemblées de neurones, à savoir les neurones qui interagissent entre eux afin de répondre à une stimulation, avec si possible le graphe de connectivité fonctionnelle les régissant. C'est pour répondre à cette problématique que nous avons proposé une procédure statistique faisant intervenir essentiellement deux outils : la modélisation de l'activité neuronale par un processus de Hawkes et une estimation parcimonieuse de l'intensité conditionnelle associée au moyen d'un critère de type Lasso.

**Mots-clés.** Processus de Hawkes, Estimation Lasso, Graphe d'indpendance locale, assemblée de neurones.

**Abstract.** One of current issues for Neurosciences is to get a better understanding of the dynamics of neuronal activity. Indeed, if for several years a synchronization phenomenon of neuronal activity was demonstrated from a practical and theoretical point of view, it remains the question of the functional connectivity. If biologically neurons are all connected, it does not mean that all are involved in the response to a stimulation. The question is therefore to be able to identify assemblies of neurons, namely neurons who interact to respond to a stimulation, with if possible the graph of functional connectivity. This is to address this problem that we proposed a statistical procedure mainly involving two tools: modeling of the neuronal activity through a Hawkes process and a parsimonious estimation of the associated conditional intensity with a lasso criteria.

**Keywords.** Hawkes process, Lasso estimate, local independence graph, neurons assembly.

# 1 Introduction

En neurosciences, l'un des défis actuels est d'acquies une meilleure compréhension de la dynamique du cerveau. Pour ce faire, les neurobiologistes étudient les potentiels d'action (spikes) considérés comme les marqueurs principaux de l'activité neuronale. Ainsi, en enregistrant simultanément les trains de spikes de plusieurs neurones, un phénomène de synchronisation a été mis en évidence lors de tâches sensori-motrices. Parmi les méthodes les plus populaires permettant de détecter ces synchronisations entre paires de neurones, nous pouvons citer la méthode des "Unitary Events", méthode développée par Grün *et al.* (2002). De nombreuses améliorations de cette méthode ont été proposées, à l'image de MTGAUE développée par Tuleau-Malot *et al.* (2014). Cependant, ces méthodes ne répondent pas à la question : "comment identifier, dans les réseaux biologiques de neurones, les graphes d'indépendance locale reflétant la connectivité fonctionnelle?" car des études par paires de neurones ne permettent pas de capturer toute l'information pertinente d'un ensemble plus complet de neurones. Biologiquement, les neurones sont physiquement tous connectés les uns aux autres, directement ou non. Pour autant, cela ne signifie pas que tous interviennent dans la réponse à une stimulation. L'objectif est donc de déterminer le graphe de connectivité fonctionnelle, qui en raison du phénomène de synchronisation peut être vu comme un graphe d'indépendance locale. Afin d'obtenir de tels graphes, nous avons développé une procédure alliant la modélisation des trains de spikes par un processus de Hawkes multivarié et l'estimation des intensités conditionnelles associées par un critère Lasso.

## 2 Modélisation

L'activité neuronale est résumée par les trains de spikes, à savoir les temps d'apparition des différents spikes au cours du temps. Il s'agit donc d'une séquence de 0 et 1 (1 traduisant la présence d'un spike) qui se modélise mathématiquement par un processus ponctuel. Dans les études menées par Grün *et al.* (2002) ou Tuleau-Malot *et al.* (2014), ces processus ponctuels sont supposés être des processus de Poisson stationnaires. Cependant, cette hypothèse ne permet pas de tenir compte des interactions qui peuvent éventuellement exister entre les neurones. C'est la raison pour laquelle, nous proposons de modéliser les trains de spikes par un processus de Hawkes multivarié en lieu et place des processus de Poisson, à l'image des travaux réalisés par Krumin *et al.* (2010).

