

Développements méthodologiques et logiciels pour l'analyse de données neurophysiologiques

Christophe Pouzat

Mathématiques Appliquées à Paris 5 (MAP5)

Université Paris-Descartes et CNRS UMR 8145

`christophe.pouzat@parisdescartes.fr`

Vendredi 14 mars 2014

Où en est-on ?

Introduction

Tri des potentiels d'action dans les cas simples

Analyse des séquences de potentiels d'action

Analyse des mesures de $[Ca^{2+}]$ par sondes fluorescentes

Conclusions et perspectives

De quoi allons-nous parler ?

Cet exposé va être consacré à des travaux sur :

- ▶ le tri des potentiels d'action ;
- ▶ l'analyse des séquences de potentiels d'action ;
- ▶ l'analyse des mesures de $[Ca^{2+}]$ par sondes fluorescentes.

Tous ces travaux portent sur des **données réelles**.

Les aspects méthodologiques et logiciels seront les seuls discutés ici, même si j'ai été directement et fortement impliqué dans les expériences elles-mêmes jusqu'en 2007.

La période couverte par cet exposé va de janvier 1999 à aujourd'hui.

Tri des potentiels d'action et analyse de leurs séquences

(1)

Les problèmes traités dans les deux premières parties de cette présentation apparaissent fréquemment et conjointement dans des études expérimentales contemporaines où les neurophysiologistes cherchent à enregistrer de **nombreux neurones simultanément**.

Les neurophysiologistes cherchent à enregistrer de nombreux neurones simultanément car :

- ▶ ils peuvent ainsi collecter plus de données par expérience ;
- ▶ ils ont de bonnes raisons de penser qu'une partie du « traitement de l'information » effectué par les neurones implique une corrélation dans l'activité de ceux-ci : c'est l'hypothèse du liage perceptif par synchronisation.

Tri des potentiels d'action et analyse de leurs séquences

(2)

Si le travail méthodologique qui va être présenté a été effectué dans le cadre qui vient d'être évoqué, les méthodes développées peuvent s'appliquer dans des contextes plus larges, comme par exemple :

- ▶ la quantification « rapide » de l'effet de drogues sur des cellules en culture ;
- ▶ le pilotage de « bras artificiels » par l'activité enregistrés par des matrices d'électrodes implantées chez un primate ou un patient.

Où en est-on ?

Introduction

Tri des potentiels d'action dans les cas simples

Analyse des séquences de potentiels d'action

Analyse des mesures de $[Ca^{2+}]$ par sondes fluorescentes

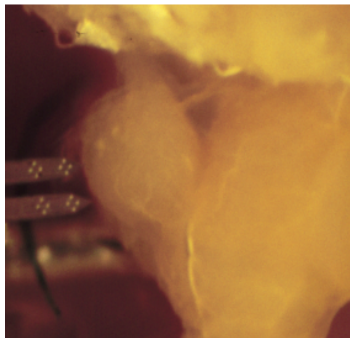
Conclusions et perspectives

Le tri des potentiels d'action : contexte expérimental

- ▶ ce travail a été généré par des études sur le système olfactif des insectes : criquet (*Schistocerca americana*), abeille (*Apis mellifera*) et blatte (*Periplaneta americana*) ;
- ▶ il a été effectué avec Ofer Mazor (étudiant en thèse à Caltech) et Antoine Chaffiol (étudiant en thèse au laboratoire de physiologie cérébrale).
- ▶ une extension très importante qui ne sera pas exposée aujourd'hui, faute de temps, a été générée par une étude des décharges des cellules de Purkinje dans des tranches de cortex cérébelleux de rat ; ce travail a été effectué avec :
 - ▶ Matthieu Delescluse, étudiant en thèse au laboratoire de physiologie cérébrale,
 - ▶ Jean Diebolt, DR CNRS, laboratoire d'analyse et de mathématiques appliquées, Université de Marne La Vallée,
 - ▶ Pascal Viot, DR CNRS (à présent Professeur), laboratoire de physique théorique de la matière condensée, Université Pierre et Marie Curie.

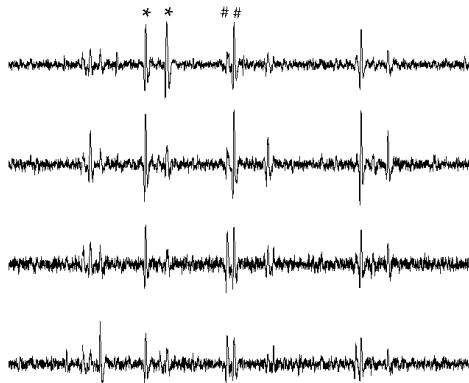
Pourquoi doit-on trier les potentiels d'action ?

Vue de l'extérieur, l'activité des neurones se manifeste par l'émission d'impulsions électriques très brèves : **les potentiels d'action** (PAs).



À gauche, le cerveau — d'un criquet — et la sonde d'enregistrement qui comporte 16 électrodes (les points brillants). La largeur d'une branche de la sonde est de $80 \mu m$. A droite, 1 sec d'enregistrement sur 4 électrodes. Les pics sont des potentiels d'action.

Pourquoi des *tetrodes*?



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

Autre illustration de la relation distance / amplitude

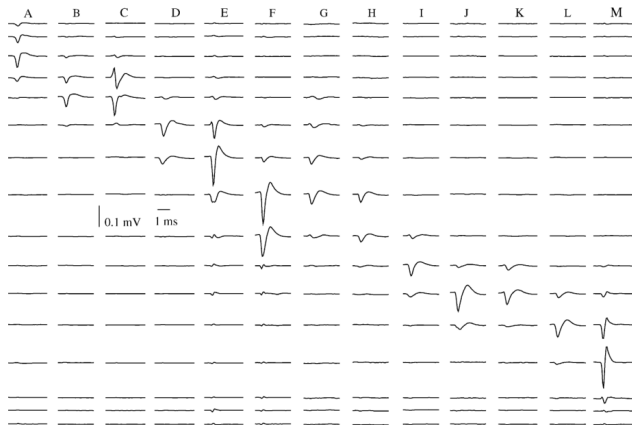


Fig. 2. Mean waveforms of the 13 spike clusters on 16 channels. Inter electrode distances were 100 μm . Each line represents a 2.5 ms long period.

Résultat du tri des PAs (cas d'une sonde avec « 16 sites linéaires ») dans le cortex auditif primaire du chat (Somogyvári, Zalányi, Ulbert et Érdi, 2005).

Le tri des PAs en une figure

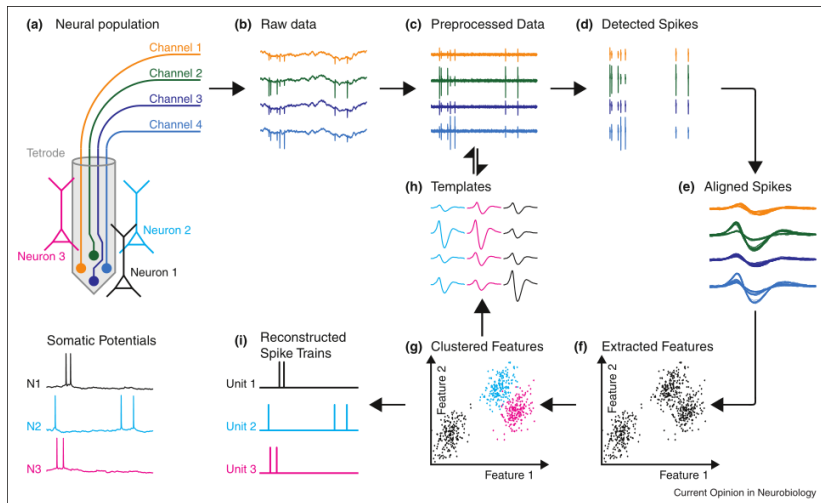


Fig. 1 de Einevoll, Franke, Hagen, Pouzat et Harris (2012).

