Des données neurophysiologiques aux modèles statistiques et aux développement de logiciels : retour d'expérience sur l'analyse des séquences de potentiels d'action et l'analyse de données de fluorescence liée au calcium.

Christophe Pouzat

Mathématiques Appliquées à Paris 5 (MAP5)

Université Paris-Descartes et CNRS UMR 8145

christophe.pouzat@parisdescartes.fr

Jeudi 19 septembre 2013

Où en est-on ?

Prendre en considération le publique à qui s'adresse le travail

Quand un « mathématicien » travaille avec un biologiste ou un médecin il devrait essayer de satisfaire à trois contraintes :

1. arriver à un résultat mathématiquement correct – c'est le travail « normal » du mathématicien – ;
2. arriver à un résultat interprétable par le collègue biologiste ou médecin – ce qui nécessite de longues discussions – ;
3. arriver à un résultat utilisable en pratique par ces derniers – ce qui requiert, en général, un développement logiciel.

Où en est-on ?

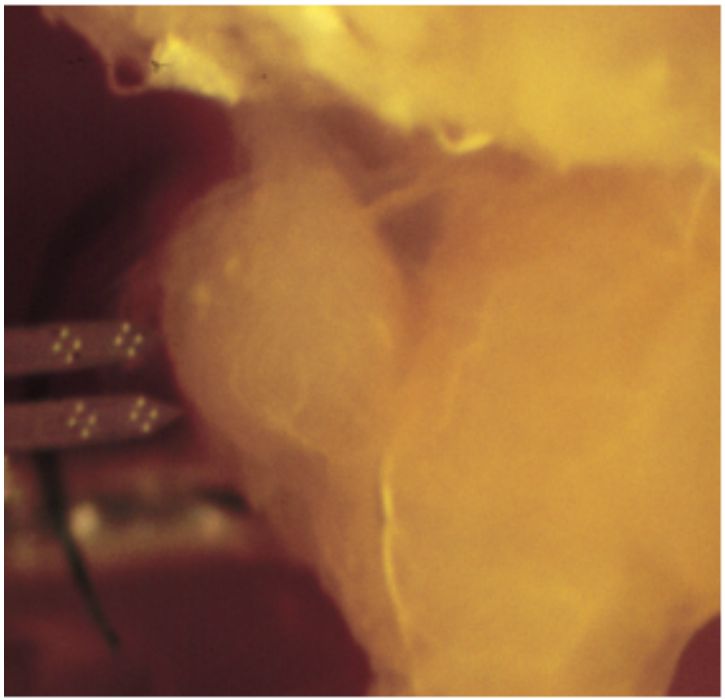
Introduction

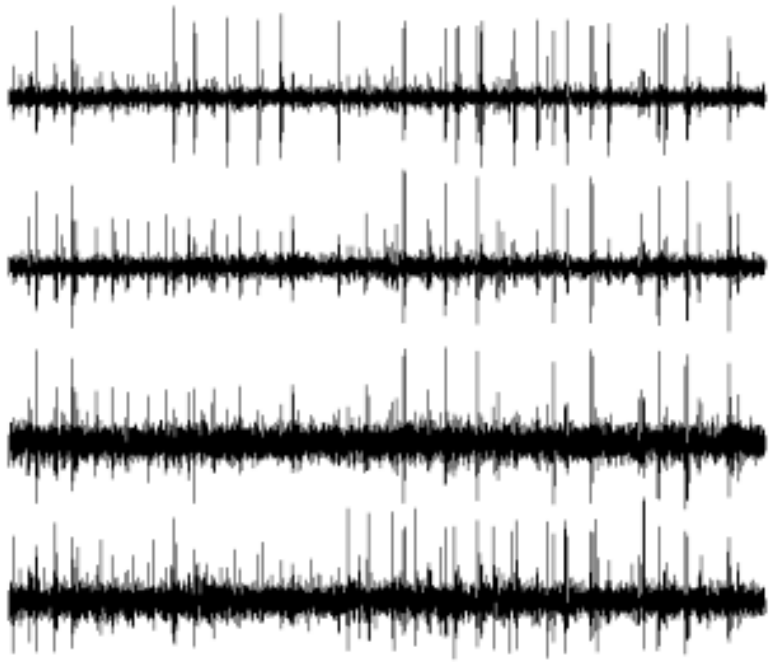
Analyse des séquences de potentiels d'action

Analyse des données de fluorescence liée au calcium

L'origine des données

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

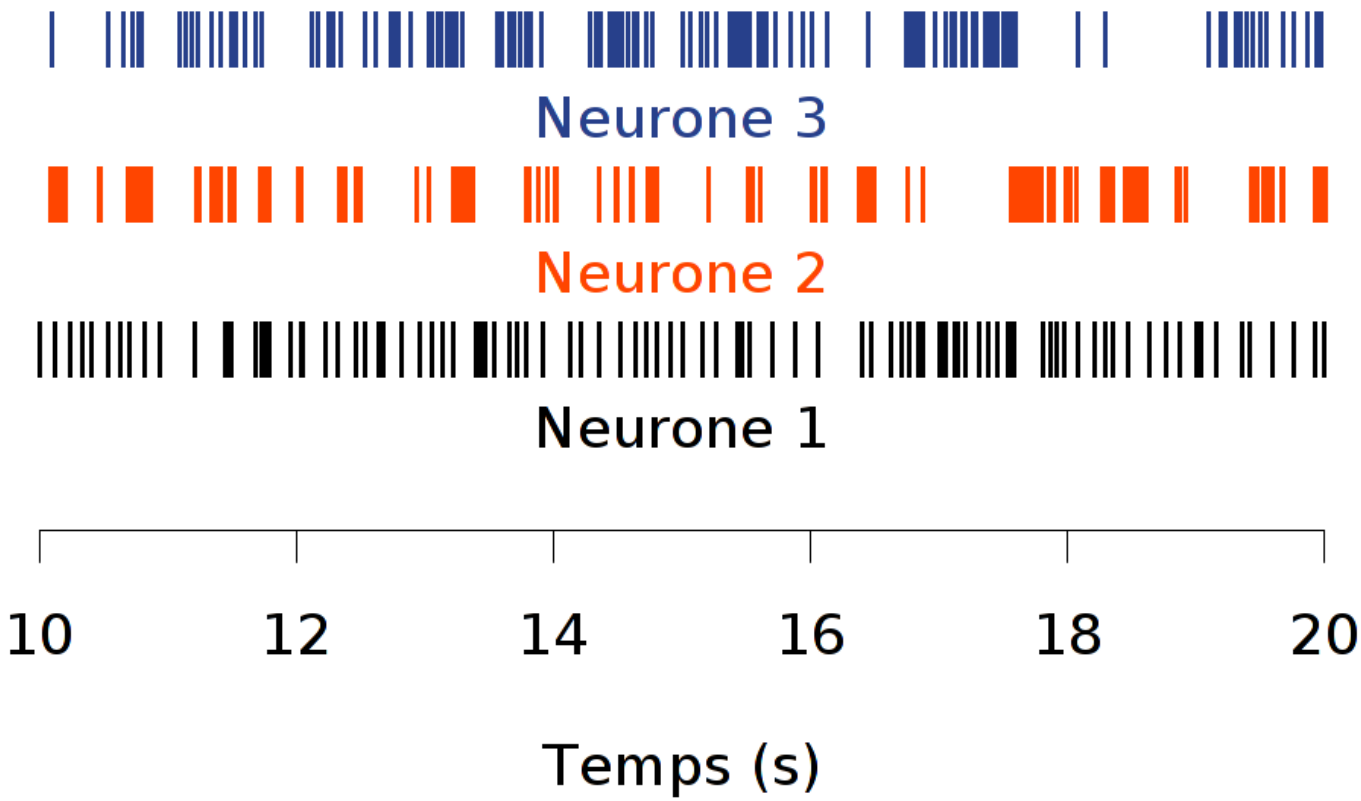




À gauche, le cerveau – d'un insecte – et la sonde d'enregistrement qui comporte 16 électrodes (les points brillants). La largeur d'une branche de la sonde est de 80 µm. A droite, 1 sec d'enregistrement sur 4 électrodes. Les pics sont des potentiels d'action.

