NenuFAR pulsar blind survey and study of the super-dispersive effects

Mark Brionne

UNIVERSITÉ D'ORLÉANS

ÉCOLE DOCTORALE Energie, Matériaux, Sciences de la Terre et de l'Univers

Laboratoire de Physique et Chimie de l'Environnement et de l'Espace

THÈSE présentée par :

Mark BRIONNE

soutenue le : 28 février 2023

pour obtenir le grade de : Docteur de l'Université d'Orléans

Discipline/ Spécialité : Astrophysique

NenuFAR pulsar blind survey and study of the super-dispersive effects

THÈSE dirigée par :

Dr. COGNARD Ismaël	Directeur de recherche, Université d'Orléans

THÈSE co-encadrée par :

Dr. GRIEßMEIER Jean-Mathias	Astronome-adjoint, Université d'Orléans

RAPPORTEURS :

Dr. HESSELS Jason	Professor, University of Amsterdam
Dr. SMITH David	Directeur de recherche, LAB, Bordeaux

JURY :

Dr. CELESTIN Sébastien	Professeur, LPC2E, Orléans, Président du jury
Dr. SMITH David	Directeur de recherche, LAB, Bordeaux,
Dr. HESSELS Jason	Professor, University of Amsterdam
Dr. BURGAY Marta	Astronomo, Osservatorio Astronomico di Cagliari
Dr. BRIAND Carine	Astronome, LESIA, Paris
Dr. COGNARD Ismaël	Directeur de recherche, LPC2E, Orléans
Dr. GRIEßMEIER Jean-Mathias	Astronome-adjoint, Université d'Orléans

Remerciements

Je tiens en premier lieu à dire à grand merci à Jean-Mathias pour son soutien et son encadrement sans failles depuis le stage de master et durant toute cette thèse. (Désolé pour tout le travail supplémentaire que je t'ai donné.) Je veux également remercier Ismaël pour sa supervision toujours pertinente, et pour toutes ces explications supplémentaires : "juste pour ta culture scientifique". Et puis merci pour votre gentillesse et votre bonne humeur ! Et plus globalement, merci a toutes les personnes qui m'ont permis de faire cette thèse.

Je voudrais également remercier tout particulièrement Louis, pour son immense aide, et pour m'avoir appris énormément de choses sur les pulsars. Merci aussi pour les discussions et tous les bons moments. Et bravo pour m'avoir supporté comme collègue de bureau pendant de nombreux mois !

Un merci également pour les différents membres de l'équipe pour leur accueil chaleureux, les sympathiques déjeuners de midi, et les pauses café autour de discussions parfois improbables. Un remerciement particulier pour Lucas, pour sa gentillesse perpétuelle et toutes les discussions intéressantes.

Enfin, un merci pour les différents stagiaires que j'ai croisé durant ma thèse : Musphar, Cyprien, El-Mehdi. Un remerciement particulier pour Raphaël, qui a lui aussi du me supporter pendant quelques mois dans le bureau. Merci pour ses bons moments, et bonne chance pour la suite. Une pensée aussi pour Anaïs, au bout de trois ans de thèse (et un même un peu plus grâce au Covid), merci pour ce soutien dans l'effort et ton humour.

Je finirai par un dernier merci pour les enseignants qui m'ont appris leurs connaissances et aidé (merci Franck Tibault pour ta dureté mais surtout ta justesse et ta confiance). Je dois aussi remercier les personnes qui m'ont fait, étant enfant, découvrir et aimer la science et plus particulièrement l'Univers, et sans qui je ne serais pas ici en train de rédiger cette thèse...

Merci maman et grand-mère
pour votre soutien indéfectible,
même quand personne n'y croyait.

Contents

List of Figures

List of Tables

Introduction

Pulsars are a particular type of star discovered in 1968 (Hewish et al. 1968). They are neutron stars featuring a high rotation velocity and a high magnetic field. In the following years, many pulsars have been discovered. Since the seventies, several sky surveys have been performed in the northern (Lazarus et al. 2015) and the southern hemisphere using different telescopes (Manchester et al. 2001). In 2022, several thousand are known.

Although the first pulsars were discovered at 81.5 MHz, the large majority of the pulsars have been discovered at frequencies higher than 300 MHz. As a result, the sky has been rarely surveyed at lower frequencies. Since the 2000s, a new generation of telescopes has had a sufficient sensitivity to be able to observe pulsars at frequencies between the ionospheric cutoff at about 10 MHz and 300 MHz. NenuFAR is a new phased-array radio telescope of almost 2 000 antennas located in France and observing from 15 to 85 MHz (Zarka et al. in prep.).

Using this new low-frequency telescope, a pulsar census has been performed a few years ago, consisting to observe all the known pulsars to elaborate the first NenuFAR pulsar catalog (Bondonneau et al. in preparation). Due to the success of the census with the detection of about 180 pulsars, a survey of pulsars in the northern hemisphere has been planned.

The NenuFAR pulsar survey of the northern sky is planned for several years. The object of this thesis was to develop and realize the first stage of the survey. The present manuscript is divided into three parts. The first part presents a review of the theory of the pulsars in the first chapter, and the second chapter is dedicated to an overview of the telescope NenuFAR. The second part of the manuscript exposes the development of the first stage of the survey starting from the definition of the sky to observe until the analysis of the first pulsar candidates found in the first months of data.

Pulsars emit pulsating radio waves which interact with the interstellar medium between the pulsar and the Earth. Since the first discovery, the effect of dispersion of the electromagnetic wave of the pulsar through the plasma in interstellar space is known. However, at the end of the 1980s, some pulsars studies proposed different super-dispersive effects, causing deviations compared to the usual cold plasma dispersion relation (Shitov & Malofeev 1985; Kuzmin 1986; Shitov et al. 1988).

The plasma dispersion relation is a frequency-dependent relation, resulting that the amplitude of potential deviations to this law being larger at low frequencies. The dispersion generates a time delay between different frequencies, leading that it is required to a large relative bandwidth to detect the small variations introduced by the super-dispersive effects. With observations at frequencies below 100 MHz in about three octaves of bandwidth, NenuFAR is a telescope particularly interesting to detect potential deviations to the usual plasma dispersion relation.

The third part of the manuscript presents a preliminary study of the super-dispersive effects on some powerful pulsars using NenuFAR. The part begins with an explanation of the context of the study and exposes the choice made to perform measurements of dispersion on single pulses. The next chapter develops the method created to realize these measurements. Finally, the part ends with the preliminary obtained results and some propositions about these results.

Résumé du chapitre :

Les pulsars sont des étoiles à neutrons découvertes en 1968 (Hewish et al. 1968), disposant d'une importante vitesse de rotation et d'un fort champ magnétique. Bien que découvert à 81,5 MHz, la grande majorité des pulsars ont été découverts au-dessus de 300 MHz. Ainsi, peu de relevés du ciel ont été effectués à plus basses fréquences. Depuis les années 2000, une nouvelle génération de télescopes dispose d'une sensibilité suffisante pour observer les pulsars entre 10 et 300 MHz. NenuFAR est un nouveau télescope à réseau phasé situé en France observant le ciel entre 10 et 85 MHz (Zarka et al. in prep.).

En utilisant ce nouveau télescope basses fréquences, un relevé des pulsars connus a été effectué ces dernières années dans le but d'élaborer le premier catalogue de pulsars de NenuFAR (Bondonneau et al. in preparation). Suite à la détection de 180 pulsars par ce premier relevé, il a été décidé de poursuivre avec un relevé aveugle du ciel de l'hémisphère nord. Le relevé aveugle de NenuFAR devant s'étendre sur plusieurs années, le but de cette thèse était de développer et réaliser la première étape de ce relevé.

Le manuscrit est divisé en trois parties. La première partie, scindée en deux chapitres, présente dans le premier la théorie des pulsars, et dans le second le télescope NenuFAR. La seconde partie du manuscrit décrit les différentes phases du développement du relevé, allant de la définition de la grille de pointage du ciel jusqu'à l'analyse des premiers candidats trouvés dans les premiers mois de données.

Les pulsars émettent des ondes radio interagissant avec le milieu interstellaire entre celui-ci et la Terre. Depuis la découverte des pulsars, l'effet de dispersion de l'onde électromagnétique au sein d'un plasma est bien connu. Cependant, depuis la fin des années 80, certaines études ont proposé des effets pouvant causer des écarts avec la loi standard de dispersion par un plasma froid (Shitov & Malofeev 1985; Kuzmin 1986; Shitov et al. 1988).

La relation de dispersion étant dépendante de la fréquence, l'amplitude des potentiels écarts doit augmenter lorsque l'on descend vers les basses fréquences. Cette dispersion génère, entre deux fréquences données, un décalage temporel, résultant de la nécessité d'avoir une bande de fréquences relativement étendue de manière à pouvoir distinguer ces éventuels fins écarts. Grâce à une largeur de bande d'environ trois octaves située en dessous de 100 MHz, NenuFAR est le parfait instrument pour tenter de détecter des effets super-dispersifs.

La troisième partie de ce manuscrit présente une étude préliminaire portant sur la recherche de ces effets super-dispersifs avec NenuFAR. Cette dernière partie commence par détailler le contexte dans lequel se place cette étude et les choix faits pour mettre en œuvre celle-ci, consistant dans la mesure de dispersion sur des impulsions individuelles. Le chapitre suivant détaille la méthode utilisée, et enfin le dernier chapitre expose les résultats préliminaires obtenus sur quelques observations avec quelques propositions pour aller au-delà.

Part I

Pulsars and NenuFAR

Chapter 1

Pulsars

In 1967, an observing program was carried out by A. Hewish and J. Bell to study the angular structure of compact radio sources, using the scintillation produced by the interstellar medium (ISM hereafter). This program observed the sky from -8 to +44°at a central frequency of 81.5 MHz with 1 MHz of bandwidth, using a radio telescope located at the *Mullard Radio Astronomy Observatory*: the *Interplanetary Scintillation Array*.

Rapidly, J. Bell noticed a weak periodic signal every 1.337 s (Figure 1.1) (Hewish et al. 1968). After investigation, the periodicity of this signal was extremely precise, with an accuracy of one part in 10^7. In the beginning, the signal was interpreted as man-made interference. However, the fixed position in the sky excluded a terrestrial origin, and the absence of a parallax showed that the source was at a distance greater than one thousand light-years from Earth.

Finally, the search of similar sources allowed to detect, less than two months after the first discovery of B1919+21, of three other pulsating sources: B0834+06, B0950+08, and B1133+16 (Pilkington et al. 1968), suggesting that this type of object is relatively common.

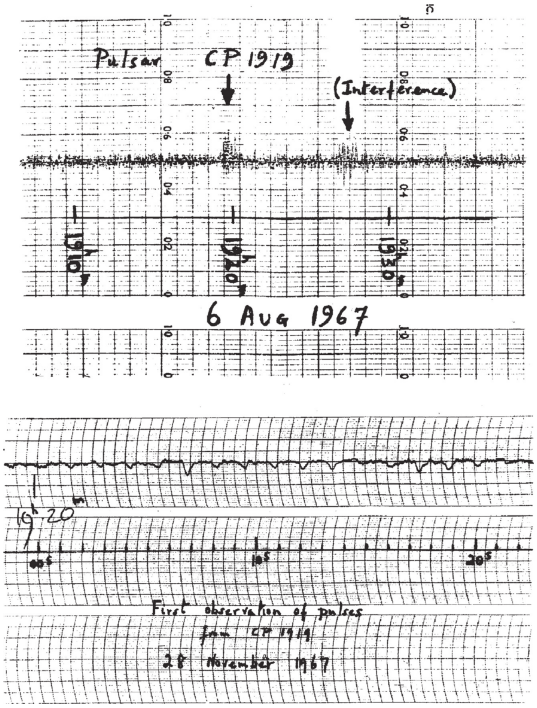


Figure 1.1: Graph of the discovery of the first pulsar B1919+21 (Hewish et al. 1968).

1.1 The astronomical object “pulsar"

1.1.1 Neutron stars

A pulsar can be described as a compact star featuring a high rotation velocity from about one millisecond to several dozens of seconds, and a high magnetic field from about 10^8 to 10^{15} G. From the beginning, for the compact star, the neutron star was a possible candidate to explain the rotation speed and magnetic field of the pulsars (Pacini 1968; Gold 1968).[1]

In 1934, Baade & Zwicky (1934) proposed that there is a link between neutron stars and supernova remnants. Some years after, based on observations of the expansion rate of the Crab nebula, Baade (1942) can't identify the type of star able to sufficiently excite the nebula. At the moment of the discovery of the first pulsars, two main theories exist to explain pulsars. In 1966, Meltzer & Thorne (1966) calculated that the radial oscillations of white dwarfs can create a periodic signal of some seconds. The second theory was proposed by Pacini (1967), which demonstrates that a neutron star with a rapid rotation and a high magnetic field can produce the needed energy for the Crab nebula. Contrary to a white dwarf, a neutron star can generate a rapid signal, featuring, in addition, a slowdown of the period. The discovery of a pulsar (PSR B0531+21) in the Crab nebula with a period of only 33 ms by Staelin & Reifenstein (1968) with the Green Bank telescope, was the first evidence in favor of the neutron star theory. Finally, the measure of a slowdown in the period of the Crab pulsar by Richards & Comella (1969) has definitely ruled out the white dwarf theory.

A neutron star is a kind of star proposed at the origin by Lev Landau (1932), which is composed, as its name indicates, of neutrons. At the end of the life of a massive star, a gravitational collapse occurs, resulting in a supernova. For stars with a mass from 8 to $18-20$ $M_\odot$, it remains a star, whose mass is theoretically contained between the Chandrasekhar mass of 1.4 $M_\odot$, and the Tolman–Oppenheimer–Volkoff limit of $2.2-2.9$ $M_\odot$ (Kalogera & Baym 1996). The Chandrasekhar mass corresponds to the mass limit to overcome the degeneracy pressure of the electrons and then passes beyond the stage of the white dwarf. On the opposite side, the Tolman–Oppenheimer–Volkoff limit corresponds to the maximum mass preventing the gravitational collapsing of the remnant star in a black hole. Due to the degeneracy pressure of the neutrons, the neutron star has an extreme density, where the matter is compacted in a small star.

