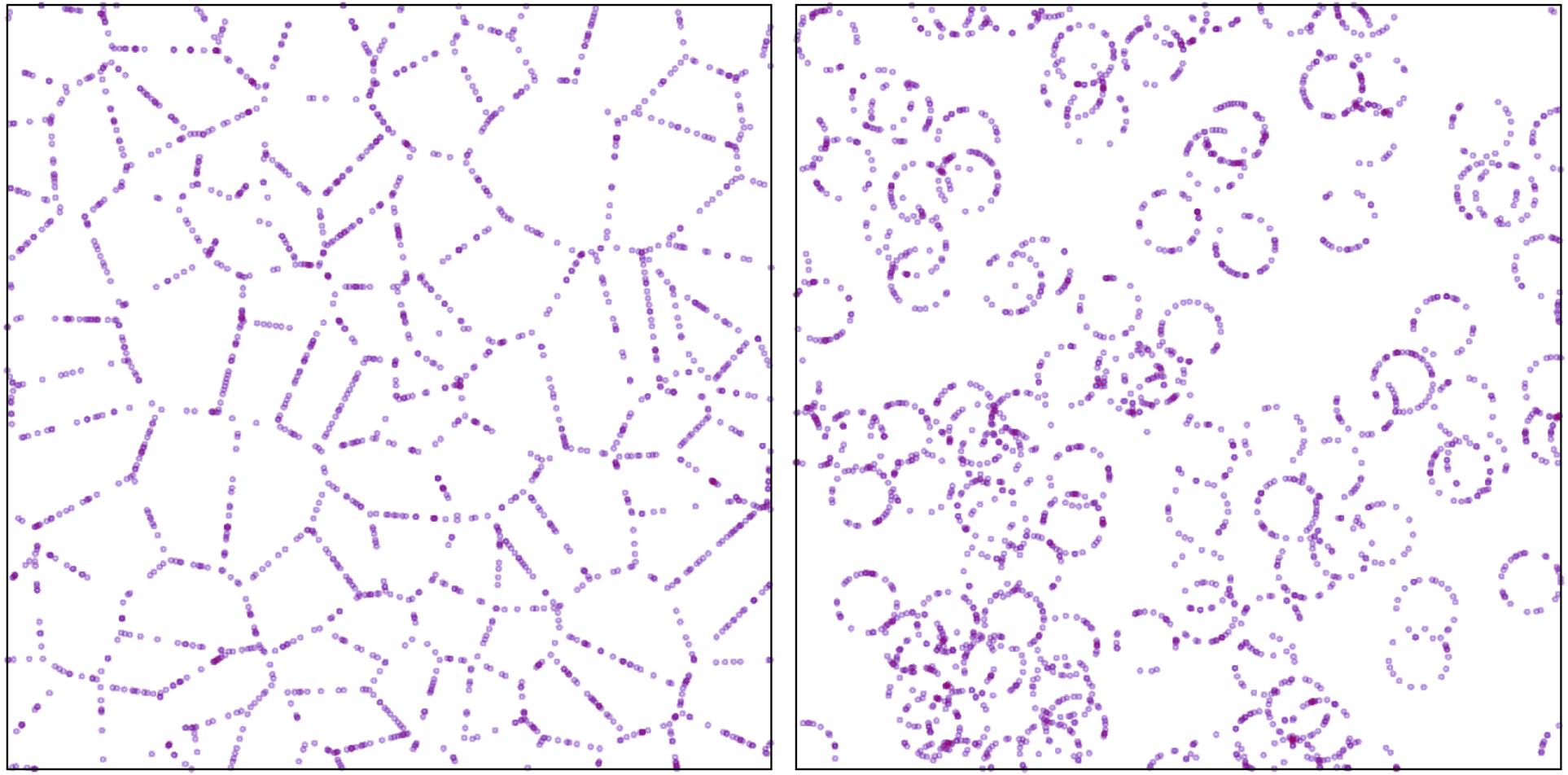
Particle gradient descent model for point process generation

Antoine Brochard, Bartek Błaszczyszyn, Sixin Zhang, and Stéphane Mallat

Journées communes GéoSto MIA
22–23 sept. 2022 Saint-Étienne-du-Rouvray

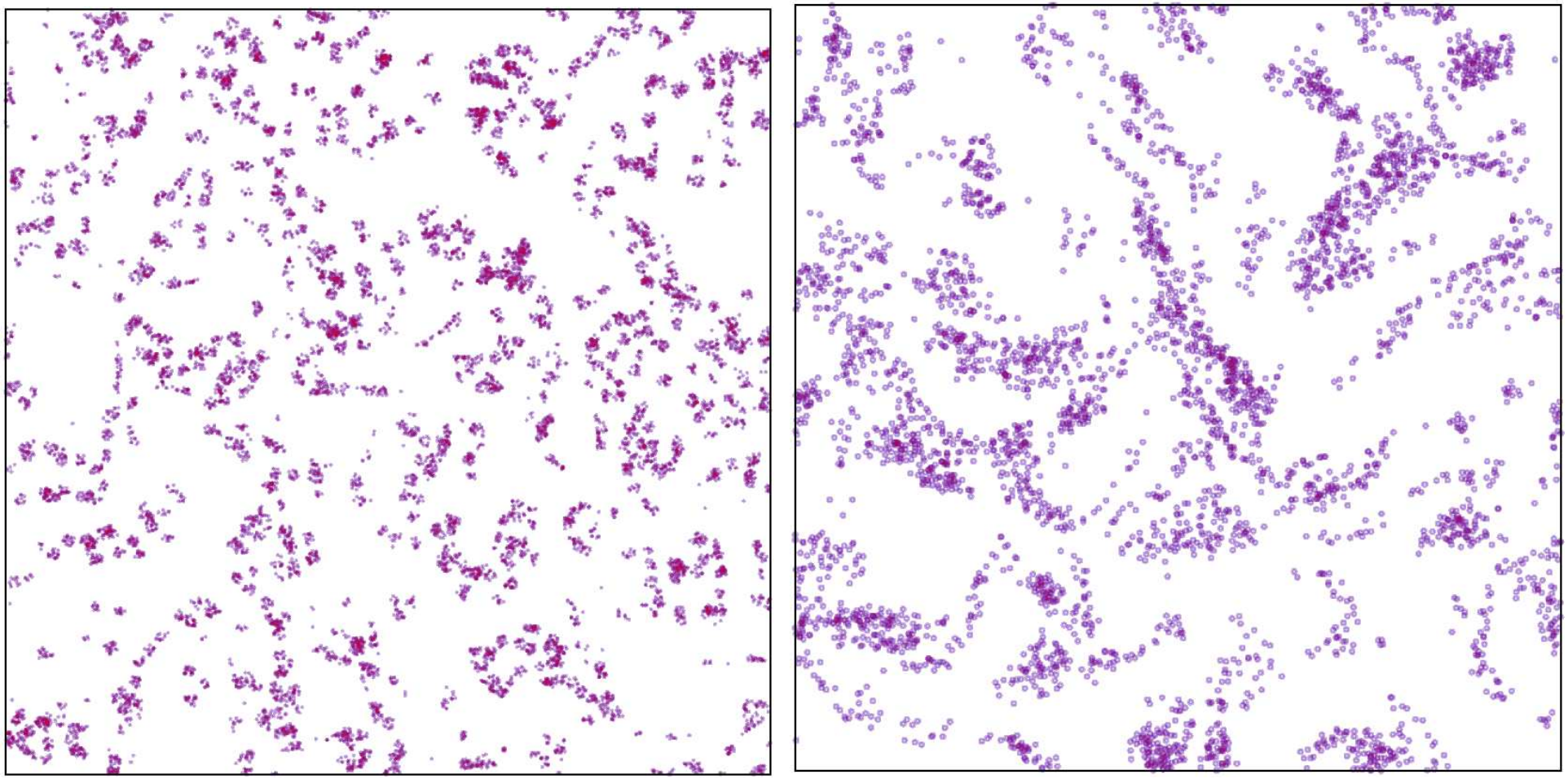
Samples from a point process. Can you recognize the model?



Sure, Cox-Voronoi and Cox-Boolean.

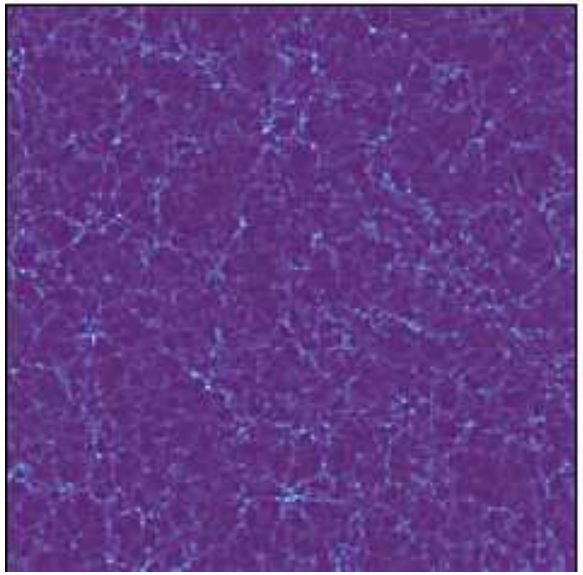
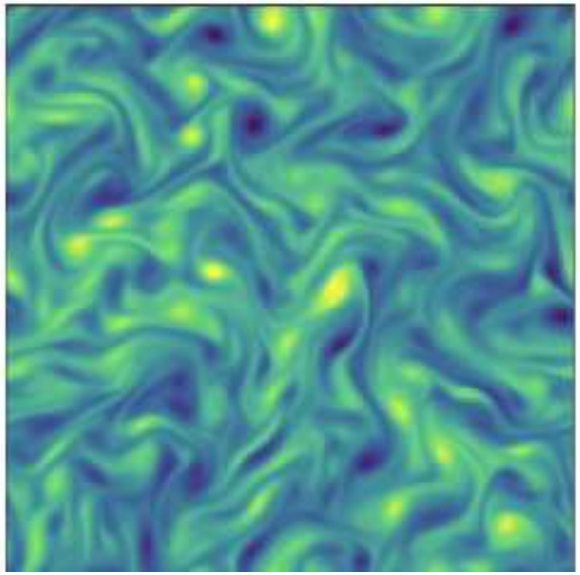
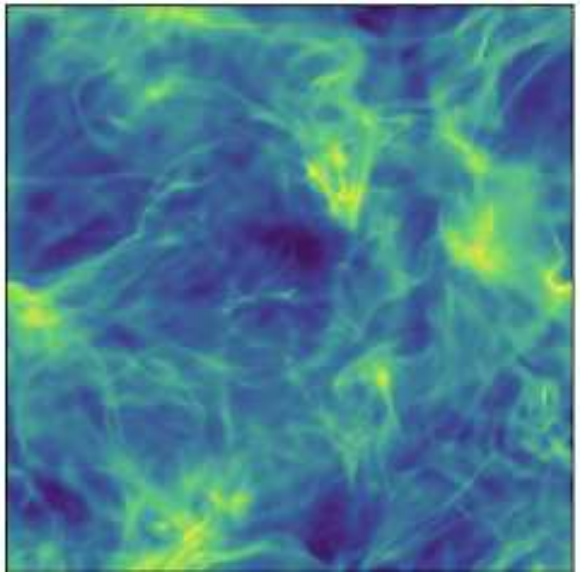
Recall: Cox = doubly stochastic Poisson process.

