MÉMOIRE

Pour obtenir le diplôme

UG**A**
Université
Grenoble Alpes

HABILITATION À DIRIGER DES RECHERCHES DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : MSTII - Mathématiques, Sciences et Technologies de l'Information, Informatique
Spécialité : MATHÉMATIQUES APPLIQUÉES
Unité de recherche : Laboratoire Jean Kuntzmann

A journey in the fields of PDE, probabilities and statistics with point processes
Voyage aux pays des EDP, probabilités et statistiques avec des processus ponctuels

Présentée par :

Julien CHEVALLIER

Rapporteurs :

Vincent RIVOIRARD
Professeur des universités, Université Paris Dauphine - PSL

Delphine SALORT
Professeure des universités, Sorbonne Université

Wilhelm STANNAT
Professeur des universités, Université technique de Berlin

Habilitation soutenue publiquement le 12 décembre 2025, devant le jury composé de :

Eric LUÇON   Examinateur
Professeur des universités, Université d'Orléans

Clémentine PRIEUR   Examinatrice
Professeure des universités, Université Grenoble Alpes

Patricia REYNAUD-BOURET   Examinatrice
Directrice de recherche, CNRS, Université Côte d'Azur

Vincent RIVOIRARD   Rapporteur
Professeur des universités, Université Paris Dauphine - PSL

Delphine SALORT   Rapportrice
Professeure des universités, Sorbonne Université

Wilhelm STANNAT   Rapporteur
Professeur des universités, Université technique de Berlin

A la grande fille dont les yeux pétillants
me donneront toujours le sourire,

A l'amoureuse dont le sourire
me fera toujours pétiller le regard.

Remerciements

Résumé

Les processus ponctuels forment une famille de modèles mathématiques couramment utilisés pour modéliser les instants d'évènements brefs (e.g. ordres boursiers, infections, potentiels d'action) ou la localisation d'objets (e.g. arbres, galaxies, fissures de matériau). Ces modèles sont utilisés, d'un point de vue probabiliste, pour expliquer de manière qualitative certains phénomènes observés sur les configurations temporelles ou spatiales observées en pratique, ou bien, d'un point de vue statistique, pour estimer de manière quantitative les paramètres des modèles à partir de configurations observées.

Ce manuscrit se place en partie dans le cadre détaillé ci-dessus, mais traite également du lien que ces modèles entretiennent avec des modèles déterministes: équations différentielles et équations aux dérivées partielles. En effet, dans un cadre de grande dimension, particulièrement intéressant pour modéliser un grand nombre de neurones en interaction, les théories du champ-moyen et de l'approximation diffusive permettent de formaliser ce lien. Ainsi, les trois chapitres principaux de ce manuscrit (Chapitres II à IV) correspondent à ces trois grands domaines des mathématiques : Équations aux dérivées partielles, Probabilités, Statistique.

D'un côté, le cadre d'application privilégié pour les processus ponctuels temporels est celui des neurosciences: chaque dimension du processus représente un neurone et chaque point représente un potentiel d'action. Des résultats d'existence et unicité pour des modèles déterministes sont détaillés dans le Chapitre II et ensuite utilisés sur des processus ponctuels dans le Chapitre III pour démontrer rigoureusement le lien entre ces deux domaines. D'un point de vue statistique, l'estimation des interactions entre les dimensions du processus est étudiée dans le Chapitre IV.

De l'autre, il n'y a pas de cadre d'application privilégié pour les processus ponctuels spatiaux, mais l'on étudie un nouveau modèle basé sur les points critiques d'un champ aléatoire gaussien. D'un point de vue probabiliste, la simulation de tels processus est étudiée dans le Chapitre III. D'un point de vue statistique, le lien entre les caractéristiques du processus ponctuel et celles du champ sous-jacent ainsi que l'estimation des paramètres de ce dernier sont étudiés, de manière préliminaire, dans le Chapitre IV.

CONTENTS

I

INTRODUCTION

$\mathbf{A}$ fter living in Grenoble for 8 years, mathematical research seems like a mountainous journey: difficult times like a steep slope or when you are lost in the middle of the forest, huge rewards when you reach the summit, and pleasant surprises like a hidden cascade or some wild animal. Two concepts accompanied me on my journey: point processes modeling and neurosciences application. And this trip took me through the fields of partial differential equations, probability and statistics.

With this idea in mind, I chose to organize this manuscript into 3 chapters, one corresponding to each of these fields. Hence, some of my research articles (perhaps most of them) are cited in two different chapters. It comes with an obvious drawback: the natural flow of these articles is cut in two. Nevertheless, I think that the following two advantages win the tradeoff: 1) my contributions to each field is clearly identified, 2) the present manuscript is not just a mere synthesis of my research articles but gives a fresh new point of view.

It may be difficult to classify a contribution as probability or statistics since their boundary is fuzzy. However there is a natural choice here: high dimensional asymptotics, related to mean-field approximations, are detailed in Chapter III (Probability) and large domain (time or space) asymptotics, related to convergence of estimators, are detailed in Chapter IV (Statistics). Each chapter contains its own introduction so I chose not to introduce them in detail here. Instead, the present chapter contains some facts that explain my writing style, an introduction to neurosciences, and finally some notations.

I.1 General facts

My journey was filled with wonderful encounters such as all my mentors or collaborators with a special mention to:

- François Delarue and especially Patricia Reynaud-Bouret who passed on his passion for point processes modeling and neurosciences application to me during my PhD;
- Eva Löcherbach and especially Guilherme Ost who took me by the hand along the most technical parts during my postdoc and since then;
- Rémy Drouilhet who helped me focus on concrete things and passed on his passion for the Julia programming language to me;
- Vincent Brault, Jean-François Coeurjolly and Sophie Achard who accompanied me to the applied statistics field.

They all had an influence on this manuscript, to a greater or lesser extent.

With respect to the "Hedgehog vs Fox" classification[1] popularized by philosopher Isaiah Berlin, I would say that I am a fox which likes to widen my research fields and follow comrades. In particular, I did not study any further mean-field limits since my PhD thesis and postdoc and

---

[1]This classification is well summarized by the sentence attributed to the Ancient Greek poet Archilochus: "A fox knows many things, but a hedgehog knows one big thing.".

I do not give perspectives in its direct lineage. However some perspectives related to diffusion approximations are connected to it.

Concerning mathematical proofs, I consider them as wild and blurry animals. Hence I often want to sketch them using heuristics and simplify them. Most often the simplification step comes with abstraction and generalization. For instance, the uniform quantization in (Chevallier, 2018) takes its roots in (Chevallier et al., 2018) and the modulus of continuity in (Chevallier, 2023) takes its roots in (Chevallier et al., 2021). This desire is also related to pedagogy which is apparent in the present manuscript in several ways.

- Since the content covers three fields of mathematics, I added three crash courses in Appendix A to help non specialist readers.
- I sketched three main objectives of mean-field approximations on some toy example. They are then detailed on two models I studied.
- I developed some heuristics, based on the so-called *chains of approximations*, for the derivation of limit fluctuation equations.
- Sometimes, one gets lost in the jungle of references to research articles or books. Hence references are specified (e.g. Theorem 3, Section 2 or page 42) when relevant in this manuscript.

Most of my early articles do not contain any figure. My interest in numerics and especially scientific reproducibility has vastly grown since then. I take this manuscript as an opportunity to produce some original figures to illustrate theoretical results. The scripts used to produce (almost) all the figures can be found on the jucheval/HDR GitHub repository. Most of the scripts are written in Julia.

Each of the three following chapters (Partial differential equations, Probability and Statistics) is organized as follows: the first section reminds some generalities which are needed to understand my contributions; the intermediate sections correspond to different domains of my contributions; the last section contains some on-going research and perspectives. Then, each of those intermediate sections is organized as follows: 1) a specific introduction with the state of the art, mostly anterior to my contributions; 2) my contributions with new plots and heuristics; 3) research articles I did not contributed to but were highly influenced by my contributions. Notice that the last item is not relevant in every section.

Finally, four appendices complement the main chapters. First, Appendix A contains three crash courses (PDE, probability, statistics). Second, Appendix B contains the exhaustive list of all vector spaces considered in this manuscript and details (at least) their topology. Then, Appendix C contains some thoughts about a simple yet rather unknown fact dealing with two-sided statistical tests. Finally, this manuscript details a selective list of my publications, with less emphasis on those related to my PhD Thesis. The list of the publications that are not detailed can be found in Appendix D.

I.2   Neurosciences

Two modeling scales are mainly considered in the present manuscript. First and mainly, the **microscopic** scale of each individual neuron. Second, the **macroscopic** scale of the whole cortical area.

Neurons form the neural network as they are connected via axons and dendrites. Through these connections, they send signals known as **spikes** following the all-or-none law. Each spike received by a neuron B from a neuron A is integrated positively or negatively whether the connection from A to B is excitatory or inhibitory. Schematically, this integration modulates the voltage of neuron B and, when its voltage reaches some specific (hard) threshold, the neuron B fires a spike.

This is the main feature of the Integrate-and-Fire model for spiking neurons. In that regard, the models considered in this manuscript correspond to a soft threshold, i.e. the firing rate of a neuron depends on its voltage denoted by Ξ or U in Chapter III. Moreover, when a neuron fires, its voltage is reset and the neuron is less sensitive for a while. This phenomenon is called the **refractory period** of a neuron. In this manuscript, it is modelled by firing rates which depends on the **age**, i.e. the time elapsed since the last spike, denoted by A in Chapter III. Finally, the sequence of the spikes of a given neuron is called a **spike train**. Spike trains may be recorded via electrophysiological or calcium imaging techniques for instance. They are commonly depicted as raster plots (see Figure I.1)
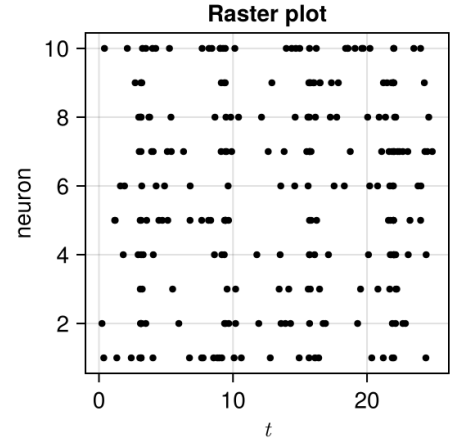
Figure I.1: Raster plot of 10 simulated neurons. Each row corresponds to the spike train of one neuron.

At a larger scale, recordings are made of averaged voltages or firing rates over voxels (i.e. small volumes) of some cortical area. Mean-field approximations are well known to deal with such kind of averages over interacting entities. Hence they are studied by neuroscientists, physicists and mathematicians to bridge the gap between microscopic and macroscopic scales. Such considerations are related to a substantial part of this manuscript.

But macroscopic models are often deterministic and may fail to reproduce some effects inherent to the finite size of the neural network. Hence an intermediate scale called **mesoscopic** scale is also considered here.

I.3   Notation

Here are some conventions on the letters used.

- f is mainly used for firing rate functions, i.e. functions that take as arguments the voltage and possibly the age of a neuron and return its firing rate.
- λ is mainly used for the intensity or firing rate of a neuron as a function of time and possibly position.
- u is used for functions that are solutions of some partial differential equation.
- φ is used for test functions or generic functions.
- Uppercase letters are mostly used for random variables but some of them are linear functionals.

Let $\mathbb{N}$ denote the set of natural integers $0, 1, \ldots$, and $\mathbb{Z}$ denote the set of all integers. Let $\mathbb{R}$ denote the set of real numbers, $\mathbb{R}_+$ denote the set of non-negative real numbers and $\mathbb{R}^d$ denote the set of d-dimensional real vectors. The Euclidean norm is denoted by $\|\cdot\|$ and the sup norm is denoted by $\|\cdot\|_\infty$. The norm, metric or topology equipped to any other space considered in this manuscript can be found in Appendix B.

The notation $:=$ is used to emphasize equalities that are in fact definitions. If a and b are some quantities that depend on some parameters θ_1 and θ_2, the notation $a \lesssim_{\theta_1} b$ means that there exists a constant κ which depends only on θ_1 (and does not depend on θ_2) such that $a \leq \kappa b$.

For any topological space E, let $\mathcal{C}(E)$ denote the space of continuous functions from E to $\mathbb{R}$. The space of k-times continuously differentiable functions is denoted by $\mathcal{C}^k(E)$ and, for non integer exponents β, let $\mathcal{C}^\beta(E)$ denote the space of $\mathcal{C}^{\lfloor \beta \rfloor}$ functions satisfying some local Hölder condition (see Appendix B). Moreover, let $\mathcal{C}^\infty(E)$ denote the space of functions with continuous derivatives of all orders. The subscript b (*respectively* c) is used to denote that any such space

is restricted to boundedly differentiable (*resp.* compactly supported) functions: e.g. $\mathcal{C}_b^k(E)$ is the space of k-times continuously differentiable functions with bounded derivatives of each order less than k. The space of càdlàg functions, i.e. continuous to the right with left limits, is denoted by $\mathcal{D}(E)$. When the target space is not $\mathbb{R}$ but some other topological space F, the notation is transformed: e.g. $\mathcal{C}(E, F)$, etc. The standard notation L^p is used for the Lebesgue spaces, and the notation L_{loc}^p is used for locally integrable functions.

The successive derivatives of a function φ defined on some interval of $\mathbb{R}$ are denoted by $\varphi', \varphi'', \varphi^{(3)}$ and so on. If φ is a function of several variables, e.g. the real variables x_1 and x_2, its partial derivatives are denoted by ∂_{x_1} and ∂_{x_2}. Finally, the k-order partial derivative with respect to x_1 is denoted by $\partial_{x_1}^k$. For a bivariate function φ from $E \times F$, and a fixed point $y^* \in F$, the section $x \mapsto \varphi(x, y^*)$ is denoted by $\varphi(\cdot, y^*)$.

For any set A, its indicator function is denoted by $\mathbf{1}_A$, i.e. $\mathbf{1}_A(x) = 1$ if $x \in A$ and 0 otherwise, and we always use the natural notation with parentheses. For two real numbers x and y, let $x \vee y$ denote their maximum and $x \wedge y$ denote their minimum. The Dirac measure at x is denoted by δ_x. Its infinitesimal is denoted by $\delta_x(\mathrm{d}y)$, and this convention is made for all measures that may be discrete.

For a measure ν on E and a function $\varphi : E \to \mathbb{R}$, the duality bracket notation $\langle \nu, \varphi \rangle :=$ $\int_E \varphi(x)\nu(\mathrm{d}x)$ is used when it makes sense. It is also used for functions $u : E \to \mathbb{R}$, i.e. $\langle u, \varphi \rangle :=$ $\int_E \varphi(x)u(x)\mathrm{d}x$.

Generic random variables with values in some space E are denoted by X. The probability distribution of X is usually denoted by $\mathrm{law}(X)$ or ν, in which case we use the notation $X \sim \nu$. For X and Y two random variables, the expectation of X is denoted by $\mathbb{E}[X] := \int_E x\,\nu(\mathrm{d}x)$ its variance is denoted by $\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$ and the covariance between X and Y is denoted by $\mathrm{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

II

PARTIAL DIFFERENTIAL EQUATIONS

M Y contributions to the field of Partial Differential Equations (PDE) are dealing with two models that are used to describe the dynamics of neural networks. The first one is the **Refractory Density Equation** (RDE), which describes the evolution of neurons by means of their age, that is the time elapsed since their last spike. As an age-structured equation, it is a transport equation with specific source and boundary terms. Moreover, these terms are nonlinear. The second one is the **Neural Field Equation** (NFE), which describes the evolution of a spatially structured neural network. It is also nonlinear but contrarily to the RDE, it contains a nonlocal interaction term.

Both models are derived from stochastic microscopic models. These derivations are based on the assumptions that: 1) the neurons interact through their statistical distribution, 2) the population is large enough. Hence the dynamics can be summarized by a global distribution of states which typically follows some PDE. These assumptions are commonly known as the **mean-field** assumption. The two derivations are loosely discussed in the present chapter but are rigorously stated in the next chapter (Chapter III). The RDE is derived from a microscopic model of age-structured spiking neurons called age dependent Hawkes process, while the NFE is derived from a microscopic model of spatially-structured spiking neurons called spatial Hawkes process.

Section II.1 provides some generalities on PDEs (and in particular transport equations) based on the book (Perthame, 2006). It is intended to provide a foundation for understanding the subsequent sections dealing respectively with the RDE and the NFE. Several footnotes are added to advise non-specialist (in PDEs) readers to read Appendix A.1 for more details.

II.1   Generalities

II.1.a)   Introduction

Let $u : (t, x) \in \mathbb{R} \times \mathbb{R}^d \mapsto u(t, x) \in \mathbb{R}$ be a function defined on the time-space (t, x) domain. A **partial differential equation** (PDE) is an equation which involves the function u and its partial derivatives. A well known example is the heat equation on some interval $I \subset \mathbb{R}$ and some open set $U \subset \mathbb{R}^d$ given by

$$\partial_t u(t, x) - \left(\partial_{x_1}^2 + \cdots + \partial_{x_d}^2 \right) u(t, x) = 0, \quad \text{for all } (t, x) \in I \times U.$$

A **transport equation** is a PDE involving u and its first order derivatives. It describes the evolution of a quantity $u(t, x)$ (e.g., mass, temperature, concentration) being driven by a velocity field $v : (t, x) \in \mathbb{R} \times \mathbb{R}^d \mapsto v(t, x) = (v_1(t, x), \dots, v_d(t, x))^\top \in \mathbb{R}^d$. First consider the transport equation written in the strong form as the Cauchy problem,

$$\begin{cases} \partial_t u(t, x) + \sum_{i=1}^d v_i(t, x) \partial_{x_i} u(t, x) = 0, & (t, x) \in \mathbb{R} \times \mathbb{R}^d, \\ u(0, x) = u^{\text{in}}(x), & x \in \mathbb{R}^d, \end{cases} \tag{TE}$$

where $u^{\mathrm{in}} : x \in \mathbb{R}^d \mapsto u^{\mathrm{in}}(x) \in \mathbb{R}$ is some prescribed initial condition. In the rest of the manuscript, the second line of (TE) is usually implicit and therefore omitted.

Remark II.1. **Age structured equations** are transport equations with constant velocity $v = 1$, where the space variable x describes the age of some entity (e.g. a neuron). Hence, the notation x is often replaced by the notation a for age.

A natural question is whether the solution of (TE) exists and is unique. It is captured by the notion of well-posedness of the PDE.

Definition II.2. A PDE is said to be **well-posed** in the space $\mathcal{G}$ for an initial condition u^{in} if there exists a unique solution $u \in \mathcal{G}$ of the PDE such that $u(0, \cdot) = u^{\mathrm{in}}$.

For instance, the transport equation (TE) is well-posed in the space $\mathcal{C}^1(\mathbb{R} \times \mathbb{R}^d)$ for any initial condition $u^{\mathrm{in}} \in \mathcal{C}^1(\mathbb{R}^d)$ under the standard framework: $b \in \mathcal{C}^1(\mathbb{R} \times \mathbb{R}^d, \mathbb{R}^d)$ satisfies the Cauchy-Lipschitz conditions (Perthame, 2006, Theorem 6.1).

Finally, a common tool used to solve PDEs is the method of characteristics[1] which expresses $u(t, x)$ by means of u^{in} and the characteristic curves of the PDE.

II.1.b)   Weak Solutions

From now on, let us assume that we are only interested in the dynamics for positive times starting from the initial condition u^{in}. The well-posedness result above can be extended to the case where the solution u is not differentiable. In such a case, the weak formulation[2] is needed.

Definition II.3. A function u is a **weak solution** of (TE) if for all $\varphi \in \mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}^d)$, the following holds:

$$- \int_{\mathbb{R}_+ \times \mathbb{R}^d} \left(\partial_t \varphi(t, x) + \sum_{i=1}^d \partial_{x_i} \left[v_i(t, x) \varphi(t, x) \right] \right) u(t, x) \, \mathrm{d}t \mathrm{d}x = \int_{\mathbb{R}^d} \varphi(0, x) u^{\mathrm{in}}(x) \mathrm{d}x.$$

The main advantage of this weak formulation is that a non differentiable function (even a distribution) can satisfy it, contrarily to the strong formulation (TE). For instance, the standard framework described above can be extended (Perthame, 2006, Theorem 6.2): if $u^{\mathrm{in}} \in \mathcal{C}^1(\mathbb{R}^d)$ is replaced by $u^{\mathrm{in}} \in L^\infty(\mathbb{R}^d)$, then (TE) is **weakly** well-posed in the space $L^\infty(\mathbb{R}_+ \times \mathbb{R}^d) \cap \mathcal{C}(\mathbb{R}_+, L^1_{\mathrm{loc}}(\mathbb{R}^d))$, in the sense that there exists a unique weak solution of (TE) in that space.

Finally, (TE) is said to be **weakly-strongly** well posed if the solution is unique in a large (measure) space but belongs to a smaller (function) space (Brenier et al., 2011).

II.1.c)   Source and boundary terms

A **source term** describing modification of the quantity being transported can be added to the transport equation. Adding a source term $s(t, x)$ to the first line of (TE) gives

$$\partial_t u(t, x) + \sum_{i=1}^d v_i(t, x) \partial_{x_i} u(t, x) + s(t, x) = 0.$$

Here, $s(t, x) < 0$ means that some quantity is created at time t and position x, while $s(t, x) > 0$ means that some quantity is destroyed. In the weak formulation, the source term takes the form $\int s(t, x) \varphi(t, x) \, \mathrm{d}t \mathrm{d}x$.

---

[1] See Appendix A.1 for more details.
[2] See Appendix A.1 for the derivation fo the weak formulation.

When the spatial domain admits a boundary, additional **boundary terms** are added to the transport equation as singular source terms. For instance, if the spatial domain is $\mathbb{R}_+$ with boundary at $x = 0$, adding a boundary term $b(t)$ to the first line of (TE) gives

$$\partial_t u(t,x) + \sum_{i=1}^{d} v_i(t,x)\partial_{x_i} u(t,x) = b(t)\delta_0(\mathrm{d}x),$$

where δ_0 is the Dirac measure at $x = 0$. Alternatively, the boundary term can be added as the additional equation $u(t,0) = b(t)$. In the weak formulation, the boundary term takes the form $\int b(t)\varphi(t,0)\,\mathrm{d}t$. Remark that the initial conditioning, second line of (TE), is in fact a boundary term at $t = 0$.

For age structured equations, the source term $s(t,a)$ is usually positive, meaning that some entities with age a are dying at time t. Likewise, the boundary term $b(t)$ is usually positive, meaning that some entities with age $a = 0$ are created at time t.

II.1.d) Fokker-Planck equations

A **Fokker–Planck equation** is a PDE that describes the time evolution of a probability density function, still denoted by u. In particular, positivity ($u(t,x) \geq 0$) and conservation of mass ($\int u(t,x)\mathrm{d}x = 1$ for all t) are expected. For ease of presentation, let us assume until the end of that section that $d = 1$.

The most common Fokker-Planck equation with constant diffusion coefficient $\sigma^2 > 0$ can be written as a "transport equation" with a source term as

$$\partial_t u(t,x) + v(t,x)\,\partial_x u(t,x) + \partial_x v(t,x)\,u(t,x) - \sigma^2\partial_x^2 u(t,x) = 0. \tag{II.1}$$

Remark that the two terms involving u and v adds up to the divergence term $\partial_x[vu] = \partial_x vu + v\partial_x u$. Integrating the equation over the space variable x in $\mathbb{R}$, the last term cancels if $\partial_x u$ vanishes at infinity and the divergence term cancels if the product vu vanishes at infinity thanks to the integration by parts formula. All in all, if derivative and integral commute, mass conservation holds: $\frac{\mathrm{d}}{\mathrm{d}t}\int u(t,x)\mathrm{d}x = \int \partial_t u(t,x)\mathrm{d}x = 0$.

From a probabilistic point of view, this equation is related to the stochastic process $(X_t)_{t\geq 0}$ solution of the stochastic differential equation

$$\mathrm{d}X_t = v(t,X_t)\mathrm{d}t + \sqrt{2}\sigma\mathrm{d}B_t,$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion. Indeed, if $u(t,\cdot)$ denotes the probability distribution of X_t, then Itô's lemma implies that u is a weak solution of Equation (II.1).

Anticipating the rest of this chapter, let us investigate mass conservation for the following transport equation on the spatial domain $\mathbb{R}_+$ with source and boundary terms,

$$\partial_t u(t,x) + \partial_x[vu](t,x) + s(t,x) = b(t)\delta_0(\mathrm{d}x).$$

Assuming once again that all boundary terms vanish at infinity and that derivative and integral commute, integration over the space variable x in $\mathbb{R}_+$ gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}_+} u(t,x)\mathrm{d}x + 0 + \int_{\mathbb{R}_+} s(t,x)\mathrm{d}x = b(t).$$

Hence, the tradeoff $b(t) = \int s(t,x)\mathrm{d}x$ ensures mass conservation. In words, it means that all the mass created or destroyed all along $\mathbb{R}_+$ is compensated at the boundary $x = 0$.

II.2   Age-structured equation

My contributions regarding age-structured equations deal with weak-strong well-posedness and sensitivity analysis. The content of this section is based on the following three articles:

- 📄 J. Chevallier. Mean-field limit of generalized Hawkes processes. *Stochastic Processes and their Applications*, 2017b.
- 📄 J. Chevallier. Fluctuations for mean-field interacting age-dependent Hawkes processes. *Electronic Journal of Probability*, 2017a.
- 📄 J. Chevallier. Stimulus sensitivity of a spiking neural network model. *Journal of Statistical Physics*, 2017c.

The first two are part of my PhD Thesis so I chose not to detail them as much as the rest.

II.2.a)   Introduction

Age-structured equations are transport equations of the general "informal" form

$$\begin{cases} \partial_t u(t,a) + \partial_a u(t,a) + f(t,a;u)u(t,a) = 0, & (t,a) \in \mathbb{R}_+ \times \mathbb{R}_+, \\ u(t,0) = b(t;u). \end{cases} \tag{ASE}$$

With respect to the previous section, the source term is $s(t,a;u) = f(t,a;u)u(t,a)$. It may depend on the solution u itself in a complex manner (local, nonlocal, linear, nonlinear, etc.) thus the notation $s(\cdot;u)$. The same remark applies to the boundary term. Exemples of such explicit dependencies are provided below. The value $u(t,a)$, *respectively* $f(t,a;u)$, describes the density, *resp.* the dying rate, of entities with age a at time t. At this point, the density $u(t,\cdot)$ is not necessarily a probability density function and its mass may not even be conserved.

The history of age-structured equations goes back to McKendrick (1925, page 122) where $f(t,a;u) = f(t,a)$ but he did not specify the boundary term. Independently, it was rediscovered by von Foerster (1959). All in all, age-structured equations are often called Kermack-McKendrick equations or von Foerster model depending on the scientific community (Keyfitz and Keyfitz, 1997).

In the neuroscience context, age structured equations of the Fokker-Planck type are used to describe the dynamics of a population of neurons. The age a is interpreted as the time elapsed since the last spike of a neuron and $u(t,\cdot)$ can be interpreted as the probability[3] density of the age of neurons at time t. The non negative function f describes the firing rate of neurons, and the boundary term $u(t,0) = b(t;u) := \int f(t,a;u)u(t,a)\mathrm{d}a$ represents the mean firing rate across the population and ensures mass conservation. In the neuroscience literature, such an equation is often called **refractory density equation** (Gerstner et al., 2014, Section 14.4).

The particular model of interest in this manuscript is a refractory density equation introduced by Pakdaman et al. (2010). It takes the following form

$$\begin{cases} \partial_t u(t,a) + \partial_a u(t,a) + f(a,\xi(t))u(t,a) = 0, \\ u(t,0) = \int_0^\infty f(a,\xi(t))u(t,a)\mathrm{d}a, \\ \xi(t) := \int_0^t h(t-s)u(s,0)\mathrm{d}s. \end{cases} \tag{RDE}$$

The parameters of the model are:

- The function $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ which describes the firing rate of neurons with respect to their age a and some interaction variable ξ that we call voltage by abuse of language.

---

[3] provided that the initial condition u^{in} is a probability density.

- The delay function $h : \mathbb{R}_+ \to \mathbb{R}$ which describes the influence of neurons' spikes. At time t, the influence of an earlier spike triggered at time $s < t$ is given by $h(t - s)$.

As the convolution between the mean firing rate $u(\cdot, 0)$ and some delay function h, the voltage variable $\xi(t)$ represents the global activity of the network perceived at time t by a neuron. In comparison with the previous paragraph, $\xi(t)$ should be denoted by $\xi(t; u)$ since it depends on the solution u itself. However I prefer to keep this notation to be consistent with the rest of the literature.

Usually, the function f is non-decreasing with respect to its second variable ξ so that positive values of the delay function h means excitation and negative values means inhibition. The function f is also non-decreasing with respect to its first variable a to model the refractory period of neurons; e.g. $f(a, \xi) = 0$ as soon as $a < a^*$ for some $a^* > 0$ models an absolute refractory period of duration a^*.

Figure II.1 shows the solution of (RDE) in two frameworks: the first one exhibits a periodic solution corresponding to synchronous firing, the second one exhibits an aperiodic solution converging to a steady state corresponding to asynchronous firing. In both heatmaps, the characteristics curves, which are lines of slope 1 here, are clearly visible. In the left panel, the dark lines fade away as the age a goes larger than the absolute refractory period $a^* = 1$.
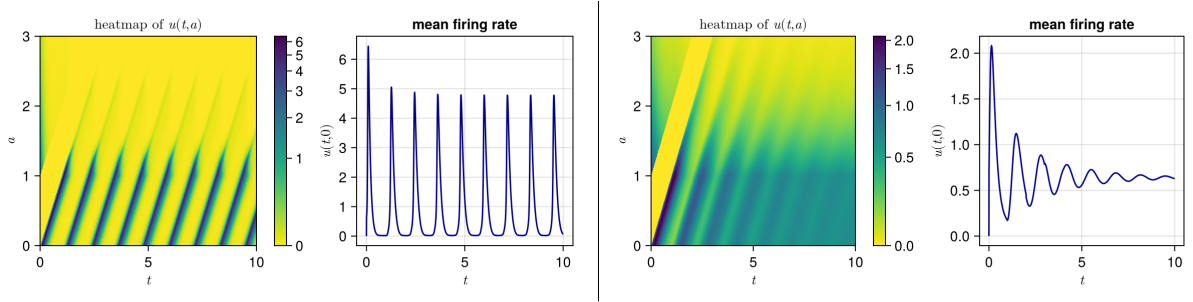
Figure II.1: Heatmap of the solution u and curve of the mean firing rate $u(t, 0)$ as a function of time. In both frameworks, $f(a, \xi) = \phi(\xi)\mathbf{1}_{[a^*, \infty)}(a)$, $h(t) = \alpha e^{-\alpha t}$ with $\alpha = 10$ and $u^{\text{in}}(a) = e^{-(a-1)}\mathbf{1}_{[1, \infty)}(a)$. Left: $\phi(\xi) = 1/2 + 10\xi^2/(\xi^2 + 1)$ as in (Sepúlveda et al., 2026, Example 2). Right: $\phi(\xi) = 1/2 + 5\xi/(\xi + 2)$.

The literature dealing with PDE analysis around the model (RDE) is restricted to the case of excitatory interaction, i.e. $h \geq 0$, with a specific emphasis on the *case without delay* where h is a multiple of the Dirac measure δ_0. In that context, $\gamma := \int_0^\infty h(t)\mathrm{d}t$ is the strength of interaction. The main areas of research are:

- Existence and uniqueness of steady states (Pakdaman et al., 2010).

- Exponential stability of the steady state in weak ($\gamma \approx 0$) and strong ($\gamma \approx \infty$) connectivity regimes. Pakdaman et al. (2010, Theorem 6.1) consider the case where $f(a, \xi) = \mathbf{1}_{a > a^*(\xi)}$ and $h = \gamma\delta_0$. Via spectral techniques, Mischler et al. (2018, Theorem 1.3) extend it to firing rates f that satisfy some Lipschitz condition w.r.t. ξ and delay function h with exponential decay. It is further extended to polynomial decay by Cáceres et al. (2025, Theorem 3) using the comparison principle for Volterra equations.

- Periodic solutions. They are expected to appear in the intermediate connectivity regime as a byproduct of Hopf bifurcations. Pakdaman et al. (2013, Theorems 3-5) proves periodicity for some solutions. Numerical examples can be found in (Pakdaman et al., 2010; Sepúlveda et al., 2026).

Finally, Torres et al. (2021) provides a nice review of the model without delay illustrated by numerical simulations.

II.2.b)   Contributions

The objective of my PhD thesis was to give a rigorous probabilistic interpretation of (RDE). This objective is postponed to Chapter III but the following well-posedness results are a crucial part of its proof.

i. Well-posedness. As a preliminary, I studied the following linearization of (RDE),

$$
\begin{cases}
\partial_t u(t,a) + \partial_a u(t,a) + f^\ell(t,a)u(t,a) = 0, \\
u(t,0) = \int_0^\infty f^\ell(t,a)u(t,a)\mathrm{d}a,
\end{cases}
\tag{ℓ-RDE}
$$

where $f^\ell : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a prescribed function. The superscript $^\ell$ is used to denote quantities related to the linear equation (ℓ-RDE). In the following, let us denote $u \in \mathcal{BC}(\mathbb{R}_+, E)$ if and only if $t \mapsto u(t,\cdot)$ belongs to $\mathcal{C}_\mathrm{b}(\mathbb{R}_+, E)$ and denote $\mathcal{P}(\mathbb{R}_+)$ the space of probability measures on $\mathbb{R}_+$. In this chapter, $\mathcal{P}(\mathbb{R}_+)$ is equipped with the Bounded-Lipschitz metric, see Appendix B. I proved the following weak-strong well-posedness result thanks to the method of characteristics.

Proposition II.4 (Chevallier (2017b, Propositions 3.1 and 3.3)). *Assume that f^ℓ is bounded and continuous with respect to its first variable (uniformly with respect to the second variable) and that $u^\mathrm{in} \in \mathcal{P}(\mathbb{R}_+)$. Then, ($\ell$-RDE) is weakly well-posed in $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ for the initial condition u^in.*

Assume furthermore that u^in admits a bounded density w.r.t. the Lebesgue measure. Then, the unique solution u of (ℓ-RDE) satisfies: i) $u \in \mathcal{BC}(\mathbb{R}_+, L^1(\mathbb{R}_+))$, ii) for all $t \geq 0$, $u(t,\cdot)$ is a bounded probability density, iii) the mean firing rate $t \mapsto u(t,0)$ is continuous.

Here is the main assumption on the firing rate function.

$\left(\mathcal{A}_\mathrm{BL}^f\right):$ | The function $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ is bounded and uniformly Lipschitz continuous with respect to the second coordinate: there exists a constant $\kappa > 0$ such that, for all $a \geq 0$, the function $\xi \mapsto f(a,\xi)$ is κ-Lipschitz.

With all these ingredients, I proved the following weak-strong well-posedness result thanks to a linearization and fixed point argument.

Theorem II.5 (Chevallier (2017b, Theorem 3.5)). *Under $(\mathcal{A}_\mathrm{BL}^f)$, assume that $h \in L^1_\mathrm{loc}$ and that u^in is a bounded probability density.*

Then, (RDE) is weakly well-posed in $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ for the initial condition u^in. Moreover, the unique solution u of (RDE) satisfies: i) $u \in \mathcal{BC}(\mathbb{R}_+, L^1(\mathbb{R}_+))$, ii) for all $t \geq 0$, $u(t,\cdot)$ is a bounded probability density, iii) the mean firing rate $t \mapsto u(t,0)$ is continuous.

ii. Finite-size effects. The refractory density equation (RDE) describes an infinite population of interacting neurons. As such, it may fail to capture some details related to finite-size effects. A way to take these effects into account is to consider a specific stochastic version of (RDE).

Here, I borrow the heuristic derivation developed by Dumont et al. (2017, page 17). Let n be the number of neurons in the population and $u^n(t,a)$ be the finite-size age distribution so that $n \times u^n(t,a)\mathrm{d}a$ gives the number of neurons with age in $[a - \mathrm{d}a, a)$ at time t. During the time increment $\mathrm{d}t$,

- most of the such neurons do not fire and their age increases by $\mathrm{d}a = \mathrm{d}t$,
- the number of such neurons that fire should be close to a Poisson variable $P_{t,a}$ with parameter $n \times f(a, \xi(t))u^n(t,a)\mathrm{d}t\mathrm{d}a$.

Since $\mathbb{E}\left[P_{t,a}\right] = \mathrm{Var}\left(P_{t,a}\right) = n \times f(a, \xi(t))u^n(t,a)\mathrm{d}t\mathrm{d}a$, there is the following Gaussian approximation

$$
P_{t,a} \approx n \times f(a, \xi(t))u^n(t,a)\mathrm{d}t\mathrm{d}a + \sqrt{n \times f(a, \xi(t))u^n(t,a)\mathrm{d}t\mathrm{d}a}Z_{t,a},
$$

where $Z_{t,a}$ is a standard Gaussian random variable. Dividing the equation above by n, the first term corresponds to the source term in (RDE) and the second term corresponds to a stochastic Gaussian noise with variance scaling as $1/n$ yielding the following (informal) nonlinear stochastic refractory density equation:

$$\begin{cases} \partial_t u^n(t,a) + \partial_a u^n(t,a) + f(a,\xi^n(t))u^n(t,a) + \sqrt{\frac{f(a,\xi^n(t))u^n(t,a)}{n}}W(t,a) = 0, \\ u^n(t,0) = \int_0^\infty f(a,\xi^n(t))u^n(t,a) + \sqrt{\frac{f(a,\xi^n(t))u^n(t,a)}{n}}W(t,a)\mathrm{d}a, \\ \xi^n(t) := \int_0^t h(t-s)u^n(s,0)\mathrm{d}s, \end{cases} \quad \text{(SRDE)}$$

where $W(t,a)$ is a space-time white noise. The rigorous derivation of this equation and its exact meaning are discussed in Chapter III.

Figure II.2 shows the solution of (SRDE) in the two frameworks considered above for (RDE).
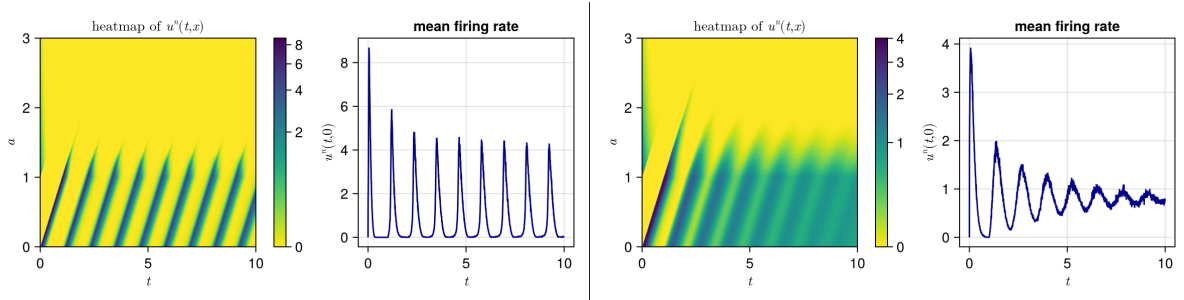
Figure II.2: Heatmap of the solution u^n and mean firing rate $u^n(t,0)$ as a function of time for $n = 100$. The two panels correspond to those of Figure II.1.

iii. Sensitivity analysis. Remind that the strength of interaction is $\gamma = \int_0^\infty h(t)\mathrm{d}t$. The stationary version of the refractory density equation (RDE) reads as

$$\begin{cases} \partial_a u_\infty(a) + f(a,\xi_\infty)u_\infty(a) = 0, \\ u_\infty(0) = \int_0^\infty f(a,\xi_\infty)u_\infty(a)\mathrm{d}a, \\ \xi_\infty := \gamma u_\infty(0). \end{cases} \quad \text{(II.2)}$$

As highlighted in Section II.2.a), the exponential convergence of $u(t,\cdot)$ to u_∞ is proved in the weak ($\gamma < \gamma_\text{weak}$) and the strong ($\gamma > \gamma_\text{strong}$) connectivity regimes.

From here, let us assume that the firing rate is affine with a strict refractory period, i.e. $f(a,\xi) = (\mu + \xi)\mathbf{1}_{[a^*,\infty)}(a)$. Furthermore, assume that the delay function h is non-negative and square integrable. The stationary solution u_∞ and in turn the steady state activity act $:= u_\infty(0)$ are fully parametrized by (μ,γ,a^*). In case the parameter μ represents some spontaneous activity related to some stimulus, it is interesting to study the sensitivity, defined by $\sigma(\mu,\gamma,a^*) := \partial_\mu\text{act}(\mu,\gamma,a^*)$. Hence, the value $\sigma(\mu,\gamma,a^*)$ can be interpreted as the sensitivity of the steady state system with parameters (μ,γ,a^*) to a small increase of the stimulus or the occurrence of a small input signal.

