

Scattering Spectra Models for Physics

Sihao Cheng (程思浩)¹*, Rudy Morel², Erwan Ally³, Brice Ménard⁴ & Stéphane Mallat^{2,5,6}

¹Institute for Advanced Study, Princeton, NJ 08540, USA

²Département d'informatique de l'Ecole normale supérieure, ENS, CNRS, PSL University, Paris, France

³Laboratoire de Physique de l'Ecole normale supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université Paris Cité, Paris, France

⁴Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA

⁵Collège de France, Paris, France

⁶Center for Computational Mathematics, Flatiron Institute, New York, USA

arXiv:2306.17210v1 [physics.data-an] 29 Jun 2023

3 July 2023

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 4th order. These 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.

1 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 non-linear 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 (Landau & Lifshitz 2013). 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 (Jaynes 1957). 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 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 to 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 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., Bougeret et al. 1995; Vielva et al. 2004; Podesta 2009) which fail to capture important non-Gaussian properties of a field. Other studies (Ha et al. 2021) 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 point-wise non-linear operators on the wavelet transforms of non-Gaussian fields to recover their high-order statistics. The scat-

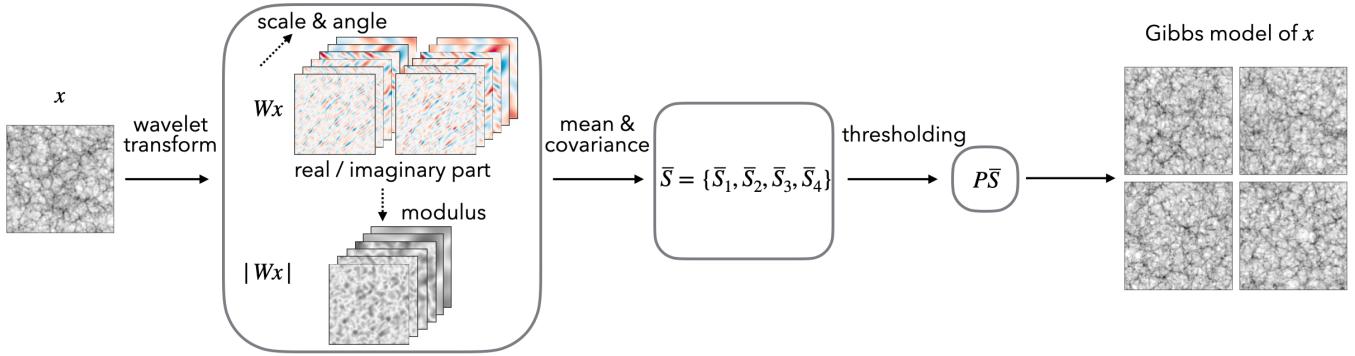


Figure 1. Steps to build a feasible model for a random field x from only one or a few realizations. We first build a low-dimension representation $\Phi(x)$ of the random field, which specifies a maximum entropy model. The representation $\Phi(x)$ is obtained by conducting the wavelet transform Wx and 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 256×256 field.

tering transform, for instance, is a representation that is built by cascading wavelet transforms and non-linear modulus (Mallat 2012; Bruna & Mallat 2013). This representation has been used in astrophysics and cosmology (Cheng & Ménard 2021), to study the interstellar medium (Allys et al. 2019; Saydjari et al. 2021), weak-lensing fields (Cheng et al. 2020), galaxy surveys (Valogiannis & Dvorkin 2022), or radio observations (Greig et al. 2022). Other representations, which are built from covariances of phase harmonics of wavelet transforms (Mallat et al. 2020; Zhang & Mallat 2021), have also been used to model different astrophysical processes (Allys et al. 2020; Jeffrey et al. 2022; Régaldo-Saint Blanchard et al. 2023). Such models, which can be built from a single image, have in turn enabled the development of new component separation methods (Régaldo-Saint Blanchard et al. 2021; Delouis et al. 2022), which can be directly applied to observational data without any particular prior model of the components of a mixture (Auclair et al. 2023; Siahkoohi et al. 2023).

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, reduced scattering covariance representations have been introduced, but only for time series, by leveraging scale invariance (Morel et al. 2022). In this paper, we present the scattering spectra, a low-dimensional representation that is able to efficiently describe a wide range of non-Gaussian processes encountered in physics. In particular, we show how it is possible to take into account the intrinsic regularity of physical fields to dramatically reduce the dimension of such representations. The first part of the paper presents maximum entropy models and scattering spectra statistics, as well as their dimensional reduction. The second part of the paper presents a quantitative validation of these models on various two-dimensional multiscale physical fields and discuss their limitations.

Notations: v^* is the complex conjugate of a scalar v . 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.

2 METHODS

2.1 Gibbs Energy of Stationary Fields

We review the properties of Gibbs energies resulting from maximum entropy models conditioned by moment values (Geman & Geman 1984; Zhu et al. 1997, 1998). 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

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

They define exponential probability models

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

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_{θ_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

$$\int \Phi(x) p_{\theta_0}(x) dx = \mathbb{E}\{\Phi(x)\}. \quad (2)$$

In statistical physics, p_{θ_0} is a macrocanonical model defined by a vector $\mathbb{E}\{\Phi(x)\}$ of observables. One can verify that θ_0 also minimizes the Kullback-Liebler divergence within the class

$$D(p \| p_{\theta_0}) = \int p(x) \log \frac{p(x)}{p_{\theta_0}(x)} dx = H(p_{\theta_0}) - H(p). \quad (3)$$

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 \tilde{x}_i . In this section, we suppose that $n = 1$. Reducing the model error given by (3) amounts to defining Φ 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 $\epsilon > 0$

$$\Omega_\epsilon = \{x \in \mathbb{R}^{L^d} : \|\Phi(x) - \Phi(\bar{x}_1)\|^2 \leq \epsilon\}. \quad (4)$$

Appendix A reviews a sampling algorithm for such model. It also explains how to extend the definition of Ω_ϵ 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 ∞ and ϵ 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 trade-off between the model error (3) and the distance between micro and macrocanonical distributions.

2.2 Fourier Polyspectra Potentials

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 average¹ 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,

$$\mathbb{E}\{\hat{x}[k]\hat{x}[k']^*\} \text{ with } k = k'. \quad (5)$$

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 (Brillinger 1965), where phase cancellation implies that for stationary fields, only the following moments are informative,

$$\mathbb{E}\{\hat{x}[k_1] \dots \hat{x}[k_n]\} \text{ with } k_1 + \dots + k_n = 0, \quad (6)$$

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 non-zero 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 (Huber 1981). For random fields with a heavy tail distribution, which is ubiquitous in complex systems (Bak et al. 1987; Bouchaud & Georges 1990; Coles & Jones 1991; Kello et al. 2010; Sor 2017), higher order moments may not even exist. Those two problems are common for high-order moments and have been demonstrated in real-world applications (Dudok de Wit 2004; Lombardo et al. 2014). In the following two sections, we introduce modifications to this approach to solve those problems.

2.3 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.

2.3.1 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 $\psi[u] = g[u] e^{i\xi \cdot u}$ where $g[u]$ is a real window whose Fourier transform $\hat{g}[k]$ is centered at $k = 0$ so that $\hat{\psi}[k] = \hat{g}[k - \xi]$ is localized in the neighborhood of the frequency ξ . Fig. A1 shows ψ and $\hat{\psi}$ for a $d = 2$ dimensional Morlet wavelet described in appendix 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

$$\psi_\lambda[u] = 2^{-jd} \psi[2^{-j}r^{-1}u] \text{ with } \lambda = 2^{-j}r\xi. \quad (7)$$

Its Fourier transform is $\hat{\psi}_\lambda[k] = \hat{g}[2^j r^{-1}(k - \xi)]$, which is centered at the frequency λ 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

$$Wx[u, \lambda] = x \star \psi_\lambda[u] = \sum_{u'} x[u'] \psi_\lambda[u - u']. \quad (8)$$

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 4 different wavelet orientations. The total number of wavelet frequencies λ is $RJ = O(\log L)$ ² as opposed to L^d Fourier frequencies.

¹ This single moment can be directly constrained, and we do not discuss it in the following.

² 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

A wavelet transform is also stable and invertible if ψ satisfies a Littlewood-Paley condition, which requires an additional convolution with a low-pass *scaling* function ψ_0 centered at the frequency $\lambda = 0$. The specifications are detailed in appendix B.

