Classification of Magnetohydrodynamic Simulations Using Wavelet Scattering Transforms

Andrew K. Saydjari1, Stephen K. N. Portillo2, Zachary Slepian3,4, Sule Kahraman5,6, Blakesley Burkhart6,7, and Douglas P. Finkbeiner1,8,9

1Department of Physics, Harvard University, 17 Oxford St., Cambridge, MA 02138, USA; andrew.saydjari@cfz.harvard.edu
2DIRAC Institute, Department of Astronomy, University of Washington, Seattle, WA 98195, USA
3Department of Astronomy, University of Florida, 211 Bryant Space Science Center, Gainesville, FL 32611, USA
4Physics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
5Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA
6Department of Physics and Astronomy, Rutgers University, 136 Frelinghuysen Rd., Piscataway, NJ 08854, USA
7Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave., New York, NY 10010, USA
8Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA

Received 2020 October 17; revised 2021 January 22; accepted 2021 February 6; published 2021 April 5

Abstract

The complex interplay of magnetohydrodynamics, gravity, and supersonic turbulence in the interstellar medium (ISM) introduces a non-Gaussian structure that can complicate a comparison between theory and observation. In this paper, we show that the wavelet scattering transform (WST), in combination with linear discriminant analysis (LDA), is sensitive to non-Gaussian structure in 2D ISM dust maps. WST-LDA classifies magnetohydrodynamic (MHD) turbulence simulations with up to a 97% true positive rate in our testbed of 8 simulations with varying sonic and Alfvénic Mach numbers. We present a side-by-side comparison with two other methods for non-Gaussian characterization, the reduced wavelet scattering transform (RWST) and the three-point correlation function (3PCF). We also demonstrate the 3D-WST-LDA, and apply it to the classification of density fields in position–position–velocity space, where density correlations can be studied using velocity coherence as a proxy. WST-LDA is robust to common observational artifacts, such as striping and missing data, while also being sensitive enough to extract the net magnetic field direction for sub-Alfvénic turbulent density fields. We include a brief analysis of the effect of point-spread functions and image pixelization on 2D-WST-LDA applied to density fields, which informs the future goal of applying WST-LDA to 2D or 3D all-sky dust maps to extract hydrodynamic parameters of interest.

Unified Astronomy Thesaurus concepts: Interstellar medium (847); Magnetohydrodynamical simulations (1966); Non-Gaussianity (1116); Convolutional neural networks (1938); Astronomy data analysis (1858)

1. Introduction

Interstellar dust and gas indirectly trace ISM turbulence, rendering them a vital observational lever on the processes that shape star formation and galaxy evolution (Elmegreen & Scalo 2004; Goodman et al. 2009; Burkhart et al. 2015; Padoan et al. 2016; Krumholz et al. 2018). Furthermore, dust in the ISM produces a foreground signal that must be removed from extragalactic measurements, such as cosmic microwave background (CMB) temperature and polarization anisotropy (Planck Collaboration et al. 2017), and the spectra of galaxies, supernovae, and stars (Cardelli et al. 1989; Corasaniti 2006). However, the complex interplay of magnetohydrodynamics (MHD), gravity, and supersonic turbulence introduces non-Gaussian correlations into the density fields, complicating both the modeling of the dust, and its subtraction for cosmology (Kandel et al. 2017; Kritsuk et al. 2018).

A useful statistical starting point for describing (real and mock) ISM density fields is the two-point correlation function (2PCF), or its Fourier-domain analog, the Fourier power spectrum (PS) (Lazarian et al. 2001; Kowal et al. 2007). However, the 2PCF cannot describe non-Gaussian processes, because the power spectrum ignores phase information, which is important for higher-order correlations (Peek & Burkhart 2019). Higher-order correlation functions, such as the three-point correlation function (3PCF) and its Fourier-domain analog, the bispectrum, represent an improvement in terms of capturing higher-order correlations (Peebles 2001). The 3PCF has had significant success in the field of cosmology, describing the non-Gaussian distribution of galaxies (Slepian et al. 2017a, 2017b, and references therein). The 3PCF (Portillo et al. 2018) and bispectrum (Burkhart et al. 2009; Burkhart & Lazarian 2016) have also been applied to MHD simulations, and show some discriminatory power, but interpretation of these higher-order statistics remains difficult.

The wavelet scattering transform (WST) provides another way to go beyond the 2PCF and exhibits several convenient properties. It is an approximately translation-invariant image representation, which is linear (Lipschitz continuous) in its deformations (Bruna & Mallat 2012; Mallat 2012). This means that if an image is slightly deformed from $i$ to $i'$, the difference in the WST coefficients describing the two images is bounded by the size of the deformation. The success of the WST in image discrimination derives from its encoding non-Gaussianity (higher-order correlation functions), which allows the WST to distinguish images with the same Fourier power spectrum (see Figure 5 from Bruna & Mallat 2012). Furthermore, unlike explicit computation of the higher-order correlations, which is susceptible to large variance arising from outliers,9 the WST coefficients are computed using non-expansive operators, and

---

9 An outlier with value $X$ will enter as $X^n$ in the N-point correlation function.
thus have lower variance. The variance of the WST coefficients in response to outliers rapidly decreases as the maximum spatial scale size, $J$, increases, because outliers are suppressed via spatial averaging.

The WST performs well with regard to standard classification benchmarks (e.g., the MNIST digit data set, LeCun & Cortes 2010), even with minimal training data. It is sufficiently robust to temporal instabilities to characterize 1D time series data, such as the measured velocity at a point in turbulent gaseous flows (Bruna et al. 2015). Both 2D and 3D formulations of the WST have been applied to density functional theory (DFT) calculations of molecular structure (Hirn et al. 2017; Eickenberg et al. 2018). In addition, progress has been made in terms of image reconstruction based on WST coefficients, using generative networks (Angles & Mallat 2018), and gradient descent (Bruna & Mallat 2018).

The general class of wavelet techniques have long found applications in astrophysics. Early works analyzed $^{13}$CO velocity fluctuations in star-forming regions via wavelet transforms (Gill & Henriksen 1990). The wavelet transform modulus maxima (WTMM) approach was used to analyze ISM emission maps, including anisotropies in H I (Khalil et al. 2006) from the Canadian Galactic Plane Survey (Taylor et al. 2003), and non-Gaussianity (Robitaille et al. 2014) in the Herschel infrared Galactic Plane Survey (Hi-GAL) (Molinari et al. 2010). However, the WST is not merely a wavelet transform. It comprises repeated convolution by wavelets, and pooling under a nonlinear function (the modulus).

While originally introduced in the deep-learning community for image discrimination and classification, the wavelet scattering transform (WST) has recently made inroads in astrophysics. The WST has found applications in cosmology, in relation to state-of-the-art, large-scale structure syntheses, and cosmological parameter inference, where it outperforms the power spectrum in the context of weak lensing (Allys et al. 2020; Cheng et al. 2020). Recently, a reduced form of the WST (RWST) was introduced, to characterize the ISM with interpretable coefficients. The RWST was subsequently applied to Herschel observations and MHD simulations (in 2D) (Allys et al. 2019), as well as polarized dust emission (Regalado-Saint Blancard et al. 2020).

The demonstrated capability of the WST to capture non-Gaussianity, its success in image classification problems, and prior applications to astrophysical data have motivated us to search for an optimal low-dimensional reduction of the WST for classification of dust density fields, in order to extract relevant hydrodynamic parameters. While convolutional neural nets (CNNs) have been trained to differentiate simulations that are sub/super-Alfvénic (Peek & Burkhart 2019), as well as entire perpendicular velocity fields from MHD simulations (Asensio Ramos et al. 2017), we show that the WST provides “training-free” classification when combined with a fast and deterministic dimensional reduction step.

2. Methods

2.1. Non-Gaussian Descriptors

2.1.1. Three-point Correlation Function (3PCF)

For a given density field, $f(x)$, correlations between three points, $f(x_1)f(x_2)f(x_3)$ for all $x_1$, $x_2$, $x_3$, can be used to construct higher-order statistics that capture non-Gaussianity. We generally seek summary statistics that are translation-invariant, so we compute the translation and rotation-averaged product over three points, which yields the 3PCF. The 3PCF then depends only on three coordinates, which can be parameterized by $r_{12}$, $r_{13}$, and $\theta$. Here, $r_{12}$ is the distance from some base point, $x_1$ to $x_2$, $r_{13}$ is the distance from $x_1$ to $x_3$, and $\theta$ is the angle between $\hat{f}_{12}$ and $\hat{f}_{13}$. Correlations between $x_1$, $x_2$, $x_3$ include contributions from the 2PCF between pairs of these three points (Groth & Peebles 1977). Setting the mean density to zero (i.e., $\int f(x)dx = 0$) removes this dependence of the 3PCF on the 2PCF (see Appendix B). By expanding the angular dependence in terms of multipole moments around $x_1$, this averaged-3PCF can be computed on a regular grid of $N_p$ points. However, the WST is not possible in 3D, because a triangle defined by $(r_{12}, r_{13}, \theta)$ is equivalent to that defined by $(r_{12}, r_{13}, \theta)'$ under rotations and translations. However, in 2D, these triangles can only be identified under reflection. Since previous works used only the even components of the 2D-3PCF coefficients (Zheng 2004; Chen & Szapudi 2005), we will provide comparisons using the real part of the coefficients in any case where we make use of the fully complex coefficients.

