Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **543,** 1092–1119 (2025) https://doi.org/10.1093/mnras/staf1466
Advance Access publication 2025 September 8

Square Kilometre Array Science Data Challenge 3a: foreground removal for an EoR experiment

A. Bonaldi [1]★ P. Hartley,[1] R. Braun,[1] S. Purser [1] A. Acharya [2] K. Ahn [3] M. Aparicio Resco [4] O. Bait [1] M. Bianco [5] A. Chakraborty [6] E. Chapman [7] S. Chatterjee [8] K. Chege [9] H. Chen,[10] X. Chen [11] Z. Chen,[12] L. Conaboy [7] M. Cruz [13] L. Darriba [14] M. De Santis,[15] P. Denzel,[16] K. Diao [17] J. Feron [7] C. Finlay,[18] B. Gehlot [9] S. Ghosh [9] S. K. Giri [19] R. Grumitt [17] S. E. Hong [20,21] T. Ito,[22] M. Jiang [23,24] C. Jordan,[25,26] S. Kim [27] M. Kim [28] J. Kim,[29] S. P. Krishna [5] A. Kulkarni [30] M. López-Caniego [4,31] I. Labadie-García [14] H. Lee [32] D. Lee [3] N. Lee,[32] J. Line,[25,26] Y. Liu,[33,34] Y. Mao [17] A. Mazumder,[35] F. G. Mertens [9,36] S. Munshi [9] A. Nasirudin [37] S. Ni [10] V. Nistane,[18] C. Norregaard,[38] D. Null [25,26,39] A. Offringa [40] M. Oh [3] S.-H. Oh [28] D. Parkinson [20] J. Pritchard [38] M. Ruiz-Granda [13,41] V. Salvador López [4] H. Shan [42,43,44] R. Sharma,[45] C. Trott [25,26] S. Yoshiura [46] L. Zhang,[47] X. Zhang [48] Q. Zheng,[42,44] Z. Zhu [42] S. Zuo [11] T. Akahori [46] P. Alberto,[49,50] E. Allys [51] T. An [42,43,44] D. Anstey,[33,34] J. Baek [20] Basavraj [52] S. Brackenhoff [9] P. Browne [53] E. Ceccotti [9,54] H. Chen,[10] T. Chen [5] S. Choudhuri [55] M. Choudhury [56] J. Coles,[33] J. Cook,[25,26] D. Cornu [36] S. Cunnington,[35] S. Das,[55] E. de Lera Acedo Acedo,[33,34] J.-M. Delouis [57] F. Deng,[11] J. Ding,[58] K. M. A. Elahi [55] P. Fernandez [59] C. Fernández [60] A. Fernández Alcázar [4] V. Galluzzi [54,61] L.-Y. Gao [48] U. Garain [62] J. Garrido [14] M.-L. Gendron-Marsolais [14,63] T. Gessey-Jones,[33,34] H. Ghorbel,[15] Y. Gong [11] S. Guo [42,43,44] K. Hasegawa,[64] T. Hayashi [65,66] D. Herranz [13] V. Holanda [59] A. J. Holloway [35] I. Hothi [51] C. Höfer [9] V. Jelić [67] Y. Jiang,[11] X. Jiang,[10] H. Kang,[29] J.-Y. Kim [32] L. V. Koopmans [9] R. Lacroix,[68] E. Lee [32] S. Leeney,[33,34] F. Levrier,[51] Y. Li [48] Y. Liu,[11] Q. Ma [69] R. Meriot [36] A. Mesinger,[37] M. Mevius [40] T. Minoda,[17] M.-A. Miville-Deschênes [51] J. Moldon [14] R. Mondal [70] C. Murmu [71] S. Murray,[72] Nirmala SR [52] Q. Niu [48] C. Nunhokee [25,26] O. O'Hara,[33,34] S. K. Pal [71] S. Pal,[73] J. Park [29] M. Parra [14] N. N. Patra,[71] B. Pindor,[26,74] M. Remazeilles [13] P. Rey [60] J. A. Rubino-Martin [75,76] S. Saha [62] A. Selvaraj,[25,26] B. Semelin,[36] R. Shah [77] Y. Shao,[48] A. K. Shaw [78] F. Shi [79] H. Shimabukuro,[80] G. Singh,[81] B. W. Sohn [20] M. Stagni [54] J.-L. Starck [82,83] C. Sui,[17] J. D. Swinbank [40] J. Sánchez [14] S. Sánchez-Expósito [14] K. Takahashi [22] T. Takeuchi,[84,85] A. Tripathi [71] L. Verdes-Montenegro [14] P. Vielva [13] F. R. Vitello [86] G.-J. Wang [87,88,89] Q. Wang,[11] X. Wang,[58] Y. Wang,[58] Y.-X. Wang,[48] T. Wiegert [14] A. Wild,[8] W. L. Williams [1] L. Wolz,[35] X. Wu [42] P. Wu,[48] J.-Q. Xia,[87] Y. Xu [11] R. Yan,[90] Y.-P. Yan,[87] Z. Yin,[90] Z. You [91] X. Yu,[90] K. Yu,[11] B. Yue [11] L. Zhang [92,93] Z. Zhao,[48] X. Zhao [17] and X. Zhou[11]

Affiliations are listed at the end of the paper

Accepted 2025 September 1. Received 2025 July 29; in original form 2025 March 14

ABSTRACT

We present and analyse the results of the Science Data Challenge 3a (SDC3a, https://sdc3.skao.int/challenges/foregrounds), an epoch of reionization (EoR) foreground-removal exercise organized by the Square Kilometre Array Observatory (SKAO) on SKA simulated data. The challenge ran for 8 months, from 2023 March to October. Participants were provided with realistic simulations of SKA-Low data between 106 and 196 MHz, including foreground contamination from extragalactic and Galactic

★ E-mail: anna.bonaldi@skao.int

emission, instrumental, and systematic effects. They were asked to deliver cylindrical power spectra of the EoR signal, cleaned from all corruptions, and the corresponding confidence levels. Here, we describe the approaches taken by the 17 teams that completed the challenge, and we assess their performance using different metrics. The challenge results provide a positive outlook on the capabilities of current foreground-mitigation approaches to recover the faint EoR signal from SKA-Low observations. The median error committed in the EoR power spectrum recovery is below the true signal for seven teams, although in some cases, there are some significant outliers. The smallest residual overall is $4.2^{+20}_{-4.2} \times 10^{-4}\,\mathrm{K^2 h^{-3} cMpc^3}$ across all considered scales and frequencies. The estimation of confidence levels provided by the teams is overall less accurate, with the true error being typically underestimated, sometimes very significantly. The most accurate error bars account for 60 ± 20 per cent of the true errors committed. The challenge results provide a means for all teams to understand and improve their performance. This challenge indicates that the comparison between independent pipelines could be a powerful tool to assess residual biases and improve error estimation.

Key words: instrumentation: interferometers – methods: data analysis – dark ages, reionization, first stars.

1 INTRODUCTION

The redshifted 21 cm signal, produced by the hyperfine spin-flip transition of neutral hydrogen, is a powerful tool for mapping the distribution of neutral gas during the cosmic evolution that followed recombination (Field 1958; Madau, Meiksin & Rees 1997). Studying this signal provides valuable insights into the astrophysical processes shaping three pivotal stages of cosmic history: the cosmic dawn (CD), the epoch of reionization (EoR), and the post-reionization era.

During the CD, neutral gas began to accumulate within dark matter haloes. As these haloes grew dense enough to undergo gravitational collapse, they formed the first luminous objects in the Universe. In the subsequent stage, the EoR, radiation from the first stars, galaxies, and quasars gradually ionized the surrounding neutral gas, transforming the Universe from predominantly neutral to mostly ionized. Finally, the Universe entered the post-reionization era, during which the intergalactic medium remained ionized, with only traces of neutral hydrogen confined to dense regions like galaxies.

Efforts to measure the redshifted 21 cm signal have led to the development of a variety of low-frequency radio facilities, which include both single-antenna instruments and interferometers. Single-antenna instruments aim to measure the globally averaged redshifted 21 cm signal across the sky, providing insights into the overall evolution of neutral hydrogen during different cosmic stages. Interferometers are designed to detect spatial variations in the redshifted 21 cm signal, enabling detailed mapping of the distribution of neutral hydrogen and ionized regions through cosmic history. These facilities, however, face several challenges, including the need to mitigate foreground contamination from Galactic synchrotron emission and extragalactic sources (Shaver et al. 1999; Jelić et al. 2008; Chapman & Jelić 2019), as well as terrestrial radio frequency interference. Additionally, achieving precise instrumental calibration is essential, as any systematic errors can mask the faint redshifted 21 cm signal and compromise its accurate measurement.

Despite the challenges, significant progress has been made over the last two decades in placing upper limits on the redshifted 21 cm signal power spectrum from the EoR using existing low-frequency interferometers. Instruments such as the Giant Metrewave Radio Telescope (Paciga et al. 2011), the Precision Array to Probe the EoR (Kolopanis et al. 2019), the LOw-Frequency ARray (LOFAR; Patil et al. 2017; Mertens et al. 2020), the Murchison Widefield Array (MWA; Barry et al. 2019; Li et al. 2019; Trott et al. 2020), and the Hydrogen Epoch of Reionization Array (HERA; Abdurashidova et al. 2022) have progressively tightened these constraints. Furthermore, several facilities, including the MWA (Ewall-Wice et al. 2016; Yoshiura et al. 2021), the Owens Valley Radio Observatory – Long Wavelength Array (Eastwood et al. 2019; Garsden et al. 2021), the LOFAR Amsterdam ASTRON Radio Transients Facility and Analysis Center (Gehlot et al. 2020), and the New Extension in Nançay Upgrading LOFAR (NenuFAR; Munshi et al. 2024), are also beginning to place constraints on the 21 cm signal from the CD. These efforts are complemented by results from single-antenna experiments, such as the Experiment to Detect the Global Epoch of Reionization Signature (Bowman, Rogers & Hewitt 2008; Bowman et al. 2018) and the Shaped Antenna measurement of the background RAdio Spectrum (Singh et al. 2017; Jishnu Nambissan et al. 2021).

The new generation of interferometers, such as the HERA (DeBoer et al. 2017) and the Square Kilometre Array – Low (SKA-Low; Koopmans et al. 2015), is poised to significantly advance the search for the redshifted 21 cm signal. These instruments, with their enhanced sensitivity, are designed to enable the most detailed observations of neutral hydrogen across a wide range of redshifts, offering unprecedented insights into its cosmic evolution. The order-of-magnitude improvement in sensitivity introduces not only new opportunities for groundbreaking science but also challenges, necessitating the development of innovative analysis techniques to handle the complex data sets produced.

This paper presents the results of the SKA Science Data Challenge 3a[1] (SDC3a), which focuses on developing and testing foreground removal techniques for an EoR experiment using realistic synthetic data sets that emulate the capabilities of the SKA-Low telescope. A major challenge lies in the lack of a detailed model for Galactic synchrotron emission at spatial resolution and frequencies of the SKA-Low, making the removal of this emission particularly complex. Additionally, source confusion from previously unobserved extragalactic radio sources further complicates the task.

Foreground mitigation strategies generally fall into three categories: subtraction, avoidance, and suppression (for an overview, see Chapman & Jelić 2019). Foreground subtraction aims to remove contamination across all k-scales in Fourier space, increasing the range of scales available for analysis. However, inaccuracies in subtraction can introduce biases across all k-scales, potentially contaminating the signal. Foreground avoidance, in contrast, sidesteps contamination by restricting analysis to well-defined 'clean' windows in cylindrical power spectrum ($k_\parallel$ versus $k_\perp$), free from foreground influence. While this approach avoids subtraction-related biases, it limits the usable scales, potentially introducing its own biases in the averaged power spectrum. Finally, foreground suppression reduces the influence of contaminated k-scales by downweighting them, mitigating the impact of residual foregrounds or subtraction errors.

[1] https://sdc3.skao.int/overview

These strategies, individually or in combination, provide a framework for addressing the critical challenge of foreground mitigation in 21 cm experiments.

In SDC3a, participants were tasked with recovering cylindrical power spectra of the redshifted 21 cm signal, which is deeply buried beneath much stronger foreground emissions, using their own mitigation methods and strategies on realistic synthetic SKA-Low data sets. The primary goal was to accurately extract the redshifted 21 cm signal while effectively mitigating foreground contamination and subtracting noise, ensuring that the results include robust error estimates. Participants were required to recover power spectra across various frequency segments within the simulation's frequency range, each corresponding to different redshifts. Where required by the teams, computing resources were offered by a list of computational facility partners to complete the challenge. Submissions were scored and ranked with a dedicated metric; additional metrics are also considered in this work for an in-depth analysis of the results.

This paper is organized as follows: Section 2 describes the simulated data; and Section 3 describes the analysis performed by the teams. The results are assessed in Section 4 and final conclusions are drawn in Section 5.

2 SIMULATED DATA

A detailed description of the simulation pipeline used to produce the data set is given in Bonaldi et al. (2025). Here, we only provide a short summary. The SDC3a data set represents the most advanced SKA-Low EoR simulation to date, in an effort to bridge the gap between simulations and realistic observational scenarios. This is particularly important, as the level of accuracy required for EoR detection demands a combined approach to the characterization and mitigation of both foreground emission and instrumental systematics (e.g. Barry et al. 2016). To address this, the SDC3a data simulation is fully performed in the visibility space; it includes both diffuse and point-source foreground contaminants, as well as the residual impact of bright, out-of-field sources that enter the field of view (FoV) through the telescope's sidelobes. Additionally, the effects of the ionosphere and calibration inaccuracies are modelled. This simulation does not include the leakage of polarized emission into Stokes *I*, which is another major challenge of EoR analysis (e.g. Asad et al. 2016; Gehlot et al. 2018; Spinelli, Bernardi & Santos 2018).

2.1 Data products summary

The simulated data set consists of the following products:

(i) Gridded visibilities, in the measurement set (MS) and UVFITS formats (7.5 TB each) produced with the OSKAR[2] (Dulwich et al. 2019) package. The telescope model makes use of the SKA-Low configuration of 512 stations presented in Dewdney & Braun (2016) thus producing 512 × 511 baselines. A 4-h duration observation was simulated, directed at (RA, Dec.) = (00 h, −30 deg) using a 10 s integration time and a 100 kHz frequency sampling between 106 and 196 MHz. The thermal noise level was finally rescaled to correspond to a target 1000 h integration. This approach mimics accumulating 250 4-h observations to produce the final data product.

(ii) Image cubes, and the associated point spread function (PSF) cubes, produced from the visibilities with the WSCLEAN task[3] (Offringa et al. 2014). Two different versions of an image and corresponding synthesized beam were produced: the first with natural weighting and no deconvolution and the second with uniform weighting followed by Gaussian tapering and a multiscale deconvolution. Each cube covers the 106–196 MHz range with 900 channels, and 2048 × 2048 pixels of 16 × 16 arcsec size, for a data size of 15 GB per data cube.

(iii) Time-averaged primary beam at all frequencies.

(iv) Test data set, consisting of an EoR + noise imaging simulated observation (with a different EoR model than what used for the challenge) and accompanying 'true' power spectrum in the 166–181 MHz frequency range, to help teams validate their power spectrum calculation pipelines.

2.2 Sky model

The sky components added to the simulations are the following:

(i) *Out-of-field extragalactic sources*. Sources above 5 Jy at 150 MHz were added over the full 2π steradians above the horizon at any time. Those are the so-called A-Team sources as well as sources from a composite GaLactic and Extragalactic All-Sky MWA (GLEAM) and Long Baseline Epoch of Reionisation Survey (LoBES) catalogue (Lynch et al. 2021).

(ii) *In-field extragalactic sources*. Within 8 × 8 deg of the field centre, sources were added from GLEAM/LoBES above 100 mJy at 150 MHz, and from a mock Tiered Radio Extragalactic Continuum Simulation (T-RECS)[4] (Bonaldi et al. 2019, 2023) catalogue for $1\ \mu\text{Jy} \leq \text{I}_{150\,\text{MHz}} < 100$ mJy. In total, over 15 million sources were added.

(iii) *Galactic emission*. We used a modified version of the GSM2016 sky model (Zheng et al. 2017), with a quadratic interpolation in log(frequency) to eliminate discontinuous derivatives within our band. Furthermore, spatial frequency content beyond the native resolution (∼1 deg at our frequencies) was added to the GSM2016 output using synthetic observations of low-frequency synchrotron emission from Bracco et al. (2022). These synthetic observations were based on magneto-hydrodynamical simulations of colliding super shells in the multiphase interstellar medium.

(iv) *CD and EoR*. The cosmological signal has been produced with the 21CMFAST[5] (Mesinger, Furlanetto & Cen 2011; Murray et al. 2020) code, for Planck18 cosmology $\Omega_m = 0.30964$, $\Omega_\Lambda = 0.69036$, and $H_0 = 67.66$ (Planck Collaboration VI 2020). The 21CMFAST reionization parameters and the corresponding brightness temperature are shown in Fig. 1. This scenario has been chosen as it yields a strong CD/EoR signal over a wide-frequency range, therefore providing good potential for detection and measurement. The simulated EoR signal, in units of cubic Mpc, has been converted into observational quantities (deg, deg, MHz) by means to the TOOLS21CM library[6] (Giri, Mellema & Jensen 2020).

2.3 Error model

In an effort to make the simulation more realistic, various instrumental errors were included.

(i) *Partial de-mixing*. The strong sources entering the field from the far sidelobes are normally modelled and removed with a so-called 'de-mixing' process within a calibration and imaging pipeline. To

[2] https://ska-telescope.gitlab.io/sim/oskar/
[3] https://gitlab.com/aroffringa/wsclean
[4] https://github.com/abonaldi/TRECS
[5] https://21cmfast.readthedocs.io/en/latest/
[6] https://tools21cm.readthedocs.io/

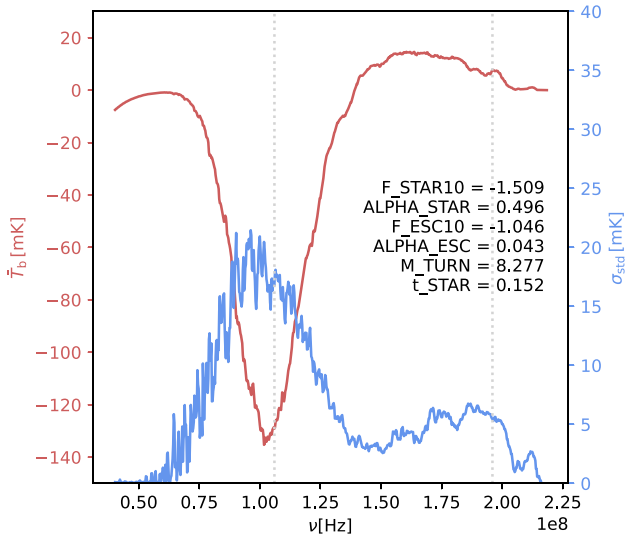

Figure 1. 21CMFAST parameters used for the generation of the EoR signal and corresponding mean (red) and standard deviation (blue) of the brightness temperature as a function of frequency.

simulate a partially successful de-mixing, the out-of-field sources were added with amplitudes attenuated by a factor of 10^{-3} (see Bonaldi et al. 2025, for more details).

(ii) *Ionospheric effects*. The ARATMOSPY[7] (Srinath et al. 2015) code was used to construct an ionospheric model that was intended to represent moderately good observing conditions, characterized by a correlation scale, $r_0 = 7$ km. The total electron counts and the corresponding phase modulation has been attenuated by the factor 10^{-2} to mimic the outcome of a successful direction dependent (DD) calibration.

(iii) *Gain calibration errors*. The effect of direction independent (DI) gain calibration errors were simulated by adding a noise term to the gain model, with standard deviations of 0.02 deg in phase and 0.02 per cent in amplitude for each of the time and the frequency domains. The time domain represents residual broad-band gain calibration errors and the frequency domain represents residual bandpass calibration errors.

(iv) *Thermal noise*. The nominal sensitivity per polarization per unit bandwidth and unit time provided by OSKAR was scaled finally to represent the average of two polarizations and a total observing time of 1000 h, instead of the 1 polarization and 4 h track that was simulated.

2.4 Systematics introduced by the simulation pipeline

Upon detailed inspection of the SDC3 data set, an error in the simulation pipeline was uncovered, which meant that the sky components in the simulated data were not as intended. The issue involved the frequency behaviour of the EoR signal, the Galactic foregrounds, and the faint extragalactic sources, which was consistently steeper than what modelled by $\Delta\alpha = -0.7$.

Although the data set has since been re-issued (Bonaldi et al. 2025, present the corrected simulation outputs and codes), the problem was discovered only after the challenge completion. This work therefore makes use of the original SDC3a data set and is scored against the corresponding 'true' EoR.