And here?



Well... ? Left: Matern cluster process driven by some turbulent field (driven by 2d Navier-Stokes equations). Right: Matern II hard core model applied to a Cox driven by the same turbulent field.

Some more patterns

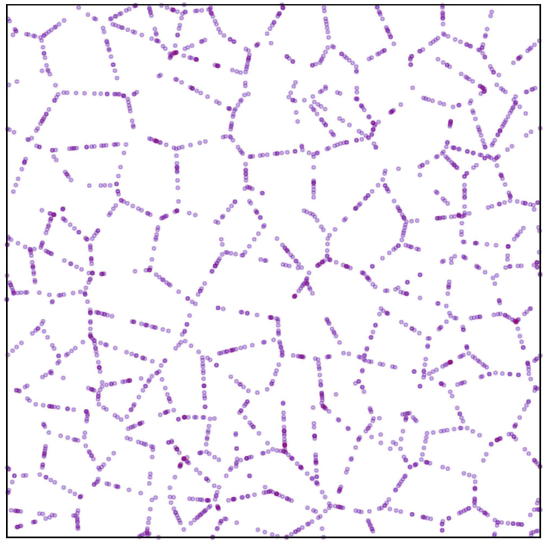
from astronomy, physics, ...
These patterns:

- Exhibit multi-scale properties (e.g. small repulsion, large cluster)
- We want model them with point process with a (very) large number of points (partcles), say $\sim \mathbf{10.000}$, in the window.
- Typically, we have only one original pattern (or, say, very few ones).
- $\Rightarrow$ Ergodic learning of point processes?

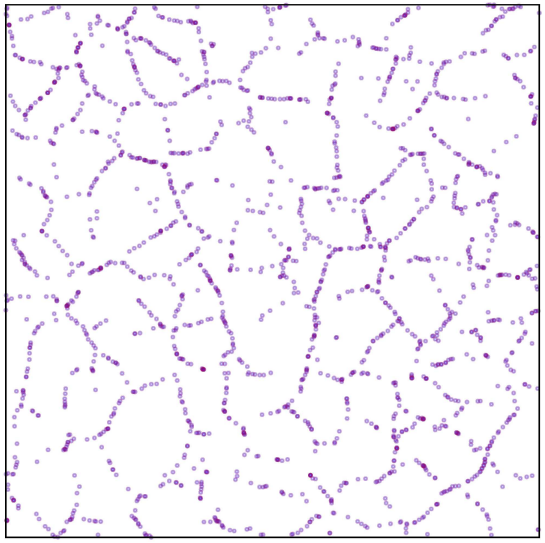
Ergodic learning of point processes

- Recall: Almost surely, any infinite realization of an ergodic point process alows one to fully characterize its distribution and thus (in principle) to sample from this distribution new realizations. $\Rightarrow$ Spatial averaging!

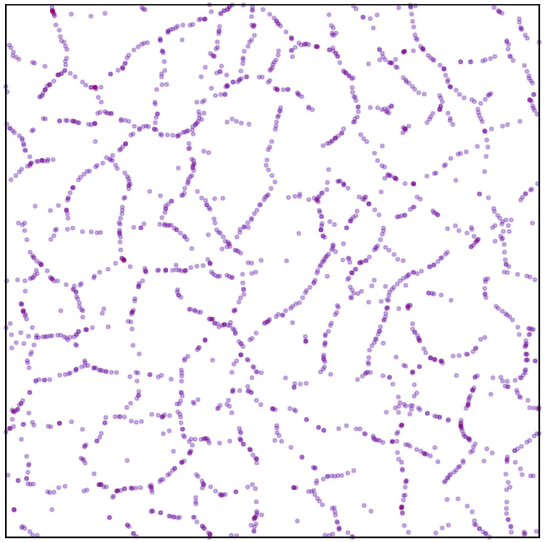
- But in practice, we have only a finite learning window. Can we get approximations of the unknown distribution?



Original image



Synthesis 1



Synthesis 2

samples from "ergodic learning model"

So how it works?

1. Choose statistics (descriptors, moments) that will "summarize" the distribution and not fully "memorize" given patterns.

2. Specify a model deriving from these statistics. Typically a type of "maximum entropy model".

3. Find a way of generating samples from this model. Not always evident!

Outline of the remaining talk

- Models (and their simulation methods)
  - Maximum entropy models (rather theoretical)
  - Particle gradient descent model $\Leftarrow$
  - Random search (benchmark; Torquato 2002, Tscheschel and Stoyan 2006)
- Spatial statistics
  - Classical spatial statistics
  - Wavelet-based representations (Mallat 2021, ...)$\Leftarrow$
- Testing results
  - Visual,
  - Spectrum,
  - Topology analysis (persistent homology) $\Leftarrow$

- MODELS
  - Maximum entropy models

Maximum entropy models

- Based on a set of statistics
  - calculate on the observed pattern $\Rightarrow$ micro-canonical model,
  - known expected values $\Rightarrow$ macro–canonical model.
- Intuitively: model is "as random as possible" under constraints based on the given statistics.

"Randomness" defined with respect to Poisson point process

- Let $\mathcal{L}_1, \mathcal{L}_2$ be two probability distributions on $\mathbb{M}$ (of point processes), 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 well defined by

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

- If $\mathcal{L}_0$ is the Poisson distribution on $\mathbb{M}$ and $\mathcal{L} \ll \mathcal{L}_0$ then

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

is called the entropy of $\mathcal{L}$ (with respect to Poisson ditribution $\mathcal{L}_0$).

Macro-canonical model

- Denote (unknown!?) point process $\Xi \sim \mathcal{L}$.
- Given (a vector of) statistics K of Ξ.
- Averaged constraints

$$\mathrm{E}(K(\Xi)) = a, \qquad \text{or some vector values } a. \tag{AC}$$

- Model:

$$\begin{aligned} &\arg\max_{\hat{\mathcal{L}}} && \mathcal{H}(\hat{\mathcal{L}}) \\ &\text{given} && \text{constraints (AC)} \end{aligned}$$

- Under some technical assumptions the solution of the macro-canonical model is given by the Gibbs point process.
- Computationally expensive: calculating solution for large dimension of K and sampling from it.

Micro-canonical model

- A given realization (of a point process) $\boldsymbol{\Xi}$.
- Given (a vector of) statistics $\boldsymbol{K}$ of $\boldsymbol{\Xi}$.
- Define the "energy" of a realization (of point measure) $\boldsymbol{\mu}$

$$\boldsymbol{E_{K(\Xi)}(\mu) := \frac{1}{2}|K(\mu) - K(\Xi)|^2.}$$

- Path-wise constraints

$$\boldsymbol{\Omega_\epsilon := \{\mu \in \mathbb{M} \,:\, E_{K(\Xi)}(\mu) \leq \epsilon\}} \qquad \text{for some } \boldsymbol{\epsilon > 0}.$$

- Model:

$$\begin{aligned} \boldsymbol{\arg\max_{\hat{\mathcal{L}}}} \quad & \boldsymbol{\mathcal{H}(\hat{\mathcal{L}})} \\ \text{given} \quad & \boldsymbol{\hat{\mathcal{L}}(\Omega_\epsilon) = 1} \end{aligned}$$

- The solution of the micro-canonical model is given by truncation of Poisson $\boldsymbol{\mathcal{L}_0}$ to $\boldsymbol{\Omega_\epsilon}$.
- Sampling computationally expensive (acceptance-rejection method!?).

- MODELS
  - Maximum entropy models
  - Particle gradient descent model

Particle gradient descent on point measures

- As for the micro-cannonical model: a given realization (of a point process) Ξ, given (a vector of) statistics K of Ξ, the "energy" of (an arbitrary) realization μ: $E_{K(\Xi)}(\mu) := \frac{1}{2}|K(\mu) - K(\Xi)|^2$.
- For some gradient step $\gamma > 0$, we define the gradient descent mapping

$$\begin{array}{rccc} F & : & \mathbb{M}^s & \longrightarrow & \mathbb{M}^s \\ & & \mu = \sum_i \delta_{x_i} & \longmapsto & \sum_i \delta_{x_i - \gamma \nabla_{x_i} E_{K(\Xi)}(\mu)}. \end{array} \tag{1}$$

- $F(\mu)$ is a the push-forward of the measure μ by the mapping $F_{\mu,K(\Xi)}(x) := x - \gamma \nabla_x E_{K(\Xi)}(\mu)$; see Molchanov and Zuyev (2002).
- For any initial point measure $\mu_0 \in \mathbb{M}^s$ we iterate the gradient descent function F:

$$\mu_n = \mu_{n,K(\Xi)}(\mu_0) = F(\mu_{n-1}), \qquad n \geq 1. \tag{2}$$

Particle gradient descent model

- For a given $n \geq 1$, and the starting distribution $\mathcal{L}_0 \sim \mu_0$, the n th iteration of F defines the law $\mathcal{L}_n \sim \mu_n = \mu_{n,K(\Xi)}(\mu_0)$.
- $\mathcal{L}_n = \mathcal{L}_{n,K(\Xi)}(\mathcal{L}_0)$ is a n th pushforward operation on $\mathcal{L}_0$ "driven" by $K(\Xi)$.
- Setting a fixed number of iterations n as a stopping rule, we consider $\mathcal{L}_n$ as model, called the gradient descent model driven by $K(\Xi)$.
- Simulation form this model is straightforward, starting from some (say Poisson) realization $\mu_0 \sim \mathcal{L}_0$.
- No theoretical guarantees for $\mathcal{L}_n \simeq \mathcal{L}(\Xi)$ except:

  Theorem. If K and $\mathcal{L}_0$ are invariant w.r.t. some set of rigid motions (translations, rotations, symmetries) on the torus, then $\mathcal{L}_n$ has the same property.

Remarks on the gradient descent model

- Inspiration from the micro-canonical model, however there is no guarantee that the optimization reaches Ω_ϵ for any $\epsilon > 0$.
- Fixed number of iterations and no configuration rejections $\Rightarrow$ model reaches a low energy level.
- Classical line-search methods in the optimization to adjust the gradient step γ so as to ensure that the energy decreases as n grows.