Exemple de séquences de potentiels d'action

Après une étape de pré-traitement appelée tri des potentiels d'action (PAs), on obtient le graphe en raster représentant les séquences de PAs – données enregistrées dans le premier relais olfactif de la blatte (*Periplaneta americana*) par Antoine Chaffiol – :



Pourquoi et comment modéliser les trains de PAs ?

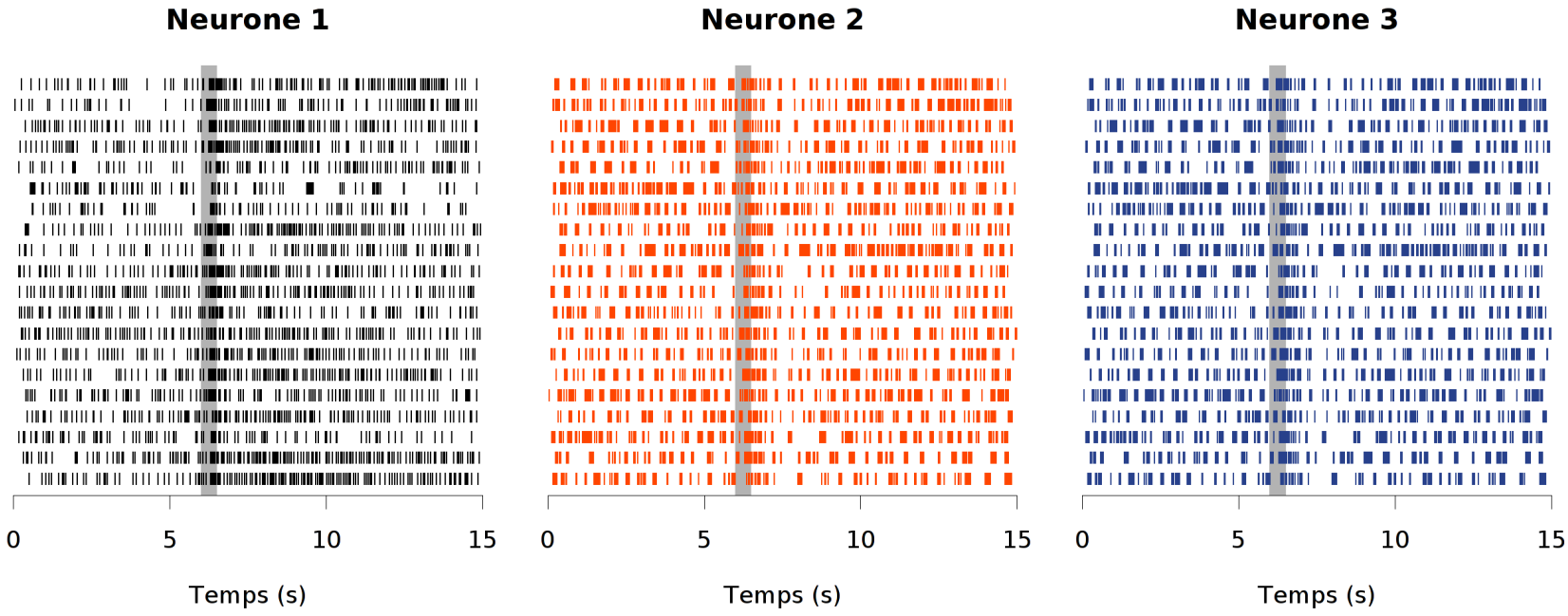
- une hypothèse de travail centrale en neurosciences est que les temps d'apparition des PAs, par opposition à leurs formes, sont le seul support de transmission de l'information entre régions du cerveau ;
- cette hypothèse légitime l'étude des trains de PAs en tant que séquences de points sur la demi-droite réelle (représentant le temps) sans nécessairement tenir compte des mécanismes biophysiques qui les génèrent.

Des questions posées par les neurophysiologistes

Les questions suivantes vont être formulées dans le cadre d'une étude du premier relais olfactif d'un insecte – mais des questions similaires se posent pour les autres systèmes sensoriels de même que dans le système moteur ; chez les insectes comme chez les vertébrés – :

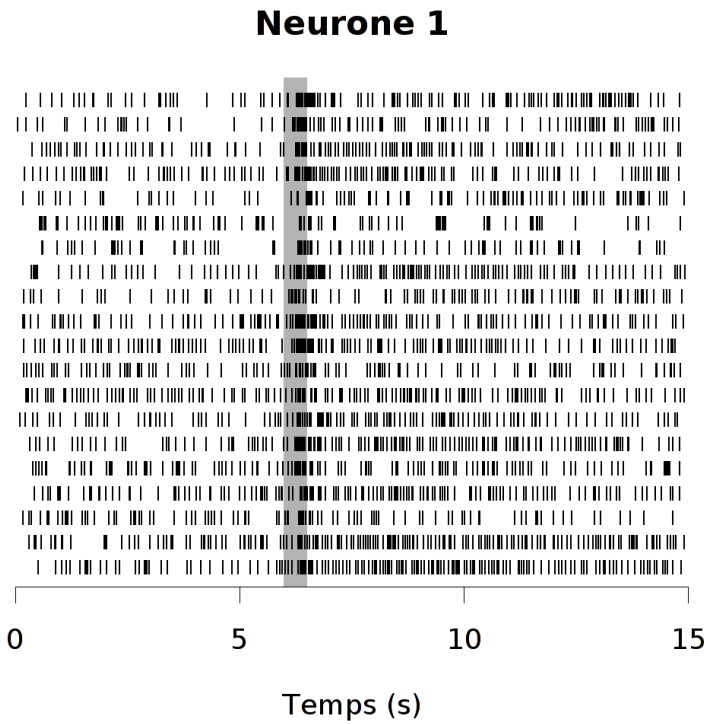
- un neurone répond-il à une stimulation (odeur) donnée ?
- un neurone répond-il différemment à différentes simulations – des stimulations peuvent être différentes parce-qu'elles sont de natures différentes : pas les mêmes odeurs ; ou parce-que leurs intensités sont différentes : pas les mêmes concentrations – ?
- deux neurones (ou plus) ont-ils des activités spontanées corrélées ?
- les corrélations d'activité en régime spontané sont-elles modifiées par une stimulation ?

