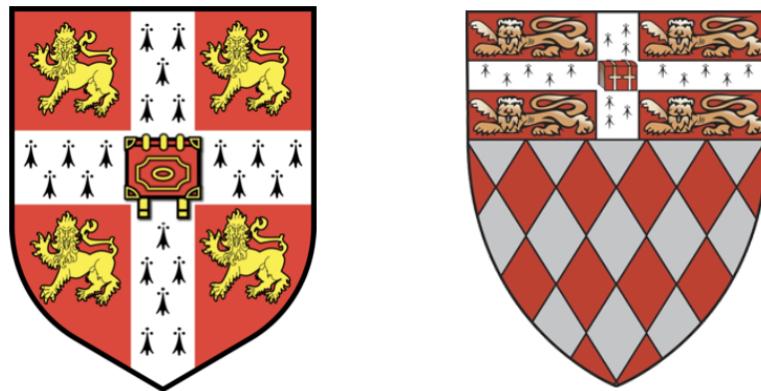


# A Machine Learning-enhanced Toolbox for Bayesian 21-cm Data Analysis and Constraints on the Astrophysics of the Early Universe



Harry Thomas Jones Bevins

Fitzwilliam College

Cavendish Astrophysics, University of Cambridge

26th June 2023

A dissertation submitted for the degree of Doctor of Philosophy



# CONTENTS

<b>Contents</b>	<b>3</b>
Declaration	7
Summary	7
Acknowledgements	8
<b>I Introduction</b>	<b>9</b>
<b>1 Introduction</b>	<b>11</b>
1.1 The significance of 21-cm Cosmology	13
1.2 21-cm Cosmology	14
1.3 Challenges facing the field	23
1.3.1 Galactic, extra-galactic foregrounds and the antenna beam	23
1.3.2 The atmosphere	27
1.3.3 Radio Frequency Interference	27
1.3.4 Unknown systematics	28
1.4 Bayesian analysis and 21-cm Cosmology	29
1.5 Machine learning and 21-cm Cosmology	33
1.6 Current experimental results	37
1.6.1 The sky-averaged 21-cm signal	38
1.6.2 The 21-cm power spectrum	40
<b>II Data Analysis Toolbox</b>	<b>43</b>
<b>2 MAXSMOOTH</b>	<b>47</b>
2.1 Introduction	48
2.2 Maximally Smooth Functions	50
2.3 Fitting Derivative Constrained Functions Using Quadratic Programming	54
2.4 Navigating Discrete Sign Spaces	57
2.4.1 Well-defined Problems and Discrete Sign Space Searches	58
2.4.1.1 The $\chi^2$ Distribution	58
2.4.1.2 The MAXSMOOTH Sign Navigating Algorithm	59
2.4.2 Ill-defined Problems and their Identification	60
2.4.3 Comparison With Basinhopping and Nelder-Mead Methods	61

<b>2.5 Applications for 21-cm Cosmology . . . . .</b>	<b>63</b>
<b>2.5.1 The Recovery of Model 21-cm Signals . . . . .</b>	<b>63</b>
<b>2.5.1.1 DCFs and 21-cm cosmology . . . . .</b>	<b>64</b>
<b>2.5.1.2 Example Residuals . . . . .</b>	<b>65</b>
<b>2.5.1.3 Identifiable 21-cm Signals and Limitations of DCFs . . . . .</b>	<b>66</b>
<b>2.5.1.4 Smooth Signal Models . . . . .</b>	<b>66</b>
<b>2.5.2 MSFs and the EDGES Data . . . . .</b>	<b>69</b>
<b>2.5.3 MSFs and the LEDA Data . . . . .</b>	<b>72</b>
<b>2.6 Conclusions . . . . .</b>	<b>75</b>
<b>3 GLOBALEMU . . . . .</b>	<b>77</b>
<b>3.1 Introduction . . . . .</b>	<b>78</b>
<b>3.2 Parameterising the Problem . . . . .</b>	<b>79</b>
<b>3.3 The Training and Test Data . . . . .</b>	<b>80</b>
<b>3.4 Data Pre-processing . . . . .</b>	<b>82</b>
<b>3.4.1 Astrophysics Free Baseline Subtraction . . . . .</b>	<b>82</b>
<b>3.4.2 Resampling of signals . . . . .</b>	<b>84</b>
<b>3.4.3 Output and Input Normalisation . . . . .</b>	<b>85</b>
<b>3.4.4 <math>x_{HI}</math> Pre-processing . . . . .</b>	<b>86</b>
<b>3.5 Neural Network Structure . . . . .</b>	<b>87</b>
<b>3.5.1 Architecture . . . . .</b>	<b>87</b>
<b>3.5.2 Loss Function and Learning rate . . . . .</b>	<b>88</b>
<b>3.5.3 Optimizer . . . . .</b>	<b>89</b>
<b>3.5.4 Activation Functions . . . . .</b>	<b>89</b>
<b>3.6 Results . . . . .</b>	<b>89</b>
<b>3.6.1 Emulation Time . . . . .</b>	<b>89</b>
<b>3.6.2 Measuring Accuracy . . . . .</b>	<b>90</b>
<b>3.6.3 Global 21-cm Signal . . . . .</b>	<b>90</b>
<b>3.6.4 Neutral Fraction . . . . .</b>	<b>92</b>
<b>3.7 Conclusions . . . . .</b>	<b>92</b>
<b>4 MARGARINE . . . . .</b>	<b>95</b>
<b>4.1 Introduction . . . . .</b>	<b>96</b>
<b>4.2 Motivation . . . . .</b>	<b>96</b>
<b>4.3 Theory . . . . .</b>	<b>97</b>
<b>4.3.1 Bayesian analysis . . . . .</b>	<b>97</b>
<b>4.3.2 Nuisance-free likelihood . . . . .</b>	<b>98</b>
<b>4.3.3 Marginal Kullback-Leibler divergence . . . . .</b>	<b>100</b>
<b>4.3.4 Marginal Bayesian Model Dimensionality . . . . .</b>	<b>100</b>
<b>4.4 Density Estimators in Practice . . . . .</b>	<b>101</b>
<b>4.4.1 Masked Autoregressive Flows . . . . .</b>	<b>101</b>
<b>4.4.2 Kernel Density Estimation . . . . .</b>	<b>101</b>

4.5 Applications of MARGARINE . . . . .	102
4.5.1 Nuisance-Free Likelihood . . . . .	102
4.5.1.1 Toy 21-cm Example . . . . .	102
4.5.1.2 Combining constraints from Planck and DES . . . . .	105
4.5.2 Prior Emulation . . . . .	106
4.5.3 Marginal Bayesian Statistics . . . . .	108
4.5.3.1 Toy Example Posteriors . . . . .	108
4.5.3.2 Radio Experiment for the Analysis of Cosmic Hydrogen . . . . .	110
4.5.3.3 Dark Energy Survey . . . . .	110
4.6 Conclusions . . . . .	111

### III Applications to Real Data 113

<b>5 Constraints on the properties of galaxies between <math>z \approx 7 - 12</math> with SARAS2</b> . . . . .	<b>117</b>
5.1 Introduction . . . . .	118
5.2 The SARAS2 data . . . . .	120
5.3 Modelling . . . . .	121
5.3.1 Noise Modelling . . . . .	122
5.3.2 Foreground Modelling . . . . .	122
5.3.3 Systematic Modelling . . . . .	124
5.3.4 Signal Modelling . . . . .	127
5.3.4.1 Standard astrophysical signals . . . . .	127
5.3.4.2 Excess radio background signals . . . . .	128
5.4 Sensitivity of the SARAS2 Data to Astrophysical Parameters . . . . .	129
5.5 Results . . . . .	130
5.5.1 Noise . . . . .	131
5.5.2 Systematics . . . . .	133
5.5.3 Disfavoured 21-cm signals from combined samples . . . . .	134
5.5.3.1 Excess Radio Background Signals . . . . .	134
5.5.3.2 Standard Astrophysical Signals . . . . .	136
5.5.3.3 Results in the context of the previous SARAS2 analysis and early HERA results	136
5.6 Conclusions . . . . .	139

### 6 Constraints on galaxies at $z \approx 15 - 25$ with SARAS3 141

6.1 Probing Cosmic Dawn . . . . .	142
6.2 Data . . . . .	143
6.3 The Cosmological 21-cm Signal . . . . .	143
6.4 Results . . . . .	144
6.5 Conclusion . . . . .	151
6.M Methods . . . . .	152
6.M.1 Noise Modelling . . . . .	152
6.M.2 Foreground Modelling . . . . .	152

6.M.3	Signal Modelling and Emulation . . . . .	153
6.M.4	Marginal KL Divergence and EDGES-like signals . . . . .	155
6.S	Supplementary Material . . . . .	156
6.S.1	Foreground Modelling: Comparison With Previous Work . . . . .	156
6.S.2	Comparison of Constraints on the High-redshift Radio Galaxy Model from SARAS2 and SARAS3 . . . . .	156
6.S.3	Synchrotron Radio Background Constraints with SARAS3 . . . . .	157
6.S.4	CMB Only Radio Background Models . . . . .	160
6.S.5	The Functional EDGES-like Posterior and Prior . . . . .	161
<b>7</b>	<b>Combining constraints from HERA and SARAS3</b>	<b>163</b>
7.1	Probing the Infant Universe . . . . .	164
7.2	Methodology . . . . .	165
7.3	Results . . . . .	167
7.4	Conclusion . . . . .	172
7.S	Supplementary materials . . . . .	173
7.S.1	Modelling the Thermal History of the Infant Universe: Power Spectrum and Global Signal Synergies . . . . .	173
7.S.2	Combining Constraints with MARGARINE . . . . .	174
7.S.3	Emulating Signals from the EoR and CD . . . . .	174
7.S.3.1	The sky-averaged 21-cm signal . . . . .	174
7.S.3.2	The 21-cm power spectrum . . . . .	175
7.S.3.3	Temperatures . . . . .	175
7.S.4	Constraints on the Astrophysical Parameters . . . . .	175
7.S.5	Constraints on the Radio and X-ray Backgrounds . . . . .	175
7.S.6	The impact of SARAS2 . . . . .	178
7.S.7	Other Power Spectrum Experiments . . . . .	180
<b>IV</b>	<b>Future work and conclusion</b>	<b>183</b>
<b>8</b>	<b>Future work and Conclusion</b>	<b>185</b>
8.1	Future work . . . . .	185
8.1.1	REACH . . . . .	185
8.1.2	Cosmology forecasts from the Dark Ages . . . . .	186
8.1.3	Breaking model dependencies and likelihood reweighting . . . . .	186
8.1.4	Exploring synergies between 21-cm cosmology and other probes of the infant Universe	187
8.2	Conclusion . . . . .	187
<b>Bibliography</b>		<b>191</b>

## Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

The antenna beam pattern shown in Fig. 1.7 was provided by John Cumner and Dominic Anstey.

Section 4.3.2 in chapter 4 was largely written by Will Handley for [1], but translated into code and visually illustrated in Fig. 4.1 and Fig. 4.2 by myself.

The top panel of Fig. 7.1 was conceptualised by Stefan Heimersheim. Section 7.S.3.2 and section 7.S.3.3 in chapter 7 were first drafted by Stefan Heimersheim for [2] and edited by myself. Similarly, section 7.S.7 and Fig. 7.S.5 were first drafted by Irene Abril-Cabezas for [2] and edited by myself. Section 7.S.7 is adapted from work submitted for Irene Abril-Cabezas master's degree.

The thesis is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

It does not exceed the prescribed word limit for the Department of Physics Degree Committee.

## Summary

The work presented in this thesis focuses on data analysis of sky-averaged (global) 21-cm experiments and is based on seven first author papers. The topic is introduced in Part I, with the scientific background of the global 21-cm signal, the challenges faced by experiments aiming to detect the signal and current experimental techniques and results being outlined. Part II introduces a series of data analysis tools that have been developed over the past few years, and Part III illustrates the application of these tools to real data sets with the aim to constrain the astrophysics of the early universe.

In Part II, chapter 2 is adapted from [Bevins et al. 2021a, 3] and discusses the development of an efficient algorithm for fitting functions with constrained derivatives, and consequently smooth properties, to data from global 21-cm experiments as a model for the smooth synchrotron and free-free Galactic and extragalactic emission. We demonstrate the application of Maximally Smooth Functions (MSFs) to data from two different global 21-cm experiments, EDGES and LEDA. Chapter 3 is based on [Bevins et al. 2021b, 4] and outlines a novel 21-cm signal emulator that significantly improves on the accuracy and runtime of the previous state-of-the-art. Finally, chapter 4 is adapted from [Bevins et al. 2022b, 5] and [Bevins et al. 2022c, 1] which outline a framework for performing marginal Bayesian analysis and efficiently combining the constraining power of different data sets. In this chapter, we demonstrate the application of the code MARGARINE to the combination of real data from Planck and DES.

In Part III, chapter 5 details the application of a variant of MSFs and the neural network signal emulator, GLOBALEMU, described in chapter 3, to data from SARAS2 and is based on [Bevins et al. 2022a, 6]. It illustrates the application of Bayesian Analysis to data from 21-cm experiments, a concept which is furthered in the following chapters, and an approach to deal with unknown systematics in data sets. This was the first time that systematics have been modelled alongside the sky-averaged 21-cm signal to derive constraints on the first galaxies. The work showed that systematics do not have to be a limiting factor in our data analysis, and pioneered concepts key to the Cambridge-based REACH experiment. Chapter 6 is adapted from [Bevins et al. 2022d, 7] and is concerned with the derivation of constraints on the properties of some of the first galaxies to form in the history of the Universe at much higher redshifts than in the previous chapter using Bayesian analysis

and GLOBALEMU. This work produced the first limits on the properties of galaxies and star formation in galaxies at  $z = 20$ . This is followed by the application of chapter 4 to data from the global 21-cm signal experiment SARAS3 and 21-cm power spectrum experiment HERA adapted from [Bevins et al. 2023, 2] in chapter 7. This was the first time a joint analysis of interferometric and global signal 21-cm data has been performed, and through the combination of the constraints from the different data sets the limits in the previous chapter are further advanced.

In Part IV, I detail future work plans including the application of Bayesian techniques to data from upcoming moon based missions for the derivation of constraints on key cosmological parameters, the future analysis of data from REACH and extensions to the work in chapter 4. I follow this with conclusions.

## Acknowledgements

Throughout my life, I have been lucky enough to have been supported by loving family members, a great group of friends and fantastic academic mentors. Without all of these people this thesis would not have happened, and I am eternally grateful to all. To those I have missed, I apologise.

Thanks first to Anna Scaife without whom I do not believe I would have reached this point. Your encouragement and support through my undergraduate study and Masters degree led me to pursue a career in academic research, for which I am grateful beyond words.

To Eloy de Lera Acedo, Anastasia Fialkov and Will Handley. Thank you for your continued patience, your guidance over the course of my PhD, your trust and your advice. I could not have asked for a better set of supervisors, and I hope one day that I can pass the advice and knowledge you have given me onto my own students.

To Lincoln Greenhill, thank you for sharing data from LEDA with me and working with me on the corresponding analysis. Similarly, thank you to Ravi Subrahmanyam and Saurabh Singh for sharing data from SARAS2 and SARAS3 with me. This PhD thesis would be far shorter without your willingness and trust.

To the members of the REACH collaboration. Thank you for creating an exciting and welcoming environment to perform research in and thank you for all the advice.

To my office mates. Thanks particularly to Dominic Anstey, Thomas Gessey-Jones, John Cumner and Stefan Heimersheim for patiently discussing the 21-cm signal, statistics and antenna design with me.

To Irene Abril-Cabezas and Karla Boxall. Thanks for being great friends and for the constant support.

To mum, dad and Reg for the unwavering emotional support, care and excitable tail wagging through the good and the bad. This one really is for you.

To my brother Howard for the tech support and enthusiastic discussions about programming and machine learning. To my brothers Joe, Sam and my sister Liz for their constant support and belief in me.

Fianlly, to my examiners, Sandro Tachella and Jackie Hewit. Thanks for the helpful discussion!

# **Part I**

# **Introduction**



CHAPTER

1

## INTRODUCTION

The field of 21-cm cosmology offers a window to the infant Universe after the formation of the Cosmic Microwave Background (CMB) and prior to the current epoch of the galaxies. We currently know very little about this period, which is divided into the Dark Ages before the first stars formed, the Cosmic Dawn when the first stars and galaxies formed and the Epoch of Reionization when the abundant neutral hydrogen in the Universe was ionized by stellar UV emission.

Although a potentially lucrative probe of the first galaxies, the 21-cm signal from neutral hydrogen in the early universe remains undetected due to complex experimental and data analysis challenges. This body of work focuses primarily on the latter and comprises two main parts.

The first part details the development of three different algorithms and techniques to improve the prospects of detecting the sky-averaged 21-cm signal from neutral hydrogen 100 - 700 million years after the Big Bang and improve our understanding of the first galaxies.

The 21-cm signal is often described as being ‘fogged’ or overwhelmed by Galactic and extragalactic foregrounds, which are five orders of magnitude bigger than the signal itself. Separation of these two components of the data is crucial, and a method for modelling the foreground is described in chapter 2 adapted from [3]. The detailed algorithm uses quadratic programming to fit functions with constrained derivatives. The algorithm is guaranteed to converge on the optimum solution and is shown to improve on traditionally used methods by a factor of two in runtime. We illustrate the application of the foreground modelling tool to data from the Experiment to Detect the Global EoR Signature (EDGES) [8] and the Large aperture Experiment to detect the Dark Ages (LEDA) [9]. The work added to the discussion surrounding the nature of the absorption feature reported as a tentative detection of the 21-cm signal in 2018 by EDGES [8]. Further, it illustrated that sinusoidal systematic structures are a prominent feature in 21-cm data sets and that these need to be appropriately accounted for where present. It also demonstrated the application of the foreground modelling technique as a tool for the detection of previously unaccounted for systematics in data sets, as discussed in [10].

The signal itself relies on a complex set of astrophysical processes that are typically modelled using semi-numerical simulations [e.g. 11–14]. While useful for gaining an understanding of how the Universe evolved during this period, these simulations can take of the order of a few hours per call, which is impractical for most kinds of data analysis. As a result, the application of neural networks, trained on the semi-numerical simulations, was proposed to speed up signal generation in computationally intensive fitting algorithms. First demonstrated in [15], chapter 3 illustrates an alternative approach to the problem of emulation and is based on [4]. The work lead to a factor of 102 improvement in emulation time and approximately a factor of two improvement in accuracy.

As more and more data becomes available, the ability to efficiently combine the constraining power of different data sets, a process which can lead to significant improvements in our understanding of the infant Universe, is becoming crucial. In chapter 4, we detail a framework for performing marginal Bayesian inference and calculating marginal Bayesian statistics using Normalising Flows and Kernel Density Estimators. The chapter is adapted from [5] and [1]. We illustrate the application of MARGARINE to the combination of constraints from Planck and the Dark Energy Survey.

The second part subsequently demonstrates the application of these tools in a Bayesian framework to data sets from several different cosmological probes (including the Shaped Antenna measurement of the background Radio Spectrum 2 and 3 (SARAS2 and SARAS3) [16–19], and the Hydrogen Epoch of Reionization Array (HERA) [20]) in an effort to improve both experimental prospects and our current understanding of the first

galaxies.

In chapter 5, the foreground modelling technique and the neural network emulator described in chapter 3 are applied to data from SARAS2 in an attempt to constrain the properties of galaxies between redshift  $z = 7 - 12$ . The analysis is based on [6] and illustrates that systematic structures in the data are not a limiting factor in our ability to perform science, provided they can be identified and modelled.

This work laid the foundations for a follow-up analysis on data from the SARAS3 instrument, which is detailed in chapter 6 and based on [7]. This analysis was the first of its kind at such high redshifts,  $z = 15 - 25$ , and has helped us learn about the X-ray and radio luminosities of the first galaxies, as well as their masses and star formation rates.

In chapter 7, we illustrate the combination of the constraining power of different data sets, as outlined in chapter 4, to data from SARAS3 and the interferometer HERA. The chapter builds on the proof of concept work which used the Planck and Dark Energy Survey data in chapter 4. The chapter, taken from [2], is the first time data from the two different types of 21-cm probes, sky-averaged and power spectrum, have ever been combined. It sets the scene for future data analysis and provides the best constraints on the properties of the first galaxies between  $z = 8 - 25$  at the time of writing.

In the following sections we introduce the field of 21-cm cosmology, the challenges that face observations, the application of Bayesian and Machine Learning techniques to the 21-cm cosmology and the state of observational results at the start of the PhD.

## 1.1 The significance of 21-cm Cosmology

Before discussing in detail the physics of the 21-cm line, it is important to motivate why we are interested in studying it. We have a well established understanding of the current universe through galaxy surveys [e.g. 21] and of the early universe through studies of the CMB [22]. A host of new and upcoming instruments, such as the Dark Energy Spectroscopic Instrument (DESI) [23] and the Atacama Cosmology Telescope [24, 25], are set to further refine our knowledge of the present and early Universe.

However, there is a gap in our knowledge about the evolution of the Universe. Outside of theoretical modelling and predictions, we know very little about how the Universe transitioned from a comparatively structureless sea of gas, dark matter and CMB photons to the structure rich zoo of galaxies we see today. The gap in our knowledge covers the formation of the first stars and galaxies, the formation of the first exotic active galaxies and binary systems and the phase transition from a neutral to ionized Universe.

Along with determining the nature of Dark Matter and Dark Energy and measuring CMB B-mode polarization, working out when the first galaxies formed and what they were like is a key question for cosmologists in the coming decades.

While the recently launched JWST [26, 27] is beginning to explore the properties of galaxies out to redshifts  $z \lesssim 13$ , 21-cm cosmology offers a window to redshifts as high as  $z \approx 200$  [28] meaning it can probe not only star formation but also cosmology and the properties of the Inter Galactic Medium (IGM) over time.

As discussed below in the introduction and throughout the thesis, detection of the 21-cm signal from the infant universe, around 200 - 1000 million years after the Big Bang, is a challenging task. First proposed by [29] and [30] a number of experimental searches for the 21-cm signal, discussed throughout the thesis, have been undertaken. The Square Kilometre Array is set to start observations in the coming decade, and tomographic

images of the 21-cm signal over time is one of its key science targets. In the meantime, efforts from radio interferometers and individual radio telescopes are being pursued to detect fluctuations in the signal and the 21-cm monopole. Significant investment is being made in the field, for example the Cambridge based REACH featured in the UK Science and Technology Facilities Council roadmap for the next decade [31] as a high priority facility. The development of novel data analysis techniques to improve our chances of observing the signature from the first stars is crucial and as discussed is the focus of this thesis.

## 1.2 21-cm Cosmology

The 21-cm line in neutral hydrogen is a potentially powerful probe of the Dark Ages, the Cosmic Dawn (CD) and the Epoch of Reionization (EoR) and is produced by the transition from one hyperfine spin state to another. The number of atoms with aligned and anti-aligned electron and proton spins is given by

$$\left(\frac{n_1}{n_0}\right) = \left(\frac{g_1}{g_0}\right) \exp\left(-\frac{T_*}{T_s}\right), \quad (1.1)$$

where  $n_i$  is the number density of atoms in the different hyperfine states,  $g_i$  is the statistical weight and the ratio  $g_1/g_0 \approx 3$ ,  $T_* = E_{10}/k_b = 68$  mK is the temperature equivalent of the energy required to transition from one state to the other and  $T_s$  is known as the spin temperature [28–30, 32–34].

The intensity of this line can subsequently be derived by solving the radiative transfer equation between two different excitation states

$$\frac{dI_\nu}{dl} = \frac{\phi(\nu)h\nu}{4\nu} [n_1 A_{10} - (n_0 B_{01} - n_1 B_{10}) I_\nu], \quad (1.2)$$

where  $dl$  is a path length,  $\nu = 1420.4$  MHz,  $A_{ij}$  and  $B_{ij}$  are the Einstein coefficients for the 21-cm transition and  $\phi(\nu)$  is the line profile chosen such that  $\int \phi(\nu)d\nu = 1$ . Writing the intensity as a brightness temperature,  $T'_b$ , at redshift  $z$  using the Rayleigh Jeans approximation to the Planck curve and subsequently substituting in the expression for  $T_s$  then gives

$$T'_b = T_s (1 - \exp(-\tau_\nu)) + T'_r \exp(-\tau_\nu), \quad (1.3)$$

where  $T'_r$  is the background radio temperature at redshift  $z$ , indicated by the prime, and  $\tau_\nu = \int \alpha ds$  is the optical depth to the 21-cm line.  $\alpha$  is the absorption coefficient given by

$$\alpha = \frac{\phi(\nu)h\nu}{4\nu} (n_0 B_{01} - n_1 B_{10}), \quad (1.4)$$

and  $\tau_\nu$  is calculated as

$$\tau_\nu = 0.009(1 + \delta)(1 + z)^{3/2} \frac{x_{HI}}{T_s} \left[ \frac{H(z)(1 + z)}{dv_{||}/dr_{||}} \right], \quad (1.5)$$

where  $x_{HI}$  is the neutral fraction of hydrogen,  $\delta$  is the fractional overdensity of baryons,  $T_s$  is in Kelvin and  $dv_{||}/dr_{||}$  is the velocity gradient along the line of sight scaled to the Hubble flow.

We measure the relative difference in the brightness temperature and the radio background on Earth as  $\delta T_b = T_b - T_r$ . This effectively involves the comparison of a hypothetical clear line of sight to the radio background and a line of sight through a hydrogen 21-cm cloud. Subtracting  $T_r$  from both sides of equation (1.3) and Taylor expanding the exponentials, since the optical depth is very small, we arrive at

$$\delta T_b = \frac{T_s - T_r}{(1 + z)} \tau_\nu. \quad (1.6)$$

Substituting in equation (1.5) we can approximate the 21-cm signal as

$$\delta T_b = 0.009(1 + \delta)(1 + z)^{1/2} x_{HI} \left(1 - \frac{T_r}{T_s}\right) \left[ \frac{H(z)(1 + z)}{dv_{||}/dr_{||}} \right]. \quad (1.7)$$

We note that in full semi-numerical simulations the exponentials are not typically Taylor expanded in an effort to be as accurate as possible. If  $T_s \ll T_r$  then the differential brightness temperature can become large and negative. The signal saturates when  $T_s \gg T_r$  and so the value of  $T_s$  is crucial.

The spin temperature is affected by the formation of the first stars, among other processes detailed below. It can be written as

$$T_s^{-1} = \frac{x_r T_r^{-1} + x_c T_k^{-1} + x_\alpha T_c^{-1}}{1 + x_r + x_c + x_\alpha}, \quad (1.8)$$

where  $x_c$  is the coupling coefficient for collisions between hydrogen atoms and other hydrogen atoms, electrons and protons.  $x_\alpha$  is the coupling coefficient for ultraviolet (UV) scattering and  $T_c$  is known as the colour temperature. Finally,  $x_r$  is the coupling coefficient between the radio background and the spin temperature, which is typically assumed to be one but can deviate under certain circumstances detailed below. If we denote the excitation and deexcitation rates for collisions as  $C_{01}$  and  $C_{10}$  then we can write

$$\frac{C_{01}}{C_{10}} = \frac{g_1}{g_0} \exp\left(\frac{-T_*}{T_k}\right). \quad (1.9)$$

Similarly, we can define the colour temperature via excitation and deexcitation rates,  $P_{01}$  and  $P_{10}$ , from UV scattering as

$$\frac{P_{01}}{P_{10}} = \frac{g_1}{g_0} \exp\left(\frac{-T_*}{T_c}\right). \quad (1.10)$$

Subsequently, we can derive equation (1.8) by assuming equilibrium and

$$n_1(C_{10} + P_{10} + A_{10} + B_{10}I_r) = n_0(C_{01} + P_{01} + B_{01}I_r), \quad (1.11)$$

where  $I_r$  is the intensity of the radio background.  $T_c$  often tends to  $T_k$  and so we can write equation (1.8) as

$$1 - \frac{T_r}{T_s} = \frac{x_r + x_c + x_\alpha}{1 + x_r + x_c + x_\alpha} \left(1 - \frac{T_r}{T_k}\right). \quad (1.12)$$

The 21-cm signal, while emitted with a rest wavelength of 21-cm, is redshifted to wavelengths of a few meters, and consequently we attempt to observe this signal using radio telescopes of different designs. We try to measure two different summary statistics of the 21-cm signal, known as the sky-averaged or global 21-cm signal and the 21-cm power spectrum. This thesis is mainly concerned with the global signal. Observations of this signal are typically attempted with a single radiometer, which produces a sky averaged measurement of both the brightness temperature of the 21-cm signal and the foregrounds (consisting of Galactic and extragalactic emission). Observations of the 21-cm power spectrum are made with interferometric arrays, and the signal quantifies how the 21-cm brightness temperature varies across angular scales on the sky and redshift.

A discussion of observational efforts in the field is given in section 1.6 and what follows is a discussion of the astrophysical processes that affect the structure of the signal. In places, we reference a common suite of parameters that form the basis of a semi-numerical simulation of the 21-cm signal between  $z \approx 5 - 50$  [11–14]. Different versions of this semi-numerical simulation are used throughout this thesis and discussed in the relevant chapters.

The semi-numerical simulations are produced on a grid of 384 comoving Mpc (cMpc) cubed with a resolution of 3 cMpc [11–14]. The density and dark matter-baryon relative velocities are evolved linearly from

the dark ages to the EoR and the simulated cube are populated with halos based on a hybrid Sheth-Tormen and Press-Schechter mass function [35]. Halos are subsequently populated with gas and stars, according to some star formation efficiency, and then the gas, radio background and spin temperatures are all evolved as a function of redshift. Each 3 cMpc pixel in the simulation has its own set of temperatures  $\{T_s, T_r, T_k\}$  mediated by various radiative processes described below that have subsequent secondary effects on the temperature values of neighbouring pixels.

After the formation of the CMB the gas temperature,  $T_k$ , and background temperature,  $T_r$ , begin to cool at different rates due to the expansion of the Universe. Relativistic material and photons cool proportionally to  $(1+z)^{-1}$  as their wavelength is increased, whereas baryonic material cools adiabatically in proportion to the rate of expansion of the Universe and  $(1+z)^{-2}$ . However, at very early times (high redshifts) Compton heating, in which energy from scattering UV CMB photons is transferred to the electron that it is being scattered off [36], drives the gas temperature to the radio background temperature. The evolution of the gas temperature is approximately governed by

$$\frac{dT_k}{dz} = \frac{10^{-7}}{\Omega_b h^2} (T_k - T_r)(1+z)^{3/2} + 2T_k(1+z)^{-1}, \quad (1.13)$$

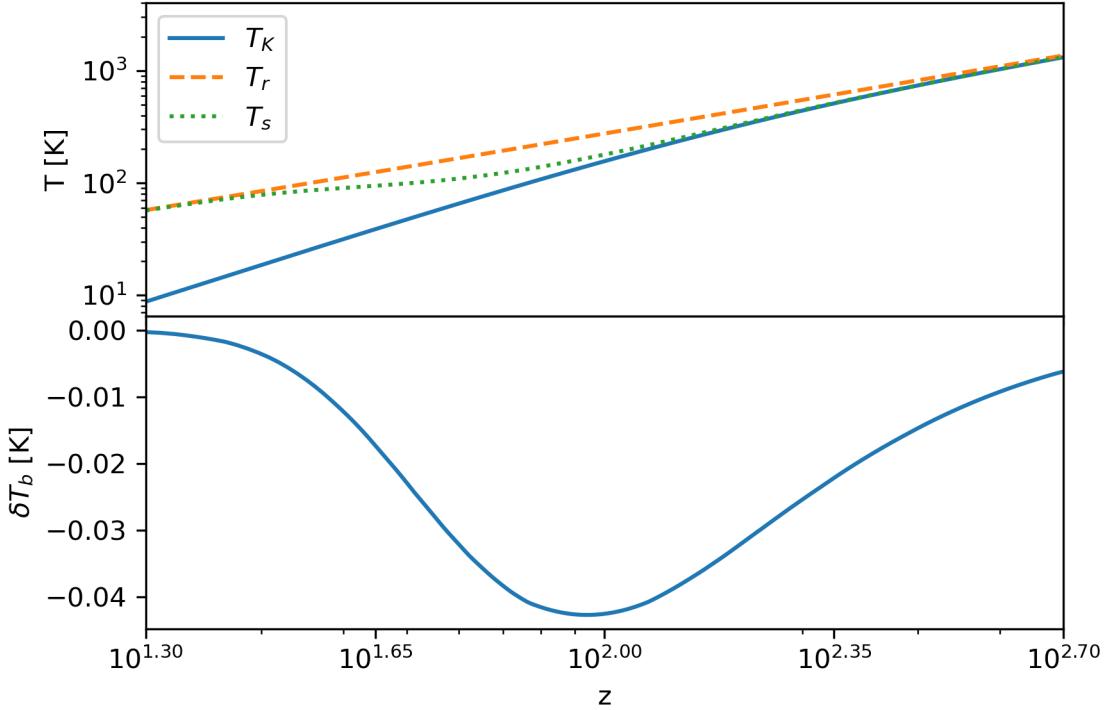
during the dark ages where  $\Omega_b$  is the baryon density parameter and  $h$  is defined by  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}$ . This approximation is valid for  $z \lesssim 200$  and more detailed modelling of Compton heating is required at higher redshifts (e.g. the 21-cm signal modelling in RECFAST++ [37]). At very high  $z$  the first term dominates the relationship and the gas temperature follows the radio background temperature. Then, at low  $z$ , since the CMB photon energy decreases with the expansion of the Universe, Compton heating becomes inefficient and the second term dominates the gas temperature. The approximate temperature evolution is shown in Fig. 1.1.

At this early time, collisional coupling drives the spin temperature to the gas temperature, which as just described is lower than the radio background temperature. This produces an absorption feature in the sky-averaged 21-cm signal shown in Fig. 1.1. The collisional coupling coefficient is given by

$$x_c = \frac{\kappa_{10} n_H T_*}{A_{10} T_r}, \quad (1.14)$$

where  $\kappa_{10}$  is the rate coefficient for spin deexcitations in collisions between neutral hydrogen atoms and is a function of the gas temperature,  $T_k$ , and  $n_H$  is the neutral hydrogen number density. This process is straightforward to model with analytic solutions. Eventually, as the density of the gas reduces, collisions become more infrequent and the spin temperature decouples from the gas and rises back up to the radio background temperature. Collisional coupling will introduce a peak in the 21-cm power spectrum at high- $z$  on scales corresponding to the mean free path for collisions. This all happens prior to the formation of the first galaxies, and a detection of this signal could inform us about key cosmological parameters. The detection of the dark ages signal is the subject of many upcoming missions to the Moon [38] and is not explored significantly in this work.

Once the first stars begin to form around  $z \approx 25 - 50$ , the spin temperature is expected to couple once again to the gas temperature via the Wouthuysen-Field effect [39, 40]. In this process, neutral hydrogen atoms absorb and re-emit Lyman- $\alpha$  photons from the first stars which leads to a change in the hyperfine state of the atom and consequently in the spin temperature. The low energy  $1_0 S_{1/2}$  hyperfine state, with anti-aligned electron and proton spins, is excited to one of the two central  $2P$  states through the absorption of a Lyman- $\alpha$  photon followed by emission and deexcitation to either of the  $n = 1$  states. This process is illustrated in Fig. 1.2.



**Figure 1.1:** The dark ages sky-averaged 21-cm signal. The top panel shows the relationship and evolution of the radio background (orange dashed line), gas temperature (blue solid line) and spin temperature (green dashed line). The bottom panel shows the dark ages 21-cm signal, which is produced by collisional coupling of the spin temperature to the gas temperature. We assume a  $\Lambda$ CDM baseline [22].

The intensity of the Lyman- $\alpha$  emission is driven by the star formation rate in early galaxies. In semi-numerical simulations, the star formation rate is modelled with a star formation efficiency,  $f_*$ , and the virial circular velocity of star forming halos,  $V_c$  [13]. The efficiency is the fraction of gas in dark matter halos that is transformed to stars.  $V_c$  is then directly related to the halo mass,  $M$ , via

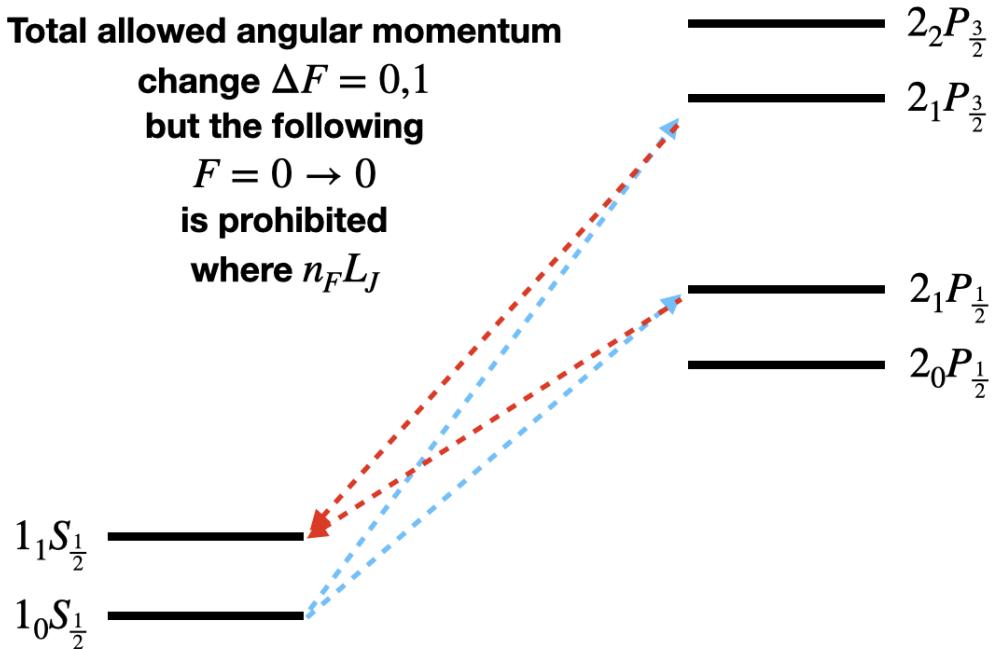
$$V_c = 23.4 \left( \frac{M}{10^8 h^{-1} M_\odot} \right)^{1/3} \left[ \frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2} \right]^{1/6} \left( \frac{1+z}{10} \right)^{1/2} \text{km s}^{-1}, \quad (1.15)$$

derived in [41]. Here  $M_\odot$  is a solar mass unit,  $\Omega_m$  is the matter density parameter,  $\Omega_m^z$  is the equivalent as a function of redshift and  $\Delta_c$  is the final overdensity relative to the critical density at the collapse redshift of the halo. The effects of varying the astrophysical parameters in the semi-numerical simulations, with a radio background above the CMB, on the global 21-cm signal are shown in Fig. 1.4.

The Wouthuysen-Field effect thus causes an absorption against the radio background between  $z \approx 20 - 30$  in the sky-averaged signal, which can be seen in Fig. 1.3, by coupling  $T_s$  and  $T_k$ . In the power spectrum, the Wouthuysen-Field effect manifests itself as a peak at high  $z$  and angular scales corresponding to the effective horizon of the coupling process (a few Mpc). In practice, the spin temperature is coupled to the colour temperature,  $T_c$ , but since the IGM is optically thick to Lyman- $\alpha$  emission during this period  $T_c \approx T_k$ . The coupling coefficient is given by

$$x_\alpha = \frac{4P_\alpha T_*}{27A_{10}T_r}, \quad (1.16)$$

where  $P_\alpha$  is the total rate at which Lyman- $\alpha$  photons are scattered per atom.  $x_\alpha$  is directly proportional to the angular averaged intensity of Lyman- $\alpha$  emission,  $J_\alpha$ , via  $P_\alpha$ . During this epoch  $T_k$ , and the coupled spin temperature, continues to cool at a faster rate than the radio background. However, various heating processes



**Figure 1.2:** The Wouthuysen-Field effect. The absorption and emission of Lyman- $\alpha$  photons by neutral hydrogen drives the spin temperature during the epoch of star formation via the Wouthuysen-Field effect. The blue dashed lines show the absorption of Lyman- $\alpha$  photons and the red show the emission.

begin when the first stars begin to form raising the gas temperature and preventing  $T_k$  from cooling adiabatically. These include Lyman- $\alpha$ , X-ray, CMB and shock heating.

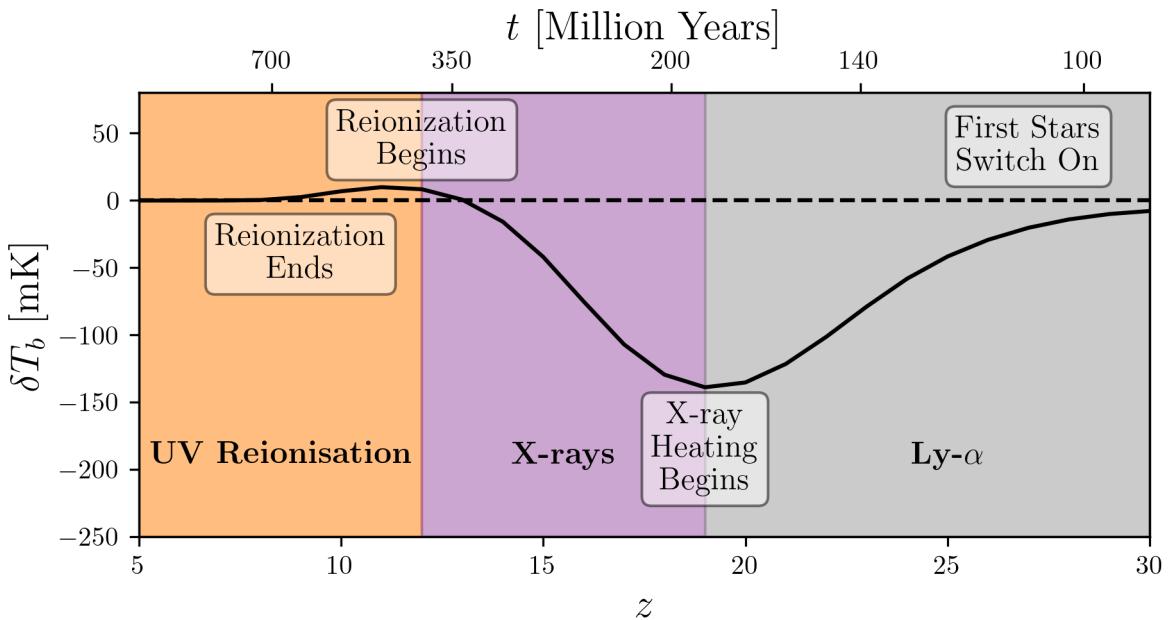
Lyman- $\alpha$  heating plays a significant role in determining the structure of the global 21-cm signal [14]. This mechanism deposits energy in the gas, increasing  $T_k$ , via the scattering of Lyman- $\alpha$  photons, that induce the Wouthuysen-Field effect, and although the energy deposited per event is small the number of scattering events is large. Since  $T_k$  and  $T_s$  are coupled at this time, both begin to move back towards the radio background temperature, reducing  $\delta T_b$ . The Wouthuysen-Field effect still dominates during this period but the heating mechanism significantly reduces the maximum allowed theoretical depth of the global signal from  $\approx -250$  mK to  $\approx -165$  mK [14] assuming the radio background is equal to the CMB temperature. This process begins as soon as the first stars form, whereas X-ray heating is delayed until quasars and X-ray binaries begin to form. The heating mechanisms in general suppress the low- $z$  power spectrum peak induced by the Wouthuysen-Field effect [14, 42–45].

A related heating mechanism is CMB heating, in which energy is transferred from the radio background to the gas via interactions with neutral hydrogen. In CMB heating, a CMB photon raises the energy state of the electron in the neutral hydrogen atom from  $1_0S_{1/2}$  to  $1_1S_{1/2}$  and the neutral hydrogen atom subsequently scatters with a Lyman- $\alpha$  photon. The scattering raises the electron into one of the 2P states and then the electron can deexcite back into the  $1_0S_{1/2}$  releasing a higher energy Lyman- $\alpha$  photon in the process. The additional energy has come from the CMB photon and is transferred by the Lyman- $\alpha$  photon to the gas thus raising the gas temperature [44]. CMB heating can reduce,  $x_r$  which is dependent on the 21-cm optical depth

$$x_r = \frac{1 - \exp(-\tau_\nu)}{\tau_\nu}, \quad (1.17)$$

below 1 [14].

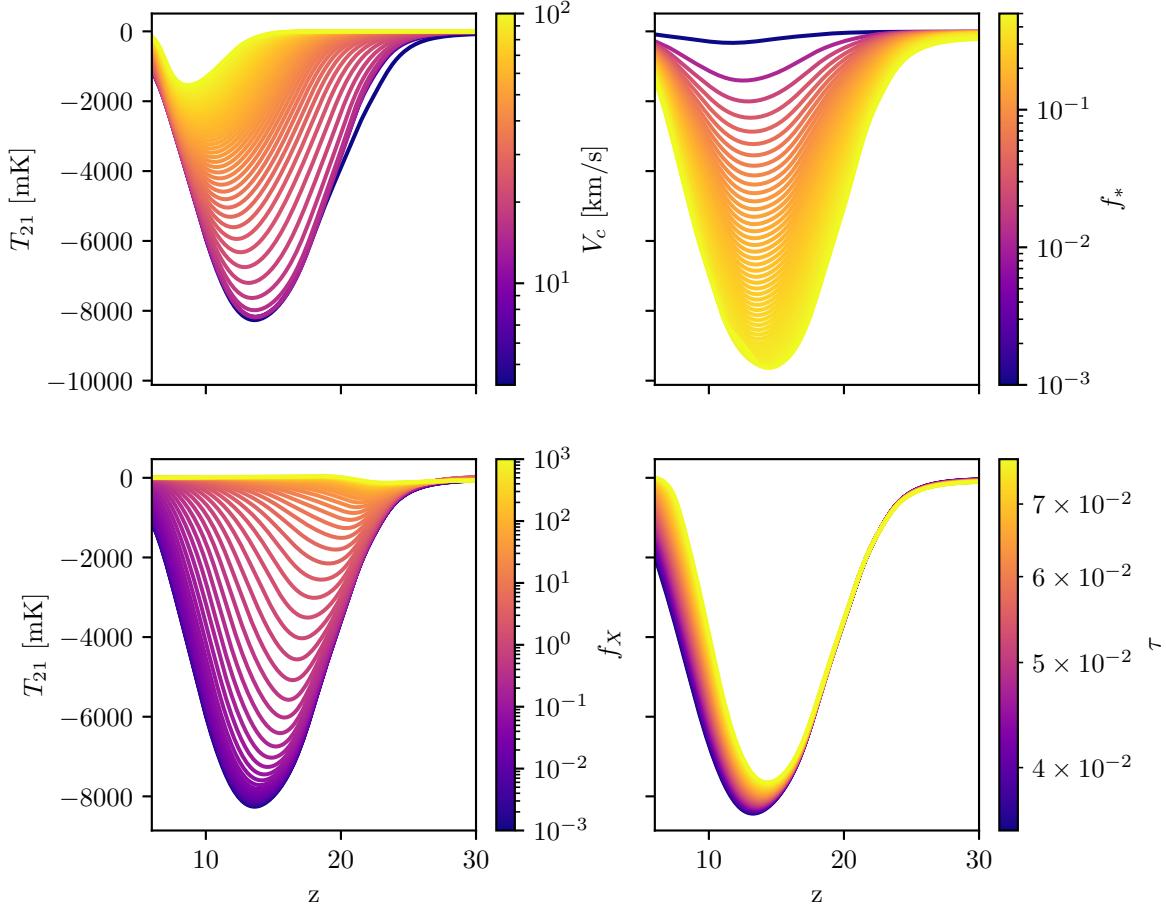
Further heating occurs through IGM shocks. As the Universe expands the density fluctuations seen in the CMB grow to form galaxies and large scale structure. However, these density fluctuations are not spherical



**Figure 1.3:** The 21-cm signal from the Cosmic Dawn and Epoch of Reionization is highly dependent on the intensity of a number of different radiative fields. The figure illustrates the dominant radiative fields as a function of redshift and time after the Big Bang and as a backdrop to an example 21-cm signal. At early redshift, when the first stars switch on, Lyman- $\alpha$  emission drives the spin temperature to the gas temperature, causing an absorption in the signal against the radio background. CMB and Lyman-alpha heating begin to play an important role in modulating the gas temperature and then at later redshifts along with X-ray heating raise the gas temperature back up to and sometimes beyond the radio background temperature. Finally, at more recent times and low redshifts UV emission from stars ionizes the neutral hydrogen and the signal disappears against the radio background. The signal shown here was generated with the neural network signal emulator GLOBALEMU, which is discussed in chapter 3, trained on simulations with a radio background equivalent to the CMB.

and so collapse into sheets or “Zel’dovich pancakes”, filaments and halos. During this process of collapse, if the velocity of the infalling gas is higher than the speed of sound, some of the gravitational infall energy is transformed into thermal energy via shocks. Shocks are only important when the temperature of the shock,  $T_{sh}$ , is greater than  $T_k$  and the temperature scales as the mass of the collapsing region to the power of six. The process plays a more significant role at high redshift but the energy transferred to the IGM through shocks reduces with increasing time and is generally thought to be minimal throughout the CD and EoR [28, 34].

Finally, X-ray heating plays a significant role in the structure of the global 21-cm signal. It produces a peak in the power spectrum as a function of redshift and at the relevant galactic and intergalactic scales. X-ray heating is expected to begin around redshifts 15-20. Low energy X-rays and some high energy X-rays with long mean free paths from Galaxies and quasars are absorbed by the IGM and deposit most of their energy as kinetic energy raising the gas temperature,  $T_k$ , which at this time is still coupled to the spin temperature via the Wouthuysen-Field effect. Further, the X-rays photoionize hydrogen and helium in the IGM with the resulting free electrons retaining the majority of the photon energy as kinetic energy. This kinetic energy is subsequently deposited into the IGM via collisional ionization and excitation of helium and hydrogen. Collisional ionization leads to a background of secondary electrons that subsequently deposit kinetic energy into the IGM and excitation of helium creates a background of photons that go on to ionize hydrogen. Collisional excitation of hydrogen in turn produces a background of Lyman- $\alpha$  photons which can heat the gas through the process described above. Finally, the free electrons created by X-ray photoreionization of hydrogen can heat the IGM through Compton collisions with other free electrons.



**Figure 1.4:** The figure shows the effect of varying four different parameters of the semi-numerical simulation detailed in the text and throughout this thesis. The base signal that is varied has strong star formation, inefficient X-ray heating and a strong radio background. The simulations have an excess radio background above the CMB from high redshift radio galaxies. The minimum virial circular velocity,  $V_c$ , and the star formation efficiency,  $f_*$ , determine the star formation rate and consequently the Lyman- $\alpha$  flux. Low values of  $V_c$  and high values of  $f_*$  lead to high Lyman- $\alpha$  fluxes, strong coupling between the spin temperature and the gas temperature, and as a result deep 21-cm signals. It also influences the onset of heating, however this is harder to disentangle from X-ray heating. X-ray heating is governed by the X-ray efficiency,  $f_X$ , in these simulations. Low values of  $f_X$  lead to inefficient X-ray heating and push the minimum of the global 21-cm signal down and to lower redshifts. Finally, we show the variation in the signal with the CMB optical depth,  $\tau$ , which has very little influence in the structure of the signal at high redshift but becomes more important at low redshift during the Epoch of Reionization. It's important to note that these effects are not independent and do not occur at distinct epochs but rather overlap, meaning that the structure of the signal as a function of redshift is complex. However, figures like this help to gain an intuition for how each parameter affects the structure of the signal. Again, the signals were produced with the neural network emulators discussed primarily in chapter 3 and then also in chapter 5 and chapter 6.

The high redshift X-ray background is poorly understood and in the semi-numerical simulations employed throughout this thesis it is assumed that the relationship between the X-ray luminosity and the star formation rate at low redshifts can be extrapolated, taking into account an increase in luminosity with decreasing metallicity, to high redshifts. Between 0.2 - 10 keV the luminosity is given by

$$L_X = 3 \times 10^{40} f_X \left( \frac{\text{SFR}}{1 M_\odot \text{yr}^{-1}} \right) \text{erg s}^{-1}, \quad (1.18)$$

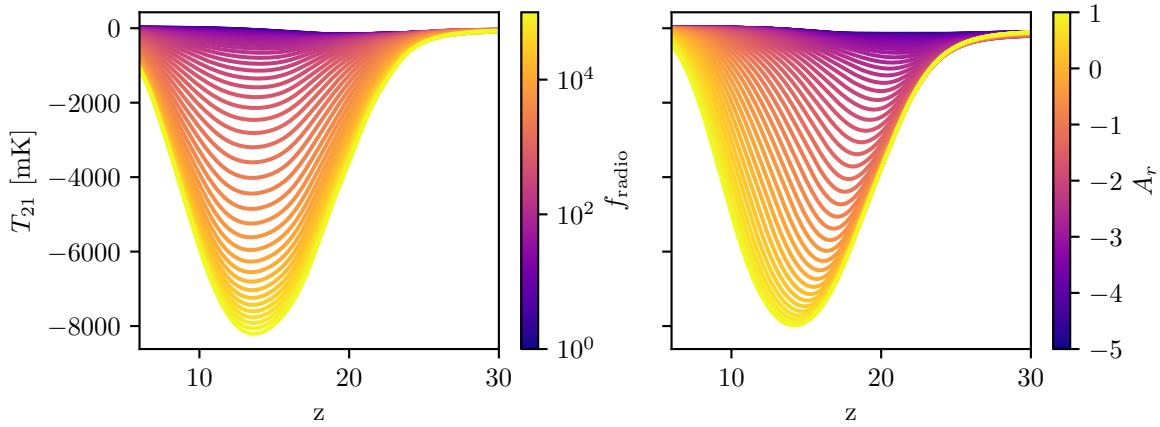
where  $f_X$  is the efficiency of X-ray production in high redshift galaxies, effectively acting as a normalisation parameter, and a free parameter in our models. We model the spectrum of X-ray sources in the simulations under the assumption that they are predominantly X-ray binaries in which stellar mass from a main sequence star accretes onto a neighbouring compact object. These systems should be prominent in the early Universe and should form as the first population of stars begins to die. We characterise the spectra of the X-ray binaries either empirically from [46] or analytically with two free parameters in our semi-numerical simulations;  $\nu_{\min}$  (occasionally referred to  $E_{\min}$ ) which is the low energy cut-off of the spectral energy density and  $\alpha$  which is the slope of the spectral energy density. In some instances, X-ray heating can be sufficiently large to raise the gas temperature and coupled spin temperature above the radio background temperature, as can be seen in Fig. 1.3.

At lower redshifts, reionization, in which neutral hydrogen is ionized via UV emission from the first stars, begins to dominate the structure of the 21-cm signal during the Epoch of Reionization. Once all the hydrogen is ionized and the neutral fraction of hydrogen in equation (1.3),  $x_{HI} = 0$ , then the signal disappears against the radio background. Reionization occurs first around the sources of UV emission and is dependent on the mean free path,  $R_{\text{mfp}}$ , of the ionizing radiation which is a free parameter in the semi-numerical simulations. The total column density of the ionizing gas is measured by the CMB optical depth  $\tau$  which is proportional to the ionizing efficiency of sources,  $\zeta$ . Again,  $\tau$  is a free parameter in the simulations. It has previously been constrained by Planck, however, its value remains uncertain with a strong dependence on the reionization model. The 21-cm line offers an alternative means by which it can be constrained. UV reionization begins as soon as the first stars form and UV photons in the Lyman-Werner band deplete the amount of neutral hydrogen that can go into stars via ionization thus suppressing star formation in a process known as Lyman-Werner feedback [13]. However, the strength of this feedback mechanism is not well understood [47].

Typically, the radio background in simulations of the 21-cm signal is set to the CMB temperature. However, we know that radio galaxies should contribute to this background. Further, measurements from the balloon experiment Absolute Radiometer for Cosmology, Astrophysics, and Diffuse Emission (ARCADE2) [48] and the Long Wavelength Array (LWA) [49] suggests that there is a radio background in excess of the CMB which can be modelled as a homogeneous synchrotron source. However, there are concerns about the galactic modelling in the two experimental measurements of the excess [50]. The reported detection from the EDGES experiment in 2018 also hints at the possibility of an excess radio background, however there are concerns about the modelling of the EDGES data (see section 1.6). A larger radio background leads to a stronger contrast with the spin temperature and consequently a deeper absorption feature in the global 21-cm signal and a larger power spectrum.

Throughout the later chapters of this thesis, we use semi-numerical simulations with two different parametrisations of an excess (over the CMB) radio background. The first is from a synchrotron source which produces a uniform radio background

$$T_r = T_{\text{CMB}} \left[ 1 + A_r^{78} \left( \frac{\nu}{78 \text{MHz}} \right)^\beta \right], \quad (1.19)$$



**Figure 1.5:** The figure shows the variation of two different parameterisations of exotic global 21-cm signals with excess radio backgrounds. In both figures the star formation rate is high (with a high star formation efficiency and low minimum virial circular velocity) and the X-ray efficiency is low, which along with a large radio background produces a deep signal. The CMB optical depth is set to 0.05 and its value is only important at low redshifts. The left-hand panel corresponds to an excess radio background above the CMB produced from radio galaxies and mediated by the radio production efficiency,  $f_{\text{radio}}$ , which is proportional to the radio luminosity per star formation rate,  $L_r/\text{SFR}$ . The right-hand panel corresponds to an excess from a homogeneous synchrotron source and its magnitude is mediated by  $A_r$ , which is effectively the fractional increase on the CMB at a given frequency (here 78 MHz). The excess radio background models were originally proposed to explain the anomalously deep EDGES absorption feature reported in 2018. More details can be found in the text. The signals were generated using the neural network emulator detailed in chapter 3, chapter 5 and chapter 6.

that decays with increasing time. Here  $A_r$  is a fractional increase on the CMB at a given frequency, and it can be scaled according to

$$A_r^{78} = A_r^{1420} \left( \frac{0.078}{1.420} \right)^\beta, \quad (1.20)$$

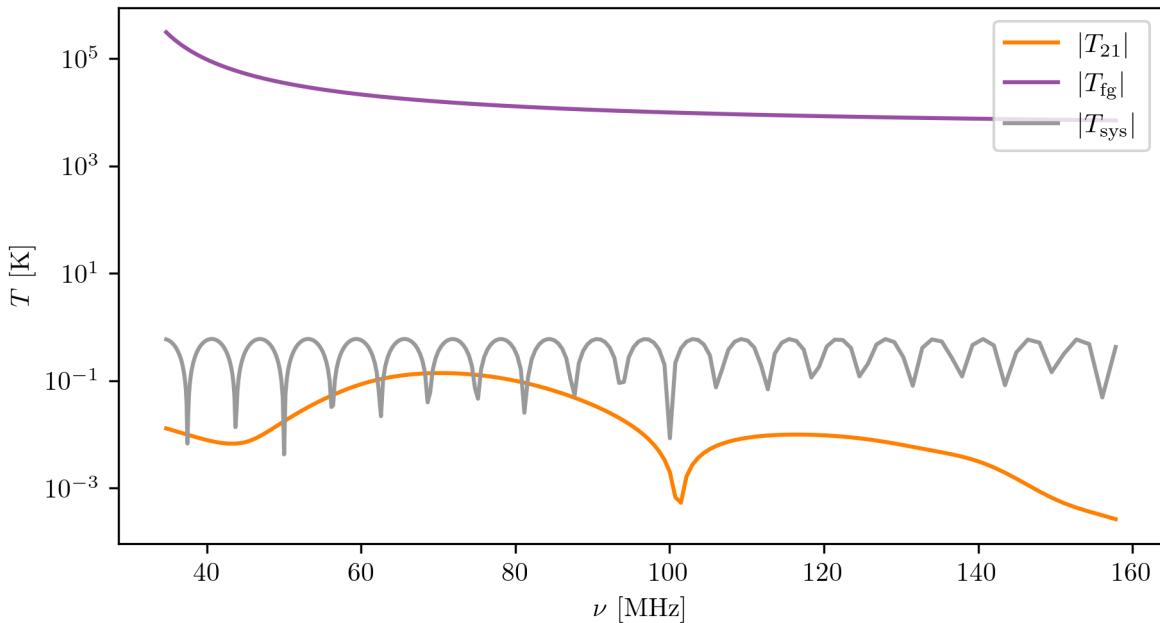
where  $\beta$  is often chosen to be equal to  $-2.6$  consistent with measurements from ARCADE2 and LWA. This model has a larger impact on the global 21-cm signal at higher redshifts and generally pushes the minimum of the global 21-cm signal to higher redshifts which can be seen in Fig. 1.5. The synchrotron excess may be sourced by superconducting cosmic strings [51] or light dark matter decays [52, 53]. Since the radio background is homogeneous, it leads to an increase in the power spectrum at all angular scales, but its impact is again more significant at high redshifts.

The second parameterisation of the radio background is from high redshift radio galaxies. In this instance the radio background is modelled via the radio production efficiency,  $f_{\text{radio}}$ , which is proportional to the radio luminosity per star formation rate

$$L_r = 10^{22} f_{\text{radio}} \left( \frac{\nu}{150 \text{ MHz}} \right)^\beta \left( \frac{\text{SFR}}{M_\odot \text{ yr}^{-1}} \right) \text{ W Hz}^{-1}, \quad (1.21)$$

where again  $\beta$  is usually set to  $-2.6$ . The level of excess background above the CMB in this model increases with decreasing redshift, in contrast to the previously discussed model, as more radio galaxies form. It is also non-uniform unlike the synchrotron model. This non-uniformity leads to an increase in the power spectrum on angular scales corresponding to galactic and intergalactic length scales. The model has a similar effect on the depth of the global 21-cm signal to the synchrotron model creating a stronger contrast between  $T_s$  and  $T_r$ . However, as can be seen in Fig. 1.5, it does not affect the redshift of the minima in the global signal.

Since the 21-cm signal is highly dependent on the amplitude of a number of different radiation fields, it is a powerful probe of the infant Universe and the first galaxies. Its detection represents one of the final frontiers



**Figure 1.6:** The graph shows the relative magnitude of different signals collected by global 21-cm experiments. The foregrounds (purple) are around five orders of magnitude larger than the 21-cm signal (orange). Sinusoidal systematics are frequently found in the data from these experiments and can be caused by cable reflections, chromaticity in the beam and discontinuities in the soil among other things discussed in the text. We show an example systematic (grey) to illustrate that these signals have similar magnitudes to the 21-cm signal and need to be corrected for.

in observational cosmology, along with studies of dark energy, dark matter and gravitational waves. In the later chapters of this thesis, we present the current best limits on the properties of early galaxies from 21-cm cosmology, but presently we describe the challenges of making a detection.

## 1.3 Challenges facing the field

There are a number of different challenges facing the field of 21-cm cosmology, and we detail several specific issues in the following subsections. This is followed by a more general discussion of ‘unknown systematics’ in which we group together additional challenges that if incorrectly dealt with can introduce non-smooth structure into the data that mimics, corrupts or hides the global 21-cm signal. We focus on the detection of the sky-averaged 21-cm signal here, as that is the main focus of the thesis, but note that there are similar challenges and unique problems in studies of the power spectrum.

### 1.3.1 Galactic, extra-galactic foregrounds and the antenna beam

Radiometers aiming to detect the sky-averaged 21-cm signal also pick up emission from the Galaxy and from other galaxies, which dominates over the order 100 mK cosmological signal. This foreground is composed of synchrotron and free-free emission and is approximately five orders of magnitude larger than the 21-cm signal, as can be seen in Fig. 1.6. The foregrounds are expected, from the theory of synchrotron and free-free emission, to be smooth and follow an approximate  $\nu^{-2.5}$  power law. SARAS3 found that the sky had a spectral index of  $-2.42$  [54], EDGES found it to be approximately  $-2.6$  as a function of local sidereal time [55], and LEDA reported a measurement of the spectral index to be  $\beta = -2.5 \pm 0.1$  [56].

Since the foreground dominates so significantly over the 21-cm signal, it is essential that the two signals can be separated from each other. To do this, we often take advantage of the fact that the foreground is smooth, and the signal is not over a wide frequency (or equivalently redshift) range. The idea is that we model the smooth structure in the data with a low order polynomial function, often in log-space,

$$\log_{10} T_{\text{fg}}(\nu) = \sum_i^N a_i \log_{10}(\nu)^i, \quad (1.22)$$

where the coefficients  $a_i$  are fitted for, to leave behind the non-smooth 21-cm signal. Alternative physically motivated polynomial models have also been proposed, however their applications have been limited [8].

There are however complications with the principle of separating smooth foregrounds from the non-smooth signal with polynomials. It requires the design of a wideband instrument, since over narrow frequency ranges the 21-cm signal is also smooth and easily modelled by low order polynomials. Whilst designing a wideband antenna is possible, this leads to a very non-smooth response from the antenna to the sky, which can be difficult to correct for. This effect in the antenna beam is known as chromaticity and is a major issue for 21-cm experiments.

In Fig. 1.7, we illustrate how the structure of the antenna beam affects the recorded data. We use an elliptical dipole above an infinite ground plane shown in the bottom right of the figure to illustrate, along the top row, how a slice through the beam at a fixed azimuthal angle changes with frequency from 50 – 200 MHz. We then posit the existence of two bright sources in the sky, shown by the purple and orange stars, which have smooth spectra that are power laws in frequency with  $\beta \approx -2.5$  as shown by the orange and purple dashed lines in the figure on the bottom left. The incoming radiation from these two bright sources is effectively along the line of sight because they are so distant that radial emission is not seen by the antenna.

The beam weights the input power from the sky,  $T_{\text{sky}}$ , based on its directivity which as can be seen in the top row is a function of angle and frequency. The solid orange and purple lines in the bottom left panel show the weighted spectra which feature chromatic distortion, where the weighting has been exaggerated for demonstration, from the two posited point sources. This can also be seen in the sum of the spectra, shown in black.

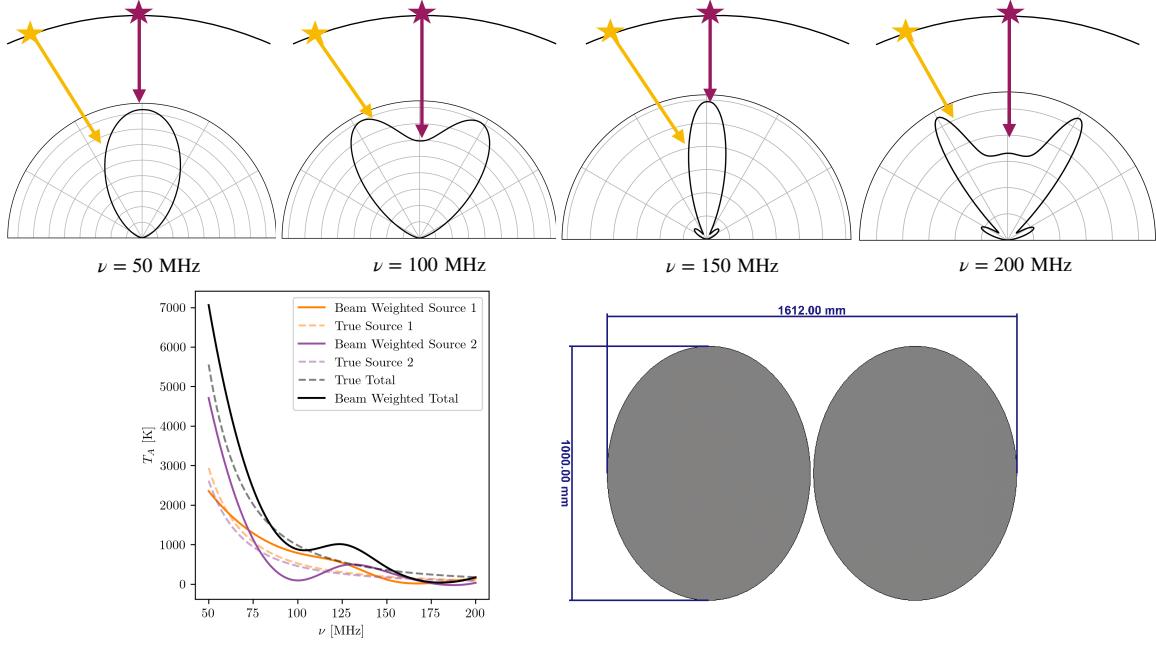
In practice, the problem is much more complicated because the beam is 3D, the sky is rich in complex structures such as the galactic plane, the north galactic spur and bright sources such as Cygnus A, and observations are taken over long periods in which the sky rotates above the beam. However, from the simple example we can clearly see how the beam can introduce non-smooth structure into the antenna temperature,  $T_A$ , and how modelling the foreground in wideband data with a smooth low-order polynomial will leave behind both the non-smooth signal but also a large non-smooth chromatic residual.

One approach to deal with chromaticity is to fit high order polynomials in an effort to model the foreground and the non-smooth chromatic structure in the data. However, this can of course fit out the 21-cm signal as well.

More complete approaches to modelling the effects of chromaticity can be taken. For example, the EDGES collaboration [55] proposed a correction given by

$$B(\nu, t) = \frac{\int_{\Omega} T_{\text{sky-model}}(\nu_{\text{ref}}, t, \theta, \phi) D(\nu, \theta, \phi) d\Omega}{\int_{\Omega} T_{\text{sky-model}}(\nu_{\text{ref}}, t, \theta, \phi) D(\nu_{\text{ref}}, \theta, \phi) d\Omega}, \quad (1.23)$$

where  $D$  is the directivity of the antenna,  $t$  is the time, and  $\theta$  and  $\phi$  are the zenith and azimuthal angles parameterising the beam. The sky temperature is modelled by the difference between a map and the CMB



**Figure 1.7:** The figure illustrates the impact of a chromatic beam on the sky-averaged 21-cm signal. The bottom right panel shows a Computer Simulation Technology (CST) model of an elliptical dipole antenna, provided by Dominic Anstey and John Cumner. Along the top row, we show slices through the beam at four fixed frequencies in the range of interest for a 21-cm experiment. The structure of the beam changes with frequency and this causes signals from the sky to be down-weighted at some frequencies relative to others. To illustrate this, we show two different point sources, which we assume to have approximate power law spectra, and how they interact with the beam. The bottom left-hand panel shows the “true spectrum” for each source as dashed lines and the recorded spectra as solid lines. We also show the sum of these two point source spectra in black. The beam introduces significant non-smooth structure into the data and this needs to be appropriately modelled so that the 21-cm signal can be recovered from the data. In practice, the problem is much more complicated because the beam is three-dimensional, the sky is full of complex structure and the sky varies with time.

scaled by a signal power law

$$T_{\text{sky-model}}(\nu_{\text{ref}}, t, \theta, \phi) = [T_{\text{base-map}} - T_{\text{CMB}}] \left( \frac{\nu_{\text{ref}}}{\nu_{\text{base-map}}} \right)^{-2.5}. \quad (1.24)$$

The corrected antenna temperature is then given by

$$T_{\text{cor}} = \int_{t_{\text{start}}}^{t_{\text{end}}} \frac{T_{\text{data}} - T_{\text{CMB}}}{B} dt + T_{\text{CMB}}. \quad (1.25)$$

However, this correction assumes an exact knowledge of the foreground and the antenna beam pattern. In practice the foreground maps are not known well enough and while we can model the antenna beam it is difficult to account fully for the environment around the antenna and for imperfections in manufacturing.

A sophisticated forward modelling of the antenna beam and the foreground has been proposed in a series of papers [57–59]. The model uses a base sky map which is divided into  $N$  regions of varying spectral index and convolved with a model of the beam

$$T_{\text{fg}} = \frac{1}{4\pi} \int_0^{4\pi} D(\nu, \theta, \phi) \int_{t_{\text{start}}}^{t_{\text{end}}} \sum_{i=0}^N M_i(\theta, \phi) [T_{\text{base-map}} - T_{\text{CMB}}] \left( \frac{\nu}{\nu_{\text{base-map}}} \right)^{-\beta_i} dt d\Omega + T_{\text{CMB}}, \quad (1.26)$$

where  $M_i(\theta, \phi)$  is the mask that divides the sky map into  $N$  regions of identical spectral index  $\beta_i$ . Work is currently ongoing on a model for any deviation in the beam from the simulations caused by environmental and manufacturing errors.

Ground plane artefacts can be introduced in the data by finite edges or discontinuities in the structure, which are usually hard to model [60]. These can produce standing waves on the finite edges of the ground plane and introduce ripples that are unaccounted for in the modelling of  $D(\nu, \theta, \phi)$ . Ground planes are typically employed for dipole antennas to prevent the antenna looking at the ground with the same efficiency that it looks at the sky, however emission from the soil can still cause problems and discontinuities in the soil structure can also distort the directivity of the antenna.

The 21-cm signal is also affected by chromaticity, but the foregrounds change on much larger length scales and the intensity of the foregrounds vary much more across the sky. We can therefore generally ignore chromatic effects in the sky-averaged 21-cm signal.

Power spectrum experiments also struggle with the effects of the antenna beams and the foregrounds. For example, a major cause of systematics in these experiments is cross-talk between different antennas in the interferometric array. Cross-talk occurs because antennas re-radiate reflected signals, and this can be picked up by neighbouring antennas in the often closely packed array. A detailed discussion of the systematics in power spectrum experiments and how these experiments deal with foregrounds is beyond the scope of this thesis. Briefly, however, one of the most effective ways of dealing with foregrounds in power spectrum experiments takes advantage of the fact that the foreground and 21-cm signals appear on different angular scales and different redshifts meaning that the foregrounds can be avoided by looking at small angular scales at late times.

In global or sky-averaged experiments we can try to avoid chromatic effects, however, this is very difficult even with a narrowband instrument. Such antennas have frequency independent beams and are known as achromatic antennas. Currently, there is only one collaboration, SARAS, attempting to build such antennas. They are relying on monopole antennas and observing over narrow bandwidths.

It has been shown via simulations that the SARAS2 instrument has an approximately achromatic response to the sky over the redshift range  $z \approx 7 - 12$  [61]. However, as shown in chapter 5 the data features sinusoidal systematics which were likely introduced by discontinuities in the ground below the antenna setting up a standing wave in the beam pattern which was not modelled and introduced chromaticity. In contrast, the SARAS3 instrument was situated on a lake and the data was found to be smooth and absent of chromatic structure over the observing band  $z \approx 15 - 25$  (see chapter 6).

Polynomial foreground models remain a useful tool if we can build and collect data with a perfectly achromatic beam. However, they can still fit out the 21-cm signal if the order is too high. This can be overcome by forcing the polynomial to be smooth by constraining the coefficients of the model so that the higher order derivatives do not cross zero

$$\frac{\delta^m T_{fg}}{\delta \nu^m} \leq 0 \text{ or } \frac{\delta^m T_{fg}}{\delta \nu^m} \geq 0. \quad (1.27)$$

Such polynomials are called ‘Maximally Smooth Polynomials’ and were originally proposed in [61] and [62]. They are the subject of chapter 2, where we expand the definition to the more general ‘Derivative Constrained Functions’ (DCFs) and introduce an efficient fitting algorithm for this non-trivial problem.

In principle, DCFs should leave behind the non-smooth signal in the residuals and indeed will leave behind any non-smooth previously unexpected chromatic residuals in the data. The application of DCFs is of course still subject to the issue of narrowband observations of smooth sections of the 21-cm signal.

### 1.3.2 The atmosphere

The atmosphere also poses a problem for observations of the global 21-cm signal, and in particular the ionosphere, which is an ionized layer of the atmosphere between  $\approx 50 - 1000$  km above sea level. As radio waves propagate through this layer of the atmosphere they can be absorbed and refracted with the effect being more severe for lower frequencies (scaling as  $v^{-2}$ )[\[63–65\]](#). The ionosphere is subdivided into multiple layers in which propagating radio waves are modulated in different ways.

The refraction of the radio waves in the F-layer of the ionosphere distorts sources so that they appear to the antenna at higher elevations than they truly are. This means that they enter the beam at a different angle to expectations and are consequently weighted by the beam to a different degree than anticipated in our modelling.

Absorption in the D-layer of the ionosphere can reduce the perceived magnitude of the 21-cm signal which has implications for our understanding of the early Universe [\[65\]](#). For example, a reduction in the depth of the global 21-cm signal implies a lower rate of star formation, a weaker radio background and strong X-ray heating.

The degree to which these effects occur can be related to the total electron content (TEC) along the line of sight. However, this quantity is difficult to measure and varies significantly with time. In addition, it has been shown that the effects of the ionosphere do not average out over time [\[65\]](#). As a result, current observations often target periods of quiet ionospheric activity so that the effects of the atmosphere are stable in the data and can be modelled with additional nuisance parameters.

A more complete approach to model the effect of the ionosphere on observations of the 21-cm signal is to modify the antenna beam [\[66\]](#). However, current efforts generally assume quiet ionospheric conditions.

### 1.3.3 Radio Frequency Interference

21-cm experiments also have to contend with man-made radio frequency interference, or RFI, which can originate from many different sources. Radio stations, for example, that use frequency modulation (FM) to transmit do so in the frequency range 88 – 108 MHz corresponding to redshifts 12 – 15. Additionally, satellite and aeroplane communications can introduce RFI in the frequency range of the 21-cm signal. Ground based communication (for example from Walkie Talkies on site) can also cause problems, and non-local to the experiment RFI can be reflected off micro-meteorites and aircraft. In practice, the receiver and instrumental backend can also introduce RFI into observations if not properly shielded.

A simple technique to minimize RFI is simply to avoid it, which is achieved by building your telescope in a radio quiet location. For example, the Radio Experiment for the Analysis of Cosmic Hydrogen (REACH) antenna is being constructed in the Karoo Radio Observatory, which is a legally protected radio quiet zone home to the Karoo Array Telescope (MeerKAT) Square Kilometer Array (SKA) precursor and HERA. Legal protection typically only covers certain frequency bands, and in some instances ‘naturally occurring’ radio quiet zones offer a viable alternative. These are often areas that are generally inhospitable, such as Marion Island located between Antarctica and South Africa where the Probing Radio Intensity at high-Z from Marion (PRIZM) experiment [\[67\]](#) and the Array of Long Baseline Antennas for Taking Radio Observations from the Sub-Antarctic (ALBATROS) [\[68\]](#) are located or remote regions such as the Timbaktu Collective in Southern India, which was radio quiet during the COVID-19 pandemic, where SARAS2 was located [\[69, 70\]](#). In addition, a high horizon around the instrument from mountains for example can shield the antenna from RFI however the shape of the horizon introduces its own problems [\[71\]](#).

Despite efforts to avoid RFI, it inevitably enters to some degree. FM radio is generally mitigated by avoidance, but for any residual FM signals in the data we can flag and remove the corresponding frequency channels easily because we expect it between 88 – 108 MHz. Things become more complicated when we do not know the exact frequencies of particular RFI contaminants, although we can take advantage of the fact that some RFI signals are time dependent. For example, satellite communications will only be present in the data when the satellites are in the beam of the antenna. We can therefore identify RFI contaminated time steps in our data and remove them before integrating our observations. Wide-band RFI can be particularly challenging to identify and deal with in our data however this is expected to be infrequent in 21-cm experiments with the majority coming from reflections off micro-meteorites and aeroplanes.

Flagging in frequency, time or a combination is often performed by setting a thresholding value for power introduced into the data. The choice of this thresholding value is usually non-trivial however, since if it is too high it can leave behind RFI in the data and if it is too low it can remove important information about the signal.

One approach to mitigate RFI is to use machine learning. Convolutional Neural Networks can be trained to identify RFI in the time vs frequency data recorded by radiometers before integration is performed, and they can therefore be used to flag corrupted time steps and frequency channels.

A recent work has shown that RFI flagging can be performed alongside signal and foreground modelling in a Nested Sampling or Markov Chain Monte Carlo (MCMC) run by modifying the likelihood. The likelihood, which is discussed in section 1.4, represents the probability of the data given the choice of the model and is the sum over a vector corresponding to the sampling of the data in frequency. The proposed algorithm essentially down-weights certain data points in this sum if their likelihood is greater than a threshold value. In essence, this allows the data to tell you which points are RFI rather than having to manually flag them [72].

In some cases, in-painting is performed on data in frequency channels where RFI has been identified. This often involves replacing flagged RFI with the average of the power in neighbouring bins in the time vs frequency observations. However, this can distort the data, and it is often better to simply remove or mask contaminated bins, some approaches used to model the data can deal with such gaps.

Significant investment is being made in performing radio astronomy and indeed 21-cm cosmology from the Moon and in lunar orbit, primarily to overcome the effects of the Earth’s atmosphere but also RFI. These observations are typically targetting the 21-cm signal from the Dark Ages at high redshifts  $> 50$  (low radio frequencies  $< 50$  MHz) that are not easily accessible to ground based instruments due to the ionosphere. However, with increased interest in lunar exploration from both private companies and the space agencies of several countries across the world, it is unclear how long the Moon will remain an RFI quiet location. Preliminary studies of the environment are planned in 2023 and 2024 with experiments such as Radiowave Observations at the Lunar Surface of the photoElectron Sheath (ROLES) and Lunar Surface Electromagnetics Experiment-lite (LuSEE-LITE) [73].

### 1.3.4 Unknown systematics

We can think of our data being made up of many different components and many different effects of those components, such as the beam weighted foreground, the ionosphere and the 21-cm signal. In practice, anything that is not the 21-cm signal is a systematic, i.e. something we do not want in our data. However, some systematics, such as the synchrotron and free-free emission from our own Galaxy and others are reasonably well understood and guaranteed to be in our data, and so we classify them as separate problems to be solved.

Some things in our data however are currently either less well understood, incorrectly handled or do not have clear causes, and we refer to these systematics as “unknown” or “unaccounted” for systematics. The ultimate goal is to understand the causes of these unknown systematics as they can have a significant impact on our ability to detect the global 21-cm signal. They are hard to mitigate, however, and they are often found to either be larger or of a similar magnitude to the global 21-cm signal, as can be seen in Fig. 1.6.

Fig. 1.6 shows a sinusoidal systematic

$$T_{\text{sys}} = A \sin\left(\frac{2\pi\nu}{P} - \phi\right), \quad (1.28)$$

where the amplitude  $A = 600$  mK,  $P = 12.5$  MHz is the period and  $\phi = 0$  rad is the phase. The model is motivated by examples seen in the literature and discussed later in this thesis (e.g. chapter 2 and chapter 5).

Sinusoidal and frequency damped sinusoidal systematics

$$T_{\text{damp. sys}} = T_{\text{sys}} \left(\frac{\nu}{\nu_0}\right)^\alpha, \quad (1.29)$$

where the frequency is normalised, can be introduced in the data as a result of cable reflections [74] or potentially from ground emission (e.g. chapter 5 and [75]). Separating out the different potential causes of such systematics is non-trivial and complicated by differing experimental set-ups, environmental conditions and the time varying nature of some components of the data. Research is ongoing into the effects of the horizon [71] and polarised foregrounds [76, 77] on the data.

## 1.4 Bayesian analysis and 21-cm Cosmology

Bayes theorem was originally proposed by Thomas Bayes in [78] published posthumously by Richard Price and takes the form

$$P(\theta|D, M) = \frac{P(D|\theta, M)P(\theta|M)}{P(D|M)}, \quad (1.30)$$

where  $\theta$  are the parameters of our model  $M$  describing the data  $D$ .

In this thesis, we are concerned with the application of Bayes Theorem to parameter inference. In this context, the prior probability  $P(\theta|M) = \pi(\theta)$  encodes existing knowledge of the parameters and is usually assumed to be uniform or log-uniform across some large range of values or Gaussian around an existing experimental result. For example, there is significant uncertainty in the efficiency of X-ray emission from the first galaxies, and we typically use a wide prior range covering 6 orders of magnitude, from 0.001 to 1000, for the X-ray efficiency of early sources [15]. However, the CMB optical depth has been measured by Planck to be  $0.054 \pm 0.007$  and occasionally we use a Gaussian prior around this value in our analysis [22].

The likelihood  $P(D|\theta, M) = \mathcal{L}(\theta)$  is the probability of the data given the choice of model and is typically assumed to be Gaussian in nature

$$\log \mathcal{L} = \sum_i \left( -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \left( \frac{T_D(\nu_i) - T_M(\nu_i)}{\sigma} \right)^2 \right), \quad (1.31)$$

where  $T_D$  is the data and  $T_M$  is the model. The standard deviation of the Gaussian noise,  $\sigma$ , is usually assumed to be uniform. However, the magnitude of the noise is expected to follow the spectral distribution of the sky being larger at low frequencies and lower at high frequencies and can be modelled with the radiometer equation [79]. In principle, a uniform approximation to the noise is sufficient because of inaccuracies in calibration among

other issues. The likelihood can take on a variety of different forms [79] and Simulation Based Inference (SBI) or Likelihood Free Inference (LFI) are undergoing increasing interest in the field [80]. The idea behind SBI is to let the data determine the distribution of the noise using numerous simulations and, for example, Neural Ratio Estimators which model the ratio of the likelihood and the evidence (e.g. [81]). The evidence  $P(D|M) = \mathcal{Z}$  is the normalisation constant in Bayes theorem and can be used for model comparison. Finally, the posterior  $\mathcal{P}(\theta|D, M)$  tells us about the probability of a given parameterisation of our model given the data.