In the standard model of a neutron star (Shapiro & Teukolsky 1983), the matter at the surface has a density of 10^9 kg.m^{-3} and is composed of degenerate electrons and iron nuclei. Below the surface, protons and electrons begin to combine to give a matter rich in neutrons up to a density of 10^{14} kg.m^{-3}. Beyond this density, the matter becomes totally composed of neutrons, with an increasing density towards the deeper layers. Finally, in the center, when the density exceeds 10^{17} kg.m^{-3}, the neutron matter becomes a sea of free superfluid neutrons.

A neutron star is an object with a mass greater or equal to the theoretical limit of 1.4 $M_\odot$. Due to the important gravitational field, the radius is more difficult to estimate. Glendenning (1992) derives a lower limit for the radius of 1.5 Schwarzschild radii R_S, corresponding to about 6.2 km for a neutron star of 1.4 $M_\odot$:

$$R_{min} \;=\; 1.5R_S \;=\; \frac{3\;G\cdot M}{c^2} \tag{1.1}$$

With G the gravitational constant, M the mass of the neutron star, and c the speed of light. The maximum radius that a neutron star can reach represents a limit of stability, corresponding to the largest radius for which the centrifugal forces can be compensated. The upper limit is then determined by the rotation period of the star P:

$$R_{max} \;=\; \left(\frac{G\cdot M\cdot P^2}{4\pi^2}\right)^{1/3} \tag{1.2}$$

Recent observations realized in X-rays with NICER and XMM-Newton space telescopes, have estimated the neutron star radius. Based on fits of the hot spot patterns, the measurements give a radius between 11.8 and 13.1 km (Miller et al. 2019).

[1] This introduction chapter is partially inspired and based on the Lorimer & Kramer (2012) and Lyne & Graham-Smith (2012)

1.1.2 Lighthouse effect

As mentioned before, a pulsar is characterized by the period of rotation of the neutron star and the slowdown of this period. Also, another main property of a pulsar is the presence of a magnetic field. In the current pulsar model, the observation of a pulsating signal is explained by the fact that the magnetic axis is unaligned with the rotation axis. The magnetic field allows acceleration of the plasma around the star, producing a beam around the magnetic axis. In the case of a crossing between this axis and the line of sight with Earth, the beam is seen once per rotation of the neutron star.

Since the first observations of pulsars, the measurements have shown a slowdown of the period with time. This slowdown can be related to the total energy emitted by the neutron star $\dot{E}$, which corresponds to the rate of lost rotational kinetic energy E_{rot}. This radiated energy can be defined relative to the two characteristics of the pulsar: the period P, and the period derivative $\dot{P}$.

$$\dot{E} \;=\; -\frac{dE_{rot}}{dt} \;=\; -\frac{d}{dt}\left(\frac{1}{2}\, I \cdot \Omega^2\right) \;=\; 4\pi^2\, I \cdot \frac{\dot{P}}{P^3} \tag{1.3}$$

Here $I = k \cdot M \cdot R^2$ represents the moment of inertia, where k depends on the shape and density profile of the star (e.g. for a sphere of uniform density k = 2/5), and $\Omega = 2\pi/P$ is the angular velocity. For a canonical pulsar with a period of one second, a period derivative of 10^{-15} s.s^{-1}, and a moment of inertia of 10^{48} kg.m^{-2}, the *spin-down luminosity* $\dot{E}$ of the pulsar is about 3.95 .10^{24} J.s^{-1}.

Furthermore, according to classical electrodynamics, one can relate this energy with the energy radiated by a rotating magnetic dipole.

$$\dot{E} \;=\; \frac{2}{3c^3} \cdot \|\vec{m}\|^2 \cdot \Omega^4 \cdot \sin^2 \alpha \tag{1.4}$$

Here $\vec{m}$ is the magnetic moment, and α is the angle between the spin axis and the magnetic moment. Combining Equations (1.3) and (1.4), we can derive a power law relation between the rotational frequency $\nu = 1/P$ and the derivative $\dot{\nu}$:

$$\dot{\nu} \;=\; -K \cdot \nu^n \tag{1.5}$$

This relation is a function of a constant K depending on the pulsar and of the *braking index* n, which has a value of three in the case of an ideal dipole. However, in reality, because of dissipation effects, the value of the braking index is between 1.4 and 2.9 (Kaspi & Helfand 2002).

Thanks to these three parameters: P, $\dot{P}$, and n, and according to the spin-down theory, it is possible to calculate some physical quantities about the pulsar. Using the differential Equation (1.5), it is possible to estimate the age of the pulsar T:

$$T \;=\; \frac{P}{(n-1)\dot{P}} \cdot \left\{1 - \left(\frac{P_0}{P}\right)^{n-1}\right\} \tag{1.6}$$

Starting with an ideal dipole, and with the assumption that the initial period P_0 is considerably shorter than the current period P, we obtain the *characteristic age* T of the pulsar. For a canonical pulsar with $P = 1$ s and $\dot{P} = 10^{-15}$ s.s^{-1}, the result gives an age of 15.8 Myr.

Another interesting quantity to know is the value of the surface magnetic field B_S. According to the relation between the magnetic field and the magnetic moment: $\|\vec{B}\| \approx \|\vec{m}\|/r^3$, we can derive an estimation of the magnetic field B_S at the surface R. Combining the Equations (1.3) and (1.4), and always using the same assumptions as for the estimation of the characteristic age, we obtain for an ideal dipole:

$$B_S \;=\; \sqrt{\frac{3c^3}{8\pi^2} \cdot \frac{I}{R^6 \cdot \sin^2 \alpha} \cdot P \cdot \dot{P}} \tag{1.7}$$

Similarly to the age, we can define a *characteristic surface magnetic field* corresponding to a canonical pulsar of 10 km of radius and with an angle α of 90°. For the previous typical values of period and period derivative, it gives us a value of 10^{12} G.

1.1.3 Pulsar magnetosphere

In 1969, Goldreich & Julian (1969) proposed a simple model for neutron star, where the magnetic dipole moment is aligned with the rotation axis. At any point inside the neutron star, the induced magnetic field produces an induced electric field. This field results in a distribution of charges which creates a balancing electric field to obtain a force-free state in the whole star.

The repartition of the charges within the star generates a surface charge layer. Due to the high strength of the electromagnetic field, the creating electric field at the surface is greater than the gravity, allowing extraction of charged particles from the neutron star surface. Outside the star, the particles follow the magnetic field lines of the dipolar external field.

The magnetic energy density being greater than the kinetic energy density, the plasma outside the neutron star is forced to move in corotation with the star. A limit can be therefore defined at the radius where the speed of rotation is equal to the speed of light. This radius R_{LC} is called the *light cylinder*, and can be expressed relative to the period of the pulsar P with the following equation:

$$R_{LC} = \frac{c}{\Omega} = \frac{c}{2\pi} \cdot P \tag{1.8}$$

Here Ω represents the angular velocity, and c is the speed of light. As illustrated in Figure 1.2, the consequence of this light cylinder is splitting between the magnetic field lines contained inside the light cylinder which are closed field lines, and these reaching the outside region which are open field lines. The

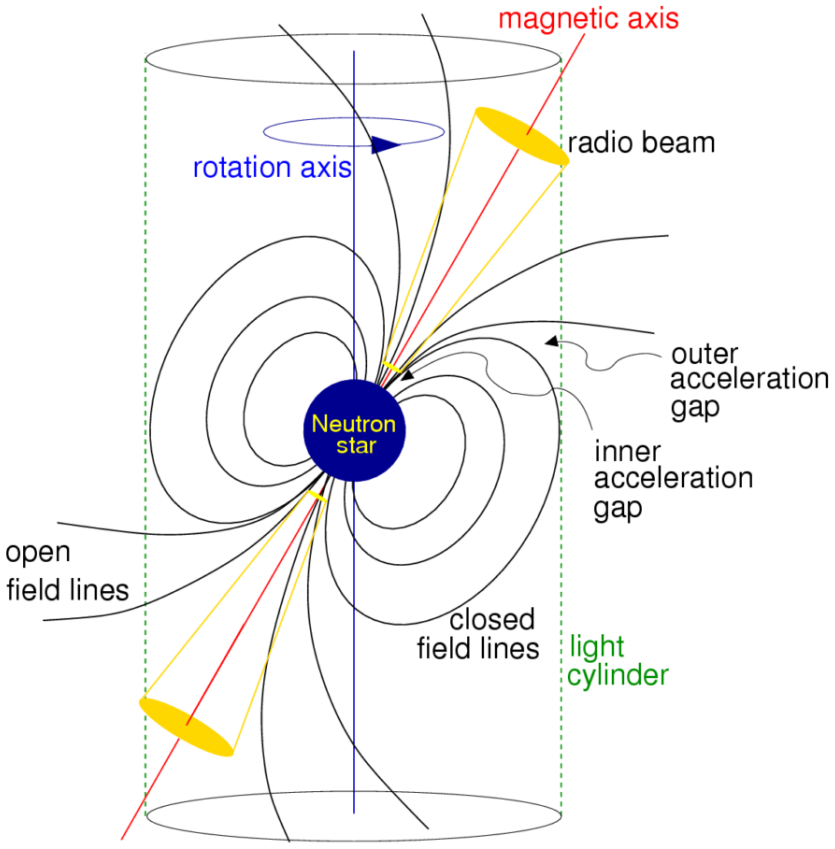


Figure 1.2: Toy model of a pulsar where the central star is a rotating neutron star featuring a misaligned magnetic axis that generates a radio beam (Lorimer & Kramer 2012). The magnetic field produces two types of field lines. The closed field lines are entirely contained inside the light cylinder, and the open field lines reach the region outside the light cylinder.

polar cap region can be then defined by the location on the surface crossed by the last open field line which is tangential to the light cylinder.

For a dipolar field, in polar coordinates relative to the center of the star and the magnetic axis, there is the property that $\sin^2(\theta)/r$ is constant for a given field line. The angular size of the polar cap θ_p can then be determined at the radius of the star R:

$$\frac{\sin^2(\theta_p)}{R} = \frac{1}{R_{LC}} \Leftrightarrow \sin(\theta_p) = \sqrt{\frac{2\pi\ R}{c \cdot P}} \tag{1.9}$$

Many models of radio emission need the presence of gap regions where the force-free state is unsatisfied. Sturrock (1971) and Ruderman & Sutherland (1975a) have proposed a model of pair cascade in a *polar gap* located above the polar cap. The particles are accelerated by the residual electric field in the gap to reach relativistic energies. Following the magnetic field lines, they can produce gamma-ray photons by curvature emission (Ruderman & Sutherland 1975a) or inverse Compton scattering (Daugherty & Harding 1986). With the intense magnetic field, the gamma photon can create an electron-positron pair. These new particles can do the same process, giving an avalanche of secondary pair plasma, which can allow producing radio emission at some distance from the neutron star.

1.1.4 Radio beam

The plasma present above the polar cap can produce radio emission, which is consequently located in a small region giving the radio beam of the pulsar. Radhakrishnan & Cooke (1969) and Komesaroff et al. (1970) have developed the standard model of a cone-shaped beam centered on the magnetic axis.

The plasma follows the magnetic field lines from the surface of the polar cap and emits photons tangentially to the field line. A radio beam is then produced in a cone around the magnetic axis (yellow cone in Figure 1.2). The detected pulse is finally caused by the passage of the line of sight within the beam, illustrated by the circle parallel to the equator of the sphere in Figure 1.3. The width of the observed pulse W, as a fraction of the rotational period, can then be geometrically expressed relative to the angular radius of the cone ρ (Gil et al. 1984).

$$\sin^2\left(\frac{W}{4}\right) = \frac{\sin^2(\rho/2) - \sin^2(\beta/2)}{\sin(\alpha) \cdot \sin(\alpha+\beta)} \tag{1.10}$$

The width of the pulse depends on two geometric parameters illustrated in Figure 1.3: α corresponding to the angle between the rotation and the magnetic axes, and the impact parameter β representing the minimum angular distance between the magnetic axis and the line of sight.

Moreover, the emission being in the region of the open field lines, using polar coordinates, it is equally possible to directly relate the angular radius of the emission cone ρ with the angle between the magnetic axis and the region of emission θ (Gangadhara & Gupta 2001).

$$\tan\theta = \frac{3}{2\tan\rho} \pm \sqrt{2 + \left(\frac{3}{2\tan\rho}\right)^2} \tag{1.11}$$

The emission beam being a cone, this angular radius is dependent on the radial position in the cone, i.e. of the emission height. Taking the property used in Equation (1.9), it is then possible to relate the width of the emission cone θ with the altitude of the emission r_{em} and the period of the pulsar P. In the case of an emission close to the magnetic axis, i.e. with ρ and θ small, we can approximate the previous relation by obtaining the following relation:

$$\rho \approx \frac{3}{2}\theta \approx \sqrt{\frac{9\pi}{2c} \cdot \frac{r_{em}}{P}} \tag{1.12}$$

This last relation shows that slow pulsars have a smaller emission cone than faster pulsars. Consequently, they have a lower probability that their emission cone crosses the line of sight.