We modified an unpublished 2D version of the code described in Portillo et al. (2018), and increased the speed. The code we used is publicly available (see Section 6). We characterize a 3PCF computation by the number of radial bins selected for $r_{12}$ and $r_{13}$, and the maximum multipole moment ($L_{\text{max}}$) retained in the expansion. A 3PCF, with $N_{\text{bins}}$ and a maximum multipole moment of $L_{\text{max}}$, therefore has $(L_{\text{max}} + 1)N_{\text{bins}}(N_{\text{bins}} + 1)/2$ coefficients; the $L_{\text{max}} + 1$ comes from multipoles running from $L = 0,\ldots,L_{\text{max}}$, and the factor of 2 comes from the symmetry between $r_{12}$ and $r_{13}$. The 3PCF is averaged over the radii spanned by each bin. For optimal comparison to the WST, we have selected eight logarithmic radial bins, bounded by $2^j$, for $j = 0,\ldots,13$.13

2.1.2. 2D Wavelet Scattering Transform (2D-WST)

In 2D, a WST is specified by $(J, L, M)$, where $J$ is the maximum scale, $L$ is the number of angular bins (i.e., the angular spacing is $180^\circ/L$), and $M$ is the maximum order of the transform. The maximum scale and number of angular bins

---

10 A non-expansive operator is an operator, $M$, such that $|Mx - My| \leq |x - y|$, where $|x|^2 = \int |x(u)|^2 du$. In the context of the WST, $M$ is simply the usual modulus for complex numbers, which is non-expansive by the triangle inequality. This property ensures stability to additive noise, because small changes in the field cannot cause more than the same small changes in the coefficients. Furthermore, an outlier with value $X$ always enters the $n$th-order scattering coefficient proportionally to $X$ (see Section 2.1.2).

11 Alfvén waves are transverse waves for which magnetic tension provides the restoring force, and the Alfvén speed scales as $\sqrt{B}$; one often measures turbulent velocity magnitudes relative to the Alfvén speed.

12 We verified that the real and imaginary parts of the 3PCF coefficients are even and odd under reflection as expected.

13 By including $j = 8$, triangles with large radii and large angles can probe pixels that are highly correlated in the presence of periodic boundary conditions.
The wavelet transform we consider here is based on a specific instance of wavelet decomposition, which is a more general technique. In fact, a similar wavelet decomposition was previously performed on the MHD simulations we study (Kowal & Lazarin 2010), although that decomposition was not used as the basis for the WST, or as inputs for a classification algorithm.

While the maximum spatial scale is $2^J$, the wavelet indices refer to the wavenumber (in pixels) in the Fourier domain. Thus $j = J - 1$ in the Fourier domain is the minimum possible scale for a spatial domain of $2^J$.

$$(J, L)$$ define the set of Morlet wavelets $\psi_{j,l}$ used in the transform. The parent Morlet wavelet is the product of a plane wave of unit wavenumber with a Gaussian window of width $\sigma$. In one dimension, the wavelet is given by

$$\psi(x) = \alpha(e^{i\frac{2\pi}{\sigma}x} - \beta)e^{-x^2/(2\sigma^2)}$$

where $\alpha$ and $\beta$ are normalization factors (Ashmead 2010). We choose $\alpha$ so that the wavelets cover the Fourier domain as uniformly as possible. We choose $\beta$ such that $\psi$ has a null average.

In 2D and above, the inverse variance, $1/\sigma^2$, of the Gaussian window becomes an inverse covariance matrix. We choose a diagonal covariance matrix, with $\sigma_{yy} = 2\sigma_{xx}$ (where $x$ is the direction of the plane wave), to enhance the directionalities of the wavelets. We choose $\sigma_{xx} = 0.8$ for our 2D-WST, which limits the number of oscillations of the plane wave, thereby enhancing filament identification. Each wavelet used in computing the WST coefficients is said to represent an orientation, indexed by $(j, l) \in \{0, \ldots, J - 1\}, \{0, \ldots, L - 1\}$, and is obtained from the parent wavelet by rotation and scaling of the argument (by $2^j$) and frequency (by $2^{-j}$). While we use $l$ as an index, we denote the corresponding angles as $\theta \in [\pi/2 - \pi/L, \ldots, (L - 1)\pi/2L, \ldots, -\pi/2]$. The wavelet, $\psi_{j,l}$, has a bandwidth of $\sim 2^{-j}$ and angle $\theta$, corresponding to $l$ in the Fourier domain (Figure 1).

The WST can be represented as a convolutional network, with layers labeled $m$ from 0 to $M$. We will outline the computation of coefficients for each layer for a network where $M = 2$. Coefficients for the $m$th layer are derived from $m$ convolutions of the image field $I(x)$, with Morlet wavelets under the usual modulus for complex numbers. Normalization of the coefficients is relative to the response to a Dirac delta distribution, $\delta_{2D}(x)$:

$$S_0 = \frac{1}{\mu_0} \int I(x) d^2x$$

Here, $\mu_0$ is the indicator function, which is 1 everywhere in the image. For $\mu_0$ and $\mu_1$, the convolution simplifies to the area of the image, and the sum of $|\psi_{j,l}|^2$ over the image, respectively. Note that first-order coefficients depend on one oriented scale $(j, l)$, while second-order coefficients depend on two oriented scales, $(j_1, l_1)$ and $(j_2, l_2)$. Since the first convolution obscures information at scales below $2^{J_0}$, only coefficients with $j_2 > j_1$ are computed, such that one is then only looking at scales $2^{J_0} > 2^J$. We have one coefficient for $m = 0$, $JL$ coefficients for $m = 1$, and $J(J - 1)L^2/2$ coefficients for $m = 2$, for a total of $J(J - 1)L^2/2 + JL + 1$. The WST coefficients at the $m$th layer depend on correlation functions up to order $2^m$ (proof in Mallat 2012). Thus, $M = 1$ is required to capture non-Gaussianity, because for $m = 1$, the WST coefficients depend predominantly on the 2PCF. For many applications, the spectral energy falls off rapidly with $m$; as such, $M = 2$ provides sufficient sensitivity to non-Gaussianity with a tractable number of coefficients.

We normalize the 2D-WST coefficients following Allys et al. (2019), and Bruna et al. (2015), as

$$\tilde{S}_0 = \log_2[S_0]$$

$$\tilde{S}_1(j_1, l_1) = \log_2[S_1(j_1, l_1)/S_0]$$

$$\tilde{S}_2(j_1, l_1, j_2, l_2) = \log_2[S_2(j_1, l_1, j_2, l_2)/S_1(j_1, l_1)].$$

This normalization makes the first- and second-order coefficients invariant under multiplication of the field by a constant factor, and makes the second-order coefficients invariant to modification of the spectrum of the field by the action of a class of linear filters (Bruna & Mallat 2018).

**2.1.3. 3D Wavelet Scattering Transform (3D-WST)**

In 3D, the WST coefficients are defined similarly, except that solid harmonic wavelets (denoted by $\psi^m_j$) are used in place of Morlet wavelets, with $m \in \{-l, \ldots, l - 1, l\}$.

$$\psi^m_j(x) = \frac{1}{\sqrt{(2\pi)^3}} e^{-|x|^2/2} |Y^m_j(x)|,$$
the first- and second-order coefficients. For first-order coefficients, this is explicitly

\[
S_i(j_i, l_i) = \frac{1}{\mu_i} \int \left( \sum_{m=\text{-}l_i}^{l_i} |l \ast \psi_{j_m l_i}(x)|^2 \right)^{q/2} \, d^3x
\]

where \( q = 1 \). For the 3D-WST, \( S_i(j_1, j_2, l_i) \) is a function of one angular scale only, we take \( l_2 = l_i \). This means that the second-order coefficients are obtained between scales at the same angles only (rather than the Cartesian product of all angles). We compute the convolution coefficients for \( q = 1/2 \) and \( q = 2 \), in addition to \( q = 1 \), which is the exclusive power used in 2D-WST, although we comment only briefly on the \( q = 1 \) cases.\(^{18}\) There are \((L + 1)(J + 1)\) coefficients for \( m = 1 \), and \((L + 1)(J + 1)/2\) coefficients for \( m = 2 \), for a total of \((L + 1)(J^2 + 3J + 2)/2\) (for each power \( q \) computed).

