
Estimation de l'intensité d'un processus ponctuel appliqué à la vision de nouveaux-nés

Laboratoire Jean Kuntzmann (LJK)

Babylab



Rapport de stage d'Antoine Sétif

Master 2 MASSS 2014-2015

Université Pierre Mendès France, Grenoble



sous la supervision d'Adeline Leclercq-Samson et Jean-François Coeurjolly

Remerciements

Dans un premier temps, je tiens à remercier Eric Bonnetier, directeur du laboratoire Jean Kuntzmann (LJK), de m'avoir accepté en tant que stagiaire au sein du laboratoire.

Je remercie tout particulièrement mes 2 tuteurs : Madame Adeline Leclercq-Samson, professeur à l'Université Joseph Fourier (UJF) et Monsieur Jean-François Coeurjolly, professeur à l'Université Pierre Mendès France (UPMF), tous deux membres du LJK, pour les précieux conseils concernant les missions évoquées dans ce rapport, pour l'aide qu'ils m'ont apportée lors des différents suivis, pour leur soutien concernant l'obtention de ce stage et pour leur grande disponibilité.

Je remercie également, David Mearly, maître de conférence au Laboratoire Psychologie et NeuroCognition de l'Université Pierre Mendès France (UPMF), mais également membre de l'équipe BabyLab (centre d'études grenoblois du nourrisson et du jeune enfant) et investigator de l'étude, pour sa disponibilité à répondre à mes questions.

Je remercie Adrian Baddeley, professeur au campus de Bentley, l'un des principaux chercheurs australiens dans la science statistique, pour le temps accordé aux questions d'interprétation portant sur le package R **spatstat** dont il est le créateur.

Remerciements à l'ensemble du personnel du laboratoire LJK pour son accueil chaleureux.

Enfin, je tiens à exprimer ma reconnaissance à toute l'équipe pédagogique et aux intervenants professionnels responsables de la formation du master MASSS.

Table des matières

| | | |
|----------|--|-----------|
| 1 | Introduction aux processus ponctuels spatiaux | 7 |
| 1.1 | Définition | 7 |
| 1.2 | Intensité d'un processus ponctuel | 8 |
| 2 | Le processus ponctuel de Poisson | 9 |
| 2.1 | Définition et propriétés | 9 |
| 2.2 | Estimation paramétrique de l'intensité d'un PPP | 10 |
| 2.2.1 | Le modèle linéaire généralisé (GLM) | 10 |
| 2.2.2 | Le modèle linéaire mixte généralisé (GLMM) | 14 |
| 3 | Application aux données de mouvements oculaires | 16 |
| 3.1 | Description du jeu de données | 16 |
| 3.2 | Visualisation du jeu de données | 17 |
| 3.3 | Distribution du processus ponctuel | 18 |
| 3.3.1 | Distribution du processus ponctuel général | 18 |
| 3.3.2 | Distribution du processus ponctuel par groupes d'âge | 20 |
| 3.4 | Covariables spatiales | 21 |
| 3.4.1 | Cartes de saillance | 21 |
| 3.4.2 | Zones cognitives | 24 |
| 3.5 | Analyse des données oculométriques par un modèle log-linéaire | 26 |
| 3.5.1 | A la découverte des fonctions ppm et mppm | 26 |
| 3.5.2 | Utilisation de la fonction MPPM aux données | 27 |
| 3.6 | Analyse des données oculométriques par un modèle log-linéaire mixte | 31 |
| A | Annexes | 33 |
| A.1 | Fixations par groupe d'âge, pour toutes les images | 33 |
| A.2 | Test des quadrats, image Rafting | 36 |
| A.3 | Fonction K_{inhom} appliquée aux 6 images | 36 |
| A.4 | Fonction K appliquée aux 5 processus de groupe d'âge, rafting | 37 |
| A.5 | Fonction K_{inhom} appliquée aux 5 processus de groupe d'âge, rafting | 37 |
| A.6 | Niveaux de saillance pour les 6 images | 38 |
| A.7 | Fixations associées au niveau de saillance moyen pour les 6 images | 40 |
| A.8 | Distances aux zones cognitives | 40 |
| A.9 | Cartes binaires pour les 6 images | 41 |
| A.10 | Intensités normalisées estimées non-paramétriquement | 41 |
| A.11 | Exemple de sortie R mppm, image Rafting | 42 |
| A.12 | Procédures R, tests multiples | 42 |
| A.13 | Tests entre groupe d'âges pour toutes les images | 43 |
| A.14 | Hyperframe contenant l'ensemble des fixations de tous les groupes, toutes les images | 45 |
| A.15 | Summary obtenu à partir de l'hyperframe images regroupées | 45 |

Table des figures

| | | |
|----|---|----|
| 1 | Exemple de schéma en quadrature | 11 |
| 2 | Les 6 images de l'expérience | 16 |
| 3 | Exemple de fixations par groupes d'âge, image rafting | 17 |
| 4 | Test des quadrats, image Rafting | 18 |
| 5 | Fonction K de Ripley appliquée à l'image rafting | 19 |
| 6 | Fonction K_{inhom} de Ripley appliquée à l'image rafting | 19 |
| 7 | Fonction K de Ripley appliquée aux processus groupe d'âge 9 mois, image rafting | 20 |
| 8 | Fonction K_{inhom} de Ripley appliquée aux processus groupe d'âge 9 mois, image rafting . . | 20 |
| 9 | Les 6 niveaux de saillance appliqués à l'image rafting | 21 |
| 10 | Niveau de saillance moyen appliqué aux 6 images | 22 |
| 11 | Fixations associées au niveau de saillance moyen, image rafting | 23 |
| 12 | Fixations associées au niveau de saillance moyen, par groupe d'âge, image rafting | 23 |
| 13 | Zones cognitives définies par les 10% des intensités adultes les plus fortes | 24 |
| 14 | Zones cognitives définies par les 5% des intensités adultes les plus fortes | 24 |
| 15 | Choix parmi 3 covariables spatiales, image Rafting | 25 |
| 16 | IC Plot appliquée aux 6 images | 28 |

Motivations du stage

Une collaboration entre 2 laboratoires

Le laboratoire **Jean Kuntzmann** est un laboratoire de Mathématiques Appliquées et d'Informatique. Il doit son nom à Jean Kuntzmann (1912-1992), pionnier de l'informatique et des mathématiques appliquées à Grenoble et pionnier du décloisonnement des sciences numériques vers l'industrie et les autres disciplines. Il regroupe des équipes de cultures assez différentes : mathématiciens, numériciens, statisticiens ou encore de spécialistes de l'informatique graphique du traitement d'images.

Le laboratoire LJK est structuré en 3 départements :

- Le département *Géométrie-Image* regroupe des équipes de modélisation géométrique, de traitement, d'analyse et de synthèse d'images et de vidéos et vision par ordinateur.
- Le département *Modèles et Algorithmes Déterministes* centre ses activités sur la modélisation (par systèmes dynamiques, par équations aux dérivées partielles) et sur des outils pour le calcul numérique et symbolique.
- Le département *Probabilités/Statistique* regroupe quant à lui des probabilistes, statisticiens et spécialistes de l'analyse des données et du traitement du signal.

Le centre d'études grenoblois du nourrisson et du jeune enfant (**BabyLab**) est spécialisé dans l'étude du développement des capacités sensorielles et de la motricité des nourrissons ou très jeunes enfants (entre 2 mois et 2 ans).

Les études réalisées par Babylab contribuent à une meilleure connaissance et compréhension de la nature humaine. Les résultats participent aux conceptions ou représentations que les adultes ont des nourrissons et de leurs compétences. Les travaux réalisés sur les nourrissons sont ensuite diffusés auprès du grand public et permettent de modifier la façon dont nous interactons avec eux.

Travaux de recherche

Au cours de la première année de vie, l'absence de langage, l'immaturité du système moteur et le manque d'attention font qu'il est nécessaire d'employer des techniques d'études qui permettent de mesurer de manière indirecte les capacités cognitives du nourrisson. Ces techniques reposent essentiellement sur l'observation visuelle, auditive ou motrice du nourrisson.

David Meary, membre de BabyLab et investigator de l'étude que nous allons analyser, porte ses travaux de recherches sur l'exploration visuelle de scènes en utilisant la mesure des mouvements oculaires pour des bébés de groupes d'âges différents (3 mois, 6 mois, 9 mois et 12 mois).

David nous a aimablement préparé et fourni l'ensemble des données. Afin de comprendre le contexte de cette application, nous reproduisons librement une partie de l'article de Barthelme [5], sur lequel nous nous sommes appuyés pour la motivation de l'utilisation de processus ponctuel et pour justifier l'ensemble de données choisi.

Mouvements oculaires

En regardant une scène statique, nos yeux exécutent une séquence de mouvements très rapides (des saccades) interrompus par des moments de relative stabilité (des fixations). L'acuité visuelle est la plus élevée au centre du regard, à savoir la fovéa, qui est donc la zone de la rétine où la vision des détails est la plus précise. Cette acuité diminue plus on s'écarte de cette zone.

Environ 3-4 saccades sont générées chaque seconde. Une saccade moyenne déplace les yeux de 4-5° pendant la perception de la scène et en fonction de l'amplitude, la durée est entre 20-50 ms.

Sur une image donnée, on retrouve généralement des groupes de fixations condensés sur certaines zones (clusters) alors que d'autres zones de l'image ne reçoivent que peu ou pas de fixations. Ce comportement a été associé à plusieurs facteurs. En effet, les caractéristiques locales des images comme le contraste, les bords ou la couleur sont supposées guider les mouvements oculaires et jouent un rôle prépondérant.

Outre ces caractéristiques de l'image, les fixations semblent être guidées par les visages, les personnes et les objets. Récemment, il a été montré que les objets sont en moyenne, plus saillants que la scène de fond suggérant que les saccades pourraient cibler principalement les objets. De plus, le comportement de l'inspection de nos yeux est de plus modulé par des connaissances spécifiques sur une scène acquise au cours des dernières fixations.

Relation entre données oculométriques et processus ponctuels spatiaux

Les processus ponctuels spatiaux sont des modèles stochastiques faisant intégrer la géométrie stochastique et la statistique spatiale. Ces processus peuvent être appliqués dans de nombreux domaines : en astrophysique pour modéliser la position des galaxies, en foresterie pour modéliser la position des arbres, en épidémiologie pour modéliser les emplacements de certaines maladies.

C'est pour cela que les processus ponctuels spatiaux sont un très bon moyen d'analyser ce genre de données de mouvements oculaires puisque pour chaque fixation que nous observerons, sera associée une entité géographique. Il existe de puissants outils statistiques capables de modéliser ces données spatiales, notamment au sein du package **spatstat** du logiciel **R**. Dans un premier temps, il serait intéressant de s'apercevoir que les fixations recueillies ne sont pas aléatoirement réparties sur l'image mais qu'il existe des concentrations de points (clusters), on se demandera si elle peut être expliquée par certaines covariables spatiales, notamment celles liées aux caractéristiques de l'image.

A partir de là, nous pourrons aisément voir le lien avec les modèles linéaires généralisés (GLM) en posant des covariables (que nous appellerons covariables spatiales dans notre cas) liées à ces caractéristiques. Aussi, nous pourrons chercher à évaluer la variabilité des coefficients (associés aux covariables) entre les différents groupes d'âge.

De plus, en cherchant à comprendre les différences de visualisation entre les sujets d'un même groupe d'âge, nous modéliserons nos données par un modèle linéaire mixte généralisé (GLMM), qui est une extension des modèles linéaires généralisés, mais qui incorpore dans le prédicteur linéaires des effets aléatoires en supplément des effets fixes.

Dans ce rapport, nous allons d'abord présenter les bases théoriques des processus ponctuels spatiaux, nous présenterons leur liens avec les modèles de type GLM et GLMM, puis nous poursuivrons avec l'analyse des données réelles.

1 Introduction aux processus ponctuels spatiaux

1.1 Définition

Dans ce chapitre, nous nous intéresserons uniquement aux processus ponctuels dans $S \subseteq \mathbb{R}^d$. Avant de donner une définition d'un processus ponctuel, nous allons commencer par donner une définition d'une réalisation (ou d'une observation) x d'un processus ponctuel X .

Une configuration de points est un ensemble non ordonné d'observations $\{x_1, \dots, x_n\}$ où n est le nombre observé de réalisations. Cette configuration peut aussi être écrite de la manière suivante : $\{n; x_1, \dots, x_n\}$.

L'ordre des réalisations n'a pas d'influence dans une telle configuration et ainsi $\{x_n, \dots, x_1\}$ est le même échantillon observé que précédemment. Il est assez trivial de constater que nous pouvons obtenir $n!$ combinaisons différemment ordonnées de l'échantillon observé $\{x_1, \dots, x_n\}$ et explique pourquoi nous retrouvons le facteur $\frac{1}{n!}$ dans de nombreuses formules.

Une réalisation, ou une observation d'un processus ponctuel est une configuration de points $\{x_1, \dots, x_n\}$, $x_i \in S$ où à la fois n et la position des points sont aléatoires, l'espace S (nommé l'espace d'état, équipé de la σ -algèbre de l'ensemble des boréliens \mathcal{B}) étant l'espace dans lequel X est défini.

Pour tout sous-ensemble $x \subseteq S$, notons $n(x)$ la cardinalité de x , en fixant $n(x) = \infty$ si x n'est pas fini. Ainsi, le sous-ensemble x est dit localement fini, si $n(x_B) < \infty$ lorsque $B \subseteq S$ est borné, où :

$x_B = x \cap B$ est la restriction d'une configuration de points x dans B

Le processus X prend donc des valeurs dans l'espace défini par :

$$N_{lf} = \{x \subseteq S : n(x_B) < \infty \text{ pour tout borné } B \subseteq S\}$$

Les éléments de N_{lf} sont appellés des configurations localement finies.
On équipe l'espace N_{lf} avec la σ -algèbre.

$$\mathcal{N}_{lf} = \sigma(\{x \in N_{lf}, n(x_B) = m\}, B \in \mathcal{B}, m \in \mathbb{N}^*)$$

Définition :

Un processus ponctuel X , défini sur S , est une application mesurable définie dans l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et prenant des valeurs dans $(N_{lf}, \mathcal{N}_{lf})$.

- La distribution P_X de X est donnée par $P_X(F) = P(\{\omega \in \Omega : X(\omega) \in F\})$ pour $F \in \mathcal{N}_{lf}$
- La mesurabilité de X est donnée par $N(B) = n(X_B)$, variable aléatoire pour tout borné $B \in \mathcal{B}$

Proposition :

- La distribution d'un processus ponctuel X est déterminée par les distributions de dimensions finies de ses fonctions de comptage, c'est à dire des distributions conjointes de $N_X(B_1), \dots, N_X(B_m)$ pour tout $B_1, \dots, B_m \in \mathcal{B}$ et pour tout $m \geq 1$.
- La distribution d'un processus ponctuel X est déterminée de manière unique par la connaissance de $P(N(B) = 0)$, $\forall B$ borné $\subseteq S$.

1.2 Intensité d'un processus ponctuel

La mesure de l'intensité est le moment de premier ordre de la fonction de comptage associée. Si cette mesure d'intensité admet une densité par rapport à la mesure sur S , alors cette densité est la fonction d'intensité.

Mesure d'intensité et fonction d'intensité

Définition : La mesure du moment d'ordre $p \geq 1$ d'un processus ponctuel X est la mesure sur $(S, \mathcal{B}(S))^p$ définie sur $B_1 \times \dots \times B_p$ par :

$$\mu_p(B_1 \times \dots \times B_p) = E(N(B_1) \dots N(B_p)).$$

En se s'intéressant qu'au moment d'ordre 1, la mesure d'intensité de X devient :

$$\lambda(B) = \mu_1(B) = E(N(B)) = E \left[\sum_{u \in X} \mathbf{1}(u \in B) \right]$$

Ou, en utilisant le théorème de Campbell et la mesure de Lebesgue, on obtient :

$$E \left[\sum_{u \in X} \mathbf{1}(u \in B) \right] = \int_S \mathbf{1}(u \in B) d\lambda(u)$$

Si la relation décrite au-dessus est vraie pour $\mathbf{1}(u \in B)$, elle l'est également pour des fonctions simples de type $\sum_i h_i \mathbf{1}(u \in B_i)$. On trouve alors :

$$E \left[\sum_{u \in X} h(u) \right] = \int_S h(u) d\lambda(u)$$

S'il existe une densité ρ de la mesure d'intensité d'un point de vue mesure de Lebesgue, alors :

$d\lambda(u) = \rho(u)du$ est appellé la fonction d'intensité.

Dans ce cas, on retrouve le premier théorème de Campbell :

$$E \left[\sum_{u \in X} h(u) \right] = \int_S h(u) \rho(u) du$$

avec l'intensité $\rho(u)du$ pouvant être interprétée comme la probabilité d'occurrence d'un point u dans une boule $B(u, du)$.