Based on observations showing variations of the pulse width with the frequency, Cordes (1978) proposed a model to explain why the width seems to decrease when the frequency increases. The *radius to*

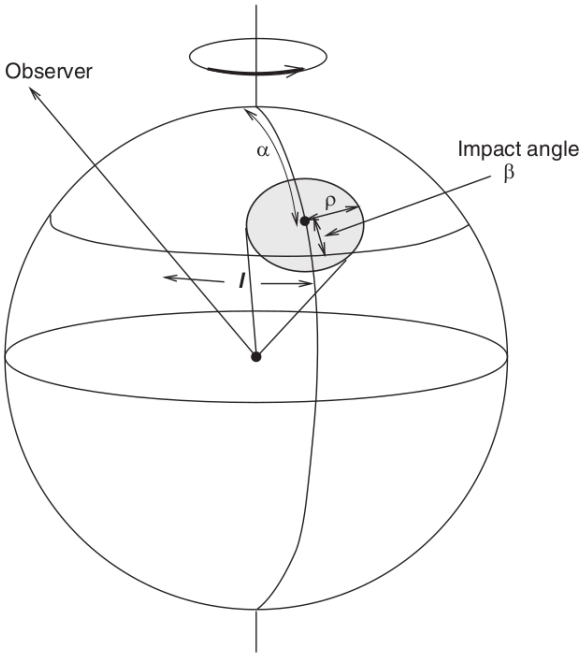


Figure 1.3: Scheme of the geometry of the emission cone of radius ρ. The center of the cone is the magnetic axis which is inclined by an angle α from the rotation axis. The line of sight crosses the beam following the circle parallel to the equator of the sphere, and β represents the minimum distance between the line of sight and the magnetic axis. (Lyne & Graham-Smith 2012)

frequency mapping model (RFM) advances that the frequency of emission is related to the altitude of emission, where the higher the emitted radio frequency, the lower the emission altitude.

In the RFM model, the altitude of emission r_{em} can be expressed, for example, with a power law relative to the frequency ν. Kijak & Gil (2003) have measured empirical values to describe the relation between the emission height, the frequency, and the two parameters period P and period derivative $\dot{P}$:

$$r_{em} \;=\; 400 \text{ km} \pm 80 \cdot \frac{(P^{0.30}/s) \cdot (\dot{P}^{0.07}/10^{-15} s.s^{-1})}{(\nu^{0.26}/GHz)} \tag{1.13}$$

Other measurements from polarisation studies have been realized and are consistent with the results of this geometrical model. The RFM consequently allows, combining with the polar gap emission model and the Goldreich-Julian model, to model the emission beam of a pulsar, explaining many differences seen in the observed pulses, such as the frequency evolution of the pulse width.

1.2 Observation of pulsars

1.2.1 Integrated profile

Because of the lighthouse effect, at each rotation of the neutron star, the beam of the pulsar points toward the Earth, and radio telescopes detect a pulse corresponding to the passing of the beam through the line of sight. The rotation of the star being extremely constant, the pulsar appears as a series of single pulses periodically spaced by a time lapse equal to the period of the neutron star.

The first pulsars, such as the first one B1919+21 by J. Bell, were found by the detection of these periodic pulses. However, except for the strongest pulsars, the majority of pulsars exhibit weak pulses which cannot be clearly noticed relative to the noise background. To remedy this, a method called *folding* is used. It consists in using the quasi-perfect periodicity of the pulsars, to perform a summation of the single pulses relative to the rotational phase of the pulsar. As a result, we obtain a time series of one average period showing an *integrated pulse profile*.

This profile represents the average of all pulses during a given observation, and by extension, provides a representation of the power emitted by the different parts of the beam. The first interesting property of the profile is that, although there are differences between individual pulses, the profile of a pulsar is very constant, without changing the shape from one observation to another. The second interesting property

of the pulsar is that the shape of the profile is complex and varies from one pulsar to another. Indeed, the profile of a pulsar is unique and is a footprint of this pulsar.

Figure 1.4 shows four different profiles resulting from observations with NenuFAR, which illustrate the variability of the shapes of profiles. B0919+06 and B1508+55 present a profile with a single peak but with a different shape. The two others correspond to multiple peak profiles with two peaks for B1133+16 and three components for B1237+25. Some pulsars can present more peaks like B1919+21, the first discovered, which has five components. There is also a particular type of profile featuring a so-called interpulse shifted by half of the period of rotation of the pulsar, corresponding to the case where the rotation and the magnetic axis are orthogonal, allowing us to see the two opposite poles.

To explain the disparity of the observed profiles, the cone-shaped model for the emission beam is insufficient, and a hollow cone model was proposed by Oster & Sieber (1976). Also, in order to explain the pulsars showing several components, this model was complexified to a nested hollow cone with a multiple ring structure (Sieber & Oster 1977; Rankin 1993), and supplemented with a core cone (as opposed to the conal components of the ring structure) located just on the magnetic axis (Backer 1976). Another model tries to explain the complexity of the profiles rather than using a cone where the emission regions are localized in many patches (Lyne & Manchester 1988). Figure 1.5 shows two schemes representing the way to generate different types of profiles. More recently, Manchester et al. (2010) and Desvignes et al. (2019) have used the precession of two pulsars in binary systems with a massive companion, to observe different parts of the beam. The obtained maps of the emission of the beams seem to be going in favor of a patchy structure rather than a simple core-cone structure.

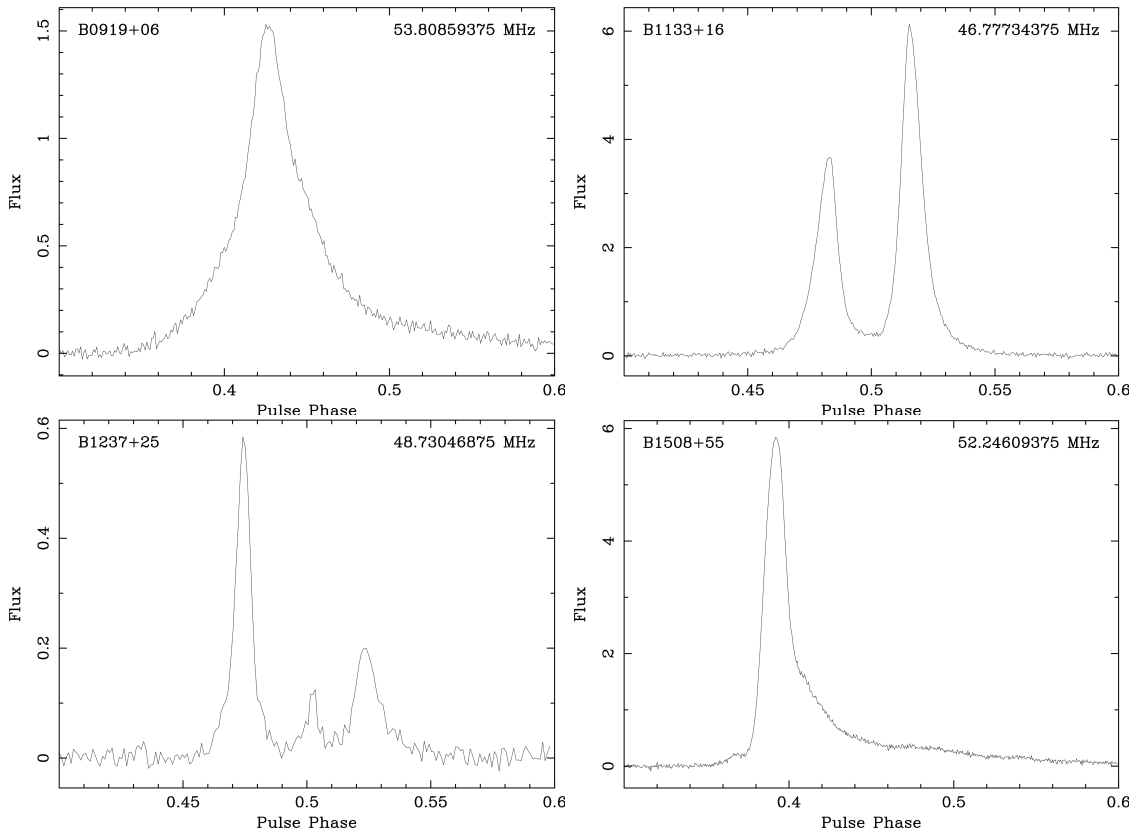


Figure 1.4: Integrated pulse profiles of four pulsars observed with NenuFAR. The profiles are zoomed around the pulse in order to see the different shapes with two single peak (B0919+06 and B1508+55), a double peak (B1133+16), and a pulsar featuring three peaks (B1237+25).

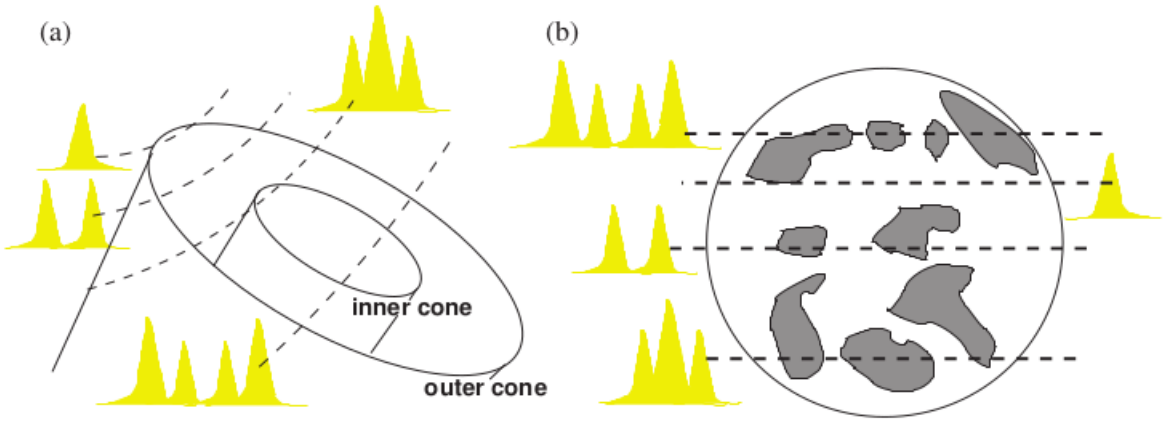


Figure 1.5: (a) Nested hollow cone model allowing to explain profiles with multiple components. (b) Model of patchy structure in the beam. (Lorimer & Kramer 2012)

The integrated pulse profile being constant from one observation to another is certainly a footprint of the pulsar. However, to model the emission beam of pulsars, the RFM model was created to take the frequency variations of the pulses into account (see Section 1.1.4). In fact, the profile remains constant for a given frequency and is distinct between two frequencies.

Moreover, the evolution in the frequency of the profile differs from one pulsar to another in multiple ways. These variations can be of various forms: change of the width of the pulse, modification of the spacing between the different components of the profile, or variation of the relative amplitude between the components. Sometimes even, the frequency evolution is more abrupt with the appearance or the disappearance of a component radically modifying the profile.

Figure 1.6 shows the evolution in the frequency of the profile of B0809+74 such as observing with NenuFAR for different frequencies between 25 and 85 MHz. B0809+74 is a good example of frequency variations, where the second component which is the minor component at higher frequencies becomes the major one at 35 MHz to be finally the last peak still visible below 15 MHz. Also, at 75 MHz, this second component is close to the major one and drifts in phase at lower frequencies.

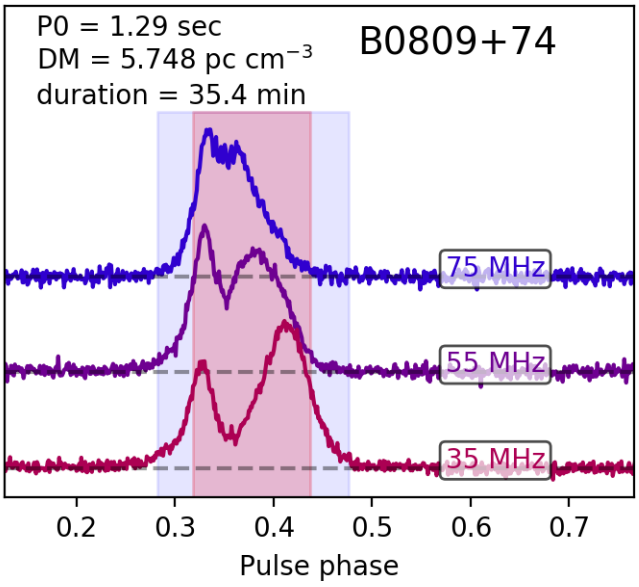


Figure 1.6: Frequency evolution of the integrated pulse profile of B0809+74 as observed with NenuFAR for three frequencies: 35, 55, and 75 MHz, integrated on 20 MHz of bandwidth. The shaded regions represent the minimum and maximum pulse widths.

1.2.2 Pulsar spectra

In addition to the frequency dependence of the integrated profile, there is also a dependence of the relative flux density on frequency. Pulsars are detected at radio wavelengths in a large continuous broadband frequency range from 10 MHz (Zakharenko et al. 2013) to 87 GHz (Morris et al. 1997). Sieber (1973) has shown that all pulsars feature a spectrum that can be modeled by power laws. A part of the pulsars follows a simple power law, and others present a low-frequency turnover, needing to add a second power law for the low frequencies. The spectra of the pulsars are different from one pulsar to another. Jankowski et al. (2018) have shown, in the case of the simple power law spectra, that the spectral indices characterizing the power laws are distributed according to a broad log-normal distribution of standard deviation 0.54.

At high frequencies, greater than 400 MHz, Maron et al. (2000) and Jankowski et al. (2018) have shown that the vast majority of the pulsars (about 80%) can be modeled with a simple power law. Based on observations of 281 pulsars between 400 MHz and 23 GHz, Maron et al. (2000) obtained a mean spectral index of -1.8 ± 0.2. Also, on a smaller range from 721 MHz to 3.1 GHz, but with a larger number of pulsars of 441, Jankowski et al. (2018) obtained a weighted mean spectral index of -1.60 ± 0.03. However, some pulsars present a modification of the spectral index, generally around 1 GHz, which need to be modeled with a broken power law. A hypothesis for these GPS pulsars (*Gigahertz Peaked Spectra*) would be the free-free absorption by an ionized surrounding (Kijak et al. 2007).