2.1.4. Reduced Wavelet Scattering Transform (RWST)

The RWST reduction (which is only defined in 2D) takes advantage of the periodicity observed in the WST coefficients to remove the angular dependence (see Appendix C). Explicitly, the RWST coefficients are obtained from the WST coefficients by a least-squares fit of the first- and second-order coefficients. For the first order,

\[
\tilde{S}_1(j_1, l_i) = S_1^{\text{iso}}(j_1) + S_1^{\text{aniso}}(j_1) \cos \left( \frac{2\pi}{L} (l_i - l_i^\text{ref}) \right)
\]

where \( S_1^{\text{iso}}(j_1), S_1^{\text{aniso}}(j_1), \) and \( l_i^\text{ref}(j_1) \) are fit coefficients. For the second order,

\[
\tilde{S}_2(j_1, l_1, l_2) = S_2^{\text{iso},1}(j_1, j_2) + S_2^{\text{iso},2}(j_1, j_2) \cos \left( \frac{2\pi}{L} (l_1 - l_2) \right)
\]

\[
+ S_2^{\text{aniso},1}(j_1, j_2) \cos \left( \frac{2\pi}{L} (l_i - l_i^\text{ref}) \right)
\]

\[
+ S_2^{\text{aniso},2}(j_1, j_2) \cos \left( \frac{2\pi}{L} (l_2 - l_b^\text{ref}) \right)
\]

where \( S_2^{\text{iso},1}(j_1, j_2), S_2^{\text{iso},2}(j_1, j_2), S_2^{\text{aniso},1}(j_1, j_2), S_2^{\text{aniso},2}(j_1, j_2), \) and \( l_b^\text{ref}(j_1, j_2) \) are fit coefficients. There are \( 3J \) coefficients for \( m = 1 \), and \( 5J(J - 1)/2 \) coefficients for \( m = 2 \), for a total of \( 5J(5J + 1)/2 \).\(^{19}\)

2.2. Dimensional Reduction Algorithms

In order to interpret these high-dimensional (>100 coefficients), non-Gaussian descriptors, we employ dimensional reduction techniques, including principal component analysis (PCA), linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA).\(^{20}\) PCA is an unsupervised technique, in which the \( i \)th principal component is orthogonal to the \( i - 1 \) principal components, and is in the direction of maximal remaining variance (Jolliffe & Cadima 2016). In contrast, LDA and QDA are supervised techniques, meaning that they take class labels (here, hydrodynamic parameters) in addition to the high-dimensional descriptor of a data point (here, the WST, RWST, or 3PCF coefficients). LDA assumes that each component of the high-dimensional descriptor is normally distributed, and that all classes share the same covariance matrix; this latter assumption is known as homoscedasticity (Bandos et al. 2009). LDA then finds the hyperplanes that best separate the \( K \) classes in the \( K - 1 \) dimensional space containing the class means. The hyperplanes are found using eigenvalue decomposition, involving the inter- and intra-class scatter matrices. This is equivalent to evaluating the Gaussian likelihood that each input is in each class, and assigning the input to the class with the greatest likelihood. Note that the LDA components (which maximize the separation between classes) need not be orthogonal. The LDA components are in units of \( \sigma_{ij} \), the on-diagonal components of the shared covariance matrix. QDA is similar to LDA, except that it drops the assumption that the covariance matrix must be the same for all classes, and thus finds conic sections that best separate the classes, rather than hyperplanes. The boundary between two neighboring classes occurs where the Gaussian likelihoods of each class are equal, which is in general a quadratic function (as in QDA), but simplifies to a linear function when the class covariances are assumed equal (as in LDA).

3. Data Sets

In this section, we introduce the observational data sets and MHD simulations used to verify the sensitivity of our method to non-Gaussianity and test its capabilities in terms of the classification of turbulence.

3.1. SFD Dust Map

Interstellar dust is an indirect tracer of interstellar turbulence, and provides an observational point of comparison to the MHD simulations discussed below. Dust grains are heated to \( \sim 200 \)K by ambient starlight, and produce thermal emissions in the far-infrared (Low et al. 1984). We use the SFD dust map (Schlegel et al. 1998), which is based on this thermal emission. The map derives dust morphology from 100 \( \mu \)m data from the Infrared Astronomical Satellite (IRAS) Sky Survey Atlas (Neugebauer et al. 1984; Wheelock et al. 1994), corrected for temperature variation using data from the Diffuse InfraRed Background Experiment (DIRBE) on the Cosmic Background Explorer (COBE) (Mather 1982; Boggess et al. 1992).

To obtain a representative sample of 2D dust images, we reproject the SFD dust map with bilinear interpolation to a TAN (gnomonic) projection, centered on random locations on the celestial sphere. Each 256\(^2 \) pixel image has a random position angle, a pixel scale of 3\(^\prime\)/0, and a point-spread function (PSF) full width at half maximum (FWHM) of 6\(^\prime\)1. The SFD bitmask flags regions with missing IRAS data (bit 128), pixels

\(^{18}\) For \( q = 1/2 \), small nonzero values are upweighted, providing a measurement of sparsity. For \( q = 1 \), the WST coefficients simply scale linearly with density. For \( q = 2 \), [density]\(^2 \) interactions are upweighted.

\(^{19}\) After the completion of this work, we became aware of PyWST, a PYTHON implementation of the RWST, available at https://github.com/bregaldo/pywst. Our own implementation can be found in Section 6. We compared the two implementations on the same images, and obtained similar results, except that PyWST includes \( S_q \) in the RWST coefficients. Importantly, we also bound the phases to a single interval, [0,2\( \pi \)]. This provided improved convergence, but also introduced minor discrepancies between the coefficients obtained by the two methods, most prominently in \( S_2^{\text{aniso},1}, S_2^{\text{aniso},2} \), and the reference angles themselves.

\(^{20}\) Previous attempts to reduce the dimensionality of the bispectrum have been made (Burkhart & Lazarian 2016), but the application was “not straightforward.”
in the Magellanic clouds or M31 (bit 64), and regions with no point-source subtraction (mostly at low Galactic latitude, |b| < 5°; bit 32). An image is discarded if any pixel has any of these mask bits set, leaving a sample of 10,222 images (Figure 2).\footnote{Since the full sky is only \(\sim 40,000\) square degrees, there are only \(\sim 200\) independent patches of sky represented in this sample, with additional images coming from translations, rotations, and partial overlaps. Since the WST is not fully rotationally and translationally invariant (the subject of our upcoming work), these additional images are useful, even if partially redundant.} By rejecting images near the Galactic plane, we focus on the morphology of relatively nearby dust (on the order a few 100 pc to a few kpc), where each image is dominated by dust at one, or a few, distances. The statistics of the dust distribution near |b| = 0 might be substantially different, owing to the large number of clouds along each line of sight, and the steep dependence on latitude. The images are normalized to have [min,max] = [0,1] for input to the WST, and zero mean with unit variance for input to the 3PCF, consistent with other images used in the paper.

3.2. Magnetohydrodynamic Simulations

We use MHD simulations of turbulent gas flows published in previous works (Cho & Lazarian 2003; Burkhart et al. 2009; Portillo et al. 2018; Bialy & Burkhart 2020; Burkhart et al. 2020), and review only key properties. The simulations solve the ideal MHD equations constraining the divergence to zero (\(\nabla \cdot \mathbf{B} = 0\)), assuming an isothermal equation of state, and periodic boundary conditions. No gravity is included, and the simulations are “scale-free.” The simulation box measures \(256^3\) voxels, and is driven isotropically by random, divergence-free forcing, with wavenumber \(k \approx 2.5\) (i.e., 1/2.5 the box size). No dissipation is assumed, other than numerical dissipation from the finite resolution of the simulation. The magnetic field is initialized uniformly in the \(x\)-direction.

Each simulation is characterized by its Alfvenic Mach number \(M_A\), and sonic Mach number \(M_S\). The Alfvenic Mach number is defined as \(M_A \equiv |v|/|v_A|\), where \(v\) is the flow velocity, and \(v_A\) the Alfven (magnetic wave) speed averaged over the simulation box. The sonic Mach number is defined as \(M_S \equiv |v|/c_s\), where \(c_s\) is the isothermal speed of sound. We use \(M_A = 0.7, 2.0\), and \(M_S = 0.7, 1.0, 2.0, 4.0\) to include sub-Alfvenic \((M_A = 0.7)\) and super-Alfvenic \((M_A = 2.0)\) cases, as well as subsonic \((M_S = 0.7)\), transonic \((M_S = 1.0)\), and supersonic \((M_S = 2.0, 4.0)\) cases (Figure 3). The images are normalized to have [min,max] = [0,1] for compatibility with the log normalization of the WST, but we have confirmed that our WST classification result is reproduced if one uses zero mean and unit variance normalization. The images are
normalized to have zero mean and unit variance when computing the 3PCF, in order to exclude any contributions from the 2PCF (see Appendix B).

4. Results and Discussion

We begin by comparing the performance of the WST, RWST, and 3PCF in capturing non-Gaussianity. We then combine these statistics with LDA to classify turbulent MHD simulations, and investigate the robustness of WST-LDA in relation to modifications of the sample images.

4.1. Sensitivity to Non-Gaussianity

We assess the performance of the WST, its reduction the RWST, and the 3PCF as high-dimensional descriptors for the characterization of non-Gaussianity in 2D images. Each of these descriptors is then used in conjunction with two dimension-reduction techniques, PCA, and LDA. PCA and LDA are used to probe the information contained in the high-dimensional parameter spaces, corresponding to the WST, RWST, and 3PCF coefficients. We begin by confirming that these descriptors are sensitive to non-Gaussianity by applying them to 2D dust map images. We compute the WST coefficients ($J = 8, L = 8, M = 2$) on 10,222 SFD images (“SFD”).

