



Universiteit  
Leiden  
The Netherlands

## The first power spectrum limit on the 21-cm signal of neutral hydrogen during the Cosmic Dawn at $z = 20-25$ from LOFAR

Gehlot, B.K.; Mertens, F.G.; Koopmans, L.V.E.; Brentjens, M.A.; Zaroubi, S.; Ciardi, B.; ... ; Yatawatta, S.

### Citation

Gehlot, B. K., Mertens, F. G., Koopmans, L. V. E., Brentjens, M. A., Zaroubi, S., Ciardi, B., ... Yatawatta, S. (2019). The first power spectrum limit on the 21-cm signal of neutral hydrogen during the Cosmic Dawn at  $z = 20-25$  from LOFAR. *Monthly Notices Of The Ras* (0035-8711), 488(3), 4271-4287. doi:10.1093/mnras/stz1937

Version: Accepted Manuscript

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/84758>

**Note:** To cite this publication please use the final published version (if applicable).

# The first power spectrum limit on the 21-cm signal of neutral hydrogen during the Cosmic Dawn at $z = 20 - 25$ from LOFAR

B. K. Gehlot<sup>1\*</sup>, F. G. Mertens<sup>1</sup>, L. V. E. Koopmans<sup>1†</sup>, M. A. Brentjens<sup>2</sup>,  
 S. Zaroubi<sup>1,3</sup>, B. Ciardi<sup>4</sup>, A. Ghosh<sup>5,6</sup>, M. Hatef<sup>1,2</sup>, I. T. Iliev<sup>7</sup>, V. Jelić<sup>8</sup>,  
 R. Kooistra<sup>1</sup>, F. Krause<sup>1,9</sup>, M. Mitra<sup>1</sup>, M. Mevius<sup>2</sup>, G. Mellema<sup>10</sup>, A. R. Offringa<sup>2</sup>,  
 V. N. Pandey<sup>1,2</sup>, M. B. Silva<sup>1</sup>, J. Schaye<sup>11</sup>, A. M. Sardarabadi<sup>1</sup>, H. K. Vedantham<sup>2</sup>,  
 and S. Yatawatta<sup>1,2</sup>

<sup>1</sup>*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700AV Groningen, the Netherlands*

<sup>2</sup>*ASTRON, PO Box 2, 7990AA Dwingeloo, The Netherlands*

<sup>3</sup>*Department of Natural Sciences, The Open University of Israel, 1 University Road, PO Box 808, Ra'anana 4353701, Israel*

<sup>4</sup>*Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Straße 1, 85748 Garching, Germany*

<sup>5</sup>*Department of Physics, University of the Western Cape, Cape Town 7535, South Africa*

<sup>6</sup>*SKA South Africa, 3rd Floor, The Park, Park Road, Pinelands, 7405 South Africa*

<sup>7</sup>*Astronomy Centre, Department of Physics and Astronomy, Pevensey II Building, University of Sussex, Brighton BN1 9QH, U.K.*

<sup>8</sup>*Ruder Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia*

<sup>9</sup>*Department of Physics and Astronomy, University College London, Gower Place, London WC1E 6BT, U.K.*

<sup>10</sup>*Department of Astronomy and Oskar Klein Centre for Cosmoparticle Physics, AlbaNova, Stockholm University, SE-106 91 Stockholm, Sweden*

<sup>11</sup>*Leiden Observatory, Leiden University, PO Box 9513, 2300RA Leiden, The Netherlands*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

Observations of the redshifted 21-cm hyperfine line of neutral hydrogen from early phases of the Universe such as Cosmic Dawn and the Epoch of Reionization promise to open a new window onto the early formation of stars and galaxies. We present the first upper limits on the power spectrum of redshifted 21-cm brightness temperature fluctuations in the redshift range  $z = 19.8 - 25.2$  (54 – 68 MHz frequency range) using 14 hours of data obtained with the LOFAR-Low Band Antenna (LBA) array. We also demonstrate the application of a multiple pointing calibration technique to calibrate the LOFAR-LBA dual-pointing observations centered on the North Celestial Pole and the radio galaxy 3C220.3. We observe an unexplained excess of  $\sim 30 - 50\%$  in Stokes  $I$  noise compared to Stokes  $V$  for the two observed fields, which decorrelates on  $\gtrsim 12$  seconds and might have a physical origin. We also show that enforcing smoothness of gain errors along frequency during calibration reduces the additional noise in Stokes  $I$  compared Stokes  $V$  introduced by the calibration on sub-band level. After subtraction of smooth foregrounds, we achieve a systematics-limited  $2\sigma$  upper limit on the 21-cm power spectrum of  $\Delta_{21}^2 < (14561 \text{ mK})^2$  at  $k \sim 0.038 \text{ h cMpc}^{-1}$  and  $\Delta_{21}^2 < (14886 \text{ mK})^2$  at  $k \sim 0.038 \text{ h cMpc}^{-1}$  for the 3C220 and NCP fields respectively. Both upper limits are consistent with each other and with the thermal noise in the data.

**Key words:** dark ages, reionization, first stars – techniques: interferometric – methods: statistical – methods: data analysis – radio lines: general – diffuse radiation

## 1 INTRODUCTION

After the Epoch of Recombination around redshift  $z \sim 1100$ , the Universe entered the ‘Dark Ages’ era during which it was completely neutral and devoid of any radiation sources.

\* E-mail: gehlot@astro.rug.nl (BKG)

† E-mail: koopmans@astro.rug.nl (LVEK)

During this period, small perturbations in matter density grew under gravitational instability, and matter started to accumulate in localized over-density peaks. The formation of the first luminous objects (stars and galaxies) in these over-dense regions marked the beginning of the so-called Cosmic Dawn (CD) era spanning the redshift range  $30 > z > 12$  (Pritchard & Furlanetto 2007). X-ray and Ultraviolet radiation from the first stars and galaxies began to heat and ionize the neutral hydrogen (HI hereafter) in the surrounding Inter-Galactic Medium (IGM), starting off the Epoch of Reionization (EoR) ( $12 > z > 6$ ) during which HI in the IGM transitioned from being fully neutral to ionized (Madau et al. 1997).

The redshifted 21-cm signal corresponding to the hyperfine transition of HI has been identified as an excellent probe of the HI distribution in the IGM during the CD and the EoR (Madau et al. 1997; Shaver et al. 1999; Furlanetto et al. 2006; Pritchard & Loeb 2012; Zaroubi 2013). A number of ongoing and upcoming experiments, such as the LOw Frequency ARray<sup>1</sup> (LOFAR; van Haarlem et al. 2013), the Giant Meterwave Radio Telescope<sup>2</sup> (GMRT; Paciga et al. 2011), the Murchison Widefield Array<sup>3</sup> (MWA; Tingay et al. 2013; Bowman et al. 2013), the Precision Array for Probing the Epoch of Reionization<sup>4</sup> (PAPER; Parsons et al. 2010), the Hydrogen Epoch of Reionization Array<sup>5</sup> (HERA; DeBoer et al. 2017), NENUFAR<sup>6</sup> (New Extension in Nançay Upgrading loFAR; Zarka et al. 2012), and the Square Kilometer Array<sup>7</sup> (SKA; Mellema et al. 2013; Koopmans et al. 2015) are seeking to detect the brightness temperature fluctuations in the cosmological 21-cm signal using statistical methods e.g. the power spectrum. Complementary to these 21-cm power spectrum measurement experiments, several efforts such as the Experiment to Detect the Global Epoch of Reionization Signature (EDGES; Bowman et al. 2018), the Large-aperture Experiment to Detect the Dark Ages (LEDA; Bernardi et al. 2016), the Shaped Antenna measurement of the background RAdio Spectrum 2 (SARAS 2; Singh et al. 2017), the Sonda Cosmológica de las Islas para la Detección de Hidrógeno Neutro (SCI-HI; Voytek et al. 2014), the Probing Radio Intensity at high  $z$  from Marion (PRIZM; Philip et al. 2018), and the Netherlands-China Low frequency Explorer<sup>8,9</sup> (NCLE) are seeking to measure the sky-averaged spectrum of the 21-cm signal.

At present, several instruments targeting the EoR redshifts have placed upper limits on the brightness temperature power spectrum of the redshifted 21-cm signal. Paciga et al. (2013) provided the first  $2\sigma$  upper limit on the brightness temperature of  $\Delta_{21}^2 < (248 \text{ mK})^2$  at wavenumber  $k \approx 0.5 \text{ h cMpc}^{-1}$  at redshift  $z = 8.6$  using the GMRT. Beard-

sley et al. (2016) used MWA to set a  $2\sigma$  upper limit of  $\Delta_{21}^2 < (164 \text{ mK})^2$  at  $k \approx 0.27 \text{ h cMpc}^{-1}$  at  $z = 7.1$ . The PAPER project also provided an upper limit of  $\Delta_{21}^2 < (22 \text{ mK})^2$  in the wavenumber range  $0.15 \leq k \leq 0.5 \text{ h cMpc}^{-1}$  at  $z = 8.4$  (Ali et al. 2015), but have recently retracted their claim due to issues with their analysis strategy (see the erratum Ali et al. 2018). The tightest  $2\sigma$  upper limit on the 21-cm power spectrum yet is  $\Delta_{21}^2 < (79.6 \text{ mK})^2$  at  $k \approx 0.053 \text{ h cMpc}^{-1}$  in the redshift range  $z = 9.6 - 10.6$  and was provided by Patil et al. (2017) using the LOFAR High Band Antenna (HBA) array. Instruments such as HERA, NENUFAR, and SKA-low which can potentially probe the CD redshifts are now in hardware roll-out stages (the latter is still in the development stage). Ewall-Wice et al. (2016) used low frequency MWA observations (75 – 113 MHz) to place an upper limit of  $\Delta^2 < (10^4 \text{ mK})^2$  at  $k \approx 0.5$  on the power spectrum of the brightness temperature fluctuations of the 21-cm signal in the redshift range  $12 \lesssim z \lesssim 18$ , which in most models corresponds to the epoch of X-ray heating during the CD (see e.g. Glover & Brand 2003; Fialkov & Barkana 2014; Ross et al. 2017).

In this work, we explore, for the first time, the possibility of observing the redshifted 21-cm signal from the CD era using the LOFAR-Low Band Antenna (LBA) array which observes in the 30 – 90 MHz frequency range. We use LOFAR-LBA dual pointing observations of the North Celestial Pole (NCP field hereafter) and an adjacent field centered on the 3C220.3 radio galaxy (3C220 field hereafter), which is  $\sim 7^\circ$  away from the NCP, to study the challenges (systematic biases) in CD studies with the LOFAR-LBA and to set the first upper limits on the 21-cm brightness temperature power spectrum in the redshift range  $z = 19.8 - 25.2$ . We also demonstrate the application of a novel dual-pointing calibration strategy to calibrate the LOFAR-LBA data, and the application of Gaussian Process Regression (GPR) as a powerful foreground removal technique in CD experiments.

The paper is organized as follows: in Section 2, we briefly describe the LOFAR-LBA system, the observational setup and preprocessing steps. In Section 3, we describe the multi-beam calibration strategy to calibrate the LOFAR-LBA data. In Section 4, we assess the noise in the observed data and address the systematic biases, such as excess noise in Stokes I versus V using various statistical methods. We describe Gaussian Process Regression (GPR) in Section 5 and its application in removing residual foregrounds in LOFAR-LBA data. In Section 6, we determine the power spectra for both fields and derive upper limits on the 21-cm power spectrum in the redshift range  $z = 19.8 - 25.2$ . Finally, in Section 7, we summarize the work and discuss future prospects.

## 2 OBSERVATIONS AND PREPROCESSING

We used the LOFAR-LBA system with dual pointing setup to simultaneously observe the NCP field and the 3C220 field, which is  $\sim 7^\circ$  away from the NCP. The NCP is the primary target field of the LOFAR-EoR KSP and has been used to set the first upper limits on the EoR power spectrum using LOFAR (see Patil et al. 2017). The observational setup and preprocessing steps are described in the following subsections.

<sup>1</sup> <http://www.lofar.org/>

<sup>2</sup> <http://gmrt.ncra.tifr.res.in/>

<sup>3</sup> <http://www.mwatelescope.org/>

<sup>4</sup> <http://eor.berkeley.edu/>

<sup>5</sup> <http://reionization.org/>

<sup>6</sup> <https://nenufar.obs-nancay.fr/>

<sup>7</sup> <http://skatelescope.org/>

<sup>8</sup> <https://www.ru.nl/astrophysics/research/radboud-radio-lab-0/projects/netherlands-china-low-frequency-explorer-ncle/>

<sup>9</sup> <https://www.astron.nl/r-d-laboratory/ncle/netherlands-china-low-frequency-explorer-ncle>

<sup>9</sup> <https://www.astron.nl/r-d-laboratory/ncle/netherlands-china-low-frequency-explorer-ncle>

## 2.1 LOFAR-Low Band Array

The LOFAR-LBA consists of 38 stations spread across the Netherlands, providing shortest baseline lengths of  $\sim 80$  m and longest baseline lengths of  $\sim 100$  km. Out of these 38 stations, 24 stations (known as core stations) are spread within a core of 2 km radius, providing a densely sampled  $uv$ -plane. The remaining 14 stations (known as remote stations) are spread across the North-Eastern part of the Netherlands. Each LOFAR station consists of 96 low band dual-polarization dipole antennas randomly spread within an area of 81 m diameter. The voltages measured with the cross dipoles are digitized using a 200 MHz sampling clock covering the frequency range of 10-90 MHz. The digitized data is beam-formed to produce a digitally steerable station beam. At a given time, only 48 out of 96 dipoles can be combined in the beam-former. This allows a user to choose from three different station configurations in LOFAR-LBA mode viz: `LBA_INNER` where the 48 innermost dipoles (array width  $\sim 30$  m) are beam-formed, `LBA_OUTER` where the 48 outermost dipoles (array width  $\sim 81$  m) are beam-formed, and `LBA_SPARSE` where half of the innermost 48 dipoles, plus half of the outermost 48 dipoles (array width  $\sim 81$  m) are beam-formed. Each configuration results in a specific Field of View (FoV) as well as different sensitivity due to mutual coupling between the dipoles. The LOFAR-LBA system has an instantaneous bandwidth of 96 MHz. However, multiple pointings in the sky can be traded against the observable bandwidth depending on the number of pointings. In case of two pointings, the bandwidth is reduced to 48 MHz per pointing. Readers may refer to [van Haarlem et al. \(2013\)](#) for more information about the observation capabilities of LOFAR.

