

The LOFAR radio environment

A. R. Offringa^{1,2}, A. G. de Bruyn^{1,3}, S. Zaroubi¹, G. van Diepen³, O. Martinez-Ruby¹, P. Labropoulos³, M. A. Brentjens³, B. Ciardi⁴, S. Daiboo¹, G. Harker⁵, V. Jelić³, S. Kazemi¹, L. V. E. Koopmans¹, G. Mellema⁶, V. N. Pandey³, R. F. Pizzo³, J. Schaye⁷, H. Vedantham¹, V. Veligatla¹, S. J. Wijnholds³, S. Yatawatta³, P. Zarka⁸, A. Alexov⁹, J. Anderson¹⁰, A. Asgekar³, M. Avruch^{11,1}, R. Beck¹⁰, M. Bell^{12,17}, M. R. Bell⁴, M. Bentum³, G. Bernardi^{13,1}, P. Best¹⁴, L. Birzan⁷, A. Bonafede¹⁵, F. Breitling¹⁶, J. W. Broderick¹⁷, M. Brügger^{18,15}, H. Butcher^{2,3}, J. Conway¹⁹, M. de Vos³, R. J. Dettmar²⁰, J. Eisloffel²¹, H. Falcke²², R. Fender¹⁷, W. Frieswijk³, M. Gerbers³, J. M. Griessmeier^{23,3}, A. W. Gunst³, T. E. Hassall^{17,24}, G. Heald³, J. Hessels³, M. Hoefl²¹, A. Horneffer¹⁰, A. Karastergiou²⁵, V. Kondratiev³, Y. Koopman³, M. Kuniyoshi¹⁰, G. Kuper³, P. Maat³, G. Mann¹⁶, J. McKean³, H. Meulman³, M. Mevius³, J. D. Mol³, R. Nijboer³, J. Noordam³, M. Norden³, H. Paas³, M. Pandey^{26,7}, R. Pizzo³, A. Polatidis³, D. Rafferty⁷, S. Rawlings²⁶, W. Reich¹⁰, H. J. A. Röttgering⁷, A. P. Schoenmakers³, J. Sluman³, O. Smirnov^{3,27}, C. Sobey¹⁰, B. Stappers²⁴, M. Steinmetz¹⁶, J. Swinbank⁹, M. Tagger²³, Y. Tang³, C. Tasse⁸, A. van Ardenne^{3,19}, W. van Cappellen³, A. P. van Duin³, M. van Haarlem³, J. van Leeuwen³, R. J. van Weeren^{7,3}, R. Vermeulen³, C. Vocks¹⁶, R. A. M. J. Wijers⁹, M. Wise³, and O. Wucknitz^{28,29,10}

(Affiliations can be found after the references)

Preprint online version: October 1, 2012

ABSTRACT

Aims. This paper discusses the spectral occupancy for performing radio astronomy with the Low-Frequency Array (LOFAR), with a focus on imaging observations.

Methods. We have analysed the radio-frequency interference (RFI) situation in two 24-h surveys with Dutch LOFAR stations, covering 30–78 MHz with low-band antennas and 115–163 MHz with high-band antennas. This is a subset of the full frequency range of LOFAR. The surveys have been observed with a 0.76 kHz / 1 s resolution.

Results. We measured the RFI occupancy in the low and high frequency sets to be 1.8% and 3.2% respectively. These values are found to be representative values for the LOFAR radio environment. Between day and night, there is no significant difference in the radio environment. We find that lowering the current observational time and frequency resolutions of LOFAR results in a slight loss of flagging accuracy. At LOFAR's nominal resolution of 0.76 kHz and 1 s, the false-positives rate is about 0.5%. This rate increases approximately linearly when decreasing the data frequency resolution.

Conclusions. Currently, by using an automated RFI detection strategy, the LOFAR radio environment poses no perceivable problems for sensitive observing. It remains to be seen if this is still true for very deep observations that integrate over tens of nights, but the situation looks promising. Reasons for the low impact of RFI are the high spectral and time resolution of LOFAR; accurate detection methods; strong filters and high receiver linearity; and the proximity of the antennas to the ground. We discuss some strategies that can be used once low-level RFI starts to become apparent. It is important that the frequency range of LOFAR remains free of broadband interference, such as DAB stations and windmills.

Key words. Instrumentation: interferometers - Methods: data analysis - Techniques: interferometric - Telescopes - Radio continuum: general

1. Introduction

The Low-Frequency Array (LOFAR) (van Haarlem et al., 2012, A&A, in prep.) is a new antenna array that observes the sky from 10–80 and 110–240 MHz. It currently consists of 41 (validated) stations, while 7 more are planned. The number of stations are likely to increase further in the future. Of the validated stations, 33 stations are located in the Netherlands, 5 in Germany and one each in Sweden, the UK and France. A Dutch station consists of 96 dipole low-band antennas (LBA) that provide the 10–80 MHz range, and one or two fields totalling 48 tiles of 4x4 bow-tie high-band antennas (HBA) for the frequency range of 110–240 MHz. The two different antenna types are shown in Fig. 1. The international stations have an equal number of LBAs, but 96 HBA tiles. For the latest information about LOFAR, we refer the reader to the LOFAR website¹.

The core area of LOFAR is located near the village of Exloo in the Netherlands, where the station density is at its highest. The six most densely packed stations are on the Superterp, an elevated area surrounded by water. It is an artificial island of about 350 m in diameter that is situated about 3 km North of Exloo. A map of LOFAR's surroundings is given in Fig. 2. Exloo is a village in the municipality of Borger-Odoorn in the province of Drenthe. Drenthe is mostly a rural area, and is sparsely populated relative to the rest of the Netherlands, with an average density of 183 persons/km² over 2,680 km² in 2011². Nevertheless, the radio-quiet zone of 2 km around the Superterp is relatively small, and households exists within 1 km of the Superterp. The distance from households to the other stations is even smaller in certain instances. Therefore, contamination of the radio environment by man-made electromagnetic radiation has been a ma-

¹ The website of LOFAR is <http://www.lofar.org/>.

² From the website of the province of Drenthe, <http://www.provincie.drenthe.nl/>.



Fig. 1: Antenna types of the Low-Frequency Array. Left image: A low-band antenna with a cabin in the background. Right image: Part of a high-band antenna station, consisting of 4×4 high-band antennas.

major concern for LOFAR (Bregman, 2000; Bentum et al., 2008). Because this radiation interferes with the celestial signal of interest, it is referred to as radio-frequency interference (RFI). Such radiation can originate from equipment that radiates deliberately, such as citizens’ band (CB) radio devices and digital video or audio broadcasting (DVB or DAB), but can also be due to unintentionally radiating devices such as cars, electrical fences, power lines and wind turbines (Bentum et al., 2010).

During the hardware design phase of LOFAR, careful consideration was given to ensure that the signal would be dominated by the sky noise (Cappellen et al., 2005; Wijnholds et al., 2005). This included placing shielding cabinets around equipment on site to minimise self-interference; making sure that RFI would not drive the amplifiers and analogue-digital converters (ADCs) into the non-linear regime; applying steep analogue filters to suppress the FM bands and frequencies below 10 MHz; and applying strong digital sub-band filters to localise RFI in frequency. Optionally, an additional analogue filter can be turned on to filter frequencies below 30 MHz.

Numerous techniques have been suggested to perform the task of RFI excision. They include using spatial information provided in interferometers or multi-feed systems to null directions (Leshem et al., 2000; Ellingson & Hampson, 2002; Smolders & Hampson, 2002; Boonstra, 2005; Kocz et al., 2010); removing the RFI by using reference antennas (Barnbaum & Bradley, 1998); and blanking out unlikely high values at high time resolutions (Weber et al., 1997; Leshem et al., 2000; Baan et al., 2004; Niamsuwan et al., 2005). During post-processing, RFI excision can consist of detecting the RFI in time, frequency and antenna space, and ignoring the contaminated data in further data processing. This step is often referred to as “data flagging”. Because of the major increase in resolution and bandwidth of observatories, leading to observations of tens of terabytes, manual data flagging is no longer feasible. Automated RFI flagging pipelines can solve this problem (Flöer et al., 2010; Offringa et al., 2010b). Alternative RFI strategies might be required for the detection of transients (Ryabov et al., 2004; Kocz et al., 2012).

Now that LOFAR deployment is nearly complete, commissioning observations have started and preliminary results show



Fig. 2: Map of the LOFAR core and its surroundings. The circular peninsula in the centre is the Superterp. Several other stations (triangular footprints) are visible as well. (source: OpenStreetMap)

that the choice of LOFAR’s site has not seriously degraded the data quality. For example, both the LOFAR-EoR project (de Bruyn et al., 2011) and the LOFAR project on pulsars and fast transients (Stappers et al., 2011) report that the data quality, in terms of the achieved sensitivity and calibratability, is as expected. Moreover, new algorithms and a pipeline have been implemented to automatically detect RFI with a high accuracy

(Offringa et al., 2010a,b). Preliminary results have shown that by using these algorithms, only a few percent of the data is lost due to RFI (Offringa et al., 2010b).

In this article, we study two 24-h RFI surveys: one for the 30–78 MHz low-band regime and one for the 115–163 MHz high-band regime. The observations were carried out in standard imaging mode in which visibilities are integrated to a time resolution of a second and have a spectral resolution of 0.76 kHz. In Sect. 2, we start by describing the relevant technical details of the LOFAR observatory. In Sect. 3, a brief analysis of the spectrum allocation situation relevant for LOFAR is presented. In Sect. 4, we describe the methods that are used to process and analyse the two data sets. Sect. 5 describes the details of the RFI observations that are used in this article. In Sect. 6 we describe the observational results of the two RFI surveys. We also compare them with other observations to assess whether they are representative in Sect. 7. In Sect. 8, we discuss the results and draw conclusions about the LOFAR RFI environment.

2. LOFAR

In this section, we will briefly describe the design details of LOFAR that are relevant for the impact of RFI. For further technical details, we refer the reader to de Vos et al. (2009) and van Haarlem et al. (2012, A&A, in prep.).

LOFAR consists of stations of clustered LBA and HBAs. The signals from the dual polarisation LBAs are amplified with low-noise amplifiers (LNA), and are subsequently transported over a coax cable to the electronics cabinet. The signals from the HBAs are amplified and processed by an analogue beamformer, which forms the beams for a tile of four times-four dipoles, before being sent to the cabinet. In the cabinet the signal from either the LBAs or the HBAs is band-pass filtered, digitised with a 12-bit ADC and one or more station beams are formed.

