

---

# The Intergalactic Medium at the Epoch of Reionization

Anshuman Acharya

---



München 2025



---

# **The Intergalactic Medium at the Epoch of Reionization**

**Anshuman Acharya**

---

Dissertation  
der Fakultät für Physik  
der Ludwig-Maximilians-Universität  
München

vorgelegt von  
Anshuman Acharya  
aus Baripada

München, den 11.03.2025

Erstgutachter: Prof. Dr. Volker Springel

Zweitgutachter: Dr. Céline Péroux

Tag der mündlichen Prüfung: 27.05.2025

*There is no effort without error and shortcoming*



# Contents

<b>Zusammenfassung</b>	<b>xvii</b>
<b>Abstract</b>	<b>xix</b>
<b>1 Epoch of Reionization: The Formation of the Firsts</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 $\Lambda$ CDM Cosmology . . . . .	4
1.3 Structure formation . . . . .	7
1.4 First Stars and Galaxies . . . . .	7
1.5 The 21-cm Signal . . . . .	10
1.5.1 Theoretical background . . . . .	10
1.5.2 Observational efforts . . . . .	12
1.5.3 The LOFAR EoR Key Science Project . . . . .	13
1.5.4 Simulations for the 21-cm Signal . . . . .	15
1.5.4.1 The Fixed & Paired Method . . . . .	18
1.5.4.2 The POLAR simulations . . . . .	19
1.6 Outline of this thesis . . . . .	20
<b>2 Machine Learning for extracting the 21-cm Signal: Testing</b>	<b>21</b>
2.1 Overview . . . . .	21
2.2 Methodology . . . . .	22
2.2.1 Simulations of the 21-sm signal . . . . .	22
2.2.1.1 GRIZZLY simulations . . . . .	22
2.2.1.2 CRASH simulations . . . . .	23

2.2.2	Gaussian process regression . . . . .	23
2.2.3	Machine learning trained 21-cm kernel . . . . .	25
2.2.4	Generating mock datasets . . . . .	27
2.2.5	Recovery with MCMC . . . . .	28
2.3	Results . . . . .	31
2.3.1	$x_{\text{HI}}$ fluctuation dominated model . . . . .	31
2.3.2	$T_{\text{S}}$ fluctuation dominated model . . . . .	33
2.3.3	CRASH simulations . . . . .	37
2.3.4	Redshift dependence . . . . .	37
2.4	Discussion . . . . .	42
2.4.1	Role of the excess noise component . . . . .	42
2.4.2	Overall performance of the VAE kernel . . . . .	43
2.4.3	Limitations . . . . .	43
2.5	Summary . . . . .	44
<b>3</b>	<b>Machine Learning for extracting the 21-cm Signal: Application</b>	<b>47</b>
3.1	Overview . . . . .	47
3.2	Methodology . . . . .	47
3.3	Results . . . . .	49
3.4	Discussion . . . . .	53
3.5	Conclusions . . . . .	55
<b>4</b>	<b>Improved simulations: boosting simulation volumes</b>	<b>57</b>
4.1	Overview . . . . .	57
4.2	Methodology . . . . .	58
4.2.1	The 21-cm signal . . . . .	58
4.2.2	Simulations . . . . .	58
4.2.2.1	The Fixed & Paired approach . . . . .	59
4.3	Analysis and Results . . . . .	61
4.3.1	Skewness . . . . .	61
4.3.2	Power Spectrum . . . . .	64

4.3.3	Bispectrum . . . . .	65
4.3.4	Bubble size distribution . . . . .	68
4.4	Discussion . . . . .	69
4.4.1	Advantage of the F&P method . . . . .	69
4.4.2	Limitations and future applications . . . . .	72
4.5	Summary . . . . .	74
<b>5</b>	<b>Improved simulations: exploring broader parameter spaces</b>	<b>77</b>
5.1	Overview . . . . .	77
5.2	Methodology . . . . .	78
5.2.1	Dark-matter simulations . . . . .	78
5.2.2	Galactic and IGM properties with POLAR . . . . .	81
5.2.2.1	Semi-Analytic modeling of galaxies . . . . .	81
5.2.2.2	Radiative transfer and the 21-cm signal . . . . .	81
5.2.3	UV luminosity functions . . . . .	82
5.2.4	Reionization history . . . . .	86
5.3	Results . . . . .	89
5.3.1	Unconstrained case . . . . .	89
5.3.1.1	Galactic properties . . . . .	89
5.3.1.2	21-cm signal from the IGM . . . . .	92
5.3.2	Constrained case . . . . .	92
5.3.2.1	Galactic properties . . . . .	93
5.3.2.2	21-cm signal from the IGM . . . . .	93
5.4	Discussion . . . . .	96
5.5	Summary . . . . .	97
<b>6</b>	<b>Final Remarks and Future Prospects</b>	<b>101</b>
6.1	Summary of the key results . . . . .	103
6.2	Future prospects . . . . .	104
<b>A</b>	<b>ML-Enhanced 21-cm Signal Extraction: Investigating a low SNR case</b>	<b>107</b>

---

<b>B</b>	<b>The F&amp;P method: Improvement factor for the equilateral bispectrum</b>	<b>109</b>
<b>C</b>	<b>POLAR: UV Luminosity Functions across redshifts</b>	<b>111</b>
<b>D</b>	<b>POLAR: Maps of <math>\delta T_b</math></b>	<b>113</b>
<b>E</b>	<b>POLAR: Fiducial model with <math>f_{\text{esc}} = 25\%</math></b>	<b>115</b>
	<b>Bibliography</b>	<b>117</b>
	<b>Acknowledgements</b>	<b>131</b>

# List of Figures

1.1	Map of the Cosmic Microwave Background (CMB) as reconstructed by the Planck satellite observations reported in Planck Collaboration et al. (2014). . . . .	4
1.2	Power spectrum of the temperature fluctuations of the CMB as measured by the Planck satellite shown with solid points from Planck Collaboration et al. (2016) along with the 6 parameter $\Lambda$ CDM cosmology model fit shown with the solid line. Source: Planck Collaboration et al. (2016, 2020). . . . .	6
1.3	Evolution of the dark matter density field shown as slices of a thickness of 15 Mpc/h from the Millennium simulations (Springel et al., 2005; Gao et al., 2005; Springel, 2005) for redshifts $z = 18.3, 5.7, 1.4$ and $0.0$ . We note that while the Universe does not show any clear features at $z = 18.3$ , by $z = 0$ , the growth of the web-like architecture of large scale structure commonly known as the “Cosmic Web” becomes apparent. . . . .	8
1.4	A representation that shows the concept of the brightness temperature ( $T_b$ ) arising from the spin temperature ( $T_S$ ) and the background CMB radiation ( $T_{\text{CMB}}$ ) which is observed by our telescopes. This figure has been adapted from Zaroubi (2013). . . . .	11
1.5	The predicted 21-cm power spectra from semi-numerical simulations (blue lines) versus observed upper limits (scattered points), along with the limits achievable with $\approx 1000$ hours of observation with LOFAR, NenuFAR and SKA. Credit: Florent Mertens. . . . .	13
1.6	The 24 core stations of LOFAR (left) and a map of the 14 remote stations across the Netherlands and the various international stations across Europe (right). Credit: ASTRON. . . . .	14
1.7	The all-sky map as observed by LOFAR, with various deep field candidates indicated. From these, the ones marked with white arrows are the finally selected ones of the North Celestial Pole (NCP) and the bright compact radio source 3C 196. Credit: Léon V. E. Koopmans. . . . .	14

1.8	Evolution of various simulated properties in the main THESAN run (Kannan et al., 2022; Garaldi et al., 2022; Smith et al., 2022; Garaldi et al., 2024) versus time going from the start of the simulation run (left) to its end (right). The top panel shows the collapse of dark matter forming the cosmic web structure, and the second panel shows the gas collapsing to create galaxies. These produce ionizing photons, which ionize the atomic Hydrogen, and the evolution of the fraction of ionized Hydrogen is shown in the third panel. The final panel shows the evolution of the temperature of the gas. . . . .	17
2.1	Power spectrum of the mock dataset (purple solid line) generated using the $x_{\text{HI}}$ fluctuation dominated model from GRIZZLY. It consists of the foreground component (intrinsic sky + mode-mixing contaminants; pink dashed), the excess noise (magenta dashed), the noise (yellow dashed) and the 21-cm signal at $z = 9.16$ (grey dashed). I also show the $1 - \sigma$ upper limit (dark-yellow dotted) achievable if the dataset is thermal noise dominated, i.e. any signal below this line has $\text{SNR} < 1$ . The top (bottom) panel refers to a case with the noise corresponding to 10 (100) nights of observation. . . . .	29
2.2	As Fig. 2.1 for the $T_{\text{S}}$ fluctuation dominated model. . . . .	30
2.3	Power spectrum recovered using the Exponential (orange dashed-dotted line) and Matern32 (green dotted) Matern class functions based covariance kernels, and the VAE-based kernel (blue solid), together with the $x_{\text{HI}}$ fluctuation dominated model 21-cm signal (grey dashed) and noise (yellow dashed) at $z = 9.16$ . The $2\sigma$ uncertainty on the recovered signal for each kernel is shown as a shaded area in the corresponding colours. The top (bottom) panel refers to a case with the noise corresponding to 10 (100) nights of observation. I also plot the estimated upper limits of power at $k = 0.075 h\text{Mpc}^{-1}$ from Mertens et al. (2020, cross). Note that this value can be significantly higher than the signal due to more complex noise in real data. . . . .	32
2.4	As Figure 2.3 for the $T_{\text{S}}$ fluctuation dominated model. . . . .	34
2.5	As Figure 2.3 for the CRASH simulation of reionization. . . . .	38
2.6	<i>Top panel:</i> input 21-cm signal for the $x_{\text{HI}}$ (dashed lines) and the $T_{\text{S}}$ (solid) fluctuation dominated model are shown for $z = 8.30$ (blue), 9.16 (orange) and 10.11 (green). <i>Middle panel:</i> recovered 21-cm signal and its associated uncertainty divided by the input 21-cm signal to give the bias and scaled uncertainty (the average values over $k$ -bins for these quantities are listed in Table 2.2) for the $x_{\text{HI}}$ fluctuation dominated model. For comparison, the line of zero bias (i.e., recovered signal perfectly matching the input signal) is also shown. <i>Bottom panel:</i> same as the middle panel, but for the $T_{\text{S}}$ fluctuation dominated model. . . . .	41

- 3.1 Upper limits on the power spectrum of the 21-cm signal from 10 nights of LOFAR observational data at  $z \approx 9.1$ . **Top:** Individual upper limits case. The upper limits on  $\Delta_{21}^2$  (blue dashes with downward arrows) are obtained using the VAE-based kernel separately from the excess noise (dashed-dotted green line). The  $2\text{-}\sigma$  confidence of the recovery by the VAE-based kernel is shaded in blue, and noted to effectively extend down to zero. I also plot the upper limits obtained by M20 (crosses with downward arrows), the noise (dashed yellow line) and the  $2\text{-}\sigma$  error on the noise (dotted brown). **Bottom:** Joint upper limits case. The upper limits on the power spectrum of the 21-cm signal and excess noise are recovered together (solid blue line and shaded region), as done in M20. The other lines and symbols have the same meaning as in the top panel. . . . . 51
- 3.2 Posterior probability distributions of the GPR model hyperparameters used:  $x_1$ ,  $x_2$ , the variances  $\sigma_{21}^2$ ,  $\sigma_{\text{sky}}^2$ ,  $\sigma_{\text{mix}}^2$  and  $\sigma_{\text{ex}}^2$ , and the coherence-scales  $l_{\text{sky}}$ ,  $l_{\text{mix}}$  and  $l_{\text{ex}}$ . The purple contours show the 68%, 95% and 99.7% confidence intervals and the diagonal plots refer to the individual posterior distributions of each of the hyperparameters listed. Note that the hyperparameters of each component are largely uncorrelated with those of other components. . . . . 54
- 3.3 Residual (black solid line) obtained by subtracting the full GPR model from the data, and its corresponding  $2\text{-}\sigma$  confidence interval (grey shaded region). The noise is shown as a yellow dashed line with  $2\text{-}\sigma$  error bars. I note an excellent agreement between them. . . . . 55
- 4.1 Evolution of the volume averaged neutral hydrogen fraction versus redshift for GIC simulations (grey solid), their ensemble average (black solid) and five F&P averages (orange, magenta, red, blue and green dashed). I note that at  $z=7$   $\langle x_{\text{HI}} \rangle$  ranges between 0.72 and 0.82 for the GIC simulations. . . . . 60
- 4.2 Middle slices of the  $\delta T_{\text{b}}$  maps of (from left to right): one of the fixed simulations, its corresponding pair, their average, and the average of 2 GIC simulations, at  $z = 10, 8.3, 7.6, 7$  (with  $\langle x_{\text{HI}} \rangle = 0.99, 0.95, 0.90, 0.80$  respectively). The average of 2 GIC simulations shows clear regions of high (and low)  $\delta T_{\text{b}}$ , unlike the F&P average. With decreasing redshift, such regions begin to form for the F&P average as well, but a closer analysis of the  $\delta T_{\text{b}}$  summary statistics is necessary to analyse the difference from GIC simulations. . . . . 62
- 4.3 Redshift evolution of skewness for 35 GIC simulations (grey solid), their ensemble average (black solid), and the five F&P averages (orange, magenta, red, blue and green; dashed). . . . . 63

- 4.4 **Top:** Power spectra of GIC simulations (grey solid), their ensemble average (black solid), two F&P pairs (purple and green, dashed and dotted) and their averages (magenta and orange, solid) for  $z = 10, 8.3, 7.6,$  and  $7$  from left to right. **Bottom:** The normalised deviation of each random simulation and F&P average from the ensemble average. . . . . 65
- 4.5 **Top panels:** Equilateral triangle bispectra ( $B_{\text{equi}}$ ) for  $0.15 \leq k/(h \text{ cMpc}^{-1}) \leq 2.0$  of GIC simulations (grey solid), their ensemble average (black solid), two F&P pairs (purple and green, dashed and dotted) and their averages (magenta and orange, solid) for  $z = 10, 8.3, 7.6,$  and  $7,$  clockwise from upper left. **Bottom panels:** The normalised deviation of each random simulation and F&P average from the ensemble average. . . . . 66
- 4.6 **Top panels:** Isosceles triangle bispectra ( $B_{\text{isoc}}$ ) for  $k_1 = k_2 = 0.2 h \text{ cMpc}^{-1}$  for  $z = 10, 8.3, 7.6,$  and  $7,$  clockwise from upper left. Colours and linestyles follow Figure 4.5. **Bottom panels:** The normalised deviation of each random simulation and F&P average from the ensemble average. . . . . 67
- 4.7 Number of ionized bubbles in different radii  $r_{\text{eff}}$  bins at  $z = 10, 8.3, 7.6,$  and  $7.$  I plot their distribution for the GIC simulations (cyan violin), the ensemble average (black circles), and the two F&P averages used in Sections 4.3.2 and 4.3.3 (magenta and orange triangles). . . . . 69
- 4.8 The standard deviation curve (grey) generated by interpolating the standard deviations of  $\langle P_{\text{ws}} \rangle_m$  GIC (grey squares) versus number of sampled GIC simulations  $m$  going from 1 to 35 for  $z = 10, 8.3, 7.6$  and  $7,$  clockwise from top left. The five F&P averages are plotted as triangles (circles) with the assumption of them being  $1\text{-}\sigma$  ( $2\text{-}\sigma$ ) away from  $\langle P_{\text{ws}} \rangle_{35}$  GIC using the same colours as used in Figures 4.1 and 4.3. Their closest matches in the standard deviation curve are shown with black dotted vertical lines gives  $m_{\text{eq}}.$  . . . . . 70
- 4.9 Redshift evolution of the improvement factor in computational expense,  $f_{\text{imp}},$  when using the F&P approach. The errorbars show the range from  $f_{\text{imp,min}}$  to  $f_{\text{imp,max}}$  for the F&P averages being  $1\text{-}\sigma$  away (yellow) or  $2\text{-}\sigma$  away (blue) from an assumed distribution of F&P averages, and the points show the average,  $\langle f_{\text{imp}} \rangle.$  The top axis shows the corresponding  $\langle x_{\text{HI}} \rangle$  and the right axis shows the number  $m_{\text{eq}}$  of GIC simulations corresponding to  $f_{\text{imp}}.$  The black dashed line indicates no improvement. . . . . 73
- 5.1 Halo mass functions for my four models: fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green) at  $z = 10$  (top panel) and  $9$  (bottom). The A- and B-series are shown with dashed and dotted lines, respectively. The corresponding fit for each simulation using the Tinker et al. (2008) model is shown with a solid grey line. . . . . 80

- 5.2 A and B-series averaged UVLFs for the unconstrained (top) and constrained (bottom) cases at  $z = 10$  and  $9$  for the fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green). I also show the dust attenuated (solid) and unattenuated (dashed) UVLFs, and JWST and HST observations (Finkelstein et al., 2015; Bouwens et al., 2015, 2021; Harikane et al., 2022; Bouwens et al., 2023a,b; Harikane et al., 2023; Leung et al., 2023; McLeod et al., 2024; Adams et al., 2024, grey circles). . . . . 84
- 5.3 Maps of  $\delta T_b$  of the A-series middle slices of single cell thickness (i.e.,  $\approx 600 h^{-1}$  ckpc) for the four cosmological models in the constrained case at  $z = 12, 10, 8,$  and  $6$  (from top to bottom). Here the dark areas represent the ionized regions with  $\delta T_b = 0$ . Note that  $\delta T_b$  cannot have negative values due to the assumption of  $T_S \gg T_{\text{CMB}}$ . . . . . 86
- 5.4 Redshift evolution of the average of the A and B-series volume-averaged neutral hydrogen fraction  $\langle x_{\text{HI}} \rangle$  for the fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green) models for the unconstrained (top) and constrained (bottom) cases. Dotted lines refer to a fit to the curves, which is used for a better estimate of the redshift of reionization when the redshift resolution is too coarse. The vertical grey dashed lines indicate the redshifts observationally relevant for LOFAR ( $z = 10.11, 9.16$  and  $8.3$ ), and the black solid line is drawn at  $\langle x_{\text{HI}} \rangle = 0.5$  to guide the eye. Grey circles are a collection of observational constraints (Fan et al., 2006b; Totani et al., 2006; Ota et al., 2008; Ouchi et al., 2010; Bolton et al., 2011; Dijkstra et al., 2011; McGreer et al., 2011; Mortlock et al., 2011; Ono et al., 2012; Chornock et al., 2013; Jensen et al., 2013; Robertson et al., 2013; Schroeder et al., 2013; Pentericci et al., 2014; Schenker et al., 2014; McGreer et al., 2015; Sobacchi & Mesinger, 2015; Choudhury et al., 2015; Mesinger et al., 2015; Greig et al., 2017; Davies et al., 2018; Mason et al., 2018; Hoag et al., 2019; Greig et al., 2019; Jones et al., 2024, collected in the CoReCON module, Garaldi 2023). . . . . 87
- 5.5 A and B-series averaged star formation rate (SFR) as a function of stellar mass for the unconstrained (top) and constrained (bottom) cases, for the fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green) models at  $z = 10$  and  $9$ . Solid lines refer to the median star formation rate in each mass bin, with 16th and 84th percentiles shown as shaded regions. Grey circles are observations from various JWST programs (Treu et al., 2023; Fujimoto et al., 2023; Looser et al., 2023; Bouwens et al., 2023b; Papovich et al., 2023; Arrabal Haro et al., 2023b,a; Long et al., 2023; Leethochawalit et al., 2023; Atek et al., 2023; Robertson et al., 2023; Heintz et al., 2023a,b; Jin et al., 2023; Helton et al., 2024; Jung et al., 2024). . . . . 90

5.6	A and B-series averaged star formation efficiency (SFE) as a function of halo mass for the unconstrained (top) and constrained (bottom) cases, for the fiducial (cyan), $h$ high (red), $\sigma_8$ low (purple) and $\sigma_8$ high (green) models at $z = 10$ and 9. Solid lines refer to the median SFE in each mass bin, with 16 <sup>th</sup> and 84 <sup>th</sup> percentiles shown as shaded regions. Grey circles are Spitzer observations from Stefanon et al. (2021), and grey triangles are abundance matching estimates from Tacchella et al. (2018). . . . .	91
5.7	A and B-series averaged normalized 21-cm signal power spectrum ( $\Delta_{21\text{cm}}^2$ ) for the unconstrained (top) and constrained (bottom) cases, for the fiducial (cyan), $h$ high (red), $\sigma_8$ low (purple) and $\sigma_8$ high (green) models at $z = 9$ . . . . .	94
5.8	Redshift evolution of the A and B-series averaged normalized 21-cm signal power spectrum ( $\Delta_{21\text{cm}}^2$ ) at $k = 0.15 \text{ h cMpc}^{-1}$ for the unconstrained (top) and constrained (bottom) cases, for the fiducial (cyan), $h$ high (red), $\sigma_8$ low (purple) and $\sigma_8$ high (green) models. The vertical dashed lines indicate the redshifts relevant for LOFAR, i.e. $z = 10.11, 9.16, \text{ and } 8.3$ . . . . .	95
6.1	Flowchart of the proposed DDPM-based network. The input merger tree “conditions” the noise into super-resolving it via the denoising process of reverse diffusion. . . . .	105
A.1	As Figure 2.3 for 21cmFAST simulations of reionization. . . . .	108
B.1	As Figure 4.8 for the equilateral triangle bispectrum ( $B_{\text{equi}}$ ). . . . .	110
B.2	As Figure 4.9 for the equilateral triangle bispectrum ( $B_{\text{equi}}$ ). . . . .	110
C.1	As Figure 5.2, but for the constrained case UVLFs at $z = 12, 10, 9, 8, 7$ and 5. . . . .	112
D.1	Same as Figure 5.3 but for the B-series of simulations. . . . .	114
E.1	As Figure 5.4, but for the fiducial model, for an escape fraction of $f_{\text{esc}} = 12.5\%$ (cyan) and 25% (brown). . . . .	115
E.2	As Figure 5.8, but for the fiducial model, for an escape fraction of $f_{\text{esc}} = 12.5\%$ (cyan) and 25% (brown). . . . .	116

# List of Tables

- 2.1 I use the results of M20 as my input ( $2^{nd}$  column) for the hyperparameters ( $1^{st}$  column) to generate a mock dataset composed of the  $\mathbf{f}_{\text{sky}}$ ,  $\mathbf{f}_{\text{mix}}$ , and  $\mathbf{f}_{\text{ex}}$  components, in addition to the  $x_{\text{HI}}(T_{\text{S}})$  fluctuation dominated models from GRIZZLY to generate the 21-cm signal component at  $z = 9.16$ . For the recovery using GPR, I keep the value of  $\eta$  fixed for the  $\mathbf{f}_{\text{sky}}$ ,  $\mathbf{f}_{\text{mix}}$ , and  $\mathbf{f}_{\text{ex}}$  components, and estimate the median and the 68% confidence intervals for the coherence-scales and variances from the MCMC-based LML maximization ( $4^{th}$  to  $7^{th}$  columns) using the priors as in M20 ( $3^{rd}$  column), i.e., I either use linear scale uniform priors  $U$ , or logarithmic scale indicated as  $-$ . As the latter is over several orders of magnitude, it has not been listed and can be assumed to be an uninformed prior from  $-\infty$  to  $+\infty$ . For  $\approx 10$  nights of observation, I use  $\sigma_{\text{noise}}^2 = 74 \times 10^3 \text{ mK}^2$  (as obtained in M20). For  $\approx 100$  nights of observation, I scale it down by a factor of 2 for the  $\mathbf{f}_{\text{sky}}$  component and by a factor of 20 for the rest of the components. The values for  $\sigma_{\text{noise}}^2$  in each case are listed in the first row for each component. Note that  $\sigma_{\text{noise}}^2$  has been scaled to the field of view of the GRIZZLY simulations, which is  $3.03^\circ \times 3.03^\circ$ . . . . . 36
- 2.2 Average bias,  $\langle \text{PS}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$ , average scaled uncertainty,  $\langle \text{Err}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$ , and average z-score,  $\langle z - \text{score} \rangle_k$ , at various redshifts. These values are obtained when the signal recovery is done employing the VAE kernel with the  $x_{\text{HI}}$  (top) and the  $T_{\text{S}}$  (bottom) fluctuation dominated model and  $\approx 100$  nights of observations. . . . . 40
- 2.3 Average bias,  $\langle \text{PS}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$ , average scaled uncertainty,  $\langle \text{Err}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$ , and average z-score,  $\langle z - \text{score} \rangle_k$ , for the  $x_{\text{HI}}$  fluctuation dominated model with  $\approx 100$  nights of observations at  $z = 9.16$ . These values are obtained when the signal recovery is done employing the VAE kernel and multiplying the coherence-scale hyperparameter,  $l_{\text{ex}}$ , or the variance,  $\sigma_{\text{ex}}^2$ , listed in the second column of Table 2.1 (and as used in M20) by a factor  $\mathbf{f}_{\text{var}}$ . . . . . 42

3.1	Comparison of hyperparameters obtained in Mertens et al. (2020) and in this work. From left to right the columns refer to the hyperparameters used, the results from M20, the results obtained in this work, and the difference between the two. As the covariance kernels used for the 21-cm signal component are not the same, I just provide the values of the hyperparameters obtained in each case. Lastly, $\sigma_{\text{noise}}^2$ is $74 \times 10^3 \text{ mK}^2$ (see M20). . . . .	49
3.2	Comparing the results of Mertens et al. (2020) with those of my joint and individual upper limits of the excess noise component and 21-cm signal power spectra. The first column refers to the $k$ bins, and the following to the recovered 21-cm signal power ( $\Delta_{21}^2$ ) and its corresponding 2- $\sigma$ upper limit ( $\Delta_{21,\text{UL}}^2$ ) for M20, the joint upper limits (bottom panel in Figure 3.1) and the individual upper limits (top panel in Figure 3.1) applied in this work. The last column shows the 2- $\sigma$ error on the noise (2- $\sigma$ error <sub>noise</sub> ). . . . .	52
4.1	$f_{\text{imp}}$ is the factor of improvement on running an F&P average over running multiple GIC simulations. I report the minimum, maximum and average value of this quantity for the two cases discussed in (iv), for the five F&P averages. . . .	72
5.1	The four $N$ -body dark matter simulation models considered in this work. From left to right, the model name, the adopted value of $h$ and $\sigma_8$ , and the reference for the values. All other cosmological parameters are fixed to the Planck Collaboration et al. (2020) values. . . . .	79
5.2	List of astrophysical parameters in L-GALAXIES (see Henriques et al., 2015, 2020, for more details) that can be tuned to control galaxy formation and evolution. I show the names of the parameters in column 1, and Henriques et al. (2015) and Henriques et al. (2020) values tuned to $z < 3$ observations in columns 2 and 3. My cosmological models (fiducial, $h$ high, $\sigma_8$ low and $\sigma_8$ high) tuned to UVLF observations at $z = 10$ and 9 are shown in columns 4, 5, 6 and 7 respectively. I highlight the parameters that differ from Henriques et al. (2015, 2020) values in bold. For the unconstrained case (see text for more details), I use the fiducial model parameters for all my $N$ -body simulations. . . . .	83
B.1	As Table 4.1, showing the $f_{\text{imp}}$ improvement factor for the equilateral triangle bispectrum ( $B_{\text{equi}}$ ). . . . .	109

# Zusammenfassung

Das Verständnis des Zustands des intergalaktischen Mediums (IGM) während der Epoche der Reionisierung (EoR) ist entscheidend, um die Entstehung der ersten Sterne, Galaxien und kosmischen Strukturen zu entschlüsseln. Diese Dissertation präsentiert eine umfassende Untersuchung, die theoretische Grundlagen, fortgeschrittene Simulationstechniken und innovative maschinelle Lernmethoden kombiniert, um das rotverschobene 21-cm-Signal aus dem IGM zu modellieren und zu extrahieren.

Im ersten Teil der Arbeit konzentriere ich mich auf die Verbesserung von Techniken zur Extraktion des 21-cm-Signalleistungsspektrums aus Radiodaten. Insbesondere konzentriere ich mich auf die Verbesserung der Vordergrundsubtraktionstechnik auf der Grundlage der Gaußschen Prozessregression (GPR) unter Verwendung von maschinellem Lernen (ML). Das heißt, anstatt mich auf die Verwendung analytischer Funktionen zu verlassen, um eine "Vorlage" für das 21-cm-Signalleistungsspektrum zu erstellen, verwende ich einen auf Variational Auto-Encoder (VAE) basierenden Algorithmus, der auf simulierten Leistungsspektren aus den GRIZZLY-Simulationen trainiert wurde. Ich habe dann simulierte Datenwürfel erstellt und 21-cm-Signalleistungsspektren nicht nur aus den GRIZZLY-Simulationen, sondern auch aus Simulationen mit völlig anderer Architektur wie CRASH und 21cmFAST injiziert und gezeigt, dass das ML-gestützte GPR das 21-cm-Signal erfolgreich innerhalb von  $2\sigma$ -Unsicherheiten extrahieren konnte, während es in allen Fällen die Form des Spektrums besser erfasste. Ich führte weitere Tests mit unterschiedlichen Rauschpegeln und Rotverschiebungen durch und untersuchte auch die Rolle des systematischen Rauschüberschusses, der bei LOFAR-Beobachtungen festgestellt wurde. Anschließend wandte ich dies auf 141 Stunden echter LOFAR-Daten an, wobei alle anderen Aspekte außerhalb der Vorlage für das 21-cm-Signalleistungsspektrum gleich blieben. Ich stelle fest, dass die überarbeitete Pipeline eine konservative  $2\sigma$ -Obergrenze für das 21-cm-Leistungsspektrum von  $\Delta_{21}^2 < (80)^2 \text{ mK}^2$  bei  $k = 0,075 h \text{ Mpc}^{-1}$  ergibt, wenn ich das überschüssige Rauschen nicht vom 21-cm-Signal trenne, wie es bei früheren Ergebnissen der Fall war. Durch die Anwendung von Bias-Korrekturen zur Beseitigung des übermäßigen Rauschens kann die Obergrenze jedoch weiter auf  $\Delta_{21}^2 < (25)^2 \text{ mK}^2$  gesenkt werden. Aufgrund des begrenzten Verständnisses der Art der in den Daten vorhandenen übermäßigen Rauschkomponente muss jedoch noch bestätigt werden, ob die erhaltenen Bias-korrigierten Obergrenzen tatsächlich das gesamte 21-cm-Signal enthalten. Dennoch stellt dieser ML-basierte Ansatz eine der vielversprechendsten Methoden zur Extraktion des 21-cm-Signalleistungsspektrums dar, was für die spätere Erkennung des 21-cm-Signals mit mehr Beobachtungsstunden von LOFAR und/oder mit zukünftigen Radioteleskopen wie SKA

von entscheidender Bedeutung wäre.

Im zweiten Teil der Arbeit konzentriere ich mich auf die Verbesserung der Modellierung des 21-cm-Signals durch die Implementierung des Fixed & Paired (F&P)-Ansatzes zur Unterdrückung der kosmischen Varianz, ohne dass massive Simulationsvolumina erforderlich sind. Herkömmliche Methoden stützen sich auf groß angelegte Simulationen, um schwache 21-cm-Fluktuationen zu erfassen, aber F&P erhöht das statistische Volumen, indem es anfängliche Dichtestörungen fixiert und Simulationen mit invertierten Phasen koppelt. Dies steigert die Effizienz um mindestens das 3,5-Fache für das Leistungsspektrum und um das 5-Fache für das gleichseitige Bispektrum, wodurch die Rechenkosten erheblich gesenkt werden. Vergleiche wichtiger statistischer Messgrößen – darunter das Leistungsspektrum, das Bispektrum, die Schiefe und die Blasengrößenverteilung – bestätigen, dass der F&P-Durchschnitt eng mit den Ensemble-Durchschnitten übereinstimmt, was ihn zu einer zuverlässigen und effizienten Methode für die Vorhersage von beobachtbaren 21-cm-Signalen macht, was für die Interpretation von Daten aus Radiointerferometern der nächsten Generation von entscheidender Bedeutung ist.

Im dritten Teil der Arbeit untersuche ich die Erweiterung des physikalischen Parameterraums von Simulationen, die zur Ableitung astrophysikalischer Eigenschaften aus dem 21-cm-Signal verwendet werden. Insbesondere muss geprüft werden, ob die Variabilität einiger Parameter zu Kompromissbeziehungen zwischen ihnen führt, sodass sich die Bandbreite der realisierbaren Modelle, die mit den Observablen übereinstimmen, erweitert. Dazu verwende ich die POLAR-Simulationen, die  $N$ -Körper-Simulationen, die mit GADGET-4 ausgeführt werden, mit dem semi-analytischen Galaxienbildungsmodell L-GALAXIES und GRIZZLY nachbearbeiten. Ich variierte zwei wichtige kosmologische Parameter – die Hubble-Konstante ( $H_0$ ) und die Materie-Cluster-Amplitude ( $\sigma_8$ ) – und passte astrophysikalische Parameter wie die Sternentstehungseffizienz und die pro Supernova-Explosion freigesetzte Energie so an, dass alle Simulationen mit den UV-Leuchtkraftfunktionen aus den neuesten JWST-Beobachtungen bei Rotverschiebungen  $z = 10$  und  $9$  übereinstimmen. Ich stelle fest, dass die vorhergesagten 21-cm-Leistungsspektren trotz der Verwendung unterschiedlicher Sätze kosmologischer und astrophysikalischer Parameter mit den aktuellen oberen Beobachtungsgrenzen übereinstimmen, was deutlich macht, dass verschiedene Modelle realisierbar sind. Dies unterstreicht das Potenzial der Kombination von galaktischen und IGM-Beobachtungsgrößen zur gemeinsamen Eingrenzung der physikalischen Prozesse, die die Reionisierung antreiben.

Insgesamt macht die Dissertation bedeutende Fortschritte sowohl bei den Simulations- als auch bei den Datenanalyseverfahren für EoR-Studien. Durch die Weiterentwicklung von ML-gestützten GPR-Methoden zur robusten Vordergrundreduzierung und den Einsatz innovativer Simulationsstrategien zur Überwindung der kosmischen Varianz und der begrenzten Parameterabdeckung trägt diese Arbeit zu einem tieferen und präziseren Verständnis des frühen Universums bei. Der hier skizzierte integrierte Ansatz bringt uns einer zuverlässigen Erkennung und Interpretation des schwer fassbaren 21-cm-Signals aus dem IGM näher und bietet ein neues Fenster zur kosmischen Strukturbildung während der Epoche der Reionisierung.

# Abstract

Understanding the state of the intergalactic medium (IGM) during the Epoch of Reionization (EoR) is pivotal to unraveling the formation of the first stars, galaxies, and cosmic structures. This thesis presents a comprehensive investigation that combines theoretical foundations, advanced simulation techniques, and innovative machine-learning methods to model and extract the redshifted 21-cm signal from the IGM.

In the first part of the thesis, I focus on improving techniques for extracting the 21-cm signal power spectrum from radio data. In particular, I focus on improving the foreground subtraction technique based on Gaussian Process Regression (GPR) using Machine Learning (ML). That is, instead of relying on using analytic functions to build a “template” for the 21-cm signal power spectrum, I use a Variational Auto-Encoder (VAE) - based algorithm trained on mock power spectra from the GRIZZLY simulations. I then built simulated datacubes and injected mock 21-cm signal power spectra not just from the GRIZZLY simulations but also from simulations with completely different architecture like CRASH and 21cmFAST and showed that the ML-enhanced GPR could successfully extract the 21-cm signal within  $2\sigma$  uncertainties while better capturing the shape of the spectrum in all cases. I performed additional tests with differing noise levels and redshifts and also explored the role of the systematic excess noise found in LOFAR observations. Subsequently, I applied this to 141 hours of real LOFAR data, keeping all other aspects beyond the template for the 21-cm signal power spectrum the same. I find that if I do not separate the excess noise from the 21-cm signal, as was done for past results, the revised pipeline yields a conservative  $2\sigma$  upper limit on the 21-cm power spectrum of  $\Delta_{21}^2 < (80)^2 \text{ mK}^2$  at  $k = 0.075 h \text{ Mpc}^{-1}$ . But on applying bias corrections to remove the excess noise, the upper limit can be further pushed down to  $\Delta_{21}^2 < (25)^2 \text{ mK}^2$ . However, due to the limited understanding of the nature of the excess noise component present in the data, it is yet to be confirmed if the bias-corrected upper limits obtained indeed contain the entirety of the 21-cm signal. Nevertheless, this ML-based approach provides one of the most promising methods of extracting the 21-cm signal power spectrum, which would be crucial to eventually detect the 21-cm signal with more hours of observation from LOFAR and/or with upcoming radio telescopes like SKA.

In the second part of the thesis, I focus on improving the modeling of the 21-cm signal by implementing the Fixed & Paired (F&P) approach to suppress cosmic variance without requiring massive simulation volumes. Traditional methods rely on large-scale simulations to capture faint 21-cm fluctuations, but F&P enhances statistical volume by fixing initial density perturbations and pairing simulations with inverted phases. This boosts efficiency by at least 3.5 times for the power

spectrum and 5 times for the equilateral bispectrum, significantly reducing computational cost. Comparisons of key statistical measures—including the power spectrum, bispectrum, skewness, and bubble size distribution—confirm that the F&P average closely matches ensemble averages, making it a reliable and efficient method for predicting 21-cm signal observables, which is crucial for interpreting data from next-generation radio interferometers.

In the third part of the thesis, I investigate the broadening of the physical parameter space of simulations used to infer astrophysical properties from the 21-cm signal. In particular, it is necessary to check if allowing some parameters to vary leads to trade-off relationships between them, such that the range of viable models that match observables broadens. For this, I use the POLAR simulations, which post-processes  $N$ -body simulations run with GADGET-4 with the semi-analytic galaxy formation model L-GALAXIES and GRIZZLY. I varied two key cosmological parameters—specifically the Hubble constant ( $H_0$ ) and the matter clustering amplitude ( $\sigma_8$ ) and adapted astrophysical parameters such as the star formation efficiency and the energy released per supernova explosion such that all simulations match the UV luminosity functions from the latest JWST observations at redshifts  $z = 10$  and 9. I find that despite adopting distinct sets of cosmological and astrophysical parameters, the predicted 21-cm power spectra remain consistent with current observational upper limits, thus making it clear that various models can be viable. This highlights the potential of combining both galactic and IGM observables for jointly constraining the physical processes driving reionization.

Overall, the thesis makes significant strides in both simulation and data-analysis techniques for EoR studies. By advancing ML-enhanced GPR methods for robust foreground mitigation and employing innovative simulation strategies to overcome cosmic variance and limited parameter coverage, this work contributes to a deeper and more precise understanding of the early Universe. The integrated approach outlined herein brings us closer to a reliable detection and interpretation of the elusive 21-cm signal from the IGM, offering a new window into cosmic structure formation during the Epoch of Reionization.

# Chapter 1

## Epoch of Reionization: The Formation of the Firsts

*Where's the glory in repeating what others have done?*

---

- Rick Riordan

### 1.1 Overview

The building of knowledge is an iterative and cumulative process. Any work that advances our understanding of cosmology necessarily builds upon foundational discoveries that attempt to address the enduring question: “How did the Universe come to be as it is today?”

Early models of the cosmos were rooted in geocentrism, placing the Earth at the center of the Universe. In this view, celestial bodies—including the Sun, planets, and a celestial sphere studded with stars—revolved around a stationary Earth. This model, epitomized by the Ptolemaic system, dominated Western thought for centuries until the heliocentric revolution spearheaded by Nicolaus Copernicus in the 16th century ([Copernicus, 1543](#)). His work was further substantiated by Galileo Galilei’s telescopic observations ([Galilei, 1610](#)) and Johannes Kepler’s laws of planetary motion ([Kepler, 1609](#)), leading to the realization that Earth was merely one of several planets orbiting the Sun.

Yet, even the Sun proved not to be special. By the early 20th century, astronomers had come to recognize that it was simply one of the billions of stars comprising the Milky Way galaxy ([Shapley, 1921](#)). Even the notion that the Milky Way encompassed the entire Universe persisted until Edwin Hubble’s groundbreaking observations at the Mount Wilson Observatory. Hubble demonstrated that distant “nebulae” were, in fact, independent galaxies far beyond the Milky Way. He further established that these galaxies were receding from us at velocities proportional to their distance—a phenomenon now known as Hubble’s Law ([Hubble, 1929](#)). This led to proposals of

an expanding Universe, with various explanations of the same, the most popular of which were that of the “Big Bang” theory and the “Steady State” theory.

The discovery of the cosmic microwave background (CMB) radiation in 1965 by Arno Penzias and Robert Wilson provided crucial evidence supporting the Big Bang theory (Penzias & Wilson, 1965). This observation, coupled with advances in theoretical modeling, solidified the modern  $\Lambda$ CDM (Lambda-Cold Dark Matter) model of the expanding Universe. This framework posits a Universe composed predominantly of cold dark matter and dark energy, the latter represented by the cosmological constant “ $\Lambda$ ” (Peebles, 1968; Planck Collaboration et al., 2014, 2016, 2020).

According to the  $\Lambda$ CDM model, the Universe originated from a singularity approximately 13.8 billion years ago in an event known as the Big Bang. This was followed by a brief but rapid phase of expansion called inflation (Guth, 1981; Linde, 1982). During this period, the Universe was extraordinarily hot and homogeneous. As it continued expanding and cooling, a plasma of quarks, leptons, photons, and gluons gradually transitioned to a state in which quarks and gluons combined to form protons and neutrons.

Roughly 380,000 years after the Big Bang, the Universe underwent the Epoch of Recombination, during which free electrons combined with protons to form neutral hydrogen and helium atoms. This dramatic reduction in free charged particles allowed photons to travel freely, leading to the decoupling of radiation from matter. The photons from this period persist today as the CMB, a faint glow that permeates the cosmos and serves as a critical observational window into the early Universe (Peebles, 1968).

Following the Epoch of Recombination, the Universe entered a phase where gravity became the dominant force driving the formation of structures. During this period, dark matter, an elusive and non-luminous component of the cosmos, began to clump together under its own gravity, forming what are known as dark matter halos. These halos acted as gravitational wells, pulling in ordinary baryonic matter—comprising protons, neutrons, and electrons—which subsequently cooled and condensed within these potential wells.

As the baryonic matter accumulated and cooled, the first stars ignited, marking the onset of a transformative era in cosmic history known as the Cosmic Dawn. These first stars, often referred to as Population III stars (Klessen & Glover, 2023), were typically massive and short-lived, composed almost entirely of hydrogen and helium, the simplest elements. Their formation is intrinsically tied to the creation of the first galaxies and other astrophysical structures (Bromm & Yoshida, 2011), fundamentally altering the dynamics and chemistry of the Universe.

The intense ultraviolet radiation emitted by these early stars played a crucial role in reionizing the intergalactic medium (IGM). This radiation had sufficient energy to ionize neutral hydrogen atoms, splitting them back into protons and electrons. As it is a reversion back to the Universe being ionized, this period where most of the neutral atomic Hydrogen was ionized is referred to as the “Epoch of Reionization” (EoR). This epoch, spanning redshifts between  $z \approx 5$  and  $z \approx 15$ , is pivotal for understanding the formation and evolution of the first luminous structures and their influence on the surrounding interstellar and intergalactic media (Ciardi & Ferrara, 2005; Morales & Wyithe, 2010; Pritchard & Loeb, 2012; Furlanetto, 2016; Liu & Shaw, 2020).

From observations of the Gunn-Peterson troughs of high- $z$  quasars (Becker et al., 2001; Fan et al., 2006a) and the optical depth for Thomson scattering of the CMB radiation (Planck Collaboration et al., 2016), we can deduce that most reionization took place in the range  $6 \lesssim z \lesssim 10$ , with recent observations suggesting an end of reionization at  $z < 6$  (see e.g. Becker et al. 2015 and Bosman et al. 2022). There exist multiple indirect probes to study this period. For example, the evolution of the observed Lyman- $\alpha$  emitter luminosity function at  $z > 6$  (Clément et al., 2012; Schenker et al., 2013), and Lyman- $\alpha$  absorption profiles to distant quasars (Mortlock, 2016; Greig et al., 2017; Davies et al., 2018). However, questions on how exactly the *first of everything we see today* formed, their timelines, rates, and exact processes are still open questions.

The goal of this thesis is to contribute to the improvements in the chances of detection of one of the most sensitive probes of this period, the fluctuations of the redshifted 21-cm line of neutral Hydrogen in the IGM against the CMB (Hogan & Rees, 1979; Madau et al., 1997; Shaver et al., 1999; Tozzi et al., 2000; Ciardi & Madau, 2003; Zaroubi, 2013). A statistical detection of the strength of these 21-cm brightness temperature fluctuations can allow us to constrain our models of the early Universe and the formation of the first stars and galaxies. This approach provides a different perspective, where instead of looking at the “sources” of ionizing photons, we look at the “sink”, which is the IGM, to understand the EoR.

However, in order to actually detect the 21-cm signal even as a power spectrum, we need to improve our detection techniques. For this, I have worked on implementing a Machine Learning-based approach. In this approach, I generate mock 21-cm signal power spectra from simulations and use them as a training set to build a “template”. This template is used with the signal extraction algorithm to identify the part of the data that arises from the 21-cm signal and allows for its subsequent extraction.

Apart from signal extraction, it is imperative to improve numerical modelling concurrently to ensure we can understand the physical processes at play that lead to the observed 21-cm signal power spectrum. As of now, the main limitations have been in the trade-offs between simulated volume, resolution, and physical modelling, forced by limits in computational power. In my PhD, I have explored possible avenues to boost the effective volume/size of simulations used, as well as broaden the parameter space of physical processes that we model.

In this chapter, I provide an overview of the standard  $\Lambda$ CDM cosmological model assumed throughout the work of this thesis in Section 1.2 and an introduction to structure formation in Section 1.3. Next, I discuss our current understanding of the formation of the first stars and galaxies in Section 1.4, and the 21-cm signal in Section 1.5. In Section 1.5, I provide a brief primer on the theoretical background of the 21-cm signal as an observable in subsection 1.5.1, the status of current observations of the 21-cm signal at the EoR in subsection 1.5.2, and the LOFAR EoR Key Science Project in subsection 1.5.3. Next, in subsection 1.5.4, I discuss various modelling efforts for understanding the EoR and the methodologies used in this thesis to improve upon them. Lastly, in Section 1.6, I chart out the structure of this thesis.

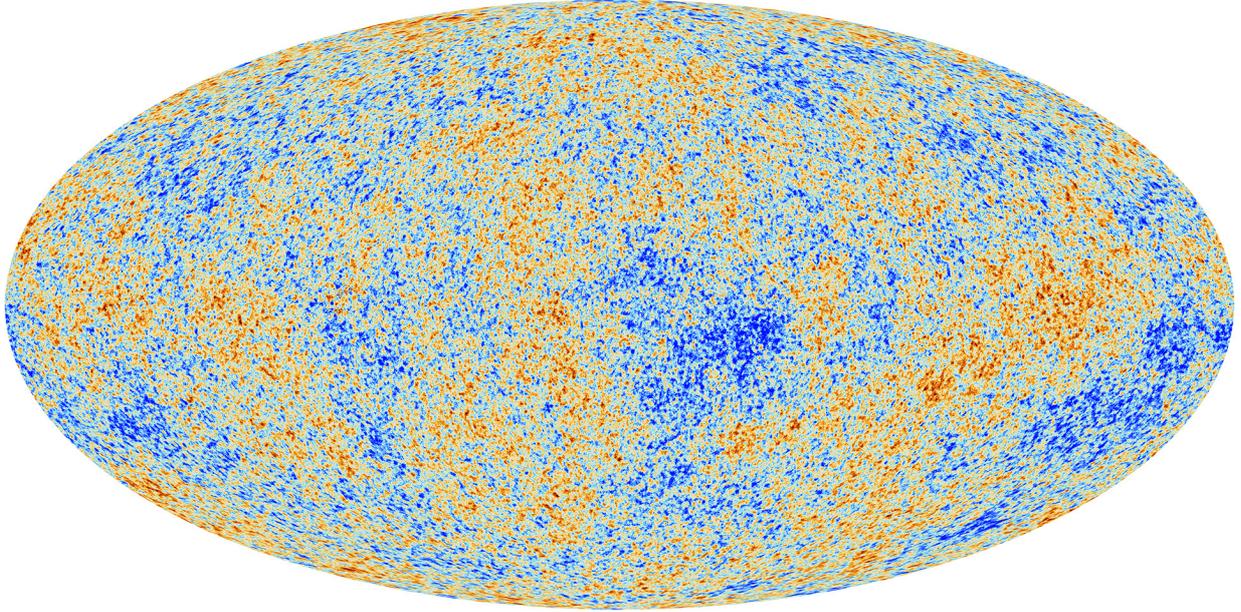


Figure 1.1: Map of the Cosmic Microwave Background (CMB) as reconstructed by the Planck satellite observations reported in [Planck Collaboration et al. \(2014\)](#).

## 1.2 $\Lambda$ CDM Cosmology

The  $\Lambda$ CDM (Lambda - Cold Dark Matter) model is the prevailing cosmological framework that describes the large-scale structure and evolution of the Universe. This model integrates the cosmological constant associated with dark energy, and cold dark matter (CDM), alongside ordinary matter and radiation. It is grounded in two broad principles: (i) the “Cosmological Principle”, and (ii) that General Relativity describes gravity correctly, which in turn governs the overall behaviour of the Universe ([Hamilton, 2014](#)). Both of these assumptions are supported by a range of observational evidence.

The first principle, i.e., the Cosmological Principle, simply states that at large enough scales, the Universe is isotropic, homogeneous, and uniformly expanding in all directions. This assumption is consistent with observations of the CMB. The CMB is found to be isotropic and can be approximated as a black-body spectrum with a temperature of  $T_{\text{CMB}} \approx 2.7$  K. Nevertheless, fluctuations in temperature are observed, at the scale of  $\Delta T/T \approx 10^{-5}$  ([Planck Collaboration et al., 2014, 2016, 2020](#)) that indicate deviations from homogeneity.

For having information on the geometry of spacetime and the evolution of structure formation, the  $\Lambda$ CDM model assumes that the geometry of the Universe can be described by the Friedman-Lemâitre-Robertson–Walker (FLRW; [Friedman, 1922](#); [Lemâitre & Eddington, 1931](#)) metric which is a solution to Einstein’s field equations of General Relativity. The FLRW metric

is used to characterise a spacetime that is expanding or contracting uniformly as:

$$ds^2 = c^2 dt^2 - a(t)^2 \left( \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right) \quad (1.1)$$

Here,  $a(t)$  represents the scale factor, which describes how distances in the Universe change over time;  $k$  denotes the curvature parameter, indicating whether the Universe is open ( $k < 0$ ), closed ( $k > 0$ ), or flat ( $k = 0$ ); and  $d\Omega^2$  encapsulates the angular part of the metric.

The dynamics of the scale factor  $a(t)$  are governed by the Friedmann equations (Friedman, 1922; Friedmann, 1924), derived from the FLRW metric and Einstein's field equations. The first Friedmann equation is thus found to be:

$$\left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \quad (1.2)$$

In this equation,  $\dot{a}$  is the time derivative of the scale factor,  $G$  is the gravitational constant,  $\rho$  is the total energy density of the Universe, and  $\Lambda$  is the cosmological constant associated with dark energy. The second Friedmann equation provides further insight into the acceleration of the Universe's expansion:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left( \rho + \frac{3p}{c^2} \right) + \frac{\Lambda c^2}{3} \quad (1.3)$$

Here  $\ddot{a}$  is the second time-derivative of the scale factor, and  $p$  represents the pressure. These equations collectively describe how the Universe's expansion rate is influenced by its energy content and geometry. On representing  $H(t) = \frac{\dot{a}}{a}$ , which is the Hubble parameter (with the present day value given as the Hubble constant,  $H_0$ ) and critical density  $\rho_c = \frac{3H_0^2}{8\pi G}$ , we obtain the dimensionless parameters:

$$\Omega_m = \rho / \rho_c \quad (1.4)$$

$$\Omega_\Lambda = \frac{\Lambda c^2}{3H_0^2} \quad (1.5)$$

$$\Omega_k = -\frac{kc^2}{a^2 H_0^2} \quad (1.6)$$

Assuming a flat Universe sets  $\Omega_k \approx 0$ . Additionally, assuming a 6 parameter  $\Lambda$ CDM model provides a good fit to observations of CMB anisotropies by space-based observatories like COBE (Lindley, 1989), WMAP (Bennett et al., 2003), and most recently, Planck (as shown in Figure 1.1; Planck Collaboration et al., 2014, 2016, 2020), by fitting the power spectrum of the temperature fluctuations as shown in Figure 1.2. These six parameters are  $\Omega_m$ ,  $\Omega_\Lambda$ ,  $H_0$ , the baryon density parameter  $\Omega_b$ , the root-mean-square mass fluctuations on  $8h^{-1}$  Mpc scale (where  $h = H/100$ )  $\sigma_8$ , and the spectral index of the primordial density fluctuation,  $n_s$ .

The latest results from Planck Collaboration et al. (2020) measure  $\Omega_m = 0.3111 \pm 0.0056$ , which is the contribution of non-relativistic matter to the cosmic energy density, and dark

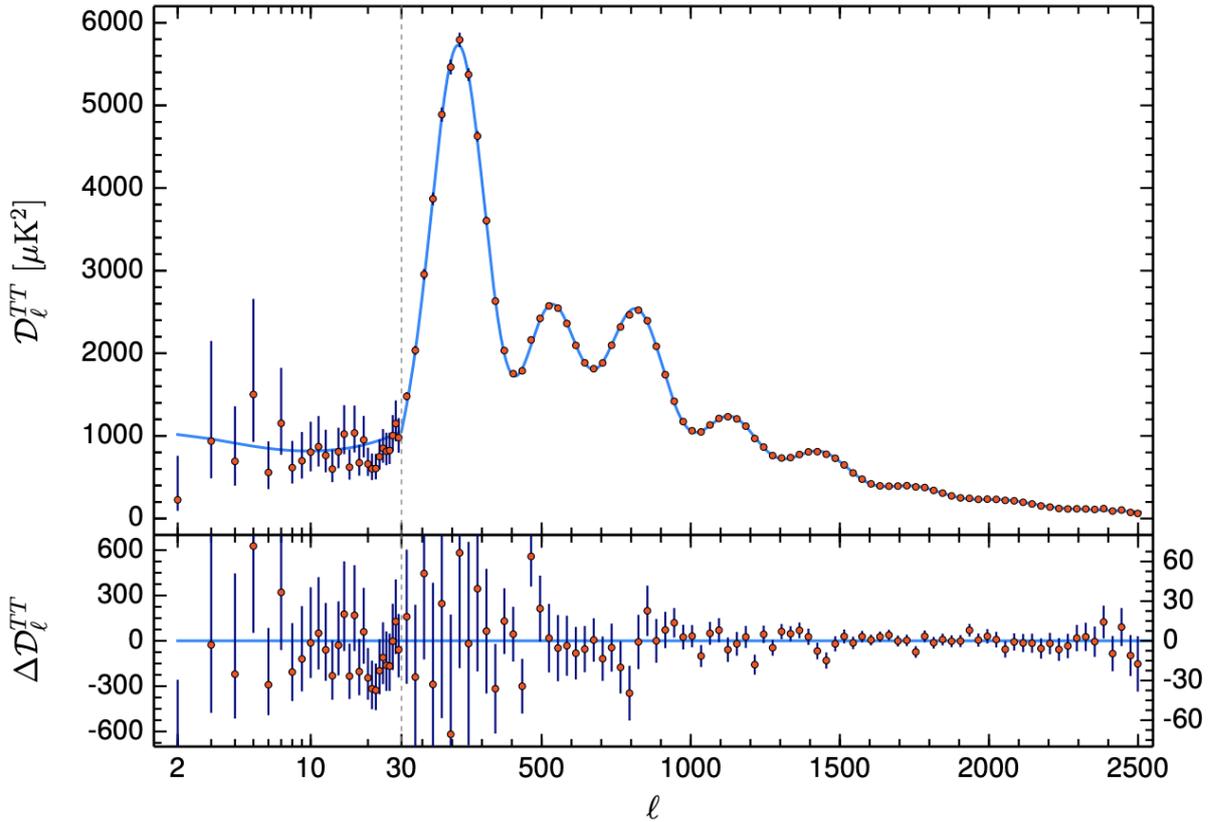


Figure 1.2: Power spectrum of the temperature fluctuations of the CMB as measured by the Planck satellite shown with solid points from [Planck Collaboration et al. \(2016\)](#) along with the 6 parameter  $\Lambda$ CDM cosmology model fit shown with the solid line. Source: [Planck Collaboration et al. \(2016, 2020\)](#).

energy contributing  $\Omega_\Lambda = 0.6889 \pm 0.0056$ . Additionally,  $H_0 = 67.66 \pm 0.42 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $n_s = 0.9665 \pm 0.0038$ ,  $\sigma_8 = 0.8102 \pm 0.0060$ , and a baryonic matter component of  $\Omega_b \approx 0.04897$ . Given the small fraction of baryonic matter, it is clear that matter in the Universe is predominantly dark. Further, we find that  $\Omega_m + \Omega_\Lambda \approx 1$ , as expected from a flat Universe. This flatness is explained by inflation ([Guth, 1981](#); [Linde, 1982](#)), which also explains the large-scale isotropy and homogeneity. The small-scale perturbations in the CMB are accounted for via quantum fluctuations in the energy density of a scalar field of inflation. These density perturbations can be described as a power spectrum,  $P_r(k) \propto k^{n_s-1}$ , and the measured value of  $n_s$  is in excellent agreement with inflationary models ([Mukhanov & Chibisov, 1981](#)).

