Geometry of (smooth) random fields excursions and statistical inference

Anne Estrade - MAP5 UMR 8145 - Université Paris Cité

Colloque "Géométries aléatoires et applications"

Collège de France, 28-29 janvier 2025

Some definitions to start with

- A random field is a map $X \; : \; \Omega \times \mathbb{R}^N \to \mathbb{R}$ such that

$$\omega \in \Omega \mapsto X(\omega, \cdot) \in \mathcal{F}(\mathbb{R}^N, \mathbb{R})$$

is measurable w.r.t. the σ-field generated by $\{$cylinders$\}$.

- The distribution of X is uniquely prescribed by the distribution of all $(X(t_1), \cdots, X(t_n)), \; t_j \in \mathbb{R}^N$. A Gaussian random field is a random field whose all finite dimensional distributions are Gaussian.

- For a level u in $\mathbb{R}$ and a rectangle domain S in $\mathbb{R}^N$, the excursion set above level u within domain S is given by

$$A(u, X, S) = \{s \in S, : \; X(s) \geq u\} \subset \mathbb{R}^N$$

What to do for?

Stochastic model for

- $N = 1$: process indexed by time (price, temperature, ...)
- $N = 2$: grayscale image, climate sciences, sea waves, ...
- $N = 3$: piece of material (rock, concrete, food, ...)
- $N + 1$: space-time phenomena
- $N >> 1$: data set in high dimension

Framework

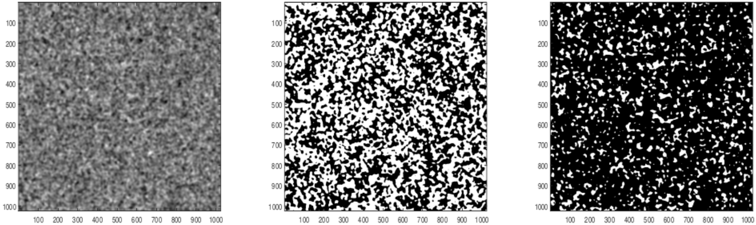
Let $X : \mathbb{R}^N \to \mathbb{R}$, random field that is stationary and isotropic, i.e. its distribution is invariant under translations and rotations in $\mathbb{R}^N$.

Specific models:

- Gaussian and Gaussian based: $F(W)$ with $W : \mathbb{R}^N \to \mathbb{R}^d$ Gaussian (Chi-square, Student, ...)
- Gaussian mixture: $\sqrt{\Lambda}W$ with W Gaussian and Λ heavy-tail random variable
- Shot-noise: $\sum_{(\xi_i, R_i)\in\Phi} \mathbf{1}\{B(\xi_i, R_i)\}$ with Φ a Poisson point process on $\mathbb{R}^N \times \mathbb{R}^+$

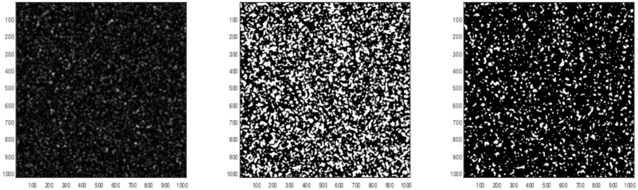
Excursion sets, examples in dimension $N = 2$

Recall that $A(u, X, S) = \{s \in S, : \; X(s) \geq u\} \subset \mathbb{R}^N$

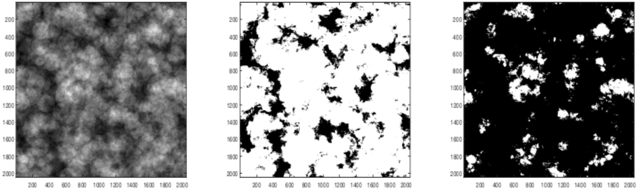


A realization of a Gaussian field with covariance function $e^{-\kappa||s||^2}$ *(left) and two excursion sets above levels* $u = 0$ *(center) and* $u = 1$ *(right)* [1]

[1] Biermé, Di Bernardino, Duval, A.E., EJS, 2019



A realization of a normalized chi-square with 2 degrees of freedom (left) and two excursion sets above levels $u = 0$ (center) and $u = 1$ (right)



A realization of a shot-noise with radius $R \in \{50, 100\}$ (left) and two excursion sets above levels $u = 7.5$ (center) and $u = 14.5$ (right)

Outline of the talk

Questions:

- How do the excursions look like?
  $\hookrightarrow$ *distribution: NO; intrinsic volumes: YES partially*
- What can we infer?
  $\hookrightarrow$ *quantitative and qualitative features characterizing the field*

Methodology:

- Lipshitz-Killing curvatures of the excursion sets
- First moment through Gaussian Kinematic type formulas
- Statistical procedures

Lipschitz-Killing curvatures (LKC)

$(\mathcal{L}_k)_{0 \leq k \leq N}$: additive functionals defined on domains in $\mathbb{R}^N$

Heuristically, in the case of dimension $N = 2$

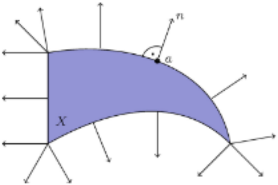
- $\mathcal{L}_2 =$ area, related to the occupation density
- $\mathcal{L}_1 = 1/2$ perimeter, related to the regularity
- $\mathcal{L}_0 =$ nb of connected components - nb of holes
  $=$ Euler characteristic, related to the connectivity

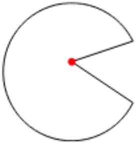
Heuristically, in dimension $N > 2$

- $\mathcal{L}_N = N$-dimensional volume (Lebesgue measure)
- $\mathcal{L}_{N-1} = N - 1$-dimensional surface measure of the border
- ...
- $\mathcal{L}_0 =$ Euler characteristic

Positive Reach sets

Intuitively, A is a positive reach set if one can roll a ball of positive radius along the exterior boundary of A keeping in touch with A.





Result.[2] *For X a "nice" random field and S a rectangle domain in $\mathbb{R}^N$, the excursion sets $A(u, X, S)$ are Positive Reach sets. It includes Gaussian based random fields with C^3 sample paths and shot-noise fields.*

[2] Thäle, Surveys in Math. and App., 2008

Gaussian Kinematic Formula (GKF)

Let $X : \mathbb{R}^2 \to \mathbb{R}$ be a Gaussian field, stationary, isotropic, standard, a.s. C^3 sample paths, non degenerate.
Result.[3] *For* $k = 0, 1, 2$, *and* $u \in \mathbb{R}$, S *rectangle* $\subset \mathbb{R}^2$,

$$\mathbb{E}\mathcal{L}_k(A(u, X, S)) = \sum_{j=0}^{2-k} \omega_{j,k}\, \lambda^{j/2}\, \mathcal{L}_{j+k}(S)\, \rho_j(u)$$

where

- $\omega_{j,k}$ *are universal constants*
- λ: *second spectral moment of* X, $Var(\nabla X(0)) = \lambda Id$
- $\mathcal{L}_2(S) = |S|$; $\mathcal{L}_1(S) = 1/2\,|\partial S|$; $\mathcal{L}_0(S) = 1$
- $\rho_0(u) = \overline{\Phi}(u)$; $\rho_{j+1}(u) = \rho_j'(u)$

[3] Adler & Taylor, Springer, 2007

More GKF-type formulas

Similar formulas do exist for

- dimension $N > 2$
- non Euclidean domains (sphere,...)
- Gaussian based fields, Gaussian mixtures
- shot-noise fields (weak formula)
- anisotropy is allowed
- ...

Statistical inference based on the excursion sets

- Very sparse observation: one single realization of $A(u, X, S)$ and only some $\mathcal{L}_k(A(u, X, T))$ (say $k = 0, N-1, N$) are observed for one or two fixed levels u and fixed domain S.
- Quantitative features characterizing X can be inferred: variance, second spectral moment, anisotropy index, ...
- Qualitative features can be tested: gaussianity, isotropy, extreme values, ...
- Theoretical knowledge given by GKF and CLT for large S

Central Limit Theorem for $\mathcal{L}_k(A(u, X, S))$

Let $X : \mathbb{R}^N \to \mathbb{R}$ be a Gaussian field, stationary, isotropic, standard, a.s. C^3 sample paths, non degenerate.

Result.[4] *For each k, under assumptions that ensure the existence of a finite and non vanishing asymptotic variance,*

$$\frac{\mathcal{L}_k(A(u, X, S)) - \mathbb{E}\mathcal{L}_k(A(u, X, S))}{|S|^{1/2}} \underset{S \nearrow \mathbb{R}^N}{\longrightarrow} \mathcal{N}(0, v_k)$$

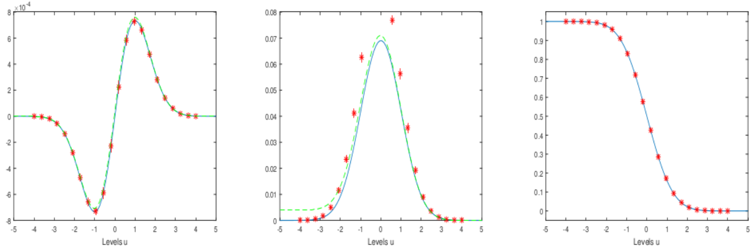
Also multivariate CLT for finitely many levels $u_1, u_2, \ldots$.