2.3.2 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 :

$$M_2 = \text{Ave}_u |Wx[u, \lambda]|^2. \quad (9)$$

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. 2(a).

2.3.3 Selected 3rd and 4th 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 non-trivial 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,

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

for $\lambda \neq \lambda'$. To create a non-zero correlation, we must realign the support of $Wx[u, \lambda]$ and $Wx[u', \lambda']$ in Fourier space through nonlinear transforms. As shown in Fig. 2(b), we may apply a square modulus to one band (shown in blue) 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 λ . 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

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

that are a priori non-zero. Furthermore, for wide classes of multiscale processes having regular power spectrum, it suffices to only keep the coefficients at $u = u'$ because of random phase fluctuation (see appendix B). For stationary random fields, they can be estimated with an empirical average over u ,

$$M_3 = \text{Ave}_u (|Wx|^2[u, \lambda] Wx[u, \lambda']^*). \quad (10)$$

to the ratio between the surface area of a $d-1$ -sphere and the volume of $d-1$ -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.

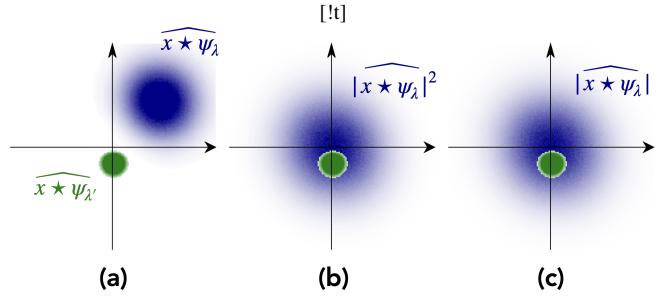


Figure 2. (a): For $\lambda \neq \lambda'$ the Fourier supports of $x \star \psi_\lambda$ (blue) and $x \star \psi_{\lambda'}$ (green) typically do not overlap. (b): The Fourier support of $|x \star \psi_\lambda|^2$ is twice larger and centered at 0 and hence overlaps with $x \star \psi_{\lambda'}$ if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

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,

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

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 :

$$M_4 = \text{Ave}_u (W|Wx|^2[u, \lambda, \gamma] W|Wx|^2[u, \lambda', \gamma]^*), \quad (11)$$

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.

2.4 Scattering Spectra

In general, the estimation of high-order moments has a high variance because high-order polynomials amplify the effect of outliers. A scattering approach (Mallat 2012; Bruna & Mallat 2013; Cheng & Ménard 2021) reduces the variance of these estimators by replacing $|Wx|^2$ by $|Wx|$. The resulting spectra 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 (Portilla & Simoncelli 2000). Their mathematical properties have been analyzed to capture non-Gaussian characteristics of random fields (Mallat et al. 2020; Zhang & Mallat 2021) in relation to scattering moments (Mallat 2012; Bruna & Mallat 2013). Scattering spectra have been defined on one-dimensional time-series (Morel et al. 2022), 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 $O(\log^3 L)$.

2.4.1 First and second wavelet moments, sparsity

For non-Gaussian fields x , wavelet coefficients $Wx[u, \lambda]$ define fields which are often sparse (Olshausen & Field 1996; Stephane 1999). 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

$$S_1(x)[\lambda] = \text{Ave}_u |Wx[u, \lambda]| \quad (12)$$

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

$$S_2(x)[\lambda] = M_2(x)[\lambda] = \text{Ave}_u (|Wx|^2[u, \lambda]). \quad (13)$$

2.4.2 Cross-Spectra between Scattering Channels

A scattering transform is computed by cascading modulus of wavelet coefficients and wavelet transforms (Mallat 2012; Bruna & Mallat 2013). Let us replace $|Wx|^2$ by $|Wx|$ in the selected third and fourth-order wavelet moments described in the previous section. The third order moments (10) become $\mathbb{E}\{|Wx[u, \lambda]| Wx[u, \lambda']^*\}$. Such moments are a priori non-zero 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 non-zero moments which can capture dependencies between different frequencies λ and λ' . The third order moment estimators given by (10) can thus be replaced by lower cross-correlations between $|Wx|$ and Wx at $|\lambda| \geq |\lambda'|$

$$S_3(x)[\lambda, \lambda'] = \text{Ave}_u (|Wx|[u, \lambda] Wx[u, \lambda']^*). \quad (14)$$

Replacing $|Wx|^2$ by $|Wx|$ in the fourth order wavelet moments (11) 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 $W|Wx|[u, \lambda, \gamma] = |x \star \psi_\lambda| \star \psi_\gamma[u]$

$$S_4(x)[\lambda, \lambda', \gamma] = \text{Ave}_u (W|Wx|[u, \lambda, \gamma] W|Wx|[u, \lambda', \gamma]^*), \quad (15)$$

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

Combining the moment estimators of Eqs. (12,13,14,15) defines a vector of scattering spectra

$$S(x) = (S_1(x), S_2(x), S_3(x), S_4(x)). \quad (16)$$

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 λ , λ' , and γ , the exact dimension of $S(x)$ is $RJ + R^2J(J-1)/8 + R^3J(J^2-1)/48$, of the order $O(\log^3 L)$.

2.4.3 Renormalization

Scattering spectra coefficients must often be renormalized to improve the sampling of maximum entropy models. Indeed, multiscale random processes often have a power spectrum that has a power law decay $\mathbb{E}\{|\tilde{x}[k]|^2\} \sim |k|^{-\eta}$ 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 λ 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 θ_0 parameters of a macrocanonical model defined in (2), for which it has been shown that renormalizing to 1 the variance of wavelet coefficients at all scales avoid numerical instabilities (Marchand et al. 2022)³.

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

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

defined by

$$\bar{S}_1(x)[\lambda] = \frac{S_1(x)[\lambda]}{\sigma[\lambda]}, \quad \bar{S}_2(x)[\lambda] = \frac{S_2(x)[\lambda]}{\sigma^2[\lambda]} \quad (17)$$

$$\bar{S}_3(x)[\lambda, \lambda'] = \frac{S_3(x)[\lambda, \lambda']}{\sigma[\lambda] \sigma[\lambda']}, \quad \bar{S}_4(x)[\lambda, \lambda', \gamma] = \frac{S_4(x)[\lambda, \lambda', \gamma]}{\sigma[\lambda] \sigma[\lambda']}.$$

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

2.5 Dimensionality reduction for physical fields

Though 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 appendix 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 (Allys et al. 2019). We refer the reader to appendix 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 pre-determined 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

³ Without such a normalization, the calculation of θ_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.

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 0-th harmonic and $F\bar{S}(x)$ is a sparse representation.

Thresholding on a sparse representation is widely used in image processing for compression (Chang et al. 2000). 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 (Stein 1956) which have lower variance and are consistent (e.g., Donoho & Johnstone 1994; Bickel & Levina 2008; Cai & Liu 2011; Fan et al. 2013). 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 standard deviations of individual coefficients of $F\bar{S}$. These adaptive thresholding estimators achieve a higher rate of convergence and are easy to implement (Cai & Liu 2011). 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., Sherman 2018). We call $P\bar{S}$ the coefficients after thresholding projection:

$$P\bar{S} = (\bar{S}_1, \bar{S}_2, P\bar{S}_3, P\bar{S}_4) = \text{thresholding } F\bar{S}. \quad (18)$$

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.

3 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. B1 for a visual assessment). All computations are reproducible with the software available on https://github.com/SihaoCheng/scattering_transform.

3.1 Dataset of Physical Fields

We use five two-dimensional 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 (Matilla et al. 2016; Gupta et al. 2018).
- (B) *Dark matter*: logarithm of 2D slices of the 3D large-scale distribution of dark matter in the Universe (Villaescusa-Navarro et al. 2020).
- (C) *2D turbulence*: turbulence vorticity fields of incompressible 2D fluid stirred at the scale around 32 pixels, simulated from 2D Navier-Stokes equations (Schneider et al. 2006).

(D) *Magnetic turbulence*: column density of 3D isothermal magnetic-hydrodynamic (MHD) turbulent simulations (Allys et al. 2019). The field is anisotropic due to a mean magnetic field in the horizontal direction.

(E) *Anisotropic turbulence*: two-dimensional slices of a set of 3D turbulence simulations (Li et al. 2008; Perlman et al. 2007). 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 conditions⁴ and normalized to have zero mean and unity standard deviation, 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.

