Searching for pulsars with LOFAR

Coenen, T.

Citation for published version (APA):

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, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER 5

AUTOMATING SEARCHES FOR SINGLE, DISPERSED PULSES

Code publicly available on Github.
https://github.com/tcoenen/brp
https://github.com/tcoenen/singlepulse-search

5.1 INTRODUCTION

Pulsar searches are generally conducted using two complementary techniques: i. the data are searched in the frequency domain using Fourier-based methods to identify the periodic signal and/or ii. the data is searched in the time domain for single bright pulses by using a set of matched-filters of varying lengths. Though the first pulsar ever discovered, PSR B1919+21, was first detected through its bright single pulses (Hewish et al., 1968), it soon became apparent that many weaker pulsars can only be detected through periodicity searches because these have the advantage of effectively integrating over many hundreds to thousands of rotational periods. Thus, until recently, there has been a bias towards using only periodicity searches and most pulsars have been found in this way.

McLaughlin et al. (2006) discovered what they coined the “Rotating Radio Transients” (RRATs). Many of the sources they found in searches of the Parkes Multibeam Survey data only displayed one or a few individual dispersed pulses per hour and were not
detectable in a Fourier Transform of the same data. Nonetheless, in most cases it was possible to derive the underlying periodicity of the pulses (and hence the rotation rate of the presumed neutron star source) by collecting enough closely spaced pulses that one could infer a greatest common denominator of the measured wait times. At the time the RRATs were discovered, they seemed like a distinctly new manifestation of radio-emitting neutron star whose radio emission is far more sporadic than canonical pulsars. Though it now seems that at least some of the RRATs are simply distant pulsars with broad pulse-energy distributions, in which we only observe the very brightest pulses (e.g. Weltevrede et al., 2006b), it still appears possible that some RRATs are genuinely intermittent emitters (e.g. related to extended nulling or mode-switching observed in some sources). Understanding their relation to the more steadily emitting radio pulsar population — in other words the full distribution of pulsar intermittency — may teach us something novel about the emission mechanism. Regardless, detecting the RRATs and characterizing their place within the spectrum of radio pulsar behavior is important for modeling the total Galactic population (Keane et al., 2011). After the discovery of RRATs, single-pulse searches again became a standard part of pulsar survey processing, and have made dozens of discoveries within the last few years. An updated catalog of RRATs, the so-called “RRATalog”\(^1\), is maintained by McLaughlin et al. and contains both the spin and burst properties of these sources as well as references to the surveys in which they were discovered.

Searching high-time-resolution pulsar data for single, dispersed radio pulses also has applications beyond the study of neutron stars. There are other known phenomena, such as solar bursts (McLean, 1985), planetary bursts and emission (e.g. from Jupiter and Saturn; Zarka (2004) and Zarka et al. 2004 Zarka et al. (2004)), and activity from flare stars which also produce sub-second radio pulses (Osten & Bastian, 2008). In general we term such sources “fast transients” because they are not easily detectable in the > 1 second integrations obtainable with standard interferometric imaging (though see Law et al., 2012; Law & Bower, 2012; Law et al., 2011). Lorimer et al. (2007) reported the detection of a very bright (∼ 30Jy) radio pulse in archival Parkes pulsar survey data of the Small Magellanic Cloud. Based on its high dispersion measure (DM; 375 cm\(^{-3}\)) pc), which cannot be accounted for by material within our galaxy and out to the Small Magellanic Cloud, the burst was proposed to be of extra-galactic origin, with an estimated distance ∼ 1Gpc. This discovery prompted new searches which yielded several more detections of this type (Keane et al., 2011; Thornton et al., 2013) (and the discovery of a previously unknown atmospheric phenomenon Burke-Spolaor et al. (2011a)). Speculation as to the origin of these bursts abounds (e.g. Keane et al., 2012), but they have so far all only been detected at one radio observatory (the Parkes

\(^1\)http://astro.phys.wvu.edu/rratalog/
telescope in Australia), despite concerted efforts with other telescopes. This is perhaps unsurprising given that the Parkes telescope is equipped with a 13-beam multi-beam receiver which gives the telescope an unmatched field-of-view (∼0.6 deg at 1.4GHz) for sensitive, high-time-resolution observing. Parkes is thus capable of acquiring a large product of field-of-view and on-sky time, and with a interesting sensitivity. Indeed, in addition to its traditional role as a unsurpassed pulsar discovery machine, the 64-m Parkes dish appears to also fill an interesting area of sensitivity/field-of-view parameter space for fast radio transients. A detection at another observatory would nonetheless be a very valuable confirmation of this phenomenon, and would probe other parts of the source populations (e.g., despite its small field-of-view, the 305-m Arecibo telescope is sensitive to much weaker fast transients than are detectable with Parkes). Furthermore, localization to about 1 arcsec precision is critical for multi-wavelength associations — e.g. associating the burst with a host galaxy. The current Parkes detections provide localization to within only ∼10 arcmin and they were only made via offline processing of the data many months after the event (also making it impossible to trigger optical and high-energy telescopes in a timely manner).