It yields consistent estimation of $\mathcal{L}_k(A(u, X, S))$ on large domains, similarly on a fixed domain with dense grid ("infill statistics").

[4] Kratz & Vadlamani, JTP, 2017

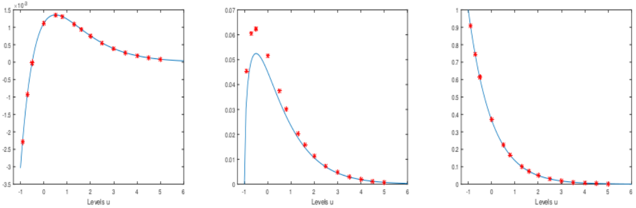
Monte-Carlo estimation of $\mathbb{E}\mathcal{L}_k(A(u,X,S))$[1]

$X : \mathbb{R}^2 \to \mathbb{R}$ Gaussian standard with covariance function $e^{-||s||^2}$

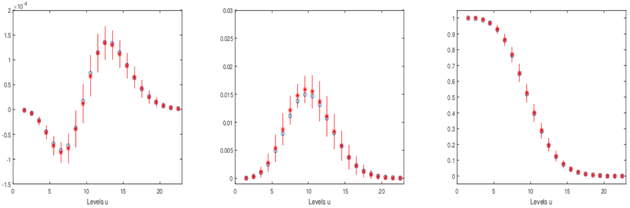


Theoretical map $u \mapsto \mathbb{E}\mathcal{L}_k(A(u,X,S))$ *(blue curve)*
and empirical average of $\mathcal{L}_k(A(u,X,S))$ *with* $n = 100$ *(red crosses)*
for $k = 0$ *(left),* $k = 1$ *(center),* $k = 2$ *(right)*

$X : \mathbb{R}^2 \to \mathbb{R}$ normalized chi-square with 2 degrees of freedom



$X : \mathbb{R}^2 \to \mathbb{R}$ shot-noise with radius $R \in \{50, 100\}$



Theoretical map $u \mapsto \mathbb{E}\mathcal{L}_k(A(u, X, S))$ *(blue curve)*
and empirical average of $\mathcal{L}_k(A(u, X, S))$ *with* $n = 100$ *(red crosses)*
for $k = 0$ *(left),* $k = 1$ *(center),* $k = 2$ *(right)*

How to get $\mathcal{L}_k(A(u, X, S))$, $k = 0, 1, 2$?

- $k = 2$: easy task ! by taking the measure of the occupied region (or counting the black pixels)
- $k = 0$: using Morse theory,
  $\mathcal{L}_0(A(u, X, S)) \approx \#(max \geq u \text{ in } S) - \#(min \geq u \text{ in } S)$
- $k = 1$: difficult task! bias coming from discretization and intrinsic drawback [5][6]

[5] Biermé & Desolneux, Ann. H. Lebesgue, 2021
[6] Cotsakis et al., Ann. Appl. Prob., 2024

Parameter estimation

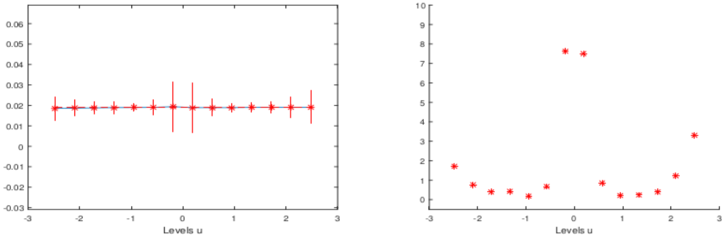
Introduce LKC densities $C_k(u) := \lim_{S \nearrow \mathbb{R}^2} \frac{\mathbb{E}\mathcal{L}_k(A(u, X, S))}{|S|}$

- unbiaised estimators for $C_k(u)$ are build by combining the equations in the GKFs, say $\widehat{C_k}(u, S)$,
- central limit theorems give confidence intervals,
- heuristically: with the observation of $\mathcal{L}_k(A(u, X, S))$ for three values of $k = 0, 1, 2$ (fixed u and S), one is able to estimate three distinct structure parameters.

