The ‘Marked’-statistics to unveil EoR 21-cm signal

M.Sc Thesis

By

Vishrut Sanjay Pandya

DEPARTMENT OF ASTRONOMY, ASTROPHYSICS & SPACE ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

May 20, 2025

The ‘Marked’-statistics to unveil the EoR 21-cm signal

A THESIS

Submitted in partial fulfillment of the
requirements for the awards of the degree
of
Master of Science

by

Vishrut Sanjay Pandya

DEPARTMENT OF ASTRONOMY, ASTROPHYSICS & SPACE ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

May 20, 2025

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **The 'Marked'-statistics to unveil the EoR 21-cm signal** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DEPARTMENT OF ASTRONOMY, ASTROPHYSICS AND SPACE ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from **June 2024** to **May 2025** under the supervision of **Dr. Suman Majumdar, Associate Professor, Department of Astronomy, Astrophysics and Space Engineering.**

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

10/05/2025
Signature of the student with date
(Vishrut Sanjay Pandya)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

12/05//2025
Signature of the Supervisor of
M.Sc. thesis (with date)
(Dr. Suman Majumdar)

Vishrut Sanjay Pandya has successfully given his/~~her~~ M.Sc. Oral Examination held on **06/05/2025**.

Signature(s) of Supervisor(s) of MSc thesis
Date: 12/05/2025

Programme Coordinator, M.Sc.
Date: 19/05/2025

Convenor, DPGC
Date: 19/05/2025

HoD, DAASE
Date:

ACKNOWLEDGEMENTS

My research experience at CSI Lab, DAASE, IIT Indore has been a wonderful journey of learning and growing. I would like to express my gratitude to all my teacher, seniors and friends for making this research journey absolutely amazing.

I would like to express my heartfelt thanks to my supervisor and mentor, **Dr. Suman Majumdar**, Associate Professor, for his constant support and guidance during my Master's journey. He trusted me with responsibilities and encouraged me to explore and lead the project. Our discussions helped me think deeply and critically. He created an open and honest environment where I could learn and grow freely as a researcher. He also gave me opportunities and platform to present my work and always motivated me to do well.

I thank my seniors—**Chandrashakher**, **Leon**, **Mohit**, and **Yashraj**—for helping me whenever I was stuck. A special thanks to **Leon**, who patiently answered my questions anytime I asked. Their support has helped me grow into a better scientist.

To my lab mates **Parth** and **Prasad**, thank you for the many interesting conversations we had. You not only made research fun and helped me see things from different angles, but also provided with your valuable feedback on my performance.

I want to thank my peers, **Vijay** for teaching me how to keep things simple and learn quickly. I'm also thankful to **Annie**, **Anushka**, and **Daisy** for listening to my ideas and giving me fresh perspectives. Moreover, letting me listen to their work and educating me on different fields of research patiently.

A big thanks to **Aryan** and **Devesh** for arranging mock presentations and helping me improve my presentation skills. I'm also grateful to **Gitaj**, **Harikrishnan**, and **Amar** for the fun chats and thoughtful discussions that helped me take a break from my own work and refresh my mind. Thanks to **Navanit** for showing me how to look at my project from a business point of view and how to share and pitch ideas better. Thank you to **Ashutosh** for letting me hang out near the observatory and for the amazing telescope sessions—you helped me live a childhood dream.

A special thanks to our new lab member **Shiriny didi** for always believing in me and cheering me on when I needed it.

A genuine thanks to **Dr. Debanjan Sarkar** and **Dr. Moh'd Kamran** for their valuable suggestions and insights on my project.

Finally, I dedicate this thesis to my mother, **Dhwani Pandya**, and my sister **Nikita Pandya**, who have always supported me with love and strength. Their constant support and care has made me everything I am right now. I owe them everything.

—— Vishrut Pandya

ABSTRACT

The 21-cm signal from neutral hydrogen in the intergalactic medium (IGM) serves as a powerful probe of the Epoch of Reionization (EoR), the period during which the first luminous sources were formed and Universe went through a phase transition from being neutral to ionized. The 21-cm signal, arising from the spin-flip transition in neutral hydrogen, traces the distribution and evolution of large-scale structures during this critical phase of cosmic history.

The EoR 21-cm brightness temperature field exhibits significant non-Gaussianity due to the complex distribution and growth of ionized regions, which is tied to the various astrophysical processes in the IGM and within the sources. As a result, the standard 21-cm power spectrum fails to capture all the relevant statistical information hidden in this field. While higher-order statistics, such as the bispectrum, can provide additional insights, they are computationally intensive and often difficult to interpret.

This project investigates the application of **marked statistics** as an alternative approach to characterize the 21-cm field. By employing non-linear transformations to the original field, the contributions from distinct types of fluctuations can be modified independently, allowing for targeted analysis of specific features. In particular, enhancing specific field properties amplifies their corresponding fluctuations, improving the ability of the **marked power spectrum** to capture non-Gaussian signatures.

Through this project, we demonstrated that it is possible to construct transformations that render the power spectrum a more optimal summary statistic for the transformed field. Marked statistics assign weights or "marks" to each spatial location based on local field properties, effectively emphasizing targeted structures. We show that this approach provides a computationally efficient framework for extracting non-Gaussian information and enables a more detailed statistical description of the 21-cm signal. The project further explores different forms of mark functions and their impact on the resulting statistics, with the broader goal of improving our understanding of the EoR.

CONTENTS

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Brief history of the Universe

In the standard cosmological model, the Big Bang is considered to be the starting point of the Universe as we observe it now. This state of the Universe was very hot, very dense, and much smaller in size. This hot, dense state was filled with a soup of various subatomic particles and radiation, coupled together. As the Universe expanded, this hot, dense state of matter and radiation had time to cool down, and highly energetic radiation decoupled from matter. This radiation is observed today in the microwave regime of the electromagnetic spectrum, known as the Cosmic Microwave Background (CMB). After some more time, baryons had the opportunity to form stable atoms. Most of the atoms formed during this stage were neutral hydrogen, along with some helium, and this gas remained in the intergalactic medium for a very long time. This era is called the Dark Ages of the Universe.

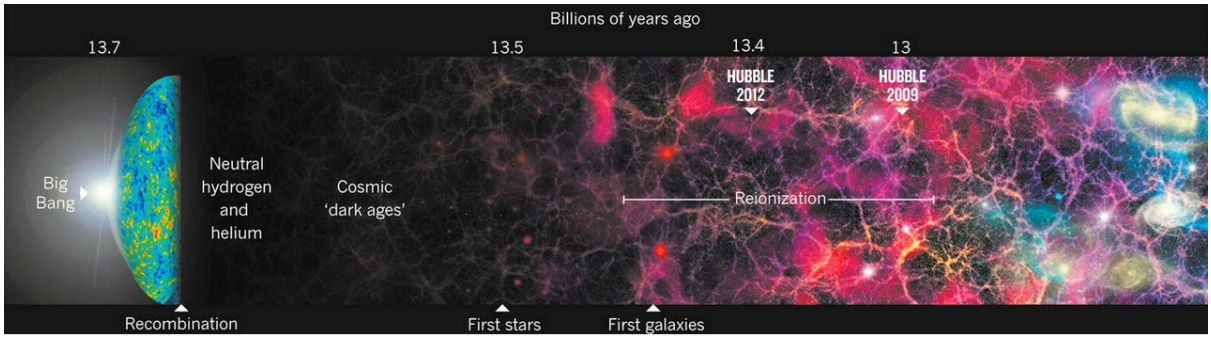


Figure 1.1: Cosmic Reionization of Hydrogen and Helium by Suk Sien Tie.

After a significant amount of time, due to gravitational instability arising from primordial fluctuations, gas began to collapse under its own gravity and formed stars—sources of photons. This part of cosmic history is known as the Cosmic Dawn. These photons were able to escape the dark matter halos, the sites where gas accumulates and stars are formed, and reionize the intergalactic medium. This gradual "phase change" of the Universe from mostly neutral to mostly ionized is called the Epoch of Reionization. The large-scale structures in this state of

the Universe are governed by the dynamics of matter and radiation and involve highly complex astrophysical processes. After the intergalactic medium became almost completely ionized, the remaining neutral matter became trapped in large, dense pockets that are optically thick enough to shield radiation from the outside. The era after the Epoch of Reionization, the Post-EoR era in which we live, contains various structures and non-linearities. In this era, most of the baryonic matter found in the universe is ionized in nature.

Each era in the cosmic history has different astrophysics and cosmology showing the evolutionary nature of the entire Universe. Studying about them, in detail, will lead to better understanding about the physics governing the largest scales known to us. This makes cosmology rich and exciting branch of astronomy.

Epoch of Reionization is considered as a gradual phase change era of the Universe which evolved from mostly neutral to mostly ionized. This transformation of the Inter-Galactic medium (IGM) is composed of different kind of complex as well as interesting physics which can reveal nature of the Cosmos itself. It not only is interesting but is also important to study aspects of this era as it explicitly governs dynamics of IGM. This is the case as the non-linearities are byproduct of the physics governing nature of the IGM as they translate directly to current distribution of large scales.

1.2 The Epoch of Reionization

As discussed in the previous section, the last scattering epoch is considered to occur right after the Cosmic Microwave Background. At this stage, the first neutral hydrogen was formed, around redshift $z = 1100$. Following this stage, for a prolonged period, the baryonic matter in the intergalactic medium (IGM) primarily existed as neutral hydrogen. During this phase, no luminous sources were present to emit detectable photons, leading to a period known as the Dark Age. After a significant amount of time, gravitational instability began to govern the dynamics, causing gas to collapse under the influence of gravity, initiating star formation. Due to the abundance of matter, luminous sources of very large scales formed. These sources produced photons that started to reionize the intergalactic medium. This period in the universe's history is known as the Epoch of Reionization. It marks an essential phase in cosmic history, where the state of the IGM transitioned from neutral to ionized. The primary evidence of this "phase transition" is the Lyman-alpha forest.

1.3 Quasar Spectra and Lyman-Alpha Forest

Quasars, in this context, are very bright sources crucial for probing the properties of the intergalactic medium. Due to their high brightness, quasars can be detected from vast distances

and, consequently, large redshifts. Their distinct spectra carry signatures of the IGM. The photons emitted by quasars travel through the intergalactic medium and interact with neutral hydrogen. These high-energy photons are redshifted due to cosmic expansion, resulting in an increase in their wavelength. At a certain point, their wavelength matches that of the Lyman-alpha transition, allowing them to be absorbed by neutral hydrogen.

At low redshifts, absorption lines are expected in the spectra. If neutral hydrogen is more diffused, the probability of absorption increases over larger distances (and hence larger redshifts), leading to a continuous dip in the quasar absorption spectra. The optical depth of diffuse gas is approximately $\tau \approx 10^5 x_{\rm HI}$, so even a small amount of neutral hydrogen can account for high optical depths at lower redshifts. However, this behavior is not observed at higher redshifts. Instead, the observed absorption lines are distinct, suggesting that most of the gas in the IGM exists in compact, dense objects known as damped Lyman-alpha systems (DLAs), rather than in a diffuse state. The corresponding well-known feature in the quasar spectra is the Gunn-Peterson trough. An example of such a quasar spectrum is shown below:

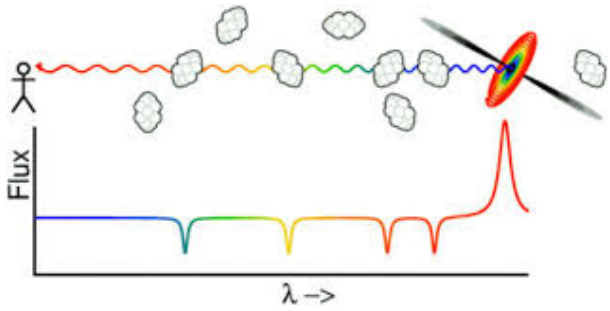


Figure 1.2: Cartoon showing idea of Lyman alpha forest (Edward L. Wright).

Consider the figure as shown below:

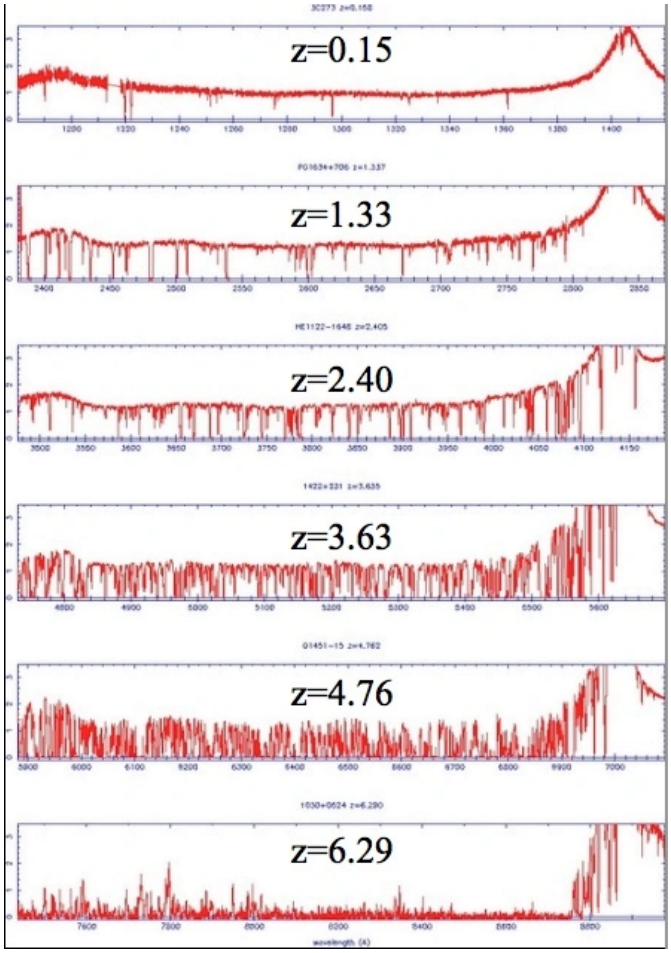


Figure 1.3: Quasar spectra at different redshifts: (From Ross McLure's STFCastro powerpoint).

As we can see clearly, as the redshift increases, the dips in the quasar spectra increases. The absorption features become more in number as redshift increases. Since neutral hydrogen absorbs radiation in discrete levels, the dips in this spectra tells us about the distribution of neutral hydrogen at different redshifts. At farther redshifts, due to cosmological expansion, light from quasar gets redshifted and at some redshift, it becomes very close to that of Lyman alpha transition. Hence, the neutral hydrogen absorbs that wavelength at that redshift and we observe dip in the quasar spectra. The cumulative effect of absorption of each different dips gives the idea about how the neutral hydrogen atom is expected to be found in close dense structures. This is the case as if the distribution of hydrogen were uniform, then it should correspond to a continuous dip in the quasar spectra.

1.4 Models of Epoch of Reionization

For Reionization to take place, the radiation should be strong enough to counteract the recombination as well as photons need to travel larger distances to ionize the inter galactic medium. The dark matter halo needs to be massive enough to host a galaxy which can support star

formation. Given this, the star formation needs to be efficient enough. Moreover, the stellar initial mass function and metalicity are also key factors deciding the dynamics of the process of reionization. The massive stars in the era on interest produce high energy photons which can ionize the medium but are very short lived stars. Hence, their nature and dynamics hold an important role in process of reionization. Moreover, the dynamics of a single halo as well as merger of halos also determine nature of reionization. Incorporating all this complicated astrophysics into a simple model is a difficult task however, all this processes can be modeled by some parameters. They are $f_\star$, the fraction of mass forming into stars, $f_{\rm esc}$, the fraction of photons escaping the star and n_γ, number of ionizing photons for a unit solar mass. These free photons propagate in the IGM. This governs nature of the radiative transfer. The escaping of photons governs the growth of the ionized regions. As discussed before, these photons need to compensate for the recombination of neutral hydrogen atom to keep the medium ionized. Hence, this gives the following rate equation:

$$\dot{n}_\gamma \Delta t = n_{\rm H} Q_{\rm HI} + \Gamma_{\rm HI} \dot{n}_{\rm HI} Q_{\rm HII} \Delta t \tag{1.1}$$

where LHS corresponds to number of ionizing photons produced in time Δt, first term on the RHS corresponds to ionizing the neutral hydrogen and the other term corresponds to keeping the ionized regions ionized against recombination. Now, since both of the processes are independent of one another, they satisfy detailed balanced condition which gives the following equation governing the dynamics of fraction of ionized volume as follows:

$$\frac{dQ_{\rm HII}}{dt} = \frac{\dot{n}_\gamma}{n_{\rm H}} - C \cdot Q_{\rm HII} \cdot \alpha_{\rm B} \cdot n_{\rm H} \cdot a^{-3} \tag{1.2}$$

where C is the clumping factor and $\alpha_{\rm B}$ corresponds coefficient of recombination and a is the scale factor.

In a halo, gas collapses under gravity and is heated to a virial temperature. In order to form stars, the gas needs to collapse under gravity but the thermal energy is hinders efficiency of this process. Hence, the gas needs to cool down a bit before star formation takes place. For very high redshift there are limited ways this cooling can take place. Hence some feedback mechanism may provide better insight to understanding the processes of Reionization.

CHAPTER 2

HI 21-cm signal as a probe of Reionization

One of the promising probes to study EoR is the HI 21-cm signal. In this era, HI 21-cm differential brightness temperature is observed as an emission line. It originates from the neutral hydrogen atom corresponding to the spin-flip transition. Neutral hydrogen in this phase of the Universe is mostly found in the intergalactic medium. Hence, studying this signal from the IGM should provide valuable information about the large-scale properties of the IGM.

Hydrogen atom is made up of a proton at the nucleus and an electron in its "orbit." The ground state (principal quantum number $n = 1$) of the hydrogen atom has no net angular momentum ($l = 0$). However, the proton and electron have quantum spins that interact with each other. This interaction results in the splitting of the ground state of the hydrogen atom into two distinct energy states: the singlet and triplet states. Each of these states corresponds to the configuration of spins of the proton and electron in anti-parallel (singlet) and parallel (triplet) directions, respectively. These distinct energy levels have a very small energy difference (0.5×10^{-6} eV). The transition of the hydrogen atom from one state to another (corresponding to the same energy difference) corresponds to a frequency of $f_o = 1420$ MHz (This is the rest frame frequency. As the Universe expands the frequency changes to $\frac{f_o}{1+z}$ where z is the redshift. Hence, to target specific redshift, corresponding frequency is to be tuned to capture the signal.) or a wavelength of $\lambda \approx 21.1$ cm ($\lambda(1 + z)$). This line is known as the HI 21-cm line. Mapping this line in the intergalactic medium can provide a powerful probe of large-scale structures at the targeted redshift. This is because, at early redshifts, most of the baryonic matter found in the IGM is in this neutral hydrogen state.

Let n_1 be the number density of atoms in the singlet state and n_2 be the same for the triplet state. Thus, the distribution of the relative population of number densities of atoms in the singlet and triplet states is given as follows:

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} \exp\left(-\frac{T_\star}{T_s}\right), \tag{2.1}$$

where g_1 is the degeneracy of the i^{th} state and $T_\star = \frac{hf_o}{k_B} \approx 0.068$ K. The quantity T_s is called the spin temperature, which quantifies the relative population of neutral hydrogen atoms in the ground state. An important fact about the triplet state is that its corresponding lifetime is about 10^{15} sec. This means it is a very long-lived state, and the transition probability from one state to another is negligible. However, given the timescale of the universe and the abundance of neutral hydrogen as baryonic matter, the possibility of transition becomes more likely. Moreover, various processes can change the level population and cause hydrogen atoms to transition from one energy level to another. The three main processes are as follows:

1. Radiation due to CMB
2. Collisional coupling
3. Lyman-alpha coupling (Wouthuysen–Field effect)

CMB This also works as a background, and hence, the signal is observed as fluctuations corresponding to $T_{\text{CMB}} = T_\gamma$. The required signal is seen as spectral distortion to the CMB signal for a given frequency. Observations at different frequencies probe different scales of the observable universe, and hence, a final 3D map can be made. Either the signal is observed as a disturbance in this way, or it is seen as a disturbance where loud radio sources are considered as the background.

Collisional Coupling The idea is more or less known, but the new information known is that as there are different species of particles present at a given time, each of them initiates spin-flip transition in their own way, and hence, these are to be treated as different approaches. The collisional coupling for system i is given as follows:

$$x_c^i = \frac{C_{10} \cdot T_\star}{A_{10} \cdot T_\gamma} = \frac{n_i k_{10}^i}{A_{10}} \frac{T_\star}{T_\gamma}, \tag{2.2}$$

Thus, total collisional coupling is modeled as the sum of all the coupling and hence is then modeled using individual rate coefficients. Note that $T_\star = \frac{hc}{\lambda_{21} k_B}$, and the species are hydrogen, electron, and proton.

Wouthuysen–Field effect The general expected spin-flip transition can also be caused by a specific set of transitions, which are discussed in previous papers. The coupling in this case is given as follows:

$$x_\alpha \propto \frac{T_\star}{T_\gamma} \frac{\chi_\alpha}{A_{10}} S_\alpha J_\alpha = \frac{S_\alpha J_\alpha}{J_\alpha^c}. \tag{2.3}$$

There is observed asymmetry in the transition line by something known as *Flow of photons*, which arises due to the finite cross-section of L_α photons interacting with HI. (Scattering of electrons produces a high population of lower energetic electrons and hence the asymmetry).

Here, the distribution of electrons, which are changed by spin flips, is neglected. Also, these L_α photons can be produced from cascading of atomic species apart from star formation.

Based on the targeted time, each process predominates over the other two and attempts to couple the spin temperature with that corresponding process. After recombination, most baryonic matter is found in the form of neutral hydrogen atoms. In the very early universe, this neutral hydrogen is abundant, increasing the likelihood of gas collisions. During this era, collisional processes govern the dynamics of the gas, and hence, the spin temperature follows the collisional temperature. After reionization progresses significantly, Lyman-alpha photons can excite hydrogen atoms from the singlet state to the first excited state, after which the atom can return to the ground state but in the triplet state. Thus, the overall change in the system of hydrogen atoms involves transitioning from the singlet state to the triplet state and vice versa. Consequently, the evolution of the spin temperature is governed by the relative activity of these processes.

We wish to measure the brightness temperature of the spin-flip transition of the neutral hydrogen gas, which may show signatures of coupling with CMB photons. The coupling of the spin-flip transition is governed by the following equation:

$$T_s^{-1} = \frac{T_\gamma^{-1} + x_g c \cdot T_g^{-1} + x_\alpha \cdot T_\alpha^{-1}}{1 + x_g + x_\alpha}. \tag{2.4}$$

Each process tries to couple the spin-flip temperature with the corresponding energy level of that temperature. Thus, based on the era of the universe, the spin-flip transition can be observed as an absorption or emission line. After recombination, T_s couples with the gas temperature T_g and the CMB temperature T_γ; therefore, fluctuations with respect to either process cannot be detected. This age is popularly known as the Age of Ignorance. After the dark ages, the coupling of T_s with T_g decreased due to adiabatic cooling caused by the expansion of the universe, allowing T_s to be observed as an absorption line. After the formation of the first luminous sources, T_s decouples from T_γ by the Wouthuysen–Field effect or Lyman-alpha pumping, which raises T_s and makes it observable as an emission line.

The observed differential brightness temperature δT_b arises from the interplay between the spin temperature, T_s, and the background radiation temperature, T_R (commonly T_γ for the CMB), along with the optical depth τ_ν. It can be expressed as:

$$\tau_\nu = \int ds \, [1 - \exp(-E_{10}/k_B T_s)] \sigma_0 \phi(\nu) n_0 \tag{2.5}$$

where E_{10} is the energy difference between the hyperfine levels, k_B is the Boltzmann constant, σ_0 is the cross-section, $\phi(\nu)$ is the line profile function, and n_0 is the number density of neutral hydrogen atoms.

