

# Quelques problèmes de statistique en neurophysiologie

Christophe Pouzat

IRMA

`christophe.pouzat@math.unistra.fr`

Séminaire Statistique, lundi 21 septembre 2020

# Outline

« Point de départ » et questions

Le tri des potentiels d'action

Que faire des séquences de potentiels d'action

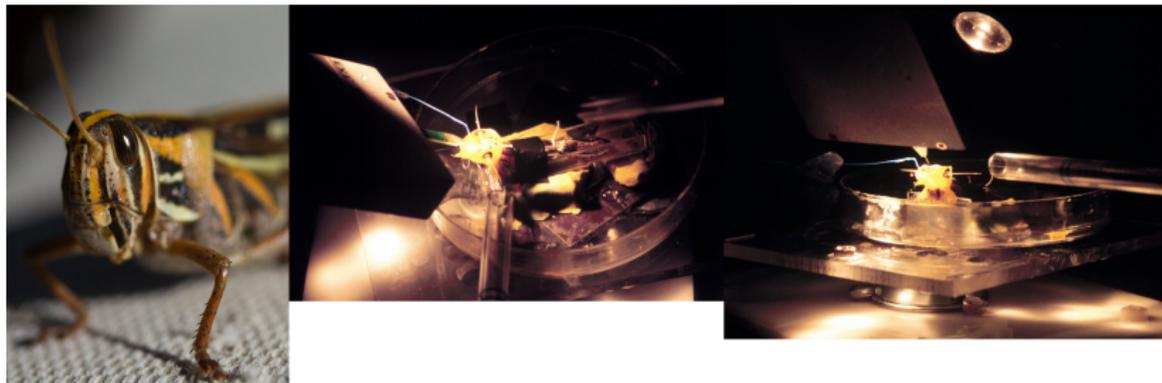
# Où en sommes nous ?

« Point de départ » et questions

Le tri des potentiels d'action

Que faire des séquences de potentiels d'action

# Modèle expérimental



Le criquet, *Schistocerca americana*, « en situation ».

# Si vous vous posez la question : pourquoi étudier les criquets ?

## Le criquet pèlerin

[FAQS](#) [NOUVELLES](#) [RESSOURCES](#)



Le criquet pèlerin (*Schistocerca gregaria*) est le ravageur migrateur le plus destructeur au monde. En réponse à des stimuli environnementaux, des essaims denses et très mobiles de criquets pèlerins peuvent se former. Ce sont des mangeurs voraces qui consomment leur propre poids par jour, ciblant les cultures vivrières et le fourrage. Un seul kilomètre carré d'essaim peut contenir jusqu'à 80 millions d'adultes qui consomment en une journée la même quantité de nourriture que 35 000 personnes. Les grands essaims constituent une menace majeure pour la sécurité alimentaire et les moyens d'existence des populations rurales. La FAO possède une expertise de longue date dans le suivi des populations de criquets pèlerins et dans l'aide aux pays pour faire face à ce ravageur qui dévaste les cultures.

### La crise du criquet pèlerin en Afrique de l'Est

Depuis le début de l'année 2020, le problème de la présence du criquet pèlerin s'est aggravé. Des conditions climatiques favorables ont en effet permis une reproduction à grande échelle du ravageur en Afrique de l'Est, en Asie du Sud-Ouest et dans la zone située autour de la mer Rouge.

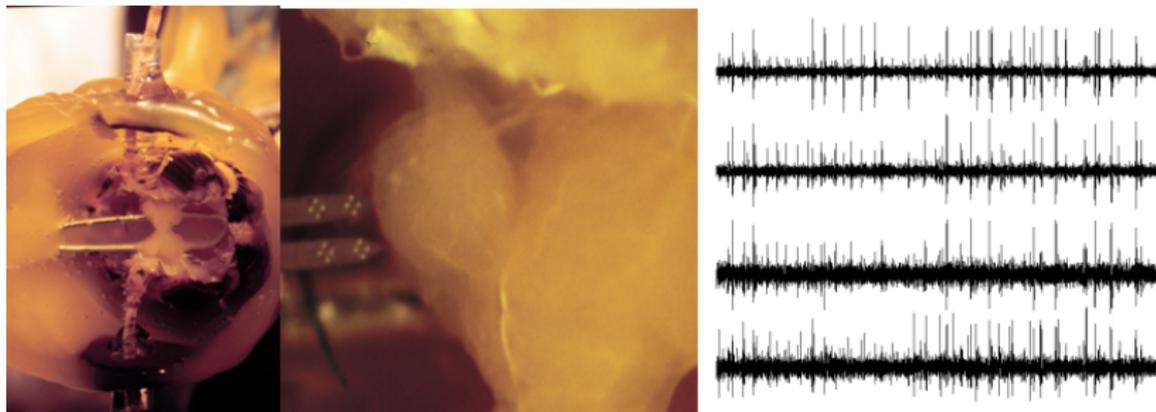
La situation est particulièrement inquiétante en Éthiopie, en Somalie et au Kenya. Les essaims de criquets pèlerins y sont extrêmement importants, très mobiles, et endommagent les cultures vivrières et fourragères. La FAO a fait de la lutte contre la recrudescence du criquet pèlerin en Afrique de l'Est l'une de ses principales priorités. L'Organisation a réagi rapidement pour aider les gouvernements à faire face à ce phénomène.

### Faits marquants



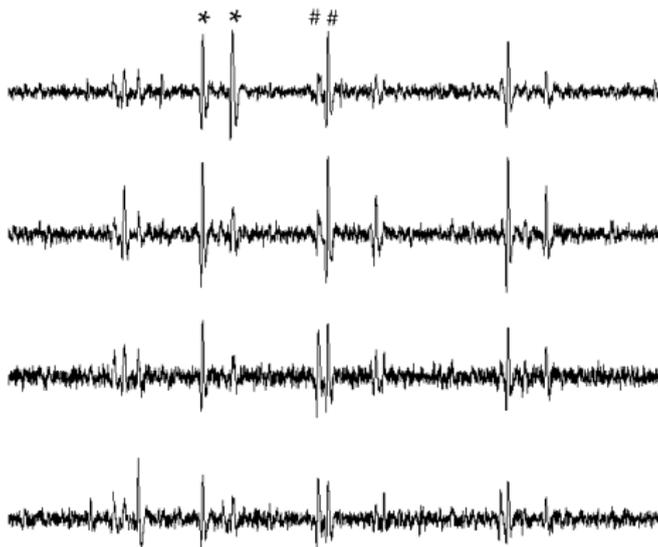
**Histoire:** Se préparer à l'invasion : La

# Les insectes ont un cerveau !



À gauche : la tête ouverte et le cerveau ; au centre : **le lobe antennaire** et la sonde d'enregistrement ; à droite : 1 sec d'enregistrements.

## Pourquoi des *tetodes*?



Les dernières 200 ms de la figure précédente. Avec **seulement** le site du haut, il serait difficile de classer correctement les deux premiers grands pics (« \* \* »). Avec seulement le site du bas, il serait difficile de classer correctement les deux pics marqués par les « # # ».

# Le tri des potentiels d'action intéresse aussi les neurologues

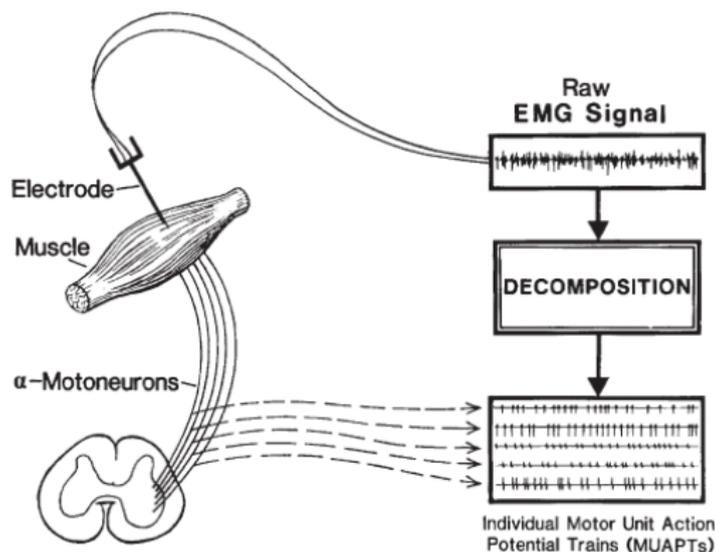
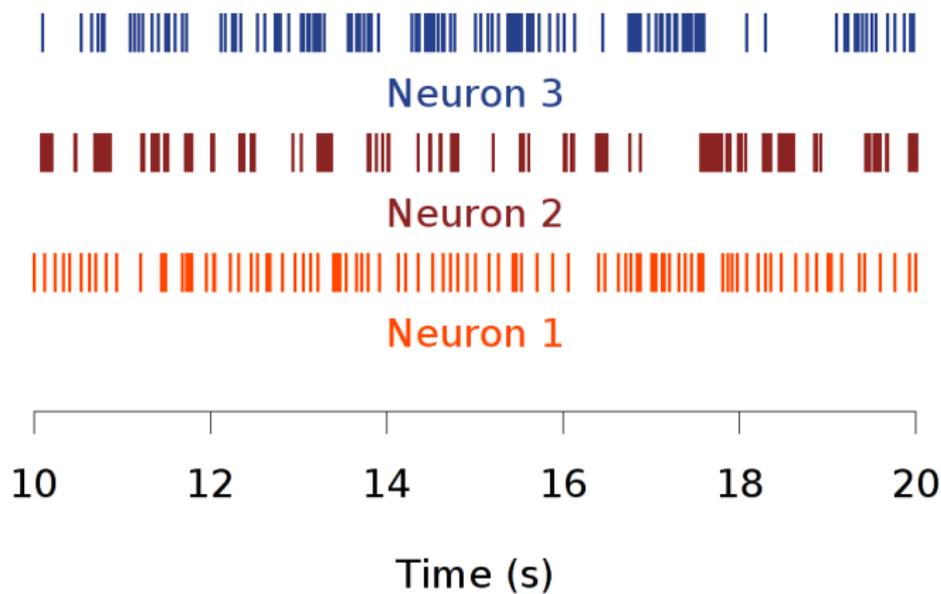