Before station beams are formed, the HBA or LBA signals are split into 512 sub-bands of 195 kHz of bandwidth, of which 244 can be selected for further processing. Other modes can optionally be processed through different signal paths. The sub-bands are formed by using a poly-phase filter (PPF) that is implemented inside the station cabinet by using field-programmable gate arrays (FPGAs). This allows for very flexible observing configurations (Romein et al., 2011). The 244 sub-band signals are transported over a dedicated wide-area network (WAN) to a Blue Gene/P (BG/P) supercomputer located in the city of Groningen. Currently, the samples are sent as 16 bit integers. However, because the transfer rate is limited to about 3 Gbit/s, the transport limits the total observed bandwidth to 48 MHz. Eight-bit and four-bit modes are scheduled to be implemented in late 2012, which would allow the transfer of 96-MHz and 192-MHz of bandwidth respectively. Multiple beams can be used, in which case the sum of the bandwidth over all beams is limited by these values.

The BG/P supercomputer applies a second PPF that increases the frequency resolution typically by a factor of 256, yielding a spectral resolution of 0.76 kHz. During this stage, the first of the 256 channels is lost for each sub-band, due to the way the PPF is implemented. Next, the BG/P supercomputer correlates each pair of stations, integrates the signal over time and applies a preliminary pass-band correction (Romein, 2008), which corrects for the response of the first (station level) poly-phase filter. Finally, the correlation coefficients are written to the discs of the LOFAR Central Processing II (CEP2) cluster.

The partitioning into sub-bands is used to distribute data over the hard discs of the computing nodes on the CEP2 clus-

ter. For storage of observations in imaging mode, LOFAR uses the CASA³ measurement set (MS) format. The first step of post-processing of the observations is RFI excision. This is performed by the AOFlogger pipeline that is described in §4.1. Further processing, such as averaging, calibration and imaging, ignores RFI contaminated data.

3. Spectrum management

In the Netherlands, the radio spectrum use is regulated by the governmental agency “Agentschap Telecom”, that falls under the Dutch Ministry of Economic Affairs, Agriculture and Innovation. This body maintains the registry of the Dutch spectrum users, which can be obtained from their website.⁴

The other countries that participate in the International LOFAR Telescope have similar bodies, and the Electronic Communications Committee⁵ (ECC), a component of the European Conference of Postal and Telecommunications Administrations (CEPT), registers the use of the spectrum at the European level. Most of the strong and harmful transmitters are allocated in fixed bands for all European countries, such as the FM radio bands, satellite communication, weather radars and air traffic communication. However, even though the allocations of the countries are similar, the usage of the allocated bands can differ. For example, several 1.792 MHz wide channels between 174 and 195 MHz are registered as terrestrial digital audio broadcasting (T-DAB) bands by the ECC. These frequencies are correspondingly allocated to T-DAB both in the Netherlands and in Germany. However, these bands are currently used in Germany, but not yet in the Netherlands. Nevertheless, the range of 216–230 MHz is actively used for T-DAB in the Netherlands. This range corresponds with T-DAB bands 11A–11D and 12A–12D, each of which is 1.792 MHz wide. These transmitters are extremely harmful for radio astronomy. Because they are wideband and have a 100% duty cycle and band usage, they do not permit radio observations. Digital video broadcasts (DVB) are similar, but occupy bands between 482 and 834 MHz (UHF channels 21–66). They are therefore outside the observing frequency range of LOFAR. Other transmitters are intermittent or occupy a narrow bandwidth, and therefore do allow radio-astronomical observations.

A short list of services with their corresponding frequencies is given in Table 1. Only a few small ranges are protected for radio-astronomy. The lowest ranges are 13.36–13.41, 25.55–25.67 and 37.5–38.25 MHz. These bands are useful for observing the Solar corona and Jovian magnetosphere, although they are too narrow, as the Sun and Jupiter emit broadband spectra. At higher LOFAR frequencies, the 150–153 MHz band is available for radio astronomy. Although the 10–200 MHz bandwidth is mostly allocated to other services, many of these — such as baby monitors — are used for short distance communication, and are therefore of low power. In addition, services such as the Citizens’ Band (CB) radio transmitters have a low duty cycle (especially during the night) and individual transmissions are of

³ CASA is the Common Astronomy Software Applications package, developed by an international consortium of scientists under the guidance of NRAO. Website: <http://casa.nrao.edu/>

⁴ The website of the Agentschap Telecom from which the spectrum registry can be obtained is <http://www.agentschaptelecom.nl/>.

⁵ The website of the Electronic Communications Committee, which registers spectrum usage at the European level, is <http://www.cept.org/ecc>, office: <http://www.ero.dk/>.

Table 1: Short list of allocated frequencies in the Netherlands in the range 10–250 MHz (source: Agentschap Telecom)

| Service type | Frequency range(s) in MHz |
|-------------------------------|------------------------------|
| Time signal | 10, 15, 20 |
| Air traffic | 10–22, 118–137, 138–144 |
| Short-wave radio broadcasting | 11–26 |
| Military, maritime, mobile | 12–26, 27–61, 68–88, 138–179 |
| Amateur | 14, 50–52, 144–146 |
| CB radio | 27–28 |
| Modelling control | 27–30, 35, 40–41 |
| Microphones | 36–38, 173–175 |
| Radio astronomy | 13, 26, 38, 150–153 |
| Baby monitor (portophone) | 39–40 |
| Broadcasting | 61–88 |
| Emergency | 74, 169–170 |
| Air navigation | 75, 108–118 |
| FM radio | 87–108 |
| Satellites | 137–138, 148–150 |
| Navigation | 150 |
| Remote control | 154 |
| T-DAB | 174–230 |
| Intercom | 202–209 |

limited bandwidth. The most problematic services for radio astronomy are therefore the FM radio (87.5–108 MHz), T-DAB (174–230 MHz) and the emergency pager (169.475–169.4875 and 169.5875–169.6 MHz) services. The FM radio range is excised by analogue filters. The emergency pager was found to be the strongest source in the spectrum. Therefore, the LOFAR signal path was designed to be able to digitise its signals correctly, i.e., without introducing non-linearities.

Around the LOFAR core, a radio-quiet zone has been established that is enforced by the province of Drenthe. The area is split into two zones. The inner zone of 2 km diameter around the core enforces full radio quietness. A “negotiation zone” with a diameter of about 10 km around the core requires negotiation before transmitters can be placed.⁶

4. Processing strategy

Processing an observation and acquiring an overview of the radio environment requires RFI detection statistics and quality assessment of the remaining data. In the following subsection, we address the detection strategy and the tools that we use for the detection. This is followed by a description of the methods for statistical analysis of RFI and data.

4.1. Detection strategy

For RFI detection, LOFAR uses the AOFlagger pipeline. This pipeline iteratively estimates the contribution of the sky by using a Gaussian high-pass filter in the time-frequency domain of a single baseline. Subsequently, the `SumThreshold` method (Offringa et al., 2010a) is used to detect line-shaped features in the same domain. A morphological operation named the scale-invariant rank (SIR) operator (Offringa et al., 2012b) is used to extent the flags into neighbouring regions that are also likely to be affected. The 4 cross-correlations (XX, XY, YX, YY) from the differently-polarised feeds are flagged individually. Finally,

⁶ The radio quiet zones are marked on “Kaart 12 — overige aanduidingen” of the environment plan of Drenthe.

if a sample is flagged in one of the cross-correlations, it is also flagged in the corresponding other cross-correlations.

The pipeline is developed in the context of the LOFAR Epoch of Reionisation key science project and was described with more detail in Offringa et al. (2010b). Compared to the strategy described there, several optimisations were made to increase the speed of the flagger. One of the changes was to use a more stable and faster algorithm to compute the morphological SIR operator (Offringa et al., 2012b). Another change was to implement several algorithms using the “streaming single-instruction-multiple-data extensions” (SSE) instruction set extension. The combined optimisations led to a decrease in the computational requirements of approximately a factor of 3, and the pipeline is input-output (IO) limited. To decrease the IO overhead, the pipeline was embedded in the “New default pre-processing pipeline” (NDPPP)⁷, which performs several tasks, such as data averaging and checking data integrity.

The AOFlagger package⁸ consists of three parts: (i) the library that implements the detection pipeline and allows for its integration into pipelines of other observatories and NDPPP; (ii) a stand-alone executable that runs the standard pipeline or a customised version; and (iii) a graphical user interface (GUI) that can be used to analyse the flagging results on a baseline-by-baseline basis and optimise the various parameters of the pipeline (see Fig. 3). The GUI was used extensively to optimise the accuracy of the pipeline. It has also been used for implementing customised strategies for data from other observatories. This has for example led to successful flagging of data from the Westerbork Synthesis Radio Telescope (WSRT) (Offringa et al., 2010a) and the Giant Metrewave Radio Telescope (GMRT) (A. D. Biggs, personal communication, Sept. 2011). Similar application of the AOFlagger on single dish data from the Parkes radio telescope also shows good initial results (J. Delhaize, personal communication, Aug. 2012).

For the data processing in this paper, we have used the original full resolution sets and applied the stand-alone flagger.

4.2. RFI and quality statistics

Assessing the quality of observations that have a volume of tens of terabyte is a non-trivial task. For example, simple operations such as calculating the mean or the root mean square (RMS) of the data are IO limited. Although these tasks can be distributed over multiple nodes if available, accessing all data of an observation still takes on the order of a few hours for large observations.

A generic solution was designed to assess the RFI situation and quality of an observation, by combining RFI statistics with other system statistics in a single platform. It consists of the following three parts: (1) a standardised storage format for the statistics; (2) software to collect the statistics; and (3) software to interpret the statistics. We will briefly describe each of these.

1. **The standardised storage format:** this was implemented as a format description of the so-called “quality tables” extension to the measurement set format⁹. The CASA measure-

⁷ See §5 of “The LOFAR Imaging Cookbook: Manual data reduction with the imaging pipeline”, ed. R. F. Pizzo et al., 2012, Astron technical document.

⁸ The AOFlagger package is distributed under the GNU General Public License version 3.0, and can be downloaded from <http://www.astro.rug.nl/rfi-software>.