The HI 21-cm signal interacts with medium in between the source it is emitted from and

hence, the observed intensity according to radiative transfer goes as follows:

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu + j_\nu, \tag{2.6}$$

Now, in the regime of radio astronomy,

$$I_\nu = \frac{2k_B T \nu^2}{c^2} \tag{2.7}$$

This gives

$$T_b^{\text{obs}} = T_{\text{ex}}\,(1 - e^{-\tau_\nu}) + T_R(\nu)e^{-\tau_\nu} \tag{2.8}$$

where τ_ν is the optical depth. For small τ_ν, this simplifies to:

$$\delta T_b = \frac{T_s - T_R}{1+z}(1 - e^{-\tau_\nu}) \tag{2.9}$$

Substituting the expression for τ_ν and incorporating cosmological parameters, we arrive at the final expression:

$$\delta T_b \approx 27 x_{\text{HI}}(1+\delta_b)\left(\frac{\Omega_b h^2}{0.023}\right)\left(\frac{0.15}{\Omega_m h^2}\frac{1+z}{10}\right)^{1/2}\left(\frac{T_s - T_R}{T_s}\right)\left(\frac{\partial_r v_r}{(1+z)H(z)}\right)\text{ mK}. \tag{2.10}$$

Here, x_{HI} is the neutral hydrogen fraction, δ_b is the baryonic density contrast, Ω_b and Ω_m are the baryon and matter density parameters, h is the dimensionless Hubble parameter, $H(z)$ is the Hubble parameter at redshift z, and $\partial_r v_r$ is the velocity gradient along the line of sight. This expression encapsulates the physical and cosmological parameters influencing the 21-cm brightness temperature. [1]

CHAPTER 3

Statistics, Non-Gaussianity and Marked Statistics

To study the field and its properties, various statistical tools are employed. This approach comes from the intrinsic nature of the fluctuations present in the field, which arise from underlying randomness of the field. This makes inherit nature of the fluctuations non-deterministic in nature, governed by the probabilistic behavior arising from the underlying physics. For example, during the Epoch of Reionization, fluctuations in the 21-cm signal result from the complex interplay of density variations, ionization fronts, and the distribution of matter and radiation in the intergalactic medium. By studying statistical measures one can study the underlying properties and evolution of the target field. These tools allow us to make sense of the inherent randomness in the fluctuations, offering insights into the physical processes shaping the field and its large-scale structure

3.1 Power Spectrum

The power spectrum is defined as the Fourier transform of the two-point correlation function. This two-point statistic provides information about the amplitude of fluctuations present at a given length scale or its corresponding k-mode. The two-point correlation function, in the context of $\zeta(x_1, x_2)$, is defined as:

$$\zeta(\mathbf{x_1}, \mathbf{x_2}) = \zeta(|x_1 - x_2|) = \langle \delta(x_1)\delta(x_2) \rangle \tag{3.1}$$

where $\delta(x)$ represents the fluctuation at position x. Here, the assumption of homogeneity and isotropy, as per the cosmological principle, implies that ζ depends only on the distance $|x_1 - x_2|$ between two points, not on their absolute positions. This correlation function encodes the probability of finding an object at x_2 given that another object is located at x_1, thereby quantifying the degree to which two points in space are correlated. To move from the correlation

function to the power spectrum, we take the Fourier transform of $\zeta(r)$ over all spatial directions which is given as follows:

$$\xi(r) = \frac{1}{(2\pi)^3} \int d^3k \, P(k) \, e^{-\mathbf{k}\cdot\mathbf{r}} \tag{3.2}$$

where

$$\xi(r) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle \tag{3.3}$$

In the case of the radio observations of the 21cm signal, signals from distant astronomical sources are observed as visibilities, which are the Fourier transform of intensities observed, and hence, working in Fourier space is convenient rather than working in real space. Since power spectrum essentially measures the relative contrast in the field with respect to mean for given length scale, it is estimated as follows:

$$\langle \delta_A(\mathbf{k_1}) \cdot \delta_A(\mathbf{k_2}) \rangle = (2\pi)^3 \delta_D(\mathbf{k_1} + \mathbf{k_2}) \mathrm{P}_{AB}(k_b) \tag{3.4}$$

where $\delta(\mathbf{k})$ is the variation in the field with respect to mean, normalized by mean at a given point c in the field and δ_D is the Dirac delta function which ensures $\mathbf{k_1} = -\mathbf{k_2}$. The wavenumber $|\mathbf{k}| = \frac{2\pi}{x}$ is the conjugate variable of the corresponding length scale $\mathbf{x}$.

Moreover, we use spherically averaged binned power spectrum which is given as follows:

$$\mathbf{P}(\mathbf{k}) = \frac{1}{N_{k_n}} \sum_{k_i} \mathbf{P}(\mathbf{k_i}) \tag{3.5}$$

where k_i represents average Fourier transform in i^{th} bin. Moreover, we also use the dimensionless power spectrum which is given as follows:

$$\Delta^2(\mathbf{k}) = \frac{k^3}{2\pi^2} P(\mathbf{k}) \tag{3.6}$$

3.2 Non-Gaussianity and Higher order statistics

The target HI 21-cm signal is governed by various complex astrophysical processes. Moreover, the target era of study, the Epoch of Reionization (EoR), is considered a phase transition of the intergalactic medium from neutral to almost completely ionized. Photons (UV, X-ray) emitted by sources gradually ionize the intergalactic medium, depending on the distribution and density of neutral gas. Consequently, the signal is influenced not only by matter density instabilities but also by the growth of ionized regions or bubbles [2]. In the very early stages of the EoR, the ionized regions are relatively small in size, and the distribution of neutral hydrogen gas is mainly governed by gravitational instabilities. This corresponds to a very low degree of

non-Gaussianity in the target signal. However, at later stages of reionization, ionized regions begin to cluster around dense pockets of matter. As reionization progresses, the number and size of these regions increase. This makes the signal non-Gaussian in nature, which is now also influenced by matter density fluctuations within these neutral hydrogen pockets. As reionization advances, ionized bubbles grow larger and eventually start to overlap, leading to percolation across larger length scales. The ionized regions become interconnected not only through overlapping but also through their filamentary structures. These structures continue to evolve as ionization progresses, while neutral regions remain interconnected by similar filamentary structures. Simultaneous evolution of these percolated structure takes place. Consequently, the intergalactic medium (IGM) transforms into a complex landscape, containing a variety of ionized and neutral regions whose topology includes various structures and interconnected networks. [3] Additionally, the IGM undergoes a phase transition from having small ionized (hot) regions surrounded by neutral (cold) regions to a state where neutral regions are reduced to dense pockets within an ionized surrounding. This critical information is captured in the bispectrum through sign changes, which cannot be detected by the standard power spectrum due to its always-positive nature. Thus, not only is the signal non-Gaussian in nature, but its degree of non-Gaussianity also varies over time as it evolves.

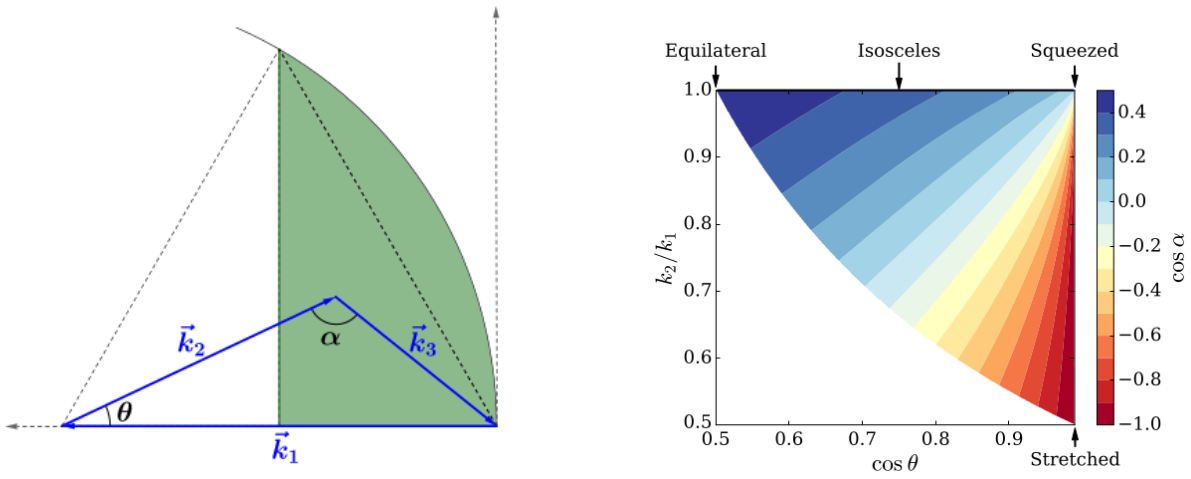
3.2.1 Bispectrum

Inflation is considered to be the reason for which the current Universe is considered flat, homogeneous, and isotropic in nature. Thereafter, the large scale structures we observe today are the reason of evolution of non-linearity. Hence, linear perturbation theory fails to explain the evolution of this non-linear regime of the Universe. Moreover, the HI 21-cm signal from EoR is highly non-Gaussian in nature as described in previous sections. Hence, for a more qualitative study, higher order statistics may be required to understand the field. The higher order correlation functions also contain valuable cosmological information. Two point correlation function is a two point statistic. Moving on to higher order statistic is the three point correlation function which quantifies how three points are spatially correlated. Its Fourier counterpart is the Bispectrum. Similarly, one defines three-point correlation and the BI-spectrum as follows:

$$\langle \delta(\mathbf{k_1})\delta(\mathbf{k_2}\delta(\mathbf{k_3}))\rangle = (2\pi)^3 \mathrm{B}(\mathbf{k_1}, \delta(\mathbf{k_2}), \delta(\mathbf{k_3}))\delta_D(\mathbf{k_1} + \mathbf{k_2} + \delta(\mathbf{k_3})) \tag{3.7}$$

The $\delta_D(\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3})$ ensures that correlation is found in between three points are in a closed triangle configuration. Hence, instead of looking at ensemble averages of product of fluctuations at different length or k modes for the case of power spectrum, the bispectrum looks at different triangle configurations, carrying more information than the standard power spectrum. In order to keep track of unique triangle configurations, Bispectrum is parameterized

using two parameters. They are $\frac{k_1}{k_2}$ and $\cos\theta = \frac{\mathbf{k_1}\cdot\mathbf{k_2}}{|k_1 k_2|}$ where k_1, k_2, k_3 are the k modes of the triangle. This way to parameterize is adopted from [4]



(a) Unique triangle configuration for bispectrum estimation.

(b) Phase diagram of the bispectrum.

Figure 3.1: Illustration of the bispectrum and its representation by parameters for a unique triangle configuration. Majumdar et.al. ([**Majumdar_2018**]).

To estimate Bispectrum is a challenging and a rather computationally heavy task. Moreover, interpreting these higher order statistics is also a rather difficult task on its own. Efforts are made to to make process of estimation of Bispectrum efficient [5] however using different summary statistic may provide a simpler and a rather computationally cheap framework.

3.3 Marked statistics to quantify Cosmology

The HI 21-cm field from neutral hydrogen atoms is a powerful tool to study the properties of the IGM [1]. However, it is quantitatively non-Gaussian in nature. Higher-order statistics such as the bispectrum are required to capture this non-Gaussianity, but they are computationally expensive. This is where the concept of a *mark* becomes useful.

The idea behind marked-based statistics is as follows: to "mark" each point of the field based on some local property, thereby enhancing or suppressing certain features in order to study different aspects of the field. Applying a mark to all points in space based on a chosen property of the field contributes additional information to the standard correlation function. This results in a new quantity known as the *marked correlation function* (MCF). For example, the study in [6] shows how marking halos based on local behavior serves as a robust test to demonstrate that the most massive halos do not populate the least dense cells and do not form at early times.

The idea of marking has been extended in [7] to study voids in galaxy surveys and constrain the mass of massive neutrinos, which are thought to remain undisturbed by virialized

surroundings. In [8], a Fisher analysis on the marked power spectra corresponding to a mark that upweights low-density regions shows how applying a mark can provide tighter constraints on cosmological parameters, especially those influenced by the local properties of the field.

The expression of the working mark in this study is inspired by [9], which demonstrates how density-dependent tests can constrain modified gravity models using marked statistics. Further motivation is drawn from [10], where marked statistics are used to constrain modified gravity models by studying low-density regions in the IGM. A recent study [11] applies the mark to the HI 21-cm signal in the post-EoR era to understand field features. One of the central aims lies in obtaining an optimum mark that captures the maximum information. This is essential since different marks may highlight different field features while still carrying comparable information, as shown in [12].

In the present study, we use the mark in two different ways: one as a non-linear transformation of the field (Equation 6.1), referred to as ***Mark***, and the other as a weight assigned to the field (Equation 6.2), referred to as ***Marked field***. The free parameters R, f, and p have the same interpretation in both cases. Parameter R is the radius of the spherical filter used to smooth the field, encoding local behavior. Parameter p controls the extent of enhancement (for negative p) or suppression (for positive p) of high-value fluctuations in the HI 21-cm differential brightness temperature $\delta T_b(\mathbf{x}, z)$. Parameter f sets the sensitivity of the mark and determines the threshold for feature selection and transformation.

The goal is to understand the features of the HI 21-cm differential brightness temperature field, particularly the nature of its fluctuations. Due to the complex nature of the target signal, we begin by applying the mark to the halo density and neutral fraction fields. After interpreting the results for these simpler fields, we extend our analysis to the observable HI 21-cm differential brightness temperature field. Using insights gained from the initial fields, we employ various mark-based statistics to extract valuable information.

The mark, which assigns weights to spatial points, is fundamentally a property of the field. Hence, the marked correlation function generalizes the two-point correlation function by measuring the clustering of marks:

$$\mathcal{M}(\mathbf{x}) = \frac{1}{n(\mathbf{x})\bar{m}^2} \sum_{i,j} m_i m_j = \frac{1+W}{1+\zeta} \tag{3.8}$$

Here, $n(\mathbf{x})$ represents the number of pairs separated by a distance $\mathbf{x}$ in the field. To compute $\mathcal{M}$, we add up the contributions from all such pairs. Each point i has a mark m_i, and the average mark across the whole sample is $\bar{m}$. The factor W is adjusted using the two-point correlation function ζ so that $\mathcal{M}$ approaches 1 at large separations, where no correlation is expected. For example, if a particle has a mark greater than $\bar{m}$, we check if its nearby particles also tend to have higher marks. This is done by comparing the average mark of nearby pairs to $\bar{m}$ as a

function of their separation.

The idea of the marked correlation function was first introduced in [13], and has since evolved substantially. Works such as [14, 15, 16, 17, 18] have studied galaxy properties such as luminosity and star formation rate (SFR) in relation to clustering. Studies such as [16, 19, 20, 21] explore the halo model and methods to break degeneracies. Marked correlation functions have also been employed to distinguish between modified gravity models [22, 23, 24, 25, 26, 27, 28, 29, 30, 31].

Analytical studies on the marked power spectrum (MPS) such as [32] investigate its relevance in the context of primordial non-Gaussianities (PNG) of the non-local type. Further developments in [33] connect the marked power spectrum to higher-order statistics. In particular, they explore a low-order polynomial form of the mark to control theoretical uncertainties and show agreement with N-body simulations for matter fields and biased tracers in redshift space. These studies demonstrate how degeneracies can be broken by including the marked auto-spectrum or its cross-spectrum with the unmarked field.

The marked power spectrum has also been used to place tighter constraints on neutrino masses [34], and to study clustering properties in galaxy surveys for improved cosmological parameter estimation [35, 36]. The recent work by [11] shows the potential of the marked power spectrum over standard power spectra for the neutral hydrogen field in the post-EoR era.

CHAPTER 4

Mark on Halo density field

Before applying marks to the 21-cm field, we aim to first analyze how marks behave in the context of the halo density field, which is instrumental in generating the HI 21-cm field. Halo density field is obtained as field at ten redshift snapshots corresponding to EoR era, from N-body simulation. The output from the N-body simulations [37] [38] halos are found using Friends of Friends (FOF) algorithm [38]. The halo density field has a large dynamic range, making it an ideal candidate for studying the behavior of marks. The strong contrast in this field allows for straightforward enhancement of high-density regions, which is a key advantage. Again as discussed in previous chapters, *mark* is defined as a transformation applied to the field, while a *marked field* is obtained by multiplying the field with the corresponding marks, effectively assigning weights to each point. By working with a field with high contrast, we can efficiently analyze both marks and marked fields, enabling better clarity in deciding the most suitable approach and improving the interpretability of the results.

$$m(\mathbf{x}, z; R, f, p) = \left[\frac{1+f}{1+f+\mathrm{HM}(\mathbf{x}, z, R)}\right]^p \tag{4.1}$$

$$\mathbf{HM}(\mathbf{x}, z; R, f, p) = m(\mathbf{x}, z; R, f, p) \cdot \mathrm{HM}(\mathbf{x}, z) \tag{4.2}$$

where $\mathrm{HM}(\mathbf{x}, z, R)$ is smoothed halo density field with a spherical filter of radius R. Here, f, R, p are known as the free parameters. In principle, varying these parameters should give relevant information in the field.

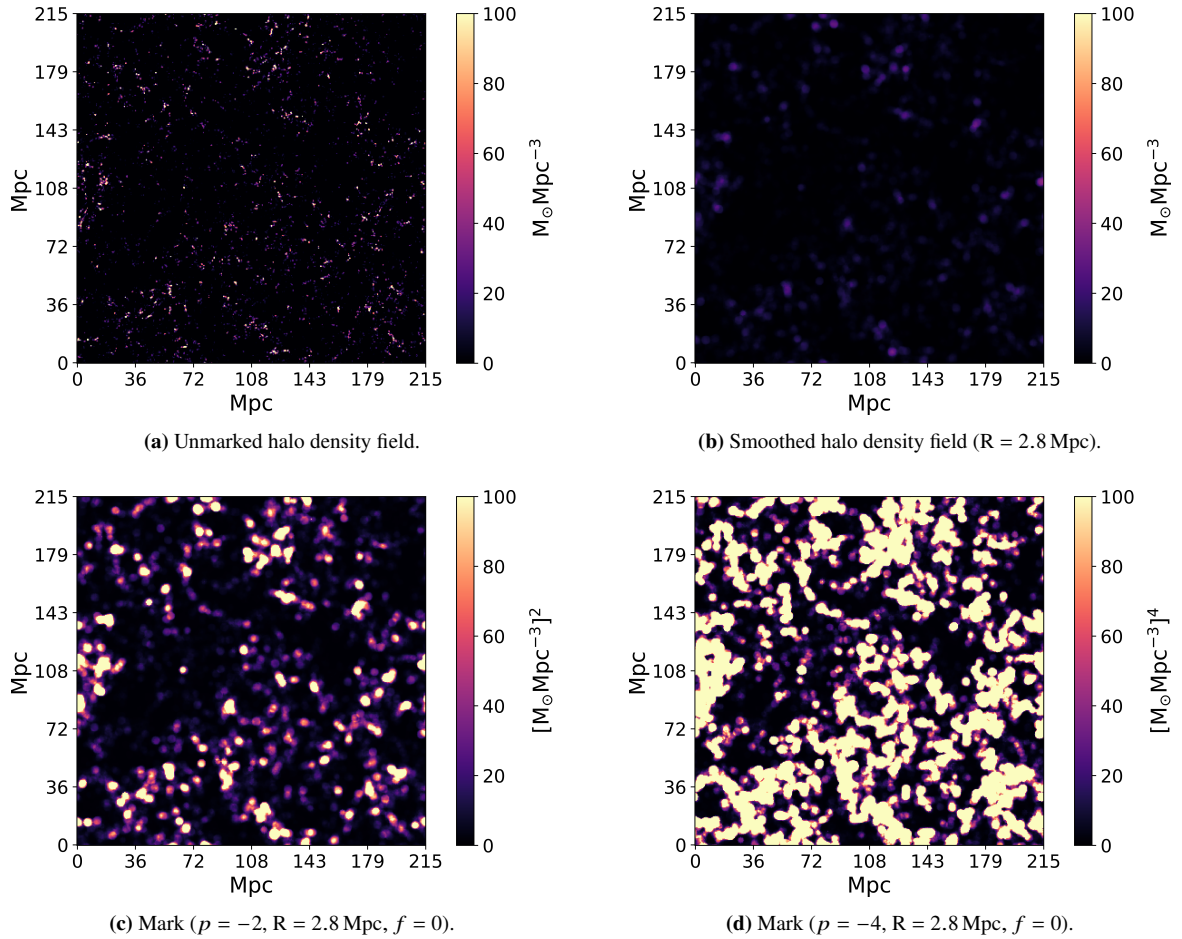
f	R **Mpc**	p
{0.0}	{1.68, 2.80, 3.92, 5.60, 8.40}	{-2,-4}

Table 4.1: Values for free parameters worked for the Halo density field.

Note that halo density field is of the units of density. Hence, the unit of the mark depends on the choice of free parameters. This may make the comparison of marks, mark field and the unmarked fields non intuitive.

4.1 Visual inspection

Looking at the expression of the mark, for negative values of p should enhance the high density regions. Hence, employing this into the procedure should make high density regions more dense and lower density regions even less denser. Hence, by doing this, the contrast in the field is increased which makes it possible to study features of high density easier. For illustration purposes, the outputs are fixed at $x_{\rm HI} = 0.5$ which corresponds to $z = 7.520$. However, what follows stays consistent for entire ionization history.



(a) Unmarked halo density field.

(b) Smoothed halo density field (R = 2.8 Mpc).

(c) Mark ($p = -2$, R = 2.8 Mpc, $f = 0$).

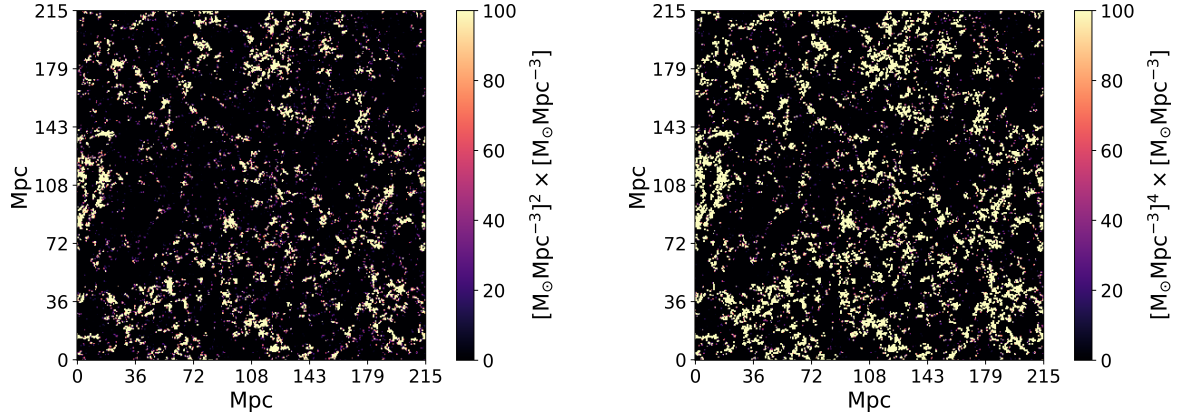
(d) Mark ($p = -4$, R = 2.8 Mpc, $f = 0$).

Figure 4.1: Top: Unmarked and smoothed halo fields. Bottom: Mark for $p = -2$ and $p = -4$. All for $z = 7.520$, R = 2.8 Mpc, $f = 0$.

As we can see clearly, the mark is making the original field more informed about a particular part of feature. This is done by enhancing the features and increasing its contribution in the field. Now, the new transformed field, Mark, is more sensitive to high dense regions. Moreover, higher the value of the exponent, higher is the contribution taken in the field.

We can use this idea of Mark as weights and can give it back to original unmarked field to get Marked field. Below is the plot for the same. Similar to the case of mark, features are picked up as expected. The only difference is the fluctuations at all length scales are preserved as weights are given back to the unsmoothed field, which is the unmarked field. One key observation

here is that for units where different m,ark corresponding to different free parameters will have different units and hence, it becomes challenging to compare the field directly. However, one can treat these Marks and Marked fields as completely different entity for analysis.



(a) Marked field ($p = -2$, R = 2.8 Mpc, $f = 0$). **(b)** Marked field ($p = -4$, R = 2.8 Mpc, $f = 0$).

Figure 4.2: Marked fields for $p = -2$ and $p = -4$ at $z = 7.520$, R = 2.8 Mpc, $f = 0$.

Visually, mark as well as mark field is able to perform as expected. The high density regions are enhanced as the contrast between different types of regions increases. In order to quantify the properties of the fields, statistics are used in the next sections.

4.2 Power Spectrum

The power spectrum, as introduced in Section 3, quantifies the amplitude of fluctuations in a field. To isolate the intrinsic structure of these fluctuations from any length-scale dependencies, we employed the normalized power spectrum defined in Equation 3.6. The resulting plots, presented below, illustrate the key trends observed across various scenarios.

As evident from the plots, increasing the contrast in the field leads to a corresponding rise in the amplitude of the power spectrum. This aligns with physical expectations: greater contrast implies stronger fluctuations, which directly translates to enhanced power across scales. Conversely, the application of smoothing operations predictably suppresses power, especially on scales smaller than the smoothing scale. This is a direct consequence of the averaging nature of smoothing, which redistributes and dampens fine-scale fluctuations.

An important observation is the influence of the field's dynamic range. Particularly in the case of the marked field, this range becomes significantly large. The marked operation—where each point in the field is multiplied by a transformation that acts as a weight—amplifies this effect. While this transformation does reduce variation in certain features of the field, such subtleties are often obscured in the power spectrum due to the overwhelming influence of the high dynamic range.

Nevertheless, the plots clearly demonstrate that the use of a mark enhances power across all k-modes, as expected. Similarly, the marked field, where the mark is applied as spatial weights, exhibits enhanced power in a manner consistent with the mark itself. This consistency affirms the interpretation and role of the free parameters discussed in Section 3.

The effect of smoothing is also coherent across different smoothing scales: as the smoothing scale increases, the suppression of power becomes more pronounced, particularly at scales below the smoothing threshold. These patterns provide strong visual confirmation of theoretical expectations.

Overall, the plots collectively validate the expected behavior of both the marked and smoothed fields. The coherence in how the free parameters influence both the mark and the marked field reaffirms the robustness of the framework. Importantly, these findings motivate further refinement of the marking procedure. Specifically, they suggest the need to design the mark in such a way that the power spectra of the marked and unmarked fields remain meaningfully comparable—even under conditions of high dynamic range. As discussed in previously, the interpretation of the mark, in particularly units, depends on the choice of free parameters used. This motivates us to modify design of mark or the processes of marking in such a way that power spectrum for unmarked and marks are comparable.

As discussed in previously, the interpretation of the mark, in particularly units, depends on the choice of free parameters used. This motivates us to modify design of mark or the processes of marking in such a way that power spectrum for unmarked and marks are comparable.

The corresponding plots are shown below.