2 Le processus ponctuel de Poisson

2.1 Définition et propriétés

Le processus ponctuel X est de type Poisson (PPP) avec une fonction d'intensité ρ , si et seulement si $\forall B \subseteq S$ tel que $\nu(B) = \int_B \rho(u)du < +\infty$ et $\forall F \subseteq N_{lf}$, on ait :

$$P(X_B \in F) = \sum_{n=0}^{\infty} \frac{e^{-\int_B \rho(u)du}}{n!} \int_{B^n} \mathbf{1}\{x \in F\} \mathbf{1}(\{x_1, \dots, x_n\}) \prod_i \rho(x_i) dx_1 \dots dx_n \quad (1)$$

Nous avons fait la remarque que $P(X_B \in F) = E[\mathbf{1}_{X_B \in F}]$, ainsi la relation (1) sera également vraie pour des fonctions h :

$$E(h(X_B)) = \sum_{n=0}^{\infty} \frac{e^{-\int_B \rho(u)du}}{n!} \int_{B^n} h(\{x_1, \dots, x_n\}) \prod_i \rho(x_i) dx_1 \dots dx_n \quad (2)$$

La façon classique de penser dans le cas d'un processus ponctuel de Poisson est la suivante :

- $\forall m \geq 1, \forall B_1, \dots, B_m$ bornés et disjoints $\subset S$, les variables aléatoires X_{B_1}, \dots, X_{B_m} sont indépendantes.
- $N(B) \sim \mathcal{P}(\int_B \rho(u)du)$ pour tout borné $B \subset S$

Propriétés :

Soit X suivant une loi de Poisson(ρ), observable dans $W \subseteq S$, alors :

- L'estimateur naturel de l'intensité ρ dans le cas Poisson homogène est : $\hat{\rho} = \frac{N(W)}{|W|}$
- La densité d'un processus Poisson inhomogène est donnée par :

$$f: x = \{x_1, \dots, x_n\} \rightarrow e^{|W| - \int_W \rho(u)du} \prod_{u \in x} \rho(u)$$

2.2 Estimation paramétrique de l'intensité d'un PPP

Il existe de nombreuses façons d'estimer les paramètres d'un tel modèle : méthode bayésienne, des moindres-carrés, par maximum de vraisemblance. Nous nous focaliserons sur ce dernier pour lequel de nombreux articles et ouvrages ont été publiés (Baddeley et Turner [4], Diggle [8], McCulloch [12], McCullagh [13] ou encore Waagepetersen [14] pour ne citer qu'eux).

Nous allons maintenant nous placer dans 2 contextes différents : le modèle linéaire généralisé (GLM) et le modèle linéaire mixte généralisé (GLMM) et faire le lien avec les processus ponctuels.

Un modèle linéaire est un modèle qui ne comprend qu'un type d'effet (effet fixe). Il peut s'écrire sous la forme $y = \theta^\top Z + \epsilon$ où θ est le vecteur des paramètres associés aux effets fixes, Z la matrice des covariables spatiales associée aux effets fixes et ϵ les résidus.

Il s'oppose au modèle linéaire mixte qui comprend des effets fixes, mais aussi aléatoires.

2.2.1 Le modèle linéaire généralisé (GLM)

Considérons le processus ponctuel de Poisson, défini sur S et observé dans une fenêtre $W \subseteq S$, avec une fonction d'intensité $\rho(u) = \rho_\theta(u) = e^{(\theta^\top z(u))}$ où θ est le vecteur des paramètres $\in \mathbb{R}^d$ et $z(u) = (z_1(u), \dots, z_p(u))^\top$ le vecteur des covariables spatiales.

Ainsi, l'équation (1) peut être réécrite de la façon suivante :

$$P(X \in F) = \sum_{n=0}^{\infty} \frac{e^{-\int_W \rho_\theta(u) du}}{n!} \int_{W^n} \mathbf{1}\{x \in F\} \prod_i \rho_\theta(x_i) dx_1 \dots dx_n$$

2.2.1.1 Vraisemblance

La vraisemblance (densité) de X s'écrit désormais :

$$f_\theta(x) = e^{|W| - \int_W \rho_\theta(u) du} \prod_{u \in x} \rho_\theta(x_i) = e^{|W| - \int_W e^{(\theta^\top z(u))} du} \prod_{u \in x} e^{(\theta^\top z(u))}$$

Ce qui peut également s'écrire en changeant les indices du produit par :

$$f_\theta(x) = e^{|W| - \int_W \rho_\theta(u) du} \prod_{i=1}^{n(X)} \rho_\theta(x_i) \text{ où } n(X) \text{ représente le nombre de points total dans } X.$$

Remarque

En lisant l'article de Melanie Bell [6], on constate qu'il existe de grandes ressemblances entre les intensités que nous cherchons à estimer et les intensités conditionnelles (Papangelou) dans le cadre de processus de Gibbs (nous ne développerons pas).

Pour rappel, $\rho_\theta(u)du$ peut être interprétée comme la probabilité d'observer un point u dans $B(u, du)$. Dans le cadre de Gibbs, $\rho_\theta(u, \mathbf{x})du$ pourrait être interprétée comme la probabilité d'observer un point u dans $B(u, du)$ sachant que les autres points sont en $(X = x)$.

La vraisemblance d'un modèle de Poisson, pour un jeu de données $x = X(\omega) = \{x_1, \dots, x_n\}$ sur W est :

$$L_W(\theta) = e^{|W| - \int_W \rho_\theta(u) du} \prod_{i=1}^{n(X)} \rho_\theta(x_i)$$

On en déduit finalement la log vraisemblance de Poisson suivante :

$$\log L_W(\theta) = |W| - \int_W \rho_\theta(u) du + \sum_{i=1}^{n(X)} \log \rho_\theta(x_i)$$

Puisque $|W|$ (constante) ne joue aucun rôle dans l'équation, on obtient :

$$\log L_W(\theta) = \sum_{i=1}^{n(X)} \log \rho_\theta(x_i) - \int_W \rho_\theta(u) du \quad (3)$$

En remplaçant $\rho_\theta(u)$ par son expression, on trouve finalement :

$$\log L_W(\theta) = \sum_{i=1}^{n(X)} \log e^{(\theta^\top z(x_i))} - \int_W e^{(\theta^\top z(u))} du$$

Estimation du maximum de vraisemblance par l'approximation de Berman-Turner

L'estimation du maximum de vraisemblance (MLE) de θ est obtenue en maximisant l'équation (3). Berman et Turner (1992) ont développé un programme en utilisant des méthodes de quadrature afin d'estimer au mieux ce maximum de vraisemblance pour un processus de type poisson inhomogène. Ces méthodes sont implantées dans le package **spatstat** de R.

Notons que ces méthodes s'adaptent également à l'estimation de l'intensité conditionnelle de Papan-gelou. Dans ce cas, on cherchera à estimer les paramètres du modèle en utilisant le critère du maximum de pseudo-vraisemblance (MPLE).

La quadrature numérique est une stratégie ayant pour optique d'approximer des intégrales numériques. En effet, $\int_W f(u) du$ d'une certaine fonction f est approché par une somme pondérée $\sum_j w_j f(u_j)$ où f est une fonction contenant u_j points quadratures ayant pour poids $w_j > 0$. Le poids de chaque point quadrature est défini comme l'aire de la cellule divisée par le nombre de points quadrature dans cette cellule, soit $w_j = a/n_j$.

La fenêtre W est divisée en une grille de $k \times k$ cellules, avec exactement un "dummy point" placé au centre de celles-ci.

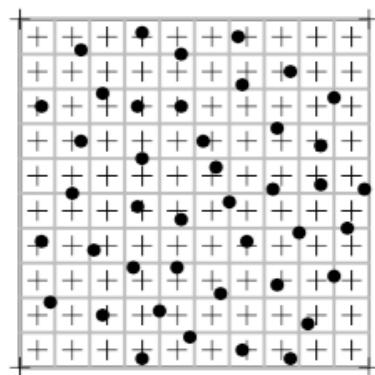


FIGURE 1 – Exemple de schéma en quadrature
Données (●), dummy points (+) et bordures des cellules (lignes grises)

L'intégrale de l'équation (3) peut alors être approchée par la quadrature :

$$\int_W \rho_\theta(u) du \approx \sum_{j=1}^m \rho_\theta(u_j) w_j$$

La log vraisemblance peut alors être exprimée par :

$$\log L_W(\theta) \approx \sum_{i=1}^{n(X)} \log \rho_\theta(x_i) - \sum_{j=1}^m \rho_\theta(u_j) w_j \quad (4)$$

Maintenant, si nous définissons une variable z_j comme étant l'indicateur des données observées (1 si le point est issu du jeu de donné, 0 si c'est un dummy point), et en posant $y_j = z_j/w_j$, l'équation (4) peut être écrite comme :

$$\log L_W(\theta) \approx \sum_{j=1}^m (y_j \log \rho_\theta(u_j) - \rho_\theta(u_j)) w_j \quad (5)$$

L'approximation de la log vraisemblance par Berman-Turner pour un modèle log-linéaire prend la forme d'une log vraisemblance pondérée Poissonnienne ayant pour fonction de lien, la fonction log.

Sur un plan pratique, et particulièrement en R, des méthodes d'estimation de modèles GLM (capables de maximiser l'équation (5)) sont déjà implantées dans certains packages R.

Nous utiliserons alors une instruction R ayant pour syntaxe :

```
glm.fit <- glm(Y ~ X, family = poisson(link = log))
```

Cependant, pour régresser un modèle de type Poisson, il serait nécessaire que la variable y_j soit de type entière. Ce n'est pas le cas ici, c'est pour cela que nous devons passer par une nouvelle approche.

2.2.1.2 Quasi-vraisemblance

On cherche ici à modéliser un modèle de type Poisson avec notre variable réponse $y_j \in \mathbb{R}^+$. Pour répondre à ce problème, nous nous sommes appuyés sur l'ouvrage de McCulloch [12] et de McCullagh [13].

Il serait intéressant d'avoir des méthodes d'inférence qui marchent aussi bien, ou presque aussi bien que le maximum de vraisemblance mais sans émettre d'hypothèses spécifiques sur les distributions des y_j , mais en connaissant leur espérance et leur fonction de variance.

C'est l'idée principale qui se dégage du terme de quasi-vraisemblance, trouver une quantité qui respecte les propriétés du maximum de vraisemblance (espérance et variance asymptotique), sans émettre d'hypothèses sur les distributions.

Dans notre contexte, notre processus ponctuel est de type Poisson, c'est à dire qu'à partir de notre échantillon iid $\underline{Y} = (Y_1, \dots, Y_m)$, où m représente le nombre de "quadrature points", nous avons $E(\underline{Y}) = \underline{\mu}$ et $V(\underline{Y}) = Diag(\underline{\mu})$.

Nous notons $Q(\mu, y)$ la log quasi-vraisemblance pour μ , basée sur les données y . Puisque les composantes de \underline{Y} sont indépendantes par hypothèse, la log quasi-vraisemblance est la somme des contributions individuelles :

$$Q(\mu, y) = \sum_i^m Q_i(\mu_i; y_i)$$

En s'appuyant sur la table 1 (librement réécrite à partir de l'ouvrage de McCullagh [13]), en sélectionnant la ligne telle que la fonction de variance $V(\mu) = \mu$, on a :

$$Q_i(\mu_i; y_i) = y_i \log \mu_i - \mu_i$$

$$\text{d'où } Q(\mu, y) = \sum_i^m (y_i \log \mu_i - \mu_i)$$

A partir de la fonction de variance connue, les équations du maximum de quasi-vraisemblance sont les mêmes que les équations du maximum de vraisemblance (équation (5)) à un terme de poids w_j près.

Dans ce contexte, on utilisera une instruction R de la forme :

```
glm.fit <- glm(Y ~ X, family = quasi(link = log, variance = mu))
```

| Variance function $V(\mu)$ | Quasi-likelihood $Q(\mu, y)$ | Canonical parameter θ | Distributions name | Range restrictions |
|----------------------------|---|------------------------------|--------------------|--------------------------------|
| 1 | $-(y - \mu)^2 / 2$ | μ | Normal | |
| μ | $y \log \mu - \mu$ | $\log \mu$ | Poisson | $\mu > 0, y \geq 0$ |
| μ^2 | $-y/\mu - \log \mu$ | $-1/\mu$ | Gamma | $\mu > 0, y > 0$ |
| μ^3 | $-y/(2\mu^2) + 1/\mu$ | $-1/(2\mu^2)$ | Inverse Gaussian | $\mu > 0, y > 0$ |
| $\mu(1 - \mu)$ | $y \log \frac{\mu}{1-\mu} + \log(1 - \mu)$ | $\log \frac{\mu}{1-\mu}$ | Binomial/m | $\mu \in]0; 1[, y \in [0; 1]$ |
| $\mu + \mu^2/k$ | $y \log \frac{\mu}{k+\mu} + k \log \frac{k}{k+\mu}$ | $\log \frac{\mu}{k+\mu}$ | Negative binomial | $\mu > 0, y \geq 0$ |

TABLE 1 – Quasi-vraisemblances associées aux fonctions de variances

Ainsi, la méthode d'estimation que nous allons utiliser tout le long de notre étude sera effectuée à partir de la quasi-vraisemblance. Nous y reviendrons à travers l'étude du jeu de données.

2.2.2 Le modèle linéaire mixte généralisé (GLMM)

Propriétés de base

Le point de départ pour un modèle linéaire (LM) est $E[y] = \theta^\top Z$, θ représentant les paramètres, et Z le vecteur des covariables spatiales des effets fixes du modèle.

Pour un modèle linéaire mixte généralisé (GLMM), nous continuons d'utiliser $\theta^\top Z$ pour les effets fixes, mais ajoutons une partie aléatoire au modèle, noté $v^\top T$, où T , représente une matrice (de covariables spatiales) connue, et v , représente le vecteur des paramètres des effets aléatoires qui se produisent dans les données y . Bien que les éléments de T soient des variables aléatoires, il est convenu dans la littérature scientifique de spécifier le modèle conditionnellement aux valeurs inobservables v .

Ainsi, nous n'écrirons pas $E[y] = \theta^\top Z + v^\top T$, mais :

$$E[y|v] = \theta^\top Z + v^\top T \text{ où } v \sim \text{une loi normale multivariée } (0, G=I\sigma_v^2)$$

Considérons le processus ponctuel de Poisson X , défini sur S et observé dans une fenêtre $W \subseteq S$, avec une fonction d'intensité conditionnellement à v , $\rho(u) = \rho_\theta(u) = e^{(\theta^\top z(u) + v^\top T(u))}$.

2.2.2.1 Vraisemblance

La vraisemblance du processus ponctuel dans ce contexte devient :

$$L_W(\theta, \Gamma) = \int_W \prod_{i=1}^{n(X)} f_{Y_i|v}(y_i|v) f_V(v) dv \text{ où } \Gamma \text{ contient les paramètres de } G$$

En notant y_{ij} le j^{e} individu pris pour le i^{e} groupe, nous pouvons écrire le modèle :

$$\begin{aligned} y_{ij}|v &\sim \text{indep. Poisson}(\mu_{ij}); i = 1, 2, \dots, m; j = 1, 2, \dots, n_i; \\ \log \mu_{ij} &= \theta^\top z_{ij} + v_i; \\ v_i &\sim \text{i.i.d } N(0, \sigma_v^2). \end{aligned}$$

Ce modèle utilise la fonction de lien log et une distribution normale pour les effets aléatoires des groupes i . Les effets aléatoires v_i sont partagés parmi les observations au sein d'un même groupe.

L'écriture de la log vraisemblance peut être écrite ainsi :

$$\log L_W(\theta, \Gamma) = \log \left(\prod_{i=1}^m \int_W \prod_{j=1}^{n_i} \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} v_i^2} dv_i \right)$$

Malheureusement, cette forme ne peut être simplifiée et il est beaucoup plus compliqué de comprendre comment les coefficients des effets fixes et aléatoires vont être estimés. En revanche, nous savons que l'approche de Berman-Turner (discrétisation de l'intralégale par la méthode de schéma en quadrature) va être de nouveau appliquée.

Remarque

Notons que nous connaissons la distribution des y_{ij} conditionnellement aux effets aléatoires (distribution Poisson inhomogène). Nous pouvons donc écrire sa vraisemblance, que nous notons L^* .

$$L_W^*(\theta, \Gamma) = f_{Y_i|v}(y_i|v) = e^{|W| - \int_W \rho_\theta(u) du} \prod_{i=1}^{n(X)} \rho_\theta(x_i)$$

Avec le même principe que dans le cas GLM, en remplaçant $\rho_\theta(u)$ par son expression, on obtient le critère conditionnel à v suivant :

$$\log L_W^*(\theta, \Gamma) = \sum_{i=1}^{n(X)} \log \rho_\theta(x_i) - \int_W \rho_\theta(u) du = \sum_{i=1}^{n(X)} \log e^{(\theta^\top z(x_i) + v^\top T(x_i))} - \int_W e^{(\theta^\top z(u) + v^\top T(u))} du$$

En utilisant de nouveau l'approximation Berman-Turner, on retrouve une nouvelle fois :

$$\log L_W^*(\theta, \Gamma) = \sum_{j=1}^m (y_j \log \rho_\theta(u_j) - \rho_\theta(u_j)) w_j$$

L'approximation de la log vraisemblance de l'intensité conditionnelle par Berman-Turner pour un modèle log-linéaire mixte prend de nouveau la forme d'une log vraisemblance pondérée Poissonnienne.

Sur un plan pratique, et particulièrement en R, des méthodes d'estimation de modèles GLMM sont déjà implantées comme **lme** (du package **nlme**) ou encore **lmer** (du package **lme4**).