La cas simple (1)

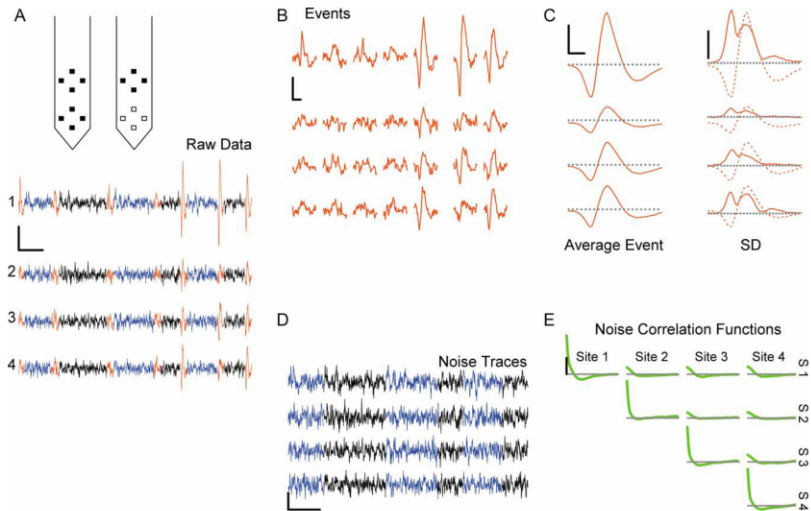


Fig. 1 de Pouzat, Mazor et Laurent (2002).

Le cas simple (2)

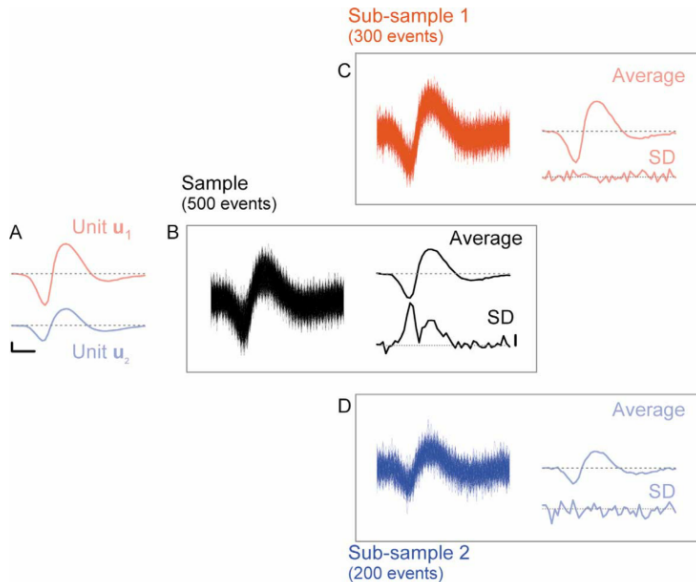
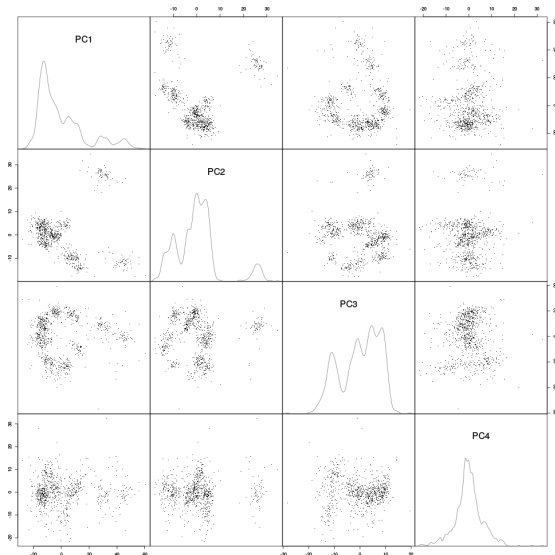


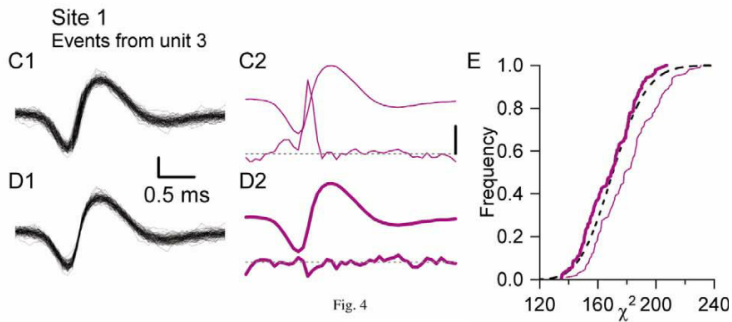
Fig. 2 de Pouzat, Mazor et Laurent (2002).

Le cas simple (3)



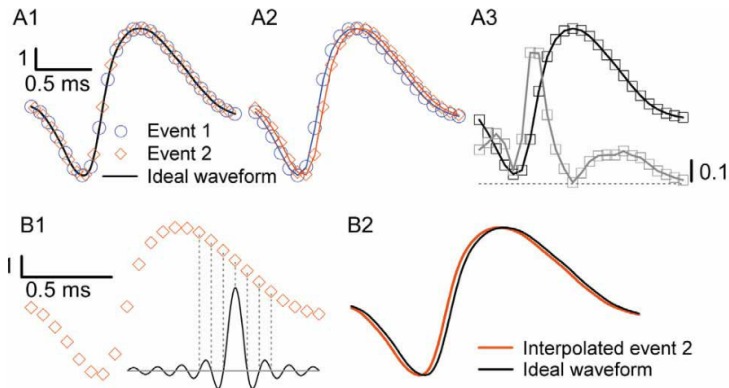
Matrice de graphes en nuages de points construite à partir des 4 premières composantes principales.

Trop simple? (1)



Partie de la figure 4 de Pouzat, Mazor et Laurent (2002).

Trop simple? (2)



Partie de la figure 4 de Pouzat, Mazor et Laurent (2002).

Une estimation efficace de la gigue électronique (1)

- ▶ soit $g(t)$ l'amplitude observée au temps t dans une fenêtre ;
- ▶ alors $g(t) = f(t + \delta) + \epsilon(t)$, où δ est notre **gigue induite par l'échantillonnage et par le bruit**, f est notre « forme mère », $\epsilon(t)$ est notre bruit ;
- ▶ nous modélisons δ comme observation d'une variable aléatoire Δ 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 δ proche de 0 :

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

Une estimation efficace de la gigue électronique (2)

- ▶ par hypothèse sur Δ 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 δ : $Eg(t) = f(t)$.
- ▶ cela signifie que nos « formes mères », estimées par $\hat{f}(t) = Eg(t)$, sont **pratiquement non biaisées si σ_{Δ}^2 ou f'' sont assez petites** ;
- ▶ Au premier ordre en δ nous avons pour la 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 δ) :

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

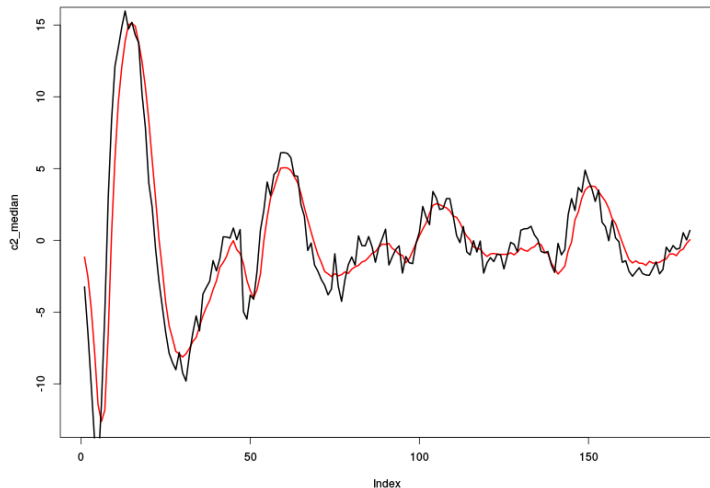
Une estimation efficace de la gigue électronique (3)

