



HAL
open science

Wavelet-based statistics for enhanced 21cm EoR parameter constraints

Ian Hothi, Erwan Allys, Benoît Semelin, François Boulanger

► **To cite this version:**

Ian Hothi, Erwan Allys, Benoît Semelin, François Boulanger. Wavelet-based statistics for enhanced 21cm EoR parameter constraints. *Astronomy & Astrophysics - A&A*, 2024, 686, pp.A212. 10.1051/0004-6361/202348444 . hal-04610255

HAL Id: hal-04610255




<https://hal.science/hal-04610255v1>

Submitted on 12 Jun 2024

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

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

Wavelet-based statistics for enhanced 21cm EoR parameter constraints

Ian Hothi^{1,2}, Erwan Allys², Benoît Semelin¹, and François Boulanger²

¹ LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Université, 75014 Paris, France
e-mail: Ian.Hothi@obspm.fr

² Laboratoire de Physique de l'ENS, ENS, Université PSL, CNRS, Sorbonne Université, Université Paris Cité, 75005 Paris, France

Received 31 October 2023 / Accepted 11 March 2024

ABSTRACT

We propose a new approach to improve the precision of astrophysical parameter constraints for the 21cm signal from the epoch of reionisation (EoR). Our method introduces new sets of summary statistics, hereafter ‘evolution compressed’ statistics, which quantify the spectral evolution of the 2D spatial statistics computed at fixed redshift. We defined such compressed statistics for power spectrum (PS), wavelet scattering transforms (WST), and wavelet moments (WM), which also characterise non-Gaussian features. To compare these different statistics, along with the 3D power spectrum, we estimated their Fisher information on three cosmological parameters from an ensemble of simulations of 21cm EoR data, both in noiseless and noisy scenarios using Square Kilometre Array (SKA) noise levels equivalent to 100 and 1000 h of observations. We also compare wavelet statistics, in particular WST, built from standard directional Morlet wavelets, as well as from a set of isotropic wavelets derived from the binning window function of the 2D power spectrum. For the noiseless case, the compressed wavelet statistics give constraints that are up to five times more precise than those obtained from the 3D isotropic power spectrum. At the same time, for 100 h SKA noise, from which it is difficult to extract non-Gaussian features, compressed wavelet statistics still give over 30% tighter constraints. We find that the wavelet statistics with wavelets derived from the power-spectrum binning window function provide the tightest constraints of all the statistics, with the WSTs seemingly performing better than the WMs, in particular when working with noisy data. The findings of this study demonstrate that evolution-compressed statistics extract more information than usual 3D isotropic power-spectra approaches and that our wavelet-based statistics can consistently outmatch power-spectrum-based statistics. When constructing such wavelet-based statistics, we also emphasise the need to choose a set of wavelets with an appropriate spectral resolution concerning the astrophysical process studied.

Key words. methods: statistical – dark ages, reionization, first stars

1. Introduction

The formation of the first stars marked a pivotal phase change in the Universe, known as the epoch of reionisation (EoR), during which the ionisation state transitioned from a predominantly neutral state to a primarily ionised one. This significant epoch serves as the precursor to the state of the intergalactic medium (IGM) that we observe today. While direct observations of the EoR are challenging, indirect constraints on its timings and progression have been obtained. Observations of the cosmic microwave background (CMB) have provided valuable insights, and suggest that a substantial portion of the reionisation process occurred after a redshift of $z = 14$ (Planck Collaboration Int. XLVII 2016; Gorce et al. 2018). Furthermore, the prominence of the Gunn-Peterson trough observed in the spectra of quasars has offered additional constraints. Early investigations by Fan et al. (2006) indicated that a considerable fraction of the Universe underwent ionisation by a redshift of $z \sim 6$. However, more recent findings based on a larger number of spectra have led to an evolution of this perspective, suggesting a later completion of reionisation around $z \sim 5.3$ (Becker et al. 2015; Bosman et al. 2018; Qin et al. 2021).

The 21cm signal emitted from the hyperfine transition in neutral hydrogen during the EoR is the most promising observable to constrain the EoR. The most common statistic used to constrain the EoR’s 21cm signal is its 3D spherical power spectrum. Several radio telescopes that are either

currently taking data or are yet to come online are designed to detect the 21cm power spectrum; for example, the Low-Frequency Array (LOFAR)¹ (van Haarlem et al. 2013); the Giant Metrewave Radio Telescope (GMRT)² (Paciga et al. 2011, 2013); the Murchison Widefield Array (MWA)³ (Tingay et al. 2013); the Hydrogen Epoch of Reionization Array (HERA)⁴ (DeBoer et al. 2017); the new extension in Nançay upgrading LOFAR (NeNuFar)⁵ (Zarka et al. 2012; Munshi et al. 2024); and the Amsterdam-ASTRON Radio Transients Facility and Analysis Center (AARTFAAC)⁶ (Prasad et al. 2016; Gehlot 2019).

The power spectrum, which quantifies signal power distribution as a function of scale, serves as a valuable two-point statistic for analysing the 21cm signal from the EoR. To probe the EoR’s non-Gaussian nature, which is related to coupling between scales, is not possible solely with the power spectrum. In this context, the three-point statistics, such as the bispectrum, has garnered significant attention within the EoR field (Shimabukuro et al. 2016; Majumdar et al. 2017; Watkinson et al. 2017, 2021; Gorce & Pritchard 2019; Hutter et al. 2020). These high-order statistics are employed to

¹ <http://www.lofar.org/>

² <http://gmrt.ncra.tifr.res.in/>

³ <http://www.mwatelescope.org/>

⁴ <http://reionization.org>

⁵ <https://nenufar.obs-nancay.fr/en/homepage-en/>

⁶ <http://aartfaac.org>

capture more of the complexities of the signal. The bispectrum, derived from the Fourier transform of three-point correlations, characterises the statistical dependence among triplets of distinct sets of Fourier modes in a signal. Consequently, it allows us to go beyond power spectrum studies. As a high-order statistic, the bispectrum encounters challenges, such as high levels of variance due, in particular, to the presence of outliers (Welling 2005). Outliers are extreme values that significantly deviate from the typical values of the data and can have a disproportionate impact on high-order moments. The bispectrum is sensitive to outliers because it is a high-order computation that involves triple products. Such high levels of variance in our summary statistics will make it more difficult to precisely infer physical parameters from limited data. To overcome these limitations, novel wavelet-based statistics based on wavelet transforms have emerged (Mallat 2011). Wavelet transforms have been used in cosmological parameter estimation in the form of wavelet moments (WM), which apply statistical moments to extract information from wavelet transform; see for example Eickenberg et al. (2022). More hierarchical representations – using a cascade of wavelet transforms and non-linearities – called wavelet scattering transforms (WST) have also been applied to cosmology and found successful applications in studying the highly non-Gaussian interstellar medium (ISM), where the interplay between gravity, magneto-hydrodynamics, and various processes within the ISM generates complex statistical patterns (Allys et al. 2019; Regalado-Saint Blancard et al. 2020; Saydjari et al. 2021). WSTs have also been used for classification and parameter inference in the context of weak lensing (Cheng & Ménard 2021) and the large-scale structure of the Universe (Valogiannis & Dvorkin 2022). These scattering transforms offer a means to probe non-Gaussianity without suffering from the shortcomings associated with high-order statistics, such as the bispectrum. Greig et al. (2022) were the first authors to apply WST to 2D images of the 21cm signal across multiple redshifts. Using a Fisher formalism for astrophysical parameter constraints, these authors demonstrated that the WST performs comparably to, and in some instances better than, a 3D power spectrum analysis on the same dataset. This latter result is a pioneering study in its application of the WST to the 21cm signal and as such should be confirmed and extended.

In their comparison of the power spectrum and the WST, Greig et al. (2022) found that certain aspects appear to favour the power spectrum, potentially skewing the results. Specifically, the authors employed a Fourier bin range for the power spectrum that effectively mitigates resolution effects within the simulations. Furthermore, they used a binning scheme that particularly benefits larger scales, which are less affected by noise. However, a comparable treatment is not extended to the WST, which may be more impacted by noise.

Another point is that the calculation of the covariance for the 21cm power spectrum, which included 1000 h of SKA noise, involved the use of 21CMSENSE⁷ (Pober et al. 2013, 2014) and thus (i) does not contain off-diagonal contributions, from cosmic variance for example (see Prelogović & Mesinger 2023), and (ii) is sensitive to the limited number of simulations used in the study. In contrast, the covariance for WST was directly derived from the limited number of fiducial simulations. This discrepancy in the covariance estimation methods introduces a potential bias.

In the present paper, we introduce 2+1 statistics – hereafter referred to as evolution-compressed statistics –, which quantify

the spectral evolution of the 2D spatial statistics computed at a fixed redshift in order to enhance the constraint on astrophysical parameters. With these evolution-compressed summary statistics, we aim to extract additional information beyond the 3D isotropic power spectrum, which is a well-established and extensively studied statistic in 21cm physics. We achieved this using recently developed wavelet-based statistics, which we built from a set of wavelets with a proper spatial frequency sampling. We validated this set of wavelets by first recovering the same information as the power spectrum in a wavelet framework, before adding non-Gaussian information through the use of non-linear statistics. We compared these wavelet statistics, built both from the WST statistic and WM, to the 2D evolution compressed power spectrum, and the 3D spherically averaged power spectrum; this comparative analysis provides valuable insights into the potential benefits and limitations of incorporating anisotropic summary statistics into parameter estimation.

The paper is structured as follows: We introduce the statistics used in this work – including WST and how its reduced form, the reduced wavelet scattering transform (RWST), is calculated – and our wavelet application in Sect. 2. Section 3 outlines the simulation used in this paper and how we set up our Fisher analysis. In Sect. 4, we show the results of comparing our different summary statistics under different noise treatments. We then summarise our findings and provide conclusions in Sect. 5.

2. Statistics

In this section, we present the various summary statistics used in this paper, including the 2D and 3D power spectrum, WM, and WST. Except for the 3D power spectrum, these statistics are calculated for fixed- z 2D spatial slices of the lightcone. Towards the end of this section, we also introduce the compression method that we use to compress the redshift evolution for 2D spatial statistics.

2.1. Power spectrum

The power spectrum is the most commonly used statistic in the 21cm field. We consider a field $I(\mathbf{x})$, where \mathbf{x} refers to the spatial location, as well as $\tilde{I}(\mathbf{k})$ its Fourier transform. The estimated power spectrum for a given bin is:

$$P_i = \int |\tilde{I}(\mathbf{k}) \cdot W_i(\mathbf{k})|^2 d\mathbf{k}, \quad (1)$$

where $W_i(\mathbf{k})$ is the spectral window function defining the i th bin. The power spectrum is then an ensemble of N_b summary statistics, with N_b being the number of bins. In this paper, we restrict ourselves to isotropic power-spectrum estimates, which correspond to spectral windows for which $W_i(\mathbf{k})$ depends only on k the modulus of \mathbf{k} . A typical choice for this window function is the top hat function, which equally weighs all k -modes in a given bin.