Although the issue described above affected the modelling of the components, it did not compromise the usefulness of the data set for foreground removal. One hypothesis that is often crucial to this part of the analysis is that the foreground components are smooth in frequency. The frequency smoothness of the components was not altered by the steepening, therefore there was no impact on foreground removal methods relying on this assumption.

While the simulated EoR signal strength was affected, there are only upper limits on this quantity, and the simulated signal is abundantly within the existing constraints. Similarly for Galactic synchrotron, which is by far the dominant diffuse foreground component at the frequencies considered, the committed error is within model uncertainties. A realistic Galactic synchrotron model should include both significant variation of the synchrotron spectral index across different regions of the sky (Giardino et al. 2002; Delabrouille et al. 2013), due to local variation in the magnetic field and electrons density, and spectral steepening with frequency, due to electron energy losses (Platania et al. 2003). However, the lack of data especially at low frequency provides limited constraints (e.g. Planck Collaboration X 2016a; Planck Collaboration XXV 2016b).

Current constraints on the faint extragalactic source population are more abundant and stringent (e.g. Mandal et al. 2021). Source-by-source spectral index variations are significant, but there is no mechanism that would result in a consistent spectral steepening across an entire region of the sky. Therefore, the faint extragalactic source population in this work is not a good representation of the real sky. Amplitudes are consistent with the model at 106 MHz but, due to the spectral steepening, are lower than what expected by about 65 per cent at the highest frequency of 196 MHz.

Additionally, the issue described above affected the test data set, although to a smaller extent due to the limited frequency coverage used for this product (up to 6 per cent at 181 MHz). Teams that had used the test data set to calibrate their final results were allowed to submit a correction after the error was discovered. Due to time constraints, only one team (HIMALAYA) submitted the correction.

2.5 The challenge defined

The challenge asked the teams to recover six cylindrical (2D) power spectra $P(k_\parallel, k_\perp)$ of the EoR signal, clean from foregrounds and noise-subtracted, and the corresponding error bars. The six spectra were to be computed each over 15 MHz intervals and without overlap, to cover the whole 90 MHz frequency range of the simulation, and over the central 4×4 deg out of the full 8×8 deg FoV to limit noise. The errors on the power spectrum were to be approximated as Gaussian and uncorrelated and provided as 1σ. The bins to be used to average the power spectrum in both $k_\parallel$ and $k_\perp$ were also provided.

The computation of the power spectrum requires converting the observational units (deg, deg, MHz) back to comoving units (Mpc, Mpc, Mpc), using again the cosmological parameters. It is worth noting that, for this inverse conversion, indications were given to adopt $H_0 = 100$ instead of $H_0 = 67.66$ as assumed in Section 2. This is because power spectrum routines and libraries available in the literature can adopt either Mpc or Mpc/h (with $h = H_0/100$) as units of comoving distance, which gives rise to inconsistent power spectrum binning and normalization. The choice of $h = 1$ allows comparing power spectra with each other and with the ground truth, whichever convention was used for the comoving distance. This is justifiable only in light of the scope of the challenge, which

[7] https://github.com/shrieks/ARatmospy

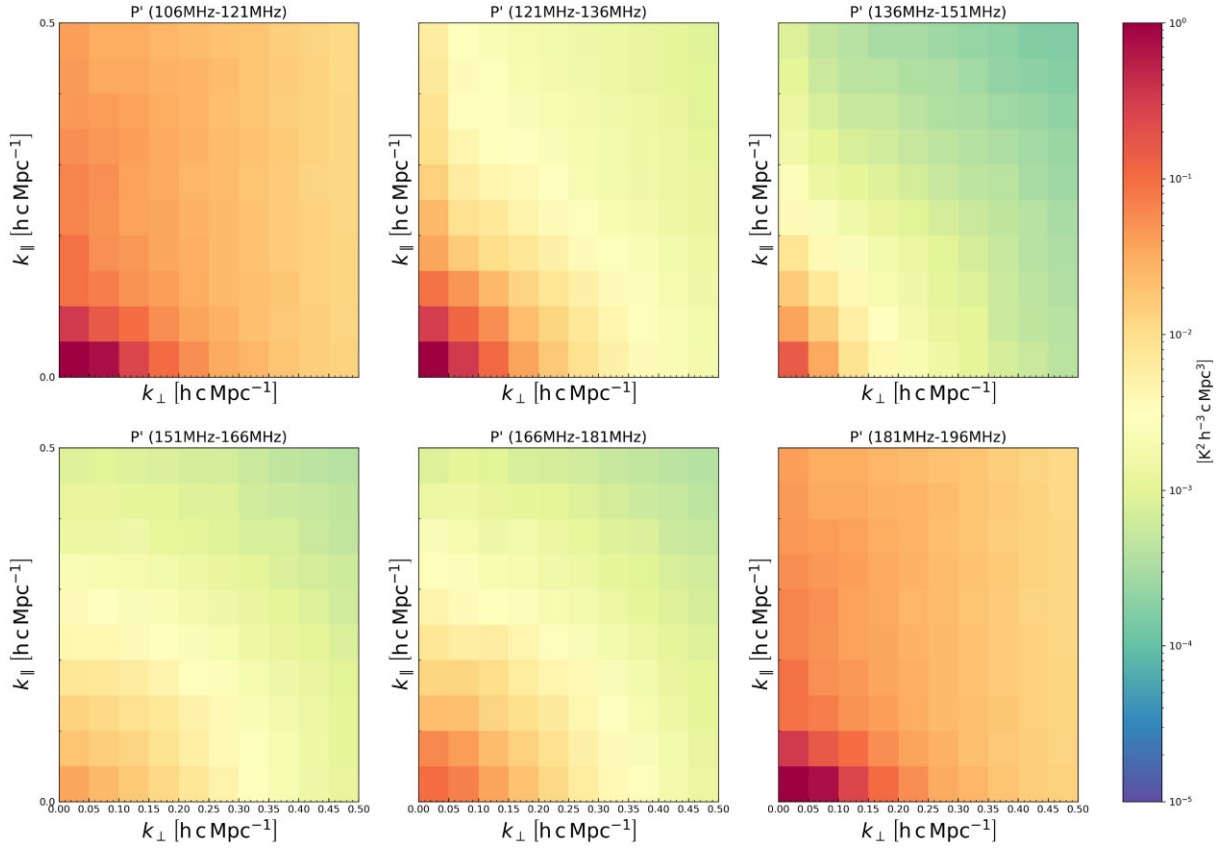

Figure 2. Cylindrical power spectrum of the true noiseless EoR, P' (computed with the $H_0 = 100$ convention).

is foreground removal only and does not include inference of the reionization properties.

The 2D power spectra for the true noiseless EoR signal were computed with the TOOLS21CM library using the frequency and $k_\parallel$, $k_\perp$ binning adopted in this challenge and the $h = 1$ convention. The resulting true EoR power spectrum for all considered frequencies in shown in Fig. 2; this has been used for the computation of the challenge score and other performance metrics.

Fig. 3 compares the EoR power spectrum (solid line) with that of the simulated data (dot–dashed line) and instrumental noise (dashed line). In order to compare them on the same plot, only the diagonal terms of the 2D power spectra ($k_\parallel = k_\perp$ elements) are shown, as representative of the others. This figure illustrates the challenge faced by the teams in reducing a foreground contamination many orders of magnitudes greater than the signal of interest. It also shows that, given the choice of a simulated 1000 h exposure, the contamination due to noise is much more limited.

2.6 Computational facility partners

Due to the computational complexity of the challenge, an effort was made to allow all interested teams to participate, by offering dedicated computational resources to teams that requested them. This was done though a network of computational facility partners, that kindly made some resources available to SDC participants. The network included various computing facilities, having in general different hardware and software infrastructure. Globally, they were able to offer a diverse set of resources, including Central Processing Unit (CPU) and Graphics Processing Unit (GPU) computation, and different specifications regarding virtual and physical memory.

No attempt was made to deliver a uniform resource allocation across different teams. Instead, teams specified their needs and were awarded the allocation that matched their needs more closely, subject to availability. Teams that worked on their own resources were able to download the data from one of the facility partners. Given the size of the visibility set (7.5 TB for each of the MS and UVFITS formats), this was much more efficient than having a single download point.

The list of facility partners in alphabetical order, and a brief description of the provision they offered, is below.

2.6.1 *ASTRON/SURF (Amsterdam, Netherlands)*

The ASTRON/SURF facility provided access to Spider,[8] a versatile high-throughput data-processing platform designed for processing large, structured data sets. Spider is a combined CPU and GPU system, located at the SURF data centre in Amsterdam, the Netherlands.

2.6.2 *China SRC (Shanghai, China)*

The China SKA Regional Centre prototype (cnSRC; An, Wu & Hong 2019; An et al. 2022) is a cutting-edge computing platform that employs a novel hybrid architecture. This advanced setup integrated general-purpose Intel x86 CPUs, GPUs, and ARM processors, offering unparalleled flexibility and efficiency for data processing tasks.

[8] https://doc.spider.surfsara.nl

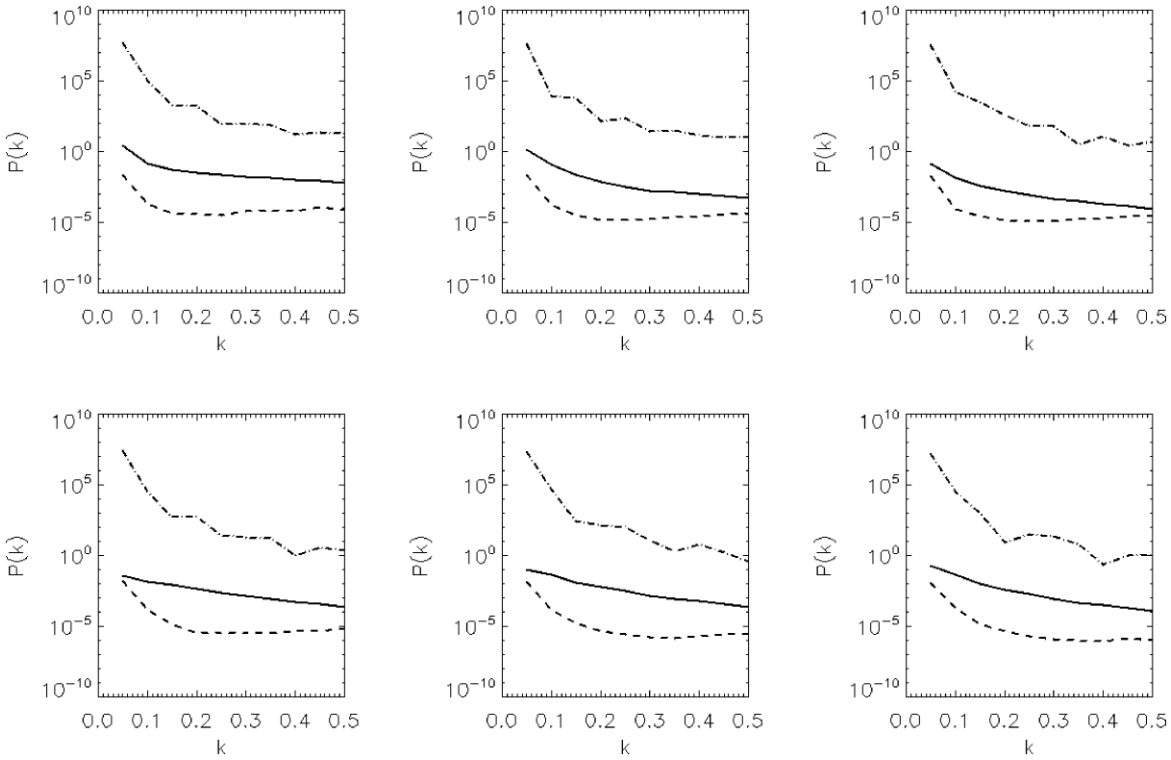

Figure 3. Diagonal terms ($k_{\parallel} = k_{\perp}$ elements only) of the power spectra or the simulation (dot–dashed line), clean EoR (solid line), and instrumental noise (dashed line) for all the frequency channels (top: 106–121, 121–136, and 136–151 MHz) from left to right. Bottom: 151–166, 166–181, and 181–196 MHz from left to right.

The prototype comprises 35 Intel x86 CPU nodes with a total of 2240 cores (up to 128 cores per node), 12 ARM CPU nodes with 1152 cores (96 cores per node), and 4 GPU nodes equipped with 16 Nvidia V100 and 8 A100 GPUs, collectively delivering approximately 800 TFlops of computing power. The CPU nodes contribute an additional 196 TFlops of computational capacity.[9] In terms of storage, cnSRC offers an approximately 9 PB of distributed storage capacity, supported by a high-speed internal network enabling data exchange at rates of 100–200 Gb s^{-1} between compute and storage nodes. Certain nodes are capable of accommodating up to 4 TB of memory, with a maximum of 36 GB per core, ensuring efficient processing of large-scale data sets. Furthermore, cnSRC is connected to a 2 Gbps transcontinental internet link, facilitating seamless international collaboration, and maintains a 200 Mbps connection with other SKA regional centre nodes to support global research initiatives.

2.6.3 Galicia Supercomputing Center/ CESGA (Santiago de Compostela, Spain)

The Galician Supercomputing Center (CESGA) facility provided access to FinisTerrae-III,[10] supercomputer, located in Santiago de Compostela, Spain. FinisTerrae-III is an HPC system integrating Intel x86 CPUs, AMD CPUs and Nvidia GPUs, with over 384 nodes, totalling more than 22 000 x86 cores and 157 GPUs with Infiniband interconnect and 6 PB of high performance storage Lustre Filesystem. For SKA SDC3a, CESGA allocated up to 100.000 core-hours and up to 2 TB of permanent storage per team.

2.6.4 GENCI/IDRIS (Orsay, France)

GENCI (*Grand Equipement National de Calcul Intensif*) granted CPU and NVIDIA V100 GPU computing resources to two teams on the Jean Zay supercomputer hosted by IDRIS (Institute for Development and Resources in Intensive Scientific Computing).

IDRIS is the national supercomputer center of the CNRS (French National Center for Scientific Research) for High Performance Computing and Artificial Intelligence. It currently hosts Jean Zay, a supercomputer comprising 720 Intel CPU nodes and 843 GPU nodes equipped with V100, A100, or H100 NVIDIA GPUs for a total of 3704 GPUs. Following three successive extensions, the cumulated peak performance of Jean Zay reached 126 Pflop/s starting in 2024 July.

2.6.5 INAF (Italy)

The Italian National Institute for Astrophysics (INAF), through its 'Unità Scientifica Centrale 8', provided computational support via the distributed HPC facility 'Pleiadi'. This infrastructure consists of 192 nodes, totalling over 7000 Intel Xeon E5-2697 V4 cores, interconnected via an Omni-Path HFI Silicon 100 Series network at 100 Gbit, with external connectivity at 1 Gb s^{-1}, and managed through the SLURM workload scheduler.

For SKA SDC3a, INAF allocated up to 8 CPU nodes (128 GB RAM each) at the Institute of Radioastronomy in Bologna and up to

[9] https://shaoska-user-guide.readthedocs.io/
[10] https://cesga-docs.gitlab.io/ft3-user-guide/overview.html

6 CPU nodes (128 GB RAM each, each with an available GPU) at the Astrophysical Observatory of Catania. Each team granted access to INAF resources had a dedicated 10 TB storage space for the duration of the challenge, ensuring seamless execution of workflows and analysis. Additionally, the challenge input data, including visibilities in both UVFITS and MS formats and image cubes, were made available in a read-only partition at both sites, facilitating efficient data access and processing.

2.6.6 JPSRC (Tokyo, Japan)

The Japanese SKA Regional Centre (JPSRC) prototype is an inter-university cooperation led by the SKA1 Japan Promotion Group (SKAJ) of the National Astronomical Observatory of Japan. JPSRC contains heterogeneous workstations with 152 CPU cores, 6 GPUs, 1 NEC SX-Aurora TSUBASA vector engine, as well as 1.86 TB DRAM memory, 3.71 TB M.2/SSD, and 338 TB HDD storage as of the SDC3a. The workstations were federated by OpenStack as virtual machines, but JPSRC solely provided one of the workstations as a stand-alone machine to segment the servers into the one for SDC3a and the others for deploying SRCNet v0.1.

The provided workstation was equipped with an AMD EPYC 7713P 64 core CPU, 2 x NVIDIA Quadro A6000 48GB, 1.024 TB DRAM memory, and 162 TB HDD, and the network speed was typically 1 Gbps. The operating system was Linux Ubuntu 22.04, and an environment for docker and singularity containerizations was supported. A secure shaell protocol (SSH) access was only permitted, and the primary data cube provided by SKAO for SDC3a was pre-stored by JPSRC at the local storage, to which a team was able to access.

2.6.7 SPSRC/IAA-CSIC (Granada, Spain)

The Spanish prototype of an SRC (a.k.a. espSRC), being developed at the Instituto de Astrofísica de Andalucía (IAA-CSIC) in Spain, is one of the 17 national initiatives contributing to the SRCNet development. The espSRC OpenStack cloud gathers 240 CPUs cores and 2.5 TB of memory across five compute hypervisors, plus 600+ TB of SSD usable storage capacity managed by Ceph. The servers are interconnected by a 100 Gbps network and the cluster is connected to RedIRIS (the Spanish National Research Network) with a 10 Gbps link.

The espSRC hosted two SDC3a teams which were provided with a virtual machine with 32 cores and 128 GB memory. 100 GB root disk using local SAS SSD and up to 5 TB of additional block storage were assigned. The Operating System and the software in the virtual machine was fully customizable by the team through pseudo-sudo access. Accessibility to the machine was flexible, provided by ssh, remote desktop, and a clientless remote desktop gateway.

2.6.8 Swiss National Computing Center/ CSCS (Lugano, Switzerland)

The Swiss National Supercomputing Center (CSCS) is a key partner in deploying and operating the Swiss SRCNet node, contributing its expertise and computational resources to support scientific research. For this project, one SDC3a team was granted access to Piz Daint, CSCS's former flagship supercomputer, which served as a testbed for Swiss SRCNet node science activities.

At its peak, the now decommissioned Piz Daint was one of the most powerful supercomputers in Europe, comprising over 5000 NVIDIA P100 GPGPUs. The team was allocated 15 000 node hours on this system, utilizing nodes equipped with one NVIDIA P100 GPU, 12 Intel Xeon CPU cores, and 65 GB of RAM. To support large-scale computations and data-intensive workflows, the project leveraged over 15 TB of long-term storage along with up to 10 PB of scratch space for high-speed temporary data access.

2.6.9 UC-LCA (Coimbra, Portugal)

The Laboratory for Advanced Computing of the University of Coimbra (UC-LCA) has provided access to its Navigator cluster, which includes CPU and GPU nodes. The resources included 20 TB of shared storage space (including image cube data) and 100 000 core-hours for the team assigned to UC-LCA. The team could use up to 1032 cores of Intel Xeon E5-2697v2 or up 400 cores of Intel Xeon Gold 6148. Each node with these processors has 96 GB RAM and the interconnect is infiniband FDR/EDR.

2.6.10 UKSRC, IRIS-CAM (Cambridge, UK)

The UK SKA regional centre (UKSRC) resource at the University of Cambridge comprises a multinode HPC/GPU cluster running a Slurm batch scheduler. An OpenStack-hosted Platform-as-a-Service Azimuth applications portal is also available on request. Each SDC3a team was allocated 100 000 core hours, 1000 GPU hours, and 20 TB of storage.

2.6.11 UKSRC, IRIS-MAN (Jodrell Bank Observatory, University of Manchester, UK)

Access was provided to the Galahad HPC cluster. This is a hybrid CPU/GPU cluster that includes large memory nodes that target wide field imaging and advanced image analysis.

3 SDC3A TEAMS

This section describes the analysis performed by the SDC3a participating teams. Table 1 presents an overview of the strategies and the techniques employed by each team, listed in alphabetical order, with more details given in separate sections. The table shows that most teams applied multiple stages of cleaning and processing. About half of them started from the visibility data, with the rest relying directly on the image data.

Subtraction of strong point sources was typically performed on the visibilities, by using existing bright source catalogues, complemented in some cases by sources detected on the SDC3a images.

On the image data, commonly adopted foreground-cleaning methods include: polynomial fitting, where a smooth function of frequency is subtracted from the maps; principal component analysis (PCA, e.g. Irfan & Bull 2021); Gaussian process regression (GPR, e.g. Mertens, Ghosh & Koopmans 2018); independent component analysis (ICA, e.g. Pedregosa et al. 2011; Chapman et al. 2012); and neural networks.

