

Single-pulse classifier for the LOFAR Tied-Array All-sky Survey

D. Michilli,^{1,2★} J. W. T. Hessels,^{1,2} R. J. Lyon,⁴ C. M. Tan,⁴ C. Bassa,² S. Cooper,⁴
V. I. Kondratiev,^{2,3} S. Sanidas,^{1,4} B. W. Stappers⁴ and J. van Leeuwen^{1,2}

¹Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, NL-1098 XH Amsterdam, The Netherlands

²ASTRON, the Netherlands Institute for Radio Astronomy, Postbus 2, NL-7990 AA, Dwingeloo, The Netherlands

³Astro Space Centre, Lebedev Physical Institute, Russian Academy of Sciences, Profsoyuznaya Str. 84/32, Moscow 117997, Russia

⁴Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK

Accepted 2018 July 27. Received 2018 July 10; in original form 2018 January 23

ABSTRACT

Searches for millisecond-duration, dispersed single pulses have become a standard tool used during radio pulsar surveys in the last decade. They have enabled the discovery of two new classes of sources: rotating radio transients and fast radio bursts. However, we are now in a regime where the sensitivity to single pulses in radio surveys is often limited more by the strong background of radio frequency interference (RFI, which can greatly increase the false-positive rate) than by the sensitivity of the telescope itself. To mitigate this problem, we introduce the Single-pulse Searcher (SPS). This is a new machine-learning classifier designed to identify astrophysical signals in a strong RFI environment, and optimized to process the large data volumes produced by the new generation of aperture array telescopes. It has been specifically developed for the LOFAR Tied-Array All-Sky Survey (LOTAAS), an ongoing survey for pulsars and fast radio transients in the northern hemisphere. During its development, SPS discovered seven new pulsars and blindly identified ~ 80 known sources. The modular design of the software offers the possibility to easily adapt it to other studies with different instruments and characteristics. Indeed, SPS has already been used in other projects, e.g. to identify pulses from the fast radio burst source FRB 121102. The software development is complete and SPS is now being used to re-process all LOTAAS data collected to date.

Key words: pulsars: general – surveys – methods: data analysis – methods: statistical.

1 INTRODUCTION

The first pulsar was discovered by recording its single pulses at ~ 80 MHz using an aperture array (Hewish et al. 1968). In later studies, folding and Fourier-based techniques were used to take advantage of the pulsar periodicity. Many pulsar observations shifted to higher observing frequencies around 1.4 GHz, where the separation between pulsar signals and sky brightness is maximum for most of the pulsar population (Clifton & Lyne 1986). Furthermore, phased arrays were generally replaced by large single dishes, which remove the complexity of signal correlation permitting an increase in telescope sensitivity and bandwidth (Garrett 2012). However, in recent years the increase in available computing power makes it possible to build phased aperture array telescopes with sensitivities and bandwidths that outperform traditional single dishes at low radio frequencies, offering a larger field-of-view (FoV) and more flexible instruments (Taylor et al. 2012; Tingay et al. 2013; van Haarlem et al. 2013). This enables an exploration of a parameter

space that is complementary to other searches: e.g. it is possible to detect sources having a spectrum steeper than the sky background (spectral index $\alpha \sim -2.5$; Mozdzen et al. 2017), which are likely too faint to be detected at higher frequencies. In addition, the larger FoV improves survey speed, and makes all-sky searches tractable.

The most sensitive phased aperture array telescope to date is the LOw Frequency ARray (LOFAR; Stappers et al. 2011; van Haarlem et al. 2013). For example, its large collecting area already enabled detailed studies of archetypal sources, and the properties of the known pulsar population. The former is exemplified by the discovery of radio/X-ray mode switching in PSR B0943+10 (Hermsen et al. 2013). Examples of the latter are the pulsar census results presented by Kondratiev et al. (2016) and Bilous et al. (2016). Beyond these known sources, many pulsars in our Galaxy remain undetected. It was recognized early on that LOFAR has great potential for discovering these (van Leeuwen & Stappers 2010). Pilot surveys placed limits on the occurrence of fast transients at low frequencies, and discovered the first two radio pulsars with LOFAR (Coenen et al. 2014). We are now performing a full, sensitive survey of the northern hemisphere called the LOFAR Tied-Array All-Sky Survey

* E-mail: danielmichilli@gmail.com

(LOTAAS; Coenen et al. 2014; Sanidas et al. 2018)¹. LOTAAS has already demonstrated its ability to find new pulsars using periodicity searches (with over 80 discoveries to date) and in this paper we focus on discoveries made through single pulses. The long dwell time (1 h) and large FoV (~ 10 sq deg per pointing) of LOTAAS also make the survey potentially well placed to discover fast radio transients, as long as they are not strongly affected by scattering or dispersive smearing.

1.1 Signal classification

An increasing issue for pulsar surveys is the presence of radio-frequency interference (RFI) produced by several devices, which can mimic the behaviour of astrophysical signals and limit survey sensitivities (e.g. Lyon et al. 2016). The large number of RFI detections makes it impractical to visually inspect and follow up all the detected signals. This is a worsening problem caused by the increasing number of devices emitting radio waves and the improvements in telescope characteristics, such as sensitivity, dwell time, and bandwidth. Therefore, this is a major challenge for next-generation radio telescopes and in particular the Square Kilometre Array (Ellingson 2004). An improvement is obtained by building telescopes in RFI-free zones, areas where the human presence is minimal. However, the radio emissions of air planes and satellites are still present.

In order to lower the number of detections to be inspected by eye, many automated classifiers have been developed for pulsar surveys (see Lyon et al. 2016 and references therein for a summary). These automatic classifiers evolved from simple heuristics and thresholds on S/N (e.g. Clifton & Lyne 1986) to semi-automated ranking algorithms (e.g. Lee et al. 2013). Also the graphical representation of the detected signals evolved, visualizing increasing information describing their parameters (e.g. Burgay et al. 2006). Machine learning (ML) techniques began to be used to evaluate heuristic performance, set threshold values, and perform the classification (e.g. Eatough et al. 2010). Most recently, significant efforts are being spent in the selection of pulsar signals to deal with the large number of detections from new, sensitive radio telescopes (e.g. Yao, Xin & Guo 2016; Ford 2017; Bethapudi & Desai 2017).

Here a branch of ML called statistical classification (Mitchell 1997) is used to filter to select astrophysical signals. This requires first acquiring a large set of candidate examples for which the ground truth origin or ‘class’ is known. When there are two classes under consideration, the classification task is termed ‘binary’. In binary problems the targets, i.e. astrophysical signals, belong to the positive class. The negative class describes all other examples (e.g. RFI or noise). In either case, the examples must be characterized via one or more variables commonly known as ‘features’. Features are numerical or textual descriptors that summarize a candidate in some relevant way (e.g. S/N, pulse width, etc.). Features must be extracted by algorithms for each candidate, and linked to their true class ‘labels’. When combined, this information forms what is known as a ‘training set’. Using supervised learning, it is possible to ‘learn’ a mathematical function from this training set that can automatically perform a similar mapping on new data. This process is known as ‘training’. The training process aims to split the training data into their respective classes, by using the inherent differences in the feature distributions to separate them. The learning process is guided via a heuristic, most often error minimization that quantifies how

many errors are made. Each correct positive classification is known as a true positive (TP), while an incorrect positive classification is known as a false positive (FP). Similarly, negative classifications can be described in terms of true-negative (TN) and false-negative (FN) predictions. Together, these four outcomes form the so-called confusion matrix, used to assess how successfully an algorithm has learned the mapping.