In this paper, we explore two applications of the power spectrum. The first application involves a first step of 2D spectral binning of the lightcone, that is, applying the power spectrum to each given frequency channel. As we consider an isotropic power spectrum, the window function, $W_i(\mathbf{k})$, is only a function of $k = \sqrt{k_x^2 + k_y^2}$, and corresponds to 2D circular shells. We chose to use a Gaussian window function (see Sect. 2.2.2 for the reasoning behind this choice), where the spectral window

⁷ [//github.com/jpober/21cmSense](https://github.com/jpober/21cmSense)

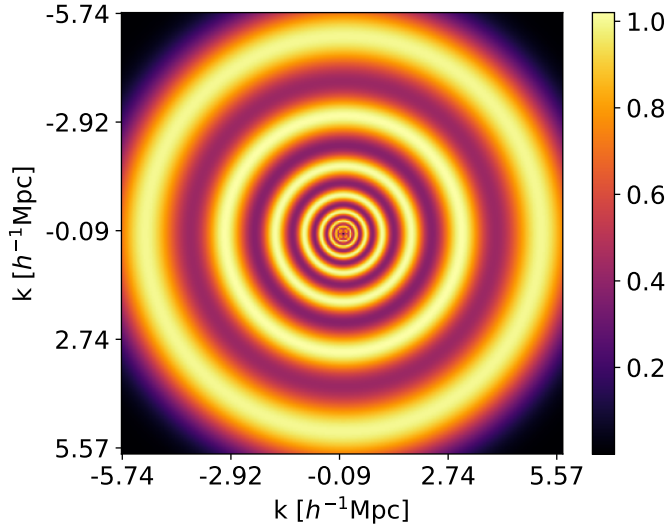


Fig. 1. Two-dimensional perspective of the binning window functions in Fourier space. We consider logarithmic binning, with the concentric circles close to the centre representing the window function of low k -bins and the concentric circles towards the edges representing the window function of high k -bins.

function of the i th-bin is defined as:

$$W_i(\mathbf{k}) = \exp\left(-\frac{(k - \bar{k}_i)^2}{2\sigma_i^2}\right), \quad (2)$$

where \bar{k}_i is the central frequency of the i th k -bin and σ_i its standard deviation. We bin our Fourier space into $N_b = 9$ logarithmic (base 10) bins for our binning scheme.

Figures 1 and 2 show these 2D spectral windows, as well as their 1D radial profile. We use PS_i^{2D} to denote the resulting power-spectrum estimates, for which we also have $N_b = 9$ coefficients. While this choice of spectra window leads to some power being distributed across adjacent bins, both sampling methods produce similar power spectra, with minor differences arising from the use of the window function. These differences are more prominent on low- k bins than on higher k -bins, and lead to a change of up to 20% in the power in a given bin.

The second application is the 3D isotropic spherically averaged power spectrum, called P_i^{3D} , which involves a 3D binning of the Fourier transformed lightcone. In this case, the window function, $W_i(k)$, is also a function of $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ only, and corresponds to 3D spherical shells. For continuity in the treatment of our binning, we chose to use the Gaussian window function defined in Eq. (2) and bin our 3D Fourier space into $N_b = 9$ logarithmic (base 10) bins.

2.2. Wavelet statistics

2.2.1. Wavelet transform and choice of wavelets

The wavelet statistics we use are built from wavelet transforms. Wavelet transforms are a mathematical tool used to process signals and images. They provide space and scale localisation of features, which makes them useful for compression, denoising, feature extraction, and pattern recognition. Unlike traditional Fourier-based techniques, wavelet transforms decompose data into maps that describe the contribution of different scales and positions within the data, allowing a localised representation at different resolutions.

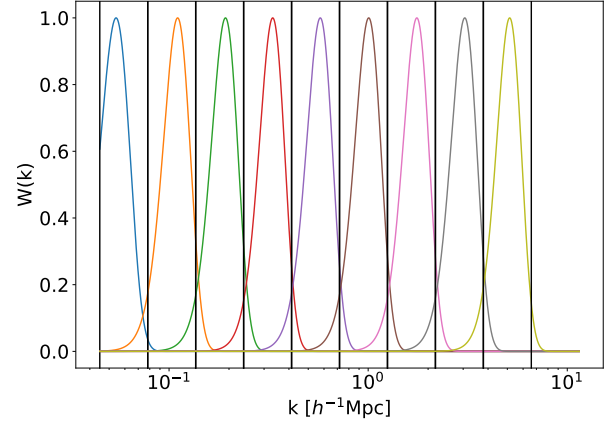


Fig. 2. One-dimensional line through our 2D Gaussian window functions, shown in Fig. 1.

A wavelet is an oscillatory function with zero mean that is localised around the origin. In two dimensions, a set of wavelets are built by dilating and rotating a single mother wavelet, $\psi(\mathbf{x})$:

$$\psi_{j,\theta}(\mathbf{x}) = 2^{-j/Q} \psi(2^{-j/Q} r_\theta^{-1} \mathbf{x}), \quad (3)$$

where r_θ^{-1} is the rotation operator of angle θ and j is the scale parameter. We usually consider integer numbers for scales j and angle θ , having J scales j between 1 and J , and dividing π into L angles. The quality factor Q controls the spectral binning of the wavelet. This binning is dyadic if $Q = 1$, that is, the integer scale j corresponds to a scale $\sim 2^j$ in pixels. We note that isotropic wavelets are only described by a scale parameter j because they are rotation invariant. In the following, we use λ to denote the ensemble of (j, θ) parameters used to identify each wavelet.

A wavelet transform is obtained by convoluting an image by a set of wavelets:

$$WT_\lambda[I](\mathbf{x}) = I * \psi_\lambda(\mathbf{x}), \quad (4)$$

where $*$ denotes a convolution. Wavelet statistics are usually constructed by successive applications of wavelet transforms and non-linear operators before an overall spatial integration. However, the choice of wavelet is critical because it drives the spatial frequency resolution of the wavelet statistics. Optimal choices depend on the application. To study this choice, we considered the following two different wavelet sets.

The first set of wavelets are dyadic directional Morlet wavelets, from which the WSTs are usually built for astrophysical applications (Allys et al. 2019; Regalado-Saint Blancard et al. 2020). These wavelets are defined as

$$\psi_m(\mathbf{x}) = \alpha (e^{i\mathbf{n}\cdot\mathbf{x}} - \beta) \cdot e^{-|\mathbf{x}|^2/(2\sigma^2)}, \quad (5)$$

where α ensures that the wavelets have a unit ℓ^2 -norm, $\beta = \exp(-\sigma^2/2)$ ensures a null average, and \mathbf{n} is a unit vector defining the oscillation direction of the mother wavelet (usually the x coordinate in a (x, y) plane). The Gaussian window, of size σ , localises the wavelet. The set of Morlet wavelets is then built using $Q = 1$ dyadic dilation and rotation, giving $\psi_{m,\lambda}$ wavelets. In this paper, we take $J_m = 6$ and $L_m = 4$ and rely on the Morlet wavelets defined in the `pywst`⁸ package.

These Morlet wavelets have a dyadic spatial frequency sampling defined on scales of 2^j , which could be coarse for EoR

⁸ <https://github.com/bregalado/pywst>

astrophysical parameter constraints. In addition, this sampling partially favours smaller scales with respect to larger ones, which could be a problem in the presence of thermal noise, which dominates at small scales. To better study the effect of the spatial frequency sampling, we also construct another set from the power spectrum sampling. An advantage of this wavelet choice is that it can also be more directly compared with power spectrum statistics as the spatial frequency sampling used by the wavelet statistics is the same.

Usual power spectrum binning for 21cm studies uses the ‘top hat’ spectral window function, whose spatial Fourier transform corresponds to a sinc function, with a poor spatial localisation. This explains our previous choice of a Gaussian window for the spatial frequency sampling, which is more spatially localised. As a second choice of wavelets, we thus use the inverse Fourier transform $\psi_{w,i} = \bar{W}_i(\mathbf{x})$ of the previous spectral windows $W_i(\mathbf{k})$ defined in Eq. (2). Contrary to the previous wavelet sets, these wavelets are isotropic and only have a scale parameter, which is called i here for consistency with previous power spectrum notations. In Fig. 3, we show the application of a window function, $I(\mathbf{x}) * \psi_{w,i}(\mathbf{x})$, on an example frequency channel in real space. These wavelets have $J_w = 9$, for a quality factor of $Q \approx 1.25$.

2.2.2. Wavelet moments

The first set of wavelet statistics we consider are WMs, which are integrated moments of the wavelet transforms (Eickenberg et al. 2022). We consider two moments, which are defined from the ℓ^1 - and ℓ^2 -norms of the wavelet convolutions⁹, yielding

$$M_1(\lambda) = \int_{\mathbb{R}^2} |I(\mathbf{x}) * \psi_\lambda(\mathbf{x})| d\mathbf{x}, \quad (7)$$

and

$$M_2(\lambda) = \int_{\mathbb{R}^2} |I(\mathbf{x}) * \psi_\lambda(\mathbf{x})|^2 d\mathbf{x}. \quad (8)$$

In this paper, these moments are constructed from the $\psi_{w,i}$ wavelets, and are therefore labelled $M_1(i)$ and $M_2(i)$.

An advantage of the $\psi_{w,i}$ choice of wavelets, which correspond to the inverse Fourier transform of the spectral window used for the power spectrum definition, is that thanks to Parseval’s identity, we directly have that $M_2(i) = P_i$. This illustrates the direct link between the choice of wavelets when building wavelet statistics and the frequency resolution achieved. A second advantage is that, as the WM is built from the M_1 and M_2 moments, they necessarily contain more Fisher information than the M_2 moments, and therefore more than the power spectrum. The M_1 moments bring additional information about the sparsity of the field. To decorrelate the M_1 moments from the M_2 moments by as much as possible, we normalise them as follows:

$$\bar{M}_1(i) = \frac{M_1(i)}{\sqrt{M_2(i)}}. \quad (9)$$

The final WM statistics are the concatenation of $\bar{M}_1(i)$ and $M_2(i)$, and thus have $2 * N_b$ coefficients.

⁹ The ℓ^p -norm $\|x\|_p$ of a vector x is defined as

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}. \quad (6)$$

2.2.3. Wavelet scattering transforms

The second set of wavelet statistics we consider is the WST Mallat (2011), Allys et al. (2019). These are constructed through a series of wavelet transforms and the application of the modulus operator, resulting in the generation of a collection of scattering coefficients. The coefficients are constructed layer by layer, and we consider only the coefficients of the first two layers, called $S_1(\lambda_1)$ and $S_2(\lambda_1, \lambda_2)$.

The first layer is constructed by convolving the 2D field $I(\mathbf{x})$ with a family of wavelets ψ_{λ_1} and applying a modulus non-linearity:

$$S_1(\lambda_1) = \frac{1}{\mu_1} \int |I * \psi_{\lambda_1}|(\mathbf{x}) d^2\mathbf{x}, \quad (10)$$

where μ_1 is a normalisation factor¹⁰. These coefficients are sometimes normalised by the mean of the field, but we did not do that here, because it is not suitable when the mean of the field is not a relevant quantity from a physical point of view.

The second layer is constructed by convolving the field again with another family of wavelets ψ_{λ_2} and applying another modulus non-linearity:

$$S_2(\lambda_1, \lambda_2) = \frac{1}{\mu_2} \int \||I * \psi_{\lambda_1}| * \psi_{\lambda_2}(\mathbf{x})\| d^2\mathbf{x}, \quad (12)$$

where μ_2 is a normalisation factor. To take into account the variability of S_2 due to the amplitude of the first wavelet convolution, we follow the usual normalisation by the first layer:

$$\bar{S}_2(\lambda_1, \lambda_2) = \frac{S_2(\lambda_1, \lambda_2)}{S_1(\lambda_1)}, \quad (13)$$

where the scale λ_2 characterised by the second wavelet should be larger than the scale λ_1 characterised by the first wavelet.

In this study, we computed the WSTs from the two sets of wavelets defined above: the directional dyadic Morlet wavelets, $\psi_{m,\lambda}$, as well as those constructed from the isotropic spectral windows used for the power spectrum, $\psi_{w,i}$, and we call these WST_m and WST_w , respectively.