3.2 Model description and visual validation

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

$$\left| \left| \text{Ave}_j \Phi(x_j) - \text{Ave}_i \Phi(\bar{x}_i) \right| \right|^2 \leq \epsilon \quad (19)$$

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 appendix 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 $\bar{M}_2 = \bar{S}_2$, $\bar{M}_3(x)[\lambda, \lambda'] = \frac{M_3(x)[\lambda, \lambda']}{\sigma^2[\lambda] \sigma[\lambda']}$ 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 (17). For the choice of wavelets, we use $J=7$ dyadic scales, and we set $R = 8$ which samples 4 orientations within π , 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 2σ Fourier thresholded representation of \bar{M} defined in exactly the same way as $P\bar{S}$ in (18). The thresholding yields $\dim P\bar{M} = 147, 286, 547, 1708, 926$ 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 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 $P\bar{S}$, which has $\dim P\bar{S} = 204, 364, 489, 615, 304$ 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.

⁴ When working without this condition, statistics can be computed by padding the images.

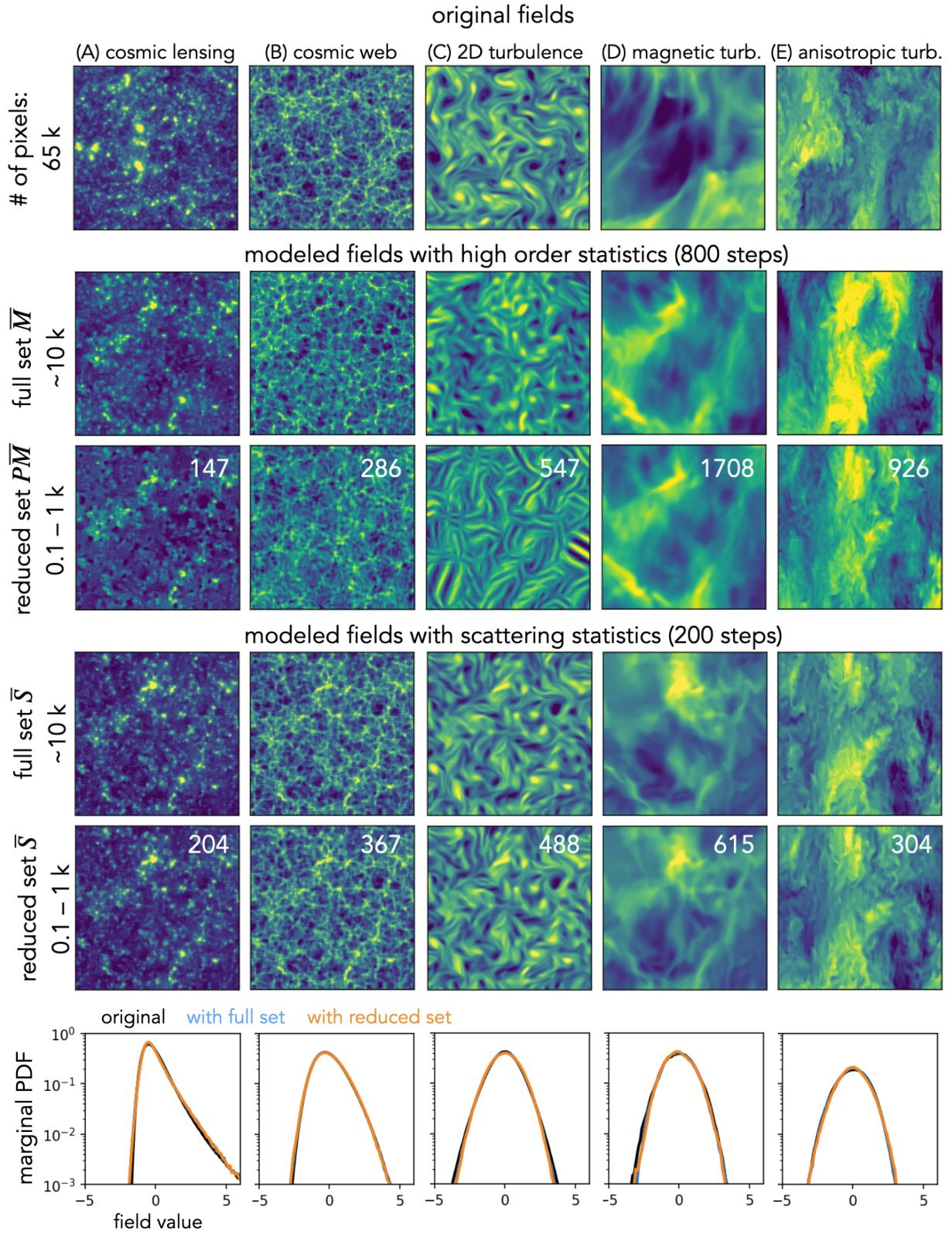


Figure 3. Visual comparison of realistic physical fields and those sampled from maximum entropy models based on wavelet higher-order moments \bar{M} and wavelet scattering spectra \bar{S} statistics. The first row shows five example fields from physical simulations of cosmic lensing, cosmic web, 2D turbulence, 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 by 128 pixels regions. Finally, to quantitatively validate the goodness of the scattering model, we show the marginal PDF (histogram) comparison in the last row.

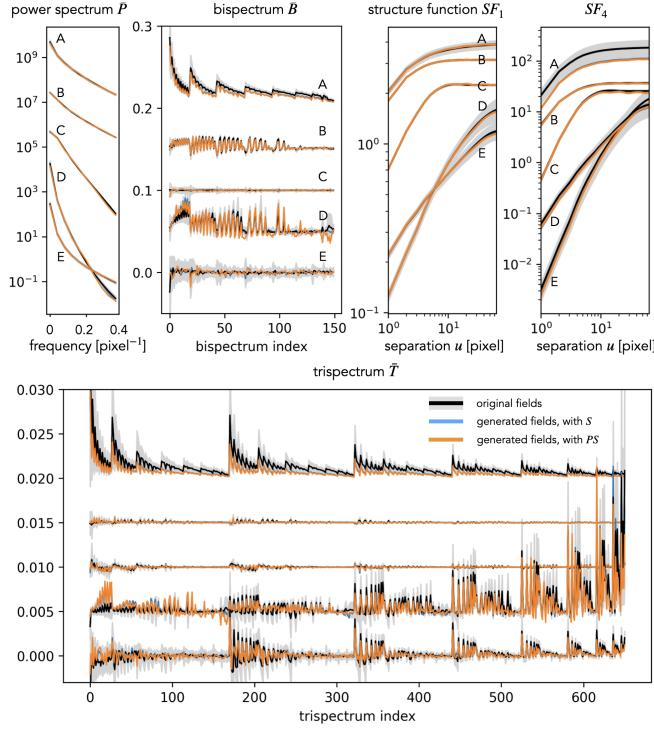


Figure 4. Validation of the scattering maximum entropy models for the five physical fields A–E by various test statistics. The curves for field E represent the original statistics and those for A–D are shifted upwards by an offset. In general, our scattering spectra models well reproduce the validation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

3.3 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 μ_{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 standard deviations σ_{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, $\mu_{P\bar{S}\text{ model}}$ are estimated on fields modeled with the reduced set $P\bar{S}$. Both these averages are estimated from the 10 fields simultaneously sampled from the corresponding microcanonical models.

3.3.1 Validation statistics

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 $P\bar{S}$, 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:

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

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

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

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

where the k_n d -dimensional wave-vectors are respectively averaged in the i_n^{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 1000 independent realizations. Details about the numbers and the ordering of \bar{B} and \bar{T} are given in appendix 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 lag

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

Initially proposed by Kolmogorov for the study of turbulent flows (Kolmogorov 1941), they are widely used to analyze non-Gaussian properties of multiscale processes (Jaffard 2004).