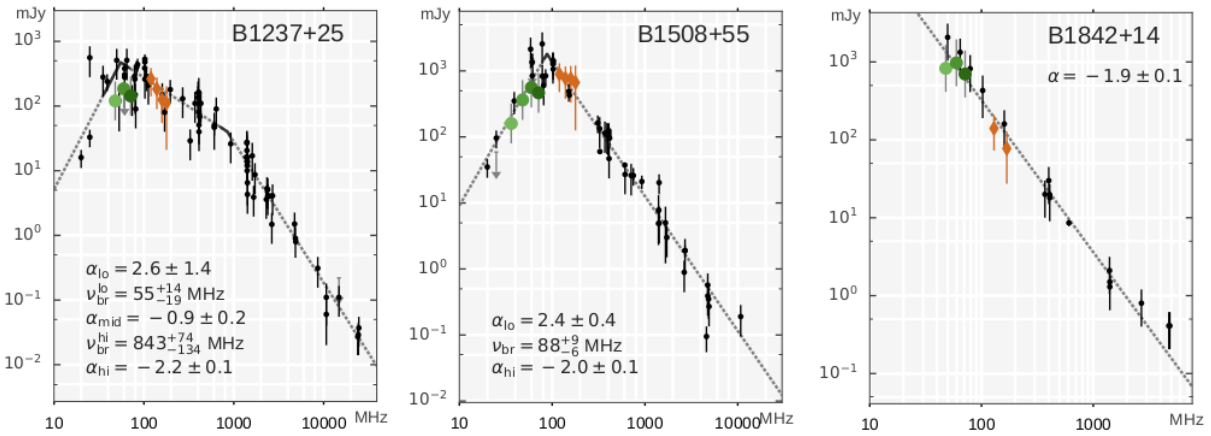


Figure 1.7: Spectra of pulsars showing three types of power law (Bilous et al. 2020). Left: double broken spectra of B1237+25 with a GHz break and a low-frequency turnover. Middle: spectrum of B1508+55 following a double power law with a low-frequency turnover. Right: single power-law spectrum of B1842+14.

At lower frequencies, some spectra of pulsars continue following a simple power law similar to the higher frequencies. As for the higher frequencies, Bilous et al. (2020) have also found more complex spectra, showing a modification of the spectral index around the frequency range $100 - 300$ MHz. Of the pulsars observed at frequencies lower than 150 MHz, about 2/3 of them indeed show a spectral turnover with a spectral index ranging from 0.1 ± 0.3 to 4.8 ± 1.4 (Bilous et al. 2020). In Figure 1.7, B1237+25 and B1508+55 present two classic examples of low-frequency turnover around $100 - 200$ MHz. In a few cases, this low-frequency changing occurs in addition to the GHz break to obtain a double broken power law (as B1237+25 on the left panel of Figure 1.7).

1.2.3 Variability of individual pulses

Generally, pulsars present a constant pulse profile in time. However, some pulsars feature variations in their emission, yielding a modification of their profile. Some different types of variations appear on different time scales.

The first type of time variation of the emission is called *giant pulse*. Some pulsars, such as the Crab pulsar (Cordes et al. 2004), have a variability of the power of their emission, emitting sometimes a huge individual pulse (related or not to the normal pulse), whose intensity can be 1 000 times larger than the

intensity of normal pulses. The left panel of Figure 1.8 shows a giant pulse of the Crab pulsar during an observation using NenuFAR. The pulse is largely more powerful than the neighboring pulses, and due to the scattering, obtains a broadening over 30 periods of the pulsar, resulting in a complete erasure of the 29 next standard pulses.

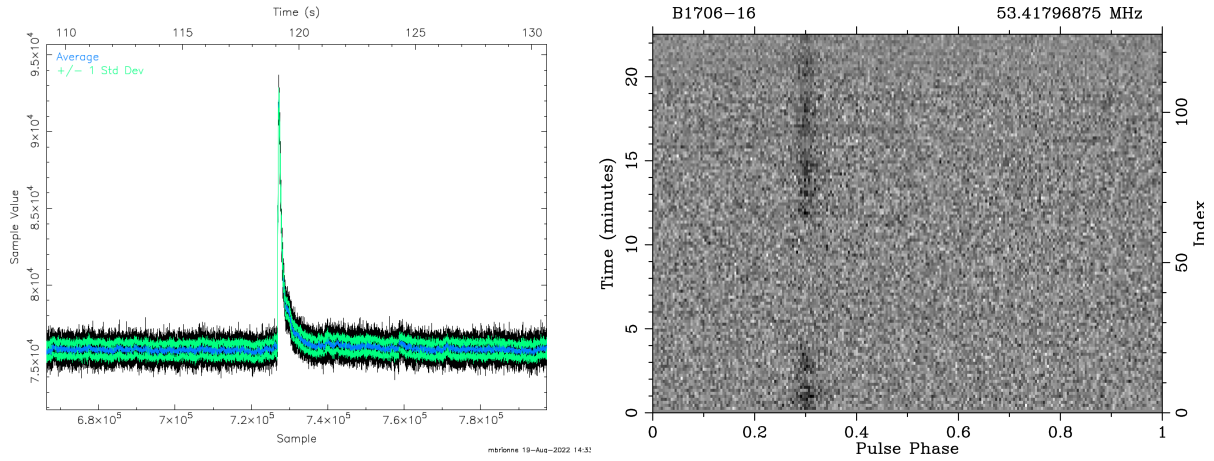


Figure 1.8: Left: time series of about 20 s during an observation using NenuFAR where appears a giant pulse emitted by the Crab pulsar. The duration of the pulse is about 1 s representing about 30 periods ($P \sim 33$ ms). The pulse is so powerful that the standard pulses around are unseen. Right: nulling of B1706-16 for about 5 minutes during an observation carried out with NenuFAR.

A second type of time variability is the *pulse nulling* (Backer 1970b), where the emission stops completely for a certain time. In extreme cases, the nulling can be sufficient to make the pulsar invisible during an observation. Figure 1.8 presents in the right panel the pulsar B1706-16 which emits for about 4 minutes before a nulling phase of about 5 minutes and finally emits again until the end of the observation.

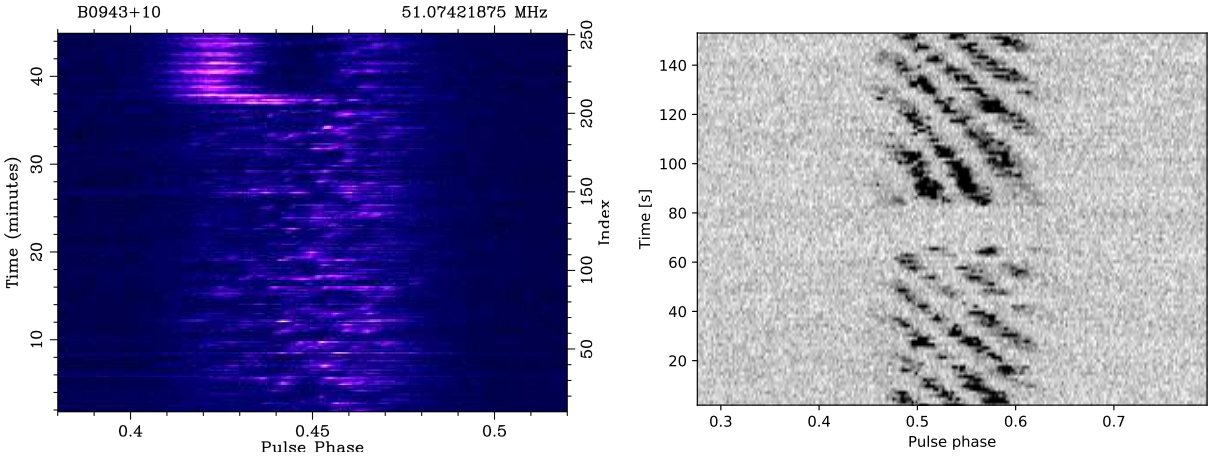
A specific type of radio source recently discovered, called *Rotating Radio Transients* (RRAT) (McLaughlin et al. 2006), reveals a similar behavior and can be seen as an extreme nulling. They are actually a type of pulsar that doesn't emit almost all the time, showing therefore a null state close to 100% of the time. In the case of nulling, the pulsar alternates between phases where the emission is continuous as a normal pulsar and phases where there is no emission. In the case of RRAT, the pulsar only emits a single pulse between period without emission from a few minutes to a few hours.

A last type of modification of the emission of a pulsar is given by the *mode changing*, where the pulsar switches between two or more emission modes. B1237+25 was the first pulsar where this behavior has been seen (Backer 1970a), resulting in the modification of the profile between the two modes. Figure 1.9 shows in the left panel an example of mode changing where B0943+10 switches between two different modes of emission during the observation.

In addition to these changes in the emission of the pulsar, it is equally possible to see a little variability in the period of the pulsar. For some pulsars, the position of the pulse relative to the rotational phase slightly changes from one rotation to another, as shown in the observation of B0943+10 on the left panel of Figure 1.9, exhibiting a cyclic drift of the pulses in time. This effect, called *drifting sub-pulse*, is illustrated by the scheme in the right panel of Figure 1.9. It can be explained using a rotating carousel of sub-beams inside the nested hollow cone (Ruderman & Sutherland 1975b).

These different variations in the time of the emission or the period, and then of the profile, result in difficulties in the observation of the pulsars. This is especially significant in the context of a survey where new pulsars are searched relying on their constancy in time on the timescales of an observation ($<$ 1 hour).

Figure 1.9: Left: observation of B0943+10 with NenuFAR showing a mode changing at the 35^{th} minute until the end of the observation. Right: drifting sub-pulse of B0809+74 on 4 min of observation, with a short nulling in the middle of the observation (Bondonneau et al. 2021).



1.3 Interactions with the ISM

Pulsars have a broadband radio spectrum and can be observed over a more or less large frequency bandwidth. In order to increase the SNR of the profile, it is then possible to integrate the observation in frequency.

However, the interstellar medium (ISM hereafter) is not empty but is filled with cold ionized plasma. During the propagation of the electromagnetic wave of the pulse through a plasma, there occur some diffusive effects dependent on the frequency of the incident wave. These diffusive effects generate differences between the pulse emitted in the magnetosphere of the pulsar and the pulse measured on Earth.

1.3.1 Dispersion

For the propagation of the pulse in the ISM, the plasma can usually be considered as sufficiently cold to neglect the thermal effects. Moreover, the magnetic fields are weak and can also be ignored in the context of the dispersion of the wave. At that time, using the Navier-Stokes and Euler equations in the adiabatic approximation, different propagation modes can be derived.

For the transverse propagation mode, the dispersion relation of the wave through this plasma relative to the refractive index μ and the frequency ν is given by:

$$\mu^2 = 1 - \frac{n_e e^2}{\epsilon_0 m_e \nu^2} = 1 - \frac{\nu_p^2}{\nu^2} \tag{1.14}$$

Where ν_p is the *plasma frequency*, which is a function of the electron density of the plasma n_e, the electron charge e, the dielectric permittivity ϵ_0, and the electron mass m_e. The plasma frequency corresponds to the cut-off frequency of the plasma, such as a wave with a frequency below ν_p cannot propagate.

The refractive index is related to the group velocity v_g of the wave associated with the pulse. The time of propagation of the pulse t from the pulsar to the Earth (at a distance D) can be therefore deduced such as:

$$t = \int_0^D \frac{dl}{v_g(l)} = \int_0^D \frac{\mu(l)}{c} \cdot dl = \frac{1}{c} \cdot \int_0^D \sqrt{1 - \frac{\nu_p^2(l)}{\nu^2}} \cdot dl \tag{1.15}$$

The time of propagation is thus a function of the frequency ν and of the electron content along the line of sight l (with $\nu_p(l) \propto n_e^{1/2}(l)$). Indeed, since the discoveries of the first pulsars, Hewish et al. (1968) have noticed the wave of the pulse should interact with the ISM and consequently yields a dispersive effect on the received light occurring a delay of the low frequencies relative to the higher frequencies.

Figure 1.10 shows two phase-frequency planes in the frequency range 40 – 65 MHz for an observation of B1133+16 with NenuFAR. The left panel presents the dispersion of the pulse where the retardation exceeds many rotational cycles. The pulse integrated in frequency is then totally spread on the entire phase and consequently cannot be seen. The right panel shows the same observation but corrected for the dispersion, where the pulse is correctly aligned in time, presenting a clear double-peaked profile.

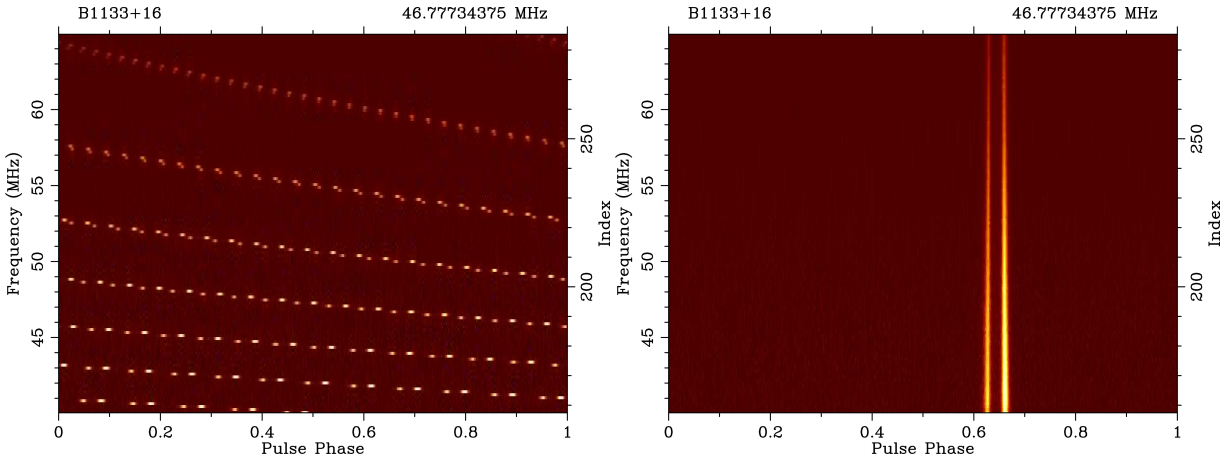


Figure 1.10: Left: dispersion of the pulse of B1133+16 in the frequency bandwidth 40 – 65 MHz obtained with NenuFAR. Right: the same pulse corrected of the dispersion, with a DM of 4.84066 pc.cm^{-3}.