Example of a parameter estimation

Let $X : \mathbb{R}^2 \to \mathbb{R}$, stationary, Gaussian, standard, with unknown second spectral moment λ.
Recall that GKF yields $C_0(u) = \lambda\,(2\pi)^{-3/2}\,ue^{-u^2/2}$.

Result.[1] $\widehat{\lambda}(u) := (2\pi)^{3/2}\,\frac{e^{u^2/2}}{u}\,\widehat{C_0}(u, S)$ *is an estimator of* λ *that satisfies a CLT as* $S \nearrow \mathbb{R}^2$.



Estimate $\widehat{\lambda}(u)$ *with associated confidence intervals for different values of u (left). Empirically estimated variance for different values of u (right).*

Example of a test: test of Gaussianity

Assume $H0$: X is Gaussian (X is stationary and standard)

- from GKF we know that $C_0(u) = \lambda\,(2\pi)^{-3/2}\,ue^{-u^2/2}$, so

$$\frac{C_0(\alpha u)}{C_0(u)} = \alpha\, e^{u^2(1-\alpha^2)/2},\ \forall\alpha > 0$$

- empirical estimator $\widehat{C_0}(u, S)$ provides a test statistic

$$R(\alpha) = \frac{\widehat{C_0}(\alpha u, S_1)}{\widehat{C_0}(u, S_2)} \text{ satisfying a CLT}$$

- power of the test can be evaluated for specific alternative hypothesis $H1$ such as chi-square, Student, shot-noise,...

In a similar way ...

- Test of symmetry (Abaach et al., EJS, 2021)
- Test of isotropy (Abaach et al., SpaSta, 2025)
- Estimation of the perturbation for $Y = X + noise$
- "Expected Euler characteristic heuristic"

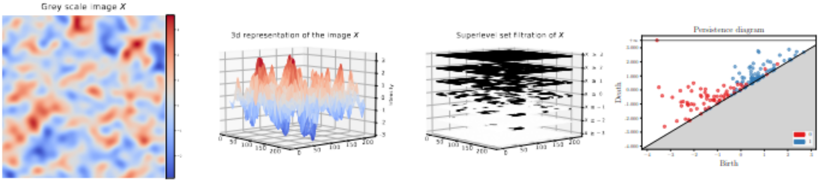
$$\mathbb{E}[\mathcal{L}_0(A(u, X, S))] \underset{u \to \infty}{\sim} \overline{\Phi}(u) = \mathbb{P}(X(s) > u)$$

(Adler & Taylor, St-Flour LNM, 2011)
- Extreme value analysis and Peak-over-threshold theory for Gaussian mixtures (Di Bernardino et al., Extremes, 2024)
- Geometrical and Topological Data Analysis

Short inside in Geometrical TDA

From a grayscale image,

- excursion sets at various levels are extracted,
- connected components and "holes" are labelled,
- birth-and-death times are recorded,
- a persistence diagram is drawn and ... ML algorithms

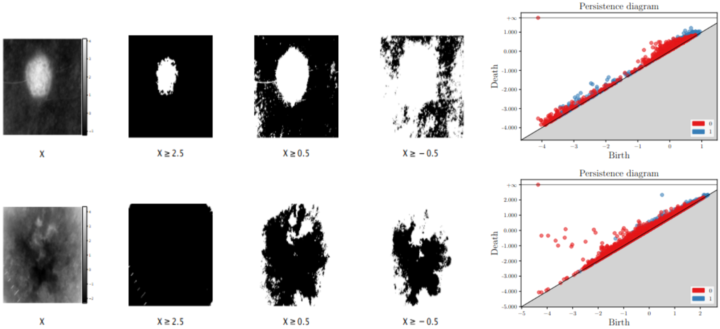


Persistence diagram extracted from a grayscale image [7]