Theorem II.6 (Chevallier (2017c, Theorem 1)). *Suppose the assumptions stated in the paragraph above.*

For every $\mu,a^* > 0$, the sensitivity σ admits a unique maximum with respect to γ. The argument of the maximum only depends on the product μa^* and is denoted by $\gamma_\text{max}(\mu a^*)$. The

function $\gamma_{\max}$ is non-increasing and

$$\begin{cases} \gamma_{\max}(\mu a^*) \to 1, & as \ \mu a^* \to 0, \\ \gamma_{\max}(\mu a^*) = 0, & if \ \mu a^* \geq 1/2. \end{cases} \tag{II.3}$$

Furthermore, $\gamma_{\max} < 1$ which means that maximal sensitivity is reached below criticality.

Sketch of Proof. The proof relies on explicit computations of the steady state activity and, in turn, of the sensitivity σ (Chevallier, 2017c, Propositions 1 and 2). Then, the proof ends by some standard optimization for univariate functions. □

Let us give some intuition on the two regimes in Equation (II.3), supported by the left panel of Figure II.3. If $\mu a^* \approx 0$, the system exhibits a sharp transition between low activity (when $\gamma < 1$) and saturation (when $\gamma > 1$). Hence, optimal sensitivity is reached near the critical value $\gamma = 1$. If μa^* is large, the system is close to saturation whatever the value of γ. Nevertheless, the unconnected regime ($\gamma = 0$) is less saturated and the system has more latitude to sense perturbations.

For biological neurons, the spontaneous activity is around one Hertz whereas the refractory period is around one millisecond. For instance, with $\mu = 2$ Hz and $a^* = 5$ ms, the product $\mu a^* = 0.01$ is small and numerical computations give $\gamma_{\max} \approx 0.97$ (see Figure II.3, right panel). This result shows nice agreement with the estimation performed on *in vivo* spiking activity by Wilting and Priesemann (2018, see Figure 3).
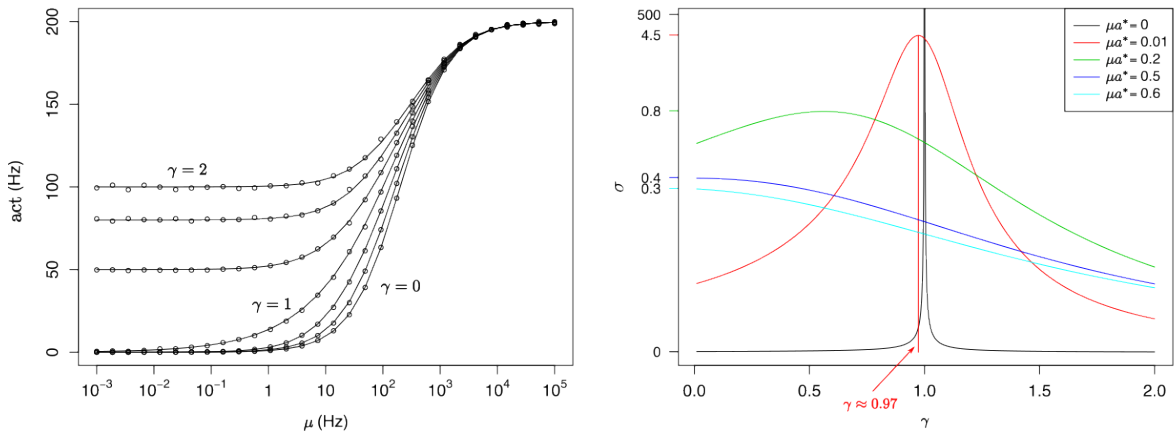
Figure II.3: Left: Steady state activity as a function of μ for fixed refractory period $a^* = 5$ ms and $\gamma \in \{0, 1/3, 2/3, \ldots, 2\}$. The curves depict the explicit expression and the circles depict the activity averaged over the 5000 first spikes of the corresponding particle system (see Section III.2). The x-axis is in log scale. Right: Sensitivity σ as a function of the interaction strength γ for $\mu a^* \in \{0, 0.01, 0.2, 0.5, 0.6\}$. The y-axis differs from one curve to an other (the maximal value of each curve is indicated on the y-axis). The red vertical line corresponds to $\gamma_{\max}(0.01)$.

II.3   Neural field equation

My contributions regarding neural field equations deal with weak-strong well-posedness and regularity of the solutions. The content of this section is based on the following two articles:

📄 J. Chevallier, A. Duarte, E. Löcherbach, and G. Ost. Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. *Stochastic Processes and their Applications*, 2018.

📄 J. Chevallier and G. Ost. Fluctuations for spatially extended Hawkes processes. *Stochastic Processes and their Applications*, 2020.

II.3.a) Introduction

Neural field equations are not transport equations but rather integral-differential equations. The study of such models started with the pioneering works of Wilson and Cowan (1972, 1973) and Amari (1977). The latter contains the first appearance of an equation similar to (NFE) defined below. General reviews on this topic can be found in (Bressloff, 2011; Ermentrout, 1998; Gerstner and Kistler, 2002).

The particular model of interest in this manuscript takes the form

$$\partial_t u(t,x) = -\alpha u(t,x) + \int_{\mathbb{R}^d} w(y,x) f(u(t,y)) \rho(\mathrm{d}y), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d. \quad \text{(NFE)}$$

Here, $u(t,x)$ represents the voltage of a typical neuron at position $x \in \mathbb{R}^d$ and time t. The parameters of the model are:

- The function $f : \mathbb{R} \to \mathbb{R}_+$ which describes the firing rate of neurons with respect to their voltage.
- The leakage rate $\alpha > 0$ which implies an exponential decay of the voltage in case of no interaction.
- The function $w : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ which represents the synaptic weight between neurons. For instance, $w(y,x)$ gives the strength of the influence of spikes from neurons at position y onto neurons at position x.
- The spatial distribution of neurons ρ in $\mathbb{R}^d$.

Usually, the function f is non-decreasing so that positive values of the synaptic weight w means excitation and negative values means inhibition. The integral term in (NFE) describes the (instantaneous) nonlocal interaction between neurons across the whole space. From a modeling point of view, notice that different choices of w and ρ may lead to the same product $w(y,x)\rho(\mathrm{d}y)$ and in turn to the same equation (NFE). In particular it means that inhomogeneous interactions in (NFE) can be considered for instance as:

- an homogeneous function w with inhomogeneous distribution ρ,
- or an inhomogeneous function w with homogeneous distribution ρ.

Figure II.4 shows the solution of (NFE) in two frameworks: the first one is homogeneous in space and exhibits a steady state solution, the second one is inhomogeneous and exhibits a wave propagating at non constant speed. For simplicity of the simulation algorithm, the inhomogeneous framework is implemented with an homogeneous distribution ρ.

Most of the mathematical literature on neural field equations is devoted to the dynamics of travelling waves (Ermentrout and McLeod, 1993) or bumps (Kishimoto and Amari, 1979) and their stability (Zhang, 2003; Veltz and Faugeras, 2010). From a modeling point of view, they are mostly used to describe the dynamics of primary visual cortex where such travelling waves or bumps may occur. For instance, they may be used to model multi-sensory rivalry (Lee et al., 2005) or merging (Forest et al., 2022).

Finally, the dynamics of travelling waves has also been studied for stochastic neural field equations by Kruger and Stannat (2014). Such equations usually take the (informal) form (Faugeras and Inglis, 2015):

$$\partial_t u(t,x) = -\alpha u(t,x) + \int_{\mathbb{R}^d} w(y,x) f(u(t,y)) \rho(\mathrm{d}y) + \sigma(u(t,x)) W(t,x),$$
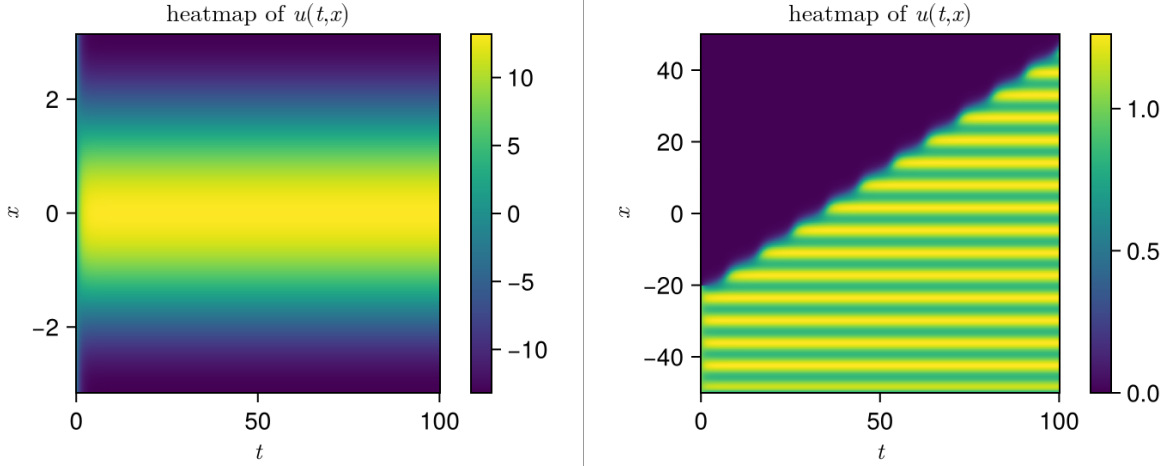
Figure II.4: Heatmap of the solution u. In both frameworks, ρ is the uniform distribution on the spatial domain, $\alpha = 1$ and $f(u) = (1 + \exp(-(u - u^*)/\kappa))^{-1}$ with $\kappa = 1/20$ is a sigmoid approximating the step function $u \mapsto \mathbf{1}_{[u^*, \infty)}(u)$. Left: the spatial domain is $[-\pi, \pi]$ with periodic boundary condition, $u^* = 1/2$, $w(y, x) = 2\pi \cos(y - x)$ and $u^{\mathrm{in}}(x) = 1.8 \cos(x)$ as in (Agathe-Nerine, 2025, Figure 2). Right: the spatial domain is $[-50, 50]$ with no flux boundary condition, $u^* = 0.3$, $w(y, x) = (1 + 0.4 \sin(y))e^{-|y-x|}/2$ and $u^{\mathrm{in}}(x) = \mathbf{1}_{[-50, -20]}(x)$ as in (Coombes and Laing, 2011, Figure 1)

where $W(t, x)$ is a white noise in space and may be colored in space.

In most cases, the spatial distribution ρ is the Lebesgue measure on a bounded or unbounded domain. However, it is crucial to note that my contributions fit in the framework where ρ is finite since this is the most natural framework arising from a microscopic model. In particular, Lebesgue measure on an unbounded domain is excluded here, but Luçon and Poquet (2025) recently treated this case - some details can be found in Subsection III.2.c) below.

II.3.b)   Contributions

The main objective of (Chevallier et al., 2018; Chevallier and Ost, 2020) was to give a rigorous probabilistic interpretation of (NFE). As for (RDE) in the previous section, the following well-posedness result is a crucial part of the proof.

i. Well-posedness. We defined the space $\mathcal{G}$ of **physical solutions** by: $u \in \mathcal{G}$ if and only if u is a continuous function such that

$$t \mapsto \int_{\mathbb{R}^d} \left\{ \int_0^t f(u(s, y)) \mathrm{d}s + \left(\int_0^t f(u(s, y)) \mathrm{d}s \right)^2 \right\} \rho(\mathrm{d}y) \ \text{ is locally bounded.}$$

Hence, a main object of interest is the spatial intensity $\lambda(t, x) := f(u(t, x))$. Here are the main sets of assumptions.

$\left(\mathcal{A}_{\mathrm{BL}}^{f, w, u^{\mathrm{in}}} \right)$: The functions f, w and u^{in} are Lipschitz continuous. Furthermore, u^{in} is bounded and there exist x_0 and y_0 in $\mathbb{R}^d$ such that both $x \mapsto w(y_0, x)$ and $y \mapsto w(y, x_0)$ are bounded.

$\left(\mathcal{A}_{\mathcal{C}^\infty}^{f, w, u^{\mathrm{in}}} \right)$: The functions f, w and u^{in} are $\mathcal{C}^\infty$ and f', f'' are bounded.

The following weak-strong well posedness do not appear verbatim in the literature.

Theorem II.7. *Under $(\mathcal{A}^{f,w,u^{\mathrm{in}}}_{\mathrm{BL}})$, assume that f is invertible.*

Then, (NFE) is weakly well-posed in $\mathcal{G}$ and the spatial intensity λ associated with the unique solution u is the unique solution of

$$\lambda(t,x) = f\left(e^{-\alpha t}u^{\mathrm{in}}(x) + \int_0^t e^{-\alpha(t-s)}\int w(y,x)\lambda(s,y)\rho(\mathrm{d}y)\mathrm{d}s\right). \qquad (\mathrm{II.4})$$

Assume furthermore that $(\mathcal{A}^{f,w,u^{\mathrm{in}}}_{\mathcal{C}^{\infty}})$ is satisfied. Then, the solution u can be identified as a function in $\mathcal{C}(\mathbb{R}_+, \mathcal{C}^{\infty}(\mathbb{R}^d))$.

Proof. Let u be a solution in $\mathcal{G}$ of Equation (NFE). Let $\phi : \mathbb{R} \to \mathbb{R}$ be a positive mollifier, that is $\phi \in \mathcal{C}^{\infty}$ is non negative, compactly supported in $[-1,1]$, $\int \phi(x)\mathrm{d}x = 1$ and, in particular $\int \varepsilon^{-1}\phi(\varepsilon^{-1}x)\varphi(x)\mathrm{d}x \to \varphi(0)$ as $\varepsilon \to 0$ for any continuous function φ. Let $\phi_\varepsilon(x) = \varepsilon^{-1}\phi(\varepsilon^{-1}x)$ denote the rescaled mollifier, and $\Phi_\varepsilon(x) = \int_{-\infty}^{x}\phi_\varepsilon(y)\mathrm{d}y$ denote its integrated version. Notice that Φ_ε is a smooth approximation of the Heaviside function.

With the short notation $\varphi = \varphi(t,x)$, the weak formulation of (NFE) reads as

$$\int (\alpha\varphi - \partial_t\varphi)\,u(t,x)\mathrm{d}x\mathrm{d}t + \int \varphi(0,x)u^{\mathrm{in}}(x)\mathrm{d}x = \int \varphi \int w(y,x)f(u(t,y))\rho(\mathrm{d}y)\mathrm{d}t\mathrm{d}x.$$

Let us fix some $t^* \geq 0$ and $x^* \in \mathbb{R}^d$ and consider the following test function

$$\varphi(t,x) = e^{\alpha t}\Phi_\varepsilon(t-t^*)\,\phi_\varepsilon(x-x^*).$$

With this specific choice of test function, the first factor in the integral simplifies to

$$\alpha\varphi(t,x) - \partial_t\varphi(t,x) = e^{\alpha t}\phi_\varepsilon(t-t^*)\,\phi_\varepsilon(x-x^*).$$

Since $u \in \mathcal{G}$, it is in particular continuous so that letting $\varepsilon \to 0$ gives

$$e^{\alpha t^*}u(t^*,x^*) = u^{\mathrm{in}}(x^*) + \int_0^{t^*} e^{\alpha t}\int w(y,x^*)f(u(t,y))\rho(\mathrm{d}y)\mathrm{d}t.$$

Dividing by $e^{\alpha t^*}$ and changing some dummy variables give that, for all $t \geq 0$, $x \in \mathbb{R}^d$, any solution u satisfies the following Volterra integral equation:

$$u(t,x) = e^{-\alpha t}u^{\mathrm{in}}(x) + \int_0^t e^{-\alpha(t-s)}\int w(y,x)f(u(s,y))\rho(\mathrm{d}y)\mathrm{d}s.$$

Hence, the intensity λ satisfies (II.4) which is exactly (Chevallier et al., 2018, Equation (3.12)) and so it is uniquely determined by Proposition 5 therein. Since f is invertible, u is uniquely determined by λ which ends the proof of the first statement.

The second statement corresponds to (Chevallier and Ost, 2020, Proposition 1)[4]. The proof is based on the following remark: u can be identified as a function in $\mathcal{C}(\mathbb{R}_+, \mathcal{C}(\mathbb{R}^d))$ denoting $u_t(x) = u(t,x)$ and it is a fixed point of the mapping F defined by

$$F(v)_t(x) := e^{-\alpha t}u^{\mathrm{in}}(x) + \int_0^t e^{-\alpha(t-s)}\int w(y,x)f(v_s(y))\rho(\mathrm{d}y)\mathrm{d}s.$$

Then, it is proved that F maps $\mathcal{C}(\mathbb{R}_+, \mathcal{C}(\mathbb{R}^d))$ into $\mathcal{C}(\mathbb{R}_+, \mathcal{C}^{\infty}(\mathbb{R}^d))$ which in turn implies that $u = F(u)$ belongs to $\mathcal{C}(\mathbb{R}_+, \mathcal{C}^{\infty}(\mathbb{R}^d))$. □

---

[4]In (Chevallier and Ost, 2020) the distribution ρ is the Lebesgue measure on $[0,1]$ but this specific result may be extended to any finite distribution.

ii. Finite-size effects. The neural field equation (NFE) describes an infinite population of neurons interacting in a nonlocal manner. As such, it may fail to capture some details related to finite-size effects. A way to capture these effects is to consider a specific stochastic version of (NFE).

Here, I adapt the heuristic derivation proposed in the previous section for (SRDE). Let n be the number of neurons in the population and $u^n(t, y)$ be the mean voltage of neurons near location y in the finite-size population. During the time increment $\mathrm{d}t$, the number of such neurons that fire should be close to a Poisson variable $P_{t,y}$ with parameter $n \times f(u^n(t,y))\mathrm{d}t\rho(\mathrm{d}y)$. Since $\mathbb{E}[P_{t,y}] = \mathrm{Var}(P_{t,y}) = n \times f(u^n(t,y))\mathrm{d}t\rho(\mathrm{d}y)$, there is the following Gaussian approximation

$$P_{t,y} \approx n \times f(u^n(t,y))\mathrm{d}t\rho(\mathrm{d}y) + \sqrt{n \times f(u^n(t,y))\mathrm{d}t\rho(\mathrm{d}y)}Z_{t,y},$$

where $Z_{t,y}$ is a standard Gaussian random variable. Multiplying the equation above by $n^{-1}w(y,x)$ and integrating it with respect to y, the first term corresponds to the integral term in (NFE) and the second term corresponds to a stochastic Gaussian noise with variance scaling as $1/n$ yielding the following (informal) nonlinear stochastic neural field equation:

$$\partial_t u^n(t,x) = -\alpha u^n(t,x) + \int_{\mathbb{R}^d} w(y,x)f(u^n(t,y))\rho(\mathrm{d}y)$$
$$+ \int_{\mathbb{R}^d} w(y,x)\sqrt{\frac{f(u^n(t,y))}{n}}W(t,y)\rho(\mathrm{d}y) \quad \text{(SNFE)}$$

where $W(t,y)$ is a space-time white noise. The rigorous derivation of this equation and its exact meaning are discussed in Chapter III.

Figure II.5 shows the solution of (SNFE) in the two frameworks considered above for (NFE). In the first one, the steady state is replaced by a wandering bump solution. In the second one, the wave still propagates but erratically and at a lower average speed (consistently over several simulations).
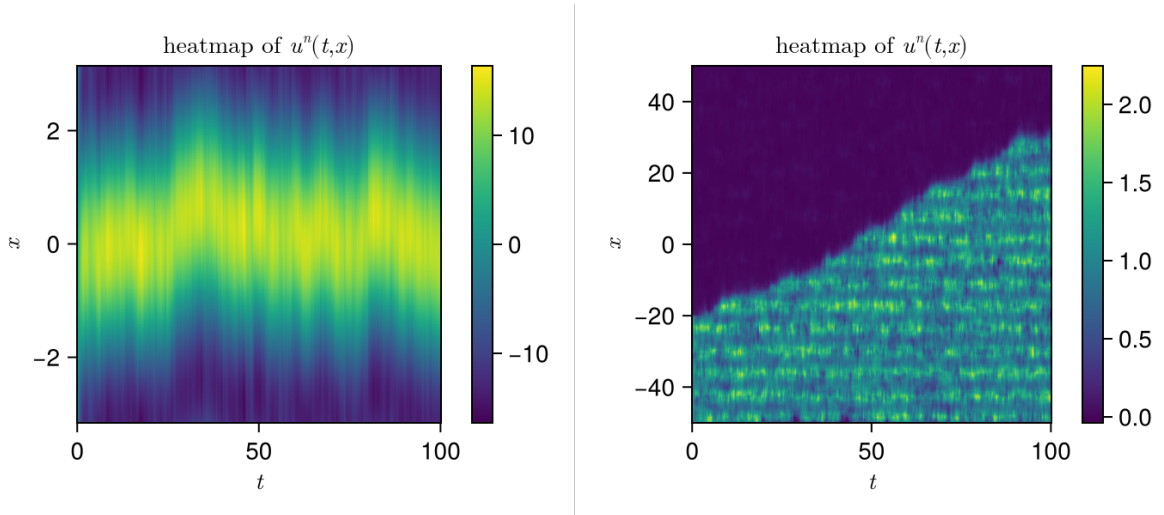
Figure II.5: Heatmap of the solution u^n for $n = 300$. The two panels correspond to those of Figure II.1.

II.4 On going research and perspectives

Blow-up for (RDE)**.** A partial differential equation is said to exhibit a **blow-up** (in finite time) if the solution u or some related quantity goes to infinity at a finite time horizon. In general, the

equation is not well-posed after the blow-up time: for instance, the equation does not make sense anymore, or there are multiple ways to extend the solution after it.

Occurrence of blow-up has been studied for PDE models used in neuroscience, for instance for the Fokker-Planck equation associated with the mean-field limit of Integrate and Fire neurons (Cáceres et al., 2011; Cáceres and Perthame, 2014; Delarue et al., 2015). This is not the case regarding (RDE). In case the firing rate f does not depend on the age a and is rapidly increasing (non Lipschitz) with respect to ξ, it is expected that the firing rate goes to infinity in finite time. Without the inhibition mechanism modeled by the age dependence, it means that every neuron fires an infinite amount of spikes in finite time which is not realistic from a biological point of view. Nevertheless, I believe that a fine interplay between the dependencies with respect to a and ξ may produce more realistic behaviors related to synchronizations. The main mathematical difficulty is to handle the extension of the solution after the blow-up time.

Out-of-equilibrium sensitivity.   In the sensitivity analysis of (RDE) detailed above, only the equilibrium sensitivity is studied. Indeed, it assumes that the system jumps from the steady state corresponding to one value of the constant input to the steady state corresponding to another value. An interesting extension would be to study what happens between the two steady states. For instance, what is the typical duration of the transient state and so the change point detection? Indeed, a system which is less sensitive but has a faster change point detection may be favored over a slower one.

I plan to study such question via numerical simulations on the three modeling scales: macroscopic (RDE), mesoscopic (SRDE) and microscopic ADHP defined in Chapter III.

III

PROBABILITY

$\mathbf{M}$Y contributions to the field of Probability mainly concern mean-field limits and a class of interacting temporal point processes known as **Hawkes processes** introduced by Hawkes (1971).

By **mean-field limit**, we mean an asymptotic framework where a large number of agents (particles, neurons, etc.) interact with each other and such that: 1) the strength of interaction goes to zero as the number of agents goes to infinity, and 2) any pair of agents interact in the same way (or with a specific structure, see Section III.2 for more insights). Since my contributions concern models with application to neuroscience in mind, the agents will be called neurons and the mean-field limit is a convenient framework to derive macroscopic neural models from microscopic spiking models of neurons (Hawkes processes in my case). My objective was to give rigorous derivations of the two equations – Refractory Density Equation (RDE) and Neural Field Equation (NFE) – introduced in Chapter II. As a by-product, I got some stochastic versions of these two equations which can be seen as mesoscopic models in the sense that they are macroscopic models complemented by a stochastic term describing the **fluctuations** coming from the microscopic spiking model. There exist several ways to deal with mean-field limits but the one I prefer is to use **coupling** arguments originally due to Sznitman (1991) and detailed below.

An other way to derive mesoscopic models is to use **diffusion approximations**. In the context of spiking neurons modeling, each spike of a neuron produces a (positive or negative) jump to its connected neurons' voltage. It is common to assume that the height of jumps and the delay between jumps go to zero as the number of neurons goes to infinity. Hence the voltage dynamics is well approximated by diffusion processes. On the one hand, most diffusion approximations are called weak since they are related to convergence in distribution. On the other hand, strong diffusion approximation results deal with **coupling** arguments. That is why I naturally looked at this topic and made some contributions.

Finally, mesoscopic models usually involve **Gaussian random fields**. Hence I studied a bit the regularity of such random fields. Independently, I recently studied the critical points of Gaussian random fields viewed as a point process and proved the efficiency of some simulation methods (detailed in the present chapter) and a central limit theorem with a view towards statistics (detailed in the next chapter).

Section III.1 provides some generalities on probability theory (and in particular stochastic processes). It is intended to provide a foundation for understanding the subsequent sections dealing respectively with mean-field limits, diffusion approximations and Gaussian random fields. Several footnotes are added to advise non-specialist (in probability) readers to read Appendix A.2 for more details.

III.1   Generalities

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a **probability space**[1] and E be a Polish space[2] equipped with its Borel σ-algebra. Appendix B gives the details of the metric or topology equipped to any of the Polish spaces considered in this manuscript.

III.1.a)   Coupling

Most of the results regarding the convergence of (random) distributions are stated in terms of some distance related to coupling. That is why this notion is central in this manuscript.

> **Definition III.1.** Let ν_1 and ν_2 be probability measures on E. A **coupling** of ν_1 and ν_2 is a probability measure π on $E \times E$ such that the marginals of π are ν_1 and ν_2.
>
> By extension, a couple (X_1, X_2) of random variables defined on a common probability space such that their joint distribution is a coupling of ν_1 and ν_2 is called a **r.v.-coupling** of ν_1 and ν_2. Namely, it means that $X_1 \sim \nu_1$ and $X_2 \sim \nu_2$.

The two distributions ν_1 and ν_2 may be close (or even equal) without implying in general that the random variables X_1 and X_2 are close (e.g. if X_1 and X_2 are independent). However, if $X_1 \approx X_2$ almost surely we expect that their respective distributions are close.

From a numerical point of view, a r.v.-coupling gives a very nice way to visualize the fact that two distributions are close (and in turn visualize convergence in distribution), especially in the case of stochastic processes, see Figures III.1 to III.4. From a theoretical point of view, it is the core idea behind coupling metrics such as Wasserstein and total variation.

> **Definition III.2.** Let $p \geq 1$ and ν_1, ν_2 be two probability measures on E with finite p-th moment. The **Wasserstein metric** of order p associated with the metric d is defined by
>
> $$\begin{aligned} W_p(\nu_1, \nu_2) \quad &:= \quad \inf_{\pi} \left(\int_{E \times E} d(x_1, x_2)^p \, \pi(\mathrm{d}x_1, \mathrm{d}x_2) \right)^{1/p}, \\ &= \quad \inf_{(X_1, X_2)} \mathbb{E}\left[d(X_1, X_2)^p \right]^{1/p}, \end{aligned}$$
>
> where the infimum of the first (resp. second) line is taken over all couplings (resp. r.v.-couplings) of ν_1 and ν_2.

> **Definition III.3.** Let ν_1, ν_2 be two probability measures on E. The **Total Variation metric** is defined by
>
> $$\begin{aligned} \mathrm{TV}(\nu_1, \nu_2) \quad &:= \quad 2 \inf_{\pi} \int_{E \times E} \mathbf{1}_{x_1 \neq x_2} \, \pi(\mathrm{d}x_1, \mathrm{d}x_2), \\ &= \quad 2 \inf_{(X_1, X_2)} \mathbb{P}\left(X_1 \neq X_2 \right), \end{aligned}$$
>
> where the infimum of the first (resp. second) line is taken over all couplings (resp. r.v.-couplings) of ν_1 and ν_2.

Since these metrics are defined as infimums over all couplings, it suffices to exhibit a nice (r.v.)-coupling of ν_1 and ν_2 to get for free a nice upper-bound of the Wasserstein or total variation distances between ν_1 and ν_2.

Kantorovich-Rubinstein's duality theorem yields the following: if ν_1 and ν_2 have bounded

---

[1]See Appendix A.2 for more details.
[2]See Appendix A.2 for more details.

support then

$$W_1(\nu_1, \nu_2) = \sup_\varphi \left| \int_E \varphi(x)\nu_1(\mathrm{d}x) - \int_E \varphi(x)\nu_2(\mathrm{d}x) \right|, \tag{III.1}$$

where the sup is taken over all test functions φ that are bounded and 1-Lipschitz.

In the next sections, some representations (stochastic differential equation, time change) of stochastic processes are given. For each of them, an informal argument is given about their utility regarding coupling.

III.1.b)   Stochastic processes and random fields

A **stochastic process** X is a collection of random variables $(X_t)_{t\in\mathcal{T}}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, indexed by a time $t \in \mathcal{T}$ (typically $\mathcal{T} = \mathbb{N}$ or $\mathbb{R}$). When it is not ambiguous, the following short notation is used $X := (X_t)_{t\in\mathcal{T}}$. Notice that a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is a particular case of a stochastic process. The following notations are used for the types of convergence[3]:

- $X_n \xrightarrow{\text{a.s.}} X_\infty$ if $(X_n)_{n\in\mathbb{N}}$ converges almost surely to X_∞;
- $X_n \xrightarrow{\text{p}} X_\infty$ if $(X_n)_{n\in\mathbb{N}}$ converges in probability to X_∞;
- $X_n \xrightarrow{\text{law}} X_\infty$ if $(X_n)_{n\in\mathbb{N}}$ converges in distribution to X_∞;

where X_∞ is some E-valued random variable.

For a tuple $(t_1, \ldots, t_n) \in \mathcal{T}^n$, the joint distribution of $(X_{t_1}, \ldots, X_{t_n})$ is called a finite-dimensional distribution of the process X. For a generic sample point $\omega \in \Omega$, the random function $X(\omega) : \mathcal{T} \to E$ is called a **sample path** of the process. Hence, the process X can be viewed as a random variable with values in the function space $E^\mathcal{T}$. Usually the sample paths are regular and the function space $E^\mathcal{T}$ is replaced by $\mathcal{C}(\mathcal{T}, E)$ or $\mathcal{D}(\mathcal{T}, E)$ for instance.

It is common to describe stochastic processes through their **mean function** $m : t \in \mathcal{T} \mapsto \mathbb{E}[X_t]$ and **covariance function** $c : (t, s) \in \mathcal{T}^2 \mapsto \mathrm{Cov}(X_t, X_s)$. Remark that the definition of the mean and covariance functions may not be trivial when the value space E is not $\mathbb{R}$ or $\mathbb{R}^d$ but a functional or distribution space.

Since the index t represents time, it is common to consider filtrations $(\mathcal{F}_t)_{t\in\mathcal{T}}$ and martingales $(M_t)_{t\in\mathcal{T}}$ which capture the unpredictable fluctuation part of a process[4]. Furthermore, let $(\langle M \rangle_t)_{t\in\mathcal{T}}$ denote the **quadratic variation** of $(M_t)_{t\in\mathcal{T}}$, that is the unique predictable process such that $M_t^2 - \langle M \rangle_t$ defines a martingale.

Finally, a **random field** is similar to a stochastic process except that the index set $\mathcal{T}$ is not related to time. Most common examples are $\mathcal{T} = \mathbb{R}^d$ or $\mathcal{T} = \mathbb{S}^2$ the unit sphere. Obviously, the concepts of filtration and martingales do not carry over to random fields since they are closely related to the natural ordering of time.

III.1.c)   Gaussian processes and fields

A **Gaussian process** (*respectively* **Gaussian random field**) $(X_t)_{t\in\mathcal{T}}$ is a stochastic process (*resp.* random field) such that all its finite-dimensional distributions are normally distributed. Most of its nice properties come from the two following facts: a) zero covariance between two coordinates of a normally distributed vector implies their independence, b) since normal distributions are characterized by their mean and covariance, a Gaussian process is characterized by its mean and covariance functions. In this manuscript, all Gaussian processes are centered in the sense that their mean function is constant equal to zero.

---

[3]See Appendix A.2 for more details.
[4]See Appendix A.2 for more details.

The main example of a Gaussian process is the (1-dimensional) **Brownian motion**, denoted by $(B_t)_{t\geq 0}$, which is the centered real-valued Gaussian process with covariance $\mathrm{Cov}(B_t, B_s) = t \wedge s$. It is well-known that, for any $\varepsilon > 0$, the sample paths of B are almost surely $(1/2 - \varepsilon)$-Hölder continuous. More generally, sample path regularity is related to covariance regularity (see Section III.4 for an example).

III.1.d) Diffusion and time change

A large class of continuous stochastic processes used in modeling and statistics consists of diffusion processes.

i. SDE representation. A **diffusion process** is a continuous-time process $(X_t)_{t\geq 0}$ with continuous sample paths which can be represented by a Stochastic Differential Equation (SDE) of the form

$$\mathrm{d}X_t = v(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, \tag{III.2}$$

where $v : \mathbb{R}^d \to \mathbb{R}^d$ is the **drift coefficient**, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ is the **diffusion coefficient** and $(B_t)_{t\geq 0}$ is an m-dimensional Brownian motion. Remind that such an equation is related to the standard Fokker-Planck equation (II.1).

The representation (III.2) may be used to get a nice coupling with respect to the Wasserstein metric. For instance, let X and $\tilde{X}$ be two processes satisfying (III.2) with respective drift / diffusion coefficients $v, \tilde{v}$ / $\sigma, \tilde{\sigma}$ and the **same** Brownian motion $(B_t)_{t\geq 0}$. If $X_0 = \tilde{X}_0$ and the values of the functions v and $\tilde{v}$ on the one side, σ and $\tilde{\sigma}$ on the other side are close, then the sample paths of X and $\tilde{X}$ should have similar increments at the same time (see Figure III.1).
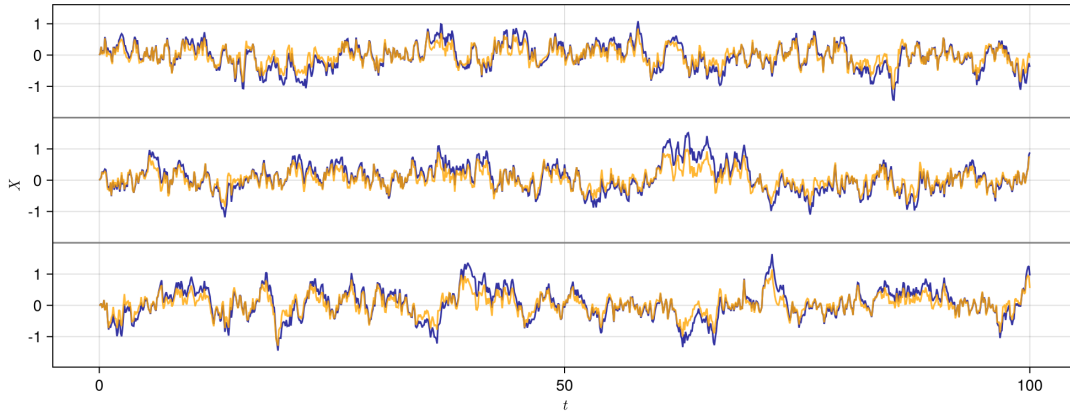
Figure III.1: Three pairs of coupled sample paths of X (in dark blue) and $\tilde{X}$ (in orange) via the SDE representation. The parameters are $v(x) = -x$, $\tilde{v}(x) = -2x$, $\sigma(x) = \tilde{\sigma}(x) = 0.3$ and the initial conditions are $X_0 = \tilde{X}_0 = 0$.

ii. Time change representation. The time change representation can be used for any continuous martingale. For simplicity, we restrict to square integrable martingales so that the quadratic variation $(\langle M \rangle_t)_{t\geq 0}$ is well-defined and the following statement is a weak reformulation of Dubbins-Schwarz theorem (Revuz and Yor, 1999, Theorem V.1.6). Remind that $t \mapsto \langle M \rangle_t$ is non decreasing and define its inverse as $\langle M \rangle_\tau^{-1} := \inf\{t \geq 0, \langle M \rangle_t \geq \tau\}$.

> **Theorem III.4** (Dubbins-Schwarz)**.** *Let $(M_t)_{t\geq 0}$ be a continuous square integrable martingale. Then, $B_\tau = M_{\langle M \rangle_\tau^{-1}}$ defines a Brownian motion and $M_t = B_{\langle M \rangle_t}$.*

This representation may be used to get a nice coupling with respect to the Wasserstein metric. However, it may be difficult to grasp in the general case so let us consider the toy example with deterministic quadratic variations. Assume that a martingale M with quadratic variation $\langle M \rangle$ and a target quadratic variation denoted by QV are given. By simplicity, assume that both are not random. The objective is to construct an other martingale $\tilde{M}$ such that $\langle \tilde{M} \rangle_t = \mathrm{QV}_t$ and $(M, \tilde{M})$ is a nice coupling. First, Let B be the Brownian motion given by Dubbins-Schwarz theorem applied to M. Revuz and Yor (1999, Proposition V.1.5) shows that QV is the quadratic variation of the martingale $\tilde{M}$ defined by $\tilde{M}_t = B_{\mathrm{QV}_t}$. Since $M_t = B_{\langle M \rangle_t}$, the absolute difference $|M_t - \tilde{M}_t| = |B_{\langle M \rangle_t} - B_{\mathrm{QV}_t}|$ is controlled by the modulus of continuity of the Brownian motion B at scale $|\langle M \rangle_t - \mathrm{QV}_t|$. Yet, the modulus of continuity of the Brownian motion B can be controlled with high probability (see Theorem III.29). Hence, if $\langle M \rangle_t \approx \mathrm{QV}_t$ for all $t \geq 0$ then $M_t \approx \tilde{M}_t$.

Figure III.2 illustrates this coupling in a more general framework. In the top panel, both paths start and stay until $t = 40$ in the zone $[0, 5]$ where $\tilde{M}$ is slower (i.e. pass through the same point at a later time). Then, they end in the zone $[15, 20]$ where $\tilde{M}$ is faster. All in all, they almost end at the same point at $t = 100$.
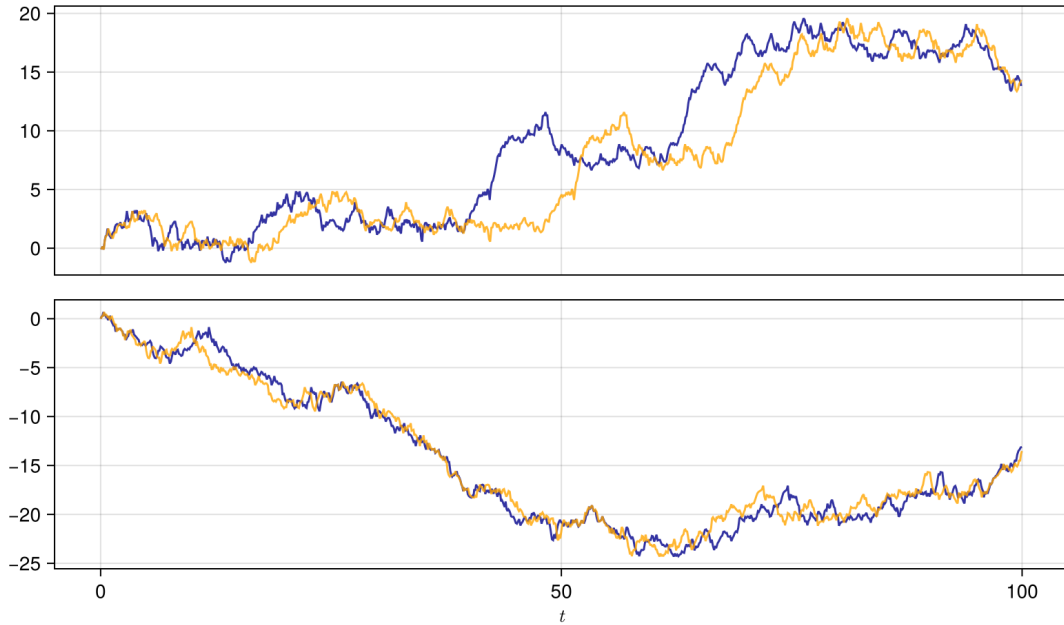
Figure III.2: Two pairs of coupled sample paths of M (in dark blue) and $\tilde{M}$ (in orange) via time change. Time is unchanged for M, i.e. $\mathrm{d}\langle M \rangle_t = \mathrm{d}t$, when $M_t \in \cup_{k \in \mathbb{Z}}[10k, 10k + 5]$ and time is slowed for M, i.e. $\mathrm{d}\langle M \rangle_t = 0.8\mathrm{d}t$ otherwise. The fast and slow zones are reversed for $\tilde{M}$.

