Mathematical modeling and statistical study of a 2-phase ageing model.

Luce Breuil (CMAP Ecole Polytechnique)
Co-supervised by
Marie Doumic (CMAP Ecole Polytechnique - Inria)
Sarah Kaakaï (Sorbonne Paris Nord)
Michael Rera (Institut Jacques Monod)

Ecole de printemps Chaire MMB - 16 Juin 2026

What is Smurfness?

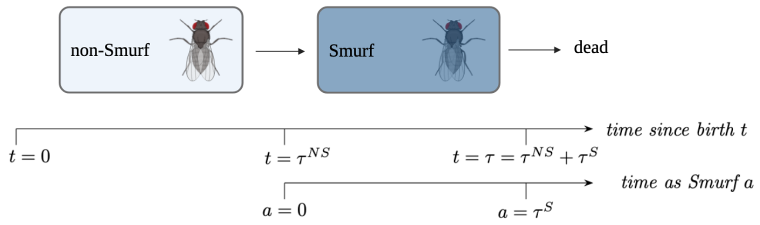
- Smurf **phenotype**: reduced mobility, increased inflammation, increased intestinal permeability
- All flies turn Smurf before dying : **indicator of old age**
- Smurf flies have reduced life expectancy ($\sim$ 2 days) : strong **predictor of death**

$\Longrightarrow$ Evidence for a **2-phased model of aging**

Rera, Clark and Walker (12)
Image Credit: M. Rera and A. Colibert

Modeling the data

Times spent non-Smurf and Smurf of 1159 synchronized drosophila flies in separate tubes.



Hypothesis in Tricoire and Rera (15): independent time in each phase and constant death hazard rate.

Mathematical model

m independent flies represented by 2 random times

- τ_i^{NS} time spent non-Smurf for fly i.
- τ_i^{S} time spent Smurf for fly i.

Mathematical model

m independent flies represented by 2 random times

- τ_i^{NS} time spent non-Smurf for fly i.
- τ_i^{S} time spent Smurf for fly i.

The laws of τ^{NS} and τ^{S} are defined by their **rates** $\approx$ **instantaneous probabilities** of occurrence.

$$\text{rate} = \frac{\text{density}}{\text{survival function}}$$

Mathematical model

m independent flies represented by 2 random times

- τ_i^{NS} time spent non-Smurf for fly i.
- τ_i^{S} time spent Smurf for fly i.

The laws of τ^{NS} and τ^{S} are defined by their **rates** $\approx$ **instantaneous probabilities** of occurrence.

$$\text{rate} = \frac{\text{density}}{\text{survival function}}$$

k_S is the **Smurf hazard rate** and k_D the **death hazard rate once Smurf** .

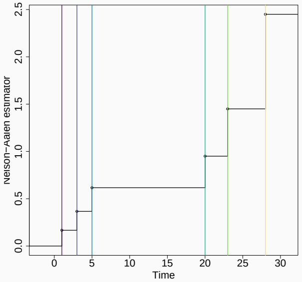
$\Longrightarrow$ **We want to estimate the hazard rates k_S and k_D to characterize 2-phase ageing in vivo.**

Non-parametric estimation of marginal laws

Nelson-Aalen estimator

$$\sum_{\tau_i \le t} \frac{1}{m - N_{\tau_i^-}},$$

with $N_t = \sum_{i=1}^{m} \mathbb{1}_{\{\tau_i \le t\}}$.

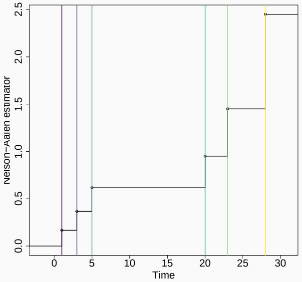


Non-parametric estimation of marginal laws

Nelson-Aalen estimator

$$\sum_{\tau_i \leq t} \frac{1}{m - N_{\tau_i^-}},$$

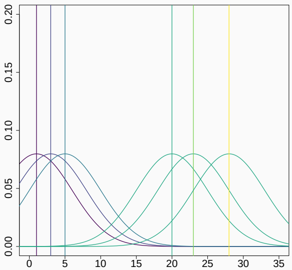
with $N_t = \sum_{i=1}^{m} \mathbb{1}_{\{\tau_i \leq t\}}$.