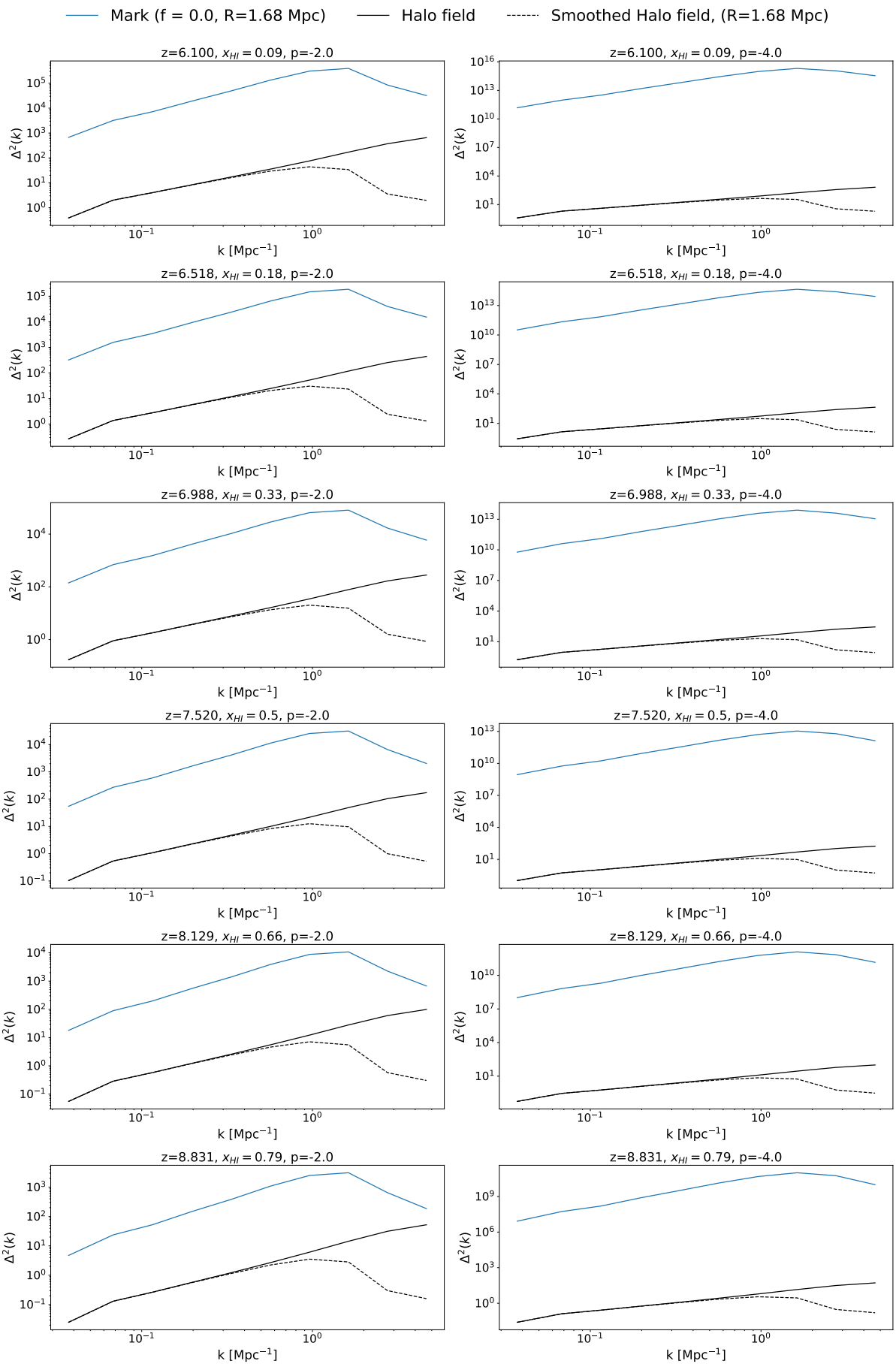


Figure 4.3: R = 1.28 Mpc, Halo Density Field.

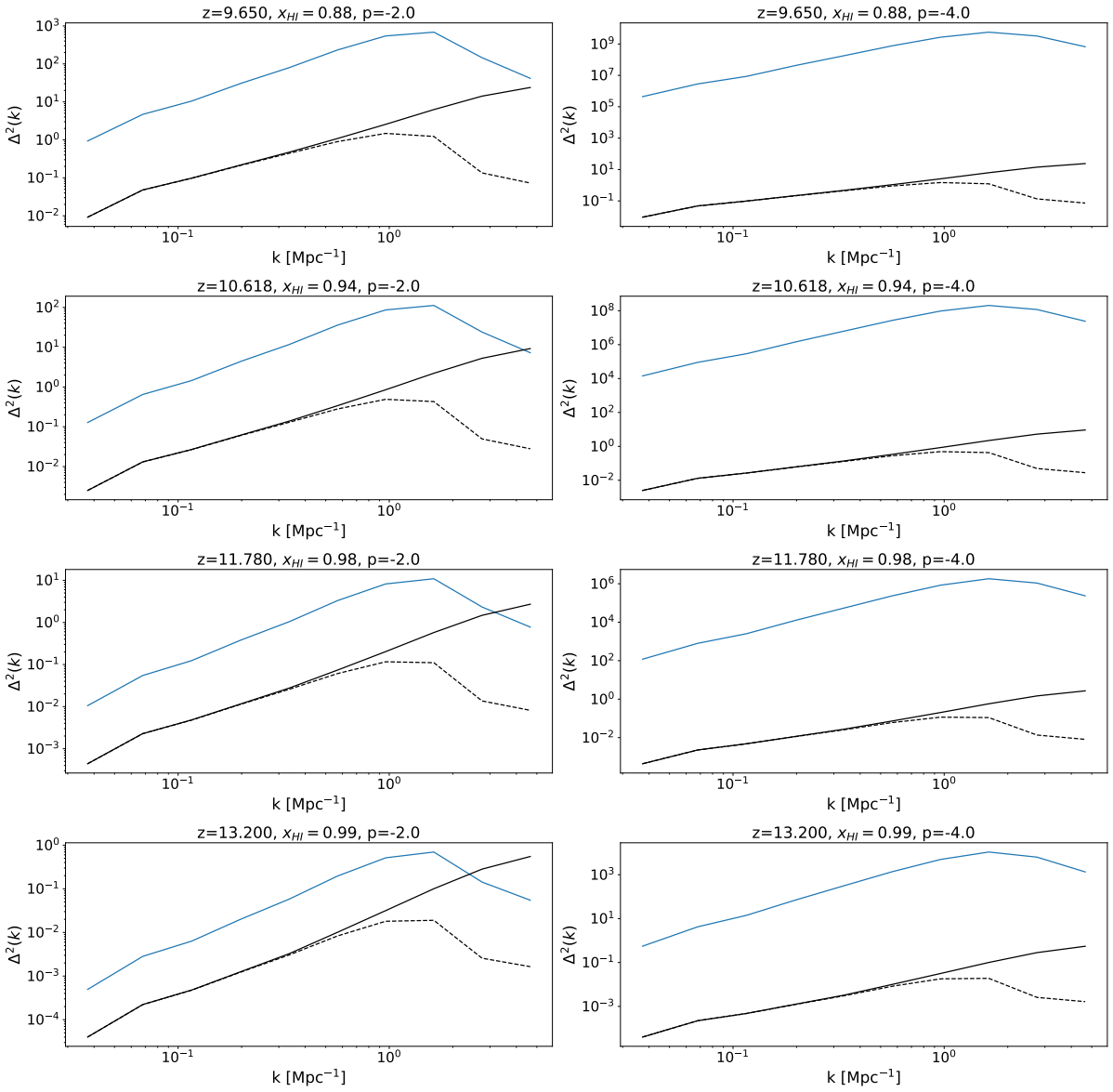


Figure 4.4: R = 1.28 Mpc, Corresponding Smoothed Field and Marks.

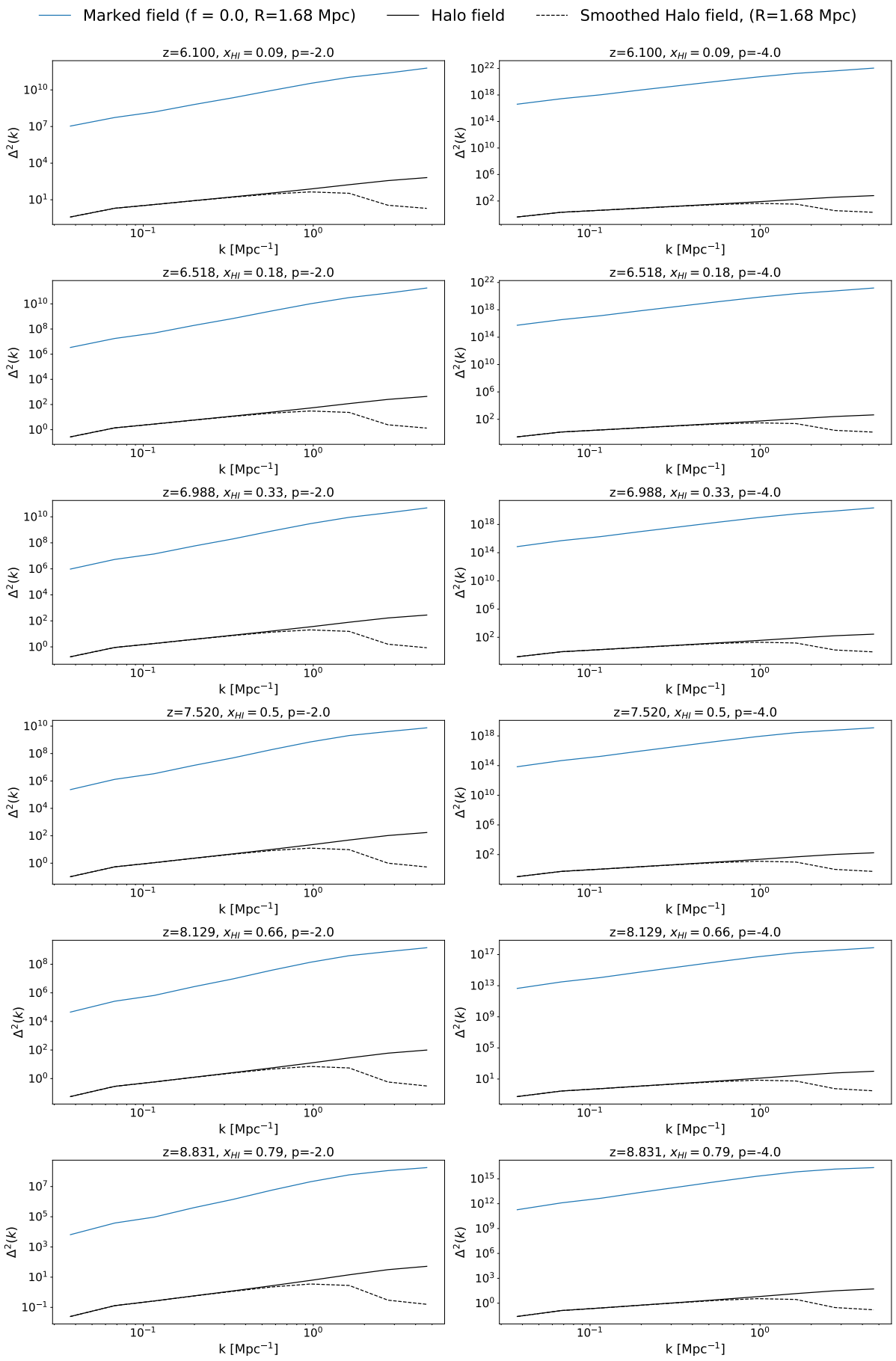


Figure 4.5: R = 5.60 Mpc, Halo Density Field.

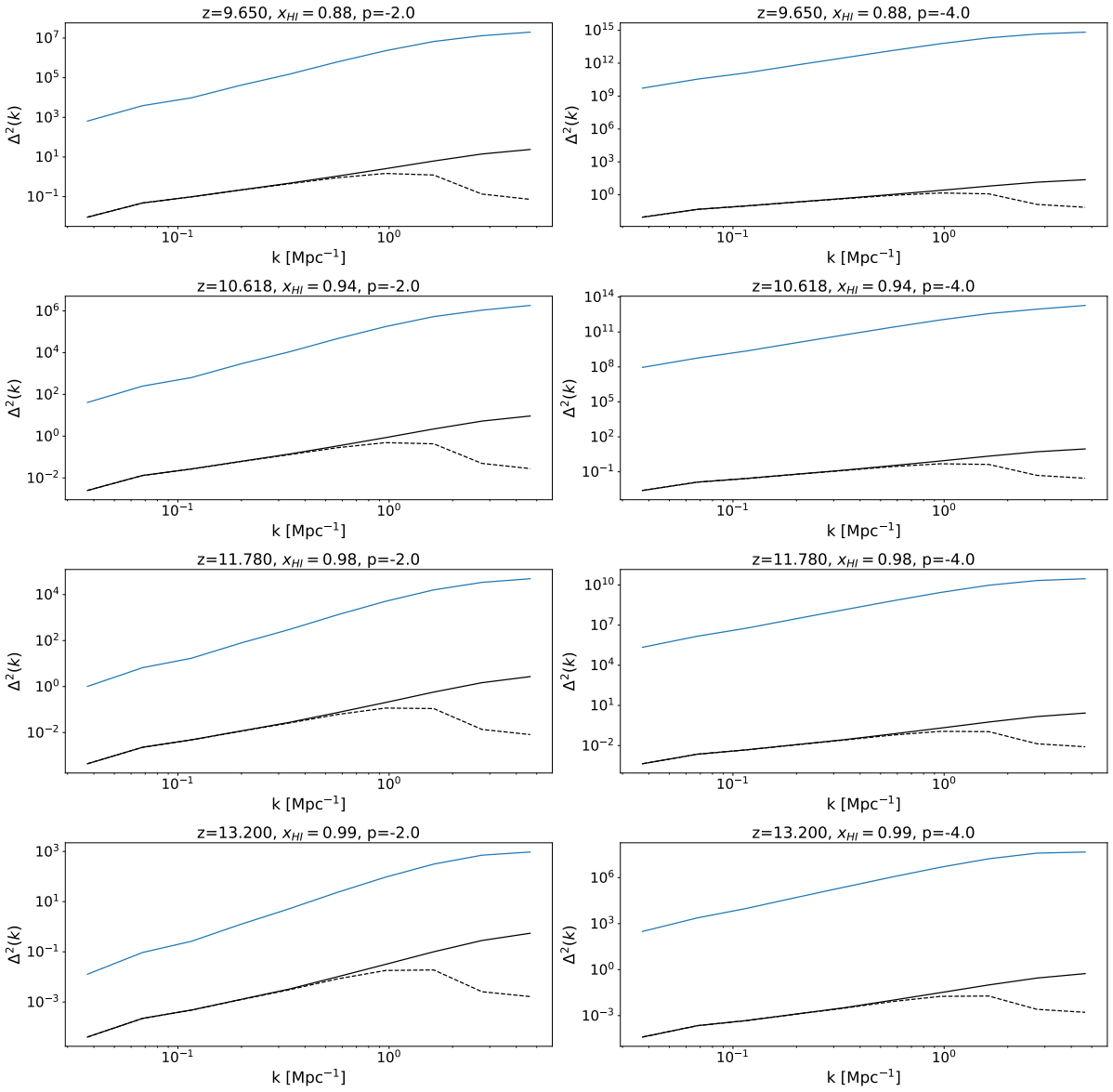


Figure 4.6: Bottom Part: R = 5.60 Mpc, Corresponding Smoothed Field and Marked Fields.

CHAPTER 5

Mark on Neutral fraction field

Moving on to the next type of field, the neutral fraction field. This field is used to identify neutral and ionized regions distinctly. Working with different fields, each with varying features and dynamic ranges, provides better insights into not only the nature of the field but also the field's response to marking. This reinforces the idea that not all fields are identical, and therefore, the mapping of features to free parameters may not be the same. To better understand the response of fields to marks and the free parameters R, f, and p of the mark, it becomes crucial to work with fields that possess different features and dynamic ranges. In this chapter, we aim to understand how marks perform in the context of the neutral fraction field. This field is inherently bimodal in nature, making it easier to isolate ionized and neutral regions. The motivation behind applying the mark, beyond understanding the response (as discussed earlier), is to highlight different features in the field. Apart from having very less dynamic range, there is one more attribute about the neutral fraction field which makes it different from halo density field in the context of applying mark. The neutral fraction field is unit-less in nature. Hence, the corresponding marks (or marked fields) are unit-less as well. This makes it easy to study as well as interpret the results of variation of free parameters on the field. The field is marked as follows:

$$m(\mathbf{x}, x; R, f, p) = \left[\frac{1+f}{1+f+x_{\rm HI}(\mathbf{x}, z, R)}\right]^p \tag{5.1}$$

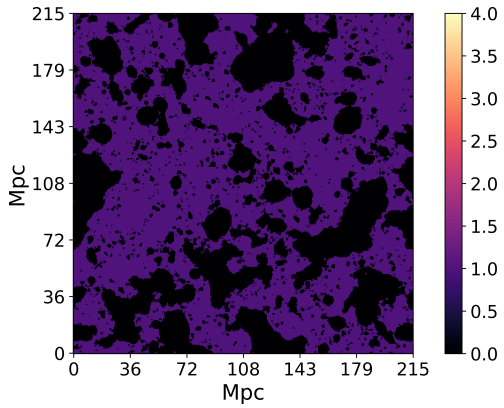
where $x_{\rm HI}(\mathbf{x}, z, R)$ is the smoothed neutral fraction field by a spherical filter of radius R.

5.1 Visual inspection

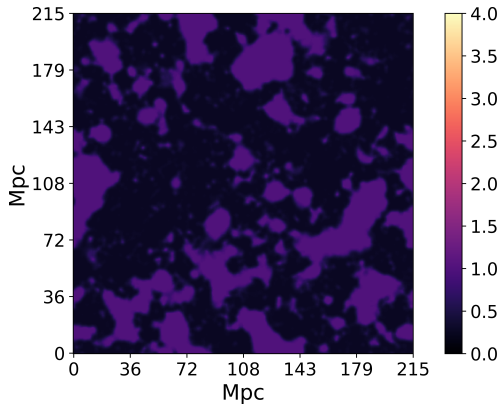
As described earlier, the sign of p determines which features of the fields are enhanced. Visually, the mark behaves as expected. A notable aspect of working with this type of field is its highly bimodal nature, composed of distinct ionized and neutral regions with a small fraction of partially ionized areas. Thus, understanding how the free parameters map to specific features

becomes important to understand for this particular field. Given the distribution's bimodality, it is particularly intriguing to examine which parts of the distribution are highlighted by different ranges of free parameter values. In other words, aim of mark remains the same but the underlying field changes and hence, its response to the whole marking operation changes.

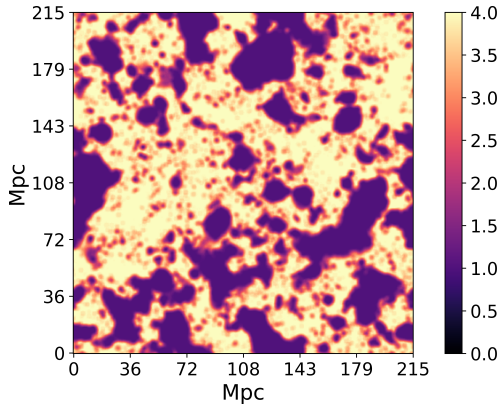
One example of this neutral fraction field are as shown below:



(a) Neutral fraction field at $z = 7.520$.



(b) Mark enhancing low-density regions (ionized).



(c) Mark enhancing high-density regions (neutral).

Figure 5.1: Representation of marks applied to the neutral fraction field at $z = 7.520$, with parameters $m(\mathbf{x}, z = 7.520; R = 1.28\,\text{Mpc}, f = 0.0, p \in \{-2, 2\})$.

In this field as well, as expected, the sign of p still follows the convention where negative and positive values enhances high density and low density regions respectively.

Note that "1" in the denominator in the mark is responsible to the enhancement of the field to such an extent. The fact that the neutral fraction field ranges from 0 to 1 explains the shooting up of values while enhancing high density regions. This is important to understand as this is not the case for halo density field clearly, precisely because of the dynamic range of the field. Here, "1" acts a different type of threshold parameter. This reveals important implication about the design of the mark and how it affects the transformation.

The values of free parameters used for further study the field are as follows:

Parameter	**Values**
f	{-0.1, -0.01, -0.001, 0.0, 0.001, 0.01, 0.1, 1.01}
R (Mpc)	{1.68, 2.80, 3.92, 5.60, 8.40}
p	{-2, 2}

Table 5.1: Values for free parameters used for the neutral fraction field.

where list of redshift snapshots worked are {6.100, 6.518, 6.988, 7.520, 8.129, 8.831, 9.650, 10.618, 11.780, 13.200}

Also, the idea of loss of variation in the context of marked field as discussed in 4 enables us to use only mark to the field while mark field can be studied separately.

5.2 Power Spectrum

The power spectrum of the neutral fraction field, as well as those of the corresponding marks, can be used to understand the variation and response of parameters relative to the unmarked field. Given the very small dynamic range of this field, interpretation is possible, but it becomes difficult to compare all power spectra of marks with different parameters directly to the neutral fraction field. The plot is as shown.

Clearly, from the plots, for all redshifts, the enhancement or suppression relative to the unmarked field remains consistent. The smoothing operation clearly shows its effect—depending on the smoothing scale, the fluctuations are washed out. A sudden jump in amplitude from $f = 0.1$ to $f = 1.01$ is observed, the underlying cause of which is yet to be explained. One immediate reason, of course, relates to the dynamic range.

This also shows that the design of the mark should complement the field in such a way that it highlights features relevant to that field. This opens the opportunity to study various mark designs and understand the corresponding response of the field for a set of free parameters. This idea is used to slightly modify the mark design to better complement the field, as shown in the next sections. In essence, the sensitivity of the structure of the mark with respect to a given field becomes an important area of study.

The fall in the amplitude of the power spectrum is as expected, being proportional to the smoothing scale. Following that, the variation in enhancement or suppression due to the mark becomes evident.

Given the dynamic range, the smoothing has a major impact on the marking operation. This is true for all fields studied here, but it is especially visible for this field due to its very small dynamic range. Hence, this serves as confirmation of the above statement. However, a thorough study is still needed to effectively design the mark.

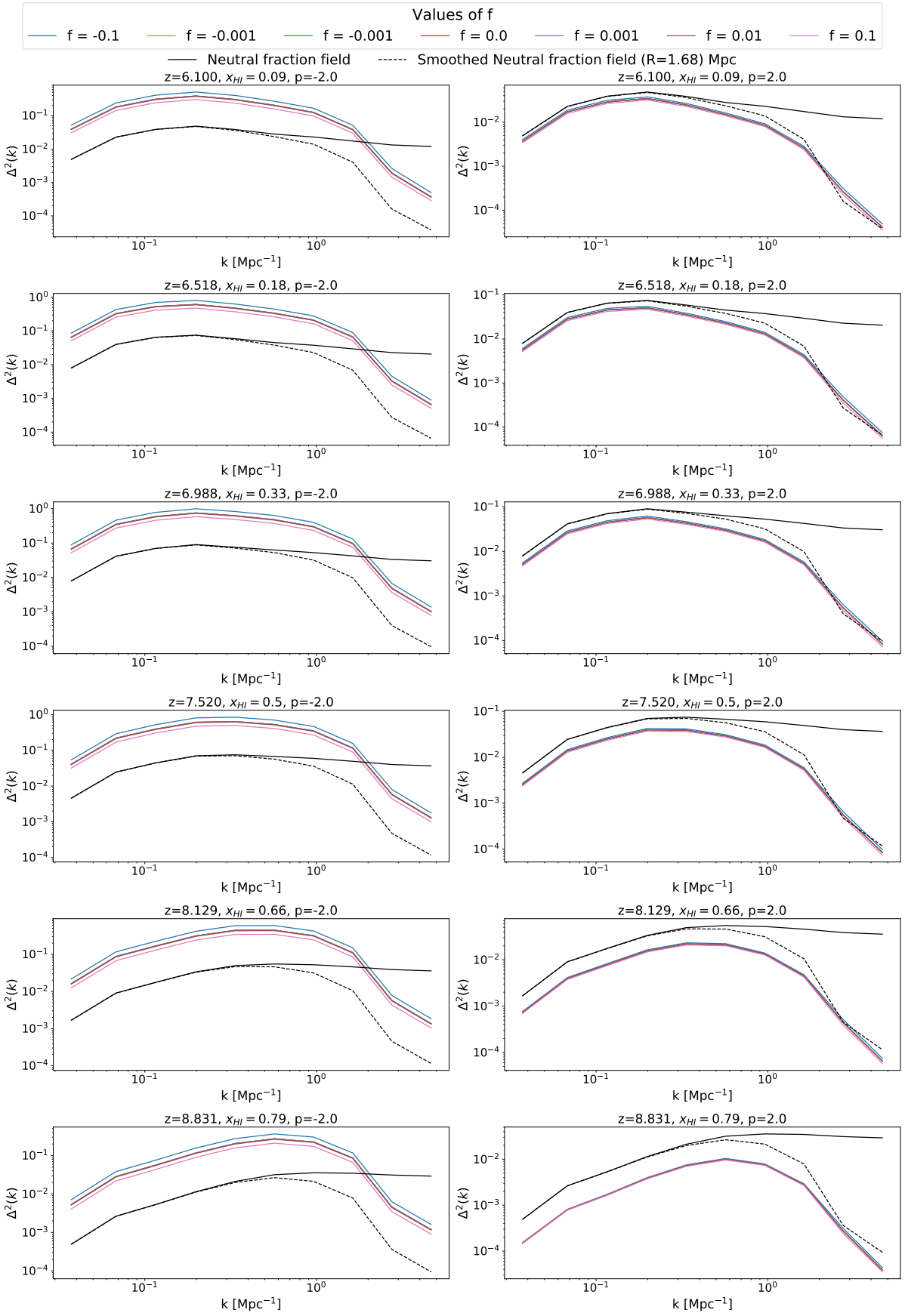


Figure 5.2: R = 1.68 Mpc, Power Spectrum of Neutral Fraction Field and Other Marks.

