# 21SSD: a new public 21-cm EoR database

## Evan Eames and Benoît Semelin

Sorbonne Universités, UPMC, Observatoire de Paris
PSL research university, CNRS
F-75014, Paris, France
email: `evan.eames@obspm.fr`

**Abstract.** With current efforts inching closer to detecting the 21-cm signal from the Epoch of Reionization (EoR), proper preparation will require publicly available simulated models of the various forms the signal could take. In this work we present a database of such models, available at `21ssd.obspm.fr`. The models are created with a fully-coupled radiative hydrodynamic simulation (LICORICE), and are created at high resolution ($1024^3$). We also begin to analyse and explore the possible 21-cm EoR signals (with Power Spectra and Pixel Distribution Functions), and study the effects of thermal noise on our ability to recover the signal out to high redshifts. Finally, we begin to explore the concepts of 'distance' between different models, which represents a crucial step towards optimising parameter space sampling, training neural networks, and finally extracting parameter values from observations.

**Keywords.** Methods: numerical, radiative transfer, dark ages, reionization

## 1. Introduction

A number of efforts are currently seeking to detect the cosmological 21-cm signal, originating from clouds of neutral hydrogen which lie in the early universe. The signal is expected to be a key probe for imaging the Epoch of Reionization (EoR), thought to begin at $z \sim 27$ and end at $z \sim 6$. In preparation for these efforts, there have been many attempts to simulate the signal. However, the form of the signal depends on a number of astrophysical parameters, some of which are poorly constrained. On account of this, it has been necessary to create databases comprising a variety of possible signals. To date, these databases have been made with semi-numerical simulations, to avoid unrealistic computation requirements. Although these models still provide ample opportunity for pertinent science, they lack much of the robustness needed to fully explore the 21-cm signal. In this vein, we present a high quality database of of simulated signals, with models created at 45 points in a 3D parameter space. These models have been made available at `21ssd.obspm.fr`

The database was created using the LICORICE code, described in Semelin & Combes 2002, Semelin *et al.* 2007, Baek *et al.* 2009, Iliev *et al.* 2009, Baek *et al.* 2010, Vonlanthen *et al.* 2011, and Semelin 2016. Three astrophysical parameters were chosen to create 21SSD: $f_\alpha$ – the Lyman band emissivity, $f_X$ – the X-ray emissivity, and $r_{H/S}$ – the hard-to-soft X-ray ratio. The full definitions of each of these are given in Semelin *et al.* 2017. Three values are chosen for $f_\alpha$ and $r_{H/S}$, while five values are chosen for $f_X$. This gives a 3D parameter space, sparsely sampled with 45 models, and for each model we generated the lightcone in the x, y, or z direction. The final result is 135 lightcones, all of which are available (at high and SKA resolutions: between angular resolutions of $\Delta\theta \sim 0.3'$ and $\sim 3' - 8'$ respectively) on the 21SSD website. As well, noisy lightcones are also provided,
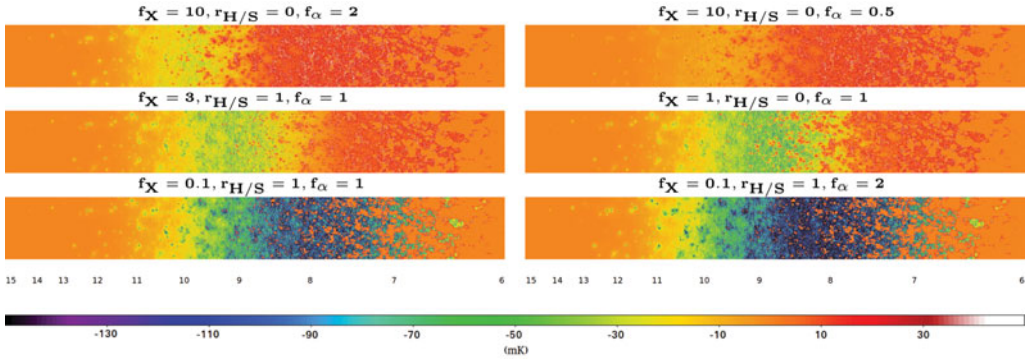
**Figure 1.** An assortment of lightcones taken from the database, with their respective parameter values. These represent slices through $1024 \times 1024 \times 8192$ rectangular lightcones, created between redshifts $z = 6 - 15$, and with angular resolution $\Delta\theta \sim 0.2' - 0.5'$.

in which realistic thermal noise is created via coverage modelling of the current SKA station layout.