⁹ described by Offringa in the technical report “Proposal for adding statistics sub-tables to a measurement set”, University of Groningen, 2011

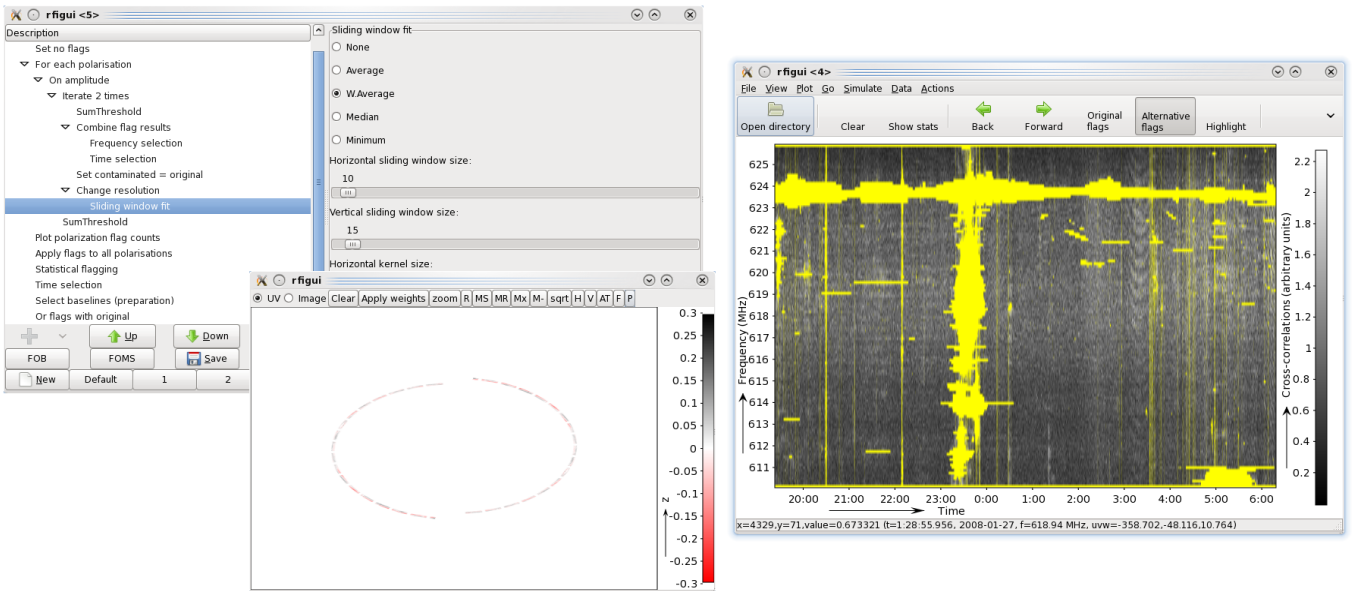


Fig. 3: Example snapshot of `rfigui`, which can be used to optimise the pipeline steps and tuning parameters. On the right is the main window showing the spectrum and flags (in yellow) of the selected baseline — in this case a GMRT data set. The left bottom window shows the uv track that this baseline covers. The upper-left window depicts the script with the actions that are performed, which can be edited interactively.

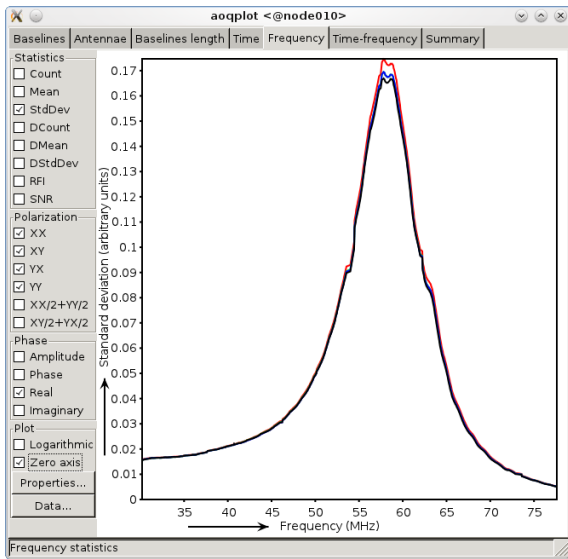


Fig. 4: The `aoqplot` tool displays the statistics interactively. In this case it shows the visibility standard deviation over frequency for a LBA observation.

ment set format allows adding custom tables, and we used this feature to add the statistics to the set. These statistics can be retrieved quickly without having to read the main data. The quality tables contain statistics as a function of frequency, time, baseline index and polarisation. The stored values allow calculation of the fraction of detected RFI in the data (RFI occupancy), the mean (signal strength), the standard deviation and the differential standard deviation as a function of time, frequency, baseline index and polarisation. The mean and standard deviation are calculated for the RFI-free samples. The differential standard deviation describes

the standard deviation of the noise by subtracting adjacent channels. Since the uncorrelated channels are only 0.76 kHz wide, the difference between adjacent channels should contain no significant contribution of the celestial signal, and are therefore a measure of the celestial and receiver noise (times $\sqrt{2}$).

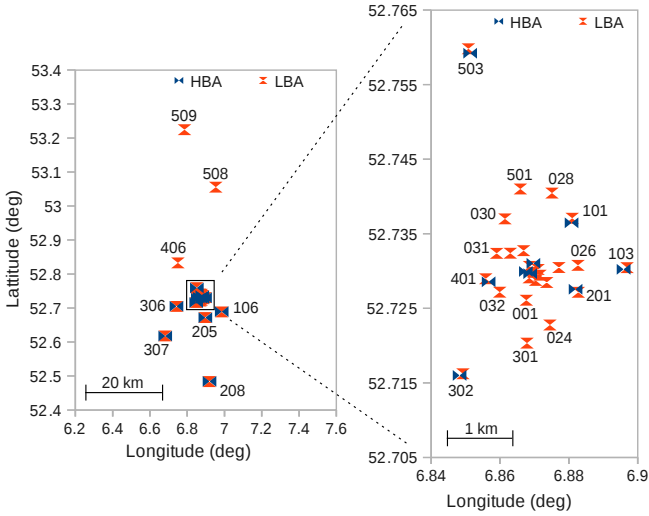
- Software to collect the statistics:** We have implemented software that collects the statistics and writes them in the described format to the measurement set. A statistics collector was added to the NDPPP averaging step. Since NDPPP is performed on most LOFAR imaging observations, all observations will thereafter have these quality tables. NDPPP is slowed down by a few per cent because the statistics have to be calculated, which is acceptable. A stand-alone tool (“`aoquality`”) is available in the AOFlogger package that can collect the statistics without having to run NDPPP.
- Software to interpret the statistics:** Once the statistics are in the described format in the tables, tools are required to read and display the quality tables. Inside the AOFlogger package is an executable (“`aoqplot`”) that performs this task: it takes either a single measurement set or an observation file that specifies where the measurement sets are located, and opens a window in which various plots can be shown and the selection can be interactively changed. An example of the plotting tool is shown in Fig. 4.

5. Description of survey data

Table 2 lists the specifications of the two 24-h RFI surveys. The number of stations used in the HBA observation was reduced to limit the volume of data. More stations were included in the LBA observation. The sets were observed at standard LOFAR time and frequency resolutions of 1 s and 0.76 kHz respectively. In both sets, the observed field was the North Celestial Pole (NCP). This field does not have a bright radio source and it is therefore easier to detect the RFI due to the absence of strong rapidly os-

Table 2: Survey data set specifications

| | LBA set | HBA set |
|----------------------|---------------|-----------------|
| Observation date | 2011-10-09 | 2010-12-27 |
| Start time | 06:50 UTC | 0:00 UTC |
| Length | 24 h | 24 h |
| Time resolution | 1 s | 1 s |
| Frequency range | 30.1–77.5 MHz | 115.0–163.3 MHz |
| Frequency resolution | 0.76 kHz | 0.76 kHz |
| Number of stations | 33 | 14 |
| Core | 24 | 8 |
| Remote | 9 | 6 |
| Total size | 96.3 TB | 18.6 TB |
| Field | NCP | NCP |

**Fig. 5:** Overview of the geometric distribution of the stations used for the RFI survey. Numbers next to the station symbols denote the station numbers.

cillating visibility fringes. Therefore, it is to be noted that if an observation is affected by very strong off-axis sources, the level of false positives might be higher than reported in this article. Only in a very few observations we see effects of strong sources that impact flagging accuracy, and this can be solved by using a customised version of the AOFlagger. The NCP field does not require tracking and fringe stopping. This might also affect the detected occupancy, since some RFI might be averaged out when applying fringe stopping. Finally, the NCP field is a good field to observe with LOFAR, because it is always at a reasonably high elevation and it is also one of the targets of the LOFAR Epoch of Reionisation project (Yatawatta et al., in prep.).

Fig. 5 shows the locations of the stations that have been used in the two surveys. For the HBA set, the stations were selected to make sure that various baseline lengths were covered and the stations had a representative geometrical coverage. Due to the inclusion of additional core stations in the LBA set, the LBA set includes more short baselines.

In the LBA set, 6 sub-bands were corrupted due to two nodes on the LOFAR CEP2 cluster that failed during observing, causing six gaps of approximately 0.2 MHz in the 48-MHz frequency span of the observation. It is expected that such losses will be less common in future observations.

6. Results

In this section, we discuss the achieved performance of the flagger, look at the RFI implications of the surveys individually and analyse their common results.

6.1. Performance

We have used the LOFAR Epoch of Reionization (EoR) cluster (see Labropoulos et al., in prep.) to perform the data analysis. This cluster consists of 80 nodes with two hyperthreaded quad-core 2.27-GHz CPUs, two NVIDIA Tesla C1060 GPU’s, 12 GB memory per node and 2 or 3 discs of approximately 2 terabyte (TB) each. The cluster is optimised for computationally intensive (GPU) tasks, such as advanced calibration and data inversion. Because it has relatively slow discs that are not in a redundant configuration (such as RAID), the cluster is not ideal for flagging, as flagging is computationally conservative and dominated by IO. To make sure the flagging would not interfere with computational tasks that were running on the cluster at that time, we chose to use only 3 CPU cores out of the 16 available cores, thus a fraction of 3/16 of the entire CPU power of the cluster. Flagging the 96-TB observation with version 2.0.1 of the AOFlagger took 40 hours, of which 32 hours were spent on reordering the observation, which consists only of reading and writing to the hard discs, and the remaining 8 hours were spent on actual flagging.

6.2. LBA survey

The default flagging pipeline found a total RFI occupancy of 2.24% in the LBA survey at a resolution of 0.76 kHz and 1 s. However, we found that the flagger had a small bias. Because the sky temperature changes due to Earth rotation, the standard deviation of the data changes over time. The flagger applies a fixed sensitivity per sub-band and per baseline, and therefore does not take into account such changes over time. This is not an issue for short observations of about less than two hours during which the sky temperature does not change significantly. However, on long observations in which the sky temperature dominates the noise level, the flagger produces more false positives when sky temperature is higher and more false negatives when the sky temperature is lower.

Unfortunately, correcting for this effect requires an accurate estimate of the sky temperature, which in turn requires the interference to be flagged. Therefore, after the first flagging run, we have applied a second run of the flagger on normalised data. In the normalised data, each timestep was divided by the standard deviation of the median timestep in a window of 15 minutes of data, thereby assuming that the first run has removed the RFI. The calculation of the standard deviation per timestep was performed on the data from all cross-correlations. Therefore, this procedure results in a very stable estimate, although the cross-correlations of longer baselines will be less affected by the Galaxy, and this method will therefore not perfectly stabilise the variance in all baselines. In this article, when we refer to a “second pass” over the data, we refer to the above described second run of the flagger. Alternatively, it is also possible to calculate the standard deviation or median of differences over a sliding window during the first run and base the detection thresholds on this quantity, but this does not match well with the SumThreshold method. The performance of the SumThreshold method would significantly decrease when it can not process the data in one consecutive run with constant sensitivity.