In the ISM, the mean electron density is about 0.03 cm^{-3}, equivalent to a plasma frequency of about 1.5 kHz. Also, the density of the Earth's ionosphere is many orders larger, resulting in a plasma frequency of about 10 MHz. Because of the small length of the ionosphere compared to the interstellar distance, the dispersion caused by the ionosphere is negligible. However, its plasma frequency represents actually a cut-off frequency, preventing wave propagation below 10 MHz. The observation frequencies being largely greater than the plasma frequency of the ISM, and an asymptotic expansion to infinity is possible to simplify the relation. The time delay between two frequencies ν_1 and ν_2 is then given by:

$$\Delta t(\nu_1, \nu_2) \;=\; \frac{e^2}{2\epsilon_0 m_e c} \cdot \left(\frac{1}{\nu_1^2} - \frac{1}{\nu_2^2}\right) \cdot \int_0^D n_e(l) \cdot dl \tag{1.16}$$

The time delay produced by the dispersion can be written using the *dispersion constant* $\mathcal{D}$, and a parameter depending on the line of sight (and then different for each pulsar): the *dispersion measure* (DM).

$$\Delta t(\nu_1, \nu_2) \;=\; \mathcal{D} \cdot DM \cdot \left(\frac{1}{\nu_1^2} - \frac{1}{\nu_2^2}\right) \tag{1.17}$$

Where the dispersion constant is a function of the fundamental constants: $\mathcal{D} = \frac{e^2}{2\epsilon_0 m_e c} \approx (4.148808 \pm 0.000003).10^3$ MHz^2.s.pc^{-1}.cm^3 (Kulkarni 2020). The dispersion measure is defined by the column density on the line of sight:

$$DM \;=\; \int_0^D n_e(l) \cdot dl \tag{1.18}$$

Hence, the DM is a parameter that is different for each pulsar, requiring it to be taken into account in the same way as the period or the period derivative. Using a model of galactic electron density, such as NE2001 (Cordes & Lazio 2002), the DM can also be used to estimate of the distance of the pulsar.

1.3.2 Scattering

For the dispersion of the wave, the propagation of the wave can be considered as following a straight path through the ISM. However, the ISM is not a constant and homogeneous medium but is instead a turbulent medium. The turbulence of the medium produces irregularities in the plasma traversed by the wave, introducing to variations in the refractive index of the medium.

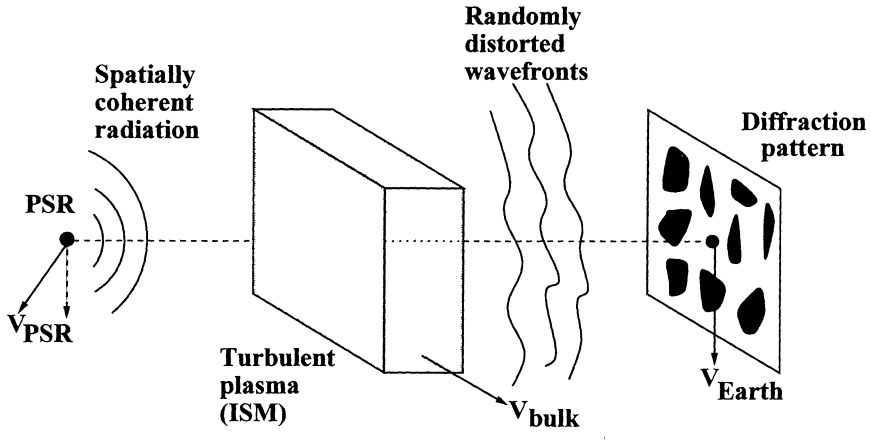


Figure 1.11: Model of a thin screen composed of a turbulent medium, explaining the scattering effect of the pulse. (Cordes 2002)

Scheuer (1968) has proposed a model of a thin screen, featuring a different electron density, midway of the line of sight (illustrated by Figure 1.11). The incident wavefront passes through this inhomogeneous screen of depth a, changing the refractive index μ, leading to a phase shifting of the wave $\Delta\phi$. At a distance D from Earth, the standard deviation of the phase variations, passing in a number of irregularities D/a, is then given by:

$$\Delta\phi \;=\; \sqrt{\frac{D}{a}} \cdot \frac{2\pi\nu \cdot \Delta\mu}{c} \cdot a \;\approx\; \frac{2\pi e^2}{\epsilon_0 m_e c} \cdot \frac{\Delta n_e \cdot \sqrt{a \cdot D}}{\nu} \tag{1.19}$$

These phase variations can be seen as the diffraction of the wave arriving on Earth in a scattering disk of angular size θ_d:

$$\theta_d \;=\; \frac{1}{2} \cdot \frac{c}{2\pi\nu} \cdot \frac{\Delta\phi}{a} \;=\; \frac{e^2}{2\epsilon_0 m_e} \cdot \sqrt{\frac{D}{a}} \cdot \frac{\Delta n_e}{\nu^2} \tag{1.20}$$

The received intensity corresponds to a diffraction pattern caused by the multipath propagation of the rays, which arrive with a deflected angle θ. The angular distribution of the intensity can thus be modeled by a Gaussian term $\exp(-\theta^2/\theta_d^2)$, which can also be interpreted as a geometric time delay Δt:

$$\theta \;=\; \frac{c\Delta t}{D} \;\Rightarrow\; I(t) \;=\; I_0 \cdot \exp\left(-\frac{c}{\theta_d^2 D} \cdot \Delta t\right) \;=\; I_0 \cdot e^{-\frac{\Delta t}{\tau_s}} \tag{1.21}$$

The measured intensity relative to the time is then a pulse extended with an exponential tail defined by the *scattering timescale* τ_s. One of the best examples of the scattering effect can be seen in the profile of B2217+47, presented in Figure 1.12 with observation at 56 MHz with NenuFAR, showing an important scattering tail extending the pulse by almost one rotational cycle.

Using the two previous relations (1.20) and (1.21), we can see that the characteristic time of the scattering shows a strong dependence on the frequency of observation: $\tau_s \;\propto\; \theta_d^2 \;\propto\; \nu^{-4}$. Moreover, τ_s is also related to the distance of the pulsar: $\tau_s \;\propto\; \theta_d^2 \cdot D \;\propto\; D^2$, and as a consequence, the scattering of a pulsar is related to its DM.

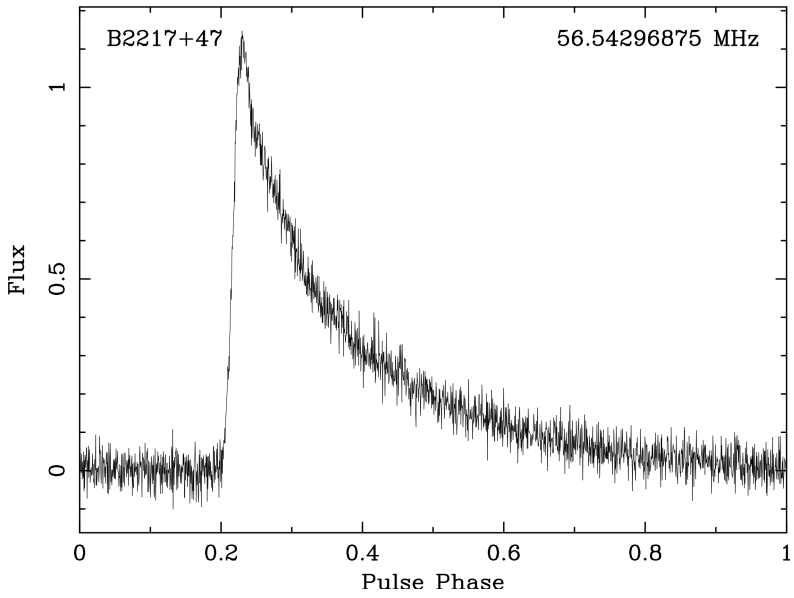


Figure 1.12: Profile of B2217+47 observed with NenuFAR for a central frequency of 56 MHz showing a clear exponential scattering tail.

Bhat et al. (2004) used observations of 81 pulsars observed between 430 MHz and 2.38 GHz to establish an empirical relation of the scattering timescale relative to the DM and the frequency. The result of the fit showed deviations from the theoretical frequency dependence in ν^{-4}, with the following relation:

$$\log \tau_s = -6.46 + 0.154 \log(DM) + 1.07(\log(DM))^2 - 3.86 \log\left(\frac{\nu}{\nu_0}\right) \tag{1.22}$$

Here, ν_0 is the reference frequency of 1 GHz. In the simple case of a midway single thin screen, the index of the frequency dependency is -4. However, in the case of multiple screens located at different positions in the line of sight, the distribution of these irregularities has a spectrum following a Kolmogorov power law (Rickett 1990), resulting in an index of the frequency dependence of -4.4. The obtained value by Bhat et al. (2004) of -3.86 indicates a smaller impact of the frequency on the scattering timescale. However, the fitting result was based on measured scattering timescales showing a large dispersion of the values.

At low frequencies, much lower than 1 GHz, the scattering timescale might be important even for low DM. In addition, this parabolic relation reveals a strong dependence on the DM, resulting in a rapid increase of the scattering timescale relative to the DM.

1.3.3 Scintillation

A turbulent thin screen on the line of sight modifies the phase of the wave. In addition, as shown in Figure 1.11, the turbulent plasma screen distorts the incident wavefront, producing interferences between different rays. Due to the proper motion of the Earth relative to the pulsar and the scattering screen, patches appear in a frequency-time diagram, called *dynamic spectrum*, as a modulation of the measured intensity. This scintillation can be characterized by a spatial scale s_F, representing the size of the first Fresnel zone, i.e. the size of the screen producing the interferences.

$$s_F = \sqrt{\frac{c \cdot D}{2\pi\nu}} \tag{1.23}$$

Here, D is the distance between the screen and Earth, c is the speed of light, and ν is the observing frequency. Adopting the diffractive model for the scattering effect, it can be defined a *coherence scale* s_0, allowing the identification of two types of scintillation by comparison of these two spatial scales.

$$s_0 = \frac{c}{2\pi\nu \cdot \theta_d} = \frac{D}{2\pi\nu \cdot \tau_s} \tag{1.24}$$

With θ_d the angular size of the scattering disk, and τ_s the scattering timescale. In the case where $s_F < s_0$, the scintillation is in the weak regime, corresponds to Fresnel diffraction, and is called *weak scintillation*. This type of scintillation is generated by a small area, resulting in small intensity variations during short timescales. The timescale of the weak scintillation is determined by the movement of the screen, causing the scintillation, relative to the line of sight. Using the velocity of the interstellar scattering screen V_{ISS}, the timescale Δt_w can be estimated by the following relation:

$$\Delta t_w = \frac{s_F}{V_{ISS}} \propto \sqrt{\frac{D}{\nu}} \tag{1.25}$$

According to this relation, weak scintillation occurs at frequencies greater than a GHz and especially for close pulsars. Because of the small size of the Fresnel zone, the timescale of the weak scintillation is rather short. According to Rickett (1990), it results in variations of the order of magnitude of the minute for frequencies above the GHz.

In the case of $s_F > s_0$, the scintillation is in the strong regime. In this context, the intensity variations are stronger than in the weak regime and appear in longer timescales. The first type of *strong scintillation*, corresponding to Fraunhofer diffraction, is the *diffractive scintillation*. Due to the constraints in terms of frequency and distance of the pulsars, diffractive scintillation is more common than weak scintillation.

The area causing the scintillation being larger than the coherence scale, the timescale of the diffractive scintillation is determined by the velocity of the interstellar scattering screen relative to the coherence scale s_0:

$$\Delta t_s = \frac{s_0}{V_{ISS}} \propto \frac{\nu^{1.2}}{D^{0.6}} \tag{1.26}$$

According to Gupta et al. (1994), the timescale of the diffractive scintillation is of the order of magnitude of hours at frequencies greater than 500 MHz and decreases in the order of magnitude of the minute for the low frequencies of NenuFAR. As a consequence, the diffractive scintillation creates patches that can be longer than the duration of an observation, resulting in important variations of the measured intensity from one observation to another.

The diffractive scintillation is caused by the movement of the region corresponding to the coherence scale relative to the line of sight. The last type of scintillation is caused by the movement of the scattering disk relative to the line of sight. The size of the scattering disk is given by $D \cdot \theta_d$ and consequently corresponds to longer timescales than for the diffractive scintillation. This type of scintillation indeed corresponds to the movement on large scale. It results in important variations of the refractive index, generating therefore a focusing effect (Rickett 1990). This so-called *refractive scintillation* has a timescale of the order of magnitude of days and can be seen as long-term variations (in some months) of the measured intensity.

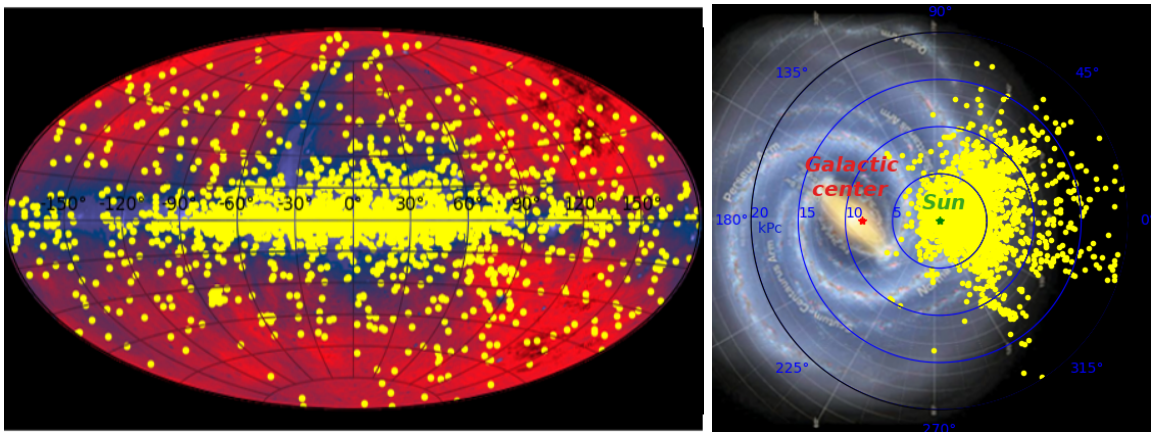
1.4 Population of pulsars

1.4.1 Galactic distribution

Since the first discoveries in 1968, a large number of pulsars have been added to the list to reach 3,341 pulsars in version 1.68 of the ATNF (Manchester et al. 2005) dating from the 5^{th} of November 2022[2]. Looking at the distribution of the known pulsars in the galactic frame, it appears that the majority of the pulsars are located in the galactic plane, and furthermore in the longitude of the center of the galaxy. This galactic distribution is indeed close to the distribution of the regions of formation of massive star (Urquhart et al. 2014), and is thus consistent with the idea of a neutron star as the cause for the pulsar phenomenon (see Section 1.1.1).

