Ergodic learning and particle gradient descent generative model for point processes



Antoine Brochard
Inria/ENS

Bartek Błaszczyszyn
Inria/ENS Paris

Sixin Zhang
University of Toulouse/IRIT

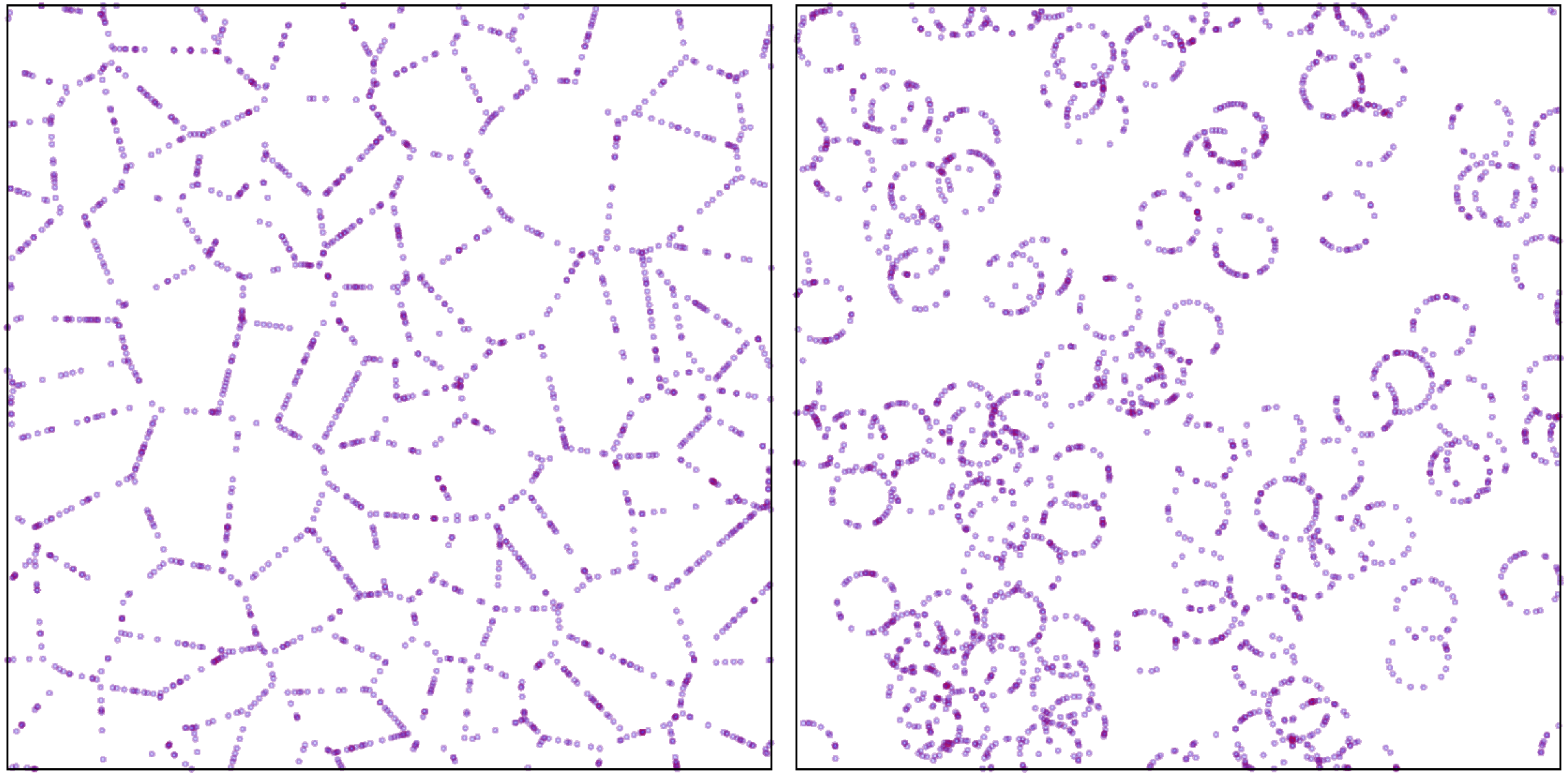
Stephane Mallat
College de France/ENS, Paris

Anzère, École d'été CUSO, September 1-4, 2024

Problem:

Learning a generative model

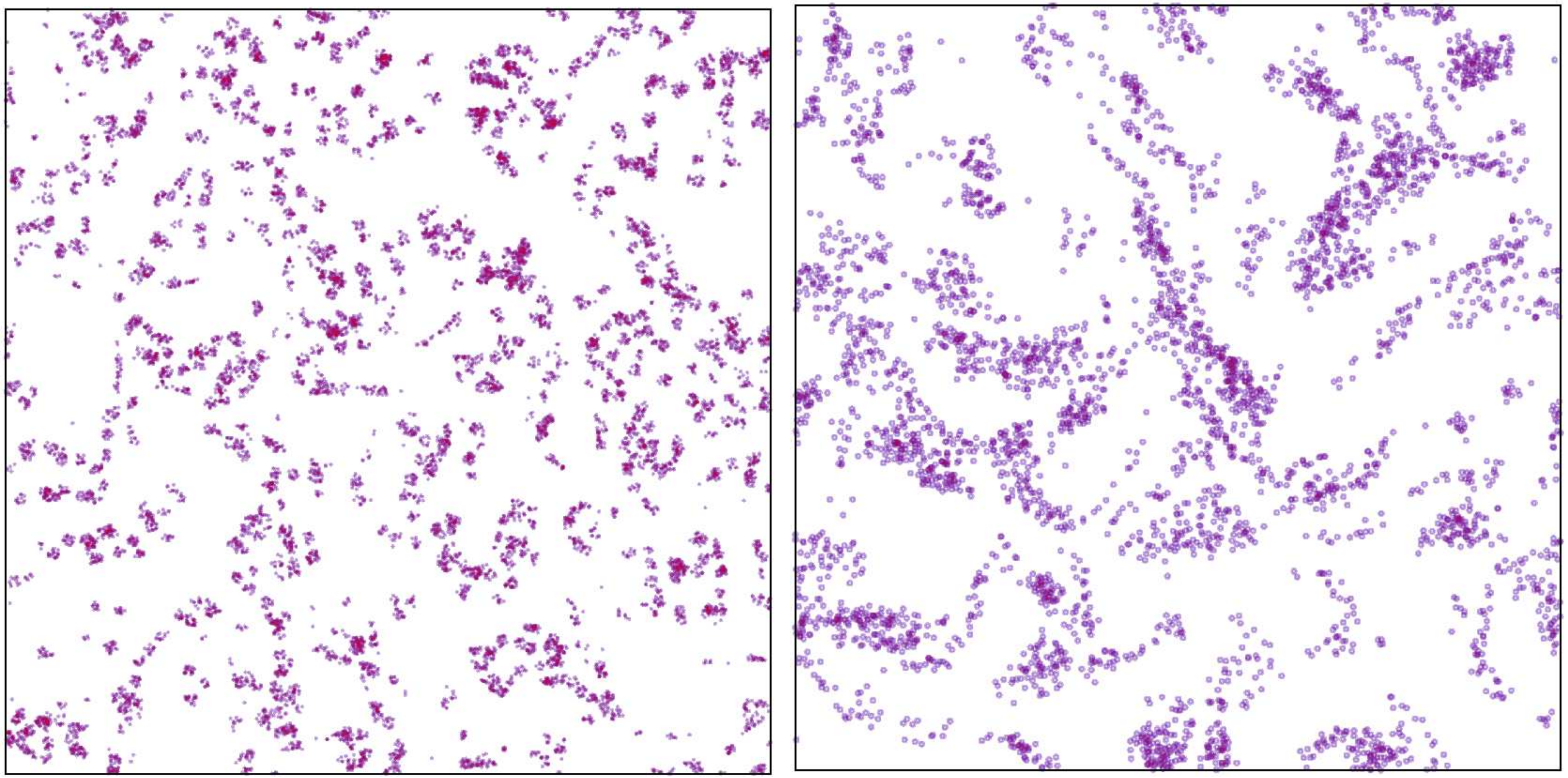
Samples from a point process. Can you recognize the model?



Sure, Cox-Voronoi and Cox-Boolean.

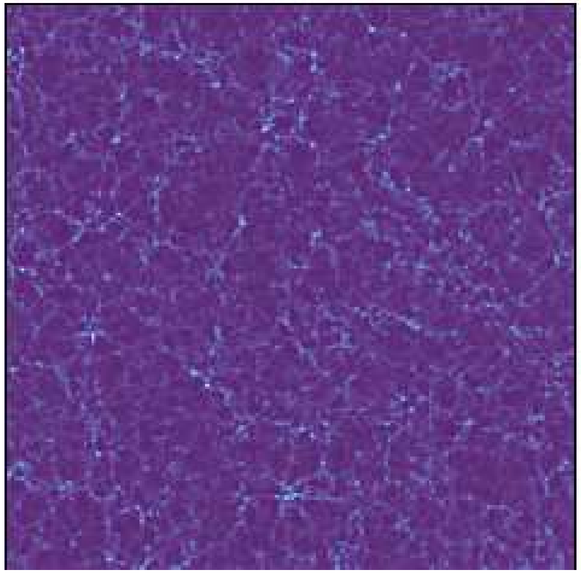
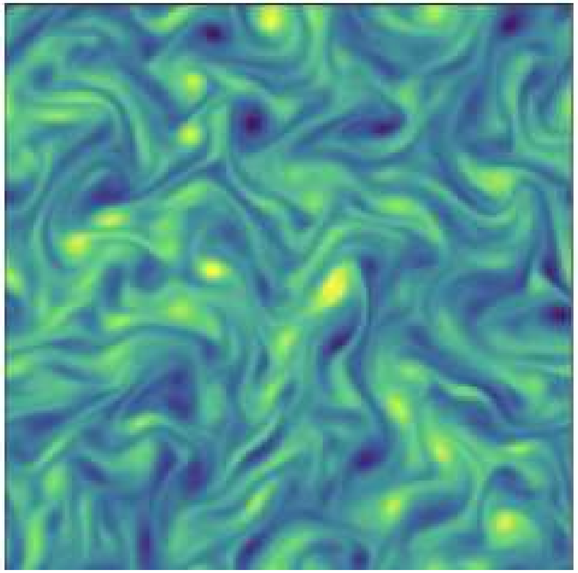
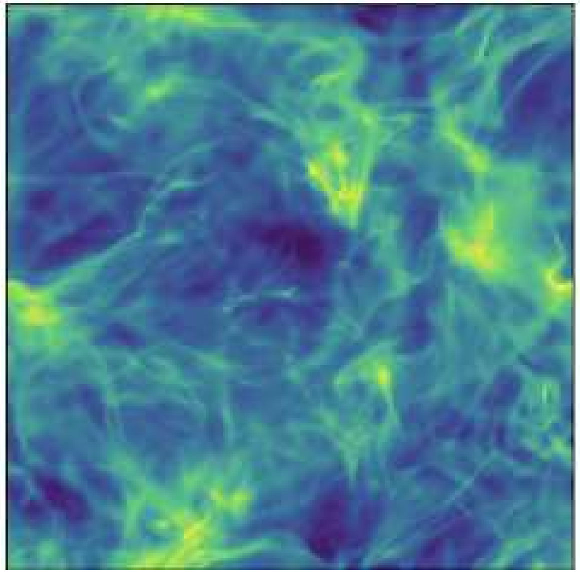
Recall: Cox $=$ doubly stochastic Poisson process.