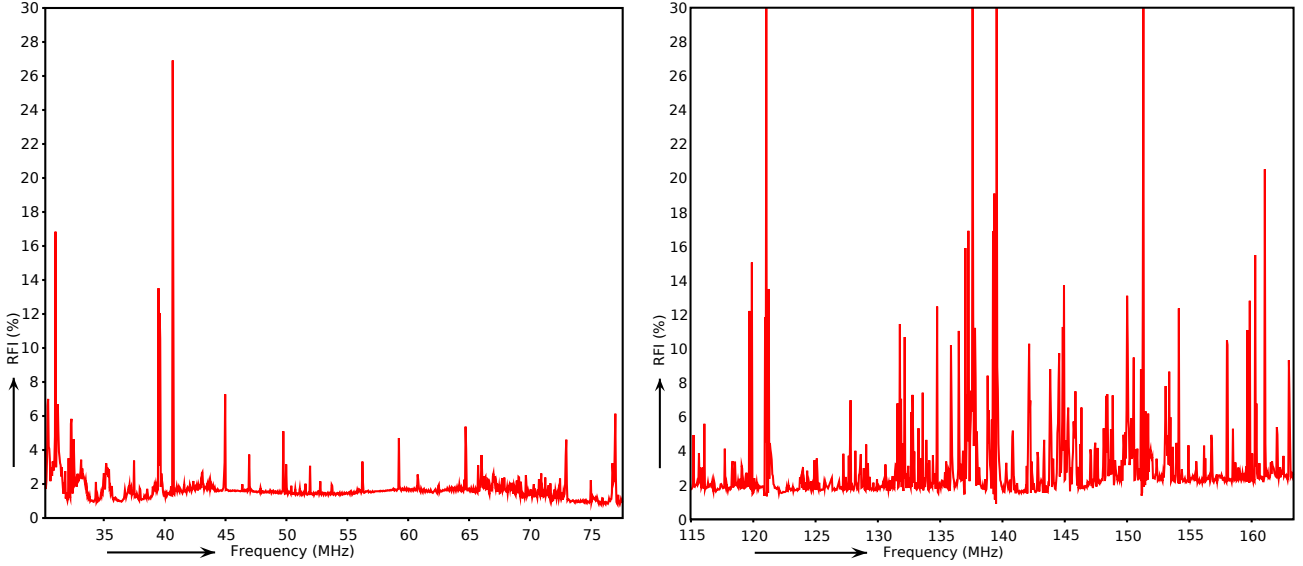


Fig. 6: The detected RFI occupancy spectra for both RFI surveys. Each data sample in the plot contains 48 kHz of data.

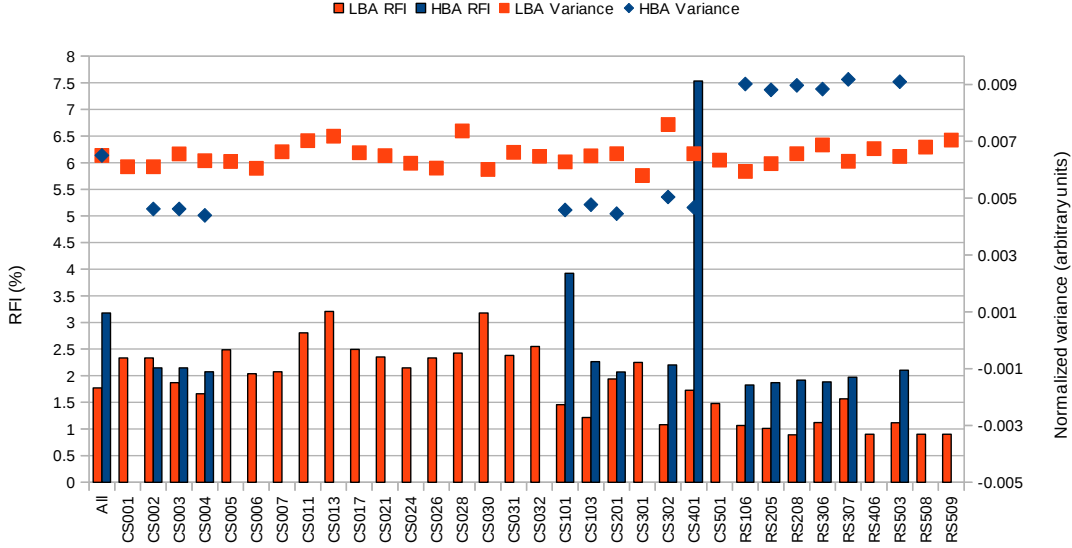


Fig. 7: The detected RFI percentages and the data variances per station, excluding auto-correlations.

The `SumThreshold` method is crucial for the accuracy of the flagger.

After having corrected for the changing sky temperature, the detected RFI occupancy is 1.77%. The RFI occupancy over frequency is plotted in Fig. 6, while Fig. 7 shows the percentages of flagged data per station. The stations with higher station numbers are generally farther away from the core, and therefore provide longer baselines. The remote stations (RS) are farthest away and for these stations, the high-band antennas are not split into two sub-stations. Fig. 7 shows that the stations closer to the core generally have a higher RFI occupancy. This can be explained by the larger number of short baselines in the central fields and the fact that RFI is decorrelated on the longer baselines. By plotting the RFI as a function of baseline length as shown in Fig. 8, it is ob-

served that the RFI decreases as a function of baseline length for lengths > 300 m, and closely follows a power law that asymptotically reaches $\sim 1.0\%$. This asymptote might be reached because of false positives and interfering sources such as satellites that do not decorrelate in the longer baselines.

Statistics in this paper are all based on cross-correlations. Detailed RFI statistics for the auto-correlations are not presented. Nevertheless, visual inspection of the auto-correlations show stronger RFI contamination and higher RFI incidence compared to the cross-correlations. Auto-correlations are typically not used for imaging or in EoR angular power spectrum measurements. However, a total power experiment using auto-correlations to detect signals from the EoR is underway, and

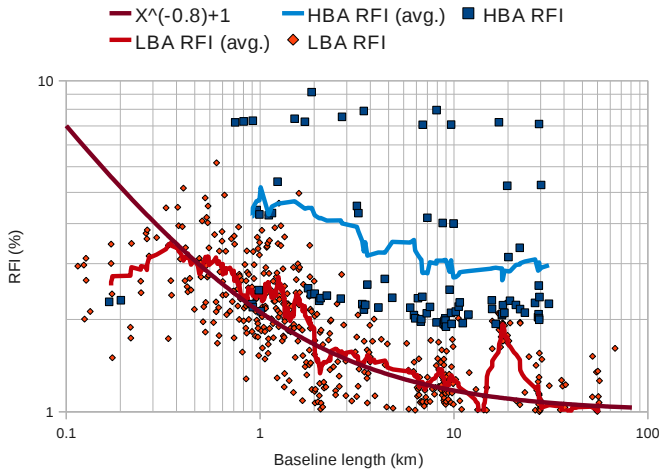


Fig. 8: RFI levels as a function of baseline length. Both axes are logarithmic. The dots represent the data (red: LBA, blue: HBA), while the lines show the trend of the points.

results from pilot observations, including RFI statistics, are in preparation (Vedantham et al., priv. com., 2012).

The LBA set contains many broadband spikes between 18:00 and 0:00 UTC. These are detected by the flagger as RFI, and are therefore visible in the dynamic RFI occupancy spectrum of Fig. 9. An example of the spikes at high resolution on a 4 km baseline is shown in Fig. 10. Individual spikes affect all samples for 1–10 seconds. Despite the relatively long baseline of 4 km, these spikes have evidently not yet become incoherent. On the 56 km baseline CS001 × RS509, the spikes are not visually present in the time-frequency plot, but some of them are still detected by the flagger because of an increase in signal to noise in these timesteps. It is assumed that they are strong ionospheric scintillations of signals from Cassiopeia A, because they correlate with its apparent position. Cas. A is 32° away from the NCP, which is the phase centre. Cygnus A might also cause such artefacts, but is 50° from the phase centre.

At the very low frequencies, around 30 MHz and 17:00–18:00 UTC, a source is visible that shows many harmonics. A high resolution dynamic spectrum is shown in Fig. 11. It is likely that this source has saturated the ADC or amplifiers. Nevertheless, its harmonics are flagged accurately, and it causes no visible effects in the cleaned data.

6.3. HBA survey

The analysis of the HBA survey shows a higher RFI occupancy of 3.18%. The increased artefacts in the RFI occupancy spectrum of the HBA in Figs. 6 and 9 also confirm that the HBA is more contaminated by interference than the LBA. However, as can be seen in Fig. 7, almost all stations have less than 2.5% RFI occupancy. Stations CS101HBA0 and CS401HBA0 are the only two exceptions, with respectively 3.9% and 7.5% RFI, and are also a cause of the higher level of RFI compared to the LBA survey. Despite the larger fraction of RFI in stations CS101HBA0 and CS401HBA0, the data variances of these are similar to the other stations. This suggests therefore the presence of local RFI sources such as a sparking electric fence or a lawn mower near these two stations, which have successfully been excised by the flagger. This RFI source seems to have been temporary, as re-

cent observations show normal RFI detection occupancies of less than 3% for data from this station. Fig. 7 also shows that the variances of the remote stations are higher. This is because these stations contain twice as many antennas.

As in the case of the LBA survey, detected RFI occupancies in the HBA are affected by the changing sky temperature. Again we have performed a second pass in which the normalised data was flagged. However, because the HBA system is far less sky noise dominated than the LBA system (Wijnholds & van Cappellen, 2011), the noise level in the HBA data is less affected by the changing sky. Consequently, the difference between the first and second pass is minor, and after the second pass the detected level of RFI is less by only 0.04%.

In Fig. 8, for the HBA it is harder to assess whether the level of RFI decreases significantly on longer baselines due to the smaller number of baselines.

6.4. Overall results

After the automated RFI detection, there are generally no harmful interference artefacts in the data at the level at which we make images at the moment. The variance over frequency and time are displayed in respectively Fig. 12 and Fig. 13, and are displayed in a time-frequency diagram in Fig. 14. While the HBA variances look clean in most frequencies, there are a few spikes of RFI that evidently have not been detected. These look like sharp features in the full spectrum, but are in fact smooth features when looking at full resolution. Because they are smooth at the raw sub-band resolution, the flagger does not detect them as RFI. Although there are interference artefacts visible in the HBA spectrum, after detection the data can be successfully calibrated and imaged. A possible second stage flagger to remove any residual artefacts will be discussed in §8. The LBA variances show only a few RFI artefacts around its higher frequencies.

The HBA spectrum contains a clearly visible ripple of about 1 MHz. This has been identified as the result of reflection over the cables, resulting from an impedance mismatch in the receiver unit. In fact, a similar phenomenon occurs in LBA observations, but because of the steeper frequency response and because not all LBA cables are of the same length, it is less apparent. The reflection is also less strong in the LBA, due to the better receiver design. A Fourier transform of the LBA variance over frequency shows slight peaks at twice the delays of the cables.

