

# Discovering Pulsars with Machine Learning

Vincent Morello

Presented in fulfillment of the requirements  
of the degree of Master of Science

May 2016

Faculty of Science, Engineering and Technology  
Swinburne University



---

# Abstract

Pulsars are magnetized and rapidly rotating neutron stars that emit a periodic, pulsed signal most often observed in the radio domain. They enable investigations of fundamental physics in extreme conditions, and truly unique high precision tests of Gravity when found orbiting a companion star. The most efficient way to discover new pulsars is to survey wide areas of the Galactic plane or even the entire sky using a radio telescope and search the recorded time series data for periodic signals using a Fourier transform. This generates up to millions of *candidates*, sets of diagnostic values and graphical plots each describing the signal from a potential new source; the overwhelming majority of them are unfortunately the manifestation of terrestrial interference, and differentiating them reliably from genuine pulsars is a subtle task that traditionally relies on visual inspection by an expert. Due to the ever increasing data volumes acquired by pulsar surveys, this approach is already unsustainable and human intervention needs to be removed without sacrificing discovery potential.

In this thesis we present a pulsar candidate classifier named SPINN that uses an artificial neural network, a type of Machine Learning (ML) algorithm. While the idea is not new, the accuracy reached by previous implementations has arguably been insufficient, either missing discoveries or still leaving an excessive amount of spurious candidates to be rejected by a human operator. Here we delve into the inner workings of ML algorithms, and from general principles governing their use on any data problem we proceed to tackle the specificities of pulsar candidate classification. We attach particular importance on how to properly represent pulsar candidates as vectors of real-valued inputs (or *features*) to the algorithm. A detailed understanding of the entire pulsar searching process is helpful, if not necessary to design optimal features. We also went one step back and entirely rewrote the software responsible for producing the candidate plots, as their quality significantly impacts the accuracy of the classifier that rates them.

Our classification system is tested on a recent pulsar survey observed with the Parkes radio telescope, and shown to reduce the amount of non-pulsar candidates to inspect by a factor of 1,000 while correctly identifying almost every known pulsar observation. A single person can now extract the value from the outputs of an entire pulsar survey in a matter of hours instead of years. More importantly, we discovered 21 new pulsars with SPINN, more than half of them in archival data previously searched and left behind due to the use of more rudimentary candidate selection tools. Among the new discoveries are the fastest spinning pulsar ( $P = 1.49$  ms) in the Galactic plane and a possible transitional

millisecond pulsar. We conclude by discussing how to achieve further improvements to the accuracy of classifiers in prevision of pulsar searches to be conducted with the first phase of the Square Kilometer Array (SKA), but point out that a system similar to SPINN could already credibly handle the associated candidate output.



# Acknowledgements

The path that eventually led to this thesis being written was an unexpected one, but finally here I am. I will first extend my gratitude to my principal supervisor Willem van Straten, first and foremost for getting me into the Swinburne pulsar group in the first place, and then for his unwavering support throughout my enrollment at Swinburne. With me being based in Europe, this involved doing many administrative chores on my behalf and having weekly meetings with me late in the evening for two hours among other things. He also gave me a lot of freedom in pursuing my own ideas and his very positive attitude certainly carried me through some of the more difficult times. I can only wish to future students to have an experience as good as the one I had with Willem. I would also like to thank my second supervisor, Ewan Barr, for those endless, overly nerdy, sometimes slightly passionate discussions about computing and data processing in all its forms. Rejoice, it looks like we will have many more in the future.

Being mainly offshore, I missed out a lot on the very enjoyable atmosphere at Swinburne, but every time I visited for a while I certainly had a good time. In no particular order: thank you Andrew Jameson for letting me access your stash of delicious Dutch waffles every afternoon around 4PM, back in the days when we shared the same office; Chris Flynn, for being extremely nice and knowledgeable, and therefore my go-to person for any question about astrophysics, especially the basic ones I should really know the answer to; Alex Codoreanu for being a delightful housemate and friend every time I visited, and those countless hours playing Pro Evolution Soccer<sup>1</sup>; and greetings to my fellow students Igor Andreoni, Dany Vohl and Chris Curtin for being excellent company as well.

And finally, my deepest gratitude goes to my beloved wife Marie, for supporting me through the many years we have been together, for her contribution in my renewed enjoyment for science, and for tolerating my enthusiastic and needlessly detailed explanations of whatever algorithm I may be interested in/trying to implement at that moment.

---

<sup>1</sup>sometimes at the expense of more pressing matters, such as writing this thesis.



## Declaration

The work presented in this thesis has been carried out in the Centre for Astrophysics & Supercomputing at Swinburne University of Technology between 2014 and 2016. This thesis contains no material that has been accepted for the award of any other degree or diploma. To the best of my knowledge, this thesis contains no material previously published or written by another author, except where due reference is made in the text of the thesis.

*Vincent Morello  
Melbourne, Victoria, Australia  
2016*

A handwritten signature in black ink, appearing to read "V. Morello".

# Contents

<b>Abstract</b>	i
<b>Acknowledgements</b>	iii
<b>Declaration</b>	v
<b>List of Figures</b>	ix
<b>List of Tables</b>	xi
<b>1 Introduction</b>	1
1.1 The Pulsar phenomenon and its observables . . . . .	1
1.1.1 Formation . . . . .	1
1.1.2 Radio emission properties . . . . .	2
1.1.3 Spin-down . . . . .	3
1.1.4 Age and magnetic field strength estimates . . . . .	5
1.1.5 Effects of the interstellar medium . . . . .	6
1.2 The known pulsar population . . . . .	7
1.2.1 Normal Pulsars . . . . .	7
1.2.2 Millisecond Pulsars . . . . .	9
1.2.3 Magnetars . . . . .	10
1.2.4 RRATs . . . . .	11
1.3 Pulsar searching: the Big Picture . . . . .	11
1.3.1 Investigating neutron star properties and evolution . . . . .	11
1.3.2 Tests of Gravity . . . . .	12
1.3.3 Gravitational wave detection . . . . .	13
1.3.4 Special Mention: Fast Radio Bursts . . . . .	14
1.4 Thesis outline . . . . .	14
<b>2 Discovering New Pulsars</b>	17
2.1 Survey Strategies . . . . .	18
2.2 Periodicity Searches for Pulsars . . . . .	19
2.2.1 De-dispersion . . . . .	19
2.2.2 Fourier Analysis . . . . .	22
2.2.3 Acceleration Searches . . . . .	25
2.2.4 Pulsar Candidate Folding . . . . .	26

---

2.3	The Pulsar Candidate Selection Problem . . . . .	29
2.3.1	Problem outline . . . . .	29
2.3.2	Crowdsourcing . . . . .	29
2.3.3	Interactive Selection Software . . . . .	31
2.3.4	Classification Algorithms . . . . .	33
2.3.5	The case of the Square Kilometre Array (SKA) . . . . .	34
<b>3</b>	<b>SPINN: A Machine Learning Algorithm for Pulsar Discovery</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.1.1	A Quick Tour of Machine Learning Tasks . . . . .	37
3.1.2	Solving a Machine Learning Problem . . . . .	38
3.2	Artificial Neural Networks . . . . .	39
3.2.1	Mathematical Model . . . . .	39
3.2.2	Important Visual Intuition . . . . .	41
3.2.3	Training . . . . .	43
3.2.4	Implementation Details . . . . .	44
3.3	Feature Design . . . . .	45
3.3.1	General Guidelines . . . . .	45
3.3.2	The Bias-Variance Dilemma . . . . .	46
3.3.3	A Short Pulsar Candidate Inspection Manual . . . . .	47
3.3.4	List of Features Used . . . . .	48
3.4	SPINN: Performance Optimization . . . . .	57
3.4.1	Training data composition . . . . .	57
3.4.2	Classification Performance Metrics . . . . .	57
3.4.3	Cross-Validation . . . . .	58
<b>4</b>	<b>Results and Discoveries</b>	<b>61</b>
4.1	Interlude: The importance of folding software . . . . .	61
4.2	The CUBR folding pipeline . . . . .	63
4.2.1	Overview . . . . .	63
4.2.2	RFI mitigation based on outlier detection . . . . .	64
4.2.3	Adaptive optimization grid . . . . .	66
4.2.4	Results . . . . .	67
4.3	Performance evaluation on the SUPERB survey . . . . .	68
4.3.1	Test setup . . . . .	68
4.3.2	Classification accuracy . . . . .	69

4.3.3	Classification speed . . . . .	73
4.4	Summary of searches conducted . . . . .	73
4.5	Discoveries . . . . .	74
<b>5</b>	<b>Conclusion</b>	<b>77</b>
5.1	Summary . . . . .	77
5.2	Future work and SKA pulsar searches . . . . .	78
	<b>Bibliography</b>	<b>89</b>



# List of Figures

1.1	Schematic view of a pulsar based on the Goldreich-Julian model . . . . .	3
1.2	A $P - \dot{P}$ diagram of the known pulsar population . . . . .	8
2.1	Galactic coordinates distribution of the known pulsar population . . . . .	19
2.2	Raw observation segment of the known pulsar J0738-4042 . . . . .	20
2.3	Dedispersed observation segment of the known pulsar J0738-4042 . . . . .	22
2.4	Principle of a Fourier based search . . . . .	24
2.5	Illustration of the period-DM optimization algorithm . . . . .	28
2.6	Examples of candidate plots . . . . .	30
2.7	Screen capture of the interactive candidate selection software REAPER . .	32
3.1	The artificial neuron model . . . . .	41
3.2	A feed-forward neural network . . . . .	42
3.3	Visualization of a neural network's output on 2-D data . . . . .	42
3.4	Simulated feature distributions with varying degrees of usefulness . . . . .	46
3.5	Illustration of the bias-variance dilemma in practice . . . . .	47
3.6	Comparison between pulsar and RFI candidate plots . . . . .	49
3.7	Illustration of the signal-to-noise algorithm . . . . .	51
3.8	Distribution of the $E_{\text{DM}}$ quantity on a sample of 139 pulsar observations.	52
3.9	Histograms of SPINN features . . . . .	55
3.10	Selected scatter plots of SPINN features . . . . .	56
3.11	Confusion Matrix of a binary classifier . . . . .	58
4.1	Candidate plot of a pulsar disturbed by interference. . . . .	62
4.2	RFI mitigation based on outlier detection . . . . .	65
4.3	Demonstration of RFI mitigation algorithms on a pulsar observation . . .	66
4.4	Adaptive Period-DM optimization grid . . . . .	67
4.5	Classification performance metrics as a function of score decision threshold	70
4.6	Lowest scoring test pulsar instance . . . . .	71
4.7	Trade-off between recall and false positive rate . . . . .	72
4.8	Sky coverage of surveys searched with SPINN . . . . .	76
5.1	False positive rate per hour, SKA extrapolation . . . . .	81



## List of Tables

2.1	Amount of candidates produced by modern surveys . . . . .	34
4.1	Folding speed benchmarks . . . . .	64
4.2	Period-DM optimization benchmarks . . . . .	68
4.3	Parameters of surveys searched with SPINN . . . . .	74
4.4	Pulsars discovered by SPINN . . . . .	75



# 1

## Introduction

Pulsar astronomy abruptly began its existence as a field of study in 1967, when a regularly spaced train of pulses with a period of 1.33 seconds was serendipitously discovered by Jocelyn Bell in radio data charts. Independent re-observations confirmed the astrophysical nature of the source. It was hypothesized to be a highly magnetized and rapidly rotating neutron star (Hewish et al., 1968), with (Gold, 1968) predicting a steady decay of its spin frequency. The discovery of the Crab pulsar a year later (Staelin & Reifenstein, 1968) led to the general acceptance of this theory by establishing an association between a pulsar and a supernova remnant and providing the first measurement of a frequency derivative of the observed periodic signal.

### 1.1 The Pulsar phenomenon and its observables

---

#### 1.1.1 Formation

Depending on their initial mass, stars follow different evolutionary tracks but, as all burn their nuclear fuel into heavier elements, they inevitably leave behind an unspent, dense core, which gravitational forces tend to contract. Electron degeneracy pressure is sufficient to oppose gravity as long as the mass of the core does not exceed the so-called Chandrasekhar limit of  $1.4 M_{\odot}$  (Chandrasekhar, 1931). In that case, the stellar remnant is a white dwarf. Otherwise, further collapse into a more compact neutron star (NS) occurs, itself balanced by neutron degeneracy pressure unless its mass lies above the Tolman-Oppenheimer-Volkoff limit (Oppenheimer & Volkoff, 1939), in which case a runaway contraction into a black hole is inevitable. The Tolman-Oppenheimer-Volkoff limit is not known accurately but is taken to be below  $3 M_{\odot}$ , with a lower bound placed at  $2.01 M_{\odot}$  from measurements of the most massive known pulsar (Antoniadis et al., 2013).

Stars with initial masses in the  $\simeq 8 - 25 M_{\odot}$  range have the potential to end up as neutron stars following a supernova explosion (e.g. Stahler & Palla, 2005). Recent work indicates that in the  $\simeq 8 - 11 M_{\odot}$  range, *isolated* stars are likely to finish their lives as white dwarfs; only those involved in mass transfers from a close binary companion during the end of their main sequence phase may turn into supernovae and leave a NS behind (Podsiadlowski et al., 2004; van den Heuvel, 2007).

Known neutron star masses are distributed around a median value of  $1.4 M_{\odot}$  (e.g. Zhang et al., 2011) which is considered the canonical pulsar mass for the purposes of most discussions. An upper bound on the radius may be derived from a simple argument, namely that at the surface of a pulsar the stellar material must be bound by gravity, which may be written as

$$R_{\max} = \frac{GMP^2}{4\pi^2} = 16.8 \text{km} \left( \frac{M}{1.4M_{\odot}} \right)^{1/3} \left( \frac{P}{\text{ms}} \right)^{2/3}, \quad (1.1)$$

where  $M$  is the pulsar's mass and  $P$  its spin period. For the fastest spinning pulsar J1748–2446ad (Hessels et al., 2006),  $P = 1.39\text{ms}$  which leads to  $R = 20.9\text{km}$ . The exact radius depends on the equation of state of dense matter; most models predict it to be close to  $10 \text{ km}$  (Lattimer & Steiner, 2014), a value usually taken as the canonical neutron star radius.

### 1.1.2 Radio emission properties

Pulsars are visible across the entire electromagnetic spectrum from radio to gamma rays, but their (arguably) most remarkable feature is their periodic, pulsed radio emission. It is observed at frequencies ranging from approximately  $100 \text{ MHz}$  to  $200 \text{ GHz}$  (Torne et al., 2015) with a spectrum best described by a power law with a median spectral index of  $-1.4$  (Bates et al., 2013). It also has a high degree of linear polarization and extreme brightness temperatures of the order of  $10^{25} \text{ K}$  (first reported by Lyne & Smith, 1968), requiring the emission process to be coherent. This naturally leads to the “lighthouse” analogy: a narrow beam of highly directional, coherent radio emission aligned with the pulsar's magnetic axis sweeps across our line of sight once per rotation (Fig. 1.1).

This simple picture accounts for the properties of the observed radiation and is sufficient to understand the basics of pulsar astronomy. However, things get vastly more complicated when investigating the emission mechanism, which is not clearly understood even today. We shall just very briefly mention a few concepts introduced in the early work of Goldreich & Julian (1969). First, that a pulsar spinning in a vacuum inevitably

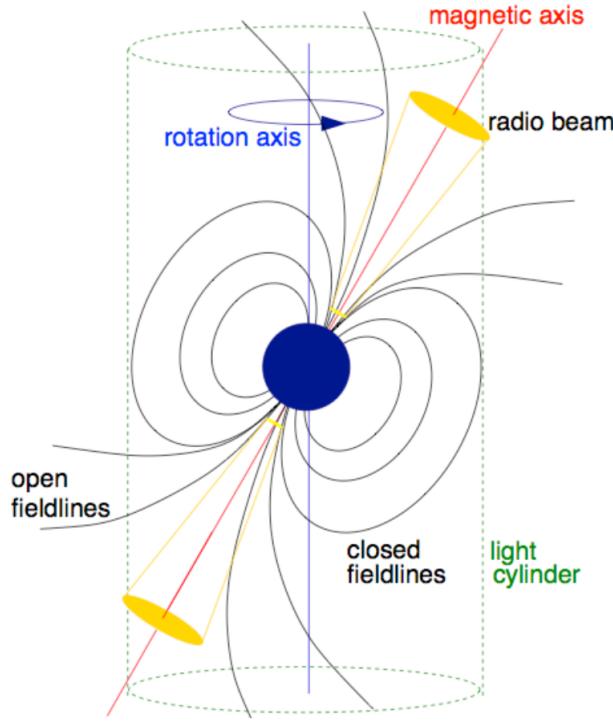


Figure 1.1 A schematic view of a pulsar based on the Goldreich-Julian model (see text). There is still no consensus on the physical processes responsible for the coherent radio emission, although the area around the magnetic poles is the most likely source. Figure from Lorimer & Kramer (2004).

ends up surrounded by a magnetosphere, as an intense induced electric field at the surface rips off charged particles. Second, that a sufficiently dense magnetosphere will reach a force-free state<sup>1</sup> causing it to co-rotate rigidly with the star, but no further than the so called “light cylinder”: an imaginary surface defined by  $r = c/\Omega$ , where  $c$  is the speed of light and  $\Omega$  the pulsar’s angular velocity, on the edge of which the outermost closed magnetic field line passes (Fig. 1.1). This defines a region of special significance in most models, the polar cap, where magnetic field lines are open; it is suspected to be the locus of electron-positron pair plasma production, ultimately leading to coherent emission (e.g Sturrock, 1971; Ruderman & Sutherland, 1975; Chen & Ruderman, 1993)

### 1.1.3 Spin-down

Even without knowledge of the local processes in a pulsar’s magnetosphere, some useful analytical relations can be obtained. The pulse period of a pulsar is observed to increase

---

<sup>1</sup>Where charges arrange themselves to create a distribution such that  $\mathbf{E} + (\boldsymbol{\Omega} \times \mathbf{r}) \times \mathbf{B} = 0$

over time, and the associated loss of kinetic energy can be explained by a number of causes, some of the most notable being magnetic dipole radiation (e.g Jackson, 1998), the emission of a relativistic particle wind (Harding et al., 1999), or magnetic field decay (e.g. Mereghetti et al., 2015). A pulsar with spin period  $P$  undergoing spin-down loses rotational energy at a rate called the *spin-down luminosity* given by

$$L_{\text{sd}} = -\frac{d}{dt} \left( \frac{1}{2} I \Omega^2 \right) = -I \Omega \dot{\Omega} = 4\pi^2 I \dot{P} P^{-3}, \quad (1.2)$$

where  $\Omega = 2\pi/P$  is the angular rotation frequency of the pulsar and  $I$  its moment of inertia. The latter can be written as  $I = kMR^2$ , where  $M$  and  $R$  stand respectively for mass and radius, and  $k$  is a constant depending on how the radial density profile of the neutron star is modeled. For a solid sphere of uniform density  $k = 2/5$ .

It is useful to work under the simplistic assumption that the loss of rotational energy is dominated by magnetic dipole radiation, in which case a number of characteristic pulsar parameters can be derived. Consider a dipole with a magnetic moment  $\mathbf{m}$ , rotating at an angular frequency  $\Omega$  with  $\alpha$  being the angle between spin and magnetic axes. It can be shown under classical electromagnetism (e.g. Jackson, 1998) that it radiates an electromagnetic wave of frequency  $\Omega$  carrying a total power of

$$E_{\text{dipole}} = \frac{2}{3c^3} |\mathbf{m}|^2 \Omega^4 \sin^2 \alpha, \quad (1.3)$$

and equating this expression with that of the spin-down luminosity yields

$$\dot{\Omega} = -\frac{2|\mathbf{m}|^2 \sin^2 \alpha}{3Ic^3} \Omega^3. \quad (1.4)$$

To take into account potential deviations from this pure dipole braking model, the last equation may be rewritten in a still simple but more general power-law form

$$\dot{\Omega} = -C\Omega^n, \quad (1.5)$$

where  $n$  is the *braking index* of a given pulsar ( $n = 3$  for pure dipole braking) and  $C$  a quantity assumed to be constant. By rewriting equation 1.5 in terms of spin frequency  $\nu = \Omega/(2\pi)$  and then taking its derivative, one obtains

$$n = \frac{\nu \ddot{\nu}}{\dot{\nu}^2}. \quad (1.6)$$

A measure of  $n$  may be therefore be obtained from the knowledge of the second derivative of a pulsar's spin frequency  $\ddot{\nu}$ . In practice, reliable measurements of  $\ddot{\nu}$  are difficult

and have been performed for less than ten pulsars that exhibit both high  $\dot{\nu}$  and spin-down evolutions exempt from irregularities such as timing noise or glitches (abrupt changes in spin period, e.g. Espinoza et al., 2011b). Known values for  $n$  range from  $-1.5$  to  $3.16$  (Espinoza et al., 2011a; Archibald et al., 2016).

#### 1.1.4 Age and magnetic field strength estimates

Equation 1.5 may be rewritten in terms of period as  $\dot{P} = KP^{2-n}$ . Postulating that  $K$  and  $n$  are both constant and integrating that expression over the lifetime  $T$  of a pulsar one gets

$$T = \frac{P}{(n-1)\dot{P}} \left[ 1 - \left( \frac{P_0}{P} \right)^{n-1} \right], \quad (1.7)$$

where  $P_0$  is the pulsar's spin period at birth. By further assuming that  $P_0 \ll P$  and that magnetic dipole braking is the sole source of spin-down ( $n = 3$ ), we obtain the *characteristic age* of the pulsar

$$\tau_c = \frac{P}{2\dot{P}} \approx 15.8 \text{Myr} \left( \frac{P}{s} \right) \left( \frac{\dot{P}}{10^{-15}} \right)^{-1}. \quad (1.8)$$

Considering the number of simplifying assumptions under which that expression is obtained, it cannot be considered accurate but remains a useful order of magnitude estimate that can readily be obtained from just  $P$  and  $\dot{P}$ . Another caveat is that the characteristic age expression is not applicable to recycled pulsars that have undergone prolonged spin-up phases via accretion of matter and angular momentum stolen from a binary companion, a subject that will be covered in the next section.

A practical estimate of the surface magnetic field strength of a pulsar can be also be obtained, using the fact that the field created by a magnetic dipole is given by  $B \approx |\mathbf{m}|/r^3$ . Combining that fact with equation 1.4 leads to

$$B_S = B(r = R) = \sqrt{\frac{3c^3 I}{8\pi^2 R^6 \sin^2 \alpha}} P \dot{P}, \quad (1.9)$$

which is the *characteristic surface magnetic field strength*, with  $R$  being the pulsar's radius and  $I$  its moment of inertia. Assuming the neutron star is a rigid sphere of uniform density, and using the canonical mass  $M = 1.4M_\odot$  and radius  $R = 10\text{km}$ , we may write

$$B_S \approx 10^{12} \left( \frac{P}{s} \right)^{\frac{1}{2}} \left( \frac{\dot{P}}{10^{-15}} \right)^{\frac{1}{2}} \text{Gauss}, \quad (1.10)$$

where  $\alpha$  has been set to  $90^\circ$ . With multiple terms of equation 1.9 being uncertain and some relevant spin-down mechanisms neglected in its derivation, it remains no more than a convenient order of magnitude estimate much like the characteristic age.

### 1.1.5 Effects of the interstellar medium

All frequencies of an electromagnetic wave propagating in a perfect vacuum obviously reach a distant observer at the same time, but this is visibly not the case as far as pulsar observations are concerned. Hewish et al. (1968) immediately noticed that lower frequencies were observed to arrive with a measurable delay, a fact accurately explained by propagation of light through an interstellar medium made of cold plasma. Such a medium has a refractive index  $\mu$  given by (e.g Huba, 2013)

$$\mu = \sqrt{1 - \left(\frac{\nu_p}{\nu}\right)^2}, \quad (1.11)$$

where  $\nu_p$  is the plasma frequency (e.g Huba, 2013) whose expression is

$$\nu_p = \sqrt{\frac{e^2 n_e}{\pi m_e}} \approx 8.97 \sqrt{\frac{n_e}{\text{cm}^{-3}}} \text{kHz}, \quad (1.12)$$

with  $e$  and  $m_e$  being respectively the electron charge and mass, and  $n_e$  the number of free electrons per unit of volume. Electromagnetic waves therefore propagate at a frequency-dependent velocity  $v = c\mu$ , with  $c$  being the speed of light, that we can approximate by the following expression, assuming  $\nu \gg \nu_p$ :

$$v \approx c \left(1 - \frac{\nu_p^2}{2\nu^2}\right). \quad (1.13)$$

Integrating the above along the line of sight of a pulsar situated at a distance  $d$ , we get the time elapsed between the arrival of a broadband pulse observed at a frequency  $\nu$  and its arrival at infinite frequency, called the *dispersion delay*

$$\Delta t = \frac{e^2}{2\pi m_e c} \frac{\text{DM}}{\nu^2}, \quad (1.14)$$

where DM, for *dispersion measure*, is the integrated free electron density along the line of sight

$$\text{DM} = \int_0^d n_e dl, \quad (1.15)$$

and is usually expressed in units of  $\text{cm}^{-3}\text{pc}$ . Replacing the fundamental constants by their

values in equation 1.14, we get the more handy relation

$$\Delta t \simeq 4.1488 \times 10^3 \left( \frac{\text{DM}}{\text{cm}^{-3}\text{pc}} \right) \left( \frac{\nu}{\text{MHz}} \right)^{-2} \text{ sec.} \quad (1.16)$$

With accurate knowledge of the Galactic distribution of free electrons, the distance to a pulsar may be numerically obtained from equation 1.15, as a function of DM and Galactic coordinates. In practice even state of the art models of the free electron distribution such as the NE2001 model (Cordes & Lazio, 2002) yield distance estimates that carry a relative uncertainty of at least 20%. Also, when using a wide-band receiver to search for pulsars in the Galactic plane, the *a priori* unknown DM of a potential source is one of the main contributors to the computational cost of a search: a range of DMs must be postulated, and corresponding delays between observed frequency channels are digitally applied to correct for dispersion, so that pulses are phase-aligned before summation across channels (see next chapter for details).

While dispersion depends only on the total integrated free electron density along the line of propagation, inhomogeneities in the ISM have a visible scattering effect, causing the light from a single radio pulse to travel via a set of slightly different paths to the observer. It can be shown (e.g. Lorimer & Kramer, 2004) that an observed pulse is expected to be broadened by a *scattering timescale*  $\tau_s \propto d^2\nu^{-4}$  where  $d$  is the distance to the pulsar and  $\nu$  the observing frequency. Bhat et al. (2004) has proposed an empirical relation between DM and  $\tau_s$ , and shows that for typical observing frequencies and lines of sight into the Galactic plane, the scattering time scale can reach several milliseconds. This negatively impacts the detectability of fast spinning pulsars at large distances.

## 1.2 The known pulsar population

---

Since some characteristics of pulsars can be obtained from simple expressions involving only their period and period derivative, a natural way to visualize the pulsar population is to display it in a  $P - \dot{P}$  diagram (Fig. 1.2) in which several clusters corresponding to different sub-species appear. In this section we will go through a quick tour of the major ones.

