Transient and variable radio sources in the LOFAR sky: an architecture for a detection framework

Scheers, L.H.A.

Publication date
2011

Citation for published version (APA):
Scheers, L. H. A. (2011). *Transient and variable radio sources in the LOFAR sky: an architecture for a detection framework*. [Thesis, fully internal, Universiteit van Amsterdam].

Chapter 1

Introduction

The excitement of new instruments
is not the old questions they will answer
but the new questions they will raise
Ron Ekers

1.1 Radio Astronomy in Historical Perspective

Karl Jansky opened the brand new field of radio astronomy in January 1932 (Kraus, 2005) by recognising that the 20.5 MHz signals received by his rotating antenna array were of extraterrestial origin. At that time he was appointed at the Bell Telephone Laboratories and given the task to study the nature of the disturbances influencing transoceanic telephone traffic. Minimising antenna responses to certain directions would then increase the signal to noise and improve the connections. More observations made Jansky (1933) realise that the signals came from a fixed location in space, the centre of our own Galaxy. The Milky Way, for the first time, showed up in radio frequencies.

The astronomical community was ignorant of the relatively new radio technology (Verschuur, 2007) and only few others were aware of the significance of the discoveries. Jansky was given no support for a proposed construction of a parabolic antenna to continue his work, after which he was not involved in radio astronomy any more. Jansky's work attracted Grote Reber, an electronics engineer, who was the first to build a parabolic-reflector antenna, a prototype of the classical dish-like radio telescopes, trying to understand more about the antenna–receiver system. Reber (1940) reported preliminary results of concentrated radiation at 160 MHz from the Galactic Plane and in 1944 he was also the first to carry out a systematic sky survey, producing the first maps of the radio sky (Reber, 1944).

During World War II, Reber's work reached the Netherlands where it was picked up by Oort, who realised that in combination with Jansky's discoveries, the radiation was merely a continuum. Seeking for line radiation, Van de Hulst (1945) concluded that a small energy transition related to the electron-spin flip in the ground state of neutral hydrogen was observable at 21.1 cm, or 1420 MHz. A few years later, searches from several sites around the world confirmed the transition in emission (e.g. Ewen & Purcell, 1951; Muller & Oort, 1951; Christiansen & Hindman, 1952), as well as in absorption (Hagen & McClain, 1954), leading to detailed maps of our Galaxy.

Further breakthroughs were lying ahead. Observations at high Galactic latitude, in apparently empty fields, carried out with the horn-reflector antenna at 4080 MHz by Penzias & Wilson (1965) measured the intensities which led to the discovery of the 3 K cosmic microwave background (CMB) radiation, of which predictions were made by Gamow (1946), Alpher, Bethe & Gamow (1948) and Dicke et al. (1965) and pointed out the validity of the Big-Bang theory. Hewish et al. (1968) serendipitously discovered radio pulsars soon after, highly magnetised, rapidly rotating neutron stars. Their instrument, consisting of 2048 dipole antennas sensitive to 80 MHz frequencies, was designed to measure the angular size of radio sources by the rapid intensity fluctuations due to scintillation in the solar corona, but it became clear that it was ideal for detecting pulsars and other detections followed.

Technological development and new design methods were able to improve the observations. An array of multiple telescopes, an interferometer, can synthesise high-resolution images by making use of the aperture synthesis method developed by Ryle & Hewish (1960). Since resolution is inversely proportional to the aperture diameter, now being the diameter of the whole array, larger arrays will have higher resolution. When all telescopes track a source, signals arrive later at those telescopes further away in the array. Correcting for this delay and then adding the signals from all the telescope-pair combinations, will amplify the signal. Furthermore, as seen from a (radio) source point of view, Earth's rotation varies the distance between telescope pairs (i.e. baselines) continuously, producing a series of measurements – the visibilities or (u, v) data – as function of the hour angle. The visibility data, however, are not in a comfortable format to visualise the sky and make comparison at other wavelengths, and therefore it is converted into image data. In creating a proper image the visibilities need to be calibrated first by removing the instrumental and atmospheric effects from the measurements. Imaging is then done by Fourier transforming the sampled calibrated visibility function, which will give the relationship between the known primary beam pattern and the unknown sky brightness (Briggs et al., 1999).

