



**HAL**  
open science

## Scattering Spectra Models for Physics

Sihao Cheng, Rudy Morel, Erwan Allys, Brice Ménard, Stéphane Mallat

► **To cite this version:**

Sihao Cheng, Rudy Morel, Erwan Allys, Brice Ménard, Stéphane Mallat. Scattering Spectra Models for Physics. PNAS Nexus, 2024, 3 (4), pp.103. 10.1093/pnasnexus/pgae103 . hal-04170733

**HAL Id: hal-04170733**

**<https://hal.science/hal-04170733v1>**

Submitted on 11 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Scattering spectra models for physics

Sihao Cheng <sup>a,\*</sup>, Rudy Morel <sup>b</sup>, Erwan Allys <sup>c</sup>, Brice Ménard <sup>d</sup> and Stéphane Mallat <sup>b,e,f</sup>

<sup>a</sup>School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA

<sup>b</sup>Département d'informatique de l'ENS, ENS, CNRS, PSL University, 75014 Paris, France

<sup>c</sup>Laboratoire de Physique de l'Ecole normale supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université Paris Cité, 75014 Paris, France

<sup>d</sup>Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA

<sup>e</sup>Collège de France, 75231 Paris, France

<sup>f</sup>Center for Computational Mathematics, Flatiron Institute, New York, NY 10010, USA

\*To whom correspondence should be addressed: Email: [scheng@ias.edu](mailto:scheng@ias.edu)

Edited By: Gigliola Staffilani

## Abstract

Physicists routinely need probabilistic models for a number of tasks such as parameter inference or the generation of new realizations of a field. Establishing such models for highly non-Gaussian fields is a challenge, especially when the number of samples is limited. In this paper, we introduce scattering spectra models for stationary fields and we show that they provide accurate and robust statistical descriptions of a wide range of fields encountered in physics. These models are based on covariances of scattering coefficients, i.e. wavelet decomposition of a field coupled with a pointwise modulus. After introducing useful dimension reductions taking advantage of the regularity of a field under rotation and scaling, we validate these models on various multiscale physical fields and demonstrate that they reproduce standard statistics, including spatial moments up to fourth order. The scattering spectra provide us with a low-dimensional structured representation that captures key properties encountered in a wide range of physical fields. These generic models can be used for data exploration, classification, parameter inference, symmetry detection, and component separation.

**Keywords:** reduced models, Gibbs energy, wavelets

## Significance Statement

Physicists need to characterize fields with a variety of structures, but building probabilistic models beyond the simple Gaussian model is often challenging, especially when the number of data samples is limited. We introduce scattering spectra models that make use of symmetry and regularity properties of physical fields and show that they can provide accurate and compact statistical descriptions for a wide range of fields. Providing both summary statistics and generative models, this representation can be used for data exploration, classification, parameter inference, symmetry detection, and component separation in analyzing the ever-growing datasets in physics and beyond.

## Introduction

An outstanding problem in statistics is to estimate the probability distribution  $p(x)$  of high-dimensional data  $x$  from few or even one observed sample. In physics, establishing probabilistic models of stochastic fields is also ubiquitous, from the study of condensed matter to the Universe itself. Indeed, even if physical systems can generally be described by a set of differential equations, it is usually not possible to fully characterize their solutions. Complex physical fields, described here as non-Gaussian random processes  $x$ , may indeed include intermittent phenomena as well as coherent geometric structures such as vortices or filaments. Having realistic probabilistic models of such fields however allows for considerable applications, for instance to accurately characterize and compare nonlinear processes, or to separate different sources and solve inverse problems. Unfortunately, no generic probabilistic model is available to describe complex physical fields

such as turbulence or cosmological observations. This paper aims at providing such models for stationary fields, which can be estimated from one observed sample only.

At thermal equilibrium, physical systems are usually characterized by the Gibbs probability distribution, also called Boltzmann distribution, that depends on the energy of the systems (1). For non-equilibrium systems, at a fixed time one may still specify the probability distribution of the field with a Gibbs energy, which is an effective Hamiltonian providing a compact representation of its statistics. Gibbs energy models can be defined as maximum entropy models conditioned by appropriate moments (2). The main difficulty is to define and estimate the moments which specify these Gibbs energies.

For stationary fields, whose probability distributions are invariant to translation, moments are usually computed with a Fourier transform, which diagonalizes the covariance matrix of the field. The resulting covariance eigenvalues are the Fourier power

**Competing Interest:** The authors declare no competing interest.

**Received:** September 17, 2023. **Accepted:** February 16, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

spectrum. However, capturing non-Gaussian properties requires to go beyond second-order moments of the field. Third- and fourth-order Fourier moments are called bispectrum and trispectrum. For a cubic  $d$ -dimensional stationary field of length  $L$ , the number of coefficients in the raw power spectrum, bispectrum, and trispectrum are  $O(L^d)$ ,  $O(L^{2d})$ , and  $O(L^{3d})$ , respectively. High-order moment estimators have high variance and are not robust, especially for non-Gaussian fields, because of potentially rare outliers which are amplified. It is thus very difficult to accurately estimate these high-order Fourier spectra from a few samples. Accurate estimations require considerably reducing the number of moments and eliminating the amplification effect of high-order moments.

Local conservation laws for mass, energy, momentum, charge, etc. result in continuity equations or transport equations. The resulting probability distributions of the underlying processes thus are typically regular to deformations that approximate the local transport. These properties have motivated many researchers to make use of a wavelet transform as opposed to a Fourier transform, which provides localized descriptors. Most statistical studies have concentrated on second-order and marginal wavelet moments, e.g. (3–5), which fail to capture important non-Gaussian properties of a field. Other studies (6) use wavelet operator for interpretation with application to cosmological parameter inference, but rely on a trained neural network model.

In recent years, new representations have been constructed by applying pointwise nonlinear operators on the wavelet transforms to recover their non-Gaussian information. The scattering transform, for instance, is a representation that is built by cascading wavelet transforms and nonlinear modulus (7, 8). This representation has been used in astrophysics and cosmology to study the interstellar medium (9, 10), weak-lensing fields (11, 12), galaxy surveys (13–15), and radio observations (16) (readers with physics background may find this review (17) useful). Other representations, which are built from covariances of phase harmonics of wavelet transforms (18, 19), have also been used to model different astrophysical processes (20–22). Such models, which can be built from a single image, have in turn enabled the development of new component separation methods (23, 24), which can be directly applied to observational data without any particular prior model of the components of a mixture (25, 26).

These models however suffer from a number of limitations: they are not very good at reproducing vortices or long thin filaments, and they require an important number of coefficients to capture dependencies between distant scales, as well as angular dependencies. Building on those previous works, Morel et al. (27) introduced a reduced scattering spectra representations for time series by leveraging scale invariance. In this paper, we present the scattering spectra for datasets with dimensions more than one, which is a low-dimensional representation able to efficiently describe a wide range of non-Gaussian processes encountered in physics. In particular, we show that it is possible to take into account the intrinsic regularity of physical fields and dramatically reduce the size of such representations. The first part of the paper presents maximum entropy models and the scattering spectra statistics, as well as their dimensional reduction. The second part of the paper presents a quantitative validation of these models on various 2D multiscale physical fields and discuss their limitations.

## Notations

$v^*$  is the complex conjugate of a scalar  $v$ .  $\text{Ave}_i$  averages values indexed by  $i$  in a finite set.  $\hat{x}[k]$  is the Fourier transform of  $x[u]$ , whether  $u$  is a continuous variable in  $\mathbb{R}^d$  or belongs to finite

periodic lattice.  $\mathbb{E}\{\Phi(x)\}$  is the expectation of  $\Phi(x)$  according to the probability distribution  $p(x)$  of a vector  $x$ .  $\log$  stands for base 2 logarithm.

## Methods

### Gibbs energy of stationary fields

We first review the properties of Gibbs energies resulting from maximum entropy models conditioned by moment values (28–30). We write  $x[u]$  a field, where the site index  $u$  belongs to a cubic  $d$ -dimensional lattice of size  $L$ . It results that  $x \in \mathbb{R}^{L^d}$ .

Assume that  $x \in \mathbb{R}^{L^d}$  has a probability density  $p(x)$  and consider Gibbs energy models linearly parameterized by a vector  $\theta = \{\theta_m\}_{m \leq M}$  over a potential vector  $\Phi(x) = \{\Phi_m(x)\}_{m \leq M}$  of dimension  $M$

$$U_\theta(x) = \langle \theta, \Phi(x) \rangle = \sum_{m=1}^M \theta_m^* \Phi_m(x). \quad (1)$$

They define exponential probability models

$$p_\theta(x) = Z_\theta^{-1} e^{-\langle \theta, \Phi(x) \rangle}. \quad (2)$$

The model class is thus defined by the potential vector  $\Phi(x)$ , which needs to be chosen appropriately.

If it exists, the maximum entropy distribution conditioned by  $\mathbb{E}\{\Phi(x)\}$  is a  $p_{\theta_0}$  which belongs to this model class. It has a maximum entropy  $H(p_{\theta_0}) = - \int p_{\theta_0}(x) \log p_{\theta_0}(x) dx$  under the expected value condition

$$\int \Phi(x) p_{\theta_0}(x) dx = \mathbb{E}\{\Phi(x)\}. \quad (3)$$

In statistical physics,  $p_{\theta_0}$  is a macrocanonical model defined by a vector  $\mathbb{E}\{\Phi(x)\}$  of observables. One can verify that  $\theta_0$  also minimizes the Kullback–Liebler divergence within the class

$$D(p \| p_{\theta_0}) = \int p(x) \log \frac{p(x)}{p_{\theta_0}(x)} dx = H(p_{\theta_0}) - H(p). \quad (4)$$

The main topic of the paper is to specify  $\Phi(x)$  in order to define accurate maximum entropy models for large classes of physical fields, which can be estimated from a small number  $n$  of samples  $\bar{x}_i$ . In this section, we suppose that  $n = 1$ . Reducing the model error given by Eq. 4 amounts to defining  $\Phi$  which reduces the excess entropy of the model. This can be done by enriching  $\Phi(x)$  and building very high-dimensional models. However, we must also take into account the empirical estimation error of  $\mathbb{E}\{\Phi(x)\}$  by  $\Phi(\bar{x}_1)$ , measured by  $\mathbb{E}\{\|\Phi(x) - \mathbb{E}\{\Phi(x)\}\|^2\}$ .