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, ...

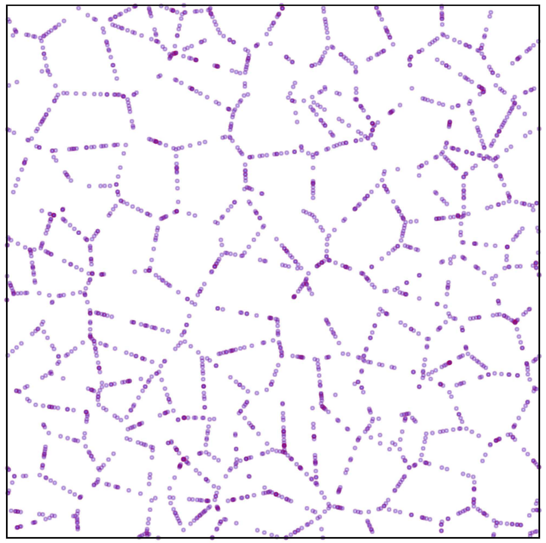
These patterns:

- 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?

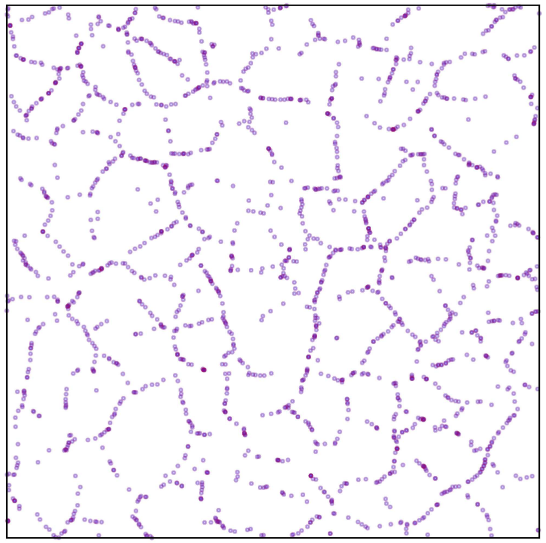
Ergodic learning of point processes

- 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!

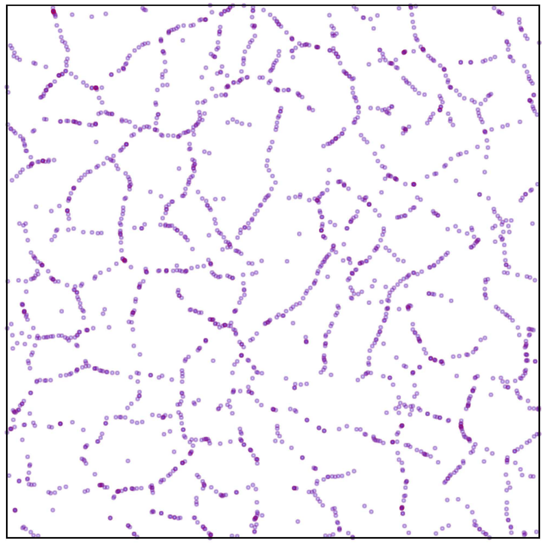
- But in practice, we have only a finite learning window. Can we get approximations of the unknown distribution?



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 (Illian, Penttinen, Stoyan, Stoyan 2008)
  - Wavelet-based representations (Mallat 2001)$\Leftarrow$
- Testing results
  - Visual,
  - Spectrum,
  - Topology analysis (persistent homology) $\Leftarrow$

- MODELS
  - Maximum entropy models

Maximum entropy models

- Based on a set of statistics to be imposed:
  - on average $\Rightarrow$ macro–canonical model.
  - path-wise $\Rightarrow$ micro-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

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

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

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

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

Macro-canonical model

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

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

- Model:

$$\underset{\hat{\boldsymbol{\mathcal{L}}}}{\arg\max} \quad \boldsymbol{\mathcal{H}}(\hat{\boldsymbol{\mathcal{L}}})$$

$$\text{given} \quad \text{constraints (AC)}$$

- 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 $\boldsymbol{K}$ 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:

$$\begin{aligned} \operatorname*{arg\,max}_{\hat{\mathcal{L}}} \quad & \mathcal{H}(\hat{\mathcal{L}}) \\ \text{given} \quad & \hat{\mathcal{L}}(\Omega_\epsilon) = 1 \end{aligned}$$

- 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: $E_{K(\Xi)}(\mu) := \frac{1}{2}|K(\mu) - K(\Xi)|^2$.
- Gradient descentin in the space of point measures [1]
  - Initialization: Generate initial homogeneous Poisson configuration of points Φ_0 (eventually conditioned on N points).
  - Gradient descent: Transport points of Φ_0 by iteratively minimizing the energy $E_{K(\Xi)}(\Phi_k)$, $k \geq 0$:
    For $\Phi_k := \sum_i \delta_{x_i^k}$ we take $\Phi_{k+1} := \sum_i \delta_{x_i^{k+1}}$ with

$$x_i^{k+1} := x_i^k - \gamma \frac{\partial E(\Phi_k)}{\partial x_i^k},$$

    where $\gamma > 0$ is some fixed gradient step.
  - Stop at some (fixed) step $k = n$.

[1] Molchanov and Zuyev (2002)

Before the remarks on the gradient descent model

Disclaimer:

Attendees of this conference with some mathematical sensitivity may feel offended by the remaining part of this talk.
The presenter is sorry for that.

A digression...

Was Babylonian mathematics as sophisticated as Greek mathematics?

History of Science and Mathematics on `hsm.stackexchange.com` :

"Greeks pay much more attention to the demonstrative side of mathematics. Computational mathematics was far more developed by Babylonians than by Greeks."

—Conifold

ChatGPT:

"Babylonian mathematics was highly advanced in practical arithmetic and algebra, while Greek mathematics was more sophisticated in developing abstract theory and formal proofs."

—

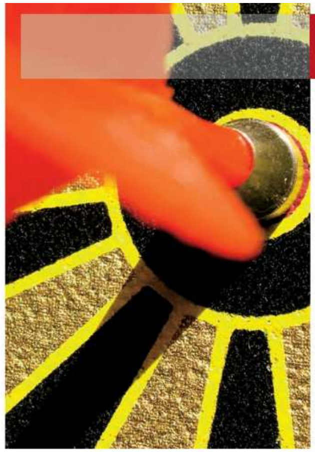
Back from Greece to Babylon?

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. XIII, 001–14 (1960)

The Unreasonable Effectiveness of Mathematics in the Natural Sciences

Richard Courant Lecture in Mathematical Sciences delivered at New York University, May 11, 1959

EUGENE P. WIGNER
Princeton University



EXPERT OPINION

1541-1672/09/$25.00 © 2009 IEEE
Published by the IEEE Computer Society
IEEE INTELLIGENT SYSTEMS

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

More in the talk "The Bridge from Mathematical to Digital, and Back" by David Donoho, Princeton University, on Mallat's 60th birthday conference, HIÉS, 2023.

I won’t monkeying to that extent

- Norvig and Co-authors point out/claim (*in talks, not paper*): *'Bayes Theorem is Empirically Wrong'*.
- They mean: when used *in a certain dataset,in a certain way*, it is **empirically outperformed** by a different rule! i.e.

not

$$P(B|A) \approx \frac{P(B)}{P(A)} \cdot P(A|B) \qquad \text{(BAYES)}$$

instead

$$P(B|A) \approx \frac{P(B)}{P(A)} \cdot P(A|B)^{1.5} \qquad \text{(NOT-BAYES)}$$

More in “Monkeying with Bayes’ theorem”, John D. Cook, 2012.

... and of the digression.

Remarks on the gradient descent model

- The (pathwise) gradient descent $\Phi_0 \rightsquigarrow \Phi_n$ defines a transport of the Poisson distribtuions $\mathcal{L}_0 := \mathcal{L}_{\Phi_0} \rightsquigarrow \mathcal{L}_{\Phi_n} =: \mathcal{L}_n$ to some point process disctribution $\mathcal{L}_n$.
- It looks like a substitute for Langevin dynamics... (?)
- The max-entropy postulate is supposed to be achieved by the Poisson initialization $\mathcal{L}_0$. However, no theoretical guarantee.
- For the approximation $\mathcal{L}_n \simeq \mathcal{L}(\Xi)$ we have rigid motions invarianace:

  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.

- 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) $\boldsymbol{\Xi}$, given (a vector of) statistics $\boldsymbol{K}$ of $\boldsymbol{\Xi}$, the "energy" of (an arbitrary) realization $\boldsymbol{\mu}$: $\boldsymbol{E_{K(\Xi)}(\mu) := \frac{1}{2}|K(\mu) - K(\Xi)|^2}$.

- Random serach on point measures
  - Initialization: Take arbitrary configuration of points $\boldsymbol{\Phi_0}$.
  - Radnom acceptence-rejection procedure to transport points of $\boldsymbol{\Phi_0}$ to minimizing the energy $\boldsymbol{E_{K(\Xi)}(\Phi_k)}$, $\boldsymbol{k \geq 0}$: For $\boldsymbol{\Phi_k := \sum_i \delta_{x_i^k}}$ choose a point uniformly at random $\boldsymbol{x_{j,k} \in \Phi_k}$, choose a now location at uniformly at random $\boldsymbol{y \in W}$ in the window and move $\boldsymbol{x_{j,k}}$ to $\boldsymbol{y}$ provided the energy is decrease:

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

- It looks like a substitute for Glauber dynamics ... (?)
- Move the point one at a time, possibly causing many 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: $\boldsymbol{K}(\boldsymbol{\Xi}) \simeq \mathbb{E}[\boldsymbol{K}(\boldsymbol{\Xi})]$ with high probability $\Rightarrow$ not to "memorize" a realization of $\boldsymbol{\Xi}$.
- Sufficiency property: $\mathbb{E}(\boldsymbol{K}(\boldsymbol{\Xi}))$ rich enough, strong (distributional) discriminate power $\Rightarrow$ "summarize" the unknown distribution.

Assuming ergodicity of $\boldsymbol{\Xi}$, a natural choice consists in spatial averaging:

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

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, $\boldsymbol{K}_i(\boldsymbol{\mu}) \simeq \mathbb{E}[\boldsymbol{K}(\boldsymbol{\Xi})]$.