3.3.2 Comparison between original and modeled fields.

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

$$|\mu_{\text{model}} - \mu_{\text{original}}|/\sigma_{\text{original}} > 2.$$

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 $P\bar{S}$. 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 $P\bar{S}$ 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

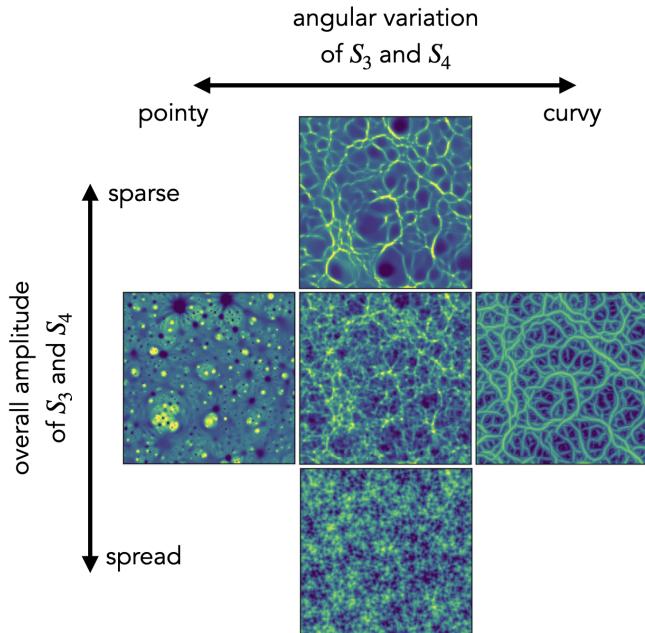


Figure 5. Visual interpretation of the scattering spectra. The central field is one realization of field B from physical simulations. The other four panels are generated fields with two simple collective modifications of the scattering spectra coefficients.

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 (Coles & Jones 1991), whose Gibbs potential includes terms in the form of $\log x$, in contrast to the form of $|x|$ in scattering covariance or x^n in high-order statistics. However, regardless of this difficulty, these outliers are all still within a 3σ 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 non-trivial 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 non-linear 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.

3.4 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 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. C1) 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 Cheng & Ménard (2021), 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 (Olah et al. 2017).

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.

3.5 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:

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

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), which consists of haloes at all scales and strengths, is closest to isotropic and scale-free. The cosmic web (B) and 2D turbulence (C) fields are isotropic but have particular physical scales above which the field becomes Gaussian, so they are not scale-free. The last two turbulence fields have anisotropic physical input, but the latter largely probes the ‘inertial regime’ of turbulence, which is scale-free.

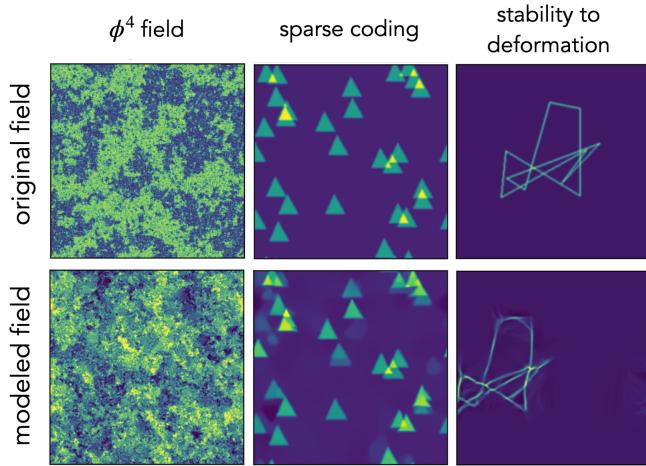


Figure 6. Example of failures and applications beyond typical physical fields. The modeled field of the central panel has been recentered for easier comparison with the original ones.

3.6 Limitations

While a broad range of physical fields satisfy the implicit priors of the scattering covariance, one does expect regimes for which the description will not be appropriate. The so-called φ^4 field in physics comes as a first problematic example. It is the maximum entropy field under the power spectrum and pointwise fourth-order moment x^4 constraints, but this characterization is unstable to specify a non-convex pdf which is a pointwise property as opposed to the delocalized Fourier moments and it is highly unstable at critical points (Marchand et al. 2022). The first column in Fig. 6 shows an original φ^4 field at its critical temperature and that generated from the full set of scattering covariance. 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. 19 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 illustrate 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 \tilde{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 multi-scale 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 (Bruna & Mallat 2013)). 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.

4 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; (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 non-stationarity limits the amount of clean recorded data. Our proposed approach enables a new range of data/simulation analyses (e.g. Regaldo-Saint Blancard et al. 2021; Delouis et al. 2022), involving extensions to the modeling of cross-regularities when multiple channels are available (e.g. Régaldo-Saint Blancard et al. 2023).

ACKNOWLEDGEMENTS

We acknowledge funding from the French government as part of the “Investissements d’avenir” program ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). SC acknowledges support from the Institute for Advanced Study. BM acknowledges support from the David and Lucile Packard Foundation.

REFERENCES

- Allys E., Levrier F., Zhang S., Colling C., Regaldo-Saint Blancard B., Boulanger F., Hennebelle P., Mallat S., 2019, *Astronomy & Astrophysics*, 629, A115
- Allys E., Marchand T., Cardoso J.-F., Villaescusa-Navarro F., Ho S., Mallat S., 2020, *Physical Review D*, 102
- Auclair C., et al., 2023, arXiv preprint arXiv:2305.14419
- Bak P., Tang C., Wiesenfeld K., 1987, *Phys. Rev. Lett.*, 59, 381
- Bickel P. J., Levina E., 2008, *The Annals of Statistics*, 36, 2577
- Bouchaud J.-P., Georges A., 1990, *Physics Reports*, 195, 127
- Bougeret J. L., et al., 1995, *Space Sci. Rev.*, 71, 231
- Brillinger D. R., 1965, *The Annals of mathematical statistics*, pp 1351–1374
- Bruna J., Mallat S., 2013, *IEEE transactions on pattern analysis and machine intelligence*, 35, 1872
- Bruna J., Mallat S., 2019, *Mathematical Statistics and Learning*, 1, 257
- Cai T., Liu W., 2011, *Journal of the American Statistical Association*, 106, 672
- Chang S. G., Yu B., Vetterli M., 2000, *IEEE transactions on image processing*, 9, 1532
- Cheng S., Ménard B., 2021, arXiv preprint arXiv:2112.01288
- Cheng S., Ting Y.-S., Ménard B., Bruna J., 2020, *MNRAS*, 499, 5902
- Coles P., Jones B., 1991, *MNRAS*, 248, 1
- Delouis J.-M., Allys E., Gauvrit E., Boulanger F., 2022, *Astronomy & Astrophysics*, 668, A122
- Donoho D. L., Johnstone I. M., 1994, *Biometrika*, 81, 425
- Dudok de Wit T., 2004, *Phys. Rev. E*, 70, 055302

- Fan J., Liao Y., Mincheva M., 2013, Journal of the Royal Statistical Society. Series B, Statistical methodology, 75
- Geman S., Geman D., 1984, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, 721
- Greig B., Ting Y.-S., Kaurov A. A., 2022, *MNRAS*, 513, 1719
- Gupta A., Matilla J. M. Z., Hsu D., Haiman Z., 2018, *Physical Review D*, 97, 103515
- Ha W., Singh C., Lanusse F., Upadhyayula S., Yu B., 2021, Advances in Neural Information Processing Systems, 34, 20669
- Huber P. J., 1981, Robust statistics. Wiley Ser. Probab. Math. Stat., John Wiley & Sons, Hoboken, NJ, doi:10.1002/0471725250
- Jaffard S., 2004, Technical report, Wavelet techniques in multifractal analysis. PARIS UNIV (FRANCE)
- Jaynes E. T., 1957, *Physical Review*, 106, 620
- Jeffrey N., Boulanger F., Wandelt B. D., Regaldo-Saint Blancard B., Ally E., Levrier F., 2022, *MNRAS*, 510, L1
- Kello C. T., Brown G. D., i Cancho R. F., Holden J. G., Linkenkaer-Hansen K., Rhodes T., Van Orden G. C., 2010, *Trends in Cognitive Sciences*, 14, 223
- Kolmogorov A. N., 1941, Cr Acad. Sci. URSS, 30, 301
- Landau L. D., Lifshitz E. M., 2013, Statistical Physics: Volume 5. Vol. 5, Elsevier
- Li Y., et al., 2008, *Journal of Turbulence*, 9, N31
- Lombardo F., Volpi E., Koutsogiannis D., Papalexiou S. M., 2014, *Hydrology and Earth System Sciences*, 18, 243
- Mallat S., 2012, *Communications on Pure and Applied Mathematics*, 65, 1331
- Mallat S., Zhang S., Rochette G., 2020, *Information and Inference: A Journal of the IMA*, 9, 721
- Mandelbrot B. B., Fisher A. J., Calvet L. E., 1997, Cowles Foundation discussion paper, p. 1412
- Marchand T., Ozawa M., Biroli G., Mallat S., 2022, arXiv e-prints, p. arXiv:2207.04941
- Matilla J. M. Z., Haiman Z., Hsu D., Gupta A., Petri A., 2016, *Physical Review D*, 94, 083506
- Meyer Y., 1992, Wavelets and Operators. Advanced mathematics. Cambridge university press
- Morel R., Rochette G., Leonarduzzi R., Bouchaud J.-P., Mallat S., 2022, arXiv preprint arXiv:2204.10177
- Olah C., Mordvintsev A., Schubert L., 2017, *Distill*
- Olshausen B. A., Field D. J., 1996, *Nature*, 381, 607
- Perlman E., Burns R., Li Y., Meneveau C., 2007, in Proceedings of the 2007 ACM/IEEE Conference on Supercomputing. SC '07. Association for Computing Machinery, New York, NY, USA, doi:10.1145/1362622.1362654, <https://doi.org/10.1145/1362622.1362654>
- Podesta J. J., 2009, *ApJ*, 698, 986
- Portilla J., Simoncelli E. P., 2000, International journal of computer vision, 40, 49
- Regaldo-Saint Blancard B., Ally E., Boulanger F., Levrier F., Jeffrey N., 2021, *Astronomy & Astrophysics*, 649, L18
- Regaldo-Saint Blancard B., Ally E., Auclair C., Boulanger F., Eickenberg M., Levrier F., Vacher L., Zhang S., 2023, *The Astrophysical Journal*, 943, 9
- Saydjari A. K., Portillo S. K., Slepian Z., Kahraman S., Burkhardt B., Finkbeiner D. P., 2021, *The Astrophysical Journal*, 910, 122
- Schneider K., Ziuber J., Farge M., Azzalini A., 2006, *Journal of Turbulence*, p. N44
- Sherman M., 2018, *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 509
- Siahkooi A., Morel R., de Hoop M. V., Ally E., Sainton G., Kawamura T., 2023, arXiv preprint arXiv:2301.11981
- 2017, Why Stock Markets Crash: Critical Events in Complex Financial Systems, rev - revised edn. Princeton University Press, <http://www.jstor.org/stable/j.ctt1h1htkg>
- Stein C., 1956, in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I. University of California Press, Berkeley-Los Angeles, Calif., pp 197–206