For a good classification, it is essential to have features that permit to separate the data into positive and negative classes. Therefore, specific algorithms must be designed to extract such features. Moreover, these algorithms need to be fast and efficient in order to keep the computing time low. These algorithms can be designed by looking at the different properties of the members of the positive and negative classes. The classification success is usually determined during a ‘testing’ phase, conducted on an independent sample of candidate examples. If the model learned during training performs well during testing (produces few FPs and FNs), it can be used to derive predictions for new previously unseen data. An algorithm will usually be successful if trained on a large representative sample of data. Different heuristics can be used to evaluate single features, such as the information gain (also known as mutual information; Brown et al. 2012), a measure of the correlation between a feature and the target variable (Lyon et al. 2016). Also, several metrics exist to evaluate the performance of the whole set of features in classifying the data. In this study, we made use of standard metrics such as the false-negative rate (FNR) and the false-positive rate (FPR), which must be as low as possible for a good classification, the true positive rate (TPR, also known as recall), the positive predictive value (PPV, also known as precision), the accuracy (ACC), the G-Measure (G-M, the geometric mean of recall and precision), and the F-Measure (F-M, the harmonic mean of recall and precision), which must be as high as possible for a good classification (e.g. Powers 2011). All these metrics assume values between 0 and 1.

1.2 Single-pulse searches

Soon after the initial discovery of pulsars, surveys began taking advantage of the inherent periodicity of pulsar signals to improve search sensitivity (Lorimer & Kramer 2004). This is usually achieved by performing a Fourier transform of the time series. However, this technique greatly decreases the sensitivity to sources whose emission is not regular over time (McLaughlin et al. 2006). Therefore, it has become standard procedure to include single-pulse searches in pulsar surveys to avoid missing sources with large pulse amplitude variations or a large null fraction (Lorimer & Kramer 2004). Moreover, two new classes of sources discovered in recent years have created new interest in single-pulse searches: rotating radio transients (RRATs; McLaughlin et al. 2006) and fast radio bursts (FRBs; Lorimer et al. 2007). The former class is composed of pulsars whose emission is so sporadic in time that they are missed by periodicity searches. Typical RRAT pulse rates range from as many as 1 per few seconds, up to one per several hours². The FRB class (Thornton et al. 2013) is composed of extra-galactic radio flashes (Tendulkar et al. 2017). To date, only one has been observed to repeat (Spitler et al. 2016) and no periodicity has been detected (Scholz et al. 2016).

A typical search for single pulses is performed after de-dispersing the data collected by the telescope. This aims to correct for the frequency-dependent delay induced by the interaction of the radio

¹<http://www.astron.nl/lotaaas>

²<http://astro.phys.wvu.edu/rratalog>

waves from the source with free electrons present along the line of sight. The amount of dispersive delay exhibited by a signal is proportional to the dispersion measure (DM, which is the column density of free electrons). Since the DM of the source is unknown *a priori*, it is necessary to de-disperse the signal at many trial values. The time series resulting from the addition of all frequency channels is then searched for single pulses. This is usually achieved by convolving each de-dispersed time series with a top-hat function of variable width W . The properties of the function and the convolution are recorded every time the resulting signal-to-noise ratio (S/N) is above a certain threshold (see Lorimer & Kramer 2004 for a detailed discussion).

1.3 LOTAAS Single-pulse Searcher

The number of signal detections arising from RFI is particularly large for LOTAAS because the LOFAR Core is in a region of high population density. In addition, the high sensitivity and long dwell time offered by the survey and the large parameter space necessary to be searched at these low frequencies increase the number of false detections. For these reasons, an automated classifier has been developed to classify the periodic signals of LOTAAS (Lyon et al. 2016; Tan et al. 2018). It uses the advantages of statistical classification to build a solid statistical framework for rejecting RFI.

The presence of RFI is particularly problematic for single-pulse searches. In fact, as opposed to periodicity searches, it is not possible here to filter out aperiodic signals. Usually, single-pulse classifiers take advantage of pulse shape in the frequency-time domain where, as opposed to RFI, astrophysical signals are expected to typically be broadband and dispersed (Lorimer & Kramer 2004). As opposed to periodicity searches discussed earlier, only a few classifiers have been reported to specifically sift single-pulse detections. Keane et al. (2010) subtracted the non-dispersed signal from their data (Eatough, Keane & Lyne 2009) and used spatial information from a multibeam survey to discover 10 new RRATs. Spitler et al. (2012) developed a multimoment technique useful for quantifying the deviation present in the pulse intensity at different frequencies. Karako-Argaman et al. (2015) presented RRATRAP, which uses manually set thresholds to discriminate astrophysical signals based on their S/N behaviour as a function of DM. Similarly, Deneva et al. (2016) developed CLUSTERRANK, which rejects RFI instances that deviate from the theoretical relation between signal strength, width, and DM (Cordes & McLaughlin 2003). Devine, Goseva-Popstojanova & McLaughlin (2016) studied different ML algorithms applied to their single-pulse classifier. For binary classification, a Random Forest (RF) trained on an oversampled set (RF_{over}^2) performed the best in their case.

Here we present a new single-pulse classifier, Single-pulse Searcher (SPS, Michilli & Hessels 2018), which is able to discriminate astrophysical signals from RFI with high-speed and accuracy. In its current implementation (LOTAAS Single-pulse Searcher, L-SPS), it has been specifically designed to process LOTAAS data. LOTAAS observations and data processing are presented in Section 2; the SPS classifier is presented in Section 3; first discoveries are presented in Section 4; and conclusions and future developments are discussed in Section 5.

2 OBSERVATIONS

LOFAR is a radio telescope composed of thousands of antennas, which are grouped into stations that are distributed across the Netherlands and other European countries (Stappers et al. 2011; van Haarlem et al. 2013). LOTAAS observations (Coenen et al.

2014) are performed using the 12 sub-stations of the Superterp, a circular area with ~ 350 -m diameter where the station concentration is highest. The High-Band Antennas (HBAs) are used to observe between 119 and 151 MHz. Signals from different stations are added both coherently ('tied-array' beams) and incoherently. The beams are divided over three sub-array pointings (SAPs, which are pointing directions formed at station level). A total of 222 simultaneous beams on sky are produced per pointing: 3 incoherent beams, a hexagonal grid of 61×3 coherent beams, and 12×3 additional coherent beams that are scattered around the central grid, and which can be pointed to known sources within the FoV. The incoherent beams have a total FoV of $\sim 30 \text{ deg}^2$, while the coherent beams have a total FoV of $\sim 10 \text{ deg}^2$ and a sensitivity $\sim \sqrt{12} = 3.5$ times higher than that of the incoherent beams. Each LOTAAS pointing has a 1-h dwell time and a time resolution of 0.492 ms. The whole survey is divided into three passes and, upon survey completion, each sky position will be covered three times by the incoherent beams and once by a coherent beam.

Data are processed using a pipeline based on PRESTO (Ransom 2001)³ described in detail by Sanidas et al. (2018), which runs on the Dutch national supercomputer Cartesius⁴. After using the RFIIND algorithm to filter RFI in the time-frequency domain, the signal is incoherently de-dispersed and frequency channels are added together using MPIPRESUBBAND in order to form time-series data. Each time-series is searched for both periodic signals and single-pulse peaks. The latter is performed using SINGLE_PULSE_SEARCH.PY, which convolves box-car functions of various lengths between 0.5 and 100 ms. Single-pulse peaks with an S/N higher than 5 are stored for further grouping and sifting.