### 1.2.1 Normal Pulsars

The dominant cluster in the center right of the  $P - \dot{P}$  diagram is that of the normal pulsars, with periods ranging from around 100 ms to a few seconds and characteristic ages under  $\simeq 1$  Gyr. They are normal in the sense that they constitute the largest fraction of

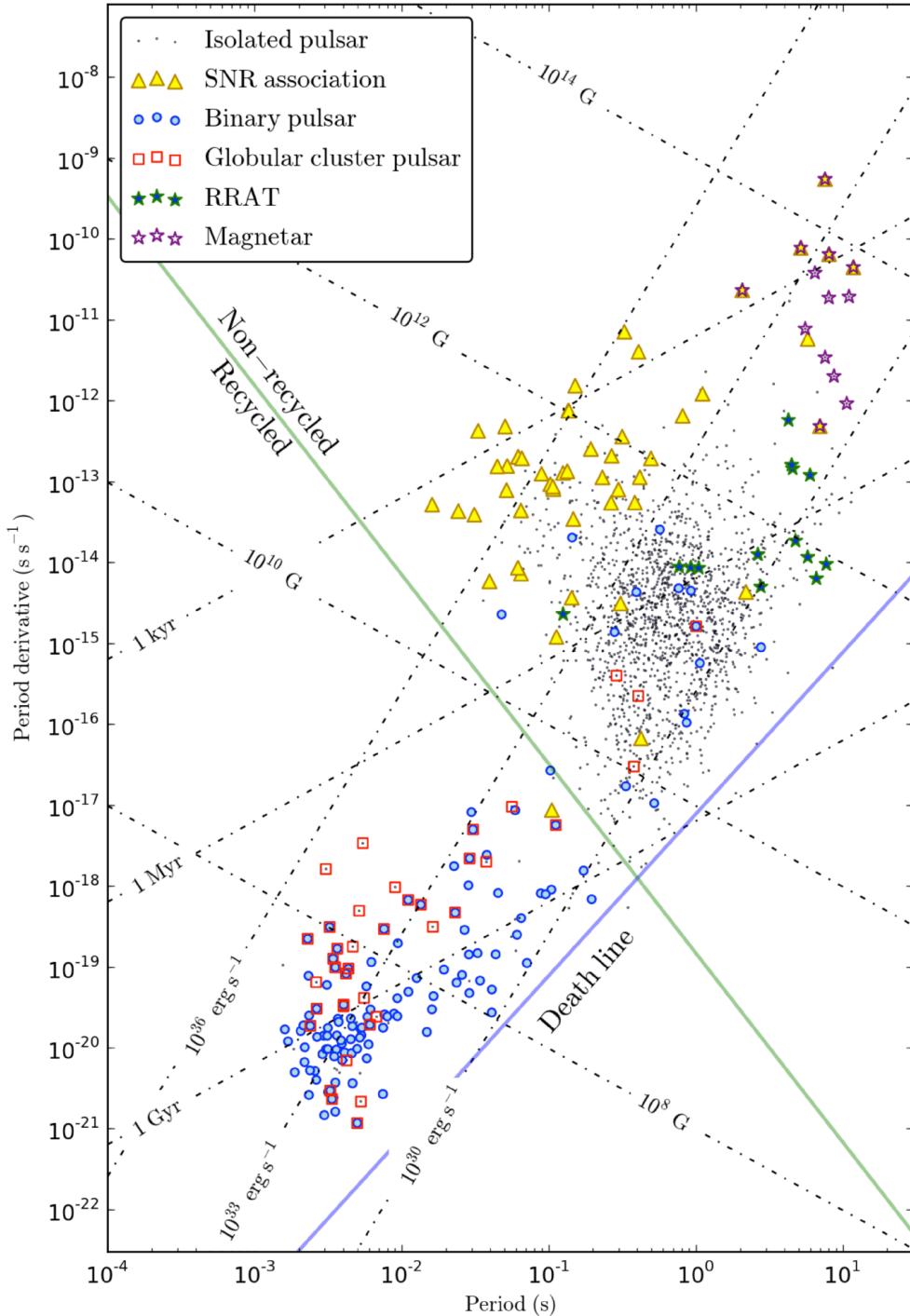


Figure 1.2  $P - \dot{P}$  diagram with 1,800 pulsars. Figure is from Barr (2012) and made with data from the ATNF pulsar catalogue. Lines of constant characteristic magnetic field strength, characteristic age and spin down luminosity are also represented. Pulsars below the so called “death line” (Chen & Ruderman, 1993) are not expected to emit in radio. The frontier shown between recycled and non-recycled pulsars is based on the empirical relation of (Lee et al., 2012). Note that the measured  $\dot{P}$  of globular cluster pulsars can be very significantly affected by their acceleration in the cluster’s gravitational potential.

the known population.

A young pulsar is expected to begin its existence at the top left of the cluster, with a high spin-down luminosity and a low characteristic age. It is therefore reassuring that a significant number of those, marked as yellow triangles on Fig. 1.2, have been associated with supernova remnants (SNRs, e.g. Camilo et al., 2002, 2009), taken to have lifetimes shorter than  $\simeq 10^5$  years. A SNR association enables an independent age estimate, either by measuring the proper motion of the pulsar and tracking it back to its birthplace (the apparent center of the remnant), or by mapping the expansion velocity of the SNR shell itself (Braun et al., 1983). If left to evolve without external influence, a pulsar will gradually spin down<sup>2</sup> until the radio emission processes cannot be sustained anymore. Chen & Ruderman (1993) derived an analytical relationship between magnetic field strength and the longest spin period above which a radio pulsar is expected to shut down, materialized by a so called “death line” (Fig. 1.2), although that model may be refined or challenged by the discovery of radio pulsars in unexpected regions of the  $P - \dot{P}$  diagram (Young et al., 1999; Zhang et al., 2000).

### 1.2.2 Millisecond Pulsars

A distinct group of objects in the lower left of the diagram is that of the millisecond pulsars (MSPs). The term was coined with the unexpected discovery of PSR B1937+21 (Backer et al., 1982) with a pulse period of 1.56 ms, far shorter than any of its known counterparts at the time. The generally accepted formation scenario for MSPs (Alpar et al., 1982; Radhakrishnan & Srinivasan, 1982) involves a binary stellar system of which the more massive component is the pulsar progenitor, reaching the end of its life first. If the resulting supernova explosion does not disrupt the binary, the other star will later reach its red giant phase, at which point it will inflate, fill its Roche lobe and a transfer of both mass and angular momentum to the pulsar will occur. During this process, known as *recycling*, the pulsar is gradually spun up and the system is observable as an X-Ray binary, as part of the gravitational potential energy of the infalling matter is released as high energy emission (Bhattacharya & van den Heuvel, 1991). The exact duration and processes involved in this accretion phase, as well as the ultimate fate of the system is mostly driven by the initial mass of the companion star  $M_C$  (e.g. Tauris & van den Heuvel, 2006, for a review), with major distinctions being made between low ( $M_C < 1M_\odot$ ), high ( $M_C > 10M_\odot$ ) and intermediate mass X-ray binaries. Evidence for this formation hypothesis includes the

---

<sup>2</sup>Assuming pure dipole braking, it will carve a path in the  $P - \dot{P}$  diagram along a line of constant characteristic magnetic field strength.

fact that almost 80% of MSPs reside in binary systems compared to approximately 2% for ordinary pulsars, and the recent discovery of *transitional millisecond pulsars*. Over time scales of a few years, these have been directly observed to cycle between a radio-quiet, bright X-ray emission state where an accretion disk is present, which then vanishes and gives way to a radio emission state with observable pulses (Archibald et al., 2009; Papitto et al., 2013).

MSPs are more difficult to discover than their ordinary counterparts for several reasons. Due to their shorter periods, their pulses are more negatively affected by interstellar scattering, in particular when looking deep into the Galactic plane. They also require observations with fast sampling times and a layer of additional, computationally expensive data processing. This is related to the adverse effects of orbital motion on the usual Fourier based search method for periodic emission, which we discuss at length in the next chapter.

### 1.2.3 Magnetars

Magnetars are characterized by their long spin periods ( $\simeq 2\text{--}12$  s) and high period derivatives implying intense surface magnetic fields  $B \simeq 10^{13} - 10^{15}$  G. They were discovered through two classes of high energy phenomena that were initially believed to be distinct. On one hand Soft Gamma Repeaters (SGRs), as their apt name indicates, distinguished themselves from standard gamma ray bursts by their ability to repeat (Mazets et al., 1979) and their softer spectrum (Atteia et al., 1987). On the other, Anomalous X-ray Pulsars (AXPs) were first thought to be located in X-ray binaries like the known similar sources at the time, but their long periods, large spin down rates and most importantly the absence of a binary companion suggested otherwise. There is now considerable evidence that they are one and the same. Over time many AXPs have been found to be SGRs and vice versa, and magnetars are the generally accepted progenitors for both (see Mereghetti et al., 2015, for a recent review).

Magnetars diverge from the rest of the population at their birth, when an internal, efficient helical dynamo is thought to form if certain initial conditions are met during supernova collapse (Duncan & Thompson, 1992), causing the unusually strong magnetic field. In contrast to ordinary pulsars, the loss of rotational energy alone cannot account for the observed strength of their high energy emission, the main power source being instead magnetic field decay (Thompson & Duncan, 1996). That being said, magnetars deserve their mention in the present introduction as four out of 23 known (Olausen & Kaspi, 2014) have been found to emit radio pulses. This indicates a link with radio pulsars, also supported by observations of a high-B rotation-powered pulsar switching to a magnetar-

like state with outbursts of high energy emission (Gavriil et al., 2008).

#### 1.2.4 RRATs

Rotating RAdio Transients or RRATs are the latest addition to the expanding pulsar menagerie. Discovered by McLaughlin et al. (2006), they are Galactic radio sources from which pulses are detectable only occasionally (one every few minutes to hours). Long term monitoring reveals that they possess an underlying periodicity and an observable spin-down rate, making the neutron star association irresistible. Whether they are a class of objects distinct from normal pulsars is a matter of debate. The working definition of RRAT (Keane & McLaughlin, 2011) is a pulsar more easily detected by its single pulses rather than the more conventional periodicity searches (see next chapter); the possibility of them having underlying weak periodic emission cannot be ruled out, which would make the distinction with ordinary pulsars telescope dependent and arbitrary. They occupy a region of the  $P - \dot{P}$  diagram not particularly remarkable, although they appear to possess longer periods and stronger magnetic fields (Keane & McLaughlin, 2011); this hints at the possibility that they are somewhere down the evolutionary track of magnetars, an idea put forward by (Lyne et al., 2009) after observations of magnetar-like behavior in RRAT J1819–1458.

### 1.3 Pulsar searching: the Big Picture

---

It turns out that pulsars are not only the most frequent manifestation of neutron stars, but also tremendously valuable laboratories to investigate fundamental physics. In this section we briefly present the main reasons why effort is being invested into increasing the known pulsar population.

#### 1.3.1 Investigating neutron star properties and evolution

The most natural way to obtain answers to the main open questions in pulsar astronomy is to discover new specimens, with a strong preference for those at the edge, or even outside the known distribution. For example the most massive pulsars bring insights into the behavior of matter at extreme densities by ruling out certain equations of state, each of which predicts a maximum neutron star mass along with a mass-radius relationship (e.g. Demorest et al., 2010, Figure 3 in particular). Fast spinning pulsars also bring their own constraints (Lattimer & Prakash, 2004); unfortunately neither the mass nor radius of

the current spin frequency record holder can be measured (Hessels et al., 2006), but the discovery of a similar, more cooperative object would prove very valuable.

Expanding the sample of MSPs should also shed some light on why some of them do not perfectly fit in the usual formation narrative briefly explained in the previous section. Indeed, recycling involves rapid circularization of the orbit and a number of highly eccentric binary MSPs in the Galactic plane are at odds with this scenario<sup>3</sup> (e.g. Champion et al., 2008). It has been suggested that they are the product of interactions in hierarchical triple systems (Freire et al., 2011; Ransom et al., 2014). Equally interesting is the  $\simeq 20\%$  fraction of isolated MSPs; they may also have been ejected from triple systems (Portegies Zwart et al., 2011), or evaporated their former companions through the emission of a particle wind during recycling (Alpar et al., 1982; Bhattacharya & van den Heuvel, 1991). The so called “black widow” systems (Roberts, 2011), i.e. tight binaries surrounded by material seemingly ablated from a low mass companion, tend to give credence to the latter idea. Finding more millisecond pulsars in these transitory states will help fill the missing pieces of the MSP evolution puzzle. The same is also true for the other end of the period distribution, where a clear link between high-B ordinary pulsars, RRATs and magnetars could emerge.

### 1.3.2 Tests of Gravity

There is no need for a perfect understanding of the physics of pulsars to leverage their immense value as a simple yet extremely regular ticking beacon. *Pulsar timing* consists in measuring pulse times of arrival (TOAs) to which a model can be fit, involving the pulsar’s intrinsic spin period and its derivatives but also any phenomena that introduce extra delays as a function of time, such as orbital motion. The achievable precision on the physical parameters derived by pulsar timing is exceptional (e.g. see Table 2 of Verbiest et al., 2008, for PSR J0437–4715), and a flagship application is to compare the measured orbital evolution of pulsars in binary systems with predictions from various theories of gravity. Observations of non-Keplerian motion in the first double neutron star system (Hulse & Taylor, 1975) turned out to match predictions from General Relativity, including the measurement of an orbital decay over time consistent with the emission of gravitational waves (Taylor & Weisberg, 1982). Those results have since been vastly improved upon, following the discovery of a double pulsar system (Kramer et al., 2006) in which both components are beaming towards Earth. Pulsar - white dwarf binaries have

---

<sup>3</sup>Although in very dense stellar environments such as globular clusters, companion exchange is a valid explanation.

also found their use in constraining alternative theories of gravity (Freire, 2013; Damour & Esposito-Farese, 1992). Finding even more extreme systems, especially in tight orbits may eventually allow to find the limits of General Relativity, if any. The most sought after prize is the first discovery of a binary involving a pulsar and a stellar mass black hole. Such pairings are expected to exist (e.g. Sigurdsson, 2003) but have proven elusive so far. They would enable precise measurements of the black hole’s mass, spin, and a brand new battery of tests of gravity (e.g. Liu et al., 2014, for a recent review).

### 1.3.3 Gravitational wave detection

While observations of known double neutron star systems have already provided indirect evidence for the existence of gravitational wave (GW) emission, direct detection is now the focus of a number of large scale efforts. Among these are the so called *pulsar timing arrays* (PTAs, Foster & Backer, 1990); they are based on the principle that a passing GW distorts the space time metric around the Earth, imprinting a characteristic signature in the measured times of arrival of pulses of an ensemble of closely monitored pulsars (Hellings & Downs, 1983; Hobbs et al., 2010). GWs are expected to be observed in a range of frequencies spanning over ten orders of magnitude, and timing arrays are maximally sensitive to signals in the  $10^{-9}$  to  $10^{-7}$  Hz range which mostly includes a stochastic background from a large number of merging supermassive black-hole binaries (see e.g. Janssen et al., 2015, for a short review of GW properties and sources, and in particular their enlightening Figure 1).

There have been a number of recent and major developments regarding PTAs. On the very bright side, the first direct gravitational wave detection (in a much higher frequency range around 100 Hz) has been made by ground based detectors and attributed to a binary black hole merger (Abbott et al., 2016). On the other hand, Shannon et al. (2015) argues that the continuing null result of PTAs should be attributed to over-optimistic modeling of the GW background amplitude, and a potential cause could be the systematic over-estimation of supermassive black hole masses (Shankar et al., 2016). We may have to wait for the Square Kilometer Array (SKA) and its improved detection capabilities to witness the first practical demonstration of a PTA (Janssen et al., 2015). In any case, the sensitivity of any timing array increases with the number of pulsars monitored and the accuracy with which times of arrival of individual pulses can be measured. For this, millisecond pulsars with short pulse lengths are particularly desirable, and provide another motivation for continued searches.

### 1.3.4 Special Mention: Fast Radio Bursts

This introduction chapter would not be truly complete without mentioning Fast Radio Bursts (FRBs). An unintended consequence of the rise of single pulse searches for RRATs was the discovery in archival data of a bright ( $\simeq 30$  Jy) radio burst lasting a few milliseconds, and showing an apparent dispersion measure pointing at a clear extra-Galactic origin (Lorimer et al., 2007). Initial concerns about them being an instrumental or interference effect specific to Parkes have been dissipating with the discovery of  $\simeq 15$  more FRBs, including one at Arecibo (Spitler et al., 2014), another with the Green Bank Telescope (Masui et al., 2015) and three more at Molonglo (M. Bailes, priv. comm.).

Models for FRB progenitors must take into account the short event durations, and extreme luminosities and brightness temperatures implied by the cosmological distances. Compact sources involving extreme physical environments are favored (Kulkarni et al., 2014, reviews plausible stellar sources, which include neutron stars). More recently, the identification of FRB150418 in real time has enabled immediate, multi-wavelength follow-up of the event which has led to the detection and accurate localization of an associated<sup>4</sup> fading radio transient originating from a galaxy at redshift  $z \simeq 0.49$  (Keane et al., 2016). The authors also reported that the low measured stellar formation rate in the host galaxy disfavors progenitor models related to young stars. On the other hand, another FRB source was reported to repeat (Spitler et al., 2016) for which a plausible explanation seems to be a nearby extra-Galactic magnetar. It is possible that we are now faced with not one but two classes of events.

Regardless of the relationship between FRBs and neutron stars or not, they are a dramatic reminder that making observations always has a chance of leading to major discoveries unrelated to the initial search purpose.

---

## 1.4 Thesis outline

---

The goal of this thesis is to develop an automated solution to reduce significantly the large amount of human intervention required in the final stage of pulsar searches, namely the selection of credible candidates to re-observe for confirmation. Accurate selection is necessary as telescope time is precious and the overwhelming majority of candidates are due to interference from artificial radio sources.

---

<sup>4</sup>The reliability of the association is still hotly debated as these lines are being written, with the argument of a coincident AGN flare being put forward (Williams & Berger, 2016), which itself has already received criticism (Li & Zhang, 2016).

In Chapter 2, we delve into the details of Fourier-based periodicity searches for pulsars, explain the pulsar candidate selection problem and describe why existing solutions must be vastly improved to properly extract the value from ongoing and future pulsar surveys, in particular those to be conducted with the Square Kilometer Array (SKA).

In Chapter 3, we provide a short introduction to Machine Learning (ML) and the specific algorithm used in this work, namely artificial neural networks. We explain general principles on how to build a ML classifier and how we applied them to the design of SPINN, our proposed solution to the pulsar candidate selection problem.

In Chapter 4, we present the major results of this work: an objective test of classification performance of SPINN and the pulsar discoveries for which it is directly responsible.

We conclude in Chapter 5 with a brief discussion of the results of SPINN, the usability of a similar classification system with the SKA and implications for future work.



# 2

## Discovering New Pulsars

Discovering radio pulsars involves searching for dispersed pulses in spectra acquired with a radio telescope. High time resolution is necessary to detect the fastest spinning MSPs, while high frequency resolution is essential to compensate for the dispersive effects of the interstellar medium on a pulsar's signal. Pulsar searching methods fall into two categories:

- **Periodicity Searches.** Even when observed with the World's largest radio telescopes, individual pulses from most known pulsars are buried in background noise. It is therefore essential to leverage the very stable periodicity of their emission in a search. Forming the fluctuation power spectrum of the observed data with a Discrete Fourier Transform (DFT) reveals periodic signals otherwise undetectable directly in the time domain.
- **Single Pulse Searches.** A subset of the pulsar population shows emission of a transient nature, RRATs and nulling pulsars (Backer, 1970) in particular. They produce only occasional isolated pulses or sequences of pulses with little to no apparent regularity over usual integration times of several minutes to hours. Periodicity searches are not sensitive to these objects and directly searching for individual pulses is more effective, by applying matched filters to the time series data. This is also how FRBs have been found so far.

In this chapter we present an overview of a periodicity search pipeline, and the operations performed to turn high time resolution radio spectra into their final data product: candidates, a set of diagnostic information describing a possible periodic signal. Candidates are produced in very large numbers, of which genuine pulsar discoveries form a negligible fraction. Carefully inspecting the entire output of a modern survey is infeasible by a single person, and it is necessary to develop accurate classification algorithms to handle the problem.

## 2.1 Survey Strategies

---

The first decision to be made when searching for pulsars is that of the areas of sky or specific targets to observe. Telescope time being a precious resource, surveys are designed with maximum efficiency in mind. We present briefly some of the most economically sensible and successful strategies.

- **Searches of the Galactic Plane.** This is where stars are formed and where most normal pulsars (age  $\leq 100$  Myr) are naturally expected to be found, therefore a radio telescope beam pointed at the Galactic plane is highly efficient in terms of number of sources covered per unit of solid angle. Surveys of the Galactic plane have been by far the most prolific, with more than 800 discoveries made in successive analyses of the Parkes Multibeam Pulsar Survey alone (Manchester et al., 2001). Nearly two thirds of all known pulsars are located within 5 degrees of the Galactic equator (see also Fig 2.1), and this narrow strip of sky remains the highest priority target of ongoing and future surveys.
- **All-sky Surveys.** Millisecond pulsars are typically several billions of years old (Kiziltan & Thorsett, 2010) and stars with that age tend to occupy the thick disk of the Milky Way, which has a scale height of  $\approx 900$  pc (e.g. Jurić et al., 2008). Both dispersion and scattering by the ISM also limit the visibility of MSPs deep into the Galactic plane. From our vantage point in the Galaxy we can therefore expect to be surrounded by a relatively isotropic population of detectable MSPs (Fig 2.1), which is an incentive to cover the entire sky. All-sky surveys are more challenging due to the larger amounts of data to be processed and sifted through (see section 2.3).
- **Targeted Searches of Globular Clusters.** Stellar densities and rates of binary interaction in Globular Clusters (GC) are much higher than in the Galactic plane, making them prime hosts of X-ray binaries in particular (e.g. Freire, 2013), the main progenitors of millisecond pulsars. Furthermore, most Globular Clusters fit entirely within a typical radio telescope beam observing at L-band, and the data processing time saved compared to an expensive all-sky search allows longer observation times of hours instead of minutes, and deeper searches for binary systems to be performed (Section 2.2.3). To date, 144 pulsars have been found in 28 different Globular Clusters. 95 % of these discoveries are Millisecond Pulsars with spin periods below 50 ms, and 60 % are in binary systems.<sup>1</sup>

---

<sup>1</sup>A catalogue of pulsars in Globular clusters is maintained by Paulo Freire at <http://www2.naic.edu/~pfreire/GCpsr.html>

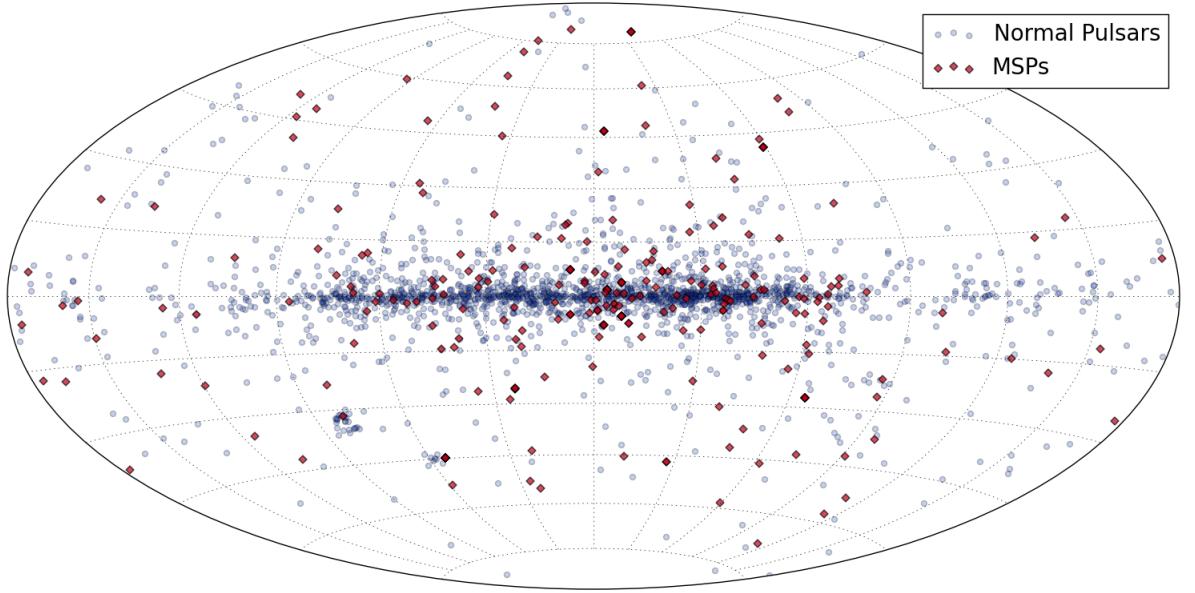


Figure 2.1 Distribution in Galactic coordinates of all 2,285 pulsars with published parameters. Here MSPs are defined as having periods shorter than 50ms. They show a much more uniform distribution than their slower spinning counterparts, which are concentrated along the Galactic plane. Figure made with data from the ATNF pulsar catalogue (Manchester et al., 2005).

