

How to evaluate the sufficiency and complementarity of summary statistics for cosmic fields: an information-theoretic perspective

CE SUI ¹, YI MAO ¹, XIAOSHENG ZHAO ², TAO JING ¹ AND BENJAMIN D. WANDELT ^{2,3}

¹*Department of Astronomy, Tsinghua University, Beijing 100084, China*

²*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA*

³*Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218, USA*

Submitted to ApJL

ABSTRACT

The advent of increasingly advanced surveys and cosmic tracers has motivated the development of new inference techniques and novel approaches to extracting information from cosmic fields. A central challenge in this endeavor is to quantify the information content carried by these summary statistics in cosmic fields. In particular, how should we assess which statistics are more informative than others and assess the exact degree of complementarity of the information from each statistic? Here, we introduce mutual information (MI) that provides, from an information-theoretic perspective, a natural framework for assessing the sufficiency and complementarity of summary statistics in cosmological data. We demonstrate how MI can be applied to typical inference tasks to make information-theoretic evaluations, using two representative examples: the cosmic microwave background map, from which the power spectrum extracts almost all information as is expected for a Gaussian random field, and the 21 cm brightness temperature map, from which the scattering transform extracts the most non-Gaussian information but is complementary to power spectrum and bispectrum. Our results suggest that MI offers a robust theoretical foundation for evaluating and improving summaries, thereby enabling a deeper understanding of cosmic fields from an information-theoretic perspective.

Keywords: Cosmology (343) — Bayesian statistics (1900) — Astrostatistics (1882)

1. INTRODUCTION

Tracing the evolution of our Universe through different cosmological observations has been a central goal of modern cosmology. Among these endeavors, the cosmic microwave background (CMB) and the large-scale structure (LSS) stand out as the most precise observational probes. On large scales, both fields are well described as Gaussian random fields (GRFs), which means that their information content is fully captured by the two-point statistics, i.e. the power spectrum (PS) in Fourier space, which is therefore the focus of standard cosmological analysis (e.g. Planck Collaboration et al. 2020; L. Anderson et al. 2012; A. G. Adame et al. 2025).

However, the anisotropies in the CMB and the non-linear features of LSS, as well as new observational windows such as small-scale galaxy clustering, weak lensing, and 21 cm intensity mapping (e.g., DES Collab-

oration et al. 2025; Euclid Collaboration et al. 2024; Square Kilometre Array Cosmology Science Working Group et al. 2020), contain additional information beyond the PS. To extract the non-Gaussian information from these datasets, significant efforts have been devoted to construct novel, informative summary statistics. Information extraction from these summaries is facilitated with the emergence of simulation-based inference (M. Ho et al. 2024; C. Hahn et al. 2023; J. Alsing et al. 2019), because these statistics can be forward-modeled by advances in theories and increasingly realistic high-resolution simulations (F. Villaescusa-Navarro et al. 2020; J. Schaye et al. 2023; N. A. Maksimova et al. 2021; R. Pakmor et al. 2023; A. Pillepich et al. 2018; D. Anbajagane et al. 2024). Examples of summary statistics include the use of cosmic voids as complementary probes to galaxy power spectra (R. Malandrino et al. 2025; S. Contarini et al. 2023; N. Hamaus et al. 2020), bispectrum (BS) analyses across multiple cosmic fields (M. M. Ivanov et al. 2023; I. Kayo et al. 2013; C. A. Watkinson et al. 2022; H. Shimabukuro et al. 2016), and

the application of Minkowski functionals to weak lensing and 21 cm maps (L. Gleser et al. 2006; A. Kapahtia et al. 2021; K. Diao et al. 2024; J. M. Kratochvil et al. 2012). More recently, techniques such as the wavelet scattering transform (ST; S. Cheng et al. 2020) have been employed to extract multiscale non-Gaussian information from various fields (S. Cheng et al. 2020; X. Zhao et al. 2024; B. Régalo-Saint Blancard et al. 2024). In addition to such physically motivated summaries, machine learning–based approaches have been developed to construct “optimal” statistics, typically by maximizing an information criterion or improving inference performance (T. L. Makinen et al. 2024; J. Alsing et al. 2018; X. Zhao et al. 2022). Beyond the inference from summary statistics, field-level inference (H. Simon-Onfroy et al. 2025; N.-M. Nguyen et al. 2024; N. Porqueres et al. 2023) is also developed along the line to fully exploit non-Gaussian information in these datasets.

In this era of thriving information, cosmologists seek to extract the *optimal* (sometimes aka *maximal* or *full*) information from different statistics, as well as providing complementary information by combining them. A key question arises: how should we assess quantitatively which statistics are more informative than others and assess the exact degree of complementarity of the information from each statistic?

Assessing or learning summaries ultimately relies on quantifying their information content with respect to the underlying model parameters. In practice, this has often been approached through Fisher analysis (D. Prelogović & A. Mesinger 2024; A. Akhmetzhanova et al. 2024), posterior comparisons (Beyond-2pt Collaboration et al. 2024; M. Gatti et al. 2024) or regression performance (N. Gillet et al. 2019; D. Prelogović et al. 2022). Among these, Fisher information (FI; M. Tegmark et al. 1997) is a widely used tool from estimation theory. However, its application in practice often relies on Gaussian likelihood assumptions and finite-difference score estimation, highlighting the need for new methods that enable more accurate estimation (C. Sui et al. 2025; J.-E. Campagne et al. 2023). In addition, FI only reflects local sensitivity around fiducial parameters. Posterior comparisons directly compare the credible region in parameter space for an inference around a fiducial test set, so they are also limited to local sensitivity. Regression-based metrics, while providing a global view, estimate the performance of point estimates, not the posteriors.

In this paper, we propose the use of mutual information (MI), a fundamental concept in information theory and modern statistical inference (Y. Polyanskiy & Y. Wu 2025; Y. Wu 2017), to assess the sufficiency and complementarity of summary statistics in cosmological data.

MI is defined to quantify the reduction in uncertainty of one random variable given knowledge of another, making it a natural metric for evaluating information content. In machine learning, MI has served as the theoretical basis for numerous representation-learning objectives (R. Devon Hjelm et al. 2018; A. van den Oord et al. 2018; R. Shwartz Ziv & Y. LeCun 2024; M. Federici et al. 2020). In cosmology, it has also been employed to design new summaries of cosmic fields (D. Sharma et al. 2024; A. Akhmetzhanova et al. 2024; T. L. Makinen et al. 2024). In this paper, however, we take a different perspective — here, MI is used not as a loss function, but as a diagnostic tool to assess and decompose the information content of cosmological fields and their summaries. The mathematical properties of MI enable a principled evaluation of summary statistics, providing insights into their effectiveness, potential improvements, and the extent to which they capture the available information.

Note that MI, FI and regression errors are related through several interesting information-theoretic inequalities (see a brief review of these relations in Appendix B). However, a key distinction between them is that while FI depends on the specific fiducial parameter, MI does *not*, because it provides a global measure of information. We emphasize that both MI and FI are useful, complementary statistical tools — MI as a metric for information assessment, and FI as a tool for direct forecast of parameter sensitivities.