III.1.e)   Point processes

A (simple) **point process** is a random collection of points in some space E. If E (*respectively* at least one dimension of E) represents time, then the point process is often called a (*resp.* marked) **temporal point process**. In that case, the mathematical theory is quite different from the pure spatial case. Even if I had some contributions to spatial point processes, I chose to present only temporal ones here. The theory of spatial point processes is postponed to Section IV.3.

For the rest of this section, assume that $E = \mathbb{R}_+$. A temporal point process N is an ordered collection of random variables $0 < T_1 < T_2 < \cdots < T_n < \ldots$. From a modeling point of view, they represent the arrival times of some events (e.g., the spikes of a neuron). That is the reason why the points of a temporal point process are called **spikes** until the end of this manuscript. Without ambiguity, we use three equivalent viewpoints on the point process N:

- as a set, $N = \{T_n, n \geq 1\}$;
- as a measure, $N(\mathrm{d}t) = \sum_{n=1}^{\infty} \delta_{T_n}(\mathrm{d}t)$ where δ_t is the Dirac measure at t;
- as a counting process, $t \in \mathbb{R}_+ \mapsto N_t := N([0,t]) = \sum_{n=1}^{\infty} \mathbf{1}_{[0,t]}(T_n)$.

In particular, the counting process belongs to $\mathcal{D}(\mathbb{R}_+)$, the space of càdlàg functions from $\mathbb{R}_+$ to $\mathbb{R}$. The most common example is the Poisson point process (of intensity $\lambda > 0$). But the notion of intensity can be generalized as follows[5].

For simplicity, assume that all the considered point processes are integrable in the sense that $\mathbb{E}[N_t] < \infty$ for all $t \geq 0$. By Daley and Vere-Jones (2008, Propositions 14.2.I and 14.3.II), if well-defined, the intensity of N with respect to some filtration $(\mathcal{F}_t)_{t \geq 0}$ is defined as the unique $\mathcal{F}_t$-predictable process $(\lambda_t)_{t \geq 0}$ such that

$$\left(N_t - \int_0^t \lambda_s \mathrm{d}s \right)_{t \geq 0} \quad \text{is a martingale with respect to } (\mathcal{F}_t)_{t \geq 0}.$$

However, most point processes can be represented by a Poisson process via a SDE or a time change. These two representations are very useful for simulation and coupling results.

i. Thinning representation. The thinning representation, also called Poisson embedding, is an SDE. It is used to construct a temporal point process N (so in dimension 1) from a Poisson point process in dimension 2. The interested reader is referred to (Daley and Vere-Jones, 2008, Section 14.7) and the following statement is a weak reformulation of (Daley and Vere-Jones, 2008, Proposition 14.7.I).

> **Theorem III.5** (Thinning). *Let $(\lambda_t)_{t \geq 0}$ be a non-negative $\mathcal{F}_t$-predictable process such that $\mathbb{E}[\int_0^t \lambda_s \mathrm{d}s] < \infty$ for all $t \geq 0$. Let Π be a $\mathcal{F}_t$-Poisson process with intensity 1 on $\mathbb{R}_+^2$. Then, $(\lambda_t)_{t \geq 0}$ is the intensity of the point process N defined by*
>
> $$N_t = \int_0^t \int_0^{\infty} \mathbf{1}_{[0,\lambda_s]}(z) \, \Pi(\mathrm{d}s, \mathrm{d}z). \tag{III.3}$$

This representation may be used to get a nice coupling with respect to the total variation metric. For instance, let N and $\tilde{N}$ be two point processes represented as (III.3) via the same Poisson process Π with respective intensities λ and $\tilde{\lambda}$. It is easy to check that any spike $s \in N \setminus \tilde{N}$ must satisfy

$$\tilde{\lambda}_s < \lambda_s \quad \text{and} \quad \exists z \in (\tilde{\lambda}_s, \lambda_s] \text{ such that } (s, z) \in \Pi.$$

Obviously, any spike in $\tilde{N} \setminus N$ must satisfy a similar condition. If $\lambda_s \approx \tilde{\lambda}_s$ for all $s \geq 0$ with high probability then the probability that a spike satisfy this kind of condition is low and in turn $N = \tilde{N}$ with high probability.

Figure III.3 illustrates this coupling in a simple inhomogeneous Poisson process framework. In the first half of the time interval, the intensity of N is larger so that all spikes of $\tilde{N}$ are spikes of N but some spikes of N are not in $\tilde{N}$. The converse holds true in the second half of the time interval.

ii. Time change representation. The time change representation is used to construct a temporal point process N from a Poisson point process in dimension 1. Let $\Lambda_t := \int_0^t \lambda_s \mathrm{d}s$ denote the cumulative intensity and $\Lambda_\tau^{-1} := \inf\{t \geq 0, \Lambda_t \geq \tau\}$ denote its inverse. The interested reader is referred to (Daley and Vere-Jones, 2008, Section 14.6) and the following statement corresponds to Proposition 14.6.III therein.

---

[5]See Appendix A.2 for some heuristics.
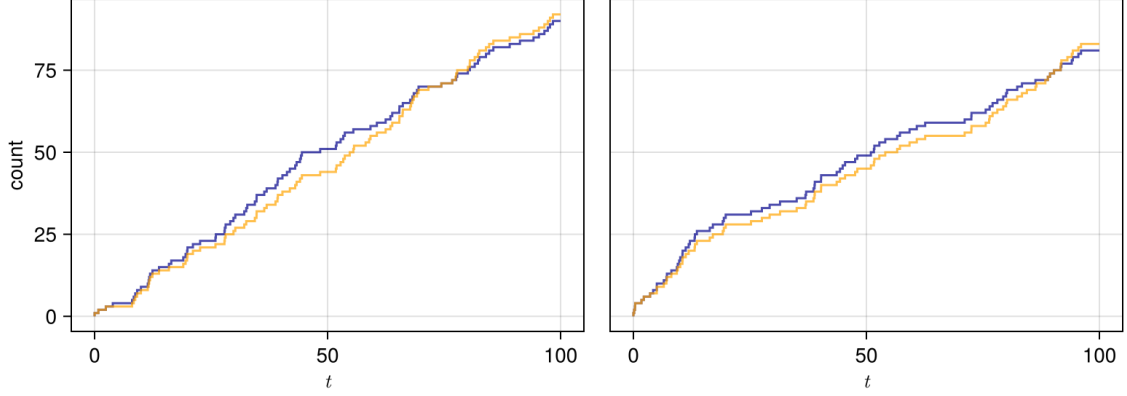
Figure III.3: Two pairs of coupled sample paths of N (in dark blue) and $\tilde{N}$ (in orange) via thinning. The intensities are respectively $\lambda_t = \mathbf{1}_{[0,50]}(t) + 0.9\mathbf{1}_{(50,100]}(t)$ and $\tilde{\lambda}_t = 0.9\mathbf{1}_{[0,50]}(t) + \mathbf{1}_{(50,100]}(t)$.

> **Theorem III.6.** *Let N be a point process with $\mathcal{F}_t$-intensity $(\lambda_t)_{t\geq 0}$. Assume that $\Lambda_t \to \infty$ almost surely as $t \to \infty$.*
>
> *Then, the time changed process Π defined by $\Pi_\tau = N_{\Lambda_\tau^{-1}}$ is a Poisson process with intensity 1 on $\mathbb{R}_+$.*

Conversely, there is the so-called time change representation: N can be viewed as a time changed Poisson process by $N_t = \Pi_{\Lambda_t}$ in the case where $(\mathcal{F}_t)_{t\geq 0}$ is the canonical filtration of N (see Gill et al., 1997, Section II.5.2.3). This representation may be used to couple two point processes in such a way that their spikes are close with high probability. Defining $N_t = \Pi_{\Lambda_t}$ and $\tilde{N}_t = \Pi_{\tilde{\Lambda}_t}$, it is easy to check that $N = \{\Lambda_\tau^{-1}, \tau \in \Pi\}$ and $\tilde{N} = \{\tilde{\Lambda}_\tau^{-1}, \tau \in \Pi\}$. In turn, if $\Lambda_\tau^{-1} \approx \tilde{\Lambda}_\tau^{-1}$ for all τ with high probability then the Hausdorff distance between N and $\tilde{N}$ is small (for all time t in N there exists a time in $\tilde{N}$ which is close to t and conversely).

Figure III.4 illustrates this coupling in the same framework as Figure III.3. In the first half of the time interval, the intensity of N is larger so that the spikes of N arrive more rapidly than those of $\tilde{N}$. The converse holds true in the second half of the time interval. Since the cumulative intensities satisfy $\Lambda_t = \tilde{\Lambda}_t$ for $t = 100$, the coupled processes end at the same point (which is not the case for the coupling induced by thinning, see Figure III.3).
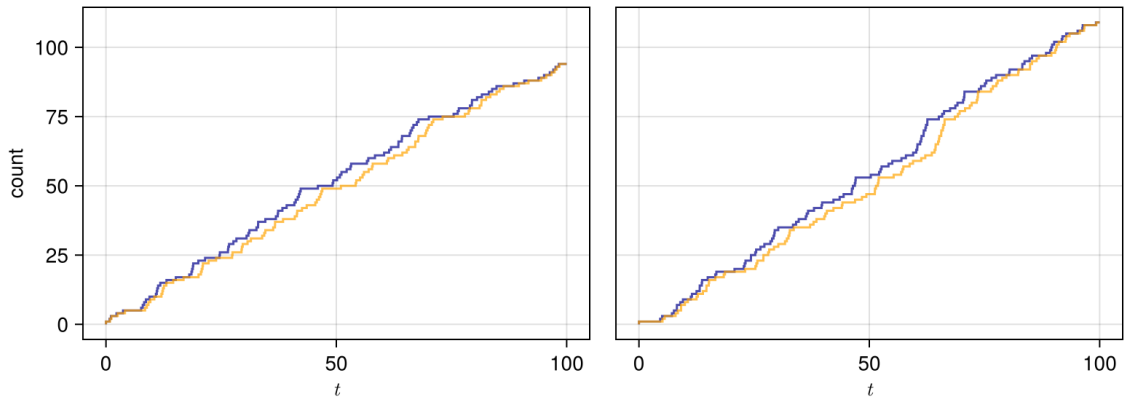
Figure III.4: Two pairs of coupled sample paths of N (in dark blue) and $\tilde{N}$ (in orange) via time change. The intensities are respectively $\lambda_t = \mathbf{1}_{[0,50]}(t) + 0.9\mathbf{1}_{(50,100]}(t)$ and $\tilde{\lambda}_t = 0.9\mathbf{1}_{[0,50]}(t) + \mathbf{1}_{(50,100]}(t)$.

iii. Hawkes process. The main model of interest in this manuscript is the Hawkes process introduced by Hawkes (1971), and its multivariate version which goes back to (Brémaud and Massoulié, 1996). With neurosciences application in mind, it models the spikes of several interacting neurons. It corresponds to point processes $N^1, \ldots, N^n$ with respective intensities, for $i = 1, \ldots, n$,

$$\lambda_t^i = f_i \left(\sum_{j=1}^n \int_0^{t-} h_{j \to i}(t-s) N^j(\mathrm{d}s) \right), \tag{III.4}$$

where $f_i : \mathbb{R} \to \mathbb{R}_+$ is the firing rate of neuron i and $h_{j \to i}$ is the delay function, similarly to Section II.2.a).

This model has been applied in a wide range of scientific domains such as seismology to model aftershocks (Ogata, 1998), genomics to model DNA location (Gusto and Schbath, 2005), finance to model selling stocks (Bacry et al., 2012).

By **linear Hawkes process**, we mean that the firing rate functions f_i are affine with positive slope. In that case, the delay functions must be non-negative to ensure that the intensity stays non negative. In particular, it cannot model inhibition.

When the delay functions are exponential, Hawkes processes enjoy a Markovian structure, therefore their simulation and theoretical study are way more efficient. For instance, if $n = 1$ and the delay function is given by $h(t) = e^{-\alpha t}$ with $\alpha > 0$, then

$$\Xi_t := \int_0^t e^{-\alpha(t-s)} N(\mathrm{d}s),$$

defines a Markov process which furthermore satisfies a simple SDE with jumps. In this example, Ξ follows an exponential decay to 0 at rate α with jumps of height 1. Figure III.5 illustrates this on a bivariate example. Furthermore, this fact extends to Erlang functions and the corresponding SDE can be found in Equation (III.43).
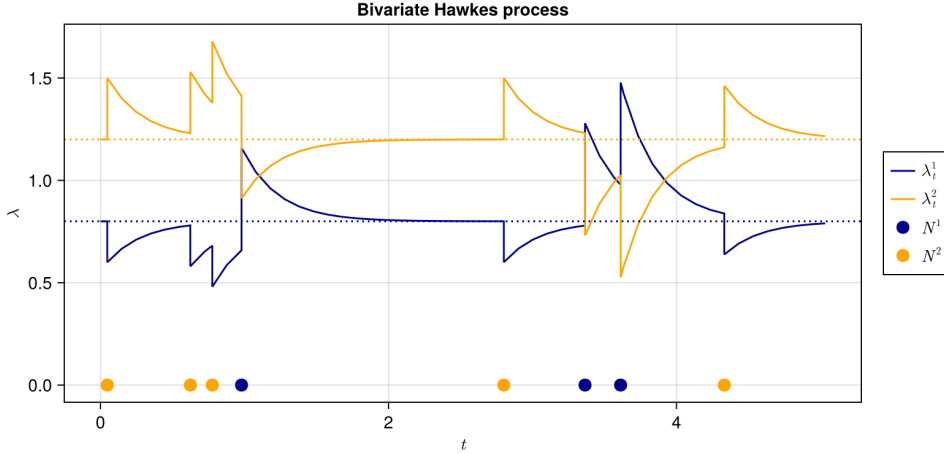
Figure III.5: Sample path of a bivariate Hawkes process (N^1, N^2). Spikes are depicted as circles on the $y = 0$ line and intensities as curves. Darkblue and orange colors respectively correspond to N^1 and N^2. The parameters are $f_i(\xi) = \max(0, \mu_i + \xi)$ with $\mu_1 = 0.8$, $\mu_2 = 1.2$, and $h_{j \to i} = \beta_{j \to i} e^{-4t}$ with $\beta_{1 \to 1} = 0.5$, $\beta_{1 \to 2} = -0.5$, $\beta_{2 \to 1} = -0.2$ and $\beta_{2 \to 2} = 0.3$. The dashed horizontal lines correspond to μ_1 and μ_2.

III.2 Mean-field limits

My contributions regarding mean-field limits deal with quantitative propagation of chaos and functional central limit theorems for high dimensional Hawkes processes. Moreover, I collect here

some minor results obtained for a binary Markov chain model which I mainly studied from a statistical point of view.

The content of this section is based on the following four articles and two preprints:

📄 J. Chevallier. Mean-field limit of generalized Hawkes processes. *Stochastic Processes and their Applications*, 2017b.

📄 J. Chevallier. Fluctuations for mean-field interacting age-dependent Hawkes processes. *Electronic Journal of Probability*, 2017a.

📄 J. Chevallier, A. Duarte, E. Löcherbach, and G. Ost. Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. *Stochastic Processes and their Applications*, 2018.

📄 J. Chevallier and G. Ost. Fluctuations for spatially extended Hawkes processes. *Stochastic Processes and their Applications*, 2020.

📄 J. Chevallier, E. Löcherbach, and G. Ost. Inferring the dependence graph density of binary graphical models in high dimension. *preprint*, 2024.

📄 J. Chevallier and G. Ost. Community detection for binary graphical models in high dimension. *preprint*, 2024.

The first two articles are part of my PhD Thesis so I chose not to detail them as much as the rest. The three groups of publications respectively correspond to the three paragraphs of Subsection III.2.b).

III.2.a) Introduction

The expression **mean-field** occurs in various mathematical domains linked with probabilities (e.g. in- or out-of-equilibrium statistical physics, Bayesian statistics). The key point that unifies this concept across all these domains is to perform a **mean-field approximation**:

Approximate dependent random variables by (well-chosen !) independent ones.

The idea first appeared in physics (statistical mechanics) in the work of Weiss (1907) about ferromagnetism. He proposed the following approximation (Weiss, 1907, page 662): a particle at position x feels a magnetic field produced by all its neighbors and that magnetic field only depends on macroscopic variables at position x. He called it **molecular field**. The expression "mean-field" comes from there: "molecular" was replaced by "mean" because the macroscopic variables mentioned above are related with statistical averages over the distribution of particles. Related ideas have been applied to a wide range of applications such as

- Bayesian inference (Blei et al., 2017) (target distributions are restricted to product distributions i.e. distributions with independent coordinates);
- queueing theory (Baccelli et al., 1992) (particles are customers);
- game theory (Lasry and Lions, 2007; Carmona et al., 2018) (particles are rational agents);
- neurosciences (Brunel and Hakim, 1999; Renart et al., 2004) (particles are neurons).

The main interest of such approximations relies on a drastic diminution in terms of complexity (either theoretical or computational). From the mean-field approximation, Weiss (1907) derived a simple equation and in turn extended Curie's law into Curie-Weiss law. Despite the approximation, this law is very accurate in a specific range of parameters.

Actually, interactions between particles depend on their microscopic state and so are high dimensional and usually considered as stochastic. The difference between the actual interactions and their mean value can be thought as a **fluctuation**. Mean-field approximation assumes that

these fluctuations are null, i.e. interactions are replaced by a "mean-field". In that context, mean-field approximation is the zeroth-order expansion of these fluctuations and can be thought as a law of large numbers on the distribution of particles.

Mean-field approximation often provides a convenient launch point for higher-order approximations. As for the standard central limit theorem, the idea is to derive the limit distribution of the correctly normalized fluctuations.

This manuscript is related with mathematical proofs of such mean-field approximations, in the context of stochastic dynamical models. The main objectives are:

1. Find the (well-chosen) approximating dynamics and derive its Fokker-Planck equation.
2. Prove that this approximation is consistent when the number of particles goes to infinity.
3. Derive a more accurate approximating dynamics.

Here are given some high level ideas for these three objectives on the framework developed by Fernandez and Méléard (1997). By simplicity, we restrict it to dimension $d = 1$. The reader interested by technical details is referred to that article. Let n be the number of particles in the system. Denote by $X_t^{n,1}, \ldots, X_t^{n,n}$ their respective positions in $\mathbb{R}$ at any time $t \geq 0$ and $\nu_t^n := n^{-1} \sum_{i=1}^n \delta_{X_t^{n,i}}$ their empirical measure. The interacting particle system reads as

$$\forall i = 1, \ldots, n, \quad \mathrm{d}X_t^{n,i} = v(X_t^{n,i}, \nu_t^n)\mathrm{d}t + \sigma(X_t^{n,i}, \nu_t^n)\mathrm{d}B_t^i, \tag{III.5}$$

where the $(B_t^i)_{t \geq 0}$'s are i.i.d. Brownian motions. Independently, assume that the initial positions $X_0^{n,1}, \ldots, X_0^{n,n}$ are i.i.d. with common distribution u^{in}.

As usual in (standard) mean-field theory, the particles are exchangeable, i.e. their joint distribution is invariant under permutation of the indices, and each particle does not feel each other particle individually but merely the empirical measure of particles ν_t^n.

Objective 1. It relies on the following simple heuristic: as $n \to \infty$, we expect the particles to behave as i.i.d. copies of some **limit process** $\overline{X} = (\overline{X}_t)_{t \geq 0}$. Hence, we expect ν_t^n to converge to the distribution of $\overline{X}_t$, denoted by $u_t := u(t, \cdot)$. The notation u is used here to anticipate the fact that it is the solution of some Fokker-Planck equation. Replacing ν_t^n by u_t in (III.5) leads to the so-called **limit equation**,

$$\mathrm{d}\overline{X}_t = v\left(\overline{X}_t, u_t\right)\mathrm{d}t + \sigma\left(\overline{X}_t, u_t\right)\mathrm{d}B_t, \tag{III.6}$$

where $(B_t)_{t \geq 0}$ is a Brownian motion and $\overline{X}_0$ is distributed according to u^{in}.

Whereas the well-posedness of the system of coupled particles is generally trivial, the well-posedness of the limit equation is not. It is due to the so called **McKean-Vlasov nonlinearity** induced by the fact that the dynamics of the process $\overline{X}$ depend on its own marginal distributions u_t. Finally, Itô's formula provides its associated Fokker-Planck equation,

$$\begin{cases} \partial_t u(t,x) + \partial_x \{v(x, u_t)u(t,x)\} - \frac{1}{2}\partial_x^2 \{\sigma(x, u_t)u(t,x)\} = 0, \ (t,x) \in \mathbb{R}_+ \times \mathbb{R}, \\ u(0,x) = u^{\mathrm{in}}(x), \quad x \in \mathbb{R}. \end{cases} \tag{III.7}$$

Objective 2. There exist two main methods to prove such a result : 1) consider the sequence $(P^n)_n$ of (random) empirical measures defined by $P^n := n^{-1} \sum \delta_{X^{n,i}}$, prove its tightness and uniqueness of its limit points; 2) couple the interacting particles with i.i.d. solutions of the limit equation. Despite being less general, the second method is less abstract and naturally provides convergence rates in terms of coupling metrics. That is why I prefer it and describe it here.

Let $(\overline{X}_0^i)_{i \geq 1}$ be i.i.d. starting positions distributed according to u^{in}. On the one side, consider the independent limit processes $\overline{X}^1, \overline{X}^2, \ldots$ satisfying

$$\mathrm{d}\overline{X}_t^i = v\left(\overline{X}_t^i, u_t\right)\mathrm{d}t + \sigma\left(\overline{X}_t^i, u_t\right)\mathrm{d}B_t^i.$$

On the other side, for any $n \geq 1$, consider the interacting processes $X^{n,1}, \ldots, X^{n,n}$ satisfying System (III.5) with initial conditions $X_0^{n,i} = \overline{X}_0^i$. For each i, the Brownian motions driving $X^{n,i}$ and $\overline{X}^i$ are the same as well as their initial conditions. Under Lipschitz continuity of v and σ, Grönwall's lemma yields the following upper-bound

$$\mathbb{E}\left[\sup_{s \in [0,t]} |X_s^{n,i} - \overline{X}_s^i|\right] \leq \frac{\kappa}{\sqrt{n}}, \qquad (\text{III.8})$$

for some constant κ independent of n. In particular, it gives the convergence: $(X^{n,1}, X^{n,2}) \xrightarrow{\text{law}} (\overline{X}^1, \overline{X}^2)$. In turn, there is the convergence of the random empirical measure $P^n \xrightarrow{\text{p}} \text{law}(\overline{X})$ thanks to the **propagation of chaos** theory, namely (Sznitman, 1991, Proposition 2.2). Finally, by continuity of the pointwise evaluation, we get a link to the Fokker-Planck equation: for all $t \geq 0$, $\nu_t^n \xrightarrow{\text{p}} u_t$ where u is the solution of (III.7).

Objective 3. Given the later convergence stated above, it is natural to look at the fluctuations, that is the difference between ν_t^n and u_t. The fluctuations must be rescaled to get a non trivial limit. In general, the order of convergence is the one of the usual law of large numbers, i.e. $n^{-1/2}$, but it may be slower: because of smoothed coefficients (Jourdain and Méléard, 1998) or singular interactions (Luçon and Stannat, 2016). In that later case, the limit fluctuations are even not Gaussian, but deterministic.

Let us define the so-called (measure-valued) **fluctuation process** $(\tilde{\nu}_t^n)_{t \geq 0}$ by

$$\tilde{\nu}_t^n = n^{1/2}\left(\nu_t^n - u_t\right).$$

Remark that it is a signed measure which integrates to 0 since ν_t^n and u_t are probability measures. One can prove that $(\tilde{\nu}^n)_n$ converges in distribution to some limit fluctuation $\tilde{\nu}$ by following the trilogy of steps:

1. prove tightness of the sequence $(\tilde{\nu}^n)_n$ in the space $\mathcal{C}([0,t], \mathcal{W})$ where $\mathcal{W}$ is some suitable space of distributions;
2. find the limit equation satisfied by any limit point of $(\tilde{\nu}^n)_n$;
3. prove uniqueness of solutions $\tilde{\nu}$ of that limit equation.

Step 1 usually requires finer upper-bounds than Equation (III.8) such as (Fernandez and Méléard, 1997, Lemma 3.2):

$$\mathbb{E}\left[\sup_{s \in [0,t]} |X_s^{n,i} - \overline{X}_s^i|^4\right] \leq \frac{\kappa}{n^2},$$

for some constant κ independent of n. Finally, the limit fluctuation process $\tilde{\nu}$ satisfies the following weak equation for test functions $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}} \varphi(x)\tilde{\nu}_t(\mathrm{d}x) - \int_{\mathbb{R}} \varphi(x)\tilde{\nu}_0(\mathrm{d}x) = \int_0^t \int_{\mathbb{R}} L_s\varphi(x)\tilde{\nu}_s(\mathrm{d}x)\mathrm{d}s + \int_{\mathbb{R}} \varphi(x)M_t(\mathrm{d}x),$$

where L_s is a linear operator acting over test functions and M is a measure-valued Gaussian process. Their precise definitions can be respectively found in (Fernandez and Méléard, 1997, page 50 and Theorem 5.2). That concludes this last objective.

Like the framework presented above, most studies focused on diffusion processes where all particles are exchangeable. Here are some extensions that were inspiring for my contributions.

- Replace diffusion processes by counting processes like Delattre et al. (2016) did with Hawkes processes.

- Say that each particle is attached to a discrete covariate (e.g. an integer between 1 and $k \ll n$). Classes correspond to particles with the same covariate. The exchangeability assumption may be relaxed to within-class exchangeability as developed by Graham (2008) with the concept of multi-class propagation of chaos. This approach has been considered by Ditlevsen and Löcherbach (2017) for Hawkes processes.
- Even continuous covariates have been considered by Luçon and Stannat (2014). It is related to the concept of spatial propagation of chaos.

III.2.b) Contributions

My contributions deal with three models: age dependent Hawkes process, spatially extended Hawkes process, and a binary Markov chain in a random environment. Each one is presented in a separate section following the three objectives presented in the previous section.

i. Age dependent Hawkes process. First of all, let us define the age associated with a point process. To any point process N, one can associate an age process A defined by, for all $t > 0$,

$$A_t := t - \sup\{T \in N, T \leq t\} = t - T_{N_t}.$$

Assume that $N \cap (-\infty, 0) \neq \emptyset$ almost surely and that T_0 is the last spike before time $t = 0$ so that the initial age A_0 is well-defined. In words, the age A_t counts the time elapsed since the last spike of N before time t. It is a càdlàg process which jumps to zero at each spike of N. For measurability reasons, the predictable version of the age must be considered when defining the intensity of a point process[6]. In that case, we use the notation $A_{t-} := t - \sup\{T \in N, T < t\}$. Finally, let $N_- := N \cap (-\infty, 0)$ and $N_+ := N \cap [0, +\infty)$.

In a first attempt (Chevallier et al., 2015), we derived the age structured equation satisfied by the distribution of the age for several models of single spike trains (e.g. Poisson, renewal, Wold, Hawkes). It appeared to be too complex in general and the good approach is the mean-field framework described below.

Let $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ and $h : \mathbb{R}_+ \to \mathbb{R}$ be respectively firing rate and delay functions. The definition of an age dependent random Hawkes process is given by providing its thinning representation.

Definition III.7. Let $(N_-^{n,i})_{i=1,..,n}$ be i.i.d. point processes on $(-\infty, 0]$. Let $(\Pi^i)_{i\geq 1}$ be i.i.d. $\mathcal{F}_t$-Poisson processes with intensity 1 on $\mathbb{R}_+^2$. A family of counting processes $(N^{n,i})_{i=1,...,n}$ on $\mathbb{R}_+$ (respectively completed by $N_-^{n,i}$ on negative times) satisfying

$$N_t^{n,i} = \int_0^t \int_0^\infty \mathbf{1}_{\left[0, f(A_{s-}^{n,i}, \Xi_s^n)\right]}(z) \, \Pi^i(\mathrm{d}s, \mathrm{d}z), \tag{III.9}$$

where $A^{n,i}$ is the age process of $N^{n,i}$ and

$$\Xi_t^n := \frac{1}{n} \sum_{j=1}^n \int_0^{t-} h(t-s) N^{n,j}(\mathrm{d}s), \tag{III.10}$$

is called an **age dependent Hawkes process** (ADHP). In particular, the intensity of $N^{n,i}$ is $\lambda_t^{n,i} = f(A_{t-}^{n,i}, \Xi_t^n)$.

Notice that it corresponds to a simple case of (Chevallier, 2017b, Definition 2.2 and Representation 2.5) where: 1) the delay functions may differ for each pair (i,j) of neurons in a random and

---

[6]In comparison, the voltage variable Ξ defined in (III.10) is already predictable to avoid additional notation.

almost i.i.d. manner, 2) some external input can be added. The exact distribution of the point processes on negative times, namely $(N_{-}^{n,i})_{i=1,..,n}$ is not of particular interest here. Nevertheless, assume that they contain at least one spike so that the age $A_0^{n,i}$ is well-defined and the distribution of $A_0^{n,i}$ is some specified initial condition u^{in}.

Figure III.6 shows a simulation of an ADHP of dimension $n = 100$ in the two frameworks considered above for (RDE).
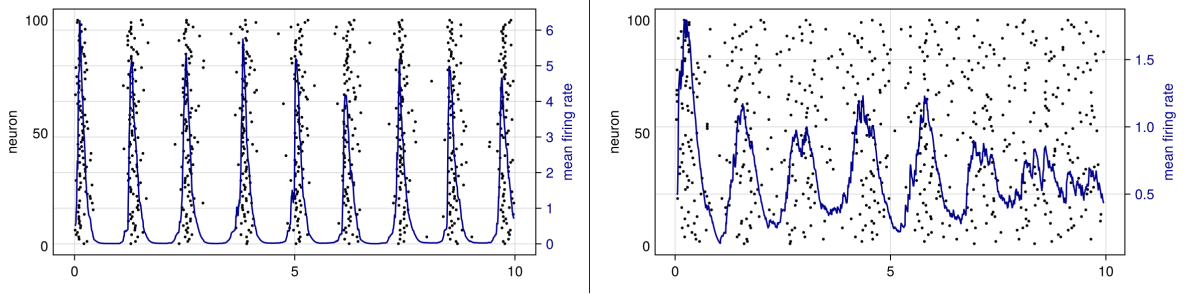
Figure III.6: Superimposition of a raster plot and mean firing rate $n^{-1} \sum_{i=1}^n f(A_{t-}^{n,i}, \Xi_t^n)$ as a function of time for $n = 100$. The two panels correspond to those of Figures II.1 and II.2.

Remind that our first objective is to find the limit equation and derive its Fokker-Planck equation. Our particles here are described by point processes so our limit process is a point process: it is denoted by $\overline{N}$ and its age process is denoted by $\overline{A}$. Let us focus on the intensity of $\overline{N}$. The empirical measure of the particles appear in Equation (III.10) therefore we expect the intensity of $\overline{N}$ to satisfy the McKean-Vlasov equation $\overline{\lambda}_t = f(\overline{A}_{t-}, \xi(t))$ where

$$\xi(t) = \int_0^{t-} h(t-s)\mathbb{E}\left[\overline{N}(\mathrm{d}s)\right].$$

As a corollary of the weak-strong well-posedness of (RDE), i.e. Theorem II.5, I proved in (Chevallier, 2017b) that the limit process is well-defined (Proposition 3.7 therein) and that the Fokker-Planck equation of the limit age process $\overline{A}$ is (RDE) (Proposition 3.9 therein).

Our second objective is to prove the consistence of the mean-field approximation by a coupling argument. Hence, we couple the processes given by (III.9) to the i.i.d. limit processes $(\overline{N}^i)_{i\geq 1}$ satisfying

$$\overline{N}_t^i = \int_0^t \int_0^\infty \mathbf{1}_{\left[0, f(\overline{A}_{s-}^i, \xi(s))\right]}(z)\Pi^i(\mathrm{d}s, \mathrm{d}z). \tag{III.11}$$

Let us mention that their initial conditions are also coupled in the sense that $N_{-}^{n,i} = \overline{N}_{-}^i$ and in particular $A_0^{n,i} = \overline{A}_0^i$. Let $\nu_t^n := n^{-1} \sum_{i=1}^n \delta_{A_t^{n,i}}$ and $P^n := n^{-1} \sum_{i=1}^n \delta_{A^{n,i}}$. They are probability measures on $\mathbb{R}_+$ and $\mathcal{D}(\mathbb{R}_+)$ respectively. Here is stated the convergence rate of the coupling and a link between the particle system and the limit Fokker-Planck equation (RDE) via Equation (III.13).

> **Theorem III.8** (Chevallier (2017b, Theorem 4.1 and Corollary 4.5)). *Under* $(\mathcal{A}_{\text{BL}}^f)$, *assume that* $h \in L_{\text{loc}}^2$ *and that* u^{in} *is a bounded probability density supported in* $[0, \kappa]$.
>   *Then, for all* $t \geq 0$,
>
> $$\mathbb{E}\left[\sup_{s\in[0,t]} |A_s^{n,i} - \overline{A}_s^i|\right] \lesssim_{t,\kappa} \mathbb{P}\left((A_{s-}^{n,i})_{s\in[0,t]} \neq (\overline{A}_{s-}^i)_{s\in[0,t]}\right) \lesssim_{t,f,h} n^{-1/2}, \tag{III.12}$$
>
> *where* $A^{n,i}$ *and* $\overline{A}^i$ *are the age processes associated with* (III.9) *and* (III.11). *In turn,*

$$\sup_{s\in[0,t]} \mathbb{E}\left[W_1(\nu_s^n, u_s)\right] \lesssim_{t,f,h,\kappa} n^{-1/2} \tag{III.13}$$

and the following convergence holds,

$$P^n \xrightarrow{\mathrm{p}} \mathrm{law}\left(\overline{A}\right).$$

Our last objective is to refine this link by studying the fluctuations. Remind that the first step is to prove tightness of the fluctuations and that finer upper-bounds are required to prove tightness. In contrast to the diffusion case where finer upper-bounds are obtained through higher moments, I found that in the jump case they are obtained through total variation: e.g.

$$\mathbb{P}\left((A_{s-}^{n,1})_{s\in[0,t]} \neq (\overline{A}_{s-}^1)_{s\in[0,t]} \text{ and } (A_{s-}^{n,2})_{s\in[0,t]} \neq (\overline{A}_{s-}^2)_{s\in[0,t]}\right) \lesssim_{t,f,h} n^{-1}. \tag{III.14}$$

Such upper-bound would be trivial if the particles were independent but can be extended to our asymptotic independence case. The precise statement can be found in (Chevallier, 2017a, Proposition 3.1). Its proof relies on fine computations involving combinatorics, martingale properties of the point processes, Rosenthal's inequality and, in fine, some Grönwall's lemma.

Here, two fluctuation processes must be considered: the age distribution fluctuations $\tilde{\nu}_t^n := n^{1/2}(\nu_t^n - u_t)$, and the voltage fluctuations $\tilde{\Xi}_t^n := n^{1/2}(\Xi_t^n - \xi(t))$. Notice that $\tilde{\nu}_t^n$ is a distribution on $\mathbb{R}_+$ whereas $\tilde{\Xi}_t^n \in \mathbb{R}$. In particular, the space of distributions in which $\tilde{\nu}_t^n$ lives must be specified. I followed an Hilbertian approach introduced by Fernandez and Méléard (1997) and considered weighted Sobolev spaces. For simplicity, let us fix some $\alpha > 1/2$ for the rest of this section and consider the weighted Sobolev space $\mathcal{W}^k := \mathcal{W}_0^{k,\alpha}$ and $\|\cdot\|_k = \|\cdot\|_{k,\alpha}$ as defined in Appendix B. The index k is an integer and $\mathcal{W}^{-k}$ is the dual of $\mathcal{W}^k$. Informally, the index k controls the regularity of the test functions when $k > 0$ or the distributions when $k < 0$. We are now in position to follow the trilogy of arguments: tightness, characterization, uniqueness.

Here is given an additional set of assumptions.

$\left(\mathcal{A}_{\mathcal{C}_b^2,\mathrm{H\ddot{o}l,BS}}^{f,h,u^{\mathrm{in}}}\right):$ | The firing rate function $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ is $\mathcal{C}_b^2$. The delay function h is Hölder continuous. The initial condition u^{in} is bounded with bounded support.

Proposition III.9 (Chevallier (2017a, Theorem 4.11 and Proposition 5.5)). *Under $(\mathcal{A}_{\mathcal{C}_b^2,\mathrm{H\ddot{o}l,BS}}^{f,h,u^{\mathrm{in}}})$, the sequence of the laws of $(\tilde{\nu}^n, \tilde{\Xi}^n)$ is tight in $\mathcal{D}(\mathbb{R}_+, \mathcal{W}^{-2} \times \mathbb{R})$. Furthermore, a random process $(\tilde{\nu}_t, \tilde{\Xi}_t)_{t\geq 0}$ distributed according to any limit law of this tight sequence takes values in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}^{-2} \times \mathbb{R})$.*

The second step is to derive the system satisfied by any limit fluctuation process $(\tilde{\nu}_t, \tilde{\Xi}_t)_{t\geq 0}$. Let us first decompose the voltage fluctuations $\tilde{\Xi}^n$. According to (RDE), remind that $\xi(t) = \int_0^t h(t-s)u_s(0)\mathrm{d}s$ and $u_t(0) = \langle u_t, f(\cdot, \xi(t))\rangle = \int_0^\infty f(a, \xi(t))u(t,a)\mathrm{d}a$ with the duality bracket notation. We have

$$\tilde{\Xi}_t^n = n^{1/2} \int_0^t h(t-s)\left\{\frac{1}{n}\sum_{i=1}^n N^{n,i}(\mathrm{d}s) - \langle u_s, f(\cdot, \xi(s))\rangle \,\mathrm{d}s\right\}.$$

The difference in curly brackets can be decomposed through the chain of approximations

$$N^{n,i}(\mathrm{d}s) \underset{(1)}{\leftrightarrow} \lambda_s^{n,i}\mathrm{d}s = f(A_{s-}^{n,i}, \Xi_s^n)\mathrm{d}s \underset{(2)}{\leftrightarrow} f(A_{s-}^{n,i}, \xi(s))\mathrm{d}s \underset{(3)}{\leftrightarrow} \langle u_s, f(\cdot, \xi(s))\rangle \,\mathrm{d}s. \tag{III.15}$$

Up to the $n^{1/2}$ factor, they lead to three contributions: (1) is related to the martingale measure associated with the point process $N^{n,i}$; (2) is approximated by $\partial_\xi f(A_{s-}^{n,i}, \xi(s))\tilde{\Xi}_s^n$ using Taylor expansion; (3) gives a sum over $i = 1, \ldots, n$ which is equal to $\langle \tilde{\nu}_s^n, f(\cdot, \xi(s))\rangle$.