- SPATIAL STATISTICS (DESCRIPTORS)
  - Good choice of statistics
  - Classical summary characteristic

Classical spatial statistics

- mean (intensity) $\mathrm{E}[\Xi(B)]/|B|$,
- correlation functions $\rho(x, y)$,
- Ripley's K-function $K(r)$,
- k-nearest neighbour distance d.f. $D_k(r)$; (Tscheschel and Stoyan 2006),
- ...,
- 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, appox. 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.

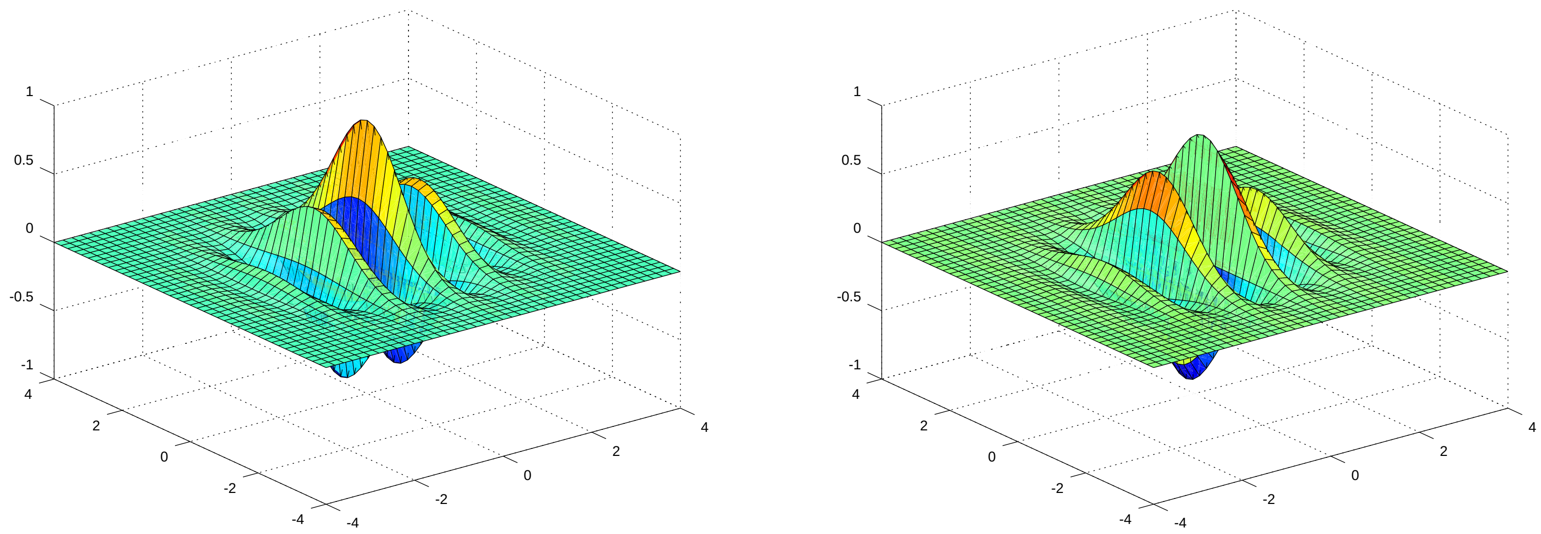
In applications $d = 1$ or 2. In this talk $d = 2$.

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 $x \in \mathbb{R}^2$.



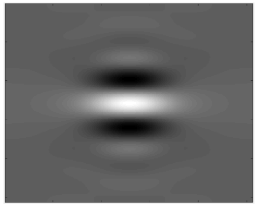
Real and Imaginary part of the Morlet wavelet with $\omega = (5.5, 0)$.

Scaling and rotating the mother wavelet

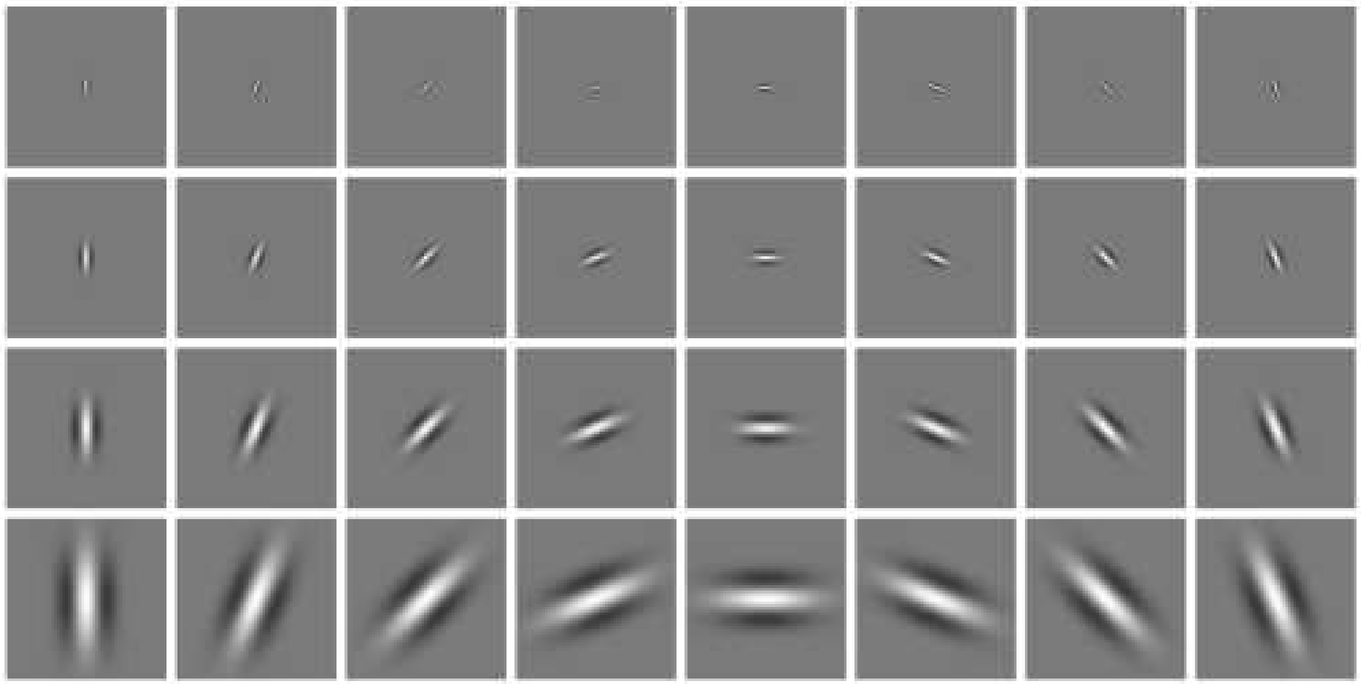
Consider a discrete family of re-scaled and rotated wavelets

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

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).



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



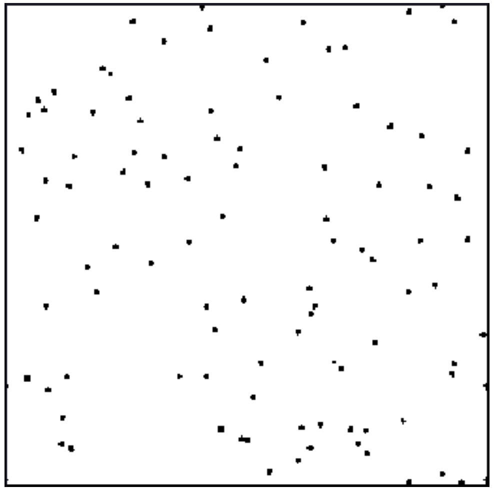
Wavelet transform of the signal

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

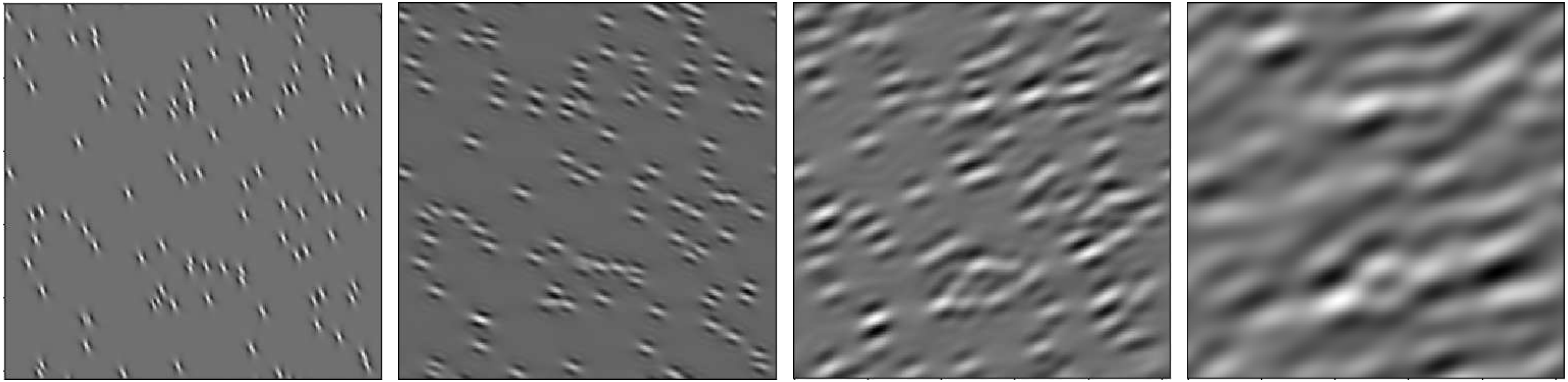
$$(\mu \star \psi_{(j,\theta)})(x) := \int_{\mathbb{R}^d} \psi_{(j,\theta)}(x-y)\,\mu(\mathrm{d}y)\,.$$

Observe: The zero average property of the mother wavelet $\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}x = 0$ implies that the wavelet transform $\mu \star \psi_{(j,\theta)}(x)$ at the scale j has larger absolute values for x where the μ is has more variability at this given scale. It (almost) vanishes where μ 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 $j \in \mathbb{Z}$, $\theta \in [0, 2\pi)$

