

LOFAR MSSS: detection of a low-frequency radio transient in 400 h of monitoring of the North Celestial Pole

A. J. Stewart,^{1,2}★ R. P. Fender,^{1,2} J. W. Broderick,^{1,2,3} T. E. Hassall,^{1,2}
 T. Muñoz-Darias,^{1,2,4,5} A. Rowlinson,^{3,6} J. D. Swinbank,^{6,7} T. D. Staley,^{1,2}
 G. J. Molenaar,^{6,8} B. Scheers,^{6,9} T. L. Grobler,^{8,10} M. Pietka,^{1,2} G. Heald,^{3,11}
 J. P. McKean,^{3,11} M. E. Bell,^{12,13} A. Bonafede,¹⁴ R. P. Breton,^{2,15} D. Carbone,⁶
 Y. Cendes,⁶ A. O. Clarke,^{2,15} S. Corbel,^{16,17} F. de Gasperin,¹⁴ J. Eislöffel,¹⁸
 H. Falcke,^{3,19} C. Ferrari,²⁰ J.-M. Grießmeier,^{17,21} M. J. Hardcastle,²² V. Heesen,²
 J. W. T. Hessels,^{3,6} A. Horneffer,²³ M. Iacobelli,³ P. Jonker,^{19,24} A. Karastergiou,¹
 G. Kokotanekov,⁶ V. I. Kondratiev,^{3,25} M. Kuniyoshi,²⁶ C. J. Law,²⁷
 J. van Leeuwen,^{3,6} S. Markoff,⁶ J. C. A. Miller-Jones,²⁸ D. Mulcahy,^{2,15} E. Orru,³
 M. Pandey-Pommier,²⁹ L. Pratley,³⁰ E. Rol,³¹ H. J. A. Röttgering,³²
 A. M. M. Scaife,¹⁵ A. Shulevski,¹¹ C. A. Sobey,³ B. W. Stappers,¹⁵ C. Tasse,^{8,10,33}
 A. J. van der Horst,³⁴ S. van Velzen,¹⁹ R. J. van Weeren,³⁵ R. A. M. J. Wijers,⁶
 R. Wijnands,⁶ M. Wise,^{3,6} P. Zarka,^{17,36} A. Alexov,³⁷ J. Anderson,³⁸ A. Asgekar,^{3,39}
 I. M. Avruch,^{11,24} M. J. Bentum,^{3,40} G. Bernardi,³⁵ P. Best,⁴¹ F. Breitling,⁴²
 M. Brüggen,¹⁴ H. R. Butcher,⁴³ B. Ciardi,⁴⁴ J. E. Conway,⁴⁵ A. Corstanje,¹⁹
 E. de Geus,^{3,46} A. Deller,³ S. Duscha,³ W. Frieswijk,³ M. A. Garrett,^{3,32}
 A. W. Gunst,³ M. P. van Haarlem,³ M. Hoeft,¹⁸ J. Hörandel,¹⁹ E. Jütte,⁴⁷ G. Kuper,³
 M. Loose,³ P. Maat,³ R. McFadden,³ D. McKay-Bukowski,^{48,49} J. Moldon,³ H. Munk,³
 M. J. Norden,³ H. Paas,⁵⁰ A. G. Polatidis,³ D. Schwarz,⁵¹ J. Sluman,³ O. Smirnov,^{8,10}
 M. Steinmetz,⁴² S. Thoudam,¹⁹ M. C. Toribio,³ R. Vermeulen,³ C. Vocks,⁴²
 S. J. Wijnholds,³ O. Wucknitz,²³ and S. Yatawatta³

Affiliations are listed at the end of the paper

Accepted 2015 November 26. Received 2015 November 24; in original form 2015 July 17

ABSTRACT

We present the results of a four-month campaign searching for low-frequency radio transients near the North Celestial Pole with the Low-Frequency Array (LOFAR), as part of the Multifrequency Snapshot Sky Survey (MSSS). The data were recorded between 2011 December and 2012 April and comprised 2149 11-min snapshots, each covering 175 deg². We have found one convincing candidate astrophysical transient, with a duration of a few minutes and a flux density at 60 MHz of 15–25 Jy. The transient does not repeat and has no obvious optical or high-energy counterpart, as a result of which its nature is unclear. The detection of this event implies a transient rate at 60 MHz of $3.9_{-3.7}^{+14.7} \times 10^{-4} \text{ d}^{-1} \text{ deg}^{-2}$, and a transient surface

*E-mail: adam.stewart@physics.ox.ac.uk

density of $1.5 \times 10^{-5} \text{ deg}^{-2}$, at a 7.9-Jy limiting flux density and ~ 10 -min time-scale. The campaign data were also searched for transients at a range of other time-scales, from 0.5 to 297 min, which allowed us to place a range of limits on transient rates at 60 MHz as a function of observation duration.

Key words: instrumentation: interferometers – techniques: image processing – radio continuum: general.

1 INTRODUCTION

The variable and transient sky offers a window into the most extreme events that take place in the Universe. Transient phenomena are observed at all wavelengths across a diverse range of objects, ranging from optical flashes detected in the atmosphere of Jupiter caused by bolides (Hueso et al. 2010) to violent Gamma-Ray Bursts (GRBs) at cosmological distances which can outshine their host galaxy (Klebesadel, Strong & Olson 1973; van Paradijs et al. 1997). Observations at radio wavelengths provide a robust method to probe these events, supplying unique views of kinetic feedback and propagation effects in the interstellar medium, which are also just as diverse in their associated time-scales. Active galactic nuclei (AGN; Matthews & Sandage 1963; Smith & Hoeffleit 1963) are known to vary over time-scales of a month or longer, whereas observations of the Crab Pulsar have seen radio bursts with a duration of nanoseconds (Hankins et al. 2003).

Historically, and still to this day, radio observations have been used to follow-up transient detections made at other wavelengths. Radio facilities generally had a narrow field of view (FoV), which made them inadequate to perform rapid transient and variability studies over a large fraction of the sky. However, blind transient surveys have been performed and have produced intriguing results. For example, Bower et al. (2007) (also see Frail et al. 2012) discovered a single-epoch millijansky transient at 4.9 GHz while searching 944 epochs of archival Very Large Array (VLA) data spanning 22 years, with three other possible marginal events. Sky surveys using the Nasu Observatory have also been successful in finding a radio transient source, with Niinuma et al. (2007) having observed a two-epoch event, peaking at 3 Jy at 1.42 GHz. Various counterparts were considered at other wavelengths, but the origin of the transient remains unknown. Lastly, Bannister et al. (2011) surveyed 2775 deg^2 of sky at 843 MHz using the Molonglo Observatory Synthesis Telescope (MOST), yielding 15 transients at a 5σ level of 14 mJy beam^{-1} , 12 of which had not been previously identified as transient or variable.

Surveys at low frequencies ($\leq 330 \text{ MHz}$) have also been completed. Lazio et al. (2010) carried out an all-sky transient survey using the Long Wavelength Demonstrator Array (LWDA) at 73.8 MHz, which detected no transient events to a flux density limit of 500 Jy. In addition, Hyman et al. (2002, 2005, 2006, 2009) discovered three radio transients during monitoring of the Galactic Centre at 235 and 330 MHz. These were identified by using archival VLA observations along with regular monitoring using the VLA and the Giant Metrewave Radio Telescope (GMRT). The transients had flux densities in the range of $100 \text{ mJy} - 1 \text{ Jy}$ and occurred on time-scales ranging from minutes to months. Lastly, Jaeger et al. (2012) searched six archival epochs from the VLA at 325 MHz centred on the *Spitzer-Space-Telescope* Wide-Area Infrared Extragalactic Survey (SWIRE) Deep Field. In an area of 6.5 deg^2 to a 10σ flux limit of $2.1 \text{ mJy beam}^{-1}$, one day-scale transient event was reported with a peak flux density of $1.7 \text{ mJy beam}^{-1}$.

Radio transient surveys are being revolutionized by the development of the current generation of radio facilities. These include

new low-frequency instruments such as the International Low-Frequency Array (LOFAR; van Haarlem et al. 2013), Long Wavelength Array (LWA; Ellingson et al. 2013) and the Murchinson Wide Field Array (MWA; Tingay et al. 2013). The telescopes listed offer a large FoV coupled with an enhanced sensitivity, with LOFAR having the capability to reach sub-mJy sensitivities and arcsecond resolutions (though this full capability is not used in this work as such modes were being commissioned at the time.). These features are achieved by utilizing phased-array technology with omnidirectional dipoles, and the above mentioned telescopes act as pathfinders for the low-frequency component of the Square Kilometre Array (SKA; Dewdney et al. 2009). With such greatly improved sensitivities at low frequencies, we have a new opportunity to survey wide areas of the sky for transients and variables, with a particular sensitivity to coherent bursts.