## 2.2 Observations

We use 14 hours of synthesis observation data of the NCP and the 3C220 fields, which were observed simultaneously with dual beam pointings using `LBA_OUTER` mode of the LOFAR-LBA system. The data were recorded during LOFAR observation Cycle 6 (ID:L557804, November 4-5, 2016). The observational details of the data are summarized in Table 1. The digitized data from beam-formed stations were correlated with 1 second time resolution and 3 kHz frequency resolution. The recorded data consists of 244 sub-bands for each field within the frequency range of 38-86 MHz. Each sub-band has a width of 195.3 kHz and consists of 64 channels. The recorded correlations (XX, XY, YX and YY) are stored in a Measurement Set (MS). The raw data volume for each field is  $\sim 18$  Terabytes and is preprocessed to reduce the data volume, as described in the next section.

## 2.3 Data selection and preprocessing

LOFAR-LBA has lower sensitivity and a relatively high RFI corruption level for frequencies above 70 MHz. Therefore, we used only 33 MHz bandwidth with the frequency range 39-72 MHz for preprocessing and further analysis. We used the standard LOFAR software pipeline (see e.g. LOFAR imaging

**Table 1.** Observational details of the data.

Parameter	value
Telescope	LOFAR LBA
Observation cycle and ID	Cycle 6, L557804
Antenna configuration	<code>LBA_OUTER</code>
Number of stations	38 (NL stations)
Observation start time (UTC)	Nov 4, 2016; 16:21:44
Number of pointings	2
Phase center ( $\alpha, \delta$ ; J2000):	
NCP field	00h00m00s, +90°00'00"
3C220 field	09h39m23s, +83°15'26"
Duration of observation	14 hours
Minimum frequency	38.08 MHz
Minimum frequency	85.54 MHz
Target bandwidth	48 MHz
Primary Beam FWHM	3.88° at 60 MHz
Field of View	12 deg <sup>2</sup> at 60 MHz
SEFD	$\sim 25$ kJy at 60 MHz
Polarization	Linear X-Y
Time, frequency resolution:	
Raw Data	1 s, 3 kHz
After flagging step 1	2 s, 12 kHz (archived)
After flagging step 2	2 s, 61 kHz

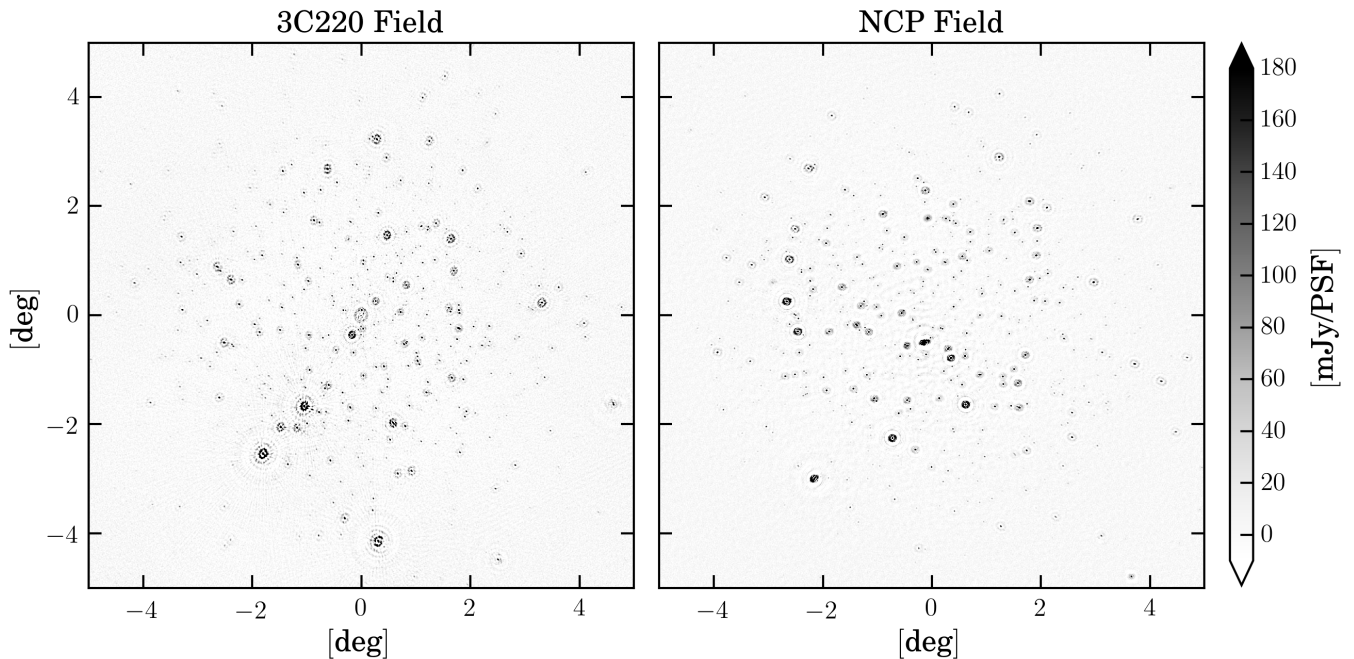
cookbook<sup>10</sup>) for preprocessing the observed raw data. Processing steps include RFI-excision and averaging the data. Flagging of RFI corrupted data is performed on the highest resolution data (1 second, 3 kHz) to minimize information loss. We use the `AOFlagger` software ([Offringa et al. 2010, 2012](#)) to flag RFI corrupted data. Two channels on both edges of every sub-band were also discarded to avoid edge effects due to the polyphase filter. The remaining data was averaged in frequency and time to an intermediate-resolution of 12 kHz and 2 seconds, resulting in 15 channels per sub-band. This intermediate resolution data is archived for future use. To reduce the data volume further, it was averaged in frequency to 61 kHz and the auto-correlations were also flagged. The resulting data consists of 3 channels of 61 kHz each per sub-band and has a time resolution of 2 seconds. We flagged the remote station RS503LBA in all sub-bands for both fields because of its proximity to a windmill, which causes strong RFI in the visibilities of the station. We also observed that CS302LBA had poor gain upon inspecting the visibilities and flagged it for both fields. The flagging and averaging was performed separately on both 3C220 and NCP field datasets, although some correlation is obviously expected.

## 3 CALIBRATION SCHEME

The visibilities recorded by LOFAR are corrupted by the instrumental (complex station gains, primary beam, instrumental bandpass, clock-drift etc.) and environmental (ionosphere) factors. Calibration of the LOFAR-LBA system involves estimating the errors that corrupt the measured visibilities, and to obtain an accurate estimate of the true visibilities from observed data. Calibration of LOFAR-LBA

<sup>10</sup> <https://www.astron.nl/radio-observatory/lofar/lofar-imaging-cookbook>





**Figure 1.** Left and right panels show Stokes  $I$  continuum images (39 – 72 MHz) of the 3C220 and NCP field respectively, after DI calibration. The images are not cleaned. These images were produced using  $\leq 2000\lambda$  baselines with the Briggs -0.1 weighting scheme. The observed image rms is  $\sigma_{\text{rms}} \sim 7$  mJy for the 3C220 field and  $\sim 5.5$  mJy for the NCP field respectively. These values are still  $\sim 10$  times higher than the expected rms ( $\sim 0.7$  mJy) calculated using SEFD (System Equivalent Flux Density) estimates for LOFAR-LBA. The values of  $\sigma_{\text{rms}}$  can be calculated from SEFD using the relation:  $\sigma_{\text{rms}} = \text{SEFD} / \sqrt{2N(N-1)\Delta\nu\Delta t}$ , where SEFD  $\sim 30$  kJy at 55 MHz,  $N = 29$  (corresponding to  $2000\lambda$  baseline range),  $\Delta\nu = 33$  MHz,  $\Delta t = 0.9 \times 14$  hours (assuming flagged data at 10% level).

data involves two major steps: (a) Direction Independent (DI) calibration and, (b) Direction Dependent (DD) calibration. DI calibration involves estimation of a single instrumental gain (represented by a complex  $2 \times 2$  Jones matrix) for each beam-formed station, and DD calibration accounts for the direction dependent errors arising from wave propagation effects through the ionosphere and the primary beam. We use SAGECa1-CO<sup>11</sup> to perform the major calibration steps. SAGECa1-CO performs calibration in a distributed way using consensus optimization (Boyd et al. 2011), which is an effective way to improve the quality of calibration of radio interferometric data. In SAGECa1-CO, the calibration problem is transformed into consensus optimization by adding frequency smoothness of systematic errors as a constraint. It uses an Alternating Direction Method of Multipliers (ADMM) algorithm to reach convergence. Readers may refer to Yatawatta (2015); Yatawatta (2016); Yatawatta et al. (2017); Yatawatta (2018) for a detailed description of the SAGECa1-CO algorithm and its capabilities.

Up on inspection of the raw visibilities, we observed that Cas A ( $\sim 30^\circ$  away from NCP) and Cyg A ( $\sim 50^\circ$  away from the NCP) superpose significant side-lobes onto both fields. It is crucial to subtract these sources before performing DI calibration to avoid errors due to these side-lobes. We use DD-calibration in SAGECa1-CO to subtract Cas A and Cyg A. Gehlot et al. (2018) (G18 hereafter) showed that the residuals after subtraction of bright sources such as Cas A and Cyg A

are significant as well as incoherent over timescales of a few minutes, depending on the strength of ionospheric scintillations. Therefore, we use a solution time and frequency interval of 30 seconds and 61 kHz to subtract Cas A and Cyg A, which is optimized to incorporate ionospheric effects while maintaining a decent signal-to-noise ratio ( $\gtrsim 10$ ) for the given solution interval. We use the Cas A and Cyg A shapelet models<sup>12</sup> as an input model for calibration and subtraction. The subtraction was performed individually on both fields.

The two fields, 3C220 and NCP, given their different pointings and gain solutions, have slightly different morphologies. The 3C220 field consists of a reasonably bright source located at the phase center (the 3C220.3 radio galaxy with a flux of  $\sim 38$  Jy at 74 MHz (Cohen et al. 2007)) which can be utilized as a bandpass calibrator, making calibration of the 3C220 field fairly straightforward. However, the NCP field does not have such relatively bright sources near the phase center, which makes it more difficult to calibrate the field. Therefore, we adopt a calibration strategy where we calibrate the 3C220 field first and then use the output calibration products to calibrate the NCP field, given that the bandpass calibration solutions should be similar between the fields because of the same electronics, and that any effect of

<sup>12</sup> Cas A and Cyg A models were derived from wide-band LOFAR-LBA and HBA observations of Cas A and Cyg A. Each source has about 200 components (shapelets and point). See Yatawatta (2011) for more details.

<sup>11</sup> <http://sagecal.sourceforge.net/>

the beam should be spectrally smooth near the phase center. A similar technique to calibrate the LOFAR-LBA data to study the ionospheric effects is shown in de Gasperin et al. (in preparation) and de Gasperin et al. (2018). Similar type of calibration strategies are more common in radio survey experiments, although in those cases it is often required to switch between sources in time.

### 3.1 Calibrating the 3C220 field

To calibrate the 3C220 field, we use a calibration strategy similar to that discussed in G18. The 3C220.3 radio galaxy is a double lobed source of  $\sim 8$  arcsec extent, but it is unresolved with the LOFAR-LBA array which has a maximum resolution of  $\sim 15$  arcsec. Therefore, we use a single point source representing 3C220.3 with 38 Jy flux at 74 MHz (Cohen et al. 2007) and a spectral index of -0.8 as a starting model for DI calibration. The major steps involved in the calibration of the 3C220 field are as follows:

(i) Calibrate the raw visibilities using the 3C220.3 point source model in NDPPP<sup>13</sup> to obtain the station gain solutions with 30 seconds and 61 kHz calibration solution intervals and subsequently apply them to the data. This step is performed separately for each sub-band (without consensus optimization). We include the primary beam<sup>14</sup> in the calibration step in NDPPP. Note that the LOFAR-LBA beam model has only been implemented in NDPPP at present. Hence, it is utilized for primary DI calibration for both fields. Note that we do not exclude any baselines during DI calibration steps for both fields.

(ii) Deconvolve (clean) and image the calibrated visibilities using the WSClean package (Offringa et al. 2014) with the following settings: cleaning threshold =  $0.5\sigma$ , weighting scheme = uniform, imaging baseline range =  $0 - 5000\lambda$ , 4th order polynomial<sup>15</sup> for fitting the source spectrum over 15 points which correspond to averaged flux over 2.2 MHz bands spread within 33 MHz bandwidth. Iterate over step (i) once more using the clean model of 3C220.3 obtained in step (ii) and perform deconvolution to obtain a more accurate 3C220.3 clean model. Further iterations were not required as the model converged.

(iii) Use SAGECa1-C0 to perform DI calibration of raw visibilities and subtract 3C220.3 using consensus optimization (7 iterations and regularization factor of 5) over a 33 MHz frequency range. We provide the final clean model of 3C220.3 obtained after step (ii) as input to SAGECa1-C0 and use a calibration solution interval of 30 seconds and 183.1 kHz. The obtained gain solutions are subsequently applied to the residual visibilities.

(iv) Repeat the deconvolution with the same settings (but with lower clean-mask =  $4\sigma$ ) in step (ii) to clean and image the residual visibilities after step (iii). The output clean model of the radio sources in the field contains 1270 components (points plus Gaussians) with flux  $> 40$  mJy at 55 MHz. We repeated Step (iii) with this updated sky-model to perform DI-calibration and subtraction of 3C220.3 from the visibilities. Using a more complete sky-model in DI calibration allows for the mitigation of calibration errors due to unmodeled sources and produces accurate calibrated visibilities. The gain solutions obtained after this step are later utilized in the calibration of the NCP field.

