Wavelet-based representations of point processes for modelling and statistical learning

Antoine Brochard

To cite this version:

Antoine Brochard. Wavelet-based representations of point processes for modelling and statistical learning. Probability [math.PR]. Ecole Normale Supérieure (Paris), 2022. English. NNT : . tel-03666508

HAL Id: tel-03666508
https://theses.hal.science/tel-03666508v1

Submitted on 12 May 2022

PSL
UNIVERSITÉ PARIS

Inria

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Inria / Ecole Normale Spérieure
Dans le cadre d'une cotutelle avec Huawei Technologies France

Wavelet-based representations of point processes for modelling and statistical learning

Soutenue par
Antoine Brochard
March 2022

École doctorale n°386
Sciences Mathématiques de Paris Centre

Spécialité
Mathématiques appliquées

Composition du jury :

Frédéric Lavancier Maître de conférences, Université de Nantes	*Rapporteur*
Kai Schneider Professeur, Université Aix-Marseille	*Rapporteur*
Djalil Chafaï Professeur, École Normale Supérieure de Paris	*Examinateur*
Agnès Desolneux Directrice de recherche, CNRS	*Présidente du jury*
Brice Ménard Professeur, Johns Hopkins University	*Examinateur*
Bartłomiej Błaszczyszyn Directeur de recherche, Inria	*Directeur de thèse*
Stéphane Mallat Professeur, Collège de France	*Co-directeur de thèse*

ENS | PSL

Je tiens à remercier tout particulièrement Sixin Zhang, qui a joué le rôle de co-directeur officieux durant mon odctorat. Il a fait preuve à mon égard d'une infinie patience, d'un soutien sans faille, et de sagesse en toute circonstance.

Je suis également reconnaissant envers mes deux directeurs de thèse, Bartek Błaszczyszyn et Stéphane Mallat, pour leurs encouragements, leur enthousiasme, leur expérience, leur rigueur, et leur curiosité. Ils ont fait de mes années de doctorat une expérience plus riche que je ne pouvais l'imaginer.

Je remercie vivement les membres du jury, et particulièrement les rapporteurs, Frédéric Lavancier et Kai Schneider, pour le temps et l'énergie qu'ils ont consacrés à la lecture de mes travaux, l'intérêt qu'ils y ont porté, et leurs judicieuses remarques. Merci à Brice Ménard pour la confiance qu'il m'a accordé, et son excellent pain au maîs.

Merci à tous les collègues de l'équipe Data: Florentin ma nounou du bureau, et ses claquements de dents dans la cour en hiver pour ne pas me laisser seul, Rudy et ses blagues sur ma calvitie naissante, Gaspar et ses conseils en menuiserie-joaillerie, Simon et son analyse ciselée de mes tenues vestimentaires, John et sa croisade pour la fresque du CSD, et maintenant Etienne, qui m'a montré qu'il y a même des poissons rouges qui jouent mieux au foot que moi. Les anciens également, Tomás, Louis, et Tanguy.

Je remercie Marc Hoffmann, pour m'avoir soutenu et guidé dans des temps difficiles.

Par dessus tout, je tiens à remercier ma famille, pour n'avoir jamais perdu espoir; et Charlotte, pour m'avoir montré qu'au fond, le plus important, c'est les pâtes à la sauce tomate.

Contents

Chapter 1

Introduction

1.1 Context and motivations

This dissertation discusses the study of different representations of random geometric objects. An elementary class of objects described by their geometry consists of point patterns, i.e. collections of points living in some metric space. In practical applications, these points usually represent the temporal or spatial location of events, or entities of interest. Without formalism, a point pattern is represented by a sum of Dirac atoms, i.e.

$$\phi = \sum_{i=1}^{N} \delta_{u_i},$$

where $N \in \mathbb{N} \cup \{\infty\}$, and the u_i represent the locations of these entities. When point patterns are considered to be observations of a random phenomenon, this phenomenon is called a point process. Point processes arise in a wide variety of contexts. Time-related processes can represent, for example, the recorded times of financial orders, the spiking times of a neuron, or the times of arrival of customers on a website. For spatial point processes, one can think of the positions of trees in forests, cells in a biological tissue, individuals in a given population, or stars, planets, galaxies in space. One can also study spatio-temporal objects such as the time *and* position of earthquakes in a given area, or the dynamics of droplets in a cloud. Depending on the phenomenon being observed, the points, or *atoms* of the process, can tend to repel each other, forming rather regular structures (for instance, trees in a forest tend to grow away from each other, in order to have the most available nutriments in the ground, see e.g. [105]). Conversely, they can also regroup together, forming *clusters* of points of possibly complex geometric structures (for example, galaxies in the universe can tend to group together, and form filamentary structures, see e.g. [107]).

Being able to understand, and adequately describe these objects is thus of central interest in many areas of science, whether it is to perform prediction of future events, estimation of biological or physical parameters, or modelling. Recent advances in statistical learning, and particularly deep learning, have led to impressive progress in the representation and manipulation of complex objects, yielding great performances on a wide range of tasks, such as classification, regression, segmentation, or generation. These methods are able to handle a vast variety of complex objects (e.g. images, audio signals, languages), including point processes (e.g. [93, 117]). However, representations defined by (deep) neural networks are often referred to as 'black boxes', as we do not fully understand what aspects of the geometry these networks learn, which make these representation lack interpretability, and possibly reliability. On the other hand, we find in the literature on point processes many different mathematical representations, called *summary characteristics*, that each capture a specific aspect of the geometry of point patterns. These characteristics usually rely either on second-order statistics (e.g. the pair correlation function, Ripley's K-function [96], or Besag's L-function [13]), or on distances between neighbouring points (such as the spherical contact distribution function or the nearest neighbours distance distribution function). None of them fully capture the entire range of geometric structures that point processes may exhibit, and it can be challenging, depending on the data that one is dealing with, and the task at hand, to find a suitable representation.

Finding concise, easily manipulable representations of complex data is also an active area of research and a central interest in the field of signal processing, where many tools have been developed (see e.g. Mallat [79]) to extract important information about the data with a relatively 'compact' (or low-dimensional) representation. A useful way to build such a representation is to leverage the possible regularities or invariances present in the data, for which a fundamental tool is the Fourier analysis. The Fourier transform builds a representation of signals based on their frequency, i.e. the repetition of patterns in time or space, with complex sinusoidal functions. However, these oscillating functions are 'delocalized' in space, so they may no be adapted to capture sudden changes, or sharp local transitions in the data, along the time or space axis. To capture such local structures in signals, one can instead compute their correlations with functions that are 'localized' both in space (or time) and frequency, such as wavelets. Wavelet analysis has been extensively studied in the literature, for the processing of many types of signals and stochastic processes, including point processes (see e.g. [1, 78], in the context of one-dimensional point processes exhibiting fractal behaviour). In this work, we shall present different wavelet-based representations of two-dimensional stationary point processes that, similarly to classical summary characteristics, are designed to capture and compactly represent the geometric structures formed by the data. We study their quality as statistical descriptors through several learning and modelling tasks.

1.2 Related works and our contributions

We give an overview of the problems considered in this dissertation, and the current methods that have been developed to address these problems. Relatively to these works, we highlight our contributions, that we shall detail in the following chapters. More specific related works can be found in the introduction of each respective chapter.

1.2.1 Fourier analysis and wavelet-based representations of point processes

A fundamental theory for the analysis of stationary stochastic processes is the Fourier analysis. It is particularly useful, for instance, to study second-order moments, as the covariance matrix of a stationary process is diagonal in the Fourier basis. In the context of point processes, this has been formalized by the covariance measure in the Fourier domain, called the Bartlett spectrum (Bartlett [9], [10, 27]), and can be estimated through the Discrete Fourier Transform of observation samples, as we show in Section 2.8.2 of this work. Figure 1.1 illustrates this estimation on a distribution for which the theoretical value of the Bartlett spectrum is known.

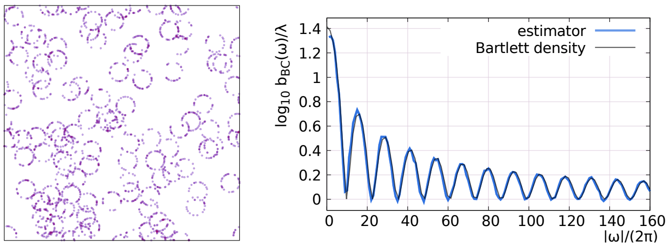


Figure 1.1: Left: sample from a Cox-cirles point process (see Example 2.8.4). Right: plot of the (theoretical) Bartlett spectrum density, and its estimation using the Discrete Fourier Transform.

In order to extract meaningful information about the structure of signals beyond second-order properties, [80] introduce the scattering moments, which compute a cascade of convolutions of the input signal with wavelets, followed by a complex modulus operator. Informally (see Section 3.2 for a more precise definition), given a zero-mean complex function ψ, a wavelet family can be constructed by dilating and rotating ψ, i.e.

$$\psi_{j,\theta}(u) = 2^{-2j}\psi(2^{-j}r_\theta u).$$

The scattering moments define a representation of an input signal ϕ, observed in a finite window W, by computing the values

$$S_1\phi(j,\theta) = \frac{1}{|W|}\int_W |\phi \star \psi_{j,\theta}|(u)du$$
$$S_2\phi(j_1,\theta_1,j_2,\theta_2) = \frac{1}{|W|}\int_W ||\phi \star \psi_{j_1,\theta_1}| \star \psi_{j_2,\theta_2}|(u)du,$$

where $\star$ denotes the convolution operator. If Φ is a stationary point process, then the scattering moments of its distribution are defined in expectation, i.e.

$$\mathbb{E}[S_1\Phi(j,\theta)]$$
$$\mathbb{E}[S_2\Phi(j_1,\theta_1,j_2,\theta_2)].$$

In [33], the authors study the properties of scattering moments of the one-dimensional homogeneous Poisson point process. We extend some of these results, in the two-dimensional case, to a broader class of stationary point processes. In particular, we study the limit behaviour of scattering moments (of first order, see Section 3.2), when the wavelet is asymptotically contracted or dilated. These limits quantify, respectively, the local interactions between points, and their macroscopic growth of variance. In Section 3.3, we show that under some technical conditions:

- When the wavelet $\psi_{j,\theta}$ is asymptotically contracted (i.e. $j \to -\infty$), then $\mathbb{E}[S_1\Phi(j,\theta)]$ is proportional to the *intensity* of the process (i.e. its expected number of atoms in a domain of unit volume). We show that this convergence is controlled by a quantity that depends on the second order moments of Φ.

- When the wavelet $\psi_{j,\theta}$ is asymptotically dilated ($j \to \infty$), the limit behaviour of $2^j\mathbb{E}[S_1\Phi(j,\theta)]$ is determined by the growth of variance of the number of atoms of Φ in domains of growing volume.

This result, illustrated in Figure 1.2 is also extended to a certain class of *marked* point process, where to each atom of a point process is attached some numerical value. Furthermore, we give an expansion formula for this class of marked point processes, in the spirit of [14, 18].

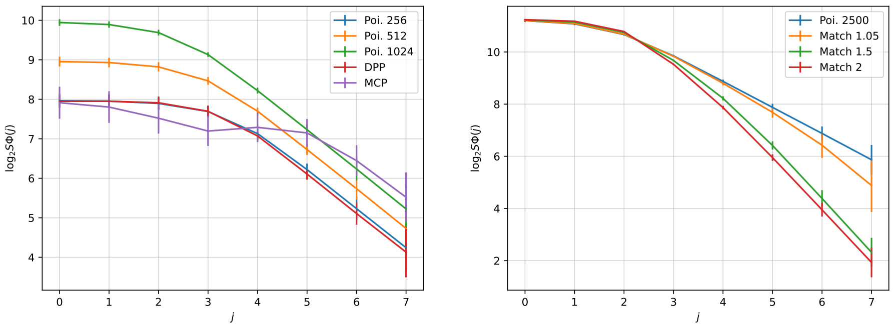


Figure 1.2: Illustration of the asymptotic behaviour of first order scattering moments. Left: Poisson processes (PPP) with different intensities, a Gaussian determinantal (DPP), and a Matérn cluster (MCP) process exhibiting the same type of macroscopic variance. Right: a Poisson process, compared to hyperuniform processes (more regular), see Section 3.3.

1.2.2 Estimating geometric marks

Estimating numerical values associated with point patterns based on their geometry is a well known problem that has been studied, for instance, in quantum chemistry to estimate the energy of molecules from the geometry of its atoms (see e.g. [50]). In this work, we shall seek to estimate the values of *marks* of individual points of a given point pattern, with these marks depending on the geometry of the pattern. More precisely, we are interested in measures of the form

$$\tilde{\phi} = \sum_i m(u_i, \phi)\delta_{u_i},$$

where $\phi = \sum_i \delta_{u_i}$ is a point pattern, and m is an unknown marking function. The goal is to estimate m based on a labeled training set. This requires to have a representation not only for the point patterns, but also for the association of the points and their marks. For this problem, we propose the following method: we build a representation $S\phi$ of the unmarked point patterns, and $S'\tilde{\phi}$ for the marked patterns. From the training set, we approximate the relation between $S\phi$ and $S'\tilde{\phi}$:

$$S'\tilde{\phi} \simeq f(S\phi).$$

Then, with the knowledge of ϕ and the approximated $S'\tilde{\phi}$, we use an optimization algorithm to *reconstruct* $\tilde{\phi}$. Our representations S and S' are the *scattering moments* of the point patterns. Figure 1.3 illustrate our proposed method.

1.2.3 Summary characteristics and probabilistic modelling

One of the tasks that we shall study in this dissertation is the modelling of distributions of point processes, using summary characteristics. When dealing with some observation

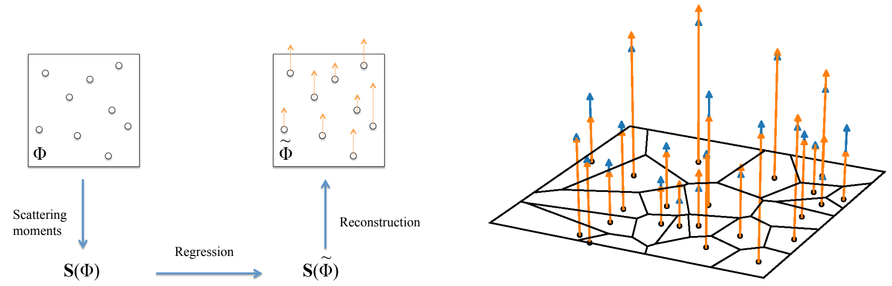


Figure 1.3: Summary of the learning of geometric marks. Left: overview of the method using the scattering moments representation. Right: an example of estimated marks, compared to the exact ones, which depend on the Voronoi cell areas of each atom (see Example 2.6.4).

data, one may find itself in need to model the underlying distribution from this data, and sample from this distribution, without using classical parametric model (e.g. if no classical parametric model fit the data satisfyingly, or if such models are hard to sample from).

Figure 1.4 (top row) shows examples of observations from distributions that we shall seek to approximate. In this situation, a possible way to model the distribution is to compute summary characteristics of the observation data, and to generate new samples that match the statistics of the observation. In more details, given an realization $\bar{\phi}$ of a point process Φ, these methods consist in computing a representation $K(\bar{\phi})$, such that

$$K(\bar{\phi}) \simeq \mathbb{E}[K(\Phi)].$$

If another point pattern ϕ satisfies $K(\phi) \simeq K(\bar{\phi})$, then $K(\phi) \simeq \mathbb{E}[K(\Phi)]$. If K is a good descriptor of the distribution of Φ, then ϕ is likely to be a realization of Φ. Such models, which can be formalized as *maximum entropy models*, also constitute a convenient way to assess wether the summary characteristics capture enough information about the geometry of the data: intuitively, if the generated samples match the statistics that were used to define the model, but do not resemble the observed data, it could indicate that these statistics fail to capture important structural information. The challenges in defining maximum entropy models thus reside in defining a descriptor K that characterizes the distribution of Φ, and finding a sampling method to generate new point patterns which descriptors match the ones of the observation $\bar{\phi}$.

In the context of point processes, the work of Tscheschel and Stoyan [110] studies the use of several classical summary characteristics, such as the k-nearest neighbours distance distribution function, to represent the distribution of point processes. In order to generate new samples, this method relies on random search, by sequentially adding and removing points from a Poisson point process sample. While their solution is efficient to generate samples containing a few hundred points, forming simple geometric structures, it is not adapted to model distributions for which samples are formed by a large number of points, exhibiting possibly long-range interactions. In this work, we propose to represent point process distributions with wavelet phase harmonics covariances. These moments were introduced in [81], and used in [120] to model non-Gaussian stationary processes. They

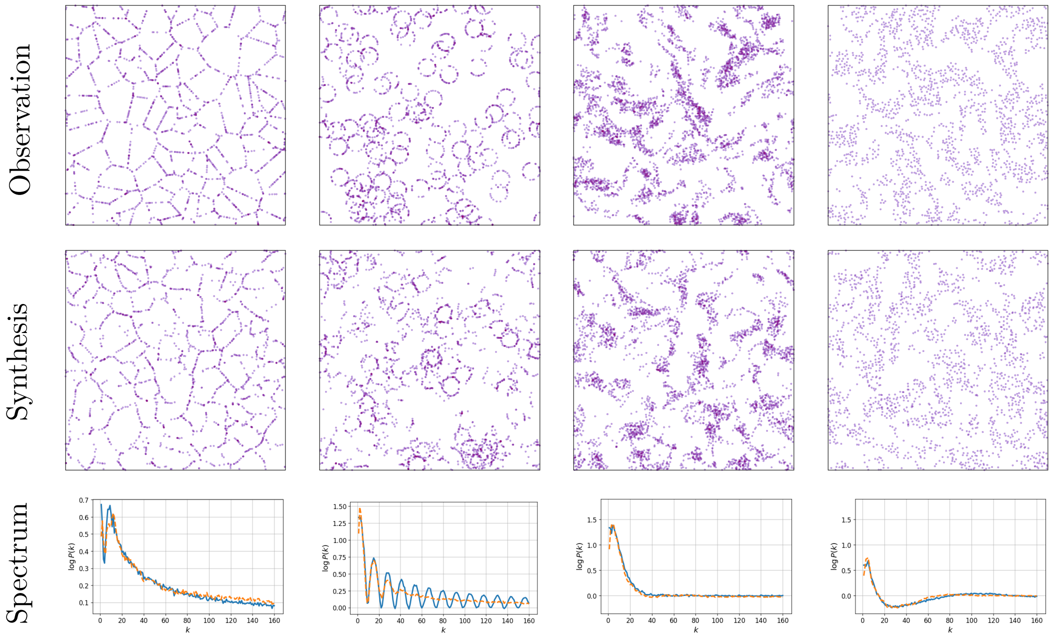


Figure 1.4: Top: samples of several point processes distributions exhibiting interactions at multiple scales. Middle: samples from our model, for each respective distribution. Bottom: Bartlett spectrum density curves for the true (full blue lines) and model (dashed orange lines) distributions.

are defined by computing the correlations between phase adjusted wavelet coefficients. The phase harmonics of a complex number z are defined by

$$[z]^k = |z|e^{ik\varphi(z)},$$

where $|z|$ denotes the complex modulus of z, and $\varphi(z)$ its argument. Wavelet phase harmonics covariances of a point pattern Φ are defined by the correlations between the phase harmonics of the wavelet coefficients of Φ, i.e. moments of the form

$$K(\phi) = \int [\phi \star \psi_{j,\theta}]^k(u)[\phi \star \psi_{j',\theta'}]^{k'}(u-\tau)du.$$

To generate new samples more efficiently than with a random search method, we propose a gradient descent algorithm, that optimizes the positions of the atoms of a sample from a Poisson distribution, in order to minimize the energy

$$E_{\bar{\phi}}(\phi) = \frac{1}{2}|K(\bar{\phi}) - K(\phi)|^2.$$

By noting $\phi = \sum_i \delta_{u_i}$, the positions u_i are updated with

$$u_i \leftarrow u_i - \zeta \nabla_{u_i} E_{\bar{\phi}}(\phi),$$

for some gradient step $\zeta > 0$. Examples of samples from this method are given in Figure 1.4 (middle row), as well as the comparison between the Bartlett densities of the true and model distributions (bottom row). We provide an open-source software implementing the described method.

The overall quality of a model can be defined by its ability to produce patterns similar to the observation, but also with a diversity of samples comparable to that of the true distribution. To evaluate our model with respect to these two aspects, we propose the use of Topological Data Analysis (see e.g. [36]). This method is based on the representation of samples from the true distribution, and samples from the model, by their *persistence diagrams*. Intuitively, persistence diagrams are morphological descriptors, that describe the size of connected components and 'holes' that samples exhibit, when their atoms are replaced by balls of varying size. This representation is then used to compute a notion of 'distance' between the different samples. It allows one to evaluate whether:

- The samples from the model are 'close' to samples from the true distribution.
- The samples from the model are not too close to the single observed sample, and exhibit intra-class distances similar to the ones from the true distribution.

A dimension reduction method is used to approximate this distance matrix with a set of two-dimensional points, in order to visually appreciate the two aforementioned properties of the model.

1.2.4 Modelling texture images

Another type of data defined by complex geometric structures, that can be seen as realizations of a two-dimensional stationary process, are spatially homogeneous images called *textures*. As for two-dimensional point processes, an approximation of the distribution underlying a given texture can be defined by a maximum entropy model, based on a set of statistical constraints. One of the principal evaluation methods of maximum entropy models for textures is visual inspection, as the human eye is often efficient in distinguishing between 'real' and generated samples. Finding a set of statistics that is able to represent and model texture images is a challenging task, due to the high complexity of geometries that they can exhibit, as illustrated in Figure 1.5 (top row).

In order to assess the ability of a given descriptor to capture geometric information, it is

Figure 1.5: Top: examples of gray-scale and color texture images. Bottom: examples of syntheses from our texture model.

thus interesting to turn to this class of processes. In this framework, the pioneer work of Portilla and Simoncelli [92] on wavelet-based representations has led to impressive results. However, recent models relying on information captured by deep convolutional neural networks (CNN) greatly outperform these representations. In this dissertation, we define a wavelet-based model allowing to generate synthesized images of similar visual quality to such CNN-based representations. This model is defined with a representation that, similarly to wavelet phase harmonics covariances, computes the correlations between modified wavelet coefficients: for a real number $\alpha \in [0, 2\pi]$, we denote the generalized rectifier of a complex number z by

$$\rho_\alpha(z) = \max(0, \mathrm{Real}(ze^{i\alpha})).$$

Given a texture image x, we define its representation by the set of coefficients

$$C(x) = \int \rho_\alpha(x \star \psi_{j,\theta})(u)\rho_{\alpha'}(x \star \psi_{j',\theta'})(u-\tau)du,$$

for an appropriate set of indices $(j, \theta, \alpha, j', \theta', \alpha', \tau)$. This set of indices, of coefficients that we choose to correlate, has a crucial impact on the model. We observe that if this set is too small, then the representation may fail to capture important structural information about the image. Conversely, if the index set is too large, the model may memorize parts of the observations, which will be reproduced in the syntheses. We illustrate this trade-off between visual quality and diversity of the syntheses, and define a model that lies between these two extreme cases. Figure 1.5 (bottom row) shows samples of syntheses from our model. We provide an open-source software implementing the described method.

1.3 Outline of the dissertation

In order to present our findings, we first need to briefly review the main notions and results from stochastic geometry on which we base our study. Chapter 2 aims to present our principal object of interest: point processes. More specifically, we shall focus on stationary and ergodic point processes on $\mathbb{R}^2$. As we take interest in representations of such processes, we review classic descriptors, such as second-order moments and other common summary characteristics. We then review maximum entropy models, a particular class of models for point processes that rely on such representations.

This leads us to present in Chapter 3 the family of representations that we study in this work, based on the wavelet transform of counting measures. We shall focus on two types of representations: scattering moments, that are constructed by applying a cascade of wavelet convolutions and modulus operators to the input signal, and phase harmonics covariances, which compute correlations between phase-adjusted wavelet transforms of the signal. We derive theoretical results for the behaviour of scattering moments to gain some insights about the information that they capture.

We can then study the adequacy of these representations in the context of statistical learning or modelling tasks. The first experiment, based on [28], and detailed in Chapter 4, consists in the estimation of *geometric marks* (i.e. numerical values, associated with each atom of a point process, that depend on the geometry formed by the atoms). The marking function is not known, and shall be estimated from a label training set of realizations of a marked point process (i.e. the data consists of several point patterns, and the label is the associated marks). We propose the use of scattering moments as a representation of both point patterns and marked point patterns. We estimate the scattering moments of marked patterns from the scattering moments of non-marked patterns, and then estimate the marks from the moments of the marked pattern. We compare our method with a more direct approach, using raw geometric information captured by local pair-wise distance matrices, and observe that our method is more adapted to recover geometric marks when the marking function has long-range dependencies.

These representations, similarly to classic summary characteristics, describe the geometry of point patterns. As such, they can also characterize aspects of the distribution of a point process, by considering them as empirical estimators of moments. Based on [29], we study in Chapter 5 a maximum entropy model built on the description of point process distributions with wavelet phase harmonics covariances. We also present a gradient descent algorithm to efficiently sample from the model. Our findings indicate that wavelet phase harmonics covariances offer a more adequate representation than classic summary characteristics, such as nearest neighbours distances, to represent the distributions of point processes in this context.

Defining a set of statistics that characterize the distribution of a stochastic process is of central importance when defining a maximum entropy model. Following [30], we study in Chapter 6 a possible improvement of the wavelet phase harmonics covariances. We focus in this chapter on textures, which are stationary random images, that can be seen as realizations of random signed measures, and present complex geometric structures. We show that, using an appropriate set of wavelet-based covariance moments, one can build a maximum entropy model that gives results of similar visual quality to models built on state-of-the-art CNN-based representations, on a wide range of texture images.

Notations

We review the principal notations that shall be frequently used in this work.

- $(E, \mathcal{B})$ shall denote the Borel space associated with some Polish space E, on which point processes are defined in Chapter 2. Our particular focus will be on $E = \mathbb{R}^2$.
- $(\Omega, \mathcal{F}, \mathbb{P})$ denotes some probability space.
- $\mathbb{M}$ denotes the measurable set of counting measures, and $\bar{\mathbb{M}}$ the measurable set of locally finite measures.
- For a Borel set $(E, \mathcal{B})$ and a measure M on this space, we shall denote by $L^p(E, M) := \{f : E \mapsto \mathbb{C}, \int_E |f|^p dM < \infty\}$. When $E = \mathbb{R}^d$, for some integer $d \geq 1$, and M is the Lebesgue measure on E, we shall simply denote this space $L^p(\mathbb{R}^d)$.
- For an integer $N \geq 1$, $\mathcal{X}_N := \{0, \cdots, N-1\}^2$ is the set on which gray-scale images are defined, as elements of $\mathbb{R}^{N \times N}$. Color images are elements of $\mathbb{R}^{3 \times N \times N}$.
- For any $s > 0$, $W_s := [-\frac{s}{2}, \frac{s}{2}]^2$. It shall denote the window of $\mathbb{R}^2$ in which our point processes will be observed.
- For any $d \geq 1$, any two functions $f, g \in L^1(\mathbb{R}^d) \cup L^2(\mathbb{R}^d)$, the convolution of f and g is denoted by $f \star g := \int_{\mathbb{R}^d} f(u) g(\cdot - u) du$.

- For any function f from $\mathbb{R}^d$ to $\mathbb{C}$, we note $\check{f}$ the function defined, for any $u \in \mathbb{R}^d$, by $\check{f}(u) := f(-u)$.
- For any $z \in \mathbb{C}$, or any complex valued function f, z^* (respectively f^*) will denote its complex conjugate.
- For any $z \in \mathbb{C}$, $|z|$ denotes its complex modulus, and $\varphi(z)$ its argument, i.e. $z = |z|e^{i\varphi(z)}$.
- For $f \in L^1(\mathbb{R}^d) \cup L^2(\mathbb{R}^d)$, its Fourier transform, noted $\hat{f}$, is defined, for $\omega \in \mathbb{R}^d$, by $\hat{f}(\omega) := \int_{\mathbb{R}^d} f(u)e^{-i\omega u}du$.

Chapter 2

Stochastic geometry

In this first chapter, we recall the basic definitions and results of point process theory that we shall need for the remainder of the work. This is by no means an exhaustive introduction to the overall theory, as it simply highlights the notions that reappear in the chapters that follow. For a more thorough review of the fundamentals about the theory of point processes and random measures, and in particular, proofs of basic results, we refer the reader to [38, 39] or [6].

2.1 Point process framework

We begin by setting up the mathematical framework that enables us to define and study the objects called point processes and random measures. Informally, a point process can be seen as a random collection of points, living in some topological space E. The aim of this section is to introduce briefly the mathematical framework used to study these objects. While the topological space E may be more general (one can consider a topological, locally compact, second countable, Hausdorff space), let us assume in this work that it is the d-dimensional Euclidean space $\mathbb{R}^d$, for some integer $d \geq 1$. More precisely, our applications presented in Chapters 3 to 6 focus on the case $E = \mathbb{R}^2$. However, most of the definitions and result remain valid for more general spaces, so we shall keep as much as possible the notation E. When the properties of the space $\mathbb{R}^2$ are needed, it will be explicitly stated in the text. E is endowed with its Borel σ-algebra noted $\mathcal{B}$. Let us denote by $\mathcal{B}_c$ the set of all topologically bounded (i.e. with compact closure) elements of $\mathcal{B}$.

To study a collection of points of E, it is convenient to view it as a sum of Dirac atoms at the locations of the points. In other words, an at most countable, possibly infinite subset $\{u_1, \ldots, u_N\} \subset E$, for some $N \in \mathbb{N}\backslash\{0\} \cup \infty$, will be identified with the measure

$$\mu = \sum_{n=1}^{N} \delta_{u_n}. \tag{2.1}$$

The collections of points considered in point process theory usually are locally finite, i.e. finite on all bounded Borel sets $B \in \mathcal{B}_c$. A counting measure on $(E, \mathcal{B})$ is a locally finite measure taking values values in $\mathbb{N} \cup \infty$. $\mathbb{M}$ denotes the set of all counting measures on $(E, \mathcal{B})$. More generally, the set of locally finite measures on $(E, \mathcal{B})$ will be noted $\bar{\mathbb{M}}$. It is known that any non-null measure $\mu \in \mathbb{M}$ can be expressed as a sum of Dirac atoms located at some points in E, without accumulation points. The link between the collection of points and the counting measure is clear: for any $u \in E$, $\mu(u)$ counts the number of points in the collection that are equal to u. One can informally write $u \in \mu$ to mean $\mu(u) \geq 1$. In order to consider random variables taking values in $\mathbb{M}$, one needs to endow $\mathbb{M}$ with a σ-algebra. Let $\mathcal{M}$ be the one generated by the mappings $\mu \mapsto \mu(B)$, $B \in \mathcal{B}_C$.

A point process is a measurable mapping from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the space $(\mathbb{M}, \mathcal{M})$. The distribution $\mathbb{P}_\Phi$ of a point process Φ is image of $\mathbb{P}$ by Φ, i.e.

$$\forall\, \Gamma \in \mathcal{M},\ \mathbb{P}_\Phi(\Gamma) = \mathbb{P}(\{\Phi \in \Gamma\}). \tag{2.2}$$

Note that a point process Φ can be viewed as a stochastic process: $\Phi = \{\Phi(B)\}_{B \in \mathcal{B}_c}$. Therefore, by the Kolomogorov extension theorem, the distribution of a point process is characterized by its finite dimensional distributions $(\Phi(B_1), \Phi(B_2), \ldots, \Phi(B_k))_{k \geq 1, B_1, \ldots, B_k \in \mathcal{B}_c}$.

This notion can be extended to define random measures: let $\bar{\mathbb{M}}$ the space of all locally finite measures on $(E, \mathcal{B})$, endowed with the σ-algebra $\bar{\mathcal{M}}$, defined similarly to $\mathcal{M}$. A random measure is a measurable mapping $\Pi : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto (\bar{\mathbb{M}}, \bar{\mathcal{M}})$.

2.2 Basic characteristics of point processes

Many tools and functions have been developed to describe the distributions of point processes. In this section, we review several of them, that shall be useful fro the remainder of this dissertation.

2.2.1 Moment measures

As point processes are random measures, it is natural to look at the expected value of such measures on Borel sets of E, or products of such measures. This can be formalized by the definitions that follow.

Definition 2.2.1 (Mean measure). The mean measure M_Φ of a point process Φ is defined on $(E, \mathcal{B})$ by

$$M_\Phi(B) = \mathbb{E}[\Phi(B)], \quad B \in \mathcal{B}. \tag{2.3}$$

This definition allows us to state one of the fundamental results in the theory of point processes.

Theorem 2.2.2 (Campbell's averaging formula). *Let Φ be a point process with mean measure M. For any measurable function $f : E \mapsto \mathbb{R}$, non-negative or integrable with respect to M, the integral $\int_{\mathbb{E}} f(u)\Phi(du)$ is a.s. well defined, and*

$$\mathbb{E}\Big[\int_E f(u)\Phi(du)\Big] = \int_E f(u) M(du). \tag{2.4}$$

The mean measure is simply the average number of atoms of Φ present in the set B. This notion can be generalized by looking at the joint number of points present in n-tuples of Borel sets.

Definition 2.2.3 (High order moment measures). For any $n \geq 1$, the n-th order moment measure of a point process Φ, noted M_{Φ^n}, is the measure on $(E^{\times n}, \mathcal{B}^{\otimes n})$ defined, for all $B_1, B_2, \cdots, B_n \in \mathcal{B}$ by:

$$M_{\Phi^n}(B_1 \times B_2 \times \cdots \times B_n) := \mathbb{E}[\Phi(B_1)\Phi(B_2) \cdots \Phi(B_n)]. \tag{2.5}$$

It is equivalent to the mean measure of the point process Φ^n on $(E^{\times n}, \mathcal{B}^{\otimes n})$ defined by $\Phi^n(B_1 \times B_2 \times \cdots \times B_n) = \Phi(B_1)\Phi(B_2)\cdots\Phi(B_n)$.

For sets $B_1, B_2, \cdots, B_n$ that are not mutually disjoints, atoms of Φ that lie in the intersections of those sets are counted more than once by the n-th order moment measure. To take this into account, one can use instead the following modification.

Definition 2.2.4 (Factorial moment measures)**.** Let $\Phi = \sum_i \delta_{u_i}$ be a point process and $n \in \mathbb{N}$, $n \geq 1$. The n-th factorial power of Φ, noted $\Phi^{(n)}$, is the random measure on $(E^{\times n}, \mathcal{B}^{\otimes n})$ defined by

$$\Phi^{(n)} := \sum_{\Delta_\Phi^n} \delta_{(u_{i,1}, u_{i,2}, \cdots, u_{i,n})}, \tag{2.6}$$

where $\Delta_\Phi^n := \{(u_{i,1}, u_{i,2}, \cdots, u_{i,n}),\ u_{i,j} \in \Phi, \forall j, k, j \neq k \Rightarrow u_{i,j} \neq u_{i,k}\}$. The n-the factorial moment measure is defined by:

$$M_{\Phi^{(n)}}(B_1 \times B_2 \times \cdots \times B_n) := \mathbb{E}[\Phi^{(n)}(B_1 \times B_2 \times \cdots \times B_n)]. \tag{2.7}$$

Furthermore, if $M_{\Phi^{(n)}}$ admits a density w.r.t. the Lebesgue measure, it shall be noted $\rho^{(n)}(x_1, \cdots, x_n)$.

The joint factorial moment measures characterize the distribution of a point process Φ, provided it has exponential moments, i.e. $\forall\, B \in \mathcal{B}_C, \exists \epsilon : \mathbb{E}[e^{\epsilon \Phi(B)}] < \infty$.

2.2.2 Other common characteristics

Other functions can be used to characterize the distribution of a point process. Among them are the two following.

Void probability The void probability v_Φ of a point process Φ is defined by

$$v_\Phi(B) = \mathbb{P}(\Phi(B) = 0), \tag{2.8}$$

for all $B \in \mathcal{B}$. The distribution of a simple point process (i.e. such that $\mathbb{P}(\forall u \in E, \Phi(\{u\}) \leq 1) = 1$, meaning that, almost surely, Φ has no multiple atoms in E) is fully characterized by the family $(v_\Phi(B))_{B \in \mathcal{B}_c}$.

Laplace transform The Laplace transform L_Φ of a point process Φ takes a non-negative, measurable function f and maps it to $\mathbb{R}_+$ by

$$L_\Phi(f) := \mathbb{E}[e^{-\int f\, d\Phi}]. \tag{2.9}$$

One can show that the Laplace transform of a point process fully characterizes its distribution.

2.3 Poisson, clustering, and repulsive distributions

As previously explained, point processes are random collections of points in some metric space. Depending on their distributions, realizations of point processes can exhibit different geometric patterns. The points can tend to regroup together or repel each other, possibly forming complex geometric structures. While there exist many classic distributions in the literature on point processes, we shall, in this section, briefly detail some of the most usual ones, in order to give a panorama of these objects. For a more complete overview of the 'zoology' of point process distributions, see for example [45].

2.3.1 Poisson point process

The most fundamental example of point process distribution is the Poisson distribution. It describes a process where points are 'placed independently of each other', and identically on the space. It is formally defined as follows.

Definition 2.3.1 (Poisson point process)**.** Let Λ be a deterministic, locally finite measure on $(E, \mathcal{B})$. A point process Φ is a Poisson point process with intensity Λ if it satisfies the following conditions:

1. For any $B \in \mathcal{B}_c$, $\Phi(B)$ is a Poisson random variable with intensity $\Lambda(B)$.
2. For any $k \geq 1$, and for any $B_1, \dots, B_k \in \mathcal{B}_c$, pairwise disjoint, the random variables $\Phi(B_1), \dots, \Phi(B_k)$ are independent.

Its simple hypotheses (independence, Poissonianity) make the Poisson distribution highly popular, and its properties have been extensively studied (see e.g. [71]).

Some properties of the Poisson point process

- Its mean measure M is equal to its intensity measure Λ. We say that the Poisson point process is *homogeneous* if its intensity measure is proportional to the Lebesgue measure on E.
- Its void probability is $\mu(B) = e^{-\Lambda(B)}$.
- Its Laplace transform is equal to $L(f) = e^{-\int_E 1-e^{-f(u)}\Lambda(du)}$.

2.3.2 Other classic distributions

The Poisson point process is sometimes said to have *complete independence*, as the number of points in any disjoint regions of the space are independent. It is usually considered as a reference process, and other distributions, that exhibit spatial structures, are usually classified in one of the two following categories: clustering or repulsive. The clustering of points refers to their tendency to form groups, well separated from each other, whereas a point process is said to be repulsive if the points tend to repel each other, and are more homogeneously spread in the space. Figure 2.1 displays examples of different point process geometries. In this section, we review several important distributions of point process exhibiting such structures. For more details about comparison methods between point process distributions, see [17].

Cox point processes Also called doubly stochastic point processes, Cox point processes constitute a rich class of processes, often used to model patterns exhibiting more clustering than the Poisson point process. Let Π be a random, locally finite, non null measure on $(E, \mathcal{B})$. A Cox point process Cox_Π on $(E, \mathcal{B})$ generated by (or directed by) Π is the mixture of Poisson point processes with respect to Π, that is, conditioned on $\Pi = \Lambda$, Cox_Π is the Poisson point process with mean (deterministic) measure Λ. A rich class of Cox processes is constituted by *Neyman-Scott processes* (see e.g. [85], or [37, Section 5.3]), where 'parent points' are drawn from a homogeneous Poisson distribution, and from each parent point p_i descend N_i children, with N_i being a Poisson r.v., and the location of the children a determined by a given density f. For instance, if this density f

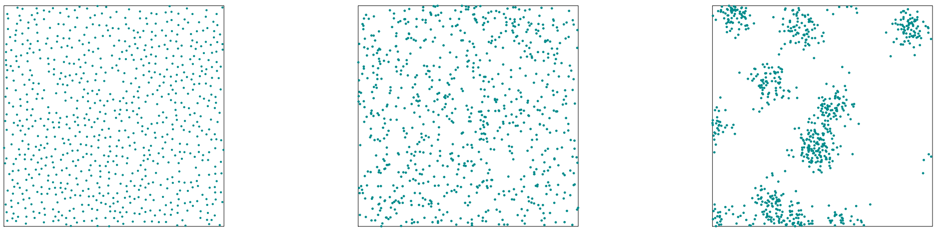


Figure 2.1: Examples of different point process distributions. Left: Matern II point process (repulsive), middle: homogeneous Poisson point process, right: modified Thomas point process (clustering).

is a Gaussian function, then the resulting process is called a modified Thomas process. If the density is uniform in a ball, it is called a Matérn cluster process.

Gibbs point processes A Gibbs process is a point process whose probability distribution has a density with respect to the probability distribution of a Poisson point process. Informally, one obtains a Gibbs point process by 'filtering' realizations of a Poisson point process, giving more probability to appear to some configuration than others. The clustering or repulsive behaviour of the distribution will depend on the filtering of the configurations. A Gibbs point process can be defined by considering a Poisson process Φ, and some measurable function from $\mathbb{M}$ to $\mathbb{R}^+$, such that $\mathbb{E}[f(\Phi)] = 1$. The point process Φ_f with distribution defined by $\mathbb{P}_{\Phi_f}(d\mu) = f(\mu)\mathbb{P}_{\Phi}(d\mu)$ is a Gibbs point process with density f with respect to Φ[1]. For a more detailed introduction to Gibbs point processes, see e.g. [44].

Hard-core point processes Hard-core processes are defined by the property that all atoms of its configurations are further away from each other than some minimal distance. They are called hard-core because the points can represent the center of solid spheres non smaller than some value. Classic examples of hard-core processes are Matérn I, II, and III point processes (see e.g. [82, 83]).

Determinantal point processes Let Φ be a simple point process on $(E, \mathcal{B})$, Λ a locally finite measure on $(E, \mathcal{B})$, and $k : E^2 \mapsto \mathbb{C}$ a measurable mapping. Φ is a determinantal point process with background measure Λ if for all $n \geq 1$, the n-th order factorial moment measure $M_{\Phi^{(n)}}$ of Φ admits a density $\rho_{\Lambda}^{(n)}$ with respect to the product measure Λ^n, which equals

$$\rho_{\Lambda}^{(n)}(u_1, \cdots, u_n) = \det\Big((k(u_i, u_j)_{1\leq i,j\leq n}\Big),$$

for Λ^n-almost all $(u_1, \cdots, u_n) \in E^n$. Interpreting Λ as the intensity measure of a Poisson point process Φ_0, $M_{\Phi^{(n)}}$ has density $\rho_{\Lambda}^{(n)}$ w.r.t. $M_{\Phi_0^{(n)}}$ ([6, Chapter 5]). Classical

---

[1] In principle, one can consider Φ to be a homogeneous Poisson point process on E. However, one cannot obtain a stationary $\Phi_f \neq \Phi$ with this construction. Indeed, in this case f has to be translation invariant. Therefore, by ergodicity, $f \equiv 1$. For this reason, such a construction is often considered in a finite window.

examples of determinantal point processes are the Gaussian determinantal point process, with a kernel of the form $k(u,v) = \alpha e^{-\frac{\|u-v\|^2}{\beta}}$, for some positive constants α, β, and $u, v \in \mathbb{R}^d$ (see e.g. [73]), and the Ginibre point process, with kernel of the form $k(u,v) = \frac{1}{\pi} e^{uv^*} e^{-(|u|^2+|v|^2)/2}$, for $u, v \in \mathbb{C}$ (see e.g. [57]), both considered w.r.t. the Lebesgue measure $\Lambda(dx) = dx$.

2.4 Palm distributions of point processes

When studying point processes, one might consider high order moments, introduced in Section 2.2, that are the expected number of points that would appear simultaneously in several given regions of the space. It may also be of interest to study the conditional number of points in a given region, knowing the number of points in the other region. Going further, on may enquire about the conditional distribution of the point process, knowing that this process has an atom at a given position $u \in E$. However, if u is not a fixed atom of Φ, $\mathbb{P}(u \in \Phi) = 0$, so one cannot directly consider conditional distribution $\mathbb{P}(\cdot | u \in \Phi)$. To overcome this issue, Palm distributions were introduced, based on the Radon-Nikodym theorem.

Definition 2.4.1 (Campbell measure and Palm distributions)**.** Let Φ be a point process on E, with Radon mean measure M. The Campbell measure of Φ is defined on $(E \times \mathbb{M}, \mathcal{B} \times \mathcal{M})$ by

$$C(B \times \Gamma) := \mathbb{E}[\int_E \mathbb{1}_B(u) \mathbb{1}_\Gamma(\Phi) \Phi(du)], \quad B \in \mathcal{B}, \Gamma \in \mathcal{M}.$$

We immediately remark that, for any $\Gamma \in \mathbb{M}$, $C(B \times \Gamma) \leq M(B)$. The Radon-Nikodym theorem tells us that there exists a measurable function $u \mapsto P_u(\Gamma)$ such that,

$$\forall B \in \mathcal{B}, \quad C(B \times \Gamma) = \int_E \mathbb{1}_B(u) P_u(\Gamma) M(du).$$

Note that for a fixed $\Gamma \in \mathbb{M}$, $P_u(\Gamma)$ is not uniquely defined, and can differ on any set of null M measure. As E is a polish space, $P_{\cdot}(\cdot)$ can be defined as a probability kernel from $(E, \mathcal{B})$ to $(\mathbb{M}, \mathcal{M})$, implying that for any $u \in E$, $P_u(\cdot)$ is a probability distribution on $(\mathbb{M}, \mathcal{M})$, called a Palm distribution of Φ.

This definition allows us to state the following results, that extends Campbell's averaging formula to stochastic processes.

Theorem 2.4.2 (Campbell-Little-Mecke (CLM))**.** *Let Φ be a point process on E, with Radon mean measure M, and Palm distributions P_u, for all $u \in E$. For all non-negative measurable functions $f : E \times \mathbb{M} \mapsto \mathbb{R}$,*

$$\mathbb{E}[\int_E f(u, \Phi) \Phi(du)] = \int_{E \times \mathbb{M}} f(u, \mu) P_u(d\mu) M(du). \tag{2.10}$$

The result extends to all functions f for which either of the two sides of the equality eq. (2.10) *is finite when f is replaced by $|f|$.*

For any $u \in E$, the quantity $\int_{\mathbb{M}} f(u, \mu) P_u(d\mu)$ shall also be noted $\mathbb{E}^u[f(u, \Phi)]$ (that is, the expectation of $f(u, \Phi)$ for some point process Φ having distribution P_u on $\mathbb{M}$).

Reduced and high-order Palm distributions The reduced version of the Palm distributions, noted $P_u^!$ can be defined as the Radon-Nikodym derivative of

$$C^{(1)}(B \times \Gamma) := \mathbb{E}[\int_E \mathbb{1}_B(u)\mathbb{1}_\Gamma(\Phi - \delta_u)\Phi(du)], \quad B \in \mathcal{B}, \Gamma \in \mathcal{M}. \tag{2.11}$$

The Palm distributions can be extended to higher order Palm distributions $P_{u_1,\cdots,u_n}$ by taking the derivative of

$$C^n(B \times \Gamma) := \mathbb{E}[\int_{E^n} \mathbb{1}_B(u)\mathbb{1}_\Gamma(\Phi)\Phi^{(n)}(du)], \quad B \in \mathcal{B}^{\otimes n}, \Gamma \in \mathcal{M}, \tag{2.12}$$

and similarly for the reduced high-order versions $P^!_{u_1,\cdots,u_n}$, by taking the derivative of:

$$C^{(n)}(B \times \Gamma) := \mathbb{E}[\int_{E^n} \mathbb{1}_B(u)\mathbb{1}_\Gamma(\Phi - \sum_{i=1}^n \delta_{u_i})\Phi^{(n)}(du)], \quad B \in \mathcal{B}^{\otimes n}, \Gamma \in \mathcal{M}. \tag{2.13}$$

The quantity $\int_{\mathbb{M}} f(u, \mu)P_{u_1\cdots,u_n}(d\mu)$ shall also be noted $\mathbb{E}^{u_1\cdots,u_n}[f(u, \Phi)]$, and the reduced versions $\mathbb{E}^{!u_1\cdots,u_n}[f(u, \Phi)]$.

2.5 Stationary point processes

For this section, we place ourselves in the case where E is some Euclidean space (our particular focus in this work shall be $E = \mathbb{R}^2$). A point process Φ is called stationary if its distribution is invariant to translation. That is, for any $t \in E$, any $n \geq 1$, and any $B_1, \cdots, B_n \in \mathcal{B}_c$, the distribution of $\{\Phi(B_1), \cdots, \Phi(B_n)\}$ is equal to that of $\{\Phi(B_1 + t), \cdots, \Phi(B_n + t)\}$, where $B + t := \{u + t,\ u \in B\}$. Another way to say this is, noting $\Phi = \sum_i \delta_{x_i}$, Φ is equal in distribution to $\sum_i \delta_{x_i - t}$. It can be convenient to consider stationary processes, when one observes a large phenomenon that appear similar in different positions, at smaller scale. For instance, the positions of trees in a forest can be regarded, when observed at a certain scale, as a stationary point process. Another example would be the positions of stars, galaxies, or other cosmological objects, projected on a sphere around the Earth. The theory of stationary point processes gives us tools to better describe their properties. Here, we review the principal definitions and results for such processes.

In order to study the properties of stationary point processes, it is useful do define a stationary framework directly on the probability space.

Definition 2.5.1 (Shift operator)**.** For any $t \in E$, and any measure μ on $(E, \mathcal{B})$, the shift operator S_t applied to μ is the measure $S_t\mu$ on $(E, \mathcal{B})$ defined by, for all $B \in \mathcal{B}$:

$$S_t\mu(B) = \mu(B + t), \tag{2.14}$$

where $B + t := \{u + t,\ u \in B\}$. If μ is a counting measure, say $\mu = \sum_i \delta_{u_i}$, then $S_t\mu = \sum_i \delta_{u_i - t}$. The shift operator can be extended to functions f with values in any space by defining $S_t f(u) := f(u + t)$.

Definition 2.5.2 (Flow on the probability space)**.** Let $(\Omega, \mathcal{F})$ a measurable space. Let us assume that there exists a family $\{\theta_t : \Omega \mapsto \Omega\}_{t \in E}$ of measurable mappings satisfying the following conditions:

- For any $t \in E$, θ_t is a bijection from Ω to Ω.

- For any $t, s \in E$, $\theta_s \circ \theta_t = \theta_{s+t}$.
- The mapping $\tilde{\theta} : (E, \Omega) \mapsto \Omega$ defined by $\tilde{\theta}(t, \varpi) := \theta_t(\varpi)$ is $\mathcal{B} \otimes \mathcal{F}$-measurable.

Such a family will be called a (measurable) flow on $(\Omega, \mathcal{F})$, and the space equipped with this flow will be noted $(\Omega, \mathcal{F}, \{\theta_t\})$.

Definition 2.5.3 (Processes compatible with the flow)**.** A point process $\Phi : (\Omega, \mathcal{F}, \{\theta_t\}) \mapsto (\mathbb{M}, \mathcal{M})$ is said to be compatible with the flow if

$$\forall t \in E,\ \Phi \circ \theta_t = S_t \Phi. \tag{2.15}$$

Example 2.5.4. *Canonical probability space with the flow. The space* $(\Omega, \mathcal{F}, \{\theta_t\}) = (\mathbb{M}, \mathcal{M}, \{S_t\})$*, is the canonical space on which we define* $\Phi(\mu) = \mu, \forall \mu \in \mathbb{M}$*, compatible with the flow.*

Definition 2.5.5 (Stationary probability)**.** Let $(\Omega, \mathcal{F}, \{\theta_t\})$ be a measurable space equipped with a flow. Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$ invariant with respect to any element in the flow, that is

$$\forall\, t \in E,\ \mathbb{P} \circ \theta_t^{-1} = \mathbb{P}. \tag{2.16}$$

$(\Omega, \mathcal{F}, \{\theta_t\}, \mathbb{P})$ is then called a stationary framework, and $\mathbb{P}$ is called the stationary probability on $(\Omega, \mathcal{F}, \{\theta_t\})$.