Let us now decompose the age fluctuations $\tilde{\nu}^n$ by considering its action $\langle \tilde{\nu}_t^n, \varphi \rangle = n^{1/2} (\langle \nu_t^n, \varphi \rangle - \langle u_t^n, \varphi \rangle)$ over test functions φ in $\mathcal{W}^2$. Age processes are always increasing with rate 1 and sometimes jumps back to 0: their dual actions respectively correspond to the differentiation $\varphi \mapsto \varphi'$ and **reset** $R : \varphi \mapsto \varphi(0) - \varphi$. On the one side, since the age process $A^{n,i}$ resets whenever t is a spike of $N^{n,i}$, we have (Chevallier, 2017a, Equation (A.5)):

$$\langle \nu_t^n, \varphi \rangle = \frac{1}{n} \sum_{i=1}^n \varphi(A_t^{n,i}) = \langle \nu_0^n, \varphi \rangle + \int_0^t \langle \nu_s^n, \varphi' \rangle \, \mathrm{d}s + \frac{1}{n} \sum_{i=1}^n \int_0^t R\varphi(A_{s-}^{n,i}) N^{n,i}(\mathrm{d}s). \qquad \text{(III.16)}$$

On the other side, writing the same kind of equation for the limit age process and taking its expectation, we have (Chevallier, 2017a, Equation (A.7)):

$$\langle u_t, \varphi \rangle = \langle u_0, \varphi \rangle + \int_0^t \langle u_s, \varphi' \rangle \, \mathrm{d}s + \int_0^t \langle u_s, f(\cdot, \xi(s)) R\varphi \rangle \, \mathrm{d}s.$$

As it appears above, the dynamics of u is governed by the linear operator[7] L_t defined by $L_t \varphi(a) = \varphi'(a) + f(a, \xi(t)) R\varphi(a)$ so that the sum of the two last terms in the previous equation writes as $\int_0^t \langle u_s, L_s \varphi \rangle \, \mathrm{d}s$. In comparison, the sum of the two last terms in (III.16) could be written as $\int_0^t \langle \nu_s^n, L_s \varphi \rangle \, \mathrm{d}s$ provided we could replace $N^{n,i}(\mathrm{d}s)$ by $f(A_{s-}^{n,i}, \xi(s)) \mathrm{d}s$. Such a difference can be decomposed through the steps (1) and (2) in the chain of approximations (III.15). In comparison with the terms involved in the decomposition of the voltage fluctuations $\tilde{\Xi}^n$, let us remark that the factor $h(t - s)$ is replaced by $R\varphi(A_{s-}^{n,i})$. Finally, the formalization of these arguments correspond to (Chevallier, 2017a, Equations (5.5) and (5.6)) with explicit rest terms corresponding to the several approximations.

Here is the last ingredient needed to derive the limit fluctuation system. It is specifically useful to study contribution (1) from the chain of approximations (III.15). Let M^n be the $\mathcal{W}^{-2}$-valued martingale defined by

$$\langle M_t^n, \varphi \rangle = n^{-1/2} \sum_{i=1}^n \int_0^t \varphi(A_{s-}^{n,i}) \left\{ N^{n,i}(\mathrm{d}s) - \lambda_s^{n,i} \mathrm{d}s \right\},$$

so that the usual martingale associated with the counting process $n^{-1/2} \sum_{i=1}^n N^{n,i}$ is $\langle M_t^n, \mathbf{1} \rangle$, where $\mathbf{1}$ denotes the constant function equal to 1. For any two test functions φ_1, φ_2, the covariance of $\langle M_t^n, \varphi_1 \rangle$ and $\langle M_t^n, \varphi_2 \rangle$ is

$$\frac{1}{n} \sum_{i=1}^n \int_0^t \varphi_1(A_{s-}^{n,i}) \varphi_2(A_{s-}^{n,i}) \lambda_s^{n,i} \mathrm{d}s,$$

which leads to the following definition.

Definition III.10. Let M be a continuous centred Gaussian process with values in $\mathcal{W}^{-2}$ and covariance given by

$$\mathbb{E} \left[\langle M_{t_1}, \varphi_1 \rangle \langle M_{t_2}, \varphi_2 \rangle \right] = \int_0^{t_1 \wedge t_2} \langle u_s, \varphi_1 \varphi_2 f(\cdot, \xi(s)) \rangle \, \mathrm{d}s.$$

Here is the limit system (with the convention that all the terms on the right hand sides are ordered according to the three steps in the chain of approximations (III.15)).

---

[7]This linear operator is the infinitesimal generator associated with the limit process $\overline{X}$. I chose not to describe what it is in this manuscript but it is related to a semi-group structure and the interested reader is referred to (Ethier and Kurtz, 1986, Chapters 1 and 4).

Proposition III.11 (Chevallier (2017a, Theorem 5.6)). *Under* $(\mathcal{A}_{\mathcal{C}_{\mathrm{b}}^2,\mathrm{H\ddot{o}l},\mathrm{BS}}^{f,h,u^{\mathrm{in}}})$, *any limit fluctuation* $(\tilde{\nu},\tilde{\Xi})$ *is a solution in* $\mathcal{C}(\mathbb{R}_+,\mathcal{W}^{-2}\times\mathbb{R})$ *of the following system (formulated*[8] *in* $\mathcal{W}^{-3}\times\mathbb{R}$*),*

$$\forall\varphi\in\mathcal{W}^3, \quad \langle\tilde{\nu}_t,\varphi\rangle - \langle\tilde{\nu}_0,\varphi\rangle - \int_0^t \langle\tilde{\nu}_s, L_s\varphi\rangle\,\mathrm{d}s = \langle M_t, R\varphi\rangle$$

$$+ \int_0^t \langle u_s, \partial_\xi f(\cdot,\xi(s))R\varphi\rangle\,\tilde{\Xi}_s\mathrm{d}s, \quad (\mathrm{III.17})$$

$$\tilde{\Xi}_t = \int_0^t h(t-s)\,\mathrm{d}\langle M_s, \mathbf{1}\rangle + \int_0^t h(t-s)\,\langle u_s, \partial_\xi f(\cdot,\xi(s))\rangle\,\tilde{\Xi}_s\mathrm{d}s$$

$$+ \int_0^t h(t-s)\,\langle\tilde{\nu}_s, f(\cdot,\xi(s))\rangle\,\mathrm{d}s. \quad (\mathrm{III.18})$$

Finally, the last step is to prove uniqueness of solutions for the limit system (see (Chevallier, 2017a, Proposition 5.11)). It implies the following central limit theorem.

Theorem III.12 (Chevallier (2017a, Theorem 5.12)). *Under* $(\mathcal{A}_{\mathcal{C}_{\mathrm{b}}^2,\mathrm{H\ddot{o}l},\mathrm{BS}}^{f,h,u^{\mathrm{in}}})$, *assume that* f *is* $\mathcal{C}_{\mathrm{b}}^4$.

Then, the sequence $(\tilde{\nu}^n, \tilde{\Xi}^n)_{n\geq 1}$ *converges in distribution in* $\mathcal{D}(\mathbb{R}_+,\mathcal{W}^{-2}\times\mathbb{R})$ *to the unique solution of the system* (III.17)-(III.18).

Whereas most of the studies regarding fluctuation processes stop with the central limit theorem, I developed a theory to provide a link to the stochastic refractory density equation (SRDE). The definition of a weak solution of (SRDE) can be found in (Chevallier, 2017a, Definition 6.1). The main result is the following.

Proposition III.13. *Under* $(\mathcal{A}_{\mathcal{C}_{\mathrm{b}}^2,\mathrm{H\ddot{o}l},\mathrm{BS}}^{f,h,u^{\mathrm{in}}})$, *assume that* f *is* $\mathcal{C}_{\mathrm{b}}^4$.

The first-order approximation, defined by $\hat{u}_t^n = u_t + n^{-1/2}\tilde{\nu}_t$, *is an "almost solution of"* (SRDE) *(in the sense of Chevallier, 2017a, Proposition 6.4).*

This last result is a bit disappointing because of its lack of precision. Nevertheless there is an analog result in the next section and that one is much more precise.

ii. Spatial Hawkes process. In this section, Hawkes processes are restricted to exponential delay functions, i.e. $h(t)$ is proportional to $e^{-\alpha t}$ for some $\alpha > 0$ which represents some leakage rate. Let $f : \mathbb{R} \to \mathbb{R}_+$, $w : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $u^{\mathrm{in}} : \mathbb{R}^d \to \mathbb{R}_+$ be respectively firing rate, synaptic weight and initial condition functions. The definition of a spatial Hawkes process is given by providing its thinning representation.

Definition III.14. Let $(\Pi^i)_{i\geq 1}$ be i.i.d. $\mathcal{F}_t$-Poisson processes with intensity 1 on $\mathbb{R}_+^2$. A family of counting processes $(N^{n,i})_{i=1,\dots,n}$ on $\mathbb{R}_+$ satisfying

$$N_t^{n,i} = \int_0^t \int_0^\infty \mathbf{1}_{[0,f(U_{s-}^{n,i})]}(z)\,\Pi^i(\mathrm{d}s,\mathrm{d}z), \quad (\mathrm{III.19})$$

where the voltage variable $U_t^{n,i}$ satisfy

$$U_t^{n,i} = e^{-\alpha t}u^{\mathrm{in}}(x_i) + \frac{1}{n}\sum_{j=1}^n w(x_j,x_i)\int_0^t e^{-\alpha(t-s)}N^{n,j}(\mathrm{d}s), \quad (\mathrm{III.20})$$

---

[8]The reason of this formulation is explained in (Chevallier, 2017a, Remark 5.7)

is called a **spatial Hawkes process** (SHP) with positions $(x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$. In particular, the intensity of $N^{n,i}$ is $\lambda_t^{n,i} = f(U_{t-}^{n,i})$.

Under $(\mathcal{A}_{\mathrm{BL}}^{f,w,u^{\mathrm{in}}})$, this process is well-defined and (Chevallier et al., 2018, Propositions 2 and 3):

$$t \mapsto \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\int_0^t \lambda_s^{n,i} \mathrm{d}s\right] + \mathbb{E}\left[\left(\int_0^t \lambda_s^{n,i} \mathrm{d}s\right)^2\right] \quad \text{is locally bounded.} \tag{III.21}$$

In comparison with the previous section, notice that the particles are not exchangeable: the behavior of the point process $N^{n,i}$ depends on its spatial covariate x_i.

Figure III.7 shows a simulation of a SHP of dimension $n = 300$ in the two frameworks considered above for (NFE).
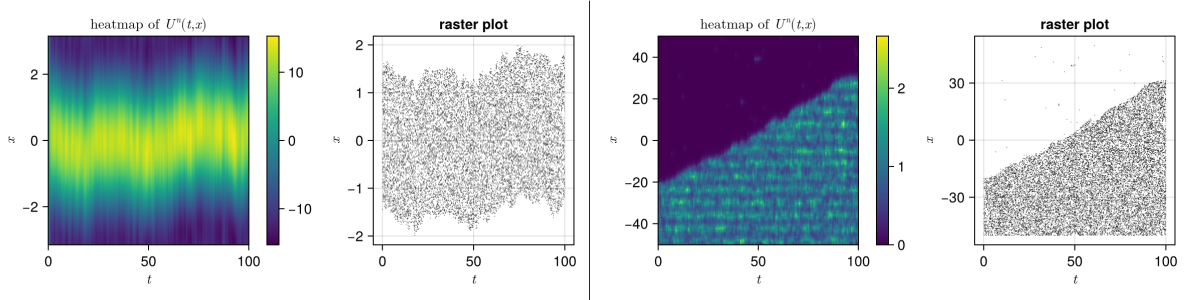
Figure III.7: Heatmap of $U^n(t, x_i) = U_t^{n,i}$ and raster plot for $n = 300$. The two panels correspond to those of Figures II.4 and II.5.

Remind that our first objective is to find the limit equation. Whereas the limit intensity is stochastic in the previous section because of the age variable A, it is purely deterministic here. However, since our particles depend on their spatial covariate, we expect a different limit process for each position $x \in \mathbb{R}^d$. All in all the limit processes are Poisson processes whose intensity depend on the time t and the position x. Let $\lambda(t, x)$ denote such intensity. Here is its heuristic derivation assuming that the empirical measure of the positions $\rho^n := n^{-1} \sum_{i=1}^n \delta_{x_i}$ converges to some limit ρ. Looking at the right hand side of Equation (III.20): 1) replace the x_i's by the generic position x; 2) replace $N^{n,j}(\mathrm{d}s)$ by its limit intensity $\lambda(s, x_j)\mathrm{d}s$; 3) replace the empirical sum over the x_j's by an integral with respect to ρ. All in all it gives

$$\lambda(t, x) = f\left(e^{-\alpha t} u^{\mathrm{in}}(x) + \int_{\mathbb{R}^d} w(y, x) \int_0^t e^{-\alpha(t-s)} \lambda(s, y) \mathrm{d}s \rho(\mathrm{d}y)\right). \tag{III.22}$$

Taking into account the bounds in Equation (III.21), any candidate limit intensity should a priori satisfy

$$\int_{\mathbb{R}^d} \left\{\int_0^t \lambda(s, x)\mathrm{d}s + \left(\int_0^t \lambda(s, x)\mathrm{d}s\right)^2\right\} \rho(dx) < \infty.$$

Remind that Theorem II.7 states that the limit intensity is uniquely determined by (III.22) as well as its connection with the neural field equation: indeed $\lambda(t, x) = f(u(t, x))$ where u is the solution of (NFE). A critical remark is that the neural field equation **is not** a Fokker-Planck equation. In particular the link between the particle system and the neural field equation is quite different from what appears in the other sections, see Equation (III.27) below.

Our second objective is to prove the consistency of the mean-field approximation by a coupling argument. Hence, we couple the process $N^{n,i}$ with the limit process $\overline{N}^i$ satisfying

$$\overline{N}_t^i = \int_0^t \int_0^\infty \mathbf{1}_{[0, f(u(t, x_i))]}(z) \Pi^i(\mathrm{d}s, \mathrm{d}z). \tag{III.23}$$

Since the spatial covariates are important, they must be included in the empirical measure considered, i.e.

$$P_t^n(\mathrm{d}\eta, \mathrm{d}x) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\left((N_s^{n,i})_{s \in [0,t]}, x_i\right)}(\mathrm{d}\eta, \mathrm{d}x). \tag{III.24}$$

They are random probability measures on $\mathcal{D}([0,t]) \times \mathbb{R}^d$. According to the heuristic derivation of the limit intensity and in comparison to P_t^n, the limit measure is expected to be

$$P_t(\mathrm{d}\eta, \mathrm{d}x) = P_t(\mathrm{d}\eta|x)\rho(\mathrm{d}x), \tag{III.25}$$

where $P_t(\mathrm{d}\eta|x)$ is the law of a Poisson process on $[0,t]$ with intensity $\lambda(\cdot, x)$ where λ satisfies (III.22). Here is the convergence rate of $P_t^n \to P_t$ as probability measures on $\mathcal{D}([0,t]) \times \mathbb{R}^d$ via a Kantorovich-Rubinstein like distance, and a link between the particle system and (NFE).

Theorem III.15 (Chevallier et al. (2018, Theorems 1, 2 and Corollary 2)). *Under $(\mathcal{A}_{\mathrm{BL}}^{f,w,u^{\mathrm{in}}})$, assume that $\int e^{\beta\|x\|}\rho(\mathrm{d}x) < \infty$ for some $\beta > 0$.*
  Then, for all $t \geq 0$,

$$\sup_{\varphi} \mathbb{E}\big[|\langle P_t^n - P_t, \varphi\rangle| \big] \lesssim_{t,f,\alpha,\beta,w,u^{\mathrm{in}}} n^{-1/2} + W_2(\rho^n, \rho). \tag{III.26}$$

where the sup *is taken over all test functions φ that are bounded and 1-Lipschitz, and*

$$\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \int_0^t |U_s^{n,i} - u(s, x_i)|\,\mathrm{d}s \right] \lesssim_{t,f,\alpha,\beta,w,u^{\mathrm{in}}} n^{-1/2} + W_2(\rho^n, \rho). \tag{III.27}$$

Sketch of Proof. As an intermediate step between P_t^n and P_t, we introduce

$$P_t^\infty(\mathrm{d}\eta, \mathrm{d}x) := P_t(\mathrm{d}\eta|x)\rho^n(\mathrm{d}x).$$

According to Equation (III.25), it corresponds to letting $n \to \infty$ for the point processes, but not for the positions. On the one side, the distance between P_t^n and P_t^∞ is controlled by the coupling $(N^{n,i}, \overline{N}^i)$. On the other side, the distance between P_t^∞ and P_t is controlled by the Wasserstein distance between ρ^n and ρ. $\qquad\square$

In words, the rate of convergence is the slowest component between: 1) the standard law of large numbers rate $n^{-1/2}$, 2) the rate of the spatial distribution convergence in terms of Wasserstein distance. Without additional assumptions on ρ, the second term is generally slower than the first one. Indeed, the optimal uniform quantization rate is given (up to a log term if $d = 2$) by $n^{-1/(d\vee 2)}$ (Chevallier, 2018, Corollary 1) for deterministic ρ^n. Another point of view is to consider that ρ^n is the empirical measure of i.i.d. positions distributed according to ρ. In that case, the rate of convergence is given (up to a log term if $d = 4$) by $n^{-1/(d\vee 4)}$ (apply Borel-Cantelli lemma to Fournier and Guillin, 2014, Theorem 2).

Up to my knowledge, the following **spatial propagation of chaos** property has not been stated in the whole literature about mean-field limits. It is an extension of the multi-class propagation of chaos introduced by Graham (2008) to continuous covariates.

Proposition III.16 (Chevallier et al. (2018, Corollary 1)). *Under $(\mathcal{A}_{\mathrm{BL}}^{f,w,u^{\mathrm{in}}})$, assume that ρ admits a $\mathcal{C}^1$ density, denoted by f_ρ and that $\int e^{\beta\|x\|}\rho(\mathrm{d}x) < \infty$ for some $\beta > 0$. Let $\phi_1, \phi_2 : \mathcal{D}([0,t]) \to [-1,1]$ and $x_1, x_2 \in \mathbb{R}^d$ such that $f_\rho(x_1) \neq 0$ and $f_\rho(x_2) \neq 0$.*
  Then, there exist "mollifiers" $\Phi_1^n, \Phi_2^n : \mathbb{R}^d \to \mathbb{R}$ such that for

$$\varphi_1^n(\eta, x) := \phi_1(\eta)\Phi_1^n(x) \quad and \quad \varphi_2^n(\eta, x) := \phi_2(\eta)\Phi_2^n(x),$$

the following convergence holds

$$\mathbb{E}\left[\langle P_t^n, \varphi_1^n \rangle \langle P_t^n, \varphi_2^n \rangle\right] \to \langle P_t(\cdot|x_1), \phi_1 \rangle \langle P_t(\cdot|x_2), \phi_2 \rangle. \tag{III.28}$$

In words, it means that the activity near position x_1 is asymptotically independent of the activity near position x_2.

Sketch of Proof. There is a tradeoff between the rate of approximation to the identity of the mollifiers and the rate of convergence of P_t^n. Hence it suffices to take mollifiers that converge slowly enough. □

Our last objective is to refine this link by studying the fluctuations. As highlighted above, the slowest component is generally the convergence of ρ^n to ρ. To avoid this case, we restrict the model to $d = 1$, ρ equals the Lebesgue measure on $[0, 1]$ and $x_i = i/n$.

Then, associate to each particle the fluctuation process $\tilde{U}_t^{n,i} = n^{1/2}(U_t^{n,i} - u(t, x_i))$, and, on the whole spatial domain $[0, 1]$, define the measure-valued fluctuation process as

$$\tilde{\nu}_t^n(\mathrm{d}x) = \frac{1}{n}\sum_{i=1}^n \tilde{U}_t^{n,i}\delta_{x_i}(\mathrm{d}x).$$

Remind that finer upper-bounds are required to prove tightness. In this framework, it is Equation (III.27) that needs refinement (Chevallier and Ost, 2020, Proposition 2):

$$\max_{i=1,\dots,n} \mathbb{E}\left[\left|U_t^{n,i} - u(t, x_i)\right|^2\right] \lesssim_{t,f,\alpha,w,u^{\mathrm{in}}} n^{-1}.$$

Its proof is less complex than the one for ADHPs.

Here, we followed a simpler approach than the one based on weighted Sobolev spaces. Let $\mathcal{S} = \mathcal{C}^\infty([0,1])$ be the Schwartz space and $\mathcal{S}'$ be its dual. In the following, $\tilde{\nu}_t^n$ is considered as an element of $\mathcal{S}'$. We are now in position to follow the trilogy of arguments: tightness, characterization, uniqueness.

Proposition III.17 (Chevallier and Ost (2020, Corollary 4)). *Under $(\mathcal{A}_{\mathcal{C}^\infty}^{f,w,u^{\mathrm{in}}})$, the sequence of the laws of $\tilde{\nu}^n$ is tight in $\mathcal{D}([0,t], \mathcal{S}')$ and a random process $(\tilde{\nu}_t)_{t\geq 0}$ distributed according to any limit law of this tight sequence takes values in $\mathcal{C}([0,t], \mathcal{S}')$.*

Sketch of Proof. Using the finer upper-bound, we can prove that: for all $\varphi \in \mathcal{S}$, the sequence of real-valued stochastic processes $(\tilde{\nu}^n(\varphi))_n$ satisfies Aldous' tightness criterion (Billingsley, 1999, Theorem 16.10). We conclude thanks to the tightness criterion developed by Mitoma (1983): tightness over all test functions in $\mathcal{S}$ is sufficient for tightness as $\mathcal{S}'$-valued processes. □

Let us derive the equation satisfied by any limit fluctuation process $(\nu_t)_{t\geq 0}$. According to Equations (III.20) and (III.22) and reminding that $\lambda(t, x) = f(u(t, x))$, we have

$$\tilde{U}_t^{n,i} = \int_0^t e^{-\alpha(t-s)}\left\{\frac{1}{n}\sum_{j=1}^n w(x_j, x_i)N^{n,j}(\mathrm{d}s) - \int_{\mathbb{R}^d} w(y, x_i)f(u(s, y))\rho(\mathrm{d}y)\mathrm{d}s\right\}.$$

The difference in curly brackets can be decomposed through the chain of approximations

$$N^{n,j}(\mathrm{d}s) \underset{(1)}{\leftrightarrow} \lambda_{s-}^{n,j}\mathrm{d}s = f(U_{s-}^{n,j})\mathrm{d}s \underset{(2)}{\leftrightarrow} f(u(s, x_j))\mathrm{d}s, \tag{III.29}$$

and

$$\frac{1}{n}\sum_{j=1}^n w(x_j, x_i)f(u(s, x_j)) \underset{(3)}{\leftrightarrow} \int_{\mathbb{R}^d} w(y, x_i)f(u(s, y))\rho(\mathrm{d}y). \tag{III.30}$$

Up to the $n^{1/2}$ factor, they lead to three contributions: (1) is related to the martingale measure associated with the point process $N^{n,i}$; (2) is approximated by $f'(u(s, x_j)) \tilde{U}_{s-}^{n,j}$ using Taylor expansion; (3) is related to Riemann sums whose convergence is faster than the rescaling so that this contribution vanishes in the limit.

Here is the last ingredient needed to derive the limit fluctuation system. It is specifically useful to study contribution (1). Let $M^{n,i}$ be the martingale associated with $N^{n,i}$, i.e. $M_t^{n,i} = N_t^{n,i} - \int_0^t f(U_{s-}^{n,i}) \mathrm{d}s$, and M^n be the $\mathcal{S}'$-valued martingale defined by

$$\langle M_t^n, \varphi \rangle = n^{-1/2} \sum_{j=1}^n \int_0^t e^{\alpha s} \langle \rho^n, w(x_j, \cdot)\varphi \rangle \, \mathrm{d}M_s^{n,j},$$

where we remind that $\langle \rho^n, w(x_j, \cdot)\varphi \rangle = n^{-1} \sum_{i=1}^n w(x_j, x_i)\varphi(x_i)$. With this notation, the contribution (1) writes as $e^{-\alpha t} \langle M_t^n, \mathbf{1} \rangle$, where $\mathbf{1}$ denotes the constant function equal to 1. For any two test functions φ_1, φ_2, the covariance of $\langle M_t^n, \varphi_1 \rangle$ and $\langle M_t^n, \varphi_2 \rangle$ is

$$\frac{1}{n} \sum_{j=1}^n \int_0^t e^{2\alpha s} \langle \rho^n, w(x_j, \cdot)\varphi_1 \rangle \langle \rho^n, w(x_j, \cdot)\varphi_2 \rangle f(U_s^{n,j}) \mathrm{d}s,$$

which leads to the following definition.

> **Definition III.18.** Let M be a continuous centred Gaussian process with values in $\mathcal{S}'$ and covariance given by
>
> $$\mathbb{E}\left[\langle M_{t_1}, \varphi_1 \rangle \langle M_{t_2}, \varphi_2 \rangle\right] = \int_0^1 \int_0^{t_1 \wedge t_2} \langle \rho, w(y, \cdot)\varphi_1 \rangle \langle \rho, w(y, \cdot)\varphi_2 \rangle f(u(s, y)) \mathrm{d}s \mathrm{d}y.$$

Here is the limit equation (with the convention that the two terms on the right hand side are ordered according to the two steps in the chain of approximations (III.29)).

Proposition III.19 (Chevallier and Ost (2020, Theorem 2)). *Under* $(\mathcal{A}_{\mathcal{C}^\infty}^{f,w,u^{\mathrm{in}}})$, *any limit fluctuation* $\tilde{\nu}$ *is a solution in* $\mathcal{C}(\mathbb{R}_+, \mathcal{S}')$ *of*

$$\forall \varphi \in \mathcal{S}, \quad \langle \tilde{\nu}_t, \varphi \rangle = e^{-\alpha t} \langle M_t, \varphi \rangle + \int_0^t e^{-\alpha(t-s)} \langle \tilde{\nu}_s, g_s \rangle \, \mathrm{d}s, \qquad \text{(III.31)}$$

where

$$g_s(y) = \langle \rho, w(y, \cdot)\varphi \rangle f'(u(s, y)).$$

Sketch of Proof. Once again, it relies on the finer upper-bound and is concluded by the continuous mapping theorem. □

Furthermore, the uniqueness of solutions for the limit system is proved in (Chevallier and Ost, 2020, Proposition 7) and it implies the following central limit theorem.

> **Theorem III.20** (Chevallier and Ost (2020, Theorem 1)). *Under* $(\mathcal{A}_{\mathcal{C}^\infty}^{f,w,u^{\mathrm{in}}})$, *the sequence* $(\tilde{\nu}^n)_{n \geq 1}$ *converges in distribution in* $\mathcal{D}(\mathbb{R}_+, \mathcal{S}')$ *to the unique solution of* (III.31).

Furthermore, there is a link with the stochastic neural field equation (SNFE). The definition of a weak solution of (SNFE) can be found in (Chevallier and Ost, 2020, Definition 3): in particular, the infinitesimal term $W(t, y)\mathrm{d}t\rho(\mathrm{d}y)$ is understand as $W(\mathrm{d}t, \mathrm{d}y)$ where $W = (W(A))_{A \in \mathcal{B}(\mathbb{R}_+ \times [0,1])}$ is a Gaussian random field with covariance $\mathbb{E}[W(A)W(B)] = |A \cap B|$. Hence (SNFE), rewrites as

$$\mathrm{d}u^n(t, x) = \left\{ -\alpha u^n(t, x) + \int_{\mathbb{R}^d} w(y, x) f(u^n(t, y)) \mathrm{d}y \right\} \mathrm{d}t + \int_{\mathbb{R}^d} w(y, x) \sqrt{\frac{f(u^n(t, y))}{n}} W(\mathrm{d}t, \mathrm{d}y).$$

Contrarily to the previous section, this nonlinear stochastic PDE is much better understood. First, it is weakly-strongly well-posed.

Proposition III.21 (Chevallier and Ost (2020, Theorems 4 and 5)). *Under* $(\mathcal{A}_{\mathcal{C}^\infty}^{f,w,u^{\mathrm{in}}})$, *assume that f is lower bounded by some positive constant.*

Then, for all $n \geq 1$, there exists a unique solution of (SNFE) *such that for all $t \geq 0$, $\sup_{s\in[0,t],x\in[0,1]} \mathbb{E}[|u^n(t,x)|^2] < \infty$. Furthermore, u^n is (β_1, β_2)-Hölder continuous for any $\beta_1 < 1/2$ in time and any $\beta_2 < 1$ in space.*

Sketch of Proof. The steps are pretty standard: existence follows a Picard iteration scheme, uniqueness is given by Grönwall's lemma, and regularity is given by a control of the covariance function (see Section III.4 for the reason why regularity of the covariance function implies regularity of paths). $\qquad\square$

Second, the limit fluctuation $\tilde{\nu}$ can be represented via a function (Chevallier and Ost, 2020, Proposition 9): there exists a continuous function, also denoted by $\tilde{\nu} : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$, by abuse of notation, such that

$$\forall \varphi \in \mathcal{S}, \quad \langle \tilde{\nu}_t, \varphi \rangle = \int_0^1 \varphi(x)\tilde{\nu}(t,x)\mathrm{d}x.$$

Hence, define by $\hat{u}^n(t,x) = u(t,x) + n^{-1/2}\tilde{\nu}(t,x)$ the first-order approximation of the particles system. It is a spatio-temporal random field which is closely related to (SNFE) as stated below (compare the n^{-1} convergence rate to the usual $n^{-1/2}$ convergence rate).

> **Theorem III.22** (Chevallier and Ost (2020, Theorem 6)). *Under* $(\mathcal{A}_{\mathcal{C}^\infty}^{f,w,u^{\mathrm{in}}})$, *assume that f is lower bounded by some positive constant.*
>
> *Then, we can couple $\hat{u}^n$ and a solution u^n of* (SNFE) *such that, for all $t \geq 0$,*
>
> $$\sup_{s\in[0,t],x\in[0,1]} \mathbb{E}\left[|\hat{u}^n(t,x) - u^n(t,x)|^2\right]^{1/2} \lesssim_{t,w,f,\alpha} \frac{1}{n}.$$

Sketch of Proof. First, represent $\hat{u}^n(t,x)$ and $u^n(t,x)$ as integrals: e.g.

$$\hat{u}^n(t,x) = e^{-\alpha t}u^{\mathrm{in}}(x) + \int_0^t e^{-\alpha(t-s)}\int_0^1 w(y,x)\sqrt{\frac{f(u(s,y))}{n}}W(\mathrm{d}s,\mathrm{d}y)$$
$$+ \int_0^t e^{-\alpha(t-s)}\int_0^1 w(y,x)\left[f(u(s,y)) + n^{-1/2}f'(u(s,y))\tilde{\nu}(s,y)\right]\mathrm{d}y\mathrm{d}s,$$

where W is the same GRF as in the definition of $u^n(t,x)$. Second, expand the square involved in the statement and apply Taylor expansion of order 2 on f at point $u(s,y)$. In turn, it gives an upper-bound involving the fourth moment of $\tilde{\nu}(t,x)$ which is controlled by (Chevallier and Ost, 2020, Proposition 9). $\qquad\square$

iii. Binary Markov chain in a random environment. In this section, a discrete time process is studied. It can be thought as the discrete time analog of the linear Hawkes process framework developed by Delattre and Fournier (2016). The main interest of the discrete time framework compared to the Hawkes one is that it enables the modeling of inhibitory neurons while remaining a linear model.

With Eva Löcherbach and Guilherme Ost, we studied this model with two statistical problems in mind (parameter estimation and community detection). The statistical procedures we came up with are detailed in Chapter IV. The proofs of their theoretical guarantees rely on large time asymptotics ($t \to \infty$) and high dimensional asymptotics ($n \to \infty$). Since the later ones are related to mean-field limits, I chose to detail them in the present chapter.

Let us first define the model and give its description later. Let $p \in [0,1]$, $\gamma \in (0,1)$ and $\mu \in [0, 1-\gamma]$ be real parameters, and $\mathcal{P}_+, \mathcal{P}_-$ be sets forming a partition of $\{1, \ldots, n\}$.

Definition III.23. Let $\Psi^n = (\Psi^n_{ij})_{1 \le i,j \le n}$ be a random matrix with i.i.d. entries distributed as $\mathrm{Ber}(p)$ for some $p \in (0,1)$. A **Markov chain in random environment** (MCRE) is a stationary process $X^n = (X^n_t)_{t \in \mathbb{Z}}$ on the state space $\{0,1\}^n$ such that, conditionally on $\Psi^n = \psi$, it follows the transition probability: for all $x, y \in \{0,1\}^n$,

$$\mathbb{P}_\psi\left(X^n_t = y | X^n_{t-1} = x\right) = \prod_{i=1}^n (q_{\psi,i}(x))^{y_i} (1 - q_{\psi,i}(x))^{1-y_i}, \qquad \text{(III.32)}$$

where

$$q_{\psi,i}(x) = \mu + \gamma \left(\frac{1}{n} \sum_{j \in \mathcal{P}_+} \psi_{ij} x_j + \frac{1}{n} \sum_{j \in \mathcal{P}_-} \psi_{ij}(1 - x_j) \right). \qquad \text{(III.33)}$$

The notation $\mathbb{P}_\psi$ is used to highlight the fact that it is a conditional probability given $\Psi^n = \psi$. Its associated expectation is denoted by $\mathbb{E}_\psi$.

In words, this process models spiking neurons in discrete time. For instance, the i-th coordinate of X^n_t, denoted by $X^{n,i}_t$, describes the state of neuron i at time t: $X^{n,i}_t = 1$ corresponds to a spike of the neuron, $X^{n,i}_t = 0$ corresponds to no activity. On the one side, the transition probability (III.32) models the fact that the spikes at time t are conditionally independent given the spikes at time $t-1$. On the other side, Equation (III.33) gives the probability of observing a spike at time t on neuron i given the spikes of the whole network at time $t-1$. Since $\psi_{ij} \in \{0,1\}$, this probability only depends on the neurons j for which $\psi_{ij} = 1$. Hence, ψ corresponds to the graph of conditional dependencies between neurons and may be thought as the graph of the interactions inside the neural network. Moreover, the neurons $j \in \mathcal{P}_+$ are excitatory since $q_{\psi,i}$ is non decreasing w.r.t. the coordinate x_j and, conversely, the neurons $j \in \mathcal{P}_-$ are inhibitory. Finally, the distribution of the random matrix Ψ^n means that the graph of interactions is an Erdős-Rényi random graph with parameter p.

The parameter μ represents some baseline activity: for instance, $q_{\psi,i}(x) = \mu$ if ψ is the null matrix. The parameter γ represents the strength of interaction in the network: for instance $\gamma = 0$ corresponds to independent neurons. Both parameters respect the condition $\mu + \gamma \le 1$ to ensure that $q_{\psi,i}$ is a probability. The subset $\mathcal{P}_+$ (resp. $\mathcal{P}_-$) is called the community of excitatory (resp. inhibitory) neurons.

Remark III.24. In the Hawkes model (III.4), the interactions are described by the delay functions $h_{j \to i}$. In (III.33), the interaction is discrete in time and localized only on the preceding time. Furthermore $h_{j \to i}$ is proportional to $\gamma n^{-1} \psi_{ij}$ if $j \in \mathcal{P}_+$ or $-\gamma n^{-1} \psi_{ij}$ if $j \in \mathcal{P}_-$. In the full model, Ψ^n is a random matrix which would correspond to a random delay function $h_{j \to i}$. Such consideration for the Hawkes model is not discussed in this manuscript but can be found in (Chevallier, 2017b, Section 2.1).

Once $\Psi^n = \psi$ is fixed, the model can be represented via a backward regeneration scheme. This representation is crucial both from a theoretical and computational point of view. In words, it works as follows. Denote $\tilde{\mu} := \mu/(1 - \gamma)$. At any given time t, neuron i first decides to update independently of anything else with probability $1 - \gamma$. If it does so, the value of $X^{n,i}_t$ is chosen according to a $\mathrm{Ber}(\tilde{\mu})$ variable. If not, it chooses uniformly one of the n neurons (including i itself): let j denote the chosen one. Then, there are three possibilities:

- if $\psi_{ij} = 0$ then $X^{n,i}_t = 0$; else
- if $j \in \mathcal{P}_+$ then $X^{n,i}_t$ copies the value of $X^{n,j}_{t-1}$;
- if $j \in \mathcal{P}_-$ then $X^{n,i}_t$ copies $1 - X^{n,j}_{t-1}$.

Such representation is called *imitation model* in the literature (see De Santis and Piccioni, 2015). To formalize this representation, consider space-time coordinates denoted by $z = (i, t)$ in $\mathcal{Z} =$

$\{1, \ldots, n\} \times \mathbb{Z}$ and let $X_z^n := X_t^{n,i}$ for all $z = (i, t) \in \mathcal{Z}$. To each space-time coordinate $z \in \mathcal{Z}$, associate a couple of independent random variables (J_z, Y_z) taking values in $\{0, 1, \ldots, N\} \times \{0, 1\}$ such that $\mathbb{P}(J_z = 0) = 1 - \gamma, \mathbb{P}(J_z = k) = \gamma n^{-1}$ for $1 \le k \le N$, and $\mathbb{P}(Y_z = 1) = 1 - \mathbb{P}(Y_z = 0) = \tilde{\mu}$. Moreover, assume that the couples $((J_z, Y_z), z \in \mathcal{Z})$ are independent. By convention, let us define $J_{(0,t)} = 0$ for all $t \in \mathbb{Z}$. Then, for any $z = (i, t) \in \mathcal{Z}$, there is a backward random walk $I^z = (I_s^z)_{s \in \mathbb{Z}}$ taking values in the state space $\{0, 1, \ldots, N, \infty\}$, where $I_s^z = \infty$ for all $s > t$, $I_t^z = i$ and, as s decreases, I^z follows the space coordinates given by the J variables, that is

$$I_{s-1}^z = J_{(I_s^z, s)}, \text{ for all } s \le t.$$

All these random walks reach the cemetery state $i = 0$ (Chevallier et al., 2024, Proposition 3.1). Hence, define the set of regenerating sites as $\mathcal{R} = \{z \in \mathcal{Z}, J_z = 0\}$. In summary, the random walk I^z equals $+\infty$ for all times $u > t$, that is, before it "starts to live". Then, the random walk lives in $\{1, \ldots, N\}$ until it reaches 0 and then remains in state 0 forever. Finally, the backward regeneration representation reads, for $z = (i, t)$, as (Chevallier et al., 2024, Equation (29)):

$$X_t^{n,i} = Y_z \mathbf{1}_{\mathcal{R}}(z) + \psi_{iJ_z} \left(X_{t-1}^{n,J_z} \mathbf{1}_{\mathcal{P}_+}(J_z) + (1 - X_{t-1}^{n,J_z}) \mathbf{1}_{\mathcal{P}_-}(J_z) \right) (1 - \mathbf{1}_{\mathcal{R}}(z)). \quad \text{(III.34)}$$

It provides a procedure to sample the desired stationary MCRE (Chevallier et al., 2024, Theorem 3.3) which I implemented in the MeanFieldGraph.jl package (see Figure III.8 below). Of course, Propp and Wilson's coupling from the past is an alternative procedure. Even if we did not try it, I believe that it is not feasible in practice because of the exponentially large state space $\{0, 1\}^n$.
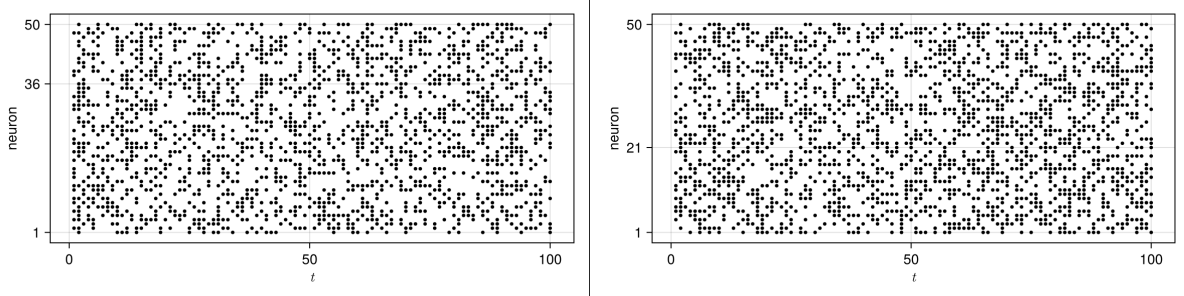
Figure III.8: Raster plots for $n = 50$ on 100 time steps. There is a dot at coordinates (t, i) if and only if $X_t^{n,i} = 1$ Left: The parameters are $\mu = 0.1$, $\gamma = 0.7$, $p = 0.8$ and $r_+ = 0.7$. Right: The parameters are $\mu = 0.3$, $\gamma = 0.3$, $p = 0.5$ and $r_+ = 0.4$. In both panels, the bottom neurons are excitatory and the top neurons are inhibitory.

Remind that our first objective is to find the limit equation. Since this is a standard mean-field framework with exchangeable particles[9] we expect that the particles behave as i.i.d. copies of some limit process $\overline{X} = (\overline{X}_t)_{t \in \mathbb{Z}}$. The temporal dependence of the Markov chain vanishes in the limit and we expect that the $\overline{X}_t$ are i.i.d. random variables with values in $\{0, 1\}$, that is $\overline{X}$ is a Bernoulli process with some parameter $m := \mathbb{P}(\overline{X}_0 = 1) \in [0, 1]$. The fact that m does not depend on time t comes from the stationarity. The only assumption on this model concerns the size of the communities and reads as follows.

$\left(\mathcal{A}_r^{\mathcal{P}} \right):$ There exist $r_+ \in (0, 1)$ and a constant κ such that, for all $n \ge 1$,

$$|\text{Card}\{\mathcal{P}_+\} - nr_+| + |\text{Card}\{\mathcal{P}_-\} - nr_-| \le \kappa,$$

where $r_- = 1 - r_+$.