The remainder of this paper is organized as follows. Section 2 introduces the core concepts and methodology. We apply MI to two representative cases in Section 3, including CMB-like GRFs and 21 cm non-Gaussian maps. We make concluding remarks in Section 4. Some technical details are left to Appendix A (on MI and Bayesian sufficiency), Appendix B (on the relationships between MI, FI, and estimation mean squared error), and Appendix C (on the definition of summary statistics). A part of preliminary results were previously summarized by us in a conference proceeding paper (C. Sui et al. 2023).

2. METHODOLOGY

MI, a central concept in information theory and statistical inference, quantifies how much information one random variable contains about another. The MI $I(x; y)$ about two random variables x and y is defined as (Y. Polyanskiy & Y. Wu 2025)

$$I(x; y) = D_{\text{KL}}(p(x, y) \parallel p(x)p(y)) \equiv \mathbb{E}_{p(x, y)} \left[\log \frac{p(x, y)}{p(x)p(y)} \right], \quad (1)$$

where $p(x, y)$, $p(x)$, $p(y)$ denote their joint and marginal distributions, respectively. Here, the Kullback–Leibler

(KL) divergence $D_{\text{KL}}(P \parallel Q)$ is defined as the expectation value of $\log(P(x, y)/Q(x, y))$ over the joint distribution $P(x, y)$. Note that MI is symmetric, i.e. $I(x; y) = I(y; x)$.

Intuitively, MI measures the discrepancy between the joint distribution and the product of the marginals using the KL divergence. In other words, MI capture how much uncertainty in a variable is reduced after the other is observed.

2.1. MI as a Metric of Information Sufficiency

Statistical inference is all about estimating the parameters θ of a physical model given the observed data x . Within the Bayesian framework, this corresponds to inferring the posterior distribution $p(\theta|x)$. However, raw data x are often high-dimensional and may contain a large amount of irrelevant or redundant information. To mitigate this, it is common to employ summary statistics, i.e., lower-dimensional representations $s = s(x)$ designed to retain the informative features of the data with respect to the parameters of interest.

If summary statistics s are sufficient, they must preserve all the information in x that is relevant to θ . The classical notion of sufficient statistics is defined by the condition,

$$p(\theta|x) = p(\theta|s(x)), \quad (2)$$

which means that sufficient statistics $s(x)$ preserve all information about θ contained in the raw observed data x . In practice, while exact sufficiency is often unattainable, statistics that closely approximate sufficiency generally yield more accurate and reliable inference.

MI provides a natural characterization of sufficiency, because the sufficiency condition (Eq. 2) is equivalent (see this detail in Appendix A) to

$$I(\theta; x) = I(\theta; s(x)). \quad (3)$$

Since the complete information from raw data x is $I(\theta; x)$, the informativeness of summary statistics $s(x)$ can be evaluated using their MI $I(\theta; s(x))$ — a higher value of MI means closer to satisfaction of the sufficiency condition and therefore the summary statistics extract more information from the data. This highlights MI as a powerful metric of information sufficiency, offering, from a solid and theoretical foundation, a practical criterion for guiding the design and selection of summary statistics in statistical inference.

2.2. Conditional MI as a Metric of Complementarity

A particularly useful feature of MI is that it naturally allows us to decompose the information content of a field into contributions from different summaries. This provides insight into how summaries can be combined to

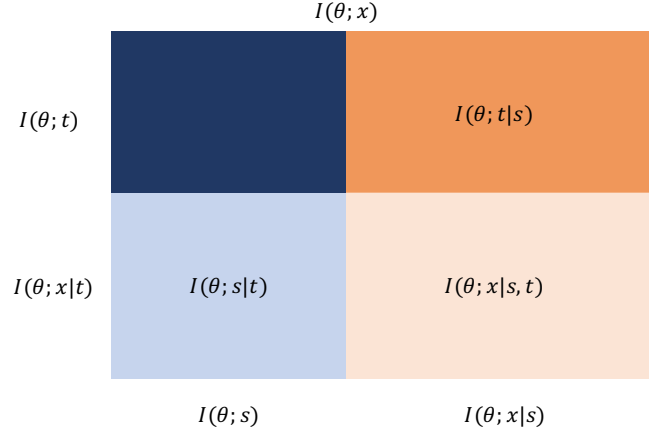


Figure 1. Decomposition of information content in a raw data x . $I(\theta; x)$ denotes the full field-level information. The left column (blue) $I(\theta; s)$ represents the information extracted by the summary s , and the right column (orange) $I(\theta; x | s)$ captures the residual information beyond s . Similarly, the upper row (dark) $I(\theta; t)$ shows the information extracted by another summary t , and the lower row (light) $I(\theta; x | t)$ captures the residual information beyond t . The upper right (dark orange) panel $I(\theta; t | s)$ quantifies the complementary information that t contributes beyond s , and the lower left (light blue) panel $I(\theta; s | t)$ quantifies the complementary information that s contributes beyond t . The upper left (dark blue) panel quantifies the common information shared by both s and t , and the lower right (light orange) panel $I(\theta; x | s, t)$ quantifies the residual information in the raw data x beyond both s and t .

maximize the captured information. This decomposition typically involves conditional MI, defined as

$$I(x; y | z) = \mathbb{E}_{p(x, y, z)} \left[\log \frac{p(x | y, z)}{p(x | z)} \right], \quad (4)$$

where x , y and z are three random variables.

The total information in the raw data x can be decomposed using the chain rule of MI,

$$I(\theta; x) = I(\theta; s) + I(\theta; x | s), \quad (5)$$

where the first term in the rhs represents the information captured by the summary statistics s , and the second term captures the residual information in the raw data x beyond s .

If two summary statistics s and t are considered, this decomposition can be further refined as

$$I(\theta; x) = I(\theta; s) + I(\theta; t | s) + I(\theta; x | s, t). \quad (6)$$

We illustrate this decomposition schematically in Figure 1. Here, $I(\theta; t | s)$, corresponds to the complementary information that t contributes beyond s . Conversely, $I(\theta; s | t)$ corresponds to the complementary information that s contributes beyond t .

These two conditional MI can be used as the metric for quantifying the degree of information complementarity between two summary statistics, i.e. providing a direct measure of the additional information contributed by one summary beyond another base summary. Rather than replacing an existing summary, it may be more effective to identify complementary summaries that capture additional information — an approach explored in T. L. Makinen et al. (2024).

2.3. Numerical Estimation of MI

Estimating MI in practice is challenging because it requires approximating high-dimensional probability densities. Classical approaches, e.g. the k -nearest neighbour estimator (A. Kraskov et al. 2004), approximate densities locally but do not scale well with dimensionality. The advent of neural networks has enabled flexible models capable of representing complex distributions or their ratios, giving rise to a range of MI estimators based on density modeling or density-ratio estimation (M. Ishmael Belghazi et al. 2018; B. Poole et al. 2019; J. Song & S. Ermon 2019). Many of these can be formulated as learning objectives that find the extrema of MI (R. Devon Hjelm et al. 2018; A. van den Oord et al. 2018; M. Federici et al. 2020; P. Cheng et al. 2020).