For Bayesian inference, we often write Bayes theorem as

$$\mathcal{P}(\theta|D, M)\mathcal{Z} = \mathcal{L}(\theta)\pi(\theta), \quad (1.32)$$

where the inputs to the inference process are defined on the right and the outputs on the left. In many implementations of Bayesian inference the evidence is not calculated and the resultant posterior not normalised

$$\mathcal{P}(\theta|D, M) \approx \mathcal{L}(\theta)\pi(\theta). \quad (1.33)$$

However, the evidence as mentioned is useful for model comparison, in that if  $\mathcal{Z}_{M_1} > \mathcal{Z}_{M_2}$  then model one is preferred by the data. This can help determine, for example, the presence of a global 21-cm signal in a data set as is done in chapter 2. In some instances where a series of models are fitted to the data a weighted sum of posterior probabilities can be taken to determine parameter constraints where the weighting is based on the evidence. This is illustrated in chapter 5.

The evidence is often referred to as the *marginal likelihood* since it is calculated via

$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)d\theta, \quad (1.34)$$

where the integral is over all of the parameters in the model. Equivalently, marginal one and two-dimensional posterior distributions can be defined by integrating the posterior over the other  $N - 1$  and  $N - 2$  parameters. This is particularly useful for visualising high dimensional posterior distributions and the technique is used throughout this thesis. It does however have the potential to hide constraints in higher dimensions, and we often need to look at functional constraints as well.

Bayesian inference is often performed with Markov Chain Monte Carlo (MCMC) methods like the Metropolis-Hastings algorithm [82] and affine invariant ensemble samplers [83]. These approaches often produce unnormalised samples of the posterior and do not directly provide an estimate of the evidence.

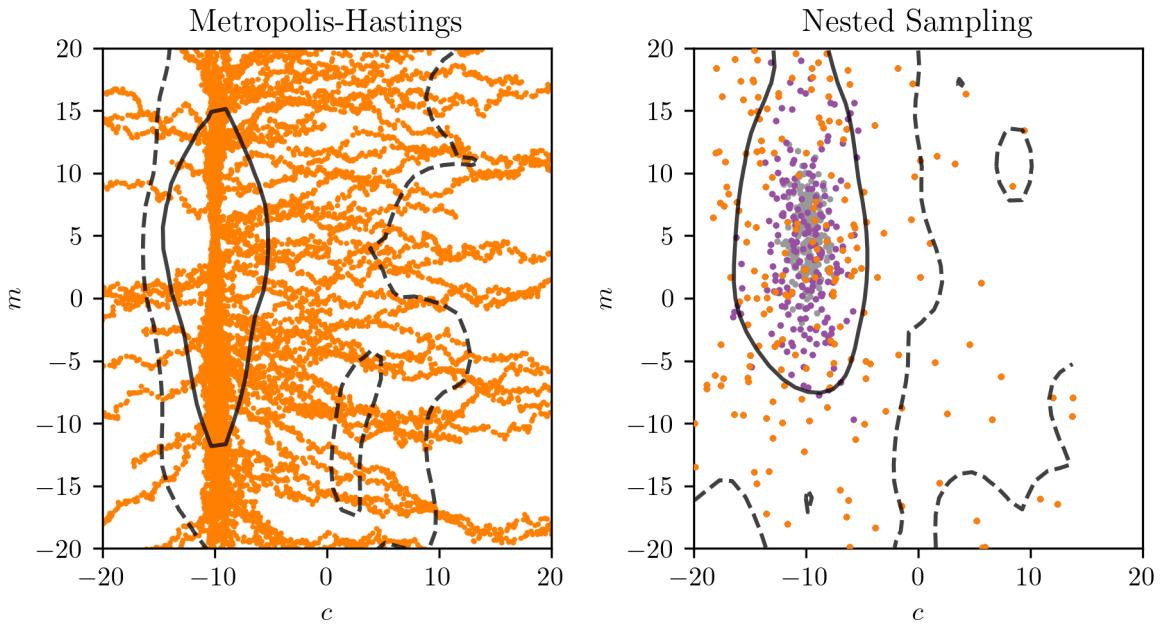
A simple example of an MCMC sampler is the Metropolis-Hastings algorithm. In the left panel of Fig. 1.8 we show the resultant posterior found when fitting the gradient,  $m$ , and intercept,  $c$ , of a straight line given by

$$y = mx + c + \epsilon, \quad (1.35)$$

where  $m = 5$  and  $c = -10$  and the noise  $\epsilon \sim \mathcal{N}(0, \sigma)$  is Gaussian distributed with a standard deviation of  $\sigma = 0.5$ .

The Metropolis-Hastings algorithm is a random walker algorithm that starts in some randomly chosen or carefully curated place,  $x_0 = \{m_0, c_0\}$ , often within some bounds, and subsequently gets perturbed in random directions. The direction of the perturbation is chosen based on some proposal distribution, which is equivalent to the prior. Here the proposal distribution of the next step,  $x_{i+1}$ , is given by a Gaussian centred around the previous step,  $x_i$

$$\pi(x_{i+1}|x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x_{i+1} - x_i)^2}{\sigma^2}\right), \quad (1.36)$$



**Figure 1.8:** The figure shows a comparison of the random walk Metropolis-Hastings (MH) sampling algorithm on the left and a rejection based Nested Sampling (NS) algorithm on the right. The MH algorithm evolves 100 walkers for 250 steps each to explore the parameter space and the NS algorithm evolves 100 live points for 600 iterations. For more details see the text.

where we have set the standard deviation,  $\sigma = 0.2$ . At each iteration, the likelihood for the new point,  $x_{i+1}$ , is calculated, for example with equation (1.31), and compared to the equivalent value for  $x_i$ . The proposal or prior weighted ratio of the two likelihoods is taken

$$\alpha = \frac{\mathcal{L}(x_{i+1})\pi(x_{i+1})}{\mathcal{L}(x_i)\pi(x_i)}, \quad (1.37)$$

and compared to a random number,  $u$ , from a uniform distribution between 0 – 1. If  $u \leq \alpha$  then the new point is accepted into the chain else  $x_i$  is perturbed again in a different random direction and the process is repeated. It is often more convenient to work in log-space because the likelihoods are very small, meaning that our inequality is given by

$$\log_{10}(u) \leq \log_{10}(\alpha) = \log_{10}(\mathcal{L}(x_{i+1})) - \log_{10}(\mathcal{L}(x_i)) + \log_{10}(\pi(x_{i+1})) - \log_{10}(\pi(x_i)). \quad (1.38)$$

The idea here is to climb up the likelihood contour but allow for occasional steps down the contours in an effort to prevent missing maxima in multi-modal problems. For example, in the case where the likelihood of step  $x_{i+1}$  is higher than for step  $x_i$  then  $\alpha$  is always greater than 1 and always greater than  $u$ . However, if the likelihood for step  $x_{i+1}$  is less than for  $x_i$  then  $\alpha < 1$  but if  $u$  is randomly chosen to be less than  $\alpha$  the step is accepted. This leads to a more complete exploration of the parameter space whilst maintaining a degree of efficiency and a focus on the maximum likelihood value. Through this process, the algorithm generates samples on the posterior distribution.

Here, the choice of  $\sigma$  in the proposal distribution has a significant impact on how the algorithm converges. In fact, the functional form of the proposal distribution strongly impacts how the space is explored and the efficiency with which the maximum likelihood value is found. The samples from HERA used in chapter 7 were derived using ensemble MCMC techniques, where the new sample points are chosen based on the distribution of all the current live points or walkers [83, 84].

In contrast, the Nested Sampling algorithm, used throughout this thesis, evolves a series of live points,  $n_{\text{live}}$ , up the likelihood contour [84, 85] in an effort to approximate the Bayesian evidence. The algorithm begins by evaluating the likelihood for the initial  $n_{\text{live}}$  points chosen from the prior distribution  $\pi(\theta)$ . The lowest likelihood point is then discarded and a new point is drawn from the prior with a higher likelihood than the discarded point. At each iteration the volume contained by the live points is compressed exponentially from an initial value of  $X_0 = 1$ , i.e. the prior, to  $X_{i+1} = \exp(-1/n_{\text{live}})X_i$ .

The evidence is therefore given by

$$\mathcal{Z} = \sum_{i=1}^{n_{\text{iter}}} w_i L_i^* \quad (1.39)$$

where  $L_i^*$  is the discarded likelihood at each iteration between  $i = 1$  and the total number of iterations  $n_{\text{iter}}$ . The weights  $w_i$  are given by

$$w_i = \frac{1}{2}(X_{i+1} - X_{i-1}). \quad (1.40)$$

The rejected sample at each iteration  $i$  has an associated likelihood, weight, evidence and prior probability, which means that the Nested Sampling algorithm generates normalised samples on the posterior. We can then interpret this to tell us which parameter values in our prior ranges best describe the data.

Nested sampling is compared with the Metropolis Hastings algorithm in Fig. 1.8 using the same two-dimensional line fitting problem discussed previously.

The simplest sampling algorithm where a new sample is randomly drawn from the prior to replace the discarded point is slow, and alternatives such as slice sampling are much more efficient [86]. Slice sampling was originally implemented in a Nested Sampling algorithm in [87] and extended in [88, 89] to work in higher dimensions.

Nested sampling algorithms typically rely on stopping criteria that determine the number of iterations to perform. It can become increasingly less rewarding to try to find replacement live points with higher likelihoods as you travel up the likelihood contours. A commonly used stopping criteria is when the fractional increase in the evidence is below some user specified tolerance

$$\Delta\mathcal{Z}/\mathcal{Z} < \text{tol}. \quad (1.41)$$

Bayesian inference techniques are well established in the field of cosmology and have been used for a number of years due to pioneering work analysing the CMB data [90]. However, it is only in the last few years, due to the increased interest in the field after the reported EDGES detection (see section 1.6), that Bayesian techniques have been applied to data from 21-cm cosmology experiments.

For example, in an analysis of data from SARAS2 the authors used likelihood ratios, comparing fits with and without signal models, to rule out certain astrophysical scenarios for the global 21-cm signal [69]. The likelihood ratio is defined by

$$\mathcal{R} = \frac{\mathcal{L}_{M,1}}{\mathcal{L}_{M,2}}, \quad (1.42)$$

and when its value is  $> 1$  then model one is preferred over model two. While likelihood ratios give an intuitive means of determining whether one model is more probable relative to another, they are a *relative* metric, unlike the Bayesian Evidence which is an absolute metric. Further, likelihood ratios do not appropriately account for the choice of prior, which the Bayesian Evidence does.

In an analysis of data from the EDGES high band instrument, the authors generated several million simulations of their data from a wide prior and calculated their corresponding likelihoods to estimate the posterior

without using MCMC or Nested Sampling [91]. Finally, a significant number of experimental efforts have employed MCMC sampling to explore the astrophysical parameter space of the 21-cm power spectrum including LOFAR [92] and MWA [93].

Forward modelling of the global 21-cm signal in a Bayesian Nested Sampling pipeline has been pioneered by the REACH collaboration [57, 58]. Bayesian inference is employed significantly in this thesis, in which the Nested Sampling algorithm, implemented with MULTINEST [94–97] and POLYCHORD [88, 89], is used to model data from EDGES, LEDA, SARAS2, SARAS3, Planck, the Dark Energy Survey (DES) and HERA.

## 1.5 Machine learning and 21-cm Cosmology

Machine learning has many applications in the field of 21-cm cosmology. It has been shown that neural networks can capture the complex relationship between the 21-cm signal and the astrophysical process in the first billion years of cosmic history to a high degree of accuracy by [15] and more recently using two different approaches in [4] (chapter 3) and [98]. The semi-numerical simulations that are used to model this signal take of order a few hours to produce per signal, whereas neural network emulators can produce estimates of these signals to varying degrees of accuracy in 10s of milliseconds. This is particularly important if we want to physically model signals in Bayesian Nested Sampling or MCMC runs where we are making thousands of calls to our signal generator, as in the latter half of this thesis.

In these scenarios, parameters, such as the star formation efficiency and virial circular velocity among others, are passed as inputs to a feed forward neural network and corresponding estimates of the 21-cm brightness temperature are output. An example network is shown in Fig. 1.9. In this example, the network takes two inputs,  $p_1$  and  $p_2$ , and converts them to an output,  $O_1$ . Each node in the network is connected to all the nodes in the following layers, producing a ‘fully connected’ feedforward network. When a numerical value is passed from one layer to the next, it is multiplied by the weight of the connection,  $w_{ij}$  and a bias,  $\beta_{ij}$  is added where  $i$  corresponds to the node in the previous layer and  $j$  the node in the next layer. For example, when  $p_1$  is passed to the first node in the first layer, the value received is  $w_{11}p_1 + \beta_{11}$ . In addition, the node receives a weighted and shifted version of input  $p_2$ , since the network is fully connected, meaning that the value at node one (subscript) in layer one (superscript) is

$$a_1^1 = w_{11}p_1 + w_{21}p_2 + \beta_1^1, \quad (1.43)$$

where we define the bias at node one in layer one to be  $\beta_1^1 = \beta_{11} + \beta_{21}$  or more generally  $\beta_j^n = \sum_i \beta_{ij}$ . The value at node one in layer one is scaled via an activation function into  $a_1^{1'}$ .

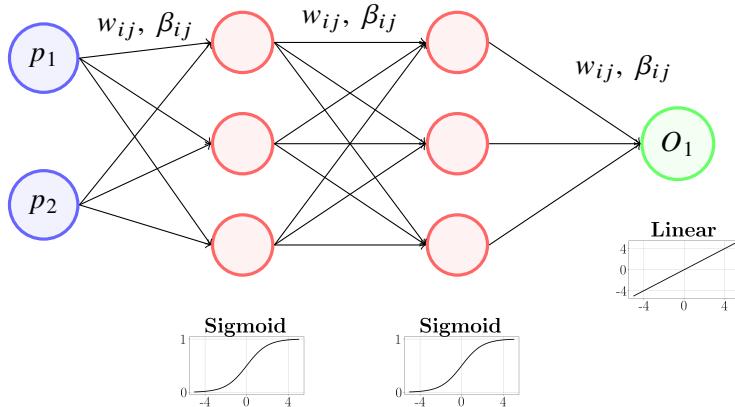
Activation functions can take many different forms and two examples, the sigmoid and linear, are shown in Fig. 1.9. The activation function is usually designed to introduce non-linearity in the network and allows the network to learn more complex relationships between the input and output. For this reason, the use of a linear output is often restricted to the final node and the sigmoid function is a common choice for the hidden layers, those between the input and output.  $a_1^{1'}$  is given by

$$a_1^{1'} = \frac{1}{1 + \exp(-a_1^1)}. \quad (1.44)$$

$a_1^{1'}$  is then passed to each node in the second layer and the process is repeated.

The number of layers and number of nodes in each layer is a choice for the user but can often be motivated by the demands of the problem as shown in chapter 3. More complex structures in theory can model more

complex relationships between the inputs and outputs of the network. However, arbitrarily increasing the size of the network can reduce the ability of the network to generalise to problems it has not seen before. This is a problem known as overfitting and is discussed in chapter 3 and [4].



**Figure 1.9:** The diagram shows a simple feed forward neural network that takes two input parameters,  $p_1$  and  $p_2$ , and transforms them to an output,  $O_1$ .  $w_{ij}$  and  $\beta_{ij}$  represent the weights and biases between layer  $i$  and layer  $j$ . The diagrams at the bottom of the figure show two different activation functions which in the hidden layers of the network, those between the input and output, are designed to introduce non-linear structure into the model.

The issue is therefore how to set the weights and biases of the neural network. These have to be learned based on some training data that we give to the network, i.e. a set of inputs and corresponding outputs, and a loss function. The loss function is chosen by the user, but typically we use a root mean squared error (or RMSE) loss given by

$$\sigma_{NN} = \sqrt{\frac{1}{N} \sum (y_{\text{true}} - y_{\text{pred}})^2}, \quad (1.45)$$

where  $y_{\text{true}}$  is the true value of the output and  $y_{\text{pred}}$  is what the network predicts.

In equation (1.45),  $y_{\text{pred}}$  is a function of the weights and biases of the network, and therefore we can say the same of the loss function,  $\sigma_{NN}(\mathbf{w}, \boldsymbol{\beta})$ . In order to optimise the values of the weights and biases, we minimise the loss function using a gradient descent algorithm, where at iteration  $i + 1$  the weights are updated according to

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \gamma \nabla \sigma(\mathbf{w}_i, \boldsymbol{\beta}_i), \quad (1.46)$$

where  $\gamma$  is known as the learning rate and determines the size of the step taken in the parameter space. The gradient effectively tells us which connections in the network have the largest impact on the loss. To evaluate the negative gradient of the loss function with respect to the weights and biases, we use the backpropagation algorithm.

The backpropagation algorithm begins by looking at how the weights and biases between the output and the previous layer need to change to improve the loss function for a particular example. To calculate the gradient of the loss with respect to each weight and bias between the output and the previous layer in the above example, we rely on the chain rule

$$\frac{\delta \sigma}{\delta w_{ij}^2} = \frac{\delta \sigma}{\delta O_1} \frac{\delta O_1}{\delta w_{ij}^2}. \quad (1.47)$$

For intermediate layers between the output and input there is an additional term accounting for the impact of the activation function. Since the output activation, the loss and the hidden layer nodes are known functions, we can use Automatic Differentiation to calculate their derivatives with respect to each weight and bias. Once

we have expressions for all the derivatives, the algorithm updates the weights according to

$$w_{ij}^2 - \gamma \frac{\delta\sigma}{\delta w_{ij}^2}. \quad (1.48)$$

starting with the weights between the input and the first hidden layer and propagating the change forward. The algorithm calculates the required changes in the weights and biases for a series of training examples, and then averages the changes over many examples to try to improve the accuracy of the network while maintaining the ability for it to generalise to unseen test examples.

We could do this for all examples in our training data and perform a gradient descent. However, this is a very computationally expensive task. For example, the network shown in Fig. 1.9 has 18 weights and biases creating a 36 dimensional space to optimize and consequently training neural networks is non-trivial and there are many tricks that can be employed to try and improve their accuracy. Some of these are discussed in chapter 3. However, the most basic trick typically employed is to divide the training data into batches and perform a stochastic gradient descent. The algorithm picks the most appropriate direction to move each weight and bias, and determines by how much to move them to improve the loss function for a specific subset of randomly chosen training examples. Doing this over many iterations for many batches, while sometimes increasing the loss for the whole training data set, eventually finds a good approximation to the true minimum or a local minimum that sufficiently optimizes the network and is much more computationally efficient than a full gradient descent.

The neural network emulator GLOBALEMU discussed in chapter 3 is a typically small feed forward network. The previous state of the art global 21-cm signal emulator 21cmGEM relies on several such networks and a decision tree [15]. The recently published 21cmVAE relies on a Variational Autoencoder to compress the input parameter space into a lower dimensional latent space in which to learn the relationship with the output 21-cm signal [98]. GLOBALEMU and 21cmVAE represent the ‘next generation’ emulators and the latter is more accurate but around 40 times slower than GLOBALEMU. Both emulators and 21cmGEM are sufficiently accurate for the current noise levels in present experimental efforts. Recently, some of the techniques used in the development of GLOBALEMU have been applied to the emulation of the 21-cm power spectrum in the HERA analysis [99].

In chapter 4, we discuss a different type of machine learning algorithm known as a Masked Autoregressive Flow (MAFs). MAFs are a form of normalising flow that string together a series of Masked Autoencoder for Distribution Estimation (MADE) [100] networks to perform density estimation. They are an example of a Neural Density Estimator. Normalising flows allow us to calculate the log-probability,  $\log(P(x))$ , for a set of samples  $\{x\}$  on some complex distribution. We assume that the multidimensional distribution  $P(x)$  can be decomposed into a set of one-dimensional conditional distributions

$$P(x) = \prod_i P(x_i|x_1, x_2, \dots, x_{i-1}), \quad (1.49)$$

where the index  $i$  represents the dimension. We model each conditional probability as a Gaussian distribution and assume a base standard normal distribution,  $\mathcal{N}(0, 1)$ , that is shifted and scaled by some standard deviation,  $\sigma_i$ , and mean,  $\mu_i$ , which can be written as

$$P(x_i|x_1, x_2, \dots, x_{i-1}) = \mathcal{N}(\mu_i(x_1, x_2, \dots, x_{i-1}), \sigma_i(x_1, x_2, \dots, x_{i-1})), \quad (1.50)$$

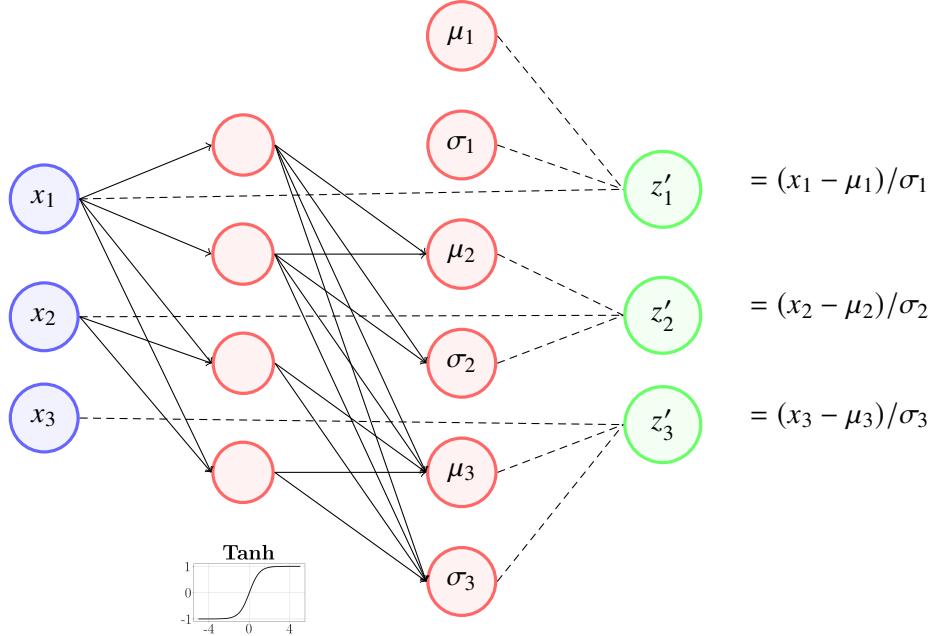
or alternatively written in terms of samples in  $x_i$

$$x_i = \sigma_i z_i + \mu_i, \quad (1.51)$$

where  $z_i \sim \mathcal{N}(0, 1)$ . The problem is how to estimate the appropriate set of  $\sigma_i$  and  $\mu_i$  to model the distribution of  $x_i$ . We do this with the MADE architecture which is shown in Fig. 1.10 (see also [101]) and the change of variables principle

$$P(x) = P(z) \left| \frac{\delta z}{\delta x} \right| \quad (1.52)$$

which describes how one probability distribution can be transformed into another. Here the magnitude of the determinant of the Jacobian represents the change in volume between  $P(x)$  and  $P(z)$  and it ensures that  $P(x)$  is normalised hence ‘normalising’ flows.



**Figure 1.10:** The MADE architecture, when trained appropriately, transforms samples from a complex probability distribution  $\{x\}$  on to samples from a standard normal distribution under the assumption that the probability distribution  $\{x\}$  can be broken into conditional one-dimensional probability distributions represented as Gaussians. Many different MADE networks can be stacked together and trained to produce a normalising flow increasing the expressivity of the density estimation. By definition normalising flows are bijective, and a trained implementation can be used to calculate  $\log(P(x))$  and draw samples from  $P(x)$ .

The MADE architecture takes in sets of samples from  $\{x\}$  and estimates the value of  $\mu_i$  and  $\sigma_i$  needed to transform them to samples on the standard normal distribution  $\{z\}$ . Doing this for many sets of samples in  $\{x\}$  adjusts the networks weights and biases appropriately to transform the distribution as a whole. The base of the MADE architecture is a fully connected neural network like that in Fig. 1.9 however certain connections are masked out to represent the fact that probability distributions and hence  $\mu_i$  and  $\sigma_i$  are conditional on samples from  $x_{i-1}$  dimensions. The masking is performed by multiplying the matrices of associated weights with a binary mask.

$\sigma_i$  and  $\mu_i$  are functions of the weights and biases in the MADE network, which can have multiple hidden layers with varying activation functions. The idea is to minimize the Kullback-Leibler (KL) divergence between the true probability distribution of the samples  $x$  and the prediction from the network  $P_\theta(x)$  given by

$$\mathcal{D}(P(x) || P_\theta(x)) = -\mathbb{E}_{P(x)} [\log P_\theta(x)] + \mathbb{E}_{P(x)} [\log P(x)], \quad (1.53)$$

where  $\theta$  represents the hyperparameters of the network. The second term in the KL divergence is independent

of the network and so for a minimization problem we can ignore this and

$$-\mathbb{E}_{P(x)}[\log P_\theta(x)] = -\frac{1}{N} \sum_{i=0}^N \log P_\theta(x_i), \quad (1.54)$$

meaning that our minimization problem becomes

$$\operatorname{argmax}_\theta \sum_{i=0}^N \log P_\theta(x_i), \quad (1.55)$$

or equivalently by a change of variables

$$\operatorname{argmax}_\theta \sum_{i=0}^N [\log P_\theta(z'_i) + \log \left| \frac{\delta z'_i}{\delta x_i} \right|], \quad (1.56)$$

where  $z'$  is a function of  $\theta$  [101]. Here the second term is just the derivative over the network and the first term is trivially calculated from the output  $\sigma$  and  $\mu$ .

We can stack many such networks together and train them in unison to transform more complicated networks under the following

$$\begin{aligned} z_0 &= \mathcal{N}(0, 1) \\ z_1 &= z_0 \sigma_1(z_0, w) + \mu_1(z_0, w) \\ &\vdots \\ x &= z_{n-1} \sigma_n(z_{n-1}, w) + \mu_n(z_{n-1}, w), \end{aligned} \quad (1.57)$$

in a normalising flow. Here  $\mu_j$  and  $\sigma_j$  are  $N$ -dimensional vectors,  $z_0$  is our multidimensional standard normal distribution and  $x$  is the multidimensional target distribution. The intermittent  $z_j$  distributions are output from one MADE and passed to the next and the weights and biases are adjusted across the whole chain at each training step. The transformation performed by a normalising flow is bijective in that there is a one-to-one relationship between samples in the base distribution and the target distribution, so we can go back and forth between the two passing the emulator samples on the standard normal distribution and recovering samples on the input posterior for example. Normalising flows and MAFs have many applications which are discussed in chapter 4.

There are many additional applications of machine learning in the field of 21-cm cosmology such as the use of Convolutional Neural Networks to accurately emulate tomographic images of the 21-cm field and modelling foregrounds in power spectrum analysis. However, we refer the reader to the literature for more details as the focus of this thesis is primarily on the global 21-cm signal.

## 1.6 Current experimental results

The 21-cm line from neutral hydrogen has long been understood to be a powerful probe of galactic physics [28, 33, 34]. However, it is only in the past 20-30 years that its potential usefulness as a cosmological probe has really been explored (see the historical review in [28]). Making tomographic images of the 21-cm signal as a function of redshift was first proposed in [29] and is a target of the upcoming Square Kilometre Array (SKA) [102].

To date, observational efforts have focused on two different but related statistics known as the 21-cm power spectrum and global or sky-averaged 21-cm signal. Detection of the 21-cm power spectrum has been

attempted by a number of different telescopes including LOFAR [LOw-Frequency ARray, 92, 103–106], MWA [Murchison Widefield Array, 93, 107, 108], HERA [20, 109] and GMRT [Giant Metrewave Radio Telescope, 110]. However, current observations are limited by our understanding of the systematics in the data.

The global 21-cm signal is in principle easier to detect than the power spectrum. Observations were first proposed in [30] and a number of ongoing experimental efforts are underway to detect this signal. While the detection of the power spectrum requires the use of an interferometer to measure the temperature variation across different angular scales on the sky, the global 21-cm signal can be detected, in principle, with a single antenna. A number of different experiments have attempted to make detections, are currently underway or have been proposed to make this measurement, including EDGES [8, 111], SARAS [16–19], LEDA [9], SCI-HI [Sonda Cosmológica de las Islas para la Detección de Hidrógeno Neutro, 112], BIGHORNS [Broadband Instrument for Global HydrOgen Reionisation Signal, 113], PRIZM [67], MIST [Mapper of the IGM Spin Temperature, 114], FARSCIDE [Farside Array for Radio Science Investigations of the Dark ages and Exoplanets, 38], and REACH [10, 57, 58, 115, 116].

In the following subsections, we explain the state of the field up to 2019, the start of this PhD, with some reference to more recent experimental results such as SARAS3 and HERA which are explored more in the main text. Part III of this thesis and the corresponding chapters detail improvements to our current understanding of the early Universe that I have been responsible for leading.

### 1.6.1 The sky-averaged 21-cm signal

A tentative detection of the global 21-cm signal was reported by the EDGES collaboration in 2018 [8]. If true, the detection implies rapid star formation, delayed and then rapid X-ray heating, an earlier than expected end to the epoch of reionization and exotic astrophysics such as an increased radio background [117–121] or interactions between Baryons and Dark Matter [122–127] to describe the depth of the signal. However, there are concerns about the data analysis.

For example, [128] analysed the effects of modelling the data with different foreground models after noting that some of the parameters for the physically motivated foreground model used in the original analysis were found to have unphysical values. In particular, the optical depth of the ionosphere and the temperature of the electrons in the ionosphere were both found to be negative, which is clearly unphysical. The authors showed that there may be a sinusoidal systematic present in the data and that the choice of model strongly impacts the ability to recover the reported absorption feature. A similar reanalysis of the data looked at the impact of modelling the foregrounds with Maximally Smooth Functions [129] and recovered a similar sinusoidal feature in the data. This analysis is repeated in chapter 2 in an effort to validate the described fitting algorithm. Further work in [60] suggested that the reported absorption feature could be produced by a discontinuity in the ground plane.

In [130] the authors fitted the EDGES data with a number of different combinations of models including physically motivated 21-cm signals, Gaussian profiles, flattened Gaussian profiles, sinusoidal systematics and different foreground models. They then used the Bayesian evidence to determine which combination of model components best described the data and consistently found evidence for a sinusoidal systematic in the data. While the work suggested the presence of some kind of signal in the data the cosmological nature of this signal is still in question.

There have been a number of works that have attempted to describe the depth of the EDGES absorption feature [e.g. 14, 122, 123, 127, 131]. One such explanation is an increased radio background above the CMB and

observations from the balloon experiment ARCADE2 [48] and the Long Wavelength Array (LWA) [49] both suggested the presence of such a background. As previously mentioned, there are concerns about the modelling of Galactic foregrounds in the analysis of the data from ARCADE2 and LWA [50]. Most explanations for the depth of the EDGES absorption feature, including those with excess radio backgrounds above the CMB, cannot explain the rapid star formation and X-ray heating that is implied.

The field is collectively working on an independent detection of the global 21-cm signal to validate or refute the cosmological nature of the EDGES absorption feature. Currently, the only independent experiment to produce accurate enough residuals is SARAS3 and in an analysis of data from this radiometer the authors ruled out the EDGES absorption feature with approximately 95% confidence [54]. However, the SARAS3 band only partially overlaps with the EDGES absorption feature and the reported signal is approximately smooth in the SARAS3 band meaning that it could easily have been fitted out by their foreground model. In chapter 6 we use the residuals from the SARAS3 data to constrain the astrophysics of the first galaxies. Prior to this work and the work in chapter 5 and chapter 7 there were a handful of constraints on the structure of the global 21-cm signal from other experiments.

LEDA reported some of the first constraints on the depth and width of a Gaussian model of the global 21-cm signal in 2016 [132]. They use a Bayesian framework and perform their analysis with MULTINEST. They found that the global 21-cm signal should have a depth  $> -890$  mK and a width  $> 6.5$  MHz at 95% confidence between  $z = 15 - 27$ .

In [69] and [70] the authors took data from the SARAS2 instrument and placed constraints on physical scenarios for the 21-cm signal. In both sets of work, the authors use a set of  $\approx 260$  theoretical models from [13] charting the parameter space of the global 21-cm signal. In the first paper, they use a likelihood ratio to determine the probability that the data favours the presence of a given signal compared to its absence. In the latter, they take each signal multiply it by a scale factor and fit the data for the value of this scale factor alongside a foreground model using a least squares algorithm. In this analysis, a scale factor of one would imply a detection, and any deviation from one tells you how unrealistic the physical scenario is. These two works are discussed more in chapter 5 however briefly they disfavoured scenarios with rapid reionization and weak X-ray heating.

In a series of three papers, data from the EDGES High band instrument (90 – 190 MHz,  $z \approx 6.5 - 15$ ) was analysed to constrain the properties of early galaxies. In the first paper, the authors placed constraints on a phenomenological Gaussian 21-cm signal models and tanh-based approximations to the evolution of the neutral fraction. They rule out weak global 21-cm signals with narrow widths and reject models in which reionization was complete within  $\Delta z \lesssim 2$  [133] in general agreement with previous results.

In the follow-up analysis [134] the authors fit the data with a set of 10,000 physical simulations of the 21-cm signal generated with the simulation code 21cmFAST. They subtract each simulation from the data and perform a maximum likelihood estimation fitting a log-polynomial foreground model to estimate the likelihood for each of the 10,000 sets of astrophysical parameters. In this work the simulations are parameterised by the minimum virial temperature of the star forming halos, the X-ray luminosity, the minimum X-ray energy that can escape a galaxy and the ionizing efficiency of sources. They disfavour high values of the virial temperature and ionizing efficiency as well as intermittent values of the X-ray luminosity. They also combine their constraints with constraints on  $\tau$  from Planck and on the hydrogen neutral fraction from quasars which leads to a much tighter constraint on the virial temperature. In general these constraints correspond to signals with late heating

and sharp reionization histories.

The final paper analysing this data [91] used a semi-Bayesian method to constrain theoretical models of the 21-cm signal from the same class described throughout this thesis. In this work they use an early version of the neural network signal emulator 21cmGEM and generate 3.2 million signal models. They then perform a similar analysis to the previous paper, where for each signal model they fit a polynomial foreground to the data minus the signal using a least squares algorithm then calculate a likelihood for each set of 21-cm signal parameters. In this work they extend on the previous analysis by multiplying the likelihood estimate by a uniform prior for the 21-cm parameters to estimate marginal 1D and 2D posteriors effectively performing a brute force Bayesian analysis. They consistently disfavour low values of the X-ray efficiency,  $f_X$ , and in the absence of additional constraints, from Planck and quasars, disfavour high values of the minimum virial velocity,  $V_c$ .

Since in the latter two works the signal parameters are fixed the analysis essentially optimizes the foreground on slices through the foreground plus signal parameter space which could result in the peak of the posterior being missed. Nevertheless, they represented some of the first works attempting to fit physical signal models in a semi-Bayesian manner to data from a 21-cm experiment and demonstrated the usefulness of neural network signal emulators for the first time. The results in chapter 5 and chapter 6 represent the state-of-the-art in both the data analysis for inferences about and constraints on the properties of the first galaxies from sky-averaged 21-cm experiments.

### 1.6.2 The 21-cm power spectrum

To date, the best upper limits on the magnitude of the 21-cm power spectrum are from HERA at  $z \approx 8$  and  $z \approx 10$ . Although a number of experiments have placed limits on its magnitude at different redshifts such as LOFAR at  $z = 9.1$  and  $z = 9.3 - 10.6$  and MWA at  $z = 6.5 - 8.7$  covering a range of different angular scales on the sky.

The focus of the thesis is on the sky-averaged 21-cm experiment. However, in chapter 7 we combine the data from HERA and SARAS3 to constrain the properties of the first galaxies. As such, our main interest here is in the derivation of constraints from the upper limit measurements of the power spectrum.

The LOFAR limit at  $z = 9.1$  was used in [92] to constrain the properties of the inter galactic medium using an MCMC exploration of models from the reionization code GRIZZLY [135, 136]. They use an erf-likelihood function around the upper limit measurement, however they calculate the exclusion likelihood as  $1 - \mathcal{L}$  and draw conclusions from this quantity. A similar work was performed with data from MWA in the range  $z = 6.5 - 8.7$  [93]. Further, analysis of the LOFAR data [106] and MWA data [108] attempted to constrain the physics of the 21-cm signal using the semi-numerical code 21cmFAST, an MCMC algorithm and in a similar fashion to the other two works an exclusion likelihood.

The exclusion likelihood or inverse likelihood used in the above works is a useful way to gain some intuition about the types of 21-cm models that the upper limits disfavour, however the interpretation of the resultant “posteriors” has to be done with care. This is because the erf-likelihood

$$\mathcal{L}(\theta_{21}) = \mathcal{P}(\text{"no detection"}|\theta) = \prod_i^{N_d} \frac{1}{2} \left( 1 + \operatorname{erf} \left[ \frac{d_i - m_i(\theta_{21})}{\sqrt{2}\sigma_i} \right] \right), \quad (1.58)$$

where  $d_i$  corresponds to the upper limit measurement and  $m_i$  represents the model, is the probability of the instrument not making a detection with respect to some parameters,  $\theta = \{\theta_{21}, d\}$  i.e.  $\mathcal{P}(\text{"no detection"}|\theta)$ . Sampling  $1 - \mathcal{L}$  gives the probability of a detection given some parameters and the upper limit  $\mathcal{P}(\text{"detection"}|\theta)$ .

The inverse likelihood therefore tells us if  $d$  was a detection, what would the parameters be. However, we should note that this is not the same as the probability that the measurement is *explained* by the set of parameters  $\theta$  equation (1.30).

The posteriors arrived at via the two likelihood functions,  $\mathcal{P}(\text{"detection"}|\theta)$  and  $\mathcal{P}(\text{"nodetection"}|\theta)$ , are therefore very different. The first tells you what parameters  $\theta$  would correspond to a detection at the upper limit. The latter tells you which parts of the parameter space are still viable given the upper limit. For this reason, I do not discuss the conclusions from the works that used the inverse likelihood function in detail.

Constraints on models with an excess radio background from a phenomenological synchrotron radio background, like those discussed in this thesis, from LOFAR were presented in [105] at  $z = 9.1$ . The authors found that the maximum contribution to the radio background allowed at  $z = 9.1$  by LOFAR is an excess of 9.6%. In a similar fashion to the analysis of the EDGES high band analysis the authors calculate the likelihood for around 8000 models again using an erf-likelihood function around the upper limit measurement. In addition to placing a constraint on the excess radio background, the LOFAR limits disfavour low values of the X-ray production efficiency in scenarios with both an excess and no excess radio background in agreement with the results from the EDGES high band analysis.

HERA recently reported an upper limit on the 21-cm power spectrum from 18 nights of data [109] and followed this up with a limit from that same 18 nights and an additional 78 nights of data [99]. The limits are at  $z \approx 8$  and  $z \approx 10$ . In both papers they explore constraints on a series of different semi-numerical signal models including from 21cmFAST and from the same set as described earlier in this introduction (e.g. [121, 131]). For the latter they use equation (1.58) and find that they rule out scenarios with large radio backgrounds and galaxies that are inefficient X-ray heaters. This is in agreement with the results presented in chapter 5 and chapter 6 which have been published around the same time as the HERA papers. In chapter 7 we discuss the HERA results in the context of chapter 5 and chapter 6 in more detail.



## **Part II**

# **Data Analysis Toolbox**



In the following three chapters, we introduce a series of novel data analysis tools designed to improve the prospects of detecting the 21-cm signal. In chapter 2, we discuss a foreground modelling tool, `MAXSMOOTH`, that uses quadratic programming to model the smooth components of data from global or sky-averaged 21-cm signal experiments. The chapter is based on the first author paper [3], which was published in Monthly Notices of the Royal Astronomical Society, and aims to address the issues raised in section 1.3.1.

We then describe a novel neural network based (see section 1.5) global 21-cm signal emulator called `GLOBALEMU` which improved on the previous state-of-the-art emulator by a factor of two in accuracy and 102 in runtime. The network uses a physically motivated preprocessing of the training data, comprised of a series of signals from semi-numerical simulations discussed in the introduction, and reparameterizes the network to take the independent variable, redshift or frequency, as an input alongside the physical parameters describing the signal. The chapter is based on the first author paper [4] published in Monthly Notices of the Royal Astronomical Society.

Finally, in chapter 4, we introduce and discuss a machine learning based preprocessing tool to analyse the outputs of Bayesian analysis pipelines called `MARGARINE`. We demonstrate a number of applications of this tool, which performs density estimation with Kernel Density Estimators and Normalising Flows (see section 1.5), such as the calculation of marginal Bayesian statistics and the efficient combination of constraints from different data sets on the same parameters. The work is based on two first author papers [1, 5] the first of which was published as proceedings for the 41st International Conference on Bayesian and Maximum Entropy methods in Science and Engineering.



CHAPTER

2

**MAXSMOOTH**

## 2.1 Introduction

Maximally Smooth Functions (MSFs), functions with no inflection points or zero crossings in higher order derivatives, were first proposed by [62] for modelling foregrounds in experiments to detect spectral signatures from the Epoch of Recombination. They are designed for modelling smooth structures in experimental data that are several orders of magnitude larger than non-smooth signals of interest and to leave behind signals and, where present, systematics in residuals [see also, 61]. MSFs can be considered part of a family of functions, which we refer to as Derivative Constrained Functions (DCFs) and includes functions with no turning points, Completely Smooth Functions (CSFs) and functions with a select number of non-zero crossing high order derivatives, Partially Smooth Functions (PSFs). We refer to the relatively high magnitude smooth components of the data as foregrounds throughout this chapter.

Our primary focus here is the application of DCFs to the field of sky-averaged or global 21-cm cosmology. [30] suggested that reionization of neutral hydrogen in the early universe would result in a sharp step in the global spectrum of the sky, and that this signal should be separable from the smooth spectrum emission that dominates the sky temperature at radio wavelengths, 70 – 240 MHz. [137] showed that the foreground emission, if modelled using a low-order polynomial, could be subtracted from the global sky spectrum to retrieve signals from the Epoch of Reionization (EoR) of order 100 mK; [138] and [139] expounded on this work, including further instrumental effects.

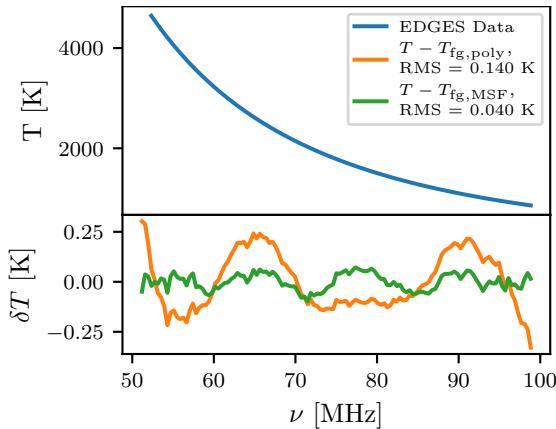
In comparison to unconstrained polynomials, DCFs are better able to separate the smooth foreground spectra from the anticipated 21-cm signals and instrumental systematics [61]. This motivates their use in global 21-cm cosmology experiments such as; REACH [115], SARAS [16], EDGES [8], LEDA [9], PRIZM [67], SCI-HI [Sonda Cosmológica de las Islas para la Detección de Hidrógeno Neutro, 112] and MIST (Mapper of the IGM Spin Temperature, <http://www.physics.mcgill.ca/mist/>).

The structure of the 21-cm signal is determined by the interplay of a number of different astrophysical processes discussed in the introduction to this thesis and by the change in the dominant processes over time. The exact timing and intensity of the signal is only broadly understood within a theoretical parameter space [13, 15, 70, 91]. Experiments that search for the global 21-cm signal are attempting to detect a signal, according to standard  $\Lambda$ CDM cosmology, a few hundred milli-Kelvin in foregrounds of up to 4 - 5 orders of magnitude brighter.

These high-magnitude foregrounds are dominated by synchrotron and free-free emission in the Galaxy and extragalactic radio sources which have smooth power law structures as discussed in chapter 1. Modelling of these foregrounds without signal loss is essential for an accurate detection, and not always possible with unconstrained polynomials. However, unconstrained polynomials remain the traditionally used foreground model in 21-cm experiments [8, 70, 91].

Studies of Gunn-Peterson troughs in quasar spectra and of the CMB anisotropies put the end of the EoR at  $z \approx 6$  [22, 140, 141]. It is predicted that the onset of star formation occurred at  $z \sim 30$  [142]. Consequently, the bandwidth of interest for 21-cm cosmology is approximately 50 – 200 MHz where the bandwidth is determined by the intrinsic frequency of the 21-cm transition, 1420.4 MHz, which is redshifted by the expansion of the universe.

Fig. 2.1 shows an example of the application of MSFs to 21-cm cosmology. Here we have fitted publicly available data from the EDGES low band experiment with an MSF and a 5<sup>th</sup> order polynomial of the form given by equation (2) in [8]. The MSF is shown to fit the foreground to a higher degree of precision and potentially



**Figure 2.1:** An example of the applications of MSFs and `MAXSMOOTH` using the publicly available global 21-cm EDGES low-band experiment data. The top panel shows the EDGES data, blue, and the bottom panel shows the residuals after fitting and removing an unconstrained polynomial, orange, and an MSF, green. The MSF fits the data to a higher degree of accuracy and reveals a systematic that has been partially removed by the polynomial as part of the foreground. The polynomial is given by equation (2) of [8] and is taken to be 5<sup>th</sup> order. We use the best fitting 11<sup>th</sup> order MSF from the built-in library in `MAXSMOOTH` to illustrate the quality of fit recovered.

reveal a sinusoidal systematic which has been previously identified in the data [128–130]. A more detailed discussion of the EDGES data can be found in section 2.5.2.

The constrained nature of DCFs, namely that specific derivatives do not cross zero in the domain of interest, makes fitting these functions a non-trivial task. While this has been historically performed with penalised optimization routines such as Basin hopping [143] and Nelder-Mead [144] we find that the use of quadratic programming [145] is considerably more computationally efficient and reliable. Our DCF code, `MAXSMOOTH` is therefore based on quadratic programming and uses the Python based convex optimisation code, CVXOPT [ConVeX OPTimization, 146]. A discussion of quadratic programming can be found in Appendix A of [3].

The constraints on a DCF are not explicitly linear but are piecewise linear with various combinations of positive and negative signs on the high order derivatives. For low order,  $N$ , DCFs testing every combination of positive and negative signs is a computationally inexpensive task. However, this becomes increasingly time-consuming with increasing  $N$  and `MAXSMOOTH` uses a ‘cascading routine’ in combination with a directional exploration to quickly search the discrete sign spaces.

DCF s can be formed from a variety of different basis functions and `MAXSMOOTH` has a built-in library. The library is not intended to be complete, and the user can implement their own basis functions. For basis functions in which the number of high order derivatives is not finite, `MAXSMOOTH` constrains derivatives up to order  $m = N - 2$ . MSFs form the basis of the analysis performed in this chapter and we focus on their uses and applications. However, the description of MSFs can be more broadly applied to DCFs.

In section 2.2 we describe MSFs in more detail. In section 2.3 we discuss the application of quadratic programming to DCF fitting with reference to CVXOPT and the piecewise linear constraints on the derivatives. Section 2.4 discusses the fitting algorithm implemented by `MAXSMOOTH` and compares its efficiency to alternative optimization routines. In section 2.5 we discuss the use of DCFs in 21-cm cosmology. The application of the fitting routine to the EDGES low band data [8] and data from LEDA [9] is illustrated in section 2.5.2 and section 2.5.3. We conclude in section 2.6.

The co-authors of [3], Will Handley, Anastasia Fialkov, Eloy de Lera Acedo, Lincoln Greenhill and Danny Price, provided comments on the manuscript that this chapter is based on.

## 2.2 Maximally Smooth Functions

MSFs are functions which feature no inflection points or zero crossings in higher order derivatives [see 61, 62].

The coefficients of the basis functions are constrained such that the  $m^{\text{th}}$  order derivative satisfies

$$\frac{d^m y}{dx^m} \geq 0 \text{ or } \frac{d^m y}{dx^m} \leq 0, \quad (2.1)$$

where  $x$  and  $y$  define the independent and dependent variables and for MSFs  $m \geq 2$ . More generally for DCFs  $m$  can be greater or equal to any value or equal to a select set of derivative orders. MAXSMOOTH features seven built-in DCFs which we use for fitting. Their functional forms and derivatives are shown in Tab. 2.1.

Name	Function	Derivatives
Normalised Polynomial	$y = y_0 \sum_{k=0}^N a_k \left(\frac{x}{x_0}\right)^k$ (2.2)	$\frac{d^m y}{dx^m} = y_0 \sum_{k=0}^{N-m} \frac{(m+k)!}{k!} a_{m+k} \left(\frac{x}{x_0}\right)^{m+k}$
Polynomial	$y = \sum_{k=0}^N a_k x^k$ (2.3)	$\frac{d^m y}{dx^m} = \sum_{k=0}^{N-m} \frac{(m+k)!}{k!} a_{m+k} x^k$
Difference Polynomial	$y = \sum_{k=0}^N a_k (x - x_0)^k$ (2.4)	$\frac{d^m y}{dx^m} = \sum_{k=0}^{N-m} \frac{(m+k)!}{k!} a_{m+k} (x - x_0)^k$
Log Polynomial	$y = \sum_{k=0}^N a_k \log_{10} \left(\frac{x}{x_0}\right)^k$ (2.5)	$\frac{d^m y}{d \log_{10}(x/x_0)^m} = \sum_{k=0}^{N-m} \frac{(m+k)!}{k!} a_{m+k} \log_{10} \left(\frac{x}{x_0}\right)^k$
Log Log Polynomial	$y = 10^{\sum_{k=0}^N a_k \log_{10}(x)^k}$ (2.6)	$\frac{d^m \log_{10}(y)}{d \log_{10}(x)^m} = \sum_{k=0}^{N-m} \frac{(m+k)!}{k!} a_{m+k} \log_{10}(x)^k$
Legendre	$y = \sum_{k=0}^N a_k P_k(z)$ (2.7)	$\frac{d^m y}{dz^m} = \sum_{k=0}^{N-m} \frac{(-1)^m P_k^m(z)}{(1-z^2)^{\frac{m}{2}}}$
Exponential	$y = y_0 \sum_{k=0}^N a_k \exp\left(-k \frac{x}{x_0}\right)$ (2.8)	$\frac{d^m y}{dx^m} = y_0 \sum_{k=0}^N \left(\frac{-k}{x_0}\right)^m a_k \exp\left(-k \frac{x}{x_0}\right)$

**Table 2.1:** The DCF models built-in to MAXSMOOTH along with expressions for their  $m^{\text{th}}$  order derivatives. For all functions  $y_0$  and  $x_0$  are pivot points in the data sets. More details on each DCF function can be found in the text.

Generally, the DCF functions can be decomposed in terms of basis functions,  $\phi$  and parameters,  $a_k$  as

$$y = \sum_{k=0}^N a_k \phi_k(x). \quad (2.9)$$

For the first DCF shown in Tab. 2.1, the Normalised Polynomial model, the basis functions are given by

$$\phi_k(x) = y_0 \left( \frac{x}{x_0} \right)^k, \quad (2.10)$$

where  $y_0$  and  $x_0$  correspond to a pivot point, defaulted to the mid-point, in the data sets. The normalised nature of this polynomial model ensures that the fit parameters,  $a_k$ , are of order unity. Here  $N$  is the order of the DCF and can take on any value. However, for a given model and data set there is a limiting value beyond which a further increase in  $N$  does not increase the quality of the fit and this is illustrated in section 2.5.2 and Fig. 2.11. The DCF model will have powers from 0 to  $N - 1$ .

Two more basis functions built-in to MAXSMOOTH are given by the Polynomial and Difference Polynomial models where the latter is based on the basis function used in [61]. The built-in set of models is not meant to be complete with the intention for it to be extended in the future.

The fourth basis function built-in to MAXSMOOTH, Log Polynomial, produces an MSF in  $y - \log_{10}(x)$  space. MAXSMOOTH is also capable of fitting a DCF in  $\log_{10}(y) - \log_{10}(x)$  space given in Tab. 2.1 as the Log Log Polynomial model. In this instance the function is constrained by derivatives in  $\log_{10}(y) - \log_{10}(x)$  space. This can be advantageous in situations where the foregrounds are expected to take on a power law structure.

The penultimate basis function in the MAXSMOOTH library of models is built from the orthogonal Legendre Polynomials,  $P_k(z)$ , where  $z$  is a variable of length  $y$  over the range [-1, 1]. The  $m^{\text{th}}$  order derivatives of this model are determined by the Associated Legendre Polynomials,  $P_k^m(z)$ . By definition the Legendre polynomials are a linear combination of the basis functions of the Normalised Polynomial model. This is true also for the Polynomial model and less trivially for the Difference Polynomial model.

Typically the basis functions are designed so that equation (2.9) has a finite number of high order derivatives and is consequently polynomial in nature. However, more elaborate models with an infinite number of derivatives are plausible if we consider these functions to be maximally smooth when all derivatives with  $2 \leq m \leq N - 2$  are constrained. The final DCF model built-in to MAXSMOOTH, the Exponential DCF, is an example of this with exponential basis functions. This model fails at high  $N$  where the exponential cannot be computationally calculated. However, it is a useful example of a DCF with infinite derivatives and performs well with low values of  $N$ . The exact value of  $N$  at which this basis begins to fail is determined by the magnitude of the  $x$  data. An alternative example of a basis with infinite derivatives that MAXSMOOTH is capable of fitting would be a polynomial function with non-integer powers.

Generally, the form of the basis function is important in determining the quality of the residuals and careful exploration of the basis functions are needed in order to draw sensible conclusions about the data set. Again, this is illustrated with an example in section 2.5.2. We also note that DCFs fitted in  $y - \log_{10}(x)$  space,  $\log_{10}(y) - \log_{10}(x)$  space,  $y - x$  space or  $y - z$  space are not equivalent since the form of the constraints and the function that we minimise are different in each case. This is discussed further in section 2.3.

With appropriate normalisation MAXSMOOTH will be able to transform any basis function into a ‘standard’ form, which can be solved easily and transformed back into the initial basis function choice. Designing and automating such a normalisation is the subject of ongoing work. Provided this ‘standard’ form is chosen well

such that it will always return the best quality fits and is computationally solvable with quadratic programming, the initial choice of basis function will largely be negated. Its form will only be determined by the need of the user to model their foreground using a specific model. For example, in 21-cm cosmology this specific model may be a linearised physical model of the data fitted as an MSF. While normalisation remains absent in MAXSMOOTH, the user has the ability to input normalised  $x$  and  $y$  data.

For quadratic programming, the method used here to fit MSFs, it is useful to reformulate equation (2.9) as a matrix equation. Explicitly we have discrete data points  $y_i$  and  $x_i$  which means that  $\phi_k(x)$  forms a two dimensional matrix,  $\Phi$ . The matrix of basis functions has dimensions  $(D \times N)$  where  $D$  is the length of  $y$  and  $N$ , as before, is the order of the function. We write this as

$$\begin{bmatrix} y_0 \\ \vdots \\ y_D \end{bmatrix} = \begin{bmatrix} \phi_{00} & \dots & \phi_{0(N-1)} \\ \vdots & \ddots & \\ \phi_{D0} & & \phi_{D(N-1)} \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_{(N-1)} \end{bmatrix}, \quad (2.11)$$

where  $\phi_{ik} = \phi_k(x_i)$ . We can summarise this as

$$\mathbf{y} = \Phi \mathbf{a}, \quad (2.12)$$

where  $\mathbf{a}$  is a column vector of length  $N$  representing the parameters. For the polynomial basis function in equation (2.3), the element  $\phi_{D0}$ , or  $\phi_0(x_D)$ , has the form  $x_D^0$  and  $\phi_{0(N-1)}$  has the form  $x_0^{(N-1)}$ .

Reformulating equation (2.1) in terms of matrices for quadratic programming is more complicated. If we take the definition of the condition with the derivative  $\leq 0$  and write this in the form of a matrix for a given derivative order  $m$  we find

$$\begin{bmatrix} \frac{d^m y}{dx^m} 0 \\ \vdots \\ \frac{d^m y}{dx^m} D \end{bmatrix} \leq \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (2.13)$$

where both matrices are columns of length  $D$ . Each row in the derivative matrix corresponds to an evaluation of the  $m^{\text{th}}$  order derivative for a given  $y_i$  and  $x_i$ .

We can expand the elements of the derivative matrix out into a matrix of derivative prefactors,  $\mathbf{G}$ , and the matrix of parameters  $\mathbf{a}$ , as in equation (2.12), which is useful for implementing quadratic programming. This is best illustrated with an example, we will look at the simple case of  $N = 3$  with one constrained derivative  $m = 2$ . We will say that our data sets have a length  $D = 4$  and choose the simplest functional form for our MSF given by equation (2.3). In this case  $G$  is given by

$$G_k^m(x_i) = \frac{(m+k)!}{k!} x_i^k, \quad (2.14)$$

for the range  $k = 0$  up to but not including  $N - m$ . For this problem  $k$  has only one value 0 which would produce a column matrix of elements of length  $D$ . However, since we need to multiply this by the column matrix  $\mathbf{a}$  of length  $N$  then the matrix  $\mathbf{G}$  should have dimensions of  $D \times N$ . The additional elements in this instance are 0 so that the product of these elements with the corresponding elements of  $\mathbf{a}$  equals 0. For example here the evaluation of the second order derivative  $\mathbf{G}\mathbf{a}$  for the first data element will be

$$\frac{d^2 y}{dx_0^2} = a_0 0 + a_1 0 + a_2 \frac{(2+0)!}{0!} (x_0)^0. \quad (2.15)$$

Generally, if the row elements of  $\mathbf{G}$  have a position from  $0 - (N - 1)$  then the elements with position  $\leq m - 1$  will be 0. The matrix  $\mathbf{G}$  for our specific problem then becomes

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & G_0^2(x_0) \\ 0 & 0 & G_0^2(x_1) \\ 0 & 0 & G_0^2(x_2) \\ 0 & 0 & G_0^2(x_3) \end{bmatrix}, \quad (2.16)$$

which when multiplied by  $\mathbf{a}$  gives us the evaluation of the derivatives as a column matrix with length  $D = 4$ .

For quadratic programming we need one matrix expression for all of the constraints on our function. Our definition of  $\mathbf{G}$  scales with the order of the DCF so that for  $N = 4$  we have

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & G_0^2(x_0) & G_1^2(x_0) \\ 0 & 0 & G_0^2(x_1) & G_1^2(x_1) \\ 0 & 0 & G_0^2(x_2) & G_1^2(x_2) \\ 0 & 0 & G_0^2(x_3) & G_1^2(x_3) \\ 0 & 0 & 0 & G_0^3(x_0) \\ 0 & 0 & 0 & G_0^3(x_1) \\ 0 & 0 & 0 & G_0^3(x_2) \\ 0 & 0 & 0 & G_0^3(x_3) \end{bmatrix}, \quad (2.17)$$

which includes the prefactors on both the  $m = 2$  and  $m = 3$  derivatives. We can, therefore, re-write equation (2.1) for  $\leq 0$  and equation (2.13) as

$$\mathbf{G}\mathbf{a} \leq \mathbf{0}, \quad (2.18)$$

where generally  $\mathbf{G}$  will have shape  $(CD) \times N$  and  $\mathbf{0}$  will have a length  $CD$ . Here  $C$  is the total number of constrained derivatives and in the two examples above  $C = 1$  and  $C = 2$  respectively. We can write the first case in equation (2.1) in one of two ways as  $\mathbf{G}\mathbf{a} \geq \mathbf{0}$  or as  $-(\mathbf{G}\mathbf{a}) \leq \mathbf{0}$ . For the implementation of quadratic programming used in MAXSMOOTH the second is the most useful and a full discussion of this can be found in section 2.3.

In some cases we find that a DCF with one or more high order derivatives free to cross zero is needed to better fit the data. It is to this effect that the potential to allow zero crossings to the fit is built-in to MAXSMOOTH. However, MAXSMOOTH will not force zero crossings and produce a Partially Smooth Function (PSF) if it can find a better solution without the need.

## 2.3 Fitting Derivative Constrained Functions Using Quadratic Programming

A brief overview of quadratic programming can be found in Appendix A of [3] and what follows is a discussion of the specific problem of fitting DCFs with quadratic programming.

When fitting a curve using the least-squares method we minimize

$$\chi^2 = \sum_k^D (y_k - y_{\text{fit}, k})^2, \quad (2.19)$$

where  $y_{\text{fit}, k}$  denotes the elements of the fitted model. We can substitute equation (2.12) for the fitted model and re-write this in terms of matrices as

$$\chi^2(\mathbf{a}) = (\mathbf{y} - \boldsymbol{\Phi}\mathbf{a})^T (\mathbf{y} - \boldsymbol{\Phi}\mathbf{a}), \quad (2.20)$$

where we are looking for solutions of the parameters  $\mathbf{a}$  that minimise  $\chi^2(\mathbf{a})$ . When expanded out this becomes

$$\chi^2(\mathbf{a}) = \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \boldsymbol{\Phi} \mathbf{a} + \mathbf{a}^T \boldsymbol{\Phi}^T \boldsymbol{\Phi} \mathbf{a}. \quad (2.21)$$

Since  $\mathbf{y}^T \mathbf{y}$  is a constant it is irrelevant for the minimisation problem and we can ignore it. We can also divide through by the factor of 2 and this leaves

$$\chi^2(\mathbf{a}) = \frac{1}{2} \mathbf{a}^T \mathbf{Q} \mathbf{a} + \mathbf{q}^T \mathbf{a}, \quad (2.22)$$

where

$$\mathbf{Q} = \boldsymbol{\Phi}^T \boldsymbol{\Phi} \text{ and } \mathbf{q}^T = -\mathbf{y}^T \boldsymbol{\Phi}. \quad (2.23)$$

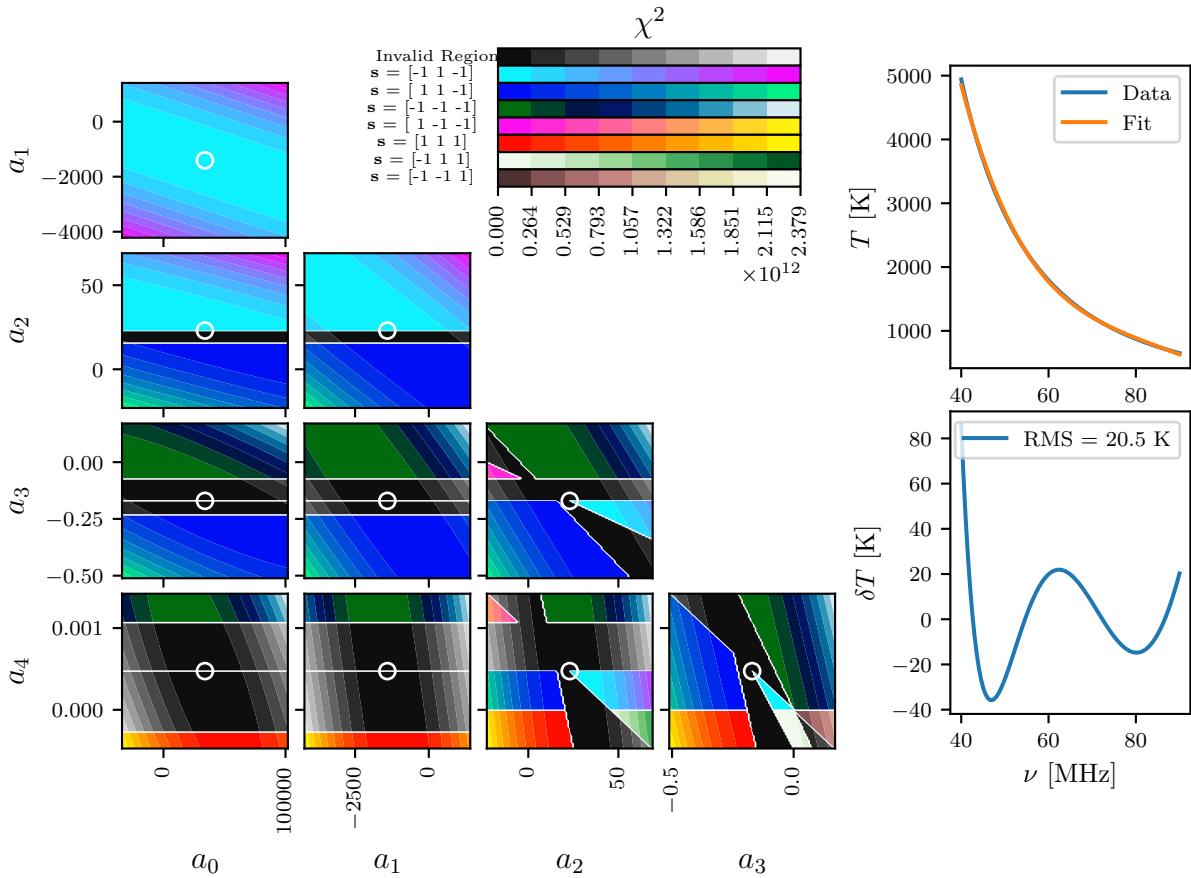
As previously discussed in section 2.2, the constraint in equation (2.1) is not explicitly linear but is two separately testable linear constraints. The quadratic program solver CVXOPT minimizes equation (2.22) subject to equation (2.18). It requires  $\mathbf{G}$  and  $\mathbf{0}$  as inputs which explains the motivation behind defining the stacked matrix of derivatives as  $\mathbf{G}\mathbf{a}$ . In CVXOPT the identity is fixed and we cannot directly constrain the problem via a greater than or equals inequality.

We can initially force all the derivatives to be positive, the first of the two conditions in equation (2.1), by multiplying each element in  $\mathbf{G}$  by a negative sign as discussed in the previous section. However, it will not necessarily be the case that the optimal DCF fit will have an entirely positive or entirely negative set of derivatives. Rather than forcing the entire matrices to produce positive derivatives we can multiply the elements of  $\mathbf{G}\mathbf{a}$  corresponding to given derivatives by a negative sign. Consequently, we have to analyse different discrete sets of sign combinations in order to find the best fit.

We refer to the combination of signs on  $\mathbf{G}$  as the `MAXSMOOTH` signs,  $\mathbf{s}$ , and we can incorporate this into our definition of  $\mathbf{G}$  so that it becomes  $\mathbf{G}(\mathbf{s})$ .  $\mathbf{s}$  is a vector of length  $C$  and each element is either given by 1 for a positive sign or  $-1$  for a negative sign. For example, in equation (2.17) since both derivatives are negative the `MAXSMOOTH` signs are  $\mathbf{s} = [1, 1]$ . For an  $N^{\text{th}}$  order MSF there are  $N - 2$  derivatives with  $m \geq 2$ , consequently, there are  $2^{(N-2)}$  sign combinations. For low order  $N$  we can explore this space exhaustively at reasonable computational cost with CVXOPT. However, as  $N$  becomes larger, the total number of sign combinations rapidly increases. While  $N = 4$  has 4 potential sign combinations, we find that  $N = 13$  has 2048. This would mean performing an exponentially increasing number of CVXOPT fits, which will become increasingly time-consuming. An alternative approach, navigating through the discrete sign spaces, is detailed in section 2.4.

We can visualise equation (2.1) by varying the parameters of an optimal MSF fit over a given range to get a better understanding of the constraints. In order to perform this analysis, we use a simulated noiseless global 21-cm foreground following  $x^{-2.5}$ . We perform a 5<sup>th</sup> order MSF fit with `MAXSMOOTH` on this data using equation (2.3) to find the optimal foreground parameters. While this fit will not return the best  $\chi^2$ , as shown in Fig. 2.2, it is sufficient to allow us to investigate how variation in the parameters affects the constraints. We vary each parameter's value 200% either side of the optimum found and sample these ranges using 100 points.

Fig. 2.2, left panel, shows the parameter space for the fit described above. Black regions in the figure are combinations of parameters for which the condition in equation (2.1) is violated. The coloured regions are regions in which the condition is upheld where their colour is related to the `MAXSMOOTH` sign combinations. Each panel in the figure shows the parameter space for two of the five parameters, and the contour lines show the values of  $\chi^2$  across the parameter ranges. While varying the parameters relevant to each panel, we maintain



**Figure 2.2: Left:** The parameter space of a 5<sup>th</sup> order MSF of the form given by equation (2.3), fit to data generated with  $y = x^{-2.5}$ . The tested parameter ranges are taken to be 200% on either side of the optimal results found by MAXSMOOTH. We maintain the optimal parameter values for three of the parameters when varying the two corresponding to each panel. Black regions of the graph show parameter combinations that violate equation (2.1) and coloured regions correspond to viable regions. The central sampled point, highlighted by the white circles (at their centres), in each panel corresponds to the optimum parameters and the optimum MAXSMOOTH sign combination on the derivatives. Transitions through regions of violation correspond to changes in the sign of one or more high order derivatives. A change in sign of parameter  $a_4$  corresponds to a change in the sign of the final constrained derivative. Since this derivative is a constant, there is no violated region between these two possible sign combinations. **Top Right:** The data and the fitted MSF, where  $T$  represents the measured sky temperature and  $\nu$  is the frequency. **Bottom Right:** The residuals after subtracting the fitted MSF from the data.

all others at their optimal values found with MAXSMOOTH. The contour lines help us to determine correlations between the parameters, and this is particularly useful when fitting a physically motivated DCF.

Transitions through a region of violation between viable regions correspond to changes in the MAXSMOOTH sign of one or more of the derivatives. This is illustrated by the use of different colour maps across the different viable regions. For example, in the panel corresponding to variation in  $a_0$  and  $a_2$  for  $a_2 \leq 15$   $s = 1$  and for  $a_2 \geq 25$   $s = -1$  for the  $m = 2$  derivative. The transitions become more complex when varying parameters  $a_2$ ,  $a_3$  and  $a_4$  because these parameters affect the magnitude and signs of multiple high order derivatives. We also see transitions between regions of different sign combinations when  $a_4$  switches sign. There is no region of violation between these viable regions because a constant value of  $a_4 = 2.5 \times 10^{-4}$  meets the MSF constraint, as will a constant value of 0 or  $a_4 = -2.5 \times 10^{-4}$ .

The equivalent graph in 5 dimensions, varying all parameters around their optimal values, would feature 5 dimensional convex faceted regions in which equation (2.1) is met with a unique set of MAXSMOOTH signs. This concept scales up and down to higher and lower dimensions of parameter space.

The parameters  $a_0$  and  $a_1$  do not affect the constrained derivatives of the MSF or the validity of the

conditions, and the associated colour map gives the optimum `MAXSMOOTH` signs. Since the central sample point of each panel corresponds to the optimum parameters for the fit, this will always be a viable region and will have the same `MAXSMOOTH` sign combination as the panel corresponding to  $a_0$  and  $a_1$ . Fig. 2.2 illustrates this point and, where visible, the central viable sample point always corresponds to derivatives of order  $m = 2, 3$  and  $4$  having  $s = [-1, 1, -1]$ .

Where the central sample point in each panel is not visible, this is a relic of the sample rate across the parameter space. For example in the panel corresponding to variation of  $a_0$  and  $a_4$ , between the two large regions of violation there is a single value of  $a_4$ , the optimum, that meets the condition given by equation (2.1). The tight constraints around the optimum values in the four panels in the bottom left corner of the figure are due to the independence of the constraints for an MSF on  $a_0$  and  $a_1$  and the strong dependence on  $a_3$  and  $a_4$ .

We performed the equivalent fit with the logarithmic basis that has derivatives constrained in  $\log_{10}(y) - \log_{10}(x)$  space. The associated parameter graph can be found in Appendix B in [3], and a comparison with the graph presented in this section shows that the constraints are much less severe in logarithmic space. The weak constraints mean that all the discrete sign combinations on the derivatives have similar minimum  $\chi^2$  values. This becomes a problem when attempting to quickly search the discrete `MAXSMOOTH` sign spaces and is discussed further in section 2.4.2.

Generally, the above conclusion will be specific to the data being fitted here, and this analysis is not a complete exploration of the basis functions available. However, since the basis functions in  $y - x$  space are all related by linear combinations of each other, we find similar parameter distributions for all. Importantly, the analysis highlights the effect that the choice of basis function has on the quality of fit and ease of fitting, as well as demonstrating the constrained nature of MSFs. These plots can be produced using `MAXSMOOTH` for any DCF fitting problem.

## 2.4 Navigating Discrete Sign Spaces

This section discusses in more detail the fitting problem, defines the `MAXSMOOTH` algorithm and compares its efficiency with an alternative fitting algorithm. To restate concisely the problem being fitted we have

$$\begin{aligned} \min_{\mathbf{a}, s} \quad & \frac{1}{2} \mathbf{a}^T \mathbf{Q} \mathbf{a} + \mathbf{q}^T \mathbf{a}, \\ \text{s.t. } & \mathbf{G}(s) \mathbf{a} \leq \mathbf{0}. \end{aligned} \tag{2.24}$$

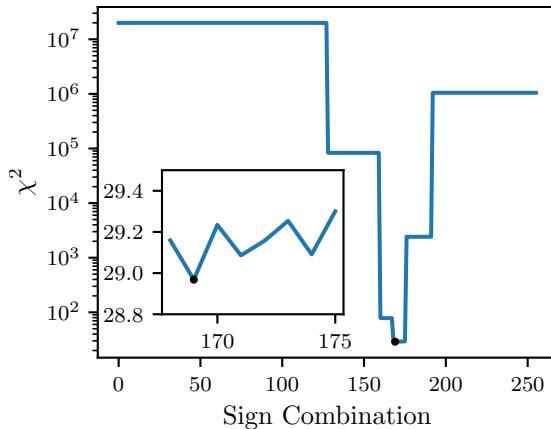
A ‘problem’ in this context is the combination of the data, order, basis function and constraints on the DCF.

With `MAXSMOOTH` we can test all possible sign combinations. This is a reliable method and, provided the problem can be solved with quadratic programming, will always give the correct global minimum. When the problem we are interested in is ‘well-defined’, we can develop a quicker algorithm that searches or navigates through the discrete `MAXSMOOTH` sign spaces to find the global minimum. Each sign space is a discrete parameter space with its own global minimum, as discussed in section 2.3. Using quadratic programming on a fit with a specific sign combination will find this global minimum, and we are interested in finding the minimum of these global minima.

A ‘well-defined’ problem is one in which the discrete sign spaces have large variance in their minimum  $\chi^2$  values and the sign space for the global minimum is easily identifiable. In contrast, we can have an ‘ill-defined’ problem in which the variance in minimum  $\chi^2$  across all sign combinations is small. This concept of ‘well-defined’ and ‘ill-defined’ problems is explored further in the following two subsections.

Sign Combination	Binary	MAXSMOOTH Signs, $s$
23	00010111	[+1, +1, +1, -1, +1, -1, -1, -1]
24	00011000	[+1, +1, +1, -1, -1, +1, +1, +1]
25	00011001	[+1, +1, +1, -1, -1, +1, +1, -1]
26	00011010	[+1, +1, +1, -1, -1, +1, -1, +1]
27	00011011	[+1, +1, +1, -1, -1, +1, -1, -1]

**Table 2.2:** The table illustrates the relationship between the binary representation of the sign combination number and the MAXSMOOTH signs,  $s$ . A 1 in the  $(N - 2)$  bit binary representation for an MSF corresponds to a negative MAXSMOOTH sign (positive derivative). The signs and their respective combination numbers are used in the fitting routine and for the visualisation of the  $\chi^2$  distribution, as shown in Fig. 2.3.



**Figure 2.3:** The  $\chi^2$  values for the discrete sign combinations on the derivatives for a 10<sup>th</sup> order MSF fit in  $y - \log(x)$  space to a simulated 21-cm foreground. Sign combination 0 corresponds to negative derivatives (positive MAXSMOOTH signs) and 255 corresponds to positive derivatives (negative MAXSMOOTH signs). The signs and sign combination numbers are related by the  $(N - 2)$  bit binary representation of the number. The global minimum is shown as a single black data point. In the insert, the distribution around the global minimum is shown, and here the axis have the same meaning as in the main plot.

## 2.4.1 Well-defined Problems and Discrete Sign Space Searches

### 2.4.1.1 The $\chi^2$ Distribution

We investigate the distribution of  $\chi^2$  values, shown in Fig. 2.3, for a 10<sup>th</sup> order MSF fit of the form given by equation (2.5) to a simulated 21-cm foreground, like that shown in Fig. 2.2. We add Gaussian noise with a standard deviation of 0.5 to the foreground. For a typical 21-cm experiment this noise is unrealistic and would mask any potential signal in the data, however, it illustrates the behaviour of the MAXSMOOTH algorithm when fitting a difficult problem.

In Fig. 2.3, a combination of all positive derivatives (negative signs) and all negative derivatives (positive signs) corresponds to sign combination numbers 255 and 0 respectively. Specifically, the MAXSMOOTH signs,  $s$ , are related to the sign combination number by its  $C$  bit binary representation, here  $C = (N - 2)$ . In binary, the sign combination numbers run from 00000000 to 11111111. Each bit represents the sign on the  $m^{\text{th}}$  order derivative, with a 1 representing a negative MAXSMOOTH sign. For example, the sign combinations surrounding number 25 are shown in Tab. 2.2.

Although we note that Fig. 2.2 corresponds to a different problem, we would expect a similar parameter space for the fit performed here. Each region in the equivalent figure would correspond to a single sign combination, and the associated minimum  $\chi^2$  value in the regions would give us the data that informs Fig. 2.3.

The distribution appears to be composed of smooth steps or shelves; however, when each shelf is studied

closer, we find a series of peaks and troughs. This can be seen in the subplot of Fig. 2.3 which shows the distribution in the neighbourhood of the global minimum found in the large or ‘global’ well. This type of distribution with a large variance in  $\chi^2$  is characteristic of a ‘well-defined’ problem. We use this example  $\chi^2$  distribution to motivate the MAXSMOOTH algorithm outlined in the following subsection.

#### 2.4.1.2 The MAXSMOOTH Sign Navigating Algorithm

Exploration of the discrete sign spaces for high  $N$  can be achieved by exploring the spaces around an iteratively updated optimum sign combination. The MAXSMOOTH algorithm begins with a randomly generated set of signs for which the objective function is evaluated, and the optimum parameters are found. We flip each individual sign one at a time, beginning with the lowest order constrained derivative first. When the objective function is evaluated to be lower than that for the current optimum sign combination, we replace it with the new set and repeat the process in a ‘cascading’ routine until the objective function stops decreasing in value.

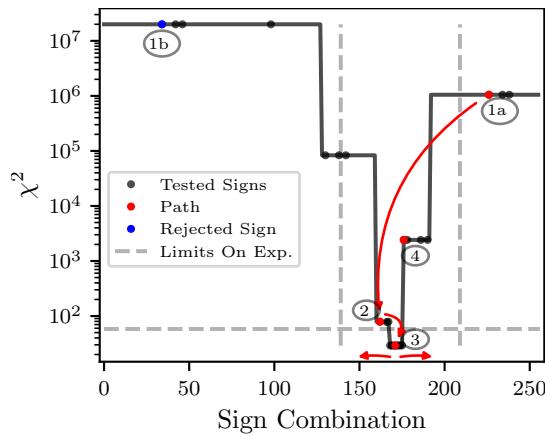
The local minima shown in Fig. 2.3 mean that the cascading algorithm is not sufficient to consistently find the global minimum. We can demonstrate this by performing 100 separate runs of the cascading algorithm on the simulated 21-cm foreground. We use equation (2.5) with  $N = 10$  to model the MSF as before. We find the true global minimum 79 times and a second local minimum 21 times. For an MSF fit to this simulated data the difference in these local minima is insignificant,  $\Delta\chi^2 = 0.12$ . However, we see the same behaviour with real data sets from EDGES and LEDA, and when performing joint fits of foregrounds and signals of interest  $\Delta\chi^2$  can greatly increase.

The abundance of local minima is determined by the magnitude and presence of signals, systematics and noise in the data. When jointly fitting a signal/systematic model with a DCF foreground, the signal/systematic parameters are estimated by another fitting routine in which MAXSMOOTH is wrapped. The initial parameter guess will not be a perfect representation of any real systematic or signal. This, along with a large noise, can produce a large difference between the true foreground and the ‘foreground’ being modelled, causing the presence of local minima to become more severe.

To prevent the routine terminating in a local minimum, we perform a complete search of the sign spaces surrounding the minimum found after the cascading routine. We refer to this search as a directional exploration and impose limits on its extent. In each direction, we limit the number of sign combinations to explore, and we limit the maximum allowed increase in  $\chi^2$  value. We prevent repeated calculations of the minimum for given signs and treat the minimum of all tested signs as the global minimum.

We run the consistency test again, with the full MAXSMOOTH algorithm, and find that for all 100 trial fits we find the same  $\chi^2$  found when testing all sign combinations. In Fig. 2.4 the red arrows show the approximate path taken through the discrete sign spaces against the complete distribution of  $\chi^2$ . Point (1a) shows the random starting point in the algorithm, and point (1b) shows a rejected sign combination evaluated during the cascade from point (1a) to (2). Point (2), therefore, corresponds to a step through the cascade. Point (3) marks the end of the cascade and the start of the left directional exploration. Finally, point (4) shows the end of the right directional exploration where the calculated  $\chi^2$  value exceeds the limit on the directional exploration.

The global well tends to be associated with signs that are all positive, all negative or alternating. We see this in Fig. 2.3 where the minimum falls at sign combination number 169 and number 170, characteristic of the derivatives for the simulated 21-cm foreground, corresponds to alternating positive and negative derivatives from order  $m = 2$ . Standard patterns of derivative signs can be seen for all data following approximate power



**Figure 2.4:** The cascading and directional exploration algorithm in practice against the entire  $\chi^2$  distribution for the fit to the simulated 21-cm experiment data. The red arrows show the approximate path of the cascade and directional exploration. The limits on the directional exploration are also shown as dashed grey lines. The point (1a) shows the initial random starting point and point (1b) shows a rejected sign combination in the cascade routine from (1a) to (2). Point (2) is an accepted step through the cascade with a  $\chi^2$  value smaller than the previous minimum. Point (3) marks the end of the cascade and the start of the left directional exploration. Finally, point (4) illustrates the end of the right directional exploration when the  $\chi^2$  value exceeds the limit on the directional exploration. The black dots mark the entirety of the searched sign combinations.

laws. All positive derivatives, all negative and alternating signs correspond to data following the approximate power laws  $y \approx x^k$ ,  $y \approx -x^k$ ,  $y \approx x^{-k}$  and  $y \approx -x^{-k}$  (see Appendix C in [3]).

The MAXSMOOTH algorithm assumes that the global well is present in the  $\chi^2$  distribution, and this is often the case. The use of DCFs is primarily driven by a desire to constrain previously proposed polynomial models to foregrounds. As a result, we would expect that the data being fitted could be described by one of the four approximate power laws highlighted above and that the global minimum will fall around an associated sign combination. In rare cases the global well is not clearly defined and this is described in the following subsection.

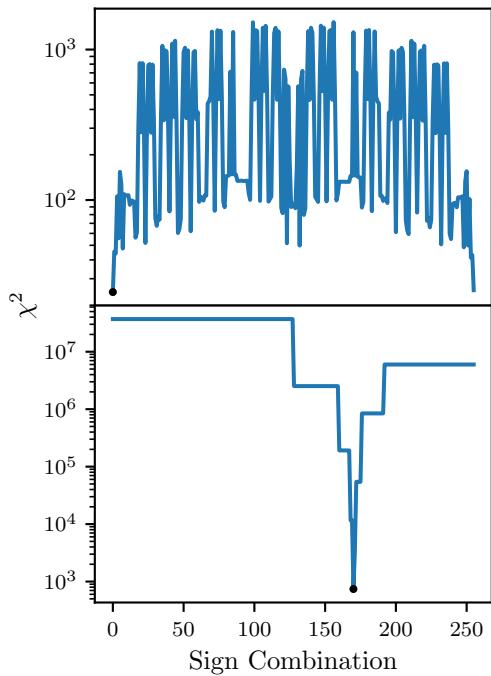
#### 2.4.2 Ill-defined Problems and their Identification

We can illustrate an ‘ill-defined’ problem, with a small variation in  $\chi^2$  across the MAXSMOOTH sign spaces, by adding a 21-cm signal into the foreground model and fitting this with a 10<sup>th</sup> order logarithmic MSF defined by equation (2.6). We take an example signal model from [13] and add Gaussian noise with a standard deviation of 20 mK more typical of a 21-cm experiment. The resultant  $\chi^2$  distribution with its global minimum is shown in the top panel of Fig. 2.5.

The global minimum, shown as a black data point, cannot be found using the MAXSMOOTH algorithm. The cascading algorithm may terminate in any of the approximately equal minima, and the directional exploration will then quickly terminate because of the limits imposed.

If we repeat the above fit and perform it with equation (2.2) we find that the problem is well-defined with a larger  $\chi^2$  variation across sign combinations. This is shown in the bottom panel of Fig. 2.5. The results, when using equation (2.6), are significantly better than when using, equation (2.2) meaning it is important to be able to solve ‘ill-defined’ problems. This can be done by testing all MAXSMOOTH signs, but knowing when this is necessary is important if you are expecting to run multiple DCF fits to the same data set. We can focus on diagnosing whether a DCF fit to the data is ‘ill-defined’ because a joint fit to the same data set of a DCF and signal of interest will also feature an ‘ill-defined’  $\chi^2$  distribution.

We can identify an ‘ill-defined’ problem by producing the equivalent of Fig. 2.5 using MAXSMOOTH and



**Figure 2.5: Top Panel:** The  $\chi^2$  distribution found when fitting simulated 21-cm experiment data with the logarithmic basis function, equation (2.6). The distribution has a noise like structure and is difficult to solve with the MAXSMOOTH sign navigating algorithm. However, the global minimum can be found by testing all sign combinations with MAXSMOOTH. The symmetric nature of the distribution stems from the symmetric nature about 0 of the high order derivatives in logarithmic space. **Bottom Panel:** The same as above using a normalised polynomial given by equation (2.2). The distribution is clearly defined and easily searchable with the sign navigating routine. The difference between this result and that shown above can be used to understand what makes a problem ‘ill-defined’.

visually assessing the  $\chi^2$  distribution for a DCF fit. Alternatively, we can use the parameter space plots to identify whether the constraints are weak or not, and if a local minimum is returned from the sign navigating routine then the global minimum in these plots will appear off-centre.