2000 images where each pixel is a uniformly drawn pseudo-random number [0,1] (“Rand”), and those same random images, convolved with a $\sigma = 4$ pixel uniform Gaussian (“Smooth”). This latter set imposes nearest-neighbor correlations, which are present in empirical images due to the continuity of real density fields and instrumental point-spread functions. To further probe that the WST captures non-Gaussianity, we compute the WST coefficients for images where the Fourier power spectra matches the SFD images, but with small higher-order correlations (“NHC”). Specifically, for each image in the SFD set, we apodize the image (i.e., multiply by a generalized Gaussian, $p = 6, \sigma = 100$), compute the Fourier transform (FT), randomize the phase (with a unique pseudo-random seed), and compute the inverse FT. The RWST and 3PCF ($N_{bias}, L = (8, 9)$) coefficients are also computed on the same four data sets for comparison.

To investigate the variance of these coefficients with respect to prototypical 2D dust, we perform PCA on the coefficients from the SFD set and then project the results from the other three control image sets (“Rand,” “Smooth,” “NHC”) onto those principal components. Each point in Figure 4 corresponds to an image from one of the data sets above.

In WST-PCA space, the SFD data set presents as an ellipsoid, with a population of outliers extending along the first PC axis in the positive direction. The “NHC,” “Smooth,” and “Rand” data sets present as progressively narrower ellipsoids overlapping the main mass of “SFD.” While slight, the different orientation of these ellipsoids may encode more detailed information differentiating the data sets. We observe

that, to first order, the diameter of the ellipsoids decreases with the progressively decreasing pixel – pixel correlations in the “SFD” > “NHC” > “Smooth” > “Rand” series. This suggests that the variance in the WST coefficients on “SFD” images primarily arises from higher-order correlations, as desired. As shown in the cumulative variance plot, most of the variance of “SFD” is described by the first few PCA components, suggesting that WST-PCA admits a low-dimensional representation of SFD dust. By cross-referencing the Galactic coordinates of the images associated with the outliers in the “SFD” data set, using GLUE, we associated most of the outliers with images of the Orion nebula. The formation of massive stars in the Orion nebula modifies the dust distribution, and is likely to give rise to outliers in the WST-PCA space. This was also true for the outliers in the RWST-PCA and 3PCF-PCA bases. The sensitivity of WST-PCA to outliers may suggest its potential application in the area of outlier detection and flagging in large survey pipelines.

When the input coefficients to PCA are not scaled, the mean and variance of the WST coefficients are sufficient to partially separate all four data sets in WST-PCA space (see the interactive version of Figure 4). This is in contrast to what is expected for a simple $N$-point correlation function, which should have an expectation value of zero for $N > 2$ on “NHC,” “Smooth,” and “Rand.” While exclusion of the $S_0$ coefficient does not visibly inhibit class separation, we cannot rule out that information regarding the mean of the density distribution does

$^{22}$The density of dust in the ISM is sometimes assumed to be a log-normal distribution; as such, we also verified these results on log [density].

$^{23}$A generalized Gaussian has the form $f(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, with an FWHM of $2\sigma \log(2) / \sqrt{2\pi} \sqrt{\ln 2}$. Note that for $p = 0.5$ this is the Laplace distribution, and for $p = 1$ this is the Gaussian distribution.

$^{24}$We confirmed that the results obtained for “SFD” do not change for apodized “SFD” images. This implies that the results for “NHC” stem from phase randomization, and are not just an effect of apodization.

$^{25}$We scale the coefficient inputs to PCA to have zero mean and unit variance, to account for any differences in magnitude between the coefficients.

$^{26}$GLUE (Beaumont et al. 2015; Robitaille et al. 2017) is a PYTHON package for data visualization and linking, available at https://glueviz.org/.
enter the higher-order coefficients, owing to the \([\text{min},\ \text{max}] = [0,1]\) normalization we chose. We therefore believe that a more careful study of the distribution and expectation values for the WST coefficients is called for in future work. The distributions of the four data sets under RWST-PCA are similar to the distribution under WST-PCA, except that the ellipsoidal distributions have lower eccentricity. The unscaled RWST-PCA basis also partially separated the four classes, and further improved the separation of “Rand” to have no overlap with the other three classes. The cumulative variance rises much more slowly as a function of the number of principal components, suggesting that the RWST does not admit as low-dimensional a representation, and that the variance in “SFD” under the RWST is similar for many directions.

We select the real part of the 3PCF coefficients for analysis by PCA. The distribution of the four data sets under 3PCF-PCA is similar to that under WST-PCA, except that “Smooth” and “Rand” have higher and lower eccentricities, respectively. The unscaled 3PCF-PCA basis separates “SFD” from the other three data sets, with little overlap. However, the expectation value of the 3PCF for “NHC,” “Smooth,” and “Rand” must be zero. As such, the centroids for all three data sets are coincident. Despite overlapping centroids, the major axes of “NHC” and “Smooth” are distinct, allowing partial separation. The distribution of “Rand” is so tight and circular that it overlaps entirely with “NHC” and “Smooth.” The cumulative variance of the 3PCF is between that of the WST and RWST, suggesting the 3PCF admits an intermediate-dimensional representation of SFD dust. We repeated this analysis, taking the modulus of the 3PCF coefficients instead. While the exact shape of the distribution for each data set changed, and the centroid of “NHC,” “Smooth,” and “Rand” shifted, the separation was not strongly affected.

The lack of complete separation in the PC basis does not mean that these descriptors do not contain the information necessary to distinguish the four classes. PCA is an unsupervised technique, and as such cannot optimize for the separation of classes, as it has no knowledge of them. To determine whether the computed coefficients can distinguish the four classes, we turn to a supervised learning technique, LDA. We split the data sets into test/train sets (80%/20%) and compute the confusion matrix for the test set between the true labels and the assignment predicted by the LDA model, which was trained on the first set. In the confusion matrix, we show the actual number of images having a true class label that were predicted to have a certain class label. The grayscale coloring of the matrix indicates the true positive percentage relative to the number of test images, with the corresponding true label (the row sum). In this way, the row sum is always the test fraction \((\sim 0.2)\) times the total number of images in each class. To quantify the predictive power of LDA on this multi-class problem, we report the precision score, which is the total fraction of correctly classified images (true positive percentage).

Figure 5 illustrates that the WST contains enough information to completely distinguish the four classes, and can thus at least partially characterize the non-Gaussianity of SFD dust. The RWST performs similarly well, apart from the presence of an outlier in the RWST-LDA space that is not associated with any of the clusters. These outliers tend to result from undertraining, i.e., failing to fully constrain non-discriminatory dimensions. While not shown, we also analyzed an RWST model, where the frequency in Equations (6) and (7) were fit in addition to the usual RWST parameters. This unfortunately created a large degeneracy between the reference angle and frequency, which caused fitting instability and massive variability that were common to all data sets (not discriminatory). Outliers may also arise when the cosinusoidal dependence is so small in amplitude that the angle is undefined, or when the cosinusoidal frequency differs from the model. These points demonstrate that more work remains to be done to regularize the RWST technique.

For improved classification, we take the modulus of the 3PCF coefficients as inputs to LDA. Under these conditions,
3PCF-LDA almost completely distinguishes the four classes, with only a few images of “SFD” confused for “NHC”, and one extreme outlier. Viewed in the interactive version of Figure 5, these confused images between “NHC” and “SFD” are associated with the “SFD” cluster, suggesting that the misclassification derives from the strictly linear boundaries to which LDA is limited. In addition, the assumption of equal class covariances (homoscedasticity) is strongly violated for the 3PCF, because the variance is strongly class-dependent (Figure 5, bottom right). To attempt to further improve classification with 3PCF, we use QDA, which relaxes the homoscedasticity condition, and removes the confusion between “NHC” and “SFD.” In all other sections of the manuscript, LDA and QDA produce nearly identical results, and so only the simpler LDA results are presented. When attempting classification with only the real part of the 3PCF coefficients, there was a large confusion between “NHC,” “Smooth,” and “Rand,” which are forced to have exactly overlapping centroids. This condition is lifted upon taking the modulus, since the expectation value of the modulus of the 3PCF coefficients need not be zero on images drawn from Gaussian distributions.

In summary, all three non-Gaussian descriptors contain enough information to distinguish “SFD,” “NHC,” “Smooth,” and “Rand,” but with the WST, the classes can be distinguished using a simple linear classifier (in contrast to classification with the 3PCF), and without outliers (in contrast to classification with the RWST).