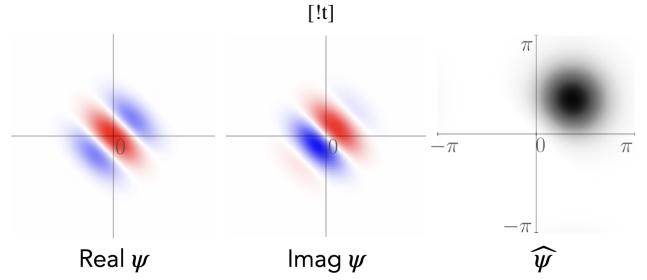


Figure A1. Real and imaginary parts of a Morlet wavelet $\psi[u]$ and its Fourier transform $\hat{\psi}[k]$, used in numerical calculations.

- Stephane M., 1999, A wavelet tour of signal processing
- Valogiannis G., Dvorkin C., 2022, *Physical Review D*, 105, 103534
- Vielva P., Martínez-González E., Barreiro R. B., Sanz J. L., Cayón L., 2004, *ApJ*, 609, 22
- Villaescusa-Navarro F., et al., 2020, *ApJS*, 250, 2
- Zhang S., Mallat S., 2021, *Applied and Computational Harmonic Analysis*, 53, 199
- Zhu S. C., Wu Y. N., Mumford D., 1997, *Neural computation*, 9, 1627
- Zhu S. C., Wu Y., Mumford D., 1998, *International Journal of Computer Vision*, 27, 107

APPENDIX A: MICROCANONICAL SAMPLING

Given n observed samples $\bar{x}_1, \dots, \bar{x}_n$ of a field, with possibly $n = 1$, the microcanonical ensemble given (4) can be extended as follow:

$$\Omega_\epsilon = \left\{ x_1, \dots, x_m \in \mathbb{R}^{L^d} : \left\| \text{Ave}_j \Phi(x_j) - \text{Ave}_i \Phi(\bar{x}_i) \right\|^2 \leq \epsilon \right\}. \quad (\text{A1})$$

Microcanonical models are maximum entropy distributions over Ω_ϵ , which have a uniform distribution over this ensemble. Increasing the number of samples n reduces the variance of $\text{Ave}_i \Phi(\bar{x}_i)$ which concentrates around $\mathbb{E}\{\Phi(x)\}$. This reduces the information about a specific realization which is contained in $\text{Ave}_i \Phi(\bar{x}_i)$, thus limiting over-fitting.

Sampling from the microcanonical model amounts to drawing a realization from a uniform distribution in Ω_ϵ . We approximate this sampling with a gradient descent algorithm studied in (Bruna & Mallat 2019). This algorithm progressively transports a white Gaussian noise distribution, which has a higher entropy than the microcanonical model, into distributions supported in Ω_ϵ . This is done with a gradient descent on $\ell(y_1, \dots, y_m) = \|\text{Ave}_j \Phi(y_j) - \text{Ave}_i \Phi(\bar{x}_i)\|^2$, where the y_j are initialized as independent realizations of white noises. At each iteration, the y_i are updated with the L-BFGS-B algorithm, which is a quasi-Newton method that uses an estimate of the Hessian matrix. In practice, we perform 200 gradient descent steps which yield a typical error $\epsilon \approx 10^{-4}$.

It is proved in (Bruna & Mallat 2019) that this algorithm converges to a distribution that has the same symmetries as $\Phi(x)$, similarly to the microcanonical one. However, it has been shown that this algorithm recovers a maximum entropy distribution in Ω_ϵ only under appropriate conditions and that such gradient descent models may differ, in general, from maximum entropy ones. Nevertheless, these algorithms provide powerful sampling methods to approximate large classes of high-dimensional stationary processes, while being much faster and computationally tractable than alternative MCMC algorithms.

APPENDIX B: WAVELETS IN \mathbb{R}^D AND SCATTERING COVARIANCES

A Morlet wavelet ψ defined on \mathbb{R}^d is the product of a Gaussian envelope with a sinusoidal wave

$$\psi[u] = g_\sigma[u](e^{i\langle \xi, u \rangle} - c) \text{ with } g_\sigma[u] = \frac{1}{(\sigma\sqrt{2\pi})^d} e^{-\frac{\|u\|^2}{2\sigma^2}},$$

where c is chosen so that $\int \psi[u] du = 0$ ⁵. Such a wavelet recovers variations around scale 2^j in the direction of ξ . It is invariant to any rotation of \mathbb{R}^d that fixes ξ . In practice we choose $\xi = (3\pi/4, 0, \dots, 0)$ and $\sigma = 0.8$. For simplifying equations in appendices we assume $\|\xi\| = 1$, without loss of generality. To recover variations at other scales and in other directions we define the wavelet filters

$$\psi_\lambda[u] = 2^{-jd}\psi[2^{-j}r^{-1}u] \text{ with } \lambda = 2^{-j}r^{-1}\xi$$

for $(j, r) \in \mathbb{R} \times SO(d)$. In Fourier, $\widehat{\psi}_\lambda$ is a Gaussian centered in λ substracted by a Gaussian centered in 0 so that $\widehat{\psi}_\lambda[0] = 0$

$$\widehat{\psi}_\lambda[k] = \widehat{\psi}[2^{-j}r^{-1}u] \text{ with } \widehat{\psi}[k] = e^{-\frac{\sigma^2}{2}\|k-\xi\|^2} - c e^{-\frac{\sigma^2}{2}\|k\|^2}$$

We shall restrict the scales 2^j to dyadic scales, hence taking j integer, and restrict the rotations to a discrete subgroup Γ of $SO(d)$ of order $2^d - 1$ (Meyer 1992). In dimension $d = 2$ such a group can be parameterized by one angle, in dimension $d = 3$ it can be parameterized by 2 angles. We write $\Lambda = \mathbb{Z} \times \Gamma$ the group that defines filters ψ_λ from ψ .