Assessment of the first derivative of the data can also help to identify an ‘ill-defined’ problem. For the example problem this is shown in Fig. 2.6 where the derivatives have been approximated using  $\Delta y / \Delta x$ . Higher order derivatives of the data will have similarly complex or simplistic structures in the respective spaces. There are many combinations of parameters that will provide smooth fits with similar  $\chi^2$  values in logarithmic space, leading to the presence of local minima. This issue will also be present in any data set where the noise or signal of interest are of a similar magnitude to the foreground in  $y - x$  space.

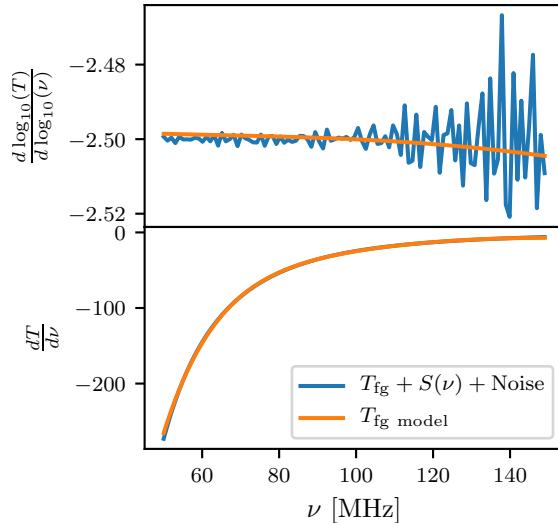
### 2.4.3 Comparison With Basin hopping and Nelder-Mead Methods

For comparison of the two methods, testing all sign combinations and navigating through sign spaces, we generate a signal  $y$  with polynomial dependence on the coordinate  $x$  and a Gaussian random noise with a standard deviation of 0.5

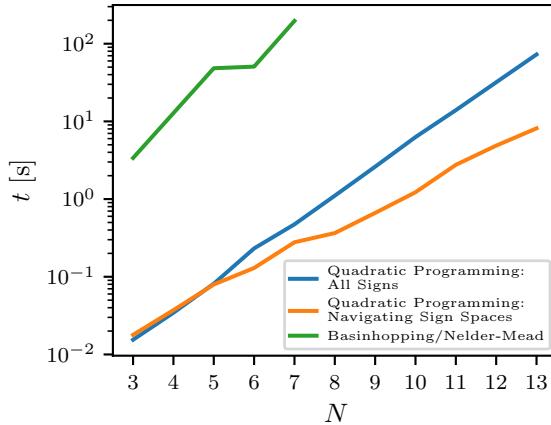
$$y = 0.6 + 2x + 4x^3 + 9x^4 + \text{noise}. \quad (2.25)$$

We fit this data with a 10<sup>th</sup> order MSF of the form described by equation (2.2) and assess the  $\chi^2$  distribution to find that the problem is well-defined. This is as expected since the data follows an approximate  $x^k$  power law, and we are fitting in linear space.

The algorithm run time becomes a significant issue when performing joint fits of foregrounds, signals of interest and/or systematics in which multiple DCF fits have to be performed. The time taken to perform both



**Figure 2.6:** The approximated first derivatives, blue lines, of the mock 21-cm experiment data in logarithmic, top panel, and linear, bottom panel, spaces which correspond to DCF fits performed using equation (2.6) and equation (2.2) respectively. Also shown as orange lines are the approximate first derivatives of the respective best fits for comparison. The large variance of the first derivative of the data in logarithmic space highlights that there are many different ways to fit this data set ‘well’ producing the  $\chi^2$  distribution we see in Fig. 2.5.



**Figure 2.7:** The time taken by MAXSMOOTH to fit MSFs of varying order  $N$  to the data, described by equation (2.25) using the two built-in quadratic programming methods. For comparison, the time taken by a Basinhopping/Nelder-Mead routine is also shown up to  $N = 7$  after which the routine fails to find the optimum solutions without adjustments to the routine parameters. All the fits were performed with equation (2.2) and on the same computing hardware.

in-built MAXSMOOTH routines is shown in Fig. 2.7. It is quicker to partially sample the available spaces for high  $N$  than testing all sign combinations, and as discussed for ‘well-defined’ problems this will return the minimum  $\chi^2$ .

The runtime of the sign navigating routine is dependent on the starting sign combination, the limits imposed on the directional exploration (which is the dominating factor) and the width of the global well. There is no consistent measure of the difference in time taken to fit the data between the two MAXSMOOTH methods. However, for the sign navigating routine we are inevitably fitting for a smaller number of the sign combinations than when testing all.

Fig. 2.7 also shows the time taken to fit the data with equation (2.2) using a Basinhopping routine followed by a Nelder-Mead algorithm, hereafter referred to as BHNM. These two algorithms have been previously used either separately or in conjunction for fitting MSFs [61, 62, 129]. We find that the BHNM method is approximately two orders of magnitude slower than MAXSMOOTH. Between  $N = 3$  and 7 we find a maximum

percentage difference in  $\chi^2$  of  $\approx 0.04\%$  when comparing the BHNM method with the results from `MAXSMOOTH`.

The primary difference in the approaches comes from the division of the parameter space into discrete sign spaces. The BHNM method attempts to search the entire parameter space and penalises parameter combinations that violate equation (2.1). However, assessment of Fig. 2.2 highlights that this is not a convenient method because across the whole parameter space there are multiple local minima with different sign combinations and transitioning from one ‘basin’ to another is not trivial for a heavily constrained parameter space. By dividing the space up into discrete sign spaces with `MAXSMOOTH`, we can test the entirety of the parameter space, unlike when using the BHNM method, guaranteeing we find the global minimum. We could perform the same division of the space and in each discrete sign space perform a Nelder-Mead or equivalent routine, however we use quadratic programming because it is designed specifically for fast and robust constrained optimisation.

For the BHNM method, we show here only fits up to  $N = 7$  after which it begins to fail without further adjustment of the routine parameters. The freedom to adjust these parameters can be considered a disadvantage that leaves the user to determine whether the routine parameters they have chosen produce a true global minimum. In contrast, `MAXSMOOTH` is designed to reliably give the optimum result, with the only adjustable routine parameters being the total number of CVXOPT iterations and the directional exploration limits.

All the fits performed in this section were done using the same computing hardware.

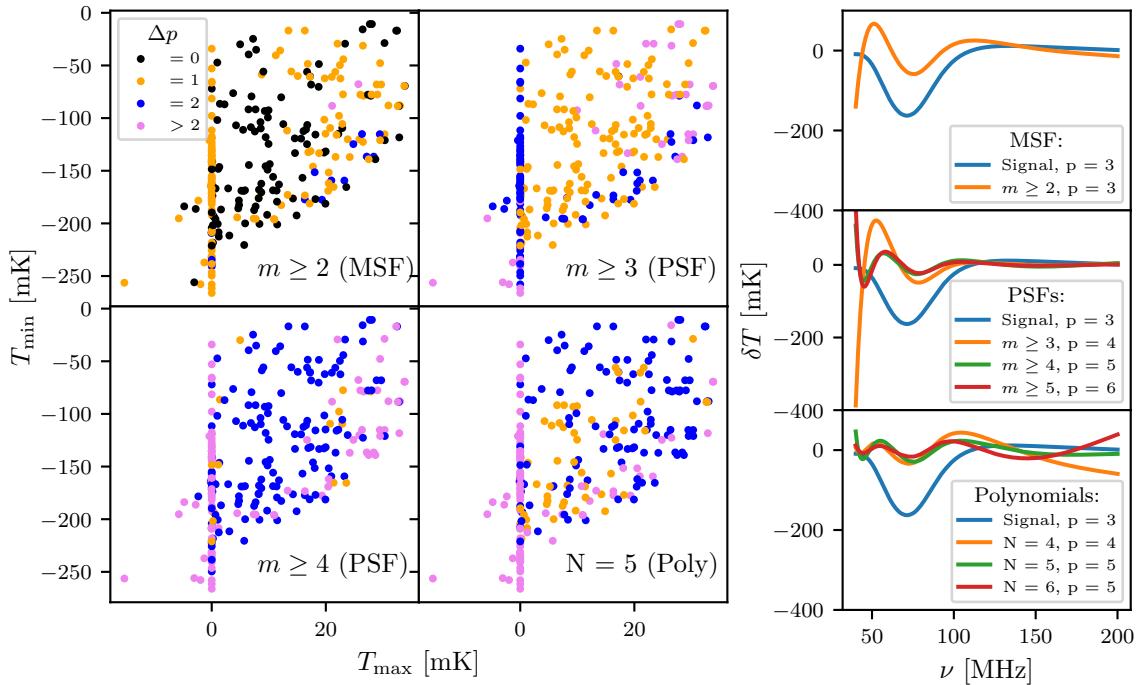
## 2.5 Applications for 21-cm Cosmology

### 2.5.1 The Recovery of Model 21-cm Signals

A discussion of MSFs and a comparison to unconstrained polynomials for global 21-cm cosmology can be found in [61]. Foreground modelling with high order MSFs is shown to accurately recover global 21-cm signals in simulated data and unconstrained polynomials are shown to introduce additional turning points, when compared to those in a mock signal. The number of additional turning points is shown to increase with polynomial order. The addition of extra turning points can obscure the signal of interest and lead to the false identification of systematics. In the following subsections, we look at fitting foregrounds for 21-cm experiments with DCFs and compare these to unconstrained polynomial fits.

Deviations from a smooth structure can be introduced in data by experimental systematics, or they can be intrinsic to the foreground. In the case of a smooth foreground, by using PSFs we can correct for non-smooth structure directly with our foreground model rather than separately fitting systematics. PSFs allow for zero crossings in the high order derivatives, but remain more constrained than traditionally used polynomial fits. However, lifting constraints on a DCF model has the potential to also result in signal loss, where the level of signal loss is dependent on the presence of non-smooth structure in the foreground and the number of lifted constraints.

In the following analysis, the quality of the fit in terms of Root Mean Square (RMS) is of secondary importance. An unconstrained high order polynomial will generally produce residuals with a lower RMS than a DCF. However, a correctly constrained DCF will leave behind the structure of a signal in the residuals, meaning it can be identified easily. As a measure of signal structure in the residuals, we take the number of turning points,  $p$  as used by [61]. The global 21-cm signal is expected to have a distinct number of turning points, 2 – 3, across the bandwidth 40 – 200 MHz determined by various astrophysical processes (see chapter 1). The successful application of DCFs in identifying global signals, being reliant on the presence of non-smooth structure in the



**Figure 2.8:** **Left:** The difference in the number of turning points between the fit residuals and the simulated signal,  $\Delta p = p_{\text{Fit Residuals}} - p_{\text{Signal}}$ , as a function of the maximum and minimum temperatures of the signal for a smooth foreground model. This is shown for four different foreground models, and in each panel the data points correspond to one of the 264 mock data sets considered. The graph shows that the MSF fit is the most likely model to return the structure of the signal, and that a correctly constrained PSF can more frequently preserve the signal than an unconstrained low order polynomial. It also shows that the fit quality is dependent on the maximum and minimum temperatures of the signal. All the DCF fits used to produce this graph were logarithmic and 13<sup>th</sup> order. **Right:** Shown are examples of how the addition of allowed zero crossings in the high order derivatives of a DCF can affect the residuals and how accurately they preserve the turning points of any signal present. We use a signal model, blue line in all panels, from the theoretically motivated set presented in [13] along with a model of a 21-cm experiment foreground. Fits with an MSF and polynomials are also shown for comparison with the number of turning points,  $p$ , displayed for each of the residuals (see legend for details). While the residuals after fitting and removing an MSF do not identically match the signal, it is the best representation of the tested foreground models. In this case, the residuals represents a smooth baseline subtracted version of the signal as discussed in [61].

data, is therefore bandwidth dependent. However, a comparison of the number of turning points in a mock signal to the residuals after removing a DCF fit from simulated data including the same mock signal will help to identify the degree to which DCFs preserve signals.

### 2.5.1.1 DCFs and 21-cm cosmology

To compare the performance of DCFs and unconstrained polynomials, we use the sample of 264 signal models,  $S(\nu)$  over the bandwidth  $\nu = 40 - 200$  MHz, presented in [13] and used by [70]. The models are provided by A. Fialkov and are publicly available at: <https://people.ast.cam.ac.uk/~afialkov/Collab.html>. We add to these models a foreground given by a  $\nu^{-2.5}$  power law to produce simplistic mock data sets. The data sets are noiseless and while this is unrealistic, we would not expect the addition of noise to obscure any larger signal structure present in the data from a 21-cm experiment with a low radiometer noise. We fit each simulated data set with an MSF, low order unconstrained polynomials and a set of PSFs. All the fitted DCF foreground models are 13<sup>th</sup> order and of the form given by equation (2.6). We test all sign combinations for the DCF fits in this section for reasons that were explained in section 2.4.2 and find that the chosen DCF model and order provides the best fits after testing the built-in MAXSMOOTH models.

Fig. 2.8, left panel, shows the difference in the number of turning points,  $\Delta p$  for the residuals,  $p_{\text{Fit Residuals}}$ ,

	$m \geq 2$ (MSF)	$m \geq 3$ (PSF)	$m \geq 4$ (PSF)	$N = 5$ (Poly)
$p_{\text{Signal}}$	116	1	0	0
$p_{\text{Signal}} + 1$	132	135	15	52
$p_{\text{Signal}} + 2$	14	88	147	116

**Table 2.3:** The table shows the total number of fits, using one of four foreground models, to the smooth foreground plus signal simulations from [13] that have residuals with the same number of turning points as the signal,  $p_{\text{Signal}}$ . The data corresponds to that shown in the left panel of Fig. 2.8. We also show instances where the recovered residuals have one or two additional turning points.

and for the signal,  $p_{\text{Signal}}$ , using four different foreground models as a function of the maximum brightness temperature,  $T_{\max}$ , and minimum temperature,  $T_{\min}$ , of the simulated signal. Each data point corresponds to one of the 264 mock data sets and  $\Delta p = 0$  signifies that the residuals have the same number of turning points as the signal. The unconstrained polynomial fits have the same functional form as the DCFs.

We can quantify the probability of returning residuals with the same number of turning points as the model signals. Tab. 2.3 shows the total number of residuals for each fit type that returned the same  $p$  as the simulated signals and 1 or 2 additional turning points. The MSF fits return  $p_{\text{Signal}}$  for 44% of the mock data sets and 99% of the time it returns at most  $p_{\text{Signal}}$  plus 2 additional turning points. They are the most likely, of the tested foreground models to return the correct number of turning points. The equivalent figures for the 5<sup>th</sup> order logarithmic unconstrained polynomial, one of the most frequently used foreground fits in 21-cm cosmology, are 0% and 64% and for the PSF with  $m \geq 3$  they are 0.004% and 85%. The statistics suggest that modelling foregrounds with MSFs and well constrained PSFs can frequently result in residuals that closely follow the structure of the signal. We include in the statistics the cases with 1 or 2 additional turning points because the signals should still be identifiable in the residuals. A joint fit of a foreground model plus a signal model in these instances should return an approximately correct parameterisation of the signal model.

### 2.5.1.2 Example Residuals

Also shown in Fig. 2.8, right panel, is an example of the fits for a given model and this is akin to Fig. 7 in [61]. We see that an MSF, top panel, while not identically recovering the signal but rather a smooth baseline subtracted version does preserve the three turning points of the model signal as expected. The example residuals from the unconstrained polynomial fits shown in the bottom right panel of Fig. 2.8 show a larger disparity with the structure of the model 21-cm signal than the MSF fit. The PSF with derivatives of order  $m \geq 3$  constrained, middle panel, produces residuals with one additional turning point. The behaviour at low frequency of the DCF fits is consistent across all the 264 tested models. It is a by-product of the basis choice, the frequency of data sampling and the steep nature of the foreground at low frequencies. We can alleviate some of these issues by increasing the sampling rate of our mock experiment and by reducing the bandwidth. However, the dominant cause is the basis function choice.

In logarithmic space, any non-smooth variations in the data and derivatives are amplified, as shown in section 2.4.2. Since the mock data set here is noiseless, the only non-smooth structure comes from the signal. However, the data is predominately smooth at high frequency and so the optimum fit tends to be an accurate representation of the foreground in this region and poorer at lower frequencies. The additional turning points when comparing the MSF and PSFs are also seen at low frequencies for the same reason. Despite the above, we maintain the full bandwidth, the same sampling rate and the logarithmic basis function in this analysis. We do

this because this basis function gives us the best fitting DCF models, and in a real experiment our knowledge of any present signal of interest will be too poor to constrain the bandwidth.

We have not analysed the full family of possible PSFs or unconstrained polynomials. Removing constraints on the lower order derivatives first will have the largest effect on the structure of the residuals, and consequently we have extensively analysed the effects of lifting the restrictions on the 2<sup>nd</sup> and 3<sup>rd</sup> order derivatives. We have, however, found evidence to suggest that MSFs and well constrained PSFs can recover signal structure to a higher degree of accuracy than unconstrained polynomials in the case of a smooth foreground. If the foreground features additional non-smooth structure, we may expect that an appropriately constrained PSF will act as an MSF. Determination of the appropriate constraints on a PSF will depend on the structure of the data, the expected structure of the signal in the data and the quality of the fit in terms of  $\chi^2$ .

### 2.5.1.3 Identifiable 21-cm Signals and Limitations of DCFs

The left panel of Fig. 2.8 is useful for 21-cm cosmology, in characterising the signals most likely to be detectable using DCFs. We find that MSFs and  $m \geq 3$  PSFs will best recover signals with approximately  $T_{\min} \geq -225$  mK and  $T_{\max} \geq 0$  mK. We would expect this to be true generally because these signal models have complex structures and feature the strongest deviations from the smooth foreground. For the coldest models with  $T_{\min} \leq -225$  mK and  $T_{\max} \leq 0$  mK, X-ray heating is negligible, and the spin temperature is always seen in absorption against the CMB. Consequently, they have the simplest structure, a weak deviation from the smooth foreground and are likely to be fitted out as part of the foreground modelling.

Smooth systematics, like smooth 21-cm signals, in the data will be removed or fitted out as part of the foreground. However, unless independent modelling of systematics is required, this can be considered an advantage. Modelling foregrounds with DCFs will help to identify non-smooth systematics in data sets where unconstrained polynomials have the potential to fit these out. This is particularly important in 21-cm cosmology, where these types of systematics need to be identified and instrumentation needs to be iteratively improved to increase the chances of a detection.

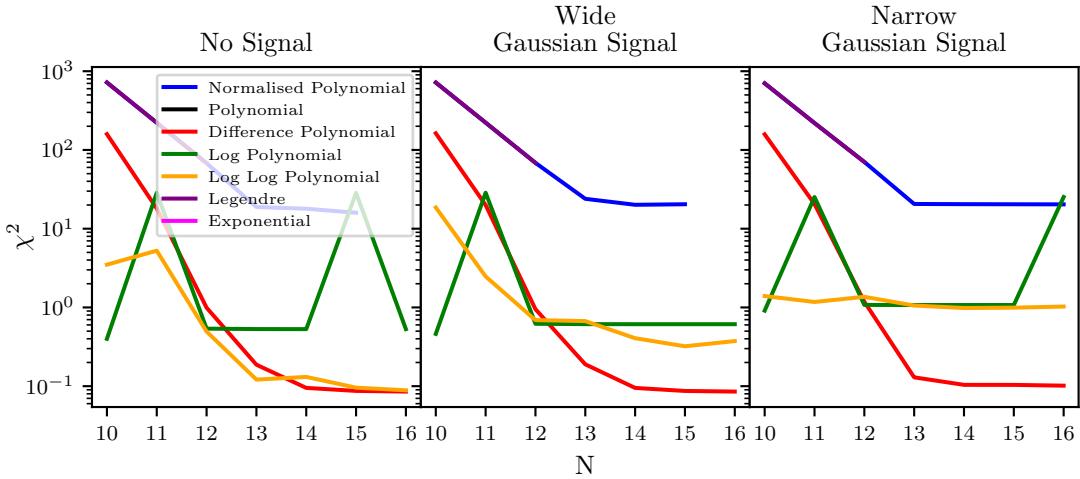
### 2.5.1.4 Smooth Signal Models

For 21-cm cosmology, it is also important to consider how the global 21-cm signal is modelled when performing joint fits. Typically, the signal is modelled as a Gaussian, flattened Gaussian or using physically motivated models. If a Gaussian model is jointly fit with a DCF foreground, then the fit is biased towards returning a ‘smooth’ Gaussian signal with a large width,  $\sigma$ , even if such a signal is not real. Similarly, incorrectly constrained DCFs can fit out ‘smooth’ signals in data sets. This can cause uncertainty in the presence of such signals, and the point is furthered in section 2.5.2.

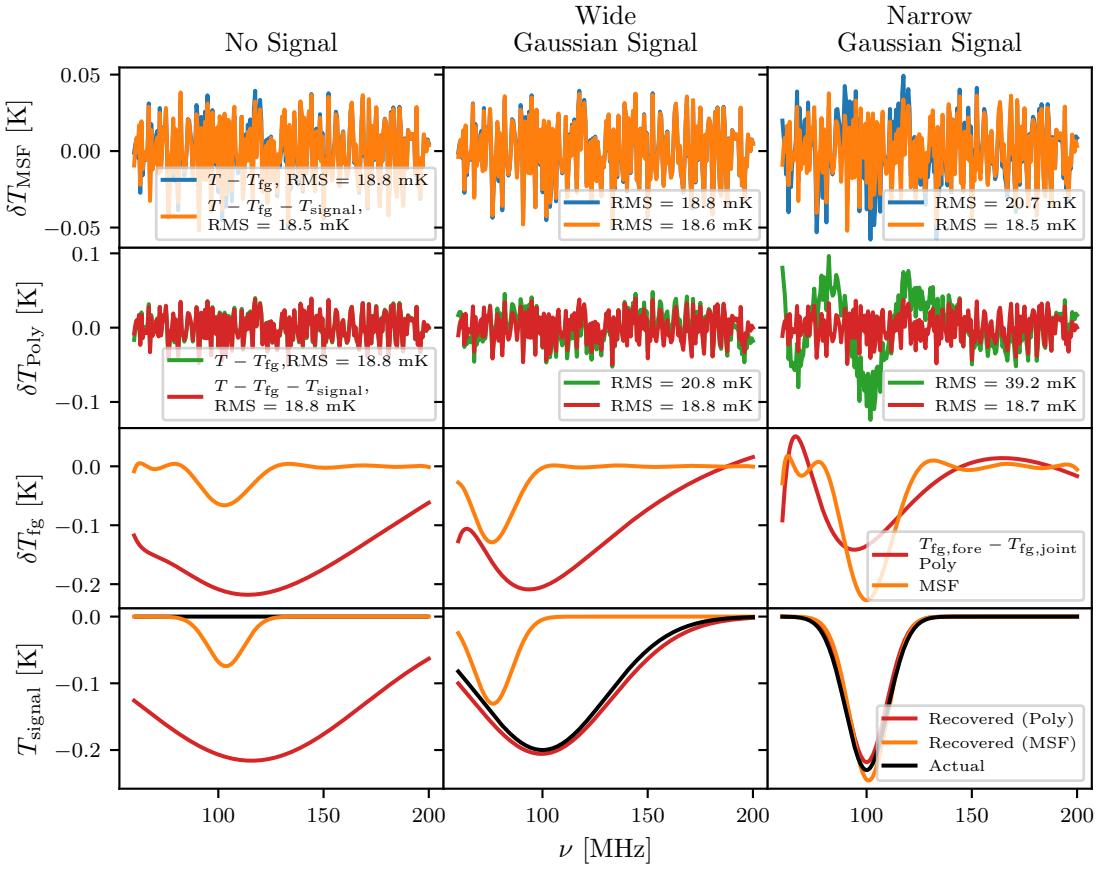
We can illustrate this by generating three different data sets, all with foregrounds following  $\nu^{-2.5}$  and the same Gaussian noise with a standard deviation of 20 mK. Into two of the three data sets, we add mock Gaussian signals with central frequencies  $\nu_c = 100$  MHz and generated using

$$T_{21} = -A \exp \left( -\frac{(\nu - \nu_c)^2}{2\sigma^2} \right), \quad (2.26)$$

where  $A$  is the amplitude. The first mock signal has an amplitude of 200 mK and a variance of 30 MHz representing a realistic wide or smooth Gaussian 21-cm signal. The second represents a narrow Gaussian signal



**Figure 2.9:** A comparison of the quality of MSF fit produced using the sign navigating algorithm and the different built-in MAXSMOOTH basis as a function of order  $N$  for three mock 21-cm experiment data sets with no signal, a wide Gaussian signal and a narrow Gaussian signal. The Difference Polynomial model, equation (2.4) of order  $N = 15$  is shown to be the optimal model for fitting these data sets. Beyond  $N = 15$  the quality of fit does not improve any further and additional terms in the model have coefficients  $a_{k \geq 14} \approx 0$ .



**Figure 2.10:** The top row shows the resultant residuals when using an MSF to fit just a foreground and to jointly fit a foreground and Gaussian signal model to three mock data sets with no signal, a wide Gaussian signal and a narrow Gaussian signal from left to right. The second row shows the equivalent for an unconstrained polynomial fit. The third row shows the change in foreground,  $\delta T_{\text{fg}}$ , between the pure foreground fit,  $T_{\text{fg,fore}}$ , and joint fit,  $T_{\text{fg,joint}}$ , for both foreground models and all three signal models. The bottom row shows the recovered signals for both foreground models alongside the true signal models in the data sets. The figure appears to show that unconstrained polynomials are better behaved than MSFs, however we would note that the signal models used here are simplistic in nature. We would expect the signals to be more complex, including emission, as with those used in Fig. 2.8 for which we showed the use of DCFs to be advantageous.

Model	$(\delta T_{\text{fg}} - T_{\text{sig}})_{\text{max}}$ [mK]	$\Delta RMS$ [mK]	$\log(Z_{\text{fore}})$	$\log(Z_{\text{joint}})$	$\Delta \log(Z)$
No Signal					
MSF	6.6	0.2	$606.42 \pm 0.03$	$606.85 \pm 0.03$	$0.08 \pm 0.04$
Poly	4.5	0.1	$604.28 \pm 0.03$	$603.38 \pm 0.03$	$0.51 \pm 0.04$
Wide Gaussian Signal					
MSF	6.0	0.2	$606.46 \pm 0.03$	$606.57 \pm 0.03$	$0.17 \pm 0.04$
Poly	17.5	2.6	$575.02 \pm 0.03$	$597.85 \pm 0.05$	$22.98 \pm 0.05$
Narrow Gaussian Signal					
MSF	18.4	1.7	$588.40 \pm 0.03$	$593.42 \pm 0.05$	$7.44 \pm 0.06$
Poly	81.4	20.3	$432.69 \pm 0.03$	$589.88 \pm 0.06$	$157.10 \pm 0.07$

**Table 2.4:** The table shows the average, across 10 repeats with different random noise, maximum difference between the recovered signals and the change in foreground for the polynomial and MSF pure foreground and joint fits,  $(\delta T_{\text{fg}} - T_{\text{sig}})_{\text{max}}$ , to the three simulated 21-cm experiment data sets with no signal, a wide signal and a narrow Gaussian signal. Also shown are the average changes in RMS,  $\Delta RMS$ , between the pure foreground and joint fits for both foreground models and all three data sets. We also provide weighted average log-evidences and values for the change in log-evidence,  $\Delta \log(Z)$ , between the joint and pure foreground fits with errors calculated from the error in each recovered  $\log(Z)$ . The comparatively large change in log-evidence for the narrow Gaussian signal data set when using MSFs provides confidence that the recovered signal is truly present in the data. Unconstrained polynomials appear to perform better than MSFs in this analysis. However, the signals used here are simplistic, and we have shown in Fig. 2.8 that DCFs behave far better for complex theoretically motivated signals that include emission.

with an amplitude of 230 mK and a variance of 10 MHz. The final data set has no additional signal in the band of 60 – 200 MHz.

We assess the best basis from the `MAXSMOOTH` library for fitting these data sets, and this is shown in Fig. 2.9. The graph shows the minimum  $\chi^2$  values when fitting MSFs of a particular form, detailed in the legend, using the `MAXSMOOTH` sign navigating algorithm with a given order  $N$  to the three data sets. We find that for all three data sets the best basis choice is the Difference polynomial, equation (2.4), with order  $N = 15$ . We consequently proceed to fit each data set with this MSF model. We compare the resultant residuals for each data set to those from a joint fit of an MSF, of the same functional form, and a Gaussian 21-cm signal model. We perform the joint fits by using `MAXSMOOTH` with the Python implementation of the Nested Sampling software `MULTINEST` [94–97]. Here, `MULTINEST` estimates the Gaussian signal parameters, `MAXSMOOTH` fits the foreground model to the data minus the estimated Gaussian signal at each iteration and `MULTINEST` minimises the data minus the foreground model plus the signal model. We also perform the equivalent fits with a 5<sup>th</sup> order unconstrained polynomial, given by equation (2.6) using a Lavenberg-Marquardt [147, 148] algorithm in place of `MAXSMOOTH`. We perform this analysis ten times, generating a new noise distribution each time. An example result is shown in Fig. 2.10 where we have provided the same theoretically motivated priors on all the Gaussian 21-cm models. Respectively from top to bottom the rows in Fig. 2.10 show the residuals after just an MSF fit and after a joint fit for comparison, the equivalent for the unconstrained polynomial fits, the change in foreground between the foreground fit and the joint fit and the recovered signal in comparison to the actual signal. The columns correspond to the case of no signal, a wide Gaussian signal and a narrow Gaussian signal from left to right.

We can see that in the absence of a signal jointly fitting with a Gaussian model and MSF returns an absorption trough. This recovered absorption trough is approximately the same as the change in the foreground model when fitting with just an MSF and the change in the RMS is small. This is illustrated in Tab. 2.4 which shows the average results across the ten repeats of this analysis. If this were a real experiment, the fitted model could

easily be misinterpreted as a real signal that had been fitted out as part of the foreground model. However, there is a very small change in log-evidence between the pure foreground and joint fit, which would suggest that the signal is not truly present in the data.

In Tab. 2.4 the values for the log-evidence for each fit and the change in log-evidence,  $\Delta \log(Z)$ , between the foreground and joint fits for each data set and each foreground model are reported. While MULTINEST returns a log-evidence for the joint fits, MAXSMOOTH is not a Bayesian algorithm. However, assuming a Gaussian likelihood, we can use MULTINEST to calculate the evidence for our pure foreground fits by having it estimate the noise and calculate the likelihood.

When jointly fitting in the presence of a wide Gaussian signal with an MSF we see a similar result to the case with no additional signal, and we could not confidently say that the signal is present in the data if we had no prior knowledge. In fact, the joint fit has recovered a poor representation of the Gaussian signal because the smooth signal in both instances, pure foreground fit and joint, has been absorbed in the foreground modelling.

Finally, we see that in the case of a narrow Gaussian signal in the data set with an MSF we get an almost exact recovery after a joint fit. There is a larger discrepancy between the difference in foreground models and the recovered signal, as shown in Tab. 2.4, and the reduction in RMS is more significant, giving us confidence that the signal is truly present. Importantly, the change in log-evidence is much larger than in the other two cases, indicating the presence of the signal.

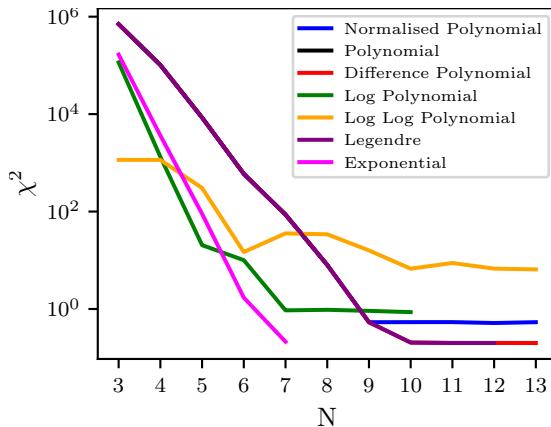
For the unconstrained polynomial, in the case where there is no signal in the data, we appear to recover a very smooth Gaussian signal. However, again the change in log-evidence between the pure foreground fit and the joint fit tells us that neither scenario is more likely, and consequently we would conclude that the signal is not present. For the other two cases, the unconstrained polynomial behaves well. However, we note that the signals induced here are simplistic representations of a global 21-cm signal. We would expect the signal to have a much more complicated structure including emission at high frequencies and similar to the signals used in Fig. 2.8 for which we showed DCFs to be advantageous.

When attempting to identify the wide Gaussian signal using DCFs, we also note it will be advantageous to fit using a CSF with derivatives of order  $m \geq 1$  constrained. For 21-cm cosmology, we find that CSFs and MSFs are generally equivalent. However, in this case the only significant non-smooth structure in the signal, aside from the inflection point, is the turning point and where an MSF may fit this out a CSF will not.

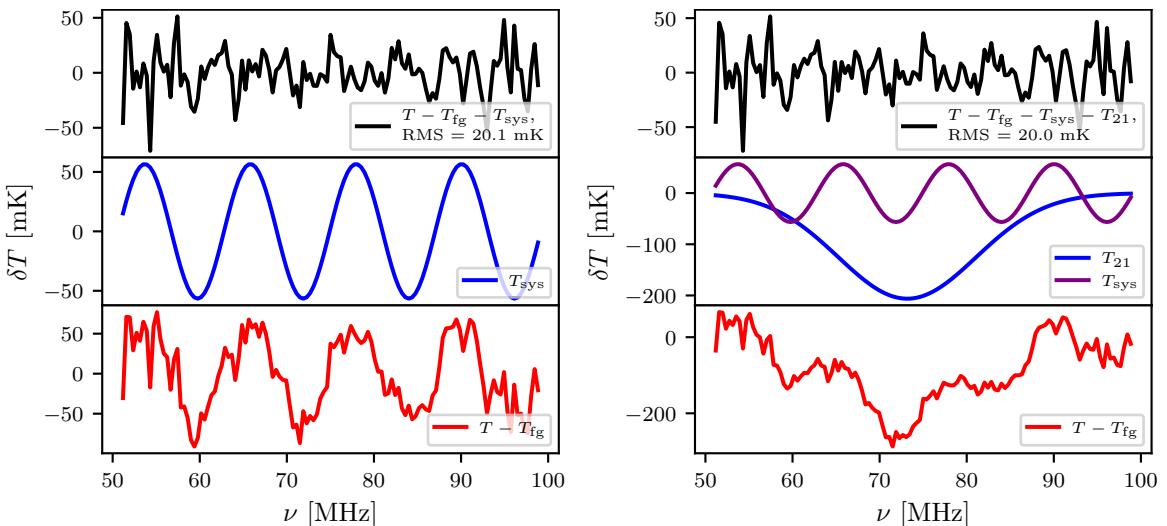
The problem of identifying and misidentifying smooth signals becomes even more prominent, particularly in the absence of any signal in the data, when modelling the foreground with a DCF in  $T - \log(\nu)$  space. Here the sample rate is non-uniform and higher at higher  $\log(\nu)$ . This makes the problem harder to fit in this region and also ‘smooths’ any signals present at low  $\log(\nu)$ . Together, these effects can make it difficult to detect signals that can already be considered ‘smooth’ in linear space. Similarly, in the absence of any signal in the data, when jointly fitting with a Gaussian signal model and large prior ranges, the routine estimating the Gaussian parameters will tend to favour ‘smooth’ signals at high frequencies in linear space because of the non-uniformity of sampling in logarithmic space.

### 2.5.2 MSFs and the EDGES Data

In 2018 the EDGES team reported the detection of an absorption trough at 78 MHz which could be interpreted as a global 21-cm signal [8]. The reported signal is  $\approx 2$  times the maximum magnitude predicted by current



**Figure 2.11:** The resultant  $\chi^2$  as a function of MSF order,  $N$ , for the `MAXSMOOTH` built-in basis functions fitted to the EDGES data using the `MAXSMOOTH` sign navigating algorithm. The Legendre, Difference polynomial, the Polynomial and Normalised Polynomial models lie on top of each other in this figure. The occasional increase in  $\chi^2$  with  $N$  for the logarithmic model is because this basis is increasingly unstable with higher  $N$  and requires all sign combinations to be tested.



**Figure 2.12: Left:** The residuals with an RMS of 20.1 mK and log-evidence of  $302.99 \pm 0.08$ , black, found after subtracting from the EDGES data a jointly fit MSF and a sinusoidal systematic. We recover a parameterisation of the systematic, shown in blue, that is consistent with previous work as discussed in the text. The red line, bottom panel, illustrates the residuals after just removing the fitted foreground from the data. **Right:** A joint fit of a Gaussian 21-cm model signal with a sinusoidal systematic and an MSF foreground to the EDGES data. The recovered 21-cm signal is smooth and the addition of the signal model has caused a large change in the recovered foreground,  $T_{fg}$ , illustrated by the decreased amplitude in the bottom panel around 70 MHz when compared to that in the left figure. The RMS, 20.0 mK, is also similar in magnitude to that found for the fit in the left panel and the log-evidence has reduced to  $292.67 \pm 0.17$ . Consequently, it is unlikely the model signal is real, and it is probable that it is produced artificially by the change in foreground.

cosmological models [13], and, in order to explain the signal as a 21-cm signal, interactions between dark matter and baryons [122, 123, 125, 127, 127] or a higher radio background [8, 117–120, 131] are needed.

While a higher radio background has been suggested by the results of the ARCADE2 experiment [48] and potentially confirmed by measurements from LWA [49] there are concerns about the analysis of the EDGES data [128–130]. These studies and the following work presented here use the publicly available integrated spectrum from the EDGES Low Band experiment, which can be found at: <https://loco.lab.asu.edu/edges/edges-data-release/>.

In [128] the authors found that recovering the absorption profile using the ‘physically motivated’ foreground

model in [8] produces unphysical negative values for the ionospheric electron temperature and optical depth. This suggests that the treatment of the foreground in [8] absorbs part of an unknown systematic. It was also found that a large change in foreground was needed when just fitting a ‘physical’ foreground to the data and when jointly fitting with a flattened Gaussian 21-cm signal profile. In [128] the authors identify the potential presence of a sinusoidal function in the EDGES data with an amplitude of  $\approx 60$  mK and a period of  $\approx 12.5$  MHz.

The authors of [130] fitted a range of models to the EDGES data, varying the 21-cm models between a Gaussian model, a flattened Gaussian model as used by [8] and physical simulations from the ARES code [149]. They vary the unconstrained polynomial order for the foreground model and examine likelihoods with and without an additional noise term and a damped sinusoidal function. They use Bayesian Evidence to quantify the most likely scenarios of an atlas of 128 models. The 21 highest evidence models all feature damped sinusoidal functions, all with a consistent amplitude of  $\approx 60$  mK and a period of  $\approx 12.5$  MHz.

An MSF fit to the foreground should leave a periodic sinusoidal function behind in the residuals if it is present because it is non-smooth in nature. This has previously been shown to be the case by [129], hereafter S19, who identified a sinusoidal feature with an amplitude of  $60 \pm 10$  mK and a period of  $12.3 \pm 0.1$  MHz. We attempt here to re-create this analysis to validate the `MAXSMOOTH` algorithm. The fitting routine used and choice of basis function are the only differences between the results presented here and in S19. The use of `MAXSMOOTH` means that our joint fit of the data and a systematic model will be computationally quicker and more reliable than the Nelder-Mead based approach to fitting taken in S19, as demonstrated in section 2.4.

We begin first by assessing the quality of fits using the various basis functions built-in to `MAXSMOOTH`. S19 used a basis function constrained in  $T - \log_{10}(\nu)$  space and although the functional form is not explicitly stated in S19 it is derived from the models in [61] and so is similar to, if not identical to, equation (2.5). Fig. 2.11 shows the resultant  $\chi^2$  as a function of MSF order for fits with varying basis functions using `MAXSMOOTH`. Our  $T - \log_{10}(\nu)$  space model cannot achieve the same RMS as that found by S19 with a similar model. With an  $N = 7$  MSF constrained in logarithmic frequency space, S19 return an RMS of 44 mK, whereas with equation (2.5) `MAXSMOOTH` returns an RMS of 87 mK. We believe this is due to the lack of normalisation in `MAXSMOOTH` and as previously discussed this is an ongoing area of development.

We use Fig. 2.11 to inform our choice of basis function and MSF order, and proceed using an 11<sup>th</sup> order MSF of the form given by equation (2.4). We find residuals with an RMS of  $\approx 40.4$  mK and a log-evidence of  $216.80 \pm 0.09$  as shown in Fig. 2.1. We note that this is in approximate agreement to results shown in S19. We also find troughs at  $\approx 70$  MHz and  $\approx 85$  MHz which correspond to those found in all the reported sinusoidal functions.

We jointly fit the data with an 11<sup>th</sup> order MSF and a sinusoidal function of the form

$$T_{\text{sys}} = p_0 \sin(p_1 \nu - p_2), \quad (2.27)$$

and the resultant residuals are shown in the left panel of Fig. 2.12. We have not included a model 21-cm signal in this fit. We use `MAXSMOOTH` along with a Lavenberg-Marquardt algorithm implemented with `SCIPY` to perform this joint fit and with initial parameters of  $p_0 = 60$  mK,  $p_1 \approx (2\pi)/12.5$  MHz<sup>-1</sup> and  $p_2 = 0$  rad. We find that the results change with the initial parameters when using the Lavenberg-Marquardt algorithm, however, the chosen initial parameters are well-informed by the previous work outlined above. We return parameters of  $p_0 \approx 56.6$  mK,  $p_1 \approx 0.52$  MHz<sup>-1</sup> or a period of  $\approx 12.1$  MHz and  $p_2 \approx 1.1$  rad in close agreement with previous analysis. We use `MULTINEST` to approximate the evidence for this fit, assuming a Gaussian likelihood and return

$\log(Z) = 302.99 \pm 0.08$ . This is a significant increase in log-evidence when compared to the pure foreground fit and would suggest strongly that the systematic is present in the data.

We find an RMS value of 20.1 mK, in close agreement with the result of 22.9 mK found in S19 when jointly fitting an MSF and the sinusoidal function. However, the RMS of the joint fit in the left panel of Fig. 2.12 is equivalent to the RMS found by S19 when jointly fitting an MSF foreground, a sinusoidal systematic and Gaussian 21-cm signal model. Their proposed Gaussian 21-cm signal model fits with standard predictions. However, the RMS of our joint fit without a Gaussian signal model may highlight some of the difficulties in detecting ‘smooth’ Gaussian signals discussed in section 2.5.1.4.

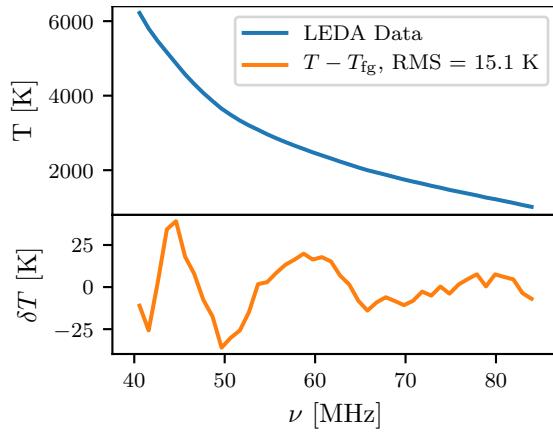
We perform a joint fit of a Gaussian 21-cm signal, a sinusoidal systematic and MSF. Due to the increased complexity of the fit and uncertainty in the fit parameters for the Gaussian model, we perform our fit using MAXSMOOTH and MULTINEST. We provide prior ranges of  $30 - 80$  mK for the amplitude,  $0 - 25$  MHz for the period and  $0 - 2\pi$  for the phase shift of the sinusoidal systematic. For our Gaussian we set realistic priors on the amplitude of  $0 - 250$  mK, on the central frequency of  $50 - 100$  MHz and on  $\sigma$  of  $0 - 20$  MHz. For the sinusoid we find an amplitude of 56.5 mK, a period of approximately 12.1 MHz and a phase of 1.1 rad. We return an amplitude of 206 mK for the Gaussian with a central frequency of 73 MHz and a FWHM  $\approx 18$  MHz. MULTINEST returns a noise parameter of 20 mK and our resultant fit, shown in the right panel of Fig. 2.12, has an RMS of approximately 20.0 mK. This is much wider and deeper than the signal reported in [129] which has a depth of  $133 \pm 60$  mK and FWHM of  $9 \pm 3$  MHz however we return the same central frequency of 73 MHz.

Noting the discussion of plausibly detectable 21-cm signals when using DCFs and the bias towards ‘detection’ of ‘smooth’ Gaussian signals in section 2.5.1.4 we assess the feasibility of the returned model signal. Here, the notion of ‘smoothness’ is relative to the bandwidth. In section 2.5.1.4 a Gaussian with FWHM  $\approx 24$  MHz is confidently identifiable as a real signal, but the bandwidth is much larger than that for the EDGES data. We can see by comparison of the bottom panels of Fig. 2.12, showing the data minus the foreground from the joint fits, that there is a large change in foreground when we include the Gaussian model as part of our joint fit. This may be because in our initial fit, without the Gaussian model, the foreground model was fitting out the smooth signal. Alternatively, the signal may not be present, and the fitting routine has returned a smooth signal by extracting it from the foreground component of the data. The reduction in RMS when the joint fit includes a Gaussian is negligible, and it is consequently challenging to determine whether the signal is present in the data. However, we can conclude from the log-evidence, which for this fit has a value of  $292.67 \pm 0.17$  and is smaller than that for the joint fit of the systematic and foreground, that the Gaussian is not likely to be real.

### 2.5.3 MSFs and the LEDA Data

LEDA, like EDGES, is a global 21-cm experiment analysing the band  $30 - 88$  MHz and aiming to detect the anticipated absorption feature [150]. The design and calibration approach of LEDA is detailed in [9]. In contrast to EDGES, the LEDA experiment comprises 5 dual-polarization radiometer antennas that are part of a larger 256-antenna interferometric array. This approach is intended to allow inter-antenna comparison and in-situ measurement of the antenna gain response. Similar to other experiments with absolute calibration, LEDA uses two noise diode references to calibrate the measured antenna temperature into units of Kelvin. Corrections are then applied to account for the impedance of the antenna and receiver, derived from vector network analyser (VNA) measurements.

As shown in [9], data are seen to vary between antennas, which are not perfectly identical, and this is



**Figure 2.13:** The LEDA data from antenna 252A taken in 2016 and averaged over 1 hour of LST shown in the top panel. Also shown, bottom panel, are the residuals after fitting the LEDA data with a 9<sup>th</sup> order MSF of the form given in equation (2.4) with a log-evidence of  $-185.45 \pm 0.09$ . We see evidence of a damped sinusoidal systematic in the data set.

attributed to minor differences in terrain, analogue component response, and physical construction. While calibrated spectra are presented, it is suggested that there are unidentified systematics in the data; work has been undertaken to better characterize and update the LEDA system with iterative improvements.

Here, we fit MSFs to data from the LEDA 2016 campaign [9]. This is the first time MSFs have been applied to LEDA data. Specifically, we fit data from antenna 252A, taken on January 26<sup>th</sup> 2016 in the Local Sidereal Time (LST) range 11:00–12:00. In this LST range, the Galactic contribution to the antenna temperature is at a minimum. The data are binned into 1.008 MHz channels, spanning 40–85 MHz.

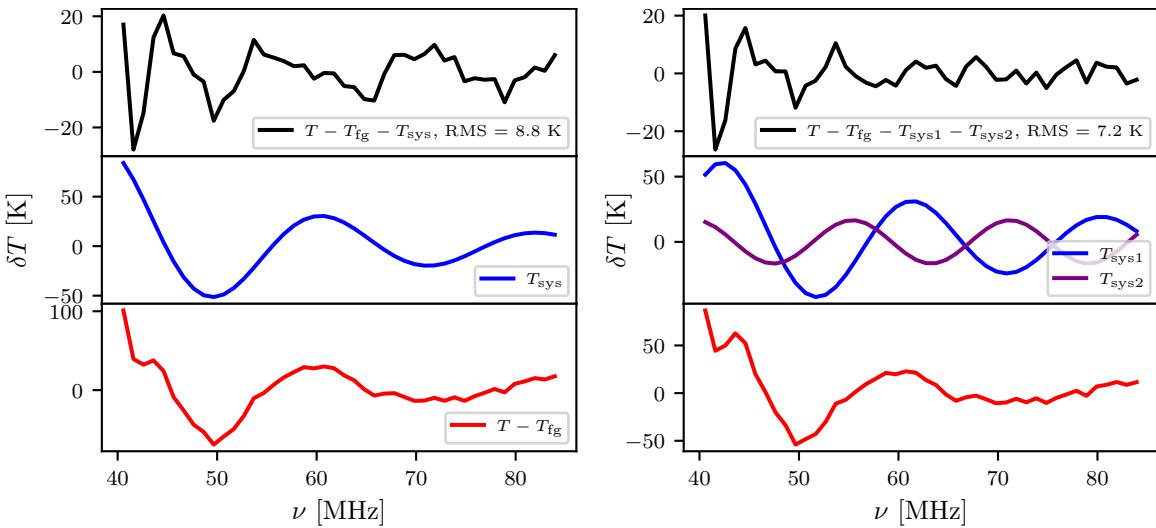
We fit an MSF of the form given in equation (2.4) to the data, as we find that this basis function returns the best fit consistently for  $N \geq 8$ . Shown in Fig. 2.13 are the resultant residuals from an  $N = 9$  MSF fit with an RMS of  $\approx 15$  K and a log-evidence of  $-185.45 \pm 0.09$ . The resultant residuals are large and would obscure a cosmological 21-cm signal. We note that as per equation (4) in [9] the radiometer noise is expected to be  $\approx 0.5$  K.

The residuals from the MSF fit appear to include a damped sinusoidal systematic structure. We proceed to fit such a systematic model given by

$$T_{\text{sys}} = \left( \frac{\nu}{\nu_0} \right)^{-p_0} p_1 \sin(p_2 \nu - p_3), \quad (2.28)$$

along with a 9<sup>th</sup> order MSF by using `MAXSMOOTH` and `MULTINEST`.  $\nu_0$  is chosen to be the central frequency of the band. We provide a prior on the power of 0 – 3 for weak damping. Prior ranges of 25 – 75 K for the amplitude of the sinusoidal function, 0 – 1 MHz<sup>-1</sup> for the period,  $P$ , which is fitted as  $p_2 = (2\pi)/P$ , 0 – 2π for the phase shift and a log uniform prior on the noise of 10<sup>-2</sup> – 10<sup>1</sup> K are also provided. The results of this fit are shown in the left panel of Fig. 2.14. We return optimal parameters of an exponent of  $\approx 2.7$ , an amplitude of  $\approx 27.9$  K, a period of 21.7 MHz and a phase shift of 3.7 rad. The residuals have an RMS of  $\approx 8.8$  K and `MULTINEST` returns a noise parameter of  $\approx 7.7$  K. We also see an increase in log-evidence, which has a value of  $-175.50 \pm 0.19$ , when compared to the pure foreground fit suggesting the systematic is present in the data.

[9] suggest, from analysis of the 2016 LEDA data, that the systematic in the data is caused by the direction-dependent gain of the antenna. A frequency dependent group delay may also be caused by a bandpass filter that could contribute unaccounted for reflections. The pattern of oscillations that form the systematic have also been found to change after rainfall. This systematic may then be caused by moisture in the surrounding soil or by changes in the electric length of the dipoles caused by moisture on the dipole itself. We also highlight the



**Figure 2.14: Left:** The residuals, black, after jointly fitting the LEDA data with an MSF and damped sinusoidal systematic with a log-evidence of  $-175.50 \pm 0.19$ . The centre panel shows the recovered systematic model, blue, and the bottom panel shows the residuals, red, after just subtracting the fitted foreground model from the data. The addition of the systematic model has reduced the RMS of the fit when compared to Fig. 2.13. **Right:** The resultant residuals, black, with a log-evidence of  $-168 \pm 0.19$  found when fitting the LEDA data with an MSF foreground, damped sinusoidal, blue, and additional sinusoidal systematic, purple. Again the bottom panel shows the residuals, red, after just subtracting the fitted foreground model from the data. The further reduction in RMS and increase in log-evidence suggests that both these systematics are present in the data, and indicates that the larger systematics in the LEDA data may be represented by the leading order terms in a damped Fourier series.

similarities in structure of the systematic with that in the EDGES data. Both have sinusoidal structures, and so similarities between the experimental setups and calibration processes may hint at larger causes of systematics across 21-cm cosmology experiments. The systematic is not likely to be associated with the sky because of the difference in periodicity and amplitude found by both experiments. EDGES does not have a bandpass filter, however it could still be affected by moisture in the surrounding environment. Although we note that this experiment is in a typically dry location (Murchison Radio-astronomy Observatory, Australia).

The residuals shown in the top left panel of Fig. 2.14 show a further sinusoidal structure after removal of the leading order damped sinusoidal systematic. We therefore attempt a joint fit to the data using an MSF foreground, a damped sinusoid and an additional sinusoid described by equation (2.27). We maintain the same priors on the original damped sinusoidal function and provide a prior of  $10 - 30$  K on the amplitude,  $0 - 1$  MHz $^{-1}$  on the period and  $0 - 2\pi$  on the phase shift of the additional systematic.

We find best fit parameters for the leading damped sinusoid of  $\approx 1.8$  for the exponent,  $\approx 30$  K for the amplitude, a period of  $\approx 19$  MHz and a phase shift of  $\approx 6.2$  rad. For the additional sinusoidal systematic we find an amplitude of  $\approx 17$  K, a period of  $\approx 16$  MHz and a phase shift of  $\approx 1.5$  rad. MULTINEST returns a noise of  $\approx 7.2$  K and the fit shown in the right panel of Fig. 2.14 has an RMS of  $\approx 7.2$  K. We find a log-evidence for this fit of  $-168.34 \pm 0.19$ .

Distinctions between the two systematics have been made in the middle right panel of Fig. 2.14 for clarity. The RMS of the residuals after removal of these two systematics is still significantly larger than the expected radiometer noise for this experiment,  $\approx 0.5$  K. However, the decrease in the RMS when these systematics are included in the fit and increase in log-evidence would suggest that both are present in the data. A further addition of sinusoidal systematics will inevitably reduce the RMS of the residuals, in the same way that the residuals after foreground removal could accurately be described by a Fourier series. Higher order terms in the series would feature smaller periods until the periodicity of the terms matched that of the noise. However,

the systematics present in the data may have a form described by the leading order terms in a damped Fourier series representing a series of reflections, for example. We leave more rigorous investigation of the additional oscillatory structure in the residuals, top right panel of Fig. 2.14, to future work.

## 2.6 Conclusions

Derivative Constrained Functions (DCFs) generally are advantageous for experiments in which the desired signal is masked by higher magnitude smooth signals or foregrounds which follow power laws. DCFs are designed to accurately replicate these power law structures by constraining individual high order derivatives of fitted polynomials. They are particularly useful when the signals of interest are expected to be several orders of magnitude smaller than the foregrounds, similar in magnitude to experimental noise and non-smooth.

We have introduced `MAXSMOOTH` as a fast and robust tool for fitting DCFs and demonstrated its abilities with examples in 21-cm cosmology. The code features an extensible library of example DCF models. Further work into the normalisation of these DCF models for `MAXSMOOTH` is anticipated with the aim to improve the quality of fitting and efficiency of the software.

`MAXSMOOTH` is shown to be  $\approx 2$  orders of magnitude faster than traditionally used Basinhopping/Nelder-Mead based algorithms [61]. This is an important improvement when jointly fitting signals, systematics and foregrounds using a Bayesian likelihood loop, as in Nested Sampling [57].

`MAXSMOOTH` is also designed to be able to cover the entire available parameter space, unlike a penalised Basinhopping/ Nelder-Mead based routine, by dividing it into discrete parameter spaces based on the different allowed combinations of signs, positive and negative, on the constrained derivatives. The extensive exploration of the parameter space provides confidence in the results and the employment of quadratic programming, a robust method for solving constrained optimisation problems, allows `MAXSMOOTH` to remain an efficient algorithm.

We have reproduced analysis of the EDGES data using `MAXSMOOTH` and analysed data from the LEDA experiment with MSFs for the first time. We have highlighted limitations of DCFs when jointly fitting for 21-cm signals and illustrated this using the EDGES data.

We show that MSFs preserve turning points of 21-cm signals more consistently than commonly used low order logarithmic unconstrained polynomial models. This is particularly true of 21-cm signals that exhibit significant non-smooth structure in the frequency band of our experiments.

We use PSFs and `MAXSMOOTH` in chapter 5 to model the foreground structure in data from the SARAS2 instrument.



CHAPTER

3

**GLOBALEMU**



### 3.1 Introduction

To recap chapter 1, the relative magnitude of the features of the 21-cm signal is determined by various astrophysical processes including; the Wouthuysen-Field effect [39, 40], Lyman- $\alpha$  heating and CMB heating [14, 42–45], X-ray heating and ionization of the hydrogen gas by UV emission [29]. The physical processes themselves and hence the global signal can be characterised by a set of astrophysical parameters (see section 3.3 and [15]): the star formation efficiency,  $f_*$ , the minimal virial circular velocity,  $V_c$ , the X-ray efficiency,  $f_X$ , the CMB optical depth,  $\tau$ , the slope of the X-ray spectral energy density (SED),  $\alpha$ , the low energy cut-off of the X-ray SED,  $\nu_{\min}$ , and the mean free path of ionizing photons,  $R_{\text{mfp}}$ .

Hybrid approaches are used to calculate realizations of the 21-cm signal, over a large cosmological volume and redshift range, which then can be averaged at every redshift separately to give the global signal [e.g. 11–14, 151]. Each simulation takes several hours to perform on a desktop [91] and though this is much faster than hydrodynamical simulations, this time is too long to allow us to constrain astrophysical parameters using data. Therefore, the desire to emulate the global signal with neural networks, trained on the results of the large scale simulations, has arisen. The neural networks can produce a realisation of the global signal in a fraction of a second by interpolating between the simulated cosmological and astrophysical models. This means they can be used to physically model the signal in, for example, Bayesian Nested Sampling loops<sup>a</sup>, as in the REACH data analysis pipeline [57], in which millions of calculations need to be made to infer cosmological parameters [153–155, Sims et al. 2021 (in prep.)].

A number of papers have considered emulation of the 21-cm power spectrum using convolutional neural networks and other techniques [e.g. 105, 156, 157]. At the time of writing 21CMGEM is the only publicly available emulator used to accurately, with a maximum normalised Root Mean Square Error (RMSE) of 10.55 %, and quickly (see section 3.6.1) emulate the global 21-cm signal [15] <sup>b</sup>. It has previously been used to provide constraints on the parameter space of the 21-cm signal using EDGES high-band data [91]. The emulator uses Principal Component Analysis [PCA, 158], the seven astrophysical parameters detailed above and in section 3.3, additional information about the mean collapsed fraction of halos as a function of redshift,  $f_{\text{coll}}(z)$ , the fraction of X-ray energy above 1 keV,  $f_{\text{XR}>1\text{keV}}$ , and 2 keV,  $f_{\text{XR}>2\text{keV}}$ , and relies on a division of the signal into 2 or 3 distinct segments defined by the turning points. It involves the application of a decision tree for classification and several regression neural networks estimating PCA components, the frequencies and temperatures of turning points as well as additional parameters such as the frequency at which the neutral fraction equals 0.16,  $\nu(x_{HI} = 0.16)$ .

In this chapter we present GLOBALEMU which uses a novel and robust approach with a single small scale neural network to emulate the global 21-cm signal given a comprehensive set of astrophysical parameters and redshift range. Where previously 21CMGEM was designed to take in astrophysical parameters and return a low dimensional representation of the global signal as a function of redshift, GLOBALEMU takes in the same astrophysical parameters and redshift and returns a signal temperature at the corresponding redshift (see Fig. 3.1). This greatly simplifies the complexity of the relationship being learnt by the neural network. It means we can achieve accurate results, with the smoothness of the signal imposed by the interpolation of the neural network between signals in the training data set, using a small network and without the need for a compressed representation of the signals like PCA where there is a potential loss of information. Additionally,

<sup>a</sup>Note that here we are not referring to a Bayesian Neural Network [see 152] but rather parameter optimisation algorithms such as POLYCHORD [see 88, 89].

<sup>b</sup>During review of this work for Monthly Notices of the Royal Astronomical Society a pre-print (now a published paper) describing the global signal emulator 21cmVAE was made public [98] in which the comparative performance of GLOBALEMU is discussed.

GLOBALEMU relies on a physically motivated pre-processing of the training data and can emulate a high resolution,  $\delta z = 0.1$  over the range  $z = 5 - 50$ , global 21-cm signal in  $\approx 1.3$  ms. GLOBALEMU will be used extensively by the REACH collaboration and has been designed to have an average accuracy less than or equal to 10 % of the expected noise in the REACH system, estimated to be  $\approx 25$  mK [10].

GLOBALEMU is written in Python using tensorflow and the KERAS backend, it is pip installable via `pip install globalemu` and available at <https://github.com/htjb/globalemu>. It is flexible enough to be retrained on any set of global 21-cm signal models, whilst maintaining the novel design and physically motivated pre-processing. We provide a demonstration of its accuracy and efficiency in this chapter using the same data used to train 21cmGEM and the corresponding trained models are publicly available on GitHub. We use GitHub actions to perform continuous integration.

In section 3.2 we describe the novel approach used to parameterise GLOBALEMU. Section 3.3 describes the training and test data used to illustrate the capabilities of GLOBALEMU in this chapter and the astrophysical parameters in the simulations of the global signal. We then describe the predominantly physically motivated pre-processing of the inputs and outputs of the neural network in section 3.4. A discussion of the neural network structure follows in section 3.5 and the quality of emulation is assessed in section 3.6. We conclude in section 3.7.

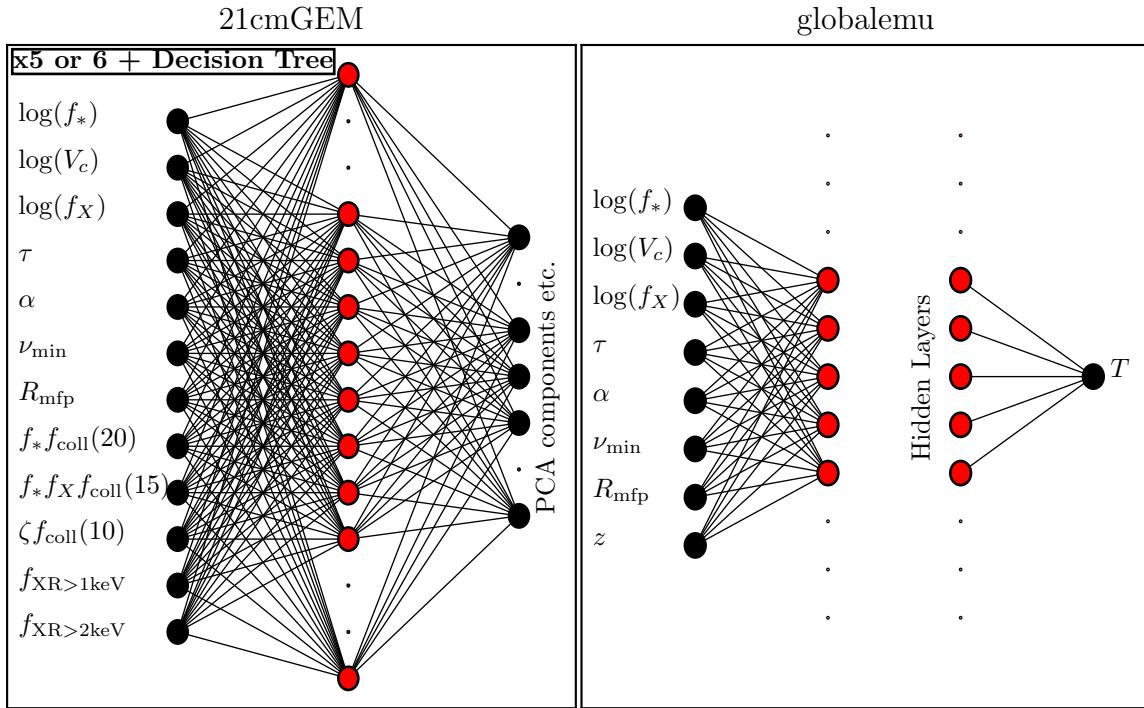
The co-authors, namely Will Handley, Anastasia Fialkov, Eloy de Lera Acedo and Kamran Javid, of [4] provided comments and feedback on the original manuscript that this chapter is modified from.

## 3.2 Parameterising the Problem

There are several approaches that can be used to emulate the global 21-cm signal with a neural network. The ultimate goal of the emulator is to take in a set of astrophysical parameters and return an estimate of the signal brightness temperature as a function of redshift,  $T_{21}(z)$ , where the relationship has been learned from detailed semi-numerical simulations. This can be done directly with a neural network that returns a value of  $T_{21}$  for each redshift data point it has been trained on. However, assuming the network is trained on high resolution signals this would result in a large number of outputs, making it hard to train, and would be limited in predictive power to specific values of redshift. The process can also be achieved by estimating, via a neural network, coefficients of a compressed representation of the signal space. For example, using PCA as with 21cmGEM [15] or learning coefficients of basis functions for polynomials or wavelets that when combined return the global signal. However, whilst this approach reduces the number of outputs compared to a direct emulation, if incorrectly designed this can result in information loss and is equally limited in predictive power.

We take the novel approach of using redshift as an input to the network alongside the astrophysical parameters and returning from the network a single corresponding temperature. This is beneficial for two reasons; the small number of inputs and outputs means that the network can retain a simple structure and second the network will be able to interpolate between the values of redshift that it has been trained on. The smooth structure of the output from the network is guaranteed by the smooth interpolation performed by the neural network and by the smooth structure of the signals it is learning. Vectorised calls to the network are used to emulate the temperature as a function of redshift.

In GLOBALEMU we also provide the ability to emulate the evolution of the neutral fraction,  $x_{HI}$ , of hydrogen as a function of redshift. We use an identical framework as when emulating the global signal to do this with a



**Figure 3.1: Left Panel:** We provide here an illustration of the regression neural networks used in 21cmGEM. Note that 21cmGEM uses either 5 or 6 of these networks and a decision tree when making predictions. For one of the regression networks used, the only input parameters are the seven astrophysical parameters detailed in section 3.3 (excluding redshift). However, for the others there are an additional five derived parameters (see text and [15]) and we illustrate the full set of 12 inputs. The number of output nodes depends on the specific application of the network, and they can correspond to either PCA components (4 nodes), additional parameters such as  $\nu(x_{HI} = 0.16)$  (1 node) and the frequencies and brightness temperatures of turning points in the signal (7 or 5 nodes). The hidden layer in all the 21cmGEM regression networks have 40 nodes. For a full illustration of the 21cmGEM algorithm and detailed description, see Fig. 11 and section 4.2 in [15]. **Right Panel:** An illustration of the GLOBALEMU neural network. Note the use of only one network to emulate the global signal, in comparison to the 5 or 6 used for 21cmGEM. Here the input layer has eight nodes (seven astrophysical parameters plus redshift) and the output layer is a single node returning the brightness temperature corresponding to the input redshift. We show a sizable hidden layer structure here with the red nodes and ‘...’ however, we note that the reduced number of inputs and outputs implies that a small architecture will be sufficient to achieve a high level of accuracy in the emulation (see section 3.5).

set of astrophysical parameters and a redshift as inputs to a neural network and a corresponding value of  $x_{HI}$  as an output.

We have built a network that can emulate the global 21-cm signal to a high degree of accuracy without the need for L1 and L2 regularisation, dropout [159], batch normalisation [160] or other similar concepts. We have achieved this by focusing on the pre-processing of the network’s inputs and outputs, with the goal to make the problem simple to solve with a basic neural network of a ‘reasonable’ size.

### 3.3 The Training and Test Data

In this chapter, we use the same model signals and corresponding astrophysical parameters used to train and test 21cmGEM [available at <https://doi.org/10.5281/zenodo.4541500>, 161]. Examples of the global signals from the training set are shown in the left panel of Fig. 3.2. In total the data set contains 27,292 training models and 2,174 test models with each model being dependent on 7 astrophysical parameters. Each global signal in the data set has 451 redshift data points, and this means that each signal corresponds to 451 training points with the same astrophysical parameters and different redshifts. Therefore, for GLOBALEMU the 27,292 training

models become 12,308,692 training data points. However, we continue throughout this chapter to talk generally of training signals rather than data points because the emulator will be used to determine the signal structure over a redshift range (see section 3.6 for more details).

The data pre-processing is explained in detail in section 3.4. Explicitly, the inputs [13, 15, 91] to the neural network when training on the 21cmGEM data are as follows (ranges are based on those in the 21cmGEM training data set, see section 2.5 of [15]);

- $f_*$ : The star formation efficiency takes values in the range 0.0001 – 0.5 and characterises the amount of gas converted to stars in the dark matter halos. A low star formation rate results in a low Lyman- $\alpha$  flux and late onset of X-ray heating, leading to a shallower absorption trough and weak or non-existent emission above the radio background in the signal.
- $V_c$ : The minimal virial circular velocity has a value in the range 4.2 – 100 km/s and is proportional to the cube root of the minimum threshold mass for star formation. A low value of  $V_c$  corresponds to a small minimum mass threshold which in turn leads to an earlier onset of Lyman- $\alpha$  coupling, responsible for the absorption feature in the global 21-cm signal, and shifts the minimum of the signal to higher redshifts.
- $f_X$ : The X-ray efficiency of sources has a range between 0 – 1000 and a high value corresponds to a high total X-ray luminosity. This leads to an earlier onset of X-ray heating which also shifts the minimum of the signal to higher redshifts, contributes to a shallower absorption and results in a significant emission feature during reionisation at low redshifts.
- $\tau$ : The CMB optical depth in the 21cmGEM data sets takes a value in the range 0.04 – 0.2 and a higher value of  $\tau$  corresponds to a higher value of the ionizing efficiency of sources,  $\zeta$ . For high  $\tau$  we would see an earlier reionisation of the hydrogen gas. We note that  $\tau$  is given as  $0.054 \pm 0.007$  by [22] and that this falls at the lower end of the range in our training and testing data set. More recent parameter studies have explored lower values of  $\tau$  in greater detail [14]. However, the 21cmGEM data is sufficient to demonstrate the abilities of GLOBALEMU.
- $\alpha$ : The power defining the slope of the X-ray Spectral Energy Density (SED) with a range given by 1 – 1.5. The global 21-cm signal is expected to have a very weak dependence on  $\alpha$  with the largest effect happening at low redshifts.
- $\nu_{\min}$ : The low energy cut-off of the X-ray SED has a range of 0.1 – 3 keV. Low values of  $\nu_{\min}$  correspond to a soft X-ray SED, efficient X-ray heating and a weak absorption feature in the 21-cm signal.
- $R_{\text{mfp}}$ : The mean free path of ionizing photons, with a range 10 – 50 Mpc.  $R_{\text{mfp}}$  is expected to have a very weak effect and only at low redshifts [see e.g. 91]. A low  $R_{\text{mfp}}$  corresponds to a slower ionization of the neutral hydrogen gas.
- $z$ : The redshift of the 21-cm brightness temperature is a measure of time and provides details about when each feature of the signal occurred. For example, the brightness temperature is expected to reach 0 mK, corresponding to the end of the EoR, at low redshifts or more recent times. It is interchangeable with frequency given that the rest frequency,  $\nu_r$ , of the 21-cm line is 1420 MHz

$$z + 1 = \frac{\nu_r}{\nu}. \quad (3.1)$$

To ensure that we make a fair comparison of our results with those found when using 21cmGEM we make the same physically motivated cuts to the test data as are detailed in section 2.4 of [15]. This equates to limits on the ionizing efficiency of sources,  $\zeta < \zeta_{\max} = 40,000 f_*$  and on the neutral fraction history at  $z = 5.9$ ,  $x_{HI}(z = 5.9) < 0.16$ . Respectively, the limits are motivated by stellar models [162] and quasar absorption

troughs [163]. We also note that some of the parameters in the testing data have different ranges and the ranges are as follows;  $f_*$ : 0.0001 – 0.5,  $V_c$ : 4.2 – 76.5 km/s,  $f_X$ : 0 – 10,  $\tau$ : 0.055 – 0.1,  $\alpha$ : 1 – 1.5,  $\nu_{\min}$ : 0.1 – 3 keV and  $R_{\text{mfp}}$ : 10 – 50 Mpc. In total, the final testing data set comprises 1703 models.

GLOBALEMU includes a simple-to-use PYTHON graphical user interface (GUI)<sup>c</sup> in which the variation of the signal with each of the astrophysical parameters listed above can be explored in more detail. We note that the GUI is a feature made possible by the speed of emulation when using GLOBALEMU (see section 3.6). There is an equivalent GUI for the neutral fraction history emulation.

As previously stated, GLOBALEMU is not limited to emulating signals modelled with the above astrophysical parameters. It is flexible enough that more complicated astrophysical relationships can also be emulated. For example, one explanation for the unexpected depth of the EDGES absorption trough is the presence of a higher than expected radio background, which can be characterised with a quantity  $f_{\text{radio}}$  determining the normalisation of the radio emissivity [assuming the source of the excess radio background is stellar, 121]. In chapter 5 we train GLOBALEMU on such models, models with a different model of the excess radio background and standard astrophysical models with updated physics in comparison to the 21cmGEM models.

For the neutral fraction,  $x_{HI}$ , we use a set of models produced as a by-product of the detailed 21cmGEM global signal simulations. The data set is smaller with 10,047 training models and 791 test models however the relationship between the astrophysical parameters and  $x_{HI}$  is expected to be (and shown to be, section 3.6) simpler. We note that for  $z \gtrsim 30$  the neutral fraction is expected to always be 1, and so we only emulate the neutral fraction over the range  $z = 5 – 30$ . The models have not been released publicly, but the parameter ranges are the same for this data set as detailed above. A subsample of the training models are shown in the right panel of Fig. 3.2.

We note that a non-uniform coverage of the parameter space in the training data set, as with the 21cmGEM global signal and neutral fraction data, may introduce bias in the neural network. The network will tend to learn regions of the astrophysical parameter space where the sampling is heavier better than others. For the purposes of illustrating the accuracy of the emulation in this chapter, this is not an issue. Training a network on a more uniform data set can alleviate this issue.

## 3.4 Data Pre-processing

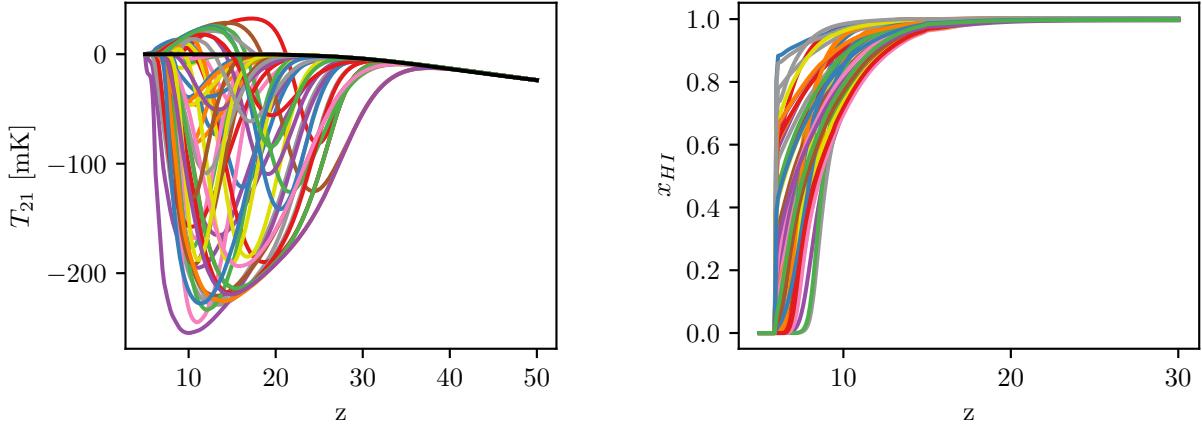
The details in the following discussion outline the pre-processing for the network predicting the global 21-cm signal. The various pre-processing steps outlined can be switched on and off by a user when training an emulator with GLOBALEMU. In section 3.4.4 we briefly discuss the pre-processing for the neutral fraction networks, which is a largely similar process. The pre-processing is summarised as a flow chart in Fig. 3.3.

### 3.4.1 Astrophysics Free Baseline Subtraction

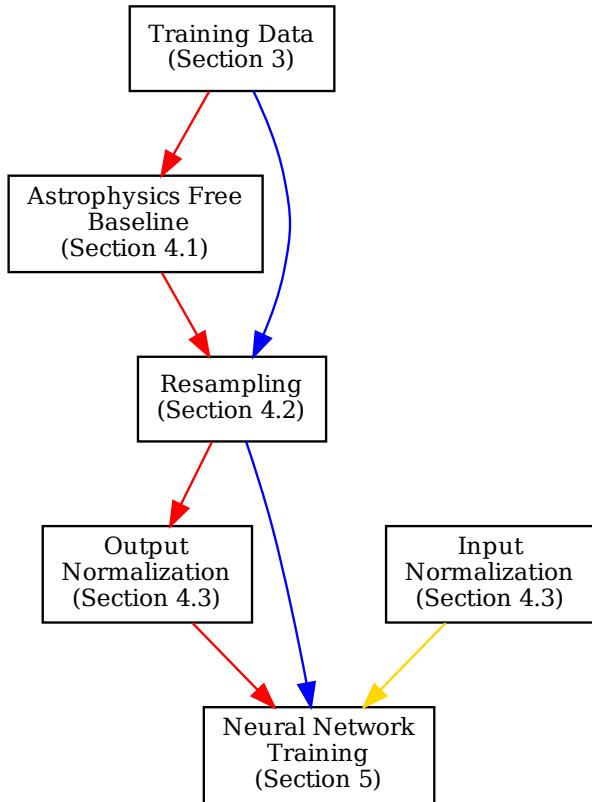
In the region where the structure of the global 21-cm signal is expected to be dominated by collisional coupling, it is independent of the 7 astrophysical parameters used here as inputs to the emulator. This means that, in the corresponding redshift range, each of the signals in our training and testing data sets have the same brightness temperatures. To prevent our network unnecessarily attempting to learn a non-trivial structure in this region,

---

<sup>c</sup>After installation via pip or from source, the GUI can be called from the terminal using the command `globalemu`. See the documentation at <https://globalemu.readthedocs.io/> for more details.



**Figure 3.2: Left Panel:** A subsample of 50 global 21-cm signals from the 21cmGEM training set used here to demonstrate the efficiency of GLOBALEMU. The signals show the expected variety of structure, with deep and shallow absorption troughs caused by Lyman- $\alpha$  coupling and terminated by X-ray heating. We also see emission against the CMB background at low redshifts in some of the models where there has been sufficient heating. Also shown in black is the Astrophysics Free Baseline (AFB) (section 3.4.1) which we model and remove from the training signals before we pass them through the neural network. Subtraction of the AFB prevents our network from attempting to learn a steadily decreasing temperature at high redshifts prior to star formation. **Right Panel:** A subsample of 50 neutral fraction histories from the training set used in this chapter. At high redshift, the hydrogen in the universe is predominantly neutral and consequently  $x_{HI} = 1$ . As the gas is ionised by UV emission from the first stars, the neutral fraction decreases until  $x_{HI} = 0$  at the end of the EoR.



**Figure 3.3:** The pre-processing applied to the training data in GLOBALEMU. Each box is outlined in more detail in the corresponding sections. The red path is the pre-processing steps used for the global 21-cm signals, the blue path for the neutral fraction histories and the gold path are steps that occur when training both neural networks.

we can treat it as an Astrophysics Free Baseline (AFB), model and remove it from the signals before they are passed to the network for training. By doing this, our network will learn a simpler relationship at high redshift between the parameters and  $T_{21}(z)$  than the existing steadily decreasing trend (see Fig. 3.2).

In Appendix A in [4] we give an approximate calculation of the AFB for the simulated signals that comprise the training data sets. The calculation follows the mean evolution of the signal and in contrast the simulations are produced over large-scale cosmological volumes evolved over cosmological time then averaged. We therefore normalise our result to the temperature of the signals in the training set at  $z = 50$  and find that this is sufficient to represent the astrophysics free component of the signals.

As stated, the AFB is then subtracted from the models before training the network and added back in after making predictions.

21cmGEM uses five extra parameters in addition to the seven astrophysical parameters used here. In principle, these parameters could be passed to the neural network during training due to the flexible nature of GLOBALEMU. However, three of these additional parameters rely on the fraction of mass contained in halos above the minimum cooling threshold,  $f_{\text{coll}}(z)$ , and help the network learn the signal structure at high redshift where collisional coupling (cosmology) dominates. Here we do not consider these parameters, which are derived from the seven astrophysical parameters described above, as we are instead removing the AFB. The final two parameters are the fractions of X-ray energy above 1 keV and 2 keV. These parameters are added to further characterise the X-ray SED, but we find with GLOBALEMU that we do not need to consider them to achieve accurate results.

### 3.4.2 Resampling of signals

The turning points, and gradients between them, of the global 21-cm signal encode all the information about the efficiency of Lyman- $\alpha$  coupling, X-ray heating and reionization. They are therefore highly dependent on the relevant astrophysical parameters and in the region where the features typically occur variation in the signals is significant. The original signal models are sampled uniformly in redshift, and there is no particular physical motivation for this. However, to improve the quality of modelling we resample the signals at a higher rate across the redshift ranges that typically correspond to the locations of the turning points and at a lower rate where the signal structure deviates less from the ‘average’ signal (e.g. above  $z \approx 30$  where the signal is free of astrophysics).

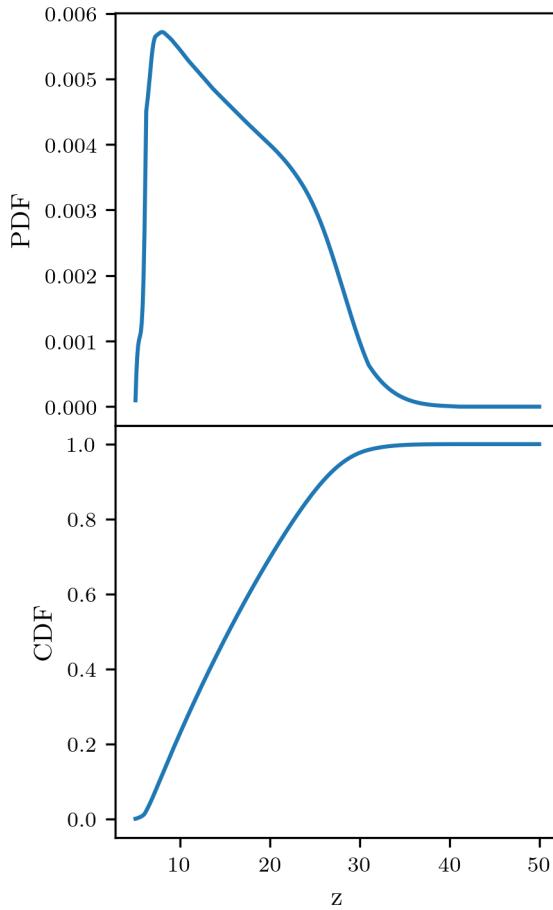
To do this we look at the variation in the signal amplitudes, after subtracting the AFB, across the training data set

$$\Delta T_{21}(z) = T_{21, \max}(z) - T_{21, \min}(z), \quad (3.2)$$

and we treat this as a probability distribution

$$P(\Delta T_{21}(z)) = \frac{\Delta T_{21}(z)}{\sum_z \Delta T_{21}(z)}. \quad (3.3)$$

Where the variation in the signal at a given redshift across the training data set is large the probability distribution is also large (see Fig. 3.4). We then calculate the corresponding cumulative distribution function (CDF) and use inverse transform sampling to produce a new redshift distribution with a high sampling rate in regions of high variation. For each signal, we can then perform an interpolation to get the corresponding  $T_{21}$  values.



**Figure 3.4: Top panel:** The probability distribution calculated from the difference between the maximum and minimum signal temperatures in the 21cmGEM training data set using equation (3.2) and equation (3.3). **Bottom panel:** The corresponding cumulative distribution function (CDF). We use this CDF to resample the 21-cm signals to better capture the variation at low redshifts and allow the emulator to more easily learn the structure of the signal.

