Foreground removal for HI using GNILC-Generalized Neetlet Internal Linear Combination

Topollo Naketsana[1], Prof Mario Santos[2], Dr Karin Fornazier[2]

[1]Stellenbosch University, Department of Physics, [2]University of the Western Cape

Introduction and Motivations

We want to study the large-scale structure of the universe and how matter is distributed in the sky. Cosmic emissions such as Neutral Hydrogen (HI) 21 cm signal, Carbon Monoxide (CO) and Hα allow us to study the history and origins of the universe - they give us hints about the properties of the galaxies, star formations and other astrophysical events during the Epoch of Re-ionization (EoR).
In this study, we focus mainly on the HI emission and trying to recover this signal from the galactic and extra-galactic foregrounds.

1. Foregrounds:

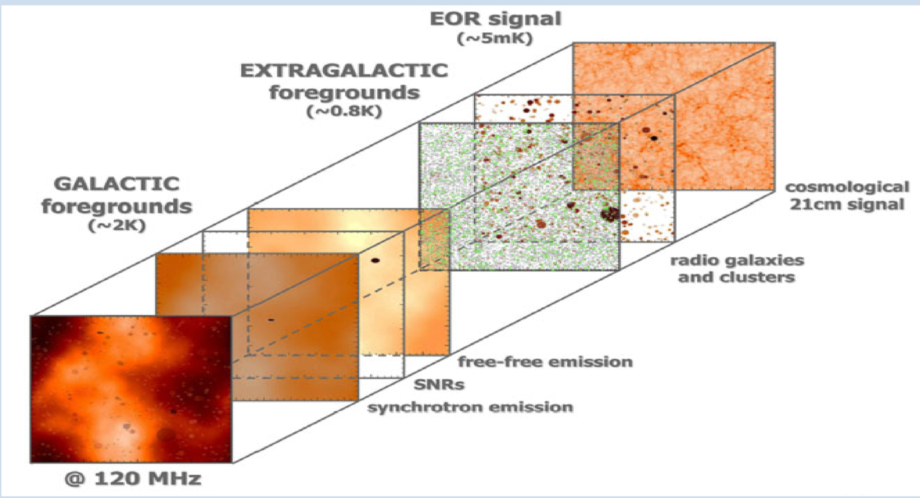

Figure 1. This image illustrates the timeline of the universe - The last layer represents the EoR where we are trying to probe.
Galactic and Extragalactic foregrounds makes it difficult for telescopes to observe the last layer because they are orders of magnitude brighter than the faint cosmological signal from earliest epoch.

2. Signal of Interest - 21 cm signal. We want to recover this signal from the dominant foregrounds. HI is the input map of temperature fluctuations δT along the projection $\hat{n}$ at redshift z,

$$\delta T(\Omega) = \int d\Omega\, W(\Omega)\, \bar{T}(\Omega)\, \delta_{HI}(\Omega) \quad (4)$$

Where W is the Window function, $\delta_{HI} = \delta\rho_{HI}/\rho_{HI}$ is the fractional density of HI.

3. Angular Power Spectrum C_ℓ. To deduce the angular power spectrum, we make use of spherical harmonics ($Y_{\ell m}$) and write temperature as $\delta T = \sum_{\ell m} a_{\ell m} Y_{\ell m}$.
The angular power power spectrum C_ℓ is defined by the two-point correlation function - ensemble average (Olivari, L. C. 2018).

$$C_\ell = \langle a_{\ell m}\, a^*_{\ell m} \rangle \quad (5)$$

Taking the average over m, the angular power spectrum C_ℓ,

$$C_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2$$

References

- Olivari, L. C. (2018). *Approach to probe the large-scale structure of the Universe.*
- Olivari, L. C (2016). *(GNILC) - Generalized Needlet Internal Linear Combination*
- De Caro, B., Carucci, et all. (2025). eGNILC approach for HI intensity mapping.
- Dai, X., Ma, Y. (2025). *Expanded Generalized Needlet Internal Linear Combination (eGNILC).*

Acknowledgements

I would like to thank my supervisors, Dr Karin and Prof Mario Santos, for their guidance and support throughout this project. I also thank the Honours Astrophysics group for their valuable discussions and contributions.

Simulations and GNILC Workflow

To impliment GNILC method in separating HI signal from the foregrounds, we create set of simulated sky maps that reproduce the same emission component that is observed by the MeerKat single dish telescope (MeerKLASS) oberving at the Ultra High Frequency band (UHF) at z = 0.4 - 1.45. The total observed map is the sum of the HI map $s_i(p)$, the astrophysical foregrounds including the systematic noise $n_i(p)$:

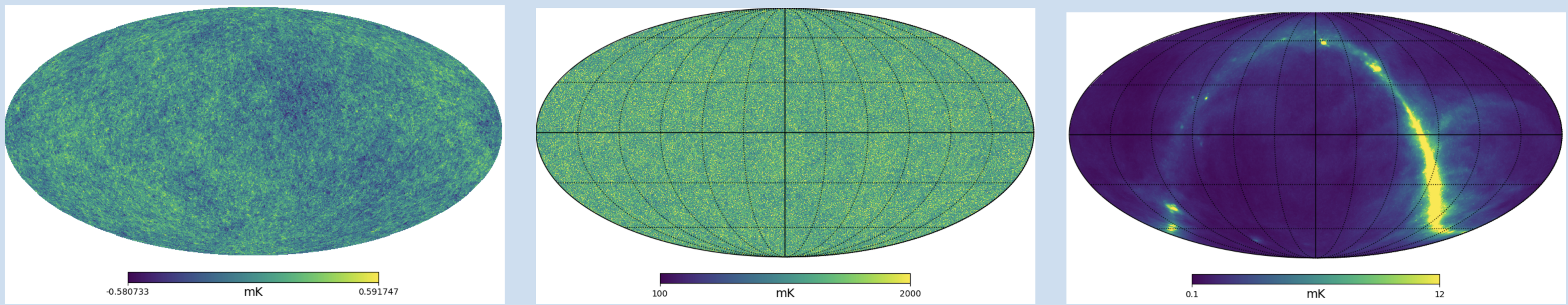

Figure 2. These are the simulated maps at frequency range 550 MHz - 1001 MHz. The first map shows the brightness temperature of the HI 21cm signal, emission from Point sources and the Galactic Synchrotron The total brightness temperature is given by the sum of all three maps stacked together as 2D $N_{ch} \times N_{pix}$

$$T(\nu, p) = T_{HI}(\nu, p) + T_F(\nu, p) + T_n(\nu, p) \quad (1)$$

Moving to spherical harmonics to show the angular power spectrum $C_{\ell m}$

GNILC:
The foreground cleaning technique developed to separate the HI from the foreground. It's an extension of the Internal Linear Combination (ILC) approach by operating in Needlet space (Olivari, 2016).
$T(\nu, p)$ represents the total input map that is used for GNILC, the HI map $T_{HI}(\nu, p)$ is used as a prior knowledge for GNILC to know how the signal should look like, and the output is extracted from the input map.

$$x(p) = s(p) + n(p) \quad (2)$$

This is what GNILC expect from the input - for all frequency channels, represented as matrix form.
The total covarience matrix from eq:2,

$$R_x(p) = \langle x(p)\, x^T(p) \rangle = R_{HI}(p) + R_n(p), \quad (3)$$

Since the spatial viariability is already captured by local covariance, we then omit the pixel dependency, $s = St$.
With independent (unphysical) templates t. Taking the covariance of the above equation,

$$R_{HI} = S R_t S^T$$

The estimation of the signal $\hat{s}$ using the linear operation - ILC,

$$\hat{s} = \sum_j^{n_{ch}} w_j\, x_j$$

With the constraint, $\sum_j w_j = 1$ to minimize the noise and maximize the eigenvalues of the signal of interest.

Results and descussion

The GNILC algorithm was applied to the simulated data cube shown above, to test its ability to recover the faint HI 21 cm signal from strong astrophysical foregrounds. The performance was evaluated by comparing the recovered HI map with the true HI simulation and analyzing their correlation, residuals, and **angular power spectra**. Reults consists of two parts, the first one is when we run GNILC using only Synchrotron as the foreground and HI map (excluding the bright extragalactic point sources) and the second part is when we include the point sources.

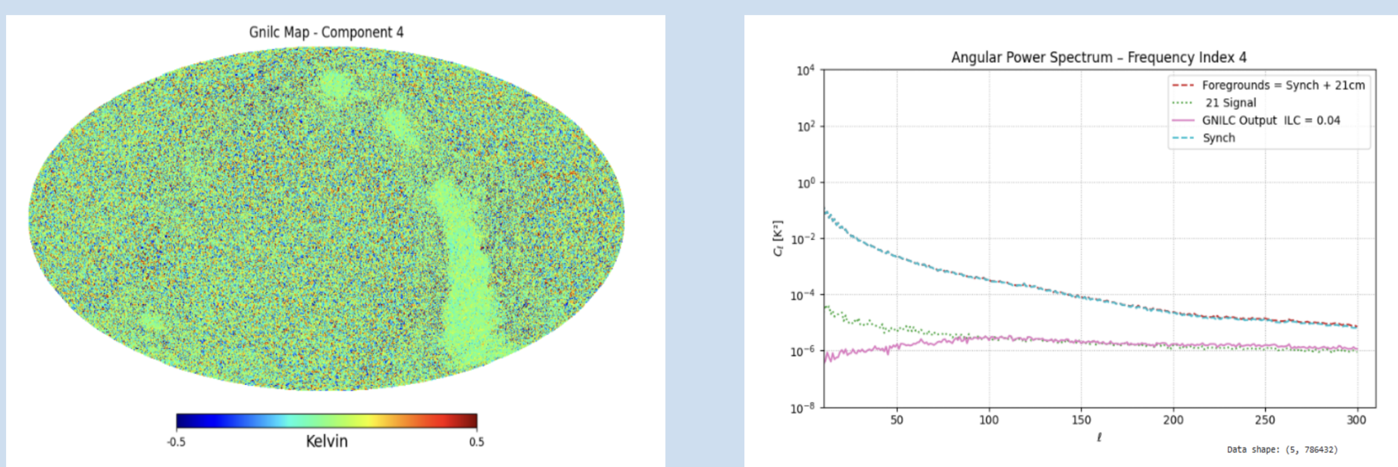

Figure 3. Comparison between the true HI map and the recovered HI map obtained after GNILC cleaning when only Galactic synchrotron emission is included as foreground

Conclusion

We successfully tested the GNILC algorithm for foreground cleaning in simulated HI 21cm signal from the data. The recovered maps showed a strong correlation ratio with the true HI signal(rp = 0.956) and a low residual RMS of 0.26 K, confirming that GNILC can efficiently separatediffuse foregrounds such as Galactic synchrotron emission. The method performs best underidealized conditions but struggles when bright point sources are included, indicating the needfor improved preprocessing and masking strategies.