3 Application aux données de mouvements oculaires

3.1 Description du jeu de données

Une expérience a été menée où les mouvements oculaires ont été enregistrés pour 5 groupes de participants qui regardent 6 images de scènes naturelles.

- Le premier groupe est composé de 43 bébés de 3 mois
- Le deuxième groupe est composé de 47 bébés de 6 mois
- Le troisième groupe est composé de 44 bébés de 9 mois
- Le quatrième groupe est composé de 46 bébés de 12 mois
- Le cinquième groupe est composé de 44 adultes

Les mouvements oculaires ont été enregistrés en utilisant l'appareil Eyelink1000 (SR Research, <http://www.sr-research.com/>) avec compensation sur les mouvements de tête, enregistrement sur un seul œil (gauche) et avec une fréquence d'échantillonnage (sampling rate) de 250 Hz. Pour chaque sujet, un étalonnage de 5 points a été effectué. Les enfants et les adultes ont passé l'expérience dans les mêmes conditions.

Les sujets sont assis sur un siège à 60 cm de l'écran (40×30 cm, résolution 1024×768), avec sur leur tête un 'sticker cible' pour la compensation des mouvements de tête. Chaque image est projetée pendant 5 secondes, et a été précédée par la visualisation d'une zone test (disque animé de 80 pixels de diamètre au centre-haut de l'écran, avec sons). Cette zone cible est projetée à l'écran jusqu'à ce que le sujet la regarde. Ensuite, l'expérimentateur appuie sur la barre d'espace et la première image apparaît pour une durée de 5 sec puis la zone cible revient pour repositionner la rétine du sujet au centre-haut de l'écran, puis la seconde image apparaît, et ainsi de suite jusqu'à la 6^e image.

Toutes les images, issues d'une série d'image Kodak (<http://www.cipr.rpi.edu/resource/stills/kodak.html>), sont de tailles égales : 1024×768 pixels. En voici un aperçu :



FIGURE 2 – Les 6 images de l'expérience

3.2 Visualisation du jeu de données

David Meary, investigator de l'étude, nous a préalablement préparé un groupe de fichiers contenant les images en format bitmap ainsi que l'ensemble des fixations de tous les sujets pour chaque image.

Il y donc eu un important travail de programmation quant à la bonne lecture de ces données à l'aide des packages **pixmap**, **bmp** et bien entendu le package principal **spatstat** qui nous suivra tout le long de l'étude. Par exemple, pour la lecture d'une image, nous avons eu besoin de passer par un format matriciel de type raster pour lire les données dans **R**. Cet objet R contiendra pour chacun de ses pixels (768×1024) un code hexadécimal, de type `#505050`.

Il est indispensable de noter que nous avons exclu toutes les saccades de notre jeu de données, c'est pour cela que nous parlons durant toute l'analyse de fixations pour désigner celles d'au moins 50 ms.

Ainsi, lors de l'importation des données, il a fallu faire très attention à supprimer les saccades de l'ensemble des fixations observées ou encore de détecter, pour chaque groupe d'âge, les sujets n'ayant pas d'observations sur une image donnée.

L'image suivante représente l'ensemble des fixations observées pour chaque groupe d'âge sur l'image du rafting (voir l'annexe A.1 pour toutes les images).

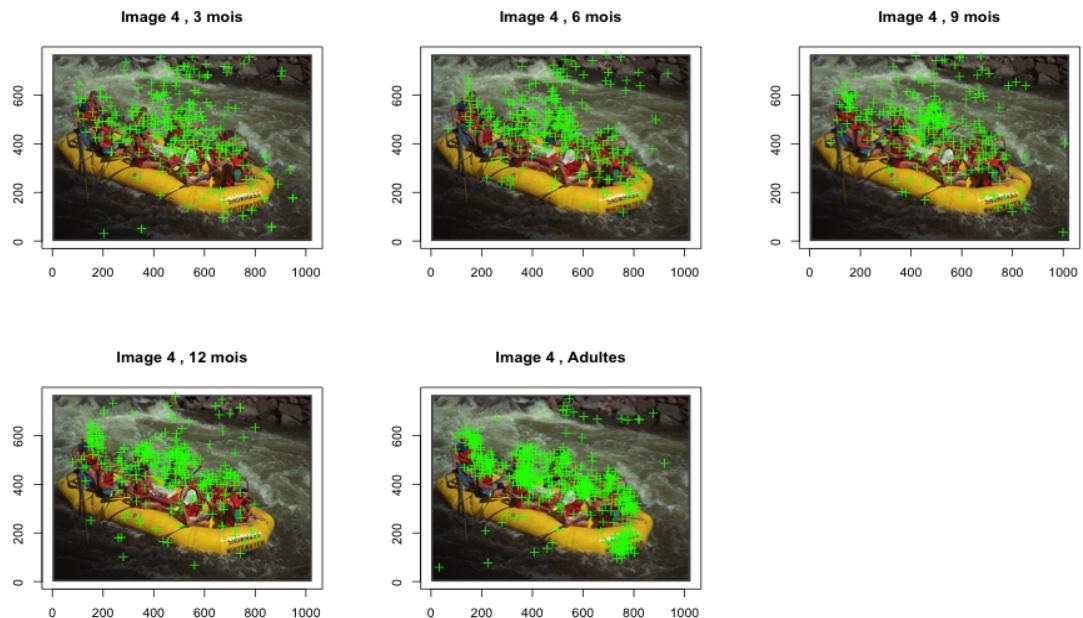


FIGURE 3 – Exemple de fixations par groupes d'âge, image rafting

A première vue, et comme on pouvait s'y attendre, on retrouve une concentration des points autour de zones bien précises définies par les caractéristiques de l'image.

Attention toutefois aux illusions d'optique, il y a beaucoup plus de fixations pour le groupe adulte (de l'ordre de 500) que pour les 4 groupes d'enfants (d'ordre 300) et ce, quelle que soit l'image.

On pourrait également espérer retrouver une concentration des points de plus en plus forte autour de ces zones, plus l'âge du sujet augmentant. Ce sera notre principal objectif et nous construirons un modèle statistique pour comprendre ce phénomène.

Avant cela, concentrons-nous sur la répartition de notre processus ponctuel.

3.3 Distribution du processus ponctuel

Pour chercher à savoir si la distribution de notre processus X (groupes d'âge confondus) peut être de nature Poissonnien homogène, nous devons présenter la fonction K de Ripley, qui est une des fonctions permettant d'analyser des structures de points, et permettant d'effectuer des tests d'hypothèses (sur l'homogénéité Poissonnienne d'un processus dans notre cas).

Définition

La fonction K de Ripley est littéralement définie, pour $r \geq 0$ par :

$$\text{Soit } K(r) = \frac{1}{\rho} E [N(B(0, r) \setminus 0) | 0 \in X]$$

où $N(B(0, r) \setminus 0) | 0 \in X$ représente le nombre d'événements supplémentaires intervenus à l'intérieur d'une boule de rayon r autour d'un événement choisi aléatoirement.

Quant à savoir si le processus est de nature Poisson inhomogène, il existe une fonction appropriée, K_{inhom} , qui testera cette inhomogénéité. Cependant, nous n'écrirons pas dans ce rapport l'écriture de la fonction K_{inhom} de Ripley par sa complexité.

Propriétés

- Dans le cas Poisson homogène, $K(r) = \pi r^2$
- Si $K(r) > \pi r^2$ (respectivement $< \pi r^2$), nous suspecterons un phénomène de clusters (respectivement de répulsion) entre les points de X .
- Les propriétés de la fonction K sont applicables à la fonction K_{inhom} .

3.3.1 Distribution du processus ponctuel général

Test de l'homogénéité poissonienne

Nous allons décrire brièvement 2 méthodes capables de tester l'homogénéité.

La première consiste à utiliser le test des quadrats (**quadrat.test()** du package spatstat). Sans entrer dans le détail, ce test a pour démarche de compter le nombre de points observés dans une division de fenêtre bien précise et comparer ce comptage au nombre de points attendu (analogie au test de khi-deux).

En prenant l'exemple de l'image du rafting, en divisant notre fenêtre en 3×3 , nous obtenons une p-value extrêmement faible (voir la sortie R en annexe A.2).

Visuellement parlant, il est très clair que les points ne sont pas homogénéiquement répartis :

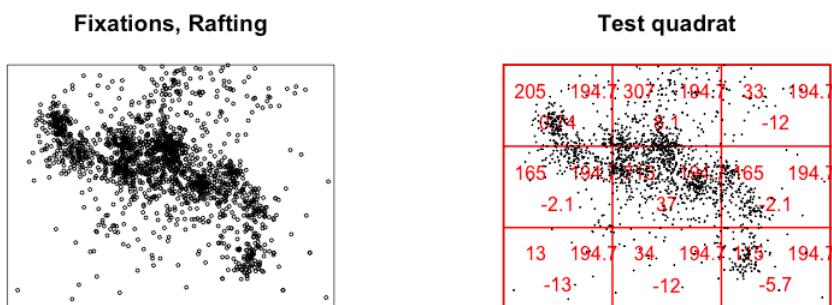


FIGURE 4 – Test des quadrats, image Rafting

La seconde méthode consiste à utiliser la fonction K de Ripley où on a simulé 100 processus de Poisson homogène (H_0) afin de construire des enveloppes de confiance (95%) et de comparer cette courbe théorique à celle observée \hat{K}_{obs} . Si cette dernière est comprise dans les enveloppes de confiance, elle pourra être considérée comme Poisson homogène.

Ce n'est clairement pas vérifié ici dans le cas où $r < 50$ pixels, puisque la courbe \hat{K}_{obs} est largement au-dessus la courbe théorique, c'est à dire d'après nos propriétés, qu'il existe un phénomène de clusters (évident) et ce quelle que soit l'image étudiée :

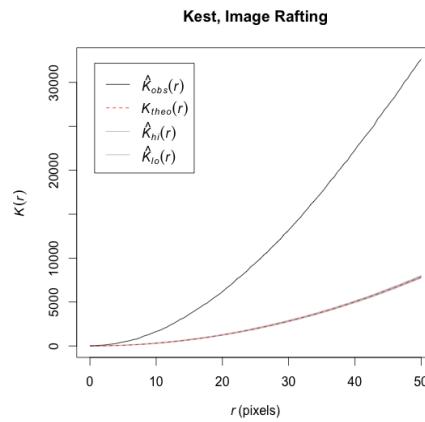


FIGURE 5 – Fonction K de Ripley appliquée à l'image rafting

Test de l'inhomogénéité poissonienne

Nous allons utiliser la fonction K_{inhom} abordée plus haut pour tester l'inhomogénéité du processus de Poisson. Pour cela, comme dans le cas homogène, nous avons simulé un certain nombre de processus de Poisson inhomogène (H_0) et comparé cette courbe théorique à celle observée \hat{K}_{inhom}^{obs} pour un $r < 200$ pixels.

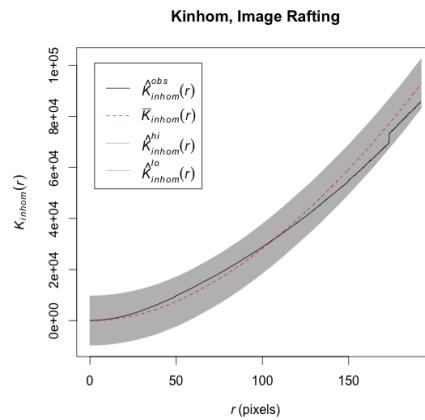


FIGURE 6 – Fonction K_{inhom} de Ripley appliquée à l'image rafting

Ici, notre processus ponctuel peut être considéré comme un processus de Poisson inhomogène. Il admet des zones privilégiées (dues aux caractéristiques de l'image) mais les points qui le composent n'ont pas de dépendance entre eux.

Cette approche est applicable à l'ensemble des images à notre disposition (voir annexe A.3).

3.3.2 Distribution du processus ponctuel par groupes d'âge

En reprenant le même raisonnement que dans le cas général, on obtient des résultats semblables si l'on regarde la distribution des 5 processus ponctuels (un par groupe d'âge).

Test de l'homogénéité poissonienne

En reprenant l'image du rafting et en appliquant par exemple, la fonction K de Ripley sur le groupe 9 mois, on trouve :

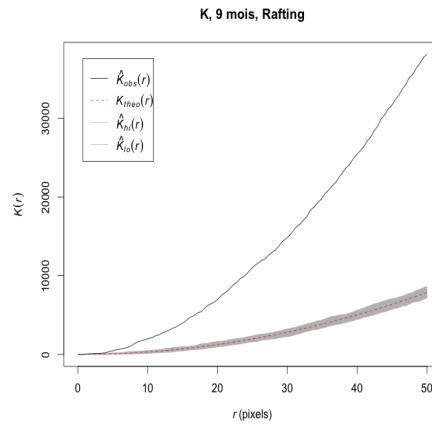


FIGURE 7 – Fonction K de Ripley appliquée aux processus groupe d'âge 9 mois, image rafting

Distribution des 5 processus ponctuels à partir de la fonction K disponibles en annexe A.4.

Test de l'inhomogénéité poissonienne

En appliquant la fonction K de Ripley sur le groupe 9 mois, on obtient les graphiques suivants :

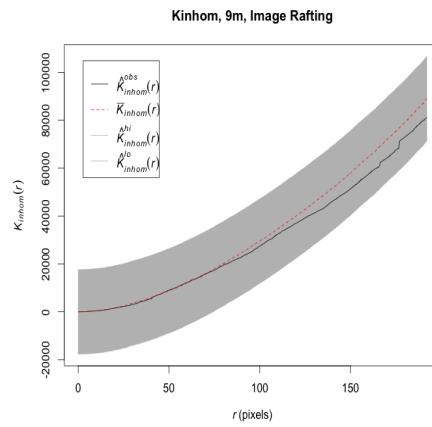


FIGURE 8 – Fonction K_{inhom} de Ripley appliquée aux processus groupe d'âge 9 mois, image rafting

Distribution des 5 processus ponctuels à partir de la fonction K disponibles en annexe A.5.

Ainsi, nos 5 processus ponctuels n'ont clairement pas l'allure de processus ponctuel de Poisson homogène mais ont une distribution qui peut être apparentée à un processus ponctuel de Poisson inhomogène.

Ces analyses peuvent être appliquées à l'ensemble de nos 6 images.

3.4 Covariables spatiales

3.4.1 Cartes de saillance

La principale approche dans la communauté de la recherche sur les mouvements de l'oeil est de produire une carte de saillance qui synthétise toutes les caractéristiques de l'image. Une haute saillance signifie une forte probabilité de fixation. Nous nous sommes servis de cette approche pour modéliser nos processus.

Pour cela David Meary nous a fourni pour chaque image un fichier contenant pour chaque pixel d'une image, 6 niveaux de saillance bien définis. Nous avons donc à disposition les cartes de saillance suivantes :

- Carte de luminance haute fréquence
- Carte de luminance basse fréquence
- Carte chromatique d'opposition Rouge/Verte haute fréquence
- Carte chromatique d'opposition Rouge/Verte basse fréquence
- Carte chromatique d'opposition Bleu/Jaune haute fréquence
- Carte chromatique d'opposition Bleu/Jaune basse fréquence

La luminance est la grandeur mesurable correspondant à la sensation visuelle de luminosité d'une surface. Autrement dit, c'est une grandeur photométrique, dépendante de la sensibilité de l'œil humain.

Visuellement parlant, sur l'image du rafting, nous obtenons :

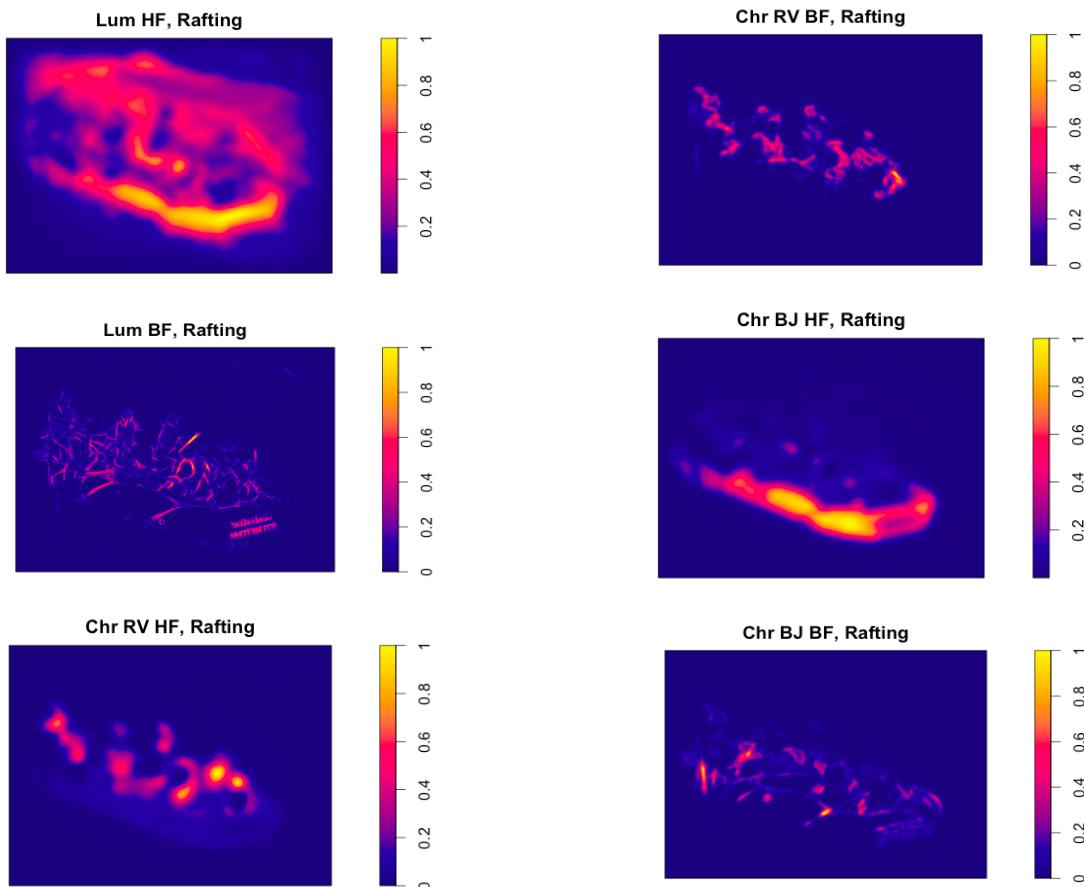


FIGURE 9 – Les 6 niveaux de saillance appliqués à l'image rafting

Pour l'ensemble de nos 6 images, les niveaux de ces 6 saillances sont consultables dans l'annexe A.6.