## 2.2 Periodicity Searches for Pulsars

### 2.2.1 De-dispersion

A first step common to most pulsar searching methods is the removal of the frequency dependent dispersion delays introduced by propagation in the ISM (section 1.1.5). In practical terms, the raw data present themselves as a two dimensional array, where each row corresponds to a frequency channel, and represents an observed intensity as a function of time sample index (Fig 2.2). Given the dispersion measure (DM) of the observed source, de-dispersion consists of shifting every frequency channel by the appropriate number of time samples. These shifts are given by rounding the following formula to the nearest integer, which directly follows from equation 1.16:

$$p_i = 4.1488 \times 10^3 \left( \frac{\text{sec}}{t_{\text{samp}}} \right) \left( \frac{\text{DM}}{\text{pc cm}^{-3}} \right) \left[ \left( \frac{\nu_0}{\text{MHz}} \right)^{-2} - \left( \frac{\nu_i}{\text{MHz}} \right)^{-2} \right] \quad (2.1)$$

where  $p_i$  is the shift to apply to frequency channel index  $i$ , with index 0 corresponding to the highest observed radio frequency,  $t_{\text{samp}}$  is the data sampling interval, and  $\nu_0$  and  $\nu_i$

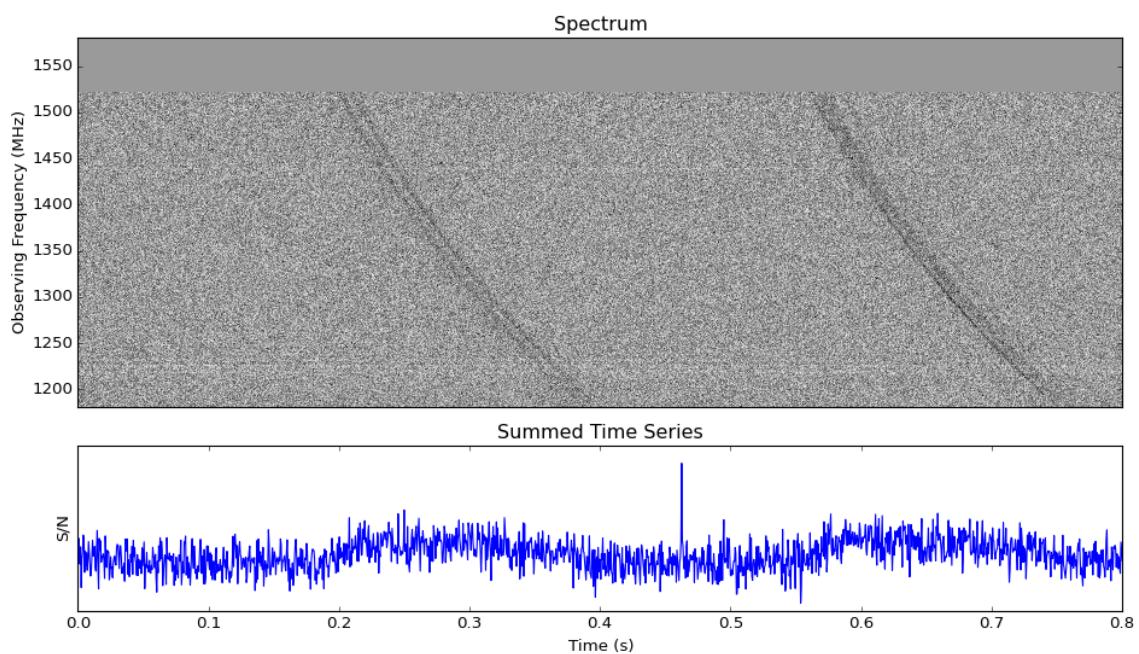


Figure 2.2 Top panel: a short segment of a high time resolution spectrum of the pulsar J0738-4042, acquired with the Parkes radio telescope. Bottom panel: sum of the spectrum across frequency channels. Two consecutive pulses are readily visible, but smeared by the summing process due to dispersion delays. The narrow spike at around 0.45 seconds is due to radio-frequency interference.

are the frequency of channels 0 and  $i$  respectively. Negative shifts are obtained if  $\nu_0 > \nu_i$ , which is the convention on Parkes data. Once these shifts have been applied, all channels are summed together to produce a one-dimensional de-dispersed time series (Fig 2.3) to be searched for periodic signals (Section 2.2.2).

In a blind pulsar survey, the dispersion measure of the yet to be discovered source is unknown and the data must be de-dispersed at many different trial DM values. Both the trial range and spacing have to be chosen in a sensible fashion. The highest trial DM should be a function of the celestial coordinates being observed, as lines of sight pointing at lower Galactic latitudes will cross larger densities of free electrons. In that case, using a model of the Galactic free electron density such as NE2001 (Cordes & Lazio, 2002) is required. Source DMs exceed  $1000 \text{ pc cm}^{-3}$  in the galactic plane, but remain lower than  $50 \text{ pc cm}^{-3}$  for most of the high Galactic latitude sky ( $|b| > 30^\circ$ ).

We can also derive an expression for the smallest reasonable DM trial step  $\Delta\text{DM}$  using a simple argument. As shown by equation 2.1, any DM value is uniquely associated with a shift between the top and bottom frequency channels of the observed band.  $\text{DM} = 0$  is the first trial value and obviously corresponds to a shift of zero time samples between extremal channels. The next trial value, equal to  $\Delta\text{DM}$ , should correspond to a top-to-bottom shift of exactly one time sample, as any smaller trial value injected in equation 2.1 will produce a nearly identical series of shifts due to them being rounded to the nearest integer. We therefore obtain

$$\Delta\text{DM} = 2.410 \times 10^{-4} \text{ pc cm}^{-3} \frac{t_{\text{samp}}}{\text{sec}} \left[ \left( \frac{\nu_0}{\text{MHz}} \right)^{-2} - \left( \frac{\nu_n}{\text{MHz}} \right)^{-2} \right]^{-1} \quad (2.2)$$

where  $\nu_0$  and  $\nu_n$  are the frequencies of extremal channels. If the observed bandwidth  $\Delta\nu = \nu_0 - \nu_n$  is small compared to the frequency of the center of the band  $\nu_c$ , one can use the more compact approximation

$$\Delta\text{DM} = 1.205 \times 10^{-4} \text{ pc cm}^{-3} \frac{t_{\text{samp}}}{\text{sec}} \frac{\nu_c^3}{\Delta\nu}. \quad (2.3)$$

In practice larger steps between trials are used to limit computing costs, based on the minimum pulse width one is interested in finding and a maximum acceptable S/N loss. For every DM trial, a one-dimensional time series is obtained and is searched independently for either single pulses or periodic signals. De-dispersing an observation in the most direct fashion requires  $n_{\text{samp}} \times n_{\text{chan}} \times n_{\text{DM}}$  additions, where  $n_{\text{samp}}$  is the number of time samples in the observation,  $n_{\text{chan}}$  the number of frequency channels and  $n_{\text{DM}}$  the number of trial DMs. Yet there is a large amount of additions common to consecutive trials, and this

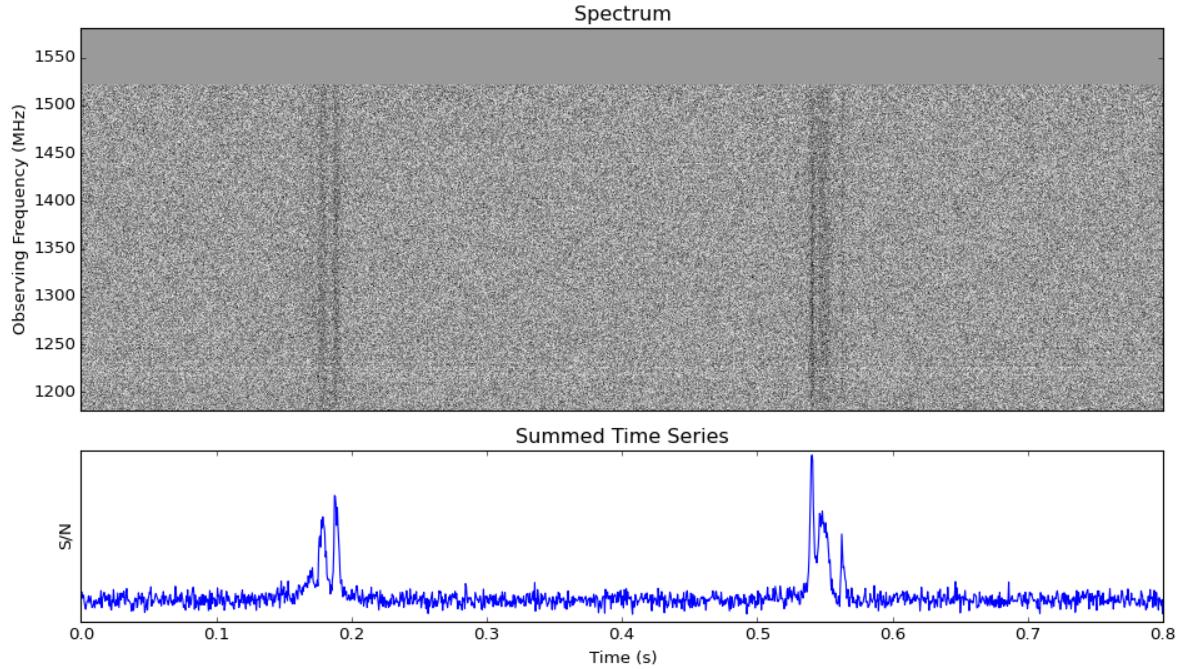


Figure 2.3 The same observation segment of J0738-4042 shown in Figure 2.2, after de-dispersion has been applied.

can be leveraged by more clever algorithms that perform multiple DM trials at a time. The Tree Dedisplacement Algorithm (Taylor, 1974) only performs an amount of additions proportional to  $n_{\text{samp}} \times n_{\text{chan}} \times \log_2(n_{\text{DM}})$ , but requires the dispersion delay to be treated as a linear function of observing frequency, an approximation valid only for small bandwidths. A recently published variant, the Fast DM Transform algorithm (Zackay & Ofek, 2014) lifts that requirement at the cost of a small increase in computational cost. However, the massive computing power brought in recent years by Graphical Processing Units (GPUs) has made the use of smarter de-dispersion algorithms somewhat superfluous (Barsdell et al., 2012), at least for the time being.

### 2.2.2 Fourier Analysis

The Discrete Fourier Transform (DFT) of an evenly sampled time series  $T_n$  containing  $N$  samples is the list of complex coefficients  $F_k$  defined by

$$F_k = \sum_{n=0}^{N-1} T_n \times e^{-2i\pi kn/N} \quad (2.4)$$

with  $0 \leq k < n$  and  $i^2 = -1$  (see e.g. Bracewell, 2000). The  $F_k$  completely describe the original time series in terms of a finite linear combination of complex sinusoids with dimensionless frequencies  $k/N$ . This fact is expressed by the Inverse DFT formula

$$T_n = \frac{1}{N} \sum_{k=0}^{N-1} F_k \times e^{+2i\pi kn/N}. \quad (2.5)$$

If we note  $t_{\text{samp}}$  the sampling interval and  $t_{\text{obs}} = Nt_{\text{samp}}$ , the frequency of sinusoid  $k$  in Hz is  $\nu_k = k/t_{\text{obs}}$  which gives the DFT a frequency resolution of  $1/t_{\text{obs}}$ . The *power spectrum* of the time series  $T_n$  is the list of complex moduli  $|F_k|^2$ .

Inspecting the power spectrum of a noisy time series allows periodic signals it contains to be identified. These would otherwise be invisible directly in the time domain (Figure 2.4). It is also a fast method, which is particularly desirable for the analysis of radio astronomical observations with typical values of  $N \approx 10^6 - 10^8$ : while equation 2.4 suggests at first glance that direct computation of a DFT requires  $\mathcal{O}(N^2)$  arithmetical operations, the highly efficient Fast Fourier Transform (FFT) algorithm (Cooley & Tukey, 1965) performs the task using only  $\mathcal{O}(N \log N)$  by taking advantage of redundant calculations.

Although forming a power spectrum and locating its peaks is a conceptually simple method, some extra post-processing steps are required to maximize its sensitivity to pulsar signals. Detailing all their intricacies is out of the scope of this introduction, so we only enumerate them briefly. For an in-depth review see Lorimer & Kramer (2004).

- **Improving the frequency response of the DFT.** The sensitivity of the DFT is optimal only for signal frequencies that are integer multiples of  $\Delta f = 1/t_{\text{obs}}$ , in which case all the power is contained in a single bin of the power spectrum. Otherwise, power spreads over to neighboring bins, with a worst case scenario being reached for a half-integer frequency  $f = (k + \frac{1}{2})\Delta f$  with an associated loss of Fourier peak amplitude of  $\approx 36\%$ . The simplest solution to reduce this effect consists in padding the original data, which increases the frequency resolution of the DFT without affecting the power spectrum. A more clever and less computationally expensive method for long time series is *Fourier interpolation* (see section 4 of Ransom et al., 2002, for details) with which one can evaluate a complex Fourier amplitude at an arbitrary, non-integer frequency. In particular, *interbinning* is a simplified form of Fourier interpolation that consists of estimating Fourier amplitudes at half-integer frequencies  $k + \frac{1}{2}$  just from their two nearest bins with the formula

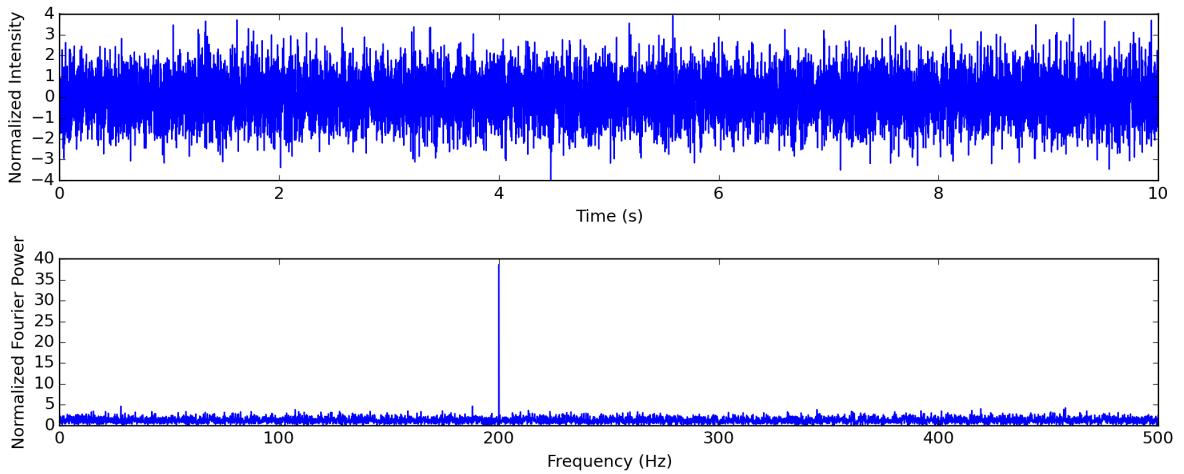


Figure 2.4 Illustration of a DFT-based search for periodic signals. Top panel: noisy time series data containing a 200 Hz sine wave. Bottom panel: power spectrum of the time series clearly revealing the presence of the signal.

$$F_{k+\frac{1}{2}} = \frac{\pi}{4}(F_k - F_{k+1}). \quad (2.6)$$

Using this simple method, the Fourier amplitude of the worst case scenario signal described above is reduced by only  $\simeq 7$  percent.

- **Spectral Whitening.** RFI and fluctuations in data acquisition systems produce noise that does not have perfectly stationary statistics, exhibiting instead a low frequency component known as *red noise*. It can be mitigated by subtracting a running median from the power spectrum.
- **Increasing sensitivity to narrow pulses.** Most pulsars do not emit a pure sine wave but rather a train of narrow pulses. The DFT of such a signal has its power spread between a main peak at the fundamental spin frequency of the pulsar, and a number of *harmonics*, i.e. secondary peaks at integer multiples of that frequency. Additional versions of the power spectrum are created, where each bin has been summed with the first  $m$  of its harmonically related counterparts. Since the actual number of harmonics is not known a priori, this process of *harmonic summing* is repeated for  $m = 2^k$  usually up to  $m = 32$ . All of these integrated power spectra are then evaluated in a final step.
- **Candidate Identification.** Lastly, the statistical significance of observed peaks in the power spectrum are evaluated. Considering a time series of Gaussian distributed

noise, both the real and imaginary parts of its DFT also follow a normal distribution (e.g. Bracewell, 2000). The power spectrum is therefore distributed according to a chi-squared distribution with 2 degrees of freedom, or  $2m$  after a sum of  $m$  harmonics has been performed. From there, the probability of a power spectrum bin to exceed a certain S/N value by chance can be evaluated. A sensible significance threshold corresponds to the case where approximately one false alarm (due to random noise) per observation is returned.

### 2.2.3 Acceleration Searches

In the previous subsection, we implicitly made the assumption that the frequency of the signal we are looking for remains constant over the course of the observation. However, a particularly interesting fraction of the pulsar population resides in binary systems. In that case, variations of the projection of a pulsar's orbital velocity along the line of sight cause the observed frequency of the pulsar signal to vary over time as a result of the Doppler effect:

$$\Delta f(t) = -f_0 \frac{v(t)}{c} \quad (2.7)$$

where  $f(t)$  is the observed signal frequency,  $f_0$  the pulsar's spin frequency,  $\Delta f(t) = f(t) - f_0$ ,  $v(t)$  its line-of-sight velocity in the observer's frame of reference ( $v > 0$  for a pulsar moving *away from* the observer), and finally  $c$  is the speed of light, assuming that  $v \ll c$ .

Properly modeling  $v(t)$  for a Keplerian orbit requires five parameters; in a blind search this would lead to probing a five dimensional space, an undertaking that carries a prohibitive computational cost in practice. The most efficient way to reduce this search space is to use a constant acceleration motion approximation  $\Delta v(t) = at$ . For pulsar searching purposes, it is considered valid as long as the observation time is less than a tenth of the pulsar's orbital period (e.g. Ransom et al., 2002). Using this simple model, and recalling that the width of one DFT bin is  $1/t_{\text{obs}}$ , we find that the pulsar signal will be spread over a number of DFT bins  $Z$  equal to

$$Z = \frac{|a|f_0 t_{\text{obs}}^2}{c}, \quad (2.8)$$

assuming that the pulsar emits a sine wave of frequency  $f_0$ . In reality, the signal contains harmonics of higher frequencies that will spread over a larger number of bins. As an example, the relativistic binary pulsar J0737-3039A (Kramer et al., 2006) has a spin frequency  $f_0 = 44.05$  Hz and reaches a maximum line-of-sight acceleration of  $a = 210 \text{ ms}^{-2}$ .

If observed in this orbital configuration for  $t_{obs} = 18$  min,  $Z$  is greater than 30. The effect is proportional to the pulsar's spin frequency and may be even more dramatic for the fastest millisecond pulsars ( $f_s > 500$ Hz).

It is therefore necessary to remove the deleterious effects of acceleration to discover most binary millisecond pulsars. There are two widely used acceleration search techniques. The most straightforward is *time domain re-sampling* (Johnston & Kulkarni, 1991) which consists of creating a numerically stretched copy of the time series that represents the data as it would be captured by a virtual observer moving with a user defined constant acceleration  $a$ . This process is repeated for many trial acceleration values, and every single time series obtained is searched with a FFT as described above. The second acceleration search method, the so-called “correlation technique” of Ransom et al. (2002) works entirely in the frequency domain. Considering a sine wave with frequency  $f$  and apparent frequency derivative  $\dot{f}$ , this technique relies on the fact that both the real and imaginary parts of its DFT will show a peak smeared in a predictable pattern. By convolving the DFT of an accelerating periodic signal with a range of complex-valued matched filters that reproduce these expected smearing patterns, the power spectrum of the signal may be reconstructed as if it had been emitted by a stationary source. A direct consequence is that only a single FFT needs to be computed for every DM trial, making the frequency domain approach computationally attractive for targeted searches that benefit from longer observation times (Globular Cluster searches in particular).

#### 2.2.4 Pulsar Candidate Folding

It turns out, and it is in fact the main focus of this thesis, that the majority of signals identified in observational data are not of astrophysical origin but rather due to terrestrial radio-frequency interference (RFI) or random noise fluctuations. Submitting an observation to Fourier analysis returns a list of significant periodic signals described by a small number of numerical properties: period, DM, acceleration and signal-to-noise ratio. This limited amount of knowledge is largely insufficient to discriminate between pulsar and non-pulsar sources, and more information has to be gathered through a final processing step, *folding*.

Folding a one-dimensional time series containing a pulsar signal consists of phase-aligning its individual pulses and summing them together. An array with  $n_{bin}$  elements is created to represent the average folded pulse profile and initialized to zero. Every sample of the time series must then be added to the appropriate profile bin. Let  $t_j = j\tau$  be the time at which the  $j$ -th sample has been recorded, where  $\tau$  is the sampling time. Considering

a pulsar with a spin frequency  $f_0$  and line-of-sight acceleration  $a_0$  ( $a_0 > 0$  if accelerating away from the observer), the sample number  $j$  has a pulse phase of

$$\phi_j = f_0 t_j \left(1 - \frac{a_0}{2c} t_j\right) \quad (2.9)$$

where phase is expressed in periods and where we have corrected for the Doppler delay. That sample must therefore be added to the profile bin index  $k_j$  given by

$$k_j = n_{\text{bin}}(\phi_j - \lfloor \phi_j \rfloor), \quad (2.10)$$

where  $\lfloor \cdot \rfloor$  denotes the floor function and  $n_{\text{bin}}$  is the number of profile phase bins. The algorithm can be easily extended to fold consecutive, equal-sized segments of the time series separately, producing a sequence of profiles called *sub-integrations*. Due to the limited resolution of the FFT, the frequency  $f$  of a candidate signal found in a search is at first known only<sup>2</sup> within plus or minus  $1/(2t_{\text{obs}})$ , which usually translates into a visible linear phase drift of the pulse across the sub-integrations (Fig. 2.5, left panel). Finding the true frequency  $f_0$  and true pulse profile of the signal requires correcting the drift by phase rotating every sub-integration. If  $f = f_0 + \Delta f$ , it can be shown using equation 2.9 that the signal of the pulsar in a sub-integration centered at time  $t$  has simply drifted by  $\Delta\phi = \Delta f \times t$ . In the worst case scenario where  $f_0$  lies at the edge of a FFT bin, we have  $\Delta f = \pm 1/(2t_{\text{obs}})$  and  $\Delta\phi = \pm 1/2$  which corresponds to a shift of the last sub-integration of  $\pm n_{\text{bin}}/2$ . The frequency optimization algorithm is therefore:

1. Pick an integer-valued circular shift of the last sub-integration in the interval  $[-n_{\text{bin}}/2, +n_{\text{bin}}/2]$  and calculate the associated signal frequency correction  $\Delta f$ .
2. Apply the corresponding shifts to all sub-integrations.
3. Compute the corrected, summed pulse profile by adding all sub-integrations together.
4. Measure the statistical significance of that summed profile, usually either the chi-squared test statistic or the signal-to-noise algorithm (see Lorimer & Kramer (2004) and section 3.3.4 for details).
5. Go back to step 1 until all shifts have been tried.

The true signal frequency is that which yields the highest profile signal-to-noise ratio (Fig. 2.5, right panel). The optimized sub-integrations array is very useful to visualize any

---

<sup>2</sup>although technically, better frequency resolution can be achieved with interbinning / Fourier interpolation

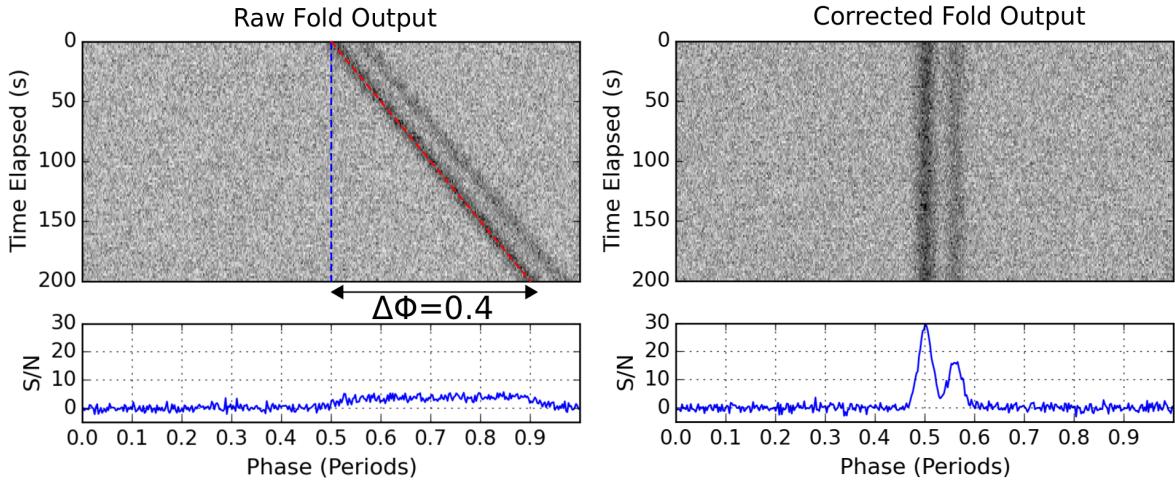


Figure 2.5 Folding a simulated pulsar signal with a true frequency of 0.498 Hz observed over  $t_{\text{obs}} = 200.0\text{s}$ . Due to limited FFT resolution, the signal would be found at an inaccurate frequency of 0.5 Hz. Top-left: sub-integrations of the signal folded at a frequency of 0.5 Hz. Bottom left: summed pulse profile, smeared due to the phase drifting of the pulse. Right panel: the folded profile and true signal frequency are recovered by applying circular shifts to individual sub-integrations.

changes of the pulse profile across time and provides valuable diagnostic information to differentiate between pulsars and interference. A signal-to-noise versus trial spin frequency curve can also be recorded.

In practice, every frequency channel is folded separately, which yields a three-dimensional array or *folded cube* made of a two-dimensional sub-integrations array per channel. By applying circular shifts to individual channels, one can also perform an optimization of the dispersion measure of the candidate following a process similar to that described above, except that shifts now compensate for phase drift due to dispersion delays. This is necessary to produce two more diagnostic plots:

- A characteristic signal-to-noise ratio versus trial DM curve.
- A sub-bands array, which is the sum of the folded cube along the time dimension, and shows changes of the pulse profile as a function of observing frequency.

Finally all the diagnostic information is grouped into large candidate plots which allow a human expert to assess it at a glance. Figure 2.6 shows examples of a known pulsar as it appears in a blind search, and of an interference signal. In general, pulsar signals have a specific signature: they are broadband, have dispersion measures that are strictly positive,

show visible emission during most of the observation and no fluctuations in signal period. RFI truly comes in all shapes and sizes but almost always breaks one of the rules above.

## 2.3 The Pulsar Candidate Selection Problem

---

### 2.3.1 Problem outline

The final step in the pulsar discovery process is to re-observe any credible candidate that has been identified. The ever limited amount of telescope time available requires making a careful selection, as simply re-observing the hypothetical source of every statistically significant periodic signal is obviously not an option. Categorizing pulsar candidates accurately at a glance is actually a relatively easy task that anyone can learn by example with a little domain knowledge (see section 3.3.3). The real problem lies both in the vast amount of candidates that must be dealt with in the case of an all-sky or Galactic plane survey, typically around *several million*, and the extremely low ratio of genuine new pulsars to non-pulsar candidates, rarely exceeding 1 in 100,000 (see Table 2.1).

We are therefore facing a proverbial and tedious needle in a haystack problem. Even at an heroic pace of one candidate classified per second for 40 hours a week, it would take a single person a year to fully sift through 5 million candidates, for a comparatively meager yield of the order of several tens of discoveries, depending on the galactic latitude and size of the sky area covered. Also, it is common practice to search past surveys for binary pulsars, as the computing power required was not available back when they were processed for the first time. In that case the expected number of new discoveries drops by an order of magnitude or more.

There are two solutions to the pulsar candidate selection problem, that are not mutually exclusive: gather more volunteers or transfer the task to a computer program. We briefly review existing methods below.

### 2.3.2 Crowdsourcing

All areas of Astronomy face rapidly increasing amounts of complex data to categorize, and so called “Citizen Science” projects can attract thousands of instances of the most powerful data classifier known to date: the Human brain. Successful examples include the Galaxy Zoo (Lintott et al., 2008), where participants from around the world are asked to classify galaxy images by morphology, or the Planet Hunters project (Fischer et al., 2012) that aims to identify planetary transits in light curves produced by the Kepler mission.

