

- **Titre du sujet : Inférence statistique pour la modélisation du vieillissement en population naturelle**
- Unité de recherche : LAGA, Institut Galilée, Université Sorbonne Paris Nord.
- Discipline : Mathématiques Appliquées
- Direction de thèse : Thanh Mai PHAM NGOC (directrice), Sarah KAAKAI (co-encadrante)
- Contact : phamngoc@math.univ-paris13.fr, kaakai@math.univ-paris13.fr
- Domaine de recherche : Statistiques, probabilités
- Mots-clés : inférence statistique, modélisation, processus ponctuels, EDP, applications au vieillissement en population naturelle

## 1 Motivations

Le vieillissement, défini comme le déclin progressif des fonctions d'un organisme entraînant un risque de mort croissant avec l'âge, a longtemps été supposé absent des populations naturelles, en raison des fortes pressions de sélection et de prédation. Un nombre croissant de données empiriques, révélant la présence d'individus physiologiquement âgés chez de nombreuses espèces [13, 2], remet aujourd'hui en cause cette hypothèse. Cela soulève des questions fondamentales : **quels mécanismes permettent le vieillissement et le maintien d'individus âgés en milieu naturel ? Comment les modéliser et les estimer ?**

Une difficulté concerne la mesure même du vieillissement. L'âge chronologique est en effet souvent difficile à mesurer, et constitue parfois un indicateur physiologique insuffisant. Des travaux récents, initialement conduits sur l'organisme modèle *Drosophila melanogaster* [15] puis étendus à d'autres organismes modèles dont la souris [3, 4], ont permis l'identification d'un phénotype appelé Smurf, qui constitue un important indicateur de vieillissement physiologique. Ce phénotype repose sur une mesure in vivo de la perméabilité intestinale à l'aide d'un colorant alimentaire bleu non toxique : les individus physiologiquement âgés absorbent ce colorant et se colorent entièrement en bleu [14, 15]. Ces individus présentent un risque de mort significativement plus élevé ainsi que plusieurs marqueurs classiques du vieillissement, notamment une altération de la signalisation insulinaire, une perte des réserves énergétiques et une diminution de l'activité spontanée [17, 18].

Le phénotype smurf offre ainsi une méthode relativement simple mais particulièrement efficace pour identifier des individus physiologiquement âgés au sein de populations naturelles. Une campagne de terrain récente a confirmé que 5 à 8% des drosophiles échantillonnées présentaient ce phénotype, résultat cohérent avec les études récentes attestant de la présence d'individus physiologiquement âgés dans la nature [2, 13].

**Objectifs de la thèse** Ces questions biologiques fondamentales motivent cette thèse, dont le but est d'établir de nouveaux résultats en inférence statistique pour des modèles de dynamiques de populations compartimentaux structurés en âge, bien adaptés pour répondre à

ces questions. A notre connaissance, ces classes de modèles, bien étudiés d'un point de vue probabiliste et déterministe, ont fait l'objet de peu de travaux concernant l'inférence statistique de leurs paramètres. Cette relative absence de résultats statistiques contraste avec la maturité des approches probabilistes et déterministes, et constitue précisément l'une des motivations principales de ce projet de thèse. Des estimateurs et des test non-paramétrique seront ainsi proposés et leurs propriétés théoriques (vitesse de convergence, vitesses de séparation) étudiées dans un paradigme minimax. L'attention sera portée sur deux régimes, celui d'une population à l'équilibre et celui d'une population hors équilibre.

## 2 Projet de recherche détaillé

Les modèles étudiés dans cette thèse sont des modèles compartimentaux structurés en âge, dans lesquels chaque individu est caractérisé à la fois par son âge chronologique et son appartenance à un groupe d'âge physiologique défini par des traits phénotypiques mesurables. Ces modèles flexibles, fondés sur des processus ponctuels ou plus généralement sur des processus à valeur mesure, trouvent des applications dans divers domaines, notamment en épidémiologie. Le projet s'articule autour de deux axes complémentaires.

**Population hors équilibre** Le premier concerne des populations sans renouvellement. Il s'agira d'estimer l'intensité des évènements ayant lieu dans la population, et de construire des procédures de tests et d'estimation. En particulier, l'un des objectifs sera de déterminer si les interactions et la compétition pour les ressources accélèrent le vieillissement des individus, à partir de données réelles d'expériences en cours à l'ECOTRON de Montpellier.

Dans un premier temps, on modélisera la dynamique simplifiée d'une population d'individus synchronisés (ayant le même âge) et en interaction, à l'aide d'un modèle de vieillissement à deux compartiments. Chaque compartiment représente un état physiologique : les individus jeunes (compartiment 1) et les individus physiologiquement âgés (compartiment 2).