For the WST_m statistics, for which coefficients are labelled by (j, θ) indices, we carry out an additional average over the direction to compute completely isotropic features. Following the results obtained with the RWST (Allys et al. 2019), we chose to do this average on the logarithm of the coefficients. This leads us to construct

$$S_1^{\text{iso}} = \left\langle \log_2 \left(S_1(j_1, \theta_1) \right) \right\rangle_{\theta_1}, \quad (14)$$

and

$$\bar{S}_2^{\text{iso}} = \left\langle \log_2 \left(\bar{S}_2(j_1, \theta_1, j_2, \theta_2) \right) \right\rangle_{\theta_1, \theta_2}. \quad (15)$$

We numerically verified that taking a logarithm before performing the angular averaging leads to better and more stable Fisher estimates. Our S_1^{iso} and $\bar{S}_2^{\text{iso},1}$ are therefore, up to \log_2 , similar to the S_1 and S_2 definitions in Greig et al. (2022).

¹⁰ It is defined as the impulse response:

$$\mu_1 = \int |\delta_d * \psi_{j_1, \theta_1}|(\mathbf{x}) d^2\mathbf{x}, \quad (11)$$

where δ_d is the dirac delta function. The following μ_2 normalisation factor is defined similarly.

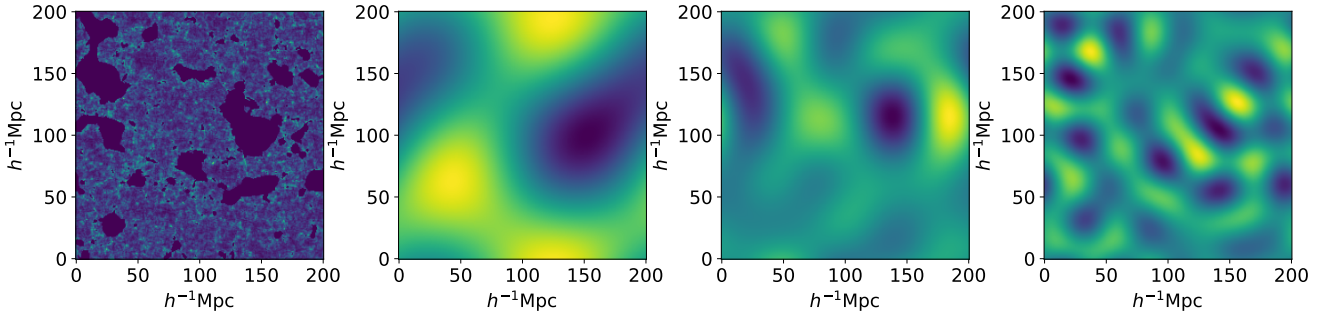


Fig. 3. Product of convolving a given frequency channel of a lightcone with our wavelet, $I(\mathbf{x}) * \bar{W}_i(\mathbf{x})$, which is the inverse Fourier transform of our Gaussian window function. Each wavelet is given by the inverse Fourier of a single concentric circle in Fig. 1. The left-most image is of the original frequency channel with no window function applied. The rest, from left to right, are from window functions centred on: $0.113 h^{-1} \text{ Mpc}$, $0.196 h^{-1} \text{ Mpc}$, and $0.342 h^{-1} \text{ Mpc}$.

Table 1. Change in number of coefficients before and after LoS summary.

Statistic	No. coefficients Before LoS summary	No. coefficients After LoS summary (maximum)
$\phi_{\text{PS}}(z)$	1152	72
$\phi_{\text{WM}}(z)$	2304	144
$\phi_{\text{WST}_m}(z)$	2688	168
$\phi_{\text{WST}_w}(z)$	5760	360

Notes. As an example, for the LoS summary, we have assumed the maximum number of scales, $j_z = 7$.

For the WST_w , no angular averaging was necessary because they are constructed from isotropic wavelets, which do not have an angular dependency. For these second statistics, we did not take the logarithm for these statistics, because it slightly worsened the conditioning of the statistic obtained through the Fisher study.

In what follows, WST_m refers to the concatenation of S_1^{iso} and \bar{S}_2^{iso} statistics for the ψ_m wavelets, while WST_w refers to the concatenation of S_1 and \bar{S}_2 for the ψ_w wavelets. For each set of statistics, we thus have $J S_1$ coefficients and $J \cdot (J - 1)/2 S_2$ coefficients. This leads to 21 coefficients for WST_m and 45 coefficients for WST_w , as $J_m = 6$ and $J_w = 9$.

2.3. Summary of evolution

In the previous section, except for the 3D power spectrum, all statistics are 2D and can be estimated on fixed- z spatial slices. In this work, we want to include the evolution of our 2D coefficients along the lightcone. This means that these 2D statistics have to be estimated for each frequency channel, giving spatial 2D statistics with a z dependency $\phi^s(z)$, which we call 2+1 statistics, where the s labels the set of statistics (PS, WM, WST_m , WST_w).

As we work with lightcones with 128 frequency channels (see Sect. 3), the total number of coefficients for our 2+1 statistics is very high, as shown in Table 1. We thus chose to compress this evolution along the line of sight (LoS). To do this, we considered the application of a continuous wavelet transform¹¹ to

the evolution of the coefficient along the lightcone via the following cosine¹² wavelet:

$$\psi_{j_z}(t) = e^{-\frac{t^2}{2^{2j_z}}} \cos\left(\frac{5t}{2^{j_z}}\right). \quad (16)$$

As these wavelets are constructed from a dyadic scaling, the maximum number of j_z scales on which we can decompose these evolutions is $\log_2(128) = 7$.

Once we apply the cosine wavelets to the redshift evolution of each coefficient, we summarise their evolution by computing integrated ℓ^1 - and square ℓ^2 -norm:

$$\bar{\phi}_{1;j_z}^s = \|\phi^s(z) * \psi_{j_z}(z)\|_1, \quad \text{and} \quad \bar{\phi}_{2;j_z}^s = \|\phi^s(z) * \psi_{j_z}(z)\|_2^2. \quad (17)$$

If we also include the mean of a coefficient's evolution, in addition to the maximum number of scale $j_z = 7$ scales, we have reduced the number of coefficients by a factor 16; see Table 1. This reduction in coefficients makes our resultant statistic far more stable. We note that this choice of wavelet compression arises from testing various methods, such as cosine decomposition and principal component analysis (PCA), to name just two, and we find the cosine wavelet to be the most stable and interpretable.

The ℓ^1 -norm and ℓ^2 -norm capture different evolution aspects of a given coefficient. Typically, the ℓ^2 -norm captures the energy or magnitude of the coefficients, while the ℓ^1 -norm captures sparsity and the presence of localised structures (Bach et al. 2011). In the case where we summarise the evolution with the ℓ^1 -norm and square ℓ^2 -norm, one incorporates both the energy and sparsity characteristics of the coefficients, which leads to a more comprehensive representation of the underlying signal. This combination should provide an informative summary of the evolution across the lightcone. We refer to our 'evolution-compressed' statistics as the evolution-compressed power spectrum (ecPS); the evolution-compressed wavelet moments (ecWM); the evolution-compressed WST_m (ec WST_m), and the evolution-compressed WST_w (ec WST_w).

We note that in the following, we concatenate only part of these evolution compressed coefficients, for which we use shorthand notation. For instance, $\bar{\phi}_{\ell^1, \ell^2; 1, 2}^{\text{WM}}$ stands for the ec-WM computed at $j_z = 1$ and $j_z = 2$ scales for both ℓ^1 and ℓ^2 norms.

¹¹ <https://pyWavelets.readthedocs.io/en/latest/ref/cwt.html>

¹² In pyWavelets this is referred to as (real) Morlet wavelets.

3. Simulation setup

3.1. 21CMFAST

The 21CMFAST (Mesinger et al. 2011) code is a widely available semi-analytical tool for simulating the 21cm signal during the EoR¹³; it offers a computationally efficient alternative to hydrodynamic simulations by employing approximations for various physical processes. The simulation creates a high-resolution density field at a redshift of $z = 300$. This field is then evolved to later redshifts using second-order Lagrangian perturbation theory (Scoccimarro 1998), which informs the spatial variations of the galaxy field through the conditional halo mass function. Dark matter halo fields are assigned galaxy properties through empirical scaling relations in accordance with the parametrisation described in Park et al. (2019). 21CMFAST employs an excursion-set-based formalism to identify ionised regions within the density field. This method involves considering spheres of decreasing radius, starting from a maximum radius R_{\max} , and determining if the number of ionising photons within the sphere exceeds the number of baryons. According to this formalism, the central pixel in a region is considered ionised if the following condition is satisfied:

$$\zeta f_{\text{coll}} \geq 1, \quad (18)$$

where f_{coll} is the fraction of collapsed matter within the spherical region under consideration. The maximum radius R_{\max} represents the farthest distance a photon can travel within the simulated field before encountering a recombined atom. It is loosely related to the characteristic mean free path. The ionising efficiency of galaxies ζ , is defined as:

$$\zeta = 30 \left(\frac{f_{\text{esc}}}{0.3} \right) \left(\frac{f_{\star}}{0.05} \right) \left(\frac{N_{\gamma}}{4000} \right) \left(\frac{2}{1 + n_{\text{rec}}} \right), \quad (19)$$

where f_{esc} represents the proportion of ionising photons that can escape into the intergalactic medium, f_{\star} denotes the fraction of gas in a galaxy that is converted into stars, N_{γ} stands for the quantity of ionising photons generated per baryon within stars, and n_{rec} indicates the typical number of times a hydrogen atom undergoes recombination.

The 21cm brightness temperature map is calculated from the ionisation field using the following expression:

$$\delta T_b = 28(1 + \delta)x_{\text{HI}} \left(1 - \frac{T_{\text{CMB}}}{T_s} \right) \left(\frac{\Omega_b h^2}{0.0223} \right) \times \sqrt{\left(\frac{0.24}{\Omega_m} \right) \left(\frac{1+z}{10} \right) \left[\frac{H}{\delta_r v_r + H} \right]}, \quad (20)$$

where x_{HI} denotes the neutral hydrogen fraction, δ represents the matter overdensity, $\delta_r v_r$ corresponds to the gradient of the line-of-sight peculiar velocity, $H(z)$ represents the Hubble parameter, and T_{CMB} and T_s are the CMB and spin temperatures, respectively. The term $\delta_r v_r$ accounts for the impact of redshift space distortions, primarily affecting small-scale overdensities. For further details of the simulations, we refer the reader to Park et al. (2019) and Murray et al. (2020).

3.2. SKA noise

In order to accurately simulate the thermal noise of the SKA, we use the PSTOOLS¹⁴ package. Our simulation methodology

¹³ <https://21cmfast.readthedocs.io/en/latest/>

¹⁴ https://gitlab.com/flomertens/ps_eor/-/wikis/Radio-interferometer-sensitivity

involves simulating the uv -coverage of SKA-Low. This means we simulate the uv tracks for each individual baseline and then grid them into a uv grid. The grid for uv -coverage is simulated based on the proposed distribution plan¹⁵ for antennae in SKA-Low.