6.5. Day and night differences

One might expect a lower RFI occupancy during the night, i.e., during 23:00–6:00 UTC (Local time is UTC+1). We use Fig. 13 to assess this possibility. The figure shows variance and RFI occupancy as a function of the hour of the day in UTC. However, after one pass of flagging, the data are highly dominated by the changing sky. Moreover, the LBA data also contain artefacts due to Cassiopeia A, which causes some spikes in the data due to strong ionospheric scintillation between 18:00 and 0:00 UTC.

Unfortunately, the biasing effect of the sky temperature is not completely removed even with a second pass over the data. There is no significant additional trend visible. This implies that there is no significant relation between the hour of the day and the RFI occupancy due to less activity at night. This is also evident in the dynamic spectra of RFI in Fig. 9, which show no obvious increase or decrease of transmitters during some part of the day, and many transmitters start and end at random times. In a few cases, the starting of a transmitter at a certain frequency

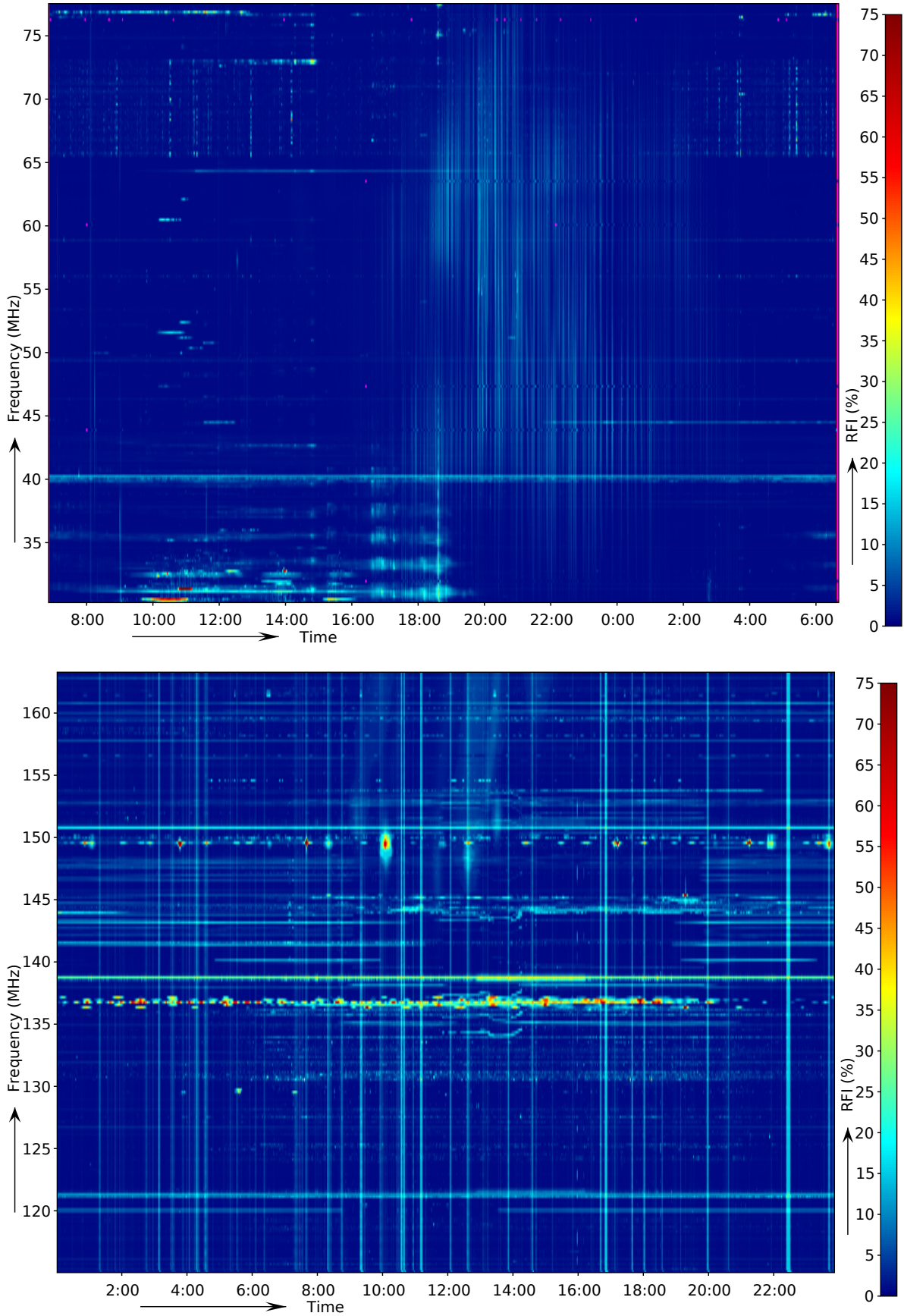


Fig. 9: Dynamic RFI occupancy spectrum for the surveys. Colour intensity represents the fraction of samples that were occupied in a specific time-frequency bin. The average over all baselines is shown. Top: LBA, bottom: HBA. The broad-band features in the LBA are likely to be ionospheric effects on Cas A.

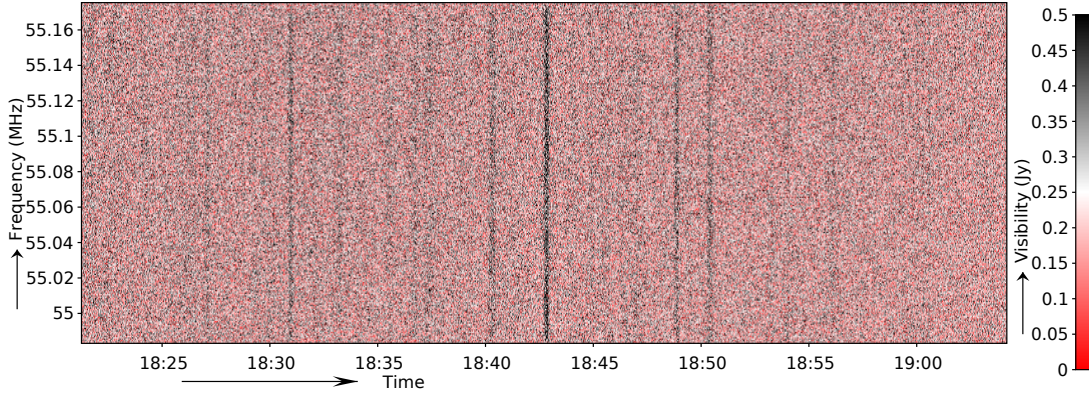


Fig. 10: Data from the LBA 4-km long baseline CS001 \times RS503 at high frequency resolution, showing strong fluctuations of 1–10 s. The flagger detects these as RFI.

coincides with the termination of a transmitter at a different frequency, suggesting that some transmitters hop to another frequency. In Fig. 9, such transmissions can be seen between 140 and 145 MHz. These transmissions end at 9:00 UTC, while at the same time several transmissions start around 135–140 MHz.

To further explore the possibility of increased RFI during daytime of the HBA set, we have performed the same analysis on a 123–137 MHz subset of the HBA observation. There are two reasons that the difference between day and night might be better visible in this frequency bandwidth: (i) all the visual peaks of detected RFI that correspond to the Sun have a frequency higher than 145 MHz; and (ii) this band corresponds to air traffic communication, which is less used during the night. Nevertheless, we still do not see a significant increase of RFI in this subset of the data.

In summary, any effect of increased activity during the day is not significant enough to be identifiable in the detected occupancies of either the LBA or the HBA data set. The post-flagging data variances are dominated by celestial effects, i.e., the Sun, the Milky Way or Cassiopeia A, and contain no clear signs of a relation between day and night time either.

6.6. Resolution & flagging accuracy

The frequency and time resolution of observations do affect the accuracy of the interference detection. It is, however, not known how significant this effect is. To quantify this, we have decreased the frequency resolution of the HBA RFI survey in several steps and reflagged the averaged set. Subsequently, the resulting flags were compared with the flags that were found at high resolution. The original high resolution flags were used as ground truth.

We found that the level of false positives is approximately linearly correlated with the decrease in resolution. Unfortunately, false positives cause samples in our ground truth to be misclassified as RFI, and will therefore show up as false negatives in the lower resolution detections. Therefore, the false positives for the ground truth data were determined by extrapolating the false-positives curve of the sets with decreased resolution. This yields a false-positives rate of 0.3%, which subsequently has been subtracted from the false negatives. The resulting curves after these corrections are plotted in Fig. 15.

Because the test is computationally expensive, we have not performed the same test on the LBA survey or for the time resolution. However, tests on small parts of the data show that

decreasing the time resolution results in similar false-negative curves compared with decreasing the frequency resolution, although it causes about 20% less false positives. Therefore, from the RFI detection perspective, it is slightly better to have higher frequency resolution compared to higher time resolution at LOFAR resolutions. It is still to be ascertained whether the small amount of data was representative enough to draw generic conclusions.

6.7. False-positives rate

If we assume that the least contaminated sub-bands in Fig. 9 are completely free of RFI on the long baselines, they can be used to determine the false-positive rate of the flagger. For the LBA set, we selected the 4-km long baseline CS001 \times RS503 and the 56-km long baseline CS001 \times RS509 of one of the best centre sub-bands at 55 MHz. For the 4-km baseline the total detected fraction of RFI is 0.75%, while for the 56-km baseline it is 0.73%. However, the 4-km baseline contains some broadband spikes around 18:40 h, as shown in Fig. 10. On the 56-km baseline CS001 \times RS509, the spikes can not be seen in the time-frequency plot, but some of them are still detected by the flagger because of an increase in signal to noise in these timesteps.

To get a more accurate estimate of the base level of false positives, we have also determined the false-positives rate by using only the last 50 min of the sub-bands. Visual inspection of this data shows indeed no RFI, except for two timesteps in the 4 km baseline that might have been affected, but these can not be assessed with certainty. The flagger does flag those timesteps, hence we ignore them in the analysis. When flagging only the 50 minutes of 4 km baseline data, thereby making sure that the threshold is based only on this 50 min of data, a fraction of 0.6% was flagged. If one assumes that the selected data contains no other RFI, then this value is the rate of falsely flagged samples. In the 56 km baseline, the same analysis leads to a slightly lower rate of false-positives of 0.5%.