A partir de ces 6 cartes, il a été construit un niveau de saillance "moyen". En voici une visualisation pour l'ensemble de nos 6 images :

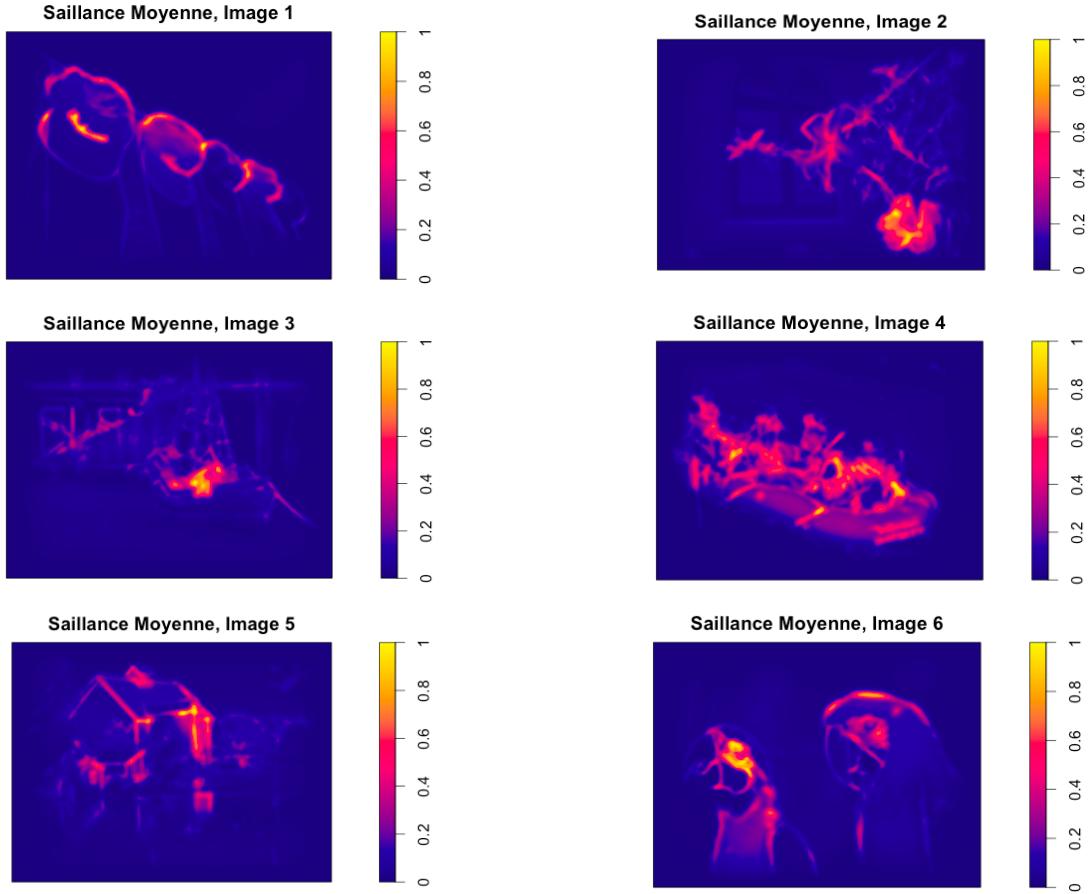


FIGURE 10 – Niveau de saillance moyen appliqué aux 6 images

Ces cartes de saillance seront utilisées pour le modèle que nous allons présenter par la suite. Elles auront le rôle de covariable spatiale dans ce dernier.

Remarque

Constatons que toutes nos cartes de saillance ont été normalisées, c'est à dire que pour chaque pixel de l'image de dimension 1024×768 , une valeur entre 0 et 1 représentant le niveau de saillance lui a été assignée.

Ce principe sera appliquée pour chaque covariable spatiale que nous définirons par la suite.

De cette manière, en mettant à la même échelle l'ensemble de nos covariables spatiales, nous éviterons tout problème d'interprétation.

Notons que, pour chaque image, nous pouvons représenter l'ensemble de nos fixations, en leur associant leur niveau de saillance moyen. Sur l'image du rafting, nous obtenons :

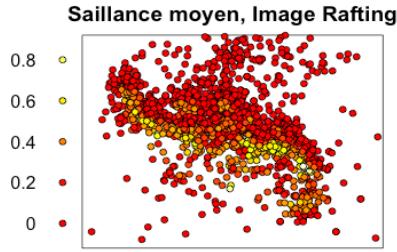


FIGURE 11 – Fixations associées au niveau de saillance moyen, image rafting

Pour l'ensemble des images, nous pourrons aller consulter l'annexe A.7.

Nous pouvons également visualiser le niveau de saillance des fixations de certains groupes en particulier. Sur l'image du rafting, nous obtenons pour les 5 groupes d'âge étudiés :

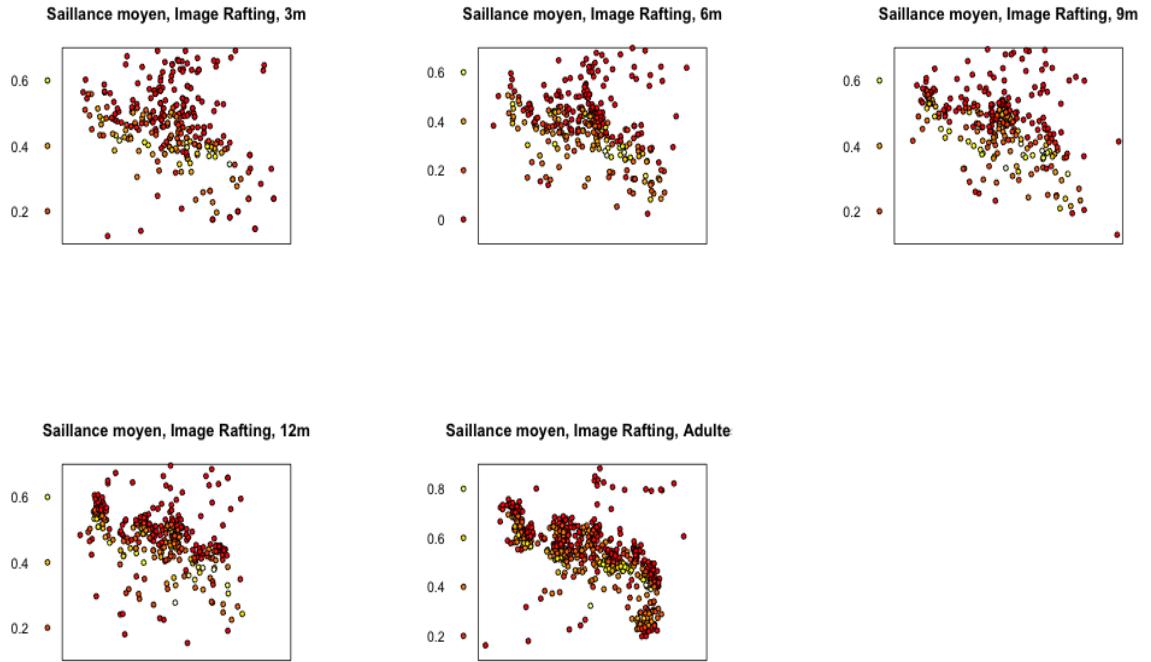


FIGURE 12 – Fixations associées au niveau de saillance moyen, par groupe d'âge, image rafting

3.4.2 Zones cognitives

3.4.2.1 Délimitation de zones cognitives

Nous avons cherché une covariable spatiale permettant de mesurer la variabilité d'acuité visuelle entre les 4 groupes d'âges des bébés. Il semblait naturel de penser que nous devions définir des zones, pour chaque image, afin d'observer une certaine maturation de cette acuité visuelle.

Pour cela, nous définissons une zone cognitive comme étant une zone qui a attiré le plus l'attention d'un adulte. Nous avons donc créé nos zones cognitives, en estimant non paramétriquement l'intensité du groupe adultes (via la fonction **density.ppp** de **R**), puis en sélectionnant 10% des pixels de l'image où l'intensité était la plus forte. Pour l'ensemble de nos 6 images, nous obtenons :

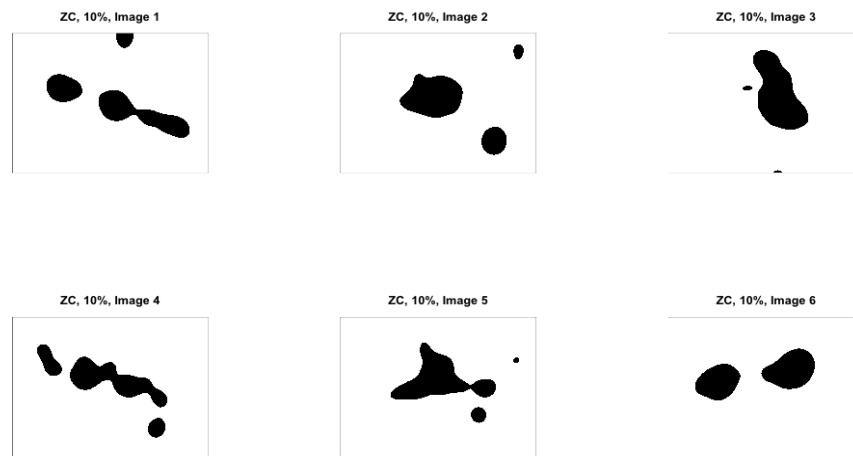


FIGURE 13 – Zones cognitives définies par les 10% des intensités adultes les plus fortes

Avec la même méthode, mais en ne sélectionnant que 5% des pixels les plus intenses.

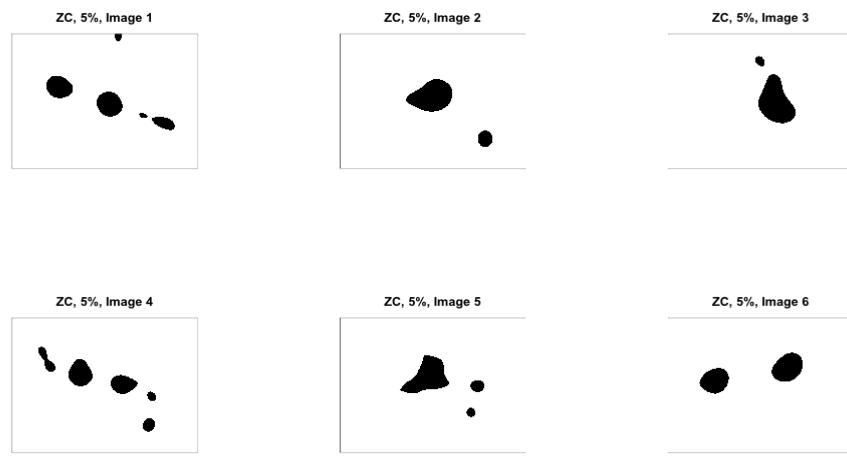


FIGURE 14 – Zones cognitives définies par les 5% des intensités adultes les plus fortes

Par un choix arbitraire, nous décidons de garder cette dernière méthode (top 5 %).

3.4.2.2 Construction d'une carte binaire

Nous avons construit plusieurs covariables spatiales à partir de ces dernières zones, et cherchons à sélectionner celle qui pourrait le mieux expliquer, d'un point de vue statistique, les différences d'acuité visuelle entre les différents groupes d'âge.

Ces covariables ont été réalisées à partir :

1. Des distances entre chaque fixation et la plus proche zone cognitive
2. D'une condition binaire (1 si la fixation est dans la zone, 0 sinon)
3. Des intensités normalisées estimées non-paramétriquement

Ainsi, pour chaque image, nous avons à disposition 3 cartes. Pour l'image du rafting, nous obtenons les covariables spatiales suivantes :

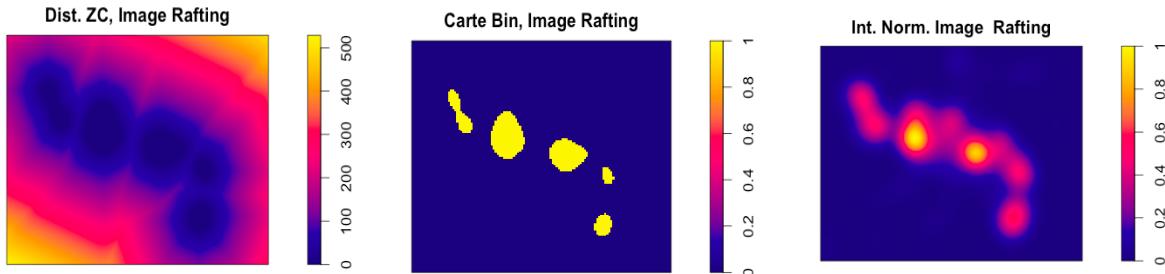


FIGURE 15 – Choix parmi 3 covariables spatiales, image Rafting

Même principe applicable à l'ensemble des images (Annexe A.8, A.9 et A.10).

A partir de résultats que nous ne présenterons pas, il s'est avéré qu'après modélisation, ces 3 cartes avaient une importance plus ou moins similaire dans l'interprétation de résultats. Il n'y a donc pas de différences significatives dans le choix de cette covariable spatiale.

Par un nouveau choix arbitraire, nous décidons de choisir le second cas : nous allons retenir la carte binaire construite à partir des 5% des pixels les plus intenses obtenus par l'intensité non-paramétrique du groupe adultes.

3.5 Analyse des données oculométriques par un modèle log-linéaire

Il est important de souligner à ce stade que les modèles que nous allons décrire sont entièrement de nature statistique : ils ne présagent en rien le comportement des mouvements oculaires générés par le cerveau. En ce sens, nos modèles sont plus proches des modèles de régression linéaire que, par exemple, des modèles biologiquement inspirés de lecture et recherche visuelle approfondies.

3.5.1 A la découverte des fonctions ppm et mppm

Au sein du package **spatstat**, il existe 2 fonctions capables d'estimer des modèles de processus ponctuels à partir de points observés.

Commençons par présenter les objets sur lesquels nous allons travailler.

Reprenons l'image du rafting, en se concentrant sur la répartition des points du groupe adultes.

```
str(X_adults_Im[[4]])  
  
## List of 6  
## $ window :List of 4  
##   ..$ type : chr "rectangle"  
##   ..$ xrange: num [1:2] 0 1024  
##   ..$ yrange: num [1:2] 0 768  
##   ..$ units :List of 3  
##     ..$ singular : chr "pixels"  
##     ..$ plural : chr "pixels"  
##     ..$ multiplier: num 1  
##     ..- attr(*, "class")= chr "units"  
##     ..- attr(*, "class")= chr "owin"  
## $ n : int 631  
## $ x : int [1:631] 443 519 368 365 222 213 154 386 367 687 ...  
## $ y : num [1:631] 348 452 421 451 487 501 582 544 531 315 ...  
## $ markformat: chr "vector"  
## $ marks : Factor w/ 43 levels "Sujet 181","Sujet 182",... : 1 1 1 1 1 1 1 1 1  
## - attr(*, "class")= chr "ppp"
```

On constate qu'il y a 43 processus ponctuels (un par sujet) cachés derrière les 631 points observés. Ainsi, le processus ponctuel non marqué sera le processus composé de l'ensemble des points observés alors que le processus ponctuel marqué (par sujets) sera composé des 43 processus ponctuels.