4.2. Classification of MHD Simulations

Since WST, RWST, and 3PCF distinguish Gaussian from non-Gaussian fields, we seek to use them to classify different physical processes that can contribute to non-Gaussianity. One application with implications for the ISM is inferring characteristic MHD parameters, such as the sonic and Alfvénic Mach number. We demonstrate the ability of WST-LDA to do this on idealized MHD simulations (Figure 3), as a proof of concept. For each of the eight conditions (sonic/Alfvén Mach number pairs) we have the full density cube (256\(^3\) pixels) at nine time steps. Because the separation between these time steps is longer than the eddy turnover time, each can be treated as an independent sample for the given condition. For each cube, we take 2D slices (256\(^2\) pixels) along each axis (x, y, z), spaced every eight pixels so as to reduce the sampling redundancy. Therefore, we computed the WST coefficients for 6,912 (8 \(\times\) 9 \(\times\) 32 \(\times\) 3) images, or 864 images per class. To ensure that the test set is not correlated with the training set, all images in the test set are selected from two independent time steps, withheld from the training set. Since each time step is an independent sample, the test and train sets are thus independent.

We achieve a precision of 97\% for WST-LDA (J = 8, L = 8, M = 2; \(N_{\text{coeff}} = 1, 857\), 98\% for RWST-LDA (\(N_{\text{coeff}} = 164\), and 87\% for 3PCF-LDA (\(N_{\text{bins}} = 8, L = 9; N_{\text{coeff}} = 360\)) (Figure 6). The precision for WST-LDA is comparable to that obtained previously for differentiating sub/super-Alfvénic simulations after training a neural net (Peek & Burkhart 2019), but notably required no training except for a deterministic LDA step. As shown in the interactive version of Figure 6, the first three LDA components separate each of the classes. There is a small amount of crosstalk between adjacent classes, due to scatter of the WST coefficients of images in the same class.

Since these MHD parameters are actually continuous variables, the variance of each class can be viewed as an indicator of the uncertainty of an inference of these MHD parameters from a single image. To obtain a rough estimate of this uncertainty, we assume \(M_S\) varies linearly as a function of separation between clusters of the same \(M_A\) value. Then, we find that \(M_S\) only varies \(\sim 10\%\) within \(1\sigma\) of the centroid of each cluster. Thus, for these simulations, we can approximately infer \(M_S \pm 10\%\). Note that the sonic Mach numbers studied here are lower than the average estimated for molecular clouds in the ISM from CO line widths (\(\sim 10–20\), Kainulainen & Tan 2013; Kainulainen & Federrath 2017), but comparable to measurements for 21-cm cold neutral medium (CNM) (Heiles & Troland 2003; Burkhart et al. 2010). However, whether this approximately linear mapping of \(M_S\) to LDA space holds at larger \(M_S\) remains an open question. We hold off on a comparison between observations and simulations until we can characterize a wider range of Mach numbers and other physical parameters.

In this WST-LDA space, the classes of images with different Alfvénic Mach number form a helical structure (Figure 3, interactive version). Along the vertical axis of the helix, the sonic Mach number varies monotonically while the separation in the plane perpendicular to this axis determines the Alfvénic Mach number. We can also see this behavior in the 2D projections, where \(LD_0\) is the helical axis (Figure 6). There is little structure as a function of Mach numbers in the 4th and higher LDA components, suggesting the classification occurs primarily in a 3D subspace of the WST-LDA basis.

The increased precision of RWST-LDA as compared to WST-LDA appears to derive from reduced scatter around the centroids of each cluster, although the centroids are separated by the same distance. Since LDA selects the optimal linear combination of the WST coefficients, the success of RWST is not simply in selecting an optimal subset of the coefficients containing the most information; the RWST is not simply performing dimensional reduction. The nonlinear combination of WST coefficients used by the RWST preferentially selects information relevant to differentiating the MHD classes.

For improved classification, we take the modulus of the 3PCF coefficients as inputs to LDA, and find partial separation of the classes. We find that the variance is strongly class-dependent, with higher sonic Mach numbers having larger variance, and lower sonic Mach numbers having smaller variance. This suggests that the variance of the 3PCF is sensitive to sharp density fluctuations. Furthermore, the spacing of the centroids is highly nonlinear. The small spacing between these centroids for \(M_S = 0.7\) is the source of most of the confused images. This small spacing suggests the contribution of an important influence from \(M_A\) in the fourth-order moments, which is present in the WST, but not in the 3PCF. As aforementioned, for cold gas in the ISM, \(M_S\) is likely to be > 0.7; as such, this lack of discriminatory power for low \(M_S\) may not be a problem for some applications. The fact that gas near young hot stars has lower \(M_S\) suggests that it may be more

---

29 Test-train precision for real part of the 3PCF coefficients was 84\%.
30 The correlation between slices (which are 1 pixel thick) does not fall to zero until there is a spacing of \(\sim 50\) pixels, so sampling every 8 slices does introduce correlations for samples within the same time step. This correlation necessarily reduces the effective sample size of our data set.
31 QDA can improve the precision to 96\%, which is to be expected, given the high heteroscedasticity.
difficult to apply WST-LDA to active star-forming regions (Gallegos-Garcia et al. 2020).

Since dust maps such as SFD project a 3D volume along the line of sight down to 2D, we investigate how WST-LDA, RWST-LDA, and 3PCF-LDA perform with respect to projected images by repeating our sampling procedure on the cumulative sum of density along a given coordinate axis. This is equivalent to sampling the column density along the line of sight in the optically thin limit. We computed the WST coefficients for the same number of images as before, i.e., 6912. Note that the line of sight sums introduce further correlations, thereby reducing the effective size of our data set, and can be seen as clusters or lines within a given class. We find a precision of 57% for WST-LDA, 83% for RWST-LDA, and 44% for 3PCF-LDA (see Appendix D, and the figure therein). The improvement provided by the nonlinear

Figure 6. Confusion matrix between MHD classes labeled by [sonic, Alfvén] Mach numbers, when applying WST-LDA (top left), RWST-LDA (middle left), and 3PCF-LDA (bottom left). Matrix entries are the number of test images with the corresponding true-predicted label pairing. The grayscale color indicates the true positive percentage relative to the number of test images with the corresponding true label. In all cases, colors correspond to true labels, as indicated on the vertical axis in the left panels. Test images are shown in the first two components of LDA space for the WST (top right), RWST (middle right), and 3PCF (bottom right). Circle markers represent images that were correctly assigned, and cross markers represent those incorrectly assigned. The RWST outlier in (4.0, 2.0) classified as (4.0, 0.7) is far outside the axes limits. An interactive version of this figure is available at https://fasm.rc.fas.harvard.edu/saydjari/RWST_2020/MHD_LDA.html, showing the images in LDA space, and allowing isolation of the train, correctly labeled test images, and incorrectly labeled test images. The WST, RWST, and 3PCF in combination with LDA all classify the MHD simulations with high precision, although we prefer the linear spacing of classes in the (R) WST-LDA basis.
transforms in the RWST becomes even clearer in the context of cumulative sums, and we anticipate that this will be instrumental in applications featuring multiple clouds along the line of sight.

Overall, the WST and RWST seem most suited to the continuous inference problem of estimating Mach numbers from measured density fields, since $M_S$ maps approximately linearly into the LDA classification space, in contrast to the highly nonlinear mapping for 3PCF-LDA. While the nonlinear transformations of WST coefficients made by the RWST improve the classification, as well as rendering the classification more robust to cumulative sums, future work should seek an optimal nonlinear transformation of WST coefficients that does not also create outliers. Since we do not yet have such an optimal transformation, we focus here on characterizing the performance of WST-LDA on several test cases.

### 4.3. 2D/3D Scale Dependence of WST-LDA

It is important to establish how many coefficients the WST needs to characterize a non-Gaussian process. However, this question is intertwined with that of the scale at which the process of interest is occurring, and how many different scales the WST needs to characterize a non-Gaussian process. We study this trade-off in our MHD test case by computing those scattering coefficients that include only a subset of the scales in the problem. Initially, we compute the WST coefficients for $L = 8$, $M = 2$, and $J = 3, 4, 5, 6, 7$, and 8 to investigate the effect of the maximum scale ($J_{\text{max}}$). We can follow a similar process to find $J_{\text{min}}$ by subsampling the $256^2$ image down to a $(2^{2} - L^2)^2$ image (i.e., binning down by a factor of $2^{-\text{min}}$), and computing the WST with $L = 8$, $M = 2$, and $J = 8, 7, 6, 5, 4, 3$. Both results are shown in Figure 7. Note that the $J_{\text{min}}$ (var) and $J_{\text{max}}$ (var) curves cross at $j = 4$. The $j = 0$ (smallest spatial scale) is likely to contain little information if the image is well sampled. Above the $j = 0$ scale, the linearity of each of the data sets suggests that there is approximately the same information content in each scale.

The number of coefficients (the dimension of the WST vector) decreases as we narrow the range of scales in the above analysis. To test whether the loss of information results explicitly from reduction in the number of coefficients versus loss of information in the image due to binning effectively eliminating certain scales, we generate a series of images of fixed size but with increasing redundancy. We leverage the fact that these simulations have periodic boundary conditions, and can therefore repeat a given image in a $2^n \times 2^n$ matrix, downsample the image to $256 \times 256$ resolution, and always compute the WST for $J = 8$, $L = 8$, $M = 2$. These results ($J_{\text{min}}$ “constant”) are always within 6% of the $J_{\text{max}}$ “variable”, so we conclude that the trend observed in the $J_{\text{min}}$ “variable” is predominantly the result of the scales excluded, and is not due to the number of coefficients.