The 0.6 and 0.5% detection rates are the result of flagging on all four cross-correlations (XX, XY, YX and YY). In the samples that have been detected as RFI, we observe that there are zero samples flagged in more than one cross-correlation for that particular time and frequency, thus they are completely uncorrelated. Each cross-correlation adds independently about 0.13–0.15% of falsely detected samples. In a simulated baseline with

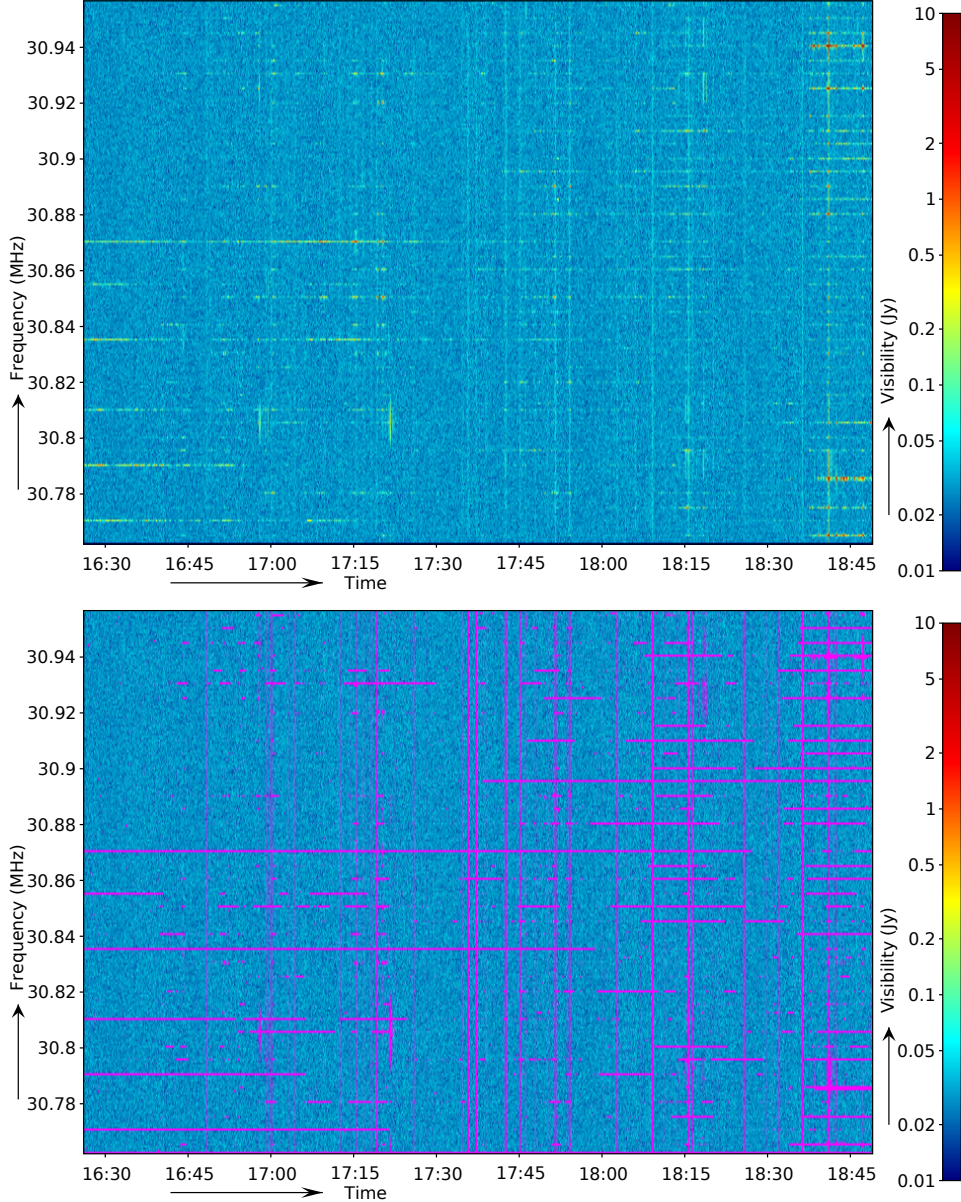


Fig. 11: A dynamic spectrum of data from one sub-band of the LBA survey, formed by the correlation coefficients of baseline $CS001 \times CS002$ at the original frequency resolution of 0.76 kHz. The displayed sub-band is one of the most affected sub-bands in terms of the detected level of RFI. The top image shows the original spectrum, while the bottom image shows with purple what has been detected as interference.

complex Gaussian noise the flagger detects 0.14% as RFI, thus these values are similar to the expected ones.

Estimating the false-negatives rate is harder to carry out, because we do not know the exact interference distribution. Because there are almost no RFI artefacts after flagging, the false-negatives can be assumed to be insignificant in most cases.

7. Comparison with other observations

Although we have analysed a substantial amount of survey time, it is useful to validate whether the two observations are representative samples for determining the LOFAR interference environment. Unfortunately, comparing the surveys with other observations is hard at this point, because often during LOFAR commissioning observations are being carried out with lower frequency

and time resolutions to reduce the data size, and the analysed 24 h surveys are the only substantial observations performed at the standard LOFAR resolution. A relative comparison can still be done for lower resolution data. There are no strong sources in the targeted NCP field, which further complicates the comparison. Fields that do have strong sources might trigger the flagger more easily, yielding higher detection rates.

To assess the differences between different observations, we have performed detection occupancy analysis of several other observations. For this purpose, we collected several LOFAR observations that were used for quality assessment. These were subsequently processed similarly to how we processed the surveys. The observations were selected independent of their quality, hence they sample the RFI situation randomly. However, it is important to note that in our experience the data quality, such

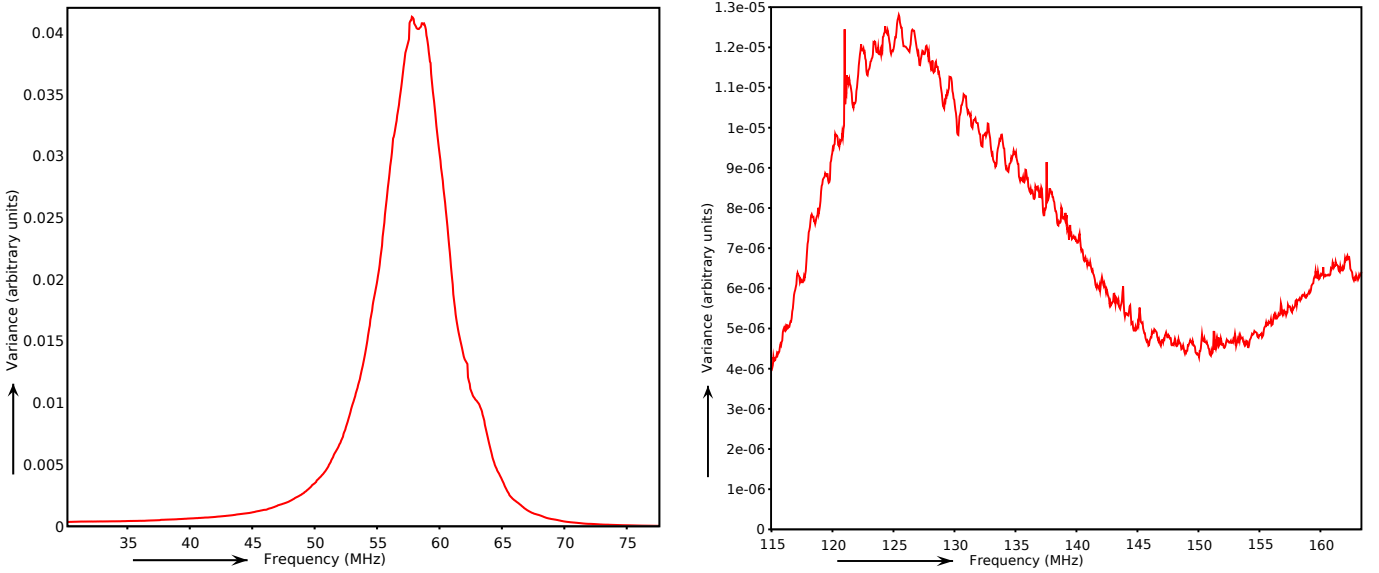


Fig. 12: The post-flagging spectra of data variances for both RFI surveys. The dominating effect is the antenna frequency response. In the HBA (right plot), a strong ripple of around 1 MHz is apparent, which is caused by reflections in the antenna cables.

Table 3: Observations and their RFI occupancy as reported by automated detection. The bold entries are the surveys analysed in this article.

| Date | Start (UTC) | Duration | Id | Target | $\Delta\nu$ (kHz) | Δt (s) | RFI ^[1] |
|--|-------------|-------------|---------------|------------|-------------------|----------------|--------------------|
| <i>LBA observations (frequency range $\approx 30 - 78$ MHz)</i> | | | | | | | |
| 2010-11-20 | 19.33 | 5 min | L21478 | Moon | 3.0 | 1 | 4.6% |
| 2010-11-20 | 19.43 | 6 h | L21479 | Moon | 3.0 | 1 | 10.3% |
| 2011-04-14 | 19.00 | 8 h | L25455 | Moon | 0.76 | 1 | 4.3% |
| 2011-10-09 | 6.50 | 24 h | L31614 | NCP | 0.76 | 1 | 1.8% |
| <i>HBA observations (frequency range $\approx 115 - 163$ MHz)</i> | | | | | | | |
| 2010-11-21 | 20.26 | 5 min | L21480 | Moon | 3.0 | 1 | 5.6% |
| 2010-12-27 | 0.00 | 24 h | L22174 | NCP | 0.76 | 1 | 3.2% |
| 2011-03-27 | 20.00 | 6 h | L24560 | NCP | 3.0 | 2 | 1.5% |
| 2011-04-01 | 16.08 | 6 h | L24837 | 3C196 | 3.0 | 2 | 2.6% |
| 2011-06-11 | 11.30 | 1.30 h | L28322 | 3C196 | 3.0 | 2 | 6.5% |
| 2011-11-17 | 18.00 | 12 h | L35008 | NCP | 3.0 | 2 | 3.6% |
| 2011-12-06 | 2.36 | 25 min | L36691 | 3C196 | 3.0 | 2 | 5.5% |
| 2011-12-06 | 8.34 | 25 min | L36692 | 3C295 | 3.0 | 2 | 8.0% |
| 2011-12-20 | 7.39 | 30 min | L39562 | 3C295 | 3.0 | 2 | 2.5% |
| 2012-01-26 | 2.00 | 5.30 h | L43786 | 3C295 | 3.0 | 2 | 3.6% |

Notes:

^[1] RFI occupancy as found by automated detection. For some targets, this is too high because of the band-edge issues that are discussed in the text, leading to approximately a 1–2% increase in 3-kHz channel observations.

as the achieved noise level of the final image, is quite independent of the detected RFI occupancy. Much more relevant is the position of the Sun in the sky, the state of the ionosphere and the stability of the station beam. These have very little effect on the detected RFI occupancy.

Table 3 lists these other observations and shows their statistics. The number of involved stations varies between the observations, but as many as possible core stations were used in all observations.

Currently, there is an issue with some LOFAR observations that causes higher RFI detection rates in fields with strong sources. This is caused by the edges of sub-bands in some cross-correlated baselines. These edges are flagged because they show time-variable changes that are very steep in the frequency direc-

tion. This effect is only observed in cross-correlations that involve exactly one Superterp station, so it is assumed that this is a bug in the station beamformer or correlator. In 64 channel observations that show this issue, the first and last sub-band channels get flagged in about half of the baselines, leading to about a 1–2% higher detected RFI occupancy. The issue only arises in fields that contain strong sources, and is consequently not affecting the 24 h RFI surveys, because there are no such sources in the NCP field. All 3C196, 3C295 and Moon observations do show the issue.