LOFAR’s large field-of-view, high sensitivity, and flexible beam-forming (see Stappers et al., 2011) is well suited to searches for rare fast radio transients (Figure 5.1). Certain types of fast transient searches are already being done commensally and in real time, see for example the AARTFAAC (Wijers, 2011) and FRATS (Ter Veen et al., 2011) projects allowing a large area of sky to be covered with a long effective dwell time. Although these projects have a large on-sky time they are constrained in sensitivity by the need to do real-time processing. This chapter is not about these new modes allowed by LOFAR, but about the searching classical pulsar survey data for single pulses. Since the data for these LOFAR pulsar surveys is largely processed off-line it is possible to search a finer grid of trial-DMs than is possible for on-line processing. The plan however, is to eventually perform the full DM grid search using an ARTEMIS-like system centrally (for ATEMIS see e.g. Armour et al., 2012).

Any standard single-pulse search starts by dedispersing the raw data to a large number of trial DMs. Each of these DMs is then searched with the aforementioned matched-filter technique and individual events above a certain statistical threshold (generally > 5 sigma) are saved. The cumulative list of events for all DM trials and for the entire observation length are then typically displayed in a diagnostic plot, an example of which is shown in Figure 5.2. For instance, by plotting trial DM versus observation time, and by marking all the saved events, it is possible to see clusters of events in DM-time space that are indicative of a dispersed burst. Unfortunately, multi-beam surveys like those conducted with LOFAR produce so many diagnostic plots that it becomes prohibitive to view them all by eye. An automated pulse detection algorithm is necessary to separate astrophysical pulses from Radio Frequency Interference (RFI)
CHAPTER 5. AUTOMATING SINGLE-PULSE SEARCHES

Figure 5.1: A comparison of the transient parameter space observable with Parkes and by using LOFAR in a number of possible configurations. By using different sets of LOFAR stations it is possible to balance the raw sensitivity versus the total field-of-view. In this way it’s possible to probe both for frequent, faint bursts as well as very rare, but bright bursts. These lines assume a spectral scaling of $-1.6$ and total observation times of a few hundred to a few thousand hours, which will be available in the coming years of observing.

and to reduce the amount of diagnostic plots that require further consideration. This also has the advantage of giving a uniform treatment to the candidate events that is free of human viewing bias. This chapter is about developing such code for the LOFAR pulsar searches described in Chapters 3 and 4.

In Section 5.2 we give a concise description of the software we used for the single-pulse searches we report in this thesis and present additional processing we performed (including automization of the search), in Section 5.3 we describe the various diagnostic plots we create for our data, in Section 5.4 we describe some of the detections we made and in Section 5.5 we describe some of the steps that can, or should, be taken to further increase the efficiency with which these searches are conducted.
5.2 SINGLE PULSE SEARCH TECHNIQUES

5.2.1 PRESTO

The PRESTO pulsar search software package\(^2\) (Ransom, 2001) contains a script, `single_pulse_search.py`, that searches dedispersed time-series for single pulses. This search is performed for each DM trial and yields a file of detected bright single pulses along with their signal-to-noise, arrival time, and duration. When all the dedispersed time-series have been searched this way, diagnostic plots can be created to check the data for interesting single pulses (see Figure 5.2).

Here we give a brief description of PRESTO’s single pulse search implementation. The search starts by de-trending the time-series in chunks of 1000 bins or when using the fast option the median is subtracted for each chunk. Next an effort is made to identify bad chunks of data by identifying chunks with very high or low standard deviations after removing the outliers in each chunk. The chunks identified as bad are subsequently ignored in further processing. The time-series is then normalized and searched for outlying values (candidate single-pulse detections). In order to do this, the time-series is convolved with a series of box-car-shaped kernels so that it can also be searched for longer duration pulses (those that span many raw data samples). Because the time-series is searched at several down-samplings there might be several detections at a particular time. Of these overlapping detections only the one with the highest signal-to-noise ratio is kept. All single-pulse candidates are then recorded in a text file, containing the pulse DM, signal-to-noise ratio, time of arrival, sample number and smoothing (down-sample factor). This process is then repeated resulting in a file of single-pulse candidates for each trial DM.