[2] http://www.atnf.csiro.au/research/pulsar/psrcat

Figure 1.13: Left: distribution of the known pulsars listed in the ATNF (Manchester et al. 2005) as yellow dots (sky in radio in the background (MPIfR, based on data of (Haslam et al. 1995))). Right: stereographic projection of the known pulsars in the ATNF in yellow dots (representation in the background of the Milky Way (© NASA)). The Sun is indicated by the green star and the galactic center by the red star at 8.5 kpc. The blue circular lines represent the distances from the Sun in kpc.



However, the distribution in the galactic latitude of the pulsars is more dispersed than for the massive stars. An explanation of this spreading out of the galactic plane is that the neutron star undergoes a kick during its birth, allowing it to obtain a high birth velocity of 450 ± 90 km.s^{-1} on average (Lyne & Lorimer 1994). The escape velocity of the galaxy being of about 533^{+54}_{-41} km.s^{-1} (Piffl et al. 2014), for a part of the stars, that can be sufficient to extract it of the galactic gravitational potential.

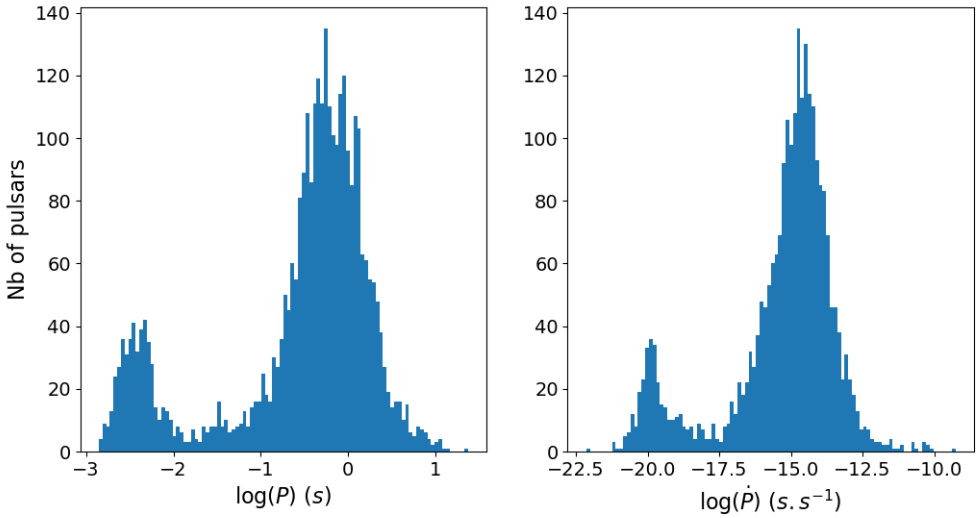
Moreover, it should be noticed that positions of the pulsars projected on the galactic plane (in the right panel of Figure 1.13) show that the majority of them are located around the Sun. This distribution around the Solar System is due, on the one hand to the decreasing flux with the square of the distance, and on the other hand to the ionized interstellar medium between the pulsars and the Earth. Consequently, these two effects make it more difficult to observe more distant pulsars. Finally, combined with the instrumental limits and the geometric constraints on the sky covered by the beam, we can definitely think the current population of the known pulsar is just a little fraction of the total number of pulsars in the galaxy. Faucher-Giguère & Kaspi (2006) estimates that the total number of pulsars in the galaxy is 1 200 000 ± 300 000, with 120 000 ± 20 000 pulsars potentially observable.

1.4.2 Classification and evolution

As mentioned in Section 1.1.2, the pulsars can be characterized by two major values related to the spin-down model, and describing the neutron star properties: the period and the period derivative. Figure 1.14 presents the distribution of the 3 320 known pulsars of version 1.68 of the ATNF catalog (when the value of the parameter is known) for these two parameters. The first remark is that the pulsars are not randomly distributed, and have globally relatively close values for each parameter. The second remark is that, for the two characteristics, the distribution is bimodal, with a major distribution on the right of each plot of Figure 1.14, and a minor distribution on the left.

To go further that the histograms of Figure 1.14, it is possible to combine the different parameters in a two-dimensional representation called the $P - \dot{P}$ *diagram*, where the x-axis corresponds to the logarithm of the period and the y-axis to the period derivative. With this representation given in Figure 1.15, the two groups of pulsars are clearly visible. Also, the two distributions of the previous histograms correspond indeed to two different populations of pulsars. The pulsars of the majority population, localized in the center, are called the *normal pulsars*, and those of the minor one, localized in the bottom left part, the *millisecond pulsars* (MSP) due to their shorter period of the order of a millisecond.

Figure 1.14: Histograms of the period (left) and period derivative (right) of all the pulsars listed in the ATNF catalog (Manchester et al. 2005) (logarithm of the two properties).



The surface magnetic field is a property of the neutron star which can be derived in the case of an ideal dipole by Equation (1.7) (see Section 1.1.2). The surface magnetic field of the pulsar is then proportional to $\sqrt{P \cdot \dot{P}}$, and corresponds on the $P - \dot{P}$ diagram to an oblique line. These lines separate the two populations in the same way as the two first parameters, meaning that MSP having lower surface magnetic fields than normal pulsars.

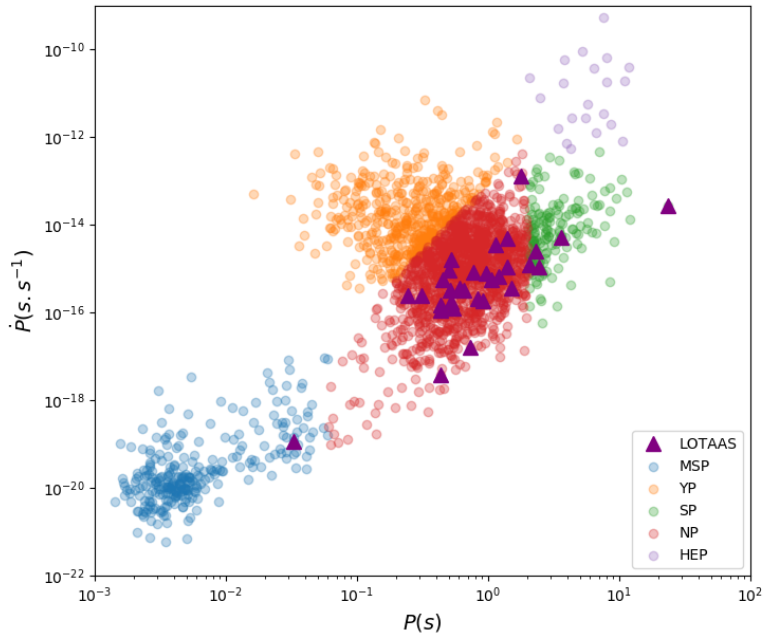
Finally, the $P - \dot{P}$ diagram allows seeing the evolution of the pulsars during their life. As for the magnetic field, the age of a pulsar is defined relative to the period and the period derivative (see Equation (1.6) in Section 1.1.2), and is proportional to $\sqrt{P/\dot{P}}$. The color bar of Figure 1.15 refers to the age of the corresponding pulsar, showing that the young pulsars are localized in the upper part of the diagram. Basically, pulsars are born with a relatively short period around some dozens of ms, a strong spin-down, and a high surface magnetic field.

A few exceptions exist with the objects in the right top corner. These are high-energy pulsars essentially detected in X-rays and gamma rays, with an ultrahigh magnetic field and a very slow period of rotation. These are generally undetected in radio and constitute mysterious types of objects such as magnetars.

Due to the strong spin-down, pulsars slow down some hundreds of millions of years. Then, as indicated by the gradient of color, they move toward the center of the graph. Mature pulsars globally feature a relatively slow period of around one second, with a weaker spin-down, and also a weaker surface magnetic field. This slowing down also means that the star emits less energy, and a limit seems to exist below which the emitted energy is insufficient to produce the pulsar phenomenon. The emitted energy being proportional to $\dot{P}/P^{-3}$, the limit of the so-called "pulsars" graveyard is a steep oblique line, with no pulsars present in the bottom right part of the $P - \dot{P}$ diagram (the grey area on Figure 1.15).

However, the second population of pulsars, i.e. the MSP population, is an exception to this trend, since it is, according to the dipolar spin-down model, constituted by the oldest known pulsars. Although they are older than the other pulsars, they continue to sufficiently emit to be detectable, drifting radically out of the main population towards short periods. However, the age determined using the standard spin-down model give, for many MSPs, an age greater than the age of the Universe. That means MSP must follow a different evolution that the normal pulsars.

Figure 1.15: $P - \dot{P}$ diagram showing all the pulsars with known P and $\dot{P}$, in the catalog (version 1.68) of the ATNF (Manchester et al. 2005). The colors correspond to the age of the pulsars (based on the dipolar spin-down model). The oblique grey dashed lines show some levels of the surface magnetic field. The grey bottom right corner represents the so-called pulsar graveyard, where the emitted energy is theoretically insufficient to be seen.



1.4.3 Properties of millisecond pulsars

The pulsars of this second population are thus drastically different from the normal pulsars in many ways as the period or the magnetic field. These specificities and the different evolution in the $P - \dot{P}$ diagram can be explained by a different origin than the normal pulsars. Although the pulsars are generally isolated, essentially due to the birth kick given by the original supernova, a minor part (about 10%) is in binary systems. Also, it appears that more than 90% of these binary pulsars are MSP, constituting the vast majority of this population (about 84%).

This particularity allows us to explain the different evolution tracks of these old pulsars which are MSP, starting from a standard binary system of two main-sequence stars (Bhattacharya & van den Heuvel 1991). The more massive star explodes in a supernova leaving a neutron star as a remnant. The majority of the systems are actually disrupted at this stage, but a few binary systems can remain intact. Due to the strong gravitational field of the neutron star, the matter of the companion is accreted. The rotation of the neutron star is thereby accelerated until it has reached the equilibrium between the magnetic pressure and the gravitational pressure. It results in a reduction of the period, and then a decrease of the magnetic field (see Equation 1.7).

In the case of a low-mass second star, the ejection of the outer layers during the end of its life, in the red giant stage, allows the accretion of matter for some hundreds of Myr, leading to finally a binary system composed of a pulsar and a white dwarf. If the companion star is in the opposite case, a massive star, the accretion is thus quicker, for many dozens of Myr until the explosion in a supernova of the companion. If the system is not disrupted, it results in a double neutron star binary system, with two completely different neutron stars as the double pulsar system J0737-3039 discovered by Burgay et al. (2003). In the other situation, the binary system is separated, leaving an isolated MSP.

This evolution as a binary system rather than as an isolated object brings some differences in their

evolution relative to the standard model used for the normal pulsars. The first obvious deviation concerns the characteristic age of the MSP, which is often much greater than the real age. According to Equation (1.6) from the spin-down theory for an ideal dipole, the obtained age is inevitably important. For some of them, it results in impossible ages, with 35 pulsars older than the Universe, such as the oldest J1801-3210 which obtains a value of 1 640 Gyr.

Another point of interest is that the light cylinder of the MSP is smaller than for the normal pulsars. The first consequence is that the polar cap is larger, leading to an emission that can be far from the magnetic axis. The simple beam model presented in Section 1.1.4, is then no longer valid to explain the observed widths and opening angles for the MSPs. The second consequence is that all the emission is produced in a small region close to the surface of the neutron star. Observations of MSPs show that the profile is quasi-independent of the frequency, leading to the fact that the RFM model can't, therefore, be applied to these pulsars.

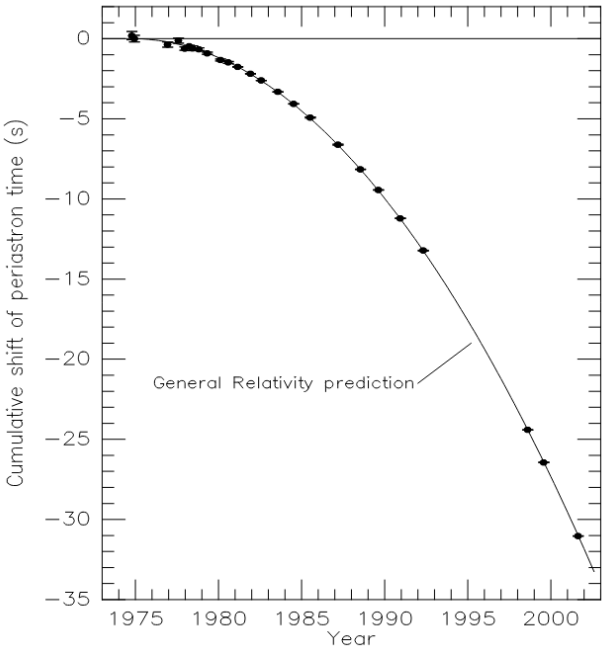


Figure 1.16: Comparison of the measured shift of the periastron of B1913+16 (black dots) with the predictions of the general relativity (Weisberg & Taylor 2003).

Although the MSPs are difficult to model, they are, like every pulsar, basically a neutron star, i.e. a compact and massive star. Also, the binary nature of the large majority of the MSP can be used to test the theories of gravity by measuring the relativistic additional terms of the Keplerian laws of the system's dynamic, called *post-Keplerian terms*. The double neutron star systems are especially useful due to the extremely strong gravitational field generated by two neutron stars. This kind of system present generally a large eccentricity which can be used, e.g. to measure the advance of the periastron predicted by the general relativity (Weisberg & Taylor 2003), as shown in Figure 1.16.