- MODELS
  - Maximum entropy models
  - Particle gradient descent model
  - Random search

Random search (Tscheschel and Stoyan 2006)

- As for the micro-cannonical model: a given realization (of a point process) Ξ, given (a vector of) statistics K of Ξ, the "energy" of (an arbitrary) realization μ: $E_{K(\Xi)}(\mu) := \frac{1}{2}|K(\mu) - K(\Xi)|^2$.

- Optimization step: at step n, given $\mu_n = \sum_{i=1}^{N} \delta_{x_{i,n}}$
  - a point chosen uniformly at random, say $x_{j,n}$, for $j \in \{1, \cdots, N\}$
  - a candidate for a new point chosen uniformly at random in the observation window $y \in W$.
  - Then

$$\mu_{n+1} := \begin{cases} \mu_n - \delta_{x_{j,n}} + \delta_y & \text{if } E_{K(\Xi)}(\mu_{n+1}) < E_{K(\Xi)}(\mu_n) \\ \mu_n & \text{otherwise.} \end{cases}$$

- Observe: move the point one at a time, possibly lot of rejected moves.

- SPATIAL STATISTICS
  - Good choice of statistics

Good choice of statistics

One usually aims at finding the (vector of) statistics $\boldsymbol{K}$ satisfying the following properties:

- Concentration property: $\boldsymbol{K}(\Xi) \simeq \mathbb{E}[\boldsymbol{K}(\Xi)]$ with high probability $\Rightarrow$ not to "memorize" a realization of Ξ.
- Sufficiency property: $\mathbb{E}(\boldsymbol{K}(\Xi))$ rich enough, strong (distributional) discriminate power $\Rightarrow$ "summarize" the unknown distribution.

Assuming ergodicity of Ξ, a natural choice consists in spatial averaging:

$$\boldsymbol{K}_i(\mu) = \frac{1}{|\boldsymbol{W}|} \int_{\boldsymbol{W}} \boldsymbol{f}_i(\boldsymbol{S}_x \mu)\, d\boldsymbol{x} \qquad \mu \in \mathbb{M},$$

for a sufficiently rich class of functions $\boldsymbol{f}$, with support not to large w.r.t. the observation window $\boldsymbol{W}$ so, by ergodicity, $\boldsymbol{K}_i(\mu) \simeq \mathbb{E}[\boldsymbol{K}(\Xi)]$.

SPATIAL STATISTICS (DESCRIPTORS)

- Good choice of statistics
- Classical summary characteristic

Classical spatial statistics

- mean (intensity) $\mathrm{E}[\Xi(B)]/|B|$,
- correlation functions $\rho(x, y)$,
- Ripleys K-function $K(r)$,
- k-nearest neighbor distances $D_k(r)$,
- ...,
- void probabilities $\mathbf{P}(\Xi(B) = 0)$; full distribution characterization,
- Laplace transform $\mathrm{E}[\exp(-\int f \mathrm{d}\Xi)]$; full distribution characterization.

□ SPATIAL STATISTICS

- Good choice of statistics
- Classical summary characteristic
- Wavelet-based representations

Wavelet

Following Bruna, Mallat, Bacry, Muzy (2015),
let ψ be a continuous, bounded, localized in space and frequency, complex valued function on $\mathbb{R}^d$ of zero average $\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}x = 0$.
Usually ψ is normalized so that $\int_{\mathbb{R}^d} |\psi(x)|\,\mathrm{d}x = 1$.

We call ψ (d-dimensional) mother wavelet.

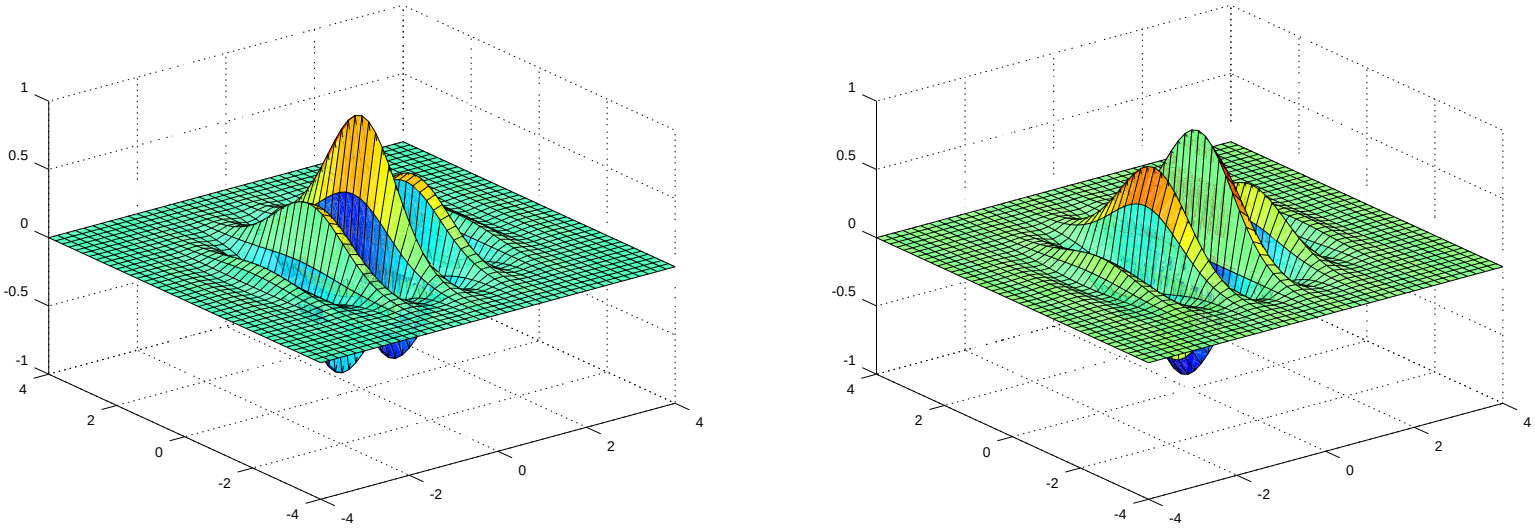
In applications $d = 1$ or 2. In this talk $d = 2$.

Example: 2D Morlet wavelet

Morlet wavelet on the plane

$$\psi(x) = \exp(i\,\omega\cdot x)\exp(-|x|^2/2),$$

where i is the imaginary unit and $\omega\cdot x$ is the scalar product of some nonzero vector parameter $\omega \in \mathbb{R}^2$, called spatial frequency, with $x \in \mathbb{R}^2$.



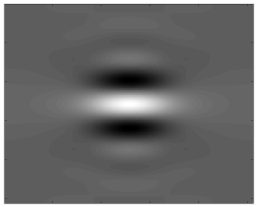
Real and Imaginary part of the Morlet wavelet with $\omega = (5.5, 0)$.

Scaling and rotating the mother wavelet

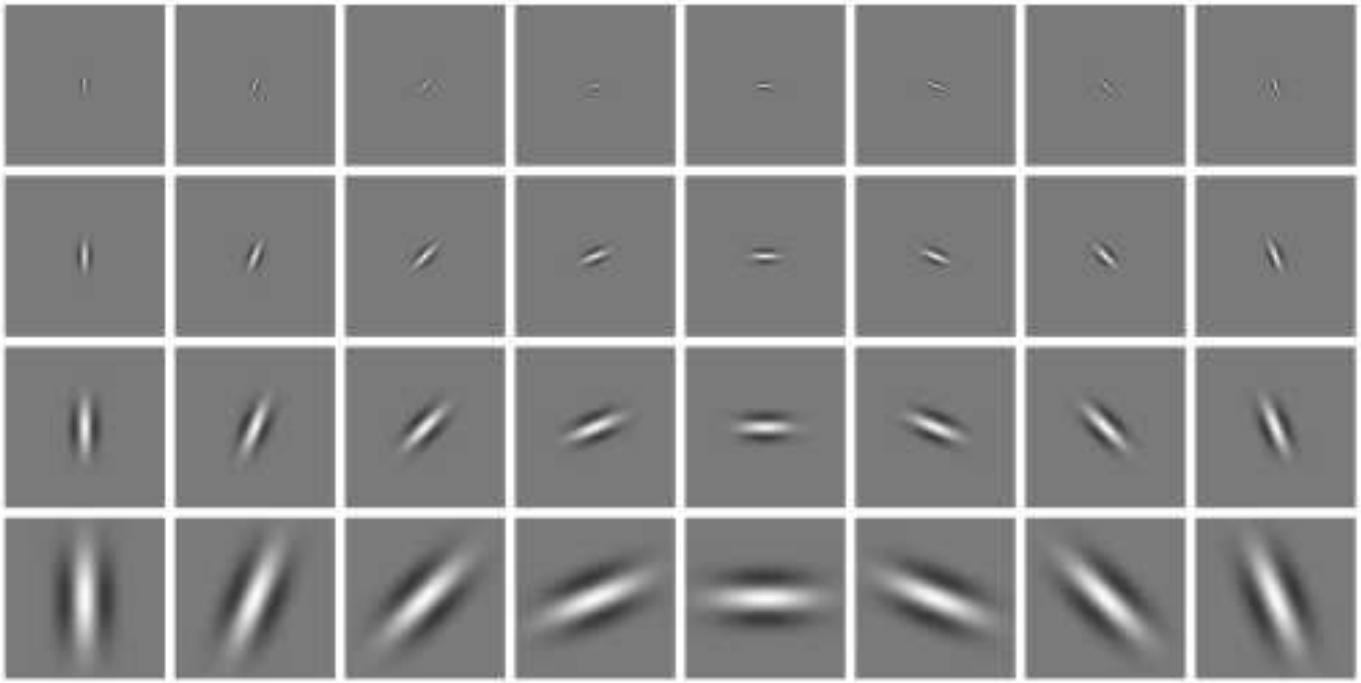
Consider a discrete family of re-scaled and rotated wavelets

$$\psi_{(j,\theta)} = \psi_{(j,\theta)}(x) := 2^{-jd}\psi(2^{-j}r_{-\theta}x),$$

with the scale parameter $j \in \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ and the rotation parameter $\theta \in [0, 2\pi)$; ($r_\theta x$ denotes the rotation of $x \in \mathbb{R}^2$ by the angle θ with respect to the origin).