Exemples de réponses de trois neurones à une odeur



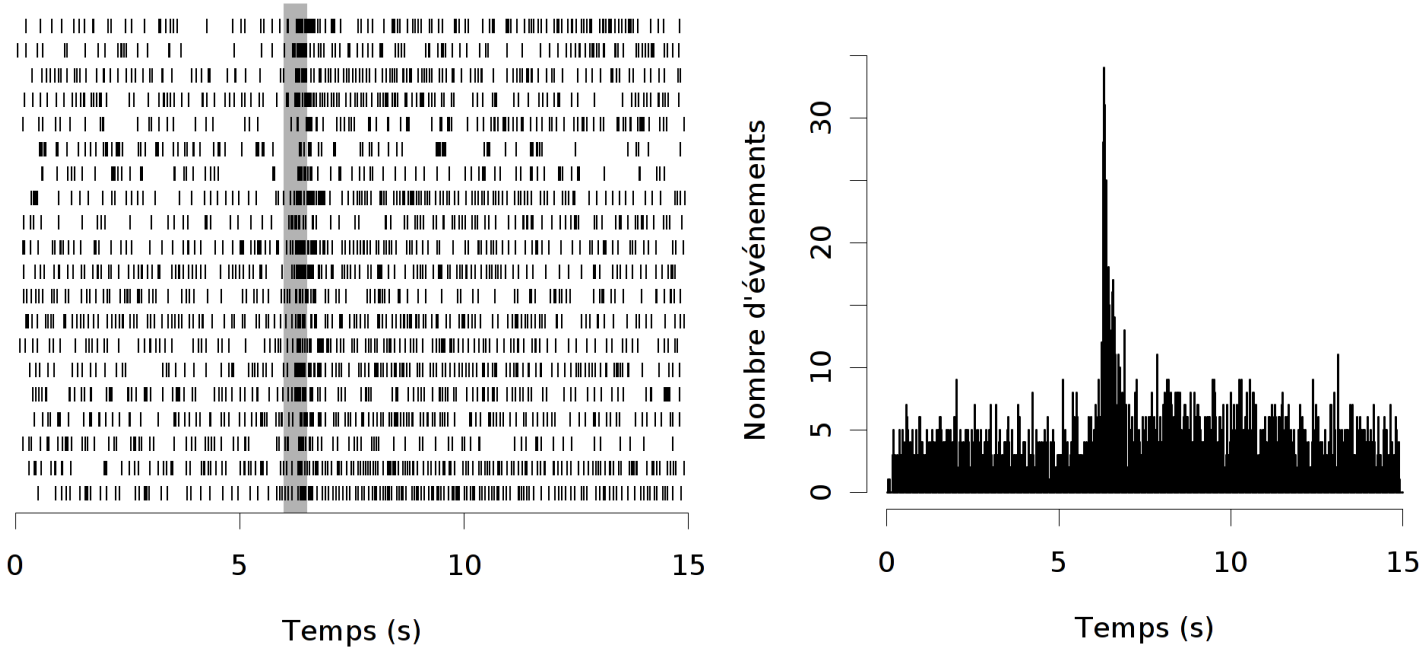
20 réponses à 20 stimulations avec du citronellal (composant principal de l'huile de citronnelle). La première réponse est en bas, le dernière est en haut (1 minute sépare les stimulations successives). La bande grisée correspond au temps d'ouverture (0,5 s) de l'électro-valve contrôlant l'arrivée de la stimulation. Données enregistrées par Antoine Chaffiol.

Estimation de la fréquence de la « réponse moyenne » (1)



On passe des données brutes…

Estimation de la fréquence de la « réponse moyenne » (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.

Estimation de la fréquence de la « réponse moyenne » (3)

- nous modélisons le « processus moyenné » comme un processus de Poisson inhomogéne d'intensité $\nu(t)$;
- l'histogramme que nous avons construit peut alors être considéré comme l'observation d'un collection de variables aléatoires de Poisson, $\{Y_1, \ldots, Y_k\}$, de paramètres :

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

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.

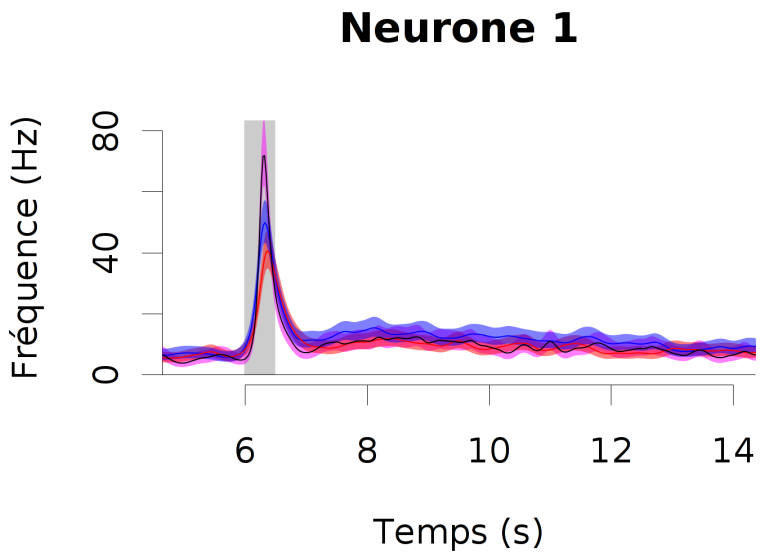
Estimation de la fréquence de la « réponse moyenne » (4)

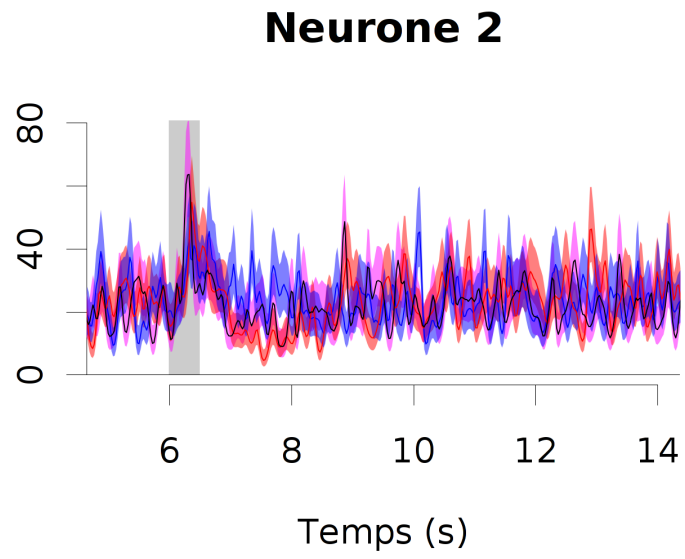
- nous estimons en fait directement : $\eta() = \log\left(N_s \nu() \delta\right)$ par régression de Poisson ;
- comme nous ne voulons pas faire d'hypothèses « fortes » sur η nous choisissons comme estimateur, $\hat{\eta}$, la fonction qui minimise la log-vraisemblance pénalisée :

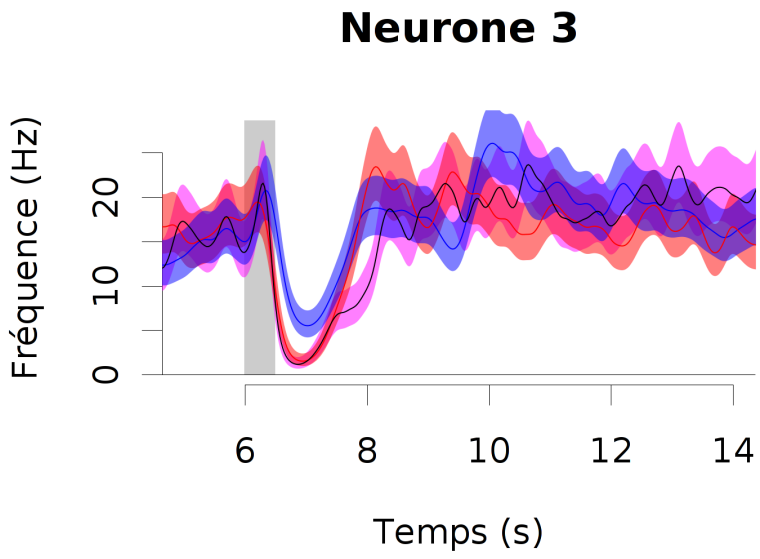
$$-\sum_{i=1}^{k}\left(y_i \eta(t_i) - \exp\left(\eta(t_i)\right)\right) + \lambda \int \left(\frac{d^2\eta(u)}{dt^2}\right)^2 du\,;$$

- le paramètre de lissage, λ, est obtenu en minimisant le critère de validation croisée généralisée ;
- cette approche, développée par Grace Wahba, fournit aussi des intervalles de confiance ;
- tous les calculs sont effectués avec le paquet `gss` (*general smoothing splines*) développé par Chong Gu pour le logiciel `R`.