## 2. Overview

The database proves useful in a number of ways. Firstly, it should be evident from Figure 1 that the choice of parameter values within our parameter space has a profound effect on the reionization history. For some, the 21-cm signal occurs as a much 'deeper' event than for others (a more negative 21-cm $< T_b >$). The beginning, end, and duration of the signal is also seen to shift to different redshifts between models.

A second immediately evident use of this database is studying effective procedures to discriminate between models, both in physical space, and redshift space. Power spectra can be created at various redshifts, an example of which can be found in Semelin *et al.* 2017. A second diagnostic – perhaps less explored – is the Pixel Distribution Function (PDF). For these, a 1D histogram is created for each redshift slice. These 1D histograms are then placed alongside each other to create a 2D histogram. Examples of these PDF can be found in Figure 2. We can see how, as with the corresponding lightcones, the PDF also capture much of the form of the 21-cm signal.

Thirdly, we can use these two diagnostics to define the concept of distance between two models. We do this using the standard $L_2$ norm:

$$\mathrm{D}_{i,j}^{\mathrm{PS}} = \sqrt{\int \left(\mathrm{PS}_i(k,z) - \mathrm{PS}_j(k,z)\right)^2 dk dz} \qquad (2.1)$$

where $\mathrm{PS}_i$ is the $i^{th}$ Power Spectrum, $k$ is the inverse distance (h·cMpc$^{-1}$), and $z$ is the redshift. This can be thought of as averaging the distance between the two Power Spectra 'surfaces'. In the case of the PDF, the calculation is similarly

$$\mathrm{D}_{i,j}^{\mathrm{PDF}} = \sqrt{\int \left(log_{10}(\mathrm{PDF}_i(T_b,z)) - log_{10}(\mathrm{PDF}_j(T_b,z))\right)^2 dk dz} \qquad (2.2)$$

where $\mathrm{PDF}_i$ is the PDF of the $i^{th}$ model. The only difference in the definition is that we use the logarithm of the PDF. This is not a mandatory convention, however, using the logarithm allows us to give more weight to the 'wings' (vertical features, shown in purple in Figure 2) of the distributions, and more easily analyse the non-Gaussian features. In
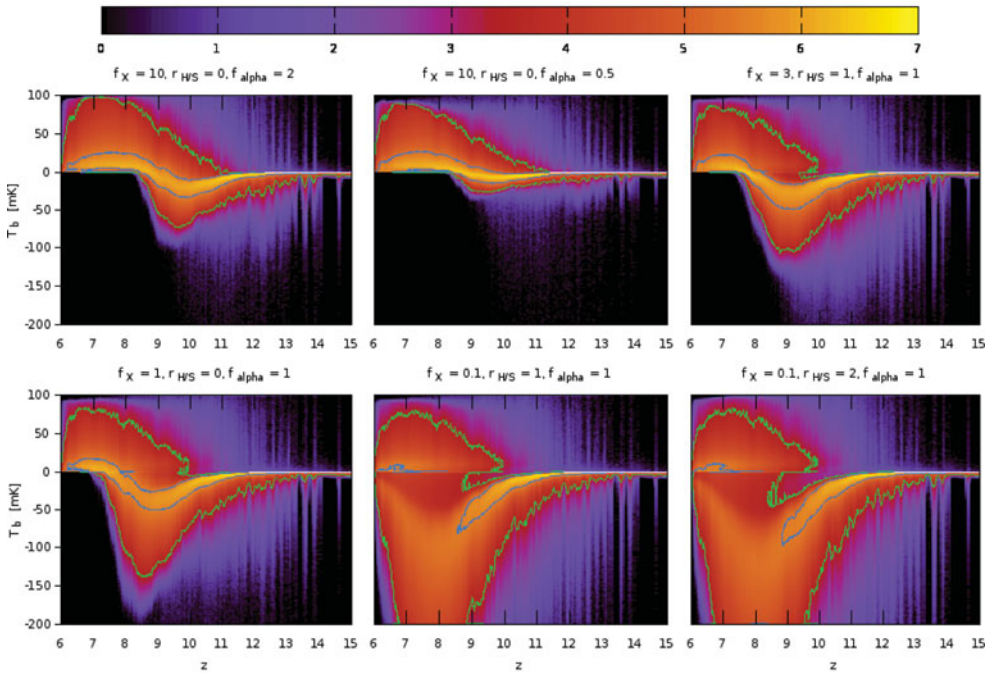
**Figure 2.** The Pixel Distribution Functions for the six lightcones in Figure 1. The colourbar units are logarithmic, and the blue and green contours enclose regions where a pixel chosen at random in the lightcone will lie with $1\sigma$ (0.682) and $3\sigma$ (0.997) confidence, respectively. Vertical spikes in the distribution are caused by sample variance.
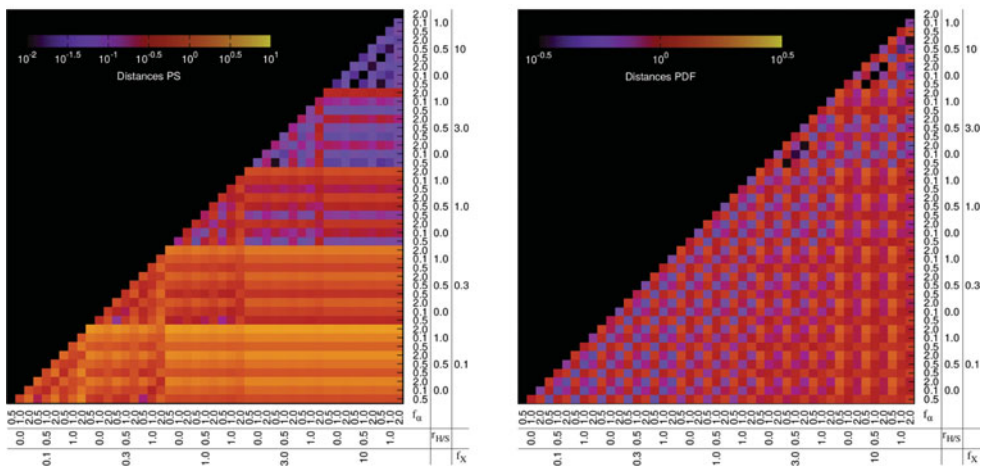