In this paper, macrocanonical models are approximated by microcanonical models, which have a maximum entropy over a microcanonical set of width  $\epsilon > 0$

$$\Omega_\epsilon = \{x \in \mathbb{R}^{L^d} : \|\Phi(x) - \Phi(\bar{x}_1)\|^2 \leq \epsilon\}. \quad (5)$$

**Supplementary material A** reviews a sampling algorithm for such model. It also explains how to extend the definition of  $\Omega_\epsilon$  for  $n > 1$  samples  $\bar{x}_i$  by replacing  $\Phi(\bar{x}_1)$  by  $\text{Ave}_i \Phi(\bar{x}_i)$ . If  $\Phi(x)$  concentrates around  $\mathbb{E}\{\Phi(x)\}$  then the microcanonical model converges to the macrocanonical model when the system length  $L$  goes to  $\infty$  and  $\epsilon$  goes to 0. The concentration of  $\Phi(x)$  generally imposes that its dimension  $M$  is small relatively to the dimension  $L^d$  of  $x$ . The choice of  $\Phi(x)$  must thus incorporate a tradeoff between the model error [4] and the distance between micro and macrocanonical distributions.

### Fourier polyspectra

Gaussian random fields are maximum entropy models conditioned on first- and second-order moments. The potential vector

$\Phi(x)$  is then an empirical estimator of first- and second-order moments of  $x$ . For stationary fields, there is only one first-order moment  $\mathbb{E}[x[u]]$  which can be estimated with an empirical average<sup>a</sup> over  $u$ :  $\text{Ave}_u x[u]$ . Similarly, the covariance matrix  $\mathbb{E}\{x[u]x[u']\}$  only depends on  $u - u'$ , so only the diagonal coefficients in Fourier space are informative, which are called the power spectrum,

$$\mathbb{E}\{\hat{x}[k]\hat{x}[k']^*\} \text{ with } k = k'. \quad (6)$$

The off-diagonal elements vanish because of phase cancellation under all possible translations, which means the second-order moments treat Fourier coefficients independently, and cannot describe relations or dependence between them. The diagonal elements, which can also be written as  $|\hat{x}[k]|^2$ , can be estimated from a single sample  $x$  by averaging  $|\hat{x}[k]|^2$  over frequency bins that are large enough to reduce the estimator variance. A uniform binning and sampling along frequencies results in power spectrum estimators with  $O(L^d)$  elements, so the Gaussian model is compact and feasible.

However, the Gaussian random field model has limited power to describe complex structures. The majority of fields encountered in scientific research are not Gaussian. Non-Gaussianity usually means dependence between Fourier coefficients at different frequencies. The traditional way goes to higher orders moments of  $\hat{x}$ , the polyspectra (31), where phase cancellation implies that for stationary fields, only the following moments are informative,

$$\mathbb{E}\{\hat{x}[k_1] \dots \hat{x}[k_n]\} \text{ with } k_1 + \dots + k_n = 0, \quad (7)$$

while other moments are zero. These polyspectra at order  $n > 2$  capture dependence between  $n - 1$  independent frequencies. As the leading term, the Fourier bispectrum specifies the nonzero third-order moments and has  $O(L^{2d})$  coefficients. However, bispectrum is usually not sufficient to characterize non-Gaussian fields. For example, it vanishes if the field distribution is symmetric  $p(x) = p(-x)$ . One must then estimate fourth-order Fourier moments, the trispectrum, which has  $O(L^{3d})$  coefficients.

There are two main problems for the polyspectra coefficients to become proper potential functions  $\Phi(x)$  in the maximum entropy models. First, the number of coefficients increases sharply with the order. Second, high-order moments are not robust and difficult to estimate from a few realizations (32). For random fields with a heavy tail distribution, which is ubiquitous in complex systems (33–37), higher order moments may not even exist. Those two problems are common for high-order moments and have been demonstrated in real-world applications (38, 39). In the following two sections, we introduce modifications to this approach to solve those problems.

## Wavelet polyspectra

Many physical fields exhibit multiscale structures induced by nonlinear dynamics, which implies regularity of  $p(x)$  in frequency. The wavelet transform groups Fourier frequencies by wide logarithmic bands, providing a natural way to compress the Fourier polyspectra. The compression not only reduces the model size but also improves estimator convergence. We use the wavelet transform to compute a compressed power spectrum estimate, as well as a reduced set of  $O(\log^2 L)$  third- and  $O(\log^3 L)$  fourth-order wavelet moments, allowing for efficient estimation of the polyspectra.

## Wavelet transform

A wavelet is a localized wave-form  $\psi[u]$  for  $u \in \mathbb{R}^d$  which has a zero average  $\int_{\mathbb{R}^d} \psi[u] du = 0$ . We shall define complex-valued wavelets  $\psi[u] = g[u] e^{i\zeta \cdot u}$ , where  $g[u]$  is a real window whose Fourier transform  $\hat{g}[k]$  is centered at  $k = 0$  so that  $\hat{\psi}[k] = \hat{g}[k - \zeta]$  is localized in the neighborhood of the frequency  $\zeta$ . Figure S1 shows  $\psi$  and  $\hat{\psi}$  for a  $d = 2$  dimensional Morlet wavelet described in [supplementary material B](#). The wavelet transform is defined by rotating  $\psi[u]$  with a rotation  $r$  in  $\mathbb{R}^d$  and by dilating it with dyadic scales  $2^j > 1$ . It defines

$$\psi_\lambda[u] = 2^{-jd} \psi[2^{-j} r^{-1} u] \text{ with } \lambda = 2^{-j} r \zeta. \quad (8)$$

Its Fourier transform is  $\hat{\psi}_\lambda[k] = \hat{g}[2^j r^{-1}(k - \zeta)]$ , which is centered at the frequency  $\lambda$  and concentrated in a ball whose radius is proportional to  $2^{-j}$ .

To decompose a field  $x[u]$  defined over a grid of width  $L$ , the wavelet is sampled on this grid. Wavelet coefficients are calculated as convolutions with periodic boundary conditions

$$Wx[u, \lambda] = x \star \psi_\lambda[u] = \sum_{u'} x[u'] \psi_\lambda[u - u']. \quad (9)$$

It measures the variations of  $x$  in a spatial neighborhood of  $u$  of length proportional to  $2^j$ , and it depends upon the values of  $\hat{x}$  in a frequency neighborhood of  $k = \lambda$  of length proportional to  $2^{-j}$ . The scale  $2^j$  is limited to  $1 \leq j \leq J$ , and for practical application to fields with a finite size  $L$ , the choice of  $J$  is limited by  $J < \log L$ . Left part of Fig. 1 illustrates the wavelet transform of an image.

The rotation  $r$  is chosen within a rotation group of cardinal  $R$ , where  $R$  does not depend on  $L$ . Wavelet coefficients need to be calculated for  $R/2$  rotations because  $Wx[u, -\lambda] = Wx[u, \lambda]^*$  for real fields. In  $d = 2$  dimensions, the  $R$  rotations have an angle  $2\pi\ell/R$ , and we set  $R = 8$  in all our numerical applications, which boils down to four different wavelet orientations. The total number of wavelet frequencies  $\lambda$  is  $RJ = O(\log L)^b$  as opposed to  $L^d$  Fourier frequencies.

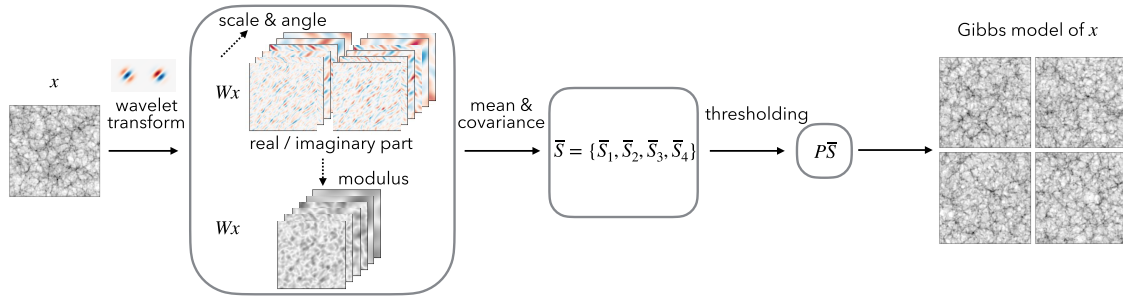
A wavelet transform is also stable and invertible if  $\psi$  satisfies a Littlewood-Paley condition, which requires an additional convolution with a low-pass scaling function  $\psi_0$  centered at the frequency  $\lambda = 0$ . The specifications are detailed in [supplementary material B](#).