5.2.2 CANDIDATES, PULSES AND PULSE TRAINS

Given a Gaussian-shaped pulse, and assuming a rectangular bandpass shape, Cordes & McLaughlin (2003) give an equation for the flux $S(\delta DM)$ at which a pulsar will be observed $\delta$ DM away from the true peak DM of the pulse. The ratio of observed flux to true peak flux $S$ is

$$\frac{S(\delta DM)}{S} = \frac{\sqrt{\pi}}{2} \zeta^{-1} \text{erf}\zeta,$$

\(^2\)Software freely available here: https://github.com/scotransom/presto

103
Figure 5.2: An example of the diagnostic plots that PRESTO produces for single pulse searches. In this case the plot shows the detection of pulsar B0525+21 in LOTAS data (the survey presented in Chapter 4 of this thesis). This pulsar was not detected in a blind periodicity search of the same data. The bottom panel shows a range of trial DMs versus observation time. Each event has been marked with a circle whose radius is proportional to the signal-to-noise of that event. A particular pulse is a collection of events and may be visible over a range of DMs, though it will peak at the specific DM of the pulsar. Roughly 8 pulses are visible as clusters of events centered at DM ≈ 51 pc cm\(^{-3}\) in the DM-time plane. The top panels provide a histogram of the signal-to-noise of all events detected in the observation and show that there is an excess of events at DM ≈ 51 pc cm\(^{-3}\).
where
\[
\zeta = 6.91 \times 10^{-3} \delta \text{DM} \frac{\Delta \nu_{\text{MHz}}}{W_{\text{ms}} v_{\text{GHz}}^3}.
\]

In this second equation \( \Delta \nu_{\text{MHz}} \) is the total observing bandwidth and \( v_{\text{GHz}} \) the observing frequency. An astrophysical, dispersed, pulse will thus generally be detected across several trial DMs. PRESTO’s single-pulse search algorithm, however, works on a single trial DM at a time and therefore has no information about the signal-to-noise versus DM profiles. After all trial DMs have been processed, diagnostic plots are created and inspected by humans. These plots show the signal-to-noise as a function of DM so manual (as opposed to automated) inspection does use the information contained in the signal-to-noise versus DM profiles. This manual inspection step is no problem for small data sets but the number of plots generated during survey processing has the potential to make manual inspection very time consuming. Even a conservative estimate for a small survey like LOTAS with about 200 observations of 19 beams, assuming only ten diagnostic plots are made for each beam in the data, still produces about 40,000 diagnostic plots that need to be inspected. The fact that PRESTO does not exploit all information available to it suggests further gains in efficiency can be made and we therefore decided to try to automate this process further. Also automation allows one to apply a consistent approach to the entire data set.

We implemented an algorithm that takes the output of PRESTO’s single-pulse search, the files containing the single-pulse candidates, and associates them across DMs — in effect, extracting pulses from the data. We then further implemented several heuristics to try to identify interesting pulses in the data. The pulse extraction goes through a process of grouping candidate detections across DMs. This grouping starts at the lowest DM in the data sets, where each candidate is turned into a pulse object and moves on to higher DMs adding candidates overlapping in time to these pulse object (or creating new pulse objects if the candidates don’t match those at lower DMs). Early versions of this algorithm used a naive implementation where each candidate was given a small box on the time-DM plane and the grouping was performed by intersecting these boxes. This implementation kept an unsorted list of candidates for each pulse. The members of these lists were potentially all checked for intersections as new candidates were read slowing the processing down as the number of spikes across DMs grew in size. The current implementation tackles this problem by keeping a set of only those candidates that potentially intersect newly read candidates (so those from the last 4 trial DMs that were read). By keeping the number of intersection tests down the processing is sped up considerably, especially for data sets that contain broad interference spikes across many trial DMs (as those contain the most individual
CHAPTER 5. AUTOMATING SINGLE-PULSE SEARCHES

candidates). Section 5.3 contains part of the Python code that does the pulse extraction for reference. After the pulses have been extracted it is possible to build scripts to use characteristics of the pulses to select the interesting ones automatically.