While the highest-resolution maps of the ISM are 2D, the resolution of 3D maps is improving (Green et al. 2019; Leike et al. 2020). Owing to the fact that the physical processes driving the structure of the ISM are happening in 3D, one may expect a measurement of non-Gaussianity that is computed in 3D to capture more information. We show that 3D-WST-LDA ($J = 6, L = 8, M = 2$) is comparable to 2D-WST-LDA in terms of separating the eight classes in our set of MHD simulations when the same scales are included (Figure 7). 3D-WST-LDA specifically outperforms 2D-WST-LDA by $\sim 5\%$ for $j = 5, 6$. Note that 3D-WST is a strictly lower dimensional representation for all studied, as we do not compute second-order coefficients between different angular bins in the 3D-WST procedure. 2D-WST and 3D-WST also use different wavelets (see Equations (1) and (4)) and so we leave a more detailed comparison of the different bases to future work. 3D-WST-LDA can clearly provide similar classifications to 2D-WST-LDA, though the increased computational expense means it is only likely to be used when 3D correlations are necessary (see Appendix A for computational costs).

### 4.4. Robustness of Classification

#### 4.4.1. Using Velocity Information

While few 3D maps of the ISM exist in position–position–position (PPP) space, and those that do exist are of relatively low resolution (1 pc; Leike et al. 2020), high-resolution maps of gas tracers (H$_1$, $^{12}$CO, $^{13}$CO) in position–position–radial...
PPV space has potential application both to a wider variety of data, and to higher-resolution 3D data. This higher-resolution PPV data may prove invaluable for mapping regions with overlapping clouds observed in dust and gas tracers. Given that a large range of velocities was required to determine the Mach number, resolving overlapping clouds may be difficult, unless the clouds are well-separated in velocity space.

4.4.2. Artifacts and Missing Data

Pattern noise is a common feature in imaging data sets. Long-term sensitivity drift in scanning detectors produced stripe artifacts in IRAS (Wheelock et al. 1994) and AKARI (Doi et al. 2015). A ground loop in the readout electronics of a charge-coupled device (CCD) can imprint nearby radio transmissions on the data. Even highly sophisticated instruments on the Hubble Space Telescope can experience electronics problems that leave subtle “herringbone” patterns in the data (Jansen et al. 2010). A desideratum of a classifier is that it be robust to such pattern noise, or at least able to give warning of its presence.

For testing purposes, we mimic pattern noise by adding a sinusoidal pattern with a random wavevector to the MHD slices. Specifically, we generate for each MHD slice (6,912 draws) a random amplitude (10%–30% of signal max), phase, period width (2–25 pixels), and angle relative to the \( x \) and \( y \) axis, and add it to the MHD slice prior to computing the WST coefficients (Figure 9). In some cases, large contiguous regions of an image may be lost due to satellite transits, diffraction spikes, or ghosts of bright stars. We model this by introducing a cutoff for eliminating pixels, based on the sinusoidal field used above (in this case with periods in the range of 43–256 pixels, and amplitudes of 10%–20%). The sinusoidal field was not added to the data, but the values in the image were set to zero for all indices where the sinusoidal field exceeded 10% (Figure 9). In both cases, we observe at most a slight reduction in precision (96% and 94%, respectively, compared to 97%) and increased variance for each class. However, the quantitative classification is evidently robust to these artifacts, and the qualitative structure of the clusters in the space of the first three LDA components was also unaffected.

4.4.3. Dependence on Point-spread Function

In observational images, the true signal is also convolved with a point-spread function (PSF). We model this as a Gaussian PSF, with a FWHM of 2.355\( \sigma \) (in pixels). Moreover, we do not assume that the PSF is rotationally symmetric, and allow for nonzero aspect ratios \( AR = \sigma_x/\sigma_y \). We investigated the following PSFs in detail: \( (\sigma, AR) = (2,1), (2,2), (4,1), (4,2), (4,3) \). The MHD classification by WST-LDA for samples from a single PSF was qualitatively unchanged, with a precision of \( \geq 94\% \). Further, when all the WST coefficients from the PSF-blurred images and original images were combined, LDA was able to classify the images on Mach numbers with a precision of 92%.

However, we must emphasize that the optimal LDA space for each individual PSF is different. For instance, the LDA space for (2,1), and the LDA space for the original images are not the same. In the interactive version of Figure 10, we demonstrate how the WST coefficients for (2,1) look in the

35 Similar results were obtained for \( y \)- and \( z \)-velocity field PPV cubes.
36 \( J = 7 \) for the 3D-WST provided too few samples for a reliable test-train split.
37 While the velocity axis and position axes do not even have the same units, the fractional extent of each axis captured by the wavelet represents the same fraction of the relevant range of spatial or velocity coordinates. The velocity axis of each cube is normalized to be \( -3 \) to \( 3\sigma_v \), where \( \sigma_v \) is the velocity dispersion.
38 One of the authors detected a Maui radio station with a Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) CCD.
LD images to lower new LDA space 39 We do not scale the WST coefficients prior to computing the LDA space, so that we can more easily study which coefficients are most affected by the PSF.

Representative slices were taken from an MHD simulation in the \((M_A, M_2) = (4, 2)\) class. Bottom: confusion matrix between MHD classes (labeled with [sonic, Alfvénic] Mach numbers) for the above noise models. Matrix entries are the number of test images with the corresponding true-predicted label pairing. The grayscale color indicates the percentage of test images with the corresponding true label in the test set. In all cases, colors correspond to true labels, as indicated on the vertical axis in the bottom left panel. In both cases, quantitative classification is evidently robust to these artifacts, because the precision is comparable to the noise-free value of 97%.

Figure 10. Effect of PSF radius and aspect ratio in a representative WST-LDA space for classification of MHD simulations. Each point comes from applying the WST to a 2D image from our MHD simulations, convolved with a PSF of the indicated radius (top) and aspect ratio (bottom), as given by the color bar. An interactive version of this figure is available at https://faun.rc.fas.harvard.edu/saydjari/RWST_2020/MHD_PSF.html, showing the PSF- and aspect ratio-dependence of the position of each point in the first three LDA coordinates. The MHD classification clusters for images containing PSF with \((2,1)\) are also shown in the LDA basis for the unblurred images, and vice versa.

LDA space for the original images, and vice versa. While these spaces are not identical, the helix that we observed in the native LDA space is still present, but shifted and compressed, or expanded. We observe that for small \(\sigma\) and AR, the old LDA space can be approximately continuously deformed into the new LDA space (not shown).

Ideally, we would find a classification space which is as insensitive as possible to the PSF. To approximate this, we take the original images, in addition to five blurred copies \((2,1), (2,2), (4,1), (4,2), (4,3)\), and define the WST-LDA space for the Mach number classification task on that entire set. To better understand how the position of an image’s WST coefficients in the LDA space depend on the \(\sigma\) and AR of the PSFs, we take four random slices from each time step of each MHD class, and compute the WST coefficients for a Cartesian product of \(\sigma\) and AR. We then plot the WST coefficients corresponding to each each image with a given PSF in the space defined above in Figure 10.\(^{59}\) We observe that increasing the PSF radius moves images to lower \(LD_2\), as compared to the original helical cluster. On this tail induced by the large PSF radii, the images are separated along \(LD_1\) by AR. As such, both \(\sigma\) and AR are encoded in the WST-LDA space. Similarly, the WST-LDA spaces for \(J_{\min} = 0\) and \(J_{\min} = 1\) are related, but not identical, suggesting that the scale of a feature relative to the minimum pixel scale will also deform the WST-LDA space. We foresee the creation of a classifier which is invariant to, or deforms in a known way in response to, changes in the PSF and pixel scale. This is likely to be the next step before WST-LDA can be applied to classify the relevant hydrodynamic parameters (or any other parameter of interest) for observational data. It is primarily for this reason that we hold off regarding any analysis of the SFD data set, in terms of the classification space derived from the MHD simulations.

4.5. Sensitivity to Magnetic Field Direction

In the MHD turbulence simulations, the original magnetic field direction in the simulation box is topologically constrained by the periodic boundary conditions of the simulation. This introduces a preferred direction in the simulation, because the characteristics and scaling of turbulence, both parallel and

\(^{59}\) We do not scale the WST coefficients to have mean zero and unit variance prior to computing the LDA space, so that we can more easily study which coefficients are most affected by the PSF.
The Astrophysical Journal, 910:122 (17pp), 2021 April 1

Saydari et al.