The other type of binary system with two compact stars, the neutron star - white dwarf system, also constitutes a useful laboratory to study the strong equivalence principle (SEP). A violation of the SEP induces that two objects with different masses fall with different velocities in an external gravitational field. The significant difference between the masses of the neutron star and the white dwarf allows us to test the SEP in the strong-field regime (see for example the study on the triple system with the pulsar J0337+1715 and two white dwarfs Voisin et al. 2020).

The extreme stability of the pulsars, cumulated with the extreme precision of the MSP due to their short period, allows also us to detect gravitational waves. Using a *pulsar timing array* (Stappers et al. 2006; Jenet et al. 2009), which is an array of well-known MSPs, a gravitational wave will produce a deviation which can be seen by comparison with the other pulsars of the array. Contrary to such interferometers as LIGO or VIRGO, these pulsar timing arrays are sensitive to low-frequency gravitational waves, allowing them to detect complementary signals unseen by these interferometers.

Résumé du chapitre :

En 1967, dans le cadre d'un programme d'observation pour étudier la structure angulaire de sources radios compactes, J. Bell identifie un signal faible avec une période très précise de 1,337 s (Hewish et al. 1968). Dans les mois qui suivirent, trois autres sources similaires ont été rapidement trouvées (Pilkington et al. 1968).

En 1932, Lev Landau propose un type d'étoiles exclusivement composé de neutrons (Landau 1932). Deux ans après, Baade & Zwicky (1934) propose que ces étoiles à neutrons soient reliées au restes de supernovae. Au moment de la découverte des pulsars, il existe alors deux théories pouvant expliquer l'observation de ces signaux radio périodiques. Meltzer & Thorne (1966) calcule que les oscillations radiales d'une naine blanche pourraient générer des signaux avec des périodicités de plusieurs secondes. En parallèle, Pacini (1967, 1968) et Gold (1968) proposent comme candidat pour expliquer les pulsars, que cela soit plutôt des étoiles à neutrons ayant des vitesses de rotation très importantes et des champs magnétiques forts, générant des signaux de périodicité plus courtes et montrant un ralentissement dans le temps. La découverte d'un pulsar de 33 ms de période à l'intérieur de la nébuleuse du Crabe par Staelin & Reifenstein (1968), avec de plus la mesure d'un ralentissement de cette période par Richards & Comella (1969), permet de définitivement exclure la naine blanche comme objet de base pour les pulsars.

Dans le modèle actuel, les pulsars sont des étoiles à neutrons d'environ 1,4 $M_{\odot}$, et ayant un rayon entre 11,8 et 13,1 km (Miller et al. 2019). Ces étoiles sont des astres extrêmement compactes avec des densités allant de 10^9 kg.m^{-3} en surface jusqu'à 10^{17} kg.m^{-3} au niveau du cœur (Shapiro & Teukolsky 1983). En outre, ces étoiles à neutrons disposent d'un puissant champ magnétique allant de 10^8 G 0 10^{15} G, dont les lignes de champ résultantes sont capables d'extraire des particules de la surface de l'étoile Goldreich & Julian (1969). Ces particules, sous forme de plasma, suivent les lignes de champ tout en étant en corotation avec l'étoile. Une limite correspondant à une vitesse de rotation égale à la vitesse de la lumière, appelée *cylindre de lumière*, peut alors être déterminée par rapport à la période de rotation de l'étoile à neutrons, permettant par la même occasion de définir deux types de ligne de champ. Le premier type est constitué par les lignes de champ fermées, qui sont celles situées à l'intérieur du cylindre de lumière, et correspondent au plasma en corotation avec l'étoile.

Enfin, le second type correspond aux lignes de champ ouvertes. Celles-ci ont pour origine une petite région de l'étoile autour de l'un des deux pôles magnétiques. Les particules situées dans ces régions sont accélérées suivant la ligne de champ, produisant ainsi un rayonnement radio tangentiel à cette dernière. Dans la très grande majorité des cas, les axes de rotation et magnétique ne sont pas alignés, induisant en conséquence un effet phare du faisceau radio produit lorsqu'il croise la ligne de visée avec la Terre. Le faisceau est ainsi observé avec une certaine largeur qui est fonction de la géométrie de la magnétosphère du pulsar. En outre, (Cordes 1978) a proposé un modèle, dénommé *radius to frequency mapping*, expliquant les variations de largeur du faisceau observées en fonction de la fréquence d'observation, par le fait que l'altitude d'émission soit également dépendante de la fréquence d'émission.

Dû à l'effet phare, les pulsars sont vus sur Terre comme une série d'impulsions individuelles périodiques. A l'exception des pulsars les plus intenses, les impulsions individuelles ont des flux faibles. Pour faciliter la détection des pulsars, une méthode appelée *folding* utilise la périodicité de ces impulsions en découpant la série temporelle en blocs d'une durée égale à la période du pulsar, pour ensuite empiler ces blocs ensemble. Il en résulte un profil d'impulsion intégré. Ce profil est constant dans le temps, et est unique pour chaque pulsar. Il existe de nombreuses formes de profil, allant d'un profil avec un unique pic, jusqu'à des profils montrant plusieurs pics plus ou moins proches. Les différents pics présents dans les profils sont caractéristiques de la géométrie de l'émission à l'intérieur du cône d'émission du pulsar. Ainsi, la théorie des cônes creux imbriqués (Rankin 1993) permet d'expliquer des profils montrant plusieurs pics différents au sein du profil.

Bien qu'étant constant dans le temps, le profil d'un pulsar n'est pas constant en fréquence, avec des composantes qui peuvent varier en forme et/ou amplitude relative par rapport aux autres composantes du profil. Aussi, en terme de flux, les pulsars montrent un spectre qui peut être modélisé par une loi puissance simple d'indice spectral négatif autour de -1,6 $\pm$ 0,03 (Jankowski et al. 2018) pour les fréquences supérieures à environ 100 MHz. Pour les fréquences plus basses, la majorité des pulsars montrent un retournement du spectre qui peut être modélisé, pour la plupart, par une seconde loi puissance simple d'indice spectral positif entre 0,1 $\pm$ 0,3 et 4,8 $\pm$ 1,4 (Bilous et al. 2020).

Pour arriver sur Terre, l'onde radio émise par le pulsar traverse le milieu interstellaire. Le milieu ionisé produit ainsi un effet de dispersion de l'onde dépendant de la fréquence, induisant un décalage temporel entre deux fréquences différentes. Cette dispersion est proportionnelle à la quantité d'électrons sur la ligne de visée, et est ainsi également plus ou moins relié à la distance du pulsar.

Dans le cas où il y a un écran de matière ionisée sur la ligne de visée, l'impulsion individuelle du pulsar subit un second effet nommé *scattering*. Cet effet est produit par la diffraction de l'onde à divers endroits de l'écran autour de la ligne de visée, générant une propagation multi-chemins de l'onde radio. Le résultat est que certaine parties de l'impulsion finalement mesurée arrive avec un certain retard, qui se traduit par un élargissement de l'impulsion suivant une forme exponentielle (Scheuer 1968).

Enfin, un dernier effet est la scintillation, qui est liée au mouvement de la ligne de visée par rapport aux différents écrans de scattering. Cet effet de scintillation cause une variation temporelle et fréquentielle du flux du pulsar, pouvant augmenter ou diminuer le flux mesuré. Aux fréquences inférieures au GHz, deux types de scintillations se voient essentiellement : la *scintillation diffractive* qui apparaît sur des échelles de l'ordre de la minute à 50 MHz, et la *scintillation réfractive* qui est relative à de plus grandes structures spatiales, et qui donc se mesure sur de plus longues échelles de temps allant de quelques jours à plusieurs mois.

Tous les pulsars actuellement connus sont situés dans la Voie Lactée. Étant issus d'explosions en supernova, leur répartition au sein de la galaxie est liée à celle de la population d'étoiles massives. En conséquence, la majorité des pulsars sont localisés proches du disque galactique. De plus, en raison des problèmes liés au milieu interstellaire, ils sont également proches du système Solaire.

Généralement, les pulsars sont caractérisés en fonction de leur période P et de leur ralentissement $\dot{P}$, et représentés au sein d'un *diagramme* $P - \dot{P}$. A cause de leur ralentissement, au cours de leur vie, les pulsars migrent au sein de ce diagramme suivant une diagonale, où les pulsars jeunes sont situés dans le coin en haut à gauche, puis dérivent vers le bas et vers la droite jusqu'à une ligne limite où plus aucun pulsar n'est détecté.

En outre, dans ce diagramme, il apparaît finalement deux populations distinctes : un groupe principal de pulsars dits normaux, et un second groupe de *pulsars millisecondes* avec des périodes très courtes entre 1 et 30 ms. Ces derniers sont en fait des pulsars âgés présents dans des systèmes binaires, et ayant subi une ré-accélération par l'accrétion de la matière de leur compagnon.

Chapter 2

NenuFAR

2.1 Overview

NenuFAR, for *New Extension in Nançay Upgrading LOFAR*, is a new radio-telescope located in France between Orléans and Vierzon at the site of the Nançay Radio Observatory (Zarka et al. in prep.). It has been inaugurated in October 2019 and was in the Early Science phase from July 2019 to December 2021.[1]

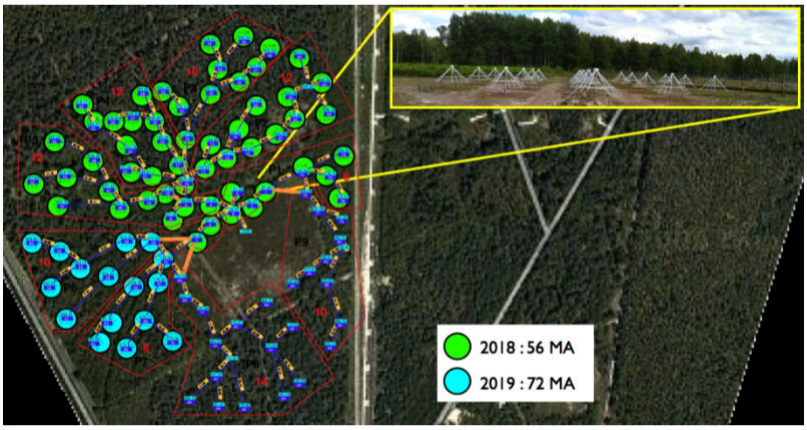


Figure 2.1: Map of the mini-arrays of NenuFAR, with a photo of a mini-array on the top right. Green points represent the mini-arrays built before the beginning of the thesis. Blue points represent the remaining mini-arrays to complete NenuFAR, whose 34 have been built in 2022. (© Observatoire Radioastronomique de Nançay)

NenuFAR is a phased array telescope, which will be composed of 96 mini-arrays (MA hereafter) spread into a disk of about 400 m in diameter. At the beginning of the present thesis, only 56 MA were built in the northern part of the final NenuFAR area in an ellipse of 200 x 400 m (green points in Figure 2.1). Since the middle of 2022, 24 new MAs are built to reach 80 MAs, and the last MAs are funded and planned to be built in 2023.

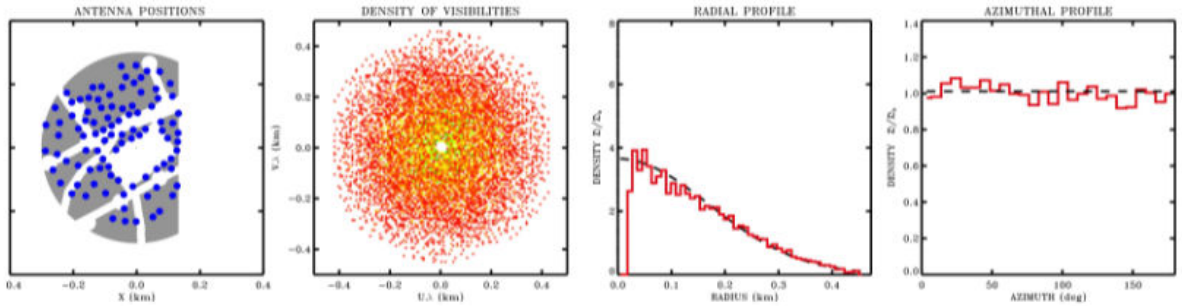
As its name indicates, NenuFAR is planned to be linked with LOFAR ((van Haarlem et al. 2013, LOw Frequency ARray)).

The 96 MA of the complete NenuFAR will be distributed following a Gaussian distribution in the radial direction and with a uniform angular distribution. This MA distribution, illustrated in Figure 2.2, allows us to obtain a Gaussian beam.

In addition to the 96 MA of the NenuFAR core, six distant MA are planned to be built to perform imaging. The distant MA will be placed about 3 km around the core while keeping an optimized coverage of the uv plane (the plane of the spatial frequencies). In 2022, four are currently built.

[1] This chapter is based on the instrument information provided by the NenuFAR webpage: https://nenufar.obs-nancay.fr/en/astronomer/

Figure 2.2: From left to right: distribution of the mini-arrays on the ground, visibilities in the uv plane, radial distribution of the mini-arrays, and angular distribution of the mini-arrays for the complete NenuFAR with 96 mini-arrays. (© Observatoire Radioastronomique de Nançay)



2.2 Mini-arrays

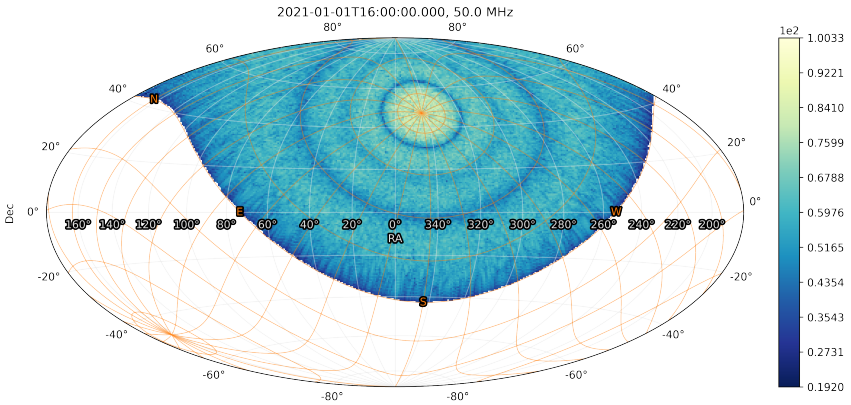
Each MA is composed of 19 antennas spaced by 5.5 m, distributed in a hexagon of 25 m in width. The used antennas, shown in Figure 2.3, are crossed dipoles in an inverted V-shape similar to the antennas of the LWA. Each antenna has an effective collecting area of $c^2/(3\nu^2) \times 9$ m^2, with ν the observing frequency and c the speed of light. All antennas are aligned at 45°on the NW-SE and NE-SW axes, but the MA are rotated at different angles: 0, 10, 20, 30, 40, and 50°.