To speed-up survey inspection we built a post-processing script that uses the algorithm described above along with a few simple heuristics to provide diagnostic plots only for potentially interesting detections. For each pulse extracted by the grouping algorithm the peak signal-to-noise ratio and the DM value at which it occurs is noted. After this step the post-processing script splits all extracted pulses into two categories; pulses brighter than a user-supplied signal-to-noise threshold that can trigger the creation of a diagnostic plot, henceforth bright pulses, and pulses below that threshold, henceforth dim pulses. Next, the bright pulses are sorted for descending peak signal-to-noise ratios. The brightest pulse is then removed from the bright pulses, along with any pulses occurring at roughly the same DM — the list of pulses extracted in this way is called a pulse train. This process is repeated until all bright pulses are assigned to a pulse train. Now a second user-supplied threshold, the minimum number of bright pulses per pulse train, is used to discard some of the pulse trains. For each pulse train, dim pulses also get added when they peak near the pulse train’s DM. For each pulse train that matches the required number of bright pulses a diagnostic plot is created (see Section 5.3.2).

5.2.3 DATA QUALITY ISSUES

Because radio data, especially early LOFAR data from the commissioning period (as used in this thesis), are often affected by instances of RFI, or in the case of early LOFAR data hardware and software issues affecting the data quality, any single-pulse search requires a robust strategy to deal with spurious detections. It was clear early on that for the LOFAR Pilot Pulsar Survey (see Chapter 3) a large fraction of the data was badly affected either by RFI or, more likely, by other issues with the telescope hardware and signal-processing chain that produce spurious single-pulse detections. The large number of single-pulse detections in bad stretches of data causes problems for PRESTO’s single-pulse search diagnostic plots. We created a more robust plotting script that can deal with a very large number of spurious detections (see Section 5.3.1). The diagnostic plots created by that new script can then be used to quickly identify bad data.
def group(spr, dms_adjacent):
    '''
    Drive the grouping algorithm by calling it for each consecutive trial DM.

    Args:
    spr: A SinglePulseReader (subclass) instance.
    dms_adjacent: An integer; the number of trial DMs below the current
                  DM to consider while extracting pulses from candidates.

    Returns:
    A list of extracted pulses.
    '''

    # Note: each Pulse instance has a 'head' against which matches can be
    # made and a 'body' that contains the rest of the candidates. When a
    # candidate is ready its head is empty and all its constituent candidates
    # are in its body. During the candidate matching candidates are removed
    # from the head if their DM is too low to still match (this is safe
    # because the grouping algorithm only moves up in DM). This trick keeps
    # the number of intersections down and makes it easy to recognize pulses
    # that are ready (since their heads are empty).

    pulses = []  # Pulses in the process of being extracted.
    done = []    # List of 'finished' i.e. extracted pulses.

    for dm in spr.dms:  # spr.dms is a sorted (small->big)
        # Find the appropriate amount of overlap in time for this trial DM.
        t_overlap = spr.get_t_overlap(dm)

        # Add this DM's candidates to the list of pulses. A candidate either
        # gets added to an existing pulse if a match is found or added as a
        # new pulse is no match was found.
        pulses = group_dm(pulses, spr.iterate_trial(dm), t_overlap, dms_adjacent)

        # Remove pulses from the list that can no longer match (because their
        # highest DM candidate is too low to still match any of the new
        # candidates, i.e. whose heads are empty).
        pulses, done = remove_done(pulses, done)

        # Finish 'in progress' pulses.
        for p in pulses:
            p.body.extend(list(p.head))
            p.head = set()

        done.extend(pulses)

    return done

Figure 5.3: The python code that is at the heart of the pulse extraction of Section 5.2.2.
During the processing of the LOFAR Pilot Pulsar Survey we started discarding stretches of single pulse search data where too many single pulse candidates were produced. We found the threshold by inspecting part of the data set and noting how many detections per 10 seconds were typical of data unaffected by RFI and how many were typical of bad data. The further processing could then be run on clean data.

The current implementation of the grouping algorithm described in Section 5.2.2 contains a limit (1,000,000) to the number of single-pulse candidates it will read before aborting. This limit is based on the number of pulses detected from the brightest pulsar detection in the LOTAS survey and observing the search script memory use deemed still acceptable on the LOFAR post-processing cluster. As such it is quite arbitrary, but prevents the processing from getting stuck whilst using large amounts of memory and making compute nodes unresponsive.