These new facilities have already produced some interesting results in this largely unexplored parameter space. Bell et al. (2014) searched an area of 1430 deg^2 for transient and variable sources at 154 MHz using the MWA. No transients were found with flux densities $> 5.5 \text{ Jy}$ on time-scales of 26 min and one year. However, two sources displayed potential intrinsic variability on a one year time-scale. Using the LWA, Obenberger et al. (2014a) detected two kilojansky transient events while using an all-sky monitor to search for prompt low-frequency emission from GRBs. They were found at 37.9 and 29.9 MHz, lasting for 75 and 100 s, respectively, and were not associated with any known GRBs. This was followed up by Obenberger et al. (2014b) who searched over 11 000 h of all-sky images for similar events, yielding 49 candidates, all with a duration of tens of seconds. It was discovered that 10 of these events correlated both spatially and temporally with large meteors (or fireballs). This low-frequency emission from fireballs was previously undetected and identifies a new form of naturally occurring radio transient foreground.

Two transient studies have now also been completed using LOFAR. Carbone et al. (2015) searched 2275 deg^2 of sky at 150 MHz, at cadences of 15 min and several months, with no transients reported to a flux limit of 0.5 Jy. Cendes et al. (2015) searched through 26, 149-MHz observations centred on the source Swift J1644+57, covering 11.35 deg^2 . No transients were found to a flux limit of 0.5 Jy on a time-scale of 11 min.

In this paper we use the LOFAR telescope to search 400 h of observations centred at the North Celestial Pole (NCP; $\delta = 90^\circ$), covering 175 deg^2 with a bandwidth of 195 kHz at 60 MHz. LOFAR is a low-frequency interferometer operating in the frequency ranges of 10–90 MHz and 110–250 MHz. It consists of 46 stations: 38 in the Netherlands and eight in other European countries. Full details of the instrument can be found in van Haarlem et al. (2013).

A previous study of variable radio sources located near the NCP field ($75^\circ < \delta < 88^\circ$) was carried out by Mingaliev et al. (2009). This study identified 15 objects displaying variability at centimetre wavelengths on time-scales of days or longer. However, the variability amplitude was found to be within seven per cent, which we would not be able to distinguish with LOFAR due to general calibration uncertainties at the time of writing. In addition, the lower

observing frequency used in this work would mean that the expected peak flux densities would be significantly lower, assuming a standard synchrotron event (e.g. van der Laan 1966), making them challenging to detect. Also, the lower frequency means that the variability would occur over even longer time-scales, again assuming that the emission arises from a synchrotron process.

The observations and processing techniques are discussed in Section 2, with a description of how the transient search was performed in Section 3. The results can be found in Section 4, which is followed by a discussion of a discovered transient event in Section 5. The implied transient rates and limits are discussed in Section 6, before we conclude in Section 7.

2 LOFAR OBSERVATIONS OF THE NCP

The monitoring survey of the NCP was performed between 2011 December 23 and 2012 April 16, resulting in a total of 2609 observations being recorded. The NCP was chosen because it is constantly observable from the Northern hemisphere, and the centre of the field is located towards constant azimuth and elevation (*az/el*) coordinates. However, this is not true for sources which lie away from the NCP, where these sources rotate within the LOFAR elliptical beam. We therefore restrict our transient search to an area around the NCP where the LOFAR station beam properties are consistent for each epoch observed, avoiding systematic errors in the light curves that might be introduced if this was not the case. It is also an advantage that the line of sight ($b = 122^{\circ}.93, l = +27^{\circ}.13$) is located towards a relatively low column density of Galactic free electrons; the maximum expected dispersion measure (DM) is 55 pc cm^{-3} according to the NE2001 model of the Galactic free electron distribution (Cordes & Lazio 2003).

The NCP measurements were taken using the LOFAR Low-Band Antennas (LBA) at a single frequency of 60 MHz; the bandwidth was 195 kHz, consisting of 64 channels. The total integration time of each snapshot was 11 min, sampled at 1 s intervals, and data were recorded using the ‘LBA_INNER’ setup, where the beam is formed using the innermost 46 LBA antennas from each station, which gives the largest possible FoV and a full width half-maximum (FWHM) of $9^{\circ}.77$.

2.1 Observation epochs

The programme piggybacked on another commissioning project being performed by LOFAR at the time, the Multifrequency Snapshot Sky Survey (MSSS) – the first major LOFAR observing project surveying the low-frequency sky (Heald et al. 2015). With every single MSSS LBA observation that took place, a beam was placed on the NCP using one subband of the full observational setup for MSSS. Fig. 1 shows a histogram of the number of NCP snapshots observed each day over the duration of the programme, in addition to a similar histogram showing the number of snapshots per hour for a particular set of days. Of the 2609 snapshots, 909 were recorded during the day and 1700 were recorded at night. The MSSS observational set-up also meant that each 11-min snapshot in the same observation block was separated by a time gap of four minutes.

2.2 Calibration and imaging

Before any processing took place, radio-frequency interference (RFI) was removed using AOFLAGGER (Offringa et al. 2010; Offringa, de Bruyn & Zaroubi 2012a; Offringa, van de Gronde & Roerdink 2012b) with a default strategy, in addition, the two channels at the highest, and lowest, frequency edges of the measurement set were

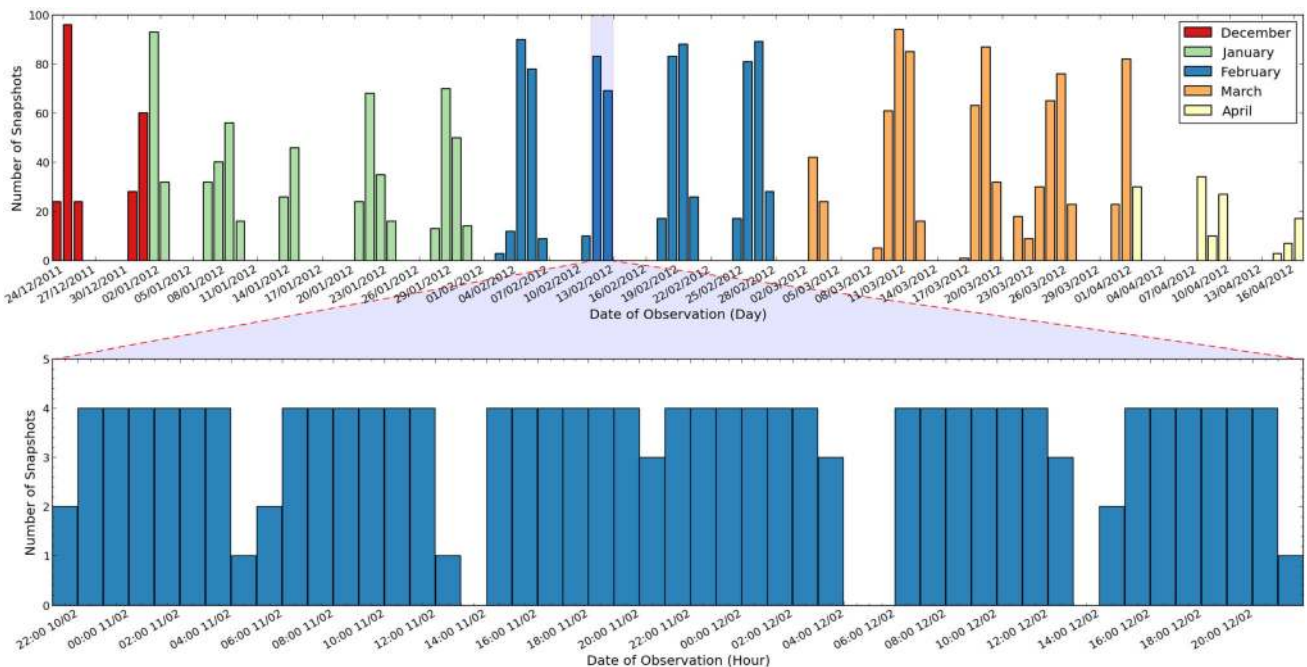


Figure 1. Histograms giving a general overview of when the 2609 11-min snapshots of the NCP were observed. The top panel contains a histogram showing how many snapshots were observed on each day over the entire 4 month period, colour coded by month, which shows the distinct observing blocks in which NCP observations were obtained. The bottom panel displays a ‘zoom-in’ of the date range 21:00 2012/02/10–21:00 2012/02/12 UTC, now showing the number of snapshots per hour. This emphasizes further the sometimes fragmented nature of the observing pattern of the NCP, with which careful consideration had to be given on how to combine the observations for the transient search.

also completely flagged, reducing the bandwidth to 183 kHz. When using an automatic flagging tool such as `AOFLAGGER`, it is important to be aware of the fact that transient sources could be mistakenly identified as RFI by the software. This is a complex issue which is beyond the scope of this work. However, an initial investigation for the LOFAR case was carried out by Cendes et al. (2015). In these tests, simulated transient sources, described by a step function, with different flux densities and time durations (from seconds to minutes), were injected into an 11 min data set. These data sets were subsequently passed through `AOFLAGGER` before calibrating and imaging as normal in order to observe how the simulated transient was affected by the automatic flagging, if at all. The authors concluded that transient signals shorter than a duration of two minutes could be partially, or in the case of \sim Jansky level sources, completely flagged. However, there are some caveats to this testing: short time-scale imaging was not tested for short-duration transients, and it remains to be determined how the automatic flagging would treat other types of transients (i.e. a non-step function event). Hence, while these results certainly suggest that transients could be affected by `AOFLAGGER`, further testing is required to completely understand how automatic flagging software can affect the detection of a transient.