Given the agreement with observations along with insights in properties of dark energy and dark matter, the  $\Lambda$ CDM model is chosen as the standard model throughout this thesis.

## 1.3 Structure formation

While the Universe is believed to have been largely featureless at the time of the CMB, i.e., at the Epoch of Recombination, the density perturbations discussed in the earlier section affect the dark matter distribution. This, in turn, leads to runaway gravitational collapse, and once the density contrast grows beyond  $\Delta\rho/\rho \approx 1$ , the gravitational attraction becomes stronger than the expansion of the Universe (often referred to as “Hubble flow”), leading to non-linear collapse and the form of virialised structures known as dark matter haloes. The virial theorem essentially states that the dark matter particles in the system contain their total energy from the kinetic energy of motion and gravitational potential energy such that:

$$E_{\text{kin}} = -\frac{1}{2}E_{\text{pot}} \quad (1.7)$$

This redistribution of energy to reach a quasi-equilibrium state is known as “violent relaxation” (Lynden-Bell, 1967). Assuming spherical collapse from an initial Gaussian random field given by inflation, we can get the number density of haloes at a given halo mass as a function of redshift from the Press-Schechter formalism (Press & Schechter, 1974):

$$M \frac{dn}{dM} = -\left(\frac{2}{\pi}\right)^{-0.5} \frac{d(\ln\sigma)}{d(\ln M)} \frac{\rho_0}{M} \nu_c e^{-\nu_c^2/2} \quad (1.8)$$

where  $\rho_0$  is the mean density at a given redshift,  $\sigma$  is the standard deviation of the density contrast smoothed through a certain window,  $\nu_c$  is the minimum number of standard deviations of a collapsed fluctuation, and  $M$  is the mass of the halo. The dark matter halo statistics obtained from cosmological simulations (e.g., the Millenium simulations as shown in Figure 1.3) are found to be in good agreement with this model, especially when elliptical collapse and halo merger statistics are taken into account (see Sheth et al., 2001). Nevertheless, some issues exist, mainly that the Press-Schechter model overpredicts the abundance of low-mass halos while at the same time underpredicts the amount of high-mass halos. Improvements to this have been explored by reducing the dependence on the exact values of the cosmological parameters used in the  $\Lambda$ CDM model (see Tinker et al., 2008), and redshift (see Watson et al., 2013).

## 1.4 First Stars and Galaxies

The matter that we can see in the Universe, be it from the naked eye or through various ground and space-based telescopes, is baryonic. Thus, while dark matter forms the majority of all matter, it is important to understand how baryons condensed to form the first stars, galaxies, and every other astrophysical object that we can see in the sky. This is especially important as the behaviour of baryons is more complex than that of dark matter. For example, dark matter behaves like a collisionless fluid, i.e., dark matter particles do not collide frequently enough and exchange energy, and thus mostly just interact gravitationally. In contrast, baryons can do so and achieve

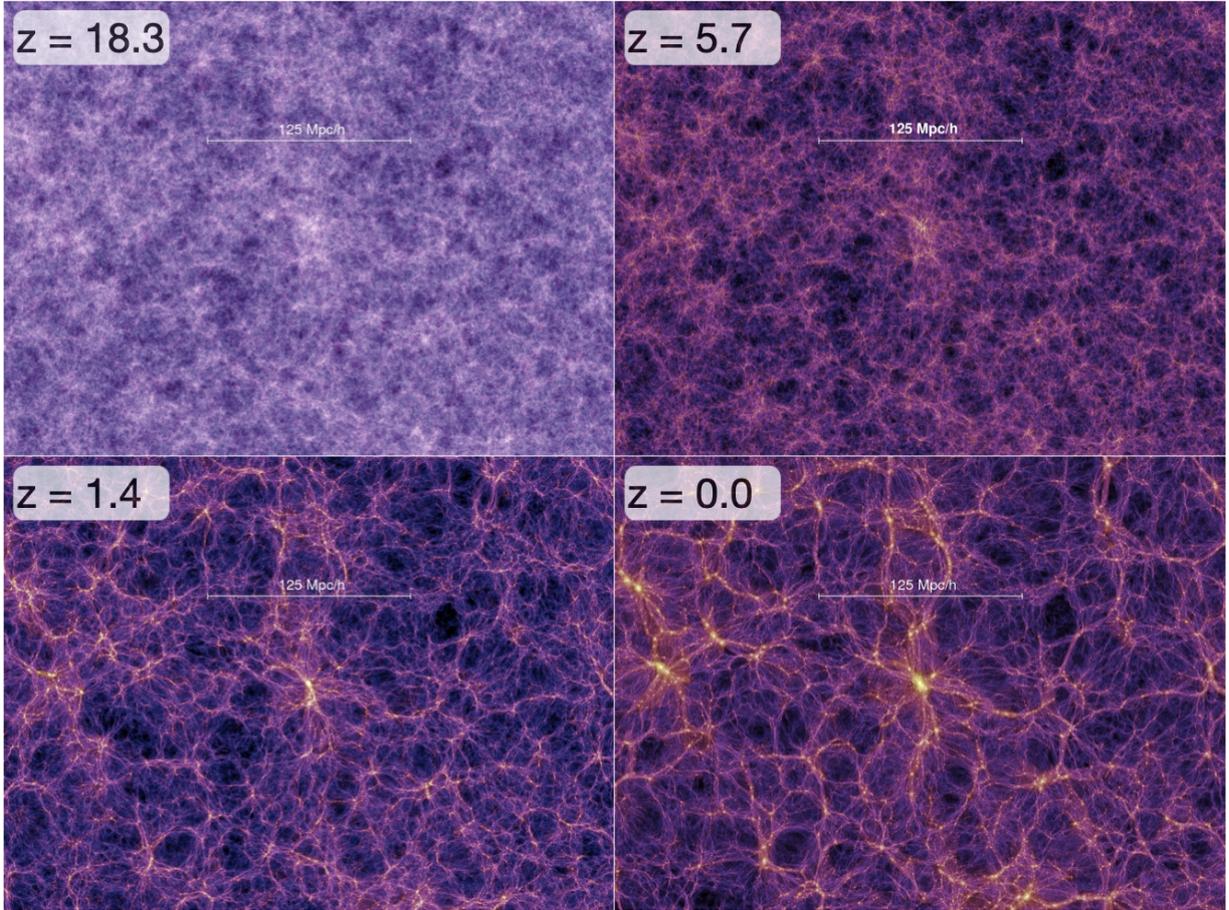


Figure 1.3: Evolution of the dark matter density field shown as slices of a thickness of 15 Mpc/h from the Millennium simulations (Springel et al., 2005; Gao et al., 2005; Springel, 2005) for redshifts  $z = 18.3$ , 5.7, 1.4 and 0.0. We note that while the Universe does not show any clear features at  $z = 18.3$ , by  $z = 0$ , the growth of the web-like architecture of large scale structure commonly known as the “Cosmic Web” becomes apparent.

thermal equilibrium. Thus, when baryons condense into the gravitational potential wells formed by dark matter, they can undergo shock heating that increases their temperature, which in turn provides pressure support against further gravitational collapse. However, in the denser regions within the halo, baryons have more frequent collisions, leading to more efficient cooling. Thus, they lose their pressure support, which drives further gravitational collapse into the potential well of the dark matter halo. Eventually, this process collects the baryons into a gas cloud at the center of the dark matter halo. This cloud finally has a balance between pressure and gravity and thus achieves hydrostatic equilibrium. The stage when the cloud reaches this state can be given by the Jeans criterion (Jeans & Darwin, 1902), which is when the sound-crossing time across the system is greater than the free-fall time. The mass required for a system to reach this stage is referred to

as Jeans mass, and it is given as:

$$M_{\text{jeans}} = \left( \frac{5kT}{G\mu m_p} \right)^{3/2} \left( \frac{3}{4\pi\rho} \right)^{1/2} \quad (1.9)$$

$$\approx 2M_{\odot} \left( \frac{c_s}{0.2 \text{ km/s}} \right)^3 \left( \frac{n}{10^3 \text{ cm}^{-3}} \right)^{-1/2}$$

where  $c_s$  is the speed of sound in the cloud,  $k$  is the Boltzmann constant,  $\rho$  is the density of the gas,  $\mu$  is the mean molecular weight,  $T$  is the temperature,  $m_p$  is the proton mass,  $G$  is the gravitational constant, and  $n$  is the number density of the gas.

This equilibrium, however, is still fairly unstable (Hoyle, 1953), with growing density leading to fragmentation due to self-gravity triggering further collisional heating in these fragments. In present-day galaxies, the presence of metals (elements beyond Helium in the periodic table) allows for more channels of cooling, such as through emission lines, but in this period, most of the gas is pristine, i.e., is made up of mostly Hydrogen and some amount of Helium. Thus, temperatures continue to be fairly higher in these fragments as compared to their surroundings, as they reach densities high enough to trigger nuclear fusion. As the Jeans mass is proportional to the temperature, this implies that the first stars, also known as Population III stars, were significantly more massive than stars observed at lower redshifts, with estimates in the range of  $10^2 - 10^3 M_{\odot}$  (Abel et al., 1999; Bromm et al., 1999; Abel et al., 2002; O'Shea & Norman, 2007; Kulkarni et al., 2019; Ventura et al., 2023). However, there is no consensus on their exact mass range, with studies incorporating turbulence (Turk et al., 2012; Wollenberg et al., 2020), radiative feedback (Hosokawa et al., 2011, 2012; Hirano et al., 2014, 2015; Hosokawa et al., 2016; Jaura et al., 2022) or simply higher resolution (Turk et al., 2012; Prole et al., 2022) show them to be similar to the next generation of stars, that is, the Population II stars.

Apart from the nuclear reactions that took place within these stars that led to the formation of some of the lighter metals, these stars are also believed to have led to the first supernovae (explosions of stars at least 8 times more massive than our Sun), neutron star mergers and supermassive black holes which produce heavier metals and large amounts of radiation. While the metals allow for more efficient cooling and thus smaller stars (Population II and eventually Population I stars), the radiative feedback processes provide heating that can mitigate gravitational collapse and thus star formation (Yoshii & Sabano, 1980; Bromm et al., 2001; Schneider et al., 2002; Smith & Sigurdsson, 2007; Tornatore et al., 2007; Xu et al., 2016; Tanaka & Hasegawa, 2021; Latif et al., 2022; Ventura et al., 2024, 2025).

However, Population III stars are yet to be observed, and thus, our knowledge of the first stars and galaxies is fairly limited. With telescopes like the *Hubble Space Telescope* (HST) and the *James Webb Space Telescope* (JWST), we are pushing the limits of the earliest galaxies observed (see for example, Finkelstein et al., 2015; Bouwens et al., 2015, 2021; Harikane et al., 2022; Bouwens et al., 2023a,b; Harikane et al., 2023; Leung et al., 2023; McLeod et al., 2024; Adams et al., 2024), and thus, a discovery of galaxies made up of Population III stars is imminent. Nevertheless, focussing on individual galaxies would invariably lead to a selection bias, i.e., the information gained from these observations would be biased towards the properties of the brightest and most

massive galaxies. The alternative to this is to rely on indirect searches, i.e., by focussing on the IGM around the first galaxies. The IGM during this period is believed to be mostly composed of neutral, atomic Hydrogen that gets ionized by UV photons produced by the first galaxies to reach the present-day state of a mostly ionized IGM. While neutral Hydrogen may usually not be expected to provide any radiation that can be used to observe these regions, the conditions so far described are conducive for the production of 21-cm line photons, which can thus become a tool to learn more about the formation of the first stars and galaxies, and broadly the Epoch of Reionization.

## 1.5 The 21-cm Signal

### 1.5.1 Theoretical background

The 21-cm line is a hyperfine transition line of atomic hydrogen in the ground state, arising due to the interaction between the electron and proton spins (Field, 1959; Hogan & Rees, 1979; Scott & Rees, 1990; Madau et al., 1997; Shaver et al., 1999; Tozzi et al., 2000; Ciardi & Madau, 2003; Zaroubi, 2013). When the electron and proton spins are parallel, it leads to the slightly higher energy triplet state, while when they are anti-parallel, it leads to the more stable singlet state. However, a transition from the triplet to a singlet state is considered as a *forbidden transition*. This is because, due to the minimal energy difference ( $\approx$  few  $\mu\text{eV}$ ), in most cases, Hydrogen atoms will flip between the two states through collisional excitation and de-excitation. In fact, the triplet state is found to have a lifetime of  $\approx 10^7$  years from decaying into the singlet state by spontaneous emission. However, the IGM at the EoR is believed to have three key properties, that make the 21-cm signal from this period observable. They are:

- **Low density:** the IGM typically has a density of  $\approx 1$  atom/ $\text{m}^3$ . This reduces the likelihood of collisional deexcitation.
- **Large amounts of neutral Hydrogen:** most of the matter in the early Universe was Hydrogen (Ewen & Purcell, 1951; Muller & Oort, 1951). Thus, despite low chances of spontaneous emission for an individual atom, as a whole, the chances are substantial.
- **Triple state boosting:** Most of the Hydrogen is boosted to the triplet state by Lyman- $\alpha$  radiation (Wouthuysen, 1952; Field, 1958) or the occasional collision.

As described above, the strength of the signal is governed by the population of neutral Hydrogen atoms in triplet and singlet states. The relation between these two states can be expressed in terms of a quantity we define as the spin temperature,  $T_S$ , as:

$$\frac{n_1}{n_0} = 3e^{(-T_*/T_S)} \quad (1.10)$$

where  $n_1$  and  $n_0$  are the number densities of electrons in the triplet and singlet states of the hyperfine level, respectively, and  $T_* = 0.00681$  K is the temperature corresponding to the 21-cm

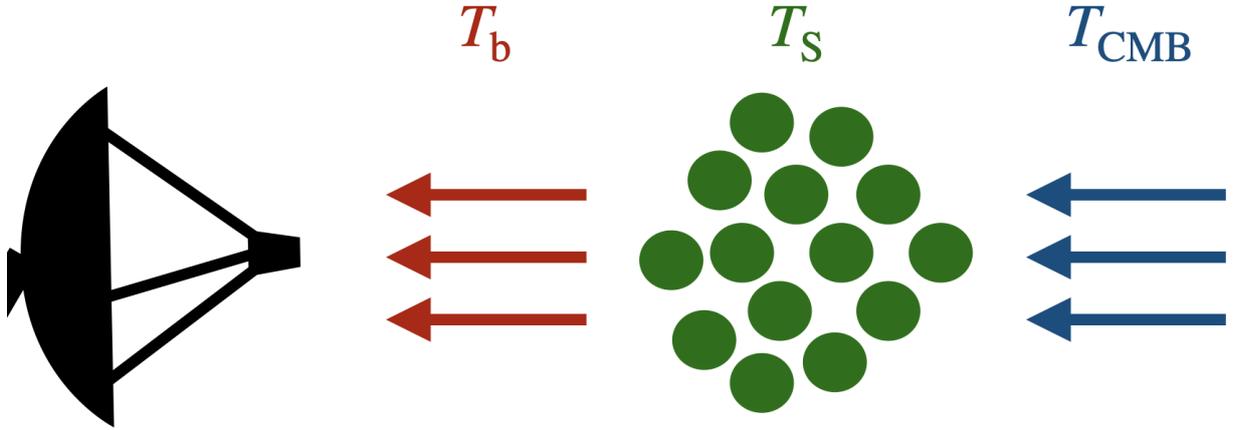


Figure 1.4: A representation that shows the concept of the brightness temperature ( $T_b$ ) arising from the spin temperature ( $T_S$ ) and the background CMB radiation ( $T_{\text{CMB}}$ ) which is observed by our telescopes. This figure has been adapted from [Zaroubi \(2013\)](#).

wavelength. The spin temperature is thus a representation of the ratio between the occupation number of the two hyperfine levels. This ratio establishes the intensity of the radiation emerging from a cloud of neutral hydrogen. In actual observations, one would have to take into account the level of background being transmitted through a given cloud as well as the amount of absorption and emission within the cloud. Thus, a careful analysis of radiative transfer is necessary. This is typically done in terms of the specific intensity or *brightness*,  $I_\nu = dI/d\nu$ , where  $\nu$  is the frequency ([Rybicki & Lightman, 1986](#)). Thus, the radiative transfer equation for a thermally emitting material at temperature  $T$  can be written in terms of the optical depth for absorption as:

$$\frac{dI_\nu}{d\tau_\nu} = -I_\nu + B_\nu T \quad (1.11)$$

where  $\tau_\nu$  is the optical depth for absorption through the cloud at a given frequency and  $B_\nu$  is the Planck function. Now, as in radio astronomy we deal with low frequencies, the Rayleigh-Jeans law is valid, and thus we can express the brightness  $I_\nu$  by its equivalent brightness temperature,  $T_b$ , which are related as:

$$T_b(\nu) = \frac{I_\nu c^2}{2\nu^2 k_B} \quad (1.12)$$

where  $k_B$  is the Boltzmann's constant and  $c$  is the speed of light. Now, as in our case, the background radiation is the CMB, Equation 1.11 reduces to:

$$\frac{dT_b}{d\tau_\nu} = -T_b + T_{\text{CMB}} \quad (1.13)$$

with  $T_{\text{CMB}}$  being the CMB temperature. Solving this for the temperature of the emergent radiation at frequency  $\nu$  yields:

$$T_b(\nu) = T_S(1 - e^{-\tau_\nu}) + T_{\text{CMB}}(\nu)e^{-\tau_\nu} \quad (1.14)$$

where  $T_S = T_b(0)$  is the brightness temperature in the absorbing cloud of neutral Hydrogen (see Figure 1.4). For a more detailed discussion, see Zaroubi (2013).

Often, the final reported quantity is the differential brightness temperature of the 21-cm signal, i.e.,  $\delta T_b = T_b - T_{\text{CMB}}$ . For any patch of the IGM, this is given by (Furlanetto et al., 2006):

$$\delta T_b = 27x_{\text{HI}}(1 + \delta_B) \left(1 - \frac{T_{\text{CMB}}}{T_S}\right) \times \left[ \left(\frac{\Omega_B h^2}{0.023}\right) \left(\frac{0.15}{\Omega_m h^2} \frac{1+z}{10}\right)^{1/2} \right] \text{mK} \quad (1.15)$$

where  $x_{\text{HI}}$  is the fraction of neutral hydrogen,  $\delta_B$  is the fractional overdensity of baryons,  $T_S$  is the spin temperature,  $T_{\text{CMB}}$  is the temperature of the CMB photons at that redshift,  $\Omega_m$  is the total matter density,  $\Omega_B$  is the baryon density,  $z$  is the redshift and  $h$  is the Hubble constant in units of  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . In this equation, the physical properties of the IGM affecting the large-scale fluctuations of  $\delta T_b$  are  $x_{\text{HI}}$ ,  $T_S$  and  $\delta_B$ .

### 1.5.2 Observational efforts

Several interferometric low-frequency radio telescopes have been designed, such as e.g. PAPER<sup>1</sup>, MWA<sup>2</sup>, LOFAR<sup>3</sup>, HERA<sup>4</sup>, NenuFAR<sup>5</sup> and the upcoming SKA<sup>6</sup>. However, the detection of the signal is a challenge, and a successful detection has not been possible yet. This is because of the fact that it is buried under foregrounds (synchrotron and free-free emissions from the Milky Way and other galaxies) that are several orders of magnitude stronger. Additionally, given the current interferometers used, the noise levels, even on observing deep fields for 1000 hours, are still of the same strength as the expected signal.

Nevertheless, over the years, the various interferometer experiments have been providing increasingly tighter upper limits on the 21-cm signal differential brightness temperature power spectrum as depicted in Figure 1.5. For example, HERA Collaboration et al. (2023) reported  $\Delta^2(k = 0.34 \text{ hMpc}^{-1}) \leq 457 \text{ mK}^2$  at  $z=7.9$  and  $\Delta^2(k = 0.36 \text{ hMpc}^{-1}) \leq 3496 \text{ mK}^2$  at  $z=10.4$  from 94 nights of observation, and Mertens et al. (2020, hereafter M20) reported  $\Delta^2(k = 0.075 \text{ hMpc}^{-1}) < 5329 \text{ mK}^2$  from 141 hours ( $\approx 10$  nights) of observation with LOFAR at  $z=9.1$ .

To address the issue of foregrounds, HERA Collaboration et al. (2023) uses the ‘‘foreground avoidance’’ technique (Kerrigan et al., 2018; Morales et al., 2019) by focusing on regions in Fourier space which are mostly foreground free, while the LOFAR EoR Key Science Project (KSP) team uses foreground modelling and removal, which allows the maximization of scales explored, as well as boosts the sensitivity up to an order of magnitude (Pober et al., 2014). This is

<sup>1</sup>Precision Array to Probe EoR, <http://eor.berkeley.edu>

<sup>2</sup>Murchison Widefield Array, <http://www.mwatelescope.org>

<sup>3</sup>Low-Frequency Array, <http://www.lofar.org>

<sup>4</sup>Hydrogen Epoch of Reionization Array, <https://reionization.org/>

<sup>5</sup>New Extension in Nançay Upgrading LOFAR, <https://nenufar.obs-nancay.fr/en>

<sup>6</sup>Square Kilometre Array, <https://www.skao.int/en>

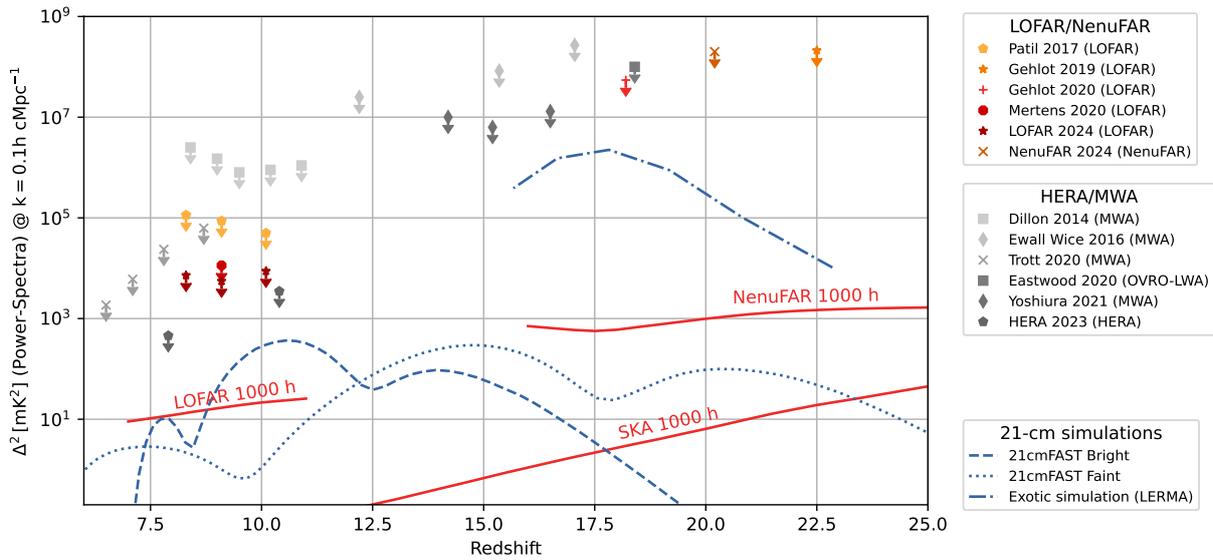


Figure 1.5: The predicted 21-cm power spectra from semi-numerical simulations (blue lines) versus observed upper limits (scattered points), along with the limits achievable with  $\approx 1000$  hours of observation with LOFAR, NenuFAR and SKA. Credit: Florent Mertens.

because no data is discarded, with a focus on modelling each component of the data as accurately as possible.

Dealing with noise is fairly straightforward in theory but is difficult to implement. Nevertheless, with improved interferometric experiments like the SKA and longer observation hours that are stacked by accounting for night-to-night variations, the contribution of noise can be reduced. The same is shown with the red lines in Figure 1.5, which shows the limits achievable with LOFAR, NenuFAR and SKA with 1000 hours of observation.

### 1.5.3 The LOFAR EoR Key Science Project

The LOFAR radio telescope comprises 24 core stations distributed within a 2 km diameter, 14 remote stations across the Netherlands, providing a maximum baseline length of approximately 100 km, and an increasing number of international stations across Europe as shown in Figure 1.7 (van Haarlem et al., 2013). The LOFAR EoR Key Science Project (KSP) team utilises observations from the High Band Antennas (HBA), operating at frequencies between 110 and 189 MHz, and targets two primary fields as shown in Figure 1.7: the North Celestial Pole (NCP) and the bright compact radio source 3C 196 (de Bruyn & LOFAR EoR Key Science Project Team, 2012). This data was recorded with a frequency range of 115 to 189 MHz, with a spectral resolution of 3.05 kHz (resulting in 64 channels per sub-band of 195.3 kHz) and a temporal resolution of 2 seconds.

So far, the LOFAR team has focussed on the NCP deep field, which has no major bright sources



Figure 1.6: The 24 core stations of LOFAR (left) and a map of the 14 remote stations across the Netherlands and the various international stations across Europe (right). Credit: ASTRON.

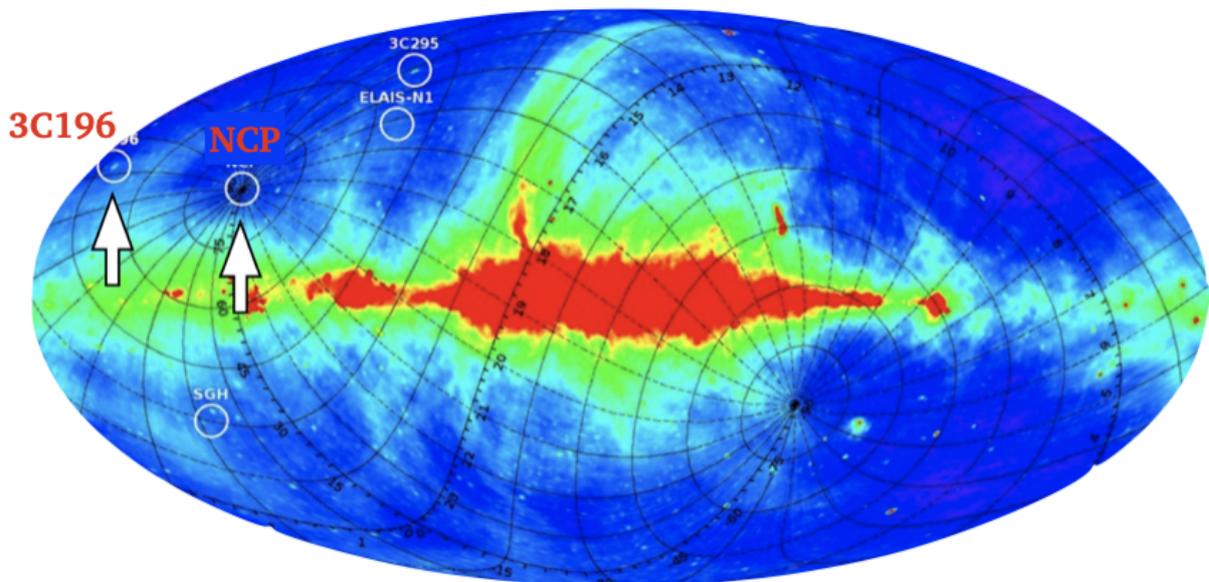


Figure 1.7: The all-sky map as observed by LOFAR, with various deep field candidates indicated. From these, the ones marked with white arrows are the finally selected ones of the North Celestial Pole (NCP) and the bright compact radio source 3C 196. Credit: Léon V. E. Koopmans.

and can be observed every night of the year. However, improvements in foreground modelling indicate that the foreground-subtracted 3C 196 field may have a lower noise contribution than the

NCP field. Thus, soon, the LOFAR team will be able to provide upper limits across two different deep fields, which can allow for cross-correlation of inferred properties.

The most stringent constraints obtained with LOFAR data were presented in M20, where Gaussian Process Regression (GPR, as described in Mertens et al., 2018; Gehlot et al., 2019; Hothi et al., 2021, and in this thesis in Section 2.2.2 of Chapter 2) was used for hyperparameter optimization with different Matern class functions (Eq. 2.6 in Chapter 2, from Stein, 1999) chosen as covariance kernels for modelling different components of the observed data and then recovering the fitted datacube (a deep field with the third side represented in terms of frequency). However, Kern & Liu (2021) pointed out some issues with this approach. Primarily, they found that given the choice of normalization and bias correction in the power-spectra estimation used in M20, misestimation of the covariance kernel for the EoR signal could lead to significant signal loss. This can have severe ramifications on the astrophysical interpretations of the estimated 21-cm signal power spectrum. Further, they show that alternative choices for normalization and weighting schemes could reduce the dependence on the choice of covariance priors, thus reducing its impact on the estimation of the 21-cm signal. However, in this thesis, I focus on improving the covariance prior, while in the future, I plan to explore other normalization and bias correction schemes to further upgrade the overall analysis pipeline.

To improve the covariance kernel, I refer to Mertens et al. (2024). They propose a Machine Learning (ML)-based approach to GPR, where the covariance kernel for the 21-cm signal is obtained by implementing a Variational Auto-Encoder (VAE)-based algorithm that learns from simulations. The specifics of the VAE-based algorithm and its implementation are discussed in detail in Section 2.2.3 of Chapter 2. The results obtained with this approach can then directly be compared against runs of the same simulation code to constrain the physical parameters used. As the covariance kernel is trained over a range of physical parameters, this would significantly reduce the chances of misestimation. Thus, it can be reliably used to derive astrophysical parameters necessary for the same simulation code to generate similar power spectra.

In particular, in this thesis, I use GRIZZLY (Ghara et al., 2015, 2018, 2020) for generating the training, test, and validation datasets. This code has been employed previously (see Ghara et al., 2020) to constrain astrophysical parameters based on the results obtained in M20. As this combines  $N$ -body simulations with 1D radiative transfer, it is more physically precise than semi-numerical algorithms, while not being as computationally expensive as codes that use 3D radiative transfer (while still performing reasonably well, as shown in Ghara et al., 2018). I tested the performance of this ML-based kernel versus covariance kernels used with GPR in previous work in Chapter 2, which is based on Acharya et al. (2024b, hereafter A24). Then, in Chapter 3, which is based on Acharya et al. (2024d), I apply this VAE-based covariance kernel to the same pipeline and same data as had been analysed by M20 and compare their results.

#### 1.5.4 Simulations for the 21-cm Signal

The observed upper limits of the 21-cm signal power spectrum have allowed us to rule out some extreme astrophysical models through comparisons with simulations (e.g., Ghara et al., 2020;

Mondal et al., 2020; Greig et al., 2021a,b; Abdurashidova et al., 2022). New methods that utilize multi-redshift power spectrum observations (Ghara et al., 2024; Choudhury et al., 2024), wavelet statistics (Hothi et al., 2024), and simulation-based inference (Saxena et al., 2023; Greig et al., 2024) have also been developed to improve the understanding of parameters governing the properties of the IGM.

To study the formation of the first galaxies responsible for driving the ionization of Hydrogen in the IGM, many hydrodynamical and/or radiative transfer simulations with different prescriptions have been utilised, such as CROC (Gnedin, 2014; Gnedin & Kaurov, 2014; Esmerian & Gnedin, 2021, 2022, 2024), FIRE (Ma et al., 2018), TECHNICOLOR DAWN (Finlator et al., 2018), SPHINX (Rosdahl et al., 2018), CRASH simulations (Eide et al., 2018, 2020; Ma et al., 2021; Kostyuk et al., 2023; Basu et al., 2024), C<sup>2</sup>RAY simulations (Mellema et al., 2006; Dixon et al., 2016; Giri & Mellema, 2021; Hirling et al., 2024; Giri et al., 2024), CoDA I,II, III (Ocvirk et al., 2016, 2020; Lewis et al., 2022), ASTRID (Bird et al., 2022), COLDSIM (Maio et al., 2022; Maio & Viel, 2023; Casavecchia et al., 2024), THESAN (see Figure 1.8 for example results; Kannan et al., 2022; Garaldi et al., 2022; Smith et al., 2022; Garaldi et al., 2024), and SPICE (Bhagwat et al., 2024). Although these simulations are ideal tools to investigate the IGM and galactic properties, they are computationally expensive and thus limited in box size and/or in the number of simulations that can be run. Thus, they cannot be used for building a wide range of models for comparison with the 21-cm power spectrum or their upper limits.

To this aim, semi-numerical approaches such as SIMFAST21 (Santos et al., 2010a,b), BEARS (Thomas et al., 2009), 21CMFAST (Mesinger & Furlanetto, 2007; Mesinger et al., 2011), GRIZZLY (Ghara et al., 2015, 2018), REIONYUGA (Mondal et al., 2017), ARTIST (Molaro et al., 2019), AMBER (Trac et al., 2022) and BEORN (Schaeffer et al., 2023) are typically employed. To incorporate a more realistic modeling of galactic properties into these methods, codes such as ASTRAEUS (Hutter et al., 2021, 2024), MERAXES (Mutch et al., 2016; Balu et al., 2023) and POLAR (Ma et al., 2023, hereafter M23) have been developed.

These are especially important, as apart from building a wide range of models, it is also important to have large-volume boxes. This is because the thermal noise is proportional to the wave-modes (Koopmans et al., 2015), and thus the best results from upcoming surveys with LOFAR, HERA, MWA and, subsequently, SKA are expected at the lowest  $k$  values, i.e.,  $0.15 \leq k/(h \text{ cMpc}^{-1}) \leq 1.5$ , after sufficient mitigation of foreground contamination. This corresponds to large physical scales, and indeed, e.g. Iliev et al. (2014) found that simulations with box sizes greater than  $100 h^{-1} \text{ cMpc}$  are necessary to accurately model the evolution and distribution of ionized regions. Kaur et al. (2020), instead, found that even larger boxes, i.e.  $\gtrsim 175 h^{-1} \text{ cMpc}$ , are necessary. However, a simulation with a box size corresponding to  $k \sim 0.15 h \text{ cMpc}^{-1}$  is still limited by sample variance. The straightforward solution to this issue is to employ larger boxes in order to increase the sampling of the long wave modes observationally relevant. For example, in Ghara et al. (2020), a side length of  $500 h^{-1} \text{ cMpc}$  has been simulated, while the semi-numerical approach of Mesinger et al. (2016) and Greig et al. (2022) has modeled Gpc scales. Such large boxes also allow us to study the topology of large ionized regions ( $\gtrsim 10 \text{ cMpc}$ ), which is especially important to understand the role of quasars. However, even these approaches have issues, as extremely large

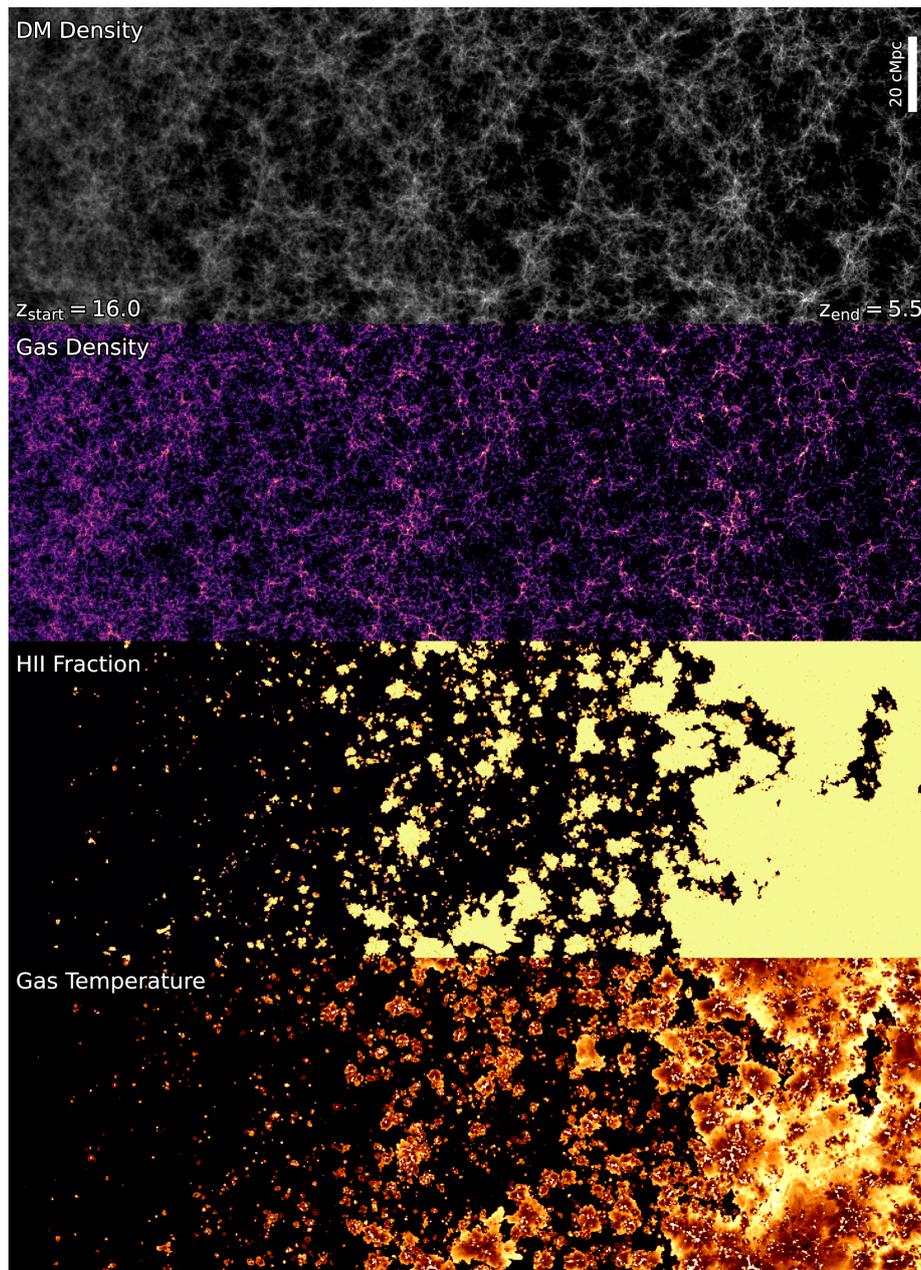


Figure 1.8: Evolution of various simulated properties in the main THESAN run (Kannan et al., 2022; Garaldi et al., 2022; Smith et al., 2022; Garaldi et al., 2024) versus time going from the start of the simulation run (left) to its end (right). The top panel shows the collapse of dark matter forming the cosmic web structure, and the second panel shows the gas collapsing to create galaxies. These produce ionizing photons, which ionize the atomic Hydrogen, and the evolution of the fraction of ionized Hydrogen is shown in the third panel. The final panel shows the evolution of the temperature of the gas.

computational resources for running the simulations and storing the generated data are required, and running large number of such simulations would only be computationally affordable with the semi-numerical approaches. While they can be appropriate for some analyses, they also have their own limitations (see for e.g., [Choudhury & Paranjape, 2018](#), for details).

Additionally, there are limitations on the extent of physics that can be explored with such simulations. For example, mini halos ( $\lesssim 10^8 M_\odot$ , see [Iliev et al., 2005](#); [Haiman et al., 2001](#)), dwarf galaxies ([Wu & Kravtsov, 2024](#)), or Lyman Limit Systems ([Georgiev et al., 2024](#); [Giri et al., 2024](#)) are expected to play a significant role in reionization, but their inclusion would either require a high enough mass resolution to be able to capture them in large boxes, or running significantly smaller boxes. The former exacerbates the computational requirements, while the latter limits the wave modes that can be sampled.

Various approaches have been proposed to tackle these problems. One approach is to run simulations with different box sizes and combine their results. A recent example of this is an analysis of galaxy populations at  $z \geq 8$  in [Kannan et al. \(2023\)](#), which, however, does not provide a truly combined picture, as the different boxes are manually re-scaled to account for resolution differences without actual convergence between simulations. There have also been suggestions to use techniques like deep learning to increase the resolution of large box simulations (see [Kodi Ramanah et al., 2020](#); [Li et al., 2021](#); [Zhang et al., 2024](#), for applications to dark matter only  $N$ -body simulations), however, these haven't been carefully tested for EoR redshifts.

Another approach is to run computationally cheaper boxes but include the properties of the more expensive, full-physics ones. Recently, [Sinigaglia et al. \(2021\)](#) developed a method to map baryonic properties of the IGM onto DM-only simulations, but its applicability to the reionization epoch is unclear.

In this thesis, I explore two methods to overcome these issues of box size and physical complexity: the Fixed & Paired (F&P) method in Chapter 4, and building a suite of viable models of the POLAR simulations in Chapter 5, which acts as a compromise between physical accuracy and computational expense. I use the F&P method for the observables of the POLAR simulations as well to provide the best possible mock observables from galactic to IGM scales.

#### 1.5.4.1 The Fixed & Paired Method

As smaller boxes can be run with all physical processes explicitly calculated, one way to deal with the issues highlighted above is to simulate the full physics of reionization in small boxes and then augment them with techniques that compensate for the limited volume. One way to do this is the Fixed and Paired (F&P) approach described in [Angulo & Pontzen \(2016\)](#) and [Pontzen et al. \(2016\)](#). It has been shown to substantially reduce the cosmic variance in the matter power spectrum (see [Maion et al., 2022](#); [Klypin et al., 2020](#); [Villaescusa-Navarro et al., 2018](#), for examples) with respect to the traditional approach for the same box size, bypassing the need to run a large number of smaller boxes. However, this reduction is substantial only for statistical properties (e.g., the stellar mass function) and not for local ones like the gas distribution of individual galaxies (check

Villaescusa-Navarro et al., 2018, for more details). The 21-cm signal is influenced by both the large-scale distribution of neutral hydrogen, and the local properties of the ionized regions around galaxies. Exploring the extent of improvement (if any) on the effective volume for estimating the summary statistics of the 21-cm signal using the F&P approach can be helpful in minimising the required computational needs. This idea was explored in Giri et al. (2023) by running a large number of realizations of the F&P approach using 21cmFAST (Mesinger & Furlanetto, 2007; Greig & Mesinger, 2015). By comparing them with randomly generated simulations, they found that F&P boxes could obtain the same precision in the 21-cm signal power spectrum as traditional boxes twice their size, allowing at least a factor of 4 reduction in computing costs. However, 21cmFAST does not take into account baryonic hydrodynamics, thus implicitly assuming that baryons track dark matter. It also does not include a proper implementation of radiative transfer (RT) or galaxy properties, and thus becomes unreliable at scales below 1 cMpc. Hence, it cannot provide an accurate picture of galaxy-scale effects on the 21-cm signal.

To increase confidence in the applicability of the F&P method for improving EoR 21-cm signal studies, it is necessary to use a more realistic framework, i.e., one that includes baryonic hydrodynamics, radiative transfer, and models galaxy properties more accurately. For this, I employ a setup similar to the THESAN simulations (Kannan et al., 2022; Garaldi et al., 2022; Smith et al., 2022; Garaldi, 2023), detailed in Chapter 4, which is adapted from Acharya et al. (2024c).

#### 1.5.4.2 The POLAR simulations

As a part of the LOFAR EoR KSP team, I have worked on designing POLAR, a semi-numeric approach which strikes a balance between speed and complexity of relevant physical processes, by post-processing  $N$ -body Dark Matter (DM) simulations with Semi-Analytic Models (SAMs) of galaxy formation and evolution, and subsequently applying the 1D radiative transfer (RT) code GRIZZLY (Ghara et al., 2015, 2018), an updated version of BEARS (Thomas et al., 2009). This allows a fast modeling of the 21-cm signal power spectrum while at the same time accounting for important galactic properties and RT effects. Consequently, observations of the 21-cm line can also be combined with data obtained in other frequency bands (e.g., from JWST) to constrain IGM and galactic properties jointly.

In Chapter 5, I investigate the impact of varying cosmological parameters on the 21-cm signal at  $z \approx 10.11, 9.16, \text{ and } 8.3$ , i.e., the three redshifts targeted by LOFAR (Mertens et al., 2025, ; submitted). I focus on those parameters whose low and high redshift measurements are in tension, i.e., the Hubble parameter  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$  and the matter clustering amplitude  $\sigma_8$ . My fiducial model is the one of Planck Collaboration et al. (2020), with  $h = 0.6766$  and  $\sigma_8 = 0.8102$ , while a higher  $h$  value is adopted to explore the low redshift measurement of  $h = 0.733$  from studies of Cepheid variables in the host galaxies of 42 Type Ia supernovae (Riess et al., 2022). Similarly, I adopt a lower  $\sigma_8$  value of 0.702 from the low redshift measurement of anisotropic galaxy clustering measurement analysis (Tröster et al., 2020). Lastly, I also adopt a  $\sigma_8 = 0.88$  from recent eROSITA results (Ghirardini et al., 2024) to consider both extremes of  $\sigma_8$  measurements. Motivated by results obtained in the context of the 21-cm signal shown by

Giri et al. (2023) and Acharya et al. (2024c, ; based on work shown in Chapter 4), I additionally implement the Fixed & Paired (F&P) approach (Angulo & Pontzen, 2016) to boost effective volumes of the simulations and suppress cosmic variance.

## 1.6 Outline of this thesis

The work in this thesis is divided into three broad parts, which are the following:

1. In Chapter 2, I describe the ML-based improvement to GPR, and its usage for building a “template” for the 21-cm signal power spectrum. For this, I train on the GRIZZLY simulations to build covariance kernels that I test across different redshifts, noise levels, and different simulations for the injected mock 21-cm signal in the simulated datacube (from GRIZZLY, CRASH and 21cmFAST).
2. In Chapter 3, I take the work in Chapter 2 a step further and apply it to real data gathered by the LOFAR telescope.
3. In Chapter 4, I explore the improvements in the effective volume of simulations used for studying the EoR with radiation magneto-hydrodynamical (RMHD) simulations by adopting the F&P method, particularly by modelling the power spectrum and bispectrum.
4. In Chapter 5, I introduce the POLAR simulations that aim to broaden the range of parameters that can be varied for building a suite of possible models of the EoR. These simulations will be useful not just for building broader training sets for Machine Learning but also for incorporating more viable models by inferring all possible combinations of model parameters that produce results that match with observations.
5. Finally, in Chapter 6, I summarise the results of this thesis and discuss future work that can be carried out to lead to an eventual detection of the 21-cm signal at the EoR and develop a deeper understanding of the physical processes that govern this period.

# Chapter 2

## Machine Learning for extracting the 21-cm Signal: Testing

*There is a lot about what is going on here that I don't understand.  
But I am participating anyway.*

---

- Martha Wells

### 2.1 Overview

This chapter covers the implementation of the Variational Auto-Encoder (VAE)-based algorithm of Machine Learning to build a covariance kernel for the 21-cm signal as proposed in [Mertens et al. \(2024\)](#) and tests its performance against analytical covariance kernels for recovering various simulated signals with different noise levels and redshifts from mock LOFAR data using Gaussian Process Regression (GPR). The work shown here was published in [Acharya et al. \(2024b\)](#), whose notation has been adapted to be consistent with the rest of the thesis.

In Section 2.2, I discuss a range of simulations used to generate mock 21-cm datasets, as well as introduce the ML-trained 21-cm kernel. I also provide a short introduction to GPR. I report the results and comparisons between kernel performances in Section 2.3. Finally, I discuss the role of the excess noise component found in [M20](#) and the overall performance of the ML-based kernel in Section 2.4.

## 2.2 Methodology

In this section, I introduce the pipeline used to implement GPR to recover the 21-cm signal from mock datasets, comparing the performance of an ML-based kernel versus kernels used in [M20](#).

### 2.2.1 Simulations of the 21-sm signal

The 21-cm differential brightness temperature relative to the CMB for any patch of the IGM is given by (see [Furlanetto et al., 2006](#)):

$$\delta T_b = 27 x_{\text{HI}} (1 + \delta_B) \left( 1 - \frac{T_{\text{CMB}}}{T_S} \right) \times \left[ \left( \frac{\Omega_B h^2}{0.023} \right) \left( \frac{0.15}{\Omega_m h^2} \frac{1+z}{10} \right)^{1/2} \right] \text{mK} \quad (2.1)$$

where  $x_{\text{HI}}$  is the fraction of neutral hydrogen,  $\delta_B$  is the fractional overdensity of baryons,  $T_S$  is the spin temperature,  $T_{\text{CMB}}$  is the temperature of the CMB photons at that redshift,  $\Omega_m$  is the total matter density,  $\Omega_B$  is the baryon density,  $z$  is the redshift and  $h$  is the Hubble constant in units of  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . In this equation, the parameters affecting the large-scale fluctuations of  $\delta T_b$  are  $x_{\text{HI}}$ ,  $T_S$  and  $\delta_B$ .

I consider a variety of simulations to generate mock 21-cm differential brightness temperature maps as discussed below. In [Section 2.2.1.1](#), I employ maps generated using GRIZZLY, where I focus on variations tied to fluctuations in  $x_{\text{HI}}$  and  $T_S$ , while assuming that the fluctuations due to  $\delta_B$  are small and can thus be ignored. In [Section 2.2.1.2](#), I do not make this assumption and employ maps generated using the reionisation simulation code CRASH. In [Appendix A](#) I also consider the additional case of using 21cmFAST ([Mesinger & Furlanetto, 2007](#); [Greig & Mesinger, 2015](#)) to generate the 21-cm differential brightness temperature maps.

#### 2.2.1.1 GRIZZLY simulations

GRIZZLY ([Ghara et al., 2015, 2018, 2020](#)) employs a 1D radiative transfer scheme in combination with cosmological density fields and halo catalogues obtained from an N-body simulation to produce brightness temperature maps of the 21-cm signal at different redshifts for a given source model. The algorithm has been shown to reproduce results similar to those obtained with 3D radiative transfer schemes with the same N-body simulation, while being at least  $10^5$  times faster ([Ghara et al., 2018](#)). Because of this, I can run a large number of GRIZZLY simulations without the process being too computationally expensive. Furthermore, it has a wide range of physical parameters that can be varied, thus allowing me to explore the role of different physical processes in generating different models of the 21-cm signal. The density fields, velocity fields and the halo lists used in this work are obtained from the same N-body simulation ( $500 h^{-1} \text{ cMpc}$  box length,  $6912^3$  particles, with a mass resolution of  $4.05 \times 10^7 M_\odot$ ) which was used in [Ghara et al. \(2020\)](#). In this study, I consider two major GRIZZLY models presented in [Sections 3.1 and 3.2 of Ghara et al. \(2020\)](#). Similar to their implementation, I use four physical parameters to

model the sources: the ionization efficiency ( $\zeta$ ), the minimum mass of the UV emitting halos ( $M_{\min}$ ), the minimum mass of the X-ray emitting halos ( $M_{\min_X}$ ) and the X-ray heating efficiency ( $f_X$ ). The emission rate of ionizing photons and X-rays per unit stellar mass from a halo are  $\zeta \times 2.85 \times 10^{45} \text{ s}^{-1} M_{\odot}^{-1}$  and  $f_X \times 10^{42} \text{ s}^{-1} M_{\odot}^{-1}$ , respectively. Further, the X-ray spectral index  $\alpha$  is fixed at 1.2, as done in [Ghara et al. \(2020\)](#). Lastly, they note that all other IGM properties can be derived from these parameters, and thus using just these to define the simulation is sufficient. The properties of the two models adopted are listed below:

- **$x_{\text{HI}}$  fluctuation dominated model:** here, I assume a uniform Ly $\alpha$  background strong enough to allow the spin temperature  $T_S$  to be fully coupled to the gas temperature  $T_K$ . Further, I adopt the following parameters:  $\zeta = 7.0$ ,  $M_{\min} = M_{\min_X} = 10^9 M_{\odot}$  and  $f_X = 100$ , which makes the gas temperature (and in turn,  $T_S$ ) significantly high compared to  $T_{\text{CMB}}$  due to strong X-ray heating. This assumption of  $T_S \gg T_{\text{CMB}}$  ensures that  $\delta T_b$  becomes insensitive to the  $(1 - \frac{T_{\text{CMB}}}{T_S})$  term from Equation 2.1. Thus, all variability of  $\delta T_b$  is tied to the fluctuation of the neutral hydrogen fraction  $x_{\text{HI}}$ .
- **$T_S$  fluctuation dominated model:** in this case, while I continue to have the assumption of a strong, uniform Ly $\alpha$  background to ensure coupling of  $T_S$  and  $T_K$ , I change my parameters to relax the condition of  $T_S \gg T_{\text{CMB}}$ . This is done by reducing the X-ray heating and ionization efficiency. Thus, I adopt the following parameters:  $\zeta = 3.0$ ,  $M_{\min} = 10^9 M_{\odot}$ ,  $M_{\min_X} = 10^{10} M_{\odot}$  and  $f_X = 1$ . This allows for greater variability tied to  $T_S$ , with regions of partial reionization and heating forming in the IGM.