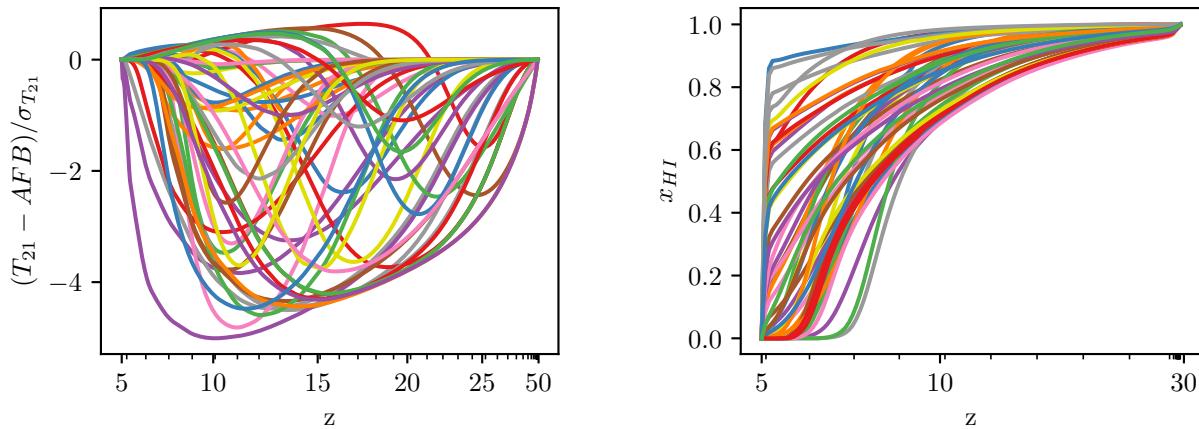
### 3.4.3 Output and Input Normalisation

Neural networks typically perform better when the outputs and inputs are of order unity and uniformly distributed. Hence, it is typical to manipulate the data sets via logarithms, normalisation and/or standardisation to improve performance.

After subtracting the AFB and resampling our signals, we also proceed to divide the signals by the standard deviation across the training data set. This type of scaling was motivated by the typically used standardisation technique, however we wanted to ensure that when scaling our signals the physically significant value of  $T_{21} = 0$  remained as 0. The signals shown in Fig. 3.2, as seen by the neural network, after pre-processing, are shown in Fig. 3.5.

For the input redshift distribution, we transform our resampled redshifts back onto a uniform distribution between 0 and 1 using the Cumulative Density Function (CDF) detailed in the previous section. It is the combination of resampling and uniform redshift input that ensures the neural network ‘sees’ ‘stretched’ signals as in the left panel of Fig. 3.5. This technique allows the neural network to interpolate the signal at redshifts it hasn’t been trained on to a higher degree of accuracy, where the signals vary greatly than if we had used uniform sampling.

For the other input astrophysical parameters we use a Min-Max normalisation scaling each feature between 0 and 1. For example, considering the distribution of the CMB optical depth,  $\tau$  in our training data as a vector



**Figure 3.5: Left Panel:** The equivalent signals from Fig. 3.2 after pre-processing. Subtraction of the AFB and the subsequent resampling mean that the important information encoding the dependence on the astrophysical parameters is retained and appropriately emphasized in the training data. Here we have plotted the resampled redshift data points as being uniformly distributed, since this is how the network is set up to interpret the input. The following division by the standard deviation across the training data set scales the signals to order unity without changing the physically significant value of  $T_{21}(z) = 0$  where the spin temperature of the neutral hydrogen is equivalent to the radio background temperature. Minor ticks are at intervals of one on the x-axis. **Right Panel:** The equivalent neutral fraction histories from Fig. 3.2 after pre-processing. For the neutral fraction histories, since the signals are already of order unity and subtraction of the equivalent AFB would not be beneficial, the pre-processing just involves resampling of the signals around regions of high variation. As with the left panel, the minor ticks are at intervals of one on the x-axis.

we normalise it such that

$$\tilde{\tau} = \frac{\tau - \tau_{\min}}{\tau_{\max} - \tau_{\min}}. \quad (3.4)$$

The decision to use this type of normalisation was arrived at after testing standardisation, Min-Max normalisation and division by the max values for the input parameters while maintaining the physically motivated pre-processing for the signal temperatures detailed above.

For  $f_X$ ,  $f_*$  and  $V_c$  the distributions are uniform in log-space, and so we perform the Min-Max normalisation on the logarithm of these variables and use these as our inputs.

### 3.4.4 $x_{HI}$ Pre-processing

As discussed, we provide provision to emulate the neutral fraction of hydrogen as a function of redshift with GLOBALEMU. The pre-processing just involves resampling of the signals since:

- The equivalent AFB for the neutral fraction has a value of 1 at all redshifts, and to subtract this from our training data set would just invert our signals, providing no benefit to training.
- The neutral fraction has a value between 0 and 1 by definition, and so we do not need to normalise the output of the network to be of order unity.

Again, resampling the signals allows the network to learn the variation in the training models and interpolation across redshift with a higher degree of accuracy. We perform the resampling with the equivalent of equation (3.2) and equation (3.3). The right panel of Fig. 3.5 shows the same set of neutral fraction histories as in Fig. 3.2 after pre-processing.

## 3.5 Neural Network Structure

As stated, the goal with GLOBALEMU is to maintain a simple network that is highly accurate without having to rely on non-trivial machine learning tools such as dropout, regularisation and batch normalisation. However, in the design of any neural network the optimizer, the architecture, loss function, activation function and learning rate are core considerations.

### 3.5.1 Architecture

Dropout [159] and the commonly used L1 and L2 regularisation are typically employed to prevent overfitting where the network learns the training data to such a high degree of accuracy that it is unable to generalise. Overfitting is generally a result of using a neural network that is too big and has an excessive number of layers and nodes. On the other hand, a network that is too small often produces poor quality predictions and consequently the aim is to produce a ‘reasonably’ sized network. The scope of what constitutes a reasonably sized network is dependent on the number of input/output nodes, the variation in the training data and the complexity of the relationship between the inputs and outputs.

By using the novel approach of having redshift as an input to the network both our global signal and neutral fraction emulators have, in the case of the 21cmGEM data, eight input nodes and one output node meaning that our network can remain small. Additionally, we have made a significant effort to simplify the problem with physically motivated pre-processing, which helps to reduce what constitutes as ‘reasonable’ size for our networks.

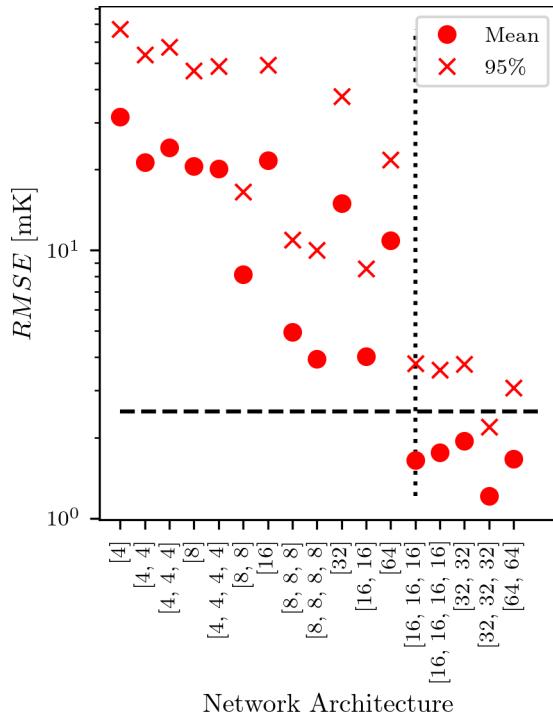
GLOBALEMU is set up in such a way that the number of layers and layer sizes can be adjusted by the user. As a result, we do not provide a prescription of what constitutes a ‘reasonable’ size, as this may not be pragmatic. We note, however, that a significant effort can be undertaken to determine the optimum ‘reasonable’ architecture that maximises accuracy and that this can also be impractical. Instead, we suggest that as a minimum requirement a ‘reasonable’ architecture for a trained GLOBALEMU model should meet the following criteria:

- The network should not over fit the training data otherwise the predictive power will be lost.
- The network should have an average accuracy that is  $\lesssim 10\%$  of the expected noise of a typical global 21-cm experiment (see section 3.6.2 for a further discussion).

Based on the above criteria, the size of our input and output layers and a minimal exploration starting from a small network, trained with the pre-processed signals, and increasing the size until our accuracy criteria was met without overfitting (see Appendix B in [4]) we use a network with 3 hidden layers all with 16 nodes for both the global signal and neutral fraction emulation in this chapter.

Fig. 3.6 illustrates the processes used to determine our architecture for the global network. We consider a set of different network sizes with one to four layers and 4, 8, 16, 32 or 64 nodes in each layer. For each of the tested networks we run a ‘full’ training of GLOBALEMU corresponding to 12 hours on a HPC equating to approximately 250 epochs using the full 21cmGEM training data. We then assess the accuracy of the trained models using the  $\approx 1,700$  testing models in the 21cmGEM dataset. We compare the mean and 95 percentile *RMSE* (see section 3.6.2) for each architecture. We find that a network of size [16, 16, 16] is the first to meet our target accuracy of on average 10 % the expected noise in a global 21-cm experiment.

While we may be able to achieve a better accuracy with a larger network this pragmatic approach leads to a sufficiently accurate network for physical signal modelling in the data analysis pipeline of a global experiment



**Figure 3.6:** The mean and 95 percentile  $RMSE$  (see section 3.6.2) for a set of different network architectures trained for 12h (approximately 250 epochs) on a HPC with the 21cmGEM global signal training data and assessed with the corresponding test data. The architectures have between 1 and 4 layers of varying sizes between 4 and 64 nodes. They are ordered based on the number of weights in the network (equivalent to the number of connections) as this is a useful measure of network size and an indication of predictive power. The graph is used to determine a ‘reasonable’ architecture considering the practical target accuracy of on average 10 % the expected noise of a global 21-cm experiment (here illustrated by the black dashed line at 2.5 mK). Throughout the rest of the chapter we use a network with 3 layers each consisting of 16 nodes which is the first to produce a mean value within our target accuracy. Our choice is highlighted with a dotted vertical line.

like REACH. We also note that a smaller network can be evaluated faster than a larger architecture and that this is important when we are making multiple evaluations inside a Nested Sampling loop.

### 3.5.2 Loss Function and Learning rate

GLOBALEMU uses the Mean Squared Error, typical for a regression network, as the loss function. In the case of the global signal network is given by

$$MSE = \frac{1}{N} \sum_{i=0}^N (T_{21,\text{sim},i}(z_i) - T_{21,\text{pred},i}(z_i))^2 \quad (3.5)$$

where  $N$  is a batch size equivalent to the number of redshift data points in each signal.  $T_{21,\text{sim}}(z)$  is the simulated signal temperature at a given redshift and  $T_{21,\text{pred}}(z)$  is the emulated equivalent. GLOBALEMU trains the neural networks in batches primarily to prevent memory related issues since the training data can be large ( $\approx 27,000$  models times 451 redshift points  $\approx 12$  million data points for the 21cmGEM data). We find that a reasonable batch size is equal to the number of redshift data points in each model.

For the 21cmGEM data and the GLOBALEMU framework, we determine an effective learning rate to be 0.001. As with the architecture, the learning rate can be adjusted by the user of GLOBALEMU to meet the requirements of the data that they are training on.

### 3.5.3 Optimizer

The neural network optimizer is used to change the networks hyperparameters to minimise the loss function. There are a number of different optimizers available [164] and the choice can be dependent on the complexity of the problem and loss surface. A more robust optimizer is less likely to fall into and get stuck in local minima when training the network, resulting in more accurate emulation. Therefore, the choice of optimizer is important in designing an effective emulator. However, since GLOBALEMU is designed to minimise the complexity of the relationship between the inputs and outputs and a Mean Squared Error (MSE) loss surface is relatively smooth<sup>d</sup> our choice is less consequential. We use, therefore, the commonly applied Adam [165] optimizer which is a momentum based modified stochastic gradient descent algorithm.

### 3.5.4 Activation Functions

For both the global signal and neutral fraction history network, we use a *tanh* activation function in the hidden layers, which can range between (-1, 1). However, for our final layer in the global signal network we use a linear activation since the pre-processed temperature can be positive or negative and range between approximately -4 and 0.5. Similar consideration is given to the output layer in the neutral fraction network, where we use a *ReLU* (Rectified Linear Unit) activation, which ensures that the output is always positive. Again, the activation functions can be changed by the user of GLOBALEMU to meet the requirements of their data. We note that the above output layer activations are designed to prevent unphysical outputs, and that this is a crucial consideration for any user.

## 3.6 Results

### 3.6.1 Emulation Time

In [15] the reported average time taken per signal with 21CMGEM is 160 ms when emulating a set of signals in a vectorised call. Here we compare the speed of 21CMGEM and GLOBALEMU by emulating the 1,703 test models and taking an average time per signal. The tests are performed with MATLAB and PYTHON respectively on the same computer with the following processors: Intel® Core™ i3-10110U CPU @ 2.10GHz × 4. For 21CMGEM we make a vectorised call to the emulator as this results in a quicker performance than repeated single calls. We use a for loop to repeatedly call GLOBALEMU which currently doesn't support such vectorised calls as they are not needed for physical signal modelling in a Nested Sampling loop<sup>e</sup>.

For GLOBALEMU we record a total time of 2.29 s and a corresponding average time per signal of  $1.3 \pm 0.01$  ms. In comparison, when emulating the same signals in a vectorised call with 21CMGEM we record a total time of 227.18 s and an average time per signal of 133 ms. We therefore achieve a factor of 102 improvement in emulation time with GLOBALEMU. We note that when using the PYMATBRIDGE (<https://github.com/arokem/python-matlab-bridge>) PYTHON wrapper for MATLAB the average time take to run a single prediction with 21CMGEM using a vectorised call is comparable to a direct call in MATLAB.

---

<sup>d</sup>This can be assessed with a plot of the loss vs epoch number during training. We find that for the results presented in section 3.6 the surface is smooth up until the loss has plateaued and training is complete, at which point we see noise like behaviour.

<sup>e</sup>Since this work was originally published the ability to perform vectorised calls has been added to GLOBALEMU.

### 3.6.2 Measuring Accuracy

In this section we primarily consider the accuracy of the global signal emulator. We note, as previously stated, that the relationship between the neutral fraction and the astrophysical parameters is expected to be simpler and therefore easier to learn.

To assess the accuracy of GLOBALEMU when emulating a global 21-cm signal simulation we use a combination of two metrics; the root mean squared error (*RMSE*) and the normalised *RMSE* given in [15] as

$$\widetilde{RMSE} = \frac{RMSE}{\max|T_{21,\text{sim}}(z)|}, \quad (3.6)$$

where

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^N (T_{21,\text{sim},i}(z_i) - T_{21,\text{pred},i}(z_i))^2}. \quad (3.7)$$

For the neutral fraction network  $\widetilde{RMSE}$  and the *RMSE* are equal since  $\max|x_{\text{sim}}^{HI}(z)| = 1$ . We assess the accuracy in the uniform redshift space, and consequently our assessment is independent of the loss function used for training.

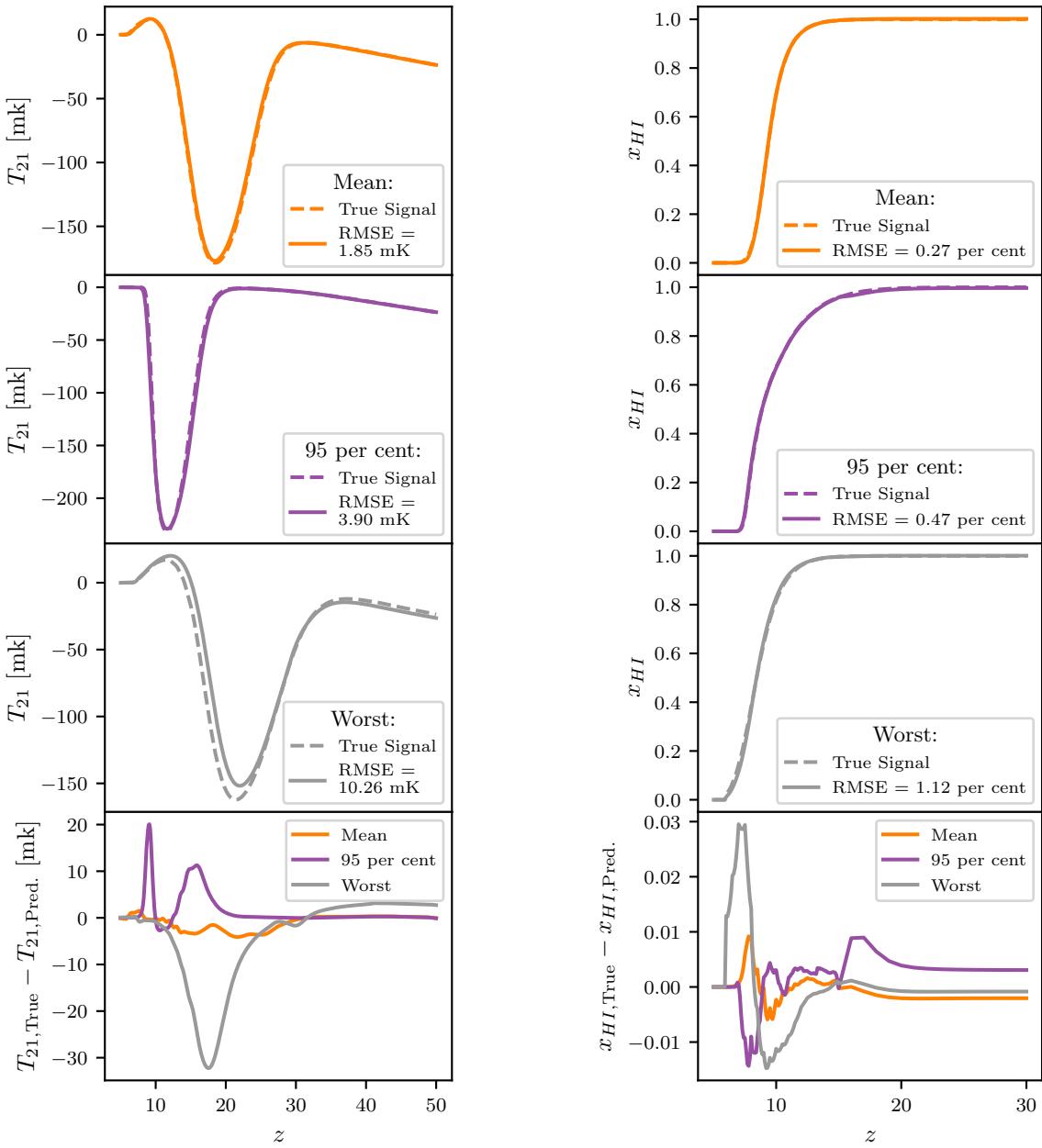
$\widetilde{RMSE}$  is a dimensionless quantity used by 21cmGEM and by assessing the quality of our network with this metric we can make direct comparisons between the two emulators. We also want our emulator to have an accuracy significantly lower than the expected noise of a global 21-cm experiment. This is required if the emulator is to be used to confidently model the 21-cm signal and draw conclusions about the astrophysics during the CD and EoR. To assess this accuracy requirement, we can use the dimensionful *RMSE* metric.

As highlighted in the previous section, we suggest an average accuracy of  $\lesssim 10\%$  the expected noise of a global 21-cm experiment such as REACH, equivalent to an *RMSE*  $\lesssim 2.5$  mK, to be a sufficient limit. Since the accuracy of emulation is a function of the bandwidth, we report the accuracy across the entire range of the simulations  $z = 5 - 50$  ( $z = 5 - 30$  for the neutral fraction network) and across the expected REACH Phase I bandwidth of  $z = 7 - 28$  [10]. The noise in a 21-cm experiment has a fundamental effect on our confidence in any astrophysical parameter values inferred from the data, and, given our target accuracy, this will likely be larger than the uncertainty introduced from GLOBALEMU.

### 3.6.3 Global 21-cm Signal

Fig. 3.7 shows that the mean *RMSE* value across the redshift range  $z = 5 - 50$  is 1.85 mK and that the maximum value is 10.26 mK. Further, Tab. 3.1 shows that performing the same calculation of the *RMSE* inside the REACH band,  $z = 7 - 28$  gives a mean value of 2.52 mK which is very close to the desired 2.5 mK limit. We also report in the table the *RMSE* value for which 95 % of the models have a value smaller than or equal to. In the REACH band, this equates to 5.37 mK. For all the reported results, the 95 percentile is significantly lower than the maximum values (a factor of 3 for the global signal and a factor of approximately 1.5 to 2 depending on bandwidth for the neutral fraction). This means that out of a set of 1,703 global signals only 85 have *RMSE* values above 3.90 mK across the band  $z = 5 - 50$  for example. We note that the values reported, averages across redshift ranges, in the REACH band are generally higher than across the whole redshift range because the REACH band excludes redshifts  $\gtrsim 30$  where the emulation is expected to be very precise.

Appendix C in [4] shows the error as a function of the explored parameter space for the global 21-cm signal in the 21cmGEM test data set.



**Figure 3.7: Left Panel:** The top three panels show the mean, 95 percentile and the worst emulations respectively, based on the  $RMSE$ , for the global 21-cm signal across the entire test set of 1,703 models. The bottom panel shows the difference between the simulations and predictions as a function of redshift. Full details of the accuracy of the emulation can be found in Tab. 3.1, and a discussion can be found in the text. **Right Panel:** The top three panels show the mean, 95 percentile and the worst emulations respectively for the neutral fraction history across the entire test set of 791 models and the difference between the simulations and predictions is shown in the bottom panel. The level of accuracy here is higher than that for the global signal, despite using a similar pre-processing and identical network, demonstrating that the relationship between the astrophysical parameters, redshift and the network output is simpler here.

In Tab. 3.1, we also report the  $\widetilde{RMSE}$  values in both the REACH band and across the whole redshift range. [15] report similar results for 21cmGEM. When training and testing on the same data sets, we recover a mean  $\widetilde{RMSE}$  of 1.12 % compared to 1.59 % when using 21cmGEM. Similarly, we report a maximum value of 6.32 % in comparison to the value of 10.55 % reported for 21cmGEM. GLOBALEMU can achieve a high degree of accuracy in its emulation.

		Global Signal		Neutral Fraction	
		$z = 5 - 50$	$z = 7 - 28$	$z = 5 - 30$	$z = 7 - 28$
$RMSE$	Minimum	0.30 mK	0.31 mK	0.09 %	0.08 %
	Mean	1.85 mK	2.52 mK	0.29 %	0.26 %
	95 <sup>th</sup> percentile	3.90 mK	5.37 mK	0.47 %	0.44 %
	Maximum	10.26 mK	15.10 mK	1.12 %	0.65 %
$\widetilde{RMSE}$	Minimum	0.21 %	0.26 %	—	—
	Mean	1.12 %	1.53 %	—	—
	95 <sup>th</sup> percentile	2.41 %	3.22 %	—	—
	Maximum	6.32 %	9.31 %	—	—

**Table 3.1:** Detailed results of the emulation using GLOBALEMU and the 21cmGEM training and test data for both the global signal and the neutral fraction history. We find that GLOBALEMU achieves the desired accuracy of on average  $\approx 10\%$  the expected noise of a typical global 21-cm experiment (equating to  $\approx 2.5$  mK in the REACH band of  $z = 7 - 28$ ). Of note are the recorded 95 % percentiles, the  $RMSE$  for which 95 % of the models have values smaller than or equal to, which are significantly lower than the maximum  $RMSE$  values. A discussion comparing the results of 21cmGEM and GLOBALEMUCan be found in the text. We find that our global signal emulator has a maximum  $\widetilde{RMSE}$  approximately half that achieved with 21cmGEM. For the neutral fraction  $RMSE = \widetilde{RMSE}$ . We find a higher degree of accuracy for the neutral fraction with an identical network and similar pre-processing, indicating a simpler relationship.

### 3.6.4 Neutral Fraction

For the neutral fraction history network, we show similar results. The right panel of Fig. 3.7 demonstrates the quality of the emulation with the mean, 95 percentile and the worst results when emulating the neutral fraction and these values are detailed in Tab. 3.1. The results generally are of higher quality than that for the global signal and, noting that the pre-processing for the two networks is near identical and the networks themselves are of the same size, this supports the understanding that the relationship between the inputs and outputs is simpler here. In the band  $z = 5 - 30$  only 39 of 791 test models have  $\widetilde{RMSE} \geq 0.47\%$ .

## 3.7 Conclusions

GLOBALEMU uses a novel approach to emulate, with neural networks, the global 21-cm signal and the evolution of the neutral fraction during the CD and EoR by considering redshift as an input to the neural networks alongside the astrophysical parameters. In tandem with this reparameterisation of the problem, we use a predominantly physically motivated pre-processing for both the global signal and neutral fraction. We subtract from the global signals an astrophysics free baseline, which obviates the need for the network to learn a non-trivial but well-understood relationship at high redshift. We then resample both the global signals and neutral fractions so that the regions which vary significantly across the training data sets can be better characterised by the networks.

The above framework allows the complex relationships between the astrophysical parameters and the global signal or neutral fraction history as functions of redshift to be effectively learnt with small neural networks. Each high resolution global signal of 451 redshift data points can be emulated in on average 1.3 ms. We note that this is a factor of approximately 102 improvement on the 133 ms we record with MATLAB when predicting the same signals on the same computer with 21cmGEM.

We demonstrate the effectiveness of GLOBALEMU by using the 21cmGEM training and testing data. This allows for a direct comparison between our results and the results of 21cmGEM. We find that GLOBALEMU can emulate to a higher degree of accuracy the global 21-cm signal than 21cmGEM with a maximum normalised  $RMSE$  of 6.32 % in comparison to 10.55 % over the range  $z = 5 - 50$ . We also demonstrate that GLOBALEMU can

emulate a global 21-cm signal to, on average, less than 10 % the expected noise of a global 21-cm experiment like REACH.

Finally, `GLOBALEMU` is a flexible `PYTHON` package that can be easily retrained on updated models with new astrophysical dependencies. For example, additional astrophysical phenomena such as Lyman- $\alpha$  heating [14] or additional radio background produced by galaxies (or an indeterminate synchrotron-like source) [121, 131] can be incorporated and easily trained upon, as shown in the second half of this thesis. While the results achieved with the 21cmGEM data are impressive, the novelty of `GLOBALEMU` is in its flexibility, incorporation of redshift as an input and physically motivated pre-processing. Particularly the final two points allow for an accurate mapping from parameters to temperature with a single neural network reducing the points of failure and need for excessive fine-tuning.

`GLOBALEMU` has been used to emulate models from the ARES[149] simulation code by members of the Network for Exploration and Space Science research group at the University of Colorado Boulder. Further, `GLOBALEMU` influenced the design of the power spectrum emulator used in the recent HERA analysis (section 7.5 [99]).



CHAPTER

4

**MARGARINE**

## 4.1 Introduction

Bayesian analysis has been applied in the inference of cosmological parameters from experiments such as the CMB mapper Planck [22], the Dark Energy Survey (DES, [166–168]), 21-cm power spectrum experiments (such as HERA [109], LOFAR [92, 105] and MWA [93, 108]) and global or sky-averaged 21-cm experiments such as REACH [57] and SARAS [6, 7].

These experiments are often plagued by ‘nuisance’ parameters that characterise foregrounds and instrumental effects in the data. This leads to high dimensional parameter spaces, of which only a few of the parameters model the core science. Consequently, drawing conclusions about the parameters of interest and making comparisons of different experiments can be challenging. For example, in the Year 1 analysis of data from DES the likelihood contains a total of 26 parameters of which 20 could be considered nuisance parameters with only 6 corresponding to cosmological parameters of interest [166]. Similarly, for REACH [10, 57] the likelihood can contain around 20–30 parameters, with only 3–7 of these corresponding to the signal.

In this chapter, we present a post-processing tool called `MARGARINE` that can be used to calculate *marginal* or *nuisance-free* posterior distributions pertaining to the core science goals of the above experiments through density estimation with Masked Autoregressive Flows (MAFs, [169]) and Kernel Density Estimators (KDEs, [170, 171]).

Due to the ubiquity of Bayesian analysis in cosmology, post-processing of samples from MCMC and Nested Sampling algorithms is an important area of research. `MARGARINE` is an invaluable addition to the Bayesian workflow and allows for direct and specific comparison of the constraining ability of different experimental approaches, rapid emulation of nuisance-free likelihoods and experimentally informed priors.

We implement our approach in `PYTHON` using `TENSORFLOW` and the `KERAS` backend and release the code as the pip installable package `MARGARINE`. The code is fully documented with worked examples, subject to continuous integration testing and does not require high performance computing to be used.

In section 4.2, we briefly outline the motivation behind the development of `MARGARINE`. In section 4.3, we reintroduce Bayesian analysis and discuss the marginal statistical quantities that can be calculated with `MARGARINE`. We discuss the two different types of density estimators used by `MARGARINE` in section 4.4. In section 4.5, we demonstrate several applications of the code base, including applications to samples from the Dark Energy Survey, Planck and mock observations with the REACH pipeline. We summarise the chapter in section 4.6.

Section 4.3.2 was first drafted by Will Handley for [1], edited by the author and visualized in Fig. 4.1 and Fig. 4.2 by the author. The co-authors, Will Handley, Pablo Lemos, Peter Sims, Eloy de Lera Acedo, Anastasia Fialkov and Justin Alsing, of [1] and [5] provided comments on the manuscripts that this chapter is adapted from.

## 4.2 Motivation

If our model  $M(\Theta)$  contains nuisance,  $\alpha$ , and cosmological signal parameters,  $\theta$ , then we can write  $\Theta = \{\alpha, \theta\}$ . Mathematically, the marginal posterior for the signal parameters,  $\mathcal{P}(\theta)$ , can be calculated by integrating out the dependence on  $\alpha$

$$\mathcal{P}(\theta) = \int \mathcal{P}(\alpha, \theta) d\alpha. \quad (4.1)$$

However,  $\mathcal{P}(\theta)$  is typically a difficult quantity to calculate. We can effectively perform this marginalisation by training density estimators on samples in  $\theta$  from MCMC or Nested Sampling chains. From these density estimators, we can generate samples from the marginal distribution  $\mathcal{P}(\theta)$  and importantly calculate  $\log \mathcal{P}(\theta)$  for a given set of  $\theta$ , which gives us access to a whole host of previously intractable marginal summary statistics.

For example, the marginal Kullback-Leibler (KL) divergence tells you how much information is gained by contracting specifically the core science prior onto the core science posterior without including contributions from correlations with or constraints on the nuisance parameters. This allows us to confidently compare the information gained from different data sets from different experiments measuring the same signal, regardless of their specific instrumental nuances. Similarly, the marginal Bayesian Model Dimensionality (BMD) can be used to compare the effective number of core science parameters constrained by different experiments and models. Both the *marginal* KL divergence and the BMD, therefore, provide a way to quantitatively determine which experimental approaches are the most informative, which can inform improvements to instrumentation in the future and prior to MARGARINE was not as easy to assess.

Using normalised density estimators to calculate the KL divergence, marginal or not, also has the added advantage that it does not require knowledge of the Bayesian evidence, which is not always accessible.

Further, since the density estimators can be used to empirically approximate the logarithm of any posterior distribution, then they can be used as a computationally inexpensive likelihood generator. In practice, we can take advantage of this to perform efficient analysis of constraints from multiple different experiments as demonstrated in this chapter and further in chapter 7.

Finally, we note that if the density estimator is set up correctly such that it is a bijective transformation between the unit hyper-cube and the target posterior, then it can be used to generate the prior on a subsequent Nested Sampling run. This allows you to use far more complex priors based on current experimental results, incorporating our current knowledge of the parameter space [172].

## 4.3 Theory

### 4.3.1 Bayesian analysis

Bayesian analysis is concerned with the estimation of posterior probabilities using Bayes theorem

$$P(\Theta|D, M) = \frac{P(D|\Theta, M)P(\Theta|M)}{P(D|M)} = \frac{\mathcal{L}(\Theta)\pi(\Theta)}{\mathcal{Z}}, \quad (4.2)$$

where  $\Theta$  are the parameters of our model  $M$  describing the data  $D$  and  $\mathcal{L}(\Theta)$  corresponds to the likelihood. The likelihood is often assumed to be Gaussian in nature (although more complicated likelihoods can be implemented [79]). The prior,  $\pi(\Theta) = P(\Theta|M)$ , encodes our existing knowledge about the parameters before any sampling is performed.  $\pi(\Theta)$  is often chosen to be uniform or log-uniform for each parameter, however we would often like to inform this with existing experimental results.

The evidence,  $\mathcal{Z}$ , is a (fully-)marginal likelihood integrated over all the parameters and weighted by the prior

$$\mathcal{Z} = \int \mathcal{L}(\Theta)\pi(\Theta)d\Theta. \quad (4.3)$$

### 4.3.2 Nuisance-free likelihood

We can re-write Bayes theorem as

$$\mathcal{L}(\theta, \alpha) \times \pi(\theta, \alpha) = \mathcal{P}(\theta, \alpha) \times \mathcal{Z}, \quad (4.4)$$

with the inputs of inference (likelihood and prior) on the left-hand side, and the outputs (posterior and evidence) on the right.

We may marginalise any probability distribution so can, in addition to the definition of the nuisance marginalised posterior, straightforwardly define the nuisance marginalised prior by integrating over  $\alpha$

$$\pi(\theta) = \int \pi(\theta, \alpha) d\alpha. \quad (4.5)$$

The nuisance marginalised version of Bayes theorem equation (4.4) takes the form

$$\mathcal{L}(\theta) \times \pi(\theta) = \mathcal{P}(\theta) \times \mathcal{Z}. \quad (4.6)$$

Here  $\mathcal{Z}$  is the original evidence, whilst  $\mathcal{L}(\theta)$  is non-trivially the *nuisance-free likelihood*

$$\mathcal{L}(\theta) \equiv \frac{\int \mathcal{L}(\theta, \alpha) \pi(\theta, \alpha) d\alpha}{\int \pi(\theta, \alpha) d\alpha} = \frac{\mathcal{P}(\theta) \mathcal{Z}}{\pi(\theta)}, \quad (4.7)$$

where the above is motivated by marginalising Bayes theorem over  $\alpha$ , and substituting the definitions in eqs. (4.1), (4.5) and (4.7) recovers the marginalised Bayes theorem equation (4.6).

The nuisance-free likelihood equation (4.7) is straightforward to compute in our framework since we (uniquely) have Nested Sampling-computed evidences  $\mathcal{Z}$  and we can generate emulators for the distributions  $\mathcal{P}(\theta)$  and  $\pi(\theta)$  with MARGARINE.

Let  $\mathcal{L}_A(\theta, \alpha_A)$  and  $\mathcal{L}_B(\theta, \alpha_B)$  be two likelihoods for distinct datasets, each with their own nuisance parameters. The nuisance-free likelihoods  $\mathcal{L}_A(\theta)$ ,  $\mathcal{L}_B(\theta)$  form a lossless compression in  $\theta$ . This means that we can recover the same (marginal) inference in combination that we would have made when performing a combined analysis with all nuisance parameters:

$$\mathcal{L}_A(\theta, \alpha_A) \mathcal{L}_B(\theta, \alpha_B) \pi_{AB}(\theta, \alpha_A, \alpha_B) = \mathcal{P}_{AB}(\theta, \alpha_A, \alpha_B) \mathcal{Z}_{AB}, \quad (4.8)$$

$$\Rightarrow \mathcal{L}_A(\theta) \mathcal{L}_B(\theta) \pi(\theta) = \mathcal{P}_{AB}(\theta) \mathcal{Z}_{AB}, \quad (4.9)$$

if their respective priors  $\pi_A(\theta, \alpha_A)$  and  $\pi_B(\theta, \alpha_B)$  satisfy the marginal consistency relations:

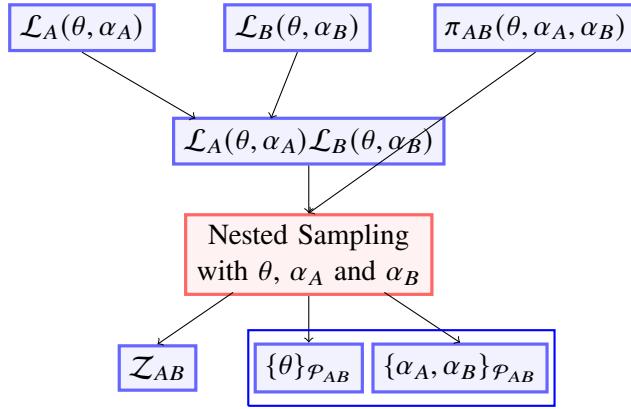
$$\pi(\theta) = \int \pi_A(\theta, \alpha_A) d\alpha_A = \int \pi_B(\theta, \alpha_B) d\alpha_B, \quad (4.10)$$

$$\int \pi_{AB}(\theta, \alpha_A, \alpha_B) d\alpha_A = \pi_B(\theta, \alpha_B), \quad \int \pi_{AB}(\theta, \alpha_A, \alpha_B) d\alpha_B = \pi_A(\theta, \alpha_A). \quad (4.11)$$

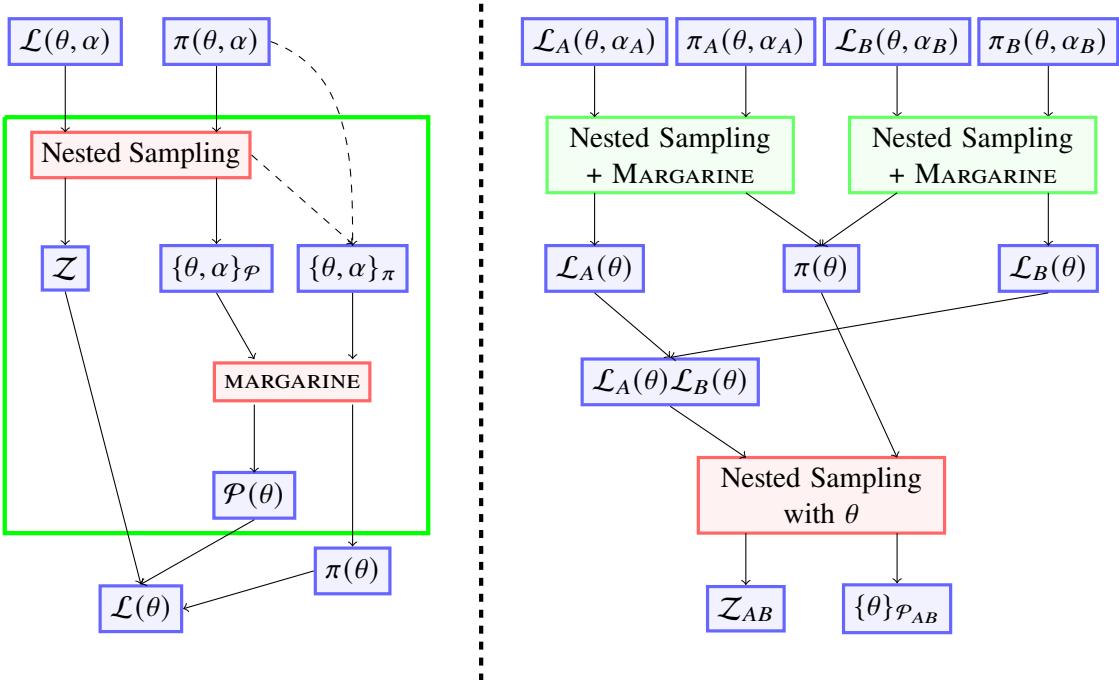
This process is represented graphically in Fig. 4.1 and Fig. 4.2.

Integrating the combined Bayes theorem equation (4.8) with respect to  $\alpha_B$ , applying the definition of the marginal posterior equation (4.5) on the right-hand side, and drawing out terms independent of  $\alpha_B$  on the left, yields

$$\mathcal{L}_A(\theta, \alpha_A) \int \mathcal{L}_B(\theta, \alpha_B) \pi_{AB}(\theta, \alpha_A, \alpha_B) d\alpha_B = \mathcal{P}_{AB}(\theta, \alpha_A) \mathcal{Z}_{AB}. \quad (4.12)$$



**Figure 4.1:** A graphical representation of combining constraints from different data sets via a full Nested Sampling run over both cosmological and nuisance parameters (equation (4.8)).



**Figure 4.2:** A graphical representation of combining constraints from two data sets via MARGARINE (equation (4.9)). Left of the dashed line illustrates the derivation of a nuisance-free likelihood function for one experimental data set.

From the definition of a nuisance-free likelihood equation (4.7), and the marginal consistency equation (4.11), we can say that the integral on the left-hand side becomes:

$$\begin{aligned}
 & \int \mathcal{L}_B(\theta, \alpha_B) \pi_{AB}(\theta, \alpha_A, \alpha_B) d\alpha_B \\
 &= \int \mathcal{L}_B(\theta, \alpha_A, \alpha_B) \pi_{AB}(\theta, \alpha_A, \alpha_B) d\alpha_B \quad [\mathcal{L}_B(\theta, \alpha_B) \equiv \mathcal{L}_B(\theta, \alpha_A, \alpha_B) \text{ since } \mathcal{L}_B \text{ indep of } \alpha_A] \\
 &= \mathcal{L}_B(\theta, \alpha_A) \int \pi_{AB}(\theta, \alpha_A, \alpha_B) d\alpha_B \quad [\text{Using equation (4.7) for } \mathcal{L}_B] \\
 &= \mathcal{L}_B(\theta) \int \pi_{AB}(\theta, \alpha_A, \alpha_B) d\alpha_B \quad [\mathcal{L}_B(\theta, \alpha_A) \equiv \mathcal{L}_B(\theta) \text{ since } \mathcal{L}_B \text{ indep of } \alpha_A] \\
 &= \mathcal{L}_B(\theta) \pi_A(\theta, \alpha_A). \quad [\text{Using marginal consistency equation (4.11)}]
 \end{aligned} \tag{4.13}$$

Substituting equation (4.13) back into equation (4.12) we find

$$\mathcal{L}_A(\theta, \alpha_A) \mathcal{L}_B(\theta) \pi_A(\theta, \alpha_A) = \mathcal{P}_{AB}(\theta, \alpha_A) Z_{AB}. \tag{4.14}$$

Proceeding with a similar manipulation to equation (4.13), marginalising with respect to  $\alpha_A$ , and applying the definition of the nuisance-free likelihood  $\mathcal{L}_A(\theta)$  equation (4.7) and the marginal prior consistency equation (4.10) we recover equation (4.9)

$$\mathcal{L}_A(\theta)\mathcal{L}_B(\theta)\pi(\theta) = \mathcal{P}_{AB}(\theta)\mathcal{Z}_{AB}.$$

Equation (4.8) represents Bayes theorem for the combined likelihood of both datasets  $\mathcal{L}_{AB}(\theta, \alpha_A, \alpha_B) = \mathcal{L}_A(\theta, \alpha_A)\mathcal{L}_B(\theta, \alpha_B)$ , using the combined prior  $\pi_{AB}(\theta, \alpha_A, \alpha_B)$ . We assume the combined prior is marginally consistent, eqs. (4.10) and (4.11), which is reasonable, merely demanding that the priors are identical in the parameter spaces where they overlap. In practice, this would usually be achieved by assuming separability between signal and nuisance parameter spaces  $\pi(\theta, \alpha) = \pi(\theta)\pi(\alpha)$ , but eqs. (4.10) and (4.11) are a slightly less restrictive requirement and therefore more general.

The upshot of this is that if you have performed inference for two datasets separately, such that you are able to compute the nuisance-free likelihoods with MARGARINE, you may discard the nuisance parameters for the next set of analyses when you combine the datasets as in Fig. 4.2.

We demonstrate applications of the nuisance-free likelihood in section 4.5.1 with a simple toy example and a real world example combining constraints on data from Planck and the Dark Energy Survey.

### 4.3.3 Marginal Kullback-Leibler divergence

The KL divergence quantifies the contraction from prior to posterior. The *marginal* KL divergence corresponding to the astrophysical parameters  $\theta$  is defined as the average Shannon Information

$$\mathcal{I}(\theta) = \log \frac{\mathcal{P}(\theta)}{\pi(\theta)} \quad (4.15)$$

over the posterior,  $\mathcal{P}(\theta)$ ,

$$\mathcal{D}(\mathcal{P}||\pi) = \int \mathcal{P}(\theta) \log \frac{\mathcal{P}(\theta)}{\pi(\theta)} d\theta = \langle \mathcal{I} \rangle_{\mathcal{P}} \approx \log \frac{V_{\pi}}{V_{\mathcal{P}}}, \quad (4.16)$$

and is a measure of how much information in bits the data provides us about the parameter space, which we can think of as the constraining power of the data. It is approximately equal to the log of the ratio of the prior volume,  $V_{\pi}$ , and the posterior volume,  $V_{\mathcal{P}}$ .  $\mathcal{D}$  is a strong function of the prior, and inherits the property of being additive for independent parameters from the Shannon Information. Calculation of the KL divergence requires the posterior distribution to be appropriately normalised and consequently the quantity is not attainable with common MCMC sampling techniques and more computationally intensive algorithms such as Nested sampling have to be implemented to attain the Bayesian evidence.

### 4.3.4 Marginal Bayesian Model Dimensionality

An alternative measure of constraint is the Bayesian Model Dimensionality [173] which gives a measure of the effective number of parameters that are being constrained and can be used to quantify tensions between different experimental results [174, 175]. The *marginal* BMD  $d$  is given by

$$\begin{aligned} \frac{d}{2} &= \int \mathcal{P}(\theta) \left( \log \frac{\mathcal{P}(\theta)}{\pi(\theta)} - \mathcal{D} \right)^2 d\theta \\ &= \text{var}(\mathcal{I})_{\mathcal{P}} \\ &= \langle (\log \mathcal{L})^2 \rangle_{\mathcal{P}} - \langle \log \mathcal{L} \rangle_{\mathcal{P}}^2, \end{aligned} \quad (4.17)$$

and a full derivation can be found in [173]. The quantity is the variance of the Shannon information and therefore a higher order statistic than the KL divergence. It is only weakly prior dependent and like the KL divergence is additive for independent parameters and invariant under a change of variables.

## 4.4 Density Estimators in Practice

In principle, you can use any type of density estimator to model subspaces in a larger parameter space and calculate the marginal Bayesian statistics discussed. The only requirements are that it can be used to resample the space and approximate the normalised log-probability associated with the posterior subspace.

### 4.4.1 Masked Autoregressive Flows

Complex densities are best estimated using expressive neural networks that have been trained to transform between some base distribution, often a standard normal, and the target probability distribution, e.g. the posterior from a Bayesian Nested Sampling run. MARGARINE uses Masked Autoregressive Flows (MAF) [169], which are discussed extensively in the introduction of this thesis, to perform this transformation.

To improve the accuracy of the density estimation, we can transform the target posterior samples that we input to our MAF into a Gaussian parameter space via the standard normal cumulative distribution function (CDF). The target distribution in this normalised space is a skewed and scaled version of the base distribution  $z_0$  which makes the transformation easier to learn.

Due to the invariance of the KL divergence and, similarly, the BMD under a change of variables, we can calculate their values in any transformed version of the original parameter space provided the prior and posterior undergo the same change of variables. In practice, MARGARINE calculates these quantities in the gaussianized space. A uniform prior in the gaussianized space corresponds to the standard normal distribution. For more complex priors, we can transform them into this space in the same way we transform the posteriors and train a prior MAF to calculate the marginal Bayesian statistics.

There are a number of tunable parameters involved when designing MAFs including the number of networks, the number of layers in each network, the number of epochs and the learning rate. All of these parameters can be explored using MARGARINE, and we use recommendations from a previous implementation [172].

### 4.4.2 Kernel Density Estimation

A simple 1D Kernel Density Estimator (KDE) is defined by

$$\mathcal{P}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (4.18)$$

where  $h$  is known as the bandwidth,  $K$  is the kernel and  $n$  the number of samples in the set  $x$ . This scales to higher dimensions where  $x$  and  $x_i$  become  $N$  dimensional vectors and  $h$  becomes a 2D matrix of bandwidths. We use a multivariate Gaussian kernel where  $h$  is equivalent to the covariance matrix and  $x_i$  is equivalent to an  $N$  dimensional vector of means. Since the KDE is a sum of known kernels with a known covariance matrix and set of means, then the probability distribution of the samples is trivially calculable, as is its logarithm.

When generating KDEs, we perform the same type of normalisation of our target distribution as is done with the MAFs, transforming it into a Gaussianised space via a standard normal CDF. The transformation allows the density estimator to better capture the edges of very flat and uniform posteriors.

We can generate samples from trained KDEs and calculate marginal Bayesian statistics, however KDEs are not strictly bijective. Adapting a KDE to be bijective is a non-linear process and requires the use of a root-finding algorithm to transform from the hypercube to the KDE space. This is a process that is not currently optimized, but is fully implemented in `MARGARINE` allowing the user to generate KDE priors.

To transform the unit hypercube into samples on the target distribution via the KDE we first note that KDEs model the target probability distribution by breaking it down into the product of conditional probabilities

$$P(x_1, x_2, \dots, x_n) = P(x_1)P(x_2|x_1)\dots P(x_n|x_1, x_2, \dots, x_{n-1}) \quad (4.19)$$

modelled by a Gaussian kernel at each data point. We then use the inverse CDFs of each Gaussian conditional probability to transform samples on the hypercube to samples on the target distribution

$$(u_1, u_2, \dots, u_n) \rightarrow (x_1, x_2, \dots, x_n), \quad (4.20)$$

via

$$\begin{aligned} x_1 &= F_1(u_1) \\ x_2 &= F_2(u_2; x_1) \\ &\vdots \\ x_n &= F_n(u_n; x_1, x_2, \dots, x_{n-1}) \end{aligned} \quad (4.21)$$

where  $F_i$  are the conditional and marginalised inverse CDFs.

For a multivariate Gaussian KDE, marginalisation is equivalent to ignoring components of  $h$  and  $x_i$  in the multivariate equivalent of equation (4.18). Conditioning the probability distribution simply involves analytical corrections to the means and standard deviations based on the corresponding values for the  $n - 1$  distributions. The equivalent transformation using MAFs is much more computationally efficient. There are fewer tunable parameters for the KDE, however we can change the bandwidth of the kernel. We use the Silverman [176] approach to calculate  $h$  but note that this can be modified with `MARGARINE`.

## 4.5 Applications of MARGARINE

### 4.5.1 Nuisance-Free Likelihood

#### 4.5.1.1 Toy 21-cm Example

To illustrate the application of the nuisance free likelihood function, we use an example from 21-cm cosmology. To a crude approximation, the 21-cm signal can be modelled with a Gaussian absorption profile. We generate two different mock sky-averaged 21-cm data sets with different bandwidths of 50 – 100 and 75 – 190 MHz with different foregrounds to simulate observations from different locations. We call these two experiments *A* and *B* and model the foreground as

$$T_{\text{fg}} = a\nu^{-\beta}, \quad (4.22)$$

where  $a$  is a common scaling factor for both experiments and  $\beta$  is set as  $-2.6$  and  $-2.5$ . The model is based on the expectation that the foreground is composed of smooth synchrotron and free-free emission and the power law is based on observational constraints from experiments like EDGES [55], SARAS3 [54], and LEDA [56].

For simplicity, we assume an achromatic beam for each experiment and thus a data set that includes a smooth foreground and Gaussian 21-cm signal. The signal in our mock data is given by

$$T_{21} = -T_{\min} \exp \left( -\frac{(\nu - \nu_c)^2}{2\Delta\nu^2} \right), \quad (4.23)$$

where  $T_{\min}$  is the signal amplitude,  $\nu$  is the frequency range of the data,  $\nu_c$  is the central frequency of the absorption feature and  $\Delta\nu$  is the signal's width. We set  $T_{\min} = 0.25$  K,  $\nu_c = 80$  MHz and  $\Delta\nu = 10$  MHz.

We explore each likelihood separately with the Nested Sampling algorithm **POLYCHORD**, equation (4.23) and a log-log polynomial foreground model given by

$$\log_{10}(T_{\text{fg}}) = \sum_{i=0}^N a_i \log_{10}(\nu)^i, \quad (4.24)$$

where  $a_i$  are coefficients to be fitted. In practice, the foreground modelling is not always consistent across different data sets and to further emphasise that the two mock experiments are observing different parts of the sky, we assume they have different complexities and fit a 3-term polynomial to experiment *A* and a 4-term polynomial to experiment *B* for the foreground. Generally, when polynomials are being used, the order of the polynomial would be optimised through the assessment of the Bayesian evidence. We inject Gaussian random noise into our data sets with standard deviations of 35 mK and 15 mK for experiments *A* and *B* respectively.

In our Nested Sampling runs, we use a Gaussian log-likelihood function

$$\log \mathcal{L} = \sum_i \left( -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \left( \frac{T_D(\nu_i) - T_{\text{fg}}(\nu_i) - T_{21}(\nu_i)}{\sigma} \right)^2 \right), \quad (4.25)$$

where  $T_D$  corresponds to the mock data and  $\sigma$  corresponds to the instrument noise and is a modelled parameter. In the top panels of Fig. 4.3 we show the sum of the noise and signal models that went into the mock datasets and the functional averages of our fitted 21-cm signals

$$\overline{T_{21}} = \frac{\sum_i^{N_{\text{samples}}} w_i T_{21}(\theta_{21,i})}{\sum_i^{N_{\text{samples}}} w_i} \quad (4.26)$$

where  $w_i$  is the weight corresponding to sample  $i$  output by **POLYCHORD**.

We also show in the center right panel of Fig. 4.3 the resultant  $\overline{T_{21}}$  found fitting both data sets simultaneously with

$$\log \mathcal{L}_{A,B}(\theta, \alpha_A, \alpha_B) = \log \mathcal{L}_A(\theta, \alpha_A) + \log \mathcal{L}_B(\theta, \alpha_B). \quad (4.27)$$

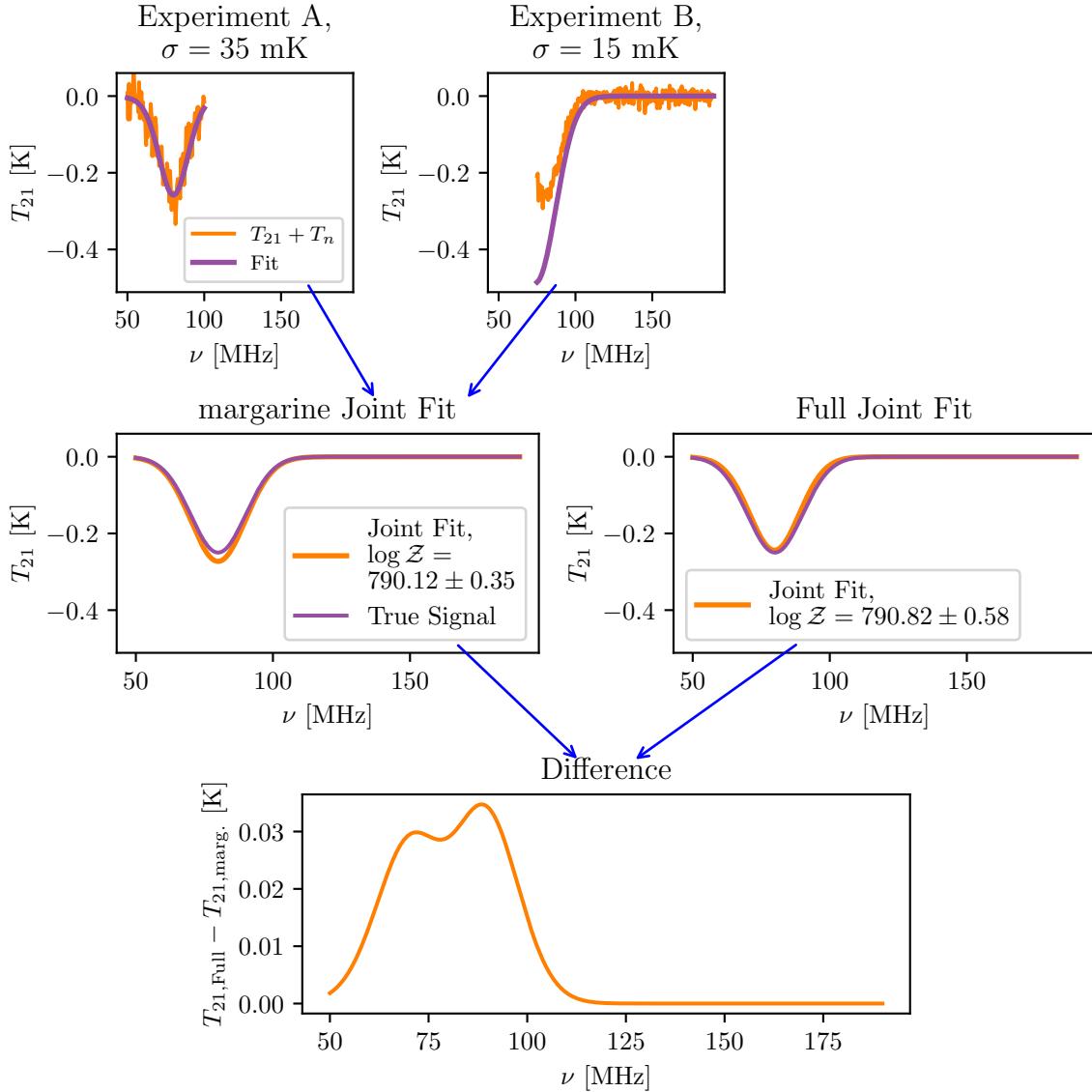
Through the combination of the two data sets we can see by visual inspection that we get a much better fit to the data than is achieved for each experiment individually as is to be expected.

However, in order to do this we have had to sample over the nuisance parameters,  $\alpha_A$  and  $\alpha_B$  (e.g. Fig. 4.1). Alternatively, with **MARGARINE** we can generate nuisance-free likelihood functions for each experiment and sample

$$\log \mathcal{L}_{A,B}(\theta) = \log \mathcal{L}_A(\theta) + \log \mathcal{L}_B(\theta), \quad (4.28)$$

over just the parameters of our 21-cm model without having to consider modelling the foregrounds (e.g. Fig. 4.2).

We show the resultant  $\overline{T_{21}}$  from the joint analysis with **MARGARINE** in the center left panel of Fig. 4.3 and we can see that the fit is consistent with the more complex analysis. In particular, we note the approximate consistency between the Bayesian evidences for each fit. The error on the evidence for the **MARGARINE** fit is



**Figure 4.3:** The top panels in the graph shows two simulated data sets comprising noise and a 21-cm signal as orange lines. In the same panels in purple, we show the averages of the functional posteriors derived from Nested Sampling fits to both of these data sets. Neither fit recovers the true signal exactly, however when we analytically combine the full likelihoods for each experiment, including foregrounds, and jointly fit the data we get a much better agreement across the whole frequency range as can be seen in the center right panel. We show the results of this joint analysis as a purple line. For comparison, we show in the center left panel, as an orange line, the average of the functional posterior from a joint fit performed with `MARGARINE` using the nuisance-free likelihood function. We can see that the joint fits are equivalent, more accurate than the fits to each individual experiment and the Bayesian evidences are similar in magnitude when using the full likelihood and the nuisance-free likelihood. We show the difference between the two fits in the bottom panel as a function of frequency.

likely underestimated due to uncertainty in the density estimation, however this is currently hard to quantify and left for future work.

In this case because the problem is trivial sampling over  $\theta$ ,  $\alpha_A$  and  $\alpha_B$  is not an inefficient or complex process, however generally speaking there are many more nuisance parameters that need to be modelled and the complexity of those models can significantly increase the evaluation time per call to the likelihood [e.g. chapter 5, 6]. In these circumstances, MARGARINE can act as a useful tool for efficiently combining constraints from different data sets [e.g. chapter 7, 2] and this is discussed further in the next section.

#### 4.5.1.2 Combining constraints from Planck and DES

Here we combine samples from the Dark Energy Survey (DES) Year 1 posterior [166] and Planck posterior [22] using MARGARINE to estimate nuisance-free likelihoods. DES surveys supernovae, galaxies and large scale cosmic structures in the universe in an effort to measure dark matter and dark energy densities and model the dark energy equation of state. In contrast, Planck mapped the anisotropies in the Cosmic Microwave Background (CMB) and correspondingly provided constraints on key cosmological parameters.

The constraints from DES and Planck have previously been combined using a full Nested Sampling run over all parameters including a multitude of ‘nuisance’ parameters in a computationally expensive exercise [174] (e.g. Fig. 4.1). The previous analysis gives us access to the combined DES and Planck evidence, which is found to have a value of  $\log(\mathcal{Z}) = -5965.7 \pm 0.3$ . In Fig. 4.4 we show the DES, Planck and joint posteriors for the six cosmological parameters derived in this work using MARGARINE and the flow chart in Fig. 4.2. The constrained parameters are the baryon and dark matter density parameters,  $\Omega_b h^2$  and  $\Omega_c h^2$ , the angular size of the sound horizon at recombination,  $\theta_{MC}$ , the CMB optical depth,  $\tau$ , the amplitude of the power spectrum,  $A_s$ , and the corresponding spectral index,  $n_s$ . These make up the set  $\theta = (\Omega_b h^2, \Omega_c h^2, \theta_{MC}, \tau, A_s, n_s)$ . We use the Nested Sampling algorithm POLYCHORD in our analysis [88, 89].

We use a uniform prior that is defined to be three sigmas around the Planck posterior mean. This is done to improve the efficiency of our Nested Sampling run. However, we subsequently have to re-weight the samples and correct the evidence for the difference between the priors used here and in the previous full Nested Sampling run [174] for comparison. If we define

$$\mathcal{Z}_A = \int \mathcal{L}(\theta) \pi_A(\theta) d\theta, \quad \mathcal{Z}_B = \int \mathcal{L}(\theta) \pi_B(\theta) d\theta, \quad (4.29)$$

where  $A$  is our uniform prior space and  $B$  is our target prior space from the previous work [174], then

$$\mathcal{Z}_B = \int \mathcal{L}(\theta) \pi_B(\theta) d\theta = \int \mathcal{L}(\theta) \pi_A(\theta) \frac{\pi_B(\theta)}{\pi_A(\theta)} d\theta = \mathcal{Z}_A \left\langle \frac{\pi_B(\theta)}{\pi_A(\theta)} \right\rangle_{\mathcal{P}_A} \quad (4.30)$$

giving

$$\mathcal{Z}_B = \mathcal{Z}_A \left\langle \frac{\pi_B(\theta)}{\pi_A(\theta)} \right\rangle_{\mathcal{P}_A}. \quad (4.31)$$

Then following similar arguments we can transform our posteriors by re-weighting the distributions with the following

$$w_B^{(i)} = w_A^{(i)} \frac{\pi_B(\theta^{(i)})}{\pi_A(\theta^{(i)})}. \quad (4.32)$$

We see from the figure and corresponding table that with our joint analysis we are able to derive a log-evidence that is approximately consistent with that found in [174],  $\log(\mathcal{Z}) = -5965.7 \pm 0.3$  as previously reported, further validating the theory discussed and its implementation with MARGARINE. Using MARGARINE

we can also derive the combined KL divergence, also reported in Fig. 4.4 and discussed further in this chapter, which we find is consistent with the result in the literature of  $\mathcal{D} = 6.17 \pm 0.36$  [173]. Similarly, we derive marginal KL divergences for the DES and Planck cosmological parameters using MARGARINE (the DES example is furthered in the following subsections). A full discussion of the implications of combining the two data sets for our understanding of cosmology can be found in the literature [e.g 173, 174].

By reducing the number of parameters that need to be sampled, we significantly reduce the Nested Sampling runtime. For POLYCHORD the runtime scales as the cube of the number of dimensions [177]. This can be seen by assessing the time complexity of the algorithm where,  $T \propto n_{\text{live}} \times \langle T\{\mathcal{L}(\theta)\} \rangle \times \langle T\{\text{Impl.}\} \rangle \times \mathcal{D}(\mathcal{P}||\pi)$ . Here  $n_{\text{live}}$  scales with the number of dimensions,  $d$ , as does the Kullback-Leibler divergence. For POLYCHORD, the specific implementation time complexity factor,  $\langle T\{\text{Impl.}\} \rangle$ , representing the impact of replacing dead points with higher likelihood live points on the runtime, scales linearly with  $d$ . Together this gives  $T \propto d^3 \times \langle T\{\mathcal{L}(\theta)\} \rangle$ . Therefore, by using nuisance-free likelihoods and sampling over 6 parameters rather than 41 parameters (cosmological plus 20 nuisance parameters for DES and 15 different nuisance parameters for Planck) we reduce the runtime by a factor of  $(41/6)^3 \approx 319$  with further improvements in  $\langle T\{\mathcal{L}(\theta)\} \rangle$ .

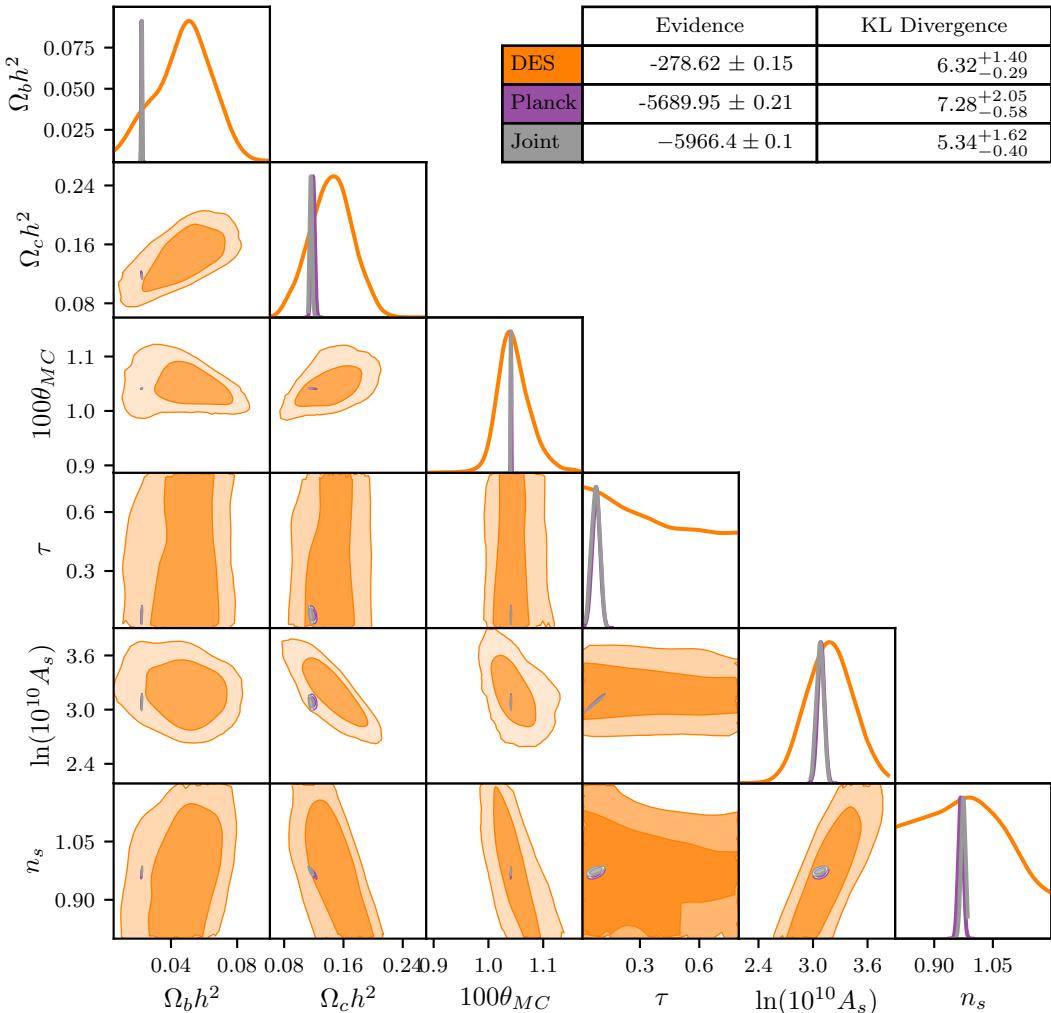
### 4.5.2 Prior Emulation

Since the 21-cm signal is dependent on many different astrophysical process, semi-numerical simulations can take several hours to produce a single model. This is impractical if we want to use Bayesian inference techniques like MCMC and Nested Sampling algorithms to fit real data. Signal emulators, like those discussed in the previous chapter, are often used as a fast alternative that can accurately model the relationship between the astrophysics and the signal structure and produce signal realisations in 10s of milliseconds. An even cheaper alternative is to approximate the signal with a Gaussian absorption feature, as alluded to in the previous section. However, it is not always clear how to set the priors on the parameters  $T_{\min}$ ,  $\nu_c$  and  $\Delta\nu$  in equation (4.23). For this, we can use MARGARINE.

The idea here is to take a large set of physically motivated signals generated using an emulator from a uniformly sampled prior on the astrophysical parameters describing the underlying semi-numerical simulation. For each physical signal we then calculate the depth corresponding to  $T_{\min}$ , the central frequency  $\nu_c$  and approximate the width  $\Delta\nu$ . Through this process we arrive at a physically motivated set of samples in the phenomenological parameters which we can train a MAF on using MARGARINE, creating a physically motivated prior on our phenomenological parameters.

We use the global 21-cm signal emulator GLOBALEMU [4], discussed in chapter 3, to model a set of 5,000 signals based on a parameterisation of the signal that includes;  $V_c$  the minimum virial circular velocity that is proportional to the cube root of the halo mass for star formation,  $f_*$  the star formation efficiency,  $f_X$  is the X-ray production efficiency which is proportional to the X-ray luminosity,  $\tau$  the CMB optical depth and  $f_r$  the radio production efficiency proportional to the radio luminosity [121]. Whilst 5,000 signals may seem like a large number of models, the equivalent number generated in a Nested Sampling or MCMC run would be far larger.

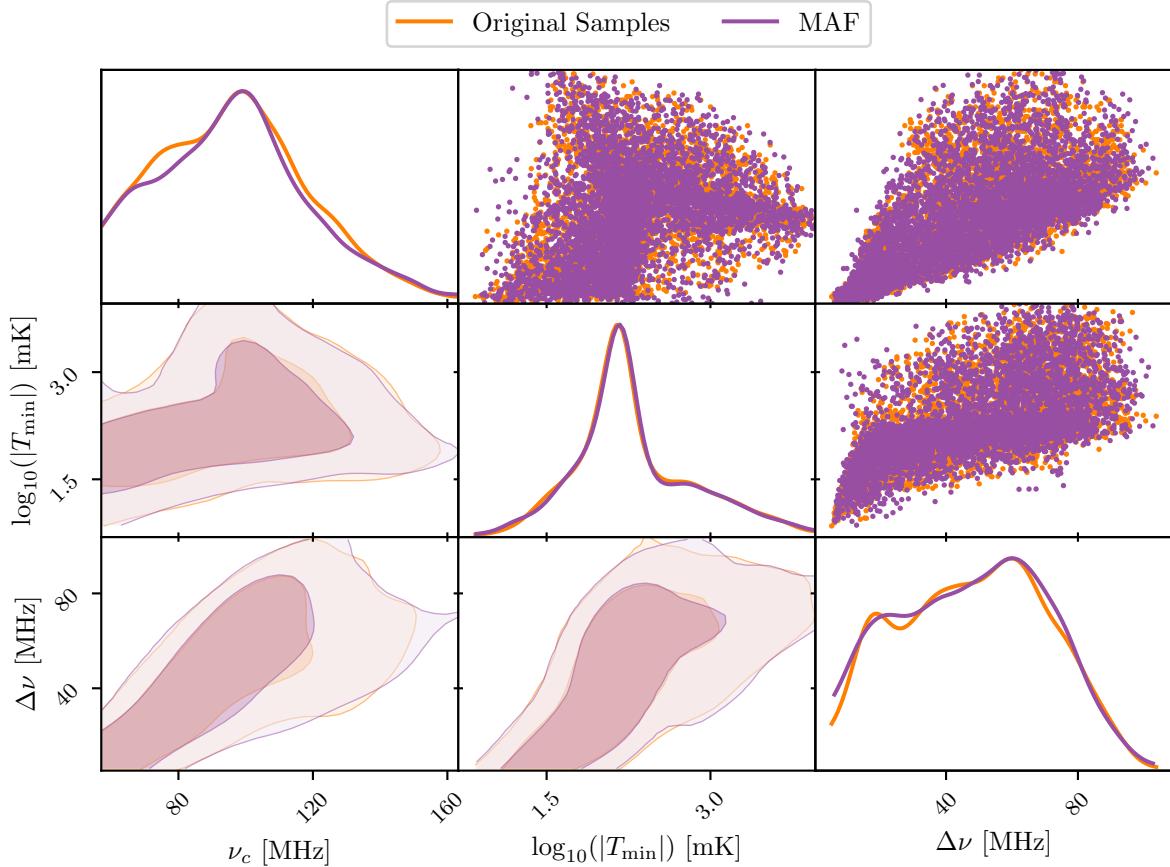
In scenarios with poor star formation, corresponding to high  $V_c$  and low  $f_*$ , and large X-ray heating corresponding to high  $f_X$  we find that the signal cannot be well approximated by a Gaussian profile because it features a weak absorption feature and strong emission. We therefore filter out these scenarios before training our MAF and of the 5,000 models we initially started with, 87% are used to produce our physically motivated prior. In Fig. 4.5 we show the distribution on  $\log_{10}(|T_{\min}|)$ ,  $\nu_c$  and  $\Delta\nu$  generated from the physical models and



**Figure 4.4:** The combined posterior (in grey) found when combining the constraints on the cosmological parameters from DES and Planck using MARGARINE. For DES and Planck, we calculate the marginal KL divergences using MARGARINE, whereas the Bayesian evidences are calculated using ANESTHETIC. The joint evidence and joint KL divergence are calculated with a combination of the two codes and are found to be approximately consistent with those found in the literature [173, 174]. Note that the error on the joint evidence is likely underestimated as it relies on evaluations of log-probabilities for various distributions, for which MARGARINE does not currently provide errors. The figure produced with ANESTHETIC [178].

samples from the corresponding MAF which can be used as a prior on the Gaussian model.

In 21-cm cosmology, instrumental effects can introduce non-smooth structure to the data. These effects are often sinusoidal [e.g. 3, 6] and if not corrected for they can affect the recovery of the 21-cm signal. For example, part of a damped sinusoidal structure in the data with a periodicity of 5 MHz can be modelled with the approximate Gaussian signal profile discussed in this chapter if the prior on the width is wide. However, from Fig. 4.5 we see that a signal profile of width 5 MHz is not likely to represent a physical scenario. The goal of the physically motivated prior is to therefore provide some information about the anticipated signal that can help to separate it from and prevent fitting systematic structures in the data when using a Gaussian profile for the signal.



**Figure 4.5:** A physically motivated prior on the phenomenological parameters of a Gaussian absorption feature used to model the sky-averaged 21-cm signal.  $\nu_c$  is the central frequency of the absorption feature,  $T_{\min}$  is the corresponding temperature, and  $\Delta\nu$  is the approximate width of the 21-cm signal. The prior is derived by approximating the phenomenological parameters for a set of physically motivated signals generated with a neural network emulator and learning the corresponding distribution, in orange, with a MAF. Samples from the MAF are shown in purple.

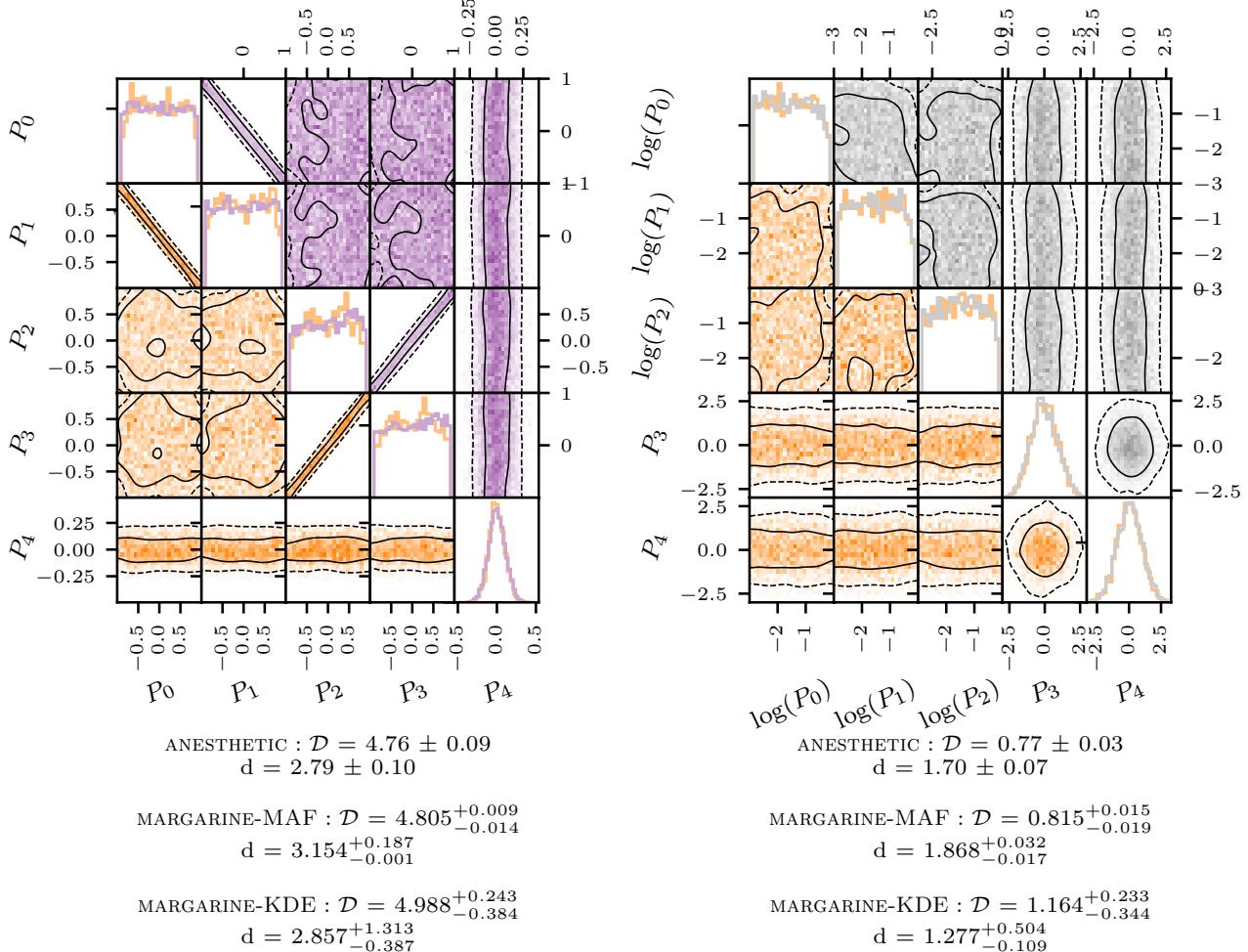
### 4.5.3 Marginal Bayesian Statistics

#### 4.5.3.1 Toy Example Posteriors

We use two different toy examples for which we know the KL divergence and BMD to demonstrate the calculation of marginal Bayesian statistics. We sample the example Gaussian likelihoods, both with five parameters, using POLYCHORD, and we show the corresponding distributions in orange in Fig. 4.6. The first has a series of correlations between the parameters (left-hand panel) and the second has a combination of uncorrelated parameters with both log-uniform and uniform priors on the parameters (right-hand panel).

We are able to use the samples from POLYCHORD and the analysis tool ANESTHETIC [178] to calculate the KL divergence and BMD for both toy examples and these values are reported, with errors, in Fig. 4.6. We show samples drawn from a KDE trained on the correlated samples and from a MAF trained on the uncorrelated samples, with values of  $\mathcal{D}$  and  $d$  listed for both types of density estimator.

The errors in  $\mathcal{D}$  and  $d$  are estimated by propagating both samples generated with the density estimator and the original samples through the density estimator to evaluate the log-probabilities and comparing the resultant statistics. If the density estimator is a perfect representation of the original distribution, then we would expect samples drawn from it to be from the same distribution as the original samples. This would lead to an approximate equivalence between the two sorted log-probability distributions as functions of the sorted



**Figure 4.6: Left Panel:** The graph shows a set of correlated nested samples from POLYCHORD in orange along with a reconstruction from a trained KDE in purple. Listed are the corresponding ‘true’ values for the KL divergence and BMD from ANESTHETIC and the estimated values from MARGARINE using both a KDE and MAF. **Right Panel:** An equivalent graph for a set of uncorrelated samples with both log-uniform and uniform priors on the parameters, shown in orange. The gray samples are taken from a trained MAF, and we report the ‘true’ Bayesian statistics values from ANESTHETIC along with estimates calculated using a MAF and KDE with MARGARINE.

associated Nested Sampling weights [179] and so any deviation we see gives us an indication of the error in our marginal statistics.

For the correlated samples, we see that the estimated KL divergence and BMD from the MAF and from the KDE are in close agreement with the ‘true’ value output from ANESTHETIC. For the uncorrelated samples, we also see that the MAF KL divergence and BMD estimates are in agreement with the ANESTHETIC values. For the KDE, the KL divergence and BMD are less consistent with the values from the MAF and from ANESTHETIC, however we find we have larger errors when using the KDEs likely because they are less expressive than the MAFs.

Typically, the upper and lower bounded range on the BMD estimates are larger than for the KL divergence estimates because it has a more complex dependence on  $\log \mathcal{L}$ . Future work is needed to investigate whether improvements can be made to the BMD estimates by modifying the MAF loss function, network configurations or KDE bandwidths among other tunable parameters.

### 4.5.3.2 Radio Experiment for the Analysis of Cosmic Hydrogen

The Radio Experiment for the Analysis of Cosmic Hydrogen (REACH) data analysis pipeline [57] uses Nested Sampling, implemented with `POLYCHORD`, to model the different components of the data. The pipeline has been extensively tested on mock observations [57, 58, 74, 79, 116]. We generate mock observational data using a realistic foreground model derived from an all-sky map and inject a Gaussian signal profile with an amplitude of  $|T_{\min}| = 155$  mK, a central frequency of  $\nu_c = 85$  MHz and a standard deviation of  $\Delta\nu = 15$  MHz.

The mock data correspond to a single snapshot observation taken from the Karoo radio observatory with a dipole antenna and modelled with a foreground model that takes advantage of the spectral structure of the sky, a correction for the non-uniform response of the antenna to the sky, Gaussian noise and a Gaussian signal profile corresponding to a 16 dimensional parameter space. In the left-hand panel of Fig. 4.7 we show the posterior distribution for the signal parameters from our fit in orange, having marginalised over the other 13 parameters, and the corresponding KDE reconstruction from `MARGARINE` is shown in purple. We see a reasonable consistency between the marginal KL divergence calculated for these samples when using both the KDE and MAF.

However, we note that there are some differences in the BMD estimate, with the MAF giving a larger value for  $d$ . From a visual inspection of the samples we would expect the BMD to be around 1-2 due to the tight constraint on  $\nu_c$  and weaker but still non-trivial constraint on  $|T_{\min}|$  meaning that both estimates are somewhat consistent with expectations.

### 4.5.3.3 Dark Energy Survey

The Year 1 DES analysis<sup>a</sup> [174], as discussed, aimed to constrain the baryon density parameter,  $\Omega_b$ , the dark matter density parameter,  $\Omega_c$ , the approximate ratio of the sound horizon to the angular diameter distance,  $\theta_{MC}$ , the optical depth of the CMB to reionization,  $\tau$  and the amplitude and tilt of the power spectrum,  $A_s$  and  $n_s$ . The samples from the cosmological subspace are shown in orange in the right-hand panel of Fig. 4.7, where we have marginalised over a set of 20 nuisance parameters, along with a MAF reconstruction of the subspace in gray.

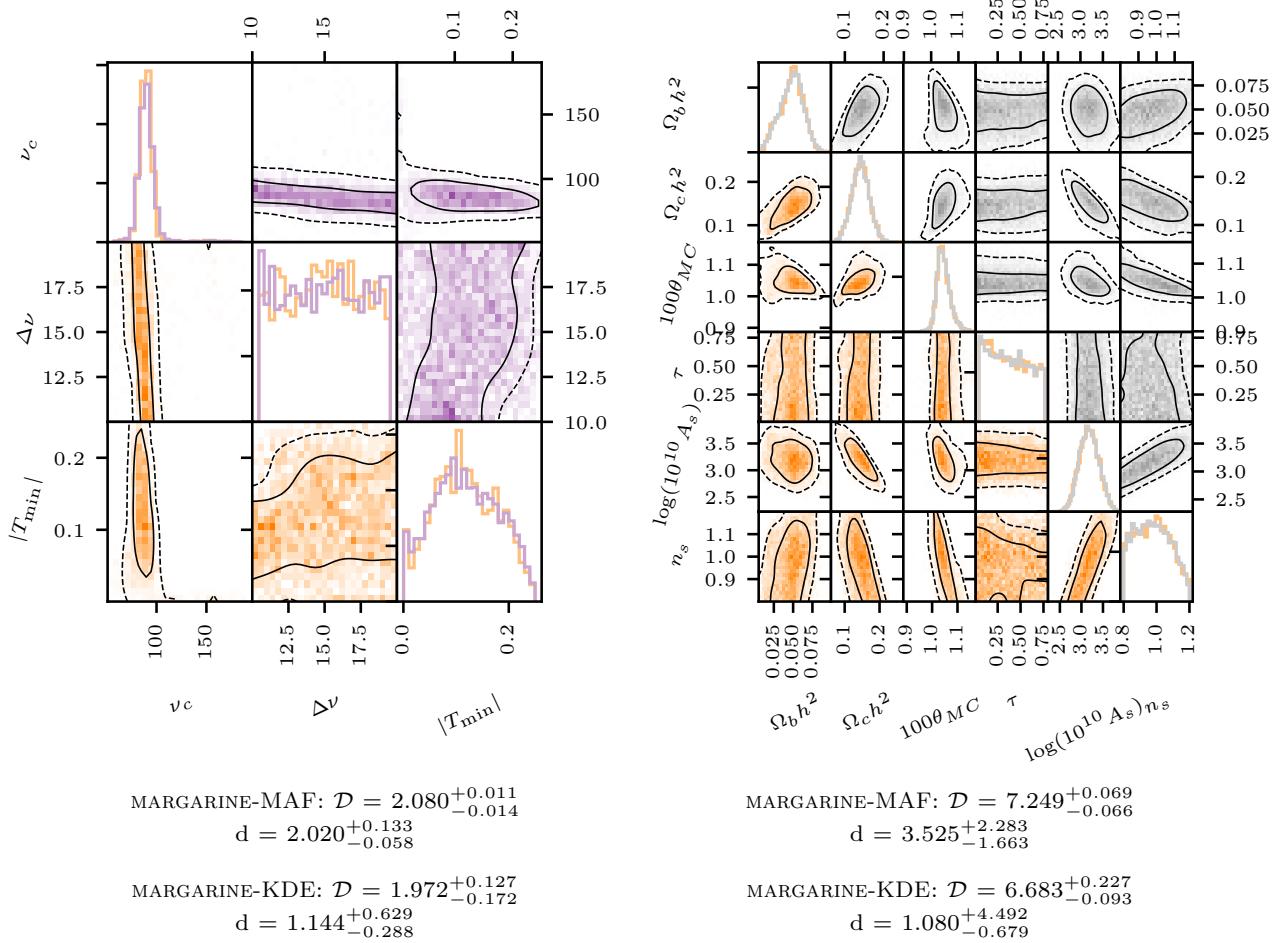
Unlike the toy examples and REACH samples, the DES cosmological samples do not have uniform priors and cannot easily be transformed into a space in which the priors are uniform. As a result, we have to use the density estimators built into `MARGARINE` to evaluate both the log-probability of the posterior and the prior if we want to calculate marginal statistics.