## Wavelet power spectrum

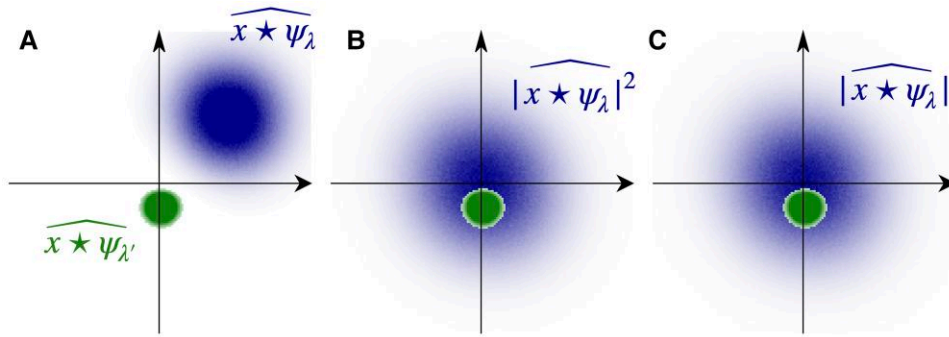
Given scaling regularity, one can compress the  $O(L^d)$  power spectrum coefficients into  $RJ = O(\log L)$  coefficients using a logarithmic binning defined by wavelets. This is obtained by averaging the power spectrum with weight functions as the Fourier transform of wavelets, which are band-pass windows,  $\text{Ave}_k (\mathbb{E}\{|\hat{x}[k]|^2\} |\hat{\psi}_\lambda[k]|^2)$ . The limited number of wavelet power spectrum coefficients has reduced estimation variance. In fact, they are also the diagonal elements of the wavelet covariance matrix,  $Wx[u, \lambda] Wx[u, \lambda]^* = |Wx[u, \lambda]|^2$ , therefore an empirical estimation can also be written as an average over  $u$ :

$$M_2 = \text{Ave}_u |Wx[u, \lambda]|^2. \quad (10)$$

Similar to the power spectrum, phase cancellation due to translation invariance means that the off-diagonal blocks, i.e. the cross-correlations between different wavelet frequency bands are nearly zero because the support of two wavelets  $\hat{\psi}_\lambda$  and  $\hat{\psi}_{\lambda'}$  are almost disjoint, as illustrated in Fig. 2a.



**Fig. 1.** Steps to build a feasible model for a random field  $x$  from only one or a few realizations. We first build a low-dimensional representation  $\Phi(x)$  of the random field, which specifies a maximum entropy model. The representation  $\Phi(x)$  is obtained by conducting the wavelet transform  $Wx$  and its modulus  $|Wx|$ , and then computing the means and covariance of all wavelet channels ( $Wx$ ,  $|Wx|$ ). Such a covariance matrix is further binned and sampled using wavelets to reduce its dimensionality, which is called the scattering spectra  $\bar{S}(x)$ . Finally, These scattering spectra are renormalized and reduced in dimension by thresholding its Fourier coefficients along rotation and scale parameters  $\Phi(x) = P\bar{S}$ , making use of the regularity properties of the field. For many physical fields, this representation can be as small as only around  $\sim 10^2$  coefficients for a  $256 \times 256$  field.



**Fig. 2.** a) For  $\lambda \neq \lambda'$ , the Fourier supports of  $x \star \psi_\lambda$  and  $x \star \psi_{\lambda'}$  typically do not overlap. b) The Fourier support of  $|x \star \psi_\lambda|^2$  is twice larger and centered at 0 and hence overlaps with  $x \star \psi_{\lambda'}$  if  $|\lambda'| \leq |\lambda|$ . c) The Fourier support of  $|x \star \psi_\lambda|$  is also centered at 0 and hence overlaps with  $x \star \psi_{\lambda'}$  if  $|\lambda'| < |\lambda|$ .

### Selected third- and fourth-order wavelet moments

One may expect to compress the polyspectra in a similar manner with a wavelet transform, taking advantage of the regularities of the field probability distribution. However, it is nontrivial to logarithmically bin the polyspectra because more than one independent frequency is involved and the phase cancellation condition needs to be considered.

To solve this problem, let us revisit the phase cancellation of two frequency bands, which causes their correlation to be zero,

$$\mathbb{E}\{Wx[u, \lambda] Wx[u', \lambda']^*\} \sim 0, \quad (11)$$

for  $\lambda \neq \lambda'$ . To create a nonzero correlation, we must realign the support of  $Wx[u, \lambda]$  and  $Wx[u', \lambda']$  in Fourier space through nonlinear transforms. As shown in Fig. 2b, we may apply a square modulus to one band in the spatial domain, which recenters its frequency support at origin. Indeed,  $|x \star \psi_\lambda|^2 = (x \star \psi_\lambda)(x \star \psi_\lambda)^*$  has a Fourier support twice as wide as that of  $x \star \psi_\lambda$ , and will overlap with another wavelet band with lower frequency than  $\lambda$ . The transformed fields  $|x \star \psi_\lambda|^2$  can be interpreted as maps of locally measured power spectra. Correlating this map with another wavelet band  $x \star \psi_{\lambda'}$  gives some third-order moments

$$\mathbb{E}\{|Wx|^2[u, \lambda] Wx[u', \lambda']^*\} \quad (12)$$

that are a priori nonzero. Furthermore, for wide classes of multiscale processes having a regular power spectrum, it suffices to only keep the coefficients at  $u = u'$  because of random phase fluctuation (see [supplementary material B](#)). For stationary

random fields, they can be estimated with an empirical average over  $u$ ,

$$M_3 = A \text{Ave}_u (|Wx|^2[u, \lambda] Wx[u, \lambda']^*). \quad (13)$$

Now we obtain a set of statistics characterizing the dependence of Fourier coefficients in two wavelet bands in a collective way, which are selected third-order moments. They can be interpreted as a logarithmic frequency binning of certain bispectrum coefficients. There are about  $R^2 j^2 = O(\log^2 L)$  such coefficients, which is a substantial compression compared to the  $O(L^{2d})$  full bispectrum coefficients. Similarly, we consider the cross correlation between two wavelet bands both transformed by the square modulus operation and obtain a wavelet binning of fourth-order moments,

$$\mathbb{E}\{|Wx[u, \lambda]|^2 |Wx[u', \lambda']|^2\} - \mathbb{E}\{|Wx[u, \lambda]|^2\} \mathbb{E}\{|Wx[u', \lambda']|^2\}. \quad (14)$$

For stationary fields, this covariance only depends on  $u - u'$ . A further reduction of such a large covariance function is possible because its Fourier transform over  $u - u'$  has two properties. First, it typically does not have higher frequency components than the initial wavelet transforms involved (see Fig. 2) as the phase fluctuations have been eliminated by the square modulus, and second, for fields with multiscale structures, it is regular and can be approximated with another logarithmic frequency binning. Thus, we can compress the large covariance function with a second wavelet transform, and estimate it by an empirical average over  $u$ :

$$M_4 = A \text{Ave}_u (|W|Wx|^2[u, \lambda, \gamma] |W|Wx|^2[u, \lambda', \gamma]^*), \quad (15)$$

where  $(W|Wx|^2)[u, \lambda, \gamma] = |x \star \psi_\lambda|^2 \star \psi_\gamma[u]$ , and the central frequencies of the second wavelets verifies  $|\lambda| \geq |\lambda'| > |\gamma|$ . There are about  $R^3 J^3 = O(\log^3 L)$  such coefficients, which is also a substantial compression compared to the  $O(L^{3d})$  full trispectrum coefficients.

## Scattering spectra

In general, the estimation of high-order moments has a high variance because high-order polynomials amplify the effect of outliers. An interesting idea learned from the scattering transform approach (7, 8) is that the multiplication  $(Wx)(W^*x) = |Wx|^2$  used in the higher order moments [10, 13, 15] can be replaced by the wavelet modulus  $|Wx|$ , which produces qualitatively similar estimators but with improved robustness and better efficiency in presence of sparse structures<sup>c</sup> (17). The resulting moments after such a replacement only depend on the mean and covariance matrix of  $(Wx, |Wx|)$ , which are low-order transforms of the original field  $x$ .

Local statistics of wavelet modulus have been studied to analyze properties of image textures (41). Their mathematical properties have been analyzed to capture non-Gaussian characteristics of random fields (18, 19) in relation to scattering moments (7, 8). Scattering spectra have been defined on 1D time series (27), from the joint covariance of a wavelet transform and its modulus:  $(Wx, |Wx|)$ . We extend it to fields of arbitrary dimension  $d$  and length  $L$ , in relation to Fourier high-order moments, and define models of dimension  $O(\log^3 L)$ .

### First and second wavelet moments, sparsity

For non-Gaussian fields  $x$ , wavelet coefficients  $Wx[u, \lambda]$  define fields which are often sparse (42, 43). This is a non-Gaussian property that can be captured by first-order wavelet moments  $\mathbb{E}[|Wx[u, \lambda]|]$ . If  $x$  is a Gaussian random field then  $Wx[u, \lambda]$  remains Gaussian but complex-valued so, and we have  $\frac{\mathbb{E}[|Wx|^2]}{\mathbb{E}[|Wx|^2]} = \frac{\pi}{4}$ . This ratio decreases when the sparsity of  $Wx[u, \lambda]$  increases. The expected value of  $|Wx|$  is estimated by

$$S_1(x)[\lambda] = \text{Ave}_u |Wx[u, \lambda]| \quad (16)$$

and the ratio is calculated with the second-order wavelet spectrum estimator

$$S_2(x)[\lambda] = M_2(x)[\lambda] = \text{Ave}_u (|Wx|^2[u, \lambda]). \quad (17)$$

### Cross-spectra between scattering channels

Let us now replace  $|Wx|^2$  by  $|Wx|$  in the selected third- and fourth-order wavelet moments described in the previous section. The third-order moments [13] become  $\mathbb{E}[|Wx[u, \lambda]|Wx[u, \lambda']^*]$ . Such moments are a priori nonzero if the Fourier transforms of  $|Wx[u, \lambda]| = |x \star \psi_\lambda|$  and  $Wx[u, \lambda'] = x \star \psi_{\lambda'}$  overlap. This is the case if  $|\lambda'| < |\lambda|$  as illustrated in Fig. 2. Eliminating the square thus preserves nonzero moments which can capture dependencies between different frequencies  $\lambda$  and  $\lambda'$ . The third-order moment estimators given by Eq. 13 can thus be replaced by lower cross-correlations between  $|Wx|$  and  $Wx$  at  $|\lambda| \geq |\lambda'|$

$$S_3(x)[\lambda, \lambda'] = \text{Ave}_u (|Wx[u, \lambda]|Wx[u, \lambda']^*). \quad (18)$$

Replacing  $|Wx|^2$  by  $|Wx|$  in the fourth-order wavelet moments [15] amounts to estimating the covariance matrix of wavelet modulus fields  $|Wx|$ . As the  $u - u'$  dependency of this covariance can also be

characterized by a second wavelet transform, this amounts in turn to estimate the covariance of scattering transforms  $W|Wx|[u, \lambda, \gamma] = |x \star \psi_\lambda| \star \psi_\gamma[u]$

$$S_4(x)[\lambda, \lambda', \gamma] = \text{Ave}_u (W|Wx|[u, \lambda, \gamma]W|Wx|[u, \lambda', \gamma]^*), \quad (19)$$

for  $|\lambda| \geq |\lambda'| \geq |\gamma|$ . It provides a wavelet spectral estimation of the covariance of  $|Wx|$ .