At this stage we also removed all data from international LOFAR stations, leaving just the Dutch stations. This was due to the complex challenges in reducing these corresponding data at the time of processing. Following this, the ‘demixing’ technique (described by van der Tol, Jeffs & van der Veen 2007) was used to remove the effects of the bright sources Cassiopeia A and Cygnus A from the visibilities. Finally, averaging in frequency and time was performed such that each observation consisted of 1 channel and an integration time of 10 s per time-step. The averaging of the data was necessary to reduce the data volume and computing time required to process the data.

This averaging has the potential to introduce effects caused by bandwidth and time smearing, which are discussed in more detail by Heald et al. (2015) in relation to MSSS data. Following Heald et al. (2015), we used the approximations given by Bridle & Schwab (1999) to calculate the magnitude of the flux loss (S/S_0) in each case, assuming a projected baseline length of 10 km. We found the bandwidth smearing factor to equal seven per cent (using a field radius corresponding to the FWHM) and a time smearing factor of 0.4 per cent. Thus, while the effect of time smearing was negligible, the impact of bandwidth smearing was potentially significant, yet remained within the calibration error margins (10 per cent; see Section 4.1).

A selection of flux calibrators, characterized by Scaife & Heald (2012),¹ were used in the main processing of the data and were observed simultaneously utilizing LOFAR’s multi-beam capability (thus the calibrator scans were also 11 min in length). The calibrators and their usage can be found in Table 1. The standard LOFAR imaging pipeline was then implemented which consists of the following steps. First, the amplitude and phase gain solutions, using XX and YY correlations, are obtained for each calibrator observation using Black Board Selfcal (BBS; Pandey et al. 2009). These solutions are direction-independent, and are derived for each time step using the full set of visibilities from the Dutch stations, as well as a point source model of the calibrator itself. Beam calibration

Table 1. List of the calibrators used for the NCP observations. It was decided early in the MSSS programme that 3C 48 and 3C 147 might not be adequate as calibrators for the LBA portion of the survey, and so these were dropped 8 and 22 d after first use, respectively. Observations using these calibrators displayed no disadvantages over those observed with other calibrators when checked in this project, and hence they were kept as part of the sample.

Calibrator source	% use	First use date	Last use date
3C 48	2%	2011 Dec 24	2012 Jan 01
3C 147	6%	2011 Dec 23	2012 Jan 14
3C 196	43%	2011 Dec 24	2012 Apr 14
3C 295	40%	2011 Dec 24	2012 Apr 01
Cygnus A	9%	2012 Jan 28	2012 Apr 16

was also enabled which accounts, and corrects, for elevation and azimuthal effects with the station beam. The amplitudes of these gain solutions were then clipped to a 3σ level to remove significant outliers, which were not uncommon in these early LOFAR data. The gain solutions were then transferred directly from the calibrators to the respective NCP observation.

Secondly, a phase-only calibration step was performed (also using BBS) to calibrate the phase in the direction of the target field. The solutions were derived using data within a maximum projected uv distance of 4000λ (20 km; 24 core + 10 remote stations). In order to perform this step, a sky model was obtained of the NCP field using data from the global sky model (GSM) developed by Scheers (2011). This model is constructed by first gathering sources which are present within a set radius from the target pointing in the 74 MHz VLA Low-Frequency Sky Survey (VLSS; Cohen et al. 2007). In the NCP case, the radius was set to 10° . From this basis, sources are then cross-correlated, using a source association radius of 10 arcsec, with the 325 MHz Westerbork Northern Sky Survey (WENSS; Rengelink et al. 1997) and the 1400 MHz NRAO VLA Sky Survey (NVSS; Condon et al. 1998) to obtain spectral index information. In those cases where no match was found, the spectral index, α (using the definition $S_\nu \propto \nu^\alpha$), was set to a canonical value of $\alpha = -0.7$. No self-calibration was performed on the data. The reader is referred to van Haarlem et al. (2013) for more LOFAR standard pipeline information.

The main MSSS project discovered that observations recorded during this 2011–2012 period potentially contained one or more bad stations, and the data quality would improve if such stations were removed. LOFAR was still very much in its infancy at the time, and, as a result, was not entirely stable; problems such as network connection issues or bad digital beam forming contributed to the poor performance of some stations. Hence, an automated tool was developed which analysed each station, identifying and flagging those that displayed a significant number of baselines with high measured noise. This tool was utilized in the NCP processing and primarily removed stations with poorly focused beam responses (Heald et al. 2015). It should be noted that present LOFAR data no longer require this tool as the issues outlined above have been rectified.

Finally, an FoV of 175 deg^2 was imaged using the `AWIMAGER` (Tasse et al. 2013), with a robust weighting parameter of 0 (Briggs 1995), and a primary-beam (PB) correction applied to each image. A maximum projected baseline length of 10 km was used in this study (2000λ ; 24 core + seven remote stations). This was chosen to obtain good uv coverage and a maximum resolution for which we were confident with the calibration. The typical resolution for the 11-min snapshots was $5.4 \times 2.3 \text{ arcmin}$.

¹ Cygnus A is not characterized by Scaife & Heald (2012), but extensive commissioning work (summarized by McKean et al. 2011 and McKean et al., in preparation) has produced a detailed source model.

2.3 Quality control

A number of bad-quality observations were detected and subsequently flagged using two methods: (i) checking the processed visibilities and (ii) inspecting the final images for each 11-min snapshot. When analysing the visibilities, poor snapshots were flagged when the calibrated visibilities had a mean value greater than the overall mean of the entire four month data set plus one standard deviation value. A slight, or indeed dramatic, rise in the mean of the visibilities does not necessarily imply a completely bad data set: an extremely bright transient (>100 Jy) could have this effect, for example. Such events may have been previously seen from flare stars at low frequencies (Abdul-Aziz et al. 1995), although at shorter time-scales than 11 min (~ 1 s). However, overall, the survey is less sensitive to extremely bright events because of this quality control step. It was beyond the scope of this project to fully investigate this possible effect, and so we decided to only use measurement sets that were deemed to be sufficiently well calibrated.

The results from the automated flagging were also checked against a manual analysis of the visibility plots and the snapshot images, the latter enabling the detection of more bad observations. In total, 460 (out of 2609) snapshots were marked as bad, and were discarded from the search. The large size of the full data set meant that there was no single common reason as to why individual snapshots were rejected, but the problems that caused rejection were mostly due to RFI or ionospheric issues. After the quality control was completed, 2149 observations (394 h) were considered in the analysis.

3 TRANSIENT AND VARIABILITY SEARCH METHOD

3.1 Time-scales searched

As the properties of the target transient population are unknown, the complete data set was split and combined in various ways to fully explore the transient parameter space available. Along with performing a search on the original snapshots, each with an integration time of 11 min, searches were also performed on images with integration times of 30 s, 2 min, 55 min and 297 min. For the longer-duration images, only those 11-min snapshots which were four minutes apart were combined together and imaged. This was to keep the visibilities as continuous as possible in the search for transients. After the quality control step described in Section 2.3, 297 min was the longest continuous integration time possible. All calibration was performed on each individual 11-min snapshot; for the longer time-scales the relevant data sets were combined and then imaged.