While this is possible, it is expected to lead to an increased uncertainty in the marginal statistics, which can be seen in Fig. 4.7. In addition, the problem is further complicated by the size of the cosmological parameter space, since we expect larger parameter spaces to be harder to replicate well with `MARGARINE`, and consequently we expect larger errors on the marginal statistics.

Again the density estimators give us different estimates for the BMD, however the errors are large. From a visual examination of the distribution, we would expect the BMD to be between  $\approx 3$  and  $\approx 4$ . There is a reasonable agreement between the KL divergence calculated for the cosmological constraints from DES with both the MAF and the KDE. We note that the KL divergence reported here for the DES samples is different from that reported in Fig. 4.4 but that the two estimates are consistent with each other. The difference stems from the fact that in each analysis we retrained the DES MAF.

---

<sup>a</sup>We note that the Year 3 DES results have been published [167, 168] but that the samples have not been made publicly available.



**Figure 4.7: Left Panel:** The signal subspace, in orange, for a fit to simulated observations of the 21-cm signal with the REACH experiment along with samples, in purple, from KDE trained on the three signal parameters effectively marginalising over the other thirteen parameters in the fit. We report the marginal KL divergence and BMD found for this set of three parameters using both a MAF and KDE. **Right Panel:** The cosmological subspace for the Year 1 DES samples shown in orange, along with a set of samples taken from a trained MAF in gray. Again, we report the corresponding marginal Bayesian statistics calculated with MARGARINE.

## 4.6 Conclusions

In this chapter we have demonstrated a number of applications of two different types of density estimator, Masked Autoregressive Flows and Kernel Density Estimators, to the calculation of marginal Bayesian statistics, the efficient modelling of multiple data sets and the derivation and emulation of physical priors. MARGARINE is a multipurpose tool that can be used to enhance our Bayesian analysis workflows.

The evaluation of marginal KL divergences and Bayesian Model Dimensionalities allows for improved comparison of the constraining power of different experiments targeting the same astrophysical or cosmological signals with different systematics or nuisance parameters. In turn, this can lead to improvements in experimental design with techniques that provide more information about the signals of interest being pursued in the future. The calculation of marginal Bayesian statistics has been illustrated in [74], [59] and chapter 6 [7].

The nuisance-free likelihood function allows for more efficient combination of constraints from different data sets, preventing the need to sample instrument specific foreground and systematic effects in joint analysis. This was demonstrated with data from the Planck and Dark Energy Survey experiments, and is explored further with data from the 21-cm power spectrum experiment HERA and sky-averaged 21-cm experiment SARAS3 in chapter 7 [2].

Finally, MARGARINE can be used to generate non-trivial priors (e.g. [172]) either from experimental results or simulations, as illustrated in this chapter. In principle, this allows us to inform the analysis of data from upcoming probes like REACH [10] or the Simons Observatory [180, 181] with results from existing experiments like SARAS3 [7] and Planck [22]. We anticipate further development of MARGARINE and additional unexpected applications that evolve as the code is used.

## **Part III**

# **Applications to Real Data**



In this section, we apply the tools outlined in the previous chapters to real data from the global 21-cm signal experiments SARAS2 and SARAS3 and the power spectrum experiment HERA. The work builds on the current experimental results outlined in section 1.6 and makes significant use of Bayesian analysis techniques outlined in section 1.4.

In chapter 5, we analyse data from the SARAS2 instrument in the redshift range  $z \approx 7 - 12$  corresponding to the epoch of reionization. We identify the presence of a sinusoidal systematic in the data and model this alongside a maximally smooth foreground (see chapter 2) and several different signal models which are emulated with GLOBALEMU (see chapter 3). The work is based on the first author paper [6] published in Monthly Notices of the Royal Astronomical Society.

We follow this with an analysis of data from the SARAS3 instrument corresponding to  $z \approx 15 - 25$  and the cosmic dawn when the first stars form. In this work, we produce the best constraints to date on the properties of the first stars using emulation of semi-numerical signals with GLOBALEMU. This chapter is based on the first author paper [7] which was published in Nature Astronomy.

Finally, in chapter 7, we demonstrate, using MARGARINE (see chapter 4), how a combined analysis of data from the power spectrum experiment HERA and data from the global signal experiment SARAS3 can lead to improved constraints in comparison to analysis of the individual data sets. This chapter is based on [2].



CHAPTER

5

**CONSTRAINTS ON THE PROPERTIES OF GALAXIES  
BETWEEN  $z \approx 7 - 12$  WITH SARAS2**

## 5.1 Introduction

Several experiments, both single antenna and interferometers, have provided constraints on the parameter space of the 21-cm signal at redshifts corresponding to the EoR including HERA [109], LOFAR [92, 105, 106], MWA [93, 108], EDGES [91, 133, 134] and SARAS2 [69, 70]. We note that the parameterisation and modelling of the signals, as well as the prior ranges, are not always consistent across the literature. However, in general the conclusions disfavour signals with deep absorption features, within the band of each instrument, from inefficient X-ray heating and a sharp reionization feature.

In this chapter, we present a re-analysis of the SARAS2 data, which targeted the EoR at low redshifts (high frequencies). Previous analysis of 63 hrs of nighttime observations, between October 2016 and July 2017, at the Timbaktu Collective in Southern India with the SARAS2 instrument concluded that scenarios with rapid reionization and weak X-ray heating were disfavoured by the data [69, 70]. In this analysis, the authors used a Bayesian likelihood ratio test to determine whether the presence of particular signal models, from a simulated set of 264, were favoured in the data or not [69]. This was followed by a detailed frequentist approach that ruled out a larger number of simulated signals from the same set of models and using the same data [70]. Of the tested scenarios, 9 were disfavoured by the data in [69] and 20, of which 15 were rejected with a significance  $> 5\sigma$ , were rejected in [70]. There was no reported detection from the analysis. In both cases high order polynomials were used to model the foreground and systematics, in the belief that any present in the data are smooth, in combination. A high level summary of the differences between the analysis in this chapter and the previous analysis of the SARAS2 data can be found in Tab. 5.1 and these are further discussed below.

We determine parameter constraints over broad prior ranges using the latest astrophysical models of the global 21-cm signal [14, 121], representing an improved understanding of the standard astrophysical picture, and the Nested Sampling algorithm [85] POLYCHORD [88, 89]. We use models that include Lyman- $\alpha$  heating [29, 42, 182, 183], CMB heating [44] and multiple scattering of Lyman- $\alpha$  photons [184–189]. The effects on the global signal of these physical processes have been understood for some time, but the magnitude of those effects over a larger parameter space were not understood until recently [14, 43, 45]. Additionally, we study astrophysical scenarios with a wide range of radio production efficiencies,  $f_{\text{radio}}$ , for early galaxies. A subset of the latter models could explain the anomalous EDGES detection [8] using an excess radio background. This is the first time that a full Bayesian analysis of data from a global 21-cm experiment has been performed with these specific astrophysical simulations. We note, however, that the value of  $f_{\text{radio}}$  has previously been constrained using the amplitude of the EDGES absorption feature [121] and more recently using upper limits on the power spectrum from the Hydrogen Epoch of Reionization Array [HERA, 109]<sup>a</sup>.

In this work we use the emulator GLOBALEMU [see chapter 3, 4] which we train on sets of signal models from [121] and [14]. We provide more details on the accuracy of each trained instance of GLOBALEMU in section 5.3.4.

---

<sup>a</sup>The updated HERA constraints presented in [99] were published after the paper associated with this chapter.

	Singh et al. 2017[69]	Singh et al. 2018 [70]	This work
Analysis Type	Likelihood testing preference of the data for the presence or absence of signals.	Frequentist approach based on that used in [133].	Bayesian Nested Sampling using POLYCHORD [88, 89].
Foreground Modelling	Unconstrained polynomials of varying orders (e.g. $N = 4\text{-}8$ in [70]).	Smooth foreground models based on Maximally Smooth Functions and implemented with MAXSMOOTH [3].	Smooth foreground models based on Maximally Smooth Functions and implemented with MAXSMOOTH [3].
Systematic Modelling	Assumed to be smooth and modelled with foreground model.	Identified through use of smooth foreground model and separately modelled with physically motivated functions.	Identified through use of smooth foreground model and separately modelled with physically motivated functions.
Noise Modelling	Gaussian distributed noise based on system attributes.	A set of Gaussian models with constant, frequency damped and relative weights based amplitudes.	A set of Gaussian models with constant, frequency damped and relative weights based amplitudes.
Signal Modelling	A library of 264 standard astrophysical models with no additional radio background above the CMB [13].	Broader study sampling across large prior ranges, for both standard astrophysical models [14] and exotic astrophysical models with excess radio backgrounds [121], using the signal emulator GLOBALEMU [4].	Broader study sampling across large prior ranges, for both standard astrophysical models [14] and exotic astrophysical models with excess radio backgrounds [121], using the signal emulator GLOBALEMU [4].

**Table 5.1:** A high level summary of the differences between the previous analysis of the SARAS2 data and the work performed in this chapter. The differences are expanded on primarily in section 5.3.

We illustrate the presence of a sinusoidal systematic in the data and attempt to physically model the structure in a manner which is independent of the foreground model. We use two separate models, each representing the introduction of a systematic at different points in the SARAS2 experiment. The motivation for each model is explained in section 5.3.3.

The identification of the systematic is driven by the application of MAXSMOOTH [see chapter 2, 3, 190] to model the foreground and smooth systematics in the data with a model that has constrained derivatives and resultant smooth properties based on Maximally Smooth Functions [MSFs, 61, 62]. The motivation behind the use of MAXSMOOTH is two-fold. Firstly, the SARAS2 antenna is designed and has been shown to have a maximally smooth reflection coefficient and efficiencies [16]. Secondly, the dominant foregrounds in global 21-cm experiments from Galactic and extragalactic synchrotron and free-free emitting sources are expected to be smooth power laws [61, 191, 192].

In section 5.2 we discuss briefly the SARAS2 experiment and the data that we are analysing. This is followed by a more detailed description of the modelling that we perform in section 5.3, a discussion about the sensitivity of the data to specific models in section 5.4 and a summary of our results in section 5.5. We draw conclusions in section 5.6.

Eloy de Lera Acedo, Anastasia Fialkov, Will Handley, Saurabh Singh, Ravi Subrahmanyam and Rennan Barkana provided comments on [6] which formed the basis of this chapter.

## 5.2 The SARAS2 data

One of the primary causes of systematics in global 21-cm experiments is chromaticity in the typically very wide beam pattern of the antenna. Sidelobes in the antenna beam, complex reflection coefficients and environmental factors can introduce frequency dependent structures in the data. However, we can attempt to build achromatic smooth antennas and this is the motivation behind the design of the SARAS2 antenna, which is a short monopole designed to have a smooth beam.

In principle, the foreground and systematics in the data from the SARAS2 experiment should both be smooth, and significant efforts were made to ensure that the efficiency and reflection coefficients in the data were smooth functions [16]. This has been explored in [61] where it was shown that simulated observations with the SARAS2 antenna of the foregrounds [produced with the Global Model for the Radio Sky Spectrum, 193] are smooth in structure to within a few mK. This is also expected generally, in the absence of ionospheric effects [65], for an achromatic beam like SARAS2.

SARAS2 was deployed in the remote radio quiet Timbaktu Collective in Southern India ( $\text{lat} = +14.^{\circ}242328$ ,  $\text{long} = 77.^{\circ}612606\text{E}$ ). The antenna comprised of a sphere mounted on top of an inverted cone resting on a circular aluminium disk. The components are smoothly joined and placed above the receiver electronics at the site. The electronics are battery powered and the site is flat and open. An optical fiber is used to connect the receiver to a signal processing unit situated 100 m away.

The beam pattern of the SARAS2 antenna is simulated, measured and shown to be frequency independent [16]. The pattern is omnidirectional and constant in azimuth, with nulls at zenith and horizon, a peak at 30 degrees in elevation and a half-power beam width of 45 degrees in elevation. A 3D visualisation can be found in Fig. 8 of [61].

Noise Model	$\sigma$	Prior	Prior Type
Constant	$A_\sigma$	$A_\sigma = 10^{-3} - 10^{-1}$ mK	Log Uniform
Frequency Damped	$A_\sigma \left( \frac{\nu}{\nu_0} \right)^{-\beta_\sigma}$	$A_\sigma = 10^{-4} - 10^{-1}$ mK	Log Uniform
		$\beta_\sigma = 0 - 5$	Uniform
Relative Weights	$A_\sigma W(\nu)$	$A_\sigma = 10^{-2} - 10^{-1}$ mK	Log Uniform

**Table 5.2:** The tested frequency dependent and independent standard deviation models for the assumed Gaussian noise in the SARAS2 data. In the frequency damped, noise model  $\nu_0$  is the central frequency in the band. The origin of the relative weights,  $W(\nu)$ , is discussed in section 5.3.1.

The antenna temperature, assuming the presence of a global 21-cm signal  $T_{21}$ , would correspond to

$$T_A = (T_{21} + T_{\text{fg}} + T_{\text{gr}})\eta_t, \quad (5.1)$$

where  $T_{\text{fg}}$  accounts for the Galactic and extragalactic foregrounds and  $\eta_t$  corresponds to the total efficiency of the SARAS2 antenna.  $T_{\text{gr}}$  refers to ground emission and for the analysis presented here we assume that the ground emission is smooth or equivalently that the ground under the antenna is homogeneous. As a result, we can subsume the ground emission term into our smooth foreground model and treat the antenna temperature as being given by

$$T_A = (T_{21} + T_{\text{fg}})\eta_t. \quad (5.2)$$

The assumption of a homogeneous ground under the antenna may not hold and that this may cause the introduction of non-smooth systematics into the data (see section 5.3.3). The sum  $T_W = (T_{21} + T_{\text{fg}})$  represents the beam-weighted sky power and  $\eta_t$  is the product of the radiation and reflection efficiency [16].  $\eta_t$  therefore, accounts for the loss due to an impedance mismatch between the antenna and the transmission line to the receiver, as well as the frequency dependent coupling of the beam-weighted sky temperature to the antenna. Estimates of  $\eta_t$  are made using the Global Model for the Radio Sky Spectrum (GMOSS) simulations and measurements of the differential antenna temperature as the sky passes through the beam. The calibration and RFI rejection are detailed in section 6 of [16] and summarised in [69]. We are assuming that the data has been calibrated to be in Kelvin units of antenna temperature and that there is no residual RFI.

The data can be seen in [69, 70] and we discuss the sensitivity of the data to the global 21-cm signal in section 5.4 after introducing the signal models in the following section.

## 5.3 Modelling

The Bayesian Nested Sampling tool **POLYCHORD** [88, 89] is used to fit two different systematic models and two different parametrisations of the global signal to the SARAS2 data. We model the foreground using the software **MAXSMOOTH** [3, 190] and emulate physical models of the global signal with **GLOBALEMU** [4].

In practice, we model the foreground as  $T_{\text{fg}}^* = T_{\text{fg}}\eta_t$ . Throughout the rest of the chapter we generally assume, unless otherwise stated, that when discussing the foreground we are including in that definition  $\eta_t$  and any additive smooth systematics. We consider the addition of non-smooth systematics,  $T_{\text{NS}}$ , into equation (5.2) in section 5.3.3. The details of the different components of our model are given in the following subsections.

We briefly discuss the reproducibility of our results in Appendix A of [6].

### 5.3.1 Noise Modelling

For all the fits performed in this chapter, we assume that the noise in the data is Gaussian distributed and use a Gaussian log-likelihood function

$$\log \mathcal{L} = \sum_i \left( -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \left( \frac{T_A(\nu_i) - T_M(\nu_i)}{\sigma} \right)^2 \right), \quad (5.3)$$

where  $T_M$  stands for the sum of the model components described below. Our assumption is supported by previous analysis of the data, which produced residuals with a Gaussian distribution after removing a high order polynomial model for the structure in the data [see 6, 69].

Typically, the noise is assumed to be frequency independent, however, in practice the noise is dependent on the system temperature which is dominated by the sky temperature and is a function of frequency. In this chapter we consider three different approximations to the standard deviation,  $\sigma$ , of the assumed Gaussian noise, each with a different frequency dependence. The first is a constant value of  $\sigma$  and the second is given by a frequency damped amplitude. The latter comes from the naive expectation that  $\sigma$  should be proportional to,  $T_W$  which means that the standard deviation should decrease with increasing frequency following the trend of the dominant foregrounds. Our third model uses the relative weights,  $W(\nu)$ , for the data which are dependent on the RFI excision, integration time and system temperature [see Fig. 4 in 70]. The noise models are summarised in Tab. 5.2 and we discuss the results when fitting with the proposed models in section 5.5.1. Previous analysis has indicated that the standard deviation of the noise is likely to be constant across the band in the calibrated and sky-averaged data [69, 70].

A detailed study of likelihood and noise modelling in global 21-cm experiments can be found in [79].

### 5.3.2 Foreground Modelling

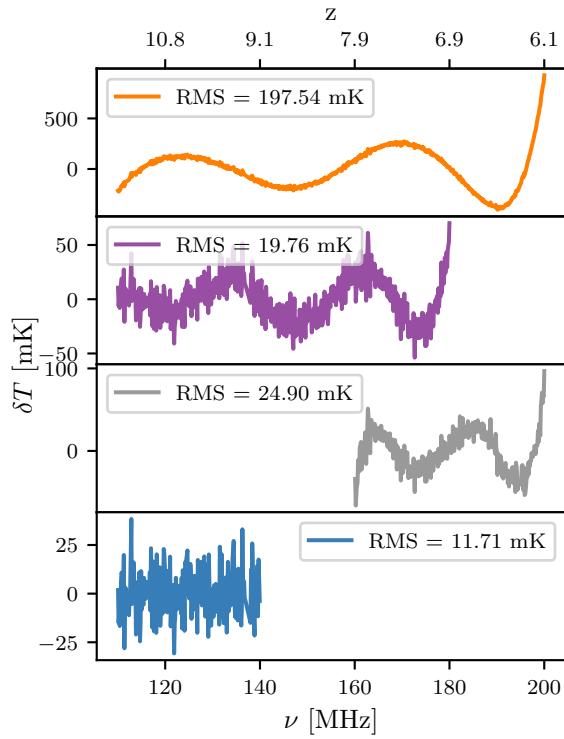
Previously, the foreground in the SARAS2 data set has been modelled in combination with systematics using a high order polynomial [ $N = 4 - 8$ , 69, 70]. However, while a polynomial will fit out both the foregrounds and smooth systematics, it could equally fit out part or all of any global signal and any non-smooth systematics in the data.

We model the foreground and smooth systematics using a variant of an MSF [61, 62] called a Partially Smooth Function [PSFs, 3]. An MSF has derivatives of order greater than or equal to two constrained so that the function does not have any inflection points or higher order non-smooth structure (i.e. the constrained derivatives do not cross zero in the band). PSFs are closely related to MSFs but more general in their definition and can have an arbitrary set of constrained derivatives.

The SARAS2 data has both a turning point and inflection point that can be attributed to the foreground multiplied by  $\eta_t$  [69, 70]. We, therefore, model the PSF foreground with derivatives of order  $m \geq 3$  constrained according to

$$\frac{d^m T_{fg}^*}{d\nu^m} \leq 0 \text{ or } \frac{d^m T_{fg}^*}{d\nu^m} \geq 0, \quad (5.4)$$

with the software `MAXSMOOTH`. This prevents the introduction of high order non-smooth structure into the model but allows the foreground model to fit for a turning point (with  $dT_{fg}^*/d\nu = 0$  at some frequency,  $\nu$ , in the band) and inflection point (with  $d^2 T_{fg}^*/d\nu^2 = 0$ ).



**Figure 5.1:** A comparison of the residuals when fitting the SARAS2 data with a  $10^{th}$  order Partially Smooth Function with derivatives of order  $m \geq 3$  constrained across the full SARAS2 band (orange) and a set of reduced bandwidths. We achieve a significantly lower RMS in the reduced bandwidths. This indicates either a poor foreground fit across the full bandwidth, which can introduce non-smooth structure, or the presence of multiple non-smooth systematics dominant at different frequencies. We proceed to perform our analysis in the reduced bandwidth 110–180 MHz based on the results presented in previous work [16, 69] and in order to retain as much data as possible.

We test the range of built-in MAXSMOOTH foreground models (see chapter 2) and find that

$$T_{\text{fg}}^* = \sum_{k=0}^{N-1} a_k (\nu - \nu_0)^k, \quad (5.5)$$

is the best fitting model with  $N \geq 10$ . Here  $\nu_0$  is the central frequency across the bandwidth. Note that MAXSMOOTH is not a Bayesian algorithm, and the model parameters  $a_k$  for the foreground are not fitted by POLYCHORD. Instead, we wrap MAXSMOOTH inside the call to POLYCHORD and at each sample point MAXSMOOTH fits the foreground parameters,  $a_k$ , to  $T_{\text{A}} - T_{21}\eta_t - T_{\text{NS}}$ .

Fig. 5.1 shows the resultant residuals, in orange, when fitting the data with the PSF model across the whole SARAS2 bandwidth. The residuals are large in magnitude and show a sinusoidal structure, which may be the result of systematics in the data and/or of inaccurate foreground modelling.

In the previous analysis of the SARAS2 data, in which parameter constraints were determined from a discrete set of signal models, the bandwidth used was optimised on a per-signal basis [69, 70]. In that work, it was frequently found that the bandwidth 110 – 180 MHz is the optimum to minimise the signal-to-noise ratio for the 264 tested signal models

In this work, the turning point and inflection point may be significantly distorting the foreground model leading to the introduction of spurious non-smooth structure in the residuals<sup>b</sup>, despite allowing for their presence in the modelling, and so we attempt to fit with a PSF in the reduced range  $\nu = 110 – 180$  MHz effectively

<sup>b</sup>This is unlikely, and it can actually be shown that Partially Smooth Functions can be effectively used to recover the noise in data sets that feature inflection points (see ‘Turning Points and Inflection Points’ in <https://maxsmooth.readthedocs.io/en/latest/maxsmooth.html>).

removing the turning point. We achieve significantly smaller residuals when fitting in this band, as shown in purple in Fig. 5.1. The difference in magnitude may suggest a better quality foreground fit, but it may also indicate the presence of multiple non-smooth systematics in the data, each of which may dominate to a different degree at different frequencies.

A further reduction in the upper bound on the frequency range leads to a further reduction in the RMS. However, we still see the same sinusoidal structure at low frequencies. In practice, we could reduce the band to 110 - 140 MHz, removing both the turning point and inflection point, and we would see a significantly lower RMS ( $\approx 12$  mK when fitting with the proposed PSF model as can be seen in the bottom panel of Fig. 5.1). This is because the systematic structure in the data in this frequency range covers one period of oscillation and is therefore smooth to a sufficient level that it is removed by the foreground model. Further, if we remove the inflection point in the data and fit in the range 160 – 200 MHz we see a sinusoidal structure that is partially consistent with the purple residuals in Fig. 5.1.

Throughout the rest of the chapter, in accordance with the previous SARAS2 analysis, to simplify the modelling and to keep as much data as possible, we use the reduced bandwidth 110 – 180 MHz. The modelling of the non-smooth systematic structure is considered in the following section.

In comparison to the predicted maximum depth of the global 21-cm signal [ $\approx 165$  mK for standard astrophysical signals in the band 110 – 180 MHz, 14], the residual RMS, 19.8 mK, appears small in magnitude. However, as highlighted above and in [69] and [70], any signal in the data will be suppressed by the total efficiency of the antenna. This is discussed further in section 5.3.4.

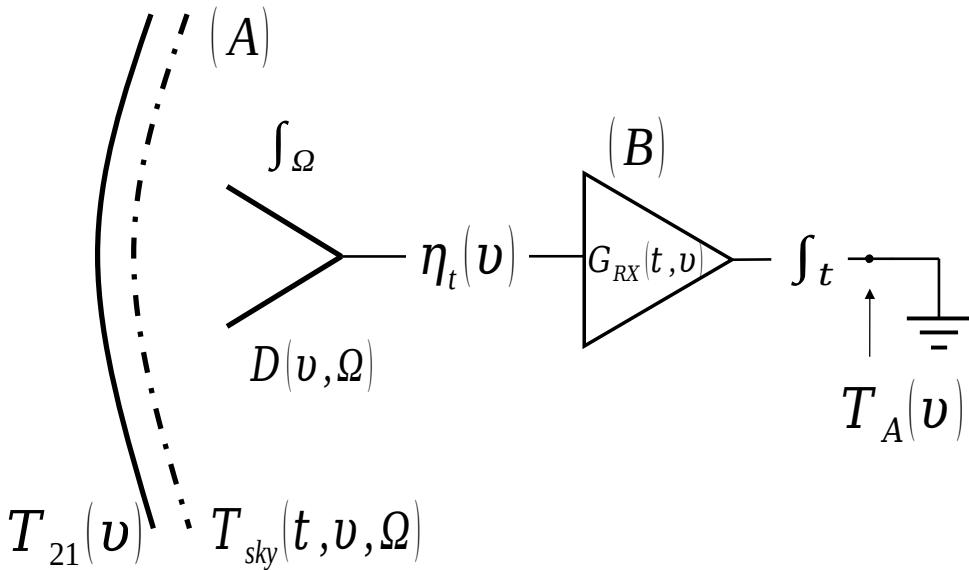
### 5.3.3 Systematic Modelling

Previous analysis has shown the potential presence of sinusoidal and damped sinusoidal systematics in data from several global 21-cm experiments using a variety of different radiometers and analysis techniques. One such example, previously mentioned in the introduction, is the EDGES data. In [128] the authors used a foreground model with spectral index characterised by a 6 term unconstrained polynomial to identify a 60 mK sinusoidal structure in the EDGES data. A subsequent re-analysis of the data in [129] and chapter 2 [3], using MSFs to model the foregrounds, identified a similar sinusoidal structure. In chapter 2 we also re-analysed data from LEDA and identified the presence of a damped sinusoidal systematic structure using MSFs [3]. Previous investigation of the LEDA data in [9], where the authors used a log-polynomial foreground model, also found sinusoidal structures.

Currently, systematics and the ambiguity of their causes pose a limiting factor in the detection and/or confidence of a detection in global 21-cm experiments. We attempt to physically motivate our model for the systematic structure in the SARAS2 data and, as a consequence of detailed systematic modelling, derive astrophysical parameter constraints on the global signal.

Specifically, we model the damped sinusoidal structure seen in the residuals after foreground subtraction with two different physically motivated systematic models representing the introduction of power at different points, (A) and (B), in the SARAS2 system as illustrated in Fig. 5.2. Any non-smooth systematic model can be included in equation (5.2) as an additive term

$$T_A = (T_{21} + T_{fg})\eta_t + T_{NS}. \quad (5.6)$$



**Figure 5.2:** A simplified schematic of a global 21-cm experiment. The figure illustrates the two positions at which systematic power could be added into  $T_A$ ; (A) prior to the antenna and (B) after the antenna in the receiver electronics, denoted by the time dependent gain,  $G_{RX}(t, \nu)$ . Note, when modelling the systematic in the SARAS2 data, we are considering these scenarios independently. The former systematic arrives with the signal from the sky,  $T_{sky}(t, \nu, \Omega)$ , and global signal,  $T_{21}(\nu)$  if present, and passes through the antenna where it is weighted by the beam directivity,  $D(\nu, \Omega)$ . The angular dependence is integrated out, and the systematic is then multiplied by the total efficiency of the antenna,  $\eta_t(\nu)$ , giving  $T_{NS,2}$  as detailed in section 5.3.3. The systematic introduced in the electronics corresponds to  $T_{NS,1}$  in section 5.3.3. After passing through the electronics, the time dependence is also integrated out of the data leaving  $T_A$ . Figure modified from Fig.2 of [116].

The sinusoidal structure may be introduced by a poor estimate of  $\eta_t$ . However, when fitting the total efficiency with a best fitting MSF of the form given in equation (5.5) we find no sinusoidal structure in the residuals. We also reiterate, for completeness, that the systematic structure may be being introduced by poor foreground modelling.

The first systematic model we use, which we refer to as the damped systematic model, corresponds to power introduced after the antenna and in the electronics [e.g. see Fig. 2 in 16]

$$T_{NS,1}(\nu) = \left( \frac{\nu}{\nu_0} \right)^{\alpha_{sys}} A \sin \left( \frac{2\pi\nu}{P} + \phi \right), \quad (5.7)$$

where  $\alpha_{sys}$  is a damping power,  $A$  is the amplitude of the systematic,  $P$  is the period,  $\phi$  is the phase and  $\nu_0$  is the central frequency in the band. We note that the prior range of  $\alpha_{sys}$  ranges from 0 – 10 (see Tab. 5.3 for all the systematic prior ranges) and as a result the model can also account for sinusoidal systematics.

The second model, referred to as the efficiency systematic, is given by

$$T_{NS,2}(\nu) = \eta_t \left( \frac{\nu}{\nu_0} \right)^{\alpha_{sys}} A \sin \left( \frac{2\pi\nu}{P} + \phi \right). \quad (5.8)$$

In this case, the systematic structure models power introduced prior to the antenna and mediated by the total efficiency which introduces some damping. Such a systematic could be explained by activity in the ionosphere over the observing period, RFI or a previously unidentified non-smooth component of the foreground.

The non-smooth structure in the data could also, as highlighted previously, be caused by emission from the ground if the assumption that this is smooth does not hold. Specifically, structure or layering in the ground at depths larger than the wavelengths of operation and below the penetration depth computed for the soil properties

	Parameter	Prior	Prior Type
Systematic	$\alpha_{\text{sys}}$	0 – 10	Uniform
	$A$	0 – 1 K	
	$P$	10 – 70 MHz	
	$\phi$	0 – $2\pi$ rad	
Signal	$\tau$	0.026 – 0.1 (STA) / 0.035 – 0.077 (ERB)	Uniform
	$\alpha$	1.3 (STA only)	
	$E_{\min}$	0.1 – 3 keV (STA only)	
	$R_{\text{rmfp}}$	30 (STA) / 40 (ERB) Mpc	Log-Uniform
	$f_*$	0.001 – 0.5	
	$V_c$	4.2 – 100 km/s	Log-Uniform
	$f_X$	0.0001 – 1000	
	$f_{\text{radio}}$	1 – 99500 (ERB only)	

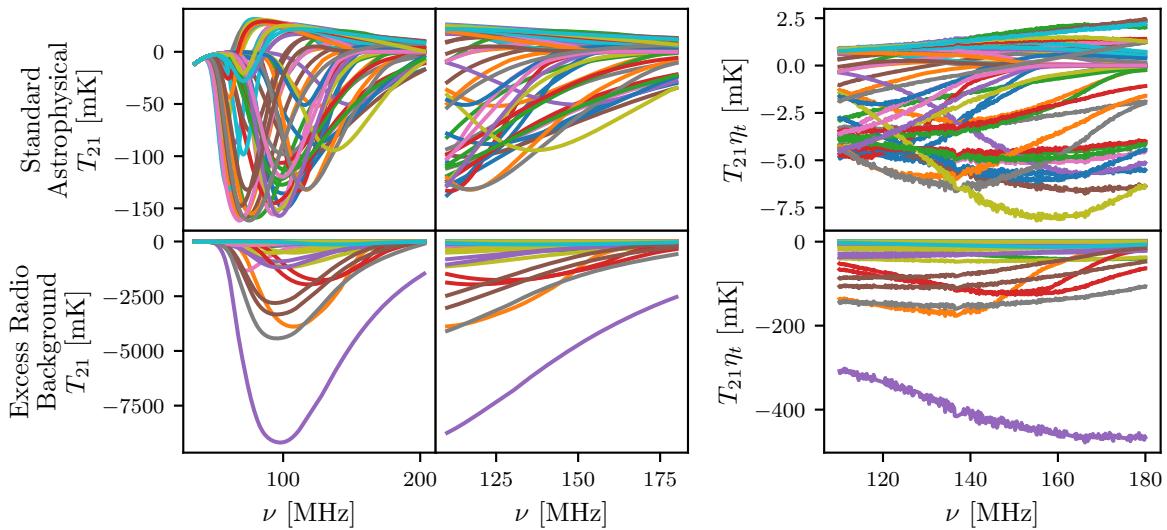
**Table 5.3:** The prior ranges and prior types used for the systematic and signal parameters fitted by POLYCHORD. Note that for the signal parameters, the prior ranges are defined by the neural network training data. For the excess radio background (ERB) signals  $R_{\text{rmfp}}$  is fixed at 40 Mpc and the X-ray SED is representative of that from X-ray binaries[46]. See section 5.3.4 for more details on each model component, the training data and the difference between the standard astrophysical (STA) and ERB models.

could introduce systematic structure. A discontinuity in the soil below the antenna such as that between the loose topsoil, caused by erosion over time, and the rock of the Deccan Plateau or from a water table could cause a damped sinusoidal structure to propagate through to the receiver noise, total efficiency and antenna temperature.

Another possible origin of a damped sinusoidal systematic could be a small clump of foliage or a root system, without significant foliage above ground, close to the deployment site. In principle, both this and ground emission are potential causes of systematics that will be mitigated in the latest iteration of the SARAS experiment, SARAS3, which has been deployed on a lake [17–19].

It is well-known that the directivity of vertical monopoles [194] naturally experiences a sinusoidal like frequency response. Such behaviour would correspond to  $T_{\text{NS},2}$ . We note, however, that the period of the sinusoidal-like structure visible in the data is faster than what would be expected from a monopole. The scale of the circular element of the antenna, 43.5 cm in radius, was chosen such that any reflections from the edges of the disks would have a period of  $\approx 350$  MHz and the observing band falls within the first resonance at 260 MHz [16, 69]. Any reflections in the beam pattern from the edges of the disk would thus be smooth across the reduced SARAS2 band and effectively subsumed by our smooth foreground model detailed in the previous subsection.

If we consider the systematic to be real and not a spurious signal introduced by an inaccurate foreground model, then our analysis could help to identify the point at which the power is introduced into the system. We note that there may be degeneracy between the two different models, but also that the efficiency is an inherent characteristic of the experiment and unlikely to mimic generic systematic properties.



**Figure 5.3:** **Top Row, left to right:** A set of 50 example signals that include Lyman- $\alpha$  heating and CMB heating, the equivalent signals in the band 110 – 180 MHz and the signals multiplied by the total efficiency of the SARAS2 antenna. **Bottom Row, left to right:** 50 example signals with different levels of excess radio background above the CMB, the same signals in the reduced SARAS2 band and finally the signals multiplied by the total efficiency of the SARAS2 antenna.

### 5.3.4 Signal Modelling

To recap, the global 21-cm signal is determined by the contrast between the spin temperature of neutral hydrogen,  $T_s$ , and the radio background,  $T_r$ , as a function of redshift

$$T_{21} = \frac{T_s - T_r}{1 + z} (1 - e^{-\tau_{21}}), \quad (5.9)$$

where  $\tau_{21}$  is the 21-cm optical depth of the IGM [34]. Calculated realisations of the global 21-cm signal are achieved by using hybrid techniques and either averaging over large modelled cosmological volumes [2D simulations e.g. 151] or directly approximating that average over redshift [1D simulations e.g. 149]. In this analysis we use the simulations detailed in [11–14, 121] and specifically the sets of global 21-cm signals presented in [121] and [14]. For each set of signals, exotic models with an excess radio background and standard astrophysics respectively, we train neural network emulators with GLOBALEMU. More details about the two sets of signal models we have used are given in the following sections, with examples shown in Fig. 5.3.

#### 5.3.4.1 Standard astrophysical signals

In this chapter, we use the recent simulations of the global signal with Lyman- $\alpha$  heating, CMB heating and multiple scattering presented in [14]. We refer to these standard astrophysical models in the text as the STA (standard astrophysical) signals. We note that, multiple scattering primarily influences the structure of the power spectrum at  $z > 20$  [14] outside the SARAS2 band.

We train GLOBALEMU on a set of 5,137 realisations of the standard astrophysical signals and test the quality of emulation with a set of 570 models. Each model spans the redshift range  $z = 6 – 39$  and the redshift spacing is given by  $\delta z = 0.1$ . The structure of the signals is determined by a set of seven astrophysical parameters [further details can be found in the introduction of the thesis and 4, 14, 15];

- The star formation efficiency,  $f_*$ : Determines the fraction of gas in dark matter halos that is converted into stars. The value of  $f_*$  drives the Lyman- $\alpha$  flux and influences the onset of X-ray heating, which both determine the depth of the global signal absorption trough and the ionizing efficiency of sources.

- The minimal circular velocity,  $V_c$ : The threshold virial circular velocity is proportional to the cube root of the minimum halo mass for star formation. Its value is determined by different cooling channels (molecular and atomic hydrogen) and star-formation-suppressing feedback mechanisms. It influences the timing of Lyman- $\alpha$  coupling, total X-ray luminosity of halos and reionization.
- The X-ray efficiency,  $f_X$ : The value of  $f_X$  affects the X-ray luminosity per star formation rate, which in turn influences the depth of the absorption trough and amplitude of any emission above the radio background during reionization.  $f_X$  also has a minor effect on reionization.
- The slope of the X-ray spectral energy distribution (SED),  $\alpha$ : The dependence of the structure of the global signal on  $\alpha$  is expected to be very weak [91] and its value plays a small role at low redshifts.
- The low energy cut-off of the X-ray SED,  $E_{\min}$ : This parameter has some influence on the efficiency of X-ray heating at low redshifts covered by the SARAS2 band. While it is not expected to affect the structure of the signal significantly, in some cases, it can have more of an impact than  $\alpha$  or  $R_{\text{mfp}}$ .
- CMB optical depth,  $\tau$ : The optical depth is directly related to the ionizing efficiency of sources,  $\zeta$  and its value strongly influences the redshift of reionization. The value of  $\tau$  has been determined by [22] to be  $0.055 \pm 0.007$  and Tab. 5.3 shows that this range is explored completely by the prior for the STA models.
- The mean free path of ionizing photons,  $R_{\text{mfp}}$ : The value of  $R_{\text{mfp}}$  affects the rate of ionization of the neutral hydrogen gas corresponding to the gradient of the signal at low redshifts. The effects of varying  $R_{\text{mfp}}$  are also not expected to influence the structure of the global signal significantly [see e.g. 91].

The primary reason  $E_{\min}$ ,  $\alpha$  and  $R_{\text{mfp}}$ , are explored in the simulations is because the simulations are also used to determine models of the power spectrum on which they have a greater influence. We train GLOBALEMU using all seven astrophysical parameters and subsequently perform fits with fixed values of  $R_{\text{mfp}} = 30$  Mpc and  $\alpha = 1.3$  [as was done with the EDGES High Band data in 91, using the contemporary standard astrophysical models and signal emulator]. We repeat the analysis in Appendix D of [6] allowing POLYCHORD to fit for all seven parameters, however we do not discuss these results in the main text as we find the effect of including the additional parameters is minimal. The ranges, equivalent to the priors, of all the parameters sampled in the training and test data sets are given in Tab. 5.3.

We assess the accuracy of the emulator across the band 110 – 180 MHz (or equivalently  $z \approx 7 - 12$ ). We use a pragmatic target accuracy when emulating the signals of on average approximately 10% of the expected noise from a global 21-cm experiment [4]. This varies based on the experiment with a value of 2.5 mK for REACH, 2 mK for EDGES [8] and approximately 1 mK for SARAS2. Therefore, when emulating the STA models, we assume a target accuracy of between 1 and 2.5 mK and use the RMSE metric given by equation (7) in [4]. The mean, 95<sup>th</sup> percentile and the worst RMSE values from the test data set of 570 models, with sampling resolution equivalent to 0.122 MHz, are 0.8, 1.9 and 6.8 mK respectively. We find that only 29 models have an RMSE larger than  $\approx 1.9$  mK indicating a high degree of accuracy. We use a fully connected network with 4 hidden layers of 16 nodes each to emulate the STA models.

#### 5.3.4.2 Excess radio background signals

Typically,  $T_r$  in equation (5.9) is assumed to be equal to the CMB temperature. However, as discussed in chapter 1, one of the possible explanations for the larger than expected absorption feature reported by EDGES [8] is an excess radio background above the CMB. In addition, a population of radio sources at high redshifts could naturally contribute to  $T_r$  [e.g 117]. While there is some evidence for a larger than expected radio

background from ARCADE2 [48] and LWA [49], there remain some concerns about the Galactic modelling in these works [50].

[121] investigated the introduction of an excess radio background from high redshift radio galaxies, and we use the models presented there in this work in an attempt to constrain the parameter  $f_{\text{radio}}$  with the SARAS2 data.  $f_{\text{radio}}$  denotes the radio production efficiency of early galaxies, a value of one corresponds to the present day and the range of  $f_{\text{radio}}$  in our training data set is given in Tab. 5.3. We refer to these exotic astrophysical models throughout the rest of the chapter as the ERB (excess radio background) models.

The simulations use a similar parameter description of the global signal presented in the previous subsection with the additional parameter  $f_{\text{radio}}$ . The value of  $R_{\text{mfp}}$  is fixed at 40 Mpc when running the simulations. The X-ray SED is assumed to be from X-ray binaries [195], and the simulations are consequently independent of  $\alpha$  and  $E_{\min}$ . The models also include the effects of Lyman- $\alpha$  heating, CMB heating and multiple scattering.

The data set contains 4,311 training models and 479 test models. We train `GLOBALEMU` with the five parameters as inputs;  $f_*$ ,  $V_c$ ,  $f_X$ ,  $\tau$  and  $f_{\text{radio}}$ . The mean, 95th percentile and the worst RMSE values are 7.3, 27.3 and 125.9 mK respectively. We use a network with 4 hidden layers each with 16 nodes as was done with the STA models and note that with this training data set the accuracy does not significantly improve if we increase the size of the network. However, the mean accuracy of emulation, 7.3 mK, is the same order of magnitude as our target accuracy of 1 – 2.5 mK. The magnitudes of the signals with the largest errors, after multiplication by  $\eta_t$ , are significant in comparison to the expected noise and as fractional accuracies the worst and 95th percentile results are reasonable.

## 5.4 Sensitivity of the SARAS2 Data to Astrophysical Parameters

When considering the sensitivity of the SARAS2 data to the global 21-cm signal, it is important to consider that any signal in the data will be multiplied by the total efficiency of the antenna. As has previously been discussed and shown in Fig. 5.3, this significantly reduces the magnitude of the signals in the data and given the expected noise of 11 mK results in a low signal-to-noise ratio. In turn, this makes any signal in the data hard to recover.

Further, the observations cover a bandwidth expected to include the EoR which has implications for the types of signals that we expect to be able to constrain. The EoR window is sensitive to the value of  $\tau$  which affects the redshift of reionization and as a result the maximum amplitude of the global signal. For example, a high value of  $\tau$  can lead to an earlier reionization. However, in the absence of efficient X-ray heating and presence of a low Lyman- $\alpha$  flux the effect of  $\tau$  is less significant in the SARAS2 bandwidth.

Similarly, the value of  $E_{\min}$  has a more significant effect on the structure of the global signal in the SARAS2 band if  $f_X$  is high. In this case, the value of  $E_{\min}$  affects the depth of the signal and the efficiency of X-ray heating. If we maintain the Lyman- $\alpha$  flux, i.e. onset of Lyman- $\alpha$  coupling, and increase the value of  $E_{\min}$  we reduce the efficiency of X-ray heating, move the minimum of the signal to lower redshifts and can create a deep global signal inside the SARAS2 band. A similar effect can occur if we decrease the value of  $f_X$  from high to low, and this is more prominent than that introduced by variation in  $E_{\min}$ . The analysis should therefore be sensitive to models with low values of  $f_X$  and high values of  $E_{\min}$ .

In addition, if  $f_X$  is high, and to a lesser extent if  $E_{\min}$  is low, then we can expect that there will be significant excess radio background which will produce a prominent and deep signal in the SARAS2 data even after multiplication by the total efficiency of the antenna.

$V_c$  and  $f_*$  determine the strength of the Lyman- $\alpha$  flux, consequent onset of Lyman- $\alpha$  coupling and position of the signal minimum. High values of  $f_*$  correspond to a high fraction of the gas in dark matter halos being converted into stars, which leads to a high Lyman- $\alpha$  flux. The top panels of Fig. 5.4 show that the strongest signals in the SARAS2 band after multiplication by the total efficiency are those with low Lyman- $\alpha$  fluxes (high  $V_c$  and low  $f_*$ ) whereas the weakest signals are those with high Lyman- $\alpha$  fluxes (low  $V_c$  and high  $f_*$ ). Here, we have defined “high” and “low” values of the parameters with respect to the middle of the log-prior ranges. Our analysis should therefore be more sensitive to models with low Lyman- $\alpha$  fluxes.

Further, dividing the two classes of signals in Fig. 5.4 based on their value of  $f_X$  we can see that we should also expect our analysis to be sensitive to low values of  $f_X$ , as expected from the discussion above.

We should also consider our sensitivity to models that have both high or both low values of  $V_c$  and  $f_*$  in combinations that do not meet the crude criteria for high and low Lyman- $\alpha$  fluxes defined above. These models are shown in black in the bottom panel of Fig. 5.4 against the backdrop of previously “classified” models in grey. From the figure we can see that these models typically also have low Lyman- $\alpha$  fluxes and have minima at more recent redshifts, particularly in comparison to the “high” Lyman- $\alpha$  flux signals discussed previously. These signals therefore have dominant structures and relatively large magnitudes in the SARAS2 band, and indicate a more general sensitivity to the values of  $f_*$  and  $V_c$ .

Finally, the data is expected to be sensitive to ERB signals with very high values of  $f_{\text{radio}}$  as these signals have deep absorption troughs of a few hundred mK even after multiplication by the total efficiency of the antenna. This is particularly true when there is also a high Lyman- $\alpha$  flux and low X-ray efficiency, which results in strong variation of the signal within the SARAS2 band.

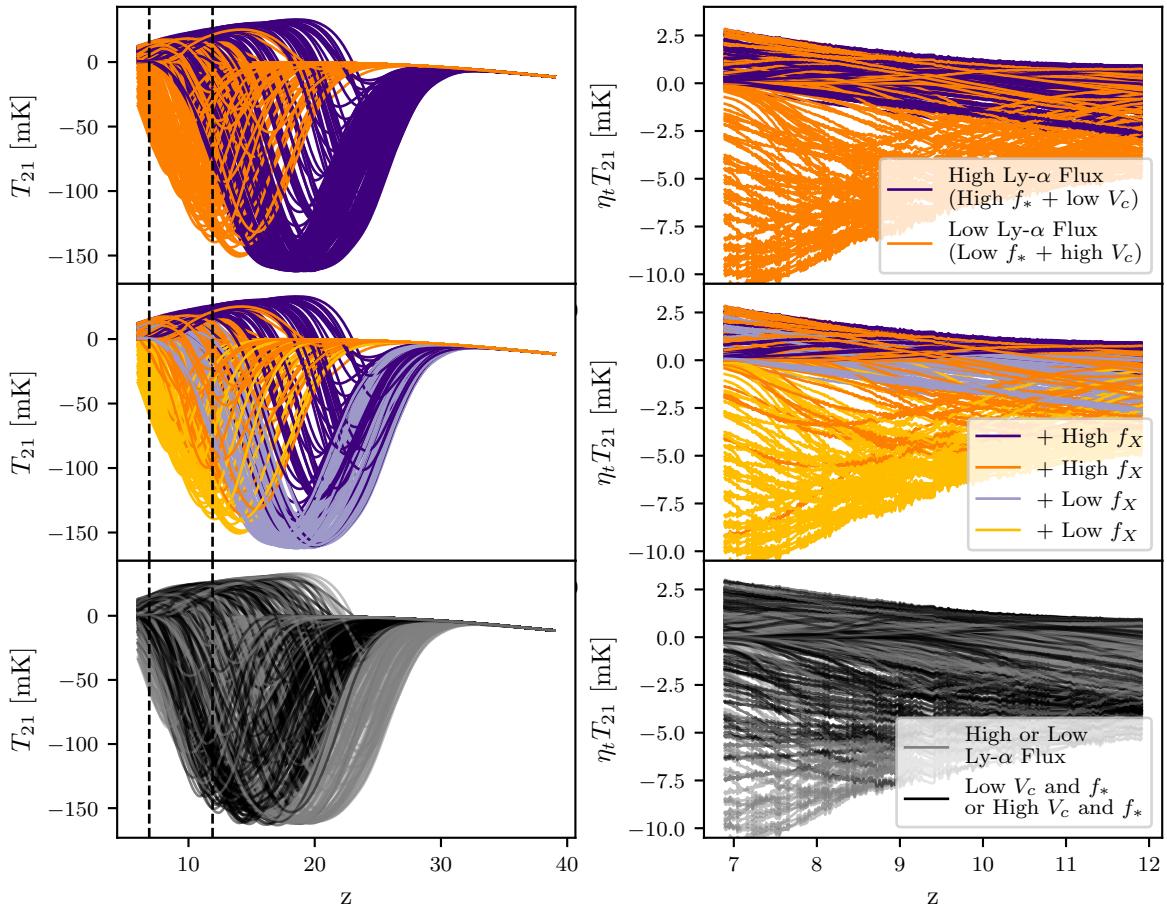
## 5.5 Results

Fig. 5.5 summarises the different combinations of signal, systematic and noise models that were fitted to the SARAS2 data. For completeness, we report fits without signals and/or systematics. The four fits with the highest Bayesian evidence are approximately equivalent (within error bars) and the relative weights based noise model is comparatively poor compared to the two alternatives considered when we include a systematic model in the fit. These points are further discussed in the following sections.

The evidence,  $Z$ , is a marginal likelihood integrated overall the fitted parameters. It quantifies the probability that the data is described by the chosen model components and is the normalising factor in Bayes theorem. A higher  $\log(Z)$  indicates a preference for that model or hypothesis as a description of the data.

Of the tested combinations of model components we find that modelling with a STA signal and no systematic leads to only a marginal increase in evidence in comparison to a foreground only fit as can be seen in the bottom panel of Fig. 5.5. The increase in evidence is larger when we model an ERB signal and no systematic, but this is to be expected as these signals can have significantly deeper absorption features than the STA signals and as a result they are better able to fit out the larger systematic structure. Fits with systematic modelling have significantly higher evidences, regardless of whether we include an STA, ERB or no global signal model, than those without systematic modelling,  $\Delta \log(Z) \approx 200 - 300$ . The data, therefore, favours the presence of a systematic model, but there is no strong indication for the presence of a signal in the data.

Appendix B in [6] shows the residuals found when fitting the data with the PSF foreground model, no signal model, the efficiency systematic and the constant noise model (fit number 4) compared with the residuals from a



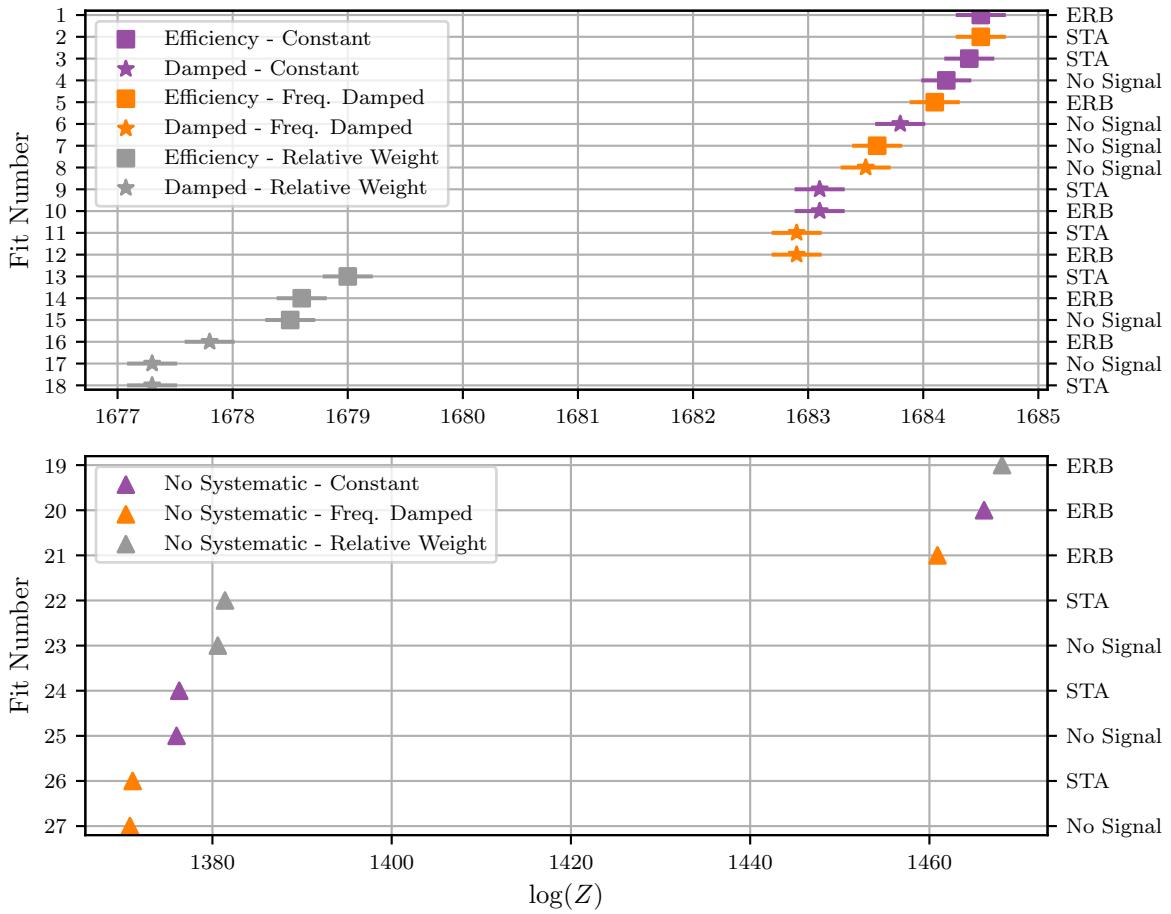
**Figure 5.4:** The figure shows 243 STA signals from the testing data coloured based on the values of  $f_*$ ,  $V_c$  and  $f_X$ . The left column shows the signals as they are in the testing data and the right panel shows the signals in the SARAS2 band, indicated on the left by the dashed lines, after multiplication by the total efficiency of the antenna. In the top panels, we show the set of signals split based on the type of Lyman- $\alpha$  flux that they correspond to. A high Lyman- $\alpha$  flux is produced by a high  $f_*$  and low  $V_c$  where we judge ‘low’ and ‘high’ with respect to the middle of the log-prior range for each parameter. In the middle panels, we further split the models based on their values of  $f_X$ . From the graph we can see that, after multiplying by the total efficiency of the SARAS2 antenna, the signals with a low Lyman- $\alpha$  flux and low value of  $f_X$  have the largest absolute magnitudes and so our analysis should be most sensitive to these models. The bottom panel shows the previously classified 243 STA signals in grey in comparison to a further 257 signals in black that do not meet our crude classification of ‘low’ and ‘high’ Lyman- $\alpha$  fluxes and with neither high  $f_*$  and low  $V_c$  or low  $f_*$  and high  $V_c$ . These models have prominent structures in the SARAS2 band, indicating a more general sensitivity to the values of  $f_*$  and  $V_c$ .

high order polynomial fit. The consistency between the two sets of residuals suggests that the complexity of the modelling used here is sufficient to describe the data. Although, a signal may still be present with an absolute maximum magnitude less than the noise after multiplication by the total efficiency of the antenna.

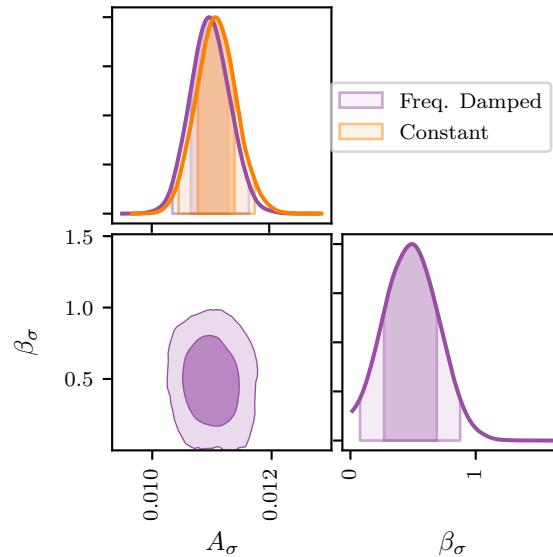
In the following sections, we discuss in more detail the results found when modelling with the different components outlined in section 5.3.

### 5.5.1 Noise

We would expect that the relative weights based noise model would provide the best representation of the noise in the data as it has been derived from system parameters. In the absence of systematic modelling, this noise model performs comparatively well because the residuals after foreground modelling are larger at higher frequencies, following the trend of the weights. However, from the top panel of Fig. 5.5 it is clear that the log-evidence for fits with this noise model and a systematic is much lower than for fits with the two alternatives.



**Figure 5.5:** A graphical summary of the different fits performed in this analysis, which demonstrates that the highest evidence fits from all of those performed are fits of equal quality. The signal type can be found on the right-hand y-axis, the fit number on the left-hand axis and the noise and systematic types in the legends. The fit numbers are used for referencing throughout the text. The top panel shows fits that include a systematic model, and the bottom panel shows fits without a systematic model. We divide the two classes of fits into separate graphs because those with a systematic have a much higher evidence,  $\Delta \log(Z) \approx 200 - 300$ , than those without. Note that a difference of  $\Delta \log(Z) \approx 10$  corresponds to a difference of  $\exp(10) \approx 22000$  or betting odds of 22000:1 in favour of the higher evidence model.



**Figure 5.6:** The posteriors for the noise parameters when fitting the SARAS2 data with STA signals and the efficiency systematic. The posteriors for the amplitude of the noise are near identical for the frequency damped (purple) and constant noise models (orange). In addition,  $\beta_\sigma \approx 0.5$  which corresponds to very weak damping. In combination, these two facts indicate a similarity between the two noise models. All the corner plots presented in this chapter were produced using ANESTHETIC [178] and the contour lines here represent 68% and 95% confidence intervals.

This is likely due to some degeneracy, that would be hard to disentangle, between the systematic model and the weights based noise model which follow the same trend in the data (increasing with frequency).

For some combinations of signal and systematic, we find that the log-evidence is comparable between the fits with constant and frequency damped standard deviation. However, in the majority of cases, the constant noise modelling gives a higher evidence. In the one case where this is not true, fitting with a STA signal and the efficiency systematic, the evidences are comparable, indicating that neither noise model is favoured over the other.

For the fit with the efficiency systematic and STA signals, we find, with the frequency damped noise,  $\log(Z) = 1684.5 \pm 0.2$ . In comparison, when modelling the noise with a constant standard deviation, we find  $\log(Z) = 1684.4 \pm 0.2$ . For these two fits, we can look in more detail at the maximum likelihood noise models, and the posteriors for the noise model parameters are shown in Fig. 5.6. For the frequency damped model the maximum likelihood parameters are  $A_\sigma = 10.9$  mK and  $\beta_\sigma = 0.45$  and for the constant noise model the maximum likelihood amplitude is  $A_\sigma = 11.0$  mK. In the former, the damping power is weak, recalling that the prior range for this parameter is 0 – 5, and the maximum likelihood amplitude for both models is near identical with a similar posterior distribution. The weak damping is a consistent feature across all the fits with the damped model, as is the similarity in the amplitude of the noise.

To summarise, when comparing the two best noise models, constant and frequency damped, we find an equality in log-evidence or preference for the former model, a consistent weak damping in the latter and similarity in amplitude. This supports the expectation from the literature that the noise should be uniform across the band in the SARAS2 data [69, 70].

### 5.5.2 Systematics

As discussed, we see a large increase in log-evidence,  $\Delta \log(Z) \approx 300$ , when joint fits with a PSF foreground and the efficiency or damped systematic models are performed as shown in Fig. 5.5. Comparing like for like fits with different systematic models, we consistently find that the efficiency systematic model is favoured over the alternative by the data. Regardless of noise modelling, the five highest evidence models are all fits with the efficiency systematic.

Both systematics rely on the same parameterisation and so a direct comparison can be made between fits with the same noise and signal modelling but different systematic models. The posterior distributions for each parameter are typically well constrained Gaussian-like distributions.  $\alpha_{\text{sys}}$  is generally centred around 0 for the efficiency systematic, in comparison to a value of approximately 3 for the damped systematic. This indicates that, in the case of the efficiency systematic, the data favours a sinusoidal systematic structure that is damped predominantly by the total efficiency of the radiometer,  $\eta_t$  (i.e. the term  $(\nu/\nu_0)^{\alpha_{\text{sys}}} \approx 1$  in equation (5.8)).

For both systematic models the period,  $P$ , and phase,  $\phi$ , have similar distributions, suggesting that any systematic in the data has a period of  $\approx 32.5$  MHz and phase of  $\approx 0$  rad. This is true generally, regardless of noise and signal model, since the systematic dominates the residuals. The systematics also have similar amplitudes but, while in the efficiency systematic this is mainly determined by  $A$ , in the damped systematic it is mainly determined by the term  $(\nu/\nu_0)^{\alpha_{\text{sys}}}$  in equation (5.7).

From the log-evidences, we can conclude that the data marginally favours a sinusoidal systematic damped by the total efficiency of the antenna. However, as with the noise modelling, the differences in log-evidence between like for like fits with different systematics are small and all the fits with the different systematic models contain

potential information about the astrophysical parameter space. We can, therefore, use all the samples from the different fits, weighted by their evidences, in a combined analysis. This approach effectively marginalises over the systematic and noise model parameters and is detailed in the following section.

### 5.5.3 Disfavoured 21-cm signals from combined samples

We can use `ANESTHETIC` [178] to combine the posterior samples from `POLYCHORD`, weighted by the fit evidence, from the various Nested Sampling runs to determine the types of global 21-cm signals disfavoured in our analysis. This is advantageous since it provides a method to deal with uncertainty in the modelling of the standard deviation on the noise and the systematic, and should be considered a conservative view of any constraints on the parameter space from the SARAS2 data.

While we note that the fits in Fig. 5.5 do not indicate a preference for the presence of a signal, we can still use the data to determine constraints on the parameter space of the global 21-cm signal, as has previously been done with data from EDGES [91].

In the previous two sections we have made the argument that the amplitude of the noise is best described by a constant standard deviation and the systematic is best modelled by the efficiency systematic. While these statements are true to an extent, the range in log-evidence between the corresponding fits with different model components is not significant. In fact, the former is largely motivated by the fact that the constant noise model is simpler than the frequency damped model and thus favourable. In practice, the data does not tell us which noise model is preferred.

Since the 2D and 1D posteriors for the astrophysical parameters,  $\theta_{21}$ , by definition marginalise over the systematic and noise parameters, we can confidently combine the samples and draw conclusions from the corresponding posteriors. As alluded to the posterior samples,  $P(\theta|D, M)$ , when combined using `ANESTHETIC` are weighted by weights,  $w$ , that are directly proportional to the fit evidence,  $Z$

$$P_{\text{combined}}(\theta_{21}|D, M) = \sum_i w_i P_i(\theta_{21}|D, M), \quad (5.10)$$

where the weights  $w_i = Z_i / \sum_j Z_j$ ,  $\theta_{21}$  is a vector of astrophysical signal parameters for fit  $i$ ,  $D$  is the data and  $M$  is the analytical model. In Appendix C in [6] we show the values of  $w_i$  for each of the different fits performed in this analysis separated by their signal type.

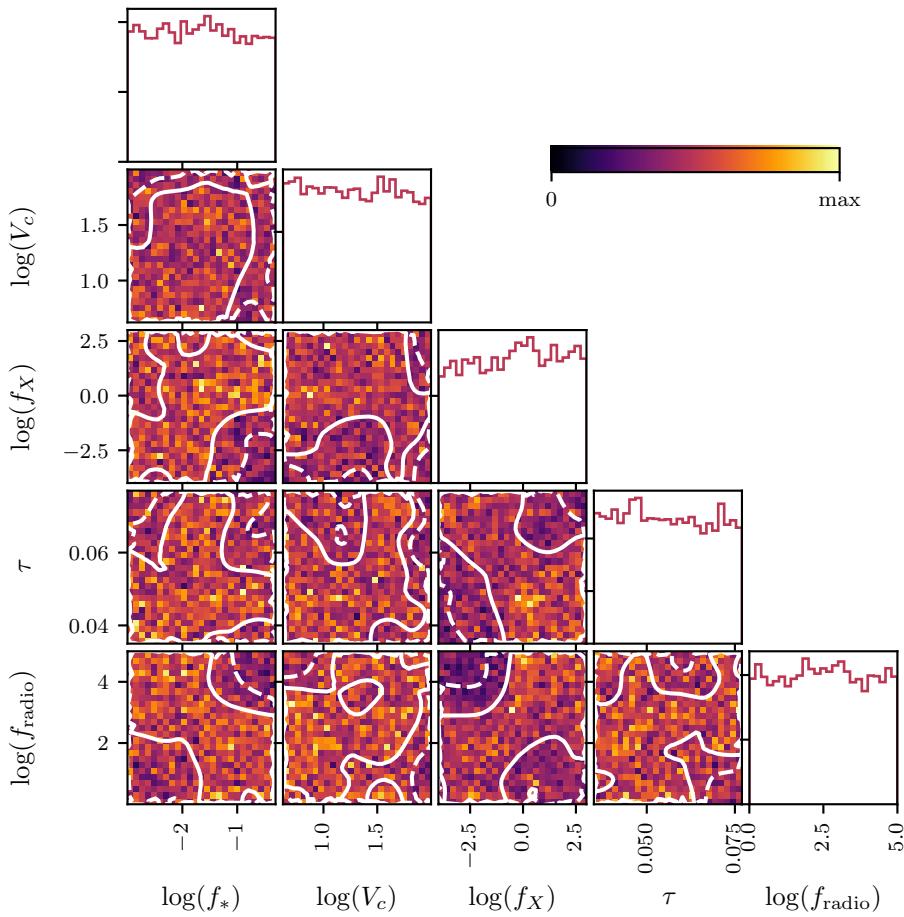
In the following subsections, we therefore discuss posteriors from combined samples for all the fits containing STA signals (see Appendix D in [6] for results with variable values of  $R_{\text{mfp}}$  and  $\alpha$ ) and fits containing ERB signals.

We briefly compare our results with those previously reported in [69, 70] and with results published in early 2022 from HERA [109] in section 5.5.3.3.

#### 5.5.3.1 Excess Radio Background Signals

We can analyse combined samples from fit numbers 1, 5, 10, 12, 14 and 16, those containing ERB signals, as shown in Fig. 5.7. The recovered 1D histograms are generally flat and do not show any significant constraints. Although regions of high  $f_{\text{radio}}$  ( $\gtrsim 407$  and  $\gtrsim 707$ ) in combination, separately, with low  $f_X$  ( $\lesssim 0.21$ ) and high  $f_*$  ( $\gtrsim 0.03$ ) are disfavoured at 68% confidence.

We use histograms to illustrate the samples in the parameter space. Binning the raw samples across the parameter space gives a clearer impression of regions of the parameter space that the Nested Sampling algorithm



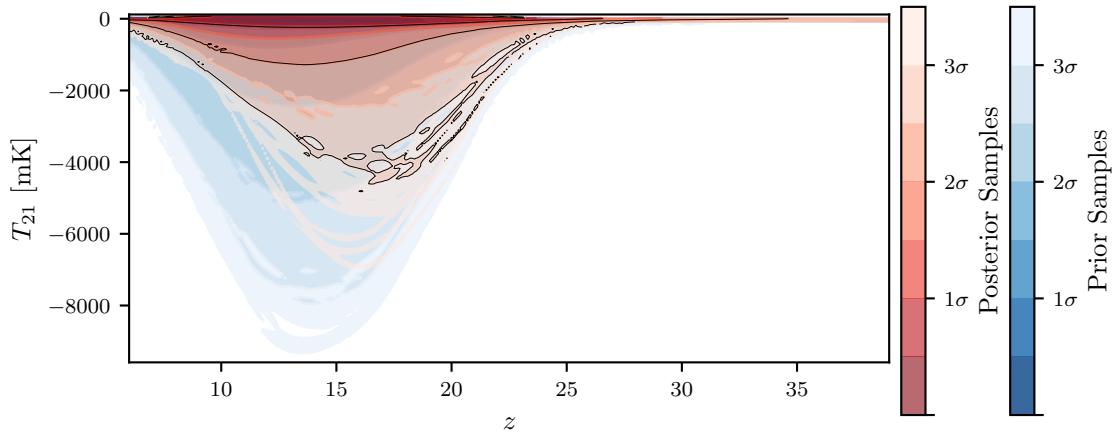
**Figure 5.7:** The 1D and 2D distributions for the combined nested samples of all fits to the SARAS2 data containing ERB signals. The solid contours in the 2D parameter space enclose 68% confidence regions, and the dashed contours enclose 95% confidence regions. Combinations of high  $f_*$  ( $\gtrsim 0.03$ ) and high  $f_{\text{radio}}$  ( $\gtrsim 707$ ) as well as low  $f_X$  ( $\lesssim 0.21$ ) and high  $f_{\text{radio}}$  ( $\gtrsim 407$ ) are less densely sampled than the rest of the parameter space and are consequently disfavoured in our analysis.

explored in greater detail (i.e. those with higher likelihoods and combinations of parameters that are favoured by the data). For example, in Fig. 5.7 favoured regions of the parameter space would be sampled in greater detail and the corresponding bins in the 2D plots would be shaded in a lighter yellow.

To quantify the strength of 2D constraints across the parameter space, the authors of [109] assess the ratio of the minimum and maximum posterior probability ( $\propto$  height of the histogram bins) across the 2D space. A ratio of 1 would indicate a perfectly flat posterior, and a low value indicates a non-uniform distribution. The metric is limited in that it does not indicate the direction of any non-uniformity and is insensitive to any pseudo-random scatter across the parameter space that may result from a fine binning of a relatively flat distribution<sup>c</sup>. However, in combination with a visual inspection of a constrained parameter space, it can be a useful metric to determine the magnitude of any directional non-uniformity. For example, the ratio between the minimum and maximum posterior probability for  $f_{\text{radio}} - f_X$  in Fig. 5.7 is 0.10. In comparison, the maximum ratio is 0.20 for  $V_c - f_{\text{radio}}$  and the minimum ratio is 0.08 for  $f_X - \tau$ .

Further, the deepest signals have high Lyman- $\alpha$  fluxes (high  $f_*$  and low  $V_c$ ), low X-ray efficiencies,  $f_X$ , and high values of  $f_{\text{radio}}$ . After multiplication by the total efficiency of the antenna these signals typically have magnitudes larger than the expected noise, and consequently we would expect to exclude these with the SARAS2 data. This can be seen clearly in Fig. 5.8 in which we plot the functional posterior samples in red on top of

<sup>c</sup>The metric is also dependent on the number of bins into which the samples are separated and so comparison across experiments is difficult.



**Figure 5.8:** The figure shows the contraction from the prior (blue) to the combined posterior samples (red) in the 5D astrophysical parameter space (i.e.  $T_{21} - z$ ) for the ERB signals. We plot the prior in blue and posterior in red on top of each other to highlight that we rule out the deepest of these exotic astrophysical models. The figure was produced using FGIVENX [196].

the prior in blue using the tool FGIVENX [196]. The tool allows us to visualize the combined posterior samples shown in Fig. 5.9 as a set of contours in the  $T_{21} - z$  space. Note that although the priors on our parameters are uniform, this does not necessarily translate to a uniform prior in the global 21-cm signal.

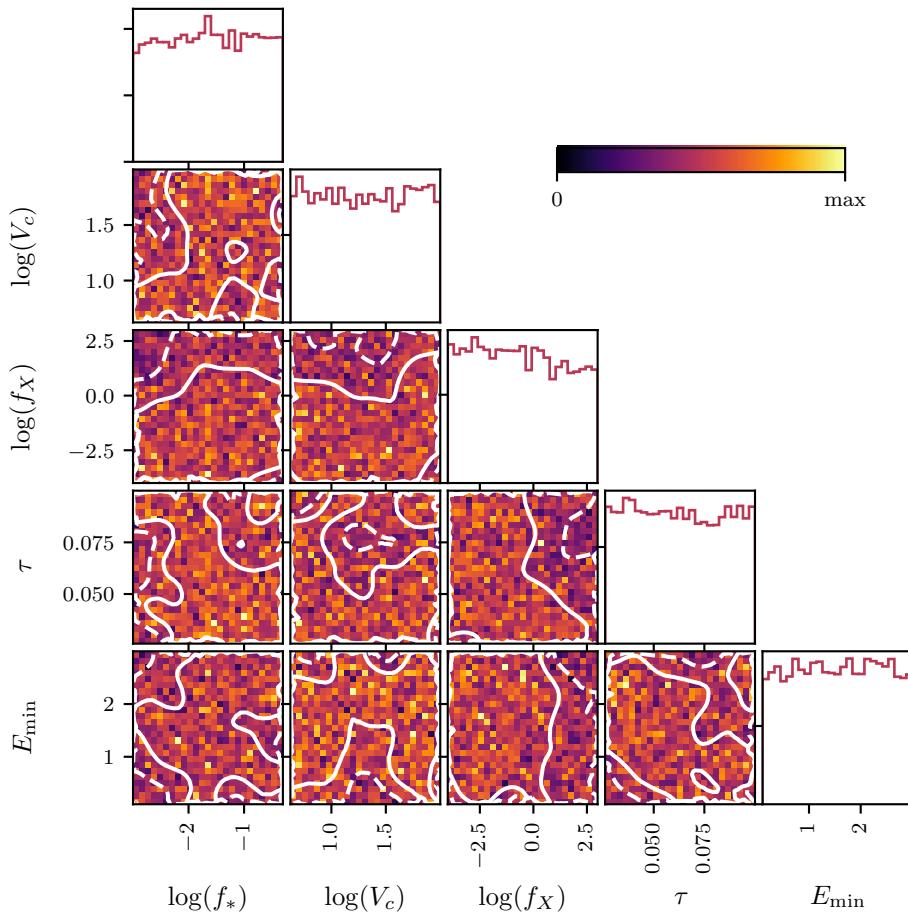
### 5.5.3.2 Standard Astrophysical Signals

Histograms produced from the combined samples for the astrophysical parameters from fit numbers 2, 3, 9, 11, 13, and 18 from Fig. 5.5, those with STA signals, are shown in Fig. 5.9. The histograms are flat, indicating that we do not significantly constrain the parameter space, and further indicating that there is no preference for a signal in the data.

Fig. 5.10 shows the functional posterior in red, along with the equivalent for samples taken from the prior in blue. There is some indication that the data disfavours (lighter shaded red regions) signals with absorption features at high redshift. For example, the preferred region (darker red and blue with significance  $< 1\sigma$ ) around the absorption minimum, where the sampling is highest, is larger and shifted to lower redshifts in the posterior than in the prior. However, the contraction from the prior to the posterior is minimal, as would be expected from the flat nature of the distributions in Fig. 5.9. Therefore, any conclusions we make about constraints on the standard astrophysical priors are by definition very weak.

### 5.5.3.3 Results in the context of the previous SARAS2 analysis and early HERA results

As discussed in section 5.1, previous analysis of the SARAS2 data, with standard astrophysical signal models, led to the conclusion that a set of specific astrophysical models with rapid reionization and weak X-ray heating (low values of  $f_X$ ) were disfavoured. We find that we very weakly disfavour standard astrophysical signals with high Lyman- $\alpha$  fluxes and weak heating, and this can be seen in Fig. 5.10. Our constraints are stronger on the ERB signals in which we also find that we disfavour signals with weak X-ray heating, in particular, although these signals were not explored in the previous SARAS2 analysis. It is important to note that the simulations used in this chapter have different parameterizations of the X-ray SEDs and the original SARAS2 analysis was confined to the study of a limited sample of 264 models.

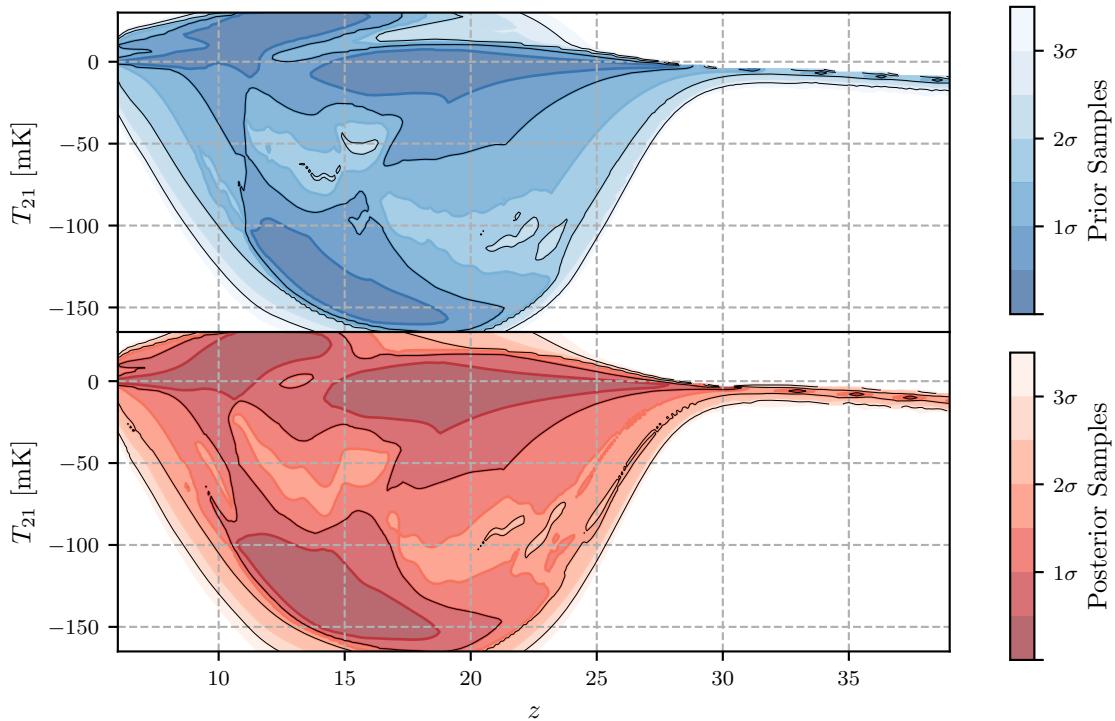


**Figure 5.9:** The 1D and 2D histogram plots from the combined nested samples of all fits to the SARAS2 data containing STA signals with fixed values of  $R_{\text{mfp}}$  and  $\alpha$ . The solid contours in the 2D posteriors enclose regions of 68% confidence, and the dashed contours enclose regions of 95% confidence. The histograms are relatively flat, and we do not disfavour values of individual parameters.

Further, the signals used in each study have significantly different dependence on the Lyman- $\alpha$  flux with the introduction of Lyman- $\alpha$  heating in the simulations used in this work. Additional heating from the CMB is also included here, which will influence the conclusions we make about the strength of the X-ray heating and makes comparison between the two sets of analysis difficult.

The additional heating mechanisms influence the position of the minima of the global signal, with a higher heating shifting the minima to higher redshifts outside the SARAS2 band and consequently reducing the magnitude of the signal at low redshift. This in turn leads to a lower signal-to-noise ratio in the data and makes signals with high heating hard to rule out. In contrast, signals with weak heating have minima at lower redshifts, have larger absolute magnitudes in the SARAS2 band and higher signal-to-noise ratios making them easier to rule out.

Lyman- $\alpha$  and CMB heating, as discussed in section 5.3.4, actually reduce the maximum allowed depth of the global signal to approximately  $\approx 165$  mK in comparison to the predictions from the previous state-of-the-art simulations, used in the original SARAS2 analysis, which could reach depths of  $\approx 250$  mK [13]. It is this difference in the maximum absolute magnitude between the two sets of signals used in each set of analysis which explains why the previous analysis is able to rule out a set of signals with such high confidence ( $> 5\sigma$  in some cases) and in this chapter we are unable to make any significant conclusions for standard astrophysical signals. The majority of the signals ruled out in the previous work have amplitudes  $\gtrsim 125$  mK and so if present



**Figure 5.10:** The figure shows the contraction from the prior (blue) to the combined posterior samples (red) in the 5D astrophysical parameter space (i.e.  $T_{21} - z$ ) for the STA signals. The graph indicates that the data weakly disfavours (lighter shaded red regions) standard astrophysical signals with absorption features at high redshift. However, we note that the contraction from the prior to the posterior is minimal. Again, the figure was produced using FGIVENX [196], and we have added grid lines to make the differences between the two distributions clearer.

in the SARAS2 data would have a higher signal-to-noise ratio than the majority of the models analysed in this chapter with amplitudes  $\lesssim 165$  mK.

Comparison with results from other experiments in the literature is generally difficult because the analysis often uses contemporary parameterizations of the global signal that are subsequently superseded by more astrophysically accurate models. However, the upper limits on the 21-cm power spectrum at  $z \approx 8$  and 10 by the HERA (Hydrogen Epoch of Reionization Array) collaboration have been used to derive parameter constraints on an excess radio background from high redshift radio galaxies using the same parameterization used in this chapter across the band  $z \approx 7 - 12$  [109, 197].

As discussed, SARAS2 data is most sensitive to and able to exclude the ERB signals with the largest radio backgrounds because their magnitudes within the SARAS2 band are larger than the experimental noise. Similarly, the authors in [109] are able to exclude models with high radio backgrounds because the corresponding power spectra are larger than the upper limits provided by HERA.

The authors perform parameter constraints using a likelihood function that marginalises over systematics, the MCMC Ensemble sampler EMCEE [83] and a neural network emulator of the power spectrum detailed in their appendix A. When investigating the parameter constraints on the ERB models the authors rule out, with a higher significance, a similar region of the  $f_{\text{radio}} - f_X^{\text{d}}$  parameter space as is done in this chapter with the SARAS2 data. Specifically, the authors rule out values of  $f_X < 0.33$  and  $f_{\text{radio}} > 391$  in comparison to our values of 0.21 and 407 respectively.

More recently, data from a larger number of nights was analysed by the HERA collaboration [99]. In this work, the parameterisation of the models was modified, making comparisons with the results in this thesis more

<sup>d</sup>Note that  $f_{\text{radio}}$  is referred to by  $f_r$  in [109].

challenging. However, the authors found that the additional data did not significantly change the constraints that could be produced with the original 18 nights of observations.

## 5.6 Conclusions

In this chapter we have reported constraints on the EoR using data from the SARAS2 experiment, the Nested Sampling algorithm **POLYCHORD**, the derivative constrained function fitting code **MAXSMOOTH** and the global signal emulator **GLOBALEMU**.

We have fitted, to data from a global 21-cm experiment for the first time, standard astrophysical signals with Lyman- $\alpha$  and CMB heating and exotic astrophysical models with an excess radio background produced from high redshift galaxies. General conclusions from our analysis are summarised below;

- We have found no conclusive evidence for the presence of a signal in the data, and fits performed with and without signal modelling have comparable evidences.
- The data generally favours the presence of noise with a constant standard deviation across the SARAS2 band over the two alternatives tested in this chapter.
- We have illustrated the presence of a damped sinusoidal systematic in the data using the smooth foreground model implemented with **MAXSMOOTH**. Our analysis suggests that this systematic is best modelled as a sinusoidal function that has been damped by the total efficiency of the antenna. This implies that the systematic is introduced as power external to the radiometer, rather than via the electronics in the receiver chain and back-end. However, we note that the log-evidence difference between the fits performed with the two different systematic models is marginal and that the non-smooth structure may have been introduced by a poor foreground model. If real, the systematic could be caused by discontinuities in the soil surrounding the antenna or shrubbery and root systems in close proximity both of which are issues that will be alleviated by the deployment of SARAS3 on a lake.
- While we do not constrain individual parameters, for the ERB signals, we disfavour combinations of high  $f_*$  ( $\gtrsim 0.03$ ) and high  $f_{\text{radio}}$  ( $\gtrsim 707$ ) and low  $f_X$  ( $\lesssim 0.21$ ) and high  $f_{\text{radio}}$  ( $\gtrsim 407$ ) which produce the deepest absorption troughs as can be seen in Fig. 5.7 and Fig. 5.8.
- The analysis with ERB signals disfavour models with weak heating, particularly X-ray heating in the latter case, in agreement with the SARAS2 data analysis in [69] and [70].
- We disfavour a similar combination of low  $f_X$  and high  $f_{\text{radio}}$  for the ERB models as was recently done using the power spectrum upper limits from 18 nights of HERA data with an identically parameterized model of the EoR [109].