Flagging of the data in delay-space or k-space, a technique sometimes called 'foreground avoidance', was further adopted by a few teams.

We note that, despite the use in some cases of the same technique, performance may vary depending on the exact implementation, parameter setting, and the combined use of additional pre- and post-processing steps. The next subsections detail the analysis from each

Table 1. Summary of the strategy employed by different teams, detailing their use of the visibility data, the image data, and flagging of the results.

Team name	Cleaning of visibility data	Cleaning of image data	Foreground avoidance	Section
Akashanga	Subtraction of a continuum emission sky model	GPR	No	3.1
Cantabrigians	No	Point-source subtraction and GPR	Yes	3.2
DOTSS-21	Subtraction of point sources from catalogues and detections on the data	GPR (on version 2 and 3 of the results)	Yes (on version 1 of the results)	3.3
ERWA	No	U-Net neural network and polynomial fitting	No	3.4
Foregrounds-FRIENDS	No	PCA	No	3.5
HAMSTER	Computation of delay power spectrum	No	Yes	3.6
Hausos	No	Transformer-based neural network based on SETR	No	3.7
HIMALAYA	No	Reconvolution and PCA	No	3.8
KORSDC	No	Point-source detection and subtraction; smoothing; ICA	No	3.9
KUSANAGI	Removal of detected point sources	Polynomial fitting/GPR used on alternative versions of the results	Yes	3.10
Nottingham-imperial	Subtraction of bright point sources from catalogues	ICA	No	3.11
Pisano Galaxy Moppers	Forward modelling of EoR and foregrounds and inference	No	No	3.12
REACTOR	No	Beam removal with semisupervised learning technique; PCA	No	3.13
Shuimu–Tianlai	No	Oriented Singular Value Decomposition	No	3.14
SKACH	Iterative subtraction of point sources from catalogues and detections on the data	Polynomial fitting and U-shaped convolutional neural network	No	3.15
SROT	No	GPR	No	3.16
Wizards of Oz	Iterative subtraction of point sources from catalogues and detections on the data; polynomial fitting	No	No	3.17

team, while performance comparison considerations are presented in Section 4.

3.1 Akashganga

The team consists of observational cosmologists based in India, as a part of the SKA-India consortium. A hybrid strategy to subtract the foreground and estimate the cosmological 21 cm power spectrum was employed.

A high-resolution image of the visibility data was first made using WSCLEAN[11] (Offringa et al. 2014). The image size used was $9.1^\circ \times 9.1^\circ$, using 2048×2048 pixels with a size of 16 arcsec, using a uniform weighting. A second-order polynomial was fitted to the clean components to account for the frequency variation of the flux of the sources. The final model image was inverted and written to the MODEL_DATA column. The UVSUB task in CASA[12] (McMullin et al. 2007) was used to subtract the continuum model from the visibility data. After continuum subtraction, a dirty image cube was produced for each sub-band using the same parameters as provided in the SDC3a data description document. Two image cubes were maid, using even and odd timestamps, and their cross-correlation was then used to estimate the power spectrum to avoid noise bias.

For the residual foreground subtraction, GPR (Mertens et al. 2018) was applied to the point-source-subtracted residual data cube. GPR builds the model covariance of the different components of the data – foreground, 21 cm signal, and noise. The hyperparameters of the covariance functions are optimized based on different spectral coherence of different signal components of the data. The foreground signal is then predicted from the model covariance of foreground and subtracted from the data cube. GPR was applied to both even and odd data cubes; the cross-correlation of the resulting data cubes was cylindrically averaged to estimate $P(k_\perp, k_\parallel)$. The signal loss was corrected for by estimating the transfer function following the process outlined in Cunnington et al. (2023).

3.2 Cantabrigians

The Cantabrigians collaboration was formed for the first time in this data challenge to develop a pipeline suitable for interferometric experiments with the SKA, leveraging prior experience in data analysis for single-antenna 21 cm cosmological experiments.

The pipeline is based on a hybrid strategy that combines foreground removal and avoidance approaches. The pipeline takes the natural weighted images as input data and removes the foregrounds from the total sky intensity by point source subtraction and a Bayesian GPR model. The GPR approach the pipeline is based on is previously established in Mertens et al. (2018, 2020). The underlying principle behind applying GPR to foreground removal is also explained in these works. The Bayesian framework within which the model is utilized is well described in Soares et al. (2022). The impact of astrophysical foregrounds is further reduced by filtering the Fourier modes of the foregrounds from those of the 21 cm signal, leaving an 'EoR window' in cylindrically averaged 2D power spectra.

As a pre-calibration, bright point sources are removed from the SDC3a image data, since it is found that the GPR models

[11] https://sourceforge.net/p/wsclean/wiki/Home/
[12] https://casa.nrao.edu

struggle with discrete radio sources scattered across the sky. The point source catalogue is a composite of the A-Team and MWA GLEAM and LoBES sources, provided in the data challenge. A sky model is created from this catalogue, applying the same flux filtering as described in Section 2. The model is then passed to the OSKAR simulator using an SKA-Low End-to-End Simulation Pipeline released in O'Hara et al. (2024). The settings for telescope layout, integration time, and frequency channels are also adjusted to maintain consistency with the SDC3a simulation. The measured visibilities generated from OSKAR are gridded and inverse Fourier transformed to yield natural weighted images for the point sources by using WSCLEAN. The desourced images are obtained by subtracting the GLEAM and LoBES sources from the SDC3a image flux. Since the subtraction can be performed equivalently either in image or visibility space, the subtraction was performed in image space for reasons of efficiency.

For the GPR foreground subtraction, three kernel functions of the Matérn class were used to represent the components in the observed sky signal: an exponential kernel for the 21 cm signal, and two radial basis function (RBF) kernels for both intrinsic and instrumentally corrupted foregrounds. The marginalized likelihood of the model is passed to a trans-dimensional nested sampler POLYCHORD (Handley, Hobson & Lasenby 2015a, b) to sample the posterior probability density and global evidence of the model. The hypeparameters with the peak posterior density are used in the model for signal separation. The global Bayesian evidence can also give a measure of how well the multikernel GPR model describes the observed data.

From the residual flux, the cylindrical power spectrum is calculated by using the `ps_eor` PYTHON code.[13] A tukey and a nuttall window function are applied to the data while transforming the images along their spatial axes and along the frequency axis respectively. The foreground-dominated regime draws a sharp boundary from the EoR window due to the frequency smoothness of radio frequency band foregrounds. This region is subsequently masked in the 2D power spectra based on the foreground avoidance strategy.

Several potential improvements to the existing pipeline are possible. First, from an algorithmic aspect, the pipeline could incorporate a baseline-dependent foreground model for the separation of the 21 cm signal. However, evaluating such a model for each baseline would incur significant computational expense. It is recommended to group the baselines by the length so that within each group the length-scale variation is limited. Secondly, the CLEAN algorithm could be employed to reduce the effects of dirty beam on the radio sources. Thirdly, uniformly weighted images may be preferable over naturally weighted ones in scenarios where the number of short baselines is significantly greater than the number of longer baselines, even though this approach comes at the cost of reduced sensitivity. Lastly, although we found gain instabilities and a phase screen to have negligible effects on the sources, corrections may be also needed in the pipeline, as they may result in source flux and position errors (Trott et al. 2018; Mertens et al. 2020).

3.3 DOTSS-21

To address the SDC3a challenge the team employed methodologies grounded in the well-established practices of the LOFAR-EoR and NenuFAR Cosmic Dawn experiments, adapted to address the unique challenges of the SKA-Low simulated data set. In essence this methodology consists of first detecting and subtracting compact and diffuse sources of foreground emission. maximum-likelihood GPR (ML-GPR) was then used to model the residual foregrounds to separate them from the 21 cm signal, and the cylindrical power spectra of the recovered 21 cm signal was finally generated.

3.3.1 *Compact sources and diffuse emission*

The strategy for subtracting compact sources and diffuse emission was executed iteratively. Initially, the brightest sources from the raw visibilities were subtracted. Based on the data set description, this included A-Team sources and prominent sources from the GLEAM catalogue (Hurley-Walker et al. 2019, 2022; Franzen et al. 2021). Using OSKAR (Dulwich et al. 2019), visibilities for the six brightest A-Team sources (Taurus A, Cygnus A, Cassiopeia A, Fornax A, Pictor A, and Hercules A) were simulated, alongside 210 point sources from the GLEAM catalogue, each exceeding 10 Jy at 200 MHz, adjusting their flux by a factor of 10^{-3} as outlined in the challenge documentation. This model of the 'brightest' sources was then subtracted from the raw visibility data.

To this end, a sky model of 'bright' compact sources within the primary beam's main lobe was produced. Using WSCLEAN (Offringa et al. 2014), images of 2500 × 2500 pixels were generated at a resolution of 15 arcsec pixel^{-1}, imposing a minimum baseline restriction of 250λ to mitigate diffuse emission impact. The PYPBSF tool was used to create an intrinsic sky model from these image cubes. The visibility model for these 'bright' sources was predicted with OSKAR and subtracted. This process was repeated on the residual data to model and subtracted 'faint' point sources. The final phase involved constructing a model for the Galactic diffuse emission using multiscale cleaning techniques and a combination of Gaussian components for accurate representation (Offringa & Smirnov 2017; Gehlot et al. 2022). To ensure that the 21 cm signal is not overfit, only seven channels spanning the observation's full bandwidth, with steps of 15 MHz were used. Employing again OSKAR, this model was subtracted from the visibilities.

3.3.2 *Residual foreground removal*

The separation of the residual foregrounds from the 21 cm signal was achieved through the application of the ML-GPR method, as described in Mertens et al. (2018); Mertens, Bobin & Carucci (2024); and Acharya et al. (2024). This method separates the different components of the observed signal – foregrounds, the 21 cm signal, and noise – by building a statistical model that exploits their distinct spectral-coherence signatures. The foreground emission was modelled using two components: one for the spectrally smooth intrinsic foregrounds and another for the 'mode-mixing' component (a consequence of the instrument's chromatic response). Time difference image cubes provided a noise estimate. For the 21 cm component, a kernel trained from 21CMFAST simulations (Mesinger et al. 2011; Park et al. 2019; Murray et al. 2020) was used as described in Mertens et al. (2024).

3.3.3 *Power-spectra calculation*

The power spectra calculation employed the `pspipe`[14] and `ps_eor`[15] packages, specifically developed for the analysis of 21 cm

[13] https://gitlab.com/flomertens/ps_eor

[14] https://gitlab.com/flomertens/pspipe

[15] https://gitlab.com/flomertens/ps_eor

observations within the LOFAR-EoR and NenuFAR Cosmic Dawn collaborations (e.g. Mertens et al. 2020; Munshi et al. 2024). The submitted three distinct results, each representing a different strategy in foreground management and signal extraction:

Avoidance strategy. Here, ML-GPR was not run, and power spectra modes predominantly influenced by foregrounds were instead filtered.

Removal strategy I. Foregrounds were removed, applying ML-GPR to an image cube after the subtraction of only compact sources.

Removal strategy II. An extension of Removal strategy I that included the subtraction of the modelled diffuse emission as well. This approach aimed to improve the reliability of ML-GPR by reducing the dynamic range of residual foregrounds, albeit introducing an element of risk.

Ultimately, the two removal strategies, which implemented a full foreground removal, scored higher than the avoidance strategy, underscoring the effectiveness of this approach.

3.4 ERWA

In scenarios where observational and instrumental effects (in particular, the frequency-dependent ones) are not considered, methods such as PCA, polynomial fitting or ICA have shown potential for recovering the EoR signal (e.g. Chapman et al. 2012, Wang et al. 2013, He et al. 2024). Thus, the adopted approach employed a U-Net neural network for denoising and deconvolution, with the goal of revealing the underlying sky map, before polynomial fitting was applied to derive the EoR map and calculate its 2D power spectrum.

3.4.1 Beam deconvolution and data denoising

The most critical part of the approach lies in the extraction of a clean sky map, free from instrumental and observational distortions, from the dirty map. To achieve this, the team proposes the construction of a U-Net network. The architecture of our U-Net network is designed to be straightforward yet effective, comprising six convolutional layers, three max-pooling layers, six transpose convolutional layers, three upsampling layers, and a final transpose convolutional layer. The network inputs the dirty map with instrumental and observational noise and outputs the corresponding clean sky map.

For the training set, the sky map was constructed following Li et al. (2019), incorporating components such as Galactic synchrotron emission, Galactic free–free emission, extragalactic point sources, radio haloes, and the EoR signal. Utilizing the OSKAR software package, the SKA-Low antenna's response and layout were simulated, incorporating thermal noise, antenna gain and phase errors, and ionospheric effects. The dirty map was generated using the WSCLEAN package, with simulation parameters such as FoV, pixel size, and frequency resolution aligned with official specifications.

The U-Net network was applied to the SDC3a natural weighting image cube to obtain the clean sky map. In the pre-training phase, bright point sources are masked. The dirty maps are then standardized across frequencies and segmented into 128×128 pixels. Training then proceeds by minimizing the pixelwise L2 norm between the network-predicted sky map and the true sky map.

3.4.2 Foreground removal and power-spectrum calculation

A pixel-by-pixel fifth-order polynomial fitting was performed on the clean sky map for each frequency range for the foreground removal.

It is important to note that the current performance is not without flaws. Some point source residues need masking, and in certain map areas, there is an absence of signal post-foreground removal. Regions with no signal (value $= 0$) and those with clear point source residue (value > 0.5 K) were masked and the 2D power spectrum for each frequency range was calculated. However, it must be acknowledged that the pseudo-power spectrum should have been computed using this mask. Given that the final results are several times lower than expected across most bins, correcting this issue is anticipated to improve the outcomes.

3.4.3 Future plans

Looking ahead, the methodology can be refined in three key areas. First, the network architecture can be enhanced to incorporate the PSF during training. Secondly, pre-training strategies can be improved to eliminate bright point sources or strong Galactic foregrounds more effectively, along with adopting more robust measures for evaluating training outcomes beyond the L2 norm. Lastly, the training set can be advanced to encompass more realistic effects, thereby optimizing the training process.

3.5 Foregrounds-FRIENDS

The Foregrounds-FRIENDS team is composed by a mixture of experts in cosmic microwave background (CMB) science and in radio astronomy. The approach combines methods developed within the framework of CMB research into the emerging field of intensity mapping. The methodology can be decomposed in three steps: point source detection, foreground removal algorithms, and power spectrum estimation.

3.5.1 Point source detection

The initial idea was to translate the expertise of part of the team in point source detection and mask creation for CMB experiments like *Planck* or Q-U-I JOint TEnerife (QUIJOTE, see e.g. Herranz et al. 2023) to this challenge. For that purpose, point source catalogues were created to use them in the foreground removal step.

The point source detection procedure is a two-step process. First, lists of candidates for every frequency in the data cube were created using a well-established source detection software, SExtractor (Bertin & Arnouts 1996). Second, convolutional neural networks (CNNs) were used to assess the reliability for each candidate. The neural network structure and detection thresholds were adjusted to accommodate the unique conditions of the challenge, and a Mexican Hat wavelet kernel was used to remove the significant diffuse emission present in the images to improve the reliability assessment. This image segmentation problem aims at creating a cube of masks tagging pixels corresponding to point sources with a 1 and background pixels with a 0. To tackle this problem, a CNN with an autoencoder architecture was implemented, divided into two main components: the encoder and the decoder.

Autoencoder architecture. The encoder consists of three main convolutional layers. The first convolutional layer receives an image with a specific number of channels (*input_channels*), applies a convolution with a 3×3 kernel, generating an output with *num_filters1* channels, while padding ensures the spatial dimensions are maintained. This process is repeated in the second and third layers with *num_filters2* and *out_channels3* channels, respectively. After each convolution, the Rectified Linear Unit (ReLU) activation function is applied, followed

by instance normalization and a MaxPool2d layer, halving the spatial dimensions. Finally, a dropout layer is included to prevent overfitting.

The decoder is responsible for reconstructing the original image from the encoded representation, using a reverse process to that of the encoder with transposed convolution layers. The final layer applies the sigmoid activation function to ensure that the output values are between 0 and 1, indicating the probability of a point source's presence.

Training and optimization. The network was trained for 65 epochs using a set of 500 images, split into 80 per cent for training and 20 per cent for testing. Adam optimizer was used with an initial learning rate of 0.001.

This methodology allowed for effective detection of point sources, facilitating foreground removal, and enhancing the exploration capabilities of the background radiation. As a result of this process, two additional products were generated: a single point source catalogue consolidating data from all frequencies, and a modified data cube where all detected point sources have been masked, facilitating the subsequent component separation of the Galactic diffuse emission and the cosmological signal.

The point source mask was not used in either of the two foreground removal algorithms explored. However, it is included in this paper as it was a significant contribution from the team.

3.5.2 *Foreground removal algorithms*

Two different foreground removal algorithms were explored aiming to use both available data sets: cubes and visibilities.

The first one is based on PCA applied to the data cubes (Irfan & Bull 2021). PCA is based on the eigenvalue and eigenvector decomposition of the frequency–frequency covariance matrix estimated using the data cubes. The modes with the highest variance were kept and subtracted from the original cube. These modes contain most of the foregrounds, so by removing them one is ideally left with the 21 cm signal. This analysis only uses frequency information and their correlation, not the spatial information of the data cube.

A four component PCA was applied to the data cube, which seemed to be optimal from simulations. The original data cube is 8×8 deg FoV, but in order to limit the noise, only the central 4×4 deg are used, as required by the power spectrum estimation.

The second approach is based on polynomial fitting of the complex visibilities, following the procedure proposed in Ghara et al. (2017). This method considers the total visibilities $\mathcal{V}(\boldsymbol{U}, \nu)$, and separately fits the real and imaginary parts as a function of the frequency ν for each Fourier space point $\boldsymbol{U}$. The fit is performed using a third-order polynomial in logarithmic space. The fitted model is then removed from the total visibilities to get the residual visibilities. The real-space maps are obtained by performing the inverse Fourier transform of the residual visibilities at each frequency channel.

After comparing the performance of the two methods, the PCA approach was adopted for this challenge.

3.5.3 *Power spectrum estimation*

The cylindrical power spectrum has been estimated with the PYTHON package `ps_eor`.[16] In the workflow, images from previous steps are converted to brightness temperature and the PSF is deconvolved. Appropriate bins that conform to the required submission format are determined to calculate the power spectrum, correcting for the primary beam. Noise-subtraction techniques have not been applied.

[16] https://gitlab.com/flomertens/ps_eor

3.5.4 *Workflow and reproducibility*

The Foreground-FRIENDS workflow aims to provide a modular and flexible implementation. Each task in the workflow (except the point source detection) is implemented either as a PYTHON script, a JUPYTER Notebook or a Bash script. A generic and reusable PYTHON script launches all the tasks sequentially. The selection and order of the tasks can easily be modified through a configuration YAML file. Each task, or step, is stored individually in a separate directory containing relevant documentation explaining the purpose and methodology followed, algorithms used and parameters of each task. For each task, the contents were divided in PYTHON scripts with the relevant functions, parameters files independent from the code, and notebooks to execute the functions and visualize the results, when possible. Therefore, the aim is to isolate the algorithms, parameters, and execution, thus facilitating reusability.

Software provenance is fixed using `conda` through an explicit `conda` environment. The repository (see Table 5) contains license and citation information, as well as a README file with installation instructions, execution instructions, a description of implemented steps and a diagram of the file structure. For persistence and findability, the workflow is stored in the long-term general archive Zenodo.

3.6 HAMSTER

The HAMSTER team developed a minimalist yet innovative approach to explore alternative methods for foreground mitigation without cleaning the foregrounds in the image cube. The aim was to sufficiently mitigate wide-field effects in the visibility data, and perform direct foreground avoidance for measuring the power spectrum (Morales et al. 2012). Therefore, the entire data analysis is performed in delay space without using the images. The analysis can be divided into two steps, which are flagging in delay space and coherent averaging for power spectrum estimation. The method is described below.

3.6.1 *Flagging in delay space*

The visibility data can be Fourier transformed along the frequency direction into delay space. The foreground emission mostly resides within the wedge defined by the primary beam FoV (Liu, Parsons & Trott 2014). However, wide-field effects such as bright sources in beam sidelobes leak into higher delay and prohibit measurements of 21 cm signal (e.g. Pober et al. 2016). Therefore the contamination at high delay is identified and excluded first.