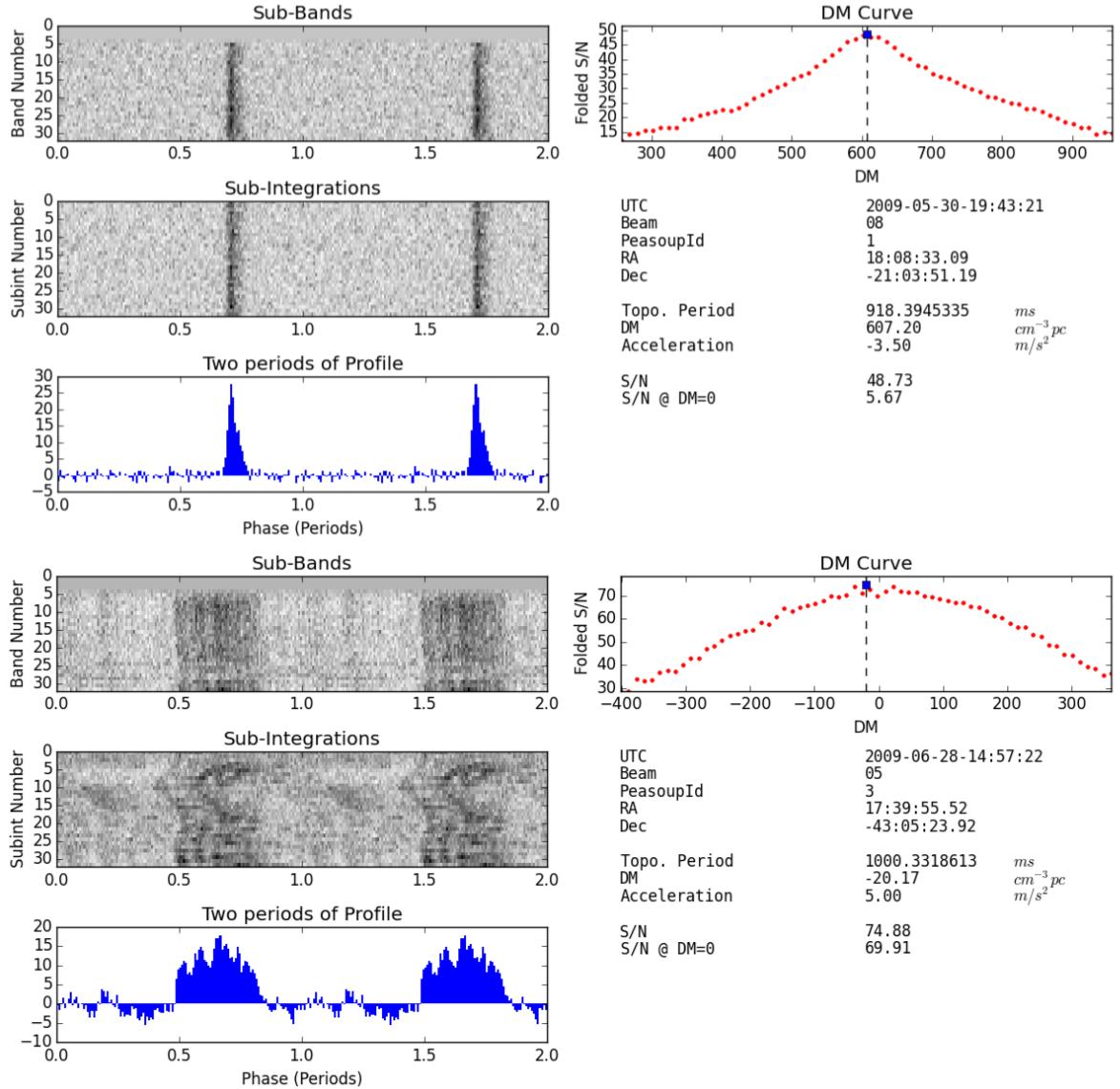


Figure 2.6 Examples of candidate plots. Top: the known pulsar J1808-2057, as found in a blind search of Parkes data. Bottom: candidate showing characteristics of terrestrial interference: dispersion measure consistent with zero and erratic fluctuations in period. Differentiating pulsars from RFI is a relatively easy task for the human eye but a hard one to automate with satisfying accuracy.

Likewise, the most direct way to tackle the pulsar candidate selection problem is to gather a large amount of volunteers and perform a full one-by-one candidate inspection. Crowdsourcing is the approach chosen by the GBNCC survey team, who has enlisted the help of high-school and undergraduate students as part of an outreach program (Stovall et al., 2014). Candidate plots are uploaded to a web-based interactive viewing interface, where each candidate plot is rated by several participants before being labeled as a genuine discovery or a spurious signal.

The main advantage of crowdsourcing is its excellent, possibly unbeatable accuracy. It also provides large sets of expert-labeled candidates that can be used as a gold standard to test automated classification algorithms. Its main downside is that it will certainly not be able to cope with the scale of future pulsar surveys (see Table 2.1).

### 2.3.3 Interactive Selection Software

Usually only a small group of students, if not a single student, is in charge of dealing with survey outputs, in which case visualizing every candidate plot is infeasible. Since at least the days of the Parkes Multibeam Pulsar Survey (PMPS, Manchester et al., 2001), various selection rules have had to be applied prior to human inspection, such as rejecting candidates with a signal-to-noise ratio below a user-defined threshold, or with a period commonly found in many survey observations (a sign of terrestrial interference). However Edwards et al. (2001), searching data also acquired at Parkes, reported that rigid selection criteria led to an unacceptably large number of genuine pulsar signals to be misinterpreted as interference. Likewise, Faulkner et al. (2004), conducting a new search of the PMPS and discovering 128 pulsars previously missed, mentions that earlier analyses of PMPS ignored 40% of the Fourier spectrum to combat interference. Both authors recognized the need to improve selection accuracy, and developed interactive tools for that purpose.

Their principle is illustrated by Fig 2.7, a screen-shot of the REAPER software (Faulkner et al., 2004). All candidates from a sequence of observations are displayed at once as points in a two-dimensional diagram of a chosen pair of parameters, such as period and signal-to-noise ratio in the example given. Candidates due to random noise fluctuations appear at all periods around statistically insignificant S/N values, while actual radio-frequency interference tends to be gathered in narrow period ranges. Genuine pulsar detections are expected to be isolated from such clusters. Selecting a promising data point in the diagram brings up the associated full candidate plot for detailed scrutiny.

Such interactive programs allow their user to implement selection rules adapted to the RFI environment present during a specific sequence of observations. However only two

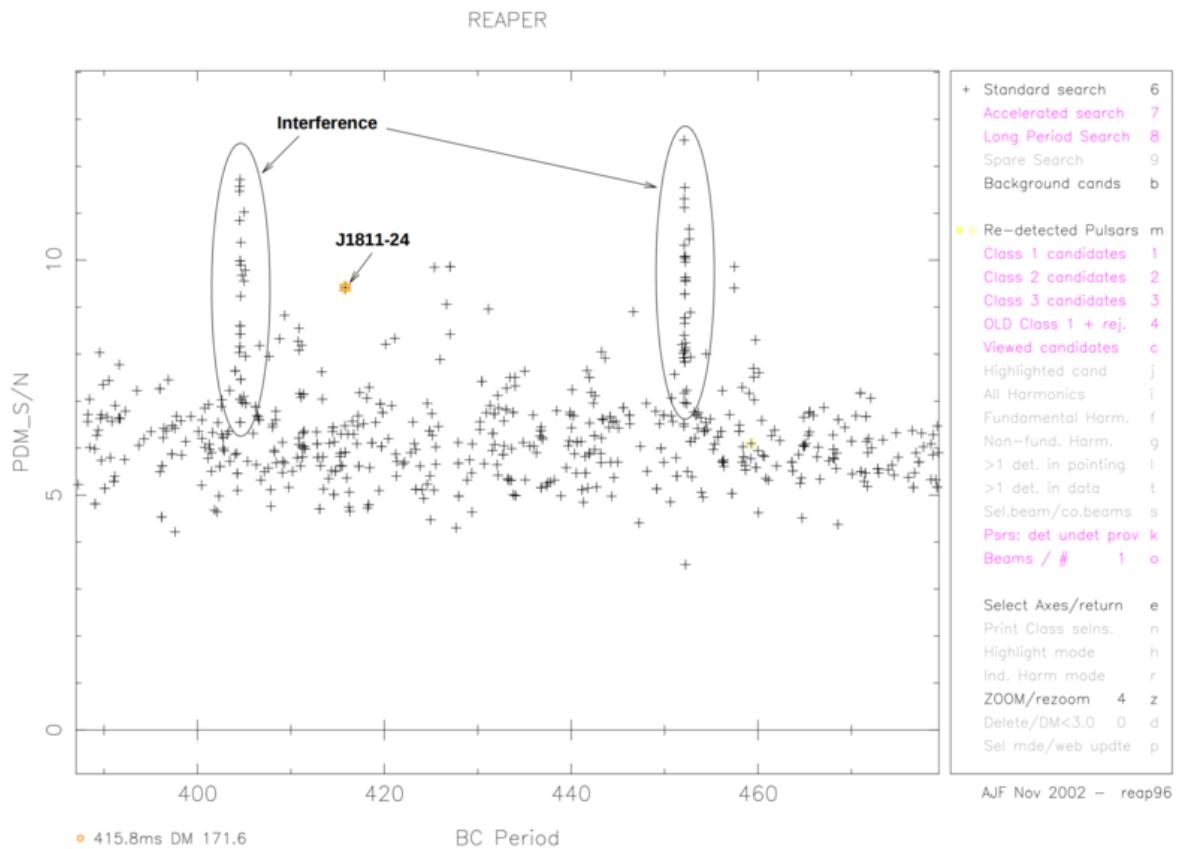


Figure 2.7 Screen capture of the REAPER software taken from Faulkner et al. (2004). Hundreds of pulsar candidates are represented in a signal-to-noise ratio versus period diagram. Interference signals accumulate around specific periods. Pulsars are expected to be clear from such clusters and have statistically significant S/N, such as the detection of J1811-24 highlighted here.

parameters can be visualized at once. The selection is also completely user-dependent and prone to be incomplete, as levels of fatigue and willingness to be exhaustive fluctuate. Neither the speed nor the accuracy of the method have ever been rigorously evaluated in the literature, although Keith et al. (2009) estimates that to really avoid missing a single detectable pulsar, the user would still have to visualize a quarter of the full candidate plots. This does not represent a significant enough reduction of the overall burden.

### 2.3.4 Classification Algorithms

The most attractive solution is to fully transfer the task to a computer program, but as mentioned above, implementing a rigid set rules is inefficient, simply because mathematically defining what makes a candidate acceptable for confirmation is at best difficult and survey-specific. Machine Learning (ML) algorithms (see next chapter) are a very attractive solution as they are able to automatically infer complex classification rules from a sample of data, imitating a process similar to a person learning the task based on examples (see next chapter). They are a very promising solution to the pulsar candidate selection problem.

The first attempt to use supervised Machine Learning on the pulsar candidate selection problem is that of Eatough et al. (2010), who used an Artificial Neural Network (ANN) to sort the output of a search of the PMPS. Candidates were turned into a vector of 12 numerical features designed by the authors, from which the algorithm tried to predict the category to which each candidate belongs. These features included basic parameters such as period, dispersion measure, signal-to noise ratio, etc. While demonstrating the potential of the approach, and discovering one pulsar that would likely have been missed using older methods, their ANN implementation did not reach a sufficient level of accuracy to replace human intervention. Tested on PMPS data, it rejected 99% of non-pulsar candidates, while correctly identifying 92 % of a sample of known pulsars as such. Following up on that work, Bates et al. (2012) tried to increase the amount of information captured by the input vector by increasing the number of features to either 22 or 27. However, no significant improvement in accuracy was achieved.

Two more recent papers have taken slightly different approaches. Zhu et al. (2014) used so-called “Deep Learning”, relying on a type of large neural network designed to process image data directly (e.g. Lecun et al., 1998). In particular they used the full sub-integrations and sub-bands plot of a candidate as inputs without pre-processing. Tested on a sample of candidates containing 56 pulsars, it was able to identify them all with a false positive rate of 3.8%. On the other hand, Lee et al. (2013) reached a nearly

Survey	Candidates	Discoveries
Swinburne Mid Latitude <sup>1</sup>	6,000,000	69
Swinburne High Latitude. <sup>2</sup>	4,500,000	26
HTRU Mid Latitude <sup>3</sup>	9,500,000	102
PMPS (2013 re-analysis) <sup>4</sup>	16,000,000	16
GBNCC <sup>5</sup>	1,200,000	67
SKA1-Mid Galactic Plane <sup>6</sup>	900,000,000	7,000
SKA1-Mid All-Sky <sup>6</sup>	15,000,000,000	11,400

<sup>1</sup> Edwards et al. (2001)

<sup>2</sup> Jacoby et al. (2009)

<sup>3</sup> Keith et al. (2010); Levin (2012). Chapter 4 describes our re-analysis of this survey.

<sup>4</sup> Eatough et al. (2013)

<sup>5</sup> Stovall et al. (2014)

<sup>6</sup> Dewdney (2013); Keane et al. (2015)

Table 2.1 Amount of candidates produced and number of previously published pulsar discoveries made by some recent and future wide-area surveys. Values for surveys to be performed with the Square Kilometer Array (SKA) are only estimates based on pulsar population studies, a beam size of  $2 \times 10^{-4}$  deg, and 50 candidates per beam. SKA1 refers to the first phase of the project that uses only the core dishes of the array.

identical level of performance with their algorithm PEACE, an expert-crafted scoring formula based on only 5 candidate features. While it does not technically fall under the banner of Machine Learning, PEACE is conceptually similar, the main difference being that the scoring formula is adjusted manually instead of being algorithmically inferred from a sample of data.

All of the previously published algorithms represent a large step forward in reducing the amount of candidates to be visually analyzed, and both PEACE and PICS have helped to discover several dozens of new pulsars. However they either miss a significant fraction of a sample of known pulsars, or let through several percent of the whole candidate population as false positives, which could still be an overwhelming amount of data to visually analyze in the case of future surveys. There is a need to push the accuracy of automated classification systems significantly further.

### 2.3.5 The case of the Square Kilometre Array (SKA)

The Square Kilometer Array (SKA) is an international, multi radio telescope project bound to revolutionize radio astronomy upon completion. One of its key science goals is to carry out wide-area pulsar surveys that will greatly increase the known Galactic pulsar population, with more than 10,000 new objects expected to be found during the first phase

of the project (with a limited number of dishes enabled, henceforth referred to as SKA1) (Keane et al., 2015). The high sensitivity achieved in short  $\simeq 10$  min integration times will be particularly favourable to the discovery of extreme binaries and pulsar-black hole systems in particular.

The SKA will bring numerous computational challenges; of particular interest is the likely impossibility of storing the raw data due to its excessive size, requiring the pulsar search process to be performed in real time, and the overwhelming amount of pulsar candidates involved. Assuming a beam size of  $2 \times 10^{-4}$  degrees at 1400MHz (Dewdney, 2013) and the production of 50 candidates per observed beam, a search for pulsars in the galactic plane conducted with the SKA is likely to generate about a billion candidates, and at least one order of magnitude more for an all-sky survey. False positive rates of 1% are completely unacceptable here. If one is to extract all the value from the data acquired by the SKA, it is a necessity to eliminate human intervention in pulsar candidate selection in present searches before the SKA even sees its first light. This is one of the primary motivations for the development of SPINN, which is described in the following chapter.



# 3

## SPINN: A Machine Learning Algorithm for Pulsar Discovery

This chapter describes the main achievement of the thesis: the pulsar candidate classifier SPINN, an acronym that stands for Straightforward Pulsar Identification using Neural Networks. But more importantly we will attempt to detail the design choices that led to its success, and some general principles that can be applied to solving other similar Machine Learning problems.

### 3.1 Introduction

---

There is no standard definition of Machine Learning (ML) despite its wide use, but in a nutshell, it is a sub-field of statistics and computer science that covers the study and design of algorithms capable of finding meaningful structure and/or making predictions on data. The best way to break the nebulous character of the previous sentence is to jump directly into concrete examples.

#### 3.1.1 A Quick Tour of Machine Learning Tasks

All learning problems start with a set of data instances of identical nature. These may be complex objects such as text snippets, images, sound recordings or pulsar candidates. As a very first step they must be turned into a set of descriptive numbers called *features*, so that an instance is now represented as a point, or vector  $\mathbf{x}_i \in \mathbb{R}^n$ . From there, a number of learning tasks can be performed, of which we provide a few common examples without trying to be exhaustive. They fall into two main categories:

- **Unsupervised Learning.** The general goal is to find hidden structure in the data set. *Clustering* is about organizing data instances into groups, by linking pairs of

nearby points, or identifying dense areas. *Anomaly detection* is the identification of points lying outside what would be a dominant statistical distribution. *Dimensionality reduction* consists in searching for a lower-dimensional space, surface, or manifold close to which most of the data are found; in this case points can be described by a smaller set of numerical parameters. Obtaining a more compact representation can be a goal in itself, or a preliminary step to facilitate the application of another learning algorithm.

- **Supervised Learning.** In the context of supervised learning, every point is associated with a desired output value  $y_i$  defined by the user. The goal is to infer a function  $f$  from the data that closely maps the input vectors  $\mathbf{x}_i$  to the  $y_i$ . The shape and complexity of  $f$  depends on the specific algorithm used. The task is called *regression* when the desired output value is a continuous variable, and *classification* when it is categorical.

Regardless of the learning problem, algorithms are first presented a sample of data on which their internal parameters are calibrated, a process called *training*, and then applied to new, previously unseen data from which they can generate sensible predictions.

### 3.1.2 Solving a Machine Learning Problem

Applying ML to a particular problem can be summarized by the following process, whose steps will be detailed later in this chapter:

1. Gather as much training data as possible. In the case of supervised learning, desired outputs must be checked carefully.
2. Determine the appropriate type of learning task (supervised or unsupervised), and the specific algorithm to be used.
3. Design features to represent the data instances. This is an absolutely critical step, as features capture all the information ever manipulated by the learning algorithm. We discuss feature design in section 3.3.
4. Train the algorithm on the set of input vectors obtained in the previous step. While all common algorithms are now packaged in ready to use libraries, a detailed understanding of its inner workings is helpful in making better design choices, and obtaining higher performance. We cover Neural Networks in section 3.2.

5. Evaluate the performance of the algorithm rigorously. Great performance is relatively easy to obtain on the specific training data set used, but what needs to be measured is the ability to make good predictions on data not seen before (see section 3.4).

Returning to the specific case of pulsar candidate classification, we are faced with rare and precious data points (pulsars) spread among an overwhelming number of undesirable ones (interference). Unsupervised learning, and anomaly detection methods in particular may look attractive at first sight. However, it is unlikely that pulsars constitute the most anomalous data points among candidates, given the much larger number and wide variety of radio-frequency interference instances. We are therefore left with supervised learning on one hand and clustering algorithms on the other, but the true goal of the latter is to organize the data points in an unspecified number of “natural categories”, while we are solely interested in distinguishing genuine pulsars from the rest. It is also easier to specify what characteristics truly make a pulsar and then define interference as anything that does not fit those criteria, rather than the other way around. For these reasons, supervised learning is the most straightforward and better choice here. Yet unsupervised learning could possibly find good uses such as categorizing spurious candidates, with a goal of identifying specific terrestrial interference sources and developing targeted counter-measures against them.

As for the algorithm, Neural Networks were chosen for two reasons: my own familiarity with them and the fact that they had already been applied in the pulsar literature to candidate classification. That way any improvement in accuracy could be more readily attributed to a better approach to the problem rather than simply a change of algorithm.

## 3.2 Artificial Neural Networks

---

Artificial Neural Networks are a family of computing models with a wide range applications in Machine Learning. In this section we cover specifically their use on supervised learning problems.

### 3.2.1 Mathematical Model

An artificial neuron is a computing unit with an arbitrary number  $n$  of real-valued inputs and a single output, inspired by its biological counterpart (Fig 3.1). A neuron performs a two-step calculation made of a weighted sum of its inputs, followed by the application of a real-valued *activation function*  $f$  so that

$$a = f \left( \sum_{i=1}^n w_i x_i + b \right) = f(\mathbf{w} \cdot \mathbf{x} + b) \quad (3.1)$$

where the  $w_i$  are the *weights* of the neuron,  $b$  its *bias*, and  $a$  its output value or activation value. This expression can be written as a more compact scalar product between a weights vector  $\mathbf{w}$  and an input vector  $\mathbf{x}$ .  $f$  should be at least non-linear<sup>1</sup>, continuous and differentiable, although on real world problems only a few different activation functions are used. Its choice is motivated by practical considerations such as computational speed, numerical stability and the nature of the learning task. Our focus is binary classification, and a commonly used activation function is the logistic sigmoid

$$f(z) = \frac{1}{1 + \exp(-z)}. \quad (3.2)$$

Its output is bound between 0 and 1 (Fig. 3.1), and these two values can be made to correspond to a negative and positive data class respectively.

Neurons can be chained into networks of any shape or size, but in most applications they are stacked into successive fully connected layers (Fig. 3.2), where computation flows in a single direction. In this configuration, computing the activation values of a whole layer of neurons on several input vectors becomes simply a matter of multiplying matrices. For a fully-connected layer of  $n$  neurons each with  $p$  inputs, equation 3.1 can be vectorized as

$$\mathbf{A} = f(\mathbf{W} \cdot \mathbf{X} + \mathbf{B}). \quad (3.3)$$

$\mathbf{W}$  is a  $n \times p$  weights matrix, each of its rows containing the weights of an individual neuron from the layer.  $\mathbf{X}$  is the input matrix of dimensions  $p \times m$ , containing an arbitrary number  $m$  of different input vectors arranged in columns.  $\mathbf{B}$  has dimensions  $n \times m$  and it has uniform rows such that  $B_{ij} = b_i$ .  $f$  is applied element-wise, and the final result is the  $n \times m$  output value matrix  $\mathbf{A}$ , where  $A_{ij}$  is the output of the  $i$ -th neuron on the  $j$ -th input vector<sup>2</sup>.

The convenience with which feed-forward networks can be represented as basic linear algebra operations is not their single strong point. It can be shown that under some mild assumptions a simple two-layer network can approximate any multivariate, real-valued

<sup>1</sup>Non-linearity is a requirement, otherwise the overall activation function of any number of neurons applied in succession will also be a linear function of the inputs, defeating the whole point of building networks of neurons.

<sup>2</sup>In practice one can append a fictitious input equal to 1 (called the intercept term) to every input vector, and treat the bias term of every neuron as an extra weight that applies to the intercept term. Then  $\mathbf{B}$  can effectively be dropped from equation 3.3

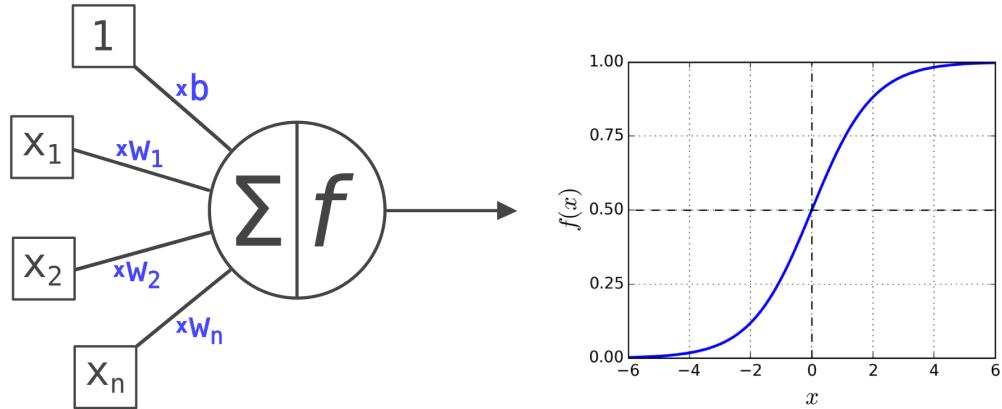


Figure 3.1 Left panel: artificial neuron model, showing its two computation steps: a linear combination of inputs followed by the application of a real-valued activation function  $f$ . Right panel: the logistic sigmoid function, a common choice for  $f$ .

function to an arbitrary level of precision provided there is no limit on the number of neurons in the first layer (Cybenko, 1989). This makes them a prime tool for supervised learning.

### 3.2.2 Important Visual Intuition

It is particularly enlightening to think of the response of a neuron in geometrical terms. Going back to equation 3.1, we notice that  $\mathbf{w} \cdot \mathbf{x} + b = 0$  defines a hyperplane  $H$  in input space, whose orientation is defined by its normal vector  $\mathbf{w}$ , and location by  $b$ . For any input vector  $\mathbf{x}$ , the output of the neuron depends only on the projection of  $\mathbf{x}$  along the direction defined by  $\mathbf{w}$ : it takes values greater than 0.5 on one side of  $H$ , and lower than 0.5 on the other. The left panel of Fig. 3.3 shows this for a logistic sigmoid neuron with two inputs, trained on a linearly separable toy dataset. It can be noted that  $\mathbf{w}$  does not only define the orientation of  $H$ : the norm of  $\mathbf{w}$  is a meaningful parameter on its own and controls the steepness with which the output value rises as  $\mathbf{x}$  crosses  $H$ , or in other words, the sharpness of the separation.

In a two-layer feed-forward network, output neurons can combine the hyperplanes defined by those in the first layer, to produce output patterns of unlimited complexity; the right panel of Fig. 3.3 provides a visual illustration of the universal approximation capability of neural networks mentioned above. But most importantly it shows that *binary classification is about carving separation boundaries in input space*: this is a crucial intuition to keep in mind when choosing how to represent data instances as feature vectors (see section 3.3).

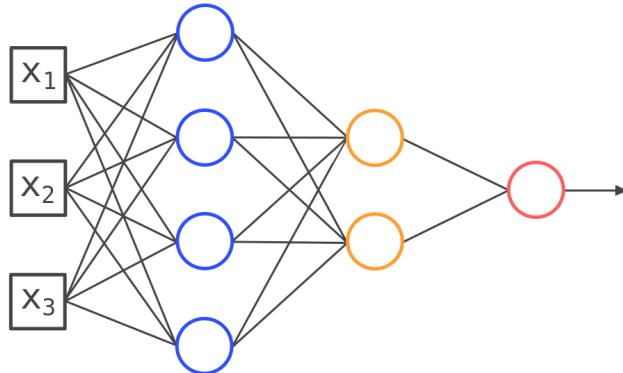


Figure 3.2 A fully connected feed-forward network, here with three inputs, one output, and three neuron layers (bias terms are not shown for clarity). Computation flows from left to right. In this network configuration, all operations can be handily represented as a succession of matrix multiplications and element-wise transformations.

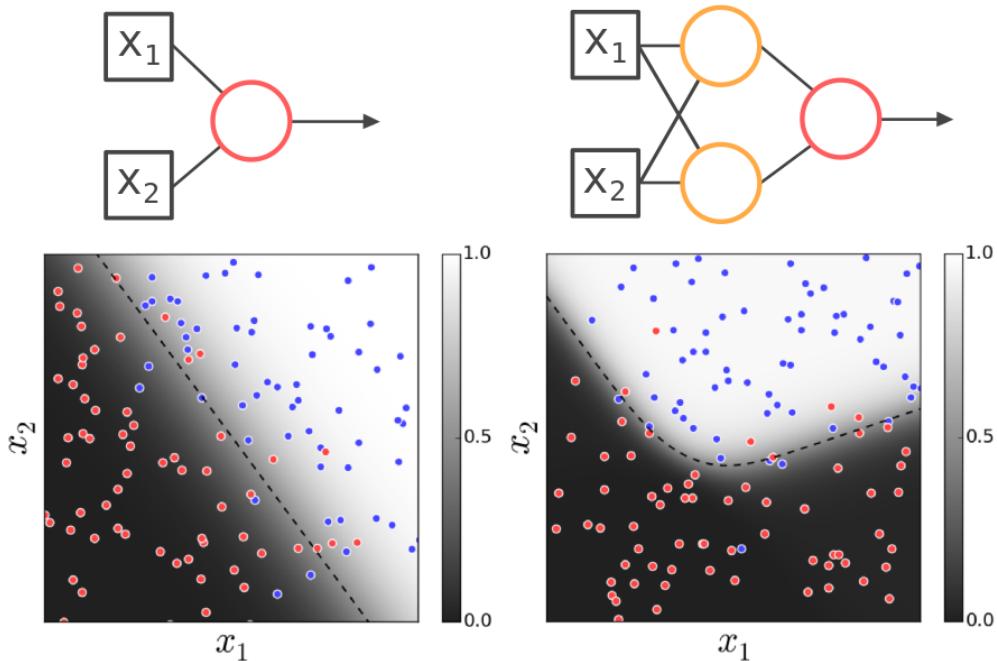


Figure 3.3 Left panel: output value of a logistic sigmoid neuron as a function of its two inputs, after having been trained on a toy dataset. Positive class points are shown in blue, negative class ones in red. Lighter background colors correspond to a higher output value, which takes its intermediate value of 0.5 on the dashed line, an illustration that a neuron defines a separation plane in input space. Right panel: output of a 2:1 feed-forward network trained on another data set. The planes defined by the neurons in the first layer are visible, and combined by the output neuron as a single boundary.

### 3.2.3 Training

Training is the process of inferring the optimal network weights and biases from a sample of data points. It is posed as a problem of minimizing a *cost function*, that measures the discrepancy between the desired and actual outputs. The simplest cost function is the mean squared error

$$C = \frac{1}{m} \sum_{j=0}^m (a_j - y_j)^2 \quad (3.4)$$

where  $m$  is the number of training data points,  $a_j$  the output of the network on input vector  $\mathbf{x}_j$  and  $y_j$  the desired output value.  $C$  must be optimized with respect to the weights and biases of the whole network, which is performed using gradient descent; starting from a random set of weights and biases, the three following steps are repeated in succession:

1. Compute the error values  $a_j - y_j$  on all training examples
2. Compute all the partial derivatives of the cost function with respect to the network parameter (weights and biases) using the *backpropagation* algorithm (see below).
3. Adjust every network parameter  $w$  with the update rule,

$$w \leftarrow w - \eta \frac{\partial C}{\partial w}, \quad (3.5)$$

where  $\eta > 0$  is a user defined parameter called the *learning rate*.

Calculating the partial derivatives of the cost function fortunately does not rely on finite difference methods. A full description of the backpropagation algorithm is out of the scope of this chapter, but can be found in a number of textbooks on neural networks such as Rojas (1996). Simply put: in a feed-forward network, it is possible to obtain the derivatives with respect to the weights of a given layer directly from those of the layer right above (or from the error values in the case of the final, output layer). In a sense, the error values computed in the first step are propagated backwards, in a series of operations that can also mostly be represented as basic linear algebra. The very existence of the backpropagation algorithm makes the training of very large networks possible.

One can note that the weights obtained at the end of training are optimal with respect to the specific sample of training data chosen, but not necessarily to new data. This has important consequences covered later.

### 3.2.4 Implementation Details

In practice, training a neural network can be particularly difficult, with a number of issues arising mostly from the use of gradient descent to optimize a cost function that is in general not convex. A large arsenal of tricks has been developed to make training converge faster to better solutions, a number of which are reviewed by LeCun et al. (2012) and briefly enumerated below.

- **Stochastic Gradient Descent.** This consists in presenting only a small and random fraction of the data to the neural network during a training step. This usually leads to much faster convergence especially with large and redundant data sets. More importantly, the introduction of randomness in the training process helps to prevent it from getting stuck in bad local minima of the cost function.
- **Choice of Learning Rate.** While small values slow the convergence of the training process, excessively large ones will lead gradient descent to overshoot minima of the cost function and eventually diverge. There is no general method for choosing an optimal learning rate except trial and error. When stochastic gradient descent is used, a useful technique is to start with a slightly excessive learning rate but make it decay over time, by multiplying it by a constant  $< 1$  after one gradient descent iteration is complete. That way a larger portion of the weight space is explored, while forcing training to eventually converge.
- **Regularization.** Networks with large numbers of neurons may fit the training data too closely, capturing not only relevant relationships between inputs but also sampling noise, and then lose predictive power on new data (see also next section). In supervised learning, regularization refers to methods of mitigating this effect. For neural networks they involve limiting the absolute values of the network weights, which intuitively corresponds to softer and less complex separation boundaries (Fig. 3.3). We used the *max-norm* constraint (e.g. Srivastava et al., 2014): after a training step, any neuron with a weights vector length exceeding a predetermined value  $w_{max} > 0$  had its norm forcibly adjusted to lie on the hypersphere of radius  $w_{max}$ . This obviously introduces another tunable network parameter. Its optimal choice is discussed in section 3.4.
- **Feature Scaling.** Training converges faster if all individual network inputs have zero mean and unit variance over the training data set. Outliers also make training slower or even unstable: a logistic neuron with just one very large input will see its

activation value saturated to 1, regardless of the other inputs, causing some components of the cost function gradient to become negligible. A reasonable guideline is that a given feature should not take values spanning several orders of magnitude.

To obtain maximum control over the training process, the neural network used by SPINN was written from scratch in Python and implements all of the techniques described above.

### 3.3 Feature Design

---

A critical part of solving any ML problem lies in devising an appropriate feature representation for data instances, as they carry all the information ever sent to the learning algorithm as inputs. We start by exposing some general principles, then move on to the problem of pulsar candidate selection, explaining the usual thought process of a human expert evaluating a candidate, and how we translated that into a feature set for SPINN.

#### 3.3.1 General Guidelines

On binary classification problems, we have seen earlier that supervised learning algorithms define separation boundaries in feature space (Fig. 3.3), whose complexity is controlled by a number of tunable parameters; in the case of neural networks those are the number of neurons, the number of neuron layers, and possibly extra regularization parameters (section 3.2.4). Based on this visualization, an ideal feature set should simply be such that the underlying distributions of both classes of data points are non-overlapping. We obviously tried to design features that were predictive of the class label, but also monotonic (see Fig. 3.4 for an illustration): that is a feature such that an increase in its value always corresponds to an increased (or decreased) probability of positive (or negative) class membership. This is not a requirement *per se*, but it greatly facilitates the training of most learning algorithms. When using an entirely monotonic feature set, one data class will be found in a single corner of the feature space, requiring a decision boundary of low complexity.

Given a large set of features, Peng et al. (2005) provides a quantitative method for selecting an optimal subset of them called *minimal-redundancy maximal-relevance*, although in the case of pulsar classification the scale of the problem was small enough that an exhaustive trial and error approach was possible. In any case, carelessly increasing the number of features is detrimental to classification accuracy.

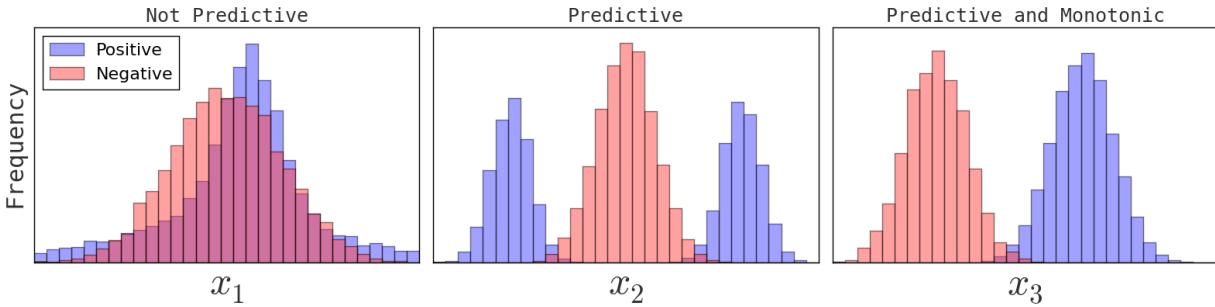


Figure 3.4 Simulated feature distributions for a binary classification problem, showing varying usefulness. Left panel: knowledge of  $x_1$  brings little information about class membership and is likely to be harmful if used as an input to learning. Middle and right panels: examples of highly predictive features. Although  $x_2$  and  $x_3$  carry comparable amounts of information about the class label,  $x_3$  is monotonic and more easily learned by most supervised algorithms including neural networks.

### 3.3.2 The Bias-Variance Dilemma

A general problem in supervised learning is the existence of two sources of error that prevent algorithms from learning a model that generalizes beyond their training set:

- **Bias** refers to systematic errors. An algorithm has high bias if it consistently fails to capture relationships between the desired output and the features, due of a lack of free model parameters. This situation is also called underfitting.
- **Variance** comes from sensitivity to noise specific to the training data at hand. An algorithm has high variance if it learns drastically different models or decision rules on different training sets, as a consequence of having too many free parameters. It is also said to be overfitting the data.

These two sources of error cannot be simultaneously minimized, a situation referred to as the Bias-Variance Dilemma or Tradeoff, reviewed in-depth in e.g. Hastie et al. (2009). The typical behavior of supervised algorithms is shown in Fig 3.5. A perfect result on the training set can be achieved with a sufficiently complex model that will only fail to generalize on new data. An inescapable source of variance is the finite nature of the sample of training data, which cannot perfectly capture its true underlying distribution. Increasing the number of features leads to a sparsely sampled feature space where the true relationship between desired output and inputs becomes ambiguous or even undefined.

Overall, the size of the training data set limits the amount of model complexity that can be used, and the pulsar candidate classification problem is arguably data-poor: any

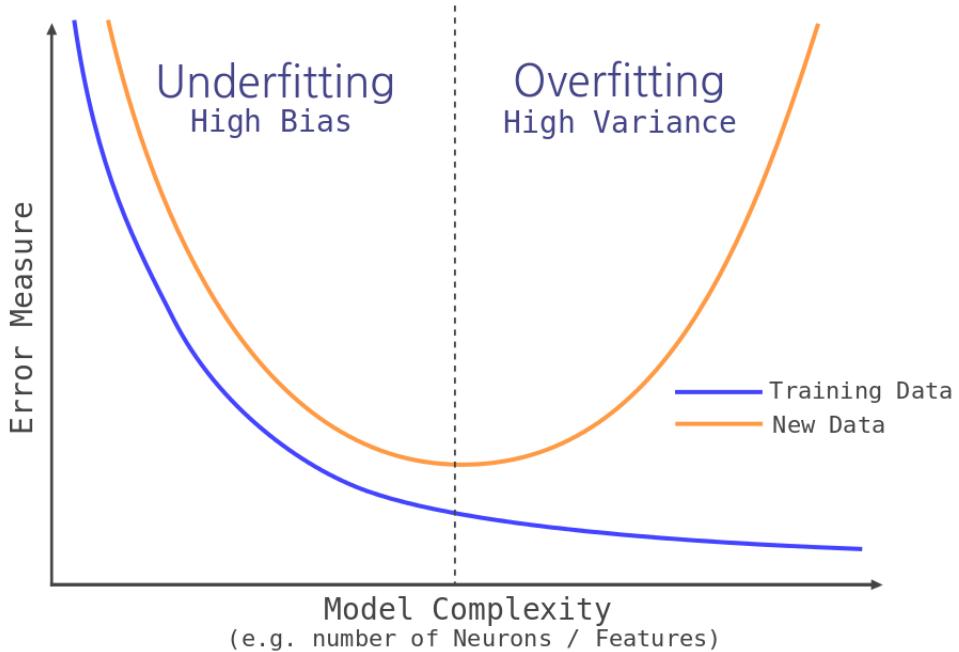


Figure 3.5 The bias-variance dilemma in practice: typical behavior of supervised learning algorithms on a given problem. In particular, adding features past a certain point is harmful, as the distribution of the data gets increasingly undersampled, causing the algorithm to model noise specific to the training set. Using more complex models requires more training data, but the supply of unique pulsars observable by any survey setup is limited.

current state-of-the-art survey setup may be able to observe at best 1,000 unique pulsars<sup>3</sup>. That amount must be put into perspective with the size of some popular data sets on which deep neural networks have had great success in recent years, for instance the CIFAR-10 (60,000 32x32 pixel images representing 10 classes of objects) or the Street View House Numbers (SVHN) data set (600,000 images of digits)<sup>4</sup>. As a consequence we decided to design a pulsar classifier as simple as possible, and in particular to keep the feature count as low as possible.

### 3.3.3 A Short Pulsar Candidate Inspection Manual

Before detailing the feature set used by SPINN it is useful to briefly explain the basic rules of visual candidate inspection. The first aspect to check is the candidate's signal-to-noise ratio (S/N, see next section for details), a measure of folded pulse profile significance

<sup>3</sup>That number was closer to 200 for past publications on the pulsar candidate selection problem. We had access to over 500 for the present work.

<sup>4</sup>Rodrigo Benenson maintains a handy web page describing the classical datasets used by the Machine Learning research community, and leader boards keeping track of the best classification accuracy achieved at [http://rodrigob.github.io/are\\_we\\_there\\_yet/build/](http://rodrigob.github.io/are_we_there_yet/build/)

related to the probability of it appearing by chance from random noise fluctuations. A reasonable threshold in S/N is set according to the number of DM, acceleration and period trials performed by the search pipeline, and adjusted based on experience from past surveys. A conventional choice is S/N = 10, with exceptions being made for the most convincing candidates.

In an ideal interference-free environment, S/N would be the only selection criterion to distinguish genuine sources from background noise. In practice, the presence of artificial signals must be taken into account. Since their sources are terrestrial, their true dispersion measures (DM) are exactly zero, which seemingly provides us with a second, simple decision rule. However the DM value attributed to a candidate is that which maximizes its S/N and it carries an uncertainty proportional to its period (see 3.3.4). Some common types of narrow-band RFI also appear to have a significantly non-zero DM. As a consequence it is necessary to consider the shape of the signal, and look for pulsar-like characteristics:

- Narrow pulse profile. Most pulsars have profiles with a single peak lasting no longer than a few percent of their periods, while RFI profiles tend to look like sine waves. There are diverse exceptions to this general trend, in particular millisecond pulsars that may have wide, multi-peaked profiles.
- Broadband signal. Pulsars emit across the whole frequency band observed. The strength of their emission may be noticeably non uniform in frequency, as a consequence of either scintillation or the usually negative spectral index of the pulsar. Signals visible in only one or two narrow bands can be safely discarded as interference.
- Stable signal. RFI signals can have wildly varying shapes over the course of an observation, while those of pulsars are always consistent when integrated over a few dozen periods.

A number of common examples are shown in Fig. 3.6

### **3.3.4 List of Features Used**

Given the considerations outlined above and knowledge of the general pulsar searching process, we enumerate and justify the list of features used by SPINN below.

1. **Best signal-to-Noise ratio (S/N) of the candidate.** S/N is a measure of folded profile significance (Lorimer & Kramer, 2004). Consider a profile with  $n$  phase bins

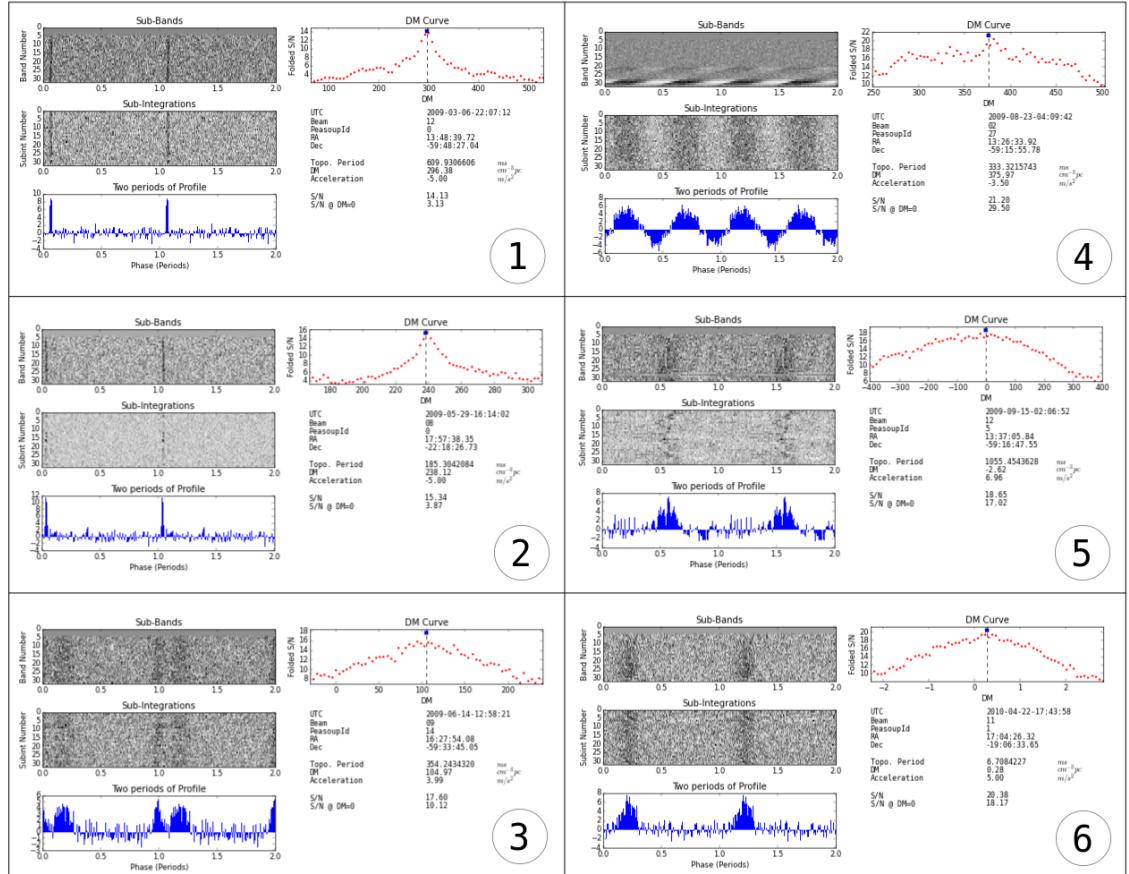


Figure 3.6 Comparison between examples of pulsar (left column) and RFI candidate plots (right column). #1 has the appearance of a typical ordinary pulsar, #2 shows significant variations in intensity over time, and #3 has a wide double-peaked profile. #4 is visible only in two narrow bands of frequency and is erroneously detected with a high DM, #5 exhibits erratically changing pulse phase over time, and #6 looks like a perfect millisecond pulsar except for its reported dispersion measure consistent with zero.

$p_i$ , whose off-pulse mean and standard deviation are respectively  $m$  and  $\sigma$ . We aim to find a subset  $P$  of contiguous bins deviating the most from this off-pulse distribution.  $P$  is defined by two parameters, a starting phase bin index  $i$  and number of bins  $W$  (width). The signal-to-noise ratio of  $P$  is defined by

$$\text{S/N}(i, W) = \frac{1}{\sigma\sqrt{W}} \sum_{j \in P} (p_j - m). \quad (3.6)$$

$m$  and  $\sigma$  are estimated either from the original time series data or from a part of the profile containing no signal. An exhaustive search in  $i$  and  $W$  is performed to find the best signal-to-noise ratio of the candidate, along with the optimal pulse width. Fig. 3.7 shows the output of the search algorithm on an example. A justification for the normalization factor in the formula above is that considering a profile where every bin  $p_i$  follows a normal distribution  $\mathcal{N}(m, \sigma^2)$  (i.e. a profile made of Gaussian noise), then for any value of  $W$  the quantity defined by equation 3.6 follows  $\mathcal{N}(0, 1)$ .

In our data, candidate S/N values range from approximately 5 to 3,000. Because of this wide spread, using them directly as an input to a neural network can cause problems (see 3.2.4). Therefore, we use the logarithm of the best candidate S/N as a feature. This extra scaling step may not be necessary when using other learning algorithms.

2. **Equivalent Duty Cycle.** From the best pulse width  $W$  obtained above we can derive a related quantity  $W_{\text{eq}}$ , the *equivalent width*. This is the width of a virtual top-hat pulse of same peak height and total area as the original pulse (see Fig. 3.7). From there, the equivalent duty cycle  $d_{\text{eq}}$  is simply the ratio of  $W_{\text{eq}}$  to the number of bins in the profile.

Typically pulsars emit short pulses with duty cycles of a few percent, while RFI often exhibit sine-shaped profiles. There are exceptions to this general trend, and this feature risks introducing a bias against high duty cycle pulsars, but its positive impact on classification accuracy is too important to ignore.

3. **CIDER.** Ultimately, our problem is about deciding if a candidate is of astrophysical origin or not, and a strictly positive DM value guarantees the former, at least in theory. However, the folding software assigns to a candidate the DM value that yields the highest profile signal-to-noise ratio, a process that is intrinsically noisy. We need to attribute a confidence index in the *true* DM value being positive.

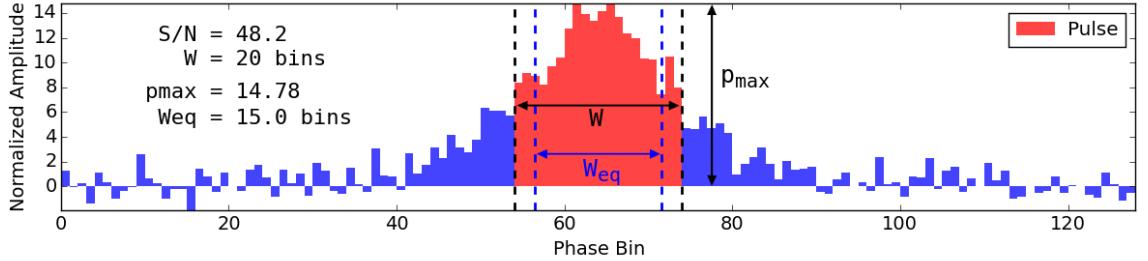


Figure 3.7 Example output of the signal-to-noise algorithm run on an artificially generated pulse profile, such that the background noise has zero mean and unit variance. If  $p_{\max}$  is the maximum of the profile, the equivalent width  $W_{eq}$  of the pulse is the quantity such that  $p_{\max} \times W_{eq}$  is equal to the pulse area in red. The equivalent duty cycle is the ratio of  $W_{eq}$  to the total number of phase bins.

The first step here is to estimate the uncertainty on the DM value of a candidate. We gathered a sample of known pulsars observations that appeared in our data, which had published “true” dispersion measures reliably obtained via pulsar timing. Our intuition was that on average, the difference between the DM reported by the folding pipeline and the true DM would be proportional to the equivalent width of the pulse. To test this, we plotted the distribution of the quantity

$$E_{DM} = \frac{DM_{fold} - DM_{true}}{W_{eq}}, \quad (3.7)$$

which is shown in Fig. 3.8 and is close to Gaussian with a standard deviation of  $\sigma_E = 235$ . This justifies defining the following confidence index that the DM returned by the folding software  $DM_{fold}$  exceeds a threshold value  $DM_{min}$  below which no pulsar is expected to be discovered:

$$I = \Phi \left( \frac{DM_{fold} - DM_{min}}{\sigma_E W_{eq}} \right) \quad (3.8)$$

where  $\Phi$  is the cumulative distribution function of the standard normal distribution. In practical terms, the quantity in brackets can be seen as a number of standard deviations separating  $DM_{fold}$  and  $DM_{min}$ . The confidence index  $I$  can then be loosely interpreted as the probability that the *true* DM exceeds  $DM_{min}$ , under the simplifying assumptions that the candidate is a pulsar and that the folding software is perfectly reliable (which is not the case, see next Chapter).

We safely set  $DM_{min} = 2.0$ , since the lowest known dispersion measure for a pulsar

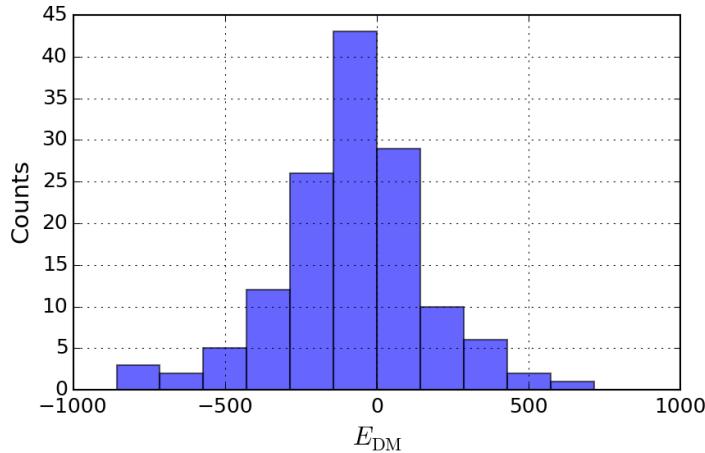


Figure 3.8 Distribution of the  $E_{\text{DM}}$  quantity defined in equation 3.7, measured on a sample of 139 pulsar observations including 19 MSPs. Spin periods in the sample range from 1.74 ms to 5 seconds.  $E_{\text{DM}}$  approximately follows a normal distribution with a standard deviation of  $\sigma_E = 235$ .

is 2.35, but higher values could reasonably be used for searches of low Galactic latitudes. This feature is extremely efficient in selecting against short period, low-DM RFI, while providing a safety net for long period pulsars whose DM has been underestimated. We named CIDER for Confidence InDex in Extra-terrestrial oRigin, because why not<sup>5</sup>.

**4. Coincidence Score.** The beams of a multi-beam receiver are arranged in such a way that their projections on the sky do not overlap, such that several areas may be surveyed in parallel. A source in the far field of the telescope is therefore visible only in a single beam, with the exception of particularly bright astronomical objects sometimes detectable in two or maybe three *adjacent* beams<sup>6</sup> at once. Any signal simultaneously detected in many non-adjacent beams is most likely the manifestation of an interference source that is either in the near field of the telescope (and out of focus) or internal to the instrumentation. In the case of pulsar searches, an efficient filtering technique consists of removing candidate groups with similar periods appearing in multiple, simultaneously observed beams (e.g. Manchester et al., 2001, applied this method to the analysis of the PMPS). This method is however not optimal as it does not take into account the adjacency of the beams in which the candidate appears, or the fact that a genuine pulsar with otherwise excellent

<sup>5</sup>Although CIAO for Confidence Index in Astrophysical Origin would work fine as well

<sup>6</sup>Such was the case of the original FRB for example (Lorimer et al., 2007)

features may be accidentally grouped with interference and not even reach the inspection stage. As a consequence, instead of using a potentially dangerous binary filter, we designed a *coincidence score* that would be one of the classification decision factors among others.