PRESTO also causes some issues for the post-processing scripts, because when it encounters bad sections in a time-series it will produce no output for them. This creates holes in the time-DM plane where the grouping algorithm operates. The edges of these holes quite often contain spurious single-pulse detections, see Figure 5.4 for a particularly bad instance in LOTAS data. Another issue that turned up in the data is that the de-dispersed time-series at adjacent DMs are not always lined-up in time correctly. At boundaries in the de-dispersion plan, i.e. where the PRESTO de-dispersion routine (mpiprepsubband) is called, there are often jumps that throw off the grouping algorithm. The current implementation of the candidate grouping algorithm contains a partial fix: it is possible to supply it with a list of delays indexed by trial DM that are added to the arrival time of each pulse.

5.3 Diagnostic plots

5.3.1 Condensed plots

PRESTO’s standard single-pulse diagnostic plots are a powerful way of looking for single dispersed pulses. However, when data contaminated by a large amount of RFI or containing a bright pulsar is plotted, the diagnostic plots tend to be overwhelmed by a large number of candidates. In that case the plots become unreadable and, being vector files, slow down the candidate inspection because they are slow to render. When processing a survey data set this is a problem.

To be able to identify bad single-pulse search data more quickly we created a condensed version of the single-pulse diagnostic plots. The version we created shows a color-coded 2D-histogram of single-pulse detection counts on the time-DM plane.
Figure 5.4: Example condensed single-pulse diagnostic plot containing RFI, showing the effect of PRESTO’s removal of bad stretches of data. This data set is part of the LOTAS survey. The main, DM-time panel gives a color coding of the single pulse candidate density in each pixel. As DM Index increases, the time-series effectively become shorter because the dispersive delay across the band is larger and more data must be discarded when the band is summed. The additional panels show the DM-time plane collapsed along both axes, as well as how DM Index maps to the true DM in units of pc cm$^{-3}$.
diagnostics remain readable when a large amount of RFI is present and are quick to render in a viewer, thereby not slowing down data inspection. Care is also taken during the creation of these plots not to read the whole data set into memory at once.

5.3.2 SINGLE PULSE DETECTION DIAGNOSTIC PLOTS

The single-pulse search post-processing script produces diagnostic plots for each pulse train that matches all criteria for an interesting detection (as described in Section 5.2.2). These plots are designed for quick viewing so as not to slow down inspection of the survey results. The main panel of these plots shows the candidates on the time-DM plane, with two side panels to the right showing respectively a SNR-DM scatter plot and histogram showing the number of candidates per trial DM. To provide some context to an interesting detection not just the candidates comprising the pulse train are shown but also part of the detections on the surrounding time-DM plane. The candidates that are part of the pulse train are drawn in a contrasting color (red). The diagnostic plots further contain text reporting the settings used when running the search script.

5.3.3 THE brp PLOTTING LIBRARY

A by-product of this thesis’ work on diagnostic plots for the single-pulse search is the Browser Plot Library (brp\(^3\)). It is a small Python library that can be used to create 2D plots in Scalable Vector Graphics (SVG). The library’s requirements were driven by the single-pulse search diagnostic plots mentioned in the previous sections and as such is not meant as a replacement for more general Python plotting packages like Matplotlib. Because SVG supports hyper-links it is possible to interlink many diagnostic plots in such a way that a large data set can be viewed quickly using a web browser.

5.3.4 LPPS 7 BEAM SINGLE PULSE DIAGNOSTIC PLOTS

During the follow-up of single-pulse candidates identified in the LPPS data it became apparent that checking the other beams of an observation was useful. When a candidate is present in all beams, or connected to some non-astrophysical signal in

\(^3\)Available from https://github.com/tcoenen/brp
one of the other beams it can be rejected. We created a script to make single-pulse diagnostic plots that covered several beams in one go as is illustrated by Figure 5.5. While this type of plots are useful in quickly identifying spurious signals when they are present in several beams, further automation is useful and is used by several other pulsar surveys such as PALFA (Deneva et al., 2009, Fig. 5) and HTRU (Burke-Spolaor et al., 2011b, Fig. 2).

5.4 ILLUSTRATIVE RESULTS

5.4.1 LPPS DETECTION OF B0154+61

During the processing of the LPPS survey we used an earlier version of the single-pulse search post-processing script, combined with thresholding to remove obviously bad data, to dramatically cut back the amount of time spent inspecting the survey results and to give a uniform treatment to the data set. We made standard PRESTO single-pulse search diagnostic plots, but they numbered in the tens of thousands and did not show enough context to quickly go through them. The post-processing script, when run with a signal-to-noise threshold of 8 and no requirements on the number of repeating bursts at the same DM produced on the order of a thousand plots (an order of magnitude fewer) with more context for each candidate detection. The majority of these plots were obviously RFI and could be quickly rejected. A single pulsar re-detection stands out, that of B0154+61. It was detected in LPPS only through its single pulse emission — emitting only 1 pulse that survived the thresholding, yet was picked up by our post-processing script (see Figure 5.6). A quick check that this pulse showed up in only one of the beams for that observation, a check of the ATNF catalog (Manchester et al., 2005) and inspection of the time-series confirmed the re-detection.