(v) Use DD-calibration with SAGECa1-C0 to subtract the clean-model obtained in step (iv). SAGECa1-C0 accounts for DD errors by obtaining the gain solutions in multiple directions. It subtracts the sources in each direction by multiplying the obtained gain solutions with the predicted visibilities and subtracting the product from the observed visibilities. We divide the 1270 components into 4 clusters using the K-means clustering algorithm (Kazemi et al. 2013) and use the centers of these clusters as four different directions. These four different directions roughly represent the four quadrants of the primary beam. We use a gain solution interval of 20 minutes and 183.1 kHz and 20 ADMM iterations for each gain solution while keeping the same regularization factor of  $\rho = 5$  (Yatawatta 2016) as in DI calibration. We discard the baselines  $\leq 200\lambda$  in the DD-calibration to avoid any bias due to unmodeled diffuse emission on shorter baselines (see e.g. Patil et al. 2016; Barry et al. 2016; Ewall-Wice et al. 2017; Gehlot et al. 2018 for more details). Using a calibration cut also mitigates the suppression of the 21-cm signal, as shown in Patil et al. (2017) and Mouri Sardarabadi & Koopmans (2018), and we will test this further in future.

(vi) Image the residual visibilities in step (v) with WSClean. We used the following settings: weighting scheme = natural, pixel size = 3 arcmin, Image dimensions =  $300 \times 300$  pixels, imaging baselines =  $15 - 200\lambda$ . Note that we do not deconvolve the final residual images. The output Stokes  $I$ ,  $V$  and Point Spread Function (PSF) images were stored for further analysis. The left panel of figure 1 shows the dirty continuum image of the 3C220 field after DI calibration where the 3C220.3 has been subtracted.

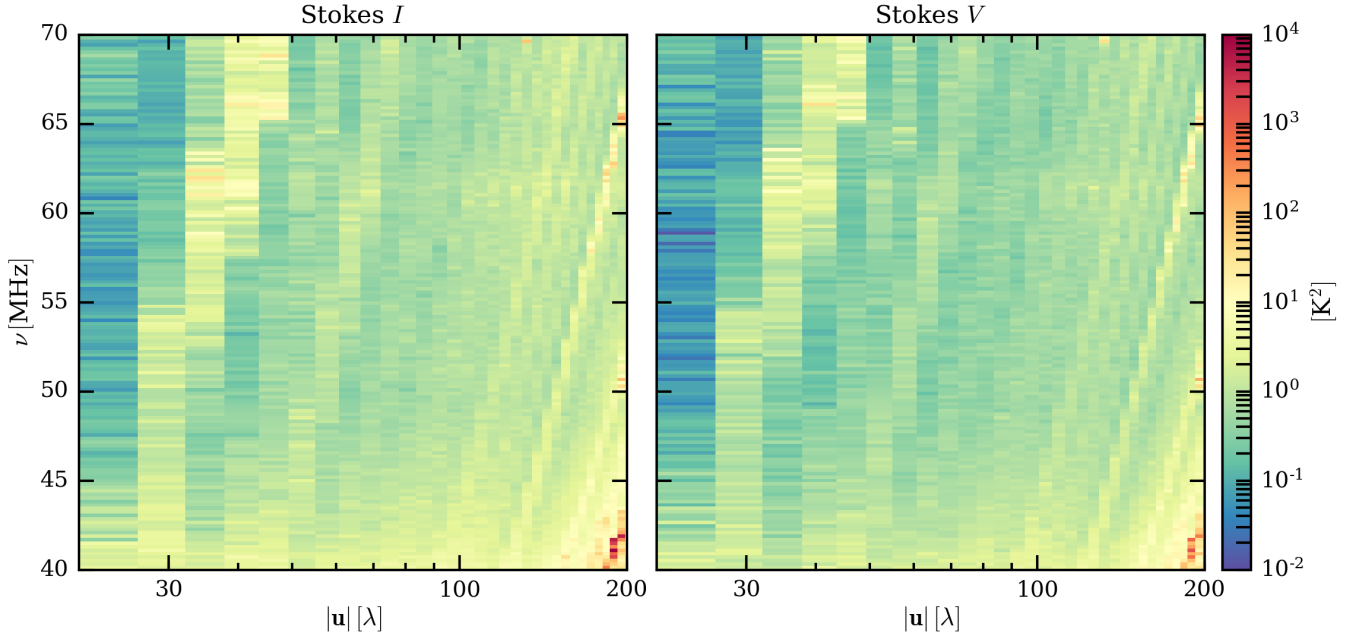
### 3.2 Calibrating the NCP field

The absence of very bright sources makes the NCP field more difficult to calibrate using the strategy we employed for the 3C220 field. Therefore, we utilize a different approach. The NCP field consists of a moderately bright source (3C061.1) which lies at the edge of the primary beam causing the source to exhibit peculiar behavior in its gain solutions. We therefore subtract 3C061.1 from the raw visibilities using DD-calibration with SAGECa1-C0 with the same settings as we employed for the Cas A and Cyg A subtractions. The 3C061.1 input model is adapted from the intrinsic model of 3C061.1 (points + shapelets, at 150 MHz) used in the LOFAR-EoR data processing pipeline (see e.g. Patil et al. 2017). The fluxes in the model were scaled properly to match the flux values quoted in Laing & Peacock (1980) and Hales et al. (1995). After subtraction of 3C061.1, visibilities were calibrated using the following steps:

<sup>13</sup> [http://www.lofar.org/operations/doku.php?id=public:user\\_software:ndppp](http://www.lofar.org/operations/doku.php?id=public:user_software:ndppp)

<sup>14</sup> Current LBA primary beam models are based on Electromagnetic (EM) simulations of the LOFAR-LBA dipoles (private communication with LOFAR Radio Observatory).

<sup>15</sup> Using log polynomials to fit source spectra is unstable in WSClean. Therefore, we use an ordinary 4th order polynomial to fit source spectra. However, SAGECa1-C0 is only compatible with log polynomials. Therefore, we separately fit the source spectra with a 3rd order log-polynomial to make it compatible with SAGECa1-C0.



**Figure 2.** Stokes  $I$  ( $P_{\Delta, I}(|\mathbf{u}|, \nu)$ ) and  $V$  ( $P_{\Delta, V}(|\mathbf{u}|, \nu)$ ) noise spectra for the 3C220 field. Left and right panels correspond to Stokes  $I$  and  $V$  respectively.

(i) Apply the DI gain solution amplitudes of the 3C220 field obtained in step (iv) in section 3.1 to the NCP field visibilities to set the amplitude scale.

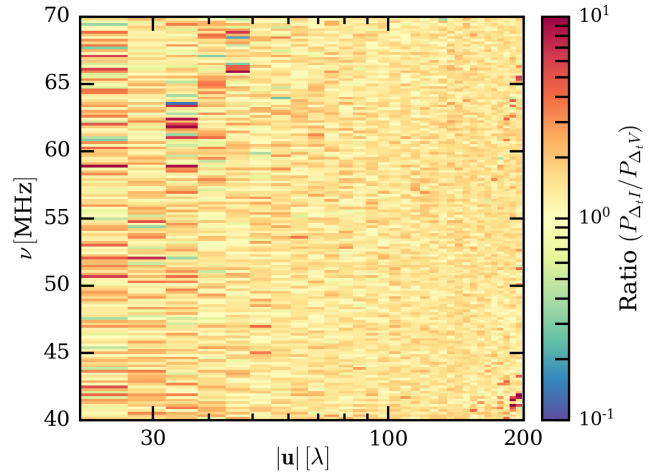
(ii) Deconvolve (clean) and image the resulting visibilities using `WSClean` with the following settings: cleaning threshold =  $0.5\sigma$ , weighting scheme = uniform, imaging baseline range =  $0 - 2000\lambda$ , 2nd order polynomial for fitting the source spectrum over 5 points which corresponds to an averaged flux over 6.6 MHz bands spread over 33 MHz.

(iii) Perform DI calibration of the visibilities with `SAGECal-CO` using consensus optimization (with same settings as in DI calibration of the 3C220 field) over the 33 MHz frequency range. The clean model obtained in step (ii) is provided as input. We use a calibration solution interval of 10 minutes and 183.1 kHz. The obtained gain solutions are subsequently applied to the visibilities. We repeat steps (ii) and (iii) in a self-cal loop with 3 iterations. The final clean model after 3 self-cal iterations contains 1470 components (points plus Gaussians) with flux  $> 40$  mJy at 55 MHz.

(iv) Perform phase calibration using `NDPPP` on the visibilities obtained after step (i). We use the final clean model obtained after step (iii) as an input and choose 30 seconds, 183.1 kHz as the calibration solution interval.

(v) Use DD-calibration with `SAGECal-CO` to subtract the clean-model obtained in step (iii). We divide 1470 components in three clusters representing three directions (which represent three non-overlapping regions within the primary beam) using the K-means clustering algorithm. We use a gain solution interval of 20 minutes and 183.1 kHz and 20 ADMM iterations for each gain solution. We discard the baselines  $\leq 200\lambda$  to avoid errors due to unmodeled diffuse emission on shorter baselines and to avoid signal suppression.

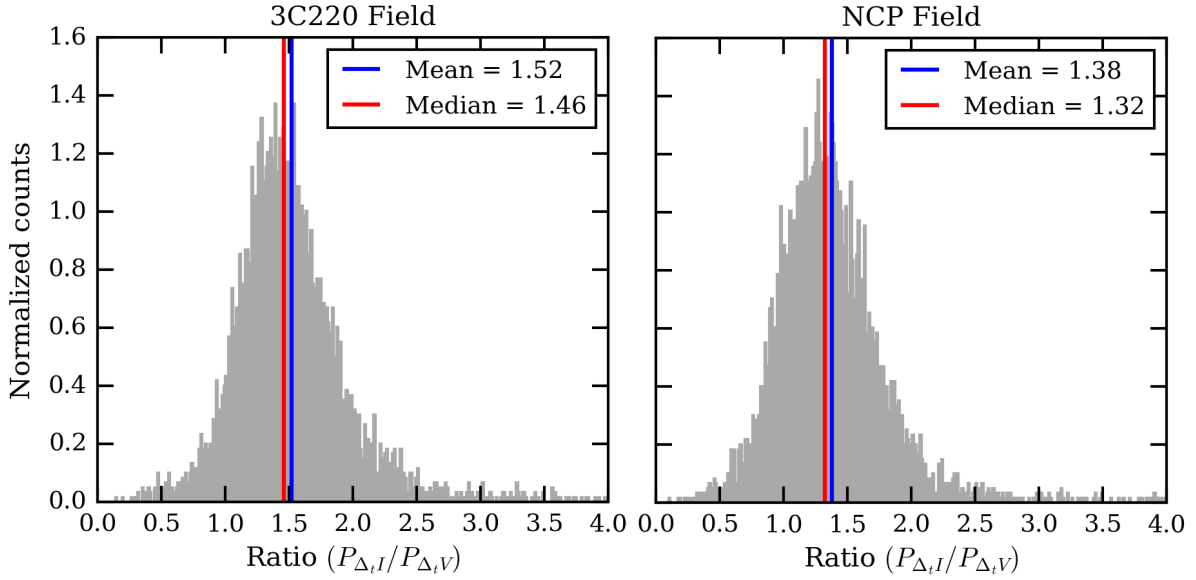
(vi) Image the residual visibilities in step (v) with `WSClean` using the following settings: weighting scheme = nat-



**Figure 3.** the ratio  $P_{\Delta, I}/P_{\Delta, V}$  of the noise spectra shown in figure 2. The ratio is flat except for a few outliers at shorter baselines.

ural, pixel-size = 3 arcmin, Image dimensions =  $300 \times 300$  pixels, imaging baselines =  $15 - 200\lambda$ . The output Stokes  $I$ ,  $V$  and PSF images were stored for further analysis. The right panel of figure 1 shows the dirty continuum image of the NCP field after DI calibration.

At this point we have residual data cubes that are DI calibrated and where the sky model has been subtracted using their DD gain solutions. These residual cubes form the input for subsequent analyses. Note that we only use the beam model during DI-calibration and deconvolution steps. We do not correct the residual images for the primary beam. In the following sections we will discuss these analyses.



**Figure 4.** The left and the right panels show normalized histograms of the distribution of the ratio values  $P_{\Delta,I}/P_{\Delta,V}$  for the 3C220 and the NCP fields. The red and blue vertical lines represent the median and the mean of the distribution respectively. For the 3C220 field, the distribution has a median value of 1.46 and a mean value of 1.54. Similarly, the median and the mean values for NCP field are 1.32 and 1.38 respectively.

#### 4 NOISE STATISTICS IN LOFAR-LBA

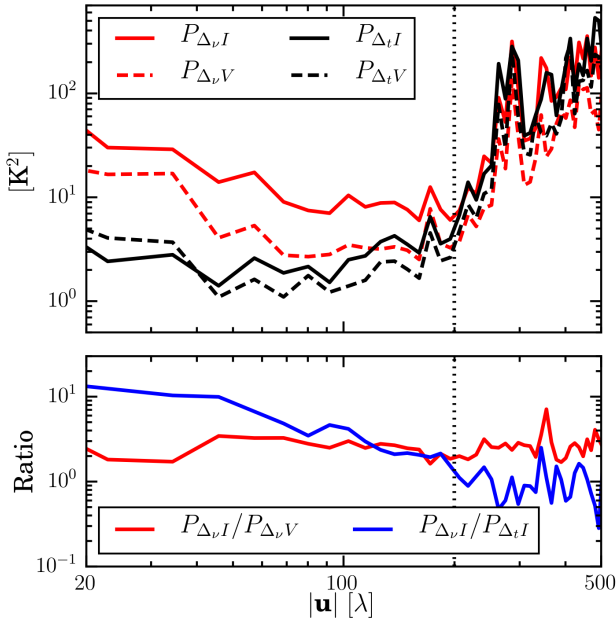
Current estimates of the average Signal Equivalent Flux Density (SEFD) per station of the LOFAR-LBA array are derived from the observations of bright sources at zenith. However, the SEFD of LOFAR varies as a function of angle from the zenith. Therefore, using zenith SEFD estimates to derive the noise on the visibilities and rms in the images (also noise power spectra) typically underestimates the SEFD for the fields away from the zenith. To avoid this bias, we estimate the noise and hence the noise spectrum (in baseline-frequency space) for the 3C220 field from the visibilities. A standard method to estimate the noise on visibilities is to subtract the un-gridded visibilities corresponding for two contiguous time-steps at the highest time resolution. However, this method is not feasible for large LOFAR-LBA datasets ( $\sim 18$  TB per dataset) because of a large number of baselines and time-steps. Therefore, we use an alternative approach where we estimate the noise spectrum from the gridded visibilities (see e.g. Jacobs et al. 2016; Beardley et al. 2016; Ewall-Wice et al. 2016). We split the DI-calibrated visibilities of the 3C220 field into even and odd samplings with 12 seconds cadence such that these samplings are interleaved in time. Note that for the baseline range  $20\lambda \leq |\mathbf{u}| \leq 200\lambda$  which we probe in our analysis, the sky and the PSF do not vary over a 12 second interval. Also, any sky leakage over 12 seconds will appear as a wedge in the cylindrically averaged power spectrum, which we do not observe in the analyses (shown in later sections). Moreover, we expect the system to be coherent over 12 seconds and only ionospheric effects are expected to change. We image these even and odd samplings using `WSClean` with the ‘natural’ weighting scheme. We Fourier Transform (FT) the even and odd image cubes and properly scale visibilities in each  $uv$ -cell with corresponding sampling density to remove the effect of gridding weights during imaging. We calculate the

azimuthally averaged (spatial) power spectrum of the difference as  $P_{\Delta,I}(|\mathbf{u}|, \nu) \equiv \langle \Delta_I \tilde{I} \rangle^2 = \langle \tilde{I}_{\text{even}} - \tilde{I}_{\text{odd}} \rangle^2 / 2$ , where  $\tilde{I}_{\text{even}}$  and  $\tilde{I}_{\text{odd}}$  are the Fourier transforms of the even and odd image cubes respectively,  $\mathbf{u} = (u, v)$  is the baseline vector (in units of wavelength) in the  $uv$ -plane and  $|\mathbf{u}| = \sqrt{u^2 + v^2}$  and  $\nu$  is the frequency.