First, we must create a realistic 10 h observation of the EoR-0 field – a proposed SKA field located at an RA of 0.00 h and Dec = -27 deg. To do this, we simulated a series of uv -tracks corresponding to this observation, with an integration time t_{int} set at 10 s. The noise within each uv -cell, and across each frequency channel ν , was generated based on a Gaussian distribution. We calculated the standard deviation of this distribution using the formula:

$$\sigma(u, v, \nu) = \frac{k_B T_{\text{sys}}(\nu)}{A_{\text{eff}}} \sqrt{\frac{1}{2\delta_{\nu} N(u, v, \nu) t_{\text{int}}}}. \quad (21)$$

Here, k_B is the Boltzmann constant, $T_{\text{sys}}(\nu)$ is the system temperature at frequency ν , A_{eff} represents the effective area, $N(u, v, \nu)$ denotes the number of baselines, δ_{ν} signifies the frequency resolution, and t_{int} stands for the integration time.

This generates a *.h5* which can be converted to images using the PSTOOLS function `simu_noise_img` – this function can also be used to scale the number of observational hours from 10 to 100, as we have done in this work.

3.3. Fisher analysis

The Fisher matrix, introduced by Fisher in 1922 (Fisher 1922), is a powerful tool for estimating the accuracy with which a statistic can constrain a parameter. Mathematically, the Fisher matrix is defined as follows:

$$\mathbf{F}_{ij} = \left\langle \frac{\partial^2 \ln \ell}{\partial \theta_i \partial \theta_j} \right\rangle, \quad (22)$$

where ℓ represents the likelihood¹⁶ function; θ_i denotes the parameter being varied, and $\langle \rangle$ represents the ensemble average, giving the expectation value. This matrix quantifies the sensitivity of the statistic to changes in the parameter. Under the assumption that the likelihood is a multivariate Gaussian, and thus the covariance matrix and mean vectors are sufficient information, we can further simplify the expression to (Tegmark et al. 1997; Carron 2013):

$$F_{ij}^{\theta} = \frac{\partial \mathbf{S}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_j}, \quad (23)$$

where \mathbf{S} represents a vector encompassing the expected values of the statistics used, such as scattering coefficients or power spectrum k -bins, and $\boldsymbol{\Sigma}$ corresponds to the covariance of these statistics, which arises from cosmic variance or the thermal noise in high-noise cases. The covariance is estimated by conducting multiple independent simulation realisations at fixed parameter values, often chosen as fiducial values. To calculate the derivative, we introduced a small perturbation to the parameter θ_i around its fiducial value and performed multiple realisations with this perturbed parameter value. According to the Cramer-Rao theorem, the variance of an unbiased estimator for a given

¹⁵ https://www.skao.int/sites/default/files/documents/d18-SKA-TEL-SKO-0000422_02_SKA1_LowConfigurationCoordinates-1.pdf

¹⁶ Likelihood of the target signal given the model parameters and the sources of stochasticity (cosmic variance and thermal noise).

parameter θ_i satisfies the following inequality:

$$\delta^2\theta_i \geq (F^{-1})_{ii}. \quad (24)$$

This Cramer-Rao bound establishes a lower bound on such a variance, indicating the smallest uncertainty achievable for an unbiased estimate of the parameter θ_i .

3.3.1. Fisher setup

The simulated lightcone has a transverse extent of $200 h^{-1}$ Mpc and consists of 256 pixels per side in each frequency channel. The lightcone spans a redshift range from $z = 8.82$ (144.60 MHz) to $z = 9.33$ (137.46 MHz), comprising 128 frequency channels.

The three parameters we considered in our Fisher analysis are: T_{vir} , which is minimum virial temperature needed for halos to host star-forming galaxies, as well as R_{Max} and ζ , as defined in the previous section. For each parameter, we varied them about the fiducial value:

1. T_{vir} : $50\,000 \pm 5000$
2. R_{Max} : 15 ± 5 Mpc
3. ζ : 30 ± 5 ,

and ran 400 simulations for each parameter value; we then used these for the derivatives. For the covariance, we used 600 realisation simulations with fiducial astrophysical parameters.

3.3.2. Fisher validation: Condition number

When using the inverse of the covariance matrix, denoted Σ^{-1} in Eq. (23), it is important to ensure that the covariance matrix is well conditioned for numerical stability during the inversion process. The condition number¹⁷ of the covariance matrix should ideally be of the order of or below 10^7 to be considered well conditioned (Park et al. 2023). If the covariance matrix is ill-conditioned, the inversion process can become unstable, leading to unreliable results in the Fisher formalism.

In Sect. 2.3, we discussed different applications of Eq. (6), which involves the use of either $p = 2$ (Euclidean norm) or a combination of $p = 1$ and $p = 2$. Though statistically limited to $j_z = 7$, because of the number of frequency channels, we must ensure that we are well-conditioned in the application of our statistics. For the ℓ^2 -norm, we find that we can use up to $j_z = 4$ before encountering issues with conditioning. However, for the combination of the ℓ^1 -norm and ℓ^2 -norm, we can only use up to $j_z = 2$ before encountering ill-conditioning.

We show a full description, including the number of coefficients, of all our statistics in Table 2. To avoid confusion between the different summaries, our evolution-compressed (ec) statistics –ecPS, ecWM, ecWST_m, and ecWST_w– are referred to here using their notation shown in Table 2.

Combining the first and second moments is crucial in order to accurately determine astrophysical parameters (Allys et al. 2019; Greig et al. 2022). However, in our analysis of the statistics, combining $M_1(i)$ and $M_2(i)$ causes the covariance matrix to be poorly conditioned due to their values being of different orders of magnitude. In order to overcome this issue, we whitened the statistics by normalising them to their standard deviation. The condition numbers of the whitened statistics are presented in Table 3. This normalisation balances the eigenvalues in the covariance matrix, making it better conditioned and

¹⁷ The condition number is a numerical measure used to assess how sensitive the output of a mathematical function or operation is to small changes in the input data. It is computed by dividing the largest eigenvalue of the covariance matrix by the smallest eigenvalue.

Table 2. Description of statistics.

Label	Details	No. terms
ϕ_{3D}^{PS}	Spherically-averaged 3D PS	9
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{PS}}$	Statistic:2D PS (ecPS) Summary: ℓ^2 -norm Scales: $j_z = 1, 2, 3, 4$	9×5
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{PS}}$	Statistic:2D PS (ecPS) Summary: ℓ^1 -norm + ℓ^2 -norm Scales: $j_z = 1, 2$	9×5
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WM}}$	Statistic:2D WM (ecWM) Summary: ℓ^2 -norm Scales: $j_z = 1, 2, 3, 4$	$(9 + 9) \times 5$
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WM}}$	Statistic:2D WM (ecWM) Summary: ℓ^1 -norm + ℓ^2 -norm Scales: $j_z = 1, 2$	$(9 + 9) \times 5$
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WST}_m}$	Statistic:2D WST _m (ecWST _m) Summary: ℓ^2 -norm Scales: $j_z = 1, 2, 3, 4$	21×5
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WST}_m}$	Statistic:2D WST _m (ecWST _m) Summary: ℓ^1 -norm + ℓ^2 -norm Scales: $j_z = 1, 2$	21×5
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WST}_w}$	Statistic:2D WST _w (ecWST _w) Summary: ℓ^2 -norm Scales: $j_z = 1, 2, 3, 4$	45×5
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WST}_w}$	Statistic:2D WST _w (ecWST _w) Summary: ℓ^1 -norm + ℓ^2 -norm Scales: $j_z = 1, 2$	45×5

Notes. Here we describe each statistic used in this work, including the number of coefficients that the application of the statistic will produce.

leading to improved numerical stability. We applied this whitening technique to all our statistics, which did not affect the results of the Fisher analysis. This also proves that our covariances are stable, as applying whitening to an ill-conditioned matrix would change the results of the Fisher analysis.

In cases where there is no noise, we do not encounter any problems. We can observe that the noiseless statistics are well conditioned by referring to Table 3, which displays the condition numbers of our various whitened statistics. However, when we introduce SKA noise into the data, we notice that the condition number for the noisy cases increases. This effect is not seen with the spherically averaged power spectrum; only with the 2+1 statistics. We find that understanding the results for 2+1 statistics is more difficult. Whether spatial statistics at various redshifts would exhibit more robust correlations concerning the EoR signal or the SKA noise remains to be seen. It is difficult to anticipate how the compression of z -information would impact the conditioning of the covariance matrix.

To ensure the numerical stability of the covariance, we performed an additional check by multiplying the inverse of the covariance by the covariance in order to verify that it is the identity matrix. Our findings reveal that the error on the off-diagonal terms – which indicates how far they stray from zero – is approximately $\sim 10^{-8}$. Given that the condition numbers are

Table 3. Condition number of the statistics.

Statistics (Results are \log_{10})	No noise	100 hours	1000 hours
ϕ_{3D}^{PS}	4.46	4.33	4.28
$\bar{\phi}_{\ell^2:1,2,3,4}^{PS}$	5.03	6.52	6.21
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{PS}$	5.08	6.49	6.17
$\bar{\phi}_{\ell^2:1,2,3,4}^{WM}$	6.17	8.10	7.62
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{WM}$	5.47	8.04	7.6
$\bar{\phi}_{\ell^2:1,2,3,4}^{WST_m}$	6.14	9.03	8.9
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{WST_m}$	6.12	8.98	8.86
$\bar{\phi}_{\ell^2:1,2,3,4}^{WST_w}$	6.19	7.66	6.92
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{WST_w}$	6.21	7.56	6.92

Notes. To ensure the robustness of our statistical analysis, it is important to assess the condition numbers of the various statistics – which have all been whitened – employed in this paper under different noise treatments. A condition number below or of the order of 10^7 is generally considered indicative of a well-conditioned covariance matrix.

Table 4. Numerical stability of each statistic.

Statistics (Results are \log_{10})	No noise	100 hours	1000 hours
ϕ_{3D}^{PS}	-12.25	-12.20	-12.51
$\bar{\phi}_{\ell^2:1,2,3,4}^{PS}$	-11.87	-9.24	-10.14
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{PS}$	-11.57	-9.13	-9.11
$\bar{\phi}_{\ell^2:1,2,3,4}^{WM}$	-10.66	-8.92	-7.97
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{WM}$	-11.20	-8.64	-9.14
$\bar{\phi}_{\ell^2:1,2,3,4}^{WST_m}$	-10.28	-7.17	-10.03
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{WST_m}$	-10.08	-7.73	-7.38
$\bar{\phi}_{\ell^2:1,2,3,4}^{WST_w}$	-10.23	-9.16	-9.70
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{WST_w}$	-10.40	-8.88	-9.80

Notes. Looking at the numerical stability of the different statistics, we multiply the inverse of the covariance by the covariance, which should be an identity matrix. We show the largest off-diagonal term, as a test for numerical stability.

of the order of $\sim 10^8$, we would expect the error to be around 10^{-8} – this is because double-precision computation allows for 16 digit representation – signifying no significant instabilities. Despite the high condition number, the covariance matrices for high-noise cases are numerically stable. We present the errors on the off-diagonal terms for all the summary statistics in Table 4.

The Fisher formalism is a valuable tool for estimating the information on parameters that is contained in a statistic. However, its effectiveness can be influenced by the difficulty in estimating ill-conditioned covariance matrices and derivatives. An alternative treatment would involve sampling the posterior distribution using a likelihood surrogate, which offers enhanced numerical stability and reliability in parameter estimations, making it a useful option. We use the Fisher matrix approach here for simplicity, but in future works, we will use more sophisticated approaches.

3.3.3. Fisher validation: Convergence

Another important check in the Fisher analysis is to ensure that our Fisher matrix is fully converged. We can check our convergence by comparing the diagonals of the Fisher matrices with different numbers of the simulation realisations (or samples) used to calculate them:

$$\text{Convergence} = \frac{F_{ii}(N_{\text{Samples}})}{F_{ii}(N_{\text{Max}})}. \quad (25)$$

We have a large number of samples with which to check this convergence. We performed two checks. In the first, we kept the number of samples used to calculate the derivatives constant at 400 and varied the number of fiducial samples to calculate the covariance. In the second, we kept the number of fiducial samples to calculate the covariance constant at 600 samples and only varied the number of samples used for the different astrophysical parameter derivatives.