Combining the moment estimators of Eqs. 16–19 defines a vector of scattering spectra

$$S(x) = (S_1(x), S_2(x), S_3(x), S_4(x)). \quad (20)$$

It provides a mean and covariance estimation of the joint wavelet and wavelet modulus vectors  $(Wx, |Wx|)$ . It resembles the second-, third-, and fourth-order Fourier spectra but has much fewer coefficients and better information concentration. Considering the conditions satisfied by  $\lambda, \lambda'$ , and  $\gamma$ , the exact dimension of  $S(x)$  is  $RJ + R^2J(J-1)/8 + R^3J(J^2-1)/48$ , of the order  $O(\log^3 L)$ .

### Renormalization

Scattering spectra coefficients must often be renormalized to improve the sampling of maximum entropy models. Indeed, multi-scale random processes often have a power spectrum that has a power law decay  $\mathbb{E}[|\hat{x}[k]|^2] \sim |k|^{-\eta}$  over a wide range of frequencies, long-range correlations corresponding to a strong decay from large to small scales. The wavelet spectrum also has a power-law decay  $\mathbb{E}[|Wx[u, \lambda]|^2] \sim |\lambda|^{-\eta}$ . This means that if we build a maximum entropy model with  $\Phi(x) = S(x)$  then the coordinate of  $\Phi(x)$  of low frequencies  $\lambda$  have a much larger amplitude and variance than at high frequencies. The microcanonical model is then dominated by low frequencies and is unable to constrain high-frequency moments. The same issue appears when computing the  $\theta_0$  parameters of a macrocanonical model defined in Eq. 3, for which it has been shown that renormalizing to 1 the variance of wavelet coefficients at all scales avoid numerical instabilities (44). Without such a normalization, the calculation of  $\theta_0$  parameters at different frequencies is ill-conditioned, which turns into a “critical slowing down” of iterative optimization algorithms. The proposed normalization is closely related to Wilson renormalization.

We renormalize the scattering spectra by the variance of wavelet coefficients,  $\sigma^2[\lambda] = \text{Ave}_u S_2(\bar{x}_i)[\lambda]$ , which can be estimated from a few samples. The renormalized scattering spectra are

$$\bar{S}(x) = (\bar{S}_1(x), \bar{S}_2(x), \bar{S}_3(x), \bar{S}_4(x)) \quad (21)$$

defined by

$$\begin{aligned} \bar{S}_1(x)[\lambda] &= \frac{S_1(x)[\lambda]}{\sigma[\lambda]}, & \bar{S}_2(x)[\lambda] &= \frac{S_2(x)[\lambda]}{\sigma^2[\lambda]} \\ \bar{S}_3(x)[\lambda, \lambda'] &= \frac{S_3(x)[\lambda, \lambda']}{\sigma[\lambda]\sigma[\lambda']}, & \bar{S}_4(x)[\lambda, \lambda', \gamma] &= \frac{S_4(x)[\lambda, \lambda', \gamma]}{\sigma[\lambda]\sigma[\lambda']}. \end{aligned} \quad (22)$$

The microcanonical models proposed in this paper are built from these renormalized statistics and/or their reduced version described below.

### Dimensionality reduction for physical fields

Although much smaller than the polyspectra representation, the scattering spectra  $\bar{S}$  representation still has a large size. Assuming isotropy and scale invariance of the field  $x$ , a first-dimensional reduction can be performed that relies on the equivariance properties of scattering spectra with respect to rotation and scaling (see [supplementary material C](#)). However, such invariances cannot be assumed in general. In this section, we propose to construct

a low-dimensional representation by only assuming regularity under rotation or scaling of the scales involved in the scattering spectra representation. A simplified version of such a dimensional reduction has been introduced in Ref. (9). We refer the reader to [supplementary material D](#) for technical details.

The goal of the reduction is to approximate the covariance coefficients  $\bar{S}_3$  and  $\bar{S}_4$ , the most numerous, using only a few coefficients. This can be seen as a covariance matrix estimation problem. To do so, we first use a linear transform to sparsify the covariance matrix and then perform a threshold clipping on the coefficients to reduce the representation. We consider a linear transform  $F\bar{S} = (\bar{S}_1, \bar{S}_2, F\bar{S}_3, F\bar{S}_4)$  with a predetermined linear transform  $F$  which stands for a 2D or 3D Fourier transform along all orientations, as well as a 1D cosine transform along scales, for  $\bar{S}_3$  and  $\bar{S}_4$ . For fields with statistical isotropy or self-similarity, all harmonics related to the action of global rotation and scaling on the field  $x$  should be consistent with zero, except for the zeroth harmonic. For general physical fields, we expect the statistics  $\bar{S}(x)$  to have regular variations to the action of rotation or scaling of the different scales involved in its computation, which implies that its Fourier harmonics  $F\bar{S}(x)$  have a fast decay away from the 0th harmonic and  $F\bar{S}(x)$  is a sparse representation.

Thresholding on a sparse representation is widely used in image processing for compression (45). We use threshold clipping on the sparse representation  $F\bar{S}$  to significantly reduce the size of the scattering spectra. Furthermore, when empirically estimating large but sparse covariance matrices such as  $F\bar{S}$ , thresholding provides Stein estimators (46) which have lower variance and are consistent, e.g. (47–50). As  $\bar{S}_1$  or  $\bar{S}_2$  are already small, we keep all of their coefficients.

There are different strategies available to set the threshold for clipping. We adopt a simple strategy which keeps those coefficients with  $\mu(F\bar{S}) > 2\sigma(F\bar{S})$ , where  $\mu(F\bar{S})$  and  $\sigma(F\bar{S})$  are the means and SDs of individual coefficients of  $F\bar{S}$ . These adaptive thresholding estimators achieve a higher rate of convergence and are easy to implement (49). With multiple realizations from simulations,  $\mu(F\bar{S})$  and  $\sigma(F\bar{S})$  can be estimated directly. In the case where only a single sample field is available,  $\sigma(F\bar{S})$  can be estimated from different patches of that sample field, e.g. (51). We call  $P\bar{S}$  the coefficients after thresholding projection:

$$P\bar{S} = (\bar{S}_1, \bar{S}_2, P\bar{S}_3, P\bar{S}_4) = \text{thresholding } F\bar{S}. \quad (23)$$

The compact yet informative set of scattering spectra  $P\bar{S}$  is the representation  $\Phi(x) = P\bar{S}(x)$  proposed in this paper to construct maximum entropy models.

## Numerical results

We have introduced maximum entropy models based on small subsets of  $O(\log^3 L)$  scattering spectra moments  $\bar{S}$  and projected moments  $P\bar{S}$ , claiming that it can provide accurate models of large classes of multiscale physical fields, and reproduce  $O(L^{3d})$  power spectrum, bispectrum and trispectrum Fourier moments. This section provides a numerical justification of this claim with five types of 2D physical fields from realistic simulations. In order to reduce the variance of the validation statistics, we consider in this section a model estimated on several realizations of a field. However, our model also produces convincing realizations when estimated on a single realization (see Fig. S2 for a visual assessment). All computations are reproducible with the software available on [https://github.com/SihaoCheng/scattering\\_transform](https://github.com/SihaoCheng/scattering_transform).

## Dataset of physical fields

We use five 2D physical fields to test the maximum entropy models. The five fields are chosen to cover a range of properties in terms of scale dependence, anisotropy, sparsity, and morphology:

- (A) *Cosmic lensing*: simulated convergence maps of gravitational lensing effects induced by the cosmological matter density fluctuations (52, 53).
- (B) *Dark matter*: logarithm of 2D slices of the 3D large-scale distribution of dark matter in the Universe (54). The logarithm allows for the filamentary cosmic web structures to stand out and thus increases the morphological diversity of our examples, which we discuss more in [supplementary material G](#).
- (C) *2D turbulence*: turbulence vorticity fields of incompressible fluid stirred at the scale around 32 pixels, simulated from 2D Navier–Stokes equations (55).
- (D) *Magnetic turbulence*: column density of 3D isothermal magnetic–hydrodynamic turbulent simulations (9). The field is anisotropic due to a mean magnetic field in the horizontal direction.
- (E) *Anisotropic turbulence*: 2D slices of a set of 3D turbulence simulations (56, 57). To create anisotropy, we have squeezed the fields along the vertical direction.

These simulations are sampled on a grid of 256×256 pixels with periodic boundary conditions<sup>d</sup> and normalized to have zero mean and unity SD, respectively. Samples of each field are displayed in the first row of Fig. 3. To clearly show the morphology of small-scale structures, we zoom in to a 128×128 region.