In cosmological inference problems, the joint distribution of physical parameters and data summaries, $p(\theta, s)$, is often intractable. Consequently, direct computation of MI is infeasible. Instead, one can employ flexible generative models to approximate the true distributions and derive variational bounds on MI. If we introduce a variational distribution $q(\theta|s)$ to approximate the intractable conditional $p(\theta|s)$, then it is straightforward to show that

$$I(\theta; s) = \mathbb{E}_{p(\theta, s)} \left[\log \frac{q(\theta|s)}{p(\theta)} \right] + \mathbb{E}_{p(s)} \left[D_{\text{KL}}(p(\theta|s) \parallel q(\theta|s)) \right]. \quad (7)$$

If $q(\theta|s)$ is sufficiently accurate such that $D_{\text{KL}}(p(\theta|s) \parallel q(\theta|s)) \approx 0$, then MI can be approximated by

$$I(\theta; s) \approx \mathbb{E}_{p(\theta, s)} [\log q(\theta|s)] + h(\theta). \quad (8)$$

The first term in the rhs of Eq. (8) can be estimated by training a flexible generative model on large sets of parameter–statistic pairs to yield a variational distribution $q(\theta|s)$ that closely approximates the true posterior. In this paper, we employ a masked autoregressive flow (MAF; G. Papamakarios et al. 2017) as a variational distribution. The second term in the rhs of Eq. (8) $h(\theta) = -\mathbb{E}_{p(\theta)} [\log p(\theta)]$ is the differential entropy of θ . This corresponds to the Barber–Agakov lower bound (D. Barber & F. Agakov 2004; B. Poole et al. 2019). In cos-

mological applications, $h(\theta)$ is typically tractable since the prior on θ is known.

The conditional MI can be estimated using chain rule,

$$I(\theta; t | s) = I(\theta; s, t) - I(\theta; s). \quad (9)$$

The two terms on the rhs of Eq. (9) can be estimated by training two generative models, respectively.

3. APPLICATIONS

3.1. Summary Statistics

Our objective is to assess how effectively different summaries capture parameter-relevant information about underlying physical parameters by estimating their MI. This includes two aspects — sufficiency (the *total* information that summary statistics contain about the parameters), and complementarity (the *additional* information that a summary provides beyond another baseline summary).

For each application case, we compute the chosen summary statistics from the field data, fit a generative model to approximate the conditional density, and then use the fitted model to estimate MI and conditional MI between summaries and parameters.

We consider two cases of application. The first case is a CMB-like GRFs. In principle, the PS contains all information in GRFs due to the Wick Theorem, thereby serving as a sanity check for MI as the metrics of information. The second application of MI is the 21 cm intensity mapping from the epoch of reionization (EoR), a non-Gaussian field for which there have been tremendous efforts to propose novel summary statistics in attempt to extract more information from the non-Gaussian 21 cm map. Here, we will use the MI to answer which statistics are more informative and investigate the complementarity between different summaries.

Among many summaries, we select three representative statistics as examples in this paper, namely, the PS, the BS, and the wavelet ST. The PS is sufficient for GRFs and remains the most widely used statistic in cosmological analysis. The BS is the higher-order statistics in extension to the PS, which can capture the non-Gaussian features and is therefore widely employed in cosmological large-scale structure. Here, we use the reduced BS $Q(k_1, k_2, k_3)$ to separate out the Gaussian information as much as possible, but without loss of clarity, we still refer to reduced BS as BS throughout this paper. The wavelet ST is a more recent statistic inspired by signal processing and designed to capture hierarchical, multiscale non-Gaussian features. We recap their definitions in Appendix C.

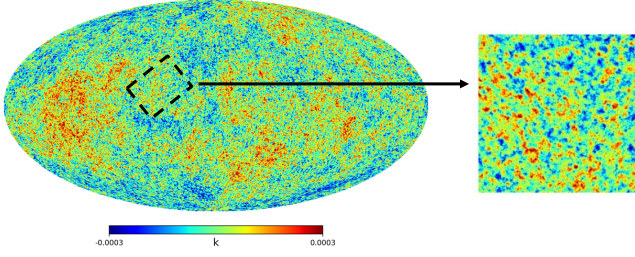


Figure 2. Left: a CMB-like map generated from a given PS. Right: a fixed square patch extracted from the map, used for computing summary statistics.

3.2. Application I: CMB-like Gaussian Field

The information content of a GRF can, in principle, be fully characterized by its PS, making it an ideal test case for validating our framework. To construct an inference problem on GRFs, we adopt standard tools developed for CMB analysis. Specifically, we generate CMB power spectra with varying cosmological parameters using `CAMB`⁴ (A. Lewis & A. Challinor 2011), and produce corresponding CMB-like maps at $n_{\text{side}} = 1024$ with `Healpy`⁵ (K. M. Górski et al. 2005; A. Zonca et al. 2019). By construction, these maps are pure GRFs. We extract a fixed square patch from each map, as illustrated in Figure 2, to calculate the statistics because some of these statistics are not defined directly on the sphere.

We vary two cosmological parameters, the dark matter density Ω_c and the Hubble parameter h . A 100×100 grid of parameter combinations is generated, with $\Omega_c \in [0.21, 0.30]$ and $h \in [0.64, 0.76]$, while all other cosmological parameters are fixed to the `CAMB` defaults. For each parameter set, we simulate the corresponding map and compute a suite of summary statistics.

To compute the PS and BS, we used the 2D polyspectra calculator implemented in the `scattering`⁶ package (S. Cheng et al. 2020). For the PS, we selected 15 logarithmically spaced k -bins within the range $[k_{\text{min}}, k_{\text{max}}]$, and used the same k -bins for the BS calculation. For the ST, we employed the Morlet wavelets implementation in `Kymatio`⁷ (M. Andreux et al. 2018), with parameters $J = 9$ and $L = 4$, chosen to maximize scale coverage.

Figure 3 shows the information content of different summary statistics for CMB-like GRFs. The MI of the BS (-0.033), and the conditional MI of the BS (0.068)

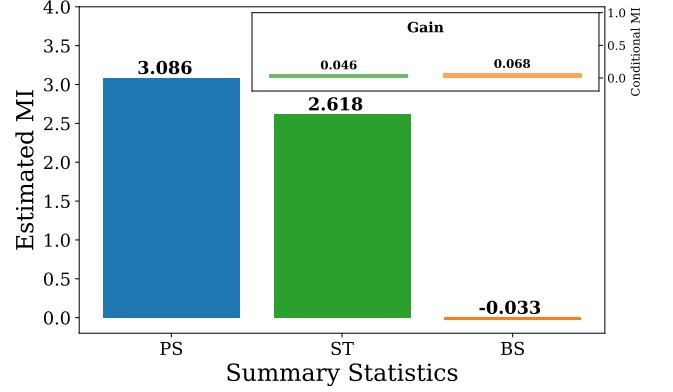


Figure 3. The MI of the PS (blue), ST (green) and BS (orange) for CMB-like GRFs. In the inset, the conditional MI of the ST (BS) given PS is shown, respectively, which measures the complementary information that the ST (BS) contributes beyond the PS.

and ST (0.046) given the PS are negligible⁸. This means that: (1) the BS contains almost no information from CMB-like GRFs; (2) while the MI of ST (2.618) is close to that of the PS (3.086), the ST contributes almost no additional information beyond the PS. In other words, the ST also captures the Gaussian information from the CMB-like GRFs. This test confirms that the MI and conditional MI correctly measure the information content in GRFs, which is consistent with the theoretical expectation that the PS contains all information in GRFs.

3.3. Application II: non-Gaussian 21 cm Maps

The 21 cm signal is non-Gaussian due to reionization patchiness, an interesting application case for studying information sufficiency and complementarity.