For each antenna pair, the visibility data are Fourier transformed. Delay power spectrum is then calculated as illustrated in Fig. 4. Excluding the region around delay $\eta = 0$ where foregrounds reside, the power spectrum in the rest of the time-delay space is used to perform iterative thresholding. The thresholding excludes 5σ of the sampling variance iteratively to find the average amplitude of the delay power, which is thermal noise dominated. The excluded delay modes, as shown in Fig. 4, have clear, stripe-like structure that is continuous in time and delay. Based on the position of excluded modes in the time-delay plane, a morphological closing of the flags was performed to identify the stripes in each antenna pair. The stripes

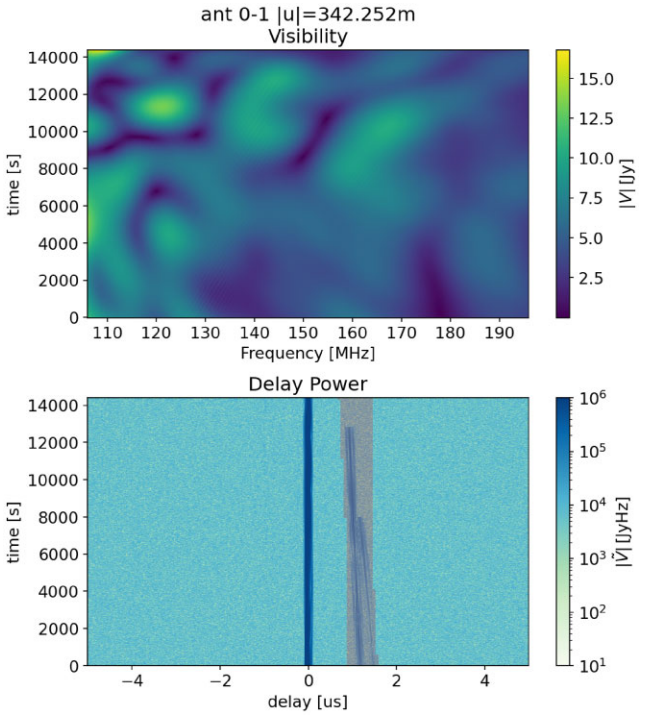

Figure 4. Illustration of the flagging method in delay space from the HAMSTER team. The top panel shows the amplitude of the visibility data for antenna pair 0 and 1. The bottom panel shows the amplitude of the delay-transformed visibility data for the same pair. Two visible stripes of excess power can be seen around the delay $\eta = 1\,\mu$s. The shaded region shows the delay modes that are excluded from the power spectrum estimation.

are then extended along the direction of time shown as the shaded region in Fig. 4.

3.6.2 *Power spectrum estimation*

The flagged, delay-transformed visibility data are averaged into u–v grids for power spectrum estimation. The visibilities are first averaged every five channels to match the maximum $k_\parallel$ required. A Blackman–Harris taper is then applied to perform the delay transform. The visibilities are then averaged into the grids with flagged delay modes excluded. We choose the grid length to be $3\,\lambda$. The data set is divided into two subset for cross-correlation to remove the noise floor (see e.g. Abdurashidova et al. 2022; Paul et al. 2023). The gridding is performed for the whole band for quality assessment, and then for each of the sub-bands required by the data challenge. The flags in delay space are downsampled to match the delay resolution of frequency sub-bands.

The 3D power spectrum is then calculated as the square of the modulus of the delay-transformed gridded visibilities. Conversion to temperature units is performed with volume renormalization based on the power-squared beam area (Parsons et al. 2014). The resulting power spectrum is shown in Fig. 5. For the whole band, the power spectrum is at $\sim 10^{-3} - 10^{-2}\,\mathrm{K^2Mpc^3}$, suggesting that signal level is being reached. However, the cylindrical power spectrum shows residual contamination exists when the same estimation is performed for each of the six sub-bands, when the measured power is at $\sim 0.1\,\mathrm{K^2Mpc^3}$. This could be due to the frequency taper not being able to suppress frequency-dependent gain errors injected at sub-band level.

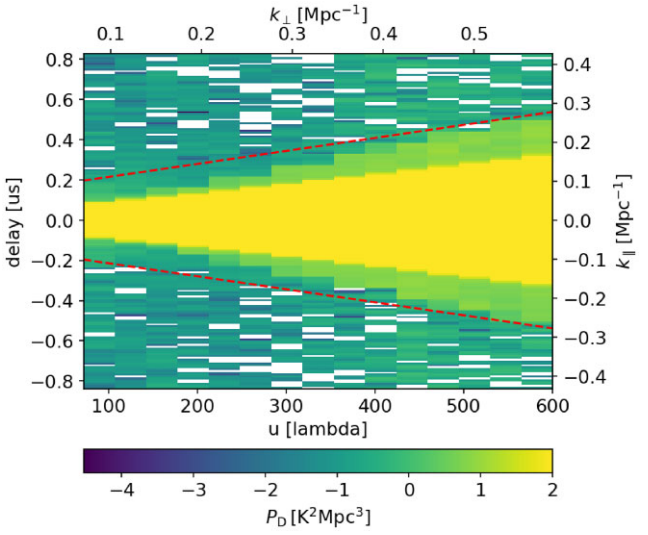

Figure 5. The estimated cylindrical power spectrum of the whole band from the HAMSTER team. The red dashed line shows the region where foregrounds dominate.

3.6.3 *Outlook*

The team plans to further develop on this method by adopting Bayesian statistics in the flagging of contamination in delay, coherently average visibilities in delay space, and power spectrum estimation. Specifically, the initial flagging and averaging can be used to calculate angular scale-dependent data covariance. The covariance can then be used to inpaint the flagged delay modes and to repeat the averaging and power spectrum estimation. This can be performed iteratively so that the data covariance for inpainting and the subsequent power spectrum estimation converge.

Foreground removal can be additionally added to the data analysis. The plan is to test frequency covariance-based cleaning such as PCA and GPR at different stages of the data analysis. This may further be implemented iteratively, where 21 cm signal covariance and foreground covariance can be separately estimated.

3.7 Hausos

The team was composed of data scientists and cosmologists based in both China and France. With the experience in sparsity-based source separation methodology, the team primarily intended to further extend the previously published Deconvolved Generalized Morphological Component Analysis (DecGMCA, Jiang, Bobin & Starck 2017), starting from visibility data, to the SDC3a data challenge. Due to the various gridding strategies and large computation costs, a deep learning-based method was finally adopted. Since the classical CNN is limited by its receptive field, a transformer-based neural network based on SEgmentation TRansformer (SETR, Zheng et al. 2021) was employed. SETR is a model that combines transformer and CNN to achieve semantic segmentation tasks. With the transformer as the encoder, global contextual features are modelled in each encoder layer, which can effectively capture long-distance features. The decoder is a CNN-based model to upsample the advanced features extracted by the encoder to the original input resolution for pixel-level prediction.

Training. The training was performed on EoR simulations provided by the Shanghai Astronomical Observatory, based on the 21CMFAST simulations (Mesinger et al. 2011; Murray et al. 2020)

and OSKAR using the same observation parameters. The training data were divided into six frequency bins, with the same convention adopted for the challenge data. In the training stage, the dirty maps and the ground-truth EoR maps were cut into patches of the 256 $\times$ 256 pixels, and the network was trained with 100 epochs from dirty patches to the corresponding ground-truth EoR patches. The optimizer was Adam, the learning rate was set to 10^{-5} and the binary cross entropy was chosen as the loss function.

Prediction. Given a frequency bin, the released natural weighting images, cut into 256×256 patches, were used as the input of the trained network, and the EoR patches were predicted for the frequency bin. Then the patches were concatenated to form the EoR images. Repeating the procedure for the six frequency bins finally yielded the EoR image cubes. During the data challenge, it was observed that the network did not perform well on the negative values due to the activation functions and the amplitude of the estimated power spectrum being biased. One reason may lie in the PSF not being modelled well enough, causing deconvolution problems. Thus, the future work will focus on incorporating the PSF modelling in the network.

3.8 HIMALAYA

The team name HIMALAYA connotes that measuring the EoR 21 cm signal is as challenging in astronomy as climbing Mount Everest. To estimate the EoR power spectra from the released SDC3a data, the team proposed a new foreground-cleaning algorithm called the 'reconvolution' method. This method involves three key steps: (1) reconvolution pre-processing of dirty maps (natural weighting) to recover the smoothness along frequency; (2) applying the PCA technique to the reconvolved dirty image; and (3) using the power spectrum correction to adjust the amplitude of the measured H I power spectrum at different k bins. The steps are summarized as follows.

3.8.1 Reconvolution

The aim of the reconvolution pre-processing is to greatly suppress the sidelobe of the PSF and reduce the so-called mode-mixing effect in interferometric observations. Without any instrumental effects, the foreground signal exhibits a very high degree of smoothness along frequency, which makes it easy to separate from the oscillating 21 cm signal. Traditional techniques such as PCA or polynomial fitting can reliably recover the 21 cm signal. However, due to incomplete uv-coverage and the mode-mixing effect from long baselines, the dirty image becomes complex and exhibits hard-to-model frequency dependence. Consequently, the measured foreground presents highly non-smooth features. The usual deconvolution methods, such as CLEAN, do not guarantee the smoothness of the sky map, potentially leading to inaccuracies in separating the 21 cm signal. Moreover, due to numerical effects, any deconvolution attempt aimed at dividing the PSF over the Fourier modes results in unavoidable artefacts that prevent the 21 cm signal recovery.

Considering the required k range of about $k < 0.5\ h\mathrm{Mpc}^{-1}$ for SDC3a, detailed measurements on small scales are not necessary in principle. Therefore, a natural way is to convolve an appropriate beam kernel, $w(r)$, for each frequency of the image. This process significantly suppresses the measured visibilities from the long baselines, while maintaining a spatial resolution sufficient to capture the Fourier modes within the desired k range. In other words, following the convolution theorem, one can design a weighting scheme for the visibilities in the Fourier domain, aiming to suppress the high-k Fourier modes while keeping the noise from being significantly amplified. After some testing, the SDC3a PSF resulted the easiest choice for the beam kernel. This choice may not be optimal, and other options, such as using a Gaussian beam kernel, are also possible. Further optimization to improve the signal-to-noise ratio, more effective removal of foreground contaminants, etc., need to be investigated through simulation.

For SDC3a, there are six frequency bins, each containing 151 frequency channels. For each frequency bin, by convolving the PSF of the first channel with the 151 dirty maps, we found that the resulting reconvolved dirty image became highly smooth along frequency. Although the small-scale structures of the images are suppressed, the frequency smoothness is recovered, which is essential for the foreground removal.

3.8.2 PCA

In the absence of instrumental effects, the eigenvalues of the frequency–frequency covariance of the sky image are expected to exponentially decrease. Typically, removing the first three eigenmodes by the PCA technique can effectively subtract most of the foreground contributions, leaving only a small foreground residual compared to the 21 cm signal level. However, the eigenvalues of the raw dirty images decrease slowly, making the PCA technique ineffective. Even after subtracting approximately 120 modes in an SDC3a dirty map (given that we only have 151 modes in each frequency bin), the residual image still had a signal amplitude too high compared to that of the expected 21 cm signal.

Surprisingly, after applying the reconvolution pre-processing, the distribution of eigenvalues becomes similar to that of the instrumental-free case. The total foreground contamination can be effectively cleaned by discarding the first 20–30 eigenmodes of the reconvolved dirty image, and the residual image pattern could reflect the correct cosmological H I pattern. After PCA, the 2D cylindrical power spectrum of the resulting image, $\tilde{P}(k_\perp, k_\parallel)$, can be estimated by averaging over the corresponding Fourier modes falling into the a given $(k_\perp, k_\parallel)$ shell. The team used their own code to do this estimation.

3.8.3 Power spectrum correction

There are four major factors that can alter the amplitude of the estimated power spectrum: primary beam, PSF, reconvolution pre-processing, and 21 cm signal loss due to the PCA subtraction. The algorithm did not correct for these effects at the map level, but at the power spectrum level, by defining a transfer function $T(k_\perp, k_\parallel) = \left\langle P(k_\perp, k_\parallel)/\tilde{P}(k_\perp, k_\parallel) \right\rangle$, where $P(k_\perp, k_\parallel)$ is the final submission. The noise bias is neglected in the analysis as it appears to be small compared to the 21 cm signal. Then, the final reconstructed H I power spectrum for SDC3a is calculated using $P(k_\perp, k_\parallel) = T(k_\perp, k_\parallel)\tilde{P}(k_\perp, k_\parallel)$. The estimation of $T(k)$ is derived from the average value of the power-spectrum ratios obtained from mock data, produced as described below. In addition, to correctly estimate the error bar of P, one can also use the mock-derived standard deviation of $T(k_\perp, k_\parallel)$, $\Delta T(k_\perp, k_\parallel)$, to propagate all statistical and systematic uncertainties i.e. $\Delta P(k_\perp, k_\parallel) = \Delta T(k_\perp, k_\parallel)\tilde{P}(k_\perp, k_\parallel)$.

The transfer function does not explicitly rely on the precise shape of the input power spectrum (at least to good approximation), as it only depends on the ratio of the input to the output. With this in mind, the 2D cylindrically averaged H I power spectrum from

the SDC3a test data set (incorporating EoR, noise, and instrumental effects across the 166–181 MHz range) was utilized. Subsequently, 50 Gaussian realizations from this spectrum were generated to create a set of H I sky maps. These maps were added to simulated foreground maps (inclusive of instrumental effects), to obtain 50 mock images. For each mock image at a given frequency, reconvolution, PCA subtraction, and power spectrum estimation were performed using the same procedure used in processing the SDC3a data, thus getting $\tilde{P}(k_\perp, k_\parallel)$. Then, the transfer function and its standard deviation were estimated from the ratios of the true H I spectrum (provided with the test data set) and the spectra derived from the mocks. By using those transfer functions, P and its associated statistical errors for SDC3a were finally determined.

3.9 KORSDC

The team is composed of members of the Korea SKA consortium (Korea SKA), which was established about a decade ago for seeking the governmental participation in the SKA project. The team used commonly known schemes on source subtraction, image smoothing, and spectral analysis, except when they were not sufficient to meet the requirements.

The starting point in this analysis was mostly the natural-weighted image cube, that is believed to preserve the large-scale information better than the uniform-weighted one. The main pipeline is composed of the following sequential steps:

(i) Removal of point sources.
(ii) Mitigation of the impact of the PSF, the primary beam, and the noise.
(iii) Removal of the foregrounds.
(iv) Calculation of the power spectrum.

Throughout the pipeline, the central $3 \times 3\,\mathrm{deg}^2$ out of the given $9 \times 9\,\mathrm{deg}^2$ were used to mitigate the impact of the noise, which is amplified after the primary beam correction as one moves away from the field centre. The following subsections describe in details of each step. It is worth noting that is was necessary to calibrate the amplitude of the power spectrum by means of a 'fudge factor' (Section 3.9.5). Overall, the impact of the PSF seemed to dominate the image even after the point-source and foreground removal.

3.9.1 Removal of point sources

The pipeline used the source extraction code SExtractor[17] to detect and remove point sources from the image data cube. In each frequency channel, the SExtractor parameter `BACK SIZE` needs to be defined to determine the background value for each pixel. Following Bertin & Arnouts (1996), a new background value for each pixel was defined based on the pixels included in `BACK SIZE`, as follows:

Non-crowded case = mean
Crowded case = 2.5 × median—1.5 × mean

According to Bertin & Arnouts (1996), for crowded regions, the average of the pixel values included in `BACK SIZE` is iterated until it reaches 3σ. If at each iteration σ decreases by 20 per cent or less, the region is considered non-crowded. Therefore, to detect point sources among irregular and extended signals, it is necessary to consider the overall brightness of extended sources close to point sources. To achieve this, the number of pixels included in `BACK SIZE` was set to 10, and the newly defined background value worked as a photometric zero-point for point sources. After separating point sources from irregular and extended sources and conducting visual inspection on residuals, all frequency channels were combined to recreate the 3D data cube.

3.9.2 Image into the brightness temperature/primary beam correction

The provided image cube was in Jy beam^{-1}. Assuming a 2D Gaussian shape with axes of 0.05 and 0.04 deg for the synthesized beam, the given flux density was divided by the beam area to obtain the specific intensity I_ν, which was then converted to the brightness temperature in the Rayleigh–Jeans regime

$$T_b = \frac{c^2}{2\nu^2 k_B} I_\nu \tag{1}$$

using routines from the `ps_eor`[18] package.

The primary beam correction was unavoidable as a substantial part of the full FoV was used. The image data plane was divided by the time-averaged beam provided at each frequency band.

3.9.3 PSF and thermal noise

To mitigate the impact of the PSF, the original strategy was to perform a 'deconvolution' the Fourier (k-) plane using the convolution theorem following equation (4) of Chen et al. (2023). However, this naive approach did not fully remove the PSF but instead left a residual fluctuation which was substantial enough to impact the resulting image and the power spectrum significantly. There was also a problem with understanding the normalization convention of the PSF. Specifically, summation of the pixel values of image cube for each frequency within the cut $3 \times 3\,\mathrm{deg}^2$ area was often negative. This problem was entangled with the thermal noise correction as well, as the noise is greatly amplified by the deconvolution. In retrospect, a better approach could have been to have smoothed the PSF itself to tame the sidelobes before the deconvolution.

In the end, it was decided to simply smooth the image to reduce the impact of the PSF and the thermal noise with a frequency-dependent smoothing kernel. The test data set was used to find an optimal choice of the smoothing kernel. By trial and error, a 2D Gaussian filter was adopted, with the smoothing angle $\sigma = 48$ arcsec at the highest frequency band (181 MHz), with the kernel in the form $\propto \exp[-\theta^2/2\sigma^2]$. The data were smoothed at each frequency band by scaling σ in proportion to the frequency inverse. This step was implemented with the multidimensional image processing packages in SciPy (Virtanen et al. 2020).

3.9.4 Foreground removal

The pipeline applied ICA on each line of sight (LoS) to the image obtained as described above, which contains diffuse foreground emission, the EoR signal, and the mitigated PSF and noise. LoSs were just treated independently, without using the intrinsic angular correlation of the diffuse foreground. Three independent foreground components were assumed, and one extra component to hopefully capture the unwanted spectral features introduced by the PSF and the noise. The four components were fitted out from each LoS

[17] https://github.com/astromatic/sextractor

[18] https://gitlab.com/flomertens/ps_eor/-/blob/master/ps_eor/datacube.py

to obtain the residual image cube. The FastICA implementation in the SCIKIT-LEARN package was used for ICA (Pedregosa et al. 2011). Unfortunately, many LoSs showed highly irregular frequency spectra even after the data had been processed (Sections 3.9.1–3.9.3), indicating that the ICA was struggling in separating the EoR signal.

3.9.5 Power spectrum estimation

The resulting image cube was first divided into six subcubes with the 15 MHz frequency interval adopted for the challenge. Then, the angular coordinates (RA, Dec.) of each subcube were converted into comoving coordinates using the specified cosmological parameters. This coarsened data cube was Fourier-transformed, and the power for each $(k_\parallel, k_\perp)$ bin was collected.

Note that a PSF normalization fudge factor of 1/3 was multiplied to the resultant power spectrum to get the final estimation. This fudge factor was obtained from the test data set as follows. The primary beam correction, angular smoothing, and the ICA with only one component (reserved for the PSF effect on an LoS spectrum) were applied to the test data set, which is free from point sources and the diffuse foreground. The power spectrum obtained after fitting out the spectra of LoSs was found to be about three times as high as the true power spectrum in amplitude but otherwise similar in the overall structure. It was concluded that the PSF effect on the actual data cube would be of the same level even after the inclusion of the point sources and the foreground.

The final result for the KORSDC team is found to be about two orders of magnitude larger than true value, indicating that the point sources and the diffuse foregrounds, combined with the PSF residual, have not been mitigated to the intended level, and still dominate the power spectrum.

3.10 KUSANAGI

KUSANAGI consisted of members of the EoR Science Working Group and Engineering Working Group in the SKA Japan (SKA-jp) consortium. This team's approach can be broken down into four parts; (1) creating a source list, (2) removing radio sources from the visibilities, (3) statistical foreground removal in image space, and (4) power spectrum estimation. Each part is described briefly below. Note that the pipeline gathered some publicly available software such as CASA, GPR4IM (Soares et al. 2022), PYPBSF (Mohan & Rafferty 2015), and WSCLEAN (Offringa et al. 2014). For other processes including point source subtraction, statistical foreground removal, and power spectrum estimation, the software was built from scratch using common PYTHON packages (e.g. NUMPY and ASTROPY).

For creating source lists used for point source removal, the pipeline started by converting the visibilities (MS format) to images with WSCLEAN using only 2 MHz of data centred at each 15 MHz band. The effect of the SKA station beam, which has significant frequency dependence, was ignored. To reduce the diffuse emission components, the baselines were limited to the range from 1000λ to 5000λ. The radio source identification was performed using PYPBSF on the CLEANed images (fits format) and made six different source lists at each 15 MHz interval.