### 2.2.1.2 CRASH simulations

As a reference, I also use the simulations of reionization described in [Eide et al. \(2018, 2020\)](#) and [Ma et al. \(2021\)](#). These are obtained by post-processing the large-scale, high-resolution hydrodynamical simulation Massive Black-II ([Khandai et al. 2015](#); box length  $100 h^{-1} \text{cMpc}$ ,  $2 \times 1792^3$  gas and dark matter particles, corresponding to a resolution of  $2.2 \times 10^6 h^{-1} M_{\odot}$  and  $1.1 \times 10^7 h^{-1} M_{\odot}$  respectively) with the multi-frequency 3D radiative transfer code CRASH ([Ciardi et al., 2001](#); [Maselli et al., 2009](#); [Graziani et al., 2013, 2018](#); [Glatzle et al., 2019](#)). Here, I make use of the ‘‘Stars’’ simulation run (which includes only stellar type sources) to generate the mock 21-cm signal data at  $z=9.18$ . I refer the reader to the original papers for more detailed information on the simulations.

## 2.2.2 Gaussian process regression

Gaussian Process Regression (GPR, [Rasmussen & Williams 2006](#); [Aigrain & Foreman-Mackey 2023](#)) can be used to model a noisy observation  $\mathbf{y} = f(\mathbf{x}) + \epsilon$ , with  $\epsilon$  Gaussian noise having variance  $\sigma_{\text{noise}}^2$ . This is achieved by modelling the Gaussian Process as a joint probability distribution for  $\mathbf{y} = \{y_i\}_{i=1, \dots, N}$ , as  $\mathbf{f}(\mathbf{x})$ , which is fully defined by its mean vector ( $\mathbf{m}$ ) and covariance matrix ( $\mathbf{K}$ ,

also called covariance “kernel”) as:

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{N}(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x})) . \quad (2.2)$$

for a set of points  $\mathbf{x}$  (independent parameters). Here, the covariance matrix  $\mathbf{K}$  gives the covariance between the function values at any two points and can be written as  $K_{ij} = \kappa(x_i, x_j, \phi) + \delta_{ij}\sigma_i^2$ , where  $\kappa(x_i, x_j, \phi)$  can be optimised by the choice of hyperparameters represented by  $\phi$ , and  $\delta_{ij}$  is the Kronecker-delta function.

When applying it to radio data to extract the 21-cm signal, I split this function into a foreground component,  $\mathbf{f}_{\text{fg}}$ , and the 21-cm signal,  $\mathbf{f}_{21}$ , giving:

$$\mathbf{y} = \mathbf{f}_{\text{sky}}(\mathbf{x}) + \mathbf{f}_{21}(\mathbf{x}) + \epsilon \quad (2.3)$$

where  $y$  is the observed data and  $x$  is the frequency. Next, following M20, I further split the foreground component into the intrinsic sky emission component ( $\mathbf{f}_{\text{sky}}$ ), which comes from the confusion-limited extragalactic sources and from the Milky Way, and the mode-mixing contaminants component  $\mathbf{f}_{\text{mix}}$ , which has contributions from the instrument chromaticity and calibration errors. Beyond the foreground, I also model the noise (represented by  $\epsilon$  in Equation 2.3) using estimates of the noise variance for  $\approx 10$  nights of observation from M20. In addition to this, M20 found a significant spectrally correlated residual, and thus I inject this “excess noise” component ( $\mathbf{f}_{\text{ex}}$ ) into my model as well. This gives an updated version of Equation 2.3:

$$\mathbf{y} = \mathbf{f}_{\text{sky}}(\mathbf{x}) + \mathbf{f}_{\text{mix}}(\mathbf{x}) + \mathbf{f}_{\text{ex}}(\mathbf{x}) + \mathbf{f}_{21}(\mathbf{x}) + \epsilon . \quad (2.4)$$

For the sake of simplicity and utilising the additive property of matrices, I can reduce Equation 2.4 to  $\mathbf{y} = \mathbf{f}(\mathbf{x}) + \epsilon$ , and represent  $\mathbf{f}(\mathbf{x})$  with its corresponding covariance kernel  $K$  (from Equation 2.2) as:

$$\mathbf{K} = \mathbf{K}_{\text{sky}} + \mathbf{K}_{\text{mix}} + \mathbf{K}_{\text{noise}} + \mathbf{K}_{\text{ex}} + \mathbf{K}_{21} . \quad (2.5)$$

M20 modelled each of these kernels using the best possible fit Matern-class functions (Eq. 2.6; Stein, 1999):

$$k_{\text{Matern}}(r) = \sigma^2 \frac{2^{1-\eta}}{\Gamma(\eta)} \left( \frac{\sqrt{2\eta}r}{l} \right)^\eta \kappa_\eta \left( \frac{\sqrt{2\eta}r}{l} \right) . \quad (2.6)$$

Note that in the Matern-class function,  $r$  is the absolute difference between the frequencies of two sub-bands,  $\kappa_\eta$  is the modified Bessel function of the second kind and  $\Gamma$  is the Gamma-function.

M20 obtained the best possible fit Matern-class function by taking different values of the hyperparameter  $\eta$ , maximising the marginal likelihood (also known as the evidence) and obtaining estimates for the coherence scale hyperparameter  $l$  and the variance  $\sigma^2$ . Then, for each kernel, they chose the  $\eta$  that led to the highest evidence by calculating the analytical integral over  $\mathbf{f}$  which is the log-marginal-likelihood (LML, see Section 2.3 in Mertens et al., 2018) and choosing the kernel that maximises its value. For calculating the hyperparameters (listed in the second column of Table 2.1), M20 used a gradient-descent based optimization algorithm for maximising the LML.

### 2.2.3 Machine learning trained 21-cm kernel

The limitations of GPR as pointed out by [Kern & Liu \(2021\)](#) mainly boil down to the choice of covariance kernel for the 21-cm signal. While the choice of hyperparameters allows a variety of functions to be accessed, the same function might not work equally well across the  $k$ -space. Having a function obtained by employing ML trained on power spectra of simulations where the sources of ionization are modelled using parameters that sample a wide range of values, allows greater flexibility, and reduces chances of misestimation. Further, it allows for a direct comparison with physical quantities, as I can reliably derive the source parameters necessary to generate simulations that produce the power spectra estimated by the ML-based kernel.

To achieve this, [Mertens et al. \(2024\)](#) uses a Variational Auto-Encoder (VAE, [Kingma & Welling, 2013, 2019](#)) algorithm. Simply put, an Auto-Encoder (AE, [Goodfellow et al., 2016](#)) is an unsupervised neural network which compresses data by reducing the number of independent parameters used to describe it into what is referred to as a “latent space” of hyperparameters. Thus, it is primarily used for data compression by filtering out independent parameters that are deemed to be unnecessary because they only slightly affect data recovery. This is a two steps process, where the first step of reducing the number of independent parameters into the latent space is called *encoding*, while the step of recovering the data given the latent space parameters is called *decoding*. Instead of taking an input of just a set of parameters  $a_1, \dots, a_n$ , a VAE ([Pinheiro Cinelli et al., 2021](#)) uses probability distributions of each parameter, thus allowing to interpolate in the latent space, and to generate a large range of new samples of reconstructed data (in my case 21-cm signal models), which are not limited by the data that the encoder was trained on. However, this also means that the VAE encoder has an inherently larger error than an AE encoder.

Thus, reconstructing the training data using the decoder with the latent space generated by the VAE encoder as input would not be an exact match to the original training data. However, the VAE is designed to optimize a trade-off between reconstruction accuracy and the fidelity of the latent space representation, by minimising the KL (Kullback-Leibler) divergence loss ([Kullback & Leibler, 1951](#)), which is a measure of the divergence between the distribution of reconstructed data and the training data. So while the reconstructed data and training data might not be an exact match, if their overall distribution is similar (i.e., divergence is minimised), the training is considered successful. Thus, in the training of a VAE algorithm, the following sources of error exist:

1. *Encoder*: the error due to sampling from a distribution for each independent parameter to build a latent space. Sampling from a distribution is expected to be noisier than choosing point values.
2. *Decoder*: the error due to deriving the independent parameters, given some value of the latent space parameters. As this does not involve sampling from distributions, the contribution to the total error is expected to be smaller.

[Mertens et al. \(2024\)](#) shows that GPR can be used to estimate the values of the latent space

parameters, after which the decoder of the VAE kernel can be used to estimate the independent parameters, and in turn to obtain the recovered 21-cm signal. In this case I would then use only the decoder part of the VAE algorithm. However, as minimising KL divergence loss requires to train the encoder and decoder together, I proceed as follows.

I start by using two hyperparameters,  $x_1$  and  $x_2$ , and an associated variance, as the latent space parameters. I train the VAE algorithm on a dataset of  $\approx 1500$  simulations (the training set), and validate it against an independent dataset of  $\approx 150$  simulations (the validation set). I refer to this as the *VAE kernel*. The training and validation datasets are generated by running GRIZZLY simulations with a wide range of values for the parameters introduced in Section 2.2.1.1. I sample them randomly at the same redshift as the targeted 21-cm signal ( $z = 9.16, 8.30$  and  $10.11$ ) and in the following ranges:  $\zeta = [-3, 3]$ ,  $M_{\min} = [9, 12]$ ,  $M_{\min_x} = [9, 12]$  and  $f_X = [-3, 2]$  in the log space. I choose these ranges to be significantly broader than necessary (i.e., the performance of the VAE kernel does not show any appreciable difference even if they are reduced by multiple orders of magnitude), and also note that the performance of the kernel is not impacted if I fix one of the parameters to a specific value and vary the remaining three. This ensures that the sampling range does not induce any major bias. Next, to train the VAE I use 2000 iterations, as I find that the KL divergence loss and the reconstruction loss for both the training and validation datasets stabilise after  $\approx 500$  iterations. The reconstruction loss is defined as the total error made when using the encoder to obtain specific values of the latent space parameters, and then employing the decoder to retrieve the data from those parameters. I then evaluate the ratio between the output and input data as a function of the wave number  $k$ , and finally calculate the median ratio, which is  $\approx 1$  across all  $k$ -bins. However, as discussed earlier, the 68% confidence interval is significant, being  $\approx 10\%$  for  $k = [0.06, 0.12] \text{ hMpc}^{-1}$ ,  $\approx 35\%$  for  $k = [0.12, 0.43] \text{ hMpc}^{-1}$  and  $\approx 27\%$  for  $k = [0.43, 1.11] \text{ hMpc}^{-1}$  at  $z = 8.30, 9.16$  and  $10.11$ .

I note a similar reconstruction error for training sets of sizes ranging from 1000 to 5000 simulations (with the validation set scaling as  $\sim 10\%$  of the training set in size), while the errors become worse when using less than 1000 simulations. I also tried to use three hyperparameters, but saw no significant improvement in the recovery error. I thus choose to use two hyperparameters to avoid overfitting.

A high reconstruction error was to be expected, as it includes also the error due to the encoding process. While this is not required for my purpose, I do need to evaluate the exact contribution from the decoder. For this, I create a testing set of  $\approx 150$  simulations, and ensure that it also includes “extreme” models from GRIZZLY (which I define as cases where the power spectrum differs by at least an order of magnitude from the mean of the power spectra from the training and validation datasets), along with some injected stochastic noise. I then apply GPR for hyperparameter optimization (using MCMC, as discussed in Section 2.2.5) to estimate  $x_1, x_2$  and the associated variance. This is then used with the decoder to obtain the recovered signal. The median ratio of the recovered and input data across all  $k$ -bins is again  $\approx 1$ . But now I obtain a 68% confidence interval of  $\approx 0.5\%$  for  $k = [0.06, 0.12] \text{ hMpc}^{-1}$ ,  $\approx 1\%$  for  $k = [0.12, 0.43] \text{ hMpc}^{-1}$  and  $\approx 3\%$  for  $k = [0.43, 1.11] \text{ hMpc}^{-1}$  at  $z = 8.28, 9.16$  and  $10.11$ , respectively. By comparing to the total error estimated above, I note that even where decoder error is high, i.e., for the highest

$k$ -bins, it is still a minor contributor. Therefore, as the overall error is  $\lesssim 5\%$ , I accept the VAE kernel as a reliable ML-based kernel for the 21-cm signal, and proceed to use it with GPR for signal recovery.

### 2.2.4 Generating mock datasets

To build a mock dataset in the gridded  $U - V$  cube domain of radio observations, I derive a full dataset  $y$  by adding each term on the right hand side of Equation 2.4. For this, I adopt the values of  $\eta$ ,  $l$  and  $\sigma^2$  for  $\mathbf{K}_{\text{sky}}$ ,  $\mathbf{K}_{\text{mix}}$  and  $\mathbf{K}_{\text{ex}}$  given in the second column of Table 2.1 to generate their corresponding  $\mathbf{f}_{\text{sky}}$ ,  $\mathbf{f}_{\text{mix}}$  and  $\mathbf{f}_{\text{ex}}$ . These values are obtained from the results of M20, where they used  $\eta_{\text{sky}} = +\infty$  for the intrinsic sky,  $\eta_{\text{mix}} = 3/2$  for the mode-mixing contaminants, and  $\eta_{\text{ex}} = 5/2$  for the excess noise to fix Matern class functions for each of these components, and then employed GPR to obtain the coherence-scale hyperparameter  $l$  and its associated variance by adopting  $\eta_{21} = 1/2$  for the 21-cm signal. Note that while  $\mathbf{f}_{\text{sky}}$  and  $\mathbf{f}_{\text{mix}}$  are not independent quantities, their recovery with GPR treats them as such. In this paper, I also generate them as independent components to build my mock datasets (unlike in real data, where the mode-mixing component depends on the true sky signal). Thus, the overall quality of the recovered  $\mathbf{f}_{\text{sky}}$  and  $\mathbf{f}_{\text{mix}}$  is better from the mock datasets, than from real data. However, the effect on the accuracy of recovery is not expected to be severe, as the impact of not factoring in the inter-dependency is insignificant, as compared to the overall power of these components. That is, even without assuming their inter-dependence, GPR can recover them with reasonably high accuracy.

Next, I also assume the value  $\sigma_{\text{noise}}^2 \sim 74 \times 10^3 \text{ mK}^2$  (from M20) to simulate the noise component  $\epsilon$  for  $\approx 10$  nights of observation with LOFAR scaled to the Field of View (FoV) corresponding to the GRIZZLY simulations, which is  $3.03^\circ \times 3.03^\circ$ . The variance for the intrinsic sky-emission, mode-mixing and excess noise components scales by the noise variance as indicated in Table 2.1.

I also consider the case of  $\approx 100$  nights of observation, to provide estimates for future observations of similar duration with LOFAR. In this case, the noise variance is expected to be reduced by a factor of 10, assuming all effects to scale uniformly from  $\approx 10$  to  $\approx 100$  nights of data. However, it should be noted that the data used in M20 was plagued with issues such as RFI flagging and bad ionospheric conditions, and it would be of little scientific value to assume the same conditions to persist for a 10 times longer observational duration. For this reason, I assume ideal conditions (for example, picking observation nights with good ionospheric conditions) and assume that the thermal noise is reduced by a factor of 20 instead, obtaining  $\sigma_{\text{noise}}^2 \sim 3.7 \times 10^3 \text{ mK}^2$ . As I expect the excess noise to behave in a similar fashion, I also reduce its variance  $\sigma_{\text{ex}}^2$  by a factor of 20. M20 noted that the mode-mixing contaminants were not decreasing when integrating over more nights of data. However, the parts of this component due to effects of ionosphere and calibration errors are uncorrelated from night to night, and thus are expected to decrease with longer integration, leading to a reduction of the mode-mixing term. Additionally, this should also decrease because of the improved UV-coverage. Here, I thus consider an ideal scenario to understand how the performance of the VAE kernel compares with previously used kernels when also the mode-mixing term's variance  $\sigma_{\text{mix}}^2$  is reduced by a factor of 20. However, I scale down

the intrinsic sky component only by a factor of 2. This is justified because (i) the confusion noise limit is set by the resolution of the LOFAR telescope, so that unresolved point sources cannot be subtracted even when integrating over more nights; (ii) modelling an increasing number of point sources is not a trivial task, and doing it accurately does not seem feasible in the short-term. Thus, I limit ourselves to an assumption of modest improvements. Because of this, I assume that the variance of  $\mathbf{f}_{\text{sky}}$  is scaled by a thermal noise variance of  $\sigma_{\text{noise}}^2 \sim 37 \times 10^3 \text{ mK}^2$ , corresponding to a factor of 2 reduction from the noise variance in the  $\approx 10$  nights case.

Following the procedure outlined above, I generate two sets of foreground, noise and excess noise components, i.e. for  $\approx 10$  and  $\approx 100$  nights of observation. I then add the 21-cm signal (see Section 2.2.1) to obtain a complete mock dataset.

The input power spectra for the overall mock dataset using GRIZZLY is shown in Figures 2.1 and 2.2 for the  $x_{\text{HI}}$  and  $T_{\text{S}}$  fluctuation dominated case, respectively. The ratio between the 21-cm signal and the  $1\text{-}\sigma$  uncertainty of the noise power spectrum is a measure of the Signal to Noise Ratio (SNR). Thus, for  $k$ -bins where the 21-cm signal's power is greater than the  $1\text{-}\sigma$  error of the noise, there is a chance of detectability with  $1\text{-}\sigma$  confidence. From Figure 2.1 I can thus conclude that for  $\approx 10$  nights of observation a detection of the  $x_{\text{HI}}$  fluctuation dominated 21-cm signal is unlikely, as the SNR is  $\ll 1$  for all  $k$ -bins. However, the chances of detectability improve for  $\approx 100$  nights of observation, as  $\text{SNR} \approx 1$ . In Figure 2.2, I see that detecting the  $T_{\text{S}}$  fluctuation dominated signal should be possible also for  $\approx 10$  nights of observations, as  $\text{SNR} > 1$  for all  $k$ -bins, while for  $\approx 100$  nights it is  $\gtrsim 10$ , assuring a detection as long as the covariance kernel used to model the 21-cm signal is correctly estimated.

While I first focus on the 21-cm signal at  $z = 9.16$  in order to make a direct comparisons with M20, I also consider  $z = 8.30$  and  $10.11$  to prepare for future LOFAR results at these redshifts.

### 2.2.5 Recovery with MCMC

As discussed in Section 2.2.2, M20 used a gradient-descent method to maximise the LML. However, in this paper I instead use MCMC sampling (Foreman-Mackey et al., 2013) to estimate the hyperparameters by sampling their posterior distributions. The benefit of MCMC sampling is that it allows me to also have a measurement of the uncertainty on the hyperparameters, which can then be propagated. I build the posterior distributions adopting flat uniform priors with broad ranges as used in M20. For the coherence-scales, I provide a smaller range for the uniform prior ( $U$ ) to improve convergence time, as done by M20. Thus, as listed in the third column of Table 2.1, I use a range of (10,100) for  $l_{\text{sky}}$  and of (1,10) for  $l_{\text{mix}}$ . To ensure that the converged value for  $l_{\text{ex}}$  remains in the  $1\text{-}\sigma$  confidence interval, I used a range of (0.1, 0.8) rather than of (0.2,0.8) as in M20. While I note that a Gamma prior for the variances of the different component leads to faster convergence, I still adopt flat priors in the logarithmic scale over several orders of magnitude (thus effectively an uninformed prior) to minimise the chances of bias.

In addition to this, I note that the coherence scale and the variance for the 21-cm signal are dependent on the baseline length. So, while to recover the intrinsic sky, mode mixing and excess

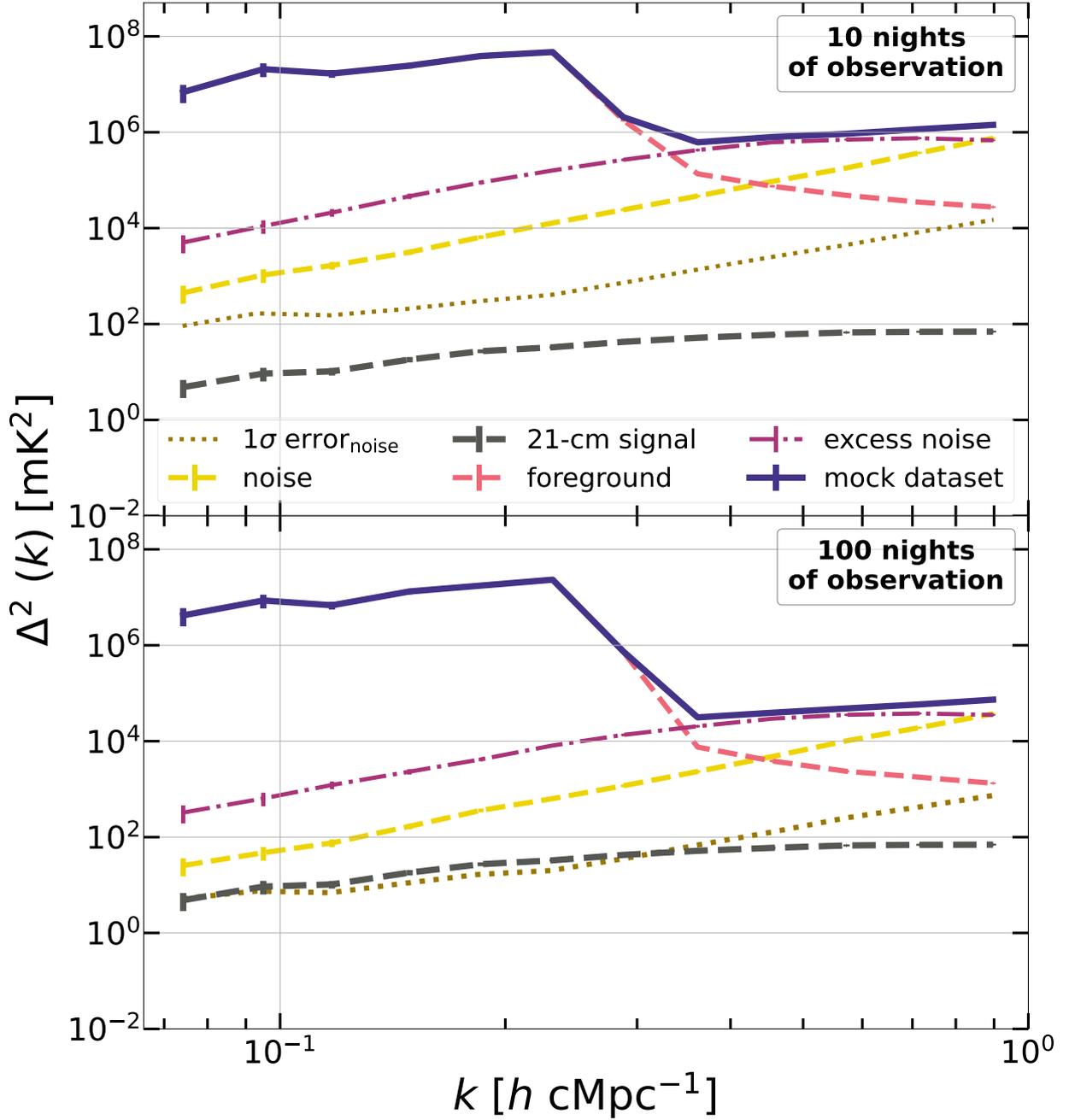


Figure 2.1: Power spectrum of the mock dataset (purple solid line) generated using the  $\chi_{\text{HI}}$  fluctuation dominated model from GRIZZLY. It consists of the foreground component (intrinsic sky + mode-mixing contaminants; pink dashed), the excess noise (magenta dashed), the noise (yellow dashed) and the 21-cm signal at  $z = 9.16$  (grey dashed). I also show the  $1 - \sigma$  upper limit (dark-yellow dotted) achievable if the dataset is thermal noise dominated, i.e. any signal below this line has  $\text{SNR} < 1$ . The top (bottom) panel refers to a case with the noise corresponding to 10 (100) nights of observation.

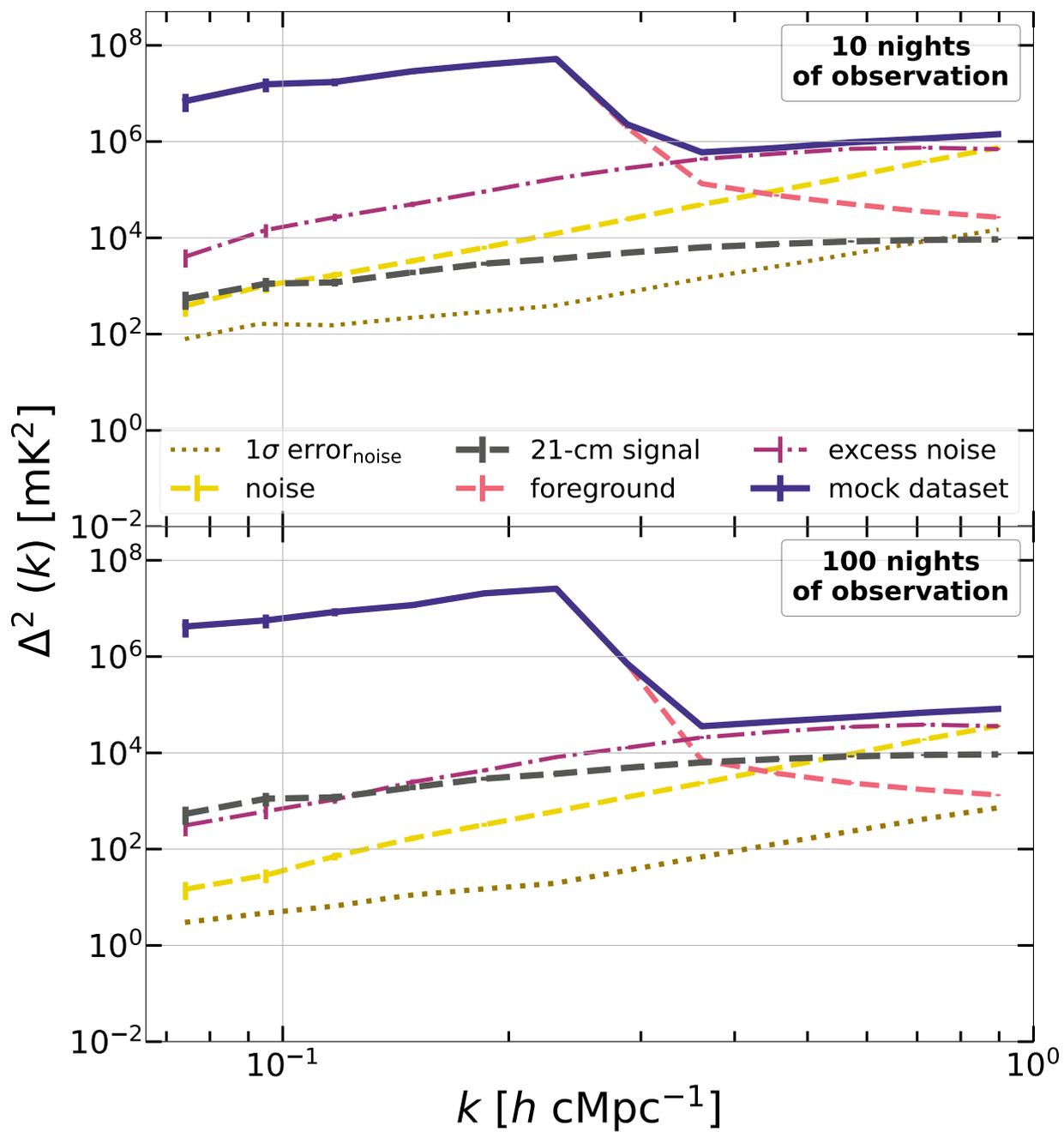


Figure 2.2: As Fig. 2.1 for the  $T_S$  fluctuation dominated model.

noise components I continue to employ the same Matern class functions used to generate them (i.e. with the same  $\eta$  values of the input, see Table 2.1), when using Matern class functions to recover the 21-cm signal, I modify the hyperparameter  $l$  and the variance as:

$$l = \frac{l_0}{1 + 0.001l_\alpha l_0(u - u_{\min})} \text{ and } \sigma^2 = \sigma_0^2 \sigma_{\text{norm}}^2 \left( \frac{u}{u_{\min}} \right)^{\sigma_\alpha^2}, \quad (2.7)$$

where  $l_0$  and  $\sigma_0^2$  are the coherence-scale parameter and variance used in M20 for baseline length  $u$  (where  $u_{\min}$  is the minimum baseline length), but now I introduce the additional parameters  $l_\alpha$  and  $\sigma_\alpha^2$  to fully define the coherence-scale hyperparameter and the variance. Further,  $\sigma_{\text{norm}}^2$  is chosen such that the mean of the variance over all baselines is  $\sigma_0^2$ . The two additional parameters (i.e.,  $l_\alpha$  and  $\sigma_\alpha^2$ ) thus allow me to encode the dependence on the baseline length into the covariance kernel for the 21-cm signal.

To recover the 21-cm signal component, I use Matern class functions with two specific values of  $\eta$ : the Exponential Matern class function with  $\eta = 1/2$  (which, as shown in M20, maximises the evidence), and the Matern32 function with  $\eta = 3/2$ . I then compare their performance for recovery to the VAE kernel using GPR. For the hyperparameters  $x_1$  and  $x_2$ , I again take an uninformed flat prior in the linear space.

## 2.3 Results

Here I discuss a qualitative comparison between the results which are shown in Figures 2.3 and 2.4 using the three aforementioned kernels (Exponential, Matern32 and VAE). I then analyse the estimated coherence-scale hyperparameter and variance values for each of the components of the mock datasets defined in Section 2.2.1.1 in Table 2.1. Further, I qualitatively compare the results obtained by using the full simulations of reionization (Section 2.2.1.2) rather than the mock datasets generated with GRIZZLY. Lastly, I explore the role of redshift on the performance of the VAE kernel, by testing cases at  $z = 8.30$  and  $10.11$ , and by comparing them against the results obtained for  $z = 9.16$ .

### 2.3.1 $x_{\text{HI}}$ fluctuation dominated model

In Figure 2.3, I compare the results from the three kernels (VAE in blue solid, Exponential in orange dashed-dotted and Matern32 in green dotted) in recovering the power spectrum of the 21-cm signal at  $z = 9.16$  (in grey dashed). I also show the  $2\text{-}\sigma$  uncertainty on the recovery from different kernels to compare their performance. I note that if the lower bound of the  $2\text{-}\sigma$  uncertainty of the recovered signal is below the uncertainty on the thermal noise (as shown in Figure 2.1) the recovery qualifies as an upper limit, otherwise it is referred to as a detection. Based on this, I note that in this case, the recovery from all kernels is going to provide upper limits, as the thermal noise uncertainty is higher than the 21-cm signal for  $\approx 10$  nights, and comparable to it for  $\approx 100$  nights.

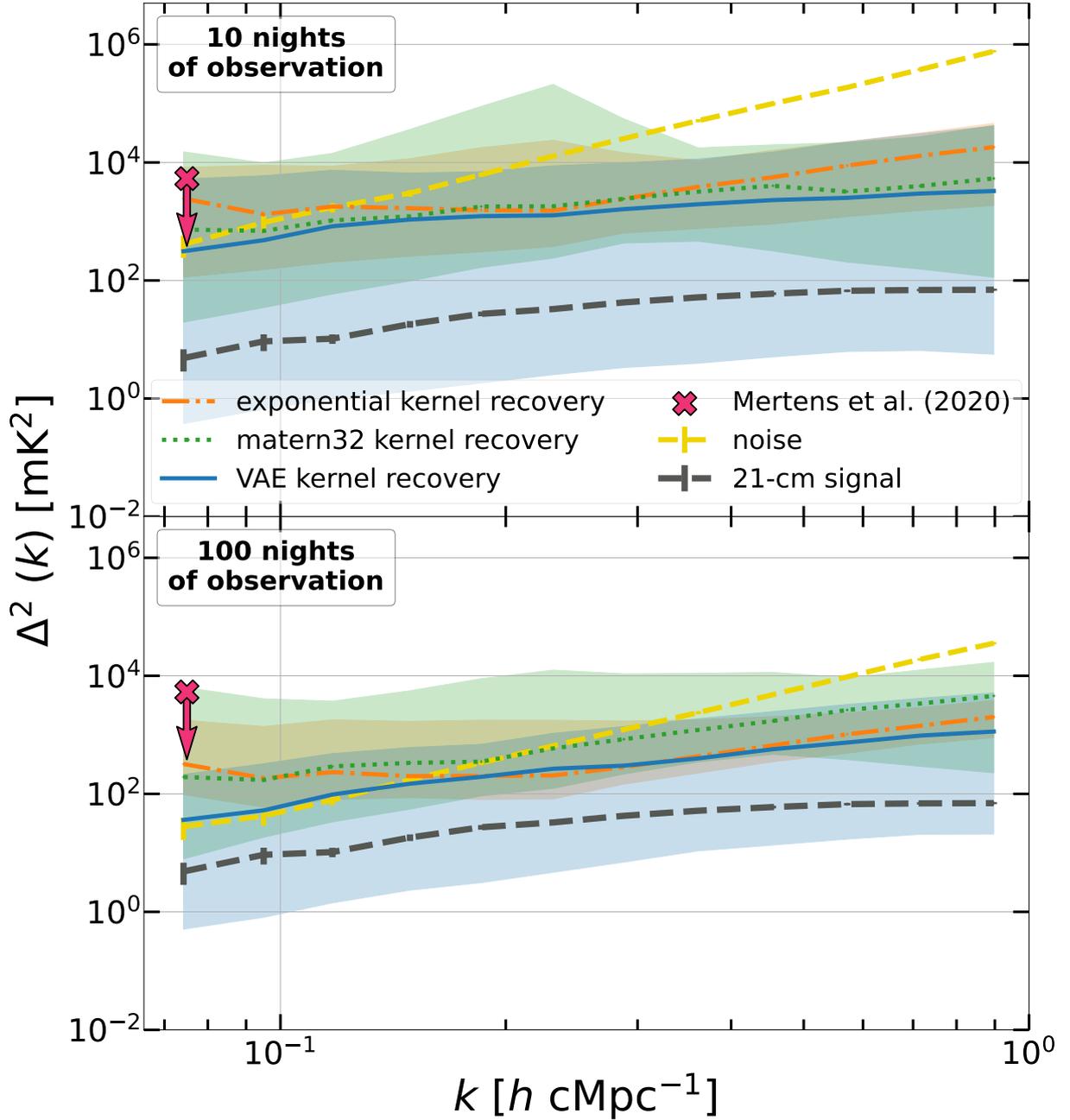


Figure 2.3: Power spectrum recovered using the Exponential (orange dashed-dotted line) and Matern32 (green dotted) Matern class functions based covariance kernels, and the VAE-based kernel (blue solid), together with the  $x_{\text{HI}}$  fluctuation dominated model 21-cm signal (grey dashed) and noise (yellow dashed) at  $z = 9.16$ . The  $2\sigma$  uncertainty on the recovered signal for each kernel is shown as a shaded area in the corresponding colours. The top (bottom) panel refers to a case with the noise corresponding to 10 (100) nights of observation. I also plot the estimated upper limits of power at  $k = 0.075 h\text{Mpc}^{-1}$  from [Mertens et al. \(2020\)](#), cross). Note that this value can be significantly higher than the signal due to more complex noise in real data.

I note that for  $\approx 10$  nights of observations, while the VAE kernel has uncertainty bands wider than the Matern-class function based kernels, the input 21-cm signal is contained within its constrained region. Thus, while the recovered signal for the three kernels is comparable, the VAE kernel is robust enough to compensate for the overestimation and to contain the signal within the  $2\text{-}\sigma$  limits of the error, although it is still an upper limit rather than a constrained detection. As discussed in Section 2.2.4, the reason for this non-detection when using Matern class functions based kernels and broad uncertainty bands for the VAE kernel is due to the low SNR, which is  $< 0.1$  across all  $k$ -bins. This, however, improves to an average SNR of  $\approx 1$  for  $\approx 100$  nights of observations, for which, as expected, I obtain tighter constraints and an improved prediction of the actual value. I see, though, that the recovered signal from Matern-class function based kernels is still over-predicted. In particular, for  $k > 0.5 \text{ hcMpc}^{-1}$  the Exponential and Matern32 kernels are unable to contain the signal even in their  $2\sigma$  uncertainty bands. On the other hand, the VAE kernel contains the signal in its  $2\sigma$  uncertainty bands across all  $k$ -bins, despite the recovered signal being about an order of magnitude higher than the input signal. The VAE kernel also does a much better job in recovering the overall shape of the power spectrum. By comparing to the estimated power at  $k = 0.075 \text{ hMpc}^{-1}$  from M20 (cross), it is clear that in this case the VAE kernel is also capable of significantly improving the 21-cm upper limits estimate. However, I highlight that the VAE kernel applied to real data is still likely to provide upper limits higher than what has been shown here, because of the more complex noise, and thus the improvement provided by the kernel might be lesser.

In the 4<sup>th</sup> and 5<sup>th</sup> columns of Table 2.1 I show the MCMC estimates for the coherence-scale hyperparameter and the variance obtained by applying GPR to the input power spectrum of the mock dataset (in indigo, Figure 2.1). The covariance kernels for the intrinsic sky, mode mixing and excess noise components are the same as those used to generate them (i.e., the value of  $\eta$  is fixed), while I adopt the VAE kernel for the 21-cm signal. Note that the variances for all components are scaled down by the corresponding value of  $\sigma_{\text{noise}}^2$ . This is equal to  $74 \times 10^3 \text{ mK}^2$  for all components for  $\approx 10$  nights of observations (see M20), while for  $\approx 100$  nights this corresponds to scaling down by a factor of 2 (i.e.,  $\sigma_{\text{noise}}^2 = 37 \times 10^3 \text{ mK}^2$ ) for the intrinsic sky component, and by a factor of 20 (i.e.,  $\sigma_{\text{noise}}^2 = 3.7 \times 10^3 \text{ mK}^2$ ) for other components (as discussed in Section 2.2.4). From the MCMC estimates, I note that the measurement of the coherence-scale for the  $\mathbf{f}_{\text{sky}}$  and  $\mathbf{f}_{\text{mix}}$  improves from  $\approx 10$  to  $\approx 100$  nights of observation. However, the variance estimates do not show an improvement, and even slightly worsen for the excess noise component. The estimates of  $x_1$  and  $x_2$  and their associated variance  $\sigma_{21}^2$  agree within error limits for both cases.

### 2.3.2 $T_{\text{S}}$ fluctuation dominated model

Here I perform a comparison between covariance kernels for the spin temperature fluctuation model, similarly to what done in the previous section. The results are shown in Figure 2.4. As seen in Figure 2.2 and discussed in Section 2.2.4, the SNR is larger than 1 also for  $\approx 10$  nights of observations, suggesting better chances of detectability and smaller uncertainty ranges. Indeed, all three covariance kernels contain the signal within the  $2\sigma$  uncertainty bands around their

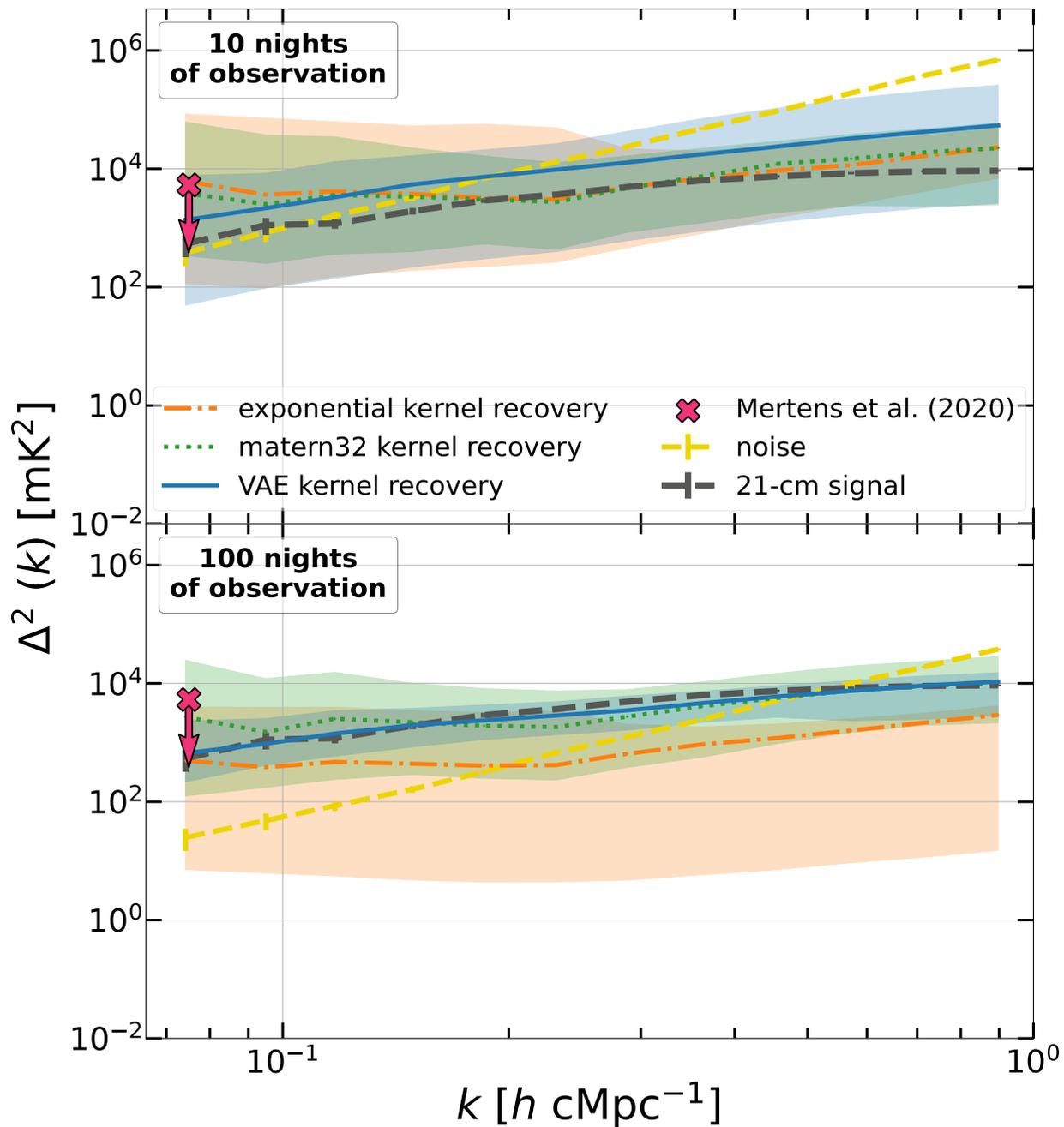


Figure 2.4: As Figure 2.3 for the  $T_S$  fluctuation dominated model.

recovered signal, and the uncertainty range for the VAE kernel is  $\approx 2$  orders of magnitude smaller than the equivalent case for the  $x_{\text{HI}}$  fluctuation dominated model. As discussed in the previous section, I compare the lower bounds of the recovered signal with the thermal noise uncertainty from Figure 2.2 to classify the recovery as a detection or an upper limit.

For  $\approx 100$  nights of observations, the SNR is  $\gg 10$  across all  $k$ -bins, so that the recovered signal is expected to reproduce the input signal with a significantly narrower  $2\sigma$  uncertainty range, provided that the covariance kernel chosen is a reliable estimate of the covariance of the input 21-cm signal. Indeed, from Figure 2.4 I note that the VAE kernel fully recovers the signal with less than one order of magnitude uncertainty. However, the Exponential kernel contains the 21-cm signal only in the lowest  $k$ -bins and shows significant bias in the estimated power at higher  $k$ -bins. Its broad  $2\text{-}\sigma$  uncertainty shows that the recovery just provides upper limits even in the low  $k$ -bins. On the other hand, the Matern32 kernel performs better, and provides a successful detection, albeit with broader uncertainty ranges than the VAE kernel recovery. This suggests that the Exponential kernel is definitely not a good match for the covariance of the input 21-cm signal, and, as highlighted by Kern & Liu (2021), would lead to significant errors in the estimated physics, if used. The Matern32 kernel is better, however the VAE kernel improves upon it even further. This problem with the Exponential kernel appears in the  $\approx 100$  and not in the  $\approx 10$  nights of observation due to the similarity of power and shape of the excess-noise and 21-cm signal components. Thus, a covariance kernel which is not a reliable estimate of the covariance of the input 21-cm signal would not be able to distinguish between the two, and may either ignore both equally, or identify one over the other purely by chance. It can also be argued that the only reason for any ‘‘detection’’ at low  $k$ -bins using the Exponential kernel could possibly just be the detection of the excess noise component, wrongly interpreted as the 21-cm signal one.

The recovered values for the coherence-scales and variances for the intrinsic sky, mode mixing and excess noise components, as well as those for the hyperparameters  $x_1$  and  $x_2$  and associated variance for the 21-cm signal are listed in the 6<sup>th</sup> and 7<sup>th</sup> columns of Table 2.1 along with their 68% confidence intervals. As expected, I find an improvement in recovery of the input values for  $\approx 100$  nights of observation in comparison to  $\approx 10$  nights, particularly for the coherence-scale hyperparameter.

While I note better estimates for  $l_{\text{sky}}$  and  $\sigma_{\text{sky}}^2/\sigma_{\text{noise}}^2$  for  $\approx 100$  nights of observations in both 21-cm signal models, in the  $x_{\text{HI}}$  fluctuation dominated model the input  $\sigma_{\text{sky}}^2/\sigma_{\text{noise}}^2$  is not included within the estimate error of the recovered values. A similar behaviour is observed in the recovery of the variance for  $\mathbf{f}_{\text{mix}}$  and  $\mathbf{f}_{\text{ex}}$  in the  $x_{\text{HI}}$  fluctuation dominated model for  $\approx 100$  nights of observations. Lastly, I also note that the  $x_1$  and  $x_2$  hyperparameters and associated variance for the 21-cm signal in both models agree within the error estimates. While the estimated variance for  $\approx 100$  nights of observations is higher, it also has a broad 68% confidence interval.

Table 2.1: I use the results of **M20** as my input (2<sup>nd</sup> column) for the hyperparameters (1<sup>st</sup> column) to generate a mock dataset composed of the  $\mathbf{f}_{\text{sky}}$ ,  $\mathbf{f}_{\text{mix}}$ , and  $\mathbf{f}_{\text{ex}}$  components, in addition to the  $x_{\text{HI}}$  fluctuation dominated models from GRIZZLY to generate the 21-cm signal component at  $z = 9.16$ . For the recovery using GPR, I keep the value of  $\eta$  fixed for the  $\mathbf{f}_{\text{sky}}$ ,  $\mathbf{f}_{\text{mix}}$ , and  $\mathbf{f}_{\text{ex}}$  components, and estimate the median and the 68% confidence intervals for the coherence-scales and variances from the MCMC-based LML maximization (4<sup>th</sup> to 7<sup>th</sup> columns) using the priors as in **M20** (3<sup>rd</sup> column), i.e., I either use linear scale uniform priors  $U$ , or logarithmic scale indicated as  $-$ . As the latter is over several orders of magnitude, it has not been listed and can be assumed to be an uninformative prior from  $-\infty$  to  $+\infty$ . For  $\approx 10$  nights of observation, I use  $\sigma_{\text{noise}}^2 = 74 \times 10^3 \text{ mK}^2$  (as obtained in **M20**). For  $\approx 100$  nights of observation, I scale it down by a factor of 2 for the  $\mathbf{f}_{\text{sky}}$  component and by a factor of 20 for the rest of the components. The values for  $\sigma_{\text{noise}}^2$  in each case are listed in the first row for each component. Note that  $\sigma_{\text{noise}}^2$  has been scaled to the field of view of the GRIZZLY simulations, which is  $3.03^\circ \times 3.03^\circ$ .

Hyperparameter	Input value from <b>M20</b>	Prior	$x_{\text{HI}}$ fluctuation dominated 10 nights	100 nights	$T_S$ fluctuation dominated 10 nights	100 nights
$\sigma_{\text{noise}}^2$ [ $\times 10^3 \text{ mK}^2$ ]			74	37	74	37
$\eta_{\text{sky}}$	$+\infty$		$+\infty$	$+\infty$	$+\infty$	$+\infty$
$l_{\text{sky}}$	47.5	$U(10, 100)$	$43.44 \pm 4.02$	$47.76 \pm 1.64$	$52.00 \pm 6.91$	$45.84 \pm 1.67$
$\sigma_{\text{sky}}^2 / \sigma_{\text{noise}}^2$	611	$-$	$665.17 \pm 29.52$	$579.91 \pm 24.11$	$598.99 \pm 27.13$	$633.43 \pm 28.23$
$\sigma_{\text{noise}}^2$ [ $\times 10^3 \text{ mK}^2$ ]			74	3.7	74	3.7
$\eta_{\text{mix}}$	3/2		3/2	3/2	3/2	3/2
$l_{\text{mix}}$	2.97	$U(1, 10)$	$2.99 \pm 0.07$	$2.97 \pm 0.08$	$3.08 \pm 0.08$	$2.90 \pm 0.08$
$\sigma_{\text{mix}}^2 / \sigma_{\text{noise}}^2$	50.4	$-$	$54.80^{+2.53}_{-2.42}$	$46.05^{+2.14}_{-2.04}$	$55.82^{+2.64}_{-2.52}$	$52.12^{+2.58}_{-2.46}$
$\sigma_{\text{noise}}^2$ [ $\times 10^3 \text{ mK}^2$ ]			74	3.7	74	3.7
$\eta_{\text{ex}}$	5/2		5/2	5/2	5/2	5/2
$l_{\text{ex}}$	0.26	$U(0.1, 0.8)$	$0.26 \pm 0.003$	$0.26 \pm 0.005$	$0.27 \pm 0.01$	$0.26 \pm 0.01$
$\sigma_{\text{ex}}^2 / \sigma_{\text{noise}}^2$	2.18	$-$	$2.15^{+0.05}_{-0.05}$	$1.91^{+0.11}_{-0.10}$	$2.25^{+0.07}_{-0.07}$	$2.32^{+0.18}_{-0.17}$
$\sigma_{\text{noise}}^2$ [ $\times 10^3 \text{ mK}^2$ ]			74	3.7	74	3.7
$x_1$	$-$	$-$	$-0.10 \pm 0.99$	$-0.23 \pm 1.02$	$-0.41 \pm 0.92$	$-0.07 \pm 1.00$
$x_2$	$-$	$-$	$-0.17 \pm 0.98$	$0.42 \pm 0.77$	$0.92 \pm 0.59$	$-0.06 \pm 0.31$
$\sigma_{21}^2 / \sigma_{\text{noise}}^2$	$-$	$-$	$0.04^{+0.21}_{-0.03}$	$0.17^{+0.37}_{-0.12}$	$0.25^{+0.33}_{-0.14}$	$1.70^{+1.07}_{-0.65}$

### 2.3.3 CRASH simulations

The power spectra resulting from the recovery using the three kernels in the case of the CRASH simulation are shown in Figure 2.5. As now the input 21-cm signal has a power which lies in between the two GRIZZLY models, this translates into an intermediate SNR across  $k$ -bins. Due to this, I am able to successfully contain the input signal in the  $2\sigma$  ranges of the recovered signals in most  $k$ -bins for  $\approx 10$  nights of observation with all kernels, but given the thermal noise uncertainty power, it is still classified as an upper limit. However, for  $\approx 100$  nights, I note that while the VAE kernel does an excellent job of recovering the signal with narrow  $2\sigma$  uncertainty bands, they still indicate that the recovery is an upper limit. On the other hand, the Matern class functions based kernels underestimate the signal and do not contain the input 21-cm signal within  $2\sigma$  uncertainty bands for some  $k$ -bins, despite them being significantly larger. When comparing to the thermal noise uncertainty, I note that they only provide upper limits for the signal in most  $k$ -bins. This result is similar to that with the  $T_S$  fluctuation dominated model with GRIZZLY, and thus is due to the same reasons discussed in Section 2.3.2.

### 2.3.4 Redshift dependence

To evaluate the performance of the VAE kernel at different redshifts, I use the simulations at  $z = 8.30, 9.16$  and  $10.11$  of both GRIZZLY models, and compare the performance of the VAE kernel for noise levels equivalent to  $\approx 100$  nights of observations. It should be noted that I use the VAE kernels trained for the respective redshifts to avoid making the assumption of the kernels being redshift agnostic.