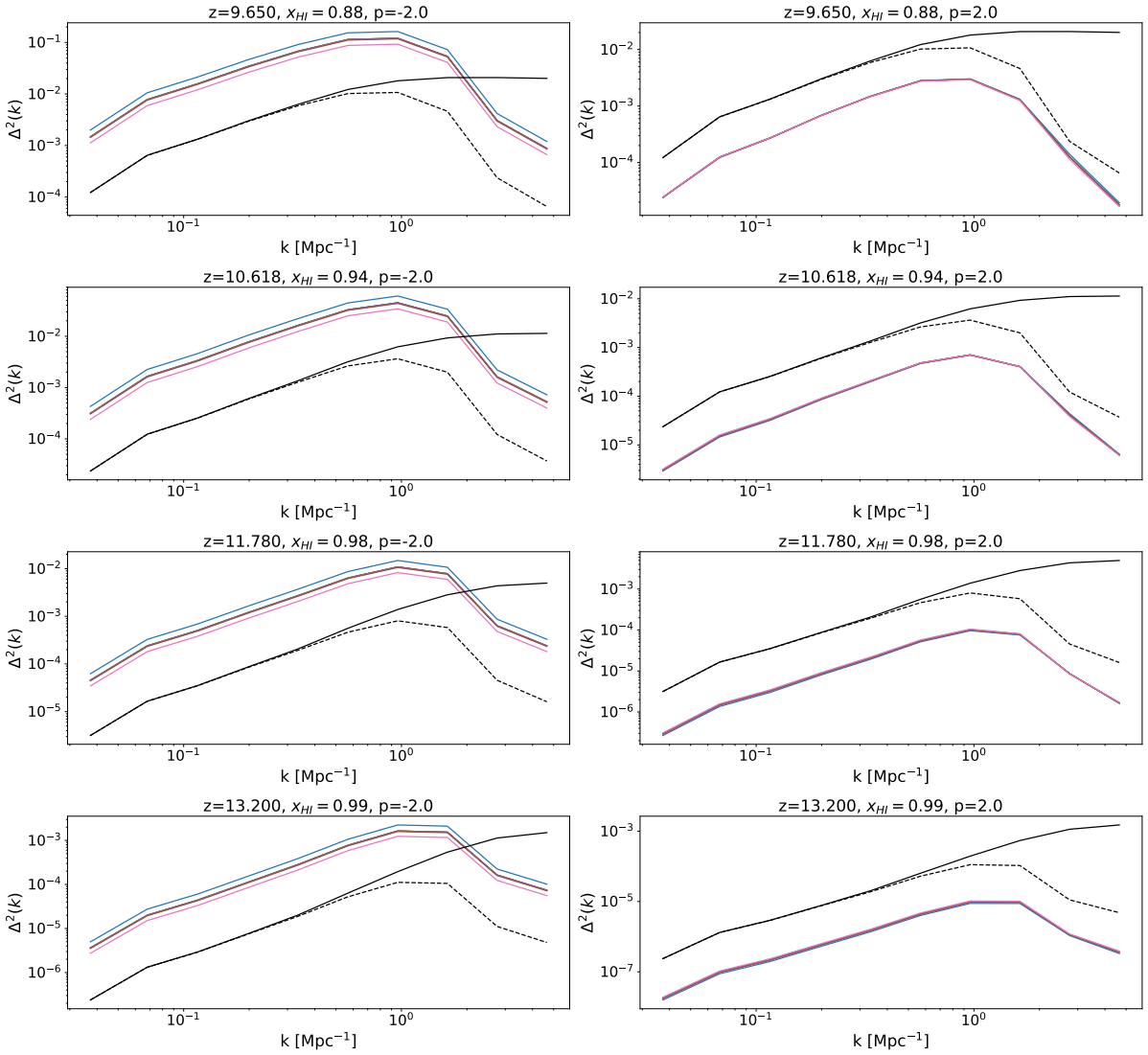


Figure 5.3: Bottom Part: R = 1.68 Mpc, Power Spectrum of Neutral Fraction Field and Other Marks.

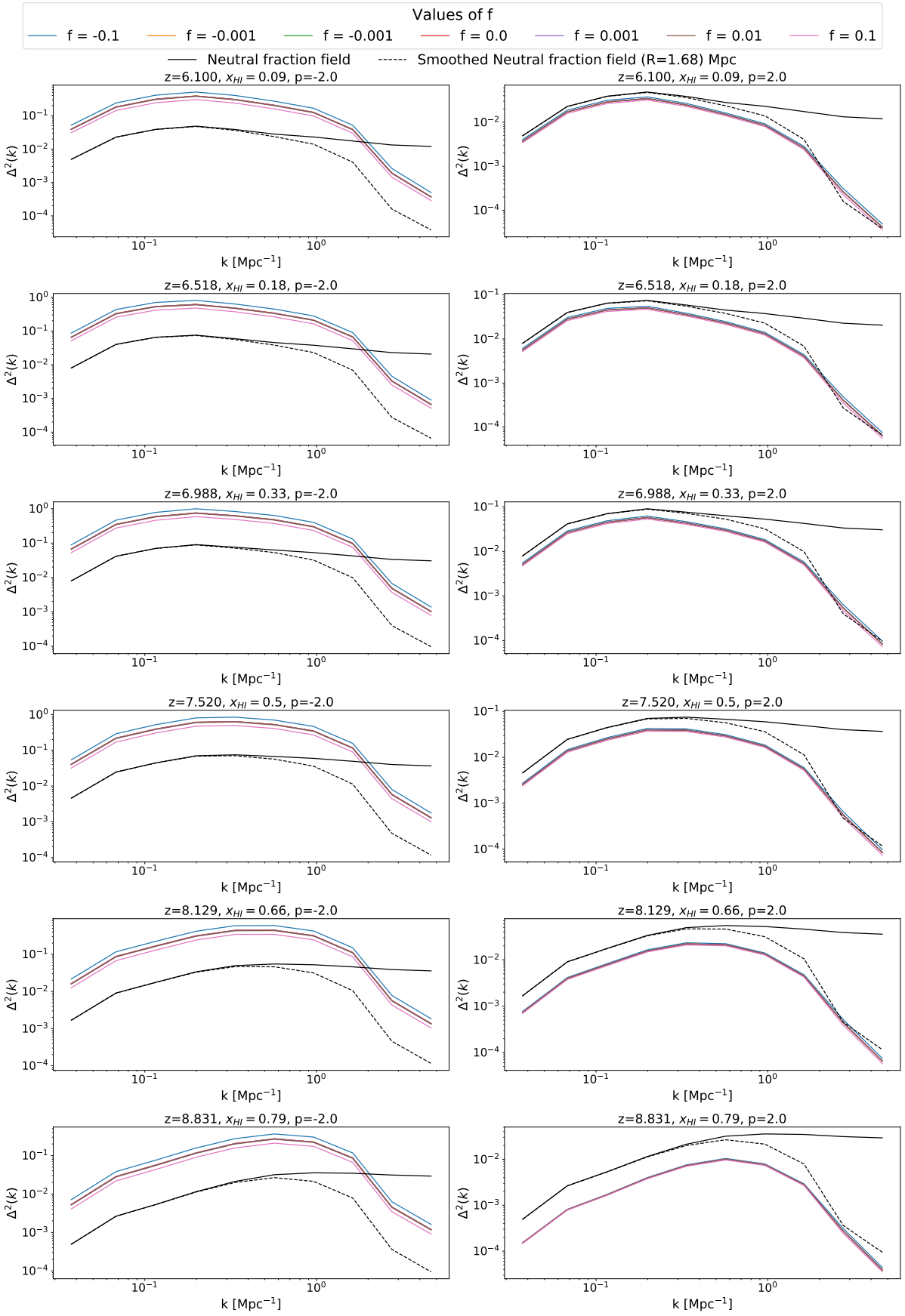


Figure 5.4: R = 2.80 Mpc, Power Spectrum of Neutral Fraction Field and Marks.

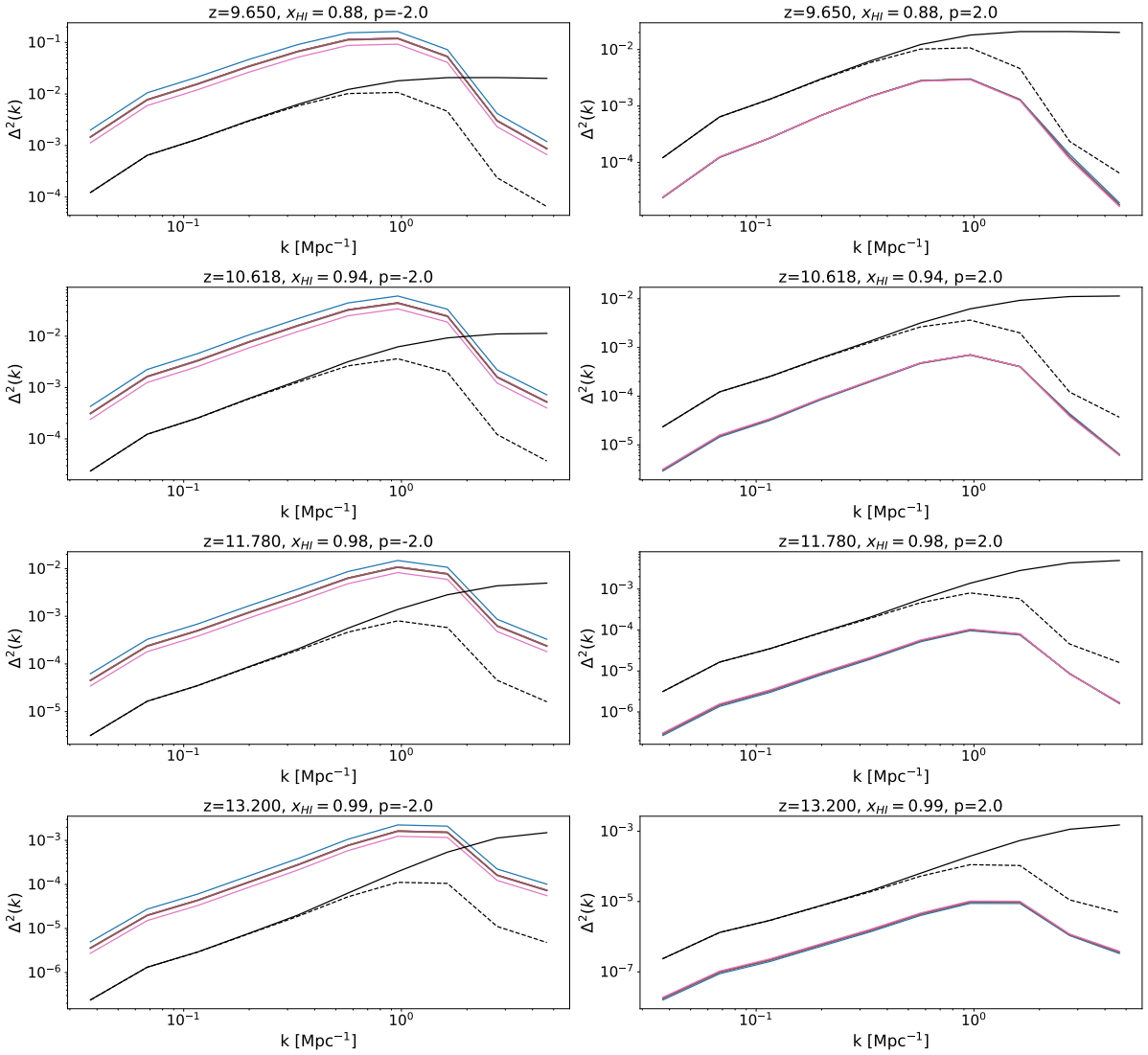


Figure 5.5: R = 2.80 Mpc, Power Spectrum of Neutral Fraction Field and Marks.

5.3 Evolution of Power Spectrum

The goal of this study also includes understanding how the field evolves. To explore this, we now examine the evolution of the power spectrum of both marked and unmarked fields at different k-modes. The k-modes of interest in this study are $k_1 = 0.2\,\mathrm{Mpc}^{-1}$, $k_2 = 0.56\,\mathrm{Mpc}^{-1}$, and $k_3 = 0.96\,\mathrm{Mpc}^{-1}$. Moreover, understanding the relationship between different free parameters and their influence on the evolution of the field can provide valuable insights into the nature of the underlying field and its sensitivity to these parameters. Plots for a fixed value of smoothing scale are shown below. These plots illustrate the evolution of the power spectrum at the fixed k-modes, showing the variation of f for a fixed p and vice versa.

One advantage of using this field lies in its very small dynamic range. As a result, not only is the amplitude of the power spectrum smaller, but the range is also narrow. For this reason, we have plotted the data on a linear scale. The plots mimic the characteristic curve corresponding to inside-out reionization. Since the power spectrum reflects contrast, the curve for a given signal is expected to show how contrast changes over time. It peaks at a certain redshift, where contrast is maximized, and then falls off in both directions towards the ionized and neutral regions as expected for higher and lower redshifts. In this context, the mark can be used to capture the ionization history. Because the neutral fraction field is bimodal, it helps isolate ionized and neutral regions. However, the dynamic range is too small, and thus the amplitude of the power spectrum remains small as expected. Consequently, the mark can be used to enhance this contrast and trace the properties of inside-out ionization by extrapolating the curve using the right parameters.

These plots allow us to observe the evolution of the power spectrum at given k-modes, as well as how it varies with different parameters. The consistency, as well as the degeneracy, between the parameters f and p, makes it a challenging task to conclude which set of parameters maps to the relevant information. However, understanding these relationships will ultimately make interpreting the results easier.

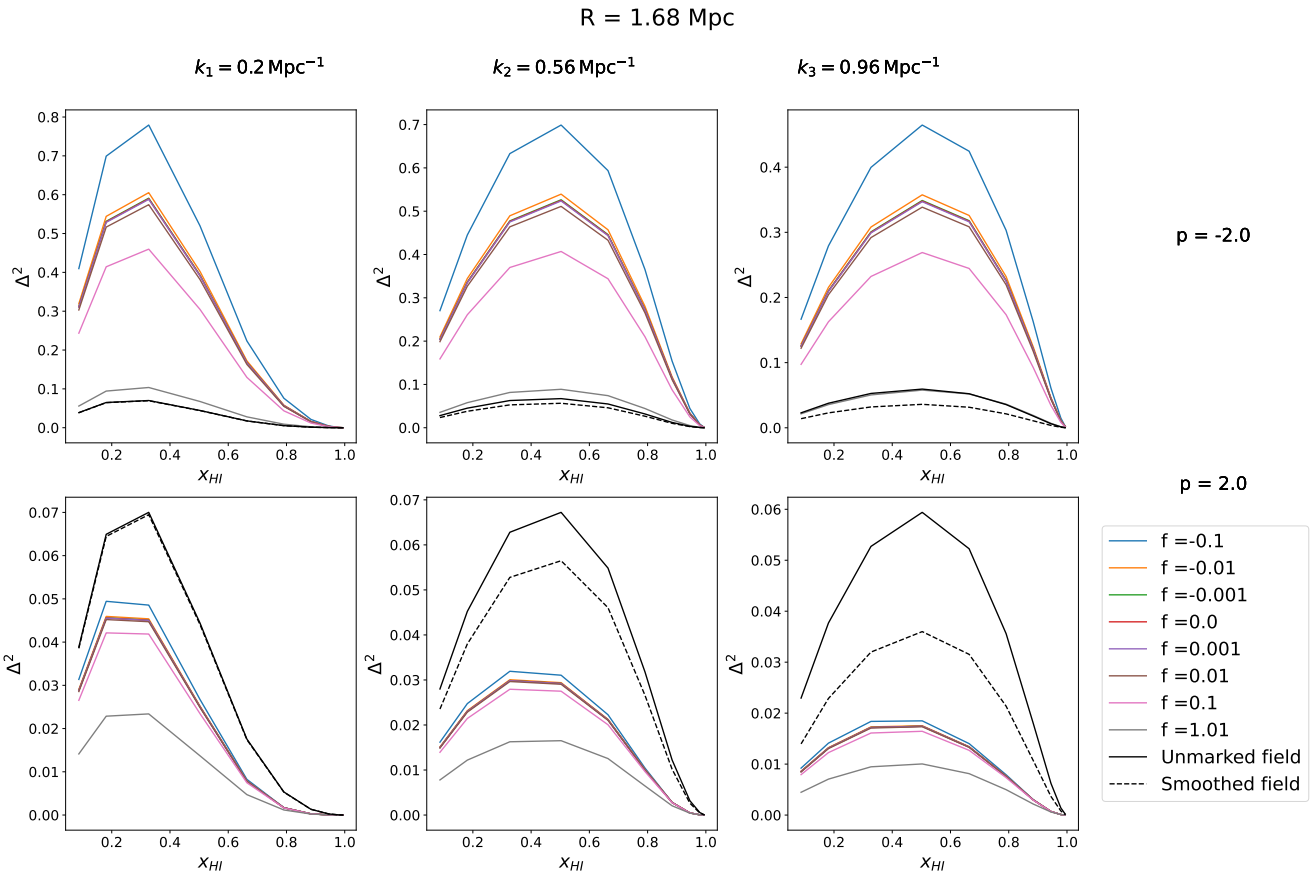


Figure 5.6: R = 1.28 Mpc: Power spectrum evolution of neutral fraction field and other marks at fixed k modes while changing p.

Once again, the separation between the power spectrum of the smoothed field and the unmarked field is as expected: it increases as the smoothing radius grows. However, the choice of smoothing scales used in this study is arbitrary. Thus, it would be interesting to explore specific smoothing scales that correspond to fixed types of fluctuations.

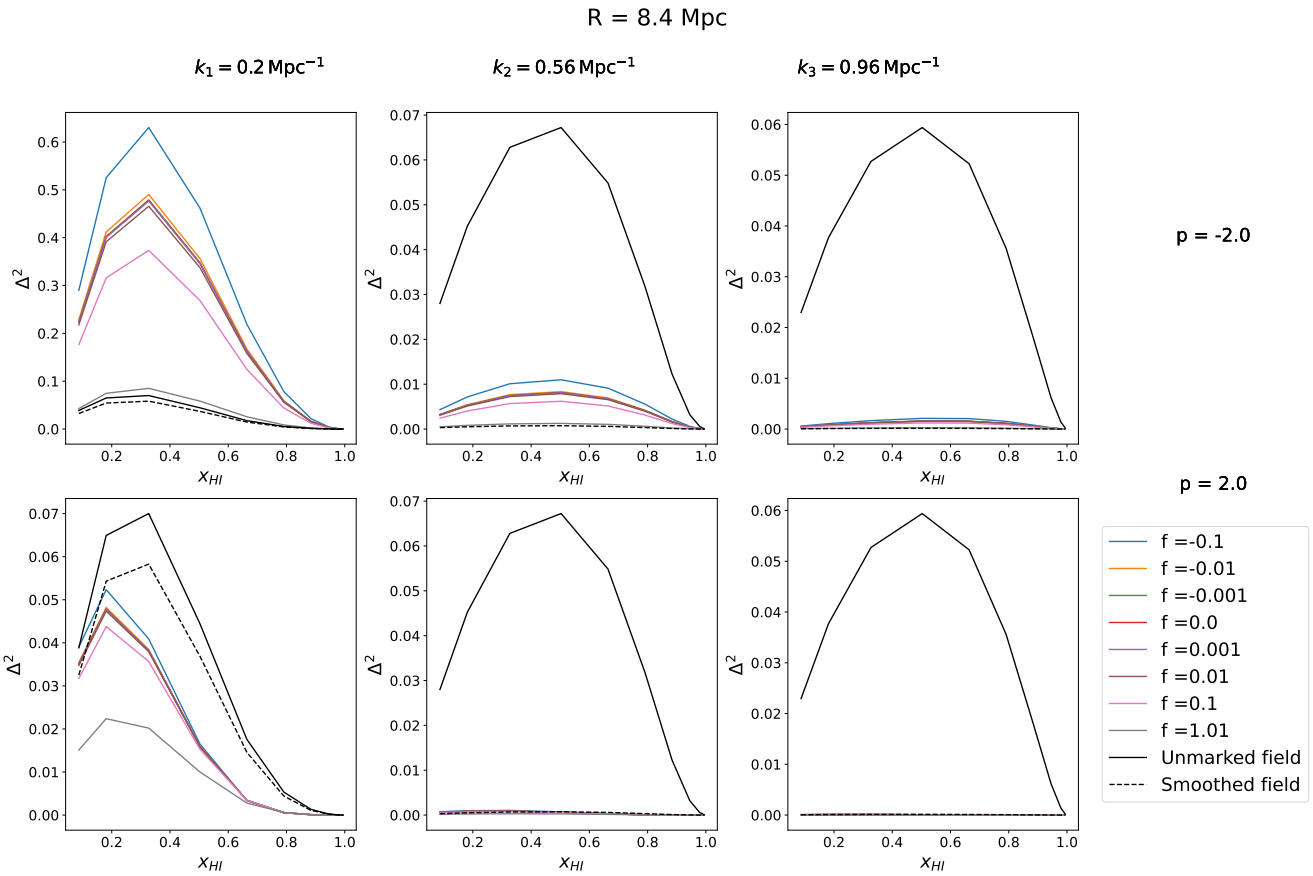


Figure 5.7: R = 8.6 Mpc: Power spectrum evolution of neutral fraction field and other marks at fixed k modes while changing p.

As previously mentioned, studies focused on special values of parameters—particularly the smoothing scale R—can be incorporated into the mark. The response of this particular parameter is already well understood. Additionally, using specific values can help us understand the response of other parameters, since the response of one parameter is known based on the nature of the fluctuations at that length scale. Specifically, the response of the threshold parameter f can be studied more thoroughly.

CHAPTER 6

Mark on HI 21-cm differential brightness temperature fluctuation field

Now since we have "played around" with the ingredient fields, we can use HI 21-cm field for the study. Just like previous cases, HI 21 cm differential brightness temperature is used as field at ten redshift snapshots corresponding to EoR era, from N-body simulation. The output from the N-body simulations [37] [38] are populated by halos using Friends of Friends (FOF) algorithm [38] and are painted by HI 21-cm signal using Reion-Yuga [39] [40] [41] and differential brightness temperature is obtained. Moreover, we convert fields into fluctuations such that before and after the operation of mark, the fields are in same unit and hence, comparison can be meaningful. The process is as follows:

1. Calculating differential brightness temperature fluctuation field

$$\delta_{\mathrm{T_b}}(\mathbf{x}, z) = \frac{\delta \mathrm{T_b}(\mathbf{x}, z)}{\langle \delta \mathrm{T_b}(z) \rangle} - 1$$

2. Smoothing field with a spherical filter of radius R.

$$\delta_{\mathrm{T_b}}(\mathbf{x}, z, \mathrm{R}) = \frac{\left[\delta \mathrm{T_b}(\mathbf{x}, z)\right]_{\mathrm{R}}}{\langle \delta \mathrm{T_b}(z) \rangle} - 1$$

3. Applying the following non-linear transformation to the field: *The Mark*

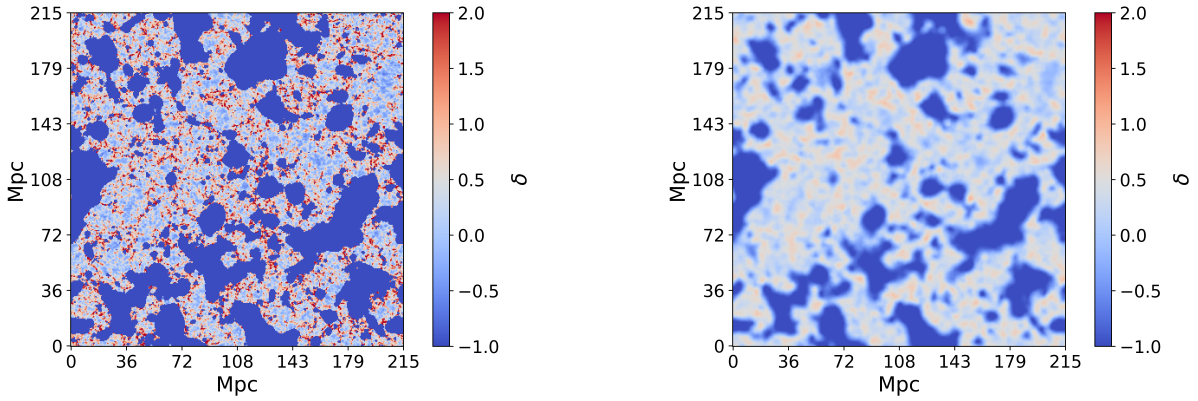
$$m(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1+f}{2+f+\delta_{\mathrm{T_b}}(\mathrm{R}, \mathbf{x}, z)}\right]^p = \left[\frac{1+f}{1+f+\frac{\left[\delta \mathrm{T_b}(\mathbf{x}, z)\right]_{\mathrm{R}}}{\langle \delta \mathrm{T_b}(z) \rangle}}\right]^p \tag{6.1}$$

4. Weighting brightness temperature field with Mark: *The Marked field*

$$\mathbf{M}(\mathbf{x}, z : \mathrm{R}, \mathrm{f}, \mathrm{p}) = \mathrm{m}(\mathbf{x}, \mathrm{z} : \mathrm{R}, \mathrm{f}, \mathrm{p}) \times \delta_{\mathrm{T_b}}(\mathbf{x}, \mathrm{z}) \tag{6.2}$$

6.1 Visual inspection

As described earlier, the idea of Mark, in current study, is used as both transformation and weight. Below given are the snapshots of the one of the field applied for a fixed parameter to illustrate the idea. Following are snapshot of HI 21-cm signal (2D slice) as output of N body simulation at redshift $z = 7.520$, ($\langle x_{HI} \rangle = 0.5$).