---

[9]Exchangeability is satisfied if the random environment Ψ^n is taken into account, but particles are not exchangeable once $\Psi^n = \psi$ is fixed.

We are now in position to derive the limit equation from (III.33) in two steps. First, substitute $q_{\psi,i}(x)$ and the x_j's by m to get

$$m \approx \mu + \gamma \left(m\frac{1}{n}\sum_{j \in \mathcal{P}_+} \psi_{ij} + (1-m)\frac{1}{n}\sum_{j \in \mathcal{P}_-} \psi_{ij} \right).$$

Second, the limit of the two empirical sums above are respectively pr_+ and pr_- so that m should satisfy

$$m = \mu + \gamma pm(r_+ - r_-) + \gamma pr_-,$$

which solves as

$$m = \frac{\mu + \gamma pr_-}{1 - \gamma p(r_+ - r_-)}. \tag{III.35}$$

Contrarily to the two previous sub-sections, the limit equation is easy to solve and in turn the well-posedness of the limit process is trivial. Furthermore, there is no link to a PDE or even an ordinary differential equation since the particles are in the stationary regime. In particular, objectives 2 and 3 developed in the introduction of this section do not make sense here.

However, as mentioned at the beginning of this sub-section, the main interest of the present model is to answer two statistical problems. The first one is the inference of the parameter p which corresponds to the density of the dependence graph. Since the parameters μ and γ are unknown in general, a critical remark is that the inference of p is not possible by using only the mean-field approximation: p only appears through the product γp in the approximation. This issue is tackled in Section IV.2 by considering some estimators based on the observation of the Markov chain between time 0 and time t, and studying their limit as $t \to \infty$. The rest of the present section contains some asymptotic results as $n \to \infty$.

In the following, the indices $i, j \in \{1, \ldots, n\}$ are referred to as space coordinates. In that respect some spatial means (i.e. averages over space coordinates) are defined below. Denote by $m^n(\psi) := (m^{n,1}, \ldots, m^{n,n})^\top = \mathbb{E}_\psi[X_0^n]$ the stationary mean vector. Taking the expectation of (III.33) and using the stationarity of the chain yields

$$m^n(\psi) = \mu\mathbf{1}_n + \gamma \left(A^n(\psi)m^n(\psi) - L^{n,\bullet-}(\psi) \right), \tag{III.36}$$

where

- $\mathbf{1}_n$ is the vector with n coordinates equal to 1;
- $A^n(\psi) = (A_{ij}^n)_{i,j}$ is a matrix with $A_{ij}^n = n^{-1}\psi_{ij}$ if $j \in \mathcal{P}_+$ and $A_{ij}^n = -n^{-1}\psi_{ij}$ if $j \in \mathcal{P}_-$;
- $L^{n,\bullet-}(\psi)$ is a vector of partial row sums of ψ with $L^{n,\bullet-}(\psi)_i = \sum_{j \in \mathcal{P}_-} A_{ij}^n$.

As soon as $\gamma < 1$, the inverse matrix $Q^n(\psi) = (I_n - \gamma A^n(\psi))^{-1}$ is well-defined (see Chevallier et al., 2024, Appendix F) and the equation above is solved as

$$m^n(\psi) = Q^n(\psi) \left(\mu\mathbf{1}_n - \gamma L^{n,\bullet-}(\psi) \right).$$

In particular, this expression shows the importance of the inverse matrix $Q^n(\psi)$. Let us define

$$m_\infty^n(\psi) := n^{-1}\sum_{i=1}^n m^{n,i} \tag{III.37}$$

the spatial mean of $m^n(\psi)$,

$$v_\infty^n(\psi) := \|m^n(\psi) - m_\infty^n(\psi)\mathbf{1}_n\|^2 = \frac{1}{n}\sum_{i=1}^n \left(m^{n,i} - m_\infty^n(\psi) \right)^2 \tag{III.38}$$

the empirical variance of $m^n(\psi)$, and

$$w^n_\infty(\psi) := \frac{1}{n} \sum_{i=1}^n (c^n_i)^2 \left(m^{n,i} - (m^{n,i})^2 \right) \tag{III.39}$$

some quantity related to temporal variance where $c^n := \mathbf{1}_n^\top Q^n(\psi)$ is the vector of the column sums of $Q^n(\psi)$. These three quantities are the large time asymptotics of three natural statistics (see Proposition IV.3 in next chapter). Their high dimensional asymptotics read as follows.

Proposition III.25 (Chevallier et al. (2024, Proposition 4.2)). *Under $(\mathcal{A}^{\mathcal{P}}_r)$,*

$$\mathbb{E}\big[|m^n_\infty(\Psi^n) - m|^2\big] \lesssim_\gamma n^{-2}, \quad \mathbb{E}\big[|v^n_\infty(\Psi^n) - v|^2\big] \lesssim_\gamma n^{-1/2}, \quad \mathbb{E}\big[|w^n_\infty(\Psi^n) - w|^2\big] \lesssim_\gamma n^{-2},$$

where

$$\begin{cases} m = \frac{\mu + \gamma p r_-}{1 - \gamma p (r_+ - r_-)}, \\ v = \gamma^2 p(1-p)((m - r_-)^2 + r_+ r_-), \\ w = m(1-m)\frac{1 + 4\gamma^2 p^2 r_+ r_-}{(1 - \gamma p(r_+ - r_-))^2}. \end{cases}$$

Sketch of Proof. The proof relies on some rates of convergence for the vectors of the row and column sums of the random matrix Ψ^n and the inverse matrix Q^n in terms of Euclidean norm. It is pretty similar to (Delattre and Fournier, 2016, Proof of Proposition 14) but there is an additional complexity here, coming from the fact that there are two communities. $\square$

The second statistical problem we have in mind is the detection of the two communities $\mathcal{P}_+$ and $\mathcal{P}_-$. Going back to Equations (III.32) and (III.33), it is natural to expect that the (one-step) conditional covariance between $X^{n,i}_1$ and $X^{n,j}_0$, denoted by $\Sigma^{(1)}_{ij} := \mathrm{Cov}_\psi(X^{n,i}_1, X^{n,j}_0)$, is related to the quantity $n^{-1}\psi_{ij}$ if $j \in \mathcal{P}_+$ and $-n^{-1}\psi_{ij}$ if $j \in \mathcal{P}_-$. This is formalized in (Chevallier and Ost, 2024, Theorem 2.1). In turn, we expect that $\sigma^n(\psi) = (\sigma^{n,1}, \dots, \sigma^{n,n})^\top$ given by

$$\sigma^{n,j} := \sum_{i=1}^n \Sigma^{(1)}_{ij}$$

is related to p if $j \in \mathcal{P}_+$ and $-p$ if $j \in \mathcal{P}_-$. This is formalized as follows.

Proposition III.26 (Chevallier and Ost (2024, Theorem 2.2)). *Under $(\mathcal{A}^{\mathcal{P}}_r)$, there exist two constants κ_1, κ_2 depending only on m, γ, p, r_+ such that*

$$\mathbb{E}\left[\big\|\sigma^n(\Psi^n) - \kappa_1\big(\kappa_2 + \mathbf{1}_{\mathcal{P}_+} - \mathbf{1}_{\mathcal{P}_-}\big)\big\|_\infty\right] \lesssim_{\gamma, p, r_+} \sqrt{\frac{\ln n}{n}},$$

where the i-th coordinate of $\mathbf{1}_{\mathcal{P}_+}$ (respectively $\mathbf{1}_{\mathcal{P}_-}$) equals $\mathbf{1}_{\mathcal{P}_+}(i)$ (resp. $\mathbf{1}_{\mathcal{P}_-}(i)$). The constants κ_1, κ_2 are explicit in (Chevallier and Ost, 2024, Theorem 2.2).

Sketch of Proof. Let $\Sigma^{(0)}$ denote the zero-step conditional covariance matrix given by $\Sigma^{(0)}_{ij} := \mathrm{Cov}_\psi(X^{n,i}_1, X^{n,j}_0)$. From the transition probabilities (III.33) and the fixed point satisfied by $m^n(\psi)$, we have $\Sigma^{(1)} = \gamma A^n(\psi)\Sigma^{(0)}$ so that it is crucial to study the asymptotics of $\Sigma^{(0)}$. A simple computation shows that: i) $\Sigma^{(0)}_{ii} = m^{n,i}(1 - m^{n,i})$ on the diagonal, ii) $\Sigma^{(0)} = \gamma^2 A^n(\psi)\Sigma^{(0)}(A^n(\psi))^\top$ out of the diagonal. The latter equation is known as Stein matrix equation and it is known that the vectorization mapping and the Kronecker product are powerful tools to solve it. We are able to prove (Chevallier and Ost, 2024, Proposition 4.2): informally, for all $i \neq j$,

$$\Sigma^{(0)}_{ij} \approx \frac{\gamma^2 p^2 m(1-m)}{1 - p^2(r_+ - r_-)^2} \times \frac{1}{n}$$

Finally, we use the rates of convergence of the row and column sums of Ψ^n and Q^n in terms of sup norm to transfer the results from $\Sigma^{(0)}$ to $\Sigma^{(1)}$ and in turn to $\sigma^n(\Psi^n)$. $\square$

When n is large, coordinates corresponding to excitatory neurons should be close to $\kappa_1(\kappa_2+1)$ whereas coordinates corresponding to inhibitory neurons should be close to $\kappa_1(\kappa_2-1)$. Hence, if the gap $2\kappa_1$ between those two values is large enough, it should be possible to discriminate between the two communities using the values of σ^n. That is the main idea underlying the community detection procedure detailed in Section IV.2.

III.2.c)   Influence

Here are described some articles to which I did not contribute but are highly related to and/or influenced by my contributions detailed above.

i. Age dependent Hawkes process.   At the same time as (Chevallier, 2017b), the independent article by Quiñinao (2016) contains the same kind of results (objectives 1 and 2) using tightness techniques instead of coupling. The main difference is that the boundedness assumption in $(\mathcal{A}_{\mathrm{BL}}^f)$ is relaxed at the cost of a restriction to exponential delay functions h.

Schmutz (2022) considered a generalization of ADHP by adding a leaky memory process $M^{n,i}$ with values in $\mathbb{R}^d$ which satisfies the following SDE

$$\mathrm{d}M_t^{n,i} = v(M_t^{n,i})\mathrm{d}t + J(M_{t-}^{n,i})N^{n,i}(\mathrm{d}t).$$

In turn the intensity $f(A_{t-}^{n,i}, \Xi_t^n)$ is replaced by $f(A_{t-}^{n,i}, M_{t-}^{n,i}, \Xi_t^n)$. Furthermore, the delay function h also depend on the age and leaky memory of the neuron at the time it emits a spike. Both extensions are motivated by two examples: i) non vanishing self-interaction, ii) short-term synaptic plasticity. Objectives 1 and 2 are treated via the coupling method and the limit Fokker-Planck equation reads as (Schmutz, 2022, Theorem 2):

$$\begin{cases} \partial_t u(t,a,m) + \partial_a u(t,a,m) + \nabla \cdot (v(m)u(t,a,m)) + f(a,m,\xi(t))u(t,a,m) = 0, \\ u(t,0) = \iota_* \left(\int_0^\infty f(a,\cdot,\xi(t))u(t,a,\cdot)\mathrm{d}a \right), \\ \xi(t) = \int_0^t \int_{\mathbb{R}^d} \int_0^\infty h(t-s,a,m)f(a,m,\xi(s))u(s,a,m)\mathrm{d}a\mathrm{d}m\mathrm{d}s., \end{cases}$$

where $\iota_*(\cdot)$ denotes the pushforward measure by the mapping $m \mapsto m + J(m)$. The computations and proofs are quite similar to (Chevallier, 2017b) but also rely on a novel path integral representation.

Heesen and Stannat (2021) considered the standard mean-field Hawkes process framework of (Delattre et al., 2016) and followed the 3 objectives. The fluctuations are not obtained via the tightness method detailed above but rather using a structure of the Volterra integral type. The main interest is that it gives a more explicit of the fluctuation process than system (III.17)-(III.18).

In (Forien et al., 2025; Zotsa Ngoufack, 2025) which are part of A. B. Zotsa Ngoufack's PhD thesis, the authors considered a stochastic epidemic model in a mean-field framework closely related to ADHP (a spiking event here corresponds to an infection event there). It would be too complex to enter into details here, but some steps of their proofs are adapted from mine. In particular, the finer upper-bounds needed for objective 3 are also of the total variation type, remind Equation (III.14), and their proof is very similar to (Chevallier, 2017a, Proposition 3.1).

ii. Spatial Hawkes process.   The PhD thesis of Z. Agathe-Nerine is focused on an extension of the SHP detailed above. The main extension lies in the replacement of the deterministic synaptic weight $w(x_j, x_i)$ in Equation (III.20) by independent Bernoulli variables. They describe the interaction graph of the neural network which can be diluted (contrarily to most standard mean-field studies). Agathe-Nerine (2022) deals with objectives 1 and 2 in that case. Some of the proofs follow (Chevallier et al., 2018) but most of them need additional tools such as graphon

theory and spectral analysis. Agathe-Nerine (2023) deals with the case where the limit equation admits a linearly stable stationary state. In that case, the particle system stays close to this limit stationary state up to polynomial times with respect to n. Finally, Agathe-Nerine (2025) deals with the case where the limit equation admits a locally stable manifold of stationary solutions (related to bump solutions). In that case, the particle system stays close to this manifold up to polynomial times with respect to n. Furthermore, its projection onto the manifold behaves asymptotically as a Brownian motion on a time scale of order n.

The framework detailed above requires that the neurons are located according to some spatial distribution ρ with finite volume. While this is sufficient for bump solutions, the theoretical study of travelling waves solutions requires translation invariance and so ρ must be the Lebesgue measure. To overcome this problem, Luçon and Poquet (2025) consider a spatial Hawkes processes located in some compact set completed by Poisson processes as boundary conditions: as n increases, both the density of the locations and the diameter of the compact set increase. In a neutral case (the wave speed is zero), they prove that the particle system stays close to the manifold of traveling wave profiles and that its projection onto the manifold behaves asymptotically as a Brownian motion on a time scale of order $n^{1/(1+\varepsilon)}$ for some small $\varepsilon > 0$.

iii. Binary Markov chain in a random environment. Guilherme Ost supervised the Master thesis of Pinto (2025) which is closely related to the binary Markov chain detailed above. Instead of characterizing the excitatory/inhibitory behavior of neurons, the communities characterize the strength of each interaction (intra- vs inter-community strengths). In particular, the analogous to the mean firing rate m, defined in Equation (III.35), and the one-step conditional covariance $\Sigma^{(1)}$ are studied.

III.3   Diffusion approximations

My contributions regarding diffusion approximations deal with strong approximations and their connection to the modulus of continuity of the Brownian motion. The content of this section is based on the following article and preprint:

- 📄 J. Chevallier, A. Melnykova, and I. Tubikanec. Diffusion approximation of multi-class Hawkes processes: Theoretical and numerical analysis. *Advances in Applied Probability*, 2021.
- 📄 J. Chevallier. Uniform in time modulus of continuity of Brownian motion. *preprint*, 2023.

III.3.a)   Introduction

The best known diffusion approximation is Donsker's invariance principle (Donsker, 1951).

Theorem III.27. *Let $(Y_i)_{i \geq 1}$ be i.i.d. centered variables with variance one and*

$$X_t^n = n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} Y_i.$$

Then, X^n converges in distribution in $\mathcal{D}(\mathbb{R}_+)$ to the Brownian motion B.

Several extensions can be found in the literature: e.g. using the notion of infinitesimal generator (Ethier and Kurtz, 1986, Chapter 7, Theorem 4.1). Such an approximation only yields a convergence in distribution and is often called weak in comparison to strong approximation results such as the seminal work by Komlós et al. (1975). Here is stated a corollary (adapted to our setting).

> **Theorem III.28** (KMT coupling[a] (Ethier and Kurtz, 1986, Chapter 7, Corollary 5.5)). *We can couple a Poisson process Π with intensity 1 on $\mathbb{R}_+$ and a Brownian motion B in a such a way that*
>
> $$\sup_{t \geq 0} \frac{|\Pi_t - t - B_t|}{\ln(2 \vee t)} \leq K,$$
>
> *where K is a random variable such that $\mathbb{E}[\exp(\beta K)] < \infty$ for some $\beta > 0$.*
>
> ──────────
> [a]The exponential moment property of K do not appear in the statement but in the proof of (Ethier and Kurtz, 1986, Chapter 7, Corollary 5.5).

This result is called strong because it gives a path-wise convergence. It is of particular interest from a theoretical and numerical point of view (see Figure III.9). Once again, several extensions can be found in the literature: e.g. Kurtz (1978) used some regularity property of the Brownian motion, namely its modulus of continuity, to extend the strong approximation to Markov chains represented by Poisson processes.
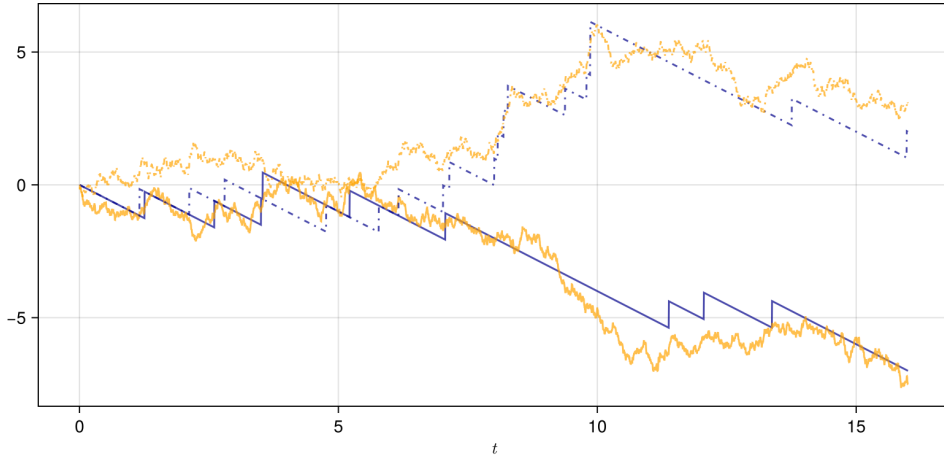
Figure III.9: Two pairs of coupled sample paths of the compensated Poisson process $t \mapsto \Pi_t - t$ (in dark blue) and B (in orange).

Let $\varphi : \mathbb{R}_+ \to \mathbb{R}$ be some function and $t > 0$ some positive time horizon. The **modulus of continuity** of φ on $[0, t]$ is defined as, for all $\eta \leq t$,

$$\omega_\varphi(t, \eta) := \sup_{s, s' \in [0, t], |s - s'| \leq \eta} |\varphi(s) - \varphi(s')|.$$

The best known control of the modulus ω_B of the Brownian motion B is Lévy continuity theorem (Lévy, 1937, page 172):

$$\omega_B(1, \eta) \sim_{\eta \to 0} \sqrt{2\eta \ln \frac{1}{\eta}} \quad \text{almost surely}.$$

Finally, let us state two results related to the next section. First, Kurtz (1978, Lemma 3.2) used the following result: for all $t \geq 0$, there exists a random variable K_t such that $\mathbb{E}[\exp(\beta K_t^2)] < \infty$ for some $\beta > 0$ and

$$\omega_B(t, \eta) \leq K_t \sqrt{\eta \left(1 + \ln \frac{t}{\eta}\right)} \quad \text{almost surely}. \tag{III.40}$$

More recently, Fischer and Nappo (2009) proved that, for all $p > 0$,

$$\mathbb{E}\left[\omega_B(t, \eta)^p\right] \lesssim_p \left[\eta \ln \left(\frac{2t}{\eta}\right)\right]^{p/2}.$$

They also proved this bound for general diffusion processes (Fischer and Nappo, 2009, Theorem 1) and in turn applied this control to the Euler numerical scheme for stochastic delay differential equations.

III.3.b)   Contributions

To study the diffusion approximation of a multi-class system of Hawkes processes, I followed the approach of Kurtz (1978). Later I turned to the issue of upper-bounding ω_B uniformly in time. It is detailed below in reverse chronological order.

i. Uniform modulus.   There exist two methods to prove (III.40): one is based on Talagrand's chaining argument, the other relies on GRR inequality (Garsia et al., 1970, Lemma 1.1). A comparison of the two methods can be found in (Talagrand, 2014, Appendix A).

I preferred the second method and used it to prove the following result.

Theorem III.29 (Chevallier (2023, Corollary 1)). *Let* $\varepsilon \in (0,1)$. *There exists a random variable* K *such that* $\mathbb{E}[\exp(\beta K^2)] < \infty$ *for some* $\beta > 0$ *and, for all* $0 < \eta < t < \infty$,

$$\omega_B(t,\eta) \le K\sqrt{\eta\left(1 + \ln\frac{t}{\eta} + \varepsilon|\ln t|\right)}.$$

Sketch of Proof. It relies on GRR inequality and is inspired from (Fischer and Nappo, 2009, Proof of Lemma 4) except that multiplicative factors are added to deal with the behavior of $\omega_B(t,\eta)$ as $t \to 0$ and $t \to \infty$. □

The main improvement with respect to (III.40) is that the constant is uniform in time at the cost of a log term. Such a result can be used to get a general stability inequality for diffusion processes (Chevallier, 2023, Proposition 3): in words, we can couple two diffusions processes so that their relative distance is controlled by the relative distances of their drift and diffusion coefficients. In turn, I applied this stability inequality to get strong approximation results in two simple frameworks (Chevallier, 2023, Corollaries 3 and 4).

ii. Multi-class Hawkes process.   The model described below is a restriction of the multi-class oscillating system introduced by Ditlevsen and Löcherbach (2017) to the case of two classes. Let $f_1, f_2 : \mathbb{R} \to \mathbb{R}_+$ be two firing rate functions and $(c_1, m_1, \alpha_1), (c_2, m_2, \alpha_2) \in \{-1, 1\} \times \{1, 2, 3, \dots\} \times (0, \infty)$ be two sets of parameters for the Erlang-type delay functions

$$h_{1\to 2}(t) = c_2 e^{-\alpha_2 t}\frac{t^{m_2}}{m_2!} \quad\text{and}\quad h_{2\to 1}(t) = c_1 e^{-\alpha_1 t}\frac{t^{m_1}}{m_1!}.$$

Remark that $m_1 = 0$ correspond to an exponential delay function. Here is the main assumption on the model.

$\left(\mathcal{A}_{\mathrm{BL}}^{f_k}\right)$: | The functions $f_1, f_2 : \mathbb{R} \to \mathbb{R}_+$ are strictly positive, bounded, non-decreasing and Lipschitz continuous.

Here, the neurons are separated in two classes $k = 1, 2$ with respective sizes n_1 and n_2. The total size is denoted by $n := n_1 + n_2$. It is crucial to note that the convention $k + 1 = 1$ for $k = 2$ is used and the dependence in n in the rest of this section is omitted in order to keep simple notation.

Definition III.30. A two-class Hawkes process (2CHP) is a multivariate point process $(N^{k,i})_{k=1,2;i=1,..,n_k}$ on $\mathbb{R}_+$ with $\mathcal{F}_t$-intensities given for all $k = 1, 2$ and $i = 1, .., n_k$ by

$$\lambda_t^{k,i} = f_k(U_{t-}^k), \tag{III.41}$$

where the voltage variable U_t^k satisfy

$$U_t^k = \frac{1}{n_{k+1}} \sum_{j=1}^{n_{k+1}} \int_0^t h_{k+1 \to k}(t-s) N^{k+1,j}(\mathrm{d}s). \tag{III.42}$$

For $k = 1, 2$, let $Z^k := n_k^{-1} \sum_{j=1}^{n_k} N^{k,j}$ and, for all $m' = 0, \ldots, m_k$,

$$U_t^{k,m'} := c_k \int_0^t e^{-\alpha_k(t-s)} \frac{(t-s)^{m_k-m'}}{(m_k - m')!} Z^{k+1}(\mathrm{d}s).$$

Notice that $U_t^{k,0} = U_t^k$ is the voltage variable of Definition III.30. Using the structure of the derivatives of Erlang functions, e.g. $h'_{k+1 \to k}(t) = -\alpha_k h_{k+1 \to k}(t) + c_k t^{m_k-1} e^{-\alpha_k t}/((m_k - 1)!)$, we get that the process $U^n = (U_t^{k,m'})_{k=1,2,m'=0,\ldots,m_k,t \geq 0}$ satisfy the following system of SDEs (Ditlevsen and Löcherbach, 2017, Equation (5.25)):

$$\begin{cases} \mathrm{d}U_t^{k,m'} = \left(-\alpha_k U_t^{k,m'} + U_t^{k,m'+1}\right) \mathrm{d}t, & 0 \leq m' < m_k \\ U^{k,m_k}(\mathrm{d}t) = -\alpha_k U_t^{k,m_k} \mathrm{d}t + c_k Z^{k+1}(\mathrm{d}t). \end{cases} \tag{III.43}$$

The process U^n is a piecewise deterministic Markov process of dimension $m_1 + m_2 + 2$ and this type of system is called a Markovian cascade (Ditlevsen and Löcherbach, 2017). The counting process Z^k has jumps of size n_k^{-1} with intensity $n_k f_k(U_{t-}^{k,0})$ so that its quadratic variation satisfies $\mathrm{d}\langle Z^k \rangle_t = n_k^{-1} f_k(U_{t-}^{k,0})$. In turn, its diffusion approximation reads

$$Z^k(\mathrm{d}t) \approx f_k(U_{t-}^{k,0})\mathrm{d}t + \sqrt{n_k^{-1} f_k(U_{t-}^{k,0})}\mathrm{d}B_t^k,$$

for some Brownian motion B^k. Notice that the independence between Z^1 and Z^2 implies independence between B^1 and B^2. Substituting $Z^k(\mathrm{d}t)$ by $f_k(U_{t-}^{k,0}) + (n_k^{-1} f_k(U_{t-}^{k,0}))^{1/2}\mathrm{d}B_t^k$ and U^n by the limit process $\overline{U}$ in the SDE system (III.43) gives (Ditlevsen and Löcherbach, 2017, Equation 5.26):

$$\begin{cases} \mathrm{d}\overline{U}_t^{k,m'} = \left(-\alpha_k \overline{U}_t^{k,m'} + \overline{U}_t^{k,m'+1}\right) \mathrm{d}t, & 0 \leq m' < m_k \\ \overline{U}^{k,m_k}(\mathrm{d}t) = -\alpha_k \overline{U}_t^{k,m_k} \mathrm{d}t + c_k \left\{ f_{k+1}(\overline{U}_t^{k+1,0})\mathrm{d}t + \sqrt{\frac{f_{k+1}(\overline{U}_t^{k+1,0})}{n_{k+1}}}\mathrm{d}B_t^{k+1} \right\}. \end{cases} \tag{III.44}$$

Notice that the informal limit of (III.44) when $n_1, n_2 \to \infty$ corresponds to the mean-field limit ODE system.

The following result do not appear verbatim in the literature. It is a refined version of (Chevallier et al., 2021, Theorem 2) which can be proved by using similar arguments as the ones in (Chevallier, 2023, Corollary 3).

Theorem III.31. *Under $(\mathcal{A}_{\mathrm{BL}}^{f_k})$, we can couple solutions of* (III.43) *with a solution of* (III.44) *such that, there exists a constant κ such that*

$$\sup_{t \geq 0} \frac{\|U_t - \overline{U}_t\|_\infty}{e^{\kappa t}} \leq K \frac{\ln n}{n},$$

where K is a random variable such that $\mathbb{E}[\exp(\beta K)] < \infty$ for some $\beta > 0$.

Sketch of Proof. Let B^1 and B^2 be the two Brownian motions involved in the SDE (III.44). By Dubbins-Schwarz theorem, there exist two Brownian motions $\hat{B}^1$ and $\hat{B}^2$ such that

$$\int_0^t \sqrt{\frac{f_{k+1}(\overline{U}_s^{k+1,0})}{n_{k+1}}} \mathrm{d}B_s^{k+1} = \frac{1}{n_{k+1}} \hat{B}_{n_k \overline{\Lambda}_t^{k+1}}^{k+1},$$

where $\overline{\Lambda}_t^k := \int_0^t f_k(\overline{U}_s^{k,0})\mathrm{d}s$ is the cumulative intensity of the limit process. Let Π^k denote the Poisson process associated with $\hat{B}^k$ according to KMT coupling - Theorem III.28. Then, the class-wise counting process Z^k and in turn the voltage variables $U^{k,m'}$ are defined via the time change representation with respect to Π^k - Theorem III.6.

Finally, considering the difference $U_t - \overline{U}$, there is a rest term of the form

$$\frac{1}{n_k} \left[\left(\Pi_{\Lambda_t^k}^k - \Lambda_t^k - \hat{B}_{\Lambda_t^k}^k \right) + \left(\hat{B}_{\Lambda_t^k}^k - \hat{B}_{\overline{\Lambda}_t^k}^k \right) \right].$$

Hence, the first contribution is controlled thanks to KMT coupling and the second one is controlled by the modulus of continuity ω_B. □

After the publication of (Chevallier et al., 2021), I came across the algorithm detailed in Mozgunov et al. (2018, page 298) and Enrico Bibbona shared its R code with me. I translated it into Julia, extended it and applied it to our framework. Thanks to this, Figure III.10 illustrates the strong diffusion approximation on the two voltage variables U^1 and U^2 in the oscillating case considered by (Ditlevsen and Löcherbach, 2017, Section 5.4). Notice that the coupled diffusion is very accurate even if the number of neurons is very small ($n = 20$). And it is not just by chance since the mean-field approximation or an independent diffusion approximation are significantly less accurate.

However, the main part of (Chevallier et al., 2021) is devoted to a numerical analysis of the limit process $\overline{U}$. Like any diffusion process, a naive numerical scheme is the Euler-Maruyama one (Blanes et al., 2024, Section 1.3). However, the limit system (III.44) can be split into two explicitly solvable subsystems: the first one contains the terms in curly brackets and the second one contains the other terms. In particular, the second subsystem is a linear ODE. Hence, splitting numerical methods such as the Lie-Trotter and the Strang methods (Blanes et al., 2024, Section 1.1) may be used. We proved the following result (we did not consider Strang method by simplicity).

Theorem III.32 (Chevallier et al. (2021, Proposition 1, Theorems 5 and 8)). *Under $(\mathcal{A}_{\mathrm{BL}}^{f_k})$, both Euler-Maruyama and Lie-Trotter schemes are mean-square convergent with order 1. Furthermore, the limit process $\overline{U}$ and the Lie-Trotter scheme are exponentially ergodic[a].*

---

[a]A process X is exponentially ergodic if it admits a unique invariant measure ν and $\mathrm{law}(X_t) \to \nu$ at exponential rate as $t \to \infty$.

Sketch of Proof. The rate of convergence relies on an order 2 Taylor expansion of the flow corresponding to the second subsystem (remind that it is a linear ODE).

Ergodicity relies on some Lyapunov condition and irreducibility (there is a path connecting any pair of points). □
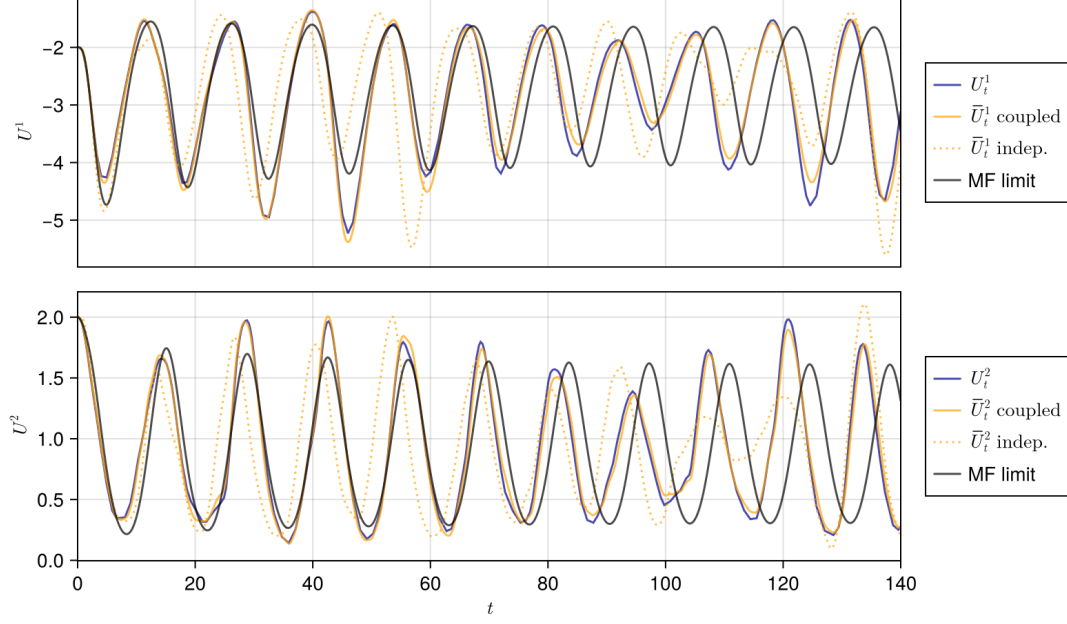
Figure III.10: The two voltage variables U^1 (top) and U^2 (bottom) as a function of time. The parameters are $f_1(x) = 10e^x \mathbf{1}_{(-\infty, \ln 20)}(x) + 400/(1 + 400e^{-2x})\mathbf{1}_{[\ln 20, +\infty)}(x)$, $f_2(x) = e^x \mathbf{1}_{(-\infty, \ln 20)}(x) + 40/(1 + 400e^{-2x})\mathbf{1}_{[\ln 20, +\infty)}(x)$, $c_1 = -1$, $c_2 = 1$, $\alpha_1 = \alpha_2 = 1$, $m_1 = 3$, $m_2 = 2$ with initial conditions $U^{1,m'} = -2$, $U^{2,m'} = 2$. In both panels, the darkblue curve corresponds to the 2CHP with $n_1 = n_2 = 10$ neurons, the solid orange curve corresponds to its coupled diffusion approximation, the dashed orange curve corresponds to an independent diffusion approximation and the black curve corresponds to the mean-field limit.

Finally, a numerical study (Cheval-lier et al., 2021, Section 5) shows that both splitting methods are better to preserve mean values than Euler-Maruyama scheme and in particular they preserve the amplitude of the oscillations. Moreover, Strang method achieves faster convergence (see Figure III.11). All in all, Strang method seems better overall.
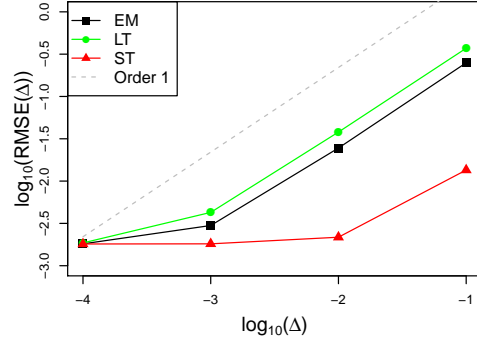
Figure III.11: Root mean square error as a function of the time step Δ averaged over 10^3 simulations. The reference solution is obtained via EM with $\Delta = 10^{-5}$. Both axes are in $\log_{10}$ scale.

III.3.c)   Influence

Bastian et al. (2025) considers a sequential change point detection problem where a stable training sample is provided. They adopt a statistical testing point of view and (Bastian et al., 2025, Theorem 2.3) states the behavior of some oracle test statistics process under the null and alternative hypotheses. In their proof of the behavior under the null, they make use of Theorem III.29 to control the regularity of that process.

III.4   Gaussian random fields

My contributions regarding Gaussian Random Fields (GRF) deal with their regularity and, if well-defined, their critical points. The content of this section is mainly based on the following preprint:

J. Chevallier, J.-F. Coeurjolly, and R. Waagepetersen. Critical point processes obtained from a Gaussian random field with a view towards statistics. *preprint*, 2025.

This preprint was initially part of Anne-Lise Porté's PhD thesis under the supervision of Jean-François and myself. Unfortunately she had to stop for medical reasons but we continued and added Rasmus as a collaborator.

III.4.a) Introduction

This section is restricted to real-valued, stationary and isotropic GRFs. Let $X = (X_t)_{t \in \mathbb{R}^d}$ denote a real-valued GRF on $\mathbb{R}^d$ ($d \geq 1$). It is assumed to be stationary and isotropic in the sense that its distribution is invariant by translation and rotation. Without loss of generality, X is assumed to be centered with unit variance (i.e. $\mathbb{E}[X_t] = 0$ and $\mathrm{Var}(X_t) = 1$). Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ denote its translation and rotation invariant covariance function. With an abuse of notation, let $c(t) = c(s, s + t)$, $t, s \in \mathbb{R}^d$, and define functions c_1 and c_2 on $\mathbb{R}_+$ by the following equalities,

$$c_2(\|t - s\|^2) = c_1(\|t - s\|) = c(t - s) = c(t, s) = \mathbb{E}[X_t X_s], \quad t, s \in \mathbb{R}^d. \tag{III.45}$$

Let F denote the spectral (probability) measure of X. By Bochner's theorem (Adler and Taylor, 2009, Theorem 5.4.1),

$$c(t) = \int_{\mathbb{R}^d} e^{\mathrm{i}\omega^\top t} F(\mathrm{d}\omega),$$

and the kth spectral moment of X is defined by

$$\lambda_k := \int_{\mathbb{R}^d} (\omega^\top e_1)^k F(\mathrm{d}\omega). \tag{III.46}$$

When k is odd, $\lambda_k = 0$ whereas when $k = 2p$ is even and $\lambda_k < \infty$, the spectral moment are related to the covariance function by (Azaïs and Delmas, 2022, Equation 2),

$$\lambda_{2p} = \mathrm{Var}\left(\partial_{t_1}^p X_0\right) = (-1)^p \frac{(2p)!}{p!} c_2^{(p)}(0),$$

provided that X and c_2 are p times differentiable at 0.

Regularity of random fields is generally stated in the mean square or the almost sure sense and has been studied extensively, especially for GRFs. It is intrinsically related to the regularity of the covariance function and the existence of spectral moments, see (Chevallier et al., 2025, Appendix A). In the following, we write "$X \in \mathcal{C}_{\mathrm{a.s.}}^\beta(\mathbb{R}^d)$" if $t \mapsto X_t(\omega)$ belongs to $\mathcal{C}^\beta(\mathbb{R}^d)$ for almost every ω. The notation is extended to the case where the superscript is β^- meaning that the condition is satisfied for all $\beta' < \beta$. Da Costa et al. (2023, Theorem 7) provide (in particular) the sufficient condition: for all integer n and $\alpha \in (0, 1/2]$, $c_1 \in \mathcal{C}^{(2n+2\beta)^-}$ implies that $X \in \mathcal{C}_{\mathrm{a.s.}}^{(n+\beta)^-}(\mathbb{R}^d)$. For instance, remind that Brownian motion sample paths are (almost) 1/2-Hölder whereas its covariance is 1-Hölder continuous, i.e. Lipschitz continuous.

Two main examples of GRFs are Matérn and random wave models. On the one side, the **Matérn** correlation function (e.g. Stein, 2012) is parameterized by a scale parameter $\sigma > 0$, a regularity parameter $\beta > 0$, and given for any $r > 0$ by

$$c_1(r) = c_2(r^2) = \frac{2^{1-\beta}}{\Gamma(\beta)} \left(\frac{r\sqrt{2\beta}}{\sigma}\right)^\beta K_\beta\left(\frac{r\sqrt{2\beta}}{\sigma}\right) \tag{III.47}$$

where Γ is the gamma function and K_β is the modified Bessel function of the second kind. The parameter β is related to the regularity of the GRF: $X \in \mathcal{C}_{\mathrm{a.s.}}^{\beta^-}(\mathbb{R}^d)$ (Da Costa et al., 2023, Proposition 10) and in particular it is at least $\lceil \beta \rceil - 1$ times continuously differentiable. In

addition, for any integer $p < \beta$, the spectral moment λ_{2p} is finite and an explicit function of p and β (see Chevallier et al., 2025, Equation 4). When $\beta = \infty$, the Matérn correlation function formally corresponds to the Gaussian correlation function $c_1(r) = \exp(-r^2/\sigma^2)$.

On the other side, the Gaussian **Random Wave Model** (RWM), see Berry (2002) in dimension $d = 2$ or Canzani and Sarnak (2019) in higher dimensions, admits a correlation function parameterized by a scale parameter $\sigma > 0$ and given by

$$c_1(r) = c_2(r^2) = \Gamma\left(\frac{d}{2}\right)\left(\frac{r\sqrt{d}}{2\sigma}\right)^{-(d/2-1)} J_{d/2-1}\left(\frac{r\sqrt{d}}{\sigma}\right) \tag{III.48}$$

where J_β is the Bessel function of the first kind. The RWM is almost surely infinitely continuously differentiable on $\mathbb{R}^d$. Its spectral moments λ_{2p} are all finite and explicit functions of p, σ and d (see Chevallier et al., 2025, Equation 6).