##### 4.1 Physical Excess Noise

Figure 2 shows  $P_{\Delta,I}$  and  $P_{\Delta,V}$  for the  $20 - 200\lambda$  baseline range for the 3C220 field. We observe that both  $P_{\Delta,I}$  and  $P_{\Delta,V}$  spectra are relatively flat. The bright tilted streaks are because of varying  $uv$ -density as a function of baseline length in LOFAR-LBA. We compare  $P_{\Delta,I}$  and  $P_{\Delta,V}$  by calculating their ratio. Figure 3 shows the ratio  $P_{\Delta,I}/P_{\Delta,V}$  of the spectra shown in figure 2. We observe that the ratio is remarkably flat, except for a few outliers at shorter baselines ( $\leq 40\lambda$ ). These outliers might arise due to imperfect calibration and slight differences in flagging of RFI affected baselines post calibration and/or gaps in the  $uv$ -coverage. Ideally, if the noise properties of Stokes  $I$  and  $V$  are statistically identical and if the sky and the PSF do not change over a 12 seconds interval,  $P_{\Delta,I}$  and  $P_{\Delta,V}$  are expected to be identical assuming that the sky has a negligible circular polarized emission component and Stokes  $V$  is virtually empty. However, we observe excess power in Stokes  $I$  compared to Stokes  $V$ , which is largely constant over the  $20 - 200\lambda$  baseline range and over the 30 MHz bandwidth. Although, the power in both Stokes  $I$  and  $V$  varies slightly with increasing baseline length, the ratio remains constant, suggesting that this slight variation is a result of varying  $uv$ -density. Hence this physical excess noise, for all practical purposes, behaves as additional white noise in Stokes  $I$  that is seemingly uncorrelated in frequency, and remains more or less the same for different baseline lengths. The left panel of figure 4 shows the normalized histogram of





**Figure 5.** The top panel shows the differential Stokes  $I$  and  $V$  power spectra calculated using residual images of the 3C220 field. The solid red curve corresponds to  $P_{\Delta\nu I}$  and the dashed red curve corresponds to  $P_{\Delta\nu V}$ . The solid and dashed black curves correspond to  $P_{\Delta I}$  and  $P_{\Delta V}$  respectively, at  $\nu = 59.95$  MHz. The bottom panel shows the ratio  $P_{\Delta\nu I}/P_{\Delta\nu V}$  (red curve) and the ratio  $P_{\Delta\nu I}/P_{\Delta I}$  (blue curve). The dotted vertical line shows the location of the  $200\lambda$  baseline cut.

the distribution of  $P_{\Delta I}/P_{\Delta V}$  values for the 3C220 field. The distribution has a median value of 1.46 and a mean value of 1.54, with most values lying within the range 1–2. The noise spectra and their ratio for the NCP field also exhibit a similar behavior as the 3C220 field that the ratio  $P_{\Delta I}/P_{\Delta V}$  is flat in frequency-baseline space. However, the distribution of the ratio values (see right panel of figure 4) has slightly lower median and mean values of 1.32 and 1.38 respectively. The cause of this excess power in  $P_{\Delta I}$  is still unknown, but it is higher for the 3C220 field which has a bright source at the center, compared to the NCP field which is devoid of relatively bright sources. We are currently investigating the cause of this excess, but given that the excess is different for the two fields, ionospheric or interplanetary scintillation noise might be a probable reason for this excess, although the rapid decorrelation with frequency remains unexplained.

#### 4.2 Comparison with sub-band level noise

We use the azimuthally averaged power spectrum of the difference of Stokes  $I$  and  $V$  images between two contiguous sub-bands (differential power spectrum) to study the behavior of noise at the inter-sub-band level (see e.g. Patil et al. 2016; Gehlot et al. 2018). This method is based on the assumption that Stokes  $I$  images are composed of total sky signal convolved with the PSF plus additive noise. Assuming that the sky signal, which is smooth in frequency

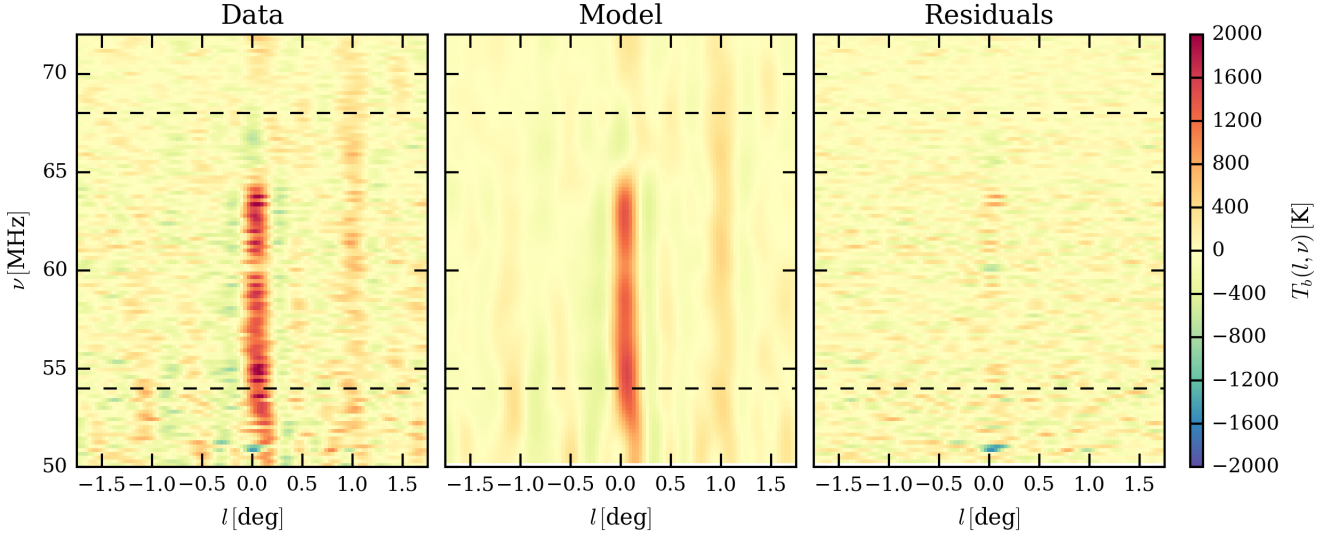
does not change<sup>16</sup> between two consecutive sub-bands 195 kHz apart, and any contribution due to the sky signal should largely drop out in the differential Stokes  $I$  images. Similarly, differential Stokes  $V$  images should contain only noise. However, effects which are non-smooth at the sub-band level are expected to leave their signature in the differential Stokes images.

We use Stokes  $I$  and  $V$  residual images of the 3C220 field ( $\nu_1 = 59.76$  MHz and  $\nu_2 = 59.95$  MHz, located at the most sensitive part of the band) after DD-calibration to estimate the differential power spectra  $P_{\Delta\nu I}$  and  $P_{\Delta\nu V}$ , and determine their ratio  $P_{\Delta\nu I}/P_{\Delta\nu V}$ . The top panel of figure 5 shows  $P_{\Delta\nu I}$  (red solid curve) and  $P_{\Delta\nu V}$  (red dashed curve). We also show a slice of  $P_{\Delta I}$  and  $P_{\Delta V}$  at 59.95 MHz in the same plot for comparison. The bottom panel of figure 5 shows the ratio  $P_{\Delta\nu I}/P_{\Delta\nu V}$  (red curve) and the ratio  $P_{\Delta\nu I}/P_{\Delta I}$  (blue curve). We observe that  $P_{\Delta\nu I}/P_{\Delta\nu V} \sim 2-3$ , which is considerably smaller than the ratio ( $P_{\Delta\nu I}/P_{\Delta\nu V} \gtrsim 10$ ) that we observed in G18. This lower noise is in part achieved because SAGEcal-CO enforces frequency smoothness of the gain solutions in the calibration process, and also because the ionospheric activity is more benign compared to the observation in G18 where frequency smoothness was not enforced in the calibration and the ionosphere behaved erratically. Furthermore, from comparison of  $P_{\Delta\nu I}$  with  $P_{\Delta I}$ , we observe that there is a sudden jump in the ratio at  $|\mathbf{u}| \sim 200\lambda$ . The ratio is  $\gtrsim 2$  for  $|\mathbf{u}| < 200$  and it continues to increase as the baseline length decreases, compared to the values ( $\sim 1-2$ ) for  $|\mathbf{u}| > 200$ . We attribute this effect to the  $200\lambda$  baseline cut used in the DD-calibration. The sky-model incompleteness or ionospheric effects can introduce random errors during the calibration step. These random errors on gain solutions when applied to the baselines excluded during the calibration step, increase the variance on the excluded baselines (Patil et al. 2016; Barry et al. 2016; Ewall-Wice et al. 2016, 2017; Gehlot et al. 2018; Mouri Sardarabadi & Koopmans 2018).

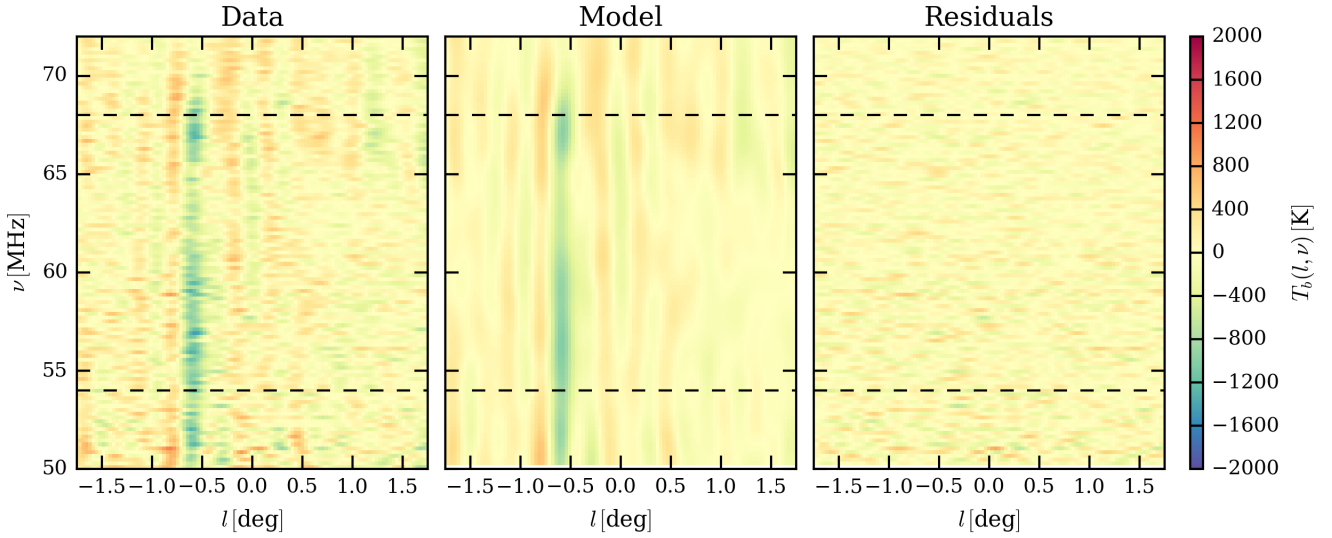
## 5 GAUSSIAN PROCESS REGRESSION

After subtracting the calibration sky-model using DD-calibration, any remaining foreground emission within the primary beam consists of unresolved sources, sources below the confusion noise, sources not included in the model, and diffuse emission. These foregrounds should vary slowly with frequency, making them separable from the 21-cm signal which has rapid spectral fluctuations. We use a Gaussian Process Regression (GPR) technique (see Mertens et al. 2018 for more details) to remove this remaining foreground emission and other spectral structures imparted on the data due to instrumental mode-mixing, such as instrumental chromaticity and imperfect calibration residuals. In this section, we briefly describe GPR and its application to LOFAR-LBA data.

<sup>16</sup> For a spectral index of  $-2.55$ , sky brightness changes at  $\sim 0.8\%$  level for 195 kHz frequency separation at 60 MHz, which has negligible contribution to the difference.



**Figure 6.** The 3C220 field Stokes  $I$  image cube slices (in brightness temperature units) across the center of a spatial axis after different processing steps. The left panel shows a slice of the image cube after DD-calibration (data). The middle panel shows the GPR model of the smooth foregrounds (intrinsic + mode-mixing) corresponding to the data. The right panel shows the residuals after subtracting the GPR model from the data. The dashed black lines represent the frequency range (54–68 MHz) used for power spectrum estimation. The residuals after GPR are noise-like except for a few outliers.



**Figure 7.** As figure 6 but for the NCP field. Similar to the 3C220 field, the residuals in the NCP field after GPR are noise-like.