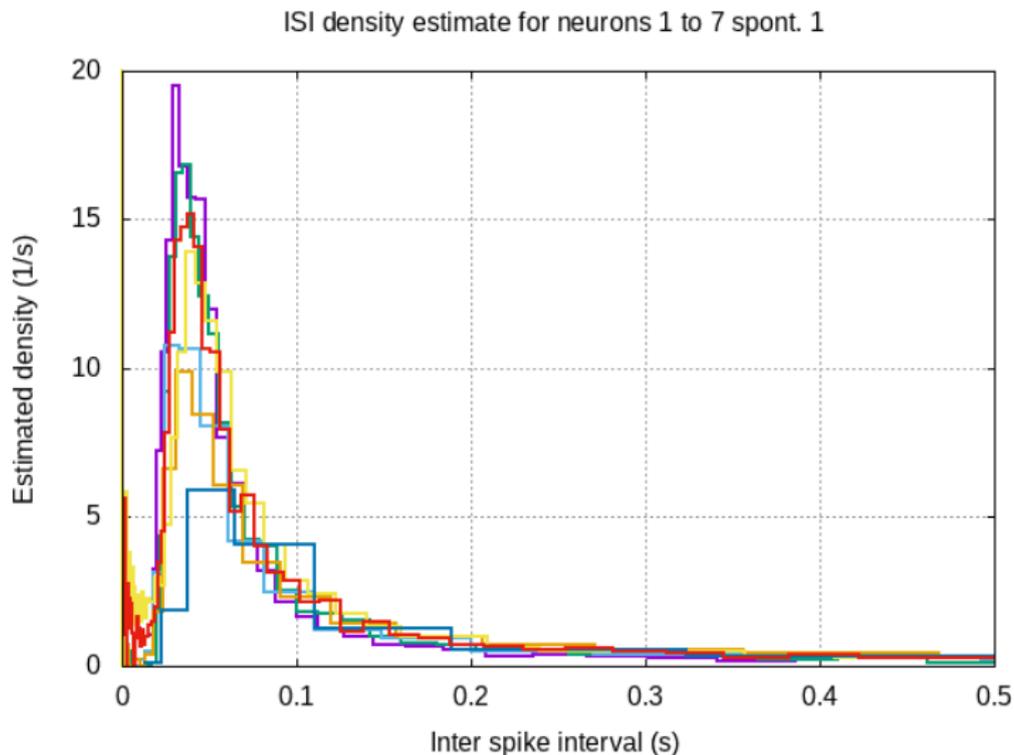
Fig. 1 de Stashuk (2001) *J Electromyogr Kinesiol* **11** :151-173.

# Séquences de potentiels d'action

Après une analyse « un peu lourde » (le **tri des potentiels d'action**), nous pouvons construire les **raster plots** représentant les séquences de PA :

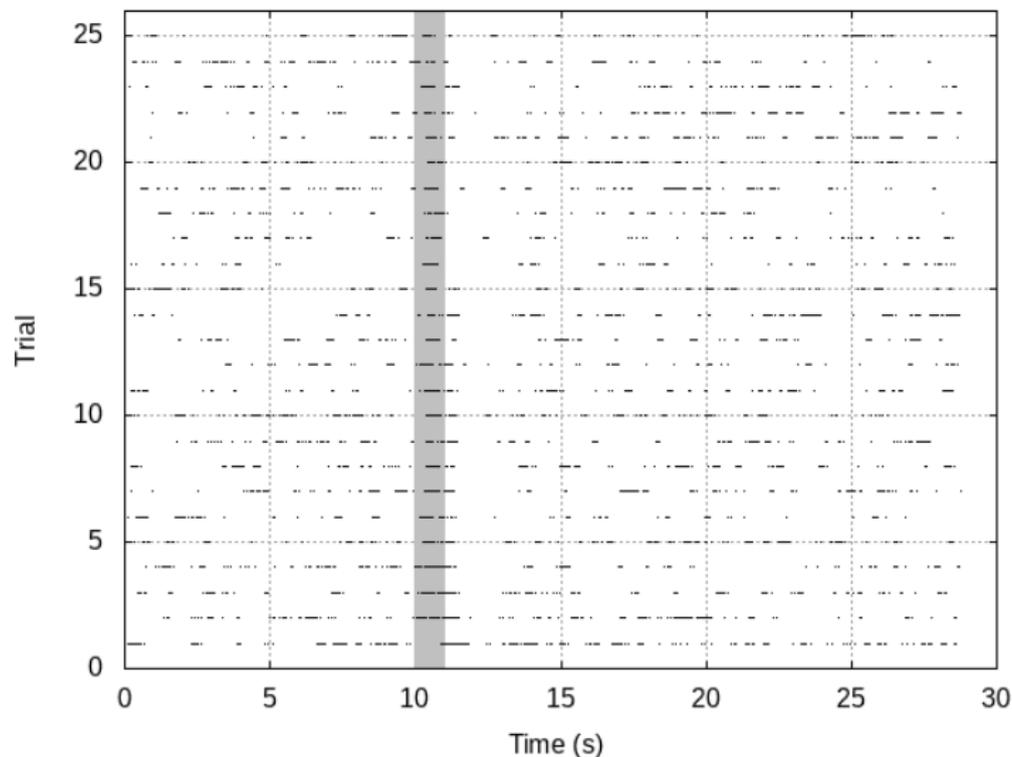


# Distributions d'intervalles



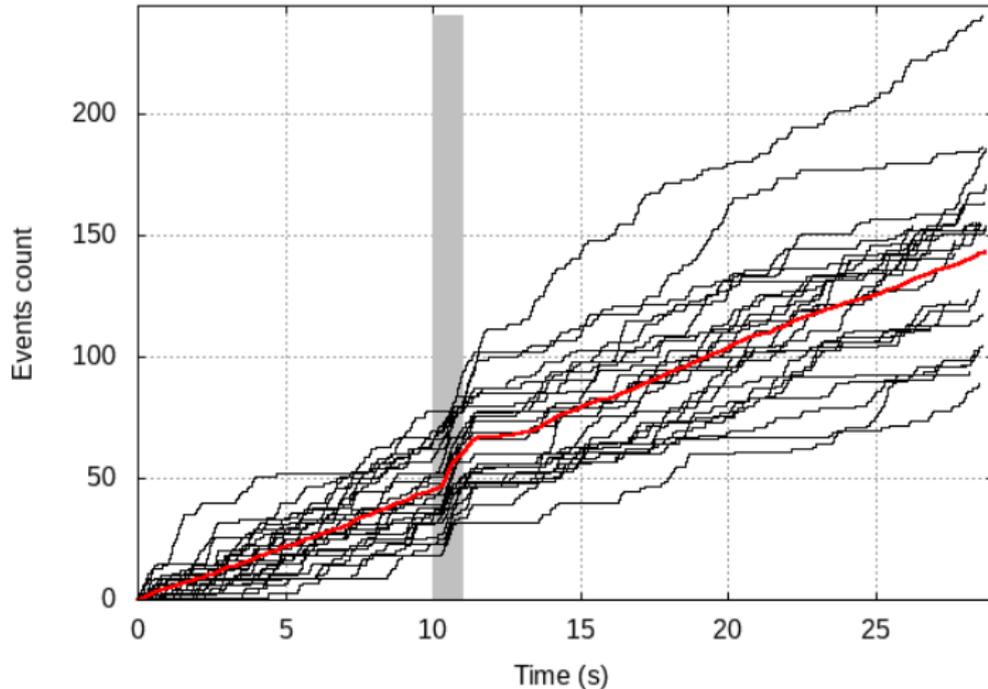
Densités estimées d'intervalles (*inter spike interval, ISI*) pour 7 neurones enregistrés simultanément.

## Réponses à des stimulations



Raster plot d'un neurone lors de 25 présentations successives (fond gris) de cis-3-hexen-1-ol.

Observed counting processes



Compteurs observés pour les 25 présentations. En rouge, la moyenne empirique des 25.

# Questions

- ▶ Comment le tri des potentiels d'actions est-il effectué ?
- ▶ Que faire avec les séquences de potentiels d'action ?