Provided that X is regular enough, let $X' = (X'_t)_{t\in\mathbb{R}^d}$ denote the vector valued gradient random field and by $X'' = (X''_t)_{t\in\mathbb{R}^d}$ the $d \times d$ matrix valued Hessian random field. Hence, X has a critical point at $t \in \mathbb{R}^d$ if $X'_t = 0$. This critical point is said to be with index ℓ for $\ell = 0, \ldots, d$, if $\iota(X''_t) = \ell$ where $\iota(M)$ denotes the number of negative eigenvalues of any squared matrix M. Thus, a critical point corresponds to a local minimum (resp. maximum) if $\ell = 0$ (resp. d). For any $\mathcal{L} \subseteq \{0, 1, \ldots, d\}$, let $\iota_\mathcal{L}(X''_t) := \mathbf{1}_\mathcal{L}(\iota(X''_t))$.

Definition III.33. For any $\mathcal{L} \subseteq \{0, 1, \ldots, d\}$, define

$$N_\mathcal{L} := \left\{ t \in \mathbb{R}^d, X'_t = 0 \text{ and } \iota_\mathcal{L}(X''_t) = 1 \right\}$$

to be the $\mathcal{L}$-**critical point process** associated with X.

The study of critical points of GRFs has a long history beginning with the pioneering works of Kac (1943) and Rice (1944). Since then it has generated a huge literature in probability theory (see the books Azaïs and Wschebor, 2009; Berzin et al., 2022) which exploits the celebrated Kac-Rice formula providing (essentially) expected values for number of critical points. This topic is detailed in Section IV.3. However, the simulation of critical point processes has not attracted much attention and it is the main subject of the next section.

Critical points associated to the latent Gaussian randomf field
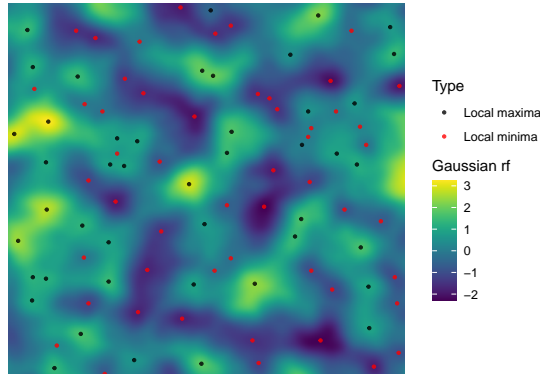
Figure III.12: Heatmap of a GRF in $[0,1]^2$ with Gaussian correlation function. The points of $N_{\{0\}}$ (in red) and $N_{\{1\}}$ (in black) are superimposed. *These points are obtained using a naive method and not the method detailed below.*

III.4.b) Contributions

i. Regularity for SPDEs. The first contribution is rather minor. It concerns the solutions of (SRDE) and (SNFE). For both I proved some regularity properties. Regarding (SRDE) it appears in (Chevallier, 2017a, Lemma A.1). Regarding (SNFE), it appears in Proposition III.21 above.

ii. Critical points. Even if they are well-defined from a theoretical point of view, critical points of a GRF may be challenging from a numerical point of view. In general, it is impossible to

derive the full distribution of $N_{\mathcal{L}}$ and its simulation must follow two steps: (i) simulate a smooth approximate sample path of the GRF X; (ii) find the critical points of this sample path. Step (ii) is essentially an optimization step. Several approaches could be considered, e.g. methods based on arithmetic intervals (Hansen and Walster, 2003) and the package IntervalArithmetic.jl, but they are not discussed in the present manuscript. Nevertheless, the crucial point is that they must be applied to a smooth field and not a discretized version.

Hence, there are two categories of simulation methods to perform Step (i): (a) simulate X on a lattice and smoothly interpolate it on the continuous space; (b) approximate a sample path of X as the "average" of smooth sample paths. Informally, each method yields a sequence of random fields $(X^n)_n$ where n represents the number of discretization points for methods (a), and the number of sample paths involved in the average for methods (b). Their associated $\mathcal{L}$-critical point processes are denoted by $N_{\mathcal{L}}^n$. Even if $X^n \to X$ as expected, the question of the convergence of $N_{\mathcal{L}}^n \to N_{\mathcal{L}}$ is not trivial. In that direction, we proved the following.

Theorem III.34 (Chevallier et al. (2025, Theorem 5)). *Let X be a GRF in $\mathcal{C}^2_{\text{a.s.}}([0,1]^d)$ such that X is a.s. a Morse function with no critical point on the boundary of $[0,1]^d$. Let $(X^n)_n$ be a sequence of random fields in $\mathcal{C}^2_{\text{a.s.}}([0,1]^d)$ such that $X^n \xrightarrow{\bullet} X$ in $\mathcal{C}^2([0,1]^d)$.*

Then, for all $\mathcal{L} \subseteq \{0,1,\ldots,d\}$,

$$N_{\mathcal{L}}^n \xrightarrow{\bullet} N_{\mathcal{L}},$$

where the convergence mode $\bullet =$ a.s, p or law.

Sketch of Proof. The proof relies on the fact that "generic" roots of a $\mathcal{C}^1$ function are moving continuously with respect to small $\mathcal{C}^1$ perturbations. It is concluded by the continuous mapping theorem. $\square$

This result is applied to one example of each methods (a) and (b) which in turn gives some conditions under which the simulation of critical point processes is consistent (Chevallier et al., 2025, Theorem 6): namely $X \in \mathcal{C}^{2+\varepsilon}_{\text{a.s.}}$ for the example of method (a) and $X \in \mathcal{C}^{3+\lfloor d/2 \rfloor}_{\text{a.s.}}$ for the example of method (b). Notice that the proof for method (b) involves the central limit theorem in a suitable functional space. Based on my experience with Sobolev spaces, described in Subsection III.2.b), we chose to work with this type of spaces. The higher required regularity $3 + \lfloor d/2 \rfloor$ comes from the application of Sobolev embedding theorem.

III.5 On going research and perspectives

Strong diffusion approximation for unbounded intensities. As noted in Section III.3, strong diffusion approximation results for a general class of models goes back to the coupling developed by Kurtz (1978). However, it relies on the main assumption that the intensity is bounded and, up to my knowledge, this assumption has not been weakened since then. The two main ingredients of its proof are KMT coupling - Theorem III.28 - and the control of the modulus of continuity of B - Equation (III.40). Even if the first one is uniform in time thus can handle unbounded intensities, the second one is local and cannot handle unbounded intensities.

Thanks to my improvement (Theorem III.29) I am adapting the proofs in (Kurtz, 1978) to deal with unbounded intensities. Most of the computations are similar to those involved in the two simple frameworks studied in (Chevallier, 2023) but the main bottleneck comes from considering an infinite state space as in (Kurtz, 1978).

Such strong diffusion approximation applied to mean-field interacting Hawkes processes as the ones studied by Heesen and Stannat (2021) gives a way to couple the particle system with the fluctuations. In turn, I believe that it should answer their conjecture which I rewrite as:

an approximation by Cox processes driven by the fluctuation intensity should be superior to an approximation via a Poisson processes driven by the limit intensity.

Recently, Prodhomme (2023) considered metastable dynamics on large time scales by refining the coupling of Kurtz (1978). The extension of the ongoing work detailed above to this refined coupling is a natural perspective and could be a subject for a postdoc with a solid background on stochastic processes.

Strong diffusion approximation in higher dimensions. As noted in the introduction of the present chapter, diffusion approximations can be used to derive mesoscopic models from microscopic ones. In turn, strong approximation results provide a coupling between the models of both scales. Up to my knowledge, such an approximation has only been studied when the macroscopic dynamics is given by an ordinary differential equation: e.g. for the two-class Hawkes process, it corresponds to the formal limit of (III.44) when $n_{k+1} \to \infty$. The case of a macroscopic PDE, like those of Chapter II, remains open. Such a strong diffusion approximation would provide a coupling between a particle system like ADHP or SHP and a stochastic partial differential equation like (SRDE) or (SNFE).

As highlighted above, the two main ingredients in dimension one are KMT coupling - Theorem III.28 - and the control of the modulus of continuity of B - Equation (III.40). Here are some ideas to extend them in higher dimensions. First, the original KMT coupling has been extended to higher dimension (Rio, 1994), but it is not obvious whether the point processes of our interest are a particular case of this abstract result. If not, it would suffice to go back the construction of KMT coupling and adapt it to our case. By the way, this step is more or less necessary in order to make numerical simulations. Second, the Brownian motion B is replaced by the Brownian sheet and Equation (III.40) has been extended to that case by Hu and Le (2013) thanks to a multiparameter GRR inequality.

A preliminary step would be to apply these results to a simple framework inspired from (Kurtz, 1978). In turn, the framework of particle systems with an SPDE mesoscopic scale would be tackled. An optional technical result would be a uniform control of the modulus of the Brownian sheet similar to Theorem III.29 by adapting the proofs in (Hu and Le, 2013).

This project could be the subject of a PhD Thesis for a student with a solid background on stochastic processes and random fields.

Numerics. Theorem III.34 and the subsequent corollary provide the theoretical foundation for algorithms to simulate critical point processes $N_{\mathcal{L}}$. Up to now, we did not implement such algorithm but Jean-François and myself are planning to do it in Julia. In particular, it would be interesting to benchmark the two simulation methods we introduced and check whether the smooth interpolation method is better than the averaging one, as predicted by the theoretical convergence result. Notice that it is not trivial to check which simulation is better in general. In our case, we know the intensity and pair correlation function of $N_{\mathcal{L}}$, see Section IV.3. They could provide a convenient way to compare the accuracy of both simulation methods.

For almost a year now, I have been involved in the development of PointProcesses.jl. It is a Julia package for temporal point process modeling, simulation and inference. It is in an early stage version: only Poisson processes are implemented for now; Hawkes processes are in the pipeline; other common models, goodness-of-fit tests and integration with the Bayesian package Turing.jl are planned. Moreover, I have some ideas and desires for future features related to the content of the present manuscript.

- Benchmark some simulation algorithms of linear Hawkes processes, especially (Møller and Rasmussen, 2005; Dassios and Zhao, 2013; Magris, 2019). The last one is related to an

explicit time change representation - Theorem III.6 - via Lambert W function. It is limited to exponential delay functions but I wonder if it could be extended to Erlang delay functions.

- Implement KMT coupling in both directions: 1) given a Brownian motion construct a coupled Poisson process (Mozgunov et al., 2018), 2) given a Poisson process construct a coupled Brownian motion. Such feature may be interesting for DiffEqNoiseProcess.jl.

- When simulating a point process via its thinning representation it is usual to simulate the next spike as a function of some bound λ_∞ of the intensity from current time $t = t_0$ to $t = \infty$. When this uniform bound is crude it implies lots of reject sampling. A less naive way is to replace the uniform bound λ_∞ by a couple $(\lambda_{\mathrm{loc}}, \tau)$ where $\tau > t_0$ is some time and λ_{loc} is a bound of the intensity on $[t_0, \tau]$. In general, λ_{loc} may be less than λ_∞ and moreover λ_{loc} is non decreasing with respect to τ. Hence there is a tradeoff between: larger τ meaning larger time coverage but more rejection, and smaller τ meaning less rejection but smaller time coverage. I would like to investigate this tradeoff in practice.

- A common way to simulate high dimensional point processes with sparse connections is to simulate the next spike of each dimension and put them into a priority queue. In that case, there is no simple way to modify these values at each new spike because of the chosen data. I would like to use binary heaps instead, as it has been done for Integrate-and-Fire neural networks by Engelken (2023). In that direction, it would be interesting to build an interface with packages dealing with spiking neural networks such as Neuroblox.jl and SpikingNeuralNetworks.jl.

- With some colleagues in my lab, I led a working group on conformal prediction which is a quite novel theory that provides distribution-free prediction regions with finite-sample coverage guarantees. Hence, I would be interested in implementing and developing some conformal methods for point processes like in (Dheur et al., 2024).

Of course, this is a long-term project in which master and/or PhD students can fit.

STATISTICS

MY contributions to the field of Statistics mainly concern the estimation of some latent graph structure and a class of spatial point processes made of the critical points of a Gaussian random field.

The latent graphs I considered are **conditional dependence graphs** underlying some multivariate stochastic processes. Each node of the graph represents a dimension of the process which itself describes the spiking behavior of neurons. Hence, the absence of an edge from node i to node j in this kind of graph means that spiking of neuron j is independent of neuron i conditionally on all the other neurons. When the dimension is low, it is possible to detect the presence of an edge between any two neurons i and j, or even estimate its strength, from the observation of the spike trains. When the dimension is high, it is not feasible in general. However, under some mean-field assumption, we can estimate some macroscopic quantities such as the **graph density**.

Critical points of Gaussian random fields have been extensively studied from a probabilistic perspective but, up to my knowledge, have not been studied from a statistical perspective or even as a **spatial point process** per se. For instance, asymptotic results mainly focus on specific statistics: the Euler characteristic (Estrade and León, 2016) or the number of critical points (Azaïs et al., 2024). As a first step towards statistical applications, such results must be extended to **more general statistics**.

Section IV.1 provides some generalities on statistics. It is intended to provide a foundation for understanding the subsequent sections dealing respectively with latent graph estimation and spatial point processes.

IV.1   Generalities

Since probability theory and statistics are closely related, this chapter borrows most of the tools and notation introduced in Chapter III. In comparison with probability theory, the main difference in (frequentist and non parametric) statistics is that the distribution of the observed variable X depends on some **parameter** θ in some **parameter space** Θ. Such a distribution is thus denoted by ν_θ. The objective is not to describe the behavior of X given that ν is fixed (like in probability theory) but rather to describe ν_θ or θ given some sample of X.

For instance, a **statistic/point estimator** $\hat{\theta}^n$ of θ is a function of a sample $(X_1, \ldots, X_n)$. Usually, asymptotic properties of $\hat{\theta}^n$ are given as the sample size n goes to infinity. Unlike the standard framework, such asymptotic is governed by:

- the observation duration t in a time process framework;
- the observation window volume $|W|$ in a spatial process framework.

Common asymptotic properties of $\hat{\theta}^n$ are:

- its consistence (law of large numbers), when $\hat{\theta}^n \to \theta$ in some sense and possibly with some convergence rate;

- its asymptotic normality (central limit theorem), when $n^p(\hat{\theta}^n - \theta)$ converges to some Gaussian distribution for some scaling index $p > 0$.

Since this chapter deals also with hypothesis testing and clustering, non-specialist (in statistics) readers are advised to read Appendix A.3 for a short introduction.

IV.2   Latent graph estimation

My contributions regarding latent graph estimation deal with detection of dependence between two point processes and inference of the latent graph for a high dimensional binary Markov chain in random environment. The content of this section is based on the following article and two preprints:

📄 J. Chevallier and T. Laloë. Detection of dependence patterns with delay. *Biometrical Journal*, 2015.

📄 J. Chevallier, E. Löcherbach, and G. Ost. Inferring the dependence graph density of binary graphical models in high dimension. *preprint*, 2024.

📄 J. Chevallier and G. Ost. Community detection for binary graphical models in high dimension. *preprint*, 2024.

The first article is part of my PhD Thesis so I chose not to detail it as much as the rest. The two groups of publications respectively correspond to the two sub-sections of Section IV.2.b).

IV.2.a)   Introduction

The latent graphs I considered are (conditional) dependence graphs associated with multivariate point processes $(N^i)_{i=1,\dots,n}$ as introduced by Didelez (2008). The nodes of the graph are the integers between 1 and n and there is an edge from j to i if the $\mathcal{F}_t^{\text{all}}$-intensity of N^i is also $\mathcal{F}_t^{(j)}$-predictable, where

$$\mathcal{F}_t^{\text{all}} := \sigma(N_s^k, k = 1, \dots, n, s \leq t) \text{ and } \mathcal{F}_t^{(j)} := \sigma(N_s^k, k \neq j, s \leq t).$$

In words, it means that the intensity of N^i do not directly depend on the spikes of N^j. In general, there is a simple characterization for multivariate Hawkes processes: there is an edge from j to i if and only if $h_{j \to i}$ is not the null function. From a neuroscience point of view, we expect that such a dependence graph can describe neuronal ensembles as connected components of the graph. Remark that the dependence graph is latent since its edges are not observed but can only be inferred from spike trains.

Historically, such studies started with the Poisson process model. For instance, it is the keystone of the popular Unitary Events (UE) method developed by S. Grün to detect synchrony patterns from a discrete time viewpoint (Grün, 1996; Grün et al., 2002) and was later adapted to the continuous time viewpoint by Tuleau-Malot et al. (2014).

The main flaw of multidimensional Poisson processes is their inability to simply describe interaction between neurons. That is the reason why most recent studies focus on Hawkes processes (or the related Generalized Linear Model) by estimating the delay functions $h_{j \to i}$. Either a discrete time viewpoint (Duarte et al., 2019) or a continuous one has been considered: using frequentist (Pillow et al., 2008; Gava et al., 2024; Reynaud-Bouret et al., 2013; Lambert et al., 2018) or Bayesian (Donnet et al., 2020; Sulem et al., 2024) methods. When the number of neurons n is high, two options exist: i) assume some sparsity or low-rank to allow exact reconstruction of the graph (Bacry et al., 2020; Wang et al., 2025); ii) assume mean-field interactions to recover the graph density (Delattre and Fournier, 2016).

IV.2.b) Contributions

i. Independence test

Let $(N^i)_{i=1,\ldots,n}$ be a multivariate point process on $[0,t]$ such that for all i, N^i is an homogeneous Poisson process. In (Chevallier and Laloë, 2015), we extended the Gaussian approximation of the Unitary Events method developed by Tuleau-Malot et al. (2014). This method is based on the so-called **coincidence count**[1] between neurons i and j, that is the number of occurrences of the pattern "a spike of i and a spike of j occur in a small time window". Such a pattern should occur frequently (resp. rarely) if i and j have synchronous (resp. asynchronous) spiking activity. In turn, we expect that the coincidence count is a good test statistic to detect synchronization.

We extended the notion of coincidence count to patterns involving more than two neurons, proved its asymptotic normality when $t \to \infty$[2] (Chevallier and Laloë, 2015, Theorem 3.2) which leads to a test with asymptotically exact type I error. Finally, we applied the Benjamini and Hochberg (1995) multiple testing procedure to answer the question of synchrony patterns detection.

ii. Binary Markov chain in a random environment

This section borrows the MCRE model introduced in Paragraph III.2.b).iii. Remind that, once the random matrix is fixed $\Psi^n = \psi$, the transition probabilities are given by Equation (III.33):

$$q_{\psi,i}(x) = \mu + \gamma \left(\frac{1}{n} \sum_{j \in \mathcal{P}_+} \psi_{ij} x_j + \frac{1}{n} \sum_{j \in \mathcal{P}_-} \psi_{ij}(1 - x_j) \right).$$

The parameters of the model are μ and γ plus the parameter p related to the random matrix Ψ^n, so that $\theta := (\mu, \gamma, p)$. The goal of this section is to address the two following statistical problems from the mere observation of a sample $X_1^n, \ldots, X_t^n$ of the Markov chain.

1. **Parameter estimation.** Can we estimate θ, knowing only the asymptotic fraction of excitatory and inhibitory components r_+ and r_-?
2. **Community detection.** Can we find the communities $\mathcal{P}_+$ and $\mathcal{P}_-$?

In particular we do not assume any prior knowledge on $\theta, \psi, \mathcal{P}_+, \mathcal{P}_-$ in any of the two cases.

In what follows, let $N_t^{n,i} := \sum_{s=1}^{t} X_s^{n,i}$ denote the number of signals emitted by the i-th component of the system in the discrete interval $\{1, \ldots, t\}$, and $\overline{N}_t^n := n^{-1} \sum_{i=1}^n N_t^{n,i}$ denote its spatial average. Set the convention $\overline{N}_0^n := 0$.

To answer the first statistical problem, three natural statistics are defined as:

$$\hat{m}_t^n := \frac{\overline{N}_t^n}{t}, \quad \hat{v}_t^n := \frac{(t+1)n}{t^3} \left[\frac{1}{n} \sum_{i=1}^n \left(N_t^{n,i} \right)^2 - \frac{t}{(t+1)} \left(\overline{N}_t^n + \left(\overline{N}_t^n \right)^2 \right) \right] \text{ and} \qquad \text{(IV.1)}$$

$$\hat{w}_t^n := 2W_{2\Delta} - W_\Delta, \quad \text{with } W_\Delta = \frac{n}{t} \sum_{k=1}^{\lfloor t/\Delta \rfloor} \left(\overline{N}_{k\Delta}^n - \overline{N}_{(k-1)\Delta}^n - \Delta \hat{m}_t^n \right)^2, \qquad \text{(IV.2)}$$

where $\Delta \in \{1, \ldots, \lfloor T/2 \rfloor\}$ is a tuning parameter. The estimator $\hat{m}_t^n$ is the **spatio-temporal mean** of the system, whereas $\hat{v}_t^n$ and $\hat{w}_t^n$ are called **spatial variance** and **temporal variance** respectively (see Chevallier et al., 2024, Subsections 2.1, 2.2 and 2.3 for 5 pages long heuristics and precisions about this denomination).

---

[1] It is called **event synchronization** in the physics literature (Quiroga et al., 2002).
[2] In fact, the time window was cut into M i.i.d. trials and the asymptotic is expressed as $M \to \infty$.

The backward regeneration representation formalized in (III.34) induces a natural (random) partitioning of the space-time state space $\mathcal{Z}$ via the equivalence relation of coalescence defined by

$$z_1 \longleftrightarrow z_2 \text{ if and only if } \exists s \in \mathbb{Z}, I_s^{z_1} = I_s^{z_2} \notin \{0, \infty\}.$$

The probability that such a coalescence appear between two or more sites z can be upper-bounded as follows.

Proposition IV.1 (Chevallier et al. (2024, Proposition 3.5)). *Let $z_k = (i_k, t_k) \in \mathcal{Z}$, for $k = 1, \ldots, 4$, be four different sites. We have*

(i) $\mathbb{P}_\psi\left(\{z_1 \longleftrightarrow z_2\}\right) \leq \left[\gamma^{|t_1 - t_2| \vee 1} / (1 - \gamma^2)\right] n^{-1}$;

(ii) $\mathbb{P}_\psi(\{z_1 \longleftrightarrow z_2 \longleftrightarrow z_3\}) \lesssim_\gamma \gamma^{(\bar{t} - \underline{t})} n^{-2}$, *where* $\bar{t} = t_1 \vee t_2 \vee t_3$ *and* $\underline{t} = t_1 \wedge t_2 \wedge t_3$;

(iii) $\mathbb{P}_\psi(\{z_1 \longleftrightarrow z_2\} \cap \{z_3 \longleftrightarrow z_4\}) \lesssim_\gamma \gamma^{|t_1 - t_2| + |t_3 - t_4|} n^{-2} + \gamma^{(\bar{t} - \underline{t})} n^{-3}$,
  where $\bar{t} = t_1 \vee t_2 \vee t_3 \vee t_4$ *and* $\underline{t} = t_1 \wedge t_2 \wedge t_3 \wedge t_4$;

and a fourth upper-bound which is not stated here by simplicity.

Sketch of Proof. From the backward random walks $(I^z)_{z \in \mathcal{Z}}$ we defined some auxiliary inhomogeneous backward Markov chain. It turns out that the coalescence probabilities can be written as probabilities of specific paths for this chain. Yet, the transition probabilities of this inhomogeneous chain are easily upper-bounded. In turn it gives upper-bounds for the probabilities of paths and so the coalescence probabilities. $\square$

Remark IV.2. Here are two simple remarks to intuitively understand the upper bounds of Proposition IV.1.

1. The exponent of the factor γ is the minimal waiting time before coalescence. For instance, in Item (ii), the first-born site (with birth time $\bar{t}$, remind that time runs backwards) must wait (hence survive) at least $\bar{t} - \underline{t}$ time steps before coalescing with the last-born site.
2. Each coalescence event adds a factor n^{-1}. For instance, in Items (ii) and (iii), at least two coalescences are needed.

In turn, it provides some upper-bounds for covariances of the form

$$\text{Cov}_\psi\left(\tilde{X}_{z_1}^n \tilde{X}_{z_2}^n, \tilde{X}_{z_3}^n \tilde{X}_{z_4}^n\right),$$

where $\tilde{X}_z^n := X_z^n - \mathbb{E}[X_z^n]$ is the centered version of X_z^n. We have identified seven different scenarios depending on which of the z_k's are equal. All of them can be found in (Chevallier et al., 2024, Lemma 3.8). These upper-bounds are crucial to prove the following large time asymptotics of the three natural statistics.

Proposition IV.3 (Chevallier et al. (2024, Proposition 4.1)). *We have*

$$\mathbb{E}_\psi\left[|\hat{m}_t^n - m_\infty^n(\psi)|^2\right] \lesssim_\gamma (nt)^{-1},$$

$$\mathbb{E}_\psi\left[|\hat{v}_t^n - v_\infty^n(\psi)|\right] \lesssim_\gamma \left(1 + \sqrt{v_\infty^n(\psi)}\right)\left(\frac{n}{t^2} + \frac{1}{\sqrt{t}} + \frac{\sqrt{n}}{t}\right),$$

$$\mathbb{E}_\psi\left[\left|\hat{w}_t^n - w_\infty^n(\psi)\right|\right] \lesssim_\gamma \left(\frac{\Delta}{t} + \frac{1}{n} + \frac{\gamma^\Delta}{\Delta}\right),$$

where $m_\infty^n(\psi)$, $v_\infty^n(\psi)$ and $w_\infty^n(\psi)$ are defined by (III.37)-(III.39).

Combining the large time asymptotics above with the high dimensional asymptotics, i.e. Proposition III.25, gives the convergence $(\hat{m}_t^n, \hat{v}_t^n, \hat{w}_t^n) \to (m, v, w)$ where we remind that

$$
\begin{cases}
m = \frac{\mu + \gamma p r_-}{1 - \gamma p (r_+ - r_-)}, \\
v = \gamma^2 p (1 - p)((m - r_-)^2 + r_+ r_-), \\
w = m(1 - m)\frac{1 + 4\gamma^2 p^2 r_+ r_-}{(1 - \gamma p(r_+ - r_-))^2}.
\end{cases}
$$

The rate of this convergence can be found in (Chevallier et al., 2024, Theorem 2.1).

Yet, the main objective is to estimate the parameter $\theta = (\mu, \gamma, p)$. Here is how an estimator $\hat{\theta}^n$ of θ can be deduced from the three statistics $(\hat{m}_t^n, \hat{v}_t^n, \hat{w}_t^n)$. Let $\Theta = \{(\mu, \gamma, p) \in (0, 1)^3 : 0 < \mu < 1 - \gamma\}$ be the parameter space, that is the set of admissible parameters. For all $\theta \in \Theta$, let us define

$$
D(\gamma, p) = 1 - \gamma p (r_+ - r_-) > 0,
$$

the denominator that appears in the expressions of m and w. Then, for $k = 1, 2, 3$, let $\Phi_k : \Theta \to \mathbb{R}$ be defined by

$$
\begin{cases}
\Phi_1(\theta) = (\mu + \gamma p r_-)/D(\gamma, p) \\
\Phi_2(\theta) = \gamma^2 p (1 - p)[(\Phi_1(\theta) - r_-)^2 + r_+ r_-] \\
\Phi_3(\theta) = \Phi_1(\theta)[1 - \Phi_1(\theta)][1 + 4\gamma^2 p^2 r_+ r_-]/D(\gamma, p)^2.
\end{cases}
$$

Finally, let $\Phi : \Theta \to \mathbb{R}^3$ be defined by the three coordinate functions above so that $(m, v, w) = \Phi(\theta)$. Whatever the value of r_+, the image set $\Phi(\Theta)$ is included in $(0, 1) \times (0, \infty)^2$ (Chevallier et al., 2024, Proposition G.3) and let us define the auxiliary function $F : (0, 1) \times (0, \infty) \to (0, \infty)$ by

$$
F(m, w) = (r_+ - r_-)^2 \frac{w}{m(1 - m)}.
$$

The function Φ can be inverted as follows.

Proposition IV.4 (Chevallier et al. (2024, Proposition 2.2)[3]). *Let $r_+ \in (0, 1)$. There exists an explicit function Φ^{-1} such that for all $\theta \in \Theta$,*

$$
\theta = \Phi^{-1} \circ \Phi(\theta)
$$

as soon as $r_+ \geq 1/2$ or $F(\Phi_1(\theta), \Phi_3(\theta)) \geq 4 r_+ r_-$.

In particular, Φ^{-1} depends on r_+ and its expression can be found in (Chevallier et al., 2024, Appendix G).

Proof. To invert the function Φ, a quadratic equation involving the quantity $F(m, w)$ must be solved. In general, we do not know which one of its two roots is the correct one. However, under the conditions stated above, only one of the two roots is admissible and in turn Φ can be inverted. $\square$

In turn, our estimator is defined as $\hat{\theta}_t^n := (\hat{\mu}_t^n, \hat{\gamma}_t^n, \hat{p}_t^n) := \Phi^{-1}(\hat{m}_t^n, \hat{v}_t^n, \hat{w}_t^n)$. Since Φ^{-1} is $\mathcal{C}^1$, its rate of convergence is obtained as a corollary of the rate of the convergence $(\hat{m}_t^n, \hat{v}_t^n, \hat{w}_t^n) \to (m, v, w)$.

Theorem IV.5 (Chevallier et al. (2024, Corollary 2.4)). *Under $(\mathcal{A}_r^{\mathcal{P}})$, assume that the condition of Proposition IV.4 is satisfied.*

---

[3]This is a simplified version of the original statement where some informations are also given when the condition is not satisfied.

Then, for all $\varepsilon \in (0, 1)$,

$$P_\theta \left(\|\hat{\theta}_t^n - \theta\|_\infty \geq \varepsilon \right) \lesssim_\gamma \left(\frac{\sqrt{n}}{t} + \frac{1}{\sqrt{n}} + \frac{\gamma^\Delta}{\Delta} + \sqrt{\frac{\Delta}{t}} \right) \varepsilon^{-1},$$

where P_θ the probability measure under which Ψ^n is distributed as i.i.d. Ber(p) variables, and the distribution of X satisfies $P_\theta(X \in \cdot | \Psi^n = \psi) = \mathbb{P}_\psi(X \in \cdot)$.

The value of the integer tuning parameter Δ must be chosen in practice. An optimization of the upper-bound above gives an optimal choice of the order $\ln(t)$. However, we noticed on numerical simulations that the temporal convergence seems better for lower values of Δ (see Figure IV.1). Overall, the temporal convergences of $\hat{m}_t^n$ and $\hat{v}_t^n$ are pretty fast as expected. Unfortunately, the temporal convergence of $\hat{w}_t^n$ is quite slow and in turn the temporal convergences of $\hat{\mu}_t^n$ and $\hat{p}_t^n$ are quite slow. Concerning the high dimensional convergences, $m_\infty^n(\psi)$ and $w_\infty^n(\psi)$ are the fastest as expected according to Proposition III.25 whereas $v_\infty^n(\psi)$ is slower. Furthermore, it is natural to define the theoretical temporal limit of $\hat{\theta}_t^n$ as

$$\theta_\infty^n(\psi) := (\mu_\infty^n(\psi), \gamma_\infty^n(\psi), p_\infty^n(\psi)) := \Phi^{-1}(m_\infty^n(\psi), v_\infty^n(\psi), w_\infty^n(\psi)).$$

Notice that the convergence of $\mu_\infty^n(\psi)$ seems to be as fast as $v_\infty^n(\psi)$ whereas $\gamma_\infty^n(\psi)$ and $p_\infty^n(\psi)$ are slower.
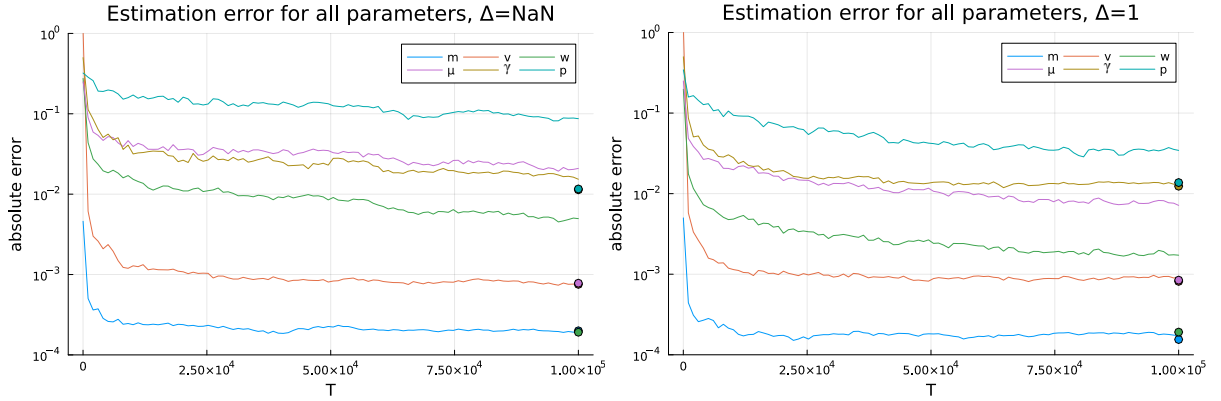
Figure IV.1: Absolute estimation error for the six estimators $\hat{m}_t^n, \ldots, \hat{p}_t^n$ (curves) and their theoretical temporal limits $m_\infty^n(\psi), \ldots, p_\infty^n(\psi)$ (marks) for $n = 500$. The y-axis is in log-scale. Each curve or mark corresponds to a median computed over 100 simulations. The parameters are $\mu = 0.25$, $\gamma = 0.5$, $p = 0.5$, $r_+ = 0.5$. The panels correspond to the choices $\Delta = \ln(t)$ and $\Delta = 1$ from left to right. *The blue, orange and olive marks are respectively hidden by the green, pink and cyan marks.*

Finally, we discussed the optimality of our estimation rate by analyzing two simple and related statistical settings (both inspired by the one considered by Delattre and Fournier (2016, Section 1.9)). This discussion can be found in (Chevallier et al., 2024, Appendix I).

Turn now to the second statistical problem. Remind that, according to Proposition III.26, the vector $\sigma^n(\psi) = (\sigma^{n,1}, \ldots, \sigma^{n,n})^\top$ given by

$$\sigma^{n,j} := \sum_{i=1}^n \mathrm{Cov}_\psi(X_1^{n,i}, X_0^{n,j})$$

can discriminate between the two communities $\mathcal{P}_+$ and $\mathcal{P}_-$. Hence, it is natural to consider its estimator $\hat{\sigma}_t^n = (\hat{\sigma}_t^{n,1}, \ldots, \hat{\sigma}_t^{n,n})$ defined as,

$$\hat{\sigma}_t^{n,j} = \sum_{i=1}^n \left[\frac{1}{t-1} \sum_{s=2}^t X_s^{n,i} X_{s-1}^{n,j} - \frac{N_t^{n,i}}{t} \frac{N_t^{n,j}}{t} \right].$$

Its rate of convergence reads as follows.

Proposition IV.6 (Chevallier et al. (2024, Theorem 2.4))**.** *For all $n \geq 2$ and $t \gtrsim_\gamma \ln n$,*

$$\mathbb{E}_\psi \left[\| \hat{\sigma}_t^n - \sigma^n(\psi) \|_\infty \right] \lesssim_\gamma \frac{n \ln(nt)}{\sqrt{t}}.$$

Proof. The proof relies on a Hoeffding type concentration inequality (Ost and Reynaud-Bouret, 2020, Theorem 3). □

Finally, a clustering algorithm applied to the estimated vector $\hat{\sigma}_t^n$ does the job. For instance, k-means algorithm with $k = 2$ applied to $\hat{\sigma}_t^n$ returns two clusters: let $\hat{\mathcal{P}}_-^{n,t}$ denote the one with the lowest values and $\hat{\mathcal{P}}_+^{n,t}$ the other one. Combining the large time asymptotics above with the high dimensional asymptotics, i.e. Proposition III.26, and Markov's inequality yields the following almost sure exact recovery result.

Theorem IV.7 (Chevallier et al. (2024, Corollary 2.5))**.** *Under $(\mathcal{A}_r^{\mathcal{P}})$ and the asymptotic regime*

$$\sqrt{\frac{\ln n}{n}} + \frac{n \ln(nt)}{\sqrt{t}} \to 0,$$

the probability of exact recovery goes to 1, i.e.

$$\lim_{n \to \infty} \mathbb{P} \left(\left\{ \hat{\mathcal{P}}_+^{n,t} = \mathcal{P}_+ \right\} \cap \left\{ \hat{\mathcal{P}}_-^{n,t} = \mathcal{P}_- \right\} \right) = 1.$$

Up to a log term, the asymptotic regime is satisfied as soon as $n \to \infty$ and $t \gg n^2$. Figure IV.2 illustrates this result via a Monte Carlo estimation of the probability of exact recovery. Obviously, the probability of exact recovery increases as a function of t for fixed n and conversely decreases as a function of n for fixed t. Finally, the separation between couples (t, n) for which exact recovery never occurs (black top-left corner) and those for which exact recovery always occurs (yellow bottom-right corner) seems compatible with the condition $t \gg n^2$ aforementioned.

Figure IV.2: Probability of exact recovery as a function of n and t (computed over 1000 simulations). The parameters are $\mu = 0.25$, $\gamma = 0.5$, $p = 0.5$, $r_+ = 0.5$.

Exact recovery implies 100% accuracy and as $n \to \infty$, it is obvious that exact recovery is much more difficult than achieving a fixed accuracy level (say 90%). Indeed (Chevallier and Ost, 2024, Figure 3) shows that the accuracy is high (above 90%) even if the probability of exact recovery is low (especially when n is large).

IV.3   Spatial point processes

My contributions regarding spatial point processes deal with the asymptotics of some (bi)-linear statistics related with characteristics of order one and two. The content of this section is based on the following preprint:

    ⬚ J. Chevallier, J.-F. Coeurjolly, and R. Waagepetersen.   Critical point processes obtained from a Gaussian random field with a view towards statistics. *preprint*, 2025.

This preprint initially was part of Anne-Lise Porté's PhD thesis under the supervision of Jean-François and me. Unfortunately she had to stop for medical reasons but we continued and added Rasmus as a collaborator.

IV.3.a)   Introduction

A spatial point process in $\mathbb{R}^d$ is a random locally finite subset N of $\mathbb{R}^d$. Equivalently, it can be seen as a random variable in the space $\mathcal{N}^{\#}$ of all boundedly finite counting measures on $\mathbb{R}^d$ (Daley and Vere-Jones, 2003, page 158).

For $B \subseteq \mathbb{R}^d$, $N \cap B$ is the restriction of N to B and let $|B|$ denote the volume of any bounded $B \subset \mathbb{R}^d$. Local finiteness of N means that the number of points $\mathrm{Card}(N \cap B)$ is finite a.s., whenever B is bounded. A point process is said to be stationary (resp. isotropic) if its distribution is invariant under translations (resp. rotations).

Assume that the kth ($k \geq 1$) factorial moment measure (Daley and Vere-Jones, 2003, Equation 5.4.4) is absolutely continuous with respect to the Lebesgue measure. Its Radon-Nikodym derivative is the kth order **intensity function** $\rho^{(k)}$ characterized by Campbell's theorem: for any nonnegative test function $\varphi : (\mathbb{R}^d)^k \to \mathbb{R}^+$,

$$\mathbb{E}\left[\sum_{t_1,\ldots,t_k \in N}^{\neq} \varphi(t_1,\ldots,t_k) \right] = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \varphi(t_1,\ldots,t_k) \rho^{(k)}(t_1,\ldots,t_k) \mathrm{d}t_1 \ldots \mathrm{d}t_k$$

where the sign $\neq$ means that the sum is defined for any pairwise distinct points $t_1,\ldots,t_k$. Roughly speaking, $\rho^{(1)}(t)\mathrm{d}t$ is the mean number of points in the vicinity of t and more generally, $\rho^{(k)}(t_1,\ldots,t_k)\mathrm{d}t_1 \ldots \mathrm{d}t_k$ can be interpreted as the probability of observing simultaneously a point in each of k infinitesimal neighborhoods around $t_1,\ldots,t_k$.