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

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

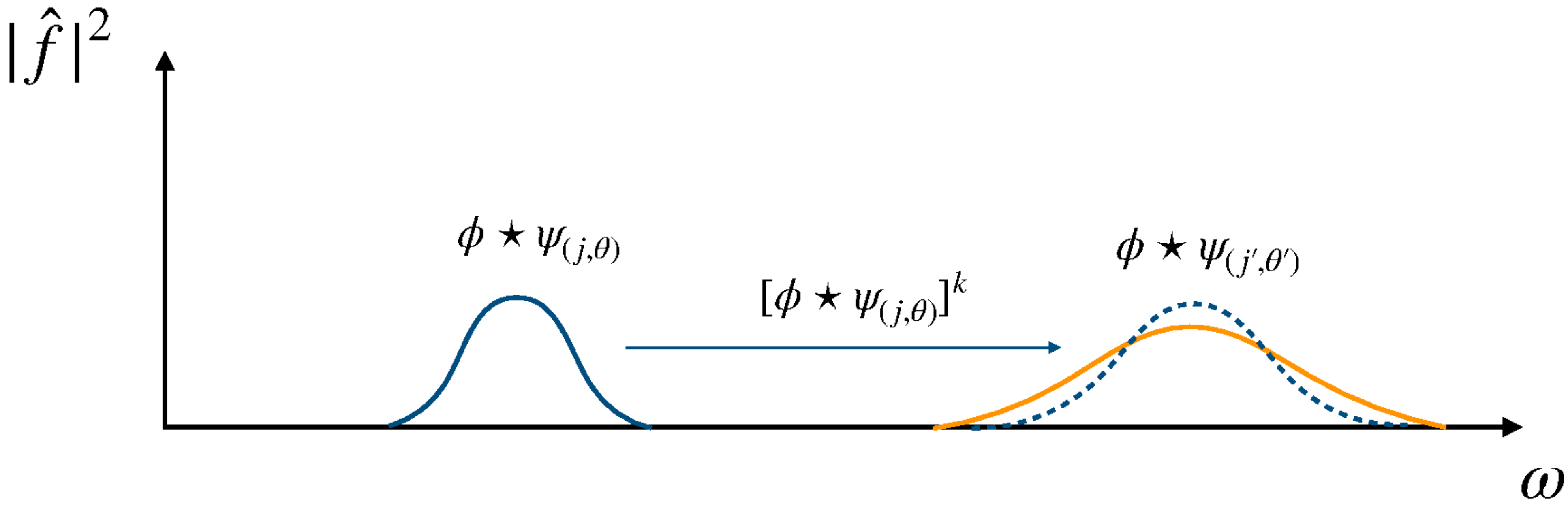
$$\hat{S}\mu(j,\theta) := \frac{1}{|W|} \int_W S_{j,\theta}\mu(x)\,\mathrm{d}x.$$

- 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.

Phase acceleration — refinement of non-linearity

for all $z \in \mathbb{C}$, $k \in \mathbb{Z}$

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

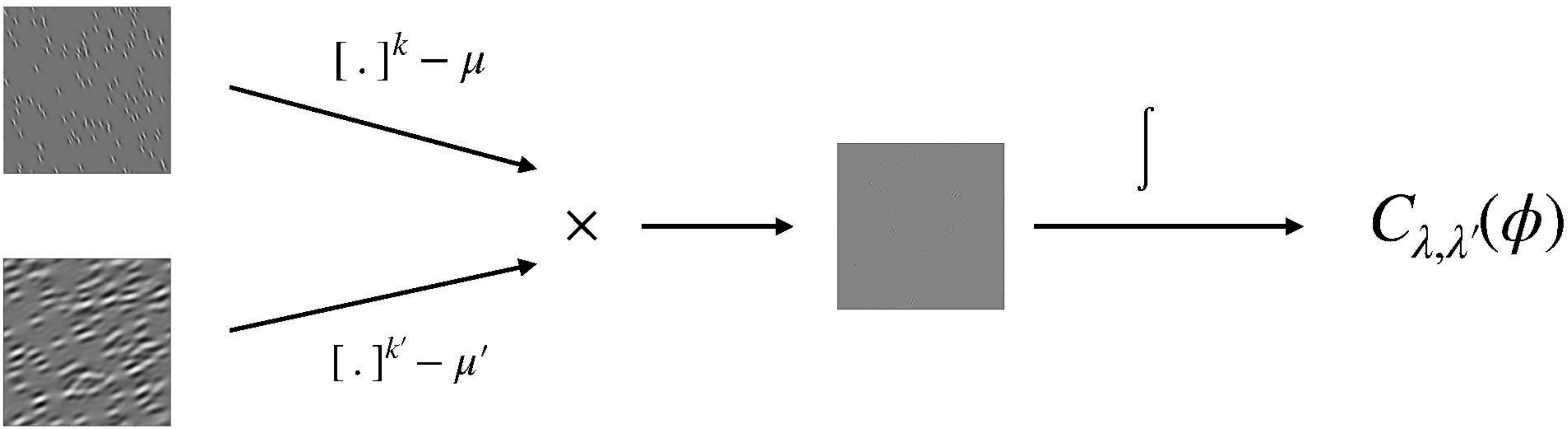


phase acceleration allowing for non-null correlation the scattering transforms

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

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

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



□ TESTING RESULTS

- Visual evaluation
- Power spectrum evaluation
- Comparison to random search (RS) with nearest neigbour distance (NND) by (Tscheschel and Stoyan 2006).
- Topological data analysis (Euler-Poincare characteristic, persistence diagrams)

Main parameters

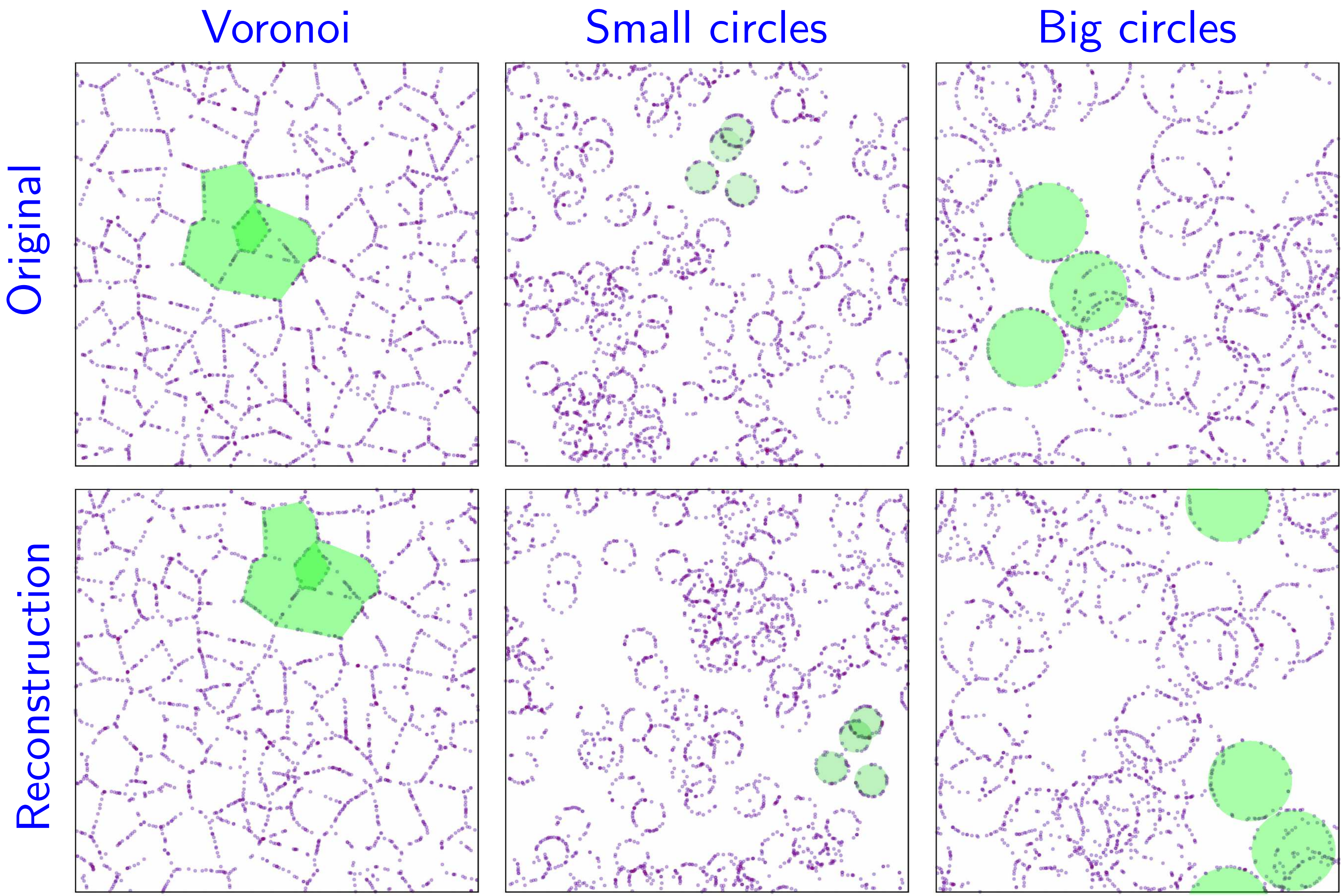
- Number of points in the given image: from $\mathbf{1900}$ up to $\mathbf{13000}$ points.

- Multi-scale statistics $\boldsymbol{K}$ (descriptors) — wavelet phase harmonics based on bump steerable wavelets (Mallat et al. 2020) dimension of $\boldsymbol{K}$):
  - Scales $\mathbf{0} \le \boldsymbol{j} < \boldsymbol{J} = \boldsymbol{J}(\boldsymbol{N})$; where
    $\boldsymbol{N}$ = size of observation window/size of pixel;
    We take $\boldsymbol{N} = \mathbf{128}$ and $\boldsymbol{N} = \mathbf{256}$.
    $\boldsymbol{J} = \log_2(\boldsymbol{N}) - \mathbf{2}$ — "memorizing image",
    $\boldsymbol{J} = \log_2(\boldsymbol{N}) - \mathbf{3}$ — good "learning distribution".
  - Total number of statistics; $|\boldsymbol{K}|$:
    $\boldsymbol{O}$(number of angles^2 $\times$ number of scales^2) $\simeq \mathbf{3\,000} - \mathbf{5\,000}$. (We take number of angles 8).

- Number of iterations (L-BFGS optimization [2]) from $\mathbf{400}$ to $\mathbf{500}$.

[2] Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm; Liu and Nocedal (1989).