To analyse my recovery technique, I check how *accurately* it recovers the input signal, and how *precise* the results it reports are. For this purpose, I use the quantities defined below, with their values listed in Table 2.2:

1. *Average bias*,  $\langle \text{PS}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$ : this is the average bias given as the ratio between the recovered ( $\text{PS}_{\text{rec}}$ ) and input 21-cm signal ( $\text{PS}_{\text{in}}$ ) power spectra, averaged across all  $k$ -bins. An accurate recovery has a value close to 1, with higher (lower) values indicating an over(under)-estimation of the input 21-cm signal. I note that the deviation of average bias from 1 increases significantly with redshift for the  $x_{\text{HI}}$  fluctuation dominated model. However, the trend is not so clear for the  $T_S$  fluctuation dominated signal, as the power and slope of the excess noise component match closely those of the input 21-cm signal at  $z = 8.30$  and  $k \leq 0.2 \text{ hMpc}^{-1}$ , which makes differentiating between them more difficult. However, I still get an average bias of  $\approx 0.70$ , suggesting only a minor under-estimation, which is primarily observed at the lowest  $k$ -bins (see Figure 2.6).
2. *Average scaled uncertainty*,  $\langle \text{Err}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$ : the  $2\sigma$  uncertainty (given by the shaded region in Figures 2.3, 2.4, 2.5, and referred to as  $\text{Err}_{\text{rec}}$  hereafter) on the recovered power spectrum ( $\text{PS}_{\text{rec}}$ ) gives the precision of recovery. However, except for cases of extremely poor recovery, the absolute value of  $\text{Err}_{\text{rec}}$  generally depends on the absolute value of the

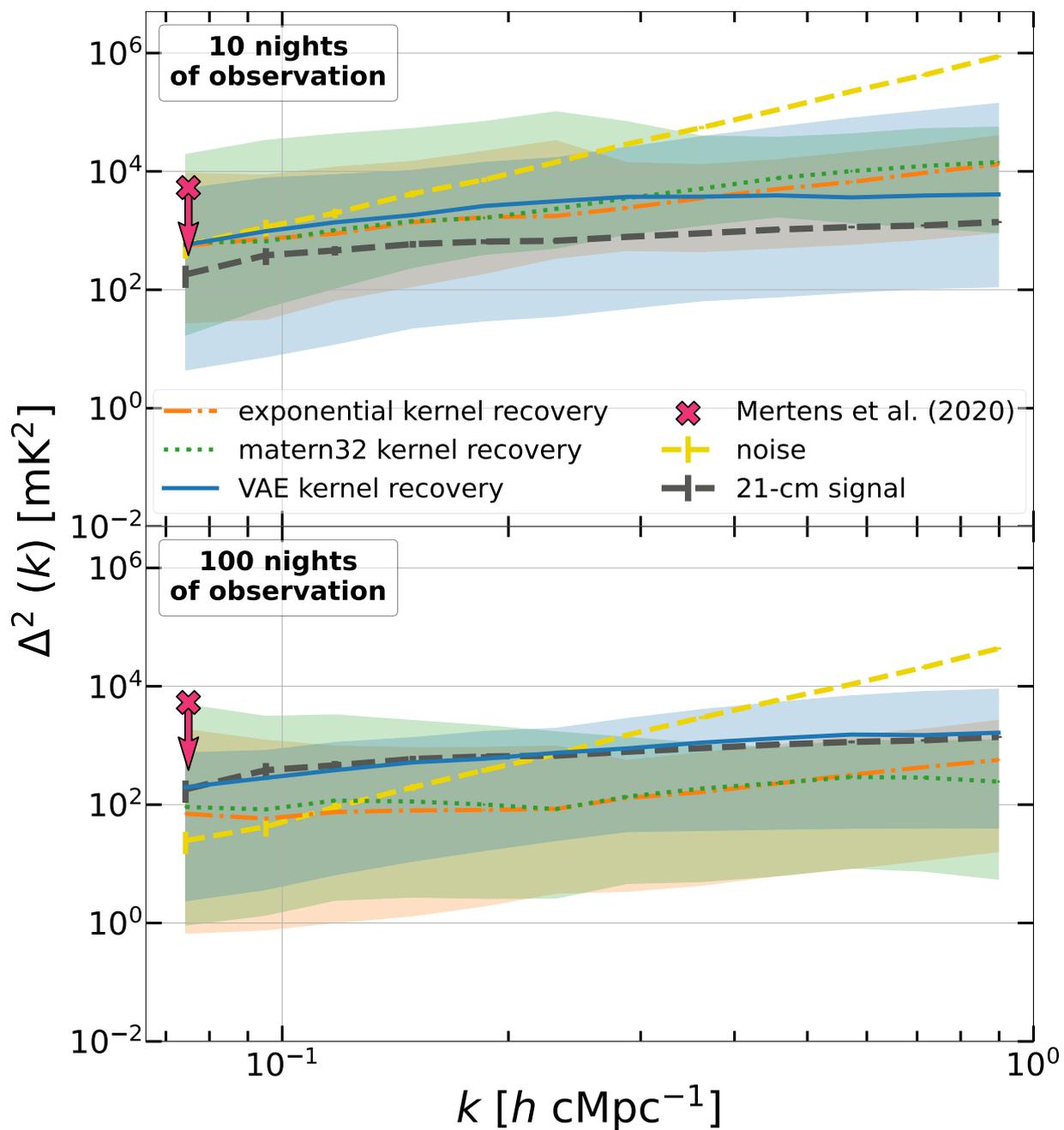


Figure 2.5: As Figure 2.3 for the CRASH simulation of reionization.

corresponding  $\text{PS}_{\text{rec}}$ . Thus, to compare different recoveries for a given  $\text{PS}_{\text{in}}$ , I need to convert it into a unitless quantity. I first tried to do this by calculating the ratio  $\text{Err}_{\text{rec}}/\text{PS}_{\text{rec}}$  as a measure of the precision of the recovery. However, this ratio contains no information on the accuracy of  $\text{PS}_{\text{rec}}$ . Indeed, one could have a precise recovery, i.e. a low  $\text{Err}_{\text{rec}}/\text{PS}_{\text{rec}}$ , but an inaccurate  $\text{PS}_{\text{rec}}$ , i.e. a bias which deviates significantly from 1 (see point above). This ratio, then, is not useful, as it does not quantify the overall quality of recovery. To overcome this issue, I use the ratio  $\text{Err}_{\text{rec}}/\text{PS}_{\text{in}}$  instead. As the magnitude of  $\text{Err}_{\text{rec}}$  depends on that of  $\text{PS}_{\text{rec}}$ , it also carries the information of the bias in recovery. Further, as I divide by  $\text{PS}_{\text{in}}$ , the scaled uncertainty becomes independent of the specific  $\text{PS}_{\text{in}}$  being recovered. This allows me to use it to compare between cases with different  $\text{PS}_{\text{in}}$ , such as at different redshifts (see top panel of Figure 2.6 where the power spectrum varies due to the time evolution of ionising bubbles) and different physical models (as discussed in Sections 2.2.1). I use this generalised comparison, and call it the *scaled* uncertainty for each  $k$ -bin. Averaging this across all  $k$ -bins provides a handy quantity to compare the precision of recovery for input 21-cm signals with different physical properties. For example, in Figure 2.6 I observe that in the  $x_{\text{HI}}$  fluctuation dominated model, while small-scale variability due to partial reionization is restricted, it still has variability tied to the large-scale distribution of neutral Hydrogen, which increases for lower redshifts. This boosts the power at large scales, corresponding to the low  $k$ -bins. The same can also be seen in the  $T_{\text{S}}$  fluctuation dominated model, although its overall power is boosted, as it allows small-scale variability in  $\delta T_{\text{b}}$  as well. Using the scaled uncertainty, I can compare the precision of recovery across redshift for both cases. The difference in  $\langle \text{Err}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$  is quite significant for the  $x_{\text{HI}}$  fluctuation dominated model, going from  $\approx 15$  at  $z = 8.30$  to  $\approx 80$  at  $z = 10.11$ .

3. *z-score*,  $\langle z\text{-score} \rangle_k$ : The z-score (Kirch, 2008) is a popular quantity to evaluate quality of recovery. In my case, it is defined as  $\frac{\text{PS}_{\text{rec}} - \text{PS}_{\text{in}}}{\sigma}$  or  $\frac{\text{PS}_{\text{rec}} - \text{PS}_{\text{in}}}{\text{Err}_{\text{rec}}/2}$  at each  $k$ -bin, and it measures how much the recovered signal deviates from the input 21-cm signal, in units of standard deviation of the recovered signal. The z-score is thus a more explicit method to combine into a single quantity the information provided by the bias along with that of the uncertainty. The only possible issue is that  $\sigma (= \text{Err}_{\text{rec}}/2)$  depends on  $\text{PS}_{\text{rec}}$  and thus their ratio would mask the accuracy of recovery as discussed in the point above (see below for an example case). Further, I note that I cannot just report an average of z-scores across all  $k$ -bins, as the distribution of z-scores is not necessarily Gaussian. Indeed, I find that while it is approximately Gaussian for the  $x_{\text{HI}}$  fluctuation dominated model, this is not the case for the  $T_{\text{S}}$  fluctuation dominated model. Thus, I report the minimum and maximum z-scores ( $z\text{-score}_{\text{min}}$  and  $z\text{-score}_{\text{max}}$ ) along with the average ( $\langle z\text{-score} \rangle_k$ ). When  $\langle z\text{-score} \rangle_k > (<) 0$ , its exact value quantifies the extent of over(under)-prediction. I note that in the  $x_{\text{HI}}$  fluctuation dominated model, the average z-score worsens with increasing redshift, consistently with the behaviour of the average bias and average scaled uncertainty. This trend is not detected for the  $T_{\text{S}}$  fluctuation dominated signal, due to the same reasons discussed above for the average bias in (i). I also note that at  $z = 10.11$ ,  $\langle z\text{-score} \rangle_k \approx -2$  and  $+1$  in the  $T_{\text{S}}$  and  $x_{\text{HI}}$  fluctuation dominated model respectively, naively suggesting a better recovery for the latter. This, though, is not correct, but simply a consequence of the very broad error bars and the

Table 2.2: Average bias,  $\langle \text{PS}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$ , average scaled uncertainty,  $\langle \text{Err}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$ , and average z-score,  $\langle z - \text{score} \rangle_k$ , at various redshifts. These values are obtained when the signal recovery is done employing the VAE kernel with the  $x_{\text{HI}}$  (top) and the  $T_{\text{S}}$  (bottom) fluctuation dominated model and  $\approx 100$  nights of observations.

$z$	$\langle \frac{\text{PS}_{\text{rec}}}{\text{PS}_{\text{in}}} \rangle_k$	$\langle \frac{\text{Err}_{\text{rec}}}{\text{PS}_{\text{in}}} \rangle_k$	$z\text{-score}_{\text{min}}$	$\langle z\text{-score} \rangle_k$	$z\text{-score}_{\text{max}}$
$x_{\text{HI}}$ fluctuation dominated					
8.30	3.90	14.88	+0.46	+0.74	+0.85
9.16	8.57	21.12	+0.73	+0.82	+0.88
10.11	29.29	78.22	+0.87	+1.02	+1.16
$T_{\text{S}}$ fluctuation dominated					
8.30	0.69	0.50	-4.00	-1.36	-1.25
9.16	0.90	0.58	-0.79	-0.60	+0.23
10.11	0.34	0.83	-3.50	-1.96	-0.92

inverse proportionality of  $\langle z\text{-score} \rangle_k$  with the error. This reasoning exposes the limitation of the z-score. Indeed, by looking at  $\langle \text{PS}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$  and  $\langle \text{Err}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$  for recovery of the  $x_{\text{HI}}$  fluctuation dominated signal (see Table 2.2), I note that the deviation from zero bias (obtained when  $\langle \text{PS}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k = 1$ ) and zero uncertainty (when  $\langle \text{Err}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k = 0$ ) is significantly higher than for the  $T_{\text{S}}$  fluctuation dominated model. In fact, these numbers suggest a better quality of recovery in the latter case, which is understandable as the SNR in this model is higher. Thus, while I report the z-score due to its popularity, I recommend using the average bias and scaled uncertainty for evaluating the quality of recovery.

The trends in various quantities discussed above are linked to the physical nature of the 21-cm power spectrum and its redshift evolution. The drop in the SNR with increasing redshift (due to a decrease in signal power as shown in Figure 2.6 and explained in (ii)), leads to a worsening of the average bias and scaled uncertainty, especially for the  $x_{\text{HI}}$  fluctuation dominated model. As already mentioned, when the excess noise and the input 21-cm signal have similar power and slope (as at  $z = 8.30$  for the  $T_{\text{S}}$  fluctuation dominated case), I observe limitations in the capability of differentiating among the two, but the effects are minor and the trends of average bias, scaled uncertainty and z-score for the  $x_{\text{HI}}$  fluctuation dominated model are also observed in the  $T_{\text{S}}$  fluctuation dominated model when going from  $z = 9.16$  to  $z = 10.11$ .

Thus, I find that the VAE kernel does not add significant biases, with its recovery and associated uncertainty largely scaling with the physical properties of the 21-cm signal.

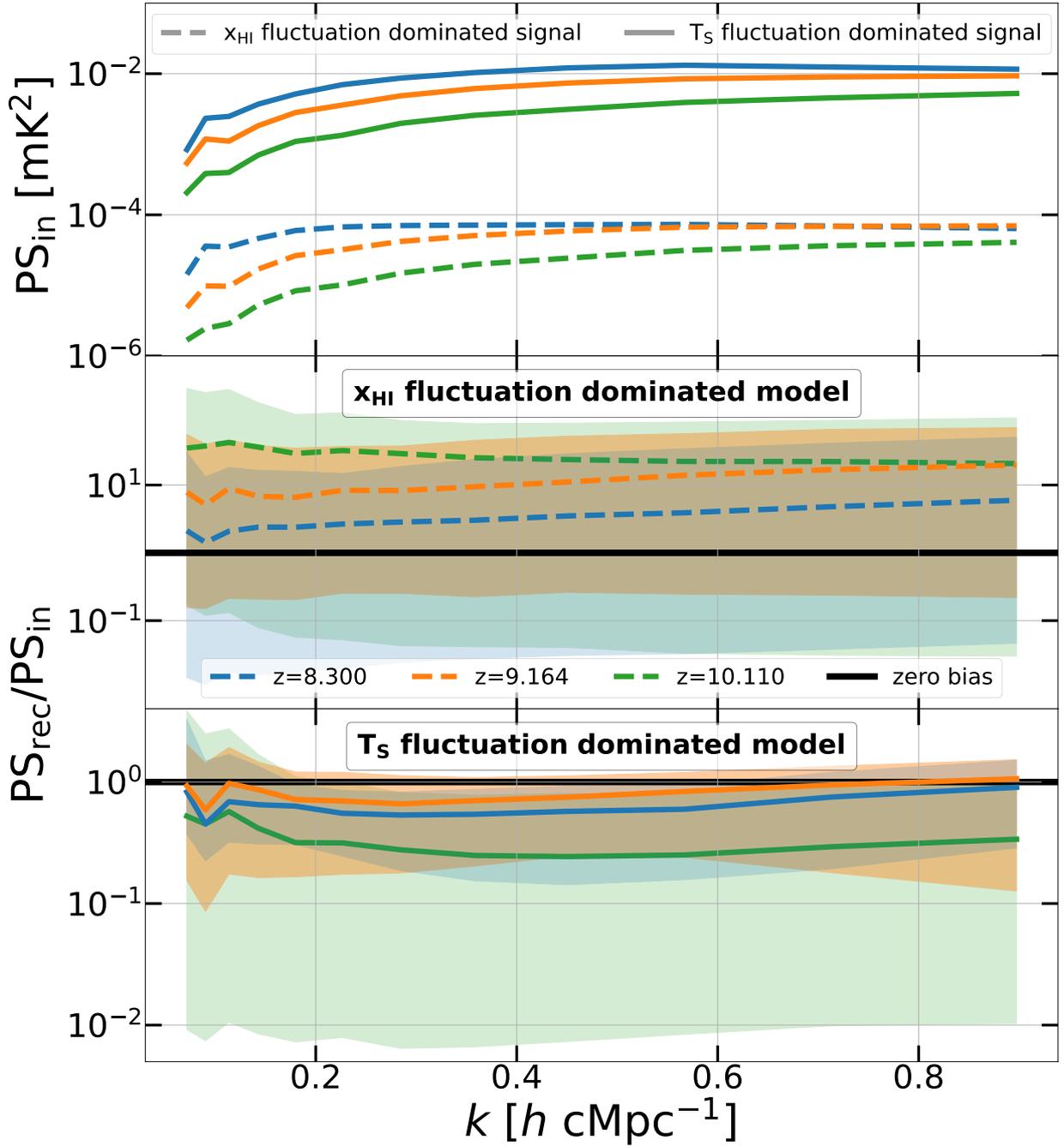


Figure 2.6: *Top panel:* input 21-cm signal for the  $x_{\text{HI}}$  (dashed lines) and the  $T_{\text{S}}$  (solid) fluctuation dominated model are shown for  $z = 8.30$  (blue),  $9.16$  (orange) and  $10.11$  (green). *Middle panel:* recovered 21-cm signal and its associated uncertainty divided by the input 21-cm signal to give the bias and scaled uncertainty (the average values over  $k$ -bins for these quantities are listed in Table 2.2) for the  $x_{\text{HI}}$  fluctuation dominated model. For comparison, the line of zero bias (i.e., recovered signal perfectly matching the input signal) is also shown. *Bottom panel:* same as the middle panel, but for the  $T_{\text{S}}$  fluctuation dominated model.

Table 2.3: Average bias,  $\langle \text{PS}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$ , average scaled uncertainty,  $\langle \text{Err}_{\text{rec}}/\text{PS}_{\text{in}} \rangle_k$ , and average z-score,  $\langle z - \text{score} \rangle_k$ , for the  $x_{\text{HI}}$  fluctuation dominated model with  $\approx 100$  nights of observations at  $z = 9.16$ . These values are obtained when the signal recovery is done employing the VAE kernel and multiplying the coherence-scale hyperparameter,  $l_{\text{ex}}$ , or the variance,  $\sigma_{\text{ex}}^2$ , listed in the second column of Table 2.1 (and as used in M20) by a factor  $f_{\text{var}}$ .

$f_{\text{var}}$	$\langle \frac{\text{PS}_{\text{rec}}}{\text{PS}_{\text{in}}} \rangle_k$	$\langle \frac{\text{Err}_{\text{rec}}}{\text{PS}_{\text{in}}} \rangle_k$	z-score <sub>min</sub>	$\langle \text{z-score} \rangle_k$	z-score <sub>max</sub>
$l_{\text{ex}}$					
0.5	12.88	31.37	+0.88	+1.04	+1.19
0.75	9.61	26.75	+0.77	+0.89	+0.97
1.0	8.57	21.12	+0.73	+0.82	+0.88
1.25	8.04	27.40	+0.75	+0.82	+0.89
1.5	7.77	15.65	+0.78	+0.84	+0.90
1.75	5.64	14.76	+0.81	+0.87	+0.93
2.0	4.55	19.38	+0.81	+0.85	+0.94
$\sigma_{\text{ex}}^2$					
1.00	8.57	21.12	+0.73	+0.82	+0.88
0.75	6.64	32.81	+0.75	+0.80	+0.88
0.50	4.77	16.89	+0.66	+0.71	+0.78
0.25	3.24	9.13	+0.58	+0.62	+0.67

## 2.4 Discussion

### 2.4.1 Role of the excess noise component

In M20 it was noted that the excess noise component was poorly constrained, and thus the combined excess noise and 21-cm signal components were jointly recovered, as separating them was not statistically justifiable. Thus, it is important to understand how well constrained the excess noise component has to be, in order to separate the 21-cm signal from it.

To explore this, I looked at the  $x_{\text{HI}}$  fluctuation dominated model of the 21-cm signal at  $z = 9.16$ . I note that in the  $\approx 100$  nights case for the chosen excess noise component, the power spectrum recovered with the VAE kernel is slightly overestimated, with the  $2\sigma$  bands on both sides of it spanning  $\approx 2$  orders of magnitude. I then generated a range of input excess noise components (by varying either the coherence-scale hyperparameter or the variance), and analysed the effects on the recovery of the 21-cm signal power spectrum using the VAE kernel by looking at the average bias, scaled uncertainty and z-score, as defined in Section 2.3.4. The results are reported in Table 2.3, where  $f_{\text{var}}$  is the factor by which the coherence-scale hyperparameter and the variance from the results of M20 were scaled. I see that, overall, the average bias and scaled uncertainty are reduced when increasing the coherence-scale hyperparameter or decreasing the variance, while no substantial difference is observed in the average z-score. This is possibly because the bias and scaled uncertainty decrease at the same rate, and thus their effects roughly cancel out.

### 2.4.2 Overall performance of the VAE kernel

Usually it is expected that the recovery of the 21-cm signal from an overall dataset with foreground components, noise and signal is not possible if  $\text{SNR} < 1$ . In Section 2.3.1 I have indeed shown that Matern class functions based kernels are unable to contain the input 21-cm signal within their  $2\sigma$  uncertainty bands when  $\text{SNR} < 1$ . However, the VAE kernel is not only able to do so, but also to recover the overall shape of the power spectrum, as seen in the top panel of Figure 2.3.

Further, as highlighted by Kern & Liu (2021), misestimation of the covariance kernel can significantly hamper signal detection given the currently used normalization and bias correction schemes. This means that as the Matern class functions model the 21-cm signal only approximately, their results can be significantly biased for more complex models of the 21-cm signal, as given e.g. by the  $T_S$  fluctuation dominated model of GRIZZLY and the 21-cm signal model from the CRASH simulations. Indeed, in these cases the Matern class functions based kernels fail to recover the signal also for noise levels equivalent to  $\approx 100$  nights of observation, even when  $\langle \text{SNR} \rangle_k \approx 10$ . The VAE kernel does not suffer from such a limitation and performs well also when used with an input signal from the CRASH simulations. This shows that the VAE covariance kernel is a more robust estimate of the covariance of the 21-cm signal, and can successfully report a detection within  $2\sigma$  uncertainty regardless of the exact physical properties of the observed 21-cm differential brightness temperature power spectrum. Lastly, I note that it performs well across all redshifts analysed here. This reconfirms the robustness of the VAE kernel in constraining the 21-cm signal, with an increase/decrease in uncertainty tied to the  $\langle \text{SNR} \rangle_k$  of the signal itself.

Overall, the  $2\sigma$  uncertainty bands given by the VAE kernel contain the signal in all cases discussed here. I consider the recovery limit of the VAE kernel in terms of SNR averaged over  $k$ -bins to be that from  $\approx 10$  nights of observation in the  $x_{\text{HI}}$  fluctuation dominated model, i.e.  $\langle \text{SNR} \rangle_k = 5 \times 10^{-2}$ . For values lower than this, I do not expect the VAE kernel based recovery to contain the input signal within its  $2\sigma$  uncertainty bands across  $k = [0.05, 1.00] \text{ hMpc}^{-1}$ . I explore one such case in Appendix A, and indeed find that the recovery does not contain the input signal within its  $2\sigma$  uncertainty bands, but instead provides upper limits. Yet it still outperforms the Matern class functions based kernels in recovering the shape of the power spectrum, spread of uncertainty on recovery, as well as the reported upper limits.

### 2.4.3 Limitations

In this work, I reduce the possibility of biases in the EoR covariance kernel by incorporating a more physically motivated covariance. I showcase the robustness of the generated VAE covariance kernel by testing it against not just mock 21-cm signals obtained with GRIZZLY, but with signals generated using very different frameworks as shown in Section 2.3.3 and in Appendix A. This has also been demonstrated in Mertens et al. (2024).

However, biases are still possible, especially if the true signal, and thus its covariance, is fundamentally different from what I obtain with my simulation codes. One way to further minimise this bias is to use mock data obtained from different codes to train the covariance kernel, which

I plan to do in further upgrades of my pipeline. In the future, I will also investigate methods to reduce the dependence on the prior by using different normalization and bias correction schemes as suggested by [Kern & Liu \(2021\)](#).

## 2.5 Summary

The LOFAR Epoch of Reionization (EoR) KSP team strives for a successful detection of the 21-cm signal from the EoR at  $z \approx 7 - 11$ . Past theoretical models indicated that at least 1000 hours of observation would be necessary to lead to a successful detection ([Mertens et al., 2018](#)), while [Mertens et al. \(2020\)](#) provided upper limits using 141 hours ( $\approx 10$  nights) of observation. In this respect, an optimal choice of the covariance kernel for the 21-cm signal component is crucial. Indeed, as shown in [Kern & Liu \(2021\)](#), given the currently used normalization and bias correction scheme, a mismatch between the adopted and the actual covariance kernel of the 21-cm signal can induce a significant signal loss, which can in turn lead to incorrect astrophysical interpretations from any “successful” detection.

To improve the choice of the 21-cm signal covariance kernel, [Mertens et al. \(2024\)](#) introduce a Machine Learning method which employs a Variational Auto-Encoder (VAE) based algorithm. As the training done using VAE is not limited by the form of the specific function (as e.g. in the case of Matern class functions), nor by the kernels of the training datasets (as in the case of a simple Auto-Encoder), it allows to reproduce the covariance kernel of the 21-cm signal with a greater flexibility. This is showcased in terms of the robustness of the VAE based kernel’s performance in comparison to previously used kernels based on Matern class functions.

I show that the result on using the VAE kernel is able to contain the input 21-cm signal within its  $2\sigma$  uncertainty band in all cases explored where  $\langle \text{SNR} \rangle_k \gtrsim 5 \times 10^{-2}$ . It is also usually better than the results from Matern class functions based covariance kernels in recovering the overall shape of the power spectrum of the signal. A key result in this paper is that Matern class functions based kernels are unable to recover the 21-cm signal for the  $T_S$  fluctuation dominated model even for  $\approx 100$  nights of observation, for which  $\langle \text{SNR} \rangle_k \approx 10$ , while a recovery with the VAE kernel is successful. A similar result is obtained also with a 21-cm signal generated using the CRASH simulations, thus clearly indicating that the Matern class functions based kernels do not correctly match the covariance of more complex models of the signal. Thus, this analysis suggests that using the VAE kernel can mitigate to a significant extent the issues highlighted by [Kern & Liu \(2021\)](#) given no change to the normalization and bias correction schemes.

Further, I show that the behaviour of the VAE kernel is consistent across all redshifts of interest, with changes in its performance strongly correlating with the neutral hydrogen distribution, which changes the strength of the resultant power spectrum, and thus the  $\langle \text{SNR} \rangle_k$ . This suggests that the VAE based kernel can be used for any choice of redshift without additional correction factors, making the algorithm developed here directly applicable to LOFAR data at  $z \approx 8$  to 10, whose results can then be compared with results from telescopes like HERA.

I also explore the effects that the properties of the excess noise component identified in [M20](#) have

on the recovery of the 21-cm signal. As expected, I find that having a higher coherence scale or a lower variance for the components leads to better recovery.

In companion papers I will apply the VAE kernel to the  $\approx 10$  nights of LOFAR data used in [M20](#), and explore the range of theoretical models which are consistent with the upper limits provided by the VAE kernel, as done in [Ghara et al. \(2020\)](#). Applying the VAE kernel to observations much longer than  $\approx 10$  nights requires a significant improvement in the modelling of the intrinsic sky component, which would eventually be limited by the confusion noise due to the angular resolution of LOFAR. Further improvements, such as noise mitigation, can be implemented by choosing data from nights with better ionospheric conditions and lesser contribution from RFI flagging. All these aspects are currently being explored by the LOFAR EoR KSP team.



# Chapter 3

## Machine Learning for extracting the 21-cm Signal: Application

*Always look where the action isn't, because that's where the action is.*

---

- Richard Osman

### 3.1 Overview

In this chapter, I use the VAE-based kernel developed in [Acharya et al. \(2024b\)](#) and described in Chapter 2 and apply it with Gaussian Process Regression on 141 hours ( $\approx 10$  nights) of data with LOFAR at  $z=9.1$ , whose upper limits with analytical covariance kernels was reported in [Mertens et al. \(2020\)](#). This work was published in [Acharya et al. \(2024d\)](#) and has been adapted to be consistent with the rest of the thesis.

In Section 3.2, I briefly highlight the covariance kernels used for the different components of the dataset; in Section 3.3, I present two cases of recovery, and compare the obtained upper limits to those presented in [M20](#). In Section 3.4, I discuss the quality of my results by comparing the residuals on applying my model to the data, and also the limitations of the method. Finally, in Section 3.5, I give my conclusions.

### 3.2 Methodology

Gaussian Process Regression (GPR; [Rasmussen & Williams, 2006](#); [Aigrain & Foreman-Mackey, 2023](#)) has been used to model radio data in frequency space as noisy observations of the form  $\mathbf{y} = \mathbf{f}(\mathbf{x}) + \epsilon$ , with  $\epsilon$  Gaussian noise vector having variance  $\sigma_{\text{noise}}^2$  (see e.g. [A24](#), [M20](#) and [Munshi et al., 2024](#)). The vector  $\mathbf{f}(\mathbf{x})$  can be split into the 21-cm signal ( $\mathbf{f}_{21}$ ) and foregrounds. To the

latter contribute an intrinsic sky component ( $\mathbf{f}_{\text{sky}}$ ) and mode-mixing contaminants ( $\mathbf{f}_{\text{mix}}$ ). Further, [M20](#) identified an additional term for bias correction, defining it as an “excess noise” component ( $\mathbf{f}_{\text{ex}}$ ), which corrects for additional systematic noise after the subtraction of foregrounds. While the exact source of this excess is yet to be identified, various possibilities have been discussed in Section 6.2 of [M20](#), and the LOFAR EoR KSP team is currently working on testing each one of them. The overall datacube  $\mathbf{y}$  can be expressed as:

$$\mathbf{y} = \mathbf{f}_{\text{sky}}(\mathbf{x}) + \mathbf{f}_{\text{mix}}(\mathbf{x}) + \mathbf{f}_{\text{ex}}(\mathbf{x}) + \mathbf{f}_{21}(\mathbf{x}) + \epsilon. \quad (3.1)$$

For a set of points  $\mathbf{x}$  (independent parameters in frequency), for each value  $x$ ,  $f(x)$  given by a Gaussian Process  $\mathcal{GP}(m, \kappa)$  is fully defined by its mean  $m$  and covariance matrix  $\kappa$ . The vector  $\mathbf{f}(\mathbf{x})$  is then fully defined by its mean vector  $\mathbf{m}$  and covariance matrix vector  $\mathbf{K}$ . The joint distribution for all random variables that share the desired covariance properties is a normal distribution vector ( $\mathcal{N}$ ) which can be represented as:

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{N}(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x})). \quad (3.2)$$

Here, the matrix  $\mathbf{K}$  gives the covariance between the function values at any two points, and can be written as  $K_{ij} = \kappa(x_i, x_j, \phi) + \delta_{ij}\sigma_{i,\text{noise}}^2$ , where  $\kappa(x_i, x_j, \phi)$  can be optimised by the choice of hyperparameters represented by  $\phi$ ,  $\delta_{ij}$  is the Kronecker-delta function and  $\sigma_{i,\text{noise}}^2$  is the noise at  $x_j$ .

If I assume that the various components are uncorrelated, I can utilise the additive property of matrices. Thus, the covariance kernels of the right-hand side of Equation 3.1 can be represented as a single covariance kernel  $K$  given as:

$$\mathbf{K} = \mathbf{K}_{\text{sky}} + \mathbf{K}_{\text{mix}} + \mathbf{K}_{\text{noise}} + \mathbf{K}_{\text{ex}} + \mathbf{K}_{21}. \quad (3.3)$$

$\mathbf{K}_{\text{sky}}$ ,  $\mathbf{K}_{\text{mix}}$ , and  $\mathbf{K}_{\text{ex}}$  are modelled using the best-fit Matern-class functions ([Stein, 1999](#)) as done by [M20](#) and [A24](#):

$$k_{\text{Matern}}(r) = \sigma^2 \frac{2^{1-\eta}}{\Gamma(\eta)} \left( \frac{\sqrt{2\eta}r}{l} \right)^\eta \kappa_\eta \left( \frac{\sqrt{2\eta}r}{l} \right), \quad (3.4)$$

with  $\eta_{\text{sky}} = +\infty$ ,  $\eta_{\text{mix}} = 3/2$  and  $\eta_{\text{ex}} = 5/2$ . Additionally,  $l$  is the coherence-scale hyperparameter, with its associated variance given by  $\sigma^2$ ,  $r$  is the absolute difference between the frequencies of two sub-bands,  $\kappa_\eta$  is the modified Bessel function of the second kind, and  $\Gamma$  is the Gamma-function.  $\eta_{\text{sky}}$ ,  $\eta_{\text{mix}}$ , and  $\eta_{\text{ex}}$  were obtained by [M20](#) by assuming different values of the hyperparameter  $\eta$ , and finding the one that maximised the marginal likelihood (or the “evidence”). This was done by calculating the analytical integral over  $\mathbf{f}$ , which is the log-marginal-likelihood (LML, see Section 2.3 in [Mertens et al. \(2018\)](#)). While [M20](#) used a gradient-descent-based optimization algorithm for maximising the LML, [A24](#), instead, introduced an MCMC sampling based approach ([Foreman-Mackey et al., 2013](#)) to estimate the hyperparameters by sampling their posterior distributions. This additionally provides a measure of the uncertainty on the hyperparameters.

The noise is modelled based on the Stokes-V visibility difference, according to the methodology laid out in [M20](#). Further, for the 21-cm signal I use the VAE-based kernel at  $z = 9.16$  built

Table 3.1: Comparison of hyperparameters obtained in [Mertens et al. \(2020\)](#) and in this work. From left to right the columns refer to the hyperparameters used, the results from [M20](#), the results obtained in this work, and the difference between the two. As the covariance kernels used for the 21-cm signal component are not the same, I just provide the values of the hyperparameters obtained in each case. Lastly,  $\sigma_{\text{noise}}^2$  is  $74 \times 10^3 \text{ mK}^2$  (see [M20](#)).

Parameter	M20	This work	Difference
$l_{\text{sky}}$	$47.5^{+3.1}_{-2.8}$	$38.6^{+2.6}_{-2.6}$	$2.3\sigma$
$\sigma_{\text{sky}}^2 / \sigma_{\text{noise}}^2$	$611^{+22}_{-19}$	$530 \pm 18$	$3.0\sigma$
$l_{\text{mix}}$	$2.97^{+0.09}_{-0.08}$	$3.05^{+0.08}_{-0.08}$	$0.7\sigma$
$\sigma_{\text{mix}}^2 / \sigma_{\text{noise}}^2$	$50.4^{+2.1}_{-1.9}$	$48.8^{+2.1}_{-2.0}$	$0.6\sigma$
$l_{\text{ex}}$	$0.26^{+0.01}_{-0.01}$	$0.32^{+0.01}_{-0.01}$	$4.2\sigma$
$\sigma_{\text{ex}}^2 / \sigma_{\text{noise}}^2$	$2.18^{+0.09}_{-0.14}$	$2.82^{+0.10}_{-0.10}$	$4.2\sigma$
$f_{21}$	$\eta_{21} = 1/2$	$x_1 = 0.49^{+1.00}_{-1.00}$ ,	–
	$l_{21} > 0.73$	$x_2 = -0.34^{+0.97}_{-0.97}$	
	$\frac{\sigma_{21}^2}{\sigma_{\text{noise}}^2} < 0.77$	$\frac{\sigma_{21}^2}{\sigma_{\text{noise}}^2} = 0.01^{+0.09}_{-0.01}$	

by [A24](#) by training on GRIZZLY simulations. This VAE-based kernel is fully defined using two hyperparameters  $x_1$  and  $x_2$  with uninformed flat priors in linear space in the range  $[-10, 10]$ , and an associated variance. The kernel has been trained on a training set of  $\approx 1500$  simulations with 4 independently variable parameters. While more hyperparameters could be employed, they do not provide any significant improvement, and thus I do not use them to avoid overfitting. [A24](#) also used an additional testing set of  $\approx 150$  simulations and found a recovery error of  $\lesssim 1\%$  for wave-modes of  $k < 0.43 h\text{Mpc}^{-1}$ . In this work, the performance of the VAE-based kernel is compared against the results obtained in [M20](#) when using a Matern-class function with  $\eta_{21} = 1/2$ .

### 3.3 Results

Following [A24](#), I recover the hyperparameters of the various signal components by applying GPR to the data using an MCMC approach ([Foreman-Mackey et al., 2013](#)). I adopt the same broad flat priors for the variances, and uniform priors for the coherence-scale parameters (see Table 1 and Section 2.5 of [A24](#)). The recovered values are listed in Table 3.1, where I compare them to those obtained by [M20](#). I find that the hyperparameters for the  $f_{\text{sky}}$  and  $f_{\text{ex}}$  components differ from the [M20](#) estimates by about  $2\text{--}4\sigma$ , while those for  $f_{\text{mix}}$  are in good agreement. This is expected with a more accurate 21-cm kernel, as it reduces the strength of the correlation of hyperparameters of the different components. This allows for better characterisation of the different components in the overall data. I discuss the details below after analysing each component’s recovered power spectra.

To have a better understanding of the results quantified above, in Figure 3.1, I show the recovered upper limits on the power spectra in two cases. In the first one (top panel), I present the excess noise component (dashed-dotted green line) and the 21-cm signal (blue dashes with downward arrows) separately, as done for the mock datasets in A24, and refer to this as the *individual upper limits case*. I additionally show the  $2\text{-}\sigma$  confidence interval on the recovery of the VAE-based kernel, the overall noise (yellow dashed line), and the  $2\text{-}\sigma$  error on the noise (brown dotted line). I compare my recovery with the results from M20 shown with magenta crosses. I note that while the upper limits are higher than the  $2\text{-}\sigma$  error on the noise, the values approach similar values at the higher  $k$  bins (see values listed in Table 3.2). Further, it is still possible that some fraction of the 21-cm signal may be assigned to the excess noise, and vice-versa. Thus, I also show a second case (bottom panel), which takes a more conservative approach of including the excess noise in the upper limits along with the 21-cm signal (solid blue line), as done in M20. I refer to this as the *joint upper limits case*.

I note that, in the individual case, the upper limits on the 21-cm signal power spectrum are significantly deeper than those of the combined power spectrum of the excess noise and 21-cm signal found by M20.

Because I depict just the extracted 21-cm signal component, I assume that the bias correction for the excess noise is accurate, with minimal loss of the 21-cm signal. This is an idealised scenario, where the excess noise component is accurately described by the covariance kernel used for it. To verify this, I need a better understanding of the source of the excess noise component. Nevertheless, I report results with my current best choice of the kernel as identified by M20, to test the extent of improvement possible with the inclusion of the VAE-based kernel for the 21-cm signal. However, in Figure 3.1, the  $2\text{-}\sigma$  confidence interval on the lower side extends to  $< 1 \text{ mK}^2$ , and thus is effectively zero. Thus, this confirms that this is just an upper limit and not a detection.

When comparing the joint upper limits cases, I find marginally higher values than those found by M20 for  $k \lesssim 0.2 h \text{ cMpc}^{-1}$ , and marginally lower for  $0.2 h \text{ cMpc}^{-1} \lesssim k \lesssim 0.5 h \text{ cMpc}^{-1}$ . This is caused by the power of the excess noise component being mildly higher in the lower  $k$ -bins, and mildly lower in the higher  $k$ -bins, compared to the results of M20. This shows up as a  $4.2\sigma$  difference of the hyperparameters  $l_{\text{ex}}$  and  $\sigma_{\text{ex}}^2/\sigma_{\text{noise}}^2$  from the results of M20. A closer inspection shows that this occurs because a small fraction of the intrinsic sky component is assigned to the excess noise component at lower  $k$ -bins, and vice-versa for the higher  $k$ -bins, due to the usage of a more accurate 21-cm kernel. While the difference in the recovered power spectrum for the foregrounds is not significant, the higher contributions at small scales for the excess noise, could allow a better characterisation of the same.

The exact values of the recovered median power spectrum of the 21-cm signal and their upper limits in each  $k$ -bin are listed in Table 3.2 for the joint and individual upper limits case discussed in this work, as well by M20. I also show the  $2\text{-}\sigma$  error of the noise (which represents the maximum sensitivity achievable) to compare against the obtained upper limits. I can summarize the results as follows:

- The excess noise is confirmed to be the dominant signal component after sky-model and

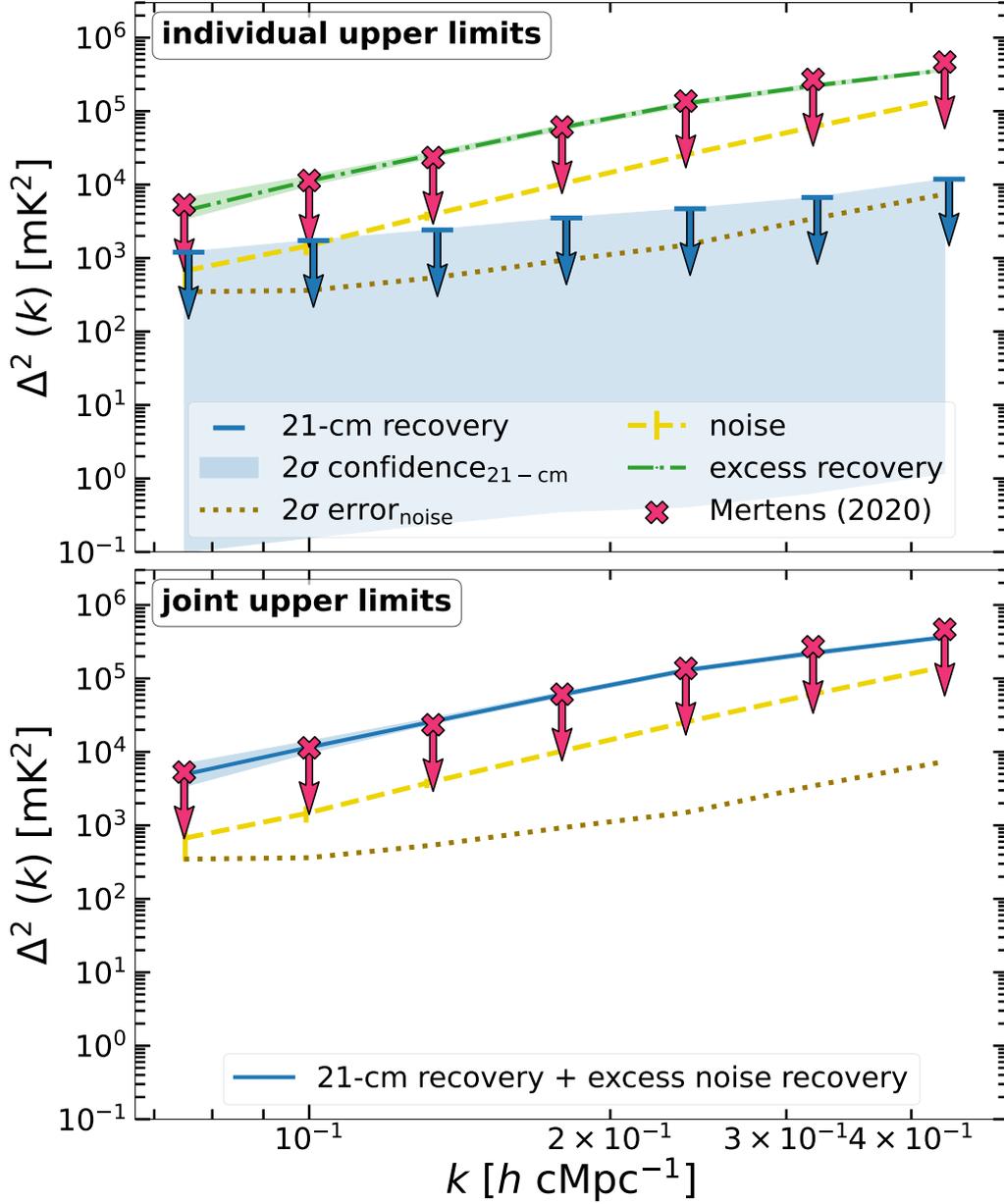


Figure 3.1: Upper limits on the power spectrum of the 21-cm signal from 10 nights of LOFAR observational data at  $z \approx 9.1$ . **Top:** Individual upper limits case. The upper limits on  $\Delta_{21}^2$  (blue dashes with downward arrows) are obtained using the VAE-based kernel separately from the excess noise (dashed-dotted green line). The  $2\text{-}\sigma$  confidence of the recovery by the VAE-based kernel is shaded in blue, and noted to effectively extend down to zero. I also plot the upper limits obtained by M20 (crosses with downward arrows), the noise (dashed yellow line) and the  $2\text{-}\sigma$  error on the noise (dotted brown). **Bottom:** Joint upper limits case. The upper limits on the power spectrum of the 21-cm signal and excess noise are recovered together (solid blue line and shaded region), as done in M20. The other lines and symbols have the same meaning as in the top panel.

Table 3.2: Comparing the results of [Mertens et al. \(2020\)](#) with those of my joint and individual upper limits of the excess noise component and 21-cm signal power spectra. The first column refers to the  $k$  bins, and the following to the recovered 21-cm signal power ( $\Delta_{21}^2$ ) and its corresponding 2- $\sigma$  upper limit ( $\Delta_{21, \text{UL}}^2$ ) for [M20](#), the joint upper limits (bottom panel in [Figure 3.1](#)) and the individual upper limits (top panel in [Figure 3.1](#)) applied in this work. The last column shows the 2- $\sigma$  error on the noise (2- $\sigma$  error<sub>noise</sub>).

$k$	<a href="#">M20</a> ( $f_{\text{ex}} + f_{21}$ ) $\Delta_{21}^2$ (mK <sup>2</sup> )	$\Delta_{21, \text{UL}}^2$ (mK <sup>2</sup> )	joint upper limits ( $f_{\text{ex}} + f_{21}$ ) $\Delta_{21}^2$ (mK <sup>2</sup> )	$\Delta_{21, \text{UL}}^2$ (mK <sup>2</sup> )	individual upper limits ( $f_{21}$ ) $\Delta_{21}^2$ (mK <sup>2</sup> )	$\Delta_{21, \text{UL}}^2$ (mK <sup>2</sup> )	2- $\sigma$ error <sub>noise</sub> $2\Delta_{\text{noise, err}}^2$ (mK <sup>2</sup> )
0.075	(58.96) <sup>2</sup>	(72.86) <sup>2</sup>	(71.33) <sup>2</sup>	(79.68) <sup>2</sup>	(6.58) <sup>2</sup>	(24.55) <sup>2</sup>	(18.63) <sup>2</sup>
0.100	(95.21) <sup>2</sup>	(106.65) <sup>2</sup>	(107.05) <sup>2</sup>	(115.77) <sup>2</sup>	(8.56) <sup>2</sup>	(29.43) <sup>2</sup>	(19.05) <sup>2</sup>
0.133	(142.17) <sup>2</sup>	(153.00) <sup>2</sup>	(158.65) <sup>2</sup>	(167.76) <sup>2</sup>	(10.48) <sup>2</sup>	(37.29) <sup>2</sup>	(23.22) <sup>2</sup>
0.179	(235.80) <sup>2</sup>	(246.92) <sup>2</sup>	(245.42) <sup>2</sup>	(261.40) <sup>2</sup>	(14.40) <sup>2</sup>	(43.92) <sup>2</sup>	(30.47) <sup>2</sup>
0.238	(358.95) <sup>2</sup>	(370.18) <sup>2</sup>	(360.08) <sup>2</sup>	(372.28) <sup>2</sup>	(16.29) <sup>2</sup>	(56.32) <sup>2</sup>	(38.70) <sup>2</sup>
0.319	(505.26) <sup>2</sup>	(520.33) <sup>2</sup>	(472.49) <sup>2</sup>	(490.22) <sup>2</sup>	(19.95) <sup>2</sup>	(73.79) <sup>2</sup>	(58.72) <sup>2</sup>
0.432	(664.23) <sup>2</sup>	(683.20) <sup>2</sup>	(601.37) <sup>2</sup>	(626.81) <sup>2</sup>	(26.61) <sup>2</sup>	(99.19) <sup>2</sup>	(86.01) <sup>2</sup>

residual foregrounds subtraction, with its recovered power spectrum being about an order of magnitude stronger than the noise.

- While the hyperparameters for the excess noise component differ by  $4.2\sigma$  from the results of M20, the recovered power spectrum and  $2\text{-}\sigma$  upper limits show only a minor deviation.
- In the case of individual upper limits, the upper limits of the 21-cm signal power spectrum lie above the  $2\text{-}\sigma$  error on the noise, and thus are not noise-dominated. However, this is not a detection, because the lower limit on the recovery by the VAE-based kernel is consistent with zero.

I conclude that with these data, performing a bias correction for the excess noise provides an individual upper limit on the 21-cm signal power spectrum that is higher than the  $2\text{-}\sigma$  error on the noise. However, the upper limits approach the  $2\text{-}\sigma$  error on the noise at higher  $k$  bins. Thus, lowering the noise threshold with longer observations could allow a greater separation between the upper limits and the noise threshold, and eventually also a detection, provided no correlation between the excess noise and the 21-cm signal is present. For this, the LOFAR EoR KSP team has  $\approx 100$  nights of data to analyse and is currently in the process of doing so. I discuss the correlations of the different components and the limitations of the overall model in the subsequent section.

## 3.4 Discussion

The individual upper limits are greater than the  $2\text{-}\sigma$  error on the noise, and thus is not noise dominated. However, it is still possible that the VAE-based kernels provide an incomplete model of the data. In this case, I would be biased towards cases where the model is enough for the 21-cm signal, and the remainder of the data is modelled with the excess and foregrounds components as defined above. For example, better characterisation of the excess noise may indeed lead to broader models for it. In my current model, what I can explore however, is the degeneracy between the VAE-based kernel and the excess noise kernel. A decrease in the correlation between them would indicate that at least the models I use for them are not degenerate. For example, the difference in the intrinsic sky and excess noise component hyperparameters with respect to M20 could be due to a decrease in correlation of the hyperparameters for these components by using the VAE-based kernel for the 21-cm signal. To have a clearer picture of this, it is important to investigate whether the hyperparameters of the different components in my overall model are correlated and whether the overall model is a complete picture of the data.

To assess this, in Figure 3.2, I show the corner plot of the hyperparameters, where the purple contours are the 68%, 95% and 99.7% confidence intervals. I see that hyperparameters of each component are largely uncorrelated with those of other components. In particular, I note that the 21-cm signal hyperparameters are completely uncorrelated from those of the excess noise. This provides greater confidence in the  $2\text{-}\sigma$  confidence intervals obtained in the individual upper limits case.

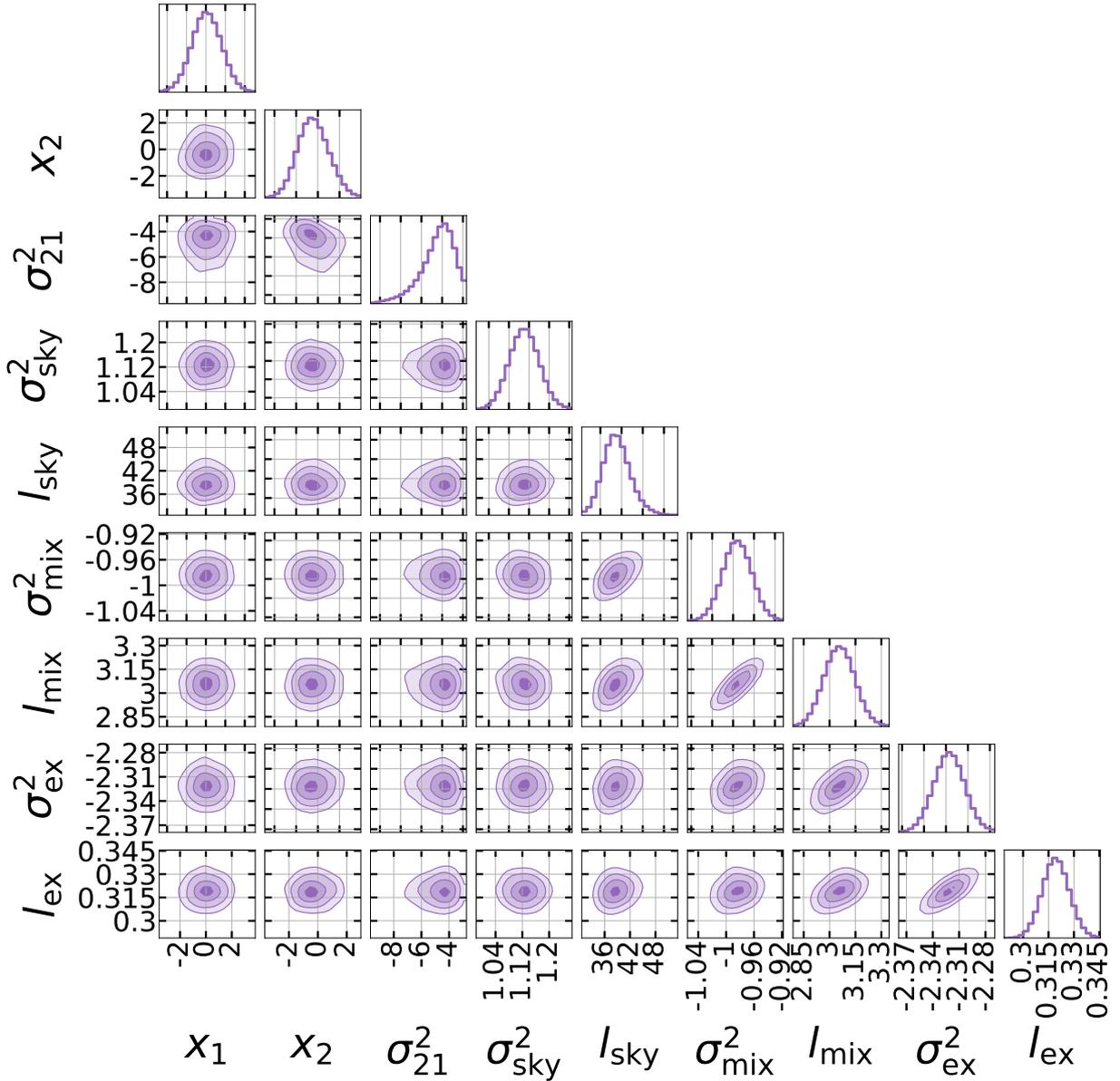


Figure 3.2: Posterior probability distributions of the GPR model hyperparameters used:  $x_1$ ,  $x_2$ , the variances  $\sigma_{21}^2$ ,  $\sigma_{\text{sky}}^2$ ,  $\sigma_{\text{mix}}^2$  and  $\sigma_{\text{ex}}^2$ , and the coherence-scales  $l_{\text{sky}}$ ,  $l_{\text{mix}}$  and  $l_{\text{ex}}$ . The purple contours show the 68%, 95% and 99.7% confidence intervals and the diagonal plots refer to the individual posterior distributions of each of the hyperparameters listed. Note that the hyperparameters of each component are largely uncorrelated with those of other components.

Furthermore, to assess the performance of my model, I plot the residual obtained by subtracting the data cube of the model from the observational data cube in Figure 3.3, and compare its power spectrum (black solid) to that of the noise data cube (yellow dashed). I find excellent agreement between them within the  $2\text{-}\sigma$  error on the noise. To numerically quantify this, I calculate the

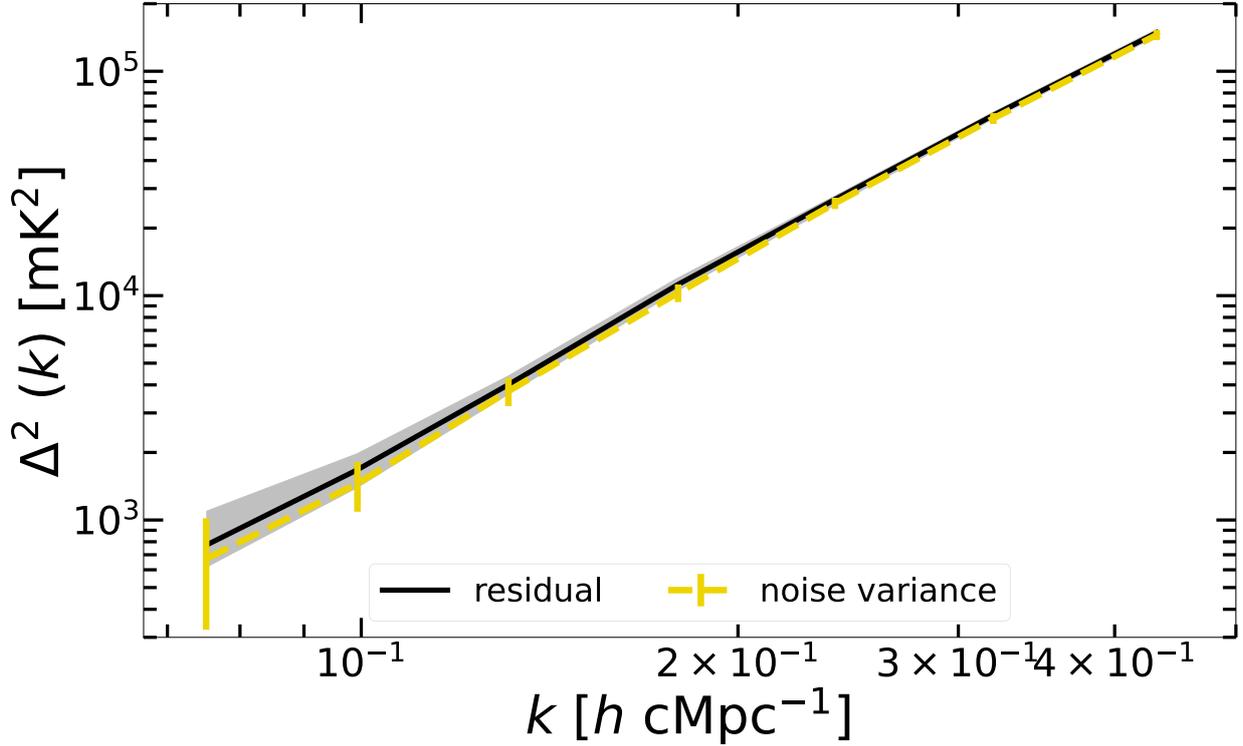


Figure 3.3: Residual (black solid line) obtained by subtracting the full GPR model from the data, and its corresponding  $2\text{-}\sigma$  confidence interval (grey shaded region). The noise is shown as a yellow dashed line with  $2\text{-}\sigma$  error bars. I note an excellent agreement between them.

$\chi^2$  of the residual versus the noise power spectrum, and find this to be  $\approx 6.12$ . With 7  $k$  bins, this gives a reduced  $\chi^2 \approx 0.87$ , thus confirming the performance of the model. However, it is still possible that some minor leakage can occur between the different signal components. Thus, better modelling of the excess noise remains necessary to improve the bias correction as more data is added. Furthermore, for an eventual detection, reduction of the noise component through longer observations is also needed.