A total of $\sim 10^4$ DM trials are performed between DM values of $0\text{--}550 \text{ pc cm}^{-3}$ with steps between DM trials of $0.01\text{--}0.1 \text{ pc cm}^{-3}$. Given a total of $\sim 7 \times 10^6$ time samples per time series, this implies a grid in the DM–time space with 7×10^{10} pixels, each one potentially containing an astrophysical signal, for each of the 222 beams in every pointing. Considering only ideal random noise, this implies an expected number of spurious detections ($\geq 5\sigma$) for each observation of ~ 45 (Cordes & McLaughlin 2003). Instead, a typical observation produces $\sim 10^7\text{--}10^8$ detections above 5σ , demonstrating the huge impact that RFI and non-stationary noise have on the single-pulse search. For comparison, the number of detections generated above 6σ is roughly an order of magnitude less and those above 7σ are normally half of that. An automated algorithm capable of sifting these detections to identify the rare astrophysical signals is thus necessary.

3 LOTAAS SINGLE-PULSE SEARCHER

L-SPS (Michilli & Hessels 2018) is a new sifting algorithm designed for single-pulse searches in a strong-RFI environment, and specifically designed for the LOTAAS survey. The software uses ML techniques to differentiate RFI from astrophysical signals, classify interesting signals, and produce diagnostic plots.

3.1 Algorithm operation

The aim of the program is to produce diagnostic plots for the very best detections in a typical pulsar single-pulse search. Three steps are necessary to achieve this result.

³<https://www.cv.nrao.edu/~sransom/presto/>

⁴<https://userinfo.surfsara.nl/systems/cartesius>

3.1.1 Events

The script `SINGLE_PULSE_SEARCH.PY` included in `PRESTO` outputs information on each detected signal, specifically the DM, the time with respect to the beginning of the time-series and the selected width of the kernel function. These values are stored for each signal as one line in a text file. We define one of such lines as an *event*. The `PRESTO`-based pipeline is run for every LOTAAS pointing. For each beam, `L-SPS` copies the events detected into the memory of the computer and stores them into a readily accessible HDF5 database⁵.

An impulsive signal having a large enough S/N, either RFI or astrophysical, will be detected in multiple time-series de-dispersed at nearby values. Therefore it will produce a number of events close in time and DM. A broadband signal exhibits a decreasing S/N as the trial DM used moves further from the actual DM value because the pulse becomes increasingly smeared in time. This smearing is not symmetric and advances or delays the events for DM values higher or lower than the actual value, respectively. Therefore, the events from a broadband signal will lie on a line in the DM–time space whose slope is given by

$$\frac{\Delta t}{\Delta \text{DM}} = \frac{k}{2} (\nu_m^{-2} - \nu_M^{-2}), \quad (1)$$

where $k = (4148.808 \pm 0.003) \text{ MHz}^2 \text{ pc}^{-1} \text{ cm}^3 \text{ s}$ is the dispersion constant (Lorimer & Kramer 2004). This equation is half the value of the DM delay between the minimum (ν_m) and maximum (ν_M) observing frequencies. In `L-SPS`, the slope defined by equation (1) is corrected so that events belonging to the same broadband signal are simultaneous.

3.1.2 Pulses

All events close in time and DM are grouped together to form a *pulse*. The grouping is performed by a friends-of-friends algorithm (e.g. Huchra & Geller 1982; Press & Davis 1982) designed to be highly efficient and able to process one million events in approximately half a second. Such an algorithm starts from the first event in the list and labels the closest other event in the DM–time space to be part of the same pulse if they are within a certain range. The thresholds on time and DM ranges between successive events were set empirically to 30 ms and 20 DM trials (i.e. between 0.2 and 2 pc cm^{-3} depending on the DM trial step size), respectively. The algorithm subsequently processes all the events with the same conditions. The characteristics of each pulse (i.e. time of arrival, DM, width, S/N) are derived from the brightest event forming that pulse. Pulses formed by fewer than five events are considered spurious or too weak for subsequent analysis and therefore removed. Weak narrow-band signals, which produce pulses with constant S/N, are mitigated by removing pulses with $\text{S/N} < 6.5$. Also, pulses with $\text{DM} < 3 \text{ pc cm}^{-3}$ are removed to avoid the contamination of broadband RFI near $\text{DM} = 0 \text{ pc cm}^{-3}$. Tests are ongoing to try to lower this threshold on the DM. ML classification is then applied to the pulses in order to discriminate RFI from astrophysical signals, as discussed in Section 3.2.

After the pulses labelled as RFI by the ML classifier are removed, those positively classified are further filtered based on spatial information. Given the sensitivity pattern of the LOTAAS beams, astrophysical signals will have a maximum S/N in the beam closest to the actual source position and decreasing S/N values in beams farther away. Given the complicated pattern of the side lobes, weaker

detections can be expected in distant beams for bright signals. For each pulse detected in a coherent beam, all the events in the other non-adjacent coherent beams of the same SAP are loaded if they are in a time window four times the pulse width and a DM window of 0.4 pc cm^{-3} around the pulse. Also, these events must have an S/N larger than half the pulse’s S/N in order to only remove signals with a nearly constant strength over many beams. If more than four additional beams contain events selected with those criteria, then the pulse is discarded. The spatial comparison is computationally expensive because of the many events to load and select. Therefore, it is implemented after the ML classifier has already removed a large fraction of the pulses.

3.1.3 Candidates

In every beam, positively classified pulses (occurring at different times) are grouped into *candidates* according to their proximity in DM. The maximum DM spread over which two pulses are considered to come from the same source is empirically set to 0.3 pc cm^{-3} , corresponding to 3–30 DM trials. Candidates in different beams within this DM range are considered to be produced by the same signal and only the brightest is retained. The characteristics of each candidate are computed, i.e. beam number, average DM, cumulative S/N, the number of pulses detected, and time of arrival if only one pulse is present, otherwise the period is calculated using the `PRESTO` routine `RRAT_PERIOD`. Candidates formed by a single pulse are considered only if they satisfy $\text{S/N} > 10$ since weaker candidates would not be visible in the dynamic spectrum, meaning that their astrophysical nature cannot be verified. For the same reason, candidates formed by multiple pulses must have a cumulative $\text{S/N} > 16$.

A maximum of 10 bright candidates (i.e. candidates with the largest cumulative S/N) are processed per beam. Typically, a lower number of candidates per beam is produced, as discussed in Section 3.3. This limit is set empirically to avoid processing too many spurious candidates, particularly in beams polluted by RFI. For the same reason, a processing limit is set by considering only the 50 brightest candidates per observation. As indicated in Lyon et al. (2016), this is not an ideal procedure. However, typical observations produce a factor of ~ 3 less candidates. Diagnostic plots are then generated for every selected candidate, as discussed in Section 3.4.

3.2 Machine learning classifier

Astrophysical signals of interest are different from RFI in that they are usually broadband, dispersed, and localized on the sky, while normally the latter is either narrow-band or not dispersed and often detected in multiple beams because the source is local. This implies that astrophysical signals produce pulses that peak in S/N in a certain beam and at a certain $\text{DM} > 0 \text{ pc cm}^{-3}$, while RFI will often have similar S/N in all the beams and constant S/N at different DM trials if narrow-band, or peaked at $\text{DM} = 0 \text{ pc cm}^{-3}$ if it is broadband. However, the real-world situation is often more complicated because of statistical and non-stationary noise (which tends to mask these differences especially for weak signals) as well as artefacts introduced by the telescope. Also, RFI can mimic a dispersive delay (e.g. Petroff et al. 2015), for example when multiple signals occur simultaneously in the de-dispersed data. An astrophysical signal can also be masked by simultaneous RFI. In addition, both RFI and astrophysical signals can have complex structures that complicate their S/N curve as a function of DM. Finally, some sources

⁵<https://www.hdfgroup.org>

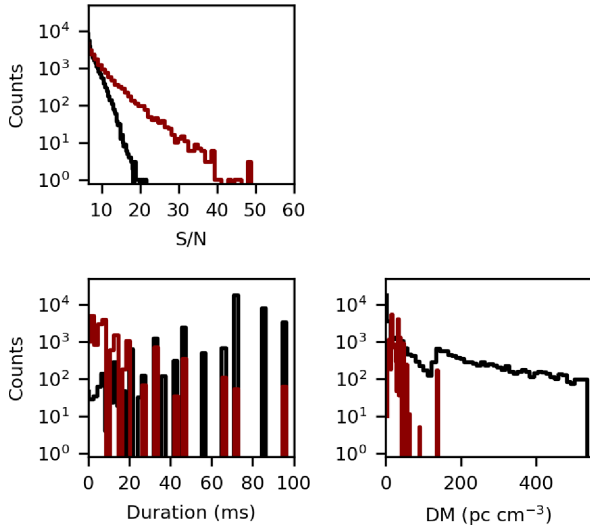


Figure 1. Comparison of the duration, S/N, and DM of the RFI (black lines) and astrophysical (red lines) instances used in the LOTAAS training set.

of interference like aeroplanes, radars, and signals reflected by the ionosphere can be beamed and thus appear localized on the sky.

We make use of ML techniques in order to effectively differentiate RFI and astrophysical signals and to have a statistical foundation to evaluate the performance of our features. The statistical algorithm chosen to build the model and perform the classification is the Gaussian-Hellinger Very Fast Decision Tree (Lyon et al. 2014), a tree learning algorithm (Mitchell 1997) based on the Very Fast Decision tree (VFDT; Hulten, Spencer & Domingos 2001). The algorithm is designed to deal with imbalanced problems, where the target class, in our case astrophysical single-pulse events, are outnumbered by the instances we wish to reject (Lyon et al. 2016). The data-mining tool WEKA⁶ was used to run this algorithm in order to evaluate classification performance, to build a classification model, and to perform the classification of new instances. The other algorithms included in the standard WEKA distribution have been tested using the training set described in the next subsection but none clearly outperformed the VFDT.

3.2.1 Building of the training set

A set of manually labelled instances was selected to evaluate the features used and to build the classification model. A total of 35 063 instances of RFI were randomly chosen from various LOTAAS observations where no astrophysical sources could be identified by eye. In addition, 18 003 pulses were chosen from 47 known pulsars selected in an equal number of beams. All the pulses in these beams were included in the training set if they were detected at the pulsar DM and were within 0.1 per cent of the expected time of arrival given the pulsar period. The same conditions applied to random DM and period values, in beams without any known pulsar, usually yielded no pulses to be selected. Given the costs associated with labelling data, the possibility of human error, and because the ground truth labels can only be confirmed via re-observing, it is possible for some training examples to be incorrectly labelled. Though this is unlikely, a small proportion of training samples may be affected in this way. Such mislabelled examples, whilst undesirable, have

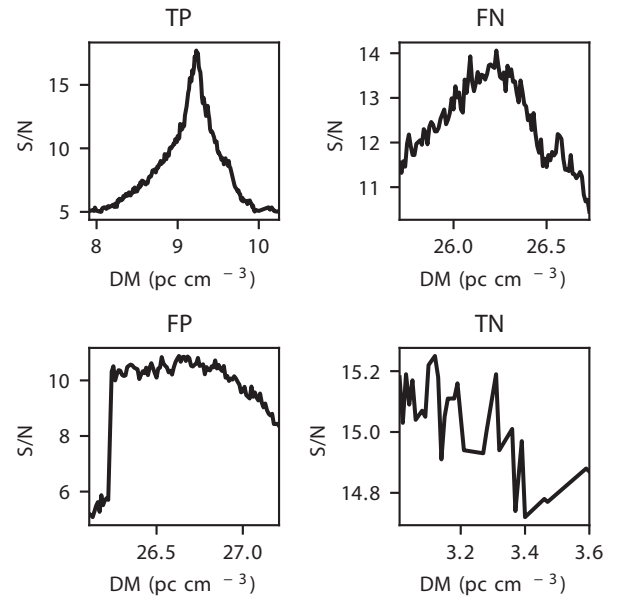


Figure 2. S/N curves along DM of four pulses with the relative classification indicated on top.

little impact on our results. This is because we used a sufficiently large collection of labelled training samples describing all classes, which compensates. In any case, such mislabelled examples can often help prevent a classifier from overfitting, which reduces real-world performance (Duda, Hart & Stork 2000; Bishop 2006). Using this selection process, also pulses affected by noise or simultaneous RFI, which are desirable to be retained, were included. In this way, we aimed to reduce potential classification bias against the rare class. The distribution of some parameters of RFI and astrophysical instances in the training set is shown in Fig. 1.

3.2.2 Feature design

We developed different algorithms to extract features used to select relevant signals. These algorithms were created by analysing individual characteristics of RFI instances. At the end of the process, individual features had been evaluated according to the value of their information gain. Features with a low information gain value were removed until a peak in the number of correctly classified instances was reached. Residual redundant features were subsequently removed via an iterative method, manually deleting all the features one by one and calculating the number of correctly classified instances for each configuration, until a peak was reached. Deneva et al. (2016) developed a classification algorithm based on equation (12) of Cordes & McLaughlin (2003) that they reported performing well. However, we could not use the same algorithm to extract new features for L-SPS because it was too slow in our implementation and required a large number of events per pulse.

At the end of this process, the following five features were found to yield the highest number of correctly classified instances. They are sorted in descending order according to their information gain (Table 1):

- (i) Pulse width, W .
- (ii) Weighted mean of the pulse DM.
- (iii) Excess kurtosis of the width curve as a function of DM.
- (iv) Excess kurtosis of the S/N curve as a function of DM.
- (v) Pulse S/N.

⁶<http://www.cs.waikato.ac.nz/ml/weka>

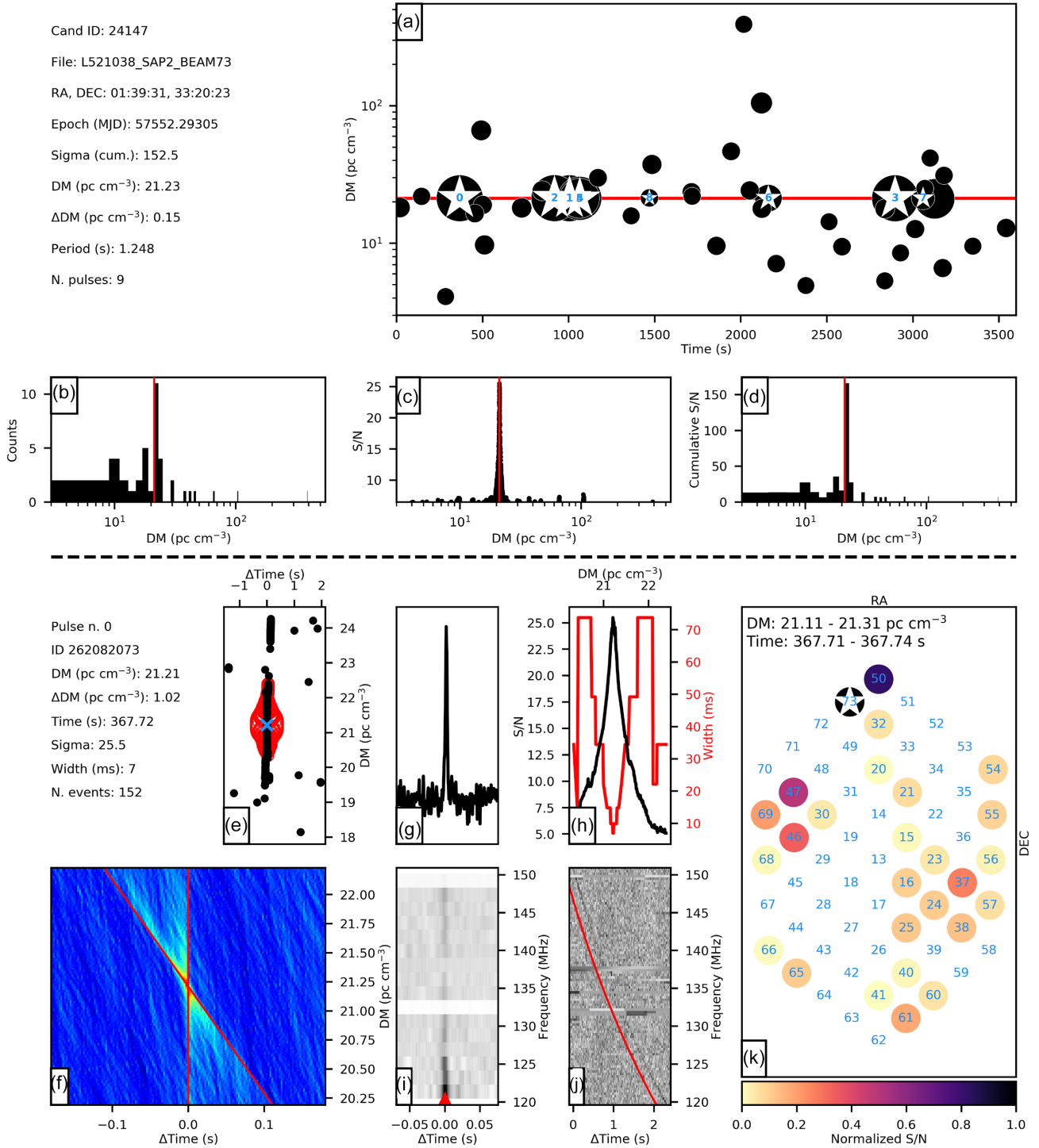


Figure 3. Discovery plot of PSR J0139+33. Diagnostic plots like this are generated for each of the positively classified candidates from the LOTAAS survey. Top: summary plots of the pulses detected in the beam. Bottom: plots of the brightest pulse forming the candidate. The description of each sub-panel is given in Section 3.4.

Features (ii), (iii), and (iv) have been adapted from Tan et al. (2018).

3.2.3 Feature evaluation

The classifier's ability to select astrophysical detections and to reject RFI instances is estimated using the manually selected sample of

pulses described in Section 3.2.1. Ideally, training and test sets should be composed of independent samples. However, we chose to use the whole sample of astrophysical detections available to train the classifier. Therefore, we used the same training set to test the classifier performance by running WEKA's 10-fold cross-validation, which randomly divides the sample into 10 groups. Nine of the groups are used to build a model and one to evaluate it.

Table 1. ML features selected for L-SPS to separate astrophysical pulses from RFI and their value of information gain. The subscript e indicates a single event within the pulse and σ is the standard deviation. The rest of the symbols are defined in the text.

	Feature	Inf. gain
(i)	$W = W_e(\max(S/N_e))$	0.74
(ii)	$\overline{DM} = \frac{\sum_e DM_e S/N_e}{\sum_e DM_e}$	0.71
(iii)	$k_W = \frac{\sum_e (DM_e - \overline{DM})^4 W_e}{\sigma^4(W_e) \sum_e W_e} - 3$	0.40
(iv)	$k_{S/N} = \frac{\sum_e (DM_e - \overline{DM})^4 W_e}{\sigma^4(S/N_e) \sum_e S/N_e} - 3$	0.29
(v)	$S/N = \max(S/N_e)$	0.10

Table 2. Confusion matrix for the ML classifier of L-SPS applied to the LOTAAS set.

	Predicted pulsar	Predicted RFI
Actual pulsar	TP=17754	FN=249
Actual RFI	FP=355	TN=34708

The process is iterated for each group and the average values are calculated. Using the features and the ML algorithm described in the previous sections, 98.9 per cent of instances in the LOTAAS set are correctly classified and the resulting confusion matrix is reported in Table 2.

A manual inspection of the FP and FN instances returned by the classifier on the training set was performed. The vast majority of the mislabelled instances would have been misclassified even after visual inspection, with high probability. It is possible that these instances were spurious detections incorrectly labelled when the training set was built (see Section 3.2.1). In rare cases, however, the instances misclassified by the ML classifier could be correctly labelled after visual inspection. This is partially because astrophysical pulses affected by RFI sometimes mimic RFI behaviour and therefore some RFI is mislabelled in order to retain such astrophysical signals. This could also be an indication that there is still space for additional features useful for discriminating RFI instances difficult to classify from true astrophysical signals. The S/N curves along DM of four pulses are reported in Fig. 2 as an example.

We have compared the performance of our classifier to others presented in the literature. However, it is important to note that statistical metrics cannot be used to compare classifiers on different samples. In fact, if we had chosen e.g. different thresholds on the pulses to build the training set (e.g. S/N or number of events) a different performance would have resulted. Therefore, it is only for reference that we compare the performance of other single-pulse classifiers. We only compare our ML classifier and not the full classification process (e.g. event grouping, beam comparison, etc.). The values obtained for the different single-pulse classifiers are reported in Table 3.

3.3 Computational performance of the L-SPS classifier

As mentioned in Section 2, the number of events that are stored at the end of the PRESTO-based pipeline is on the order of 10^7 – 10^8 above 5σ for a typical LOTAAS observation. By only selecting one pulse for each group of events and by removing all the pulses formed by fewer than six events, having $S/N < 6.5$ or $DM < 3 \text{ pc cm}^{-3}$, as discussed in Section 3.1.2, the number of instances decreases

to $\sim 10^6$. At this point, pulses are selected using the ML classifier described in Section 3.2. For a typical observation, the ML selection takes roughly a minute to process the whole observation. At the end of the ML process $\sim 10^4$ pulses remain. The spatial comparison of different beams on the sky removes roughly half of these pulses. The grouping of pulses into candidates, the selection of the brightest for each DM range, and the thresholds on their cumulative S/N values discussed in Section 3.1.3 typically leave ~ 20 diagnostic files to inspect per observation. In total, L-SPS takes around 30 min to process one observation on one Cartesius node (powered by 2×12 cores 2.6 GHz Intel Xeon E5-2690 v3 Haswell). Multimoment analysis could further reduce the number of RFI instances and tests are ongoing.

3.4 Diagnostic plots

Despite the huge progress in ML classification in recent years, a final human inspection is essential. Bright signals can be easily identified by machines and by humans, but for weak signals the classification becomes uncertain even after human inspection. For this reason, we developed a set of diagnostic plots to help visually identify astrophysical signals even when they are buried in the noise. Once an observation has been processed, a file containing the diagnostic plots is generated for each candidate. An example is shown in Fig. 3, which represents the discovery plot of PSR J0139+33. For comparison, its detection in a confirmation observation is shown in Fig. A1, while a diagnostic plot for an RFI candidate is shown in Fig. A2.