Estimation de la fréquence de la « réponse moyenne » (5)







Bilan

- nous n'avons illustré qu'une méthode d'estimation entièrement automatique de la fréquence de la réponse moyenne d'un neurone à une stimulation ;
- cette méthode fourni des intervalles de confiances ;
- nous n'avons pas le temps ici d'illustrer nos méthodes d'estimation de l'intensité conditionnelle de nos processus ponctuels – une extension non-paramétrique de l'approche de David Brillinger (1988) – ;
- nous avons également proposé un nouveau test d'adéquation qui complète ceux proposés par Yosihiko Ogata (1988) ;
- tout cela est disponible sous forme du paquet `R` : `STAR` (*Spike Train Analysis with R*) effectivement utilisé par plusieurs expérimentateurs.

Où en est-on ?

Introduction

Analyse des séquences de potentiels d'action

Analyse des données de fluorescence liée au calcium

Pourquoi s'intéresser au calcium ?

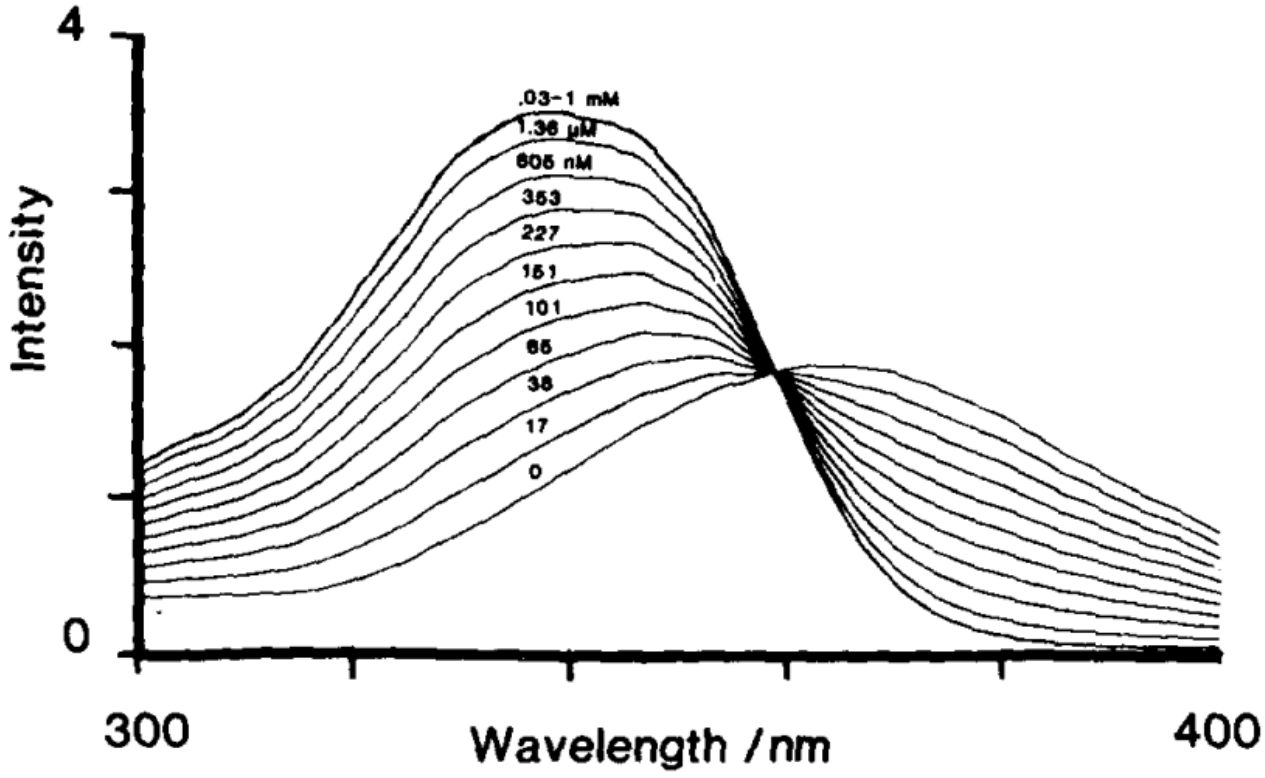
Les ions calcium (Ca^{2+}) contrôlent ou influencent des fonctions – au sens biologique du terme ! – aussi variées que :

- la mobilité cellulaire ;
- la mitose ;
- la contraction musculaire ;
- l'exocytose.

Mesurer la concentration calcique

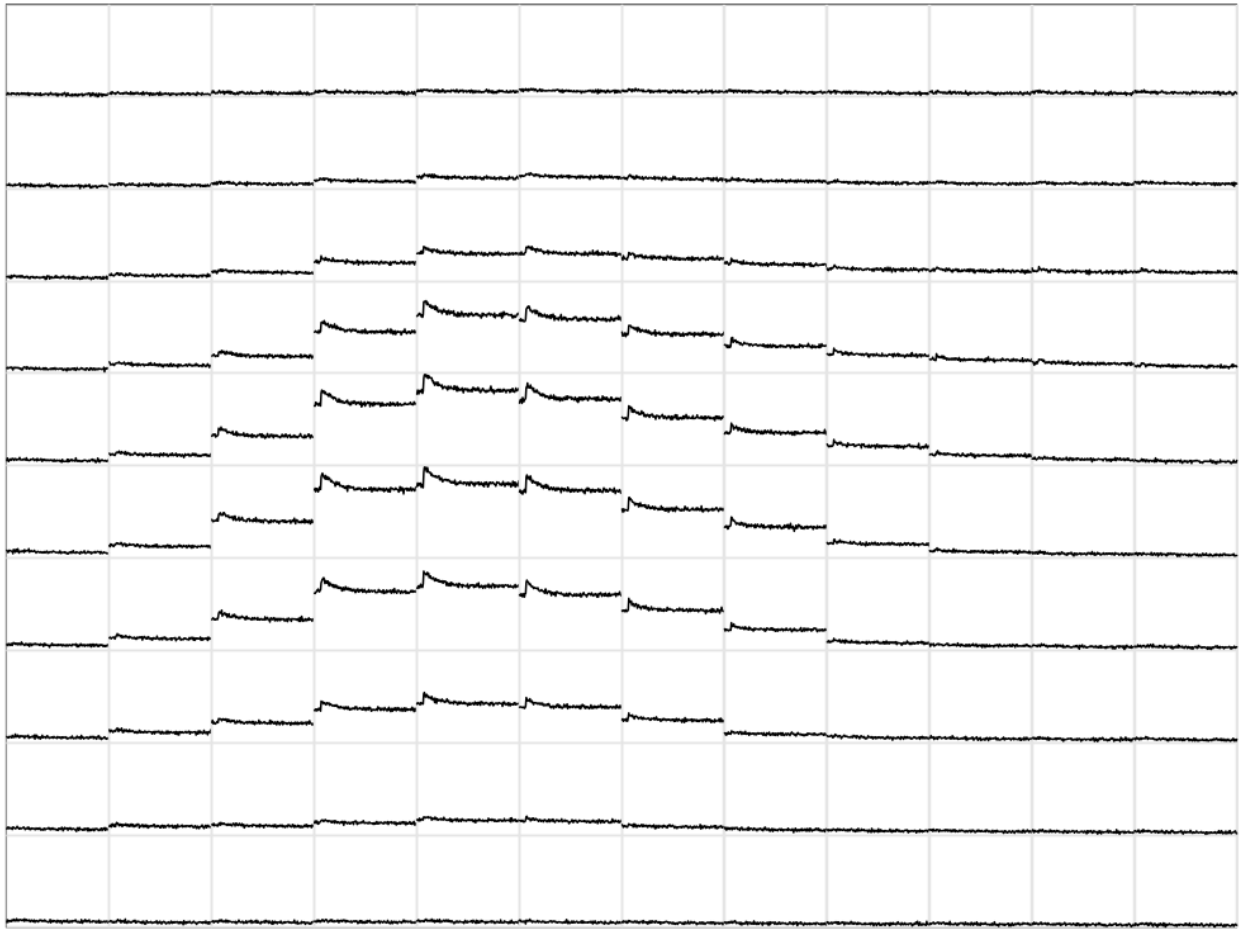
- les sondes / colorants / fluorophores, molécules dont les propriétés d'absorption ou d'émission changent quand elles se lient (réversiblement) au Ca^{2+} constituent la base des méthodes « modernes » de mesure de la concentration calcique ;
- l'emploi de ce type de technique permet une approche « quantitative » lorsqu'on dispose d'un microscope équipé de capteurs « adéquats » comme des capteurs CCD ou des photomultiplicateurs ;
- dans la suite nous allons nous concentrer sur un type de colorant : le Fura-2 ; dont la fluorescence sera enregistrée au moyen d'un capteur CCD (c.-à-d. un capteur du même type que celui d'un appareil photo numérique).