It directly follows from the above definition that a point process Φ defined on a stationary framework $(\Omega, \mathcal{F}, \{\theta_t\}, \mathbb{P})$ is stationary.

Example 2.5.6. *The distribution of a homogeneous Poisson point process is invariant to any shift, and can equip the canonical probability space.*

Palm probabilities in the stationary framework Let Φ be a point process defined on the stationary framework, compatible with the flow. A first observation is that its mean measure is proportional to the Lebesgue measure. The coefficient of proportionality is called the intensity of the process.

Definition 2.5.7 (Campbell-Matthes' measure)**.** Let $\mathcal{C}$ be a measure on $(E \times \Omega, \mathcal{B} \times \mathcal{F})$ defined by, for all $B \in \mathcal{B}$ and all $A \in \mathcal{F}$

$$\mathcal{C}(B \times A) := \mathbb{E}\Big[\int_E \mathbb{1}_B(u)\mathbb{1}_A(\theta_u)\Phi(du)\Big] \tag{2.17}$$

Note that Campbell-Matthes' measure is an extension of the mean measure of Φ: $\mathcal{C}(B \times \Omega) = M_\Phi(B)$. Moreover, for all $A \in \mathcal{F}$, $\mathcal{C}(B \times A) \leq M_\Phi(B)$, which implies that $\mathcal{C}(\cdot \times A)$ admits a Radon-Nikodym derivative $\frac{d\mathcal{C}(\cdot \times A)}{dM_\Phi(\cdot)}(u)$ w.r.t. M_Φ. Additionally, similarly to M_Φ, $\mathcal{C}(\cdot \times A)$ is invariant to any shift in E, which gives us that it is proportional to the Lebesgue measure on E $(= \mathbb{R}^d)$. Thus, $\frac{d\mathcal{C}(\cdot \times A)}{dM_\Phi(\cdot)}(u)$ is a constant, depending only on A and λ. Seeing it as a function of A, one can easily check that it constitutes a probability measure on $(\Omega, \mathcal{F})$, leading us to the following definition.

Definition 2.5.8 (Palm probability)**.** Let Φ be a point process defined on the stationary framework $(\Omega, \mathcal{F}, \{\theta_t\}, \mathbb{P})$, compatible with the flow, with finite non-null intensity λ. The Palm probability of Φ is the unique measure $\mathbb{P}^0$ on $(\Omega, \mathcal{F})$ defined by

$$\mathbb{P}^0(A) := \frac{1}{\lambda|B|}\mathbb{E}\Big[\int_E \mathbb{1}_B(u)\mathbb{1}_A(\theta_u)\Phi(du)\Big], \quad \forall A \in \mathcal{F}, \tag{2.18}$$

for any $B \in \mathcal{B}$ on finite, non-null Lebesgue measure $|B|$. We shall denote by $\mathbb{E}^0$ the expectation under $\mathbb{P}^0$.

Theorem 2.5.9 (Campbell-Little-Mecke-Matthes)**.** *Let Φ be a point process defined on the stationary framework $(\Omega, \mathcal{F}, \{\theta_t\}, \mathbb{P})$, compatible with the flow, with finite non-null intensity λ. Denote $\mathbb{P}^0$ the Palm probability of Φ. For any non-negative measurable function f on $E \times \Omega$,*

$$\mathbb{E}\Big[\int_E f(u, \theta_u\cdot)\Phi(du)\Big] = \lambda\int_E \mathbb{E}^0[f(u,\cdot)]du. \tag{2.19}$$

The results extends for any function f for which either one of the two sides of the above equation is finite when f is replaced by $|f|$.

Corollary 2.5.10. *Under the assumptions of Theorem 2.5.9, $0 \in \Phi$ $\mathbb{P}^0$-almost surely.*

Definition 2.5.11 (Reduced second-order density)**.** Let Φ be a point process defined on the stationary framework $(\Omega, \mathcal{F}, \{\theta_t\}, \mathbb{P})$, compatible with the flow, with finite non-null intensity λ. Suppose that $M_{\Phi^{(2)}}$ is σ-finite, and has a density w.r.t. the Lebesgue measure. As Φ is stationary, we can write $M_{\Phi^{(2)}}(d(u,v)) = \lambda\kappa(v-u)dudv$, and for any $B \in \mathcal{B}$, $\mathbb{E}^0[(\Phi - \delta_0)(B)] = \int_B \kappa(u)du$. The function κ is called the reduced second order density of Φ.

2.6 Geometric marks for stationary point processes

One can extend the notion of point process, by considering counting measures with values attached to each atom. Such values, possibly random, are usually called the *marks* of the atoms, and the associated random measure is called a marked point process. In this work, we shall take interest in a particular class of marks, described in what follows. In this section, we consider $E = \mathbb{R}^2$.

Definition 2.6.1 (Geometric marks)**.** Let Φ be a point process defined on the stationary framework $(\Omega, \mathcal{F}, \{\theta_t\}, \mathbb{P})$, compatible with the flow, with finite non-null intensity λ. Let $m : \mathbb{R}^2 \times \mathbb{M} \mapsto \mathbb{C}$ be a measurable function, satisfying, for any $u, v \in \mathbb{R}^2$, and any $\phi \in \mathbb{M}$,

$$m(u - v, S_v\phi) = m(u, \phi).$$

We say that the function m is translation invariant, and shall be called a geometric mark of Φ. With the convention we consider, the associated marked point process[2], noted

$$\tilde{\Phi} = \sum_{u\in\Phi} m(u,\Phi)\delta_u,$$

is stationary. Note that a point process without marks is the special case of $m \equiv 1$.

---

[2] Rigorously, a marked point process is a sum of atoms $\tilde{\Phi} = \sum_i \delta_{(u_i,m_i)}$, with $(m_i)_i$ being a sequence of random variables. When the $m_i \in \mathbb{C}$, we adopt the convention $\tilde{\Phi} = \sum_{u\in\Phi} m(u,\Phi)\delta_u$, which is formally not a marked point process.

In what follows, we briefly recall a few basic examples of geometric marks, that will be considered in this work (more particularly in Sections 3.3 and 3.4 and Chapter 4).

Example 2.6.2 (Shot-noise)**.** *For* $u \in \mathbb{R}^2$ *and* $\Phi = \sum_i \delta_{u_i} \in \mathbb{M}$*, we can define*

$$m(u, \Phi) := \sum_i \mathbb{1}(u \neq u_i)\ell(|u - u_i|),$$

a shot-noise *functional, with some non-negative response function* ℓ*, that depends on the distance* $|u - u_i|$ *between* u *and* $u_i \in \phi$*, e.g.* $\ell(r) = r^\beta$*, for some* $\beta > 2$*.*

Example 2.6.3 (Nearest-neighbour distance)**.** *We can define the nearest neighbour distance*

$$m_{nn}(u, \Phi) = \min\{||u - u_i||, u_i \in \Phi\}.$$

Example 2.6.4 (Voronoi cell area and moment of inertia)**.** *The Voronoi cell of an atom* $u \in \Phi$ *is defined by*

$$V(u) = \{v \in \mathbb{R}^2 : \forall\, u' \in \Phi, \|v - u\| \leq \|v - u'\|\}.$$

One can consider the Voronoi cell volume

$$m_V(u, \Phi) = |V(u)|, \tag{2.20}$$

or its moment of inertia

$$m_I(u, \Phi) := \int_V |v - u|^2 \, dv.$$

Example 2.6.5 (Voronoi shot-noise)**.** *One can consider more sophisticated marks, such as*

$$m_Z(u, \Phi) = \sum_i \mathbb{1}(u_i \neq u)\ell(|u - u_i|)m_V(u, \Phi).$$

2.7 Ergodicity

Ergodic theory studies the conditions under which spatial averages of functions of point processes converge to the expectation of such functions, in order to bridge the gap between the theoretical point of view of probabilistic theory, and real-world measurements. We saw in Section 2.5 that a point process is stationary if it appears similar at different locations in the space. Ergodicity goes even further, saying that one can average a function over all different locations in the space, for a single realization, and by doing so, estimate the expected value of this function. In a sense, it can be viewed as a kind of Law of Large Numbers. This property is useful when we are dealing with a single, large observation, which can happen when collecting the data is difficult.

Recalling the notions of Section 2.5, let $(\Omega, \mathcal{F}, \{\theta_t\}_{t\in E}, \mathbb{P})$ a stationary framework (in this section, E is some Euclidean space). We begin by defining invariant events under this framework.

Definition 2.7.1 (Invariant events)**.** An event $A \in \mathcal{F}$ is said to be $(\{\theta_t\}_{t\in E}, \mathbb{P})$-invariant if, for all $t \in E$,

$$\mathbb{P}(A \Delta \theta_t A) = 0,$$

where Δ is the symmetric difference: $A\Delta B := (A \cup B)\backslash(A \cap B)$. The space

$$\mathcal{I} := \{A \in \mathcal{F} : A \text{ is } (\{\theta_t\}_{t\in E}, \mathbb{P})\text{-invariant}\}$$

is a σ-algebra, called the invariant σ-algebra.

The fundamental theorem in ergodic theory (in the context of point processes), which proof can be found in e.g. [67], states the following.

Theorem 2.7.2 (Birkhoff's individual ergodic theorem)**.** *Let* $(\Omega, \mathcal{F}, \{\theta_t\}_{t\in E}, \mathbb{P})$ *be a stationary framework,* $\mathcal{I}$ *its invariant* σ*-algebra. Let* Φ *be a point process on* E*, compatible withe the flow, with intensity* $0 < \lambda < \infty$*. Let* $\{B_n\}_{n\in\mathbb{N}}$ *be a convex averaging sequence, i.e. an sequence of bounded, convex Borel sets such that for all* $n \in \mathbb{N}$*,* $B_n \subset B_{n+1}$*, and* $\sup\{r \geq 0 : B_n \text{ contains a ball of radius } r\} \xrightarrow[n\to\infty]{} \infty$*. Let* f *be a measurable, integrable function from* $\mathbb{M}$ *to* $\mathbb{C}$*. Then*

$$\lim_{n\to\infty} \frac{1}{|B_n|} \int_{B_n} f(\theta_t \Phi)\, dt = \mathbb{E}[f(\Phi)|\mathcal{I}], \quad \mathbb{P}\text{-}a.s.$$

We now need the following notions:

Definition 2.7.3 (Metrically transitive, ergodic, or mixing framework)**.** We say that the stationary framework is metrically transitive if $\mathcal{I}$ is trivial, i.e. for all $A \in \mathcal{F}$, $\mathbb{P}(A) \in \{0, 1\}$. The framework is ergodic if for all $A_1, A_2 \in \mathcal{F}$,

$$\lim_{a\to\infty} \frac{1}{(2a)^d} \int_{[-a,a]^d} \mathbb{P}(A_1 \cap \theta_t A_2) dt = \mathbb{P}(A_1)\mathbb{P}(A_2),$$

where d is the dimension of the Euclidean space E. The framework is said to be mixing if for all $A_1, A_2 \in \mathcal{F}$,

$$\lim_{|t|\to\infty} \mathbb{P}(A_1 \cap \theta_t A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).$$

The following result, which proof can be found in e.g. [6], states the relation between the three definitions from above.

Proposition 2.7.4. *Let* $(\Omega, \mathcal{F}, \{\theta_t\}_{t\in E}, \mathbb{P})$ *be a stationary framework. Then, the following relations are true*

$$\textit{mixing} \Rightarrow \textit{ergodic} \Leftrightarrow \textit{metrically transitive.}$$

The mixing and ergodic properties are useful because they are easier to verify in practice than the metrical transitivity. These notions allow us to state the main result:

Corollary 2.7.5. *Under the conditions of Theorem 2.7.2, and if the stationary framework is metrically transitive, then*

$$\mathbb{E}[f(\Phi)|\mathcal{I}] = \mathbb{E}[f(\Phi)]. \tag{2.21}$$

2.8 Linear operators and second-order properties for stationary point processes

Second order characteristics offer a natural way to describe geometric structures in point processes, by studying correlations between pairs of points. In this section, we shall place ourselves in $E = \mathbb{R}^2$. There have been extensively studied in the literature (see e.g. [96, 9, 27]) and constitute some of the most widely used tools to analyse point processes.

We call a shot noise of a point process Φ any random variable defined by $\Phi(f) = \int_{\mathbb{R}^2} f(u)\Phi(du)$, for any measurable, integrable function $f : \mathbb{R}^2 \mapsto \mathbb{C}$. In this section, we shall study second order moments based on shot noises for stationary processes, define the Bartlett spectrum, and show how we can estimate it using the Discrete Fourier Transform on a finite window.

2.8.1 Bartlett spectrum of a point process

The Bartlett spectrum of a stationary point process is the measure that allows one to express the covariance between shot-noises as a function of their Fourier transform. More formally,

Proposition 2.8.1. *Let Φ be a point process defined on the stationary framework $(\Omega, \mathcal{F}, \{\theta_t\}, \mathbb{P})$, compatible with the flow, with finite non-null intensity λ. Suppose that $\gamma(u) := \lambda\kappa(u) - \lambda^2 \in L^2(\mathbb{R}^2)$, where κ is the reduced second-order density of Φ (cf. Definition 2.5.11). Let us denote L^2_Φ the set of complex valued, measurable, integrable functions such that $\mathbb{E}[\Phi(|f|)^2] < \infty$. For any two functions $f, g \in L^2_\Phi$ such that $f \star \check{g} \in L^2(\mathbb{R}^2)$,*

$$Cov(\Phi(f), \Phi(g)) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{f}(\omega)\hat{g}^*(\omega) b_\Phi(\omega)(d\omega),$$

where $b_\Phi(\omega) := \hat{\gamma}^*(\omega) + \lambda$.

Proof. First remark that $L^2_\Phi \subseteq L^1(\mathbb{R}^2, M_\Phi) \cap L^2(\mathbb{R}^2, M_\Phi)$, so $\mathrm{Cov}(\Phi(f), \Phi(g)) := \mathbb{E}[\Phi(f)\Phi(g^*)] - \mathbb{E}[\Phi(f)]\mathbb{E}[\Phi(g^*)]$ is well defined, and both terms in the r.h.s. of the equality are finite. Then, we decompose

$$\begin{aligned}\mathrm{Cov}(\Phi(f), \Phi(g)) = &\int_{\mathbb{R}^2 \times \mathbb{R}^2} f(u) g^*(v) M_{\Phi^{(2)}}(d(u,v)) \\ &+ \lambda \int_{\mathbb{R}^2} f(u) g^*(u) du \\ &- \lambda^2 \int_{\mathbb{R}^2 \times \mathbb{R}^2} f(u) g^*(v) du dv.\end{aligned}$$

This gives us, with the change of variables $z = v - u$,

$$\begin{aligned}\mathrm{Cov}(\Phi(f), \Phi(g)) &= \lambda \int_{(\mathbb{R}^2)^2} f(u)\check{g}^*(-u-z)(\kappa(z) - \lambda) du dz \\ &\quad + \lambda \int_{(\mathbb{R}^2)^2} f(u)\check{g}^*(-u-z) du \delta_0(dz) \\ &= \lambda \int_{\mathbb{R}^2} f \star \check{g}^*(-u)(\kappa(u) - \lambda) du + \lambda \int_{\mathbb{R}^2} f \star \check{g}^*(-u)\delta_0(du) \\ &= \int_{\mathbb{R}^2} f \star \check{g}^*(u) \Gamma_\Phi(du),\end{aligned}$$

where

$$\Gamma_\Phi(du) := \lambda\kappa(u)du - \lambda^2 du + \lambda\delta_0(du) \tag{2.22}$$

is called the covariance measure of Φ. Recall that $\gamma(u) = \lambda\kappa(u) - \lambda^2 \in L^2(\mathbb{R}^2)$, with κ being symmetric, so by Plancherel's equality,

$$\begin{aligned}\mathrm{Cov}(\Phi(f), \Phi(g)) &= \int_{\mathbb{R}^2} f \star \check{g}^*(u)\gamma(u) du + \lambda \int_{\mathbb{R}^2} f \star \check{g}^*(-u)\delta_0(du) \\ &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{f}(\omega)\hat{g}^*(\omega)\hat{\gamma}^*(\omega) d\omega + \frac{\lambda}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{f}(\omega)\hat{g}^*(\omega) d\omega \\ &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{f}(\omega)\hat{g}^*(\omega) b_\Phi(\omega)(d\omega). \end{aligned} \tag{2.23}$$

□

Definition 2.8.2 (Bartlett spectral measure)**.** Let Φ be a stationary point process, satisfying the conditions of Proposition 2.8.1. Let $b_\Phi(\omega)$ be the function defined in Proposition 2.8.1. Then,

$$B_\Phi(d\omega) := b_\Phi(\omega)d\omega$$

is called the Bartlett spectral measure (or Bartlett spectrum) of Φ. Note that the Bartlett spectrum can be defined for a wider class of point processes, called wide sense stationary processes (see e.g. [27]). The function $b_\Phi(\omega)$ is the density of the Bartlett spectrum.

2.8.2 Discrete Fourier Transform

Let Φ be a stationary point process, satisfying the conditions of the previous section. Let $s > 0$ be some positive real value, and $W_s := [-s/2, s/2[^2$. The Discrete Fourier Transform of Φ on W_s is defined, for any $m \in 2\pi\mathbb{Z}^2$, as

$$F_m(\Phi) := \int_{W_s} e^{-iu.\frac{m}{s}}\Phi(du). \tag{2.24}$$

Note that, by Campbell's averaging formula (Theorem 2.2.2), if $m = (0, 0)$, $F_m(\Phi) = \Phi(W_s)$, otherwise $\mathbb{E}[F_m(\Phi)] = 0$. We shall study

$$U_{m,s}(\Phi) := \frac{1}{|W_s|}|F_m(\Phi)|^2 \tag{2.25}$$

as an estimator of the Bartlett spectrum of Φ.

Proposition 2.8.3. *Let Φ be a stationary point process, satisfying the conditions of Proposition 2.8.1. Let $B_\Phi(d\omega) := b_\Phi(\omega)d\omega$ be its Bartlett spectrum. Let $U_{m,s}$ defined as in eq.* (2.25)*. Then,*

$$\mathbb{E}[U_{m,s}(\Phi)] = \int_{\mathbb{R}^2} G_s(\frac{m}{s} - \omega)b_\Phi(\omega)d\omega = G_s \star b_\Phi(\frac{m}{s}),$$

with[3] $G_s(u) := \frac{1}{(2\pi)^2}s^2 \operatorname{sinc}^2(\frac{s}{2}u_1)\operatorname{sinc}^2(\frac{s}{2}u_2)$, *and* $\operatorname{sinc}(u) := \frac{\sin(u)}{u}$*. Furthermore, if b_Φ is continuous at $\xi \in E$, then, if $m \to \infty$, $s \to \infty$, such that $\frac{m}{s} = \xi \neq 0$, $\mathbb{E}[U_{m,s}(\Phi)] \to b_\Phi(\xi)$.*

Proof. We write, using Proposition 2.8.1, and the fact that $\mathbb{E}[U_{m,s}] = 0$ for $m \neq 0$,

$$\begin{aligned}\mathbb{E}[U_{m,s}(\Phi)] &= \frac{1}{|W_s|}\operatorname{Var}(F^*_m(\Phi)) \\ &= \frac{1}{(2\pi s)^2}\int_{\mathbb{R}^2}|\hat{f}_m|^2(\omega)b_\Phi(\omega)d\omega,\end{aligned}$$

with f_m being the product of a complex exponential and the indicator function of the window W_s, so $\hat{f}_m(\omega) = s^2 \operatorname{sinc}(\frac{s}{2}(\omega_1 - \frac{m_1}{s}))\operatorname{sinc}(\frac{s}{2}(\omega_2 - \frac{m_2}{s}))$, where $\omega = (\omega_1, \omega_2)$, and $m = (m_1, m_2)$. As the sinc function is pair, we obtain

$$\mathbb{E}[U_{m,s}(\Phi)] = \int_{\mathbb{R}^2} G_s(\frac{m}{s} - \omega)b_\Phi(\omega)d\omega = G_s \star b_\Phi(\frac{m}{s}).$$

that is, $\mathbb{E}[U_{m,s}(\Phi)]$ is the convolution of the Bartlett spectral measure with the function G_s, evaluated at $\frac{m}{s}$.

---

[3] G_s is the square modulus of the Fourier transform of $\mathbb{1}_{W_s}$.

Suppose now that b_Φ is continuous at $\xi \in E$. Recall that $\int_E G_s(\xi - \omega)d\omega = 1$. Then, let V_ϵ be a neighbourhood of ξ such that, for any $\omega \in V_\epsilon$, $|b_\Phi(\omega) - b_\Phi(\xi)| \leq \epsilon$. We can write

$$\begin{aligned}|G_s \star b_\Phi(\xi) - b_\Phi(\xi)| &\leq |\int_{V_\epsilon} G_s(\xi - \omega)(b_\Phi(\omega) - b_\Phi(\xi))d\omega| \\ &\quad + |\int_{V_\epsilon^c} G_s(\xi - \omega)(b_\Phi(\omega) - b_\Phi(\xi))d\omega| \\ &\leq \epsilon \int_{V_\epsilon} G_s(\xi - \omega)d\omega \\ &\quad + |\int_{V_\epsilon^c} G_s(\xi - \omega)(b_\Phi(\omega) - b_\Phi(\xi))d\omega|,\end{aligned}$$

with $\int_{V_\epsilon} G_s(\xi - \omega)d\omega \leq 1$, and

$$|\int_{V_\epsilon^c} G_s(\xi - \omega)(b_\Phi(\omega) - b_\Phi(\xi))d\omega| \leq |\int_{V_\epsilon^c} G_s(\xi - \omega)b_\Phi(\omega)d\omega| + |b_\Phi(\xi)\int_{V_\epsilon^c} G_s(\xi - \omega)d\omega|. \tag{2.26}$$

Moreover, for any $s \geq 1$

$$\begin{aligned}G_s(\xi - \omega)\mathbb{1}_{V_\epsilon^c}(\omega) &\leq \frac{1}{s^2(\xi_1 - \omega_1)^2(\xi_2 - \omega_2)^2}\mathbb{1}_{V_\epsilon^c}(\omega) \\ &\leq \frac{1}{(\xi_1 - \omega_1)^2(\xi_2 - \omega_2)^2}\mathbb{1}_{V_\epsilon^c}(\omega) \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2),\end{aligned}$$

and $G_s(\xi - \omega) \xrightarrow[s\to\infty]{} 0$. Therefore, the dominated convergence theorem applies. ☐

Example 2.8.4 (Cox-circles point process)**.** *The Bartlett spectrum can be calculated for the Cox point process with Poisson points of linear intensity* λ' *on circles in* $\mathbb{R}^2$ *of radius* R *and the (total) intensity of points denoted by* λ *(cf. [37, Example 10.6]).*

$$\kappa_{BC}(u) = \lambda + \begin{cases} \frac{2\lambda'}{\pi|u|}\frac{R}{\sqrt{4R^2 - |u|^2}} & \text{for } |u| \leq 2R \\ 0 & \text{otherwise,}\end{cases}$$

yielding

$$b_{BC}(\omega) = \lambda(1 + 2\pi R\lambda' J_0^2(R|\omega|)), \tag{2.27}$$

where J_0 *is the* 0*-th Bessel function of the first kind. Figure 2.2 shows the theoretical Bartlett spectrum, as well as its DFT estimator, averaged over 10 realizations in a window of size 1, for frequencies (averaged in angles) such that* $|\omega| \in 2\pi\mathbb{Z}^2$*.*

2.9 Spatial statistics and summary characteristics

In spatial statistics, one is dealing with the observation of one or several collections of points (or more generally signed measures), lying on a bounded window. These collections of points, called point patterns, or configurations, are assumed to be realizations of a point process (or a random signed measure). In order to perform a statistical analysis of the point patterns, it is often necessary to extract important information about this data by representing it in a relatively concise way. In other words, one usually define

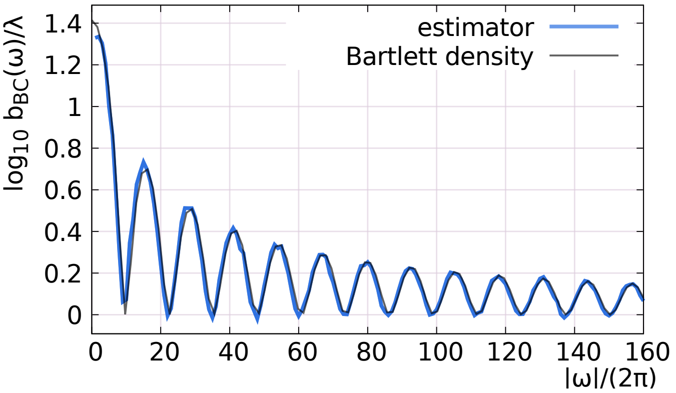


Figure 2.2: Theoretical (black full line) and estimated (blue dotted line) Bartlett spectrum for the Cox-circles point process with parameters $R = 10/256$ and $2\pi R\lambda' = 25$. The density of the Bartlett spectrum is normalized by the total intensity λ, thus, by (2.27), the curve is invariant to the density of circle centers (a particular numerical scenario considered here is presented in Figure 1.1 left). Observe in particular, $\log_{10} b_{\mathrm{BC}}(0)/\lambda = \log_{10}(1 + 2\pi R\lambda') = \log_{10}(26) \approx 1.415$.

several functionals of the point patterns, that shall be considered as statistical estimators of moments of the underlying point process, that aim to characterize (fully or partially) its distribution (in Section 2.10, we saw how such estimators could be used to define a model for the underlying distribution). For example, when observing a point pattern ϕ in a window W, a very simple functional that can be considered is the average number of points per volume unit in ϕ, i.e. $f(\phi) = \frac{1}{|W|}\phi(W)$, where $|W|$ denotes the volume of W. Here, if ϕ is assumed to be a realization of a stationary point process Φ, then $f(\phi)$ (strictly speaking $f(\Phi)$) can be viewed as an estimator of the intensity of Φ. This value alone obviously doesn't fully characterize the process Φ, and there exist many other classic functionals in the literature on point processes, called summary characteristics, that capture important geometric information about point patterns. In this section, we shall briefly review several of them, which are the most broadly used in the literature. While these functions can be defined for point processes on $\mathbb{R}^d$, for any $d \geq 1$, we shall focus on the planar case ($d = 2$).

2.9.1 Second-order statistics

A natural way to extract information beyond the mere intensity of a point process is to look at second-order moments, that capture the pairwise interactions between the atoms of the process. To this end, on can consider Ripley's K-function ([96]), defined for a stationary process Φ with intensity $0 < \lambda < \infty$, as

$$K_\Phi(r) := \frac{1}{\lambda}\mathbb{E}^0[\Phi(B_0(r)) - 1], \quad r \geq 0,$$

which counts the average number of points present in a ball of radius r around a typical point of Φ. Alternatives to Ripley's K-function are Besag's L-function ([13]), defined as $L_\Phi(r) := \sqrt{\frac{K_\Phi(r)}{\pi}}$, and the pair correlation function $g_\Phi(r) := \frac{K'_\Phi(r)}{2\pi r}$. These three functions offer different representations of the same information about the process. For more details about these functions and their comparisons, see for example [64], which also presents many other summary characteristics, among them the ones we shall present in the next section. Note that the Discrete Fourier Transform, defined in Section 2.8.2, also belongs to the family of second-order statistics.

2.9.2 Empty set and nearest neighbour statistics

Other statistics can be defined, relying on the size of sets that do not contain any points, or on the distance between neighbouring points. The two most commonly found statistics are the spherical contact distribution function (d.f.), defined as

$$H_s(r) := \mathbb{P}(\Phi(B_0(r)) > 0), \quad r \geq 0, \tag{2.28}$$

and the nearest neighbour distance d.f.,

$$D(r) := \mathbb{P}^0(\Phi(B_0(r)) > 1), \quad r \geq 0. \tag{2.29}$$

Both the spherical contact d.f. and the nearest neighbour d.f. measure the probability that at least one point is encountered in the ball of radius r, either centered at a typical point of the plane (for H_s) or at a typical point of Φ (for D). The nearest neighbour distance d.f. can be extended to an arbitrary number of neighbours:

$$D_k(r) := \mathbb{P}^0(\Phi(B_0(r)) > k), \quad r \geq 0,\ k \geq 1. \tag{2.30}$$

Remark. Note that these functions are defined by mathematical expectations, and are therefore theoretical. We find in the literature many estimators for these statistics, see e.g. [64].

2.10 Maximum entropy models

In Section 2.3.2, we presented a few classical distributions of point processes. When dealing with observed data, it is often useful to try to fit a model (usually a well known distribution) to this data. However, one can find itself in the situation where classic parametric models fail to represent the data in a satisfying way. Suggested by statistical physics, maximum entropy models have been introduced in [65] to build models of probability distributions, based on partial knowledge of it, for instance by estimating moments of functionals from the observed data (in the case of point processes, one can use statistics called summary characteristics, that we present in Section 2.9). Intuitively, this means that the model is 'as random as possible' under some prescribed constraints. Two standard models are the macro-canonical model, with expectations constraints, and the micro-canonical model, with path-wise constraints. We review these two models in the following sections.

2.10.1 Macro-canonical model

Let us fix $0 < \lambda < \infty$, and denote $\mathcal{L}_0$ a Poisson distribution on $(E, \mathcal{B})$. We define

$$\mathbb{M}_0 := \{\mathcal{L} : \mathcal{L} \ll \mathcal{L}_0\},$$

where '$\ll$' means 'absolutely continuous with respect to'.

The notion of entropy is well defined for discrete random variables. It can be extended to stochastic processes on Euclidean spaces by the notion of differential entropy, which is the Kullback-Leibler divergence ([70]) with respect to the Lebesgue measure. In this work, we shall use the following definition of entropy for point processes.

Definition 2.10.1 (Kullback-Leibler divergence and entropy of a point process)**.** Let $\mathcal{L}_1, \mathcal{L}_2$ be two probability distributions on $\mathbb{M}$, such that $\mathcal{L}_1 \ll \mathcal{L}_2$. The Kullback-Leibler divergence (or KL divergence) of $\mathcal{L}_1$ w.r.t. $\mathcal{L}_2$ is defined by

$$\mathrm{KL}(\mathcal{L}_1 || \mathcal{L}_2) := \int_{\mathbb{M}} \log(\frac{d\mathcal{L}_1}{d\mathcal{L}_2}) d\mathcal{L}_1.$$

It is well defined, as shown in Lemma 2.10.2. If $\mathcal{L}_0$ is the Poisson distribution on $(E, \mathcal{B})$, and $\mathcal{L} \ll \mathcal{L}_0$, then

$$\mathcal{H}(\mathcal{L}) := -\mathrm{KL}(\mathcal{L} || \mathcal{L}_0)$$

is called the entropy of $\mathcal{L}$.

We shall need the following result:

Lemma 2.10.2. *Let $\mathcal{L}_1, \mathcal{L}_2$ be two probability distributions on $\mathbb{M}$, such that $\mathcal{L}_1 \ll \mathcal{L}_2$. Then,*

$$KL(\mathcal{L}_1 \| \mathcal{L}_2) \geq 0, \tag{2.31}$$

and the inequality becomes an equality if and only if $\mathcal{L}_1 = \mathcal{L}_2$.

Proof. It follows from Jensen's inequality. By definition,

$$1 = \mathbb{E}_{\mathcal{L}_1}[1] = \int_{\mathbb{M}} d\mathcal{L}_1 = \int_{\mathbb{M}} \frac{d\mathcal{L}_1}{d\mathcal{L}_2} d\mathcal{L}_2 = \mathbb{E}_{\mathcal{L}_2}[\frac{d\mathcal{L}_1}{d\mathcal{L}_2}].$$

The function $f(u) = u\log(u)$ being convex, $0 = f(\mathbb{E}_{\mathcal{L}_2}[\frac{d\mathcal{L}_1}{d\mathcal{L}_2}]) \leq \mathbb{E}_{\mathcal{L}_2}[f(\frac{d\mathcal{L}_1}{d\mathcal{L}_2})]$, with

$$\begin{aligned}
\mathbb{E}_{\mathcal{L}_2}[f(\frac{d\mathcal{L}_1}{d\mathcal{L}_2})] &= \mathbb{E}_{\mathcal{L}_2}[\log(\frac{d\mathcal{L}_1}{d\mathcal{L}_2})\frac{d\mathcal{L}_1}{d\mathcal{L}_2}] \\
&= \int_{\mathbb{M}} \log(\frac{d\mathcal{L}_1}{d\mathcal{L}_2})\frac{d\mathcal{L}_1}{d\mathcal{L}_2} d\mathcal{L}_2 \\
&= \int_{\mathbb{M}} \log(\frac{d\mathcal{L}_1}{d\mathcal{L}_2}) d\mathcal{L}_1 \\
&= \mathrm{KL}(\mathcal{L}_1 \| \mathcal{L}_2).
\end{aligned}$$

Since the function f is strictly convex, we have an equality iff $\frac{d\mathcal{L}_1}{d\mathcal{L}_2} = 1$ $\mathcal{L}_2$-almost everywhere. ☐

Let $K : \mathbb{M} \mapsto \mathbb{R}^N$, for some $N \geq 1$, be a vector of statistics, and $a \in \mathbb{R}^N$. Note

$$\mathcal{A} := \{\mathcal{L} \in \mathbb{M}_0 : \mathbb{E}_{\mathcal{L}}[K] = a\}.$$

We are interested in solutions, if any exists, to the following problem:

$$\mathcal{L}^* \in \arg\max_{\mathcal{L} \in \mathcal{A}} \mathcal{H}(\mathcal{L}). \tag{2.32}$$

We are going to show the following result:

Proposition 2.10.3. *Suppose there exists some* $\gamma \in \mathbb{R}^N$ *such that* $0 < Z_\gamma := \mathbb{E}_{\mathcal{L}_0}[e^{\gamma.(K-a)}] < \infty$*, and define* $\mathcal{L}_\gamma$ *such that*

$$\frac{d\mathcal{L}_\gamma}{d\mathcal{L}_0} = \frac{1}{Z_\gamma} e^{\gamma.(K-a)}.$$

If $\gamma.(K-a)$ *is integrable w.r.t* $\mathcal{L}_\gamma$*, and* $\mathbb{E}_{\mathcal{L}_\gamma}[K-a] = 0$*, then* $\mathcal{L}_\gamma$ *is the unique solution of eq.* (2.32)*.*

Proof. Let $\mathcal{L} \in \mathcal{A}$. First, note that $\mathcal{L} \ll \mathcal{L}_\gamma$. Indeed, for any $\Gamma \in \mathcal{M}$, $\mathcal{L}_\gamma(\Gamma) = \frac{1}{Z_\gamma} \int_\Gamma e^{\gamma.(K-a)} d\mathcal{L}_0$, with $e^{\gamma.(K(\mu)-a)} > 0$ for all $\mu \in \mathbb{M}$. Therefore, if $\mathcal{L}_\gamma(\Gamma) = 0$, then $\mathcal{L}_0(\Gamma) = 0$ which implies $\mathcal{L}(\Gamma) = 0$.

Then,

$$\begin{aligned}
-\mathcal{H}(\mathcal{L}) &= \int_{\mathbb{M}} \log(\frac{d\mathcal{L}}{d\mathcal{L}_0}) d\mathcal{L} \\
&= \int_{\mathbb{M}} \log(\frac{d\mathcal{L}}{d\mathcal{L}_\gamma}\frac{d\mathcal{L}_\gamma}{d\mathcal{L}_0}) d\mathcal{L} \\
&= \int_{\mathbb{M}} \log(\frac{d\mathcal{L}}{d\mathcal{L}_\gamma}) d\mathcal{L} + \int_{\mathbb{M}} \log(\frac{d\mathcal{L}_\gamma}{d\mathcal{L}_0}) d\mathcal{L} \\
&= \mathrm{KL}(\mathcal{L}\|\mathcal{L}_\gamma) + \log(\frac{1}{Z_\gamma}) \int_{\mathbb{M}} d\mathcal{L} + \gamma. \int_{\mathbb{M}} (K-a) d\mathcal{L} \\
&= \mathrm{KL}(\mathcal{L}\|\mathcal{L}_\gamma) - \log(Z_\gamma) \\
&\geq -\log(Z_\gamma), \qquad (2.33)
\end{aligned}$$

with the last inequality given by Lemma 2.10.2 and the fact that $\mathcal{L} \ll \mathcal{L}_\gamma$. Remark that

$$\begin{aligned}
\mathrm{KL}(\mathcal{L}_\gamma, \mathcal{L}_0) &= \int_{\mathbb{M}} \log(\frac{d\mathcal{L}_\gamma}{d\mathcal{L}_0}) d\mathcal{L}_\gamma \\
&= \int_{\mathbb{M}} \log(\frac{1}{Z_\gamma} e^{\gamma.(K-a)}) d\mathcal{L}_\gamma \\
&= -\log(Z_\gamma) + \gamma.\mathbb{E}_{\mathcal{L}_\gamma}[K-a]. \qquad (2.34)
\end{aligned}$$

Inserting eq. (2.34) in eq. (2.33) gives us

$$-\mathcal{H}(\mathcal{L}) \geq \mathrm{KL}(\mathcal{L}_\gamma\|\mathcal{L}_0) - \gamma.\mathbb{E}_{\mathcal{L}_\gamma}[K-a]. \qquad (2.35)$$

It follows from eq. (2.35) and the assumption $\mathbb{E}_{\mathcal{L}_\gamma}[K-a] = 0$ that $\mathcal{H}(\mathcal{L}) \leq \mathcal{H}(\mathcal{L}_\gamma)$, with equality iff $\mathrm{KL}(\mathcal{L}\|\mathcal{L}_\gamma) = 0$, that is, $\mathcal{L} = \mathcal{L}_\gamma$. ☐

Under some conditions, we can further express the solution of eq. (2.32) as the minimum of Z_γ.

Proposition 2.10.4. *Suppose that, for all* $\gamma \in \mathbb{R}^{*N}$, $\mathbb{E}_{\mathcal{L}_0}[e^{\gamma.(K-a)}] < \infty$*, and* $\mathbb{P}_{\mathcal{L}_0}(\gamma.(K-a) \neq 0) > 0$*. Then*

$$\exists!\, \gamma^* \in \arg\min_{\gamma \in \mathbb{R}^N} Z_\gamma.$$

Moreover, suppose that there exists a neighbourhood V *of* γ^* *in which* $\gamma \mapsto Z_\gamma$ *is differentiable, and* $\nabla Z_\gamma = \mathbb{E}_{\mathcal{L}_0}[(K-a)e^{\gamma.(K-a)}] < \infty$*. Then,*

$$\gamma^* \in \arg\min_{\gamma \in \mathbb{R}^N} Z_\gamma \Leftrightarrow \mathcal{L}_{\gamma^*} \in \arg\max_{\mathcal{L} \in \mathcal{A}} \mathcal{H}(\mathcal{L}).$$

Proof. If for all $\gamma \in \mathbb{R}^N$, $\mathbb{E}_{\mathcal{L}_0}[e^{\gamma\cdot(K-a)}] < \infty$, then $\gamma \mapsto Z_\gamma$ is defined on $\mathbb{R}^N$. It is easy to see that this function is convex. If $\forall\, \gamma \in \mathbb{R}^N$, $\mathbb{P}_{\mathcal{L}_0}(\gamma.(K-a) \neq 0) > 0$ (implied e.g. by $Z_\gamma \neq 1$), the function is strictly convex, and the uniqueness of the minimum follows directly. If the function is differentiable on a neighbourhood of γ^*, then γ^* is the unique solution of $\nabla Z_\gamma = 0$. Moreover, as $\nabla Z_\gamma = \mathbb{E}_{\mathcal{L}_0}[(K-a)e^{\gamma\cdot(K-a)}]$, $\nabla Z_\gamma = 0$ is equivalent to $0 = \mathbb{E}_{\mathcal{L}_0}[(K-a)e^{\gamma^*\cdot(K-a)}] = Z_{\gamma^*}\mathbb{E}_{\mathcal{L}_{\gamma^*}}[K-a]$. We know that $Z_{\gamma^*} > 0$, which gives us the result. ☐

2.10.2 Micro-canonical model

The micro-canonical model is defined similarly to the macro-canonical model, but with path-wise constraints rather than constrains in expectations. More formally, let us fix some $\epsilon > 0$, and define

$$A := \{\mu \in \mathbb{M} : \|K(\mu) - a\| \leq \epsilon\}.$$

Suppose that $\mathbb{P}_{\mathcal{L}_0}(A) > 0$, and define

$$\mathcal{M}_A := \{\mathcal{L} \ll \mathcal{L}_0 : \mathbb{P}_{\mathcal{L}}(A) = 1\}.$$

We are interested in solutions, if any exists, to the following problem:

$$\mathcal{L}^* \in \arg\max_{\mathcal{L}\in\mathcal{M}_A} \mathcal{H}(\mathcal{L}). \tag{2.36}$$

Proposition 2.10.5. *The solution to eq.* (2.36) *is given by $\mathcal{L}^*$ such that*

$$\frac{d\mathcal{L}^*}{d\mathcal{L}_0} = \frac{\mathbb{1}_A}{\mathbb{P}_{\mathcal{L}_0}(A)}.$$

Proof. It is direct to show that $\mathcal{L}^* \in \mathcal{M}_A$. Note that

$$\begin{aligned}
\mathrm{KL}(\mathcal{L}^*||\mathcal{L}_0) &= \int_A \log(\frac{\mathbb{1}_A}{\mathbb{P}_{\mathcal{L}_0}(A)})d\mathcal{L}^* \\
&= \int_A \log(\frac{\mathbb{1}_A}{\mathbb{P}_{\mathcal{L}_0}(A)})\frac{\mathbb{1}_A}{\mathbb{P}_{\mathcal{L}_0}(A)}d\mathcal{L}_0 \\
&= -\log(\mathbb{P}_{\mathcal{L}_0}(A))\frac{1}{\mathbb{P}_{\mathcal{L}_0}(A)}\int_A d\mathcal{L}_0 \\
&= -\log(\mathbb{P}_{\mathcal{L}_0}(A)).
\end{aligned}$$

Now, let $\mathcal{L} \in \mathcal{A}$. For any $\Gamma \in \mathcal{M}$, $\mathbb{P}_{\mathcal{L}^*}(\Gamma) = 0 \Rightarrow \mathbb{P}_{\mathcal{L}_0}(\Gamma \cap A) = 0 \Rightarrow \mathbb{P}_{\mathcal{L}}(\Gamma \cap A) = 0 \Rightarrow \mathbb{P}_{\mathcal{L}}(\Gamma) = 0$, so $\mathcal{L} \ll \mathcal{L}^*$. One can then easily verify that

$$\frac{d\mathcal{L}}{d\mathcal{L}_0} = \frac{d\mathcal{L}}{d\mathcal{L}^*}\frac{\mathbb{1}_A}{\mathbb{P}_{\mathcal{L}_0}(A)}, \quad \mathcal{L}\text{-a.s.}$$

Then,

$$\begin{aligned}
\mathrm{KL}(\mathcal{L}||\mathcal{L}_0) &= \int_A \log(\frac{d\mathcal{L}}{d\mathcal{L}_0})d\mathcal{L} \\
&= \int_A \log(\frac{d\mathcal{L}}{d\mathcal{L}^*}\frac{\mathbb{1}_A}{\mathbb{P}_{\mathcal{L}_0}(A)})d\mathcal{L} \\
&= \mathrm{KL}(\mathcal{L}||\mathcal{L}^*) + \int_A \log(\mathbb{1}_A)d\mathcal{L} - \log(\mathbb{P}_{\mathcal{L}_0}(A))\int_A d\mathcal{L} \\
&= \mathrm{KL}(\mathcal{L}||\mathcal{L}^*) - \log(\mathbb{P}_{\mathcal{L}_0}(A)) \\
&= \mathrm{KL}(\mathcal{L}||\mathcal{L}^*) + \mathrm{KL}(\mathcal{L}^*||\mathcal{L}_0) \\
&\geq \mathrm{KL}(\mathcal{L}^*||\mathcal{L}_0).
\end{aligned}$$

$\square$

Maximum entropy models are defined by the description of the distribution, through the function K. This function, or *descriptor*, can be chosen, for instance, to be the summary characteristics mentioned in Section 2.9. In the next chapter, we shall study another class of descriptors, defined by a multi-scale representation of the geometry, using wavelets.

Chapter 3

Wavelet-based representations

As previously described, the aim of summary characteristics in spatial statistics is to concisely describe the main geometric features of a point pattern, for instance its tendency to exhibit clusters of points, or to form more regular patterns. Applied to realizations of a point process, they allow one to infer information about its distribution, by estimating their expected values. They can, for instance, be used to regress values associated to a point pattern, or to define a maximum entropy model.

Each statistic presented in Section 2.9 describe a particular aspect of a point process, and none of them fully characterizes its distribution. For instance, it is known that second order statistics do not characterize the distribution of a point process. For simple point processes, the void probability of all Borel sets characterize its distribution, but considering only balls may not be enough, especially for anisotropic distributions. Depending on the data and the task at hand, it is therefore a challenge to find the appropriate representation.

In this work, we present a class of representations of stationary point processes, constructed from a family of shot-noise processes, i.e. from a family

$$\{\int \psi_\xi(u - v)\Phi(du)\}_{\xi \in \Xi},$$

where the ψ_ξ are measurable, integrable functions, and Ξ shall be specified later. This family, that we shall present in the next chapter, is designed to separate different relevant information about the geometric structures of processes. More precisely, these functions are wavelets, which we shall introduce in the following section.

3.1 Wavelet transform

Informally speaking, a wavelet ψ is a function averaging to zero, that is localized both is the spatial and the frequency domain (meaning that both ψ and $\hat{\psi}$ have most of their energy concentrated in compact domains). A wavelet ψ can be constructed in such a way that, by products of dilations and translations of this wavelet, one obtains an orthonormal basis of $L^2(\mathbb{R}^2)$ (see e.g. [79]). Wavelets have been extensively studied in the literature (e.g. [53, 35, 60, 41, 42]), as they constitute a popular tool in many areas of science (see e.g. [59, 51, 79, 98]). In this section, we briefly recall the basic definitions of wavelets, and the wavelet transform, in the context of our work.

While the definition of wavelets differ depending on the context, we shall consider the following:

Definition 3.1.1 (Wavelet family)**.** Let $\psi : \mathbb{R}^2 \mapsto \mathbb{C}$ be a complex valued function such that $\psi \in L^2(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$, such that $\int_{\mathbb{R}^2} \psi(u)du = 0$, and $\|\psi\|_1 = 1$. For any $j \in \mathbb{Z}$, and any $\theta \in [0, 2\pi]$, define

$$\psi_{j,\theta}(u) := 2^{-2j}\psi(2^{-j} r_\theta u). \tag{3.1}$$

If the function ψ satisfies the following condition

$$\int_{\mathbb{R}^2} \frac{|\hat{\psi}|^2(\omega)}{|\omega|^2} d\omega < \infty, \tag{3.2}$$

then ψ is called a wavelet. For any discrete subset $\Theta \subset [0, 2\pi]$, the family of functions $\{\psi_{j,\theta}\}_{j\in\mathbb{Z},\theta\in\Theta}$ is called a wavelet family. The condition eq. (3.2) is called the *admissibility condition.* A wavelet family also includes a low-pass filer (i.e. local averaging function), that we shall denote ψ_0. It can be chosen, for instance, to be an isotropic Gaussian function.

Usually, one defines ψ such that $\hat{\psi}$ is centered at a frequency $\xi_0 \in \mathbb{R}^2$, and has fast decay in both the spatial and frequency domains. Informally, we can consider that there exists constants C, C', and ϵ such that $|\psi|(u) \leq \epsilon$ if $|u| > C$, and $|\hat{\psi}|(\omega) \leq \epsilon$ if $|\omega - \xi_0| > C'$. A wavelet family is obtained by dilation and rotation of the wavelet ψ (often called the mother wavelet). Let $\theta \in [0, 2\pi]$, r_θ the rotation by angle θ and $j \in \mathbb{Z}$. The wavelet at scale j and angle θ can also be indexed by its central frequency $\xi := 2^{-j} r_{-\theta}\, \xi_0$, that is

$$\psi_\xi(u) = 2^{-2j}\psi(2^{-j} r_\theta u) \quad \Rightarrow \quad \hat{\psi}_\xi(\omega) = \hat{\psi}(2^j r_\theta \omega).$$

Since $\hat{\psi}(\omega)$ is centered around ξ_0, it results that $\hat{\psi}_\xi(\omega)$ is centered around the frequency ξ. The wavelet ψ_ξ at scale j has negligible amplitude for $|u| > 2^j C$, and its Fourier transform $\hat{\psi}_\xi$ has negligible amplitude for ω such that $|\omega - \xi| > 2^{-j}C'$.

By indexing the wavelets by their central frequency, we shall sometimes use the notation $\{\psi_\xi\}_{\xi\in\Xi}$ for the wavelet family, where Ξ denotes the set of frequencies corresponding to $(j, \theta) \in \mathbb{Z} \times \Theta$.

Example 3.1.2 (Morlet wavelet)**.** *A classic example is the Morlet wavelet (see [58]), defined as*

$$\psi(u) = \frac{1}{2\pi\sigma^2} e^{-\frac{|u|^2}{2\sigma^2}} (e^{i\xi_0 . u} - c_\psi), \tag{3.3}$$

where c_ψ *is a constant such that* $\int_{\mathbb{R}^2} \psi(u)du = 0$*. Figure 3.1 shows the real and imaginary parts of a Morlet wavelet, for a given* ξ_0*, and Figure 3.2 displays a Morlet wavelet family in the spatial and frequency domains.*

Example 3.1.3 (Bump steerable wavelet)**.** *Another example of wavelet that we shall consider in this work is the bump steerable wavelet, introduced in [81]. It is defined in the frequency domain using polar coordinates. Let* L *be the even number of angles, and denote its central frequency by* $\xi_0 = (\omega_0, 0)$*. Then*

$$\hat{\psi}(\omega) = c_\psi e^{-\frac{(|\omega|-\omega_0)^2}{\omega_0^2 - (|\omega|-\omega_0)^2}} \mathbb{1}_{[0,2\omega_0]}(|\omega|) \cos^{\frac{L}{2}-1}(\varphi(\omega)) \mathbb{1}_{\varphi(\omega)<\frac{\pi}{2}}, \tag{3.4}$$

where c_ψ *is a normalizing constant. Figure 3.3 shows the real and imaginary parts of a bump steerable wavelet, for a given* ξ_0*, and Figure 3.4 displays a bump steerable wavelet family in the spatial and frequency domains.*

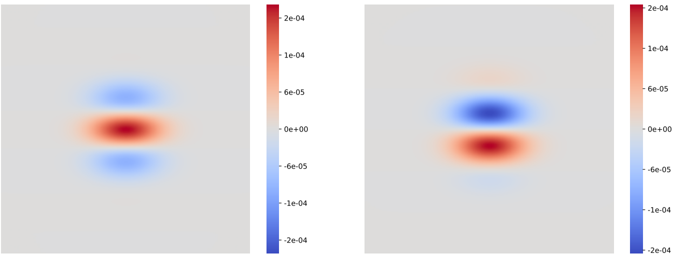


Figure 3.1: A Morlet wavelet. Left: Real part, right: imaginary part.