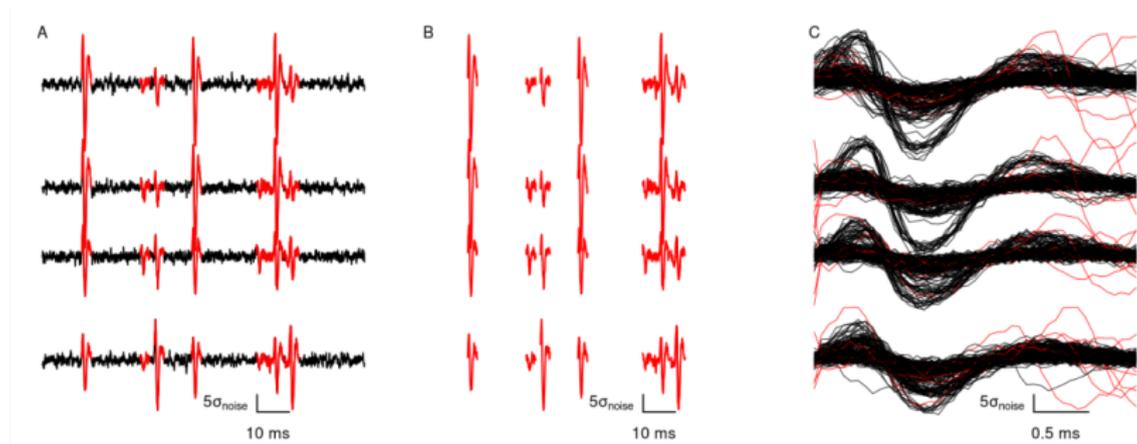
# Où en sommes nous ?

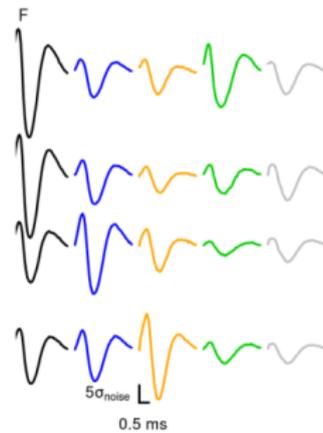
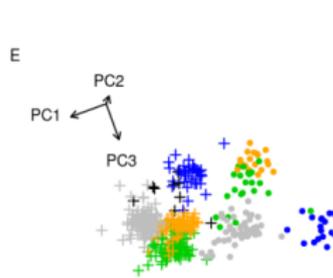
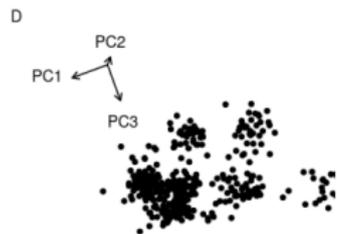
« Point de départ » et questions

Le tri des potentiels d'action

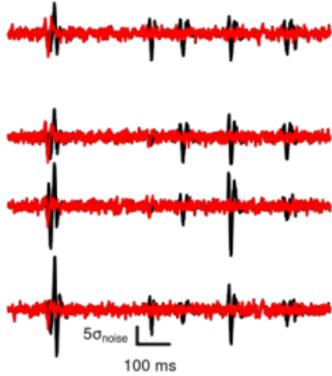
Que faire des séquences de potentiels d'action