3.2 The transients pipeline

The analysis of the data and search for radio transients were performed using software developed by the LOFAR Transients Key Science Project, named the Transients Pipeline (TRAP). It is built to search for transients in the image plane, whilst also storing light curves and variability statistics of all detected sources. Moreover, it is designed to cope with large data sets containing thousands of sources such as this NCP project. A full and detailed overview of the TRAP can be found in Swinbank et al. (2015).² In brief it performs the following steps:

² The work presented in this paper primarily used TRAP release 1.0. However, the data were re-processed once TRAP release 2.0 was available, which is the version described by Swinbank et al. (2015), to confirm results.

Table 2. The average image sensitivity and number of epochs for each time-scale at which a transient search was performed. The accepted epochs column defines how many of the total number of images passed the TRAP image quality control.

Time (min)	Average rms (mJy beam ⁻¹)	Typical resolution (arcmin)	Total no. of epochs	Accepted no. of epochs
0.5	3610	4.8×2.2	47 970	41 340
2	2110	4.7×2.1	10 739	9 262
11	790	5.4×2.3	2 149	1 897
55	550	4.9×2.1	371	328
297	250	3.1×1.4	34	32

(i) Input images are passed through the TRAP quality control which examines two features of the images. First, the rms of the map is compared against the expected theoretical rms of the observation, and if the ratio between the observed and theoretical rms is above a set threshold then the image is flagged as bad. In this case, the threshold was set to the mean ratio value of each time-scale plus one standard deviation. The second test involves checking that the beam is not excessively elliptical by comparing the ratio of the major and minor axes. If this value is over a set threshold then the image is also flagged as bad. All bad images are then rejected and are not analysed by the TRAP (see Rowlinson et al., in preparation for methods of setting these thresholds). The number of images accepted by the TRAP compared to the total entered can be seen in Table 2.

(ii) Sources are extracted using PYSE – a specially developed source extractor for use in the TRAP (Spreeuw 2010, Carbone et al., in preparation). Importantly, all sources are initially extracted as unresolved point sources, which would be expected from a transient event.

(iii) For each image, the source extraction data are analysed to associate each source with previous detections of the same source, such that a light curve is constructed. In cases where no previous source is associated with an extraction, the source is flagged as a potential ‘new source’ and is continually monitored from the detection epoch onwards.

For the source extraction, we define an island threshold, which defines the region in which source fitting is performed, and a detection threshold where only islands with peaks above this value are considered. These island and detection thresholds were set to 5σ and 10σ , respectively. While the use of a 10σ detection threshold may seem very conservative, we agree with the arguments presented by Metzger, Williams & Berger (2015) (hereafter MWB15) who advocate these criteria when identifying a transient source. In their paper, the authors’ main motivation for this high threshold is the significant possibility of spurious signals such as those seen in previous radio transient searches (Gal-Yam et al. 2006; Ofek et al. 2010; Croft et al. 2011; Frail et al. 2012; Aoki et al. 2014), arising from calibration artefacts, residual sidelobes and other similar issues. We share these concerns, in addition to being generally cautious as this survey is one of the first conducted with the new LOFAR telescope. As also stated by MWB15, previous surveys have used 5σ as a detection threshold, which will of course increase the number of potential transient detections; however, this will also yield a high number of false detections, especially with the large number of epochs being used in this survey. Thus, minimizing false detections and obtaining a manageable number of transient candidates were further motivations to use a 10σ detection threshold. We refer the reader to MWB15 for further discussion on this topic.

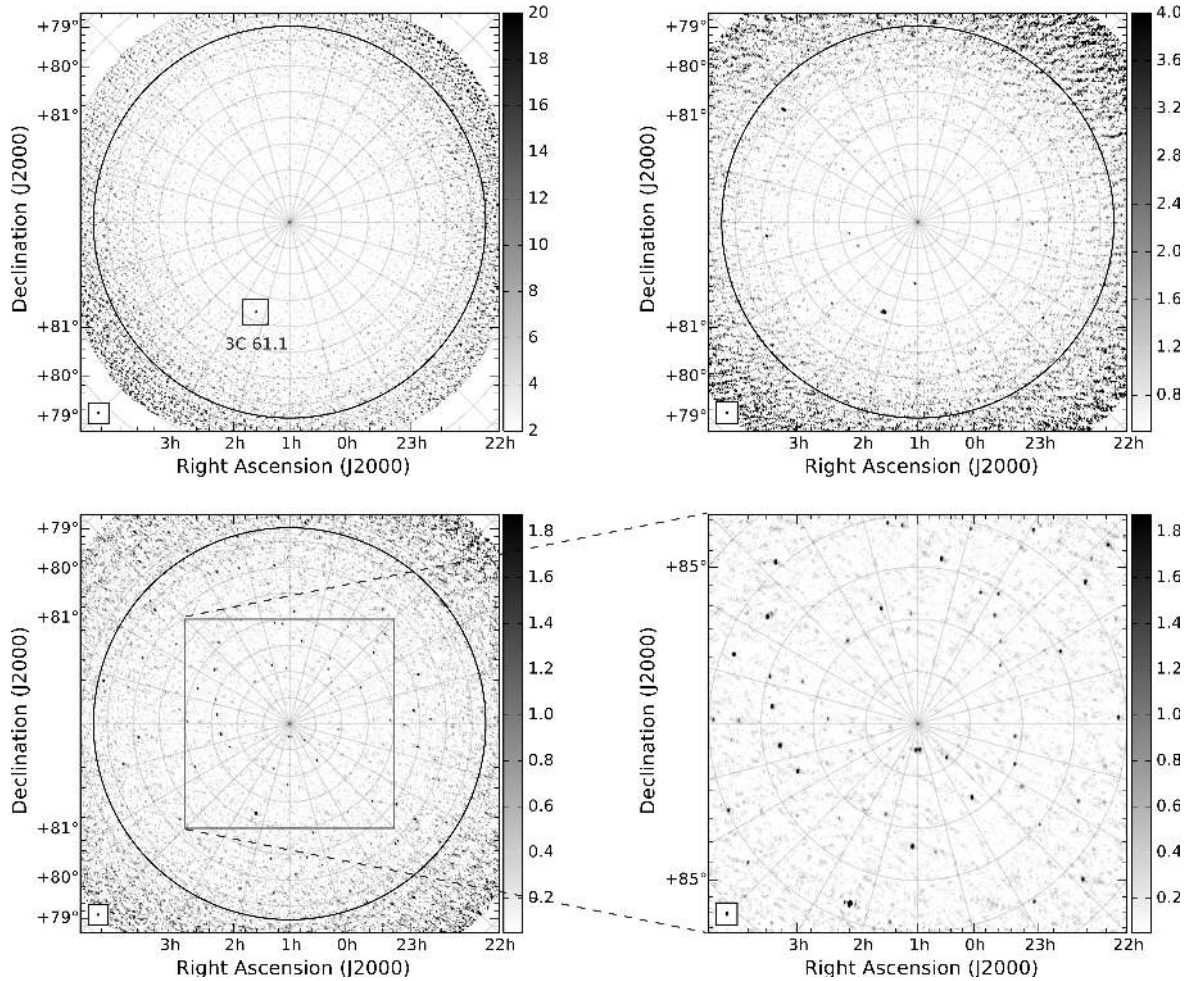


Figure 2. Examples of the NCP field maps at different time-scales. Where present, the area within the black circle indicates the portion of the image searched for transients. This was the same for each time-scale and had a radius of 7.5 . Upper left panel: an image on the 30 s time-scale which was observed on 2012 January 9. Using projected baselines of up to 10 km, the map has a resolution of 4.2×2.3 arcmin (synthesized beam position angle [BPA] -39°) with a noise level of 1.9 Jy beam $^{-1}$. Only the source 3C 61.1 is detected at a 10σ level, and this source is marked on the image. Upper right panel: an 11 min snapshot observed on 2011 December 31. The noise level is 320 mJy beam $^{-1}$ and the resolution is 5.6×3.6 arcmin (BPA 43°). The number of detected sources at a 10σ level is now ~ 15 . Lower left panel: an example of the longest time-scale images available of 297 min, constructed by concatenating and imaging 27, 11-min sequential snapshots. Observed on 2012 February 4, this image has a resolution of 3.5×2.0 arcmin (BPA -6°) and a noise level of 140 mJy beam $^{-1}$, with ~ 50 sources now detected at a 10σ level. Lower right panel: a magnified portion of the lower left panel image. The colour bar units are Jy beam $^{-1}$.

The transient search was also constrained to within a circular area of radius 7.5 from the centre of the image. This was to avoid the outer part of the image which was much noisier and did not have reliable flux calibration.