- ▶ nous avons en fait : $g(t_i) = f(t_i + \delta) + \epsilon_i$, où les ϵ_i sont des observations de variables aléatoires (corrélées) d'espérance nulle et de variance finie ;
- ▶ au second ordre nous avons :
 $g(t_i) \approx f(t_i) + \delta f'(t_i) + \frac{\delta^2}{2} f''(t_i) + \epsilon_i$;
- ▶ nous estimons δ par $\hat{\delta}$ défini comme l'argument qui minimise le critère des moindres carrés :

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

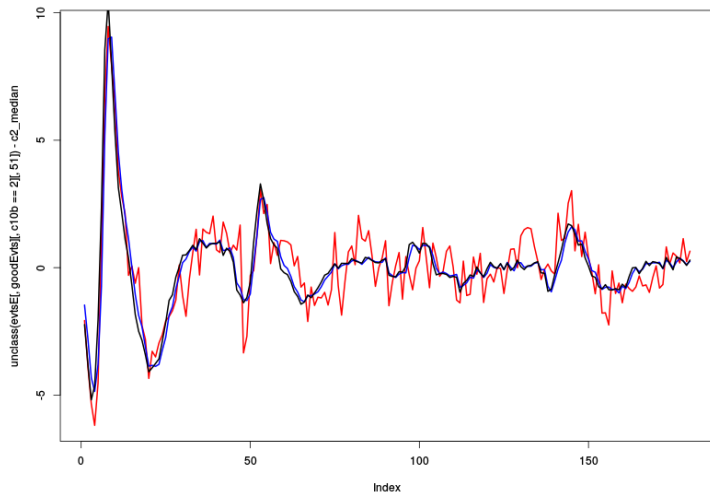
- ▶ le calcul est très rapide dès lors que des estimations de f , f' and f'' sont disponibles ;
- ▶ il est (beaucoup) plus sûr d'utiliser la médiane, plutôt que la moyenne pour ces dernières.

Illustration (1)



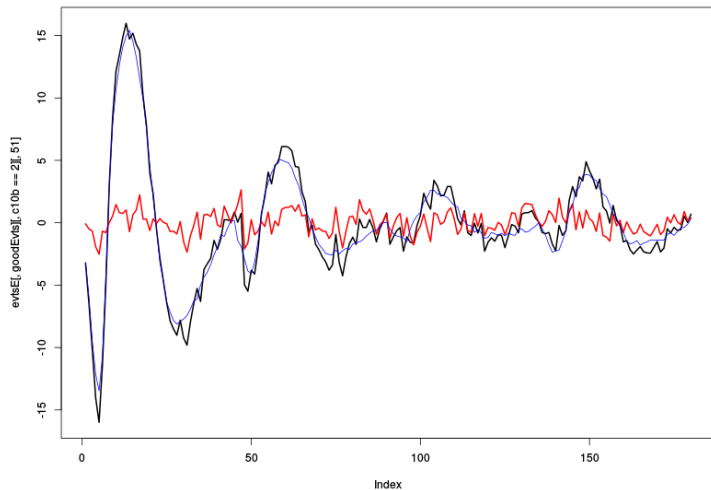
Un événement (en noire) et « sa forme mère » (en rouge).

Illustration (2)



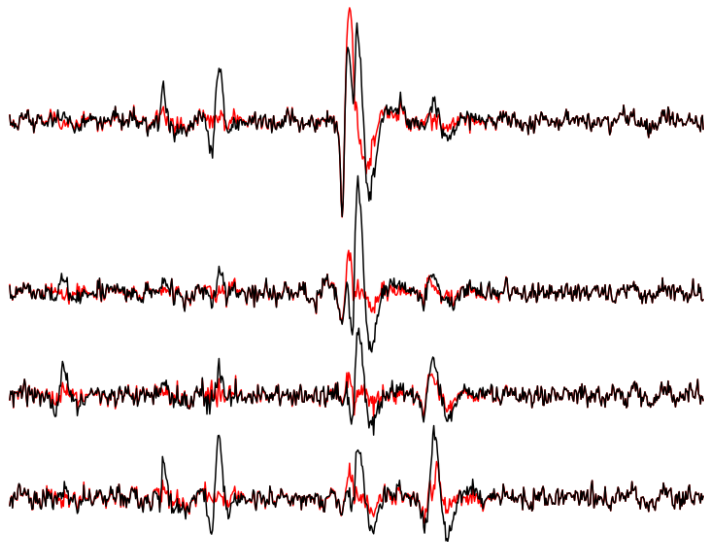
En rouge, l'événement dont la forme mère a été soustraite ; en bleu, $1,5 \times f'$; en noir, $1,5 \times f' + 1,5^2/2 \times f''$.

Illustration (3)



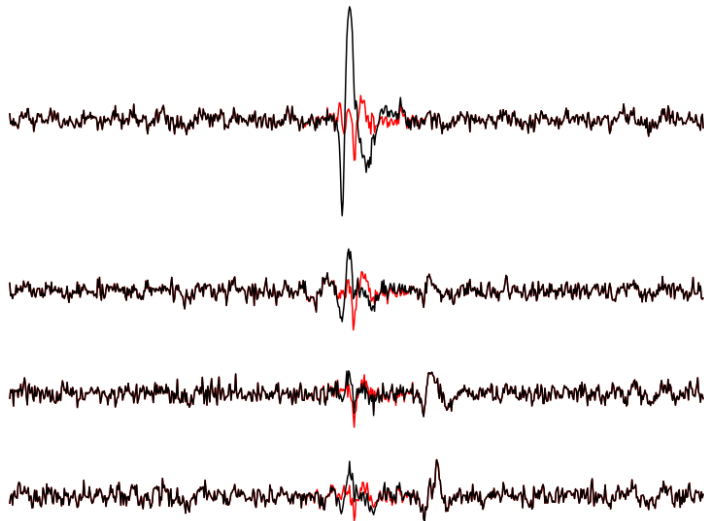
En noire, l'événement (g) ; en bleu, $f + \hat{\delta}f' + \hat{\delta}^2/2f''$; en rouge, $g - f - \hat{\delta}f' - \hat{\delta}^2/2f''$.

Résolution des superpositions par « épiluchage » (1)



En noir, les données brutes ; en rouge, après le premier
« épiluchage ».

Résolution des superpositions par « épiluchage » (2)



En noir, après le premier « épiluchage » ; en rouge, après le deuxième.

Où en est-on ?

Introduction

Tri des potentiels d'action dans les cas simples

Analyse des séquences de potentiels d'action

Analyse des mesures de $[Ca^{2+}]$ par sondes fluorescentes

Conclusions et perspectives

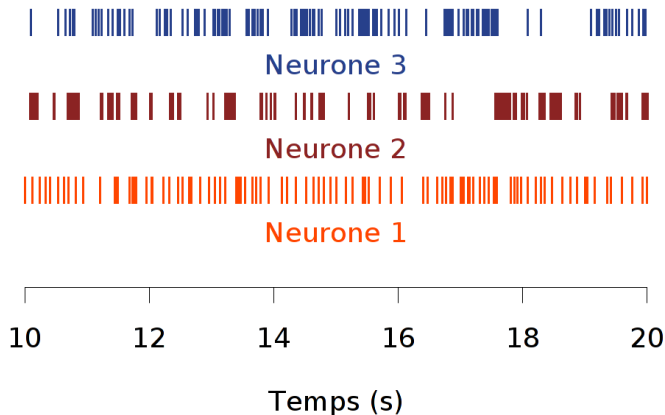
Contexte

Ce travail a été effectué avec :

- ▶ Antoine Chaffiol, étudiant en thèse ;
- ▶ Pierre Bohec, étudiant du M2 Systèmes Complexes de Jussieu ;
- ▶ Chong Gu, Professeur de statistique à l'Université de Purdue.