When a point process is stationary (resp. stationary and isotropic), $\rho^{(1)}(s) = \rho$ is constant and $\rho^{(2)}(s,t)$ depends only on $t - s$ (resp. on $\|t - s\|$). The **pair correlation function** (PCF) is defined by $g(s,t) := \rho^{(2)}(s,t)/\{\rho^{(1)}(s)\rho^{(1)}(t)\}$. In the stationary and isotropic case let, for $r > 0$,

$$g(r) = \frac{\rho^{(2)}(r)}{\rho^2} = \frac{\rho^{(2)}(re_1)}{\rho^2},$$

by abuse of notation where e_1 is any unit d-dimensional vector. The PCF measures deviations from the Poisson point process which is the reference model without any interactions between points. In particular, the kth order intensity of a Poisson point process exists and equals $\prod_{i=1}^k \rho(t_i)$ which leads to $g \equiv 1$. Hence, when $g(r) < 1$ (resp. $g(r) > 1$) two points at distance r are less likely (resp. more likely) to appear jointly than under a Poisson point process. A point process is said to be **repulsive** or **clustered** (resp. purely repulsive or purely clustered) if $g(0) < 1$ or $g(0) > 1$ (resp. $g(r) < 1$ or $g(r) > 1$ for any r). An other common second order characteristics is **Ripley's K-function**. It is a cumulative version of the PCF since (Daley and Vere-Jones, 2003, Equation 8.1.23):

$$K(r) = \int_{B(0,r)} g(\|t\|)\mathrm{d}t.$$

The repulsive and clustered properties as well as the PCF and Ripley's K function as illustrated by Figure IV.3.

Finally, let us mention the cross second-order characteristics which extends the previous definitions starting from the cross second-order intensity function $\rho^{(2)}_{N,N'}$ between two point processes N and N' which satisfies

$$\mathbb{E}\left[\sum_{s \in N, t \in N'}^{\neq} \varphi(s,t) \right] = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(s,t) \rho^{(2)}_{N,N'}(\|s - t\|)\mathrm{d}s\mathrm{d}t.$$
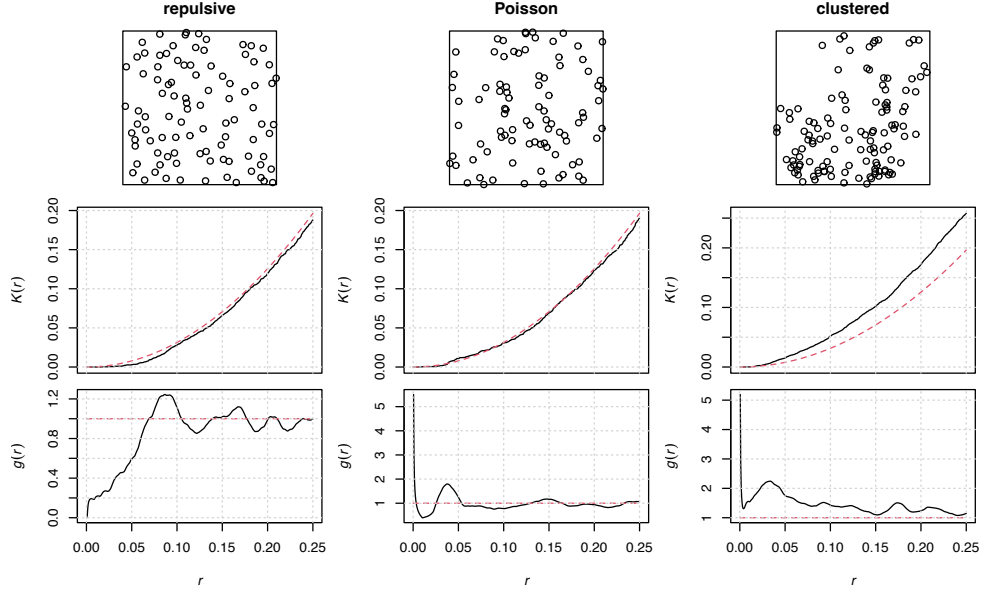
Figure IV.3: Top row: simulated point patterns. Middle row: Ripley's K-function estimated from the point pattern (black curve) and theoretical corresponding to a Poisson process (red dashed curve). Bottom row: PCF estimated from the point pattern (black curve) and theoretical corresponding to a Poisson process (red dashed curve). Left column: Bessel type determinantal process (repulsive model). Middle column: Poisson process (neutral model). Right column: Matérn cluster process (clustered model). *Made with the spatstat R package (Baddeley et al., 2015).*

Obviously, the standard second-order characteristics are particular cases of the cross ones, e.g. $\rho^{(2)} = \rho^{(2)}_{N,N}$.

Remind the $\mathcal{L}$-critical point process associated with a Gaussian random field (GRF) defined in Definition III.33. Denote by $\rho_{\mathcal{L}}$ the first-order intensity of $N_{\mathcal{L}}$. Let $g_{\mathcal{L},\mathcal{L}'}$ and $K_{\mathcal{L},\mathcal{L}'}$ denote respectively the cross pair correlation and cross Ripley's K-functions between $N_{\mathcal{L}}$ and $N_{\mathcal{L}'}$. The standard ones are denoted by $g_{\mathcal{L}} := g_{\mathcal{L},\mathcal{L}}$ and $K_{\mathcal{L}} := K_{\mathcal{L},\mathcal{L}}$.

As mentioned in Section III.4, the study of $N_{\mathcal{L}}$ and the references below heavily relies on Kac-Rice formula (Azaïs and Wschebor, 2009, Theorem 6.2). The intensity parameter $\rho_{\mathcal{L}}$ is expressed in terms of the correlation function of the GRF by Cheng and Schwartzman (2018). Next, Kratz and León (2006) gave conditions ensuring the finiteness of second order moments while Gass and Stecconi (2024) considered higher order moments. Dimension-dependent repulsion and clustering properties of critical point processes are studied by (Beliaev et al., 2020; Azaïs and Delmas, 2022; Ladgham and Lachièze-Rey, 2023) via the asymptotic characterization of the PCF $g_{\mathcal{L}}(r)$ as $r \to 0$. Finally, Estrade and León (2016), and Azaïs et al. (2024) proved central limit theorems, in any dimension, for the number of critical points and related quantities. Remark that most assumptions of the aforementioned papers are related to regularity and non-degeneracy of the GRF.

IV.3.b)   Contributions

We provided some numerical illustrations of previously obtained theoretical results and a central limit theorem for (bi)-linear statistics of $N_{\mathcal{L}}$.

This section borrows the GRF notations introduced in Section III.4. For instance, remind that X' and X'' are respectively the gradient and Hessian random fields. Here are given fundamental assumptions[4,5].

$\left(\mathcal{A}^{X}_{\mathcal{C}}[\nu]\right)$: | The GRF X is centered, stationary, isotropic and $X \in \mathcal{C}^{\nu}_{\text{a.s.}}(\mathbb{R}^d)$.

---

[4]Assumption $(\mathcal{A}^{X'}_{\text{nd}})$ is a strong combination of (Chevallier et al., 2025, Assumptions $(\mathcal{C}.3)$ and $(\mathcal{C}.5)$).
[5]See Appendix B for the multi-index notation ∂^{α}.

$\left(\mathcal{A}_{\mathrm{nd}}^{X}[p]\right):$ | The Gaussian random vector $(\partial^{\alpha}X_t)_{1\leq|\alpha|\leq p}$ is non-degenerate for any $t\in\mathbb{R}^d$.

$\left(\mathcal{A}_{\mathrm{nd}}^{X'}\right):$ | The Gaussian random vector $[(X_0')^{\top},(X_0'')^{\top},(X_{re_1}')^{\top},(X_{re_1}'')^{\top}]^{\top}$ is non-degenerate for all $r>0$, where e_1 is any unit d-dimensional vector.

$\left(\mathcal{A}_{L^1}^{c_2}\right):$ | There exists a decreasing function ϕ on $\mathbb{R}_+$, such that $\max_{p=1,\ldots,4}\{r^p|c_2^{(p)}(r^2)|\}\leq \phi(r)$ and $\int_0^{\infty}r^{d-1}\phi(r)\mathrm{d}r<\infty$.

Remark that the first two assumptions above depend on a parameter. For instance, Assumption $\mathcal{A}_{\mathcal{C}}^{X}[2]$ implies that X is $\mathcal{C}^2$ almost surely. The interested reader is referred to the lengthy discussion of these assumptions in (Chevallier et al., 2025, Sections 3.1 and 5.2). Up to our knowledge, the literature only studied the (cross-)PCF near the radius $r=0$, and we gathered these results into (Chevallier et al., 2025, Table 2).

Nevertheless, the following slight extension of (Azaïs and Delmas, 2022, Proposition 1.5) can be used to get numerical approximations of $g_{\mathcal{L}}(r)$ for any $r>0$.

Proposition IV.8 (Chevallier et al. (2025, Theorem 2)). *Under* $(\mathcal{A}_{\mathcal{C}}^{X}[2])$, $(\mathcal{A}_{\mathrm{nd}}^{X}[1])$ *and* $(\mathcal{A}_{\mathrm{nd}}^{X'})$, *for all* $\mathcal{L},\mathcal{L}'\subseteq\{0,1,\ldots,d\}$ *and* $r>0$, *we have*

$$g_{\mathcal{L},\mathcal{L}'}(r)=\frac{1}{\rho_{\mathcal{L}}\,\rho_{\mathcal{L}'}}\,f_{V(r)}(0,0)\mathbb{E}\Big[|\det\{X_0''\}|\times|\det\{X_{re_1}''\}|\times$$

$$\iota_{\mathcal{L}}\{X_0''\}\,\iota_{\mathcal{L}'}\{X''(re_1)\}\,\Big|\,X_0'=X_{re_1}'=0\Big],$$

where $f_{V(r)}$ *is the probability density of* $V(r)=[(X_0')^{\top},(X_{re_1}')^{\top}]^{\top}$. *Furthermore,* $g_{\mathcal{L},\mathcal{L}'}$ *is continuous.*

Sketch of Proof. The proof relies on a quite standard application of Kac-Rice formula (Azaïs and Wschebor, 2009, Theorem 6.4). □

The conditional expectation above seems hard to deal with. However, Azaïs and Delmas (2022, Lemma 8) computed the conditional distribution of $[(X_0'')^{\top},(X_{re_1}'')^{\top}]^{\top}$ given $X_0'=X_{re_1}'=0$. Therefore, an approximation of $g(r)$, for any $r>0$, can be obtained using Monte Carlo simulations of this conditional distribution. This is exploited in Figure IV.4. A lengthy discussion of this figure can be found in (Chevallier et al., 2025, Section 3.4 and Appendix B), especially in comparison to the asymptotic results gathered in Table 2 therein. Here are two key facts:

- for Matérn random fields, repulsion increases as the regularity parameter β increases;
- the RWM is more repulsive at low scales, its PCF oscillates a lot and tends slowly to 1 which indicates a near lattice behavior.

Considering $N_{\mathcal{L}}$ as a statistical model, it is important to investigate the asymptotics of its (bi)-linear statistics. The asymptotic takes place as the observation domain W increases. To ease the presentation, consider the sequence of bounded domains $W_n=[-n/2,n/2]^d$ (with volume n^d). Estrade and León (2016), and Azaïs et al. (2024) considered consistency and asymptotic normality for: i) normalized versions of $N_{\mathcal{L}}(W_n)$, ii) the number of critical points above a level u, and iii) linear combinations of these numbers for different sets $\mathcal{L}$. Their results can be significantly extended by considering (bi)-linear statistics as follows.

Definition IV.9. Let $\mathcal{L}\subseteq\{0,\ldots,d\}$ and let $\{\varphi_{1,n}\}_{n\geq1}$ and $\{\varphi_{2,n}\}_{n\geq1}$ be two sequences of measurable test functions with $\varphi_{1,n}:\mathbb{R}^d\to\mathbb{R}$ and $\varphi_{2,n}:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$. Then,

$$\Phi_{1,n}=\sum_{t\in N_{\mathcal{L}}\cap W_n}\varphi_{1,n}(t)\quad\text{and}\quad\Phi_{2,n}=\sum_{(t,s)\in(N_{\mathcal{L}}\cap W_n)^2}^{\neq}\varphi_{2,n}(t,s),$$
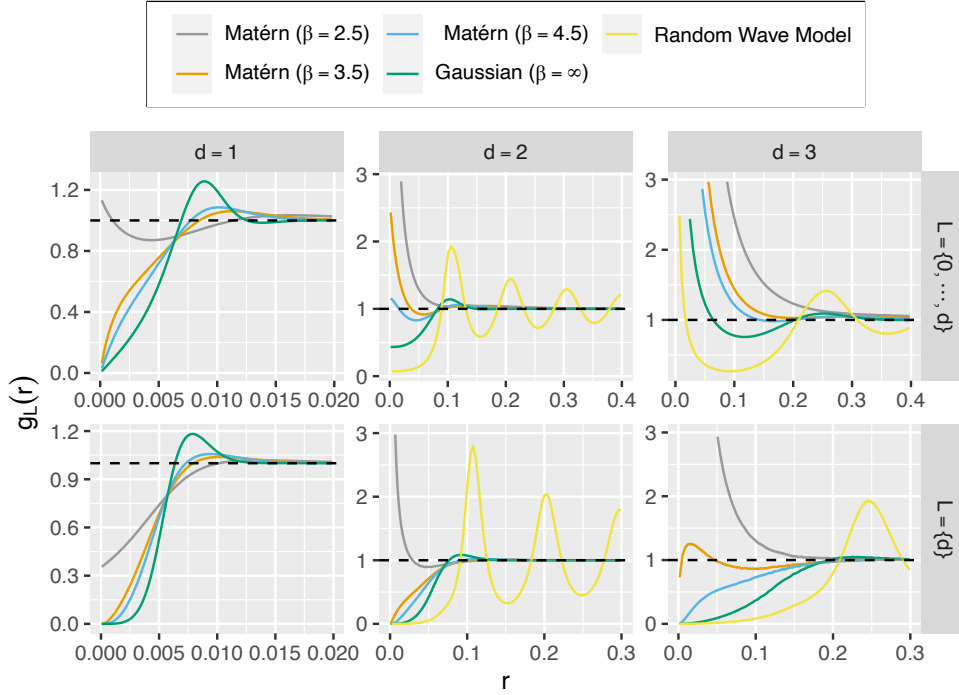
Figure IV.4: Monte Carlo estimations (with 10^7 simulations) of PCF for critical points with indices $\mathcal{L} = \{0, \dots, d\}$ and $\mathcal{L} = \{d\}$ based on Matérn random fields with regularity parameter $\beta = 2.5, 3.5, 4.5, \infty$ and on the Random Wave Model (for $d > 1$). The scale parameter σ is set such that $\rho_{\mathcal{L}} = 100$. Envelopes corresponding to 95% pointwise confidence intervals for the Monte Carlo estimates are represented but invisible.

are respectively called **linear** and **bilinear statistics** of $N_{\mathcal{L}}$.

Here are one example of each. On the one hand, $\varphi_{1,n}(t) = \mathbf{1}_{W_n}(t)/|W_n|$ yields the linear statistic $\Phi_{1,n} = \operatorname{Card}(N_{\mathcal{L}} \cap W_n)/|W_n| =: \hat{\rho}_{\mathcal{L}}$ which is the natural unbiased estimator of the intensity parameter $\rho_{\mathcal{L}}$. On the other hand, for some technical reasons[6], we restricted attention to functions $\varphi_{2,n}$ with support included in the diagonal cylinder

$$D_{\eta,R} := \left\{ (t, s) \in (\mathbb{R}^d)^2 : \eta \leq \|t - s\| \leq R \right\}, \tag{IV.3}$$

for some $0 < \eta < R$ fixed until the end. Now, for any $r > \eta$, consider the function

$$\varphi_{2,n}(t, s) = \frac{\mathbf{1}_{D_{\eta,r}}((t, s))}{\rho_{\mathcal{L}}^2 |W_n \cap (W_n)_{t-s}|}, \tag{IV.4}$$

where $\Delta_t = \{s + t, s \in \Delta\}$ for any domain Δ and $t \in \mathbb{R}^d$. By the second-order Campbell's Theorem, its associated bilinear statistic $\Phi_{2,n}$ satisfies $\mathbb{E}[\Phi_{2,n}] = K_{\eta,\mathcal{L}}(r)$ with

$$K_{\eta,\mathcal{L}}(r) := \int_{B(0,\eta,r)} g_{\mathcal{L}}(\|t\|)\mathrm{d}t \tag{IV.5}$$

where $B(0, \eta, r) = B(0, r) \backslash B(0, \eta)$ is the shell between the hyperspheres of radius η and r centered at 0. Obviously, $K_{0,\mathcal{L}} = K_{\mathcal{L}}$ is the standard Ripley's K-function but we were not able to put $\eta = 0$ in the proof of the central limit theorem. Then, a natural estimator of $K_{\eta,\mathcal{L}}(r)$ is

$$\hat{K}_{\eta,\mathcal{L}}(r) := \Phi_{2,n}\ \frac{\rho_{\mathcal{L}}^2}{\hat{\rho}_{\mathcal{L}}^2}. \tag{IV.6}$$

To study its asymptotic behavior, the first step is clearly to derive results for $\hat{\rho}_{\mathcal{L}} = \Phi_{1,n}$ and for $\Phi_{2,n}$ defined above. The second step is to investigate the joint distribution of $(\Phi_{1,n}, \Phi_{2,n})$. Both steps are tackled in the following central limit theorem.

---

[6]See (Chevallier et al., 2025, Section 5.1) for details.

Theorem IV.10 (Chevallier et al. (2025, Theorem 8)). *Under* $(\mathcal{A}_{\mathcal{C}}^{X}[5])$, $(\mathcal{A}_{\mathrm{nd}}^{X}[4])$, $(\mathcal{A}_{\mathrm{nd}}^{X'})$ *and* $(\mathcal{A}_{L^1}^{c_2})$, *let* $m \geq 1$, $\eta \leq r_1 < \cdots < r_m$ *and define*

$$\zeta_{\mathcal{L}} := \left[\hat{\rho}_{\mathcal{L}} - \rho_{\mathcal{L}}, \hat{K}_{\eta,\mathcal{L}}(r_1) - K_{\eta,\mathcal{L}}(r_1), \ldots, \hat{K}_{\eta,\mathcal{L}}(r_m) - K_{\eta,\mathcal{L}}(r_m)\right]^{\top}.$$

Then, there exists an explicit matrix Σ such that

$$n^{d/2}\zeta_{\mathcal{L}} \xrightarrow{\mathrm{law}} \mathcal{N}\left(0_{m+1}, \Sigma\right),$$

where 0_{m+1} is the null vector of $\mathbb{R}^{m+1}$. The matrix Σ is explicit in the sense that it is written as an explicit function of Hermite coefficients related to the GRF X.

Sketch of Proof. We followed the proofs in (Estrade and León, 2016). The main difficulty is to extend it to bilinear statistics (which has not treated in the literature up to my knowledge). First, the formalism needs to be extended and then the most difficult part is to control the Hermite coefficients related to the bilinear case (which are fundamentally different from the linear case).

First, the statistics $\Phi_{1,n}$ and $\Phi_{2,n}$ can be represented as limits of continuous integrals involving the GRF (Chevallier et al., 2025, Proposition 1): e.g.

$$\Phi_{1,n} = \lim_{\varepsilon \to 0} \int \varphi_{1,n}(t) |\det\left(X_t''\right)| \iota_{\mathcal{L}}\left(X_t''\right) \frac{\mathbf{1}_{[0,\varepsilon]}\left(\|X_t'\|\right)}{|B(0,\varepsilon)|} \, \mathrm{d}t,$$

almost surely and in the L^2-sense. The almost sure convergence is quite standard and can be seen as an application of (Adler and Taylor, 2009, Theorem 11.2.3) which relies on the inverse mapping theorem. The L^2 convergence for the bilinear statistic requires the continuity of some fourth moments which appears in the recent article (Gass and Stecconi, 2024, Remark 1).

Second, this representation can be used to derive the Hermite expansions of $\Phi_{1,n}$ and $\Phi_{2,n}$ and their limit variances and covariances. Their expressions are quite complex and writing them here would serve no purpose. However, their proof relies on: 1) a control of the moments of scaled Hermite polynomials, 2) a somehow surprising property of Gaussian vectors: rescaling a conditional distribution by a matrix that standardizes the full distribution gives a "rather standard" Gaussian vector in the sense that its covariance is a projection matrix.

Third, a central limit theorem involving multiple Wiener-Itô integrals (Peccati and Taqqu, 2011, Theorem 11.8.1) gives the asymptotic normality of $\Phi_{1,n}$ and $\Phi_{2,n}$.

Finally, the stated result follows from Slutsky's lemma and the multivariate delta method. $\quad\square$

IV.4   On going research and perspectives

Some of my ongoing works and perspectives are related to the **Stochastic Block Model** (SBM) so I introduce some notation here. Let n and q be some integers with $q < n$. The number of observations or neurons is n and they are grouped into q classes. Class membership is summarized by the $n \times q$ matrix Z such that $Z_{ik} = 1$ if neuron i is in class k. Given Z, the SBM adjacency matrix ψ is given by independent random variables and $\psi_{ij} \sim \mathrm{Ber}(p_{k\ell})$ where $Z_{ik} = Z_{j\ell} = 1$. The Erdős-Rényi model corresponds to the SBM with $q = 1$ class.

Graph observed through a fixed point projection. Let Ψ be the random matrix used as a random environment for a binary Markov chain in Sections III.2 and IV.2. Conditionally on $\Psi = \psi$, consider the vector $X \in \mathbb{R}^n$ which satisfies

$$X = \mu\mathbf{1}_n + n^{-1}\gamma\psi X$$

This equation is similar to the one satisfied by the stationary mean vector $m^n(\psi)$ - Equation (III.36). Hence, this model is related to the MCRE but in some sense simpler since there is no stochastic dynamics.

I am now working on a combination of the model introduced above with a Stochastic Block Model structure. In the most general case, it writes as

$$X = Z\boldsymbol{\mu} + n^{-1}(Z\boldsymbol{\gamma}) \odot (\psi Z\boldsymbol{\eta} \odot X),$$

where $\boldsymbol{\mu}, \boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ are parameters in $\mathbb{R}^q$ and $\odot$ denotes Hadamard product. The full model is expected to be non identifiable but I am searching for some restrictions that may be identifiable. As an estimation procedure, I am considering quantities analog to the statistics $\hat{m}_t^n$ and $\hat{v}_t^n$ introduced in Subsection IV.2.b) combined with some contrast minimization.

High dimensional Hawkes processes. Most statistical procedures involving Hawkes processes do not scale with the dimension (i.e. the number of neurons). Recently, some studies considered some sparsity or low-rank assumption (Bacry et al., 2020; Wang et al., 2025). Another line of research is to assume an SBM structure on the nodes (e.g. neurons). Here are two examples.

- Consider a n-dimensional Hawkes process $(N^i)_{i=1,\dots,n}$ whose parameters follow the class structure. To be precise, the firing rate functions f_i and the delay functions $h_{j\to i}$ do only depend on the respective classes of i and j. If the classes are known then the estimation problem is a lot easier and is solved by standard procedures. If the classes are unknown, estimation and clustering must be performed at the same time by an EM algorithm (Dempster et al., 1977) for instance. This line of research is part of the ANR proposal *MIAMI* described below.
- Consider n^2 independent univariate Hawkes processes $(N^{i,j})_{i,j=1,\dots,n}$ whose parameters follow the class structure. To be precise, the firing rate functions f_i and the delay functions $h^{i,j}$ do only depend on the respective classes of i and j. This model is more suited to the modeling of message exchanges on social media. In that case, the spikes of $N^{i,j}$ are the times at which user j sends a message to user i. Once again, the objective is to combine estimation and clustering. This line of research was the subject of Martin Combelles' internship. It gave some promising results using EM algorithm.

Here is a brief overview of the ANR proposal *MIAMI* led by Elodie Fino and submitted to the AAPG (*Appel à projets générique*) 2025. It is an interdisciplinary project combining experimental neurobiologists and statisticians (S. Achard and S. Jaffard). The objective is to understand how multisensory integration is computed, by combining *in vivo* calcium imaging recordings and data analysis identifying the structure of neuronal connectivity networks.

From the biological point of view, a mouse is running on a self-paced treadmill enriched with sensory cues (pads, LED, etc.). A camera is held in place on its head and longitudinal in vivo experiments are recorded. Neuronal activity is timely aligned with behavior and cues stimulations. These stimulations can be unimodal (only visual or tactile) or multimodal (both visual and tactile). Hence, the main question of interest is: which neuronal ensembles are responsible for uni- vs multisensory integration? Our proposal is to answer this question by inferring and comparing neuronal connectivity networks in different stimulus configurations.

From the methodological/mathematical point of view, there are 3 bottlenecks that we want to tackle.

- The time resolution of calcium recordings is way less than electrophysiological recordings which are usually used to extract spike trains. Hence the spike inference for calcium recording is not trivial and state-of-the-art techniques must be adapted to our recordings.

- Calcium imaging records hundreds of neurons simultaneously. Hence, statistical methods that scale with this high dimension are necessary. Assume that the spike inference described above is successful. The n-dimensional Hawkes process with SBM structure described above is one line of research, but we also propose an alternative with the use of variational autoencoders (Yang and Zha, 2024). If these two methods fail to cluster neurons in a satisfactory way, we will turn to the whole calcium recordings and treat them as univariate time series.
- The recorded field correspond to the dorsomedial striatum which is not an isolated neural network. Hence, stimuli coming from the cortex or the thalamus are latent covariates for our observations. A promising line of research is to follow Jaffard (2025, Chapter 6) which considers a minimum contrast estimator based on first and second-order moment equations.

Latent GRF estimation. From a statistical viewpoint, the main objective of our study of $\mathcal{L}$-critical point processes is to infer some characteristics of the underlying (latent) GRF. Up to now, we only established a link between the first two intensity functions of a critical point process and some spectral moments of the latent GRF.

On the one side, such a link can be exploited into an estimation procedure via some contrast minimization like (Waagepetersen and Guan, 2009) for instance. Furthermore, asymptotic results for the minimum contrast estimates could be obtained using Theorems IV.10 in combination with a standard central limit for the Monte Carlo approximation of the Ripley's K-function.

On the other side, there are important, yet standard, summary functions to be established for critical point processes, such as the nearest-neighbor distribution function and the void probabilitiesto name a few.

These two lines of research are part of the ANR proposal *ESCaPe*. It is led by Jean-François Coeurjolly and submitted to the AAPG (*Appel à projets générique*) 2025. Here is a brief overview of the project.

The project aims at investigating mathematical and statistical properties of particular random sets $\mathcal{S}$ obtained from a random field X. Critical point processes, excursion sets and nodal sets are some examples of such sets. The main objective of the project is to answer the following question: given a single observation of a set $\mathcal{S}$, potentially difficult to characterize both theoretically and numerically, how can one infer information about the latent random field X?

This objective is tackled through 3 axes:

1. Extend the existing probability literature to models more suited to data: e.g. non stationary fields, fields with spatial covariates, discrete fields;
2. Simulation of $\mathcal{S}$ and/or X: e.g. computation of $\mathcal{S}$ from X, conditional simulation of a field X with prescribed set $\mathcal{S}$, reconstruction of X from $\mathcal{S}$;
3. Estimation of the latent random field: e.g. the perspective detailed above, anisotropy estimation, Gaussianity test.

CRASH COURSES

Here are gathered some basic facts on the three mathematical domains involved in this manuscript. It is intended to help non specialist readers.

A.1 Partial differential equations

A.1.a) Method of characteristics

The **characteristics** associated with the transport equation (TE) are a family of curves $X_x : t \in \mathbb{R} \mapsto X_x(t) \in \mathbb{R}^d$ indexed by $x \in \mathbb{R}^d$ such that

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}X_x(t) = v(t, X_x(t)), & t \in \mathbb{R} \\ X_x(0) = x. \end{cases}$$

The so-called method of characteristics is used to reduce the PDE to an ordinary differential equation. The core idea is that the solution of (TE) is constant along each characteristic trajectory,

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t, X_x(t)) = \partial_t u(t, X_x(t)) + \sum_{i=1}^{d} v_i(t, X_x(t))\partial_{x_i} u(t, X_x(t)) = 0.$$

In turn, the unique solution of (TE) is given by

$$u(t, x) = u^{\mathrm{in}}(X_x^{-1}(t)).$$

Obviously, this method requires that X_x is invertible.

In the simple case where v is constant, the characteristic curves are lines in the time-space domain. In the equations above, the space domain does not have any boundary, but this method can be extended to that case. For instance, if the space domain is $\mathbb{R}_+$ with boundary at $x = 0$ the characteristic curves which reach the boundary $x = 0$ instead of the boundary $t = 0$ involve the time-section $u(\cdot, 0)$ instead of the space-section $u(0, \cdot) = u^{\mathrm{in}}$. Such a case is illustrated by Figure A.1 in dimension $d = 1$ and $v \equiv 1$ which is inherent of age-structured equations. Here, the time and space sections are known for simplicity but in general the time-section is unknown and may depend on u itself in a non-linear way.
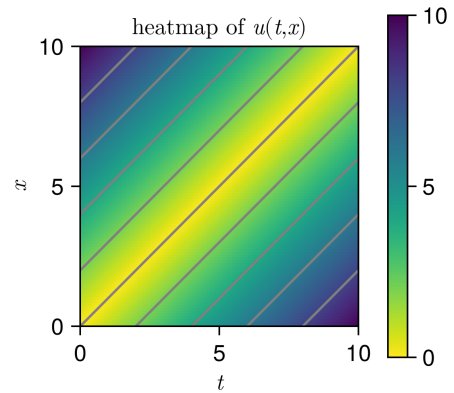

Figure A.1: Heatmap of the solution of $\partial_t u(t, x) + \partial_x u(t, x) = 0$ with initial condition $u^{\mathrm{in}}(x) = x$ and boundary condition $u(t, 0) = t$. The characteristic curves are superimposed as gray lines.

A.1.b) Weak formulation

Let $\varphi : (t, x) \in \mathbb{R} \times \mathbb{R}^d \mapsto \varphi(t, x) \in \mathbb{R}$ be any test function in $\mathcal{C}_c^\infty(\mathbb{R} \times \mathbb{R}^d)$. Multiplying the first line of (TE) by φ and integrating over the time-space domain yields

$$\int_{\mathbb{R}_+ \times \mathbb{R}^d} \left(\partial_t u(t, x) + \sum_{i=1}^d v_i(t, x) \partial_{x_i} u(t, x) \right) \varphi(t, x) \, \mathrm{d}t \mathrm{d}x = 0.$$

To remove the differential operators acting on u, the core idea is to use integration by parts. Since φ is compactly supported, all boundary terms are null except at $t = 0$, so that any solution of (TE) satisfies

$$-\int_{\mathbb{R}_+ \times \mathbb{R}^d} \left(\partial_t \varphi(t, x) + \sum_{i=1}^d \partial_{x_i} \left[v_i(t, x) \varphi(t, x) \right] \right) u(t, x) \, \mathrm{d}t \mathrm{d}x = \int_{\mathbb{R}^d} \varphi(0, x) u^{\mathrm{in}}(x) \mathrm{d}x. \qquad \text{(A.1)}$$

Hence, the definition of a weak solution (Definition II.3).

If the equation is a Fokker-Planck equation related to the stochastic process $(X_t)_{t \geq 0}$, the terms appearing in (A.1) are interpreted as expectations involving X_t, for instance

$$\int_{\mathbb{R}_+ \times \mathbb{R}} \partial_t \varphi(t, x) u(t, x) \mathrm{d}t \mathrm{d}x = \int_{\mathbb{R}_+} \mathbb{E} \left[\partial_t \varphi(t, X_t) \right] \mathrm{d}t.$$

A.2   Probability

A.2.a)   Probability space

Let Ω be an arbitrary set, usually called **sample space**. From a probabilistic point of view, the elements ω of Ω are called **sample points** and the subsets of Ω are called **events**. This space is equipped with a σ-algebra denoted by $\mathcal{F}$ and a probability measure denoted by $\mathbb{P}$. The σ-algebra $\mathcal{F}$ is a collection of events satisfying some properties and the probability measure $\mathbb{P}$ is a function that assigns a probability to each event in $\mathcal{F}$ and satisfies some properties. The triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**. Generally, the sample space Ω can be chosen according to the problem at hand: e.g. $\Omega = \{\text{head}, \text{tail}\}$ for a coin toss, $\Omega = \mathbb{R}^3$ for the position of a particle at a fixed time or $\Omega = \mathcal{C}(\mathbb{R}_+, \mathbb{R}^3)$ for its (continuous) trajectory. However, Ω may be larger than necessary without any harm as soon as $\mathbb{P}$ is chosen accordingly. That is why the sample space is often of little importance but the focus is rather on random variables and especially their distribution.

A random variable with values in the measurable space $(E, \mathcal{E})$ is a measurable function X from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$. Its distribution is nothing else but the pushforward of $\mathbb{P}$ onto $(E, \mathcal{E})$ induced by X. More precisely, it is the probability measure ν on E defined by $\nu(A) = \mathbb{P}(X^{-1}(A))$ for all $A \in \mathcal{E}$, and we denote $X \sim \nu$. As for the sample space, the value space E may for instance be discrete, continuous or even functional. Notice that for any measurable function φ from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$, $\varphi(X)$ is also a random variable.

It is common to consider the σ-algebra generated by X, that is $\sigma(X) = \{X^{-1}(A), A \in \mathcal{E}\}$ the smallest σ-algebra equipped to Ω such that X is measurable. From a probabilistic point of view, it represents the information given by X about the sample point ω. Indeed, if one observes $X(\omega) \in A$ he knows that $\omega \in X^{-1}(A)$.

A.2.b)   Polish space

For several technical reasons[1], it is convenient to assume that E is a Polish space, that is a separable topological space (namely it contains a countable dense subset) such that its topology can be induced by a complete metric (in the sense that Cauchy sequences converge). In the following, E is a Polish space equipped with the Borel σ-algebra $\mathcal{E} = \mathcal{B}(E)$, that is the σ-algebra generated by the open sets of E. Most of the time, the metric is obvious (for instance the Euclidean metric for $E = \mathbb{R}^d$). If not obvious, see Appendix B.

A.2.c)   Types of convergence

Denote by d_E some metric on the Polish space E. We say that

- $(X_n)_{n \in \mathbb{N}}$ converges **almost surely** to X_∞ if and only if $\mathbb{P}(\lim_{n \to \infty} X_n = X) = 1$;
- $(X_n)_{n \in \mathbb{N}}$ converges **in probability** to X_∞ if and only if $\mathbb{P}(d_E(X_n, X_\infty) > \varepsilon) \to 0$ for all $\varepsilon > 0$;
- $(X_n)_{n \in \mathbb{N}}$ converges **in distribution** to X, if and only if $\mathbb{E}[\varphi(X_n)] \to \mathbb{E}[\varphi(X)]$ for all bounded continuous functions $\varphi : E \to \mathbb{R}$.

Notice that the convergence in distribution is related to the weak-$*$ convergence in the context of functional analysis where the space of distribution is viewed as the dual of the space of continuous functions vanishing at infinity (instead of bounded). The reader is referred to the Portemanteau theorem for equivalent definitions of the convergence in distribution.

A.2.d)   Filtrations

Since the index t represents time, it is natural to consider a stochastic process as a dynamic arrival of random variables at each time step. In this context, the information of an observer is increasing with respect to time and we often consider the collection of σ-algebras given by $\mathcal{F}_t^X = \sigma(X_s, s \leq t)$. More generally, increasing sequences of σ-algebras are called **filtrations**.

Given a filtration $(\mathcal{F}_t)_{t \in \mathcal{T}}$, it is common to decompose, in a unique manner, a stochastic process $(X_t)_{t \in \mathcal{T}}$ into a predictable part and a martingale part, $X_t = A_t + M_t$. In words, it means that $(A_t)_{t \in \mathcal{T}}$ captures the predictable mean dynamics of the process while $(M_t)_{t \in \mathcal{T}}$ captures the unpredictable fluctuations. More formally, in the case $\mathcal{T} = \mathbb{N}$, $(A_t)_{t \in \mathbb{N}}$ is predictable meaning that A_t is $\mathcal{F}_{t-1}$-measurable, $(M_t)_{t \in \mathbb{N}}$ is a martingale meaning "essentially" that $\mathbb{E}[M_t \,|\, \mathcal{F}_{t-1}] = M_{t-1}$ and the decomposition is called Doob decomposition theorem. In continuous time, the definitions are more technical and the result is called Doob–Meyer decomposition theorem.

Let $(M_t)_{t \in \mathcal{T}}$ be a square integrable martingale and $(\langle M \rangle_t)_{t \in \mathcal{T}}$ denote the unique predictable process given by Doob-Meyer decomposition applied to the square process $(M_t^2)_{t \in \mathcal{T}}$, that is $M_t^2 - \langle M \rangle_t$ defines a martingale. The process $(\langle M \rangle_t)_{t \in \mathcal{T}}$ is called the **quadratic variation** of $(M_t)_{t \in \mathcal{T}}$. Jensen's inequality implies that the quadratic variation must be non decreasing.

A.2.e)   Heuristics for the intensity

Equivalently, a point process is characterized by its **inter-spike-intervals** (ISIs), defined by $S_n = T_n - T_{n-1}$ for $n \geq 1$ with the convention $T_0 = 0$. The most common example is the Poisson point process (of intensity $\lambda > 0$) where the ISIs are i.i.d. (independent and identically distributed) with distribution $\mathcal{E}(\lambda)$. Remind that the probability density and survival functions of $\mathcal{E}(\lambda)$ are given by $f_\lambda(t) = \lambda e^{-\lambda t}$ and $S_\lambda(t) = e^{-\lambda t}$ on $\mathbb{R}_+$. Commonly used in survival analysis, the

---

[1]For instance, i) if E is not separable then its cardinality may be larger than $2^{\aleph_0}$ (the cardinality of $\mathbb{R}$) and the product measures of such large spaces behave badly, ii) a metric is necessary to define the notion of convergence in probability, iii) completeness is really convenient when dealing with convergences.

hazard function h associated with some probability density f and survival function S is defined by, for all $t \geq 0$ such that $S(t) > 0$,

$$h(t) = \frac{f(t)}{S(t)}.$$

It is related to the probability that an event occurs at time t given that it has not occurred yet. One key fact is that the hazard function associated with the exponential distribution $\mathcal{E}(\lambda)$ is constant equal to λ (which is the intensity of the corresponding Poisson point process).

More generally, an observer at time t knowing the event times before time t, say $0 < t_1 < \cdots < t_n < t$, could characterize the conditional distribution of the next time T_{n+1} given that $T_1 = t_1, \ldots, T_n = t_n$ by its hazard function. This gives rise to the iterative process $(\lambda_t)_{t \geq 0}$:

- between 0 and T_1, $\lambda_t = h_{T_1}(t)$, the hazard function of T_1;
- between T_1 and T_2, $\lambda_t = h_{T_2|T_1}(t)$, the conditional hazard function of T_2 given T_1;
- and so on...

This process corresponds to the intuition underlying the **intensity** of the point process. However the rigorous definition of the intensity is more technical and usually involves martingales[2].

A.3   Statistics

A.3.a)   Hypothesis testing

The most simple hypothesis test takes the following form: one wants to discriminate between two hypotheses on some parameter θ,

- $\mathcal{H}_0 : \theta = \theta_0$,
- $\mathcal{H}_1 : \theta \neq \theta_0$,

where $\theta_0 \in \Theta$ is some prescribed value. The decision of a test is either to reject or not reject the null hypothesis $\mathcal{H}_0$. There are two possible types of errors:

- **Type I error:** rejecting $\mathcal{H}_0$ when it is true;
- **Type II error:** failing to reject $\mathcal{H}_0$ when it is false.

The construction of an hypothesis test relies on finding a good statistic S to discriminate between $\mathcal{H}_0$ and $\mathcal{H}_1$: for instance, $S \approx 0$ under $\mathcal{H}_0$, $S \ll 0$ when $\theta < \theta_0$ and $S \gg 0$ when $\theta > \theta_0$. Then, a control of the distribution of S under $\mathcal{H}_0$ provides the good threshold value t such that the type I error of the decision "reject $\mathcal{H}_0$ when $|S| > t$" is controlled.

The hypothesis tests considered in this manuscript are independence tests therefore non parametric. They do not follow the parametric formulation as above, but the underlying idea is the same.

A.3.b)   Clustering

The standard **clustering** framework reads as: given data points $X_1, \ldots, X_n$ in $\mathbb{R}^d$, the goal is to partition the dataset into groups (clusters) such that points in the same cluster are similar and points in different clusters are dissimilar. It is common to model data by a mixture model. Unlike classification, clustering is an unsupervised task: there is no training set on which you know the true labels (i.e. the mixture component from which data has been generated).