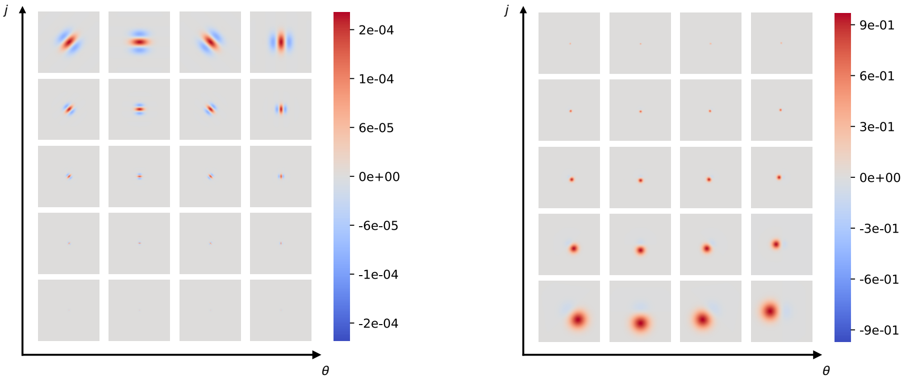


Figure 3.2: A Morlet wavelet family. Left: spatial domain (real part), right: frequency domain, centered around 0 (real part).

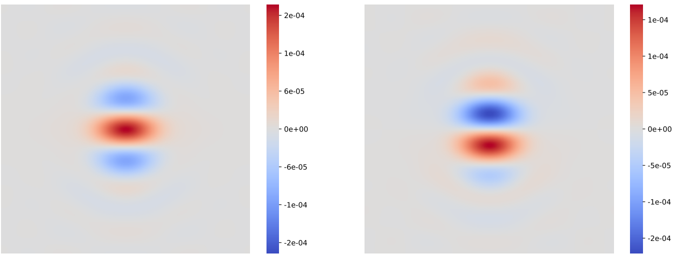


Figure 3.3: A bump steerable wavelet. Left: Real part, right: imaginary part.

Definition 3.1.4 (Wavelet transform)**.** Let μ be a counting measure on $(\mathbb{R}^2, \mathcal{B})$, and $\Psi = \{\psi_0, \psi_{j,\theta}\}_{j\in\mathbb{Z},\theta\in[0,2\pi]}$ be a wavelet family such that for all $\psi \in \Psi$, and all $t \in \mathbb{R}^2$, $\psi(t - \cdot) \in L^1(\mathbb{R}^2, \mu)$. The wavelet transform of μ is defined by the convolutions

$$\mathcal{W}\mu := \{\mu \star \psi_0,\ \mu \star \psi_{j,\theta}\}_{j\in\mathbb{Z},\theta\in[0,2\pi]}, \tag{3.5}$$

with

$$\mu \star \psi_{j,\theta}(t) := \int_{\mathbb{R}^2} \psi_{j,\theta}(t-u)\mu(du), \tag{3.6}$$

and similarly for ψ_0. The family of functions $\mathcal{W}\mu$ defined by the wavelet transform of μ are called the *wavelet coefficients* (at scale j and orientation θ). In this work, for simplicity, we shall only consider wavelets verifying $\psi(u) = \psi^*(-u)$, for $u \in \mathbb{R}^2$.

For a random measure Φ, the wavelet transform is well defined if all $\psi \in \Psi$ are integrable with respect to the mean measure M_Φ of Φ. If Φ is stationary with finite

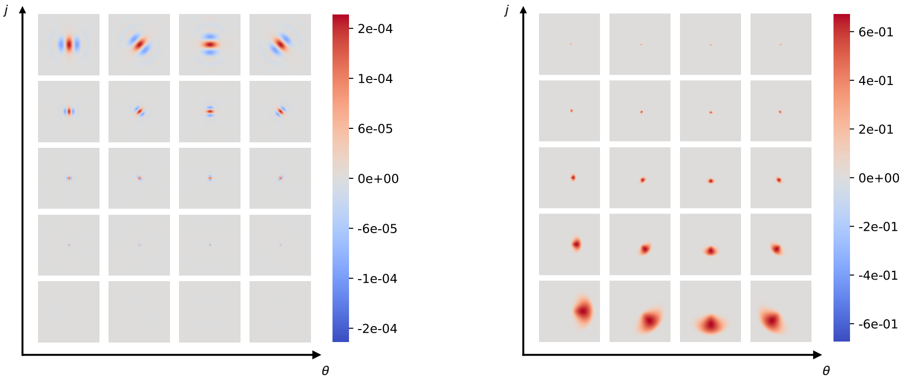


Figure 3.4: A bump steerable wavelet family. Left: spatial domain (real part), right: frequency domain, centered around 0 (real part).

non-null intensity, then the wavelet transform is always well defined, due to Campbell's averaging formula.

Note that, if Φ is a.s. finite, then for any $\omega \in \mathbb{R}^2$,

$$\widehat{\Phi \star \psi_{j,\theta}}(\omega) = \hat{\psi}_{j,\theta}(\omega)\hat{\Phi}(\omega),$$

where $\hat{\Phi}(\omega) = \frac{1}{2\pi}\int_{\mathbb{R}^2} e^{-2i\pi t.\omega}\Phi(dt)$. This highlights the fact that, if $\psi_{j,\theta}$ is localized in frequency, then so is $\Phi \star \psi_{j,\theta}$. In other words, the wavelet transform separates the information in Φ into components at different scales and orientations, or equivalently, in separate domains in the frequency plane.

3.2 Scattering moments

The wavelet transform is a linear operator that represents a measure by a family of measurable functions on $\mathbb{R}^2$. When performing statistical analysis or learning tasks, one is usually looking for non-linear representations of the input data. Moreover, if Φ is a stationary process (with non-null finite intensity λ), taking the expectation of the wavelet coefficients would give, for any $u \in \mathbb{R}^2$,

$$\mathbb{E}[\Phi \star \psi_{j,\theta}(u)] = \lambda \int_{\mathbb{R}^2} \psi_{j,\theta}(t)dt = 0.$$

This shows that, in order to extract meaningful information about the geometry of the process, one needs to apply a non-linear operator to the wavelet transform of Φ. This can be done, for instance, by applying the complex modulus operator to its wavelet coefficients. Then, taking the expectancy of the resulting functions gives us a description of the point process distribution. This defines the (first order) scattering transform of Φ, inspired by [80]:

Definition 3.2.1 (First order scattering moments)**.** Let Φ be a stationary point process (or random measure) on $(\mathbb{R}^2, \mathcal{B})$, and $\Psi = \{\psi_0, \psi_{j,\theta}\}_{j\in\mathbb{Z},\theta\in[0,2\pi]}$ be a wavelet family such all wavelet coefficients are almost surely integrable stochastic processes. The first order scattering moments of Φ, noted $S_1\Phi$, are defined, for all $j \in \mathbb{Z}$, $\theta \in [0, 2\pi]$, by

$$S_1\Phi(j,\theta) := \mathbb{E}[|\Phi \star \psi_{j,\theta}|], \tag{3.7}$$

which does not depend on $u \in \mathbb{R}^2$ by stationarity of Φ. Intuitively, the first order scattering moments compute 'how much geometric structure at scale j and orientation θ is present in Φ'. For example, if Φ exhibits a certain regularity at large scales (that is, from a certain scale, the process appears uniform, with no irregularities), then the larges scales scattering moments will have small values. Conversely, if Φ is composed of clusters of atoms, well separated from each other, then the values of the scattering moments at this scale will have large value.

If $S_1\Phi(j, \theta)$ detects the average regularity or clustering behaviour of Φ at a certain scale and orientation, on can enquire about the regularity or clustering of $S_1\Phi(j, \theta)$ at larger scales (i.e. the regularity/clustering of the regularity/clustering). This can be done by computing the scattering transform of $S_1\Phi(j, \theta)$. It defines the second order scattering moments of Φ.

Definition 3.2.2 (Second order scattering moments)**.** Let Φ be a stationary point process on $(\mathbb{R}^2, \mathcal{B})$. Let us denote $I := \{(j_1, j_2, \theta_1, \theta_2) \in \mathbb{Z}^2 \times [0, 2\pi]^2 : j_2 > j_1\}$. If, for all $(j_1, j_2, \theta_1, \theta_2) \in I$, the random process $|\Phi \star \psi_{j_1,\theta_1}| \star \psi_{j_2,\theta_2}$ is integrable, then Φ admits second order scattering moments, defined by

$$S_2\Phi(j_1, \theta_1, j_2, \theta_2) := \mathbb{E}\Big[||\Phi \star \psi_{j_1,\theta_1}| \star \psi_{j_2,\theta_2}|\Big], \tag{3.8}$$

for $(j_1, j_2, \theta_1, \theta_2) \in I$.

Note that we only select indices j_2 in I such that $j_2 > j_1$. This is heuristically explained by the fact that the modulus operator in the first order scattering transform 'kills' the high frequency modulations of the wavelet coefficients $\Phi \star \psi_{j_1,\theta_1}$. Therefore, the Fourier transform of the $|\Phi \star \psi_{j_1,\theta_1}|$ are roughly supported on the low frequencies, which implies that the coefficients $|\Phi \star \psi_{j_1,\theta_1}| \star \psi_{j_2,\theta_2}$ are negligible when $j_2 \le j_1$.

Remark. One can define high order scattering moments, by cascading convolutions with the wavelet family, composed with a modulus operator.

Additionally, the definition of the wavelet transform in Section 3.1 can be extended to stationary marked point processes, by defining, for some wavelet ψ, and any $v \in \mathbb{R}^2$,

$$\widetilde{\Phi} \star \psi(v) := \int_{\mathbb{R}^2 \times \mathbb{R}} \psi(v - u) m(u, \Phi) \Phi(du) \tag{3.9}$$

The wavelet transform of Section 3.1 is a particular case, where $m \equiv 1$. Similarly, under integrability conditions, one can extend the definition of first and second order scattering moments (Section 3.2) to (geometrically) marked point processes, and they do not depend on $u \in \mathbb{R}^2$.

3.3 Asymptotic properties of scattering moments

To gain more insights about what information this representation captures, we shall study the limit of first order scattering moments when the scale parameter j goes to $\pm\infty$. For simplicity, in this section we shall assume that the wavelets have compact support, included in some bounded set $B \in \mathcal{B}$, and are symmetric, i.e. $\psi(-u) = \psi(u)$[1]. The

---

[1] More precisely, the wavelets we shall consider, such as Morlet wavelets, verify $\check{\psi} = \psi^*$. The results remain valid for these wavelets, and we make the symmetric assumption for mere notation simplicity.

results are true for any $\theta \in [0, 2\pi]$, and for notation simplification, we omit the subscript θ in the following section.

We consider here random, possibly signed measures of the form (cf. Section 2.6)

$$\widetilde{\Phi} = \sum_{u \in \Phi} m(u, \Phi)\delta_u,$$

where Φ is a stationary point process, with intensity $0 < \lambda < \infty$, and the functions m are stationary, in the sense that, for any $u, v \in \mathbb{R}^2$,

$$m(u - v, S_v\Phi) = m(u, \Phi).$$

Typically, m, called a marking function, is a measurable function of the geometry of the points in Φ, such as the ones presented in what follows.

3.3.1 Small scale limit

The small scale limit behaviour ($j \to -\infty$) of first order scattering moments can be described by the following result.

Proposition 3.3.1. *Let Φ be a stationary point process on $\mathbb{R}^2$ admitting a reduced second-order density κ (cf. Definition 2.5.11), ψ be a bounded wavelet with support included in B, and a marking function m such that $\int_B \mathbb{E}^{0,u}[|m(0, \Phi)|]|\kappa(u)|du < \infty$. Then,*

$$S_1\widetilde{\Phi}(j) = \lambda\mathbb{E}^0[|m(0, \Phi)|] + O(\lambda 2^{2j+1}\|\psi\|_\infty \int_{B^2} \mathbb{E}^{0,2^j(v-u)}[|m(0, \Phi)|]\kappa(2^j(v - u))dudv).$$

Proof. We can write the following decomposition:

$$S_1\widetilde{\Phi}(j) := \mathbb{E}[|\int_{\mathbb{R}^2} \psi_j(u)\widetilde{\Phi}(du)|] = \mathbb{E}[\int_{\mathbb{R}^2} |\psi_j(u)||\widetilde{\Phi}|(du)] + e_j,$$

where $|\widetilde{\Phi}| := \sum_{u\in\Phi} |m(u, \Phi)|\delta_u$, and $e_j := S_1\widetilde{\Phi}(j) - \mathbb{E}[\int_{\mathbb{R}^2} |\psi_j(u)||\widetilde{\Phi}|(du)]$.

Remark also that $m(u, \Phi) = m(0, S_u\Phi) = m(0, \Phi \circ \theta_u)$. Then, as Φ is stationary, we can apply the CLMM theorem (eq. (2.19)):

$$\begin{aligned}\mathbb{E}[\int_{\mathbb{R}^2} |\psi_j(u)||\widetilde{\Phi}|(du)] &= \mathbb{E}[\int_{\mathbb{R}^2} |\psi_j(u)m(u, \Phi)|\Phi(du)] \\ &= \lambda \int_{\mathbb{R}^2} |\psi_j(u)|\mathbb{E}^0[|m(0, \Phi)|]du. \qquad \text{(i)}\end{aligned} \tag{3.10}$$

Furthermore, note that ψ_j is supported in $B_j := \{2^j u : u \in B\}$. Then,

$$\begin{aligned}
|e_j| &= |\mathbb{E}[|\int_{B_j} \psi_j(u)m(u,\Phi)\Phi(du)| - \int_{B_j} |\psi_j(u)m(u,\Phi)|\Phi(du)]| \\
&=: |\mathbb{E}[Z]| \\
&\leq \mathbb{E}[|Z|] \\
&= \mathbb{E}[|Z|\mathbb{1}_{\{\Phi(B_j)>1\}}] \\
&\leq 2\,\mathbb{E}[\int_{B_j} |\psi_j(u)m(u,\Phi)|\Phi(du)\mathbb{1}_{\{\Phi(B_j)>1\}}] \\
&\leq 2 \underbrace{||\psi_j||_\infty}_{2^{-2j}||\psi||_\infty} \mathbb{E}[\int_{B_j} |m(u,\Phi)|\Phi(du)\mathbb{1}_{\{\Phi(B_j)>1\}}]. \qquad \text{(ii)}
\end{aligned}$$

Then we write:

$$\begin{aligned}
\mathbb{E}[\int_{B_j} |m(u,\Phi)|\Phi(du)\mathbb{1}_{\{\Phi(B_j)>1\}}] &\leq \mathbb{E}[\int_{B_j} |m(u,\Phi)|\Phi(du)(\Phi(B_j)-1)] \\
&= \mathbb{E}[\int_{B_j\times B_j} |m(u,\Phi)|(\Phi-\delta_u)(dv)\Phi(du)] \\
&= \int_{B_j\times B_j} \mathbb{E}^{u,v}[|m(u,\Phi)|]\rho^{(2)}(u,v)dudv \\
&= \lambda\int_{B_j\times B_j} E^{0,v-u}[|m(u,S_u\Phi)|\kappa(v-u)dudv \\
&= \lambda\int_{B_j\times B_j} E^{0,v-u}[|m(0,\Phi)|\kappa(v-u)dudv \\
&= \lambda 2^{4j}\int_{B\times B} \mathbb{E}^{0,2^j(v-u)}[|m(0,\Phi)|]\kappa(2^j(v-u))dudv, \\
&\qquad \text{(iii)}
\end{aligned}$$

where the fourth line stems from the stationarity of Φ, and the fifth from the translation invariance property of m. Combining eqs. (i) to (iii) completes the proof. □

Remark. In the case of a Poisson point process with $f \equiv 1$ we find the result from [33]:

$$\bar{S}\Phi(j) = \lambda(||\psi||_1 + O_{j\to-\infty}(2^{2j}\lambda))$$

3.3.2 Large scale limit

We consider here the case $j \to +\infty$. The results of this section rely on some notion of central limit theorem, because the convolution of a wavelet ψ with a marked point process $\widetilde{\Phi}$ operates a sum over the atoms of $\widetilde{\Phi}$. Intuitively, as the scale of the wavelet increases, the sum extends to a wider subset of $\mathbb{R}^2$. Therefore, in order to apply the CLT, we need some notion of near independence of the positions of the atoms and their marks, as the distance between them increases. Hence, we need to introduce several definitions before stating the main result, based on [22].

Definition 3.3.2. Let $p \in [1,\infty)$, we say that the pair (m,Φ) satisfies the p-moment condition if

$$\sup_{1\leq p'\leq\lfloor p\rfloor}\ \sup_{u_1,\cdots,u_{p'}\in\mathbb{R}^d} \mathbb{E}^{u_1,\cdots,u_{p'}}[|m(u_1,\Phi)|^p] < \infty,$$

where sup signifies here *ess* sup with respect to $M_{\Phi^{(p)}}$.

Definition 3.3.3. We call the correlation functions of $\widetilde{\Phi}$ the functions defined for all $p \in \mathbb{N}$, and all $(k_1, \cdots, k_p) \in \mathbb{N}^p$ by:

$$m^{(k_1,\cdots,k_p)}(u_1,\cdots,u_p) := \mathbb{E}^{u_1,\cdots,u_p}[m(u_1,\Phi)^{k_1}\cdots m(u_p,\Phi)^{k_p}]\,\rho^{(p)}(u_1,\cdots,u_p)$$

Definition 3.3.4. The correlation functions of $\widetilde{\Phi}$ are said to be ω-mixing if there exists a decreasing function $\tilde{\omega} : \mathbb{N} \times \mathbb{R}^+ \to \mathbb{R}^+$ such that, $\forall\ n \in \mathbb{N}$, $\lim_{\|u\|\to\infty} \tilde{\omega}(n,u) = 0$ and, $\forall\, p,q \in \mathbb{N}$ s.t. $p < q, \forall\ (u_1,\cdots,u_q) \in \mathbb{R}^q$, $\forall\ (k_1,\cdots,k_q) \in \mathbb{N}^q$,

$$|m^{(k_1,\cdots,k_q)}(u_1,\cdots,u_q) - m^{(k_1,\cdots,k_p)}(u_1,\cdots,u_p)m^{(k_{p+1},\cdots,k_q)}(u_{p+1},\cdots,u_q)| \leq \tilde{\omega}(K,d),$$

where $K := \sum_{i=1}^{q} k_i$, and $d := \inf_{i\leq p,\, p<j\leq q} |u_j - u_i|$.

Definition 3.3.5. $\widetilde{\Phi}$ is said to have fast decay of correlations if its correlation functions are $\tilde{\omega}$-mixing, with $\tilde{\omega}(n,u) = \tilde{C}_n f(\tilde{c}_n u)$ for some rapidly decreasing function $f : \mathbb{R}^+ \to [0,1]$ and some constants $\tilde{c}_n, \tilde{C}_n \in \mathbb{R}^+$.

For a stationary point process Φ on $\mathbb{R}^2$ with intensity λ and a marking function m, we define

$$\sigma^2(m,\Phi) := \lambda\Big(\mathbb{E}^0[m^2(0,\Phi)] + \int_{\mathbb{R}^2} (\mathbb{E}^{0,u}[m(0,\Phi)m(u,\Phi)]\kappa(u) - \mathbb{E}^0[m(0,\Phi)]^2\lambda)du\Big)$$

provided the integral exists. We observe that, in the spirit of (2.22), $\sigma^2(m,\Phi) \geq 0$ is the total mass of the covariance measure of $\tilde{\Phi}$. For a wavelet function $\psi = \mathrm{Real}(\psi) + i\mathrm{Im}(\psi)$, denote by $\mathcal{N}(\psi)$ the complex Gaussian random variable with mean 0 and covariance matrix (between real and imaginary parts)

$$\Sigma(\psi) := \begin{pmatrix} \|\mathrm{Real}(\psi)\|_2^2 & \langle \mathrm{Real}(\psi), \mathrm{Im}(\psi)\rangle \\ \langle \mathrm{Real}(\psi), \mathrm{Im}(\psi)\rangle & \|\mathrm{Im}(\psi)\|_2^2 \end{pmatrix},$$

where $\langle\cdot,\cdot\rangle$ denotes the dot product in $L_2(\mathbb{R}^2)$.

Proposition 3.3.6. *Let Φ be stationary point process on $\mathbb{R}^2$ with intensity λ, and m a marking function, and ψ a compactly supported wavelet. Suppose that, for all $p \in (1,\infty)$, the pair (m,Φ) satisfies the p-moment condition, $\widetilde{\Phi}$ has fast decay of correlations. Then, $0 \leq \sigma^2(m,\Phi) < \infty$ and, if furthermore $\sigma^2(m,\Phi) > 0$, then*

$$\lim_{j\to\infty} 2^j S_1\widetilde{\Phi}(j) = \sigma(m,\Phi)\mathbb{E}[|\mathcal{N}(\psi)|]. \tag{3.11}$$

If the (complex) wavelet is such that its real and imaginary parts are orthogonal, with equal L_2 norm (e.g. for Morlet wavelets) then $\mathbb{E}[|\mathcal{N}(\psi)|] = \|\psi\|_2\frac{\sqrt{\pi}}{2}$.

Proof. Let us denote

$$\begin{aligned} Y_j &:= \psi_j \star \widetilde{\Phi}(0) = \int \psi_j d\widetilde{\Phi} \\ &= 2^{-2j} \sum_{u\in\Phi\cap 2^j B} \psi(2^{-j}u)m(u,\Phi). \end{aligned}$$

Considering the mother wavelet ψ as a test function (with support contained in a square window), from [22, Theorem 1.12 (ii)] we know that

$$\lim_{j\to\infty} 2^{2j}\mathrm{Var}(\mathrm{Real}(Y_j)) = \sigma^2(m,\Phi)\|\mathrm{Real}(\psi)\|_2^2 \tag{3.12}$$

and similarly for $\mathrm{Var}(\mathrm{Im}(Y_j))$. Using CLMM (Theorem 2.5.9),

$$\mathbb{E}[Y_j] = 2^{-2j}\lambda\mathbb{E}^0[m(0,\Phi)]\int_{\mathbb{R}^2}\psi(2^{-j}u)du = 0.$$

Hence, by [22, Theorem 1.13 with Remark (vii) *Multivariate central limit theorem*] (since ψ is a complex function)

$$2^jY_j \xrightarrow[j\to\infty]{d} \sigma(m,\Phi)\mathcal{N}(\psi).$$

We then conclude using the following lemma, from [40], for which the conditions are satisfied, due to the p-moment condition:

Lemma 3.3.7. *Let $(X_j)_{j\geq 1}$ be a sequence of random variables, such that $X_j \xrightarrow[j\to\infty]{d} X$. If, for some $\delta > 0$,*

$$\sup_j \mathbb{E}[|X_j|^{1+\delta}] < \infty,$$

then

$$\mathbb{E}[|X_j|] \xrightarrow[j\to\infty]{} \mathbb{E}[|X|].$$

For a specific wavelet ψ, such that $\langle\mathrm{Real}(\psi),\mathrm{Im}(\psi)\rangle = 0$ and $\|\mathrm{Real}(\psi)\|_2^2 = \|\mathrm{Im}(\psi)\|_2^2$, observe that real and imaginary parts of $\mathcal{N}(\psi)$ are independent Gaussians $\frac{\|\psi\|_2}{\sqrt{2}}\mathcal{N}(0,1)$. Consequently $|\mathcal{N}(\psi)| \sim \frac{\|\psi\|_2}{\sqrt{2}}\mathcal{R}(1)$, where $\mathcal{R}(1)$ is Rayleigh distribution (i.e. square root of $\chi^2(2)$ — χ-distribution with parameter 2) giving

$$\mathbb{E}[|\mathcal{N}(\psi)|] = \frac{\|\psi\|_2}{\sqrt{2}}\mathbb{E}[\mathcal{R}(1)] = \frac{\|\psi\|_2}{\sqrt{2}}\sqrt{\frac{\pi}{2}} = \|\psi\|_2\frac{\sqrt{\pi}}{2}.$$

□

Remark. In [22], the authors give several detailed conditions concerning the point process Φ and marking function m under which $\widetilde{\Phi}$ satisfies the fast decay of correlations and p-moment conditions, including $\sigma^2(m,\Phi) > 0$, (requested in Proposition 3.3.6). The processes satisfying these conditions include unmarked Poisson, Gaussian determinantal, and cluster Poisson processes with fixed range dependence (such as Matérn cluster processes, cf. Section 2.3.2), that we study in Example 3.3.8. These processes admit some local marking functions m, such as statistics of Voronoi cells, or k-nearest neighbors.

Other, more regular distributions (such as the Ginibre point process) are shown in [22] to satisfy the fast decay of correlations and p-moment conditions, but with $\sigma^2(m,\Phi) = 0$. In this case, it is possible to obtain the Central Limit Theorem under some technical conditions, and we refer the reader to [22] for more details.

There also exist Central Limit Theorem results for Y_i involving more clustering point process, not satisfying conditions of Proposition 3.3.6, such as intersections of Poisson line process (see [88, 62]) .

More generally, observe that the asymptotic of the variance of $Y_j = \int \psi_j d\widetilde{\Phi}$ depends on the wavelet ψ and the marked point $\tilde{\Phi}$. In the case of $m \equiv 1$, using Proposition 2.8.1, we can express the variance of Y_j as

$$2^{2j}\mathrm{Var}(Y_j) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}|\hat{\psi}(\omega)|^2 b_\Phi(2^{-j}\omega)d\omega, \tag{3.13}$$

where b_Φ is the density of the Bartlett spectrum of Φ. If Φ is a Poisson point process, then $b_\Phi(\omega) = 1$, confirming that $2^{2j}\mathrm{Var}(Y_j) \to \|\psi\|_2^2 b_\Phi(0)$, as shown in [33]. If b_Φ is continuous at 0, with $0 < b_\Phi(0) < \infty$ (e.g. Gaussian determinantal point processes and Matérn cluster processes) then we obtain the same limit of the variance $2^{2j}\mathrm{Var}(Y_j)$. If $\lim_{|\omega|\to 0} b_\Phi(\omega) = 0$ or $\lim_{|\omega|\to 0} b_\Phi(\omega) = \infty$, the asymptotic of $S_1\Phi(j)$ may be inferior or superior to 2^{-j} respectively, which the numerical experiments of Figure 3.6 indicate.

Example 3.3.8 (Poisson, Gaussian determinantal and Matérn cluster processes)**.** *Figure 3.5 illustrates asymptotics of scattering moments for three different distributions of* Φ *verifying* $\sigma^2(1, \Phi) > 0$ *with no marking* ($m = 1$)*. It shows the* $\log_2$ *of scattering moments (averaged in angle) for Poisson point processes of different intensities, a Gaussian determinantal point process, and a Matérn cluster process. We can observe that when* $j \to -\infty$ ($j = 0$ *is the smallest scale in our implementation), the scattering moments converge towards* $\log_2(\lambda)$ *(since* $\log_2(\|\psi\|_1) = 0$*), and when* $j \to \infty$*, the slope goes to -1, independently of the distribution of* Φ*. However the value at the origin of this affine limit changes from one distribution to another. Let us recall that this value is:* $a := \frac{1}{2}\log_2\left(\lambda\frac{\pi}{4}\|\psi\|_2^2(1 + \int_{\mathbb{R}^2}(\kappa(u) - \lambda)du)\right)$*. Recall from eq.* (2.22) *that the covariance measure of* Φ *is defined as* $\Gamma_\Phi(du) = \lambda\delta_0 + \lambda(\kappa(u) - \lambda)$*. Therefore*

$$a = \frac{1}{2}\log_2\left(\frac{\pi}{4}\|\psi\|_2^2\,\Gamma_\Phi(\mathbb{R}^2)\right)$$

measures the total mass of the covariance measure of Φ*, indicating the general repulsion of points* ($\Gamma_\Phi(\mathbb{R}^2) < \lambda$) *or their attraction* ($\Gamma_\Phi(\mathbb{R}^2) > \lambda$)*. In Figure 3.5, we observe that while the Gaussian determinantal, Poisson, and Matérn cluster point processes have the same intensity* $\lambda = 256$*, and the same slope of scattering moments when* j *goes to* ∞*, they exhibit different values of* a*, that are increasing according to the repulsion, Poisson-like independence, and clustering behaviour. For Poisson distributions,* $\int_{\mathbb{R}^2}(\kappa(u) - \lambda)du = 0$*, so* $a = \frac{1}{2}\log_2\lambda\frac{\pi}{4}\|\psi\|_2^2$*. In our simulations, using a Morlet mother wavelet with central frequency* $(3\pi/4, 0)$*,* $\|\psi\|_2 \simeq 169$*, giving for a Poisson point process with intensity* $\lambda = 256$*,* $a = \log_2(256\frac{\pi}{4}169) \simeq 11.2$*. For* $j = 7$*, we find* $\log_2(S_1\Phi(j)) \simeq 4.2$*, so* $j + \log_2(S_1\Phi(j)) \simeq 11.2$*.*

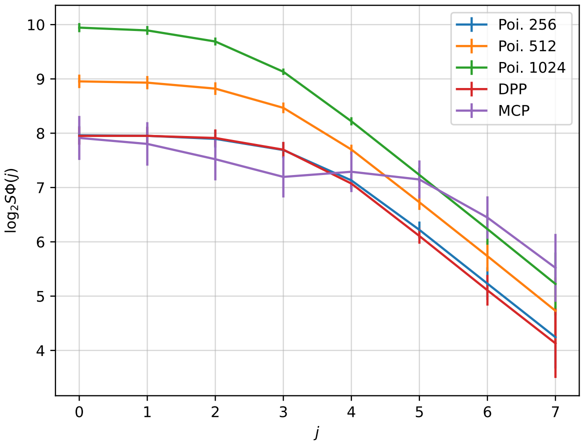


Figure 3.5: Asymptotics of scattering moments for Poisson, Gaussian determinantal (DPP), and Matérn cluster (MCP) processes, estimated on 500 realizations.

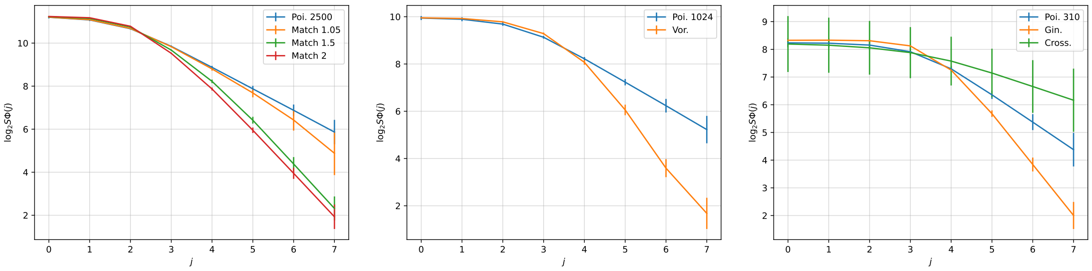


Figure 3.6: Asymptotics of scattering moments for various processes, estimated on 500 realizations. Left, a Poisson process is compared with a matching process, middle: with a marked Poisson process, with Voronoi area marks, right: with a Ginibre process and a crossing process.

Example 3.3.9 (Hyperuniform and hyper clustering processes)**.** *In Figure 3.6, we compare a Poisson distribution with different distributions that may not satisfy the assumptions of Proposition 3.3.6: a hyperuniform matching process (see [4]), a Ginibre point process (which we approximate using eigenvalues of Hermitian matrices with complex Gaussian entries, see e.g. [57]), a Poisson lines crossing process (the points of the process are the intersections of random lines uniformly and independently drawn in the window), and a Poisson point process geometrically marked with the areas of its Voronoi cells (cf. Example 2.6.4).*

3.4 Factorial moment expansion for scattering moments

In this section, we shall extend the result from [14], expressing functionals of a point process Φ through its factorial moments measures. In the context of first order scattering moments of marked point process, we consider functions of the form

$$F(\Phi) = |\int_{\mathbb{R}^2} \psi(u) m(u, \Phi) \Phi(du)|.$$

The idea of the factorial moment expansion is to express the function $F(\Phi)$ as a telescoping sum depending on the number of atoms in Φ. In analogy to the Taylor expansion for functions of Euclidean spaces, on can think of this telescoping sum as a sum of derivatives, expressed at the origin. However in our case, the origin is the point measure with no atoms, noted $\emptyset$. For some marking functions, it is not well defined to apply them to null, or even finite measures (think of the Voronoi cell of an atom with no points around it). To remedy this problem, we shall express such functions in the form $F(\Phi) = f(\Phi, \Phi)$, where f is a function from $\mathbb{M} \times \mathbb{M}$ to $\mathbb{R}$. This shall allow us to circumvent the telescoping for the value of the marking function.

For any $u \in E$, and $\mu \in \mathbb{M}$, we note $\mu_u \in \mathbb{M}$ the counting measure defined by:

$$\forall B \in \mathcal{B},\ \mu|_u(B) = \mu(B \cap B_0(u)),$$

where $B_0(u) = \{v \in \mathbb{R}^2 : \|v\| < \|u\|\}$. Let $f : \mathbb{M} \times \mathbb{M} \mapsto \mathbb{R}$. We say that f is continuous at ∞ if $\forall \nu, \mu \in \mathbb{M}$, and $\forall (u_n)_n \in (\mathbb{R}^2)^{\mathbb{N}}$ s.t. $\|u_n\| \xrightarrow[n\to\infty]{} \infty$, $f(\nu, \mu|_{u_n}) \xrightarrow[n\to\infty]{} f(\nu, \mu)$.

We define the first order differential operator of f at u by:

$$\begin{aligned} f_u^{(1)} : \mathbb{M} \times \mathbb{M} &\mapsto \mathbb{R} \\ (\nu, \mu) &\to f(\nu, \mu|_u + \delta_u) - f(\nu, \mu|_u). \end{aligned}$$

We recursively define the nth order differential operator by:

$$f^{(n)}_{u_1,\cdots,u_n}(\nu,\mu) = \left(f^{(n-1)}_{u_1,\cdots,u_{n-1}}\right)^{(1)}_{u_n}(\nu,\mu).$$

We remark that if f is continuous at ∞, then so is $f^{(n)}_{u_1,\cdots,u_n}$, for all $n \geq 1$ and $(u_1, \cdots, u_n) \in (\mathbb{R}^2)^n$ (see e.g [14, 18] for a proof of this fact).

Observe that, if $\mu = \sum_{i=1}^n \delta_{u_i}$, with $\|u_1\| \leq \cdots \leq \|u_n\|$, then, for any $\nu \in \mathbb{M}$,

$$\begin{aligned} f(\nu,\mu) &= f(\nu,\emptyset) + \sum_{i=1}^n f(\nu,\mu|_{u_i} + \delta_{u_i}) - f(\nu,\mu|_{u_i}) \\ &= f(\nu,\emptyset) + \int_{\mathbb{R}^2} (f(\nu,\mu|_u + \delta_u) - f(\nu,\mu|_u)\mu(du) \\ &= f(\nu,\emptyset) + \int_{\mathbb{R}^2} f_u^{(1)}(\nu,\mu)\mu(du). \end{aligned}$$

Now, let $\mu \in \mathbb{M}$, and f continuous at ∞ and such that $\int_E |f_u^{(1)}(\mu,\mu)|\mu(du) < \infty$. Then, for any $(v_n)_n \in E^n$ such that $\|v_n\| \xrightarrow[n\to\infty]{} \infty$, we can write

$$\begin{aligned} f(\nu,\mu) &= \lim_{n\to\infty} f(\nu,\mu|_{v_n}) \\ &= f(\nu,\emptyset) + \lim_{n\to\infty} \int_{\mathbb{R}^2} f_u^{(1)}(\nu,\mu|_{v_n})\mu|_{v_n}(du) \\ &= f(\nu,\emptyset) + \lim_{n\to\infty} \int_{\mathbb{R}^2} f_u^{(1)}(\nu,\mu)\mu|_{v_n}(du) \\ &= f(\nu,\emptyset) + \int_{\mathbb{R}^2} f_u^{(1)}(\nu,\mu)\mu(du), \end{aligned}$$

where the last equality follows by dominated convergence.

Remark. As we can see, the measure ν is unchanged in the r.h.s. of the above equality. This shall allow one to compute the value of the marking function on the entire measure Φ.

Let Φ be a point process such that $\int_{\mathbb{R}^2} \mathbb{E}^{!u}[|f_u^{(1)}(\Phi+\delta_u,\Phi)|]M_\Phi(du) < \infty$, where M_Φ denotes the mean measure of Φ. Then, using the Campbell-Little-Mecke theorem, we obtain

$$\begin{aligned} \mathbb{E}[f(\Phi,\Phi)] &= \mathbb{E}[f(\Phi,\emptyset)] + \int_{\mathbb{R}^2} \mathbb{E}^u[f_u^{(1)}(\Phi,\Phi)]M_\Phi(du) \\ &= \mathbb{E}[f(\Phi,\emptyset)] + \int_{\mathbb{R}^2} \mathbb{E}^{!u}[f_u^{(1)}(\Phi+\delta_u,\Phi)]M_\Phi(du), \end{aligned} \tag{3.14}$$

This leads us to the following result:

Proposition 3.4.1. *Let Φ be a simple point process, and let f be a measurable function, continuous at ∞, such that, for any $j \in \{1, \cdots, n+1\}$,*

$$\int_{(\mathbb{R}^2)^j} \mathbb{E}^{!u}[|f_u^{(j)}(\Phi + \delta_u, \Phi)|] M_{\Phi^{(j)}}(du) < \infty. \tag{3.15}$$

Then,

$$\begin{aligned} \mathbb{E}[f(\Phi, \Phi)] =& \mathbb{E}[f(\Phi, \emptyset)] + \sum_{j=1}^{n} \int_{(\mathbb{R}^2)^j} \mathbb{E}^{!u}[f_u^{(j)}(\Phi + \delta_u, \emptyset)] M_{\Phi^{(j)}}(du) \\ &+ \int_{(\mathbb{R}^2)^{n+1}} \mathbb{E}^{!u}[f_u^{(n+1)}(\Phi + \delta_u, \Phi)] M_{\Phi^{(n+1)}}(du), \end{aligned} \tag{3.16}$$

with, for simplicity, δ_u denoting $\sum_i \delta_{u_i}$. If furthermore,

$$\int_{(\mathbb{R}^2)^n} \mathbb{E}^{!u}[f_u^{(n)}(\Phi + \delta_u, \Phi)] M_{\Phi^{(n)}}(du) \xrightarrow[n \to \infty]{} 0, \tag{3.17}$$

then

$$\mathbb{E}[f(\Phi, \Phi)] = \mathbb{E}[f(\Phi, \emptyset)] + \sum_{j=1}^{\infty} \int_{(\mathbb{R}^2)^j} \mathbb{E}^{!u}[f_u^{(j)}(\Phi + \delta_u, \emptyset)] M_{\Phi^{(j)}}(du). \tag{3.18}$$

Proof. Equation (3.14) tells us that the result holds for $n = 0$. Suppose that it is true for some $n \in \mathbb{N}$, and eq. (3.15) holds for $n + 1$.

Let $u \in (\mathbb{R}^2)^n$, $v \in \mathbb{R}^2$, and let Φ_u a point process with distribution $P_u^!$.

$$\begin{aligned} &\int_{u \in (\mathbb{R}^2)^n} \int_{v \in \mathbb{R}^2} \mathbb{E}^{!v}[|(f_u^{(n)})_v^{(1)}(\Phi_u + \delta_u + \delta_v, \Phi_u)|] M_{\Phi_u}(dv) M_{\Phi^{(n)}}(du) \\ &= \int_{(u,v) \in (\mathbb{R}^2)^{n+1}} \mathbb{E}^{!u,v}[|f_{(u,v)}^{(n+1)}(\Phi + \delta_u + \delta_v, \Phi)|] M_{\Phi^{(n+1)}}(d(u,v)) < \infty \end{aligned}$$

Thus, $\mathbb{E}^{!v}[|(f_u^{(n)})_v^{(1)}(\Phi_u + \delta_u + \delta_v, \Phi_u)|] M_{\Phi_u}(dv) < \infty$, $M_{\Phi^{(n)}}$-a.s. We can therefore apply eq. (3.14) to $f_u^{(n)}(\Phi_u + \delta_u, \Phi_u)$, and obtain the result for $n + 1$. ☐

We can now apply this result to first order scattering moments of marked point processes.

Corollary 3.4.2. *Les Φ be a simple stationary point process, ψ_j a wavelet, $m(u, \Phi)$ a marking function, and*

$$f(\nu, \mu) = |\int_{\mathbb{R}^2} \psi_j(u) m(u, \nu) \mu(du)|,$$

such that the conditions of Proposition 3.4.1 are satisfied, including the condition given by eq. (3.17)*. Then,*

$$\begin{aligned} S_1 \widetilde{\Phi}(j) = & \lambda \mathbb{E}^0[|m(0, \Phi)|] \\ &+ \int_{\|v\| < \|u\|} \mathbb{E}^{u,v}[|\psi_j(u) m(u, \Phi) + \psi_j(v) m(v, \Phi)|] \rho^{(2)}(u, v) du dv \\ &- \int_{\|v\| < \|u\|} \mathbb{E}^{u,v}[|\psi_j(u) m(u, \Phi)| + |\psi_j(v) m(v, \Phi)|] \rho^{(2)}(u, v) du dv \\ &+ R, \end{aligned}$$

with

$$R = \sum_{n=3}^{\infty} \frac{1}{n!} \sum_{J \in \{2^n\}} (-1)^{n-|J|} \int_{(\mathbb{R}^2)^n} \mathbb{E}^{u_J}[|\sum_{i \in J} \psi_j(u_i) m(u_i, \Phi)|] \rho^{(n)}(u) du,$$

noting $u = (u_i)_{i=1,\cdots,n}$*, and* $u_J = (u_i)_{i \in J}$*.*

Proof. The proof follows directly from Proposition 3.4.1, and the following Lemma ([14]).

Lemma 3.4.3. *Let* f *be a measurable function, continuous at* ∞*. Let* $n \geq 1$*, and* $u = (u_1, \cdots, u_n) \in E^n$*. Then, noting* $\{2^n\}$ *the set of all subsets of* $\{1, \cdots, n\}$*, for any* $\nu, \mu \in \mathbb{M}$*, if* $\|u_n\| < \cdots < \|u_1\|$*,*

$$f_u^{(n)}(\nu, \mu) = \sum_{J \in \{2^n\}} (-1)^{n-|J|} f(\nu, \mu|_{u_n} + \sum_{i \in J} \delta_{u_i}), \tag{3.19}$$

otherwise $f_u^{(n)}(\nu, \mu) = 0$*.*

□

Example 3.4.4 (Voronoi marks of a determinantal point process)**.** *We check the conditions of Proposition 3.4.1 for a determinantal point process (cf. Section 2.3.2) with the marking function* $m(u, \Phi) = V(u)$ *(cf. eq.* (2.20)*), with Morlet wavelets* ψ_j *(cf. Example 3.1.2).*

First, note that eq. (3.19) *also reads, for* u *such that* $\|u_n\| < \cdots < \|u_1\|$*,*

$$\begin{aligned} f_u^{(n)}(\nu, \mu) = \sum_{J \in \{2^n\} \backslash \{1\}} & (-1)^{n-1-|J|} f(\nu, \mu|_{u_n} + \delta_{u_1} + \sum_{i \in J} \delta_{u_i}) \\ & - (-1)^{n-1-|J|} f(\nu, \mu|_{u_n} + \sum_{i \in J} \delta_{u_i}), \end{aligned} \tag{3.20}$$

where $\{2^n\} \backslash \{1\}$ *denotes all subsets of* $\{2, \cdots, n\}$*. Therefore,*

$$|f_u^{(n)}(\nu, \mu)| \leq \sum_{J \in \{2^n\} \backslash \{1\}} |f(\nu, \mu|_{u_n} + \delta_{u_1} + \sum_{i \in J} \delta_{u_i}) - f(\nu, \mu|_{u_n} + \sum_{i \in J} \delta_{u_i})|,$$

which gives us,

$$\begin{aligned} |f_u^{(n)}(\nu, \mu)| \leq & \sum_{J \in \{2^n\} \backslash \{1\}} || \int_{\mathbb{R}^2} \psi_j(u) |V(u)| (\mu_J + \delta_{u_1})(du)| \\ & \qquad - | \int_{\mathbb{R}^2} \psi_j(u) |V(u)| (\mu_J)(du)|| \\ \leq & \sum_{J \in \{2^n\} \backslash \{1\}} | \int_{\mathbb{R}^2} \psi_j(u) |V(u)| (\mu_J + \delta_{u_1} - \mu_J)(du)| \\ \leq & \sum_{J \in \{2^n\} \backslash \{1\}} |\psi_j(u_1)||V(u_1)| \\ \leq & \, 2^{n-1} |\psi_j(u_1)||V(u_1)|, \end{aligned} \tag{3.21}$$

where we noted $\mu_J = \mu|_{u_n} + \sum_{i \in J} \delta_{u_i}$*. Applying eq.* (3.21)*, we obtain*

$$\int_{(\mathbb{R}^2)^p} \mathbb{E}^{!u}[|f_u^{(p)}(\Phi + \delta_u, \Phi)|] M_{\Phi^{(p)}}(du) \leq \frac{2^{p-1}}{(p-1)!} \int_{(\mathbb{R}^2)_1^p} |\psi_j(u_1)| \mathbb{E}^u[|V(u_1)|] M_{\Phi^{(p)}}(du), \tag{3.22}$$

where we note $(\mathbb{R}^2)_1^p = \{(u_1, \cdots, u_p) : \forall\, i \neq 1,\ \|u_i\| < \|u_1\|\}$*.*

Now, in order to bound the expected value of the Voronoi cell surface $\mathbb{E}^u[|V(u_1)|]$*, we need to explicitly write its dependency on* Φ*:*

$$\begin{aligned}
\mathbb{E}^u[|V(u_1)|] &= \mathbb{E}^{!u}[|V_{\Phi+\delta_u}(u_1)|] \qquad (3.23)\\
&\leq \mathbb{E}^{!u}[|V_{\Phi+\delta_{u_1}}(u_1)|] \qquad (3.24)\\
&= \int_{\mathbb{R}^+} \mathbb{P}^!_u(|V_{\Phi+\delta_{u_1}}(u_1)| > t)dt\\
&= \int_{\mathbb{R}^+} \mathbb{P}^!_u(|V_{\Phi+\delta_{u_1}}(u_1)| > \pi r^2) 2\pi r dr\\
&\leq \int_{\mathbb{R}^+} \mathbb{P}^!_u(\exists i \in \{1, \cdots, 6\} : \Phi(\angle_i(r)) = 0) 2\pi r dr \qquad (*)\\
&\leq \sum_{i=1}^{6} \int_{\mathbb{R}^+} \mathbb{P}^!_u(\Phi(\angle_i(r)) = 0)) 2\pi r dr, \qquad (3.25)
\end{aligned}$$

where $\{\angle_i(r)\}_{i=1,\cdots,6}$ *forms a partition of* $B_{u_1}(r)$*, with every* $\{\angle_i(r)\}$ *having the same volume. The explanation for eq. (*) is geometric: if every subset* $\{\angle_i(r)\}$ *of* $B_{u_1}(r)$ *contains an atom, then the Voronoi cell of* u_1 *is contained in* $B_{u_1}(r)$[2]*.*

Let Φ *be a Poisson point process on* $(\mathbb{R}^2, \mathbb{B})$*, with intensity* $0 < \lambda < \infty$*. We know that, for all* $B \in \mathcal{B}$*,*

$$\mathbb{P}(\Phi(B) = 0) = e^{-\mathbb{E}[\Phi(B)]}.$$

Furthermore, we know that, for any $N \geq 1$*, and any* $u \in (\mathbb{R}^2)^N$*, the reduced Palm version of* Φ *at* u*, noted* $\Phi^!_u$*, is equal in distribution to* Φ*. Thus, for all* $B \in \mathcal{B}$*,*

$$\mathbb{P}(\Phi^!_u(B) = 0) = e^{-\mathbb{E}[\Phi(B)]}.$$

Now, let Φ^K *be a stationary determinantal point process, with Hermitian kernel* K*, and intensity* λ *(that is, for all* $B \in \mathcal{B}$*,* $\mathbb{E}[\Phi^K(B)] = \mathbb{E}[\Phi(B)]$*). Then, for all* $B \in \mathcal{B}$*, we know that ([16]),*

$$\mathbb{P}(\Phi^K(B) = 0) \leq \mathbb{P}(\Phi(B) = 0) = e^{-\mathbb{E}[\Phi(B)]} = e^{-\mathbb{E}[\Phi^K(B)]}.$$

Let $u = (u_1, \cdots, u_p) \in \mathbb{R}^p$*. One can show that the reduced Palm version of* Φ^K *at* u*, noted* $\Phi^{K!}_u$*, is also a determinantal point process, which kernel shall be noted* K_u*. Denote also* $\Phi^{K_u!}_P$ *the Poisson point process with the same mean measure as* $\Phi^{K!}_u$*. Then, for all* $B \in \mathcal{B}$ *([16]),*

$$\begin{aligned}
\mathbb{P}^!_u(\Phi^K(B) = 0) &= \mathbb{P}(\Phi^{K!}_u(B) = 0)\\
&\leq \mathbb{P}(\Phi^{K_u!}_P(B) = 0)\\
&= e^{-\mathbb{E}[\Phi^{K_u!}_P(B)]}\\
&= e^{-\mathbb{E}[\Phi^{K!}_u(B)]}\\
&\leq e^{-(\mathbb{E}[\Phi^K(B)] - p)}\\
&= e^{-(\mathbb{E}[\Phi(B)] - p)},
\end{aligned}$$

---

[2] Indeed, the partition $\{\angle_i(r)\}_{i=1,\cdots,6}$ can be defined by 6 points $v1, \cdots, v_6$, such that for all i, the angle $\widehat{v_i u_1 v_{i+1}}$ equals $\frac{\pi}{3}$. Let v be the atom of Φ lying inside the subset of $B_{u_1}(r)$ inside $\{\angle_i(r)\}$. Since $\widehat{v_i u_1 v_{i+1}} = \frac{\pi}{3}$, and $\|v - u_1\| \leq r$ the bisector of $[u_1, v]$ intersects $[u_1, v_i]$ and $[u_1, v_{i+1}]$.

where the last inequality comes from the fact that $\mathbb{E}[\Phi_u^{K!}(B)] \geq \mathbb{E}[\Phi^K(B)] - p$ *(see [22, eq.(1.6) in the Supplemental content], for a proof).*

We can then write, for any $r \geq 0$, *and any* $i \in \{1, \cdots, 6\}$,

$$\begin{aligned} \mathbb{P}_u^!(\Phi^K(\angle_i(r)) = 0) &\leq e^{-(\mathbb{E}[\Phi(\angle_i(r))] - p)} \\ &\leq e^{p - \lambda \frac{\pi r^2}{6}}. \end{aligned}$$

Injecting this inequality in eq. (3.25), *it gives us*

$$\begin{aligned} \mathbb{E}^{!u}[V_{\Phi+\delta_u}(u_1)] &\leq 6 \int_{\mathbb{R}^+} e^{p - \lambda \frac{\pi r^2}{6}} 2\pi r dr \\ &= c_p, \end{aligned}$$

where $c_p = 6e^p \int_{\mathbb{R}^+} e^{-\lambda \frac{\pi r^2}{6}} 2\pi r dr = c e^p < \infty$.

Going back to eq. (3.22), *we obtain*

$$\begin{aligned} \int_{E^p} \mathbb{E}^{!u}[|f_u^{(p)}(\Phi + \delta_u, \Phi)|] M_{\Phi^{(p)}}(du) &\leq c \frac{2^{p-1}}{(p-1)!} e^p \int_{(\mathbb{R}^2)_1^p} |\psi_j(u_1)| M_{\Phi^{(p)}}(du) \\ &\leq c \frac{2^{p-1}}{(p-1)!} e^p \lambda^p \int_{(\mathbb{R}^2)_1^p} |\psi_j(u_1)| du \\ &= c \frac{2^{p-1}}{(p-1)!} e^p \lambda^p \pi^{p-1} \int_{\mathbb{R}^2} |\psi_j(u)| |u|^{2(p-1)} du, \quad (3.26) \end{aligned}$$

where the second inequality comes from the fact that, for a determinantal point process, for any $u \in E^p$, $\rho^{(p)}(u) \leq \lambda^p$.