"Memorizing" samples modulo random translation



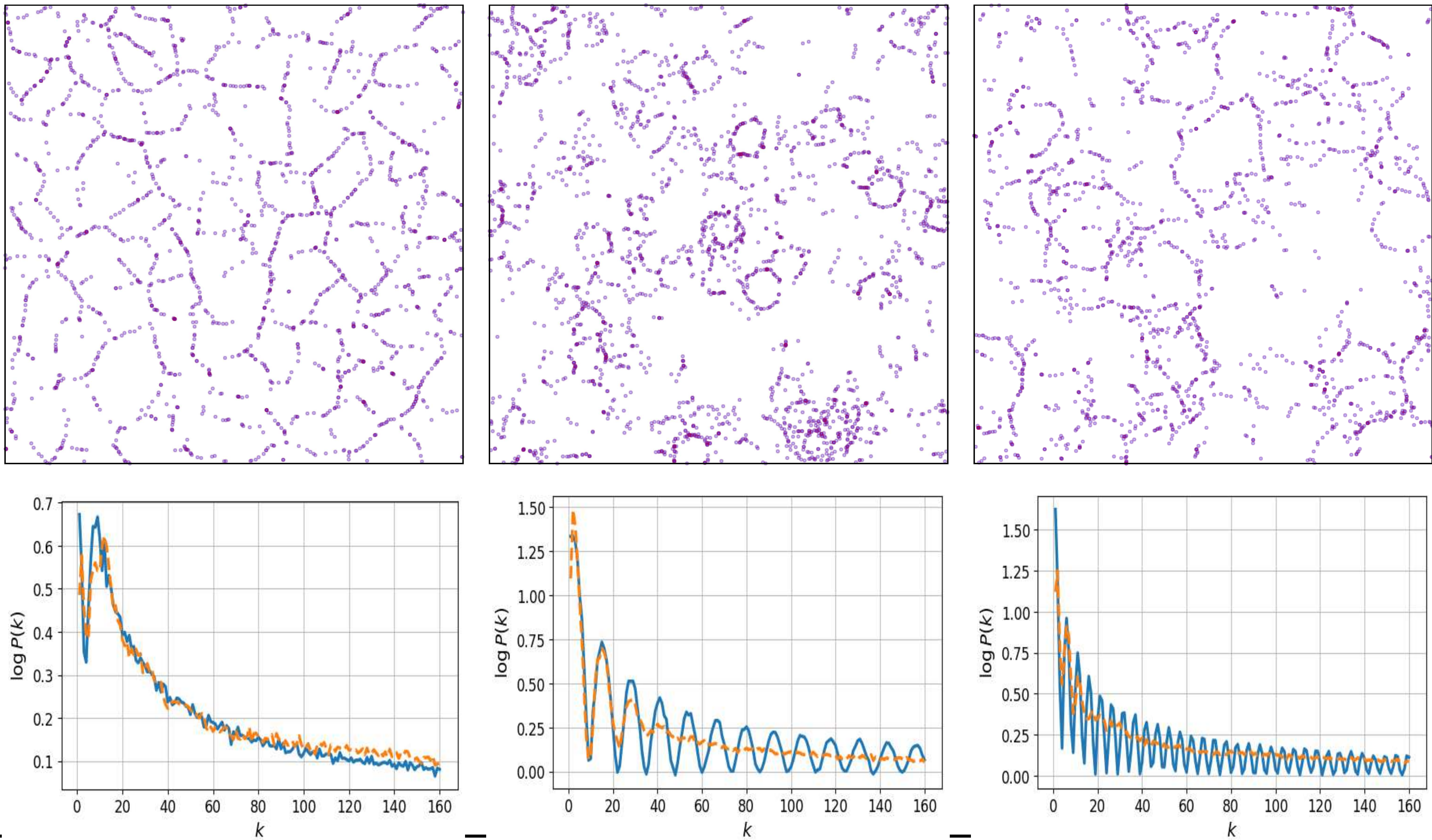
Statistics $\boldsymbol{K}$ based on wavelets of all scales, up to the size of the window.

… removing descriptors in K with too large scales

Voronoi | Small circles | Big circles

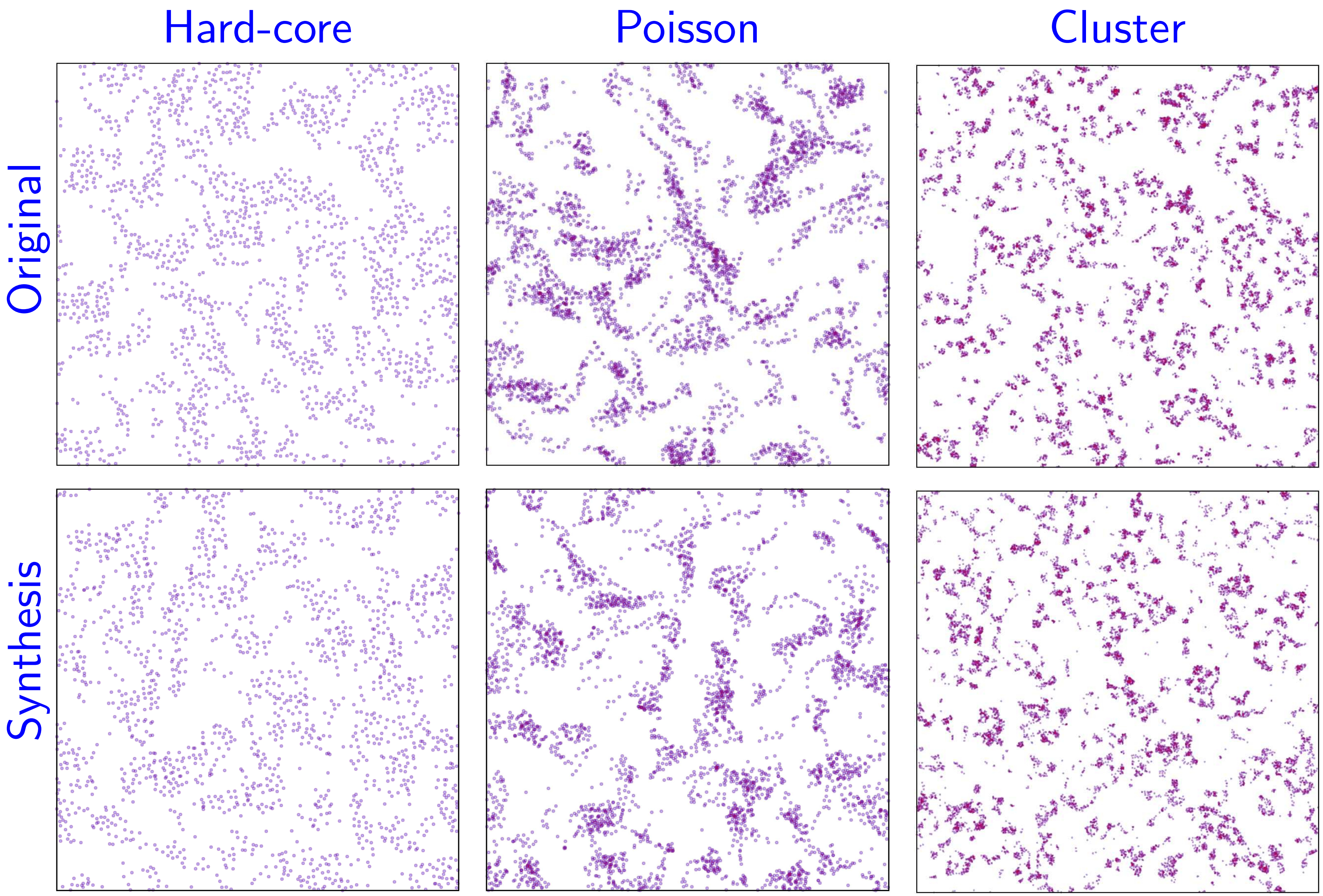
Synthesis

Power spectrum [3]

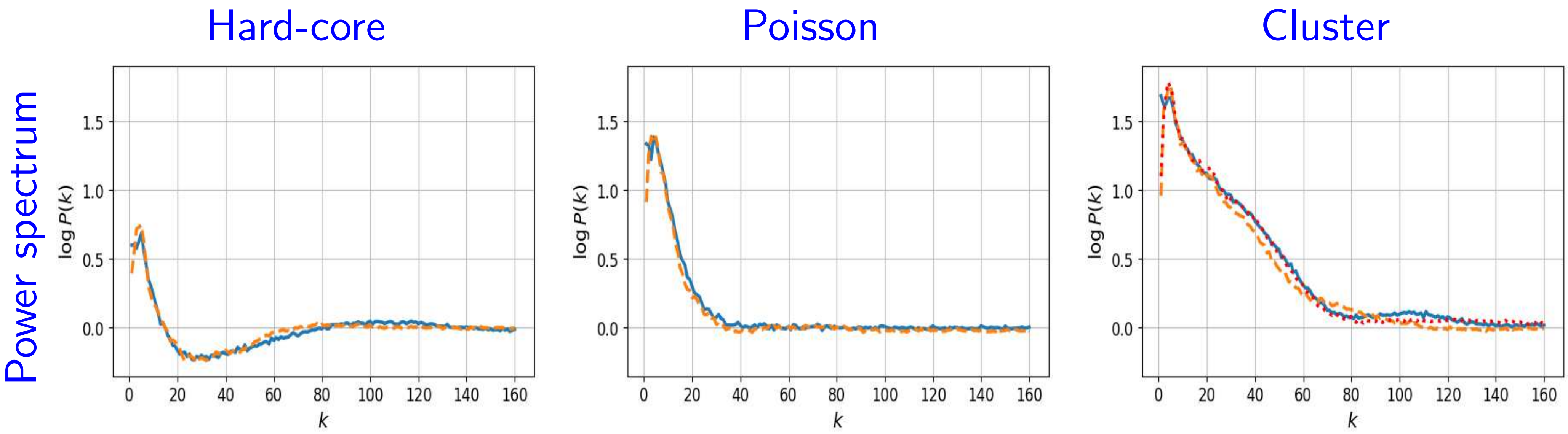


[3] Discrete Fourier transform (estimator of Bartlett's spectrum) — circularly averaged modulus for the frequencies of radius k.

Originals and samples from model for turbulence processes



Turbulence models — spectrum comparison



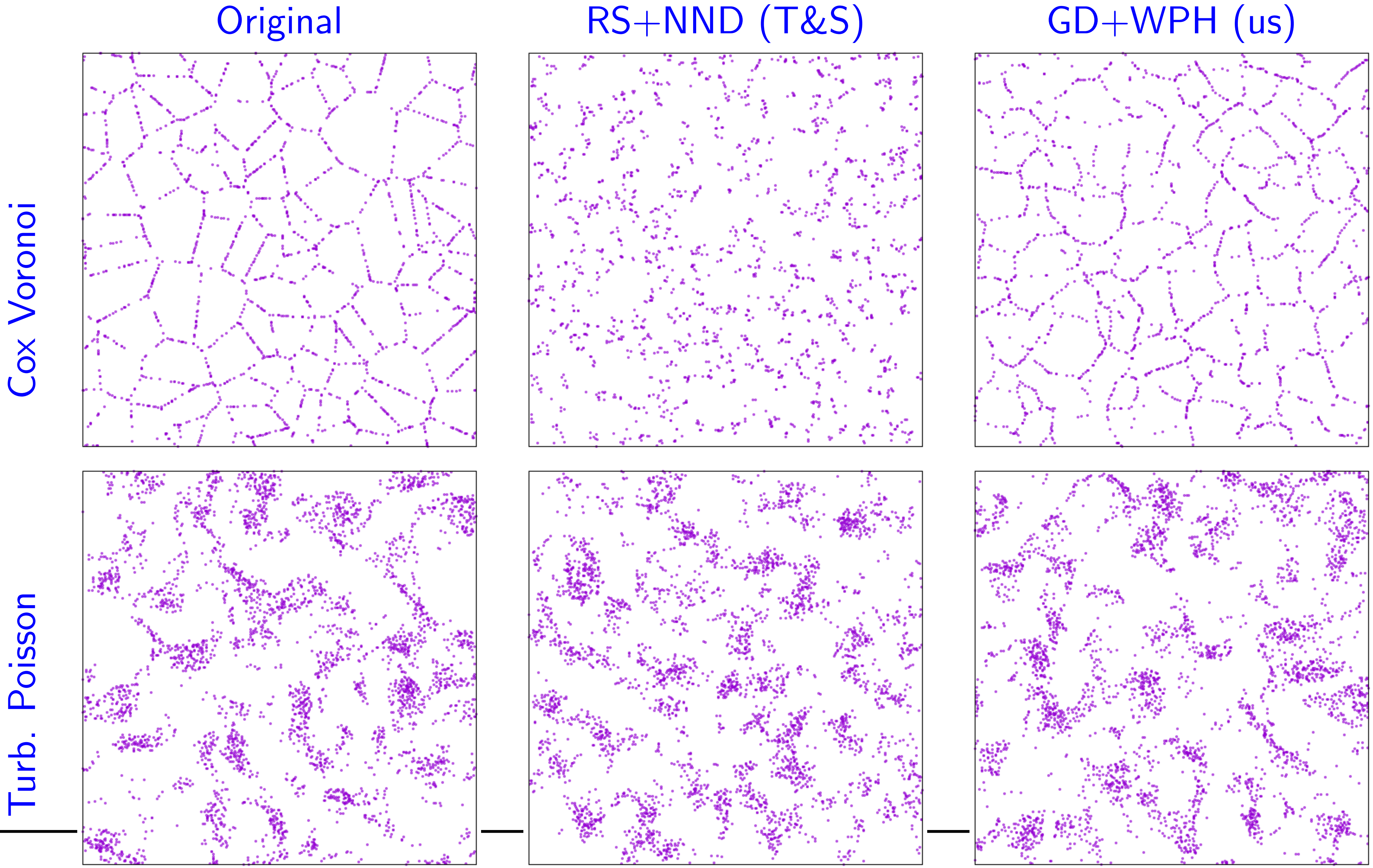
Random search (RS) vs Gradient descent (GD) model