5. Conclusions

We have introduced WST-LDA as a new low-dimensional classifier for non-Gaussianity. In comparison to the 3PCF, we find the WST and RWST can more readily differentiate non-Gaussian from Gaussian fields, as well as classify MHD simulation parameters. RWST-LDA appears more robust than WST-LDA to sums along the line of sight, and achieves higher precision, but is susceptible to outliers in its current form. To the best of our knowledge, this is the first classification of MHD Mach numbers in which ~97% precision has been achieved without training a neural network. We show that the power of WST-LDA comes mostly from the number (and the relevance) of scales included in the scattering transform, rather than the number of coefficients. We similarly demonstrate 3D-WST-LDA and its application to PPV space, a case in which 3D correlations are essential to high-precision classification. Classification in PPV space may be improved by the development of anisotropic 3D wavelets preferring the velocity axis. The classification using WST-LDA was robust to additive sinusoidal noise, partial data loss, and convolution with various point-spread functions. However, the WST-LDA space in which the classification occurs was not invariant to the aforementioned effects, or to the scale of features relative to the minimum pixel scale. As such, modifying WST-LDA to be sensitive only to the MHD parameters and not to details of the resolution and pixel scale of measurements will be the next step prior to the classification of observational data.

6. Code and Data Availability

Data products associated with this paper are publicly available at doi:10.5281/zenodo.4057156 (41 GB). This includes the computed WST, RWST, and 3PCF coefficients, our selected subset of SFD, PPV density cubes from the MHD simulations, and our code for computing the 2D-3PCF. Reuse of our 3PCF code should cite Slepian & Eisenstein (2015), Slepian & Eisenstein (2016), Portillo et al. (2018), and S. Portillo et al. (2020, in preparation). JUPYTER notebooks containing code to reproduce all figures in the text, some minimal working examples, and how to run these computations on a cluster are also included. The full MHD simulation runs are available from the Catalog for Astrophysical Turbulence Simulations (CATS) at http://www.mhdturbulence.com (Burk hart et al. 2020).

A.S. gratefully acknowledges the support of a National Science Foundation Graduate Research Fellowship (DGE-1745303). D.F. acknowledges the support of NSF grant AST-1614941, Exploring the Galaxy: 3-Dimensional Structure and Stellar Streams. B.B. is grateful for support from the Simons Foundation. S.K.N.P. acknowledges support from the DIRAC Institute in the Department of Astronomy at the University of Washington. The DIRAC Institute is supported through generous gifts from the Charles and Lisa Simonyi Fund for Arts and Sciences, and the Washington Research Foundation. The authors wish to thank Catherine Zucker and Michael Foley for helpful discussions. A.S. wishes to thank Sophia Sánchez-Maes for helpful discussion and much support. Computations in this paper were run on the FASRC Cannon cluster, supported by the FAS Division of Science Research Computing Group at Harvard University.

Facilities: IRAS, COBE.

Software: ASTROPY (Astropy Collaboration et al. 2013), DUSTMAPS (Green 2018), GLUE (Beaumont et al. 2015; Robitaille et al. 2017), H5PY (Collette 2013), IPYTHON (Perez & Granger 2007), KYT MATIO (Andreux et al. 2018), MATHPL TLIB (Hunter 2007), NUMPY (van der Walt et al. 2011), PYTORCH (Paszke et al. 2019), SCIPY (Virtanen et al. 2020), SCIKIT-LEARN (Pedregosa et al. 2012), SCIKIT-IMAGE (van der Walt et al. 2014), YT (Turk et al. 2011).

Appendix A
Computational Costs

In order to provide an idea of the relative computational cost for each of the non-Gaussian descriptors used, we present the number of coefficients and computational time per coefficient for the 2D-WST, 3D-WST, and 3PCF (Table 1). The computation of the RWST coefficients from the WST coefficients simply involves the usual cost of least-squares optimization (and is relatively fast compared to the WST computation). Note that the definition of $J$ for each method is

---

Figure 11. Confusion matrix between slices taken from the $x$, $y$, and $z$ axes for the sub-Alfvénic MHD classes ($M_A = 0.7$) (left). Matrix entries are the number of test images with the corresponding true-predicted label pairing. The grayscale color indicates the percentage of test images with the corresponding true label in the test set. In all cases, colors correspond to true labels, as indicated on the vertical axis in the left panel. Test and train images are plotted in the first two components of LDA space (right). Circular markers represent test images, and cross markers represent test images that were incorrectly assigned. WST-LDA can clearly distinguish the magnetic field axis, in the absence of magnetic field wandering, based solely on the density field.
slightly different. The 2D-WST, computed up to a scale, J, includes \( j = 0, \ldots, J - 1 \), while the 3D-WST and 3PCF include \( j = 0, \ldots, J \). For the 2D and 3D-WST we used \( L = 8 \), and for the 3PCF we used \( L = 9 \). For \( J \geq 7 \), the memory load for 3D-WST exceeds 100 MiBs, and needs to be considered. For 3D-WST and 3PCF, the coefficients were calculated on an image of uniform grid size \( 2^J \). For 2D-WST only, the computational time reflects the calculation of the coefficients on \( 2^{8-J} \times 2^{8-J} \) subsets of the full \( 256^2 \) images of size \( 2^J \times 2^J \). These experiments were executed on the FASRC Cannon cluster at Harvard University, on a computer node with water-cooled Intel 24-core Platinum 8268 Cascade Lake CPUs, with 192 GB RAM running 64 bit CentOS 7. Cascade Lake cores have dual AVX-512 fused multiply add (FMA) units. Code scaling is reported on a single core in a PYTHON environment, specified by the YAML file in Section 6.

### Appendix B

**Effect of Mean Zero on 3PCF**

Defining a density fluctuation field with mean zero as \( \delta(x) \equiv \rho(x) / \bar{\rho} - 1 \), with \( \bar{\rho} \) being the average density, the 3PCF is

\[
\langle \delta(x) \delta(x + r_1) \delta(x + r_2) \rangle
\]

where the angle brackets denote averaging over translations and rotations. Substituting in the definition of \( \delta \), and defining the density-2PCF, as \( \zeta_\rho(s) \equiv \langle \rho(x) \rho(x + s) \rangle \), we find

\[
\langle \delta(x) \delta(x + r_1) \delta(x + r_2) \rangle = \zeta_\rho(r_1, r_2, \tilde{r}_1 \cdot \tilde{r}_2) - \left[ \zeta_\rho(r_1) + \zeta_\rho(r_2) + \zeta_\rho(r_3) \right]
\]

(B2)

where \( \zeta_\rho \) is the 3PCF of the density field, and we have used the identity that \( |r_2 - r_1| = r_3 \). Therefore, as claimed, taking a zero-mean field results in subtracting out the 2PCF contribution to the 3PCF.

### Appendix C

**RWST Angle Dependence**

While we show that the RWST transformation of the WST coefficients can improve the performance of RWST-LDA, as compared to WST-LDA, the origin of this improvement is still not fully understood. To illustrate this, we compute the WST coefficients \( J = 2, L = 256, \) and \( M = 2 \) for two images measuring \( 16 \times 16 \) pixels: a subset of an SFD image, and an image with random pixels. The SFD image at this pixel scale provides only an example of a real, slowly varying field, in contrast with the random image, in which neighboring pixels are totally uncorrelated. The angle dependence of the second-order WST coefficients for an internal \( 4 \times 4 \) subset of pixels is shown as a function of \( \theta_1 \) and \( \theta_2 \) in Figure 12. A horizontal line cut near the top of the WST coefficient map is also shown, in order to illustrate the oscillation envelope fit by the cosine in the RWST model. While detailed aspects of the WST coefficient maps are different, both the continuously varying field and the random image display clear oscillations. In this case, we attribute the oscillations to the fact that, for all scales, the wavelets are not well sampled in the Fourier domain. It is possible that much of the relevant information enters into the amplitude of the WST coefficients, but more work is needed to understand the optimal parameterization of these oscillations, and to better understand their interpretation.

### Table 1

| \( J \) | Time/Coeff (core-milliseconds) | Number of Coefficients |
|--------|-------------------------------|-------------------------|
|        | 2D-WST | 3D-WST | 3PCF | 2D-WST | 3D-WST | 3PCF |
| 8      | 3.4    | 2600   | 1.2  | 1857   | 1215   | 360  |
| 7      | 1.4    | 140    | 0.95 | 1401   | 972    | 280  |
| 6      | 0.75   | 12     | 0.73 | 1009   | 756    | 210  |
| 5      | 0.47   | 1.5    | 0.56 | 681    | 567    | 150  |
| 4      | 0.30   | 0.35   | 0.39 | 417    | 405    | 100  |
| 3      | 0.23   | 0.12   | 0.27 | 217    | 270    | 60   |
Appendix D