The source lists were used to make modelled visibilities at each channel based on the information of the original SKA visibilities (UVFITS format) and SKA station beam model. A CUDA scheme developed in the real-time system (RTS, Mitchell et al. 2008) was employed for the modelling. The spectral index of radio sources was set to −0.7. The modelled visibilities were directly subtracted from the SKA visibilities. The residual visibilities data are converted to MS format using CASA. Images were made with WSCLEAN from the residual visibilities at each channel. At this stage, the natural weighted dirty map were made using a limited range of baseline from 30λ to 250λ. The images consist of image cubes of 512 × 512 × 150 grids at each 15 MHz. Two sets of images were made at even and odd time samples. Thus, there were 12 image cubes (6 bands × 2 time-steps). The images were dominated by diffuse emission and some residual bright sources.

Two different statistical foreground removal methods were then applied in the image space. One was the traditional fitting with a third-order polynomial function using the 'polyfit' method in NUMPY. The spectrum at each pixel along the frequency axis was simply fitted for each image cube. After subtracting the fitted values, the residual image cubes were available for power spectrum estimation. Another method was based on GPR in the image space. There were three covariance kernels (RBF, Matern3/2, and exponential). The hyperparameters were optimized by a GPy implemented optimizer. The mean value of the foreground was estimated by using RBF and Merten3/2 kernels at each pixel along the frequency axis and was removed from the image cubes. The results of the polynomial fitting and the GPR were referred to as KUSANAGI-poly and KUSANAGI-GPR, respectively.

Finally, the power spectrum was estimated from the residual image cubes. The procedure performed here was motivated by the method described in Mertens et al. (2020). The image cubes were converted to gridded visibilities cubes by the Fast Fourier Transform (FFT). The Blackman–Harris window function was applied along the frequency axis before the Fourier transform. The cross-power spectrum was evaluated using two different time-step data at each Fourier mode. The median of cylindrical power spectrum value was evaluated at each $k_\perp$ and $k_\parallel$. To match the SDC3a required format, values were averaged in wider k-bins (0.05 hMpc^{-1}). The variance of samples σ^2 and the number of samples (N) were also calculated in each wider k-bin. The quantity $\sigma/\sqrt{\mathrm{N}}$ was submitted as the error on each bin of the power spectrum. Even after the foreground removal, the residual had significant power, especially at the foreground wedge (Morales et al. 2012). Following the foreground avoidance approach, the k-bin in the foreground wedge were removed assuming the maximum contamination from 10 deg from the field centre.

3.11 Nottingham–Imperial

The pipeline started from the visibility files and used the OSKAR visibilities simulation package[19] to remove the bright point sources from the data set. This was done by aligning the brightest point sources with the GLEAM catalogue and using this bright point source map to feed into OSKAR to remove them.

Using the RA, Dec, and minimum flux of the point sources (100 mJy at 151 MHz) located within the data cubes, the GLEAM catalogue (Hurley-Walker et al. 2017) was queried in VizieR[20] to extract an approximate catalogue of corresponding point sources, which was then combined into an OSKAR sky model. OSKAR was used to simulate visibilities over the entire frequency range, in steps of $\mathrm{d}\nu = 0.1$ MHz. Each simulated observation was centred on RA $= 0°$ and Dec $= -30°$, had a total length of 4 h, a 10 s integration time and used the provided telescope model. These simulated visibilities were directly subtracted from the data challenge visibilities to produce a

[19] https://github.com/OxfordSKA/OSKAR
[20] https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=VIII/100/gleamegc

set of point-source-subtracted visibilities, which formed the basis for the rest of the analysis.

The next step applied the FastICA blind source separation technique to remove the remaining diffuse foregrounds (Chapman et al. 2012). This technique requires careful choice of the number of components and the frequency bins over which to apply FastICA, in order to model diffuse foregrounds while avoiding over fitting. After application of FastICA to obtain diffuse foreground subtracted residuals, Pearson correlation coefficients were used to compare the similarity of the residual data cube to the foreground model and the bright source removed data cube. The minimum correlation was found by first varying the size of the frequency cube and then the number of components. It was found a binning of 7.5 MHz allowed the best foreground removal and easily allowed the power spectrum to be made in 15 MHz bins, as was the required output. Different component numbers, ranging from 4–8 components, were obtained for each frequency bin in this way.

For the submitted analysis, the bright source subtracted data cube was first cut to be 1024 × 1024 instead of the full 2048 × 2048 so only the region within the primary beam was present. This was intended to negate some of the sidelobe effects and improve FastICA performance. The data was then split into 7.5 MHz sections and each section was put through FastICA using the optimum component number. The option of imposing a fixed resolution at all frequencies was also explored, by convolving the frequency sections with a beam of fixed resolution, before applying FastICA. This has been suggested as a way of allowing FastICA to better identify foreground structure. This did not appear to give significant improvement, so ultimately it was chosen not to sacrifice resolution in this way.

After converting the residual data cubes from Jy beam^{-1} to Kelvin, using standard relations, the TOOLS21CM[21] PYTHON package was used to calculate the 2D power spectrum for each 7.5 MHz section of data (12 total with corresponding error 2D power spectrum). The $k_{\parallel}$ and $k_{\perp}$ bins given by the SKA were fed directly into the calculation for the power spectrum. Uncertainties were calculated based on a thermal noise model converted to a 2D power spectrum added to an estimate of sample variance.

There is a plan to further develop this method by optimizing the selection of the number of FastICA components in an automated way for blind data where the truth is not known. Much of the challenge for this analysis lay in accurate removal of point sources.

3.12 Pisano Galaxy Moppers

This team utilized the foreground avoidance method in a 21 cm inference setting to 'recover' the underlying cosmic signal. 21CMMC (Park et al. 2019) was used to generate 21 cm light-cones and perform the Bayesian inference, while the data reduction and likelihood computation was done on-the-fly using `py21cmmc-fg`, the plug-in code to 21CMMC. To maximize the time spent on sampling the parameter space, several assumptions and simplifications were made:

(i) In both data and model, instead of gridding the visibilities, the delay power spectrum was calculated. The conversion to power spectrum was made once the final mean and uncertainty for the frequency bins were calculated.

(ii) The visibility data were split into odd/even time-steps, the cross-(delay) power was calculated between them, and all power in each frequency bin was then averaged.

[21] https://tools21cm.readthedocs.io/

(iii) Only the sky over the central 4 × 4 deg FoV was modelled. To generate the sky model, steps similar to those used to generate the SDC3a data described in Section 2.2 were undertaken. Instead of T-RECS, a power-law relation with parameters from Intema et al. (2011) was used for the faint extragalactic source population.

3.13 REACTOR

This team have developed an innovative algorithm for deconvolution of astronomical images, namely `PI-AstroDeconv`, integrating physical information with semisupervised learning techniques (Ni et al. 2024). Its design philosophy has been further developed based on previous research achievements (Ni et al. 2022).

Beam removal operations on the astronomical images are first performed; subsequently, PCA is applied to the images for foreground subtraction. Next, the `astropy.FlatLambdaCDM` and `scipy.states` libraries were utilized to calculate the dimensions (dims) of the 3D cube and the average value within each 2D region after binning, respectively.

3.13.1 Beam effects on foreground removal

In previous research by this team, it was noted that intricate beam effects had a considerable impact on the outcomes of PCA (Ni et al. 2022). To emulate full-sky neutral hydrogen observational data, the Cosmological Realizations for Intensity Mapping Experiments (CRIME) simulator was used, superimposing multiple components before convolving them with two distinct beam types: an ideal Gaussian beam and the cosine beam employed by the MeerKAT telescope (Matshawule et al. 2021). The results demonstrated that PCA could successfully remove foreground components affected by the Gaussian beam, yet it struggled to eradicate those influenced by the cosine beam. Nevertheless, the integration of the UNet neural network significantly enhanced performance.

Based on the aforementioned conclusions, further validation was conducted. The UNet model was first applied to the observational data for deconvolution to counteract the beam smearing effect, with the input being the sum of the convolved foreground and signal, and the label being the sum of the original, unconvolved foreground, and signal. Following this, PCA was sued to process the data and eliminate the foreground components. The results indicated that this method remains effective. It can be represented by the following formula:

$$\begin{aligned} &\mathrm{UNet}\{\mathrm{PCA_{res}}[\mathrm{Beam_{sm}}(\mathrm{FG}+\mathrm{HI})]\} \approx \mathrm{HI} \Leftrightarrow \\ &\mathrm{PCA_{res}}\{\mathrm{UNet}[\mathrm{Beam_{sm}}(\mathrm{FG}+\mathrm{HI})]\} \approx \mathrm{HI}, \end{aligned} \tag{2}$$

where $\mathrm{PCA_{res}}$ denotes the residual values after PCA processing, UNet refers to the results processed through the UNet network, $\mathrm{Beam_{sm}}$ indicates the convolutional beam operation applied to the signal, FG represents the cumulative sum of the individual component foregrounds, and H I stands for the neutral hydrogen signal.

It is evident from the aforementioned formula that the order of applying PCA followed by the UNet network or vice versa does not affect the final analytical outcome. This demonstrates that both sequences of processing are capable of effectively removing the foreground signals from the data, thus achieving the desired analytical objectives. Based on this, the team proposes a semisupervised learning algorithm that integrates physical information, aimed at mitigating the effects of beam patterns in observational data.

3.13.2 PI-AstroDeconv model

The pivotal innovation of the `PI-AstroDeconv` (Ni et al. 2024) netowrk lies in the integration of physical information from astronomical observations – specifically, the telescope PSF – into the model training process, as depicted in fig. 1 of Ni et al. (2024). Within the network, the intermediate prediction layer acts as the output layer, corresponding to the 'prediction' section shown in fig. 1 of Ni et al. (2024). The training objective of this network is to achieve complete identity between the input and output. In this context, the convolution of the 'prediction' layer with the PSF represents the observed data, while the UNet is responsible for pixel-level restoration. Thus, it can be deduced that the 'prediction' output from the intermediate layer equates to the observation image after the removal of convolutional effects. Ultimately, each frequency band was trained separately, resulting in a 3D data cube.

It is worth noting that the deconvolution was performed on the original images, which have dimensions of 2048×2048 pixels. This choice leads to a significant slowdown in the training speed due to the mathematical convolution operation added in the last layer of the network. To effectively address this issue, FFT technology was innovatively employed to redesign the mathematical convolution process, thereby significantly improving the training efficiency.

Building on this, the impact of the beam effect was further mitigated, the PCA algorithm was employed to perform foreground subtraction on the network. Through this process, six major components were removed, obtaining the residual image after foreground removal, which was then used to calculate the cylindrical (2D) power spectrum.

3.13.3 Power spectrum estimation

Following the standards set by SDC3a for simulated data description and parameter configuration, the central 512×512 region of data for each frequency band was selected for analysis. To estimate the cylindrical power spectrum, the image cube was first subdivided into six subcubes with a frequency bandwidth separation of 15 MHz. The `box_dims` size was computed using the `FlatLambdaCDM` model from the ASTROPY library, followed by applying the `binned_statistic_2d` function from the SCIPY library to perform binned statistical analysis on the 2D data. Subsequently, the data were subdivided according to the bins recommended by SDC3a and specific statistical operations were performed within each partition to generate the corresponding cylindrical power spectra. The same process was followed for the error calculation step, but the standard deviation was substituted for the mean to obtain an error estimate.

3.14 Shuimu–Tianlai

The Shuimu–Tianlai team comprises experts in cosmology, radio astronomy, and data sciences from Tsinghua University, National Astronomical Observatories, Chinese Academy of Sciences, and a number of other institutions in China. The team used a new method called oriented singular value decomposition (O-SVD; Zeng & Ng 2020) for the foreground subtraction task.

The O-SVD of a third-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ with rank $R_3 = \text{rank}_3(\mathcal{A})$ is

$$\mathcal{A} = (\mathcal{U} *_3 \mathcal{S} *_3 \mathcal{V}) \times_3 U^{(3)},$$

where the orthogonal matrix $U^{(3)} \in \mathbb{R}^{I_3 \times I_3}$, three tensors $\mathcal{U} \in \mathbb{R}^{I_1 \times I_1 \times I_3}$, $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, $\mathcal{V} \in \mathbb{R}^{I_2 \times I_2 \times I_3}$ such that

(1) $\mathcal{U}(:, :, i), \mathcal{V}(:, :, i)$ are orthogonal and $\mathcal{S}(:, :, i)$ is a non-negative diagonal matrix for $i = 1, \cdots, R_3$;

(2) $\mathcal{U}(:, :, i)$, $\mathcal{V}(:, :, i)$, and $\mathcal{S}(:, :, i)$ are all zero matrices for $i = R_3 + 1, \cdots, I_3$.

The diagonal elements s_{jji} of each frontal slice of $\mathcal{S}$ are called the singular values of the pair $(\mathcal{A}, \mathcal{S})$.

The given image cube can be taken as a third-order tensor, therefore the O-SVD method can be applied. Similar to the matrix SVD method, all the singular values can be arranged in descending order, and the largest SVD modes can be truncated as foregrounds to be subtracted from the data.

The data used are the uniform-weighted image cube. The `radio_beam`[22] package was first used to convert the unit of the data from Jy beam^{-1} to K with the BMAJ and BMIN in the FITS file header, and the `FlatLambdaCDM` model in `astropy.cosmology`[23] was used to convert the angle and frequency scale to physical size in unit of Mpc. Such an image cube can then be foreground subtracted with the O-SVD method and used for cylindrical 2D power spectrum estimation. The code in TOOLS21CM[24] was used to compute the cylindrical power spectrum, and the 1σ error of the power spectrum was given by the standard deviation of the values that fall into a specific $(k_\perp, k_\parallel)$ bin. The technique of transfer function was not used to compensate for signal loss by the foreground subtraction method, and possibly as a result the power in small k-modes of the submitted results is lower than the true power, indicating there is indeed some signal loss at large scales. The complete code used for this analysis was open-sourced in zenodo[25] and Github.[26]

One of the primary challenges in implementing the method lies in determining the optimal number of foreground modes to be truncated, as the singular values plot does not exhibit an obvious gap for reference. Currently, this number is determined through visual comparison of the residual images with test images that lack foreground. As an improvement over this approach, the team plans to explore model selection techniques such as the Akaike information criterion, the Bayes information criterion, generalized cross-validation, among others, to derive a more objective and reliable criterion.

3.15 SKACH

The team comprised cosmologists, radio interferometry experts and data scientists members of the SKA Switzerland consortium (SKACH.[27]). Their approach consisted of three pre-processing steps followed by a deep-learning framework that recovers the 21 cm signal from tomographic data with residual foreground contamination and noise.

For the pre-processing, the composite GLEAM and LoBES catalogue were first employed to remove the brightest extragalactic compact sources. The corresponding raw visibilities, V_{exgf}, were computed with OSKAR, then subtracted from the data challenge visibilities, V_{SDC3a}, to obtain $V_{\text{res1}} = V_{\text{SDC3a}} - V_{\text{exgf}}$. Once they were removed, dirty images were created at each observed frequency

[22] https://pypi.org/project/radio-beam/
[23] https://docs.astropy.org/en/stable/cosmology/index.html
[24] https://github.com/sambit-giri/tools21cm/blob/master/src/tools21cm/power_spectrum.py
[25] https://zenodo.org/records/10124117
[26] https://github.com/zuoshifan/ska_sdc3a_pipeline
[27] https://skach.org

with WSCLEAN.[28] PYPBSF was employed on each image to find faint sources in the sky, and their flux distribution was modelled with an ellipsoid. The resulting catalogue was used to compute the raw visibilities to be subtracted from the previous step $V_{\rm res2} = V_{\rm res1} - V_{\rm PyPBSF}$. This process was repeated until reaching the flux limit for unresolved sources of 100 mJy. To finalize the first pre-processing step, *uv*-sampling filtering for points with $U > 200\,\lambda_{\rm obs}$, where $\lambda_{\rm obs}$ correspond to the observed wavelength, was applied to $V_{\rm res2}$. The resulting visbilities were finally used to create a tomographic dirty image.

The second step consists of a spatial-domain decomposition of the previously mentioned tomographic data. For each pixel in the image, a polynomial function was fit along the LoS to remove the large contribution of the residual foreground emission, thus producing a dirty image, $I_{\rm res}$, with shape 1024^2 and angular resolution of 14 arcsec in which ideally most of the foreground contamination is removed.

An U-shaped CNN based on a modified version of the `SegU-Net` framework presented by Bianco et al. (2021, 2024) was then employed. This network was meant for binary segmentation of neutral and ionized regions from 2D tomographic images of the 21 cm signal during the late stage of reionization, $z < 11$ ($\nu_{\rm obs} > 118$ MHz). The original architecture was modified to process the tomographic data, and it is similar, but for 3D inputs with shape (128, 128, 16), to the `RecU-Net` architecture presented in Bianco et al. (2025).

A data set of approximately 20 simulations with different astrophysical parameters and initial conditions was created to train and test the network. The simulated pipeline employed the 21CMFAST seminumerical code (Mesinger et al. 2011) to simulate the 21 cm signal and followed the description of the challenge to create a mock observation. The simulation of the Galactic synchrotron foreground signal followed the method in Choudhuri et al. (2014) and employed the OSKAR code (Dulwich et al. 2019) to simulate the raw visibilities for a 4-h observation, including the ionospheric effect with the ARATMOSPY (Srinath et al. 2015) code. The same time-evolving phase screen was applied for all the data. Additionally, systematic noise was included for the same observational length and decreased in level by a factor of 250. With the same approach, DI gain error was applied following the model by Wang et al. (2025). In this training data set, foreground contamination was not included, as it was assumed that the pre-processing steps would be able to almost completely remove the point sources and synchrotron galactic emission.

The network was trained for approximately 250 epochs and only for simulated data within 166 and 181 MHz. In previous studies by this team, the neural network showed an improved performance at frequencies where the 21 cm standard deviation is maximized. For this reason, it was decided to focus efforts on this frequency range. Moreover, the size of the data set and the frequency coverage were limited by computational resources. Because of the limited GPU memory, it was not possible to train a network that manages an input with a mesh size of 1024. Instead, the neural network trains and predicts only a small portion of the entire image, with a size of 128, so that they can be patched together later. The trained network was applied on the residual dirty image $I_{\rm res}$. 128^2 regions were selected to overlap with each other, such that it was possible to calculate the average and standard deviation of the predicted 21 cm signal based on the number of overlaps in each pixel. The 2D power spectra for the corresponding frequency range were finally calculated with the TOOLS21CM software (Giri et al. 2020). The resulting average and standard deviation were then submitted as the final result and error, respectively.

3.16 SROT

SROT (Space Radio Observation and Testing) studies the signals from the cosmos through its in-house design of antenna systems and code pipeline design.

Soares et al. (2022) and Hothi et al. (2021) find that GPR is a better-performing solution at lower frequencies than alternatives, e.g. the PCA method. Hence, GPR was used to solve the proposed challenge, specifically the GPR4IM PYTHON package,[29] which was written for MeerKAT's H I observation data.

The pipeline utilized the GPR4IM package and made needful changes to achieve the results. Image cube data were used in both natural and uniform weighted, but results were presented on the natural-weighted data. The GPR4IM package supports fetching the cylindrical power spectrum, as requested by the challenge. The result of these changes and the adjustment to run the whole pipeline within the resources available, i.e. 32 cores and 125 GB of RAM, is presented in the code repository.[30]

3.17 Wizards of Oz 3D

The team, composed of Australia-based members of the MWA EoR project, chose to approach the SDC3a challenge with the same methodology used for the MWA, to explore how well the same methods were applicable to these data. As such, the same calibration and compact source subtraction approach were adopted, and any sophisticated foreground fitting was not attempted.

The test data set was used to define the normalization of our results, but the team was unable to resubmit the results after the issue described in Section 2.4 was discovered, due to loss of team members. As such, the results retain the bias inherent in the test data set.

The team used a custom version of `mwa_hyperdrive`, which is the primary calibration software for MWA (Tingay et al. 2013; Wayth et al. 2018) EoR data in Australia (Jordan et al. 2025). HYPERDRIVE is open source and licensed under the Mozilla Public License version 2.0. A number of improvements and accommodations to HYPERDRIVE were made for SDC3a:

(i) The ability to read, calibrate, and write single-polarization data (with full-polarization data, Jones matrices are assumed as the 'unit' of a visibility throughout the code. Using only a single polarization meant that calibration would always fail, because all Jones matrices were singular.).

(ii) The ability to remove precessed UVWs UVFITS; MS formats expect their UVWs to be in the J2000 frame, but OSKAR appears to only write UVWs in the observation frame.

(iii) Inclusion of an 'airy disk' beam model, which was used for the SDC3a data.