## 5.1 Methodology

The visibilities observed by an interferometer ( $\mathcal{V}_{\text{obs}}(\mathbf{u}, \nu)$ ) can be written as a sum of different components viz. the signal of interest ( $\mathcal{V}_{21}(\mathbf{u}, \nu)$ ), the foreground contribution ( $\mathcal{V}_{\text{sky}}(\mathbf{u}, \nu)$ ), instrumental mode-mixing ( $\mathcal{V}_{\text{mix}}(\mathbf{u}, \nu)$ ) and noise ( $\mathcal{V}_{\text{n}}(\mathbf{u}, \nu)$ ), i.e.

$$\mathcal{V}_{\text{obs}}(\mathbf{u}, \nu) = \mathcal{V}_{21}(\mathbf{u}, \nu) + \mathcal{V}_{\text{sky}}(\mathbf{u}, \nu) + \mathcal{V}_{\text{mix}}(\mathbf{u}, \nu) + \mathcal{V}_{\text{n}}(\mathbf{u}, \nu). \quad (1)$$

Each of these components has a distinct spectral behavior which is exploited by GPR to separate them from each other and eventually remove the foreground compo-

nents from the observed visibilities, leaving residuals with the signal of interest buried below the noise (Mertens et al. 2018). GPR models these different components with Gaussian Processes (GP). A GP ( $f \sim \mathcal{GP}(m, \kappa)$ ) is the joint distribution of a collection of normally distributed random variables and is defined by its mean  $m$  and covariance function  $\kappa$ . Values of  $\kappa$  specify the covariance between pairs of points at different frequencies and determine the structure of the function (e.g. its smoothness in frequency) which can be modeled with a GP. GPs are often described by parameterized priors in GPR, and the GP prior is selected such that it maximizes

the Bayesian evidence, estimated by conditioning these priors to the observations (see [Rasmussen & Williams 2005](#) for an extensive review). The parameters of the covariance functions (also known as ‘hyper-parameters’) can be estimated using standard optimization or MCMC algorithms. The observed data  $\mathbf{d}$ , being a stacked set of gridded visibilities as a function of frequency, can be modeled as

$$\mathbf{d} = \mathbf{f}_{\text{fg}}(\nu) + \mathbf{f}_{21}(\nu) + \mathbf{n}, \quad (2)$$

where  $\mathbf{f}_{\text{fg}}(\nu)$  corresponds to the foreground component,  $\mathbf{f}_{21}(\nu)$  corresponds to the signal of interest and  $\mathbf{n}$  is the noise. The 21-cm signal is expected to decorrelate over frequency scales  $> 1$  MHz, whereas foregrounds are expected to be smooth on 1 MHz scales and decorrelate over a much larger frequency range. The covariance function  $K \equiv \kappa$  for this model can be written as a sum of foreground covariance function  $K_{\text{fg}}$  and a 21-cm signal covariance function  $K_{21}$ , i.e.

$$K = K_{\text{fg}} + K_{21}. \quad (3)$$

$K_{\text{fg}}$  can further be decomposed into  $K_{\text{int}}$ , which corresponds to intrinsic foregrounds (large-scale correlation of  $\sim 10 - 100$  MHz) and  $K_{\text{mix}}$ , which corresponds to instrumental mode mixing such as sidelobe noise (decorrelates within  $\sim 1 - 5$  MHz). The joint probability distribution for the observed data  $\mathbf{d}$  and function values  $\mathbf{f}_{\text{fg}}$  of the foreground model at the same frequency  $\nu$  is then given by

$$\begin{bmatrix} \mathbf{d} \\ \mathbf{f}_{\text{fg}} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K_{\text{fg}} + K_{21} + \sigma_n^2 I & K_{\text{fg}} \\ K_{\text{fg}} & K_{\text{fg}} \end{bmatrix} \right), \quad (4)$$

where  $\sigma_n^2$  is the noise variance,  $I$  is the identity matrix and  $K \equiv K(\nu, \nu)$ . After GPR, the foreground model can be retrieved as

$$E(\mathbf{f}_{\text{fg}}) = K_{\text{fg}} \left[ K + \sigma_n^2 I \right]^{-1} \mathbf{d}, \quad (5)$$

$$\text{cov}(\mathbf{f}_{\text{fg}}) = K_{\text{fg}} - K_{\text{fg}} \left[ K + \sigma_n^2 I \right]^{-1} K_{\text{fg}}, \quad (6)$$

where  $E(\mathbf{f}_{\text{fg}})$  and  $\text{cov}(\mathbf{f}_{\text{fg}})$  are the expectation values and covariance of the foregrounds respectively. The residuals  $\mathbf{d}_{\text{res}}$  after foreground model subtraction are

$$\mathbf{d}_{\text{res}} = \mathbf{d} - E(\mathbf{f}_{\text{fg}}). \quad (7)$$

Readers may refer to [Mertens et al. \(2018\)](#) for a detailed description of the GPR technique and its application as a novel method for foreground removal and 21-cm signal estimation.

## 5.2 Application of GPR to the LOFAR-LBA data

We use GPR to remove remaining foreground emission from the residual visibilities after DD calibration. We test various covariance functions as kernels to model different components of the residual visibilities in GPR. To model the intrinsic foreground emission (unmodeled sources and diffuse emission) we choose a RBF (Radial Basis Function) covariance function as kernel. The RBF kernel is essentially a square exponential or Gaussian function defined as:

$$\kappa_{\text{RBF}}(\nu_p, \nu_q) = \exp \left( \frac{-|\nu_q - \nu_p|^2}{2l^2} \right) \quad (8)$$

where  $l$  is the characteristic coherence scale in frequency. We use 5–100 MHz as the prior for the range of coherence scales

of the intrinsic foregrounds. To model the mode-mixing component of the foregrounds, we choose a Rational Quadratic (RQ) covariance function defined as:

$$\kappa_{\text{RQ}}(\nu_p, \nu_q) = \left( 1 + \frac{|\nu_q - \nu_p|^2}{2\alpha l} \right)^{-\alpha}, \quad (9)$$

where  $l$  is the coherence scale and  $\alpha$  is the so-called power-parameter. RQ functions can be understood as infinite sums of Gaussian covariance functions with characteristic coherence scales ([Rasmussen & Williams 2005](#)). We use 1–8 MHz as prior values for the coherence scales and  $\alpha = 0.1$  for the mode-mixing component. To account for the 21-cm signal, we use an Exponential covariance function, which is a special case of a Matern class covariance function ([Stein 1999](#)) defined as:

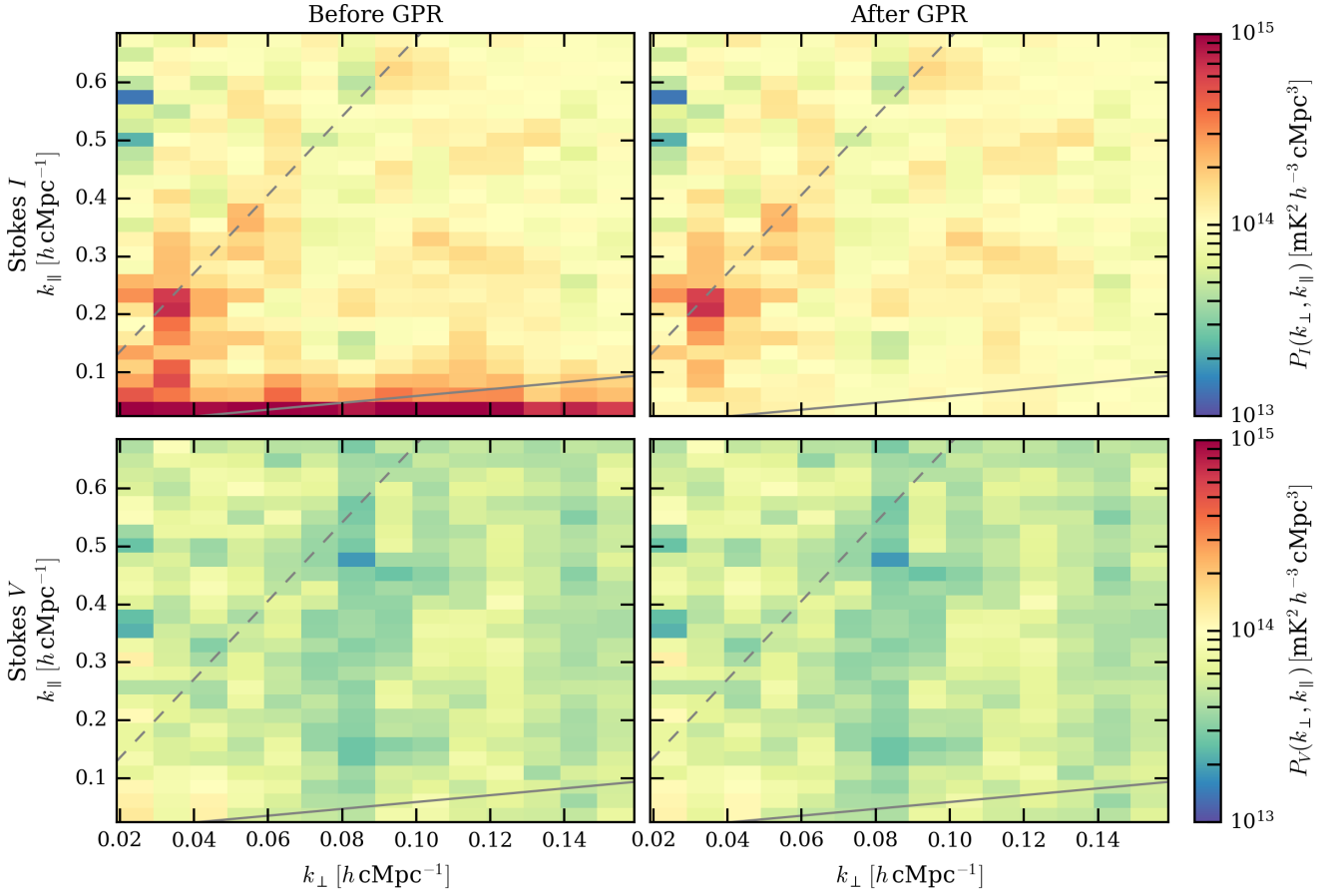
$$\kappa_{\text{matern}}(\nu_p, \nu_q) = \frac{2^{1-\eta}}{\Gamma(\eta)} \left[ \frac{\sqrt{2\eta}|\nu|}{l} \right]^\eta K_\eta \left( \frac{\sqrt{2\eta}|\nu|}{l} \right), \quad (10)$$

where  $|\nu| = |\nu_q - \nu_p|$  and  $K_\eta$  is the modified Bessel function of the second kind. The ‘hyper-parameter’  $l$  represents the characteristic coherence scale. Special classes of Matern covariance functions can be obtained by choosing various values for  $\eta$ , e.g. choosing  $\eta = 1/2$  gives an exponential kernel. We use a frequency coherence scale of 0.01–1.5 MHz for the 21-cm signal with an initial value of 0.5 MHz. Using 21cm-FAST simulations ([Mesinger & Furlanetto 2007](#); [Mesinger et al. 2011](#)), [Mertens et al. \(2018\)](#) have shown that these coherence scales covers a wide range of possible 21-cm signal models.

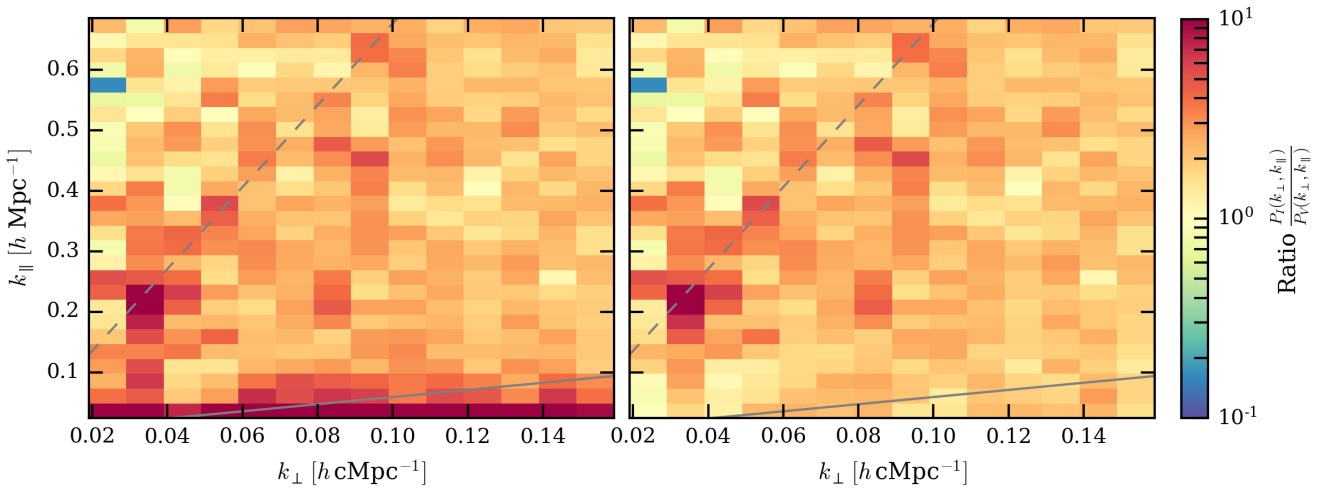
We use the residual image-cubes obtained after DD-calibration for foreground removal. We perform GPR foreground removal on the inner  $3.5^\circ \times 3.5^\circ$  region of the image cubes (which is slightly smaller than the primary beam FWHM  $\sim 4^\circ$ ) to limit sky curvature and primary beam effects. We selected the 50–72 MHz frequency range for GPR foreground removal, which is 8 MHz wider than the power spectrum estimation window, for better foreground fitting and removal. Figure 6 shows slices through the Stokes  $I$  image cubes for the 3C220 field across the center of one of the two spatial axes before GPR (left panel), the GPR foreground fit (middle panel) and the residuals after subtracting the foreground model obtained with GPR from the data (right panel). Similarly, figure 7 shows the slices of Stokes  $I$  image cubes for the NCP field. We observe that the residuals after foreground removal with GPR basically look noise-like for both the 3C220 and NCP fields. In the following section, we use these residuals after GPR to create cylindrically and spherically averaged power spectra for the 3C220 and the NCP fields.