To guarantee that the wavelet transform W (defined in (8)) is invertible and satisfies an energy conservation, we impose that the ψ_λ satisfy the following Littlewood-Paley inequality for $0 < \delta < 1$

$$\forall k \neq 0, \quad 1 - \delta \leq \sum_{\lambda \in \Lambda} |\widehat{\psi}_\lambda[k]|^2 \leq 1 + \delta. \quad (\text{B1})$$

For fields defined on a cubic d -dimensional lattice of length L , the wavelets ψ_λ are discretized accordingly. The wavelet transform is computed up to the largest scale 2^J which is smaller than length L so as to achieve a reasonable estimate of low-frequency moments, even on a single realization. The lower frequencies of x in the ball $|k| \leq 2^J$ are captured by a low-pass filter ψ_0 which is a Gaussian centered in $k = 0$ in Fourier $\widehat{\psi}_0[k] = c_0 \exp(-\sigma_0^2 \|k\|^2/2)$ with $\sigma_0 = \sigma 2^{J-1}$. The Littlewood-Paley inequality now reads:

$$\forall k \neq 0, \quad 1 - \delta \leq |\psi_0[k]|^2 + \sum_{|\lambda|^{-1} \leq 2^J} |\widehat{\psi}_\lambda[k]|^2 \leq 1 + \delta.$$

By applying the Parseval formula we derive that for all x

$$(1 - \delta)\|x\|^2 \leq \|Wx\|^2 \leq (1 + \delta)\|x\|^2$$

which insures that W preserves the norm of x , up to a relative error of δ , and is therefore invertible, with stable inverse. For the wavelet used for syntheses of physical fields in this paper, we have $\delta \approx 0.8$.

Covariance of wavelet coefficients $Wx[u, \lambda]$ can be written from the power spectrum $P(x)$ of x

$$\mathbb{E}\{Wx[u, \lambda]Wx[u', \lambda']^*\} = \frac{1}{2\pi} \int P(x)[k]\widehat{\psi}_\lambda[k]\widehat{\psi}_{\lambda'}[k]e^{i\langle u-u', k \rangle} dk.$$

It implies that this correlation is zero if the supports of $\widehat{\psi}_\lambda$ and $\widehat{\psi}_{\lambda'}$ do not overlap. For the specified wavelets, as soon as $\lambda \neq \lambda'$, these supports barely overlap and $\mathbb{E}\{Wx[u, \lambda]Wx[u', \lambda']^*\} \approx 0$. Moreover, since x is stationary, the covariance $\mathbb{E}\{Wx[u, \lambda]Wx[u', \lambda']^*\}$ only

⁵ In practice, the envelope g_σ is an elliptical Gaussian window to increase the angular resolution of ψ , but this does not virtually modify our discussion.

depends on $u - u'$ and have a fast decay when the power spectrum $P(x)$ is regular. Thus, even if dependencies across separate scales may exist, they are not captured by correlation.

Taking the modulus of wavelet coefficients removes complex phase oscillations and thus recenter the frequency support of $Wx[u, \lambda]$. Indeed, the power spectrum $P_\lambda(x)$ of $x \star \psi_\lambda$ is mostly supported in a ball $\|k - \lambda\| \leq 2^{-j}\sigma^{-1}$ which does not overlap with the Fourier support of the power spectrum $P_{\lambda'}(x)$ of $x \star \psi_{\lambda'}$. Taking a modulus on $x \star \psi_{\lambda'}$ eliminates the phase which oscillates at the central frequency λ' . As a consequence, the power spectrum of $|x \star \psi_{\lambda'}|$ is centered at $k = 0$ and its energy is mostly concentrated in $\|k\| \leq 2^{-j}\sigma^{-1}$ which now may overlap with the support of $P_W(x)[\lambda]$ as can be seen in Fig. 2. The power spectra of $|Wx[u, \lambda]|$ and $|Wx[u, \lambda']|$, both centered at zero, also overlap.

We now justify taking $u = u'$ in order 3 moments given by (10). The cross spectrum $P_{\lambda, \lambda'}(x)$ between $Wx[u, \lambda]$ and $|Wx[u, \lambda']|$ is assumed regular for the fields considered in this paper. In that case one can approximate such cross-spectrum using wavelets, which gives the moments $\mathbb{E}\{WWx[u, \lambda, \gamma]W|Wx|[u, \lambda', \gamma]\}$. However, the left-hand-side $WWx[u, \lambda, \gamma]$ is negligible when $\lambda \neq \gamma$ because Fourier support of wavelets ψ_λ and ψ_γ barely overlap. The resulting coefficients

$$\mathbb{E}\{WWx[u, \lambda, \lambda]W|Wx|[u, \lambda', \lambda]\} = \frac{1}{2\pi} \int P_{\lambda, \lambda'}(x)[k]|\widehat{\psi}_\lambda|^2 dk$$

average $P_{\lambda, \lambda'}(x)[k]$ in a ball $\|k\| \leq 2^{-j}\sigma^{-1}$ through $|\widehat{\psi}_\lambda|^2$. However, $P_{\lambda, \lambda'}(x)$ is already concentrated in this ball. We thus remove $|\widehat{\psi}_\lambda|^2$ which yields $\mathbb{E}\{Wx[u, \lambda]Wx|[u, \lambda']\}$.

The following proposition shows that Scattering Spectra reveal non-Gaussianity in a field x .

Proposition 1. Let x be a stationary process.

(i) If x is Gaussian then for any separate scales λ, λ' , meaning that $\widehat{\psi}_\lambda \widehat{\psi}_{\lambda'} = 0$

$$\mathbb{E}\{\bar{S}_1(x)\} = \frac{\pi}{4},$$

$$\mathbb{E}\{\bar{S}_3(x)[\lambda, \lambda']\} = 0 \text{ and } \mathbb{E}\{\bar{S}_4(x)[\lambda, \lambda', \gamma]\} = 0.$$

(ii) If x is symmetric i.e. $p(-x) = p(x)$ then

$$\mathbb{E}\{\bar{S}_3(x)\} = 0.$$

(iii) If x is invariant by rotation of angle π i.e. $p(x[-u]) = p(x[u])$ then

$$\text{Im } \mathbb{E}\{\bar{S}_3(x)\} = 0 \text{ and } \text{Im } \mathbb{E}\{\bar{S}_4(x)\} = 0.$$

Proof. If x is Gaussian then $Wx[u, \lambda]$ is also Gaussian and the ratio between its first and second order moment is $\pi/4$. If $\widehat{\psi}_\lambda \widehat{\psi}_{\lambda'} = 0$ then $Wx[u, \lambda]$ and $Wx[u, \lambda']$ are decorrelated, since $(Wx[u, \lambda], Wx[u, \lambda'])$ is Gaussian, this implies that $Wx[u, \lambda]$ and $Wx[u, \lambda']$ are independent. Thus, $Wx[u, \lambda]$ and $|Wx[u, \lambda']|$ are independent, so are $W|Wx|[u, \lambda, \gamma]$ and $W|Wx|[u, \lambda', \gamma]$ which proves 1. Point 2. is proved by observing that $S_3(-x) = S_3(x)$ and point 3. by observing that $S_3(x[-u]) = S_3(x)^*$ and $S_4(x[-u]) = S_4(x)^*$. \square

For the physical fields studied in this paper, such coefficients are non-zero, thus revealing their non-Gaussianity Fig.C1.

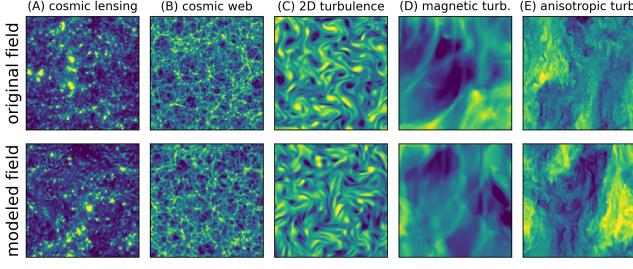


Figure B1. Visual assessment of our model based on \bar{S} with 11 641 coefficients estimated on a single realization (top). Generated fields (bottom) show very good visual quality.

APPENDIX C: EQUIVARIANCE AND INVARIANCE TO ROTATIONS AND SCALING

The scattering spectra are computed from wavelet transforms, which are equivariant to rotations and scalings. We show that scattering spectra inherit these equivariance properties. If $p(x)$ is isotropic or self-similar, then one can build isotropic or self-similar maximum entropy models by averaging renormalized scattering spectra over rotations or scales, which reduces both the variance and dimensionality of \bar{S} .

To avoid discretization and boundary issues for rotations and scaling, we consider fields $x[u]$ defined over continuous variables $u \in \mathbb{R}^d$, and establish the mathematical results in this framework. For this purpose, the sum in the wavelet transform defined in (8) is replaced by an integral over \mathbb{R}^d . Wavelets are dilated by 2^j for $j \in \mathbb{Z}$ and rotated by r in a rotation group G of cardinal R . In dimension $d = 2$, these rotations have an angle $2\pi\ell/R$.