The analysis presented here serves to highlight that non-smooth systematics, if effectively identified with tools like **MAXSMOOTH** and modelled, do not prevent the derivation of constraints on the astrophysics of the early universe.



CHAPTER

6

**CONSTRAINTS ON GALAXIES AT  $z \approx 15 - 25$  WITH  
SARAS3**

## 6.1 Probing Cosmic Dawn

Understanding the early Universe, when the first stars and galaxies formed, is one of the major science goals of a number of new observatories. The recently launched James Webb Space Telescope (JWST) will directly image these early galaxies in deep near-infrared surveys. Its increased sensitivity, in comparison to the previous generation of telescopes, will allow JWST to target faint high-redshift galaxies existing during the first few hundred million years of cosmic history all the way out to the cosmological redshift of  $z \sim 20$  [26]. Future confirmed and proposed X-ray missions, such as ATHENA [Advanced Telescope for High Energy Astrophysics, 198], Lynx [199] and AXIS [Advanced X-ray Imaging Satellite, 200], will supplement this exploration by observing the hot gas in the Universe. Radio telescopes aim at complementing this picture by mapping the neutral intergalactic gas across the first billion years of cosmic history via observations of the 21-cm spin-flip transition of atomic hydrogen seen against the radio background radiation, which is usually assumed to be the Cosmic Microwave Background (CMB) [29, 34].

Upper limits on the 21-cm signal from the Epoch of Reionization (EoR,  $z \sim 6 - 15$ ) are already available measured by both the radiometers [16, 111], which probe the sky-averaged (global) 21-cm signal, and large interferometric arrays [9, 20, 103, 104, 107] targeting fluctuations. These data have recently allowed constraints to be derived on the astrophysical processes at the EoR [6, 69, 70, 91, 93, 105, 106, 109, e.g. chapter 5,] and are in a broad agreement with other probes of reionization history such as high redshift quasars and galaxies [201–205].

The cosmological origin of the EDGES signal at  $z \sim 17$  [8] was recently disputed by the SARAS3 collaboration, who conducted an independent experiment and reported a non-detection of the EDGES best-fit profile in their data [17–19, 54].

The non-detection by SARAS3 of the EDGES profile increases the likelihood of the anomalous absorption feature being non-cosmological and brings the focus back to the more conventional astrophysical scenarios. In this work, we use the SARAS3 data to provide constraints on the astrophysical processes at CD. We consider a potential population of high-redshift radio-luminous galaxies which contribute to the radio background, thus affecting the 21-cm signal. In general, radio galaxies are expected in standard astrophysical scenarios [120, 121], it is only for extremely high values of their radio luminosity that their contribution is large enough to explain the EDGES signal [8, 117–121, 131]. Here, we consider a wide selection of models varying astrophysical properties of high-redshift galaxies over a broad range. We repeat our analysis for two additional scenarios (shown in section 6.S): one with the CMB as the radio background radiation [14] and the other with a phenomenological synchrotron radio background in addition to the CMB [131].

In section 6.2 we discuss in more detail the SARAS3 data. Section 6.3 introduces the different modelled components in our analysis and discusses how we determine constraints on the astrophysical processes at CD, with further details given in section 6.M. Our constraints on high-redshift radio galaxies are discussed in section 6.4. We conclude in section 6.5. Additional astrophysical models are discussed in section 6.S.

Anastasia Fialkov, Eloy de Lera Acedo, Will Handley, Saurabh Singh, Ravi Subrahmanyan and Rennan Barkana provided comments on [7] which formed the basis of this chapter.

## 6.2 Data

SARAS3 is a radiometer based on a monococone antenna that has made observations of the sky from a location in Southern India in the band  $43.75 - 87.5$  MHz, targeting the cosmological 21-cm signal from  $z \sim 15 - 32$  [17–19]. The experiment is the first global 21-cm experiment of its kind to take observations whilst floating on a body of water, which is expected to improve the total efficiency of the antenna. The total efficiency quantifies how the sky radiation is coupled to the antenna, including losses in the local environment of the antenna, such as ground or water beneath the antenna [18]. Further deployment on a lake should prevent the introduction of non-smooth systematics caused by stratified ground emission that can impede the detection of a global signal [6].

15 hours of observations were integrated in the frequency range  $55 - 85$  MHz ( $z \sim 15 - 25$ ), reduced after radio frequency interference filtering, with corrections made for emission from the water beneath the antenna and receiver noise temperature. The data were then appropriately scaled, given an estimate of the total efficiency, to produce an average measurement of the sky temperature,  $T_{\text{sky}}$ , which we expect to be the sum of the Galactic and extra-Galactic foregrounds,  $T_{\text{fg}}$ , and the cosmological 21-cm signal,  $T_{21}$ .

Previously, a log-log polynomial foreground model was fitted to the data in combination with the phenomenological best-fit EDGES absorption profile multiplied by a scale factor,  $s$ , using a Markov Chain Monte Carlo analysis. The data were shown to reject the presence of the EDGES signal (i.e.  $s = 0$ ) with 95.3% confidence and a series of EDGES-like signals, representing the likelihood distributions of uncertainties in the profile parameters, were rejected with a significance of 90.4% [54]. The SARAS3 measurement of the sky-averaged radio spectrum thus represents a non-detection of the EDGES absorption feature, with the potential to constrain astrophysical scenarios that result in signals larger in magnitude than the instrument noise.

## 6.3 The Cosmological 21-cm Signal

To provide constraints on the astrophysical processes at CD using the SARAS3 data, we need to model the global 21-cm signal. To recap, theoretical predictions of the signal are made difficult owing to the non-local impact of the non-uniform radiative fields produced by a distribution of luminous sources. Either numerical or semi-numerical methods are required to calculate the three-dimensional 21-cm signal and evolve it with time. The global signal can then be calculated as the spatial average [11, 12, 14, 195, 206], although one-dimensional radiative transfer codes are also used to calculate the global signal [207]. As astrophysical processes at high redshifts are poorly understood, a range of theoretical predictions for the 21-cm signal need to be computed for different astrophysical scenarios, which can then be constrained by data.

The calculation takes into account important processes that shape the 21-cm signal: The baryonic matter in the early Universe is predominantly composed of atomic hydrogen and, as the first stars and black holes form at  $z \sim 30$  [208, 209], they affect both the total intensity and the fluctuations of the hydrogen signal. Radiation from the first stars in the Lyman band plays a fundamental role as it enables observations of the 21-cm signal against the radio background by coupling the characteristic temperature of the spin-flip transition, the spin temperature  $T_S$ , to the kinetic temperature of the gas,  $T_K$  [39, 40]. At CD, typically, radiation is warmer than the gas, and the signal appears as an absorption feature against the radio background. Heating of the intergalactic medium subsequently raises the gas temperature too and, perhaps, above the radio background [12, 14, 195], resulting in emission at low redshifts. This evolution culminates at the EoR when ultraviolet radiation from stars ionizes

the neutral hydrogen in the intergalactic medium and the signal disappears. If it exists at high redshifts, any additional radio background above the CMB would contribute to the 21-cm signal by increasing  $T_{\text{rad}}$  and, thus, deepening the absorption profile [117, 119–121, 131].

Starting from a set of the simulated signals, we use neural networks [4] to interpolate the astrophysical parameter space (see *Methods* and Tab. 6.M.1 for a discussion of the network training and accuracy).

To extract constraints on the global signal, we also need to model the foreground in the data. We do this using the same log-log polynomial as in the previous analysis of the SARAS3 data [54] for consistency (see section 6.M).

As discussed, the theoretical CD 21-cm signal is sensitive to the process of star formation, thermal history and the temperature of the radio background radiation (see *Methods*). The root mean squared (RMS) of the appropriately weighted SARAS3 residuals after foreground modelling and removal is 213 mK at their native spectrum resolution of 61 kHz [54]. Signals that are within the sensitivity of the instrument and would have been detected are those with deeper absorption troughs that have strong variation within the band. Typically, such signals are created in scenarios with high intensity of the Lyman- $\alpha$  photons, corresponding to a combination of low minimum virial circular velocities  $V_c$  and high star formation efficiencies  $f_*$ , and a strong contrast between the gas temperature and the temperature of the radio background, i.e. for low values of the X-ray production efficiency  $f_X$  and high radio production efficiencies  $f_{\text{radio}}$ . Therefore, we expect the SARAS3 data to constrain these model parameters. In our analysis, we marginalize parameters, including the CMB optical depth,  $\tau$ , and the mean free path of ionizing photons,  $R_{\text{mfp}}$ , that determine the structure of the signal during the EoR because they are not relevant to the SARAS3 band.

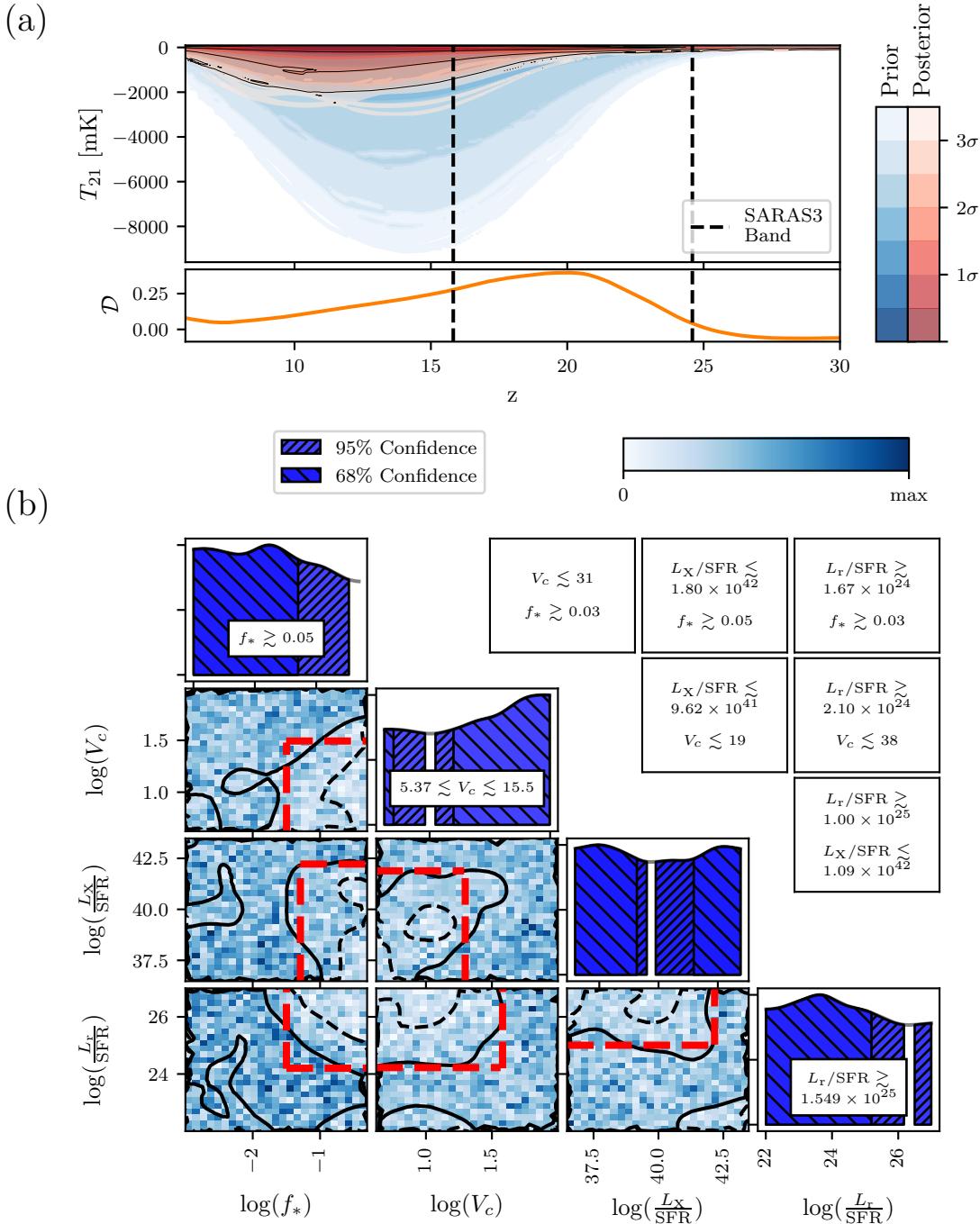
We use the Bayesian Nested Sampling algorithm to perform our model fitting [85] (see chapter 1).

## 6.4 Results

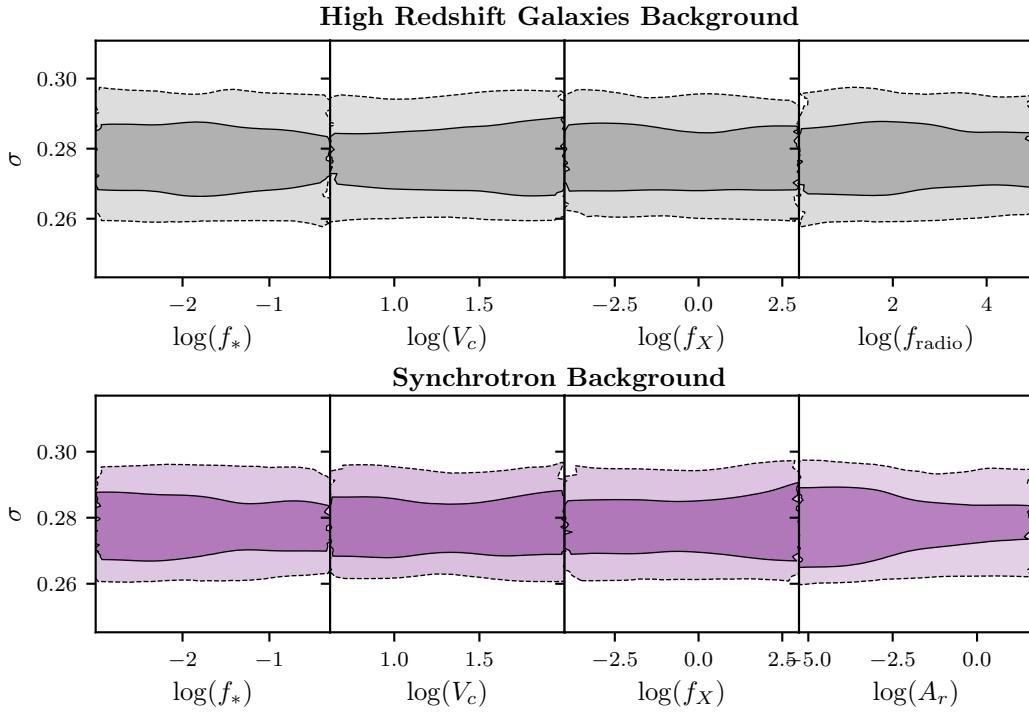
In this section, we discuss the SARAS3 constraints in the redshift range  $z \sim 15 - 25$  on a population of high-redshift radio-luminous galaxies and show the main results in Fig. 6.1. Constraints on models with the CMB-only background and an excess radio background from a phenomenological synchrotron source [131] are discussed in section 6.S.

We calculate the posterior distribution (using the Nested Sampling algorithm discussed in chapter 1), which is a multivariate probability distribution for the thirteen parameters that describe the foreground (seven polynomial coefficients), noise (one parameter, the standard deviation of the Gaussian noise in the data, see section 6.M) and the cosmological 21-cm signal (five astrophysical parameters,  $f_{\text{radio}}$ ,  $f_*$ ,  $V_c$ ,  $f_X$ , and  $\tau$  the CMB optical depth). We then marginalize over the foreground parameters, noise and  $\tau$  which allows us to calculate the likelihood of different astrophysical 21-cm signals. The standard deviation of the noise is shown to be independent of the astrophysical parameters in Fig. 6.2 for excess background models. We subsequently derive limits on the parameters related to star formation, heating, and the excess radio background above the CMB. In section 6.M and section 6.S, we discuss the foreground model in more detail [54]. Although we note here that in fits with both a foreground and 21-cm signal model we found no correlation between the two sets of parameters as can be seen in Fig. 6.3.

We find no evidence for an astrophysical signal in the data. The foreground-only fit consistently has a larger log-evidence by approximately 10 log units when compared to fits with signal profiles. Therefore, any



**Figure 6.1: SARAS3 constraints on high-redshift radio galaxies.** The data disfavour deep global signals, as can be seen by comparing the functional prior (blue) with the posterior (red) in panel (a). At the bottom of this panel we show the Kullback-Leibler (KL) divergence,  $\mathcal{D}$ , as a function of redshift between the functional prior and posterior. The KL divergence gives a measure of the information gain when moving from one to the other and illustrates the constraining power of the SARAS3 data, which peaks at around  $z \approx 20$ . Panel (b) shows the corresponding 1D and 2D posteriors for the astrophysical parameters found when fitting the foreground and a global 21-cm signal to the data. From left to right (see text for details): fraction of gas that turns into stars,  $f_*$ ; circular velocity in units of  $\text{km s}^{-1}$ ,  $V_c$ ; radio luminosity per unit Star Formation Rate (SFR),  $L_r/\text{SFR}$ , in units of  $\text{W Hz}^{-1} \text{M}_\odot^{-1} \text{yr}$  calculated at 150 MHz; X-ray luminosity per unit SFR,  $L_x/\text{SFR}$ , in  $\text{erg s}^{-1} \text{M}_\odot^{-1} \text{yr}$ . The corresponding 1D posterior with marked 68% disfavoured regions are shown in the top of each column. The colour of the 2D posteriors (see colour bar) reflects the magnitude of the 2D posterior probabilities. The dashed black lines encapsulate the 95% confidence regions. The solid black lines show the 68% confidence regions for which we make a conservative approximation with the dashed red lines (to guide the eye) with the corresponding numerical values summarized in the inverted triangle table. Figures produced with ANESTHETIC [178] and FGIVENX [196].

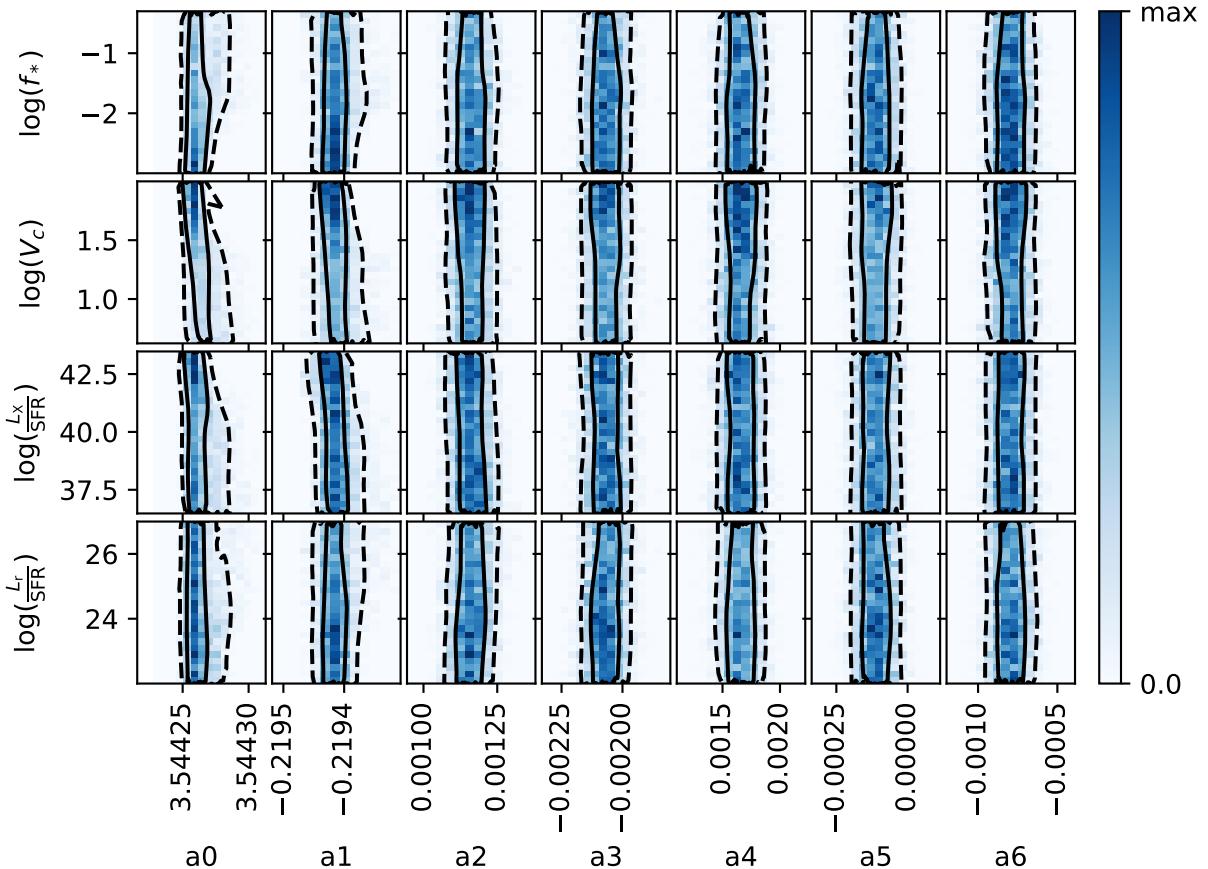


**Figure 6.2: The relationship between the astrophysical parameters and the noise.** The figure shows that the standard deviation of the assumed Gaussian noise is uncorrelated with the astrophysical parameters, and consequently we would expect that a full treatment of any frequency dependence of the noise in the data, which is left for future work, will have little impact on the derived parameter constraints.

cosmological signal present in the data is likely to be undetectable in the residuals (with an RMS of  $\approx 213$  mK), after foreground modelling and subtraction. Since the predicted amplitude of the 21-cm signal is expected to be lower than  $\lesssim 165$  mK [14] in the case of the standard scenario with the CMB as the only source of the background radiation, the SARAS3 constraints on this scenario are very weak (see section 6.S). However, in the case of an excess radio background the predicted signals can be much stronger, allowing us to disfavour regions of the astrophysical parameter space.

Panel (a) of Fig. 6.1 shows the contraction from the initial set of possible 21-cm signals (prior, blue region, details of the astrophysical priors can be found in section 6.M and Tab. 6.1) to the set of scenarios that are allowed by the data (i.e., functional posterior, shown in red). The functional posterior and prior are produced by taking the posterior samples returned from our Nested Sampling run and a representative set of prior samples are transformed into realizations of the global 21-cm signal using the trained neural networks. The figure illustrates that, as anticipated, the deepest signals and the signals with the strongest variation in the SARAS3 band,  $z = 15 - 25$ , are disfavoured. Signals with a modest variation within the band and signals with minima at  $z \lesssim 15$  are indistinguishable from the foregrounds and, thus, cannot be ruled out.

We can estimate the quantitative contribution of the SARAS3 measurement to our understanding of the high-redshift Universe by calculating the information gain using the Kullback-Leibler (KL) divergence [210],  $\mathcal{D}$ , between the functional prior and posterior (bottom of panel (a)). The KL divergence by definition has arbitrary scaling but should always be  $\gtrsim 0$ . We see that  $\mathcal{D}$  is highest, meaning the information gain is largest, at  $z \sim 20$  which corresponds to the middle of the SARAS3 band. Owing to the dependence of the 21-cm signal on the star formation and heating histories, we see that the constraining power of the SARAS3 measurement extends to lower redshifts outside the SARAS3 frequency band (KL divergence is non-vanishing). On the other hand,  $\mathcal{D}$  is approximately zero at  $z \gtrsim 30$  where the global 21-cm signal is dominated by cosmology rather than



**Figure 6.3:** The 2D posterior distributions between the astrophysical and foreground parameters found when fitting the data with the radio galaxy radio background models. We see no clear correlations between the two sets of parameters, indicating that they are independent of each other.

astrophysics.

We next consider the astrophysical parameter constraints and show the corresponding 1D and 2D posteriors in panel (b) of Fig. 6.1. The visualization of the constraints on the signal parameters is non-trivial and is achieved in 2D and 1D via marginalization as in the previous chapter. Marginalization involves integrating out of the posterior distribution the dependence on the  $N - 1$  or  $N - 2$  parameters to leave 1D and 2D distributions for the astrophysical parameters. Key numerical results and comparison with SARAS2 [6] and HERA [109] constraints are summarized in Tab. 6.2. To guide the eye, we show the approximate 2D constraints (red dashed lines) roughly corresponding to the 68% confidence contours (solid black lines) and list these limits in Fig. 6.1 (inverted triangle table). Note that there are regions of low probability outside these guides. From the figure it is clear that SARAS3 data most strongly constrain the process of primordial star formation (clear trends in the 1D posterior probabilities of  $V_c$  and  $f_*$ ) and the strength of the radio background (limits on the radio luminosity per star formation rate at 150 MHz,  $L_r/\text{SFR}$ ) with somewhat weaker sensitivity to X-ray heating, which, however, is clearly constrained in combination with the strong radio background.

Parameter	Radio Background	Range
$f_*$	CMB Only, Synchrotron, Radio Galaxies	0.001 - 0.5
$V_c$	CMB Only, Synchrotron, Radio Galaxies	4.2 - 100 km/s
$f_X$	CMB Only, Synchrotron, Radio Galaxies	0.001 - 1000
$f_{\text{radio}}$	Radio Galaxies	1.0 - 99,500
$A_r^{1420}$	Synchrotron	0 - 47
$\tau$	CMB Only	0.026 - 0.103
	Synchrotron	0.016 - 0.158
	Radio Galaxies	0.035 - 0.077
$\alpha$	CMB Only	1.0 - 1.5
$E_{\min}$	CMB Only	0.1 - 3.0 keV
$R_{\text{mfp}}$	CMB Only, Synchrotron, Radio Galaxies	Fixed at 30, 40 and 40 Mpc

**Table 6.1: The astrophysical priors.** The prior ranges on the parameters for the CMB only, synchrotron and high-redshift radio galaxy background global 21-cm signal models fitted in this paper. The definitions of the parameters are given in the text. The prior ranges are designed to encompass the current uncertainty in the properties of the high-redshift Universe. The emulators are unreliable outside these bounds. Note that  $\tau$  is not an important parameter in the SARAS3 band, however, we train the models with this parameter as an input, perform fits with it and then marginalize over it. Similarly,  $R_{\text{mfp}}$  is only important at lower redshifts outside the SARAS3 band. The global signal only has a weak dependence on this parameter, and so we fix its value at 40 Mpc for the radio galaxies and radio synchrotron backgrounds, while in the CMB-only case it was fixed to  $R_{\text{mfp}} = 30$  Mpc.

	SARAS3	HERA	SARAS2
Signal type	Global	Power Spectrum	Global
Redshift range	$z \approx 15 - 25$	$z \approx 8 \text{ and } \approx 10$	$z \approx 7 - 12$
$L_r/\text{SFR}$	$\gtrsim 1.549 \times 10^{25}$	$\gtrsim 4.00 \times 10^{24}$	—
$L_r/\text{SFR} \cap L_X/\text{SFR}$	$\gtrsim 1 \times 10^{25} \cap \lesssim 1.09 \times 10^{42}$	$\gtrsim 4.00 \times 10^{24} \cap \lesssim 7.60 \times 10^{39}$	$\gtrsim 4.07 \times 10^{24} \cap \lesssim 6.3 \times 10^{39}$
$M$	$4.4 \times 10^5 \lesssim M \lesssim 1.1 \times 10^7$	—	—
$f_*$	$\gtrsim 0.05$	—	—
$f_* \cap M$	$\gtrsim 0.03 \cap \lesssim 8.53 \times 10^8$	—	—

**Table 6.2: Summary of key constraints from SARAS3 (this work), early results from HERA [109] and SARAS2 [6] experiments.** We specify the signal type measured by each instrument (either global signal or power spectrum); redshift range targeted by each experiment; constraints on the value of  $L_r/\text{SFR}$  expressed in units of  $\text{W Hz}^{-1}\text{M}_\odot^{-1}$  yr at 150 MHz; limits on  $L_r/\text{SFR}$  in combination with  $L_X/\text{SFR}$  (calculated between 0.2 and 95 keV and expressed in units of  $\text{erg s}^{-1}\text{M}_\odot^{-1}$  yr); limits on the mass of star forming halos,  $M$ , given in solar masses at  $z = 20$ ; star formation efficiency  $f_*$  and, finally, constraint on  $f_*$  in combination with the halo mass. Limits on the individual parameters correspond to the regions that are disfavoured (with 68% confidence) in the 1D posteriors, combined constraints approximately correspond to the 68% confidence limits in the 2D posteriors. Note that: SARAS2 is unable to constrain individual parameters; HERA targets the power spectrum in comparison to the two SARAS experiments; SARAS3 is at much higher redshifts than the other two experiments; while HERA provides individual constraints on  $L_r/\text{SFR}$  and  $L_X/\text{SFR}$  [109], here we only quote the individual constraint on  $L_r/\text{SFR}$  and the combined constraint with  $L_X/\text{SFR}$ , which is done to ease the comparison with SARAS3 and SARAS2.

From the 1D posterior distribution, we see that the data constrain the radio production efficiency of the early sources, with the values of  $L_r/\text{SFR} \gtrsim 1.549 \times 10^{25} \text{ W Hz}^{-1}\text{M}_\odot^{-1} \text{ yr}$  at 150 MHz ( $f_{\text{radio}} \geq 1549$ ) being disfavoured at 68% confidence. Moreover, we expect a significant correlation between the impact of radio background and that of the thermal history on the global 21-cm signal, as both a strong radio emission and weak X-ray heating contribute in the same direction deepening the absorption trough. Considering the 2D posterior probability in the plane  $L_r/\text{SFR} - L_X/\text{SFR}$  and the corresponding approximate contours in red, we see that high-redshift galaxies that are both efficient in producing radio photons with  $L_r/\text{SFR} \gtrsim 1.00 \times 10^{25} \text{ W Hz}^{-1}\text{M}_\odot^{-1} \text{ yr}$  and inefficient at producing X-ray photons with an X-ray luminosity per star formation rate of  $L_X/\text{SFR} \lesssim 1.09 \times 10^{42} \text{ erg s}^{-1}\text{M}_\odot^{-1} \text{ yr}$  are disfavoured.  $L_r/\text{SFR}$  and  $L_X/\text{SFR}$  are proportional to  $f_{\text{radio}}$  and  $f_X$  respectively and defined in *Methods*.

The data disfavour (at 68% confidence) models with early onset of efficient star formation which is characterized by low values of  $5.37 \lesssim V_c \lesssim 15.5 \text{ kms}^{-1}$  (note that the lower limit of the prior range is  $V_c = 4.2 \text{ kms}^{-1}$ ), corresponding to small typical dark matter halos of  $4.4 \times 10^5 \text{ M}_\odot \lesssim M \lesssim 1.1 \times 10^7 \text{ M}_\odot$  at  $z = 20$  [e.g. 121], and high values of star formation efficiency  $f_* \gtrsim 0.05$ , interpreted as a large fraction of collapsed gas that turns into stars. Each one of these criteria individually, as well as their combination, would guarantee efficient Lyman- $\alpha$  coupling, resulting in a deep high-redshift absorption profile. Considering the 2D posterior distribution, we find, using the approximate red contours on Fig. 6.1, that  $f_* \gtrsim 0.03$  together with galaxies hosted in dark matter halos of masses  $M \lesssim 8.53 \times 10^8 \text{ M}_\odot$  at  $z = 20$  ( $V_c \lesssim 31 \text{ km/s}$ ) are disfavoured. We also find that combinations of high  $f_*$  (and low  $V_c$ ) with both low X-ray efficiency and high radio efficiency are disfavoured. We note that when fitting the data with the phenomenological synchrotron radio background model, we disfavour similar combinations of  $V_c$  and  $f_*$  (see *Supplementary Material*).

The derived constraints can be compared to the recently published constraints from SARAS2 [e.g. chapter 5, 6], see Tab. 6.2 and Fig. 6.S.2 in section 6.S. SARAS2 probes a lower redshift range,  $z = 7 - 12$ , and, thus, is complementary to SARAS3, being more sensitive to the process of heating and ionization. However, the experiment has a comparatively low signal-to-noise ratio, meaning that any constraints derived from it are likely to be weaker. For example, one particular signal may have a magnitude lower than the SARAS3 noise floor but higher than the SARAS2 noise floor. This means that were that particular signal real, SARAS3 would have detected it but SARAS2 would not of, and hence, given the non-detection in the SARAS3 data, the corresponding combination of astrophysical parameters will produce a lower posterior probability for SARAS3,  $\mathcal{P}(\theta|D_{\text{SARAS3}}, M_{\text{SARAS3}})$ , than SARAS2,  $\mathcal{P}(\theta|D_{\text{SARAS2}}, M_{\text{SARAS2}})$  (see *Methods* for a discussion on the posterior probability). Previously, it was found that SARAS2 disfavours (at approximately 68% confidence) early galaxies with X-ray luminosity of  $L_X/\text{SFR} \lesssim 6.3 \times 10^{39} \text{ erg s}^{-1}\text{M}_\odot^{-1} \text{ yr}$  in combination with  $L_r/\text{SFR} \gtrsim 4.07 \times 10^{24} \text{ W Hz}^{-1}\text{M}_\odot^{-1} \text{ yr}$  [6]. This corresponds to disfavouring  $\approx 23\%$  of the available parameter space in the  $L_X - L_r$  plane at approximately 68% confidence compared to  $\approx 32\%$  for SARAS3, although we note that both experiments disfavour slightly different regions of the parameter space.

The same set of astrophysical models, used here, has recently been constrained with an upper limit on the 21-cm power spectrum measured by HERA [109]. HERA disfavours at 68% confidence level values of  $L_r/\text{SFR} \gtrsim 4 \times 10^{24} \text{ W Hz}^{-1}\text{M}_\odot^{-1} \text{ yr}$  as well as  $L_X/\text{SFR} \lesssim 7.6 \times 10^{39} \text{ erg s}^{-1}\text{M}_\odot^{-1} \text{ yr}$ . We find that SARAS3 provides a similar constraint in the plane  $L_r/\text{SFR}$  and  $L_X/\text{SFR}$  with a weaker limit on  $L_r/\text{SFR}$  but a stronger limit on  $L_X/\text{SFR}$  than HERA. Similarly, SARAS2 gives a comparable constraint in the  $L_r/\text{SFR} - L_X/\text{SFR}$  plane. We note that both SARAS2 and HERA probe the 21-cm signal at much lower redshifts than SARAS3,

thus the experiments potentially probe different populations of sources. Moreover, HERA constraints come from the limit on the 21-cm power spectrum, rather than the global signal. Constraints on the 21-cm power spectrum are also available from the MWA [93, 107] and LOFAR [104–106] interferometers. However, these limits are slightly weaker than those from HERA. We discuss this further in chapter 7.

In the context of verifying the EDGES Low-Band detection, we assess the constraining power of the SARAS3 data on physical models that could, in principle, describe the reported absorption feature. Although, we note that none of our models can fit the flattened EDGES absorption signal well. We define EDGES-like signals using a conditional equation that ensures the models have approximately the same central frequency, width and depth as the EDGES absorption feature but does not strictly enforce the flattened Gaussian shape of the EDGES profile [14, 131]. In our analysis so far, the broad prior range was determined by our poor understanding of the high-redshift Universe. Now we use the restricted EDGES-like space as our prior, which is shown in Fig. 6.S.6. We perform the fitting procedure and penalize models that do not meet the EDGES-like criteria by setting the likelihood to zero. The volume contraction from prior to posterior gives a quantitative measure of the level of consistency between the EDGES-like prior and the SARAS3 data and can be estimated using a marginal KL divergence [210]. This effectively allows us to say that if EDGES is true and indicative of a physical scenario, then a given percentage of the physical EDGES-like parameters space is inconsistent or ruled out by the SARAS3 data. We find that the volume of the EDGES-like posterior, when fitting with the radio galaxy models, is 60% of the EDGES-like prior volume. In other words, 60% of the physical EDGES-like parameter space is consistent with the SARAS3 data. See section 6.M and section 6.S for details.

Finally, we find that the data provide interesting limits on the amplitude of the synchrotron radio background in excess of the CMB disfavouring contributions of  $\gtrsim 6\%$  at a reference frequency of 1.42 GHz with 68% confidence. The constraints from SARAS3 can be compared to the excess backgrounds inferred from ARCADE2 [48] and LWA [49] experiments, assuming that the excess is cosmological and is not due to incorrect calibration of the Galactic foregrounds [50]. We find that the 68% confidence limit on  $T_{\text{rad}}$  is significantly lower than the reported deductions from the two experiments (see section 6.S).

## 6.5 Conclusion

We provide astrophysical constraints on the Universe at  $z \sim 20$ , corresponding to  $\sim 200$  million years after the Big Bang, using upper limits on the sky-averaged 21-cm signal measured by the SARAS3 radiometer in the frequency range 55 and 85 MHz,  $z \sim 15 - 25$ . These are the first astrophysical limits of their kind. The only other existing constraining data (from EDGES) revealed a controversial flattened absorption profile, which is awaiting verification by an independent experiment. The residuals observed in SARAS3 data, after modelling for foregrounds, do not provide evidence for a detected 21-cm signal, including the EDGES profile, and they allow for the first time constraints of astrophysics at cosmic dawn.

We fit the data with a log-log polynomial foreground model, as in the original SARAS3 data analysis paper, together with astrophysically motivated models for the global 21-cm signal, showing that deep global signals are disfavoured by the data. These constraints are then mapped into the astrophysical parameter space using a fully Bayesian analysis. We find that the SARAS3 data provide constraints on the processes that are linked to the formation of first stars and galaxies, production of radio photons at high redshifts as well as heating of the intergalactic medium. We disfavour, at 68% confidence, a population of radio galaxies with luminosity per

SFR of  $L_r/\text{SFR} \gtrsim 1.549 \times 10^{25} \text{ W Hz}^{-1}\text{M}_\odot^{-1}$  yr at 150 MHz, i.e. a factor of a thousand brighter than their low-redshift counterparts, and a synchrotron radio background in excess of the CMB of  $\gtrsim 6\%$  at 1.42 GHz. We also find correlation between the constraints on the radio background and on the thermal history of the global 21-cm signal, showing that galaxies which are both luminous in the radio band and inefficient at producing X-ray photons are disfavoured. Finally, the non-detection of the 21-cm signal in the SARAS3 data can be used to derive constraints on the properties of the first star forming regions. We find that, as an approximation to the 68% confidence constraint, the data disfavour efficient star formation at high redshifts with a minimum mass of star forming halos of  $M \lesssim 8.53 \times 10^8 \text{ M}_\odot$  at  $z = 20$  in which  $\gtrsim 3\%$  of the gas is converted into stars.

Lessons learned from the SARAS3 analysis can be contrasted with those from other instruments, specifically with the EDGES Low-Band detection at  $z \sim 17$  as well as the astrophysical limits derived from the SARAS2 data at  $z \sim 7 - 12$  and the limits from HERA on the 21-cm power spectrum at  $z \sim 8 - 10$ , MWA at  $z \approx 6.7 - 8.5$  [93, 107] and LOFAR at  $z \approx 9$  [104, 106]. For example, by conditioning the prior parameter space to be compatible with the EDGES detection and neglecting the steep walls of the feature, we find that  $\sim 60\%$  of the available parameter space is still consistent with the SARAS3 data.

Although the SARAS3 constraints on the high-redshift astrophysical processes are weak, the analysis presented here demonstrates the potential of the 21-cm line as a probe for the early Universe. The cosmic dawn constraints are expected to tighten in the next few years as the new low-frequency 21-cm experiments are coming online [10, 57, 211].

## 6.M Methods

### 6.M.1 Noise Modelling

Throughout our analysis we assume a Gaussian likelihood function,  $\mathcal{L}$ , and a Gaussian noise distribution with a constant standard deviation,  $\sigma$ ,

$$\log \mathcal{L} = \sum_i \left( -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \left( \frac{T_D(\nu_i) - T_{fg}(\nu_i) - T_{21}(\nu_i)}{\sigma} \right)^2 \right), \quad (6.1)$$

where  $T_D$  is the SARAS3 data,  $T_{21}$  is the global 21-cm signal model and  $T_{fg}$  is the foreground model. In practice, the noise in a global 21-cm experiment is expected to be larger at low frequencies and decrease with increasing frequencies, following the general trend of the sky temperature. A full treatment of any frequency dependence in the noise is left for future work. However, we find, see Fig. 6.2, that the posterior probability for the constant standard deviation on the assumed Gaussian noise is uncorrelated with the astrophysical parameters, and we would therefore expect a full treatment of the noise to have little impact on the derived parameter constraints for the two excess background models. We expect a full treatment of the noise to be more important in future experiments that provide tighter constraints on the astrophysical process during CD.

### 6.M.2 Foreground Modelling

The foreground model used here is identical to the one employed in the original SARAS3 analysis [54]. The log-log polynomial foreground model is given by

$$\log_{10} T_{fg} = \sum_{i=0}^{i=6} a_i (\mathcal{R}(\log_{10} \nu))^i, \quad (6.2)$$

where  $a_i$  are the fitted coefficients,  $\mathcal{R}$  is a normalizing function that scales its argument,  $\log_{10} \nu$ , linearly between -1 and +1 and  $\nu$  is frequency in MHz. When fitting the model with POLYCHORD we provide a uniform prior of -10 to 10 on each of the foreground model coefficients,  $a_i$ . In addition to the foregrounds, the model is designed to account for any residual systematics from the calibration process.

### 6.M.3 Signal Modelling and Emulation

At the high redshifts of CD the most important factors that drive the 21-cm signal are the intensity of the Lyman- $\alpha$  background which determines the efficiency of the coupling between  $T_S$  and  $T_K$ , the temperature of the radio background,  $T_{\text{rad}}$ , and the thermal history of the gas. The dependence of the 21-cm signal on these processes is as follows: The earlier star formation starts, the lower will be the frequency of the absorption profile; the stronger the Lyman- $\alpha$  background, the steeper and deeper will be the resulting 21-cm signal; the colder is the gas, relative to the background radiation, the deeper will be the absorption feature. The resulting 21-cm signal can be written as

$$T_{21} = \frac{T_S - T_{\text{rad}}}{1+z} [1 - \exp(-\tau_{21})] \propto 1 - \frac{T_{\text{rad}}}{T_S}, \quad (6.3)$$

where we assumed that the Universe is largely neutral at the high redshifts of CD.

One potential source of radio photons at CD are early radio galaxies [120]. The radio background contribution created by such sources is proportional to the star formation rate (SFR), thus increasing with time, and is non-uniform following the distribution of galaxies. The radio luminosity spectrum as a function of frequency  $\nu$  produced by a star forming region and calculated per SFR in units of  $\text{W Hz}^{-1}$  is given by

$$L_r = f_{\text{radio}} 10^{22} \left( \frac{\nu}{150 \text{ MHz}} \right)^{-\alpha_{\text{radio}}} \frac{\text{SFR}}{\text{M}_\odot \text{yr}^{-1}}, \quad (6.4)$$

where  $f_{\text{radio}}$  is an efficiency factor that measures radio photon production in high-redshift galaxies compared to their present day counterparts and  $\alpha_{\text{radio}} = 0.7$  is the spectral index in the radio band [121]. The temperature of the radio background produced by such galaxies at redshift  $z$  is calculated by integrating over the contribution of all galaxies within the past light-cone [121] and is added to the temperature of the CMB to give the total radio background temperature. We quote constraints on the radio luminosity per SFR,  $L_r/\text{SFR}$ , at a reference frequency of 150 MHz in section 6.4.

We take into account several heating and cooling mechanisms, such as cooling due to the expansion of the Universe and heating due to structure formation, Lyman- $\alpha$  [29, 42] and CMB [44] heating, as well as heating by first X-ray binaries [195]. In our model, the first four effects are fully determined by cosmology and star formation, whereas heating by X-ray binaries invokes new astrophysical processes (such as black hole binary formation and X-ray production by the high redshift sources). Therefore, X-ray heating requires independent parameterization, and we model X-ray luminosity per star formation rate [46] as

$$L_{X,0.2-95\text{keV}} = 3 \times 10^{40} f_X \frac{\text{SFR}}{\text{M}_\odot \text{yr}^{-1}} \quad (6.5)$$

calculated in units of  $\text{erg s}^{-1}$  between 0.2 and 95 keV, where  $f_X$  is the efficiency of X-ray photon production. Gas thermal history is then evaluated at every redshift by integrating over the contribution of all galaxies within the past light-cone to find the corresponding heating rate and then solving a differential equation to evolve the gas temperature.

Both the radio luminosity and the thermal history of the gas depend on the SFR, which is not well constrained for early galaxies. Therefore, our model also includes free parameters that regulate star formation. One is the star formation efficiency of high-redshift galaxies,  $f_*$ , which measures the fraction of collapsed gas in star forming regions that turns into stars, and the other is the minimum mass of star forming halos, or, equivalently, the minimum circular velocity of star forming halos,  $V_c$  [41]. This quantity depends on the local environment of each star forming region and is affected by factors such as the local intensity of the radiative background in the Lyman-Werner band [206, 212] or the relative velocity between dark matter and gas [208, 212, 213].

In order to physically model the global 21-cm signal, we rely on neural network-based emulation with the PYTHON package GLOBALEMU [e.g. chapter 3, 4] trained on the results of the full semi-numerical simulations of the global 21-cm signal [11–14, 121, 131].

For each global signal model, we have a series of testing and training signals. Tab. 6.1 shows the ranges of the parameters in each of the training and testing data sets for the different models of the global 21-cm signals used in this work. The boundaries correspond to the broadest possible ranges allowed for each one of the parameters from the astrophysical principles and existing (weak) observational constraints. Outside these ranges the emulators are unreliable and consequently the ranges act as the prior bounds for the Nested Sampling code POLYCHORD. The parameters are sampled either uniformly or log-uniformly between the ranges in the training and test data, and we use appropriate prior probability distributions for each parameter when running the fits.

For all three signal emulators, we use the same neural network architecture with 4 hidden layers of 16 nodes each. The same radio galaxies and CMB only emulators were used in the analysis of the SARAS2 data [e.g. chapter 5, 6]. We note that the network for the CMB only radio background models, however, has seven astrophysical inputs compared to the radio galaxy and synchrotron radio background networks which both have five. For the CMB only model the X-ray spectral energy density (SED) is characterized by the slope of the spectrum,  $\alpha$ , and a low energy cut-off,  $E_{\min}$ ; while for the other two models the X-ray SED is fixed to that of high-redshift X-ray binaries [46]. Further, parameters related to reionization have very modest effect on the 21-cm signal in the SARAS3 range. Therefore, we fix the mean free path of ionizing photons. For the radio galaxies and radio synchrotron backgrounds, the mean free path was fixed to 40 Mpc, while in the CMB-only case it was fixed to  $R_{\text{mfp}} = 30$  Mpc.

We assess the accuracy of the neural networks in the SARAS3 band,  $z \approx 15 - 25$ , using the root mean squared error (RMSE) when emulating the test data after training. The synchrotron radio background network has a mean RMSE when emulating 1034 test models, after training on 9304 models, of 7.98 mK, a 95<sup>th</sup> percentile RMSE of 23.06 mK and a worst RMSE of 85.65 mK. The CMB only background network is trained on 5137 models and tested on 570 models. In the SARAS3 band the mean RMSE for the test data is 0.78 mK, the 95<sup>th</sup> percentile is 2.67 mK and the worst is 13.36 mK. Finally, when trained on a data set of 4311 models and tested on 479 models the radio galaxy radio background neural network emulator is found to have a mean RMSE of 5.11 mK, a 95<sup>th</sup> percentile RMSE of 20.53 mK and a worst RMSE of 81.70 mK in the SARAS3 band. All the trained networks have 95<sup>th</sup> percentile RMSEs well below the RMS found after modelling and subtracting the log-log polynomial foreground model from the SARAS3 data. The numbers are summarized in Tab. 6.M.1.

When fitting all three signal models with the foreground, we see no correlation between the astrophysical parameters and the foreground parameters. An example of this can be seen in the 2D posteriors, which are shown in Fig. 6.3, between the astrophysical and foreground parameters from the fit with the radio galaxies

background model.

Background	Radio Galaxies	Synchrotron	CMB Only
Training Models	4311	9304	5137
Testing Models	479	1034	570
Mean RMSE	5.11	7.98	0.78
95 Percentile RMSE	20.53	23.06	2.67
Worst RMSE	81.70	85.65	13.36

**Table 6.M.1: The neural network emulation.** The table summarizes the number of models used to train and test the three different neural network emulators used in this paper, along with summary statistics for their accuracies. Temperatures RMSEs are given in mK.

#### 6.M.4 Marginal KL Divergence and EDGES-like signals

To determine the volume of the EDGES-like parameter space that the SARAS3 data rules out, we calculate a marginal Kullback-Leibler (KL) divergence,  $\mathcal{D}$ . To illustrate the types of signals that we are selecting by constraining our parameter space to be EDGES-like, we show the corresponding functional prior and posterior in section 6.S.

The KL divergence is a measure of the information gained when contracting a prior onto a posterior. For our purposes we are interested in the EDGES-like parameter space and as a result we would consider the foreground parameters to be nuisance parameters and need to integrate them out.

To calculate the marginal KL divergence,  $\mathcal{D}$ , we therefore need to evaluate the log-probabilities associated with the signal parameters in the EDGES-like prior,  $\pi$ , and the corresponding posterior,  $\mathcal{P}$ ,

$$\mathcal{D}(\mathcal{P}||\pi) = \int \mathcal{P}(\theta) \log_e \left( \frac{\mathcal{P}(\theta)}{\pi(\theta)} \right) d\theta = \left\langle \log_e \left( \frac{\mathcal{P}(\theta)}{\pi(\theta)} \right) \right\rangle_{\mathcal{P}}. \quad (6.6)$$

We use a Gaussian kernel density estimators (KDE) to replicate the samples in the signal sub-spaces of our EDGES-like prior and posterior using MARGARINE discussed in chapter 4. A multivariate Gaussian KDE, implemented with SCIPY, is produced by summing over multiple multivariate Gaussian profiles with known standard deviation centred around each sample point, consequently the log of the probability density function is easily tractable making the KL divergence easily calculable.

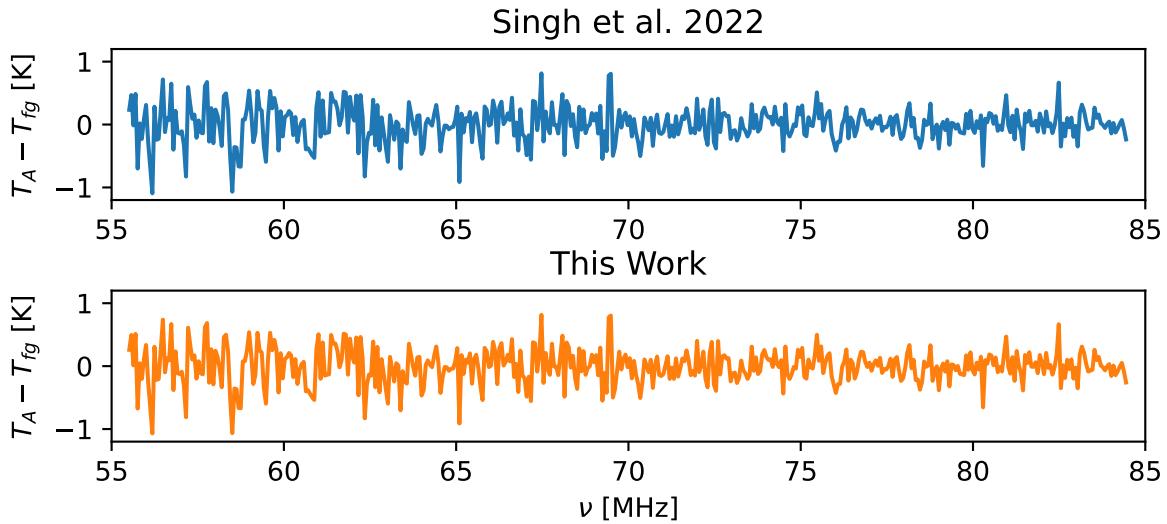
It can be shown that the KL divergence is related to the volume fraction between the posterior and the prior via

$$\mathcal{D}(\mathcal{P}||\pi) \approx \log_e \left( \frac{V_{\pi}}{V_{\mathcal{P}}} \right), \quad (6.7)$$

and therefore

$$\exp(-\mathcal{D}(\mathcal{P}||\pi)) \approx \frac{V_{\pi}}{V_{\mathcal{P}}}, \quad (6.8)$$

can be used to determine the volume of the prior contained in the posterior or in our case the volume of the EDGES-like prior that is still consistent with the SARAS3 data after fitting.



**Figure 6.S.1:** A comparison of foreground models used here and in the original analysis. The figure shows the residuals found when subtracting the foreground model parameterized by the MCMC run in the original SARAS3 analysis in comparison to the equivalent using the parameterization from our Nested Sampling run. We see that the two sets of residuals are consistent with each other, showing that our implementation of the foreground is consistent with the previously implemented MCMC run.

## 6.S Supplementary Material

### 6.S.1 Foreground Modelling: Comparison With Previous Work

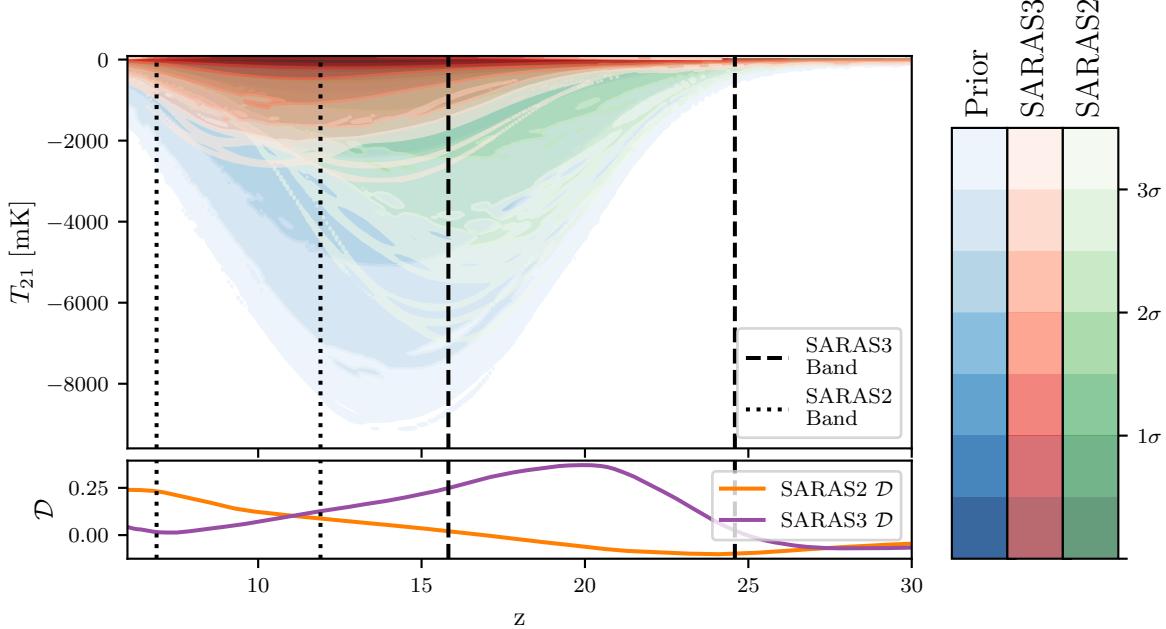
In Fig. 6.S.1, we show the residuals found when subtracting the foreground model parameterized by the Nested Sampling results in this work and parameterized by the MCMC run in the original work [54]. The two sets of residuals are found to be consistent with each other, indicating that our implementation of the foreground modelling is also consistent with that in the original SARAS3 analysis.

### 6.S.2 Comparison of Constraints on the High-redshift Radio Galaxy Model from SARAS2 and SARAS3

We note that, previous analysis of the SARAS2 data used similar techniques to those detailed in this paper and fitted the SARAS2 data, in the range  $z \approx 7 - 12$ , with the radio galaxy global signal models, maximally smooth foreground models and models for non-smooth systematics (chapter 5 and [6]). Although the signal-to-noise ratio is lower for the SARAS2 data, the analysis disfavoured the deepest signals corresponding to the combinations of parameters: high  $f_{\text{radio}}$  with high  $f_*$  and low  $f_X$ . Comparison of the functional posterior plots, shown in Fig. 6.S.2, demonstrates that the two experiments are complimentary, with the peak constraining power of SARAS3 at  $z \gtrsim 10$  and that of SARAS2 at  $z \lesssim 10$ .

Looking at the 2D posterior distributions of the astrophysical parameters, we see that SARAS2 disfavours (at approximately 68% confidence) galaxies with  $f_X \lesssim 0.21$  ( $L_X/\text{SFR} \lesssim 6.3 \times 10^{39} \text{ erg s}^{-1} \text{M}_\odot^{-1} \text{ yr}$ ) in combination with  $f_{\text{radio}} \gtrsim 407$  ( $L_r/\text{SFR} \gtrsim 4.07 \times 10^{24} \text{ W Hz}^{-1} \text{M}_\odot^{-1} \text{ yr}$ ). SARAS3 constrains a similar combination of parameters, with  $f_{\text{radio}} \gtrsim 1000$  ( $L_r/\text{SFR} \gtrsim 1.00 \times 10^{25} \text{ W Hz}^{-1} \text{M}_\odot^{-1} \text{ yr}$ ) and  $f_X \lesssim 109$  ( $L_X/\text{SFR} \lesssim 1.09 \times 10^{42} \text{ erg s}^{-1} \text{M}_\odot^{-1} \text{ yr}$ ) being disfavoured at  $\approx 68\%$  confidence. SARAS2 was also found to disfavour galaxies with a combination of  $f_* \gtrsim 0.03$  and  $f_{\text{radio}} \gtrsim 707$  ( $L_r/\text{SFR} \gtrsim 7.07 \times 10^{24} \text{ W Hz}^{-1} \text{M}_\odot^{-1} \text{ yr}$  at 150 MHz) at  $\approx 68\%$  confidence which is again comparable to the constraints from SARAS3,  $f_* \gtrsim 0.03$  and  $f_{\text{radio}} \gtrsim 167$

( $L_r/\text{SFR} \gtrsim 1.67 \times 10^{24} \text{ W Hz}^{-1} \text{M}_\odot^{-1} \text{ yr}$  at 150 MHz), except that SARAS3 again provides a tighter constraint on the radio luminosity. A major difference between the constraining power of the two experiments is that SARAS3 provides constraints on the individual parameters (as can be seen from the 1D posterior distributions) which SARAS2 is not able to do. The constraints of SARAS2 are manifested in higher-dimensional posterior distributions, e.g. in 2D.



**Figure 6.S.2:** A comparison of the constraints from the SARAS2 analysis [6] and the SARAS3 analysis presented here when fitting both data sets with the radio galaxy background models. Note that the two sets of analysis have a number of differences, and so direct comparisons like this have to be made with caution. Namely, the SARAS2 functional posterior is derived from a number of fits to allow for uncertainty in the modelling of the noise and a systematic, whereas the SARAS3 results are derived from a single fit of a foreground plus signal model. Both sets of analysis share the same prior. Nonetheless, we can see from the KL divergence in the bottom panel that the SARAS3 data provides tighter constraints on this astrophysical model at  $z \gtrsim 10$  and the SARAS2 data at  $z \lesssim 10$ .

### 6.S.3 Synchrotron Radio Background Constraints with SARAS3

In addition to our main model, where the excess radio background over the CMB is created by high-redshift radio galaxies, we consider a phenomenological synchrotron radio background [131]. The total temperature of the radio background is

$$T_{\text{rad}} = T_{\text{CMB}} \left[ 1 + A_r^{78} \left( \frac{\nu}{78 \text{MHz}} \right)^\beta \right], \quad (6.9)$$

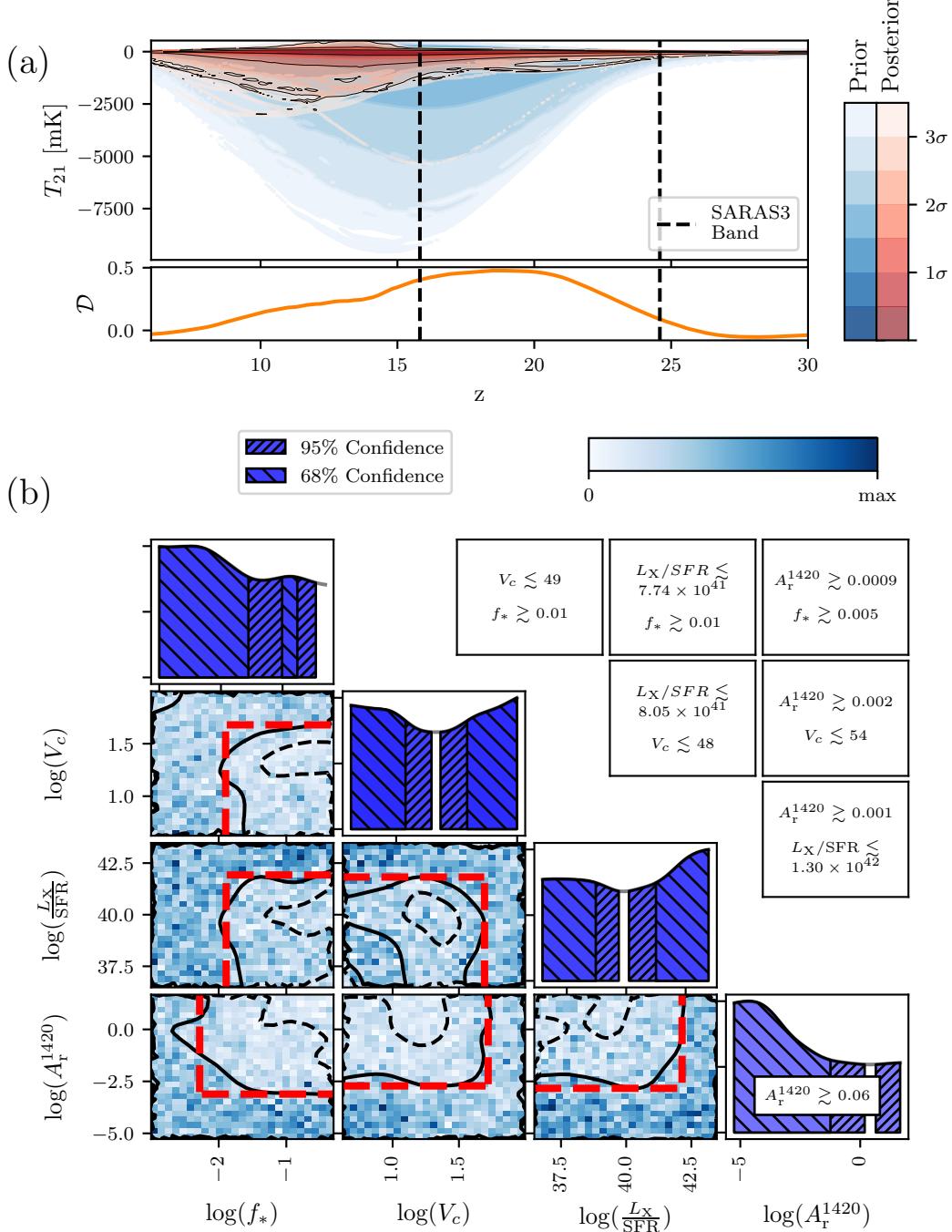
where  $T_{\text{CMB}}$  is the CMB contribution, and we add the synchrotron term with amplitude  $A_r^{78}$  calculated at the reference frequency of 78 MHz.  $\beta = -2.6$  is the power law index which is chosen in agreement with the observations of the radio background today by LWA [49] and ARCADE2 [48] low frequency radio experiments. Throughout this work we also use the value of the amplitude at 1.42 GHz,  $A_r^{1420}$ , and the relationship between the two amplitudes is given as

$$A_r^{78} = A_r^{1420} \left( \frac{0.078}{1.420} \right)^\beta, \quad (6.10)$$

where the superscript gives the reference frequency in MHz.

The two main differences between the radio background created by galaxies and the synchrotron radio background case is that the former is non-uniform and grows with time, following galaxy formation. In contrast,

the latter is assumed to be uniform and decays with time, by construction. The different evolution of the two types of backgrounds is the key feature that leads to distinct astrophysical constraints with the SARAS3 data. We repeat our analysis for the synchrotron radio background model and show the results in Fig. 6.S.3.



**Figure 6.S.3: SARAS3 constraints on early Universe scenarios with a synchrotron radio-background.** We disfavour the deepest synchrotron background signals, and the contraction from the functional prior (blue) to the functional posterior (red) is shown in panel (a) of this figure. We see that the KL divergence  $D \approx 0$  at high redshifts as expected and that the divergence is highest in the SARAS3 band around  $z \approx 20$ . Panel (b) shows that the parameter  $A_r^{1420}$  is tightly constrained to  $\lesssim 0.06$  at 68% confidence and at a reference frequency of 1.42 GHz, corresponding to a radio background in excess of the CMB by  $\lesssim 6\%$ . High values of  $f_*$ , low values of  $V_c$  in combination with high  $A_r$  and similarly low values of  $L_X/SFR$  in combination with high values of  $A_r$  are all disfavoured. Again we show the approximate 2D 68% disfavoured regions in the inverted corner plot and the 68% 1D constraints in their respective sub-panels.

Panel (a) of Fig. 6.S.3 shows the functional prior (blue) for the synchrotron radio background model in comparison to the functional posterior (red) corresponding to the constraints in panel (b). We see that, as with

the radio galaxies background model, the SARAS3 data disfavours the deepest astrophysical models in the band  $z \approx 15 - 25$  but allows for some deeper models with minima at  $z \lesssim 15$ . The corresponding KL divergence,  $\mathcal{D}$ , between the functional prior and posterior is highest at  $z \approx 20$  and vanishes at high redshift where the signals are independent of the astrophysics and are defined by the cosmological model thus sharing a common structure.

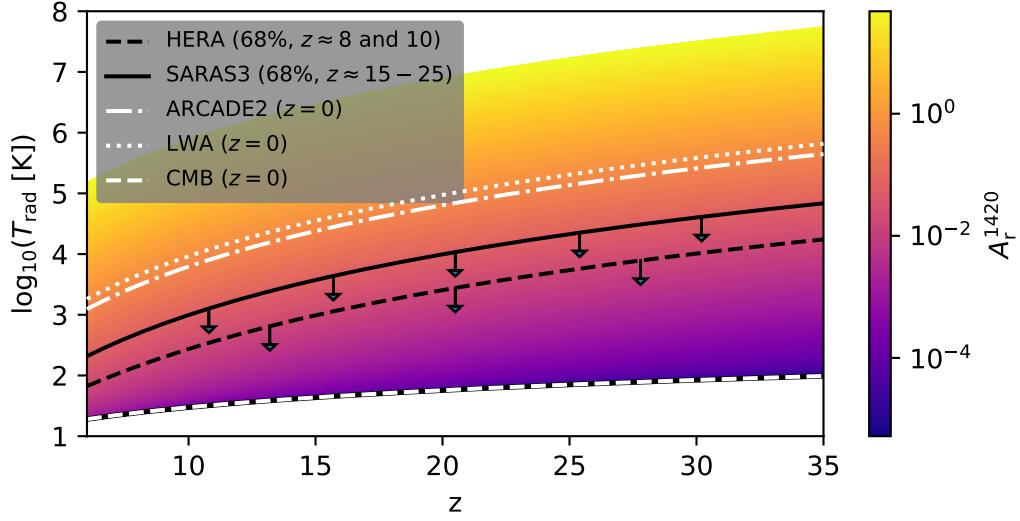
Panel (b) of Fig. 6.S.3 shows the 1D and 2D posterior distributions for the astrophysical parameters  $f_*$ ,  $V_c$ ,  $L_X$  and  $A_r^{1420}$  when fitting the SARAS3 data with the synchrotron radio background signal model and log-log polynomial foreground model. From the 1D posterior probability of the parameters, we find that the SARAS3 data disfavours with 68% confidence a radio background in excess of the CMB by  $\gtrsim 6\%$  at 1.42 GHz. Recent analysis of the HERA power spectrum upper limits using the same set of models disfavoured a radio background in excess of the CMB of  $\gtrsim 1.6\%$  at 1.42 GHz at 68% confidence [109]. This is a stronger constraint compared to that by SARAS3. However, this claim is model-dependent and assumes that the same astrophysical model that describes CD is valid during the EoR. In reality, the population of sources might evolve and, because HERA is observing at much lower redshifts, it might be probing a different population of astrophysical sources.

Concentrating on the 2D posterior probabilities, we find from the 68% confidence levels (solid lines in Fig. 6.S.3) which we approximate by the red dashed lines, that we disfavour different combinations of parameters. Notably, values of  $A_r^{1420} \gtrsim 0.002$  in combination with low values of  $f_X \lesssim 43$  are disfavoured. These values correspond to a radio background in excess of the CMB by  $\gtrsim 0.2\%$  at 1.4 GHz and an X-ray luminosity of  $L_{X,0.2-95\text{kev}} = 1.30 \times 10^{42} \text{ erg s}^{-1}\text{M}_\odot^{-1} \text{ yr}$ . Similarly, we disfavour  $A_r^{1420} \gtrsim 9 \times 10^{-4}$  (an excess of  $\gtrsim 0.09\%$  at 1.42 GHz) in combination with  $f_* \gtrsim 5 \times 10^{-3}$  at approximately 68% confidence and, separately, high  $f_* \gtrsim 0.01$  with low  $V_c \lesssim 49 \text{ km/s}$ . The latter constraint can be translated to the limit on the dark matter halo masses  $M \lesssim 3.37 \times 10^8 \text{ M}_\odot$  at  $z = 20$ .

There are some differences between the constraints for the two considered excess radio background models. For example, the posterior for  $L_X/\text{SFR}$  is flatter when fitting with the high-redshift radio galaxies model in comparison to the synchrotron background model. As highlighted above, the discrepancy can be attributed to the difference in the evolution of the radio backgrounds with redshift. Since, by construction, the synchrotron radio background decays with time, its impact at higher redshifts is stronger than at lower redshifts. In the case of the radio galaxies model, the radio background builds up with time as stellar populations evolve. Therefore, the constraints are weaker at higher redshifts than at lower.

Owing to the simple evolution of the radio background temperature,  $T_{\text{rad}}$ , equation (6.9), we can link the SARAS3 68% constraints at  $z \sim 15 - 25$  to the local measurements of the excess radio background by LWA [49] and ARCAE2 [48] made at  $z = 0$  (note, however, the concerns about the Galactic modelling in the ARCAE2 and LWA analysis [50]). Fig. 6.S.4 shows the evolution of  $T_{\text{rad}}$  as a function of  $z$  and the value of  $A_r^{1420}$ . The solid black line shows the radio background corresponding to the 68% confidence constraint on  $A_r^{1420}$  from the SARAS3 analysis. This limit can be compared to the values observed with ARCAE2 and LWA (shown with white lines). We use the reported values of the radio background at 310 MHz,  $A_r^{310} = 30.4$  from the LWA measurements [49] and  $A_r^{310} = 21.1$  from the ARCAE2 measurements [48], and scale them appropriately to 78 MHz, assuming a spectral index of  $-2.6$ , to calculate the corresponding values of  $T_{\text{rad}}$ . The measurements from ARCAE2 and LWA are then extrapolated from  $z = 0$ , while the SARAS3 constraint is extrapolated outside the band  $z \approx 15 - 25$ . We show the radio background from the CMB ( $A_r^{1420} = 0$ ) for reference. Owing to the simplicity of the adopted model, we can connect observations at different redshifts. In reality, as mentioned above, the different experiments may be probing different physical processes because

they measure the background at different redshifts and this comparison should be taken as a qualitative one. We find that the SARAS3 limit on the strength of the excess radio background is much lower than the background measured by the ARCADE2 and LWA. Therefore, in this scenario, it is unlikely that the ARCADE2/LWA measurement is explained by a contribution of high-redshift radio sources. We also show the constraint from HERA for reference [109].



**Figure 6.S.4:** The radio background temperature,  $T_{\text{rad}}$ , for a set of synchrotron radio background models coloured with respect to their value of  $A_r^{1420}$ . We also show the radio backgrounds corresponding to the 68% confidence limits (black) for SARAS3 extrapolated outside  $z \approx 15 - 25$ , 68% confidence limits (black dashed) from HERA extrapolated from  $z \approx 8$  and  $z \approx 10$ , and the corresponding backgrounds from the ARCADE2 and LWA measurements (white), both extrapolated from  $z = 0$ . We show the CMB temperature for reference.

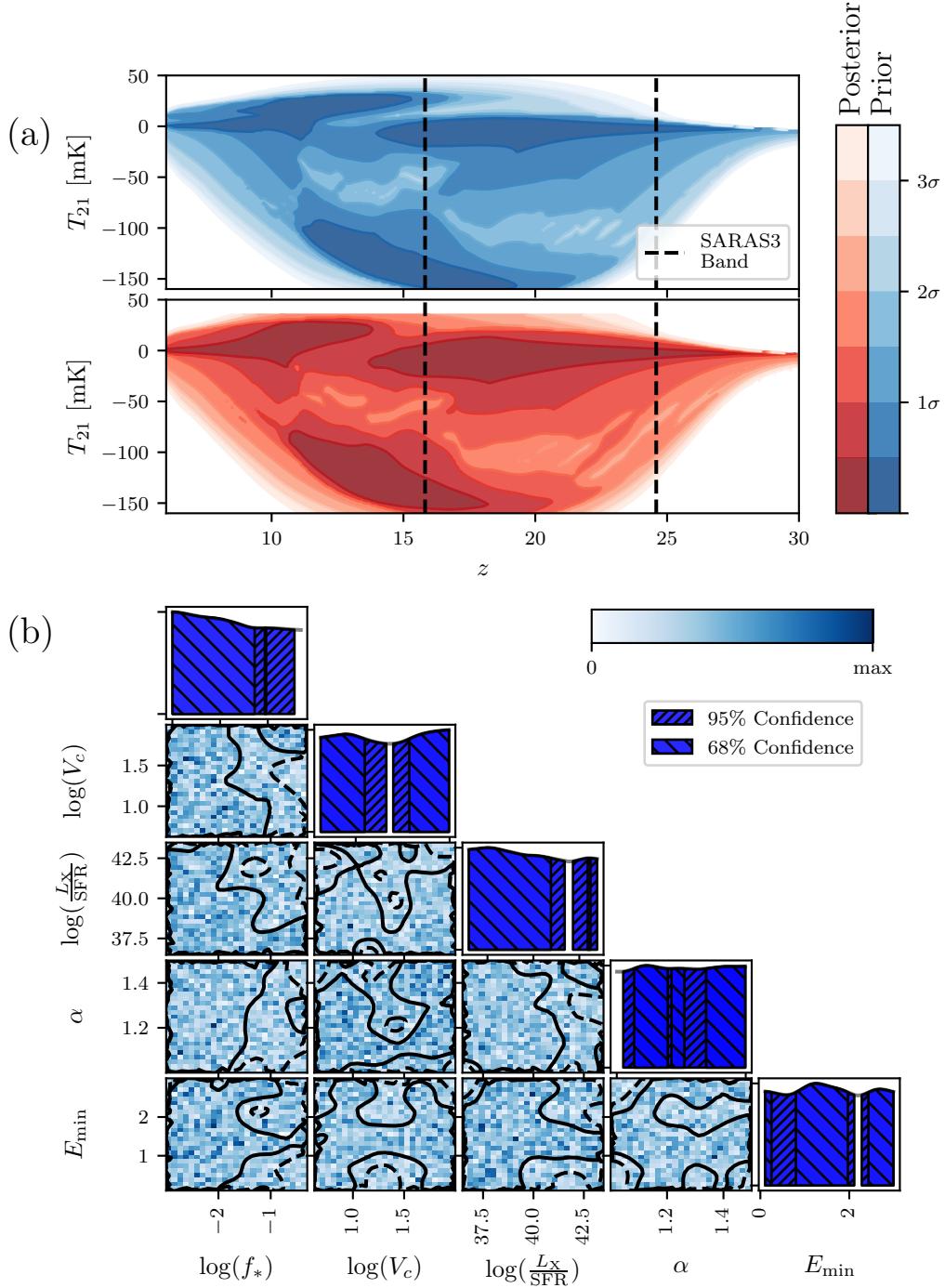
Finally, we use the synchrotron radio background models to check the constraining power of the SARAS3 data in the context of the EDGES Low Band detection. We repeat the analysis described in the main text, fitting the data with EDGES-like physical 21-cm signals with a synchrotron radio background global 21-cm signal in combination with the polynomial foreground model. We find that, when selecting for EDGES-like models, the posterior encapsulates only 57% of the EDGES-like prior volume. Note that as we are implicitly constraining our prior to produce EDGES-like models, in this piece of analysis and the corresponding analysis in the main text we are assuming a high level of accuracy in the EDGES absorption feature. As a result, the analysis is not intended to ‘rule out’ EDGES-like signals but rather to say that if EDGES is taken at face value to be true then 57% of the corresponding physical parameter space is consistent with SARAS3 (or 43% is inconsistent).

#### 6.S.4 CMB Only Radio Background Models

Finally, we consider the ‘standard’ astrophysical case where  $T_{\text{rad}} = T_{\text{CMB}}$ . Since CMB is a rather weak background, compared to the other models that we considered in this work, we expect no strong constraints from the SARAS3 data. The deepest global signals in this model reach  $\sim 165$  mK [14].

Indeed, the functional posterior and prior plots in panel (a) of Fig. 6.S.5 show that there is no significant constraint on the global 21-cm signal for this particular class of models from SARAS3. Further, the corresponding 2D and 1D posteriors, shown in panel (b), exhibit flat distributions and little constraint on the parameter space of the CMB-only models. The models are characterized by a set of astrophysical parameters: the star formation efficiency,  $f_*$ , the minimum circular velocity of star forming halos,  $V_c$ , the X-ray efficiency,  $f_X$ , the slope of the

X-ray SED,  $\alpha$ , the low energy cut-off of the X-ray SED,  $E_{\min}$ , the mean free path of ionizing photons,  $R_{\text{mfp}}$ , and the CMB optical depth,  $\tau$ . Again, as SARAS3 is a high-redshift instrument, we marginalized out  $\tau$  and for these models we fixed  $R_{\text{mfp}}$  to 30 Mpc.

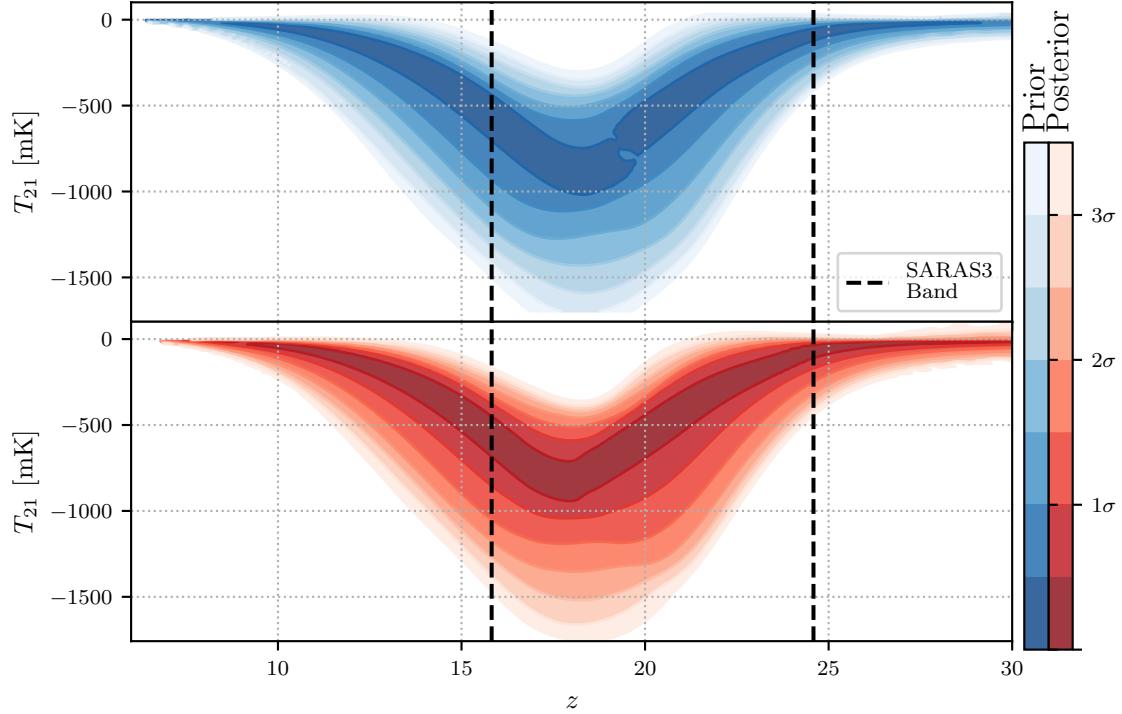


**Figure 6.S.5: SARAS3 constraints on early galaxies assuming a CMB only radio background.** Panel (a) shows the functional prior and posterior found when fitting the SARAS3 data with the CMB only background models, and panel (b) shows the corresponding 1D and 2D posterior distributions for the astrophysical parameters. The distributions are flat and show no constraints, as expected from the magnitude of the SARAS3 residuals after foreground modelling and subtraction.

## 6.S.5 The Functional EDGES-like Posterior and Prior

In Fig. 6.S.6, we show the functional EDGES-like physical posterior (red) and prior (blue) found when fitting the data with the radio galaxies background 21-cm signals. The posterior and prior are used to determine what

percentage of the EDGES-like physical signals, those with similar depths and central frequencies to the reported absorption feature, is consistent with the SARAS3 data. This is discussed in detail in the main text and in section 6.M.



**Figure 6.S.6:** The functional EDGES-like prior and the corresponding posterior found when fitting the data using the radio galaxies background models. The prior is constrained by a conditional equation that selects models with a similar central frequency and depth as the reported EDGES profile [131].

CHAPTER

7

## COMBINING CONSTRAINTS FROM HERA AND SARAS3

## 7.1 Probing the Infant Universe

As discussed, the infant Universe, corresponding to cosmic time between 200 and 700 million years after the Big Bang, remains largely unexplored. This epoch is traced by some of the most remarkable events in cosmic history, such as the birth of primordial stars, formation of the very first black holes and assembly of early galaxies. The nature of the first bright objects to form and the exact timing of these events are yet to be constrained by observations. Theoretical studies and numerical simulations suggest that first stars form between  $z \sim 20 - 60$ , i.e. around 35 – 200 million years after the Big Bang[209, 214, 215], and, thus, are currently out of reach of the modern telescopes like JWST.

Although the controversial EDGES signal is the only detection of the high-redshift 21-cm signal reported to date, advances have been made by other existing radio telescopes, as alluded to in the introduction, with upper limits reported by both the interferometers such as PAPER [Precision Array for Probing the Epoch of Reionization, 216], MWA [107, 217], LOFAR [218, 219], AARTFAAC [Amsterdam-ASTRON Radio Transients Facility And Analysis Center, 220] and HERA [99, 109, 197] on the 21-cm power spectrum, which quantifies the variation in the 21-cm brightness field as a function of angular scale and time, and on the global 21-cm signal by SARAS [6, 7, 69, 70, e.g. the previous two chapters], EDGES High Band [91, 133, 134] and LEDA [132]. These upper limits are becoming increasingly constraining and have already being used to put limits on the properties of high-redshift luminous objects [6, 7, 69, 70, 99, 105, 109]. The ongoing and upcoming experiments including SARAS [54], MIST [114], REACH [10], LOFAR [103], NenuFAR [ New Extension in Nançay Upgrading loFAR, 211], HERA [20], the SKA and future proposed space based missions such as DAPPER and FAR SIDE [38] aim to further improve our understanding of the infant Universe.

Existing upper limits on the 21-cm signal provide the first very weak constraints on the astrophysical sources at a broad range of redshifts. In this chapter, we take the current tightest publicly available upper limits from the HERA interferometer on the magnitude of the 21-cm power spectrum at redshifts  $z \approx 8$  and  $\approx 10$  which provides a window to the EoR when ultraviolet photons emitted by first massive galaxies efficiently ionize the neutral hydrogen gas [109], and the SARAS3 radiometer [54] that probes the sky-averaged 21-cm emission at higher redshifts between  $z \approx 15 - 25$  when the first stars and X-ray emitting objects are expected to have formed in small galaxies during the Cosmic Dawn. We combine these two data sets for the first time to improve constraints on the properties of the first galaxies and the state of the neutral gas. We develop the methods to perform the joint analysis using a machine learning enhanced Bayesian workflow and pave the way for future discoveries. We find that when considered together, these two experiments provide a better leverage on theoretical scenarios that bridge across the wide redshift range compared to the constraints from each individual experiment. In synergy, the two experiments leave only  $64.9_{-0.1}^{+0.3}\%$  of the explored broad theoretical parameter space to be consistent with the joint data set, in comparison to  $92.3_{-0.1}^{+0.3}\%$  for SARAS3 and  $79.0_{-0.2}^{+0.5}\%$  for HERA alone. We use the joint analysis to limit star formation efficiency, minimum halo mass for star formation, X-ray luminosity of early emitters and the radio luminosity of early galaxies. The joint analysis disfavours at 68% confidence a combination of galaxies with X-ray emission that is  $\lesssim 33$  and radio emission that is  $\gtrsim 32$  times as efficient as present day galaxies.

In section 7.2 we review the data and the specifics of the experiments incorporated in our analysis. This is followed by a discussion about the synergies between the power spectrum and sky-averaged 21-cm experiments in the same section. We present the implications of our work for the astrophysical constraints and the validity of the EDGES absorption feature as a sky-averaged 21-cm signal in section 7.3. This is followed by a summary in

section 7.4. Additional information about the methodology and additional results can be found in section 7.S.