Fura-2



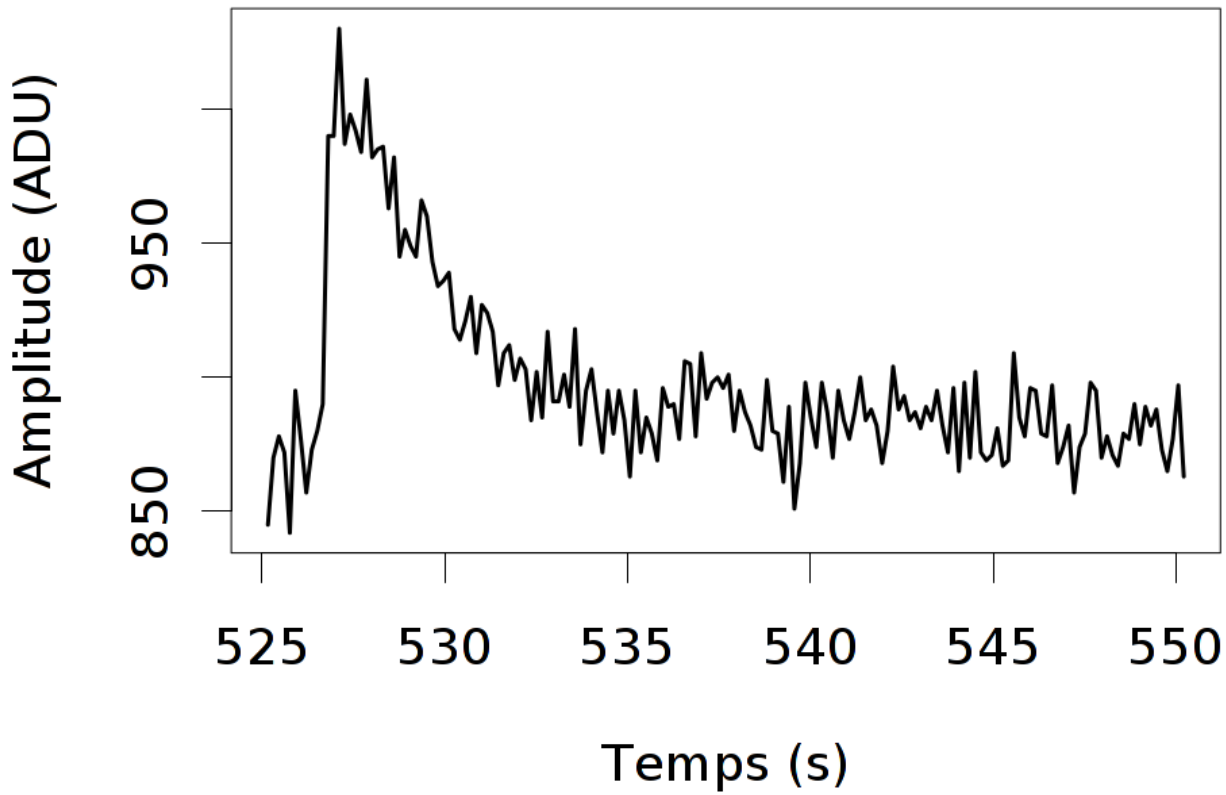
Spectres d'absorption du Fura-2 en fonction de la $[Ca^{2+}]$. Les mesures sont effectuées à 510 nm ; le spectre d'émission est indépendant de la $[Ca^{2+}]$. Les données que vous allez voir par la suite ont été enregistrées avec une longueur d'onde d'excitation de 340 nm, c.-à-d. que la fluorescence augmente avec la $[Ca^{2+}]$.
Source : Grynkiewicz, Poenie et Tsien (1985).

Exemple de données brutes (1)



Données brutes (en ADU, *Analog to Digital Unit*). 25,05 s de données sont montrées (avec une échelle uniforme) dans chaque pixel. Données enregistrées par Andreas Pippow, laboratoire de Peter Kloppenburg, Université de Cologne.

Exemple de données brutes (2)



Les « fluctuations » inhérentes à ce type de données sont clairement visibles sur cet agrandissement d'un des pixels centraux de l'image précédente.