We show all of our convergence plots in Appendix A. When looking for convergence, we look to have less than a 10% error in our convergence check – which we indicate here via the red-shaded region. We look to have our convergences stably inside this region, that is for at least approximately 100 samples¹⁸. We find that all of our Fisher matrices – that is, for all statistics and noise cases – are converged.

4. Results

We then carried out a Fisher analysis on our different statistics. We show corner plots comparing the best applications of our different statistics to the different noise cases of our data data, including ϕ_{3D}^{PS} as is traditionally used in 21cm parameter inference. We show all of the statistics in the Cramer-Rao bounds of this work for each parameter in this study in tables. Greig et al. (2022) previously compared wavelet scattering transforms to the power spectrum. In their work, the authors simulated a 21CMFAST lightcone within the redshift range of $5.9 \leq z \leq 27.4$. They broke the lightcone into 12 chunks, applying the spherically averaged power spectrum to each chunk and applied the 2D WST to the central frequency slice of each light-cone chunk.

In this work, we looked to simulate our lightcones in a single redshift band and at both the SKA angular resolution and frequency resolution. In addition to the ϕ_{3D}^{PS} statistic, we also used Eq. (16) to summarise the line-of-sight evolution of our different 2D statistics. This line-of-sight information can further constrain parameters.

4.1. Noiseless

From Fig. 4 and Table 5, we see that overall, $\bar{\phi}_{\ell^2:1,2,3,4}^{WM}$ produces the tightest constraints on the astrophysical parameters in our contour plots, closely followed by $\bar{\phi}_{\ell^2:1,2,3,4}^{WST_w}$ and $\bar{\phi}_{\ell^1, \ell^2:1,2}^{WST_w}$. In their analysis of 2D WST versus the 3D power spectrum, Greig et al. (2022) found that the WST produced slightly more precise constraints. Comparing Table 5, we see that all the evolution-compressed statistics outperform the 3D power spectrum. Additionally, our wavelet-based statistics produce the tightest Cramer-Rao bounds. This is expected, as they combine

¹⁸ If we used 399 samples out of 400 it would be within this 10% region. Once within the 10% region, we can only be sure of convergence if we remain within this region as we increase the number of samples. Hence by seeing if the Fisher matrix in this region for a greater number of samples used, we can say that it has converged.

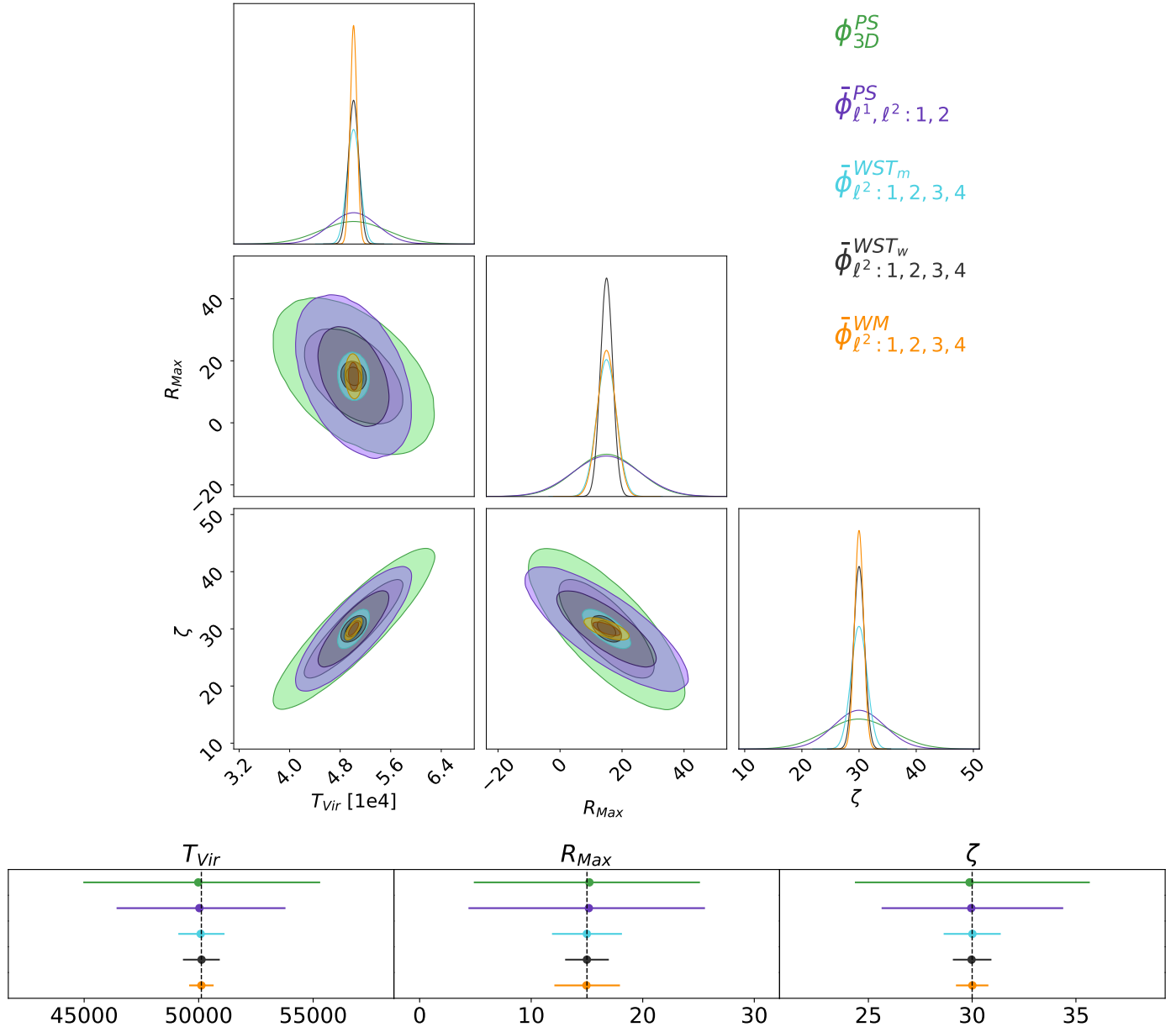


Fig. 4. Results from the Fisher analysis of the three different summary statistics of the 21cm signal, when three astrophysical parameter are varied, as we have noiseless data, considering cosmic variance as the only source of variance. *Top:* corner plot of our noiseless Fisher analysis, showing that $\bar{\phi}_{\ell^2: 1, 2, 3, 4}^{WM}$ provides the tightest contours. *Bottom:* the $\pm 68\%$ credibility intervals of our different astrophysical parameters, for each statistic. The ordering of the statistics is based on their performance, going from least constraining statistic (top) to most constraining (bottom).

the richer 2D information, such as that in the scattering transform as seen in Greig et al. (2022), while explicitly decomposing the line-of-sight information.

The evolution of the power spectrum along the lightcone exhibits high sensitivity to parameter changes, with the large-scale modes of the power spectrum being the most sensitive. Performing spherical averaging results in a loss of sensitivity to this line-of-sight information, as we mix both perpendicular and line-of-sight information. It is worth noting, however, that these conclusions are made for a single redshift band.

For each of our 2+1 statistics, we performed two LoS decompositions: one summarises the evolution with the ℓ^2 -norm on scales $j_z = 1, 2, 3, 4$ and the other with the ℓ^1 -norm + ℓ^2 -norm on scales $j_z = 1, 2$. From Table 5, it is difficult to make direct comparisons between the two LoS decompositions given that they operate on different scales. However, it should be noted that ℓ^2 -

norm on scales $j_z = 1, 2, 3, 4$ is expected to perform better than ℓ^2 -norm on scales $j_z = 1, 2$. The similarity in the performance of the two decompositions shown in Table 5 suggests that ℓ^1 -norm + ℓ^2 -norm is more informative than ℓ^2 -norm, given that it takes two scales rather than four to produce comparable results. We are only restricting ourselves to small scales in the redshift domain – the ℓ^1 -norm with ℓ^2 -norm only look at scales of size 2^1 and 2^2 – rather than looking for large-scale information on the evolution. As mentioned above, we tried to include this large-scale information, but when we do, the condition number of our covariance matrix becomes too high.

The contours in Fig. 4 show that the astrophysical parameter constraints derived from the wavelet-based statistics and power spectra present similar degeneracies. Specifically, the S_1^{iso} information alone from the WST_m , and the first layer of WST_w , are akin to that obtained from the power spectrum (Allys et al. 2019;

Table 5. No noise case, Cramer-Rao bounds.

Statistics (Results are \log_{10})	T_{Vir}	R_{Max}	ζ
$\phi_{3\text{D}}^{\text{PS}}$	7.42	2.01	1.50
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{PS}}$	7.25	2.05	1.37
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{PS}}$	7.13	2.04	1.28
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WM}}$	5.45	0.93	-0.22
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WM}}$	6.25	1.00	0.43
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WST}_m}$	6.00	0.98	0.27
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WST}_m}$	5.99	1.01	0.30
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WST}_w}$	5.81	0.58	-0.07
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WST}_w}$	5.79	0.60	-0.05

Notes. Cramer-Rao bounds for all of our summary statistics in the case where we have no noise. The bound establishes a lower bound on the variance, i.e. the smallest uncertainty achievable for an unbiased estimate on a given parameter.

Table 6. 100 h of SKA noise case, Cramer-Rao bounds.

Statistics (Results are \log_{10})	T_{Vir}	R_{Max}	ζ
$\phi_{3\text{D}}^{\text{PS}}$	9.22	3.60	2.97
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{PS}}$	8.90	3.52	2.80
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{PS}}$	8.68	3.24	2.56
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WM}}$	8.16	2.84	2.27
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WM}}$	8.09	2.76	2.20
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WST}_m}$	9.31	3.88	3.21
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WST}_m}$	9.22	3.78	3.12
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WST}_w}$	7.31	1.73	1.29
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WST}_w}$	7.28	1.71	1.27

Cheng et al. 2020). However, applying WST_m (and WST_w) here introduces an additional coefficient¹⁹ from $S_2^{\text{iso}1}$, which helps further constrain the parameters. Notably, although we still observe comparable degeneracies with the power spectrum, the degeneracy between R_{Max} and T_{vir} is somewhat alleviated.

4.2. One hundred hours of SKA noise

We now look to see how robust these statistics are to thermal noise by comparing how well they constrain these parameters with the inclusion of two levels of noise. We start by looking at 100 h of SKA observation level noise. We show the Cramer-Rao bounds of each statistic and for each parameter in Table 6, and show the same subset of each statistic in Fig. 5, as shown in Fig. 4.

In the noiseless case, we see similar results between summarising the evolution of the statistics with either the ℓ^2 -norm alone or a combination of the ℓ^1 -norm and ℓ^2 -norm. Now we see in Table 6 that for all of evolving statistics, the two scales summarised with the ℓ^1 -norm and ℓ^2 -norm outperform the four scales summarised with the ℓ^2 -norm. When there is noise in our

¹⁹ Though we restricted ourselves to the isotropic RWST components, there also exist anisotropic components.

data, the noise can cause extra variations and alter the magnitude characteristics of the coefficients. This means that relying solely on the ℓ^2 -norm may be more dependable, as there are a significant number of scales not affected by the noise. The ℓ^1 -norm, on the other hand, is able to extract more information from a smaller subset of scales, which could be less affected by the noise.

