Intensity Mapping Foreground Cleaning with Generalized Needlet Internal Linear Combination

L. C. Olivari, M. Remazeilles and C. Dickinson

Jodrell Bank Centre for Astrophysics, Alan Turing Building, School of Physics & Astronomy,
The University of Manchester,
Oxford Road, Manchester, M13 9PL, U.K.
email: `lucas.olivari@postgrad.manchester.ac.uk`

Abstract. Intensity mapping (IM) is a new observational technique to survey the large-scale structure of matter using spectral emission lines. IM observations are contaminated by instrumental noise and astrophysical foregrounds. The foregrounds are at least three orders of magnitude larger than the searched signals. In this work, we apply the Generalized Needlet Internal Linear Combination (GNILC) method to subtract radio foregrounds and to recover the cosmological HI and CO signals within the IM context. For the HI IM case, we find that GNILC can reconstruct the HI plus noise power spectra with 7.0% accuracy for $z = 0.13 - 0.48$ ($960 - 1260\,\mathrm{MHz}$) and $\ell \lesssim 400$, while for the CO IM case, we find that it can reconstruct the CO plus noise power spectra with 6.7% accuracy for $z = 2.4 - 3.4$ ($26 - 34\,\mathrm{GHz}$) and $\ell \lesssim 3000$.

Keywords. methods: data analysis, cosmology: observations, cosmology: large-scale structure of universe

1. Introduction

In recent years, a new way of performing redshift surveys in the radio that does not require the detection of any particular galaxy, the so-called intensity mapping (IM) technique, has been proposed (see Kovetz *et al.* (2017) for a review of the technique). Its first rationale has been made for the 21 cm spectral line of neutral hydrogen (HI). The fact that the 21 cm line is so weak means that it requires large collecting areas so that galaxies with significant luminosity can be detected in an acceptable period of time (Wilkinson 1991). Such large experimental structures, however, can be avoided if the full intensity field $T(\theta, \phi, \nu)$ of the HI line is used to measure, on the physical scales of interest, the spatial fluctuations that arise from emitting HI galaxies.

The IM technique, however, is not limited to the HI line. It can also be used with other spectral lines, such as the [CII], Lyα, Hα, and CO. Observations with these lines can then be used in conjunction with HI observations to either minimize systematics biases and uncertainties or to enhance the understanding of certain types of phenomena such as star and black hole formation (Fonseca *et al.* 2017).

The main advantages of IM when compared to galaxy surveys are four: (i) a large volume of the Universe can be surveyed within a relatively short observing time since no galaxy needs to be resolved; (ii) all the signal is recorded, including the gas in between galaxies; (iii) since high angular resolution is not required, it is more economical than traditional surveys; and (iv) the redshift comes for free due to known Doppler shift of the observed emission line.

The IM technique, however, does have some challenges. The main difficulties are the astrophysical contamination (the so-called foregrounds) and the systematics effects that

are present in the observed signal. At GHz frequencies, the most relevant foregrounds are the Galactic synchrotron emission and the background emission of extragalactic point sources (Ichiki 2014). These emissions are at least three orders of magnitude larger than the searched signals. Examples of sources of systematic errors in radio astronomy include: $1/f$ noise, standing waves, calibration errors, polarization leakage, atmosphere absorption or scattering, and radio-frequency interference (RFI) (Condon & Ransom 2016).

In this work, we apply the Generalized Needlet Internal Linear Combination (GNILC) method to both HI and CO IM simulated data and study its ability to clean the contaminating foregrounds and reconstruct the cosmological signals. GNILC is a non-parametric method that has been developed by Remazeilles *et al.* (2011) in the cosmic microwave background (CMB) context and that has been adapted to be used in the HI IM context by Olivari *et al.* (2016). It has also been successfully applied to *Planck* data to disentangle the cosmic infrared background (CIB) from the Galactic thermal dust emission (Planck Collaboration *et al.* 2016). As for the experiments we use the BAO from Integrated Neutral Gas Observations (BINGO) (Battye *et al.* 2013) for our HI IM simulations and the CO Mapping Array Pathfinder (COMAP) (Li *et al.* 2016) for our CO IM simulations.