### 3.5 Conclusions

To improve the modelling of the 21-cm signal covariance kernel, [Mertens et al. \(2024\)](#) introduced a Machine Learning based method that uses a Variational Auto-Encoder (VAE) based algorithm. This was trained on GRIZZLY simulations ([Ghara et al., 2015, 2018, 2020](#)) of the 21-cm signal and tested against mock datasets by [Acharya et al. \(2024b\)](#). In this work, I use the VAE-based kernel trained at  $z = 9.16$  to obtain an updated upper limit on the 21-cm signal from 141 hours of LOFAR data ( $\approx 10$  nights) at  $z \approx 9.1$ . I compare my results to past efforts, which used analytic functions instead of Machine Learning ([Mertens et al., 2020](#)), and find that they are consistent with errors, although the new upper limits are slightly increased at  $k \lesssim 0.2 h \text{ cMpc}^{-1}$ , and mildly decreased

for  $0.2 h \text{ cMpc}^{-1} \lesssim k \lesssim 0.5 h \text{ cMpc}^{-1}$ . Overall, I report a  $2\text{-}\sigma$  upper limit of  $\Delta_{21}^2 < (80)^2 \text{ mK}^2$  at  $k = 0.075 h \text{ Mpc}^{-1}$  in the conservative case, when jointly recovering the 21-cm signal and excess noise components.

I also investigate the case in which the upper limits of the 21-cm signal are evaluated including a bias correction for the excess noise component. In this case, I report a  $2\text{-}\sigma$  upper limit of  $\Delta_{21}^2 < (25)^2 \text{ mK}^2$  at  $k = 0.075 h \text{ Mpc}^{-1}$ . I confirm that the VAE kernel hyperparameters are uncorrelated from the excess component ones, thus providing confidence for the estimated upper limits. This is strengthened by the fact that the residuals obtained by subtracting the full model from the data are consistent with the noise with  $0.87\sigma$ . Although the bias correction is promising, I still caution against using the bias-corrected upper limits, as better modelling of the excess noise remains necessary to improve confidence on its separation from the 21-cm signal, as the data improves and the signal to noise increases. In future work, the LOFAR EoR KSP team will be improving the characterisation of the excess noise component and its corresponding covariance kernel, and assess whether an excess noise bias correction can reliably be applied as this work suggests it can. As demonstrated here, this approach would substantially reduce the current upper limits on the 21-cm signal, although this should still be considered as an upper limit and not a detection for the given data, as it is already reaching the sensitivity limit due to the noise, given by the  $2\text{-}\sigma$  error on the noise. This improvement could offer enhanced astrophysical constraints compared to [Ghara et al. \(2020\)](#), allowing for the rejection of a significantly greater number of cold IGM scenarios and achieving stricter constraints on the sources of X-ray heating. This shall be carried out in future work. Additionally, the team will make use of VAE-based covariance kernels as developed in [Acharya et al. \(2024b\)](#) across multiple redshifts with data from the LOFAR telescope.

# Chapter 4

## Improved simulations: boosting simulation volumes

*When stupid ideas work, they become genius ideas.*

---

- Andy Weir

### 4.1 Overview

In this chapter, I move on to dealing with improvements in numerical simulations of the 21-cm signal observables, including the skewness, power spectrum, bispectrum, and bubble size distribution. In particular, I focus on the improvement in terms of the effective volume of simulations possible by implementing the Fixed & Paired (F&P) method. I test this with Radiation-Magneto-Hydrodynamic (RMHD) simulations to carefully account for radiative transfer and baryonic hydrodynamics, which in turn impact galaxy properties. Using these simulations ensures we account for the impact of small-scale properties (below one comoving Megaparsec) on the 21-cm signal. The contents of this chapter were published in [Acharya et al. \(2024c\)](#) and have been adapted to be consistent with the rest of the thesis.

The simulations used are detailed in Section 4.2. I discuss the results of various summary statistics in Section 4.3 and the improvement on using the F&P approach in Section 4.4, while I give my conclusions in Section 4.5.

## 4.2 Methodology

### 4.2.1 The 21-cm signal

The brightness temperature fluctuations of the 21-cm signal are given relative to the CMB temperature for any patch of the IGM as (see [Furlanetto et al., 2006](#)):

$$\delta T_b = 27 x_{\text{HI}} (1 + \delta_B) \left( 1 - \frac{T_{\text{CMB}}}{T_S} \right) \times \left[ \left( \frac{\Omega_B h^2}{0.023} \right) \left( \frac{0.15}{\Omega_m h^2} \frac{1+z}{10} \right)^{1/2} \right] \text{mK} \quad (4.1)$$

where  $x_{\text{HI}}$  is the fraction of neutral hydrogen,  $\delta_B$  is the fractional overdensity of baryons,  $T_S$  is the hydrogen spin temperature,  $T_{\text{CMB}}$  is the temperature of the CMB photons at redshift  $z$ ,  $\Omega_m$  is the total matter density,  $\Omega_B$  is the baryon density, and  $h$  is the Hubble constant in units of  $100 \text{ km s}^{-1} \text{ cMpc}^{-1}$ . Above I assume that the spin temperature is coupled to the gas temperature, i.e.,  $T_S = T_{\text{gas}}$ , where  $T_{\text{gas}}$  is the gas temperature self-consistently calculated in the simulations.

In the next section, I introduce the simulation used to generate mock differential brightness temperature ( $\delta T_b$ ) maps.

### 4.2.2 Simulations

My setup is inspired by the THESAN simulations. I run a suite of radiation-magneto-hydrodynamic simulations that utilize the moving-mesh hydrodynamics code AREPO ([Springel, 2010](#); [Weinberger et al., 2020](#)), which includes a gravity solver based on the hybrid Tree-PM method ([Barnes & Hut, 1986](#)), a quasi-Lagrangian Godunov method ([Godunov & Bohachevsky, 1959](#)) based hydrodynamics solver implemented on an unstructured Voronoi mesh grid ([Vogelsberger et al., 2020](#)) and the radiative transfer extension AREPO-RT ([Kannan et al., 2019](#)) for a self-consistent treatment of ionizing radiation. I include the production and propagation of ionizing photons in three energy bins relevant for hydrogen and helium photoionization ( $[13.6, 24.6, 54.4, \infty]$  eV). Further, I employ a non-equilibrium thermochemistry solver to model the coupling of radiation fields to gas. For the luminosity and spectral energy density of stars, I use a complex function of age and metallicity calculated using the Binary Population and Spectral Synthesis models (BPASS v2.2.1; [Eldridge et al., 2017](#)), modeling the unresolved birth cloud with a uniform escape fraction of  $f_{\text{esc}} = 0.37$ . I note that I do not perform a recalibration of this parameter with respect to THESAN since (i) it mostly impacts the final phases of reionization, while I am interested in the initial ones, and (ii) my goal is to compare methods for initial condition generation, so a slightly-inaccurate reionization history is not expected to affect at all my results. For further details on the THESAN simulations, see [Kannan et al. \(2019, 2022\)](#).

All my simulations have a box-size of  $L = 95.5 \text{ cMpc}$ , and  $N = 2 \times 525^3$  particles, giving a dark matter and baryonic particle mass of  $m_{\text{DM}} = 2.0 \times 10^8 M_{\odot}$  and  $m_{\text{gas}} = 4.7 \times 10^6 M_{\odot}$ , respectively. The gravitational softening length for the star and dark matter particles is set to  $6.0 \text{ ckpc}$ , which is also the minimum value for the adaptively softened gas cells according to

cell radius. The cosmological parameters are taken from [Planck Collaboration et al. \(2016\)](#) as  $h = 0.6774$ ,  $\Omega_m = 0.3089$ ,  $\Omega_\Lambda = 0.6911$ ,  $\Omega_b = 0.0486$ ,  $\sigma_8 = 0.8159$  and  $n_s = 0.9667$ . In the initial conditions the gas is assumed to follow the DM distribution perfectly, with a primordial hydrogen and helium fractions of  $X = 0.76$  and  $Y = 0.24$ , respectively. I start the simulations from  $z_{\text{ini}} = 49$ , generating 54 snapshots between  $z = 20$  to  $7$ .

Due to the chosen particle number, my dark matter halo masses are  $\gtrsim 10^{10} M_\odot$ , leading to reionization being driven by relatively massive galaxies. The resulting topology of reionization therefore somewhat differs at small scales from the one that would be produced including lower mass, more abundant, galaxies. This will impact the 21-cm signal and slow down the reionization process (which is driven by  $M_{\text{star}} \sim 10^7 M_\odot$  galaxies as shown in [Rosdahl et al., 2022](#); [Yeh et al., 2023](#); [Kostyuk et al., 2023](#)). However, these effects will equally affect all simulations, so will be factored out by my comparative analysis. A higher mass resolution would be immensely computationally expensive to run the number of simulations necessary to perform the statistical analyses discussed in subsequent sections of this work. As discussed earlier, the necessity of simulating large physical scales in order to compare with upcoming surveys prevents me from employing smaller (and thus computationally cheaper) boxes.

Once a significant fraction of neutral hydrogen is reionized, the improvement of using the F&P approach to study the 21-cm signal is expected to saturate. This is depicted in Figures 4.8 and B.1, showing the improvement factor calculated according to the methodology of Section 4.4.1. Thus, I only run the simulations down to  $z_{\text{fin}} = 7$ , which corresponds to  $\langle x_{\text{HI}} \rangle \approx 0.8$ .

Lastly, as done in [Kannan et al. \(2022\)](#), I also save Cartesian data output at a higher cadence (243 outputs between  $z = 16$  to  $z = 7$ ) by gridding the simulation data onto a regular Cartesian grid employing a first order Particle-In-Cell approach. I use a  $256^3$  grid, i.e. each cell is  $\sim 372$  ckpc in size.

#### 4.2.2.1 The Fixed & Paired approach

The F&P approach was first proposed by [Angulo & Pontzen \(2016\)](#) and is based on the creation of a special pair of initial conditions (ICs) that, when employed together, significantly reduce the impact of cosmic variance. In the traditional approach to ICs creation, an initially uniform and isotropic distribution of tracers particles is perturbed (following e.g. [Zel'dovich, 1970](#)) to induce density perturbations:

$$\delta(\mathbf{k}, z_{\text{ini}}) = \sqrt{P(\mathbf{k}, z_{\text{ini}})} e^{i\theta_{\mathbf{k}}} \quad (4.2)$$

where the phase  $\theta_{\mathbf{k}}$  is sampled from a flat distribution in the range  $[0, 2\pi]$ , and the power spectrum amplitude  $P(\mathbf{k}, z_{\text{ini}})$  is sampled from a Gaussian distribution centered on its expectation value  $E[P(\hat{\mathbf{k}}, z_{\text{ini}})]$  (necessary to produce a Gaussian random field).

In the F&P approach, the power spectrum modes are fixed to their expectation values. This produces the *fixed* initial conditions, while the *paired* one is obtained by reversing the phase associated to each particle displacement. By combining simulations run with these two sets of ICs, the variance-induced fluctuations (which in traditional ICs mainly affect the large-scale

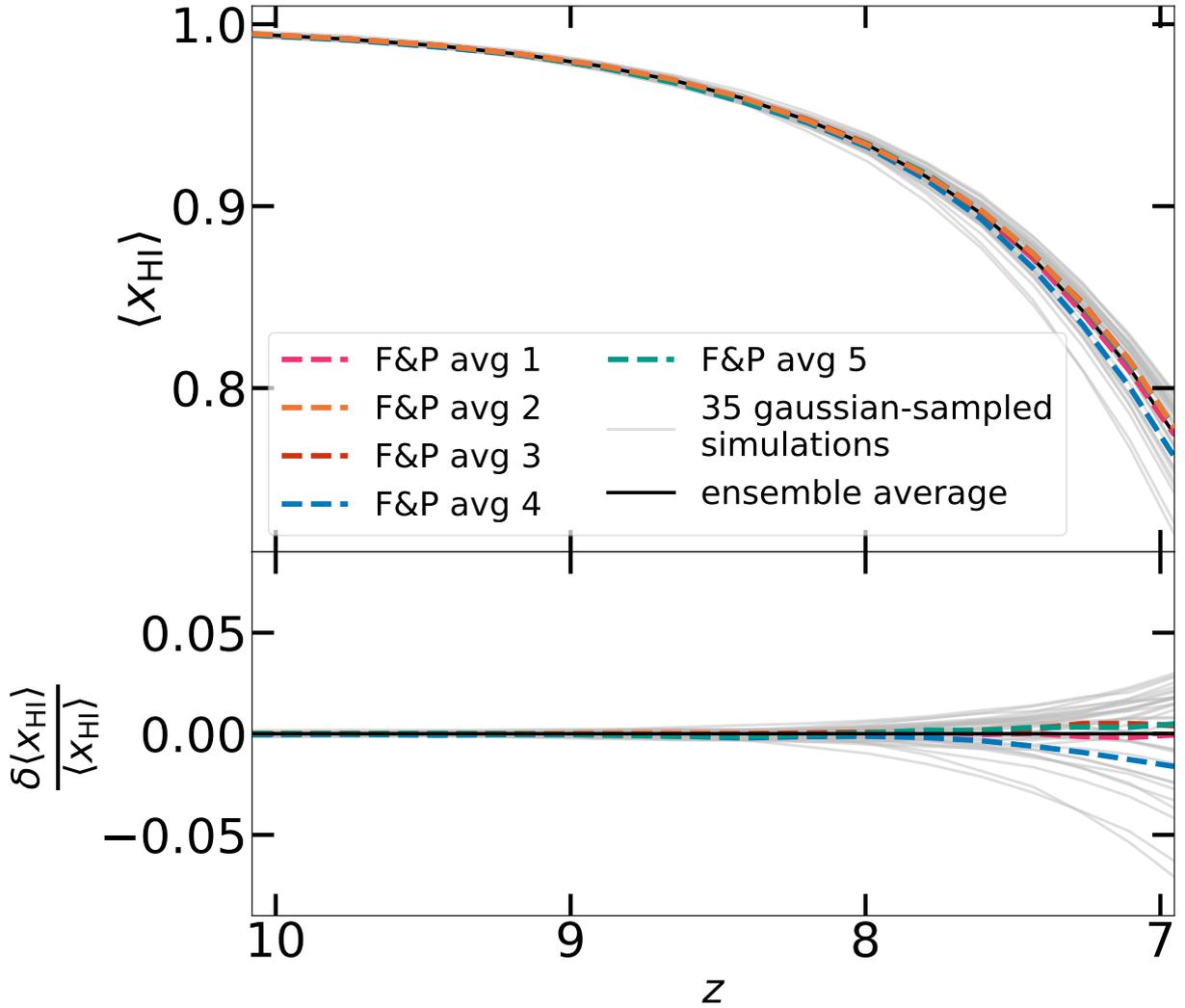


Figure 4.1: Evolution of the volume averaged neutral hydrogen fraction versus redshift for GIC simulations (grey solid), their ensemble average (black solid) and five F&P averages (orange, magenta, red, blue and green dashed). I note that at  $z=7$   $\langle x_{\text{HI}} \rangle$  ranges between 0.72 and 0.82 for the GIC simulations.

modes, where the sampling of the power spectrum amplitude is scarce and therefore more susceptible to deviations from its expectation value) are suppressed up to the fourth perturbative order (Angulo & Pontzen, 2016). Additionally, despite breaking the Gaussianity of the generated field, this approach does not induce unwanted features, as shown by e.g., Angulo & Pontzen (2016) and Chartier et al. (2021) for the matter power spectrum, bispectrum and the halo mass function, by Anderson et al. (2019) for Lyman- $\alpha$  forest power spectra, and for a variety of other quantities by Villaescusa-Navarro et al. (2018) and Klypin et al. (2020).

In this work, I generate 5 such pairs of F&P simulations, in order to explore the effects of the

random sampling of density perturbations. To minimise the effect of randomness, for each pair I fix the seed for the random number generator used for stochastic algorithms like the star formation prescription. Further, I run all the simulations described here on the same machine with the same hardware configurations. As AREPO is coded to be binary identical in such conditions, it allows me to avoid floating point errors building up and biasing my results.

I compare the averages of these 5 F&P pairs against 35 traditional Gaussian-sampled initial conditions based simulations (hereafter referred to as GIC). As an example, in Figure 4.1 I show the reionization histories of the 5 pairs (magenta, orange, red, blue, green dashed), as well as that of the GIC simulations (grey solid). From the figure, I note that the reionization histories of the F&P averages cluster close to the average of those of the GIC simulations. This is a consequence of the fact that the F&P method ensures that the F&P averages closely match the halo mass function, while GIC simulations can spuriously have an excess/dearth of very bright sources.

In Figure 4.2, I show maps of  $\delta T_b$  at different redshifts ( $z = 10, 8.3, 7.6, 7$ ; these correspond to  $\langle x_{\text{HI}} \rangle = 0.99, 0.95, 0.9, 0.8$ ) for one of the fixed simulations, its corresponding pair, their average, and the average of two randomly chosen GIC simulations. I choose two GIC simulations at random to provide a visual comparison of results obtained from averaging them as opposed to averaging an F&P pair. Note that due to the phase inversion used for generating the initial conditions of the pair, the regions of high  $\delta T_b$  in the fixed simulation roughly overlaps with regions of low  $\delta T_b$  in the pair. This is evident in their average, which does not have specific regions of high (or low)  $\delta T_b$ , unlike the average of the two GIC simulations. With decreasing redshift, I see such regions beginning to form for the F&P average as well, but a closer analysis of the  $\delta T_b$  summary statistics is necessary to analyse the difference from GIC simulations.

In the next section, I compare the ensemble average of the GIC simulations (taken to be the “true” value) against the F&P average simulations. I show qualitative comparisons between two of the F&P averages and the true value in Section 4.3, and a more quantitative analysis using all five of the averaged simulations in Section 4.4.

## 4.3 Analysis and Results

In this section, I compare the various summary statistics of the 21-cm signal for the GIC simulations and their ensemble average versus the F&P averages. In particular, I focus on the skewness, the power spectrum, and the bispectrum. In Sections 4.3.1, 4.3.2 and 4.3.3 I analyse their behaviour at various redshifts, focusing on  $z = 10, 9, 8$ , and 7. I also compare the ionized bubble size distribution across these redshifts in Section 4.3.4.

### 4.3.1 Skewness

The skewness ( $\tilde{\mu}_{3,b}$ ) is a statistical measure of asymmetry in a distribution, i.e., as the name suggests, it quantifies how skewed a distribution is around its mean value.  $\tilde{\mu} = 0$  indicates a

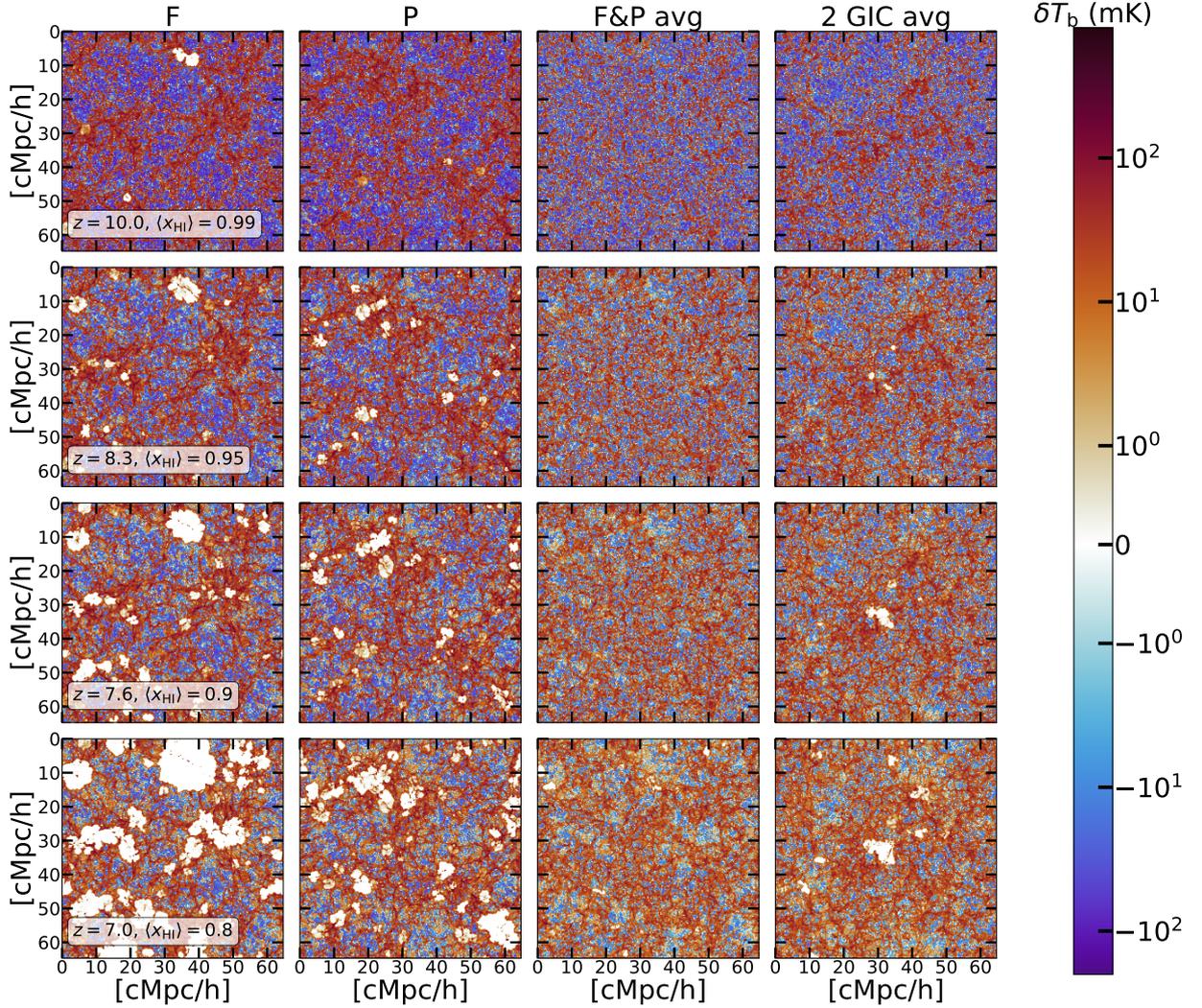


Figure 4.2: Middle slices of the  $\delta T_b$  maps of (from left to right): one of the fixed simulations, its corresponding pair, their average, and the average of 2 GIC simulations, at  $z = 10, 8.3, 7.6, 7$  (with  $\langle x_{\text{HI}} \rangle = 0.99, 0.95, 0.90, 0.80$  respectively). The average of 2 GIC simulations shows clear regions of high (and low)  $\delta T_b$ , unlike the F&P average. With decreasing redshift, such regions begin to form for the F&P average as well, but a closer analysis of the  $\delta T_b$  summary statistics is necessary to analyse the difference from GIC simulations.

symmetric distribution. Here, I use the definition of skewness from [Ma et al. \(2021\)](#), given as:

$$\tilde{\mu}_{3,b} = \frac{\mu_3(\delta T_b)}{\mu_2(\delta T_b)^{3/2}} = \text{E} \left[ \frac{(\delta T_b - \langle \delta T_b \rangle)^3}{\sigma_{\delta T_b}^3} \right] \quad (4.3)$$

where  $\mu_i$  is the  $i$ -th central moment, and  $\text{E}[\cdot]$  is the expectation value at each snapshot where I calculate the skewness. In my case, this would be the volume average.

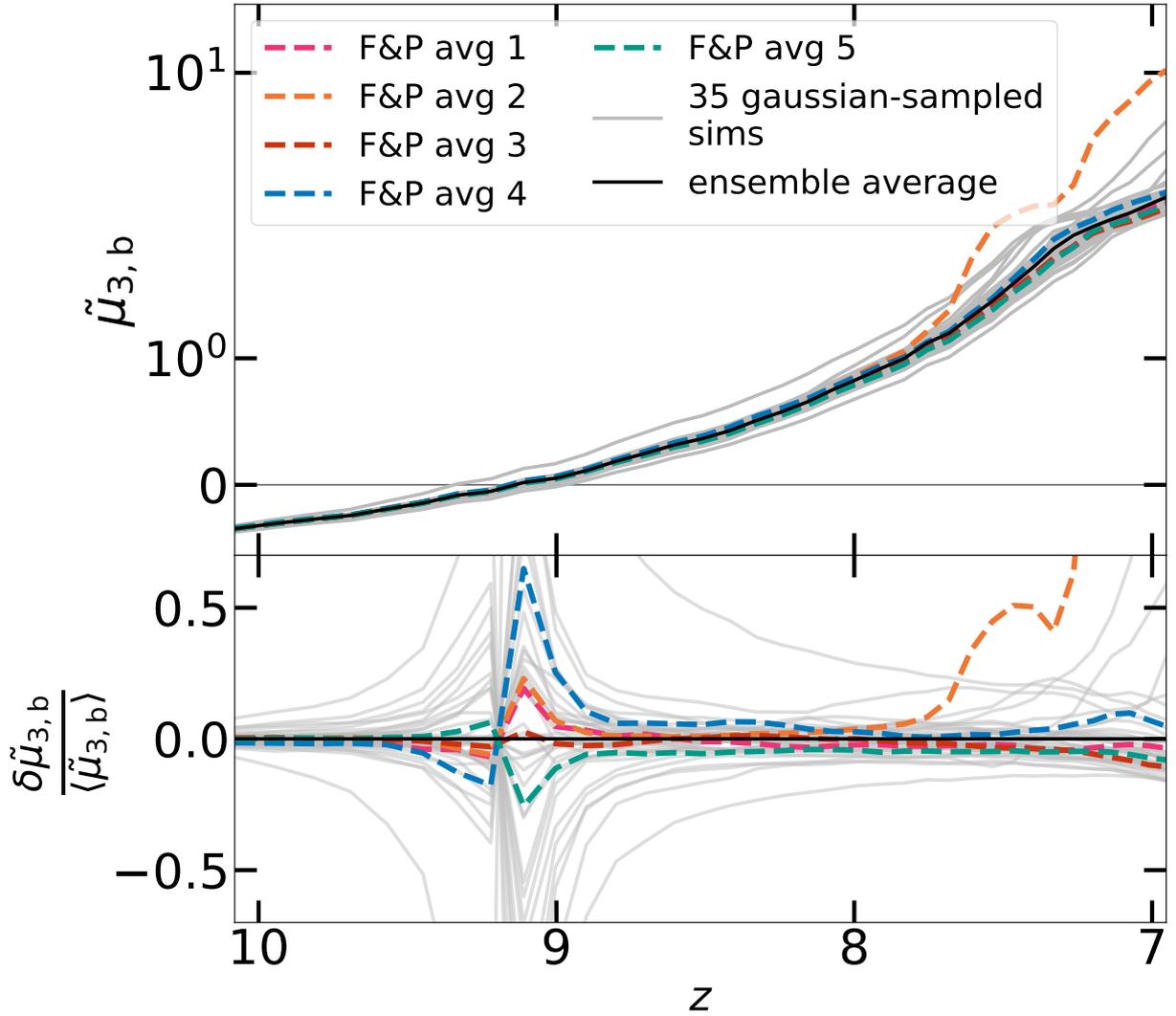


Figure 4.3: Redshift evolution of skewness for 35 GIC simulations (grey solid), their ensemble average (black solid), and the five F&P averages (orange, magenta, red, blue and green; dashed).

In Figure 4.3, I compare the evolution of skewness versus redshift from  $z = 10$  to 7 for the 35 GIC simulations, their ensemble average, and the 5 F&P average simulations. I note that the F&P averages are a good estimate for the true value of the skewness in most cases. For the 2nd F&P average, the deviation is larger at  $z \lesssim 8$ , which is indicative of more inhomogeneity in the  $\delta T_b$ . I find that this anomalous skewness is caused by the fixed simulation in the second pair having a chance association of galaxies with strong black hole feedback and high output of ionizing photons. While such cases are rare, they are physically possible and highlight how the relevance of galactic processes in the production of the 21-cm signal hinders the improvements granted by the F&P approach. I also note that such cases are not completely captured by e.g. the 21cmFAST simulation used in [Giri et al. \(2023\)](#).

This anomalous skewness however, is not expected to cause a major effect on other summary statistics like the power spectrum and bispectrum, because the effect is localised around few simulated galaxies. Thus, not only the scales involved are significantly smaller than the scales I consider in the next sections, but this effect is driven by a statistically-insignificant number of objects. To ensure this, I checked that by replacing in Equation 4.3 the volume average with the median, the anomalous behavior vanishes. To make this more evident, in all following figures concerning summary statistics ( $0.15 \leq k/(h \text{ cMpc}^{-1}) \leq 2.0$ ) I elected to show the F&P pair that most closely follows the ensemble-averaged skewness and the pair that most deviates from it. There is no significant difference between these two.

### 4.3.2 Power Spectrum

The power spectrum is expected to be the first detectable statistic of the 21-cm signal during the EoR, and thus is of particular interest. So analysing the improvement in modelling, if any, with the F&P method is important. The power spectrum of  $\delta T_b$  as defined in Equation 4.1 is given by:

$$P_{21\text{cm}}(\mathbf{k}) = \delta_D(\mathbf{k} + \mathbf{k}') \langle \delta T_b(\mathbf{k}) \delta T_b(\mathbf{k}') \rangle \quad (4.4)$$

where  $\delta_D$  is the Dirac function,  $\delta T_b(\mathbf{k})$  is the differential brightness temperature in Fourier space, and  $\langle \dots \rangle$  is the ensemble average. In this work, I use the normalized form of the power spectrum given by:

$$\Delta_{21\text{cm}}^2 = \frac{k^3}{2\pi^2} \times P_{21\text{cm}}. \quad (4.5)$$

In Figure 4.4 I show the power spectra of the GIC simulations (in grey), their ensemble average (in black), and the F&P averages (magenta and orange) for 2 pairs (purple and green, with the fixed simulation in dashed, and their corresponding pairs in dotted lines) out of the 5 generated in Section 4.2.2.1. While I have analysed results for all of the five generated F&P pairs (see Section 4.4.1 for a quantitative comparison), for the sake of visual clarity I show only two out of the five pairs. I explicitly choose the second F&P pair average (orange) to check if its deviation in skewness leads to any difference in behaviour as compared to one of the other four pairs (magenta).

I note that the F&P averages match the ensemble average across all redshifts. While their deviation from the ensemble average increases with decreasing redshift, it is still less than that of the individual GIC simulations (see lower panels of Figure 4.4). This is understandable as a combination of two effects. First, the F&P approach is expected to yield improvements on the predictions of features that are primarily governed by the large-scale structure (as they are tied to the matter power spectrum), while it does not provide a statistical improvement on the prediction of galaxy properties (Villaescusa-Navarro et al., 2018), such as ionizing photons output, as they are entirely dominated by local physics. Secondly, the period of emergence of ionized regions is one where the 21-cm signal becomes increasingly non-Gaussian. Thus, the information content in the power spectrum is reduced in this period as compared to earlier and later redshifts, which have a homogenized distribution of neutral and ionized hydrogen, respectively. While the contribution

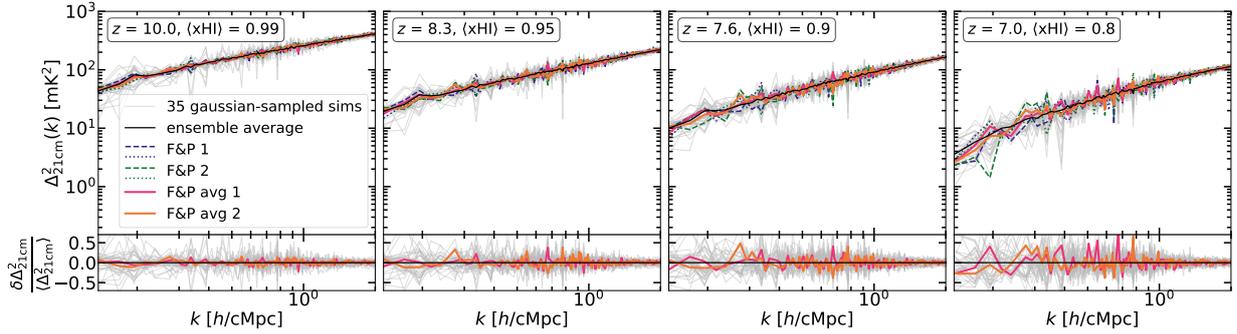


Figure 4.4: **Top:** Power spectra of GIC simulations (grey solid), their ensemble average (black solid), two F&P pairs (purple and green, dashed and dotted) and their averages (magenta and orange, solid) for  $z = 10, 8.3, 7.6,$  and  $7$  from left to right. **Bottom:** The normalised deviation of each random simulation and F&P average from the ensemble average.

from the large-scale structure remains the same, the fluctuations of  $\delta T_b$  are increasingly affected by the presence of ionized gas, which in my simulation is dominated by large and isolated ionized regions. While this mostly affects the power spectrum at small scales, the cumulative fluctuations over large scales can also show up. Thus for the F&P method I observe deviations in the 21-cm signal power spectrum which are larger than those in the matter power spectrum as redshift decreases. However, I note that these deviations are still smaller than those of the individual GIC simulations. Nevertheless, it is necessary to check if an F&P pair average power spectrum is more likely to minimize cosmic variance as opposed to an average of two random GIC simulations. Thus, I quantify the improvement on using the F&P method in Section 4.4.1, where I consider all my F&P pair averages.

While discrepancies between my results and [Giri et al. \(2023\)](#) could possibly be due to the difference in methodologies adopted for comparing the F&P method versus traditional generation of initial conditions, contributions from better handling of galaxy-driven physics in my simulations are also possible. It is however difficult to discern the extent of the effect galaxy-driven physics has on my results. Finally, it is also possible that my ensemble average is biased because it is an average of just 35 simulations. Using significantly larger number of GIC simulations may rule out this issue. However, I note no appreciable difference in the ensemble average once more than 20 GIC simulations have been used, and thus refrain from running additional simulations.

### 4.3.3 Bispectrum

As discussed in the previous section, the non-Gaussianity of the  $\delta T_b$  increases with the growth of ionized regions as redshift decreases. This means that a statistic that focuses only on the Gaussian parts of the signal, like the power spectrum, encapsulates less and less information as I move to lower redshift, into the regime of  $\langle x_{\text{HI}} \rangle \lesssim 0.9$ . Therefore, I turn to higher-order statistics to assess whether the F&P average is still a good approximation of the ensemble average of simulations in this regime, since they are able to capture the non-Gaussian features of the  $\delta T_b$ . While skewness

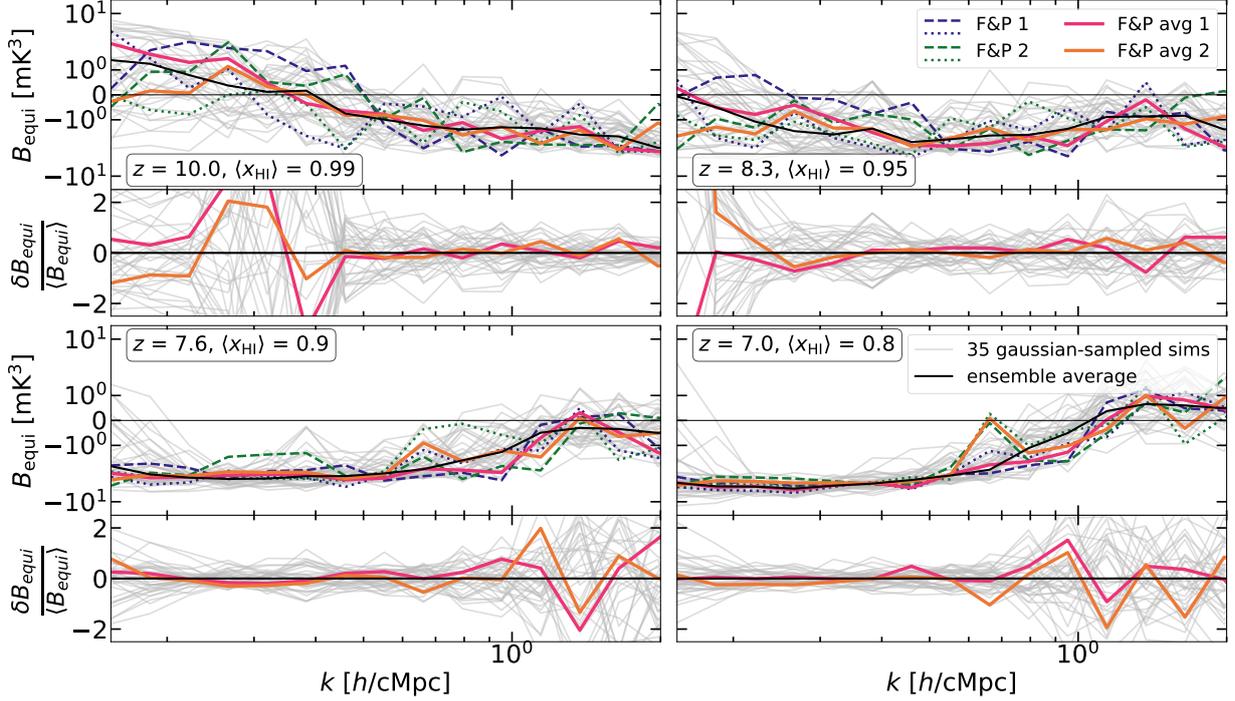


Figure 4.5: **Top panels:** Equilateral triangle bispectra ( $B_{\text{equi}}$ ) for  $0.15 \leq k/(h \text{ cMpc}^{-1}) \leq 2.0$  of GIC simulations (grey solid), their ensemble average (black solid), two F&P pairs (purple and green, dashed and dotted) and their averages (magenta and orange, solid) for  $z = 10, 8.3, 7.6,$  and  $7,$  clockwise from upper left. **Bottom panels:** The normalised deviation of each random simulation and F&P average from the ensemble average.

as discussed in Section 4.3.1 would show some broad non-Gaussian features of the signal, it is still a one point statistic, i.e. it will not quantify the correlation of the signal between different Fourier modes. Thus, I now focus on the bispectrum as defined in Majumdar et al. (2018):

$$b_{21\text{cm}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \langle \delta T_b(\mathbf{k}_1) \delta T_b(\mathbf{k}_2) \delta T_b(\mathbf{k}_3) \rangle \quad (4.6)$$

where  $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3$  are the Fourier space wave numbers,  $\delta_D$  is the Dirac delta function and  $\langle \delta T_b(\mathbf{k}_1) \delta T_b(\mathbf{k}_2) \delta T_b(\mathbf{k}_3) \rangle$  is a measure of the number of triangles (weighted by the  $\delta T_b$  values at their vertices) of different configurations formed by wave numbers  $\mathbf{k}_1, \mathbf{k}_2$  and  $\mathbf{k}_3$ . The different triangles can be formed by varying the magnitude of the wave numbers.

To evaluate the bispectra from my simulations, I use the publicly available BiFFT package (Watkinson et al., 2017), which employs a Fourier transform based technique (as described in Scoccimarro, 2015; Sefusatti et al., 2016) much faster rather than the more traditional approach of counting individual triangles, while still providing consistent results. Following Majumdar et al. (2020), I normalize  $b_{21\text{cm}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$  as  $B_{21\text{cm}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = k_2^3 k_3^3 / (2\pi^2)^2 b_{21\text{cm}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ .

As done in Figure 4.4, I focus on wave numbers between  $0.15 \leq k \leq 2h \text{ cMpc}^{-1}$  and show only two of the five F&P averages for the sake of visual clarity. Further, I consider only two reference

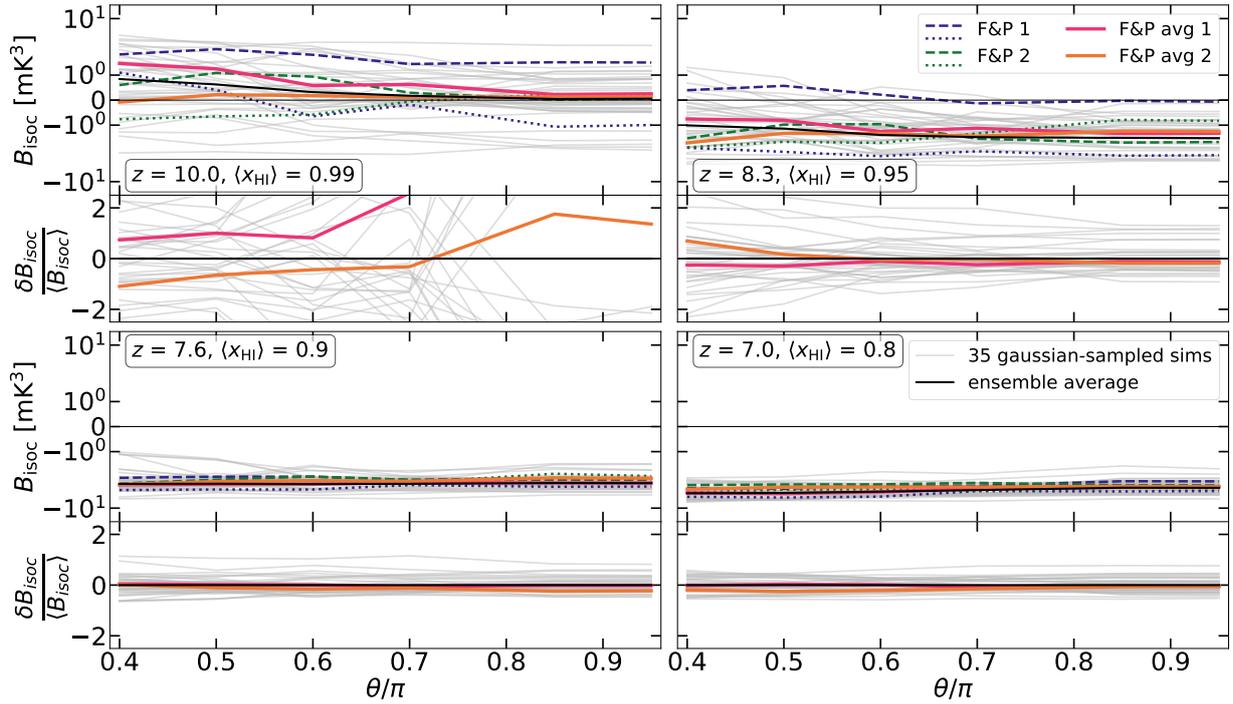


Figure 4.6: **Top panels:** Isosceles triangle bispectra ( $B_{\text{isoc}}$ ) for  $k_1 = k_2 = 0.2 h \text{ cMpc}^{-1}$  for  $z = 10, 8.3, 7.6$ , and  $7$ , clockwise from upper left. Colours and linestyles follow Figure 4.5. **Bottom panels:** The normalised deviation of each random simulation and F&P average from the ensemble average.

cases:

- **Equilateral triangles ( $B_{\text{equi}}$ ):** Here I set  $k_1 = k_2 = k_3 = k$ , where  $k$  goes from  $0.15$  to  $2 h \text{ cMpc}^{-1}$ . This allows me to explore the non-Gaussian features of the signal across various physical scales. The results are shown in Figure 4.5, with the same colours and linestyles as Figure 4.4. I note that the F&P averages are a close match to the ensemble average across all redshifts. The apparent large deviations seen at some wave modes of the F&P average in comparison to the ensemble average (lower panels of Figure 4.5) arise because at those scales the bispectra approach 0. The normalisation by the ensemble average thus exaggerates the small differences significantly. Giri et al. (2023) carried out a similar analysis for  $z = 9$  (corresponding to  $\langle x_{\text{HI}} \rangle = 0.8$  in their simulations), and found the F&P averages to be a close match to the ensemble average for  $k > 0.1 h \text{ cMpc}^{-1}$ . Thus my results are consistent with their conclusions. I also quantify the improvement on using the F&P method for  $B_{\text{equi}}$  in Appendix B, using the methodology of Section 4.4.1, considering all my F&P pair averages.
- **Isosceles triangles ( $B_{\text{isoc}}$ ):** I set  $k_1 = k_2 = k = 0.2 h \text{ cMpc}^{-1} \neq k_3$  to explore large physical scales more thoroughly. I plot  $B_{\text{isoc}}$  versus the opening angle between the vectors  $\mathbf{k}_1$  and  $\mathbf{k}_2$  given as  $\theta = \cos^{-1}(\mathbf{k}_1 \cdot \mathbf{k}_2 / (k_1 k_2))$  in Figure 4.6, with the same colours and linestyles as

Figure 4.4. As expected, the GIC simulations show large sample variance, which reduces with decreasing redshift. However, the F&P averages are an even closer match to the ensemble average as compared to the GIC simulations, and thus continue to provide an improvement across all four of the redshift bins used. Similarly to what observed for  $B_{\text{equi}}$ , the large deviations seen at some scales at  $z = 10$  are due to  $B_{\text{isoc}}$  approaching 0.

I also repeat the process for other  $B_{\text{isoc}}$ , by varying  $k_1 = k_2 = k$  from  $0.15$  to  $1.5 h \text{ cMpc}^{-1}$ , and find that the trends hold across all physical scales. This is an interesting result, as it provides a useful statistic for modelling the 21-cm signal using F&P averages at redshifts with  $\langle x_{\text{HI}} \rangle \leq 0.9$ . It thus works well for comparing with observations too, as at these redshifts the bispectrum is a more useful statistic than the power spectrum.

However, it is necessary to verify that the improvement in estimating the “true” bispectrum when using an F&P average indeed correlates with statistical estimates of the physical properties (i.e., the distribution of neutral hydrogen) at the observed redshifts. A good way to check this, is to analyse the properties and distribution of the source of the non-Gaussian features of the  $\delta T_b$ , i.e., the ionized regions.

#### 4.3.4 Bubble size distribution

There is no universal consensus on how to identify ionized regions, with several methods having been used in literature (see [Giri et al., 2018](#), for a detailed comparison of different methods). Here, firstly, I choose to define cells with  $x_{\text{HI}} \lesssim 0.5$  as ionized. Next, I use a Friend-of-Friends algorithm based on the `ndimage` package of SciPy ([Virtanen et al., 2020](#)) to identify regions with clusters of such ionized cells, which are then considered as “bubbles”. From their volume, I derive the radius of an equivalent sphere and use it as an effective bubble radius ( $r_{\text{eff}}$ ). The threshold value of  $x_{\text{HI}}=0.5$  chosen to identify a cell as neutral or ionized is arbitrary. However, I find that varying this value does not affect my qualitative results.

The maximum value of  $r_{\text{eff}}$  is determined by the box size, while I ignore bubbles equivalent to an individual cell as they are resolution limited. I then find radii in the range  $0.16 \lesssim r_{\text{eff}}/(h^{-1} \text{ cMpc}) \lesssim 40$ . As my simulations are run only down to  $z = 7$  and quasars are not included as sources of ionizing photons, I do not expect many bubbles with  $r_{\text{eff}} > 5 h^{-1} \text{ cMpc}$ , and thus do not need to worry about the other extreme of the resolution range. I thus limit my analysis to the range  $0.3 \leq r_{\text{eff}}/(h^{-1} \text{ cMpc}) \leq 5$ , and build histograms of number of bubbles ( $n_{\text{bubbles}}$ ) binned according to  $r_{\text{eff}}$ . In Figure 4.7, I construct violin plots (cyan) for the different values of  $n_{\text{bubbles}}$  for the GIC simulations, and also show the ensemble average (black circles). To compare this with the two F&P averages used in Section 4.3.3, I plot the averages for the number of bubbles for the individual fixed and paired simulations using the same colour scheme (magenta and orange triangles). Again, I only show two out of the generated five averages for the sake of easy visual comparison, as all five give similar results. I note that both F&P averages are a good match for the ensemble average, and even when they deviate, they remain well within the violins, showcasing the ranges of the GIC simulations. This confirms that the improvement

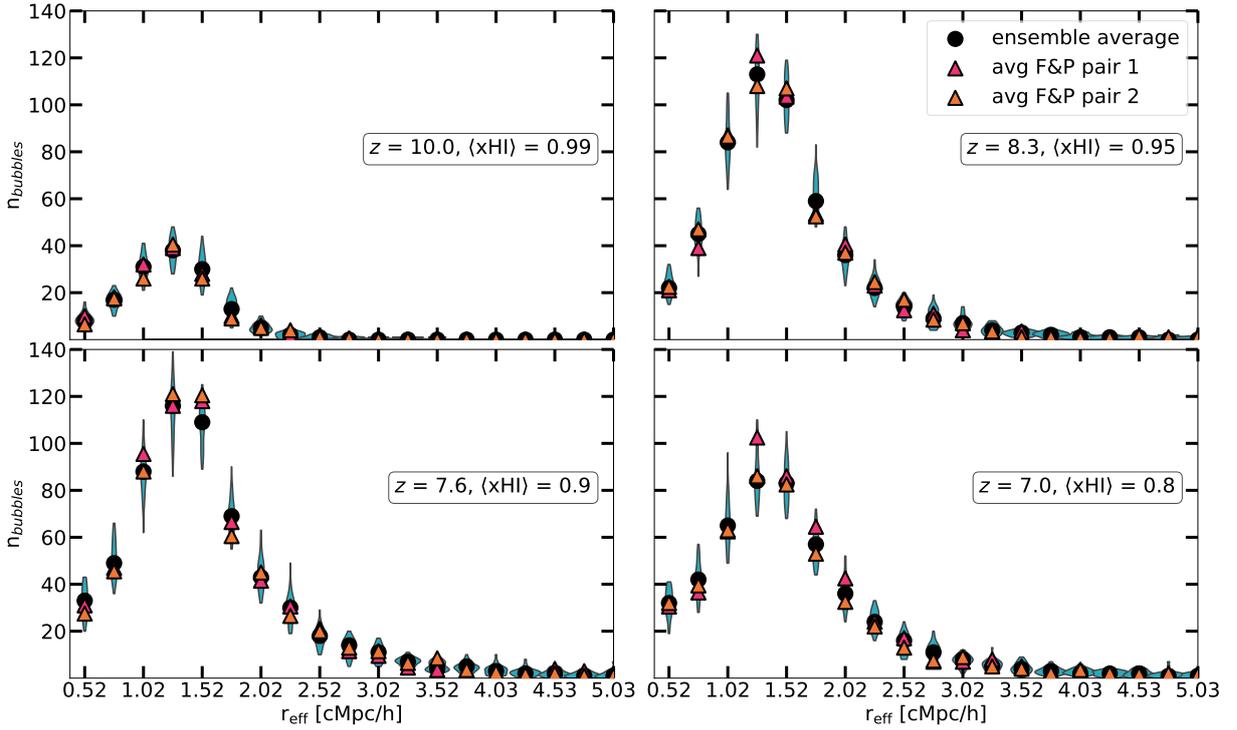


Figure 4.7: Number of ionized bubbles in different radii  $r_{\text{eff}}$  bins at  $z = 10, 8.3, 7.6,$  and  $7$ . I plot their distribution for the GIC simulations (cyan violin), the ensemble average (black circles), and the two F&P averages used in Sections 4.3.2 and 4.3.3 (magenta and orange triangles).

in the bispectrum noted when using the F&P average is obtained because they closely match the number and sizes of ionized regions of the ensemble average.

Further, as expected, I see that the violin plots get narrower as I approach lower redshifts, as small scale variability between the GIC simulations grows with decreasing redshift. Lastly, I note that the number of smaller bubbles grows when going from  $z = 10$  to  $z = 7.6$ , but by  $z = 7$ , many of them would have begun merging, leading to a fall in the number of bubbles with  $1.0 \leq r_{\text{eff}}/(h^{-1} \text{ cMpc}) \leq 2.0$ .

## 4.4 Discussion

### 4.4.1 Advantage of the F&P method

Qualitatively, I note that the F&P average provides a closer estimate of the ensemble average, as compared to any individual GIC simulation, for statistics like the skewness and the power spectrum, at least for  $\langle x_{\text{HI}} \rangle \geq 0.9$ . For  $\langle x_{\text{HI}} \rangle \lesssim 0.9$ , the improvements for the power spectrum is reduced and the bispectrum becomes the better statistic to use.

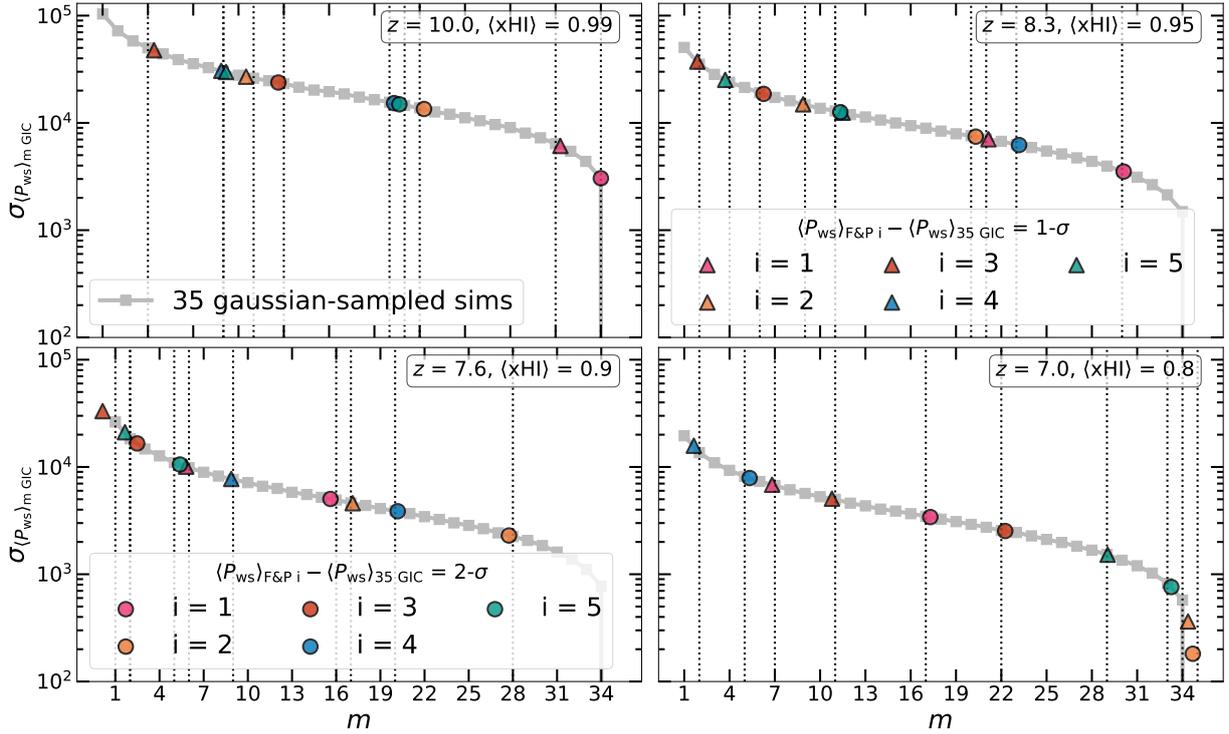


Figure 4.8: The standard deviation curve (grey) generated by interpolating the standard deviations of  $\langle P_{ws} \rangle_m \text{ GIC}$  (grey squares) versus number of sampled GIC simulations  $m$  going from 1 to 35 for  $z = 10, 8.3, 7.6$  and  $7$ , clockwise from top left. The five F&P averages are plotted as triangles (circles) with the assumption of them being  $1-\sigma$  ( $2-\sigma$ ) away from  $\langle P_{ws} \rangle_{35 \text{ GIC}}$  using the same colours as used in Figures 4.1 and 4.3. Their closest matches in the standard deviation curve are shown with black dotted vertical lines gives  $m_{\text{eq}}$ .