## 6 POWER SPECTRA RESULTS

In this section we present the cylindrically and spherically averaged power spectra for the 3C220 and NCP fields. Cylindrically averaged power spectra (or 2D cosmological power spectra) are widely used in 21-cm experiments to assess various 21-cm signal contaminants such as foregrounds, side-lobe noise and systematic biases. Cylindrically averaged power spectra ( $P(k_\perp, k_\parallel)$ ) are defined as ([Parsons et al. 2012](#); [Thya-](#)



**Figure 8.** The cylindrically averaged Stokes  $I$  and  $V$  power spectra for the 3C220 field. The top row shows  $P_I(k_\perp, k_\parallel)$  before (left panel) and after (right panel) foreground removal with GPR. The bottom row shows  $P_V(k_\perp, k_\parallel)$  before (left panel) and after (right panel) foreground removal with GPR. The solid gray lines correspond to a  $5^\circ$  field of view which is slightly larger than the primary beam FWHM at 60 MHz. The dashed gray lines correspond to the instrumental horizon.



**Figure 9.** The ratio of the cylindrically averaged Stokes  $I$  and  $V$  power spectra for the 3C220 field. The left panel shows  $P_I/P_V$  before foreground removal with GPR. The right panel shows  $P_I/P_V$  after foreground removal with GPR.



garajan et al. 2015):

$$P(k_{\perp}, k_{\parallel}) = \frac{X^2 Y}{\Omega_A B} \langle |\mathcal{V}(\mathbf{u}, \tau)|^2 \rangle, \quad (11)$$

where  $\mathcal{V}(\mathbf{u}, \tau)$  is the FT of the visibilities in the frequency direction,  $\Omega_A$  is the primary beam solid angle and  $B$  is the bandwidth of the visibility cube.  $X$  and  $Y$  are the conversion factors from angle and frequency to transverse co-moving distance ( $D(z)$ ) and the co-moving depth along the line of sight ( $\Delta D$ ), respectively. The wave numbers  $k_{\perp}$  and  $k_{\parallel}$  are related to baseline vector ( $\mathbf{u} = (u, v)$  in units of wavelength) and delay ( $\tau$ ) as:

$$k_{\perp} = \frac{2\pi(|\mathbf{u}|)}{D(z)}, \quad k_{\parallel} = \frac{2\pi\nu_{21}H_0E(z)}{c(1+z)^2}\tau, \quad (12)$$

where  $\nu_{21}$  is the rest frame frequency of the 21-cm spin flip transition of HI,  $z$  is the redshift corresponding to the observation frequency,  $c$  is the speed of light,  $H_0$  is the Hubble constant and  $E(z) \equiv [\Omega_M(1+z)^3 + \Omega_k(1+z)^2 + \Omega_{\Lambda}]^{1/2}$  is a function of the standard cosmological parameters. The spherically averaged dimensionless power spectrum can be estimated from  $P(k_{\perp}, k_{\parallel})$  as:

$$\Delta^2(k) = \frac{k^3}{2\pi^2} P(k), \quad (13)$$

where  $k = \sqrt{k_{\perp}^2 + k_{\parallel}^2}$ . We determine  $P(k_{\perp}, k_{\parallel})$  for both the 3C220 and NCP fields using the gridded visibility cubes of the  $3.5^\circ \times 3.5^\circ$  region of the image cubes with 14 MHz bandwidth (54–68 MHz), corresponding to the redshift range  $z = 19.8 - 25.2$ . We weigh the  $uv$ -cells using an empirical weighting scheme<sup>17</sup> where  $uv$ -cells in Stokes  $I$  and  $V$  with higher noise are down-weighted. We use a Least Square Spectral Analysis (LSSA) method (full least squares FT-matrix inversion, see e.g. [Barning 1963](#); [Lomb 1976](#); [Stoica et al. 2009](#); [Trott et al. 2016](#)) to Fourier transform the visibilities along the frequency direction. We apply a ‘Hann’<sup>18</sup> window function to the data prior to the frequency transform to control the side-lobes along the delay axis, although this window function somewhat increases the noise. The resulting  $\mathcal{V}(\mathbf{u}, \tau)$  cubes are scaled accordingly with the conversion factors  $X$  and  $Y$  calculated using  $\Lambda$ CDM cosmological parameters that are consistent with the Planck 2016 results ([Planck Collaboration et al. 2016](#)), and then cylindrically and spherically averaged to obtain  $P(k_{\perp}, k_{\parallel})$  and  $\Delta^2(k)$ , respectively.

### 6.1 The 3C220 field: cylindrical power spectra

In this section, we examine the cylindrical power spectra for the 3C220 field. The top row of figure 8 shows  $P(k_{\perp}, k_{\parallel})$

<sup>17</sup> The weights are scaled by the inverse of the per-visibility variance in each  $uv$ -cells estimated by computing robust variance statistics of Stokes  $V$  along the frequency direction. While the per-visibility variance should theoretically be identical in each  $uv$ -cell, it is not in the presence of systematics that can affect baselines differently. This empirical weighting scheme allows to reduce the impact of those systematics. Note that only Stokes  $V$  is used in determining those weights.

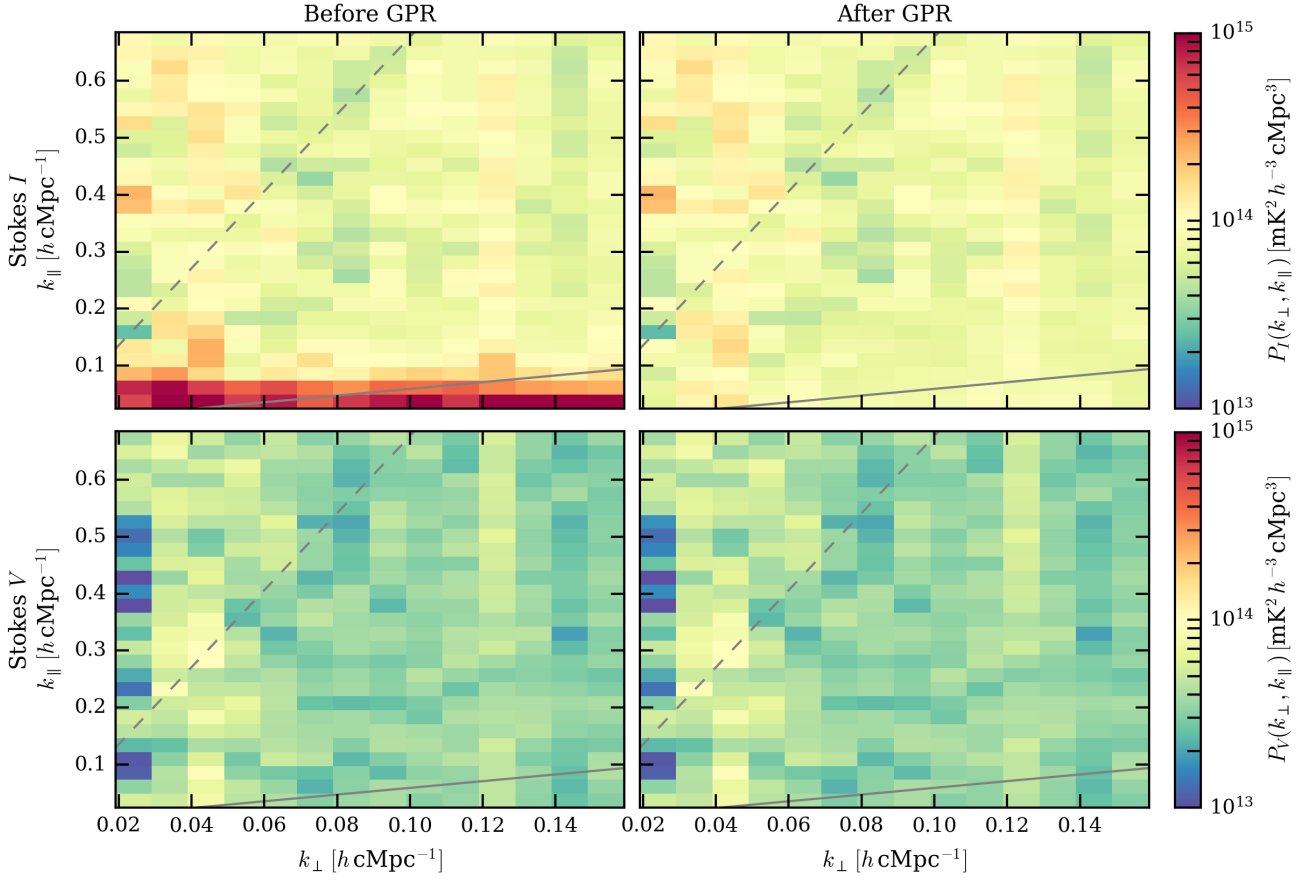
<sup>18</sup> The ‘Hann’ window is defined as  $W(n) = 0.5 - 0.5 \cos \left[ \frac{2\pi n}{(M-1)} \right]$ , where  $0 \leq n \leq M-1$  (see e.g. [Blackman & Tukey 1958](#); [Harris 1978](#)).

for Stokes  $I$  before and after foreground removal. We observe that the lowest  $k_{\parallel}$  bins in Stokes  $I$  are dominated by smooth foreground emission (see top left panel) even after subtraction of the sky-model during DD-calibration. This foreground emission is effectively removed by the GPR foreground removal method (see top right panel). We also observe an excess power around the horizon delay line in Stokes  $I$  prior to GPR, which is not removed during GPR, suggesting that this excess power has much lower spectral smoothness or decorrelates quickly over time between gridded visibilities and cannot be modeled with a GP with current settings. This structure is reminiscent of the ‘pitchfork’ observed in G18 and is possibly caused by the residuals after Cas A and Cyg A subtraction. An inaccurate source model, ionospheric effects, time variation of the primary beam, or a combination of these effects might explain these residuals. We expect the modeling errors to be negligible as their corresponding models are derived from high spatial resolution images and also because Cas A and Cyg A appear as compact sources on shorter baselines. Ionospheric effects, however, become stronger at lower elevations due to projection effects and subtraction of Cas A and Cyg A at 30 seconds and 61 kHz calibration resolution might not be sufficient to correct for ionospheric effects, especially on the shorter baselines. Also, the primary beam changes with time as the 3C220 field is tracked. Therefore, a combination of ionospheric effects, beam errors and time variation of the primary beam is likely capable of producing such an effect.

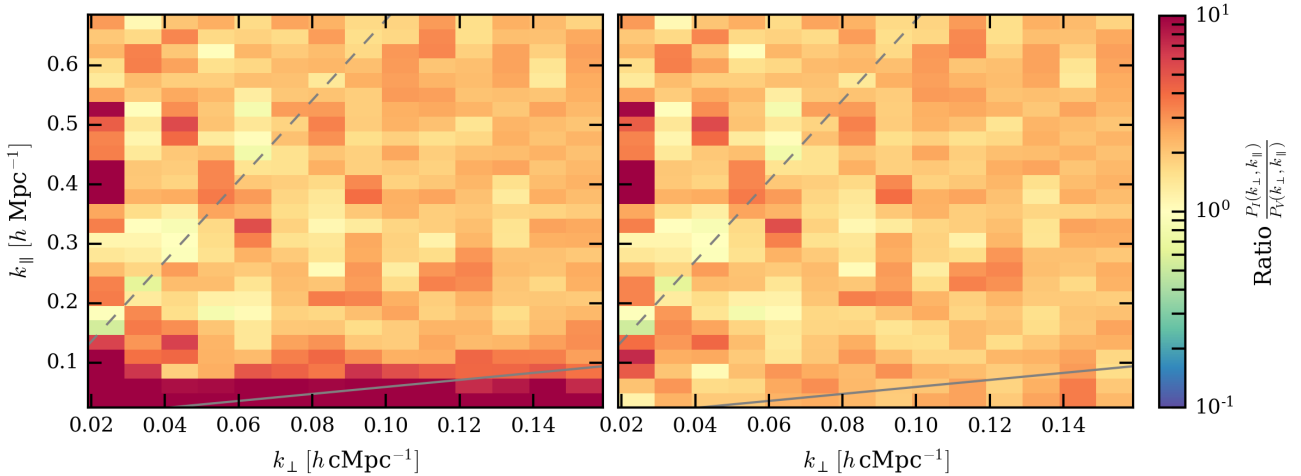
The bottom row of figure 8 shows  $P(k_{\perp}, k_{\parallel})$  for Stokes  $V$  before and after GPR foreground removal. We observe that the Stokes  $V$  power spectrum is noise-like before and after foreground removal, which suggests that any foreground emission/leakage in Stokes  $V$  is lower than the noise in the current data. We also do not observe any visible signature of Cas A and Cyg A residuals. The vertical bands in Stokes  $I$  and  $V$  near  $k_{\perp} \approx 0.08$  and  $0.14$  are due to varying  $uv$ -density and drop out in the ratio  $\frac{P_I}{P_V}$ . The ratio after foreground removal, as shown in the right panel of figure 9, is relatively flat compared to the one before foreground removal, except for the above-mentioned region near the horizon. The ratio has a median value of 2.07 which is higher than the median value ( $\sim 1.46$ ) observed in the ratio  $P_{\Delta, I}/P_{\Delta, V}$  for the 3C220 field (see section 4.1). However, it is consistent with the excess at the sub-band level (see section 4.2) caused by the use of a baseline cut in the DD-calibration.

### 6.2 The NCP field: cylindrical power spectra

In this section, we assess the cylindrical power spectra for the NCP field. Figure 10 shows  $P(k_{\perp}, k_{\parallel})$  for Stokes  $I$  and  $V$ . The top left panel shows the power spectrum after DD-calibration, where low  $k_{\parallel}$  modes are dominated by the power due to foreground emission. Similar to the 3C220 field, this power is effectively removed with GPR (see top right panel). We do not observe a ‘pitchfork’ in Stokes  $I$  or  $V$  (presumably) due to Cas A and Cyg A residuals opposed to the 3C220 field. This might be primarily because the NCP is stationary on the sky and therefore the beam does not change (only rotates) as the observation progresses. It is also likely that the Cas A and Cyg A are closer to the null for the NCP, causing the power on/around the structure to be below the noise. Similar to the 3C220 field, Stokes  $V$  power spectra for



**Figure 10.** The cylindrically averaged Stokes  $I$  and  $V$  power spectra for the NCP field. The top row shows  $P_I(k_\perp, k_\parallel)$  before (left panel) and after (right panel) foreground removal with GPR. The bottom row shows  $P_V(k_\perp, k_\parallel)$  before (left panel) and after (right panel) foreground removal with GPR.