2. The GNILC Method

The GNILC method can be divided into two main steps. First, using a prior on the wanted signal power spectra, the local ratio between the wanted signal and the total signal is determined. This ratio is then used to perform a Principal Component Analysis (PCA) of the data so that the effective dimension of the wanted signal subspace can be determined. The second step is a multimensional Internal Linear Combination (ILC) filter, which is done within the wanted signal subspace found in the previous step. The ILC filter minimizes the foreground contamination that may still be present in this subspace. In the PCA step, the number of principal components of the observation covariance matrix is estimated locally both in space and in angular scale by using a wavelet (needlet) decomposition of the observations. To make the selection of the principal components of the observation covariance matrix, a statistical criterion, the Akaike Information Criterion, is used. For more details of the formalism, see Olivari *et al.* (2016).

3. Application to HI IM

The GNILC method presented above is now applied to BINGO, which is a single-dish experiment that aims to map the HI emission at frequencies from 960 to 1260 MHz ($z = 0.13 - 0.48$) (Battye *et al.* 2013). To have a stronger test of the GNILC ability to clean the foregrounds, instead of using the limited sky coverage of BINGO, we use full-sky maps, so that all the complexity of the observed sky is present in the resulting analysis. For the foregrounds we assume them to be given by the Galactic synchrotron and free-free emissions and by the 1.4 GHz population of point sources. The models used to simulate these emissions can be found in Olivari *et al.* (2018a) and references therein. In generating our maps, we use the HEALPix pixelization scheme (Górski *et al.* 2005). We use a Galactic mask, the GAL070 *Planck* HFI Galactic mask, which limits us to 71% of the sky. The number of frequency channels is 40.

An example of a GNILC reconstructed HI plus noise power spectrum is shown in Fig. 1. This figure shows two properties of GNILC: (i) it does not significantly affect the wanted signal statistics, i.e., there is no artifact in the reconstructed power spectrum; and (ii) it underestimates by a few percentage the wanted signal. The consequence of these two facts to the resulting cosmological analysis is discussed in Olivari *et al.* (2018a).
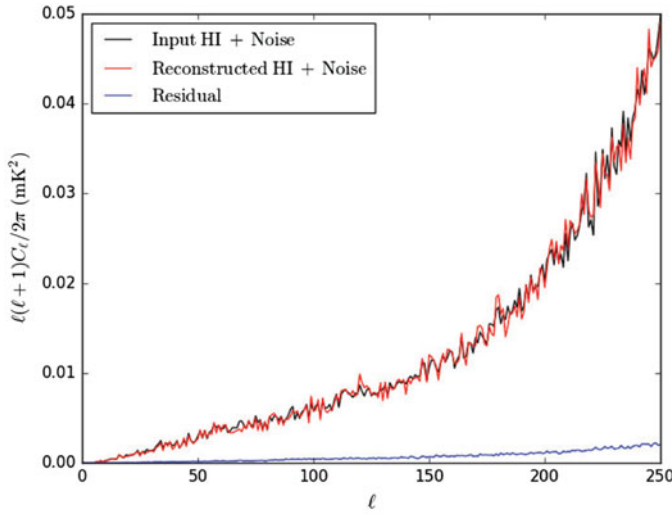
Figure 1. Angular power spectra for the input HI plus noise signal (*black*), the GNILC reconstructed HI plus noise signal (*red*), and the residual (input minus reconstructed HI plus noise) signal (*blue*) at 1.095 GHz ($z \simeq 0.30$). For this particular channel, $N_{\mathrm{GNILC,\,abs}} = 6.4\%$.
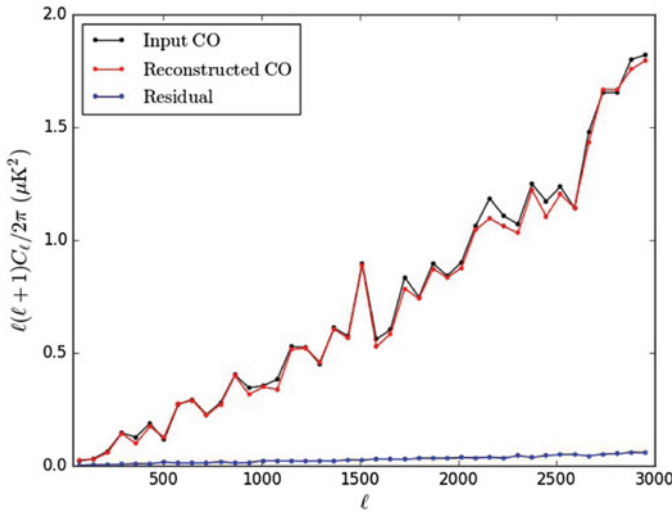

Figure 2. Angular power spectra for the input CO signal (*black*), the GNILC reconstructed CO signal (*red*), and the residual signal (*blue*) at 29.8 GHz ($z \simeq 2.9$). For this particular channel, $N_{\mathrm{GNILC,\,abs}} = 3.9\%$.