La fonction **mppm** nécessite dans un premier temps de construire des hyperframes :

```
## [[1]]  
## Hyperframe:  
##   ppp group ZCBin sail n ## Hyperframe:  
## 1 (ppp) 3mois (im) (im) 309 ## 1 (ppp) 3mois (im) (im) 260 ## Hyperframe:  
## 2 (ppp) 6mois (im) (im) 338 ## 2 (ppp) 6mois (im) (im) 223 ## 2 (ppp) 6mois (im) (im) 276  
## 3 (ppp) 9mois (im) (im) 299 ## 3 (ppp) 9mois (im) (im) 240 ## 3 (ppp) 9mois (im) (im) 228  
## 4 (ppp) 12mois (im) (im) 323 ## 4 (ppp) 12mois (im) (im) 269 ## 4 (ppp) 12mois (im) (im) 355  
## 5 (ppp) Adulte (im) (im) 632 ## 5 (ppp) Adulte (im) (im) 583 ## 5 (ppp) Adulte (im) (im) 609  
##  
## [[2]]  
## Hyperframe:  
##   ppp group ZCBin sail n ## Hyperframe:  
## 1 (ppp) 3mois (im) (im) 265 ## 1 (ppp) 3mois (im) (im) 288 ## Hyperframe:  
## 2 (ppp) 6mois (im) (im) 279 ## 2 (ppp) 6mois (im) (im) 292 ## 2 (ppp) 6mois (im) (im) 294  
## 3 (ppp) 9mois (im) (im) 246 ## 3 (ppp) 9mois (im) (im) 280 ## 3 (ppp) 9mois (im) (im) 274  
## 4 (ppp) 12mois (im) (im) 318 ## 4 (ppp) 12mois (im) (im) 334 ## 4 (ppp) 12mois (im) (im) 354  
## 5 (ppp) Adulte (im) (im) 561 ## 5 (ppp) Adulte (im) (im) 631 ## 5 (ppp) Adulte (im) (im) 549  
##  
## [[3]]  
## Hyperframe:  
##   ppp group ZCBin sail n ## Hyperframe:  
## 1 (ppp) 3mois (im) (im) 260 ## 1 (ppp) 3mois (im) (im) 265 ## Hyperframe:  
## 2 (ppp) 6mois (im) (im) 276 ## 2 (ppp) 6mois (im) (im) 276 ## 2 (ppp) 6mois (im) (im) 276  
## 3 (ppp) 9mois (im) (im) 228 ## 3 (ppp) 9mois (im) (im) 228 ## 3 (ppp) 9mois (im) (im) 228  
## 4 (ppp) 12mois (im) (im) 355 ## 4 (ppp) 12mois (im) (im) 355 ## 4 (ppp) 12mois (im) (im) 355  
## 5 (ppp) Adulte (im) (im) 609 ## 5 (ppp) Adulte (im) (im) 609 ## 5 (ppp) Adulte (im) (im) 609  
##  
## [[4]]  
## Hyperframe:  
##   ppp group ZCBin sail n ## Hyperframe:  
## 1 (ppp) 3mois (im) (im) 288 ## 1 (ppp) 3mois (im) (im) 288 ## Hyperframe:  
## 2 (ppp) 6mois (im) (im) 294 ## 2 (ppp) 6mois (im) (im) 294 ## 2 (ppp) 6mois (im) (im) 294  
## 3 (ppp) 9mois (im) (im) 274 ## 3 (ppp) 9mois (im) (im) 274 ## 3 (ppp) 9mois (im) (im) 274  
## 4 (ppp) 12mois (im) (im) 354 ## 4 (ppp) 12mois (im) (im) 354 ## 4 (ppp) 12mois (im) (im) 354  
## 5 (ppp) Adulte (im) (im) 549 ## 5 (ppp) Adulte (im) (im) 549 ## 5 (ppp) Adulte (im) (im) 549  
##  
## [[5]]  
## Hyperframe:  
##   ppp group ZCBin sail n ## Hyperframe:  
## 1 (ppp) 3mois (im) (im) 265 ## 1 (ppp) 3mois (im) (im) 265 ## Hyperframe:  
## 2 (ppp) 6mois (im) (im) 276 ## 2 (ppp) 6mois (im) (im) 276 ## 2 (ppp) 6mois (im) (im) 276  
## 3 (ppp) 9mois (im) (im) 228 ## 3 (ppp) 9mois (im) (im) 228 ## 3 (ppp) 9mois (im) (im) 228  
## 4 (ppp) 12mois (im) (im) 355 ## 4 (ppp) 12mois (im) (im) 355 ## 4 (ppp) 12mois (im) (im) 355  
## 5 (ppp) Adulte (im) (im) 609 ## 5 (ppp) Adulte (im) (im) 609 ## 5 (ppp) Adulte (im) (im) 609  
##  
## [[6]]  
## Hyperframe:  
##   ppp group ZCBin sail n ## Hyperframe:  
## 1 (ppp) 3mois (im) (im) 288 ## 1 (ppp) 3mois (im) (im) 288 ## Hyperframe:  
## 2 (ppp) 6mois (im) (im) 294 ## 2 (ppp) 6mois (im) (im) 294 ## 2 (ppp) 6mois (im) (im) 294  
## 3 (ppp) 9mois (im) (im) 274 ## 3 (ppp) 9mois (im) (im) 274 ## 3 (ppp) 9mois (im) (im) 274  
## 4 (ppp) 12mois (im) (im) 354 ## 4 (ppp) 12mois (im) (im) 354 ## 4 (ppp) 12mois (im) (im) 354  
## 5 (ppp) Adulte (im) (im) 549 ## 5 (ppp) Adulte (im) (im) 549 ## 5 (ppp) Adulte (im) (im) 549
```

Pour chacun des 6 hyperframes, on retrouve pour chaque ligne un processus ponctuel, ainsi que les cartes (saillance, zones cognitives) construites précédemment.

On notera qu'un tel objet R pourra facilement être très volumineux, puisque nous travaillons sur des cartes pixelisées en guise de covariables spatiales. Tout particulièrement lorsque nous travaillerons sur les hyperframes marquées par sujet.

On a voulu savoir si les 2 fonctions estimaient de manière semblable un modèle où seul l'intercept joue un rôle de variable explicative.

```

fit1 <- ppm(unmark(X_adults_Im[[4]]),~1) # PPM non marqué par sujet
fit2 <- mpmm(ppp~1,data=hyp.Im[[4]][5]) # MPPM non marqué par sujet
fit3 <- ppm(X_adults_Im[[4]],~1) # PPM marqué par sujet
c(summary(fit1)$coe[1,1],
  summary(fit2)$Fit$coeff[1,1],
  summary(fit3)$coe[1,1]+log(length(levels(X_adults_Im[[4]]$marks))))
## [1] -7.127956 -7.127956 -7.127956

```

En comparant les 2 méthodes dans le cas de processus non marqué, on constate que les coefficients estimés sont les mêmes. En travaillant sur le processus marqué par sujets, on obtient une différence de $\log(\text{nombre de marques})$ par rapport au cas non marqué.

Comme nous allons utiliser par la suite, les processus ponctuels marqués par sujets dans le cadre du modèle log-linéaire mixte, nous nous concentrerons sur la fonction **mpmm** au détriment de **ppm**.

Remarque

En cherchant à estimer ces modèles à partir de covariables spatiales, les coefficients estimés sur celles-ci seront les mêmes quelle que soit la nature du processus (marqué ou non). Seule la valeur estimée de l'intercept variera en fonction du nombre de marques.

3.5.2 Utilisation de la fonction MPPM aux données

3.5.2.1 Ecriture du modèle

Dans cette partie, notre objectif est de pouvoir évaluer les différences d'intensités (différences d'emplacement des fixations) entre les 4 groupes d'âge des bébés. Pour cela, pour chaque image, nous allons construire le modèle suivant :

$$\log \rho(u, m, \theta) = \theta_1 Saillance(u) + \sum_{m'=1}^4 (\theta_0^{m'} + \theta_1^{m'} ZCBin(u)) \mathbb{1}_{m'=m}$$

où :

- u est un point de R^2
- m est un groupe d'âge
- $Saillance$ et $ZCBin$ sont les covariables spatiales précédemment décrites

Nous n'avons pas représenté le groupe adultes, puisque nous avons construit la variable $ZCBin$ à partir des données de ce groupe, qui nous sert de groupe référence.

En appliquant la fonction d'estimation choisie précédemment, nous construisons les modèles :

```

fit <- list()
for(i in 1:6) {
  fit[[i]] <- mpmm(ppp~sail+group*ZCBin-ZCBin,data=hyp.Im[[i]])
}

```

Pour l'image du rafting, la sortie R nous donne l'estimation des coefficients suivants :

```
summary(fit[[4]])$coef

##      (Intercept)          sail    group3mois   group6mois
## -8.23463633  2.03504193 -0.01874119 -0.05988180
## group9mois group12mois:ZCBin group3mois:ZCBin group6mois:ZCBin
## -0.07942860  1.44696303  0.85964130  1.17161475
## group9mois:ZCBin
## 1.01592721
```

Voir la totalité du résumé statistique dans l'annexe A.11.

Notons que nous avons 9 coefficients par sortie graphique : la covariable *sail*, les 4 interactions entre le groupe et la covariable *ZCBin* et les 4 intercepts isolés pour chaque groupe. Notons également que nous retrouvons la méthode d'estimation utilisée par la fonction **mppm**, la quasi-vraisemblance.

```
## Call:
## glm(formula = fmla, family = quasi(link = log, variance = mu),
##     data = moadf, weights = .mpl.W, subset = (.mpl.SUBSET ==
##     "TRUE"), control = glm.control(maxit = 50))
```

Sur les 9 coefficients que nous avons à disposition, nous allons retenir seulement les 4 nous permettant de mesurer la variabilité entre les groupes d'âge. En traçant, à l'aide du package **plotrix**, sur un même graphique les coefficients estimés avec leur intervalle de confiance (à 95%), nous obtenons :

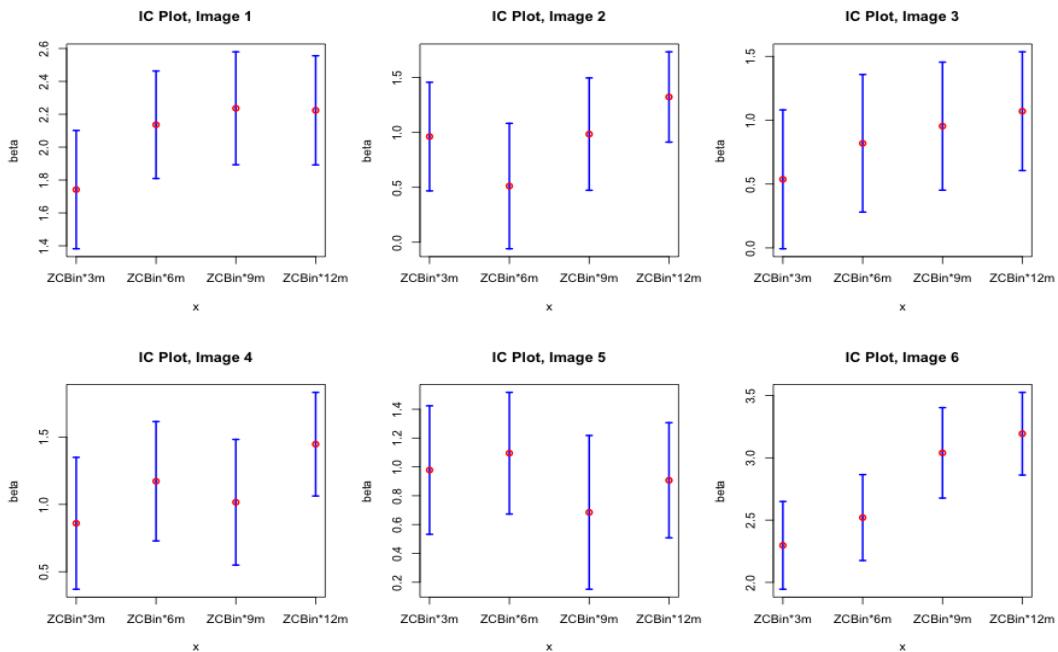


FIGURE 16 – IC Plot appliquée aux 6 images

Par la nature de la construction de la carte binaire *ZCBin* (1 dans la zone cognitive, 0 en dehors), il est naturel d'espérer retrouver une hausse (plus ou moins graduelle) des coefficients du groupe 3 mois au groupe 12 mois.

A première vue, l'image 6 (perroquets) semble être la plus pertinente à analyser.

D'un point de vue statistique, il est nécessaire d'appliquer des tests pour comparer les différences d'intensités entre certains groupes.

3.5.2.2 Tests multiples

Nous avons voulu savoir s'il existait des différences significatives dans 3 situations :

(1) groupe 12 mois contre groupe 3 mois, (2) groupe 9 et 12 mois réunis contre groupe 3 et 6 mois réunis et (3) groupe 12 mois contre groupe 3, 6 et 9 mois réunis.

Nous ne devons pas tester individuellement ces 3 cas, mais bien poser une hypothèse H_0 commune comprenant les 3 hypothèses à tester. Il existe des procédures qui permettent de contrôler le fait que l'on fait 3 tests en même temps, et qui en particulier vont réhausser les p-valeurs que l'on observerait si on faisait les tests individuellement.

En effet, lorsque l'on fait des tests de contraste corrigés image par image, nous faisons 3 tests et les p-valeurs de ces trois tests sont corrigées selon un critère d'ajustement, qui est par défaut le FWER (familywiser error rate).

Cette stratégie de correction permet d'interpréter de manière plus fiable la valeur des p-valeurs obtenues.

Pour cela, nous allons utiliser des outils du package **multcomp**. Nous construisons dans un premier temps une matrice K, en définissant des poids (sur les coefficients associés selon ce que nous souhaitons tester) puis nous appellons cette matrice à l'aide de la fonction **glht** du package, qui réalise les tests simultanés. Pour chacun des 3 cas, nous ferons un test unilatéral, en testant la supériorité du groupe contenant le groupe d'âge 12 mois (voir procédures R annexe A.12).

Pour l'image rafting, nous obtenons un test significatif si nous comparons le groupe 12 mois au 3 mois ou encore en comparant le groupe 12 mois aux groupes 3, 6 et 9 mois réunis.

```
## [[4]]
##
##   Simultaneous Tests for General Linear Hypotheses
##
##   Linear Hypotheses:
##                   Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0      0.5873    0.2016  2.914 0.00425 **
## 9/12m - 3/6m <= 0  0.4316    0.2904  1.487 0.13843
## 12m-3/6/9m <= 0   1.2937    0.4538  2.851 0.00527 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

En se concentrant sur l'image des perroquets, on s'aperçoit que les fixations se concentrent de plus en plus autour de la zone cognitive plus l'âge du bébé augmente.

```
## [[6]]
##
##   Simultaneous Tests for General Linear Hypotheses
##
##   Linear Hypotheses:
##                   Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0      0.8966    0.1633  5.489 < 1e-06 ***
## 9/12m - 3/6m <= 0  1.4160    0.2360  6.001 < 1e-06 ***
## 12m-3/6/9m <= 0   1.7235    0.3941  4.373 1.06e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

En revanche, pour l'image 5, les tests sont tous non-significatifs (p-valeurs toutes supérieures à 70%).

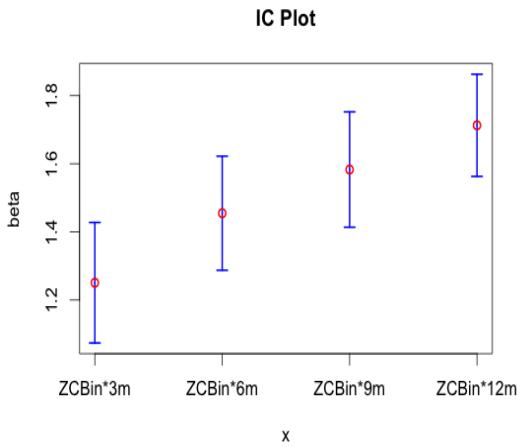
```
## [[5]]
##
##   Simultaneous Tests for General Linear Hypotheses
##
##   Linear Hypotheses:
##                   Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0      -0.07103   0.21975 -0.323  0.819
## 9/12m - 3/6m <= 0 -0.48153   0.33275 -1.447  0.985
## 12m-3/6/9m <= 0   -0.03643   0.53186 -0.068  0.736
## (Adjusted p values reported -- single-step method)
```

En conclusion, nous pouvons dire d'après les tests, que les caractéristiques de certaines images, comme celle des perroquets, nous permettent de dire que l'acuité visuelle des bébés se rapproche de celle de l'adulte, l'âge du sujet augmentant. Ce n'est clairement pas le cas de toutes les images.

L'ensemble des résultats de ces 3 tests sur les 6 images sont disponibles dans l'annexe A.13.

Images regroupées

En souhaitant obtenir les résultats pour les 6 images réunies, il est nécessaire de construire une unique hyperframe (et non plus une liste d'hyperframes), noté **hyp1** (voir annexe A.14), comprenant l'ensemble des fixations de tous les groupes sur les 6 images. En appliquant le modèle à cette nouvelle hyperframe, nous obtenons l'IC plot suivant à partir du résumé du modèle obtenu (voir annexe A.15) :



En effectuant le même test, nous obtenons de bons résultats :

```
##   Simultaneous Tests for General Linear Hypotheses
##
##   Linear Hypotheses:
##                   Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0      0.46213   0.07533  6.135 <1e-06 ***
## 9/12m - 3/6m <= 0  0.59036   0.10820  5.456 <1e-06 ***
## 12m-3/6/9m <= 0   0.85010   0.17411  4.882 <1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Remarque

Nous devons cependant être critiques par rapport au dernier test, puisque les p-valeurs ne sont pas corrigées en tenant compte de l'hyperframe construit (images confondues).

3.6 Analyse des données oculométriques par un modèle log-linéaire mixte

Effet aléatoire sujet

Ici, nous cherchons à savoir s'il existe des différences significatives entre l'acuité visuelle entre les sujets d'un même groupe sur une image donnée.

Pour cela, nous avons besoin de construire une nouvelle hyperframe volumineuse, puisque nous obtenons pour toutes les images, pour tous les groupes, l'ensemble des fixations détaillé de tous les sujets. Ainsi, nous obtenons une hyperframe comprenant 1344 lignes (processus ponctuels).