From Fig. 5, we observe a notable contrast in the results compared to Figure 4. Specifically, with $\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WST}_w}$ producing the tightest constraints, $\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WST}_m}$ is now the least constraining statistic, with not only $\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{PS}}$ producing tighter constraints but also $\bar{\phi}_{3\text{D}}^{\text{PS}}$. The $\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{PS}}$ statistic produces, on average, 15% tighter parameter constraints compared to $\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WST}_m}$, and $\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WM}}$, on average, produces 40% tighter parameter constraints.

The power spectra are understood to outperform WST_m – with its current choice of wavelets – in high-noise situations due to its averaging effect over multiple Fourier modes during binning as well as the fact that it probes the scale resolutions of interest, which enhances the signal-to-noise ratio. The binning strategy used by the power spectrum is more sensitive to large-scale structures, whereas thermal noise typically affects the smallest scales. Consequently, even with averaging, the power spectrum primarily captures structural information where noise is less likely to dominate. In contrast, WST_m coefficients here are predominantly associated with smaller scales, making them more susceptible to noise. The abundance of coefficients for smaller scales amplifies the influence of noise, which dominates at those scales. We find the three largest k -bins of the power spectra, which are also used by the WM application, to be severely noise dominated. These are associated with the scales 2^1 and 2^2 for WST_m , which contain a total of 55 evolved coefficients, roughly half of the number of coefficients for WST_m .

For the WMs, we resolved this by leveraging the binning strategy of the power spectrum using the inverse Fourier transform of its binning window function as the wavelet of choice in order to exploit its spectral resolution to probe structural information where noise is unlikely to dominate. As expected, the WMs outperform the 2D power spectrum. We see that including the $M_1(i)$ in the WMs introduces a statistical measure that exhibits greater sensitivity to parameter changes than noise. We further demonstrate the importance of a more appropriate wavelet with the results of WST_w . For WST_w , we use a similar²⁰ scattering transform formalism but with the same wavelets as those used with the WMs. In this case, we benefit from the informative scattering transform on scales less affected by the noise, as now fewer of its components are in the less informative and noise-dominated bands, as they are with WST_m . For both LoS decompositions, WST_w provides the tightest constraints (see Table 6) and the Cramer-Rao bounds are orders of magnitude better than those of the other statistics.

4.3. One thousand hours of SKA noise

We now consider 1000 h of SKA noise, and show Cramer-Rao bounds of the different statistics for each parameter in Table 7. We show the best result of each statistic in Fig. 6, for 1000 h of SKA noise.

In this low-noise regime, ecWST_w still outperforms the other evolution-compressed statistics – with both wavelet-based techniques using the wavelet derived from the power spectrum binning scheme performing the best. We see that the results of

²⁰ As the wavelets used are isotropic, with no directional component.

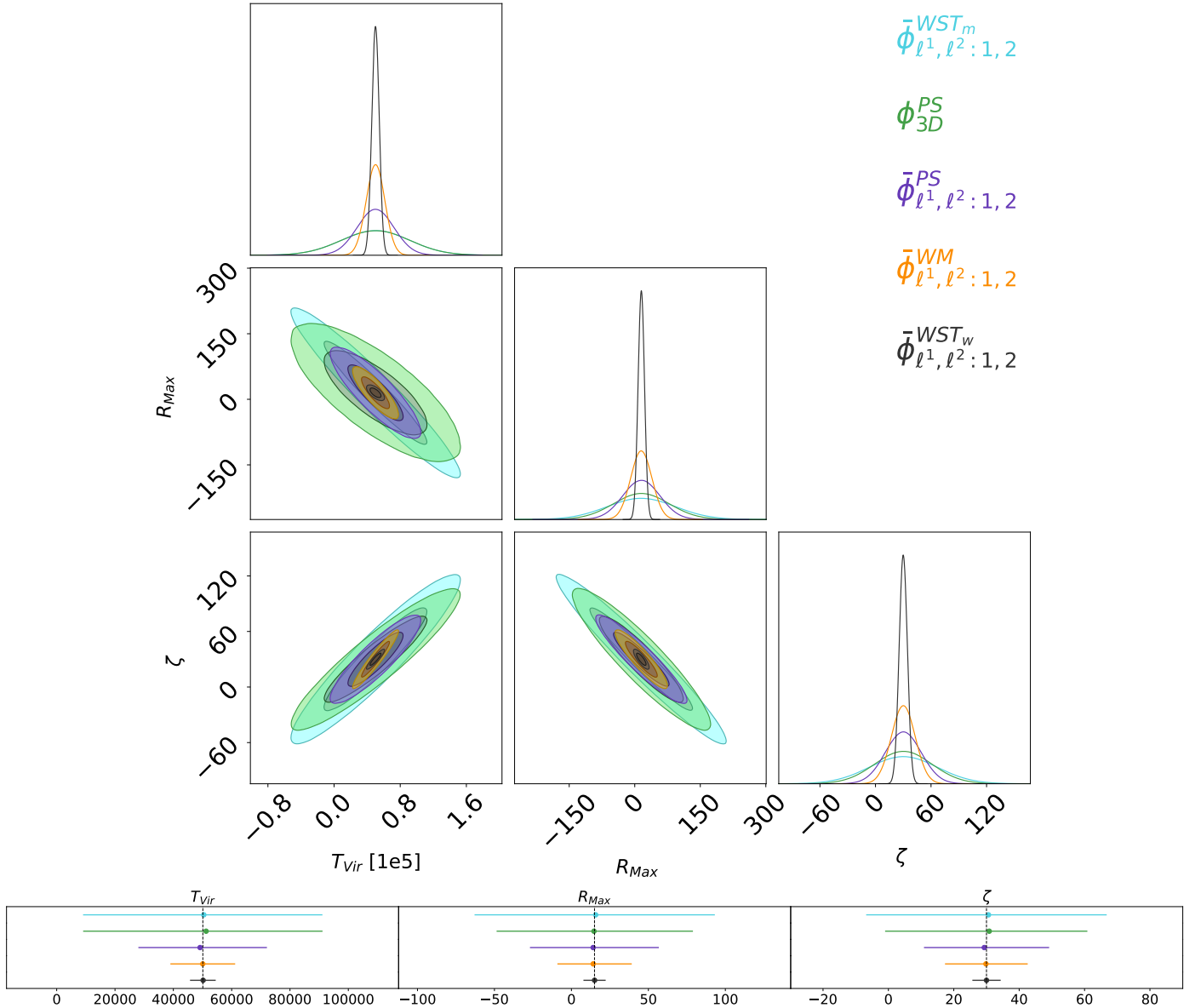


Fig. 5. As in Fig. 4, but for 100h of SKA noise, where the noise is the dominant source of variance. We see now that WST_m is the statistic with the worst performance, and that WST_w , with its evolution along the lightcone summarised by the ℓ^1 -norm and ℓ^2 -norm on scales $j_z = 1, 2$, which utilises wavelets derived from the power spectra binning, provides the tightest contours.

$ecWST_m$ have significantly improved in this lower-noise case. The precision difference between $ecWST_m$ and the other statistics has been significantly reduced in this case. In this lower-noise regime, the smaller scales are less impacted by the noise, allowing the coefficients of $ecWST_m$ to have a higher signal-to-noise ratio and a better statistical constraining power.

Whereas before the ℓ^1 and ℓ^2 combination produced the highest precision constraints for our evolution-compressed statistics, Table 7 shows that now, at least for the wavelet-based statistics, the ℓ^2 LoS summary performs at least as well as the ℓ^1 and ℓ^2 summary. We also see that the 2+1 evolution-compressed statistics still produce constraints of higher precision overall than ϕ_{3D}^{PS} . However, for R_{Max} , we see that the posterior for ϕ_{3D}^{PS} produces tighter constraints than that for $ecWST_m$ for both LoS decompositions. We also see, for the first time, a significant change in the degeneracies of our statistics. For $R_{Max}-T_{Vir}$ and $\zeta-T_{Vir}$, we see that the two evolution-compressed statistics have a different degeneracy compared to that of ϕ_{3D}^{PS} .

In real-world observations, there are no cases that are completely free of noise. We have indeed gained the advantages intended by choosing a set of wavelets more suited to the spectral domain of interest for our signal. In these noise cases, we are able to probe the important large scales whilst limiting the effect of noise and utilise the WMs to enhance our constraints.

5. Conclusions

In this paper, we present an approach to enhance the accuracy of astrophysical parameter constraints when analysing the 21cm signal from the epoch of reionisation (EoR). We focus on the use of 2+1 statistics, which include spectral evolution of the 2D spatial statistics. In addition to comparing different summary statistics such as the power spectrum, WMs, and the WSTs, we also explore the importance of the choice of wavelet basis to improve sensitivity to astrophysical parameters. For WM and WST_w , we aim to use the binning strategy usually used for power spectra

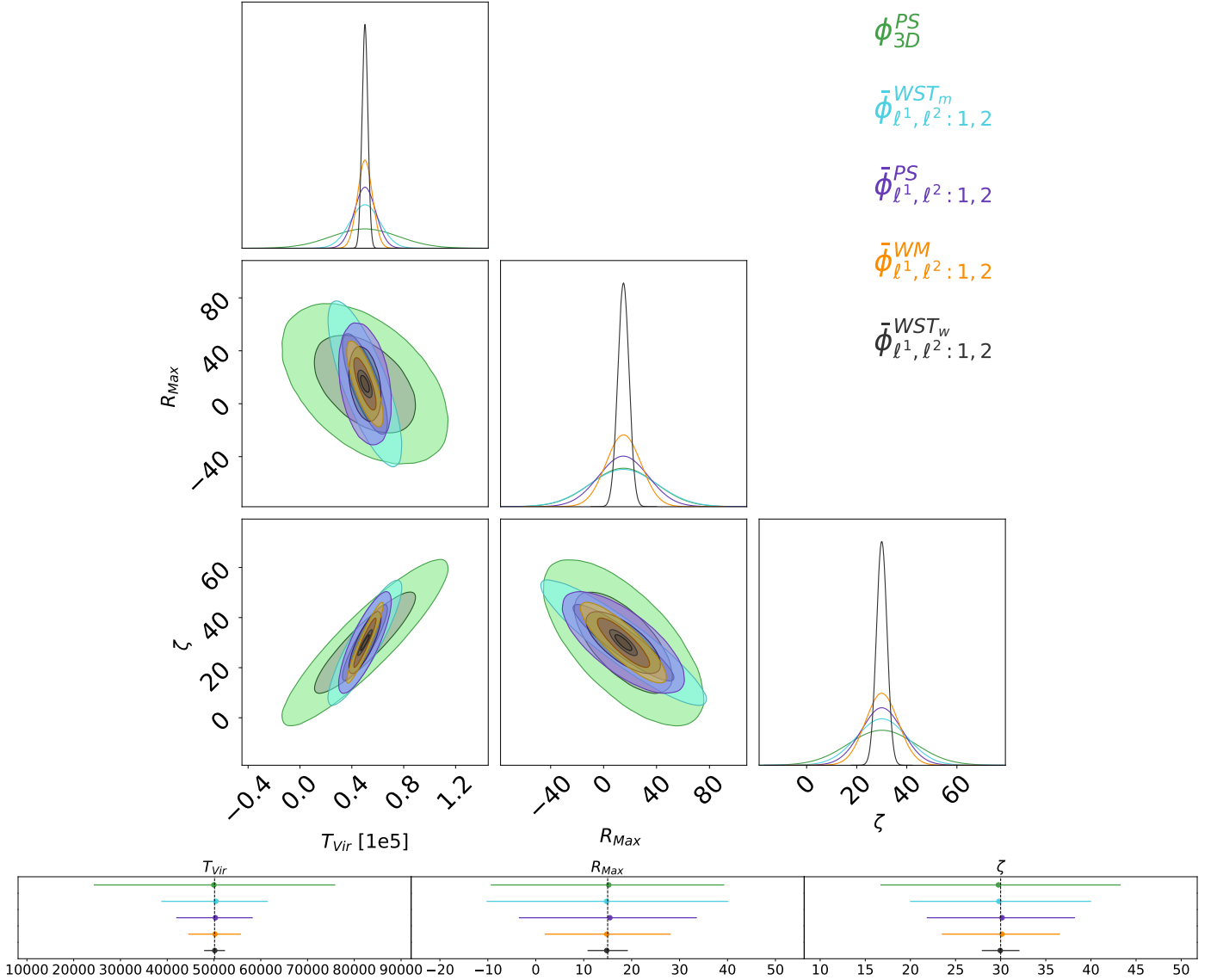


Fig. 6. Same as Fig. 5, but now we have 1000 h of SKA noise. WST_m is still the worst performing evolution-compressed statistic in the low-noise case, but it is now closer to the precision of the power spectrum. The WST_w statistic still provides the most precise constraints.

in EoR studies. This is expected to yield more accurate parameter estimations by choosing a frequency sampling relevant to the EoR signal while decreasing the contribution of noise to the summary statistics. By way of comparison, for WST_m we retained the dyadic wavelet sampling used in previous astrophysical studies. We investigated the performance of these statistics in both noiseless and noisy scenarios, with noise levels corresponding to 100 and 1000 h of observations with the SKA.