Un processus de Hawkes se décrit par la donnée de son intensité conditionnelle notée  $\lambda(\cdot)$ . Concrètement, étant donné le passé du processus avant le temps  $t$ ,  $\lambda(t).dt$  représente la probabilité conditionnelle qu'un nouveau point survienne au temps  $t$  pour ce processus. Dans le cadre qui nous occupe, nous nous intéressons à  $M$  trains de spikes enregistrés simultanément sur  $M$  neurones différents. Ces  $M$  trains de spikes, notés  $(N^{(1)}, \dots, N^{(M)})$ , sont modélisés par un processus de Hawkes multivarié défini à travers ses intensités con-

ditionnelles notées respectivement  $\lambda^{(1)}, \dots, \lambda^{(M)}$ , intensités qui représentent les taux de décharge instantanés. Ces intensités ont pour définition :

$$\forall m \in \{1, \dots, M\}, \quad \lambda^{(m)}(t) = \nu^{(m)} + \sum_{k=1}^M \int_{-\infty}^{t-} h_k^{(m)}(t-u) dN^{(k)}(u)$$

où :

- $\nu^{(m)}$  représente l'activité spontanée du neurone  $m$  ou l'activité engendrée par les neurones non enregistrés;
- $h_k^{(m)}$  représente l'interaction du neurone  $k$  sur le neurone  $m$ ;
- $dN^{(k)}$  représente la mesure ponctuelle du neurone  $k$ .

Cette modélisation apparaît comme réaliste au regard d'une part des travaux existants et d'autre part des tests d'adéquation proposés par Reynaud-Bouret *et al.* (2014).

### 3 Estimation Lasso

A présent que l'on dispose d'une modélisation des trains de spikes, la question est de parvenir à en déduire le graphe de connectivité fonctionnelle pour lequel chaque sommet représente un neurone. Toute la difficulté est de déterminer les arêtes orientées composant le graphe. La fonction  $h_i^{(j)}$  représente l'action du neurone  $i$  sur le neurone  $j$ . Notre idée est de placer une arête de  $i$  vers  $j$  si la fonction  $h_i^{(j)}$  n'est pas identiquement nulle. Il faut donc pour cela estimer l'ensemble des fonctions  $h_i^{(j)}$  de manière parcimonieuse.

Pour cette étape d'estimation, nous disposons de  $n$  enregistrements de  $(N^{(1)}, \dots, N^{(M)})$  notés  $\{(N_i^{(1)}, \dots, N_i^{(M)})\}_{i=1, \dots, n}$ .

Notons  $\eta = (\eta^{(1)}, \dots, \eta^{(M)})$  le vecteur des intensités conditionnelles candidates, avec :

$$\forall i \in \{1, \dots, M\}, \quad \eta^{(i)}(t) = \mu^{(i)} + \sum_{j=1}^M \int_{-\infty}^{t-} g_j^{(i)}(t-u) dN^{(j)}(u)$$

où les fonctions  $g_j^{(i)}$  sont des fonctions constantes par morceaux sur une grille fine à  $K$  éléments, à savoir  $g_j^{(i)} = \sum_{k=1}^K a_{i,j,k} \delta^{-1/2} 1_{[(k-1)\delta, k\delta]}$ , avec  $\delta > 0$  la longueur des éléments de la grille fine,  $K\delta = A$  et  $A$  la portée des fonctions d'interaction.

Les candidats  $\eta$  sont d'autant meilleurs que  $\sum_{i=1}^M -2 \int_0^T \eta^{(i)}(t) \lambda^{(i)}(t) dt + \int_0^T \eta^{(i)}(t)^2 dt$  sera petite ( $T$  étant la durée d'enregistrement de l'activité neuronale).

Cependant, puisque  $\lambda^{(i)}$  est une quantité inconnue, la minimisation du critère précédent

est impossible. Mais, de par sa définition,  $\lambda^{(i)}(t).dt$  fluctue autour de  $dN_t^{(i)}$ , ce qui nous amène à considérer le critère :

$$\gamma(\eta) = \sum_{i=1}^M -2 \int_0^T \eta^{(i)}(t) dN_t^{(i)} + \int_0^T \eta^{(i)}(t)^2 dt. \quad (1)$$

qui n'est autre que le contraste des moindres carrés associé à un processus ponctuel possédant une intensité conditionnelle.