The average detected RFI occupancies are 5.4 and 4.3% with standard deviations 3.5 and 2.0% for the LBA and HBA observations respectively. Therefore, it appears that the analysed 24 h RFI surveys, with 2.4 and 3.2% RFI occupancy in the low and

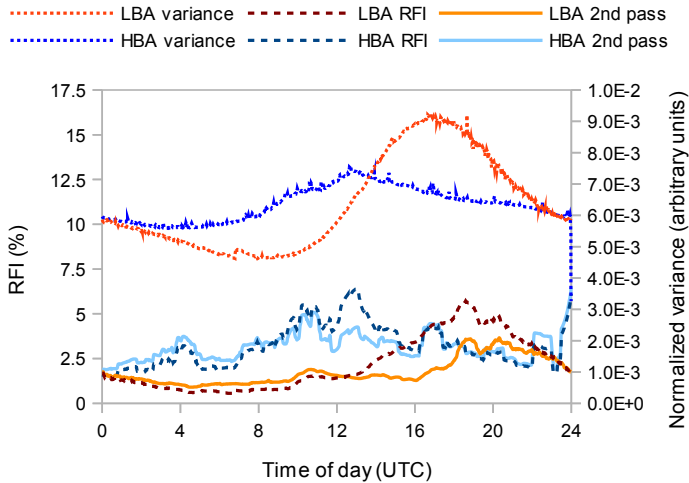


Fig. 13: RFI levels and variances as a function of the time of day. The RFI percentages are smoothed. Although there is some variation in the detected RFI during the observation, this is likely not because of a different occupation of RFI between day and night. Instead, they are likely caused by the changing sky, since they correlate with the variance of the data and visual celestial artefacts in the dynamic spectra.

high bands respectively, are less affected by RFI than the average observation. If one however assumes that the observations with lower time and frequency resolutions have an approximately 1.0% RFI increase, which seems to be a reasonable estimate according to Fig. 15, and taking into account that the subband-edge issue causes another 1.5% RFI increase on average in the fields with strong sources, the averages after correction for these effects become 3.7 and 2.4%. Therefore, the RFI occupancies of the 24 h surveys seem to be reasonably representative for the RFI occupancy of LOFAR at its nominal resolution of 0.76 kHz with 1 s integration time. On the other hand, it also shows that 3 kHz channels may well suffice for regular LOFAR observations.

Visual inspection of the same data agreed with this observation: the RFI environment is not significantly different between different observations. The only exception was the Moon observation of 2010-11-20, which seems to contain unusual broadband interference over the entire duration of the observation. Note that the moon is known to reflect some of the RFI, but such reflections are too faint to trigger the flagger. The shape and frequency at which the interference occurred is not like in any other observation. Therefore, we suspect that either something went wrong during this particular observation or ionospheric conditions were exceptional. According to weather reports, it was observed at the day with highest humidity of the year, although we have no explanation why this would influence the RFI occupancy.

8. Discussion & conclusions

We have analysed 24-h RFI surveys for both the high-band and low-band frequency range of LOFAR. Both sets show a very low contamination of detectable interference of 1.8 and 3.2% for the LBA and HBA respectively. In the considered frequency ranges,

these are predicted to be representative quantities for what can be expected when LOFAR starts its regular observing with resolutions of 0.76 kHz and 1 s. Therefore, the LOFAR radio environment is relatively benign, and is not expected to be the limiting factor for deep field observing. However, it remains important that the spectrum is not used for broadband transmitters such as DAB stations. Also strong local interference can become a problem. For example, it is currently not clear what the effect of windmills close to the LOFAR stations might be, since these can potentially reflect and generate additional and time-varying interference. We have also not considered LOFAR's entire frequency range, but instead focused on the most sensitive region. This region is probably the least contaminated by RFI, because the RFI situation is worse below 30 MHz and above 200 MHz. We have focused on the RFI situation for imaging observations. The RFI situation might be different when observing with a much higher time resolution, as is done for the LOFAR transient key science project.

Almost all visible interference is detected after the single flagging step at highest resolution, and RFI that leaks through is very weak. This agrees with the first imaging results, which are thought to be limited by beam and ionospheric calibration issues and system temperature, but not by interference. However, whether this will still be the case for long integration times of tens of nights, as will be done as part of the Epoch of Reionization project, remains to be seen. In that case, one might find that weak, stationary RFI sources add up coherently, and might at some point become the limiting factor. Nevertheless, the situation looks promising: our first-order flagging routines use only per-baseline information, but remove in most cases all RFI that is visible in the spectra. The resulting integrated statistics of 24 hours show very few artefacts of interference, and these are causing no obvious issues when calibrating and imaging the data.

If RFI does become a problem, there are many methods at hand to further excise it. The interference artefacts still present can be flagged with a second stage flagger. In such a stage, the flagger could use the information from the entire observation, and such a strategy would be more sensitivity for weak stationary sources. Moreover, the Fourier transform used for imaging is a natural filter of stationary interference. Without fringe stopping, a single baseline will observe a stationary source as a constant source. Therefore, the contribution of stationary sources would end up at the North Pole. With sufficient uv-coverage, the sidelobe of this source at the NCP will be benign. Furthermore, if necessary these can be further attenuated with filtering techniques, such as low-pass filters that remove contributions in the data with a fringe frequency faster than can be generated by on-axis sources (Offringa et al., 2012a). Therefore, we believe that RFI will not keep LOFAR from reaching its planned sensitivity.

Unexpectedly, we found that the RFI occupancy is not significantly different between day and night. In both the system temperature of the instrument and the detected RFI occupancy, the setting of the Galaxy and the Sun overshadow the influence caused by true RFI sources, and this is the only structured variation over time that is apparent in the data. Therefore, RFI is not a factor for deciding whether to observe at day or night. Of course, there are other reasons to conduct low-frequency observations at night, especially because of the stronger effect of the ionosphere and the presence of the Sun during the day, which both make successful calibration more challenging.

We estimate the false-positives rate of the AOFlagger pipeline to be 0.5–0.6%, based on the level of falsely detected samples in clean-appearing data. The resulting loss in sensitiv-

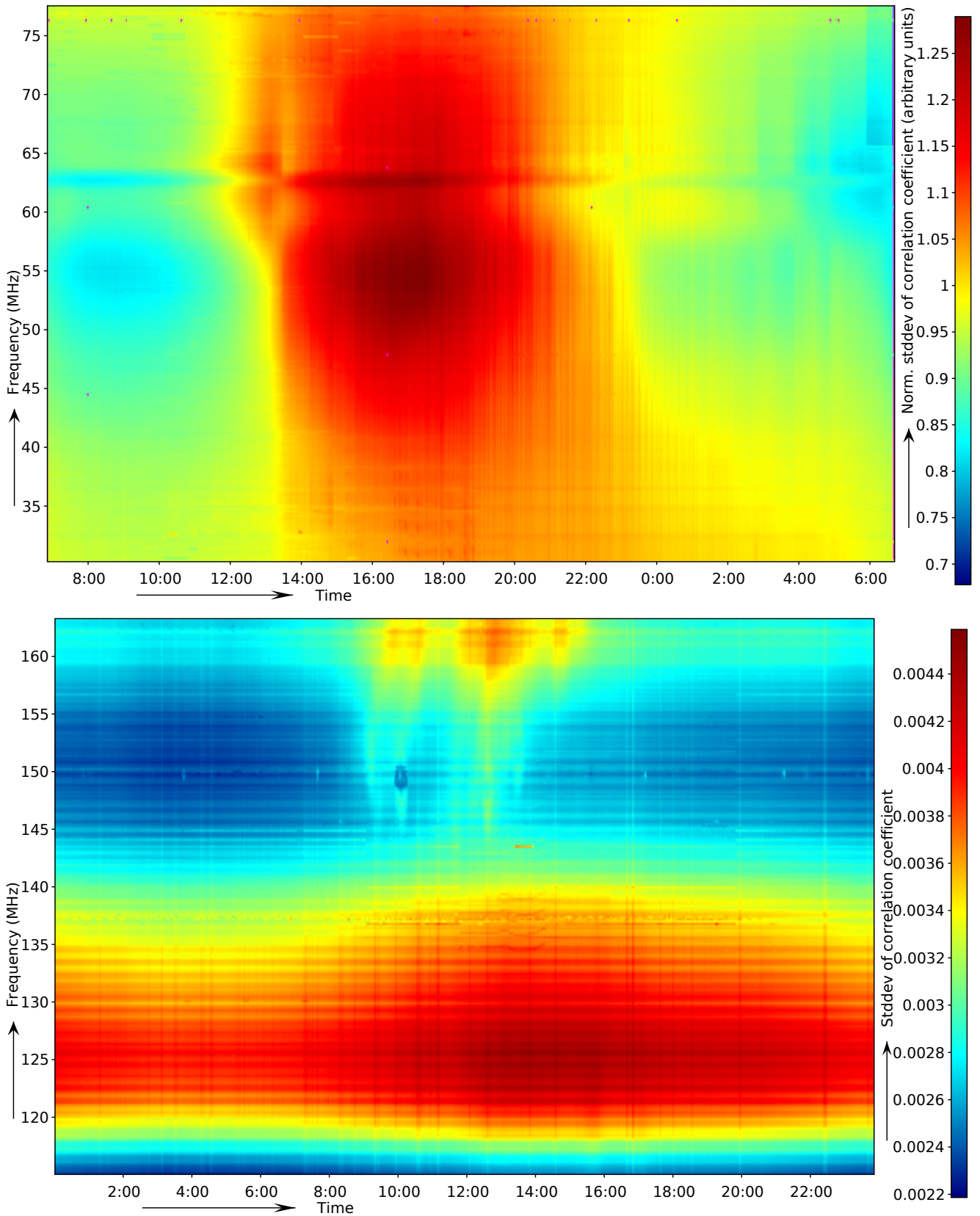


Fig. 14: The standard deviation over time and frequency during the surveys. In the LBA set, the individual statistics of each sub-band were divided by the Winsorized mean of the sub-band, to correct for the antenna response on first order. In the LBA set, no residual RFI is visible, except some weak residuals near the edges of the band. A few purple dots can be seen in the data, which denotes missing data. The HBA set shows a bit more undetected, but weak RFI residuals.