Exemple de séquences de potentiels d'action

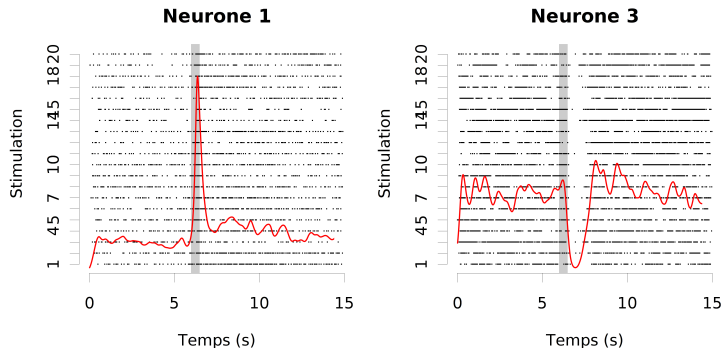
Le tri des PAs effectués, nous obtenons le **graphe en raster** représentant les séquences ou trains de PAs (données enregistrées et PAs triés par Antoine Chaffiol) :



Pourquoi et comment modéliser les trains de PAs ?

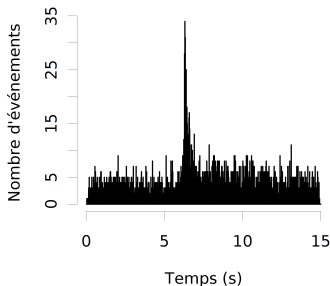
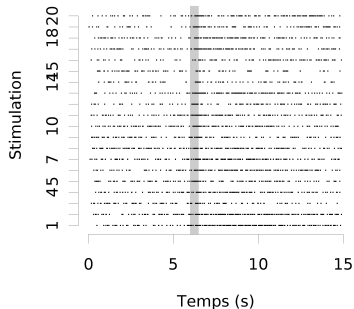
- ▶ une hypothèse de travail centrale en neurosciences est que les temps d'apparition des PAs, par opposition à leurs formes, sont le seul support de transmission de l'information entre régions du cerveau (Adrian et Zotterman, 1926) ;
- ▶ cette hypothèse légitime l'étude des trains de PAs en tant que 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**).

Régime non-homogène : réponses aux odeurs (1)



20 réponses à des stimulations de 500 ms avec du citronnellal. En rouge : une estimation « lisse » de la fréquence de décharge moyennée sur les stimulations (échelles identiques sur les deux graphes). La fréquence spontanée du neurone 1 est de 6.6 Hz et celle du neurone 3 est de 16.8 Hz.

Régime non-homogène : estimation de la fréquence moyennée (2)



On passe des données brutes à un histogramme construit avec un petit pas de temps (25 ms) ; ce qui nous donne, à une normalisation près, un estimateur peu biaisé mais de forte variance de la fréquence moyennée.

Régime non-homogène : estimation de la fréquence moyennée (3)

- ▶ On modélise le « processus moyenné » comme un **processus de Poisson inhomogène** d'intensité $\lambda(t)$.
- ▶ L'histogramme que nous avons construit peut alors être considéré comme l'observation d'une collection de variables aléatoires de Poisson, $\{Y_1, \dots, Y_k\}$, de paramètres :

$$N_s \int_{t_i - \delta/2}^{t_i + \delta/2} \lambda(u) du \approx N_s \lambda(t_i) \delta, \quad i = 1, \dots, k,$$

où t_i est le temps central d'une classe, δ est la largeur des classes, N_s est le nombre de stimulations et k est le nombre de classes.

Régime non-homogène : estimation de la fréquence moyennée (4)

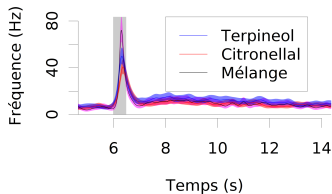
- ▶ Nous estimons en fait directement : $\eta() = \log(N_s \lambda() \delta)$ par régression de Poisson.
- ▶ Comme nous ne voulons pas faire d'hypothèses « fortes » sur η nous choisissons comme estimateur, $\hat{\eta}$, la fonction qui minimise la vraisemblance pénalisée :

$$-\sum_{i=1}^k \left(y_i \eta(t_i) - \exp(\eta(t_i)) \right) + \rho \int \left(\frac{d^2 \eta(u)}{dt^2} \right)^2 du.$$

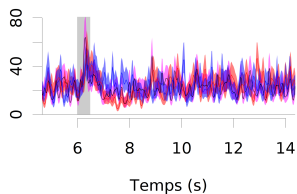
- ▶ Le paramètre de lissage, ρ , est obtenu en minimisant le **critère de validation croisée généralisée**.
- ▶ Tous les calculs sont effectués avec le paquet gss (**general smoothing splines**) développé par Chong Gu pour le logiciel R.

Régime non-homogène : estimation de la fréquence moyennée (5)

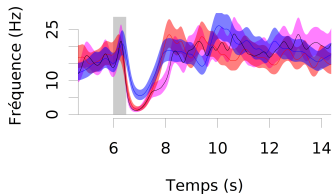
Neurone 1



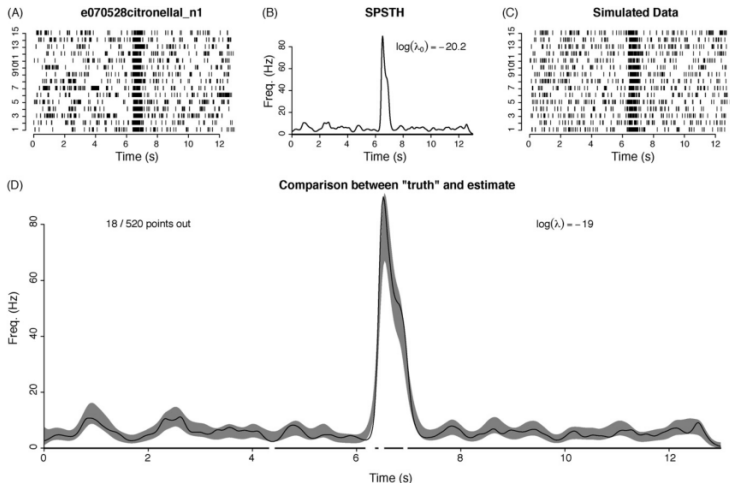
Neurone 2



Neurone 3



Régime non-homogène : estimation de la fréquence moyennée (6)



Les intervalles de confiance ont un sens ! Figure 5 de Pouzat et Chaffiol (2009).

Test d'adéquation

- ▶ nous disposons d'un processus de comptage observé et d'une estimation $\hat{\lambda}$ de l'intensité λ de notre processus de Poisson inhomogène, une question se pose : « Les deux sont-ils compatibles ? » ;
- ▶ nous disposons également de plusieurs $\hat{\lambda}$ pour un même neurone dans plusieurs conditions (stimulations avec le terpéneol, le citronellal et le mélange des deux) et on se demande si les réponses sont différentes, ou, formulé différemment, on se demande si les différences observées entre les $\hat{\lambda}$ ($\hat{\lambda}_{\text{terpéneol}}$, $\hat{\lambda}_{\text{citronellal}}$, $\hat{\lambda}_{\text{mélange}}$) sont significatives ;
- ▶ nous avons développé un test pour répondre à ces questions.

Intensité et intensité cumulée

Nous associons à λ , l'**intensité cumulée** :

$$\Lambda(t) \equiv \int_0^t \lambda(u) du.$$

Il n'est alors pas trop difficile de montrer que :

- ▶ **si notre modèle est bon**, la loi des intervalles entre PAS successifs après transformation du temps :

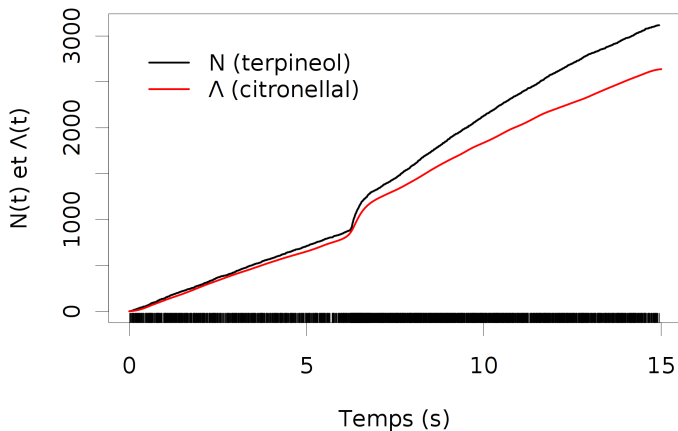
$$\{t_1, \dots, t_n\} \rightarrow \{\Lambda(t_1) \equiv \Lambda_1, \dots, \Lambda(t_n) \equiv \Lambda_n\}$$

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

- ▶ le processus ponctuel observé $\{\Lambda_1, \dots, \Lambda_n\}$ est donc l'observation d'un **processus de Poisson homogène de paramètre 1**.

Illustration de la transformation du temps (1)

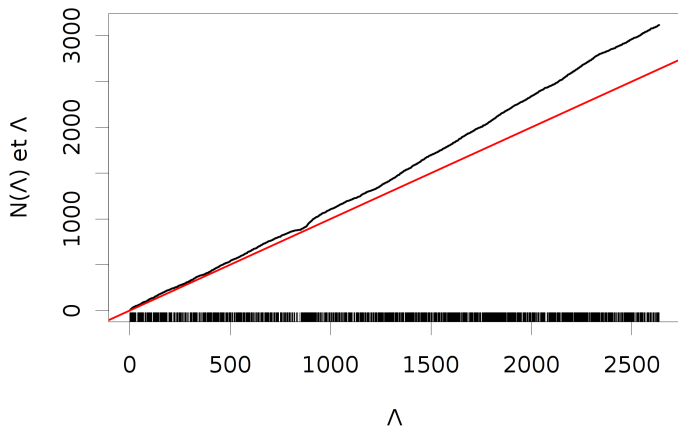
Neurone 1



En noir, le processus de comptage observé en terpeneol (20 réponses agrégées) ; en rouge l'intégrale de $\hat{\Lambda}$ (obtenue avec le citronellal) multipliée par le nombre de stimulations (20).

Illustration de la transformation du temps (2)

Neurone 1



Les tests d'Ogata

- ▶ si, pour un bon modèle, la séquence « transformée » de PAs, $\{\Lambda_1, \dots, \Lambda_n\}$, 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 $\{\Lambda_1, \dots, \Lambda_n\}$ à un Poisson homogène ;
- ▶ c'est ce qu'a proposé Yoshihiko Ogata en 1988 (Statistical models for earthquake occurrences and residual analysis for point processes, Journal of the American Statistical Association, 83 : 9-27) ;
- ▶ une observation basée sur de nombreuses simulations nous a néanmoins amené à proposer un test basé sur le théorème de Donsker...

Théorème de Donsker

Théorème

Soit X_1, X_2, \dots une suite de variables aléatoires IID telles que $EX_i = 0$ et $EX_i^2 = 1$, alors la suite de processus :

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{i=0}^{\lfloor nt \rfloor} X_i, \quad 0 \leq t \leq 1, \quad X_0 = 0$$

converge en loi vers le mouvement brownien canonique.

Preuve

Pour une preuve du théorème de Donsker voir :

- ▶ R Durrett (2009) *Probability : Theory and Examples*. CUP. Sec. 7.6, pp 323-329 ;
- ▶ P Billingsley (1999) *Convergence of Probability Measures*. Wiley. p 121.

Application du théorème de Donsker

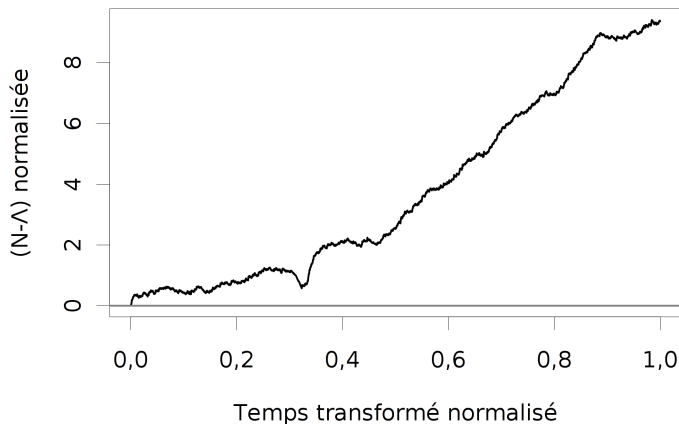
- ▶ les X_i de l'énoncé du théorème peuvent être donnés par : $N(k+1)-N(k)-1$, où $k \in \mathbb{N}$ et où le compteur est défini sur l'échelle de temps transformé (loi de Poisson de paramètre 1 suivie par le nombre d'événements générés par un processus de Poisson de paramètre 1 sur une durée de 1).
- ▶ le graphe de :

$$\frac{N(t) - \Lambda(t)}{\sqrt{[\Lambda(t_{max})]}} \text{ en fct. de } \frac{\Lambda(t)}{[\Lambda(t_{max})]},$$

où $0 \leq t \leq t_{max}$, va ainsi « ressembler de plus en plus » au chemin d'un brownien canonique sur $[0,1]$.

Illustration

Neurone 1



$(N_{terpi}(t) - \hat{\Lambda}_{citron}(t)) / \sqrt{[\hat{\Lambda}_{citron}(t_{max})]}$ en fonction de $\hat{\Lambda}_{citron}(t) / [\hat{\Lambda}_{citron}(t_{max})]$.

Alors quoi de neuf?

- ▶ notre proposition — qui semble être originale — est d'utiliser la convergence en loi du processus de comptage « compensé » et exprimé sur une nouvelle échelle de temps vers un mouvement brownien canonique, **quand le modèle est bon**, pour tester si notre modèle est effectivement bon ;
- ▶ pour cela il nous faut être capables de reconnaître un mouvement brownien canonique sur $[0,1]$ lorsque nous en voyons un...

Région de prédiction de surface minimale

- ▶ dans un article (pas franchement trivial) Kendall, Marin et Robert (Statistics and Computing, 17 : 1-10, 2007) ont montré qu'un domaine dont la frontière « supérieure » est donnée par :

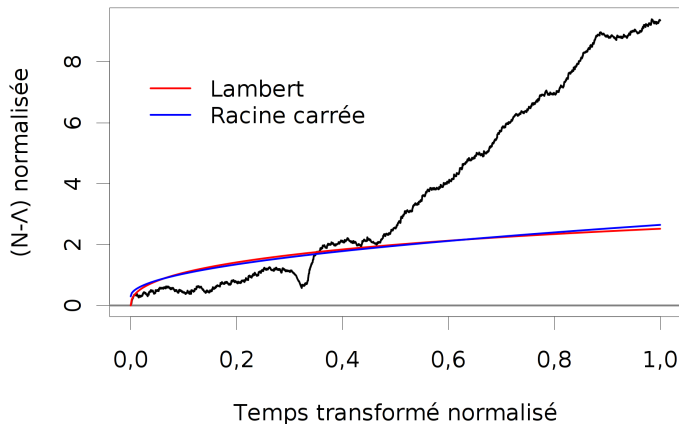
$$u^*(t) \equiv \sqrt{-W_{-1}(-(\kappa t)^2)} \sqrt{t}, \quad \text{pour } \kappa t \leq 1/\sqrt{e}$$

où W_{-1} est la « branche réelle secondaire » de la fonction W de Lambert, définie comme la solution de : $W(z) \exp W(z) = z$, contient **la totalité d'un brownien observé** avec une probabilité p contrôlée par le paramètre κ et que **ce domaine est de surface minimale** ;

- ▶ ils ont aussi montré qu'un domaine dont la frontière supérieure est donnée par $f(t) = a + b\sqrt{t}$, (où a et b sont ajustés pour avoir la proba. désirée) est presque de surface minimale ;
- ▶ un article de Loader et Deely (J. Statist. Comput. Simulation, 27 : 95-105, 1987) donne un algorithme très efficace pour ajuster κ ou a et b ;
- ▶ notre logiciel STAR (*Spike Train Analysis with R*) rend tout cela directement accessible dans R.

Illustration : régions à 95 % (1)

Neurone 1



Attention : ici c'est $\hat{\Lambda}$ et pas Λ qui est utilisée. La probabilité de faux positifs va donc être plus grande que 5 % et il faudrait « calibrer » avec des simulations.

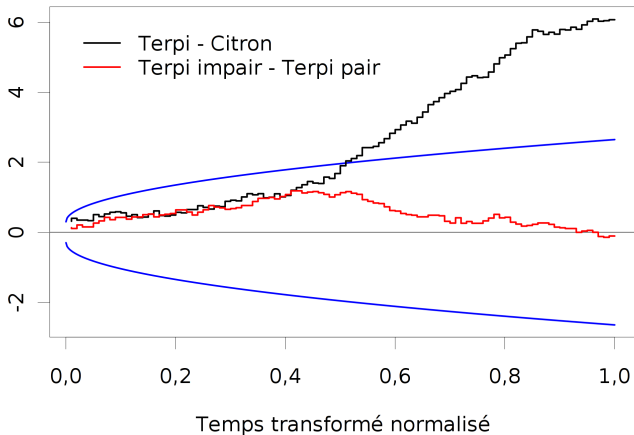
Illustration : régions à 95 % (2)

On contourne le problème précédent comme suit :

- ▶ on suppose, hypothèse nulle, que $\lambda(t)$ est identique dans les deux conditions (terpineol et citronellal) ;
- ▶ on découpe l'intervalle des temps [0 s, 15 s] en 100 classes de longueur 150 ms ;
- ▶ on compte combien d'événements tombent dans chaque classe pour chacune des 2 conditions (on construit deux histogrammes) ;
- ▶ on forme la différence des deux conditions pour chaque classe et on divise cette différence par la racine carrée de la somme des comptes dans chaque classe (approximation de la loi des comptes par une loi de Poisson) ;
- ▶ on obtient ainsi les X_i de Donsker.

Illustration : régions à 95 % (3)

Neurone 1



Où en est-on ?

Introduction

Tri des potentiels d'action dans les cas simples

Analyse des séquences de potentiels d'action

Analyse des mesures de $[Ca^{2+}]$ par sondes fluorescentes

Conclusions et perspectives

Analyse des mesures de $[Ca^{2+}]$ par sondes fluorescentes

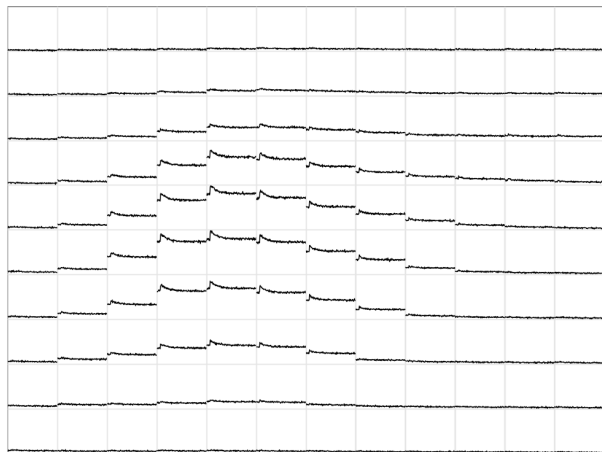
Ce travail a été effectué en collaboration avec :

- ▶ Peter Kloppenburg et Andreas Pippow de la faculté de zoologie de l'Université de Cologne ;
- ▶ Sébastien Joucla, stagiaire post-doctoral au laboratoire.

Les données qui vont être présentées ont été enregistrées dans le noyau arqué de l'hypothalamus de souris. Les souris ont été modifiées génétiquement dans le laboratoire de Jens Brüning (Institut de génétique, Université de Cologne) de sorte que les neurones sécrétant un peptide, la pro-opiomélanocortine ou POMC, expriment la GFP (*Green Fluorescent Protein*) et soient ainsi facilement identifiables sur les tranches par les physiologistes.

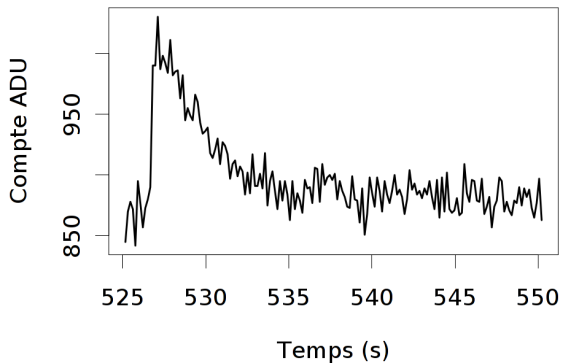
C'est la dynamique de la concentration des ions calcium ($[Ca^{2+}]$) qui intéresse les physiologistes. Ils mesurent cette dernière au moyen d'une sonde fluorescente, le Fura-2, qui se lie réversiblement au calcium et dont l'absorption change suivant qu'elle est liée ou non au calcium.

La variabilité inhérente aux mesures de fluorescence (1)



Données brutes d'un capteur CCD (comptes ADU) suivant l'excitation du Fura-2 à 340 nm. Chaque carré correspond à un pixel. 25.05 s de données sont montrées. L'échelle est la même dans chaque pixel. 12x10 pixels / 60x80 pixels sont montrés.

La variabilité inhérente aux mesures de fluorescence (2)



Un des pixels centraux de la figure précédente.

Les questions (1)

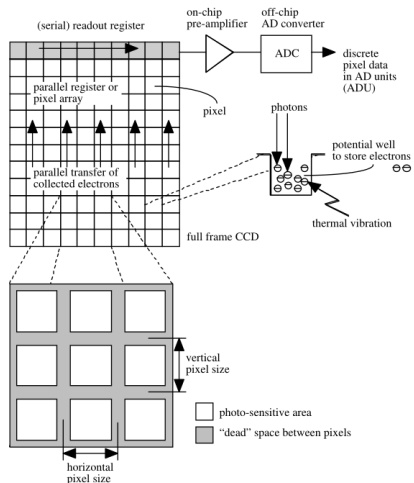
À partir d'un jeu de données comme celui que nous venons de présenter, les physiologistes vont souhaiter estimer des paramètres comme :

- ▶ l'amplitude au pic ;
- ▶ la ou les constantes de temps de décroissance ;
- ▶ le niveau de base ;
- ▶ le décours temporel complet — qui est une fonction.

Les questions (2)

- ▶ la variabilité inhérente aux mesures de fluorescence implique que nos estimations de paramètres seront nécessairement effectuées avec une précision finie ;
- ▶ nous souhaitons donc développer des méthodes qui, en plus d'une estimation des paramètres, fournissent une évaluation de la précision de ces estimations ;
- ▶ sans cela, il est impossible de comparer les résultats de différentes expériences (entre laboratoires, comme au sein d'un même laboratoire) ;
- ▶ pour progresser, nous allons avoir besoin d'un modèle précis de la variabilité de nos mesures de fluorescence.

Schéma de fonctionnement d'un capteur CCD



Source : L. van Vliet et col. (1998) Digital Fluorescence Imaging Using Cooled CCD Array Cameras (figure 3).

Les principales sources de « bruit » d'un capteur CCD

- ▶ le **bruit photonique** ou « bruit de grenaille » (*shot noise* en anglais) vient du fait que mesurer une intensité de fluorescence, λ , implique un **comptage de photons** dont la loi est une loi de Poisson – à part « changer les lois de la Physique », il n'y a rien à faire pour diminuer ce « bruit » – :

$$\Pr\{N = n\} = \frac{\lambda^n}{n!} \exp -\lambda, \quad n = 0, 1, \dots, \quad \lambda > 0;$$

- ▶ le **bruit de lecture** vient de la conversion d'un nombre de photo-électrons en tension équivalente, il suit une loi normale dont la variance est indépendante de la moyenne (tant qu'on n'effectue pas de lecture à trop haute fréquence).

Un modèle simple de CCD

- ▶ lorsqu'un grand nombre de photons est détecté, c.-à-d. lorsque λ est grand :

$$\Pr\{N = n\} = \frac{\lambda^n}{n!} \exp -\lambda \approx \mathcal{N}(\lambda, \lambda),$$

ce que nous écrirons : $N \sim \lambda + \sqrt{\lambda} \epsilon$ ($\epsilon \sim \mathcal{N}(0, 1)$);

- ▶ le bruit de lecture (L) s'ajoute indépendamment et suit une loi normale de moyenne nulle et de variance σ_L^2 :

$$M \equiv N + L \sim \lambda + \sqrt{\lambda + \sigma_L^2} \epsilon;$$

- ▶ un gain $G < 1$ est enfin appliqué pour donner :

$$Y \equiv G \cdot M \approx G\lambda + \sqrt{G^2(\lambda + \sigma_L^2)} \epsilon.$$

Calibration d'un capteur CCD (1)

- ▶ notre problème devient donc : Comment tester $Y \sim G\lambda + \sqrt{G^2(\lambda + \sigma_L^2)}\epsilon$? Où comment faire varier λ ?
- ▶ un examen du système d'imagerie nous permet d'écrire, **pour un pixel (i,j)** :

$$Y_{i,j} \sim Gp_{i,j}\tau + \sqrt{G^2(p_{i,j}\tau + \sigma_L^2)}\epsilon_{i,j},$$

où τ est le temps d'exposition et $p_{i,j}$ un paramètre spécifique au pixel ;

- ▶ nous donc avons demandé à nos collaborateurs P. Kloppenburg et A. Pippow de :
 - ▶ choisir 10 temps d'expositions,
 - ▶ pour chacun des 10 temps effectuer 100 expositions successives,
 - ▶ pour chacune des 10 x 100 expositions, enregistrer la valeur y_{ij} prise par la v.a. Y_{ij} au pixel i, j du capteur CCD.

Calibration d'un capteur CCD (2)

- ▶ si notre modèle est correct alors pour chaque pixel i, j nous devons avoir, pour un temps d'exposition donné, une moyenne empirique :

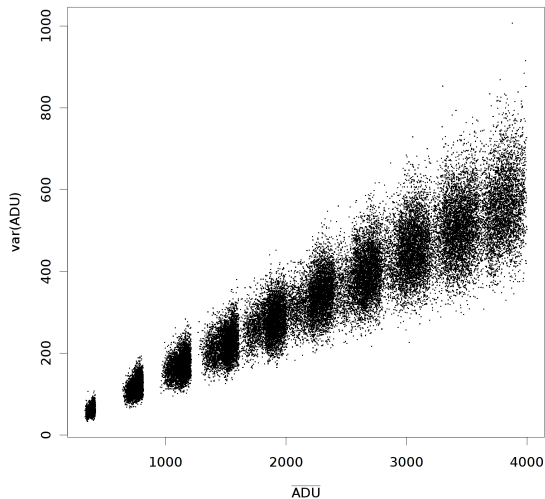
$$\bar{y}_{i,j} = \frac{1}{100} \sum_{k=1}^{100} y_{i,j,k} \approx G p_{i,j} \tau ;$$

- ▶ un variance empirique :

$$S_{i,j}^2 = \frac{1}{99} \sum_{k=1}^{100} (y_{i,j,k} - \bar{y}_{i,j})^2 \approx G^2 (p_{i,j} \tau + \sigma_L^2) ;$$

- ▶ et le graphe de $S_{i,j}^2$ en fonction de $\bar{y}_{i,j}$ devrait être une droite de pente G et d'ordonnée à l'origine $G^2 \sigma_L^2$.

Calibration d'une CCD (3)



La variance en fonction de la moyenne.

Stabilisation de la variance (1) : propagation des erreurs

- ▶ Soient deux variables aléatoires Y et Z telles que :
- ▶ $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ ou $Y = \mu_Y + \sigma_Y \epsilon$ ($\epsilon \sim \mathcal{N}(0, 1)$) ;
- ▶ $Z = f(Y)$, f différentiable.
- ▶ Un développement de Taylor au premier ordre donne :

$$\begin{aligned} Z &= f(\mu_Y + \sigma_Y \epsilon) \\ &\approx f(\mu_Y) + \sigma_Y \epsilon \frac{df}{dY}(\mu_Y) \end{aligned}$$

- ▶ $EZ \approx f(\mu_Y) = f(EY)$
- ▶ $\text{Var}Z \equiv E[(Z - EZ)^2] \approx \sigma_Y^2 \frac{df}{dY}^2(\mu_Y)$
- ▶ $Z \approx f(\mu_Y) + \sigma_Y \left| \frac{df}{dY}(\mu_Y) \right| \epsilon$

Stabilisation de la variance (2)

- ▶ notre modèle de CCD nous donne (pour un pixel donné) :

$$Y \sim G\lambda + \sqrt{G^2(\lambda + \sigma_R^2)}\epsilon = \mu_Y + \sqrt{G\mu_Y + G^2\sigma_R^2}\epsilon;$$

- ▶ alors si $Z = f(Y)$, il vient :

$$Z \approx f(\mu_Y) + |f'(\mu_Y)| G\sqrt{\mu_Y/G + \sigma_R^2}\epsilon;$$

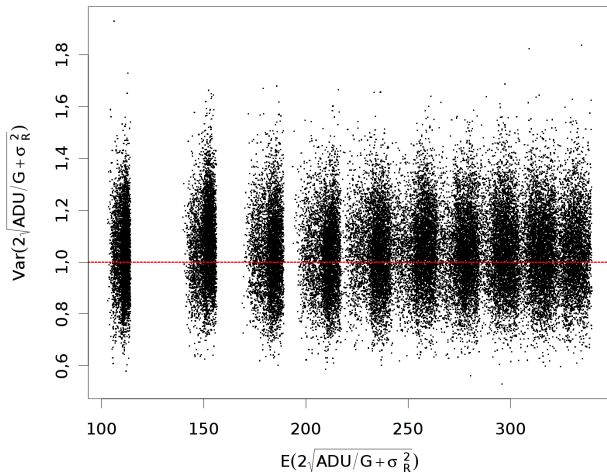
- ▶ qu'arrive-t-il si nous choisissons : $f(x) = 2\sqrt{x/G + \sigma_R^2}$?
- ▶ nous obtenons :

$$f'(x) = \frac{1}{G\sqrt{x/G + \sigma_R^2}};$$

- ▶ soit :

$$Z \approx 2\sqrt{\mu_Y/G + \sigma_R^2} + \epsilon.$$

Stabilisation de la variance (2) : exemple



Stabilisation de la variance appliquée aux données de calibration.

Retour au point de départ



Tout le capteur (60x80 pixels) est cette fois montré.

Détermination rapide de ROI (*region of interest*) (1)

- ▶ après avoir stabilisé le variance : $Z_{i,j,k} = 2 \sqrt{ADU_{i,j,k} / G + \sigma_R^2}$, la variance en chaque pixel (i, j) à chaque temps, k , devrait être égale à 1 ;
- ▶ si un pixel ne contient pas de signal dépendant du temps, la statistique suivante :

$$RSS_{i,j} \equiv \sum_{k=1}^K (Z_{i,j,k} - \bar{Z}_{i,j})^2 \quad \text{où} \quad \bar{Z}_{i,j} \equiv \frac{1}{K} \sum_{k=1}^K Z_{i,j,k}$$

devrait suivre une loi du χ^2 avec $K - 1$ degrés de liberté ;

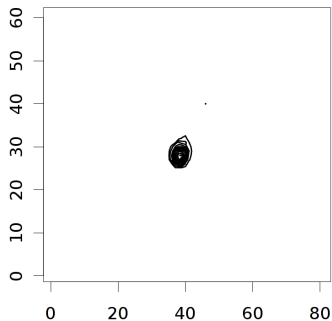
- ▶ nous pouvons donc calculer la complémentaire de la fonction de répartition théorique :

$$1 - F_{\chi_{K-1}^2}(RSS_{i,j})$$

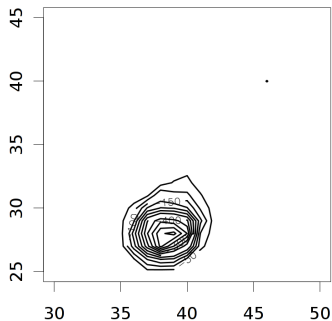
et chercher les valeurs très petites.

Détermination rapide de ROI (2)

Capteur complet



Zoom



Graphes en lignes de niveau de $\log\left(1 - F_{\chi_{K-1}^2}(\text{RSS}_{i,j})\right)$

Estimation ponctuelle du décours temporel (1)

- ▶ nous décidons d'être prudents et groupons les pixels tels que :
 $\log\left(1 - F_{\chi_{K-1}^2}(RSS)\right) \leq -300$;
- ▶ il nous reste 12 pixels ;
- ▶ nous modélisons l'intensité de chaque pixel par :

$$S_{i,j}(t) = \phi_{i,j}f(t) + b,$$

où $f(t)$ est le décours temporel commun à chaque pixel de la ROI, $\phi_{i,j}$ est un paramètre spécifique à chaque pixel et b est l'auto-fluorescence supposée commune à chaque pixel.

- ▶ le temps t est en fait une variable discrète, $t = \delta k$ ($\delta = 150$ ms) et nous cherchons une estimation ponctuelle : $\{f_1, f_2, \dots, f_K\}$ ($K = 168$) où $f_k = f(\delta k)$.
- ▶ nous avons donc $12 (\phi_{i,j}) + 168 (f_k) + 1 (b) = 181$ paramètres pour $12 \times 168 = 2016$ observations.

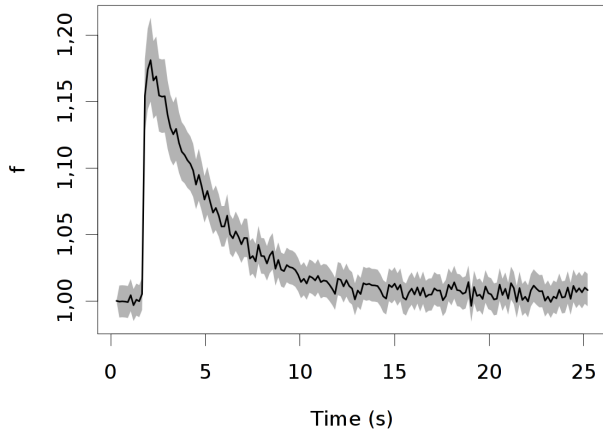
Estimation ponctuelle du décours temporel (2)

- ▶ nous devons introduire une contrainte car notre modèle ($S_{i,j,k} = \phi_{i,j}f_k + b$) n'est pas identifiable ; nous pouvons en effet multiplier tous les $\phi_{i,j}$ par 2 et diviser tous les f_k par 2 et avoir la même prédiction ;
- ▶ nous choisissons de fixer $f_1 = 1$; ainsi notre estimation ponctuelle est liée à ce que les physiologistes font habituellement avec ce type de données puisqu'ils travaillent avec :

$$\Delta S(t)/S(1) = \frac{S(t) - S(1)}{S(1)} = f(t) - 1 + \text{bruit}.$$

- ▶ remarquez qu'aucune estimation indépendante de l'auto-fluorescence n'est utilisée dans notre approche.

Estimation ponctuelle du décours temporel (3)



Remarquez les intervalles de confiance (ici à 95%).

Où en est-on ?

Introduction

Tri des potentiels d'action dans les cas simples

Analyse des séquences de potentiels d'action

Analyse des mesures de $[Ca^{2+}]$ par sondes fluorescentes

Conclusions et perspectives

La suite : tri des PAs

- ▶ finir le recodage des méthodes en Python pour avoir une meilleur intégration avec OpenElectrophy (S. Garcia) ;
- ▶ rajouter les méthodes MCMC et les rendre plus « flexibles » ;
- ▶ mettre en œuvre des méthodes type *sequential importance sampling* ;
- ▶ intégrer des modèles plus récents d'intensité (statistique de décharges) au niveau du tri des PAs.

La suite : analyse des séquence de PAs

- ▶ faire une comparaison systématique des intensités (en régime stationnaire comme en régime stimulé) sur mes deux gros jeux de données : criquet et blatte ;
- ▶ intégrer les processus de Hawkes (Reynaud-Bouret, Tuleau-Malot, Rivoirard, Grammont) à STAR ;
- ▶ travailler sur les tests ;
- ▶ améliorer les temps de calculs (avec Python + Cython ou Clojure).

La suite : étude de la dynamique de la $[Ca^{2+}]$

- ▶ tester les modèles basés sur équations différentielles ordinaires considérés habituellement ;
- ▶ aller vers des modèles prenant en compte la diffusion du calcium -> EDP ;
- ▶ améliorer les sélections de ROIs avec la méthode d'Yves Rozenholc ;
- ▶ utiliser des versions stochastiques (processus de « naissance et mort ») des équations différentielles lorsque le volume du neurite observé n'autorise plus l'emploi du concept de concentration (épine dendritique) ;
- ▶ intégrer les méthodes à un outil type imageJ (via Jython ou Clojure).

Je remercie

- ▶ les membres du jury;
- ▶ mes étudiants et collaborateurs cités au cours de cet exposé;
- ▶ les membres de mes laboratoires successifs : Caltech, physiologie cérébrale, MAP5;
- ▶ leurs directeurs : Gilles Laurent, Alain Marty et Annie Raoul;
- ▶ le CNRS pour la grande liberté dont j'ai bénéficié;
- ▶ les développeurs de R, Python, GGobi, MixMod, emacs + org, gcc, etc.