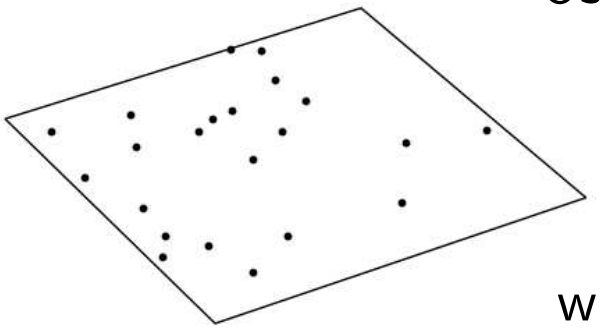
$\psi \longmapsto \psi_{(j,\theta)}$



Signal

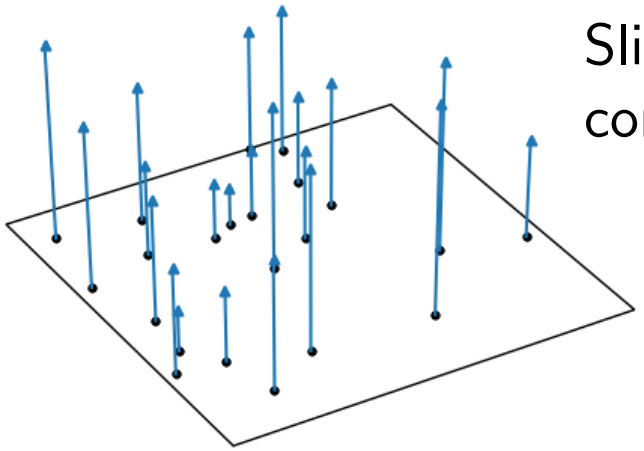
In this talk we are mostly interested in purely atomic signals $\boldsymbol{\Phi}$.

Using the formalism of point processes



$$\boldsymbol{\Phi} = \sum_{i} \boldsymbol{\delta}_{\boldsymbol{x}_i}$$

where $\boldsymbol{\delta}_{\boldsymbol{x}}$ is the Dirac measure describing the unit atom at $\boldsymbol{x}$.



Slightly more generally, atoms (points) can have marks $\boldsymbol{M_i}$ considered as "weights"

$$\tilde{\boldsymbol{\Phi}} = \sum_{i} \boldsymbol{M_i} \boldsymbol{\delta}_{\boldsymbol{X}_i}.$$

Slightly more general, signal is modeled as a (possibly signed) measure $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\mathrm{d}\boldsymbol{x})$ on $\mathbb{R}^{\boldsymbol{d}}$.

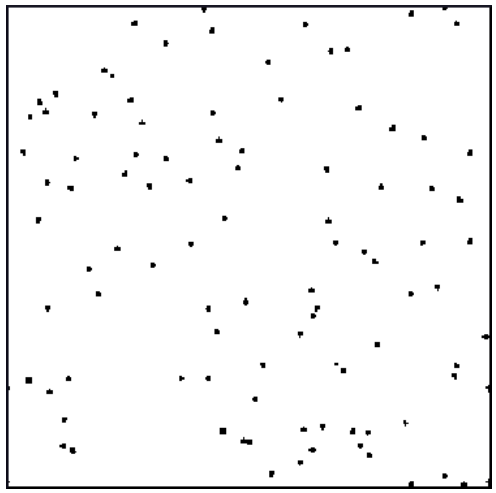
Wavelet transform of the signal

Wavelet transform of (a realization of) $\boldsymbol{\Lambda}$ at scale $\boldsymbol{2^j}$ and angle $\boldsymbol{\theta}$, is a (random) filed on $\mathbb{R}^{\boldsymbol{d}}$ defined as a convolution of $\boldsymbol{\Lambda}$ with the wavelet $\boldsymbol{\psi_{(j,\theta)}}$:

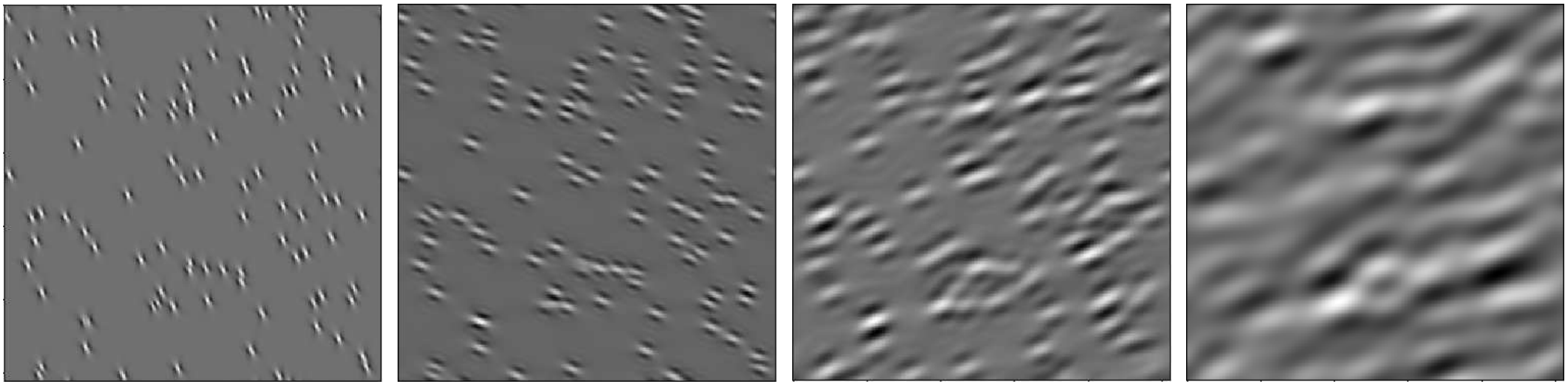
$$(\boldsymbol{\Lambda} \star \boldsymbol{\psi_{(j,\theta)}})(\boldsymbol{x}) := \int_{\mathbb{R}^{\boldsymbol{d}}} \boldsymbol{\psi_{(j,\theta)}}(\boldsymbol{x}-\boldsymbol{y})\, \boldsymbol{\Lambda}(\mathrm{d}\boldsymbol{y})\,.$$

Observe: The zero average property of the mother wavelet $\int_{\mathbb{R}^{\boldsymbol{d}}} \boldsymbol{\psi}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \boldsymbol{0}$ implies that the wavelet transform $\boldsymbol{\Lambda} \star \boldsymbol{\psi_{(j,\theta)}}(\boldsymbol{x})$ at the scale $\boldsymbol{j}$ has larger absolute values for $\boldsymbol{x}$ where the $\boldsymbol{\Lambda}$ is has more variability at this given scale. It (almost) vanishes where $\boldsymbol{\Lambda}$ is (almost) uniform at this scale.

Wavelet transforms of a point pattern



signal



signal wavelet transforms at different scales

Scattering moments: introducing non-linearity and averaging

- Define the scattering fields as the modulus of the (complex valued) wavelet transforms for $j \in \mathbb{Z}$, $\theta \in [0, 2\pi)$

$$S_{j,\theta}\tilde{\Phi}(x) := |\tilde{\Phi} \star \psi_{(j,\theta)}(x)|, \quad x \in \mathbb{R}^d.$$

- Define (empirical) scattering moments as the averages of the scattering fileds over x in a given observation window W

$$\hat{S}\tilde{\Phi}(j,\theta) := \frac{1}{|W|} \int_W S_{j,\theta}\tilde{\Phi}(x)\, \mathrm{d}x.$$

- In practice, the scale parameter is restricted to a finite window $j \in [j_{\min}, j_{\max}]$ such that the support of $\psi_{(j_{\min},\theta)}$ "separates points" and this of $\psi_{(j_{\max},\theta)}$ covers the whole window. Some discrete set of angles $\theta_1, \ldots, \theta_{\max} \in [0, 2\pi)$ is considered.

Scattering moments — invariance properties

- Scattering moments provide some information about the given signal pattern (data).
- They are Lipschitz-continuous with respect smooth signal deformations.
- They are (can be made) invariant with respect to some "distributional symmetries" of the signal. If the signal can be assumed to have the same symmetry properties then it can be captured, up to these symmetries, in a more concise way.

Assuming stationary model

- Assume some underlying stationary point process model for the signal $\tilde{\mathbf{\Phi}}$ on $\mathbb{R}^d$. That is, the observed realization of the signal (data) is assumed to be sampled from the assumed distribution and restricted to the observation window.
- One considers also mathematical scattering moments as the expected values of the scattering fields. By the stationairty of the model, one can calculate them, without loss of generality, at the origin $\mathbf{0} \in \mathbb{R}^d$

$$\bar{S}\tilde{\mathbf{\Phi}}(j,\theta) := \mathsf{E}[S_{j,\theta}\tilde{\mathbf{\Phi}}(\mathbf{0})].$$

- If moreover the point process $\tilde{\mathbf{\Phi}}$ model is assumed isotropic (rotation-invariant distribution) the mathematial scattering moments do not depend on the angle: $\bar{S}\tilde{\mathbf{\Phi}}(j,\theta) = \bar{S}\tilde{\mathbf{\Phi}}(j)$.

Properties of the mathematical scattering moments

- Mathematical scattering moments are characteristics of the assumed signal distribution (mathematical model).
- The non-linearity produced by the modulus $|\cdot|$ make them (a priori) depend on all correlation functions (which would not be the case if the square $|\cdot|^2$ of the norm is taken).
- Open question: to what extend mathematical scattering moments characterize the correlation functions of $\tilde{\Phi}$ (and thus its distribution)? Asymptotic results when $j \to -\infty$ and $j \to \infty$ (at small and large scales) are known. Small scales characterize the intensity (density) of the point process, large scale characterize some long-range variance properties (hyper-uniformity, hyper-fluctuations) $\Rightarrow$ PhD thesis A. Brocahrd (2022)

Stationarity, isotropy — first invariance properties

- Scattering moments are invariant with respect to all periodic translations of the signal in the square window. Consequently a stationary signal can captured with smaller description dimension (no need to describe the position the signal with respect to, say, the center of the window).
- If an isotropic model of the signal can be assumed (rotation invariant distribution), then averaging the scattering moments over different angles one can further reduce the dimension of the signal descriptor.

Ergodicity — more invariance properties

- Recall: $\tilde{\Phi}$ is ergodic when the spatial averaging of one sample over a large window is close to the mathematical expectation. ("One sees all local patters on one big sample.")
- In this case, for small scale j with respect to the observation window W

$$\hat{S}\tilde{\Phi}(j,\theta) := \frac{1}{|W|}\int_W S_{j,\theta}\tilde{\Phi}(x)\,\mathrm{d} \approx \mathrm{E}[S_{j,\theta}\tilde{\Phi}(0)] =: \bar{S}\tilde{\Phi}(j,\theta).$$

  Small-scale scattering moments are close to the mathematical constants and thus invariant with respect to different realizations coming from the same distribution. $\Rightarrow$ Can be used to construct generative models of point patterns. (in this talk)

- Larger scales carry additional information about the given data pattern up to its exact characterization. $\Rightarrow$ All scattering moment can be used to statistical learning of signal. (Brochard, A., BB, Mallat, S. and Zhang, S. 2019).

Refinement I: Higher order scattering moments

The second order scattering transforms of $\tilde{\mathbf{\Phi}}$ at scales $2^{j_1}, 2^{j_2}$ and angles θ_1, θ_2 are defined as the first order scattering moments at scale 2^{j_2} and angle θ_2 of the first order scattering transform $S_{j_2,\theta_2}\tilde{\mathbf{\Phi}}(x)$ taken as the signal.

