

# DriPP : Processus ponctuels pilotés pour la modélisation des signaux M/EEG modulés par stimulation

Cédric ALLAIN, Alexandre GRAMFORT, Thomas MOREAU

Parietal, INRIA Saclay, Université Paris-Saclay  
1 rue Honoré Estienne d’Orves, 91120 Saclay, France  
cedric.allain@inria.fr, alexandre.gramfort@inria.fr  
thomas.moreau@inria.fr

**Résumé** – L’analyse quantitative des signaux électrophysiologiques non invasifs est rendue possible par l’identification de motifs récurrents, tant dans le temps que dans l’espace cérébral. Une question naturelle est d’arriver à estimer dans quelle mesure les occurrences de ces derniers sont modulées par certaines tâches cognitives et manipulations expérimentales. Pour y répondre, nous proposons une approche par processus ponctuels où la fonction d’intensité est liée à un processus ponctuel déterministe correspondant aux événements de stimulation. Nous proposons une méthode rapide basée sur l’algorithme espérance-maximisation (EM) afin d’estimer les paramètres de ce modèle.

**Abstract** – The quantitative analysis of non-invasive electrophysiology signals is made possible by the identification of recurring patterns, both in time and in brain space. A natural question is to estimate how their occurrences are modulated by certain cognitive tasks and experimental manipulations. To address it, we propose an approach based on point processes where the intensity function is linked to a deterministic point process corresponding to stimulation events. We derive a fast expectation-maximization based algorithm to estimate the parameters of this model.

## 1 Introduction

L’analyse statistique des enregistrements neuronaux humains est au cœur de la recherche moderne en neurosciences. Les méthodes d’enregistrement non invasives telles que la magnéto- et l’électroencéphalographie (M/EEG) permettent de capturer des signaux neuronaux à l’échelle de la milliseconde lorsque des sujets humains effectuent différentes tâches cognitives. L’apprentissage par dictionnaire convolutif (*Convolutional dictionary learning*, CDL) [1] décompose efficacement et de manière non supervisée les signaux M/EEG en combinaisons de motifs invariants dans le temps, appelés *atomes*, avec leurs déclenchements respectifs, appelés *activations*. Cela conduit à une description des données basée sur des événements discrets et aléatoires, rendant l’utilisation des processus ponctuels (PP) pertinente sur les signaux M/EEG.

Dans cette étude, on souhaite évaluer comment les stimuli externes affectent les activations des atomes, afin de déterminer si un atome est déclenché ou non par un type spécifique de stimulus et, le cas échéant, de quantifier dans quelle mesure et avec quelle latence. Les modèles de PP existants n’abordent pas directement cette question, à savoir la caractérisation de l’influence d’un PP déterministe (les déclenchements des stimuli) sur un PP stochastique (les activations des atomes), avec notamment la modélisation de la latence entre les deux, là où les modèles classiques de PP s’intéressent principalement à des effets d’excitation immédiats. Nous proposons donc une nouvelle méthode de PP – appelée DriPP (pour *Driven temporal Point*

*Processes*) –, et nous dérivons un algorithme efficace basé sur l’espérance-maximisation (EM) pour estimer les paramètres du modèle. La méthode est évaluée sur des simulations ainsi que sur trois jeux de données MEG. Ce travail est une courte version de [2].

## 2 Driven temporal Point Processes

Un processus ponctuel temporel est un processus stochastique dont la réalisation consiste en des événements discrets  $\{t_i\}$  se produisant en temps continu,  $t_i \in \mathbb{R}^+$  [3]. Dans le cas où la probabilité qu’un événement se produise au moment  $t$  dépend uniquement des événements passés  $\mathcal{F}_t := \{t_i, t_i < t\}$ , les PP sont généralement caractérisés par la fonction d’intensité conditionnelle  $\lambda : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  :

$$\lambda(t|\mathcal{F}_t) := \lim_{dt \rightarrow 0} \frac{\mathbb{P}(N_{t+dt} - N_t = 1|\mathcal{F}_t)}{dt}, \quad (1)$$

où  $N_t := \sum_{i \geq 1} \mathbb{1}_{\{t_i \leq t\}}$  désigne le processus de comptage associé au PP. Cette fonction correspond, en espérance, au taux infinitésimal auquel les événements se produisent au temps  $t$  étant donné les temps d’arrivée des événements passés [3].

Le modèle proposé DriPP est adapté du processus de Hawkes (HP; [4]), car l’occurrence d’un événement passé sur le processus pilote (*driver*) augmente la probabilité d’occurrence d’événements d’activation dans un futur proche. Cependant, nous supposons ici que le processus ponctuel stochastique de notre modèle d’activations neuronales n’a pas le comportement auto-

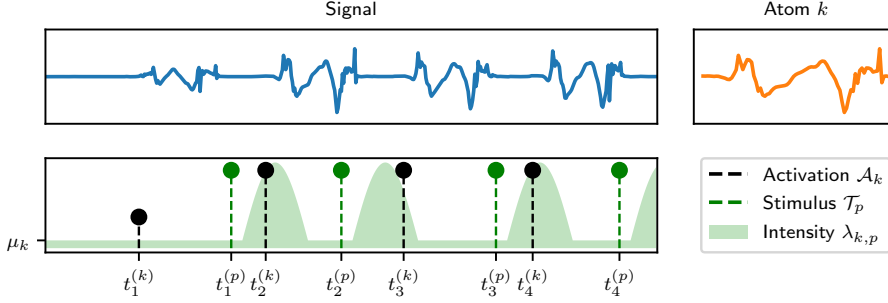


FIGURE 1 – **Haut** : *Convolutional dictionary learning* (CDL) appliqué à un signal univarié (bleu) le décompose comme la convolution d'un motif temporel (orange) et d'un signal d'activation *sparse* (noir). **Bas** : Fonction d'intensité  $\lambda_{k,p}$  définie par sa baseline  $\mu_k$  et les événements du stimulus  $\mathcal{T}_p$  (vert). L'intensité augmente après événements de stimulation avec une certaine latence.

excité caractéristique d'un HP. Au lieu de cela, les sources d'activation dans DriPP sont soit les *drivers*, soit une activité basale spontanée, mais pas ses propres activations précédentes. Plus précisément, dans DriPP, l'intensité au temps  $t$  entre un processus stochastique  $k$  (dont l'ensemble des événements est noté  $\mathcal{A}_k$ ) et un ensemble non vide de *drivers*  $\mathcal{P}$  (dont les événements sont notés  $\mathcal{T}_p := \{t_1^{(p)}, \dots, t_{n_p}^{(p)}\}, p \in \mathcal{P}$ ) est composé d'une intensité de base (*baseline intensity*)  $\mu_k \geq 0$  et de noyaux (*triggering kernels*)  $\kappa_{k,p} : \mathbb{R}^+ \rightarrow \mathbb{R}$  :

$$\lambda_{k,\mathcal{P}}(t) = \mu_k + \sum_{p \in \mathcal{P}} \sum_{i, t_i^{(p)} \leq t} \alpha_{k,p} \kappa_{k,p}(t - t_i^{(p)}) \quad , \quad (2)$$

où  $\alpha_{k,p} \geq 0$  est un coefficient qui contrôle l'importance relative du *driver*  $p$  sur l'occurrence des événements du processus stochastique  $k$ . Notons que lorsque les processus *drivers* sont connus, la fonction d'intensité est déterministe et correspond donc à l'intensité d'un processus de Poisson inhomogène [3]. Le coefficient  $\alpha_{k,p}$  est fixé à une valeur non négative afin de ne modéliser que les effets excitateurs, car les événements sur le *driver* ne font qu'augmenter la probabilité d'occurrence de nouveaux événements sur le processus stochastique. Les effets d'inhibition sont ici supposés inexistantes. La **figure 1** illustre comment les événements  $\mathcal{T}_p$  sur le *driver* influencent la fonction d'intensité après une certaine latence.

Une paramétrisation importante de ce modèle est le choix des noyaux  $\kappa_{k,p}$ . Pour modéliser au mieux la latence, nous utilisons une distribution normale tronquée de moyenne  $m_{k,p} \in \mathbb{R}$  et d'un écart-type  $\sigma_{k,p} > 0$ , avec support  $[a; b] \subset \mathbb{R}^+, b > a$  :

$$\kappa_{k,p}(x) := \frac{1}{\sigma_{k,p}} \frac{\phi\left(\frac{x - m_{k,p}}{\sigma_{k,p}}\right)}{\Phi\left(\frac{b - m_{k,p}}{\sigma_{k,p}}\right) - \Phi\left(\frac{a - m_{k,p}}{\sigma_{k,p}}\right)} \mathbb{1}_{\{a \leq x \leq b\}} \quad , \quad (3)$$

où  $\phi$  (resp.  $\Phi$ ) est la fonction de densité de probabilité (resp. la fonction de répartition) de la distribution normale standard. Cette paramétrisation diffère du noyau exponentiel qui capture les réponses immédiates tel que communément utilisé en HP. Les valeurs de troncature  $a, b \in \mathbb{R}^+$  sont supposées indépendantes du processus stochastique et des *drivers*, elles sont donc identiques pour tous les noyaux  $p \in \mathcal{P}$ . En effet, dans le contexte de cet article, ces valeurs délimitent l'intervalle de temps pendant lequel une réponse neuronale pourrait se produire suite à un stimulus externe. En d'autres termes, l'intervalle  $[a; b]$  dénote l'ensemble des valeurs de latence possibles.

Dans ce qui suit,  $T := T^{(k)}$  désigne la durée du processus  $k$ .

### 3 Inférence par algorithme EM

Nous proposons d'inférer les paramètres du modèle  $\Theta_{k,\mathcal{P}} = (\mu_k, \alpha_{k,\mathcal{P}}, m_{k,\mathcal{P}}, \sigma_{k,\mathcal{P}})$ , où nous désignons en gras la version vectorielle du paramètre, *i.e.*,  $\mathbf{x}_{k,\mathcal{P}} = (x_{k,p})_{p \in \mathcal{P}}$ , par maximum de vraisemblance en utilisant un algorithme espérance-maximisation (EM) [5]. L'algorithme EM [6] est un algorithme itératif qui permet de trouver les estimations du maximum de vraisemblance (EMV) des paramètres d'un modèle probabiliste lorsque celui-ci dépend de variables latentes non observables. Tout d'abord, à partir de (2), nous dérivons la log-vraisemblance négative du modèle [3, chap. 2.1] :

$$\begin{aligned} \mathcal{L}_{k,\mathcal{P}}(\Theta_{k,\mathcal{P}}) &= \mu_k T + \sum_{p \in \mathcal{P}} \alpha_{k,p} n_p \\ &- \sum_{t \in \mathcal{A}_k} \log \left( \mu_k + \sum_{p \in \mathcal{P}} \sum_{i, t_i^{(p)} \leq t} \alpha_{k,p} \kappa_{k,p}(t - t_i^{(p)}) \right) \end{aligned}$$

**Étape E (espérance)** Pour une estimation donnée, l'étape E vise à calculer l'assignation des événements, c'est-à-dire la probabilité qu'un événement provienne soit du noyau, soit de l'intensité de base. À l'itération  $n$ , soit  $P_k^{(n)}(t) := P_k^{(n)}(t; p, k)$  la probabilité que l'activation au temps  $t \in [0; T]$  ait été déclenchée par l'intensité de base du processus stochastique  $k$  et  $P_p^{(n)}(t) := P_p^{(n)}(t; p, k)$  la probabilité que l'activation au temps  $t$  ait été déclenchée par le *driver*  $p$ . Par la définition de notre modèle d'intensité (2), nous avons :

$$\begin{aligned} P_k^{(n)}(t) &= \frac{\mu_k^{(n)}}{\lambda_{k,\mathcal{P}}^{(n)}(t)} \\ \forall p \in \mathcal{P}, P_p^{(n)}(t) &= \frac{\alpha_{k,p}^{(n)} \sum_{i, t_i^{(p)} \leq t} \kappa_{k,p}^{(n)}(t - t_i^{(p)})}{\lambda_{k,\mathcal{P}}^{(n)}(t)} \quad , \end{aligned}$$

où  $\theta^{(n)}$  désigne la valeur du paramètre  $\theta$  à l'étape  $n$  de l'algorithme, et de la même manière, si  $f$  est une fonction de paramètre  $\theta$ ,  $f^{(n)}(x; \theta) := f(x; \theta^{(n)})$ . Notez que  $\forall t \in [0; T]$ ,  $P_k^{(n)}(t) + \sum_{p \in \mathcal{P}} P_p^{(n)}(t) = 1$ .

**Étape M (maximisation)** Une fois cette assignation calculée, il faut mettre à jour les paramètres du modèle en utilisant l'EMV. Pour obtenir les équations de mise à jour, on fixe les probabilités  $P_k^{(n)}$  et  $P_p^{(n)}$ , et on annule les dérivées de la log-vraisemblance négative par rapport à chaque paramètre. Pour des valeurs données des probabilités  $P_k^{(n)}(t)$  et  $P_p^{(n)}(t)$ , nous donnons ici la mise à jour des paramètres  $\mu$  et  $\alpha$  :

$$\mu_k^{(n+1)} = \frac{1}{T} \sum_{t \in \mathcal{A}_k} P_k^{(n)}(t) \text{ et } \alpha_{k,p}^{(n+1)} = \pi_{\mathbb{R}^+} \left( \frac{1}{n_p} \sum_{t \in \mathcal{A}_k} P_p^{(n)}(t) \right)$$

où  $\pi_{\mathcal{E}}(\cdot)$  désigne l'opérateur de projection sur l'ensemble  $\mathcal{E}$ . Ces deux mises à jour reviennent à maximiser les probabilités que les événements attribués au *driver* ou à l'intensité de base restent attribués au même processus de génération. Ensuite, nous donnons les équations de mise à jour pour  $m$  et  $\sigma$ , ce qui correspond à des estimations paramétriques de chaque paramètre du noyau Gaussien tronqué avec des événements affectés au noyau :

$$m_{k,p}^{(n+1)} = \frac{\alpha_{k,p}^{(n)} \sum_{t \in \mathcal{A}_k} \sum_{i, t_i^{(p)} \leq t} \frac{(t-t_i^{(p)}) \kappa_{k,p}^{(n)}(t-t_i^{(p)})}{\lambda_{k,p}^{(n)}(t)}}{\sum_{t \in \mathcal{A}_k} P_p^{(n)}(t)} - \sigma_{k,p}^{(n)2} \frac{\nabla_m C(m_{k,p}^{(n)}, \sigma_{k,p}^{(n)}, a, b)}{C(m_{k,p}^{(n)}, \sigma_{k,p}^{(n)}, a, b)}$$

$$\sigma_{k,p}^{(n+1)} = \pi_{[\varepsilon; +\infty)} \left( \frac{C(m_{k,p}^{(n)}, \sigma_{k,p}^{(n)}, a, b)}{\nabla_{\sigma} C(m_{k,p}^{(n)}, \sigma_{k,p}^{(n)}, a, b)} \right)$$

$$\times \left( \frac{\alpha_{k,p}^{(n)} \sum_{t \in \mathcal{A}_k} \sum_{i, t_i^{(p)} \leq t} \frac{(t-t_i^{(p)}(t) - m_{k,p}^{(n)})^2 \kappa_{k,p}^{(n)}(t-t_i^{(p)})}{\lambda_{k,p}^{(n)}(t)}}{\sum_{t \in \mathcal{A}_k} P_p^{(n)}(t)} \right)^{1/3}$$

où,  $C(m, \sigma, a, b) := \int_a^b \exp\left(-\frac{1}{2} \frac{(u-m)^2}{\sigma^2}\right) du$ . Ici,  $\varepsilon > 0$  est prédéterminé pour garantir que  $\sigma$  reste strictement positif. En pratique, nous fixons  $\varepsilon$  de manière à éviter le sur-apprentissage qui peut se produire lorsque la masse du noyau est trop concentrée. Notez qu'une fois les valeurs initiales des paramètres déterminées, l'algorithme EM est entièrement déterministe. Pour cela, nous proposons une stratégie d'initialisation par méthode des moments, où les paramètres sont initialisés en fonction de leur rôle dans le modèle (voir la version longue [2] pour le détail).

## 4 Expériences

Nous avons évalué notre modèle sur plusieurs expériences, en utilisant des données MEG synthétiques et empiriques. Nous avons utilisé Python et ses bibliothèques scientifiques, ainsi que `alphasc` pour le CDL avec des contraintes de rang 1 sur les données MEG [1] et nous avons utilisé `MNE` [7] pour charger et manipuler les ensembles de données MEG.

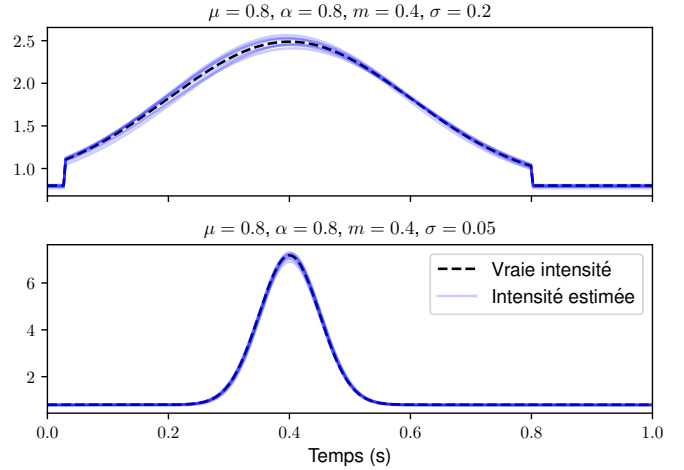


FIGURE 2 – Fonctions d'intensité réelle et estimée après un stimulus au temps zéro pour deux noyaux différents, sur des données synthétiques. **Haut** : noyau « large » avec  $\sigma = 0.2$ . **Bas** : noyau « étroit » avec  $\sigma = 0.05$ . Sur les données synthétiques, l'algorithme EM retrouve avec succès les vraies valeurs des paramètres, pour les deux formes de noyaux.

**Évaluation de l'EM sur données simulées** Pour un ensemble de paramètres  $\Theta_{k,p}$  fixé ainsi qu'une durée de processus  $T$ , on sélectionne un ensemble de *timestamps* possible sur une grille afin de générer les ensembles des événements  $\mathcal{T}_p$ , avec une séparation temporelle minimale entre les stimuli de 1 s pour le noyau large et de 1,4 s pour le noyau étroit. La fonctionne d'intensité  $\lambda_{k,p}$  est alors entièrement définie, et il est possible de simuler les événements associés à l'aide de l'algorithme de seuillage de Lewis [8] (voir [2] pour plus de détails). La **figure 2** illustre l'estimation de la fonction d'intensité avec deux *drivers* de formes différentes considérés ensemble : un avec noyau « large » ( $\sigma = 0.2$ ) et l'autre avec un noyau « étroit » ( $\sigma = 0.05$ ). Ne sont représentés sur la figure que les parties de l'intensité au niveau d'un noyau, pour 8 simulations indépendantes et une durée de processus de  $T = 1 \times 10^4$  s. L'algorithme EM est exécuté avec 50 itérations en utilisant la stratégie d'initialisation par méthode des moments. Ici, le caractère aléatoire provient uniquement de la génération des données, car l'algorithme EM utilise une initialisation déterministe. La **figure 2** montre bien que l'algorithme EM est capable d'estimer avec succès les paramètres pour les deux formes de noyaux. D'autres résultats sur données simulées, disponibles dans la version longue [2], montrent que la précision des estimations EM augmente avec des processus plus longs et plus denses.

**Caractérisation des effets évoqués et induits dans les données MEG** Les expériences sur les données MEG ont été réalisées sur deux jeux de données provenant du package Python MNE [9] : *sample* obtenu lors de stimulations audio-visuelles et *somato* obtenu lors de stimulation du nerf médian<sup>1</sup>. Ces jeux de données ont été sélectionnés car ils présentent deux types

1. Tous deux disponibles sur [https://mne.tools/stable/overview/datasets\\_index.html](https://mne.tools/stable/overview/datasets_index.html)

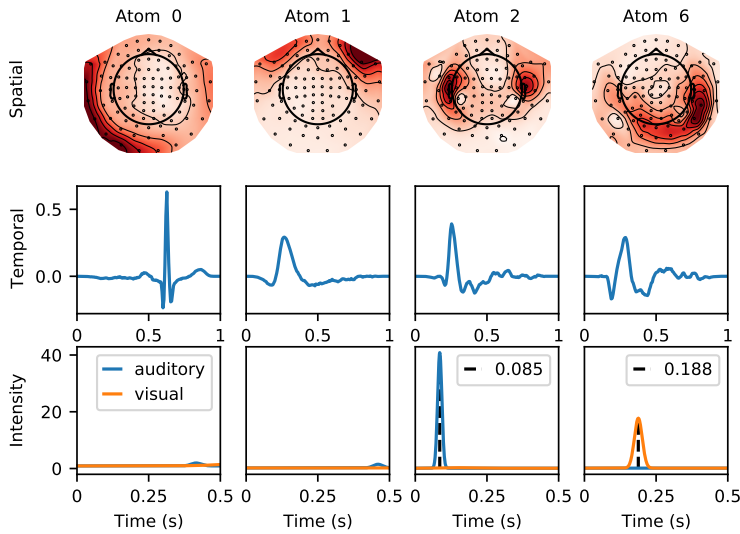


FIGURE 3 – Représentations spatiales et temporelles de 4 atomes extraits de *MNE sample*, et leurs fonctions d’intensité estimées respectives après un stimulus auditif ou visuel au temps  $t = 0$  s. Les artefacts de battement de cœur et de clignement des yeux ne sont liés à aucun stimulus. Un stimulus auditif (resp. visuel) induit une réponse neuronale similaire à celle de l’atome 2 (resp. 6), avec une latence moyenne de 85 ms (resp. 188 ms).