**Figure 11.** The ratio of the cylindrically averaged Stokes  $I$  and  $V$  power spectra for the NCP field. The left panel shows  $P_I/P_V$  before foreground removal with GPR. The right panel shows  $P_I/P_V$  after foreground removal with GPR.

the NCP field appear noise like before and after foreground removal (see figure 10).

The behavior of the ratio  $\frac{P_I}{P_V}$  (see figure 11) is also similar to that of the 3C220 field. The ratio becomes relatively flat after foreground removal except for a few outliers at the small  $k_{\perp}$  values. The ratio has a median value of 2.10, which is almost equivalent to the median we observed for the 3C220 field, but higher than the median of the ratio  $P_{\Delta_I I}/P_{\Delta_I V}$  for the NCP field (Section 4.1). This excess can also be attributed to the baseline cut in DD-calibration as discussed in Section 4.2, which we know causes excess power.

### 6.3 Comparison with noise power spectra

We determine the cylindrically averaged noise power spectra  $P_I^N$  and  $P_V^N$  corresponding to the Stokes  $I$  and  $V$  difference cubes  $\Delta_I \tilde{I}$  and  $\Delta_I \tilde{V}$  respectively for the 3C220 field (e.g. see section 4), by passing these cubes through the power spectrum estimation pipeline. Note that we do not perform foreground removal on these data-cubes because we expect the sky component to drop out on time scales of 12 seconds. Figure 12 shows  $P_I^N$  (left panel) and  $P_V^N$  (right panel). We observe that power in both Stokes  $I$  and  $V$  is lower for small  $k_{\perp}$  values and higher for larger  $k_{\perp}$  because the  $uv$ -density of LOFAR-LBA decreases with increasing baseline length and drops out in the ratio  $\frac{P_I^N}{P_V^N}$  shown in figure 13. From comparison of  $P_V(k_{\perp}, k_{\parallel})$  for the 3C220 and NCP fields with  $P_V^N$ , we notice that  $P_V(k_{\perp}, k_{\parallel})$  deviates from  $P_V^N$  for lower  $k_{\perp}$  ( $< 0.1$ ) values. This deviation in  $P_V(k_{\perp}, k_{\parallel})$  compared to  $P_V^N$  can be attributed to the baseline cut used in the DD-calibration, which increases the noise on the baselines excluded from the calibration.

Moreover,  $\frac{P_I^N}{P_V^N}$  has a median value of 1.51, which is consistent with the median value of 1.46 for the ratio  $P_{\Delta_I I}/P_{\Delta_I V}$ . The NCP field, however, has a slightly lower median value  $\sim 1.3$ . We observe that this excess power in Stokes  $I$  for both the 3C220 and the NCP fields at 12 seconds level does not depend on the calibration, and is present at the same level throughout the analysis even after DD-calibration, foreground removal and also in the auto-correlations (results not shown here). This excess is different from the calibration cut induced noise and might have a physical origin. To account for this physical excess noise in the estimation of the spherically averaged power spectrum, we multiply the residual Stokes  $V$  gridded visibilities after DD-calibration (since Stokes  $V$  is an independent estimator of the thermal noise of the instrument) with the square-root of the median of the ratio  $P_{\Delta_I I}/P_{\Delta_I V}$  (calculated in section 4.1) to obtain an estimate of the noise in Stokes  $I$ . We use the median instead of the mean because the median is a more robust representative of the central tendency of the distribution, whereas the mean is sensitive to outliers and becomes biased in the presence of strong outliers. This excess noise bias corrected Stokes  $V$  is used as an estimator for the noise in the data in the foreground removal and spherically averaged power spectrum estimation steps.

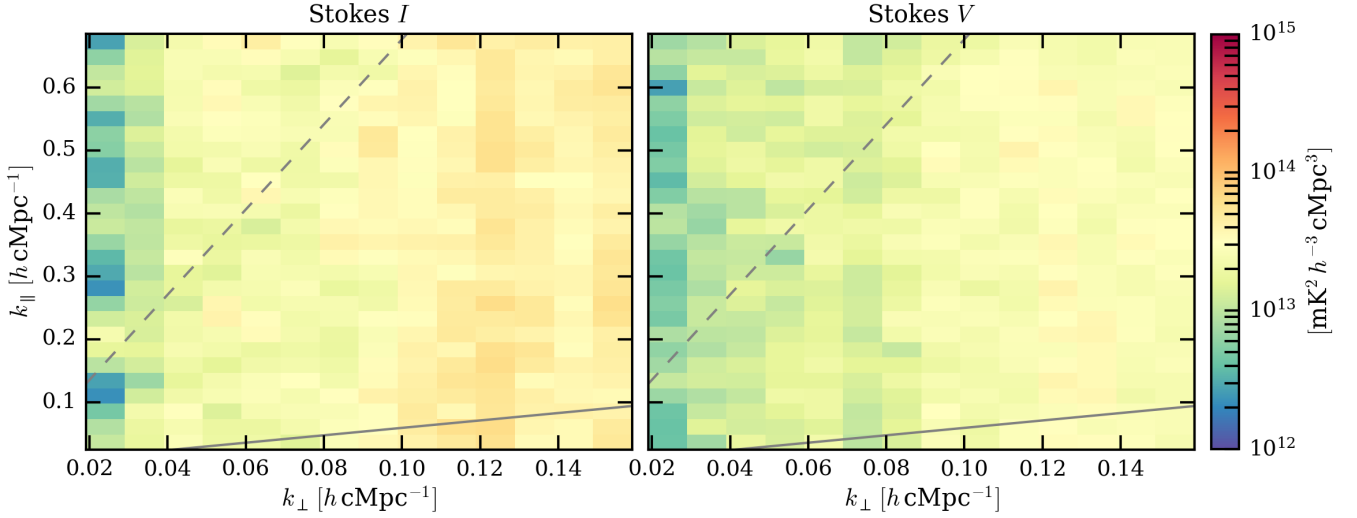
### 6.4 Spherically averaged power spectra

We finally determine the Stokes  $I$  and  $V$  spherically averaged power spectra ( $\Delta^2(k)$ ) for both the 3C220 and NCP fields before and after foreground removal for the redshift range  $z = 19.8 - 25.2$ . Figure 14 shows Stokes  $I$  power spectra  $\Delta_I^2$  and Stokes  $V$  power spectra  $\Delta_V^2$  for both the 3C220 (left panel) and NCP fields (right panel) before and after foreground removal. We use the physical excess noise bias corrected (using the median values from Section 4.1) Stokes  $V$  visibilities as an estimator of the noise component in the Stokes  $I$  power spectrum ( $\Delta_{I,n}^2$ ), in order to account for the physical excess noise in Stokes  $I$  compared to Stokes  $V$ . The dashed gray curves in figure 14 represent the excess bias corrected Stokes  $V$  power spectra  $\Delta_{I,n}^2$ . For both fields, we observe that the power on smaller  $k$  modes is dominated by large-scale foreground emission which comprises diffuse emission, unmodeled sources and sources below the confusion noise prior to foreground removal. Residual Stokes  $I$  power on the smallest  $k$  modes after foreground removal is an order of magnitude lower than the former. However, GPR does not remove any power from Stokes  $V$ , which means that any structure in Stokes  $V$  is spatially and spectrally incoherent and behaves as uncorrelated noise. For the 3C220 field, Stokes  $I$  residuals approach  $\Delta_{I,n}^2$  at smaller  $k$  modes, however these are still higher than  $\Delta_{I,n}^2$  by  $\sim 30\%$  on large  $k$  modes. On the other hand, Stokes  $I$  residuals for the NCP field are higher than  $\Delta_{I,n}^2$  by  $\sim 50\%$  on most  $k$  modes except for the lowest one. This remaining excess power, after correcting for the physical excess noise bias, is likely due to the baseline cut used during the DD-calibration.

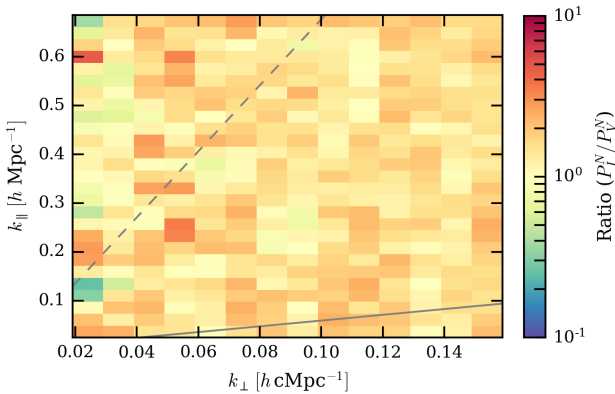
Assuming that the physical noise properties of Stokes  $I$  and  $V$  are statistically identical, we use the post GPR excess noise bias corrected Stokes  $V$  power spectrum ( $\Delta_{I,n}^2$ ) to remove the noise component in the residual Stokes  $I$  power spectrum. The noise bias corrected power spectrum  $\Delta_I^2 - \Delta_{I,n}^2$  for the 3C220 (blue circles) and the NCP field (red crosses) are shown in figure 15. The dashed curves show thermal noise power spectrum estimate  $\Delta_N^2$  estimated from  $\Delta_I \tilde{V}$  for the 3C220 (‘skyblue’ colored) and the NCP field (‘coral’ colored). We observe that  $\Delta_I^2 - \Delta_{I,n}^2$  for both fields are consistent with each other within the  $2\sigma$  uncertainty for modes  $k \lesssim 0.2 h \text{ cMpc}^{-1}$  and deviate slightly on  $k \gtrsim 0.2 h \text{ cMpc}^{-1}$ . This is possibly due to different morphologies of the two fields on small angular scales. The  $\Delta_I^2 - \Delta_{I,n}^2$  for both fields, within  $2\sigma$  uncertainty, agree with their respective noise estimate  $\Delta_N^2$  (determined from  $\Delta_I \tilde{V}$ ) which is a more accurate estimator of the thermal noise of the system. The  $\Delta_N^2$  for both fields show power-law like behavior and agree with each other on all  $k$  modes. We find a  $2\sigma$  upper limit of  $\Delta_{21}^2 < (14561 \text{ mK})^2$  at  $k \sim 0.038 h \text{ cMpc}^{-1}$  for the 3C220 field and  $\Delta_{21}^2 < (14886 \text{ mK})^2$  at  $k \sim 0.038 h \text{ cMpc}^{-1}$  for the NCP field in the redshift range  $z = 19.8 - 25.2$ . Both upper limits are consistent with each other within 2%.

## 7 SUMMARY AND OUTLOOK

In this work, we have explored the possibility of a statistical measurement of the redshifted 21-cm signal of neutral hydrogen from the Cosmic Dawn using the LOFAR-Low Band Antenna system. We have presented the first upper limits on



**Figure 12.** The cylindrically averaged noise power spectra  $P_I^N$  and  $P_V^N$  (corresponding to Stokes  $I$  and  $V$ ) for the 3C220 field determined from the difference cubes  $\Delta_t \bar{I}$  and  $\Delta_t \bar{V}$  respectively. The left panel shows  $P_I^N$  and the right panel shows  $P_V^N$ .



**Figure 13.** The ratio  $\frac{P_I^N}{P_V^N}$  for the 3C220 field. We observe that the ratio has a median value of 1.51.

the power spectrum of the 21-cm signal in the high redshift range of  $z = 19.8 - 25.2$  using LOFAR-LBA data with dual-pointing setup pointed at the NCP and the radio galaxy 3C220.3 simultaneously. Our main conclusions are:

(i) For the 3C220 field, after 14 hours of integration, a  $2\sigma$  upper limit of  $\Delta_{21}^2 < (14561 \text{ mK})^2$  at  $k = 0.038 \text{ h cMpc}^{-1}$  is reached on the power spectrum of 21-cm brightness temperature fluctuations. Similarly, for the NCP field, we reach a  $2\sigma$  upper limit of  $\Delta_{21}^2 < (14886 \text{ mK})^2$  at  $k = 0.038 \text{ h cMpc}^{-1}$  in the redshift range  $z = 19.8 - 25.2$ . The power spectra are still dominated by the systematics and noise. Both upper limits are consistent with each other within 2% level.

(ii) We demonstrate the application of a multiple pointing method to calibrate LOFAR-LBA dual pointing observations.

(iii) We observe excess of noise in the ratio of Stokes  $I$  and  $V$  noise spectra over short time-scales (12 seconds) in baseline-frequency space, derived from the Stokes  $I$  and  $V$

difference image-cubes created from even and odd visibility samplings at 12 second level. This excess is independent of frequency and baseline length and is also not affected by calibration. This excess noise is different from that introduced during calibration and already exists before DI and after DD calibration and does not change during those steps. The excess is different for the two fields and seems to have no clear origin. We suspect it to be caused by (diffractive) ionospheric scintillation noise, but we leave this analysis for the future.