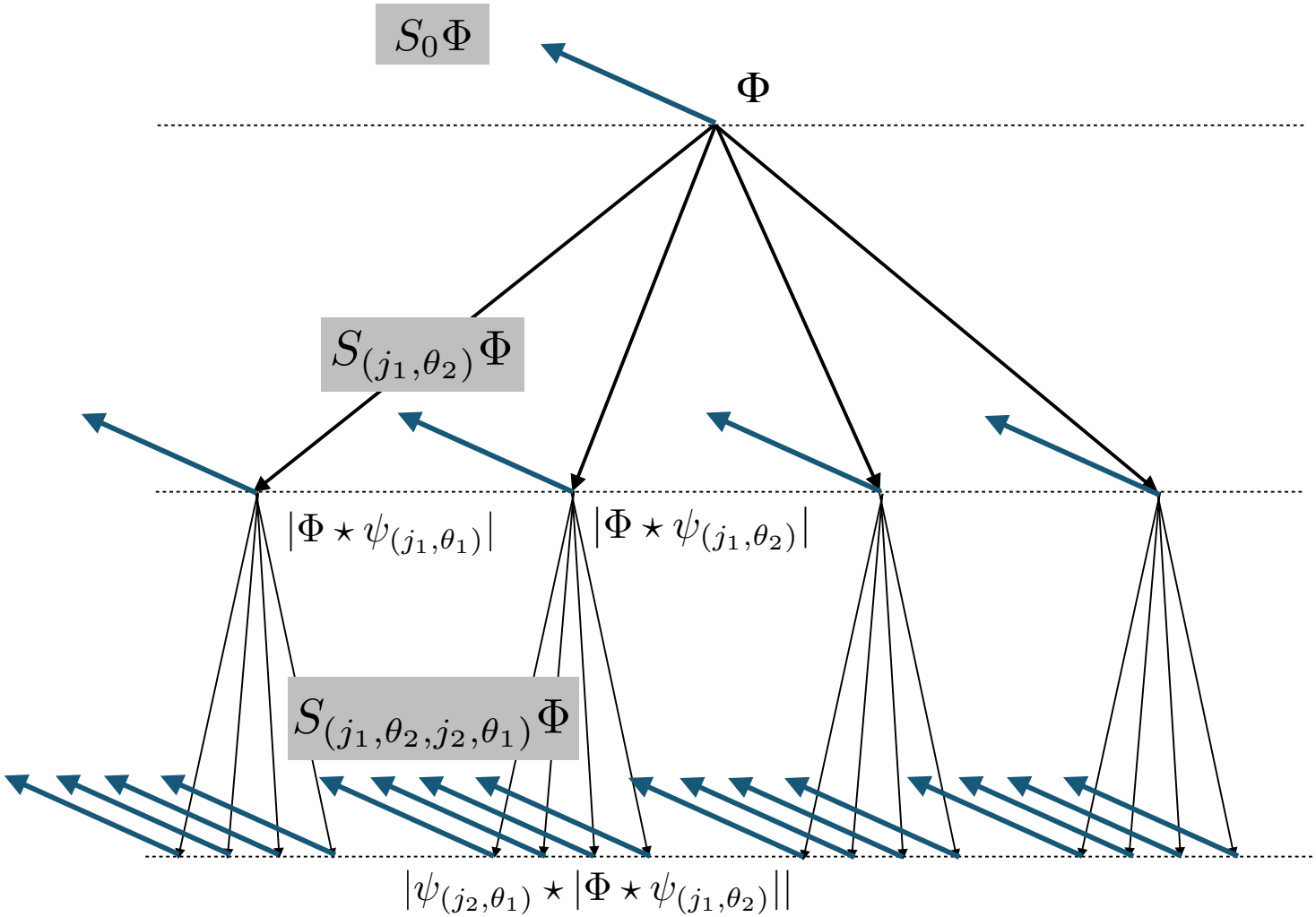
$$S_{j_1,\theta_1,j_2,\theta_2}\tilde{\mathbf{\Phi}}(x) := ||\tilde{\mathbf{\Phi}} \star \psi_{(j_1,\theta_1)}| \star \psi_{(j_2,\theta_2)}(x))| \,.$$

By empirical averaging over the observation window (or taking mathematical expectation) one defines the corresponding second order empirical (or mathematical) scattering moments.

$$\hat{S}\tilde{\mathbf{\Phi}}(j_1,\theta_1,j_2,\theta_2) := \frac{1}{|W|}\int_W S_{j_2,\theta_2}\tilde{\mathbf{\Phi}}(x)\,\mathrm{d}x,$$
$$\bar{S}\tilde{\mathbf{\Phi}}(j_1,\theta_1,j_2,\theta_2) := \mathsf{E}[S_{j_2,\theta_2}\tilde{\mathbf{\Phi}}(0)].$$

Higher order scattering moments are defined similarly by the induction.

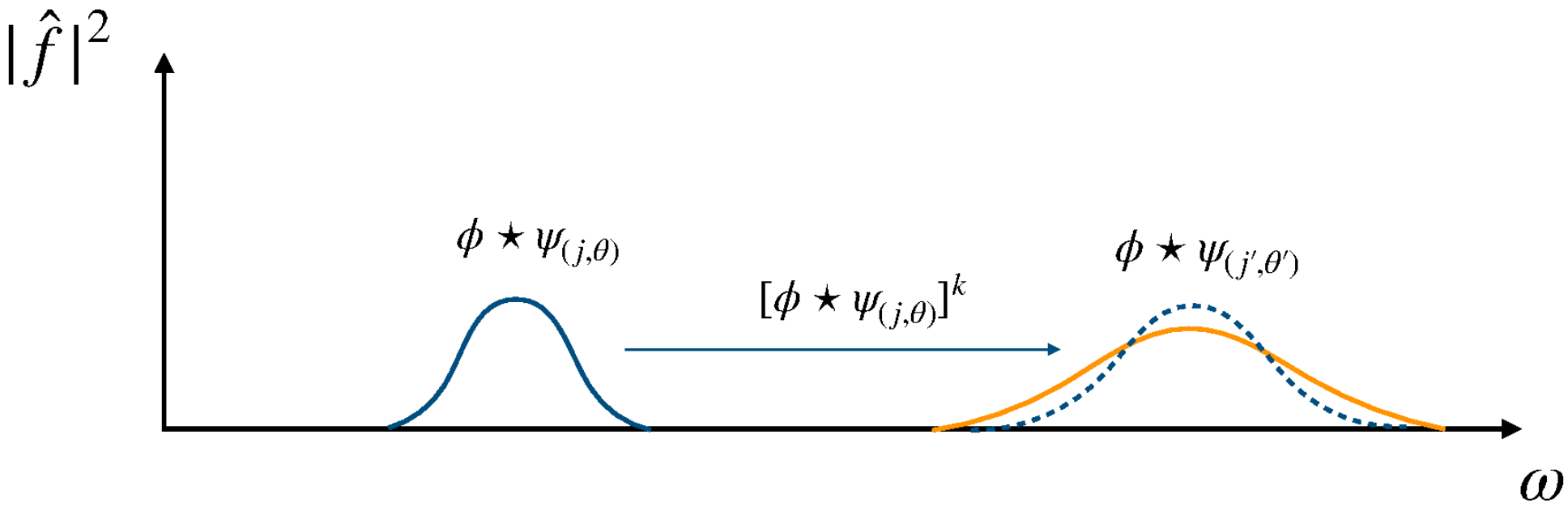
Cascade of higher order scattering moments



Refinement II: wavelet phase harmonics

Goal: capture correlations between scales and orientations.
Phase acceleration — refined non-linearity: for all $z \in \mathbb{C}$, $k \in \mathbb{Z}$

$$[z]^k := |z| e^{ik\varphi(z)}, \quad \text{where } \varphi(z) \text{ is the complex argument of } z.$$



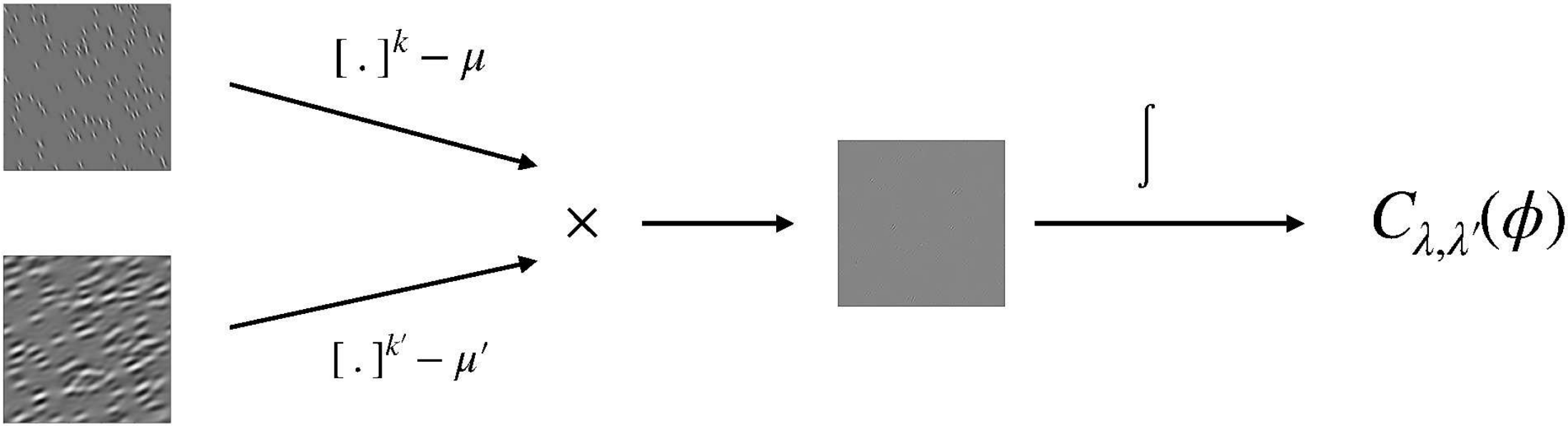
phase acceleration allowing for non-null correlation the scattering transforms

$$j, \theta, j', \theta' :$$

Refinement II: wavelet phase harmonics

Covariance between wavelet transforms at different scales and orientations:
For $\lambda := (j, \theta)$, $\mu_{\lambda,k} := \int_W [\tilde{\Phi} \star \psi_\lambda]^k(x)\, \mathrm{d}x$.
Similarly $\mu_{\lambda',k'}$ $\lambda' := (j', \theta')$. Wavelet phase harmonics

$$C_{\lambda,\lambda'}(\Phi) := \int_W \left([\tilde{\Phi} \star \psi_\lambda]^k(x) - \mu_{\lambda,k}\right) \left([\tilde{\Phi} \star \psi_{\lambda'}]^{k'}(x) - \mu_{\lambda',k'}\right)^* \mathrm{d}x.$$