In the top part, four summary plots of detections in the beam are present, along with meta-data containing information on the beam where the candidate was detected. In Fig. 3(a), all the pulses positively classified in the beam are plotted with a size proportional to their S/N as a function of time of arrival and DM. The pulses belonging to the candidate shown are highlighted with stars and the 10 brightest are numbered in order of decreasing S/N. Fig. 3(b) reports the number of positively classified pulses per DM range. Fig. 3(c) shows the distribution of S/N versus DM of the events forming the pulses. Fig. 3(d) shows the cumulative S/N of the pulses for each DM range. The red lines indicate the candidate DM.

In the bottom part, the sub-panels show information related to the brightest pulse associated with the candidate source. Meta-data about the pulse is reported at the top-left. Fig. 3(e) is a close-up of the brightest pulse from Fig. 3(a). It shows all the single events in the DM and time ranges plotted as black dots, including those considered to be noise or RFI. Events forming the pulse are highlighted with red circles, where the size is proportional to the S/N. The blue cross marks the most likely DM and time of the burst. Fig. 3(f) shows the time series around the pulse time and DM. Red lines highlight the expected smearing of a broad-band signal. Fig. 3(g) shows the profile of the pulse, i.e. the time series around the pulse's time of arrival de-dispersed at the pulse's DM. Fig. 3(h) represents the S/N (black line) and width (red line) of the events forming the pulse as functions of DM. Fig. 3(i) and (j) represent the pulse spectrum around the pulse's time of arrival. These two plots are generated using WATERFALLER.PY⁷. In Fig. 3(i), the spectrum is de-dispersed at the pulse DM, down-sampled in frequency by a factor of 162 to 16 sub-bands and smoothed in time with a boxcar function of the same width as the pulse. The red triangle indicates the expected position of the signal. Fig. 3(j) shows the dispersed data around the

⁷<https://github.com/plazar/pyrpulsar>

Table 3. Metrics to evaluate the performance of different single-pulse classifiers. The ML algorithm of L-SPS is applied to the LOTAAS training set. The metrics used can assume any value between 0 and 1 and are described in the text. While the FNR and FPR assume lower values for a better classification, the rest of the metrics assume higher values for a better classification. RRATRAP and CLUSTERRANK do not use ML and only rough estimations are available. The metrics are evaluated on different data sets and thus they are reported only for reference (see the main text for details). Values for RF_{OVER}^2 are from Devine et al. (2016), for RRATRAP are from Karako-Argaman et al. (2015), and for CLUSTERRANK are from Deneva et al. (2016).

Classifier	FNR	FPR	TPR	ACC	PPV	G-M	F-M
L-SPS	0.014	0.010	0.986	0.989	0.980	0.983	0.983
RF_{OVER}^2	0.282	0.011	0.718	–	–	–	0.716
RRATRAP	0.2	0.09	–	–	–	–	–
CLUSTERRANK	0.3	0.07	–	–	–	–	–

Table 4. Pulsars discovered by early versions of L-SPS in the LOTAAS survey. The level of pulse brightness variability during the observation is described qualitatively.

Name	Period (s)	DM (pc cm ⁻³)	Variability
PSR J0139+33	1.248	21.2	Extreme
PSR J0301+20	1.207	19.0	Large
PSR J0317+13	1.974	12.9	Large
PSR J0454+45	1.389	20.8	Some
PSR J1340+65	1.394	30.0	Some
PSR J1404+11	2.650	18.5	Some
PSR J1849+15	2.233	77.4	Extreme

pulse, down-sampled in frequency by a factor of 27–96 sub-bands, and in time to 3 times the pulse width. The red curve represents the expected DM sweep of the signal. The RFI visible in Figs 3(i) and (j) as horizontal stripes is the cause of the weak detections around the pulse in Fig. 3(e) and of the short-duration events in the pulse tails in Fig. 3(h). Fig. 3(k) reports the maximum S/N of the events detected in each of the coherent beams of the SAP, within the DM and time ranges indicated at the top. The S/N is normalized between 0 and 1 and represented as a hot-shades colour scale. A white star indicates the beam where the candidate was detected and beams are numbered in blue. In this case, the signal is strongest in beam 73, where the pulse is detected. The complicated sensitivity pattern and RFI are responsible for the (apparent) weaker detections in other beams.

4 EARLY SINGLE-PULSE DISCOVERIES FROM THE LOTAAS SURVEY

The L-SPS classifier has been developed and tested using data from the LOTAAS survey. The latest version of the program is presented in this paper and it will be used in a complete reanalysis of the survey data. Around 80 known pulsars have been blindly identified by the different preliminary versions of L-SPS. Some of these were used to produce the training set described in Section 3.2.1. Seven new pulsars have been discovered to date using L-SPS — dozens have also been discovered using the periodicity searches and associated classifier (Tan et al. 2018). A summary of their properties is reported in Table 4. PSR J0139+33 is an RRAT that was not detectable in periodicity searches in the discovery pointing, nor in successive, targeted observations. PSRs J0301+20, J0317+13, and J1849+15 manifest a strong variability between single pulses. For this reason, the first two were initially detected through their bright single pulses and subsequently also found using a periodicity search in follow-up observations. The remaining pulsars are bright enough to be detected in both periodicity and single-pulse searches. The timing

solutions of the sources presented here will be reported in future papers.

5 CONCLUSIONS AND FUTURE DEVELOPMENTS

We have presented L-SPS, a new classifier for searches of single-radio pulses in the LOTAAS survey. It employs an ML algorithm to discriminate astrophysical signals from RFI, with high accuracy. During its development, the algorithm has discovered seven new pulsars and blindly identified ~ 80 known sources. A full reprocessing of the LOTAAS data with the latest version of L-SPS, as presented here, is under way.

Future improvements to the software include testing of multi-moment analysis and development of additional features. Also, we only made use of binary classification, i.e. instances were divided into astrophysical and RFI. The use of multiclass classification (e.g. distinguishing broad-band and narrow-band RFI) could improve the performance (Tan et al. 2018). In addition, a larger training set with positive instances better distributed (e.g. more high-DM pulsars) will be used in future. Finally, deep learning techniques could significantly improve the classification performance (Deng & Yu 2014). However, deep learning algorithms typically require larger training sets. Therefore, they could possibly be used to reprocess the survey data when a sufficient number of discoveries and re-detections are achieved.

5.1 Portability

Although L-SPS has been designed specifically for the LOTAAS survey, efforts are ongoing to produce a more general software (SPS; Michilli & Hessels 2018) that can be readily adapted to other projects. The aim is to create a program that is user-friendly, simple to customize, and robust to different observation characteristics, in order to be easily used in a generic study. This is achieved by designing a modular software where the different tasks discussed in Section 3 are performed by different modules executed in sequence by a main script. Therefore, the sequence can be easily modified and each of the modules can be tailored to a specific study. A first release of this software is freely available on [github](https://github.com/danielemichilli/SPS)⁸. At the time of writing, while the correct operation of SPS has been extensively tested, some of the features developed specifically for the LOTAAS survey still need to be included, such as the capability to process multiple telescope beams in parallel over different computer cores.