Some well-known radio interferometers take advantage of the above-mentioned techniques; they were constructed some decades ago and are still operational and highly successful. Examples are the Westerbork Synthesis Radio Telescope (WSRT) in the Netherlands (since 1970), the Very Large Array (VLA) in New Mexico, USA (1980), the Giant Metre-wave Radio Telescope (GMRT) in Pune, India (1988) and the Australia Telescope Compact Array (ATCA) in Narrabri, Australia (1988).

Until the end of the previous century, this was the proven technique of making radio observations. With the increasing power of computing resources becoming available, it became feasible to start designing arrays of less costly telescope elements, and digitising the signals at the earliest stages. The achievable higher sensitivities and spectral, spatial and temporal resolutions, while relying on embedded electronic engineering and software techniques, make astronomy shift into a new paradigm.

Very good historic reviews are given in, e.g., Kraus (2005), Thompson, Moran & Swenson (2004), and Verschuur (2007).

1.2 LOFAR: The Next-Generation Radio Telescope

The LOw-Frequency ARray, or LOFAR, is scientifically as well as technologically a next-generation radio telescope. It is sensitive in a not yet well-explored domain of the electromagnetic spectrum, the low frequencies at 30 – 240 MHz. Relatively simple antennas (see Fig. 1.1) are connected to large computer facilities for further signal processing such as calibration and imaging. In combination with the unprecedented spatial and temporal resolution, sensitivity and large field of view, this makes LOFAR an excellent instrument that is very flexible to carry out observations of many different kinds. Below we will briefly describe the telescope and the radio sources and phenomena that will be studied by the Transients Key Science Project.

Figure 1.1: Left: The Low Band Antennas operate at 30 – 80 MHz (picture T. Krieg). In the computer cabinet, at the top left, all the antenna signals in the field are combined, before the data are transported to the central processing units in Groningen. Right: The High Band Antennas operate at 120 – 240 MHz.

1.2.1 The International LOFAR Telescope

The new radio telescope LOFAR is technically of fundamentally different design. Numerous, *ordinary*, dual-dipole antennas are mounted on the ground and grouped into stations. The antenna elements do not have any steerable parts like dish telescopes, but can be electronically directed by how their signals are combined to specific sky locations. Two different types of antenna design are operational, at frequencies below and above the radio broadcasting FM band, repsectively. The former being the Low Band Antennas (LBAs) sensitive at 30 – 80 MHz and the latter the High Band Antennas (HBAs) performing at 120 – 240 MHz. Fig. 1.1 shows the two types of antennas.

The stations are distributed over large areas. A dense core of 24 stations is grouped into an area of 2 – 3 km across, of which 6 stations are located on the circular superterp, having a diameter of 340 m; this area is near the town of Exloo in the province of Drenthe at geographical coordinates $52^\circ\,54'\,32''$ N, $6^\circ\,52'\,8''$ E. Fig. 1.2 shows the lay out of a core station and the superterp.

Remote stations are spread over the northern part of the Netherlands and have baselines up to $\sim$ 100 km, whereas the international European stations have baselines as large as $\sim$ 1000 km. Fig. 1.3 is a representation of the sites of funded and planned LOFAR stations throughout Europe. Currently funded international stations are located in Germany (Effelsberg, Garching, Jülich, Potsdam, Tautenburg), France (Nançay), England (Chilbolton) and Sweden (Onsala).

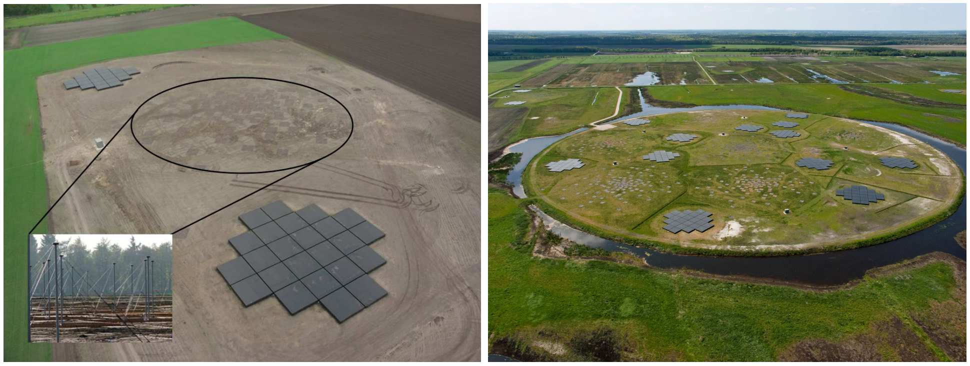

Figure 1.2: Left: Aerial photo of a LOFAR core station on the superterp near Exloo, the Netherlands. A core station consists of two HBA fields made up of 24 tiles, a tile being an assembly of 16 HB Antennas arranged on a 4×4 grid, and one LBA field with 96 LB Antennas (see inset). Right: An aerial photo of the LOFAR superterp with diameter of 340 m and its 6 core stations. The superterp is located in the heart of the LOFAR core area, having a diameter of 2 – 3 km and containing 22 core stations. Photo credits: ASTRON.

Figure 1.3: Artist's impression of the sites of LOFAR stations throughout Europe (Spektrum der Wissenschaft/Emde-Grafik).

Signals from the antenna fields are digitised and combined locally at a station level, before they are transported over a dedicated 10 Gb/s ethernet link to the central processing supercomputer in Groningen, that correlates the data coming from all the stations. Then, on the post-processing computing cluster, further operations run on the data streams (still at about 2 Gb/s) in the automated calibration and imaging software pipelines (Wijnholds, 2010). At this point the LOFAR Key Science Projects hook in their software pipelines for dedicated processing.

1.2.2 The Transients Key Science Project

The development of LOFAR is largely driven by the Key Science Projects (KSPs), that will exploit LOFAR to meet their science goals. The Transients Key Science Project aims to search for, detect, and study all the transient events in the LOFAR radio sky. Therefore, the Transients Key Project (TKP) will be using LOFAR intensively to survey the whole sky on a daily basis, monitor large areas, and observe single sources. Fast hemispherical surveys for detecting transients at the earliest stage and full-array modes for observing and tracking targeted sources in detail. All observed sources and flux measurements will be stored in a light-curve database, gradually building up the largest radio catalogue. Internally it will be used for classifying sources and light curves and it will become available to the astronomical community as soon as possible. Sending and receiving triggers for follow-up strategies with multi-messenger instruments via the Virtual Observatory Event (VOEvent) network are set up for further scientific analysis. For transient sources, any follow-up with other telescopes or particle detectors need to be fast; therefore, LOFAR has been set up to send and receive alerts of new events via the so-called VOEvent network (e.g., Seaman et al., 2006; Swinbank, 2009).

One of the key observational modes of LOFAR for the TKP will be the Radio Sky Monitor (RSM; Fender et al., 2007). This mode will look at large patches of the radio sky for temporal changes in source brightnesses, by comparing current with previous observations, separated logarithmically in time. Multiple beams can be pointed simultaneously from the Core Stations to tile out a large fraction of the sky, while separate beams from the Remote Stations might be formed to observe simultaneously in targeted mode as is shown in Fig. 1.4. Using multiple $(100-600)$ pointings, a Rapid All-Sky Monitor is able to survey the visible sky once a day, going as deep as $2-180\,\mathrm{mJy}$ with resolutions of $3-7\,\mathrm{arcmin}$ at the HB and LB frequencies, respectively. Another transient survey mode is the Zenith Monitor (ZM), when again a pattern of beams tiles out a large patch of the sky instantaneously. Using Earth's rotation and keeping the antennas pointing at the zenith, stable beams survey a large strip of declination once a day, including parts of the Galactic Plane, and reaching sensitivities of $0.7-8\,\mathrm{mJy}$ and resolutions of $3-7\,\mathrm{arcmin}$ for HB and LB frequencies, respectively. New transients will be picked up quickly in these modes, while swift full-array follow-up observations will reveal more details about the evolution of a detected event.

Of particular interest to the TKP is the class of jet sources, which show radio emission due to synchrotron radiation produced by magnetised relativistic particle outflows interacting with the environment. Many types of objects are known to belong to this class, among which are the black hole and neutron star X-ray binaries (Fender, Belloni & Gallo, 2004), gamma-ray bursts (GRBs; e.g. van Paradijs, Kouveliotou & Wijers, 2000; Mészáros, 2006) and their radio afterglow phases detectable at LOFAR frequencies (van der Horst et al., 2008). Jets are manifested at the largest, galactic, scales as well, in Active Galactic Nuclei (AGNs) where they emanate from the central supermassive black hole. Observing different types of jet sources offers unique insights into the physics that plays a role in the evolution of these jets.

Other expected transient events to be caught by LOFAR are flares from soft gamma-ray repeaters (SGRs), a class of slowly rotating neutron stars

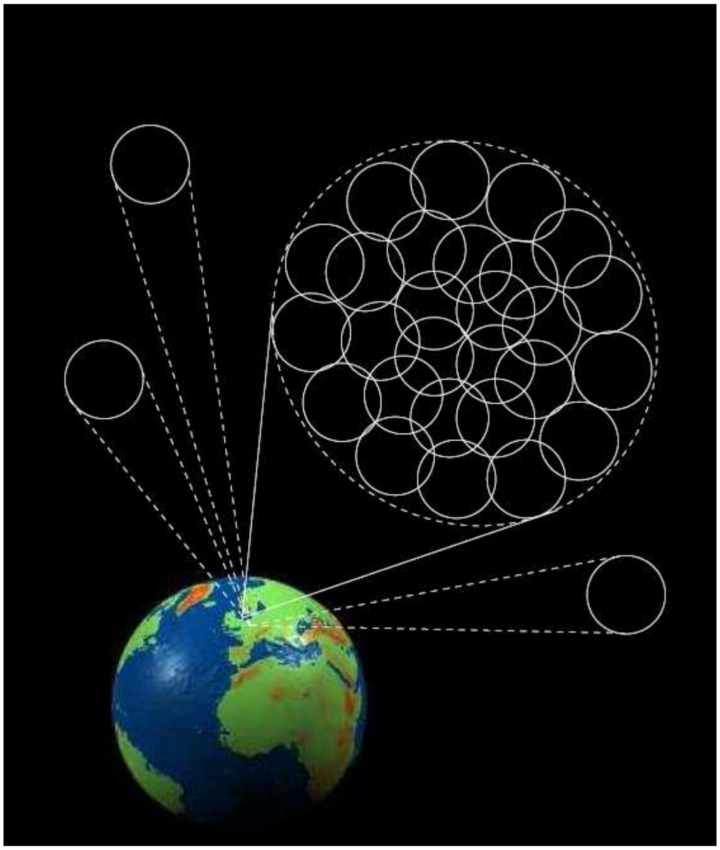

Figure 1.4: Artist's impression of the Radio Sky Monitor (RSM) mode of LOFAR. Multiple Core Station beams operate together to observe a large patch of the sky instantaneously. Single beams observe for targeted sources. The RSM will be one of the envisaged strategies to find transient events at an early stage.

with high magnetic fields (Gaensler et al., 2005a) and polarised radiation from the radio nebulae produced by these explosions (e.g. Taylor et al., 2005; Spreeuw, Scheers & Wijers, 2010). In addition, the TKP will use LOFAR in its full-array mode in order to carry out systematic deep pulsar surveys that are expected to find many new pulsars and to deliver the true pulsar population in the nearby Galaxy (van Leeuwen & Stappers, 2010). Radio emission from pulsars has a steep spectrum ($S_\nu \propto \nu^\alpha; \alpha \sim -2$) in frequencies above the LOFAR High Band (see e.g., Burke & Graham-Smith, 2002), whereas a turnover at frequencies below 100 MHz might or might not be seen (e.g. Navarro et al., 1995), right in the LOFAR Bands. Pulsar models will be improved by the large sample of pulsars for which the measured spectral indices of the pulse components and continuum radiation and the turn-over break frequencies are important parameters. Extragalactic pulsar surveys are planned as well. Furthermore, LOFAR will visit some known extrasolar planets that are good candidates for predicted low-frequency detections (Grießmeier, Zarka & Spreeuw, 2007).

A primary goal of the Transients Key Project is to fill in the gaps at the relatively unexplored lower end of the radio spectrum for the sources detected by LOFAR. Dedicated wide-field monitors, commensally processing all LOFAR data taken, and targeted source observations with the full array will then provide us with extended light-curve information which will be permanently stored in a database that is accessible and queryable by the astronomical community. Fig. 1.5 displays spectra of some known classes of radio sources, showing that sources of synchrotron emission with or without turn-over frequencies due to synchrotron self-absorption are best observable in the LOFAR Bands.

The large number of classes of sources with hitherto covered behaviour in the low-frequency and time domains makes the quest for a new low-

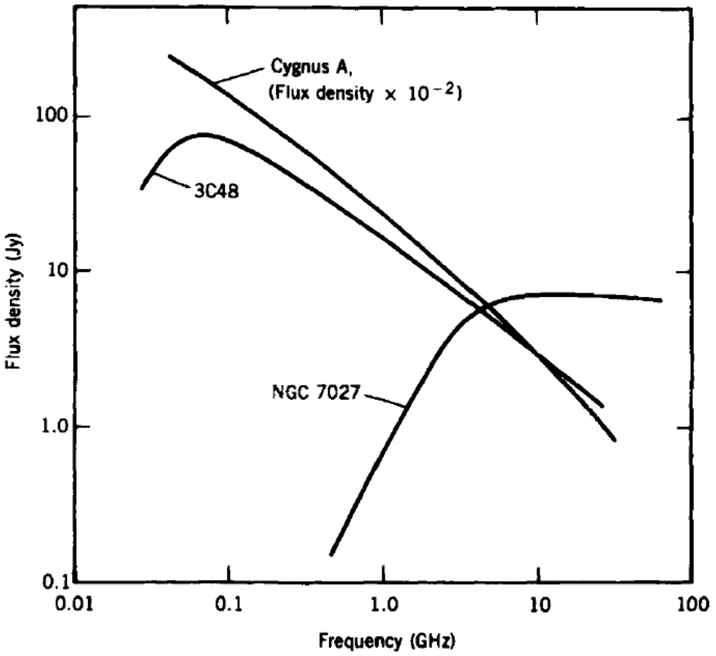

Figure 1.5: Spectra of known classes of synchrotron sources and a thermal source (NGC 7027). Synchrotron radiation is produced by relativistic electrons spiraling in magnetised fields in the radio galaxy Cygnus A and the quasar 3C48, where synchrotron self-absorption in the latter source causes a turnover in its spectrum. NGC 7027 is a planetary nebula, ionised by the radiation of the central star nebula within our Galaxy. The unbound electron–ion (free–free) collisions produce the radio emission, opaque at the lower frequencies (making it invisible in the LOFAR Bands), following the Rayleigh–Jeans law and becoming transparent at the higher frequencies, giving rise to a flat spectrum. Figure adopted from Thompson, Moran & Swenson (2004).

frequency telescope clear. As a software telescope, LOFAR is capable of exploring the new parameter spaces at a wide range of frequencies, timescales, resolutions and sensitivities. The quick reconfiguration of the instrument in flexible modes will optimise observational strategies for detecting events at the earliest stages. Based on experiences with previous new telescopes (see Section 1.1) new types of sources or phenomena are very likely to be discovered, and exploration by LOFAR of the new spectral window on short as well as long timescales, the new parameter space, will almost certainly lead to serendipitous discoveries.

1.3 Expectations of LOFAR

1.3.1 Source Counts and Data Rates

The Transients Key Project inspects the LOFAR data streams of calibrated images for transient sources by means of an automated software pipeline that runs in a Python-based framework (Swinbank et al., 2007). Source fitting routines extract sources and related properties from the images (see Spreeuw, 2010), after which the results are stored into a database. It needs to be emphasised that only extracted data from the images will be stored in the TKP database and not the raw image data. Inside the database, the results are then compared with known, catalogued, source properties. Significant differences are reported back, and triggers for further specific follow-up actions may be sent. Using a database is necessary to have fast access to the large archive of source properties and light-curve data that the TKP gradually builds up.

To optimise the TKP pipeline, we need to think ahead about the specific observation modes, the number of distinct sources likely to be detected, the number of times a source will be observed, and at which pace the data input is streaming. These expected numbers determine the amount of disk space that has to be allocated and the data growth over time. Related to these are the accessibility of the database, containing millions of measurements of millions of sources. Also the response times of the transient detection algorithms should be quick enough to keep pace with the shortest integration times. It was one of the tasks of this thesis to investigate these restrictions and analyse detection algorithms and their applicability in the TKP software pipeline.

Most sources out of the Galactic Plane that LOFAR will detect, are of extragalactic origin and obey, on average, the flux-density–frequency scaling law: $S_\nu \propto \nu^{-0.7}$. Scaling the LOFAR sensitivities to existing source-count models at non-LOFAR frequencies (1.4 GHz, Huynh et al., 2005) will give us predictions on the numbers of sources likely to be detected in certain observation modes.

Observation Mode	N $[\times 10^6]$	m $[\mathrm{s}^{-1}]$	d [GB/day]
Zenith Monitor	0.6	2400	62
Rapid All-Sky Monitor	1.1	350	9.1
Full-NL	9.2	4400	114
MSSS (commensal)	3.4	1800	47

Table 1.1: Expected numbers for some observation modes of LOFAR. We assume a 5σ detection threshold in these modes and a 100% allocation time per mode. N is the number of distinct sources likely to be detected, m is the averaged number of measurements to be stored every second, where a (source) measurement is about 300 Bytes. d is the growth of the TKP database disk size per day assuming a 100% time allocation. The RSM Zenith Monitor and Rapid All-Sky Monitor modes and the Full-NL mode will be frequently used by the TKP, whereas the Million Sources Sky Survey (MSSS) is an example of commensally processing LOFAR data.

Some observation modes of LOFAR that will be frequently used by the TKP were briefly mentioned in Section 1.2.2 and are listed in Table 1.1 with their expected number of sources, measurements per second, and data growth per day. From Table 1.1 it can be seen that the highest peaks of the data rate are to be expected from the Full-NL mode.

1.3.2 Finding Transients with LOFAR

Successful detection of transient events and variable sources depends on the monitoring strategies as well as on the software algorithms applied to the data streams and it was one of the tasks of this thesis to find out which database system and algorithms fit best to these requirements.

Positions of all the sources detected by LOFAR will be maintained in a database. Known sources from existing catalogues will be loaded as well for positional cross matching. Flux measurements and related properties (e.g. 1σ errors, timestamp, frequency, resolution) will be added to the database every time an image has been processed. This means that flux measurements at a given location on the sky might have an association with a known source position in the database. A match will add the measurement to the light curve of the source, whereas no match at all needs further investigation before a reliable alert is sent out.