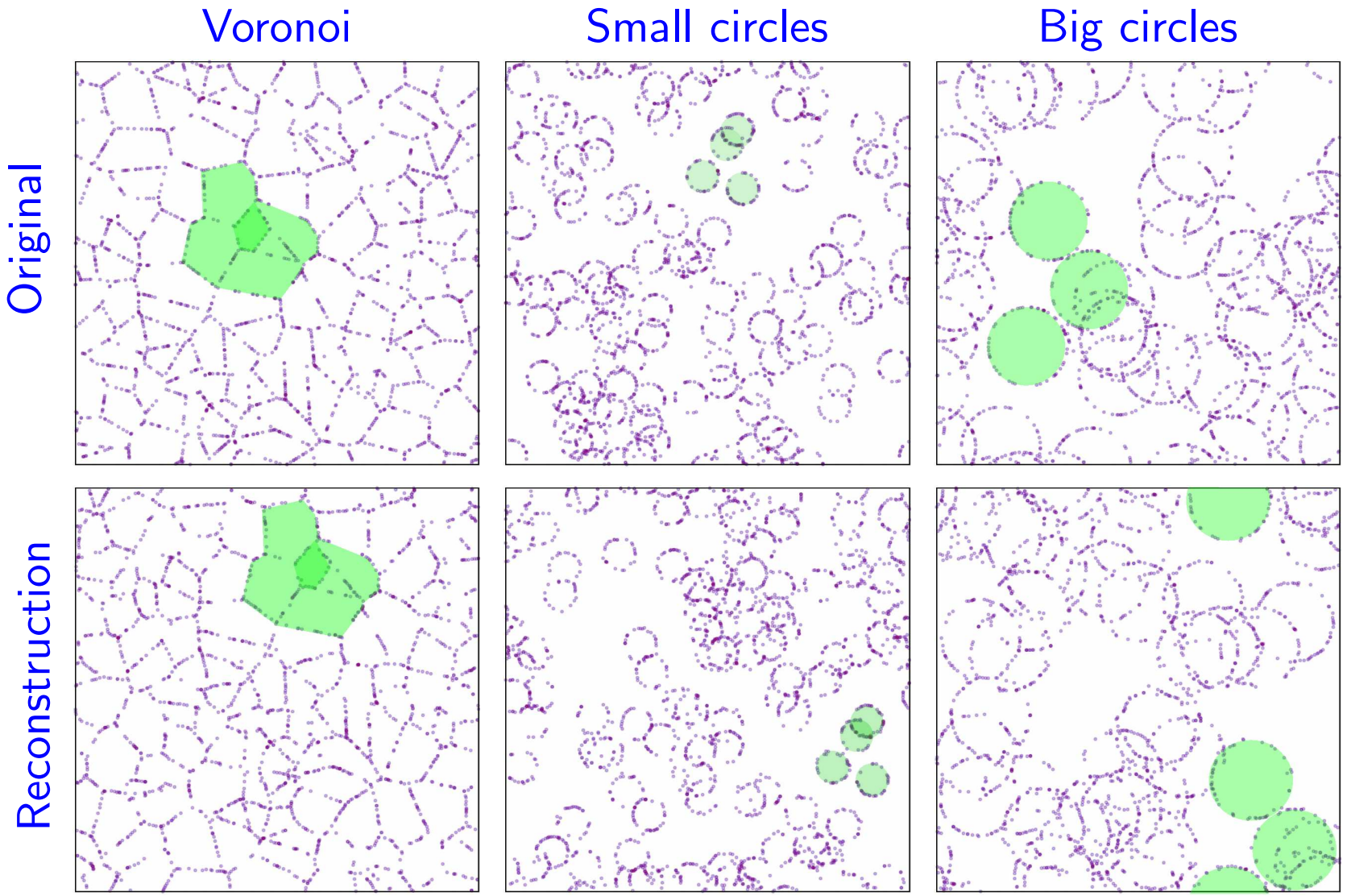


- TESTING RESULTS
  - Visual evaluation

Main parameters

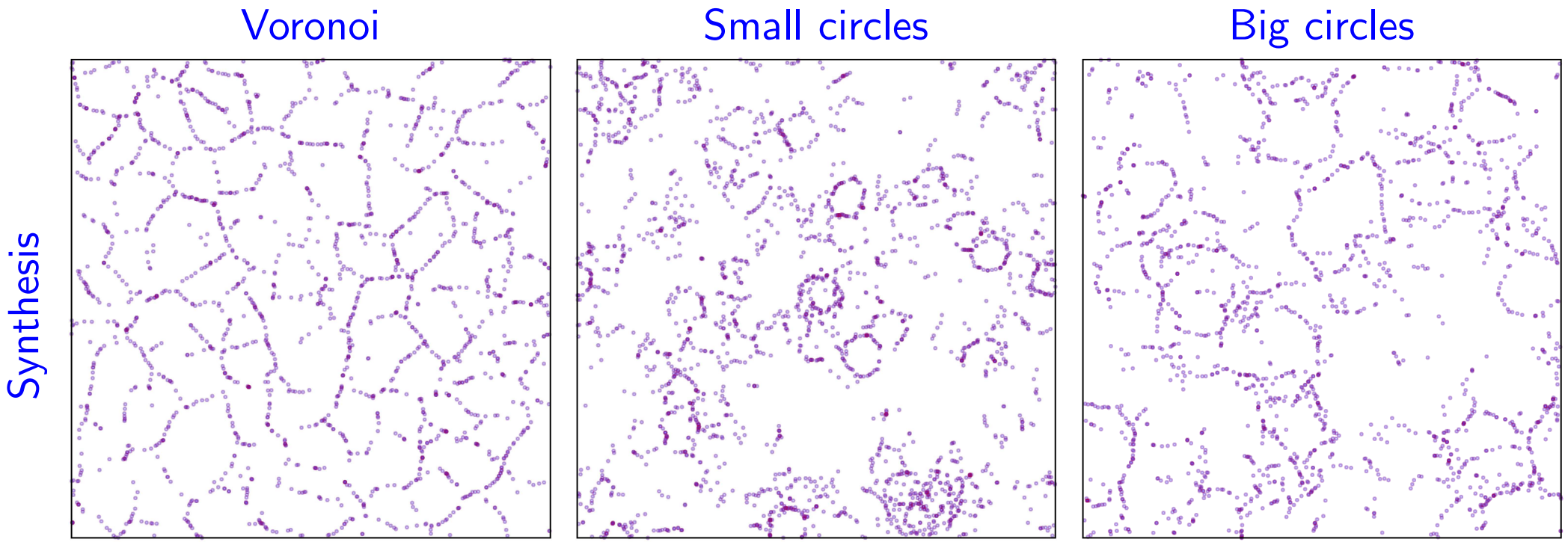
- Number of points in the given image: from $\mathbf{1900}$ up to $\mathbf{13000}$ points.
- Statistics $\boldsymbol{K}$ (descriptors) Wavelet phase harmonics based on bump steerable wavelets (Mallat et al. 2020).
- Scales $\mathbf{0} \leq j < \boldsymbol{J} = \boldsymbol{J}(\boldsymbol{N})$; where
  $\boldsymbol{N}$ = size of observation window/size of pixel;
  We take $\boldsymbol{N} = \mathbf{128}$ and $\boldsymbol{N} = \mathbf{256}$.
  $\boldsymbol{J} = \mathbf{log_2}(\boldsymbol{N}) - \mathbf{2}$ — "memorizing image",
  $\boldsymbol{J} = \mathbf{log_2}(\boldsymbol{N}) - \mathbf{3}$ — good "learning distribution".
- Number of statistics (dimension of $\boldsymbol{K}$)
  $\boldsymbol{O}(\text{number of angles}^{\mathbf{2}} \times \text{number of scales}^{\mathbf{2}}) \simeq \mathbf{3\ 000 - 5\ 000}$. (We take number of angles 8).
- Number of iterations (L-BFGS optimization) from $\mathbf{400}$ to $\mathbf{500}$.

"Memorizing" samples modulo random translation



Statistics $\boldsymbol{K}$ based on wavelets of all scales, up to the size of the window.

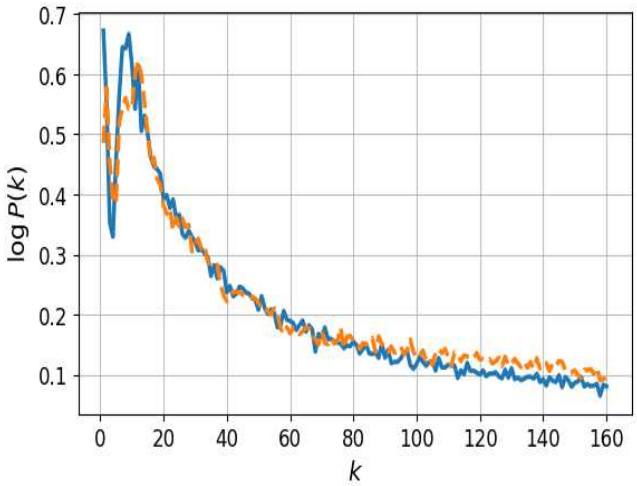
Samples from our random when removing too large scales