## Model description and visual validation

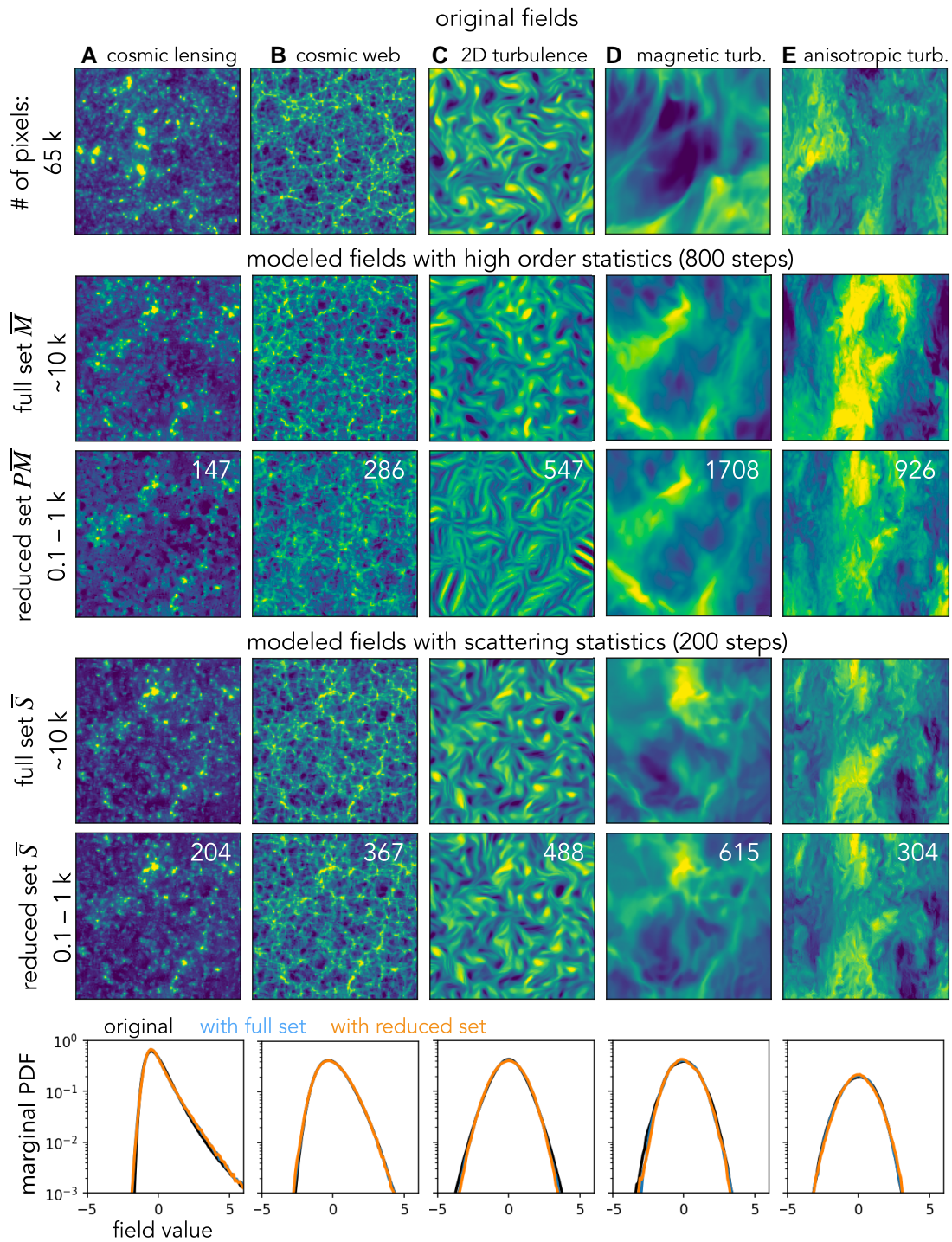
We fit our maximum entropy model using wavelet polyspectra and scattering spectra, respectively, with the following constraint:

$$\|\text{Ave}_j \Phi(x_j) - \text{Ave}_i \Phi(\bar{x}_i)\|^2 \leq \epsilon, \quad (24)$$

where the second average is computed on an ensemble of 100 realizations  $\bar{x}_i$  for each physical simulation (for field D, we use only 20 realizations due to the availability of simulations), and the field generation is performed simultaneously for 10 fields  $x_j$ , making our microcanonical model closer to its macrocanonical limit. The microcanonical sampling algorithm is described in [supplementary material A](#).

Examples of field generation results are given in Fig. 3. The second row shows samples generated based on the high-order normalized wavelet moments  $\Phi(x) = \bar{M}(x) = (\bar{M}_2(x), \bar{M}_3(x), \bar{M}_4(x))$ , where  $\bar{M}_2 = \bar{S}_2$ ,  $\bar{M}_3(x)[\lambda, \lambda'] = \frac{M_3(x)[\lambda, \lambda']}{\sigma^2[\lambda]\sigma^2[\lambda']}$ , and  $\bar{M}_4(x)[\lambda, \lambda'] = \frac{M_4(x)[\lambda, \lambda']}{\sigma^2[\lambda]\sigma^2[\lambda']}$  are defined similarly to  $\bar{S}$  in Eq. 22. For the choice of wavelets, we use  $J = 7$  dyadic scales, and we set  $R = 8$  which samples 4 orientations within  $\pi$ , resulting in  $\dim \bar{M} = 11$ , 677 coefficients for  $\bar{M}$ . The third row in Fig. 3 shows results from a reduced set  $\Phi(x) = P\bar{M}(x)$ , which is a  $2\sigma$  Fourier thresholded representation of  $\bar{M}$  defined in exactly the same way as  $P\bar{S}$  in Eq. 23. The thresholding yields  $\dim P\bar{M} = 147, 286, 547, 1,708, 926$  for fields A–E, respectively. A visual check shows that these models fail to recover all morphological properties in our examples especially when a thresholding reduction is applied. This issue is a manifestation of the numerical instability of high-order moments.

In the fourth row, we present sample fields modeled with the scattering spectra  $\bar{S}$  with  $\dim P\bar{S} = 11, 705$  for  $J = 7$  and  $R = 8$ . A visual check reveals its ability to restore coherent spatial



**Fig. 3.** Visual comparison of realistic physical fields and those sampled from maximum entropy models based on wavelet higher order moments  $\bar{M}$  and wavelet scattering spectra  $\bar{S}$  statistics. The first row shows five example fields from physical simulations of cosmic lensing, cosmic web, 2D turbulence, magnetic turbulence, and squeezed turbulence. The second and third rows show syntheses based on the selected high-order wavelet statistics estimated from 100 realizations. They are obtained from a microcanonical sampling with 800 steps. The fourth and fifth rows show similar syntheses based on the scattering spectra statistics, with only 200 steps of the sampling run. This figure shows visually that the scattering spectra can model well the statistical properties of morphology in many physical fields, while the high-order statistics either fail to do so or converge at a much slower rate. To clearly show the morphology of structures at small scales, we show a zoom-in of  $128 \times 128$  pixels regions. Finally, to quantitatively validate the goodness of the scattering model, we show the marginal PDF (histogram) comparison in the last row.

structures including clumps, filaments, curvy structures, etc. The low-order nature and numerical stability of  $\bar{S}$  also significantly fasten the sampling compared to the high-order moments  $\bar{M}$  (200 vs. 800 steps to converge). The last row shows sample fields modeled by a much smaller set  $\bar{P}\bar{S}$ , which has

$\dim \bar{P}\bar{S} = 204, 364, 489, 615, 304$  coefficients for fields A-E, respectively. This model is  $\sim 10^2$  times smaller, while generating samples visually indistinguishable from the full set model with  $\Phi(x) = \bar{S}(x)$ . In addition, the ratio between the dimensionality of the field  $\dim x = L^d$  (the number of pixels) and the model  $\dim \Phi$



is more than 100. For interested readers, we also present the improvement of modeling from using power spectrum alone to the full scattering spectra in [supplementary material F](#).

## Statistical validation

We now quantify the consistency between the scattering spectra models and the original fields using a set of validation statistics  $V(x)$  defined below, including marginal PDF, structure functions  $SF_n$ , power spectrum  $P$ , and normalized bispectrum  $\bar{B}$  and trispectrum  $\bar{T}$ . The validation statistics are shown in Figs. 3 and 4, where black curves represent the expected value  $\mu_{\text{original}}$  of these statistics, estimated from 100 realizations  $\bar{x}_i$  of the original simulated fields (except for field D for which we have only 20 realizations). Gray regions around the black curves represent the SDs  $\sigma_{\text{original}}$  of those statistics estimated on the original fields. Blue curves are statistics  $\mu_{\bar{S}, \text{model}}$  estimated on fields modeled with  $\bar{S}$ . Similarly,  $\mu_{\bar{P}\bar{S}, \text{model}}$  are estimated on fields modeled with the reduced set  $\bar{P}\bar{S}$ . Both these averages are estimated from the 10 fields simultaneously sampled from the corresponding microcanonical models.

The marginal probability distribution function (PDF) is measured as the histogram of sample fields and shown in Fig. 3. It averages out all spatial information and keeps only the overall asymmetry and sparsity properties of the field. The marginal information is not explicitly encoded in the scattering spectra, but for all the five physical fields we examine here, it is recovered even with the reduced model  $\bar{P}\bar{S}$ , where only  $\sim 10^2$  scattering spectra coefficients are used.

Given that the high dimensionality of the full set of polyspectra coefficients, as well as the computational cost of estimating them properly, we adopt an isotropic shell binning for the power spectrum, bispectrum, and trispectrum. Although this reduces the number of coefficients as well as their variance, working with isotropic statistics prevents the characterization of anisotropic features, for instance in fields D and E, unlike with scattering spectra. Validation results with these isotropic polyspectra are given in Fig. 4.

The shell binning is defined as follow. We first divide the Fourier space into 10 annuli with the frequencies linearly spaced from 0 to 0.4 cycles/pixel. Then, we average the power and poly spectra coefficients coming from the same annulus combinations. For instance, the power spectrum yields

$$P[i] = \text{Ave}_{k \text{ in annuli } i} \hat{x}[k]\hat{x}[-k]. \quad (25)$$

To decorrelate the information from the power spectrum and higher orders, we normalized the binned bi- and tri-spectra by  $P[i]$ :

$$\bar{B}[i_1, i_2, i_3] = \frac{\text{Ave}_{k_n \text{ in annuli } i_n} \hat{x}[k_1]\hat{x}[k_2]\hat{x}[k_3]}{\sqrt{P[i_1]P[i_2]P[i_3]}}, \quad (26)$$

$$\bar{T}[i_1, i_2, i_3, i_4] = \frac{\text{Ave}_{k_n \text{ in annuli } i_n} \hat{x}[k_1]\hat{x}[k_2]\hat{x}[k_3]\hat{x}[k_4]}{\sqrt{P[i_1]P[i_2]P[i_3]P[i_4]}}, \quad (27)$$

where the  $k_n$   $d$ -dimensional wave-vectors are respectively averaged in the  $i_n^{\text{th}}$  frequency annuli, and satisfy  $\sum_n k_n = 0$ . To clearly reveal the diversity of different type of physical fields, the trispectrum  $\bar{T}$  coefficients shown in Fig. 4 are subtracted by the reference value of Gaussian white noise, evaluated numerically on 1,000 independent realizations. Details about the numbers and the ordering of  $\bar{B}$  and  $\bar{T}$  are given in [supplementary material E](#).