1.4 This Thesis

One of the main scientific goals of this thesis was related to low-frequency behaviour of variable sources in the LOFAR radio sky by reducing existing interferometry data of multiple observations of the decaying intensities from gamma-ray burst afterglows and soft gamma-ray repeaters. LOFAR will outscale the classical observing techniques by making many measurements per source on timescales as short as 1 second. Data will eventually be processed, without manual intervention, in an automated software pipeline optimised to detect transient events and variable sources. This served as the framework for the other scientific software goals: design of a transient database schema that is capable of storing all the source measurements at the high rate at which LOFAR observes. Design and implement fast search and detection algorithms with response times comparable to the image input rate. *Chapters 2, 3 & 4* are related to the software development of the Transients Key Project and *Chapters 5 & 6* to types of variable sources that are detectable at LOFAR frequencies.

Chapter 2 describes the main observation modes that are relevant to the TKP. Expected numbers for source counts are worked out down to the one-second timescale and up to the envisaged maximum integration times per mode. The TKP will try to commensally process all LOFAR data and for this the Million Sources Sky Survey (MSSS) is discussed as an example. The schema of the database, the main tables and algorithms are elaborated. Benchmarking the speed of the main algorithms concerning source association and transient and variability detection showed a clear preference for the column-store oriented database MonetDB to be implemented in the TKP pipeline.

The source association and transient and variability detection algorithms are described in *Chapter 3*. Three parameters – distance, weighted positional difference and likelihood ratio – are determined for every pair of sources that is considered as a source association candidate. A reliable and fast cut-off association parameter of two genuinely associated sources is the weighted positional difference, which has a Rayleigh distribution. Furthermore, two variability indices define the absolute flux change and the weighted flux change, i.e., the significance of any variability. This significance gives a quantitative measure for the likelihood of having seen a truly variable event, and thus can be the basis for deciding whether to initiate any follow-up of the source (along with measured source characteristics).

We applied the algorithms to simulated and real data. For the simulated data we used a sample of 1000 VLA 325 MHz noise maps and inserted 64 fixed source positions in each map, meaning that a source position is Rayleigh distributed due to the local noise. The 1000 images were processed to assess the critical values for genuine source associations. These were applied to WSRT observations of the field of GRB 030329 made during the years 2003 and 2007 at several frequencies, resolutions, and sensitivities to search for transient and/or variable sources. We found two variable sources.

Chapter 4 describes the usage of multiple major radio catalogues in the TKP database. These catalogues were produced by surveys that were carried out at different frequencies and resolutions. LOFAR will create a catalogue as well. The cross correlation of catalogues is based on the same principles as the association of sources as described in *Chapters 2 & 3*. We derive source lists that include fitted spectral indices, which are available to the calibration processes of LOFAR and the TKP transient detection pipeline. We cross-identified the sources in the VLSS (74 MHz), WENSS (325 MHz) and NVSS (1400 MHz) catalogues and present a high-quality source list of nearly 2800 bright sources (> 1 Jy) at declinations between 44° and 64°, that have unique counterparts in all three catalogues.

Chapter 5 reports on the Galactic Centre Radio Transient source, GCRT J1745–3009. It was discovered at 325 MHz by Hyman et al. (2005), showing five Jy-level bursts of 10 min duration and with 77 min recurrence time. Follow-up observation campaigns showed only two redetections of faint bursts (Hyman et al., 2006, 2007). Reanalysis of the VLA discovery data set improved the burst recurrence times and spectral indices, making clear that such bursts are detectable by LOFAR.

Results of early radio follow-up observations of the giant flare produced by the soft gamma-ray repeater SGR 1806-20 are described in *Chapter 6*. Low-frequency polarimetry and total intensity measurements of the radio nebula revealed the absence of depolarisation at these frequencies and probably probed different substructures in the nebula as compared to higher frequencies. Again, the results show that LOFAR in full array mode would have been able to detect this burst and follow the evolution of the radio nebula and the magnetic fields.