# Le tri des PAs en trois figures

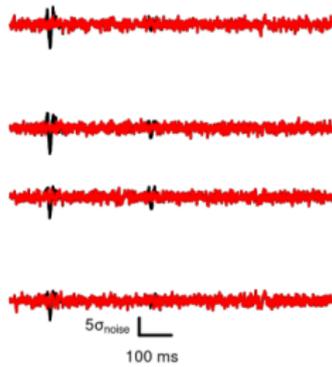




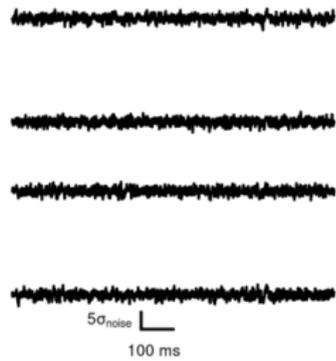
G1



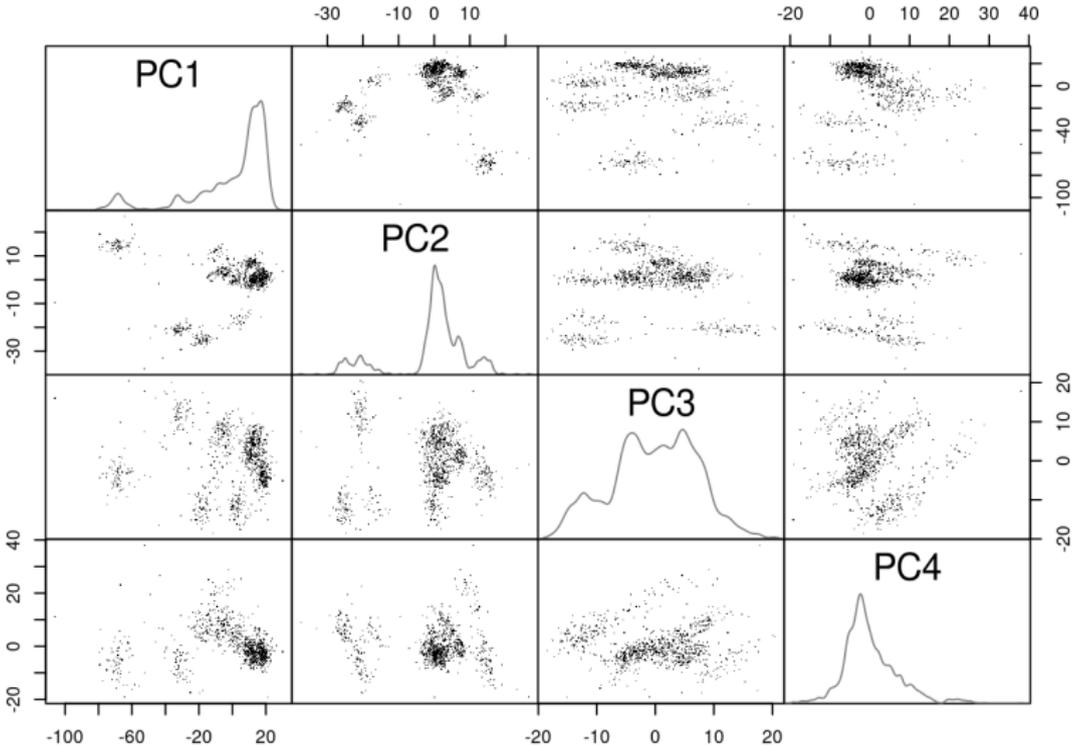
G2

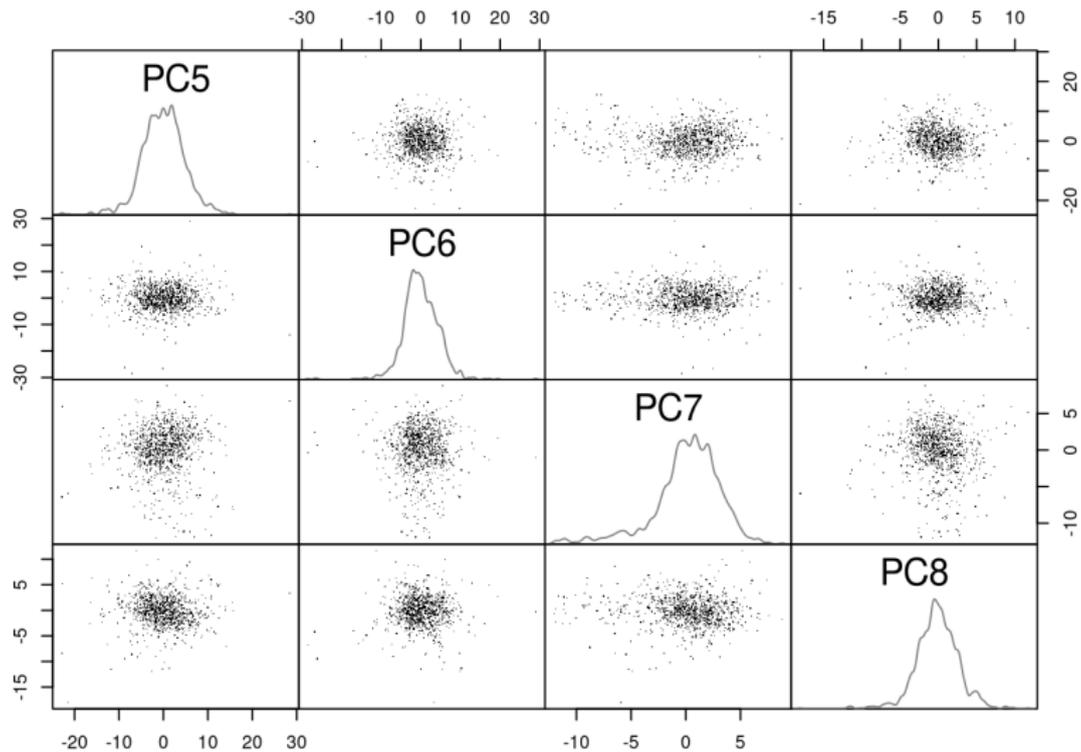


G3

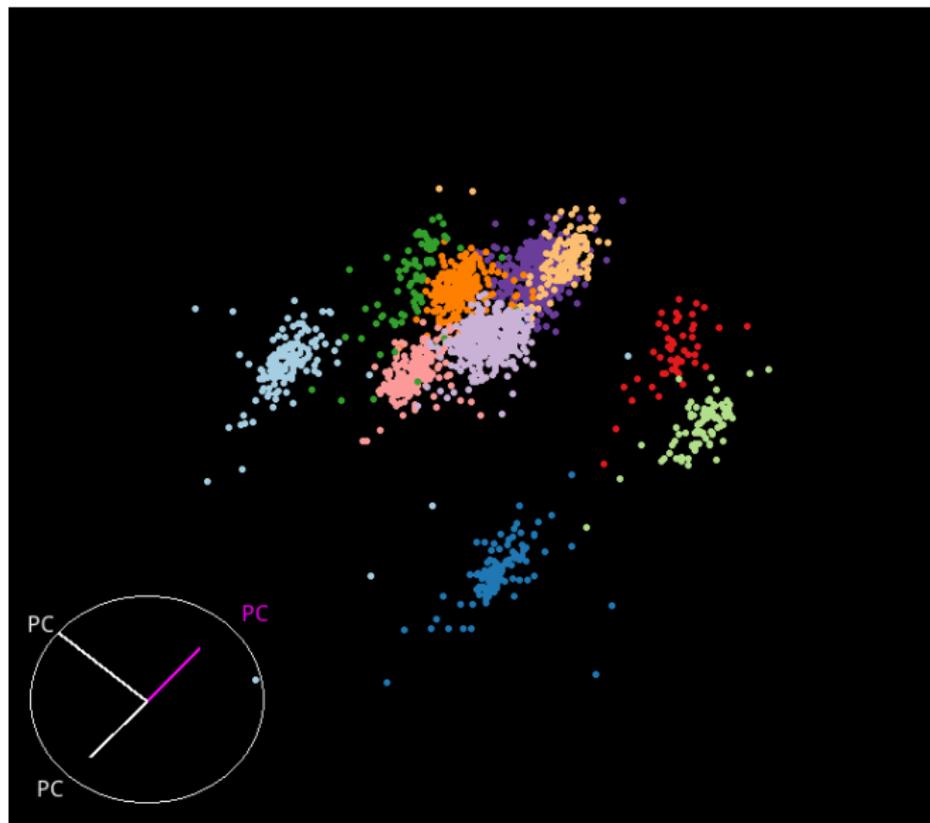


# De la classification non supervisée classique ?





Ne pas oublier le plus important GGobi



# La gigue électronique

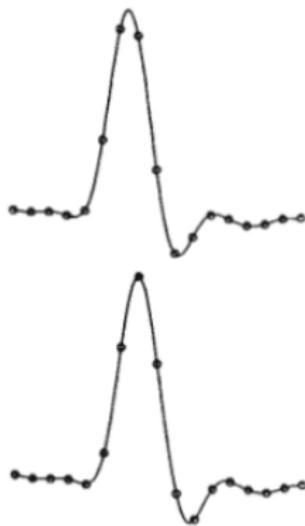


Fig. 1. Two copies of the same waveform, sampled near the Nyquist rate. Sampling at the Nyquist rate only sketches the underlying continuous waveform, and two renditions are apt to appear quite different if the samples are out of phase.

Fig. 1 de McGill et Dorfman (1984) *IEEE Trans Biomed Eng*  
31 :462. **EMG**.

# Correction de la gigue – version « lourde »

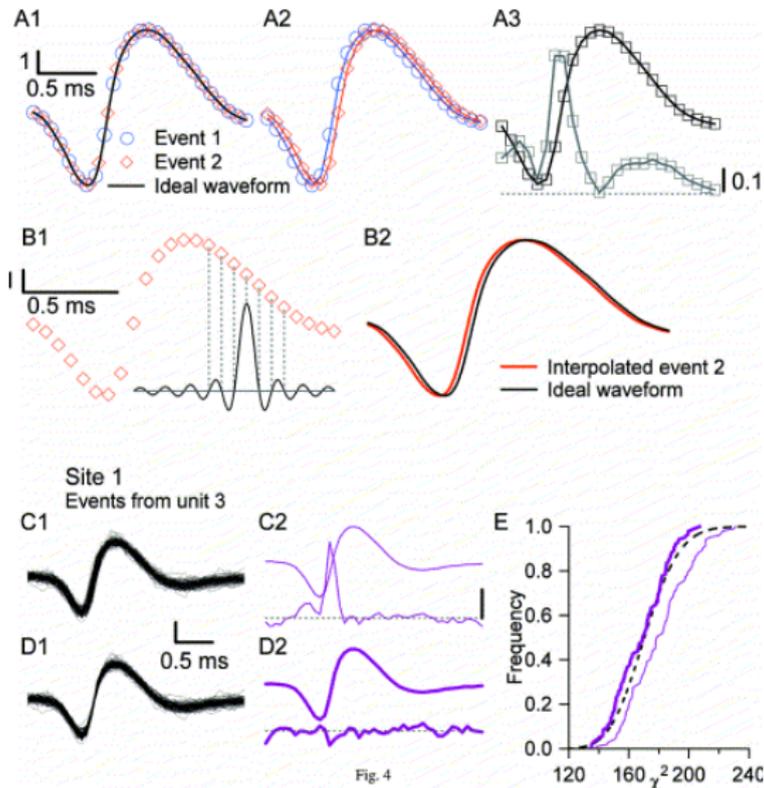


Fig. 4 de Pouzat et col. (2002) *J Neurosci Methods* 122 :43.

## Correction de la gigue – version « légère »

- ▶ soit  $g(t)$  l'amplitude observée au temps  $t$  dans une fenêtre ;
- ▶ alors  $g(t) = f(t + \delta) + \epsilon(t)$ , où  $\delta$  est notre **gigue induite par l'échantillonnage et par le bruit**,  $f$  est notre « forme mère »,  $\epsilon(t)$  est notre bruit ;
- ▶ nous modélisons  $\delta$  comme observation d'une variable aléatoire  $\Delta$  d'espérance,  $E\Delta = 0$ , et de variance finie,  $E(\Delta - E\Delta)^2 = E\Delta^2 \equiv \sigma_{\Delta}^2$  ;
- ▶ en supposant qu'au moins les deux premières dérivées de  $f$  existent (c'est le cas car les données ont été filtrées), nous avons pour  $\delta$  proche de 0 :

$$g(t) = f(t) + \delta f'(t) + \frac{\delta^2}{2} f''(t) + o(\delta^3) + \epsilon(t).$$

- ▶ par hypothèse sur  $\Delta$  nous avons :

$$Eg(t) = f(t) + \frac{\sigma_{\Delta}^2}{2} f''(t) + o(E\Delta^3),$$

- ▶ c'est-à-dire qu'au premier ordre en  $\delta$  :  $Eg(t) = f(t)$ .
- ▶ cela signifie que nos « formes mères », estimées par  $\hat{f}(t) = Eg(t)$ , sont **pratiquement non biaisées si  $\sigma_{\Delta}^2$  ou  $f''$  sont assez petites** ;
- ▶ Au premier ordre en  $\delta$  nous avons pour le variance de  $g(t)$  :

$$E(g(t) - Eg(t))^2 = E(\delta f'(t))^2 = \sigma_{\Delta}^2 f'^2(t),$$

- ▶ ce qui donne notre écart type (au premier ordre en  $\delta$ ) :

$$\sigma_{g(t)} = \sigma_{\Delta} |f'(t)|.$$

- ▶ nous avons en fait :  $g(t_i) = f(t_i + \delta) + \epsilon_i$ , où les  $\epsilon_i$  sont des observations de variables aléatoires (corrélées) d'espérance nulle et de variance finie ;

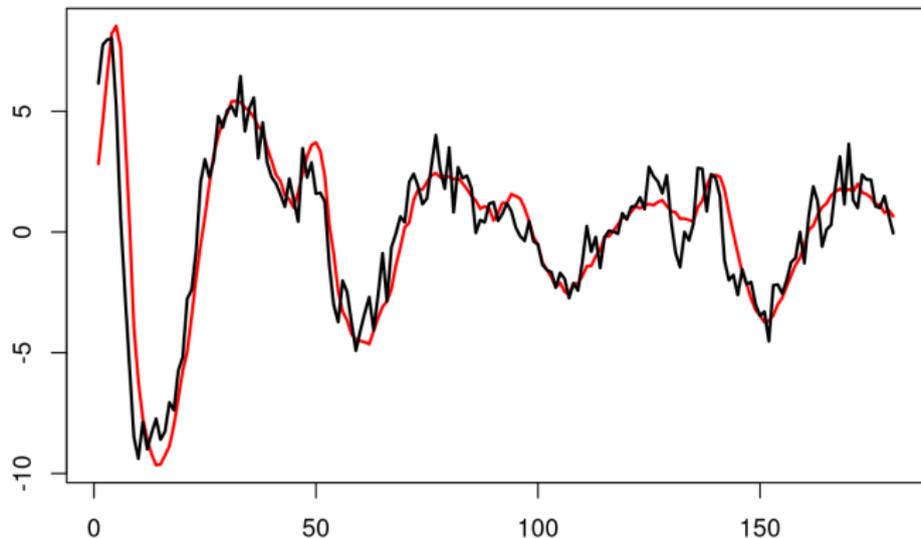
- ▶ au second ordre nous avons :  
 $g(t_i) \approx f(t_i) + \delta f'(t_i) + \delta^2/2 f''(t_i) + \epsilon_i$  ;

- ▶ nous estimons  $\delta$  par  $\hat{\delta}$  défini comme l'argument qui minimise le critère des moindres carrés :

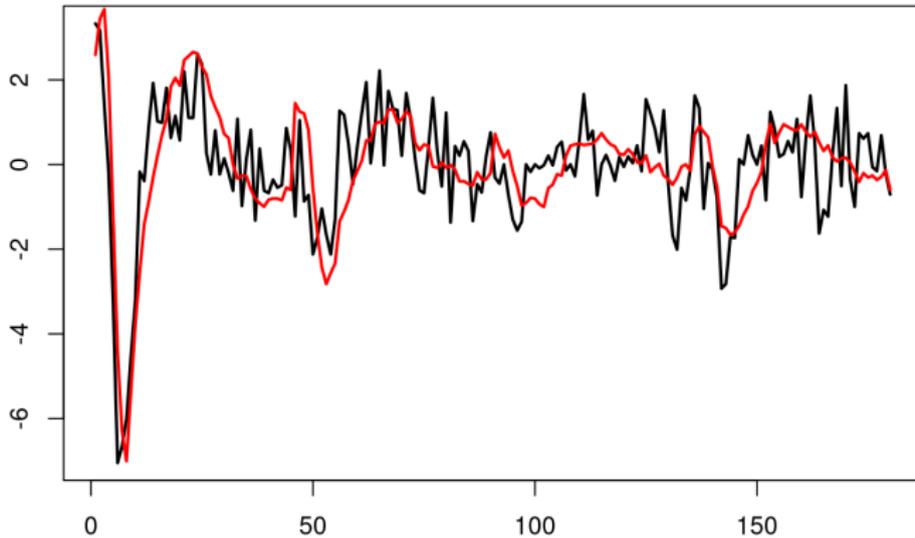
$$\hat{\delta} \equiv \operatorname{argmin}_{\delta} \sum_i \left( g(t_i) - f(t_i) - \delta f'(t_i) - \frac{\delta^2}{2} f''(t_i) \right)^2 .$$

- ▶ le calcul est très rapide dès lors que des estimations de  $f$ ,  $f'$  and  $f''$  sont disponibles ;
- ▶ il est (beaucoup) plus sûr d'utiliser la médiane, plutôt que la moyenne pour ces dernières.