Proposition 2. For $r \in G$ with $x_r[u] = x[r^{-1}u]$ one has

$$S(x_r)[\lambda, \lambda', \gamma] = S(x)[r\lambda, r\lambda', r\gamma]. \quad (\text{C1})$$

For $j \in \mathbb{Z}$ with $x_j[u] = x[2^{-j}u]$ one has

$$S(x_j)[\lambda, \lambda', \gamma] = S(x)[2^j\lambda, 2^j\lambda', 2^j\gamma]. \quad (\text{C2})$$

Proof. It follows from the equivariance of wavelet coefficients, $Wx_r[u, \lambda] = Wx[r^{-1}u, r\lambda]$ and $Wx_j[u, \lambda] = Wx[2^{-j}u, 2^j\lambda]$.

Isotropic fields x have a distribution that is invariant to rotation $x_r \stackrel{d}{=} x$ for all $r \in G$. Self-similar fields x have a distribution that is invariant to scaling, up to random multiplicative factors $x_j \stackrel{d}{=} A_j x$ for all $j \geq 0$ (Mandelbrot et al. 1997). For such fields, we show that the expected scattering spectra exhibit invariance to rotation or scaling of their indices, and thus have a lower-dimensional structure. For that purpose we used normalized scattering spectra coefficient $\bar{S}(x)$ defined (17), where the normalization is done by $\sigma^2[\lambda] = \mathbb{E}\{|Wx[u, \lambda]|^2\}$.

Proposition 3. If x is isotropic then for any $r \in G$

$$\mathbb{E}\{\bar{S}(x)[r\lambda, r\lambda', r\gamma]\} = \mathbb{E}\{\bar{S}(x)[\lambda, \lambda', \gamma]\}. \quad (\text{C3})$$

If x is self-similar at scales $2^j \leq 2^J$ then

$$\mathbb{E}\{S_1(x)[\lambda]\} = c_1|\lambda|^{-\zeta_1}, \quad \mathbb{E}\{S_2(x)[\lambda]\} = c_2|\lambda|^{-\zeta_2} \quad (\text{C4})$$

$$\mathbb{E}\{\bar{S}_3(x)[2^j\lambda, 2^j\lambda']\} = \mathbb{E}\{\bar{S}_3(x)[\lambda, \lambda']\} \quad (\text{C5})$$

$$\mathbb{E}\{\bar{S}_4(x)[2^j\lambda, 2^j\lambda', 2^j\gamma]\} = \mathbb{E}\{\bar{S}_4(x)[\lambda, \lambda', \gamma]\} \quad (\text{C6})$$

Proof. Let us assume x is isotropic $x_r \stackrel{d}{=} x$. It implies that $\mathbb{E}\{S(x_r)\} =$

$\mathbb{E}\{S(x)\}$. Thanks to the equivariance property of (C1) one gets the invariance property on S : $\mathbb{E}\{S(x)[r\lambda, r\lambda', r\gamma]\} = \mathbb{E}\{S(x)[\lambda, \lambda', \gamma]\}$. We obtain (C3) by dividing this equation by $\mathbb{E}\{S_2(x)[r\lambda]\} = \mathbb{E}\{S_2(x)[\zeta]\}$.

Let us assume x is self-similar, $x_j \stackrel{d}{=} A_j x$. In that case one has $A_{j+j'} \stackrel{d}{=} A_j A_{j'}$, taking order 1 and order 2 moments, this implies $\mathbb{E}\{A_j\} = 2^{-j\zeta_1}$ and $\mathbb{E}\{A_j^2\} = 2^{-j\zeta_2}$ for certain power-law exponents ζ_1, ζ_2 . Now from self-similarity and equivariance property given by (C2) one has $\mathbb{E}\{S_1(x)[2^j\lambda]\} = \mathbb{E}\{A_j\}\mathbb{E}\{S_1(x)[\lambda]\} = 2^{-j\zeta_1}\mathbb{E}\{S_1(x)[\lambda]\}$. Taking $2^{-j} = |\lambda|$ one obtains $\mathbb{E}\{S_1(x)[\zeta]\} = c_1|\lambda|^{-\zeta_1}$ with $c_1 = \mathbb{E}\{S_1(x)[|\lambda|^{-1}\lambda]\}$ independent on $|\lambda|$. With the same reasoning on S_2 we obtain (C4). From self-similarity and equivariance property given (C2), we get similarly: $\mathbb{E}\{S_3(x)[2^j\lambda, 2^j\lambda']\} = 2^{-j\zeta_2}\mathbb{E}\{S_3(x)[\lambda, \lambda']\}$. Dividing by $\mathbb{E}\{S_2(x)[\lambda]\} = c_2|\lambda|^{-\zeta_2}$ yields (C5). We obtain (C6) similarly, which proves the proposition.

The wavelet coefficient renormalization is necessary to ensure that the scattering spectra are invariant to scaling. As explained in (Marchand et al. 2022), it is directly related to Wilson renormalization, which yields macrocanonical parameters (physical couplings) that remain constant across scales (fixed point) at phase transitions, where the field becomes self-similar.

If x is isotropic, then (C3) implies that

$$\text{Ave}_{r \in G} \bar{S}(x)[r\lambda, r\lambda', r\gamma] \quad (\text{C7})$$

is an unbiased estimator of $\mathbb{E}\{\bar{S}(x)[\lambda, \lambda', \gamma]\}$ with lower variance than $\bar{S}(x)[\lambda, \lambda', \gamma]$. Choosing $\Phi(x) = \text{Ave}_{r \in G} \bar{S}(x)[r\lambda, r\lambda', r\gamma]$ also reduces the dimension of our model by a factor R . Since this representation is invariant to rotations of x in G , the macrocanonical and microcanonical models defined from it are also invariant to these rotations.

Similarly, if x is self similar on a range of scales $2^j \leq 2^J$, then (C5) and (C6) implies that

$$\text{Ave}_j \bar{S}_3(x)[2^j\lambda, 2^j\lambda'] , \quad \text{Ave}_j \bar{S}_4(x)[2^j\lambda, 2^j\lambda', 2^j\gamma] \quad (\text{C8})$$

where the average is taken on all scales j such that $(2^j|\lambda|)^{-1} \leq 2^J, (2^j|\lambda'|)^{-1} \leq 2^J, (2^j|\gamma|)^{-1} \leq 2^J$, are unbiased estimators of $\mathbb{E}\{\bar{S}_3(x)[\lambda, \lambda']\}$ and $\mathbb{E}\{\bar{S}_4(x)[\lambda, \lambda', \gamma]\}$ with lower variance than $\bar{S}_3(x)$ and $\bar{S}_4(x)$. Choosing $\Phi(x) = (\bar{S}_1(x), \bar{S}_2(x), \text{Ave}_j \bar{S}_3(x), \text{Ave}_j \bar{S}_4(x))$ reduces the dimension of our model by at most a factor $\log L$. The resulting maximum entropy model is not necessarily self-similar due to the presence of scale-dependent moments $\mathbb{E}\{\bar{S}_1(x)\}$ and $\mathbb{E}\{\bar{S}_2(x)\}$. However, if $\bar{S}_1(x)[\zeta]$ and $\bar{S}_2(x)[\lambda]$ have a power-law decay along λ our model becomes self-similar.

APPENDIX D: DIMENSION REDUCTION WITH FOURIER THRESHOLDING

We give in this appendix the details of the dimensional reduction of \bar{S} into $P\bar{S}$, which is done by Fourier projectors of $\bar{S}(x)$ along rotations and scales, estimated by thresholding. This dimensional reduction based on regular variations of the dependence of \bar{S} on different scales, allows for a representation of lower variance, bringing the microcanonical and macrocanonical models closer together.