In Fig. 4, we also show the validation with structure functions, which are  $n$ th order moments of the field increments as a function of the position separation  $\Delta u$

$$SF_n[\Delta u] = \text{Ave}_u |x[u] - x[u + \Delta u]|^n. \quad (28)$$

In our 2D case, we further average over  $\Delta u$  with different orientations to obtain a structure function only depending on the magnitude of the separation  $|\Delta u|$ . Initially proposed by Kolmogorov for the study of turbulent flows (58), they are widely used to analyze non-Gaussian properties of multiscale processes (59).

We quantify the discrepancy between the model and original field distributions by the outlier fraction of validation statistics outside the  $2\sigma$  range,

$$|\mu_{\text{model}} - \mu_{\text{original}}|/\sigma_{\text{original}} > 2. \quad (29)$$

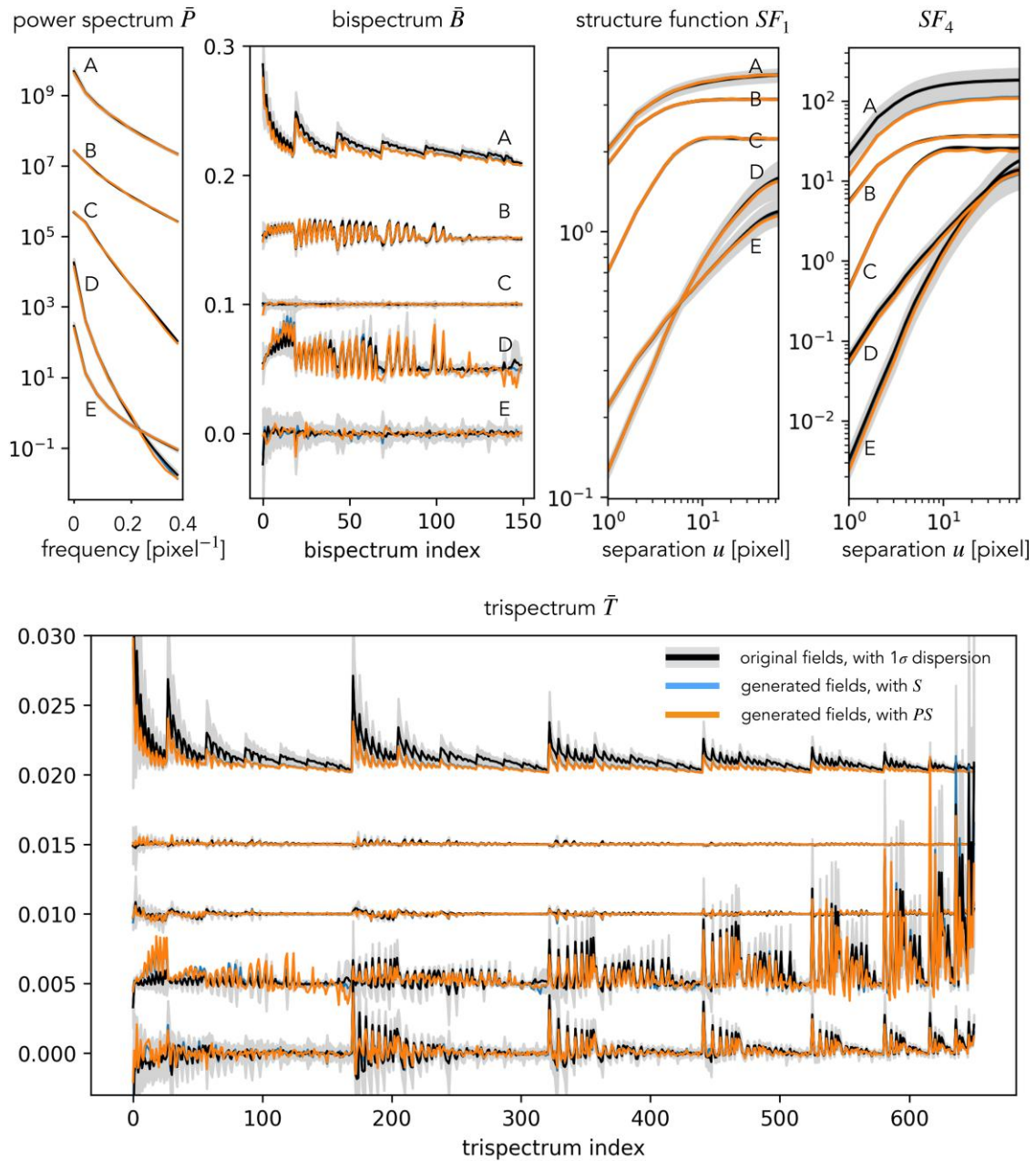
For each of the five types of fields, we observe the following fractions. The binned power spectrum has fractions of  $P$ : 0, 0, 20, 0, 0% for the models using all  $\bar{S}$  statistics and 0, 10, 40, 10, 0% for the thresholding models with  $\bar{P}\bar{S}$ . The power spectrum deviation of field C is likely caused by the longer convergence steps required by smooth fields, as our generative models start from white noise with strong small-scale fluctuations. Indeed increasing the steps to 800 reduces the outlier fraction of the  $\bar{P}\bar{S}$  model to 10%. For  $\bar{B}$  and  $\bar{T}$ , the outlier fractions are all below 5% except for the models of field A, where the bispectrum coefficients have 13% of outliers. Those outliers all have the smallest scale involved, and disappear if the high-frequency cut is moved from 0.4 to 0.35 cycles/pixel. The low fractions demonstrate consistency between our maximum entropy models and ensembles of the original physical fields.

For field A, a similar deviation is also observed in high-order structure functions. For this field, it can be seen from Fig. 4 that even though many coefficients are not defined as outliers, they all tend to have a lower value than the original ones. This effect may originate from the log-normal tail of the cosmic density field (35), whose Gibbs potential includes terms in the form of  $\log x$ , in contrast to the form of  $|x|$  in scattering spectra or  $x^n$  in high-order statistics. However, regardless of this difficulty, these outliers are all still within a  $3\sigma$  range, demonstrating that the scattering spectra provide a good approximation though not exact model for fields with such heavy tails.

The marginal PDF, structure functions, power spectrum, and polyspectra probe different aspects of the random field  $p(x)$ . The polyspectra especially probe a huge variety of feature configurations. For all the validation statistics, we observe general agreement between the model and original fields. Such an agreement is a nontrivial success of the scattering spectra model, as those statistics are not generically constrained by the scattering spectra for arbitrary random fields. They indeed significantly differ from the scattering spectra in the way they combine spatial information at different frequencies and in the nonlinear operation adopted. The agreement implies, as we have argued, that symmetry and regularity can be used as strong inductive bias for physical fields and the scattering spectra, with those priors build-in, can efficiently and robustly model physical fields.

## Visual interpretation of scattering spectra coefficients

The key advantage of the scattering spectra compared to usual convolutional neural networks is their structured nature: their computation corresponds to the combination of known scales and orientations in a fixed way. Beyond the limited number of symmetries, the structured nature of the scattering spectra



**Fig. 4.** Validation of the scattering maximum entropy models for the five physical fields A–E by various test statistics. The curves for field E represent the original statistics and those for A–D are shifted upwards by an offset. In general, our scattering spectra models well reproduce the validation statistics of the five physical fields.

allows us to both quantify and interpret the morphology of structures, which is one of the original goals to design these statistics.

The values of scattering spectra can be shown directly (see Fig. S3) to analyze non-Gaussian properties of the field. Moreover, the meaning of its coefficients can also be visualized through our maximum entropy generative models. As one gradually changes the value of some summary statistics, the morphology of structures in the generated fields also changes. A similar exploration for a smaller set of scattering transform coefficients has been explored in Ref. (17), and we show such results with the much more expressive scattering spectra coefficients in Fig 5. Such exploration using synthesis is also similar to the feature visualization efforts for convolutional neural networks (60).

The central panel is a realization of field B from physical simulations. The other four panels are generated fields with two

collective modifications of the scattering spectra: the vertical direction shows the effect of multiplying all  $\bar{S}_3$  and  $\bar{S}_4$  coefficients by a factor of 1/3 or 3. It indicates that the amplitude of  $\bar{S}_3$  and  $\bar{S}_4$  controls the overall non-Gaussian properties of the field and in particular the sparsity of its structures. The horizontal direction corresponds to adjusting the orientation dependence. We set the coefficients with parallel wavelet configurations (i.e.  $\bar{S}_3[|\lambda|, |\lambda'|, l_1 = l_2]$  and  $\bar{S}_4[|\lambda|, |\lambda'|, |\gamma|, l_1 = l_2 = l_3]$ ) as references and keep them unchanged. Then, we make the difference from other coefficients to those references to be 2 times or  $-2$  times the original difference. Visually, it controls whether structures are more point-like or more curvy-like in the field. In this experiment, the generated field is initialized with the original field instead of white noise, in order to clearly show the correspondence between the field structure and scattering spectra coefficients.