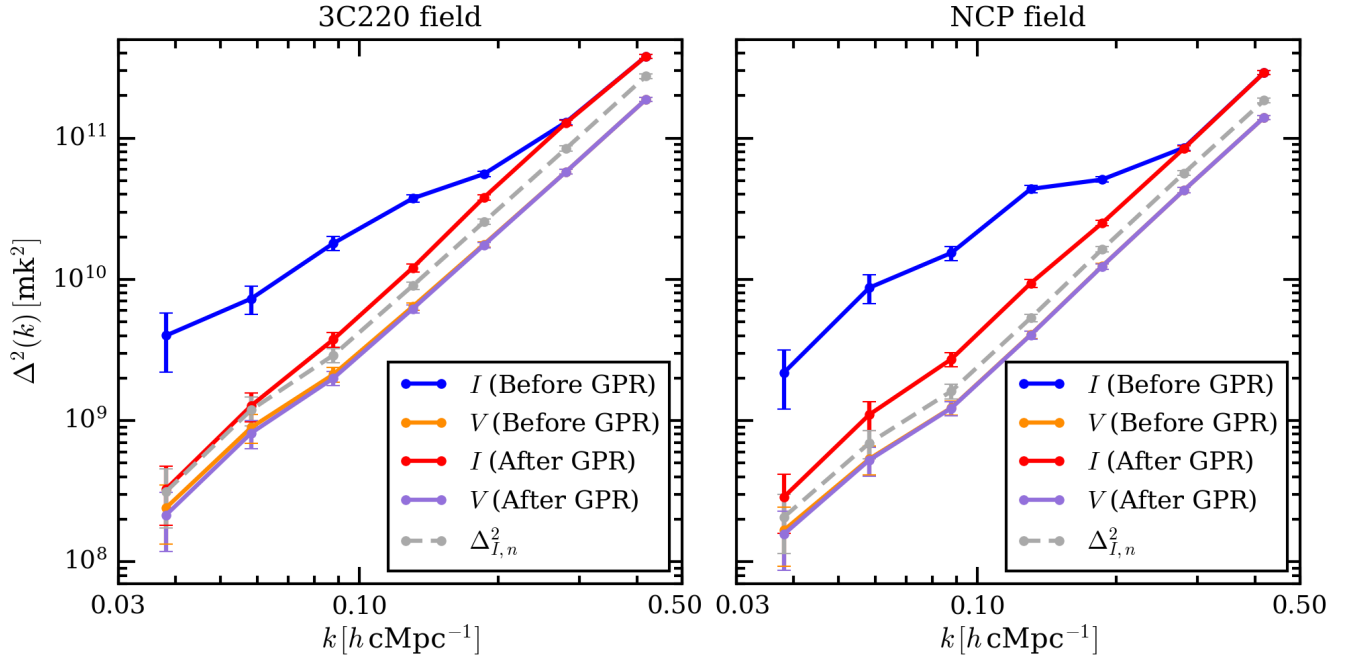
(iv) We show that introducing frequency smoothness of instrumental gains as a constraint in both Direction Independent and Direction Dependent calibration of LOFAR-LBA data greatly reduces the calibration induced excess variance on the sub-band level in Stokes  $I$  compared to Stokes  $V$  in contrast to the unconstrained case presented in G18, where we found an excess by a factor  $\sim 10$ . However, an excess of  $\sim 2 - 3$  still remains, which can be explained by the exclusion of short baselines during DD-calibration as shown for LOFAR-HBA data calibration as well in Patil et al. (2016) and Mouri Sardarabadi & Koopmans (2018).

(v) After foreground removal using Gaussian Process Regression, the Stokes  $I$  power spectrum is  $\sim 2$  times that of Stokes  $V$  for both fields and is noise-like on most scales. However, we observe a ‘pitchfork’ like structure in the 3C220 field at low  $k_\perp$  near the horizon delay line. We expect this structure to be caused by Cas A and Cyg A residuals as seen by G18.

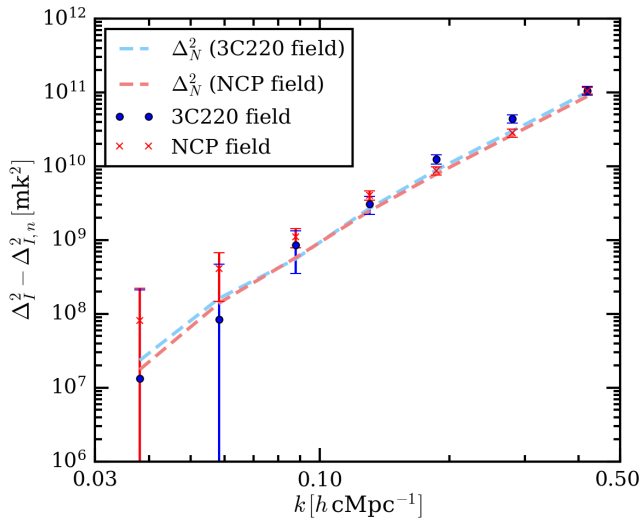
## 7.1 Outlook

Detection of the redshifted 21-cm HI signal from Cosmic Dawn and the Epoch of Reionization promises to be an excellent probe to study the early phases of the evolution of the universe and has the potential to unveil exotic astrophysical phenomena. With the analysis shown in this work, a CD experiment with LOFAR-LBA will require  $> 10^4$  hours of integration (power spectrum sensitivity of  $\sim (100 \text{ mK})^2$  in CD





**Figure 14.** The spherically averaged Stokes  $I$ ,  $V$  and excess noise bias corrected Stokes  $V$  power spectra. The left panel shows  $\Delta_I^2$  and  $\Delta_V^2$  for the 3C220 field before (blue and orange curves respectively) and after (red and purple curves respectively) foreground removal. Similarly, the right panel shows  $\Delta_I^2$  and  $\Delta_V^2$  for the NCP field using the same color scheme. The dashed gray curves represent  $\Delta_{I,n}^2$  for the corresponding fields. The errorbars represent the  $2\sigma$  errors on the power spectra.



**Figure 15.** Noise bias corrected spherically averaged Stokes  $I$  power spectra ( $\Delta_I^2 - \Delta_{I,n}^2$ ) for the 3C220 and NCP fields. Blue circles represent the 3C220 field and red crosses represent the NCP field. The dashed ‘lightblue’ and ‘coral’ colored curves represent the thermal noise power spectrum estimate  $\Delta_V^2$  for the 3C220 and the NCP field respectively. The errorbars correspond to the  $2\sigma$  errors on the power spectra.

redshift range) in order to constrain the optimistic CD X-ray heating and baryon-Dark Matter scattering models (see e.g. Fialkov et al. 2018; Cohen et al. 2018). We plan to improve the analysis in the future by improving the enforcement of

spectral smoothness in calibration, better modeling of the instrument (improving beam models) and by using the new Image Domain Gridded (IDG) combined with `WSClean` (see e.g. Veenboer et al. 2017; van der Tol, Sebastiaan et al. 2018) to subtract off-axis sources. The upcoming LOFAR 2.0 upgrade will also increase the sensitivity of the LOFAR-LBA system. The combination of all these improvements will allow us improve the CD 21-cm power spectrum sensitivity significantly.

Moreover, recently a deep absorption feature ( $-0.5$  K deep) centered at  $\sim 78$  MHz ( $z \sim 17$ ) in the averaged sky spectrum was presented by Bowman et al. (2018), possibly being the sought-after 21-cm signal absorption feature seen against the Cosmic Microwave Background during the CD era. The suggested absorption feature is considerably ( $\sim 2.5$  times) stronger and wider than predicted by standard astrophysical models (Barkana 2018). If confirmed, such a strong signal will lead to a large increase in the 21-cm brightness temperature fluctuations in the redshift range  $z = 17 - 19$  corresponding to the deepest part of the absorption profile (Barkana 2018; Fialkov et al. 2018), making it possible to detect its signal in a much shorter integration time compared to what was previously expected. Motivated by this, we have commenced a large scale program to use the Amsterdam-ASTRON Radio Transients Facility And Analysis Center (AARTFAAC) correlator based on LOFAR, to obtain wide-field data for statistical detection of the 21-cm brightness temperature fluctuations within the redshift range of the absorption feature. The techniques discussed in this paper and lessons learned here will be useful in understanding and mitigating the challenges in AARTFAAC data processing and analysis, as well as in the HERA and the upcoming SKA-low,

which can also observe the similar redshift range. The SKA-low will also support multi-beam observations, and thus also can take advantage of the dual-beam calibration strategy we have demonstrated.

## ACKNOWLEDGEMENTS

BKG and LVEK acknowledge the financial support from a NOVA cross-network grant. FGM acknowledge support from a SKA-NL roadmap grant from the Dutch Ministry of OCW. LOFAR, the Low Frequency Array designed and constructed by ASTRON, has facilities in several countries, that are owned by various parties (each with their own funding sources), and that are collectively operated by the International LOFAR Telescope (ILT) foundation under a joint scientific policy.

## REFERENCES

- Ali Z. S., et al., 2015, *ApJ*, **809**, 61
- Ali Z. S., et al., 2018, *ApJ*, **863**, 201
- Barkana R., 2018, *Nature*, **555**, 71
- Barning F. J. M., 1963, *Bull. Astron. Inst. Netherlands*, **17**, 22
- Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, *MNRAS*, **461**, 3135
- Beardsley A. P., et al., 2016, *ApJ*, **833**, 102
- Bernardi G., et al., 2016, *MNRAS*, **461**, 2847
- Blackman R. B., Tukey J. W., 1958, *Bell System Technical Journal*, **37**, 185
- Bowman J. D., et al., 2013, *Publ. Astron. Soc. Australia*, **30**, e031
- Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, *Nature*, **555**, 67
- Boyd S., Parikh N., Chu E., Peleato B., Eckstein J., 2011, *Found. Trends Mach. Learn.*, **3**, 1
- Cohen A. S., Lane W. M., Cotton W. D., Kassim N. E., Lazio T. J. W., Perley R. A., Condon J. J., Erickson W. C., 2007, *AJ*, **134**, 1245
- Cohen A., Fialkov A., Barkana R., 2018, *MNRAS*, **478**, 2193
- DeBoer D. R., et al., 2017, *PASP*, **129**, 045001
- Ewall-Wice A., et al., 2016, *MNRAS*, **460**, 4320
- Ewall-Wice A., Dillon J. S., Liu A., Hewitt J., 2017, *MNRAS*, **470**, 1849
- Fialkov A., Barkana R., 2014, *MNRAS*, **445**, 213
- Fialkov A., Barkana R., Cohen A., 2018, *Physical Review Letters*, **121**, 011101
- Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Phys. Rep.*, **433**, 181
- Gehlot B. K., et al., 2018, *MNRAS*, **478**, 1484
- Glover S. C. O., Brand P. W. J. L., 2003, *MNRAS*, **340**, 210
- Hales S. E. G., Waldram E. M., Rees N., Warner P. J., 1995, *MNRAS*, **274**, 447
- Harris F. J., 1978, *Proceedings of the IEEE*, **66**, 51
- Jacobs D. C., et al., 2016, *ApJ*, **825**, 114
- Kazemi S., Yatawatta S., Zaroubi S., 2013, *MNRAS*, **430**, 1457
- Koopmans L., et al., 2015, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, **p. 1**
- Laing R. A., Peacock J. A., 1980, *MNRAS*, **190**, 903
- Lomb N. R., 1976, *Ap&SS*, **39**, 447
- Madau P., Meiksin A., Rees M. J., 1997, *ApJ*, **475**, 429
- Mellema G., et al., 2013, *Experimental Astronomy*, **36**, 235
- Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, *MNRAS*, **478**, 3640
- Mesinger A., Furlanetto S., 2007, *ApJ*, **669**, 663
- Mesinger A., Furlanetto S., Cen R., 2011, *MNRAS*, **411**, 955
- Mouri Sardarabadi A., Koopmans L. V. E., 2018, preprint, ([arXiv:1809.03755](https://arxiv.org/abs/1809.03755))
- Offringa A. R., de Bruyn A. G., Biehl M., Zaroubi S., Bernardi G., Pandey V. N., 2010, *MNRAS*, **405**, 155
- Offringa A. R., van de Gronde J. J., Roerdink J. B. T. M., 2012, *A&A*, **539**, A95
- Offringa A. R., et al., 2014, *MNRAS*, **444**, 606
- Paciga G., et al., 2011, *MNRAS*, **413**, 1174
- Paciga G., et al., 2013, *MNRAS*, **433**, 639
- Parsons A. R., et al., 2010, *AJ*, **139**, 1468
- Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012, *ApJ*, **756**, 165
- Patil A. H., et al., 2016, *MNRAS*, **463**, 4317
- Patil A. H., et al., 2017, *ApJ*, **838**, 65
- Philip L., et al., 2018, preprint, ([arXiv:1806.09531](https://arxiv.org/abs/1806.09531))
- Planck Collaboration et al., 2016, *A&A*, **596**, A108
- Pritchard J. R., Furlanetto S. R., 2007, *MNRAS*, **376**, 1680
- Pritchard J. R., Loeb A., 2012, *Reports on Progress in Physics*, **75**, 086901
- Rasmussen C. E., Williams C. K. I., 2005, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press
- Ross H. E., Dixon K. L., Iliev I. T., Mellema G., 2017, *MNRAS*, **468**, 3785
- 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
- Stein M., 1999, *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics, Springer New York, [https://books.google.nl/books?id=5n\\_XuL2Wx1EC](https://books.google.nl/books?id=5n_XuL2Wx1EC)
- Stoica P., Li J., He H., 2009, *IEEE Transactions on Signal Processing*, **57**, 843
- Thyagarajan N., et al., 2015, *ApJ*, **804**, 14
- Tingay S. J., et al., 2013, *Publ. Astron. Soc. Australia*, **30**, e007
- Trott C. M., et al., 2016, *ApJ*, **818**, 139
- Veenboer B., Petschow M., Romein J. W., 2017, in 2017 IEEE International Parallel and Distributed Processing Symposium (IPDPS). pp 545–554, [doi:10.1109/IPDPS.2017.68](https://doi.org/10.1109/IPDPS.2017.68)
- Voytek T. C., Natarajan A., Jáuregui García J. M., Peterson J. B., López-Cruz O., 2014, *ApJ*, **782**, L9
- Yatawatta S., 2011, in 2011 XXXth URSI General Assembly and Scientific Symposium. pp 1–4, [doi:10.1109/URSIGASS.2011.6051224](https://doi.org/10.1109/URSIGASS.2011.6051224)
- Yatawatta S., 2015, *MNRAS*, **449**, 4506
- Yatawatta S., 2016, in 2016 24th European Signal Processing Conference (EUSIPCO). pp 265–269, [doi:10.1109/EUSIPCO.2016.7760251](https://doi.org/10.1109/EUSIPCO.2016.7760251)
- Yatawatta S., 2018, preprint, ([arXiv:1805.00265](https://arxiv.org/abs/1805.00265))
- Yatawatta S., Diblen F., Spreeuw H., 2017, preprint, ([arXiv:1710.05656](https://arxiv.org/abs/1710.05656))
- Zarka P., Girard J. N., Tagger M., Denis L., 2012, in Boissier S., de Laverny P., Nardetto N., Samadi R., Valls-Gabaud D., Wozniak H., eds, SF2A-2012: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. pp 687–694
- Zaroubi S., 2013, in Wiklind T., Mobasher B., Bromm V., eds, *Astrophysics and Space Science Library* Vol. 396, *The First Galaxies*. p. 45 ([arXiv:1206.0267](https://arxiv.org/abs/1206.0267)), [doi:10.1007/978-3-642-32362-1\\_2](https://doi.org/10.1007/978-3-642-32362-1_2)
- de Gasperin F., Mevius M., Rafferty D. A., Intema H. T., Fallows R. A., 2018, *A&A*, **615**, A179
- van Haarlem M. P., et al., 2013, *A&A*, **556**, A2
- van der Tol, Sebastiaan Veenboer, Bram Offringa, Andr   R. 2018, *A&A*, **616**, A27

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.