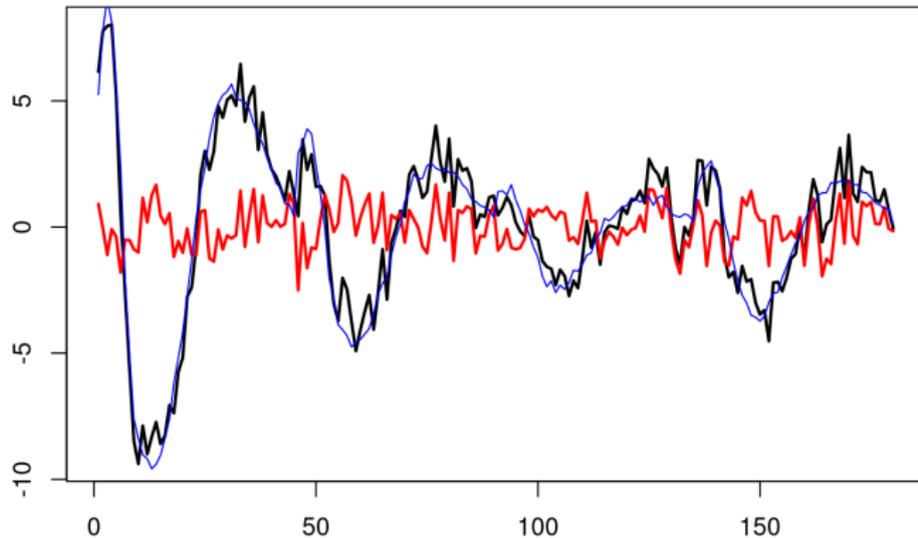
# Illustration



Un événement (en noire) et « sa forme mère » (en rouge).

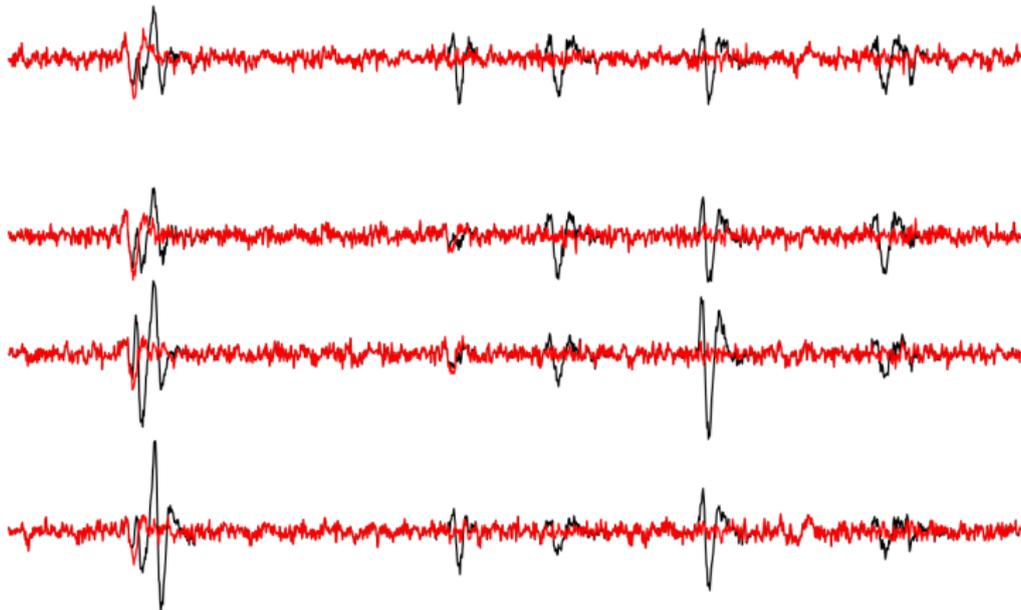


En noir, l'événement dont la forme mère a été soustraite ; en rouge,  $f'$ .



En noire, l'événement ( $g$ ); en bleu,  $f + \hat{\delta}f' + \hat{\delta}^2/2f''$ ; en rouge,  $g - f - \hat{\delta}f' - \hat{\delta}^2/2f''$ .

# L'épluchage en action





## Pour aller plus loin

- ▶ Mes données de post-doc sont toutes disponibles sur zenodo : Pouzat, C., Mazor, O., & Laurent, G. (2015). Extracellular recordings from the locust, *Schistocerca americana*, olfactory pathway [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.21589> ;
- ▶ une description très détaillée de la méthode de tri avec les codes R, etc est disponible sur GitLab, allez sur le serveur GitLab (<https://gitlab.com/>), cherchez mon nom, puis `etude-de-cas-m1` ; tout est dans le sous répertoire `TriDesPA` ;
- ▶ une vidéo YouTube décrit et discute la procédure pas à pas.
- ▶ le code (R) du trie de mes données de post-doc est aussi disponible sur GitHub.

# Où en sommes nous ?

« Point de départ » et questions

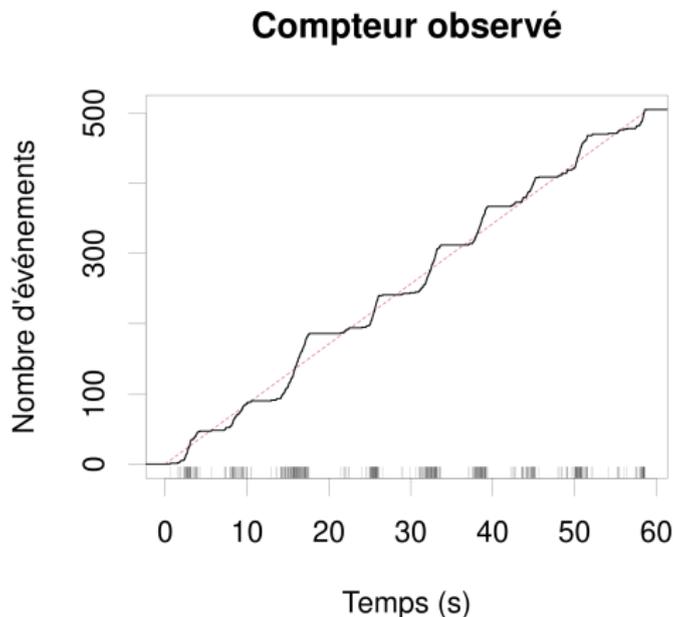
Le tri des potentiels d'action

Que faire des séquences de potentiels d'action

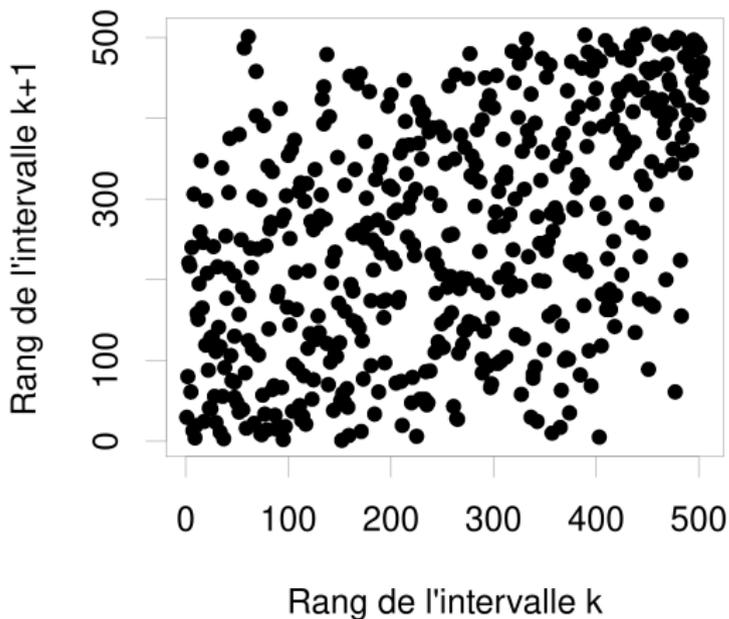
# Pourquoi et comment modéliser les trains de PAs ?

- ▶ une hypothèse de travail centrale en neurosciences est que les temps d'apparition des PAs, par opposition à leurs formes, sont le seul support de transmission de l'information entre régions du cerveau ;
- ▶ cette hypothèse légitime l'étude des trains de PAs en tant que séquences de points sur la demi-droite réelle (représentant le temps) sans nécessairement tenir compte des mécanismes biophysiques qui les génèrent ;
- ▶ Dans ce qui suit nous allons assimiler un train de PAs à un processus ponctuel auquel nous allons associer un compteur (ou processus de comptage).