The 21 cm brightness temperature from the EoR relative to the CMB in position $\mathbf{x}$ is given by (S. R. Furlanetto et al. 2006):

$$T_{21}(\mathbf{x}, z) = \tilde{T}_{21}(z) x_{\text{HI}}(\mathbf{x}) [1 + \delta(\mathbf{x})] \left(1 - \frac{T_{\text{CMB}}}{T_s} \right), \quad (10)$$

where $\tilde{T}_{21}(z) = 27 \sqrt{[(1+z)/10](0.15/\Omega_m h^2)(\Omega_b h^2/0.023)}$ mK, $x_{\text{HI}}(\mathbf{x})$ is the neutral hydrogen fraction, and $\delta(\mathbf{x})$ is the matter overdensity. To simulate the cosmic 21 cm signal from the EoR, we generate coeval boxes using the publicly available `21cmFAST`⁹ code (A. Mesinger & S. Furlanetto 2007; A. Mesinger et al. 2011), a semi-numerical simulation code for simulating the EoR 21 cm signal. Each simulation is run in a cubic volume of 128

⁴ <https://github.com/cmbant/CAMB>

⁵ <https://github.com/healpy/healpy>

⁶ https://github.com/SihaoCheng/scattering_transform/blob/master/scattering/polyspectra_calculators

⁷ <https://www.kymat.io/>

⁸ The negative MI value of the BS indicates small numerical artifact in its estimation.

⁹ <https://github.com/andreimesinger/21cmFAST>

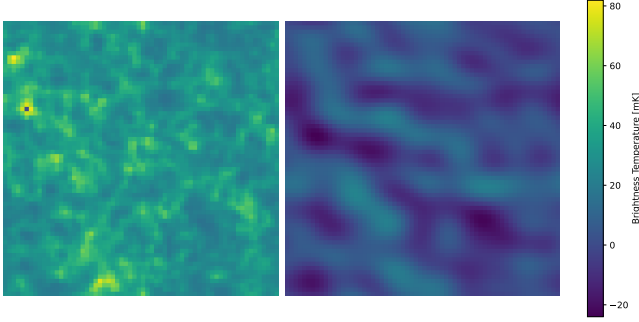


Figure 4. Illustration of the 21 cm maps. Left: cosmic 21 cm signal from the EoR. Right: mock observation (cosmic 21 cm signal with telescope noise) with SKA 1,000h observation.

comoving Mpc on each side, with 64^3 grid cells. In principle, tomographic, lightcone 21 cm field that covers the entire EoR will be needed to measure all the information, but for the simple purpose of the demonstration of MI, we only simulate coeval boxes at redshift $z = 12$. Also, we assume the spin temperature $T_S \gg T_{\text{CMB}}$, a valid approximation after the onset of the EoR.

We vary two reionization parameters, the ionizing efficiency ζ in the range $10 \leq \zeta \leq 250$, and the minimum virial temperature of halos hosting ionizing sources T_{vir} in the range $4 \leq \log_{10}(T_{\text{vir}}/\text{K}) \leq 6$. Cosmological parameters are fixed to the Planck results (Planck Collaboration et al. 2016), $(\Omega_\Lambda, \Omega_m, \Omega_b, n_s, \sigma_8, h) = (0.692, 0.308, 0.0484, 0.968, 0.815, 0.678)$.

In addition to simulating the cosmic 21 cm signal from the EoR, we also consider the observational effects with the assumption of the Square Kilometre Array (SKA) configuration. We employ the `Tools21cm`¹⁰ code (S. K. Giri et al. 2020) to simulate instrumental noise for the SKA1-Low configuration. The uv coverage is computed at the observed frequency corresponding to $z = 12$, and used to model the telescope response and suppress thermal noise in uv space. The system temperature is modeled as $T_{\text{sys}} = 60 (\nu/300 \text{ MHz})^{-2.55} + 100 \text{ K}$. Also, we assume a total of 1,000h observation, with 6 hours per day and 10 s integration time. The resulting noise cube is added to the signal cube to generate the mock observation. We illustrate an example of the cosmic 21 cm map and mock observation, respectively, in Figure 4.

To calculate PS and BS, we employ the `Tools21cm` code. Both PS and BS use the same 30 linearly uniform k -bins within $[k_{\text{min}}, k_{\text{max}}]$. For the ST, we use the `Kymatio` code with solid harmonic wavelets. The parameters are $J = 5$ scales, $L = 5$ angles, coefficients up to second order, and modulus powers $q = [0.5, 1, 2]$.

¹⁰ <https://github.com/sambit-giri/tools21cm>

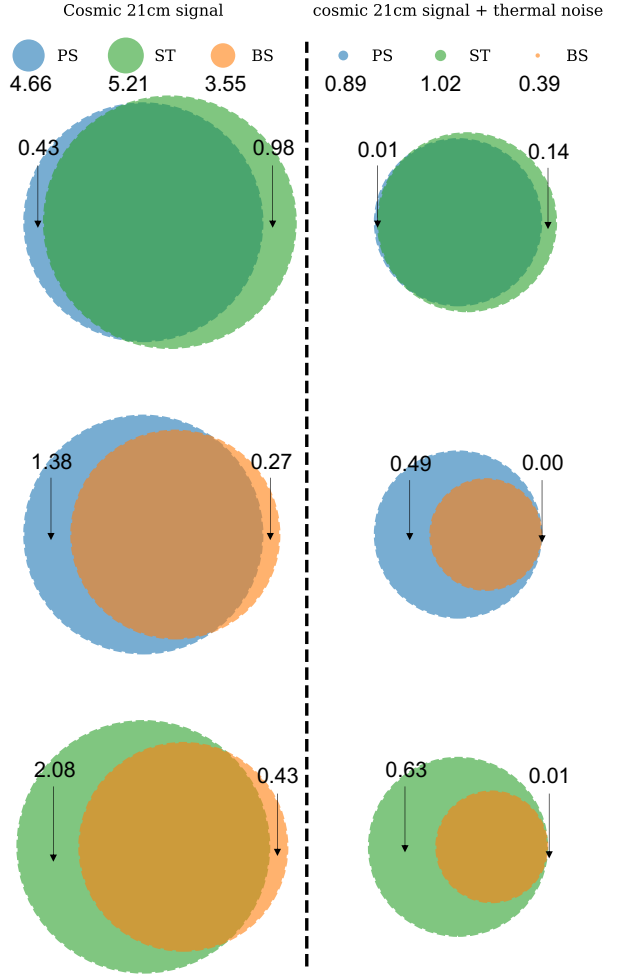


Figure 5. Venn diagram for the MI of summary statistics for the 21 cm maps from the EoR. We consider three summaries, PS (blue), ST (green) and BS (orange), and visualize their respective MI value (as marked) by the scaled circle size in the legend. We consider two cases, (left) the 21 cm map with only cosmic EoR signal, and (right) mock observations with 21 cm EoR signal and telescope noise with SKA 1,000h observation. For each case, we visualize the conditional MI (as marked with arrows) by the non-overlapped regions between two circles in Venn diagram.

The coefficients are averaged over different orientations, following X. Zhao et al. (2024).