For each light curve, two values are calculated in order to define whether a source is a likely transient or variable: V_v , a coefficient of variation, and η_v , the significance of the variability (Scheers 2011). V_v is defined as

$$V_v = \frac{s_v}{\bar{I}_v} = \frac{1}{\bar{I}_v} \sqrt{\frac{N}{N-1} (\bar{I}_v^2 - \bar{I}_v^2)}, \quad (1)$$

where s is the unbiased sample flux standard deviation, \bar{I} is the arithmetic mean flux of the sample, and N is the number of flux measurements obtained for a source. The significance value, η_v , is based on reduced χ^2 statistics and indicates how well a source light curve is modelled by a constant value. It is given by

$$\eta_v = \frac{N}{N-1} \left(\frac{\omega \bar{I}_v^2}{\bar{\omega}} - \frac{\bar{\omega} \bar{I}_v^2}{\bar{\omega}} \right), \quad (2)$$

where ω is a weight which is inversely proportional to the error of a given flux measurement ($\omega = 1/\sigma_{I_v}^2$). Throughout this paper we define these parameters as the ‘variability parameters’. For more detailed discussion on these parameters we refer the reader to Scheers (2011) and Swinbank et al. (2015).

To define a transient or variable source, a histogram of each parameter for the sample was created and fitted with a Gaussian in logarithmic space. Any source which exceeds a 3σ threshold on these plots is flagged as a potential candidate. Rowlinson et al. (in preparation) will offer an in-depth discussion on finding transient and variable sources using these methods.

4 RESULTS

4.1 Image quality

Examples of the 30 s, 11 min and 297 min time-scale images can be found in Fig. 2. Note that imaging the NCP can sometimes cause confusion when displaying the right ascension (RA) and declination

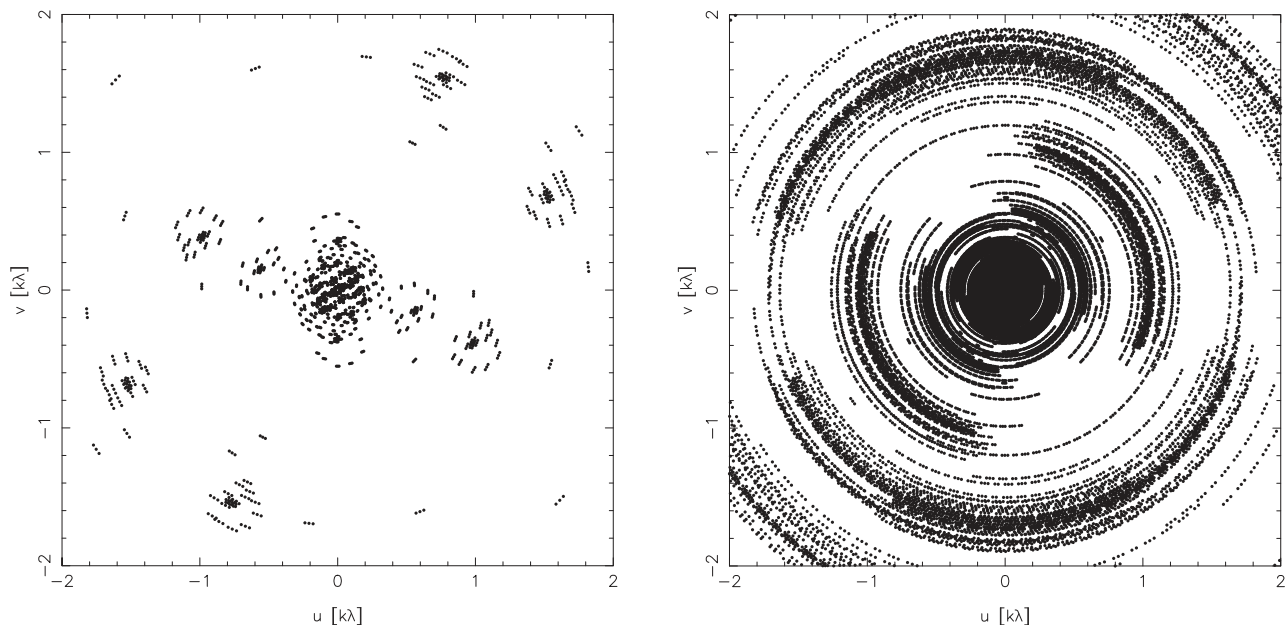


Figure 3. Left panel: the uv coverage obtained with an 11 min snapshot. Right panel: the improved uv coverage gained when combining 27 snapshots (297 min). In each case the uv range is limited to ± 2 $k\lambda$ (10 km).

(Dec) on the image axis, as the grid lines become circular. The grid lines are shown in all figures to help demonstrate this. The obtained uv coverage of the 11 and 297 min observations can be viewed in Fig. 3. The average sensitivity reached with each time-scale is summarized in Table 2, along with the number of epochs available after the quality control described in Sections 2 and 3.

It is important to note that, as a consequence of the PB correction, search areas centred on the NCP do not have a uniform noise level. Larger search areas include noisier regions further from the phase centre, and hence the flux density threshold at which we could detect a transient across the full search area will be higher. Fig. 4 shows an example of a PB map from one of the NCP observations. In order to obtain a noise estimate accounting for the variation caused by the beam, for each image at each time-scale we split the area into four annuli, equally spaced in radius. These four regions are also marked in Fig. 4. The rms for each annulus was then measured, using a clipping technique, with the area-weighted average of these four values providing the single value rms estimate for the individual image. We then took the average of each time-scale, which are used as our sensitivity levels in Table 2. Fig. 5 shows that these measured rms values of the different time-scales approximately follow a $1/\sqrt{t}$ relation, where t represents the integration time of the observation. We note that the longer time-scale rms values appear to lie above the $1/\sqrt{t}$ relation. We believe this is caused by the clipping technique being less accurate at measuring the rms of the longer time-scale images annuli. This in itself is due to the presence of many more sources compared to the relatively source free-short time-scale images. In addition to this, it is possible that the CLEAN algorithm was not applied to a deep enough level in some cases. Hence, the combination of these two methods means that the longer time-scale rms values are likely to be slightly overestimated, but not at a concerning level in the context of this investigation.

We could have limited the transient search to a smaller region with the deepest sensitivity; however, when calculating the figure of merit (FoM; $\propto \Omega s^{-3}$ where Ω is the FoV and s is the sensitivity) it can be shown that it is more beneficial to extend the area of

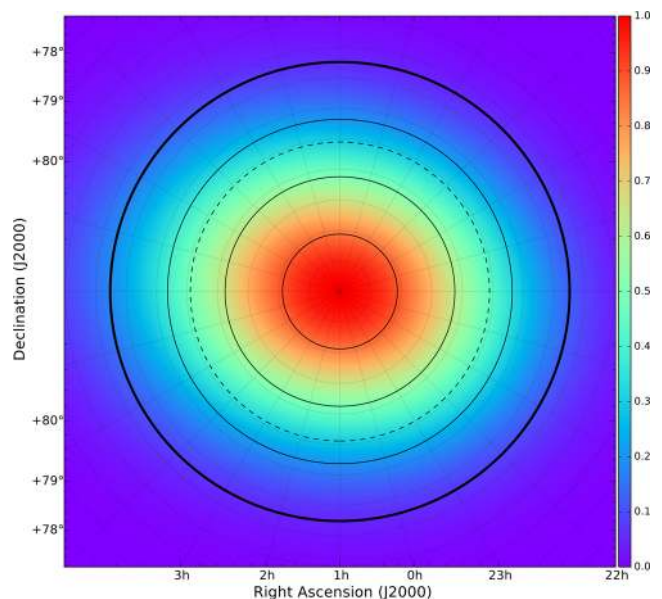


Figure 4. An example of a normalized PB map from one of the NCP observations, which has been scaled to 1.0. The bold, outer solid-line circle represents the full extent of the area for which the transient search was performed (radius of 7.5). The inner solid-line circles show how the area was divided in order to gain an estimate of the average rms for each image accounting for the PB. The dashed-line circle indicates the position of the PB half-power point.

the search, despite the increase in average rms. This can easily be demonstrated as the full area is 16 times larger but the weighted sensitivity only drops by a factor of about 2; hence the FoM is around five times better, illustrating the motivation for searching wide area. We refer the reader to Macquart (2014) for an in-depth discussion of the FoM in the context of transient surveys.