Now, as ψ_j *is a Morlet wavelet (cf. eq.* (3.3)*), we can give an upper bound for* $|\psi_j(u)|$:

$$|\psi_j(u)| \leq \frac{1}{2\pi\sigma^2} 2^{-2j} e^{-2^{-2j}|u|^2/(2\sigma^2)} |e^{i2^{-j}\xi_0 . u} - c_\psi| \leq 2^{-2j} c'_\psi e^{-2^{-2j}|u|^2/(2\sigma^2)},$$

where c'_ψ *is a positive constant. Therefore,*

$$\begin{aligned} \int_{\mathbb{R}^2} |\psi_j(u)| |u|^{2(p-1)} du &\leq 2^{-2j} c'_\psi \int_{\mathbb{R}^2} e^{-2^{-2j}|u|^2/(2\sigma^2)} |u|^{2(p-1)} du \\ &\leq 2^{-2j} c'_\psi \int_{[0,2\pi]} d\theta \int_{\mathbb{R}^+} e^{-2^{-2j} r^2/(2\sigma^2)} r^{2p-1} dr \\ &\leq 2^{-2j} c'_\psi 2\pi (2^{2j} 2\sigma^2)^p \int_{\mathbb{R}^+} e^{-r^2} r^{2p-1} dr \\ &\leq 2^{-2j} c'_\psi 2\pi (2^{2j} 2\sigma^2)^p \frac{1}{2} \int_{\mathbb{R}^+} e^{-r} r^{p-\frac{1}{2}} r^{-\frac{1}{2}} dr \\ &\leq 2^{-2j} c'_\psi \pi (2^{2j} 2\sigma^2)^p \int_{\mathbb{R}^+} e^{-r} r^{p-1} dr \\ &\leq \pi c'_\psi 2^{2j(p-1)} \sigma^{2p} 2^p (p-1)! \end{aligned}$$

Injecting this into eq. (3.26), *we obtain*

$$\int_{E^p} \mathbb{E}^{!u}[|f_u^{(p)}(\Phi + \delta_u, \Phi)|] M_{\Phi^{(p)}}(du) \leq c_1 (4\pi\sigma^2 e \lambda 2^{2j})^p, \quad (3.27)$$

where c_1 is a positive constant, independent of p. This shows that, for all p,

$$\int_{(\mathbb{R}^2)^p} \mathbb{E}^{!u}[|f_u^{(p)}(\Phi + \delta_u, \Phi)|] M_{\Phi^{(p)}}(du) < \infty,$$

so the condition Equation (3.15) *of proposition 3.4.1 is satisfied. If furthermore* $4\pi\sigma^2 e\lambda 2^{2j} < 1$*, then*

$$\int_{(\mathbb{R}^2)^p} \mathbb{E}^{!u}[|f_u^{(p)}(\Phi + \delta_u, \Phi)|] M_{\Phi^{(p)}}(du) \xrightarrow[p\to\infty]{} 0,$$

and eq. (3.18) *holds. In conclusion, factorial moment expansions of any order exist for all scattering moments $S_1\tilde{\Phi}(j)$ of any scale j, for a determinantal process marked by the volumes of the Voronoi cells. Also, the infinite expansion exists for all sufficiently small scales with respect to the intensity of the determinantal process. Remark that the fact that Φ is a determinantal point process introduces the factor e^p in the bound eq.* (3.27) *w.r.t. a Poisson point process.*

Remark. Note that the wavelet transform (eq. (3.9)) is a linear function (shot-noise functional) of the (marked) point process. In other words, it is a first-order U-statistics of Φ^m; cf [71, Section 12.3]. The fact that scattering moments $S_1\Phi(j,\theta)$ are defined via the modulus of the wavelet transform makes them dependent on all higher-order correlation functions of Φ: the factorial moment expansions of $S_1\Phi(j,\theta)$ involve all moment measures (in contrast to the square norm of the wavelet transforms, which can be represented using the first and second-order U-statistics and thus their expansion involves only the first two correlation functions). Consequently, the first order scattering moments are supposed to capture more information regarding the intrinsic dependence of the points than just the pairwise correlations. We shall see in our numerical study that this information allows one to efficiently recover some geometric marks which do depend on higher order correlations, e.g, the Voronoi cell characteristics.

3.5 Wavelet phase harmonics covariance

We shall now take interest in another class of representations based on the wavelet transform of point processes. Scattering moments decompose the information of a point process into different features at different scales, and those features in different sub-features at different scales. However, this representation does not take into account the dependencies between the different features at different scales. Furthermore, the modulus operator introduced in the definition of scattering moments remove the phase information of the wavelet coefficients, which can contain important information about the geometry of the observed pattern, as illustrated in [92, 120].

For this reason, in [120], the authors introduce covariance moments between wavelet coefficients, which phases are adjusted by a non-linear operator. In this section, we shall review this representation, for stationary random measures defined on $(\mathbb{R}^2, \mathcal{B})$.

As stated in Section 3.1, for a wavelet ψ with central frequency ξ_0, one builds a wavelet family by dilating and rotating ψ. The wavelet at scale j and angle θ has central frequency $\xi := 2^{-j} r_{-\theta}\,\xi_0$. In this section, we shall, for notations simplicity, identify the central frequency ξ with the indices of the wavelets, (j,θ). We are interested, in this section, in capturing the dependencies between wavelet coefficients of a stationary point process Φ, at possibly different scales, orientations, and spatial positions. To study such dependencies, one can naturally look at the covariance moments between wavelet coefficients, that is, moments of the form

$$C\Phi(\xi,\xi',\tau) = \mathbb{E}[\Phi \star \psi_\xi(u)\, \Phi \star \psi_{\xi'}(u-\tau)^*], \tag{3.28}$$

for $(\xi,\xi') \in (\mathbb{Z}\times[0,2\pi])^2$, $\tau \in \mathbb{R}^2$. Note that in this case, we do not need to subtract the expected averages, since the wavelets ψ_ξ have zero means. Furthermore, as Φ is stationary, the value in the right hand side of eq. (3.28) does not depend on u, that can be set to 0.

Suppose that Φ satisfies the conditions of Section 2.8, that is, it is stationary, has a non-null finite intensity, and with a σ-finite second-order factorial moment measure $M_{\Phi^{(2)}}$ absolutely continuous w.r.t. the Lebesgue measure. Then, using eq. (2.23), we obtain

$$C\Phi(\xi,\xi',\tau) = \int_{\mathbb{R}^2} \hat{\psi}_\xi(\omega)\hat{\psi}^*_{\xi'}(\omega)e^{-i\omega.\tau}b_\Phi(\omega)d\omega. \tag{3.29}$$

Therefore, computing $C\Phi(\xi,\xi',\tau)$, for $\tau\in\mathbb{R}^2$, is equivalent to computing the Fourier transform of $\hat{\psi}_\xi\hat{\psi}^*_{\xi'}b_\Phi$.

Recall also that the wavelets considered in this work are localized in frequency, i.e. there exists a constant C' such that $\hat{\psi}(\omega) \simeq 0$ if $|\omega-\xi_0| > C'$. This means that, the (approximate) frequency support of ψ_ξ is contained in a ball centered around $2^{-j}r_{-\theta}\xi_0$ and of radius $2^{-j}C'$. If C' is small enough, then for $\xi \neq \xi'$, ψ_ξ and $\psi_{\xi'}$ approximately have disjoint frequency supports, meaning that

$$\hat{\psi}_\xi(\omega)\hat{\psi}^*_{\xi'}(\omega) \simeq 0, \quad \forall\, \omega\in\mathbb{R}^2.$$

This implies that the covariance between wavelet coefficients will be close to zero, even though they may not be independent.

To remedy this problem, one can apply a non-linear operator to the wavelet coefficients, in order to modify their frequency support, so that they may overlap. The phase harmonics operator, that we shall now define, is particularly suited for that purpose.

Definition 3.5.1 (Phase harmonics operator)**.** Let $z\in\mathbb{C}$, and denote $|z|$ its modulus, and $\varphi(z)$ its phase, i.e. $z = |z|e^{i\varphi(z)}$. The phase harmonics operator is defined, for every $k\in\mathbb{Z}$, by

$$[z]^k := |z|e^{ik\varphi(z)}. \tag{3.30}$$

Note that, for any $z\in\mathbb{C}$, $[z]^0 = |z|$, $[z]^1 = z$, and $[z]^{-1} = z^*$. More generally, for any $k\in\mathbb{Z}$, $([z]^k)^* = [z]^{-k}$ and $|[z]^k| = |z|$ for $k\in\mathbb{Z}$.

For $k\in\mathbb{Z}$ and $z\in\mathbb{C}$, the phase harmonics operator has the same effect on the phase of z than taking z to the power k. However, using estimators of high order moments can prove problematic, due to their possibly large variance. This operator, however, does not change the modulus of z.

We can apply the phase harmonics operator to the wavelet coefficients of $\mu\in\mathbb{M}$ to adjust their phase.

Definition 3.5.2 (Wavelet phase harmonics)**.** Let μ be a measure on $(\mathbb{R}^2,\mathcal{B})$, and $\psi\in L^1(\mathbb{R}^2,\mu)$. We define the wavelet phase harmonics (WPH) of μ by

$$[\mu\star\psi_\xi]^k = |\mu\star\psi_\xi|e^{ik\varphi(\mu\star\psi_\xi)}, \tag{3.31}$$

for $\xi\in\Lambda = \mathbb{Z}\times[0,2\pi]$, and $k\in\mathbb{Z}$.

The phase harmonics operator was introduced in [81], and used in [120] to build a class of representations, and define maximum entropy models for non-Gaussian stationary processes. The idea of multiplying the phase of complex wavelet coefficients to compute their dependencies was already present in [92], although it was not conceptualized in an operator, that can be computed for any $k \in \mathbb{Z}$.

Frequency transposition The phase harmonics operator modifies the phase of complex wavelet coefficients, without modifying its modulus. In [81], the authors explain that such an operator, called a frequency transposition, performs a dilation in the Fourier domain. They show that, in the frequency domain, if $\hat{\psi}_{j_1,\theta_2}$ is approximately supported on the ball $B_\xi(C')$ of radius C', centered at $\xi \in \mathbb{C}$, then $[\hat{\psi}_{j_1,\theta_2}]^k$ is approximately supported on $B_{k\xi}(kC')$. This property, illustrated in [120] in the case of realizations of stationary processes, allows us to capture dependencies of wavelet coefficients across scales and angles, as we shall detail next.

Definition 3.5.3 (WPH covariance moments)**.** Let Φ be a stationary point process on $(\mathbb{R}^2, \mathcal{B})$, and $\Psi = \{\psi_0, \psi_\xi\}_{\xi \in \Xi = \mathbb{Z} \times [0,2\pi]}$ be a wavelet family such that all wavelet coefficients have finite second-order moments. For any $\xi, \xi' \in \Xi$, $k, k' \in \mathbb{Z}$, $\tau \in \mathbb{R}^2$, let us note $\gamma = (\xi, \xi', k, k', \tau)$. The WPH covariance of Φ is defined by

$$K_\gamma \Phi := \operatorname{Cov}\Big([\Phi \star \psi_\xi]^k(u), [\Phi \star \psi_{\xi'}]^{k'}(u - \tau) \Big). \tag{3.32}$$

Note that, as Φ is stationary, the r.h.s. of eq. (3.32) does not depend on $u \in \mathbb{R}^2$.

Note that the WPH covariance representation does not need to include the covariance between coefficients for all indices $\xi, \xi' \in \Xi$, $k, k' \in \mathbb{Z}$, $\tau \in \mathbb{R}^2$, because some of these covariance moments are nearly zero (as explained by the frequency transposition property). A way to choose the coefficients in the representation shall be detailed in Section 5.4, in the case where this representation is used to define a maximum entropy model.

3.6 Estimation and coefficient selection

Estimation The wavelet-based representations presented in Sections 3.2 and 3.5 capture information about the distribution $\mathcal{L}_\Phi$ of Φ by computing mathematical expectancies of non-linear functions applied to the the wavelet transform of Φ (cf. Section 3.1). In other words, we build representations of the form

$$R(\Phi) = \mathbb{E}[f(\Phi)],$$

for non-linear functions f.

If Φ is a stationary, ergodic point process, these expectations can be estimated from a realization $\bar{\phi}$ of Φ, observed on a finite window W, by spatial averaging over the window. That is, we build our estimators as

$$\hat{R}_W(\Phi) := \frac{1}{|W|} \int_W f(\Phi|_W)(u) du,$$

where $\Phi|_W = \Phi \cap W$. Note that, if we were given Φ over the whole space $\mathbb{R}^2$ instead of the bounded window W, we could define

$$\hat{R}_{W,\mathbb{R}^2}(\Phi) := \frac{1}{|W|} \int_W f(\Phi)(u) du,$$

and this estimator would be unbiased. Furthermore, Birkhoff's ergodic theorem (cf. eq. (2.21)) would tell us that

$$\lim_{|W|\to\infty} \hat{R}_{W,\mathbb{R}^2}(\Phi) = R(\Phi), \quad \mathbb{P}\text{-a.s.}$$

As we only observe $\Phi|_W$, that is the points of Φ that lie inside W, the wavelet coefficients $\Phi|_W \star \psi_{j,\theta}(u)$ at points $u \in W$ near the border of W cannot take into account points outside W. One can therefore consider that the observation sample has no point outside W, or consider other types of border corrections. For instance, in this work, we shall usually perform a periodic boundary correction, i.e; consider that the sample repeats itself outside W, when W is a square window. This can be done by using periodic wavelets $\{\psi_\xi^s\}$, defined as

$$\psi_\xi^s(u) := \sum_{n\in\mathbb{Z}^2} \psi_\xi(u + sn), \tag{3.33}$$

where $s > 0$ is the size of the square window W. We shall however often make the approximation that $|W|$ is large enough, so that border effects are negligible, and the ergodic theorem holds.

Finite representation In Section 3.1, the wavelet transform has been defined with an infinite wavelet family $\{\psi_{j,\theta}\}$, with $j \in \mathbb{Z}$ and $\theta \in [0, 2\pi]$. For practical purposes, we need to select a finite number of such coefficients.

For the angular parameter θ, this can be done by simply selecting discrete, evenly spaced angles in $[0, 2\pi]$. Thus θ shall be parametrized by $\theta = 2\pi\frac{l}{L}$, with $l \in \{0, \cdots, L-1\}$, for some $L \geq 1$. In our simulations, we shall usually take $L = 4$ or $L = 8$.

The scale parameter j will be taken in $\{0, \cdots, J-1\}$, for some $J \geq 1$. The highest frequency captured by the wavelet transform is therefore determined by the central frequency ξ_0 of the mother wavelet. Equivalently, the wavelet family will not capture structures smaller than C, the size of the approximate spatial support of ψ. The parameter ξ_0 must therefore be chosen according to the task at hand. In practical applications, one deals with samples observed inside a bounded window W. The maximal scale parameter J is typically chosen so that the approximate support of the wavelets $\psi_{J-1,\theta}$ (those with maximal support) is contained in the observation window, i.e. $2^J C \leq s$, for a square window W of size s. Scales equal or larger than J are carried by the low-pass filter ψ_0, whose frequency support is centered at $\xi = 0$.

Properties The estimators of scattering moments and wavelet phase harmonics covariance moments enjoy properties that make them suitable for statistical learning of non-linear functions of a (possibly marked) point pattern, a point process, or its distribution. First, being defined with a global spatial average, these representations are invariant to translations, which can be a desirable property (e.g. if the function to approximate is itself also translation invariant). Another property that is sometimes required of such descriptors is some notion of stability. In other words, if a point pattern undergoes some kind of small deformation, then the associated value of the function to approximate might not change too much. Thus, the descriptor of this point pattern should not be too affected by this deformation either. In the case of estimators of first order scattering moments, such stability property can be easily proven.

Proposition 3.6.1. *Let* $\phi \in \mathbb{M}$ *be a finite counting measure, i.e.* $\phi = \sum_{i=1}^N \delta_{u_i}$*. Let* $m \in \mathbb{R}^N$*,* $m = (m_1, \cdots, m_N)$*. We define* $\widetilde{\phi} = \sum_{i=1}^N m_i \delta_{u_i}$*. Let us also define the marks and positions perturbations:* $u_\epsilon \in (\mathbb{R}^2)^N$ *and* $m_\epsilon \in \mathbb{R}^N$*. We note* $\phi_\epsilon = \sum_{i=1}^N \delta_{u_i + u_{\epsilon,i}}$*, and* $\widetilde{\phi}_\epsilon = \sum_{i=1}^N (m_i + m_{\epsilon,i}) \delta_{u_i + u_{\epsilon,i}}$*. Finally, note*

$$\hat{S}_1 \widetilde{\phi} = \int_{\mathbb{R}^2} |\widetilde{\phi} \star \psi|(u) du,$$

for some wavelet ψ*, and similarly for* $\widetilde{\phi}_\epsilon$*. Suppose that* ψ *is integrable, and such that* $\forall\, u, t \in \mathbb{R}^2$*,* $|\psi(u+t) - \psi(u)| \leq \|t\| \varsigma(u)$*, for some integrable function* ς*. Then,*

$$|\hat{S}_1 \widetilde{\phi} - \hat{S}_1 \widetilde{\phi}_\epsilon| \leq \|\varsigma\|_1 \|m\|_2 \||u|_\epsilon\|_2 + \|\psi\|_1 \|m_\epsilon\|_1$$

where $|u|_\epsilon = (\|u_{\epsilon,i}\|)_{i=1,\cdots,N}$*.*

Proof. Let us denote $\tilde{\xi} = \sum_{i=1}^N m_i \delta_{u_i + u_{\epsilon,i}}$. Using the fact that, for all $z, z' \in \mathbb{C}$, $||z| - |z'|| \leq |z - z'|$, we write

$$\begin{aligned}
|\hat{S}_1 \widetilde{\phi} - \hat{S}_1 \widetilde{\phi}_\epsilon| &= |\hat{S}_1 \widetilde{\phi} - \hat{S}_1 \tilde{\xi} + \hat{S}_1 \tilde{\xi} - \hat{S}_1 \widetilde{\phi}_\epsilon| \\
&\leq |\hat{S}_1 \widetilde{\phi} - \hat{S}_1 \tilde{\xi}| + |\hat{S}_1 \tilde{\xi} - \hat{S}_1 \widetilde{\phi}_\epsilon| \\
&\leq \sum_i |m_i| \int_{\mathbb{R}^2} |\psi(u - u_i) - \psi(u - u_i - u_{\epsilon,i})| du + \sum_i |m_{\epsilon,i}| \int_{\mathbb{R}^2} |\psi(u - u_i - u_{\epsilon,i})| du \\
&\leq \|\varsigma\|_1 \sum_i |m_i| \|u_{\epsilon,i}\|_E + \|\psi\|_1 \sum_i |m_{\epsilon,i}| \\
&\leq \|\varsigma\|_1 \|m\|_2 \||u|_\epsilon\|_2 + \|\psi\|_1 \|m_\epsilon\|_1.
\end{aligned}$$

□

In this chapter, we have presented two classes of representations of stationary point processes, based on a multi-scale decomposition of the geometric information with a wavelet transform.

In the remainder of this work, we shall study the adequacy of these representations to perform statistical learning and probabilistic modelling tasks.

Chapter 4

Scattering moments for point process marks regression

In this chapter, we propose to address the problem of learning marks of point processes: an unknown function, called a marking function, takes as argument a point pattern, and produces numerical values, called marks, associated with each point of the pattern. Our goal is to learn this function, from examples of point patterns with observed marks, in order to predict unknown marks for new point patterns.

Our particular motivation comes from the problem of learning how the cell loads in wireless networks depend on the geometry of the base stations (and possibly of the traffic demand). One may want to learn this dependence directly from real data, collected in the existing operational networks, to predict the loads of base stations for different configurations of base stations, and/or different traffic demand (see more details in the Related works paragraph). In this work, we consider some simple, generic marks, often appearing in this context. They are produced by the standard shot-noise interference model, the nearest neighbour distance, and some characteristics of the Voronoi cells (cf. Section 2.6 Figure 4.1). Our goal is to to understand the efficiency of our proposed method to represent the dependence of the marks on the geometry of point patterns, rather than to solve the original cell load problem for some particular real network data.

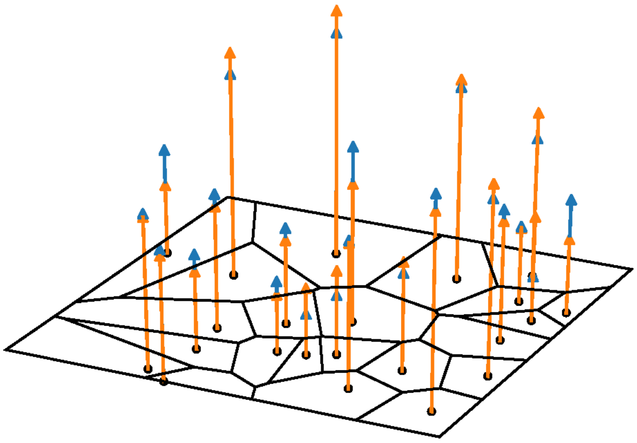


Figure 4.1: Exact (blue) and reconstructed (orange) marks being the surface areas of the Voronoi cells.

Learning geometric marks via local or global representation The problem of learning the marking function, as posed above, does not seem amenable to any direct statistical approach, due to the structure of the space of marked point measures. To overcome this difficulty, we need some suitable representation of the geometry of point patterns, and we propose two approaches in this regard.

As a baseline, we propose to estimate the mark of each point using a statistical regression model based on the *local distance matrix* of a suitable chosen vicinity of this point. In this relatively simple approach, the training data set consists of an ensemble of central marked points, surrounded by some number, say K, of their (non-marked) neighbours. Observe that the dimension of the representation of the local geometry of the point (local distance matrix) increases as K^2. For highly non-local marking functions (e.g. related to power-law shot-noise function), one might need to chose a large number of neighbours K, proportional to the total number of points, possibly making this baseline approach inefficient.

When the marks have non-local dependence, or when we have no prior knowledge of the dependence range, we further propose to use the *scattering moments* representation (cf. Section 3.2) to capture the geometry of marked point patterns. Recall that the scattering moments are defined as the spatial average of a cascade of convolutions with a wavelet family, followed by a complex modulus operator. The property of the modulus operator makes this representation locally invariant to small translations ([31]), and the global averaging implies a global translation invariance. Furthermore, they have been shown ([80, 31], Proposition 3.6.1) to be stable to small deformations of the input signal. These invariance and stability properties make them useful in signal processing, in particular in relation to statistical learning. Indeed, if the information content of a signal is typically not (strongly) affected by translations, rotations, and small deformations, similar properties of the signal representation allows one to capture this content in a more concise way, and hopefully learn its intrinsic structure from a smaller number of signal samples. The pertinence of this approach has already been demonstrated in various contexts (see e.g. [31, 100, 63]).

Using scattering moments to represent marked point processes, our learning problem can be addressed in the following two steps (also depicted on Figure 4.2):

- We build a statistical regression model on scattering moments of the marked point patterns, with the explanatory variables being the scattering moments of the non-marked point patterns. This model is computed on the training data consisting of point patterns with observable marks, and is meant to be used to estimate the (marked) scattering moments of new point patterns, for which marks are not observed.

- We estimate (reconstruct) the marks of new point patterns, for which marks are not observed, from their estimated scattering moments. It is a non-convex optimization problem, that we numerically solve using the L-BFGS-B algorithm [34, 121].

Compared to our baseline approach, for which the size of the representation of the local geometry of a point grows quadratically with its number of neighbours, the scattering representation grows logarithmically with the neighbouring size. However, it may suffer from possible errors, introduced in the reconstruction phase (absent in the benchmark approach). The overall benefits of the scattering approach become significant for highly non-local marking functions.

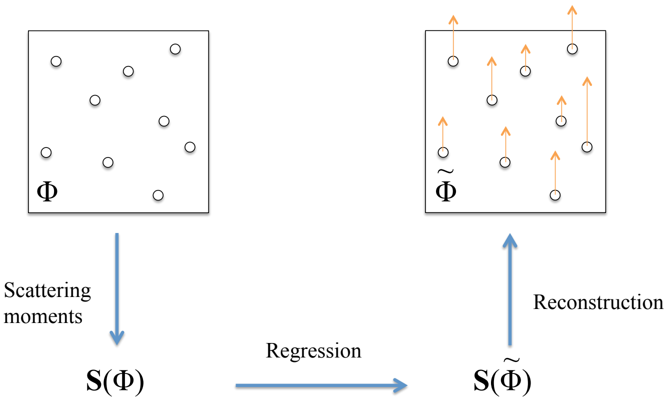


Figure 4.2: An overview of our method for the reconstruction of geometric marks via scattering moments.

In general, the quality of the regression of the scattering moments depends on the sensitivity of the marking function to small point pattern deformations. On the other side, in the reconstruction phase (recovering of marks from the true or estimated scattering moments) significant errors consist in swapping a large and a small mark of two neighbouring points (not leading to a significant modification of the scattering moments). The quality of this phase of the approach depends thus on the existence of clusters of points in point patterns — a problem already studied in the literature; cf Section 4. The Poisson point process, considered in this paper, exhibits a baseline type of clustering.

Related works Stochastic-geometric study of cellular networks expanded rapidly in recent years, primarily through analytic results regarding Poisson network models ([21]). Performance evaluation of operational wireless cellular networks, in particular the quality of service perceived by users in function of the traffic demand, is a complex problem involving stochastic and geometric modelling of several network layers. A key element of this problem is the analysis of the cell loads, which non trivially depend on the geometric configuration of serving base stations and their traffic demands, and capture in a concise way the quality of service offered by the individual cells. A detailed physical model of cell loads was proposed in [103] and revisited in [19, 20, 15], including the validation with respect to some data collected in operational networks.

Recently, the prediction of the cell loads has been the subject of a machine learning study published in [5], though not as the geometric problem posed in our paper. Indeed, in this earlier paper, the number and the locations of base stations are fixed, and the problem consists in learning the loads of this given configuration of stations in function of their traffic demands. This (non-geometric) problem is cast and solved in the framework of monotone interpolation of Lipschitz functions.

The problem formulated and studied in our paper allows one to address the orthogonal question of learning the cell loads, given constant traffic demand, in function of the geometry of the network. Combining the two approaches, left for future work, may allow one to predict the performance of new geometric configurations of base stations and traffic demand, by collecting the cell load data in existing operational networks, thus offering an alternative to building and solving complex physical models.

Observe that, even for a fixed number of base stations, the two dimensional location of each base station on the plane makes the dimension of our input data two times higher than the one considered in [5]. The use of wavelet scattering transforms allows one to leverage geometric invariants, such as translation and rotation, and thus significantly reduce the dimension of the representation. Scattering moments have been proven useful in this context in a number of tasks such as quantum molecular energy prediction [63, 49] or texture classification [100].

Geometric marks (also called score functions) have received a lot of attention in stochastic geometry and spatial statistics, where they represent some interaction of a given point with the whole point pattern. If this interaction is local in some sense, and the underlying point process exhibits some decay of correlations, then it is possible to establish asymptotic results (including central limit theorems) regarding the sums of the geometric marks in increasingly large windows (cf e.g. [90, 11, 22]. These results can be used to study large-scale asymptotics of the scattering moments, as in [33].

We mentioned earlier in this introduction that the quality of reconstruction of the marks depends on whether point process is regular, or exhibits clusters of points. These notions are formalised in [16, 17], and it is argued that Poisson point process can be considered as a reference type of point clustering, to which more regular point processes (e.g. determinantal ones) and more clustering ones (e.g. permanental and Cox processes) can be compared.

The remaining part of this chapter, based on [28], is organized as follows: in Section 4.1, we describe several models of geometric marks and formulate the problem of their statistical learning. We then present our main approach to solving this problem, based on scattering transforms of marked point processes. We also describe the benchmark approach, to which our method shall be compared. The numerical results of both approaches applied to the considered mark models are presented in Section 4.2.

4.1 Statistical learning of geometric marks

In this section, we formally present the problem of learning geometric marks of point processes, and formulate possible solutions to address this problem. Our main approach, described in Section 4.1.4, uses the scattering moments of marked point patterns, which are presented in Section 4.1.3. An alternative method is proposed in Section 4.1.5, based on a distance matrix representation of point neighbourhoods. This alternative approach is used as a comparative benchmark for our numerical study in Section 4.2.

4.1.1 Geometric characteristics of wireless networks

In stochastic-geometric modeling of networks (e.g. wireless networks), one usually represents locations of objects (e.g. transmitters/receivers) as *points* of a point process, and their characteristics as *marks*. These characteristics can depend not only on the given object, but also on the geometry of the network (at least locally, e.g. the surface of the cell served by the base station, the extra-cell interference). These marks, called in what follows *geometric marks* (cf.Section 2.6), are the object of interest in this study.

In this work, we shall consider consider stationary marks. Recall from Section 2.6, that they are defined as mappings $m : \mathbb{R}^2 \times \mathbb{M} \mapsto \mathbb{R}$ such that, for any $u, v \in \mathbb{R}^2$, and any $\phi \in \mathbb{M}$, $m(u-v, S_v\phi) = m(u,\phi)$.

We denote by $M_i := m(u_i, \Phi)$ the geometric mark of the point u_i of Φ. Note that the process $\tilde{\Phi} := \sum_i \delta_{(u_i,M_i)} \sim \sum_i M_i \delta_{u_i}$ is a stationary marked point process.

In the remainder of this chapter, we shall consider the basic examples presented in Section 2.6. For a better understanding of the context, we interpret the points u_i as locations of base stations of some cellular network. The shot-noise mark (cf. Example 2.6.2) can be considered as the total power (usually interpreted as interference) received at a given base station in the network, transmitting with unit power. The Voronoi cell (cf. Example 2.6.4) is a fundamental model for cellular networks. Its area and moment of inertia can be interpreted as very simple proxys for the traffic demand and the cell load of a base station (see [21, Section 4.1.8]). The Voronoi shot-noise mark (cf. Example 2.6.5), can be also interpreted as the interference, however with the stations transmitting the signals with a power proportional to their Voronoi cells. (Similar kinds of dependence, with more complex expressions, can be recognized in the cell load model of [103].)

4.1.2 Problem formulation

Let us now formulate the main problem studied in this chapter, that is the problem of the *learning of geometric marks.* Suppose the marking function m is not known explicitly. One observes only some realizations of the marked point process $\tilde{\Phi}$ with points restricted to some finite observation window W. Denote these realizations by

$$\widetilde{\phi}_k = \sum_i m_i(k) \delta_{u_i(k)},$$

with $u_i(k) \in W$, $k = 1, \ldots, n$, forming a training set of data. The problem consists in learning the function m from the training set, so as to be able to calculate approximations of the unobserved marks $m_i = m(\phi, u_i)$ for a new realization $\phi = \sum_i \delta_{u_i}$ of Φ.

4.1.3 Scattering moments of marked point processes

Recall from Section 3.2, the wavelet transform of a marked point process is defined by

$$\begin{aligned}\tilde{\Phi} \star \psi : &= \int_{\mathbb{R}^2} \psi(v-u) m(\Phi, u) \tilde{\Phi}(du) \\ &= \sum_{u_i \in \Phi} M_i \psi(u - u_i),\end{aligned}$$

and that first and second order scattering moments (Section 3.2) can similarly be extended to marked point processes. Scattering moments have not yet been extensively theoretically studied in the context of the theory of point processes. Some asymptotic properties at small and large scale ($j \to -\infty$ and $j \to \infty$ respectively) have been presented in section 3.3, extending results of [31].

As explained in Section 3.6, for a given realization $\widetilde{\phi}$ of $\tilde{\Phi}$ observed in a finite window W, empirical scattering moments $\hat{S}_1\tilde{\Phi}(j,\theta)$ and $\hat{S}_2\tilde{\Phi}(j_1,\theta_1,j_2,\theta_2)$ are obtained by replacing the expectation by a spatial averaging of the fields $|\widetilde{\phi} \star \psi_{(j,\theta)}(u)|$ and $||\widetilde{\phi} \star \psi_{(j_1,\theta_1)}| \star \psi_{(j_2,\theta_2)}(u))|$, respectively, over $u \in W$. When W increases, these (empirical) moments become asymptotically non-biased estimators of $S_1\widetilde{\Phi}(j,\theta)$ and $S_2\widetilde{\Phi}(j_1,\theta_1,j_2,\theta_2)$, respectively, provided $\widetilde{\Phi}$ is ergodic (under some technical conditions about the wavelet

ψ), (cf ergodic theorem for point processes [39, Theorem 13.4.III], or Section 2.7). For some other properties of the scattering moment estimators see [33, Section 5.1].

The empirical scattering moments are calculated over a finite square window W by periodizing the input signal. As a consequence, the value of the mark is calculated as if the point process were periodic.

As explained in Section 3.6, in order to define a finite wavelet family, we need to choose a finite number of scales $j \in \{0, \cdots, J-1\}$. Assume that there is also a wavelet at scale $j_{\min}$ small enough so that it separates all the points in every pattern (theoretically corresponding to $j_{\min} \sim -\infty$). At this scale, $\hat{S}\widetilde{\Phi}(j_{\min}, \theta)$ is proportional to the empirical mean $\frac{1}{|W|}\int_{W^2} m\phi(du)$, and does not depend on θ.

The first and second empirical scattering moments calculated on $\widetilde{\Phi}$ form finite-dimensional vectors. We denote by $\hat{\mathbf{S}}\widetilde{\Phi}$, and $\widehat{\mathbf{S2}}\widetilde{\Phi}$, respectively, the vector of the first moments and the *joint* vector of the first and second moments. If there is no ambiguity, for simplicity, in what follows we shall simply call them scattering moments of $\widetilde{\Phi}$ (without distinction between first and second order).

4.1.4 Learning of marks via scattering moments

Recall from Section 4.1.2 that our goal is to learn the marking function m from the training set of data, which consists of examples of realizations of a marked point process $\tilde{\Phi}$ in a finite window. Note that this problem consists in the interpolation of the function m on the space $\mathbb{R}^2 \times \mathbb{M}$, and, due to the complexity of this space, it is not amenable to any direct statistical approach. To overcome this difficulty, we map this original problem to some finite dimensional regression problem, and solve it using classical tools.

More specifically, in our main approach, we shall capture the function m through the relation between the vector of the first order scattering moments $\hat{\mathbf{S}}\widetilde{\phi}$ of the marked point pattern, and the two order moments $\widehat{\mathbf{S2}}\phi$ of the non-marked one. This relation is established using a regression model described in Section 4.1.4. In order to be able to use it to estimate the marks $m_i = m(u_i, \phi)$, we need next to solve an inverse problem described in Section 4.1.4. It consists in reconstructing marks from the regressed scattering moments, knowing the positions of the points in ϕ.

Linear regression

Let $X_k := \widehat{\mathbf{S2}}\phi_k$ and $Y_k := \hat{\mathbf{S}}\widetilde{\phi}_k$, $k = 1, \ldots, n$, be the vectors of the (empirical) scattering moments, calculated for the training data set consisting of n realizations of the point process $\widetilde{\Phi}$, where $\phi_k = \sum_i \delta_{u_i(k)}$ and $\widetilde{\phi}_k = \sum_i m_i(k)\delta_{u_i(k)}$ (cf. Sections 4.1.2 and 4.1.3). Our goal is to find a common relation between X_k and Y_k for all samples k, and the simplest possible one is a linear relation, represented by some matrix $\mathbf{B}$ and vector β_0 such that

$$\mathbf{B}X_k + \beta_0 \approx Y_k \quad \text{for all } k = 1, \ldots, n. \tag{4.1}$$

If the linear function does not allow one to capture the dependence, we can use kernel regression, or more advanced machine learning methods. For simplicity, in this study, we focus on the linear ridge regression, detailed in what follows.

To find the relation (4.1), we will use linear ridge model; cf [97, Section 7.5]. For $\lambda = (j, \theta)$, denote by $\boldsymbol{\beta}(\lambda)$ the line of the matrix $\mathbf{B}$ (4.1), and similarly the component

$\beta_0(\lambda)$ for the vector $\boldsymbol{\beta}_0$. Let $Y_k(\lambda) := \hat{\mathbf{S}}\widetilde{\phi}_k(\lambda)$ be the (j,θ)-component of $\hat{\mathbf{S}}\widetilde{\phi}_k$. The ridge model consists in minimizing the regularized sum of the squared residuals

$$\sum_{k=1}^{n}[\boldsymbol{\beta}(\lambda)X_k + \beta_0(\lambda) - Y_k(\lambda)]^2 + \gamma(\lambda)||\boldsymbol{\beta}(\lambda)||^2\,,$$

for some regularization parameter $\gamma(\lambda) \geq 0$, chosen by cross-validation (to minimize this squared residuals on the validation set: a subset of the training set), where $\|\cdot\|$ is the Euclidean norm. This model admits a well known explicit solution in the form

$$[\hat{\beta}_0(\lambda), \hat{\boldsymbol{\beta}}(\lambda)]^t = (\mathbf{X}^t\mathbf{X} + \gamma(\lambda)\mathbf{I})^{-1}\mathbf{X}^t\mathbf{Y}(\lambda), \tag{4.2}$$

where, for all $k = 1, \cdots, n$, $\mathbf{X}(k,\cdot) = [1, X_k(\cdot)]$, $\mathbf{Y}(\lambda)(k) = Y_k(\lambda)$, and $\mathbf{I}$ is the identity matrix of dimension $n+1$.

Using (4.2), one can calculate approximations $\hat{S}\widetilde{\phi}(\lambda)$ of the scattering moments of a new marked configuration $\widetilde{\phi} = \sum_i m_i(k)\delta_{u_i}$ observing only its points $\phi = \sum_i \delta_{u_i}$

$$\hat{\hat{S}}\widetilde{\phi}(\lambda) := \hat{\boldsymbol{\beta}}(\lambda)\,\widehat{\mathbf{S2}}\phi + \hat{\beta}_0(\lambda),$$

where $\widehat{\mathbf{S2}}\phi$ is the vector of the first and second order scattering moments calculated on ϕ (without marks). Remember, expression (4.2) requires the tuning of the regularization parameters $\gamma(\lambda) \geq 0$, usually needed in high dimensional regression problems when the matrix $\mathbf{X}^t\mathbf{X}$ is not invertible. The ordinary least square (OLS) estimator corresponding to $\gamma(\lambda) = 0$ is usually not performing well in this case.

Reconstruction

Using linear ridge, we calculate approximations $\hat{\hat{S}}\widetilde{\phi}(j,\theta)$ of the scattering moments of a new marked configuration $\widetilde{\phi}$, observing only ϕ. Denote the whole vector of $\hat{\hat{S}}\widetilde{\phi}(j,\theta)$ by $\hat{\hat{\mathbf{S}}}\widetilde{\phi}$. From $\hat{\hat{\mathbf{S}}}\widetilde{\phi}$ we estimate (reconstruct) unknown marks m_i of $\widetilde{\phi}$ by looking for a solution to the following minimization problem

$$\arg\min_{\widetilde{\phi}':\phi'=\phi} ||\hat{\mathbf{S}}\widetilde{\phi}' - \hat{\hat{\mathbf{S}}}\widetilde{\phi}||^2, \tag{4.3}$$

where we minimize over all arbitrarily marked configurations $\widetilde{\phi}'$ having the same atoms as ϕ, and $\hat{\mathbf{S}}\widetilde{\phi}'$ denotes the scattering moment calculated for $\widetilde{\phi}'$. It should be noted that (4.3) is a non convex optimization problem. To solve it, we use the L-BFGS-B algorithm, which is a limited-memory algorithm for solving large nonlinear optimization problems subject to simple bounds on the variables (cf. [77, 121]).

4.1.5 Learning via local distance representation

In the following approach, taken as a benchmark, we consider each marked point $(u_i(k), m_i(k))$ of each realization $\widetilde{\phi}_k$, $k = 1, \ldots, n$, of the training set, *along with some neighborhood*, as one element of the new training data set. Therefore, in this case, the training set consists of point patterns having a marked point in their center, surrounded by some neighboring points. Using the linear ridge regression method described in Section 4.1.4, we regress the marks of those central points with respect to the vectors containing all inter-point distances in the considered point neighborhood, ordering the points according to the distance to the central point. Note that, in this approach, there is no reconstruction phase, as the marks are directly approximated. The main parameter of this approach is the number of neighbouring points K of each central point.

4.2 Numerical results

In this section, we provide the details of our numerical study of the problem considered in this chapter. We begin by describing some general assumptions and procedures.

4.2.1 General numerical framework

Scattering moments approach

For our numerical experiments, for each specific geometric model, we create a data set (denote it by $\mathcal{X}$) of Poisson point patterns ϕ_k (with constant intensity to be specified) using the R software and its package *Spatstat* for point process analysis [7]. The points are considered in the unit square window, and their marks are analytically computed (according to the given model) using *Spatstat* with the window mapped to the torus, thus leading to the marked point patterns $\widetilde{\phi}_k$, $k = 1, \ldots, n$. The size of the data set $\mathcal{X}$ is $n = 10$k marked point patterns.

The (empirical) scattering moments are computed on these point patterns (with and without marks) using *ScatNet* software [3] developed in Matlab (no *Spatstat* implementation is available yet). It uses the Morlet wavelet ([58], cf. eq. (3.3)). Because this software uses raster images, we convert each marked point pattern of $\mathcal{X}$ into images of size $2^7 \times 2^7$ pixels (removing images with points corresponding to the same pixel).

The following family of scattering moments are computed. Recall from Section 4.1.3 that there is a minimal scale $j_{\min}$ such that, at this scale, the first order scattering moments correspond to the empirical mean measure, and do not depend on the angle θ (we can therefore take only one). We further choose $J = 7$, and $L = 8$ (i.e. $\theta \in \{0, \cdots, 7\pi/8\}$). Thus, there are $1 + 8 \times 7 = 57$ first order scattering moments (the dimension of the regressed vectors $\hat{\mathbf{S}}\tilde{\phi}_k$ is 57), and $8 \times 8 \times \frac{7\times 6}{2} = 1344$ second order scattering moments (thus making the dimension of the explanatory vectors $\widehat{\mathbf{S2}}\phi_k$ equal to $57 + 1344 = 1401$).

We use the linear ridge regression, described in Section 4.1.4, on the data set $\mathcal{X}$. To optimize the regression parameters, we use a 5-fold cross-validation [97, Section 1.4.8] on $\mathcal{X}$.

Having calculated the estimators $\hat{\hat{\mathbf{S}}}\widetilde{\phi}$ of the first order scattering moments $\hat{\mathbf{S}}\widetilde{\phi}$ for the point patterns in the test set, we use the L-BFGS-B algorithm to solve the inverse problem (4.3), i.e. to reconstruct the marks. This is a steepest descent algorithm for which it is important to optimize (ideally via cross-validation) the number of iterations; to be explained in Section 4.2.3. Using a rule of thumb (no formal cross-validation), we fix the number of steps so as to minimize the mean square error (MSE) on the test data set.

Benchmark

As explained in Section 4.1.5, for every image in the data set $\mathcal{X}$, we consider each point of the image, along with its K neighbours as an element of a new data set $\mathcal{X}'$. More precisely, we take the first 20k points of the images of $\mathcal{X}$, and create $\mathcal{X}'$ with K neighbours of the chosen point. The value of K depends on the range of dependence of a given mark model; it is experimentally discovered for each mark, as will be explained in Section 4.2.3. We use $\mathcal{X}'$ to regress the marks of the central points directly, using linear ridge regression with respect to the vector of dimension $K(K-1)$ of the local distance matrix. Note that there is no reconstruction phase in the benchmark method.

Validation methodology

To test our main approach, we produce an independent data set of 100 marked point pattern realizations, and we use the first 100 points of this set for the benchmark approach.

For all points in the respective test sets, we compute Q-Q plots, the root mean square error (RMSE), the normalized RMSE with normalization by the range ($\max - \min$) of the marks (NRMSE1), and the normalized RMSE with normalization by the mean of the marks (NRMSE2).

In order to study the quality of the representation of the mark point patterns by their first order scattering moments (independently of the regression), we also perform the reconstruction of marks 4.1.4 directly from the exact (and not regressed) scattering moments $\hat{\mathbf{S}}\widetilde{\phi}$. Note, these latter are not exact marks but marks reconstructed from exact empirical scattering moments of the given point pattern. The gap between them and the marks reconstructed from the regressed moments allows one to apprehend the error introduced by the regression and the one introduced by the reconstruction.

4.2.2 Results

We now present our numerical study of different mark models, presented in Section 2.6. The observed results are discussed in Section 4.2.3.

Shot-noise

We consider shot-noise marks (cf. Example 2.6.2), with the response function $\ell(r) := \max(10r, 0.6)^{-3}$. For this example the data set $\mathcal{X}$ consists of 10k realizations of Poisson point process with intensity 40 (recall, we consider the unit square window). In order to observe how the performance of our approach depends on the size of the data set, we first use only 5k elements of $\mathcal{X}$ as the training set for the linear ridge regression method. Next, we use the whole training set, with linear ridge regression. For the benchmark method, we choose a neighbourhood of $K = 15$ points. The number of iterations for the reconstruction from the exact scattering moments and estimated ones is taken, respectively, 30 and 4; see Section 4.2.3 for the explanation.

Figure 4.3 presents an example of the reconstruction of marks of one given image, with our method, and Figure 4.4 shows Q-Q plots for various reconstruction variants. Table 4.1 gives the reconstruction errors. Recall that the Q-Q plot (c) and the last column of the table represents the error of the reconstruction of the marks from the exact (and not regressed) scattering moments.

Nearest neighbour distance

The data set for this example is also made of 10k realizations of a Poisson point process with intensity 40. In this case, in Figures 4.5 and 4.6 and Table 4.2, we only show the results for the entire training set. For this mark model we do not consider the benchmark,

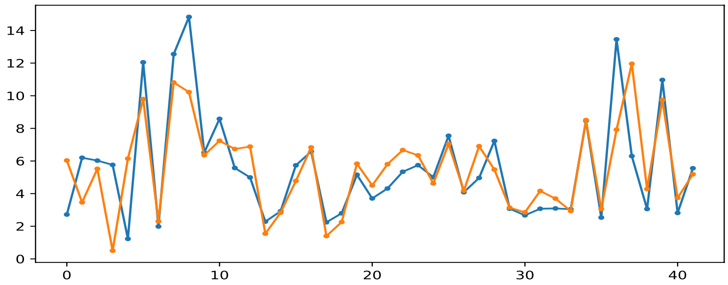
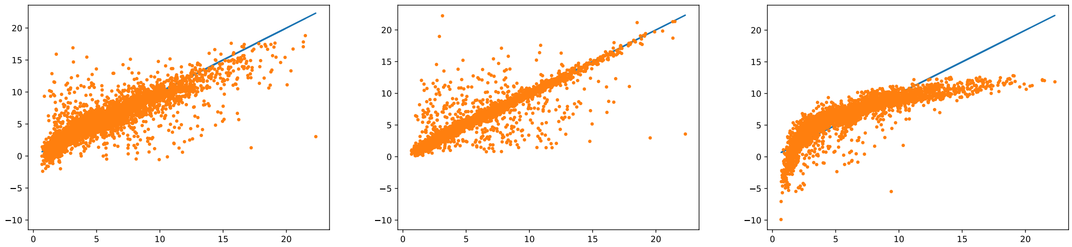


Figure 4.3: Reconstructed image example (mapped to 2D; the peaks of the lines correspond to the values of marks of points numbered in lexicographic order; orange curve — reconstruction, blue — exact values).



(a) (b) (c)

Figure 4.4: Shot-noise reconstruction. Q-Q plots (reconstructed mark in function of its true value) for 100 test images (a) estimated scattering moments, (b) exact scattering moments, (c) benchmark.

method	scattering 5 000	scattering 10 000	exact scatt.	benchmark
RMSE	1.99	1.98	1.64	1.98
NRMSE1	9.21e-2	9.15e-2	7.56e-2	9.14e-2
NRMSE2	3.32e-1	3.29e-1	2.75e-1	3.29e-1

Table 4.1: Shot-noise reconstruction errors using different methods.

because the nearest neighbour distance is equal to the first element of the distance matrix. Thus, the scattering moment approach cannot do better. The number of iterations for the reconstruction from the exact scattering moments and estimated ones is taken, respectively, 250 and 8.

method	estimated scatt.	exact scatt.	benchmark
RMSE	3.14e-2	2.06e-2	–
NRMSE1	1.21e-1	7.94e-2	–
NRMSE2	3.94e-1	2.58e-1	–

Table 4.2: Nearest neighbour distance; reconstruction error

Voronoi cell surface area

For this example, the training set $\mathcal{X}$ consists of 10 000 realizations of Poisson point process with intensity 30. The reconstruction results are presented on Figures 4.7 and 4.8

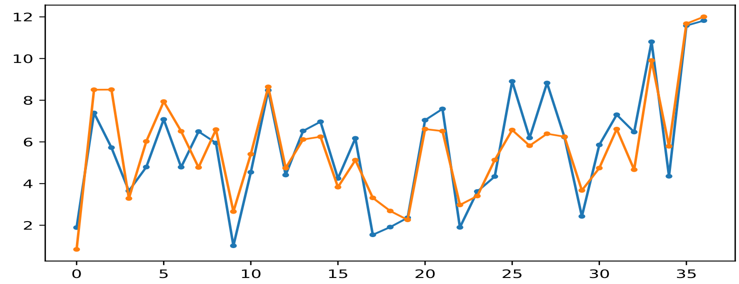
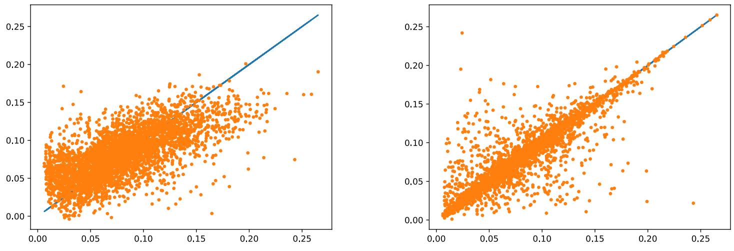


Figure 4.5: Reconstructed image example for the nearest neighbour model.



(a) (b)

Figure 4.6: Nearest neighbour distance reconstruction; example and Q-Q plots using: (a) estimated scattering moments, (b) exact scattering moments.

and in Table 4.3. The neighborhood for the benchmark is very large $K = 35$ (more than the average number of points) because the sum of the areas of the Voronoi cells in the finite window is constant equal to the total window surface area, which introduces a strong global dependence for this mark. The number of iterations for the reconstruction from the exact scattering moments and estimated ones is taken, respectively, 30 and 6.