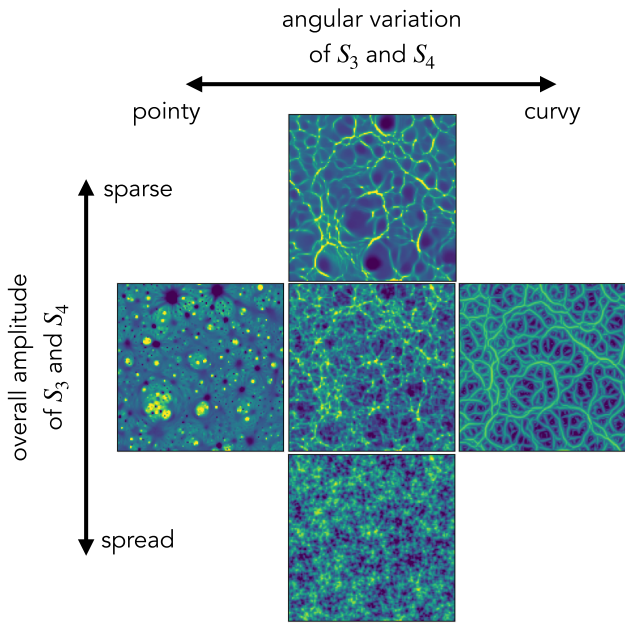
### Application to identifying symmetry

As an expressive representation whose coefficients are equivariant under standard group transformation, the scattering spectra can also be used to detect and identify the various statistical invariances commonly present in physical fields. Besides the aforementioned rotation and scaling invariance, more can also be included, such as the flipping of coordinate or field values.

The simplest way to check asymmetry to a transformation like rotation or flip is to check if the scattering spectra  $S$  are changed after applying such a transform. A more sophisticated way that can also quantify partial symmetries is to linearly decompose  $\bar{S}$  into symmetric and asymmetric parts and then compute the fraction of asymmetric coefficients surviving the thresholding reduction. We further normalize this fraction by that in the full set to eliminate the dependence on image size:

$$\text{asymmetry index} = \frac{\dim(P\bar{S}_{\text{asym}})}{\dim(P\bar{S})} / \frac{\dim(\bar{S}_{\text{asym}})}{\dim(\bar{S})}. \quad (30)$$

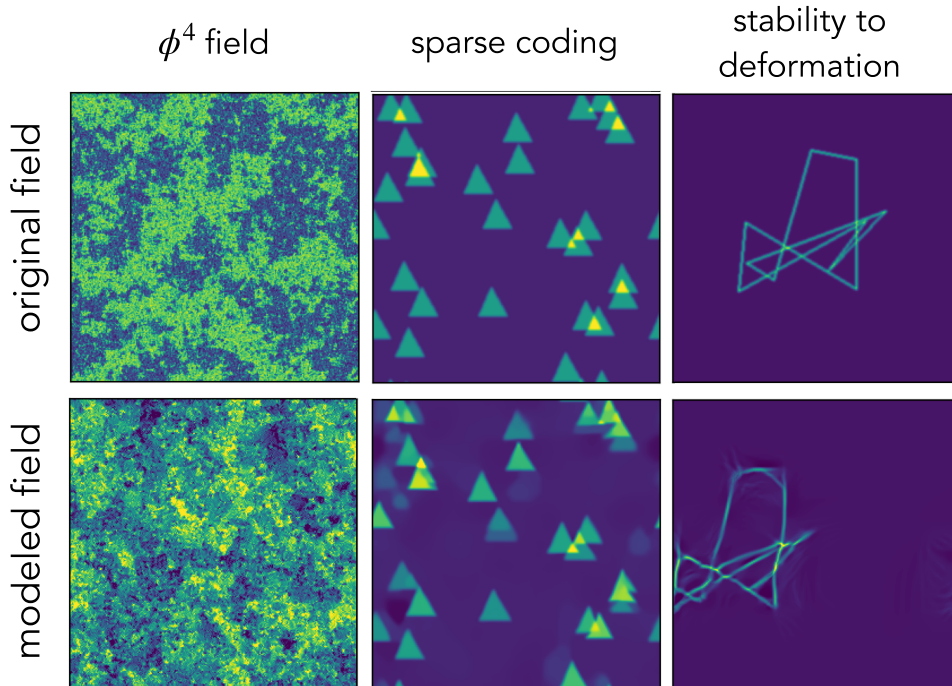
When it is zero, the random field  $p(x)$  should be invariant to the transform up to the expressivity of our representation. For the five random fields analyzed in this study, we measure their asymmetry indices with respect to rotation and scaling. The corresponding anisotropy and scale dependence indices are (A) 0, 0.16; (B) 0, 0.53; (C) 0, 0.66; (D) 0.32, 0.45; (E) 0.28, 0.29. As expected, the cosmic lensing field (field A) is closest to isotropic and scale-free, because the scale range of the simulated field (approximately 80 Mpc in physical size) falls in the nonlinear scale of cosmic structure formation and thus consists of peaks with all sizes and strengths. The cosmic web (B) and 2D turbulence (C) fields are isotropic but not scale-free, because they have particular physical scales above which the field becomes Gaussian: for cosmic web it is around 150 Mpc (25 pixel), and for turbulence it is the scale of driving force (32 pixel), both in the middle of the scale range of our simulations. The last two turbulence fields have anisotropic physical input, but the latter largely probes the “inertial regime” of turbulence, which is scale-free.



**Fig. 5.** Visual interpretation of the scattering spectra. The central field is one realization of field B in physical simulations. The other four panels are generated fields with two simple collective modifications of the scattering spectra coefficients.

### Limitations

While a broad range of physical fields satisfy the implicit priors of the scattering spectra, one does expect regimes for which the description will not be appropriate. The so-called  $\phi^4$  field in physics comes as a first problematic example. It is the maximum entropy



**Fig. 6.** Example of failures and applications beyond typical physical fields. The modeled field of the central panel has been recentered for easier comparison with the original ones.

field under the power spectrum and pointwise fourth-order moment  $x^4$  constraints, but this characterization is unstable to specify a nonconvex pdf which is a pointwise property as opposed to the delocalized Fourier moments and it is highly unstable at critical points (44). The first column in Fig. 6 shows an original  $\phi^4$  field at its critical temperature and that generated from the full set of scattering spectra. In contrast to previous examples, this type of field is not successfully reproduced.

On the other hand, when built based on one example field  $x_1$  and generating only one realization  $\bar{x}_1$  (i.e. in Eq. 24 both  $i$  and  $j$  are 1), our model has a risk of over-fitting: it almost exactly copies the original field with an arbitrary translation and does not provide enough randomness. It can also be seen as a transition from generative modeling regime into a coding regime. This is related to the fact that for maximum entropy models, when the number of constraints amounts to a considerable fraction of the number of total degree of freedom, the microcanonical distribution deviates significantly from the macrocanonical distribution, and has a much lower entropy. The middle panel of Fig. 6 illustrates this effect, where the relative position of triangles of the modeled field is exactly copied from the original field. It happens only when the field is sparse, and when the full set  $\bar{S}$  is used. This problem can be avoided by increasing the number of input fields or generated fields, or an early stop in the microcanonical sampling.

For physical fields with multiscale structures, it is expected that the distribution function  $p(x)$  does not change much under a slight deformation. When modeling such fields, it is important to have a representation that has the same property. Being built from wavelet decomposition and contracting operator, the scattering spectra also linearize small deformation in the field space, which plays an important role in lowering its variance (see Ref. (8)). However, when modeling structured fields whose distribution functions are not regular under deformation, this means that the generative model will simply produce structures that are close enough up to small deformations. This typical type of failure is shown in the third example of Fig. 6.

## Conclusion

We build maximum entropy models for non-Gaussian random fields based on the scattering spectra statistics. Our models provide a low-dimensional structured representation that captures key properties encountered in a wide range of stationary physical fields, namely: (i) stability to deformations as a result of local conservation laws in Physics for mass, energy, momentum, charge, etc.; (ii) invariance and regularity to rotation and scaling; and (iii) scale interactions typically not described by high-order statistics. Those are the priors included in the scattering spectra.

Our models provide a practical tool for generating mock fields based on some example physical fields. In sharp contrast to neural network models, our representation has the key advantage of being interpretable and can be estimated on a few realizations. This is crucial in Physics where generating fields in experiments or simulations is costly or when nonstationarity limits the amount of clean recorded data. Our proposed approach enables a new range of data/simulation analyses, e.g. (23, 24), involving extensions to the modeling of cross-regularities when multiple channels are available, e.g. (22).

## Notes

<sup>a</sup>This single moment can be directly constrained, and we do not discuss it in the following.

<sup>b</sup>Here, we assume the choice of  $R$  is independent of field dimension  $d$ . Another possible choice is to require a constant ratio between the radial and tangential sizes of the  $d$ -dimension oriented wavelets. Then,  $R$  is proportional to the ratio between the surface area of a  $d-1$ -sphere and the volume of a  $d-1$ -ball, proportionally to  $\Gamma(n/2 + 1/2)/\Gamma(n/2)$ . It results in an approximate scaling of  $RJ = O(d \log L)$  when  $d$  is small and  $O(\sqrt{d} \log L)$  when  $d$  is large.

<sup>c</sup>More than a 100 years ago, astronomer Eddington (40, p. 147) expressed in his book a favor of mean-modulus error over mean-square error for similar reasons.

<sup>d</sup>When working without this condition, statistics can be computed by padding the images.

## Acknowledgments

S.C. thanks Siyu Yao for her constant inspiration and encouragement.

## Supplementary Material

Supplementary material is available at PNAS Nexus online.

## Funding

The authors acknowledge funding from the French government as part of the “Investissements d’avenir” program ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). S.C. acknowledges the support of the Martin A. and Helen Chooljian Member Fund, fund from Zurich Insurance Company, and the Fund for Natural Sciences at the Institute for Advanced Study. B.M. acknowledges the support from the David and Lucile Packard Foundation.

## Author Contributions

S.C. developed part of the algorithm, most of the implementation and numerical experiments, introduced the correspondence to higher order moments, and had an important writing contribution. R.M. developed some of the core mathematics, and algorithms related to self-similarity and their implementation, and wrote the corresponding sections and appendices. E.A. introduced transformations related to rotation symmetries and reduced representations and participated in numerical experiments and their expositions. B.M. provided his physical expertise to define the problem, evaluate solutions, and present results in the paper. S.M. provided an important part of the mathematical framework of the representation and organized the teamwork and written paper.

## Preprints

A preprint of this article is published at <https://doi.org/10.48550/arXiv.2306.17210>.