Plus précisément, à  $t = 0$ , on considère  $N_0$  individus d'âge 0, initialement dans le compartiment 1. Chaque individu  $i \in \{1, \dots, N_0\}$  passe d'abord un temps aléatoire  $\tau_i^1$  dans le compartiment 1, puis un temps  $\tau_i^2$  dans le compartiment 2, avant de mourir. La dynamique de population est décrite par le processus de comptage  $N^1$  comptant les transitions du compartiment 1 vers le compartiment 2, ainsi que par un processus à valeur mesure  $Z^2$  décrivant le nombre d'individus dans le second compartiment et le temps qu'ils y ont passé. On s'intéressera en particulier à l'intensité du processus  $N^1$ , décrivant les transitions du compartiment 1 vers le compartiment 2 et représentant le vieillissement des individus :

$$\lambda_t^1 = \left( f_{12}(t) + c_1(t)(N_0 - N_t^1) + \int_0^\infty U^2(t, u)Z^2(du) \right) (N_0 - N_t^1). \quad (1)$$

Une question biologique essentielle est de savoir s'il y a ou non une accélération du vieillissement due à la compétition et aux interactions entre les individus. Sur le plan de l'inférence statistique cela se traduit par tester si  $U^2$  et/ou  $c_1$  sont nulles. Nous pourrions dans un premier temps considérer le modèle simplifié où  $U^2$  est nulle. Une fois les tests établis, nous pourrions nous pencher sur les questions d'estimation de ces deux paramètres.

Un premier cadre d'observations à partir desquelles l'inférence statistique sera menée consistera en les  $N$  trajectoires indépendantes du processus  $(N^{1,i}, Z^{2,i})_{1 \leq i \leq N}$  partant de  $N$  conditions initiales  $(N_0^i)_{1 \leq i \leq N}$ . Un second cadre d'observations correspondra à un régime de grande population où la population initiale  $N_0$  dépend d'un paramètre  $K$  avec  $N_0$  tendant vers l'infini quand  $K$  tend vers l'infini. Il faudra alors considérer les processus renormalisés par ce paramètre  $K$ . Ceci nous amènera naturellement à la seconde partie de la thèse où l'on considèrera une population à l'équilibre.

**Population à l'équilibre** Le second axe du projet de thèse porte sur des dynamiques de populations avec renouvellement (naissances, entrées dans la population), supposées à l'équilibre. La question centrale est la suivante : à partir d'un échantillonnage d'individus à l'équilibre, peut-on identifier, même partiellement, les paramètres démographiques cohérents avec certaines mesures observées ?

Les processus à valeur mesure modélisant des dynamiques de populations structurées par âge et par caractéristiques phénotypiques ont fait l'objet de travaux approfondis d'un point de vue probabiliste (voir par exemple [16, 12, 11]) et des outils efficaces existent pour les simuler [7]. En revanche, leur étude sous l'angle statistique reste à ce jour peu développée (voir par exemple [1] pour des applications en démographie).

Une approche possible consiste à considérer la population à l'équilibre dans un régime en grande population. Plus précisément, on s'intéressera aux solutions stationnaires du système d'EDP suivant, modélisant la dynamique d'une population structurée en deux compartiments : les individus du compartiment 1 (physiologiquement jeunes), de densité de population  $n_1(a, t)$  indexée par l'âge chronologique  $a$ , et les individus du compartiment 2 (physiologiquement âgés), de densité  $n_2(u, t)$  indexée par l'âge physiologique  $u$  (temps passé dans le compartiment 2).

$$\begin{cases} (\partial_a + \partial_t)n_1(a, t) = -c_d(\bar{n}_1(t), \bar{n}_2(t))n_1(a, t) - (f_{12}(a) + c_1(a)\bar{n}_1(t) + \int_0^\infty U^2(a, u)n_2(u, t)du)n_1(a, t), \\ (\partial_u + \partial_t)n_2(u, t) = (-k_d(u) - \tilde{c}_d(\bar{n}_1(t), \bar{n}_2(t)))n_2(u, t), \\ n_1(0, t) = \int_0^\infty (b_1(a)n_1(a, t) + b_2(a)n_2(a, t))da, \\ n_2(0, t) = \int_0^\infty (f_{12}(a) + c_1(a)\bar{n}_1(t) + \int_0^\infty U^2(a, u)n_2(u, t)du)n_1(a, t)da, \end{cases} \quad (2)$$

où  $\bar{n}_x(t) = \int_0^\infty n_x(a, t) da$  pour  $x = 1$  ou  $2$ . L'étude de ce système d'équations non linéaires en toute généralité n'est pas triviale et fait l'objet de travaux en cours. On sera ainsi amené à considérer des cas simplifiés pour lesquels l'équilibre est explicitement connu, mais qui donnent néanmoins lieu à des problèmes inverses non standard pour l'estimation des paramètres démographiques à partir d'un échantillonnage de la population stationnaire.