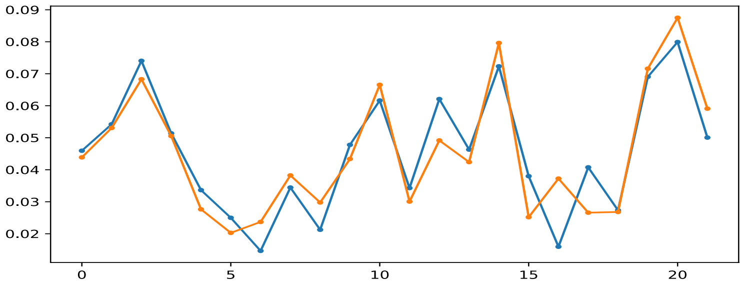
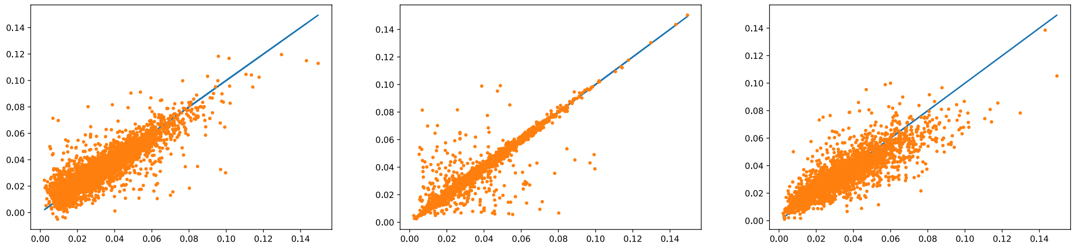


Figure 4.7: Reconstructed image example for the Voronoi cell surface model.

method	estimated scatt.	exact scatt.	benchmark
RMSE	9.71e-3	7.33e-3	9.60e-3
NRMSE1	6.60e-2	4.98e-2	6.52e-2
NRMSE2	2.87e-1	2.16e-1	2.85e-1

Table 4.3: Voronoi cell surface reconstruction errors.



(a) (b) (c)

Figure 4.8: Voronoi cell surface area reconstruction; example and Q-Q plots using: (a) estimated scattering moments, (b) exact scattering moments, (c) benchmark.

Voronoi cells moment of inertia

For this example, similarly to the previous example, the training set $\mathcal{X}$ consists of 10k realizations of a Poisson point process with intensity 30. The results are presented on Figures 4.9 and 4.10 and in Table 4.4. The neighborhood for the benchmark is $K = 15$ (note that the global dependence specific for the Voronoi surface area does exist here). The number of iterations for the reconstruction, from the exact scattering moments and the estimated ones, are chosen as 150 and 8, respectively.

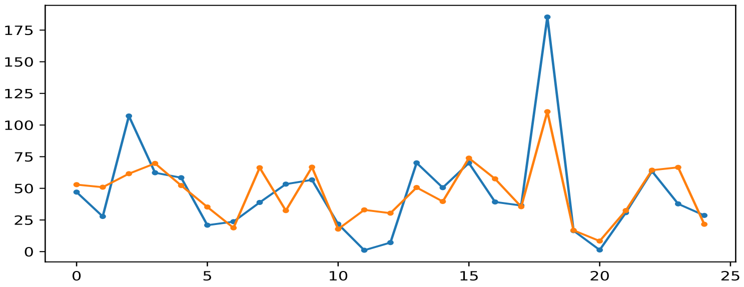
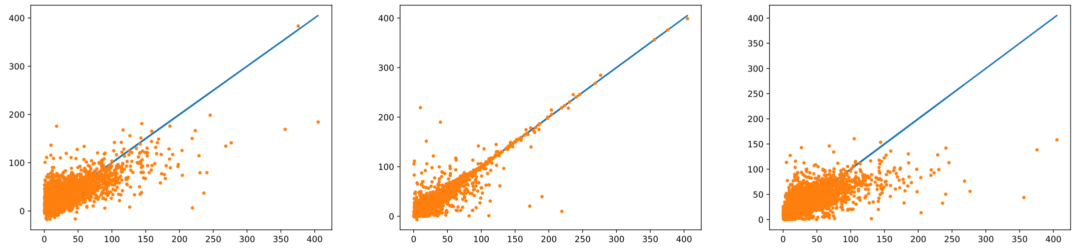


Figure 4.9: Reconstructed image example for the Voronoi cell moment of inertia model.



(a) (b) (c)

Figure 4.10: Voronoi cell moment of inertia reconstruction; example and Q-Q points using (a) estimated scattering moments, (b) exact scattering moments, (c) benchmark.

method	estimated scatt.	exact scatt.	benchmark
RMSE	2.42e-4	1.37e-4	2.65e-4
NRMSE1	5.98-2	3.39e-2	6.53e-2
NRMSE2	7.12e-1	4.04e-1	7.74e-1

Table 4.4: Voronoi cell moment of inertia reconstruction errors.

Voronoi shot-noise

For this example, similarly to the previous example, the training set $\mathcal{X}$ consists of 10k realizations of a Poisson point process with intensity 30. The results are presented in Figures 4.11 and 4.12 and in Table 4.5. The neighborhood for the benchmark is $K = 15$. The number of iterations for the reconstruction, from the exact scattering moments and the estimated ones, are chosen as 50 and 5, respectively.

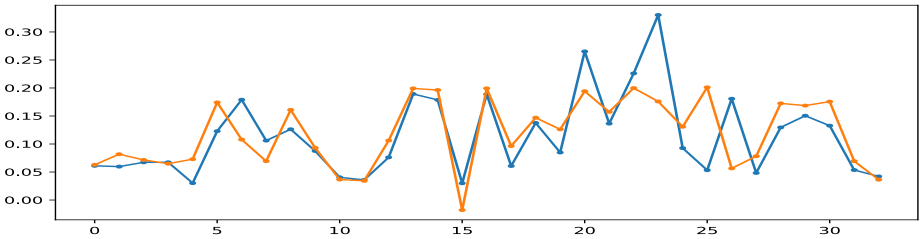
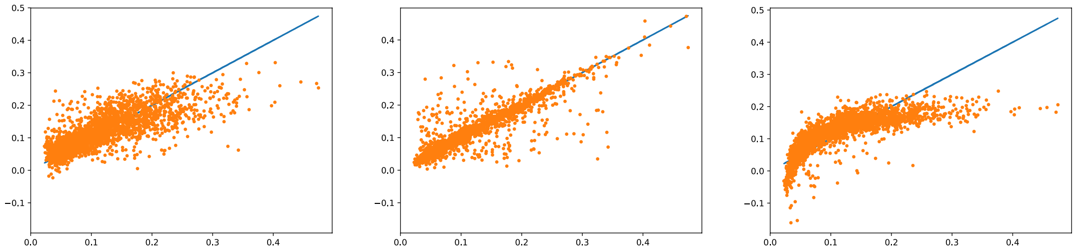


Figure 4.11: Reconstructed image example for the Voronoi shot noise model.



(a)

(b)

(c)

Figure 4.12: Voronoi shot-noise reconstruction; example and Q-Q points using (a) estimated scattering moments, (b) exact scattering moments, (c) benchmark.

method	estimated scatt.	exact scatt.	benchmark
RMSE	4.18e-2	3.02e-2	4.29e-2
NRMSE1	9.24-2	6.69e-2	9.50e-2
NRMSE2	3.76e-1	2.72e-1	3.86e-1

Table 4.5: Voronoi shot-noise reconstruction errors.

4.2.3 Discussion

The following remarks can be formulated regarding the observed results.

Reconstruction from the exact and estimated moments

Observing the Q-Q plots and (N)RMSE's of the marks reconstructed from the exact first order scattering moments, we see 'how much information' they effectively carry regarding the marking function. While all marks are relatively well reproduced in this way, the quality of the reconstruction depends on the type of dependencies represented by a given mark. For example, the shot-noise and the surface areas of the Voronoi cell are more easy to represent than the nearest neighbour. This can be explained by different sensitivity (stability) to small deformations of the point pattern, with a precise formulation yet to be theoretically studied on the ground of point processes.

A typical significant error in the signal reconstruction, both from the exact and the estimated scattering moments, consists in the swap of a large and a small mark of two neighbouring points (e.g. the points number 36 and 37 on Figure 4.3), not leading to a significant modification of the considered scattering moments. This effect can be seen also on the Q-Q plots where many points significantly far from the diagonal appear in symmetric pairs. We believe these swapping errors should occur less often for more regular (less clustering) point processes than the considered Poisson one. Examples of such point processes are determinantal point processes; see [16, 17] for a clustering comparison theory. Future studies should investigate this issue.

As already mentioned, it is important to properly tune the number of iterations of the steepest descent algorithm used in the reconstruction phase, preventing it from going too deeply into potential local minima. We observe the following *local-global reconstruction quality trade-off*: while some number of initial iterations makes all the marks approach their right values, further iterations improve the quality of approximation of some subset of marks at the price of degrading this quality for the remaining ones, as illustrated in Figure 4.13. This trade-off can be observed on average by watching the RMSE on the test set, which first decreases and then increases. We use this observation to choose an optimal number of iterations. It is larger for the reconstruction from the exact scattering moments than for the reconstruction from the estimated ones, where this effect is less visible.

Choice of neighbourhood

A crucial benchmark parameter is the number of neighbors to be considered in the local representation. They should be selected in function of the type of mark dependence. If no a priori information is available, this can be done observing the RMSE on the test set as shown in Table 4.6. Observe that the constant sum of the Voronoi cell areas makes them globally dependent, unlike the Voronoi moments of inertia.

model K	10	15	20	35
Shot-noise	2.00	1.98	1.98	–
Voronoi area	1.04e-2	1.00e-2	9.80e-3	9.60e-3
Voronoi inertia	2.76e-4	2.65e-4	2.80e-4	–
Voronoi shot-noise	4.32e-2	4.30e-2	4.29e-2	–

Table 4.6: Choice of the number of neighbours K for the benchmark approach and the corresponding RMSE.

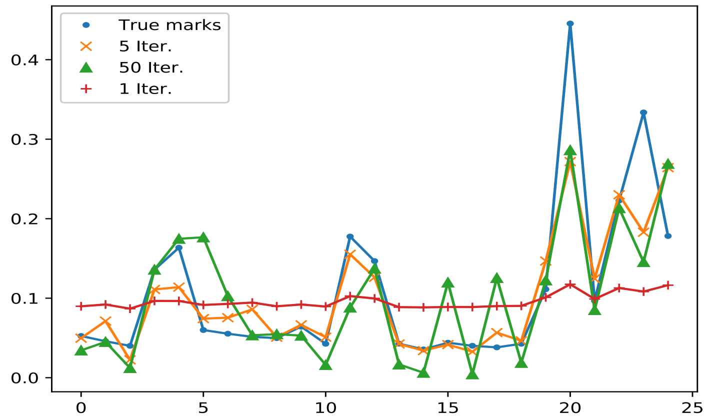


Figure 4.13: Reconstruction of a sample of the Voronoi shot-noise marks from the estimated scattering moments after 1, 5 and 50 iterations. 5 iterations give the smallest RMSE of 4.17e-2, while 50 iterations used for the reconstruction from the exact scattering moments improve the reconstruction of some marks but give worse RMSE of 5.26e-2.

Quality of the regression

The regression relative errors

$$\epsilon^m(j,\phi) := \frac{|\hat{\hat{S}}\widetilde{\phi}(j,\theta) - \hat{S}\widetilde{\phi}(j,\theta)|}{\hat{S}\widetilde{\phi}(j,\theta)}$$

calculated on the training and test set, presented on Figure 4.14, show that there is no overfitting in the regression of the scattering moments.

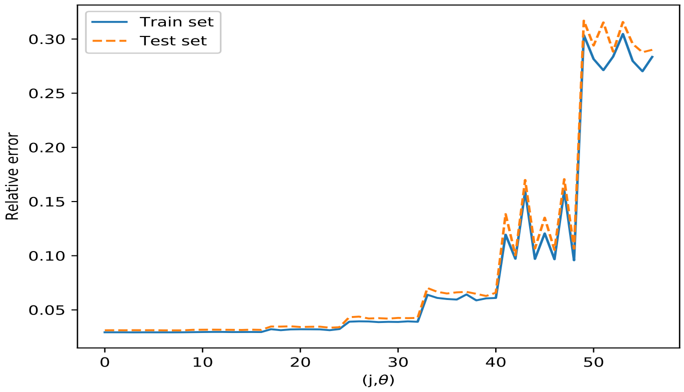


Figure 4.14: Relative errors $\epsilon^m(j,\theta)$ for (j,θ) in order $(0,0),(1,\pi/8),\ldots,(2,0),\ldots,(7,7\pi/8)$; calculated on training and test set.

Scattering moments versus benchmark

With respect to the RMSE, the scattering moments approach outperforms the benchmark for all marks except for the Voronoi surface area with the neighborhood taking almost all points and, for obvious reasons, the nearest neighbor distance. A significant difference can be observed on the Q-Q plots of the shot-noise example in Figure 4.4, and the Voronoi shot-noise example in Figure 4.12, where the benchmark essentially fails to capture the marking function. This shows that the benchmark approach might not be appropriate in the case of long-range dependent marks.

4.3 Discussion

Motivated by the stochastic-geometry problems related to wireless networks, in this study we have discussed how to learn a point-marking function dependent on the configurations of points. We propose two different approaches to address the problem, using tools from statistical learning. The baseline approach extracts the geometric information for each point based on the matrix distance of its nearby points. It is then solved using the ridge linear regression method, and the difficulty lies in the choice of the number of nearby points. The other approach uses multi-scale wavelet scattering moments to define a global feature vector for all the points in the domain, and another feature vector for the points with marks. The relation between the two feature vectors is also learned using linear regression. We then solve a non-convex optimization problem to reconstruct the marks, from the predicted feature vector with marks. These feature vectors are translation-invariant and stable to deformation of the domain.

Depending on the nature of the clustering of the points and the regularity of the marking function, we find that the scattering moments predict as good as the baseline approach, showing that they capture well the geometric property. In case where marks depend on points in a non-local way, the scattering moments method outperforms the benchmark approach.

Regarding practical applications, it seems promising to combine geometry with local demand, in order to obtain an operational method for cell load prediction in cellular networks.

Future directions to be explored may include a combination of the two approaches, e.g. by representing the neighbourhood of each point not by its distance matrix, but by its scattering moments. Replacing the linear regression by others such as kernel regression or neural network approaches are also of potential interest. Additionally, the impact of regularity or clustering of the underlying point patterns on the quality of scattering representation ought to be better understood.

Limitations of scattering moments This work demonstrated the ability of wavelet scattering moments to adequately capture important geometric structures in point patterns, and their amenability to be used in a learning task involving point processes. However, they do not fully characterize these processes. Indeed, scattering moments do not capture the dependencies between wavelet coefficients at different scales and orientations. Interpreting the first order scattering moments as measuring 'how much structure at each scale and orientation is present in the data', second order coefficients measure the partition of these structures between larger scales, but not the joint presence of structures at different scales. Additionally, due to the wavelet symmetries and the use of the modulus operator, this representation is invariant to changes of sign in the marking function, and to space reversal ($u \to -u$), which probably explains the error due to the swapping of marks between nearby points. Other wavelet-based representations, such as the WPH covariance presented in Section 3.5, may help capture this kind of missing information. To study the quality of WPH covariance moments as descriptors of the geometry in point processes, we shall use them to define a model for stationary point process distributions, as detailed in the next chapter.

Chapter 5

Point process synthesis using wavelet phase harmonics and gradient descent

In order to generate new realizations of a stochastic process of which we have only one realization, one needs to build a probabilistic model which approximates the distribution of this process, and from which we can efficiently sample from. In this chapter, we are interested in generative models for stationary, ergodic point processes. Such models are of interest in a wide range of applications [64, Section 6], for instance biology [8, 46], ecology [116], turbulent flows in atmosphere science [48, 87, 84, 47], or cosmology [108, 104]. In many of these domains, the observed patterns exhibit complex structures, with a large number of particles (such as filaments in cosmology, or vortices in turbulent flows).

More particularly, we seek to generate realizations formed by a large number of particles, with both short and long range interactions. Figure 5.1 shows some examples of distributions that we shall consider. Currently, for such diverse geometries exhibiting long-range interactions, which naturally appear e.g. in cosmology or turbulent flows in physics and atmosphere science, no model has been proposed in the literature on point processes. To address this problem, we shall introduce a statistical model inspired from the maximum-entropy principle ([65], cf. Section 2.10), to approximate such point process distributions and simulate new realizations.

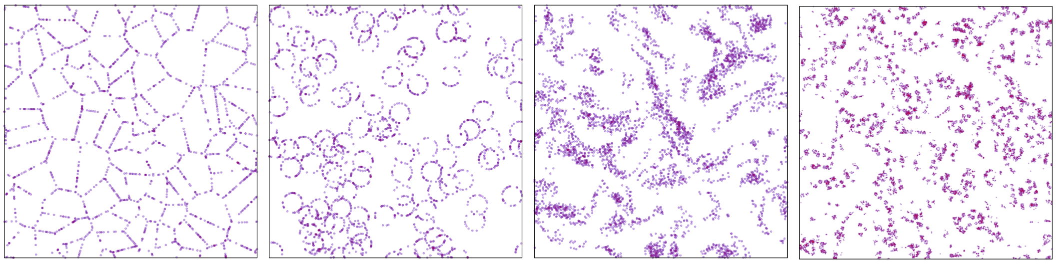


Figure 5.1: Samples of point processes of various geometries. The number of points ranges from 1000-13000.

Recall from Section 2.10, maximum entropy models are based on the description of the distribution with a set of moments, or their estimators. Intuitively, this means that the model is 'as random as possible' under certain constraints, based on the information captured by the moments.

5.1 Generation of point processes with maximum entropy models

When defining a generative model based on the maximum entropy principle, there are three underlying problems:

1. Choosing the moments that will describe the distribution. They should be informative enough to capture the geometric structures characterizing the distribution. On the other hand, they should be accurately estimated from a single observation, so the number of moments should not be too large.

2. Specifying a model deriving from these moments. This can be done by defining a maximum entropy model such as the macro-canonical model (maximizing the entropy under expectation constraints), or the micro-canonical model (maximizing the entropy under path-wise constraints).

3. Generating new samples from the model. In the micro-canonical setup, this can be done by minimizing an energy, that defines the set of admissible realizations. The minimization method must make it possible to generate diverse low energy samples without being too costly in terms of calculation.

Recall from Section 2.10, under some technical assumptions (in particular, having density with respect to the reference Poisson distribution), the solution of the macro-canonical model is given by the Gibbs point process. Sampling from the macro-canonical model is usually computationally very expensive (c.f. [32]). Therefore, we shall place our model in the micro-canonical setup, as detailed in Section 5.2. The main challenges then reside in the problems 1 and 3.

In Section 5.3, we present our method to address the problem of generating new samples: we minimize the energy of a new sample by moving the particles of an initial random configuration using the gradient of its energy with respect to the particles positions. In the point process literature, a classical method (see [110]) consists in updating an initial random configuration by successively replacing the particles one by one, with new particles located at random positions (we shall call this method random search). The major drawback of this method is its computational cost, as the optimization, which does not use gradient information to minimize the energy, requires a large number of energy evaluations. In fact, this method has been applied to generate point processes formed by a few hundred particles. On the other hand, advanced methods in the modelling of textures and non-Gaussian stationary processes allow for fast sampling by first drawing from an initial Gaussian distribution, and minimizing an energy by gradient descent on the amplitudes of the pixels of the image (see e.g. [92, 54, 32, 120]). Our approach leverages the efficiency of this sampling method, while ensuring that the resulting samples are atomic measures. The idea of moving the points according to their gradient is often used in molecular dynamics (e.g. [118, 119]), however it requires knowledge of the physical mechanisms behind the underlying process. Our statistical modeling approach has a potential to simulate new, complex particle configurations directly from one observation, when the underlying physical phenomena are very complicated to model.

This brings us to the other challenge that we address in this work: choosing the moments that we shall use to characterize the distribution. We shall use the wavelet phase harmonics (WPH) covariance moments, presented in Section 3.5, and briefly reviewed in Section 5.4. Recall that, similarly to scattering moments (cf. Section 3.2, and used in Section 4.1.3), they are spatial statistics based on coefficients computed from a wavelet transform of atomic measures. It is known that the covariance between the wavelet coefficients capture only second-order correlations, which are equivalent to the Bartlett spectrum (cf. eq. (3.29) and Section 2.8). To capture information beyond second-order correlations, we apply a non-linear phase harmonic operator to the wavelet coefficients. This operator acts on the complex phase of the wavelet coefficients, without changing their amplitude. The covariance between the resulting coefficients allows one to capture particles interactions across different scales.

Compared to high-order correlation functions, (see [109, Section 12.4.2]), our moments have the potential to define a sufficient set of statistics, while maintaining a small estimation error, which is similar to the second-order statistics. Other statistics often used in the point process literature (cf. Section 2.9), such as the k nearest neighbour distribution function suggested in [110], have a number of elements that grows with the intensity. Since there is only one observation, the number of moments should be limited, in order to control their estimation variance. The wavelet transform allows for direct control over the scales of the structures that we wish to capture, regardless of the intensity of the process. This property allows one to model point processes formed by a large number of particles with a limited number of moments.

The wavelet phase harmonic covariance descriptors are defined as spatial averages evaluated over a point process realization. In practice, the calculation of such descriptors can be done by discretization of the observation window in the form of a grid of pixels. However, making these descriptors differentiable with respect to the positions of the particles remains a challenge. In this regard, we describe in Section 5.5 a complete numerical scheme allowing one to solve this problem. It is based on a differentiable discretization of atomic measures. We further present a multi-scale optimization in the gradient descent, intended to avoid unwanted shallow minima of the energy.

In Section 5.6, we evaluate our model on some distributions exhibiting various geometric structures, like Cox point processes on the edges of Poisson-Voronoi tessellations and on the Boolean model with circular grains. Other processes we consider are Matern hard-core and cluster processes driven by Poisson processes with turbulent intensities. Their intensities are sampled from a turbulent flow simulated from Navier-Stokes equations [99]. Besides the visual inspection of the samples from our generative model, we evaluate second order correlations and compare the persistent homology diagrams, that has been proven useful for topological data analysis (see e.g. [36]).

To evaluate the possible benefits of our methods relatively to the well established model of [110], Section 5.7 presents a comparison between the two methods, in terms of speed of convergence of the optimization methods, and of quality of the models.

For simplicity, we present our model with strong assumptions about the data (stationarity, periodicity, square window). In practical cases, these assumptions would probably not be fully met, so we discuss in Section 5.9 how to relax these assumptions.

Finally, in Section 5.8, we propose a modified version of our model, that consists in adding a regularization term intended to improve the diversity of the generated samples. Moreover, we give a proof of a result stating the translation invariance of our model (see Section 5.3.2).

5.2 Micro-canonical framework for point processes

This section details the specific context of this study, and the definitions that we shall need.

We consider here stationary, ergodic point processes on $E = \mathbb{R}^2$ (cf. Sections 2.5 and 2.7). Recall that a point process Φ is stationary if its distribution is invariant to translation, i.e.

$$\mathcal{L}_{S_t\Phi} = \mathcal{L}_{\Phi},$$

where $\mathcal{L}_{\Phi}$ denotes the distribution of Φ, and $S_t\Phi(B) := \Phi(B + t)$, for all $B \in \mathcal{B}$, and all $t \in \mathbb{R}^2$. Note that S_t is a simple example of a push-forward operator on Φ, which is generally noted $F_{\#}\mu$ for a point measure μ and a measurable function $F : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, and is simply the displacement of its atoms by the function F

$$F_{\#}\mu = \sum_i \delta_{F(u_i)}.$$

Recall also that Φ is said to be ergodic if the empirical averages of integrable functions f, over windows of increasing sizes $W_s = [-s, s[\times[-s, s[$, converge almost surely to the expectation of f, i.e.

$$\lim_{s\to\infty} \frac{1}{|W_s|} \int_{W_s} f(S_u\Phi)\, du = \mathbb{E}[f(\Phi)] = \int_{\mathbb{M}} f(\mu)\, \mathcal{L}_{\Phi}(d\mu), \quad \mathbb{P}\text{-a.s.} \tag{5.1}$$

For a given $s > 0$, we denote by $\mathbb{M}^s$ the set of counting measures on W_s, and $\mathcal{M}^s$ its induced σ-algebra. We will consider W_s with addition and scalar multiplication modulo W_s. Also we shall denote by $\bar{S}_u$ the corresponding shift operator on $\mathbb{M}^s$ with torus correction on the window W_s.

For a point process Φ on $\mathbb{R}^2$, one can only observe realizations of Φ on bounded subsets of $\mathbb{R}^2$. For the remainder of this chapter, we shall consider realizations of point processes observed on a finite square window $W_s = [-s, s]^2$, for some $s > 0$. We denote by $\bar{\Phi}$ the restriction of Φ to W_s, that is $\bar{\Phi}$ is a point process on W_s such that, $\forall\, n \in \mathbb{N}, \forall\, (B_1, ..., B_n) \in \mathcal{B}(W_s)^n$, $(\Phi(B_1), ..., \Phi(B_n)) = (\bar{\Phi}(B_1), ..., \bar{\Phi}(B_n))$ in distribution (where $\mathcal{B}(W_s)$ stands for the Borel σ-algebra on W_s). A realization of Φ observed on W_s is therefore a realization of $\bar{\Phi}$, and will be noted $\bar{\phi}$.

As mentioned in Section 5.1, we shall place ourselves in the context of maximum entropy models (cf. Section 2.10). Such models are based on the following intuitive idea: given an observation pattern, we aim at finding new patterns that are similar to, but different from the observation. To this end, we define a notion of similarity by choosing a set of statistics that will be computed on the observation and on the new patterns. The two will be considered similar if their statistics match. Furthermore, if the chosen statistics describe sufficiently well the point process behind our observation, we do not want to add any more constraints, that is, we want to find new patterns 'as random as possible', under the constraints defined by the statistics. This can be formalized by maximizing the entropy of the model. They are used in large classes of stochastic models [56], and will inspire our particle gradient descent model. In this work, we place ourselves in the micro-canonical setting.

Consider a mapping

$$K : \mathbb{M}^s \mapsto \mathbb{C}^d,$$

for some $d \geq 1$, that shall define our set of statistics (e.g. statistics presented in Section 2.9, or wavelet-based statistics in Chapter 3).

Micro-canonical model The micro-canonical model is defined by replacing the expectation constraints $\mathbb{E}(K(\Xi)) = a$ with pathwise constraints. Let $\bar{\phi} \in \mathbb{M}^s$ be our observation sample, of unknown distribution. For all $\mu \in \mathbb{M}^s$, we define the energy of μ as:

$$E_{\bar{\phi}}(\mu) := \frac{1}{2}|K(\mu) - K(\bar{\phi})|^2. \tag{5.2}$$

The micro-canonical set of level ϵ, for some $\epsilon > 0$, is defined as

$$\Omega_\epsilon := \{\mu \in \mathbb{M}^s \, : \, E_{\bar{\phi}}(\mu) \leq \epsilon\}. \tag{5.3}$$

The micro-canonical model is defined as the distribution $\mathcal{L}$ that minimizes the KL divergence with respect to the reference distribution $\mathcal{L}_0$ under *pathwise constraints* requiring $\mathcal{L}$ to be supported on Ω_ϵ:

$$\arg\min_{\mathcal{L}} \quad \mathrm{KL}(\mathcal{L}, \mathcal{L}_0) \tag{5.4}$$

$$\text{given} \quad \int_{\mathbb{M}^s} \mathbb{1}(\mu \in \Omega_\epsilon)\mathcal{L}(d\mu) = 1, \tag{5.5}$$

where $\mathbb{1}(\cdot)$ is the indicator function. As stated in Proposition 2.10.5, if $\mathcal{L}_0(\Omega_\epsilon) > 0$, the solution to this problem (5.4) (5.5), is the measure $\mathcal{L}$ having a uniform density on $\mathbb{M}_s$ given by $\dfrac{d\mathcal{L}}{d\mathcal{L}_0}(\mu) = \dfrac{1}{\mathcal{L}_0(\Omega_\epsilon)}\mathbb{1}(\mu \in \Omega_\epsilon)$, $\mathcal{L}_0$-a.s.

In order to consider the micro-canonical model as a good approximation of the observation distribution, one usually aims at finding K satisfying the following properties:

(P1) *Concentration property*: The value of $K(\bar{\Phi})$ should concentrate around its mean, i.e. $K(\bar{\Phi}) \simeq \mathbb{E}[K(\bar{\Phi})]$ with high probability. A natural assumption is that the variance of $K(\bar{\Phi})$ is small.

(P2) *Sufficiency property*: The moments $\mathbb{E}[K(\bar{\Phi})]$, should *characterize the unknown distribution* as completely as possible. It requires that K has a strong (distributional) discriminate power.

A natural framework allowing one to address (P1) and (P2) consists in defining the descriptors $K = (K_1, \ldots, K_d)$ as a vector of empirical averages

$$K_i(\mu) = \frac{1}{|W_s|}\int_{W_s} f_i(\bar{S}_u\mu)\, du \qquad \mu \in \mathbb{M}^s, \tag{5.6}$$

for a sufficiently rich class of functions f_i on $\mathbb{M}^s$, and relying on the ergodic assumption (5.1) regarding Φ. These properties are needed in order to have a model that reproduces typical geometric structures in Φ, and generates diverse samples.

In this study, we shall consider that the number of points of our model in W_s is fixed. In such case it is customary to take homogeneous Poisson point process distribution conditioned to have exactly n points in W_s as the reference measure, which is equivalent to n points sampled uniformly, independently in W_s. We will note this distribution $\mathcal{L}_0^n$.

Sampling from the uniform density on Ω_ϵ efficiently remains an open problem. In the literature of stochastic process modelling, most sampling algorithms rely on the following method: one first samples from an initial, high-entropy measure, and iteratively minimize the energy (cf. (5.2)) of this sample until it reaches Ω_ϵ. By choosing a high entropy initial measure, one hopes that the resulting model also has a high entropy. Recall that the micro-canonical model has the highest entropy supported on Ω_ϵ. Contrary to the classical method in the point process literature [110, 69], which relies on random search, popular methods in image modelling use gradient descent to perform fast sampling in the micro-canonical set. However, optimizing the values of the image pixels does not guarantee that the resulting sample is an atomic measure. For these reasons, in what follows we propose a model based on the transport of a Poisson point process via a gradient descent algorithm.

5.3 Particle gradient descent model

In Section 5.3.1, we introduce the particle gradient descent model, that uses gradient descent on the positions of the particles of the sample. This model consists in using the gradient of a prescribed energy to move the points of an initial random configuration, until we obtain a pattern similar (in an informal sense) to the observation. We then present in Section 5.3.2 a theorem stating that this model preserves some basic invariances of the original distribution. This result, allowing us to gain some understanding about the entropy of our model, extends the results of [32]. While we focus here on planar point processes, our approach can readily be extended to any dimensions.

5.3.1 Particle gradient descent model

As in Section 5.2, let $\bar{\phi} \in \mathbb{M}^s$ be our observation sample of unknown distribution, and $K : \mathbb{M}^s \longrightarrow \mathbb{C}^d$, for some $d \geq 1$, a mapping defining our descriptors. We note the resulting energy $E_{\bar{\phi}}$ (cf. eq. (5.2)).

Let $\bar{\phi}_0$ sample from an initial distribution that we choose as $\mathcal{L}_0^{\bar{\phi}(W_s)}$ (i.e. the same number of particles as $\bar{\phi}$, drawn uniformly and i.i.d.). We minimize the energy of $\bar{\phi}_0$ through its gradient with respect to the particles positions. More precisely, we define the mapping

$$\begin{array}{cccc} F & : & \mathbb{M}^s & \longrightarrow & \mathbb{M}^s \\ & & \mu = \sum_i \delta_{u_i} & \longmapsto & \displaystyle\sum_i \delta_{u_i - \zeta \nabla_{u_i} E_{\bar{\phi}}(\mu)} \end{array} \tag{5.7}$$

for some gradient step $\zeta > 0$. The measure $F(\mu)$ can be seen as the push-forward $F_{\mu\#}\mu$ of the measure μ by the mapping $F_\mu(u) := u - \zeta \nabla_u E_{\bar{\phi}}(\mu)$. Note that the function F_μ depends on the measure μ which is pushed forward. For any initial point measure $\bar{\phi}_0 \in \mathbb{M}^s$ we define the successive point measures:

$$\bar{\phi}_n := F_{\bar{\phi}_{n-1}\#}\bar{\phi}_{n-1}, \qquad n \geq 1. \tag{5.8}$$

Pushforward of the point process distributions The pushforward operation $F_{\#\mu}$ on $\mathbb{M}^s$ induces the corresponding pushforward operation on the probability measures on $\mathbb{M}^s$, which are distributions of point processes. We denote this latter by $\mathcal{F}_\#$: For a probability law $\mathcal{L}$ on $\mathbb{M}^s$ $\mathcal{F}_\#\mathcal{L}(\Gamma) := \mathcal{L}(\{\mu \in \mathbb{M}^s : F_{\mu\#}\mu \in \Gamma\})$, for any $\Gamma \in \mathcal{M}^s$. Then, for an initial probability law $\mathcal{L}_{\bar{\phi}_0}$ on $\mathbb{M}^s$ we define the successive probability laws

$$\mathcal{L}_{\bar{\Phi}_n} := \mathcal{F}_\#\mathcal{L}_{\bar{\Phi}_{n-1}}, \qquad n \geq 1. \tag{5.9}$$

Note that $\mathcal{L}_{\bar{\Phi}_n}$ is the distribution of the point process $\bar{\Phi}_n$ obtained by n iterations of (5.8) starting from $\bar{\Phi}_0$ having law $\mathcal{L}_{\bar{\Phi}_0} = \mathcal{L}_0^{\bar{\Phi}(W_s)}$. Our model is defined by setting a fixed number of iterations as a stopping rule. Figure 5.2 illustrates the particle gradient descent method to generate samples in Ω_ϵ.

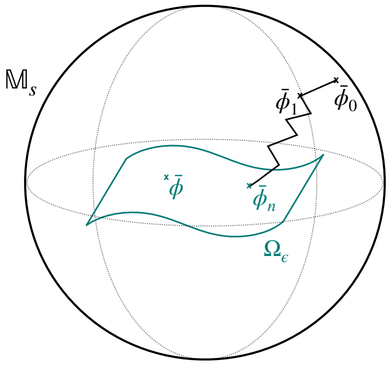


Figure 5.2: Overview of the particle gradient descent method: starting from $\bar{\phi}_0 \in \mathbb{M}_s$, we iteratively modify the configuration until reaching Ω_ϵ.

Observe that our model takes inspiration from the micro-canonical model, however there is no guarantee that our model reaches Ω_ϵ (defined in (5.3)), for any $\epsilon > 0$. By setting a fixed number of iterations and not rejecting any configuration, we make the implicit assumption that our model reaches a low energy level.

5.3.2 Leveraging invariances

One can leverage some a priori known invariance properties of Φ (for instance stationarity or isotropy), by building a model that satisfies the same invariance properties as Φ. By using the descriptor K with the same invariance, the particle gradient model respects these invariance properties. In particular, we obtain a stationary point process model when K is defined by the empirical averaging (5.6).

This requires some explanation, since invariance properties of the distribution of Φ do not, in general, imply any natural invariance of its restriction $\bar{\Phi}$ to W_s. Indeed, while some invariances can be observed on the torus for the distribution of Φ on $\mathbb{R}^2$ (the most popular being translation invariance), it does not imply the same for $\bar{\Phi}$ with respect to the translation on W_s. The latter, that we call *circular stationarity*, requires also Φ to be periodic. However, circular stationarity of the generated point process on large window W_s (as a distributional approximation of $\bar{\Phi}$) can be considered as a desirable ersatz of the stationarity of Φ. Indeed, in what follows we shall formulate a result saying that, when K and the distribution of $\bar{\Phi}_0$ are invariant with respect to some subset of *rigid circular transformations* on W_s, then the resulting model satisfies this property as well.

More specifically, a *rigid circluar transformation* on W_s is an invertible operator T on W_s of the form $Tu := Au + u_0$ for some orthogonal matrix A with entries in $\{-1, 0, 1\}$ and $u_0 \in W_s$. Note that the matrix A is restricted in integer entries for T to be a well defined invertible operator. It encapsulates translations, flips, and orthogonal rotations.

We say that:

- The initial probability law $\mathcal{L}_{\bar{\Phi}_0}$ of the model is invariant to the action of T if $\forall\, \Gamma \in \mathbb{M}^s$, $\mathcal{L}_{\bar{\Phi}_0}(T_{\#}^{-1}(\Gamma)) = \mathcal{L}_{\bar{\Phi}_0}(\Gamma)$.
- The descriptor K is invariant to the action of T if $\forall\, \mu \in \mathbb{M}^s$, $K(T_{\#}\mu) = K(\mu)$.

Theorem 5.3.1. *Let T be a rigid circular transformation. Let $\bar{\Phi}_0$ be a point process on W_s such that its distribution $\mathcal{L}_{\bar{\Phi}_0}$ is invariant to the action of T and let K be a descriptor invariant to the action of T. Then, for all $n \in \mathbb{N}$, $\mathcal{L}_{\bar{\Phi}_n}$ defined as the push-forward of $\mathcal{L}_{\bar{\Phi}_0}$ by* (5.9) *is invariant to the action of T. This result also holds true for the distributions of our model, defined by section 5.3.1.*

The proof is given in Section 5.8, for a more general model.

The result itself is inspired from [32], where the preservation of invariance is proven for the gradient descent model in the pixel domain. Observe, the invariance of the distribution of the point process $\bar{\Phi}_n$ increases the diversity of the generative model samples. Any descriptor taking the form of eq. (5.6), using the periodic shift operator $\bar{S}_u$, is invariant to all circular translations.

A drawback is that such a boundary condition introduces a statistical bias to the spatial average (5.6) as an estimator of $\mathbb{E}[K(\bar{\Phi})]$ in the case of a non periodic Φ over W_s. One can expect, however, that when the window size is large enough and spatial correlations of the patterns are not too large, this border effect becomes negligible.

5.4 Wavelet phase harmonic descriptors

Classical descriptors for spatial point process usually include statistics more or less directly related to the pair correlation function, such as Ripley's K-function, Besag's L-function, or the radial distribution function (see [37, Section 4.5]). All of these functions only capture second order correlations of the process. Other usual functions are the empty space function or the k-nearest neighbors function; (see Section 2.9, and [37, Section 2.3.4 and 4.1.7]). In [110], the authors advocate the use of the k-nearest neighbors distribution function, with a k significantly greater than 1. In addition to being non-differentiable, these moments suffer from another drawback. If one wants to capture geometric structures formed by the particles, up to a fixed scale, the number of moments (i.e. the k nearest neighbours) will grow linearly with the number of particles forming such structures. This can become a problem if the intensity of the process is large, both computationally, and from a statistical point of view, as the variance of the moments may become large when estimated from a single observation.

For this reason, we choose to use descriptors for which the spatial range of structure captured is independent of the intensity of the process, and the computational time is linear in the number of points. As a result, this method would become much faster for large samples, as the number of statistics would remain constant. In this work, we shall use the wavelet phase harmonics (WPH) covariance moments (cf. eq. (3.32)) to represent the distributions of our point processes. Recall from Section 3.5, these moments are based on the wavelet transform of a point process (eq. (3.5)), and in this study we shall used periodic wavelets for the observation window W_s (cf. eq. (3.33)). The WPH covariance moments use a non-linear operator to adjust the phase of the wavelet coefficients, in order to capture dependencies between coefficients at different scales and orientations.

More precisely, to capture edge-like structures in a planar point process, we shall use bump steerable wavelets. Recall from Example 3.1.3, they are defined by the translations, dilations and rotations of a complex analytic function $\psi(u) \in \mathbb{C}$ with compact frequency support. The finite wavelet family is defined as $\{\psi_\xi\}_{\xi \in \Xi}$, for $\xi = (j, \theta)$, with $0 < j < J$, and J such that all wavelets have (approximate) support contained in the observation window W_s. Scales equal or larger than J are carried by a low-pass filter ψ_0 whose frequency support is centered at 0. The angles are chosen such that $\theta = 2\pi \frac{\ell}{L}$, with $\ell \in \{0, \cdots, L-1\}$. For any $\xi, \xi' \in \Xi$, $k, k' \in \mathbb{Z}$, $\tau \in W_s$, we note $\gamma = (\xi, \xi', k, k', \tau)$, and estimate

$$K_\gamma \bar{\Phi} = \mathrm{Cov}\left([\bar{\Phi} \star \psi_\xi]^k(u), [\bar{\Phi} \star \psi_{\xi'}]^{k'}(u - \tau) \right), \tag{5.10}$$

which does not depend on $u \in W_s$, as the process is (circular) stationary. Recall that the phase harmonics operator $[\cdot]^k$ is defined, for all $z \in \mathbb{C}$, as

$$[z]^k = |z| e^{ik\varphi(z)}.$$

As explained in Section 3.5, the overlap between the frequency support of $[\bar{\Phi} \star \psi_\xi]^k$ and that of $[\bar{\Phi} \star \psi_{\xi'}]^{k'}$ is necessary for the wavelet phase harmonic covariance to be large. Due to the frequency transposition property of the wavelet phase harmonics, it is empirically verified that the covariance at $k\xi \simeq k'\xi'$ is often non-negligible when the process is non-Gaussian (i.e. has structures beyond second order correlations). We shall follow this empirical rule to define a set of WPH coefficients to correlate.

Let $v_{\xi,k} = \mathbb{E}([\bar{\Phi} \star \psi_\xi(u)]^k)$. We define the descriptors $K(\mu)$ using (5.6), as empirical estimators of moments. Additionally, let us denote $\mu_{\xi,k}(u) := [\mu \star \psi_\xi(u)]^k - v_{\xi,k}$. Taking the spatial average (5.6) gives the descriptor of the form:

$$K(\mu) = \left(\frac{1}{|W_s|} \int_{W_s} \mu_{\xi,k}(u)\, \mu_{\xi',k'}(u - \tau')^* du \right)_{(\xi,k,\xi',k',\tau') \in \Upsilon}. \tag{5.11}$$

As $\bar{\Phi}$ is circular-stationary, it suffices to choose $\tau' \neq 0$ to measure spatial shifts (i.e., there is no need to consider $u - \tau$, for some non null τ). Note also that $K(\mu)$ is invariant with respect to any circular translation $\bar{S}_u$ of $\mu \in \mathbb{M}^s$ on $u \in W_s$.

In the numerical computation, we shall replace $v_{\xi,k}$ in (5.11) by $\bar{v}_{\xi,k} = \frac{1}{|W_s|} \int_{W_s} [\bar{\phi} \star \psi_\xi(u)]^k du$ as a plug-in estimator for the first-order moment $v_{\xi,k}$. The $K(\bar{\phi})$ modified in this way becomes an empirical estimator of the covariances in eq. (5.10). This is a good approximation of $K(\mu)$ as the estimation variance of the covariance moments is typically much larger than that of the first-order moments.

Note that the wavelet phase harmonics do not increase the amplitude of the wavelet coefficients with $k > 1$. This approach may thus significantly reduce the variance of the descriptor K (to satisfy (P1)) compared to the higher order correlations, while still capturing information beyond second-order correlations (to satisfy (P2)).

Choice of the covariance set Υ in (5.11) Rather than detailing the full list of elements in Υ, we provide an intuitive way to choose the set Υ. For the full list, see [29, Section 4]. Overall, the total number of elements in Υ is in the order of $O(L^2 J^2)$. Note that the smallest structures that the descriptors can capture depend on the spatial support C of the wavelet ψ. Information about structures smaller than C can be added in a post-processing step that will be explained in Section 5.

- Choice of J: The covariance set Υ depends on the parameter J, which is the maximal scale of the wavelet transform. A suitable choice for this parameter J would be one allowing for a good trade-off between satisfying the sufficiency of K, while maintaining the concentration property (cf. properties (P1) and (P2) from Section 2.10).
- The parameter τ' is chosen so that each wavelet is translated in a particular direction in order to capture correlations along nearby edges in the observation.
- Choice of (ξ, ξ', k, k'): These parameters are chosen in order to capture 2nd-order correlations, as well as dependencies between wavelet coefficients at different scales and orientations, both with and without phase information, based on a rule of thumb that $k\xi \approx k'\xi'$ due to the frequency transposition property of the wavelet phase harmonics.

5.5 Numerical scheme for particle gradient descent

Calculating the wavelet phase harmonic covariances can be computationally demanding (due to the calculation of two integrals). In order to gain some efficiency, we can perform the computations in a discrete domain. However, the energy needs to remain differentiable with respect to the positions of the points in the pattern. We propose a method, consisting of a Gaussian smoothing of the configuration of points, to address this problem. Building on that method, we then present two technical aspects of the sampling method.

In this section, we discuss a complete numerical scheme to generate samples from the particle gradient descent model, defined with the wavelet phase harmonic descriptors presented in Section 5.4. It is composed of the following ideas:

- *Discretization* for an approximate calculation of the covariance of the wavelet phase harmonics: necessary to accelerate the calculation of the descriptor and the gradients.
- *Multiscale optimization*: allowing one to avoid shallow local minima in the gradient descent model. At each scale, we use a quasi-Newton gradient-descent method for greater efficiency.
- Final *blurring* (optional): to add a priori information on structures whose size is smaller than C into the model samples. It helps to get rid of some clusterisation (clumping) artifact caused by the initial discretization.

5.5.1 Discretization

Differentiable discretization of atomic measures

To compute the descriptor K in (5.11) for a point measure μ, we need to integrate functions over the observation window W_s (first for the convolution operators, then for the averages). Computationally efficient integration requires discretization of the atomic measure. The main difficulty is to do it in such a way that the (periodic) convolutions of the discretized atomic measures with wavelets, as in (6.2), remain differentiable with respect to the positions of the original atoms in μ, so that we can still perform gradient descent. Classical finite element methods may not achieve this goal efficiently.

We are going to approximate our atomic measures on W_s by matrices (images) of given size $N \times N$ (the image resolution), and then use the automatic differentiation software Pytorch [89] to perform the following operations. It allows one to compute the derivative of a modified energy w.r.t. any point u_i in μ. The following paragraph details this discretization:

We first map a given point measure μ on W_s to a continuous function μ_σ by the convolution

$$\mu_\sigma(u) := \mu \star g_\sigma(u) = \sum_{u_i \in \mu} e^{-\frac{|u-u_i|^2}{2\sigma}}, \quad x \in W_s \tag{5.12}$$

with a (periodized) Gaussian function g_σ of given standard deviation σ. Then we evaluate μ_σ on the $N \times N$ regular grid inside W_s and denote the resulting matrix μ_σ^N, with entries called (values of) *pixels*. The convolution with a Gaussian function makes each entry of μ_σ^N smoothly depend on the atom positions of μ. We then compute $\bar{K}(\mu_\sigma^N)$ instead of $K(\mu)$, where $\bar{K}$ is this discrete analogy of the descriptor (5.11) (cf. [120][1]). Note that, because the value of a pixel continuously depends on the positions of the atoms, this discretization makes our descriptor only invariant to discrete translations (multiple of the pixel size $2\frac{s}{N}$), for which Theorem 5.3.1 applies. The gradient of the energy $|\bar{K}(\mu_\sigma^N) - \bar{K}(\bar{\phi}_\sigma^N)|^2$ with respect to each atom position of μ can therefore be computed using automatic differentiation (with the Pytorch software [89]). Indeed, we know that $\bar{K}$ is differentiable w.r.t. each entry of μ_σ^N, as a combination of linear and non-linear operators. Moreover, for any $i, j \in \{1, N\}^2$, noting $\tilde{i} = -s + 2si/N$, $\tilde{j} = -s + 2sj/N$, (5.12) gives us that

$$\mu_\sigma^N(i,j) = \mu \star g_\sigma(\tilde{i},\tilde{j}) = \sum_{u_i \in \mu} e^{-\frac{|(\tilde{i},\tilde{j})-u_i|^2}{2\sigma}}, \tag{5.13}$$

which is differentiable w.r.t. any u_i in μ. This discretization step is illustrated in the Figure 5.3.

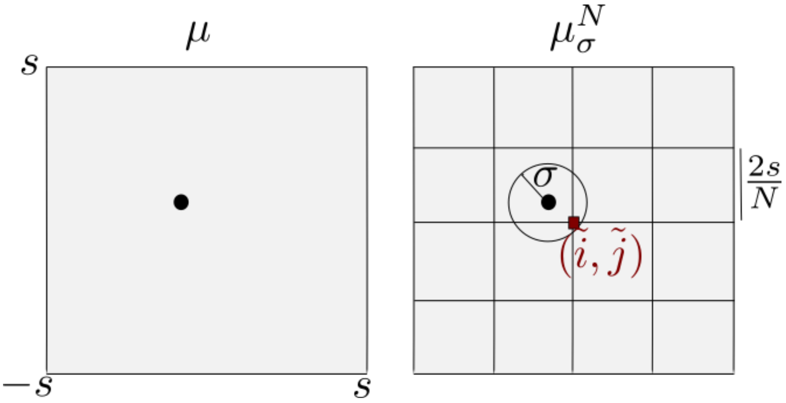


Figure 5.3: The differentiable discretization of a courting measure μ into a point-image μ_σ^N on a $N \times N$ regular grid.

In signal processing, the Gaussian function acts as a low-pass filter. It is needed to cut-off high frequency information of μ so that μ_σ can be discretized into an image with negligible aliasing effect. This means that μ_σ carries the information on the positions

---

[1] Computations are produced using a software available at `https://github.com/kymatio/phaseharmonics`

of μ up to some precision which depends on σ. The subsequent evaluation of μ_σ on the grid $N \times N$ in μ_σ^N implies that σ cannot be taken too small. Indeed, we take $\sigma_{\min} = \frac{s}{N}$ as the lowest value of σ.

Wavelet discretization and choice of scales

As stated in Section 5.4, the family of wavelets used in our descriptor is constructed by dilating the mother wavelet ψ in the range of the scales $0 \leq j < J$. Based on the choice of N, we set $C = \frac{2s}{N}$. In this way, the spatial support of ψ has a radius C of one pixel of the image. As a consequence, this smallest-scale wavelet ψ can also be discretized (without significant aliasing) in order to compute the discretized descriptor $\bar{K}$.

The choice of the largest scale J can be decided based on the visual structures in the observation. For example, if we want to model structures whose spatial size is close to the size of the window $[0, 1/8]^2 \subset W_s$, we shall set $2^J C = 1/8$, i.e. $J = \log_2(N) - 3 - \log_2(2s)$.

5.5.2 Multiscale optimization

Phase harmonic covariance moments of point process images (i.e. point patterns converted into regular pixel grids, as described above) may have large values at high frequencies (large values of $|\xi|, |\xi'|$ in (5.11)), due to the fact that the point-images are composed of local spikes when σ is small. This implies that these high frequency statistics have an important impact on the gradient of $\bar{K}$, which in turn can lead to the gradient descent model being trapped at shallow local minima, where only the high frequencies are well optimized to match the observation.

This optimization issue can be overcome by matching the descriptors from low frequency to high frequency in a sequential order, through an appropriate modulation of the parameter $\sigma \in \sigma_0, \sigma_1, ..., \sigma_{J-1}$ of the Gaussian functions used to discretize μ, introduced in Section 5.5.1.

Indeed, since Gaussian functions are low-pass filters, we can interpret the convolution in (5.12) as a blurring, limiting the space localization of Dirac measures. When such smoothing of the point pattern is done by a Gaussian function that has a large σ, the high frequencies of the signal function are close to 0 and the same holds true for the phase harmonics, because wavelets are localized in frequency. Therefore the wavelet phase harmonics are dominated by the low frequencies. Thus, by smoothing the observed sample and generating the optimal one with high variance Gaussian function, we create a new objective leading, in the gradient descent optimization, to a point configuration for which only low frequencies moments (small values of $|\xi|, |\xi'|$ in (5.11)) are matched with the ones of our observed sample. Thus, we propose a *multiscale* gradient descent procedure that consists in choosing first a high value for precision parameter σ, run the optimization algorithm, and then reduce the value of σ to run the optimization again, starting from the result of the previous run (and repeat this operation until $\sigma = \sigma_{J-1}$). We choose $\sigma_j := \frac{s}{N} 2^{J-j-2}$. Note that σ_{J-1} is equal to $\sigma_{\min}$. For numerical efficiency, we perform the gradient descent procedure using the L-BFGS optimization algorithm [76].

Empirical evidence in Section 5.6.2 shows that the above multi-scale optimization procedure allows one to reconstruct (modulo translation) the observed sample when using $\bar{K}$ defined with $J = \log_2(N) - 2$, which is not the case when simultaneously optimizing all frequencies. In order to preserve the ability to reproduce geometric structures at all scales, we shall also apply this multiscale optimization method to our model defined with

$J = \log_2(N) - 3$.