Ce que veulent les physiologistes

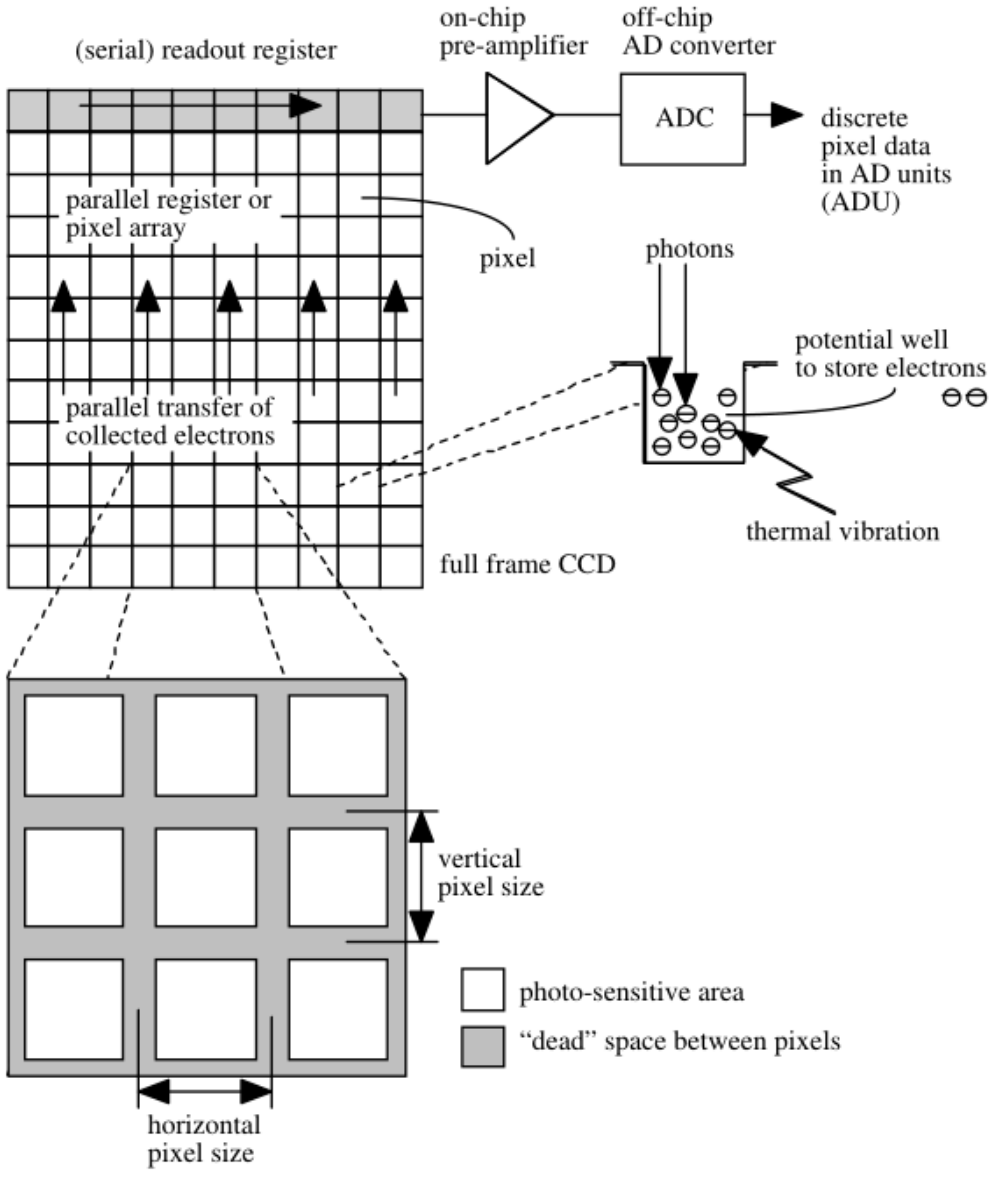
À partir des données que nous venons de présenter, les physiologistes vont souhaiter estimer des « paramètres » comme :

- l'amplitude au pic ;
- la, ou les, constantes de décroissance ;
- le niveau de base du signal ;
- le décours temporel complet – *stricto sensu*, une fonction.

Pour cela nous aurons besoin :

- d'une description des sources de « bruit » (fluctuations) dans nos enregistrements ;
- d'un modèle liant la fluorescence mesurée au calcium – si nous souhaitons « traduire » nos mesures de fluorescence en terme de concentration calcique .

Origine des fluctuations (1)



Le principe de fonctionnement d'une caméra CCD. Source : L. van Vliet et col. (1998, figure 3).

Origine des fluctuations (2)

- le bruit photonique (*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 :

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

- le bruit thermique vient du fait que l'agitation thermique peut faire tomber des électrons dans les puys de potentiels, ce bruit suit aussi une loi de Poisson mais il peut être rendu négligeable en *refroidissant* le capteur ;
- 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 que la lecture n'est pas effectuée à trop haute fréquence).

Un modèle simple de CCD

Les mesures en un pixel donné sont modélisées comme réalisations d'une variable aléatoire Y ; avec les « ingrédients » suivants :

- la convergence (en loi) d'une variable aléatoire de Poisson de paramètre λ vers une variable aléatoire normale de moyenne et variance λ ;
- les sources de bruits (photonique et lecture) sont prises en compte ;

nous aboutissons à :

$$Y \alpha \mathcal{N}\left(G\lambda, G^2\left(\lambda + \sigma_L^2\right)\right) \quad \text{ou} \quad Y = G\lambda + \sqrt{G^2\left(\lambda + \sigma_L^2\right)}\, Z\,,$$

où G est le gain de la caméra, σ_L^2 est la variance du bruit de lecture et Z est une variable aléatoire normale centrée, réduite.
On remarque que :

$$\mathrm{Var}\, Y = G(G\lambda) + G^2\,\sigma_L^2 = G(\mathrm{E}\, Y) + G^2\,\sigma_L^2,$$

c'est-à-dire que la variance est une fonction linéaire de la moyenne.

Stabilisation de la variance (1)

En utilisant la *méthode delta* ou *méthode de propagation des erreurs*, on voit qu'en travaillant avec W :

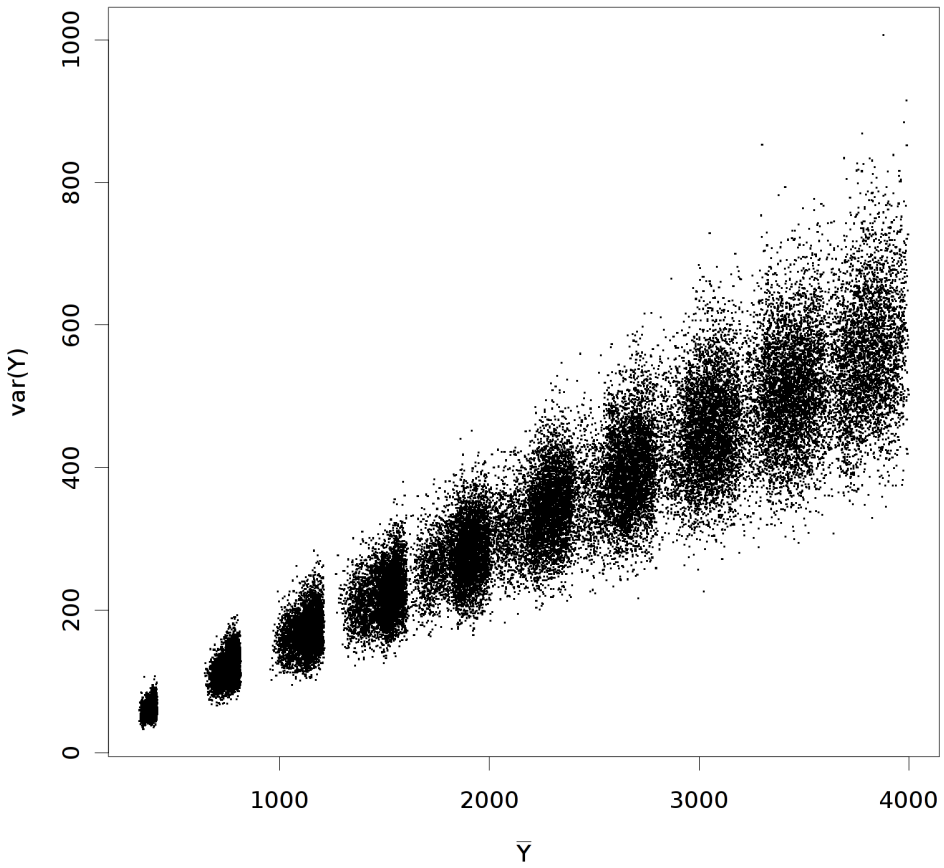
$$W = 2\sqrt{Y/G + \sigma_L^2}$$

nous avons :