In Figure 5, we visualize the MI of three summary statistics for the 21 cm maps from the EoR and, in particular, visualize the conditional MI, i.e. complementary information of one summary beyond another, using the Venn diagram. In the case with only cosmic 21 cm signal (i.e. no thermal noise), while the PS contains substantial information (i.e. $I(\theta; \text{PS}) = 4.66$), the ST captures the most (i.e. $I(\theta; \text{ST}) = 5.21$). Here, $\theta = \{\zeta, \log_{10}(T_{\text{vir}}/\text{K})\}$ is the set of reionization parameters. The ST con-

tributes additional non-Gaussian information beyond the PS (i.e. $I(\theta; \text{ST} \mid \text{PS}) = 0.98$). In comparison, the BS extracts less information than ST and PS (i.e. $I(\theta; \text{BS}) = 3.55$), and contributes less additional information beyond the PS or ST (i.e. $I(\theta; \text{BS} \mid \text{PS}) = 0.27$ and $I(\theta; \text{BS} \mid \text{ST}) = 0.43$).

Note that each pair of PS, ST and BS has nonzero conditional MI beyond each other for the cosmic 21 cm field. This implies that these three summaries extract different, albeit not completely different, aspects of information which are indeed complementary to each other.

For the case of mock observation that is applied with thermal noise assuming SKA 1,000h observation, the MI of these summaries for the mock observation is significantly smaller (i.e. $I(\theta; \text{PS}) = 0.89$, $I(\theta; \text{ST}) = 1.02$ and $I(\theta; \text{BS}) = 0.39$) than that for the case of cosmic signal. Basically, thermal noise is independent of reionization parameters, suppressing the inference effectiveness of the total 21 cm signal.

Note that the conditional MI, $I(\theta; \text{PS} \mid \text{ST})$, $I(\theta; \text{BS} \mid \text{PS})$ and $I(\theta; \text{BS} \mid \text{ST})$, are all nearly zero in this case. At face value, this means that the information of the BS is a subset of the PS, and the information of the PS is also a subset of the ST, when thermal noise at the SKA sensitivity level is applied. This interesting result serves as a proof-of-concept for MI as a metric of information content under our simplified simulations and mock observations, but it should not be overly interpreted. For example, the MI of BS might be underestimated herein because the calculation of BS only takes into account the configuration of equilateral triangles. More realistic assumptions on instrumental effects and systematics than in this work might also find a different complementarity between these summaries (see, e.g. C. A. Watkinson et al. 2022).

In summary, for all cases, the ST successfully extracts the most information (see also, X. Zhao et al. 2024), the PS ranks second, and the BS provides the least information. For the cosmic 21 cm map which is non-Gaussian, ST also provides significantly complementary information beyond the PS, whereas the BS adds a little beyond the PS. For the mock 21 cm observation with thermal noise, the information extracted by the ST contains that by the PS which also contains that by the BS.

These results demonstrate that the MI framework can successfully provide a quantitative assessment of both the total and complementary information content of different summary statistics. In addition, the application

to the 21 cm map highlights the advantage of ST in capturing non-Gaussian features from the 21 cm data.

4. CONCLUSIONS

In this paper, we demonstrate that MI provides a powerful and principled metric for evaluating the effectiveness of summary statistics in cosmological inference. Using GRFs as a validation case, we confirm that MI correctly identifies the optimal summaries. We then applied the approach to 21 cm simulations, showing how MI can quantify both the absolute information content of different summaries and their complementary contributions, enabling the systematic selection of summaries that enhance a base set.

Beyond evaluation, MI offers a natural target for learning informative summaries. By incorporating MI into loss functions, neural networks can be trained to produce summaries that maximize relevant information for inference tasks, offering a path toward automated and optimized summary design.

Current MI estimators, however, face challenges in very high-dimensional settings, such as field-level data. Advancing estimation techniques in this regime will allow us to fully quantify the maximal extractable information from complex cosmological fields, bridging the gap between theory and practical data analysis. Overall, MI provides both a theoretical foundation and a practical framework for designing and assessing summary statistics in modern cosmology.

ACKNOWLEDGMENTS

This work is supported by National SKA Program of China (grant No. 2020SKA0110401). X.Z. acknowledges the support through a grant from the Schmidt Sciences. We acknowledge the Tsinghua Astrophysics High-Performance Computing platform at Tsinghua University for providing computational and data storage resources that have contributed to the research results reported within this paper.

Software: 21cmFAST (A. Mesinger & S. Furlanetto 2007; A. Mesinger et al. 2011), NumPy (C. R. Harris et al. 2020), Matplotlib (J. D. Hunter 2007), SciPy (P. Virtanen et al. 2020), scikit-learn (F. Pedregosa et al. 2011), Python3 (G. Van Rossum & F. L. Drake 2009), Kymatio (M. Andreux et al. 2018), Tools21cm (S. K. Giri et al. 2020), healpy (K. M. Górski et al. 2005; A. Zonca et al. 2019), camb (A. Lewis & A. Challinor 2011), ltutil (M. Ho et al. 2024), Pylians (F. Villaescusa-Navarro 2018)

APPENDIX

A. MI AND BAYESIAN SUFFICIENCY

Let $s = f(x)$ be a deterministic statistic. Then $(\theta \rightarrow x \rightarrow s)$ forms a Markov chain. By the chain rule for MI,

$$I(\theta; x, s) = I(\theta; s) + I(\theta; x | s). \quad (\text{A1})$$

Since s is a function of x , we also have $I(\theta; x, s) = I(\theta; x)$. Hence,

$$I(\theta; x) = I(\theta; s) + I(\theta; x | s). \quad (\text{A2})$$

Thus $I(\theta; x) = I(\theta; s)$ if and only if $I(\theta; x | s) = 0$. Now, $I(\theta; x | s) = 0$ holds if and only if $\theta \perp x | s$, i.e.

$$p(\theta | x, s) = p(\theta | s) \quad (\text{A3})$$

Because $s = f(x)$, conditioning on (x, s) is the same as conditioning on x , so $p(\theta | x, s) = p(\theta | x)$. Therefore

$$I(\theta; x | s) = 0 \iff p(\theta | x) = p(\theta | s(x)) \quad (\text{A4})$$

Combining (A2) and (A4) yields

$$p(\theta | x) = p(\theta | s(x)) \iff I(\theta; x) = I(\theta; s(x)).$$

□

B. RELATIONSHIPS BETWEEN MI, FI, AND ESTIMATION MSE

For a likelihood $p(x|\theta)$, the FI matrix is defined as

$$\mathcal{I}_F(\theta) \equiv \mathbb{E}_{p(x|\theta)} [\nabla_\theta \log p(x|\theta) \nabla_\theta \log p(x|\theta)^\top]. \quad (\text{B5})$$

According to the Cramér–Rao bound, no unbiased estimator can achieve a smaller variance than the inverse of the FI matrix. This makes FI a natural tool to quantify how much information an observable carries about the parameters.

For a likelihood model $p(x|\theta)$ and a smooth prior $p(\theta)$, in addition to the FI matrix defined in Equation (B5), we define the *prior information matrix* as

$$\mathcal{J} \equiv \mathbb{E}_{p(\theta)} [\nabla_\theta \log p(\theta) \nabla_\theta \log p(\theta)^\top], \quad (\text{B6})$$

which quantifies the informativeness of the prior.