5.5.3 Final blurring

We observed that the contrast between the continuous nature of our objects and the discrete approximation described in Section 5.5.1 creates undesired artificial structures at frequencies higher than the image resolution: when the number of particles in a configuration is large with respect to the number of pixels in the image, or if the configuration exhibits strong clustering behaviour, several pixels may contain more than one particle. In such cases, our algorithm produces samples having an artificial clustering structure *inside* each of these pixels (see [29, Section 5] for an illustration of this phenomenon).

To remove this artificial clustering, we chose to force these high frequencies to be "as random as possible", i.e. to have Poisson-like structure. To this end, we introduce a uniform i.i.d. perturbation of the positions of points after the last optimization run. It can be viewed as an additional, this time stochastic, measure transport, following the deterministic one from the particle gradient descent. This final randomization can be viewed as enforcing a-priori information on high frequency structures of the process: Poisson-like structure.

5.6 Numerical experiments

In this section we present numerical experiments involving our generative model. We begin by presenting in Section 5.6.1 our numerical settings, in particular the distributions of point processes whose samples are used as original point patterns. We next evaluate how well our generative model with the phase harmonic covariance descriptor can generate samples similar to those given by the original point processes. In Section 5.6.2, we evaluate these models by comparing samples from the original distributions to samples from our models, visually as well as by estimating their power spectrum (defined as the average over angles of the squared modulus of the Discrete Fourier Transform, cf. Section 2.8). The power spectrum gives information equivalent to the second order correlation function of the process (cf. [26]), which captures clustering or repulsive behaviour between atoms of a realization. Such information cannot always be detected visually. In order to further quantify how well our model captures visual geometric structures, and to gain some insight into the ability of our model to produce diverse samples, we shall use the topological data analysis (TDA), derived from the theory of persistent homology. The comparison will be done in Section 5.6.3.

5.6.1 Numerical settings

We first describe the original point processes on which we shall evaluate the particle gradient-descent model, then specify the parameters of the model in the numerical experiments.

Original point process distributions

For our experiments, we choose point process distributions that show complex geometric structures, for which we can visually recognize geometric structures. We begin by presenting results for *Cox (double-stochastic Poisson) processes* with Poisson points living on one dimensional structures generated by two famous stochastic geometric models, namely edges of the Voronoi tessellation, and the Boolean model with circular grains of fixed radius considered in Example 2.8.4. Both underlying geometric models are generated by a Poisson parent process within the observation window W_s, and we construct these models in a periodic way to avoid border effects. We call the respective Cox processes *Voronoi* and *Circle* processes. Note that, for these two processes, Poisson points live on different geometric shapes: polygons for the Voronoi and possibly overlapping circles for other one. Additionally, we consider two different radii of circles.

Then, we take interest in distributions having *turbulent intensity* (derived from the simulations of a decaying isotropic turbulent vorticity field driven by 2d Navier-Stokes equations, see e.g. [99]). Such fields exhibit complex mulsticale structures, and are representations of physical phenomena, known to be difficult to model faithfully. Furthermore, the distributions we consider have much greater intensities that the previous Cox models. From the turbulent intensity, we sample three different processes, exhibiting distinct microscopic structures (repulsive, independent or clustering): a Matern *cluster* process, a Poisson point process and a Matern II *hard-core* process, see [37, Example 5.5 and Section 5.4, respectively]. We study the ability of our model to reproduce simultaneously the macroscopic (i.e. the turbulent intensity) and microscopic structures (i.e. at small scales) of the process.

The number of points in the Cox Voronoi, Small circles, Big circles, and the Turbulent Hardcore, Poisson, ad Cluster processes are around, respectively, 1 900, 2 500, 2 000, 1 700, 3 800 and 13 000. Note that, for comparison, point patterns considered in [110] have around 400 points.

Image resolution, number of iterations and computation time

As discussed in Section 5.5.1, point configurations are convoluted with Gaussian densities and evaluated on $N \times N$ grid (as images) in order to efficiently compute our descriptors, and move the particles with gradient descent. For simplicity, we fix $s = 1/2$ for all the examples that we shall consider. The ultimate Gaussian variance (precision) of this mapping is thus $\sigma_{\min} = \frac{1}{2N}$. The larger N is, the more information we are able to keep (in high frequencies), but the larger the computation time. We chose for our experiments a resolution of $N = 128$. We show one example where a higher resolution, $N = 256$, is used to capture most of the high frequency information. The number of iterations of the L-BFGS optimization is chosen to be 100 for each scale σ_j (a total of 400 iterations for $N = 128$, and 500 iterations for $N = 256$). The number of angles in the steerable wavelets is $L = 8$. The average computation time on 4 GPU (Nvidia Tesla P100) for a sample for a turbulent process having roughly 13 000 points with resolution $N = 256$ is between 5 and 10 minutes while the same task at the resolution $N = 128$ takes between 1 and 2 minutes.

5.6.2 Visual evaluation and spectrum comparison

We evaluate the ability of our model to capture and reproduce geometric structures exhibited by realizations of the point processes described in Section 5.6.1. A natural first method to assess the sufficiency of a generative model (property (P2)) is visual evaluation, which is widely used in image analysis but subjective. We then compare

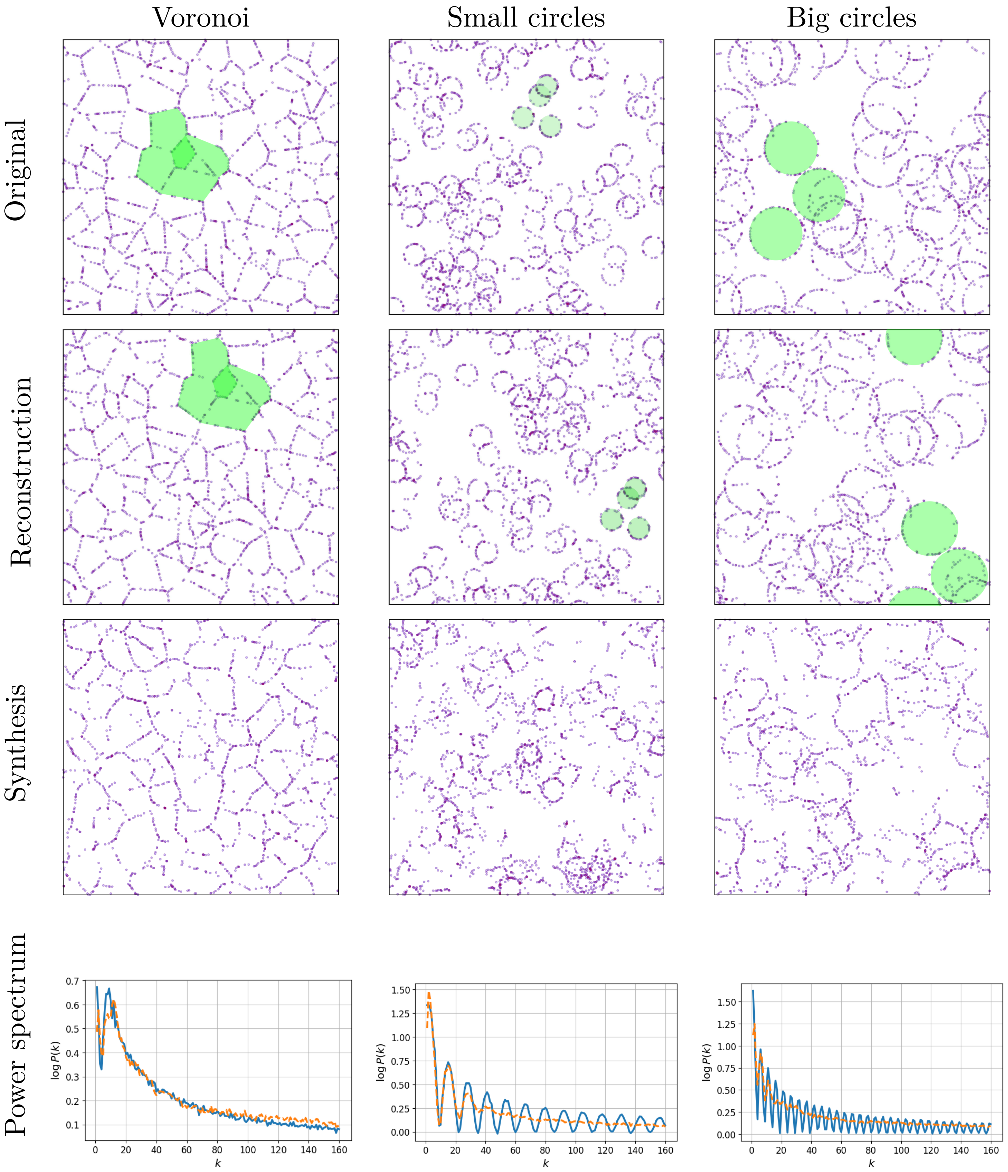


Figure 5.4: Three Cox models; original sample, reconstruction, and synthesis. For the power spectrum plots, the full lines correspond to the original distributions, and the dashed lines correspond to the models. The y-axis is presented in log scale.

the power spectra of our models and the original distributions. To estimate the power spectra, we generate (for each original distribution) 10 i.i.d. samples from the same model (i.e. from the same observation sample $\bar{\phi}$, but with different initial configurations $\bar{\phi}_0$). We average the power spectra of the 10 syntheses, and compare it to the average of 10 i.i.d. samples from the original distribution. All these samples will also serve in Section 5.6.3 to compare their geometric similarities.

Figure 5.4 shows a study of our three Cox distributions. The first line presents samples from the original distributions. The second line presents samples from the model using our descriptor with $J = \log(N) - 2$. In this setup, the concentration property is not satisfied (see property (P1) in Section 5.2), and the result is the memorization of

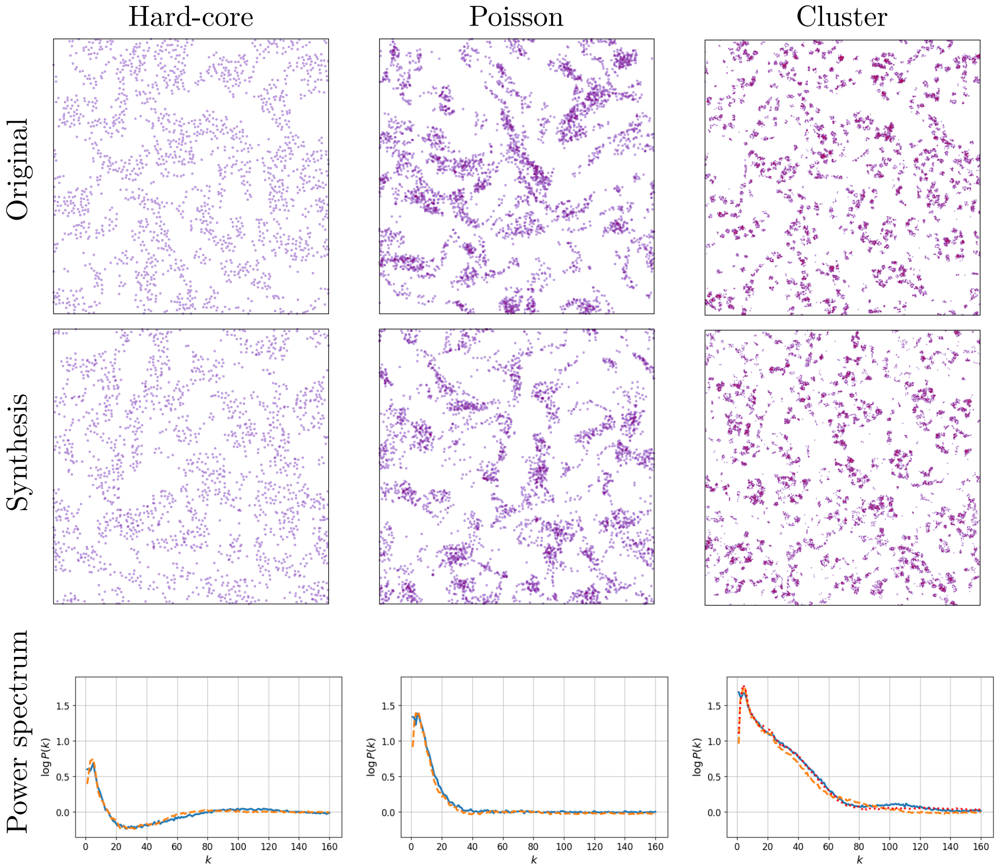


Figure 5.5: Synthesis of turbulence processes with various microscopic structures (Hardcore, Poisson, and Cluster), and their power spectrum plots. Full lines correspond to the original processes, dashed lines correspond to the model. For the Cluster distribution, the dotted line corresponds to our model defined with the resolution $N = 256$.

the observation sample $\bar{\phi}$. Indeed, this line shows quite faithful reconstructions of the original samples subjected to a periodic translation, up to some precision error due to a finite image resolution N. This is however not a good model because it essentially only contains the observation $\bar{\phi}$. It suggests that we need to improve the concentration property (P1) of the descriptors in order to enlarge the ensemble Ω_ϵ. (Note that, in the work of [110], the authors use the term 'reconstruction' to refer to random sampling method, which we call in this work 'synthesis'.)

In order to improve (P1), we shall reduce the parameter J in the wavelet transform. The third line of Figure 5.4 shows realizations sampled using $J = \log(N) - 3$ for different original distributions. Our analysis in Section 5.5.1 suggests that this range of J can model structures whose spatial size is at most 1/8 of the window W_s. Observe that most polygons and circles are well reproduced in the synthesis of Voronoi and Small circles. The Big Circles are harder to model since the size of each circle is slightly larger than 1/8 of W_s.

In the last line of Figure 5.4, we present the power spectrum (cf. Equation (2.25)) for $k \in \mathbb{N} \cap [1, 128[$ from the original distributions and as well as from our model. Larger errors can be observed at k near zero (say k=1,2 and 3). This is because only the average spectral information is captured (and matched) using the low-pass filter ψ_0 in the wavelet transform, which is included in the descriptor $K(\mu)$ (c.f. (5.11)). Moreover, the variance of the empirical information at small k can create extra error since it can be far away from its expectation. Similarly, because the wavelet convolutions average the spectral information over different frequency bands when using a reduced number of τ' in $K(\mu)$, our descriptor does not capture fast oscillations in the power spectra in the range of $k \leq N/2 = 64$ (see [120] for more details about how to capture these oscillations). This is observed in the cases of Small and Big Circles Cox processes. See example 2.8.4 for a theoretical formula of the power spectrum in the case of (Small or Big) circles. When $k > N/2 = 64$, the descriptor $\bar{K}$ (cf. 5.5.1) does not contain accurate spectral information due to a finite image resolution N. In this regime, we observe a smooth decay of the (log) power spectrum towards 0. We observed that if we apply the final blurring (cf. Section 5.5.3), then the error of the model spectrum becomes larger. Therefore, for these three processes, no final blurring has been applied.

All models discussed up to now are Cox processes, with Poisson (hence independent) points sitting on some random macroscopic structures. Figure 5.5 presents our analysis of three turbulent point processes having different microscopic structures: a hardcore, a non-correlated (Poisson) and a clustering one. We see that our generated samples capture to some extent this microscopic structure. For the clustering model, the presented synthesis is done with $N = 256$. We see that our model (using $J = \log(N) - 3$) can generate samples with similar macroscopic and microscopic structure. The power spectrum at small k has larger errors, as we have observed in the Cox models. However, since the power spectra are mostly smooth in these Turbulent processes, we observe a relatively small spectrum error over a wide range of k. This is also due to the use of the final blurring which helps to remove some artificial spectrum errors for $k > N/2$. For the clustering model, we also compare the power spectrum of two models with different resolutions: $N = 128$ and $N = 256$. We see that setting a higher resolution reduces significantly the error, allowing to match the spectrum up to $k \simeq 80$. we still observe some small error when $k \geq 80$, probably because of the final blurring (cf. Section 5.5.3), which may also impact the high frequencies that we optimize. Overall, both the visual and the spectral analysis suggest that our model can generate well various Turbulent points processes.

5.6.3 Persistent homology and topology analysis

As previously mentioned, power spectrum evaluation corresponds to the comparison of second order moments, which only partially capture geometric structures. Visual evaluation can be more discriminate, but is subjective. To evaluate more precisely the ability of our model to capture the geometric structures of the given distributions, we shall use a representation of objects derived from persistent homology theory, which is a powerful algebraic tool for studying the topological structure of shapes, functions, or in our case point clouds. We shall perform this evaluation by comparing the *persistence diagrams* of the generated samples to those of the original ones. Furthermore, this representation allows us to evaluate in a simple way the ability of our model to produce diverse samples.

We begin by a brief, intuitive presentation of persistence diagrams, and the whole comparison method that will be simply referred to as topology data analysis (TDA). For more details we refer the reader to [23, Section 11.5]. We then present the TDA of our point process distributions and models. TDA can be seen as a complementary tool with respect to the spectrum analysis, being more consistent with visual perception (see [29, Appendix C] for more details about this link).

Persistence diagram Persistent homology theory describes a way to encode the topological structure of a point cloud through a representation called *persistent diagram* (PD). It is constructed, for a given point configuration $\phi \in \mathbb{M}^s$, from the family $(G_r)_{r\geq 0}$ of Gilbert graphs, where the vertices are the positions of atoms of ϕ, and the edges are pairs of points closer to each other than r. (In our case we use the periodic metric.) Then, we fill-in the triangles (triplets of points joined by edges) of the graph. Points, edges and filled-in triangles constitute the so-called 2-skeleton of the Vietoris-Rips (VR) complex. For any $r \geq 0$, we study two characteristics of the skeleton: its *connected components*, and its *holes* (this latter notion is well formalized in the algebraic topology, in our case they correspond to the natural idea of a hole). Each connected component "is born" at time (radius) $r = 0$ and it "dies" at some time $r > 0$ when it is merged with another connected component. Similarly, each hole has a birth time ($r > 0$) corresponding to the minimal radius at which it appears, and a (larger) death time corresponding to the minimal radius for which the hole is completely filled-in by the triangles. The persistence diagram of ϕ is the collection of pairs of birth and death times of the connected components and holes. It is hence a point process in the positive orthant of the plane, offering a multiscale (as our wavelet-base descriptor) description of the topology of ϕ. As our descriptor, it is also stable to small deformations of ϕ. It is hence interesting to use this alternative tool to evaluate our generative model.

Topological data analysis Our approach in this matter is inspired by [36], and we refer the reader to this paper for a more detailed description. We use the 'holes' birth-death process, as it appears more relevant to capture information in the Cox distributions, such as the polygons and the circles.

In order to compare the distributions of our models to the original distributions, we compute the PDs of our samples from each distribution (cf. Section 5.6.2 for a description of these samples). Recall, these PDs can be viewed again as point clouds in two dimensions. Therefore, a distance between two PDs can be computed, and we use in this regard a periodic version of the Wasserstein distance between two point clouds on the plane (we found that the bottleneck distance also suggested in [36] is not sufficiently discriminating for our point patterns). We obtain in this way a distance matrix between different PDs (reflecting topological similarities or differences of the point processees realizations for which PDs were calculated). We then apply a standard dimension reduction algorithm (namely Multi Dimensional Scaling) to this distance matrix, to represent every PD (and hence the corresponding sample) as one point on the plane, and we visualize the representation of all samples.

TDA of our experiments In the plots on the first line of Figure 5.6, we study separately the Cox Voronoi and Big circles processes, and the turbulent hardcore process. In each plot, we observe 20 *dots* (having different shapes), each representing one configuration of points in W_s (the term "dot" is used to avoid confusion with points in W_s). For each model there are 10 dots representing i.i.d. realizations of the original distribution and 10 representing realizations from the generative model. For each plot, the sample additionally marked with a black dot represents the observation used in our model to

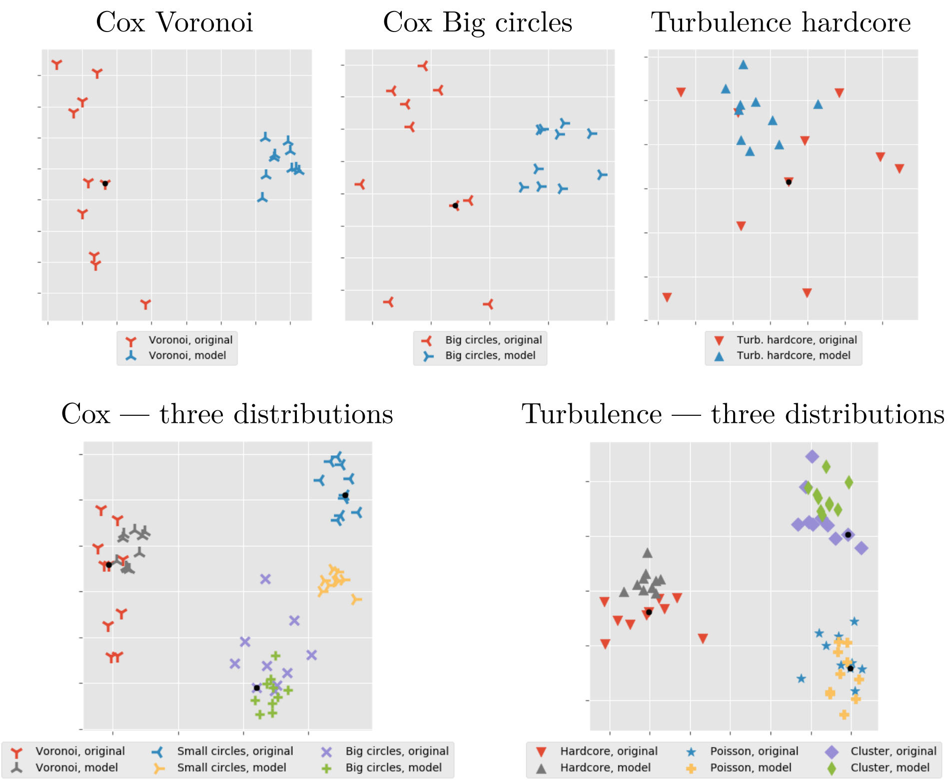


Figure 5.6: TDA of the distributions presented in this study, and their respective models.

produce the 10 syntheses.

We see in the first two Cox examples a clear separation between the original process and the model, implying a lack of sufficiency (P2) in our model. This is probably because, in order to satisfy (P1) and produce diverse samples, we have chosen to reduce J, and therefore lose some information about large scale structures of the process. On the other hand, we observe that this separation is smaller for the Turbulence hard-core case, where there is a better balance between (P1) and (P2). The error in the Cox models is probably due to the difficulty to reproduce highly constrained structures (perfect circles or convex polygons). These observations agree with our visual evaluation of the syntheses: the Voronoi and (more particularly so) the Big circles models are easily discriminated from their original distributions (their highly constrained structures are not perfectly reproduced in the syntheses), but this discrimination is harder for the Turbulent hardcore case. Moreover, these figures indicate, by the spread of the dots representing the syntheses, that our model reproduces, to some extent, the diversity in the samples of the original distributions (suggesting a certain entropy in our model). To further reduce the distances between the model samples and the original samples of a process, while maintaining a similar diversity (and hence a similar entropy between the model distribution and the original distribution) remains an interesting problem for future works.

The two plots in the second line present the TDA of the three Cox distributions together, and the three turbulent distributions together. We observe that, for the Cox distributions, the Small circles model is about as close to the original Small circles distribution as it is to the original Big circles distribution. Nevertheless, the original distribution of the Big circles and the Small circles are well separated, suggesting that there is a topological distinction that is not well respected in the small circles. However, this kind of error is hard to perceive visually. On the other hand, the Voronoi case is well separated from the other two, suggesting that the model is better than the ones of the circles distributions. For the turbulence case, we observe that the three distributions (both original and model) are well separated, and each respective model is closer to its original distribution than to the other distributions. This agrees with our earlier visual and spectral analysis, suggesting that our model is able to capture complex geometric structures formed by a large amount of points.

Remark: We used the R packages `TDAstats` [115] to calculate the PDs of our point patterns and `TDA` [52] to calculate their Wasserstein distances. Due to memory constraints, for second line, the analysis was done using a random thinning to reduce the number of points of each sample to 2 000, which could artificially impact the results. The experiments were repeated several times and the variability in the random thinning did not impact our conclusions.

5.7 Numerical comparisons

In this section, in order to illustrate the advantages of the method presented here, we present a brief comparison between the method presented in [110] and ours.

5.7.1 Differences between the two methods

The method developed in [110] and ours are based on the same idea: to produce similar but different point patterns to a given observation, one first defines what should be 'similar' between the observation and the synthesis, by choosing a set of statistical constraints, computed on the observation. Then, starting from an initial random configuration of points, one iteratively modifies this configuration in order to match the set of prescribed statistics (by minimizing an energy E, related to the square difference between the statistics of the original and the synthesised point patterns). If the set of statistics does not describe the observation itself, but rather its underlying distribution, then the output of the optimization procedure should be a new point pattern, similar but different to the observation.

However, the two methods differ on two points:

- First, the optimization method to match the set of statistics. In [110], the optimization steps can be described as follows: given the point configuration being synthesized $\phi_k = \sum_{i=1}^{N} \delta_{u_{i,k}}$ at some step k, a point in the configuration is chosen uniformly at random, say $u_{j,k}$, for $j \in \{1, \cdots, N\}$. A candidate for a new point $v \in W$ is chosen uniformly at random in the observation window. Then, if the energy of $\tilde{\phi}_k := \phi_k - \delta_{u_{j,k}} + \delta_v$ is lower than the energy of ϕ_k, we define $\phi_{k+1} = \tilde{\phi}_k$. Otherwise, $\phi_{k+1} = \phi_k$. We call this optimization *random search* (RS).

  In our method, $\phi_{k+1} = \sum_{i=1}^{N} \delta_{u_{i,k} - \zeta \nabla_{u_{i,k}} E(\phi_k)}$, where $\nabla_{u_{i,k}} E(\phi_k)$ is the gradient of the energy with respect to the point $u_{i,k}$ of ϕ_k, see (5.7). This optimization method will be noted (GD).

- Second, the set of statistical constraints used to describe the geometry of the point patterns. In [110], the authors use the k nearest neighbour distance distribution functions (d.f.) $D_k(r)$, for $k \in \{1, \cdots, k_{max}\}$, $k_{max} \geq 1$, and evaluated at a sequence of radii $r \in \{r_0, \cdots, r_{max}\}$, $r_{max} > 0$, (see [106, p. 267]). We call this statistical descriptor *nearest neighbour distances* (NND).

  Our statistics are based on the covariance between phase harmonics of the wavelet phase harmonics coefficients (WPH) of the point patterns (see 5.4).

5.7.2 Preliminary discussion

Before presenting a numerical comparison between the two methods, we briefly explain what the limitations of the method in [110] are, and why our method could overcome such limitations.

The optimization method in [110] is based on random search. This implies that for some configuration ϕ_k at some step k, there may be a lot of failing new candidates to replace some point in ϕ_k before finding one that reduces the energy of ϕ_k. This means that there may be a lot of energy evaluations before updating the current configuration. Furthermore, each iteration, requiring one energy evaluation, only moves one point in the configuration. Conversely, at each iteration, our algorithm computes the energy as well as the gradient of the energy, and all the points are moves according to the gradient. This implies that, for some energy level $e > 0$, the gradient descent method may reach this level in less iterations than [110].

Moreover, the statistical constraints used in [110] are based on 3 parameters: the number of neighbours for the points in the configuration, the maximal radius at which to evaluate whether or not there is a neighbour, and the number of radii between 0 and this maximal radius. While this latter relates to the precision of the d.f.'s, and can be pretty much fixed for any observation, this is not the case for the first two parameters. Indeed, k_{max} and r_{max} have to be chosen so as to describe the geometry formed by the points, up to some scale. The maximal radius r_{max} can be seen as the maximal scale up to which the statistical constraints describe the geometric structure. This can be fixed depending on the observation, but should not be too large, in order to satisfy the ergodic averaging property. However, for a fixed sequence of radii, the parameter k_{max} can change significantly depending on the observation. Even if two configuration exhibit structures up to similar scales, the number of nearest neighbours inside some ball may differ depending on the intensity of the process. For instance, consider the Cox Circles distribution, where points are located on circles of fixed radius r_0, with the center of those circles forming a Poisson point process. If the observed pattern has around 10 points per circle, one would probably need to fix $r_{max} = 2r_0$, and $k_{max} = 10$. However, if the circles contain an average of 100 points, then one would have to increase k_{max} up to 100, even if the circles have the same size as before. This would increase significantly the number statistics to compute at every step. Conversely, our descriptors only depend on a number of scales at which we compute the wavelet coefficients, which does not depend on the intensity of the process. The only parameters to fix are the maximal scale J, and the minimal scale, set by the resolution N, which relates to the precision set by the number of radii in the k nearest neighbours case. In the above Cox Circles example, if the points have Poisson distribution on the circles, then the size of the descriptor will not change between the two setups (10 or 100 points per circles).

5.7.3 Numerical comparison

In this section we present a numerical comparison between the method of [110] and ours. This comparison aims at illustrating the two following points:

1. For the same energy, the gradient descent optimization method reaches low energy levels in less time (i.e. less evaluations of energy value) than the optimization method used in [110]. To highlight this point, we shall consider the Cox Voronoi example, and use the WPH descriptors (c.f. (5.11)) to define the energy, and compare the RS and GD optimization methods.
2. The amount of information captured by the kth nearest neighbours d.f.'s depends on the intensity of the process, regardless of the scales of the structures. Considering the Turbulent Poisson example with different intensities, we shall see that, for a fixed descriptor (i.e. fixed k_{max}, r_{max}, number of r), the quality of the syntheses decreases with the intensity of the process.

These two points are illustrated in the two following sections, respectively. A last section will present an overall comparison of the two methods, on the Cox Voronoi and Turbulence Poisson examples.

Comparison between RS and GD (Cox Voronoi example)

For this experiment, we study the Cox Voronoi example, and define the energy from the wavelet phase harmonics covariances, presented in Section 5.4. Let K be our descriptor (defined in (5.11)), $\bar{\phi}$ our observation sample, and $E_{\bar{\phi}}(\cdot) = \frac{1}{2}|K(\cdot) - K(\bar{\phi})|^2|$ the corresponding energy. We define the relative energy by

$$e(\cdot) = 2\frac{E_{\bar{\Phi}}(\cdot)}{|K(\bar{\phi})|^2} = \frac{|K(\cdot) - K(\bar{\phi})|^2}{|K(\bar{\phi})|^2}. \tag{5.14}$$

We ran the optimization of the energy with the random search method from [110], and observe the relative energy of the syntheses (for 10 syntheses), after $n = 19870$ and $n = 29805$ iterations, i.e. respectively 10 and 15 iterations per point. After $n = 19870$ iterations, the algorithm reaches a relative energy of $e = 9,00.10^{-4}$ (with a std of $9,23.10^{-5}$), and after $n = 29805$ iterations, we found $e = 4,76.10^{-4}$ (std$= 2,47.10^{-5}$). It took an average of 1h04min and 1h36min respectively. We observed the respective relative energies, and ran our gradient descent optimization algorithm (without the multi-scale procedure) until the relative energy reaches the levels from the random search method. The results and comparisons with our method are summarized in Table 5.1.

Dependence of WPH and NND on the intensity of points (Turbulence Poisson example)

To illustrate our second point, for NND descriptor we fix the parameters of the kth nearest neighbours d.f.'s to $k_{max} = 16$, $r_{max} = .125$ (on a window of size 1), and discretize $D_k(r)$, $r \in (0, r_{\text{max}}]$, regularly by 250 values of radii r. With this fixed descriptor, we perform a synthesis using RS optimization for three different observations: the Turbulence Poisson observation randomly thinned to have 500 points, the same observation thinned

	Random search	Gradient descent
$e = 9,00.10^{-4}$	19870 (1h04m)	52 (0m35s)
$e = 4,76.10^{-4}$	29805 (1h36m)	69 (0m45s)

Table 5.1: Speed comparison between random search and gradient descent, in number of iterations (computation time in parenthesis) for the synthesis of Poisson Voronoi patterns. The time per iteration in the gradient descent method is larger, due to the possible several energy (and gradient) evaluations for the line search. However, the total amount of time is much lower. The computations have been run on a single GPU Nvidia Tesla P100.

to 2000 points, and the raw observation, which contains 3784 points. We set the number of iterations to 400 iterations per point. Figure 5.7 shows examples of syntheses for the 3 different patterns, as well as syntheses from our method using WPH descriptor with GD multiscale optimization. We observe that the method using the RS+NND fails to reproduce the geometric structures in the example with the largest number of points. Figure 5.7 also presents the estimations of the spherical contact distribution function (SCDF), defined for a point process Ξ as $H_s(r) := 1 - \mathbb{P}(\Xi \cap B(0,r) = \emptyset)$, where $B(0,r)$ denotes the ball of radius r, centered at 0. For the observation with the largest number of points, the model from [110] presents a larger error than ours.

Direct comparison

On the two distributions considered above, we shall illustrate the overall performance of both methods. We ran 10 simulations of syntheses from the method of [110], with $k_{max} = 64$ for the Cox Voronoi example, and $k_{max} = 128$ for the Turbulence Poisson example, with 400 iterations per point. For both models we perform also the syntheses from our multiscale gradient descent method (with 400 iterations).

The two families of examples of syntheses share the same corresponding, original pattern of points coming from the original distributions. We also simulate other 9 patterns for the original distributions to compare the averaged statistics and the diversity among the original distribution and the models.

We compare the different distributions with visual evaluation (Figure 5.8), TDA (Figure 5.9), and by estimation of the SCDF (Figure 5.10). For the evaluation with TDA, in addition to the visualization of a 2-dimensional representation of the distance matrix between (the PD of) all original and synthesized point patterns (cf. Section 5.6.3), we also computed the average Wasserstein distance between all pairs of point patterns belonging to different distributions. In more details, for a given distribution (Cox Voronoi or Turbulence Poisson), let $M := M_{orig/GD+WPH}$ be the 10×10 distance matrix between the 10 realizations of the original distribution and the 10 realizations of our model. We compute $d_{orig/GD+WPH} = \frac{1}{100}\sum_{i,j} M_{i,j}$. Similarly, we compute $d_{orig/RS+NND}$ for the RS+NND model. We obtained, for the Cox Voronoi example, $d_{orig/GD+WPH} = 0.75$, and $d_{orig/RS+NND} = 1.52$, showing a significant advantage to our method. Our experiments on the Turbulence Poisson example gave $d_{orig/GD+WPH} = 0.61$, and $d_{orig/RS+NND} = 0.62$. These results are coherent with the visual evaluation, that indicates a better performance of our model, especially for the Cox Voronoi example. The better performance of our method for the Turbulence Poisson example is best illustrated in Figure 5.10, that shows a greater error for the RS+NND model than ours on the estimation of the spherical

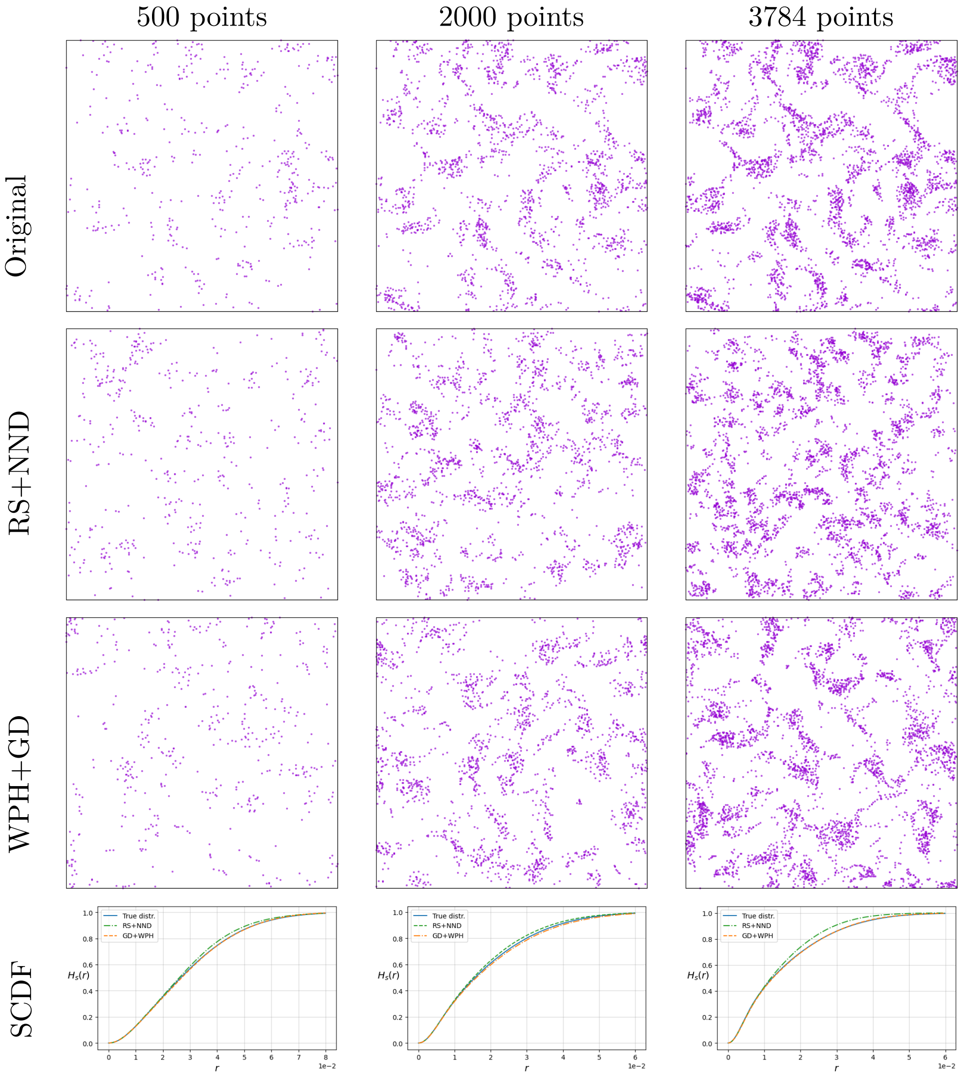


Figure 5.7: Syntheses from the model of [110] and our model, from observations containing 500, 2000 and 3784 points. The last line shows estimations of the spherical contact distribution function $H_s(r)$, for the corresponding models and methods, averaged over 10 realizations. For the three plots, the full line (blue) corresponds to the observation distribution, the (green) dashed line to the method (RS+NND), and the (orange) dash-dotted line to (GD+WPH).

contact distribution function.

5.8 Regularized model and proof of Theorem 5.3.1

Note that eq. (5.2) admits at least one global minimum, which is the observation $\bar{\phi}$ itself. By building the model defined in Section 5.3, one hopes to generate new configurations $\bar{\phi}'$ such that $E_{\bar{\phi}}(\bar{\phi}') \simeq 0$, and $\bar{\phi}' \neq \bar{\phi}$. If the descriptor K is well chosen, then local minima of $E_{\bar{\phi}}$ correspond to configurations that are likely to have been drawn from $\mathcal{L}_{\bar{\Phi}}$ (i.e. the model is a good approximation of the true distribution).

The landscape of $E_{\bar{\phi}}$ is defined by the descriptor K. One might find itself in the

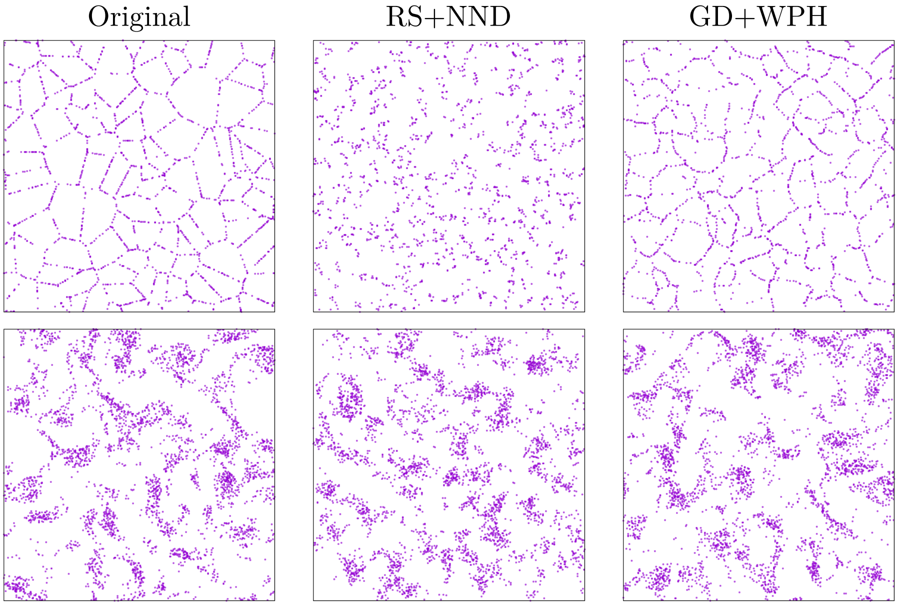
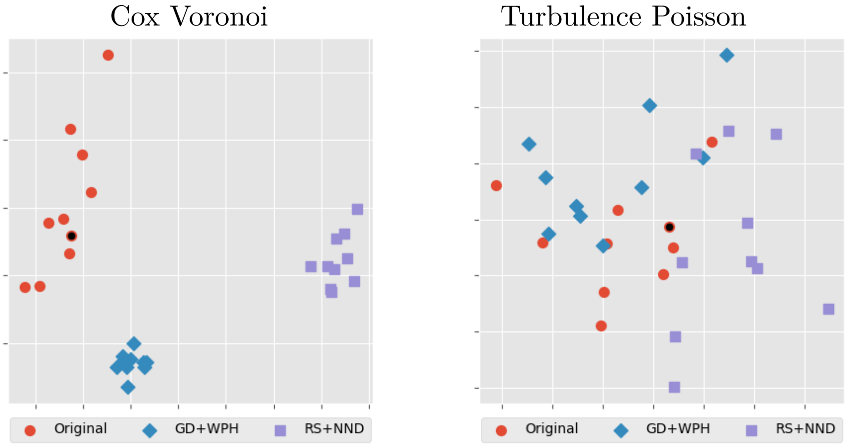


Figure 5.8: Visualization of syntheses for the Cox Voronoi and Turbulence Poisson examples. Left: observations, middle: RS+NND, right: GD+WPH.



$$d_{orig/RS+NND} = 1.52$$
$$d_{orig/GD+WPH} = 0.75$$

$$d_{orig/RS+NND} = 0.62$$
$$d_{orig/GD+WPH} = 0.61$$

Figure 5.9: Visualization of the TDA of the three distributions (Original, GD+WPH, RS+NND), for the Cox Voronoi example (left), and the Turbulence Poisson example (right). The black point represents the observation pattern used for the syntheses. The averaged (true) distances between the original and synthesized patterns (via Wasserstein distance of persistence diagrams) are given as well.

situation where a certain descriptor K gives poor local minima, (i.e. K does not satisfy the sufficiency property (P2), so samples from the model defined with K are not similar to $\bar{\phi}$), but incorporating additional constrains to K result in a model for which the only local minimum is $\bar{\phi}$ itself. This is the case, for instance, when adding new constraints to K leads to a violation of the concentration property (P1), due to the high variance of

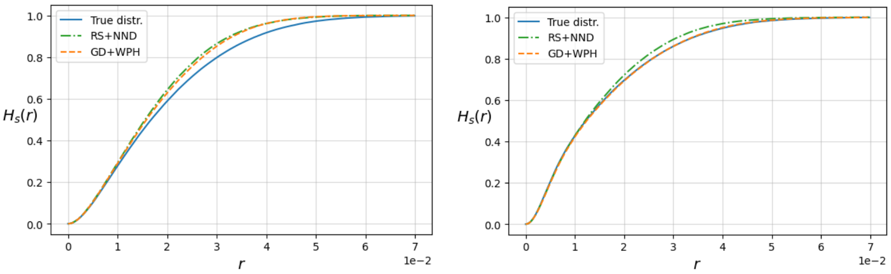


Figure 5.10: Estimation of the spherical contact distribution functions (SCDF), for the Cox Voronoi example (left), and the Turbulence Poisson example (right). For the two plots, the full (blue) line corresponds to the observation distribution, the dashed-dotted (green) line to the method (RS+NND), and the dash-dotted (orange) line to (WPH+GD).

the new descriptor, as it will be illustrated in Figure 5.4[2].

To address this issue, we propose an alternative model allowing for the use of a descriptor K that may not satisfy (P1). It consists in adding a regularization term to the energy $E_{\bar{\phi}}(\mu)$ in (5.2), so as to prevent the gradient descent algorithm converge to the global minimum which is $\bar{\phi}$. This can be realized, for example, using a Wassertein-type distance between the current and the initial configuration of points, as described in what follows.

The regularization that we introduce in this section aims at preventing the successive iterations of the gradient descent model considered in Section 5.3.1 to move too far away from its initial configuration. It is achieved by adding the distance between the initial configuration and optimized configuration as a penalization term in the objective. It thus forces the model to explore local minima of $E_{\bar{\phi}}(\cdot)$ around the initial configuration. As all the configurations are measures with equal mass, it is natural to consider the Wasserstein distance between them. However, as this distance is computationally expensive to compute, we choose to replace it with the sliced Wesserstein distance [24]. For counting measures with same mass, the 2-Wasserstein distance is defined as:

$$\mathcal{W}_2(\mu,\nu) = \min_{\sigma\in\mathfrak{S}_N} \left(\sum_{i=1}^{N} ||u_i - v_{\sigma(i)}||^2 \right)^{\frac{1}{2}},$$

where the minimum is taken over the set $\mathfrak{S}_N$ of permutations of $(1,\ldots,N)$.

The square $\mathcal{SW}_2$ distance between two planar counting measures μ and ν is then defined as follows:

$$\mathcal{SW}_2^2(\mu,\nu) := \int_{\theta\in\mathbb{S}^1} \mathcal{W}_2^2(\theta_{\#}\mu, \theta_{\#}\nu) d\theta,$$

where $\theta_{\#}$ designates the pushforward by the orthogonal projection operator in the direction θ.

The use of $\mathcal{SW}_2$ instead of $\mathcal{W}_2$ is motivated by the fact that applying the projection operators leads us to compute $\mathcal{W}_2$ on $\mathbb{R}$ instead of $\mathbb{R}^2$, which is much faster, as we simply need to sort the Dirac measures to obtain the optimal matching. It has been shown [25] that $\mathcal{SW}_2$ is indeed a distance, and that it induces the same topology as $\mathcal{W}_2$ for compact

---

[2] Note that the violation of (P1) does not imply memorizing $\bar{\phi}$.

domains. One can approximate $\mathcal{SW}_2$ by choosing a certain number of fixed directions [24].

We use this distance with respect to the initial configuration $\bar{\phi}_0 \in \mathbb{M}^s$ as a regularization term in the minimization of $E_{\bar{\phi}}(\cdot)$ defined in Eq. (5.2). More precisely, we consider the following optimization problem

$$\arg \min_{f \in \mathcal{F}(\mathbb{R}^2)} \frac{1}{2}|K(f_{\#}\bar{\phi}_0) - K(\bar{\phi})| + \eta \frac{1}{2}\mathcal{SW}_2^2(f_{\#}\bar{\phi}_0, \bar{\phi}_0), \tag{5.15}$$

where the minimization is done over the set $\mathcal{F}(\mathbb{R}^2)$ of measurable functions from $\mathbb{R}^2$ into itself, and $\eta \geq 0$ is a regularization parameter.

This new optimization problem implies that we do not try to sample from the microcanonical set here, as we suppose that it does not contain enough realizations of $\bar{\Phi}$.

Note, in (5.15) we optimize the transport $f_{\#}$ of a given initial configuration $\bar{\phi}_0$ towards $\bar{\phi}$, by minimizing the value of the function

$$\begin{aligned} E_{\bar{\phi}}^{\bar{\phi}_0}(f) &= E_{\bar{\phi}}(f_{\#}\phi_0) + \eta \frac{1}{2}\mathcal{SW}_2^2(f_{\#}\bar{\phi}_0, \bar{\phi}_0) \\ &= \frac{1}{2}|K(f_{\#}\bar{\phi}_0) - K(\bar{\phi})| + \eta \frac{1}{2}\mathcal{SW}_2^2(f_{\#}\bar{\phi}_0, \bar{\phi}_0). \end{aligned}$$

In a similar way as eq. (5.7), for any initial measure $\bar{\phi}_0$ and any measurable function f, we define a function from $\mathbb{R}^2$ to $\mathbb{R}^2$

$$x \longmapsto \nabla_u E_{\bar{\phi}}^{\bar{\phi}_0}(f) := \eta \nabla_u \mathcal{SW}_2(f_{\#}\bar{\phi}_0, \bar{\phi}_0) + \nabla_u E_{\bar{\phi}}(f_{\#}\bar{\phi}_0).$$

The sequence of point-configuration transports induced by gradient descent algorithm is defined recursively by taking an initial configuration $\bar{\phi}_0$, $f_0 := I_d$ (identity) and for $n \geq 0$

$$\begin{cases} f_{n+1} : x \longmapsto u - \zeta \nabla_u E_{\bar{\phi}}^{\bar{\phi}_0}(f_n \circ \ldots \circ f_0) \\ \bar{\phi}_{n+1} = f_{n+1\#}\bar{\phi}_n = (f_n \circ \ldots \circ f_0)_{\#}\bar{\phi}_0. \end{cases}, \tag{5.16}$$

where $\zeta > 0$ is some gradient step size.

In analogy to (5.8) and (5.9), for all $n \geq 1$, $\bar{\phi}_n$ transports the initial distribution $\mathcal{L}_0$ (of $\bar{\phi}_0$) to some $\mathcal{L}_n$. Note $\mathcal{L}_k = \mathcal{L}_{\bar{\Phi}_n}$ is the distribution of the point process $\bar{\Phi}_n$ obtained by n iterations of (5.16) started of $\bar{\Phi}_0$ having law $\mathcal{L}_0 = \mathcal{L}_{\bar{\Phi}_0}$.

Proof of Theorem 5.3.1

Proof. In order to prove Theorem 5.3.1, we need to formally define Eq. (5.7). Recall that in this section and in what follows, W_s is interpreted as endowed with the addition and scalar multiplication modulo W_s.

For $\mu \in \mathbb{M}^s$ and any $u \in \mathrm{Supp}(\mu)$[3], we define the following functions:

$$\begin{array}{rccc} h_u^\mu & : & \mathbb{R}^2 & \longrightarrow & \mathbb{M}^s \\ & & v & \longmapsto & \mu - \delta_u + \delta_{u+v}, \end{array}$$

$$\begin{array}{rccc} K_u^\mu & : & \mathbb{R}^2 & \longrightarrow & \mathbb{C}^d \equiv \mathbb{R}^{2d} \\ & & v & \longmapsto & K \circ h_u^\mu(v), \end{array}$$

---

[3] For any counting measure μ, $\mathrm{Supp}(\mu)$ is defined as the union of all singletons where μ has an atom, i.e. $\mathrm{Supp}(\mu) = x : \mu(x) > 0$.

$$\begin{array}{rcccc} E_u^{\mu} & : & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^+ \\ & & v & \longmapsto & E_{\bar{\phi}} \circ h_u^{\mu}(v). \end{array}$$

The function K_u^{μ} can be complex valued. However, as our energy function is the square Euclidean norm, it is equivalent to consider that K_u^{μ} has values in $\mathbb{R}^{2d}$. Moreover, we assume in what follows that the function K is such that for all $\mu \in \mathbb{M}^s$ and all $u \in \mathrm{Supp}(\mu)$, K_u^{μ} is differentiable. We can then define from chain rule, for any $\mu \in \mathbb{M}^s$ and any $u \in W_s$

$$\nabla_u K(\mu) := \begin{cases} Jac[K_u^{\mu}](0) & \text{if } u \in \mathrm{Supp}(\mu) \\ 0 & \text{otherwise} \end{cases} \tag{5.17}$$

and

$$\nabla_u E_{\bar{\phi}}(\mu) := \begin{cases} Jac[E_u^{\mu}](0) & \text{if } u \in \mathrm{Supp}(\mu) \\ 0 & \text{otherwise} \end{cases} \tag{5.18}$$

where $Jac[f]$ denotes the Jacobian matrix of the function f. When $u \in \mathrm{Supp}(\mu)$, the chain-rule gives $Jac[E_u^{\mu}](0) = (\nabla_u K(\mu))^t (K(\mu) - K(\bar{\phi}))$.