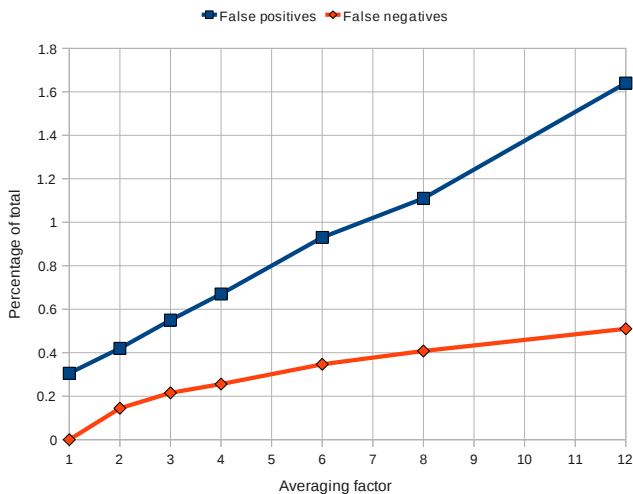


Fig. 15: This plot shows the RFI detection accuracy as a function of frequency resolution, using data from the LBA survey. The frequency resolution is 0.76 kHz at an averaging factor of 1. The resolution is lowered by averaging the samples in adjacent channels. The time resolution is fixed at 1 s.

ity is therefore negligible. We have seen that during long observations, in which the system temperature changes due to the setting of the Galaxy and the Sun, time ranges with increased variance result in higher levels of false detections. Therefore, it would be a good practice to apply the correction method that was used for the LBA set: by (temporarily) dividing the samples by an accurate estimate of the standard deviation before flagging the data, the rate of false-positives will become constant for timesteps with a different sky temperature. This requires two runs of the flagger: one run to be able to estimate the variance on clean data, and one more to flag the data with the normalised standard deviation. This decreases the level of false-positives by about 0.5% (a total detected rate of 1.77% instead of 2.24%) on LBA sets and will also decrease the number of false negatives in areas of low variance, but because of the smaller field of view of the HBA array, the improvement is less significant there. It is computationally twice as expensive, and is not necessary for short observations that do not show a significant change in sky temperature.

Up to now, interference detection was often performed manually and ad-hoc by the observer. Consequently, few statistics are available in the literature that describe the amount of data loss in cross-correlated data due to interference for a particular observatory and frequency range, but some studies have been performed. A systematic analysis of interference at the Mauritius Radio Telescope showed an average RFI occupancy of 10% (Pandey & Shankar, 2005). In general, compared to data losses achieved with common RFI excision strategies, the loss in LOFAR data is low. This is especially surprising considering the fact that LOFAR is built in a populated area and operates at low frequency. Several reasons can be given for the small impact of RFI on LOFAR:

- Many interfering sources contaminate a narrow frequency range or short duration. LOFAR’s high time and frequency resolutions, of 1 s and 0.76 kHz respectively, minimise the amount of data loss caused by such interfering sources. Since

the current loss of data is small, it seems unnecessary to go to even higher resolutions.

- LOFAR is the first telescope to use many novel post-correlation detection methods, such as the scale-invariant rank operator and the `SumThreshold` techniques, which allow detection with high accuracy.
- LOFAR’s hardware is designed to deal with the strong interfering sources that are found in its environment. The receiver units remain in linear state in the neighbourhood of such sources, and the strong band-pass filters spectrally localise the sources. Consequently, almost no interfering source will cause ramifications in bands that are adjacent to their transmitting frequency. The only exception is at very low frequencies, where we do see a very strong source saturate the ADCs when ionospheric conditions are bad. This source and its harmonics are successfully removed during flagging.
- Propagation models for Earth-bound signals show a strong dependence on the height of the receiver (e.g., Hata (1980)). In contrast to dishes with feeds in the focal point, the receiving elements of LOFAR are close to the ground.
- LOFAR is remotely controlled, and the in situ cabins with electronics are shielded. We have found no post-correlation contamination that is caused by self-generated interference. This is in contrast with for example the WSRT, where the dishes close to the control room (which contains the correlator, but it is operated from elsewhere) are known to observe more interference. In the LOFAR auto-correlations, every now and then we do see some artefacts that suggest local interference, but these do not visibly contaminate cross-correlations. It might be that forming station beams before correlation helps reducing such RFI as well.

Given the low impact of RFI on LOFAR, we can conclude that the interference environment should not have an absolute weight in site selection of future (low-frequency) radio telescopes — or its substations — for example for the Square Kilometre Array. Instead, it should be carefully weighted against the non-negligible costs of logistics that are involved in building and maintaining a telescope in a remote area, and when dealing with low frequencies, against the quality of the ionosphere for performing radio astronomy.

In this article, we have not yet looked at the Gaussianity of the signal and the implications of the statistical distribution of RFI. Such statistical properties of RFI sources might have implications on long integrations, such as the LOFAR EoR project. We will deal with this in future work.

References

- Baan, W. A., Fridman, P. A., & Millenaar, R. P. 2004, *AJ*, 128, 933
 Barnbaum, C. & Bradley, R. F. 1998, *AJ*, 115, 2598
 Bentum, M., Boonstra, A.-J., Millenaar, R., & Gunst, A. 2008, in *URSI General Assembly 2008 (Chicago: URSI)*
 Bentum, M. J., Boonstra, A. J., & Millenaar, R. P. 2010, *Proc. of RFI2010*
 Boonstra, A. J. 2005, PhD thesis
 Bregman, J. D. 2000, in *SPIE Proc. Astr. Telescopes and Instr.*, ed. H. R. Butcher, Vol. 4015 (SPIE), 19–32
 Cappellen, W. A., Bregman, J. D., & Arts, M. J. 2005, *Exp. Astr.*, 17, 101
 de Bruyn, A. G., Brentjens, M. A., Koopmans, L. V. E., et al. 2011, in *General Assembly and Scientific Symposium, 2011 XXXth URSI*, 1–4
 de Vos, M., Gunst, A. W., & Nijboer, R. 2009, *Proceedings of the IEEE*, 97, 1431
 Ellingson, S. W. & Hampson, G. A. 2002, *IEEE Trans. on Antennas & Propagation*, 50, 25
 Flöer, L., Winkel, B., & Kerp, J. 2010, in *Proc. of RFI2010*
 Hata, M. 1980, *IEEE Trans. on Vehicular Technology*, VT-29, 317
 Kocz, J., Bailes, M., Barnes, D., Burke-Spolaor, S., & Levin, L. 2012, *Monthly Notices of the Royal Astronomical Society*, 420, 271

- Kocz, J., Briggs, F. H., & Reynolds, J. 2010, *AJ*, 140, 2086
- Leshem, A., van der Veen, A.-J., & Boonstra, A.-J. 2000, *ApJS*, 131, 355
- Niamsuwan, N., Johnson, J. T., & Ellingson, S. W. 2005, *Radio Science*, 40
- Offringa, A. R., de Bruyn, A. G., Biehl, M., & Zaroubi, S. 2010b, in *Proc. of RFI2010, Astron (PoS)*
- Offringa, A. R., de Bruyn, A. G., Biehl, M., et al. 2010a, *MNRAS*, 405, 155
- Offringa, A. R., de Bruyn, A. G., & Zaroubi, S. 2012a, *MNRAS*, 422, 563
- Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M. 2012b, *A&A*, 539
- Pandey, V. N. & Shankar, N. U. 2005, in *Proc. of XXVIIIth URSI General Assembly*
- Romein, J. W. 2008, *Bandpass correction in LOFAR*, Tech. rep., ASTRON
- Romein, J. W., Mol, J. D., van Nieuwpoort, R. V., & Broekema, P. C. 2011, in *URSI General Assembly, 2011 (IEEE)*, 1–4
- Ryabov, V., Zarka, P., & Ryabov, B. 2004, *Planetary and Space Science*, 52, 1479
- Smolders, B. & Hampson, G. 2002, *IEEE Antennas & Propagation magazine*, 44, 13
- Stappers, B. W., Hessels, J. W. T., Alexov, A., et al. 2011, *A&A*, 530, A80
- Weber, R., Faye, C., Biraud, F., & Dansou, J. 1997, *A&AS*, 126, 161
- Wijnholds, S. J., Bregman, J. D., & Boonstra, A.-J. 2005, *Exp. Astr.*, 17, 35
- Wijnholds, S. J. & van Cappellen, W. A. 2011, *IEEE Trans. on Antennas and Propagation*, 59, 1981
-
- ¹ University of Groningen, Kapteyn Astronomical Institute, PO Box 800, 9700 AV Groningen, The Netherlands. e-mail: offringa@mso.anu.edu.au
 - ² Mount Stromlo Observatory, RSAA, Cotter Road, Weston Creek, ACT 2611, Australia.
 - ³ ASTRON, PO Box 2, 7990 AA Dwingeloo, The Netherlands.
 - ⁴ Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, 85748 Garching bei München, Germany.
 - ⁵ Center for Astrophysics and Space Astronomy, University of Colorado, 389 UCB, Boulder, Colorado 80309-0389, USA.
 - ⁶ AlbaNova University Center, Department of Astronomy, SE-106 91 Stockholm, Sweden.
 - ⁷ Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands.
 - ⁸ Observatoire de Paris, FR 92195 Meudon, France.
 - ⁹ University of Amsterdam, Astronomical Institute Anton Pannekoek, PO Box 94249, 1090 GE Amsterdam, The Netherlands.
 - ¹⁰ Max-Planck Institute for Astrophysics, P.O. Box 20 24, D-53010 Bonn, Germany.
 - ¹¹ SRON Netherlands Institute for Space Research, PO Box 800, 9700 AV Groningen, The Netherlands.
 - ¹² Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW 2006, Australia.
 - ¹³ Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138
 - ¹⁴ Royal Observatory Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK.
 - ¹⁵ Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany.
 - ¹⁶ Astrophysical Institute Potsdam, An der Sternwarte 16, 14482 Potsdam, Germany.
 - ¹⁷ University of Southampton, University Road, Southampton SO17 1BJ, UK.
 - ¹⁸ University of Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany.
 - ¹⁹ Chalmers University of Technology, SE-412 96 Gothenburg, Sweden.
 - ²⁰ Ruhr-University Bochum, Universitätsstrae 150, 44801 Bochum, Germany.
 - ²¹ Thüringer Landessternwarte, Tautenburg Observatory, Sternwarte 5, D-07778 Tautenburg, Germany.
 - ²² Radboud University Nijmegen, Faculty of NWI, PO Box 9010, 6500 GL Nijmegen, The Netherlands.
 - ²³ Centre national de la recherche scientifique, 3 rue Michel-Ange, 75794 Paris cedex 16, France.
 - ²⁴ University of Manchester, Oxford Road, Manchester, M13 9PL, UK.
 - ²⁵ University of Oxford, Wellington Square, Oxford OX1 2JD, UK.
 - ²⁶ Centre de Recherche Astrophysique de Lyon, Observatoire de Lyon, 9 av Charles André, 69561 Saint Genis Laval Cedex, France.
 - ²⁷ Rhodes University, RATT, Dep. Physics and Electronics, PO Box 94, Grahamstown 6140, South Africa.
 - ²⁸ University of Bonn, Regina-Pacis-Weg 3, D-53012 Bonn, Germany.
 - ²⁹ Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany.