



## UvA-DARE (Digital Academic Repository)

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

Scheers, L.H.A.

**Publication date**  
2011

[Link to publication](#)

#### **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].

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

# LOFAR's Transients and Variability Detection Algorithms

---

B. Scheers, H. Spreeuw, A. Kamble, A.J. van der Horst, E. Rol,  
R.A.M.J. Wijers & the LOFAR Transients Key Science Project  
**to be submitted**

### Abstract

LOFAR, the LOw Frequency ARray, will produce vast amounts of radio data that will be inspected by the Transients Key Science Project to search for and study transient and variable sources. An automated software pipeline that processes the data streams of calibrated images of 2 Gb/s is under development, with the aim to identify, classify and monitor astronomical transients in nearly real-time. An implemented light-curve database keeps track of all the sources detected by LOFAR and compares them with previous LOFAR observations and preloaded source lists from external catalogues. To detect transient and variable sources, the responsible algorithms are investigated. Algorithms concerning source association and variability detection are developed inside the database engine as an integrated part of the software pipeline. To quantify the reliability of a source association, we use a likelihood ratio method, whereas for detecting source flux changes we use two variability indices. The methods are applied to simulated data and the intensively monitored field of GRB 030329. The criteria we found for classifying a source association as reliable and for detecting a source as variable, were applied to the real data sets of GRB 030329. The Transients Key Project software pipeline identified several sources of interest, besides the target object, in the GRB 030329 field. Besides GRB 030329, we found eight sources to be variable in one or more bands of which two are serious candidates for multiwavelength follow-up observations.

### 3.1 Introduction

The International LOFAR Telescope currently has 28 stations out in the fields that are operational<sup>1</sup>. Core and remote stations are located in the Netherlands with baselines up to 3 and 100 km respectively, whereas the European stations have baselines up to 1000 km. A station field consists of two types of antennas, the Low Band Antennas (LBAs) and the High Band Antennas (HBAs). Both types are dual-dipole antennas that are sensitive in the low-frequency regime of 30–80 MHz and 110–240 MHz, respectively. Submillijansky sensitivities in the High Bands can be achieved in less than 500 seconds of integration time in full-array mode, with fields of view of the order of 10 deg<sup>2</sup>. Larger fields of view, > 100 deg<sup>2</sup>, are formed in the Low Bands, where sensitivities reach approximately 20 mJy for similar integration times. The individual signals are preprocessed and beam-formed at the station site before they are sent to the central processing supercomputer. Here, the data will be filtered and correlated before they are moved to a dedicated cluster for calibration and imaging processing (e.g., de Vos et al., 2009; Wijnholds, 2010). An automated software pipeline, currently under development by the Transients Key Project, is connected to the imaging pipeline in order to inspect and search the calibrated images for transient events and variable sources.

In the near future the Million Sources Sky Survey (MSSS) is planned. It will provide LOFAR with a Global Sky Model (GSM) of the northern hemisphere. The GSM will improve and evolve in time and serves the calibration and imaging process once the source parameters are validated and the astrometry is better than 0.5 arcsec to meet LOFAR's resolution of the longest baselines. Furthermore, this survey will give more technical insight into the next phases and prepares a fully operational LOFAR for future surveys.

Other dedicated configurations of LOFAR will serve as strategic modes for detecting transient and variable sources. For instance, two Radio Sky Monitor (RSM) modes, the Rapid All-Sky Monitor (RASM) and the Zenith Monitor (ZM) will both scan the northern hemisphere with large fields of view in continuous mode at distinct frequency bands, roughly once a day. Fender et al. (2007) describe the RSM modes, the main scientific foci of the Transients Key Project (TKP) and divides the sources likely to be detected by LOFAR into two types. To the first belongs the class of jet sources, incoherent synchrotron emitters, which is divided into subclasses of Cataclysmic Variables, X-ray binaries, Gamma-Ray Bursts (GRBs), Supernovae, and Active Galactic Nuclei (AGN). The second type contains the classes of the coherent radio emitters, among which are the flare stars, brown dwarfs, active binaries, extrasolar planets, and extragalactic radio bursts. Radio pulsars are also included and van Leeuwen & Stappers (2010) investigate

---

<sup>1</sup>Current status can be followed at <http://www.lofar.org>

detection strategies for LOFAR.

Analysis of the correlated and calibrated imaged data will be carried out by an automated software pipeline that will detect, classify, monitor and store new transients and variable sources on time scales down to one second. Pulsar observations will be processed in a dedicated pipeline, which is not described here. Internal as well as external triggers for follow-up observations should be handled. All the detected source measurements will be stored, causing a gradually evolving light-curve catalogue, that becomes available to the astronomical community at an early stage.

To achieve this, the TKP pipeline is attached to the standard imaging pipeline. The TKP pipeline is built in a Python based framework as described by Swinbank et al. (2007). The python package `scipy` now incorporates the source extraction routines from `stsci_python` developed at the STScI<sup>2</sup>. Spreeuw (2010) extended and improved the routines in performance concerning speed, astrometry and photometry. Extracted sources and related relevant data are stored into the column oriented database system `MonetDB`<sup>3</sup>. Source association and transients/variability detection algorithms are shifted *into* the programmable database engine, enabling fast access to catalogued and light-curve data.

In this Chapter we will describe the main algorithms and their database implementation in Section 2. The characteristics and criteria of the association and variability algorithms of simulated data are described in Section 3, and in Section 4 they are applied to the intensively monitored field of GRB 030329. In the last section we discuss the results.

## 3.2 Database and Internal Algorithms

### 3.2.1 Column Oriented Storage Model

Two of the main goals of LOFAR and the TKP are to detect transient and variable sources, and to build a catalogue that contains all the individual measurements of the sources observed by LOFAR, as raw material for further study of the transients.

The growth and volume of this catalogue database and the ability to do data mining at the same time, in combination with user ad-hoc query requests, rule out most of the mainstream relational database systems.

A fundamentally different design of a database system is the column oriented model adopted in `MonetDB` (Boncz, 2002). In this model the columns of a relational table are represented by binary tables mapping a unique object ID to a single attribute, thereby only retrieving relevant columns. Standard TPC-H Benchmarking for large volumes of data show significant

---

<sup>2</sup>See <http://www.scipy.org/SciPyPackages/Ndimage> for documentation

<sup>3</sup><http://monetdb.cwi.nl>

differences<sup>4</sup> between relational and column oriented systems. When sorted in a dense sequence, efficient positional lookups are enabled. Idreos, Kersten & Manegold (2002) further optimised the query processing architecture by implementing a cracking scheme that can speed up repetitive queries by two orders of magnitude. Furthermore, MonetDB's kernel is a programmable relational algebra machine operating on array-like structures, which allows even further optimisation of the pipeline software. The queries that are executed from within the TKP pipeline are known and *fixed*, thereby benefitting most from the characteristics mentioned above.

#### 3.2.2 Database Schema

The implemented database schema of the TKP pipeline is relatively simple, and therefore we will only highlight the most relevant tables. Calibrated images that are produced by the imaging pipeline will serve as input for the TKP pipeline. Raw image data are temporarily stored on the storage cluster with a capacity of about a Petabyte. Archival and temporary data storage will not be discussed here. Before the images are processed, an observation entry containing the characteristics is created in the database. Sources from the major catalogues like VLSS, WENSS, NVSS and eventually LOFAR itself are loaded into the `catalogedsources` table for association purposes. Meta-data from an image, like observation timestamps, frequencies, synthesised beam shapes, etc., are stored in the `images` table, having a reference to the observation to which it belongs. Sources detected in an image by the Python source extraction modules will be stored directly in the `extractedsources` table, with a reference to the `images` table. For these sources, the database will be searched for matching detections in previous images (i.e. in the `extractedsources` table) and in the `catalogedsources` table. Only those sources that lie within the search area of an extracted source are taken into account. For all these candidate association pairs found, the association properties that will be explained in Section 3.2.3 are attached to the pair. An extracted–extracted pair will be appended to the `assocxtrsources` table, and analogously the found extracted–catalogued couples to the `assoccatsources` table. The source association procedures run inside the database, together with auxiliary functions. The association tables are defined such, that retrieval of various kinds of source lists is fairly easy, avoiding multiple and/or recursive joins. These tables can in fact be regarded as the light-curve tables.

Furthermore, newly detected sources and sources flagged as transient or variable by the algorithms described in Section 3.2.4, will be copied to a separate table. Writing to this table activates triggers that undertake a broad range of further actions (deeper evaluation, fitting, classification),

---

<sup>4</sup><http://monetdb.cwi.nl/SQL/Benchmark/TPCH/>

before alerts will be sent out. This last part is work in progress.

### 3.2.3 Source Association

To keep track of sources and fluxes at positions of interest in the sky the TKP pipeline must be able to match sources by position in time as well as in frequency. For this purpose, a large catalogue of sources and their light-curve characteristics must be available all the time with response times less than the shortest integration times. The source association supports the creation of a long-term catalogue, containing millions of distinct sources with their light-curve properties for scientific analysis and serving the calibration of LOFAR.

A simple method to associate two point sources is if their positions coincide within their positional errors. This was explored by Richter (1975). Further development by including spectral information and local source density was done by De Ruiter, Willis & Arp (1977) and Sutherland & Saunders (1992), who calculated and defined probability ratios of true and chance associations. Applications and modifications were made by Rutledge et al. (2000), who cross-correlated the ROSAT/Bright Source Catalog of X-ray Sources with the USNO A-2 Optical Point Sources. They compared on-source fields with source-offset background fields, in order to calculate for each source association pair a probability of the association being unique. Attached probabilities of source associations made at multiple frequencies, will give more precise information about the fluxes extrapolated to the LOFAR bands of the sources that serve as calibrators.

Following mainly Sutherland & Saunders (1992) and Rutledge et al. (2000) we will investigate the association parameters that set the criteria of a reliable association. The distance in arcsec on the sky between source  $i$  and its association counterpart  $j$ ,  $\theta_{ij}$ , is defined as the first association parameter, whereas the second is the normalised distance, weighted by the positional errors of both sources

$$r_{ij} = \sqrt{\frac{(\Delta\alpha)_{ij}^2}{\sigma_{\Delta\alpha,ij}^2} + \frac{(\Delta\delta)_{ij}^2}{\sigma_{\Delta\delta,ij}^2}}, \quad (3.1)$$

where

$$\sigma_{\Delta\alpha,ij}^2 = \sigma_{\alpha,i}^2 + \sigma_{\alpha,j}^2, \quad (3.2)$$

and  $\sigma_{\Delta\delta,ij}^2$  is defined analogously.  $\sigma_{\alpha(\delta),i}$  are the  $1\sigma$  rms uncertainties for source  $i$  in right ascension (declination), respectively. In these calculations we assume that the  $\sigma_\alpha$  and  $\sigma_\delta$  are uncorrelated.

Genuine associations have positional differences due to measurement errors, which follow a Rayleigh distribution (e.g., De Ruiter, Willis & Arp, 1977), that gives the probability of having a true association between  $r_{ij}$  and  $r_{ij} + dr_{ij}$

$$dp_{\text{true},ij} = r_{ij} \exp(-r_{ij}^2/2) dr_{ij}. \quad (3.3)$$

From the Rayleigh probability density distribution,  $r \exp(-r^2/2)$ , the derived integral probability of finding a source association at  $r \geq \rho$  is theoretically given by  $p_r(r \geq \rho) = \int_{r=\rho}^{\infty} r \exp(-r^2/2) dr = \exp(-\rho^2/2)$ . This may be used for determining the search radius,  $r_s$ , of the area that will be scanned for possible counterparts. If we allow missing  $10^{-4}$  counterparts, this corresponds to setting the search radius to  $r_s \leq 4.29$ , whereas missing a factor of  $10^{-7}$  limits the search radius to  $r_s \leq 5.68$ . All counterparts falling in the search area will be taken into account as a candidate for association.

The probability of a chance association occurring between  $r_{ij}$  and  $r_{ij} + dr_{ij}$  with a background source follows the Poisson distribution and depends on the local source density,  $n_L$ , of the search area

$$dp_{\text{chance},ij} = 2\pi r_{ij} dr_{ij} \sigma_{\Delta\alpha,ij} \sigma_{\Delta\delta,ij} n_L, \quad (3.4)$$

where  $n_L$  depends on frequency, region in the sky and the flux level, and where  $\sigma_{\Delta\alpha,ij}$  and  $\sigma_{\Delta\delta,ij}$  are defined as in Eq. 3.2.

We can now define the third association parameter as the likelihood ratio, given by

$$LR_{ij} = \frac{dp_{\text{true},ij}}{dp_{\text{chance},ij}} = \frac{\exp(-r_{ij}^2/2)}{2\pi \sigma_{\Delta\alpha,ij} \sigma_{\Delta\delta,ij} n_L}. \quad (3.5)$$

Because  $LR$  decreases exponentially with increasing  $r$  it is often more convenient to use the logarithmic value of  $LR$  as the value of the third association parameter. The reliability of a source association pair can be determined by evaluating the  $\log LR_{ij}$  value.

In short the processing is as follows. All images are processed in the TKP pipeline, and the detected sources are stored in the database. For every source that is inserted, the database will be searched for candidate sources that lie in an area of radius  $r_s$  around the current source. Previous images as well as the major catalogues (VLSS, WENSS and NVSS) will be included in the association search. For all source pairs found (current source—candidate source) the values defined above will be calculated and the pair will be appended to the association table. A source having multiple counterparts in its search area, will thus have multiple table entries. This table is defined in such a way that light-curve data points can be generated quite easily per source. Simple aggregate functions and selection clauses on the association parameters of this table are used to give statistical information for these sources.

### 3.2.4 Transient and Variability Detection

To inspect the detected sources on variability or transient behaviour, we need to have some indicators that will alert us when a flux is changing with time. We will compare the current state of a source with available information stored in the catalogue database. For this, summarised and/or averaged source properties might be useful when the cadence is too high to explore all the individual values. Here we will define two variability indices. The first is based on the flux deviation from the average and can be interpreted as the magnitude of variability. The other index weighs the deviation from the weighted average with the errors and can be seen as the significance of the variability.

The magnitude of the flux variability of a source can be expressed as the ratio of the sample flux standard deviation,  $s_\nu$ , and the sample arithmetic mean flux  $\overline{I}_\nu$ . The sample standard deviation is defined as

$$s_\nu = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (I_{\nu,i} - \overline{I}_\nu)^2}, \quad (3.6)$$

where  $\overline{I}_\nu = \sum_{i=1}^N I_{\nu,i}/N$ ,  $I_{\nu,i}$  is the  $i$ -th flux measurement of the source at frequency  $\nu$ , and  $N$  is the number of measurements. The first indicator, the magnitude of the flux variability of a source, written in aggregate form, which is easy to handle in database queries on bulk data, is then defined as

$$V_\nu \equiv \frac{s_\nu}{\overline{I}_\nu} = \frac{1}{\overline{I}_\nu} \sqrt{\frac{N}{N-1} (\overline{I}_\nu^2 - \overline{I}_\nu^{*2})}. \quad (3.7)$$

The second indicator, which expresses the significance of the flux variability, is based on reduced  $\chi^2$  statistics. We assume that the weighted average flux value is a fitted parameter, so that the number of degrees of freedom is  $N - 1$ . It is given by the sum of the squared deviations from the weighted average weighted by the errors, and divided by the number of degrees of freedom

$$\eta_\nu \equiv \chi_{N-1}^2 = \frac{1}{N-1} \sum_{i=1}^N \frac{(I_{\nu,i} - \overline{I}_\nu^*)^2}{\sigma_{I_{\nu,i}}^2}, \quad (3.8)$$

where now,  $\overline{I}_\nu^*$  is the *weighted* average; if we define the weight as  $w_{\nu,i} = 1/\sigma_{I_{\nu,i}}^2$ , with  $\sigma_{I_{\nu,i}}$  the  $1\sigma$  error of the  $i$ -th flux measurement then

$$\overline{I}_\nu^* = \frac{\sum_{i=1}^N w_{\nu,i} I_{\nu,i}}{\sum_{i=1}^N w_{\nu,i}}. \quad (3.9)$$

In Eq. 3.8,  $\eta_\nu$  is the well-known  $\chi^2$  probability distribution (see, e.g., Barlow, 1989; Bevington & Robinson, 2003), which, in our case, is expressed as

$p_{\eta_\nu}(\eta_\nu, N - 1)$ . Furthermore, we choose to test the null hypothesis,  $H_0$ , that the source under consideration is not variable. Contributing terms to  $\eta_\nu$  in the sum will be of the order of unity, giving a value of roughly one after  $N$  measurements. With the integral probability,  $p_{\eta_\nu} = \int_{\eta_\nu'=\eta_\nu}^{\infty} p_{\eta_\nu}(\eta_\nu', N - 1)d\eta_\nu'$ , we can quantify the probability  $p_{\eta_\nu}$  of having a value equal to or larger than the  $\eta_\nu$  obtained from the measurements. This probability  $p_{\eta_\nu}$  will justify a rejection of  $H_0$  once the confidence level is set. It also determines the expected rate of false positives based on the source counts models.

Again, the significance of the flux variability,  $\eta_\nu$  in Eq. 3.8, can be written as a simple aggregate function

$$\eta_\nu = \frac{N}{N - 1} \left( \frac{wI_\nu^2}{w} - \frac{\overline{wI_\nu^2}}{w} \right). \quad (3.10)$$

$\eta_\nu$  is close to unity when a source does not show significant variability. When a source has an outlier in one of its flux measurements  $V_\nu$  will be large, but depending on the error of the flux measurement, the significance of variability  $\eta_\nu$  will be either small when  $\sigma_{I_\nu}$  is large or large when  $\sigma_{I_\nu}$  is small.  $V_\nu$  and  $\eta_\nu$  will be helpful in the source classification routines. The classification might discriminate between four separate regions of source variability. Those where the magnitude  $V_\nu$  is large or small in combination with a large or small significance  $\eta_\nu$ .

Other descriptive light-curve parameters, like the probability of having a series of consecutive flux measurements deviating above or below from the average, and long-term variability fits are not investigated here, but will be investigated in future work.

### 3.3 Simulated Data

#### 3.3.1 The Data

To analyse the behaviour of the source association and variability parameters defined in Sections 3.2.3 and 3.2.4, we used a sample of 1000 images of simulated data. These were processed sequentially in the TKP pipeline. Sources in the images are at fixed positions and only vary due to the background noise. All the source associations that are made are therefore genuine and can be considered as true. These will give us the boundaries of  $r$  and  $\log LR$  in which reliable source associations fall. Furthermore, we inserted a transient source that went into outburst at some random point in time, followed by an exponential decay. By monitoring the variability indices, this source should be picked up by the TKP pipeline as a transient source. The variability indices  $V_\nu$  and  $\eta_\nu$  (or  $p_{\eta_\nu}$ ) of the sources in the sample may guide us in setting the constraints on the transient alert system.

The images are created from VLA archival data of the discovery observa-

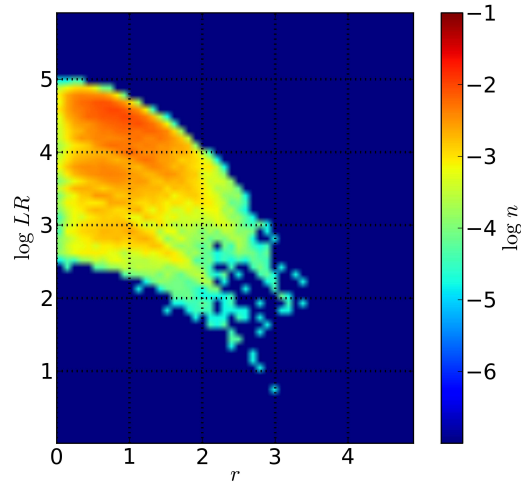
tion of GCRT J1745-3009 at 325 MHz by Hyman et al. (2005). As described in Chapter 2 of Spreeuw (2010), we only used the  $(u, v)$  coverage. Gaussian noise was added to the visibilities, by using the data reduction package AIPS and its tasks UVMOD and WTMOD and the adverb FLUX. An average rms noise of 10 mJy/beam was generated in each map. This was chosen to have it comparable to the 325 MHz WENSS  $5\sigma$  flux limit which is at 18 mJy as reported by Rengelink et al. (1997). The final 1000 noise maps were created with the AIPS task IMAGR in natural-weighting mode, with a pixel size of 12 arcsec in both directions, resulting in images that are  $256 \times 256$  pixels or somewhat less than  $1^\circ \times 1^\circ$  in size.

In every noise map we inserted 64 sources on a rectangular grid. The inserted position of a single source is the same in all the images, although fluctuations occur due to the different local noise in each map. The fluxes of the sources are linearly spaced between 1 mJy and 1 Jy. The lower limit was chosen to have sources below the average noise level as well as sources that have fluxes at the same level of the local noise in a single image, causing this source not to be detected in all of the images. Like the position, the flux of a single source is assumed to fluctuate due to different local noise in the images. The time between two consecutive images is of course arbitrary, but for the purpose of illustration it is set to 1 h.

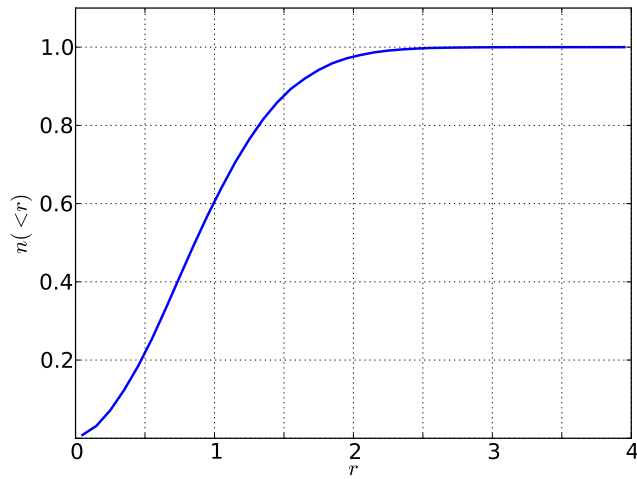
At a random point in time one of the existing sources is substituted by a transient source. This transient has a random peak flux between 1 and 5 Jy and an exponential decay time chosen randomly to lie between 10 and  $10^5$  s.

### 3.3.2 Source Association, $r$ and $\log LR$

We know that all the source associations made in the sample are true and that if a source could not be associated it is because its flux is below the detection threshold or disappears in the local noise of the image. The normalised number distribution of the dimensionless distance,  $r_{ij}$ , and the logarithm of the likelihood ratio,  $\log LR_{ij}$ , of the (true) association pairs of the sample are shown in Fig. 3.1. The numbers peak towards low  $r$  and high  $\log LR$ . The minimum value of  $\log LR$  is at  $\log LR_{\min} = 0.80$ , whereas the maximum value of  $r_{ij}$  for a source association in the sample is at  $r_{\max} = 3.48$ , corresponding to a probability of 0.2%. Fig. 3.2 shows the normalised cumulative distribution of  $r$  of the associations found in the sample. As can be seen, classifications for genuine associations fall well below the theoretical cutoffs as given in Section 3.2.3. From the sample values associations may be classified as genuine for cutoffs at  $r \leq 3.3$  or at  $\log LR > 1.4$ , in which cases we omit 0.01% of the true associations. These numbers differ slightly from the given theoretical values in Section 3.2.3, because here we compare the positional deviations in  $\alpha$  and  $\delta$  of a source association to the first stored detection of a source, which does not change in time during the run, whereas the Rayleigh distribution expects both values to be randomly chosen from



**Figure 3.1:** Histogram of the normalised number distribution of the association pairs in the sample of 1000 images. On the horizontal axis the dimensionless normalised distance,  $r_{ij}$ , of an association pair is plotted, and on the vertical axis the logarithm of the corresponding likelihood ratio ( $\log LR_{ij}$ ). The cell sizes are  $dr \times d\log LR = 0.1 \times 0.1$ . The color bar represents the logarithm of the normalised number of association pairs in a cell.

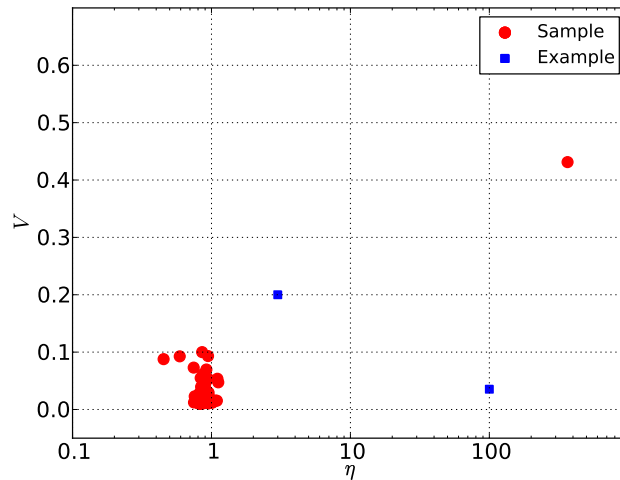


**Figure 3.2:** Normalised cumulative distribution of the normalised distance,  $r$ , of the source association pairs in the sample as described in Section 3.3.1. The bin width of  $r$  is fixed to 0.1.

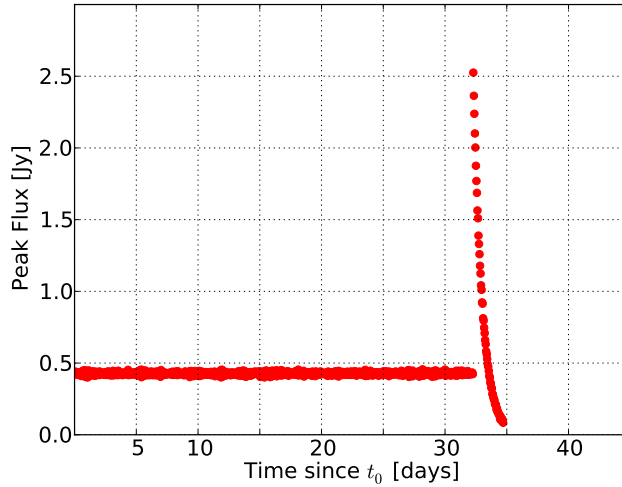
their Gaussian distributions.

### 3.3.3 Variability

The variability indices  $V_\nu$  and  $\eta_\nu$  of a source are calculated using all its  $N$  flux measurements. From now on we drop the subscript  $\nu$  since we work in a single frequency band. The variability indices of all 64 sources in the sample of 1000 images are plotted in Fig. 3.3. As briefly mentioned in Section 3.2.4 it may be divided into four characteristic regions. The lower left containing the sources that do not show pronounced and significant variability. Fluxes may fluctuate, but only in the margins of the rms noise. Sources residing in the lower right part show low, but significant, variability. A source showing periodic variability of, say 1%, where the deviations from the average are well above the rms will be located here. Going to the upper left domain we will encounter sources that show variability, but at low significance. Large deviations from the mean, in combination with large flux errors will result in a lower level of significance. Sources falling in the upper right quarter, display large *and* significant variability. As can be seen in Fig. 3.3, the transient is located in the upper right area at  $V = 0.431$  and  $\eta = 365.58$ . The  $\eta$ 's of the other sources, with  $\eta_{\max} = 1.119$ , all fall in the regime where the probabilities of these values for  $\eta$  are  $p_\eta > 0.5\%$ , meaning that we cannot reject  $H_0$  for these sources. Stricter rejection criteria are met when



**Figure 3.3:** The significance of variability  $\eta_\nu$  against the magnitude of variability  $V_\nu$  for the 64 sources in the sample. Each source is represented by a red circle, whereas the number of associations that was made for a source is not shown here. The red circle in the upper right part of the plot is the inserted transient. The two blue squares are plotted for explanatory purposes (see text).



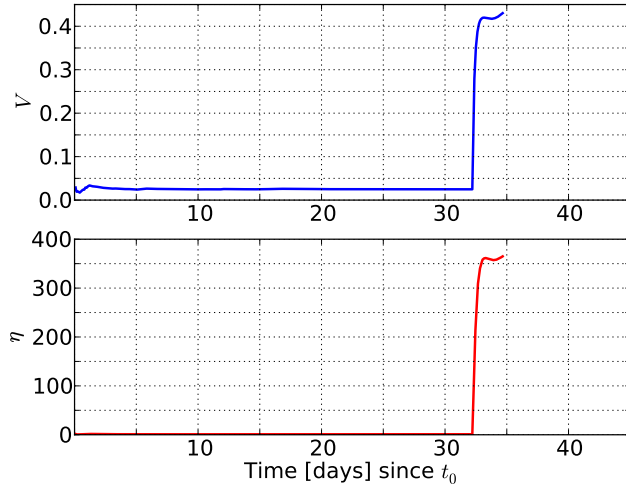
**Figure 3.4:** The light curve of the source with the largest variability indices ( $V_\nu = 0.431$ ,  $\eta_\nu = 365.58$ ) in the sample. The horizontal axis shows the number of days elapsed since  $t_0$  (time of first image) for the entire sample. The vertical axis shows the measured peak flux (Jy). Note that the average rms noise is 10 mJy/beam in each image, causing the light curve to drop below the detection threshold around day 36.8.

we set cutoffs for the probability at  $p_\eta < 10^{-3}$ . In our data sample, with  $N - 1 = 999$ , this would imply finding sources with  $\eta_\nu \geq 1.144$ . We have to note that the variability indices of the sources, as plotted in Fig. 3.3, were calculated after all the images were processed.

The light curve of the transient as found by the TKP pipeline is plotted in Fig. 3.4, from which it can be seen that the flux was constant during the first 32 days of the sample (774 h after  $t_0$ ), then increased significantly, dropped exponentially until it disappeared in the noise and became undetectable after day 34.7.

Fig. 3.5 shows the development of the variability indices with time of this transient source. Both parameters are fairly constant in time, apart from the beginning when they need to settle. The sudden steep rise occurs at the same moment of outburst of the transient source (775 h after  $t_0$ ) and continues until the end of detection.

Let us assume that we have an intrinsically variable source with an average flux of  $\bar{I}$ , and that all the measurements were taken with the same error as the rms of the noise  $\sigma_{\text{rms}}$ , in order to apply a constant weight  $w$ . For large  $N$ , Eqs. 3.7 and 3.10 can be rewritten to have for this source  $\bar{V} = \sqrt{d\bar{I}^2/\bar{I}}$  and  $\bar{\eta} = w d\bar{I}^2$ , respectively. The latter may be expressed in terms of the rms noise or with the use of  $d\bar{I} = n\sigma_{\text{rms}}$  as  $\bar{\eta} = n^2$ . This means that a well-sampled variable source with an amplitude of the flux variations at 5% of the



**Figure 3.5:** The development of the two variability indices as function of time for the monitored transient source.  $V$  and  $\eta$  rise steeply at the time of flux enhancement.

average flux ( $A = a\bar{I}$ , with  $a = 0.05$ ) and its  $dI$  all at the  $10\sigma_{\text{rms}}$  level will be located in the lower right region at  $\eta = n^2 = 100$  and  $V = a/\sqrt{2} \approx 0.035$ . This example source is plotted in Fig. 3.3 as one of the blue squares. Another example is a source having  $\eta = 3$ ,  $V = 0.2$  (see Fig. 3.3). Depending on the number of degrees of freedom,  $N - 1$ , the probability  $p_\eta$  is used to accept or reject  $H_0$ . For six data points,  $N - 1 = 5$  and  $p_\eta = 1.04 \times 10^{-2}$ , which is not that unlikely as could be seen from the sample. However, if we had a larger number of measurements for these values, and thus more degrees of freedom, the probabilities would decrease further, making the rejection of  $H_0$  very plausible.

The fastest transient and variability triggers will be based on the evaluation of the (simple) variability indices,  $V_\nu$  and  $\eta_\nu$ . Other simple triggers are included when the flux of a known, i.e., catalogued, source, deviates more than  $n$  percent from the average value. As soon as one of the parameters rises significantly, the triggers will activate other inspection modules. These might consist of fitting and classifying the light curve, sending alerts (internal as well as external) when a series of consecutive rising data points in the variability curves are observed. Reconfiguration of the telescope to observe with higher time and angular resolution might then be one of the follow-up strategies.

## 3.4 GRB 030329 Field

### 3.4.1 The Data

GRB 030329 was one of the nearest gamma-ray bursts detected and was among the brightest and best-sampled radio, optical and X-ray afterglows ever. This led to still ongoing long-term multifrequency radio follow-up campaigns spanning more than seven years now. The afterglow was followed by the Westerbork Synthesis Radio Telescope (WSRT) at frequencies ranging from 350 MHz to 8.4 GHz (e.g., Van der Horst et al., 2008; Kamble et al., 2010, and references therein). The intensively monitored field of view of GRB 030329 makes it a good candidate to search for transients and variable sources.

For this, we collected already reduced and analysed WSRT data from observations taken in the period from the end of 2003 to the end of 2007. We refer to Van der Horst et al. (2008) for the details of the data reduction. In order to further optimise the WSRT data we performed the AIPS task `PBCOR` to have the images corrected for the primary beam attenuations at the different frequencies. The selection of the datasets for our sample is listed in Table 3.1.

To search for variable sources we apply the algorithms described in Sections 3.2.3 and 3.2.4 while processing all images in the TKP pipeline. We set the source extraction detection threshold to seven times the local rms noise. Spreeuw (2010) showed that this value minimises the number of false source detections near the edges of the images and still finds the fainter sources. The characteristic rms noise in the WSRT 1.4 GHz images is of the order of 0.1 mJy/beam, with a synthesised beam of  $62'' \times 21''$ . Extracted sources from an image were associated with detected sources from previous processed images in the sample. In this way data points are appended to the light curves of the sources already found in the sample, or when detected for the first time a new branch for light-curve data points is created. The association algorithm only couples sources by positional match and does not take into account any assumptions on the spectral correlation of the associated pair. This enables the routines to find association pairs across multiple frequency bands by just evaluating their positions. The optical or X-ray counterparts positions of AGNs might be offset from the radio up to the level of arcmins. Here we will fix the search to the radio domain, but this should be taken into account when optical and/or non-radio sources are included in the search. In this sample the VLSS (74 MHz) and NVSS (1.4 GHz) catalogues were also looked up to find possible counterparts for the extracted WSRT sources. The WENSS (325 or 352 MHz) catalogue was excluded because it has no overlapping areas with the GRB 030329 field.

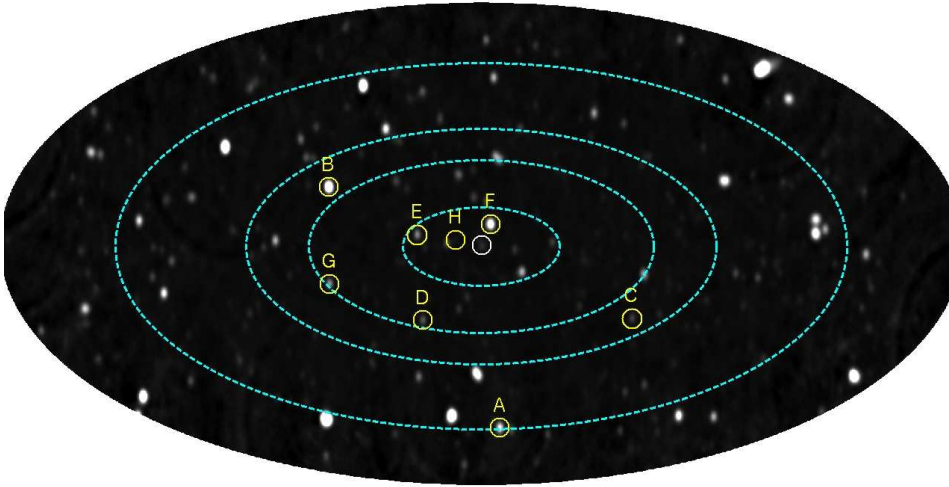
Sources far away from the pointing centre will be affected by the primary beam attenuation, which is frequency dependent. Pointing errors may in-

Observation start date	Time since burst [days]	Time on target [hours]	Frequency [MHz]
2003-12-22	268.577	6.5	2300
2003-12-25	270.612	4.4	1400
2004-01-25	302.599	12.0	350
2004-01-29	306.408	3.3	1400
2004-01-30	306.753	3.3	2300
2004-02-10	318.555	12.0	840
2004-03-26	363.816	6.4	4800
2004-03-27	364.424	11.7	1400
2004-03-28	365.524	7.0	2300
2004-04-11	379.267	6.2	840
2004-05-02	400.331	12.0	350
2004-05-19	417.342	9.2	1400
2004-05-22	420.168	6.8	2300
2004-07-03	462.078	8.0	2300
2004-07-04	463.076	8.0	4800
2004-08-01	490.974	6.7	1400
2004-09-25	546.013	8.1	4800
2004-11-02	583.683	5.0	2300
2004-11-11	592.721	8.0	1400
2004-11-12	593.802	12.0	840
2005-03-24	726.439	12.0	840
2005-03-26	728.433	12.0	350
2005-04-09	742.395	12.0	1400
2005-04-10	743.892	12.0	2300
2005-05-14	777.242	9.2	4800
2005-11-27	974.759	12.0	4800
2005-12-07	984.163	13.8	2300
2005-12-09	986.132	12.5	1400
2006-04-09	1107.399	11.8	840
2006-04-30	1128.338	12.0	4800
2006-10-08	1288.899	12.0	1400
2006-10-29	1309.842	12.0	4800
2006-11-12	1323.803	12.0	840
2007-11-03	1679.828	12.0	4800
2007-11-18	1694.787	12.0	1400
2007-11-25	1701.768	12.0	2300

**Table 3.1:** Log of WSRT observations of GRB 030329 that are used in our sample. The indicated time on target is the difference between the logarithmic average of the start and end of the integration and the time of the burst.

Frequency [MHz]	Resolution [arcsec]	Sensitivity [ $\mu$ Jy]	FWHM [degrees]	$r_{\text{field}}$ [arcsec]
840	26	400	0.83	1400
1400	13	100	0.5	900
2300	7.8	50	0.37	660
4800	3.7	30	0.17	300

**Table 3.2:** Properties of the fields that were searched in our sample. Typical values for the spatial resolution and FWHM of the primary beam were adopted from the WSRT Guide at <http://www.astron.nl>.



**Figure 3.6:** The GRB 030329 field at 1.4 GHz as observed on 27 March 2004 by WSRT. The ellipses in the outward direction represent the search areas centred around GRB 030329 at 4.8, 2.3, 1.4 GHz, and 840 MHz, respectively. The inner ellipse resembles the search area at 4.8 GHz and has major and minor axes of 300 arcsec. As a reference, the sources that are discussed in the text are marked by the labeled circles A through H, where the white circle in the centre represents GRB 030329. The synthesised beam size is  $62'' \times 20.7''$ .

introduce fluctuations of 5 to 10% in the flux of sources at the edge of the full width half maximum (FWHM) of the primary beam (T. Oosterloo, private communication). We therefore select down to only those sources that lie within the FWHM of the primary beam, of which the pointing centre is at the GRB location. Table 3.2 reports the radii of the search areas centred at the pointing centre that will be searched for variables at the corresponding frequencies. Resolutions are listed as well because sources might get resolved towards higher frequencies, causing difficulties in fitting multiple Gaussians, which may in turn lead to higher values in  $V_\nu$  and  $\eta_\nu$  as will be demonstrated in the next section.

Fig. 3.6 shows the field of GRB 030329 at 1.4 GHz on 27 March 2004. The dashed ellipses resemble the search areas at the frequencies 4.8, 2.3, 1.4 GHz, and 840 MHz (in order of increasing size). The pointing centres of the observations are all at RA =  $10^{\text{h}}44^{\text{m}}50.0^{\text{s}}$  and Dec =  $21^{\circ}31'17.8''$ , which is 0.6 arcsec to the northeast of the VLBI position of GRB 030329, RA =  $10^{\text{h}}44^{\text{m}}49.96^{\text{s}}$ , Dec =  $21^{\circ}31'17.44''$  as reported by Taylor et al. (2007). The labeled circles mark the sources that are variable to some extent and will be discussed below.

### 3.4.2 Results

The values for  $r$  and  $\log LR$  as determined in Section 3.3.2 for genuine associations obey the ideal cases. The images in the sample of the GRB 030329 observations all have different characteristics and will affect the association parameters, causing the ideal values to be too strict in some cases (see, e.g., Source B in Section 3.4.2).

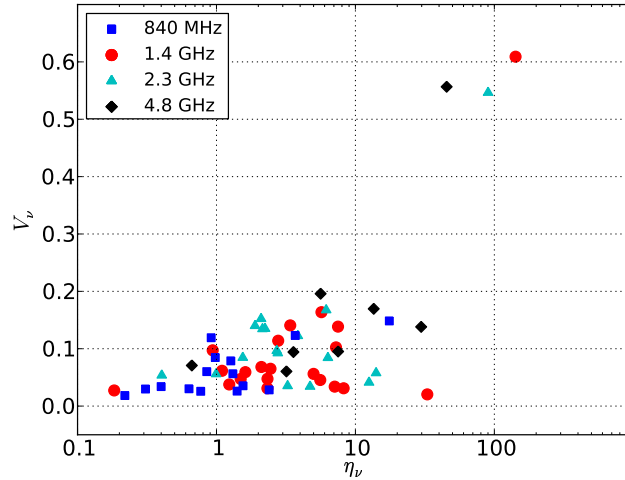
### Variability

The variability indices  $V_{\nu}$  and  $\eta_{\nu}$  of the sources are determined per frequency band, as *a priori* the spectral behaviour across the frequency bands is unknown. Fig. 3.7 shows the relation between the two variability indices for the sources that fall within the FWHM of the beam of the corresponding frequency bands.

Associations in the 350 MHz band were excluded in this plot, because the three observations were too few for a statistical interpretation, although this band was included in the source association routines in order to have full coverage of a source's light curve. A single symbol corresponds to a source and all its associations that were made in the same frequency band. It does, however, not depict the number of associations that was found for the source. Due to some poor  $(u, v)$  coverage images and therefore higher noise levels, a source might not get detected in the whole time series. We selected sources having at least 5, 7, 6, and 5 associations in the frequency bands of 840 MHz, 1.4, 2.3, and 4.8 GHz, respectively.

There were no single detections of sources having fluxes larger than 0.5 mJy in the different frequency bands. We focus on the sources that stand out in the variability plot of Fig. 3.7 and combine the fluxes of the associated sources across the frequency bands to get their spectral light curves. An overview of these sources is presented in Table 3.3.

The sources that are alphabetically labeled in Table 3.3 are represented by the circles in Fig. 3.6. They all showed significant variability in at least one band with  $\eta_{\nu}$  values having probabilities of  $p_{\eta} \leq 10^{-6}$ . Sources that did show variability in just one band with  $p_{\eta} > 10^{-6}$  were left unlabeled.



**Figure 3.7:** The variability indices for the sources in the GRB 030329 field at the different frequency bands of the WSRT observations. A single symbol corresponds to a source and all its associations. The number of times a source could be associated is not displayed here. GRB 030329 reveals itself by the three data points in the upper right region.

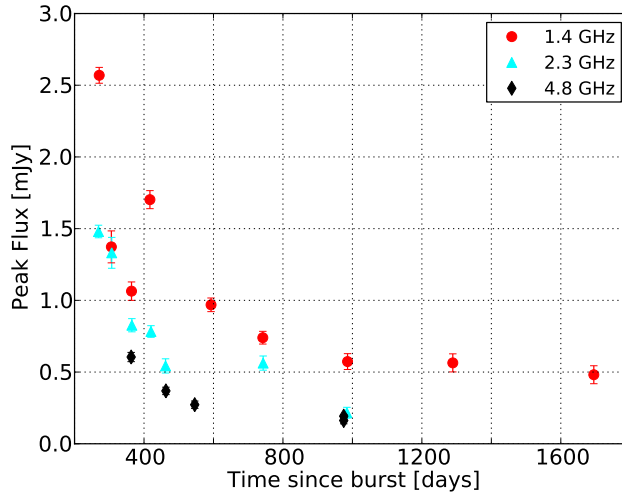
### Individual Sources and Light Curves

GRB 030329. The three data points in the upper right region of Fig. 3.7 originate from the same source: GRB 030329. Fig. 3.8 shows the light curve of the GRB 030329. Because of the higher resolution in the 4.8 GHz observation in November 2005 (approximately 975 days after the burst) in combination with a slightly offset positional fit of GRB 030329 in the initial image, which is considered as the base point of a light curve, two sources could be associated to the GRB. One has a peak flux of  $192 \pm 12 \mu\text{Jy}$  and an integrated flux of  $121 \pm 15 \mu\text{Jy}$  at 50.1 arcsec from the VLBI position of GRB 030329. The other candidate has a peak flux of  $162 \pm 23 \mu\text{Jy}$  and an integrated flux of  $149 \pm 38 \mu\text{Jy}$  at a distance of 0.7 arcsec from the VLBI position of GRB 030329. The integrated flux of the latter source was reported by Van der Horst et al. (2008). As mentioned, the measured position of GRB 030329 in the 1.4 GHz images to which these sources were matched is roughly in between the two association candidates. This arises from the fact that the first positional fit was done in a lower-resolution 1.4 GHz image, resulting in a distance of nearly 30 arcsec from the VLBI position<sup>5</sup>. Both association candidates'  $r$  (3.32 and 2.94, resp.) and  $\log LR$  (0.60 and 1.1,

<sup>5</sup>The current schema that is in place avoids this first-measurement bias. A *running* catalogue is maintained with mean weighted positions and errors from the associated sources. In effect this is analogous to the static positions of the sources in the external catalogues.

Source Label	Distance [arcsec]	Frequency Band Variability												RA & Dec		
		840 MHz			1.4 GHz			2.3 GHz			4.8 GHz			Catalogue	Source Name	
		$N$	$V_\nu$	$\eta_\nu$	$N$	$V_\nu$	$\eta_\nu$	$N$	$V_\nu$	$\eta_\nu$	$N$	$V_\nu$	$\eta_\nu$			
GRB030329																
A	1396	5*	0.148*	17.5*	9	0.609	142.6	7	0.547	90.16	5	0.557	45.37	$10^{\text{h}}44^{\text{m}}49.96^{\text{s}}$	$21^{\circ}31'17.44''$ †	
B	739	6	0.012	5.12	8	0.020	32.91							$10^{\text{h}}44^{\text{m}}44.95^{\text{s}}$	$21^{\circ}08'01.83''$	
C	799	4	0.149	2.54	10	0.138	7.50							NVSS J104444+210801		
D	609	4	0.030	0.132	9	0.038	1.24	9	0.084	6.34	6	0.095	7.51	SDSS J104444.97+210801.8		
E	262	6	0.060	0.85	10	0.045	5.58	6	0.034	4.72	6	0.095	7.51	$10^{\text{h}}45^{\text{m}}31.87^{\text{s}}$	$21^{\circ}38'49.96''$	
F	174	4	0.005	0.07	8	0.031	8.23	8	0.041	12.53	4	0.138	29.73	NVSS J104531+213850		
G	649	5	0.018	0.22	9	0.034	7.10	8	0.058	14.09	8	0.170	13.55	$10^{\text{h}}44^{\text{m}}08.72^{\text{s}}$	$21^{\circ}22'04.61''$	
H	112													$10^{\text{h}}45^{\text{m}}06.01^{\text{s}}$	$21^{\circ}21'51.27''$	
														$10^{\text{h}}45^{\text{m}}07.62^{\text{s}}$	$21^{\circ}32'48.93''$	
														NVSS J104507+213249		
														$10^{\text{h}}44^{\text{m}}47.46^{\text{s}}$	$21^{\circ}34'07.91''$	
														NVSS J104447+213409		
														$10^{\text{h}}45^{\text{m}}31.64^{\text{s}}$	$21^{\circ}26'29.32''$	
														NVSS J104532+212630		
														$10^{\text{h}}44^{\text{m}}57.08^{\text{s}}$	$21^{\circ}32'09.55''$	
														SDSS J104457.09+213210.1		

**Table 3.3:** Sources that showed variability in the sample. The source labels correspond to the labels in Fig. 3.6. The Distance column lists the (unweighted) average distance of all the position measurements of a source to the pointing centre.  $N$  is the number of associations that could be made for a source,  $V_\nu$  the magnitude and  $\eta_\nu$  the significance of the variability, all three specified per frequency band. The corresponding plot of the variability indices of all the well-detected sources in the sample is shown in Fig. 3.7. The last column lists the weighted average source position in Right Ascension and Declination as determined by the source extraction procedures and the corresponding catalogue source name if an association could be established. No VLSS counterparts were found for these sources. The SDSS associations were found manually via the SkyServer webform. \*: If we include the measurements of Source A that fell outside the 840 MHz FWHM search area, the new variability parameters are:  $N = 6$ ,  $V_{840\text{MHz}} = 0.137$  and  $\eta_{840\text{MHz}} = 15.5$ . †: From Taylor et al. (2007).



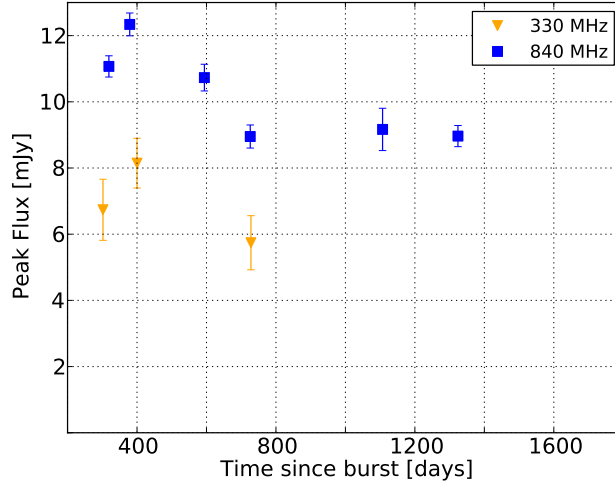
**Figure 3.8:** The spectral light curve of GRB 030329.

resp.) values are comparable, so that neither source could be ruled out *a priori* as a genuine association, although the second candidate is more likely to be the counterpart. Therefore, the fluxes of both sources are plotted in GRB 030329's light curve in Fig. 3.8.

Knowing the GRB position in hindsight with VLBI accuracy of sub-arcsec level one of the sources can be ruled out. This demonstrates the importance of reliable and up-to-date catalogues being available during observation runs, and can save further time consuming processing analysis once a catalogue counterpart is known.

Furthermore, from the large values of the variability indices, this type of transient would have been picked up by the TKP pipeline.

**Source A.** Source A is so close to the edge of the search area that one of the associations was not taken into account because its distance to the pointing centre was slightly larger than  $r_{\text{field}}$  at 840 MHz. Because the search area was chosen conservatively and falls within the FWHM of the beam this data point is included in the light curve that is shown in Fig. 3.9. If this measurement is also included in the calculations of the variability indices, the reported values from Table 3.3 change to  $N = 6$ ,  $V_{840\text{ MHz}} = 0.137$  and  $\eta_{840\text{ MHz}} = 15.5$ . Source A falls outside the 1.4 GHz search area, and therefore only the 330 MHz and 840 MHz data points are plotted. Both probabilities of having  $\eta_{840\text{ MHz}} \geq 17.5$  for five or  $\eta_{840\text{ MHz}} \geq 15.5$  for six measurements is  $p_{\eta} < 10^{-13}$ , making it very plausible to reject  $H_0$  for this source, even though the time sampling is sparse and the number of data points are few. Furthermore the 330 MHz data points behave similarly,



**Figure 3.9:** The light curve of Source A. The 1.4 GHz data are omitted due to the fact that the source lies outside the 1.4 GHz FWHM area. The first data point in the 840 MHz band was omitted in the calculations of the variability indices in Table 3.3, because its distance to the pointing centre was 1402 arcsec.

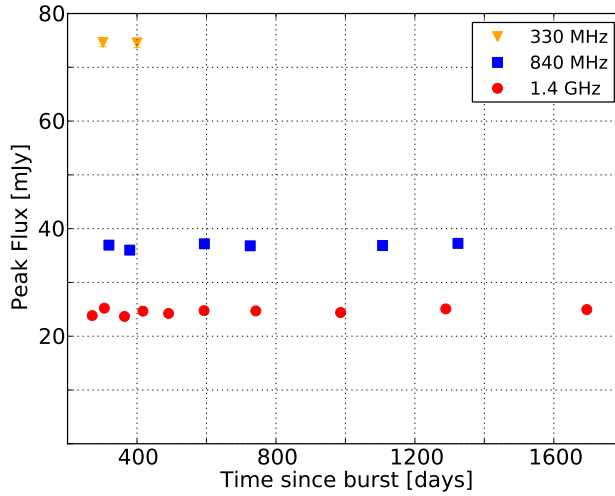
suggesting this source underwent an outburst.

The NVSS counterpart found for source A is J104444+210801 (see Table 3.3), which could be associated for all the detections in the GRB 030329 field, with an average value of the normalised association distance of  $r = 4.0$ . We searched the optical catalogue of the Sloan Digital Sky Survey (SDSS) and found a nearest-neighbour source<sup>6</sup>, SDSS J104444.97+210801.8, in the search area ( $r_s = 90$  arcsec) centred at Source A at a distance of 0.40 arcsec, but no further (spectral) information about the source is known.

**Source B.** From Table 3.3 and Fig. 3.7 it can be seen that Source B has the second largest value for  $\eta_{1.4\text{GHz}}$ . Its spectral light curve is shown in Fig. 3.10, where we excluded the 2.3 GHz detections because they were outside the FWHM beam; it was not detected at 4.8 GHz. It is a relatively bright source and therefore the flux variations around day 400 are large when weighted by their relatively small errors and contribute much to the sum in  $\eta_{1.4\text{GHz}}$ .

Closer visual investigation reveals that some associations were left out because of the strict selection of only those pairs having  $\log LR > 2$ . For these pairs the normalised distance  $r$  was large due to small positional errors relative to the distance between the sources of the pair, leading to a negative value for  $\log LR$  for a genuine association. If we relax the criterion (only in this case) to also selecting pairs having  $\log LR > -10$ , we get for the

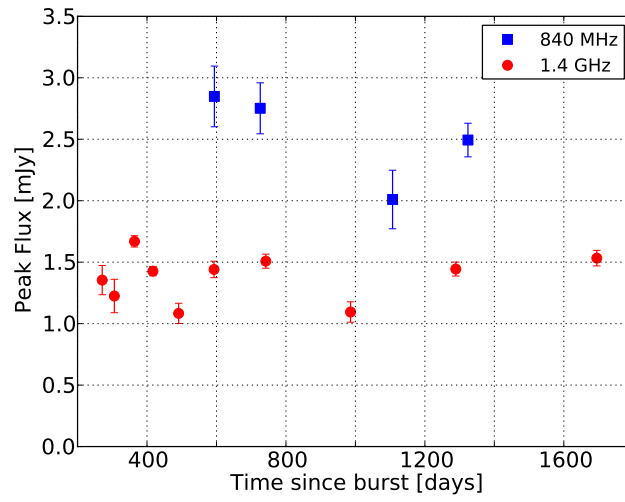
<sup>6</sup><http://cas.sdss.org/astrodr7/en/tools/explore/obj.asp?id=587742014348001931>



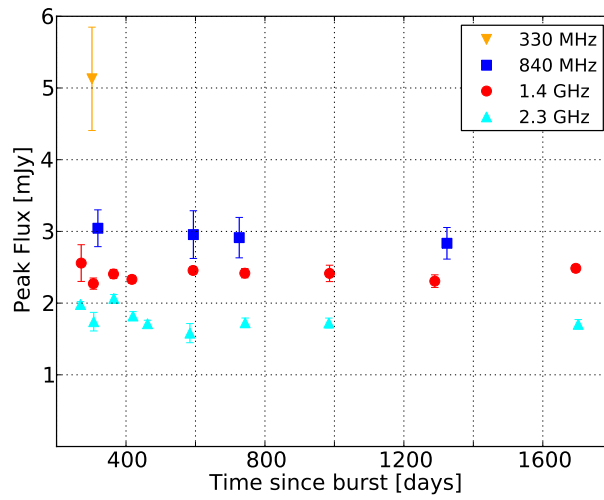
**Figure 3.10:** The light curve of Source B. At 1.4 GHz, the first measurement and the one near day 600 could be included in a less strict selection criterion for  $\log LR$ .

1.4 GHz variability indices  $V_{1.4\text{GHz}} = 0.021$  and  $\eta_{1.4\text{GHz}} = 31.27$  for  $N = 10$  data points (while previously the values were  $V_{1.4\text{GHz}} = 0.020$  and  $\eta_{1.4\text{GHz}} = 32.91$ , see Table 3.3). The corresponding probability is too low to accept  $H_0$ . The lower value for  $\eta_{840\text{MHz}}$  has a probability of  $p_\eta \approx 10^{-4}$ , which is too high to justify a rejection of  $H_0$ . Based on these results, this source shows pronounced variability in the 1.4 GHz band, whereas variability in the other bands could not be measured significantly enough.

**Source C.** The light curve of Source C is shown in Fig. 3.11. Since there were only four detections in the 840 MHz band, the significance of the variability is well below the threshold level to declare it variable. At 1.4 GHz the probability for the measured  $\eta_{1.4\text{GHz}} = 7.5$  is  $p_\eta \approx 10^{-10}$ , small enough to reject  $H_0$ . Source B and C are nearly diametrically opposite to each other, but do not show evidence for correlated opposite flux variations due to beam attenuation effects. Both sources did not have closely neighbouring sources to test the primary beam attenuation effects in more detail. No reliable counterpart could be found in the SDSS Catalogue.



**Figure 3.11:** The light curve of Source C.



**Figure 3.12:** The light curve of Source D.

Source D. The light curve of Source D is shown in Fig. 3.12. This source does not show significant variability outside the 2.3 GHz band, where the probability for the  $\eta_{2.3\text{GHz}}$  value is  $p_\eta < 10^{-7}$ . Source D lies well within the FWHM, and therefore the variations in flux of 20% from average are not likely caused by primary beam attenuations. Although the fluxes in the other bands change similarly as in the 2.3 GHz band, the significance of variability was too low for a  $H_0$  rejection. There were no counterparts within the search radius of 90 arcsec in the VLSS, NVSS and SDSS Catalogues.

Source E. This source shows variability in three of the four bands in which it was detected. The flux measurement at 1.4 GHz at day 491 deviates significantly from the average (see Fig. 3.13) and causes the variability index to rise. It has to be noted that the quality of this image was quite low, presumably due to the poor  $(u, v)$  coverage. The probabilities for the variability indices in the 1.4 and 4.8 GHz bands are both  $p_\eta < 10^{-7}$ , making a rejection of  $H_0$  plausible. At 2.3 GHz the probability of having  $\eta_{2.3\text{GHz}}$  does not allow us to reject  $H_0$ . Therefore, Source E showed significant variability in two bands.

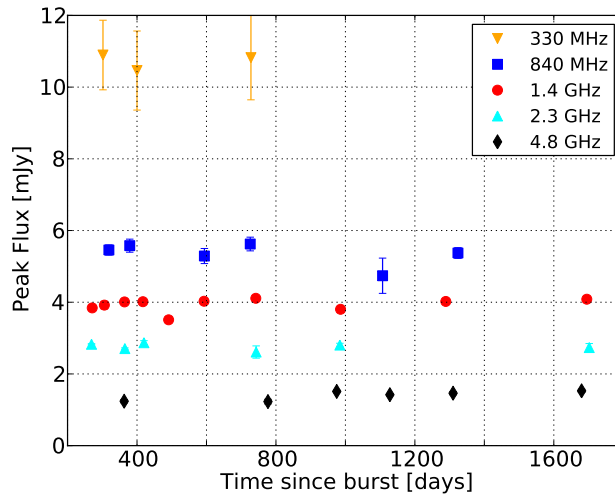
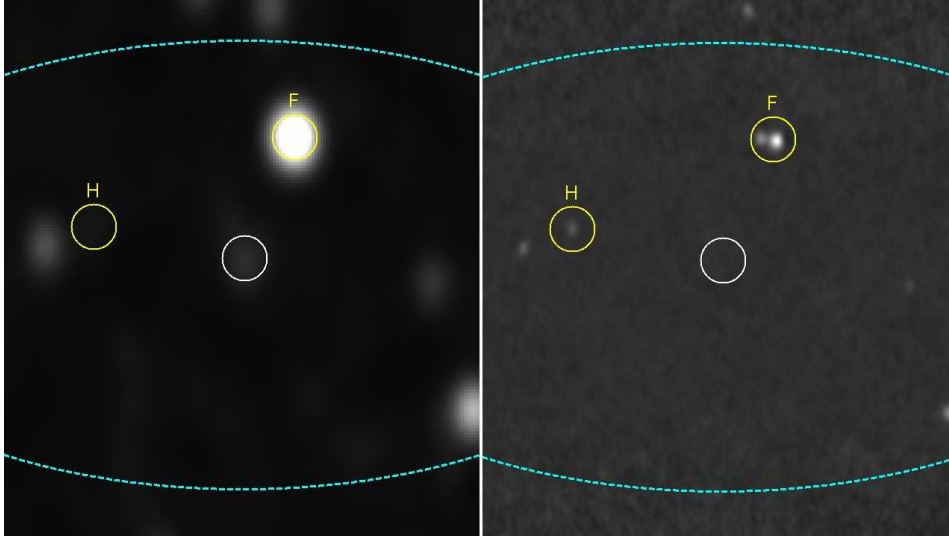
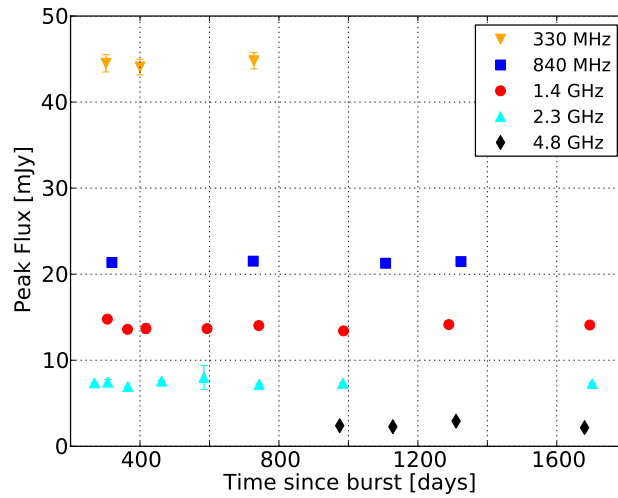


Figure 3.13: The light curve of Source E.

Source F. The large  $\eta_\nu$  at 2.3 and 4.8 GHz is explained by the higher resolution at these frequencies, where the source is resolved into two components. This can be seen in Fig. 3.14 where the GRB 030329 field is shown at comparable size for the 1.4 (left) and 4.8 GHz (right) images. The resolution of the latter is of the order of  $4''$ , three times higher than the former.



**Figure 3.14:** The left panel shows the GRB 030329 field at 1.4 GHz from Fig. 3.6, but now zoomed in to the search area of the 4.8 GHz primary beam. The right panel shows the 4.8 GHz observation at 2007, November 3rd. The dashed ellipses both represent the search area of the 4.8 GHz primary beam with a radius of  $300''$ . The resolution of  $4''$  at 4.8 GHz is three times higher than at 1.4 GHz, causing Source F to be resolved into two components. Source H was detected at 4.8 GHz, but not in the 1.4 GHz images.



**Figure 3.15:** The light curve of Source F. At 2.3 GHz and higher frequencies the source gets resolved into two components, of which only one was detected and could not be fitted well, leading to incorrect flux values.

Spreeuw (2010) reports the limiting deblending capacities of the current state of the Source Extraction software, making it difficult to fit multiple Gaussians simultaneously. This causes less accuracy on position and flux fitting, giving rise to the flux variations as seen for Source F. Its spectral light curve is shown in Fig. 3.15.

The same effect is present in the 1.4 GHz flux due to the fact that it is barely resolvable at this frequency causing small positional shifts that introduce variations in the fluxes. The variability at 1.4 GHz might be due to the complexity of the source being just unresolved along with poor  $(u, v)$  coverage.

It is clear that this type of sources, known to be unresolved in one and resolved in another band, should be flagged for further analysis to avoid false alerts, which is part of the work in progress.

**Source G.** Source G lies near the edge of the 2.3 GHz FWHM field. The flux variations are modest ( $V_{2.3\text{GHz}} \leq 0.1$ ), but two measurements deviate between 8 and 10% from the average. The spectral light curve of Source G is shown in Fig. 3.16. It is not unlikely that the fluctuations for a source at the edge of the FWHM arose from beam attenuation effects in combination with the poor  $(u, v)$  coverage of the early observations. The flux variations could be caused by instrumental effects. As before, the 1.4 GHz flux measurement at day 491 was influenced by the poor image quality, resulting in the same effect that caused the flux of Source E to drop. Excluding this single suspicious data point reduces the variability indices to  $V_{1.4\text{GHz}} = 0.019$  and  $\eta_{1.4\text{GHz}} = 2.20$  for  $N = 8$ , which are too low to reject  $H_0$ , making it not plausible to claim this source variable.

**Source H.** Source H was not detected at and below 1.4 GHz, but the higher resolutions and sensitivities at the higher frequencies make it detectable as can be seen in Fig. 3.14. The spectral light curve of Source H is shown in Fig. 3.17. At both frequencies the magnitude of the flux variability is relatively high at a significant level, giving probabilities  $p_\eta < 10^{-6}$ . These values make it plausible to reject  $H_0$  and classify this source as variable. No counterparts in the major radio and the FIRST and USNO catalogues were found, but the Sloan Digital Sky Survey catalogue has an optical candidate, SDSS J104457.09+213210.1 that is at a distance of 0.61 arcsec. According to the SDSS archive<sup>7</sup> this source has too few good detections to be classified and it has not been reported in the literature.

---

<sup>7</sup><http://cas.sdss.org/astrodr7/en/tools/explore/obj.asp?id=587742014884938668>

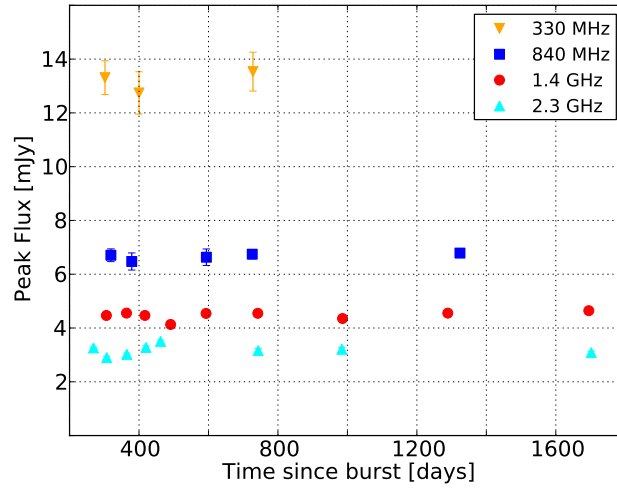


Figure 3.16: The light curve of Source G.

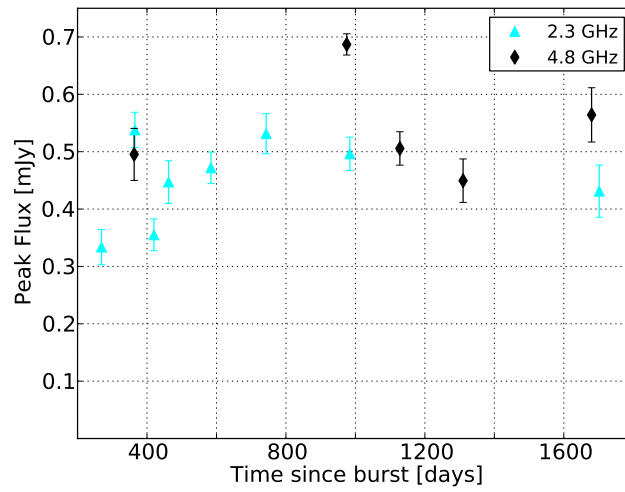


Figure 3.17: The light curve of Source H.

### 3.5 Discussion and Conclusions

The source association algorithm presented in Section 3.2.3 is based on a weighted positional difference between two sources and does not take into account any spectral source information, because there is no *a priori* knowledge of a source's spectral behaviour. The likelihood ratio, as defined by Eq. 3.5, on the other hand, is frequency (and rms noise) dependent with the use of the (local) source density  $n_L$ . The source density is determined by the catalogue properties, but existing extragalactic source-count models (e.g., Huynh et al., 2005) may predict the number of sources down to the sub-mJy level, and should then be scaled to the observing frequencies and the expected rms noise. Eventually LOFAR will produce these source count predictions from its own catalogue by usage of the Local and Global Sky Models. A more dynamic determination of  $n_L$  will further refine the criteria of the sources that are considered as an association. This is even more needed when the images are confusion limited. Restricting the number of candidate associations can also be done by a search radius  $r_s$  that depends on the resolution of the observation.

Continuous monitoring of the variability indices of the extracted and then associated sources will enable LOFAR to detect transients in nearly real time. As was shown in Section 3.4.2 this should definitely include more elaborate analysis of sources that have multiple components at higher frequencies in order to limit the number of false alerts. In principle, the source association is independent of the image resolution. In practice, the current source fitting routines have difficulty in deblending and fitting multiple Gaussians, giving rise to shifted positions and incorrect fluxes for sources near the resolution limit. This also means that the ideal criteria for the association parameters ( $r$  and  $\log LR$ ) as determined in Section 3.3.2 are too strict for the practical cases where we have images produced with different beam sizes,  $(u, v)$  coverage, resolution, pointing errors, etc. Again, if this is not handled properly, it may cause false alerts.

Furthermore, sources that lie in the outer regions or near the edges of the FWHM of the primary beam may vary due to pointing errors or individual antenna defects. These effects are small on average for the WSRT primary beam, but may be up to 5–10%. A way to overcome this is to do local calibrations by comparing similarities in the flux variations of neighbouring sources.

The sample of WSRT observations of GRB 030329 was processed in the TKP pipeline similar in the way the LOFAR images will be. GRB 030329 was detected as a transient source based on the (large) values of the variability indices. The light curve of Source A showed a burst-like rise and decay of its flux in two bands. Source H displayed significant variability in two bands. After closer inspection, Sources B, C, D, and E showed significant variability in only one band. At higher frequencies Source F was resolved

but fitting problems caused the variations in flux to have large contributions to the variability indices. Source G turned out to be suspicious by a bad data point from a low-quality image.

Source H is a good candidate for follow-up observations by LOFAR or other telescopes. A counterpart was detected in the Sloan Digital Sky Survey, where too few good detections could not classify it. The high resolution of the HBA station of the full array will enable a detection provided that the spectral index is less than zero for this source ( $S_\nu \propto \nu^{-\alpha}$ ). A redetection of the burst-like phenomena in Source A is possible if the spectral index is not too steep towards the higher frequencies.

The long-term WSRT observations of the sources from Table 3.3 revealed variability in about 25% of the well-detected sources in the bands at 1.4, 2.3 and 4.8 GHz; at 840 MHz this is about 10%. Although the sources and their counterparts are not classified as known types in any of the other catalogues, LOFAR observations can make a large contribution in doing so. The large fields of view of the full Dutch array (Full-NL; see Section 2.3.4 of Chapter 2) is able to reach in less than three hours of observing, resolutions of  $\lesssim 10''$  and sensitivities of 4 and 0.2 mJy at 60 and 150 MHz, respectively. This mode will be very suitable to yield more spectral information about the sources.

## Acknowledgements

We thank Tom Oosterloo for helpful discussions about the WSRT beam attenuation effects. The Westerbork Synthesis Radio Telescope is operated by ASTRON (Netherlands Foundation for Research in Astronomy) with support from the Netherlands Foundation for Scientific Research (NWO). This research was supported by NWO NOVA project 10.3.2.02 (BS) and by NWO VICI grant C.2320.0017 (AK) and by NWO VICI 639.043.302 (RAMJW). AJvdH was supported by an appointment to the NASA Post-doctoral Program at the MSFC, administered by Oak Ridge Associated Universities through a contract with NASA.