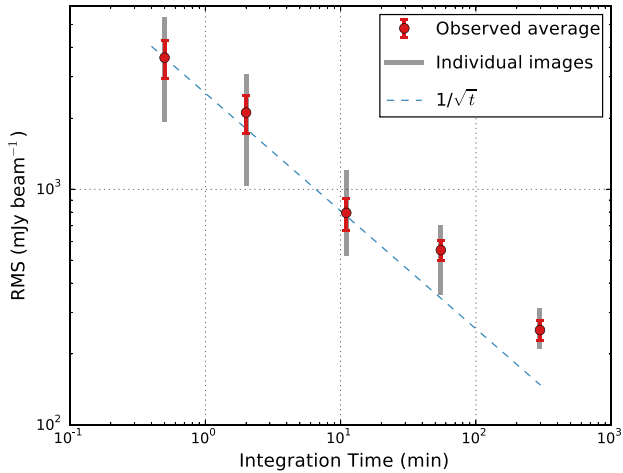


Figure 5. The average rms obtained from the images produced by combining and splitting the data set. Also plotted in light grey are the range of noise values for the individual images at their respective time-scales, in addition to the $1/\sqrt{t}$ relation where t is the integration time of the observation. It can be seen that the average rms values approximately follow this relation; the longer time-scale values are likely to be slightly overestimated due to the methods used to estimate the rms. The errors shown on the average points are one standard deviation of the rms measurements from the respective time-scale.

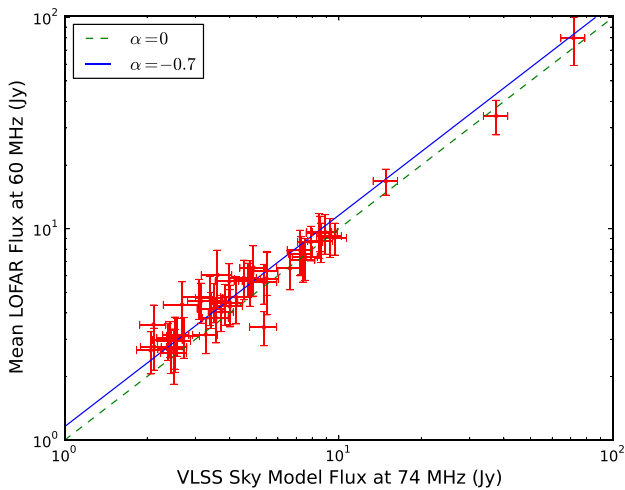


Figure 6. Plot of the mean extracted flux of sources from the 297 min NCP survey at 60 MHz against the cross-matched VLSS survey at 74 MHz. The solid line represents the expected LOFAR flux density assuming a spectral index of $\alpha = -0.7$. For illustrative purposes a dashed-line representing $\alpha = 0$ (a 1:1 ratio) is also shown.

The 55 and 297 min time-scale images offered the best flux calibration stability from image to image due to the better uv coverage achieved on these time-scales. An example of the general flux calibration quality can be seen in Fig. 6, which shows the averaged measured flux across all the 297 min snapshots of sources detected at 60 MHz, cross-matched with the VLSS catalogue at 74 MHz. It shows a general agreement with the fluxes that would be expected assuming an average spectral index of $\alpha = -0.7$. If we assume that all sources have this spectral index and calculate the expected VLSS 60 MHz flux for each source, we find that the average ratio of this expected VLSS flux against the measured LOFAR flux is 1.00 ± 0.17 .

Overall, there was a typical scatter of 10 percent in each light curve of sources detected, which was measured by the TRAP. It was common that fainter sources ($<10\sigma$) would appear to ‘blink’ in and out of images; this was especially apparent in the 11 min snapshots. This was likely due to a mixture of varying rms levels and the ionosphere causing phase calibration issues. Such behaviour was a further reason why a 10σ source detection limit was used in the transient search. The sensitivities of the shortest time-scale maps, 30 s and 2 min, were such that only the brightest source in the field, 3C 61.1, was detectable. The LOFAR and VLSS source positions were also consistent within 5.1 arcsec on average; the typical resolution in the LOFAR band is 3.1×1.4 arcmin for the 297 min time-scale.

It was also important to determine whether the images produced for the transient search are confusion limited. In order to calculate an estimate of the confusion noise for the average resolutions presented in Table 2, we followed the same approach as Heald et al. (2015), using VLSS C-configuration estimates (see Cohen 2004) which we extrapolate to 60 MHz using a typical spectral index of -0.7 . We also alter the formula to account for the non-circular beams:

$$\sigma_{\text{conf, VLSS}} = 29 \left(\frac{\theta_1 \times \theta_2}{1''} \right)^{0.77} \left(\frac{60 \text{ MHz}}{74 \text{ MHz}} \right)^{-0.7} \mu\text{Jy beam}^{-1}, \quad (3)$$

where θ_1 is the synthesized beam size major axis and θ_2 is the minor axis. For the five time-scales used in the transient search shown in Table 2, beginning with 30 s, we calculate the confusion noise estimates to be 113, 107, 128, 111 and 57 mJy beam $^{-1}$, respectively. Thus, due to our simple reduction strategy, our images, at best, are approximately 4 times the confusion noise level and hence would not affect our transient search.

Along with these cadences, a deep map was constructed by using all the available 297 min images, reaching a sensitivity of 71 mJy beam $^{-1}$ (this value was measured using the weighted average method discussed above in this section.). This map can be seen in Fig. 7. This, however, had to be produced by means of image stacking as opposed to direct imaging due to the amount of data involved. A total of 150 sources were detected at a 10σ level within the same 7:5 radius circle used for the transient search, with the map primarily being used as a deep reference image for the field. We can, however, use this deep map to verify our calibration and imaging procedures by comparing our detected source counts to the VLSS. First, using a spectral index of -0.7 , $S_{60} = 710$ mJy corresponds to a flux density at 74 MHz of $S_{74} = 613$ mJy. Using this flux density limit, there are 263 catalogued VLSS sources within 7:5 of the phase centre. Cross-correlating the VLSS with our LOFAR 60 MHz detections, we find that 41 percent of the VLSS sources have a LOFAR match. The factor of ~ 2 discrepancy can be shown to be simply due to the PB attenuation in our deep map. Hence, we were satisfied that the calibration and imaging results were valid and consistent with previous studies, and therefore would not negatively impact any transient searches.

This map was also further analysed for any previously uncatalogued radio sources, but none were found. However, the direct comparison to VLSS revealed that one source, located at $02^{\text{h}}13^{\text{m}}28^{\text{s}}+84^{\circ}04'18''$, has apparently significantly different 60 and 74 MHz flux densities: the VLSS-integrated flux density is 1.49 Jy (possibly put in the error), whereas in the LOFAR band it is detected at the 8σ level with an integrated flux density of 236 mJy. There are no detections of the source in WENSS or NVSS. However, this source is located within a stripe feature in the VLSS image, and the

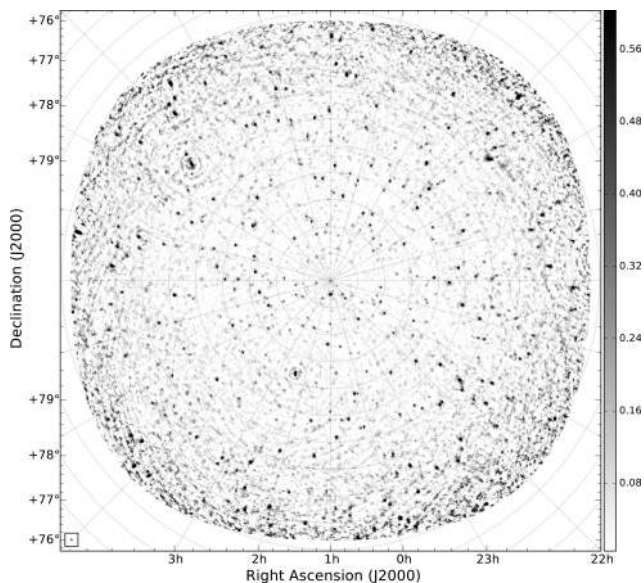


Figure 7. The deepest map produced of the NCP field from the survey. It was constructed by averaging all 31 of the 297-min-duration images together in the image plane, using inverse-variance weighting. It has a noise level of 71 mJy beam^{-1} and a resolution of $3.1 \times 1.4 \text{ arcmin}$ (BPA 42°). A total of 150 sources are detected at a 10σ level within a radius of 7.5 from the centre of the map. While none of these sources are previously undetected, it provided a detailed reference map to check any transient candidates. The colour bar units are Jy beam^{-1} .

source is not present in the VLSS Redux catalogue (VLSS; Lane et al. 2014); hence we do not pursue this source further. The full MSSS survey will offer further insight into this potential source, confirming its flux density and spectral index, if it is real.