(iv) A small offset between the supplied UVWs and those generated by the pipeline was noticed. An additional time offset was used to robustly match the supplied UVWs.

[28] Using the same parameters as described in the SDC3a data product descriptions.

[29] https://github.com/paulassoares/gpr4im

[30] https://github.com/AkashRadio/SKA_SDC3

After creating our 12 time-steps, full bandwidth files and finding the time offset according to the UVWs, the team investigated calibration. No gain error was found, so calibration was abandoned. Additionally, ionospheric offsets were found to be present but too small to be noticeable. Most of the time was instead focused on refining the sky model to improve the quality of the sky-model-subtracted visibilities. The SDC3a-supplied LoBES sky model did not work as well as another version of LoBES available. After stripping LoBES sources that did not appear to be in the SDC3a data, the aegean source finder (Hancock, Trott & Hurley-Walker 2018) was run on a cleaned image to obtain a sky model for the T-RECS sources. This process was iterated until the subtracted visibilities looked good in images as well as power spectra.

The power spectrum estimation methodology followed that used by the CHIPS software for MWA data (Trott et al. 2016), with updated parameters relevant for SKA. The data were split into even/odd time-steps for power spectrum estimation, and to remove noise power bias. The calibrated and subtracted visibilities were gridded onto the *uv*-plane using a Blackman–Harris gridding kernel matched to the instrument FoV, along with a separate weights grid. For each *uv*-cell, a fourth-order polynomial was fitted to the real and imaginary components of the spectrum and subtracted, in order to reduce residual diffuse emission. A fourth-order polynomial has sufficiently long coherence lengths to not affect the cosmological signal. The cubes were then split into an individual cube for each redshift range. The data were Fourier Transformed along the spectral direction, after weighting with a Blackman–Harris window. Visibility cubes were normalized by their weights, and then cylindrically averaged and squared to produce the final power spectra. Noise was calculated using the difference set of visibilities. Data analysis was performed using resources of the Pawsey Supercomputing Research Centre.

4 ASSESSMENT OF RESULTS

In this section, we assess the performance of the participating teams by means of different metrics, starting from the SDC3a score (Section 4.1) and then moving to power spectrum residuals (Section 4.2), and error bar accuracy (Section 4.3).

4.1 SDC3a score

The SDC3a results submission consists of a list of power spectrum values with error bars $P_i(k_\parallel, k_\perp) \pm \Delta P_i(k_\parallel, k_\perp)$ where $i = 1, 6$ represents the power spectra covering the 90 MHz frequency interval in 15 MHz slices, and $(k_\parallel, k_\perp)$ are the bins in the k-modes parallel and perpendicular to the LoS.

The SDC3a score has been designed to take into account the accuracy of both the EoR power spectrum P and of the error bars ΔP, over the full range of scales $k = 0.05$–0.5 and frequencies $\nu = 106 - 196$ MHz. Since we are neglecting correlation between bins in the error characterization, all power spectrum values are treated as independent. To simplify the notation, we can introduce a new index j that runs through all power spectra and k bins. In this notation, $P_j \pm \Delta P_j$ is the recovered result and P'_j is the true input (shown in Fig. 2).

In the Gaussian approximation, for each j, we can compute the probability of the true value P'_j given the measurement $P_j \pm \Delta P_j$ as

$$\Pr(P'_j) = \frac{1}{\sqrt{2\pi}\Delta P_j} \exp[-(P'_j - P_j)^2/2\Delta P_j^2]. \quad (3)$$

Table 2. List of teams that employed foreground avoidance as part of their strategy, and percentage of data submitted per frequency interval (lowest frequency from left to right).

$\nu_{\min}$[MHz]	106	121	136	151	166	181
$\nu_{\max}$[MHz]	121	136	151	166	181	196
Cantabrigians	64	60	62	64	65	70
DOTSS-21cm_Avoidance	66	67	68	70	70	71
HAMSTER	56	58	65	60	56	55
KUSANAGI-poly	19	20	25	25	27	30
KUSANAGI-GPR	19	20	25	25	27	30

Table 3. SDC3a teams ordered by the score of equation (4).

Rank	Team	Score
1	DOTSS-21cm Advanced ML-GPR	240 226
2	DOTSS-21cm ML-GPR	228 445
3	HIMALAYA	134 752
4	DOTSS-21cm Avoidance	109 567
5	Shuimu–Tianlai	98 128
6	ERWA	71 885
7	Wizards of Oz 3D	59 513
8	Akashanga	40 224
9	SKACH	37 202
10	Hausos	26 315
11	REACTOR	
12	KUSANAGI-poly	
13	Cantabrigians	
14	Nottingham–Imperial	
15	KUSANAGI-GPR	
16	Pisano Galaxy Moppers	
17	HAMSTER	
18	Foregrounds-FRIENDS	
19	KORSDC	
20	SROT	

The final score was obtained by adding all the probabilities together

$$\mathrm{SDC3a_{FOM}} = \sum_j \Pr(P'_j). \quad (4)$$

A possible strategy for solving the challenge is to avoid the most contaminated modes, therefore not submitting all power spectrum entries. In the event that P, ΔP are not submitted for some j, a broad, pre-defined prior $\tilde{P} \pm \Delta\tilde{P}$ was used. With the spectrum $\tilde{P}$, we wanted to capture the possible sources of error, due to noise and foreground contamination. We defined it as

$$\tilde{P} = P' + P_{\mathrm{foreg}} \times f_{\mathrm{foreg}}^2 + P_{\mathrm{noise}} \quad (5)$$

where P_{foreg} and P_{noise} are the cylindrical power spectra of the foregrounds and of the noise, respectively, and f_{foreg} is a scaling factor to somewhat reduce the foreground contamination. We used $f_{\mathrm{foreg}} = 0.1$, which represents a reduction in the foreground map intensity by one order of magnitude. We further adopted $\Delta\tilde{P} = \tilde{P}$ to reflect a non-detection of this mode. Table 2 lists the teams which employed foreground avoidance as part of their strategy, and the percentage of data entries submitted as a function of frequency.

Table 3 presents the SDC3a teams raked by the SDC3 score. Scores are included for the top 10 teams. The best score is achieved by the team DOTSS-21cm with the Advanced ML-GPR and the ML-GPR foreground removal strategies, followed by HIMALAYA.

In the rest of the section, we explore a variety of metrics and figures to further discuss the challenge outcome. Given the choice of multiple frequency intervals and 2D power spectra as challenge

Table 4. Percentage of the SDC3a score achieved within each of the six frequency bins.

$\nu_{\rm min}$[MHz]	106	121	136	151	166	181
$\nu_{\rm max}$[MHz]	121	136	151	166	181	196
Akashganga	1	6	39	15	17	19
Cantabrigians	3	6	0	1	7	80
DOTSS-21cm	1	5	7	19	24	42
ERWA	0	24	24	27	15	6
Foregrounds-FRIENDS	0	0	0	0	0	99
HAMSTER	2	6	15	18	22	33
Hausos	0	3	79	12	4	0
HIMALAYA	0	30	8	15	16	27
KORSDC	0	0	9	16	38	35
KUSANAGI	5	0	0	0	44	50
Nottingham–Imperial	0	0	0	4	95	0
Pisano Galaxy Moppers	1	4	10	19	28	35
REACTOR	7	18	18	18	18	18
SKACH	0	0	56	17	7	18
SROT	16	16	16	16	16	16
Shuimu–Tianlai	0	2	9	12	28	45
Wizards of Oz 3D	1	0	7	26	29	35

submission, and the high number of teams that completed the challenge, a thorough inspection of the results requires a large number of figures. In the rest of this section, we include only some of them, for illustration and discussion purposes. The complete set of figures is made available in an online repository.[31] In all team comparison plots, to reduce figure complexity, we show only the best-scoring entry for the DOTSS-21cm and KUSANAGI teams that submitted multiple results. Whenever results are shown for a single frequency bin rather than the whole 106–196 MHz range, we include here the 151–166 MHz bin, while the other ones are available in the online repository. The reason for this choice is that, for most teams, scores steadily improve with higher frequency, as foreground emission gets fainter, such that the central frequencies represent the in-between scenario. However, some teams exhibit different and sometimes much more marked trends with ν, which means the central bin is not a good representation of the overall performance. This is quantified in Table 4, which shows how the total score is accumulated between the six frequency bins for all teams.

Fig. 6 shows the submission and the comparison with the true input at 151–166 MHz for the top five scoring teams. The first column shows attempts made by different teams to recover the true power spectrum in the bottom-left panel of Fig. 2. The error patterns (see also the third column) are quite different, with some team displaying a stronger trend with $k_\parallel$ than others. Error bars also exhibit a significant variation between teams.

4.2 True versus estimated signal

In Fig. 7, we show the diagonal terms of the recovered $P(k_\parallel, k_\perp)$ for the frequency range 151–166 MHz. For comparison, the true power spectrum (black diamonds) and the data before foreground removal (black dot–dashed line) are also shown. To aid figure clarity, teams' results are divided into three panels based on alphabetical order.

Lines that appear incomplete indicate a foreground avoidance strategy. The team KUSANAGI performed an aggressive masking which included the power spectrum diagonal, which means their results do not appear in this figure. SROT's entry is not visible as below the lower limit of the plot. All teams achieve a suppression of the foreground power by at least one order of magnitude, but typically many more. An excess power suggests an incomplete foreground removal, while a power deficit could indicate that some of the EoR signal has been removed. Since the challenge submission involves both foreground mitigation and power spectrum estimation, however, errors could be introduced by the power spectrum computation as well. Five teams (Akashanga, DOTSS-21cm, ERWA, HIMALAYA, and REACTOR) get the closest power spectrum recovery at the frequency and scales shown in the figure.

To investigate errors across the whole sample (all scales, all frequencies submitted), Fig. 8 represents the distribution of the errors $P - P'$ in terms of the median (diamonds) and the 16 per cent and 84 per cent percentiles (horizontal bar) for each team. Positive and negative values correspond to the power spectrum being overestimated and underestimated, respectively. Some distributions are quite asymmetric, which results in an offset between the median and the centre of the bar.

As expected, the largest absolute errors, shown in the top panel of the figure, correspond to the signal being overestimated, which is indicative of a partial foreground removal. The bottom panel presents a zoom-in on those teams that are the closest to the $P - P' = 0$ line.

DOTSS-21cm and HIMALAYA are confirmed as the teams doing the best in terms of recovery of P'; ERWA, Hausos, REACTOR, SKACH, SROT, and Shuimu–Tianlai also perform relatively well in terms of $P - P'$. No team's result is completely unbiased; some biases are positive and some are negative, with amplitudes (as measured by the absolute value of the median) ranging from 4.2×10^{-4} to 1.7×10^{-3} $\mathrm{K}^2 h^{-3}\mathrm{cMpc}^3$ in those eight teams. By considering different teams together, one would be able to significantly reduce any bias, as different teams are biased in different direction. Seven teams (Akashanga, DOTSS-21cm, ERWA, Hausos, HIMALAYA, SKACH, and Shuimu–Tianlai) achieve a value of the median fractional error $-1 < (P - P')/P' < 1$ across all scales and frequencies, although there are significant outliers in some cases.

4.3 True versus estimated error bar

Fig. 9 shows an assessment of error bar estimation on the same frequency and scales of Fig. 7 ($k_\parallel = k_\perp$ elements only), by means of the submitted error bars $\Delta P(k_\parallel, k_\perp)$ normalized by the absolute value of the true error committed $|P(k_\parallel, k_\perp) - P'(k_\parallel, k_\perp)|$. Values below and above 1 represent an underestimation and an overestimation of the error bars, respectively. This figure demonstrates how accurate error estimation is generally an issue for the teams, with the ratio between estimated and true error in some cases exceeding one order of magnitude.

Fig. 10 compares the distributions of $\log(\Delta P/|P - P'|)$ for all teams at all scales and frequencies, in terms of the median (diamonds) and the 16 per cent and 84 per cent percentiles (error bar). The logarithmic scale has been necessary to compress the very high dynamic range of this metric. Values below and above zero correspond to the error bars being under- and overestimated, respectively.

While some teams have been conservative with their error bars (e.g. Akashanga, and REACTOR), most commonly error bars have been underestimated. Seven teams (DOTSS-21cm, ERWA, HAMSTER, HIMALAYA, KUSANAGI, Pisano Galaxy Moppers, and REACTOR) manage to keep $\Delta P/|P - P'|$ within 0.1 and 10. Pisano Galaxy Moppers performs the best overall, with $\Delta P/|P - P'| =$ 0.4–0.8 and a median of 0.6. The mixed performance in this metric reflects the difficulty of estimating and propagating systematic errors.

[31] https://tinyurl.com/SDC3a

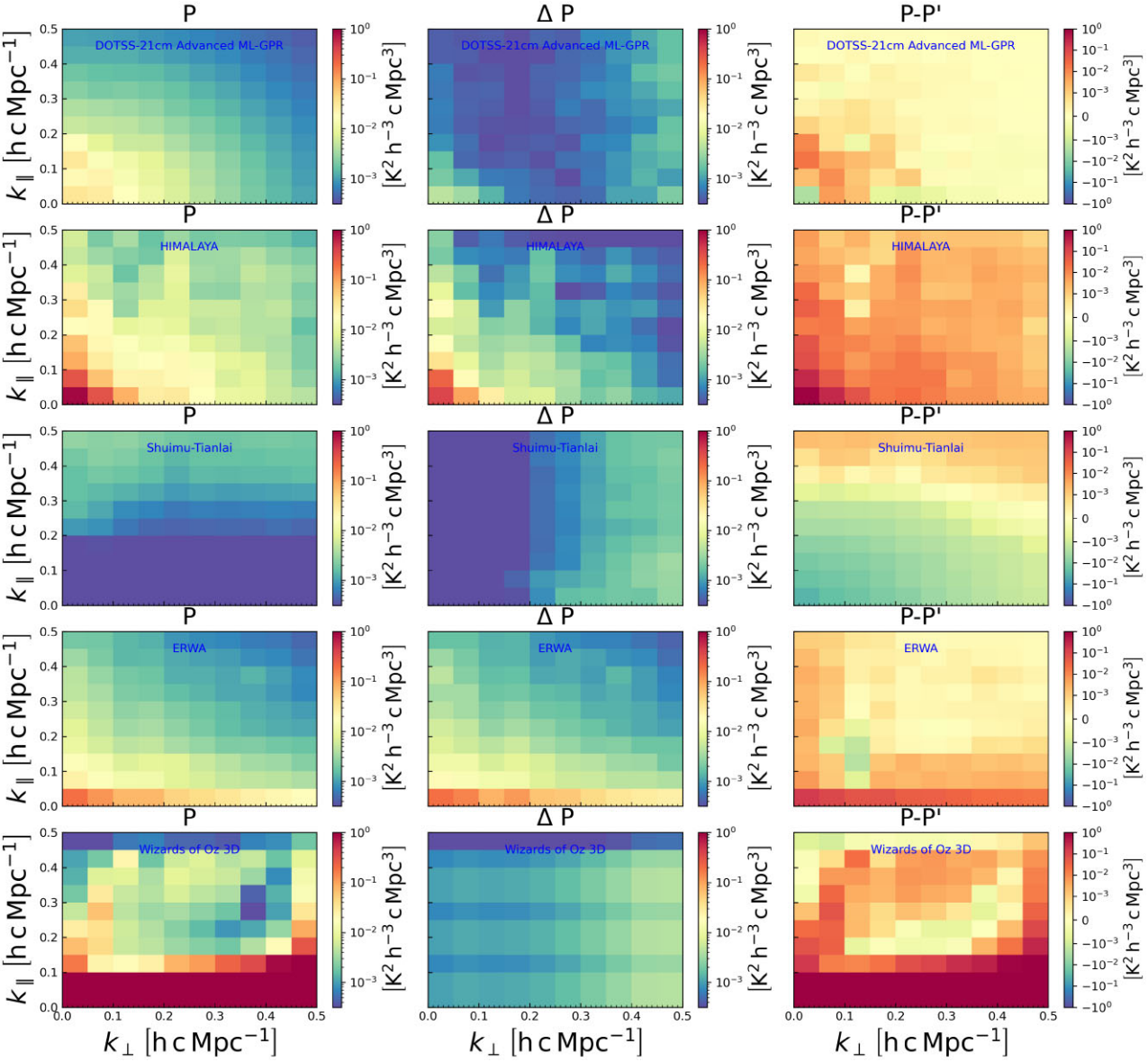

Figure 6. SDC3a submissions for the 151–166 MHz frequency and assessment for the top five teams (with only one result per team) according to the SDC3a score. Left: power spectrum submission P, to be compared with the corresponding true power spectrum P' in Fig. 2. Centre: submitted 1σ error bar ΔP. Right: true error committed, $P - P'$.

5 DISCUSSION AND CONCLUSIONS

SDC3a aimed at bringing simulation work close to the realistic scenario, by including both diffuse and point-like foreground contamination, as well as the residual effect of out-of-field strong sources that enter the FoV due to the telescope's sidelobes, ionospheric effects, and calibration errors. This work does not include systematics induced by polarization, for example polarization leakage. The increased realism mandates the use of visibility data, rather than image-plane data, for the data simulation and, typically, for part of the data analysis performed by teams. The resulting analysis is more complex and more computationally demanding than some previous exercises, which needs to be taken into account when comparing with previous results.

There were 20 submissions of results at the end of the challenge, coming from teams all around the world. As in previous SKA SDCs, different teams had different level of expertise in this analysis, and employed pipelines ranging from 'tried and tested' to still under development, both of which contribute to the difference in teams' performance. Some results exhibit systematics that suggest an issue with the actual computation of the power spectrum (e.g. the binning in k, the correction of the PSF and/or primary beam) rather than the foreground reduction, which complicates the comparison of the results. Several performance metrics are presented in this work for all competing teams, giving feedback and indicating paths for improvement. Additionally, the release of the ground truth after the challenge completion allows teams to further inspect their results.

All teams achieved a significant suppression of the initial contamination from foreground emission, however in some cases the result is still significantly biased, mostly – but not exclusively – in excess of the true signal. One of the greatest challenges with real data, and,

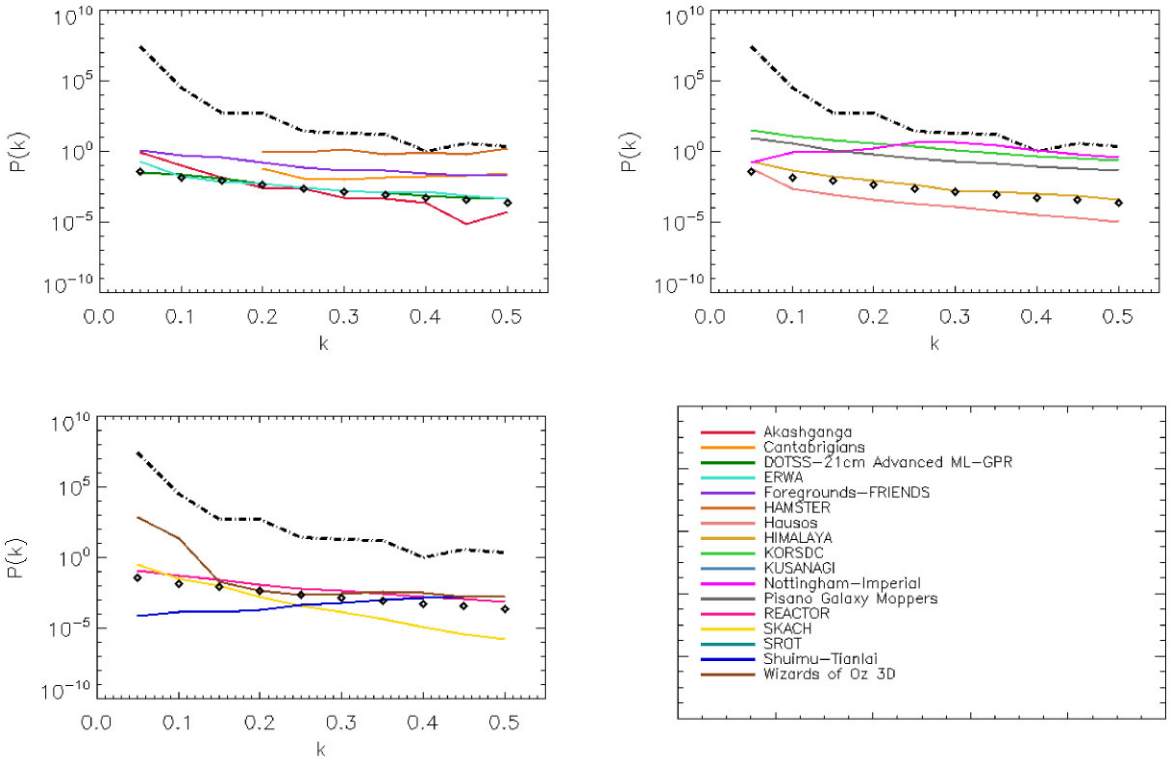

Figure 7. Diagonal terms ($k_\parallel = k_\perp$ elements only) of the 2D power spectrum $P(k_\parallel, k_\perp)$ for the 151–166 MHz frequency range. Diamonds: true EoR power spectrum P'; dot–dashed lines: true total power spectrum, including foregrounds. Solid lines: EoR power spectra P recovered by teams, as detailed in the legend, spread into three panels for clarity.

to some extent, with this simulation, is the interaction between the instrumental PSF including sidelobes and strong point-like sources, which creates ripples across the whole image. Those frequency-dependent ripples affect the smoothness of the foreground emission in frequency space, and renders foreground-cleaning methods much less effective. As shown in Table 1, all teams applied multiple stages of cleaning to solve the challenge, with pre-processing steps typically designed to reduce this extra complexity. Some of the best results in terms of foreground reduction seem to stem from strategies that adopted either an iterative detection and subtraction of point sources from the visibilities, or an explicit step to mitigate the effects of the PSF (convolution/deconvolution). Either approach could have reduced the ripples in the data significantly, and made the subsequent application of a foreground-cleaning method much more successful. The best performance in terms of EoR power spectrum recovery, from the team DOTSS-21cm, is $P - P' = 4.2^{+20}_{-4.2} \times 10^{-4}\,\mathrm{K}^2 h^{-3}\mathrm{cMpc}^3$ over all frequencies and scales probed by the challenge. For comparison, the true EoR power spectrum across all frequencies and scales is $P' = 1.7^{+14}_{-1.3} \times 10^{-3}\,\mathrm{K}^2 h^{-3}\mathrm{cMpc}^3$, which means that the residual error is about four times lower. Six more teams (ERWA, HAMSTER, HIMALAYA, KUSANAGI, Pisano Galaxy Moppers, and REACTOR) also achieve a median residual error below the true EoR power spectrum.