However, it is necessary to quantify this improvement as compared to the average of multiple GIC simulations. For this reason, here I lay out the methodology to find the number of GIC simulations needed to match the performance of one F&P average with respect to the power spectrum. For this, I define “performance” as the extent of deviation of the average of multiple GIC simulations or of an F&P pair from the power spectrum of the ensemble average. If the number of GIC simulations required to match the performance of an F&P average is greater than 2, this means more of them need to be run to achieve the same performance. In this case, using the F&P average which just needs 2 simulations to be run would reduce computational costs.

Ideally, for this comparison one should run a large number of GIC simulations as well as F&P averages, and compare the extent of their deviation from the ensemble average at specific wave-modes. However, as radiation-hydrodynamical simulations are computationally expensive, I utilise the 35 GIC simulations discussed in previous sections, and compare them to all 5 F&P averages generated in Section 4.2.2.1. To mimic large number statistics, I proceed as follows:

1. **Wave-mode window:** The availability of a large number of simulations allows the investi-

gation of specific wave-modes, while the same cannot be done when I am limited to just a small number of them as in this case individual wave-modes are affected as well by random noise. I thus choose a window of wave-modes and sum all the power spectra at  $ks$  contained within it for each GIC simulation ( $P_{ws}$ ), their ensemble average ( $\langle P_{ws} \rangle_{35 \text{ GIC}}$ ), as well as each F&P average ( $\langle P_{ws} \rangle_{\text{F\&P } i}$ , where  $i$  goes from 1 to 5). I choose this window to be  $0.15 \leq k/(h \text{ cMpc}^{-1}) \leq 0.4$  to ensure that the largest physical scales covered correspond to those at which the current and next-generation radio telescopes are most sensitive.

2. **Choosing GIC simulations:** To evaluate the ‘effective volume’ of the F&P simulations, I need to determine the number  $m_{\text{eq}}$  of GIC ones that match the statistical power of a single pair of F&P simulations. However, different subsets of  $m < 35$  GIC simulations will produce different results, especially for small values of  $m$ . Therefore, in the following I consider all the  ${}^{35}C_m$  possible subsets, which is the number of combinations of  $m$  objects out of 35. But for computational reasons, I cap the number of combinations to  $10^4$ .
3. **Standard deviation curve:** For every  $m$ , I compute the average  $P_{ws}$  for all  ${}^{35}C_m$  combinations. These  $\langle P_{ws} \rangle_m \text{ GIC}$  values form a Gaussian distribution centred around the ensemble average value  $\langle P_{ws} \rangle_{35 \text{ GIC}}$ . I measure the standard deviation of this Gaussian distribution ( $\sigma_{\langle P_{ws} \rangle_m \text{ GIC}}$ ) for each  $m$ . I linearly interpolate these values to generate the grey curves in Figure 4.8, with the grey squares referring to  $\sigma_{\langle P_{ws} \rangle_m \text{ GIC}}$ .
4. **Standard deviation of the F&P runs:** Although the above procedure should be repeated for the F&P runs, this is prohibitively expensive from a computational point of view. Therefore, I explicitly parameterize my ignorance of the true width of the distribution of  $\langle P_{ws} \rangle_{\text{F\&P}}$  by assuming that each F&P average lies exactly  $1\text{-}\sigma$  away from its center. In other words, I assume that  $\langle P_{ws} \rangle_{\text{F\&P } i} - \langle P_{ws} \rangle_{35 \text{ GIC}}$  is a measure of the width of the Gaussian distribution of  $\langle P_{ws} \rangle_{\text{F\&P}}$ . This is done independently for each F&P run, and is shown in Figure 4.8 with triangles. I then repeat this procedure but assuming that each F&P average lies exactly  $2\text{-}\sigma$  away from the center of the distribution, and show the results with circles.
5. **Closest  $m$  matching:** For the F&P runs, I compare their standard deviation with the curve generated from the GIC simulations in step (iii) and determine the closest  $m$  value (black dotted vertical lines in Figure 4.8). This is my estimate of  $m_{\text{eq}}$  for each F&P simulation pair.
6. **Improvement factor ( $f_{\text{imp}}$ ):** Finally, I define the improvement factor  $f_{\text{imp}} = m_{\text{eq}}/2$  for all 5 F&P runs. This corresponds to the ratio between the number of simulation runs (i.e. simulated volumes, since all my runs have the same box size) necessary with GIC and F&P ICs, for which I report the minimum, maximum and average values in Table 4.1.

From Figure 4.9, I note that the extent of improvement provided by the F&P method reduces with decreasing redshift, as was expected from the qualitative results of Section 4.3.2. However, interestingly, this trend seems to stop at  $z = 7.6$ , with higher values of  $f_{\text{imp}}$  noted at  $z = 7$ . This indicates that while the F&P average may be performing worse at lower redshifts, it still does

Table 4.1:  $f_{\text{imp}}$  is the factor of improvement on running an F&P average over running multiple GIC simulations. I report the minimum, maximum and average value of this quantity for the two cases discussed in (iv), for the five F&P averages.

$z$	1 - $\sigma$ away			2 - $\sigma$ away		
	$f_{\text{imp,min}}$	$f_{\text{imp,max}}$	$\langle f_{\text{imp}} \rangle$	$f_{\text{imp,min}}$	$f_{\text{imp,max}}$	$\langle f_{\text{imp}} \rangle$
10.0	2.0	15.5	6.4	6.5	17.0	10.9
8.3	1.0	10.5	4.7	3.0	15.0	9.0
7.6	0.5	8.5	3.5	1.0	14.0	7.1
7.0	1.0	17.0	8.3	2.5	17.5	11.2

better than an average of a few GIC simulations. In fact, I note that  $\langle f_{\text{imp}} \rangle$  is  $\sim 6$  at  $z = 10$  (thus one F&P average is better than running 12 GIC simulations), but  $\sim 8$  at  $z = 7$  (equivalent to 16 GIC simulations) for the case of F&P averages being 1- $\sigma$  away. The lowest average improvement is 3.5 at  $z = 7.6$ , indicating that at least 7 GIC simulations are needed to match an F&P average.

As I sum the power spectra across the aforementioned wave-mode window, a direct comparison between my results and those of [Giri et al. \(2023\)](#) is difficult. Nevertheless, I note that in the range  $10 \geq z \geq 7$ , I obtain  $f_{\text{imp}} \geq 3.5$ , which agrees with their result of an improvement of at least a factor of 4 at  $k = 0.1 h \text{ cMpc}^{-1}$  at  $z = 9$ ,  $\langle x_{\text{HI}} \rangle = 0.8$ . The worst possible value of  $f_{\text{imp}}$  is 0.5, which corresponds to the case when I have a single GIC simulation perform better. This is expected, as pure randomness does allow such chance events. However, as the average improvement is above 2, I believe this still supports the use of F&P averages for modelling the 21-cm signal power spectrum rather than running GIC simulations.

A similar analysis can be run for the equilateral triangles bispectrum. I present this in more details in [Appendix B](#), where I find  $f_{\text{imp}} \geq 5.0$ . This result showcases that the F&P method is even better for the bispectrum, and thus using multiple summary statistics for the 21-cm signal can allow me to maximise the interpretation of the 21-cm signal without requiring large effective volumes. Note that similar analyses can be performed for bispectra generated for different values of  $k_1$ ,  $k_2$  and  $k_3$ , by using a window over the opening angle instead of the wave-mode.

#### 4.4.2 Limitations and future applications

As I have shown in the previous Section, the F&P method can be a powerful tool to extend the statistical accuracy of limited-volume simulations of the EoR 21-cm signal. In combination with the advancement of computational techniques and hardware, this method can deliver accurate predictions over volumes of interest for the study of the 21-cm signal, especially in the early phases of cosmic reionization. However, I caution that the F&P method does not improve my predictive ability concerning individual galaxy properties, as they are dominated by local processes and environment. Therefore, the improvement granted on statistical quantities (e.g. the power spectrum and bispectrum) does not necessarily mirror exactly on different observables that might be more affected by individual objects (see e.g. the discussion in [Sec. 4.3.1](#)).

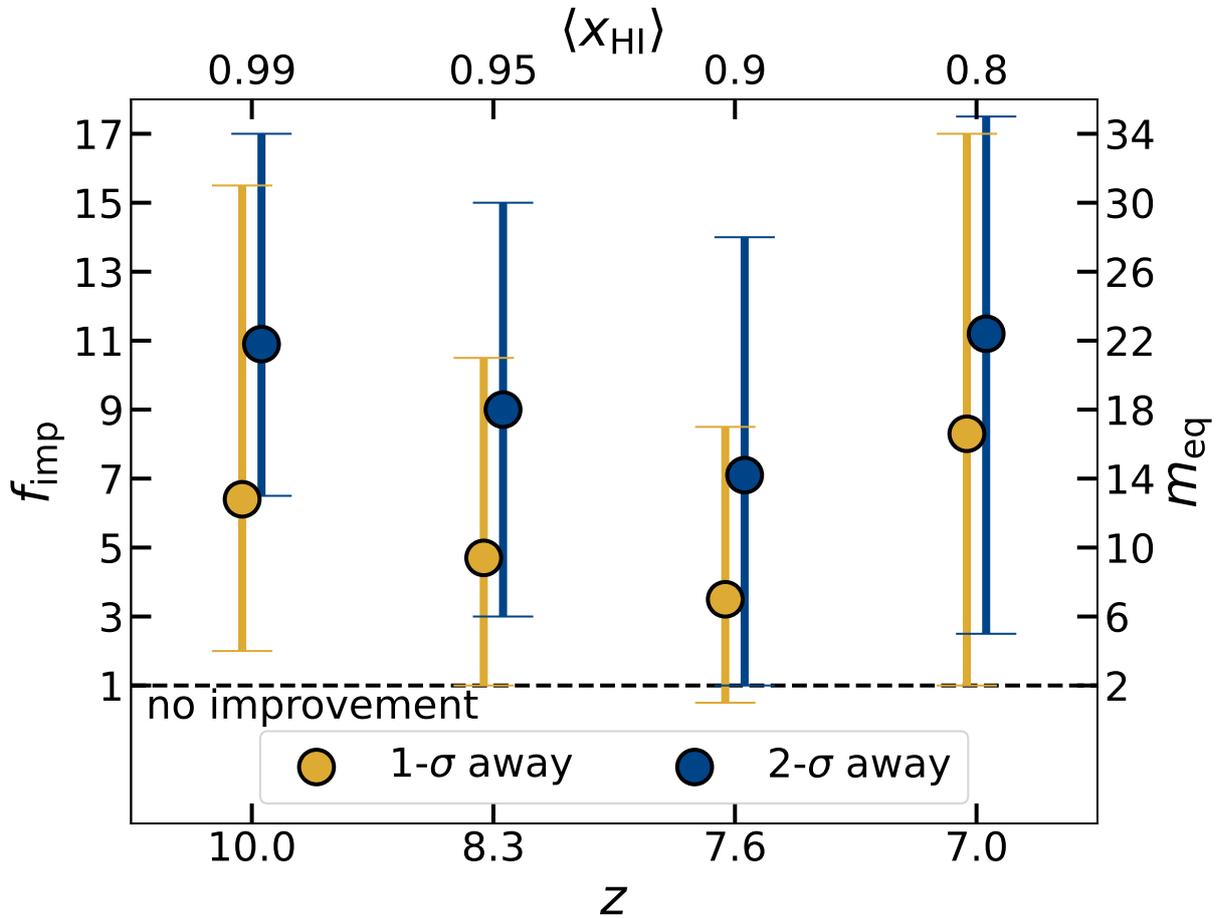


Figure 4.9: Redshift evolution of the improvement factor in computational expense,  $f_{\text{imp}}$ , when using the F&P approach. The errorbars show the range from  $f_{\text{imp,min}}$  to  $f_{\text{imp,max}}$  for the F&P averages being 1- $\sigma$  away (yellow) or 2- $\sigma$  away (blue) from an assumed distribution of F&P averages, and the points show the average,  $\langle f_{\text{imp}} \rangle$ . The top axis shows the corresponding  $\langle x_{\text{HI}} \rangle$  and the right axis shows the number  $m_{\text{eq}}$  of GIC simulations corresponding to  $f_{\text{imp}}$ . The black dashed line indicates no improvement.

I foresee numerous application of this technique. For instance, the lower computational cost required (with respect to traditional approaches) entails e.g. that larger volumes, as well as a broader range of cosmologies and/or physical models can be explored with accurate simulations, greatly improving the reliability of predictions and – eventually – inference from 21-cm signal data. While basing the entire inference process on RMHD simulations remains prohibitive, they will be needed in order to confirm and improve constraints obtained through computationally-cheaper less-accurate methods. Additionally, they are necessary to explore the coupling between small and large scales, e.g. spatial correlations between galaxies and the 21-cm signal, that the advent of SKA will enable. My results suggest that such accurate predictions can be obtained from RMHD simulations even in statistically-significant volumes of the Universe.

Finally, the F&P approach operates orthogonally to super-resolution techniques (Kodi Ramanah et al., 2020; Li et al., 2021, as discussed in Section 1.5.4); therefore these two approaches should be considered complementary rather than in opposition.

## 4.5 Summary

Running simulations for EoR is computationally very expensive. This is exacerbated when aiming at resolving low-mass galaxies (whose importance has been recently shown observationally in Atek et al., 2023) in volumes large enough to be statistically significant for reionization studies. In this work, I explored the Fixed & Paired (F&P) approach (Angulo & Pontzen, 2016; Pontzen et al., 2016) to investigate the possibility of reducing the number of simulations needed (and thus the overall computational expense) to produce unbiased models of the 21-cm signal. While past efforts have used semi-numerical approaches to implement this, I have shown more rigorous results by using radiation hydrodynamic simulations that model more accurately galaxy-scale effects on the 21-cm signal. I focus on the wave modes in the range  $0.15 \leq k/(h \text{ cMpc}^{-1}) \leq 2$ , as the best measurements from present and upcoming radio telescopes like LOFAR, HERA, MWA and SKA are expected in this regime. Further, I focus on redshifts  $10 \geq z \geq 7$ , which in my case correspond to  $1.0 > \langle x_{\text{HI}} \rangle \geq 0.8$ .

To explore the improvement with respect to various 21-cm signal statistics obtained by adopting the F&P approach rather than running Gaussian-sampled initial conditions based (GIC) simulations, I use a setup similar to that of the THESAN project (Kannan et al., 2022; Garaldi et al., 2022; Smith et al., 2022; Garaldi, 2023). In particular, I investigate the impact on the skewness, power spectrum, bispectrum and the ionized region size distribution, and introduce a novel method to quantify the improvement.

I find that the skewness and power spectrum are well-estimated by the F&P averages for  $\langle x_{\text{HI}} \rangle \geq 0.9$ , and their performance is good also in the range  $0.9 > \langle x_{\text{HI}} \rangle \geq 0.8$ , with an improvement in computational cost better than 3.5 for the generation of the power spectrum. I find that the bispectrum is well estimated for  $\langle x_{\text{HI}} \rangle \geq 0.8$ , with the scaled deviation between the F&P averages and the ensemble average being below 1. The only exception are those modes where the ensemble average bispectrum approaches zero, artificially increasing the small differences between the F&P

average and the ensemble average. In fact, I find that the improvement in computational cost is better than a factor of 5 for the equilateral triangle bispectrum. This confirms that the F&P average bispectrum is a great complement to the power spectrum for studies of the 21-cm signal when  $\langle x_{\text{HI}} \rangle \geq 0.8$ . Finally, I show that in this regime the F&P averages also provide a good estimate of the HII regions size distribution, with the F&P averages being within  $2 - \sigma$  deviation of the ensemble average for bubbles with radius  $\leq 5h^{-1}$  cMpc.

Thus, the F&P method can be used to model the 21-cm signal summary statistics with significantly reduced effective volume and computational expense.



# Chapter 5

## Improved simulations: exploring broader parameter spaces

*You've got a whole plateful of "maybe a little less wrong".*

---

- James S. A. Corey

### 5.1 Overview

In this chapter, I look at another major component of improving simulations of the IGM at the Epoch of Reionization by running a suite of the POLAR simulations. In particular, I explore the necessity for broadening the range of physical parameters (and thus the complexity of processes) included by showcasing that the interplay of cosmological and astrophysical parameters governing simulations can lead to multiple viable models that are in agreement with IGM and galactic observables. This also indicates the need for multi-wavelength parameter inference across a wider range of observable quantities to constrain the space of viable parameters. This work has been submitted for publication to the Monthly Notices of the Royal Astronomical Society and is currently available publicly at [Acharya et al. \(2024a\)](#).

In Section 5.2, I discuss the setup of the POLAR simulations with different cosmologies. In Section 5.3, I present the resulting galactic and IGM properties, and in Section 5.4, I discuss the implications of varying astrophysical and cosmological parameters for inference modeling from 21-cm signal observations, jointly constrained with other observables of the EoR. Finally, I summarize my results in Section 5.5.

## 5.2 Methodology

Building on M23, I utilize a similar setup by running  $N$ -body DM simulations, and then post-processing them with the L-GALAXIES SAM (Barrera et al., 2023; Henriques et al., 2015, 2020) to model the formation and evolution of galaxies, and with the 1D radiative transfer code GRIZZLY (Ghara et al., 2015, 2018) to model the gas ionization and the 21-cm signal from neutral hydrogen. In Sections 5.2.1 and 5.2.2, I highlight the key parameters used for setting up the simulations and the analysis done in this work. Further, in Section 5.2.3 I propose two cases: one where I keep the astrophysical parameters the same for all four cosmological models, and one in which I tune them to match UV luminosity functions (UVLFs) observed with JWST and HST at  $z = 10$  and 9. I refer to the first and second case as “unconstrained” and “constrained”, respectively.

### 5.2.1 Dark-matter simulations

I use the GADGET-4 code (Springel et al., 2021) for running  $N$ -body DM simulations with a volume of  $(150h^{-1} \text{ cMpc})^3$  and  $2048^3$  DM particles. This translates into a DM particle mass of  $5 \times 10^7 M_{\odot}$ , matching the particle resolution of larger simulations used by the LOFAR EoR KSP team (see for example Giri et al., 2019b,a). The box size is chosen to probe wave-modes of  $k \leq 1.0h \text{ cMpc}^{-1}$ , where the best results from observations with LOFAR, HERA, MWA (and eventually SKA) are expected (Koopmans et al., 2015). This condition requires simulations with box sizes  $> 100h^{-1} \text{ cMpc}$  (Iliev et al., 2014), although Kaur et al. (2020) suggests box sizes  $> 175h^{-1} \text{ cMpc}$  being necessary when accounting for X-ray heating of the IGM. To sidestep this issue, I assume that at my redshifts of interest, the IGM has already been heated above the CMB temperature (see Section 5.2.2.2). I additionally employ the F&P approach (Angulo & Pontzen, 2016) to mitigate sample variance, and thus run two realisations of each DM simulation. Each pair of realisations has the mode amplitudes fixed to the square root of the initial matter power spectrum, and the phases of the second realisations (B-series) are obtained by mirroring those of the first realisations (A-series). The F&P approach of averaging observables of these two realisations has been shown to boost the statistical precision of the matter power spectrum, bispectrum and halo mass function (Angulo & Pontzen, 2016; Chartier et al., 2021; Maion et al., 2022; Villaescusa-Navarro et al., 2018), Lyman- $\alpha$  power spectra (Anderson et al., 2019), 21-cm signal power spectrum and bispectrum (Giri et al., 2023; Acharya et al., 2024c) and several other quantities derived from simulations (Villaescusa-Navarro et al., 2018; Klypin et al., 2020). Beyond this, I use the same random seed to minimize the effect of randomized initial conditions for both the A and B-series of simulations.

For my reference simulation, I assume a “fiducial”  $\Lambda$ -Cold Dark Matter ( $\Lambda$ CDM) cosmological model based on Planck Collaboration et al. (2020, specifically the TT,TE,EE+lowE+lensing+BAO case), setting  $\Omega_{\Lambda} = 0.6889$ ,  $\Omega_{\text{m}} = \Omega_{\text{b}} + \Omega_{\text{dm}} = 0.3111$ ,  $\Omega_{\text{b}} = 0.04897$ ,  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$  with  $h = 0.6766$ ,  $\sigma_8 = 0.8102$  and  $n_s = 0.9665$ , where the symbols have their usual meaning. Further, I consider three additional cosmologies, where I vary  $h$  and  $\sigma_8$ , while keeping the other parameters fixed to the above-mentioned values. This is necessary to consider the maximum

Table 5.1: The four  $N$ -body dark matter simulation models considered in this work. From left to right, the model name, the adopted value of  $h$  and  $\sigma_8$ , and the reference for the values. All other cosmological parameters are fixed to the [Planck Collaboration et al. \(2020\)](#) values.

Model	$h$	$\sigma_8$	Reference
fiducial	0.6766	0.8102	<a href="#">Planck Collaboration et al. (2020)</a>
$h$ high	0.7330	0.8102	<a href="#">Riess et al. (2022)</a>
$\sigma_8$ low	0.6766	0.7020	<a href="#">Tröster et al. (2020)</a>
$\sigma_8$ high	0.6766	0.8800	<a href="#">Ghirardini et al. (2024)</a>

impact of changing cosmological parameters across the range of observed values. First, I adopt  $h = 0.7330$  (from [Riess et al., 2022](#)), and refer to this as the “ $h$  high” model. In addition, I explore two extreme values of  $\sigma_8$ , namely the “ $\sigma_8$  low” case with  $\sigma_8 = 0.702$  ([Tröster et al., 2020](#)), and the “ $\sigma_8$  high” case with  $\sigma_8 = 0.88$  ([Ghirardini et al., 2024](#)). The details of the four simulation models are listed in Table 5.1.

The initial conditions for all simulations are generated at  $z = 199$  with a second-order Lagrangian perturbation theory based on the NGENIC algorithm implemented into GADGET-4, using the same linear theory power spectrum as [Hernández-Aguayo et al. \(2023\)](#). I generate 90 snapshots between  $z = 20$  and  $z = 5$ , for which I save full snapshot information. However, in order to ensure a maximum step size of 10 Myrs between snapshots used to build halo merger trees with the Friend-Of-Friends (FOF) group finding algorithm ([Springel et al., 2001](#)), I perform a finer gridding to generate a total of 156 output time steps for FOF groups. I use a standard linking length of 0.2 times the mean particle spacing and a minimum group size of 64 DM particles (corresponding to a minimum halo size of  $\approx 3 \times 10^9 M_\odot$ ). Substructures within halos are identified using the SUBFIND-HBT algorithm ([Han et al., 2018](#); [Springel et al., 2021](#)). Finally, the gravitational softening length is set to 0.025 of the mean particle spacing, i.e.  $\approx 1.83h^{-1}$  ckpc.

To evaluate the performance of the simulations, in Figure 5.1 I present the halo mass function (HMF) as  $M_h \frac{dn_h}{dM_h}$ , where  $M_h$  and  $n_h$  are the halo mass and number density respectively, and compare them with fits by [Tinker et al. \(2008\)](#), computed using the Python package HMF ([Murray, 2014](#)) at  $z = 9$  and 10. As in this study, [Tinker et al. \(2008\)](#) takes into account the non-universality of the HMF by considering the impact of varying the cosmological parameters in the  $\Lambda$ CDM model. I note that the fiducial and  $h$  high models at these redshifts produce very similar HMFs, despite some noticeable differences in their corresponding fits. A deeper exploration indicates that my simulation setup is less sensitive to differences in the Hubble parameter, given the choice of the starting redshift, and that the initial conditions were generated using second-order Lagrangian perturbation theory. A significantly higher starting redshift or third/fourth-order Lagrangian perturbation theory may thus be required to better match the fits. Nevertheless, some differences do exist, especially at the extremities of the HMF, which lead to the differences noted in the subsequent sections.

Further, I also find a mild under-prediction of the HMF at the low mass end and an over-prediction at the high mass end. This may be due to the choice of the linking length of 0.2, which

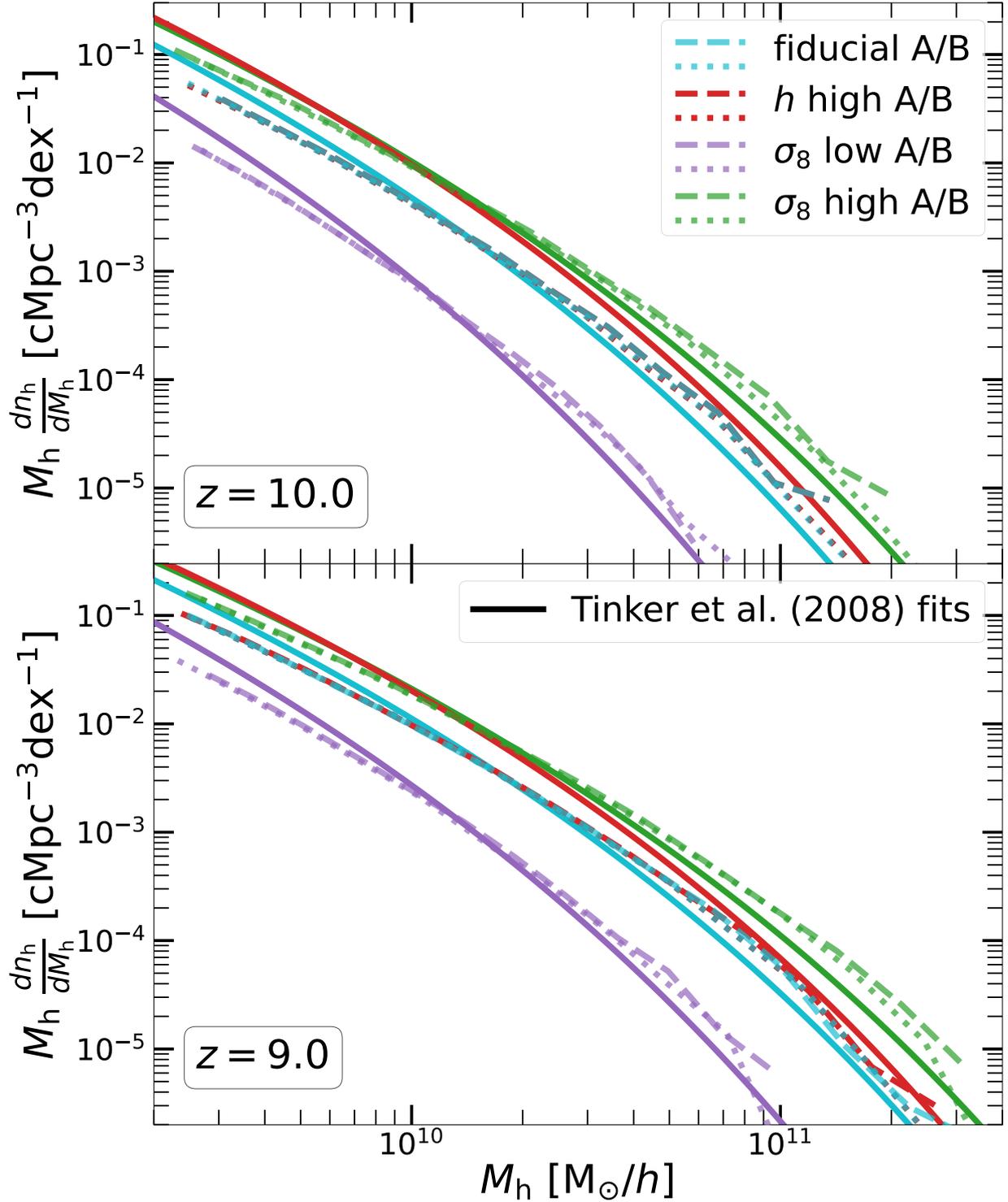


Figure 5.1: Halo mass functions for my four models: fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green) at  $z = 10$  (top panel) and 9 (bottom). The A- and B-series are shown with dashed and dotted lines, respectively. The corresponding fit for each simulation using the Tinker et al. (2008) model is shown with a solid grey line.

at significantly high redshifts may lead to overlinking. A lower value of 0.17, as suggested by [Watson et al. \(2013\)](#), may resolve this issue. However, I still use 0.2 for minimizing differences between the setups used for past applications of L-GALAXIES (e.g., [Henriques et al., 2015, 2020](#); [Barrera et al., 2023](#), and [M23](#)) and my setup. Lastly, I note that the B-series has slightly more massive halos in all models. This is purely because of the choice of the random seed, which happens to lead to some regions of massive halo formation in the B-series where the A-series have voids.

## 5.2.2 Galactic and IGM properties with POLAR

POLAR is a semi-numerical model designed to obtain high- $z$  galaxy properties and the 21-cm signal from the IGM in a fast and robust manner. It combines the semi-analytical galaxy formation and evolution model of L-GALAXIES ([Henriques et al., 2015, 2020](#); [Barrera et al., 2023](#)) with the one-dimensional radiative transfer code GRIZZLY ([Ghara et al., 2015, 2018](#)).

### 5.2.2.1 Semi-Analytic modeling of galaxies

I use L-GALAXIES as described by [Barrera et al. \(2023\)](#), as a post-processing module of GADGET-4. This is an updated version of the publicly available L-GALAXIES 2020 that was used by [M23](#). L-GALAXIES implements most major physical processes of gas cooling, star formation, galaxy mergers, supernovae feedback, black hole growth, AGN feedback, and dust attenuation. While [Barrera et al. \(2023\)](#) largely builds on [Henriques et al. \(2015\)](#), I also consider the parameters used by [Henriques et al. \(2020\)](#) before adapting them for my purposes. In particular, I focus on those that control star formation efficiency ( $\alpha_{\text{SF}}$ ), star formation efficiency during galaxy mergers ( $\alpha_{\text{SF,burst}}$ ), AGN accretion rate ( $k_{\text{AGN}}$ ), reheating of cold gas by star formation ( $\epsilon_{\text{reheat}}$ ,  $V_{\text{reheat}}$ ), and the energy released by each supernova ( $E_{\text{SN}}$ ).

I tune the parameters to match photometric observations of the UVLFs from JWST and HST at  $z = 10$  and  $9$ , as these redshifts are observationally relevant for the LOFAR EoR KSP. I note that L-GALAXIES assumes a 100% escape fraction for the UV photons, and thus I also implement the dust attenuation approach of [Henriques et al. \(2015\)](#). However, this is a simplified model and more complex ones may lead to greater suppression of the UV luminosity function. In [Table 5.2](#) I list all the possible parameters available in L-GALAXIES, and their values set to match UVLF observations as discussed in [Section 5.2.3](#).

### 5.2.2.2 Radiative transfer and the 21-cm signal

Modeling the EoR requires the inclusion of radiative transfer to describe the hydrogen ionization and heating. For this, I take the results of the  $N$ -body simulations from [Section 5.2.1](#) and the semi-analytic modeling of galaxies from [Section 5.2.2.1](#), and post-process them with the 1D radiative transfer code GRIZZLY, as done by [M23](#). GRIZZLY uses pre-computed ionization and

temperature profiles of gas for different source and density properties at various redshifts to model the ionization and heating processes and the differential brightness temperature of the 21-cm signal ( $\delta T_b$ ).

More specifically, GRIZZLY requires as input the gridded matter density field and dark matter halo masses from the  $N$ -body simulation, as well as the corresponding galactic stellar masses and stellar ages obtained with L-GALAXIES. My reference simulation has a grid of  $256^3$  cells, resulting in a cell size of  $\approx 600 h^{-1}$  ckpc, but I use a range of different sizes (from  $64^3$  to  $512^3$  cells) to assure convergence of output values in the range  $12 > z > 5$ . Next, I assume that the gas density scales by a factor of  $\Omega_b/\Omega_{\text{dm}}$  with the dark matter density. Lastly, as done in M23, I use the Binary Population and Spectral Synthesis (BPASS; Stanway & Eldridge, 2018) code to model the Spectral Energy Distributions (SEDs) of stellar sources. In future work, I will also explore the impact of other source types such as X-ray binaries, shock-heated interstellar medium, and accreting black holes (as done in Eide et al., 2018, 2020; Ma et al., 2021).

GRIZZLY computes  $\delta T_b$  as follows (see Furlanetto et al., 2006):

$$\delta T_b = 27x_{\text{HI}}(1 + \delta_B) \left(1 - \frac{T_{\text{CMB}}}{T_S}\right) \times \left[ \left(\frac{\Omega_b h^2}{0.023}\right) \left(\frac{0.15}{\Omega_m h^2} \frac{1+z}{10}\right)^{1/2} \right] \text{mK} \quad (5.1)$$

where  $x_{\text{HI}}$  is the fraction of neutral hydrogen,  $\delta_B$  is the fractional overdensity of baryons,  $T_S$  is the hydrogen spin temperature,  $T_{\text{CMB}}$  is the temperature of the CMB photons at redshift  $z$ , and  $\Omega_m$  is the total matter density. As done by M23, I assume  $T_S \gg T_{\text{CMB}}$  which is valid when the IGM has been sufficiently heated by X-ray sources and expected to be the case in the range of redshift of interest here. Additionally, I also ignore the impact of redshift space distortions. Thus Equation 5.1 is reduced to

$$\delta T_b = 27x_{\text{HI}}(1 + \delta_B) \left[ \left(\frac{\Omega_b h^2}{0.023}\right) \left(\frac{0.15}{\Omega_m h^2} \frac{1+z}{10}\right)^{1/2} \right] \text{mK}. \quad (5.2)$$

I define the power spectrum of  $\delta T_b$  as:

$$P_{21\text{cm}}(\mathbf{k}) = \delta_D(\mathbf{k} + \mathbf{k}') \langle \delta T_b(\mathbf{k}) \delta T_b(\mathbf{k}') \rangle, \quad (5.3)$$

where  $\delta_D$  is the Dirac delta function,  $\delta T_b(\mathbf{k})$  is the differential brightness temperature in Fourier space, and  $\langle \dots \rangle$  is the ensemble average. In the following, I will report my results in terms of the normalized form of the power spectrum, given by (see Peacock, 1999, for details):

$$\Delta_{21\text{cm}}^2(k) = \frac{k^3}{2\pi^2} \times P_{21\text{cm}}(k). \quad (5.4)$$

### 5.2.3 UV luminosity functions

Because Henriques et al. (2015, 2020) tuned the L-GALAXIES parameters to the low-redshift Universe ( $z < 3$ ), it is necessary to adapt them to match the redshift regime that I are interested



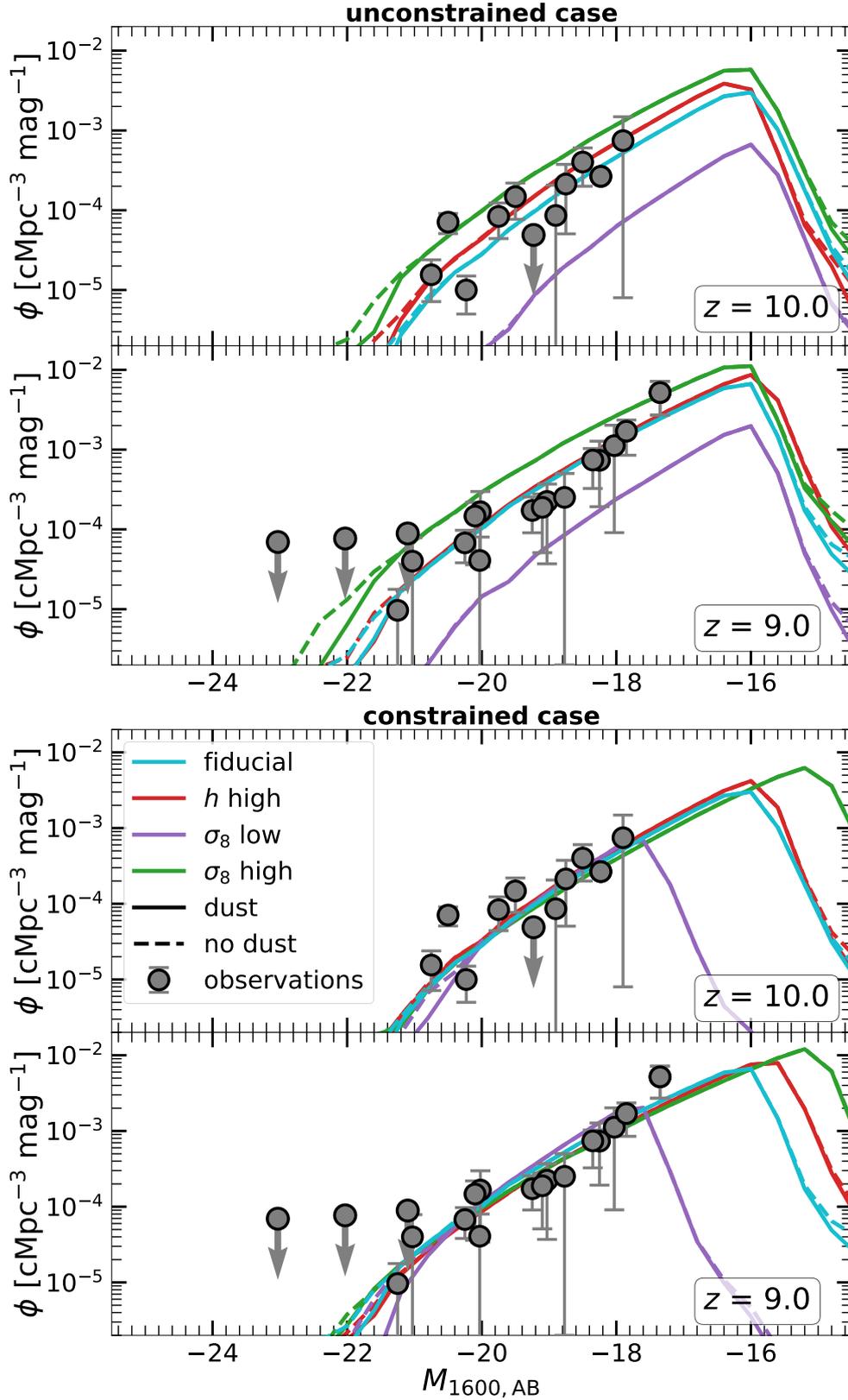


Figure 5.2: A and B-series averaged UVLFs for the unconstrained (top) and constrained (bottom) cases at  $z = 10$  and  $9$  for the fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green). I also show the dust attenuated (solid) and unattenuated (dashed) UVLFs, and JWST and HST observations (Finkelstein et al., 2015; Bouwens et al., 2015, 2021; Harikane et al., 2022; Bouwens et al., 2023a,b; Harikane et al., 2023; Leung et al., 2023; McLeod et al., 2024; Adams et al., 2024, grey circles).

in (see Vani et al., 2024, on discrepancies from observations at higher redshifts). In particular, for the LOFAR EoR KSP team, focussing on  $10 > z > 8.5$  is crucial, as this is the regime in which the LOFAR telescope is most sensitive. Thus, I change the values of the parameters listed in Section 5.2.2.1 to match the observations of the UVLFs from HST legacy fields and JWST programs (Finkelstein et al., 2015; Bouwens et al., 2015, 2021; Harikane et al., 2022; Bouwens et al., 2023a,b; Harikane et al., 2023; Leung et al., 2023; McLeod et al., 2024; Adams et al., 2024) at  $z = 10$  and 9. I have also varied other parameters, but their impact on the UVLFs is minimal (see M23, for a detailed analysis of the impact of different parameters). In future work, I will broaden the range of astrophysical parameters and analyse their effect on other observables.

First, I tune the  $\alpha_{\text{SF}}$ ,  $\alpha_{\text{SF,burst}}$ , and  $E_{\text{SN}}$  values in order for the fiducial case to match the observed high- $z$  UVLFs. Additionally, I reduce the value of the fraction of AGN formed ( $k_{\text{AGN}}$ ) by a small amount, which improves the agreement at the brightest magnitudes ( $M_{1600,\text{AB}} < -20$ ). Similarly, the supernova feedback-based heating efficiency parameters  $\epsilon_{\text{reheat}}$  and  $V_{\text{reheat}}$  are changed to improve the agreement for  $M_{1600,\text{AB}} > -19$ .

I consider two more cases. In the first one, the same L-GALAXIES parameters of the fiducial model above are adopted for all  $N$ -body simulations, to investigate the impact of different cosmologies on the 21-cm signal independently from observations of galaxies at high redshifts. I refer to this model as “unconstrained”. This case allows me to have a clearer picture of the impact of changing cosmologies on the 21-cm signal, without astrophysical processes potentially cancelling out their effects on observables. In the second case, called “constrained”, the L-GALAXIES parameters (specifically  $\alpha_{\text{SF}}$ ,  $\alpha_{\text{SF,burst}}$ , and  $E_{\text{SN}}$ ) are changed for each cosmological model to match the observed UVLFs. This is potentially more interesting, as it allows for a joint constraint of astrophysical and cosmological parameters based on multi-frequency observations. The L-GALAXIES parameters adopted are listed in Table 5.2.

I show the UVLFs for the unconstrained (top) and constrained case (bottom) in Figure 5.2 for all cosmological models, along with observations at  $z = 10$  and 9. In the unconstrained case, I note that the fiducial and  $h$  high models have similar UVLFs, while the  $\sigma_8$  low model underpredicts the UVLF as compared to observations. On the other hand, the UVLF in the  $\sigma_8$  high model is mildly overpredicted. These results are in agreement with the impact that these parameters are expected to have on matter clustering. In the constrained case, the boost in star formation due to higher  $\alpha_{\text{SF}}$  and  $\alpha_{\text{SF,burst}}$  values and a lower  $E_{\text{SN}}$  value, allows the  $\sigma_8$  low model to match the bright end of the UVLF. However, due to lower matter clustering, it does not produce as many faint galaxies as the other cases, and sees a steep drop at  $M_{1600,\text{AB}} > -18$ . On the other hand, the high  $E_{\text{SN}}$  in the  $\sigma_8$  high model strongly suppresses star formation and thus matches other cosmological models, as well as JWST and HST observations. Additionally, it produces more faint galaxies even at  $M_{1600,\text{AB}} > -16$ . Between the  $h$  high and fiducial models, the difference is minimal at  $z = 10$  due to the mildly higher  $E_{\text{SN}}$  in the  $h$  high model. However, at  $z = 9$  the impact of higher matter clustering in the  $h$  high model shows up in the form of more galaxies at the low luminosity end.

Lastly, I note that in the fiducial model the brightest galaxy at  $z = 14$  has  $M_{1600,\text{AB}} = -20.36$ , while at  $z = 13$  it has  $M_{1600,\text{AB}} = -20.60$ , which is in agreement with the recent spectroscop-

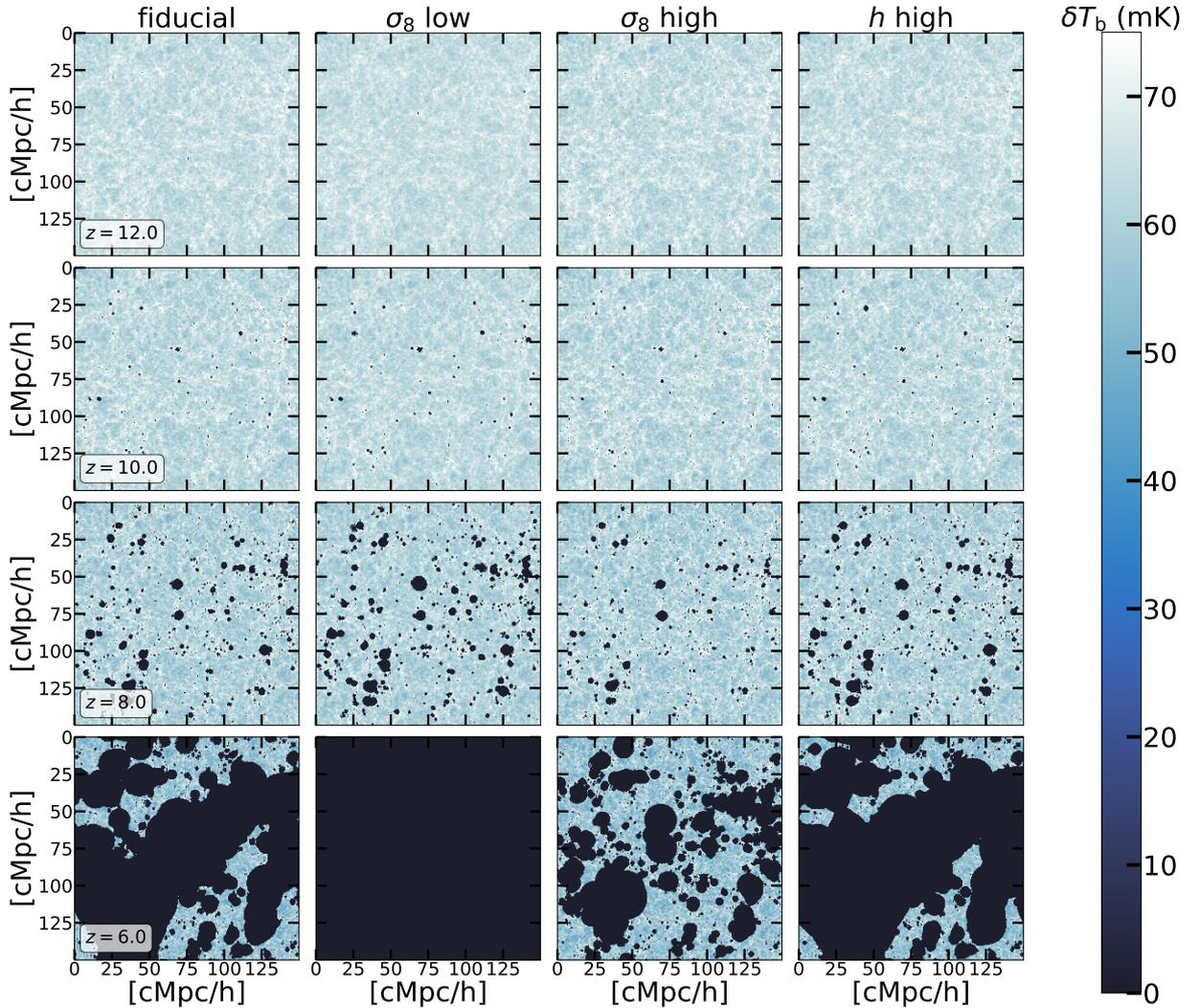


Figure 5.3: Maps of  $\delta T_b$  of the A-series middle slices of single cell thickness (i.e.,  $\approx 600 h^{-1}$  ckpc) for the four cosmological models in the constrained case at  $z = 12, 10, 8,$  and  $6$  (from top to bottom). Here the dark areas represent the ionized regions with  $\delta T_b = 0$ . Note that  $\delta T_b$  cannot have negative values due to the assumption of  $T_S \gg T_{\text{CMB}}$ .

ically confirmed galaxies at  $z = 14.32$  and  $13.90$  (Carniani et al., 2024). For completeness, in Appendix C I also show the UVLFs in the range  $12 > z > 5$  for the constrained case. I find that my models reasonably agree with UVLF observations across this redshift range.

### 5.2.4 Reionization history

Before discussing the reionization history in more detail, I note that the escape fraction of UV photons had been set to  $f_{\text{esc}} = 12.5\%$  in order for the fiducial model to reionize completely

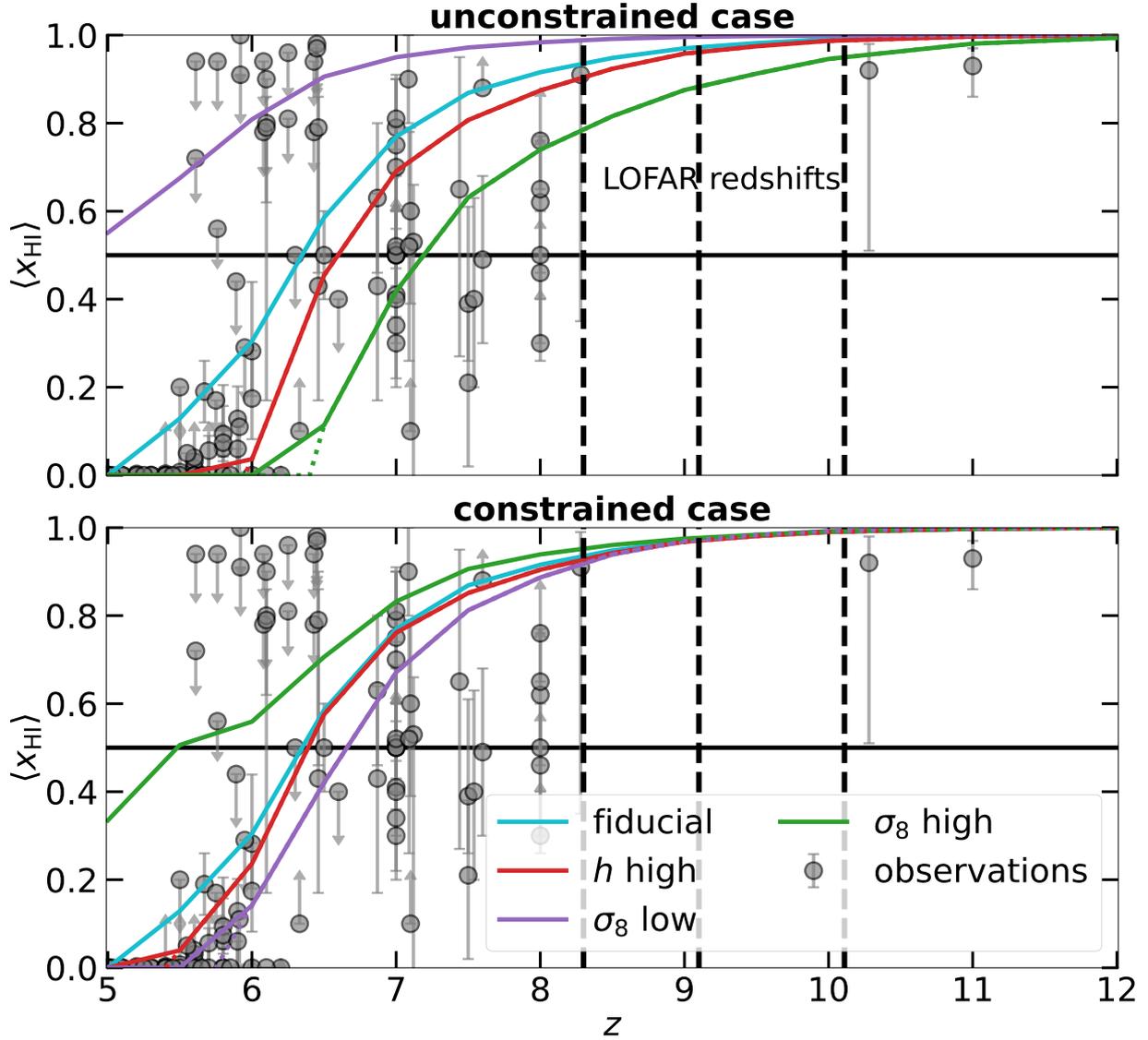


Figure 5.4: Redshift evolution of the average of the A and B-series volume-averaged neutral hydrogen fraction  $\langle x_{\text{HI}} \rangle$  for the fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green) models for the unconstrained (top) and constrained (bottom) cases. Dotted lines refer to a fit to the curves, which is used for a better estimate of the redshift of reionization when the redshift resolution is too coarse. The vertical grey dashed lines indicate the redshifts observationally relevant for LOFAR ( $z = 10.11, 9.16$  and  $8.3$ ), and the black solid line is drawn at  $\langle x_{\text{HI}} \rangle = 0.5$  to guide the eye. Grey circles are a collection of observational constraints (Fan et al., 2006b; Totani et al., 2006; Ota et al., 2008; Ouchi et al., 2010; Bolton et al., 2011; Dijkstra et al., 2011; McGreer et al., 2011; Mortlock et al., 2011; Ono et al., 2012; Chornock et al., 2013; Jensen et al., 2013; Robertson et al., 2013; Schroeder et al., 2013; Pentericci et al., 2014; Schenker et al., 2014; McGreer et al., 2015; Sobacchi & Mesinger, 2015; Choudhury et al., 2015; Mesinger et al., 2015; Greig et al., 2017; Davies et al., 2018; Mason et al., 2018; Hoag et al., 2019; Greig et al., 2019; Jones et al., 2024, collected in the CoRECON module, Garaldi 2023).

(i.e. the average neutral fraction  $\langle x_{\text{HI}} \rangle$  becomes zero) by  $z = 5$ . For the other models, though, full reionization is reached at different redshifts. In the future I plan to consider  $f_{\text{esc}}$  as a free parameter as well, although, as a reference, in Appendix E I show the reionization history of the fiducial model with  $f_{\text{esc}} = 25\%$ .

For a qualitative view of the reionization history, in Figure 5.3 I show the A-series (i.e. with “fixed” initial conditions in the F&P pair of simulations) middle slices of the  $\delta T_b$  cubes for the four cosmological models in the constrained case at  $z = 12, 10, 8$  and  $6$ . I note that there are no major differences until  $z = 8$ , when more and larger ionized regions appear in the  $\sigma_8$  low model, while less and smaller ones are present in the  $\sigma_8$  high model. This leads to the  $\sigma_8$  low model being completely ionized by  $z = 6$ , while the  $\sigma_8$  high model still has large neutral regions. This is because, contrary to the fiducial model, in these cases the choice of the parameters controlling the astrophysical processes has an impact stronger than the one of  $\sigma_8$  on the matter clustering. However, the differences between the astrophysical parameters of the  $h$  high and the fiducial model are smaller, so that the differences in the size of the ionized regions from  $z = 10$  is mainly driven by matter clustering. Nevertheless, the choice of both astrophysical and cosmological parameters leads to significantly different speeds at which reionization happens across all four models. Indeed, despite the relatively similar maps observed at  $z = 8$ , by  $z = 6$  the differences are much more evident. For completeness, in Appendix D I also show the B-series middle slices.

In Figure 5.4, I present the redshift evolution of the average of the A and B-series volume-averaged neutral hydrogen fraction,  $\langle x_{\text{HI}} \rangle$ , in the unconstrained (top) and constrained (bottom) cases, together with observational constraints. As in some models the redshift of reionization is not properly captured due to the coarse redshift resolution<sup>1</sup>, I additionally provide a fit to these curves to estimate more accurately the end of reionization. I note that in the unconstrained case, all models reionize at different redshifts. Reionization is the fastest (at  $z = 6.4$ ) in the  $\sigma_8$  high model, as structure formation happens earlier. The  $\sigma_8$  low model is still only 45% reionized by  $z = 5$  because of the lack of sources, and shows the poorest agreement with observational constraints. In the constrained case, the differences are reduced, but still significant. In particular, I note that the  $\sigma_8$  low model reionizes the fastest at  $z = 5.75$ , while the  $\sigma_8$  high model is only  $\approx 65\%$  ionized by  $z = 5$ . This reversal of reionization histories with respect to the unconstrained case is due to the impact of different astrophysical parameters. Specifically, the significantly higher  $E_{\text{SN}}$  in the  $\sigma_8$  high model blows away gas and thus suppresses star formation, which consequently reduces the production of ionizing photons. On the other hand, the higher  $\alpha_{\text{SF}}$  and  $\alpha_{\text{SF,burst}}$  in the  $\sigma_8$  low model along with the lower  $E_{\text{SN}}$  significantly boosts star formation.

As the four models reionize at different redshifts and the end of the EoR can be observationally constrained (e.g., detections of Gunn-Peterson troughs in the Lyman- $\alpha$  forest as shown in Becker et al. 2015, Qin et al. 2021 and Bosman et al. 2022), this in turn limits the possible choice of astrophysical and cosmological parameters. I note that to model more realistically the final phases of reionization, the unresolved Lyman limit systems which govern absorption during the final stages of the reionization process should be accounted for (see Georgiev et al., 2024; Giri et al.,

<sup>1</sup>While I have a large number of snapshots as discussed in Section 5.2.1 for the sake of simplicity, to run POLAR I adopt a uniform step size of  $\Delta z = 0.5$ .