Afin de résoudre le problème de minimisation, introduisons les notations suivantes : pour les neurones  $i$  et  $j$  et l'enregistrement  $l$ ,

$$\begin{aligned} (\mathbf{a}^{(i)})' &= (\mu^{(i)}, a_{i,1,1}, \dots, a_{i,1,K}, a_{i,2,1}, \dots, a_{i,M,K}), \\ (\mathbf{c}_t^{(i,l)})' &= (N_t^{(i)}([t - \delta, t]), \dots, N_t^{(i)}([t - K\delta, t - (K - 1)\delta])), \\ (\mathbf{Rc}_t^{(l)})' &= (1, \delta^{-1/2}(\mathbf{c}_t^{(1,l)})', \dots, \delta^{-1/2}(\mathbf{c}_t^{(M,l)})'), \\ \mathbf{n}_{ij}^{(l)} &= \left( \int_0^T N_t^{(j)}([t - k\delta, t - (k - 1)\delta]) dN_t^{(i)}(t) \right)_{k=1, \dots, K}, \\ (\mathbf{b}^{(i,l)})' &= (N_t^{(i)}([0, T]), \delta^{-1/2} \mathbf{n}'_{i,1}, \dots, \delta^{-1/2} \mathbf{n}'_{i,M}), \\ \mathbf{G}^{(l)} &= \int_0^T \mathbf{Rc}_t^{(l)} (\mathbf{Rc}_t^{(l)})' dt. \end{aligned}$$

La solution de la minimisation du contraste des moindres carrés (1) est donnée par :

$$\hat{a}^{(j)} = \left( \sum_{l=1}^n \mathbf{G}^{(l)} \right)^{-1} \left( \sum_{l=1}^n \mathbf{b}^{(j,l)} \right).$$

Cette solution ne permet pas d'obtenir de la parcimonie au niveau de l'estimation des fonctions d'interaction, ce qui a tendance à privilégier un graphe de connectivité complet. Afin, d'obtenir une estimation parcimonieuse, on considère un contraste pénalisé de type Lasso en lieu et place du contraste des moindres carrés.

Si l'on se réfère aux travaux de Hansen *et al.* (2012), on montre que la solution est alors définie par :

$$\tilde{a}^{(j)} \in \underset{\mathbf{a}^{(j)}}{\operatorname{argmin}} \left( -2(\mathbf{a}^{(j)})' \left( \sum_{l=1}^n \mathbf{b}^{(j,l)} \right) + (\mathbf{a}^{(j)})' \left( \sum_{l=1}^n \mathbf{G}^{(l)} \right) \mathbf{a}^{(j)} + 2(\mathbf{d}^{(j)})' |\mathbf{a}^{(j)}| \right)$$

où  $d^{(j)}$  est un vecteur de poids spécifique et défini en fonction des données.

## 4 Applications sur données simulées

Afin de vérifier la pertinence de la procédure proposée, nous l'avons éprouvée sur de nombreux réseaux de neurones obtenus par simulations. Nous avons ainsi considéré dans un

premier temps des réseaux de neurones simulés à l'aide de processus de Hawkes multivariés. Puis afin de ressembler davantage à la réalité biologique, nous avons travaillé sur des réseaux issus de modèles integrate and fire. Ainsi, sur le réseau présenté en Figure (1),

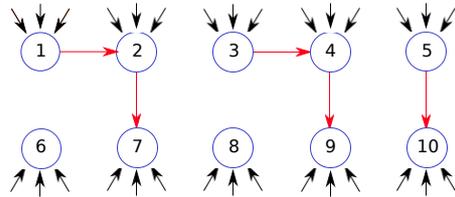


Figure 1: réseau neuronal composé de 10 neurones obtenu à partir d'un modèle integrate and fire et présentant un effet de chaîne

avec un choix convenable des paramètres de la procédure, nous retrouvons parfaitement le réseau alors que l'utilisation de cross-corrélogrammes fait apparaître des arêtes en trop. Pour ce qui est du réseau présenté en Figure (2), la reconstruction du réseau n'est pas

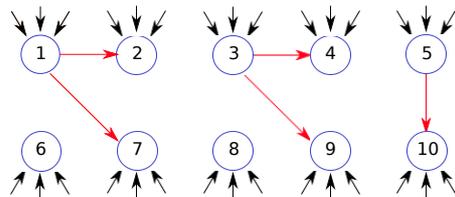


Figure 2: réseau neuronal composé de 10 neurones obtenu à partir d'un modèle integrate and fire avec une entrée commune

parfaite. Cependant, on a remarqué que les arêtes supplémentaires sont associées à des fonctions d'interaction  $h$  avec une énergie ( $\int h^2$ ) faible, comme le montre la Figure (3).