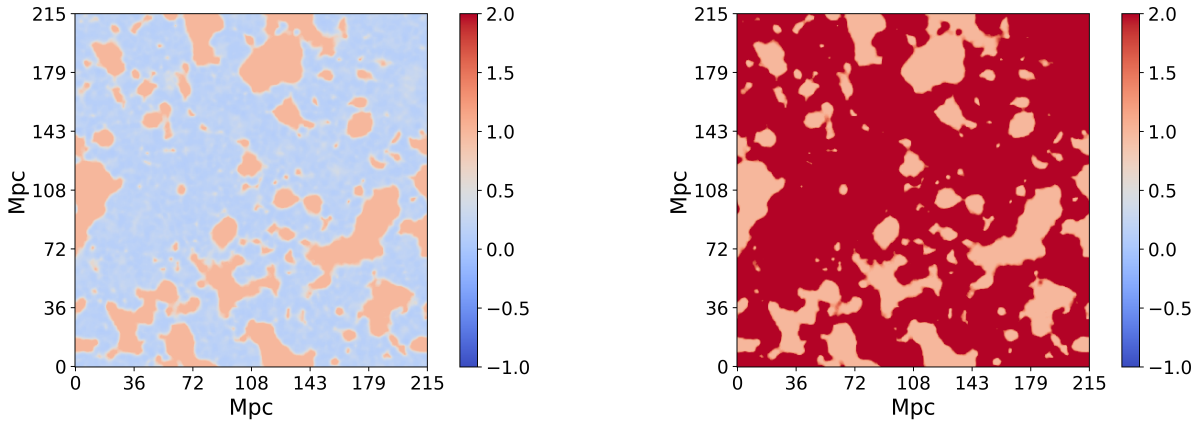


(a) Fluctuation field.

(b) Smoothed fluctuation field (R = 2.8 Mpc).

Figure 6.1: Original and smoothed fluctuation fields at $z = 7.520$.

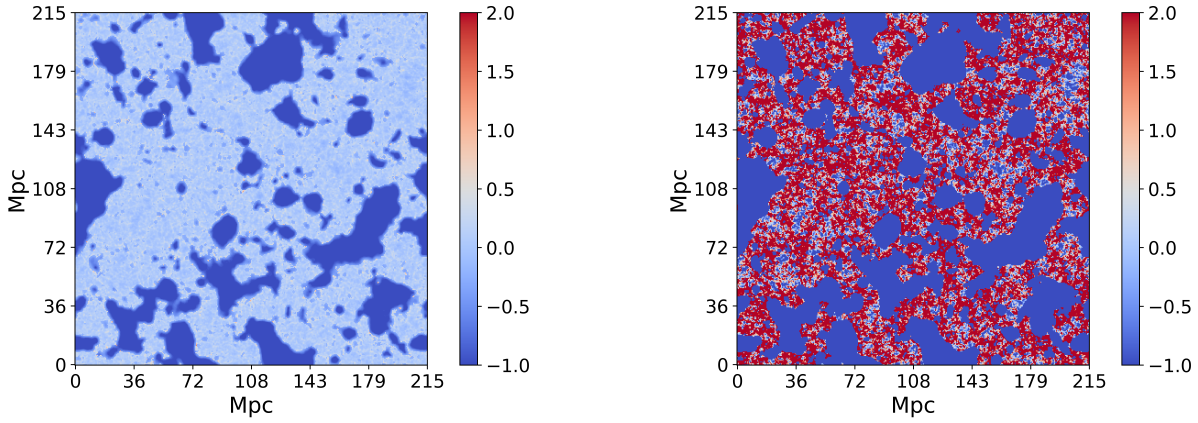
Note that based on 6.1, positive values of p corresponds to an operation which enhances the low "dense" regions (here, lower values of $\delta \mathrm{T_b}(\mathbf{x}, z, \mathrm{R})$) and suppresses high density regions. In contrast to negative values for p, high density regions get enhanced while low density regions are suppressed.



(a) Mark: $p = 2$.

(b) Mark: $p = -2$.

Figure 6.2: Mark fields for R = 2.8 Mpc, $f = 0$, and $p \in \{-2, 2\}$ at $z = 7.520$.



(a) Marked field: $p = 2$.

(b) Marked field: $p = -2$.

Figure 6.3: Marked fields for R = 2.8 Mpc, $f = 0$, and $p \in \{-2, 2\}$ at $z = 7.520$.

Visually, the transformation is working as expected. All plots are plotted at fixed range to compare with original field. As seen from the figure, for a set of free parameters, mark is able to pick up properties of the map from where, information can be extracted. Particularly, Mark is able to make the distinction between ionized and neutral regions clearer whereas the mark field is amplifying the fluctuations in the map. In the following sections, we will use both of these idea independently for different problem statement.

6.2 Power Spectrum

One of the immediate quantities that can be calculated after applying this mark, just like in previous sections, is non other than the power spectrum. The power spectrum for mark $f = 0, p = \pm 2$ and different R is as follows. For reference, power spectrum of the smoothed field, in dashed line, is also plotted to understand how differently mark is affecting the field than smoothing the field.

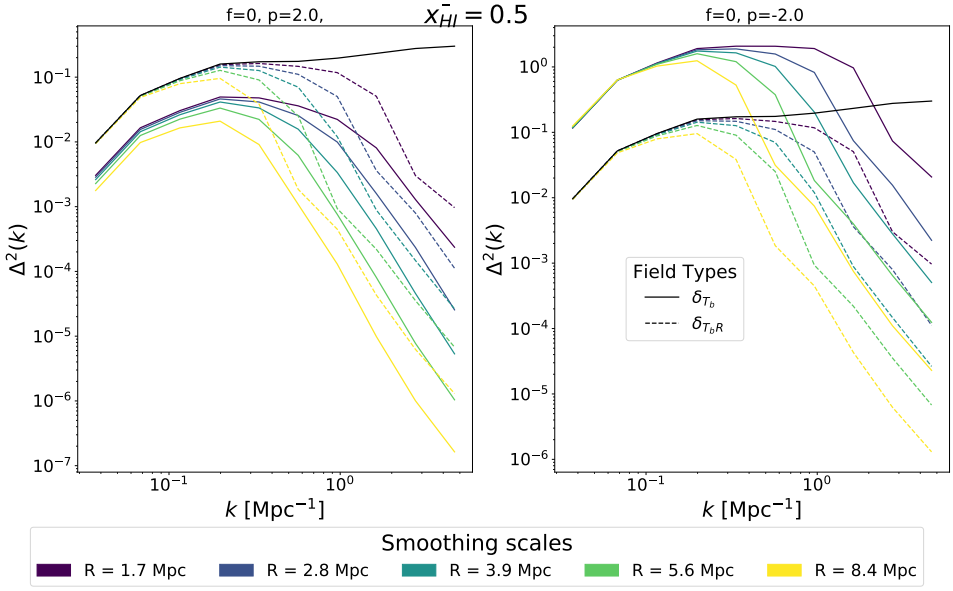


Figure 6.4: Comparison of Normalized power spectra of $\delta \mathrm{T_b}(\mathbf{x}, z)$ (solid black line), smoothed field $\delta \mathrm{T_b}(\mathbf{x}, z, \mathrm{R})$ (dashed lines) and corresponding Marks (solid lines).

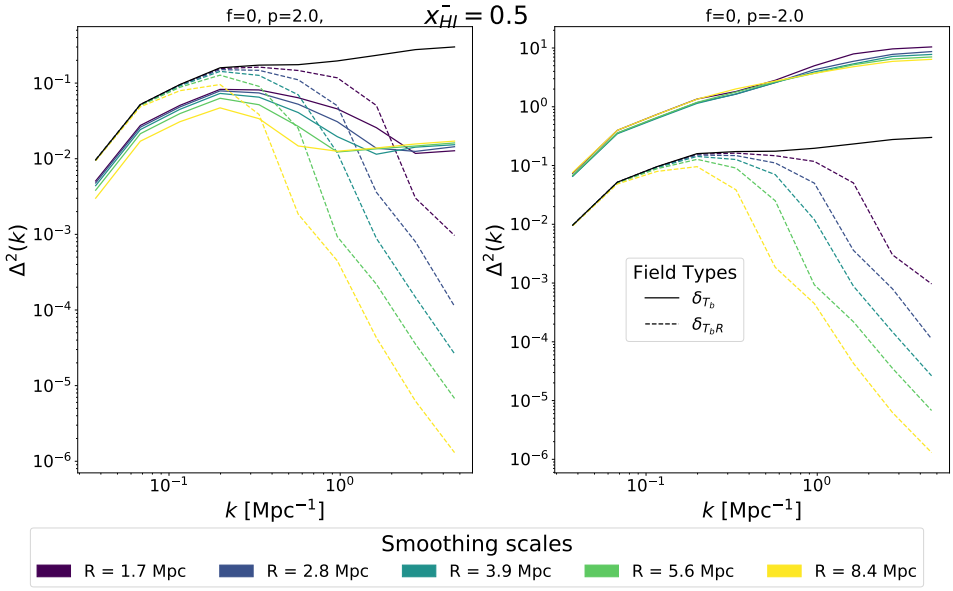


Figure 6.5: Comparison of Normalized power spectra of $\delta \mathrm{T_b}(\mathbf{x}, z)$ (solid black line), smoothed field $\delta \mathrm{T_b}(\mathbf{x}, z, R)$ (dashed lines) and corresponding Marked fields (solid lines).

The solid black line is the power spectra of $\delta T_b(\mathbf{x}, z)$ and dashed line are the power spectra at smoothed field with different $\delta T_b(\mathbf{x}, z, R)$ radius corresponding to smoothing radius used in Mark. As clearly seen from the plots, the Mark (as well as the Marked field) is able to enhance and suppress features of the field, resulting in a change in the power. However, since the marked field is the result of the product of the mark with the original field, loss of variation is observed in the power spectrum of the Marked field as it follows the shape of the unmarked field. This is evident here in this field because of relatively smaller dynamic range which was not visible in 4 section.

The rest or power spectrum for all redshifts shows that the shape of power spectrum follows that for smoothed field however, it is able to vary the amplitude of power spectrum consistently. Hence, it is expected to maintain similar trend of consistency throughout the evolution of signal which is discussed in following sections. Moreover, negative values of f tend to increase the fluctuations. Thus, a clear degeneracy is observed between choice of f and p. This may lead to very similar power spectrum form for very different values of parameters. This once again points to an important fact that about how finding the right parameters about the field which can map to unique, relevant and new information lies at the heart of the study. This may seem a bit problematic but if we recall the interpretations of f and p, enhancing amplitude too much can significantly increase the noise as well. However, if cleverly chosen, free parameters may be able to map the same features by not enhancing noise that much. The fluctuation may be able pick out in the same way but since the parameters are different, interpretation about the field changes. Hence, interpreting the free parameters is also important to understand the response of fields to those maps. However, we leave incorporating noise for future studies.

6.3 Evolution of Power spectrum

The goal of this study is to understand the attributes of the HI 21-cm differential brightness temperature and hence, now we look at the evolution of power spectrum of marked and unmarked fields at different k modes. The k modes within the interest of study are $k_1 = 0.2\text{Mpc}^{-1}$, $k_2 = 0.56\text{Mpc}^{-1}$, $k_3 = 0.96\text{Mpc}^{-1}$. Moreover, to understand which free parameter leads to what kind of evolution can also provide useful insights not only about the nature of the underlying field but also its sensitivity to free parameters. Plots for a fixed value of smoothing scale is as shown. Moreover, the list of free parameters used for this field are as follows:

f	R **Mpc**	p
{-0.1, -0.01, -0.001, 0.0, 0.001, 0.01, 0.1, 1.01}	{1.68, 2.80, 3.92, 5.60, 8.40}	{-2, 2}

Table 6.1: Values for free parameters worked for the HI 21-cm brightness temperature fluctuation.

where list of redshift snapshots worked are {6.100, 6.518, 6.988, 7.520, 8.129, 8.831, 9.650, 10.618, 11.780, 13.200}

Clearly, from the plot, it is evident that the mark is working consistently. In other words, the enhancement or suppression of the field which corresponds to increase and decrease in amplitude of power spectrum is consistent. Each set of parameter f, p , at least for small smoothing scales, with respect to unmarked field and other marked field changes the amplitude of power spectrum in a fixed way. Moreover, the unmarked field is has dynamic range from -1 to 2. For this dynamic range, the negative f values increases the amplitude of the power spectrum with respect to positive values. Concluding that this is the general case for any type of field corresponding to any dynamic range need not be accurate. The response of field to the mark clearly depends on the nature of the field as well as the design of the mark. Hence, for this particular design of mark, in further sections, we have applied mark to Neutral fraction field having very small dynamic range and Halo density field, having very large dynamic range. Hope is that studying about these fields for similar design of mark should provide the response of the field to given mark. Moreover, the shape of power spectrum while observing response of field to different p values for a given f value stays more or less consistent. The plot for a single smoothing scale is as follows. The results and key observation hold valid for all smoothing scales used to analyze as given in 6.1

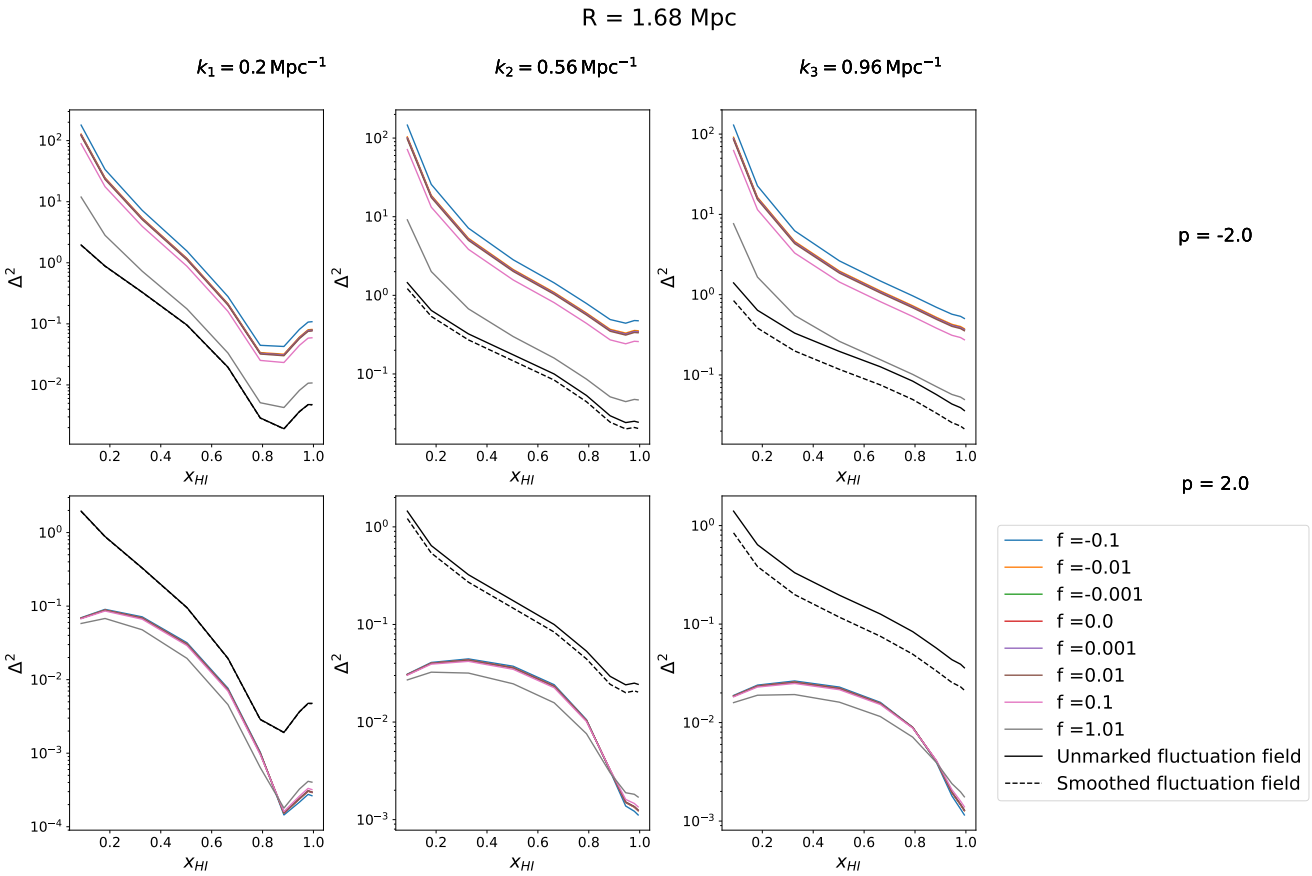


Figure 6.6: R = 1.68Mpc: Comparison of evolution of Normalized power spectra of $\delta \mathrm{T_b}(\mathbf{x}, z)$ (solid black line), smoothed field $\delta T_b(\mathbf{x}, z, R)$ (dashed lines) and corresponding Marks (solid lines). for a given smoothing scale at fixed k modes. Observing variation in Power spectrum by changing f.

Note that these plots are plots of normalized power spectrum as a function of mean neutral fraction, for a given k mode which shows how power spectrum at a given k mode evolves with time. Clearly, as suspected, the shape of power spectrum remains more or less fixed while changing f for a given R value where $p = -2$.

Ideally, one would expect the characteristic curve in the evolution of power spectrum for inside-out reionization as in 5. This is clearly not the case here particularly because of the functional form of the mark where we are dividing by the mean of the brightness temperature field. This division at later stages of EoR corresponds to division to small number and results into rise in power as shown. Moreover, one expects the same for Marked field as defined in 6.1 as we are giving weights to fluctuation field as well. But this artifact corresponds to structure of the mark and not to marked field. To show this, we use the same mark but use the mean subtracted field to give weights to. The plot is as shown:

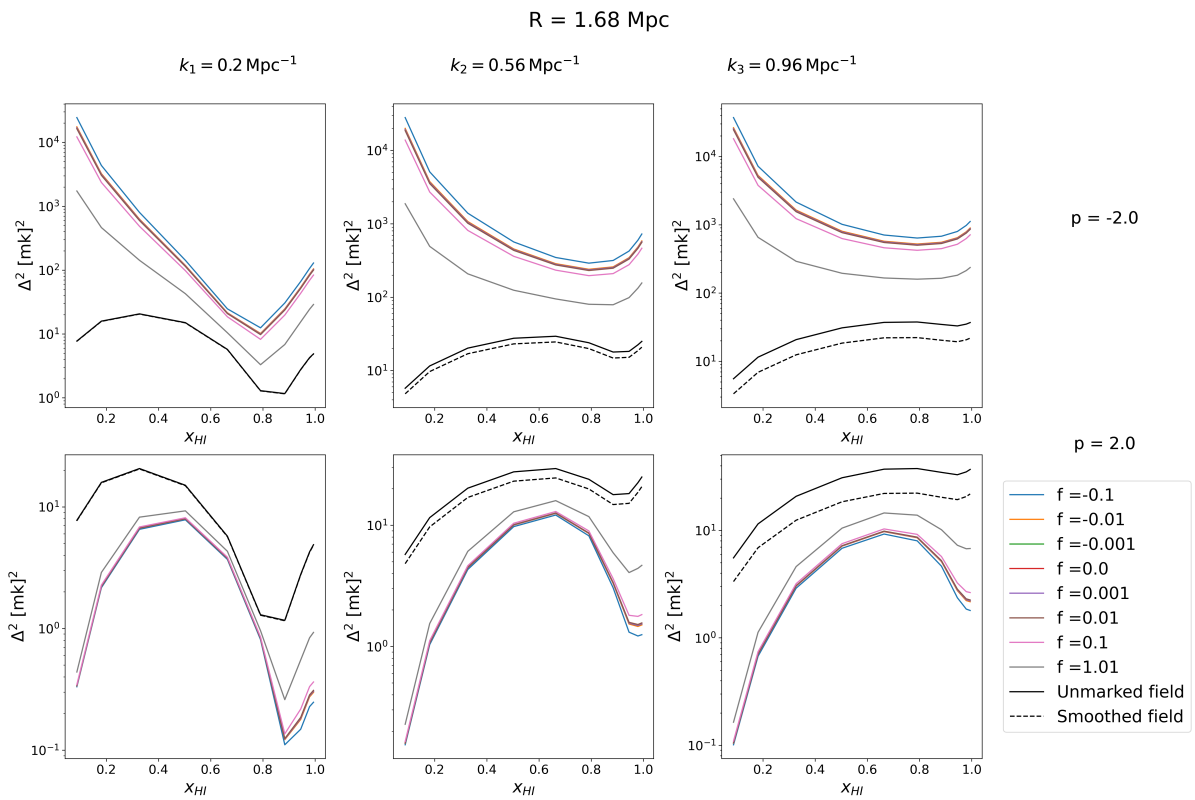


Figure 6.7: R = 1.68Mpc: Comparison of evolution of Normalized power spectra of $\delta \mathrm{T_b}(\mathbf{x}, z)$ (solid black line), smoothed field $\delta T_b(\mathbf{x}, z, R)$ (dashed lines) and corresponding Marked fields to mean subtracted field. (solid lines). for a given smoothing scale at fixed k modes. Observing variation in Power spectrum by changing f.

This motivates us to use different functional forms of the mark depending on the problem statement. This makes the study of marked statistics rich, comprising varied problem statements and their solutions.

6.4 Exploring different functional forms for Mark

In principle, the Mark can take any functional form and hence there is no reason to use a fixed functional form. Moreover, the functional form in 6.1 included dividing by the mean of the brightness temperature field at a given neutral fraction. This leads to the loss of characteristic curve in the evolution of power spectrum at specific. This motivates us to extend the idea and include the functional form of mark and to try such functional forms where dividing by the mean is avoided. This gives us an opportunity to understand more about the general functional form and hence, to understand the mark and marked field as completely different field.

Moreover, the analysis in 4 and 5 shows the importance of using the idea of Mark such that units before and after the transformation of field retain their consistency. Doing this, one can ensure that the comparison between fields is meaningful.

We take these ideas into consideration and propose the following functional forms which are studied both as Mark and Marked field.

$$m_1(\mathbf{x}, z : R, f, p) = \left[\frac{1+f}{1+f+\left[\mathrm{dT}(\mathbf{x}, z)\right]_{\mathrm{R}}}\right]^p \tag{6.3}$$

$$m_2(\mathbf{x}, z : \mathrm{R}, f, p) = \left[\frac{1+f}{1+f+\left[\delta\mathrm{T_b}(\mathbf{x}, z)\right]_{\mathrm{R}}}\right]^p \tag{6.4}$$

$$m_3(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1+f\cdot\langle\delta\mathrm{T_b}(z)\rangle}{f\cdot\langle\delta\mathrm{T_b}(z)\rangle+\left[\mathrm{dT}(\mathbf{x}, z)\right]_{\mathrm{R}}}\right]^p \tag{6.5}$$

where $\mathrm{dT}(\mathbf{x}, z) = \delta\mathrm{T_b}(\mathbf{x}, z) - \langle\delta\mathrm{T_b}(z)\rangle$ is the mean subtracted brightness temperature field.

Note that 6.3 and 6.4 are functional forms such that the transformed field's unit depends on the choice of free parameter, enabling us to explore a more general functional form of mark and opportunity to study mark a completely different field. We use the functional forms as given by 6.5 and 6.6 in context of both Mark and Marked field precisely because of the fact that the units of these functional forms are independent of the choice of free parameters. These functional forms are obtained to make the mark unit less to use them as weights while not dividing by the mean value like in 6.1 and hence, the fields before and after can be compared. In the context of Marked field, we weigh these functional forms to mean subtracted field. This functional form is very close to that of used in [11].

In addition, the transformation process begins with smoothing the field. This step helps us isolate fluctuations occurring above a certain length scale. Based on this idea, we can consider assigning marks—not to the raw mean-subtracted field—but to its *smoothed* version. This approach allows us to selectively enhance fluctuations at specific scales by up-weighting them after identifying them through smoothing.

Each of these functional forms for the mark can be interpreted as a *Taylor expansion* in powers of the field. Choosing specific values for parameters like f and p is effectively equivalent to selecting particular coefficients in this expansion.

The functional form that best highlights features in the field is the one that offers the most effective control over the amplitudes of these expansion coefficients—much like adjusting a dial. By tuning these "dials" or exploring different combinations of them, we can target and enhance specific features in the field map.

In addition to these, we have also worked with another functional form which is very similar to that in 6.5. It is given as follows:

$$m_4(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1 + f \cdot \mathrm{dT}(\mathbf{x}, z)}{f \cdot \mathrm{dT}(\mathbf{x}, z) + \left[\mathrm{dT}(\mathbf{x}, z)\right]_{\mathrm{R}}} \right]^p \tag{6.6}$$

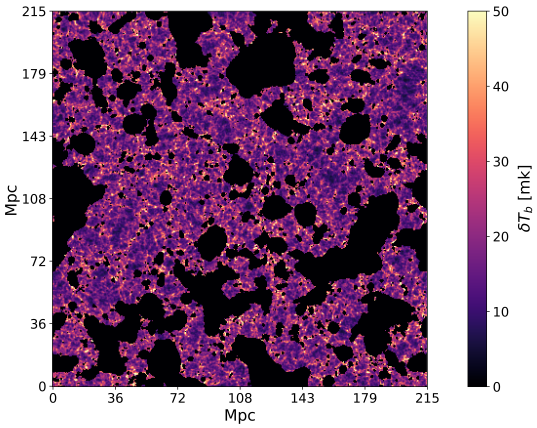
where $\mathrm{dT}(\mathbf{x}, z) = \delta \mathrm{T_b}(\mathbf{x}, z) - \langle \delta \mathrm{T_b}(z) \rangle$ is the mean subtracted brightness temperature field.

We utilize both the concepts of the *mark* and the *marked field* in the context of Figures 6.5 and 6.6. As described earlier, the marked field is constructed by assigning marks as weights to a given field.

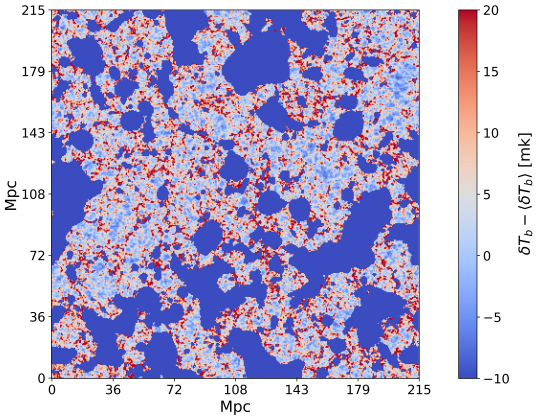
A key step in defining the mark involves smoothing the field and then applying this mark back to the original field. This process effectively restores fluctuations that were lost during the smoothing step, thereby "reinstalling" them into the field. This type of functional form can help us probe the correlation between a point in the intergalactic medium (IGM) and its surroundings.

On the other hand, if we want to intentionally suppress fluctuations below a certain length scale, we can apply the mark to the *smoothed field* itself rather than the original one. This ensures that only large-scale fluctuations are retained in the resulting marked field. We explore both of these approaches in our analysis.

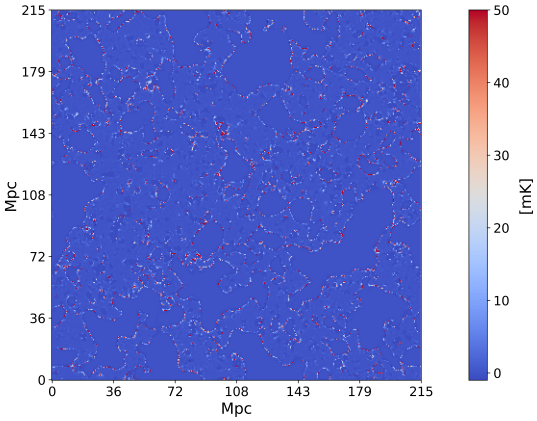
The following plots shows visual inspection of these marks for a set of free parameters:



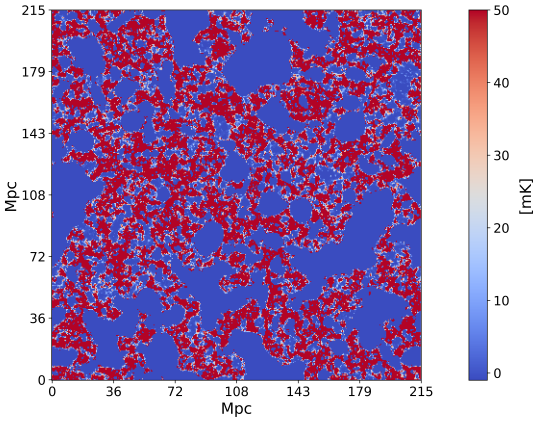
(a) Brightness temperature field



(b) Mean subtracted field

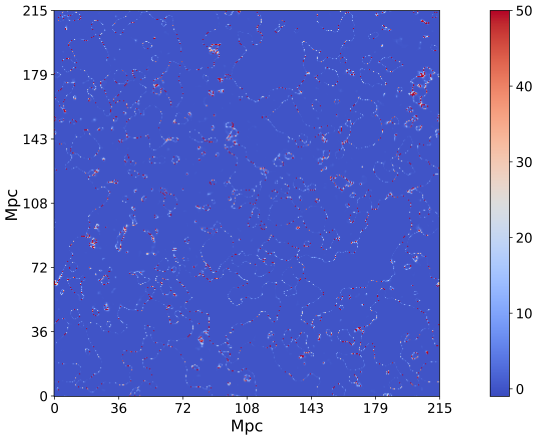


(c) 6.5 for $f = 0.01$, $p = 2$ on marked mean-subtracted field

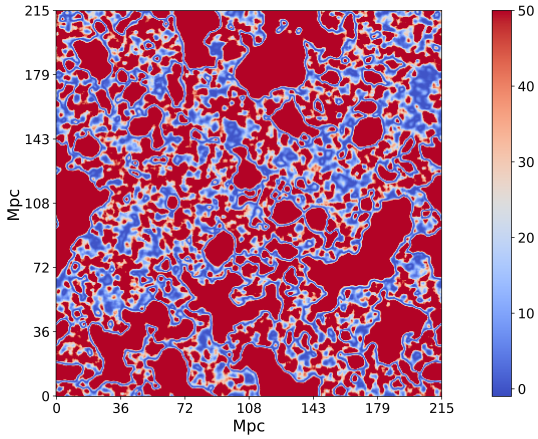


(d) 6.5 for $f = 0.01$, $p = -2$ on marked mean-subtracted field

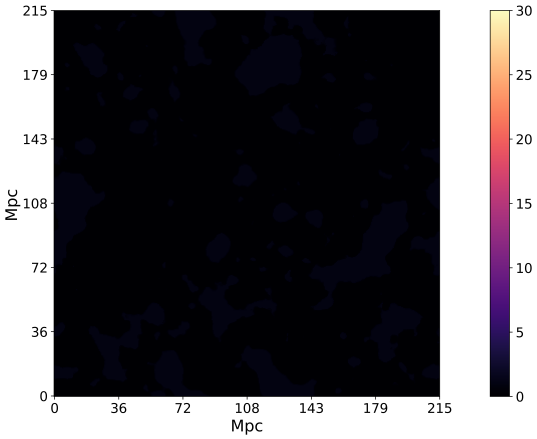
Figure 6.8: Overview of the brightness temperature field, mean-subtracted field, and 6.5 applied to the marked mean-subtracted field for R = 1.68 Mpc.



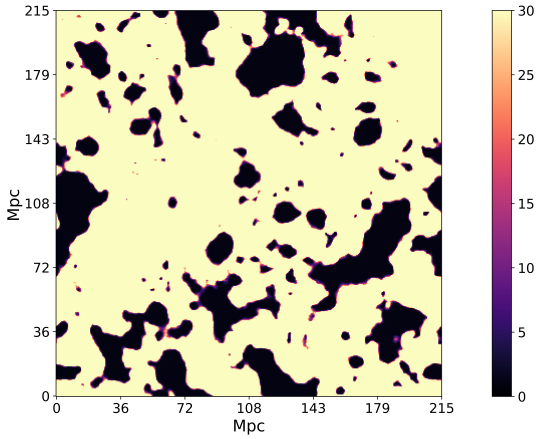
(a) 6.3 for $f = 0$, $p = 2$.



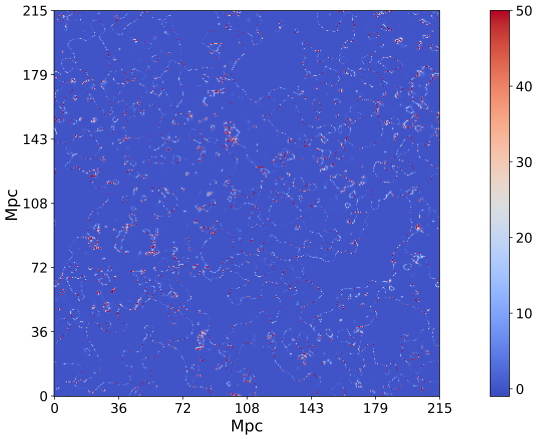
(b) 6.3 for $f = 0$, $p = -2$.



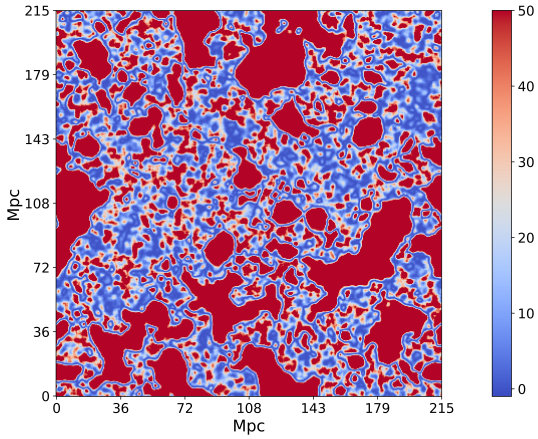
(c) 6.4 for $f = 0$, $p = 2$.



(d) 6.4 for $f = 0$, $p = -2$.



(e) 6.5 for $f = 0.01$, $p = 2$.



(f) 6.5 for $f = 0.01$, $p = -2$.

Figure 6.9: Visual representation of 6.3, 6.4, and 6.5 variants for R = 1.68 Mpc.

Here, in the cases of 6.3, 6.4, we treat the mark as a separate entity with its own units. This allows us to use the marked power spectrum as a fundamentally different kind of statistic. To maintain consistency in units and improve interpretability, we use the form in 6.5—which applies the mark to a mean-subtracted field. This helps us better understand the effect of the

mark.

As seen in **??**, the mark has a clear impact: it enhances the amplitude of fluctuations. This is crucial because it not only preserves unit consistency, but also highlights features in the field that carry physical information. By doing so, we make it possible to build an estimator that can infer astrophysical parameters from the data.

The reasoning behind this is as follows: the fluctuations in the signal (captured by the power spectrum) are already shaped by underlying astrophysical parameters $M_{min}, N_{ion}, R_{mfp}$. So, in principle, the power spectrum can be used as an estimator for these parameters. The role of the mark is to amplify those fluctuations, making the power spectrum more directly sensitive to the physics we aim to study. This gives the Marked Power Spectrum a potential advantage over the standard power spectrum.

However, boosting fluctuations can also be done in the approach described in 6.1, where the mark is applied to a mean-subtracted field. This raises an important question: among these different functional forms of the mark, which one performs better for our specific science case?

To answer this, we need a way to quantify how much information is carried by each mark and marked field. For this, we turn to a Fisher information analysis.

6.5 Fisher information analysis

To quantify the relevant information in Marks and Marked fields, we perform Fisher metric analysis. This is done for different function forms of marks and marked fields, for the chosen set of free parameters.

The Fisher matrix at a fixed redshift z, summing only over k-modes, with its relation with the covariance matrix as described in [42] given by:

$$\mathcal{F}_{\alpha\beta}(z) = \sum_{i,j} \left[\frac{\partial \Delta^2_{21}(k_i, z)}{\partial q_\alpha} \left[\mathbf{C}_{ij}(z)\right]^{-1} \frac{\partial \Delta^2_{21}(k_j, z)}{\partial q_\beta} \right] \tag{6.7}$$

Assuming uncorrelated errors between k-bins, the covariance matrix becomes diagonal:

$$\mathbf{C}_{ij}(z) = \delta_{ij}\, \sigma_{\Delta^2}(k_i, z) \cdot \sigma_{\Delta^2}(k_j, z) \tag{6.8}$$

The Cramér-Rao theorem [43],[44] states that any unbiased estimator for the parameters will produce a covariance matrix that is no more accurate than $\mathcal{F}^{-1}$.

$$\sigma^2(\theta_\alpha) \geq (\mathcal{F}_{\alpha\alpha})^{-1} \tag{6.9}$$

The lower the marginalized error, the better a certain parameter can be constrained by the considered statistic(s). Thus, the Fisher matrix gives the variance of an optimal unbiased estimator for the parameter θ_α

We use Fisher information analysis to forecast the minimum errors in estimating the astrophysical parameters as follows by the Marked Power Spectrum as an estimator of these astrophysical parameters for various functional forms of mark and corresponding mark field. This analysis serves as a metric to quantify the relevant information content in the mark.

Description of astrophysical parameters:

M_{min}: The minimum halo mass to host and sustain star formation. Higher the value of this parameter, lesser the number of halos in the simulation will support star formation and the reionization will be delayed as there are less photons at an instant to ionize the IGM complementary to the case where its value is relatively small.

N_{ion}: This parameter corresponds to the number of ionizing photon per unit baryon. More generally, the "simulation model" assumes the number of ionizing photon leaving the halo is directly proportional to the mass of halo. Hence, this parameter servers as a proportionality constant composed of other different astrophysical parameter such as such as the star formation efficiency $f_\star$, escape fraction of ionizing photons from a halo f_{esc}, given as follows:

$$N_\gamma(M_h) = N_{ion} \frac{M_h}{\Omega_b} \frac{m_p}{\Omega_m} \tag{6.10}$$

Here, N_{ion} is a dimensionless constant quantifying the number of ionizing photon per unit baryon as given in [45]. Higher the number, more is the number of ionizing photons which are available to ionize the medium.

R_{mfp}: This astrophysical parameter quantifies the mean free path traveled by the ionizing photons in the IGM, a free streaming length. This parameter gives approximate size of H_{II} region at an instant before they overlap.

Below given table summarizes values to perform Fisher matrix analysis.

Ionizing Efficiency $\mathbf{N_{ion}}$ -	**Minimum Halo Mass** $\mathbf{M_{h,\,min}}$ $[\log_{10}(M_\odot)]$	**Mean Free Path** $\mathbf{R_{mfp}}$ [Mpc]
23.21	10	8.28
23.71	11	9.96
24.21	12	11.64
24.71	13	13.32
25.21	**14**	**15.00**
25.71	15	16.68
26.21	16	18.36
26.71	17	20.04
27.21	18	21.72

Table 6.2: Parameter space for Fisher matrix analysis. Fiducial values are shown in bold.

In order to calculate the Fisher matrix $\mathcal{F}_{\alpha\beta}$, we need the 21-cm PS error covariance $\mathbf{C}_{ij}$ and $\frac{\partial P(k)}{\partial q_\alpha}$ which is the partial derivatives of the bin-averaged 21-cm Power spectrum and Marked Power spectrum. We use increment in the astrophysical parameters as the fractional errors in these parameters. To compute the derivative, we use the six point numerical derivative.

Moreover, for our analysis, we sum over only selective parts of k modes. Below are the triangle plots for Fisher information analysis for all functional forms of the Mark and Marked field for some set of free parameters. Since smoothing at larger length scale washes fluctuations below certain length scale, we expect information loss. Hence, we keep the smoothing scale fixed for all functional forms even if other free parameters are varied.

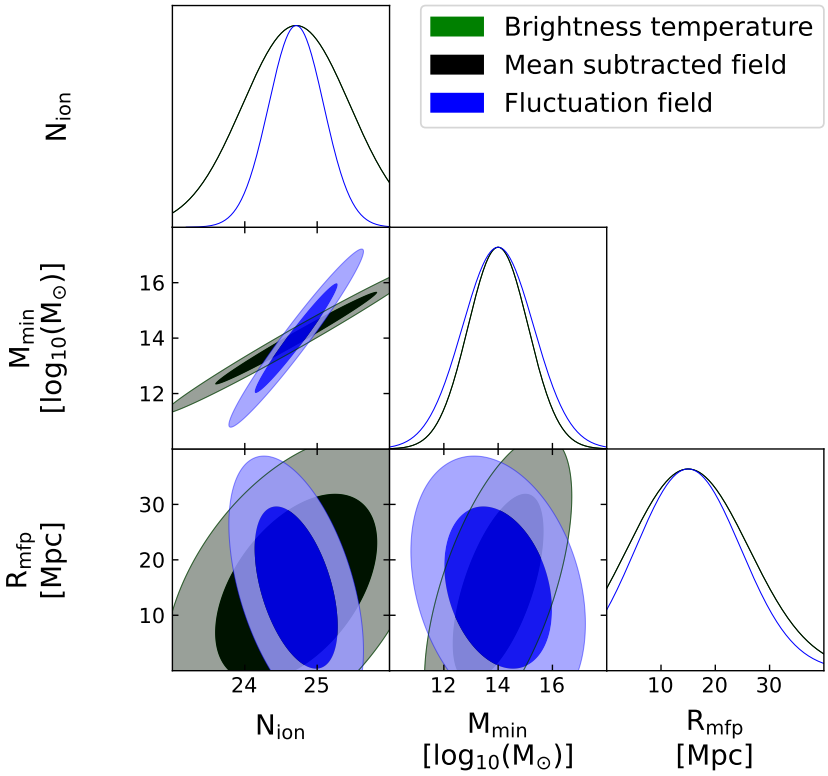


Figure 6.10: Fisher information analysis for the Power spectrum at $\bar{x}_{HI} = 0.5$. As expected, there is complete overlap of brightness temperature field and corresponding mean subtracted field as shifting the mean does not modify the distribution of fluctuations.

We compute the derivative of the Power spectra for different functional forms of mark and marked field for various free parameters (the mark parameters).

We perform this analysis for all functional forms of marks and below are the results for the same.

$$m_0(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1+f}{2+f+\delta_{\mathrm{T_b}}(\mathrm{R}, \mathbf{x}, z)} \right]^p$$

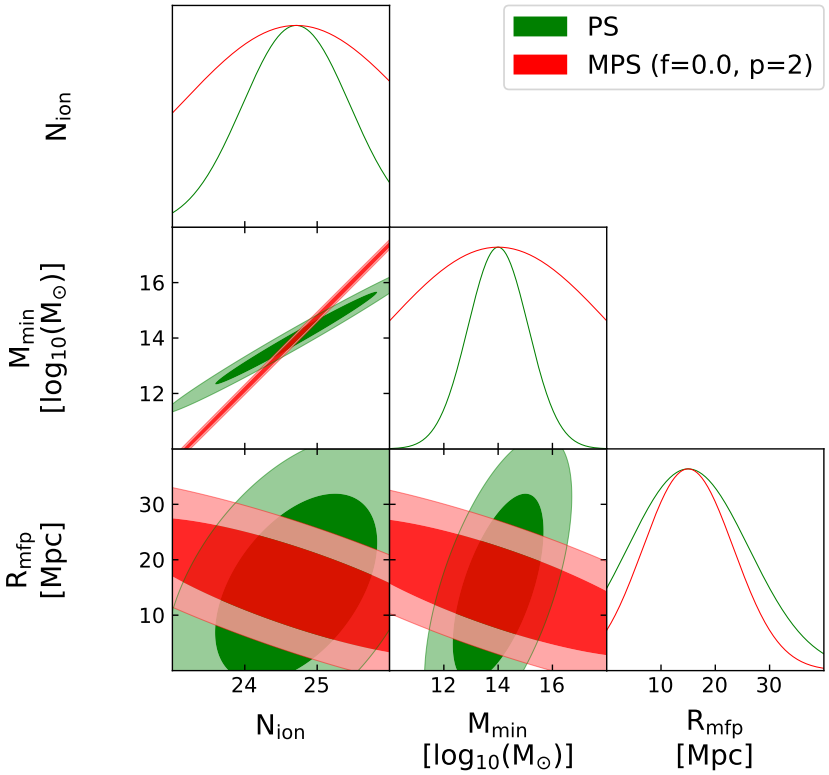


Figure 6.11: Fisher information analysis for the Marked Power spectrum at $\bar{\mathrm{x}}_{\mathrm{HI}} = 0.5$ (z=7.520) for given functional form).

As discussed earlier, this functional form alone can not be used interdependent to make better estimator.

$$m_0(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1+f}{2+f+\delta_{\mathrm{T_b}}(\mathrm{R}, \mathbf{x}, z)}\right]^p$$

and

$$\mathbf{M}_0(\mathbf{x}, z; \mathrm{R}, f, p) = m_0(\mathbf{x}, z; \mathrm{R}, f, p) \times \mathrm{dT}(\mathbf{x}, z)$$

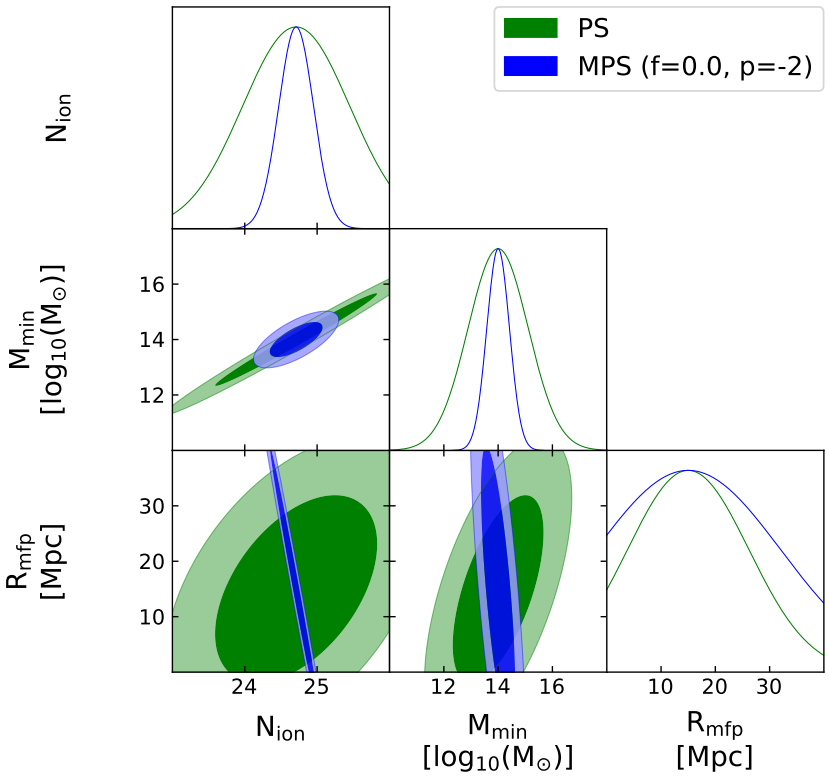


Figure 6.12: Fisher information analysis for the Marked Power spectrum at $\bar{x}_{\mathrm{HI}} = 0.5$ (z=7.520) for given functional form).

There is significantly better performance of this functional form when reassigned as weights. But we can do better as shown in next plots.

$$m_1(\mathbf{x}, z : \mathrm{R}, f, p) = \left[\frac{1+f}{1+f+\left[\mathrm{dT}(\mathbf{x}, z)\right]_{\mathrm{R}}} \right]^p$$

where $\mathrm{dT}(\mathbf{x}, z) = \delta \mathrm{T_b}(\mathbf{x}, z) - \langle \delta \mathrm{T_b}(z) \rangle$ is the mean subtracted brightness temperature field.

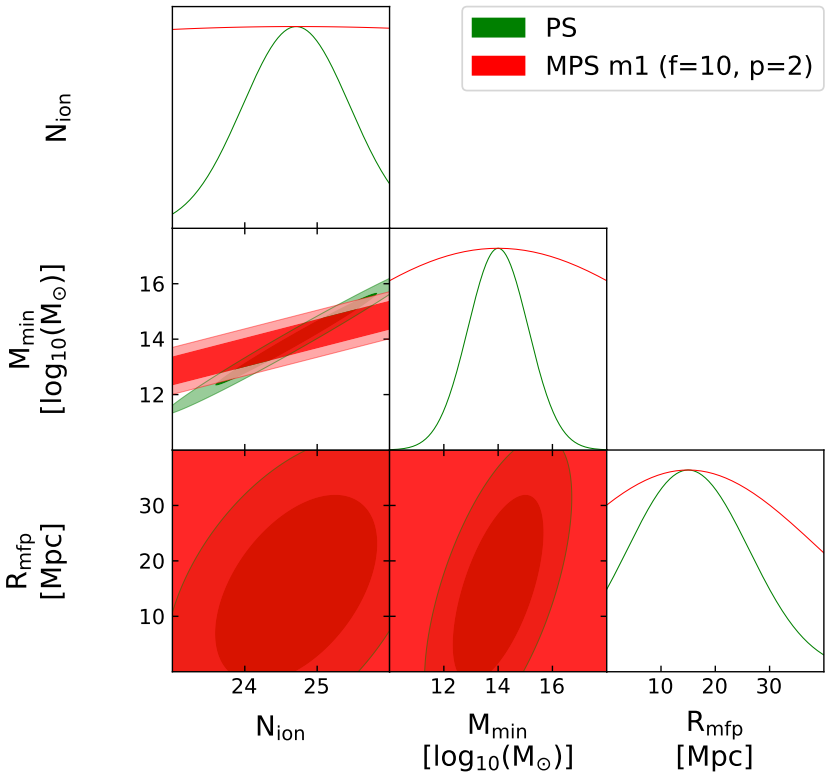


Figure 6.13: Fisher information analysis for the Marked Power spectrum at $\bar{\mathrm{x}}_{\mathrm{HI}} = 0.5$ (z=7.520) for given functional form).

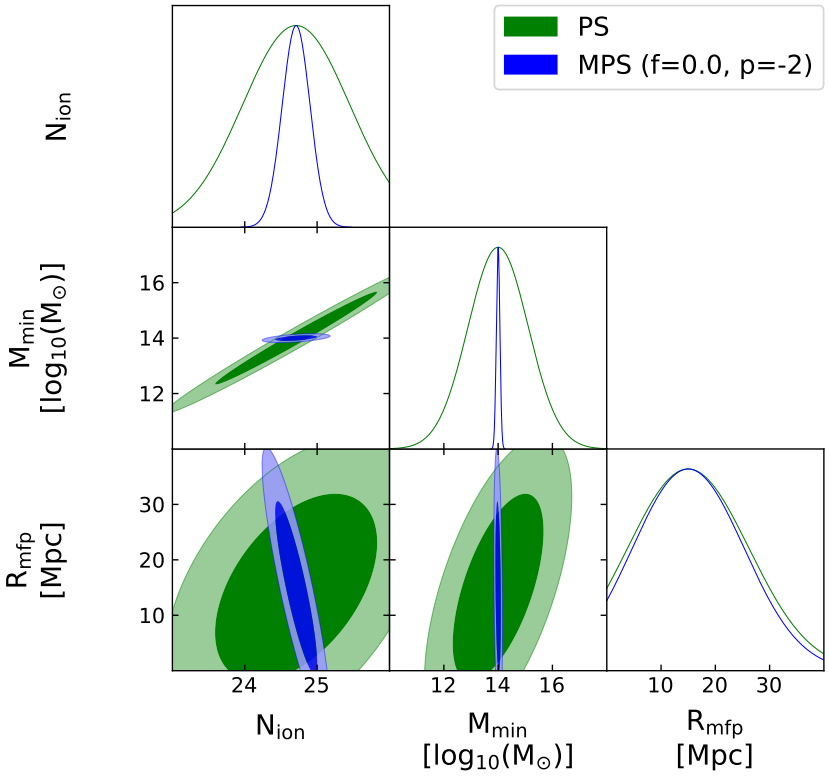


Figure 6.14: Fisher information analysis for the Marked Power spectrum at $\bar{x}_{HI} = 0.5$ (z=7.520) for given functional form).

There clearly is better constraints as show for a set of free parameter value. The "funny" units again impose challenge to interpret results as expected. These plots also show how various free parameters for the same functional form of the mark can give very different sensitivity of the Marked Power Spectrum with the astrophysical parameters.

$$m_2(\mathbf{x}, z : \mathrm{R}, f, p) = \left[\frac{1+f}{1+f+\left[\delta \mathrm{T_b}(\mathbf{x}, z)\right]_{\mathrm{R}}} \right]^p$$

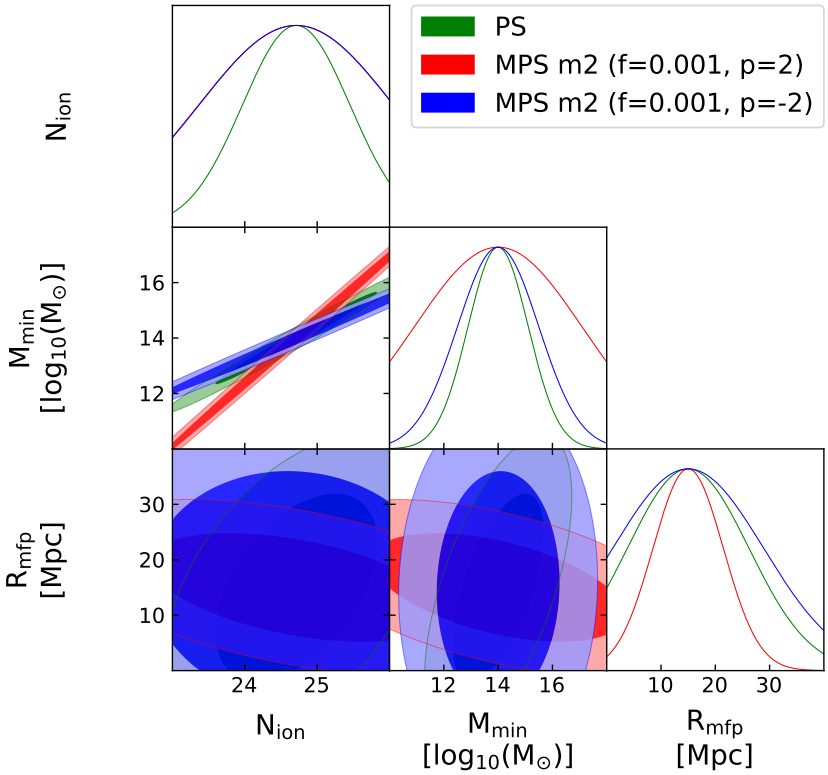


Figure 6.15: Fisher information analysis for the Marked Power spectrum at $\bar{\mathrm{x}}_{\mathrm{HI}} = 0.5$

Very similar conclusion can be made for this functional form with that for 6.3.

$$m_3(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1 + f \cdot \langle \delta \mathrm{T_b}(z) \rangle}{f \cdot \langle \delta \mathrm{T_b}(z) \rangle + \left[\mathrm{dT}(\mathbf{x}, z) \right]_{\mathrm{R}}} \right]^p$$

where $\mathrm{dT}(\mathbf{x}, z) = \delta \mathrm{T_b}(\mathbf{x}, z) - \langle \delta \mathrm{T_b}(z) \rangle$ is the mean subtracted brightness temperature field.

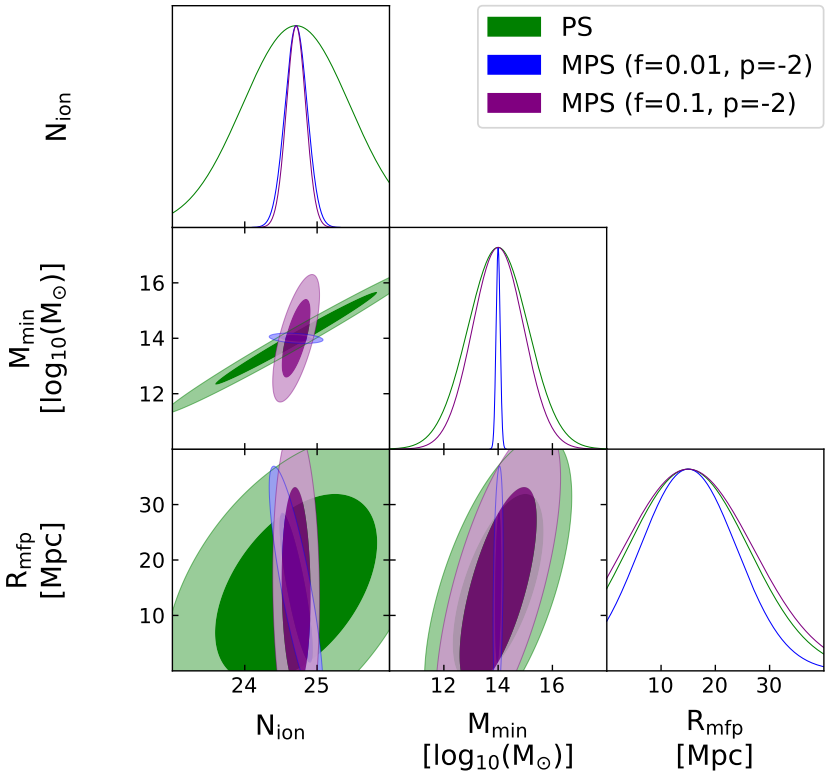


Figure 6.16: Fisher information analysis for the Marked Power spectrum at $\bar{\mathrm{x}}_{\mathrm{HI}} = 0.5$ (z=7.520) for given functional form).

$$m_3(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1 + f \cdot \langle \delta \mathrm{T_b}(z) \rangle}{f \cdot \langle \delta \mathrm{T_b}(z) \rangle + \left[\mathrm{dT}(\mathbf{x}, z) \right]_{\mathrm{R}}} \right]^p$$

and

$$\mathbf{M}_3(\mathbf{x}, z; \mathrm{R}, f, p) = m_3(\mathbf{x}, z; \mathrm{R}, f, p) \times \mathrm{dT}(\mathbf{x}, z)$$

where $\mathrm{dT}(\mathbf{x}, z) = \delta \mathrm{T_b}(\mathbf{x}, z) - \langle \delta \mathrm{T_b}(z) \rangle$ is the mean subtracted brightness temperature field.

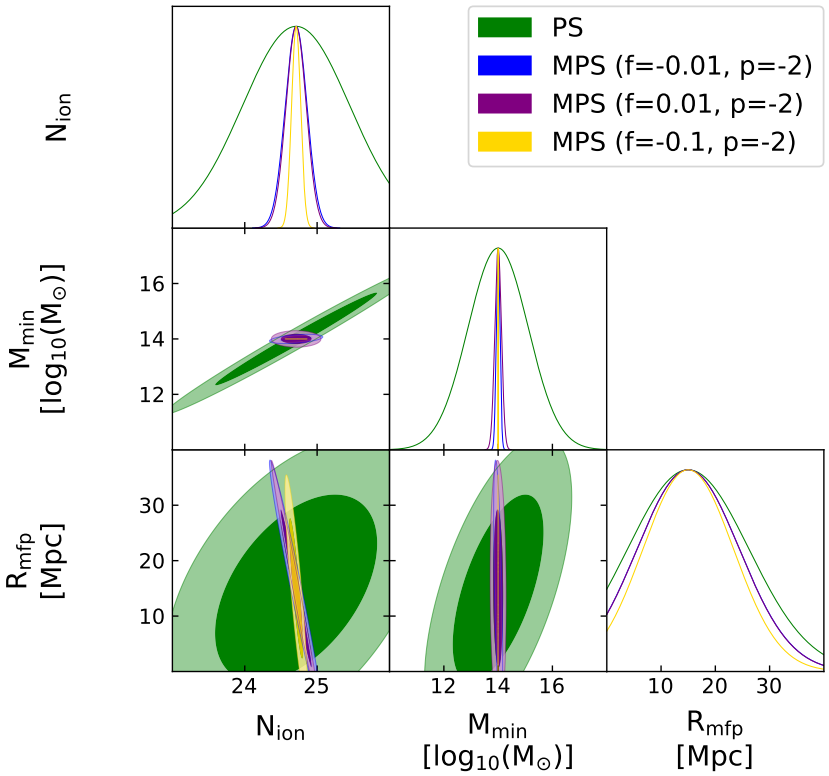


Figure 6.17: Fisher information analysis for the Marked Power spectrum at $\bar{\mathrm{x}}_{\mathrm{HI}} = 0.5$ (z=7.520) for given functional form).

$$m_3(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1 + f \cdot \langle \delta \mathrm{T_b}(z) \rangle}{f \cdot \langle \delta \mathrm{T_b}(z) \rangle + \left[\mathrm{dT}(\mathbf{x}, z) \right]_{\mathrm{R}}} \right]^p$$

and

$$\mathbf{M}_3(\mathbf{x}, z; \mathrm{R}, f, p)_{\mathrm{R}} = m_3(\mathbf{x}, z; \mathrm{R}, f, p) \times \left[\mathrm{dT}(\mathbf{x}, z) \right]_{\mathrm{R}}$$

where $\mathrm{dT}(\mathbf{x}, z) = \delta \mathrm{T_b}(\mathbf{x}, z) - \langle \delta \mathrm{T_b}(z) \rangle$ is the mean subtracted brightness temperature field.

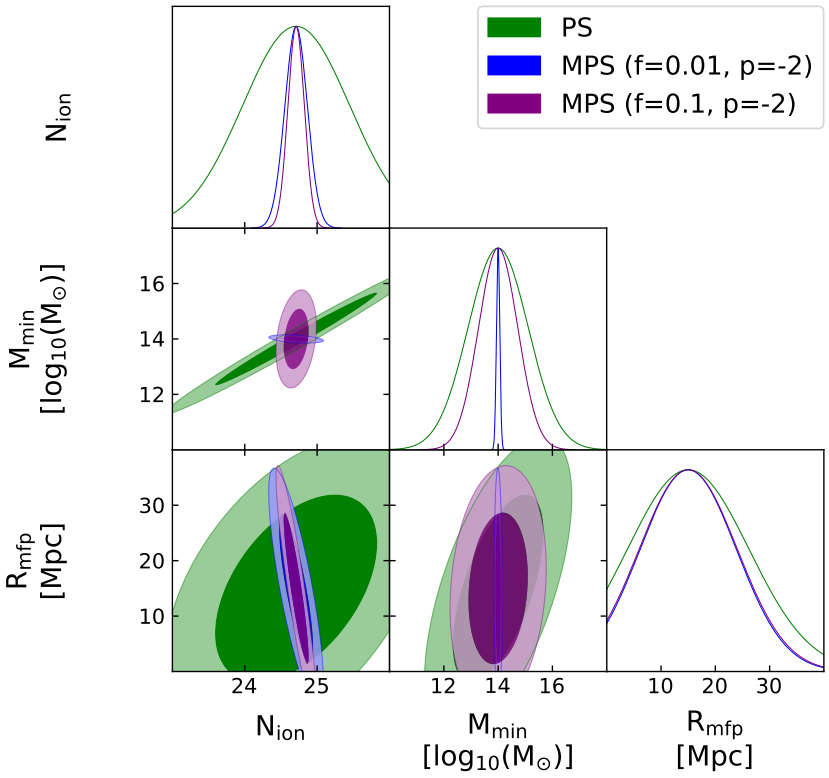


Figure 6.18: Fisher information analysis for the Marked Power spectrum at $\bar{\mathrm{x}}_{\mathrm{HI}} = 0.5$ (z=7.520) for given functional form).

As seen clearly for functional form 6.5 for both Mark and marked field, there is significantly less error in Marked Power Spectrum (MPS) in estimating astrophysical parameters. The degeneracy in these free parameters is evident too as described in previous sections. This shows versatility of Mark and its functional form and how it can be implemented to extract maximum information from the field. Moreover, certain Mark can be used to remove covariance in the astrophysical parameters. Also, their combination can be used to put tighter constraints.

$$m_4(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1 + f \cdot \langle \mathrm{dT}(z) \rangle}{f \cdot \langle \mathrm{dT}(z) \rangle + \left[\mathrm{dT}(\mathbf{x}, z) \right]_{\mathrm{R}}} \right]^p$$

where $\mathrm{dT}(\mathbf{x}, z) = \delta \mathrm{T_b}(\mathbf{x}, z) - \langle \delta \mathrm{T_b}(z) \rangle$ is the mean subtracted brightness temperature field.

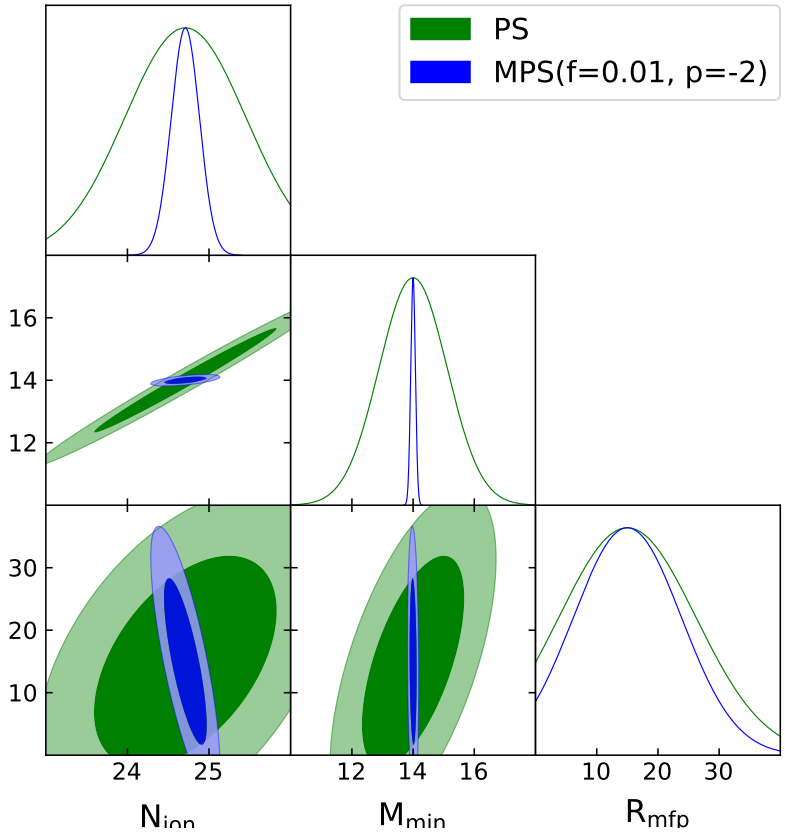


Figure 6.19: Fisher information analysis for the Marked Power spectrum at $\bar{\mathrm{x}}_{\mathrm{HI}} = 0.5$ (z=7.520) for given functional form).

$$m_4(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1 + f \cdot \langle \mathrm{dT}(z) \rangle}{f \cdot \langle \mathrm{dT}(z) \rangle + \left[\mathrm{dT}(\mathbf{x}, z) \right]_{\mathrm{R}}} \right]^p$$

and

$$\mathbf{M}_4(\mathbf{x}, z; \mathrm{R}, f, p) = m_4(\mathbf{x}, z; \mathrm{R}, f, p) \times \mathrm{dT}(\mathbf{x}, z)$$

where $\mathrm{dT}(\mathbf{x}, z) = \delta \mathrm{T_b}(\mathbf{x}, z) - \langle \delta \mathrm{T_b}(z) \rangle$ is the mean subtracted brightness temperature field.

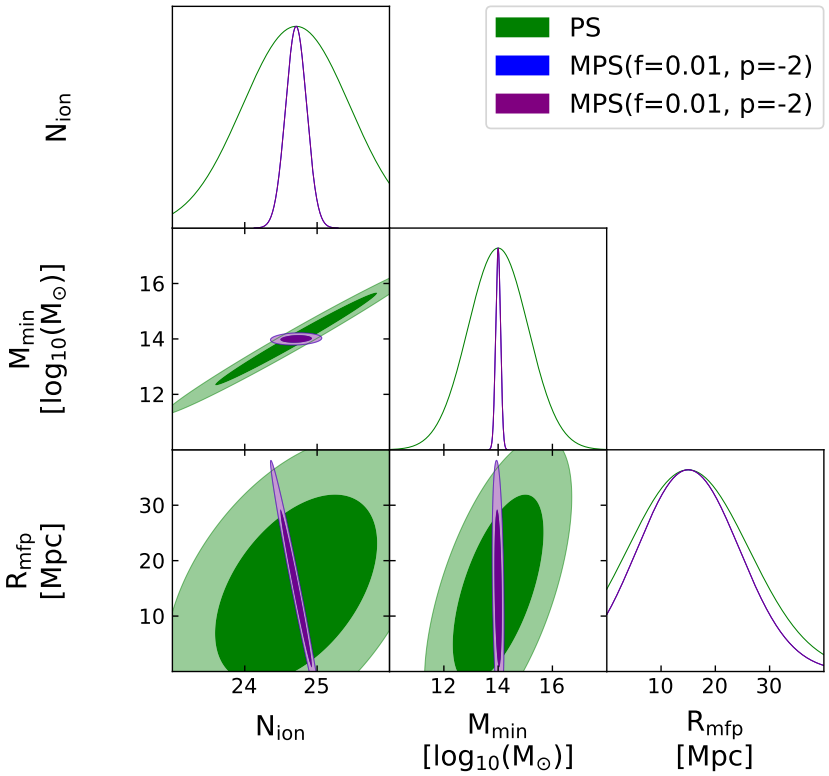


Figure 6.20: Fisher information analysis for the Marked Power spectrum at $\bar{\mathrm{x}}_{\mathrm{HI}} = 0.5$ (z=7.520) for given functional form).

$$m_4(\mathbf{x}, z; \mathrm{R}, f, p) = \left[\frac{1 + f \cdot \langle \mathrm{dT}(z) \rangle}{f \cdot \langle \mathrm{dT}(z) \rangle + \left[\mathrm{dT}(\mathbf{x}, z) \right]_{\mathrm{R}}} \right]^p$$

and

$$\mathbf{M}_4(\mathbf{x}, z; \mathrm{R}, f, p)_{\mathrm{R}} = m_4(\mathbf{x}, z; \mathrm{R}, f, p) \times \left[\mathrm{dT}(\mathbf{x}, z) \right]_{\mathrm{R}}$$

where $\mathrm{dT}(\mathbf{x}, z) = \delta \mathrm{T_b}(\mathbf{x}, z) - \langle \delta \mathrm{T_b}(z) \rangle$ is the mean subtracted brightness temperature field

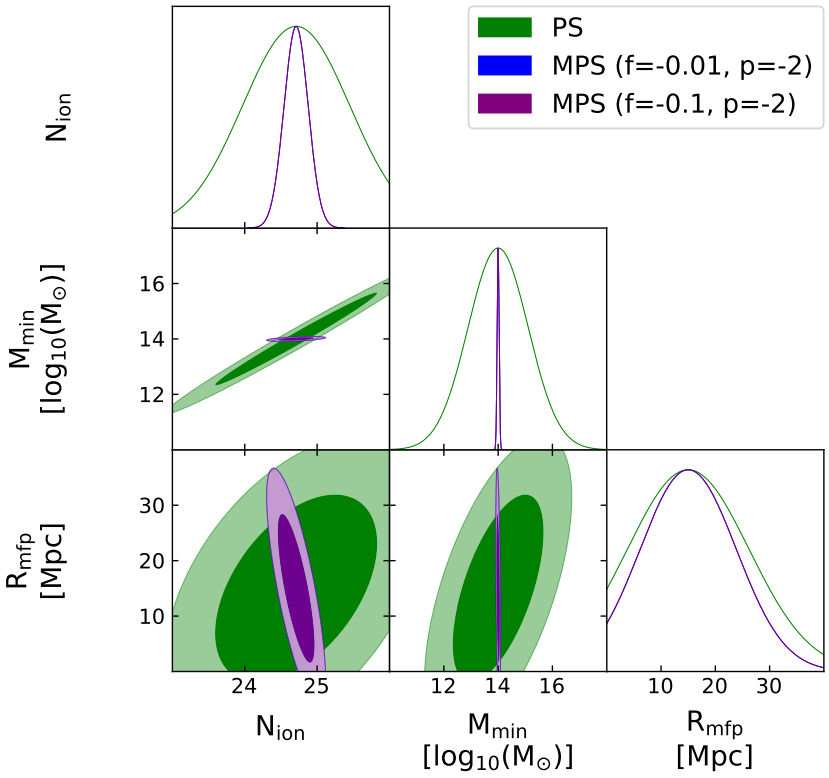


Figure 6.21: Fisher information analysis for the Marked Power spectrum at $\bar{\mathrm{x}}_{\mathrm{HI}} = 0.5$ (z=7.520) for given functional form).

Since the functional form of this is very similar to that of 6.5, the results and conclusions are too similar as expected. This particular functional for is worth trying since the interferometer measures mean subtracted field. So this mark, motivated from previous marks, can be used to understand about the signal's statistical properties. Another interesting observation is as follows: Some functional forms of marks (particularly the corresponding Marked power spectrum) for a a set of free parameters are able to provide less error in astrophysical parameters with change in the degeneracy. Hence, combination of Power spectrum and Marked Power spectrum can be considered as an interesting estimator of the astrophysical parameters giving lesser error. We leave this exploration to future study.

The marginalized errors as the function of Largest kmode for different functional form of mark for a set of free parameters is given as follows:

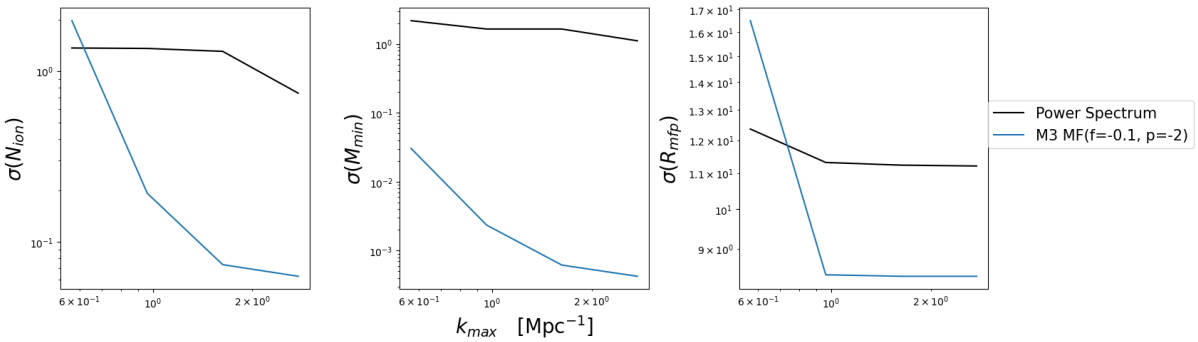


Figure 6.22: Marginalized error for a specific functional form of mark.

As seen clearly this functional form has significantly less error. Here, m3 means 6.5 and MF stands for Marked field. Hence, this functional form corresponds to usage of 6.5 in context of marked field with weights to mean subtracted field.

We extend this analysis to different types of functional form and to compare performance of different functional form. The plot is as shown:

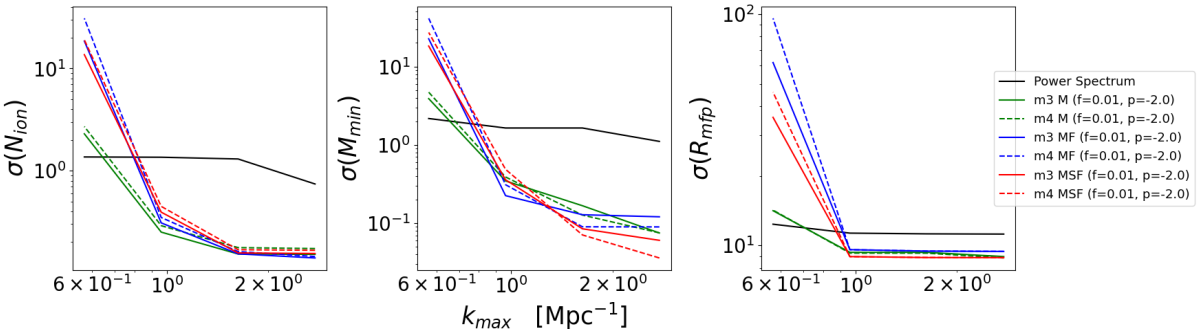


Figure 6.23: Marginalized error in astrophysical parameters as function of largest k-mode for various functional form of marks.

There is clear shoot up in error if we take less number of k modes into consideration but if we increase number of k modes, the error in estimation of astrophysical parameters can be reduced significantly. Here, M is Mark, MF is Marked field and MSF is Marked filed where weight is given to smoothed fluctuation field (Marked Smoothed Field).

6.6 Largest Cluster Statistic

One important statistic applied to the marked field is the Largest Cluster Statistic (LCS). The expression for LCS, specifically when targeting under-dense regions (i.e., low values of fluctuation in the HI 21-cm signal) and therefore ionized regions, is as follows:

$$\text{LCS} = \frac{\text{Volume of Largest Ionized Region}}{\text{Volume of Total Ionized Region}} \tag{6.11}$$

LCS represents the fraction of the ionized volume that resides in the largest ionized region within the intergalactic medium (IGM). Using this statistic, the IGM can be analyzed during the epoch of reionization (EoR). During the initial stages of reionization, when luminous sources are largely absent and the IGM consists predominantly of neutral hydrogen atoms, the value of LCS is very small. As ionization advances, more luminous sources form and contribute to ionizing increasingly larger portions of the IGM, thereby increasing the LCS value. Over time, as these ionized regions grow and begin to overlap, there is a rapid increase in the value of LCS, a phenomenon known as percolation. This process of percolation, wherein smaller ionized regions merge into larger ones, marks a phase transition in the IGM from a neutral to an ionized state on very large scales. Therefore, LCS is an effective tool for studying this phase change in the IGM.

Since the marked statistic, by design, distinguishes between neutral and ionized regions with greater clarity, it should, with the appropriate parameters, be capable of accurately tracing the reionization history. Examining the ionization history of the IGM, we observe that due to percolation (the merging of smaller ionized regions into a single large ionized region), the LCS value spikes at a specific time (or redshift). Comparing LCS from the marked field with that of the differential brightness temperature provides insights into the percolation threshold and the morphological evolution of the IGM as traced by the 21-cm signal.

6.6.1 Impact of smoothing radius R

Since 6.1 clearly makes the distinction between ionized and neutral regions more and more clear, we use this functional form for the analysis. It is evident in 6.2a. When smoothing the field with a spherical filter, the number of partially ionized regions increases, introducing a bias in the ionization history. This bias can be observed by plotting the LCS for the smoothed field, as discussed in [46]. Comparing this plot to the LCS of the marked field reveals additional information about the structure of the field. The resulting plot is as follows:

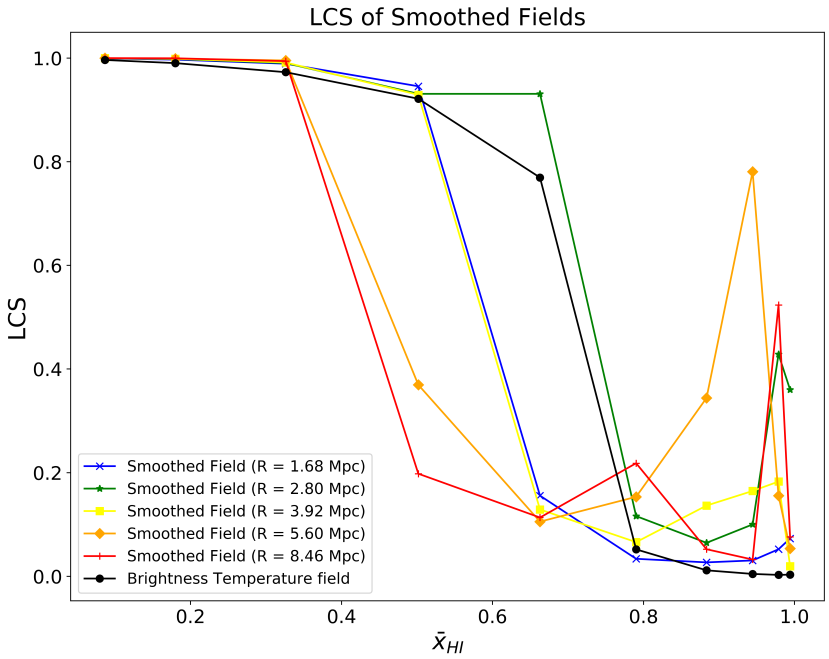


Figure 6.24: LCS smoothed field.

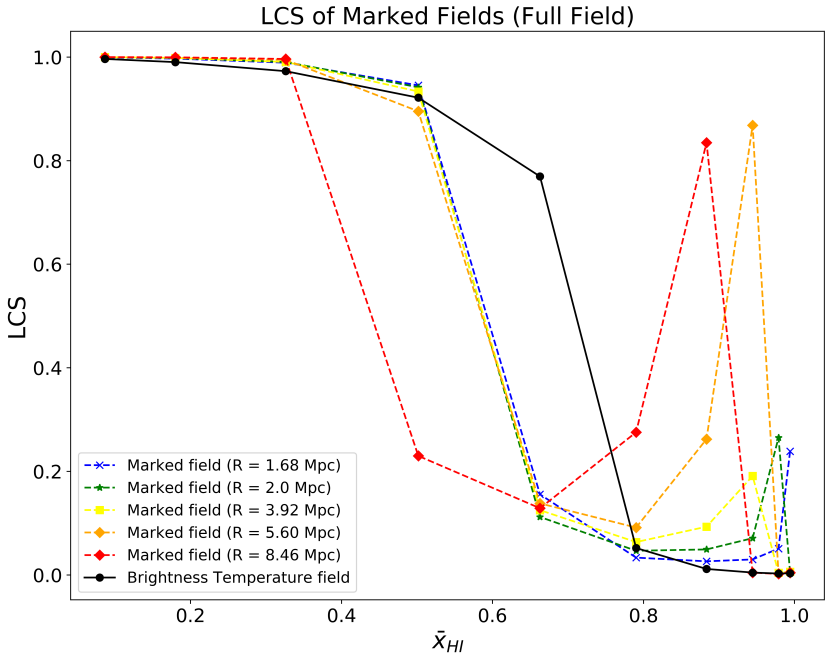


Figure 6.25: LCS for Mark.

Clearly, for both the cases for mark and marked field, there is a a bias in the ionization history which corresponds to smoothing operation which is as expected. By smoothing the fields, averaging over values over space is done which results into more and more number of partially ionized region. These regions are considered neutral or ionized based on amount of averaging is done, in other words, length scale of smoothing.

6.6.2 Impact of exponent p

Positive values of p tends to make the unmarked field more and more bimodal. Moreover, due to filamentary, it becomes difficult to apply study of LCS to neutral fraction field. Hence, if the field is made more and more bimodal, distinction between ionized and neutral regions becomes more and more clear which comes at a cost of making the field more and more biased towards ionized region. Hence, for very large values of p, LCS can be applied and. The following plot shows variation of p with LCS for a fixed smoothing radius R. The plot is as shown:

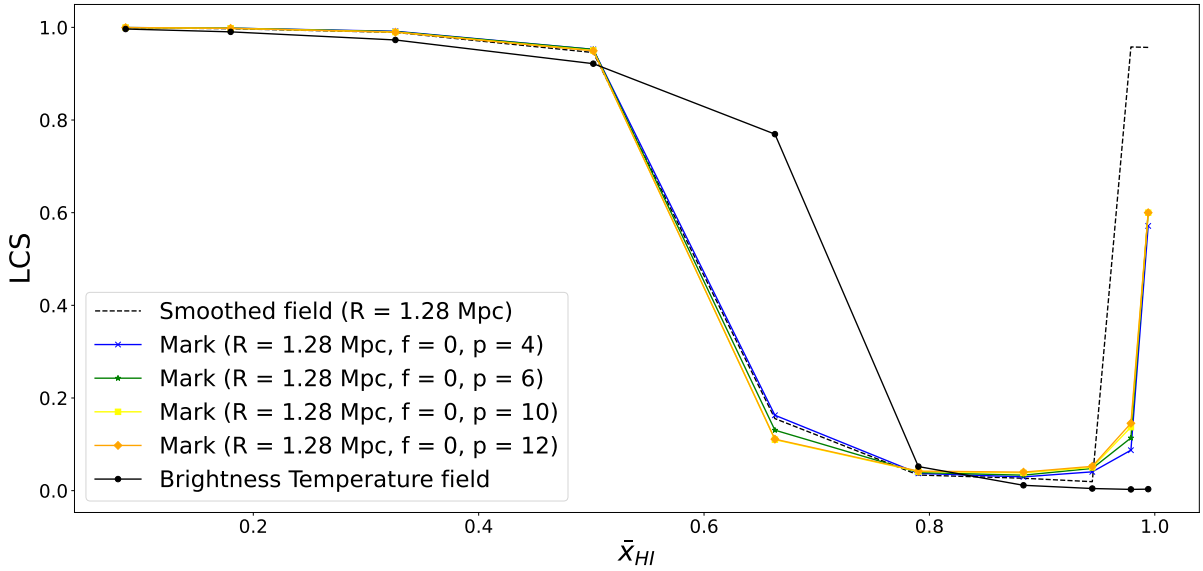


Figure 6.26: Variation of LCS for brightness temperature field with exponent p for 6.1.

Since p makes the field more and more bimodal in nature, the distinction from ionized to neutral regions becomes clear. However, this distinction after certain point becomes saturated and it becomes irrelevant to apply higher values of p as they lead to very similar response. Hence, nothing new can be inferred from the statistic applied to corresponding marks which is evident in the above plot. Hence, this once again serves as a justification about how looking for right parameters to mark the field lies at the heart of the study.

CHAPTER 7

Summary and Future Plan

7.1 Summary

The process of marking the field aims to effectively visualize key differences in the field, but multiplying the mark with the original field (Marked field) can reduce variations for this type of setup. Finding the right parameters for the mark to study IGM in EoR is the central objective of the study since the information content of a mark isn't unique. To extract maximum insight, different statistics and the information content of each mark needs thorough study, especially at given redshifts to identify fluctuations separately at specific scales that drive 21-cm field variations. Hence, quantifying unique and relevant information is what is required in this study. Moreover, the response of the mark to each field is different. The reason may be the dynamic range but more importantly, difference in features existing in the fields.

However, challenges remain in the selection of parameters for the design of the mark and its interpretation. Different sets of free parameters can map to similar information, creating a degeneracy that makes the process of drawing unique inferences difficult. Addressing this issue is critical for understanding the nature of the field and interpreting results more effectively. Hence, a thorough study is required, which can lead to better insights about the field in question.

The fact that different functional forms can be used to approach different problem statements makes the study of Marked Power Spectrum in the context of intensity mapping a very rich and interesting field of study.

7.2 Future scope

The versatility of Marked Statistics opens several avenues for advancing our understanding of the Epoch of Reionization (EoR) and cosmic evolution. Building on the current research, the following directions are proposed to further explore and refine this framework:

1. Theoretical Modeling and Simulation Studies

- **Marked Power Spectrum on detailed maps:** We plan to extend the application of MPS to larger and more detailed simulations by including additional astrophysical parameters like gas content and radiative transfer. This would help in capturing the environmental dependence of ionizing sources and further validate the mark design on both large- and small-scale structures.
- **Information comparison with higher-order statistics:** A systematic comparison between the MPS and the bispectrum will help quantify the non-Gaussian information captured by the Mark. This study will identify when the MPS suffices and when more computationally intensive statistics are necessary.
- **Marked Cross Power Spectrum:** Moreover, development of the concept of cross-marking by applying marks from one tracer field (e.g., CO or CII lines) to another (e.g., HI 21-cm) can also be done. This can amplify complementary signals and tighten astrophysical constraints.

2. Bayesian Inference and Statistical Analysis

- **Bayesian pipeline development:** We plan to construct a Bayesian inference framework using Marked fields. By performing parameter estimation directly on the Marked statistics, we can potentially obtain tighter constraints on astrophysical and cosmological parameters compared to traditional methods.

3. Observational Applications and Systematics

- **Noise and foreground suppression:** We aim to design marks tailored to reduce specific types of noise or foreground contamination in observational data. By down-weighting known systematic contributions, this method could improve signal extraction and enhance the reliability of observational studies.

4. Machine Learning Integration

- **ML-based mark optimization:** We also plan to employ machine learning algorithms to learn optimal functional forms for marks. This bypasses the need for trial-and-error approaches and accelerates the identification of marks that maximize information extraction.
- **Classification using Marked fields:** Using the Marked fields as enhanced inputs for ML-based classification tasks can also be implemented. For example, distinguishing

ionized and neutral regions in the IGM becomes more efficient when marks enhance feature contrast.

References

[1] Jonathan R. Pritchard and Abraham Loeb. "21 cm Cosmology in the 21st Century". In: *arXiv preprint arXiv:1109.6012* (2012). URL: https://arxiv.org/abs/1109.6012.

[2] Suman Majumdar. "Quantifying the Non-Gaussianity in the EoR 21-cm Signal through Bispectrum". In: *Journal of Cosmology and Astroparticle Physics* (2018). URL: https://arxiv.org/abs/1804.00004.

[3] Steven R. Furlanetto and S. Peng Oh. "21-cm Cosmology". In: *Monthly Notices of the Royal Astronomical Society* 457 (2016), p. 1813. DOI: 10.1093/mnras/stw059. URL: https://academic.oup.com/mnras/article/457/2/1813/2584299.

[4] Somnath Bharadwaj, Arindam Mazumdar, and Debanjan Sarkar. "Quantifying the redshift space distortion of the bispectrum I: primordial non-Gaussianity". In: *Monthly Notices of the Royal Astronomical Society* 493.1 (Feb. 2020), pp. 594–602. ISSN: 1365-2966. DOI: 10.1093/mnras/staa279. URL: http://dx.doi.org/10.1093/mnras/staa279.

[5] Oliver H E Philcox et al. "encore: an O (Ng2) estimator for galaxy N-point correlation functions". In: *Monthly Notices of the Royal Astronomical Society* 509.2 (Oct. 2021), pp. 2457–2481. ISSN: 0035-8711. DOI: 10.1093/mnras/stab3025. eprint: https://academic.oup.com/mnras/article-pdf/509/2/2457/41227438/stab3025.pdf. URL: https://doi.org/10.1093/mnras/stab3025.

[6] Ravi K. Sheth and Giuseppe Tormen. "On the environmental dependence of halo formation". In: *Monthly Notices of the Royal Astronomical Society* 350.4 (June 2004), pp. 1385–1390. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2004.07733.x. eprint: https://academic.oup.com/mnras/article-pdf/350/4/1385/3089294/350-4-1385.pdf. URL: https://doi.org/10.1111/j.1365-2966.2004.07733.x.

[7] Elena Massara, Simone Ferraro, and Uroš Seljak. "Constraining Neutrino Mass with Voids in Galaxy Surveys Using Marked Correlation Functions". In: *Journal of Cosmology and Astroparticle Physics* (2020). URL: https://arxiv.org/abs/2011.00005.

[8] Elena Massara and Tobias Baldauf. "Fisher Forecasts for Marked Power Spectra: Improving Constraints with Environmental Information". In: *Monthly Notices of the Royal Astronomical Society* 512.1 (2022), pp. 1008–1019. URL: https://arxiv.org/abs/2201.00006.

[9] Martin White and Lam Hui. "Marked Correlation Functions for Large-Scale Structure". In: *Physical Review D* 94.8 (2016), p. 083528. URL: https://arxiv.org/abs/1608.06118.

[10] Alejandro Aviles and Marco Baldi. "Marked Statistics for Modified Gravity: An Efficient Tool to Probe Dark Energy and Gravity Theories". In: *Journal of Cosmology and Astroparticle Physics* (2021). URL: https://arxiv.org/abs/2102.04545.

[11] Mohd Kamran et al. *The re-markable 21-cm power spectrum I: Probing the HI distribution in the post-reionization era using marked statistics*. 2024. arXiv: 2409.05187 [astro-ph.CO]. URL: https://arxiv.org/abs/2409.05187.

[12] Jessica Cowell, Tiziana Di Matteo, and Rupert Croft. "Information Content of Marked Power Spectra: A Comparison of Different Marks". In: *Journal of Cosmology and Astroparticle Physics* (2024). URL: https://arxiv.org/abs/2401.00001.

[13] Dietrich Stoyan. "On Correlations of Marked Point Processes". In: *Mathematische Nachrichten* 116.1 (1984), pp. 197–207. DOI: https://doi.org/10.1002/mana.19841160115. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mana.19841160115. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/mana.19841160115.

[14] Ravi K. Sheth, Andrew J. Connolly, and Ramin Skibba. "Marked correlations in galaxy formation models". In: (2005). arXiv: astro-ph/0511773 [astro-ph]. URL: https://arxiv.org/abs/astro-ph/0511773.

[15] Claus Beisbart and Martin Kerscher. "Luminosity- and Morphology-dependent Clustering of Galaxies". In: *The Astrophysical Journal* 545.1 (Dec. 2000), p. 6. DOI: 10.1086/317788. URL: https://dx.doi.org/10.1086/317788.

[16] Claus Beisbart, Martin Kerscher, and Klaus Mecke. "Mark correlations: relating physical properties to spatial distributions". In: (2002). arXiv: physics/0201069 [physics.data-an]. URL: https://arxiv.org/abs/physics/0201069.

[17] Ramin Skibba et al. "The luminosity-weighted or 'marked' correlation function". In: *Monthly Notices of the Royal Astronomical Society* 369.1 (May 2006), pp. 68–76. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2006.10196.x. eprint: https://academic.oup.com/mnras/article-pdf/369/1/68/18664730/mnras0369-0068.pdf. URL: https://doi.org/10.1111/j.1365-2966.2006.10196.x.

[18] Ramin A. Skibba et al. "Galaxy Zoo: the environmental dependence of bars and bulges in disc galaxies: Environmental dependence of bars and bulges". In: *Monthly Notices of the Royal Astronomical Society* 423.2 (Apr. 2012), pp. 1485–1502. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2012.20972.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2012.20972.x.

[19] S. Gottlöber et al. "Spatial distribution of galactic halos and their merger histories". In: *Astronomy amp; Astrophysics* 387.3 (May 2002), pp. 778–787. ISSN: 1432-0746. DOI: 10.1051/0004-6361:20020339. URL: http://dx.doi.org/10.1051/0004-6361:20020339.

[20] R. K. Sheth. "The halo-model description of marked statistics". In: *Monthly Notices of the Royal Astronomical Society* 364.3 (Dec. 2005), pp. 796–806. ISSN: 1365-2966. DOI: 10.1111/j.1365-2966.2005.09609.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2005.09609.x.

[21] Martin White and Nikhil Padmanabhan. "Breaking halo occupation degeneracies with marked statistics". In: *Monthly Notices of the Royal Astronomical Society* 395.4 (June 2009), pp. 2381–2384. ISSN: 1365-2966. DOI: 10.1111/j.1365-2966.2009.14732.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2009.14732.x.

[22] Gong-Bo Zhao, Baojiu Li, and Kazuya Koyama. "Testing Gravity Using the Environmental Dependence of Dark Matter Halos". In: *Phys. Rev. Lett.* 107 (7 Aug. 2011), p. 071303. DOI: 10.1103/PhysRevLett.107.071303. URL: https://link.aps.org/doi/10.1103/PhysRevLett.107.071303.

[23] Hans A. Winther, David F. Mota, and Baojiu Li. "ENVIRONMENT DEPENDENCE OF DARK MATTER HALOS IN SYMMETRON MODIFIED GRAVITY". In: *The Astrophysical Journal* 756.2 (Aug. 2012), p. 166. DOI: 10.1088/0004-637X/756/2/166. URL: https://dx.doi.org/10.1088/0004-637X/756/2/166.

[24] Lucas Lombriser, Fergus Simpson, and Alexander Mead. "Unscreening Modified Gravity in the Matter Power Spectrum". In: *Phys. Rev. Lett.* 114 (25 June 2015), p. 251101. DOI: 10.1103/PhysRevLett.114.251101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.114.251101.

[25] Difu Shi, Baojiu Li, and Jiaxin Han. "Environmental screening of dark matter haloes in f(R) gravity". In: *Monthly Notices of the Royal Astronomical Society* 469.1 (Apr. 2017), pp. 705–715. ISSN: 0035-8711. DOI: 10.1093/mnras/stx865. eprint: https://academic.oup.com/mnras/article-pdf/469/1/705/49203330/mnras_469_1_705.pdf. URL: https://doi.org/10.1093/mnras/stx865.

[26] Joaquín Armijo et al. "Testing modified gravity using a marked correlation function". In: *Monthly Notices of the Royal Astronomical Society* 478.3 (May 2018), pp. 3627–3632. ISSN: 0035-8711. DOI: 10.1093/mnras/sty1335. eprint: https://academic.oup.com/mnras/article-pdf/478/3/3627/25072375/sty1335.pdf. URL: https://doi.org/10.1093/mnras/sty1335.

[27] Siddharth Satpathy et al. "Measurement of marked correlation functions in SDSS-III Baryon Oscillation Spectroscopic Survey using LOWZ galaxies in Data Release 12". In: *Monthly Notices of the Royal Astronomical Society* 484.2 (Jan. 2019), pp. 2148–2165. ISSN: 0035-8711. DOI: 10.1093/mnras/stz009. eprint: https://academic.oup.com/mnras/article-pdf/484/2/2148/27620577/stz009.pdf. URL: https://doi.org/10.1093/mnras/stz009.

[28] César Hernández-Aguayo et al. "Large-scale redshift space distortions in modified gravity theories". In: *Monthly Notices of the Royal Astronomical Society* 485.2 (Feb. 2019), pp. 2194–2213. ISSN: 1365-2966. DOI: 10.1093/mnras/stz516. URL: http://dx.doi.org/10.1093/mnras/stz516.

[29] Joaquin Armijo et al. "A new test of gravity – I. Introduction to the method". In: *Monthly Notices of the Royal Astronomical Society* 529.3 (Mar. 2024), pp. 2866–2876. ISSN: 1365-2966. DOI: 10.1093/mnras/stae719. URL: http://dx.doi.org/10.1093/mnras/stae719.

[30] Martin White. "A marked correlation function for constraining modified gravity models". In: *Journal of Cosmology and Astroparticle Physics* 2016.11 (Nov. 2016), pp. 057–057. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2016/11/057. URL: http://dx.doi.org/10.1088/1475-7516/2016/11/057.

[31] Alejandro Aviles. "Testing modified gravity theories with marked statistics". In: (2021). arXiv: 2110.13767 [gr-qc]. URL: https://arxiv.org/abs/2110.13767.

[32] Marco Marinucci et al. "The constraining power of the Marked Power Spectrum: an analytical study". In: (2024). arXiv: 2411.14377 [astro-ph.CO]. URL: https://arxiv.org/abs/2411.14377.

[33] Haruki Ebina and Martin White. "An analytically tractable marked power spectrum". In: *Journal of Cosmology and Astroparticle Physics* 2025.01 (Jan. 2025), p. 150. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2025/01/150. URL: http://dx.doi.org/10.1088/1475-7516/2025/01/150.

[34] Elena Massara et al. "Using the Marked Power Spectrum to Detect the Signature of Neutrinos in Large-Scale Structure". In: *Phys. Rev. Lett.* 126 (1 Jan. 2021), p. 011301. DOI: 10.1103/PhysRevLett.126.011301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.126.011301.

[35] Elena Massara et al. "Cosmological Information in the Marked Power Spectrum of the Galaxy Field". In: *The Astrophysical Journal* 951.1 (July 2023), p. 70. ISSN: 1538-4357. DOI: 10.3847/1538-4357/acd44d. URL: http://dx.doi.org/10.3847/1538-4357/acd44d.

[36] Elena Massara et al. SIMBIG*: Cosmological Constraints using Simulation-Based Inference of Galaxy Clustering with Marked Power Spectra*. 2024. arXiv: 2404.04228 [astro-ph.CO]. URL: https://arxiv.org/abs/2404.04228.

[37] Somnath Bharadwaj and Pennathur Sridharan Srikant. "HI fluctuations at large redshifts: III — Simulating the signal expected at GMRT". In: *Journal of Astrophysics and Astronomy* 25.1–2 (Mar. 2004), pp. 67–80. ISSN: 0973-7758. DOI: 10.1007/bf02702289. URL: http://dx.doi.org/10.1007/BF02702289.

[38] Rajesh Mondal et al. "The effect of non-Gaussianity on error predictions for the Epoch of Reionization (EoR) 21-cm power spectrum". In: *Monthly Notices of the Royal Astronomical Society: Letters* 449.1 (Feb. 2015), pp. L41–L45. ISSN: 1745-3925. DOI: 10.1093/mnrasl/slv015. URL: http://dx.doi.org/10.1093/mnrasl/slv015.

[39] Tirthankar Roy Choudhury, Martin G. Haehnelt, and John Regan. "Inside-out or outside-in: the topology of reionization in the photon-starved regime suggested by Ly forest data". In: *Monthly Notices of the Royal Astronomical Society* 394.2 (Apr. 2009), pp. 960–977. ISSN: 1365-2966. DOI: 10.1111/j.1365-2966.2008.14383.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2008.14383.x.

[40] Suman Majumdar et al. "On the use of seminumerical simulations in predicting the 21-cm signal from the epoch of reionization". In: *Monthly Notices of the Royal Astronomical Society* 443.4 (Aug. 2014), pp. 2843–2861. ISSN: 0035-8711. DOI: 10.1093/mnras/stu1342. URL: http://dx.doi.org/10.1093/mnras/stu1342.

[41] Rajesh Mondal, Somnath Bharadwaj, and Suman Majumdar. "Statistics of the epoch of reionization (EoR) 21-cm signal – II. The evolution of the power-spectrum error-covariance". In: *Monthly Notices of the Royal Astronomical Society* 464.3 (Oct. 2016), pp. 2992–3004. ISSN: 1365-2966. DOI: 10.1093/mnras/stw2599. URL: http://dx.doi.org/10.1093/mnras/stw2599.

[42] A. Repp et al. "The impact of non-Gaussianity upon cosmological forecasts". In: *Monthly Notices of the Royal Astronomical Society* 454.4 (Oct. 2015), pp. 3533–3541. ISSN: 0035-8711. DOI: 10.1093/mnras/stv2212. eprint: https://academic.oup.com/mnras/article-pdf/454/4/3533/18509342/stv2212.pdf. URL: https://doi.org/10.1093/mnras/stv2212.

[43] C.R. Rao. "Information and the Accuracy Attainable in the Estimation of Statistical Parameters". In: *Bulletin of Calcutta Mathematical Society* 37 (1945), pp. 81–91.

[44] Harald Cramér. *Mathematical Methods of Statistics*. English. Accessed via National Library of Australia on 01 May 2025. Princeton University Press, 1946. URL: https://nla.gov.au/nla.cat-vn81100.

[45] Suman Majumdar, Somnath Bharadwaj, and T. Roy Choudhury. "Modelling the 21-cm signal from the Epoch of Reionization". In: *Monthly Notices of the Royal Astronomical Society* 443.3 (2014), pp. 2843–2854. DOI: 10.1093/mnras/stu1342. URL: https://doi.org/10.1093/mnras/stu1342.

[46] Saswata Dasgupta et al. "Interpreting the Hi 21-cm cosmology maps through Largest Cluster Statistics. Part I. Impact of the synthetic SKA1-Low observations". In: *Journal of Cosmology and Astroparticle Physics* 2023.05 (May 2023), p. 014. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2023/05/014. URL: http://dx.doi.org/10.1088/1475-7516/2023/05/014.