2024). These, though, are not included here.

Finally, the four models produce a Thomson scattering optical depth in the range (0.041-0.067) and (0.048- 0.060) for the unconstrained and constrained case respectively, assuming an Helium (He) II fraction equal to the HII fraction, and an instantaneous HeII reionization at  $z = 3$ . These numbers are in agreement with *Planck* observations (Planck Collaboration et al., 2020; de Belsunce et al., 2021; Giarè et al., 2024).

## 5.3 Results

In this section, I present the results from my simulations with respect to various galactic and IGM properties for the unconstrained and constrained case. To reduce the effect of cosmic variance, I always consider the A and B-series averages when comparing to observations.

### 5.3.1 Unconstrained case

In the unconstrained case, I discuss the galactic and IGM observables for models with the UVLFs reported in the top panel of Figure 5.2.

#### 5.3.1.1 Galactic properties

In Figure 5.5, I present the mass-binned star formation rate (SFR) of galaxies in the four models at  $z = 10$  and  $9$ , with the top panel showing the unconstrained case. The solid lines are the median SFR, with the shaded regions referring to the 16th to 84th percentiles. The star formation main sequence (SFMS) in each case is additionally compared with results from various recent JWST programs represented by grey circles with a binning of  $\pm 0.25$  around the redshift. As all parameters affecting the star formation rate were kept the same across all four models, I note that the overall trends are similar, and agree with observations. The impact of matter clustering is seen at the low-mass end, where a larger clustering allows for the formation of a statistically significant sample of smaller mass galaxies in the  $\sigma_8$  high and  $h$  high models, while the smallest mass galaxies formed in the  $\sigma_8$  low case are an order of magnitude more massive. Further, at the high mass end, I note that the  $\sigma_8$  high model produces the most massive galaxies, which also have the highest SFR.

To confirm the agreement in star formation rates, in Figure 5.6 I additionally look at the global star formation efficiency (SFE) at the same redshifts with the top panel showing the unconstrained case. I define SFE as the ratio of the stellar and halo mass scaled by  $f_b = \Omega_b/\Omega_m$ , i.e.  $M_{\text{star}}/M_{\text{halo}}f_b$ , as a function of the halo mass  $M_{\text{halo}}$ . I also compare this to Spitzer observations at  $z = 7$  (Stefanon et al., 2021) and to abundance matching estimates from Tacchella et al. (2018). As mentioned above, the parameters controlling the star formation efficiency are the same across all models, and thus the corresponding global SFE is the same as well. The only difference is in the largest

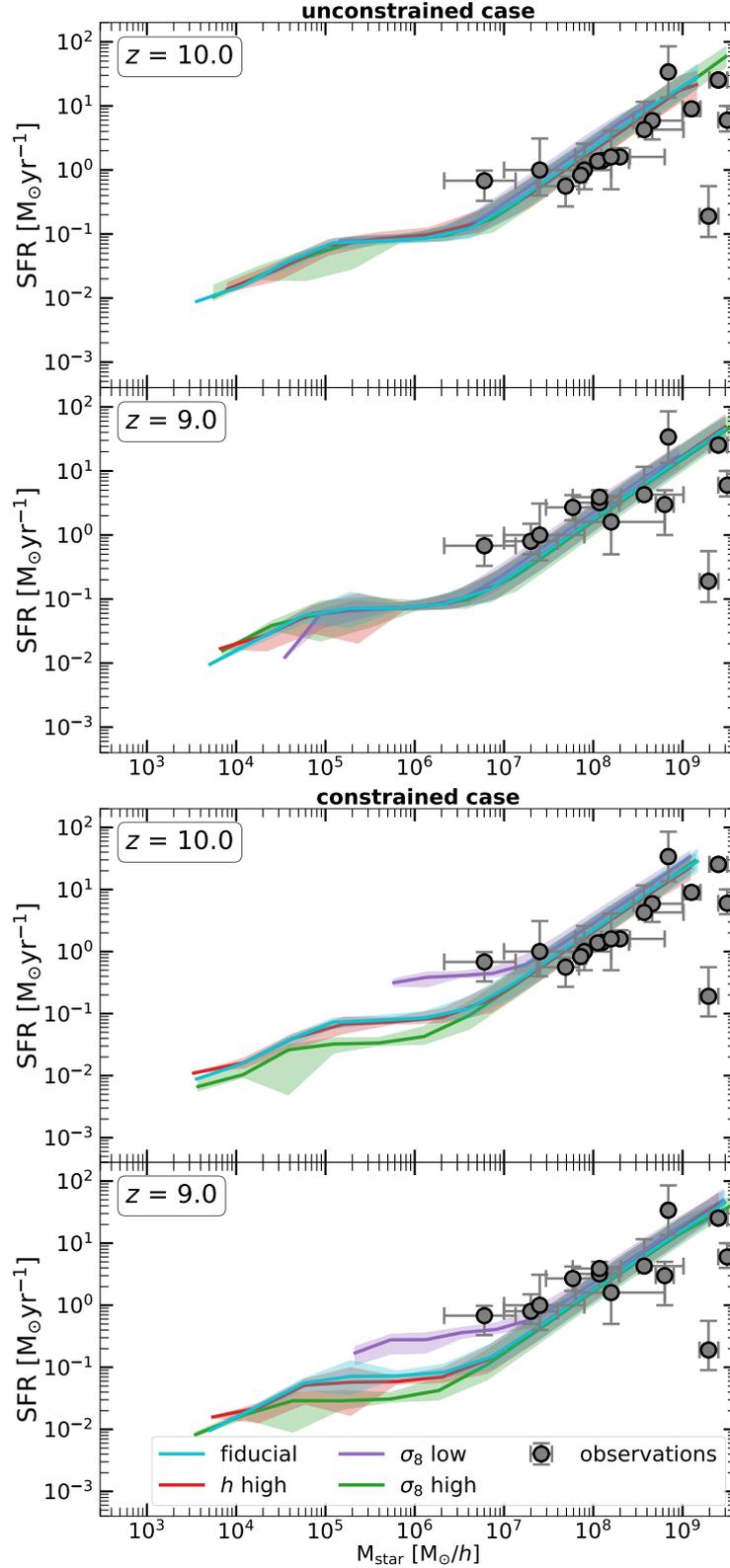


Figure 5.5: A and B-series averaged star formation rate (SFR) as a function of stellar mass for the unconstrained (top) and constrained (bottom) cases, for the fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green) models at  $z = 10$  and 9. Solid lines refer to the median star formation rate in each mass bin, with 16th and 84th percentiles shown as shaded regions. Grey circles are observations from various JWST programs (Treu et al., 2023; Fujimoto et al., 2023; Looser et al., 2023; Bouwens et al., 2023b; Papovich et al., 2023; Arrabal Haro et al., 2023b,a; Long et al., 2023; Leethochawalit et al., 2023; Atek et al., 2023; Robertson et al., 2023; Heintz et al., 2023a,b; Jin et al., 2023; Helton et al., 2024; Jung et al., 2024).

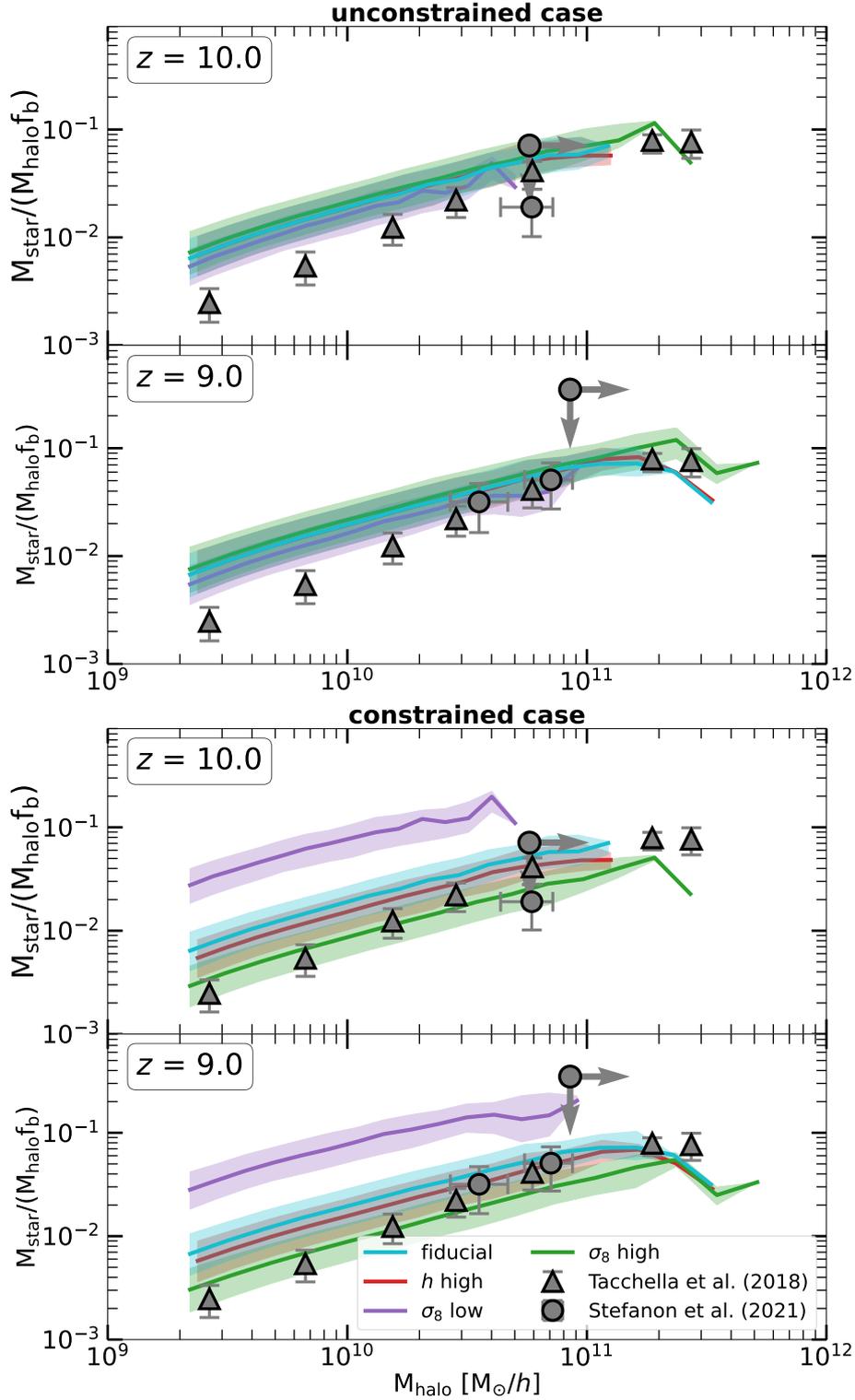


Figure 5.6: A and B-series averaged star formation efficiency (SFE) as a function of halo mass for the unconstrained (top) and constrained (bottom) cases, for the fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green) models at  $z = 10$  and 9. Solid lines refer to the median SFE in each mass bin, with 16<sup>th</sup> and 84<sup>th</sup> percentiles shown as shaded regions. Grey circles are Spitzer observations from [Stefanon et al. \(2021\)](#), and grey triangles are abundance matching estimates from [Tacchella et al. \(2018\)](#).

galaxies formed, which is directly proportional to  $\sigma_8$  as it controls the clustering of mass. I note that the SFE is marginally higher than the abundance matching estimates for the lower mass halos, but the difference is  $\lesssim 0.3$  dex. Further, the weak mass dependence of the global SFE trend is in agreement with zoom-in hydrodynamical simulations like FIREbox<sup>HR</sup> (Feldmann et al., 2024) and COLDSIM (Maio & Viel, 2023).

### 5.3.1.2 21-cm signal from the IGM

I note that despite the disagreement in UVLFs for the four cosmological models shown in the top panel of Figure 5.2, the models still reproduce fairly well observable quantities such as the SFR and global SFE, as shown in Section 5.3.1.1. Further, they are broadly in agreement with available data. I now look at IGM observables, and more specifically at the 21-cm signal power spectrum (calculated according to Equation 5.4), shown in the top panel of Figure 5.7 for the unconstrained case at  $z = 9$ . At this redshift, I note from the top panel of Figure 5.4 that the reionization history is primarily governed by the matter clustering. Thus, the  $\sigma_8$  low model having the highest  $\langle x_{\text{HI}} \rangle$  suggests a significant contrast between its large neutral and ionized regions, thus leading to a higher power amplitude up to  $k = 0.40 \text{ hcMpc}^{-1}$  as compared to the other three models. On the other hand, the greater matter clustering in the  $\sigma_8$  high and  $h$  high models leads to lower contrast at these scales due to more numerous ionized regions, and thus less power. However, at smaller scales, having more ionized regions allows for greater contrast, and thus the  $h$  high model catches up with the fiducial model, and the  $\sigma_8$  high model has the highest power overall at  $k > 0.40 \text{ hcMpc}^{-1}$ .

For a better understanding of the 21-cm signal power spectrum evolution, in Figure 5.8 I additionally present its redshift dependence at large scales, i.e.  $k = 0.15 \text{ hcMpc}^{-1}$ , with the top panel showing the unconstrained case. This allows me to compare how the 21-cm signal evolves in the four cosmological models across redshifts relevant to LOFAR observations. As suggested by the reionization histories, I note that the  $\sigma_8$  low model has a larger contrast due to slower reionization, and thus has a higher amplitude of  $\Delta_{21\text{cm}}^2$  ( $k = 0.15 \text{ hcMpc}^{-1}$ ) at the beginning of reionization. However, the  $\sigma_8$  high and  $h$  high models reach a higher amplitude by  $z \approx 8$  due to the formation of larger ionized regions.

### 5.3.2 Constrained case

As in this case I have also changed parameters that control star formation efficiency and supernova feedback, it is crucial to understand their impact on observables other than the UVLFs shown in the bottom panel of Figure 5.2.

### 5.3.2.1 Galactic properties

As in Section 5.3.1.1, in the bottom panel of Figure 5.5 I look at the A and B-series averaged SFMS and compare them with observations from JWST at  $z = 10$  and 9 for the constrained case. I note that despite differences in the energy of supernovae ( $\alpha_{\text{SF}}$  and  $\alpha_{\text{SF,burst}}$  also differ for the  $\sigma_8$  low model), the high mass ends of the SFMS of the four models still agree with each other and with observations. This is because the most massive galaxies are also the most luminous, and matching their UVLFs in Section 5.2.3 consequently matches their star formation properties. However, at the low-mass end, the SFMS of the four models shows some deviation at both redshifts. This is because at lower masses, the impact of astrophysical parameters starts to dominate, and while the  $\sigma_8$  high model forms galaxies with masses smaller than the  $\sigma_8$  low model, the median SFR of these low mass galaxies is an order of magnitude smaller than the lowest mass galaxies in the  $\sigma_8$  low model, because the impact of the energy released per supernova is sufficient to clear out gas and suppress star formation in the low-mass galaxies.

To further explore the impact of the energy released by supernovae, in the bottom panel of Figure 5.6 I also look at the global SFE for the constrained case. Here, I notice that, despite similar SFRs at the high-mass end, the efficiency of star formation in each case differs. The SFE for the  $\sigma_8$  low model is significantly boosted compared to observations, while the fiducial and  $h$  high models only mildly overpredict the observed SFE. The  $\sigma_8$  high case, on the other hand, is the best match. I note that the global SFE is inversely proportional to  $E_{\text{SN}}$  for all halo masses. The weak mass dependence trend of the SFE is the same of the unconstrained case, and in agreement with the trends of the FIREbox<sup>HR</sup> simulations (Feldmann et al., 2024) and COLDSIM (Maio & Viel, 2023) for the fiducial,  $\sigma_8$  high and  $h$  high models, while the  $\sigma_8$  low model is an order of magnitude higher but with the same slope.

### 5.3.2.2 21-cm signal from the IGM

While the UVLFs at  $M_{1600,\text{AB}} < -18$  in the constrained case are similar for all four cosmological models, the SFR and global SFE show differences, and in some cases do not match observations. In such a scenario, it becomes interesting to analyze the impact on the 21-cm signal, which is presented in the bottom panel of Figure 5.7 at  $z = 9$  for the constrained case. Unlike the unconstrained case where the differences in matter clustering contributed to differences not only in the overdensity term but also in the neutral hydrogen fraction term in equation 5.2, here the changes in astrophysical parameters lead to a drastic reduction in the differences between the value of  $\langle x_{\text{HI}} \rangle$  across the four models. Thus, the differences in the 21-cm signal power spectra of the four models are governed primarily by the overdensity term. This leads to a higher power across all wave modes for the  $\sigma_8$  high model due to greater matter clustering. While one would expect the  $h$  high model to have slightly higher power than the fiducial one due to the dependence on overdensity, the neutral-fraction term in Equation 5.2 reduces the contrast due to a marginally faster reionization process, so that the power in the two cases becomes similar. Even the  $\sigma_8$  low model matches the fiducial and  $h$  high models' power spectra due to a faster reionization process.

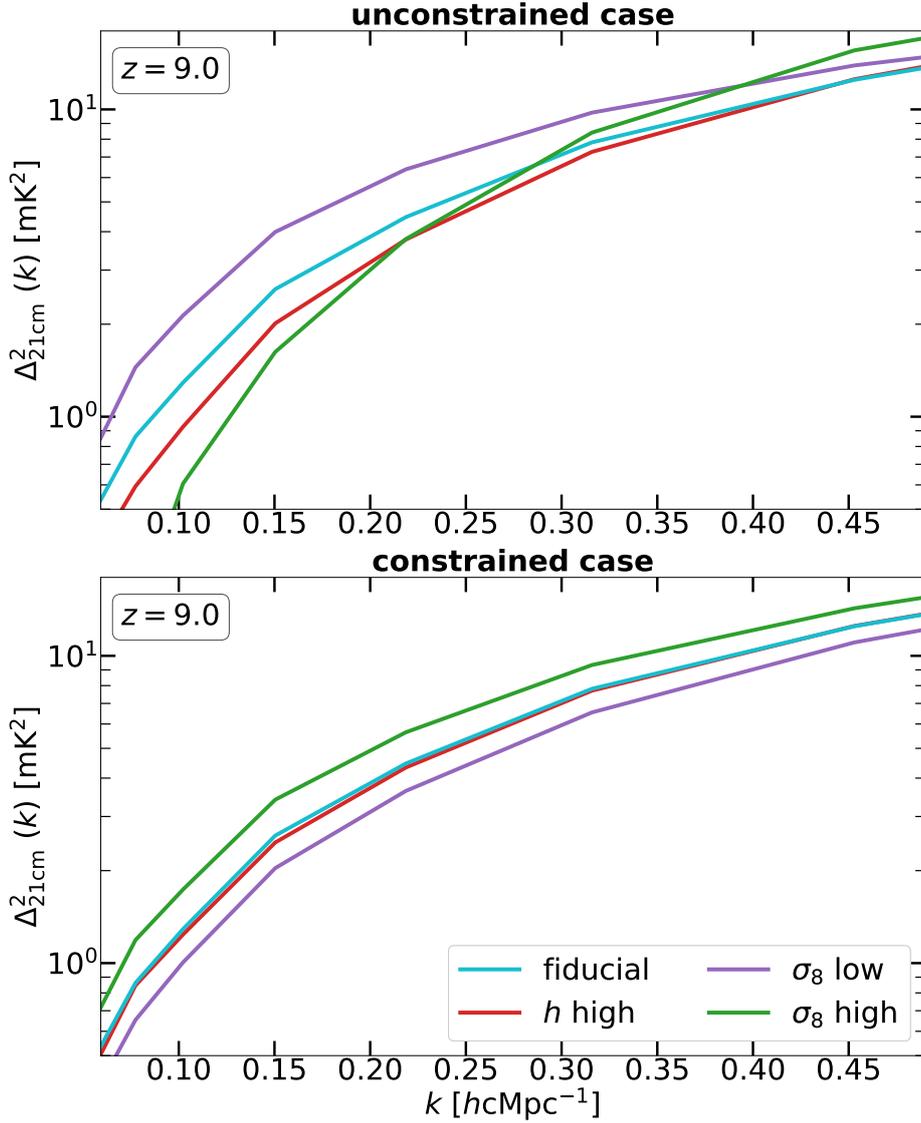


Figure 5.7: A and B-series averaged normalized 21-cm signal power spectrum ( $\Delta_{21\text{cm}}^2$ ) for the unconstrained (top) and constrained (bottom) cases, for the fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green) models at  $z = 9$ .

In the bottom panel of Figure 5.8, I present the redshift evolution of the 21-cm signal power spectrum at  $k = 0.15 \text{ hcMpc}^{-1}$  for the constrained case and notice a significant reduction in the differences of  $\Delta_{21\text{cm}}^2(k = 0.15 \text{ hcMpc}^{-1})$  values between the various cosmological models at  $12 > z > 5$  compared to the extent of differences in the unconstrained case (top panel of Figure 5.8). Indeed, at  $10 > z > 8.5$ , where LOFAR observations are focused, variations in the  $\sigma_8$  parameter lead to differences in the power spectrum of only a few  $\text{mK}^2$ . While the exact magnitude of the difference is subject to the position of the peak of  $\Delta_{21\text{cm}}^2(k = 0.15 \text{ hcMpc}^{-1})$ , which in turn is governed by modeling assumptions like the choice of  $f_{\text{esc}}$  (see Appendix E for

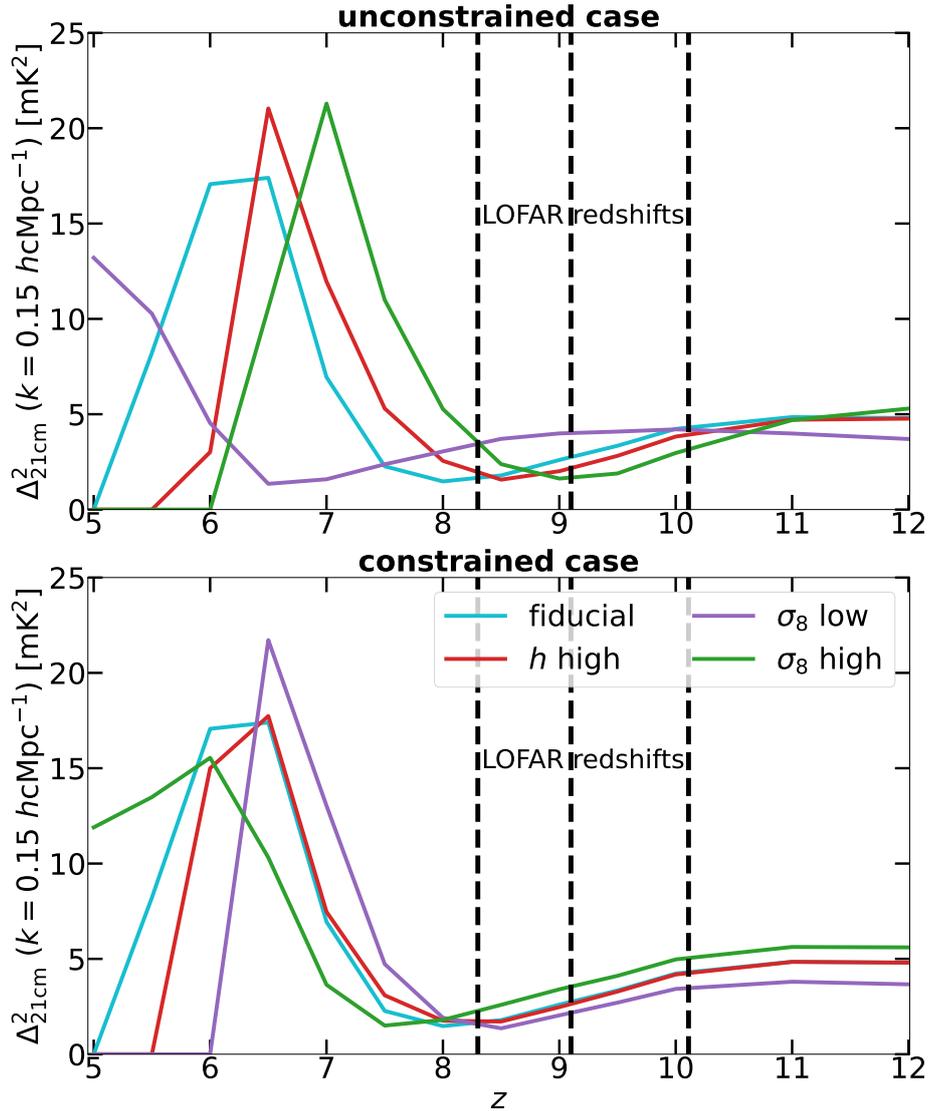


Figure 5.8: Redshift evolution of the A and B-series averaged normalized 21-cm signal power spectrum ( $\Delta_{21\text{cm}}^2$ ) at  $k = 0.15 \text{ hcMpc}^{-1}$  for the unconstrained (top) and constrained (bottom) cases, for the fiducial (cyan),  $h$  high (red),  $\sigma_8$  low (purple) and  $\sigma_8$  high (green) models. The vertical dashed lines indicate the redshifts relevant for LOFAR, i.e.  $z = 10.11, 9.16,$  and  $8.3$ .

a comparison between  $f_{\text{esc}} = 12.5\%$  and  $25\%$ ), the key takeaway is that *significantly different choices of cosmological and astrophysical parameters can still lead to similar 21-cm signal observables, even when constrained by the UVLFs from JWST and HST observations*. Thus, to explore the parameter space of models that agree with upper limits of 21-cm signal power spectra observations at the redshifts of interest to LOFAR, one should still consider various cosmological and astrophysical parameters as free. However, jointly constraining these models with a more diverse set of galactic observables from JWST and upcoming surveys with Euclid and SPHEREx

can allow me to narrow down the choice of viable ones. Additionally, tomographic imaging of the distribution of neutral Hydrogen with the upcoming SKA-Low could also be used for additional constraints on IGM properties from images such as those shown in Figure 5.3 (see for e.g. Bianco et al., 2021).

## 5.4 Discussion

By analyzing various galactic and IGM observables, I note that the interplay of astrophysical and cosmological parameters can lead to similarities as well as differences. For example, in the constrained case, while agreement with observed UVLFs leads to similar SFR for the high-mass galaxies across all models, there are significant differences in the SFR and global SFE for low-mass galaxies. On the other hand, while in the unconstrained case the SFR and global SFE show a very similar dependence on stellar mass in all models because the parameters that regulate such quantities are the same, the UVLFs exhibit significant differences.

For the IGM, I note that the 21-cm signal power spectrum at the redshifts of interest for LOFAR is very similar across all four models in the constrained case (bottom panel, Figure 5.7). This is due to the choice of astrophysical parameters, which reduces the impact of the cosmological ones, and consequently the differences between the reionization histories in comparison to the unconstrained case. In the unconstrained case, I note significant differences in reionization histories simply because the choice of cosmological parameters changes the overall matter clustering.

These similarities in terms of 21-cm observables have a profound impact on inference modeling, as models with very different cosmological and astrophysical parameters may still produce 21-cm signal power spectra in agreement with current upper limits (for example from Mertens et al., 2020; Trott et al., 2020; HERA Collaboration et al., 2023; Acharya et al., 2024d), or possibly also with an eventual measurement. This is because, while the end of reionization may differ among models, at  $z > 7$  the extent and speed of reionization are still largely similar. This means that fixing a priori any of these parameters could exclude viable models. While introducing cosmological and astrophysical parameters as free parameters during inference modeling has been attempted in several earlier studies (see e.g. Kern et al., 2017; Schneider et al., 2023), these have used only approximate reionization simulation codes due to the significant computational costs involved in building such a high-dimensional parameter space. Indeed, in a more physical approach, it is necessary to run a large number of dark-matter only simulations with different cosmologies, each post-processed with a large set of astrophysical parameters. The problem worsens if one would like to boost the resolution of the simulations to take into account the role of mini halos (see Haiman et al., 2001; Iliev et al., 2005), dwarf galaxies (Wu & Kravtsov, 2024), or Lyman Limit Systems (Georgiev et al., 2024; Giri et al., 2024) which likely play a significant role in the EoR. While faster semi-numerical codes and emulators can be employed, they are not ideal tools to model galaxy-scale physics and radiative transfer effects.

Because of this, boosting the resolution of low-resolution simulations using analytic techniques based on smaller boxes with higher resolution has been proposed (see for example Nasirudin

et al., 2020; Barsode & Choudhury, 2024). Newer methods based on Machine Learning, such as Generative Adversarial Networks (GANs, as done in Zhang et al., 2024), bypass the requirement of a higher resolution simulation. However, it needs to be explored whether such methods are robust enough to factor in small differences in individual parameters. In subsequent work with POLAR, I intend to include resolution-boosting techniques to resolve halos and galaxies down by at least two orders of magnitude in mass. The development and implementation of such techniques will additionally allow me to create more diverse training sets of power spectra for building Machine Learning kernels for Gaussian Process Regression based signal extraction, as shown by Mertens et al. (2024) and Acharya et al. (2024b,d).

## 5.5 Summary

In this work, I have investigated the impact of different cosmological models on the 21-cm signal and some galactic properties using POLAR, which combines  $N$ -body dark matter simulations run with GADGET-4 with the semi-analytic model of galaxy formation L-GALAXIES, and the 1-D radiative transfer code GRIZZLY. I have applied the framework to four different cosmological models: a “fiducial” model, which adopts values of the cosmological parameters from Planck Collaboration et al. (2020); a “ $h$  high” model, with  $h = 0.733$  based on results from studies of Cepheid variables in the host galaxies of 42 Type Ia supernovae (Riess et al., 2022); a “ $\sigma_8$  low” case, with  $\sigma_8 = 0.702$  from an anisotropic galaxy clustering measurement analysis done by Tröster et al. (2020); and a “ $\sigma_8$  high” case, with  $\sigma_8 = 0.880$  according to recent eROSITA results (Ghirardini et al., 2024). I additionally used the Fixed & Paired approach (Angulo & Pontzen, 2016) to suppress cosmic variance by boosting the effective volumes of the simulations. For all the quantities analyzed, I took an average of the fixed initial conditions case (which I refer to as the A-series) and its corresponding pair (referred to as the B-series).

I then choose astrophysical parameters in L-GALAXIES such that the fiducial model matches UV luminosity functions from HST legacy fields and JWST programs at  $z = 10$  and  $9$ . To investigate the effect of adopting different values for  $h$  and  $\sigma_8$  on other galactic and IGM observables, I use the same astrophysical parameters also for the other three cosmological models. I refer to this as the “unconstrained” case, as by not changing astrophysical parameters for the different cosmological models the resulting UVLFs will not necessarily be consistent with the observed ones. I also built a “constrained” case, where I instead choose the astrophysical parameters such that the UVLFs obtained in each cosmological model are consistent with those observed at  $z = 10$  and  $9$ . For this, I increased the star formation efficiency and reduced the energy released per supernova in the  $\sigma_8$  low model to boost star formation and, in turn, its UVLF. On the other hand, I increased the energy released per supernova for the  $h$  high and  $\sigma_8$  high models to suppress star formation and mitigate the impact of increased matter clustering as compared to the fiducial model. For the radiative transfer calculations, I only took into account the stellar sources while modeling the spectral energy distributions and chose a global escape fraction of 12.5% across all redshifts for all four models.

Our results can be summarized as follows:

- **Reionization history:** In the unconstrained case, the values of the cosmological parameters strongly influence the matter clustering, which in turn leads to significant differences in how reionization progresses in each model. While the  $\sigma_8$  high model reionizes by  $z \approx 6.4$ , the  $\sigma_8$  low model is only 45% reionized even by  $z = 5$ . In the constrained case, instead, the impact of astrophysical parameters is significant, and the  $\sigma_8$  low model is the first to reionize, at  $z = 5.75$ , while the  $\sigma_8$  high model is only 70% reionized by  $z = 5$ .
- **Star Formation Rate (SFR):** In the unconstrained case, no significant differences are observed as the parameters controlling star formation have been kept the same. However, in the constrained case the SFR in low-mass galaxies is impacted by the required energy released per supernova, and thus the SFR is boosted for the  $\sigma_8$  low model and suppressed for the  $\sigma_8$  high model. In both cases, all models broadly agree with observations at  $z = 10$  and 9.
- **Global Star Formation Efficiency (SFE):** As the parameters controlling star formation are the same in the unconstrained case, the global SFE at all masses is similar in all models. The SFE is also in agreement with observations and abundance matching estimates. However, in the constrained case, due to a higher value of the parameter regulating the star formation efficiency, the global SFE in the  $\sigma_8$  low model is also higher. The other three models show minor differences, governed by the differences in the energy released per supernova.
- **21-cm signal power spectrum:** In the unconstrained case, the differences between the four models are not only due to the difference in overdensity but also to those in the neutral fraction and its redshift evolution. However, in the constrained case, the neutral fraction at  $z > 8$  is similar in all models, as the impact of matter clustering on the neutral fraction is canceled out by the impact of setting different values for the astrophysical parameters controlling star formation. Thus, the power spectrum at higher redshifts is mostly dictated by the overdensity term, and because of this it shows smaller differences among the four models.

Overall, I conclude that different values of cosmological and astrophysical parameters can lead to differences in some observables (e.g. low-mass SFR and global SFE), while others are largely unaffected (e.g. UVLFs and 21-cm signal power spectrum). In particular, I note that despite significantly different galactic processes and reionization histories, the 21-cm power spectra are very similar in power across  $k$ -bins and in agreement with current observational upper limits (Acharya et al., 2024d; Mertens et al., 2020; Trott et al., 2020; HERA Collaboration et al., 2023). Due to this, when doing inference modeling, it is essential to consider all cosmological and astrophysical parameters as free parameters, with other observational constraints serving as priors. While a limited exploration of astrophysical parameters can be done with excursion set algorithms, a semi-analytic model provides a significantly more rigorous and physical approach to modeling galactic properties. For varying cosmological parameters though, it is necessary to run a large number of  $N$ -body simulations to populate the prior parameter space. However, running

so many high-resolution simulations would be prohibitively expensive in terms of computational resources and time. To address this issue, in future work I intend to incorporate techniques for boosting the resolution of less costly, low-resolution simulations using either analytic or machine learning techniques. Such an implementation would also additionally allow me to build broader training sets of power spectra for the Machine Learning kernels used with Gaussian Process Regression by the LOFAR EoR Key Science Project team (as done by [Acharya et al., 2024b](#)).

Lastly, to achieve a stronger constraining power on a broadened parameter space, it is essential to jointly employ a variety of galactic observables, e.g. from JWST, Euclid, and upcoming SPHEREx, as well as tomographic imaging of neutral Hydrogen in the IGM with the future SKA-Low.



# Chapter 6

## Final Remarks and Future Prospects

*Deadlines just aren't real to me until I'm staring one in the face.*

---

- Rick Riordan

In this thesis, I have provided a two-pronged approach towards using the 21-cm line of neutral Hydrogen to study the evolution of properties of the Intergalactic Medium (IGM) during the Epoch of Reionization (EoR).

First, I focus on improvements in techniques used for extracting the 21-cm signal power spectrum from deep-field radio interferometric observations. In particular, as a part of the LOFAR EoR Key Science Project (KSP) team, I focus on improving their foreground subtraction technique using Gaussian Process Regression (GPR). Recognizing that misestimation of the 21-cm signal's covariance kernel (which serves as a “template” for the power spectrum) can cause substantial signal loss, I introduced a novel covariance kernel derived via a Variational Auto-Encoder (VAE)-based algorithm. The VAE is trained on a large suite of GRIZZLY simulations, enabling it to learn the complex covariance properties of the 21-cm signal across a range of astrophysical scenarios. Apart from tests with mock signals from GRIZZLY itself, I also tested this with various other simulations, such as CRASH and 21cmFAST. I also examined the influence of the “excess noise” component (see Section 2.4.1) and explored potential redshift dependencies (see Section 2.3.4), thereby paving the way for more robust signal extraction in future EoR experiments.

It is also crucial for us to understand the similarities and differences in results between analytical templates used in the past and this Machine Learning (ML)-based kernel when applied to real data. For this, I apply the ML-based kernel to the same  $\approx 10$  nights of data as analysed by [Mertens et al. \(2020\)](#), keeping all other components and steps of the signal extraction process the same in Chapter 3. I also show the possibility of separating the upper limits on the 21-cm signal power spectrum from the excess noise component for the first time. Thus, this ML-based approach provides one of the most promising methods of extracting the 21-cm signal power spectrum, which would be crucial to eventually detect the 21-cm signal with more hours of observation from LOFAR and/or with upcoming radio telescopes like SKA.

Being on the cusp of detection means that the theoretical predictions of the 21-cm signal power spectrum, other observables, and cross-correlations with observations at other wavelengths need to be improved in order to infer the properties of the Universe during the EoR soundly. Thus, secondly, I focus on improving the modelling of simulations used to study the 21-cm signal and cosmic reionization.

One of the major issues in the development of simulations of the 21-cm signal power spectrum is that the nature of the observed signal is such that it traditionally requires modelling of  $O(10^{7-8} \text{ Mpc}^3)$  volumes to suppress the impact of cosmic variance. To achieve similar results in smaller volumes, the Fixed & Paired (F&P) approach was proposed, which uses carefully crafted simulation pairs. However, the results are significant only for observables of large-scale features, while the 21-cm signal is affected by both large-scale (overall distribution of neutral Hydrogen in the IGM) and small-scale (growth of ionized bubbles around individual galaxies) features. Thus, in Chapter 4, I tested the applicability of and improvement granted by this technique to different observables of the 21-cm signal from the EoR. I employed radiation-magneto-hydrodynamics simulations to ensure the most realistic physical description of this epoch, greatly improving over previous studies using a semi-numerical approach without accurate galaxy formation physics and radiative transfer. Through this, I explored the statistical improvement granted by the F&P technique on predictions of the skewness, power spectrum, bispectrum, and ionized regions size distribution of the 21-cm signal at redshift  $7 \leq z \leq 10$  (corresponding to  $\geq 80\%$  of the gas being neutral).

Beyond boosting effective volumes, it is also crucial to expand the parameter space of simulations used to infer astrophysical properties from this 21-cm signal. In particular, it is necessary to check if allowing some parameters to vary leads to trade-off relationships between them, such that the range of viable models that match observables broadens. This becomes even more interesting when multi-wavelength observations, such as those of the 21-cm signal and JWST observations, are still explainable with various combinations of parameters. Having these models would also be useful for broadening the training set of mock 21-cm signals used for building the ML-based kernel described earlier. Thus, in Chapter 5, I ran four models of the POLAR simulations, with differing cosmological parameters (in particular, the Hubble constant  $H_0$  and the matter clustering amplitude  $\sigma_8$ ) whose values, as provided by measurements at different redshifts, are in tension. Additionally, I vary astrophysical parameters, such as the star formation efficiency and the energy released per supernova explosion, in order to match with observations of the UV luminosity function at redshifts  $z = 10$  and  $9$  from JWST. In turn, I look at how much of a difference it leads to in other observables, including, most importantly, the 21-cm signal power spectrum at these redshifts.

Below, I have summarised the key findings of each of these projects undertaken as a part of this thesis.

## 6.1 Summary of the key results

### 1. Enhanced Foreground Removal for 21-cm Signal Extraction Using Machine Learning

The first step in this thesis was to work towards implementing a significant upgrade to the GPR framework for removing foregrounds in LOFAR observations aimed at detecting the faint 21-cm signal from the EoR in Chapter 2. Extensive tests on mock datasets by simulating noise levels equivalent to both  $\approx 10$  nights (141 hours) and  $\approx 100$  nights (1410 hours) of LOFAR observations demonstrate that the VAE-based kernel recovers the 21-cm power spectrum within  $2\sigma$  uncertainties for a signal-to-noise ratio (SNR) averaged over  $k$ -bins being  $\geq 5 \times 10^{-2}$  for  $k = [0.05, 1.00] h\text{Mpc}^{-1}$ , while better capturing both its shape and amplitude compared to traditional analytic kernels. I also find that across redshifts, the recovery is affected purely by the changes to the SNR of the 21-cm signal, with no additional significant biases added by the VAE kernel. This is supported further by the fact that better constraints on the excess noise component are shown to improve the quality of recovery.

### 2. Revised LOFAR Upper Limits on the 21-cm Signal at $z \approx 9.1$

After confirming the robustness of the ML-enhanced covariance kernel as a template for the 21-cm signal, I applied it to real LOFAR data in Chapter 3, specifically targeting the 21-cm signal at  $z \approx 9.1$  using the same data and processing pipeline as [Mertens et al. \(2020\)](#). By incorporating the ML-based covariance kernel within the GPR framework, I achieved a more accurate modeling of the data components, including the dominant foregrounds and a systematic “excess noise” term. Analyzing 141 hours of LOFAR observations, the revised pipeline yields a conservative  $2\sigma$  upper limit on the 21-cm power spectrum of  $\Delta_{21}^2 < (80)^2 \text{ mK}^2$  at  $k = 0.075 h \text{ Mpc}^{-1}$  when jointly considering the 21-cm signal and excess noise. After applying bias corrections to remove the excess noise, an even deeper limit of  $\Delta_{21}^2 < (25)^2 \text{ mK}^2$  is reported (see Table 3.2 for values across wave modes). These findings demonstrate that the ML-enhanced GPR approach not only reduces signal loss but also provides tighter constraints on the EoR 21-cm signal. However, due to the limited understanding of the nature of the excess noise component present in the data, it is yet to be confirmed if the bias-corrected upper limits obtained indeed contain the entirety of the 21-cm signal. Nevertheless, this ML-based approach provides one of the most promising methods of extracting the 21-cm signal power spectrum, which would be crucial to eventually detect the 21-cm signal with more hours of observation from LOFAR and/or with upcoming radio telescopes like SKA.

### 3. Suppressing Cosmic Variance in EoR 21-cm Signal Simulations Using the F&P Approach

Moving into modelling improvements in Chapter 4, I presented a novel technique to mitigate cosmic variance in radiation–hydrodynamic simulations of the reionization-era 21-cm signal by employing the F&P approach. Traditional methods require simulating enormous volumes to accurately capture the faint 21-cm fluctuations amid large-scale density variations. By fixing the amplitudes of the initial density perturbations to their expected values and pairing each simulation with one that uses inverted phases, the F&P method effectively boosts the statistical volume by at least a factor of 3.5 for measuring the 21-cm signal power spectrum, and by at least a factor

of 5 for measuring the equilateral bispectrum (see Sections 4.4.1 and B for the methodology to calculate these improvement factors). This improvement directly translates to an improvement in computational cost by the same factor, both in terms of time and memory. Detailed comparisons of key statistical measures (power spectrum, bispectrum, skewness, and bubble size distributions) between F&P averages and conventional simulations with randomly sampled initial conditions show that the F&P approach yields results that closely match ensemble averages and thus the expected “true” value. This enhanced simulation strategy provides a more efficient and reliable pathway to predict the 21-cm signal observables, which is critical for interpreting upcoming observations from next-generation radio interferometers.

#### 4. Exploring the effect of different cosmologies on the Epoch of Reionization 21-cm signal

I finally investigate how variations in key cosmological parameters—specifically the Hubble constant ( $H_0$ ) and the matter clustering amplitude ( $\sigma_8$ )—affect the predicted 21-cm signal from the Epoch of Reionization. Using  $N$ -body simulations run with GADGET-4, I post-process the results with the POLAR reionization pipeline, which couples the semi-analytic galaxy formation model L-GALAXIES with 1D radiative transfer via GRIZZLY. I then compared the simulated UV luminosity functions with the latest JWST observations at redshifts  $z = 10$  and  $9$  to determine the astrophysical parameters necessary for consistency and then explored the impact of different cosmologies on the 21-cm power spectrum in the redshift range probed by LOFAR ( $z \approx 8.5 - 10$ ). Despite adopting distinct sets of cosmological and astrophysical parameters, the predicted 21-cm power spectra remain consistent with current observational upper limits, highlighting the potential of combining both galactic and IGM observables to constrain the physical processes driving reionization more tightly.

## 6.2 Future prospects

In the efforts to improve the modelling of the 21-cm signal, in this thesis, I have explored improvements in effective volumes and physical complexity. However, as shown from the results in Chapter 5, it would be crucial to perform multi-wavelength parameter inference in order to put tighter constraints on the range of viable models of the Universe. For this, large simulations required for modelling the 21-cm signal will also have to have the resolution to analyze galaxy observations from JWST and upcoming missions like SPHEREx (Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer). While smaller simulation boxes can achieve higher resolution while having similar computational costs, all state-of-the-art cosmological simulations built so far tend to focus on only the brightest galaxies. The argument in support of this is simple: due to the associated computational costs, simulations only resolve up to the resolution of galaxy masses that are observable. However, this induces a bias in the sampling of galaxies. While telescopes observe only a subset of galaxies out of all galaxies present at any redshift, the corresponding simulations *only* generate a sample of the observable galaxies. Thus, the physical processes in these simulations would be biased towards not producing any of the fainter, unobserved galaxies.

The immediate question that arises is whether this matters at all. And the answer is **yes**. Analytical models show that smaller, fainter galaxies can contribute as much as 50% of ionizing photons during reionization (Wu & Kravtsov, 2024). This oversight could be one of the reasons behind recent debates on “crises in cosmology,” particularly concerning photon abundance during reionization (Davies et al., 2021; Muñoz et al., 2024).

Thus, current simulations are unable to fully reconcile theoretical predictions with observational data by design. Furthermore, the high computational cost of running these simulations requires fixing several cosmological and astrophysical parameters, thereby narrowing the range of viable early Universe models that can be tested against upcoming observational data. This limitation hampers efforts to resolve critical questions, including the Hubble parameter and  $\sigma_8$  tensions and the properties of dark matter. These challenges persist even as next-generation telescopes prepare to detect signals from the IGM and investigate the first galaxies. Because of these issues, the next step in my work will be to develop methodologies to boost the resolution without compromising on simulation box sizes and physical complexity. However, brute-forcing is also not computationally feasible. Therefore, to address these constraints, I propose developing a novel, physics-informed spatiotemporal super-resolution technique. The core principle would be to upscale inexpensive, low-resolution simulations by operating on merger trees of dark matter halos. Merger trees are just data structures that store the information on dark matter halo mergers across time, which leads to the growth of halo masses.

This upscaling would essentially increase the branching of the merger trees at each time step but can eventually also be used to extrapolate the number of time steps beyond that of the training sets. As for the exact methodology, while Generative Adversarial Networks (GANs) have been used in a similar context (e.g., Zhang et al., 2024), they often suffer from mode collapse, producing

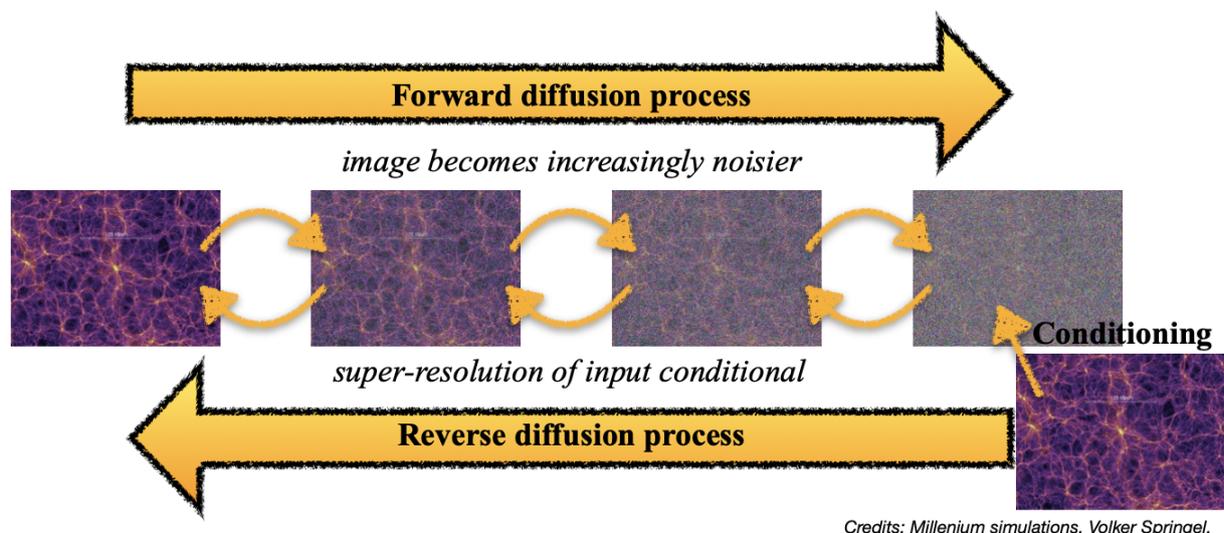


Figure 6.1: Flowchart of the proposed DDPM-based network. The input merger tree “conditions” the noise into super-resolving it via the denoising process of reverse diffusion.

insufficient diversity in generated outcomes despite variations in the input. To overcome this, I propose the use of **denoising diffusion probabilistic models** (DDPMs, which have demonstrated robustness and versatility in astrophysical applications such as [Ono et al., 2024](#); [Nguyen et al., 2024](#)). They are more stable than GANs as they produce diverse outputs while preserving the input information. Further, past methods have relied on training on particle data, requiring massive computational resources for their training data generation. Using merger trees as a training set instead reduces the computational load, as a single simulation can generate thousands of them containing the required halo data.

A DDPM works in two stages: adding noise and then removing it. In the first stage, called forward diffusion, the model gradually adds random noise to the data, making it more and more distorted until it becomes almost unrecognizable. The second stage, reverse diffusion, is where the model learns to remove the noise step by step, restoring the original data. By guiding the model with an input merger tree, I will then improve the input tree's resolution by essentially "denoising" it and generating halos beyond its resolution limit. The process is shown as a flowchart in [Figure 6.1](#). The super-resolved output will be constructed back into the format required by L-GALAXIES by rebuilding the merger trees from the super-resolved halo data. This can then, in turn, be used for the POLAR simulations.

Combining super-resolution with the boost in effective volumes provided by the F&P method enables the development of EoR simulations that are large-scale, high-resolution, and physically complex, all without incurring prohibitive computational costs. This unified approach minimizes systematic biases that arise when comparing physics inferred from different observables by offering a singular platform for interpreting multi-wavelength data. Thus, with the groundwork laid in this thesis and the planned future steps, I will not only be able to explore a broader parameter space but also provide tighter constraints on the physical properties of the Universe during cosmic reionization. Ultimately, this work allows for a comprehensive understanding of the EoR by providing deep insights into the formation of the first luminous structures and their impact on the IGM.

# Appendix A

## ML-Enhanced 21-cm Signal Extraction: Investigating a low SNR case

I use the semi-numerical code 21cmFAST (Mesinger & Furlanetto, 2007; Greig & Mesinger, 2015) to generate a mock 21-cm signal at  $z = 9.1$  within a box of length 400 cMpc. I adopt the values  $\zeta = 30\%$ ,  $R_{\text{mfp}} = 15$  cMpc and  $T_{\text{vir}}^{\text{feed}} = 5 \times 10^4$  K (see Greig & Mesinger 2015 for more details).

The power spectra recovered using the three kernels are shown in Figure A.1. As in this model the input 21-cm signal is much weaker than in those discussed in the main text, for  $\approx 10$  nights of observation ( $\langle \text{SNR} \rangle_k \approx 1.6 \times 10^{-2}$ ) the input signal is never contained within the  $2\sigma$  uncertainty bounds of the recovered signal with the exception of the highest  $k$ -bins, where the SNR is higher. When comparing with the thermal noise uncertainty, I find that all three kernels manage to provide upper limits, with the VAE kernel also recovering the overall shape. A detection is not possible because of the low SNR, with the uncertainty on the thermal noise being almost two orders of magnitude larger than the input 21-cm signal. For  $\approx 100$  nights ( $\langle \text{SNR} \rangle_k \approx 0.3$ ), the signal is within the  $2\sigma$  uncertainty bounds for all kernels, while still providing upper limits and not a detection because the SNR is still low despite the increased integration time. The VAE kernel has the tightest constraints on the uncertainty. While the VAE kernel overestimates the input signal, it does an excellent job in recovering its overall shape.

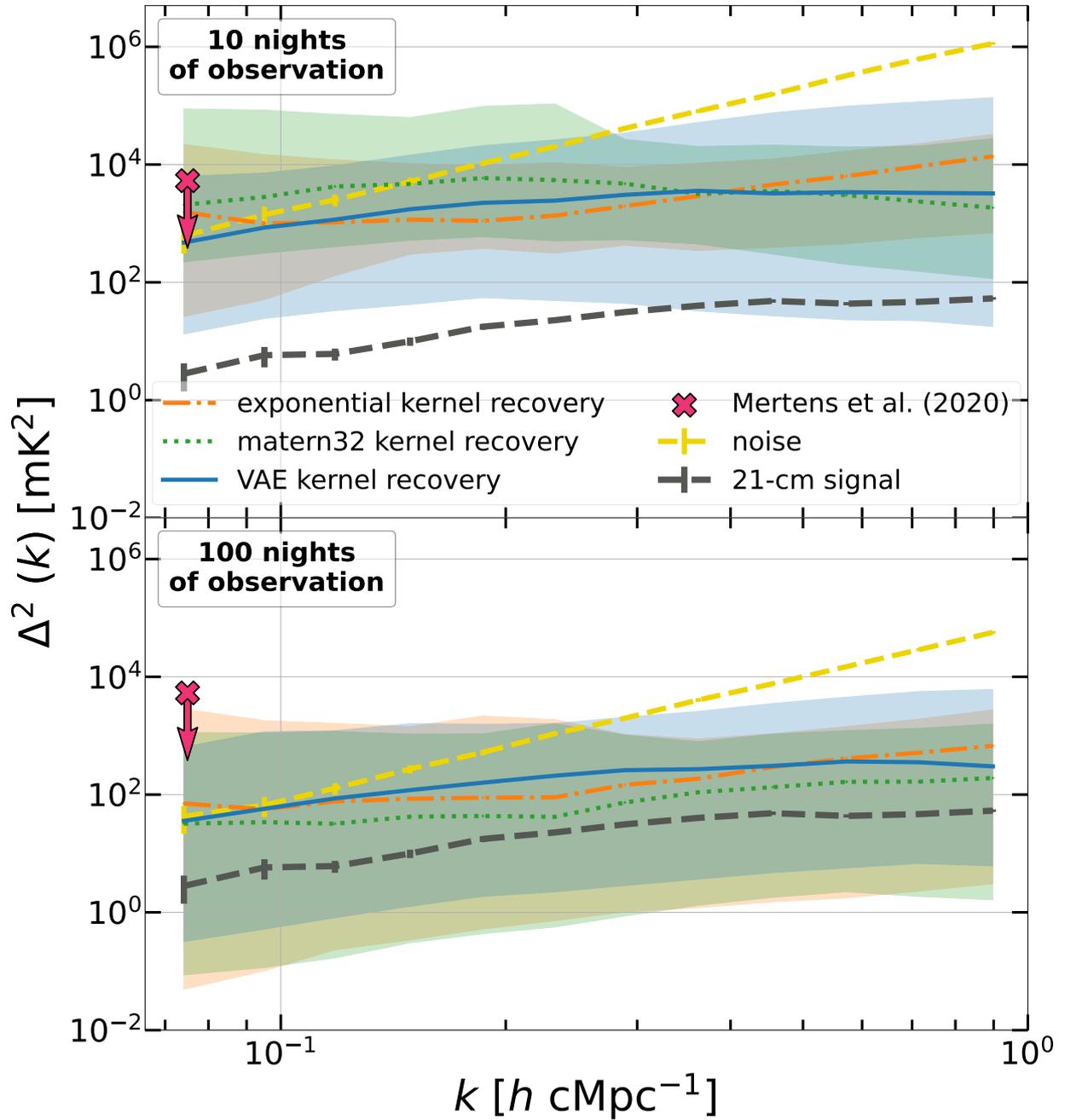


Figure A.1: As Figure 2.3 for 21cmFAST simulations of reionization.