Time computing comparison

Energy error	Random search	Gradient descent
$\boldsymbol{e = 9,00.10^{-4}}$	19870 [10 per point] (1h04m)	52 (0m35s)
$\boldsymbol{e = 4,76.10^{-4}}$	29805 [15 per point] (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.

Visual comparison to (Tscheschel & Stoyan 2006)

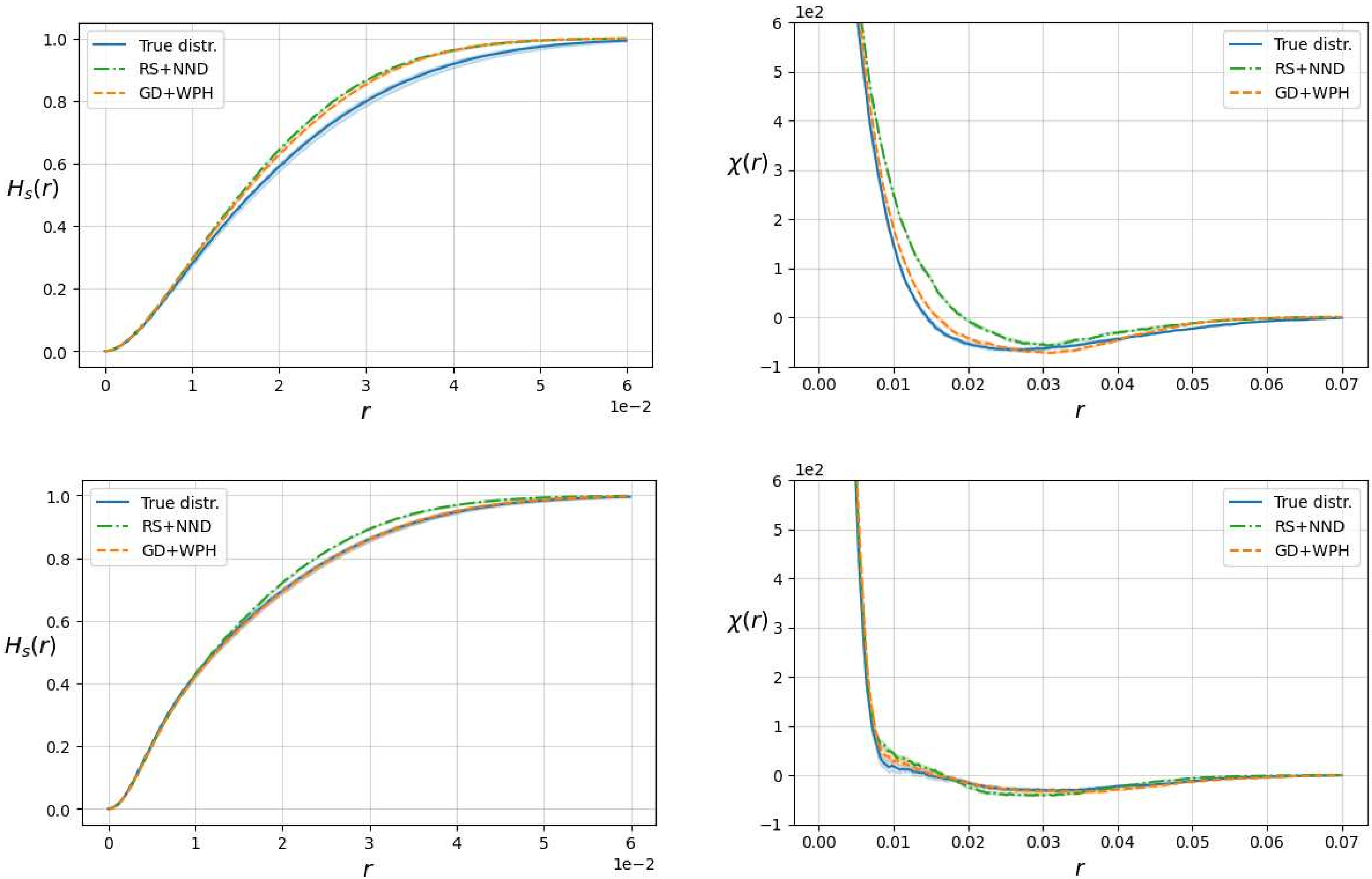


Contact distribution and Euler-Poincaré; T&S vs our model

SCDF [4] | Euler-Poincaré [5]

Cox Voronoi

Turb. Poisson



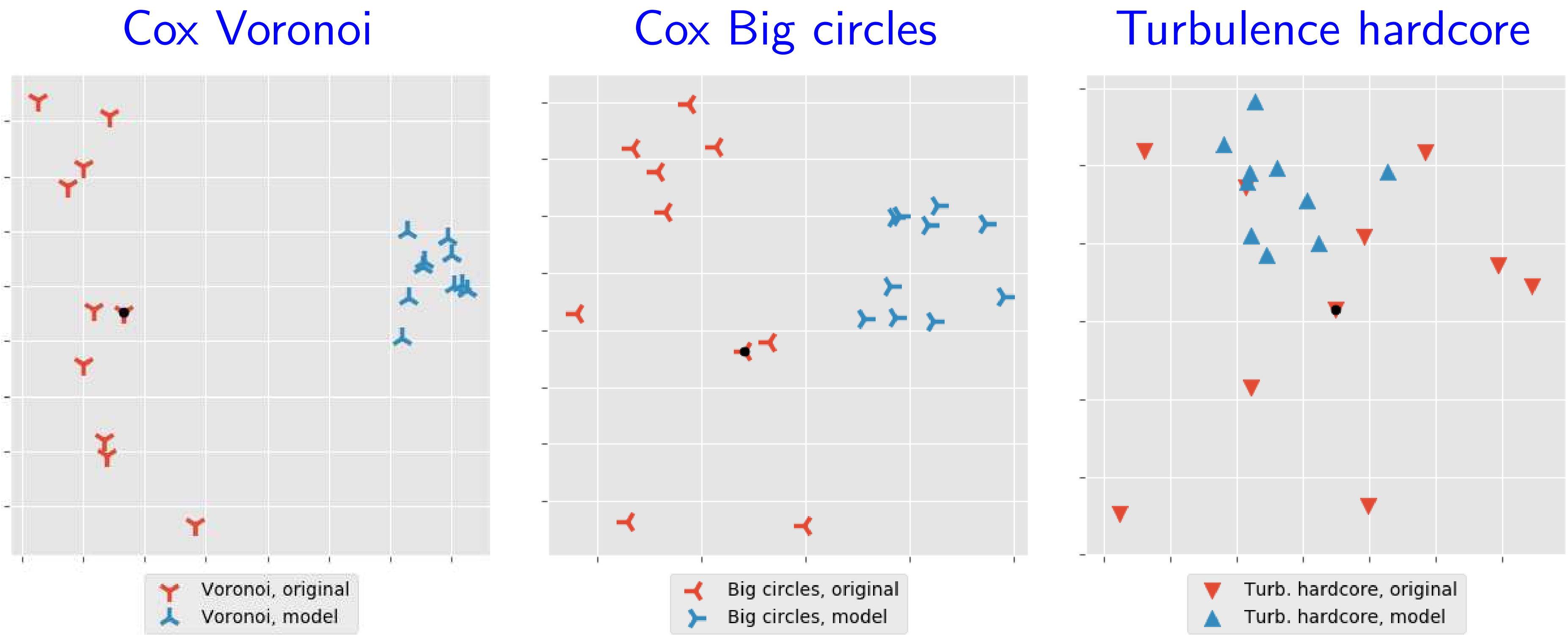
[4] Comparison using Spherical contact distribution function

[5] Number of connected components minus the number of holes in Vietoris-Rips of radius r

Persistent homology (in topology data analysis, TDA)

- Visual evaluation can be more discriminateting, 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).