Des avancées théoriques récentes ont permis des progrès en inférence non-paramétrique sur les taux de division et des noyaux de fragmentation pour les populations cellulaires structurées en taille ou en âge [6, 10, 5, 9, 8]. Ces résultats ouvrent une direction de recherche prometteuse pour l'inférence de paramètres démographiques en population naturelle avec vieillissement. Leur extension à des populations avec interactions et groupes d'âge physiologique constitue cependant un problème ouvert, qui ne se réduit pas à une simple adaptation des méthodes existantes.

## References

- [1] A. Boumezoued, M. Hoffmann, and P. Jeunesse. Nonparametric adaptive inference of birth and death models in a large population limit. *Math. Stat. Learn.*, 3:1–69, 2021.
- [2] A.K. Brunet-Rossinni and S.N. Austad. Senescence in wild populations of mammals and birds. In *Handbook of the Biology of Aging*, pages 243–266. Elsevier, 2005.
- [3] C. Cansell, V. Goepp, F. Bain, N. Todd, V. Douard, M. Monnoye, C. Sanchez, N. Pietran-costa, C. Rovere, R.G. Denis, S. Luquet, and M. Rera. Two phases model of ageing in mice: towards a better identification of age-related and late-life metabolic decline [Reg-istered Report Stage 1 Protocol]. *BMC Biol.*, 2023.
- [4] E. Dambroise, L. Monnier, L. Ruisheng, H. Aguilaniu, J.-S. Joly, H. Tricoire, and M. Rera. Two phases of aging separated by the smurf transition as a public path to death. *Sci. Rep.*, 6:23523, 2016.
- [5] M. Doumic and M. Hoffmann. Individual and population approaches for calibrating division rates in population dynamics: Application to the bacterial cell cycle. In *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 1–81. World Scientific, 2023.
- [6] M. Doumic, M. Hoffmann, P. Reynaud-Bouret, and V. Rivoirard. Nonparametric es-timation of the division rate of a size-structured population. *SIAM J. Numer. Anal.*, 50:925–950, 2012.
- [7] D. Giorgi, S. Kaakai, and V. Lemaire. Efficient simulation of individual-based population models: The R package IBMPopSim. *Computo*, 2024.
- [8] V.H. Hoang, P. T. Nguyen, T. M. Pham Ngoc, V. Rivoirard, and V.C. Tran. Goodness-of-fit testing for the stationary density of a size-structured pde,. *To appear in VIASM Lecture Notes "Resonances in the Mathematical World"*, 2026.
- [9] V.H. Hoang, T.M. Pham Ngoc, V. Rivoirard, and V.C. Tran. Nonparametric estimation of the fragmentation kernel based on a partial differential equation stationary distribution approximation. *Scand. J. Stat.*, 2022.
- [10] Marc Hoffmann and Adélaïde Olivier. Nonparametric estimation of the division rate of an age dependent branching process. *Stochastic Processes and their Applications*, 126(5):1433–1471, 2016.
- [11] S. Méléard, M. Rera, and T. Roget. A birth–death model of ageing: from individual-based dynamics to evolutive differential inclusions. *J. Math. Biol.*, 2019.
- [12] S. Méléard and V.C. Tran. Trait substitution sequence process and canonical equation for age-structured populations. *J. Math. Biol.*, 58:881–921, 2009.
- [13] D.H. Nussey, H. Froy, J.-F. Lemaitre, J.-M. Gaillard, and S.N. Austad. Senescence in natural populations of animals: Widespread evidence and its implications for bio-gerontology. *Ageing Res. Rev.*, 12:214–225, 2013.

- [14] M. Rera, S. Bahadorani, J. Cho, C.L. Koehler, M. Ulgherait, J.H. Hur, W.S. Ansari, T. Lo, D.L. Jones, and D.W. Walker. Modulation of longevity and tissue homeostasis by the *Drosophila* PGC-1 homolog. *Cell Metab.*, 14:623–634, 2011.
- [15] M. Rera, R. Clark, and D. Walker. Intestinal barrier dysfunction links metabolic and inflammatory markers of aging to death in *Drosophila*. *PNAS*, 2012.
- [16] V.C. Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM Probab. Stat.*, 12:345–386, 2008.
- [17] F. Zane, C. MacMurray, C. Guillermain, C. Cansell, N. Todd, and M. Rera. Ageing as a two-phase process: theoretical framework. *Front. Aging*, 5, 2024.
- [18] C. Zhang, J. Wang, T. Yao, J. Hu, F. Sun, C. Feng, Z. Sun, Y. Shao, Z. Wang, J. Wu, and Y. Huang. Proteomic analysis across aged tissues reveals distinct signatures and the crucial involvement of midgut barrier function in the regulation of aging. *Aging Cell*, 24, 2025.