4.2 Variability search results

The four month data set provides an opportunity to search for variable sources as well as transient sources. We define variables as sources which are present throughout the entire data set, taking into consideration varying sensitivity, whose light curve displays significant variability over the period. This is opposed to transient sources, which we define as sources that appear or disappear during the time spanned by the data set, again taking into account the varying sensitivity. Consulting historical catalogues also helps with the distinction between variables and transients. Due to the higher level of image quality, the variability search was limited to the two longest time-scales of 55 and 297 min. For each detected source in these two sets of images, variability parameters (V_ν and η_ν) were calculated by the TRAP. Fig. 8 shows the respective distributions of the variability parameters for each time-scale plotted in logarithmic space. In each case, the central panel shows η_ν plotted against V_ν for each detected source. The top panel displays a histogram representing the distribution of η_ν of all the sources along with a fitted Gaussian curve. The right panel contains the distribution and fitted Gaussian curve for V_ν . The dashed lines represent a 3σ threshold for each value; any sources with variability parameters exceeding one or both of these values are considered as potentially variable. Candidates also had to show a variability of significantly more than 10 per cent, which was the calibrator error of the measurements. This was set at a level of 2σ from this value. An ideal transient would appear in the top-right-hand corner of the central panel scatter plot, exceeding the threshold in each parameter.

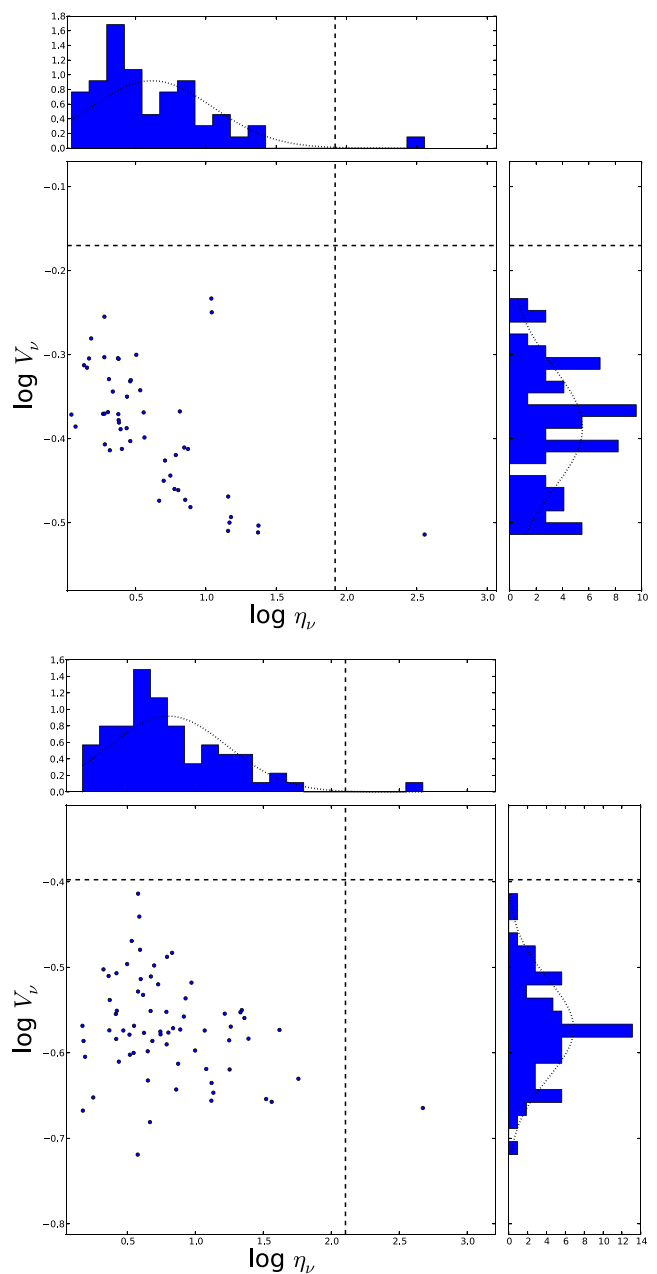


Figure 8. This figure shows the distribution of values obtained for the variability parameters V_ν , a coefficient of variation, and η_ν , the significance of the variability (see text for full definitions) for each light curve detected. The upper panel shows the 55-min image results and the lower panel shows the 297-min time-scale results. In each case, the central panel plots the two values against each other for each source, with the top panel and right side panel displaying the histogram showing the distribution of the η_ν and V_ν values, respectively, for all sources. The dotted lines represent a 3σ threshold for each parameter. A very likely variable or transient source would appear in the top-right of the plot, exceeding a 3σ level in each parameter. At both time-scales, one source (3C 61.1) is found to have a significant value in η_ν . However, this is likely to arise from fluctuations caused by calibration issues.

It can be seen that at both time-scales, no sources exhibit variable behaviour in V_ν above a 3σ level, but one source has a significant η_ν value. This source is 3C 61.1, which dominates the field. While the result points towards low-level variability of 3C 61.1, the source is a well-resolved radio galaxy (Leahy & Perley 1991) whose flux

is dominated by 100-kpc-scale lobes, making it very unlikely that we would detect any intrinsic variability. It is more likely that this is the result of calibration errors and the source extraction and subsequent calculation of η_v itself. The model for 3C 61.1 used during this investigation is quite basic for such a complex source. This, along with ionospheric effects and the general calibration accuracy of the instrument at the time, can have quite a substantial effect on such a bright source, with such calibration errors not included in this analysis. The source is also spatially extended, but the extraction treats it as a point source (as mentioned in Section 3), and this will therefore also have a significant impact on the recorded flux. Removing the point source fitting constraint does indeed move the data point closer back towards the 3σ threshold, but only marginally by 0.1 dex in η_v . As for the η_v value, this parameter is weighted by the flux errors of the source extraction. Bright sources, such as 3C 61.1, are well fitted when they are extracted, which means they have small associated statistical flux errors. This in turn then causes η_v to rise. If we discount 3C 61.1, no sources displayed any significant variability at the 55 and 297 min time-scales.

4.3 Transient search results

Using the TRAP and a manual analysis of its results, searches performed on the time-scales of 0.5, 2, 55 and 297 min found no transient candidates. However, nine transient candidates emerged from the analysis of the 11-min time-scale. At first, it appeared strange to achieve nine candidates at one time-scale but none at any other. However, the sensitivity of the shorter time-scales was such that only bright transients (>25 Jy) would have been confidently detected, and as previously stated no other source, or even artefact, was detected at these flux levels other than 3C 61.1. At the longer time-scales, the improved uv coverage meant that the images improved substantially in quality. This reduced the number of imaging artefacts that could spawn false detections and sources were consistently detected throughout the epochs (as opposed to many sources blinking in and out as discussed in Section 4.1). Any sources that were defined as ‘new’ by the TRAP (these are sources that appeared in later images but were not detected in the first image searched) were in fact association errors and not transient sources.

While the nine candidates could point towards the 11 min images meeting the required sensitivity and time-scale of a transient population, these images are also the most likely to exhibit misleading artefacts due to the limited uv coverage. Hence, the nine reported candidates were subjected to a series of tests to determine whether they were spurious sources. The following tests were performed:

- (i) Subtraction of 3C 61.1 from the visibilities using the clean component model from the deconvolution process. The visibilities were then re-imaged.
- (ii) Applying an extra round of RFI removal using AOFLAGGER.
- (iii) Re-running the automated tool to remove perceived bad LOFAR stations from the observations, followed by a manual check.
- (iv) Imaging the data using different weighting schemes and baseline cutoffs.

The tests were applied in the above order, meaning that if one method definitely succeeded in removing the candidate the latter tests were not performed. Only one of the nine candidates completely survived all the tests; three were inconclusive but quite doubtful, whereas four were definite artefacts. One other source was very marginal in passing all the tests; hence this event is not presented in this paper, but will be discussed in a future publica-

tion. The surviving candidate was thus a potential real astrophysical event and is the subject of the following Section 5.

5 TRANSIENT CANDIDATE ILT J225347+862146

The only candidate to have passed all the validity checks was found in a single 11 min snapshot taken on 2011 December 24 at 04:33 UTC. The source was extracted by the TRAP with a flux of 7.5 Jy (14σ detection in individual image), at coordinates $22^{\text{h}}53^{\text{m}}47^{\text{s}}.1+86^{\circ}21'46''.4$, with a positional error of 11 arcsec. It was only seen in this one snapshot with no detection of the source in the preceding or subsequent snapshots. The observation can be seen in Fig. 9. Nothing was present at the candidate position in either the relatively deep image constructed from the longer time-scale images (see Section 4.1) or the very deep image of the field from the LOFAR Epoch of Reionization (EoR) group (Yatawatta et al. 2013). Note that the EoR project uses the LOFAR high-band antennas, and hence it is at a higher frequency range of 115–163 MHz.