Additionally, for $\phi_0 \in \mathbb{M}$, $u \in \phi_0$, and any $f \in \mathcal{F}(\mathbb{R}^2)$, we define the functions

$$\begin{array}{rcccc} h_{f,u}^{\bar{\phi}_0} & : & \mathbb{R}^2 & \longrightarrow & \mathbb{M} \\ & & v & \longmapsto & f_{\#}\mu - \delta_{f(u)} + \delta_{f(u)+v}, \end{array}$$

and

$$\begin{array}{rcccc} \mathcal{SW}_2{}_{f,u}^{\bar{\phi}_0} & : & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^+ \\ & & v & \longmapsto & \mathcal{SW}_2(h_{f,u}^{\bar{\phi}_0}(v), \phi_0). \end{array}$$

For any $u \in \bar{\phi}_0$, and any $f \in \mathcal{F}(\mathbb{R}^2)$, $\mathcal{SW}_2{}_{f,u}^{\bar{\phi}_0}$ is differentiable, so we can define a function from $\mathbb{R}^2$ to $\mathbb{R}^2$

$$u \longmapsto \nabla_u \mathcal{SW}_2(f_{\#}\bar{\phi}_0, \bar{\phi}_0) := \begin{cases} J[\mathcal{SW}_2{}_{f,u}^{\bar{\phi}_0}](0) & \text{if } u \in \mathrm{Supp}(f_{\#}\bar{\phi}_0) \\ 0 & \text{otherwise.} \end{cases}$$

The result follows from the fact that, for each $n \geq 0$, the function $u \longmapsto f_n(u) = f_n(u, \bar{\phi}_0)$ defined in (5.16) satisfies

$$f_n(Tu; T_{\#}\bar{\phi}_0) = T f_n(u; \bar{\phi}_0), \quad \bar{\phi}_0 \in \mathbb{M}^s \tag{5.19}$$

and, by consequence, the same holds for the composition $F_n(u, \bar{\phi}_0) := f_n \circ \ldots \circ f_1(u, \bar{\phi}_0)$.

Indeed, if (5.19) holds, then for $\bar{\Phi}_n := F_{n\,\#}\bar{\Phi}_0 = F_n(\cdot, \bar{\Phi}_0)_{\#}\bar{\Phi}_0$,

$$\begin{aligned} T_{\#}\bar{\Phi}_n &= \sum_{u \in \bar{\Phi}_0} \delta_{TF_n(u, \bar{\Phi}_0)} \\ &= \sum_{u \in \bar{\Phi}_0} \delta_{F_n(Tu, T_{\#}\bar{\Phi}_0)} \\ &= \sum_{u \in T_{\#}\bar{\Phi}_0} \delta_{F_n(u, T_{\#}\bar{\Phi}_0})} \\ &= F_{n\#}(T_{\#}\bar{\Phi}_0). \end{aligned}$$

Then for any $\Gamma \in \mathcal{M}^s$, as $\bar{\Phi}_0$ is invariant to the action of T:

$$\begin{aligned}
\mathbb{P}(\bar{\Phi}_n \in T_{\#}^{-1}(\Gamma)) &= \mathbb{P}(T_{\#}\bar{\Phi}_n \in \Gamma) \\
&= \mathbb{P}(F_{n\#}(T_{\#}\bar{\Phi}_0) \in \Gamma) \\
&= \mathbb{P}(F_{n\#}\bar{\Phi}_0 \in \Gamma) \\
&= \mathbb{P}(\bar{\Phi}_n \in \Gamma),
\end{aligned}$$

where the second to last equality comes from the invariance of the distribution of $\bar{\Phi}_0$ to the action of T.

It remains to prove (5.19), which can be done by induction. It is trivially true for $n = 0$, as $f_0 = Id$. Assume now that (5.19) is satisfied for all f_k, $k = 1, \ldots, n$ and, consequently, for $F_n = f_n \circ \ldots \circ f_1$. Then, it remains to show that, for all $\bar{\phi}_0 \in \mathbb{M}^s$,

$$f_{n+1}(Tu; T_{\#}\bar{\phi}_0) = Tf_{n+1}(u; \bar{\phi}_0).$$

Let us fix, without loss of generality, $\zeta = \eta = 1$, then

$$f_{n+1}(u, \bar{\phi}_0) := u - \nabla_u \mathcal{SW}_2(F_n(\cdot, \bar{\phi}_0)_{\#}\bar{\phi}_0, \bar{\phi}_0) - \nabla_u E_{\bar{\phi}}(F_n(\cdot, \bar{\phi}_0)_{\#}\bar{\phi}_0). \tag{5.20}$$

If $u \notin F_n(\cdot, \bar{\phi}_0)_{\#}\bar{\phi}_0$, then the gradients are null, so the result is trivial.

If $u \in F_n(\cdot, \bar{\phi}_0)_{\#}\bar{\phi}_0$, then there exists $u_i \in \bar{\phi}_0$ such that $u = F_n(u_i, \bar{\phi}_0)$. We study the second and third term of the r.h.s. of eq. (5.20) separately. Let us note A the linear component of T, i.e. $Tu + Av = T(u+v)$, for $u, v \in \mathbb{R}^2$. In what follows, we shall simplify the notation replacing $F_n(\cdot, \bar{\phi}_0)_{\#}\bar{\phi}_0$ by F_n and $F_n(u_i, \bar{\phi}_0)$ by $F_n(u_i)$.

We can write, as K is invariant to T,

$$\begin{aligned}
\nabla_{TF_n(u_i)} E_{\bar{\phi}}(TF_n) &= \Big(\nabla_{TF_n(u_i)} K(TF_n)\Big)^t \Big(K(TF_n) - K(T_{\#}\bar{\phi})\Big) \\
&= \Big(\nabla_{TF_n(u_i)} K(TF_n)\Big)^t \Big(K(F_n) - K(\bar{\phi})\Big),
\end{aligned} \tag{5.21}$$

with

$$\nabla_{TF_n(u_i)} K(TF_n) = J[K^{TF_n}_{TF_n(u_i)}](0) = J[K \circ h^{TF_n}_{TF_n(u_i)} \circ A](0)A^t, \tag{5.22}$$

where we used the fact that $A^{-1} = A^t$. Furthermore, using the fact that K is invariant to T,

$$\begin{aligned}
&K \circ h^{TF_n}_{TF_n(u_i)} \circ A(v) \\
&= K\Big(TF_n - \delta_{TF_n(u_i)} + \delta_{TF_n(u_i)+Av}\Big) \\
&= K\Big(TF_n - \delta_{TF_n(u_i)} + \delta_{T(F_n(u_i)+v)}\Big) \\
&= K\Big(T_{\#}\Big(F_n - \delta_{F_n(u_i)} + \delta_{F_n(u_i)+v}\Big)\Big) \\
&= K\Big(F_n - \delta_{F_n(u_i)} + \delta_{F_n(u_i)+v}\Big) \\
&= K \circ h^{F_n}_{F_n(u_i)}.
\end{aligned} \tag{5.23}$$

Using (5.22) and (5.23) in (5.21) we then obtain

$$\begin{aligned}\nabla_{TF_n(u_i)} E_{\bar{\phi}}(TF_n) &= \Big(\nabla_{F_n(u_i)} K(F_n) A^t\Big)^t \Big(K(F_n) - K(\bar{\phi})\Big) \\ &= A\Big(\nabla_{F_n(u_i)} K(F_n)\Big)^t \Big(K(F_n) - K(\bar{\phi})\Big) \\ &= A\nabla_{F_n(u_i)} E_{\bar{\phi}}(F_n).\end{aligned}$$

The computations for the regularization term follow the same lines, as the gradient of the regularization function is equivariant to T and, the regularization itself is invariant. Finally, we obtain

$$\begin{aligned}f_{n+1}(Tu, T_\#\bar{\phi}_0) &= Tu - A\nabla_u \mathcal{SW}_2(F_n(\cdot,\bar{\phi}_0)_\#\bar{\phi}_0, \bar{\phi}_0) - A\nabla_u E_{\bar{\phi}}(F_n(\cdot,\bar{\phi}_0)_\#\bar{\phi}_0) \\ &= T(u - \nabla_u \mathcal{SW}_2(F_n(\cdot,\bar{\phi}_0)_\#\bar{\phi}_0, \bar{\phi}_0) - \nabla_u E_{\bar{\phi}}(F_n(\cdot,\bar{\phi}_0)_\#\bar{\phi}_0)) \\ &= Tf_{n+1}(u, \bar{\phi}_0),\end{aligned}$$

which concludes the proof. ☐

5.9 Relaxing the assumptions on the data

In order to present our model in a simple setting, strong theoretical assumptions have been made on the data. However, in real world applications, the data will most likely not satisfy these assumptions. This section presents ideas on how to adapt our model in such cases.

Non-periodic boundaries Recall that our descriptor, defined in (5.11), applies periodic boundary correction to point patterns in a square window. If the structure of the observed pattern is not periodic, one can modify the descriptor by applying non-periodic integrals in (5.11) over some smaller window. In particular, we suggest a *scale-dependent reduction of the integration window*, pertinent when the wavelet ψ has a compact (or approximately compact) spatial support. Specifically, we consider a new descriptor $\tilde{K}$ by considering the integrals in (5.11) with $i = (\xi, k, \xi', k', \tau') \in \Upsilon$ over smaller windows $W_{s_i} \subset W_s$, such that boundary effects are negligible. Our current software can also handle such non-periodic boundary conditions.

More general observation windows Additionally, we considered that the observed pattern lies in a square observation window. If this is not the case, one could use a similar idea to the non-periodic case: embed the observation window in a square window and considering integrals in (5.11) over the observation window.

Non stationary process In [69], the authors focus on building a model for non stationary point processes inspired by [110]. Similarly, one might adapt our method to model non stationary processes. This could be done by modifying two aspects of the method. First, the initial distribution Φ_0 (cf. Section 5.3.1) could be chosen as a non stationary Poisson point process, estimating the intensity with a kernel estimator, such as in [69]. In addition, as pointed out in [69], the descriptor should be adapted not to be translation invariant. This could be done, for example, by applying local integrals over patches of the observation window in (5.11) (this requires some notion of "local stationarity" of the process). Another method that may be useful in this scenario is the regularization proposed in Section 5.8, where a regularization term is added to the energy. This term consists of the (Sliced Wasserstein) distance [95] between the initial configuration and the current configuration (the one being optimized). By adding this regularization term to the energy, the points of the configuration are forced not to move too far away from the initial configuration, which could help preserve the non stationarity of the initial distribution in the distribution of the model.

5.10 Discussion

In this chapter, we presented a particle gradient descent model to simulate stationary and ergodic point processes, based on a single observation in a square window. This model is able to synthesize processes formed by a large number of points, exhibiting interactions at multiple scales. Our method is built upon recent works on gradient descent methods to approximate the micro-canonical model. To characterize complex geometric point patterns, we use the wavelet phase harmonic descriptors that allow to explicitly control the scales of the structures to model. Compared to the classical approaches developed in [109, 110], our approach brought a new perspective to the modeling of point processes, through the lens of wavelet analysis and image modeling.

While numerical results on Cox and Turbulent distributions illustrate the ability of the model to capture various geometric structures in the observation, these distributions are not perfectly well approximated by our model, as shown in Figure 5.6. In order to study what geometric information is missing from our wavelet covariance descriptor, we shall take interest in the modelling of texture images, as they can exhibit complex structures, often more easily perceptible.

Chapter 6

Generalized rectifier wavelet covariance models for texture synthesis

In Chapter 5, we used wavelet phase harmonics covariance statistics (cf. Section 3.5) to represent the distribution of point processes, and built a maximum entropy model based on this representation. We shall now study what possible improvements can be made to this set of descriptors, by focusing on texture images, which can be view as stationary random fields, and can exhibit more complex geometric structures than point processes. The model developed in this chapter can, however, be readily applied to random measures and stationary point processes.

6.1 Modelling texture images

Textures ares spatially homogeneous images, consisting of similar patterns forming a coherent ensemble. In texture modeling, one of the standard approaches to synthesize textures relies on defining a maximum entropy model [65] using a single observed image [94]. Recall from 2.10, maximum entropy models consist of computing a set of prescribed statistics from the observed texture image, and then generating synthetic textures producing the same statistics as the observation. If the statistics correctly describe the structures present in the observation, then any new image with the same statistics should appear similar to the observation. A major challenge of such methods resides in finding a suitable set of statistics, that can generate both high-quality and diverse synthetic samples. This problem is fundamental as it is at the heart of many texture related problems. For example, in patch re-arrangement methods for texture modeling, these statistics are used to compute high-level similarities of image patches [75, 94]. Such statistics are also used in texture interpolation for probing visual perception [114], style transfer and image inpainting [55, 72].

A key question along this line of research is to find *what it takes to generate natural textures.* This problem was originally posed in [66], in which the author looks for a statistical characterization of textures. In the classical work of [92], the authors presented a model whose statistics are built on the wavelet transform of an input texture image. These statistics were carefully chosen, by showing that each of them captured a specific aspect of the structure of the image. This model produces satisfying results for a wide range of textures, but fails to reproduce complex geometric structures present in some natural texture images. Figure 6.1 presents a typical example composed of radishes, and synthetic images from three state-of-the-art models developed over the last few decades. To address this problem, the work of [54] proposes to use statistics built on the correlations between the feature maps of a deep convolutional neural network (CNN), pre-trained on the ImageNet classification problem [43, 102]. While this model produces visually appealing images, these statistics are hard to interpret. The work of [112] made a significant simplification of such statistics, by using the feature maps of a one-layer rectifier CNN with random filters (without learning). A crucial aspect of this simplification relies on using multi-scale filters, which are naturally connected to the wavelet transform. In this chapter, we propose a wavelet-based model, more interpretable than CNN-based models (with learned or random filters), to synthesize textures with complex geometric structures. It allows one to bridge the gap between the classical work of [92], and state-of-the-art models.

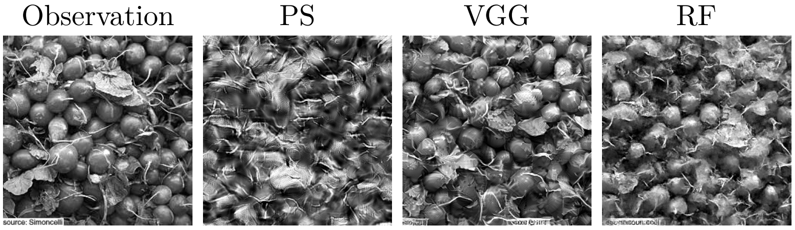


Figure 6.1: Example of syntheses from three texture models in chronological order. From left to right: the observed texture in gray-scale, synthesis from PS ([92]), from VGG ([54]), and from RF ([112]).

This model is built on the recent development of the phase harmonics for image representations and non-Gaussian stationary process modeling, introduced in [81, 120], and presented in Section 3.5 in the stochastic-geometric framework. Recall that the phase harmonics are non-linear transformations that adjust the phase of a complex number. In [92, 120], the authors illustrate that the phase dependencies between wavelet coefficients across scales contain important information about the geometric structures in textures and turbulent flows, and that they can be captured by applying the phase harmonics to complex wavelet coefficients. Remarkably, [81] show that the phase harmonics admit a dual representation, closely related to the rectifier non-linearity in CNNs.

In this work, we present the following results:

- We develop a family of texture models based on the wavelet transform and a generalized rectifier non-linearity, that significantly improves the visual quality of the classical wavelet-based model of [92] on a wide range of textures. It relies on introducing spatial shift statistics across scales to capture geometric structures in textures.
- By changing the number of statistics in our models, we show explicitly the trade-off on the quality and diversity of the synthesis. When there are too many statistics, our model tends to memorize image patches. We further investigate such memorization effects on non-stationary images and find that it sometimes relies on what statistics are chosen, rather than on how many.

- Through the modeling of geometric structures in gray-scale textures, our model indicates the possibility of reducing significantly the number of statistics in the works of [54] and [112], to achieve a similar visual quality.

The rest of this chapter is organized as follows: Section 6.2 reviews the framework of microcanonical maximum-entropy models, build upon a general family of covariance statistics. We then present our model for both gray-scale and color textures in Section 6.3. Section 6.4 shows synthesis results of our model, compared with state-of-the-art models. Finally, in Section 6.5, we discuss possible improvements of our model.

Notations For the remainder of this chapter, N denotes a positive integer. A gray-scale image x is an element of $\mathbb{R}^{N\times N}$, i.e. $x = x(u)$, $u \in \mathcal{X}_N$, with $\mathcal{X}_N := \{0, \cdots, N-1\}^2$. Recall, a color image $x = \{x^c\}_{c=1,2,3}$ is an element of $\mathbb{R}^{3\times N\times N}$, or equivalently, each $x^c \in \mathbb{R}^{N\times N}$. We shall denote $\bar{x}$ the observed texture (observation), which is assumed to be a realisation a random vector X. For any complex number $z \in \mathcal{C}$, z^* is the complex conjugate of z, Real(z) its real part, $|z|$ its modulus, and $\varphi(z)$ its phase. In this work, for the sake of clarity, the model proposed in [92] shall be abbreviated PS, the model presented in [112] shall be noted RF (as they use random filters), and the VGG-based model of [54] shall be referred to as VGG.

6.2 Microcanonical covariance models

We briefly review the standard framework of micro-canonical maximum-entropy models for textures. To reliably estimate the statistics in these models, we assume that a texture is a realization of a stationary and ergodic process (restricted to $\mathcal{X}_N$). We then review a special family of statistics that are used in the state-of-the-art texture models (mentioned in Figure 6.1), based on covariance statistics of an image representation.

6.2.1 Framework

Given a observation texture $\bar{x}$, we aim at generating new texture images, similar but different from $\bar{x}$. To that end, a classical method is to define a set of statistics $C\bar{x}$, computed on the observation, and try to sample from the *microcanonical set*

$$\{x : \|Cx - C\bar{x}\| \leq \epsilon\},$$

where $\|\cdot\|$ denotes the L^2 norm.

Under the stationary and ergodic assumption of X, one can construct Cx as a statistical estimator of $\mathbb{E}(CX)$, from a complex-valued representation $Rx(\gamma, u) \in \mathbb{C}$ as a function of (γ, u) in an index set $\Gamma \times \mathcal{X}_N$. The set of covariance statistics Cx of a model can then be constructed by computing an averaging over the spatial variable u, i.e.

$$Cx(\gamma, \gamma', \tau) := \frac{1}{|\mathcal{X}_N|} \sum_{u \in \mathcal{X}_N} Rx(\gamma, u) Rx(\gamma', u - \tau)^*, \tag{6.1}$$

for $(\gamma, \gamma', \tau) \in \Upsilon \subseteq \Gamma \times \Gamma \times \mathcal{X}_N$. The statistics $Cx(\gamma, \gamma', \tau)$ can be interpreted as estimating the covariance (resp. correlations) between $RX(\gamma, u)$ and $RX(\gamma', u - \tau)$ for zero-mean RX (resp. non-zero mean RX). The ergodicity assumption ensures that when N is large enough, the approximation $C\bar{x} \simeq \mathbb{E}(CX)$ over Υ should hold with high probability. Under these conditions, it makes sense to sample the microcanonical set in order to generate new texture samples. To draw the samples, we follow gradient-based sampling algorithms, suitable in high-dimensions [32], to minimize the objective function $\|Cx - C\bar{x}\|^2$, starting from an initial sample from a normal distribution. This sampling method is similar to the one presented in Section 5.3 to sample from point processes (recall that the method proposed in Section 5.3 use a differentiable mapping from point patterns to images, cf. Section 5.5.1).

This framework encompasses a wide range of state-of-the-art texture models, such as [92, 54, 112, 120]. In particular, the PS model developed in [92] takes inspiration from the human early visual system to define a multi-scale representation based on the *wavelet transform* of the image [61]. We next review a family of covariance model which generalizes the statistics in the PS model. We write C^{M} the statistics for a specific model M that uses the representation R^{M}.

6.2.2 Wavelet phase harmonic covariance models

We review a family of microcanonical covariance models defined by a representation built upon the wavelet transform and phase harmonics (Sections 3.1 and 3.5). The wavelet phase harmonics covariance representation is defined for texture images in the same way as for point processes. It defines a class of covariance statistics that capture dependencies between wavelet coefficients across scales.

Wavelet transform

For texture modeling, we consider oriented wavelets to model geometric structures in images at multiple scales. They include the Morlet wavelets (cf. Example 3.1.2) and steerable wavelets (such as Example 3.1.3), proposed in [58, 101, 111]. In particular, the Simoncelli steerable wavelets have been used to model a diverse variety of textures in [92].

Oriented wavelets are defined by the dilation and rotation of a complex function $\psi : \mathbb{R}^2 \mapsto \mathbb{C}$ on a plane. Let r_θ denote the rotation by angle θ in $\mathbb{R}^2$. They are derived from ψ with dilations by factors 2^j, for $j \in \{0, 1, \cdots, J-1\}$, and rotations r_θ over angles $\theta = \ell\pi/L$ for $0 \leq \ell < L$, where L is the number of angles in $[0, \pi)$. The wavelet at scale j and angle θ is defined by

$$\psi_{j,\theta}(u) = 2^{-2j}\psi(2^{-j} r_\theta u), \quad u \in \mathbb{R}^2$$

Scales equal or larger than J are carried by a low-pass filter ϕ_J. These continuous wavelets are then discretized (with periodic boundary condition) to define a discrete wavelet transform in the spatial domain $\mathcal{X}_N$.

The wavelet transform of an image $x \in \mathbb{R}^{N \times N}$ is a family of functions obtained by the convolution of x with the discrete wavelets. Let $\Lambda := \{0, \cdots, J-1\} \times \frac{\pi}{L}\{0, \cdots, L-1\}$ be the index set of the wavelet coefficients. For notation convenience, we denote the resulting coefficients by

$$x \star \psi_{j,\theta}(u) = \sum_{v \in \mathcal{X}_N} x(u-v)\psi_{j,\theta}(v), \quad u \in \mathcal{X}_N, \quad (j,\theta) \in \Lambda. \tag{6.2}$$

The low-pass coefficients $x \star \phi_J$ are defined similarly.

Wavelet phase harmonics and the PS model

To model natural textures, it has been shown ([92, 120]) that it is crucial to capture statistical dependencies between wavelet coefficients across scales. This can be achieved by using a wavelet phase harmonic representation, which is defined by the composition of a linear wavelet transform of x, and a non-linear phase harmonic transform.

In [81], the authors introduce the phase harmonics to adjust the phase of a complex number $z \in \mathbb{C}$ (cf. Section 3.5). Recall that the phase harmonics $\{[z]^k\}_{k\in\mathbb{Z}}$ of a complex number $z \in \mathbb{C}$ are defined by multiplying its phase $\varphi(z)$ of z by integers k, while keeping the modulus constant, i.e. $\forall\, k \in \mathbb{Z}$, $[z]^k := |z|e^{ik\varphi(z)}$..

The wavelet phase harmonic representation (WPH) is then defined by

$$R^{\mathrm{WPH}}x(\gamma, u) = [x \star \psi_{j,\theta}(u)]^k - \mu_\gamma, \quad \gamma = (j, \theta, k) \in \Gamma = \Lambda \times \mathbb{Z}, \tag{6.3}$$

where μ_γ is defined as the spatial average of $[\bar{x} \star \psi_{j,\theta}]^k$.

It is shown in [120] that the PS model can be regarded as a low-order wavelet phase harmonics covariance model, which considers only a restricted number of pairs (k, k'). In the next section, we shall use a dual representation of the phase harmonic operator to define a covariance model to capture high-order phase harmonics.

6.3 Generalized rectifier wavelet covariance model

In the previous section, we presented a class of models, built from the wavelet phase harmonic representation. A dual representation of the phase harmonic operator $[\cdot]^k$ can be defined via a generalized rectified linear unit, that we review in Section 6.3.1. We then discuss in Section 6.3.2 how to define an appropriate index set of Γ for gray-scale textures. Section 6.3.3 extends the model to color textures.

6.3.1 From phase harmonics to the generalized rectifier

The generalized rectified linear unit of a complex number z, with a phase shifted by $\alpha \in [0, 2\pi]$, is defined by

$$\rho_\alpha(z) = \rho(\mathrm{Real}(e^{i\alpha} z)), \tag{6.4}$$

where ρ is a rectified linear unit, i.e. for any $t \in \mathbb{R}$, $\rho(t) := \max(0, t)$. In [81], it is shown that applying a Fourier transform on $\rho_\alpha(z)$ along the variable α results in the phase harmonics of z (up to some normalization constant). This suggests an alternative model, defined by coefficients of the form

$$R^{\mathrm{ALPHA}}x(\gamma, u) = \rho_\alpha(x \star \psi_{j,\theta}(u)) - \mu_\gamma, \quad \gamma = (j, \theta, \alpha), \tag{6.5}$$

for $\gamma \in \Gamma = \Lambda \times [0, 2\pi]$, and, as before, μ_γ is defined as the spatial average of $\rho_\alpha(\bar{x} \star \psi_{j,\theta}(u))$ over $u \in \mathcal{X}_N$.

Relation with high-order phase harmonics Based on the duality between the phase harmonics $k \in \mathbb{Z}$ and the phase shift variable $\alpha \in [0, 2\pi]$, we now present the relation between C^{ALPHA} and the high-order phase harmonics in C^{WPH}. The following result was first proved in [81],

Proposition 6.3.1. *There exists a complex-valued sequence* $\{c_k\}_{k\in\mathbb{Z}}$ *such that for all* $(j,\theta,\alpha)\in\Gamma$, $(j',\theta',\alpha')\in\Gamma$, *and all* $\tau\in\mathcal{X}_N$,

$$C^{\mathrm{ALPHA}}x((j,\theta,\alpha),(j',\theta',\alpha'),\tau)=\sum_{(k,k')\in\mathbb{Z}^2}c_kc_{k'}^*C^{\mathrm{WPH}}x((j,\theta,k),(j',\theta',k'),\tau)e^{i(k\alpha-k'\alpha')}.$$

Proof. Using the fact that

$$\begin{aligned}\rho(\mathrm{Real}(ze^{i\alpha}))&=\rho(\mathrm{Real}(|z|e^{i(\varphi(z)+\alpha)}))\\&=|z|\rho(\cos(\alpha+\varphi(z)),\end{aligned}$$

and taking the Fourier transform of $\rho_\alpha(z)$ in the variable α, we obtain

$$\begin{aligned}\mathcal{F}(\rho_\alpha(z))(k)&:=\frac{1}{2\pi}\int_{[0,2\pi]}\rho_\alpha(z)e^{-ik\alpha}d\alpha\\&=|z|\frac{1}{2\pi}\int_{[0,2\pi]}\rho(\cos(\alpha+\varphi(z))e^{-ik\alpha}d\alpha\\&=|z|e^{ik\varphi(z)}c_k\\&=[z]^kc_k,\end{aligned}$$

where c_k is the Fourier transform of $h(.):=\rho(\cos(.))$ at the frequency k. The function $\alpha\mapsto\rho_\alpha(z)$ being periodic in α, we have its decomposition in Fourier series

$$\begin{aligned}\rho_\alpha(z)&=\sum_{k\in\mathbb{Z}}\mathcal{F}(\rho_\alpha(z))(k)e^{ik\alpha}\\&=\sum_{k\in\mathbb{Z}}c_k[z]^ke^{ik\alpha}.\end{aligned}$$

We can then write, for any $z,z'\in\mathbb{C}$, and $\alpha,\alpha'\in[0,2\pi]$,

$$\rho_\alpha(z)\rho_{\alpha'}(z')^*=\sum_{k,k'\in\mathbb{Z}^2}c_kc_{k'}^*[z]^k[z']^{-k'}e^{i(k\alpha-k'\alpha')}.$$

Replacing z and z' by any two wavelet coefficients $x\star\psi_{j,\theta}(u)$ and $x\star\psi_{j',\theta'}(u-\tau)$, we thus obtain the relation in Proposition 6.3.1. ☐

We remark that the sequence $\{c_k\}_{k\in\mathbb{Z}}$ is uniquely determined by the rectifier non-linearity ρ, and they are non-zero if k is even ([81]). This result shows that for a suitable choice of (α,α'), the covariance statistics $C^{\mathrm{ALPHA}}x$ can implicitly capture $C^{\mathrm{WPH}}x$ with a wide range of k and k'.

Relation with second order statistics Using a simple decomposition of wavelet coefficients into their positive, negative, real and imaginary parts, we can further show that the covariance statistics $C^{\mathrm{ALPHA}}x$ capture the classical second order statistics of wavelet coefficients, also used in the PS model (with phase harmonic coefficients $k=k'=1$).

Proposition 6.3.2. *Let* $I=\{0,\frac{\pi}{2},\pi,\frac{3\pi}{2}\}$. *There exists a finite complex-valued sequence* $\{w_{\alpha,\alpha'}\}_{(\alpha,\alpha')\in I^2}$ *such that for all* $(j,\theta)\in\Lambda$, *and all* $\tau\in\mathcal{X}_N$,

$$\sum_{(\alpha,\alpha')\in I^2}w_{\alpha,\alpha'}C^{\mathrm{ALPHA}}x((j,\theta,\alpha),(j',\theta',\alpha'),\tau)=\frac{1}{\mathcal{X}_N}\sum_{u\in\mathcal{X}_N}\Big(x\star\psi_{j,\theta}(u)\Big)\Big(x\star\psi_{j',\theta'}(u-\tau)\Big)^*.\tag{6.6}$$

Proof. Let $z \in \mathbb{C}$, and recall from eq. (6.4), that $\rho_\alpha(z) = \rho(\mathrm{Real}(ze^{i\alpha}))$. Note that we have the following relation

$$z = \rho_0(z) - \rho_\pi(z) - i(\rho_{\frac{\pi}{2}}(z) - \rho_{\frac{3\pi}{2}}(z)). \tag{9}$$

We can then write

$$\begin{aligned} zz'^* &= \Big(\rho_0(z) - \rho_\pi(z) - i(\rho_{\frac{\pi}{2}}(z) - \rho_{\frac{3\pi}{2}}(z))\Big)\Big(\rho_0(z') - \rho_\pi(z') - i(\rho_{\frac{\pi}{2}}(z') - \rho_{\frac{3\pi}{2}}(z'))\Big) \\ &= \sum_{\alpha,\alpha' \in I^2} w_{\alpha,\alpha'} \rho_\alpha(z) \rho_{\alpha'}(z'), \end{aligned}$$

with $I = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. Replacing z with $x \star \psi_{j,\theta}(u)$, z' with $x \star \psi_{j',\theta'}(u - \tau)$, and injecting this relation in eq. (6.1) gives us the desired result. □

This shows that using only four α uniformly chosen between $[0, 2\pi]$ is sufficient to capture second order statistics. Because the wavelet transform is an invertible linear operator (on its range space), computing the r.h.s of eq. (6.6) for all (j, θ, τ), as well as the low-pass coefficients carried out by Φ_J, is equivalent to computing the correlation matrix of x.

Relation with the RF model Setting aside the subtraction by the spatial mean μ_γ, the RF model can be viewed as a particular case of models defined by eq. (6.5). Indeed, the statistics of the RF model take the form of eq. (6.1), with

$$R^{\mathrm{RF}} x(f, u) = \rho(x \star \psi_f(u)),$$

where $\{\psi_f\}$ being a family of multi-scale random filters (samples from a uniform distribution, see [112] for more details). By writing $\rho_\alpha(x \star \psi_{j,\theta}(u - \tau)) = \rho(x \star \mathrm{Real}(\psi^\tau_{j,\theta} e^{i\alpha})(u))$, with $\psi^\tau_{j,\theta}$ denoting the translation of $\psi_{j,\theta}$ by τ, we see that the models are similar, the difference being that our models use wavelet-based filters instead of random ones.

6.3.2 Defining an appropriate Υ

The choice of the covariance set Υ is of central importance in the definition of the model. Intuitively, a too small set of indices will induce a model that could miss important structural information about the texture that we want to synthesize. Conversely, if Υ contains too many indices, the syntheses can have good visual quality, but the statistics of the model may have a large variance, leading to the memorization of some patterns of the observation. There is a trade-off between these two aspects: one must capture enough information to get syntheses of good visual quality, but not much, so as not to reproduce parts of the original image. To illustrate this point, we shall study the model ALPHA defined with three different sets Υ : A smaller model $\mathrm{ALPHA_S}$ with a limited amount of elements in Υ, an intermediate model $\mathrm{ALPHA_I}$, and a larger model $\mathrm{ALPHA_L}$.

To precisely define these models, let us note $\mathfrak{J} := \{0, \cdots, J-1\}$, $\Theta := \frac{\pi}{L}\{0, \cdots, L-1\}$, and $\mathcal{A}_A = \frac{2\pi}{A}\{0, \cdots, A-1\}$. Let us also define the set $\mathfrak{T} := \{0\} \cup \{2^j(\cos(\theta), \sin(\theta)) : 0 \leq j < J,\ \theta \in \frac{\pi}{L}\{0, \cdots, 2L-1\}\}$, from which the spatial shift shall be selected. Table 6.1 summarizes the conditions that all parameters have to satisfy to be contained in these sets. Additionally, these models include large scale information through the covariance of a low-pass filter, i.e. the spatial average of $x \star \phi_J(\cdot) x \star \phi_J(\cdot - \tau)$, for $\tau \in \mathfrak{T}$. To count the size of Υ without redundancies, in Section 6.4.1 we provide an upper bound on the non-redundant statistics in our models. This upper bound is used to count the number of statistics in our models. To keep this number from being too large, instead of taking all shifts in a square box, such as in [92], we choose to select only shifts of dyadic moduli, and with the same orientations as the wavelets.

Model	Scales	Angles	Phase shift	Spatial shift	Size of Υ
ALPHA$_\mathrm{S}$	$(j,j')\in\mathfrak{J}^2$ $\lvert j'-j\rvert\leq 1$	$(\theta,\theta')\in\Theta^2$	$(\alpha,\alpha')\in\mathcal{A}_4\times\mathcal{A}_1$	$\tau\in\mathfrak{T}$ if $(j,\theta)=(j',\theta')$ $\tau=0$ otherwise.	$(\ J\lvert\Theta\rvert^2+J\lvert\Theta\rvert\lvert\mathfrak{T}\rvert)\lvert\mathcal{A}_4\rvert$
ALPHA$_\mathrm{I}$	$(j,j')\in\mathfrak{J}^2$	$(\theta,\theta')\in\Theta^2$	$(\alpha,\alpha')\in\mathcal{A}_4\times\mathcal{A}_1$	$\tau\in\mathfrak{T}$	$J^2\lvert\Theta\rvert^2\lvert\mathcal{A}_4\rvert\lvert\mathfrak{T}\rvert$
ALPHA$_\mathrm{L}$	$(j,j')\in\mathfrak{J}^2$	$(\theta,\theta')\in\Theta^2$	$(\alpha,\alpha')\in\mathcal{A}_4^2$	$\tau\in\mathfrak{T}$	$J^2\lvert\Theta\rvert^2\lvert\mathcal{A}_4\rvert^2\lvert\mathfrak{T}\rvert$

Table 6.1: List of indices in Υ for different ALPHA models.

ALPHA$_\mathrm{S}$ **vs.** ALPHA$_\mathrm{I}$ The small model ALPHA$_\mathrm{S}$ is inspired from the PS model, as it only takes into account of the interactions between nearby scales (i.e. $|j'-j|\leq 1$), and the spatial shift correlations are only considered for $(j,\theta)=(j',\theta')$. There are two notable differences in the statistics included in the ALPHA$_\mathrm{S}$ and ALPHA$_\mathrm{I}$ models. The first one is the range of scales being correlated. It has been shown in [120] that constraining correlation between a wider range of scales induces a better model for non-Gaussian stationary processes, and a better estimation of cosmological parameters from observed data [2]. The second difference, which has a significant impact on the number of statistics (it increases the model size by a factor $\sim$10), is the number of spatial shifts in the correlations. In the ALPHA$_\mathrm{I}$ model, spatially shifted correlations are computed for all pairs of coefficient (γ,γ'). For both stationary textures and non-stationary images in gray-scale, shape and contours of salient structures and objects are better reproduced with ALPHA$_\mathrm{I}$, as illustrated in Figure 6.2 and Figure 6.3.

ALPHA$_\mathrm{I}$ **vs.** ALPHA$_\mathrm{L}$ As we observe in Figure 6.2, the ALPHA$_\mathrm{I}$ model, containing 4 times less coefficients than the ALPHA$_\mathrm{L}$, suffers less from memorization effects, while still capturing most of the geometric information in the images. This small loss of information can be partially explained by the frequency transposition property of the phase harmonics operator [81], for compactly supported wavelets in the frequency domain, as detailed in the following.

First, remark that the simple linear relation

$$\Big(\sum_{\alpha\in\mathcal{A}_4}\rho_\alpha(z)e^{ik\alpha}\Big)\rho_0(z)=\sum_{\alpha\in\mathcal{A}_4}\rho_\alpha(z)\rho_0(z)e^{ik\alpha} \tag{6.7}$$

tells us that computing all correlations for $\alpha\in\mathcal{A}_4$ and $\alpha'=0$ gives us at least the information contained in the r.h.s. of eq. (6.7).

Furthermore, if $\alpha\in\mathcal{A}_4=\{0,\cdots,\frac{3\pi}{4}\}$, we make the following approximation

$$\sum_{\alpha\in\mathcal{A}_4}\rho_\alpha(z)e^{ik\alpha}\simeq\int_{[0,2\pi]}\rho_\alpha(z)e^{-ik\alpha}d\alpha=2\pi c_k[z]^k.$$

Recall also from the proof of Proposition 6.3.2, that $\mathcal{F}(\rho_\alpha(z))(k) = [z]^k c_k$, where the Fourier transform is taken along the variable α. Therefore,

$$\rho_0(z) = \sum_{k\in\mathbb{Z}} c_k [z]^k.$$

Therefore,

$$\begin{aligned}\Big(\sum_{\alpha\in\mathcal{A}_4} \rho_\alpha(z) e^{ik\alpha}\Big)\rho_0(z) &\simeq \Big(\int_{[0,2\pi]} \rho_\alpha(z) e^{-ik\alpha} d\alpha\Big)\rho_0(z) \\ &= 2\pi c_k [z]^k \Big(\sum_{k'\in\mathbb{Z}} c_{k'} [z']^{k'}\Big) \\ &= 2\pi c_k \sum_{k'\in\mathbb{Z}} c_{k'} [z]^k [z']^{k'}.\end{aligned}$$

Then, replacing z and z' with wavelet coefficients, we get

$$\sum_{\alpha\in\mathcal{A}_A} e^{-ik\alpha} C^{\mathrm{ALPHA}} x((j,\theta,\alpha),(j',\theta',0),\tau) \simeq 2\pi c_k \sum_{k'\in\mathbb{Z}} c_{k'} C^{\mathrm{WPH}} x((j,\theta,k),(j',\theta',k'),\tau).$$

Using Plancherel's theorem, we can write that

$$C^{\mathrm{WPH}} x((j,\theta,k),(j',\theta',k'),\tau) = \frac{1}{|\mathcal{X}_N|} \sum_{\omega\in\frac{2\pi}{N}\mathcal{X}_N} \mathcal{F}([x\star\psi_\xi]^k)(\omega)\mathcal{F}([x\star t_\tau\psi_\xi]^{-k'})(\omega),$$

where the Fourier transform of an image x is defined by $\mathcal{F}(x)(\omega) := \sum_{u\in\mathcal{X}_N} x(u) e^{-i\omega u}$, and t_τ denotes the translation by τ, i.e. $t_\tau f(\cdot) = f(\cdot - \tau)$.

Now, suppose that the wavelets ψ_α have disjoint compact frequency support, in balls $B_\xi(2^{-j}C')$, where $\xi = 2^{-j} r_{-\theta}\xi_0$, and ξ_0 is the central frequency of the mother wavelet ψ (cf. Section 3.1). Suppose also that frequency transposition property of the phase harmonics operator (cf. [81]) is such that $[x\star\psi_\xi]^k$ has (approximately) frequency support in $B_{k\xi}(k2^{-j}C')$. Then, for all ξ,ξ', and all $k\in\mathbb{Z}$, there exists only one k^* such that the frequency supports of $[x\star\psi_\xi]^k$ and $[x\star\psi_{\xi'}]^{-k^*}$ are not disjoint, i.e. such that $C^{\mathrm{WPH}} x((j,\theta,k),(j',\theta',k'),\tau) \neq 0$. This tells us that

$$\sum_{\alpha\in\mathcal{A}_A} e^{-ik\alpha} C^{\mathrm{ALPHA}} x((j,\theta,\alpha),(j',\theta',0),\tau) \simeq c_k c_{k^*} C^{\mathrm{WPH}} x((j,\theta,k),(j',\theta',k^*),\tau).$$

Thus, computing all correlations for $\alpha\in\mathcal{A}_4$, and $\alpha' = 0$ gives us (approximately) all the information contained in WPH coefficients for any pair k, k'.

This result lies on several approximations, and strong assumptions about the wavelets, which are not fully met in practice. For this reason, setting $\alpha' = 0$ instead of $\alpha' \in \mathcal{A}_4$ effectively reduced the amount of information captured by the statistics, and therefore increases the diversity of the model. However, as we observe in Figure 6.2, there is not too much information lost, and the resulting model still captures most of the important geometric structures in texture images. In order to avoid this memorization effect, we shall, in the rest of the chapter, consider only the intermediate model.

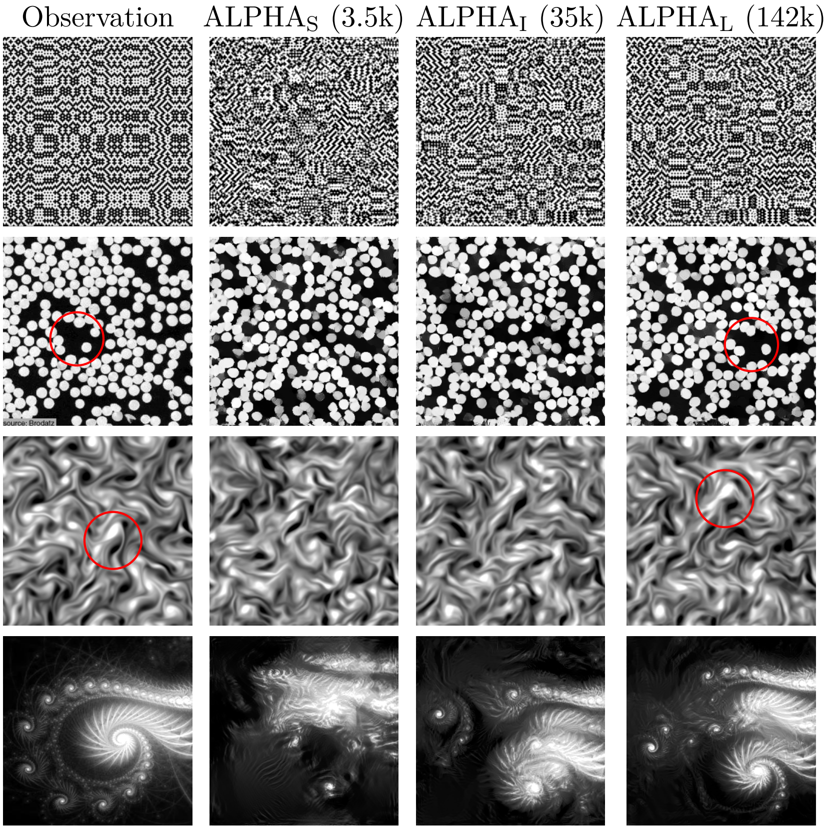


Figure 6.2: Examples of syntheses from the ALPHA models defined in Table 6.1. Visually similar image patches in textures are highlighted by red circles. The number of statistics is given in brackets next to each model name. From top to bottom: a Julesz counterexample, a stationary texture image, a stationary turbulent field and a non-stationary image. We see that the model ALPHA$_\mathrm{I}$ achieves a balance between the visual quality and diversity on these examples.

6.3.3 Modelling color interactions

In order to generate color images, the covariance model ALPHA$_\mathrm{I}$ defined in Section 6.3.2 could be directly applied to each R, G and B color channel independently. However, it would not take into account the color coherence in the structures of the observation.

To capture color interactions in the observation image, we shall impose the covariance between the coefficients of (6.5) for all indices in Υ and all color channels. More precisely, let $x = \{x^c\}_{c=1,2,3}$ be a color image, with the parameter c representing the color channel. The ALPHA$_\mathrm{C}$ color model is defined by correlations between coefficients of the form:

$$R^{\mathrm{ALPHA_C}} x(\gamma, u) = \rho_\alpha(x^c \star \psi_{j,\theta})(u) - \mu_\gamma, \quad \gamma = (j, \theta, \alpha, c). \tag{6.8}$$

The set of indices is defined as

$$\Upsilon^{\mathrm{ALPHA_C}} := \{(\gamma, \gamma', \tau) : ((j, \theta, \alpha), (j', \theta', \alpha'), \tau) \in \Upsilon^{\mathrm{ALPHA_I}}, (c, c') \in \{1, 2, 3\}^2\}.$$

Reduced ALPHA$_\mathrm{C}$ The model ALPHA$_\mathrm{C}$ has a large size as it computes correlations between all coefficients for all color channels. This size can be significantly reduced by

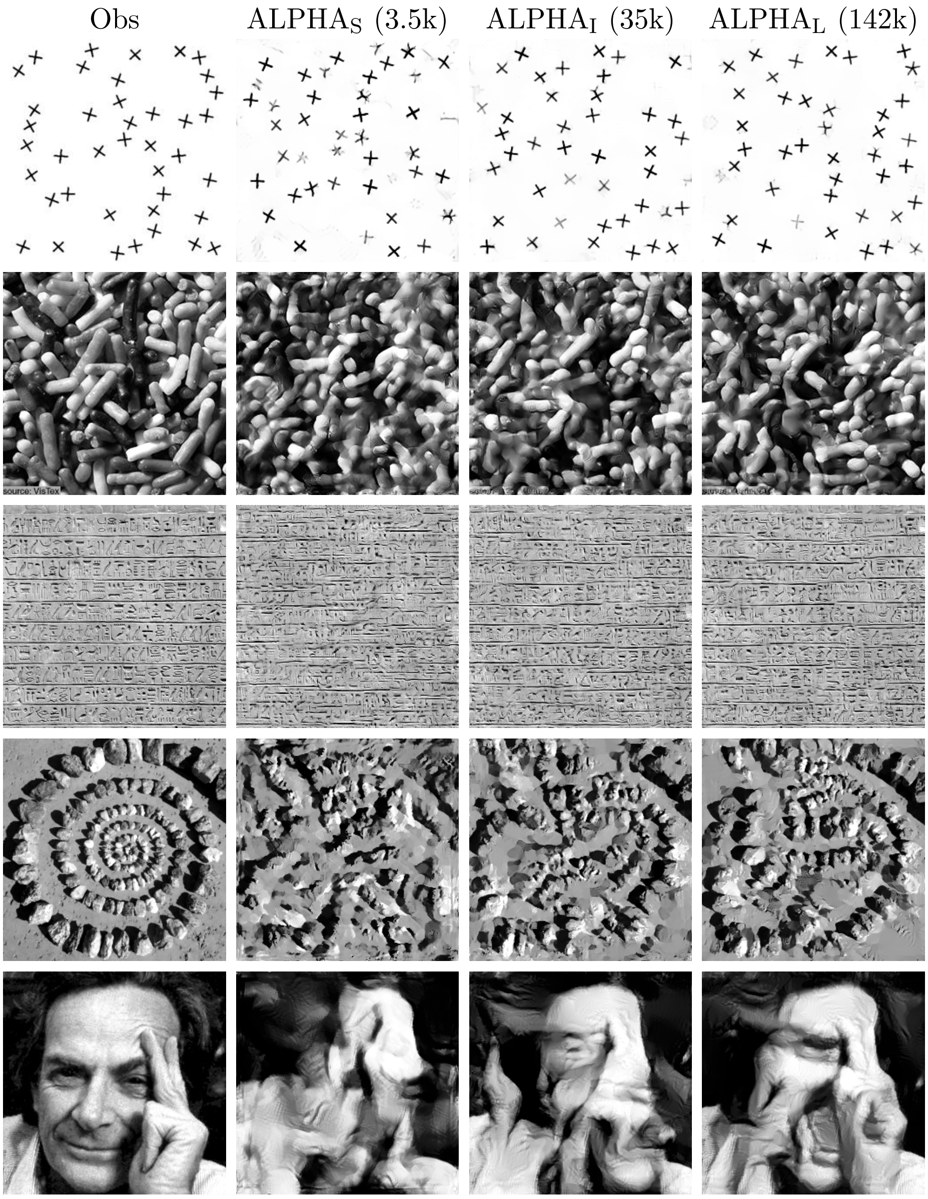


Figure 6.3: Visual comparison of syntheses from the ALPHA$_{\mathrm{S}}$, ALPHA$_{\mathrm{I}}$ and ALPHA$_{\mathrm{L}}$ models.

computing spatially shifted coefficients only for the same color channels (to capture their geometries).

More precisely, it is defined by the following index set: $\Upsilon := \{(\gamma, \gamma', \tau) : ((j, \theta, \alpha), (j', \theta', \alpha'), \tau) \in \Upsilon^{\mathrm{ALPHA_I}}, c = c' \in \{1, 2, 3\}\} \cup \{(\gamma, \gamma', 0) : ((j, \theta, \alpha), (j', \theta', \alpha'), 0) \in \Upsilon^{\mathrm{ALPHA_I}}, (c, c') \in \{1, 2, 3\}^2\}$. This gives a model of size about three times smaller ($\sim$ 113k), with little degradation of the visual quality, as shown in Figure 6.4.