```
hyp.glob[c(1:3, 645:647, 1342:1344),]

## Hyperframe:
##      ppp     groupe   image    sujet sailmoy ZCBin n
## 1  (ppp) Baby3month Image1  Sujet1   (im)  (im) 11
## 2  (ppp) Baby3month Image1  Sujet2   (im)  (im)  2
## 3  (ppp) Baby3month Image1  Sujet3   (im)  (im)  7
## 645 (ppp) Baby9month Image3 Sujet107 (im)  (im)  7
## 646 (ppp) Baby9month Image3 Sujet108 (im)  (im)  0
## 647 (ppp) Baby9month Image3 Sujet109 (im)  (im) 11
## 1342 (ppp)      Adult Image6 Sujet222 (im)  (im) 11
## 1343 (ppp)      Adult Image6 Sujet223 (im)  (im)  9
## 1344 (ppp)      Adult Image6 Sujet224 (im)  (im) 12
```

Dans cette sélection, nous pouvons voir que certains sujets n'ont aucune fixation sur certaines images.

Ainsi, sur l'image rafting, à l'appel de la fonction **mppm**, en mettant un effet aléatoire du sujet sur les covariables et en ne tenant pas compte des sujets n'ayant aucune fixation sur l'image, nous obtenons :

```
fitai <- mppm(ppp~sailmoy+ZCBin, random=~sailmoy+ZCBin|sujet,
                 data=subset(hyp.glob[hyp.glob$groupe=="Baby3month" & hyp.glob$image=="Image4"], npoints(ppp)>0))
```

On constate que les effets aléatoires des sujets sont très faibles, et ce quelle que soit la covariable retenue.

```
random.effects(fitai)[2:10,]

##          (Intercept)      sailmoy      ZCBin
## Sujet10  5.736569e-09 -4.050660e-08 -1.030605e-05
## Sujet11 -6.440928e-09  2.427843e-08  1.825700e-05
## Sujet12 -2.165114e-09  1.404328e-08  8.600959e-06
## Sujet13  1.541687e-09  1.202146e-08 -9.173356e-07
## Sujet14 -2.675526e-09  6.671103e-09  8.591307e-06
## Sujet15 -1.054407e-09 -6.585642e-08 -8.123355e-07
## Sujet16 -1.556714e-09 -4.633904e-08 -8.713266e-07
## Sujet17  5.278193e-10 -5.754879e-09 -9.544204e-07
## Sujet18 -4.188996e-09 -3.305945e-08  8.517291e-06
```

De plus, à la vue de la matrice de variance-covariance résiduelle, on s'aperçoit que la variabilité résiduelle est extrêmement faible.

```
## Random effects:
## Formula: ~sailmoy + ZCBin | sujet
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev     Corr
## (Intercept) 2.276325e-05 (Intr) sailmy
## sailmoy     3.532580e-04 -0.005
## ZCBin       3.077865e-03 -0.053  0.002
## Residual    1.000000e+00
```

En conclusion, nous ne pouvons pas déceler de différences significatives entre les sujets d'un même groupe pour une image donnée.

Résumé

Les processus ponctuels spatiaux sont des modèles stochastiques faisant intégrer la géométrie stochastique et la statistique spatiale. Des techniques d'estimation (paramétriques) de la fonction d'intensité, noté $\rho(u; \theta)$, (interprété comme la probabilité d'observer un point au voisinage de u) sont connues et des modèles de la forme suivante peuvent donc être approchés :

$$\log \rho(u; \theta) = \theta_0 + \theta_1 z_1(u) + \dots + \theta_p z_p(u)$$

où $\theta = (\theta_0, \dots, \theta_p)^\top$ est le vecteur des estimations et z_1, \dots, z_p les covariables spatiales

Durant le stage, sont utilisées des données oculométriques (fixations oculaires de la rétine) observées pour différents sujets, différents groupes de sujets (5), et différentes images (6).

Les mouvements oculaires sont des alternances de saccades (20-50ms) et de fixations (250-300ms). L'emplacement des fixations est en partie guidé par les caractéristiques locales de l'image : contrastes, bordures et couleurs.

A l'aide du package **spatstat** implémenté dans R, nous avons pu déterminer que la distribution des points était de type Poisson inhomogène et qu'il n'existe pas de structures particulières (de type clusters ou répulsion) entre les points. On note simplement des zones privilégiées liées aux caractéristiques de l'image.

Nous avons ensuite construit des covariables spatiales pouvant expliquer au mieux l'intensité : le niveau de saillance moyen (*Saillance*) et la construction d'une carte binaire (*ZCBin*) correspondant à la délimitation de zones cognitives à partir des 5% des pixels les plus intenses, estimées non paramétriquement à partir du groupe adulte.

Notre jeu de données comprend donc l'ensemble des fixations (saccades exclues) de l'ensemble de nos sujets sur les 6 images et de différentes cartes pixelisées définissant nos covariables spatiales.

Pour répondre à notre problématique, nous avons comparé les intensités pour les différents groupes d'âge des bébés (au nombre de 4), pour une image donnée, en retenant le modèle log-linéaire suivant :

$$\log \rho(u, m, \theta) = \theta_1 Saillance(u) + \sum_{m'=1}^4 \left(\theta_0^{m'} + \theta_1^{m'} ZCBin(u) \right) \mathbb{1}_{m'=m}$$

où :

- u est un point de R^2
- m est un groupe d'âge
- *Saillance* et *ZCBin* sont les covariables spatiales retenues

A l'aide de tests statistiques, nous obtenons des résultats variables selon l'image retenue. Certaines images montrent un mouvement net des fixations se déplaçant vers les zones cognitives plus l'âge du bébé évolue. Sur d'autres images, nous ne pouvons rien déceler de significatif.

Enfin, nous avons voulu savoir s'il était possible de comparer les intensités des sujets au sein d'un même groupe pour une image donnée. En appliquant un modèle de type log-linéaire mixte, à la vue des coefficients aléatoires et de la matrice de variance-covariance qui en résultent, nous en avons conclu qu'une grande variabilité entre sujets ne pouvait pas être retenue.

Perspectives

Afin d'affiner encore mieux nos modèles, d'autres sujets de recherche sur les mouvements de l'oeil pourraient être encore approfondis. Par exemple, sur nos données, serions-nous capables d'estimer l'intensité de nos processus ponctuels en ajoutant un effet aléatoire de l'image sur certains groupes de covariables. Aujourd'hui, nous n'en sommes pas capables pour des problèmes de non-convergence de l'algorithme de maximisation de la vraisemblance.

A Annexes

A.1 Fixations par groupe d'âge, pour toutes les images

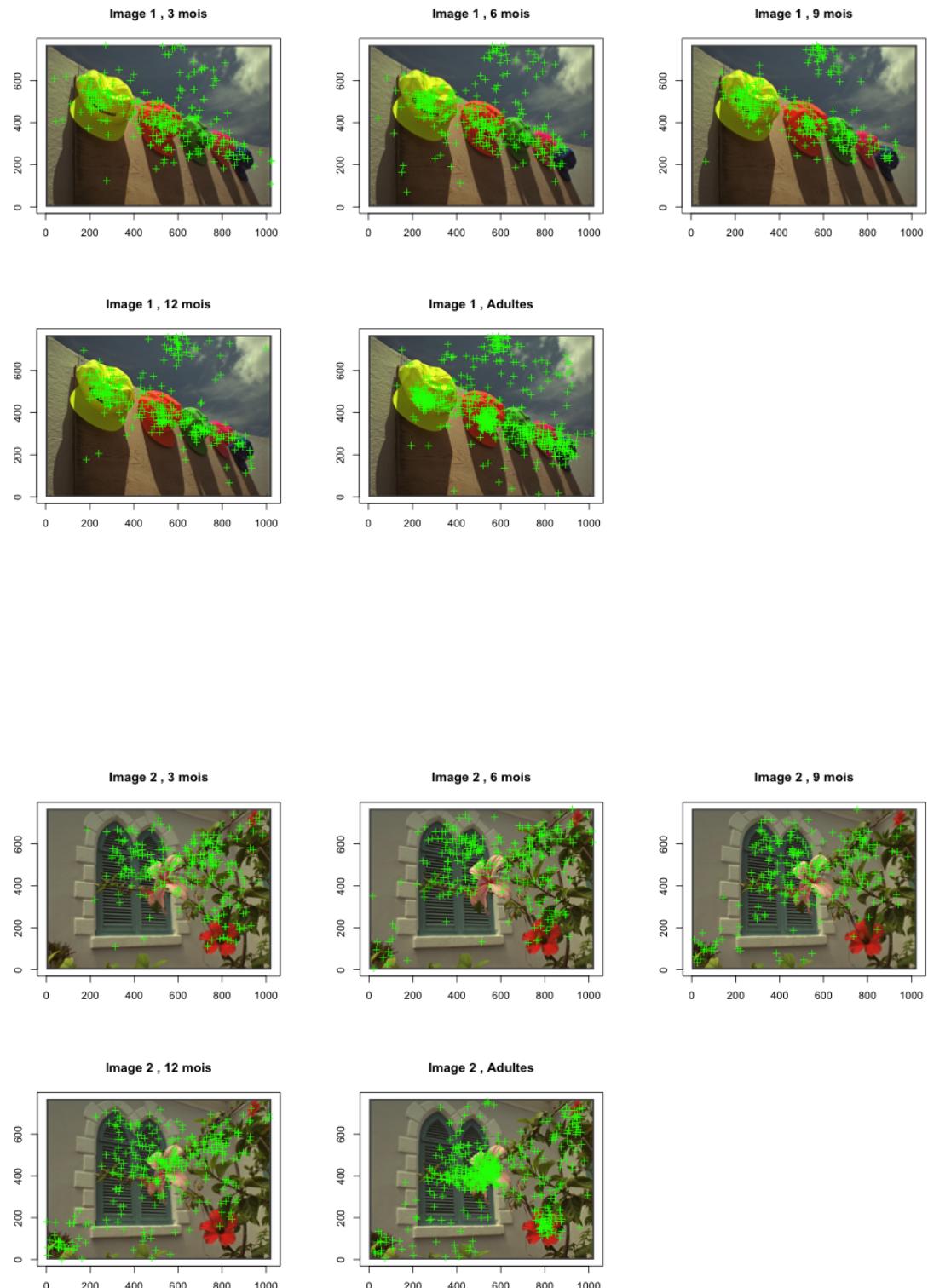


Image 3 , 3 mois

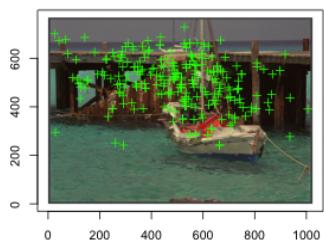


Image 3 , 6 mois

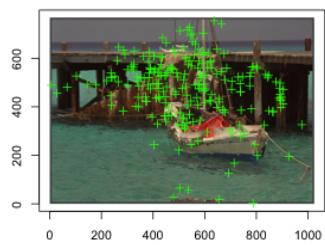


Image 3 , 9 mois

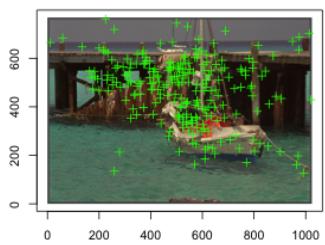


Image 3 , 12 mois

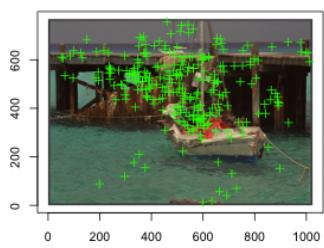


Image 3 , Adultes

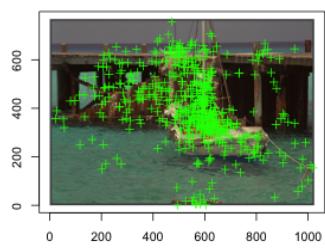


Image 4 , 3 mois

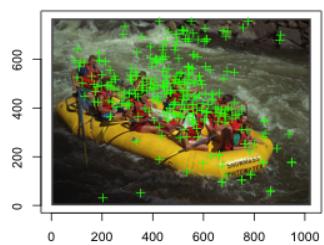


Image 4 , 6 mois

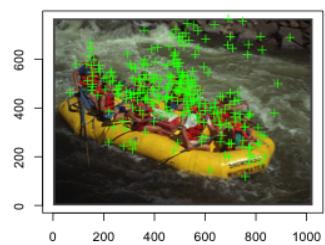


Image 4 , 9 mois

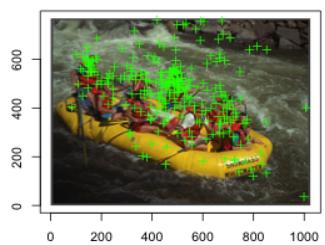


Image 4 , 12 mois

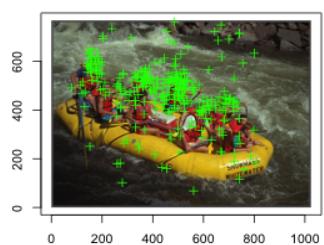
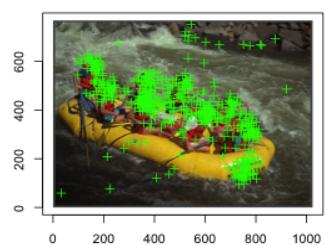
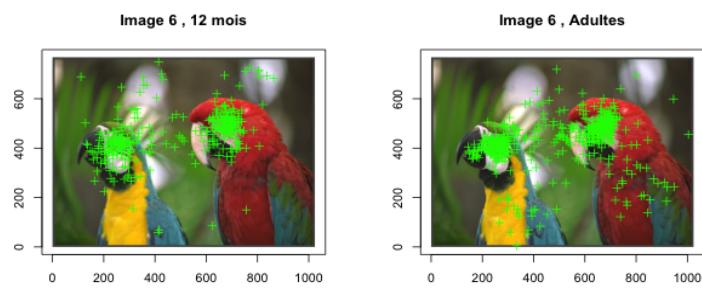
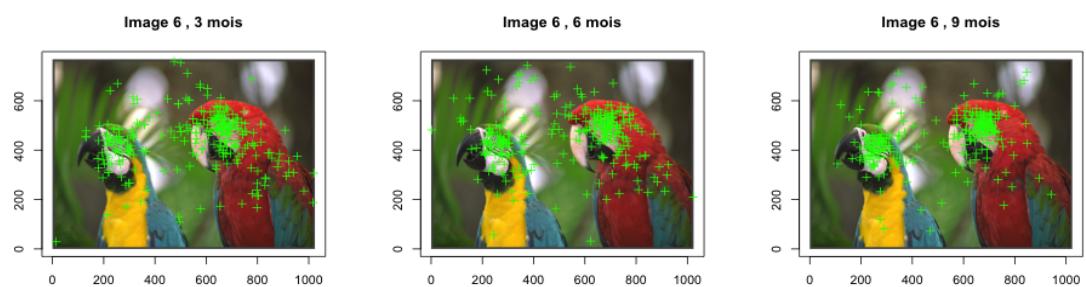
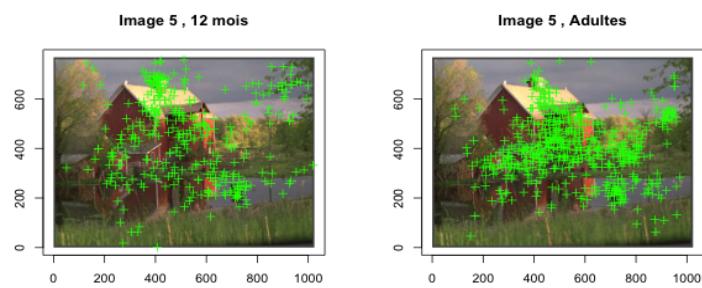
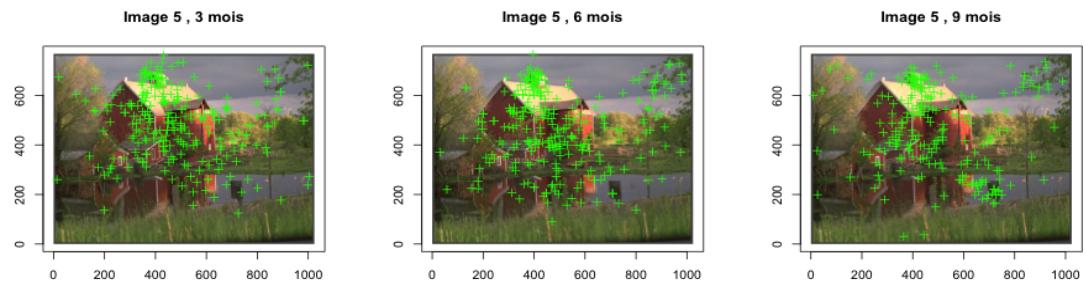


Image 4 , Adultes



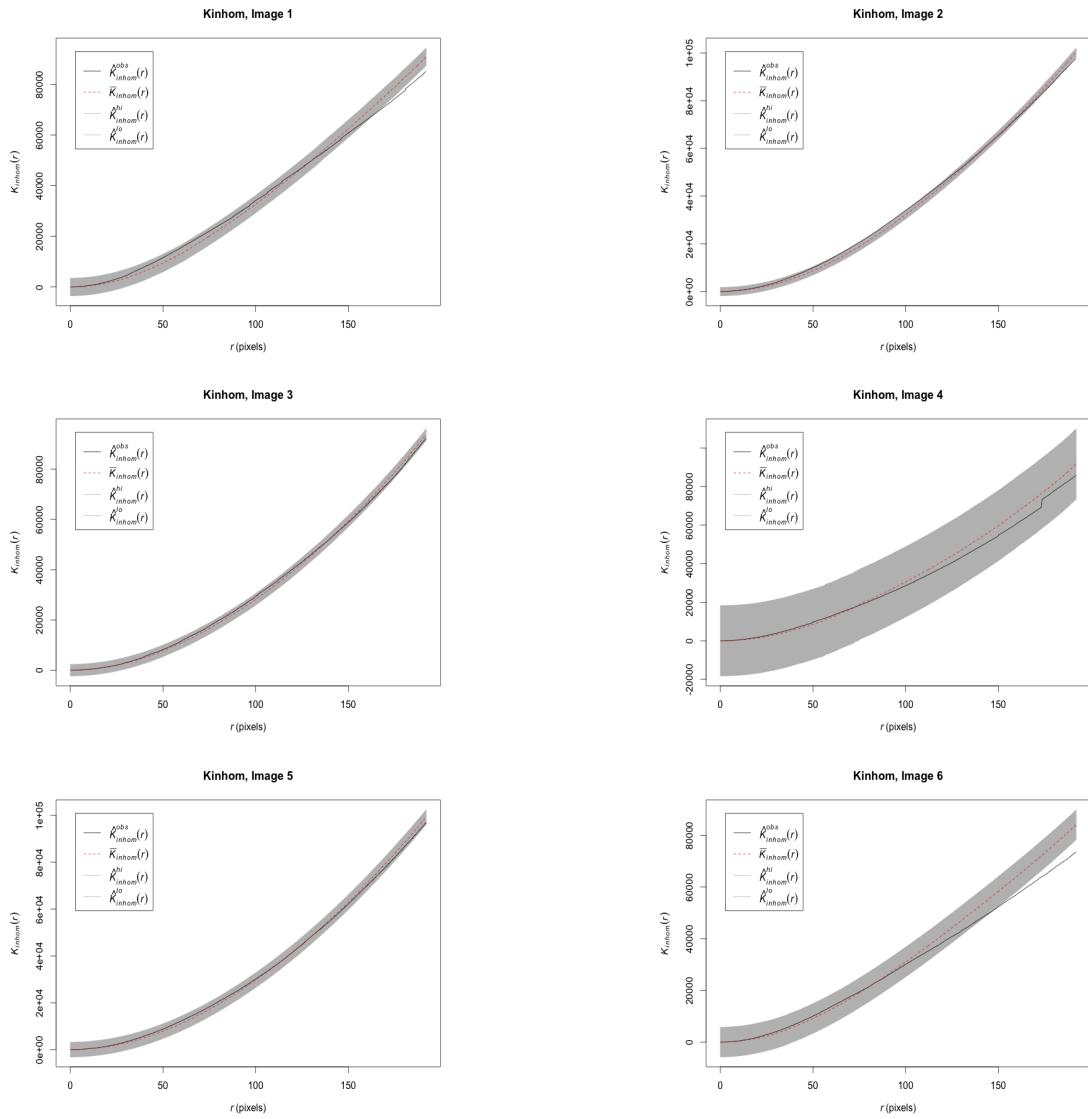


A.2 Test des quadrats, image Rafting

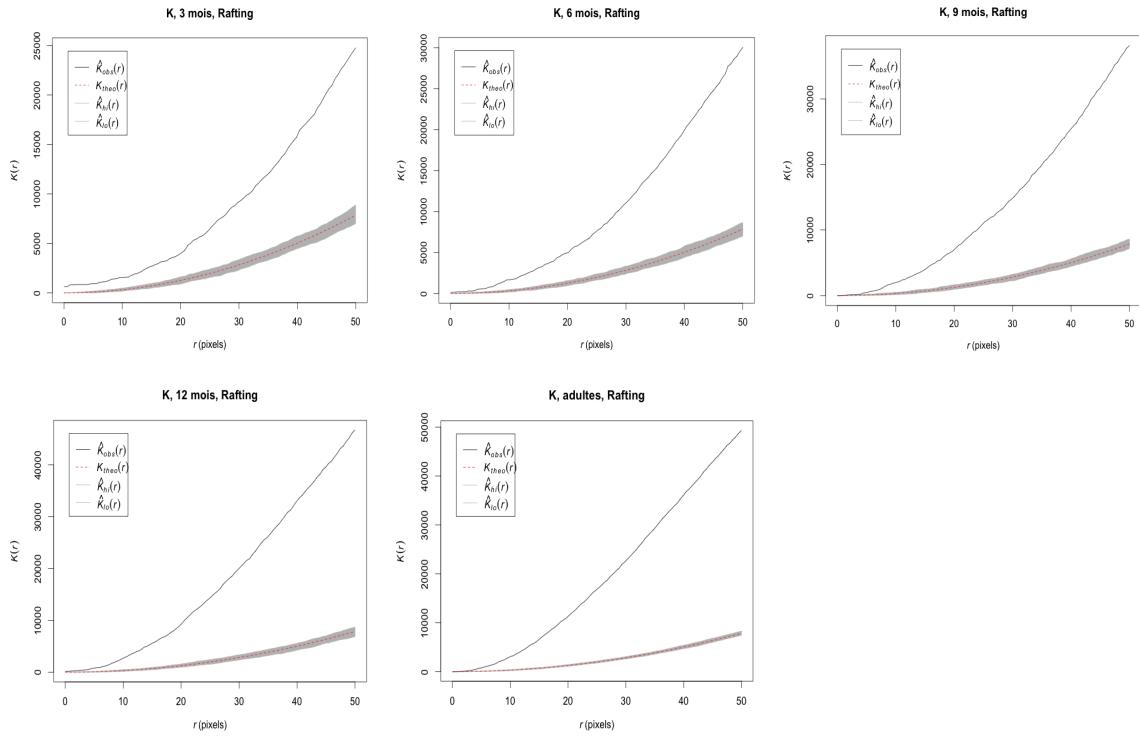
```
quadrat.test(unmark(X_Im4), nx=3)