---

[2]In fact, if the filtration is the canonical filtration of N then the definition via hazard rates is correct and coincides with the definition via martingales.

The best known clustering algorithm is k-means. Its objective is to (iteratively) solve the following optimization problem

$$\underset{c_1,\dots,c_k}{\arg\min} \sum_{i=1}^{n} \min_{1\leq\ell\leq k} \|X_i - c_\ell\|^2,$$

where $c_1,\dots,c_k$ are k cluster centers.

In our simulations, the true labels (i.e. the population of each neuron) are known. Hence, there are several metrics to evaluate a clustering procedure like its accuracy (the proportion of good classification) or the probability of exact recovery.

SPACES

Here are gathered all the spaces used in the present manuscript. For each space, either its topology or metric or norm is specified. Concerning Polish spaces, remark two well-known facts: i) any separable Banach (and therefore Hilbert) space is a Polish space, ii) a countable product of Polish spaces is a Polish space.

B.1 Continuous functions

The following is based on (Adams and Fournier, 2003, Chapter 1). Let us introduce the multi-index notation for derivatives. Let $\alpha = (\alpha_1, \ldots, \alpha_d)$ be a d-tuple of non-negative integers. Its size is denoted by $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and $\partial^\alpha := \partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}$ is a differential operator of order $|\alpha|$.

Let U be an open set of $\mathbb{R}^d$. The space $\mathcal{C}_{\mathrm{b}}^k(U)$ is a Banach space with norm given by

$$\|\varphi\|_{\mathcal{C}_{\mathrm{b}}^k(U)} := \max_{|\alpha| \leq k} \sup_{x \in U} |\partial^\alpha \varphi(x)|.$$

Let $\bar{U}$ denote the closure of U. Then, let $\mathcal{C}^k(\bar{U})$ denote the space of functions $\varphi \in \mathcal{C}^k(U)$ with bounded and uniformly continuous derivatives on U (continuously extended to $\bar{U}$). It is a closed subspace of $\mathcal{C}^k(U)$ and so a Banach space with the same norm $\|\cdot\|_{\mathcal{C}_{\mathrm{b}}^k(U)}$. Moreover, if U is bounded then $\mathcal{C}^k(\bar{U})$ is separable as a consequence of Stone-Weierstrass' theorem. In that case, the space $\mathcal{C}^\infty(\bar{U})$ is a Polish space equipped with the inductive limit metric given by

$$d_{\mathcal{C}^\infty(\bar{U})}(\varphi_1, \varphi_2) := \sum_{k=0}^\infty 2^{-k} \left(1 \wedge \|\varphi_1 - \varphi_2\|_{\mathcal{C}^k(\bar{U})} \right).$$

For non-integer exponents $\beta > 0$, let $\mathcal{C}^\beta(\bar{U})$ denote the space of functions $\varphi \in \mathcal{C}^{\lfloor \beta \rfloor}(\bar{U})$ satisfying the local Hölder condition: for all compact $\Delta \subset U$, there exists $\kappa > 0$ such that

$$|\partial^\alpha f(x) - \partial^\alpha f(y)| \leq \kappa \|x - y\|^{\beta - \lfloor \beta \rfloor}$$

for all $\alpha \in \mathbb{N}^d$ such that $|\alpha| = \lfloor \beta \rfloor$ and $x, y \in \Delta$.

The following is based on (Billingsley, 1999, Chapter 3) and deals with the space of càdlàg functions. For a non decreasing function $\tau : [0,1] \to [0,1]$ such that $\tau(0) = 0$ and $\tau(1) = 1$, let

$$\|\tau\|^\circ := \sup_{s<t} \left| \ln \frac{\tau(t) - \tau(s)}{t - s} \right|.$$

Such a function is used to allow small time deformation in order to accommodate for discontinuities. The space $\mathcal{D}([0,1])$ is a Polish space equipped with the metric

$$d_{\mathcal{D}([0,1])}(\varphi_1, \varphi_2) := \inf_\tau \left(\|\tau\|^\circ \vee \|\varphi_1 - \varphi_2 \circ \tau\|_{\mathcal{C}((0,1))} \right),$$

with the abuse of notation $\| \cdot \|_{\mathcal{C}((0,1))}$ even if $\varphi_1 - \varphi_2 \circ \tau \notin \mathcal{C}((0,1))$. The space $\mathcal{D}(\mathbb{R}_+)$ is also a Polish space equipped with a metric obtained as the inductive limit of the metrics $d_{\mathcal{D}([0,t])}$ as $t \to \infty$ (see Billingsley, 1999, Section 16). Analogously, $\mathcal{C}(\mathbb{R}_+)$ is a Polish space equipped with the inductive limit norm given by

$$\|\varphi\|_{\mathcal{C}(\mathbb{R}_+)} := \sum_{t=1}^{\infty} 2^{-t} \left(1 \wedge \|\varphi\|_{\mathcal{C}([0,t])} \right).$$

These definitions are naturally extended to the case when the target space is not $\mathbb{R}$ but a normed space.

B.2 Integrable functions

The following is based on (Adams and Fournier, 2003, Chapters 2 and 3). For any $p \in [1, \infty]$, the space $L^p(\mathbb{R}^d)$ is a Banach space with the norm given by

$$\|\varphi\|_p := \left(\int_{\mathbb{R}}^d |\varphi(x)|^p \mathrm{d}x \right)^{1/p} \text{ for } p < \infty \text{ and } \|\varphi\|_\infty := \operatorname*{ess\,sup}_{x \in \mathbb{R}^d} |\varphi(x)|.$$

It is separable when $p < \infty$.

For any integer k and $\alpha > 0$, the space $\mathcal{W}^{k,\alpha}$ is defined as the completion of $\mathcal{C}_c^\infty(\mathbb{R}_+)$ for the following norm

$$\|\varphi\|_{k,\alpha} := \left(\sum_{j=0}^{k} \int_{\mathbb{R}_+} \frac{|\varphi^{(j)}(x)|}{1 + |x|^{2\alpha}} \mathrm{d}x \right)^{1/2},$$

where $\varphi^{(j)}$ is the j-th derivative of φ. Then, $\mathcal{W}^{k,\alpha}$ equipped with this norm is a separable Hilbert space. Its dual space is denoted by $\mathcal{W}^{-k,\alpha}$ and its norm is denoted by $\| \cdot \|_{-k,\alpha}$.

B.3 Measures

The following is based on (Daley and Vere-Jones, 2003, Appendix 2). Let E be a Polish space. The space of probability measures on E is denoted by $\mathcal{P}(E)$. It is a Polish space when equipped with the topology of weak convergence (a.k.a. convergence in distribution). For instance, it can be generated by the bounded Lipschitz metric given by

$$d_{\mathrm{BL}}(\nu_1, \nu_2) := \sup_{\varphi} \left| \int_E \varphi(x)\nu_1(\mathrm{d}x) - \int_E \varphi(x)\nu_2(\mathrm{d}x) \right|$$

where the sup is taken over all test functions φ that are bounded by 1 and 1-Lipschitz. Remark that it looks like Kantorovich-Rubinstein metric, Equation (III.1), but it is coarser.

Finally, taking an inductive limit like in Section B.1 yields a separable and complete metric for the space $\mathcal{N}^\#$ of all boundedly finite counting measures so that it is Polish.

C

TWO-SIDED TESTS

After years of teaching to students in biology or mathematics and discussions with colleagues, I am convinced that statistical tests are one of the most misunderstood mathematical concepts despite being quite simple and commonly used. I think that one of the main problems lies in the choice between one- and two-sided tests. And this problem arises at all academic levels: undergraduate or postgraduate students and even some professors in statistics may be misleading on this topic.

For many years, I have been convinced that a good principle should be to teach only two-sided tests (at first) because they are less biased: they avoid the biased choice of a one-sided test based on some overview of the data. An inspiring discussion with R. Drouilhet made me realize that there is some kind of fierce battle between the partisans of one-sided tests versus partisans of two-sided tests. The fruit of this discussion is detailed below in perhaps the most simple framework: Student's t-test with respect to the specified mean value $\mu_0 = 0$.

C.1 The two-sided test is dead...

The null hypothesis of the two-sided test is $\mathcal{H}_0 : \mu = 0$. The fact that the null hypothesis is a point in a continuous space is quite disturbing.

For instance, from a Bayesian viewpoint it is common to assume that μ follows a continuous distribution which in turn means that $\mathcal{H}_0$ is false almost surely. In that regard, the problem of testing whether $\mathcal{H}_0$ is true or false seems meaningless. Since the test is consistent, it gives 100% rejection rate even if $\mu \approx 0$ as soon as the sample size n goes to infinity. I think that even a pure frequentist may be quite troubled by this fact.

Obviously, the one-sided tests $\mathcal{H}_0 : \mu < 0$ and $\mathcal{H}_0 : \mu > 0$ do not suffer from this issue.

C.2 Long live the two-sided test

Like a phoenix rising from the ashes, the usual two-sided test can be viewed as a multiple test made from the 2 one-sided tests. In particular, the family-wise error rate of this multiple test corresponds to the prescribed level of the two-sided test. But, does it really make any difference?

Yes! A rather philosophical one and a practical one. First, the test should not be thought as $\mathcal{H}_0 : \mu = 0$ versus $\mathcal{H}_1 : \mu \neq 0$ but rather

$$\mathcal{H}_0 : \mu = 0 \quad \text{versus} \quad \mathcal{H}_1^- : \mu < 0 \text{ or } \mathcal{H}_1^+ : \mu > 0.$$

Second, in case $\mathcal{H}_0$ is rejected the decision of the test is much more interesting than the one usually taught. It should not be "accept $\mathcal{H}_1 : \mu \neq 0$" but rather "accept $\mathcal{H}_1^- : \mu < 0$" or "accept $\mathcal{H}_1^+ : \mu > 0$" depending on the sign of the test statistics (which is in fact the intuition that many students have when they apply their first two-sided test).

C.3 What about naturally one-sided tests?

Of course, the discussion above does not apply to any statistical test. Some of them are naturally one-sided such as goodness-of-fit and independence tests. Are they?

Indeed, a two-sided goodness-of-fit test could be used to avoid data that seems too good to be true such as Mendel's data (Weeden, 2016).

Scientific activity

Here are gathered some informations about my scientific activity since my PhD that are not covered by the main part of the present manuscript.

D.1 Other scientific productions

- J. Chevallier and P. Reynaud-Bouret (participation of M. Monticelli for the animations). Nos neurones se synchronisent-ils ?, *Image des Mathématiques (website)*, 2017.
  A series of two popular articles (part 1 and part 2) related to the coincidence count described in Section IV.2.
- J. Chevallier. Approximation par champ-moyen: le couplage à la Sznitman pour les nuls, *lecture notes*, 2017.
  These notes detail the first two objectives of mean-field approximations described in Subsection III.2.a) on some toy example.
  They were written (in french) as a supplementary material to my presentation at the fifth edition of the Young Statisticians and Probabilists days. I am planning to translate them in english and add a section dealing the third objective of mean-field approximations.
- J. Chevallier. Uniform decomposition of probability measures: quantization, clustering and rate of convergence, *Journal of Applied Probability*, 2018.
  Given a probability measure on $\mathbb{R}^d$, it provides a discrete measure supported on n points with equals mass n^{-1}. This quantization measure yields sharp Wasserstein W_p distances for generic probability measures with finite moments of any order. The general case is treated via a suboptimal truncation method which has been refined by (Seeger, 2025, Theorem 1).
- P. Grazieschi, M. Leocata, C. Mascart, J. Chevallier, F. Delarue, and E. Tanré. Network of interacting neurons with random synaptic weights, *ESAIM: ProcS*, 2019.
  Theoretical and numerical study of a mean-field limit of stochastic leaky integrate-and-fire neurons with random synaptic weights and in particular the case when the number of non-null synaptic weights is sparse. This publication was the result of a research project led by F. Delarue E. Tanré and I at CEMRACS 2017.
- P. J. Gerrish, B. Galeota-Sprung, F. Cordero, P. Sniegowski, A. Colato, N. Hengartner, V. Vejalla, J. Chevallier, and B. Ycart. Natural selection and the advantage of recombination, *preprint*, 2020.
  We show, on some toy example, that natural selection can amplify poorly-matched gene combinations and thus create negative associations.
- P. J. Gerrish, B. Galeota-Sprung, P. Sniegowski, A. Colato, J. Chevallier, and B. Ycart. Natural selection promotes the evolution of recombination 1: between the products of natural selection, *preprint*, 2021.
  It is analog to the previous item but on a more elaborate model.

D.2 Research projects

- *IRGA IDEX* – STePP-Meteo – 2022/24 – PI : Jean-François Coeurjolly
  This project focuses on more or less realistic models for lightning strikes. One of the objectives is to implement statistical methods on very large (and very sparse) datasets of lightning strikes at the national level (or more reasonably at the level of the Alps) since 2012.

- *ANR* – ChaMaNe – 2020/25 – PI : Delphine Salort

  The aim is to create a research group to make significant advances in the field of mathematics for neuroscience. Some key areas of research are: the intrinsic dynamics of a neuron, the dynamics that emerge from large neural networks taking interactions between neurons, memory effects, and spatial structure into account. This research group gathered experts in PDEs, probability and statistics.

- *IRS IDEX* – GraTweet – 2017/19 – PI : Vincent Brault

  This project focuses on inferring the connectivity graph, and more specifically the underlying communities, in the Twitter social network. The combined use of Hawkes processes and stochastic block models allows for the temporal dynamics of retweets and the community structure of Twitter to be taken into account.

- *PEPS JCJC CNRS* – MaNHawkes – 2017/18 – PI : Julien Chevallier

  This project aims to improve understanding of the Hawkes process model in the context of its application to neuroscience. It has funded visits by Mads Bonde Raad (study of the stability of age-dependent Hawkes processes) and Guilherme Ost (study of fluctuations in Hawkes processes with spatial dependence) which gave the article (Chevallier and Ost, 2020).

D.3   Supervision

- Research project – *CEMRACS* – 2017 (1 month, co-supervised with F. Delarue and E. Tanré)

  Subject : Network of interacting neurons with random synaptic weights. It produced one of the publication listed above.

- **Kevin Polisano** – *Postdoc* – 2018/19 (1 year, co-supervised with E. Gaussier and A. Leclercq-Samson)

  Subject : Joint modeling of tweets and follow for information dynamics on Twitter.

- **Anne-Lise Porté** – *PhD* – 2021/2022(co-supervised with J.-F. Coeurjolly)

  Subject : Point processes generated by critical points of a latent Gaussian field. Unfortunately Anna-Lise had to stop for medical reasons but we continued and it led to the preprint (Chevallier et al., 2025).

- **Marjolaine Demouge** – *M1 internship* – 2022 (co-supervised with J.-F. Coeurjolly)

  Subject : Continuous testing for spatial point process intensity estimation.

- **Martin Combelles** – *M1 internship* – 2025 (co-supervised with V. Brault)

  Subject : Social media interactions inference via Hawkes processes and Stochastic Block Model.

D.4   Contribution to scientific life

Refereeing

- Reviewer for scientific journals : *Annals of Applied Probability* (4), *Acta Applicandae Mathematicae* (1), *Annales de l'IHP (B) Probabilités et Statistique* (3), *ALEA* (1), *Applied Probability Journals* (1), *ESAIM: Probability and Statistics* (1), *Electronic Communications in Probability* (1), *Electronic Journal of Probability* (2), *Journal of Approximation Theory* (1), *Journal of Mathematical Neuroscience* (2), *Journal of Statistical Physics* (1), *Journal of Theoretical Probability* (1), *Kinetic and Related Models* (2), *Scandinavian Journal of Statistics* (1), *Stochastic Processes and their Applications* (6), *Statistics and Probability Letters* (2), *Transactions on Modeling and Computer Simulation* (1).

- Assistant professor selection committees (Université de Strasbourg - 2020, ENSIMAG - 2022).

Conference organization

- Co-organisation of MathEnJeans 2019 in Grenoble (March 2019).

- Co-organisation of the conference BioHasard 2021 (online due to Covid-19).
- Co-organisation of the conference SSIAB15 (May 2023).
- Co-organisation of the conference GeoSto25 (June 2025).

Local animation

- Co-organisation of weekly seminar of DATA departement in LJK (2017 - 2022).
- Leader of the SVH research team, 13 permanent members (2022 - 2025).

BIBLIOGRAPHY

R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003. ISBN 0-12-044143-8.

R. J. Adler and J. E. Taylor. *Random fields and geometry*. Springer Science & Business Media, 2009.

Z. Agathe-Nerine. Multivariate Hawkes processes on inhomogeneous random graphs. *Stochastic Processes and their Applications*, 152:86–148, 2022.

Z. Agathe-Nerine. Long-term stability of interacting Hawkes processes on random graphs. *Electronic Journal of Probability*, 28:1–42, 2023.

Z. Agathe-Nerine. Stability of wandering bumps for Hawkes processes interacting on the circle. *Stochastic Processes and their Applications*, 182:104577, 2025.

S.-i. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2): 77–87, 1977. doi: 10.1007/BF00337259.

J.-M. Azaïs and C. Delmas. Mean number and correlation function of critical points of isotropic Gaussian fields and some results on GOE random matrices. *Stochastic Processes and their Applications*, 150:411–445, 2022. ISSN 0304-4149. doi: 10.1016/j.spa.2022.04.013.

J.-M. Azaïs and M. Wschebor. *Level sets and extrema of random processes and fields*. John Wiley & Sons, 2009.

J.-M. Azaïs, F. Dalmao, and C. Delmas. Multivariate CLT for critical points. *arXiv preprint arXiv:2401.09117*, 2024.

F. Baccelli, F. Karpelevich, M. Y. Kelbert, A. Puhalskii, A. Rybko, and Y. M. Suhov. A mean-field limit for a class of queueing networks. *Journal of statistical physics*, 66(3):803–825, 1992.

E. Bacry, K. Dayri, and J.-F. Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B-Condensed Matter and Complex Systems*, 85 (5):1–12, 2012. doi: 10.1140/epjb/e2012-21005-8.

E. Bacry, M. Bompaire, S. Gaïffas, and J.-F. Muzy. Sparse and low-rank multivariate Hawkes processes. *Journal of Machine Learning Research*, 21(50):1–32, 2020.

A. Baddeley, E. Rubak, and R. Turner. *Spatial point patterns: methodology and applications with R*. CRC press, 2015.

P. Bastian, T. Kutta, R. Basu, and H. Dette. Monitoring time series for relevant changes. *arXiv preprint arXiv:2509.01756*, 2025.

D. Beliaev, V. Cammarota, and I. Wigman. No repulsion between critical points for planar Gaussian random fields. *Electronic Communications in Probability*, 25:1–13, 2020.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 57(1):289–300, 1995. ISSN 0035-9246. doi: 10.1111/j.2517-6161.1995.tb02031.x.

M. V. Berry. Statistics of nodal lines and points in chaotic quantum billiards: perimeter corrections, fluctuations, curvature. *Journal of Physics A: Mathematical and General*, 35(13):3025, 2002.

C. Berzin, A. Latour, and J. León. Kac-Rice formula: a contemporary overview of the main results and applications. *arXiv preprint arXiv:2205.08742*, 2022.

P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.

S. Blanes, F. Casas, and A. Murua. Splitting methods for differential equations. *Acta Numerica*, 33:1–161, 2024. doi: 10.1017/S0962492923000077.

D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.

P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Annals of Probability*, 24(3):1563–1588, 1996. ISSN 0091-1798. doi: 10.1214/aop.

Y. Brenier, C. De Lellis, and L. Székelyhidi. Weak-strong uniqueness for measure-valued solutions. *Communications in Mathematical Physics*, 305(2):351, 2011. ISSN 1432-0916. doi: 10.1007/s00220-011-1267-0.

P. C. Bressloff. Spatiotemporal dynamics of continuum neural fields. *Journal of Physics A: Mathematical and Theoretical*, 45(3):033001, 2011. doi: 10.1088/1751-8113/45/3/033001.

N. Brunel and V. Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural computation*, 11(7):1621–1671, 1999. doi: 10.1162/089976699300016179.

M. J. Cáceres and B. Perthame. Beyond blow-up in excitatory integrate and fire neuronal networks: refractory period and spontaneous activity. *Journal of theoretical biology*, 350:81–89, 2014. doi: 10.1016/j.jtbi.2014.02.005.

M. J. Cáceres, J. A. Carrillo, and B. Perthame. Analysis of nonlinear noisy integrate & fire neuron models: blow-up and steady states. *Journal of Mathematical Neuroscience*, 1:Art. 7, 33, 2011. ISSN 2190-8567. doi: 10.1186/2190-8567-1-7.

M. J. Cáceres, J. A. Cañizo, and N. Torres. Comparison principles and asymptotic behavior of delayed age-structured neuron models. *Acta Applicandae Mathematicae*, 199(1):10, 2025.

Y. Canzani and P. Sarnak. Topology and nesting of the zero set components of monochromatic random waves. *Communications on Pure and Applied Mathematics*, 72(2):343–374, 2019.

R. Carmona, F. Delarue, et al. *Probabilistic theory of mean field games with applications I-II*, volume 3. Springer, 2018.

D. Cheng and A. Schwartzman. Expected number and height distribution of critical points of smooth isotropic Gaussian random fields. *Bernoulli*, 24(4B):3422, 2018.

J. Chevallier. Fluctuations for mean-field interacting age-dependent Hawkes processes. *Electronic Journal of Probability*, 22:1–49, 2017a. ISSN 1083-6489. doi: 10.1214/17-EJP63.

J. Chevallier. Mean-field limit of generalized Hawkes processes. *Stochastic Processes and their Applications*, 2017b. doi: 10.1016/j.spa.2017.02.012.

J. Chevallier. Stimulus sensitivity of a spiking neural network model. *Journal of Statistical Physics*, pages 1–9, 2017c. doi: 10.1007/s10955-017-1948-y.

J. Chevallier. Uniform decomposition of probability measures: quantization, clustering and rate of convergence. *Journal of Applied Probability*, 55(4):1037–1045, 2018. doi: 10.1017/jpr.2018.69.

J. Chevallier. Uniform in time modulus of continuity of Brownian motion. *arXiv preprint arXiv:2312.15931*, 2023.

J. Chevallier. MeanFieldGraph.jl, 2024.

J. Chevallier and T. Laloë. Detection of dependence patterns with delay. *Biometrical Journal*, 57(6):1110–1130, 2015. ISSN 1521-4036. doi: 10.1002/bimj.201400235.

J. Chevallier and G. Ost. Fluctuations for spatially extended Hawkes processes. *Stochastic Processes and their Applications*, 130(9):5510–5542, 2020. ISSN 0304-4149. doi: https://doi.org/10.1016/j.spa.2020.03.015.

J. Chevallier and G. Ost. Community detection for binary graphical models in high dimension. *arXiv preprint arXiv:2411.15627*, 2024.

J. Chevallier, M. J. Cáceres, M. Doumic, and P. Reynaud-Bouret. Microscopic approach of a time elapsed neural model. *Mathematical Models and Methods in Applied Sciences*, 25(14):2669–2719, 2015. doi: 10.1142/S021820251550058X.

J. Chevallier, A. Duarte, E. Löcherbach, and G. Ost. Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. *Stochastic Processes and their Applications*, 2018. doi: https://doi.org/10.1016/j.spa.2018.02.007.

J. Chevallier, A. Melnykova, and I. Tubikanec. Diffusion approximation of multi-class Hawkes processes: Theoretical and numerical analysis. *Advances in Applied Probability*, 53(3):716–756, 2021. doi: 10.1017/apr.2020.73.

J. Chevallier, E. Löcherbach, and G. Ost. Inferring the dependence graph density of binary graphical models in high dimension. *arXiv preprint arXiv:2406.07066*, 2024.

J. Chevallier, J.-F. Coeurjolly, and R. Waagepetersen. Critical point processes obtained from a Gaussian random field with a view towards statistics. *arXiv preprint arXiv:2507.04753*, 2025.

S. Coombes and C. Laing. Pulsating fronts in periodically modulated neural field models. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):011912, 2011.

N. Da Costa, M. Pförtner, L. Da Costa, and P. Hennig. Sample path regularity of gaussian processes from the covariance kernel. *arXiv preprint arXiv:2312.14886*, 2023.

D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2003. ISBN 0-387-95541-0. Elementary theory and methods.

D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. II.* Probability and its Applications (New York). Springer, New York, second edition, 2008. ISBN 978-0-387-21337-8. doi: 10.1007/978-0-387-49835-5. General theory and structure.

G. Dalle, J. Chevallier, and J. Kling. PointProcesses.jl, 2021.

A. Dassios and H. Zhao. Exact simulation of Hawkes process with exponentially decaying intensity. *Electronic Communications in Probability*, 2013.

E. De Santis and M. Piccioni. One-dimensional infinite memory imitation models with noise. *Journal of Statistical Physics*, 161(2):346–364, 2015.

F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanré. Global solvability of a networked integrate-and-fire model of McKean–Vlasov type. *The Annals of Applied Probability*, 25(4):2096–2133, 2015. doi: 10.1214/14-AAP1044.

S. Delattre and N. Fournier. Statistical inference versus mean field limit for Hawkes processes. *Electronic Journal of Statistics*, 10(1):1223–1295, 2016. doi: 10.1214/16-EJS1142.

S. Delattre, N. Fournier, and M. Hoffmann. Hawkes processes on large networks. *Annals of Applied Probability*, 26(1):216–261, 02 2016. doi: 10.1214/14-AAP1089.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22, 1977.

V. Dheur, T. Bosser, R. Izbicki, and S. Ben Taieb. Distribution-free conformal joint prediction regions for neural marked temporal point processes. *Machine Learning*, 113(9):7055–7102, 2024.

V. Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 70(1):245–264, 2008. doi: 10.1111/j.1467-9868.2007.00634.x.

S. Ditlevsen and E. Löcherbach. Multi-class oscillating systems of interacting neurons. *Stochastic Processes and their Applications*, 127(6):1840–1869, 2017. doi: 10.1016/j.spa.2016.09.013.

S. Donnet, V. Rivoirard, and J. Rousseau. Nonparametric Bayesian estimation for multivariate Hawkes processes. *The Annals of Statistics*, 48(5):2698–2727, 2020. doi: 10.1214/19-AOS1903.

M. D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 6:12, 1951. ISSN 0065-9266,1947-6221.

A. Duarte, A. Galves, E. Löcherbach, and G. Ost. Estimating the interaction graph of stochastic neural dynamics. *Bernoulli*, 25(1):771–792, 2019.

G. Dumont, A. Payeur, and A. Longtin. A stochastic-field description of finite-size spiking neural networks. *PLOS Computational Biology*, 13(8):e1005691, 2017.

R. Engelken. Sparseprop: Efficient event-based simulation and training of sparse recurrent spiking neural networks. *Advances in Neural Information Processing Systems*, 36:3638–3657, 2023.

B. Ermentrout. Neural networks as spatio-temporal pattern-forming systems. *Reports on progress in physics*, 61 (4):353, 1998. doi: 10.1088/0034-4885/61/4/002.

G. B. Ermentrout and J. B. McLeod. Existence and uniqueness of travelling waves for a neural network. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 123(3):461–478, 1993.

A. Estrade and J. R. León. A central limit theorem for the Euler characteristic of a Gaussian excursion set. *Annals of Probability*, 44(6):3849–3878, 2016. doi: 10.1214/15-AOP1062.

S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. ISBN 0-471-08186-8. doi: 10.1002/9780470316658. Characterization and convergence.

O. Faugeras and J. Inglis. Stochastic neural field equations: a rigorous footing. *Journal of mathematical biology*, 71(2):259–300, 2015. doi: 10.1007/s00285-014-0807-6.

B. Fernandez and S. Méléard. A Hilbertian approach for fluctuations on the McKean-Vlasov model. *Stochastic Processes and their Applications*, 71(1):33–53, 1997. ISSN 0304-4149. doi: 10.1016/S0304-4149(97)00067-7.

M. Fischer and G. Nappo. On the moments of the modulus of continuity of Itô processes. *Stochastic Analysis and Applications*, 28(1):103–122, 2009. doi: 10.1080/07362990903415825.

S. Forest, J.-C. Quinton, and M. Lefort. A dynamic neural field model of multimodal merging: application to the ventriloquist effect. *Neural Computation*, 34(8):1701–1726, 2022.

R. Forien, G. Pang, E. Pardoux, and A.-B. Zotsa-Ngoufack. Stochastic epidemic models with varying infectivity and waning immunity. *The Annals of Applied Probability*, 35(3):2175–2216, 2025.

N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, pages 1–32, 2014. doi: 10.1007/s00440-014-0583-7.

A. M. Garsia, E. Rodemich, H. Rumsey, and M. Rosenblatt. A real variable lemma and the continuity of paths of some Gaussian processes. *Indiana University Mathematics Journal*, 20(6):565–578, 1970. doi: 10.1512/iumj. 1971.20.20046.

L. Gass and M. Stecconi. The number of critical points of a Gaussian field: finiteness of moments. *Probability Theory and Related Fields*, pages 1–31, 2024.

G. P. Gava, L. Lefèvre, T. Broadbelt, S. B. McHugh, V. Lopes-dos Santos, D. Brizee, K. Hartwich, H. Sjoberg, P. V. Perestenko, R. Toth, et al. Organizing the coactivity structure of the hippocampus from robust to flexible memory. *Science*, 385(6713):1120–1127, 2024.

H. Ge, K. Xu, and Z. Ghahramani. Turing.jl, 2018.

W. Gerstner and W. M. Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002.

W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

R. D. Gill, N. Keiding, and P. K. Andersen. *Statistical models based on counting processes*. Springer, 1997.

C. Graham. Chaoticity for multiclass systems and exchangeability within classes. *Journal of Applied Probability*, 45(4):1196–1203, 2008. doi: 10.1239/jap/1231340243.

S. Grün. *Unitary joint events in multiple neuron spiking activity: detection, significance, and interpretation.* PhD thesis, Ruhr-Universitat Bochum, 1996.

S. Grün, M. Diesmann, and A. Aertsen. Unitary Events in Multiple Single-Neuron Spiking Activity: I. Detection and Significance. *Neural Computation*, 14(1):43–80, 2002. doi: 10.1162/089976602753284455.

G. Gusto and S. Schbath. FADO: A Statistical Method to Detect Favored or Avoided Distances between Occurrences of Motifs using the Hawkes' Model. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 2005. doi: 10.2202/1544-6115.1119.

E. Hansen and G. W. Walster. *Global optimization using interval analysis: revised and expanded*, volume 264. CRC press, 2003.

A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971. doi: 10.1093/biomet/58.1.83.

S. Heesen and W. Stannat. Fluctuation limits for mean-field interacting nonlinear Hawkes processes. *Stochastic Processes and their Applications*, 139:280–297, 2021.

Y. Hu and K. Le. A multiparameter Garsia–Rodemich–Rumsey inequality and some applications. *Stochastic Processes and their Applications*, 123(9):3359–3377, 2013. ISSN 0304-4149.

S. Jaffard. *Spiking neural networks : learning as point processes.* PhD thesis, Université Côté d'Azue, 2025.

B. Jourdain and S. Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 34(6):727–766, 1998. ISSN 0246-0203. doi: 10.1016/S0246-0203(99)80002-8.

M. Kac. On the average number of real roots of a random algebraic equation. *Bulletin of the American Mathematical Society*, 49:314–320, 1943.

B. L. Keyfitz and N. Keyfitz. The McKendrick partial differential equation and its uses in epidemiology and population study. *Mathematical and Computer Modelling*, 26(6):1–9, 1997. doi: 10.1016/S0895-7177(97)00165-9.

K. Kishimoto and S.-i. Amari. Existence and stability of local excitations in homogeneous neural fields. *Journal of Mathematical Biology*, 7:303–318, 1979.

J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32(1):111–131, 1975.

M. F. Kratz and J. R. León. On the second moment of the number of crossings by a stationary Gaussian process. *The Annals of Probability*, 34(4):1601–1607, 2006.

J. Kruger and W. Stannat. Front propagation in stochastic neural fields: a rigorous mathematical framework. *SIAM Journal on Applied Dynamical Systems*, 13(3):1293–1310, 2014. doi: 10.1137/13095094X.

T. G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*, 6(3):223–240, 1978. doi: 10.1016/0304-4149(78)90020-0.

S. Ladgham and R. Lachièze-Rey. Local repulsion of planar Gaussian critical points. *Stochastic Processes and their Applications*, 166:104221, 2023. doi: 10.1016/j.spa.2023.09.008.

R. C. Lambert, C. Tuleau-Malot, T. Bessaih, V. Rivoirard, Y. Bouret, N. Leresche, and P. Reynaud-Bouret. Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *Journal of neuroscience methods*, 297:9–21, 2018.

J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007. doi: 10.1007/s11537-007-0657-8.

S.-H. Lee, R. Blake, and D. J. Heeger. Traveling waves of activity in primary visual cortex during binocular rivalry. *Nature neuroscience*, 8(1):22–23, 2005.

P. Lévy. Théorie de l'addition des variables aléatoires. *Gauthier-Villars, Paris*, 1937.

E. Luçon and C. Poquet. Neural Field Equations and Hawkes processes: long-term stability of traveling wave profiles in the neutral case. *arXiv preprint arXiv:2507.19236*, 2025.

E. Luçon and W. Stannat. Mean field limit for disordered diffusions with singular interactions. *The Annals of Applied Probability*, 24(5):1946–1993, 2014. doi: 10.1214/13-AAP968.

E. Luçon and W. Stannat. Transition from Gaussian to non-Gaussian fluctuations for mean-field diffusions in spatial interaction. *The Annals of Applied Probability*, 26(6):3840–3909, 2016. doi: 10.1214/16-AAP1194.

M. Magris. On the simulation of the Hawkes process via Lambert-W functions. *arXiv preprint arXiv:1907.09162*, 2019.

A. G. McKendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, 1925. doi: 10.1017/S0013091500034428.

S. Mischler, C. Quiñinao, and Q. Weng. Weak and strong connectivity regimes for a general time elapsed neuron network model. *Journal of Statistical Physics*, 173:77–98, 2018.

I. Mitoma. Tightness of probabilities on $\mathcal{C}([0,1];\mathcal{S}')$ and $\mathcal{D}([0,1];\mathcal{S}')$. *The Annals of Probability*, pages 989–999, 1983.

J. Møller and J. G. Rasmussen. Perfect simulation of Hawkes processes. *Advances in Applied Probability*, 37(3): 629–646, 2005. ISSN 0001-8678. doi: 10.1239/aap.

P. Mozgunov, M. Beccuti, A. Horvath, T. Jaki, R. Sirovich, and E. Bibbona. A review of the deterministic and diffusion approximations for stochastic chemical reaction networks. *Reaction Kinetics, Mechanisms and Catalysis*, 123(2):289–312, 2018.

L. R. Mujica-Parodi, C. Organtzidis, A. Pathak, D. Hofmann, M. Protter, A. G. Chesbro, A. Driussi, G. Abrevaya, A. Botond, C. Rackauckas, R. Granger, E. K. Miller, A. Edelman, and H. H. Strey. Neuroblox.jl, 2024.

Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998. doi: 10.1023/A:1003403601725.

G. Ost and P. Reynaud-Bouret. Sparse space-time models: Concentration inequalities and Lasso. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(4):2377–2405, 2020. doi: 10.1214/19-AIHP1042.

K. Pakdaman, B. Perthame, and D. Salort. Dynamics of a structured neuron population. *Nonlinearity*, 23(1):55, 2010. doi: 10.1088/0951-7715/23/1/003.

K. Pakdaman, B. Perthame, and D. Salort. Relaxation and self-sustained oscillations in the time elapsed neuron network model. *SIAM Journal on Applied Mathematics*, 73(3):1260–1279, 2013. doi: 10.1137/110847962.

G. Peccati and M. S. Taqqu. *Wiener Chaos: Moments, Cumulants and Diagrams: A survey with computer implementation*, volume 1. Springer Science & Business Media, 2011.

B. Perthame. *Transport equations in biology*. Springer Science & Business Media, 2006. doi: 10.1007/978-3-7643-7842-4.

J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008. doi: 10.1038/nature07140.

C. C. Pinto. Analysis of firing rate and correlation of spike trains from a brain network model with clusters. Master's thesis, Universidade Federal do Rio de Janeiro, 2025.

A. Prodhomme. Strong Gaussian approximation of metastable density-dependent Markov chains on large time scales. *Stochastic Processes and their Applications*, 160:218–264, 2023.

A. Quaresima. SpikingNeuralNetworks.jl, 2019.

C. Quiñinao. A microscopic spiking neuronal network for the age-structured model. *Acta Applicandae Mathematicae*, 146:29–55, 2016. doi: 10.1007/s10440-016-0056-3.

R. Q. Quiroga, T. Kreuz, and P. Grassberger. Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Physical review E*, 66(4):041904, 2002.

C. Rackauckas and Q. Nie. DiffEqNoiseProcess.jl, 2017.

A. Renart, N. Brunel, and X.-J. Wang. Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks. In J. Feng, editor, *Computational Neuroscience: A comprehensive approach.* Chapman & Hall/CRC Mathematical Biology and Medicine Series, 2004.

D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion (Grundlehren der mathematischen Wissenschaften).* Springer-Verlag, 3rd edition, 1999. ISBN 3540643257.

P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Inference of functional connectivity in Neurosciences via Hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*, pages 317–320, 2013. doi: 10.1109/GlobalSIP.2013.6736879.

S. O. Rice. Mathematical analysis of random noise. *The Bell System Technical Journal*, 23(3):282–332, 1944.

E. Rio. Local invariance principles and their application to density estimation. *Probability Theory and Related Fields*, 98(1):21–45, 1994. doi: 10.1007/BF01311347.

D. P. Sanders and L. Benet. IntervalArithmetic.jl, 2014.

V. Schmutz. Mean-field limit of age and leaky memory dependent Hawkes processes. *Stochastic Processes and their Applications*, 149:39–59, 2022.

B. Seeger. Error estimates for deterministic empirical approximations of probability measures. *arXiv preprint arXiv:2510.03451*, 2025.

M. Sepúlveda, N. Torres, and L. M. Villada. Well-posedness and numerical analysis of an elapsed time model with strongly coupled neural networks. *Communications in Nonlinear Science and Numerical Simulation*, 152:109144, 2026. ISSN 1007-5704.

M. L. Stein. *Interpolation of spatial data: some theory for kriging.* Springer Science & Business Media, 2012.

D. Sulem, V. Rivoirard, and J. Rousseau. Bayesian estimation of nonlinear Hawkes processes. *Bernoulli*, 30(2): 1257–1286, 2024.

A.-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991. doi: 10.1007/BFb0085169.

M. Talagrand. *Upper and lower bounds for stochastic processes*, volume 60. Springer, 2014.

N. Torres, M. J. Cáceres, B. Perthame, and D. Salort. An elapsed time model for strongly coupled inhibitory and excitatory neural networks. *Physica D: Nonlinear Phenomena*, 425:132977, 2021.

C. Tuleau-Malot, A. Rouis, F. Grammont, and P. Reynaud-Bouret. Multiple Tests Based on a Gaussian Approximation of the Unitary Events Method with delayed coincidence count. *Neural Computation*, 26:7, 2014.

R. Veltz and O. Faugeras. Local/global analysis of the stationary solutions of some neural field equations. *SIAM Journal on Applied Dynamical Systems*, 9(3):954–998, 2010.

H. von Foerster. Some remarks on changing populations. *The kinetics of cellular proliferation*, pages 382–407, 1959.

R. Waagepetersen and Y. Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(3):685–702, 2009.

X. Wang, M. Kolar, and A. Shojaie. Statistical inference for networks of high-dimensional point processes. *Journal of the American Statistical Association*, 120(550):1014–1024, 2025.

N. F. Weeden. Are Mendel's data reliable? The perspective of a pea geneticist. *Journal of Heredity*, 107(7): 635–646, 2016.

P. Weiss. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *J. Phys. Theor. Appl.*, 6(1):661–690, 1907.

H. R. Wilson and J. D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1, 1972. doi: 10.1016/S0006-3495(72)86068-5.

H. R. Wilson and J. D. Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2):55–80, 1973.

J. Wilting and V. Priesemann. Inferring collective dynamical states from widely unobserved systems. *Nature communications*, 9(1):2325, 2018.

S. Yang and H. Zha. A Variational Autoencoder for Neural Temporal Point Processes with Dynamic Latent Graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(15):16343–16351, 2024.

L. Zhang. On stability of traveling wave solutions in synaptically coupled neuronal networks. *Differential and Integral Equations*, 16(5):513–536, 2003.

A. B. Zotsa Ngoufack. Functional central limit theorems for epidemic models with varying infectivity and waning immunity. *ESAIM: Probability and Statistics*, 29:45–112, 2025.