## 5 Perspectives

Si la procédure proposée semble augurer de bons résultats tant sur des données simulées que sur des données réelles, il subsiste un certain nombre défis à relever :

- la modélisation par un processus de Hawkes multivarié tient compte des interactions entre les neurones. Cependant la modélisation proposée ne prend pas en considération la non stationarité de l'activité neuronale. Un travail est actuellement en cours avec P. Reyanaud-Bouret et F. Picard afin d'améliorer cela;

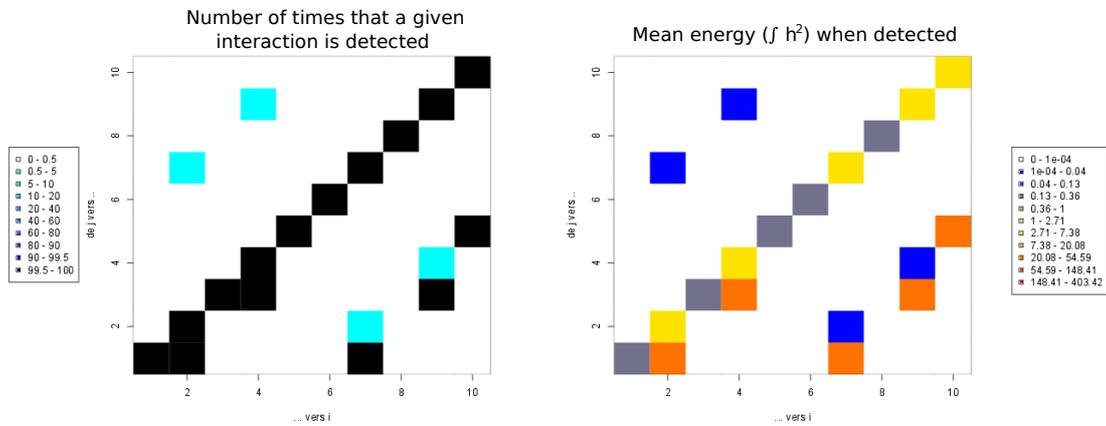


Figure 3: résultats de la procédure sur le réseau de la Figure (2) : en ordonnée apparaît le numéro du neurone qui agit alors qu’en abscisse figure le numéro du neurone sur lequel on agit.

- il semble que de fausses arêtes puissent apparaître dans les graphes estimés du fait de la détection de faux positifs. Une idée pourrait être, afin d’améliorer la procédure, de procéder à des étapes de simulation et estimation supplémentaires;
- dans les études, seul un petit nombre de neurones sont enregistrés alors qu’il y a en réalité un très grand nombre de neurones. Comment intégrer le fait que l’information dont on dispose est partielle?

## Bibliographie

- [1] Grün, S., Diesman, M. et Aertsen, A.M. (2002), Unitary events in multiple single-neuron spiking activity: I. Detection and Significance, *Neural Computation*, 14, 43–80.
- [2] Tuleau-Malot, C., Rouis, A., Grammont, F. et Reynaud-Bouret, P. (2014), Multiple tests based on a Gaussian approximation of the Unitary Events method with delayed coincidence count, *Neural Computation*, 26 (7), 1408–1454.
- [3] Krumin, M., Reutsky, I. et Shoham, S. (2010), Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input, *Frontiers in Computational Neuroscience*, 4(article 147).
- [4] Reynaud-Bouret, P., Rivoirard, V., Grammont, F. et Tuleau-Malot, C. (2014), Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis, *Journal of Mathematical Neuroscience*, 4:3.
- [5] Hansen, N.R., Reynaud-Bouret, P. et Rivoirard, V. (à paraître), Lasso and probabilistic inequalities for multivariate point processes.