##
## Chi-squared test of CSR using quadrat counts
## Pearson X2 statistic
##
## data: unmark(X_Im4)
## X2 = 1934.24, df = 8, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 3 by 3 grid of tiles
```

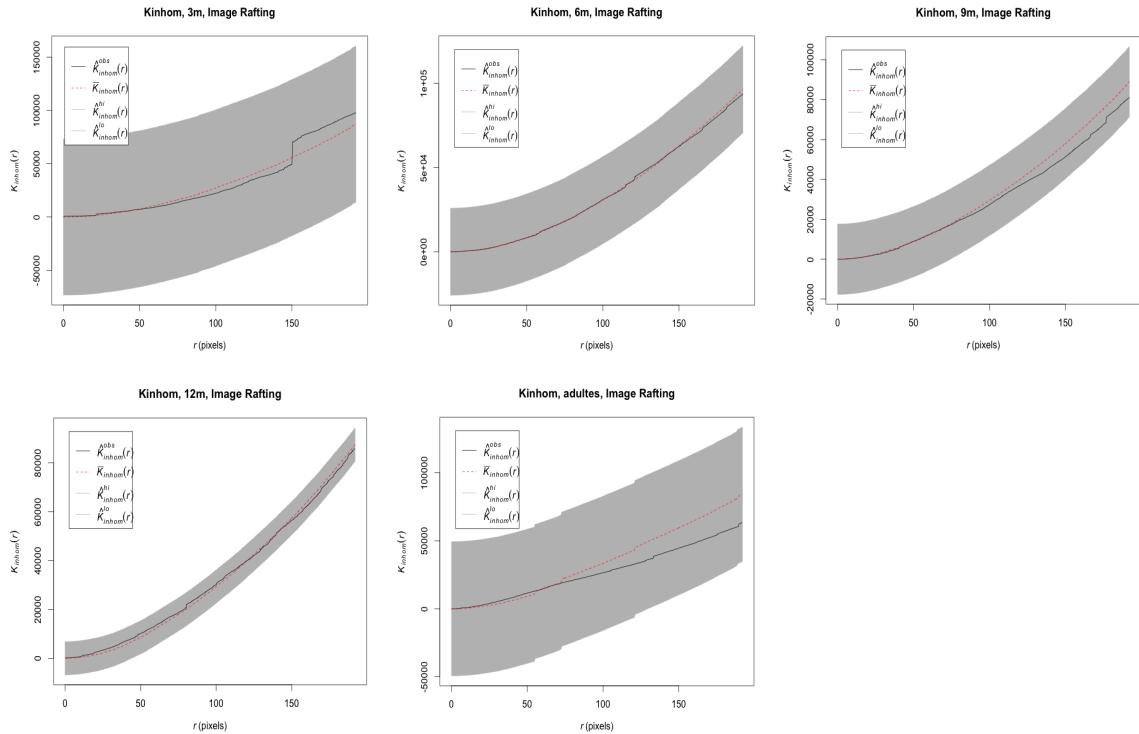
A.3 Fonction K_{inhom} appliquée aux 6 images



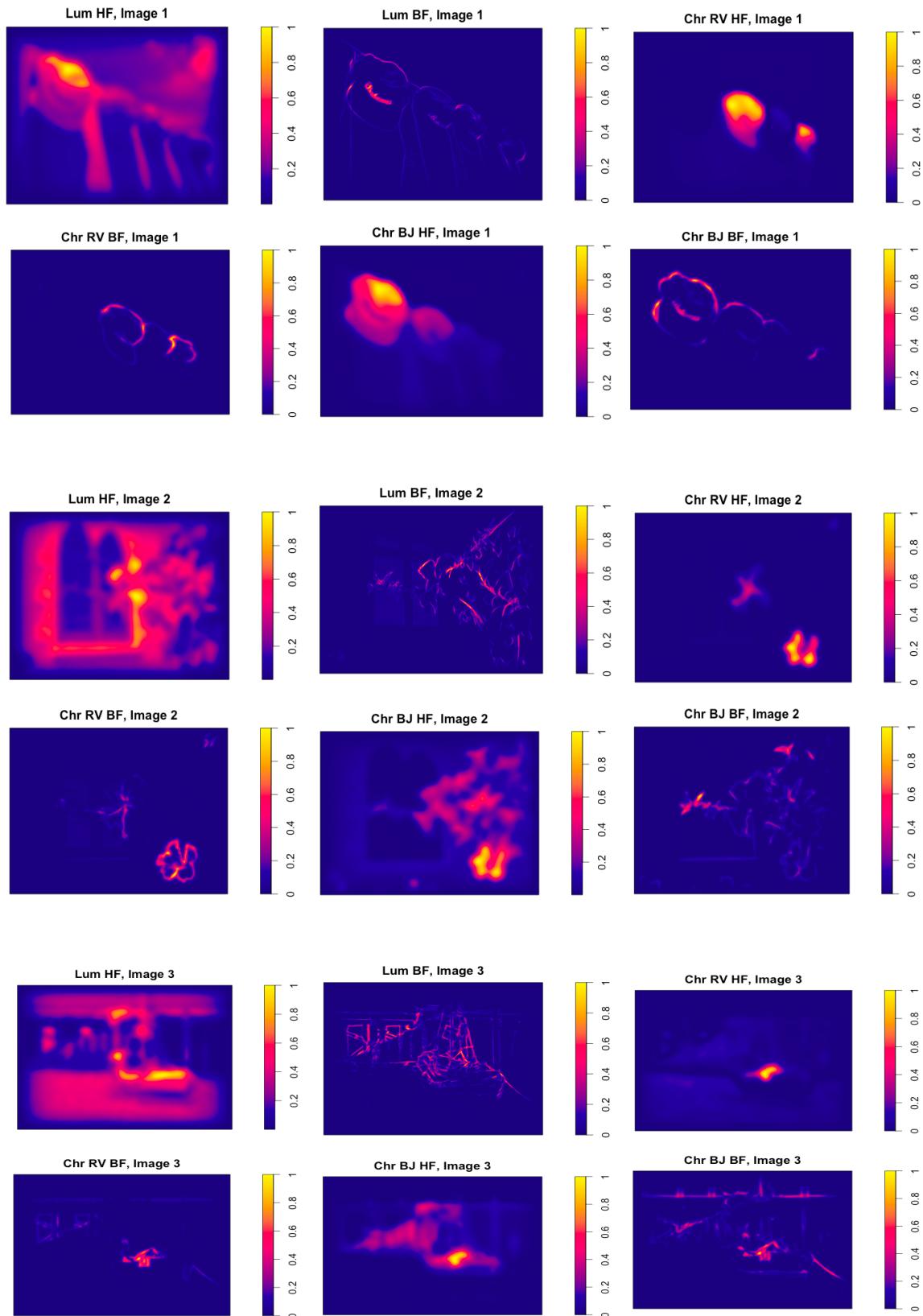
A.4 Fonction K appliquée aux 5 processus de groupe d'âge, rafting

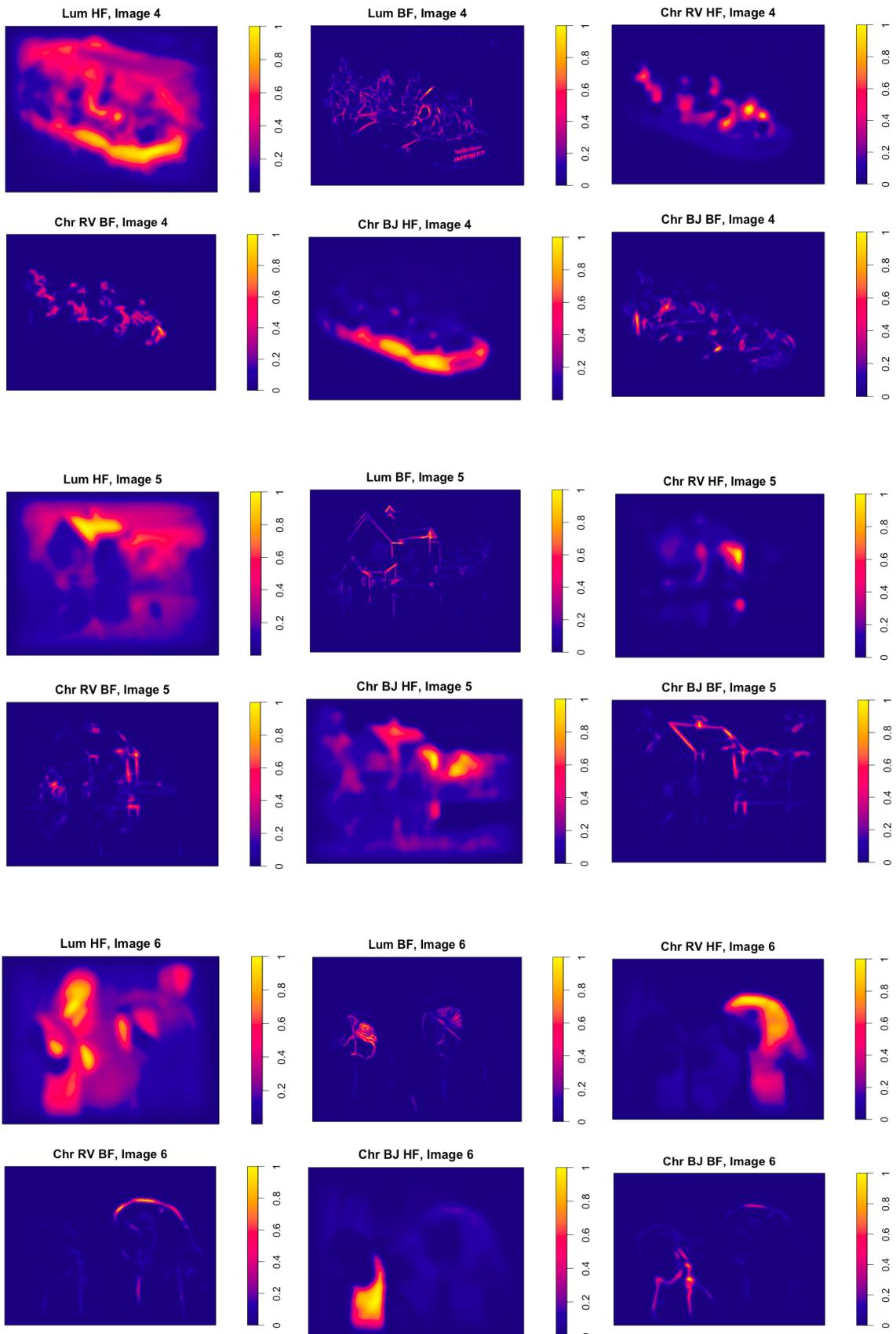


A.5 Fonction K_{inhom} appliquée aux 5 processus de groupe d'âge, rafting

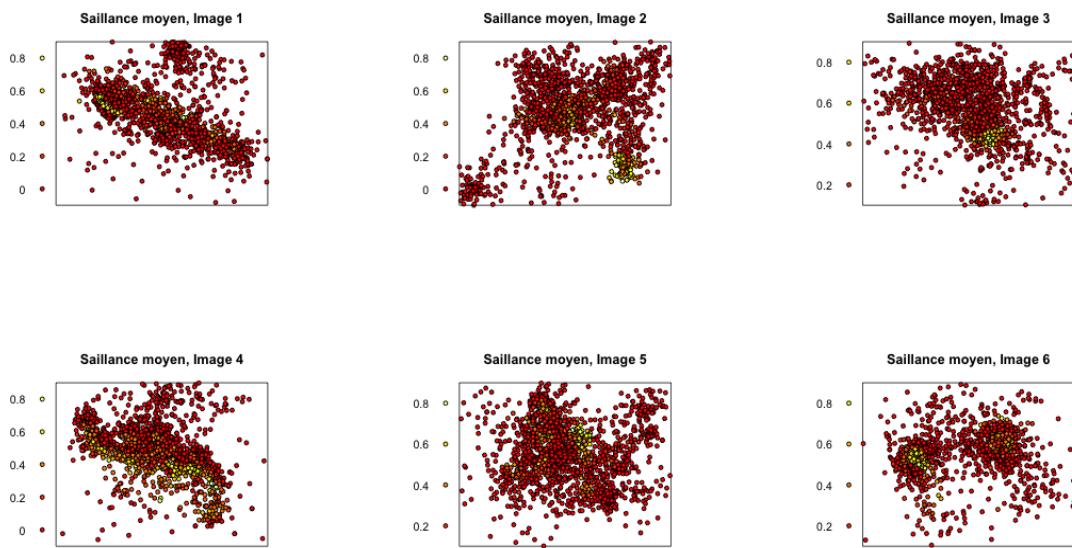


A.6 Niveaux de saillance pour les 6 images

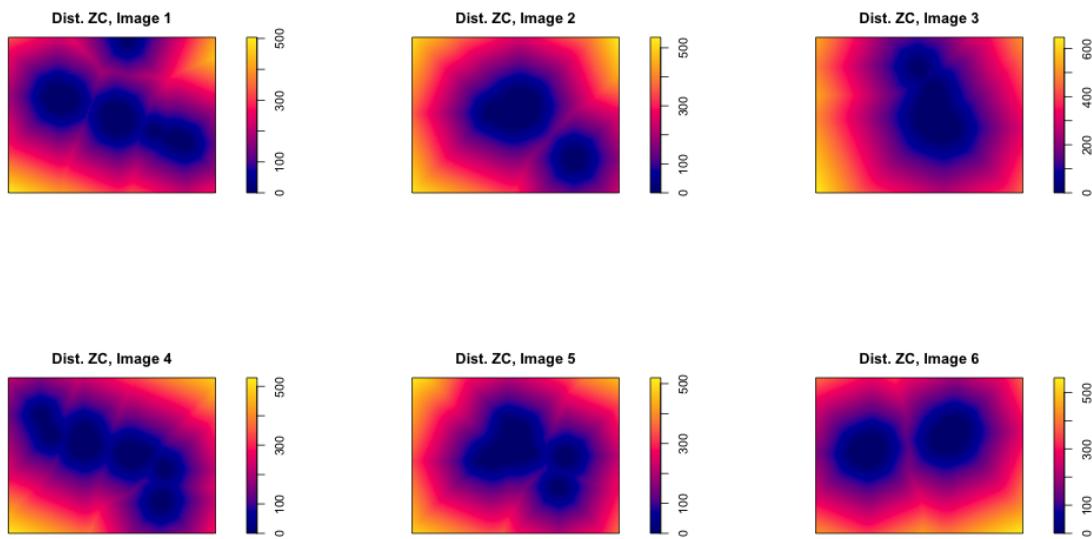




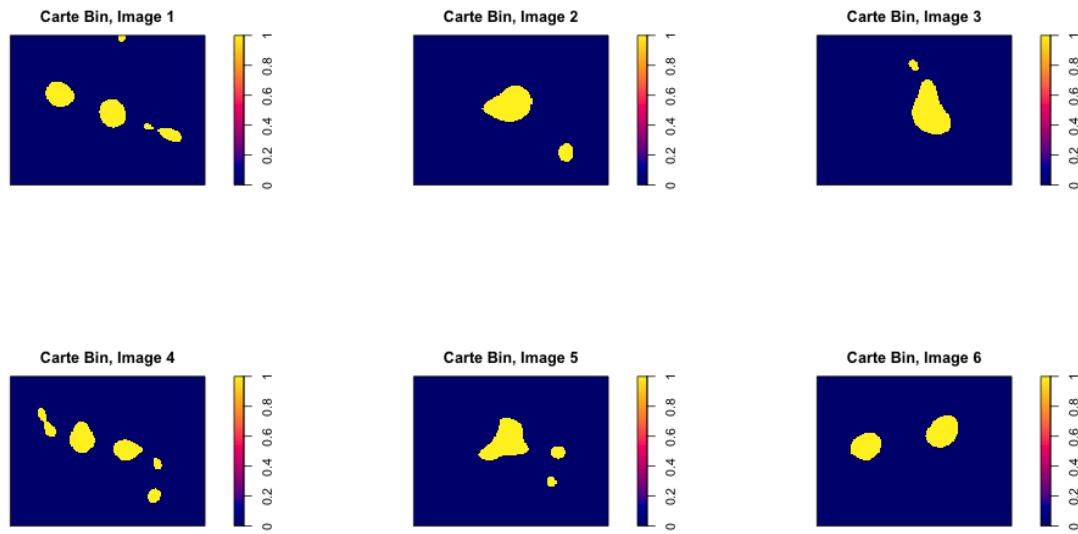
A.7 Fixations associées au niveau de saillance moyen pour les 6 images



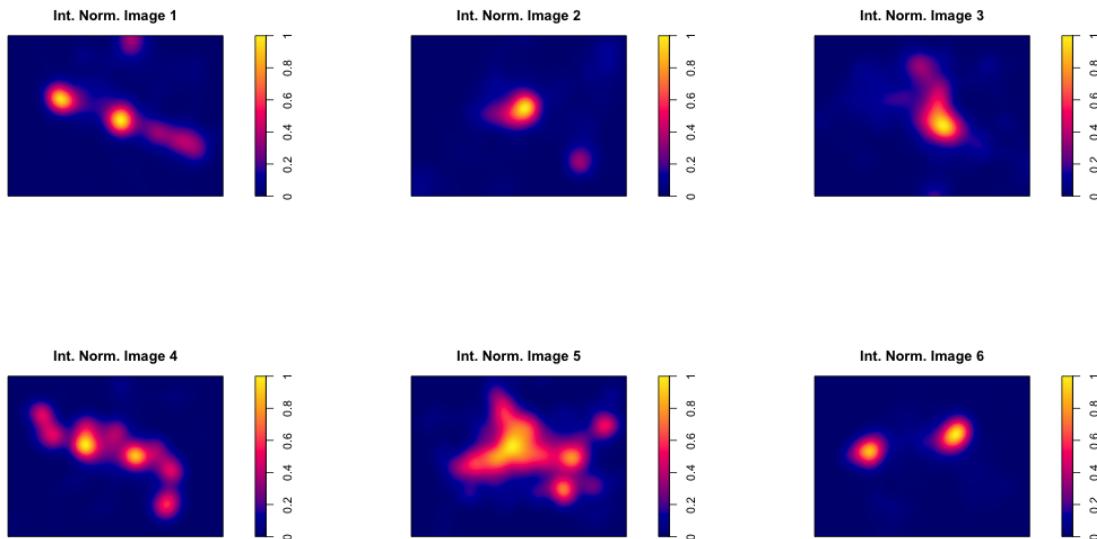
A.8 Distances aux zones cognitives



A.9 Cartes binaires pour les 6 images



A.10 Intensités normalisées estimées non-paramétriquement



A.11 Exemple de sortie R mppm, image Rafting