[7] Abaach & Morilla, ArXiv, 2023

Application to skin mole images

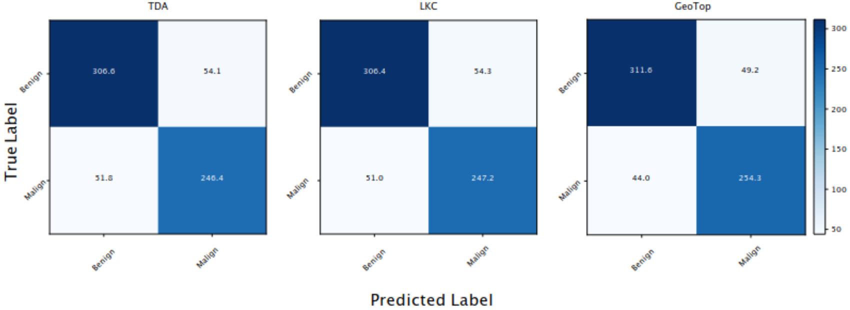
Classification between benign and malignant moles



Benign (top row) and malignant (bottom row) mole images with various excursions and associated persistence diagram

GeoTop algorithm

- classical TDA algorithms use persistence diagrams
- adding geometrical features of excursion sets provides more meaningfull information



Comparison of confusion matrix on 660 images for three methods: TDA (left), LKC (center), GeoTop (right) [7]

[7] Abaach & Morilla, ArXiv, 2023

Conclusion and open problems

The expected LKC of random fields excursion sets

- are nice summaries to characterize the geometry of a field,
- are nice tools to characterize extreme value behaviour
- provide theoretical formulas for statistical inference (estimation, tests)
- improve performance and understanding of TDA algorithms

but they need

- to be explicitely computed: extended models?
- to be consistently estimated: CLT ?
- to be evaluated on a grid: discretization ?
- to be parameter-free: normalization?

Advertising - Coming soon !

"Including Geometry in Topological Data Analysis", PhDInFrance program (CoFund, FSMP), dead line February 14

Stochastic Geometry Days 2025, Grenoble, June 23-27

"Stochastic Geometry: Percolation, Tesselations, Gaussian Fields and Point Processes", Springer LNM with lectures given during last annual meetings of the french research group GeoSto (now MAIAGES), to be published in 2025 (?)

RT MAIAGES

https://rt-maiages.math.cnrs.fr/

Thank you for your attention

and many thanks to my co-authors

- Mariem Abaach (PhD, Univ. Paris Cité)
- Hermine Biermé (Univ. Tours)
- Elena Di Bernardino (Univ. Nice Cote d'Azur)
- Federico Dalmao (Univ. de la República, Uruguay)
- Céline Duval (Sorbonne U.)
- Julie Fournier (Univ. Paris Saclay)
- José León "Chichi" (Univ. de la República, Uruguay)
- Thomas Opitz (INRAE, Avignon)
- Maurizia Rossi (Univ. Milano Bicocca)

M. Abaach, H. Biermé, E. Di Bernardino (2021). Testing marginal symmetry of digital noise images through the perimeter of excursion sets, *EJS*, **15**(2), 6429–6460.

Abaach M., Biermé H., Di Bernardino E., Estrade A. (2025). Local isotropy test based on oriented perimeter for digital images, *Spatial Stat.*, **65**.

Abaach M., Morilla I. (2023). GeoTop: Advancing Image Classification with Geometric-Topological Analysis, *Preprint*, Hal-archives ouvertes-2311.16157.

Adler R.J., Taylor J.E. (2009). *Random Fields and Geometry*, Springer, New York.

Adler R.J., Taylor J.E. (2011). *Topological complexity of smooth random functions*, Ecole d'été de Probabilités de Saint-Flour XXXIX-2009, Springer.

Biermé H., Desolneux A. (2021). The effect of discretization on the mean Geometry of a 2D random fields. *Annales Henri Lebesgue*, **4**, 1295-1345.

- Biermé H., Di Bernardino E., Duval C., Estrade A. (2019) Lipschitz-Killing curvatures of excursion sets for two dimensional random fields, *Electron. J. Statist.* **vol 13**; 536-581.
- Cotsakis R., Di Bernardino E., Duval C. (2024). Surface area and volume of excursion sets observed on point cloud based polytopic tessellations, *Ann. of Appl. Probab.*, **34**(3), 3093-3124.
- Di Bernardino E., Estrade A., Opitz T. (2024), Spatial extremes and stochastic geometry for Gaussian-based peaks-over-threshold processes, *Extremes*, **27**, 397–435.
- Kratz M., Vadlamani S. (2017). Central Limit Theorem for Lipschitz–Killing Curvatures of Excursion Sets of Gaussian Random Fields. *J. Theor. Probab.* **26** (3), 1–30.
- Thäle C. (2008). 50 years sets with positive reach - a survey. Surveys in Mathematics and its Applications, 3:123–165, 2008.