Section 7.S.3.2 and section 7.S.3.3 were led by Stefan Heimersheim for [2] and similarly the work in section 7.S.7 was led by Irene Abril-Cabezas. The co-authors, Stefan Heimersheim, Irene Abril-Cabezas, Anastasia Fialkov, Eloy de Lera Acedo, Will Handley, Saurabh Singh and Rennan Barkana, of [2] provided comments on the manuscript that this chapter is based on.

## 7.2 Methodology

The third iteration of the SARAS radiometer experiment took 15 hours of data, integrated in the frequency range 55 – 85 MHz, from a lake in Southern India. After corrections were made for the receiver noise temperature, radio frequency interference and emission from the water beneath the antenna, the data was appropriately scaled by the total efficiency of the system to produce a measurement of the average sky temperature which is expected to include the Galactic and extra-galactic foregrounds as well as the 21-cm signal. SARAS3 reported a null detection of the EDGES absorption feature with a confidence of 95.3% [54] and more recently the data has been used to place constraints on the properties of the first galaxies which we discuss in more detail below [chapter 6, 7].

SARAS2, the previous iteration of the instrumentation, recorded data at higher frequencies (lower redshifts) of 110 – 200 MHz and was found to contain a non-smooth systematic structure possibly caused by emission from the ground that the antenna was placed on. The data has been used to derive constraints on galaxies in the infant Universe, initially by fitting the foreground and systematic structure together with high order polynomials [69, 70] and subsequently by modelling the two components separately [chapter 5, 6]. In section 7.S we combine constraints from the latter with constraints from SARAS3 and HERA.

To date, interferometric experiments have only observed upper limits of the cosmological 21-cm power spectrum, which still allow for a large range of realistic astrophysical scenarios. The tightest constraints are derived from the data from the HERA interferometer [197], followed by MWA in the redshift range  $z = 6.5 - 8.7$  [107] and LOFAR, which provide the tightest upper limits at redshifts  $z \sim 9.1$  [218] and  $z \sim 9.3 - 10.6$  [219]. The HERA telescope is a radio interferometer located in the Karoo Desert of South Africa [20] and has conducted its first observing campaign from 2017 to 2018. Already, this first public data release delivered the strongest constraints on the 21-cm power spectrum to date. The publicly available HERA data<sup>a</sup> from the analysis of Internal Data Release 2 that we use in this chapter is based on 18 nights of observations, with 39 antennas operating at science quality level. For our analysis, we employ the publicly available spherically averaged power spectra derived from this data [197], in the wavenumber range  $k = 0.128 \text{ } h\text{Mpc}^{-1}$  to  $0.960 \text{ } h\text{Mpc}^{-1}$  and from the two bands focusing on redshifts  $z = 7.9$  and  $10.3$ . In section 7.S we consider the implications of including the upper limits on the power spectrum from MWA and LOFAR in our joint analysis.

In 21-cm cosmology, data analysis efforts are increasingly employing Bayes theorem

$$\mathcal{P}(\theta|D, \mathcal{M}) = \frac{\mathcal{L}(\theta)\pi(\theta)}{\mathcal{Z}}, \quad (7.1)$$

to derive constraints on the astrophysical scenarios of the early Universe. To recap, the likelihood,  $\mathcal{L}(\theta)$ , represents the probability that we observe the data,  $D$ , from SARAS3 or HERA given a particular model,  $\mathcal{M}$ .

<sup>a</sup>Note that an analysis of Internal Data Release 3 (IDR3) was recently published [99] which, at the time of writing, is not publicly available. IDR3 included the data used to arrive at the earlier limits on the power spectrum [109] used here, along with an additional 76 nights of observations. Analysis of the data on the class of models explored in this work showed led to only a small improvement in the astrophysical constraints.

The prior,  $\pi(\theta)$ , represents our assumed knowledge before we consider any data and the evidence,  $\mathcal{Z}$ , is a normalisation constant. The posterior,  $\mathcal{P}(\theta|D, \mathcal{M})$ , tells us which parts of the parameter space,  $\theta$ , given the data and chosen model, are more probable than others. The evaluation of Bayes theorem is usually performed with Nested Sampling or Markov Chain Monte Carlo (MCMC) algorithms (see section 7.S). In many 21-cm experiments,  $\theta$  is composed of parameters that describe instrumental effects,  $\theta_I$ , foregrounds,  $\theta_{fg}$ , and the astrophysical processes that influence the 21-cm signal,  $\theta_{21}$ . We typically refer to the set of  $\theta_I$  and  $\theta_{fg}$  as the nuisance parameters  $\alpha$ . Since we are only interested in the 21-cm signal, we have to approximate the marginal or nuisance-free posterior  $\mathcal{P}(\theta_{21}|D, \mathcal{M})$  given samples on  $\mathcal{P}(\theta|D, \mathcal{M})$ . We can do this with normalising flows and the marginal Bayesian statistics code MARGARINE [1, 5]. The code is described in chapter 4, in summary it allows us to calculate the marginal posterior and in turn the nuisance-free likelihood,  $\mathcal{L}(\theta_{21})$ , for HERA and SARAS3 for a given set of parameters via a trained density estimator. Since the experiments are independent, we can subsequently multiply the two likelihoods together to produce a joint likelihood

$$\mathcal{L}_{\text{joint}}(\theta_{21}) = \mathcal{L}_{\text{HERA}}(\theta_{21}) \mathcal{L}_{\text{SARAS3}}(\theta_{21}), \quad (7.2)$$

and sample this in equation (7.1) in a subsequent Nested Sampling or MCMC run to produce joint constraints on early astrophysics. We recap the astrophysical processes included in the modelling, and the definition of  $\theta_{21}$ , next.

In order to realistically model the range of time covered by the Cosmic Dawn and Epoch of Reionization, we need a consistent modelling of the cosmological and astrophysical processes from redshift 60, when star formation might have begun, all the way to redshift 5 at the end of the EoR. To that end we employ semi-numerical simulations [11, 12, 14, 195, 206] that have previously been used in this thesis to analysis data from SARAS2 [6] and SARAS3 [7] as well as to analyse data from HERA [109], LOFAR [105] and EDGES High-Band [91] data. We model the three-dimensional 21-cm field as a function of cosmic time during the infant Universe taking into account important astrophysical process including the Wouthuysen-Field (WF) effect [12, 39, 40], heating of the intergalactic medium by X-ray [195], Lyman- $\alpha$  [14] and CMB photons [131], multiple scattering of Lyman- $\alpha$  photons, relative velocity between dark matter and gas [11], feedback of Lyman-Werner radiation on star formation [206], and radio emission from galaxies [121]. The key parameters in the astrophysical model are: the star formation efficiency,  $f_*$ , which quantifies the percentage of the baryonic mass in the star forming halos that is converted into stars; the minimum virial circular velocity,  $V_c$ , which is proportional to the cube root of the halo mass,  $M$ ; the X-ray production efficiency,  $f_X$ , which is directly proportional to the X-ray luminosity per star formation rate,  $L_X/\text{SFR}$  measured in  $\text{erg s}^{-1} \text{M}_\odot^{-1} \text{yr}$ , between 0.2 and 95 keV; the CMB optical depth,  $\tau$ ; finally, we model the contribution of high redshift radio-luminous galaxies to the radio background by specifying a radio production efficiency,  $f_{\text{radio}}$ , which is proportional to the radio luminosity per star formation rate,  $L_r$ , measured in  $\text{W Hz}^{-1} \text{M}_\odot^{-1} \text{yr}$  at 150 MHz, and normalized such that it has a value of one for the present day population of radio galaxies [121]. The X-ray spectral energy density is modelled based on a population of X-ray binaries, as in [46]. In our Bayesian analysis,  $\theta_{21}$ , therefore, comprises the set of parameters  $\{f_*, V_c, f_X, \tau, f_{\text{radio}}\}$  or equivalently  $\{f_*, M, L_X/\text{SFR}, \tau, L_r/\text{SFR}\}$ .

We explore wide prior ranges on all the parameters in an attempt to let the data inform us about the high-redshift astrophysical processes. Specifically, the model for radio-luminous galaxies that we employ here is not conventionally considered in 21-cm cosmology, where typically the CMB is assumed to be the only source of radio background photons. Here, we expect that early galaxies will contribute to the radio background, thus increasing the amplitude of both the sky-averaged 21-cm signal [117] and the power spectrum [121, 131]. We

explore a broad range of radio luminosities that produce radio excesses above the CMB of between  $\approx 0.5 - 270$  times the CMB temperature at  $z = 20$  and  $\approx 1 - 32000$  times the CMB at  $z = 10$ . While moderate radio emission might be expected [120], the extreme values of radio production efficiencies are usually invoked to explain the anomalously deep EDGES signal [8, 117–121, 131] but struggle to explain the rapid star formation and rapid heating of the gas [221] that is implied by the shape of the absorption feature.

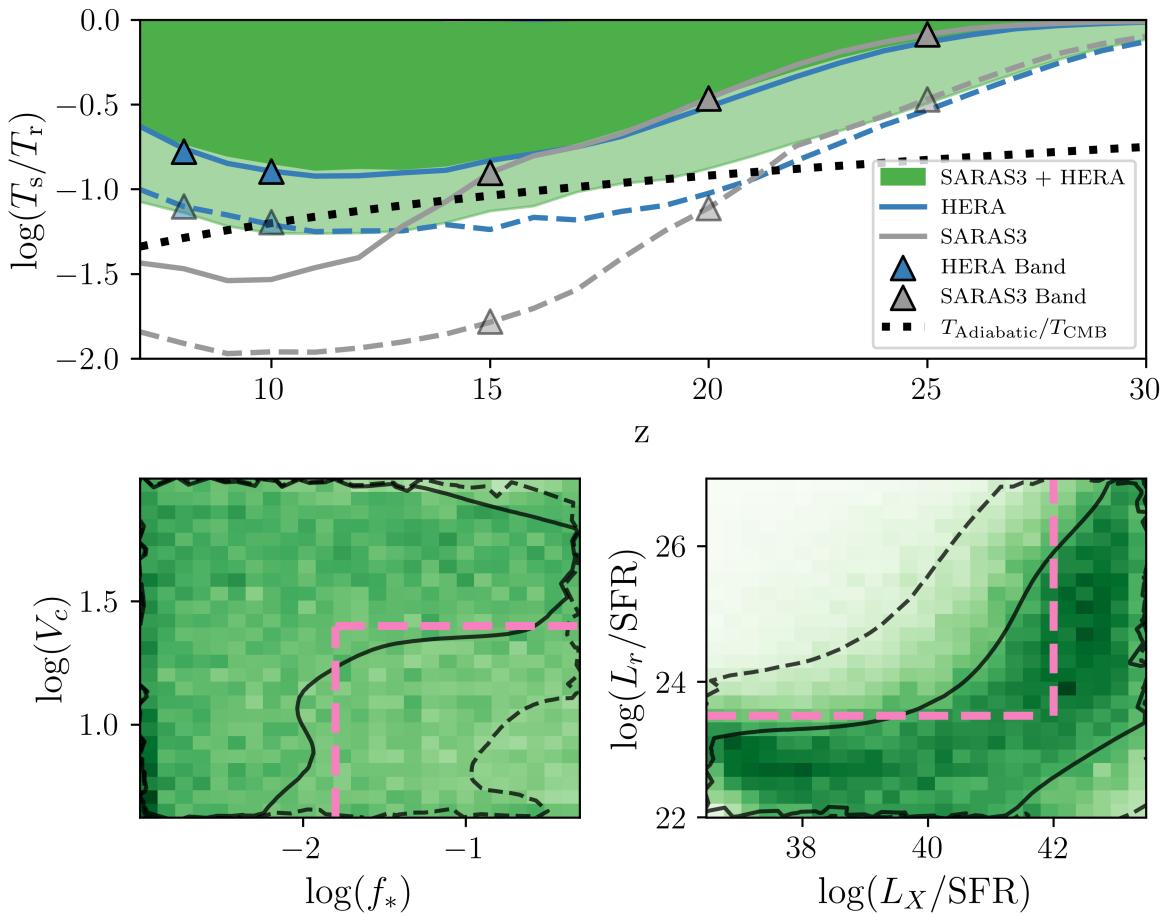
The computed cubes in our semi-numerical simulations of 21-cm signals are used to infer both the sky-averaged 21-cm signal and the 21-cm power spectrum. Both types of signals rely on the same underlying physics, and constraints from experiments targeting the different probes can be effectively combined to improve our knowledge of the infant Universe. To model the 21-cm signal, we train neural networks on outputs of the detailed simulations. The specific emulators used in this analysis and more details regarding the signal modelling can be found in section 7.S.

## 7.3 Results

Although there are a number of well-developed experimental efforts targeting the 21-cm signal, the field is still young. Consequently, current constraints on the properties of the first galaxies, including those presented here, are still weak. However, through the novelty of combining the previously reported HERA and SARAS3 constraints we produce the tightest constraints to date on the properties of the infant Universe as detailed below. This is the first time a joint analysis between a global signal data and interferometric limits has been attempted.

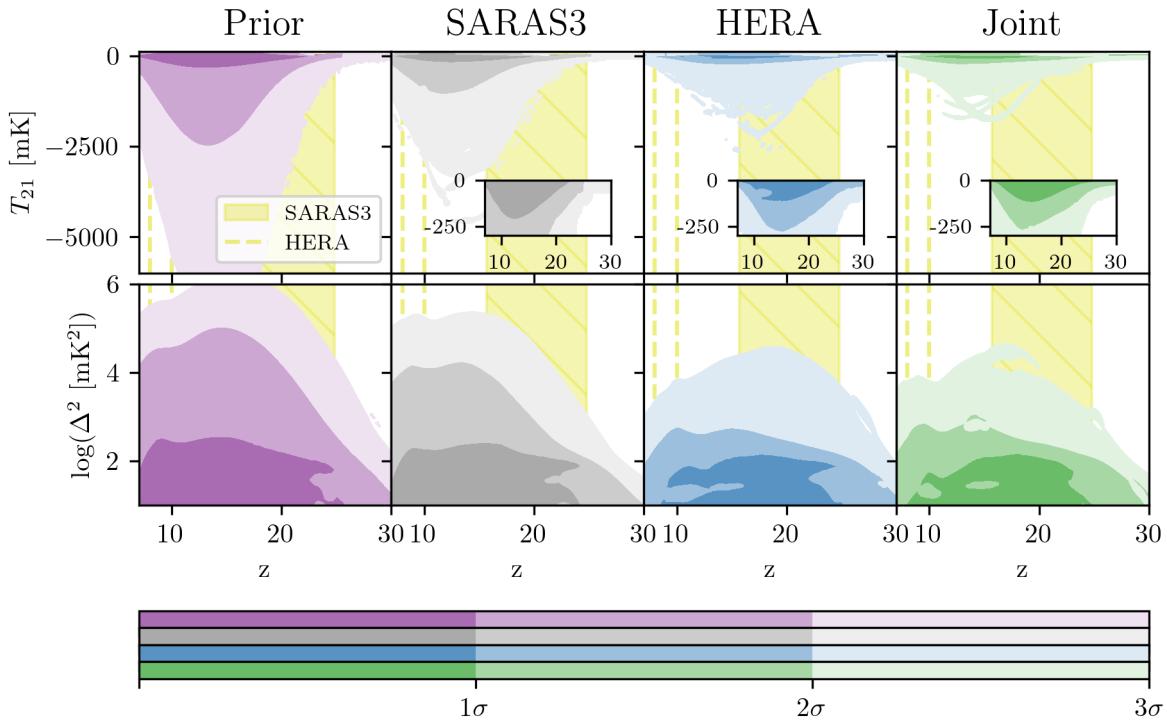
To visualize the importance of combining the constraining power of HERA and SARAS3 we show, in the top panel of Fig. 7.1, constraints on the ratio of the spin temperature of neutral hydrogen and background radiation temperature. The background radiation temperature is the sum of the CMB temperature and radio background from galaxies. The ratio determines the maximum absorption of the sky-averaged 21-cm signal, the smaller the ratio the larger the signal can be. The SARAS3 limits on the 21-cm signal (grey markers) correspond to lower limits on  $T_s/T_r$ , with the corresponding 1 and 2  $\sigma$  contours shown as lines and extrapolated out of the SARAS3 band. Our simulations provide a natural link between the power spectra and the global quantities, e.g.  $T_s/T_r$ , meaning that we can use the limits on the power spectrum from HERA to derive an equivalent constraint on  $T_s/T_r$ . These constraints are shown in Fig. 7.1 (blue markers and lines). The joint constraint, as shown by the green contours, provides the strongest constraints on the ratio, and in particular at redshifts  $z = 15 - 20$ , gives better constraints than either experiment alone. Further, the combination of the two experimental data sets improves the constraints at intermediate redshifts over a pure extrapolation of each one of the sets of constraints. As a guideline, the dashed black line in the figure shows the ratio for  $T_r = T_{CMB}$ , i.e. in the absence of radio emission from galaxies, and assuming adiabatically cooled gas in an expanding Universe in the absence of any astrophysical heating sources but with saturated coupling between the 21-cm spin temperature and the gas kinetic temperature. This limit is often used in the literature to give context to the constraints (e.g. [99, 109]). However, we note that for the models tested here the gas does not cool adiabatically because of the CMB and Lyman- $\alpha$  heating at early times, and we have an excess radio background above the CMB.

Next, we explore the functional constraints in the  $T_{21} - z$  and  $\Delta_{21} - z$  planes as shown in Fig. 7.2. These are calculated by taking the samples of  $\theta_{21}$  output from our fits and using the neural network emulators to produce corresponding global signals and power spectra. Using our theoretical models, we can easily map between the constraints on the power spectra and global signals. We see that for both the power spectra and the sky-averaged



**Figure 7.1: Key results from the joint analysis.** **Top Panel:** The ratio of the spin temperature of neutral hydrogen,  $T_s$ , and the radio background temperature,  $T_r$ , as a function of redshift for the joint HERA and SARAS3 analysis in green. We show the HERA and SARAS3 68% and 95% confidence constraints in blue and grey respectively as triangles at the relevant redshifts and solid and dashed lines. As a guideline, we show the ratio for  $T_r = T_{CMB}$  and assuming adiabatically cooled gas in an expanding Universe in the absence of any heating but with saturated coupling between  $T_s$  and the gas kinetic temperature (dashed black line). **Bottom Left:** The 2D probability distribution function (PDF) from the joint analysis on the minimum virial circular velocity,  $V_c$ , in combination with the star formation efficiency,  $f_*$ , marginalising over  $f_{\text{radio}}$ ,  $f_X$  and  $\tau$ . The solid black line shows the 68% contour, approximated by the pink dashed line, and the black dashed line shows the 95% contour. The joint analysis disfavours low values of  $V_c$  and high  $f_*$  corresponding to efficient star formation. **Bottom right:** The constraint on the X-ray and radio luminosities from the joint analysis marginalising over  $V_c$ ,  $f_*$  and  $\tau$ . The joint analysis disfavours at 68% confidence low X-ray efficiencies in combination with high radio production efficiencies.

signals, although more clearly for the latter, the range of plausible models is reduced by our joint analysis. The signals that are inconsistent with the data typically have strong power spectra and corresponding deep absorption trough in the global signal, and belong to the models with weak X-ray heating and strong radio luminosity. Further, we see that the  $2\sigma$  region for the functional constraints correspond to signals with the power spectra  $\lesssim 10^{2.1} \text{ mK}^2$  at  $z = 25$ , which via our modelling maps to global 21-cm signals shallower than  $-277 \text{ mK}$ . We note that the  $3\sigma$  limit on the power spectrum at  $z = 25$  of  $\lesssim 10^{3.2} \text{ mK}^2$  is approximately equivalent to the expected sensitivity of NenuFAR from 1000 hours of observations [222]. For the global 21-cm signal,  $3\sigma$  limit on the magnitude reduces from  $\approx -2630 \text{ mK}$  from HERA to  $\approx -1770 \text{ mK}$  from the joint analysis. Remarkably, we find that for the sky-averaged signal, the  $2\sigma$  limit is very close to the minimum depth of the ‘standard astrophysical’ models,  $\approx -165 \text{ mK}$  [14], where the radio background is equated to the CMB, the contributions from radio galaxies and X-ray heating sources are assumed to be negligible, while the CMB and Lyman- $\alpha$  heating are present. The relatively tight constraints on the global signal and power spectrum suggest that future improved



**Figure 7.2: Functional constraints on  $T_{21}$  and  $\Delta^2$ .** The functional prior (purple), SARAS3 (grey), HERA (blue) and joint (green) posteriors for the sky-averaged 21-cm signal (top row) and power spectrum (bottom row). The yellow shaded region shows the SARAS3 band, and the dashed yellow lines show the HERA redshifts. The functional prior and posteriors are calculated by taking representative samples from the corresponding probability distributions for the astrophysical parameters and generating the corresponding signals using neural networks. We see that by combining the constraining power of HERA and SARAS3, we significantly reduce the  $3\sigma$  (lightest shaded regions) constraints on the magnitude of both signals from  $-2630 \text{ mK}$  to  $-1770 \text{ mK}$  at  $z = 15$  for the global signal and  $10^{3.7} \text{ mK}^2$  to  $10^{3.2} \text{ mK}^2$  for the power spectrum at  $z = 25$ . The figure is produced with FGIVENX [196].

measurements will allow us to dig deeper into the models with weak excess radio background radiation. It is clear from our analysis, that the joint constraint improves limits on the range of plausible global signals and power spectra.

We can now interpret the constraints on the temperatures in terms of the limits on the properties of the astrophysical sources modelled in the simulation. The key constraints from the joint analysis are shown in the bottom panels of Fig. 7.1 including the limits on the high-redshift star formation which drives the high-redshift portion of the 21-cm signal via the process of Lyman- $\alpha$  coupling (constraints in the planes  $V_c - f_*$ ) and the constraints on the luminosity of X-ray and radio sources ( $L_r - L_X$ ) which primarily regulates the depth of the absorption trough. The full marginalised 1D and 2D posteriors corresponding to the joint analysis are shown in Fig. 7.S.1 and the key numerical results are summarized in Tab. 7.1 where we list the individual constraints from SARAS3 and HERA along with the joint analysis. We find that the combination of the two experiments leads to stronger constraints in the two-dimensional probability distribution of  $L_r - L_X$  than either of the two experiments individually. Where HERA constrains the population of high redshift radio-luminous galaxies to be  $\lesssim 400$  times brighter in the radio band than the current population, the combination of the data sets constrains the galaxies to be  $\lesssim 300$  times brighter when marginalising over the other parameters ( $V_c$ ,  $f_*$ ,  $f_X$  and  $\tau$ ). Similarly, HERA disfavours at 68% confidence galaxies with an X-ray luminosity  $\lesssim 0.25$  times the present day value in combination with the radio luminosity of galaxies in the early universe that is  $\gtrsim 400$  times the present day value. The joint analysis provides a stronger constraint, ruling out scenarios where the X-ray

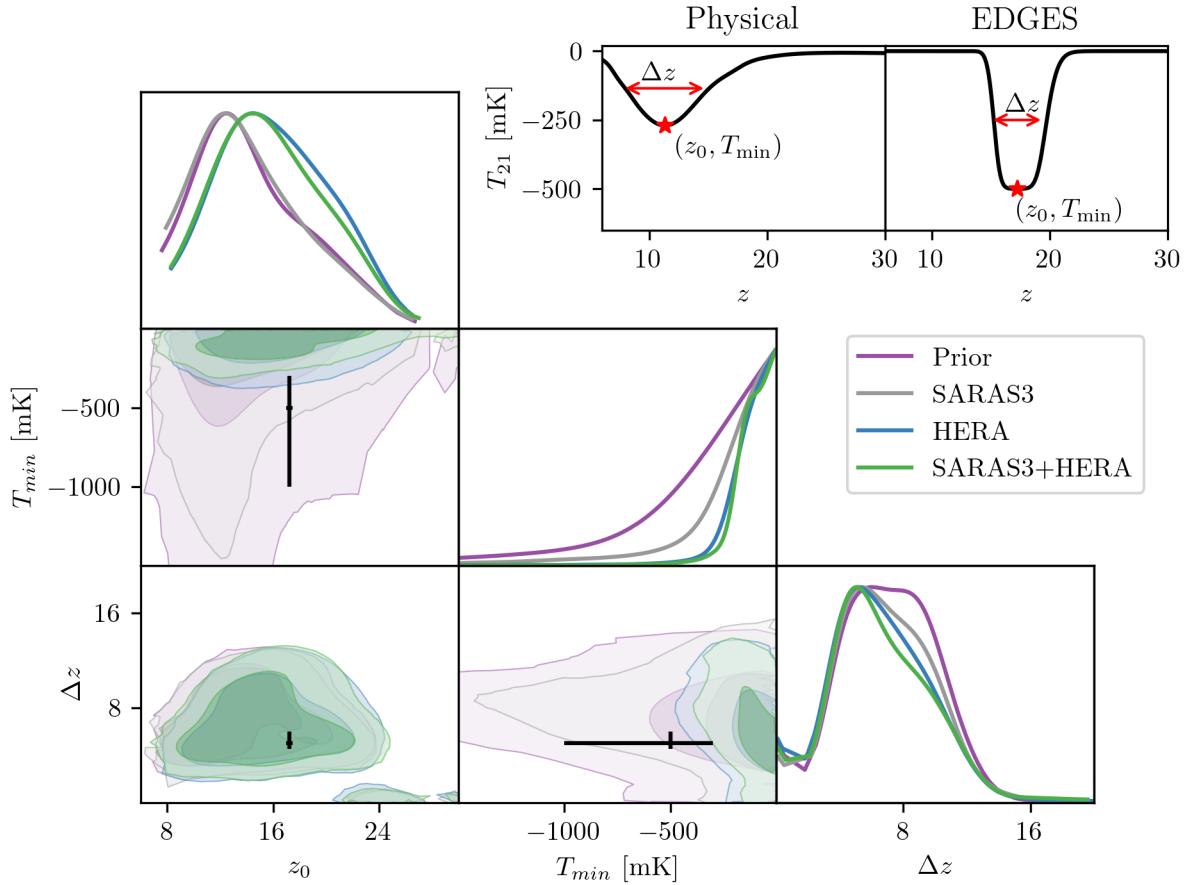
	SARAS3	HERA	SARAS3 + HERA
Signal	Sky-averaged	Power Spectrum	Both
$z$	$\approx 15 - 25$	$\approx 8 \& \approx 10$	$\approx 8, \approx 10 \&$ $\approx 15 - 25$
$L_r/\text{SFR}$	$\gtrsim 1.55 \times 10^{25}$	$\gtrsim 4.00 \times 10^{24}$	$\gtrsim 3.31 \times 10^{24}$
$L_X/\text{SFR}$	–	$\lesssim 7.60 \times 10^{39}$	$\lesssim 3.71 \times 10^{39}$
$L_r/\text{SFR} \& L_X/\text{SFR}$	$\gtrsim 1.00 \times 10^{25} \&$ $\lesssim 1.09 \times 10^{42}$	$\gtrsim 4.00 \times 10^{24} \&$ $\lesssim 7.60 \times 10^{39}$	$\gtrsim 3.16 \times 10^{23} \&$ $\lesssim 1.00 \times 10^{42}$
$M$	$4.40 \times 10^5 \lesssim M \lesssim 1.10 \times 10^7$	–	$2.55 \times 10^5 \lesssim M \lesssim 7.04 \times 10^6$
$f_*$	$\gtrsim 0.05$	–	$\gtrsim 0.06$
$f_* \& M$	$\gtrsim 0.03 \&$ $\lesssim 8.53 \times 10^8$	–	$\gtrsim 0.02 \&$ $\lesssim 4.50 \times 10^7$

**Table 7.1: Key parameter constraints from SARAS3, HERA and the joint analysis.** Here the SARAS3 and HERA limits are taken from chapter 6 and [109]. In the top two rows, we show the type of signal targeted by each set of analysis and the corresponding redshifts. The joint analysis produces improved constraints on the radio and X-ray backgrounds, while retaining the constraining power of SARAS3 on the star formation properties of early galaxies.  $L_r$  is measured in  $\text{W Hz}^{-1} \text{M}_\odot^{-1} \text{yr}$  at a reference frequency of 150 MHz,  $L_X$  in  $\text{erg s}^{-1} \text{M}_\odot^{-1} \text{yr}$ , and is calculated by integrating X-ray spectral distribution of sources between 0.2 and 95 keV assuming an X-ray SED consistent with that for X-ray binaries [46]. The halo mass,  $M$ , is measured in solar masses. These constraints are derived from Kernel Density estimates (KDE) of the 1D and 2D posterior distributions.

luminosity is  $\lesssim 33$  times the present day value and the radio luminosity of the first galaxies is  $\gtrsim 32$  times the present day value at 68% confidence.

We find comparable constraints on  $f_*$  and minimum mass of star forming halos,  $M$ , as was found with SARAS3 alone [chapter 6, 7] when combining the data sets, suggesting that HERA has no influence on our ability to determine high-redshift star formation efficiencies or dark matter halo masses (or, equivalently,  $V_c$ ). Marginalising over the radio and X-ray luminosities, we disfavour at 68% confidence galaxies in which  $\gtrsim 2\%$  of the gas is converted into stars and the minimum mass for star forming halos is  $\lesssim 45$  million solar masses.

Further, we explore the structure of the global 21-cm signals which are consistent with the data from SARAS3, HERA and the two sets together. We do this by looking at the minimum temperature,  $T_{\min}$ , the corresponding central redshift,  $z_0$ , and an approximate full width at half max,  $\Delta z$ , of each absorption trough as defined in the top right of Fig. 7.3. More specifically, for each parameter set  $\theta_{21}$  (either from the prior parameter range, or from the posterior ranges consistent with SARAS3, HERA, and the joint analysis) we generate a global signal using the neural network emulator GLOBALEMU [chapter 3, 4]. We then measure the values of  $T_{\min}$ ,  $z_0$ , and  $\Delta z$  for each signal, producing probability distributions for each parameter corresponding to the constraints from each data and showing the extent of the prior (see results in Fig. 7.3). We compare these distributions with the values used to parameterise the phenomenological EDGES flattened Gaussian signal with the EDGES 99% confidence ranges shown by the black crosses in the corner plot in Fig. 7.3. We see that when considered individually, the SARAS3 and HERA experiments allow for astrophysically motivated sky-averaged 21-cm signals that have a minimum temperature, location and width in agreement with the EDGES detection (i.e. the EDGES measurement overlaps with the 95% contours), while the joint analysis rules out models with



**Figure 7.3: Phenomenological constraints.** The triangular plots shows the prior (purple) and posteriors (grey for SARAS3, blue for HERA, green for joint) of the features of a typical global absorption signal: the central redshift,  $z_0$ , the corresponding minimum temperature,  $T_{\min}$ , and the width of the signal,  $\Delta z$ , as is depicted in the top right corner. Darker shaded regions show  $1\sigma$  constraints, lighter shaded regions show  $2\sigma$  constraints. Overlaid on the posterior distributions are the  $99\%$  confidence intervals, black crosses, reported for the corresponding phenomenological parameterisation of the EDGES absorption feature in [8]. Note that this is not the same as the physical EDGES-like distribution explored in Fig. 7.S.2. We see that individually the experiments allow for signals with depths that are consistent with EDGES. However, the combination of the two data sets disfavours these signals with greater than  $2\sigma$  significance. We do not disfavour signals with the same width or central frequency as EDGES, but note that the joint analysis indicates a preference for shallower and narrower signals with higher central redshifts as can be seen in the 1D PDFs.

a depth that is consistent with EDGES at greater than a  $2\sigma$  significance. The joint analysis has a preference for shallower (lower values of  $|T_{\min}|$ ) and narrower signals with higher central frequencies, as can be seen in the corresponding 1D PDFs, which is driven largely by the HERA constraints.

We also consider the impact of SARAS2 on our joint constraints as well as that of MWA and LOFAR (with the latter two providing weaker constraints and thus having no effect), see section 7.S for details. We find that the addition of SARAS2 leads to an increased uncertainty in the star formation properties of galaxies in the infant Universe. This is likely because the SARAS2 data increases the envelope of plausible models at higher redshifts, where star formation dominates the structure of the 21-cm signal. However, we note that even with its inclusion, we find a tight constraint on the combination of  $L_r - L_X$ . We focus in the main text on SARAS3 and HERA, due to the uncertainty in the modelling of and presence of systematic structures in the SARAS2 data, which is discussed in more detail in section 7.S.

## 7.4 Conclusion

Through a combination of constraints on fluctuations and the sky-averaged 21-cm signal of neutral hydrogen, we have improved our understanding of the first galaxies that formed in the infant Universe between 200 and 700 million years after the Big Bang. This is the first time the data from the two different 21-cm probes have been combined to derive constraints on the astrophysical properties of the early galaxies. Even though the existing constraints are weak, we develop novel methodology and outline the approach which will become increasingly more useful as the next-generation experiments deliver stronger observational constraints.

Considering a wide space of plausible astrophysical models including high-redshift sources of ultraviolet, X-ray and radio photons which affect the 21-cm signal, we calculate corresponding sky-averaged spectra as well as the power spectra of fluctuations. Using an upper limit on the fluctuations from the HERA interferometer and non-detection of the global 21-cm signal by the SARAS3 radiometer, we find that only  $64.9^{+0.3}_{-0.1}\%$  of the explored theoretical parameter space is consistent with the joint SARAS3 and HERA constraint, which is a significant improvement over the individual values of  $92.3^{+0.3}_{-0.1}\%$  and  $79.0^{+0.5}_{-0.2}\%$  respectively.

Using the newly developed methodology we place the tightest constraints to date on the properties of cosmic gas, such as the spin temperature of the 21-cm hydrogen line (closely related, but not equal, to the gas kinetic temperature) and the radio background temperature,  $T_r$ , as well as on the radio and X-ray luminosities of the first galaxies disfavouring at 68% confidence galaxies that are approximately 32 times more efficient radio emitters than present galaxies and simultaneously are less than 33 times bright in the X-ray band. This work reports an increased degree of confidence over a wider range of redshifts than previous works which typically extrapolate outside the redshifts targeted by individual experiments, while here we interpolated between the observations of SARAS3 at  $z = 15 - 25$  and the HERA limits at lower redshifts  $z \sim 8$  and 10.

In this work, we also considered the addition of interferometric data sets from MWA and LOFAR to our analysis, which led to a negligible improvement in the results. These experiments probe the physics of the EoR covering a similar redshift range to HERA (between  $z \approx 6 - 10$ ), with the current HERA data providing the tightest constraints on our models. Of the current global or sky-averaged 21-cm experiments only SARAS2, SARAS3 and EDGES were able to place limits on the astrophysics of the infant Universe. Our main focus is on the SARAS3 limits, as there are concerns surrounding the cosmological nature of the signal reported in EDGES and a degree of uncertainty in the modelling of systematics in the SARAS2 data. We note that the SARAS2 data covers a similar redshift range as the HERA data and based on our analysis does not lead to a significant improvement in our constraints.

The new methodology developed in this chapter will allow for synergies to be explored between the upcoming observations, e.g. of the power spectrum from cosmic dawn measured by the NenuFAR [211], HERA, LOFAR or the SKA [102], as well as the measurements of the global signal by the wide-band REACH experiment covering the redshift range  $z \approx 7 - 28$  [10], PRIZM [67], MIST [114] and missions to the moon [38] among others.

## 7.S Supplementary materials

### 7.S.1 Modelling the Thermal History of the Infant Universe: Power Spectrum and Global Signal Synergies

At high redshifts around  $z \approx 20 - 30$ , the first stars begin to form and produce Lyman- $\alpha$  photons that interact with the baryonic matter, predominantly composed of neutral hydrogen, in the Universe. Neutral hydrogen atoms absorb and remit ambient Lyman- $\alpha$  photons in a process known as the Wouthuysen-Field (WF) effect [39, 40] that drives the relative number of atoms with aligned and anti-aligned proton and electron spins. This process couples the spin temperature of the neutral hydrogen,  $T_s$ , which describes the distribution of hydrogen atom spins, to the gas temperature,  $T_k$ , which is cooling at a faster rate than the radio background as the Universe expands. Further, interactions between the neutral hydrogen and Lyman- $\alpha$  emission result in the transfer of kinetic energy that raises the gas temperature, and coupled spin temperature, in a process known as Lyman- $\alpha$  heating [29, 42] preventing the gas from cooling adiabatically. However, despite the heating, the dominant WF effect produces an absorption feature in the sky-averaged 21-cm signal, which is measured relative to the radio background. Further, it leads to a peak in the power spectrum at high  $z \approx 25$  on angular scales corresponding to the effective horizon of the Lyman- $\alpha$  emission and the distribution of galaxies, which disappears when the coupling becomes saturated [223]. The intensity and spatial fluctuations of the Lyman- $\alpha$  emission evolve with the population of early galaxies, and consequently it is dependent on their star formation rate and the minimum halo mass for star formation. We parameterize these quantities with the star formation efficiency,  $f_*$ , which quantifies the percentage of the baryonic mass in the star forming halos that is converted into stars, and the minimum virial circular velocity,  $V_c$ , which is proportional to the cube root of the halo mass,  $M$ .

At intermediate redshifts of  $z \approx 10 - 20$  the gas is further heated by X-ray binaries [46, 195], continuing Lyman- $\alpha$  heating, CMB heating [44] and heating through structure formation. X-ray heating is dependent on processes such as X-ray production in high redshift galaxies and black hole binary formation. This means that it has to be separately parameterized and in our semi-numerical simulations, we model the X-ray production efficiency,  $f_X$ , which is directly proportional to the X-ray luminosity per star formation rate,  $L_X/\text{SFR}$  measured in  $\text{erg s}^{-1} M_{\odot}^{-1} \text{yr}$ , between 0.2 and 95 keV. The heating affects the redshift of the minimum and depth of the sky-averaged 21-cm signal. If sufficiently efficient it can raise the temperature of the gas, and coupled 21-cm brightness temperature, above the radio background resulting in emission in the sky-averaged signal at low redshifts. The heating rate has a direct impact on how fast the brightness temperature transforms from absorption to 0 K or emission. In the power spectrum, due to the non-uniformity of the heating, the various mechanisms can produce a peak in the signal at around  $z \approx 15$ . Although this redshift is model dependent (see [223]) and in some cases when heating is done by hard X-rays (energies  $> 1$  keV with long mean free paths) X-ray heating is smooth and no peak is imprinted in the power spectrum [12, 195].

Finally, at more recent times,  $z \approx 5 - 15$ , ultraviolet emission from the first massive galaxies begins to ionize the neutral hydrogen, stripping the abundant gas of its electrons. This reduces the sky-averaged 21-cm signal and when reionization is complete, the signal disappears. The process produces a peak in the power spectrum at the scales corresponding to the typical size of ionized bubbles, but again the signal is destroyed once reionization is complete. The process of reionization is highly dependent on the ionizing efficiency of sources,  $\zeta$ , which in the models explored here is normalized by the CMB optical depth,  $\tau$ . While  $\tau$  has been weakly constrained by cosmological experiments such as Planck [22], we treat it as a free parameter and note that 21-cm cosmology

offers a means by which to break degeneracies between  $\tau$  and other cosmological parameters.

### 7.S.2 Combining Constraints with MARGARINE

We use the marginal Bayesian statistics code **MARGARINE** [5] to combine constraints from different data sets. **MARGARINE** is outlined in chapter 4 but to recap it uses neural networks known as Masked Autoregressive Flows (MAFs) to model the probability distribution,  $\mathcal{P}(\theta|D, \mathcal{M})$ , of a set of samples, here  $\theta_{21}$ , marginalising over nuisance parameters describing the instrumental systematics,  $\theta_I$ , and foregrounds,  $\theta_{fg}$  in the process. It does this by shifting and scaling a base standard normal distribution, making the probability of the distribution easily tractable, to replicate the target samples, where the shifting and scaling are determined by the outputs of the MAFs.

This can subsequently be used to calculate a nuisance-free likelihood given a prior distribution and the Bayesian evidence using

$$\mathcal{L}(\theta_{21}) \equiv \frac{\int \mathcal{L}(\theta_{21}, \alpha) \pi(\theta_{21}, \alpha) d\alpha}{\int \pi(\theta_{21}, \alpha) d\alpha} = \frac{\mathcal{P}(\theta_{21}|D, \mathcal{M}) \mathcal{Z}}{\pi(\theta_{21})}, \quad (7.3)$$

where  $\alpha = \{\theta_I, \theta_{fg}\}$  [1]. In instances when the prior is flat, then the following is true up to a normalisation constant

$$\mathcal{L}(\theta_{21}) \approx \mathcal{P}(\theta_{21}|D, \mathcal{M}). \quad (7.4)$$

With **MARGARINE** we can subsequently evaluate  $\log \mathcal{L}(\theta_{21})$  for any set of  $\theta_{21}$  for any existing posterior distribution,  $\mathcal{P}(\theta|D, \mathcal{M})$ , from previous analysis of a data set like HERA or SARAS3

$$\theta = \{\theta_I, \theta_{fg}, \theta_{21}\} \rightarrow \{\theta_{21}\} \rightarrow \text{MARGARINE} \rightarrow \log \mathcal{P}(\theta_{21}|D, \mathcal{M}) \rightarrow \log \mathcal{L}(\theta_{21}), \quad (7.5)$$

where the log is base 10. The likelihood evaluations can then be combined,

$$\log \mathcal{L}_{\text{joint}}(\theta_{21}) = \log \mathcal{L}_{\text{HERA}}(\theta_{21}) + \log \mathcal{L}_{\text{SARAS3}}(\theta_{21}), \quad (7.6)$$

as discussed in the main text, to be sampled over using MCMC methods or in our case Nested Sampling implemented with **POLYCHORD** [88, 89].

### 7.S.3 Emulating Signals from the EoR and CD

#### 7.S.3.1 The sky-averaged 21-cm signal

To emulate the sky-averaged 21-cm signal in models with an excess radio background from high redshift radio-luminous galaxies, we use the publicly available emulator **GLOBALEMU** described in chapter 3 and [4] trained on sets of astrophysical simulations. The simulations include the effects of Lyman- $\alpha$  and CMB heating. The mean free path of ionizing photons,  $R_{\text{mfp}}$ , is fixed at 40 Mpc and X-ray heating is powered by a population of X-ray binaries with a realistic spectral energy distribution [195]. The models correspond to the parameterisation detailed above and in [121].

The emulator has previously been used in the individual analysis of data from SARAS2 [6] and SARAS3 [7] and detailed in the corresponding papers and chapters. We therefore only briefly summarise the accuracy of the neural network here. The original set of simulations comprise approximately 10,000 models however the explored range of  $\tau$ , the CMB optical depth, is large and when training our neural network we filter out models

that have a value of  $\tau$  outside the range given by Planck  $\pm 3\sigma$ . This results in a training set of approximately 4,300 models, and the network is tested on approximately 500 models. A root mean squared error (RMSE) of 5.11 mK is found and a 95 percentile RMSE of 20.53 mK indicating a high level of accuracy (see chapter 6).

### 7.S.3.2 The 21-cm power spectrum

For comparing models with the HERA measurements we use the same 21-cm power spectrum emulator used in the HERA analysis [109], based on the same suit of simulations of the sky-averaged signal emulator used in the SARAS2 and SARAS3 analysis. Based on an input of the 5 model parameters, this emulator returns the 21-cm power spectrum with a relative accuracy of 20% at the wave numbers and redshifts observed by HERA. For HERA the impact of  $\tau$  is more significant because the data corresponds to lower redshifts and so the full prior range on the parameter is used for training. This results in a training set of  $\sim 8,000$  simulations and another 2,000 independent samples for testing. Full details and tests of this emulator can be found in the HERA analysis paper [109], Appendix B.

We note that when performing our joint analysis we use the narrow prior on  $\tau$  defined by Planck, since the SARAS3 posterior is not defined in the original analysis for values outside this range.

### 7.S.3.3 Temperatures

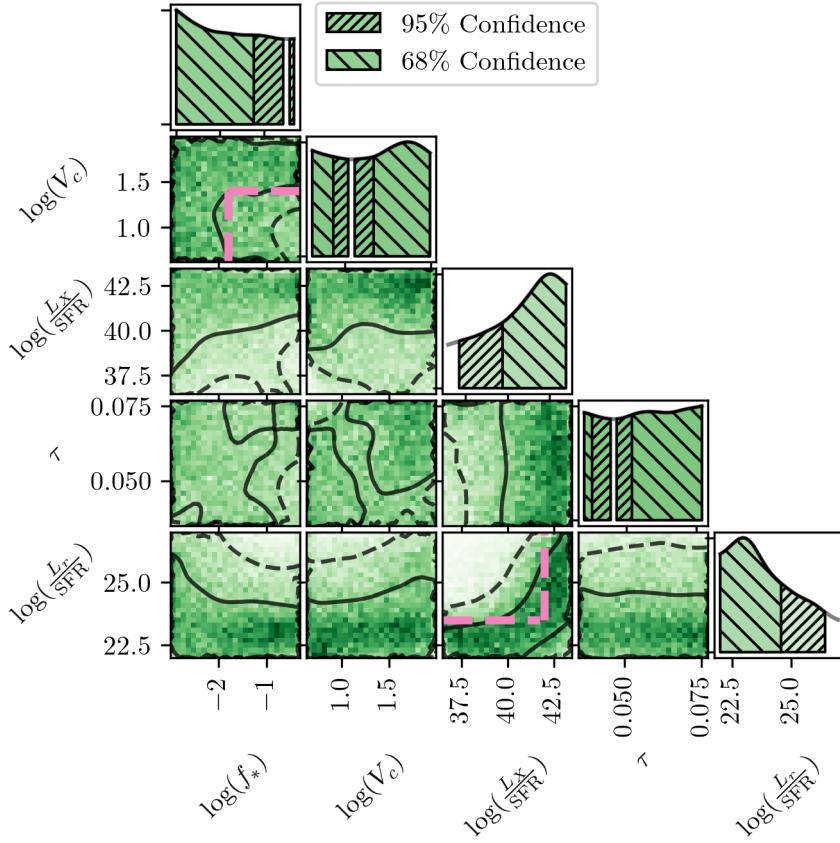
To emulate physical properties such as the spin temperature and temperature of the radio background (as seen in Figure 7.1) we use a GLOBALEMU-style emulator. In this framework, the emulator takes in the five astrophysical parameters and a single redshift and returns a single corresponding temperature. In practice, this means that vectorised calls have to be made to emulate the spin temperature,  $T_s(z)$ , and the radio background temperature,  $T_r(z)$ , as a function of redshift, but the method is found to be more accurate, quicker and allows for interpolation at a range of different redshifts. We use the same training and test sets as used for the 21-cm power spectrum emulator, and a similar architecture of 4 layers, with a reduced size of 100, 30, 10, and 5 nodes per layer. This allows us to emulate the spin temperature  $T_s$  within  $\pm 6\%$  accuracy (95% confidence interval), and the radio background temperature  $T_r$  within  $\pm 4\%$  accuracy (95% confidence interval). These neural networks and the power spectrum network were developed by Stefan Heimersheim.

## 7.S.4 Constraints on the Astrophysical Parameters

Fig. 7.S.1 shows the 1D and 2D marginal posteriors found for the joint analysis of SARAS3 and HERA. The graph shows that the combined constraining power of the two experiments leads to a strong constraint on the combination of the radio and X-ray luminosities per star formation rate for an early population of galaxies when marginalising over the other astrophysical parameters. We also see constraints in the plane of  $V_c - f_*$  when marginalizing over the radio background and X-ray heating parameters. The constraints are summarized in detail in the main text.

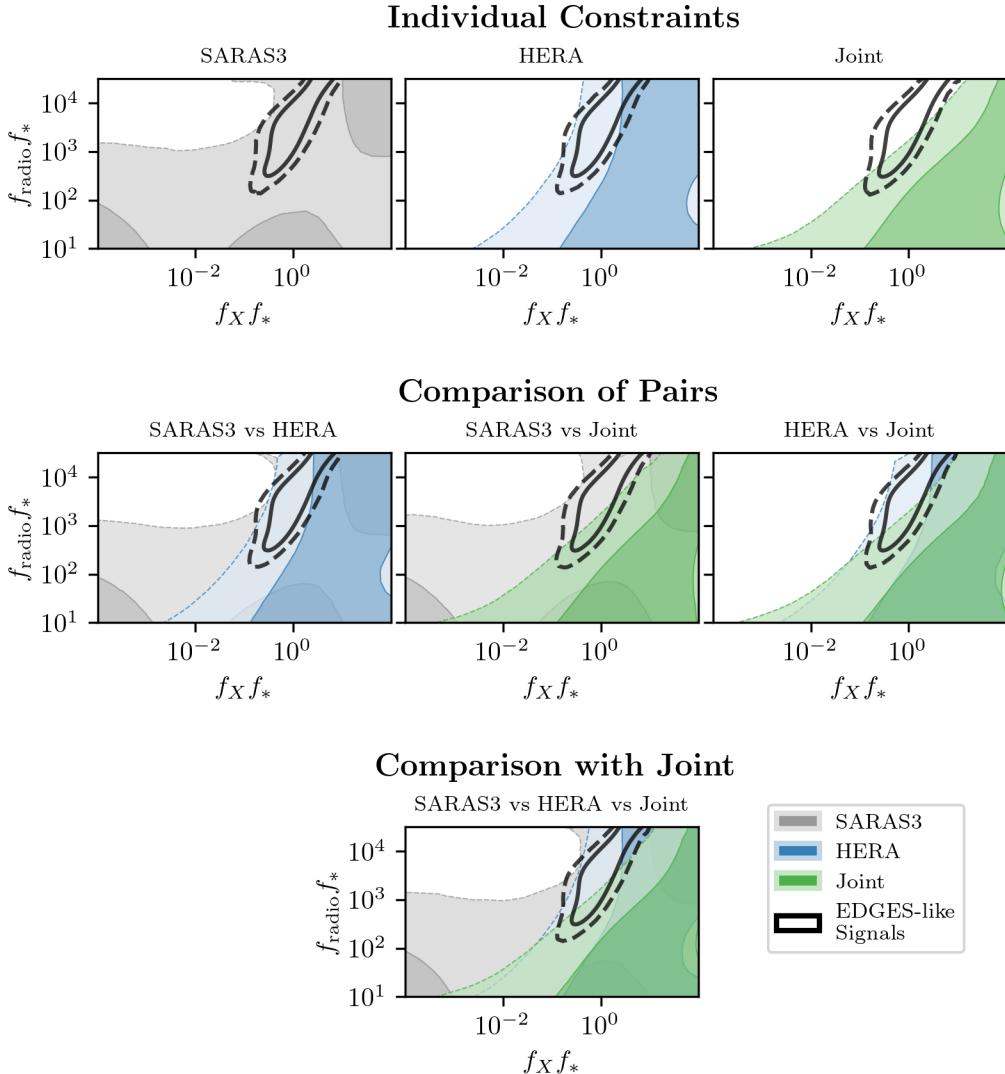
## 7.S.5 Constraints on the Radio and X-ray Backgrounds

The product  $f_* f_{\text{radio}}$  is proportional to the total radio background created by radio-luminous galaxies, and, equivalently,  $f_* f_X$  is a proxy for the total X-ray background created by the early population of X-ray sources. This is demonstrably true for our model parameterisation as the star formation rate is proportional to  $f_*$  and the



**Figure 7.S.1: Parameter constraints from the joint analysis.** The astrophysical parameter constraints on models with excess radio background in addition to the CMB derived when combining an upper limit on the 21-cm power spectrum at  $z \approx 8$  and  $\approx 10$  from HERA with data from the 21-cm sky-averaged experiment SARAS3 in the band  $z \approx 15 - 25$ . Through the combination of these two data sets probing different statistical properties of the 21-cm signal at different redshifts, we are able to improve constraints on the radio and X-ray luminosities of early radio-luminous galaxies and maintain constraints provided by SARAS3 on the star formation properties of these early galaxies.  $L_r$  is measured in units of  $\text{W Hz}^{-1} \text{M}_\odot^{-1} \text{yr}$  at 150 MHz and  $L_X$  is in units of  $\text{erg s}^{-1} \text{M}_\odot^{-1} \text{yr}$  calculated between 0.2 and 95 keV assuming a realistic SED of an early X-ray binary population. The pink dashed lines approximate regions that are disfavoured with 68% confidence.

radio luminosity and X-ray luminosities per star formation rate are proportional to  $f_{\text{radio}}$  and  $f_X$  respectively. Both X-ray and radio backgrounds are responsible for regulating the depth of the absorption feature, and they can also be observed independently by other telescopes (e.g. observations of the unresolved X-ray background by Chandra [224] and of the low-frequency radio background by ARCADE2/LWA [48, 49]). In Fig. 7.S.2 we show constraints on the values of  $f_* f_X$  and  $f_* f_{\text{radio}}$  achieved by SARAS3, HERA and the joint analysis. Since these combinations of the parameters regulate the absorption depth of the global 21-cm signal, we can also condition our prior on the astrophysical parameters to produce signals with the same central frequency and depth as the absorption feature found in the EDGES data [7, 14, 131]. In each panel of Fig. 7.S.2, we show black contours corresponding to these EDGES-like physical signals. We see that while HERA and SARAS3 allow for combinations of  $f_X f_*$  and  $f_{\text{radio}} f_*$  that could partially explain EDGES, the combination of the two experiments, which produces a tighter constraint on the X-ray and radio luminosities of early galaxies, disfavours a large portion of the EDGES-like parameter space (i.e. most of the EDGES-like parameter space is beyond the 95% contours of the joint constraints, while it is well within the 95% contours for SARAS3 and HERA



**Figure 7.S.2: Background constraints.** The figure shows constraints on the radio and X-ray backgrounds, parameterized by  $f_* f_{\text{radio}}$  and  $f_* f_X$  respectively, from SARAS3 (grey), HERA (blue) and the joint analysis (green). The SARAS3 and HERA posteriors are based on the results presented in [7] and [109] respectively. We show each distribution individually on the top row, overlaid pairs of distributions for comparison in the middle, and all three on the same figure in the bottom row. In all panels, we show 68% and 95% contours (black solid and dashed lines respectively) for physical signal models that have similar depths and central frequencies as the EDGES absorption feature as defined by the inequality in [131]. These physical EDGES-like models have previously been explored in the literature in [121, 131]. We note that while individually both HERA and SARAS3 allow for astrophysically motivated signal models that could explain the depth of the EDGES feature, together they rule the corresponding parameter space out with approximately greater than  $2\sigma$  confidence, although some EDGES-like signals are still viable. We stress, again, that the explored physical models cannot fully explain the shape of the EDGES signal.

individually). This demonstrates further the power of combining different data sets. However, we note that the explored theoretical signals do not fit EDGES data well, as none of the models closely reproduce the flattened Gaussian-like feature found in the data [8].

### 7.S.6 The impact of SARAS2

Previous analysis of the SARAS2 data revealed some weak constraints, most notably in the plane of  $L_X - L_r$  in agreement with HERA and SARAS3, on the properties of galaxies in the infant Universe [e.g chapter 5, 6]. SARAS2 is at much lower redshifts than SARAS3 but overlaps with the redshifts probed by HERA having recorded observations in the band  $z \approx 7 - 12$ . The data is contaminated by a sinusoidal systematic and a number of different models were fitted to this feature. The models correspond to a signal introduced prior to the antenna possibly from ground emission or some unknown component of the foreground, and separately a signal introduced in the system electronics potentially from cable reflections. The sinusoidal systematic was fitted alongside a signal model generated with GLOBALEMU and a foreground model that is conditioned to be smooth preventing it fitting out non-smooth systematics or signals in the data [3]. Here we take the best fitting model, with a systematic from ground emission or a non-smooth component from the foreground, with the highest evidence from the original analysis [see chapter 5 and 6] and combine the corresponding constraints on the astrophysical parameters  $V_c$ ,  $f_*$ ,  $f_X$ ,  $f_{\text{radio}}$  and  $\tau$  with the joint constraints from HERA and SARAS3 to assess the impact.

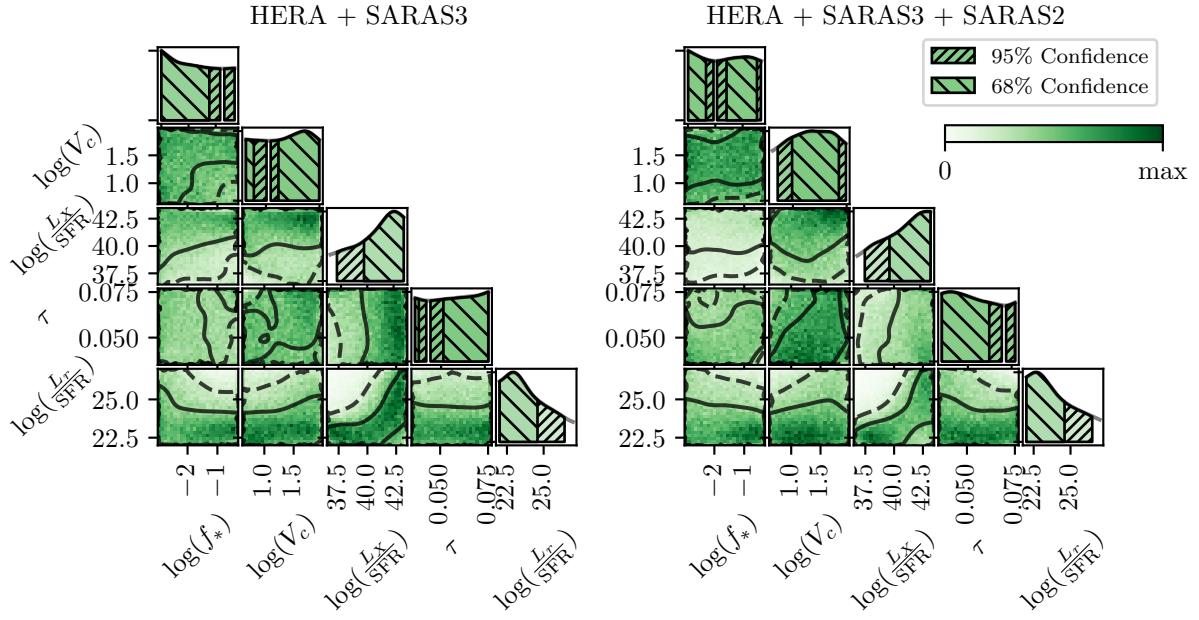
As can be seen in Fig. 7.S.3, the addition of SARAS2 to our analysis washes out the constraint in the plane  $f_* - V_c$ . One possible explanation for this is that the addition of SARAS2, while constraining the properties that affect the signal at low redshifts, increases the envelope of possible models at higher redshifts, where star formation is more important, that are plausible even given the constraints from the SARAS3 data. Despite this, we note that we maintain the constraint in the plane  $L_X - L_r$  when we add SARAS2 into our analysis.

We can quantify the impact of SARAS2 on our analysis by looking at the percentage of the astrophysical prior volume which is consistent with the different combinations of the three different data sets. To calculate this percentage, we use MARGARINE to calculate the marginal Kullback-Liebler divergence,  $\mathcal{D}$ , between the flat prior on the five astrophysical parameters in the set  $\theta_{21}$  and the corresponding posteriors. The KL divergence is related to the percentage via

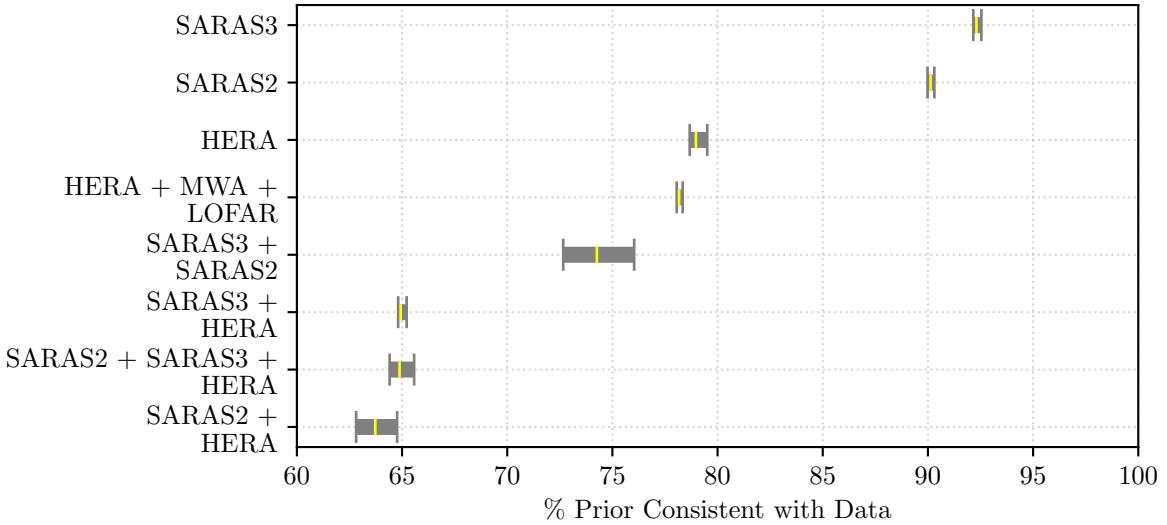
$$\% = 100 \times \exp(-\mathcal{D}) \approx 100 \times \frac{V_\rho}{V_\pi}, \quad (7.7)$$

where  $V_\pi$  is the prior volume and  $V_\rho$  is the posterior volume. This quantity is useful as it quantifies the constraining power of the different data sets in all five dimensions, including correlations that may not be visible in the one and two-dimensional projections used to produce the corner plots in this chapter and in the literature.

We show in Fig. 7.S.4 the percentage of the astrophysical parameter prior volume that is consistent with different combinations of the data sets discussed in this work (including additional interferometric measurements of the power spectrum discussed in section 7.S.7). We see that the combination of either or both of the SARAS data sets with HERA lead to a percentage consistency with the data of  $\approx 63 - 65\%$  and this is likely dominated by HERA. Individually, HERA allows for  $\approx 80\%$  of the astrophysical parameter space, SARAS2 for  $\approx 90\%$  and SARAS3 for  $\approx 92\%$ . Due to the uncertainty in the modelling of the systematics in the SARAS2 analysis, we leave SARAS2 out of the main results. We note also that while the combinations of SARAS3 and HERA and separately SARAS2 and HERA constrain similar percentages of the parameter space, that this quantity does not tell us whether the two pairs of experiments constrain the same parts of the parameter space.



**Figure 7.S.3: The impact of SARAS2 data on the astrophysical constraints.** We show the joint posterior distributions for HERA and SARAS3 on the left panel (identical to Figure 7.S.1, but shown here for comparison) and for HERA, SARAS3 and SARAS2 on the right panel. SARAS2 covers the band  $z \approx 7 - 12$  and therefore has some overlap with HERA but not with SARAS3. The addition of SARAS2 to the joint analysis washes out the constraint on star formation properties,  $V_c$  and  $f_*$ , because it leads to increased uncertainty in the structure of the signals at high redshifts. However, we still see a consistent disfavouring of a population of radio galaxies with high radio and low X-ray luminosities. The one dimensional posteriors for  $\tau$  appear to be in disagreement, however, we note that these are basically flat. We exclude SARAS2 from our main results in the text because of uncertainty in the modelling of systematics in the data.



**Figure 7.S.4: Constraining power of different data sets.** The percentage of the wide astrophysical parameter prior that is found to be consistent with the different data sets and different combinations of data sets explored in this work. A lower value indicates a better set of constraints, although a difference of a few percents does not necessarily translate into significant differences in the parameter constraints, as can be seen when comparing the results from HERA and HERA + LOFAR + MWA.

	HERA	LOFAR	MWA	LOFAR + MWA + HERA
$z$	$\approx 8 \& \approx 10$	9.1 & 9.3 – 10.6	6.5 – 8.7	Discrete and continuous ranges of $z$ between 6.5 – 10.6
$k [\text{h Mpc}^{-1}]$	$0.128 - 0.960$	0.075 – 0.432 & 0.053	0.070 – 3.000	Discrete and continuous ranges of $k$
$L_r/\text{SFR}$	$\gtrsim 4.00 \times 10^{24}$	$\gtrsim 1.20 \times 10^{25}$	$\gtrsim 1.58 \times 10^{25}$	$\gtrsim 4.00 \times 10^{24}$
$L_X/\text{SFR}$	$\lesssim 7.60 \times 10^{39}$	$\lesssim 8.70 \times 10^{38}$	$\lesssim 1.16 \times 10^{39}$	$\lesssim 1.58 \times 10^{40}$
$L_r/\text{SFR} \& L_X/\text{SFR}$	$\gtrsim 4.00 \times 10^{24} \& \lesssim 7.60 \times 10^{39}$	$\gtrsim 3.16 \times 10^{25} \& \lesssim 1.00 \times 10^{40}$	$\gtrsim 1.00 \times 10^{25} \& \lesssim 1.00 \times 10^{40}$	$\gtrsim 4.00 \times 10^{24} \& \lesssim 1.58 \times 10^{40}$

**Table 7.S.1: Constraints from interferometers.** The table shows the various constraints on the radio and X-ray luminosities for HERA, MWA, LOFAR and the combination of all three along with their respective wavenumbers and redshift ranges. The joint analysis only marginally improves our understanding of the infant universe.

### 7.S.7 Other Power Spectrum Experiments

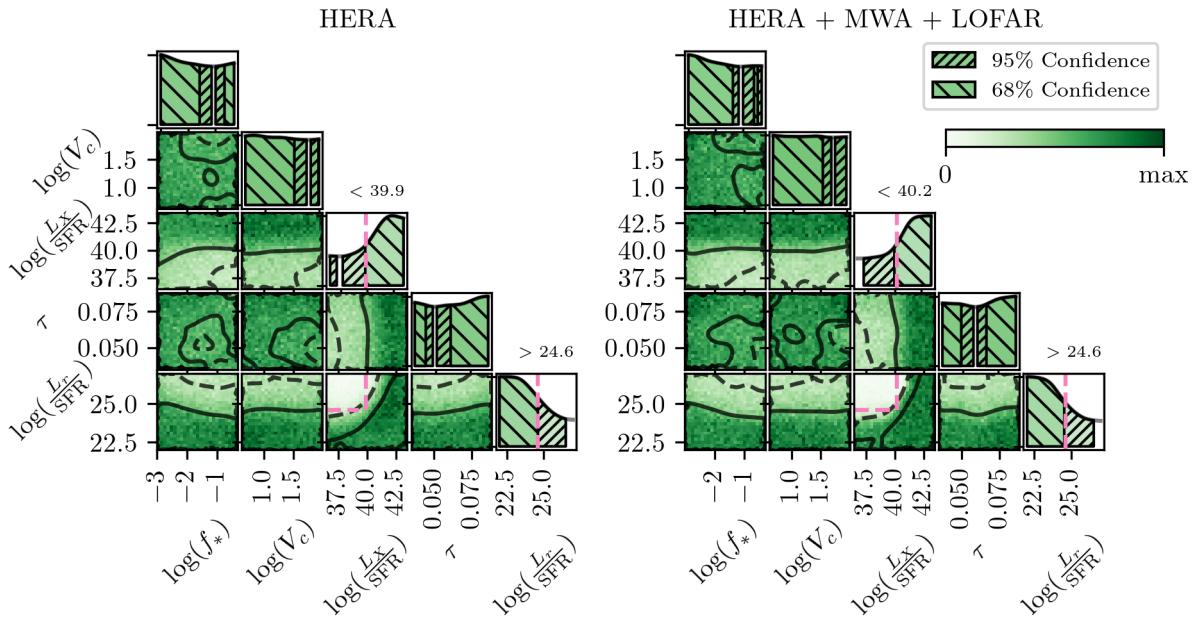
In Fig. 7.S.5, we show the projected posteriors derived using HERA data alone (left panel) and HERA, MWA and LOFAR together (right panel). We note that the constraints from the different interferometers are all at low redshifts between  $z \approx 6 - 10$  and varying wavenumbers or angular scales. These are detailed in Tab. 7.S.1 along with the constraints from the individual experiments on key parameters.

We derived parameter constraints from the MWA and LOFAR data using the approach taken in the original HERA analysis. Specifically, we take the measured upper limits, the mean power spectrum and uncertainty, and treat it as a measurement of cosmological signal plus systematics. As in HERA [109] we take this uncertainty to be Gaussian and marginalize over a uniform prior on the systematics, yielding the likelihood

$$\mathcal{L}(\theta_{21}) = \prod_i^{N_d} \frac{1}{2} \left( 1 - \text{erf} \left[ \frac{d_i - m_i(\theta_{21})}{\sqrt{2}\sigma_i} \right] \right), \quad (7.8)$$

where  $N_d$  represents the number of data points,  $d_i$  and  $\sigma_i$  correspond to the mean and standard deviation in each data point, and  $m_i(\theta_{21})$  is the model prediction for that redshift and wave number. Thus, a model prediction  $m \gg d$  gives  $\mathcal{L} \approx 0$  while  $m \ll d$  gives a constant. This likelihood is effectively a step function that disfavours models above a given amplitude. A full discussion of its derivation can be found in [109].

We performed this joint analysis using a full analytic likelihood approach (independent of MARGARINE) since there are no nuisance parameters describing the systematics. We find that each of the experiments disfavours individually similar regions of the  $L_r - L_X$  plane. However, the joint analysis does not improve the results derived from HERA data alone (as we summarise in Tab. 7.S.1 and is further illustrated in Fig. 7.S.4) which motivates our decision to only use HERA in the main text.



**Figure 7.S.5: The impact of MWA and LOFAR on the parameter constraints.** Projected posterior distribution functions (PDFs) for the 5 simulation parameters, obtained by assuming flat priors and combining different observations: HERA alone (left, as in [109]) and LOFAR, HERA and MWA (right). Solid (dashed) lines correspond to regions containing the highest 68% (95%) probability. We see that HERA constraints are not significantly improved by adding the published limits on the power spectrum from other interferometers.



## **Part IV**

### **Future work and conclusion**



## FUTURE WORK AND CONCLUSION

### 8.1 Future work

#### 8.1.1 REACH

In the coming year, deployment, commissioning and the first season of data collection from the Cambridge based REACH experiment is expected to have taken place. As an active member of the collaboration, I will be involved in the data analysis and many of the techniques outlined in this thesis will be applied to the data;

- The primary contribution from this thesis to the data analysis for REACH will be chapter 3 on the emulation of the global 21-cm signals. The signal emulator, GLOBALEMU, will be an integral part of the forward modelling in the REACH pipeline and allow the collaboration to derive key astrophysical information about the early universe. The emulator needs to be regularly updated to maintain consistency with the evolving state-of-the-art theoretical modelling, and was designed with this in mind. In practice, we will be able to train GLOBALEMU on different sets of models from semi-numerical and 1D radiative transfer codes such as 21cmFAST [151], ARES [149], GRIZZLY [135, 136] and the code used throughout this thesis to allow us to compare parameter constraints on each model or in the event of a detection check for consistency in the recovered physics.
- The work in chapter 5 and chapter 6, detailing the analysis of the SARAS2 and SARAS3 data sets, helped to develop the parameter inference framework that will be applied to the data from REACH. The former work demonstrated that systematics if present in the data do not necessarily prevent our ability to do useful physics, which is a key principle behind REACH and was discussed in [10] and [74].
- Chapter 4 and chapter 7 set out a marginal Bayesian framework which will allow the results from REACH to be easily combined with results from other probes of the infant Universe such as the 21-cm power spectrum and constraints on the CMB optical depth from Planck. Forecasts for such an analysis are currently been worked on by members of the collaboration (see Sims et al. in prep 2023a, b).
- The REACH analysis will primarily be performed via the forward modelling pipeline discussed in [57–59, 66, 74, 79] and via a Singular Value Decomposition pipeline based on the work presented in [225] and subsequent works and applied to REACH in [226]. In principle, an additional standalone analysis

using an EDGES style chromaticity correction and `MAXSMOOTH` could be performed. Although there are limitations to the EDGES style chromaticity correction, discussed briefly in chapter 1, it may be sufficient to model the dominant chromatic structure in the data and in combination with `MAXSMOOTH` produce residuals that can be used to constrain the properties of the first galaxies via the 21-cm signal. Future work could also be performed to extend the EDGES-style correction to fit for errors in the sky model and beam directivity in tandem with the 21-cm signal [e.g. 227].

### 8.1.2 Cosmology forecasts from the Dark Ages

Many of the techniques employed in this thesis can also be applied to studies of the 21-cm signal from other experiments and from the Dark Ages. The Dark Ages signal is of particular interest because it is independent of astrophysics, strongly dependent on Cosmology and potentially dependent on dark matter. As such, it can act as a very high redshift probe of the baryon,  $\Omega_b$ , and dark matter,  $\Omega_c$ , densities.

In principle, many of the techniques developed in this thesis and applied in chapter 5 and chapter 6 can be used to make predictions about the constraining power of a detection of the dark ages signal on key cosmological parameters. Future work is planned to do this for the Cambridge based experiment CosmoCube and to assess whether the constraints from the dark ages signal could break degeneracies in or improve the constraints from Planck and DES, for example, using `MARGARINE`.

### 8.1.3 Breaking model dependencies and likelihood reweighting

Extensions to `MARGARINE` are also planned to improve the accuracy of the density estimation and extend the code to emulate densities at discrete steps between the prior and posterior from the original analysis.

Synergies between `MARGARINE` and likelihood reweighting can also be explored in the future. Likelihood reweighting has been used in the study of Gravitational Waves to translate posterior samples for model A with a fast likelihood to the posterior for model B with a slower likelihood using the following relationship

$$P_B(\theta|D, M_B) = P_A(\theta|D, M_A) \frac{L_B(\theta|D, M_B)}{L_A(\theta|D, M_A)} \quad (8.1)$$

where we have assumed the same prior for model A and model B on the same set of parameters [228].

The idea is then that you can fully sample model A with MCMC or Nested Sampling, and then you only need to call the slow likelihood a few thousand times rather than a few tens of thousands of times to millions of times. In practice, however, this approach can produce too few samples on  $P_B$  for it to be accurately described. One way around this is to use `MARGARINE` to emulate the posterior distribution and likelihood for model A which would allow you to generate an arbitrary number of samples on the posterior  $P_B$ .

One current issue facing the field of 21-cm cosmology is the number of different semi-numerical simulations and 1D radiative transfer codes modelling the 21-cm signal in circulation and being applied to different data sets. This makes it difficult to directly compare the astrophysical constraints from one analysis to those from another that used a different modelling of the 21-cm signal. Breaking this model dependency is essential for validating conclusions across different experimental efforts.

The degeneracy appears in two forms. Firstly, across contemporary analyses that use different simulations with different parametrisations of the astrophysics in the early Universe. Secondly, in the comparison of contemporary and historical analyses where it is more likely that two data sets were analysed with variants of

the same simulation parameterised by the same quantities but where the contemporary work relies on a slightly modified physical picture.

Likelihood reweighting will allow us to begin to break the second of these degeneracies. For example, imagine that the analysis in chapter 5 was performed with a model that propagated Lyman- $\alpha$  photons in straight lines and between then and the publication of chapter 6 the models were updated to include the scattering of Lyman- $\alpha$  photons. We can refer to the first model as model A and the second as model B, and we want to compare and maybe combine the constraints from SARAS2 and SARAS3 antennas on model B. We therefore need the posterior  $P_B(\theta|D_{\text{SARAS2}}, M_B)$  but we have  $P_A(\theta|D_{\text{SARAS2}}, M_A)$  and we can translate between the two using equation (8.1).

### 8.1.4 Exploring synergies between 21-cm cosmology and other probes of the infant Universe

Finally, it will become increasingly important, as better data becomes available, to pursue synergies between 21-cm cosmology and other probes of the infant Universe, such as observations of distant galaxies with JWST. While JWST and experiments like REACH and HERA do not directly target the same signals, they target similar periods in cosmic history and observations of individual galaxies and the 21-cm signal can be used to constrain quantities like the neutral fraction of hydrogen as a function of redshift,  $x_{HI}(z)$ .

In practice, we can use neural networks to learn the relationship between parameters like the star formation rate and the CMB optical depth and the neutral fraction of hydrogen, as illustrated in chapter 3. This allows us to translate the posterior from HERA and SARAS3 for the parameters describing the 21-cm signal outlined in the previous chapter into a posterior on the redshift of reionization,  $z_{\text{EoR}}$  when  $x_{HI} \approx 0.5$  and the duration of reionization  $\Delta z = z(x_{HI} = 0.95) - z(x_{HI} = 0.05)$ . Learning this distribution with MARGARINE in turn gives us access to  $\log(\mathcal{P}_{\text{H+S}}(\theta|D, M))$  where  $\theta = \{\Delta z, z_{\text{EoR}}\}$ . Similarly, we can use observations of Lyman- $\alpha$  absorption in galaxy spectra from JWST to begin to constrain the properties of the epoch of reionization [229]. We can imagine deriving similar posterior emulators using MARGARINE for individual galaxies  $\log \mathcal{P}_{\text{JWST,GAL-i}}(\theta|D, M)$  and defining a combined likelihood

$$\log \mathcal{L}(\theta) = \log \mathcal{P}_{\text{H+S}}(\theta|D, M) + \sum_i \log \mathcal{P}_{\text{JWST,GAL-i}}(\theta|D, M), \quad (8.2)$$

where the sum is over observations of different galaxies. In practice, we would sample this likelihood by sampling the astrophysical parameters describing star formation rates, radio and X-ray luminosities and the CMB optical depth and translating the resultant samples to estimates of  $z_{\text{EoR}}$  and  $\Delta z$  using the neural network emulator for  $x_{HI}$  to be passed to the MARGARINE emulators inside our likelihood function.

Similar approaches can be taken to relate constraints on UV luminosity functions for example from JWST and 21-cm probes using MARGARINE and I intend to investigate this in the future.

## 8.2 Conclusion

In this thesis, we have outlined the development of several novel techniques and tools for data analysis in 21-cm cosmology. Each has applications outside the field, which are generally alluded to in the relevant chapters.

In chapter 2 we describe an algorithm for fitting smooth functions to data sets and discuss this in the context of 21-cm cosmology. However, we note that this can be applied to any field where the separation of non-smooth structure from smooth data is crucial.

In chapter 3 we discuss the development of a novel neural network 21-cm signal emulator and demonstrate its accuracy and speed. The preprocessing techniques, in particular the outlined resampling, can be applied to many fields, as can inputting the independent variable into the network to produce small and efficient emulators. One example is the emulation of the CMB power spectrum.

In chapter 4 we build the framework to perform marginal Bayesian statistics and demonstrate the calculation of key statistical quantities on sub spaces of larger distributions. This framework has numerous applications in any field where Bayesian inference is employed.

In chapter 5, chapter 6 and chapter 7 we apply these tools to data from the SARAS2, SARAS3 and HERA. In doing so, we demonstrate several key results;

- Systematics do not need to be a limiting factor in our ability to constrain the properties of the infant Universe, if appropriately recovered in the data and modelled. Although we note that in the event of a detection, the systematics need to be physically well understood and modelled for that detection to be accepted by the community.
- We can combine constraints to marginalise over the uncertainty in systematic modelling based on the evidence for each fit.
- Combining constraints across different data sets looking for the global 21-cm signal and the power spectrum leads to significant improvement in our understanding of the first galaxies and improvements in our constraints on the magnitude of each signal.
- Combining constraints from data sets across a wide range of redshifts improves our knowledge of the history of the infant Universe in comparison to extrapolations of constraints outside narrow redshift ranges.

The SARAS3 constraints are the first of their kind at such high redshift and although weak we are able to disfavour at 68% confidence galaxies with high star formation efficiencies above 3% in combination with minimum halo masses of less than  $\approx 9 \times 10^8$  solar masses at  $z = 20$ . In addition, we disfavour scenarios where galaxies have high radio luminosities per star formation rate in combination with low X-ray luminosities per star formation rate at  $z = 20$ . These limits are further improved through the combination of constraints from HERA and SARAS3.

Throughout the thesis, we draw several conclusions about the absorption feature reported by the EDGES collaboration [8]. We find in chapter 2, for example, that the data contains a sinusoidal systematic in line with the findings of [60, 128–130]. We also find in chapter 7 that while the HERA and SARAS3 data individually allow for physically motivated models, that have excess radio backgrounds above the CMB from radio galaxies, with depths consistent with the EDGES absorption feature the combination of the constraints from the two experiments rules these signals out with greater than  $2\sigma$  significance.

In the thesis, we explore a series of models with excess radio backgrounds above the CMB produced separately by radio galaxies and from a high redshift synchrotron source. While both models are partially motivated by the EDGES absorption feature, there is some indication of an excess radio background from measurements by ARCADE2 [48] and LWA [49]. The modelling of the excess radio background in the ARCADE2 and LWA work is similar to that used here for the high redshift synchrotron excess. As such, we compare our constraints on this model of the excess from SARAS3 in chapter 6 in the context of constraints from HERA. We find that while our results from SARAS3 disfavour the LWA and ARCADE2 measurements at greater than 68% confidence, there is still a non-zero probability that the ARCADE2 and LWA measurements are genuine excesses

despite concerns about the foreground modelling in those works [50]. We also find that our constraints on this radio background model are not as strong as those from HERA. However, we caution that our constraints are at different redshifts to HERA and the measurements from LWA and ARCADE2 meaning that each instrument may be probing a different population of sources.

The work in this thesis advances the library of available tools for 21-cm analysis, effectively demonstrates the application of Bayesian inference in the field, and advances our knowledge of the first galaxies.