# Un cas compliqué



L'espérance du compteur d'un processus de Poisson homogène – de même fréquence moyenne — est figurée en rouge (pointillés). Elle est utilisée comme « test » de stationnarité.



Un processus de renouvellement est ici inadéquat : le rang des intervalles inter-PA successifs **sont corrélés**.

# Cahier des charges du modèle

Notre modèle devrait nous permettre de travailler avec :

- ▶ le temps écoulé depuis le dernier PA du neurone (suffisant pour un processus de renouvellement homogène) ;
- ▶ des variables correspondant à l'histoire de la décharge du neurone, comme la durée de l'intervalle entre les deux derniers PAs ;
- ▶ des variables associées au temps écoulé depuis le dernier PA d'un neurone « fonctionnellement associé » ;
- ▶ le temps écoulé par-rapport au début d'une stimulation.

# Filtration, histoire et intensité conditionnelle

- ▶ Les probabilistes des processus introduisent la filtration ou l'histoire : une famille de tribus croissantes,  $(\mathcal{F}_t)_{0 \leq t \leq \infty}$ , telle que toute l'information relative au processus au temps  $t$ , peut-être représentée par un élément de  $\mathcal{F}_t$  ;
- ▶ l' **intensité conditionnelle** du compteur  $N(t)$  est alors définie par :

$$\lambda(t | \mathcal{F}_t) \equiv \lim_{h \downarrow 0} \frac{\text{Prob}\{N(t+h) - N(t) = 1 | \mathcal{F}_t\}}{h} .$$

- ▶ avec  $\lambda$ , nous obtenons une **description exhaustive** de notre processus / séquence de PAs.

# Les deux problèmes à résoudre

Maintenant que nous avons adopté un formalisme basé sur l'intensité conditionnelle, nous devons :

- ▶ trouver un estimateur  $\hat{\lambda}$  de  $\lambda$  ;
- ▶ trouver des tests de qualité d'ajustement de nos modèles à nos données.

## Que faire de $\lambda$ : un résumé

Après avoir associé à  $\lambda$ , l' **intensité cumulée** :

$$\hat{\Lambda} = \int_0^t \hat{\lambda}(u | \mathcal{F}_u) du,$$

il est facile – mais un peu long dans ce type de présentation – de prouver les résultats suivants :

- ▶ **si notre modèle est bon** ( $\hat{\lambda} \approx \lambda$ ), la densité des intervalles entre PAs successifs après transformation du temps :

$$\{t_1, \dots, t_n\} \rightarrow \{\hat{\Lambda}(t_1) = \hat{\Lambda}_1, \dots, \hat{\Lambda}(t_n) = \hat{\Lambda}_n\}$$

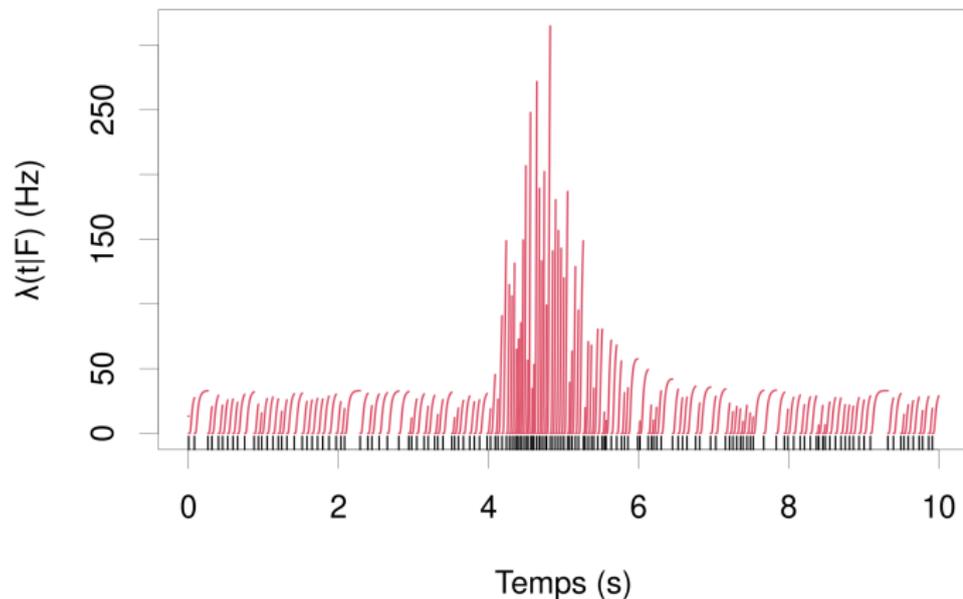
est une **densité exponentielle de paramètre 1** ;

- ▶ le processus ponctuel observé  $\{\hat{\Lambda}_1, \dots, \hat{\Lambda}_n\}$  est donc l'observation d'un **processus de Poisson homogène de paramètre 1**.

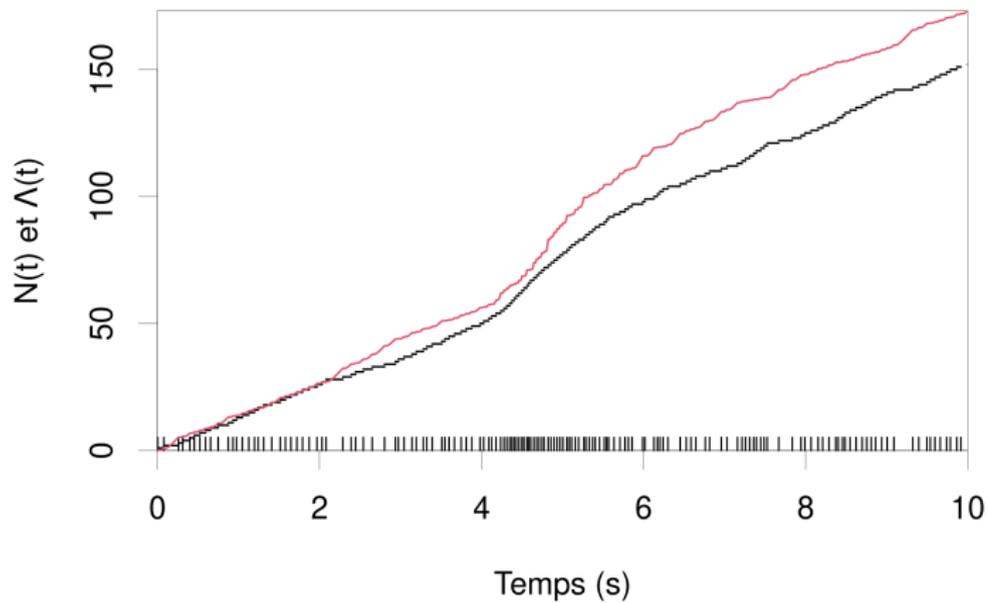
Les quelques diapos suivantes vont illustrer ces résultats.

# Illustration de la transformation du temps sur données simulées

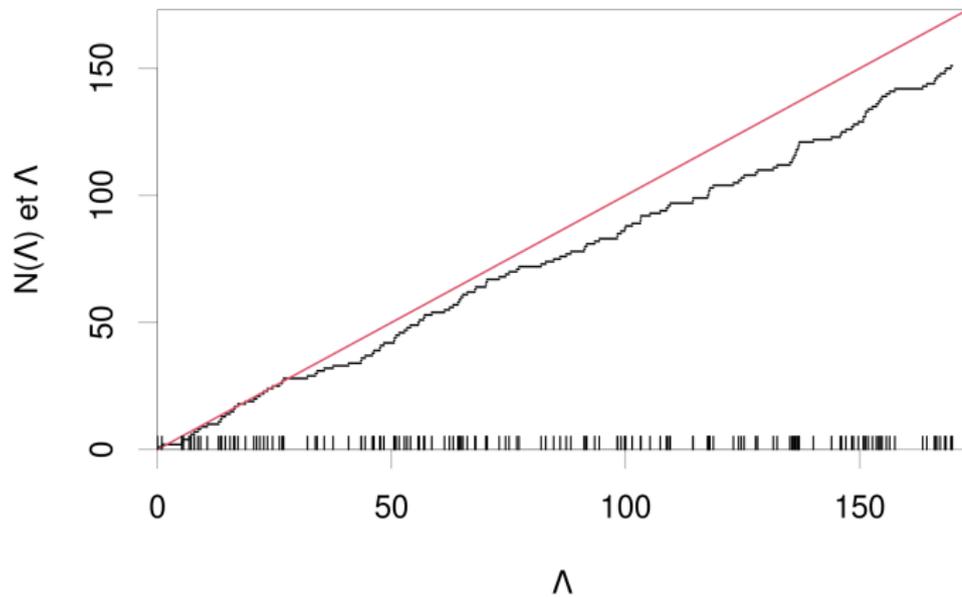
## Processus d'intensité et séquence d'événements