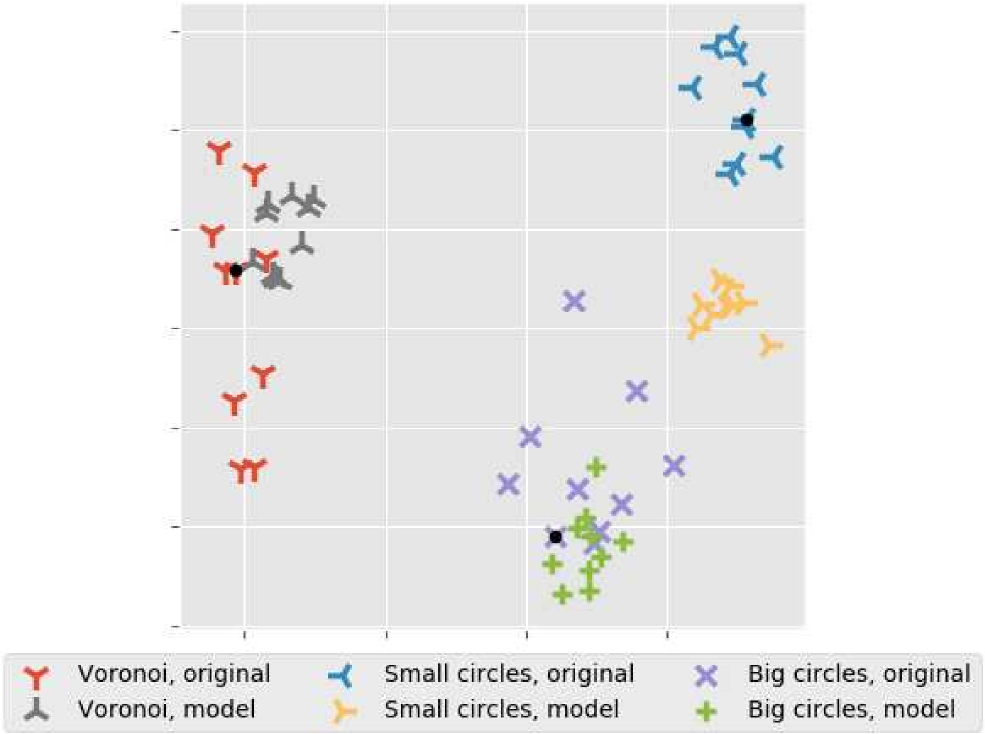
Generative model vs original distribution via TDA



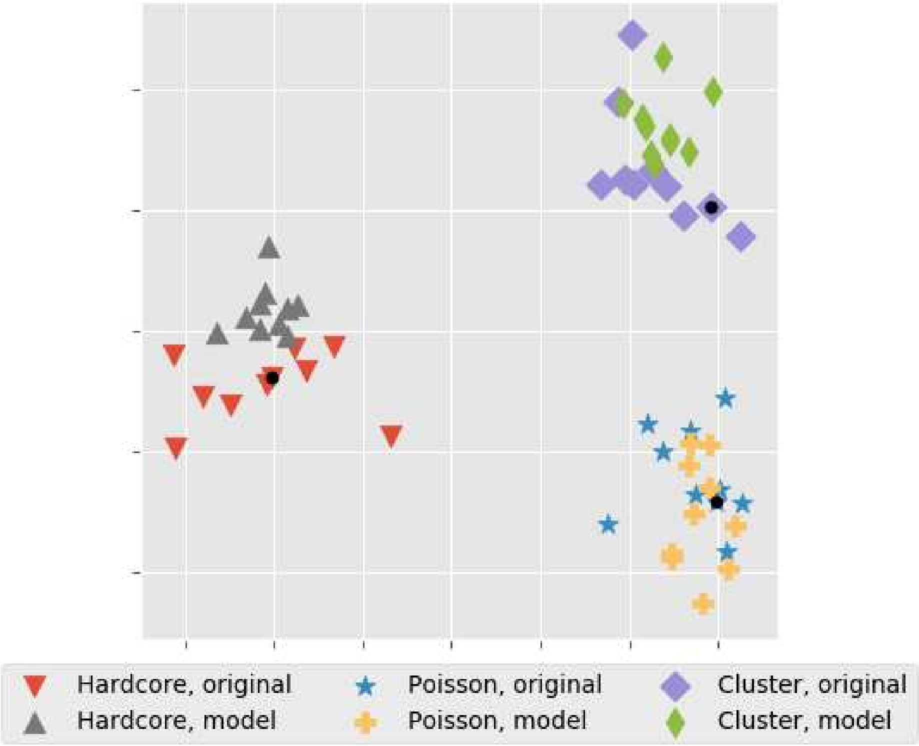
For each model there are 10 “dots” representing (via TDA analysis) i.i.d. realizations of the original distribution and 10 “dots” representing i.i.d. realizations from the generative model estimated on one of the original realizations (marked by the black dot).

Generative model vs original distribution via TDA

Cox — three distributions

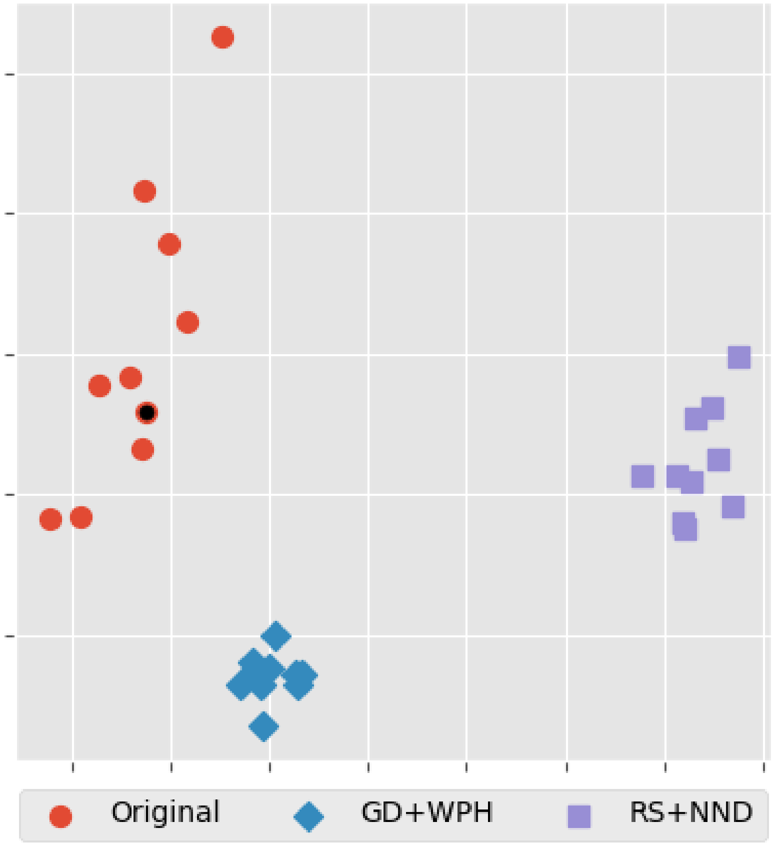


Turbulence — three distributions

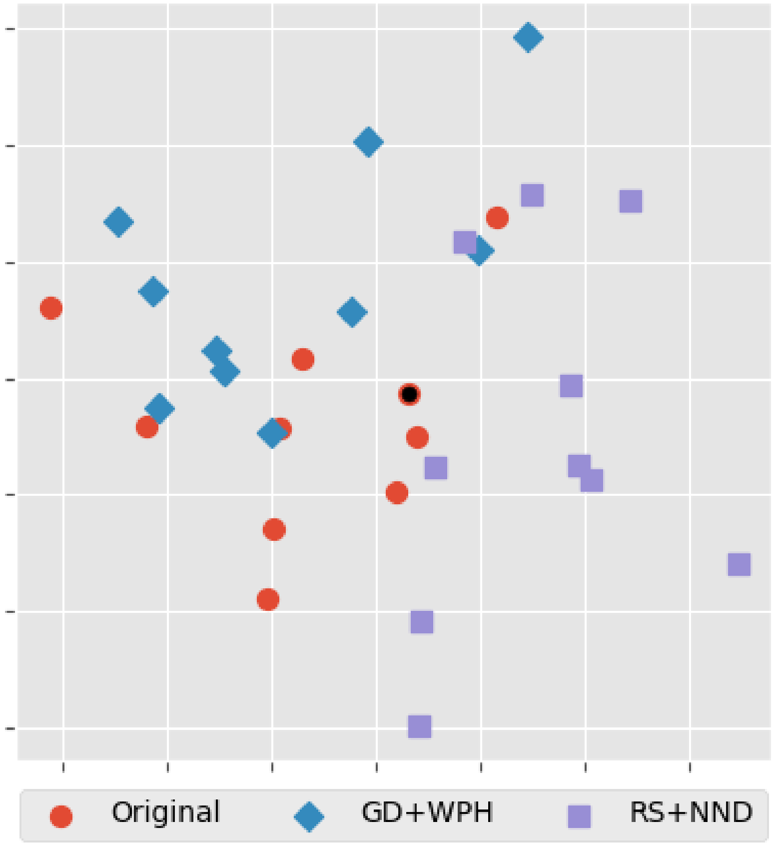


TDA; T&S vs our model

Cox Voronoi



Turbulence Poisson



$$d_{orig/RS+NND} = 1.52$$
$$d_{orig/GD+WPH} = 0.75$$

$$d_{orig/RS+NND} = 0.62$$
$$d_{orig/GD+WPH} = 0.61$$

Averaged distance between the Original models and Generated models (Wasserstein distance in the persistence diagrams).

Conclusions

- Training a generative model for ergodic point processes.
  - Unique realization, large enough in terms of points, provides an approximation of the unknown distribution.
  - Captured by the values of the ergodic estimators evaluated at several scales.
  - Sampling new realizations by pushing Poisson configuration towards a target configuration of points via a gradient descent involving the values of the estimators. (Mimicking Langevin dynamics?)
- Model similar to convolutional neural networks (CNN) with only a few layers and convolutions involving well-understood wavelet methods.
- Gain in the simplicity and interpretation of the CNNs.
- Special feature of our generative model: learning in continuous space (vector graphics) vs classical discrete spaces (raster graphics).

For more details, see:

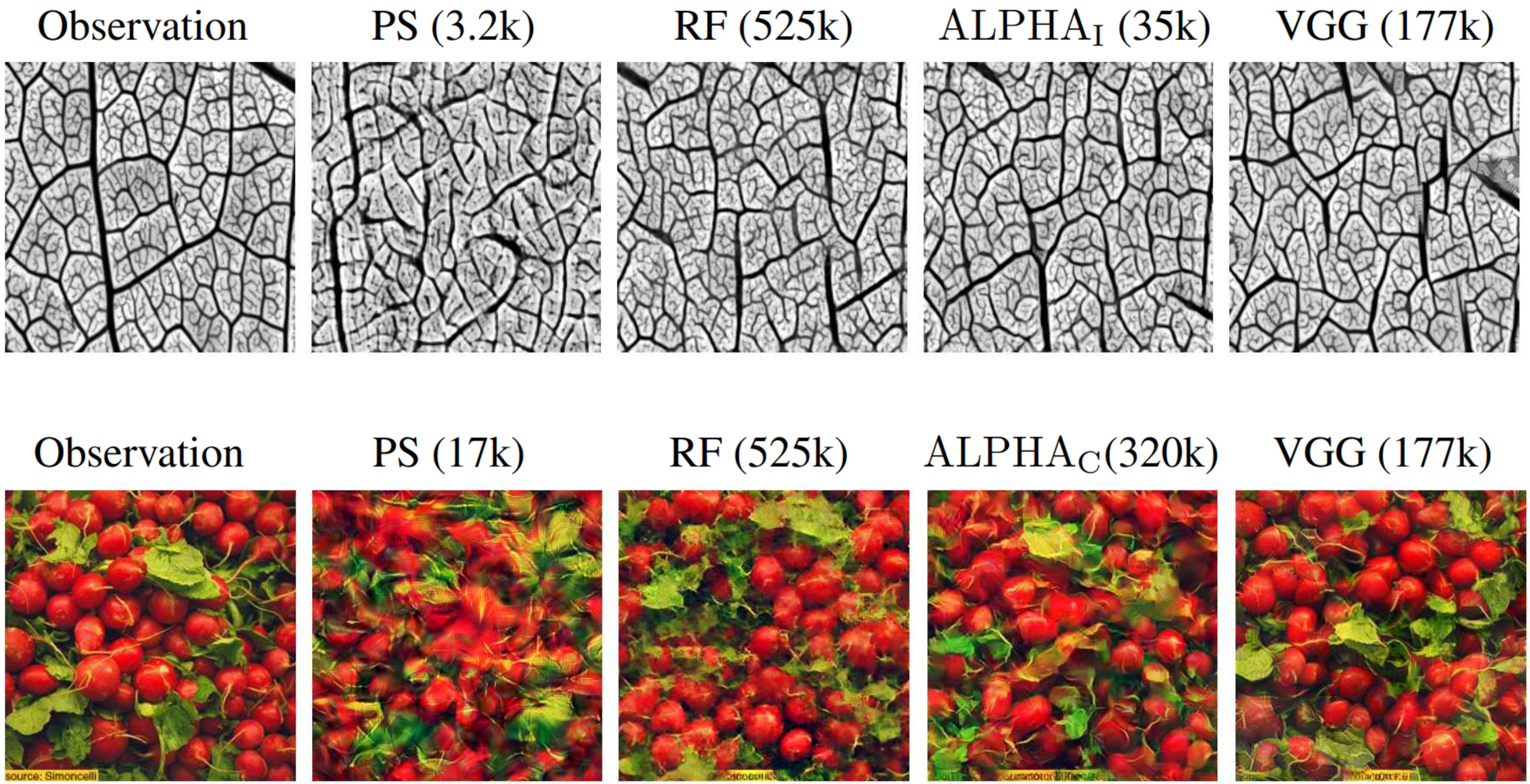
- Brochard, A., BB, Mallat, S. and Zhang, S. (2022). Particle gradient descent model for point process generation. Statistics and Computing; https://arxiv.org/abs/2010.14928
- Brochard, A., S. Zhang and S. Mallat. (2022) Generalized Rectifier Wavelet Covariance Models For Texture Synthesis. ICLR; https://arxiv.org/abs/2203.07902
- Brochard, A. Wavelet-based representations of point processes for modelling and statistical learning PhD thesis, (2022) https://tel.archives-ouvertes.fr/tel-03666508
- Illian, J., Penttinen, A., Stoyan, H., Stoyan, D. (2008) Statistical analysis and modelling of spatial point patterns John Wiley & Sons
- Mallat, S. (2001) A Wavelet Tour of Signal Processing: The Sparse Way, Academic Press
- Mallat, S., Zhang, S., Rochette, G. (2020) Phase harmonic correlations and convolutional neural networks. Information and Inference: A Journal of the IMA
- Molchanov, I., Zuyev, S. (2002) Steepest descent algorithms in a space of measures. Statistics and Computing
- Tscheschel, A., Stoyan, D. (2006) Statistical reconstruction of random point patterns. Computational statistics & data analysis

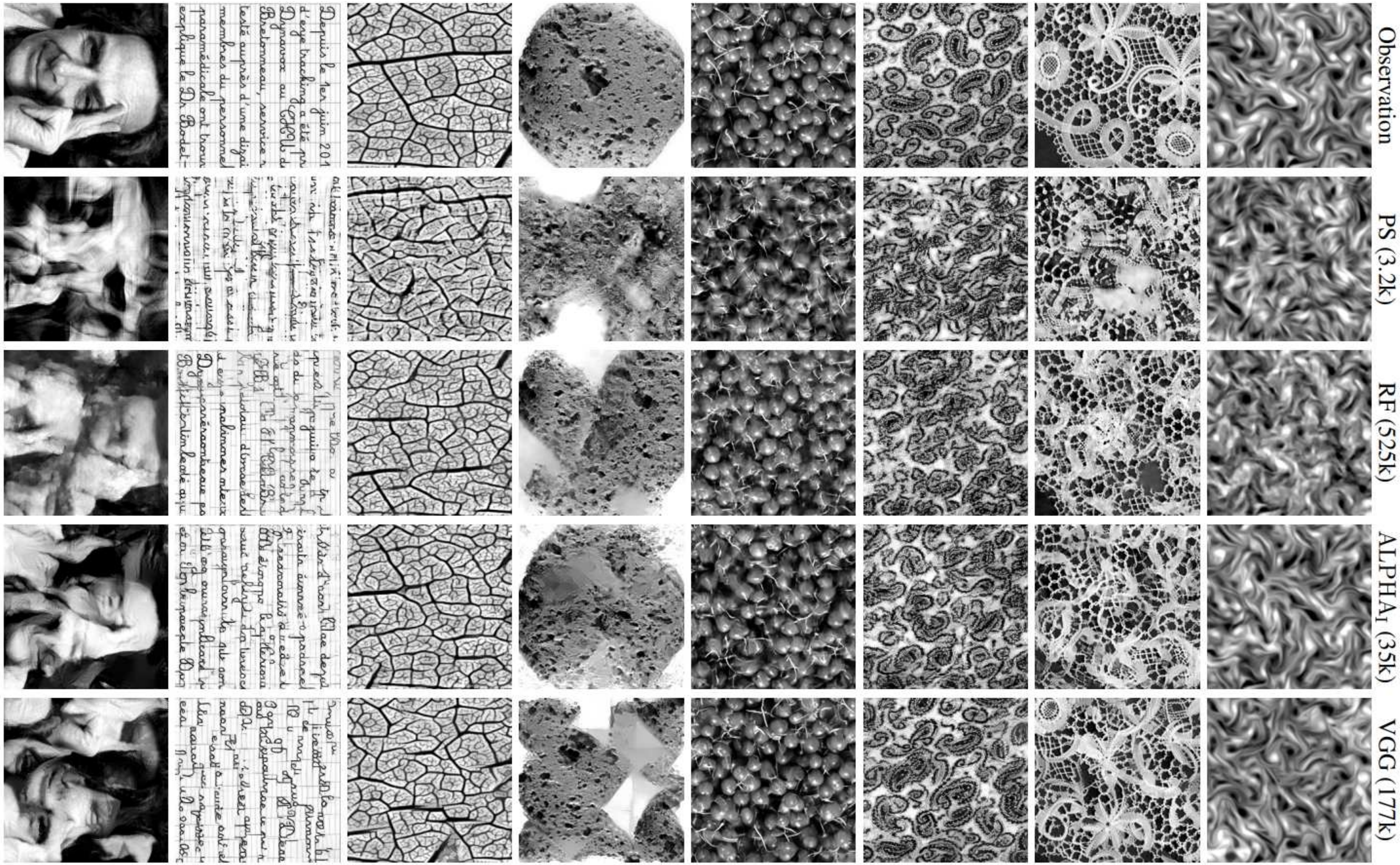
Thank you!

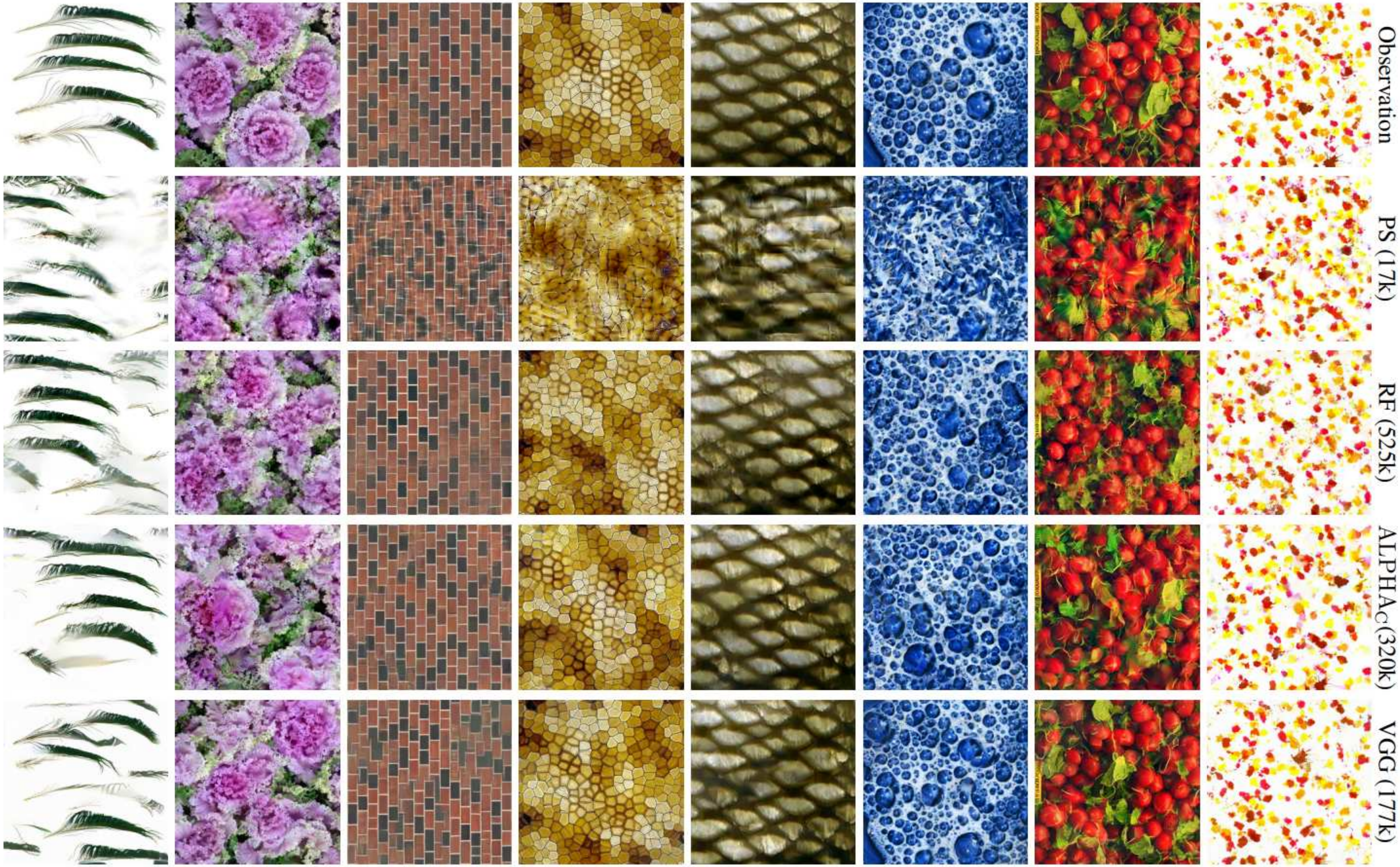
Additional result (raster graphics)

From: Brochard, A., S. Zhang and S. Mallat. (2022) Generalized Rectifier Wavelet Covariance Models For Texture Synthesis. ICLR

Generating textures (raster graphics)




Observation
PS (3.2k)
RF (525k)
ALPHA$_I$ (35k)
VGG (177k)


Observation
PS (17k)
RF (525k)
ALPHA$_C$(320k)
VGG (177k)