$$W \approx 2\sqrt{\mathrm{E}Y/G + \sigma_L^2} + Z,$$

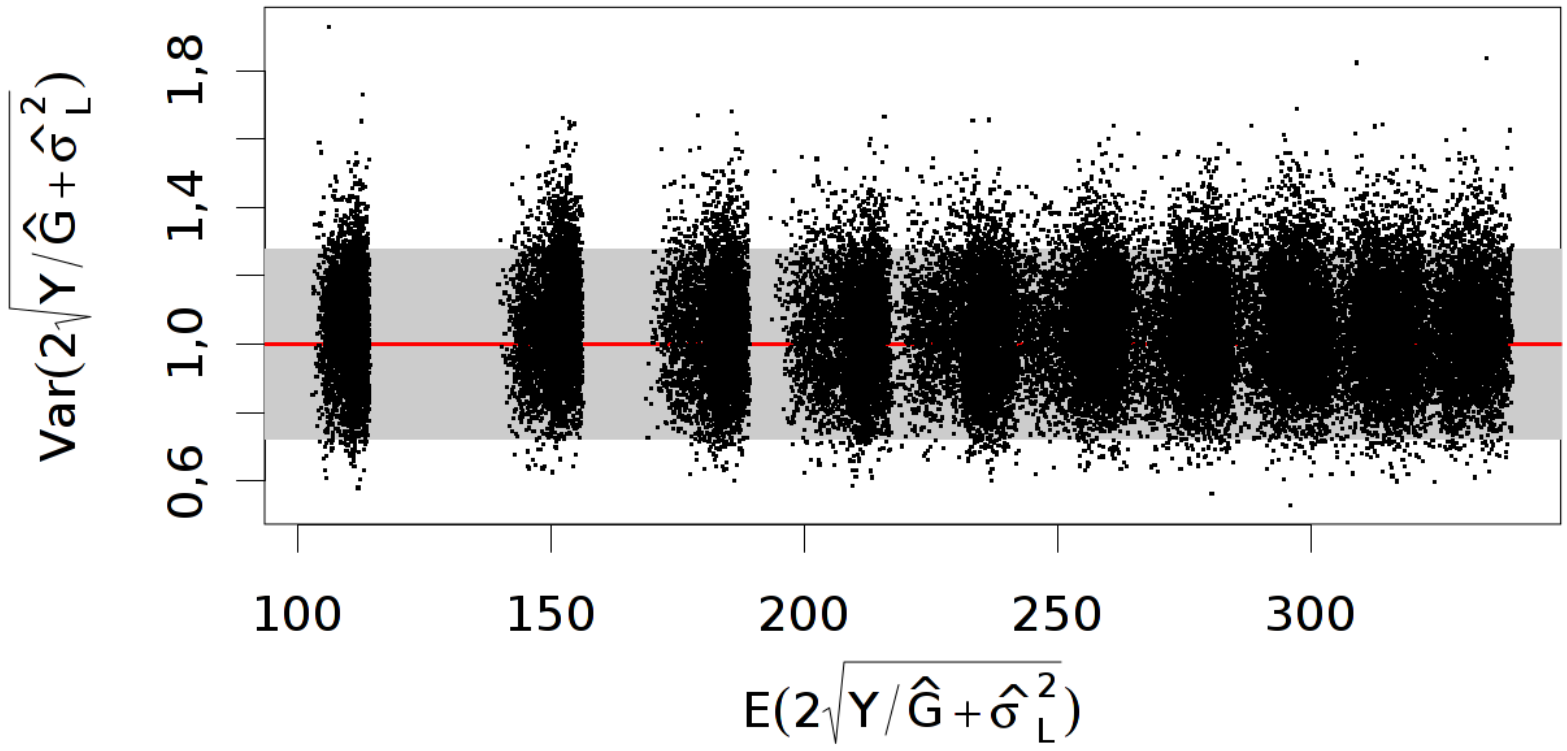
c'est-à-dire que W est (approximativement) de loi normale de moyenne $2\sqrt{\mathrm{E}Y/G + \sigma_L^2}$ et de variance 1.

Stabilisation de la variance (2)



Expérience de calibration. Chaque point correspond au couple (moyenne, variance) estimé sur 100 mesures, pour un pixel et un temps d'exposition donnés. 10 temps d'exposition ont été employés. La relation linéaire entre variance et moyenne est clairement visible. Données de A. Pippow et P. Kloppenburg (Univ. de Cologne).

Stabilisation de la variance (3)



Données précédentes transformées pour stabiliser la variance. Le gain G et la variance de lecture σ_L^2 ont été estimés à partir des données par regression linéaire. Bande grise : intervalle de confiance à 95 % (empiriquement, 93 % des points sont dedans) ; ligne rouge : valeur théorique.

Construction automatique des régions d'intérêt (1)

- après stabilisation de la variance avec $W_{i,j,k} = 2\sqrt{Y_{i,j,k}/\widehat{G} + \hat{\sigma}^2_L}$, la variance de chaque pixel (i,j) at chaque temps k devrait être égale à 1 ;
- si un pixel ne contient pas de signal dépendant du temps, la statistique suivante :

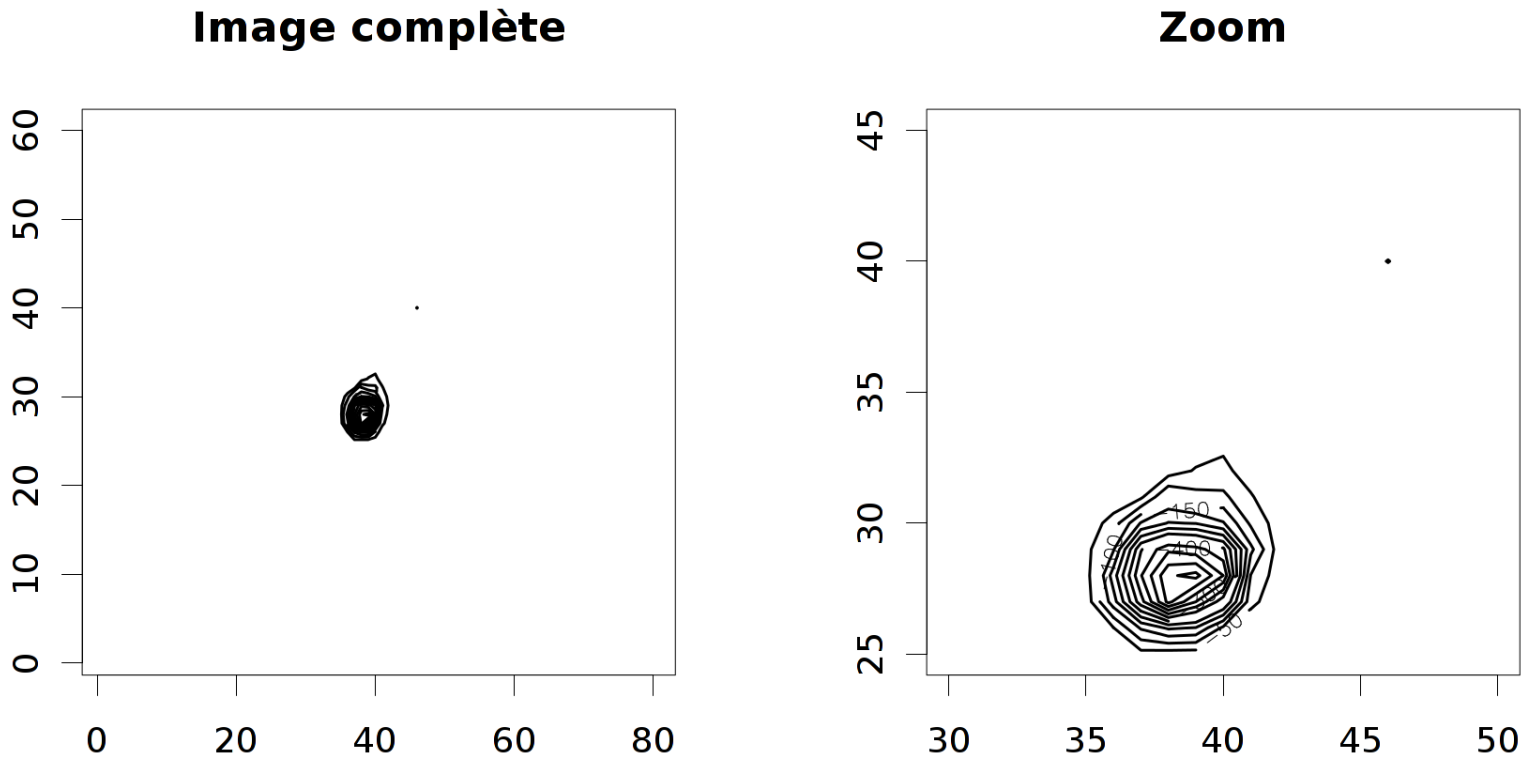
$$RSS_{i,j} \equiv \sum_{k=1}^{K} (W_{i,j,k} - \overline{W}_{i,j})^2 \quad \text{où} \quad \overline{W}_{i,j} \equiv \frac{1}{K} \sum_{k=1}^{K} W_{i,j,k}$$