5.1 A mirrored ghost source

On closer inspection, the transient candidate appeared to have a secondary-associated positive ‘ghost’ source mirrored across the brightest source in the field, 3C 61.1 (the transient lies at an angular distance of $3^{\circ}2'$ from 3C 61.1.), which can also be seen in Fig. 9. This ghost was not detected by TRAP due to the higher rms value in that region, and like the transient candidate it was a ‘new’ source with no previous or subsequent detections. In fact, the ghost source was actually nominally brighter than the transient source with a flux density of 13 Jy. However, in the non-PB-corrected map the candidate has a higher peak flux density (9 Jy) than the ghost (6 Jy). This was not the first time we had witnessed this type of effect in LOFAR observations, with previous commissioning data we had obtained in 2010 showing a similar situation. Currently, the exact explanation of why ghosts of this nature, including specifically the ghost presented in this work, are generated in LOFAR data is unknown. It should be noted that none of the other eight transient candidates detailed previously had an associated ghost source. In the following discussions we refer to the original detected transient source ILT J225347+862146, to the west of 3C 61.1, as the ‘transient candidate’ and the source to the east of 3C 61.1 as the ‘ghost’ source (refer to Fig. 9).

5.1.1 Ghost artefacts in radio interferometry

Calibration artefacts presenting themselves as spurious ‘ghost’ sources are not an entirely new topic to radio interferometry. The topic of ‘spurious symmetrization’ is discussed in Cornwell & Fomalont (1999); in brief, if a point source model is used for a slightly resolved source, a single iteration of self-calibration can result in features of the image being reflected relative to the point-like object. However, this can be corrected with further iterations of self-calibration which would cause the spurious features to disappear. As will be discussed in Section 5.1.2, the ghost presented in this work can be seen before initiating any kind of self-calibration of the target field, i.e. any calibration using a target field sky model. Therefore, it is highly unlikely that the spurious symmetrization previously described is the sole cause of the ghost. However, this is not to say that the effect plays no role in its creation.

More recently, Grobler et al. (2014) (hereafter ‘G14’) began a series of investigations dedicated to ghost phenomena. This first study

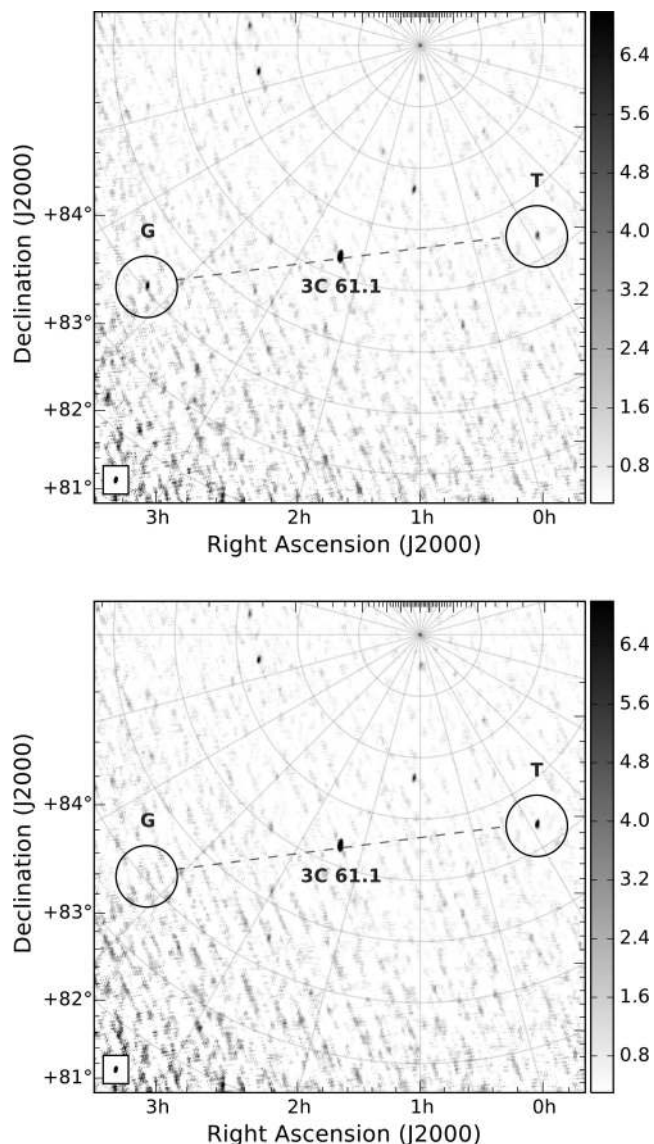


Figure 9. Upper panel: illustration of how the transient source (labelled ‘T’), ILT J225347+862146, was originally detected in the image, along with the associated ghost source (labelled ‘G’) across from 3C 61.1. Lower panel: now the measurement set as been re-calibrated with the transient included in the sky model; the ghost source has vanished. Upon closer inspection, other faint, source-like features also disappear from the re-calibrated image. These are most likely fainter ghost features which are reduced when the data were calibrated with a more complete sky model. The colour bar units are Jy beam^{-1} .

concentrated on ghosts seen in data from the Westerbork Synthesis Radio Telescope (WSRT). In these data, ghost sources appeared as strings of (usually) negative point sources passing through the dominant source(s) in the field. The arrangement of these negative point sources appeared quite regular, along with the fact that the positions were not affected by frequency. In their investigation, G14 were successful in deriving a theoretical framework to predict the appearance of ghosts in WSRT data for a two-source scenario, and were able to confirm what previous work had suggested concerning these ghost sources (see text in G14).

In brief, the main features about ghosts to note are as follows: (i) they are associated with incomplete sky models, for example missing or incorrect flux; (ii) in the WSRT case, the ghosts always

formed in a line passing through the poorly modelled or unmodelled source(s) and the dominant source(s) in the field; (iii) the ghosts are mostly negative in flux, while positive ghosts are rare and weaker; and (iv) the general ghost mechanism can also explain the observed flux suppression of unmodelled sources.

G14 also concluded that the simple East–West geometry of the WSRT array is the reason for ghosts appearing in a regular, straight line, pattern. This becomes more complex when a fully 2D/3D array is considered such as LOFAR, where the ghost pattern is expected to become a lot more scattered and noise-like. This subject will be the focus of Paper II (Wijnholds et al., in preparation) in the series on ghost sources. However, G14 did note that regardless of the array geometry, ghosts are expected to occur at the $n\phi_0$ positions, where ϕ_0 represents the angular separation between the respective bright source and unmodelled source, and n is an integer number. Usually the strongest ghost responses are the $n = 0$ and $n = 1$ positions, i.e. the suppression ghosts that sit on top of the sources in question. However the case discovered in this work, and also two independent cases (de Bruyn, private communication; Clarke, private communication) in LOFAR data suggest that the $n = -1$ position could also generate a strong response. What is significant about the transient presented in this work, however, is that the ghost appears as a positive source.

5.1.2 Investigating the NCP ghost

Returning to the situation detailed in this paper, we were presented with two sources for which either could be the real (transient) source or the ghost. We attempted to simulate the situation within real data, in order to investigate how the different stages of calibration would react to a bright transient, and if we could also generate a positive ghost source. This was done by taking a different NCP observation and inserting a simulated transient source into the visibilities (the transient was set to be ‘on’ for the entire 11 min.) before any calibration had taken place. The snapshot was then calibrated as normal, but importantly the inserted source was *not* included in the NCP sky model used for the phase-only calibration step (refer to Section 2.2). This test was repeated using various different sky positions and flux densities for the inserted source. We found that we could produce a significant positive ghost source only if the flux of the simulated transient was relatively bright, ~ 40 Jy. An example can be seen in Fig. 10. We observed that it was common for the total flux to be shared approximately equally between the simulated source and its associated ghost. However, not every position on the sky at which the transient was inserted produced a ghost source, a feature that we cannot currently explain. Yet, when a transient was inserted at the position of ILT J225347+862146, this did produce a ghost source. We were then able to test what happened when the simulated source was included in the sky model. We observed that when the simulated source was accounted for perfectly in the sky model, the ghost source disappeared. If the sky model component was instead inserted at the location of the ghost source, while the ghost appeared brighter, the simulated transient never fully disappeared.

In light of the results from the simulations, we performed the same sky model test with the transient candidate and ghost in order to determine which source was the ‘real’ source. Recalling that the total flux of the transient candidate and ghost was ~ 7 Jy + ~ 13 Jy ≈ 20 Jy, we began by inserting a 20 Jy point source into the NCP sky model at the position of the transient candidate and re-calibrated the data set. We found that in this case the flux of the ghost was significantly reduced, by ~ 70 per cent, and the candidate