## Trajectoires de N et $\Lambda$



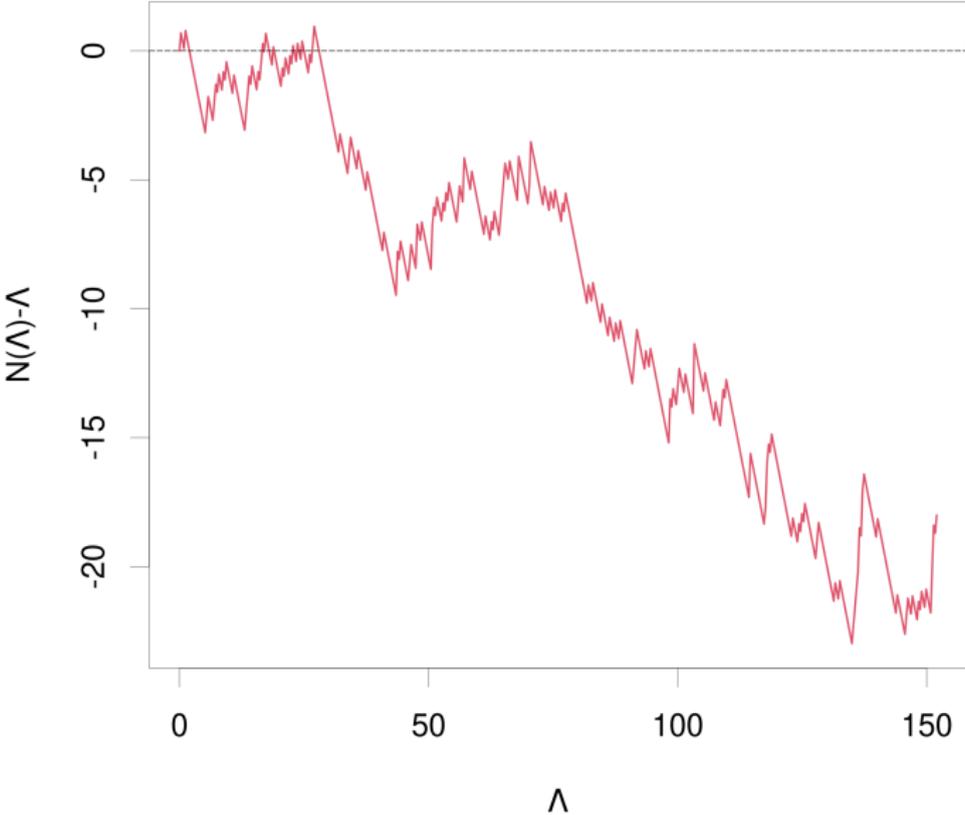
## Trajectoires de $N$ et $\Lambda$ en fonction de $\Lambda$



## Les tests de Lewis / Ogata

- ▶ si, pour un bon modèle, la séquence « transformée » de PAs,  $\{\hat{\Lambda}_1, \dots, \hat{\Lambda}_n\}$ , est une réalisation d'un processus de Poisson homogène de paramètre 1, il est possible de tester l'adéquation d'un modèle ajuster en comparant  $\{\hat{\Lambda}_1, \dots, \hat{\Lambda}_n\}$  à un Poisson homogène ;
- ▶ c'est ce qu'ont proposé Peter Lewis (1973) et Yoshihiko Ogata (1988) ;
- ▶ une observation suggère, néanmoins, qu'un autre type de test pourrait s'appliquer à notre problème...

# Un mouvement brownien ?

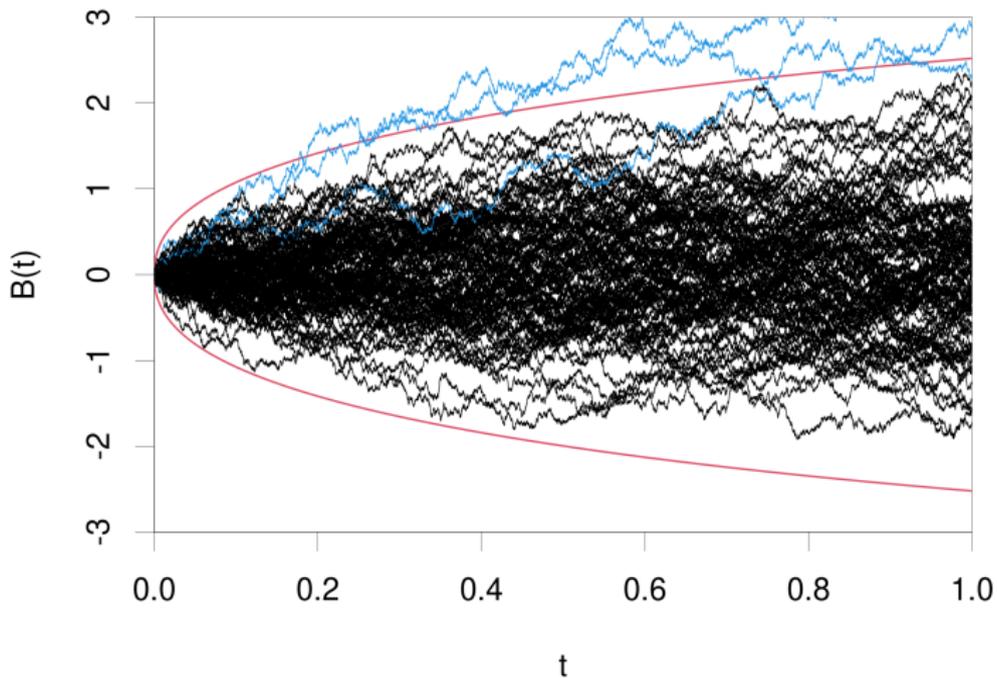


# Théorème de Donsker et région de surface minimale

- ▶ l'intuition d'une convergence – d'une version correctement normalisée – du processus  $N(\Lambda) - \Lambda$  vers un mouvement brownien est correcte ;
- ▶ c'est le théorème de Donsker, comme me l'a indiqué Vilmos Prokaj sur la *mailing list* de  $\mathbb{R}$  ;
- ▶ il est de plus possible de définir des régions de surfaces minimales ayant des probabilités données de contenir toutes les réalisations d'un mouvement brownien canonique (Kendall, Marin et Robert, 2007 ; Loader et Deely, 1987) ;
- ▶ nous obtenons ainsi un nouveau « test » de qualité d'ajustement.

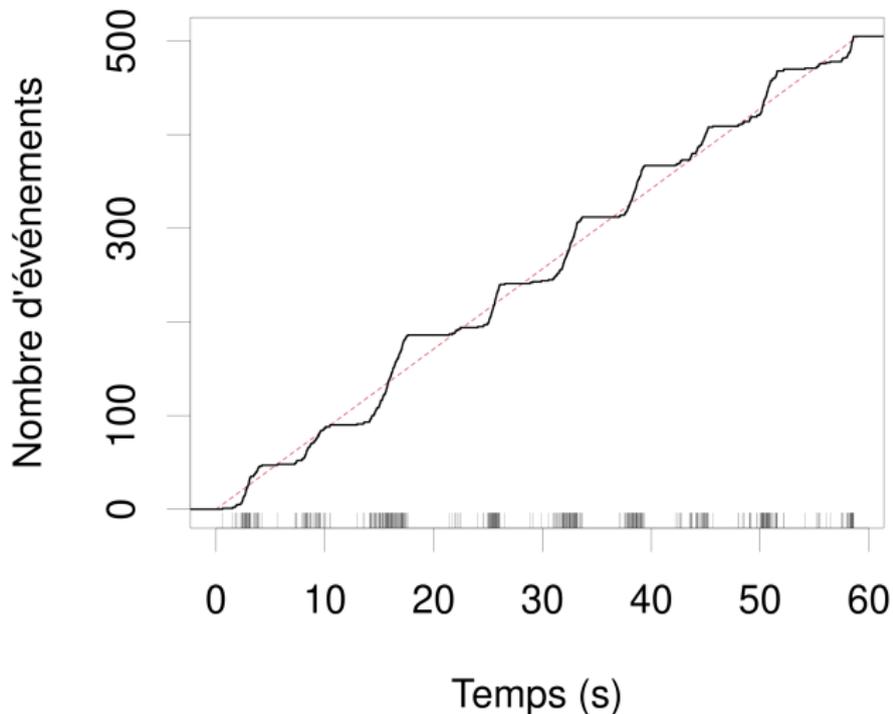
# Région de prédiction à 95 % de surface minimale

**n = 100**



## Retour sur le cas « compliqué »

### Compteur observé



L'analyse exploratoire précédente nous impose de considérer un modèle minimal du type :

$$\lambda(t|\mathcal{F}_t) = f(t - t_d, t_d - t_{ad}),$$

où  $t_d$  est le temps du dernier PA précédent  $t$  et  $t_{ad}$ , celui de l'avant dernier.

# Approche de David Brillinger

- ▶ nous suivons Brillinger (1988) qui commence par discrétiser l'axe des temps en blocs de durée  $h$ ;  $h$  étant suffisamment petit pour avoir au plus un PA par bloc;
- ▶ nous sommes ainsi ramenés à un problème de **régression binomiale**;
- ▶ les « données discrétisées » sont alors considérées comme une observation d'une collection de variables aléatoires de Bernoulli  $\{Y_1, \dots, Y_k\}$  de paramètres :  
 $f(t - t_d, t_d - t_{ad}) h$ ;
- ▶ nous allons directement estimer :

$$\log \left( \frac{f(t - t_d, t_d - t_{ad}) h}{1 - f(t - t_d, t_d - t_{ad}) h} \right) = \eta(t - t_d, t_d - t_{ad}).$$

## Les données discrétisées

	event	time	neuron	lN.1	i1
14604	0	58.412	1	0.012	0.016
14605	1	58.416	1	0.016	0.016
14606	0	58.420	1	0.004	0.016
14607	1	58.424	1	0.008	0.016
14608	0	58.428	1	0.004	0.008
14609	0	58.432	1	0.008	0.008
14610	1	58.436	1	0.012	0.008
14611	0	58.440	1	0.004	0.012

**event** est la séquence de PAs discrétisée ; **time** est le temps du centre des blocs ; **neuron** est le neurone auquel **event** « appartient » ; **lN.1** est  $t - t_d$  ; **i1** est  $t_d - t_{ad}$  . Ici,  $h$  est égal à 4 ms.

## Splines de lissage

- ▶ comme la biophysique cellulaire ne nous donne que peu d'informations sur la forme fonctionnelle de  $\eta$ , nous allons employer des fonctions splines et nous allons pénaliser la vraisemblance (Wahba, 1990 ; Green et Silverman, 1994 ; Eubank, 1999 ; Gu, 2002) ;
- ▶ nous effectuons les calculs avec le paquet gss de Chong Gu ;
- ▶  $\eta(t - t_d, t_d - t_{ad})$  est décomposée de façon unique en :

$$\eta_{\emptyset} + \eta_1(t - t_d) + \eta_2(t_d - t_{ad}) + \eta_{1,2}(t - t_d, t_d - t_{ad}),$$

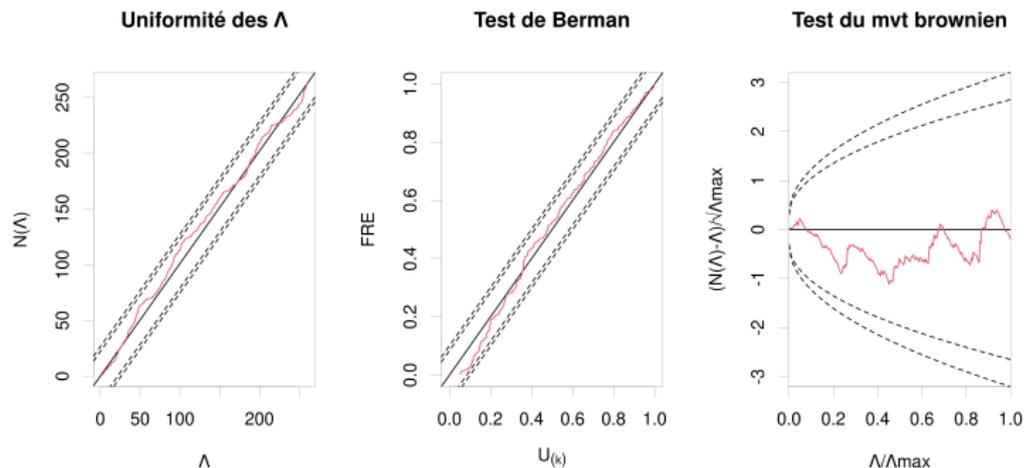
où les variables :  $t - t_d$  et  $t_d - t_{ad}$  ont été transformées (linéairement) pour avoir leurs domaines définitions égaux à  $[0,1]$  ;

- ▶ la décomposition est unique parce-qu'on impose des conditions du type :  $\int_0^1 \eta_i = 0$ .

## Remarque sur les tests

- ▶ les tests de Lewis / Ogata, tout comme le test « du mouvement brownien » que nous proposons, supposent qu'aucun paramètre n'est dépendant des données ;
- ▶ en général, quand la taille de l'échantillon est grande et que le nombre de paramètres est « petit », cela ne pose pas de problème ;
- ▶ mais nous sommes ici dans un contexte non-paramétrique : le nombre de paramètres augmente avec la taille de l'échantillon ;
- ▶ nous adoptons donc la stratégie suivante : nous ajustons le (ou les) modèle(s) sur une moitié des données et nous testons sur l'autre ; nous échangeons les rôles des deux moitiés.

# Modèle sans interaction : ajustement début / test fin

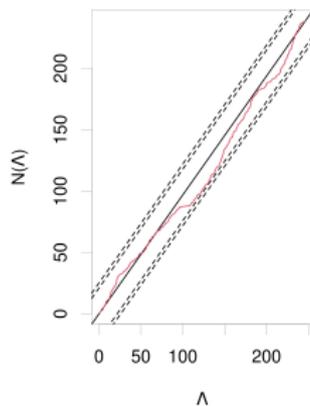


Le modèle est ici :

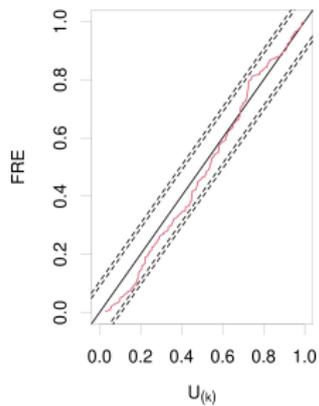
$$\eta(t - t_d, t_d - t_{ad}) = \eta_0 + \eta_1(t - t_d) + \eta_2(t_d - t_{ad}).$$

# Modèle sans interaction : ajustement fin / test début

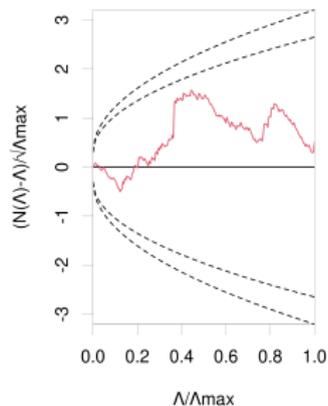
**Uniformité des  $\Lambda$**



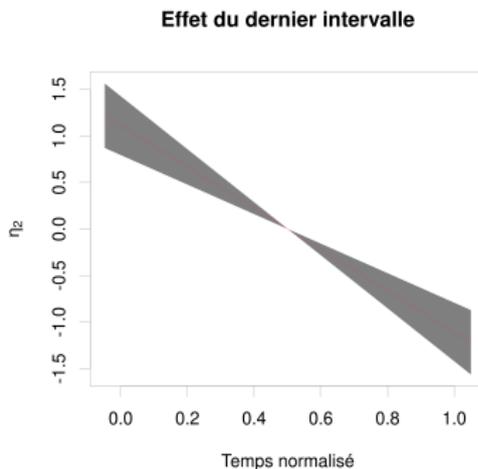
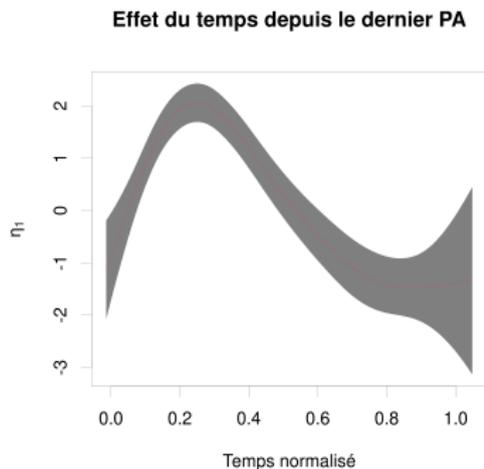
**Test de Berman**



**Test du mvt brownien**

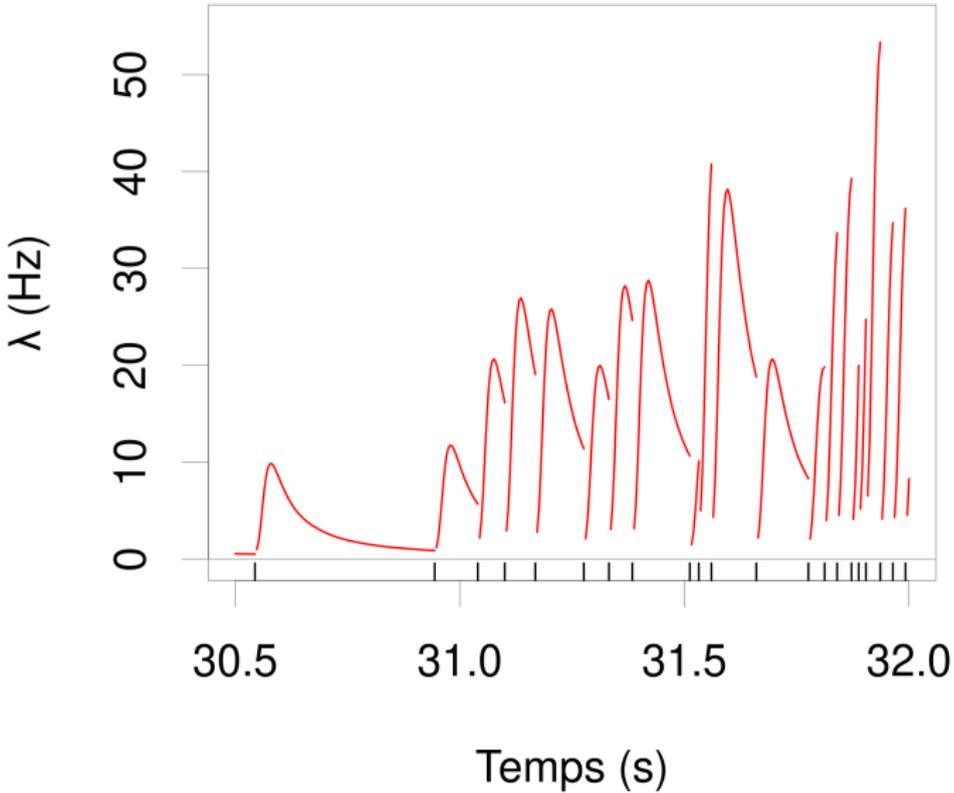


$\eta_1$  et  $\eta_2$



Les bandes grises correspondent à des intervalles de confiance à 95 % ponctuels.

Données et  $\hat{\lambda}$



Merci !

Merci pour l'invitation et votre attention.