"Differentation" $\longrightarrow$

Empirical hazard

$$\sum_{i=1}^{m} \frac{\delta_{t=\tau_i}}{m - N_{\tau_i^-}}$$

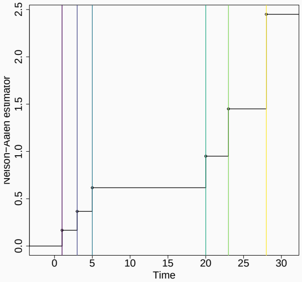


Non-parametric estimation of marginal laws

Nelson-Aalen estimator

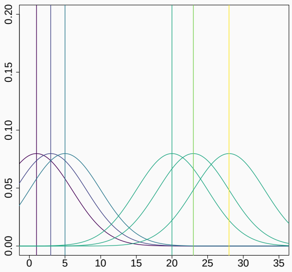
$$\sum_{\tau_i \leq t} \frac{1}{m - N_{\tau_i^-}},$$

with $N_t = \sum_{i=1}^{m} \mathbb{1}_{\{\tau_i \leq t\}}$.



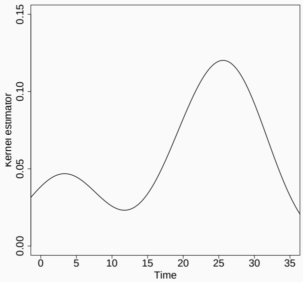
Empirical hazard

$$\sum_{i=1}^{m} \frac{\delta_{t=\tau_i}}{m - N_{\tau_i^-}}$$



Kernel estimator

$$\sum_{i=1}^{m} \frac{\kappa_{t,b}(\tau_i)}{m - N_{\tau_i^-}},$$



"Differentation" $\longrightarrow$

Smoothing $\longrightarrow$

Kernel estimation

Classical kernels $\kappa_{t,b}(x) = \frac{1}{b}\kappa(\frac{x-t}{b})$.
Boundary bias problem because they do not integrate to 1 on the support of the hazard rate.

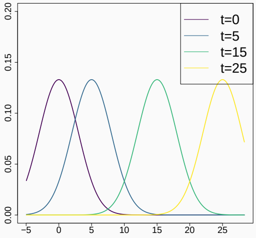


Figure 1: Gaussian kernel.

Kernel estimation

Classical kernels $\kappa_{t,b}(x) = \frac{1}{b}\kappa(\frac{x-t}{b})$.
Boundary bias problem because they do not integrate to 1 on the support of the hazard rate.

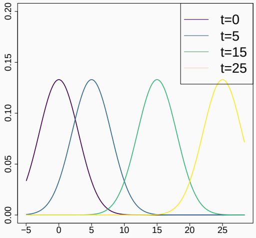


Figure 1: Gaussian kernel.

A popular solution: **associated kernels**, Beta and Gamma kernels introduced in Chen (99), Chen (00). Not supported on $\mathbb{R}$, asymmetric, no generic dependence in t and b.

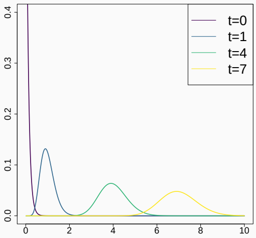


Figure 2: Gamma kernel.

New kernels !

Definition (Continuous associated kernel)

Let $b > 0$ be the bandwidth. An associated kernel $(\kappa_{t,b})$ is a parametrized probability density function $\kappa_{t,b}$ such that for any $y \in \mathbb{R}_+$, $t \mapsto \kappa_{t,b}(y)$ is continuous and verifying for all $t \in \mathbb{R}_+$

$$\Lambda(t,b) := \mathbb{E}(Z_{t,b}) - t \xrightarrow[b\to 0]{} 0 \quad \text{and} \quad Var(Z_{t,b}) \xrightarrow[b\to 0]{} 0, \tag{1}$$

where $Z_{t,b}$ denotes the random variable with pdf $\kappa_{t,b}$.

Ensures that $\kappa_{t,b}$ approximates correctly the Dirac at t.

Theoretical study of the associated kernel hazard rate estimator

Under certain assumptions on the kernel and the hazard rate we have for some $\gamma > 0$ and a sequence of bandwidths $b_m \xrightarrow[m\to+\infty]{} 0$.