The idea here is to group together all candidates observed simultaneously, then consider all possible pairs of them. First we need a measure of distance in terms of candidate periods: if  $C_i$  and  $C_j$  are two candidates, they are separated by a number of FFT bins  $x_{ij} = T|\nu_i - \nu_j|$  where  $T$  is the observation time in seconds,  $\nu_i$  and  $\nu_j$  their respective signal frequencies in Hz. From there we define a measure of “undesirable association” between the pair as

$$U_{ij} = \exp[-x_{ij}^2] \times \left(\frac{\alpha_{ij}}{\alpha_0}\right)^2, \quad (3.9)$$

where  $\alpha_{ij}$  is the angular separation on the sky of the two beams in which the candidates were detected ( $\alpha_{ij} = 0$  if they were found in the same beam) and  $\alpha_0$  the angular separation between adjacent beams. The first term will be significantly non-zero only if  $x_{ij}$  is not much greater than 1, pointing at an association between  $C_i$  and  $C_j$ . The second term is a large penalty factor for appearing in beams that are distant on the sky. Note that the  $\alpha_0$  normalization factor is not necessary, but makes the score human interpretable, as a pair of candidates with identical periods found in adjacent beams get a score of 1, an occasional occurrence for a bright pulsar.

The final coincidence score of a candidate  $C_i$  is simply

$$\text{CScore}(C_i) = \sum_{j=1}^n U_{ij}, \quad (3.10)$$

where  $n$  is the number of candidates grouped together in this observation. We take the extra precaution that a pair of candidates is considered for undesirable association only if they are both statistically significant, which we defined as  $S/N > 9$  here, to avoid chance associations with noise.

5. **Correlation through time.** Here we simply assess the persistence and consistency of the signal through time using the sub-integrations array. We compute the linear correlation coefficients  $\rho_{P,I_k}$  between the folded profile  $P$  and every single sub-integration  $I_k$ , and use their mean as a feature:

$$C_T = \frac{1}{n_{\text{sub}}} \sum_{k=0}^{n_{\text{sub}}} \rho_{P,I_k} \quad (3.11)$$

An important note is that the value of  $C_T$  depends on the number of phase bins of the profile, which must therefore be the same across all candidates.  $C_T$  is both a measure of the consistency in shape of the signal as well as its level of persistence through time.

6. **Correlation through frequency.** Exactly as above, except that we measure the linear correlation between the average profile and every sub-band. This is a measure of signal consistency through the frequency domain. In particular a signal visible in a single band only will obtain a low score.
7. **Z-ratio.** Some types of RFI may be erroneously detected with very significantly non-zero dispersion measures in the Fourier domain. We decided to fold all candidates at both the DM value returned by the search code  $\text{DM} = d$  and also at  $\text{DM} = 0$ . We recorded the profile S/N value in both cases and used the ratio of the two as a feature:

$$Z = \frac{\text{S/N}_{\text{DM}=d}}{\text{S/N}_{\text{DM}=0}} \quad (3.12)$$

Figure 3.9 shows the distributions of every feature for both data classes on our training data set, and Figure 3.10 is a collection of scatter plots of the training data in selected two dimensional cross sections of the feature space. The set of features presented here is very effective on candidates from Parkes data and should generalize well. However many other equivalent sets could be devised and lead to similar classification performance, as long as the same relevant information about a candidate is captured: in particular the formulas for CIDER and coincident score might possibly be simplified. The RFI distribution is also strongly dependent on the location of the observatory, observing frequency, instrument used and data processing pipeline. The feature set should always be adapted to target particularly common classes of spurious candidates.

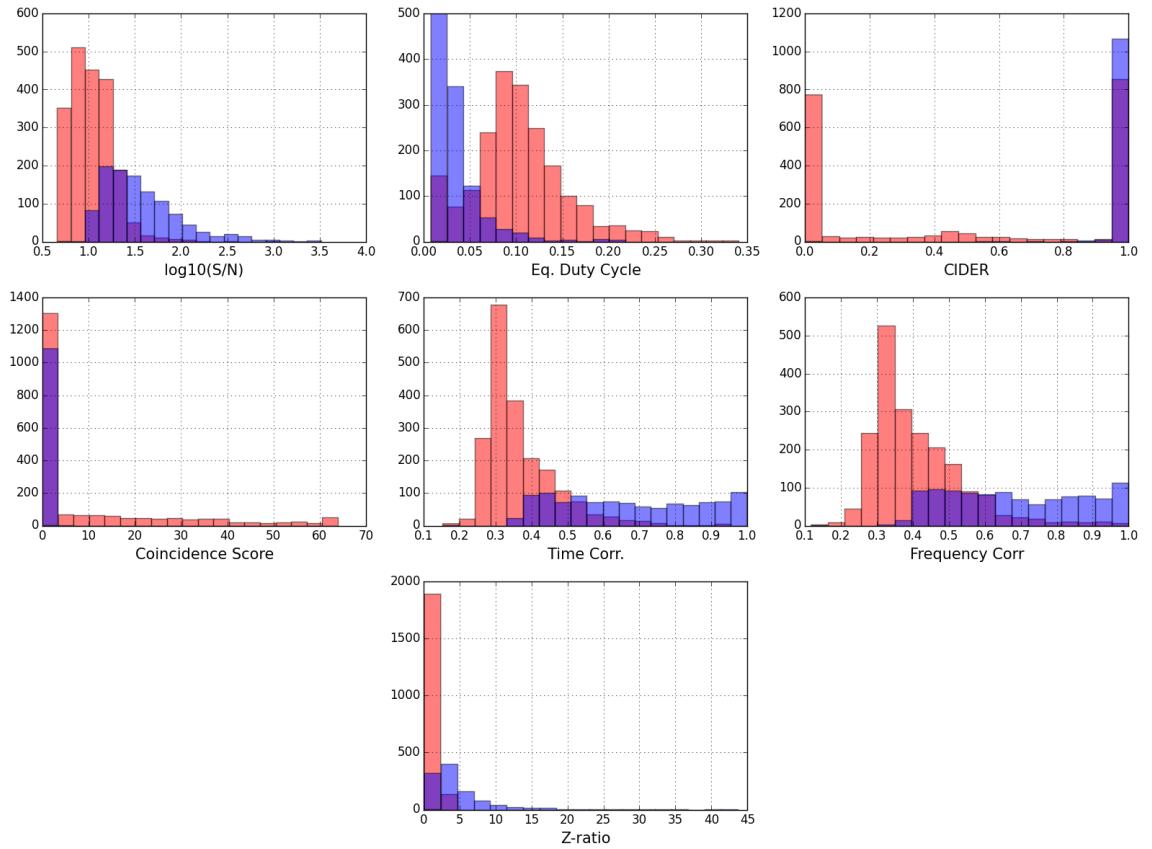


Figure 3.9 Histograms of SPINN features evaluated on our training data set. Pulsars are in blue, non-pulsars in red. Z-ratio may not seem particularly useful in general but efficiently targets a very specific subset of pulsar-like RFI with Z-ratio values below 1.

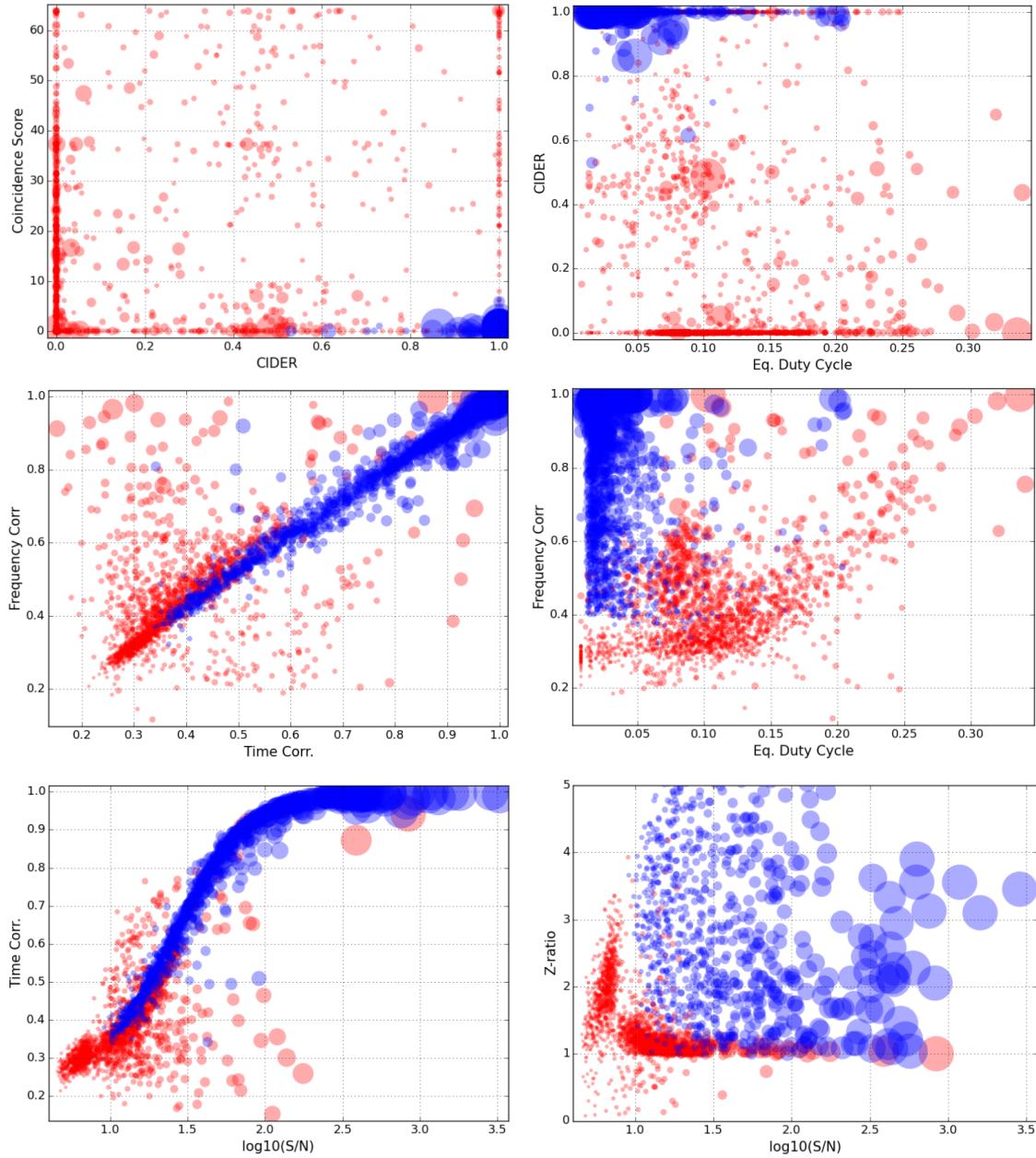


Figure 3.10 Training data projected in a selection of two-dimensional scatter plots where each axis represents a feature (they all appear once at least). Pulsar training examples are in blue, negative examples in red. The surface area of data points is proportional to their signal-to-noise ratio.

## 3.4 SPINN: Performance Optimization

---

### 3.4.1 Training data composition

Gathering as many candidate plots of known pulsars as possible was our obvious priority. We had access to the raw data acquired during the HTRU Intermediate Latitude Survey (Keith et al., 2010), which covered the southern part of the Galactic plane. Fortunately, this survey used the exact same instrument and similar integration times as the more recent surveys we planned to search using SPINN. We ran our whole search pipeline on it and visually identified 1089 candidate plots corresponding to 544 distinct known pulsars. Those that were seen multiple times showed different signal-to-noise ratios and levels of RFI corruption, making every instance useful.

We also had access to several million of negatives (non-pulsar candidates), from which we had to select a sample of similar size to that of our set of positives. But we did not pick an entirely random sample: more than 80% of the negatives had a S/N that was not statistically significant. A classifier trained on a random sample of those would have limited predictive power on actual, high S/N interference. We therefore build a sample of 2000 negatives, half of which were sampled at random, and the other half picked only among negatives with a S/N value higher than 10, guaranteeing that they were not the result of background noise fluctuations.

### 3.4.2 Classification Performance Metrics

Pulsar candidate classification is a problem with very large *class imbalance*; that is there are far fewer instances of one class of data points (pulsars) than the other. In that case simply measuring the accuracy of a classifier is not appropriate: an algorithm predicting that every candidate is a non-pulsar would have near-perfect accuracy but would be quite useless. The right approach when testing a classifier on a given data set is to evaluate its confusion matrix (Fig. 3.11), which keeps track of the right and wrong predictions on each data class separately. From there a number of metrics can be computed, two of which we are particularly interested in optimizing:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$
(3.13)

Recall is the fraction of pulsars properly identified as such, and False Positive Rate

		True Label	
		P	N
Prediction	P	True Positive <b>TP</b>	False Positive <b>FP</b>
	N	False Negative <b>FN</b>	True Negative <b>TN</b>

Figure 3.11 A Confusion Matrix fully describes the types of errors and correct predictions made by a binary classification algorithm on a given data set. From the values of the matrix a number of performance metrics can be derived (see 3.4.2).

(FPR) the fraction of negatives erroneously identified as pulsars. None of these metrics are sensitive to the class imbalance (ratio of positives to negatives) of the particular data set on which they are measured, which make them a good choice for comparing the performance of classifiers across different data sets.

While lowering the false positive rate is a requirement to minimize human intervention, we focused more heavily on optimizing recall for two reasons. Firstly, we planned to use SPINN on survey data that had been searched before, likely limiting new discoveries to a subset of pulsars missed by human inspection and therefore requiring an algorithm as sensitive as possible. Secondly, recall can only be rigorously evaluated on a set of known pulsars, and not on a sample of yet-to-be discovered ones. This is an important distinction, as the population of visible but undiscovered pulsars is on average fainter and possibly different; previous candidate selection methods may have biases that carry over to the known pulsar population. For example, it was reported in previous work that pulsars with wide or multi-peaked pulse profiles were more difficult to identify, both with Machine Learning approaches (Eatough et al., 2010; Bates et al., 2012) and even visual inspection software that may use pulse width as a pre-selection criterion (Keith et al., 2009). For these reasons we set a goal of exceeding 99% recall.

### 3.4.3 Cross-Validation

We have previously discussed that solving any Machine Learning problem requires choosing an appropriate feature set (section 3.3) and tweaking of various free parameters of the

algorithm chosen (section 3.2). We have also seen that obtaining great performance on a given training set is easy to achieve but does not guarantee that the algorithm will generalize well to new data (3.3.2).

Cross-Validation is a set of methods that allow to estimate the results of a statistical model on an *independent* data set without actually requiring the access to one. *k-fold cross-validation* consists of splitting the training data into  $k$  equal-sized parts, holding out one subset for testing while training the algorithm on the remaining  $k - 1$ . The process is repeated  $k$  times by holding out a different part of the data each time. This ensures that the testing and training data are drawn from the same distribution while testing the algorithm on every data instance exactly once. We used 10-fold cross validation to optimize the following:

- The feature set. A number of features were designed from which we selected an optimal subset.
- Neural network architecture. We ended up selecting a 8:1 network, that is two layers of neurons with 8 units in the first and one in the second.
- The max-norm regularization parameter (see 3.2.4) for each layer .

In general, cross-validation can be used for model selection but not for rigorous model evaluation, which requires estimating its performance on new data. The reason is that the choice of the model parameters is indirectly influenced by *all* the data at hand, despite the fact that the algorithm is never tested on data it has been trained upon (Hastie et al., 2009). This may lead to overfitting and therefore an optimistic evaluation of the model. In our case we also encountered several extra issues:

- Some of our positive candidates originated from observations of the same pulsar. This meant that there was some potential leakage of information between testing and training sets during every iteration of cross-validation.
- Our sample of negative candidates is not representative of the underlying population, containing an artificially increased fraction (50 %) of bright and difficult examples. This would lead to an overestimated false positive rate.
- It turns out that the false positive rate of SPINN ended up being too low to be properly evaluated on a sample of several thousand candidates, requiring two orders of magnitude more.

We therefore evaluated SPINN on a very large set of candidates from a more recent survey, covering an entirely different portion of the sky and ensuring a fully independent test. The results are presented in the next chapter.

# 4

## Results and Discoveries

### 4.1 Interlude: The importance of folding software

---

It may seem obvious that the accuracy of a classifier is very much dependent on the implementation quality of all the earlier stages of the search pipeline. Pursuing a philosophy of “no pulsar left behind” requires thinking critically about the interactions between those processing stages and the classifier itself. In particular we ran into a significant barrier to pushing the accuracy of SPINN to the limit: the quality of the candidate plots themselves.

Historically, the first version of SPINN processed candidates that were produced by the PSRCHIVE software package (Hotan et al., 2004). Delving into details, this involved running four programs in succession on every candidate independently:

- `dspsr` folded the filterbank data at a given period, dispersion measure and acceleration returned by the FFT search code. As an output, it wrote a data cube with 1,024 channels, 32 sub-integrations and 128 phase bins.
- `paz` set to zero a list of channels known to be often or permanently affected by interference. This was the only RFI mitigation step.
- `pam` added the 1,024 channels into 32 sub-bands to accelerate the final optimization step.
- `pdmp` ran the period and DM optimization algorithm described earlier in section 2.2.4, and wrote a XML file containing the usual candidate parameters, profile, sub-band and sub-integration arrays.

The first obvious issue was the constant reading and writing of small files on the lustre filesystem<sup>1</sup> used by Swinburne’s computing cluster gSTAR. At that time SPINN spent

---

<sup>1</sup>A distributed file system particularly fast at reading and writing large (1GB+) files, but also particularly inefficient on tiny ones.

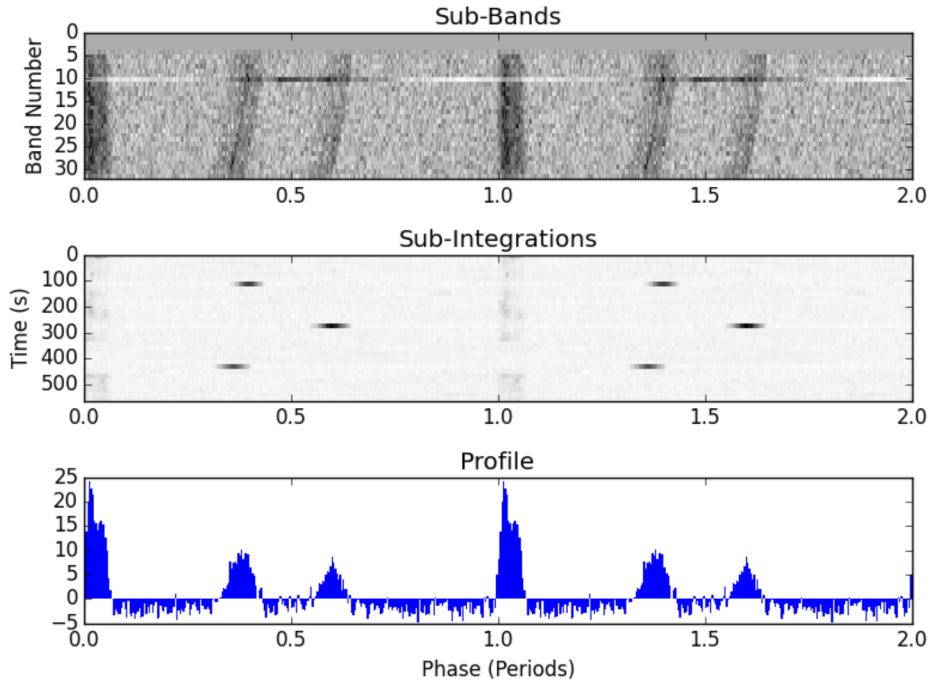


Figure 4.1 Example candidate plot of a pulsar disturbed by two types of interference. The pulsar’s signal is the vertical trail at a phase of 1 period (two periods are shown for readability), and is demonstrably nulling between  $\simeq 330$  and  $450$  seconds into the observation. Top panel: a strong, periodic sine-wave like RFI is present in frequency band no. 10. Middle panel: additionally, three bright dashes correspond to brief bursts of broadband interference; they are also visible in the sub-bands plot as curved trails, and generate secondary pulses in the overall folded profile (bottom panel).

two orders of magnitude more time reading  $\simeq 50$  KB candidate XML files than evaluating them. A look at the source code of `pdmp` also revealed a number of significant speed improvement opportunities as well. But more importantly, a fraction of our training candidates containing a pulsar were visibly affected by interference. It would manifest itself either as a set of scrambled frequency channels that were not included in the usual suspects list used by `paz`, or by short 1–100 ms broadband, non-dispersed bursts showing up as bright dashes in the sub-integrations plot on top of the legitimate pulsar signal. A clear example is shown in Figure 4.1. Such RFI may have two major consequences:

- **Critical failure of period-DM optimization.** The algorithm consists, as explained in section 2.2.4, of applying circular shifts to individual profiles in the data cube then summing them into a global folded profile. The one that has maximal signal-to-noise ratio identifies the best period and DM. The algorithm is foolproof only if there is just a pulsar signal on a white noise background; the end result can

very well be the product of an unlucky alignment of two short bursts of RFI, or the optimization of a brighter narrowband signal at a neighboring period. In these cases the output period and dispersion measure are erroneous, while the pulsar may still be visible as a signal drifting across the sub-integrations and sub-bands plots, much like in the left panel of Figure 2.5.

- **Incorrect SPINN feature values.** An underlying assumption of the feature space design is that the folding process is reliable. Cases described above usually result in a smeared pulse profile with increased width, lower DMs and lower correlation coefficients between profile and sub-bands/sub-integrations. In other words, the derived features exhibit the trademarks of a non-pulsar.

An expert eye may still be able to spot a poorly folded pulsar; but to achieve the same, a machine learning algorithm requires a diverse sample of folding pipeline mistakes and a more complex feature space capable of dealing with a signal drifting across the sub-integrations and sub-bands plots. **A folding pipeline of poor quality vastly increases the dimensionality of the ML problem, which should be avoided** for reasons explained in the previous chapter. The most economical way to tackle the problem was to rewrite the folding software, add an RFI mitigation step before period-DM optimization, and solve the speed issues along the way.

## 4.2 The CUBR folding pipeline

---

### 4.2.1 Overview

CUBR<sup>2</sup> is a package that handles filterbank folding, data cube manipulation and period-DM optimization, and outputs candidates in a format practical for large-scale evaluation by SPINN. It is written mostly in the Python language which enables the use of the powerful NumPy, SciPy and Matplotlib libraries. A number of performance critical functions are written in C, using the OpenMP library for multi-threading. CUBR is also specifically designed to fold a large number of candidates from the same observation at once, and performs the task of `dspsr` three times faster in our typical data processing conditions (Table 4.1). This is a consequence of assigning one thread per candidate and having to make only a single pass over the observation data.

---

<sup>2</sup>because it makes data cubes and the name sounds all right

	Total (s)	Time / cand (s)
<b>dspsr</b>	156	6.5
CUBR	51	2.1

Table 4.1 Folding speed comparison between **dspsr** and CUBR. Using 6 threads, we measured the time used to fold 24 candidates from a 540 second Parkes observation containing  $2^{23}$  time samples with 1,024 frequency channels. The benchmark was performed on a node of Swinburne’s gSTAR cluster, equipped with Intel Xeon E5 2.2GHz CPUs.

#### 4.2.2 RFI mitigation based on outlier detection

After folding one obtains a data cube: a three-dimensional array  $I_{\nu t \phi}$  representing an observed intensity as a function of frequency channel number  $\nu$ , sub-integration index  $t$  and phase bin index  $\phi$ . It can also be thought as a set of folded profiles indexed by  $\nu$  and  $t$ . At this stage, no dispersion delay shifts have been applied yet and the cube represents the observation folded at a certain period  $P$  and a dispersion measure of zero.

We identify profiles in the cube that are affected by RFI with a simple outlier detection method. The first step is to compute a number of numerical features describing every profile, we chose:

- Variance.
- Peak-to-peak difference.
- Absolute value of the second bin of the profile FFT. This is a proxy for measuring how similar to a sine wave the profile is.

The second step is to identify which profiles stand out from the majority in either of those three dimensions. We use Tukey’s rule for outliers (e.g. Chandola et al., 2009): given a set of values, one is considered anomalous if it falls out of the interval  $[Q_1 - 1.5R, Q_3 + 1.5R]$  where  $Q_1$  is the 25-th percentile of the set,  $Q_3$  the 75-th percentile and  $R = Q_3 - Q_1$  is called the interquartile range. None of these statistics are distorted by the presence of outliers, which make them a good choice as opposed to the mean or the standard deviation. See Figure 4.2 for a visualization of the process on the same pulsar observation as in Figure 4.1. Finally, any anomalous profile has all its values forcibly set to zero in the data cube.