## BIBLIOGRAPHY

- [1] Harry Bevins, Will Handley, Pablo Lemos, Peter Sims, Eloy de Lera Acedo, and Anastasia Fialkov. Marginal Bayesian Statistics Using Masked Autoregressive Flows and Kernel Density Estimators with Examples in Cosmology. *arXiv e-prints*, page arXiv:2207.11457, July 2022.
- [2] Harry T. J. Bevins, Stefan Heimersheim, Irene Abril-Cabezas, Anastasia Fialkov, Eloy de Lera Acedo, William Handley, Saurabh Singh, and Rennan Barkana. Joint analysis constraints on the physics of the first galaxies with low frequency radio astronomy data. *arXiv e-prints*, page arXiv:2301.03298, January 2023.
- [3] H. T. J. Bevins, W. J. Handley, A. Fialkov, E. de Lera Acedo, L. J. Greenhill, and D. C. Price. MAXSMOOTH: rapid maximally smooth function fitting with applications in Global 21-cm cosmology. *MNRAS*, 502(3):4405–4425, April 2021.
- [4] H. T. J. Bevins, W. J. Handley, A. Fialkov, E. de Lera Acedo, and K. Javid. GLOBALEMU: a novel and robust approach for emulating the sky-averaged 21-cm signal from the cosmic dawn and epoch of reionization. *MNRAS*, 508(2):2923–2936, December 2021.
- [5] Harry T. J. Bevins, William J. Handley, Pablo Lemos, Peter H. Sims, Eloy de Lera Acedo, Anastasia Fialkov, and Justin Alsing. Removing the fat from your posterior samples with margarine. *arXiv e-prints*, page arXiv:2205.12841, 2022.
- [6] H. T. J. Bevins, E. de Lera Acedo, A. Fialkov, W. J. Handley, S. Singh, R. Subrahmanyam, and R. Barkana. A Comprehensive Bayesian re-analysis of the SARAS2 data from the Epoch of Reionization. *arXiv e-prints*, page arXiv:2201.11531, January 2022.
- [7] H. T. J. Bevins, A. Fialkov, E. de Lera Acedo, W. J. Handley, S. Singh, R. Subrahmanyam, and R. Barkana. Astrophysical constraints from the saras3 non-detection of the cosmic dawn sky-averaged 21-cm signal. *Nature Astronomy*, December 2022.
- [8] Judd D. Bowman, Alan E. E. Rogers, Raul A. Monsalve, Thomas J. Mozdzen, and Nivedita Mahesh. An absorption profile centred at 78 megahertz in the sky-averaged spectrum. *Nature*, 555(7694):67–70, March 2018.
- [9] D. C. Price, L. J. Greenhill, A. Fialkov, G. Bernardi, H. Garsden, B. R. Barsdell, J. Kocz, M. M. Anderson, S. A. Bourke, J. Craig, M. R. Dexter, J. Dowell, M. W. Eastwood, T. Eftekhar, S. W. Ellingson, G. Hallinan, J. M. Hartman, R. Kimberk, T. Joseph W. Lazio, S. Leiker, D. MacMahon, R. Monroe, F. Schinzel, G. B. Taylor, E. Tong, D. Werthimer, and D. P. Woody. Design and characterization of the Large-aperture Experiment to Detect the Dark Age (LEDA) radiometer systems. *MNRAS*, 478(3):4193–4213, 2018.
- [10] E. de Lera Acedo, D. I. L. de Villiers, N. Razavi-Ghods, W. Handley, A. Fialkov, A. Magro, D. Anstey, H. T. J. Bevins, R. Chiello, J. Cumner, A. T. Josaitis, I. L. V. Roque, P. H. Sims, K. H. Scheutwinkel, P. Alexander, G. Bernardi, S. Carey, J. Cavillot, W. Croukamp, J. A. Ely, T. Gessey-Jones, Q. Gueuning,

- R. Hills, G. Kulkarni, R. Maiolino, P. D. Meerburg, S. Mittal, J. R. Pritchard, E. Puchwein, A. Saxena, E. Shen, O. Smirnov, M. Spinelli, and K. Zarb-Adami. The REACH radiometer for detecting the 21-cm hydrogen signal from redshift  $z \approx 7.5 - 28$ . *Nat. Astron.*, pages 1–15, July 2022. Publisher: Nature Publishing Group.
- [11] Eli Visbal, Rennan Barkana, Anastasia Fialkov, Dmitriy Tseliakhovich, and Christopher M. Hirata. The signature of the first stars in atomic hydrogen at redshift 20. *Nature*, 487(7405):70–73, July 2012.
  - [12] Anastasia Fialkov and Rennan Barkana. The rich complexity of 21-cm fluctuations produced by the first stars. *MNRAS*, 445(1):213–224, November 2014.
  - [13] Aviad Cohen, Anastasia Fialkov, Rennan Barkana, and Matan Lotem. Charting the parameter space of the global 21-cm signal. *MNRAS*, 472(2):1915–1931, December 2017.
  - [14] Itamar Reis, Anastasia Fialkov, and Rennan Barkana. The subtlety of Ly  $\alpha$  photons: changing the expected range of the 21-cm signal. *MNRAS*, 506(4):5479–5493, October 2021.
  - [15] Aviad Cohen, Anastasia Fialkov, Rennan Barkana, and Raul A. Monsalve. Emulating the global 21-cm signal from Cosmic Dawn and Reionization. *MNRAS*, 495(4):4845–4859, June 2020.
  - [16] Saurabh Singh, Ravi Subrahmanyam, N. Udaya Shankar, Mayuri Sathyaranayana Rao, B. S. Girish, A. Raghunathan, R. Somashekar, and K. S. Srivani. SARAS 2: a spectral radiometer for probing cosmic dawn and the epoch of reionization through detection of the global 21-cm signal. *Exp. Astron.*, 45(2):269–314, April 2018.
  - [17] B. S. Girish, K. S. Srivani, Ravi Subrahmanyam, N. Udaya Shankar, Saurabh Singh, T. Jishnu Nambissan, Mayuri Sathyaranayana Rao, R. Somashekar, and A. Raghunathan. SARAS CD/EoR Radiometer: Design and Performance of the Digital Correlation Spectrometer. *J. of Astron. Instrum.*, 9(2):2050006–70, January 2020.
  - [18] Agaram Raghunathan, Ravi Subrahmanyam, N. Udaya Shankar, Saurabh Singh, Jishnu Nambissan, K. Kavitha, Nivedita Mahesh, R. Somashekar, Gaddam Sindhu, B. S. Girish, K. S. Srivani, and Mayuri S. Rao. A Floating Octave Bandwidth Cone-Disc Antenna for Detection of Cosmic Dawn. *arXiv e-prints*, page arXiv:2104.03522, April 2021.
  - [19] Jishnu Nambissan T., Ravi Subrahmanyam, R. Somashekar, N. Udaya Shankar, Saurabh Singh, A. Raghunathan, B. S. Girish, K. S. Srivani, and Mayuri Sathyaranayana Rao. SARAS 3 CD/EoR radiometer: design and performance of the receiver. *Experimental Astronomy*, January 2021.
  - [20] David R. DeBoer, Aaron R. Parsons, James E. Aguirre, Paul Alexander, Zaki S. Ali, Adam P. Beardsley, Gianni Bernardi, Judd D. Bowman, Richard F. Bradley, Chris L. Carilli, Carina Cheng, Eloy de Lera Acedo, Joshua S. Dillon, Aaron Ewall-Wice, Gcobisa Fadana, Nicolas Fagnoni, Randall Fritz, Steve R. Furlanetto, Brian Glendenning, Bradley Greig, Jasper Grobbelaar, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, MacCalvin Kariseb, Saul A. Kohn, Telalo Lekalake, Adrian Liu, Anita Loots, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary Martinot, Nathan Mathison, Eunice Matsetela, Andrei Mesinger, Miguel F. Morales, Abraham R. Neben, Nipanjana Patra, Samantha Pieterse, Jonathan C. Poher, Nima Razavi-Ghods, Jon Ringuelette, James Robnett, Kathryn Rosie, Raddwine Sell, Craig Smith, Angelo Syce, Max Tegmark, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. Hydrogen Epoch of Reionization Array (HERA). *PASP*, 129(974):045001, April 2017.
  - [21] James E. Gunn, Walter A. Siegmund, Edward J. Mannery, Russell E. Owen, Charles L. Hull, R. French Leger, Larry N. Carey, Gillian R. Knapp, Donald G. York, William N. Boroski, Stephen M. Kent,

- Robert H. Lupton, Constance M. Rockosi, Michael L. Evans, Patrick Waddell, John E. Anderson, James Annis, John C. Barentine, Larry M. Bartoszek, Steven Bastian, Stephen B. Bracker, Howard J. Brewington, Charles I. Briegel, Jon Brinkmann, Yorke J. Brown, Michael A. Carr, Paul C. Czarapata, Craig C. Drennan, Thomas Dombeck, Glenn R. Federwitz, Bruce A. Gillespie, Carlos Gonzales, Sten U. Hansen, Michael Harvanek, Jeffrey Hayes, Wendell Jordan, Ellyne Kinney, Mark Klaene, S. J. Kleinman, Richard G. Kron, Jurek Kresinski, Glenn Lee, Siriluk Limmongkol, Carl W. Lindenmeyer, Daniel C. Long, Craig L. Loomis, Peregrine M. McGehee, Paul M. Mantsch, Jr. Neilsen, Eric H., Richard M. Neswold, Peter R. Newman, Atsuko Nitta, Jr. Peoples, John, Jeffrey R. Pier, Peter S. Prieto, Angela Prosapio, Claudio Rivetta, Donald P. Schneider, Stephanie Snedden, and Shu-i. Wang. The 2.5 m Telescope of the Sloan Digital Sky Survey. *aj*, 131(4):2332–2359, April 2006.
- [22] Planck Collaboration VI. Planck 2018 results - vi. cosmological parameters. *A&A*, 641:A6, 2020.
- [23] Brenna Flaugher and Chris Bebek. The Dark Energy Spectroscopic Instrument (DESI). In Suzanne K. Ramsay, Ian S. McLean, and Hideki Takami, editors, *Ground-based and Airborne Instrumentation for Astronomy V*, volume 9147 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 91470S, July 2014.
- [24] J. W. Fowler, M. D. Niemack, S. R. Dicker, A. M. Aboobaker, P. A. R. Ade, E. S. Battistelli, M. J. Devlin, R. P. Fisher, M. Halpern, P. C. Hargrave, A. D. Hincks, M. Kaul, J. Klein, J. M. Lau, M. Limon, T. A. Marriage, P. D. Mauskopf, L. Page, S. T. Staggs, D. S. Swetz, E. R. Switzer, R. J. Thornton, and C. E. Tucker. Optical design of the Atacama Cosmology Telescope and the Millimeter Bolometric Array Camera. *ao*, 46(17):3444–3454, June 2007.
- [25] Frank J. Qu, Blake D. Sherwin, Mathew S. Madhavacheril, Dongwon Han, Kevin T. Crowley, Irene Abril-Cabezas, Peter A. R. Ade, Simone Aiola, Tommy Alford, Mandana Amiri, Stefania Amodeo, Rui An, Zachary Atkins, Jason E. Austermann, Nicholas Battaglia, Elia Stefano Battistelli, James A. Beall, Rachel Bean, Benjamin Beringue, Tanay Bhandarkar, Emily Biermann, Boris Bolliet, J Richard Bond, Hongbo Cai, Erminia Calabrese, Victoria Calafut, Valentina Capalbo, Felipe Carrero, Julien Carron, Anthony Challinor, Grace E. Chesmore, Hsiao-mei Cho, Steve K. Choi, Susan E. Clark, Rodrigo Córdova Rosado, Nicholas F. Cothard, Kevin Coughlin, William Coulton, Roohi Dalal, Omar Darwish, Mark J. Devlin, Simon Dicker, Peter Doze, Cody J. Duell, Shannon M. Duff, Adriaan J. Duivenvoorden, Jo Dunkley, Rolando Dünner, Valentina Fanfani, Max Fankhanel, Gerrit Farren, Simone Ferraro, Rodrigo Freundt, Brittany Fuzia, Patricio A. Gallardo, Xavier Garrido, Vera Gluscevic, Joseph E. Golec, Yilun Guan, Mark Halpern, Ian Harrison, Matthew Hasselfield, Erin Healy, Shawn Henderson, Brandon Hensley, Carlos Hervás-Caimapo, J. Colin Hill, Gene C. Hilton, Matt Hilton, Adam D. Hincks, Renée Hložek, Shuay-Pwu Patty Ho, Zachary B. Huber, Johannes Hubmayr, Kevin M. Huffenberger, John P. Hughes, Kent Irwin, Giovanni Isopi, Hidde T. Jense, Ben Keller, Joshua Kim, Kenda Knowles, Brian J. Koopman, Arthur Kosowsky, Darby Kramer, Aleksandra Kusiak, Adrien La Posta, Alex Lague, Victoria Lakey, Eunseong Lee, Zack Li, Yaqiong Li, Michele Limon, Martine Lokken, Thibaut Louis, Marius Lungu, Niall MacCrann, Amanda MacInnis, Diego Maldonado, Felipe Maldonado, Maya Mallaby-Kay, Gabriela A. Marques, Jeff McMahon, Yogesh Mehta, Felipe Menanteau, Kavilan Moodley, Thomas W. Morris, Tony Mroczkowski, Sigurd Naess, Toshiya Namikawa, Federico Nati, Laura Newburgh, Andrina Nicola, Michael D. Niemack, Michael R. Nolta, John Orlowski-Scherer, Lyman A. Page, Shivam Pandey, Bruce Partridge, Heather Prince, Roberto Puddu, Federico Radiconi, Naomi Robertson, Felipe Rojas, Tai Sakuma, Maria Salatino, Emmanuel Schaan, Benjamin L. Schmitt, Neelima Sehgal, Shabbir Shaikh,

- Carlos Sierra, Jon Sievers, Cristóbal Sifón, Sara Simon, Rita Sonka, David N. Spergel, Suzanne T. Staggs, Emilie Storer, Eric R. Switzer, Niklas Tampier, Robert Thornton, Hy Trac, Jesse Treu, Carole Tucker, Joel Ulluom, Leila R. Vale, Alexander Van Engelen, Jeff Van Lanen, Joshiwa van Marrewijk, Cristian Vargas, Eve M. Vavagiakis, Kasey Wagoner, Yuhan Wang, Lukas Wenzl, Edward J. Wollack, Zhilei Xu, Fernando Zago, and Kaiwen Zhang. The Atacama Cosmology Telescope: A Measurement of the DR6 CMB Lensing Power Spectrum and its Implications for Structure Growth. *arXiv e-prints*, page arXiv:2304.05202, April 2023.
- [26] Rogier A. Windhorst, Seth H. Cohen, Rolf A. Jansen, Chris Conselice, and Haojing Yan. How JWST can measure first light, reionization and galaxy assembly. *Nature*, 50(1-3):113–120, March 2006.
- [27] Marcia Rieke and the JADES Collaboration. JADES Initial Data Release for the Hubble Ultra Deep Field: Revealing the Faint Infrared Sky with Deep JWST NIRCam Imaging. *arXiv e-prints*, page arXiv:2306.02466, June 2023.
- [28] Steven R. Furlanetto, S. Peng Oh, and Frank H. Briggs. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. *Phys. Rep.*, 433(4-6):181–301, October 2006.
- [29] Piero Madau, Avery Meiksin, and Martin J. Rees. 21 Centimeter Tomography of the Intergalactic Medium at High Redshift. *ApJ*, 475(2):429–444, February 1997.
- [30] P. A. Shaver, R. A. Windhorst, P. Madau, and A. G. de Bruyn. Can the reionization epoch be detected as a global signature in the cosmic background? *A&A*, 345:380–390, May 1999.
- [31] Stephen Serjeant, James Bolton, Poshak Gandhi, Ben Stappers, Paolo Mazzali, Aprajita Verma, and Noelia E. D. Noël. UK Astronomy Science and Technology Roadmap: STFC Astronomy Advisory Panel Roadmap 2022. *arXiv e-prints*, page arXiv:2301.05457, January 2023.
- [32] J. R. Pritchard and A. Loeb. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, August 2012.
- [33] Rennan Barkana. The rise of the first stars: Supersonic streaming, radiative feedback, and 21-cm cosmology. *Phys. Rep.*, 645:1–59, July 2016.
- [34] Andrei Mesinger. *The Cosmic 21-cm Revolution; Charting the first billion years of our universe*. 2019.
- [35] Rennan Barkana and Abraham Loeb. Unusually Large Fluctuations in the Statistics of Galaxy Formation at High Redshift. *apj*, 609(2):474–481, July 2004.
- [36] Steven R. Furlanetto. The global 21-centimeter background from high redshifts. *MNRAS*, 371(2):867–878, September 2006.
- [37] Sandeep Kumar Acharya, Jiten Dhandha, and Jens Chluba. Can accreting primordial black holes explain the excess radio background? *MNRAS*, 517(2):2454–2461, December 2022.
- [38] Jack Burns, Stuart Bale, Richard Bradley, Z. Ahmed, S. W. Allen, J. Bowman, S. Furlanetto, R. MacDowall, J. Mirocha, B. Nhan, M. Pivovaroff, M. Pulupa, D. Rapetti, A. Slosar, and K. Tauscher. Global 21-cm Cosmology from the Farside of the Moon. *arXiv e-prints*, page arXiv:2103.05085, March 2021.
- [39] S. A. Wouthuysen. On the excitation mechanism of the 21-cm (radio-frequency) interstellar hydrogen emission line. *AJ*, 57:31–32, 1952.
- [40] G. B. Field. The Spin Temperature of Intergalactic Neutral Hydrogen. *ApJ*, 129:536, May 1959.
- [41] R. Barkana and A. Loeb. In the beginning: the first sources of light and the reionization of the universe. *Phys. Rep.*, 349(2):125–238, July 2001.
- [42] Leonid Chuzhoy and Paul R. Shapiro. Heating and Cooling of the Early Intergalactic Medium by Resonance Photons. *ApJ*, 655(2):843–846, February 2007.

- [43] Pablo Villanueva-Domingo, Olga Mena, and Jordi Miralda-Escudé. Maximum amplitude of the high-redshift 21-cm absorption feature. *Physical Review D*, 101(8):083502, April 2020. Publisher: American Physical Society.
- [44] Tejaswi Venumadhav, Liang Dai, Alexander Kaurov, and Matias Zaldarriaga. Heating of the intergalactic medium by the cosmic microwave background during cosmic dawn. *Phys. Rev. D*, 98(10):103513, November 2018.
- [45] Shikhar Mittal and Girish Kulkarni. Ly  $\alpha$  coupling and heating at cosmic dawn. *MNRAS*, 503(3):4264–4275, May 2021.
- [46] T. Fragos, B. D. Lehmer, S. Naoz, A. Zezas, and A. Basu-Zych. ENERGY FEEDBACK FROM x-RAY BINARIES IN THE EARLY UNIVERSE. *ApJ*, 776(2):L31, oct 2013.
- [47] Eli Visbal, Zoltán Haiman, and Greg L. Bryan. Direct collapse black hole formation from synchronized pairs of atomic cooling haloes. *MNRAS*, 445(1):1056–1063, November 2014.
- [48] D. J. Fixsen, A. Kogut, S. Levin, M. Limon, P. Lubin, P. Mirel, M. Seiffert, J. Singal, E. Wollack, T. Villela, and C. A. Wuensche. ARCADE 2 MEASUREMENT OF THE ABSOLUTE SKY BRIGHTNESS AT 3-90 GHz. *The Astrophysical Journal*, 734(1):5, May 2011. Publisher: American Astronomical Society.
- [49] Jayce Dowell and Greg B. Taylor. The Radio Background below 100 MHz. *The Astrophysical Journal*, 858(1):L9, May 2018. Publisher: American Astronomical Society.
- [50] Ravi Subrahmanyan and Ramanath Cowsik. IS THERE AN UNACCOUNTED FOR EXCESS IN THE EXTRAGALACTIC COSMIC RADIO BACKGROUND? *The Astrophysical Journal*, 776(1):42, September 2013. Publisher: American Astronomical Society.
- [51] Robert Brandenberger, Bryce Cyr, and Rui Shi. Constraints on superconducting cosmic strings from the global 21-cm signal before reionization. *Journal of Cosmology and Astroparticle Physics*, 2019(09):009–009, sep 2019.
- [52] Sean Fraser, Andi Hektor, Gert Hütsi, Kristjan Kannike, Carlo Marzo, Luca Marzola, Antonio Racioppi, Martti Raidal, Christian Spethmann, Ville Vaskonen, and Hardi Veermäe. The edges 21 cm anomaly and properties of dark matter. *Physics Letters B*, 785:159–164, 2018.
- [53] Maxim Pospelov, Josef Pradler, Joshua T. Ruderman, and Alfredo Urbano. Room for new physics in the rayleigh-jeans tail of the cosmic microwave background. *Phys. Rev. Lett.*, 121:031103, Jul 2018.
- [54] Saurabh Singh, Nambissan T. Jishnu, Ravi Subrahmanyan, N. Udaya Shankar, B. S. Girish, A. Raghunathan, R. Somashekar, K. S. Srivani, and Mayuri Sathyaranayana Rao. On the detection of a cosmic dawn signal in the radio background. *Nature Astronomy*, 6:607–617, February 2022.
- [55] T. J. Mozdzen, J. D. Bowman, R. A. Monsalve, and A. E. E. Rogers. Improved measurement of the spectral index of the diffuse radio background between 90 and 190 MHz. *MNRAS*, 464(4):4995–5002, February 2017.
- [56] M. Spinelli, G. Bernardi, H. Garsden, L. J. Greenhill, A. Fialkov, J. Dowell, and D. C. Price. Spectral index of the Galactic foreground emission in the 50-87 MHz range. *MNRAS*, 505(2):1575–1588, August 2021.
- [57] Dominic Anstey, Eloy de Lera Acedo, and Will Handley. A General Bayesian Framework for Foreground Modelling and Chromaticity Correction for Global 21cm Experiments. *arXiv:2010.09644 [astro-ph]*, November 2020. arXiv: 2010.09644.
- [58] Dominic Anstey, John Cumner, Eloy de Lera Acedo, and Will Handley. Informing antenna design for sky-averaged 21-cm experiments using a simulated Bayesian data analysis pipeline. *MNRAS*, 509(4):4679–4693, February 2022.

- [59] Dominic Anstey, Eloy de Lera Acedo, and Will Handley. Use of Time Dependent Data in Bayesian Global 21cm Foreground and Signal Modelling. *arXiv e-prints*, page arXiv:2210.04707, October 2022.
- [60] Richard F. Bradley, Keith Tauscher, David Rapetti, and Jack O. Burns. A Ground Plane Artifact that Induces an Absorption Profile in Averaged Spectra from Global 21 cm Measurements, with Possible Application to EDGES. *ApJ*, 874(2):153, April 2019.
- [61] Mayuri Sathyaranayana Rao, Ravi Subrahmanyam, N. Udaya Shankar, and Jens Chluba. Modeling the Radio Foreground for Detection of CMB Spectral Distortions from the Cosmic Dawn and the Epoch of Reionization. *ApJ*, 840(1):33, May 2017.
- [62] Mayuri Sathyaranayana Rao, Ravi Subrahmanyam, N. Udaya Shankar, and Jens Chluba. On the detection of spectral ripples from the recombination epoch. *ApJ*, 810(1):3, 2015.
- [63] H. K. Vedantham, L. V. E. Koopmans, A. G. de Bruyn, S. J. Wijnholds, B. Ciardi, and M. A. Brentjens. Chromatic effects in the 21 cm global signal from the cosmic dawn. *MNRAS*, 437(2):1056–1069, January 2014.
- [64] Abhirup Datta, Richard Bradley, Jack O. Burns, Geraint Harker, Attila Komjathy, and T. Joseph W. Lazio. Effects Of The Ionosphere On Ground-Based Detection Of The Global 21 CM Signal From The Cosmic Dawn And The Dark Ages. *arXiv e-prints*, page arXiv:1409.0513, September 2014.
- [65] Emma Shen, Dominic Anstey, Eloy de Lera Acedo, Anastasia Fialkov, and Will Handley. Quantifying ionospheric effects on global 21-cm observations. *MNRAS*, 503(1):344–353, May 2021.
- [66] Emma Shen, Dominic Anstey, Eloy de Lera Acedo, and Anastasia Fialkov. Bayesian data analysis for sky-averaged 21-cm experiments in the presence of ionospheric effects. *MNRAS*, 515(3):4565–4573, September 2022.
- [67] L. Philip, Z. Abdurashidova, H. C. Chiang, N. Ghazi, A. Gumba, H. M. Heilgendorff, J. M. Jáuregui-García, K. Malepe, C. D. Nunhokee, J. Peterson, J. L. Sievers, V. Simes, and R. Spann. Probing Radio Intensity at High-Z from Marion: 2017 Instrument. *Journal of Astronomical Instrumentation*, 8(2):1950004, January 2019.
- [68] H. C. Chiang, T. Dyson, E. Egan, S. Eyono, N. Ghazi, J. Hickish, J. M. Jáuregui-Garcia, V. Manukha, T. Ménard, T. Moso, J. Peterson, L. Philip, J. L. Sievers, and S. Tartakovsky. The Array of Long Baseline Antennas for Taking Radio Observations from the Sub-Antarctic. *Journal of Astronomical Instrumentation*, 9(4):2050019–564, January 2020.
- [69] Saurabh Singh, Ravi Subrahmanyam, N. Udaya Shankar, Mayuri Sathyaranayana Rao, Anastasia Fialkov, Aviad Cohen, Rennan Barkana, B. S. Girish, A. Raghunathan, R. Somashekar, and K. S. Srivani. First Results on the Epoch of Reionization from First Light with SARAS 2. *ApJ Lett.*, 845(2):L12, August 2017.
- [70] Saurabh Singh, Ravi Subrahmanyam, N. Udaya Shankar, Mayuri Sathyaranayana Rao, Anastasia Fialkov, Aviad Cohen, Rennan Barkana, B. S. Girish, A. Raghunathan, R. Somashekar, and K. S. Srivani. SARAS 2 Constraints on Global 21 cm Signals from the Epoch of Reionization. *ApJ*, 858(1):54, May 2018.
- [71] Neil Bassett, David Rapetti, Keith Tauscher, Bang D. Nhan, David D. Bordenave, Joshua J. Hibbard, and Jack O. Burns. Lost Horizon: Quantifying the Effect of Local Topography on Global 21 cm Cosmology Data Analysis. *ApJ*, 923(1):33, December 2021.
- [72] S. A. K. Leeney, W. J. Handley, and E. de Lera Acedo. A Bayesian approach to RFI mitigation. *arXiv e-prints*, page arXiv:2211.15448, November 2022.

- [73] Jack O. Burns, Robert MacDowall, Stuart Bale, Gregg Hallinan, Neil Bassett, and Alex Hegedus. Low Radio Frequency Observations from the Moon Enabled by NASA Landed Payload Missions. *The Planetary Science Journal*, 2(2):44, April 2021.
- [74] K. H. Scheutwinkel, E. de Lera Acedo, and W. Handley. Bayesian evidence-driven diagnosis of instrumental systematics for sky-averaged 21-cm cosmology experiments. *Publ. Astron. Soc. Australia*, 39:e052, October 2022.
- [75] M. Spinelli, G. Kyriakou, G. Bernardi, P. Bolli, L. J. Greenhill, A. Fialkov, and H. Garsden. Antenna beam characterization for the global 21-cm experiment LEDA and its impact on signal model parameter reconstruction. *MNRAS*, 515(2):1580–1597, September 2022.
- [76] M. Spinelli, G. Bernardi, and M. G. Santos. Simulations of Galactic polarized synchrotron emission for Epoch of Reionization observations. *MNRAS*, 479(1):275–283, September 2018.
- [77] Marta Spinelli, Gianni Bernardi, and Mario G. Santos. On the contamination of the global 21-cm signal from polarized foregrounds. *MNRAS*, 489(3):4007–4015, November 2019.
- [78] Thomas Bayes and Richard Price. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.
- [79] K. H. Scheutwinkel, W. Handley, and E. de Lera Acedo. Bayesian evidence-driven likelihood selection for sky-averaged 21-cm signal extraction. *arXiv e-prints*, page arXiv:2204.04491, April 2022.
- [80] Anchal Saxena, Alex Cole, Simon Gazagnes, P. Daniel Meerburg, Christoph Weniger, and Samuel J. Witte. Constraining the X-ray heating and reionization using 21-cm power spectra with Marginal Neural Ratio Estimation. *arXiv e-prints*, page arXiv:2303.07339, March 2023.
- [81] Benjamin K Miller, Alex Cole, Patrick Forré, Gilles Louppe, and Christoph Weniger. Truncated marginal neural ratio estimation. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 129–143. Curran Associates, Inc., 2021.
- [82] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.
- [83] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC Hammer. *PASP*, 125(925):306, March 2013.
- [84] Greg Ashton, Noam Bernstein, Johannes Buchner, Xi Chen, Gábor Csányi, Andrew Fowlie, Farhan Feroz, Matthew Griffiths, Will Handley, Michael Habeck, Edward Higson, Michael Hobson, Anthony Lasenby, David Parkinson, Livia B. Pártay, Matthew Pitkin, Doris Schneider, Joshua S. Speagle, Leah South, John Veitch, Philipp Wacker, David J. Wales, and David Yallup. Nested sampling for physical scientists. *Nature Reviews Methods Primers*, 2(1):39, 2022.
- [85] John Skilling. Nested Sampling. *AIP Conference Proceedings*, 735(1):395–405, 2004.
- [86] Radford M. Neal. Slice Sampling. *arXiv e-prints*, page physics/0009028, September 2000.
- [87] Stuart Aitken and Ozgur Akman. Nested sampling for parameter inference in systems biology: Application to an exemplar circadian model. *BMC systems biology*, 7:72, 07 2013.
- [88] W. J. Handley, M. P. Hobson, and A. N. Lasenby. PolyChord: nested sampling for cosmology. *MNRAS*, 450:L61–L65, June 2015. arXiv: 1502.01856.
- [89] W. J. Handley, M. P. Hobson, and A. N. Lasenby. PolyChord: next-generation nested sampling. *MNRAS*, 453:4385–4399, November 2015. arXiv: 1506.00171.

- [90] Roberto Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49(2):71–104, March 2008.
- [91] Raul A. Monsalve, Anastasia Fialkov, Judd D. Bowman, Alan E. E. Rogers, Thomas J. Mozdzen, Aviad Cohen, Rennan Barkana, and Nivedita Mahesh. Results from EDGES High-Band. III. New Constraints on Parameters of the Early Universe. *ApJ*, 875(1):67, April 2019.
- [92] R Ghara, S K Giri, G Mellema, B Ciardi, S Zaroubi, I T Iliev, L V E Koopmans, E Chapman, S Gazagnes, B K Gehlot, A Ghosh, V Jelić, F G Mertens, R Mondal, J Schaye, M B Silva, K M B Asad, R Kooistra, M Mevius, A R Offringa, V N Pandey, and S Yatawatta. Constraining the intergalactic medium at  $z \approx 9.1$  using LOFAR Epoch of Reionization observations. *MNRAS*, 493(4):4728–4747, 02 2020.
- [93] Raghunath Ghara, Sambit K Giri, Benedetta Ciardi, Garrett Mellema, and Saleem Zaroubi. Constraining the state of the intergalactic medium during the epoch of reionization using MWA 21-cm signal observations. *MNRAS*, 03 2021. stab776.
- [94] F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *MNRAS*, 384(2):449–463, February 2008.
- [95] F. Feroz, M. P. Hobson, and M. Bridges. MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS*, 398(4):1601–1614, October 2009.
- [96] Farhan Feroz, Michael P. Hobson, Ewan Cameron, and Anthony N. Pettitt. Importance Nested Sampling and the MultiNest Algorithm. *The Open Journal of Astrophysics*, 2(1):10, November 2019.
- [97] J. Buchner, A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni, M. Salvato, J. Donley, and D. Kocevski. X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *A&A*, 564:A125, April 2014.
- [98] Christian Hellum Bye, Stephen K. N. Portillo, and Anastasia Fialkov. 21cmVAE: A VAE-based Emulator of the 21-cm Global Signal. *arXiv e-prints*, page arXiv:2107.05581, July 2021.
- [99] The HERA Collaboration, Zara Abdurashidova, Tyrone Adams, James E. Aguirre, Paul Alexander, Zaki S. Ali, Rushelle Baartman, Yanga Balfour, Rennan Barkana, Adam P. Beardsley, Gianni Bernardi, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Daniela Breitman, Philip Bull, Jacob Burba, Steve Carey, Chris L. Carilli, Carina Cheng, Samir Choudhuri, David R. DeBoer, Eloy de Lera Acedo, Matt Dexter, Joshua S. Dillon, John Ely, Aaron Ewall-Wice, Nicolas Fagnoni, Anastasia Fialkov, Randall Fritz, Steven R. Furlanetto, Kingsley Gale-Sides, Hugh Garsden, Brian Glendenning, Adélie Gorce, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Stefan Heimersheim, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, Nicholas S. Kern, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, David Lewis, Adrian Liu, Anita Loots, Yin-Zhe Ma, David H. E. MacMahon, Lourence Malan, Keith Malgas, Cresshim Malgas, Matthys Maree, Bradley Marero, Zachary E. Martinot, Lisa McBride, Andrei Mesinger, Jordan Mirocha, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Julian B. Muñoz, Steven G. Murray, Vighnesh Nagpal, Abraham R. Neben, Bojan Nikolic, Chuneeta D. Nunhokee, Hans Nuwegeld, Aaron R. Parsons, Robert Pascua, Nipanjana Patra, Samantha Pieterse, Yuxiang Qin, Nima Razavi-Ghods, James Robnett, Kathryn Rosie, Mario G. Santos, Peter Sims, Saurabh Singh, Craig Smith, Hilton Swarts, Nithyanandan Thyagarajan, Michael J. Wilensky, Peter K. G. Williams, Pieter van Wyngaarden, and Haoxuan Zheng. Improved Constraints on the 21 cm EoR Power Spectrum and the X-Ray Heating of the IGM with HERA Phase I Observations. *arXiv e-prints*, page arXiv:2210.04912, October 2022.

- [100] Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. MADE: Masked Autoencoder for Distribution Estimation. *arXiv e-prints*, page arXiv:1502.03509, February 2015.
- [101] Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *MNRAS*, 488(3):4440–4458, September 2019.
- [102] Garrelt Mellema, Léon V. E. Koopmans, Filipe A. Abdalla, Gianni Bernardi, Benedetta Ciardi, Soobash Daiboo, A. G. de Bruyn, Kanan K. Datta, Heino Falcke, Andrea Ferrara, Ilian T. Iliev, Fabio Iocco, Vibor Jelić, Hannes Jensen, Ronniy Joseph, Panos Labropoulos, Avery Meiksin, Andrei Mesinger, André R. Offringa, V. N. Pandey, Jonathan R. Pritchard, Mario G. Santos, Dominik J. Schwarz, Benoit Semelin, Harish Vedantham, Sarod Yatawatta, and Saleem Zaroubi. Reionization and the Cosmic Dawn with the Square Kilometre Array. *Exp. Astron.*, 36(1-2):235–318, August 2013.
- [103] L. V. E. Koopmans and LOFAR EoR KSP Team. Current Status of the LOFAR EoR Key Science Project. In Vibor Jelić and Thijs van der Hulst, editors, *Peering towards Cosmic Dawn*, volume 333, pages 71–76, May 2018.
- [104] B. K. Gehlot, F. G. Mertens, L. V. E. Koopmans, M. A. Brentjens, S. Zaroubi, B. Ciardi, A. Ghosh, M. Hatef, I. T. Iliev, Jelić, V., R. Kooistra, F. Krause, G. Mellema, M. Mevius, M. Mitra, A. R. Offringa, V. N. Pandey, A. M. Sardarabadi, J. Schaye, M. B. Silva, H. K. Vedantham, and S. Yatawatta. The first power spectrum limit on the 21-cm signal of neutral hydrogen during the Cosmic Dawn at  $z = 20\text{--}25$  from LOFAR. *MNRAS*, 488(3):4271–4287, September 2019.
- [105] R. Mondal, A. Fialkov, C. Fling, I. T. Iliev, R. Barkana, B. Ciardi, G. Mellema, S. Zaroubi, L. V. E. Koopmans, F. G. Mertens, B. K. Gehlot, R. Ghara, A. Ghosh, S. K. Giri, A. Offringa, and V. N. Pandey. Tight constraints on the excess radio background at  $z = 9.1$  from LOFAR. *MNRAS*, 498(3):4178–4191, November 2020.
- [106] Bradley Greig, Andrei Mesinger, Léon V E Koopmans, Benedetta Ciardi, Garrelt Mellema, Saleem Zaroubi, Sambit K Giri, Raghunath Ghara, Abhik Ghosh, Ilian T Iliev, Florent G Mertens, Rajesh Mondal, André R Offringa, and Vishambhar N Pandey. Interpreting LOFAR 21-cm signal upper limits at  $z \approx 9.1$  in the context of high- $z$  galaxy and reionization observations. *MNRAS*, 501(1):1–13, 11 2020.
- [107] Cathryn M. Trott, C. H. Jordan, S. Midgley, N. Barry, B. Greig, B. Pindor, J. H. Cook, G. Sleap, S. J. Tingay, D. Ung, P. Hancock, A. Williams, J. Bowman, R. Byrne, A. Chokshi, B. J. Hazelton, K. Hasegawa, D. Jacobs, R. C. Joseph, W. Li, J. L. B. Line, C. Lynch, B. McKinley, D. A. Mitchell, M. F. Morales, M. Ouchi, J. C. Pober, M. Rahimi, K. Takahashi, R. B. Wayth, R. L. Webster, M. Wilensky, J. S. B. Wyithe, S. Yoshiura, Z. Zhang, and Q. Zheng. Deep multiredshift limits on Epoch of Reionization 21 cm power spectra from four seasons of Murchison Widefield Array observations. *MNRAS*, 493(4):4711–4727, apr 2020.
- [108] Bradley Greig, Cathryn M Trott, Nichole Barry, Simon J Mutch, Bart Pindor, Rachel L Webster, and J Stuart B Wyithe. Exploring reionization and high- $z$  galaxy observables with recent multiredshift MWA upper limits on the 21-cm signal. *MNRAS*, 500(4):5322–5335, 11 2020.
- [109] Zara Abdurashidova, James E. Aguirre, Paul Alexander, Zaki S. Ali, Yanga Balfour, Rennan Barkana, Adam P. Beardsley, Gianni Bernardi, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Philip Bull, Jacob Burba, Steve Carey, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matt Dexter, Eloy de Lera Acedo, Joshua S. Dillon, John Ely, Aaron Ewall-Wice, Nicolas Fagnoni, Anastasia Fialkov, Randall Fritz, Steven R. Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Stefan Heimersheim, Jacqueline N. Hewitt,

- Jack Hickish, Daniel C. Jacobs, Austin Julius, Nicholas S. Kern, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Telalo Lekalake, David Lewis, Adrian Liu, Yin-Zhe Ma, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E. Martinot, Eunice Matsetela, Andrei Mesinger, Jordan Mirocha, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Julian B. Muñoz, Steven G. Murray, Abraham R. Neben, Bojan Nikolic, Chuneeta D. Nunhokee, Aaron R. Parsons, Nipanjana Patra, Samantha Pieterse, Jonathan C. Pober, Yuxiang Qin, Nima Razavi-Ghods, Itamar Reis, Jon Ringuette, James Robnett, Kathryn Rosie, Mario G. Santos, Sudipta Sikder, Peter Sims, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. HERA Phase I Limits on the Cosmic 21 cm Signal: Constraints on Astrophysics and Cosmology during the Epoch of Reionization. *ApJ*, 924(2):51, January 2022.
- [110] Gregory Paciga, Tzu-Ching Chang, Yashwant Gupta, Rajaram Nityanada, Julia Odegova, Ue-Li Pen, Jeffrey B. Peterson, Jayanta Roy, and Kris Sigurdson. The GMRT Epoch of Reionization experiment: a new upper limit on the neutral hydrogen power spectrum at  $z \approx 8.6$ . *mnras*, 413(2):1174–1183, May 2011.
- [111] Raul A. Monsalve, Alan E. E. Rogers, Judd D. Bowman, and Thomas J. Mozdzen. Calibration of the EDGES High-band Receiver to Observe the Global 21 cm Signature from the Epoch of Reionization. *ApJ*, 835(1):49, January 2017.
- [112] T. C. Voytek, A. Natarajan, J. M. J. García, J. B. Peterson, and O. López-Cruz. PROBING THE DARK AGES AT  $z \sim 20$ : THE SCI-HI 21 cm ALL-SKY SPECTRUM EXPERIMENT. *ApJ*, 782(1):L9, jan 2014.
- [113] M. Sokolowski, S. E. Tremblay, R. B. Wayth, S. J. Tingay, N. Clarke, P. Roberts, M. Waterson, R. D. Ekers, P. Hall, M. Lewis, M. Mossammaparast, S. Padhi, F. Schlagenhauf, A. Sutinjo, and J. Tickner. BIGHORNS - Broadband Instrument for Global HydrOgen Reionisation Signal. *Publ. Astron. Soc. Australia*, 32:e004, February 2015.
- [114] R. Monsalve. MIST Global 21cm Experiment, 2022.
- [115] E. de Lera Acedo. Reach: Radio experiment for the analysis of cosmic hydrogen. In *2019 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, pages 0626–0629, 2019.
- [116] J. Cumner, E. De Lera Acedo, D. I. L. de Villiers, D. Anstey, C. I. Kolitsidas, B. Gurdon, N. Fagnoni, P. Alexander, G. Bernardi, H. T. J. Bevins, S. Carey, J. Cavillot, R. Chiello, C. Craeye, W. Croukamp, J. A. Ely, A. Fialkov, T. Gessey-Jones, Q. Gueuning, W. Handley, R. Hills, A. T. Josaitis, G. Kulkarni, A. Magro, R. Maiolino, P. D. Meerburg, S. Mittal, J. R. Pritchard, E. Puchwein, N. Razavi-Ghods, I. L. V. Roque, A. Saxena, K. H. Scheutwinkel, E. Shen, P. H. Sims, O. Smirnov, M. Spinelli, and K. Zarb-Adami. Radio antenna design for sky-averaged 21 cm cosmology experiments: the REACH case. *arXiv e-prints*, page arXiv:2109.10098, September 2021.
- [117] C. Feng and G. Holder. Enhanced Global Signal of Neutral Hydrogen Due to Excess Radiation at Cosmic Dawn. *ApJ Lett.*, 858(2):L17, May 2018.
- [118] R. Jana, B. B. Nath, and P. L. Biermann. Radio background and IGM heating due to Pop III supernova explosions. *MNRAS*, 483(4):5329–5333, 2019.
- [119] A. Ewall-Wice, T. C. Chang, J. Lazio, O. Doré, M. Seiffert, and R. A. Monsalve. Modeling the Radio Background from the First Black Holes at Cosmic Dawn: Implications for the 21 cm Absorption Amplitude. *ApJ*, 868(1):63, November 2018.

- [120] J. Mirocha and S. R. Furlanetto. What does the first highly redshifted 21-cm detection tell us about early galaxies? *MNRAS*, 483(2):1980–1992, February 2019.
- [121] Itamar Reis, Anastasia Fialkov, and Rennan Barkana. High-redshift radio galaxies: a potential new source of 21-cm fluctuations. *MNRAS*, 499(4):5993–6008, December 2020.
- [122] J. B. Muñoz and A. Loeb. A small amount of mini-charged dark matter could cool the baryons in the early universe. *Nature*, 557:684–686, 2018.
- [123] E. D. Kovetz, V. Poulin, V. Gluscevic, K. K. Boddy, R. Barkana, and M. Kamionkowski. Tighter limits on dark matter explanations of the anomalous EDGES 21 cm signal. *Phys. Rev. D*, 98(10):103529, November 2018.
- [124] Rennan Barkana. Possible interaction between baryons and dark-matter particles revealed by the first stars. *Nature*, 555(7694):71–74, March 2018.
- [125] T. R. Slatyer and C. Wu. Early-Universe constraints on dark matter-baryon scattering and their implications for a global 21 cm signal. *Phys. Rev. D*, 98(2):023013, July 2018.
- [126] A. Berlin, D. Hooper, G. Krnjaic, and S. D. McDermott. Severely Constraining Dark-Matter Interpretations of the 21-cm Anomaly. *prl*, 121(1):011102, July 2018.
- [127] R. Barkana, N. J. Outmezguine, D. Redigol, and T. Volansky. Strong constraints on light dark matter interpretation of the EDGES signal. *Phys. Rev. D*, 98:103005, November 2018.
- [128] R. Hills, G. Kulkarni, P. D. Meerburg, and E. Puchwein. Concerns about modelling of the EDGES data. *Nature*, 564(7736):E32–E34, 2018.
- [129] S. Singh and R. Subrahmanyan. The Redshifted 21 cm Signal in the EDGES Low-band Spectrum. *ApJ*, 880(1):26, 2019.
- [130] Peter H. Sims and Jonathan C. Pober. Testing for calibration systematics in the EDGES low-band data using Bayesian model selection. *MNRAS*, 492(1):22–38, 2020.
- [131] Anastasia Fialkov and Rennan Barkana. Signature of excess radio background in the 21-cm global signal and power spectrum. *MNRAS*, 486(2):1763–1773, June 2019.
- [132] G. Bernardi, J. T. L. Zwart, D. Price, L. J. Greenhill, A. Mesinger, J. Dowell, T. Eftekhari, S. W. Ellingson, J. Kocz, and F. Schinzel. Bayesian constraints on the global 21-cm signal from the Cosmic Dawn. *MNRAS*, 461(3):2847–2855, September 2016.
- [133] Raul A. Monsalve, Alan E. E. Rogers, Judd D. Bowman, and Thomas J. Mozdzen. Results from EDGES High-band. I. Constraints on Phenomenological Models for the Global 21 cm Signal. *ApJ*, 847(1):64, September 2017.
- [134] Raul A. Monsalve, Bradley Greig, Judd D. Bowman, Andrei Mesinger, Alan E. E. Rogers, Thomas J. Mozdzen, Nicholas S. Kern, and Nivedita Mahesh. Results from EDGES High-band. II. Constraints on Parameters of Early Galaxies. *ApJ*, 863(1):11, August 2018.
- [135] Raghunath Ghara, T. Roy Choudhury, and Kanan K. Datta. 21 cm signal from cosmic dawn: imprints of spin temperature fluctuations and peculiar velocities. *MNRAS*, 447(2):1806–1825, 01 2015.
- [136] Raghunath Ghara, Garrelt Mellema, Sambit K Giri, T Roy Choudhury, Kanan K Datta, and Suman Majumdar. Prediction of the 21-cm signal from reionization: comparison between 3D and 1D radiative transfer schemes. *MNRAS*, 476(2):1741–1755, 02 2018.
- [137] J. R. Pritchard and A. Loeb. Constraining the unexplored period between the dark ages and reionization with observations of the global 21 cm signal. *Phys. Rev. D*, 82(2):023006, July 2010.

- [138] Geraint J. A. Harker, J. R. Pritchard, J. O. Burns, and J. D. Bowman. An MCMC approach to extracting the global 21-cm signal during the cosmic dawn from sky-averaged radio observations. *MNRAS*, 419(2):1070–1084, January 2012.
- [139] G. Bernardi, M. McQuinn, and L. J. Greenhill. Foreground Model and Antenna Calibration Errors in the Measurement of the Sky-averaged  $\lambda$ 21 cm Signal at  $z \sim 20$ . *ApJ*, 799(1):90, January 2015.
- [140] R. H. Becker, Xiaohui Fan, Richard L. White, Michael A. Strauss, Vijay K. Narayanan, Robert H. Lupton, James E. Gunn, James Annis, Neta A. Bahcall, J. Brinkmann, A. J. Connolly, István Csabai, Paul C. Czarapata, Mamoru Doi, Timothy M. Heckman, G. S. Hennessy, Željko Ivezić, G. R. Knapp, Don Q. Lamb, Timothy A. McKay, Jeffrey A. Munn, Thomas Nash, Robert Nichol, Jeffrey R. Pier, Gordon T. Richards, Donald P. Schneider, Chris Stoughton, Alexander S. Szalay, Aniruddha R. Thakar, and D. G. York. Evidence for Reionization at  $z \sim 6$ : Detection of a Gunn-Peterson Trough in a  $z=6.28$  Quasar. *AJ*, 122(6):2850–2857, December 2001.
- [141] D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. *ApJS*, 170(2):377–408, June 2007.
- [142] Tom Abel, Greg L. Bryan, and Michael L. Norman. The Formation of the First Star in the Universe. *Science*, 295(5552):93–98, January 2002.
- [143] D. J. Wales and J. P. K. Doye. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116, 1997.
- [144] J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 01 1965.
- [145] J. Nocedal and S. J. Wright. *Numerical Optimisation*. Springer, New York, NY, 2006.
- [146] M. S. Anderson, J. Dahl, and L. Vandenberghe. Cvxopt: A python package for convex optimisation, version 1.2. available at <https://github.com/cvxopt/cvxopt>, 2018.
- [147] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944.
- [148] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.
- [149] Jordan Mirocha. Decoding the X-ray properties of pre-reionization era sources. *MNRAS*, 443(2):1211–1223, 07 2014.
- [150] L. J. Greenhill and G. Bernardi. HI Epoch of Reionization Arrays. *arXiv e-prints*, page arXiv:1201.1700, January 2012.
- [151] Andrei Mesinger, Steven Furlanetto, and Renyue Cen. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. *MNRAS*, 411(2):955–972, February 2011.
- [152] Kamran Javid, Will Handley, Mike Hobson, and Anthony Lasenby. Compromise-free Bayesian neural networks. *arXiv e-prints*, page arXiv:2004.12211, April 2020.
- [153] Adrian Liu and J. Richard Shaw. Data Analysis for Precision 21 cm Cosmology. *PASP*, 132(1012):062001, June 2020.

- [154] Florian List and Geraint F. Lewis. A unified framework for 21 cm tomography sample generation and parameter inference with progressively growing GANs. *MNRAS*, 493(4):5913–5927, April 2020.
- [155] Atrideb Chatterjee, Tirthankar Roy Choudhury, and Sourav Mitra. CosmoReionMC: a package for estimating cosmological and astrophysical parameters using CMB, Lyman- $\alpha$  absorption, and global 21 cm data. *MNRAS*, 507(2):2405–2422, October 2021.
- [156] C. J. Schmit and J. R. Pritchard. Emulation of reionization simulations for Bayesian inference of astrophysics parameters using neural networks. *MNRAS*, 475(1):1213–1223, March 2018.
- [157] W D Jennings, C A Watkinson, F B Abdalla, and J D McEwen. Evaluating machine learning techniques for predicting power spectra from reionization simulations. *MNRAS*, 483(3):2907–2922, 11 2018.
- [158] K Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.
- [159] N. Srivastava, G. Hinton, A Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.
- [160] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR.
- [161] A. Cohen, A. Fialkov, R. Barkana, and R. A. Monsalve. 21cmgem training and testing data sets, February 2021. Uploaded by Harry Bevins with permission from Anastasia Fialkov.
- [162] Volker Bromm, Rolf P. Kudritzki, and Abraham Loeb. Generic Spectrum and Ionization Efficiency of a Heavy Initial Mass Function for the First Stars. *ApJ*, 552(2):464–472, May 2001.
- [163] Ian D. McGreer, Andrei Mesinger, and Valentina D’Odorico. Model-independent evidence in favour of an end to reionization by  $z \approx 6$ . *MNRAS*, 447(1):499–505, 12 2014.
- [164] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv e-prints*, page arXiv:1609.04747, September 2016.
- [165] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [166] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen, A. Amara, J. Annis, J. Asorey, S. Avila, D. Bacon, E. Balbinot, M. Banerji, N. Banik, W. Barkhouse, M. Baumer, E. Baxter, K. Bechtol, M. R. Becker, A. Benoit-Lévy, B. A. Benson, G. M. Bernstein, E. Bertin, J. Blazek, S. L. Bridle, D. Brooks, D. Brout, E. Buckley-Geer, D. L. Burke, M. T. Busha, A. Campos, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, N. Chen, M. Childress, A. Choi, C. Conselice, R. Crittenden, M. Crocce, C. E. Cunha, C. B. D’Andrea, L. N. da Costa, R. Das, T. M. Davis, C. Davis, J. De Vicente, D. L. DePoy, J. DeRose, S. Desai, H. T. Diehl, J. P. Dietrich, S. Dodelson, P. Doel, A. Drlica-Wagner, T. F. Eifler, A. E. Elliott, F. Elsner, J. Elvin-Poole, J. Estrada, A. E. Evrard, Y. Fang, E. Fernandez, A. Ferté, D. A. Finley, B. Flaugher, P. Fosalba, O. Friedrich, J. Frieman, J. García-Bellido, M. Garcia-Fernandez, M. Gatti, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, M. S. S. Gill, K. Glazebrook, D. A. Goldstein, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. Hamilton, W. G. Hartley, S. R. Hinton, K. Honscheid, B. Hoyle, D. Huterer, B. Jain, D. J. James, M. Jarvis, T. Jeltema, M. D. Johnson, M. W. G. Johnson, T. Kacprzak, S. Kent,

- A. G. Kim, A. King, D. Kirk, N. Kokron, A. Kovacs, E. Krause, C. Krawiec, A. Kremin, K. Kuehn, S. Kuhlmann, N. Kuropatkin, F. Lacasa, O. Lahav, T. S. Li, A. R. Liddle, C. Lidman, M. Lima, H. Lin, N. MacCrann, M. A. G. Maia, M. Makler, M. Manera, M. March, J. L. Marshall, P. Martini, R. G. McMahon, P. Melchior, F. Menanteau, R. Miquel, V. Miranda, D. Mudd, J. Muir, A. Möller, E. Neilsen, R. C. Nichol, B. Nord, P. Nugent, R. L. C. Ogando, A. Palmese, J. Peacock, H. V. Peiris, J. Peoples, W. J. Percival, D. Petracca, A. A. Plazas, A. Porredon, J. Prat, A. Pujol, M. M. Rau, A. Refregier, P. M. Ricker, N. Roe, R. P. Rollins, A. K. Romer, A. Roodman, R. Rosenfeld, A. J. Ross, E. Rozo, E. S. Rykoff, M. Sako, A. I. Salvador, S. Samuroff, C. Sánchez, E. Sanchez, B. Santiago, V. Scarpine, R. Schindler, D. Scolnic, L. F. Secco, S. Serrano, I. Sevilla-Noarbe, E. Sheldon, R. C. Smith, M. Smith, J. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, G. Tarle, D. Thomas, M. A. Troxel, D. L. Tucker, B. E. Tucker, S. A. Uddin, T. N. Varga, P. Vielzeuf, V. Vikram, A. K. Vivas, A. R. Walker, M. Wang, R. H. Wechsler, J. Weller, W. Wester, R. C. Wolf, B. Yanny, F. Yuan, A. Zenteno, B. Zhang, Y. Zhang, J. Zuntz, and Dark Energy Survey Collaboration. Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 98(4):043526, August 2018.
- [167] O. Friedrich, F. Andrade-Oliveira, H. Camacho, O. Alves, R. Rosenfeld, J. Sanchez, X. Fang, T. F. Eifler, E. Krause, C. Chang, Y. Omori, A. Amon, E. Baxter, J. Elvin-Poole, D. Huterer, A. Porredon, J. Prat, V. Terra, A. Troja, A. Alarcon, K. Bechtol, G. M. Bernstein, R. Buchs, A. Campos, A. Carnero Rosell, M. Carrasco Kind, R. Cawthon, A. Choi, J. Cordero, M. Crocce, C. Davis, J. DeRose, H. T. Diehl, S. Dodelson, C. Doux, A. Drlica-Wagner, F. Elsner, S. Everett, P. Fosalba, M. Gatti, G. Giannini, D. Gruen, R. A. Gruendl, I. Harrison, W. G. Hartley, B. Jain, M. Jarvis, N. MacCrann, J. McCullough, J. Muir, J. Myles, S. Pandey, M. Raveri, A. Roodman, M. Rodriguez-Monroy, E. S. Rykoff, S. Samuroff, C. Sánchez, L. F. Secco, I. Sevilla-Noarbe, E. Sheldon, M. A. Troxel, N. Weaverdyck, B. Yanny, M. Aguena, S. Avila, D. Bacon, E. Bertin, S. Bhargava, D. Brooks, D. L. Burke, J. Carretero, M. Costanzi, L. N. da Costa, M. E. S. Pereira, J. De Vicente, S. Desai, A. E. Evrard, I. Ferrero, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, J. Gschwend, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, D. J. James, K. Kuehn, O. Lahav, M. Lima, M. A. G. Maia, F. Menanteau, R. Miquel, R. Morgan, A. Palmese, F. Paz-Chinchón, A. A. Plazas, E. Sanchez, V. Scarpine, S. Serrano, M. Soares-Santos, M. Smith, E. Suchyta, G. Tarle, D. Thomas, C. To, T. N. Varga, J. Weller, R. D. Wilkinson, R. D. Wilkinson, and DES Collaboration. Dark Energy Survey year 3 results: covariance modelling and its impact on parameter estimation and quality of fit. *MNRAS*, 508(3):3125–3165, December 2021.
- [168] T. M. C. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis, S. Avila, D. Bacon, E. Baxter, K. Bechtol, M. R. Becker, G. M. Bernstein, S. Bhargava, S. Birrer, J. Blazek, A. Brandao-Souza, S. L. Bridle, D. Brooks, E. Buckley-Geer, D. L. Burke, H. Camacho, A. Campos, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, A. Chen, R. Chen, A. Choi, C. Conselice, J. Cordero, M. Costanzi, M. Crocce, L. N. da Costa, M. E. da Silva Pereira, C. Davis, T. M. Davis, J. De Vicente, J. DeRose, S. Desai, E. Di Valentino, H. T. Diehl, J. P. Dietrich, S. Dodelson, P. Doel, C. Doux, A. Drlica-Wagner, K. Eckert, T. F. Eifler, F. Elsner, J. Elvin-Poole, S. Everett, A. E. Evrard, X. Fang, A. Farahi, E. Fernandez, I. Ferrero, A. Ferté, P. Fosalba, O. Friedrich, J. Frieman, J. García-Bellido, M. Gatti, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, G. Giannini, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, I. Harrison, W. G. Hartley, K. Herner, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, E. M. Huff, D. Huterer, B. Jain, D. J. James,

- M. Jarvis, N. Jeffrey, T. Jeltema, A. Kovacs, E. Krause, R. Kron, K. Kuehn, N. Kuropatkin, O. Lahav, P. F. Leget, P. Lemos, A. R. Liddle, C. Lidman, M. Lima, H. Lin, N. MacCrann, M. A. G. Maia, J. L. Marshall, P. Martini, J. McCullough, P. Melchior, J. Mena-Fernández, F. Menanteau, R. Miquel, J. J. Mohr, R. Morgan, J. Muir, J. Myles, S. Nadathur, A. Navarro-Alsina, R. C. Nichol, R. L. C. Ogando, Y. Omori, A. Palmese, S. Pandey, Y. Park, F. Paz-Chinchón, D. Petracick, A. Pieres, A. A. Plazas Malagón, A. Porredon, J. Prat, M. Raveri, M. Rodriguez-Monroy, R. P. Rollins, A. K. Romer, A. Roodman, R. Rosenfeld, A. J. Ross, E. S. Rykoff, S. Samuroff, C. Sánchez, E. Sanchez, J. Sanchez, D. Sanchez Cid, V. Scarpine, M. Schubnell, D. Scolnic, L. F. Secco, S. Serrano, I. Sevilla-Noarbe, E. Sheldon, T. Shin, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, M. Tabbutt, G. Tarle, D. Thomas, C. To, A. Troja, M. A. Troxel, D. L. Tucker, I. Tutzus, T. N. Varga, A. R. Walker, N. Weaverdyck, R. Wechsler, J. Weller, B. Yanny, B. Yin, Y. Zhang, J. Zuntz, and DES Collaboration. Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 105(2):023520, January 2022.
- [169] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked Autoregressive Flow for Density Estimation. *arXiv e-prints*, page arXiv:1705.07057, May 2017.
- [170] Murray Rosenblatt. Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, 27(3):832–837, September 1956. Publisher: Institute of Mathematical Statistics.
- [171] Emanuel Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, September 1962. Publisher: Institute of Mathematical Statistics.
- [172] Justin Alsing and Will Handley. Nested sampling with any prior you like. *MNRAS*, 505(1):L95–L99, July 2021.
- [173] Will Handley and Pablo Lemos. Quantifying dimensionality: Bayesian cosmological model complexities. *Phys. Rev. D*, 100(2):023512, July 2019.
- [174] Will Handley and Pablo Lemos. Quantifying tensions in cosmological parameters: Interpreting the DES evidence ratio. *Phys. Rev. D*, 100(4):043504, August 2019.
- [175] Aaron Glanville, Cullan Howlett, and Tamara M. Davis. Full-Shape Galaxy Power Spectra and the Curvature Tension. *arXiv e-prints*, page arXiv:2205.05892, May 2022.
- [176] Bernard W Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.
- [177] Aleksandr Petrosyan and William James Handley. SuperNest: accelerated nested sampling applied to astrophysics and cosmology. *arXiv e-prints*, page arXiv:2212.01760, December 2022.
- [178] Will Handley. anesthetic: nested sampling visualisation. *Journal of Open Source Software*, 4(37):1414, 2019.
- [179] Diana Harrison, David Sutton, Pedro Carvalho, and Michael Hobson. Validation of Bayesian posterior distributions using a multidimensional Kolmogorov-Smirnov test. *MNRAS*, 451(3):2610–2624, August 2015.
- [180] Peter Ade, James Aguirre, Zeeshan Ahmed, Simone Aiola, Aamir Ali, David Alonso, Marcelo A. Alvarez, Kam Arnold, Peter Ashton, Jason Austermann, Humna Awan, Carlo Baccigalupi, Taylor Baildon, Darcy Barron, Nick Battaglia, Richard Battye, Eric Baxter, Andrew Bazarko, James A. Beall, Rachel Bean, Dominic Beck, Shawn Beckman, Benjamin Beringue, Federico Bianchini, Steven Boada, David Boettger, J. Richard Bond, Julian Borrill, Michael L. Brown, Sarah Marie Bruno, Sean Bryan, Erminia Calabrese, Victoria Calafut, Paolo Calisse, Julien Carron, Anthony Challinor, Grace Chesmore, Yuji Chinone, Jens Chluba, Hsiao-Mei Sherry Cho, Steve Choi, Gabriele Coppi, Nicholas F. Cothard, Kevin

Coughlin, Devin Crichton, Kevin D. Crowley, Kevin T. Crowley, Ari Cukierman, John M. D'Ewart, Rolando Dünnér, Tijmen de Haan, Mark Devlin, Simon Dicker, Joy Didier, Matt Dobbs, Bradley Dober, Cody J. Duell, Shannon Duff, Adri Duivenvoorden, Jo Dunkley, John Dusatko, Josquin Errard, Giulio Fabbian, Stephen Feeney, Simone Ferraro, Pedro Fluxà, Katherine Freese, Josef C. Frisch, Andrei Frolov, George Fuller, Brittany Fuzia, Nicholas Galitzki, Patricio A. Gallardo, Jose Tomas Galvez Ghersi, Jiansong Gao, Eric Gawiser, Martina Gerbino, Vera Gluscevic, Neil Goeckner-Wald, Joseph Golec, Sam Gordon, Megan Gralla, Daniel Green, Arpi Grigorian, John Groh, Chris Groppi, Yilun Guan, Jon E. Gudmundsson, Dongwon Han, Peter Hargrave, Masaya Hasegawa, Matthew Hasselfield, Makoto Hattori, Victor Haynes, Masashi Hazumi, Yizhou He, Erin Healy, Shawn W. Henderson, Carlos Hervias-Caimapo, Charles A. Hill, J. Colin Hill, Gene Hilton, Matt Hilton, Adam D. Hincks, Gary Hinshaw, Renée Hložek, Shirley Ho, Shuay-Pwu Patty Ho, Logan Howe, Zhiqi Huang, Johannes Hubmayr, Kevin Huffenberger, John P. Hughes, Anna Iijas, Margaret Ikape, Kent Irwin, Andrew H. Jaffe, Bhuvnesh Jain, Oliver Jeong, Daisuke Kaneko, Ethan D. Karpel, Nobuhiko Katayama, Brian Keating, Sarah S. Kernasovskiy, Reijo Keskitalo, Theodore Kisner, Kenji Kiuchi, Jeff Klein, Kenda Knowles, Brian Koopman, Arthur Kosowsky, Nicoletta Krachmalnicoff, Stephen E. Kuenstner, Chao-Lin Kuo, Akito Kusaka, Jacob Lashner, Adrian Lee, Eunseong Lee, David Leon, Jason S. Y. Leung, Antony Lewis, Yaqiong Li, Zack Li, Michele Limon, Eric Linder, Carlos Lopez-Caraballo, Thibaut Louis, Lindsay Lowry, Marius Lungu, Mathew Madhavacheril, Daisy Mak, Felipe Maldonado, Hamdi Mani, Ben Mates, Frederick Matsuda, Loïc Maurin, Phil Mauskopf, Andrew May, Nialh McCallum, Chris McKenney, Jeff McMahon, P. Daniel Meerburg, Joel Meyers, Amber Miller, Mark Mirmelstein, Kavilan Moodley, Moritz Munchmeyer, Charles Munson, Sigurd Naess, Federico Nati, Martin Navaroli, Laura Newburgh, Ho Nam Nguyen, Michael Niemack, Haruki Nishino, John Orlowski-Scherer, Lyman Page, Bruce Partridge, Julien Peloton, Francesca Perrotta, Lucio Piccirillo, Giampaolo Pisano, Davide Poletti, Roberto Puddu, Giuseppe Puglisi, Chris Raum, Christian L. Reichardt, Mathieu Remazeilles, Yoel Rephaeli, Dominik Riechers, Felipe Rojas, Anirban Roy, Sharon Sadeh, Yuki Sakurai, Maria Salatino, Mayuri Sathyaranayana Rao, Emmanuel Schaan, Marcel Schmittfull, Neelima Sehgal, Joseph Seibert, Uros Seljak, Blake Sherwin, Meir Shimon, Carlos Sierra, Jonathan Sievers, Precious Sikhosana, Maximiliano Silva-Feaver, Sara M. Simon, Adrian Sinclair, Praveen Siritanasak, Kendrick Smith, Stephen R. Smith, David Spergel, Suzanne T. Staggs, George Stein, Jason R. Stevens, Radek Stompor, Aritoki Suzuki, Osamu Tajima, Satoru Takakura, Grant Teply, Daniel B. Thomas, Ben Thorne, Robert Thornton, Hy Trac, Calvin Tsai, Carole Tucker, Joel Ullom, Sunny Vagnozzi, Alexander van Engelen, Jeff Van Lanen, Daniel D. Van Winkle, Eve M. Vavagiakis, Clara Vergès, Michael Vissers, Kasey Wagoner, Samantha Walker, Jon Ward, Ben Westbrook, Nathan Whitehorn, Jason Williams, Joel Williams, Edward J. Wollack, Zhilei Xu, Byeonghee Yu, Cyndia Yu, Fernando Zago, Hezi Zhang, Ningfeng Zhu, and Simons Observatory Collaboration. The Simons Observatory: science goals and forecasts. *J. Cosmology Astropart. Phys.*, 2019(2):056, February 2019.

- [181] Adrian Lee, Maximilian H. Abitbol, Shunsuke Adachi, Peter Ade, James Aguirre, Zeeshan Ahmed, Simone Aiola, Aamir Ali, David Alonso, Marcelo A. Alvarez, Kam Arnold, Peter Ashton, Zachary Atkins, Jason Austermann, Humna Awan, Carlo Baccigalupi, Taylor Baildon, Anton Baleato Lizancos, Darcy Barron, Nick Battaglia, Richard Battye, Eric Baxter, Andrew Bazarko, James A. Beall, Rachel Bean, Dominic Beck, Shawn Beckman, Benjamin Beringue, Tanay Bhandarkar, Sanah Bhimani, Fed-

erico Bianchini, Steven Boada, David Boettger, Boris Bolliet, J. Richard Bond, Julian Borrill, Michael L. Brown, Sarah Marie Bruno, Sean Bryan, Erminia Calabrese, Victoria Calafut, Paolo Calisse, Julien Carron, Fred. M. Carl, Juan Cayuso, Anthony Challinor, Grace Chesmore, Yuji Chinone, Jens Chluba, Hsiao-Mei Sherry Cho, Steve Choi, Susan Clark, Philip Clarke, Carlo Contaldi, Gabriele Coppi, Nicholas F. Cothard, Kevin Coughlin, Will Coulton, Devin Crichton, Kevin D. Crowley, Kevin T. Crowley, Ari Cukierman, John M. D'Ewart, Rolando Dünner, Tijmen de Haan, Mark Devlin, Simon Dicker, Bradley Dober, Cody J. Duell, Shannon Duff, Adri Duivenvoorden, Jo Dunkley, Hamza El Bouhargani, Josquin Errard, Giulio Fabbian, Stephen Feeney, James Ferguson, Simone Ferraro, Pedro Fluxa, Katherine Freese, Josef C. Frisch, Andrei Frolov, George Fuller, Nicholas Galitzki, Patricio A. Gallardo, Jose Tomas Galvez Ghersi, Jiansong Gao, Eric Gawiser, Martina Gerbino, Vera Gluscevic, Neil Goeckner-Wald, Joseph Golec, Sam Gordon, Megan Gralla, Daniel Green, Arpi Grigorian, John Groh, Chris Groppi, Yilun Guan, Jon E. Gudmundsson, Mark Halpern, Dongwon Han, Peter Hargrave, Kathleen Harrington, Masaya Hasegawa, Matthew Hasselfield, Makoto Hattori, Victor Haynes, Masashi Hazumi, Erin Healy, Shawn W. Henderson, Brandon Hensley, Carlos Hervias-Caimapo, Charles A. Hill, J. Colin Hill, Gene Hilton, Matt Hilton, Adam D. Hincks, Gary Hinshaw, Renee Hložek, Shirley Ho, Shuay-Pwu Patty Ho, Thuong D. Hoang, Jonathan Hoh, Selim C. Hotinli, Zhiqi Huang, Johannes Hubmayr, Kevin Huffenberger, John P. Hughes, Anna Iijas, Margaret Ikape, Kent Irwin, Andrew H. Jaffe, Bhuvnesh Jain, Oliver Jeong, Matthew Johnson, Daisuke Kaneko, Ethan D. Karpel, Nobuhiko Katayama, Brian Keating, Reijo Keskitalo, Theodore Kisner, Kenji Kiuchi, Jeff Klein, Kenda Knowles, Anna Kofman, Brian Koopman, Arthur Kosowsky, Nicoletta Krachmalnicoff, Akito Kusaka, Paul La Plante, Jacob Lashner, Adrian Lee, Eunseong Lee, Antony Lewis, Yaqiong Li, Zack Li, Michele Limon, Eric Linder, Jia Liu, Carlos Lopez-Carballo, Thibaut Louis, Marius Lungu, Mathew Madhavacheril, Daisy Mak, Felipe Maldonado, Hamdi Mani, Ben Mates, Frederick Matsuda, Loïc Maurin, Phil Mauskopf, Andrew May, Nialh McCallum, Heather McCarrick, Chris McKenney, Jeff McMahon, P. Daniel Meerburg, James Mertens, Joel Meyers, Amber Miller, Mark Mirmelstein, Kavilan Moodley, Jenna Moore, Moritz Munchmeyer, Charles Munson, Masaaki Murata, Sigurd Naess, Toshiya Namikawa, Federico Nati, Martin Navaroli, Laura Newburgh, Ho Nam Nguyen, Andrina Nicola, Mike Niemack, Haruki Nishino, Yume Nishinomiya, John Orlowski-Scherer, Luca Pagano, Bruce Partridge, Francesca Perrotta, Phumlani Phakathi, Lucio Piccirillo, Elena Pierpaoli, Giampaolo Pisano, Davide Poletti, Roberto Puddu, Giuseppe Puglisi, Chris Raum, Christian L. Reichardt, Mathieu Remazeilles, Yoel Rephaeli, Dominik Riechers, Felipe Rojas, Aditya Rotti, Anirban Roy, Sharon Sadeh, Yuki Sakurai, Maria Salatino, Mayuri Sathyaranayana Rao, Lauren Saunders, Emmanuel Schaan, Marcel Schmittfull, Neelima Sehgal, Joseph Seibert, Uros Seljak, Paul Shellard, Blake Sherwin, Meir Shimon, Carlos Sierra, Jonathan Sievers, Cristobal Sifon, Precious Sikhosana, Maximiliano Silva-Feaver, Sara M. Simon, Adrian Sinclair, Kendrick Smith, Wuhyun Sohn, Rita Sonka, David Spergel, Jacob Spisak, Suzanne T. Staggs, George Stein, Jason R. Stevens, Radek Stompor, Aritoki Suzuki, Osamu Tajima, Satoru Takakura, Grant Teply, Daniel B. Thomas, Ben Thorne, Robert Thornton, Hy Trac, Jesse Treu, Calvin Tsai, Carole Tucker, Joel Ullom, Sunny Vagnozzi, Alexander van Engelen, Jeff Van Lanen, Daniel D. Van Winkle, Eve M. Vavagiakis, Clara Vergès, Michael Vissers, Kasey Wagoner, Samantha Walker, Yuhan Wang, Jon Ward, Ben Westbrook, Nathan Whitehorn, Jason Williams, Joel Williams, Edward Wollack, Zhilei Xu, Siavash Yasini, Edward Young, Byeonghee Yu, Cyndia Yu, Fernando Zago, Mario Zannoni, Hezi Zhang, Kaiwen Zheng, Ningfeng Zhu, and Andrea

- Zonca. The Simons Observatory. In *Bulletin of the American Astronomical Society*, volume 51, page 147, September 2019.
- [182] Xuelei Chen and Jordi Miralda-Escudé. The Spin-Kinetic Temperature Coupling and the Heating Rate due to Ly $\alpha$  Scattering before Reionization: Predictions for 21 Centimeter Emission and Absorption. *ApJ*, 602(1):1–11, February 2004.
- [183] Steven R. Furlanetto and Jonathan R. Pritchard. The scattering of Lyman-series photons in the intergalactic medium. *MNRAS*, 372(3):1093–1103, November 2006.
- [184] Semelin, B., Combes, F., and Baek, S. Lyman-alpha radiative transfer during the epoch of reionization: contribution to 21-cm signal fluctuations. *A&A*, 474(2):365–374, 2007.
- [185] Smadar Naoz and Rennan Barkana. Detecting early galaxies through their 21-cm signature. *MNRAS Lett.*, 385(1):L63–L67, 03 2008.
- [186] Baek, S., Di Matteo, P., Semelin, B., Combes, F., and Revaz, Y. The simulated 21 cm signal during the epoch of reionization: full modeling of the Ly-ping. *A&A*, 495(2):389–405, 2009.
- [187] Eli Visbal and Matthew McQuinn. The Impact of Neutral Intergalactic Gas on Ly  $\alpha$  Intensity Mapping during Reionization. *ApJ Lett.*, 863(1):L6, August 2018. Publisher: American Astronomical Society.
- [188] Margherita Molaro, Romeel Davé, Sultan Hassan, Mario G Santos, and Kristian Finlator. Artist: fast radiative transfer for large-scale simulations of the epoch of reionization. *MNRAS*, 489(4):5594–5611, November 2019.
- [189] Itamar Reis, Rennan Barkana, and Anastasia Fialkov. Mapping discrete galaxies at cosmic dawn with 21-centimeter observations. *arXiv e-prints*, page arXiv:2008.04914, August 2020.
- [190] Harry T. J. Bevins. maxsmooth: Derivative constrained function fitting. *Journal of Open Source Software*, 5(54):2596, 2020.
- [191] G. Bernardi, A. G. de Bruyn, M. A. Brentjens, B. Ciardi, G. Harker, V. Jelić, L. V. E. Koopmans, P. Labropoulos, A. Offringa, V. N. Pandey, J. Schaye, R. M. Thomas, S. Yatawatta, and S. Zaroubi. Foregrounds for observations of the cosmological 21 cm line. I. First Westerbork measurements of Galactic emission at 150 MHz in a low latitude field. *A&A*, 500(3):965–979, June 2009.
- [192] I. C. Nițu, H. T. J. Bevins, J. D. Bray, and A. M. M. Scaife. An updated estimate of the cosmic radio background and implications for ultra-high-energy photon propagation. *Astroparticle Physics*, 126:102532, March 2021.
- [193] Mayuri Sathyanarayana Rao, Ravi Subrahmanyam, N. Udaya Shankar, and Jens Chluba. GMOSS: ALL-SKY MODEL OF SPECTRAL RADIO BRIGHTNESS BASED ON PHYSICAL COMPONENTS AND ASSOCIATED RADIATIVE PROCESSES. *The Astronomical Journal*, 153(1):26, December 2016. Publisher: American Astronomical Society.
- [194] V. Trainotti and G. Figueroa. Vertically polarized dipoles and monopoles, directivity, effective height and antenna factor. *IEEE Transactions on Broadcasting*, 56(3):379–409, 2010.
- [195] Anastasia Fialkov, Rennan Barkana, and Eli Visbal. The observable signature of late heating of the Universe during cosmic reionization. *Nature*, 506(7487):197–199, February 2014.
- [196] Will Handley. fgivenx: A python package for functional posterior plotting. *Journal of Open Source Software*, 3(28):849, 2018.
- [197] Zara Abdurashidova, James E. Aguirre, Paul Alexander, Zaki S. Ali, Yanga Balfour, Adam P. Beardsley, Gianni Bernardi, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Philip Bull, Jacob Burba,

- Steve Carey, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matt Dexter, Eloy de Lera Acedo, Taylor Dibblee-Barkman, Joshua S. Dillon, John Ely, Aaron Ewall-Wice, Nicolas Fagnoni, Randall Fritz, Steven R. Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, Nicholas S. Kern, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Telalo Lekalake, David Lewis, Adrian Liu, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E. Martinot, Eunice Matsetela, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Steven G. Murray, Abraham R. Neben, Bojan Nikolic, Chuneeta D. Nunhokee, Aaron R. Parsons, Nipanjana Patra, Robert Pascua, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Peter Sims, Saurabh Singh, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K. G. Williams, Haoxuan Zheng, and HERA Collaboration. First Results from HERA Phase I: Upper Limits on the Epoch of Reionization 21 cm Power Spectrum. *ApJ*, 925(2):221, February 2022.
- [198] The Athena Collaboration. The Hot and Energetic Universe: A White Paper presenting the science theme motivating the Athena+ mission. *arXiv e-prints*, page arXiv:1306.2307, June 2013.
- [199] The Lynx Team. The Lynx Mission Concept Study Interim Report. *arXiv e-prints*, page arXiv:1809.09642, September 2018.
- [200] R. Mushotzky. AXIS: a probe class next generation high angular resolution x-ray imaging satellite. In Jan-Willem A. den Herder, Shouleh Nikzad, and Kazuhiro Nakazawa, editors, *Space Telescopes and Instrumentation 2018: Ultraviolet to Gamma Ray*, volume 10699 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 1069929, July 2018.
- [201] Andrei Mesinger. Was reionization complete by  $z \sim 5 - 6$ ? *MNRAS*, 407(2):1328–1337, September 2010.
- [202] Joshua Schroeder, Andrei Mesinger, and Zoltán Haiman. Evidence of Gunn–Peterson damping wings in high- $z$  quasar spectra: strengthening the case for incomplete reionization at  $z \sim 6 - 7$ . *MNRAS*, 428(4):3058–3071, 11 2012.
- [203] Masami Ouchi, Yuichi Harikane, Takatoshi Shibuya, Kazuhiro Shimasaku, Yoshiaki Taniguchi, Akira Konno, Masakazu Kobayashi, Masaru Kajisawa, Tohru Nagao, Yoshiaki Ono, Akio K. Inoue, Masayuki Umemura, Masao Mori, Kenji Hasegawa, Ryo Higuchi, Yutaka Komiyama, Yuichi Matsuda, Kimihiko Nakajima, Tomoki Saito, and Shiang-Yu Wang. Systematic Identification of LAEs for Visible Exploration and Reionization Research Using Subaru HSC (SILVERRUSH). I. Program strategy and clustering properties of  $\sim 2000$  Ly $\alpha$  emitters at  $z = 6-7$  over the  $0.3-0.5$  Gpc $^2$  survey area. *PASJ*, 70:S13, January 2018.
- [204] Alexa M. Morales, Charlotte A. Mason, Sean Bruton, Max Gronke, Francesco Haardt, and Claudia Scarlata. The evolution of the lyman-alpha luminosity function during reionization. *ApJ*, 919(2):120, oct 2021.
- [205] Bradley Greig, Andrei Mesinger, Frederick B. Davies, Feige Wang, Jinyi Yang, and Joseph F. Hennawi. IGM damping wing constraints on reionization from covariance reconstruction of two  $z \gtrsim 7$  QSOs. *MNRAS*, 512(4):5390–5403, June 2022.
- [206] Anastasia Fialkov, Rennan Barkana, Eli Visbal, Dmitriy Tseliakhovich, and Christopher M. Hirata. The 21-cm signature of the first stars during the Lyman-Werner feedback era. *MNRAS*, 432(4):2909–2916, July 2013.

- [207] Jordan Mirocha. ARES: Accelerated Reionization Era Simulations. Astrophysics Source Code Library, record ascl:2011.010, November 2020.
- [208] Anastasia Fialkov, Rennan Barkana, Dmitriy Tseliakhovich, and Christopher M. Hirata. Impact of the relative motion between the dark matter and baryons on the first stars: semi-analytical modelling. *MNRAS*, 424(2):1335–1345, August 2012.
- [209] Ralf Klessen. Formation of the first stars. In Muhammad Latif and Dominik Schleicher, editors, *Formation of the First Black Holes*, pages 67–97. 2019.
- [210] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951. Publisher: Institute of Mathematical Statistics.
- [211] Philippe Zarka, Andrée Coffre, Laurent Denis, Cédric Dumez-Viou, Julien Girard, Jean-Mathias Grießmeier, Alan Loh, and Michel Tagger. The low-frequency radiotelescope nenufar. In *2018 2nd URSI Atlantic Radio Science Meeting (AT-RASC)*, pages 1–1, 2018.
- [212] Anna T. P. Schauer, Simon C. O. Glover, Ralf S. Klessen, and Paul Clark. The influence of streaming velocities and Lyman-Werner radiation on the formation of the first stars. *MNRAS*, 507(2):1775–1787, October 2021.
- [213] Dmitriy Tseliakhovich and Christopher Hirata. Relative velocity of dark matter and baryonic fluids and the formation of the first structures. *Phys. Rev. D*, 82(8):083520, October 2010.
- [214] H. E. Bond. Where is population III ? *ApJ*, 248:606–611, September 1981.
- [215] Volker Bromm and Richard B. Larson. The first stars. *Annual Review of Astronomy and Astrophysics*, 42(1):79–118, 2004.
- [216] Daniel C. Jacobs, Jonathan C. Pober, Aaron R. Parsons, James E. Aguirre, Zaki S. Ali, Judd Bowman, Richard F. Bradley, Chris L. Carilli, David R. DeBoer, Matthew R. Dexter, Nicole E. Gugliucci, Pat Klima, Adrian Liu, David H. E. MacMahon, Jason R. Manley, David F. Moore, Irina I. Stefan, and William P. Walbrugh. Multiredshift Limits on the 21 cm Power Spectrum from PAPER. *ApJ*, 801(1):51, March 2015.
- [217] Matthew Kolopanis, Jonathan Pober, Daniel C. Jacobs, and Samantha McGraw. New EoR Power Spectrum Limits From MWA Phase II Using the Delay Spectrum Method and Novel Systematic Rejection. *arXiv e-prints*, page arXiv:2210.10885, October 2022.
- [218] A. H. Patil, S. Yatawatta, L. V. E. Koopmans, A. G. de Bruyn, M. A. Brentjens, S. Zaroubi, K. M. B. Asad, M. Hatef, V. Jelić, M. Mevius, A. R. Offringa, V. N. Pandey, H. Vedantham, F. B. Abdalla, W. N. Brouw, E. Chapman, B. Ciardi, B. K. Gehlot, A. Ghosh, G. Harker, I. T. Iliev, K. Kakiichi, S. Majumdar, G. Mellema, M. B. Silva, J. Schaye, D. Vrbanec, and S. J. Wijnholds. Upper Limits on the 21 cm Epoch of Reionization Power Spectrum from One Night with LOFAR. *ApJ*, 838(1):65, March 2017.
- [219] F. G. Mertens, M. Mevius, L. V. E. Koopmans, A. R. Offringa, G. Mellema, S. Zaroubi, M. A. Brentjens, H. Gan, B. K. Gehlot, V. N. Pandey, A. M. Sardarabadi, H. K. Vedantham, S. Yatawatta, K. M. B. Asad, B. Ciardi, E. Chapman, S. Gazagnes, R. Ghara, A. Ghosh, S. K. Giri, I. T. Iliev, V. Jelić, R. Kooistra, R. Mondal, J. Schaye, and M. B. Silva. Improved upper limits on the 21 cm signal power spectrum of neutral hydrogen at  $z \approx 9.1$  from LOFAR. *MNRAS*, 493(2):1662–1685, April 2020.
- [220] B K Gehlot, F G Mertens, L V E Koopmans, A R Offringa, A Shulevski, M Mevius, M A Brentjens, M Kuiack, V N Pandey, A Rowlinson, A M Sardarabadi, H K Vedantham, R A M J Wijers, S Yatawatta, and S Zaroubi. The AARTFAAC Cosmic Explorer: observations of the 21-cm power spectrum in the EDGES absorption trough. *MNRAS*, 499(3):4158–4173, 10 2020.

- [221] Shikhar Mittal and Girish Kulkarni. Implications of the cosmological 21-cm absorption profile for high-redshift star formation and deep JWST surveys. *MNRAS*, July 2022.
- [222] F. G. Mertens, B. Semelin, and L. V. E. Koopmans. Exploring the Cosmic Dawn with NenuFAR. In A. Siebert, K. Baillié, E. Lagadec, N. Lagarde, J. Malzac, J. B. Marquette, M. N'Diaye, J. Richard, and O. Venot, editors, *SF2A-2021: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, pages 211–214, December 2021.
- [223] Aviad Cohen, Anastasia Fialkov, and Rennan Barkana. Charting the parameter space of the 21-cm power spectrum. *MNRAS*, 478(2):2193–2217, August 2018.
- [224] B. D. Lehmer, Y. Q. Xue, W. N. Brandt, D. M. Alexander, F. E. Bauer, M. Brusa, A. Comastri, R. Gilli, A. E. Hornschemeier, B. Luo, M. Paolillo, A. Ptak, O. Shemmer, D. P. Schneider, P. Tozzi, and C. Vignali. The 4 Ms Chandra Deep Field-South Number Counts Apportioned by Source Class: Pervasive Active Galactic Nuclei and the Ascent of Normal Galaxies. *ApJ*, 752(1):46, June 2012.
- [225] Keith Tauscher, David Rapetti, Jack O. Burns, and Eric Switzer. Global 21 cm Signal Extraction from Foreground and Instrumental Effects. I. Pattern Recognition Framework for Separation Using Training Sets. *ApJ*, 853(2):187, February 2018.
- [226] Anchal Saxena, P. Daniel Meerburg, Eloy de Lera Acedo, Will Handley, and Léon V. E. Koopmans. Sky-averaged 21-cm signal extraction using multiple antennas with an SVD framework: the REACH case. *arXiv e-prints*, page arXiv:2212.07415, December 2022.
- [227] Peter H. Sims, Judd D. Bowman, Nivedita Mahesh, Steven G. Murray, John P. Barrett, Rigel Cappallo, Raul A. Monsalve, Alan E. E. Rogers, Titu Samson, and Akshatha K. Vydula. A Bayesian approach to modelling spectrometer data chromaticity corrected using beam factors – I. Mathematical formalism. *arXiv e-prints*, page arXiv:2212.03875, December 2022.
- [228] Ethan Payne, Colm Talbot, and Eric Thrane. Higher order gravitational-wave modes with likelihood reweighting. *Phys. Rev. D*, 100:123017, Dec 2019.
- [229] Emma Curtis-Lake, Stefano Carniani, Alex Cameron, Stephane Charlot, Peter Jakobsen, Roberto Maiolino, Andrew Bunker, Joris Witstok, Renske Smit, Jacopo Chevallard, Chris Willott, Pierre Ferruit, Santiago Arribas, Nina Bonaventura, Mirko Curti, Francesco D'Eugenio, Marijn Franx, Giovanna Giardino, Tobias J. Looser, Nora Lützgendorf, Michael V. Maseda, Tim Rawle, Hans-Walter Rix, Bruno Rodriguez del Pino, Hannah Übler, Marko Sirianni, Alan Dressler, Eiichi Egami, Daniel J. Eisenstein, Ryan Endsley, Kevin Hainline, Ryan Hausen, Benjamin D. Johnson, Marcia Rieke, Brant Robertson, Irene Shivaei, Daniel P. Stark, Sandro Tacchella, Christina C. Williams, Christopher N. A. Willmer, Rachana Bhatawdekar, Rebecca Bowler, Kristan Boyett, Zuyi Chen, Anna de Graaff, Jakob M. Helton, Raphael E. Hvding, Gareth C. Jones, Nimisha Kumari, Jianwei Lyu, Erica Nelson, Michele Perna, Lester Sandles, Aayush Saxena, Katherine A. Suess, Fengwu Sun, Michael W. Topping, Imaan E. B. Wallace, and Lily Whitler. Spectroscopy of four metal-poor galaxies beyond redshift ten. *arXiv e-prints*, page arXiv:2212.04568, December 2022.