We concentrate on the two-dimensional case $d = 2$ corresponding to numerical applications. The rotation group is then Abelian and defined by a single angle parameter, which simplifies the Fourier

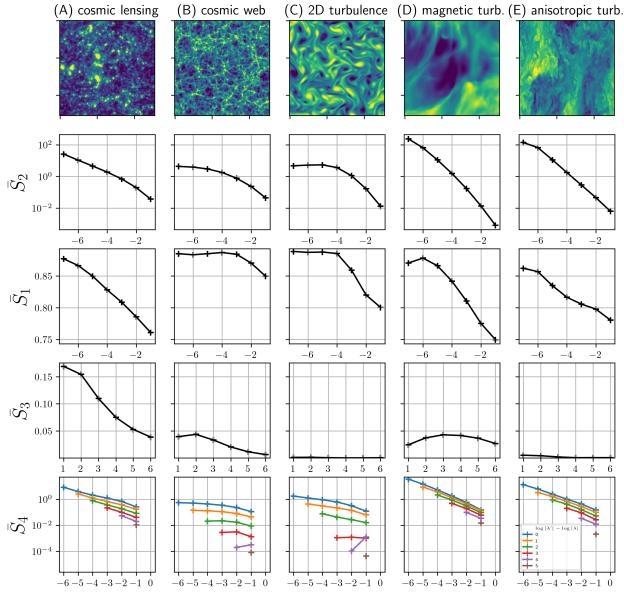


Figure C1. Visualization of Scattering Spectra \bar{S} for different physical fields. Power-spectrum S_2 and sparsity factors \bar{S}_1 are averaged along all angles (amount to taking the 0-th angle Fourier harmonic). We only show the 0-th angle Fourier harmonic and 0-th scale Fourier harmonic for order 3 and order 4-moment estimators \bar{S}_3 and \bar{S}_4 . Thus, the quantities that are shown are invariant to the rotation of the field, and the last two rows (\bar{S}_3 , \bar{S}_4 are furthermore invariant to scaling). Non-zero coefficients \bar{S}_3 show that the cosmic lensing and cosmic web fields are not invariant to sign flip. This is due to the presence of high positive peaks on the former and filaments on the latter. The large amplitude of envelope coefficients \bar{S}_4 on the last 2 fields indicate long-range spatial dependencies as evidenced by the presence of structures at the level of the map.

transform calculation. However, the same approach applies to non-commutative groups G of rotations in \mathbb{R}^d for $d > 2$, with their Fourier transform. Each wavelet frequency is defined in (7) by $\lambda = 2^{-j} r_\ell \xi$, where r_ℓ is a rotation of angle $2\pi\ell/R$. To guarantee that the scattering spectra frequencies satisfy $|\lambda| \leq |\lambda'| < |\gamma|$, we write

$$\lambda = 2^{-j_1} r_{\ell_1} \xi, \quad \lesssim' = 2^{-j_1-a} r_{\ell_2} \xi, \quad \gamma = 2^{-j_1-b} r_{\ell_3} \xi$$

with $0 \leq a < b \leq J - j_1$ and $J < \log L$. It leads to a scale and angle reparametrization of the scattering spectra:

$$\bar{S}(x)[\lambda, \lambda', \gamma] = \bar{S}(x)[j_1, a, b, \ell_1, \ell_2, \ell_3].$$

If $\bar{S}(x)$ has regular variations as a function of rotations then its three-dimensional Fourier transform along the (ℓ_1, ℓ_2, ℓ_3) has coefficients of negligible amplitude at high frequencies, which can thus be eliminated. One can also take advantage of regularities along scales. Since $1 \leq j_1 \leq J$ varies on an interval without periodicity, the Fourier transform is replaced by a cosine transform along j_1 for a and b fixed. We could also perform a cosine transform along the scale shift a and b , but this is not done in numerical applications because their range of variations is small and j -dependent. The Fourier transforms along j_1 is however sufficient to identify scale-invariance, since one then expects \bar{S} to only depend on a and b , see appendix C. We write $F \bar{S}(x)$ the Fourier transform of $\bar{S}(x)$ along (ℓ_1, ℓ_2, ℓ_3) and its cosine transform along j_1 .

Since F is unitary, it preserves the estimator variance:

$$\sigma_{\bar{S}}^2 = \mathbb{E}\{\|Aye F \bar{S}(x_i) - \mathbb{E}\{F \bar{S}(x)\}\|^2\}. \quad (\text{D1})$$

Ideally, the estimation error of $\mathbb{E}\{F \bar{S}(x)\}$ is reduced by eliminating

its coefficients whose squared amplitude is smaller than the variance of the empirical estimation error. It amounts to suppressing all coefficients having a variance that is larger than the bias resulting from their elimination. However, we can not implement this optimal "oracle" decision because we do not know $\mathbb{E}\{F \bar{S}(x)\}$. In this paper, we instead apply an approximate thresholding algorithm, which eliminates small amplitude coefficients of $\bar{S}(x)$ below a threshold proportional to their standard deviation, as discussed in the main text. This thresholding algorithm is adaptive and the selected coefficients vary from one process to another. For each process studied, an ensemble of between 20 to 100 samples $\{x_i\}$ were used to empirically estimate the average and variance of $F \bar{S}$, called $\mu(F \bar{S})$ and $\sigma(F \bar{S})$. The coefficients which have been kept are those that individually verify $\mu(F \bar{S}) > 2\sigma(F \bar{S})$.

A projected scattering spectra

$$\Phi(x) = P \bar{S}(x)$$

is computed with a linear Fourier projection P which eliminates all coefficients of $F \bar{S}(x)$ corresponding to coefficients of $\text{Ave}_i F \bar{S}(x_i)$ below their threshold. The efficiency of this projected scattering is the variance reduction ratio $\sigma_{P \bar{S}}^2 / \sigma_{\bar{S}}^2$ with

$$\sigma_{P \bar{S}}^2 = \mathbb{E}\{\|Aye P \bar{S}(x_i) - \mathbb{E}\{P \bar{S}(x)\}\|^2\}. \quad (\text{D2})$$

If $p(x)$ is isotropic or self-similar then we expect that P is a low-frequency projector along global rotations (which act similarly on all l_i coordinates) or scalings (which act on j), which corresponds to the averages described in (C7) and (C8). The Fourier projection P is however much more general and can adapt to unknown regularities of $p(x)$ along rotations and scales.

APPENDIX E: NUMBER OF COEFFICIENTS FOR SHELL BINNED POLYSPECTRA

For a 2D field, there are originally $O(L^4)$ bispectrum coefficients in total, as there are two independent frequencies in the bispectrum and each has two dimensions. If we take N_{bin} linear frequency bins along each side of L lattice points, the coefficients to be estimated is reduced to $O(N_{\text{bin}}^4)$. A rotation and parity average will further reduce and better estimate the bispectrum coefficients, which eliminates one dimension and leads to $\sim \frac{1}{A_3^3} \cdot \frac{1}{2} N_{\text{bin}} \cdot \frac{3}{4} N_{\text{bin}}^2 \cdot \frac{1}{2} = \frac{1}{32} N_{\text{bin}}^3$ binned coefficients, where $1/A_3^3 = \frac{1}{6}$ is the repeated counting of the three-frequency combinations in bispectrum, $\frac{1}{2} N_{\text{bin}}$ is the number of choice of k_1 , given rotation invariance, $\frac{3}{4}$ is the number of choice of k_2 given the requirement that each k is within the $L \times L$ lattice in Fourier space and $k_1 + k_2 + k_3 = 0$, and the factor $\frac{1}{2}$ comes from parity average.

For 2D fields, the shell-binned bispectrum is essentially a fast way to compute the rotation and parity average of the bispectrum. It does not mix very different configurations, because a given set of $|k_1|, |k_2|, |k_3|$ combined with the condition $k_1 + k_2 + k_3 = 0$ uniquely set the configuration up to free rotations. The number of coefficients is of the order $\sim \frac{1}{8} N_{\text{bin}}^3$ (the scaling power is 3 rather than $2d = 4$ because the orientation average eliminates one degree of freedom). For our choice of $N_{\text{bin}} = 10$, there are 151 shell-binned bispectrum coefficients. Similarly, the shell-binned trispectrum \bar{T} has $651 \sim \frac{1}{16} N_{\text{bin}}^4$ coefficients. Note that the shell-binning for trispectrum is more aggressive, because in 2D the same set of $|k_1|, |k_2|, |k_3|, |k_4|$ may come from different combinations k_1, k_2, k_3, k_4 even if the condition $k_1 + k_2 + k_3 + k_4 = 0$ is applied.

The ordering of \bar{B} and \bar{T} shown in Fig. 4 is determined in a nested

way. The frequency annuli are labeled by i from small to large $|k|$. To remove redundant coefficients, we require $i_1 \leq i_2 \leq i_3 (\leq i_4)$ and order them first by i_1 in increasing order; when two binning configurations have the same i_1 , they are then ordered by i_2 and so on.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.