On our data, this method has proven to be very efficient at removing strong narrow-band interference, but brief impulses of broadband RFI similar to those visible in Figure 4.1 can remain. We add one more processing step to get rid of them. For that, we consider the sub-integrations array at  $DM = 0$ :  $S_{t\phi} = \sum_{\nu=1}^{n_{\text{chan}}} I_{\nu t \phi}$ , which is the sum of the data

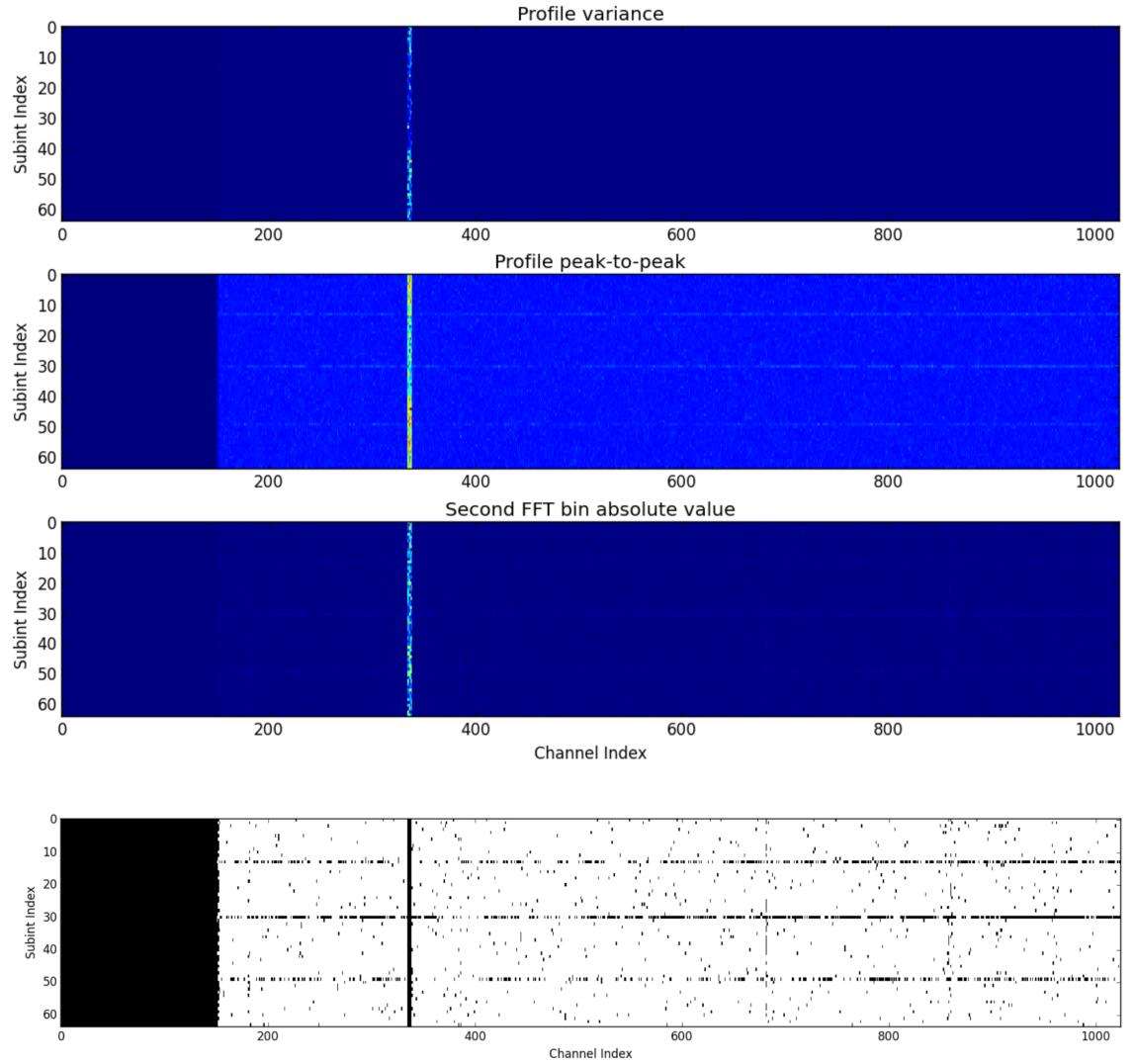


Figure 4.2 RFI mitigation process based on outlier detection. We use the same pulsar observation as in Figure 4.1, folded as a cube of 1,024 frequency channels  $\times$  64 sub-integrations  $\times$  256 phase bins. Top panel: each pixel in the image represents the variance of a profile in the folded data cube. Middle two panels: same exercise for the two other features (see text). Bottom panel: any profile that is anomalous in either of those three dimensions according to Tukey's rule is flagged (shown as a black pixel) and all its values are set to zero in further processing steps. Both the interference-ridden narrow band of frequency and the three disturbed sub-integrations are properly identified.

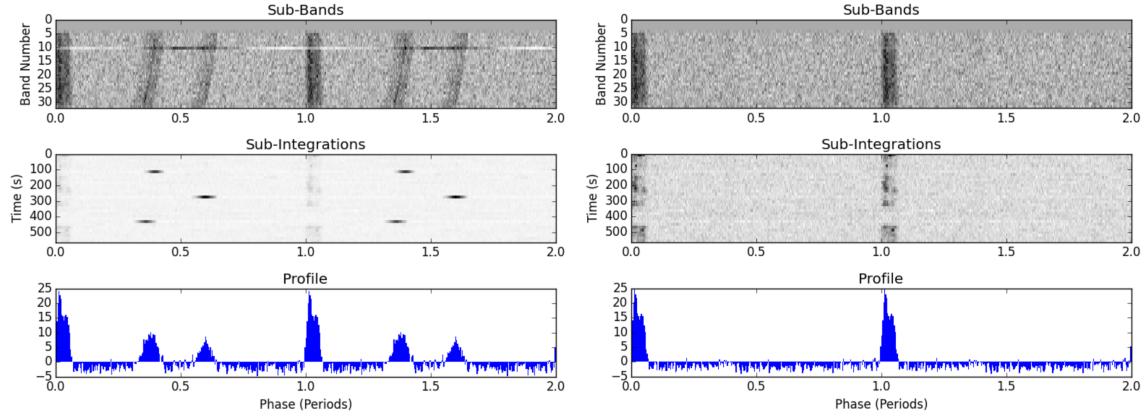


Figure 4.3 The pulsar observation from Figure 4.1 before (left column) and after (right column) the application of RFI mitigation algorithms described in section 4.2.2. Interference has been effectively erased.

cube along the frequency dimension with no dispersion shifts applied. Every column of  $S_{t\phi}$  corresponds to a given phase and we apply Tukey’s rule to all of them independently, thereby identifying anomalous  $(t, \phi)$  pairs. For any such pair we substitute associated data cube values with the following replacement:

$$I_{\nu t\phi} = \frac{1}{n_{\text{chan}} \sum_{t \in [0, n_{\text{int}}]} \text{median}\{S_{t\phi}\}} \quad (4.1)$$

where  $n_{\text{int}}$  is the number of sub-integrations and  $n_{\text{chan}}$  the number of frequency channels. The overall effect of the two cleaning algorithms described here is shown on our pulsar observation example in Figure 4.3. The effective deletion of many instances of interference at this stage of the processing is instrumental in the consistent success of the period-DM optimization algorithm and makes the ML classification problem easier. We finally note that running both cleanup algorithms on a candidate takes less than a second for our typical processing parameters.

### 4.2.3 Adaptive optimization grid

We also re-implemented the period-DM optimization stage (explained in 2.2.4), as we wanted to use an exhaustive implementation of the S/N algorithm (see 3.3.4) that tries every possible pulse width, and also achieve faster speeds than `pdmp`. Beyond various code optimizations, the most significant change was the introduction of an adaptive period-DM trial grid. Since in most of the two dimensional search space a pulsar signal has little to no power, it is not economical to search its entirety at a high resolution. Instead we start by

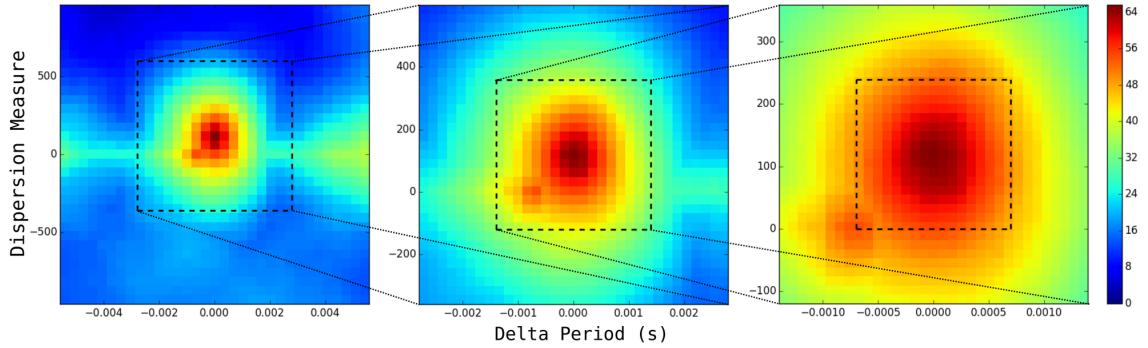


Figure 4.4 Adaptive Period-DM optimization grid demonstrated on a real pulsar observation. Left panel: the signal-to-noise ratio of the candidate is computed on a coarse  $32 \times 32$  Period-DM trial grid and an approximate set of candidate parameters is found. Middle and Right panels: that initial guess is then iteratively refined by performing a new  $32 \times 32$  grid of trials with the search ranges halved, and centered around the best solution previously found. In three iterations here, we achieve the same period and DM accuracy as if we used a single  $128 \times 128$  grid, but at less than 20% the computational cost.

exploring the full Period and DM ranges at a coarse resolution, find an approximate set of parameters for which S/N is maximized, then iteratively refine that solution by performing higher resolution searches around it (Figure 4.4). Overall, we achieved a speedup factor in excess of 10 over pdmp which was the main bottleneck of the old pipeline (Table 4.2).

#### 4.2.4 Results

The goal of improving the speed at which candidate plots are produced has been reached, with the new folding pipeline being approximately 5 times faster than the one previously used. The output format has been changed so that all the outputs from an observation are stored in a single file that contains of the order of 300 candidates and directly stores their data in practical and ready to use Python objects and NumPy arrays.

We historically had to build the largest possible training data set from the outputs of both pipelines, a process that required us to visually inspect the candidates from the observations of the HTRU intermediate latitude survey known to contain a known pulsar. This gave us a measure of output quality: the number of candidate plots associated to a known pulsar that would be selected for confirmation in a blind search. That number of recognizable pulsar observations went from 1,028 to 1,089 after we switched to the CUBR pipeline, and none presented signs of imperfect period-DM optimization. This made the effort invested in CUBR worthwhile, as it is not sensible to pursue high ( $\geq 98\%$ ) levels of recall for a classifier if nearly 6% of pulsar instances are being lost in earlier processing.

Program	Trial grid size	Iterations	Time / cand (s)
pdmp	32×32	1	8.6
pdmp	128×128	1	17.6
CUBR	32×32	3	1.3

Table 4.2 Candidate optimization speed comparison between `pdmp` and CUBR. We measured the time used to find the best period and DM of a candidate with 1,024 frequency channels  $\times$  32 sub-integrations  $\times$  128 phase bins. 3 iterations of CUBR with a  $32 \times 32$  grid achieve the same resolution as `pdmp` with a  $128 \times 128$  one (Fig. 4.4). The benchmark was performed on a single Intel Xeon E5 2.2GHz CPU.

## 4.3 Performance evaluation on the SUPERB survey

We present in this section an evaluation of the classification accuracy of SPINN, when processing CUBR candidates specifically. In light of the above, it should be clear that folding and classification software can be evaluated only as a whole.

### 4.3.1 Test setup

After building our training set of candidates from the HTRU Intermediate Latitude survey (Keith et al., 2010), we tested SPINN using the first year of observations of the recent SUPERB survey (Keane et al., in prep.). Both data sets were acquired using the same observing setup at Parkes, including identical 9 minute integration times on the sky, and were processed with the same infrastructure. The data were searched using the fully GPU-based PEASOUP<sup>3</sup> code (Barr, in prep.) and the 24 most significant candidates in every beam folded using CUBR. There are only two noteworthy differences between SUPERB and HTRU-Mid:

1. **Sky region observed.** HTRU-Mid covered a strip of the Galactic plane with  $|b| < 15^\circ$ , while SUPERB spent the majority of its time at  $|b| > 25^\circ$ . See Figure 4.8.
2. **Data acquisition dates.** HTRU-Mid was observed during 2008–2010, and the SUPERB test data between April 2014 and February 2015.

There was absolutely no overlap between survey regions. This would guarantee an objective test of SPINN, with the added difficulty of a changed, more recent RFI environment than that on which it was trained. For every known pulsar in the SUPERB region, we checked if it was detected by PEASOUP in observed beams less than 14 arcmin<sup>4</sup> away.

<sup>3</sup><https://github.com/ewanbarr/peasoup>

<sup>4</sup>The FWHM of the beam response on the Parkes multibeam receiver

We identified 139 detections of 85 distinct known pulsars, which made our positive test set. Our sample of negatives was made of the entire output of CUBR on all 60,450 beams that contained neither a known pulsar nor any of our discoveries, for a total of 1,418,598 negative candidates. This number is large enough that missed discoveries, if any, would not affect false positive rate measurements.

### 4.3.2 Classification accuracy

Since SPINN is a neural network, it returns a real-valued score between 0 and 1 for every candidate, which is particularly useful in producing a natural ranking in which to examine them. However that score must be converted into a binary class by applying a score decision threshold above which a candidate is classified as positive. We computed a confusion matrix (Fig. 3.11) as a function of score threshold on our test sample, and from there the following metrics:

$$\begin{aligned} \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{FNR} &= \frac{\text{FN}}{\text{TP} + \text{FN}} = 1 - \text{Recall} \\ \text{FPR} &= \frac{\text{FP}}{\text{TN} + \text{FP}} \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \end{aligned} \tag{4.2}$$

*Recall* is the fraction of true pulsars found, or equivalently the *false negative rate* (FNR) is the fraction of pulsars missed. The *false positive rate* (FPR) is the fraction of negatives incorrectly reported as pulsars. *Precision* is the fraction of genuine pulsars among the candidates tagged as positives by SPINN. All of the above metrics are presented as a function of score threshold in Fig. 4.5. Note that we measured the false positive rate on the entire sample of 1.42 million negatives (red line), and also measured it using only the more difficult subset of 0.43 million negatives with a S/N exceeding 10 (black line). This guarantees that they are the product of actual radio-frequency interference and not of random noise fluctuations, which are easy to reject based on their low statistical significance alone. More importantly, we plotted the false positive rate as a function of recall to illustrate clearly the trade-off between the two in Fig. 4.7.

It is apparent that performance is most heavily affected by a few percent of difficult pulsars, and discovering them requires going through two orders of magnitude more false positives. One particular pulsar instance that received the lowest score of 0.19 requires

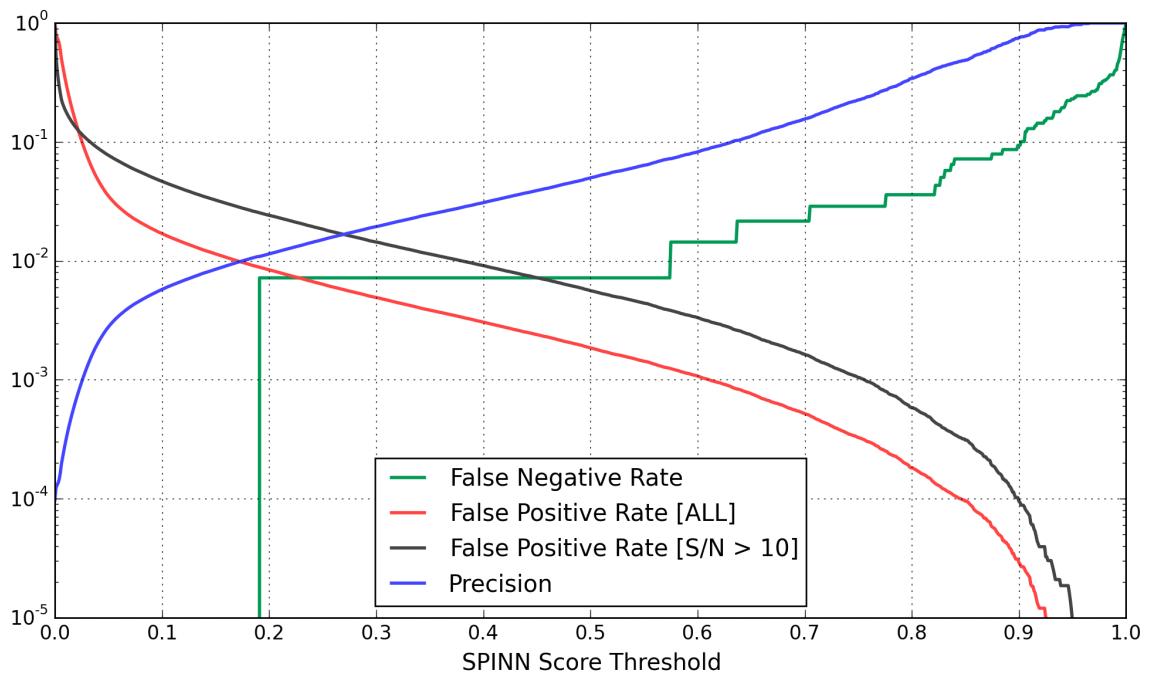


Figure 4.5 Classification performance metrics of SPINN evaluated on the SUPERB data set, as a function of score decision threshold. Higher thresholds lead to a lower false positive rate but an increased fraction of genuine pulsars being missed (false negative rate). We also measured the false positive rate on a subset of negative candidates that exclude noise fluctuations (see text), conservatively selected as those with a S/N exceeding 10.

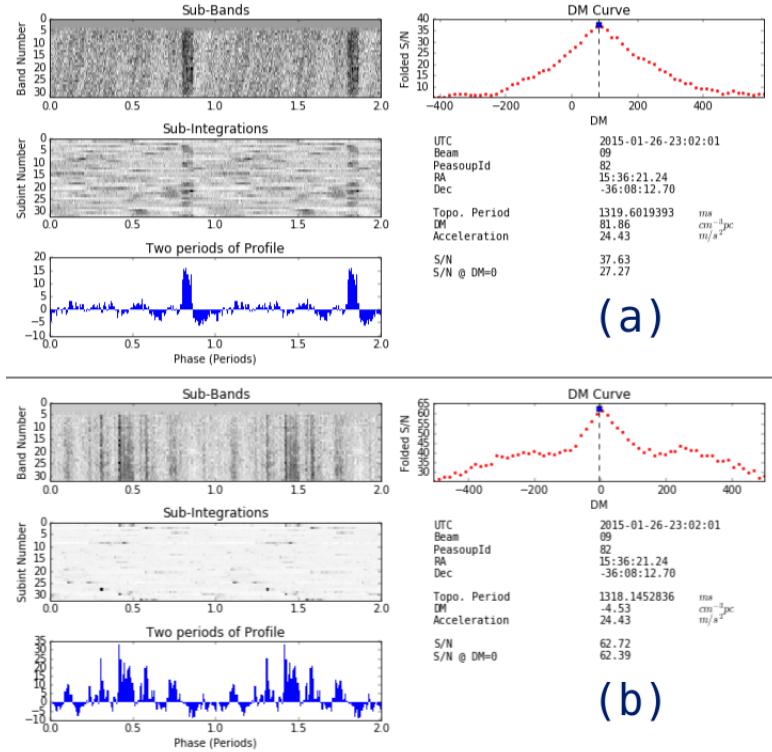


Figure 4.6 Top panel (a): lowest scoring candidate plot of a pulsar in the test sample as produced by the CUBR folding software. Heavy signs of interference remain. Bottom panel (b): the same candidate plot when produced without CUBR’s RFI mitigation algorithms, ending up with the wrong period and DM and unrecognizable overall.

a closer inspection. Its candidate plot, shown in Fig. 4.6, is heavily affected by RFI, leading its correlation through time and frequency features to be unusually low. Interestingly enough<sup>5</sup>, this pulsar observation is completely unrecognizable by eye without the application of CUBR’s interference cleanup algorithms, and would not have been part of the candidate test sample had they not been used. Improving RFI mitigation further is probably the best solution to increase the detectability of such edge cases.

With that being said, 75% of pulsar observations are ranked higher than any of the 1.42 million negative candidates. Achieving 96% recall requires going through about 150 false positives per million candidates, a negligible amount that consumes no more than a few minutes of extra inspection time. There is no need for false positive rates much lower than  $10^{-3}$  on current surveys, but SPINN can definitely achieve that and would therefore scale to candidate numbers of up to a billion if a recall rate of  $\simeq 90\%$  can be considered acceptable.

<sup>5</sup>and perhaps ironically

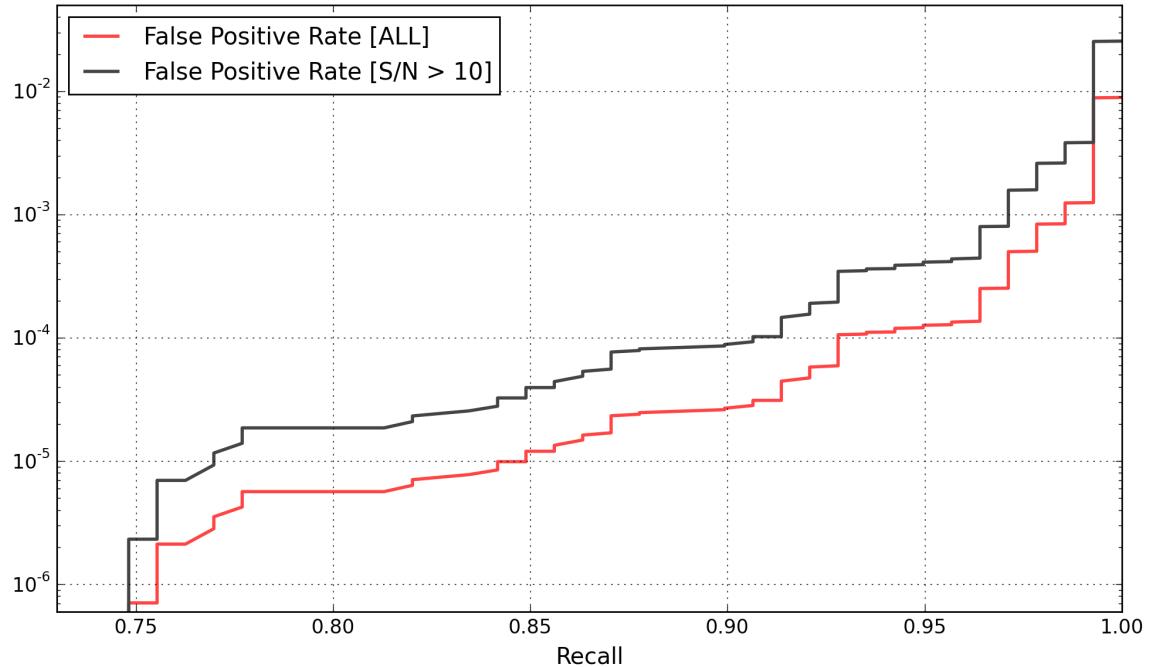


Figure 4.7 False positive rate as a function of recall, illustrating the trade-off between the two metrics. 75% of all the known pulsar candidates are ranked higher than any non-pulsar over a 1.42 million candidate sample. SPINN reduces the amount of negative candidates to inspect by a factor between 300 and 10,000 depending on the desired completeness of the search. Higher reduction factors are unnecessary on current surveys with  $\leq 10$  million candidates, but they can be achieved and would certainly be useful on next generation telescopes such as the SKA.

### 4.3.3 Classification speed

The scoring speed of SPINN on a single Intel Xeon E5 2.2GHz CPU is approximately 180 candidates per second, or 650,000 per hour. That includes reading candidate files, computing feature vectors, activating the neural network and writing the results to a file. Scoring candidates is an embarrassingly parallel problem that can be divided in identically sized batches and assigned to multiple CPUs, each running an instance of the classifier. At least on our sets of a few million candidates we have observed the scoring speed to scale linearly with the number of processors used. On Swinburne’s gSTAR cluster, using 64 CPUs in parallel, the entire SUPERB test sample described above has been scored in 90 seconds.

---

## 4.4 Summary of searches conducted

---

We combined PEASOUP and SPINN into a fast and accurate search system to process all of the survey data that was available to us. A summary of surveys processed can be seen in Table 4.3, and their sky coverage in Figure 4.8. The entirety of HTRU Mid and around 50% of HTRU high latitude had been searched before, although not for binary pulsars as the computing power for acceleration searches was not available at the time. Since graphical tools had been used for candidate selection, these data sets afford an opportunity to test our ability to detect pulsars that were previously left behind, if any. We visually inspected all candidates down to a score of 0.5 to ensure a high level of completeness, which involved approximately 20,000 plots to inspect for all the searches combined<sup>6</sup>. The false positive rates on HTRU were considerably lower than on SUPERB, due to obviously apparent worsening of the RFI environment at Parkes over time.

The SUPERB survey has also been a testing ground for the very first real-time search pipeline with automated candidate classification. This was made possible by the ability of PEASOUP to search the data in significantly less than  $T_{\text{obs}}$ , the speed improvements brought by CUBR and the extremely low false positive rates achievable by SPINN. This so called “fast” pipeline searches more limited acceleration ranges due to time constraints, and candidates scoring higher than a conservative threshold of 0.9 are immediately uploaded to an interactive web based viewing tool, ensuring that a large majority ( $\simeq 80\%$ , see *precision* in Figure 4.5) of reported candidates are genuine pulsars. The delay between the end of an observation and a potential pulsar alert is approximately 20 minutes and can

---

<sup>6</sup>Although the thorough development and testing process of SPINN and CUBR probably involved looking at a few extra tens of thousands as well

Survey	HTRU Mid	HTRU High	SUPERB
$T_{\text{obs}}$ (s)	540	270	540
Observation years	2008–2010	2010–2013	2014–2016
Beams searched	90,000	357,500	227,600
$ a_{\text{max}} $ (m/s <sup>2</sup> )	50	50	250
$\Delta a$ (m/s <sup>2</sup> )	0.92	3.66	0.92
Candidates	4,300,000	13,000,000	5,500,000

Table 4.3 Parameters of surveys searched.  $|a_{\text{max}}|$  and  $\Delta a$  are the maximum trial acceleration value and the acceleration trial step respectively. Part of SUPERB was also searched with a real-time pipeline with accurate candidate selection integrated (see text), for which  $|a_{\text{max}}|$  was reduced to 25 due to processing time constraints.

be optimized further. Most importantly 4 new pulsar discoveries (Table 4.4) have been made with the fast pipeline, which makes it much more than a simple proof of concept. This is a significant result as pulsar searches with the SKA may be performed in real time on this model because the raw data are too large to be stored permanently.

## 4.5 Discoveries

Overall, a total of 21 new discoveries have been made using SPINN so far, which we summarize in Table 4.4. To these can be added 5 unpublished pulsars detected blindly and subsequently found to have been reported earlier by survey teams that were operating other telescopes. The efficiency of our folding and classification system is most clearly demonstrated by the results obtained on the HTRU Mid and High Latitude surveys. 122 objects had previously been discovered in these data, to which we have now contributed an extra 16 new pulsars. Of these, 13 are isolated and searches previously conducted should have been able to detect them. Yet to be re-observed for confirmation are an extra 13 high quality HTRU candidates, that appear to be mostly isolated, slow spinning normal pulsars and two potential MSPs.