5.4.2 LOTAS SINGLE-PULSE SEARCHES

The second LOFAR pulsar commissioning survey, LOTAS (see Chapter 4), was also searched for single, dispersed pulses. Data processing of this survey was also based on the PRESTO software suite and automated with a customized search pipeline written in Python. While a single-pulse search was run on each trial-DM, no diagnostic plots were created to prevent problems caused by strong RFI, which could possibly affect some of the observations. After the initial search, we created condensed
Figure 5.5: An LPPS 7-beam plot showing a detection of the HAMSAT satellite. This satellite is used by radio amateurs and produces intrinsically frequency-swept signals that mimic dispersed signals at DMs of 13 and 16 pc cm\(^{-3}\). The non-astrophysical source of the signal is immediately apparent by its presence in all 7 beams at once.
Figure 5.6: Diagnostic plot created by a recent version of the single-pulse search post-processing script showing a single bright pulse from PSR B0154+61 detected in the LPPS data.

single-pulse diagnostic plots and inspected them to get a feel for the data quality. This inspection yielded the re-detection of 8 pulsars that were clearly visible even in condensed plots, but it also became evident there were misalignments in time of blocks of trial-DMs Figure 4.4 in Chapter 4 shows a clear example of the misalignment problem, it must be stressed that this figure shows a badly affected data set and that most LOTAS data is much cleaner.

After an initial pass over the LOTAS data, we ran our single-pulse extraction script on the entire LOTAS data set, much as we did for LPPS. After the LPPS processing, the extraction script had been much improved. The implementation of the grouping algorithm, as described in Section 5.2.2, ran much faster than the one used for LPPS, and consequently was able to process data sets with more RFI than the earlier version. We therefore did not attempt to remove the RFI-affected stretches of data. Our single-pulse extraction script produced a very large number of diagnostic plots. Inspecting all of them was not realistic given the time constraints and furthermore the fact that some beams produced roughly one thousand plots. This indicated that the handling of RFI in the single-pulse extraction script needed to be improved. Because some data sets created very large plot files, which were impossible to quickly inspect in a web browser, we implemented a rasterized fallback for the single-pulse extraction diagnostics. The RFI excision in the single-pulse extraction script also needed to be improved. We had planned to implement a simple heuristic that could detect RFI-affected data by checking for bright (in the sense of many single-pulse candidates) streaks across many
CHAPTER 5. AUTOMATING SINGLE-PULSE SEARCHES

DMs. This turned out to be problematic as the LOTAS data was affected by the
described misalignment in time (likely caused by a problem in the program
used to dedisperse the data). The automated search through the LOTAS data did not
yield any extra single-pulse (re-)detections.

Aside from its use in blind searches of large data sets, the single-pulse extraction script
is also a convenient way of extracting known sources from the data. In this mode we
were able to extract all pulsars detected by the quick inspection of the condensed plots.

Figure 5.7 shows an example of PSR B1831+05 extracted using our script.

5.5 FUTURE WORK

The efficiency and sensitivity of our single pulse searches can be further increased.
The first step in the single pulse search, that currently performed by PRESTO, only
uses box-car shaped kernels during the smoothing steps. For LOFAR in particular
that means that highly scattered pulses could be missed. By changing the shape of the
kernels to take into account scattering we should be able to increase our sensitivity
to scattered single pulses. This change could be made without needing to change
the post-processing scripts. The post-processing script currently does not attempt to
differentiate between astro-physical pulses and RFI-caused pulses by their signal-to-
5.5. Future work

noise versus DM shape or pulse width (available as the down-sample factors of the individual candidates). This technique was used by Rubio-Herrera et al. (2013) in their single pulse searches using the WSRT. No attempt is currently made to discover noise-spikes in the data that cause many spurious detections across many trial DMs. In the LOTAS survey it is clear that these noise spikes are the major reason that diagnostic plots are made. By rejecting pulses as non-astrophysical and by detecting and rejecting noise spikes it should be possible to cut back on the number of diagnostic plots (and therefore time spent inspecting them) by an order of magnitude if newer LOFAR data suffers from some of the same problems as the LOTAS data does.