# Appendix B

## The F&P method: Improvement factor for the equilateral bispectrum

I repeat the analysis of the improvement factor when using the F&P method as done in Section 4.4.1 for the bispectrum. While this can be performed for any type of bispectra, for the sake of simplicity here I discuss only  $B_{\text{equi}}$ . For a direct comparison with the power spectrum improvement factor, I use the same wave-mode window of  $0.15 \leq k/(h \text{ cMpc}^{-1}) \leq 0.4$ .

Further, for the standard deviation curve for  $B_{\text{equi}}$  shown in Figure B.1, I do not consider combinations of GIC simulations with  $m < 4$  or  $m > 32$  because of sampling bias, which is significant because unlike the power spectrum, the equilateral bispectrum can take negative values. This, combined with the fact that the values of  $B_{\text{equi}}$  are close to 0 mean that minor deviations from the ensemble average can lead to severe deviations in the average in cases of poor sampling. As done in Section 4.4.1, I report the minimum, maximum and average improvement factor ( $f_{\text{imp}}$ ) values in Table B.1 and depict the same graphically in Figure B.2. I note that the average improvement factor for  $B_{\text{equi}}$  is consistent across redshifts  $10 \geq z \geq 7$ .

Table B.1: As Table 4.1, showing the  $f_{\text{imp}}$  improvement factor for the equilateral triangle bispectrum ( $B_{\text{equi}}$ ).

$z$	1 - $\sigma$ away			2 - $\sigma$ away		
	$f_{\text{imp,min}}$	$f_{\text{imp,max}}$	$\langle f_{\text{imp}} \rangle$	$f_{\text{imp,min}}$	$f_{\text{imp,max}}$	$\langle f_{\text{imp}} \rangle$
10.0	2.5	17.0	5.7	6.0	18.5	9.6
8.3	2.0	15.0	7.5	6.0	17.5	11.2
7.6	2.5	12.5	5.0	5.5	16.0	9.3
7.0	0.5	17.0	7.4	3.0	18.0	10.8

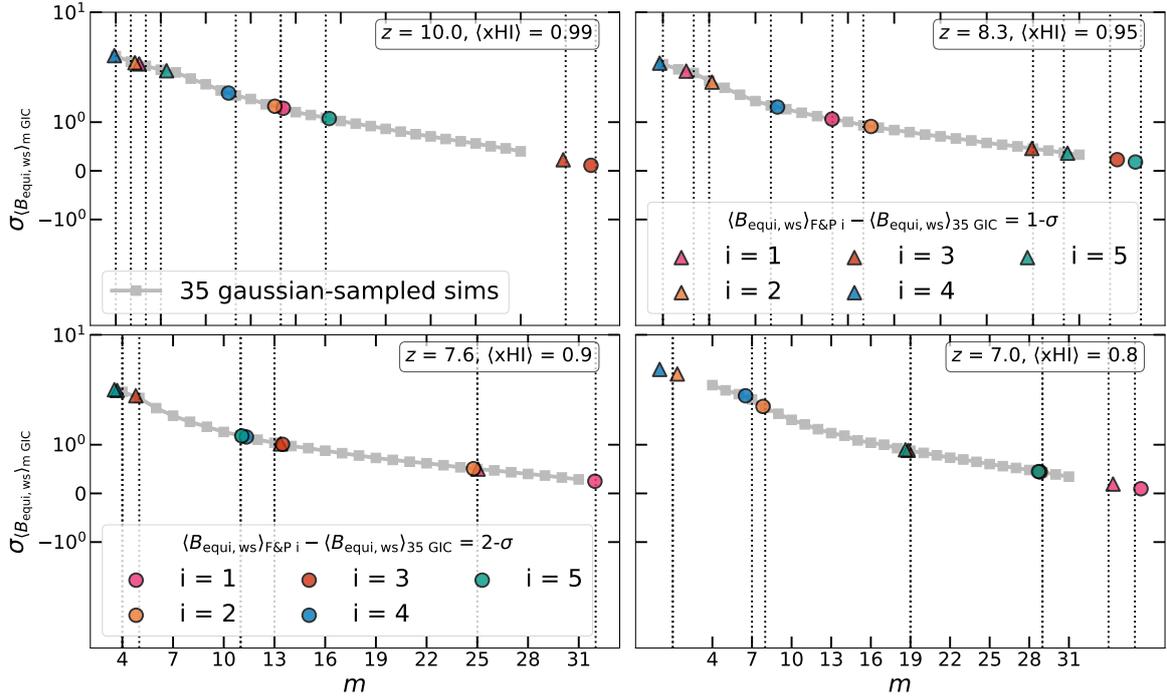


Figure B.1: As Figure 4.8 for the equilateral triangle bispectrum ( $B_{\text{equi}}$ ).

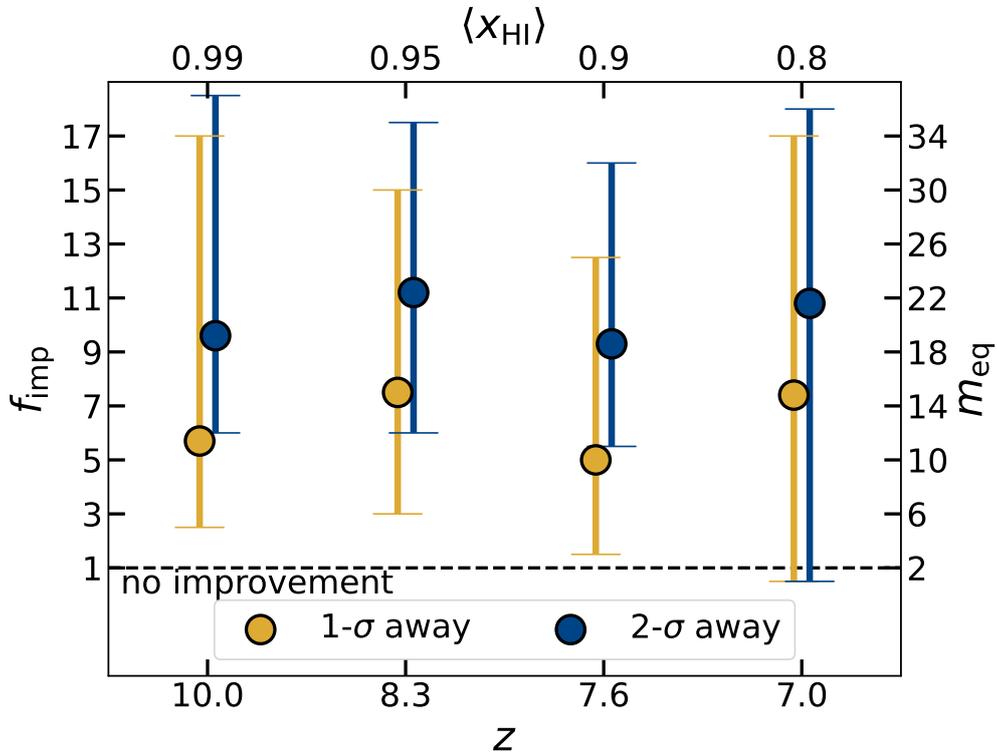


Figure B.2: As Figure 4.9 for the equilateral triangle bispectrum ( $B_{\text{equi}}$ ).

# Appendix C

## POLAR: UV Luminosity Functions across redshifts

As I tune the astrophysical parameters for the constrained case only to match the UVLFs at  $z = 10$  and 9, it is interesting to explore how the four cosmological models compare to the UVLFs observed at other redshifts. In Figure C.1, I show results at  $z = 12, 10, 9, 8, 7, 5$ . I note that while the models mildly overestimate observational data at  $z = 5$ , this could be because astrophysical parameters may need to be evolved with redshift, while here they are kept constant. Interestingly, the  $\sigma_8$  low model tends to reproduce the bright end much better at lower redshifts.

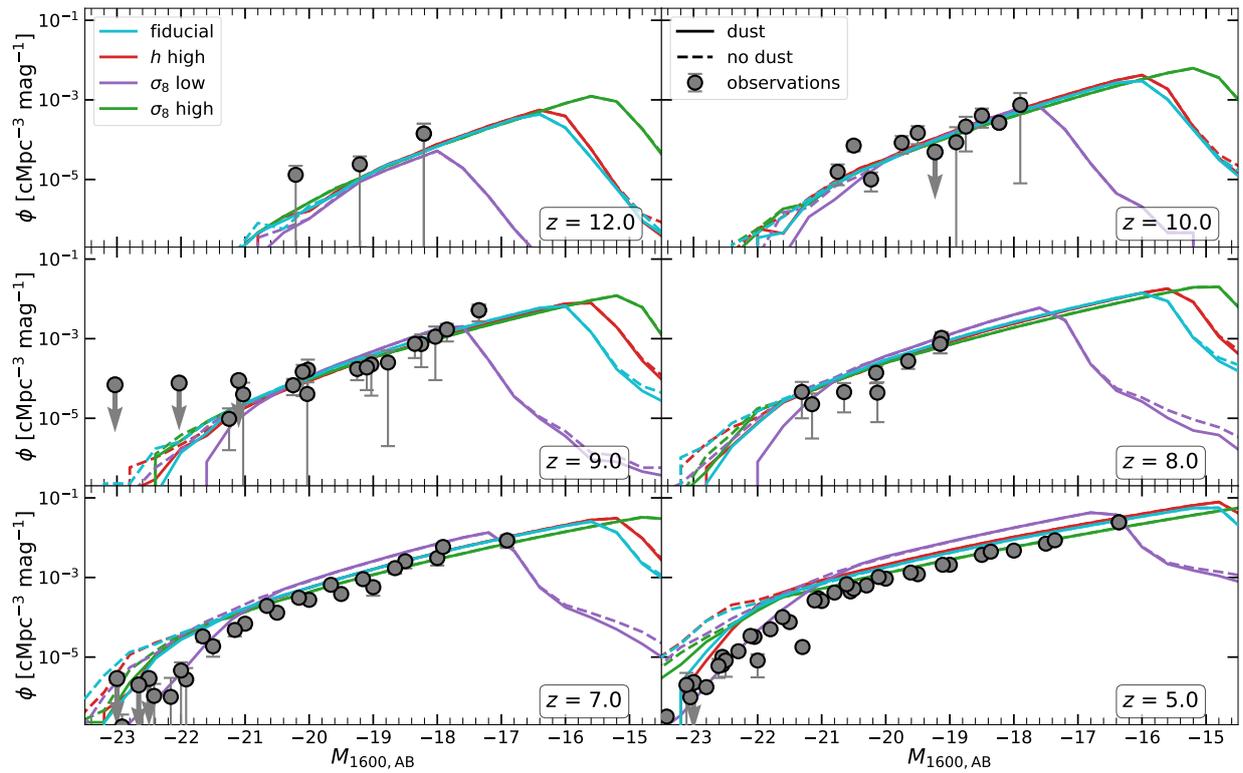


Figure C.1: As Figure 5.2, but for the constrained case UVLFs at  $z = 12, 10, 9, 8, 7$  and  $5$ .

# Appendix D

## POLAR: Maps of $\delta T_b$

In Figure D.1, I show the maps of  $\delta T_b$  of the B-series middle slices (the simulations with “paired” initial conditions in the F&P pair of simulations) for the four cosmological models in the constrained case at  $z = 12, 10, 8,$  and  $6$  (from top to bottom). Here the darkest regions represent the ionized regions with  $\delta T_b = 0$ . Note that  $\delta T_b$  cannot assume negative values, due to the assumption of  $T_S \gg T_{\text{CMB}}$ . I note that the ionized regions of the A-series middle slices shown in Figure 5.3, correspond to the most neutral regions of Figure D.1. This is because of the nature of the Fixed & Paired approach, as regions of matter clustering in the A-series should correspond to voids in the B-series by construction.

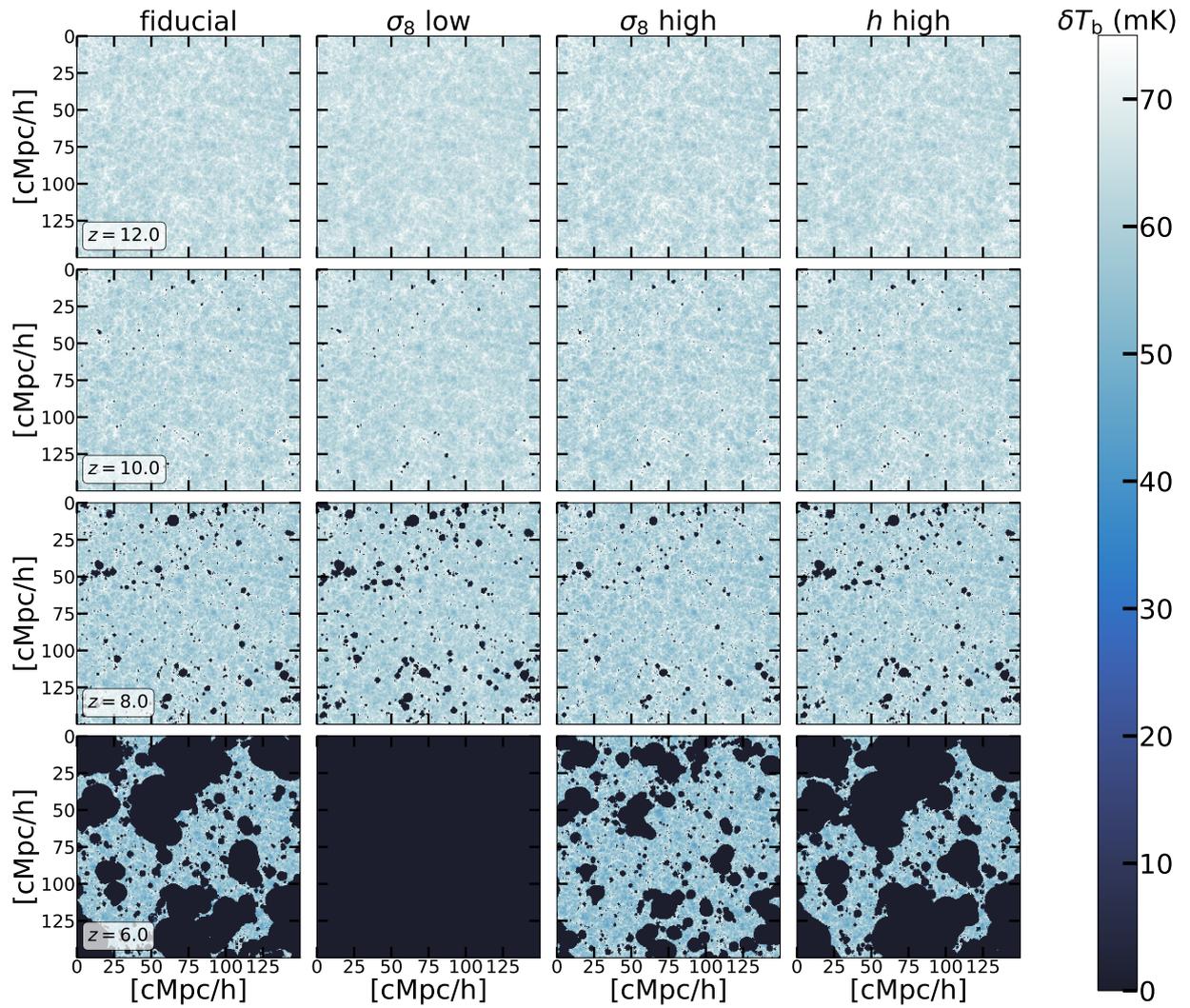


Figure D.1: Same as Figure 5.3 but for the B-series of simulations.

# Appendix E

## POLAR: Fiducial model with $f_{\text{esc}} = 25\%$

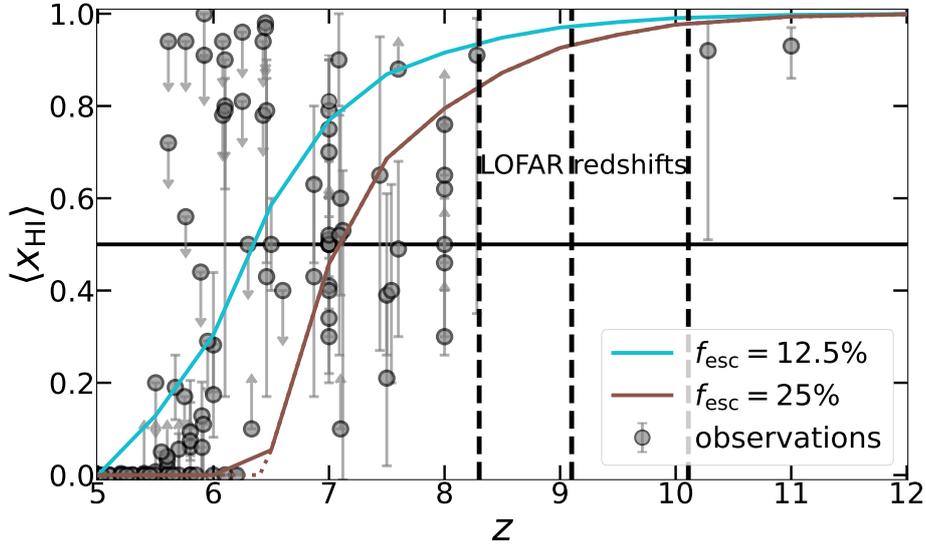


Figure E.1: As Figure 5.4, but for the fiducial model, for an escape fraction of  $f_{\text{esc}} = 12.5\%$  (cyan) and  $25\%$  (brown).

In Figure E.1, I show the redshift evolution of the average of the A and B-series volume-averaged neutral hydrogen fraction  $\langle x_{\text{HI}} \rangle$  for an escape fraction of  $f_{\text{esc}} = 12.5\%$  (cyan) and  $25\%$  (brown). I note that as expected, doubling the escape fraction leads to a faster rate of reionization, getting 50% reionized by  $z \approx 7$  for  $f_{\text{esc}} = 25\%$  as compared to  $z \approx 6.4$  for  $f_{\text{esc}} = 12.5\%$ . However, both cases are in agreement with observations, except for redshifts  $z \leq 6$ , where the lower  $f_{\text{esc}}$  case has slightly higher values. Nevertheless, it is still within the margin of error of observations. In Figure E.2 I also show the redshift evolution of the normalized 21-cm signal power spectrum ( $\Delta_{21\text{cm}}^2$ ) at  $k = 0.15 \text{ hcMpc}^{-1}$ . I note that just like the other parameters as shown in Figure 5.8, the choice of  $f_{\text{esc}}$  impacts where  $\Delta_{21\text{cm}}^2 (k = 0.15 \text{ hcMpc}^{-1})$  peaks as well as its overall trend across redshift.

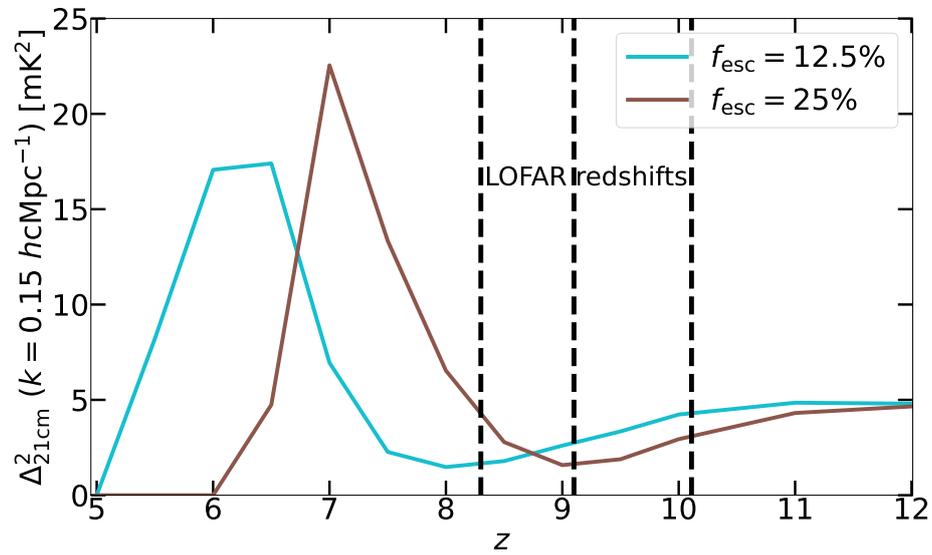


Figure E.2: As Figure 5.8, but for the fiducial model, for an escape fraction of  $f_{\text{esc}} = 12.5\%$  (cyan) and 25% (brown).

# Bibliography

- Abdurashidova Z., et al., 2022, [ApJ](#), 924, 51
- Abel T., Norman M. L., Madau P., 1999, [The Astrophysical Journal](#), 523, 66
- Abel T., Bryan G. L., Norman M. L., 2002, [Science](#), 295, 93–98
- Acharya A., et al., 2024a, [arXiv e-prints](#), p. [arXiv:2410.11620](#)
- Acharya A., et al., 2024b, [MNRAS](#), 527, 7835
- Acharya A., Garaldi E., Ciardi B., Ma Q.-b., 2024c, [MNRAS](#), 529, 3793
- Acharya A., Mertens F., Ciardi B., Ghara R., Koopmans L. V. E., Zaroubi S., 2024d, [MNRAS](#), 534, L30
- Adams N. J., et al., 2024, [ApJ](#), 965, 169
- Aigrain S., Foreman-Mackey D., 2023, [ARA&A](#), 61, 329
- Anderson L., Pontzen A., Font-Ribera A., Villaescusa-Navarro F., Rogers K. K., Genel S., 2019, [ApJ](#), 871, 144
- Angulo R. E., Pontzen A., 2016, [MNRAS](#), 462, L1
- Arrabal Haro P., et al., 2023a, [Nature](#), 622, 707
- Arrabal Haro P., et al., 2023b, [ApJ](#), 951, L22
- Atek H., et al., 2023, [MNRAS](#), 524, 5486
- Balu S., Greig B., Qiu Y., Power C., Qin Y., Mutch S., Wyithe J. S. B., 2023, [MNRAS](#), 520, 3368
- Barnes J., Hut P., 1986, [Nature](#), 324, 446
- Barrera M., et al., 2023, [MNRAS](#), 525, 6312
- Barsode A., Choudhury T. R., 2024, [arXiv e-prints](#), p. [arXiv:2407.10585](#)
- Basu A., Garaldi E., Ciardi B., 2024, [MNRAS](#), 532, 841

- Becker R. H., et al., 2001, *AJ*, 122, 2850
- Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, *MNRAS*, 447, 3402
- Bennett C. L., et al., 2003, *ApJ*, 583, 1
- Bhagwat A., Costa T., Ciardi B., Pakmor R., Garaldi E., 2024, *MNRAS*, 531, 3406
- Bianco M., Giri S. K., Iliev I. T., Mellema G., 2021, *MNRAS*, 505, 3982
- Bird S., Ni Y., Di Matteo T., Croft R., Feng Y., Chen N., 2022, *MNRAS*, 512, 3703
- Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, *MNRAS*, 416, L70
- Bosman S. E. I., et al., 2022, *MNRAS*, 514, 55
- Bouwens R. J., et al., 2015, *ApJ*, 803, 34
- Bouwens R. J., et al., 2021, *AJ*, 162, 47
- Bouwens R., Illingworth G., Oesch P., Stefanon M., Naidu R., van Leeuwen I., Magee D., 2023a, *MNRAS*, 523, 1009
- Bouwens R. J., et al., 2023b, *MNRAS*, 523, 1036
- Bromm V., Yoshida N., 2011, *ARA&A*, 49, 373
- Bromm V., Coppi P. S., Larson R. B., 1999, *The Astrophysical Journal*, 527, L5
- Bromm V., Ferrara A., Coppi P., Larson R., 2001, *Monthly Notices of the Royal Astronomical Society*, 328, 969
- Carniani S., et al., 2024, *arXiv e-prints*, p. arXiv:2405.18485
- Casavecchia B., Maio U., Péroux C., Ciardi B., 2024, *A&A*, 689, A106
- Chartier N., Wandelt B., Akrami Y., Villaescusa-Navarro F., 2021, *MNRAS*, 503, 1897
- Chornock R., Berger E., Fox D. B., Lunnan R., Drout M. R., Fong W.-f., Laskar T., Roth K. C., 2013, *ApJ*, 774, 26
- Choudhury T. R., Paranjape A., 2018, *MNRAS*, 481, 3821
- Choudhury T. R., Puchwein E., Haehnelt M. G., Bolton J. S., 2015, *MNRAS*, 452, 261
- Choudhury M., et al., 2024, *arXiv e-prints*, p. arXiv:2407.03523
- Ciardi B., Ferrara A., 2005, *Space Sci. Rev.*, 116, 625

- Ciardi B., Madau P., 2003, *ApJ*, 596, 1
- Ciardi B., Ferrara A., Marri S., Raimondo G., 2001, *MNRAS*, 324, 381
- Clément B., et al., 2012, *A&A*, 538, A66
- Copernicus N., 1543, *De revolutionibus orbium coelestium*. Johannes Petreius
- Davies F. B., et al., 2018, *ApJ*, 864, 142
- Davies F. B., et al., 2021, *ApJ*, 918, L35
- Dijkstra M., Mesinger A., Wyithe J. S. B., 2011, *MNRAS*, 414, 2139
- Dixon K. L., Iliev I. T., Mellema G., Ahn K., Shapiro P. R., 2016, *MNRAS*, 456, 3011
- Eide M. B., Graziani L., Ciardi B., Feng Y., Kakiichi K., Di Matteo T., 2018, *MNRAS*, 476, 1174
- Eide M. B., Ciardi B., Graziani L., Busch P., Feng Y., Di Matteo T., 2020, *MNRAS*, 498, 6083
- Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, *Publ. Astron. Soc. Australia*, 34, e058
- Esmerian C. J., Gnedin N. Y., 2021, *ApJ*, 910, 117
- Esmerian C. J., Gnedin N. Y., 2022, *ApJ*, 940, 74
- Esmerian C. J., Gnedin N. Y., 2024, *ApJ*, 968, 113
- Ewen H. I., Purcell E. M., 1951, *Nature*, 168, 356
- Fan X., et al., 2006a, *AJ*, 131, 1203
- Fan X., et al., 2006b, *AJ*, 132, 117
- Feldmann R., et al., 2024, *arXiv e-prints*, p. arXiv:2407.02674
- Field G. B., 1958, *Proceedings of the IRE*, 46, 240
- Field G. B., 1959, *ApJ*, 129, 551
- Finkelstein S. L., et al., 2015, *ApJ*, 810, 71
- Finlator K., Keating L., Oppenheimer B. D., Davé R., Zackrisson E., 2018, *MNRAS*, 480, 2628
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Friedman A., 1922, *Zeitschrift für Physik*, 10, 377
- Friedmann A., 1924, *Zeitschrift für Physik*, 21, 326

- Fujimoto S., et al., 2023, *ApJ*, 949, L25
- Furlanetto S. R., 2016, *The 21-cm Line as a Probe of Reionization*. Springer International Publishing, Cham, pp 247–280, doi:10.1007/978-3-319-21957-8\_9, [https://doi.org/10.1007/978-3-319-21957-8\\_9](https://doi.org/10.1007/978-3-319-21957-8_9)
- Furlanetto S. R., Oh S. P., Pierpaoli E., 2006, *Phys. Rev. D*, 74, 103502
- Galilei G., 1610, *Sidereus Nuncius*. Thomas Baglioni
- Gao L., Springel V., White S. D. M., 2005, *Monthly Notices of the Royal Astronomical Society: Letters*, 363, L66–L70
- Garaldi E., 2023, *The Journal of Open Source Software*, 8, 5407
- Garaldi E., Kannan R., Smith A., Springel V., Pakmor R., Vogelsberger M., Hernquist L., 2022, *MNRAS*, 512, 4909
- Garaldi E., et al., 2024, *MNRAS*, 530, 3765
- Gehlot B. K., et al., 2019, *MNRAS*, 488, 4271
- Georgiev I., Mellema G., Giri S. K., 2024, *arXiv e-prints*, p. arXiv:2405.04273
- Ghara R., Choudhury T. R., Datta K. K., 2015, *MNRAS*, 447, 1806
- Ghara R., Mellema G., Giri S. K., Choudhury T. R., Datta K. K., Majumdar S., 2018, *MNRAS*, 476, 1741
- Ghara R., et al., 2020, *MNRAS*, 493, 4728
- Ghara R., et al., 2024, *A&A*, 687, A252
- Ghirardini V., et al., 2024, *arXiv e-prints*, p. arXiv:2402.08458
- Giarè W., Di Valentino E., Melchiorri A., 2024, *Phys. Rev. D*, 109, 103519
- Giri S. K., Mellema G., 2021, *MNRAS*, 505, 1863
- Giri S. K., Mellema G., Dixon K. L., Iliev I. T., 2018, *MNRAS*, 473, 2949
- Giri S. K., Mellema G., Aldheimer T., Dixon K. L., Iliev I. T., 2019a, *MNRAS*, 489, 1590
- Giri S. K., D’Aloisio A., Mellema G., Komatsu E., Ghara R., Majumdar S., 2019b, *J. Cosmology Astropart. Phys.*, 2019, 058
- Giri S. K., Schneider A., Maion F., Angulo R. E., 2023, *A&A*, 669, A6
- Giri S. K., Bianco M., Schaeffer T., Iliev I. T., Mellema G., Schneider A., 2024, *MNRAS*, 533, 2364

- Glatzle M., Ciardi B., Graziani L., 2019, *MNRAS*, 482, 321
- Gnedin N. Y., 2014, *ApJ*, 793, 29
- Gnedin N. Y., Kaurov A. A., 2014, *ApJ*, 793, 30
- Godunov S. K., Bohachevsky I., 1959, *Matematičeskij sbornik*, 47(89), 271
- Goodfellow I., Bengio Y., Courville A., 2016, *Deep Learning*. MIT Press
- Graziani L., Maselli A., Ciardi B., 2013, *MNRAS*, 431, 722
- Graziani L., Ciardi B., Glatzle M., 2018, *MNRAS*, 479, 4320
- Greig B., Mesinger A., 2015, *MNRAS*, 449, 4246
- Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, *MNRAS*, 466, 4239
- Greig B., Mesinger A., Bañados E., 2019, *MNRAS*, 484, 5094
- Greig B., Trott C. M., Barry N., Mutch S. J., Pindor B., Webster R. L., Wyithe J. S. B., 2021a, *MNRAS*, 500, 5322
- Greig B., et al., 2021b, *MNRAS*, 501, 1
- Greig B., Wyithe J. S. B., Murray S. G., Mutch S. J., Trott C. M., 2022, *MNRAS*, 516, 5588
- Greig B., Prelogović D., Qin Y., Ting Y.-S., Mesinger A., 2024, *MNRAS*, 533, 2530
- Guth A. H., 1981, *Phys. Rev. D*, 23, 347
- HERA Collaboration et al., 2023, *ApJ*, 945, 124
- Haiman Z., Abel T., Madau P., 2001, *ApJ*, 551, 599
- Hamilton J. C., 2014, *Studies in the History and Philosophy of Modern Physics*, 46, 70
- Han J., Cole S., Frenk C. S., Benitez-Llambay A., Helly J., 2018, *MNRAS*, 474, 604
- Harikane Y., et al., 2022, *ApJS*, 259, 20
- Harikane Y., et al., 2023, *ApJS*, 265, 5
- Heintz K. E., et al., 2023a, *arXiv e-prints*, p. arXiv:2306.00647
- Heintz K. E., et al., 2023b, *Nature Astronomy*, 7, 1517
- Helton J. M., et al., 2024, *ApJ*, 962, 124
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, *MNRAS*, 451, 2663

- Henriques B. M. B., Yates R. M., Fu J., Guo Q., Kauffmann G., Srisawat C., Thomas P. A., White S. D. M., 2020, [MNRAS](#), **491**, 5795
- Hernández-Aguayo C., et al., 2023, [MNRAS](#), **524**, 2556
- Hirano S., Hosokawa T., Yoshida N., Umeda H., Omukai K., Chiaki G., Yorke H. W., 2014, [ApJ](#), **781**, 60
- Hirano S., Hosokawa T., Yoshida N., Omukai K., Yorke H. W., 2015, [MNRAS](#), **448**, 568
- Hirling P., Bianco M., Giri S. K., Iliev I. T., Mellema G., Kneib J. P., 2024, [Astronomy and Computing](#), **48**, 100861
- Hoag A., et al., 2019, [ApJ](#), **878**, 12
- Hogan C. J., Rees M. J., 1979, [MNRAS](#), **188**, 791
- Hosokawa T., Omukai K., Yoshida N., Yorke H. W., 2011, [Science](#), **334**, 1250
- Hosokawa T., Omukai K., Yorke H. W., 2012, [ApJ](#), **756**, 93
- Hosokawa T., Hirano S., Kuiper R., Yorke H. W., Omukai K., Yoshida N., 2016, [ApJ](#), **824**, 119
- Hothi I., et al., 2021, [MNRAS](#), **500**, 2264
- Hothi I., Allys E., Semelin B., Boulanger F., 2024, [A&A](#), **686**, A212
- Hoyle F. S., 1953, *The Astrophysical Journal*, **118**, 513
- Hubble E., 1929, [Proceedings of the National Academy of Science](#), **15**, 168
- Hutter A., Dayal P., Yepes G., Gottlöber S., Legrand L., Ucci G., 2021, [MNRAS](#), **503**, 3698
- Hutter A., Cueto E. R., Dayal P., Gottlöber S., Trebitsch M., Yepes G., 2024, [arXiv e-prints](#), p. [arXiv:2410.00730](#)
- Iliev I. T., Shapiro P. R., Raga A. C., 2005, [MNRAS](#), **361**, 405
- Iliev I. T., Mellema G., Ahn K., Shapiro P. R., Mao Y., Pen U.-L., 2014, [MNRAS](#), **439**, 725
- Jaura O., Glover S. C. O., Wollenberg K. M. J., Klessen R. S., Geen S., Haemmerlé L., 2022, [MNRAS](#), **512**, 116
- Jeans J. H., Darwin G. H., 1902, [Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character](#), **199**, 1
- Jensen H., Laursen P., Mellema G., Iliev I. T., Sommer-Larsen J., Shapiro P. R., 2013, [MNRAS](#), **428**, 1366
- Jin S., et al., 2023, [A&A](#), **670**, L11

- Jones G. C., et al., 2024, [arXiv e-prints](#), p. [arXiv:2409.06405](#)
- Jung I., et al., 2024, [ApJ](#), **967**, 73
- Kannan R., Vogelsberger M., Marinacci F., McKinnon R., Pakmor R., Springel V., 2019, [MNRAS](#), **485**, 117
- Kannan R., Garaldi E., Smith A., Pakmor R., Springel V., Vogelsberger M., Hernquist L., 2022, [MNRAS](#), **511**, 4005
- Kannan R., et al., 2023, [MNRAS](#), **524**, 2594
- Kaur H. D., Gillet N., Mesinger A., 2020, [MNRAS](#), **495**, 2354
- Kepler J., 1609, *Astronomia Nova*. Henricus Laurentius
- Kern N. S., Liu A., 2021, [MNRAS](#), **501**, 1463
- Kern N. S., Liu A., Parsons A. R., Mesinger A., Greig B., 2017, [ApJ](#), **848**, 23
- Kerrigan J. R., et al., 2018, [ApJ](#), **864**, 131
- Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, [MNRAS](#), **450**, 1349
- Kingma D. P., Welling M., 2013, [arXiv e-prints](#), p. [arXiv:1312.6114](#)
- Kingma D. P., Welling M., 2019, [arXiv e-prints](#), p. [arXiv:1906.02691](#)
- Kirch W., ed. 2008, z-Score. Springer Netherlands, Dordrecht, pp 1484–1484, [doi:10.1007/978-1-4020-5614-7\\_3826](#), [https://doi.org/10.1007/978-1-4020-5614-7\\_3826](https://doi.org/10.1007/978-1-4020-5614-7_3826)
- Klessen R. S., Glover S. C. O., 2023, [ARA&A](#), **61**, 65
- Klypin A., Prada F., Byun J., 2020, [MNRAS](#), **496**, 3862
- Kodi Ramanah D., Charnock T., Villaescusa-Navarro F., Wandelt B. D., 2020, [MNRAS](#), **495**, 4227
- Koopmans L., et al., 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*. p. 1 ([arXiv:1505.07568](#)), [doi:10.22323/1.215.0001](#)
- Kostyuk I., Nelson D., Ciardi B., Glatzle M., Pillepich A., 2023, [MNRAS](#), **521**, 3077
- Kulkarni M., Visbal E., Bryan G. L., 2019, [The Astrophysical Journal](#), **882**, 178
- Kullback S., Leibler R. A., 1951, *The annals of mathematical statistics*, **22**, 79
- Latif M. A., Whalen D., Khochfar S., 2022, [The Astrophysical Journal](#), **925**, 28

- Leethochawalit N., et al., 2023, *ApJ*, 942, L26
- Lemaître A. G., Eddington A. S., 1931, *Monthly Notices of the Royal Astronomical Society*, 91, 490
- Leung G. C. K., et al., 2023, *ApJ*, 954, L46
- Lewis J. S. W., et al., 2022, *MNRAS*, 516, 3389
- Li Y., Ni Y., Croft R. A. C., Di Matteo T., Bird S., Feng Y., 2021, *Proceedings of the National Academy of Science*, 118, e2022038118
- Linde A., 1982, *Physics Letters B*, 108, 389
- Lindley D., 1989, *Nature*, 342, 329
- Liu A., Shaw J. R., 2020, *PASP*, 132, 062001
- Long A. S., et al., 2023, *arXiv e-prints*, p. arXiv:2305.04662
- Looser T. J., et al., 2023, *arXiv e-prints*, p. arXiv:2306.02470
- Lynden-Bell D., 1967, *Monthly Notices of the Royal Astronomical Society*, 136, 101
- Ma X., et al., 2018, *MNRAS*, 478, 1694
- Ma Q.-B., Ciardi B., Eide M. B., Busch P., Mao Y., Zhi Q.-J., 2021, *ApJ*, 912, 143
- Ma Q.-B., Ghara R., Ciardi B., Iliev I. T., Koopmans L. V. E., Mellema G., Mondal R., Zaroubi S., 2023, *MNRAS*, 522, 3284
- Madau P., Meiksin A., Rees M. J., 1997, *ApJ*, 475, 429
- Maio U., Viel M., 2023, *A&A*, 672, A71
- Maio U., Péroux C., Ciardi B., 2022, *A&A*, 657, A47
- Maion F., Angulo R. E., Zennaro M., 2022, *J. Cosmology Astropart. Phys.*, 2022, 036
- Majumdar S., Pritchard J. R., Mondal R., Watkinson C. A., Bharadwaj S., Mellema G., 2018, *MNRAS*, 476, 4007
- Majumdar S., Kamran M., Pritchard J. R., Mondal R., Mazumdar A., Bharadwaj S., Mellema G., 2020, *MNRAS*, 499, 5090
- Maselli A., Ciardi B., Kanekar A., 2009, *MNRAS*, 393, 171
- Mason C. A., et al., 2018, *ApJ*, 857, L11
- McGreer I. D., Mesinger A., Fan X., 2011, *MNRAS*, 415, 3237

- McGreer I. D., Mesinger A., D’Odorico V., 2015, *MNRAS*, 447, 499
- McLeod D. J., et al., 2024, *MNRAS*, 527, 5004
- Mellema G., Iliev I. T., Alvarez M. A., Shapiro P. R., 2006, *New Astron.*, 11, 374
- Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, *MNRAS*, 478, 3640
- Mertens F. G., et al., 2020, *MNRAS*, 493, 1662
- Mertens F. G., Bobin J., Carucci I. P., 2024, *MNRAS*, 527, 3517
- Mertens F. G., et al., 2025, *A&A*, submitted.
- Mesinger A., Furlanetto S., 2007, *ApJ*, 669, 663
- Mesinger A., Furlanetto S., Cen R., 2011, *MNRAS*, 411, 955
- Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, *MNRAS*, 446, 566
- Mesinger A., Greig B., Sobacchi E., 2016, *MNRAS*, 459, 2342
- Molaro M., Davé R., Hassan S., Santos M. G., Finlator K., 2019, *MNRAS*, 489, 5594
- Mondal R., Bharadwaj S., Majumdar S., 2017, *MNRAS*, 464, 2992
- Mondal R., et al., 2020, *MNRAS*, 498, 4178
- Morales M. F., Wyithe J. S. B., 2010, *ARA&A*, 48, 127
- Morales M. F., Beardsley A., Pober J., Barry N., Hazelton B., Jacobs D., Sullivan I., 2019, *MNRAS*, 483, 2207
- Mortlock D., 2016, *Quasars as Probes of Cosmological Reionization*. Springer International Publishing, Cham, pp 187–226, doi:10.1007/978-3-319-21957-8\_7, [https://doi.org/10.1007/978-3-319-21957-8\\_7](https://doi.org/10.1007/978-3-319-21957-8_7)
- Mortlock D. J., et al., 2011, *Nature*, 474, 616
- Muñoz J. B., et al., 2024, *arXiv*, p. arXiv:2404.07250
- Mukhanov V. F., Chibisov G. V., 1981, *ZhETF Pisma Redaktsiiu*, 33, 549
- Muller C. A., Oort J. H., 1951, *Nature*, 168, 357
- Munshi S., et al., 2024, *A&A*, 681, A62
- Murray S., 2014, HMF: Halo Mass Function calculator, Astrophysics Source Code Library, record ascl:1412.006

- Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, *MNRAS*, 462, 250
- Nasirudin A., Iliev I. T., Ahn K., 2020, *MNRAS*, 494, 3294
- Nguyen T., et al., 2024, *arXiv*, p. [arXiv:2409.02980](https://arxiv.org/abs/2409.02980)
- Ocvirk P., et al., 2016, *MNRAS*, 463, 1462
- Ocvirk P., et al., 2020, *MNRAS*, 496, 4087
- Ono Y., et al., 2012, *ApJ*, 744, 83
- Ono V., et al., 2024, *ApJ*, 970, 174
- Ota K., et al., 2008, *ApJ*, 677, 12
- Ouchi M., et al., 2010, *ApJ*, 723, 869
- O'Shea B. W., Norman M. L., 2007, *The Astrophysical Journal*, 654, 66
- Papovich C., et al., 2023, *ApJ*, 949, L18
- Peacock J. A., 1999, *Cosmological Physics*
- Peebles P. J. E., 1968, *ApJ*, 153, 1
- Pentericci L., et al., 2014, *ApJ*, 793, 113
- Penzias A. A., Wilson R. W., 1965, *ApJ*, 142, 419
- Pinheiro Cinelli L., Araújo Marins M., Barros da Silva E. A., Lima Netto S., 2021, *Variational Autoencoder*. Springer International Publishing, Cham, pp 111–149, [doi:10.1007/978-3-030-70679-1\\_5](https://doi.org/10.1007/978-3-030-70679-1_5), [https://doi.org/10.1007/978-3-030-70679-1\\_5](https://doi.org/10.1007/978-3-030-70679-1_5)
- Planck Collaboration et al., 2014, *A&A*, 571, A16
- Planck Collaboration et al., 2016, *A&A*, 596, A108
- Planck Collaboration et al., 2020, *A&A*, 641, A6
- Pober J. C., et al., 2014, *ApJ*, 782, 66
- Pontzen A., Slosar A., Roth N., Peiris H. V., 2016, *Phys. Rev. D*, 93, 103519
- Press W. H., Schechter P., 1974, *Astrophys. J.*, 187, 425
- Pritchard J. R., Loeb A., 2012, *Reports on Progress in Physics*, 75, 086901
- Prole L. R., Clark P. C., Klessen R. S., Glover S. C. O., 2022, *MNRAS*, 510, 4019

- Qin Y., Mesinger A., Bosman S. E. I., Viel M., 2021, *MNRAS*, **506**, 2390
- Rasmussen C. E., Williams C. K. I., 2006, Gaussian processes for machine learning. Adaptive computation and machine learning, MIT Press
- Riess A. G., et al., 2022, *ApJ*, **934**, L7
- Robertson B. E., et al., 2013, *ApJ*, **768**, 71
- Robertson B. E., et al., 2023, *Nature Astronomy*, **7**, 611
- Rosdahl J., et al., 2018, *MNRAS*, **479**, 994
- Rosdahl J., et al., 2022, *MNRAS*, **515**, 2386
- Rybicki G. B., Lightman A. P., 1986, Radiative Processes in Astrophysics
- Santos M., Ferramacho L., Silva M., Amblard A., Cooray A., 2010a, SimFast21: Simulation of the Cosmological 21cm Signal, Astrophysics Source Code Library, record ascl:1010.025
- Santos M. G., Ferramacho L., Silva M. B., Amblard A., Cooray A., 2010b, *MNRAS*, **406**, 2421
- Saxena A., Cole A., Gazagnes S., Meerburg P. D., Weniger C., Witte S. J., 2023, *MNRAS*, **525**, 6097
- Schaeffer T., Giri S. K., Schneider A., 2023, *MNRAS*, **526**, 2942
- Schenker M. A., et al., 2013, *ApJ*, **768**, 196
- Schenker M. A., Ellis R. S., Konidakis N. P., Stark D. P., 2014, *ApJ*, **795**, 20
- Schneider R., Ferrara A., Natarajan P., Omukai K., 2002, *The Astrophysical Journal*, **571**, 30
- Schneider A., Schaeffer T., Giri S. K., 2023, *Phys. Rev. D*, **108**, 043030
- Schroeder J., Mesinger A., Haiman Z., 2013, *MNRAS*, **428**, 3058
- Scoccimarro R., 2015, *Phys. Rev. D*, **92**, 083532
- Scott D., Rees M. J., 1990, *MNRAS*, **247**, 510
- Sefusatti E., Crocce M., Scoccimarro R., Couchman H. M. P., 2016, *MNRAS*, **460**, 3624
- Shapley H., 1921, Harvard College Observatory Bulletin, **758**, 1
- Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, *A&A*, **345**, 380
- Sheth R. K., Mo H. J., Tormen G., 2001, *MNRAS*, **323**, 1

- Sinigaglia F., Kitaura F.-S., Balaguera-Antolínez A., Nagamine K., Ata M., Shimizu I., Sánchez-Benavente M., 2021, *ApJ*, **921**, 66
- Smith B. D., Sigurdsson S., 2007, *The Astrophysical Journal*, 661, L5
- Smith A., Kannan R., Garaldi E., Vogelsberger M., Pakmor R., Springel V., Hernquist L., 2022, *MNRAS*, **512**, 3243
- Sobacchi E., Mesinger A., 2015, *MNRAS*, **453**, 1843
- Springel V., 2005, *Monthly Notices of the Royal Astronomical Society*, 364, 1105–1134
- Springel V., 2010, *MNRAS*, **401**, 791
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, **328**, 726
- Springel V., et al., 2005, *Nature*, **435**, 629
- Springel V., Pakmor R., Zier O., Reinecke M., 2021, *MNRAS*, **506**, 2871
- Stanway E. R., Eldridge J. J., 2018, *MNRAS*, **479**, 75
- Stefanon M., Bouwens R. J., Labbé I., Illingworth G. D., Gonzalez V., Oesch P. A., 2021, *ApJ*, **922**, 29
- Stein M. L., 1999, *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media
- Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, *ApJ*, **868**, 92
- Tanaka T., Hasegawa K., 2021, *MNRAS*, **502**, 463
- Thomas R. M., et al., 2009, *MNRAS*, **393**, 32
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, *ApJ*, **688**, 709
- Tornatore L., Ferrara A., Schneider R., 2007, *MNRAS*, **382**, 945
- Totani T., Kawai N., Kosugi G., Aoki K., Yamada T., Iye M., Ohta K., Hattori T., 2006, *PASJ*, **58**, 485
- Tozzi P., Madau P., Meiksin A., Rees M. J., 2000, *ApJ*, **528**, 597
- Trac H., Chen N., Holst I., Alvarez M. A., Cen R., 2022, *ApJ*, **927**, 186
- Treu T., et al., 2023, *ApJ*, **942**, L28
- Tröster T., et al., 2020, *A&A*, **633**, L10

- Trott C. M., et al., 2020, *MNRAS*, 493, 4711
- Turk M. J., Oishi J. S., Abel T., Bryan G. L., 2012, *ApJ*, 745, 154
- Vani A., Ayromlou M., Kauffmann G., Springel V., 2024, *arXiv e-prints*, p. arXiv:2408.00824
- Ventura E. M., Trinca A., Schneider R., Graziani L., Valiante R., Wyithe J. S. B., 2023, *Monthly Notices of the Royal Astronomical Society*, 520, 3609–3625
- Ventura E. M., Qin Y., Balu S., Wyithe J. S. B., 2024, *Mon. Not. Roy. Astron. Soc.*, 529, 628
- Ventura E. M., Qin Y., Balu S., Wyithe J. S. B., 2025, Semi-analytic modelling of Pop. III star formation and metallicity evolution - II. Impact on 21cm power spectrum (arXiv:2502.08971), <https://arxiv.org/abs/2502.08971>
- Villaescusa-Navarro F., et al., 2018, *ApJ*, 867, 137
- Virtanen P., et al., 2020, *Nature Methods*, 17, 261
- Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, *Nature Reviews Physics*, 2, 42
- Watkinson C. A., Majumdar S., Pritchard J. R., Mondal R., 2017, *MNRAS*, 472, 2436
- Watson W. A., Iliev I. T., D’Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, *Monthly Notices of the Royal Astronomical Society*, 433, 1230–1245
- Weinberger R., Springel V., Pakmor R., 2020, *ApJS*, 248, 32
- Wollenberg K. M. J., Glover S. C. O., Clark P. C., Klessen R. S., 2020, *MNRAS*, 494, 1871
- Wouthuysen S. A., 1952, *AJ*, 57, 31
- Wu Z., Kravtsov A., 2024, *The Open Journal of Astrophysics*, 7, 56
- Xu H., Norman M. L., O’Shea B. W., Wise J. H., 2016, *The Astrophysical Journal*, 823, 140
- Yeh J. Y. C., et al., 2023, *MNRAS*, 520, 2757
- Yoshii Y., Sabano Y., 1980, *Publications of the Astronomical Society of Japan*, 32, 229
- Zaroubi S., 2013, *The Epoch of Reionization*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 45–101, doi:10.1007/978-3-642-32362-1\_2, [https://doi.org/10.1007/978-3-642-32362-1\\_2](https://doi.org/10.1007/978-3-642-32362-1_2)
- Zel’dovich Y. B., 1970, *A&A*, 5, 84
- Zhang X., Lachance P., Ni Y., Li Y., Croft R. A. C., Matteo T. D., Bird S., Feng Y., 2024, *MNRAS*, 528, 281
- de Belsunce R., Gratton S., Coulton W., Efstathiou G., 2021, *MNRAS*, 507, 1072

de Bruyn A. G., LOFAR EoR Key Science Project Team 2012, in American Astronomical Society Meeting Abstracts #219. p. 214.05

van Haarlem M. P., et al., 2013, [A&A](#), 556, A2

# Acknowledgements

I am deeply thankful to my supervisor, Benedetta Ciardi, for not only providing me with this opportunity to work at MPA but also to be a part of the wonderful LOFAR EoR KSP team. Her support and guidance for working effectively with people within and beyond our team was crucial for me to learn the ropes of collaborative research. Further, she ensured that I learned to balance hard work with necessary breaks, an invaluable lesson. Speaking of the LOFAR team, I also thank: Florent Mertens, who taught me GPR, Machine Learning, and the intricacies of observational data; Sambit Giri, for assisting with my NORDITA visiting PhD application where I developed the project with the POLAR simulations; and Qing-bo Ma and Raghunath Ghara for helping me adapt L-GALAXIES and GRIZZLY for POLAR. I am also thankful to the senior members of the collaboration, Saleem Zaroubi, Garrelt Mellema, Léon Koopmans, and Ilian Iliev for their consistent guidance and support.

At MPA, I am firstly thankful to the EoR group. In particular, I am thankful to Enrico Garaldi, who taught me not just the basics of cosmological simulations but also how to effectively communicate science. The support from everyone in the EoR group (Aniket, Arghyadeep, Benedetta Casavecchia, Ivan, Katyayani, and Shubham) has been essential for navigating the ups and downs of research. In research, I extend my thanks to Volker Springel and Rüdiger Pakmor for helping me navigate the usage of HPC systems and helping me optimize my code. I additionally thank Volker and Célin Péroux for examining my thesis and Alex and Maria W. for proofreading the German version of my thesis abstract.

Ups and downs were also there in the daily life in a new country. I would like to deeply thank the MPA secretaries: Maria, Gabi, Sonja, Cornelia, and later Solvejg, Isabel, and Lena, who ensured it was always a smooth process. I would also like to thank Sonia in HR, Claudia at the travel department, Annette at IMPRS, and Anna-Serena at the Dean's office for patiently helping me out with the various administrative tasks throughout the PhD. I am also thankful for MPA's generous travel funding support, which could make the attendance of various relevant conferences, meetings, and workshops possible. Those were essential not only for showcasing my work but also for networking with researchers from around the world.

And, of course, all of this would not be possible without the amazing community I had around me. I am thankful to Hitesh for being a great flatmate and being a familiar face right from the start. I am also thankful to Silvia for always being up to singing our favourite songs in the Common Room and to Abinaya, Akash, Arghyadeep, Géza, Joanne, and others for always being up for

afternoon coffee discussions, rants, and impromptu dinner plans. The post-lunch crossword-solving tradition started by Katyayani (and joined in by many people from the institute!) was always a great way to relax. Additionally, the “Astro-dancing” classes organised by Alena, Aleksandra, and Christoph were an amazing way to learn a new hobby, enjoy, and connect with various people from across the Garching campus. The meme group started by Géza, Ivan, and Miha was also an essential platform to unwind. I would also like to thank Abhijeet, Ilkham, Naveen, Pavan, Periklis, Vyoma, and others who helped me find my footing when I first arrived in Munich. I also extend my thanks to the other PhD students who started with me: Abinaya, Arghyadeep, Beatriz, Fulvio, Han, Hitesh, Lazaros and Matteo who reminded me that we are all on this journey together. Also, a big thanks to all the postdocs at MPA: Anna, Adam, Daniela, Deepika, Earl, Jan, Iker, Maria W., Philipp, Ryan, Seok-Jun, Taeho, Tiara, and more, whose advice was crucial through every step of this journey. I am also very thankful for Komal and for her constant support and encouragement and for constantly pushing me to make the most out of this PhD program while also taking care of my physical as well as mental health.

I would also like to thank the vibrant social culture of MPA, MPE, and ESO: the daily morning coffee discussions, the parties, the barbecues, the various talks, seminars, and discussions, and the general camaraderie among the people across the institutes provided a wonderful atmosphere to grow not just as a researcher, but also as a person. I am also thankful to the city of Munich, where I could walk, explore, and stumble upon talented street musicians whenever I needed to relax. On a similar note, I am also thankful for some of my favourite bands: Green Day, Linkin Park, and Fall Out Boy, for releasing amazing albums that I could listen to while writing this thesis.

Lastly, I would love to thank my parents and grandparents for their constant love, support, advice, and encouragement. My maternal grandmother’s figuring out how to do video calls so that she could check on me every day, even with just a 2-minute call, was often the push I needed. Their unwavering support was crucial for my successful completion of this thesis.