As the focus of the thesis has been the design of an accurate candidate classifier, we do not provide full descriptions and timing solutions of the discoveries, which will be done in future publications. However a number of the more significant finds deserve a brief mention. J1804–29 is the fastest spinning pulsar in the Galactic plane, exceeding the rotational frequency of the “original” MSP (P=1.55ms, Backer et al., 1982). It is seemingly isolated but shows a significantly and steadily *decreasing* spin period over two years of timing, so the presence of a companion in a long period orbit (several 10<sup>3</sup> days) remains a possibility. Like J1755+01, also found by SPINN in HTRU Mid survey data,

Name	Survey	Period (ms)	DM ( $\text{cm}^{-3}\text{pc}$ )	S/N	Notes
J1804–29	HTRU Mid	1.4927	232.4	10.3	
J1705–19	HTRU Mid	2.4802	57.5	22.7	b
J1755+01	HTRU Mid	4.4108	70.3	9.8	b
J1558–68	HTRU Mid	267.27	105.4	10.8	
J1842–28	HTRU Mid	815.27	61.6	14.8	
J1703–19	HTRU Mid	1270.2	45.7	15.4	
J1655–27	HTRU Mid	1623.7	129.2	17.6	
J1307–41	SUPERB	2.2046	35.0	23.3	<b>R</b> , b
J1421–44	SUPERB	6.3859	54.6	12.1	b
J1914+09 <sup>1</sup>	SUPERB	440.07	293.1	13.0	
J1126–39 <sup>2</sup>	SUPERB	877.62	51.1	18.3	<b>R</b>
J1604–32	SUPERB	883.89	57.9	19.6	
J0750–69	SUPERB	915.19	48.1	18.1	<b>R</b>
J1337–45	SUPERB	1257.5	101.8	18.1	<b>R</b>
J2155–28 <sup>2</sup>	SUPERB	1343.4	40.8	21.9	
J1406–43	SUPERB	2436.4	61.1	18.3	
J2228–65	SUPERB	2746.0	64.9	24.7	
J1948–19	HTRU High	2.6032	25.2	16.0	b
J0125–23 <sup>2</sup>	HTRU High	3.6759	9.6	32.9	
J0637–31	HTRU High	3.9458	15.5	10.7	
J1517–32 <sup>2</sup>	HTRU High	64.402	25.5	35.7	
J1843–41 <sup>3</sup>	HTRU High	324.19	65.9	15.7	
J1404–03 <sup>3</sup>	HTRU High	362.63	30.7	20.8	
J1941+03 <sup>3</sup>	HTRU High	1232.2	90.3	25.8	
J1942+02 <sup>3</sup>	HTRU High	1405.0	150.5	15.1	
J0754–08 <sup>3</sup>	HTRU High	2093.6	38.0	15.7	

[R] Discovered in real time

[b] Binary pulsar

<sup>1</sup> Previously found in the PALFA survey

<http://www.naic.edu/~palfa/newpulsars/>

<sup>2</sup> Previously found in the GBNCC survey

<http://arcc.phys.utb.edu/gbncc/>

<sup>3</sup> Co-discovered with students Sally Cooper and Robert Lyon

Table 4.4 The 26 confirmed pulsar discoveries made by SPINN so far. We provide only discovery parameters including the folded S/N ratio of the detection; the names are provisional until full timing solutions and accurate positions are obtained. Five of these pulsars, although unpublished, have been reported earlier by other survey teams; they are nonetheless mentioned in this list as they were found blindly. A further 14 high quality candidates are yet to be re-observed (13 in HTRU and 1 in SUPERB).

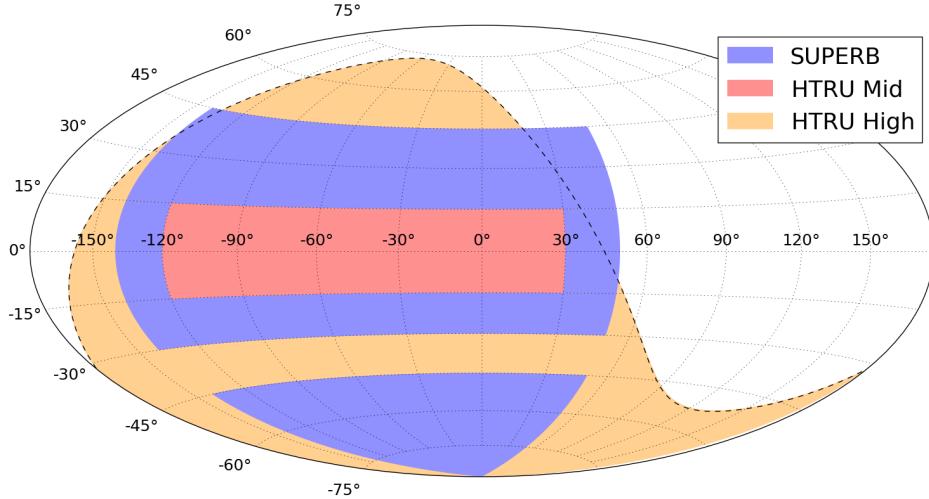


Figure 4.8 Galactic coordinates covered by the surveys that have been searched using SPINN. SUPERB and HTRU intermediate latitude do not overlap by design, while HTRU high latitude covers the whole sky below a declination of  $10^\circ$ , materialized by a dashed line.

it was detected with a very low S/N  $\simeq 10$  and the fact that SPINN could identify it is a comforting result on its own. J1705–19 is a binary MSP with a low mass companion; it has an extremely narrow pulse width of less than  $40\mu\text{s}$  and is overall a very stable rotator with a timing residual RMS of a few hundred nanoseconds. It may be a very valuable addition to pulsar timing arrays.

Finally, the most exciting case is that of SUPERB discovery J1307–41. It was discovered in real-time as an apparently bright accelerating MSP, and seen a second time half an hour later in a nearby survey pointing with a significantly different detection acceleration. This confirmed its existence and binary nature, while the measured change in acceleration hinted at an orbital period of  $\simeq 1$  day should its orbit be circular. In a most puzzling fashion, it has since disappeared from the radio sky, and has not been detected again after observing the discovery coordinates for more than 10 hours over several months, mostly in 30 to 60-minute integrations. J1307–41 could be a new specimen of transitional millisecond pulsar, and future re-observations in X-rays may help to confirm this hypothesis if it is seen to have entered a new accretion phase.

# 5

## Conclusion

### 5.1 Summary

---

In this thesis we have presented SPINN, a Machine Learning based pulsar candidate classifier that is both fast and highly accurate. We tested it on data completely independent from those on which it was trained, and showed its ability to reduce the number of spurious candidates to examine by a factor of the order of 1,000 with almost no pulsar loss compared to a full one-by-one visual inspection by a human expert<sup>1</sup>. The output of an entire modern pulsar survey can be now sifted through by a single person in a matter of hours.

A practical demonstration of the efficiency of SPINN has been made as we used it to discover 26 new pulsars with minimal inspection effort; another 13 promising candidates still are awaiting confirmation. Most importantly, the majority of these discoveries were made in data that were previously searched using interactive candidate selection tools described in section 2.3.3; that these were arguably missed demonstrates the superiority of a well-designed ML approach not only in terms of reduction of effort but also completeness of the search. Finally we also used SPINN as the final stage of a real-time search system with an essentially zero false positive rate that discovered four new pulsars.

A number of considerations for present and future pulsar searches should be taken away from the present work:

- **Feature design is critical.** As a consequence of the Bias-Variance tradeoff in Machine Learning, the scarcity of pulsars in the training data requires the feature space to remain low-dimensional and to capture as much relevant information as possible. Detailed knowledge about the whole pulsar searching process has been

---

<sup>1</sup>An approach rarely taken anyway considering the sheer amount of tedium involved

instrumental in designing high quality features. However a substantial increase in training examples of pulsars could possibly change the situation (see next section).

- **Importance of candidate plot quality.** The effort of developing of a human-level accuracy classifier is justified only if candidate plots associated to a pulsar reliably come out of the previous stages of the pipeline in a recognizable form. We showed that, if care is not taken, that a few percent of pulsars can realistically be lost before they even reach the classifier, automated or human. Candidate plots of consistently high quality also significantly reduce the difficulty of the classification problem, indirectly reducing either the number of features to be used or the amount of training data required. We invested significant effort to improve candidate plot quality; in particular, two RFI mitigation algorithms were added to our folding software. We advocate evaluating all pipeline stages coming after the FFT search as a whole.
- **Classification speed is a marginal issue at best.** As long as wasteful I/O operations are avoided, the time required to evaluate one candidate is three orders of magnitude lower than that consumed to produce it in both the search and folding process. In fact the scoring speed of SPINN on a single CPU is 650,000 candidates per hour, which matches the candidate output rate of the all-sky pulsar search planned for SKA1-Mid (Dewdney, 2013):

$$2,222 \text{ beams} \times 6 \text{ obs./hour} \times 50 \text{ cand./beam} = 666,600 \text{ cand./hour.}$$

This speed is unsurprising since the candidate features used are simple and fast to compute, and activating the neural network is essentially two small matrix multiplications. However, there is certainly a lot of room for the use of more complex candidate pre-processing, new features and more advanced ML algorithms should the need arise.

## **5.2 Future work and SKA pulsar searches**

---

We suggest two directions of investigation to further improve the accuracy of candidate classifiers. The first is the addition of improved RFI mitigation methods to previous stages of processing. We found that pulsars observed during interference outbursts were the most likely to be mislabeled by SPINN, as their candidate plots inevitably share characteristics with those in the negative class of the training data. The approach of putting the entire burden of RFI rejection on the final candidate selection stage is not optimal as a significant fraction of pulsars can be lost in earlier stages of the search process. Techniques that deal

with interference directly in the raw data have already been used with some success, such as the zero-DM filter of Eatough et al. (2009), or the `rfifind` routine of the PRESTO software package (Ransom, 2001). When observing with an array of receivers, *spatial filtering* can be applied, where the incident direction and signature of an interference source can be inferred from short-term matrices of covariances between the recorded signals (e.g. Leshem et al., 2000; Kocz et al., 2012). The interference-free signal can then in theory be reconstructed.

Another very substantial improvement can be obtained with a realistic training example generator. Such a system would have to properly simulate an underlying pulsar population, including observed properties of their signal and ISM propagation effects, and inject the resulting train of pulses in raw data acquired with a certain instrument affected by its usual RFI environment. It would then be searched with the same pipeline as would be used in a blind survey. The availability of a large amount ( $\simeq 10^6$ ) of training pulsar examples would lift the small feature set size requirement that we adopted here; in particular, the approach of performing direct pattern recognition on candidate plots using deep neural networks (already attempted by Zhu et al., 2014, but with a small training data set) is very likely to become superior. And beyond the welcome improvement in accuracy, there are more reasons to adopt training data generation in preparation for SKA pulsar searches:

1. **Cost effectiveness.** Observing time on the SKA especially in its early days will be in high demand and precious. Ideally, training data should be obtained by re-observing every known pulsar under the same conditions as in a blind search survey; however, this may prove too costly.
2. **Avoid training data biased towards bright pulsars.** The vastly increased sensitivity of SKA compared to previous instruments means that the pulsar population known at the time it sees its first light would be seen with signal-to-noise ratios far above the detection limit. As all supervised learning algorithms inevitably carry all the selection biases present in their training data, the detectability of faint pulsars will suffer accordingly.
3. **Building a reference test data set.** The test data set we used here remains too small to accurately evaluate recall rates, or the frequency with which mislabeled edge cases such as interference-ridden pulsar observations occur. A larger sample of fake but realistic pulsars would allow to properly understand and compare the sensitivity of classification algorithms as a function of a number of candidate parameters (S/N,

period, DM for example), and help their development process.

While the performance of SPINN can and will certainly be exceeded, we wish to point out as a closing remark that it reaches levels of accuracy that could nearly handle the output of a survey conducted with SKA1-Mid. In Figure 5.1 we display the false positive rate per observed hour and per beam on our test sample. This can in turn be extrapolated to a total number of false positives per hour that would be reported on a fictitious 2,222-beam receiver installed at Parkes in the same test conditions. In particular, note that 10 false alarms per hour correspond to a recall rate of  $\simeq 90\%$ . However, we can speculate that the observing conditions will be more favorable at the SKA, because unlike the Parkes multibeam receiver, it will be a long baseline interferometer with the ability to discriminate against near-field sources of interference, and will be installed in a radio quiet zone. Therefore, SPINN or a similar classifier is arguably ready for the first phase of the SKA.

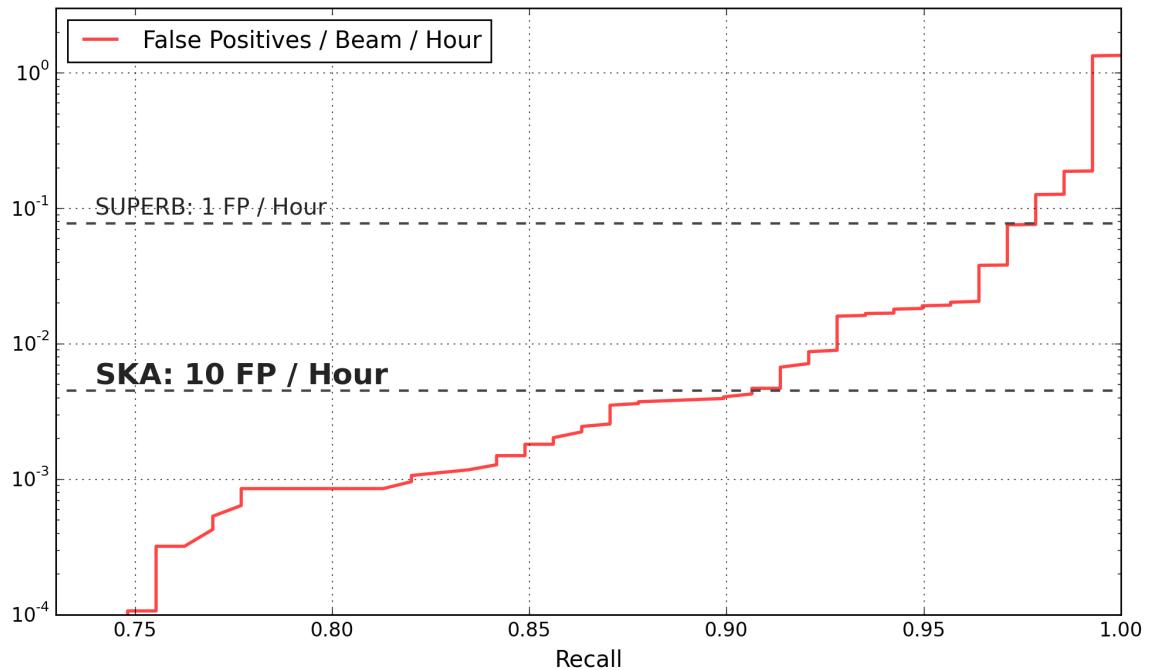


Figure 5.1 False positive rate (FPR) per observation hour and per beam on the SUPERB data set. Upper dashed line: FPR / hour / beam that must be achieved to face only a total of one false positive per hour with the 13-beam Parkes receiver, used for the SUPERB survey. Lower dashed line: if instead 2,222 beams were observed concurrently, as in the SKA1-Mid pulsar search configuration, this corresponds to a total of 10 false positives reported per hour. A classifier similar to SPINN could possibly handle SKA pulsar searches both in terms of false positive rate and classification speed.



## Bibliography

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Physical Review Letters, 116, 061102
- Alpar, M. A., Cheng, A. F., Ruderman, M. A., & Shaham, J. 1982, Nature, 300, 728
- Antoniadis, J., Freire, P. C. C., Wex, N., et al. 2013, Science, 340, 448
- Archibald, A. M., Stairs, I. H., Ransom, S. M., et al. 2009, Science, 324, 1411
- Archibald, R. F., Gotthelf, E. V., Ferdman, R. D., et al. 2016, ApJ Letters, 819, L16
- Atteia, J.-L., Boer, M., Hurley, K., et al. 1987, ApJ Letters, 320, L105
- Backer, D. C. 1970, Nature, 228, 42
- Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M., & Goss, W. M. 1982, Nature, 300, 615
- Barr, E. D. 2012, PhD thesis, University of Bonn
- . in prep.
- Barsdell, B. R., Bailes, M., Barnes, D. G., & Fluke, C. J. 2012, MNRAS, 422, 379
- Bates, S. D., Lorimer, D. R., & Verbiest, J. P. W. 2013, MNRAS, 431, 1352
- Bates, S. D., Bailes, M., Barsdell, B. R., et al. 2012, MNRAS, 427, 1052
- Bhat, N. D. R., Cordes, J. M., Camilo, F., Nice, D. J., & Lorimer, D. R. 2004, ApJ, 605, 759
- Bhattacharya, D., & van den Heuvel, E. P. J. 1991, Phys. Rep., 203, 1
- Bracewell, R. N. 2000, The Fourier Transform and Its Applications, McGraw-Hill, 3rd edition, 2000. ISBN 0-07-303938-1.
- Braun, R., Goss, W. M., Danziger, I. J., & Boksenberg, A. 1983, in IAU Symposium, Vol. 101, Supernova Remnants and their X-ray Emission, ed. J. Danziger & P. Gorenstein, 159–164
- Camilo, F., Manchester, R. N., Gaensler, B. M., Lorimer, D. R., & Sarkissian, J. 2002, ApJ Letters, 567, L71
- Camilo, F., Ng, C.-Y., Gaensler, B. M., et al. 2009, ApJ Letters, 703, L55

- Champion, D. J., Ransom, S. M., Lazarus, P., et al. 2008, *Science*, 320, 1309
- Chandola, V., Banerjee, A., & Kumar, V. 2009, *ACM Comput. Surv.*, 41, 15:1
- Chandrasekhar, S. 1931, *ApJ*, 74, 81
- Chen, K., & Ruderman, M. 1993, *ApJ*, 402, 264
- Cooley, J. W., & Tukey, J. W. 1965, *Mathematics of Computation*, 19, 297
- Cordes, J. M., & Lazio, T. J. W. 2002, ArXiv Astrophysics e-prints
- Cybenko, G. 1989, *Mathematics of Control, Signals and Systems*, 2, 303
- Damour, T., & Esposito-Farese, G. 1992, *Classical and Quantum Gravity*, 9, 2093
- Demorest, P. B., Pennucci, T., Ransom, S. M., Roberts, M. S. E., & Hessels, J. W. T. 2010, *Nature*, 467, 1081
- Dewdney, P. 2013, Document number SKA-TEL-SKO-DD-001 Revision 1
- Duncan, R. C., & Thompson, C. 1992, *ApJ Letters*, 392, L9
- Eatough, R. P., Keane, E. F., & Lyne, A. G. 2009, *MNRAS*, 395, 410
- Eatough, R. P., Kramer, M., Lyne, A. G., & Keith, M. J. 2013, *MNRAS*, 431, 292
- Eatough, R. P., Molkenthin, N., Kramer, M., et al. 2010, *MNRAS*, 407, 2443
- Edwards, R. T., Bailes, M., van Straten, W., & Britton, M. C. 2001, *MNRAS*, 326, 358
- Espinoza, C. M., Lyne, A. G., Kramer, M., Manchester, R. N., & Kaspi, V. M. 2011a, *ApJ Letters*, 741, L13
- Espinoza, C. M., Lyne, A. G., Stappers, B. W., & Kramer, M. 2011b, *MNRAS*, 414, 1679
- Faulkner, A. J., Stairs, I. H., Kramer, M., et al. 2004, *MNRAS*, 355, 147
- Fischer, D. A., Schwamb, M. E., Schawinski, K., et al. 2012, *MNRAS*, 419, 2900
- Foster, R. S., & Backer, D. C. 1990, *ApJ*, 361, 300
- Freire, P. C. C. 2013, in *IAU Symposium*, Vol. 291, Neutron Stars and Pulsars: Challenges and Opportunities after 80 years, ed. J. van Leeuwen, 243–250
- Freire, P. C. C., Bassa, C. G., Wex, N., et al. 2011, *MNRAS*, 412, 2763

- Gavriil, F. P., Gonzalez, M. E., Gotthelf, E. V., et al. 2008, *Science*, 319, 1802
- Gold, T. 1968, *Nature*, 218, 731
- Goldreich, P., & Julian, W. H. 1969, *ApJ*, 157, 869
- Harding, A. K., Contopoulos, I., & Kazanas, D. 1999, *ApJ Letters*, 525, L125
- Hastie, T., Tibshirani, R., & Friedman, J. 2009, *The elements of statistical learning: data mining, inference and prediction*, 2nd edn. (Springer)
- Hellings, R. W., & Downs, G. S. 1983, *ApJ Letters*, 265, L39
- Hessels, J. W. T., Ransom, S. M., Stairs, I. H., et al. 2006, *Science*, 311, 1901
- Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. 1968, *Nature*, 217, 709
- Hobbs, G., Archibald, A., Arzoumanian, Z., et al. 2010, *Classical and Quantum Gravity*, 27, 084013
- Hotan, A. W., van Straten, W., & Manchester, R. N. 2004, *PASA*, 21, 302
- Huba, J. D. 2013
- Hulse, R. A., & Taylor, J. H. 1975, *ApJ Letters*, 195, L51
- Jackson, J. D. 1998, *Classical Electrodynamics*, 3rd Edition, 832
- Jacoby, B., Bailes, M., Ord, S. M., Edwards, R. T., & Kulkarni, S. R. 2009, *The Astrophysical Journal*, 699, 2009
- Janssen, G., Hobbs, G., McLaughlin, M., et al. 2015, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 37
- Johnston, H. M., & Kulkarni, S. R. 1991, *ApJ*, 368, 504
- Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, *ApJ*, 673, 864
- Keane, E., Bhattacharyya, B., Kramer, M., et al. 2015, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 40
- Keane, E. F., Barr, E. D., Jameson, A., & Morello, V. in prep.

- Keane, E. F., & McLaughlin, M. A. 2011, Bulletin of the Astronomical Society of India, 39, 333
- Keane, E. F., Johnston, S., Bhandari, S., et al. 2016, Nature, 530, 453
- Keith, M. J., Eatough, R. P., Lyne, A. G., et al. 2009, MNRAS, 395, 837
- Keith, M. J., Jameson, A., van Straten, W., et al. 2010, MNRAS, 409, 619
- Kiziltan, B., & Thorsett, S. E. 2010, ApJ, 715, 335
- Kocz, J., Bailes, M., Barnes, D., Burke-Spolaor, S., & Levin, L. 2012, MNRAS, 420, 271
- Kramer, M., Stairs, I. H., Manchester, R. N., et al. 2006, Science, 314, 97
- Kulkarni, S. R., Ofek, E. O., Neill, J. D., Zheng, Z., & Juric, M. 2014, ApJ, 797, 70
- Lattimer, J. M., & Prakash, M. 2004, Science, 304, 536
- Lattimer, J. M., & Steiner, A. W. 2014, ApJ, 784, 123
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proceedings of the IEEE, 86, 2278
- LeCun, Y. A., Bottou, L., Orr, G. B., & Müller, K.-R. 2012, Neural Networks: Tricks of the Trade: Second Edition (Berlin, Heidelberg: Springer Berlin Heidelberg), 9–48
- Lee, K. J., Guillemot, L., Yue, Y. L., Kramer, M., & Champion, D. J. 2012, MNRAS, 424, 2832
- Lee, K. J., Stovall, K., Jenet, F. A., et al. 2013, MNRAS, 433, 688
- Leshem, A., van der Veen, A.-J., & Boonstra, A.-J. 2000, ApJS, 131, 355
- Levin, L. 2012, PhD thesis, Swinburne University of Technology
- Li, Y., & Zhang, B. 2016, ArXiv e-prints
- Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179
- Liu, K., Eatough, R. P., Wex, N., & Kramer, M. 2014, MNRAS, 445, 3115
- Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, Science, 318, 777
- Lorimer, D. R., & Kramer, M. 2004, Handbook of Pulsar Astronomy

- Lyne, A. G., McLaughlin, M. A., Keane, E. F., et al. 2009, MNRAS, 400, 1439
- Lyne, A. G., & Smith, F. G. 1968, Nature, 218, 124
- Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993
- Manchester, R. N., Lyne, A. G., Camilo, F., et al. 2001, MNRAS, 328, 17
- Masui, K., Lin, H.-H., Sievers, J., et al. 2015, Nature, 528, 523
- Mazets, E. P., Golentskii, S. V., Ilinskii, V. N., Aptekar, R. L., & Guryan, I. A. 1979, Nature, 282, 587
- McLaughlin, M. A., Lyne, A. G., Lorimer, D. R., et al. 2006, Nature, 439, 817
- Mereghetti, S., Pons, J. A., & Melatos, A. 2015, Space Science Reviews, 191, 315
- Olausen, S. A., & Kaspi, V. M. 2014, ApJ Suppl., 212, 6
- Oppenheimer, J. R., & Volkoff, G. M. 1939, Physical Review, 55, 374
- Papitto, A., Ferrigno, C., Bozzo, E., et al. 2013, Nature, 501, 517
- Peng, H., L., F., & Ding, C. 2005, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27, 1226
- Podsiadlowski, P., Langer, N., Poelarends, A. J. T., et al. 2004, ApJ, 612, 1044
- Portegies Zwart, S., van den Heuvel, E. P. J., van Leeuwen, J., & Nelemans, G. 2011, ApJ, 734, 55
- Radhakrishnan, V., & Srinivasan, G. 1982, Current Science, 51, 1096
- Ransom, S. M. 2001, PhD thesis, Harvard University
- Ransom, S. M., Eikenberry, S. S., & Middleditch, J. 2002, AJ, 124, 1788
- Ransom, S. M., Stairs, I. H., Archibald, A. M., et al. 2014, Nature, 505, 520
- Roberts, M. S. E. 2011, in American Institute of Physics Conference Series, Vol. 1357, American Institute of Physics Conference Series, ed. M. Burgay, N. D'Amico, P. Esposito, A. Pellizzoni, & A. Possenti, 127–130
- Rojas, R. 1996, Neural Networks - A Systematic Introduction (Springer-Verlag, Berlin, New-York)

- Ruderman, M. A., & Sutherland, P. G. 1975, ApJ, 196, 51
- Shankar, F., Bernardi, M., Sheth, R. K., et al. 2016, MNRAS
- Shannon, R. M., Ravi, V., Lentati, L. T., et al. 2015, Science, 349, 1522
- Sigurdsson, S. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 302, Radio Pulsars, ed. M. Bailes, D. J. Nice, & S. E. Thorsett, 391
- Spitler, L. G., Cordes, J. M., Hessels, J. W. T., et al. 2014, ApJ, 790, 101
- Spitler, L. G., Scholz, P., Hessels, J. W. T., et al. 2016, Nature, 531, 202
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, J. Mach. Learn. Res., 15, 1929
- Staelin, D. H., & Reifenstein, E. C. 1968, Science, 162, 1481
- Stahler, S. W., & Palla, F. 2005, The Formation of Stars, 865
- Stovall, K., Lynch, R. S., Ransom, S. M., et al. 2014, ApJ, 791, 67
- Sturrock, P. A. 1971, ApJ, 164, 529
- Tauris, T. M., & van den Heuvel, E. P. J. 2006, Formation and evolution of compact stellar X-ray sources, 623–665
- Taylor, J. H. 1974, Astron. Astrophys. Suppl., 15, 367
- Taylor, J. H., & Weisberg, J. M. 1982, ApJ, 253, 908
- Thompson, C., & Duncan, R. C. 1996, ApJ, 473, 322
- Torne, P., Eatough, R. P., Karuppusamy, R., et al. 2015, MNRAS, 451, L50
- van den Heuvel, E. P. J. 2007, in American Institute of Physics Conference Series, Vol. 924, The Multicolored Landscape of Compact Objects and Their Explosive Origins, ed. T. di Salvo, G. L. Israel, L. Piersant, L. Burderi, G. Matt, A. Tornambe, & M. T. Menna, 598–606
- Verbiest, J. P. W., Bailes, M., van Straten, W., et al. 2008, ApJ, 679, 675
- Williams, P. K. G., & Berger, E. 2016, ArXiv e-prints
- Young, M. D., Manchester, R. N., & Johnston, S. 1999, Nature, 400, 848

- Zackay, B., & Ofek, E. O. 2014, ArXiv e-prints
- Zhang, B., Harding, A. K., & Muslimov, A. G. 2000, ApJ Letters, 531, L135
- Zhang, C. M., Wang, J., Zhao, Y. H., et al. 2011, A&A, 527, A83
- Zhu, W. W., Berndsen, A., Madsen, E. C., et al. 2014, ApJ, 781, 117