  devrait être un χ^2 à $K-1$ degrés de liberté ;
- nous pouvons alors calculer la valeur de la complémentaire de la fonction de répartition d'une loi du χ^2_{K-1} :

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

  et chercher les très faibles valeurs.

Construction automatique des régions d'intérêt (2)



Lignes de niveaux de $\log\left(1 - F_{\chi^2_{K-1}}(RSS_{i,j})\right)$

Estimation du signal (1)

- nous allons être « prudents » et ne garder dans notre région d'intérêt que les pixels tels que : $\log\left(1 - F_{\chi^2_{K-1}}(RSS)\right) \leq -300$;
- il nous reste alors 12 pixels ;
- nous allons modéliser l'intensité de fluorescence de ces pixels par :

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

où $f()$ est le signal commun à chaque pixel de la région, $\phi_{i,j}$ est un paramètre spécifique à chaque pixel et où b est l'auto-fluorescence supposée commune à chaque pixel ;

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

Estimation du signal (2)

- nous avons besoin d'une contrainte car avec notre modèle :

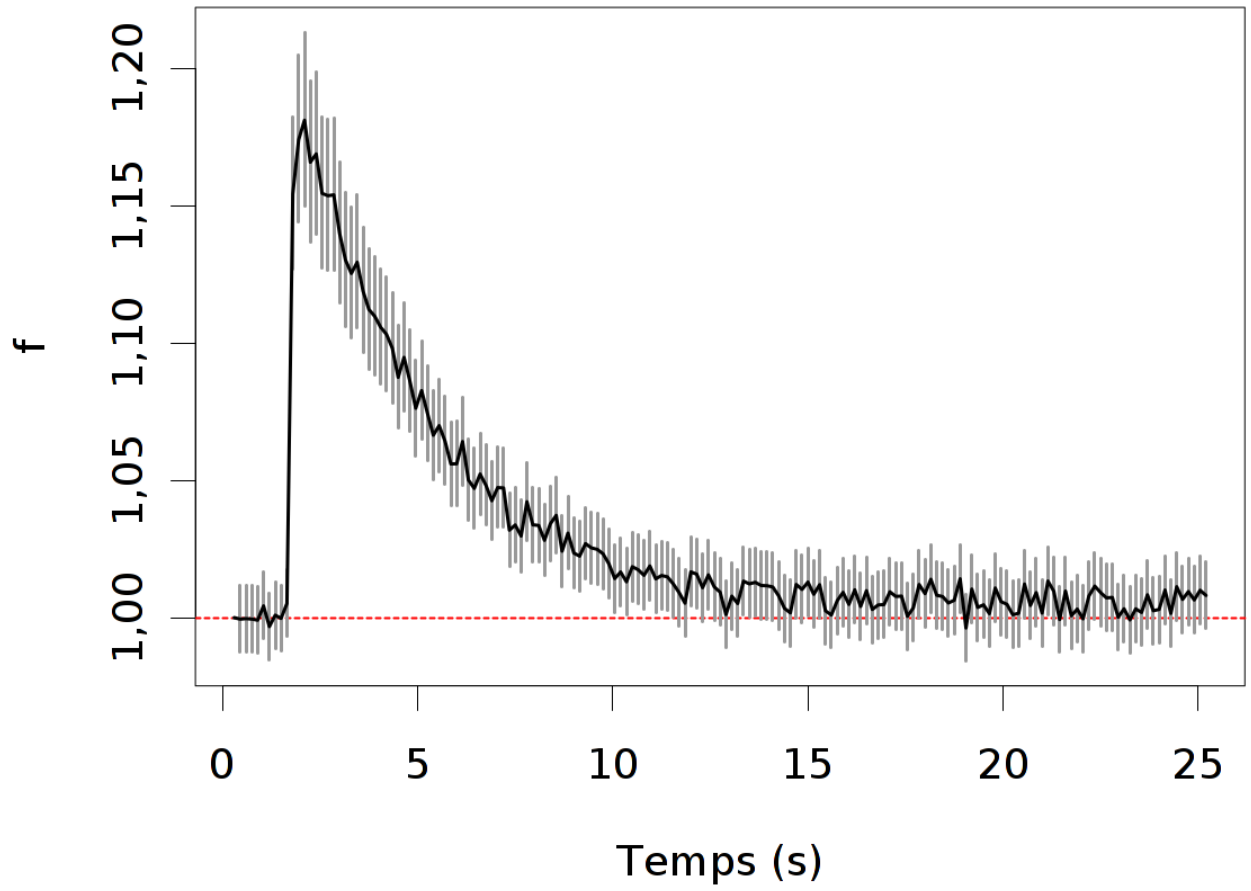
$$\lambda_{i,j,k} = \phi_{i,j} f_k + b\,,$$

  nous pouvons multiplier tous les $\phi_{i,j}$ par 2 et diviser tous les f_k par 2 en conservant la même prédiction ;
- nous allons donc poser $f_1 = 1$; notre estimation est alors liée à ce que les physiologistes font habituellement avec ce type de données, $\Delta\lambda(t)/\lambda(1) = (\lambda(t) - \lambda(1))/(\lambda(1)) = f(t) - 1 + \text{bruit}$;
- nous n'avons pas besoin de mesure indépendante de l'auto-fluorescence ;
- l'estimation se fait par régression non-linéaire en minimisant (avec l'algorithme BFGS) :

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(W_{i,j,k} - 2\sqrt{\left(\phi_{i,j} f_k + b\right)/\widehat{G} + \hat{\sigma}_L^2}\right)^2\,,$$

  il est aussi possible de prendre en compte l'incertitude sur les paramètres calibrés.

Estimation du signal (3)



Signal estimé avec intervalles de confiance à 95 %.

Remerciements

- Antoine Chaffiol, ancien étudiant en thèse, pour les données du système olfactif de la blatte ;
- Chong Gu, Prof. à l'Université de Purdue, pour son aide sur l'application des splines de lissage à l'analyse des séquences de potentiels d'action ;
- Andreas Pippow et Peter Kloppenburg de l'Université de Cologne, pour les données de fluorescence liées au calcium ;
- Sébastien Joucla et Romain Franconville, anciens post-doc et thésard, pour le travail commun sur l'analyse des données d'Andreas et Peter ;
- les organisateurs pour m'avoir invité ;
- vous, pour m'avoir écouté.