Efroimovich’s inequality (K.-Y. Lee 2022) states that

$$\frac{1}{2\pi e} e^{\frac{2}{d} h(\theta|x)} \geq \frac{1}{\det \left(\mathcal{J} + \mathbb{E}_{p(\theta)} [\mathcal{I}_F(\theta)] \right)^{1/d}}, \quad (\text{B7})$$

or

$$-\frac{2}{d} h(\theta|x) \leq \log \left[(2\pi e)^{-1} \det \left(\mathcal{J} + \mathbb{E}_{p(\theta)} [\mathcal{I}_F(\theta)] \right)^{1/d} \right], \quad (\text{B8})$$

where d is the dimension of θ . Here, $h(\theta|x) \equiv -\mathbb{E}_{p(\theta,x)} [\log p(\theta|x)]$ is the conditional entropy, and we have $I(\theta; x) = -h(\theta|x) + h(\theta)$, where $h(\theta)$ is the differential entropy of θ . Therefore, ignoring $h(\theta)$ (since it is independent of the field x), the Efroimovich’s inequality implies that the FI integrated over the prior $p(\theta)$ provides an upper bound on the MI.

On the other hand, assume we have an estimator $\hat{\theta}(x)$ of the parameter θ . Using the data processing inequality and the fact that entropy is maximized by a Gaussian, we have

$$\begin{aligned} \frac{2}{d} h(\theta|x) &\leq \frac{2}{d} h(\hat{\theta} - \theta) \\ &\leq \log \left((2\pi e) \det(\Sigma)^{1/d} \right) \\ &\leq \log \left[(2\pi e) \frac{1}{d} \mathbb{E}_{p(\theta,x)} \|\hat{\theta} - \theta\|^2 \right], \end{aligned} \quad (\text{B9})$$

or

$$-\frac{2}{d} h(\theta|x) \geq -\log \left[(2\pi e) \frac{1}{d} \mathbb{E}_{p(\theta,x)} \|\hat{\theta} - \theta\|^2 \right]. \quad (\text{B10})$$

This means that the Bayesian estimation error provides a lower bound on the MI. These inequalities (Equations B8 and B10) reveal interesting relationships between different information metrics: MI is upper bounded by FI, and lower bounded by the mean squared error (MSE) of the estimation.

Combining Equations (B8) and (B10) yields

$$\mathbb{E}_{p(\theta,x)} \|\hat{\theta} - \theta\|^2 \geq \frac{d}{\det \left(\mathcal{J} + \mathbb{E}_{p(\theta)} [\mathcal{I}_F(\theta)] \right)^{1/d}}. \quad (\text{B11})$$

This is known as *van Trees inequality* (H. L. Van Trees 2004; A. B. Tsybakov 2008), or the Bayesian Cramér–Rao bound (R. D. Gill & B. Y. Levit 1995).

C. SUMMARY STATISTICS

We summarize the definitions and main configuration parameters of summaries used in this paper.

C.1. PS

The PS $P(\mathbf{k})$ is the most widely used statistic in cosmology. For a 3D field, it is defined as

$$\langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle = (2\pi)^3 \delta^D(\mathbf{k} + \mathbf{k}') P(\mathbf{k}), \quad (\text{C12})$$

where $\delta(\mathbf{k})$ is the Fourier transform of the field, δ^D is the Dirac delta function, and $\langle \cdot \rangle$ denotes an ensemble average.

In practice, ensemble averaging is replaced by binning over spherical shells in Fourier space, assuming statistical isotropy:

$$P(k) \approx \frac{\sum_{\mathbf{k} \in k} |\delta(\mathbf{k})|^2}{N_k V}, \quad (\text{C13})$$

where V is the survey volume and N_k is the number of modes in the spherical k -shell $[k - \Delta k/2, k + \Delta k/2]$. This procedure yields the spherically averaged power spectrum. While spherical harmonic methods are often used for CMB and other sky-projected fields, in this paper we restrict our PS calculations to 2D and 3D flat fields.

The key parameters for the PS calculation are the binning scheme and the accessible k -range. For a d -dimensional field with box size L and N grid cells per dimension, the minimum and maximum Fourier modes are

$$k_{\min} = \frac{2\pi}{L}, \quad k_{\max} = \frac{\pi N \sqrt{d}}{L}. \quad (\text{C14})$$

The bins are typically spaced either linearly or logarithmically within this range.

C.2. BS

Non-Gaussian cosmological fields motivate the use of higher-order correlation functions. The BS is the Fourier transform of the three-point correlation function. It is defined as

$$\langle \delta(\mathbf{k}_1) \delta(\mathbf{k}_2) \delta(\mathbf{k}_3) \rangle = (2\pi)^3 \delta^D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \times B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3). \quad (\text{C15})$$

Similar to PS, the bispectra are spherically averaged over triangle orientations thanks to isotropy, so that the BS depends only on triangle shape and size. The reduced BS is also defined to separate information from

the PS,

$$Q(k_1, k_2, k_3) \equiv \frac{B(k_1, k_2, k_3)}{P(k_1)P(k_2) + P(k_2)P(k_3) + P(k_3)P(k_1)}. \quad (\text{C16})$$

In general, the BS requires sampling over all closed triangles of wavevectors, which is computationally demanding. For tractability, in this paper we restrict ourselves to the configuration of equilateral triangles, parameterized by a single k value. This choice keeps the parameter dependence consistent with that of the PS, with the key parameter choice again being the k -binning.

C.3. ST

Introduced by [S. Mallat \(2011\)](#), the ST provides a hierarchical representation of fields by iteratively combining wavelet convolutions, nonlinearities, and spatial averaging. For a field x , the order- m scattering coefficient at scales $(\lambda_1, \dots, \lambda_m)$ is

$$Sx[\lambda_1, \dots, \lambda_m] = |\psi_{\lambda_m} \star \dots \star \psi_{\lambda_1} \star x| \dots|, \quad (\text{C17})$$

where $\star$ denotes convolution and $\{\psi_\lambda\}$ is a family of wavelets. For 3D fields, an additional dimensionality reduction step is used, in which the modulus outputs are raised to the power q and spatially averaged into scalar coefficients.

The main configuration parameters of the ST include (1) the choice of wavelet family (e.g., Morlet, solid harmonic); (2) the index for the scale J and the orientation L ; (3) the maximum order m (typically 1 or 2); (4) the modulus exponent q , controlling sensitivity to large vs small fluctuations. These hyperparameters jointly determine the richness and dimensionality of the resulting summary.

The ST captures multiscale, multiorientation features while enforcing translation and rotation invariance. It has been demonstrated to recover non-Gaussian information and has applied to parameter inference in 3D reionization fields ([S. Cheng et al. 2020](#); [X. Zhao et al. 2024](#)).