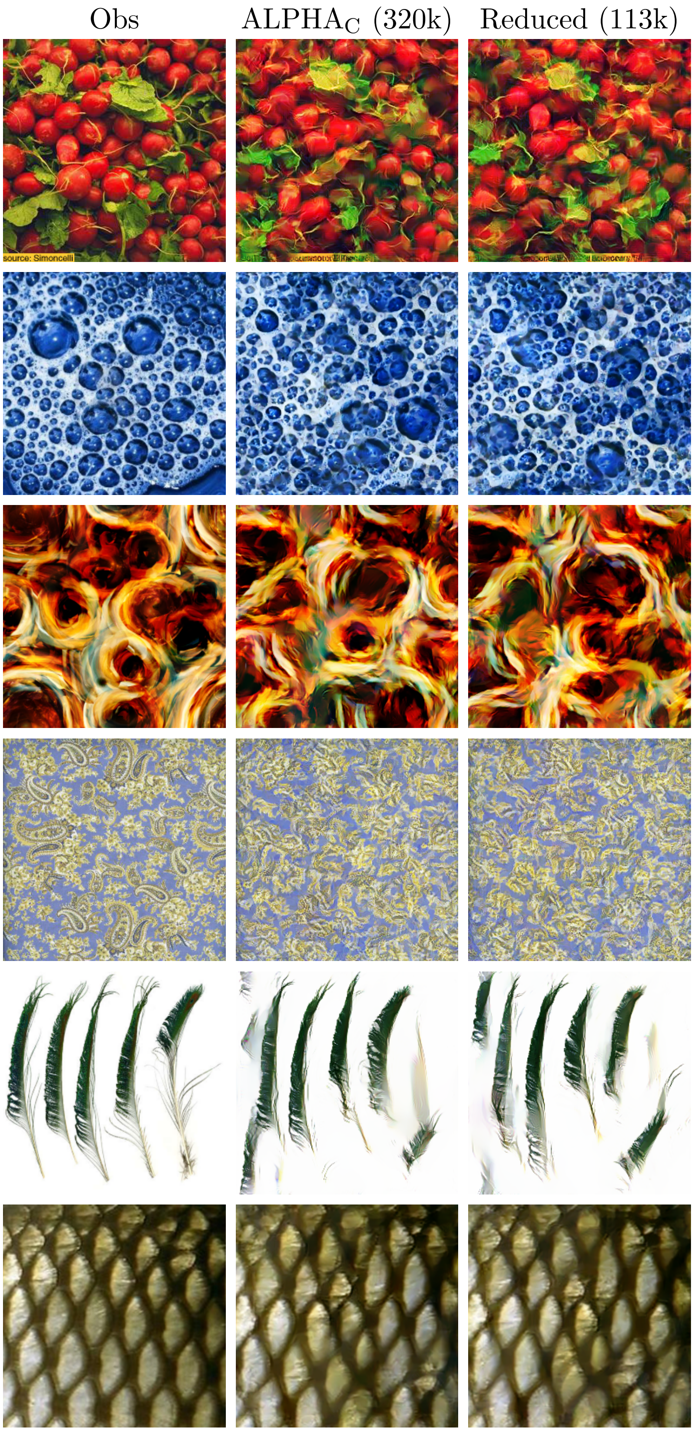


Figure 6.4: Visual comparison of syntheses from ALPHA_C model and its reduced version.

6.4 Numerical experiments

In this section, we compare our intermediate model to the state-of-art models (PS, RF and VGG) on both gray-scale and color textures. We first specify the experimental setup and detail the size of each model. We then present the synthesis results of various examples, and discuss their quality through visual inspection. A quantitative evaluation of the quality of the syntheses, based on the synthesis loss of the VGG model, and proposed in [112], is also discussed here.

6.4.1 Experimental Setup

For our experiments, we choose gray-scale and color textures with various visual structures. Our natural texture examples were obtained from the following three sources: CNS NYU[1], Textures.com[2], Describable Textures Dataset model[3] and the Github page of [12] [4]. In the gray-scale examples, we also include a turbulent vorticity flow, which is simulated from Navier-Stokes equations in two dimensions ([99]). These examples all contain complex geometric structures that are hard to model by the classical PS model.

The texture images presented in this work have a size of $N = 256$, giving a number of pixels of $\sim$65k. For all the ALPHA models, we use Morlet wavelets. This choice differs from the PS model, which uses Simoncelli steerable wavelets. As illustrated in Figure 6.5, this choice can have a visible impact on the quality of the textures. We observe that, while on the first example, the coherence of the structures appear similar for the three wavelet families, the second example shows that the wavelets used in [92] are less efficient in reproducing the contours of the objects (small pebbles). While in our experiments, we chose to use the classical Morlet wavelets, an optimal choice for the wavelet family remains an open problem.

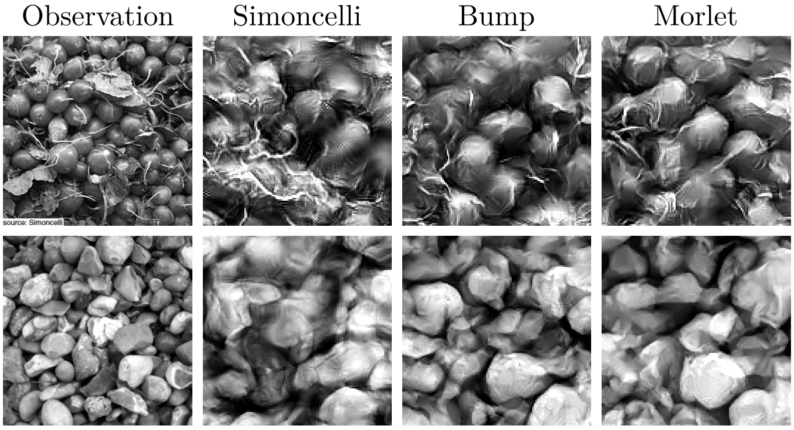


Figure 6.5: Comparison between different wavelets families used in the ALPHA$_\text{I}$ model. Central zooms of syntheses using the same covariance model, with three different wavelet families. From left to right: observation, Simoncelli steerable wavelets, bump steerable wavelets (cf. Example 3.1.3), and Morlet wavelets (Example 3.1.2).

The number of orientations is set to $L = 4$, and the maximal scale is $J = 5$. The maximal scale parameter also plays an important role in the definition of the wavelet transform. It determines the scales of the structures being captured by the transform. If this parameter is too small, large structures in the observation image might not be captured and reproduced in the model syntheses. Conversely, if J is too large, then

---

[1] `http://www.cns.nyu.edu/~lcv/texture/`

[2] `https://textures.com/`

[3] `https://www.robots.ox.ac.uk/~vgg/data/dtd/index.html`

[4] `https://github.com/guillaumebrg/texture_generation`

the large scale statistics may have a high variance, inducing a memorization effect in the syntheses. Figure 6.6 illustrates this point on two examples from Section 6.4.2. By setting $J = 4$ (i.e. the maximal range of structures captured by the wavelets is of size $2^4 = 16$), we observe on the first example that the larger structures (bubbles) are not well reproduced. When J is set to 6, the observation is almost identically reproduced by the synthesis. Similarly on the second examples, several parts of the synthesis appear very similar to ones in the observation. We found that a suitable trade-off consists in setting $J = 5$ for images of size $N = 256$.

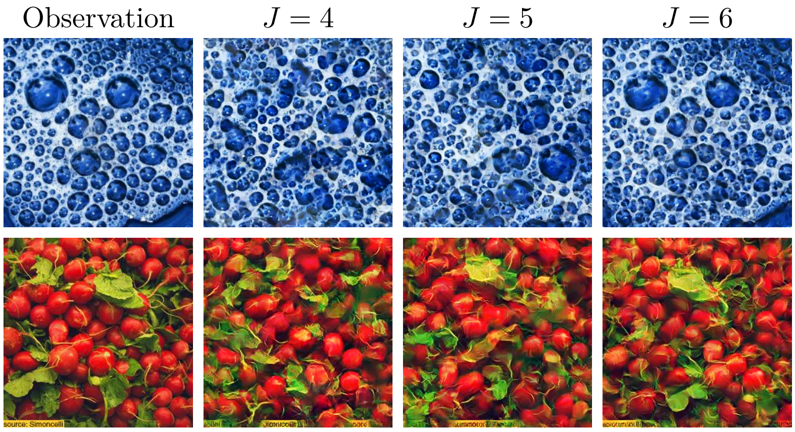

Observation $J = 4$ $J = 5$ $J = 6$

Figure 6.6: Syntheses from the $\mathrm{ALPHA}_{\mathrm{C}}$ model defined with three different maximal scale parameter $J \in \{4, 5, 6\}$.

Similarly to [54] and [112], we use the L-BFGS algorithm [86] for the optimization of the objective. As in the VGG model ([54]), we further apply a histogram matching procedure as post-processing after optimization.

Number of statistics Here, we detail the number of statistics in the different ALPHA models. We begin by giving the formula for each model (note that we do not include the low-pass statistics, which numbers are negligible). To (partially) avoid redundancy in the coefficients, for all models, we compute only the correlations for indices in Υ such that $j_2 \geq j_1$. This gives us the following formulas for the number of statistics:

- $\#(\mathrm{ALPHA}_S) = (2J - 1)|\Theta_4|^2|\mathcal{A}_4| + J|\Theta_4||\mathcal{A}_4||\Upsilon|$
- $\#(\mathrm{ALPHA}_{\mathrm{I}}) = \frac{1}{2}J(J+1)|\Theta_4|^2|\mathcal{A}_4||\Upsilon|$
- $\#(\mathrm{ALPHA}_{\mathrm{L}}) = \frac{1}{2}J(J+1)|\Theta_4|^2|\mathcal{A}_4|^2|\Upsilon|$
- $\#(\mathrm{ALPHA}_{\mathrm{C}}) = \frac{9}{2}J(J+1)|\Theta_4|^2|\mathcal{A}_4||\Upsilon|$

Note that, for all ALPHA models, we also compute first order statistics μ_γ, i.e. the spatial averages of $R^{\mathrm{ALPHA}}\bar{x}(\gamma, u)$. There are $J|\Theta_4|^2$ statistics of this sort in every model, which is negligible with respect to the total number of second order statistics.

Note also that there are still some redundancies in these statistics, as for $j = j'$, all correlations for $\theta, \theta', \alpha, \alpha', \tau$ are counted twice. The number of such statistics is[5], for all

---

[5] The number of such moments is of the order of $J|\Theta|^3$ for the small model, $J|\Theta|^4$ for the intermediate and large models, and $9J|\Theta|^4$ for the color model.

model, superior to the number of first order statistics, which shows that our formula is in fact an upper bound for the exact number of statistics.

This leads us to the following descriptor sizes: for gray-scale textures, the number of the statistics used in the $\mathrm{ALPHA_I}$ model is ~35k. The small model $\mathrm{ALPHA_S}$ has ~3.5k statistics, while the large model $\mathrm{ALPHA_L}$ has ~142k statistics. For comparison, the PS, RF and VGG models have ~7k, ~525k , ~177k statistics respectively. For color textures, the number of statistics of the $\mathrm{ALPHA_C}$ model is ~320k, while for the PS, RF and VGG models, it is ~36k, ~525k, ~177k respectively.

Non-periodic boundaries in natural images The convolution operation in the wavelet transform ((6.2)) is performed using the Fast Fourier Transform. Additionally, recall from Section 6.2.1 that spatial shifts are defined with periodic boundary conditions. This implies periodicity of the input image x. However, natural texture images are not periodic, so one needs to adapt the computation of coefficients to take into account possible border effects. To that end, instead of averaging over all $u \in \mathcal{X}_N$ as in eq. (6.1), each correlation coefficient is averaged over a sub-window inside $\mathcal{X}_N$ (also proposed for the point process synthesis model, cf. Section 5.9), which size depends on the scales of the coefficients being correlated. More precisely, let $\gamma = (j, \theta, \alpha)$ and $\gamma' = (j', \theta', \alpha')$. Note $j_m := \max(j, j')$. We define $\mathcal{X}_{j_m} := \{u = (u_1, u_2) \in \mathcal{X}_N : 2^{j_m} \leq u_i < N - 2^{j_m}, \ i = 1, 2\}$. Then, for non periodic images, we compute

$$C^{\mathrm{ALPHA}}x(\gamma, \gamma', \tau) = \frac{1}{|\mathcal{X}_{j_m}|} \sum_{u \in \mathcal{X}_N} \mathbb{1}_{\mathcal{X}_{j_m}}(u) \mathbb{1}_{\mathcal{X}_{j_m}}(u - \tau) R^{\mathrm{ALPHA}}x(\gamma, u) R^{\mathrm{ALPHA}}x(\gamma', u - \tau), \tag{6.9}$$

where the spatial shifts are defined periodically. Note that the spatial averages μ_γ and $\mu_{\gamma'}$ are also performed on $\mathcal{X}_{j_m}$.

6.4.2 Results

In Figure 6.7, we present examples of syntheses from the $\mathrm{ALPHA_I}$ (or $\mathrm{ALPHA_C}$), PS, RF and VGG models, for both gray-scale and color textures, as well as for non-stationary images. We observe that our model $\mathrm{ALPHA_I}$ produces texture syntheses of similar visual quality to the RF and VGG models. It also significantly outperforms the PS model in terms of the visual quality, without introducing visible memorization effects. As the model PS uses the statistics closer to $\mathrm{ALPHA_S}$ compared to $\mathrm{ALPHA_I}$, the performance of PS is somehow expected. More synthesis examples can be found in Figures 6.9 and 6.10.

Note that for the tiles example (the fifth row) in Figure 6.7, the VGG model produces less convincing textures, because the long-range correlations present in the image (aligned tiles) are not reproduced. To remedy this issue, it has been proposed in [12] to add spatial shifts to the correlations of the network feature maps. These shift statistics are similar to the parameter τ in our model. We also observe that, in the case of the sixth row example (flowers), all models fail to reproduce complex structures at object-level. Possible improvements of such models is further discussed in Section 6.5.

For non-stationary images, we find that certain image patches can be more or less memorized by the RF, VGG and $\mathrm{ALPHA_I}$ models, as illustrated in the seventh row example. Understanding such memorization effect of non-stationary images is a subtle topic, as we find that in some binary images ($\bar{x}(u) \in \{0, 1\}$), only the PS model can reproduce the observation, even though it has a much smaller number of statistics (the last row example). We find that this is related to the spatial correlation statistics in PS (non-zero τ). By removing this constraint, $\bar{x}$ is no longer always reproduced. This simple example suggests that sometimes it is very important to choose the right statistics to capture specific geometric structures. The non-stationary nature also appear in some logo near the boundary of some textures (e.g. bottom left in the observation of the first and fourth rows). Although this logo is reproduced by RF, VGG, ALPHA_I and ALPHA_C, it is a very local phenomenon, as we do not find visible copies of the textures when there is no logo, and it is likely related to the way one addresses the boundary effect (cf. Section 6.4.1).

VGG score In [112], the authors proposed to use the synthesis loss of the VGG model to evaluate the quality of syntheses from any model. The goal is to define a quantitative, and more objective evaluation method than mere visual inspection. Since the VGG model produces syntheses almost indistinguishable from real textures, it is natural to consider its loss to asses the quality of a synthesis. We computed this loss for the examples presented in Figure 6.7, and report it in Table 6.2. Note however that this loss is not exactly the same as the one used in [112], as the layers selected to compute the loss are different. In this work, we chose to use the layers suggested in [54], i.e. 'conv1_1', 'pool1', 'pool2', 'pool3', and 'pool4' of the VGG-19 network ([102]), and compute the relative VGG loss[6].

We notice that this score is not always consistent with visual inspection, as there are texture examples and models for which the syntheses do not look much like the observation image, yet produce a small VGG loss (see e.g. the first and last rows of Figure 6.7, the RF model syntheses have the smallest loss). It should also be noted that the VGG loss reported on the VGG syntheses is *not* the synthesis loss after optimization, as a histogram matching (HM) procedure is performed as post-processing after optimization. We observed that the VGG loss of the syntheses from the VGG model after HM was considerably higher than the one for syntheses before it, while being visually very similar as illustrated in Figure 6.8. These observations suggest that the VGG score suffers from instabilities after reaching a certain level (that is, if the VGG loss is small enough, small perturbations of the values of the image pixels might have a strong impact on the loss).

Data / Model	ALPHA_I	VGG	PS	RF
Radishes	5.02e-05	1.87e-05	2.37e-04	1.13e-05
Cherries	4.86e-05	1.47e-06	6.68e-04	9.65e-06
Gravel	5.97e-05	3.08e-06	7.25e-04	1.29e-05
Turbulence	5.59e-05	5.97e-05	2.42e-04	3.95e-05

Table 6.2: Relative VGG loss of each model on examples in Figure 6.7. Each row corresponds to the same row in Figure 6.7.

---

[6] Using the code from `https://github.com/ivust/random-texture-synthesis/blob/master/vgg_loss.py` (function style_loss_relative).

6.5 Discussion

In this work, we presented a new generative model for texture images built on a wavelet-based representation of geometric structures. Inspired from the classical work of [92], this model incorporates a wider range of statistics, by computing covariances between rectified wavelet coefficients, at different scales, phases and positions. We showed that this model is able to capture and reproduce complex geometric structures, present in natural textures or physical processes, producing syntheses of similar quality to state-of-the-art models that use CNN-based representations.

Being defined with a wavelet family instead of multi-scale random filters, the proposed model uses less statistics than the RF model for color textures. For the gray-scale textures, our model has about 15 times less statistics, as it focuses on capturing the geometric structures present in images. However, further reduction of the number of statistics remains an open problem. For instance, the size of the model could be reduced with a more sophisticated selection of the spatial shift parameter τ, depending on the scales of the coefficients being correlated. Additionally, our color model is defined by computing covariances between coefficients of all different color channels, and is therefore nine times larger than the gray-scale model. In model of [92] extended to color images performs a Principal Component Analysis on the color channels ([113]). A similar idea could also be applied to our model for color images. Another line of research to reduce the size of the model could be to incorporate a second layer of wavelet transform ([74, 100]). Such statistics could also capture additional information to improve the quality of the synthesized textures.

Finding a minimal set of statistics to define a texture model is important because a large number of statistics can result in a high variance of the estimators, and the associated model can suffer from memorization effects. That is, parts of the observation image can be reproduced in the syntheses. This is a problem because the aim of the model is to approximate the underlying distribution of the observation, and therefore produce diverse textures. In this regard, the mere visual evaluation of the synthetic textures can fail to take this aspect of the model into account. A more general evaluation of texture models is an interesting line of research, left for future works.

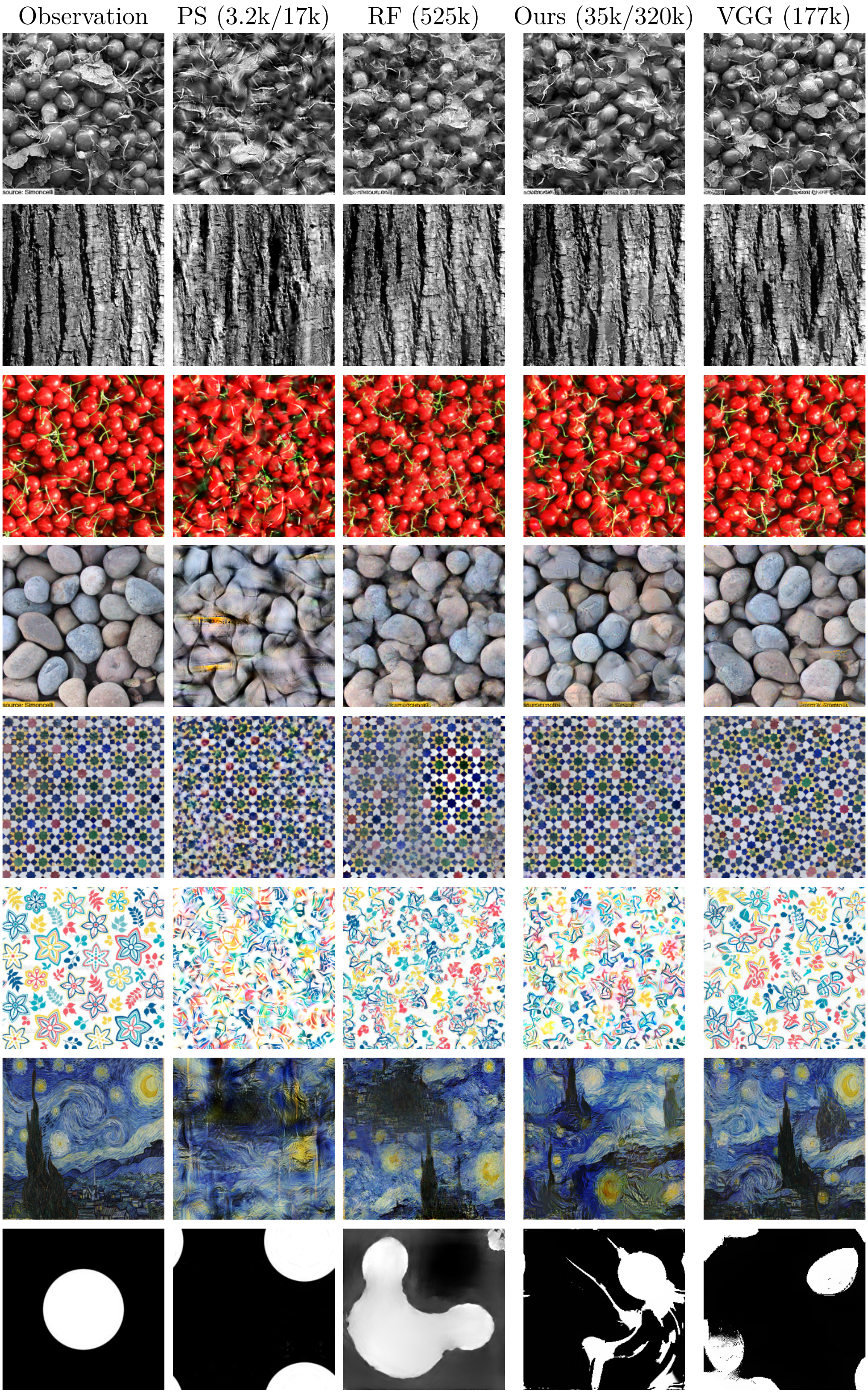


Figure 6.7: Visual comparison between the gray-scale and color models. 'Ours' denotes the ALPHA_I model for gray-scale images and the ALPHA_C model for color images.

Figure 6.8: Visual comparison of syntheses from the VGG model, before and after histogram matching. Before, the relative VGG loss is 2.22e-08, while after, the loss is 5.38e-05.

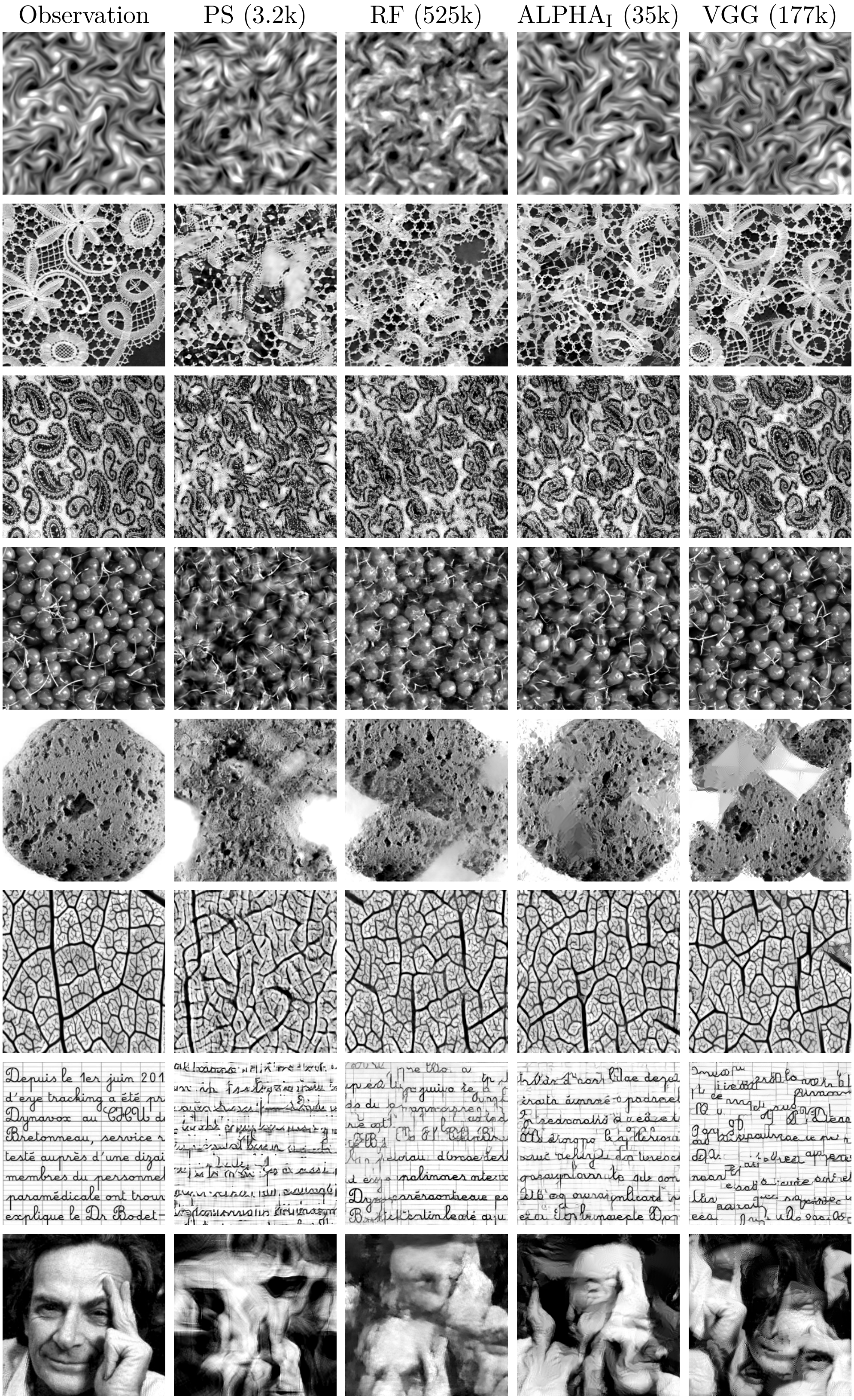


Figure 6.9: Further visual comparison between different texture models on gray-scale images.

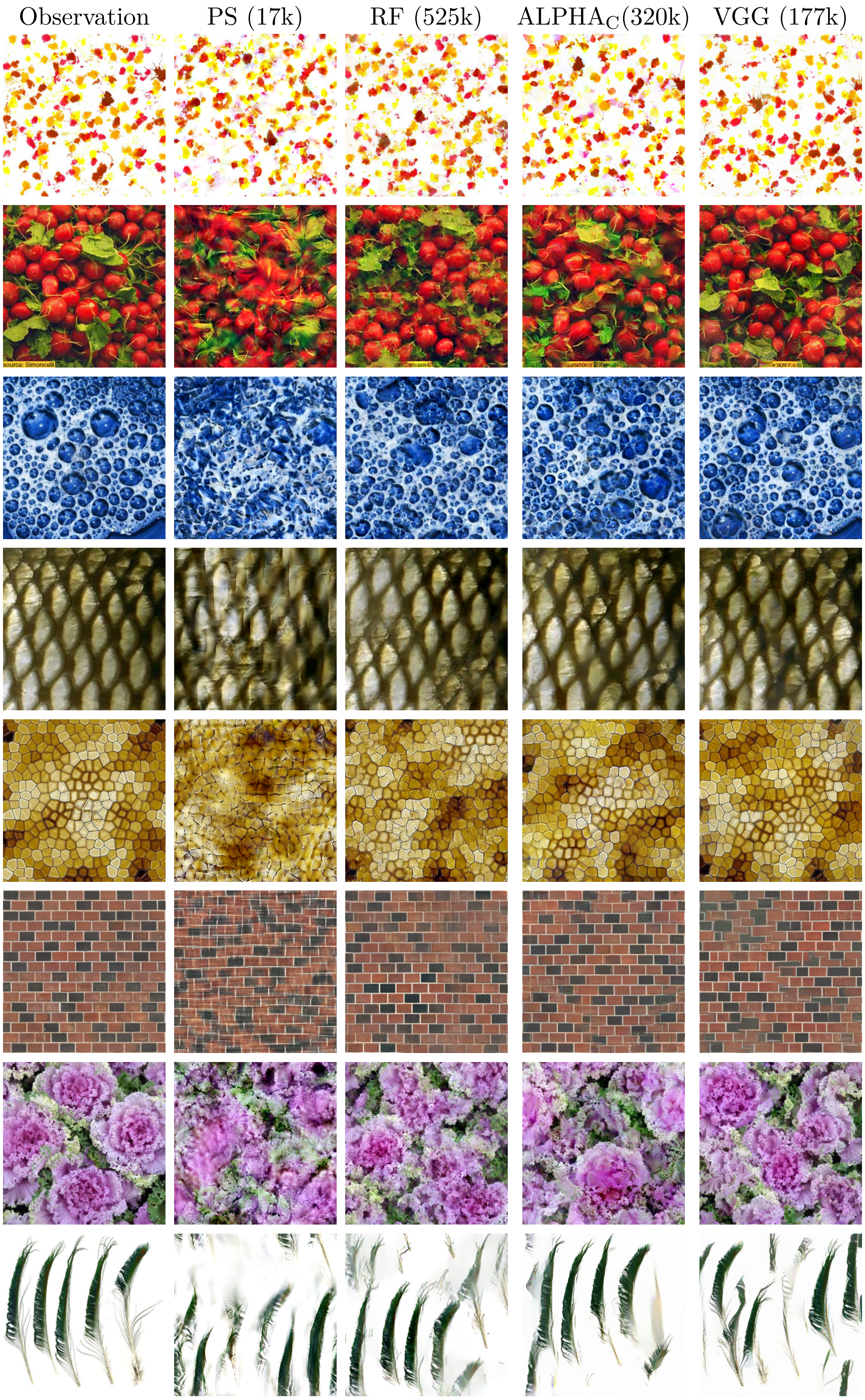


Figure 6.10: Further visual comparison between different texture models on color images.

Chapter 7

Conclusion

This dissertation studied different wavelet-based representations of random geometric processes, and in particular point processes. Similarly to classical summary characteristics, these representations capture geometric structures by applying non-linear transformations to the data to extract relevant information about the inter-point distances. However, the wavelet transform separates the structures present in the data at different scales, which allows one to decompose the information into different parts, making such representations suitable for statistical tasks such as regression or probabilistic modelling.

7.1 Summary of findings

We first studied the scattering moments representation of point processes (cf. Section 3.2), which decompose the geometric structures into different frequency bands, and those bands into different frequency sub-bands. We observed that this relatively concise representation (1k statistics) is able to efficiently represent the geometry of point processes, and is suitable to perform the task we considered, which was to estimate different geometric marks associated with point patterns.

The scattering moments, however, do not capture the dependencies between the different frequency parts of the patterns. Additionally, the complex modulus operator discards the phase of the wavelet coefficients, and induces symmetries which may lead to the loss of important information about the data. For these reasons, we took interest in the wavelet phase harmonics covariance moments, which compute correlations between phase adjusted wavelet coefficients at different frequency bands. We found that this representation is able to capture the geometry of various point process distributions, and allows to efficiently model these distributions, compared to the summary characteristics usually considered in the literature.

To further improve the representation of complex geometric objects, we turned to the modelling of texture images, where state-of-the-art methods are based on the representation of images with convolutional neural networks. By noticing the important amount of information carried by spatially shifted wavelet coefficients, we built a wavelet covariance model that performs similarly to these CNN-based state-of-the-art models, illustrating that such wavelet-based model can capture most of the geometric information present in random processes such as natural texture images.

7.2 Future perspectives

While this work presents empirical evidence that wavelet-based moments can adequately describe point patterns and point processes, little is known about their mathematical properties, such as their discriminating power, their limit behaviour, their stability to deformations, their dependency on different geometric marks, or their ability to fully characterize a certain class of point processes. Recent advances have been made in this area (see e.g. [91]), and further efforts in this direction is an interesting line of research.

The optimization method presented in Chapter 5 differs from classic gradient-descent models, such as in Chapter 6, as it moves mass around in space, rather than changing the values of each pixel. Preliminary results indicate the possibility of synthesizing physical fields (such as a turbulent flows, or an observation of cosmic gas) with this method, by approximating this field by a sum of Gaussian functions. More precisely, we generate $\tilde{x}(u) = \sum_i \exp(-(u-u_i)^2/2\sigma_i^2)/(\sigma_i\sqrt{2\pi})$, such that $Cx \sim C\tilde{x}$, by optimizing the parameters x_i and σ_i, that is, the positions and variances of the Gaussian functions. It could be interesting to study if this method can improve synthesis models, e.g. by avoiding shallow local minima of the objective function.

Another promising line of research is the use of Topological Data Analysis (cf. Section 5.6.3) to evaluate the error between a distribution and a model. This method seems to be able to distinguish between various types of point process distributions, but also between different samples from a same distribution, enabling to evaluate the diversity of a model. TDA has been proposed for analysing images in learning tasks (see e.g. [68], so it could be promising to study this tool as an evaluation of texture synthesis models.

The scattering moments and phase harmonics covariances offer representations of somewhat different nature: the former captures rich information of processes at each scale, with an architecture similar to these of CNNs by using a cascade of linear and non-linear operators, while the other efficiently captures correlations between the different scales. It is easy to see that spatially shifted correlations, as in eq. (6.1), have a direct link with a second layer of convolution, by noting that $Rx(\gamma,\cdot) \star \psi(u) = \int Rx(\gamma, u-\tau)\psi(\tau)d\tau$. A unifying way to encode the information contained in both types of moments (scattering and phase harmonics covariance) could be an interesting way to further improve the representation of geometric objects.

Bibliography

[1] P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *IEEE transactions on information theory*, 44(1):2–15, 1998.

[2] E. Allys, T. Marchand, J.-F. Cardoso, F. Villaescusa-Navarro, S. Ho, and S. Mallat. New interpretable statistics for large-scale structure analysis and generation. *Physical Review D*, 102(10):103506, 2020.

[3] J. Andén, L. Sifre, S. Mallat, M. Kapoko, V. Lostanlen, and E. Oyallon. Scatnet. `http://www.di.ens.fr/data/software/scatnet`, 2014.

[4] M. Andreas Klatt, G. Last, and D. Yogeshwaran. Hyperuniform and rigid stable matchings. *Random Structures & Algorithms*, 57(2):439–473, 2020.

[5] D. A. Awan, R. L. Cavalcante, and S. Stanczak. A robust machine learning method for cell-load approximation in wireless networks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2601–2605. IEEE, 2018.

[6] F. Baccelli, B. Błaszczyszyn, and M. Karray. Random measures, point processes, and stochastic geometry, 2020.

[7] A. Baddeley and R. Turner. Spatstat: an r package for analyzing spatial point patterns. *Journal of statistical software*, 12(1):1–42, 2005.

[8] A. Baddeley, A. Jammalamadaka, and G. Nair. Multitype point process analysis of spines on the dendrite network of a neuron. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, pages 673–694, 2014.

[9] M. S. Bartlett. The spectral analysis of point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 25(2):264–281, 1963.

[10] M. S. Bartlett. The spectral analysis of two-dimensional point processes. *Biometrika*, 51 (3/4):299–311, 1964.

[11] Y. Baryshnikov and J. E. Yukich. Gaussian limits for random measures in geometric probability. *The Annals of Applied Probability*, 15(1A):213–253, 2005.

[12] G. Berger and R. Memisevic. Incorporating long-range consistency in cnn-based texture generation. In *International Conference on Learning Representations, ICLR 2017*. OpenReview.net, 2017.

[13] J. Besag. Contribution to the discussion on dr ripley's paper. *JR Stat. Soc.*, 39:193–195, 1977.

[14] B. Błaszczyszyn. Factorial moment expansion for stochastic systems. *Stochastic processes and their applications*, 56(2):321–335, 1995.

[15] B. Błaszczyszyn and M. K. Karray. Performance analysis of cellular networks with opportunistic scheduling using queueing theory and stochastic geometry. *IEEE Transactions on Wireless Communications*, 18(12):5952–5966, 2019.

[16] B. Błaszczyszyn and D. Yogeshwaran. On comparison of clustering properties of point processes. *Advances in Applied Probability*, 46(1):1–20, 2014.

[17] B. Błaszczyszyn and D. Yogeshwaran. Clustering comparison of point processes, with applications to random geometric models. In *Stochastic Geometry, Spatial Statistics and Random Fields*, pages 31–71. Springer, 2015.

[18] B. Błaszczyszyn, E. Merzbach, and V. Schmidt. A note on expansion for functionals of spatial marked point processes. *Statistics & probability letters*, 36(3):299–306, 1997.

[19] B. Błaszczyszyn, M. Jovanovicy, and M. K. Karray. How user throughput depends on the traffic demand in large cellular networks. In *2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pages 611–619. IEEE, 2014.

[20] B. Błaszczyszyn, R. Ibrahim, and M. K. Karray. Spatial disparity of qos metrics between base stations in wireless cellular networks. *IEEE Transactions on Communications*, 64(10): 4381–4393, 2016.

[21] B. Błaszczyszyn, M. Haenggi, P. Keeler, and S. Mukherjee. *Stochastic geometry analysis of cellular networks*. Cambridge University Press, 2018.

[22] B. Błaszczyszyn, D. Yogeshwaran, and J. E. Yukich. Limit theory for geometric statistics of point processes having fast decay of correlations. *The Annals of Probability*, 47(2):835–895, 2019.

[23] J.-D. Boissonnat, F. Chazal, and M. Yvinec. *Geometric and topological inference*, volume 57. Cambridge University Press, 2018.

[24] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

[25] N. Bonnotte. Unidimensional and Evolution Methods for Optimal Transportation, 2013.

[26] P. Brémaud. *Mathematical Principles of Signal Processing: Fourier and Wavelet Analysis*. Springer Science & Business Media, 2002.

[27] P. Brémaud. Fourier analysis of stochastic processes. In *Fourier Analysis and Stochastic Processes*, pages 119–179. Springer, 2014.

[28] A. Brochard, B. Błaszczyszyn, S. Mallat, and S. Zhang. Statistical learning of geometric characteristics of wireless networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 2224–2232. IEEE, 2019.

[29] A. Brochard, B. Błaszczyszyn, S. Mallat, and S. Zhang. Particle gradient descent model for point process generation. *arXiv preprint arXiv:2010.14928*, 2020.

[30] A. Brochard, S. Zhang, and S. Mallat. Generalized rectifier wavelet covariance models for texture synthesis. In *International Conference on Learning Representations*, 2022.

[31] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013.

[32] J. Bruna and S. Mallat. Multiscale sparse microcanonical models. *Mathematical Statistics and Learning*, 1(3):257–315, 2019.

[33] J. Bruna, S. Mallat, E. Bacry, and J.-F. Muzy. Intermittent process analysis with scattering moments. *The Annals of Statistics*, 43(1):323–351, 2015.

[34] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208, 1995.

[35] A. P. Calderón. Intermediate spaces and interpolation, the complex method. *Studia Mathematica*, 24(2):113–190, 1964.

[36] F. Chazal and B. Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*, 2017.

[37] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

[38] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003.

[39] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume II*. Springer, 2007.

[40] A. DasGupta. *Asymptotic theory of statistics and probability*. Springer Science & Business Media, 2008.

[41] I. Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

[42] I. Daubechies. *The wavelet transform, time-frequency localization and signal analysis*. Princeton University Press, 2009.

[43] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[44] D. Dereudre. Introduction to the theory of gibbs point processes. In *Stochastic Geometry*, pages 181–229. Springer, 2019.

[45] P. J. Diggle. *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press, 2013.

[46] P. J. Diggle, S. J. Eglen, and J. B. Troy. Modelling the bivariate spatial distribution of amacrine cells. In *Case Studies in Spatial Point Process Modeling*, pages 215–233. Springer, 2006.

[47] L. Ducasse and A. Pumir. Intermittent particle distribution in synthetic free-surface turbulent flows. *Physical Review E*, 77(6):066304, 2008.

[48] L. Ducasse and A. Pumir. Inertial particle collisions in turbulent synthetic flows: quantifying the sling effect. *Physical Review E*, 80(6):066312, 2009.

[49] M. Eickenberg, G. Exarchakis, M. Hirn, and S. Mallat. Solid harmonic wavelet scattering: Predicting quantum molecular energy from invariant descriptors of 3d electronic densities. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6543–6552, 2017.

[50] M. Eickenberg, G. Exarchakis, M. Hirn, S. Mallat, and L. Thiry. Solid harmonic wavelet scattering for predictions of molecule properties. *The Journal of chemical physics*, 148(24): 241732, 2018.

[51] M. Farge. Wavelet transforms and their applications to turbulence. *Annual review of fluid mechanics*, 24(1):395–458, 1992.

[52] B. T. Fasy, J. Kim, F. Lecci, C. Maria, and V. Rouvreau. TDA: statistical tools for topological data analysis. *Software available at https://cran. r-project. org/package= TDA*, 2014.

[53] D. Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26): 429–441, 1946.

[54] L. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, 28:262–270, 2015.

[55] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.

[56] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6 (6):721–741, 1984.

[57] J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *Journal of*

Mathematical Physics, 6(3):440–449, 1965.

[58] P. Goupillaud, A. Grossmann, and J. Morlet. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, 23(1):85–102, 1984. ISSN 0016-7142. Seismic Signal Analysis and Discrimination III.

[59] P. Goupillaud, A. Grossmann, and J. Morlet. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, 23(1):85–102, 1984.

[60] A. Grossmann and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*, 15(4):723–736, 1984.

[61] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 229–238, 1995.

[62] L. Heinrich, H. Schmidt, and V. Schmidt. Central limit theorems for poisson hyperplane tessellations. *The Annals of Applied Probability*, 16(2):919–950, 2006.

[63] M. Hirn, S. Mallat, and N. Poilvert. Wavelet scattering regression of quantum chemical energies. *Multiscale Modeling & Simulation*, 15(2):827–863, 2017.

[64] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons, 2008.

[65] E. T. Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[66] B. Julesz. Visual pattern discrimination. *IRE transactions on Information Theory*, 8(2): 84–92, 1962.

[67] O. Kallenberg. *Foundations of modern probability*, volume 2. Springer, 1997.

[68] K. Kim, J. Kim, M. Zaheer, J. Kim, F. Chazal, and L. Wasserman. Pllay: Efficient topological layer based on persistent landscapes. *Advances in Neural Information Processing Systems*, 33, 2020.

[69] K. Koňasová and J. Dvořák. Stochastic reconstruction for inhomogeneous point patterns. *Methodology and Computing in Applied Probability*, 23(2):527–547, 2021.

[70] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[71] G. Last and M. Penrose. *Lectures on the Poisson process*, volume 7. Cambridge University Press, 2017.

[72] P. Laube, M. Grunwald, M. Franz, and G. Umlauf. Image inpainting for high-resolution textures using cnn texture synthesis. In *Proceedings of the Conference on Computer Graphics & Visual Computing*, pages 103–107, 2018.

[73] F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77 (4):853–877, 2015.

[74] R. Leonarduzzi, G. Rochette, J.-P. Bouchaud, and S. Mallat. Maximum-entropy scattering models for financial time series. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5496–5500. IEEE, 2019.

[75] C. Li and M. Wand. Combining markov random fields and convolutional neural networks for image synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2479–2486, 2016.

[76] D. C. Liu and J. Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989.

[77] D. C. Liu and J. Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989.

[78] S. B. Lowen and M. C. Teich. Estimation and simulation of fractal stochastic point processes. *Fractals*, 3(01):183–210, 1995.

[79] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way, 3rd Edition*. Academic Press, 2001.

[80] S. Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[81] S. Mallat, S. Zhang, and G. Rochette. Phase harmonic correlations and convolutional neural networks. *Information and Inference: A Journal of the IMA*, 9(3):721–747, 2020.

[82] B. Matern. Spatial variation. stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden fran Statens Skogsforskningsinstitut*, 49(5), 1960.

[83] B. Matérn. Lecture notes in statistics. In *Spatial variation*, volume 36. Springer-Verlag Berlin, 1986.

[84] K. Matsuda and R. Onishi. Turbulent enhancement of radar reflectivity factor for polydisperse cloud droplets. *Atmospheric Chemistry and Physics*, 19(3):1785–1799, 2019.

[85] J. Neyman and E. L. Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(1):1–29, 1958.

[86] J. Nocedal. Updating quasi-newton matrices with limited storage. *Mathematics of computation*, 35(151):773–782, 1980.

[87] T. Oujia, K. Matsuda, and K. Schneider. Divergence and convergence of inertial particles in high-reynolds-number turbulence. *Journal of Fluid Mechanics*, 905, 2020.

[88] K. Paroux. Quelques théoremes centraux limites pour les processus poissoniens de droites dans le plan. *Advances in Applied Probability*, 30(3):640–656, 1998.

[89] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8026–8037, 2019.

[90] M. D. Penrose and J. E. Yukich. Weak laws of large numbers in geometric probability. *The Annals of Applied Probability*, 13(1):277–303, 2003.

[91] M. Perlmutter, J. He, and M. Hirn. Scattering statistics of generalized spatial poisson point processes. *arXiv preprint arXiv:1902.03537*, 2019.

[92] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, 40(1):49–70, 2000.

[93] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[94] L. Raad, A. Davy, A. Desolneux, and J.-M. Morel. A survey of exemplar-based texture synthesis. *Annals of Mathematical Sciences and Applications*, 3(1):89–148, 2018.

[95] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011.

[96] B. D. Ripley. The second-order analysis of stationary point processes. *Journal of applied probability*, 13(2):255–266, 1976.

[97] C. Robert. Machine learning, a probabilistic perspective, 2014.

[98] K. Schneider and O. V. Vasilyev. Wavelet methods in computational fluid dynamics. *Annual review of fluid mechanics*, 42:473–503, 2010.

[99] K. Schneider, J. Ziuber, M. Farge, and A. Azzalini. Coherent vortex extraction and

simulation of 2d isotropic turbulence. *Journal of Turbulence*, 7(44):N44, 2006.

[100] L. Sifre and S. Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1233–1240, 2013.

[101] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *Proceedings., International Conference on Image Processing*, volume 3, pages 444–447 vol.3, Oct 1995. doi: 10.1109/ICIP.1995.537667.

[102] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[103] I. Siomina and D. Yuan. Analysis of cell load coupling for lte network planning and optimization. *IEEE Transactions on Wireless Communications*, 11(6):2287–2297, 2012.

[104] R. S. Stoica, V. J. Martinez, J. Mateu, and E. Saar. Detection of cosmic filaments using the candy model. *Astronomy & Astrophysics*, 434(2):423–432, 2005.

[105] D. Stoyan and A. Penttinen. Recent applications of point process methods in forestry statistics. *Statistical science*, pages 61–78, 2000.

[106] D. Stoyan and H. Stoyan. *Fractals, random shapes and point fields: methods of geometrical statistics*, volume 302. Wiley-Blackwell, 1994.

[107] E. Tempel, R. Stoica, V. J. Martinez, L. Liivamägi, G. Castellan, and E. Saar. Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the sdss. *Monthly Notices of the Royal Astronomical Society*, 438(4):3465–3482, 2014.

[108] E. Tempel, R. S. Stoica, R. Kipper, and E. Saar. Bisous model—detecting filamentary patterns in point processes. *Astronomy and Computing*, 16:17–25, 2016.

[109] S. Torquato. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties.* Springer, New York,, 2002.

[110] A. Tscheschel and D. Stoyan. Statistical reconstruction of random point patterns. *Computational statistics & data analysis*, 51(2):859–871, 2006.

[111] M. Unser and N. Chenouard. A unifying parametric framework for 2d steerable wavelet transforms. *SIAM Journal on Imaging Sciences*, 6(1):102–135, 2013.

[112] I. Ustyuzhaninov, W. Brendel, L. A. Gatys, and M. Bethge. What does it take to generate natural textures? In *International Conference on Learning Representations, ICLR 2017*. OpenReview.net, 2017.

[113] J. Vacher and T. Briand. The portilla-simoncelli texture model: towards understanding the early visual cortex. *Image Processing On Line*, 11:170–211, 2021.

[114] J. Vacher, A. Davila, A. Kohn, and R. Coen-Cagli. Texture interpolation for probing visual perception. *Advances in Neural Information Processing Systems*, 33, 2020.

[115] R. R. Wadhwa, D. F. Williamson, A. Dhawan, and J. G. Scott. TDAstats: R pipeline for computing persistent homology in topological data analysis. *Journal of open source software*, 3(28):860, 2018.

[116] T. Wiegand and K. A. Moloney. *Handbook of spatial point-pattern analysis in ecology.* CRC press, 2013.

[117] S. Xiao, M. Farajtabar, X. Ye, J. Yan, L. Song, and H. Zha. Wasserstein learning of deep generative point process models. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper/2017/file/f45a1078feb35de77d26b3f7a52ef502-Paper.pdf`.

[118] G. Zhang, F. H. Stillinger, and S. Torquato. Ground states of stealthy hyperuniform potentials: I. entropically favored configurations. *Physical Review E*, 92(2):22119, 2015.

[119] G. Zhang, F. H. Stillinger, and S. Torquato. Ground states of stealthy hyperuniform potentials. ii. stacked-slider phases. *Physical Review E*, 92(2):022120, 2015.

[120] S. Zhang and S. Mallat. Maximum entropy models from phase harmonic covariances. *Applied and Computational Harmonic Analysis*, 53:199–230, 2021.

[121] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560, 1997.

RÉSUMÉ

Cette dissertation présente une classe de représentations de processus ponctuels. Inspirés par le succès des méthodes d'ondelettes en traitement du signal, ces descripteurs sont basés sur la convolution d'un processus ponctuel avec un famille d'ondelettes. À partir de ces convolutions sont construits des ensembles de descripteurs statistiques de processus ponctuels stationnaires, en appliquant des opérateurs non linéaires, suivis d'un moyennage spatial. Tout comme les caractéristiques classiques pour les processus ponctuels, ces statistiques sont conçues pour extraire les informations contenues dans le processus en un nombre relativement faible de valeurs numériques, en décrivant sa géométrie. Leur but est de décrire la façon dont les atomes du processus ont tendance à se repousser, ou bien se regrouper, et de la sorte former des formes géométriques complexes. De par leur construction, ces descripteurs bénéficient de plusieurs propriétés qui les rendent adaptés à des tâches d'apprentissage et d'analyse statistique. Afin d'illustrer la qualité de ces représentations en tant que descripteurs statistiques, nous étudions plusieurs problèmes impliquant l'analyse statistique de processus ponctuels. Dans un première expérience, nous cherchons à estimer une fonction inconnue qui prend en argument une configuration de points, et renvoie une version marquée de cette configuration, c'est-à-dire pour laquelle une valeur numérique est associée à chaque atome de la configuration. Nous utilisons une représentation en ondelettes de ces configurations pour estimer la relation entre leur versions non marquées et marquées. Dans un second temps, nous étudions la capacité de ces descripteurs à modéliser certaines distributions de processus ponctuels, en définissant un modèle de maximum d'entropie défini par des statistiques d'ondelettes, calculées sur une unique observation. Pour ces deux problèmes, nous observons que les représentations que l'on propose amènent de meilleures performances que les statistiques classiques couramment utilisées dans la littérature sur les processus ponctuels. Enfin, pour étudier à quel point de telles représentations peuvent capturer les structures géométriques présentes dans les textures, nous définissons un model de maximum d'entropie qui s'appuie sur des statistiques d'ondelettes similaires, produisant des synthèses de qualité comparable à celles de l'état de l'art, dont les modèles sont basés sur des representations à partir de réseaux convolutionels profonds.

MOTS CLÉS

processus ponctuels, traitement du signal, apprentissage statistique, synthèse

ABSTRACT

This dissertation presents a class of representations of spatial point processes. Inspired from the success of wavelet methods in signal processing, these descriptors rely on the convolution of a point process with a family of wavelet filters. From these convolutions are built sets of statistical descriptors of stationary point process, by applying non-linear operators, followed by a spatial averaging. Much like classical summary characteristics for point processes, these statistics are designed to extract information about the process with a relatively small number of numerical values, by describing its geometry. Their goal is to describe whether the atoms of the process tend to repel each other, or cluster together, and by doing so, form possibly complex geometric shapes. By construction, these descriptors enjoy several properties that make them suitable for statistical analysis and learning tasks. To illustrate the quality of these representations as statistical descriptors, we study several problems involving statistical analysis of point processes. In a first experiment, we seek to estimate an unknown function that takes as input a point pattern, and returns a marked version of this pattern, where a numerical value is associated to each atom of the pattern. We use a wavelet-based representation of point patterns to estimate the relation between their non-marked and marked version. We then study, in a second experiment, the ability of such representations to model the distribution of a point process, by defining a maximum entropy model defined by a set of wavelet-based statistics, computed on a single observation. For these two problems, we observe that our representations lead to better performance than summary statistics commonly used in the literature on point processes. Finally, to study to what extent such representations can capture geometric structures of texture images, we define a maximum entropy model relying on similar wavelet statistics, yielding syntheses of similar visual quality to state-of-the-art models based on deep convolutional neural networks representations.

KEYWORDS

point processes, signal processing, statistical learning, synthesis