- $\mathbb{E}\left[\int_I (\hat{k}_m(t) - k(t))^2 \,\mathrm{d}t\right] = O(b_m^{4\gamma} + m^{-1} b_m^{-\gamma})$ = convergence of the Mean Integrated Square Error (MISE): measure of the "quality" of the estimator.
- $\frac{\hat{k}_m(t) - \mathbb{E}[\hat{k}_m(t)]}{\sqrt{Var(\hat{k}_m(t))}} \xrightarrow[m\to+\infty]{} \mathcal{N}(0,1)$= central limit theorem, allows to build confidence intervals.
- Theoretical guarantees for a practical bandwidth choice method.

See B., Kaakaï (26+)

Results of the non-parametric estimation

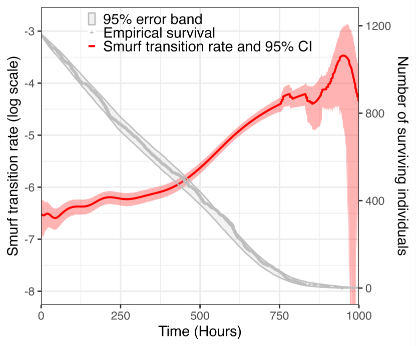


Figure 3: Smurf transition rate.

Gompertz-Makeham rate $k_S(t) = f + ge^{ht}$

Results of the non-parametric estimation

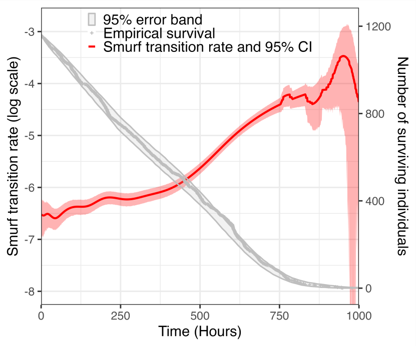


Figure 3: Smurf transition rate.

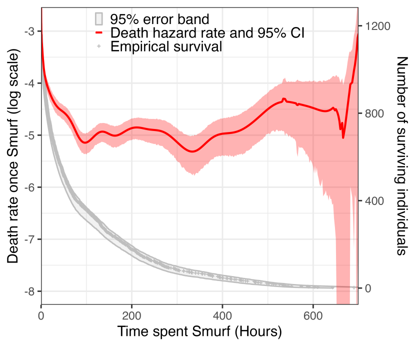


Figure 4: Death hazard rate once Smurf.

Gompertz-Makeham rate $k_S(t) = f + ge^{ht}$

Decreasing exponential $k_D(a) = c + de^{-ga}$

function of time spent Smurf !

Some quantitative indicators of dependence

Time spent non Smurf	Correlation coefficient	p-value of Wald test
All	-0.164	$5 \cdot 10^{-8}$

Table 1: Statistical indicators of dependence between jumping times on $\tau^{NS} \leq 200$ and $\tau^{NS} > 200$.

Some quantitative indicators of dependence

Time spent non Smurf	Correlation coefficient	p-value of Wald test
All	-0.164	$5 \cdot 10^{-8}$
≤ 200	-0.077	0.04
> 200	-0.28	$6.3 \cdot 10^{-16}$

Table 1: Statistical indicators of dependence between jumping times on $\tau^{NS} \leq 200$ and $\tau^{NS} > 200$.

Death hazard rate once Smurf

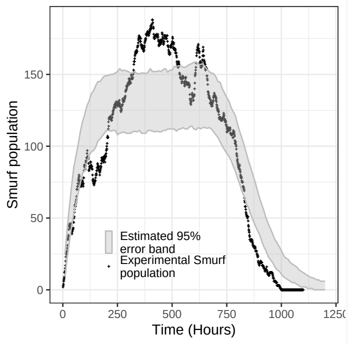
We model dependence with a semi-parametric **Cox Model**.