The estimation of the error bars ΔP (approximated as Gaussian and uncorrelated) presented a challenge for the teams, with the true error committed being in some cases orders of magnitude away from the estimation. This indicates the need for robust error estimation strategies, since over- or underestimated error bars can mean the difference between having a detection of the EoR signal or an upper limit, and can lead to wrong conclusions in terms the inference of the reionization properties of the Universe. The best performance in terms of error estimation, from the team Pisano Galaxy Moppers, is $\Delta P/|P - P'| = 0.6^{+0.2}_{-0.2}$.

The difficulty with estimating errors in foreground removal is that they are mostly systematic errors, stemming from effects that the analysis either does not consider or models only partially. As such, they cannot easily be quantified nor propagated. A possible strategy involves relying on parallel analysis on simulations, where the ground truth is known. The success of this strategy for error estimation relies on the accuracy of the simulation, both in terms of the sky signal and the instrumental response. Another approach is the comparison of results on real data between independent pipelines. Fig. 8 shows a promising convergence of several teams around the truth, which indicates that such an approach would be feasible.

ACKNOWLEDGEMENTS

The authors thank the referee for useful comments and corrections, which resulted in an improved paper.

Akashanga. The Akashanga team acknowledges the contribution from its members: Arnab Chakraborty, Suman Chaterjee, Samir Choudhuri, Samit Pal, Narendra Nath Patra, Asif Elahi, Madhurima Choudhury, Anshuman Tripathi, Chandrashekhar Murmu, Srijita Pal, Santanu Das, Rajesh Mondal, Abinash Kumar Shaw, Rahul Shah, Gurmeet Singh, Soumadeep Saha, and Utpal Garain.

Cantabrigians. The Cantabrigians team acknowledges the contribution from its members: Yuchen Liu, Oscar Sage David O'Hara, Eloy de Lera Acedo, Jeremy Coles, Samuel Alan Kossoff Leeney, Thomas Gessey-Jones, and Dominic Anstey.

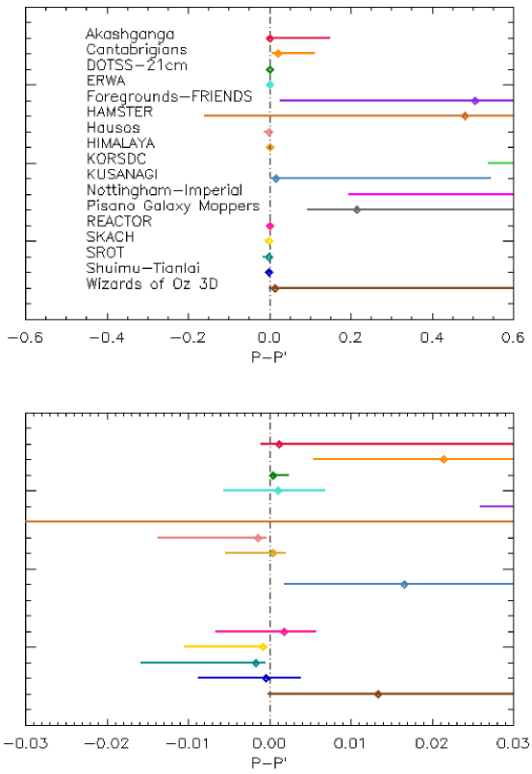

Figure 8. Median (diamonds) and 16 per cent–84 per cent percentiles (bar) of the estimated – true power spectra, $P - P'$, over all frequencies and scales for all the teams, as detailed in the legend. The value on the y-axis has no meaning and it is used to displays the teams in alphabetical order. Top: expanded error range to include all teams; and bottom: zoom-in around the perfect result, $P - P' = 0$.

DOTSS-21. The DOTSS-21 team acknowledges the contribution from its members: F. Mertens, K. Chege, A. Offringa, B. Gehlot, S. Munshi, S. Ghosh, A. Acharya, B. Semelin, L. Koopmans, E. Allys, S. Brakenhoff, E. Ceccotti, D. Cornu, J.M. Delouis, H. Gan, C. Höfer, I. Hothi, F. Levrier, R. Meriot, M. Mevius, and M. A. Miville-Deschênes.

ERWA. The ERWA team acknowledges the contribution from its members: Z. Zhu, H. Shan, and Q. Zheng. The ERWA team acknowledges funding from the National Natural Science Foundation of China (no. 12203085) and the National SKA Project of China (no. 2020SKA0110100).

Foregrounds-FRIENDS. The Foregrounds-FRIENDS team acknowledges the contribution from its members: M. Ruiz-Granda, I. Labadie-García, M. Cruz, M. Aparicio, A. Fernandez, M. López-Caniego, V. Salvador, L. Darriba, J. Moldón, J. Garrido, D. Herranz, M.-L. Gendron-Marsolais, M. Remazeilles, J. A. Rubiño-Martin, S. Sánchez, L. Verdes-Montenegro, P. Vielva, and T. Wiegert. The Foregrounds-FRIENDS acknowledges the Spanish Prototype of an SRC (SPSRC) service and support funded by the Spanish Ministry of Science, Innovation and Universities, by the Regional Government of Andalusia, by the European Regional Development Funds and by the European Union NextGenerationEU/PRTR.

LVM, SS, JG, JM, TW, MGM, LD, and IL acknowledge financial support from the grant CEX2021-001131-S funded by MCIU/AEI/10.13039/501100011033 and from the grant PID2021-123930OB-C21 funded by MCIU/AEI/ 10.13039/501100011033 and by ERDF/EU. LVM, JG, TW, and MGM acknowledge financial support from the coordination of the participation in SKA-SPAIN, funded by the Ministry of Science, Innovation and Universities (MCIU). LVM, SS, JG, LD, and IL acknowledge financial support from the grant TED2021-130231B-I00 funded by MCIU/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. IL acknowledges financial support from the grant PRE2021-100660 funded by MCIU/AEI/10.13039/501100011033 and by ESF+. JM acknowledges financial support from grant PID2023-147883NB-C21, funded by MCIU/AEI/ 10.13039/501100011033 and by ERDF/EU.

MRG, DH, MC, MR, and PV have been supported by MCIN/AEI/10.13039/501100011033, project refs. PID2019-110610RB-C21PID2022-139223OB-C21 (funded also by European Union NextGenerationEU/PRTR), and from Universidad de Cantabria and Consejería de Educación, Formación Profesional y Universidades del Gobierno de Cantabria, via the 'Actividad estructural para el desarrollo de la investigación del Instituto de Física de Cantabria'. MRG acknowledges financial support from the Formación del Profesorado Universitario program of the Spanish Ministerio de Ciencia, Innovación y Universidades. MRG and DH thank the Spanish Agencia Estatal de Investigación (AEI, MICIU) for the financial support provided under the project with reference PID2022-140670NA-I00.

HAMSTER. The HAMSTER team acknowledges the contribution from its members: Zhaoting Chen, Steven Cunnington, Aishrila Mazumder, Amadeus Wild, and Laura Wolz.

Hausos The Hausos team acknowledges the contribution from its members: Ming Jiang, Zhenzhen You, Jean-Luc Starck, Huanyuan Shan, and Feng Shi. MJ acknowledges funding support from the National Natural Science Foundation of China (program no. 12203038) and the Guangdong Basic and Applied Basic Research Foundation (program no. 2021A1515110057).

HIMALAYA. The contribution from the HIMALAYA team is authored by Le Zhang.

KORSDC. The KORSDC team acknowledges the contribution from its members: Kyungjin Ahn, Minji Oh, Jaebeom Kim, David Parkinson, Namuk Lee, Dahee Lee, Minsu Kim, Shinna Kim, Sungwook E. Hong, Se-Heon Oh, Hyeseung Lee, Eunyu Lee, Jae-Young Kim, Junhyun Baek, Jaehong Park, Bong Won Sohn, and Hyunwoo Kang.

KUSANAGI. The KUSANAGI team acknowledges the contribution from its members: S. Yoshiura, T. Ito, T. Minoda, T. T. Takeuchi, T. J. Hayashi, K. Takahashi, H. Shimabukuro, K. Hasegawa, and T. Akahori. The KUASAGI team thanks the JP-SRC funded by SKAJ and cooperating by Kumamoto University and Nagoya University for providing computational resources used their analysis.

Nottingham-Imperial. The Nottingham–Imperial team acknowledges the contribution from its members: Emma Chapman, Luke Conaboy, Jennifer Feron, Carina Norregaard, and Jonathan Pritchard. The team acknowledges the use of the University of Nottingham HPC Augusta.

Pisano Galaxy Moppers. The Pisano Galaxy Moppers team acknowledges the contribution from its members: A. Nasirudin, S. Murray, and A. Mesinger. The team gratefully acknowledges computational resources of the Center for High Performance Computing (CHPC) at Scuola Normale Superiore (SNS) and the Pleiadi infrastructure by INAF.

REACTOR. The REACTOR team acknowledges the contribution from its members: Shulei Ni, Huaxi Chen, Hao Chen, and Xuejian Jiang.

Shuimu–Tianlai. The Shuimu–Tianlai team acknowledges the contribution from its members: Shifan Zuo, Kangning Diao, Richard Grumitt, Yi Mao, Xuelei Chen, Furen Deng, Yan Gong, Yuer

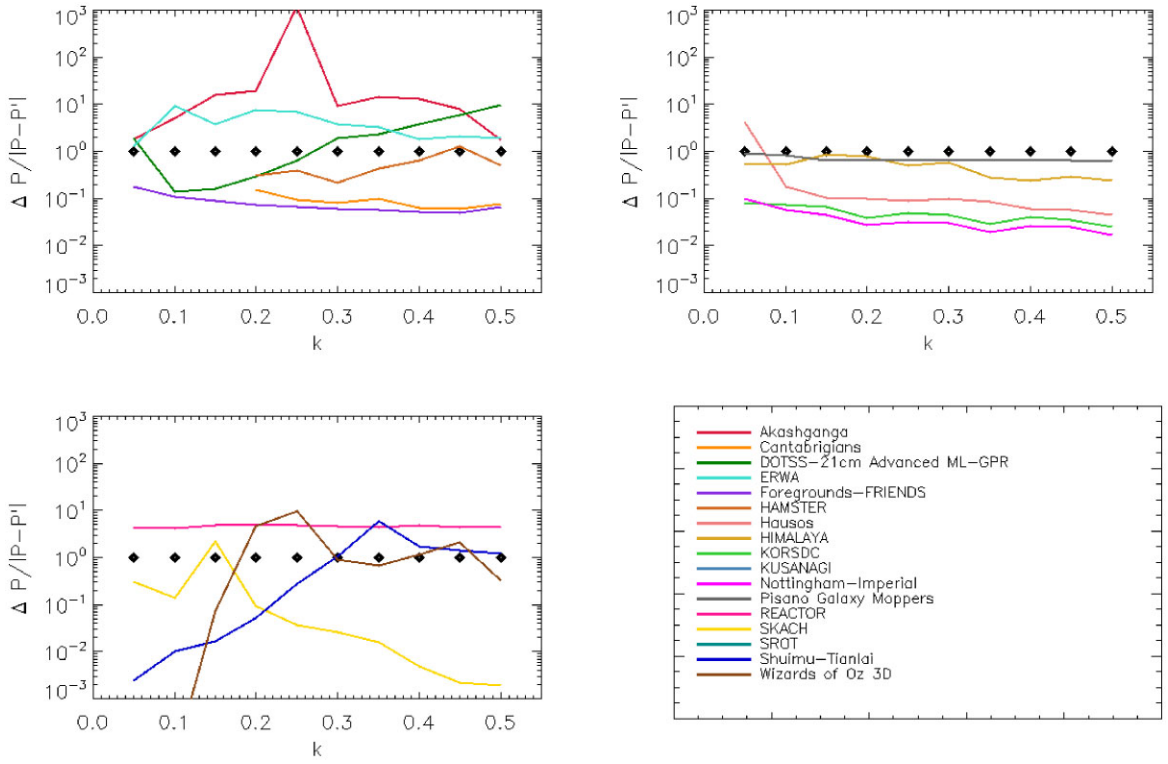

Figure 9. Diagonal terms ($k_\parallel = k_\perp$ elements only) of the submitted error bar $\Delta P(k_\parallel, k_\perp)$ divided by the absolute value of the true error $|P(k_\parallel, k_\perp) - P'(k_\parallel, k_\perp)|$ for the 151–166 MHz frequency range. Lines: EoR power spectra recovered by teams, as detailed in the legend, spread into three panels for clarity. Diamonds: fiducial value corresponding to perfectly accurate error bars; value below and above the diamonds indicate that the error bars have been underestimated and overestimated, respectively.

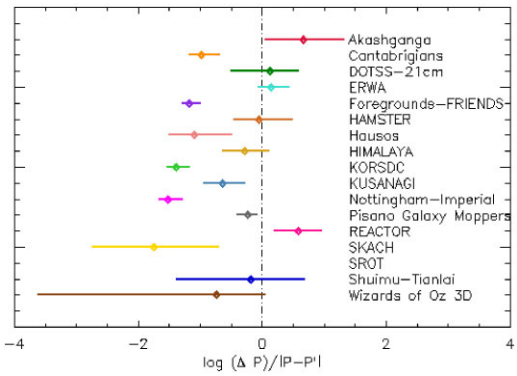

Figure 10. Median (diamonds) and 16 per cent–85 per cent percentiles (bar) of the distribution of $\log(\Delta P/|P - P'|)$ over all frequencies and scales for all the teams. This quantity represents the estimated error bar normalized by the true error committed, and the perfect result is therefore $\log(\Delta P/|P - P'|) = 0$, represented by the vertical dot–dashed line. Points to the left (right) of this line correspond to underestimated (overestimated) error bars. The value on the *y*-axis has no meaning and it is used to displays the teams in alphabetical order.

Jiang, Yichao Li, Yingfeng Liu, Qingbo Ma, Hayato Shimabukuro, Tian-Yang Sun, Qiao Wang, Xin Wang, Yu-Xin Wang, Jun-Qing Xia, Yidong Xu, Ruiqing Yan, Ye-Peng Yan, Zongyao Yin, Kaifeng Yu, Xianchuan Yu, Bin Yue, Li Zhang, Xin Zhang, and Xingchen Zhou. The Shuimu–Tianlai team acknowledges funding from the National SKA Program of China (grant nos 2020SKA0110400, 2020SKA0110401, 2020SKA0110402, 2022SKA0110100, 2022SKA0110101, 2022SKA0110200, and 2022SKA0110203), and the support of the National Natural Science Foundation of China (grant nos 12473001 and 12473091) and the Fundamental Research Funds for the Central Universities (grant no. N2405008).

SKACH. The SKACH team acknowledges the contribution from its members: M. Bianco, S. K. Giri, R. Sharma, S. Krishna, T. Chen, C. Finlay, V. Nistane, P. Denzel, M. De Santis, and H. Ghorbel.

SROT. The SROT team acknowledges the contribution from its members: Akash Kulkarni, Nirmala S. R., and Basawaraj. The SROT team acknowledges the Spanish Prototype of an SRC (SPSRC) service and support funded by the Spanish Ministry of Science, Innovation and Universities, by the Regional Government of Andalusia, by the European Regional Development Funds and by the European Union NextGenerationEU/PRTR.

Wizards of Oz. The Wizards of Oz team acknowledges the contribution from its members: C.H. Jordan, J.H. Cook, J.B. Line, D. Null, C. Nunhokee, B. Pindor, A. Selvaraj, and C.M. Trott. The team ackowledges the use of resources of the Pawsey Supercomputing Research Centre.

ChinaSRC. This work is supported by 100 101 Key Laboratory of Radio Astronomy and Technology (Chinese Academy of Sciences) and used resources of the China SKA Regional Centre prototype funded by the Ministry of Science and Technology of the People's Republic of China and Chinese Academy of Sciences, National SKA Program of China (grant nos 2022SKA0130103 and 2022SKA0120102). Shaoguang Guo is supported by the international partnership program of the Chinese Academy of Sciences (grant no. 018GJHZ2024025GC).

Table 5. Software delivery made available by some of the competing teams.

Team name	Pipeline	Reference
Cantabrigians	https://github.com/ycliu23/Cambridge-SKA-SDC3-Foregrounds	Liu et al. (2023)
DOTSS	https://gitlab.com/flomertens/dotss21_sdc3_pipeline	Mertens et al. (2023)
ERWA	https://github.com/zzh0616/SKA-DECONV	Zhu, Shan & Zheng (2023)
FOREGROUNDS-FRIENDS	https://github.com/espsrc/FOREGROUNDS-FRIENDS	Ruiz-Granda et al. (2023)
Hausos	https://github.com/CEA-jiangming/Hausos-sdc3a/tree/v1.0	Jiang, Starck & Shan (2023)
HIMALAYA	https://github.com/553445316/HIMALAYA.git	Zhang (2023)
KORSDC	https://github.com/KJ-Ahn/KORSDC_FGremove	Ahn et al. (2023)
SROT	https://github.com/AkashRadio/SKA_SDC3	Kulkarni, SR & Patil (2023)
Wizards of Oz	https://github.com/d3v-null/sdc3-pipeline	Jordan et al. (2023)

CESGA. The supercomputer FinisTerrae III and its permanent data storage system have been funded by the NextGeneration EU 2021 Recovery, Transformation and Resilience Plan, ICT2021-006904, and also from the Pluriregional Operational Programme of Spain 2014–2020 of the European Regional Development Fund (ERDF), ICTS-2019-02-CESGA-3, and from the State Programme for the Promotion of Scientific and Technical Research of Excellence of the State Plan for Scientific and Technical Research and Innovation 2013–2016 State subprogramme for scientific and technical infrastructures and equipment of ERDF, CESG15-DE-3114.

ASTRON/SURF. Access to this system was supported by FuSE, the Fundamental Sciences e-Infrastructure, Dutch Research Council (NWO) Roadmap Project 184.035.004, and by SURF Innovation Funds.

SPSRC. The SPSRC acknowledges financial support from the State Agency for Research of the Spanish MCIU through the 'Center of Excellence Severo Ochoa' award to the Instituto de Astrofísica de Andalucía (SEV-2017-0709) and from the grant CEX2021-001131-S funded by MCIU/AEI/ 10.13039/501100011033 (Garrido et al. 2021).

JS and MP acknowledge financial support from the grant CEX2021-001131-S funded by MCIU/AEI/ 10.13039/501100011033 and from the grant PID2021-123930OB-C21 funded by MCIU/AEI/ 10.13039/501100011033 and by ERDF/EU. JS and MP acknowledge financial support from the grant TED2021-130231B-I00 funded by MCIU/AEI/ 10.13039/501100011033 and by the European Union NextGenerationEU/PRTR.