Classification of Cumulative Sums

In order to mimic 2D dust maps that measure dust integrated along the lines of sight, we sampled the cumulative sum of our MHD density cubes along each axis, and repeated the classification based on sonic and Alfvénic Mach numbers, using LDA combined with the WST, RWST, and 3PCF (Figure 13). WST-LDA achieves a precision of only 54%, and shows a large broadening in the variance for each class, which leads to significant crosstalk between multiple adjacent classes. However, the confusion matrix is still predominately diagonal, with most of the density within two classes from the correct label. As such, sampling the cumulative sum can be viewed as causing an increased uncertainty in the Mach number assignments. RWST-LDA shows a much smaller decrease in precision (to 81%) and much smaller class variance on the cumulative sum images, as compared to WST-LDA. This further supports the claim that the nonlinear transformation of the WST coefficients made by the RWST is not only useful in terms of dimension reduction, but also makes the classifier more robust against variations that are uninformative with regard to distinguishing the MHD classes. While the 3PCF-LDA precision drops to approximately to 51%, near that for WST-LDA, the classification is much worse, in that many images overlap in the LDA space, and are all predicted to be in the same class. This is a function of the nonlinear spacing of classes in the 3PCF-LDA space. The comparatively poor performance of 3PCF also suggests that the specific combination of up to fourth-order moments included in the WST is well chosen, since the 3PCF alone is so strongly modified.
Figure 13. Confusion matrix between MHD classes, labeled with [sonic, Alfvénic] Mach numbers when applying WST-LDA (top left), RWST-LDA (middle left), and 3PCF-LDA (bottom left) for images drawn from cumulative sums along the MHD simulation density cubes. Matrix entries are the number of test images with the corresponding true-predicted label pairing. The grayscale color indicates the true positive percentage, relative to the number of test images with the corresponding true label. Test images are shown in the first two components of LDA space for the WST (top right), RWST (middle right), and 3PCF (bottom right). Circle markers represent images that were correctly assigned, and cross markers represent those that were incorrectly assigned. In all cases, colors correspond to true labels. An interactive version of this figure is available at https://faun.rc.fas.harvard.edu/saydjari/RWST_2020/MHD_LDA_cumsum.html, showing the images in 3D LDA space, and allowing isolation of the train, correctly labeled test images, and incorrectly labeled test images.
Gallegos-Garcia, M., Burkhart, B., Rosen, A. L., Naiman, J. P., & Beaumont, C., Goodman, A., & Green

The Astrophysical Journal, 910:122 (17pp), 2021 April 1

Saydjari et al.

Beaumont, C., Goodman, A., & Greenfield, P. 2015, in ASP Conf. Ser. 495, Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV), ed. A. R. Taylor & E. Rosolowsky (San Francisco, CA: ASP), 101

Bialy, S., & Burkhart, B. 2020, ApJL, 894, L2

Boggess, N. W., Mather, J. C., Weiss, R., et al. 1992, ApJ, 397, 420

Bruna, J., & Mallat, S. 2012, arXiv:1203.1513

Bruna, J., & Mallat, S. 2018, arXiv:1801.02013

Bruna, J., Mallat, S., Bacity, E., & Muzy, J.-F. 2015. AnSta, 43, 323

Burkhart, B., Appel, S. M., Bialy, S., et al. 2020, ApJ, 905, 14

Burkhart, B., Falceta-Gonçalves, D., Kowal, G., & Lazarian, A. 2009. ApJ, 693, 250

Burkhart, B., & Lazarian, A. 2016. ApJ, 827, 26

Burkhart, B., Lee, M.-Y., Murray, C. E., & Stanimirović, P. 2020, ApJ, 905, 14

Chen, G., & Szapudi, I. 2005, ApJ, 635, 743

Cho, J., Lazarian, A., & Vishniac, E. T. 2002, ApJ, 564, 291

Clark, S. E., & Hensley, B. S. 2019, ApJ, 887, 136

Collette, A. 2013, Python and HDF5 (Seabastopol, CA: O'Reilly & Associates)

Corasaniti, P. S. 2006, MNRAS, 372, 191

Dame, T. M., Hartmann, D., & Thaddeus, P. 2001, ApJ, 547, 792

Doi, Y., Takita, S., Ootsubo, T., et al. 2015, PASJ, 67, 50

Eckner, M., Exarchakis, G., Hirn, M., Mallat, S., & Thiry, L. 2018, JChPh, 148, 241732

Elmegreen, B. G., & Scalo, J. 2004, ARA&A, 42, 211

Gallegos-Garcia, M., Burkhart, B., Rosen, A. L., Naiman, J. P., & Ramirez-Ruiz, E. 2020, ApJL, 899, L30

Gill, A. G., & Henriksen, R. N. 1990, ApJL, 365, L27

Goldreich, P., & Sridhar, S. 1995, ApJ, 438, 763

Goodman, A. A., Rosolowsky, E. W., Borkin, M. A., et al. 2009, Nat, 457, 63

Green, G. M. 2018, JOSS, 3, 695

Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, ApJ, 887, 93

Groth, E. J., & Peebles, P. J. E. 1977, ApJ, 217, 385

Heiles, C., & Troland, T. H. 2003, ApJS, 145, 329

Hirn, M., Mallat, S., & Poilvert, N. 2017, Multiscale Model. Simul., 15, 827

Hunter, J. D. 2007, CSE, 9, 90

Jansen, R. A., Windhorst, R., Kim, H., et al. 2010, in 2010 Space Telescope Science Institute Calibration Workshop: Hubble after SM4. Preparing JWST, ed. S. Deustua & C. Oliveira (Baltimore, MD: STScI), 50

Jolliiffe, I. T., & Cadima, J. 2016. RSPTA, 374, 20150202

Kainulainen, J., & Federrath, C. 2017, A&A, 608, L3

Kainulainen, J., & Tan, J. C. 2013, A&A, 549, A53

Kandel, D., Lazarian, A., & Pogosyan, D. 2017, MNRAS, 472, L10

Khalil, A., Joncas, G., Nekka, F., Kestener, P., & Armeodo, A. 2006, ApJS, 165, 512

Kowal, G., & Lazarian, A. 2010, ApJ, 720, 742

Kowal, G., Lazarian, A., & Beresnyak, A. 2007, ApJ, 658, 423

Kritsuk, A. G., Pfaumer, R., & Ustyugov, S. D. 2018, PhRvL, 121, 021104

Krumholz, M. R., Burkhart, B., Forbes, J. C., & Crocker, R. M. 2018, MNRAS, 477, 2716

Lazarian, A., Pogosyan, D., Vázquez-Semadeni, E., & Pichardo, B. 2001, ApJL, 555, 130

Lazarian, A., & Vishniac, E. T. 1999, ApJ, 517, 700

LeCun, Y., & Cortes, C. 2010, The MNIST Database, http://yann.lecun.com/exdb/mnist/

Leike, R. H., Glazte, M., & Enßlin, T. A. 2020, A&A, 639, A138

Low, F. J., Beintema, D. A., Gautier, T. N., et al. 1984, ApJL, 278, L19

Mallat, S. 2012, Commun. Pure Appl. Math., 65, 1331

Mather, J. C. 1982, OptEn, 21, 769

Molinari, S., Swinyard, B., Bally, J., et al. 2010, PASP, 122, 314

Neugebauer, G., Habing, H. J., van Duinen, R., et al. 1984, ApJL, 278, L1

Padoan, P., Pan, L., Haugbolle, T., & Nordlund, A. 2016, ApJL, 822, 11

Paszke, A., Gross, S., Massa, F., et al. 2019, arXiv:1912.01703

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, CSE, 13, 22

Perez, F., & Granger, B. E. 2007, CSE, 9, 21

Pineda, J. E., Caselli, P., & Goodman, A. A. 2008, ApJ, 679, 481

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2017, A&A, 607, A95

Portillo, S. K. N., Sleipan, Z., Burkhart, B., Kahraman, S., & Finkbeiner, D. P. 2017, ApJL, 862, 119

Regalado-Saint Blancard, B., Levrier, F., Allys, E., Bellomi, E., & Boulanger, F. 2020, A&A, 642, A217

Ridge, N. A., Di Francesco, J., Kirk, H., et al. 2006, AJ, 131, 2921

Robitaille, T. F., Joncas, G., & Miville-Deschênes, M. A. 2014, MNRAS, 440, 2726

Robitaille, T., Beaumont, C., Qian, P., Borkin, M., & Goodman, A. 2017, glueviz v0.13.1: multidimensional data exploration, Zenodo, doi:10.5281/zenodo.1237692

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Slepian, Z., & Eisenstein, D. J. 2015, MNRAS, 454, 4142

Slepian, Z., & Eisenstein, D. J. 2016, MNRAS, 455, L31

Slepian, Z., Eisenstein, D. J., Beutler, F. et al. 2017a, MNRAS, 468, 1070

Slepian, Z., Eisenstein, D. J., Brownstein, J. R., et al. 2017b, MNRAS, 469, 1738

Taylor, A. R., Gibson, S. J., Percival, W. J. et al. 2003, AJ, 125, 3145

Turk, M. J., Smith, B. D., Oishi, J. S., et al. 2011, ApJS, 192, 9

van der Walt, S., Schönberger, J. L., & Peres, E. 2014, PeerJ, 2, e453

Virtanen, P., Gommers, R., et al. 2020, Nature, 570, 23

Wheelock, S. L., Gautier, T. N., Chillemi, J., et al. 1994, IRAS Sky Survey Atlas: Explanatory Supplement, NASA STI/Recon Technical Report

Zheng, Z. 2004, ApJ, 614, 527