```

summary(fit[[4]])

## Point process model fitted to 4 point patterns
## Call: mppm(ppp ~ sail + group * ZCBin - ZCBin, data = hyp.Im[[i]])
## Log trend formula: ~sail + group * ZCBin - ZCBin
## Fitted trend coefficients:
##          (Intercept)      sail    group3mois    group6mois
## -8.23463633  2.03504193 -0.01874119 -0.05988180
## group9mois group12mois:ZCBin group3mois:ZCBin group6mois:ZCBin
## -0.07942860  1.44696303  0.85964130  1.17161475
## group9mois:ZCBin
##           1.01592721
## All fitted coefficients:
##             Estimate Std. Error   t value   Pr(>|t|)
## (Intercept) -8.23463633  0.1004463 -81.9805018 0.000000e+00
## sail         2.03504193  0.2312250  8.8011336 1.656024e-18
## group3mois  -0.01874119  0.1358175 -0.1379881 8.902535e-01
## group6mois  -0.05988180  0.1372880 -0.4361765 6.627210e-01
## group9mois  -0.07942860  0.1380655 -0.5752964 5.651079e-01
## group12mois:ZCBin 1.44696303  0.1923232  7.5236007 5.939759e-14
## group3mois:ZCBin  0.85964130  0.2447601  3.5121790 4.470296e-04
## group6mois:ZCBin  1.17161475  0.2213908  5.2920668 1.243196e-07
## group9mois:ZCBin  1.01592721  0.2331962  4.3565338 1.338743e-05
##
## Interaction for all patterns:
## Poisson process
##
##
## --- Gory details: ---
## Combined data frame has 7610 rows
##
## Call:
## glm(formula = fmla, family = quasi(link = log, variance = mu),
##      data = moadf, weights = .mpl.W, subset = (.mpl.SUBSET ==
##            "TRUE"), control = glm.control(maxit = 50))
##
## Deviance Residuals:
##       Min     1Q Median     3Q    Max
## -2.2109 -0.5257 -0.5068 -0.4909  2.4535
## 
```

A.12 Procédures R, tests multiples

```

K <- rbind("12m-3m"=c(rep(0,5), 1, -1, 0, 0),
           "9/12m - 3/6m"=c(rep(0,5), 1, -1, -1, 1),
           "12m-3/6/9m"=c(rep(0,5), 3, -1, -1, -1))

K

##          [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
## 12m-3m      0    0    0    0    0    1   -1    0    0
## 9/12m - 3/6m 0    0    0    0    0    1   -1   -1    1
## 12m-3/6/9m  0    0    0    0    0    3   -1   -1   -1

t <- list()
for(i in 1:6) {
  t[[i]] <- glht(fit[[i]], linfct = K, alternative = c("greater"))
}

# Résultats
res1 <- list()
for(i in 1:6) {
  # Single method
  res1[[i]] <- summary(t[[i]])
}

```

A.13 Tests entre groupe d'âges pour toutes les images

```
## [[1]]
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Linear Hypotheses:
##                         Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0           0.4818    0.1667  2.891 0.0046 **
## 9/12m - 3/6m <= 0     0.5817    0.2320  2.508 0.0146 *
## 12m-3/6/9m <= 0      0.5570    0.3940  1.414 0.1519
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
##
##
## [[2]]
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Linear Hypotheses:
##                         Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0           0.3606    0.2173  1.660 0.10367
## 9/12m - 3/6m <= 0     0.8340    0.3408  2.447 0.01756 *
## 12m-3/6/9m <= 0      1.5106    0.5171  2.921 0.00474 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
##
##
## [[3]]
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Linear Hypotheses:
##                         Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0           0.5334    0.2307  2.312 0.0243 *
## 9/12m - 3/6m <= 0     0.6677    0.3313  2.015 0.0485 *
## 12m-3/6/9m <= 0      0.9020    0.5324  1.694 0.0936 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```

## [[4]]
##
##   Simultaneous Tests for General Linear Hypotheses
##
##   Linear Hypotheses:
##                   Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0      0.5873    0.2016  2.914 0.00425 **
## 9/12m - 3/6m <= 0  0.4316    0.2904  1.487 0.13843
## 12m-3/6/9m <= 0   1.2937    0.4538  2.851 0.00527 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
##
##
## [[5]]
##
##   Simultaneous Tests for General Linear Hypotheses
##
##   Linear Hypotheses:
##                   Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0      -0.07103   0.21975 -0.323  0.819
## 9/12m - 3/6m <= 0 -0.48153   0.33275 -1.447  0.985
## 12m-3/6/9m <= 0   -0.03643   0.53186 -0.068  0.736
## (Adjusted p values reported -- single-step method)
##
##
## [[6]]
##
##   Simultaneous Tests for General Linear Hypotheses
##
##   Linear Hypotheses:
##                   Estimate Std. Error z value Pr(>z)
## 12m-3m <= 0      0.8966    0.1633  5.489 < 1e-06 ***
## 9/12m - 3/6m <= 0  1.4160    0.2360  6.001 < 1e-06 ***
## 12m-3/6/9m <= 0   1.7235    0.3941  4.373 1.06e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)

```

A.14 Hyperframe contenant l'ensemble des fixations de tous les groupes, toutes les images

```
hyp1  
## Hyperframe:  
##      ppp   image groupe sail ZCBin    n  
## 1 (ppp) Image1     3m (im)  (im) 309  
## 2 (ppp) Image2     3m (im)  (im) 265  
## 3 (ppp) Image3     3m (im)  (im) 260  
## 4 (ppp) Image4     3m (im)  (im) 288  
## 5 (ppp) Image5     3m (im)  (im) 265  
## 6 (ppp) Image6     3m (im)  (im) 288  
## 7 (ppp) Image1     6m (im)  (im) 338  
## 8 (ppp) Image2     6m (im)  (im) 279  
## 9 (ppp) Image3     6m (im)  (im) 223  
## 10 (ppp) Image4    6m (im)  (im) 292  
## 11 (ppp) Image5    6m (im)  (im) 276  
## 12 (ppp) Image6    6m (im)  (im) 294  
## 13 (ppp) Image1    9m (im)  (im) 299  
## 14 (ppp) Image2    9m (im)  (im) 246  
## 15 (ppp) Image3    9m (im)  (im) 240  
## 16 (ppp) Image4    9m (im)  (im) 280  
## 17 (ppp) Image5    9m (im)  (im) 228  
## 18 (ppp) Image6    9m (im)  (im) 274  
## 19 (ppp) Image1   12m (im)  (im) 323  
## 20 (ppp) Image2   12m (im)  (im) 318  
## 21 (ppp) Image3   12m (im)  (im) 269  
## 22 (ppp) Image4   12m (im)  (im) 334  
## 23 (ppp) Image5   12m (im)  (im) 355  
## 24 (ppp) Image6   12m (im)  (im) 354
```

A.15 Summary obtenu à partir de l'hyperframe images regroupées

```
fit.hyp1 <- mppm(ppp~sail+groupe*ZCBin-ZCBin,data=hyp1)  
  
summary(fit.hyp1)$coef  
  
##      (Intercept)          sail       groupe3m       groupe6m  
## -9.89210986    2.26453925   -0.02287345   -0.05755915  
## groupe9m groupe12m:ZCBin groupe3m:ZCBin groupe6m:ZCBin  
## -0.17838034    1.71284619    1.25071236    1.45475358  
## groupe9m:ZCBin  
##           1.58297709
```

Références

- [1] Baddeley, A., Rubak, E. and Turner, R., *Spatial point pattern analysis with R*, 2015.
- [2] Baddeley, A., *Analysing spatial point patterns in R.*, Technical report, CSIRO, Version 4, 2010.
Available at csiro.au/resources/pf16h.html
- [3] Baddeley, A. and Turner, R., *Spatstat : an R package for analyzing spatial point patterns*, Journal of Statistical Software, 12(6) :1–42, 2005. Available at <http://www.jstatsoft.org>, ISSN : 1548-7660.
- [4] Baddeley, A. and Turner, R., *Practical maximum pseudolikelihood for spatial point patterns (with discussion)*, Australian and New Zealand Journal of Statistics, 42(3) :283–322, 2000.
- [5] Barthelme, S., Trukenbrod, H., Song, Y. and Engbert, R., *Modelling fixation locations using spatial point processes*, Journal of Vision, 13, 2013.
- [6] Bell, M. and Grunwald, G., *Mixed models for the analysis of replicated spatial point patterns*, Biostatistics, 5 :633–648, 2008.
- [7] Cressie, N., *Statistics for spatial data*, John Wiley and Sons, Inc, 1993.
- [8] Diggle, P.J., *Statistical Analysis of Spatial Point Patterns*, Arnold London, second edition, 2003.
- [9] Guyon, X., *Random Fields on a Network*, Springer-Verlag, New York, 1991.
- [10] Ho-Phuoc, T., Guyader, N. and Guérin-Dugué, A., *A Functional and Statistical Bottom-Up Saliency Model to Reveal the Relative Contributions of Low-Level Visual Guiding Factors*, Springer Science Business Media, LLC, 2010.
- [11] Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D., *Statistical Analysis and Modelling of Spatial Point Patterns*, Statistics in Practice, Wiley, Chichester, 2008.
- [12] McCulloch, C.E., Searle, S.R. and Neunhaus, J.M., *Generalized Linear, and Mixed Models*, Wiley Series in probability and statistics, Wiley, 2nd edition, 2008.
- [13] McCullagh, P. and Nelder J.A., *Generalized Linear Models*, Chapman & HALL/CRC, Monographs on Statistics and Applied Probability 37 2nd edition, 1989.
- [14] Møller, J. and Waagepetersen, R., *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC, Boca Raton, 2004.
- [15] Pinheiro, J.C. and Bates, D.M., *Mixed-effects models in S and S-PLUS*, Statistics and Computing, Springer, 2000.