Arguably the most critical part of the program is the final selection of astrophysical signals. In fact, due to the need for a large data set of labelled detections, an ML analysis can be difficult or

⁸<https://github.com/danielemichilli/SPS>

impossible to perform in the case of small studies. Therefore, a set of filters that do not rely on ML techniques was created for these situations. While such filters have been designed to keep the false-negative rate low, the FPR will be higher than in the ML approach, though still manageable for small projects. It is difficult to assess the general performance of these filters since they depend on the characteristics of the specific observations. A rough estimate is obtained from the study of FRB 121102 with Arecibo. The features used typically reduce the number of candidates by ~ 80 – 90 per cent. Of the remaining candidates, typically ~ 25 per cent were found to be real after visual inspection, though this fraction varied between 1 per cent and 64 per cent in different observations (depending on the severity of RFI).

ACKNOWLEDGEMENTS

This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. Computing time was provided by NWO Physical Sciences (project SH-242-15). The LOFAR facilities in the Netherlands and other countries, under different ownership, are operated through the International LOFAR Telescope foundation (ILT) as an international observatory open to the global astronomical community under a joint scientific policy. DM and JETH acknowledge funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Starting Grant agreement no. 337062 ('DRAGNET'). JETH also acknowledges funding from an NWO Vidi fellowship. JvL acknowledges funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. 617199. Different PYTHON modules were used in this study, specifically NUMPY, MATPLOTLIB, PANDAS, and ASTROPY.

REFERENCES

- Bethapudi S., Desai S., 2018, *Astronomy and Computing*, 23, 15
 Bilous A. V. et al., 2016, *A&A*, 591, A134
 Bishop C. M., 2006, *Pattern Recognition and Machine Learning*. Springer-Verlag, Berlin
 Brown G., Pocock A., Zhao M.-J., Luján M., 2012, *J. Mach. Learn. Res.*, 13, 27
 Burgay M. et al., 2006, *MNRAS*, 368, 283
 Clifton T. R., Lyne A. G., 1986, *Nature*, 320, 43
 Coenen T. et al., 2014, *A&A*, 570, A60
 Cordes J. M., McLaughlin M. A., 2003, *ApJ*, 596, 1142
 Deneva J. S. et al., 2016, *ApJ*, 821, 10
 Deng L., Yu D., 2014, *Foundations and Trends in Signal Processing*, 7, 3–4, Now Publishers Inc., Hanover, MA, USA
 Devine T. R., Goseva-Popstojanova K., McLaughlin M., 2016, *MNRAS*, 459, 1519
 Duda R. O., Hart P. E., Stork D. G., 2000
 Eatough R. P., Keane E. F., Lyne A. G., 2009, *MNRAS*, 395, 410
 Eatough R. P., Molkenhuth N., Kramer M., Noutsos A., Keith M. J., Stappers B. W., Lyne A. G., 2010, *MNRAS*, 407, 2443
 Ellingson S. W., 2004, *Exp. Astron.*, 17, 261
 Ford J. M., 2017, PhD thesis, Nova Southeastern Univ.
 Garrett M. A., 2013, in *Proceedings of "From Antikythera to the Square Kilometre Array: Lessons from the Ancients"*. PoS, Trieste, 170, 41
 Hermsen W. et al., 2013, *Science*, 339, 436
 Hewish A., Bell S. J., Pilkington J. D. H., Scott P. F., Collins R. A., 1968, *Nature*, 217, 709
 Huchra J. P., Geller M. J., 1982, *ApJ*, 257, 423
 Hulten G., Spencer L., Domingos P., 2001, in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '01. ACM, New York, NY, p. 97
 Karako-Argaman C. et al., 2015, *ApJ*, 809, 67
 Keane E. F., Ludovici D. A., Eatough R. P., Kramer M., Lyne A. G., McLaughlin M. A., Stappers B. W., 2010, *MNRAS*, 401, 1057
 Kondratiev V. I. et al., 2016, *A&A*, 585, A128
 Lee K. J. et al., 2013, *MNRAS*, 433, 688
 Lorimer D. R., Kramer M., 2004, *Handbook of Pulsar Astronomy*. Cambridge University press, Cambridge, UK
 Lorimer D. R., Bailes M., McLaughlin M. A., Narkevic D. J., Crawford F., 2007, *Science*, 318, 777
 Lyon R., Brooke J., Knowles J., Stappers B., 2014, *2014 22nd International Conference on Pattern Recognition (ICPR)*, p. 1969
 Lyon R. J., Stappers B. W., Cooper S., Brooke J. M., Knowles J. D., 2016, *MNRAS*, 459
 McLaughlin M. A. et al., 2006, *Nature*, 439, 817
 Michilli D., Hessels J. W. T., 2018, *SpS: Single-pulse Searcher*. Astrophysics Source Code Library. Astrophysics Source Library Code, record ascl:1806.013
 Mitchell T. M., 1997, *Machine Learning*. 1 edn. McGraw-Hill, Inc., New York, NY
 Mozden T. J., Bowman J. D., Monsalve R. A., Rogers A. E. E., 2017, *MNRAS*, 464, 4995
 Petroff E. et al., 2015, *MNRAS*, 451, 3933
 Powers D. M. W., 2011, *J. Mach. Learn. Technol.*, 2, 37
 Press W. H., Davis M., 1982, *ApJ*, 259, 449
 Ransom S. M., 2001, PhD thesis, Harvard Univ.
 Sanidas S. et al., 2018, in preparation
 Scholz P. et al., 2016, *ApJ*, 833, 177
 Spitler L. G., Cordes J. M., Chatterjee S., Stone J., 2012, *ApJ*, 748, 73
 Spitler L. G. et al., 2016, *Nature*, 531, 202
 Stappers B. W. et al., 2011, *A&A*, 530, A80
 Tan C. M., Lyon R. J., Stappers B. W., Cooper S., Hessels J. W. T., Kondratiev V. I., Michilli D., Sanidas S., 2018, *MNRAS*, 474, 4571
 Taylor G. B. et al., 2012, *J. Astron. Instrument.*, 1, 1250004
 Tendulkar S. P. et al., 2017, *ApJ*, 834, L7
 Thornton D. et al., 2013, *Science*, 341, 53
 Tingay S. J. et al., 2013, *PASA*, 30, e007
 van Haarlem M. P. et al., 2013, *A&A*, 556, A2
 van Leeuwen J., Stappers B. W., 2010, *A&A*, 509, A7
 Yao Y., Xin X., Guo P., 2016, in *Proceedings of the 12th International Conference on Computational Intelligence and Security (CIS)*. Springer Nature Switzerland, Basel, p. 120

APPENDIX A: RFI DIAGNOSTIC PLOT

Examples of two diagnostic plots generated by L-SPS are presented. Fig. A1 reports the detection of PSR J0139+33 during its confirmation observation. Fig. A2 shows an example of a typical diagnostic plot for a candidate produced by RFI. The explanation of the sub-plots is given in Section 3.4.