In the noiseless scenario, we discovered that all the 2+1 evolution-compressed statistics perform better than the spherically averaged power spectrum. The wavelet statistics outperform the ecPS, with ecWM – with the ℓ^2 LoS decomposition on redshift scales $j_z = 1, 2, 3, 4$, and both ec WST_w – providing the tightest limits of all the statistics. This shows that a simpler 2+1 statistical method can provide precise parameter constraints. These wavelet statistics can effectively identify non-Gaussian information beyond two-point statistics. WST_m combines S_1^{iso} and $S_2^{\text{iso},1}$, while WST_w combines the first and second layer – as also shown by Greig et al. (2022). For WM, combining the $M_2(i)$

coefficients – which are similar to the 2D power spectrum – with the $M_1(i)$ coefficients allows us to investigate sparsity.

In both cases of noise, that is, the high-noise case (100 h of SKA noise) and the low-noise case (1000 h of SKA noise), overall, the 2+1 evolution-compressed statistics, including the ecPS, outperform the spherically averaged power spectrum. This suggests that the 2+1 wavelet-based statistics proposed in this paper form an effective alternative to the direct introduction of 3D statistics.

We also see that ec WST_w far outperforms other statistics in these noisy regimes. This demonstrates that proper choices of wavelet and scale sampling are critical when building wavelet statistics. Indeed, the ec WST_m under-performed here in high-noise cases due to its improper sampling of the largest scales, which are less contaminated by noise. Interestingly, both LoS summaries of ec WST_w perform similarly, whereas the other statistics see a marked improvement in their constraints in high-noise cases by using the ℓ^1 and ℓ^2 combination for the LoS summary.

Table 7. 1000 h of SKA noise case, Cramer-Rao bounds.

Statistics (Results are \log_{10})	T_{Vir}	R_{Max}	ζ
ϕ_{3D}^{PS}	8.82	2.77	2.24
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{PS}}$	7.85	2.74	1.94
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{PS}}$	7.82	2.53	1.82
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WM}}$	7.56	2.36	1.70
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WM}}$	7.50	2.23	1.62
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WST}_m}$	8.10	2.80	2.00
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WST}_m}$	8.11	2.79	2.00
$\bar{\phi}_{\ell^2:1,2,3,4}^{\text{WST}_w}$	6.74	1.24	0.65
$\bar{\phi}_{\ell^1, \ell^2:1,2}^{\text{WST}_w}$	6.69	1.24	0.63

Real-world observations are subject to additional sources of nuisance and complexity. Various factors can affect the data, such as residual gain errors, imperfect excision of radio frequency interference (RFI), and other instrumental artefacts. Therefore, it is crucial to incorporate these elements into the analysis framework. This step is necessary to make astrophysical parameter estimation more robust and reflective of the challenges that arise in practical applications.

A word of caution is warranted regarding the limitations of the Fisher approach. The Fisher approach, which approximates the likelihood by a single multivariate Gaussian at the fiducial point, simplifies the parameter space into ellipses; these poorly represent the true shapes of real parameter posteriors, which exhibit complex, often non-Gaussian shapes, reflecting the degeneracies and correlations between astrophysical parameters. Recognising this limitation should drive future research towards more sophisticated statistical methodologies that can capture the full complexity of the parameter space (Park et al. 2023).

Finally, we restricted ourselves to the isotropic components of our data. In reality, the 21cm signal is highly non-Gaussian in nature, and the performance of the WST statistics would likely have been enhanced if we had included its anisotropic components (see Allys et al. 2019). The results of this paper highlight the potential of 2+1 statistics and wavelet-based statistics for future 21cm intensity mapping studies in cosmology. We have created a GitHub repository²¹ containing the scripts used to produce the statistics in this paper, and detailing how to reproduce the simulations used in this work.

Acknowledgements. The post-doctoral contract of Ian Hothi was funded by Sorbonne Université in the framework of the Initiative Physique des Infinis (IDEX SUPER). We would like to thank Florent Mertens and Andrea Bracco for their useful discussions and input. We also thank our referee Prof. Yuan-Sen Ting for their comments to make the paper more balanced and their discussion points. This research made use of astropy, a community-developed core Python package

for astronomy (Astropy Collaboration 2018); scipy, a Python-based ecosystem of open-source software for mathematics, science, and engineering (Virtanen et al. 2020) – including numpy (Walt 2011); matplotlib, a Python library for publication quality graphics (Hunter 2007).

References

- Allys, E., Levrier, F., Zhang, S., et al. 2019, *A&A*, **629**, A115
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, **156**, 123
- Bach, F., Jenatton, R., Mairal, J., & Obozinski, G. 2011, ArXiv e-prints [arXiv:1108.0775]
- Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, *MNRAS*, **447**, 3402
- Bosman, S. E. I., Fan, X., Jiang, L., et al. 2018, *MNRAS*, **479**, 1055
- Carron, J. 2013, *A&A*, **551**, A88
- Cheng, S., & Ménard, B. 2021, *MNRAS*, **507**, 1012
- Cheng, S., Ting, Y.-S., Ménard, B., & Bruna, J. 2020, *MNRAS*, **499**, 5902
- DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, *PASP*, **129**, 045001
- Eickenberg, M., Allys, E., Moradinezhad Dizgah, A., et al. 2022, ArXiv e-prints [arXiv:2204.07646]
- Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, *AJ*, **132**, 117
- Fisher, R. 1922, *Philos. Trans. R. Soc.*, **222**, 309
- Gehlot, B. 2019, PhD Thesis, University of Groningen, The Netherlands
- Gorce, A., & Pritchard, J. R. 2019, *MNRAS*, **489**, 1321
- Gorce, A., Douspis, M., Aghanim, N., & Langer, M. 2018, *A&A*, **616**, A113
- Greig, B., Ting, Y.-S., & Kurov, A. A. 2022, *MNRAS*, **513**, 1719
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- Hutter, A., Watkinson, C. A., Seiler, J., et al. 2020, *MNRAS*, **492**, 653
- Majumdar, S., Pritchard, J. R., Mondal, R., et al. 2017, *MNRAS*, **476**, 4007
- Mallat, S. 2011, ArXiv e-prints [arXiv:1101.2286]
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, *MNRAS*, **411**, 955
- Munshi, S., Mertens, F. G., Koopmans, L. V. E., et al. 2024, *A&A*, **681**, A62
- Murray, S., Greig, B., Mesinger, A., et al. 2020, *J. Open Source Software*, **5**, 2582
- Paciga, G., Chang, T.-C., Gupta, Y., et al. 2011, *MNRAS*, **413**, 1174
- Paciga, G., Albert, J. G., Bandura, K., et al. 2013, *MNRAS*, **433**, 639
- Park, J., Mesinger, A., Greig, B., & Gillet, N. 2019, *MNRAS*, **484**, 933
- Park, C. F., Allys, E., Villaescusa-Navarro, F., & Finkbeiner, D. 2023, *ApJ*, **946**, 107
- Planck Collaboration Int. XLVII. 2016, *A&A*, **596**, A108
- Pober, J. C., Parsons, A. R., DeBoer, D. R., et al. 2013, *AJ*, **145**, 65
- Pober, J. C., Liu, A., Dillon, J. S., et al. 2014, *ApJ*, **782**, 66
- Prasad, P., Huizinga, F., Kooistra, E., et al. 2016, *J. Astron. Instrum.*, **5**, 1641008
- Prelogović, D., & Mesinger, A. 2023, *MNRAS*, **524**, 4239
- Qin, Y., Mesinger, A., Bosman, S. E. I., & Viel, M. 2021, *MNRAS*, **506**, 2390
- Regaldo-Saint Blancard, B., Levrier, F., Allys, E., Bellomi, E., & Boulanger, F. 2020, *A&A*, **642**, A217
- Saydjari, A. K., Portillo, S. K. N., Slepian, Z., et al. 2021, *ApJ*, **910**, 122
- Scoccimarro, R. 1998, *MNRAS*, **299**, 1097
- Shimabukuro, H., Yoshiura, S., Takahashi, K., Yokoyama, S., & Ichiki, K. 2016, *MNRAS*, **468**, 1542
- Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, *ApJ*, **480**, 22
- Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, *PASA*, **30**, e007
- Valogiannis, G., & Dvorkin, C. 2022, *Phys. Rev. D*, **105**, 103534
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, *A&A*, **556**, A2
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, **17**, 261
- Walt, S., & v. d., Colbert, S. C., & Varoquaux, G., 2011, *Comput. Sci. Eng.*, **13**, 22
- Watkinson, C. A., Majumdar, S., & Pritchard, J. R. 2017, *MNRAS*, **472**, 2436
- Watkinson, C. A., Trott, C. M., & Hothi, I. 2021, *MNRAS*, **501**, 367
- Welling, M. 2005, in Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, eds. R. G. Cowell, & Z. Ghahramani, *Proc. Mach. Learn. Res.*, **R5**, 405
- Zarka, P., Girard, J. N., Tagger, M., Denis, L., et al. 2012, in SF2A-2012: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, eds. S. Boissier, P. de Laverny, N. Nardetto, et al., 687

²¹ https://github.com/ihothi/Wavelet_Stats

Appendix A: Fisher convergences

In this section, we analyse the convergence of different statistics using equation (25). We use the Fisher matrix as a measure of convergence, which is calculated using a varying number of simulations. Our objective is to determine when convergence is achieved when the Fisher matrix calculated using a certain number of simulations is identical to the Fisher matrix calculated using the maximum number of simulations. To this end, we plot the convergence ratio between the Fisher matrix calculated for a given number of simulations and the one calculated using the maximum number of simulations. This plot helps us to determine the convergence region, which is the range where the ratio is considered to have converged within a 10% error. We allow for a convergence error because complete convergence is challenging to achieve. This task can be easily misinterpreted. As we increase the number of simulations to calculate the Fisher, we are approaching ‘convergence’ of our Fisher matrix. For instance, if we vary the number of simulations used to calculate the derivatives, the ratio may appear within the convergence region at 250 simulations. One might assume that convergence has been achieved and stop there. However, at 255 simulations, the ratio may diverge from convergence. Therefore, when we perform such an analysis, we must ensure that the Fisher matrix is within the convergence zone for many simulations (used in calculations) after the supposed convergence. This approach helps us to confirm that we have truly reached convergence.