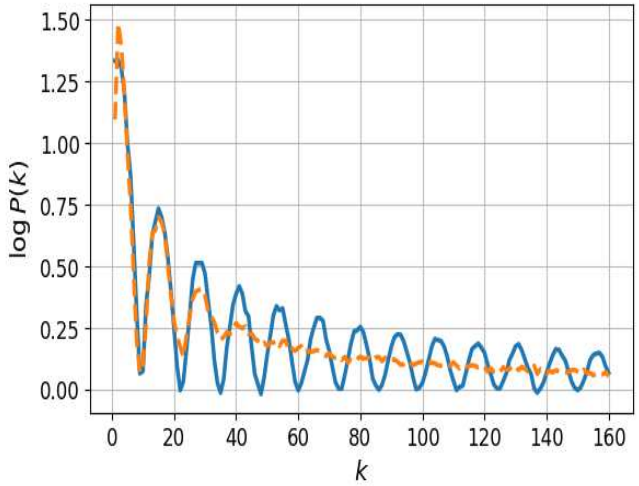
Spectrum comparison

Power spectrum

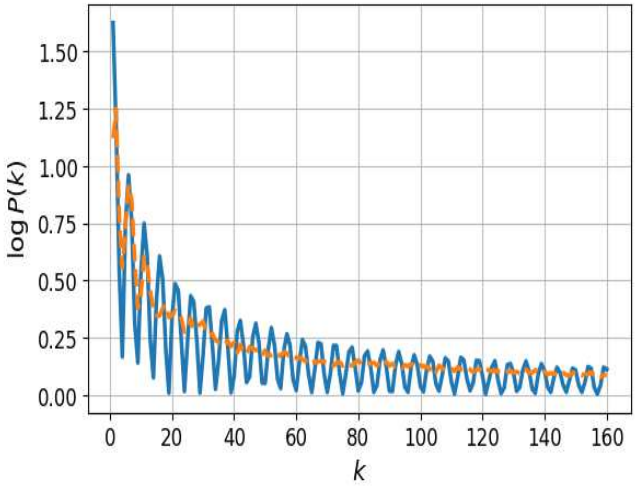
Voronoi



Small circles



Big circles

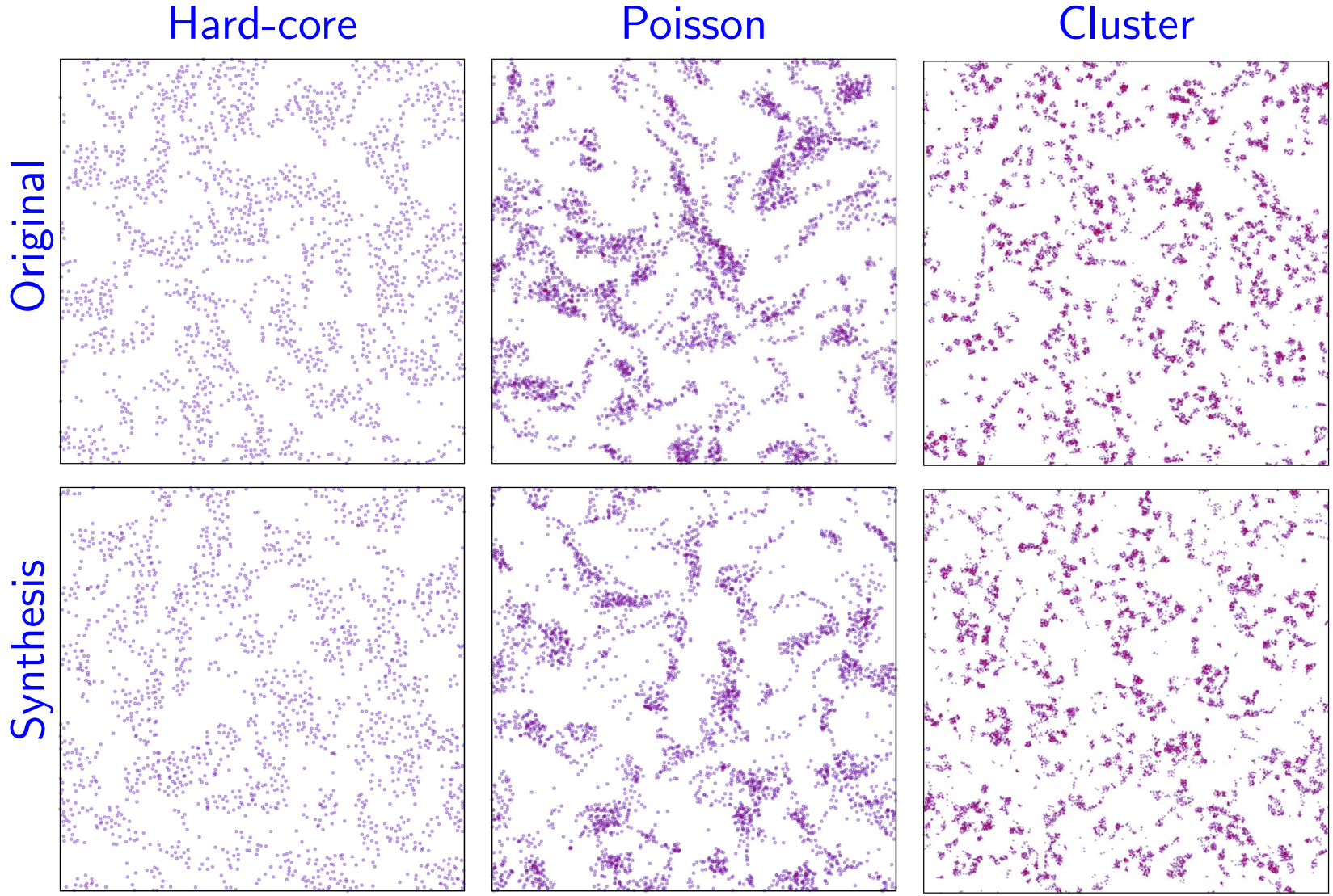


Discrete Fourier transform (DFT) of a counting measure $\mu = \sum_u \delta_{x_u} \in \mathbb{M}$ on the (square) window $W_s = [-s, s[^2$ at integer frequency $m \in \mathbb{Z}^2$

$$F_m(\mu) := \int_{W_s} e^{-i\pi m x/s}\, \mu(dx) = \sum_u e^{-i\pi m x_u/s}.$$

This is an asymptotically (when $s \to \infty$) non-biased estimator of density of the Bartlett spectrum of point processes.

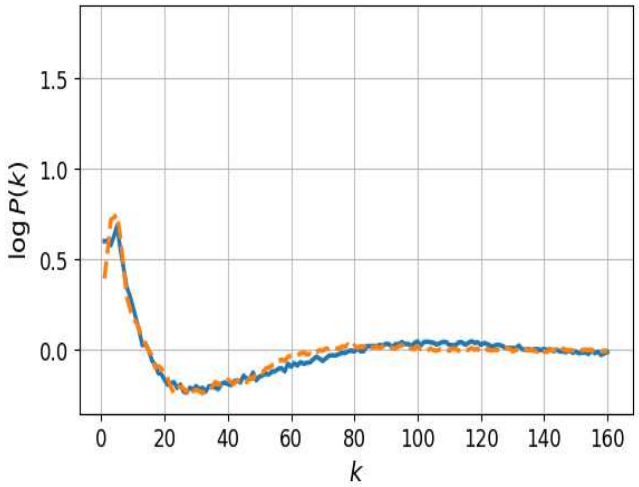
Originals and samples from model for turbulence processes



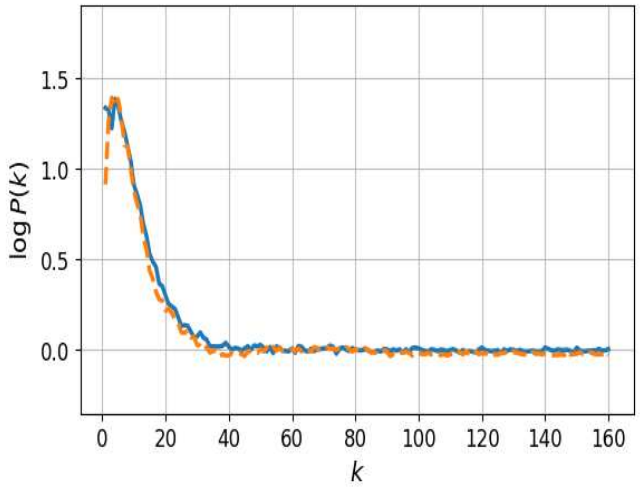
Turbulence models — spectrum comparison

Power spectrum

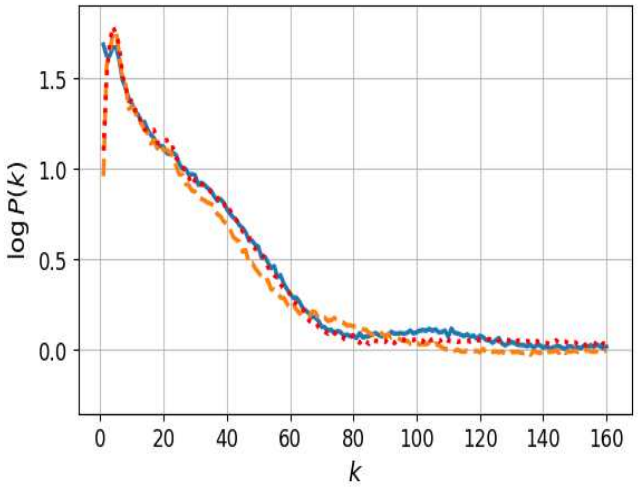
Hard-core



Poisson



Cluster

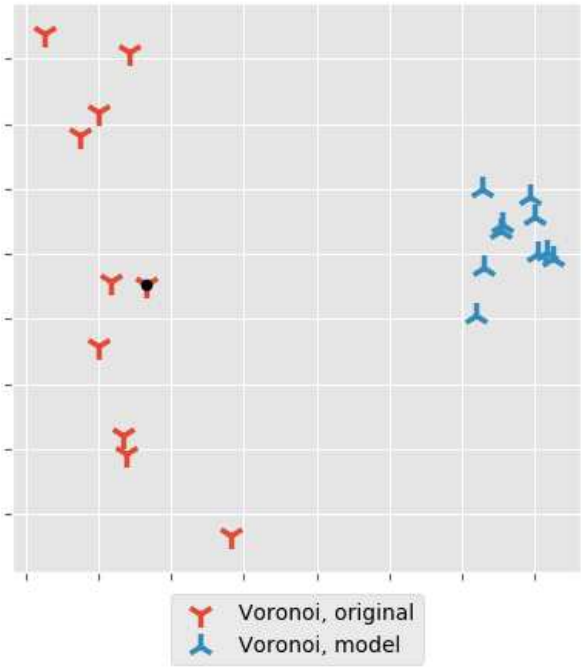


Persistent homology (in topology data analysis, TDA)

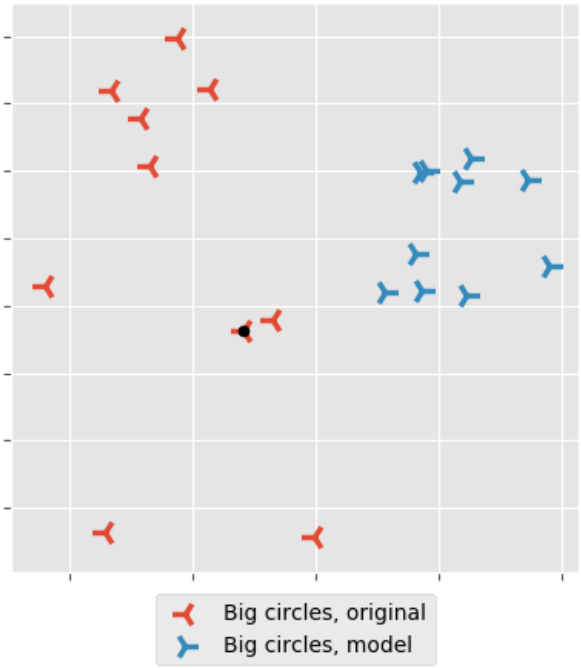
- Visual evaluation can be more discriminate, but is subjective. We need a tool to capture the geometric structures $\Rightarrow$ persistence diagrams (Boissonnat, Chazal, Yvinec M 2018).
  - For all radius $r > 0$ on construct the Gilbert graph connecting points of μ closer to each other than r.
  - "Fill-in" the triangles (triplets of points joined by edges) $\Rightarrow$ 2-skeleton of the Vietoris-Rips (VR) complex on μ.
  - Observe holes formed when radius r grows from $r = 0$: each hole has a birth radius $r > 0$) and a (larger) death radius (when completely filled-in by the triangles).
- "Our" persistence diagram of μ is the collection of pairs of radii: (birth,death) of holes $\Rightarrow$ "diagram points" in positive orthant on $\mathbb{R}^2$.
- For two patterns μ_1, μ_2 we calculate their Wasserstein distances between their corresponding persistent diagrams;`TDAstats` and/or `TDA` soft.
- For many patterns μ_i, represent every persistent diagram as a "dot" on the plane (using standard Multi Dimensional Scaling).