distincts d’activations neuronales liées à des événements : des réponses évoquées dont la latence est quasi-fixe et des réponses induites qui présentent des latences plus variables. Des expériences complémentaires ont été réalisées sur le jeu de données Cam-CAN [10]. La présentation détaillée des jeux de données ainsi que l’ensemble des résultats obtenus sont présents dans la version longue [2].

Sur les données MNE sample, un prétraitement des données est appliqué, et 40 atomes d’une durée d’1 s sont extraits à l’aide d’un CDL (voir [2] pour plus de détails). Pour chacun des atomes ainsi extraits, on estime la fonction d’intensité définie avec les deux types de stimuli conjointement (stimulus auditif et visuel) avec 400 itérations d’EM et une initialisation par méthode des moments, et où chacun des deux noyaux a pour support l’intervalle  $[0,03\text{ s}; 0,5\text{ s}]$ . Les résultats sont présentés en figure 3, où il est reporté, pour 4 atomes bien choisis, les représentations spatiales et temporelles, ainsi que la fonction d’intensité estimée au niveau des noyaux en fonction du stimulus associé. Les deux premiers atomes (0 et 1) peuvent aisément être associés respectivement aux artefacts dus aux battements du cœur et aux clignements des yeux (à noter qu’une ICA permet généralement de les retrouver). La fonction d’intensité est plate pour les deux types de stimulus, ce qui conforte l’intuition que l’activation de ces atomes n’est pas liée à l’apparition de tels stimuli. En revanche, les représentations de l’atome 2 (resp. 6) nous permet de le lier à une réponse auditive bilatérale (resp. visuelle), ce qui est confirmé par la fonction d’intensité pour le stimulus associé qui forme un pic étroit autour de 85 ms (resp. 188 ms) post-stimulus, tandis que celle pour l’autre stimulus reste plate, avec un coefficient  $\alpha$  nul. Ces réponses peuvent être associées respectivement à la réponse M100 et P200.

## 5 Conclusion

Sans aucun réglage ou sélection manuelle, DriPP offre une approche unifiée pour extraire les formes d’onde et sélection-

ner automatiquement celles qui sont susceptibles d’être déclenchées par les stimuli considérés, en capturant la vague d’activations associées aux événements externes, fournissant ainsi une caractérisation statistique directe de quand et comment chaque stimulus est responsable de l’apparition des réponses neuronales. Les travaux futurs exploreront la modélisation des effets d’inhibition fréquemment observés dans les données neuroscientifiques.

## Références

- [1] T. Dupré la Tour, T. Moreau, M. Jas, and A. Gramfort. Multivariate convolutional sparse coding for electromagnetic brain signals. *Advances in Neural Information Processing Systems*, 31 :3292–3302, 2018.
- [2] C. Allain, A. Gramfort, and T. Moreau. Dripp : Driven point processes to model stimuli induced patterns in m/eeeg signals. *10th International Conference on Learning Representations*, 2022.
- [3] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Volume 1 : Elementary theory and methods*. Probability and Its Applications. Springer-Verlag New York, 2003.
- [4] A. G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society : Series B (Methodological)*, 33 (3) :438–443, 1971.
- [5] E. Lewis and G. Mohler. A nonparametric em algorithm for multiscale hawkes processes. *Journal of Nonparametric Statistics*, 1(1) :1–20, 2011.
- [6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society : Series B (Methodological)*, 39(1) :1–22, 1977.
- [7] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in neuroscience*, 7, 2013.
- [8] PA W Lewis and G. S. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval research logistics quarterly*, 26(3) :403–413, 1979.
- [9] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen. MNE software for processing MEG and EEG data. *Neuroimage*, 86 :446–460, 2014.
- [10] M. A. Shafto, L. K. Tyler, M. Dixon, J. R. Taylor, J. B. Rowe, R. Cusack, A. J. Calder, W. D. Marslen-Wilson, J. Duncan, T. Dalgleish, et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol : a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC neurology*, 14(1) :1–25, 2014.