In Appendix A.1, we investigate the convergence of the number of fiducial simulations used to calculate the covariance for our fisher analysis. We keep the number of simulations used to calculate the derivatives fixed at 400. For each Fisher matrix calculated using a given number of simulations/samples, as shown in equation (25), we divided it by the Fisher matrix using the total number of fiducial simulations. This ratio is plotted as a function of the number of simulations/samples.

In Appendix A.2, we calculate the Fisher matrix, keeping the number of fiducial simulations fixed at 600 while varying the number of simulations or ‘samples’ used to calculate the derivatives. We aim to plot the Fisher as a function of the number of simulations or samples. In each of our plots, we look for the error—indicated with a red shaded region—to be below 10%.

A.1. Covariance convergence

In this test, we keep the number of simulations (or samples) used to calculate the derivatives constant at 400, and vary the

number of fiducial simulations used to calculate the covariance from 1 to 600. We can see from Fig. A.1a that most of our different statistics are fully converged by 300 samples. We see that our evolution-compressed statistics require at least 300 samples before converging, whilst the spherically averaged power spectrum is convergent after 200 samples. WST_w needs around 400 samples in the no-noise case before it becomes fully convergent. In Fig. A.1b, we show the convergence plots of our statistics in a high-noise scenario. Whereas the noiseless case will demonstrate the point of convergence with respect to cosmic variance, the noisy cases will probe convergence of our noise realisations. We see that most of the statistics reach convergence after 200 samples. The power spectrum statistics converge most readily, converging after 150 samples, whereas most of the wavelet-based statistics need a few hundred more samples to be convergent. WST_w requires close to 450 samples before it is fully convergent. The convergence plots for our statistics under 1000 hours of SKA noise are presented in Fig. A.1c. We observe a similar trend to that in Fig. A.1b, where noise has a significant impact and we need approximately 200 samples to achieve noise convergence. We can see that our wavelet-based statistics typically require more samples than the power-spectra-based statistics before they are fully convergent. This is likely due to the non-Gaussian information they contain, requiring more samples.

A.2. Derivative convergence

In this test, we keep the number of fiducial simulations used to calculate the covariance constant at 600, and vary the number of simulations used to calculate the derivatives from 1 to 400.

We can see from Fig. A.2a that all of our different statistics are fully converged by 400 samples, with T_{Vir} and ζ being the most readily convergent parameters, converging after only 50 or so samples. However, R_{Max} requires more samples before it converges towards the ‘target range’—though it is at the edge of the convergence region for 200 or so samples. In Fig. A.2b, we show the convergence plots of our statistics in a high-noise scenario. We see that all the statistics reach convergence after 200 samples. This is likely due to the fact that this is a noise-dominated regime, and so after only 100 samples, for all statistics, the Fisher matrix is at the 10% error before needing 100 samples more to fully converge. In Fig. A.2c, we show the convergence plots of our statistics for 1000 hours of SKA noise. We see a similar trend to that in Fig. A.2b, where the regime is still noise dominated in nature, and around 200 samples are required to reach noise convergence.

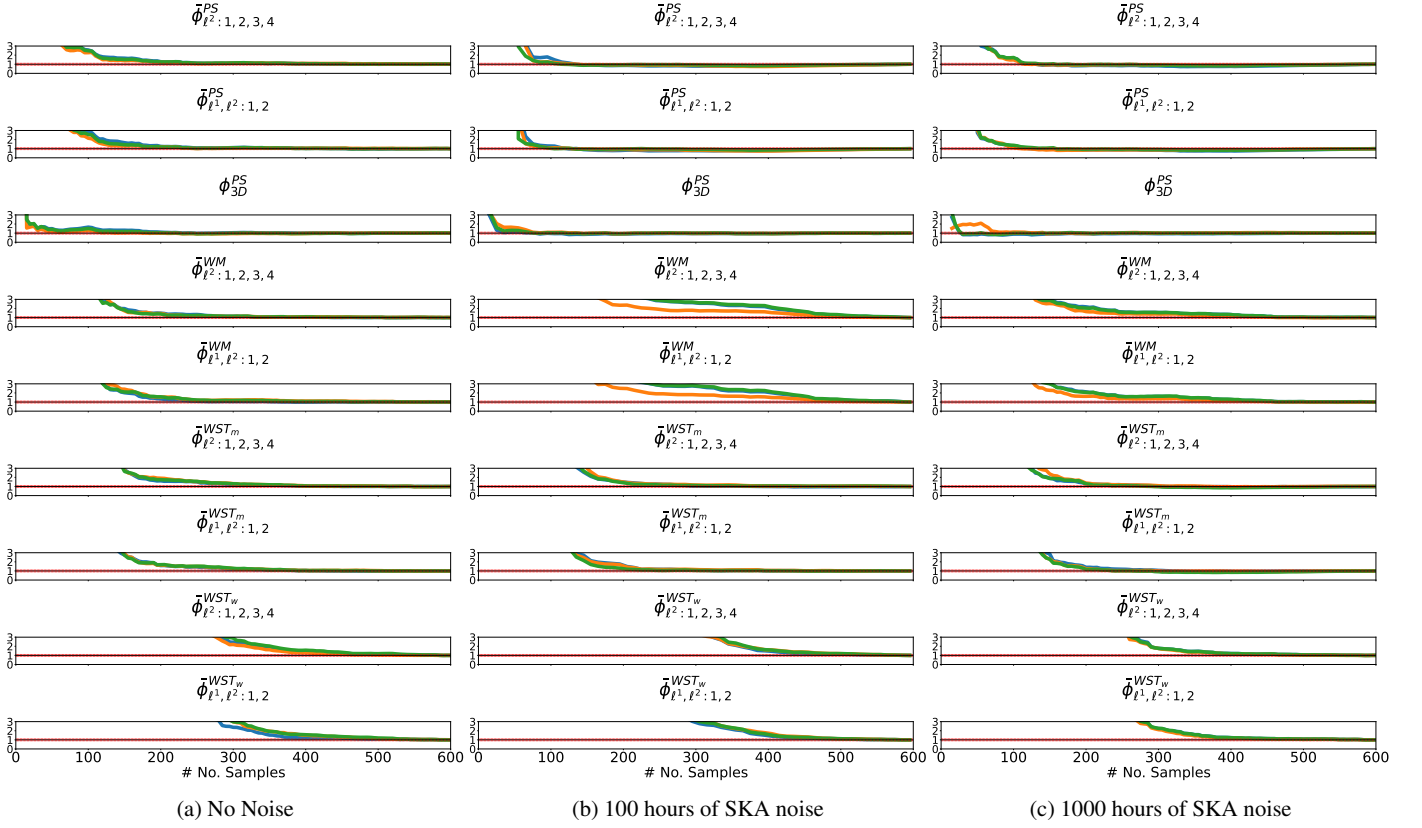


Fig. A.1. Convergence plots using equation (25), where we have kept the number of simulations used for our derivatives constant at 400 and are solely varying the number of fiducial simulations used for the covariance. We consider the different noise cases: (a) no noise, (b) 100 hours of SKA noise, and (c) 1000 hours of SKA noise. We see that by 200 samples all of our statistics are fully convergent, falling within the 10%, shown as the red shaded region.

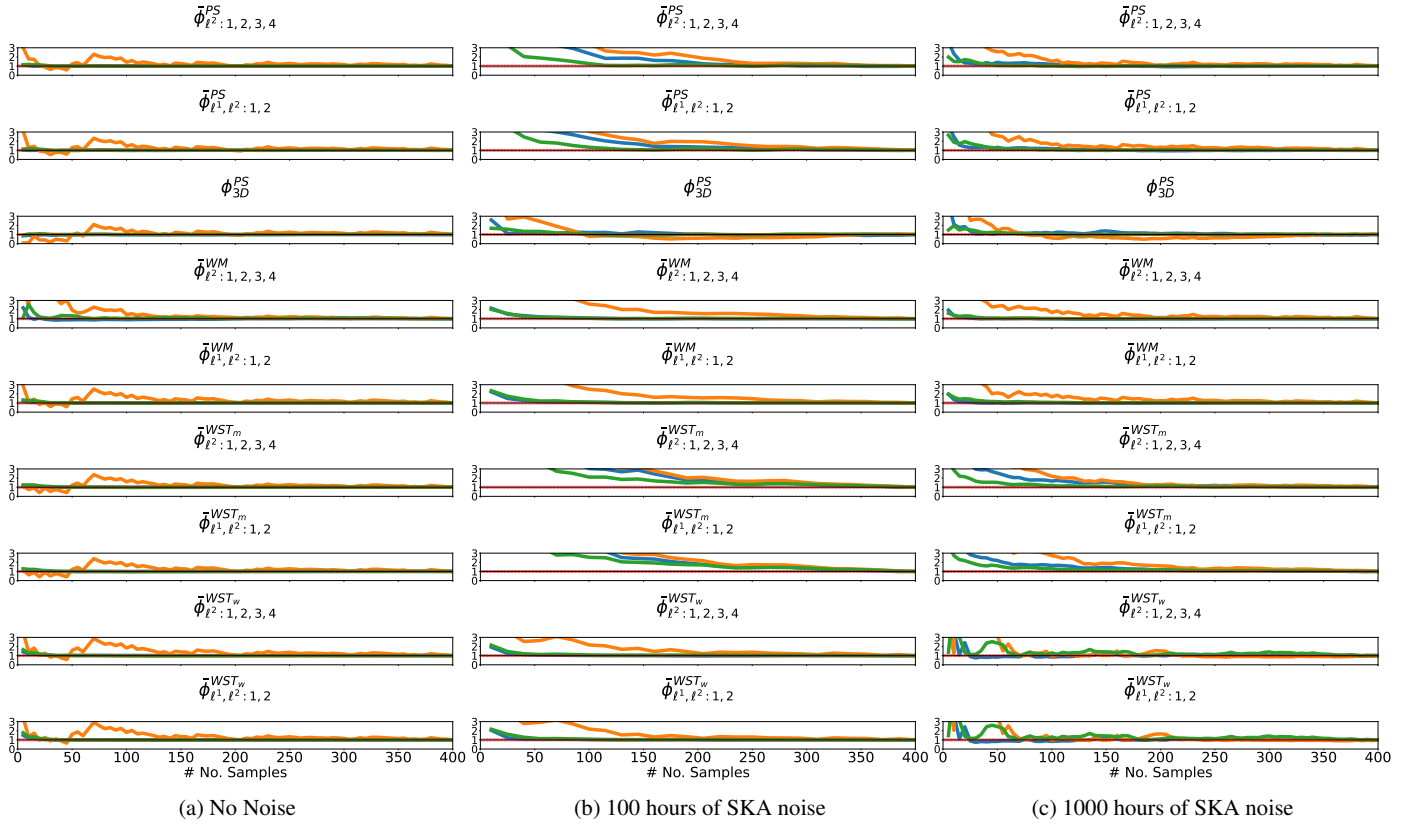


Fig. A.2. Convergence plots using equation (25), where we have kept the number of fiducial simulations for our covariance constant at 600 and are solely varying the number of samples used for the derivatives. We consider the different noise cases: (a) no noise, (b) 100 hours of SKA noise, and (c) 1000 hours of SKA noise. We see that by 400 samples all of our statistics are fully convergent, falling within the 10%, shown as the red shaded region.