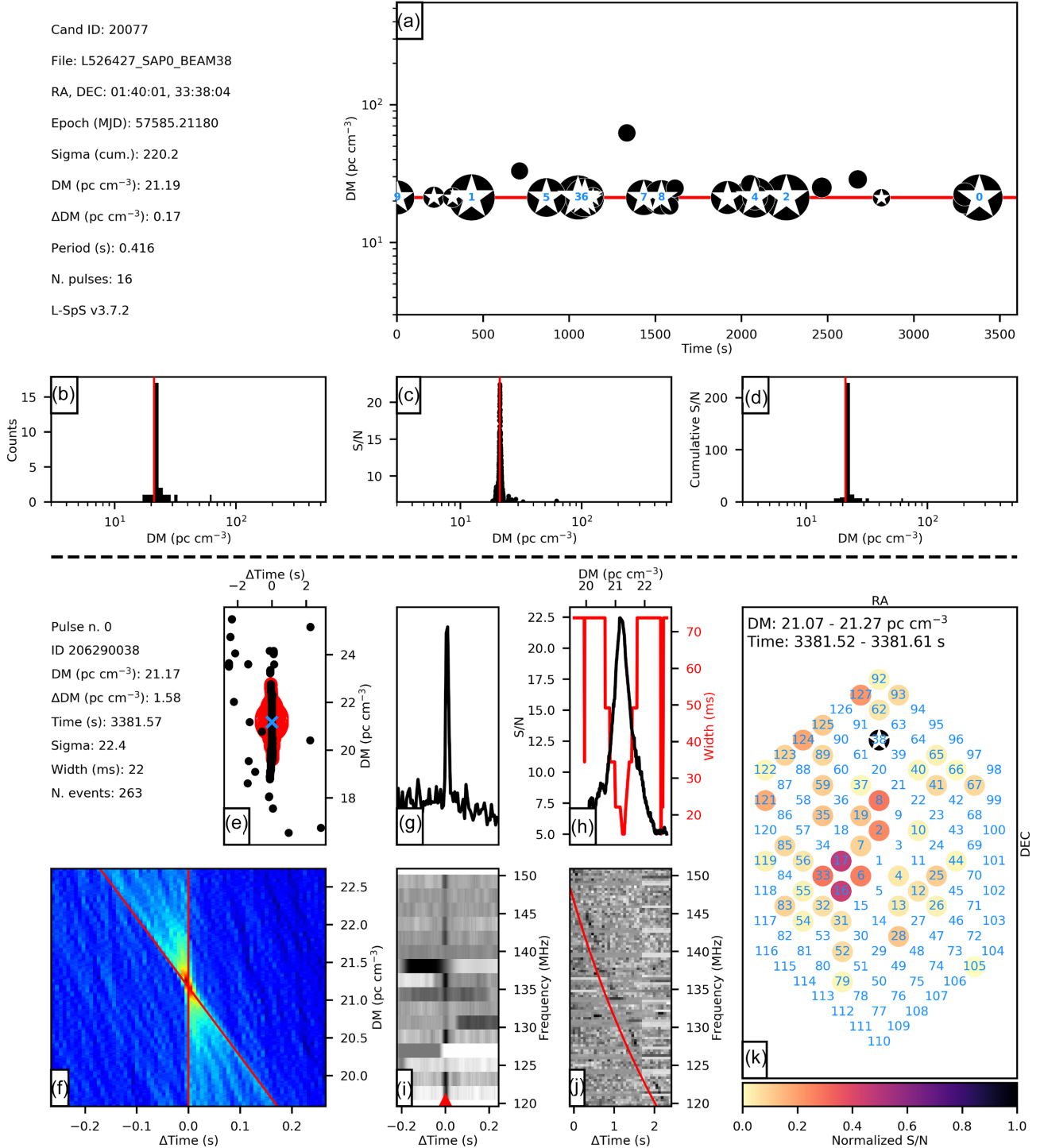


Figure A1. Example of diagnostic plots of the confirmation observation of PSR J0139+33. The same plots as Fig. 3 are represented and they are discussed in the text. The number of beams visible in panel (k) is different for confirmation observations. The strong interference visible as horizontal stripes in panels (i) and (j) did not prevent to detect the bright pulses from the RRAT.

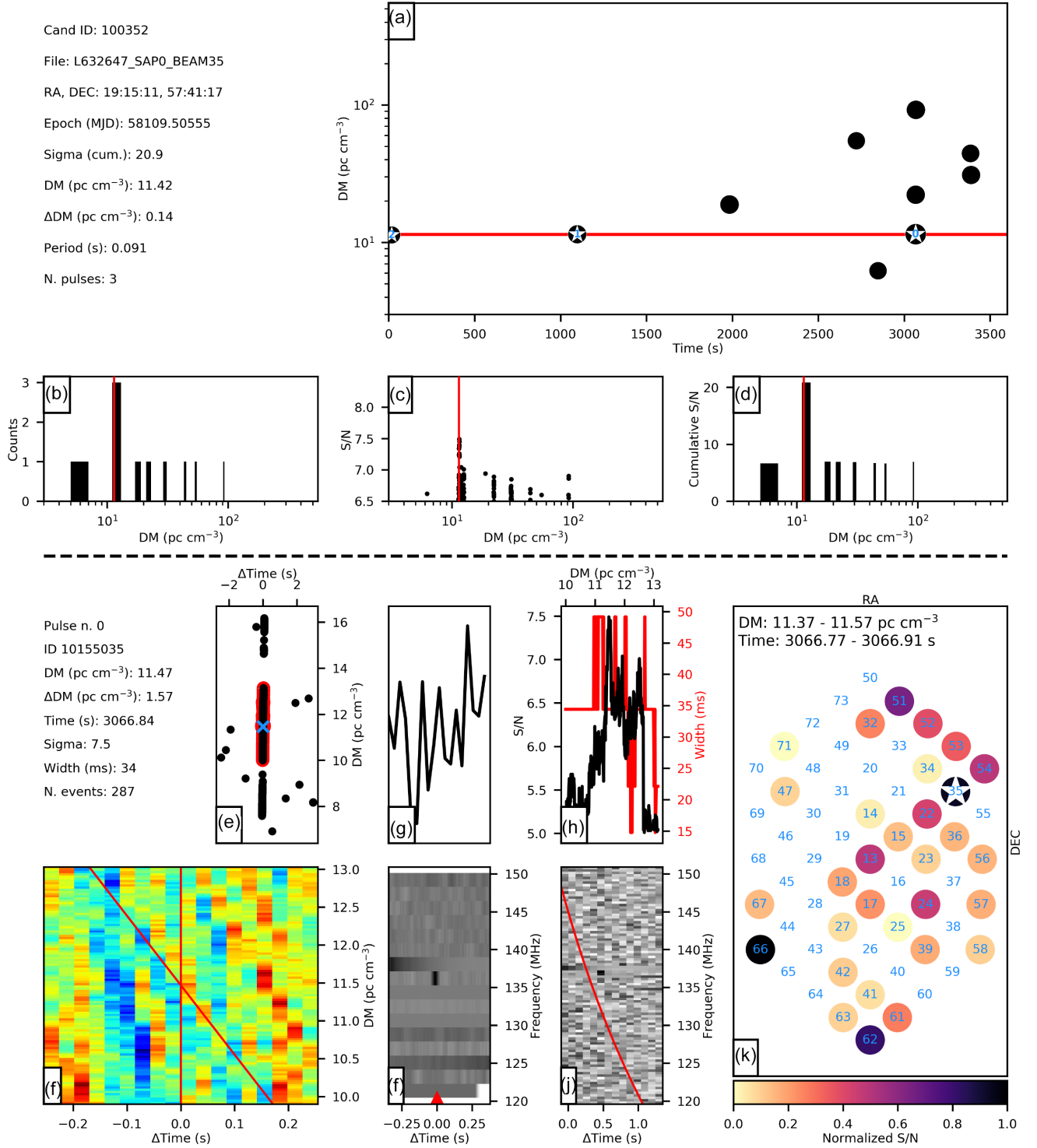


Figure A2. Example of diagnostic plots for an RFI candidate. The same plots as in Fig. 3 are represented and they are discussed in the text.

This paper has been typeset from a \LaTeX file prepared by the author.