CSCS. This effort was made possible through the Performance Contract 2021–2024 in support to the activities of the Swiss CTAO Collaboration and the Swiss SKAO-Consortium funding with the Swiss State Secretariat for Education, Research and Innovation, SERI.

UKSRC, IRIS-CAM and IRIS-MAN. Support comes from the Science and Technology Facilities Council (STFC) and the IRIS project.

DATA AVAILABILITY STATEMENT

The SDC3a simulated data and code used to produce it is made available in Bonaldi et al. (2025). Table 5 reports details of the code used to analyse the SDC3a data that have been made public by the participating teams. Each pipeline was evaluated by an expert panel against the pre-defined criteria, following the recommendations of the The Software Sustainability Institute (SSI)[32] (Crouch et al. 2013).

[32]https://www.software.ac.uk/

REFERENCES

Abdurashidova Z. et al., 2022, ApJ, 925, 221
Acharya A. et al., 2024, MNRAS, 527, 7835
Ahn K., Kim J., Oh M., Parkinson D., 2023, KORSDC SDC3a Software. Available at: https://github.com/KJ-Ahn/KORSDC_FGremove
An T., Wu X.-P., Hong X., 2019, Nat. Astron., 3, 1030
An T., Wu X., Lao B., Guo S., Xu Z., Lv W., Zhang Y., Zhang Z., 2022, Sci. China Phys. Mech. Astron., 65, 129501
Asad K. M. B. et al., 2016, MNRAS, 462, 4482
Barry N. et al., 2019, ApJ, 884, 1
Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, MNRAS, 461, 3135
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bianco M. et al., 2025, MNRAS, 541, 234
Bianco M., Giri S. K., Iliev I. T., Mellema G., 2021, MNRAS, 505, 3982
Bianco M., Giri S. K., Prelogović D., Chen T., Mertens F. G., Tolley E., Mesinger A., Kneib J.-P., 2024, MNRAS, 528, 5212
Bonaldi A. et al., 2025, preprint (arXiv:2506.09533)
Bonaldi A., Bonato M., Galluzzi V., Harrison I., Massardi M., Kay S., De Zotti G., Brown M. L., 2019, MNRAS, 482, 2
Bonaldi A., Hartley P., Ronconi T., De Zotti G., Bonato M., 2023, MNRAS, 524, 993
Bowman J. D., Rogers A. E. E., Hewitt J. N., 2008, ApJ, 676, 1
Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67
Bracco A. et al., 2022, A&A, 663, A37
Chapman E. et al., 2012, MNRAS, 423, 2518
Chapman E., Jelić V., 2019, The Cosmic 21-cm Revolution, Foregrounds and Their Mitigation. IOP Publishing, Bristol, US, p. 2514
Chen Z., Chapman E., Wolz L., Mazumder A., 2023, MNRAS, 524, 3724
Choudhuri S., Bharadwaj S., Ghosh A., Ali S. S., 2014, MNRAS, 445, 4351
Crouch S. et al., 2013, Comput. Sci. Eng., 15, 74
Cunnington S. et al., 2023, MNRAS, 523, 2453
DeBoer D. R. et al., 2017, PASP, 129, 045001
Delabrouille J. et al., 2013, A&A, 553, A96
Dewdney P. E., Braun R., 2016, Technical Report. SKA-TEL-SKO-0000422. SKA Organization, Macclesfield, UK
Dulwich F., Mort B. J., Salvini S., Zarbi Adami K. Jones M. E., 2019, in Torchinsky S. A., van Ardenne A., van den Brink-Havinga T., van Es A. J. J., Faulkner A. J., eds, OSKAR: Simulating Digital Beamforming for the SKA Aperture Array. SISSA, Trieste, PoS(SKADS 2009)031
ZarEastwood M. W. et al., 2019, AJ, 158, 84
Ewall-Wice A. et al., 2016, MNRAS, 460, 4320
Field G. B., 1958, Proc. IRE, 46, 240
Franzen T. M. O., Hurley-Walker N., White S. V., Hancock P. J., Seymour N., Kapińska A. D., Staveley-Smith L., Wayth R. B., 2021, Publ. Astron. Soc. Aust., 38, e014
Garrido J. et al., 2021, J. Astron. Telesc. Instrum. Syst., 8, 1
Garsden H. et al., 2021, MNRAS, 506, 5802
Gehlot B. K. et al., 2018, MNRAS, 478, 1484
Gehlot B. K. et al., 2020, MNRAS, 499, 4158
Gehlot B. K. et al., 2022, A&A, 662, A97

Ghara R., Choudhury T. R., Datta K. K., Choudhuri S., 2017, MNRAS, 464, 2234
Giardino G., Banday A. J., Górski K. M., Bennett K., Jonas J. L., Tauber J., 2002, A&A, 387, 82
Giri S., Mellema G., Jensen H., 2020, J. Open Source Softw., 5, 2363
Hancock P. J., Trott C. M., Hurley-Walker N., 2018, Publ. Astron. Soc. Aust., 35, e011
Handley W. J., Hobson M. P., Lasenby A. N., 2015a, MNRAS, 450, L61
Handley W. J., Hobson M. P., Lasenby A. N., 2015b, MNRAS, 453, 4384
He M., Zheng Q., Guo Q., Shan H., Zhu Z., Xie Y., Huang Y., Zhao F., 2024, MNRAS, 529, 3140
Herranz D. et al., 2023, MNRAS, 519, 3526
Hothi I. et al., 2021, MNRAS, 500, 2264
Hurley-Walker N. et al., 2017, MNRAS, 464, 1146
Hurley-Walker N. et al., 2019, Publ. Astron. Soc. Aust., 36, e047
Hurley-Walker N. et al., 2022, Publ. Astron. Soc. Aust., 39, e035
Intema H. T., van Weeren R. J., Röttgering H. J. A., Lal D. V., 2011, A&A, 535, A38
Irfan M. O., Bull P., 2021, MNRAS, 508, 3551
Jelić V. et al., 2008, MNRAS, 389, 1319
Jiang M., Bobin J., Starck J.-L., 2017, SIAM J. Imag. Sci., 10, 1997
Jiang M., Starck J.-L., Shan H., 2023, Hausos SDC3a Software. Available at: https://zenodo.org/records/10230697
Jishnu Nambissan T. et al., 2021, Exp. Astron., 51, 193
Jordan C. H. et al., 2025, In Proceedings of the 7th 2025 URSI Asia-Pacific Radio Science Conference – AP-RASC. Sydney, Australia
Jordan C. H., Null D., Trott C., Line J., Nunhokee C. D., Selvaraj A., Cook J., Pindor B., 2023, Wizards of Oz SDC3 Software, Zenodo. Available at: https://doi.org/10.5281/zenodo.10148809
Kolopanis M. et al., 2019, ApJ, 883, 133
Koopmans L. et al., 2015, in Tyler L. B. et al., eds, The Cosmic Dawn and Epoch of Reionisation with SKA. SISSA, Trieste, PoS(AASKA14)001
Kulkarni A., SR N., Patil B., 2023, SROT_SKA_SDC3a GPR4im. Available at: https://github.com/AkashRadio/SKA_SDC3.git
Li W. et al., 2019, ApJ, 887, 141
Liu A., Parsons A. R., Trott C. M., 2014, Phys. Rev. D, 90, 023018
Liu Y., O'Hara O. S. D., de Lera Acedo E., Coles J., Dulwich F., Leeney S. A. K., Gessey-Jones T., Anstey D., 2023, SKAO SDC3a Cantabrigians. Available at: https://github.com/ycliu23/Cambridge-SKA-SDC3-Foregrounds
Lynch C. R. et al., 2021, Publ. Astron. Soc. Aust., 38, e057
Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429
Mandal S. et al., 2021, A&A, 648, A5
Matshawule S. D., Spinelli M., Santos M. G., Ngobese S., 2021, MNRAS, 506, 5075
McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, p. 127
Mertens F. G. et al., 2020, MNRAS, 493, 1662
Mertens F. G., Bobin J., Carucci I. P., 2024, MNRAS, 527, 3517
Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, MNRAS, 478, 3640
Mertens F., Chege K., Munshi S., Gehlot B., Offringa A., The DOTSS-21 team, 2023, DOTSS-21 SDC3 pipeline (1.1). Zenodo. Available at: https://doi.org/10.5281/zenodo.10263162
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Mitchell D. A., Greenhill L. J., Wayth R. B., Sault R. J., Lonsdale C. J., Cappallo R. J., Morales M. F., Ord S. M., 2008, IEEE J. Sel.Topics Signal Process, 2, 707
Mohan N., Rafferty D., 2015, Astrophysics Source Code Library, record ascl:1502.007
Morales M. F., Hazelton B., Sullivan I., Beardsley A., 2012, ApJ, 752, 137
Munshi S. et al., 2024, A&A, 681, A62
Murray S., Greig B., Mesinger A., Muñoz J., Qin Y., Park J., Watkinson C., 2020, J. Open Source Softw., 5, 2582
Ni S., Li Y., Gao L.-Y., Zhang X., 2022, ApJ, 934, 83
Ni S., Qiu Y., Chen Y., Song Z., Chen H., Jiang X., Chen H., 2024, preprint (arXiv:2403.01692)
O'Hara O. S. D., Dulwich F., de Lera Acedo E., Dhandha J., Gessey-Jones T., Anstey D., Fialkov A., 2024, MNRAS, 533, 2876
Offringa A. R. et al., 2014, MNRAS, 444, 606
Offringa A. R., Smirnov O., 2017, MNRAS, 471, 301
Paciga G. et al., 2011, MNRAS, 413, 1174
Park J., Mesinger A., Greig B., Gillet N., 2019, MNRAS, 484, 933
Parsons A. R. et al., 2014, ApJ, 788, 106
Patil A. H. et al., 2017, ApJ, 838, 65
Paul S., Santos M. G., Chen Z., Wolz L., 2023, preprint (arXiv:2301.11943)
Pedregosa F. et al., 2011, J. Mach. Learn. Res., 12, 2825
Planck Collaboration X, 2016a, A&A, 594, A10
Planck Collaboration XXV, 2016b, A&A, 594, A25
Planck Collaboration VI, 2020, A&A, 641, A6
Platania P., Burigana C., Maino D., Caserini E., Bersanelli M., Cappellini B., Mennella A., 2003, A&A, 410, 847
Pober J. C. et al., 2016, ApJ, 819, 8
Ruiz-Granda M. et al., 2023, Foreground-Friends Participation in the SKA Data Challenge 3 (1.1). Zenodo. Available at: https://doi.org/10.5281/zenodo.10234314
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Singh S. et al., 2017, ApJ, 845, L12
Soares P. S., Watkinson C. A., Cunnington S., Pourtsidou A., 2022, MNRAS, 510, 5872
Spinelli M., Bernardi G., Santos M. G., 2018, MNRAS, 479, 275
Srinath S., Poyneer L. A., Rudy A. R., Ammons S. M., 2015, Opt. Express, 23, 33335
Tingay S. J. et al., 2013, Publ. Astron. Soc. Aust., 30, e007
Trott C. M. et al., 2016, ApJ, 818, 139
Trott C. M. et al., 2018, ApJ, 867, 15
Trott C. M. et al., 2020, MNRAS, 493, 4711
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Wang H. et al., 2025, Phys. Rev. D, 111, 103531
Wang J. et al., 2013, ApJ, 763, 90
Wayth R. B. et al., 2018, Publ. Astron. Soc. Aust., 35, e033
Yoshiura S. et al., 2021, MNRAS, 505, 4775
Zeng C., Ng M. K., 2020, Numerical Linear Algebra with Applications. Wiley, Hoboken, NJ, USA, p. e2290
Zhang L., 2023, HIMALAYA SDC3a Software. Available at: https://github.com/553445316/HIMALAYA.git
Zheng H. et al., 2017, MNRAS, 464, 3486
Zheng S. et al., 2021, Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. IEEE, Nashville, TN, USA, p. 6881
Zhu Z., Shan H., Zheng C., 2023, ERWA SDC3a Software. Available at: https://github.com/zzh0616/SKA-DECONV

[1]*SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield, SK11 9FT, UK*
[2]*Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Straße 1, D-85748 Garching, Germany*
[3]*Department of Earth Sciences, Chosun University, Gwangju 61452, Korea*
[4]*Universidad Europea de Madrid, E-28670 Madrid, Spain*
[5]*Laboratoire d'Astrophysique, Ecole Polytechnique Federale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290, Versoix, Switzerland*
[6]*Department of Physics and Trottier Space Institute, McGill University, 3600 rue University, Montréal, QC H3A 2T8, Canada*
[7]*School of Physics and Astronomy, The University of Nottingham, University Park, Nottingham, NG7 2RD, UK*
[8]*Department of Physics and Astronomy, University of the Western Cape, 7535 Bellville, Cape Town, South Africa*
[9]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, The Netherlands*
[10]*Research Center for Astronomical Computing, Zhejiang Laboratory, Hangzhou 311121, China*
[11]*National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China*

[12]*Institute for Astronomy, The University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK*
[13]*Instituto de Física de Cantabria (CSIC - Universidad de Cantabria), Avda. de los Castros s/n, E-39005 Santander, Spain*
[14]*Instituto de Astrofísica de Andalucía (IAA - CSIC), Glorieta de la Astronomía s/n, E-18008, Granada, Spain*
[15]*Haute Ecole Arc Ingénierie, University of Applied Sciences and Arts Western Switzerland (HES-SO), CH-2610 Saint-Imier, Switzerland*
[16]*Centre for Artificial Intelligence, ZHAW Zurich University of Applied Sciences, Technikumstrasse 71, CH-8400 Winterthur, Switzerland*
[17]*Department of Astronomy, Tsinghua University, Beijing 100084, China*
[18]*Département de Physique Théorique and Center for Astroparticle Physics, Université de Geneve, 24 quai Ernest Ansermet, CH-1211 Geneve 4, Switzerland*
[19]*Nordita, KTH Royal Institute of Technology and Stockholm University, Hannes Alfv´ens väg 12, SE-10691 Stockholm, Sweden*
[20]*Korea Astronomy and Space Science Institute, Daejeon 34055, Korea*
[21]*Astronomy Campus, University of Science and Technology, Daejeon 34055, Korea*
[22]*International Research Organization for Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan*
[23]*National Key Laboratory of Radar Signal Processing, Xidian Universtiy, Xi'an, 710071, China*
[24]*Guangzhou Institute of Technology, Xidian University, Guangzhou 510555, China*
[25]*International Centre for Radio Astronomy Research, Curtin University, 6102, Bentley WA, Australia*
[26]*ARC Centre of Excellence for All-Sky Astrophysics in 3D, Australia*
[27]*ICRAR M468, The University of Western Australia, WA 6009, Crawley (Perth), Australia*
[28]*Department of Physics and Astronomy, Sejong University, Seoul 05006, Korea*
[29]*Korea AeroSpace Administration, Sacheon-si, 52535 Gyeongsangnam-do, Korea*
[30]*Indian Institute of Technology Dharwad, Karnataka 580011, India*
[31]*Aurora Technology for the European Space Agency, Camino bajo del Castillo, s/n, Urbanización; Villafranca del Castillo, Villanueva de la Cañada, E-28692 Madrid, Spain*
[32]*Department of Physics, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea*
[33]*Cavendish Astrophysics, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK*
[34]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[35]*Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK*
[36]*LUX, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Université, F-75014 Paris, France*
[37]*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[38]*Blackett Laboratory, Imperial College London, Department of Physics, RoadPrince Consort Road, London, SW7 2AZ, UK*
[39]*Australian SKA Regional Centre (AusSRC), Curtin University, Bentley, WA, Australia*
[40]*ASTRON Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, NL-7991 PD Dwingeloo, The Netherlands*
[41]*Departamento de Física Moderna, Universidad de Cantabria, Avda. los Castros s/n, E-39005 Santander, Spain*
[42]*Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai 200030, P. R. China*
[43]*School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, P. R. China*
[44]*State Key Laboratory of Radio Astronomy and Technology, Chinese Academy of Sciences, A20 Datun Road, Chaoyang District, Beijing 100101, P. R. China*
[45]*Space, Planetary and Astronomical Sciences and Engineering (SPASE), Indian Institute of Technology, Kanpur, 208016 Uttar Pradesh, India*
[46]*Mizusawa VLBI Observatory, National Astronomical Observatory Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*
[47]*School of Physics and Astronomy, Sun Yat-sen University, Zhuhai 519082, P. R. China*
[48]*Liaoning Key Laboratory of Cosmology and Astrophysics, College of Sciences, Northeastern University, Shenyang 110819, China*
[49]*CFisUC, Departamento de Física, Universidade de Coimbra, P-3000-056 Coimbra, Portugal*
[50]*Laboratory for Advanced Computing of the University of Coimbra, P-3000-056 Coimbra, Portugal*
[51]*Laboratoire de Physique de l'Ecole Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France*
[52]*KLE Tecnological University, Hubli, Karnataka 580031, India*
[53]*University of Cambridge, Research Computing Services, Roger Needham Building, 7 J J Thomson Avenue, West Cambridge Site, Cambridge CB3 0RB, UK*
[54]*INAF – Istituto di Radioastronomia, via P. Gobetti 101, I-40129 Bologna, Italy*
[55]*Centre for Strings, Gravitation and Cosmology, Department of Physics, Indian Institute of Technology Madras, Chennai 600036, India*
[56]*Center of Fundamental Physics of the Universe, Department of Physics, Brown University, Providence, RI 02912, USA*
[57]*IFREMER, Université Brest, CNES, CNRS, IRD, Laboratoire d'Océanographie Physique et Spatiale, 29280 Plouzané, France*
[58]*School of Physics and Astronomy, Sun Yat-Sen University, No.2 Daxue Road, Zhuhai 519082, China*
[59]*Swiss National Supercomputing Center, ETH Zurich, CH-6900 Lugano, Switzerland*
[60]*Fundacion Centro Tecnologico de Supercomputación de Galicia (CESGA), Avda. de Vigo s/n. Campus Sur., E-15705 Santiago de Compostela, Spain*
[61]*INAF, Italian Centre for Astronomical Archives, via G.B. Tiepolo 11, I-34143 Trieste, Italy*
[62]*Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, 700108 Kolkata, India*
[63]*Département de Physique, de génie Physique et d'Optique, Université Laval, Québec (QC) G1V 0A6, Canada*
[64]*Department of Mechanical Engineering, Suzuka National College of Technology, Shiroko-cho, Suzuka, Mie 510-0294, Japan*
[65]*National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Division of Science, Japan*
[66]*Azabu Junior and Senior High School, 2-3-29 Motoazabu, Minato, Tokyo 106-0046, Japan*
[67]*Ruđer Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia*
[68]*IDRIS, CNRS, Université Paris-Saclay, F-91403 Orsay, France*
[69]*School of Physics and Electronic Science, Guizhou Normal University, Guiyang 550001, China*
[70]*Department of Physics, National Institute of Technology Calicut, Calicut, 673601 Kerala, India*
[71]*Department of Astronomy, Astrophysics, and Space Engineering, Indian Institute of Technology Indore, Madhya Pradesh 453552, India*
[72]*School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287-1404, USA*
[73]*Indian Institute of Science, CV Raman Rd, Bengaluru, Karnataka 560012, India*
[74]*School of Physics, The University of Melbourne, 3010 Victoria, Australia*
[75]*Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*
[76]*Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain*
[77]*Physics and Applied Mathematics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700 108, India*
[78]*Department of Computer Science, University of Nevada Las Vegas, 4505 S. Maryland Pkwy., Las Vegas, NV 89154, USA*
[79]*School of Aerospace Science and Technology, Xidian University, Xi'an, 710126, China*
[80]*Yunnan University, SWIFAR, No. 2 North Green Lake Road, Kunming, Yunnan Province 650500, China*
[81]*Indian Institute of Technology, G66M+W8P, Kalyanpur, Kanpur, Uttar Pradesh 208016, India*

[82]*Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, F-91191 Gif-sur-Yvette, France*
[83]*Institutes of Computer Science and Astrophysics, Foundation for Research and Technology Hellas (FORTH), GR 70013 Heraklion, Crete, Greece*
[84]*Division of Particle and Astrophysical Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464–8602, Japan*
[85]*The Research Center for Statistical Machine Learning, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190–8562, Japan*
[86]*INAF – Osservatorio Astrofisico di Catania, via Santa Sofia 78, I-95123 Catania, Italy*
[87]*School of Physics and Astronomy, Beijing Normal University, Beijing 100875, China*
[88]*Department of Physics, Stellenbosch University, Matieland 7602, South Africa*
[89]*National Institute for Theoretical and Computational Sciences (NITheCS), Merensky Building, Merriman Street, Stellenbosch, 7600, South Africa*
[90]*School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China*
[91]*School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, 710048, China*
[92]*College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China*
[93]*State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China*

This paper has been typeset from a TEX/LATEX file prepared by the author.