To quantify the GNILC performance we use the average absolute normalized difference between the input, $C_\ell^S(\nu_i)$, and the reconstructed power spectrum, $C_\ell^R(\nu_i)$, of the wanted signal,

$$N_{\mathrm{GNILC,\,abs}} = \frac{1}{N_\ell n_{\mathrm{ch}}} \sum_i^{n_{\mathrm{ch}}} \sum_\ell^{\ell_{\mathrm{max}}} \left| \frac{C_\ell^R(\nu_i) - C_\ell^S(\nu_i)}{C_\ell^S(\nu_i)} \right|. \tag{3.1}$$

This quantity measures the absolute dispersion around the mean. For the present exercise, we find it to be $N_{\mathrm{GNILC,\,abs}} = 7.0\%$. As the brightness temperature fluctuations of the foregrounds are (at 1 degree resolution) $\Delta T_{\mathrm{b}} \simeq 700$ mK and the HI plus noise, $\Delta T_{\mathrm{b}} \simeq 0.2$ mK, we can say that 99.97% of the observed signal has been removed to a 7.0% level.

4. Application to CO IM

We now apply the GNILC method to COMAP, which is a single-dish experiment that aims to map the CO emission at frequencies from 26 to 34 GHz, which corresponds to $z = 2.4 - 3.4$ for the CO(1–0) line (Li *et al.* 2016). As the field-of-view of COMAP is very small ($\Omega_{\mathrm{sur}} = 25 \deg^2$ for the full experiment), the use of Cartesian projections are more efficient than HEALPix projections. The GNILC machinery, therefore, has to be slightly modified in this case, i.e., instead of localizing the observed signal in the spherical harmonic space, we have to localize it in the standard Fourier space. In this exercise, the foregrounds are given by Galactic synchrotron, free-free, and anomalous microwave emission (AME), CMB, and the 30 GHz population of point sources. The COMAP observational parameters are those of the full experiment (Li *et al.* 2016). In Fig. 2, we show the GNILC reconstructed CO power spectrum for a case without the inclusion of thermal noise. In this exercise, we have $N_{\mathrm{GNILC, abs}} = 6.9\%$. If we include thermal noise and reconstruct the CO plus noise signal, we obtain $N_{\mathrm{GNILC, abs}} = 6.7\%$. The application of GNILC to COMAP is going to be discussed with more details in Olivari *et al.* (2018b)

5. Final Remarks

The above results show that GNILC is a powerful tool for cleaning the foregrounds, which are at least three orders of magnitude larger than the searched signals. For both HI and CO IM, it is capable of recovering the wanted signal power spectra with a few percent accuracy over a large range of multipoles and frequencies. It is also worth to say that, although not shown here, GNILC can be applied to other scientific contexts, such as CMB B-modes and epoch of reionization (EoR) observations.

6. Acknowledgments

References

Battye, R. A., Browne, I. W. A., Dickinson, C., Heron, G., Maffei, B., & Pourtsidou, A. 2013, *MNRAS*, 434, 1239

Condon, J. J. & Ransom, S. M. 2016, *Essential Radio Astronomy*, Princeton University Press

Fonseca, J., Silva, M. B., Santos, M. G., & Cooray, A. 2017, *MNRAS*, 464, 1948

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. 2005, *ApJ*, 622, 759

Ichiki, K. 2014, *Progress of Theoretical and Experimental Physics*, 06B109

Kovetz, E. D. *et al.* 2017, *Submitted to Physics Reports [arXiv: 1709.09066]*

Li, T. Y., Wechsler, R. H., Devaraj, K., & Church, S. E. 2016, *ApJ*, 817, 169

Olivari, L. C., Remazeilles, M., Dickinson, C. 2016, *MNRAS*, 456, 2749

Olivari, L. C., Dickinson, C., Battye, R. A., Ma, Y.-Z., Costa, A. A., Remazeilles, M., & Harper, S. 2018a, *MNRAS*, 473, 4242

Olivari, L. C., *et al.* 2018b, *in Preparation*

Planck Collaboration *et al.* 2016, *A&A*, 596, A109

Remazeilles, M., Delabrouille, J., & Cardoso, J.-F 2011, *MNRAS*, 418, 467

Wilkinson, P. N. 1991, *in Astronomical Society of the Pacific Conference Series, Vol. 19, IAU Colloq. 131: Radio Interferometry. Theory, Techniques, and Applications*, 428