The investigated models are:

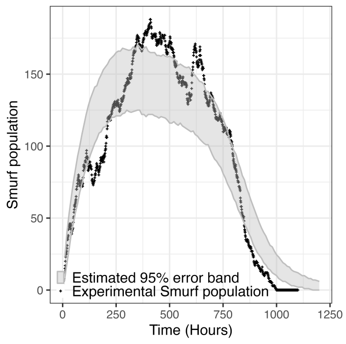
1. **No dependence** $k_D(a)$
2. **Single dependence** $k_D(a, \tau^{NS})$
3. **Piece-wise dependence** $k_D(a, \tau^{NS})\mathbb{1}_{\{\tau^{NS}>200\}} + k_D(a)$

Simulations of Smurf population over time

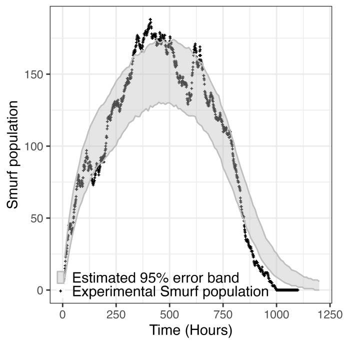
Without dependence



With a single dependence



With a piecewise dependence



Populations simulated with **IBMPopSim** - D.Giorgi, S.Kaakaï, V.Lemaire

Main conclusions

- Exponentially **decreasing death hazard rate** : Smurf phenotype is a strong predictor of death.
- **Dependence** between both jumping times with "viable" and "non-viable" flies.

Nonparametric hazard rate estimation with associated kernels and minimax bandwidth choice - B., Kaakaï (2026+) (stat theory)

Acute Smurf mortality and age-dependence in a two-phase ageing model - B., Doumic, Kaakaï, Rera (2026+) (application to data)

Main conclusions

- Exponentially **decreasing death hazard rate** : Smurf phenotype is a strong predictor of death.
- **Dependence** between both jumping times with "viable" and "non-viable" flies.

Nonparametric hazard rate estimation with associated kernels and minimax bandwidth choice - B., Kaakaï (2026+) (stat theory)

Acute Smurf mortality and age-dependence in a two-phase ageing model - B., Doumic, Kaakaï, Rera (2026+) (application to data)

What about a **wild population** model?

$$\begin{cases} \frac{\partial n_1(t,a)}{\partial t} + \frac{\partial n_1(t,a)}{\partial a} = -\left(k(a) + c_1^{\text{tot}} N_1(t) + c_2^{\text{tot}} N_2(t)\right) n_1(t,a) \\ n_1(0,t) = \int_0^{+\infty} b(a) n_1(t,a)\, \mathrm{d}a \\ \frac{\mathrm{d}N_2(t)}{\mathrm{d}t} = -\left(d + \tilde{c}_1 N_1(t) + \tilde{c}_2 N_2(t)\right) N_2(t) \\ \qquad + \int_0^{+\infty} \left(k(a) + \eta_1 N_1(t) + \eta_2 N_2(t)\right) n_1(t,a)\, \mathrm{d}a \\ n_1(a,0) = n_{1,0}(a) \in L^1(\mathbb{R}_+) \\ N_2(0) = N_{2,0} \geq 0. \end{cases}$$

Stability analysis and long-time convergence of a partial differential equation model of two-phase ageing - B. (2026+)

Comparison on classic death rate

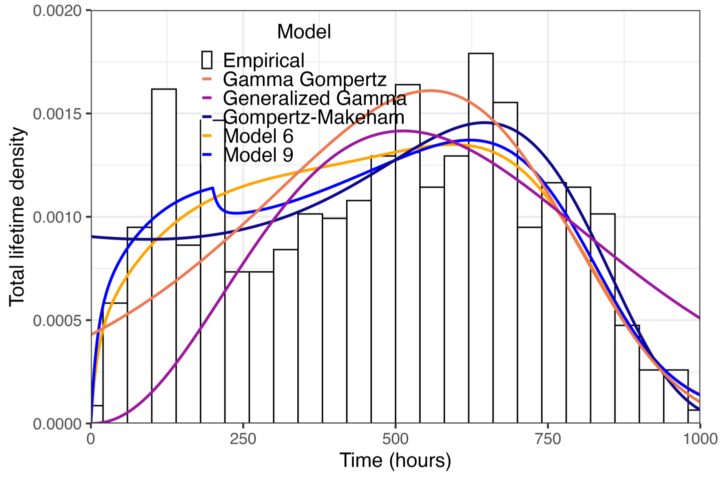


Figure 5: Comparison of parametric classic death rates