**Figure 3.** Distances between models, in which the distances were calculated using the Power Spectra (eq. 2.1) on the left, and the Pixel Distribution Functions (eq. 2.2) on the right.

order to compare how these two distance diagnostics perform we illustrate the distances between all 45 models in Figure 3.

Notable is that the two methods of calculating distance prove to be strikingly different diagnostics, as is evident by the different forms of the left and right images in Figure 3. For the Power Spectra Distances, it is seen that variations between each $9 \times 9$ 'block' of simulations made with equal $f_X$ are easily apparent. However, within each of these

blocks, the variations are more subtle. This seems to suggest that using the power spectra to calculate distances is an efficient method for reconstructing the values of $f_X$, but less powerful for $r_{H/S}$ and $f_\alpha$, especially for higher $f_X$ values. For the PDF distances, the form is quite different, and we see that there is much less variation between different 9×9 'blocks' of constant $f_X$, nor between 3×3 blocks of constant $r_{H/S}$, yet there is quite a noticeable difference between simulations made with different $f_\alpha$ values, which means that the PDF is a good probe of this value between models. This is seen in the varying magnitudes of neighbouring pixels. Therefore, it appears that the two methods are complementary, in that they discriminate well for different variables (although neither one is especially powerful for distinguishing between models made with different $r_{H/S}$ values).

## 3. Implications

*Database of 21-cm signal lightcones*

We have presented 21SSD, a database of both high resolution, as well as SKA resolution, lightcones. A sparse sampling of a 3D parameter space gives 45 models, which have been used to create 135 lightcones.

*Quantifying 21-cm signal histories using Pixel Distribution Functions*

In addition to the Power Spectra, PDFs have been explored as a novel way of characterizing different Epoch of Reionization scenarios.

*Distance measures*

Both the Power Spectra, as well as the PDFs, are used to quantify the 'distance' between different models. They prove to be somewhat complementary diagnostics.

A future work will aim to use these distance measures to explore a choice of optimized parameter space. Such an optimization would aid in parameter extraction, and training neural networks – complimentary to the work of Shimabukuro 2017 and Mesinger 2015.

## References

Baek *et al.* 2009, *A&A*, 495, 389
Baek *et al.* 2010, *A&A*, 523, A4+
Iliev *et al.* 2009, *MNRAS*, 400, 1283
Mesinger *et al.* 2016, *MNRAS*, 459, 2342
Semelin B. 2016, *MNRAS*, 455, 962
Semelin B. & Combes F. 2002, *A&A*, 495, 389
Semelin B. *et al.* 2007, *A&A*, 495, 389
Semelin B. *et al.* 2017, *MNRAS*, 472, 4520
Shimabukuro H. & Semelin B. 2017, *MNRAS*, 468, 3869
Vonlanthen P., Semelin B., Baek S., & Revaz Y 2010, *A&A*, 532, A97+