## Data Availability

The code used in this paper can be found on the github: [https://github.com/SihaoCheng/scattering\\_transform](https://github.com/SihaoCheng/scattering_transform). Data used for this work were previously published, including the Dark Matter dataset (52, 53) from the Columbia lensing group accessible at <http://columbialensing.org>, the Quijote Simulations (54) available at <https://quijote-simulations.readthedocs.io>, the turbulence simulations published in Refs. (9, 55), and the forced isotropic

turbulence simulation (56, 57) from the Johns Hopkins Turbulence Database accessible at <http://turbulence.pha.jhu.edu>.

## References

- Landau LD, Lifshitz EM. 2013. *Statistical physics*. Vol. 5. Oxford: Elsevier.
- Jaynes ET. 1957. Information theory and statistical mechanics. *Phys Rev*. 106(4):620–630.
- Bougeret JL, et al. 1995. Waves: the radio and plasma wave investigation on the wind spacecraft. *Space Sci Rev*. 71(1–4):231–263.
- Vielva P, Martínez-González E, Barreiro RB, Sanz JL, Cayón L. 2004. Detection of non-Gaussianity in the Wilkinson microwave anisotropy probe first-year data using spherical wavelets. *Astrophys J*. 609(1):22–34.
- Podesta JJ. 2009. Dependence of solar-wind power spectra on the direction of the local mean magnetic field. *Astrophys J*. 698(2): 986–999.
- Ha W, Singh C, Lanusse F, Upadhyayula S, Yu B. 2021. Adaptive wavelet distillation from neural networks through interpretations. *Adv Neural Inf Process Syst*. 34:20669–20682.
- Mallat S. 2012. Group invariant scattering. *Commun Pure Appl Math*. 65(10):1331–1398.
- Bruna J, Mallat S. 2013. Invariant scattering convolution networks. *IEEE Trans Pattern Anal Mach Intell*. 35(8):1872–1886.
- Allys E, et al. 2019. The RWST, a comprehensive statistical description of the non-Gaussian structures in the ISM. *Astron Astrophys*. 629:A115.
- Saydjari AK, et al. 2021. Classification of magnetohydrodynamic simulations using wavelet scattering transforms. *Astrophys J*. 910(2):122.
- Cheng S, Ting Y-S, Ménard B, Bruna J. 2020. A new approach to observational cosmology using the scattering transform. *Mon Not R Astron Soc*. 499(4):5902–5914.
- Cheng S, Ménard B. 2021. Weak lensing scattering transform: dark energy and neutrino mass sensitivity. *Mon Not R Astron Soc*. 507(1):1012–1020.
- Valogiannis G, Dvorkin C. 2022. Towards an optimal estimation of cosmological parameters with the wavelet scattering transform. *Phys Rev D*. 105(10):103534.
- Valogiannis G, Dvorkin C. 2022. Going beyond the galaxy power spectrum: an analysis of BOSS data with wavelet scattering transforms. *Phys Rev D*. 106(10):103509.
- Valogiannis G, Yuan S, Dvorkin C. 2023. Precise cosmological constraints from BOSS galaxy clustering with a simulation-based emulator of the wavelet scattering transform, arXiv, arXiv:2310.16116, preprint: not peer reviewed.
- Greig B, Ting Y-S, Kurov AA. 2022. Exploring the cosmic 21-cm signal from the epoch of reionization using the wavelet scattering transform. *Mon Not R Astron Soc*. 513(2):1719–1741.
- Cheng S, Ménard B. 2021. How to quantify fields or textures? A guide to the scattering transform, arXiv, arXiv:2112.01288, preprint: not peer reviewed.
- Mallat S, Zhang S, Rochette G. 2020. Phase harmonic correlations and convolutional neural networks. *Inf Inference J IMA*. 9(3): 721–747.
- Zhang S, Mallat S. 2021. Maximum entropy models from phase harmonic covariances. *Appl Comput Harmon Anal*. 53:199–230.
- Allys E, et al. 2020. New interpretable statistics for large-scale structure analysis and generation. *Phys Rev D*. 102(10):103506.
- Jeffrey N, et al. 2022. Single frequency CMB B-mode inference with realistic foregrounds from a single training image. *Mon Not R Astron Soc*. 510(1):L1–L6.
- Régaldo-Saint Blancard B, et al. 2023. Generative models of multi-channel data from a single example-application to dust emission. *Astrophys J*. 943(1):9.
- Regaldo-Saint Blancard B, Allys E, Boulanger F, Levrier F, Jeffrey N. 2021. A new approach for the statistical denoising of planck interstellar dust polarization data. *Astron Astrophys*. 649:L18.
- Delouis J-M, Allys E, Gauvrit E, Boulanger F. 2022. Non-Gaussian modelling and statistical denoising of planck dust polarisation full-sky maps using scattering transforms. *Astron Astrophys*. 668:A122.
- Auclair C, et al. 2024. Separation of dust emission from the cosmic infrared background in Herschel observations with wavelet phase harmonics. *Astron Astrophys*. 681:A1.
- Siahkoobi A, et al. 2023. Unearthing insights into Mars: unsupervised source separation with limited data, arXiv, arXiv:2301.11981, preprint: not peer reviewed.
- Morel R, Rochette G, Leonarduzzi R, Bouchaud J-P, Mallat S. 2022. Scale dependencies and self-similar models with wavelet scattering spectra, arXiv, arXiv:2204.10177, preprint: not peer reviewed.
- Geman S, Geman D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell*. PAMI-6(6):721–741.
- Zhu SC, Wu YN, Mumford D. 1997. Minimax entropy principle and its application to texture modeling. *Neural Comput*. 9(8): 1627–1660.
- Zhu SC, Wu Y, Mumford D. 1998. Filters, random fields and maximum entropy (frame): towards a unified theory for texture modeling. *Int J Comput Vis*. 27(2):107–126.
- Brillinger DR. 1965. An introduction to polyspectra. *Ann Math Statist*. 36(5):1351–1374.
- Huber PJ. 1981. *Robust statistics*. Wiley series in probability and statistics. Hoboken (NJ): John Wiley & Sons.
- Bak P, Tang C, Wiesenfeld K. 1987. Self-organized criticality: an explanation of the 1/f noise. *Phys Rev Lett*. 59(4):381–384.
- Bouchaud J-P, Georges A. 1990. Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. *Phys Rep*. 195(4–5):127–293.
- Coles P, Jones B. 1991. A lognormal model for the cosmological mass distribution. *Mon Not R Astron Soc*. 248(1):1–13.
- Kello CT, et al. 2010. Scaling laws in cognitive sciences. *Trends Cogn Sci*. 14(5):223–232.
- Sornette D. 2017. *Why stock markets crash: critical events in complex financial systems*. Revised ed. Princeton (NJ): Princeton University Press.
- de Wit TD. 2004. Can high-order moments be meaningfully estimated from experimental turbulence measurements? *Phys Rev E*. 70(5):055302.
- Lombardo F, Volpi E, Koutsoyiannis D, Papalexiou SM. 2014. Just two moments! a cautionary note against use of high-order moments in multifractal models in hydrology. *Hydrol Earth Syst Sci*. 18(1):243–255.
- Eddington AS. 1914. *Stellar movements and the structure of the universe*. London: Macmillan and Co. Ltd.
- Portilla J, Simoncelli EP. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int J Comput Vis*. 40(1):49–70.
- Olshausen BA, Field DJ. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. 381(6583):607–609.
- Mallat S. 2009. *A wavelet tour of signal processing*. 3rd ed. Burlington (MA): Elsevier.

- 44 Marchand T, Ozawa M, Biroli G, Mallat S. 2023. Multiscale data-driven energy estimation and generation. *Phys Rev X*. 13:041038.
- 45 Chang SG, Yu B, Vetterli M. 2000. Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans Image Process*. 9(9):1532–1546.
- 46 Stein C. 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability 1954–1955*. Vol. I. Berkeley-Los Angeles (CA): University of California Press. p. 197–206.
- 47 Donoho DL, Johnstone IM. 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. 81(3):425–455.
- 48 Bickel PJ, Levina E. 2008. Covariance regularization by thresholding. *Ann Stat*. 36(6):2577–2604.
- 49 Cai T, Liu W. 2011. Adaptive thresholding for sparse covariance matrix estimation. *J Am Stat Assoc*. 106(494):672–684.
- 50 Fan J, Liao Y, Mincheva M. 2013. Large covariance estimation by thresholding principal orthogonal complements. *J R Stat Soc Series B Stat Methodol*. 75(4):603–680.
- 51 Sherman M. 2018. Variance estimation for statistics computed from spatial lattice data. *J R Stat Soc Series B Stat Methodol*. 58(3): 509–523.
- 52 Matilla JMZ, Haiman Z, Hsu D, Gupta A, Petri A. 2016. Do dark matter halos explain lensing peaks? *Phys Rev D*. 94(8):083506.
- 53 Gupta A, Matilla JMZ, Hsu D, Haiman Z. 2018. Non-Gaussian information from weak lensing data via deep learning. *Phys Rev D*. 97(10):103515.
- 54 Villaescusa-Navarro F, et al. 2020. The quijote simulations. *Astrophys J Suppl Ser*. 250(1):2.
- 55 Schneider K, Farge M, Azzalini A, Ziuber J. 2006. Coherent vortex extraction and simulation of 2D isotropic turbulence. *J Turbulence*. 7(44):N44.
- 56 Li Y, et al. 2008. A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. *J Turbulence*. 9:N31.
- 57 Perlman E, Burns R, Li Y, Meneveau C. 2007. Data exploration of turbulence simulations using a database cluster. In: *Proceedings of the 2007 ACM/IEEE Conference on Supercomputing, SC'07*. New York (NY): Association for Computing Machinery. p. 1–11.
- 58 Kolmogorov AN. 1941. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Cr Acad Sci URSS*. 30:301–305.
- 59 Jaffard S. 2004. Wavelet techniques in multifractal analysis. Technical Report. Paris University, France.
- 60 Olah C, Mordvintsev A, Schubert L. 2017. Feature visualization. *Distill*. 2(11):e7.