Generative model vs original distribution via TDA

Cox Voronoi



Cox Big circles

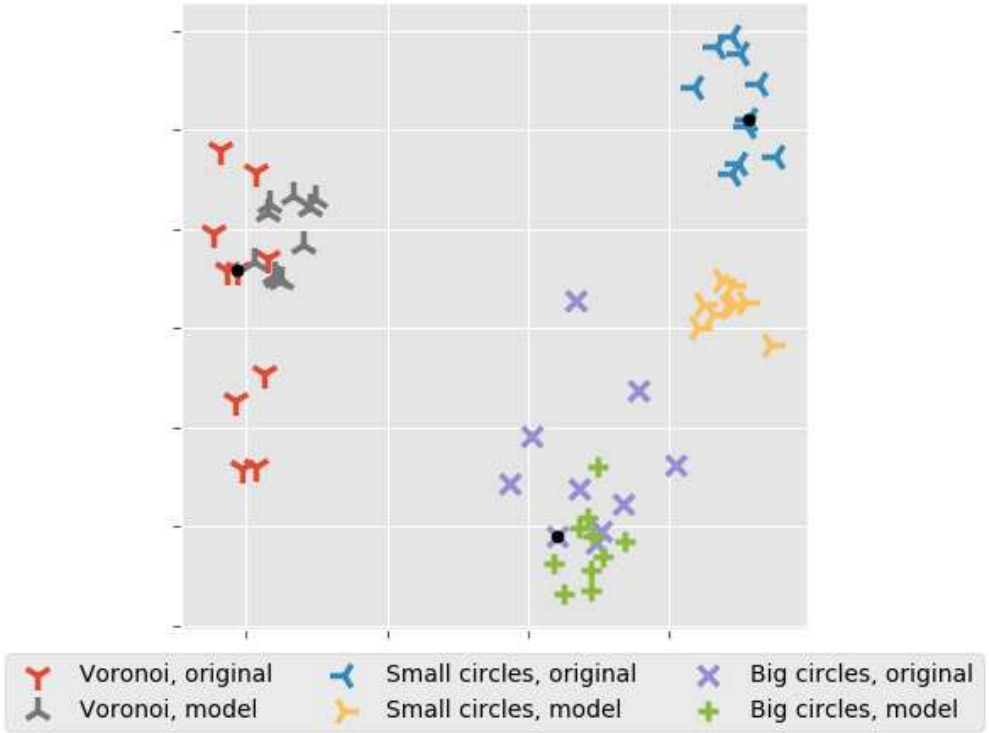


Turbulence hardcore

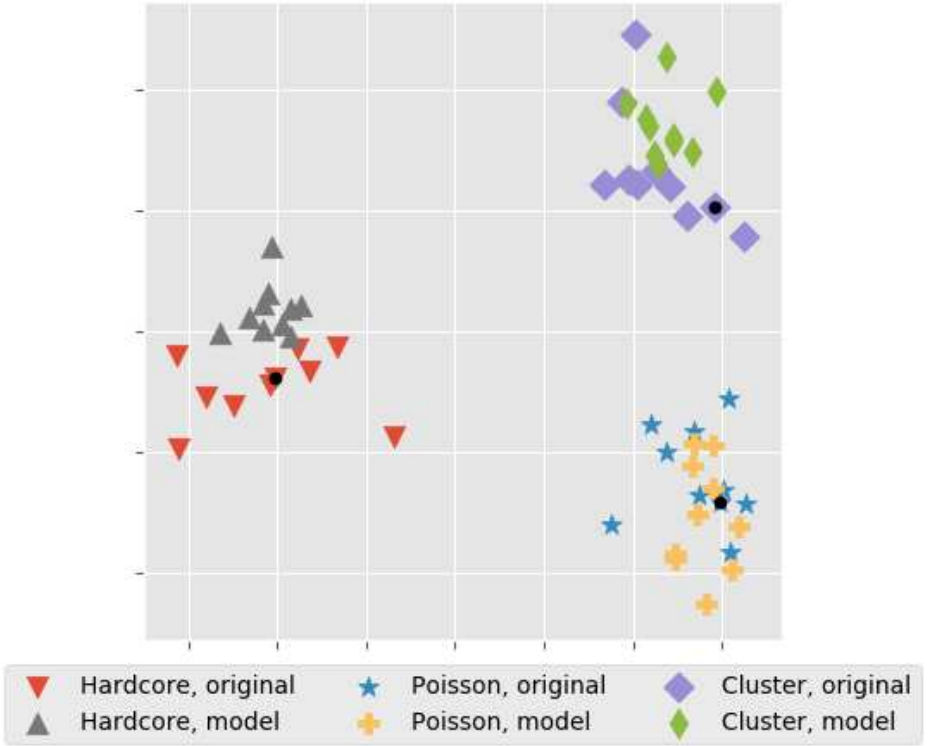
For each model there are 10 "dots" representing (via TDA analysis) i.i.d. realizations of the original distribution and 10 "dots" representing i.i.d. realizations from the generative model estimated on one of the original realizations (marked by the black dot).

Generative model vs original distribution via TDA

Cox — three distributions



Turbulence — three distributions



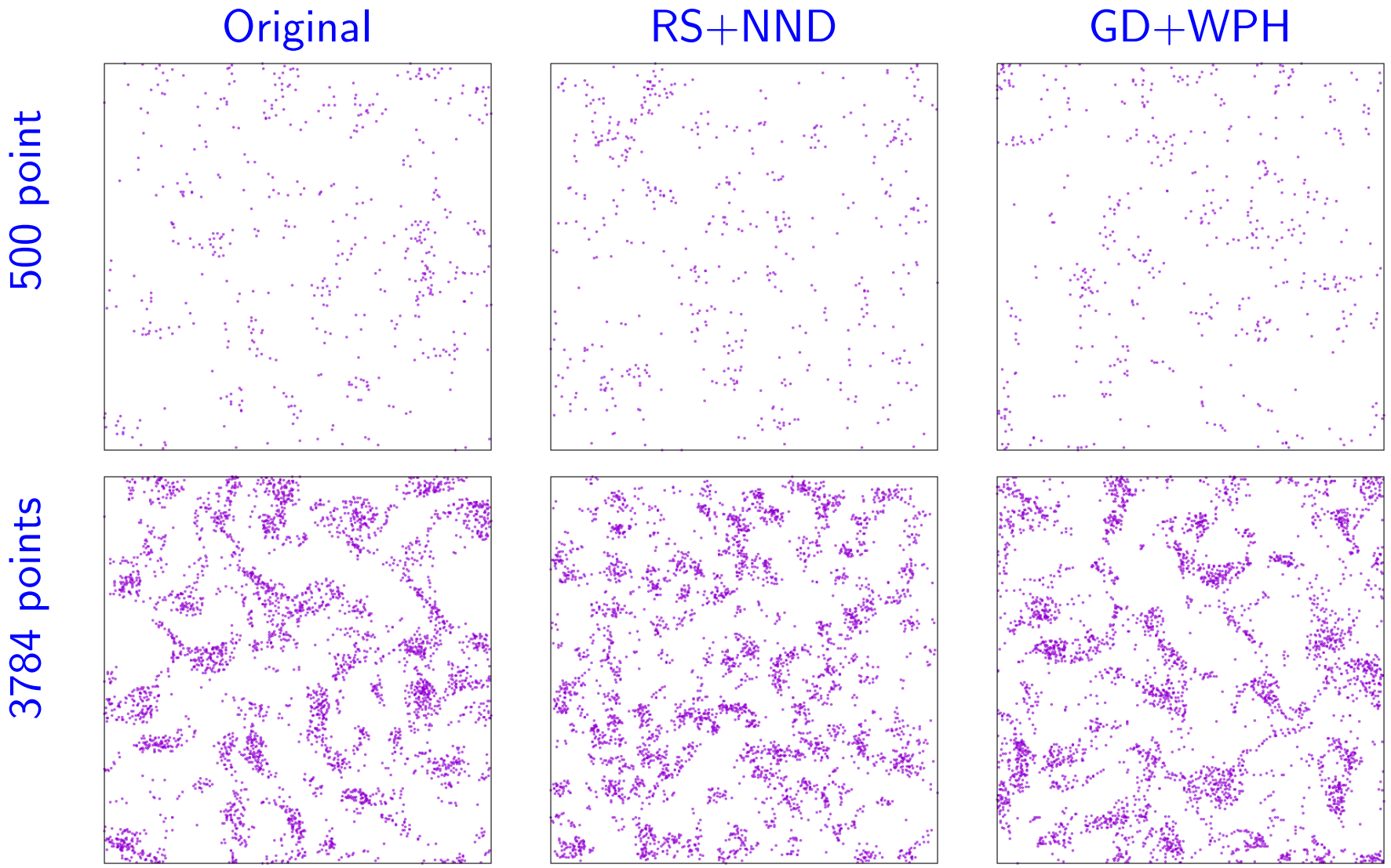
Random search (RS) vs Gradient descent (GD) model

Time computing comparison

Energy error	Random search	Gradient descent
$e = \mathbf{9,00.10^{-4}}$	19870 (1h04m)	52 (0m35s)
$e = \mathbf{4,76.10^{-4}}$	29805 (1h36m)	69 (0m45s)

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.

RS+NND vs GD+WPH models — visual comparison



RS+NND Random search with nearest neighbor distance; Tscheschel and Stoyan (2006)

WPH+GD Wavelet phase harmonics with gradient descent.

For more details, see:

- Brochard, A., BB, Mallat, S. and Zhang, S. (2022). Particle gradient descent model for point process generation. Statistics and Computing **32**, 1-25, https://arxiv.org/abs/2010.14928
- Brochard, A. Wavelet-based representations of point processes for modelling and statistical learning PhD thesis, (2022) https://tel.archives-ouvertes.fr/tel-03666508

Thank you!