REFERENCES

- Adame, A. G., Aguilar, J., Ahlen, S., et al. 2025, JCAP, 2025, 021, doi: [10.1088/1475-7516/2025/02/021](https://doi.org/10.1088/1475-7516/2025/02/021)
- Akhmetzhanova, A., Mishra-Sharma, S., & Dvorkin, C. 2024, MNRAS, 527, 7459, doi: [10.1093/mnras/stad3646](https://doi.org/10.1093/mnras/stad3646)
- Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, MNRAS, 488, 4440, doi: [10.1093/mnras/stz1960](https://doi.org/10.1093/mnras/stz1960)
- Alsing, J., Wandelt, B., & Feeney, S. 2018, MNRAS, 477, 2874, doi: [10.1093/mnras/sty819](https://doi.org/10.1093/mnras/sty819)
- Anbajagane, D., Pandey, S., & Chang, C. 2024, The Open Journal of Astrophysics, 7, 108, doi: [10.33232/001c.126788](https://doi.org/10.33232/001c.126788)
- Anderson, L., Aubourg, E., Bailey, S., et al. 2012, MNRAS, 427, 3435, doi: [10.1111/j.1365-2966.2012.22066.x](https://doi.org/10.1111/j.1365-2966.2012.22066.x)
- Andreux, M., Angles, T., Exarchakis, G., et al. 2018, arXiv e-prints, arXiv:1812.11214, <https://arxiv.org/abs/1812.11214>

- Barber, D., & Agakov, F. 2004, *Advances in neural information processing systems*, 16, 201
- Beyond-2pt Collaboration, :, Krause, E., et al. 2024, arXiv e-prints, arXiv:2405.02252, doi: [10.48550/arXiv.2405.02252](https://doi.org/10.48550/arXiv.2405.02252)
- Campagne, J.-E., Lanusse, F., Zuntz, J., et al. 2023, *The Open Journal of Astrophysics*, 6, 15, doi: [10.21105/astro.2302.05163](https://doi.org/10.21105/astro.2302.05163)
- Cheng, P., Hao, W., Dai, S., et al. 2020, arXiv e-prints, arXiv:2006.12013, doi: [10.48550/arXiv.2006.12013](https://doi.org/10.48550/arXiv.2006.12013)
- Cheng, S., Ting, Y.-S., Ménard, B., & Bruna, J. 2020, *MNRAS*, 499, 5902, doi: [10.1093/mnras/staa3165](https://doi.org/10.1093/mnras/staa3165)
- Contarini, S., Pisani, A., Hamaus, N., et al. 2023, *ApJ*, 953, 46, doi: [10.3847/1538-4357/acde54](https://doi.org/10.3847/1538-4357/acde54)
- DES Collaboration, Abbott, T. M. C., Aguena, M., et al. 2025, arXiv e-prints, arXiv:2503.13632, doi: [10.48550/arXiv.2503.13632](https://doi.org/10.48550/arXiv.2503.13632)
- Devon Hjelm, R., Fedorov, A., Lavoie-Marchildon, S., et al. 2018, arXiv e-prints, arXiv:1808.06670, doi: [10.48550/arXiv.1808.06670](https://doi.org/10.48550/arXiv.1808.06670)
- Diao, K., Chen, Z., Chen, X., & Mao, Y. 2024, *ApJ*, 974, 141, doi: [10.3847/1538-4357/ad6c40](https://doi.org/10.3847/1538-4357/ad6c40)
- Euclid Collaboration, Congedo, G., Miller, L., et al. 2024, *A&A*, 691, A319, doi: [10.1051/0004-6361/202450617](https://doi.org/10.1051/0004-6361/202450617)
- Federici, M., Dutta, A., Forré, P., Kushman, N., & Akata, Z. 2020, arXiv e-prints, arXiv:2002.07017, doi: [10.48550/arXiv.2002.07017](https://doi.org/10.48550/arXiv.2002.07017)
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, *Physics Reports*, 433, 181, doi: [10.1016/j.physrep.2006.08.002](https://doi.org/10.1016/j.physrep.2006.08.002)
- Gatti, M., Jeffrey, N., Whiteway, L., et al. 2024, *PhRvD*, 109, 063534, doi: [10.1103/PhysRevD.109.063534](https://doi.org/10.1103/PhysRevD.109.063534)
- Gill, R. D., & Levit, B. Y. 1995, *Bernoulli*, 1, 59
- Gillet, N., Mesinger, A., Greig, B., Liu, A., & Ucci, G. 2019, *MNRAS*, 484, 282, doi: [10.1093/mnras/stz010](https://doi.org/10.1093/mnras/stz010)
- Giri, S. K., Mellema, G., & Jensen, H. 2020, *Journal of Open Source Software*, 5, 2363, doi: [10.21105/joss.02363](https://doi.org/10.21105/joss.02363)
- Gleser, L., Nusser, A., Ciardi, B., & Desjacques, V. 2006, *MNRAS*, 370, 1329, doi: [10.1111/j.1365-2966.2006.10556.x](https://doi.org/10.1111/j.1365-2966.2006.10556.x)
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, 622, 759, doi: [10.1086/427976](https://doi.org/10.1086/427976)
- Hahn, C., Lemos, P., Parker, L., et al. 2023, arXiv e-prints, arXiv:2310.15246, doi: [10.48550/arXiv.2310.15246](https://doi.org/10.48550/arXiv.2310.15246)
- Hamaus, N., Pisani, A., Choi, J.-A., et al. 2020, *JCAP*, 2020, 023, doi: [10.1088/1475-7516/2020/12/023](https://doi.org/10.1088/1475-7516/2020/12/023)
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2)
- Ho, M., Bartlett, D. J., Chartier, N., et al. 2024, *The Open Journal of Astrophysics*, 7, 54, doi: [10.33232/001c.120559](https://doi.org/10.33232/001c.120559)
- Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Ishmael Belghazi, M., Baratin, A., Rajeswar, S., et al. 2018, arXiv e-prints, arXiv:1801.04062, doi: [10.48550/arXiv.1801.04062](https://doi.org/10.48550/arXiv.1801.04062)
- Ivanov, M. M., Philcox, O. H. E., Cabass, G., et al. 2023, *PhRvD*, 107, 083515, doi: [10.1103/PhysRevD.107.083515](https://doi.org/10.1103/PhysRevD.107.083515)
- Kapahtia, A., Chingangbam, P., Ghara, R., Appleby, S., & Choudhury, T. R. 2021, *JCAP*, 2021, 026, doi: [10.1088/1475-7516/2021/05/026](https://doi.org/10.1088/1475-7516/2021/05/026)
- Kayo, I., Takada, M., & Jain, B. 2013, *MNRAS*, 429, 344, doi: [10.1093/mnras/sts340](https://doi.org/10.1093/mnras/sts340)
- Kraskov, A., Stögbauer, H., & Grassberger, P. 2004, *PhRvE*, 69, 066138, doi: [10.1103/PhysRevE.69.066138](https://doi.org/10.1103/PhysRevE.69.066138)
- Kratochvil, J. M., Lim, E. A., Wang, S., et al. 2012, *PhRvD*, 85, 103513, doi: [10.1103/PhysRevD.85.103513](https://doi.org/10.1103/PhysRevD.85.103513)
- Lee, K.-Y. 2022, PhD thesis, University of California, Berkeley
- Lewis, A., & Challinor, A. 2011, *CAMB: Code for Anisotropies in the Microwave Background*, Astrophysics Source Code Library, record ascl:1102.026
- Makinen, T. L., Sui, C., Wandelt, B. D., Porqueres, N., & Heavens, A. 2024, arXiv e-prints, arXiv:2410.07548, doi: [10.48550/arXiv.2410.07548](https://doi.org/10.48550/arXiv.2410.07548)
- Maksimova, N. A., Garrison, L. H., Eisenstein, D. J., et al. 2021, *MNRAS*, 508, 4017, doi: [10.1093/mnras/stab2484](https://doi.org/10.1093/mnras/stab2484)
- Malandrino, R., Lavaux, G., Wandelt, B. D., McAlpine, S., & Jasche, J. 2025, arXiv e-prints, arXiv:2507.06866, doi: [10.48550/arXiv.2507.06866](https://doi.org/10.48550/arXiv.2507.06866)
- Mallat, S. 2011, arXiv e-prints, arXiv:1101.2286, doi: [10.48550/arXiv.1101.2286](https://doi.org/10.48550/arXiv.1101.2286)
- Mesinger, A., & Furlanetto, S. 2007, *ApJ*, 669, 663, doi: [10.1086/521806](https://doi.org/10.1086/521806)
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, *MNRAS*, 411, 955, doi: [10.1111/j.1365-2966.2010.17731.x](https://doi.org/10.1111/j.1365-2966.2010.17731.x)
- Nguyen, N.-M., Schmidt, F., Tucci, B., Reinecke, M., & Kostić, A. 2024, *PhRvL*, 133, 221006, doi: [10.1103/PhysRevLett.133.221006](https://doi.org/10.1103/PhysRevLett.133.221006)
- Pakmor, R., Springel, V., Coles, J. P., et al. 2023, *MNRAS*, 524, 2539, doi: [10.1093/mnras/stac3620](https://doi.org/10.1093/mnras/stac3620)
- Papamakarios, G., Pavlakou, T., & Murray, I. 2017, arXiv e-prints, arXiv:1705.07057, doi: [10.48550/arXiv.1705.07057](https://doi.org/10.48550/arXiv.1705.07057)
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Pillepich, A., Springel, V., Nelson, D., et al. 2018, *MNRAS*, 473, 4077, doi: [10.1093/mnras/stx2656](https://doi.org/10.1093/mnras/stx2656)
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, 594, A13, doi: [10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830)

- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6, doi: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910)
- Polyanskiy, Y., & Wu, Y. 2025, *Information Theory: From Coding to Learning*, doi: [10.1017/9781108966351](https://doi.org/10.1017/9781108966351)
- Poole, B., Ozair, S., van den Oord, A., Alemi, A. A., & Tucker, G. 2019, arXiv e-prints, arXiv:1905.06922, doi: [10.48550/arXiv.1905.06922](https://doi.org/10.48550/arXiv.1905.06922)
- Porqueres, N., Heavens, A., Mortlock, D., Lavaux, G., & Makinen, T. L. 2023, arXiv e-prints, arXiv:2304.04785, doi: [10.48550/arXiv.2304.04785](https://doi.org/10.48550/arXiv.2304.04785)
- Prelogović, D., & Mesinger, A. 2024, *A&A*, 688, A199, doi: [10.1051/0004-6361/202449309](https://doi.org/10.1051/0004-6361/202449309)
- Prelogović, D., Mesinger, A., Murray, S., Fiameni, G., & Gillet, N. 2022, *MNRAS*, 509, 3852, doi: [10.1093/mnras/stab3215](https://doi.org/10.1093/mnras/stab3215)
- Régaldo-Saint Blancard, B., Hahn, C., Ho, S., et al. 2024, *PhRvD*, 109, 083535, doi: [10.1103/PhysRevD.109.083535](https://doi.org/10.1103/PhysRevD.109.083535)
- Schaye, J., Kugel, R., Schaller, M., et al. 2023, *MNRAS*, 526, 4978, doi: [10.1093/mnras/stad2419](https://doi.org/10.1093/mnras/stad2419)
- Sharma, D., Dai, B., & Seljak, U. 2024, *JCAP*, 2024, 010, doi: [10.1088/1475-7516/2024/08/010](https://doi.org/10.1088/1475-7516/2024/08/010)
- Shimabukuro, H., Yoshiura, S., Takahashi, K., Yokoyama, S., & Ichiki, K. 2016, *MNRAS*, 458, 3003, doi: [10.1093/mnras/stw482](https://doi.org/10.1093/mnras/stw482)
- Shwartz Ziv, R., & LeCun, Y. 2024, *Entropy*, 26, 252, doi: [10.3390/e26030252](https://doi.org/10.3390/e26030252)
- Simon-Onfroy, H., Lanusse, F., & de Mattia, A. 2025, arXiv e-prints, arXiv:2504.20130, doi: [10.48550/arXiv.2504.20130](https://doi.org/10.48550/arXiv.2504.20130)
- Song, J., & Ermon, S. 2019, arXiv preprint arXiv:1910.06222
- Square Kilometre Array Cosmology Science Working Group, Bacon, D. J., Battye, R. A., et al. 2020, *PASA*, 37, e007, doi: [10.1017/pasa.2019.51](https://doi.org/10.1017/pasa.2019.51)
- Sui, C., Pandey, S., & Wandelt, B. D. 2025, arXiv e-prints, arXiv:2507.07833, doi: [10.48550/arXiv.2507.07833](https://doi.org/10.48550/arXiv.2507.07833)
- Sui, C., Zhao, X., Jing, T., & Mao, Y. 2023, in *Machine Learning for Astrophysics*, 30, doi: [10.48550/arXiv.2307.04994](https://doi.org/10.48550/arXiv.2307.04994)
- Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, *ApJ*, 480, 22, doi: [10.1086/303939](https://doi.org/10.1086/303939)
- Tsybakov, A. B. 2008, *Introduction to Nonparametric Estimation*, 1st edn. (Springer Publishing Company, Incorporated)
- van den Oord, A., Li, Y., & Vinyals, O. 2018, arXiv e-prints, arXiv:1807.03748, doi: [10.48550/arXiv.1807.03748](https://doi.org/10.48550/arXiv.1807.03748)
- Van Rossum, G., & Drake, F. L. 2009, *Python 3 Reference Manual* (Scotts Valley, CA: CreateSpace)
- Van Trees, H. L. 2004, *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory* (John Wiley & Sons)
- Villaescusa-Navarro, F. 2018, *Pylians: Python libraries for the analysis of numerical simulations*, *Astrophysics Source Code Library*, record ascl:1811.008 <http://ascl.net/1811.008>
- Villaescusa-Navarro, F., Hahn, C., Massara, E., et al. 2020, *ApJS*, 250, 2, doi: [10.3847/1538-4365/ab9d82](https://doi.org/10.3847/1538-4365/ab9d82)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- Watkinson, C. A., Greig, B., & Mesinger, A. 2022, *MNRAS*, 510, 3838, doi: [10.1093/mnras/stab3706](https://doi.org/10.1093/mnras/stab3706)
- Wu, Y. 2017, *Lecture Notes for ECE598YW (UIUC)*, 16, 15
- Zhao, X., Mao, Y., Cheng, C., & Wandelt, B. D. 2022, *ApJ*, 926, 151, doi: [10.3847/1538-4357/ac457d](https://doi.org/10.3847/1538-4357/ac457d)
- Zhao, X., Mao, Y., Zuo, S., & Wandelt, B. D. 2024, *ApJ*, 973, 41, doi: [10.3847/1538-4357/ad5ff0](https://doi.org/10.3847/1538-4357/ad5ff0)
- Zonca, A., Singer, L., Lenz, D., et al. 2019, *Journal of Open Source Software*, 4, 1298, doi: [10.21105/joss.01298](https://doi.org/10.21105/joss.01298)