Figure 2.3: Photo of an antenna of NenuFAR. (© Observatoire Radioastronomique de Nançay)

The 19 antennas of a MA are linked by different physical delay lines. The signals of each antenna are combined by applying a time delay which allows pointing to a specific sky position. The combination of the signals of the antennas generates an analog beam with a full-width-half-maximum of about 8° at the highest frequency of 85 MHz and 46° at the lowest frequency of 15 MHz. Due to the limited number of delay lines, the number of positions is discrete and follows a grid of 16 384 available positions. To track a source, the tracking adjustment of the analog beam occurs each 6 min.

The combined signals of the different MA are then combined using a digital continuous time delay to form the digital beam of NenuFAR. Contrary to the analog beam where the time delay is discrete, this continuous time delay allows tracking the source with a shorter tracking adjustment of the digital beam every 10 s. This digital beam has a width of 0.5° at 85 MHz, and of 2.9° at 15 MHz. An simulation of the digital beam for a pointing at the Zenith for a frequency of 50 MHz using the complete NenuFAR (i.e. with 96 MAs) is presented in Figure 2.4.

Figure 2.4: Model of digital beam for pointing at the zenith for a frequency of 50 MHz (A.Loh). Orange lines represent the frame in altitude-elevation.



2.3 Observing characteristics

Because of the limited number of available delays for the pointing of the analog beam, there are sky positions not able to be observed at low elevations. Moreover, the collecting area is dependent on the elevation of the source. As a consequence, NenuFAR can efficiently observe elevations between 20 and 90°, corresponding to declinations from -23 to 90°.

Antennas of NenuFAR are designed to observe at very low frequencies between the ionospheric cutoff of 10 MHz and the lower edge of the FM radio at 85 MHz. Using a custom-designed low-noise preamplifier, NenuFAR has a relatively flat bandpass in a large fraction of the observable bandwidth. Figure 2.5 shows the superposition of the bandpass of all the MA of NenuFAR, where the bandpass is approximately flat between 20 and 80 MHz.

The signals of the MA are collected and digitized by the receiver *LaNewBa* (LOFAR super station Advanced NEW BAckend). The two signals of each MA, corresponding to the two linear polarizations measured by the antennas, are coherently summed to generate a beamformed signal. The beamformed signals are channelized in 384 subbands of 195.3125 MHz, allowing a time resolution of 5.12 μs.

In addition to the elevation dependency of the effective area, there is also a frequency dependence illustrated in Figure 2.6. It results from this dependency in a variation of the effective area (with the current configuration using 80 MAs) at the zenith from 69 000 m^2 at 15 MHz to 6 300 m^2 at 85 MHz. As a consequence, the minimum flux sensitivity varies from about 130 mJy at 15 MHz to 9 mJy at 85 MHz for observation at the zenith of one hour with 10 MHz of bandwidth.

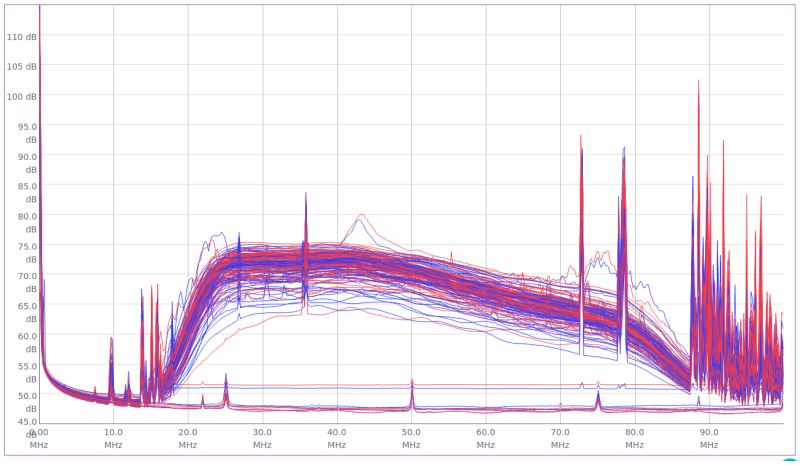
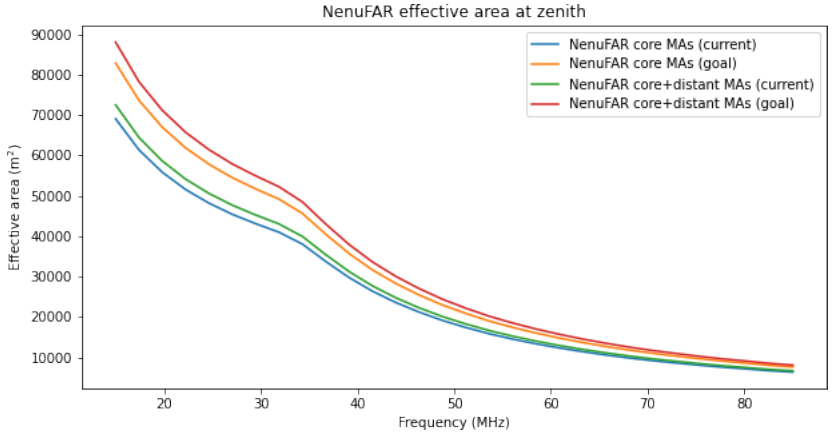


Figure 2.5: Superposition of the bandpass of each mini-array of NenuFAR (the red and blue colors are for the two measured polarizations).

Figure 2.6: Effective area of NenuFAR as a function of the frequency for observation at the zenith. The areas are shown for four configurations depending on the number of mini-arrays: current core = 80, goal core = 96, current core+distant = 80+4, goal core+distant = 96+6. (© Observatoire Radioastronomique de Nançay)



2.4 Sensitivity

From these effective areas, it is possible to determine the gain of NenuFAR using the relation:

$$G = \frac{A_{eff}}{2k_B} \tag{2.1}$$

With k_B the Boltzmann constant of $1.38.10^{-23}$ J.K^{-1} and A_{eff} the effective area of the telescope. For an observation at the zenith, the gain of NenuFAR ranges from 2.28 K.Jy^{-1} at 15 MHz to 25 K.Jy^{-1} at 85 MHz.

The sensitivity of the telescope determines the minimum flux S_{min} that a pulsar must have to be detected. It can be estimated for a minimum SNR SNR_{min} based on this gain G and the radiometer equation (Dewey et al. 1985):

$$S_{min} = \frac{SNR_{min} \cdot T_{sys}}{G \cdot \sqrt{n_p \cdot \Delta t \cdot \Delta f}} \cdot \sqrt{\frac{\delta_{cyc}}{1 - \delta_{cyc}}} \tag{2.2}$$

This equation takes many parameters as input. There are three observing parameters: n_p the number of polarizations, Δt the duration of the observation, and Δf the observation bandwidth. There is also a parameter specific at the observed pulsar δ_{cyc} representing the duty cycle of the pulsar, i.e. the width of the pulse compared to the period.

Finally, there is a last parameter T_{sys} corresponding to the system temperature. At higher frequencies, it is typically the temperature of the electronics of the telescope. However, at the NenuFAR frequencies, the major component is indeed directly the sky whose temperature is considerably greater than the telescope temperature. Also, below 100 MHz, the range of temperature in the sky is substantial. According to the sky map of Haslam et al. (1995) with the scaling index of -2.87 determined by Lawson et al. (1987), a “cold region” (where the background radio emission is low) has a temperature of about 3 000 K, while a “hot region” (towards the galactic center) reaches up to 155 000 K.

As a consequence, the sensitivity of NenuFAR approximately varies of a factor 50 following the observed sky position. Thereby, for an observation in total intensity of one hour using the total bandwidth of 75 MHz, the sensitivity of detection is indicated in Table 2.1. Thus, at 15 MHz towards a “cold region”, the sensitivity of detection of a pulsar with a typical duty cycle of 0.1 is 0.27 mJy, and considerably increases up to 154 mJy in the case of the observation of the galactic center at 85 MHz. Moreover, adding the elevation dependency, these sensitivities may decrease approximately by a factor of three at the maximum.

Table 2.1: Gain and sensitivity of NenuFAR (with 80 mini-arrays) at observation frequencies of 15 and 85 MHz. The sensitivities are calculated for an observation in total intensity of 1 hour over 75 MHz at the zenith, and towards a "cold region" (T_{sky} = 3 000 K) or the galactic center (T_{sky} = 155 000 K).

ν (MHz)	G (K.Jy^{-1})	S_{min} (T_{sky} = 3 000 K) (mJy)	S_{min} (T_{sky} = 155 000 K) (mJy)
15	25	14	154
85	2.82	0.27	3

2.5 Observing mode

The beamformed data produced by LaNewBA are processed by two identical CPU/GPU calculators *UnDySpuTed* (Unified Dynamic Spectrum Pulsar and Time Domain receiver) (Bondonneau et al. 2021). The data are divided into lanes representing data in full polarization with 37.5 MHz of bandwidth beamformed at a specific position. Each node can process in real-time up to two lanes, allowing to perform observations corresponding to four lanes at the maximum. The maximum observing possibilities correspond to three different cases:

- 4 different sky positions observed on half-bandwidth (37.5 MHz);
- 3 different sky positions with one source observed on full bandwidth (75 MHz) and the two others in half bandwidth;
- 2 different sky positions observed on full bandwidth.

Each of the lanes sent to the UnDySpuTed nodes is processed using one of four different observing modes:

- pulsar – folded mode;
- pulsar – single pulse mode;
- dynamic spectra mode;
- waveform mode.

The *pulsar – folded* mode is a specific observing mode for the pulsars. In this mode, the data within the lane are coherently dedispersed for the intra-channel dispersion in real-time at the DM of the pulsar. Data are next incoherently dedispersed for the inter-channel dispersion, and time sub-integrations of 10.737 s are folded using the ephemeride of the pulsar.

The *pulsar – single pulse* mode is the other observing mode specific to the pulsars. As for the previous pulsar mode, the data are coherently dedispersed for the intra-channel dispersion in real-time. However, in this mode, the data are unfolded, and for the inter-channel dispersion, a DM search is performed around the reference DM.

In addition to these two pulsar modes, the *dynamic spectra* mode returns a time-frequency plane without dedispersion or folding. The dynamic spectra mode allows performing additional channelization of the data to increase the frequency resolution between 0.10 and 12.20 kHz, with a time integration between 0.3 and 84 ms depending on the desired frequency resolution.

The last observing mode is the *raw waveform* mode, consisting of directly returning the raw data produced by LaNewBa. This mode allows obtaining data with the initial time resolution of 5.12 μs. However, this observing mode is rarely used because of the large required storage of about 1 TB for one hour of observation.

2.6 Virtual controol room

To schedule the observations, NenuFAR is managed using a graphical user web interface: the VCR for Virtual Control Room. This web interface allows us seeing the planning of observation or the time slots of visibility of a source with NenuFAR. To schedule observations, the VCR permits two ways: a graphical way using different windows to select all the observing parameters, or with the import of configuration files called *parset_user*. Figure 2.7 shows the graphical window of a scheduled observation, allowing us to see information about the analog beam, the different digital beams, and many configuration parameters which can be modified directly in this window.

The VCR features different tools to check information about the telescope, such as the measured bandpass, some dynamic spectra of the observed sky, or the presence of thunderstorms. Figure 2.8 shows the dashboard page of the VCR allowing us to follow in real-time the celestial hour, a map of the observed sky, some statistics about the measured data, and a log of the telescope activity.

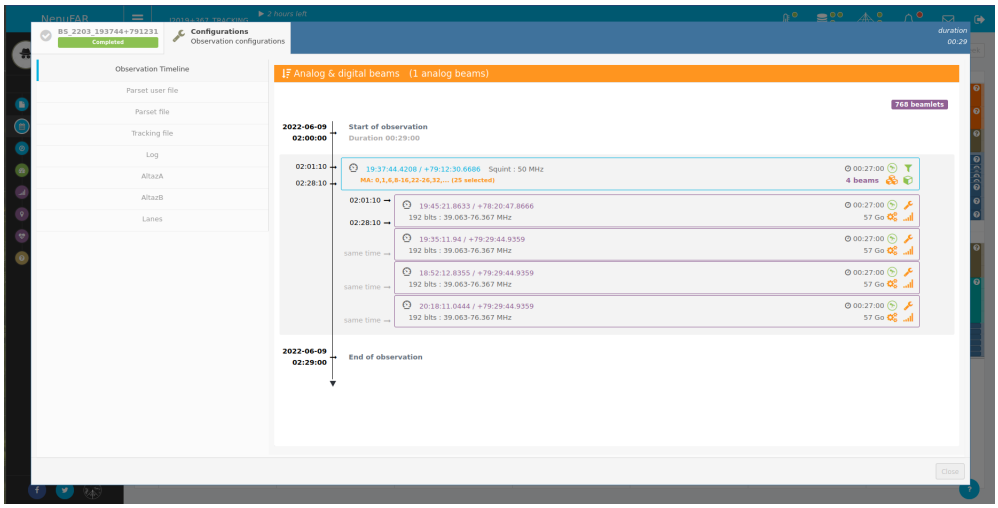


Figure 2.7: Graphical window presenting different information about the analog beam, the digital beams, and different information about the observation.

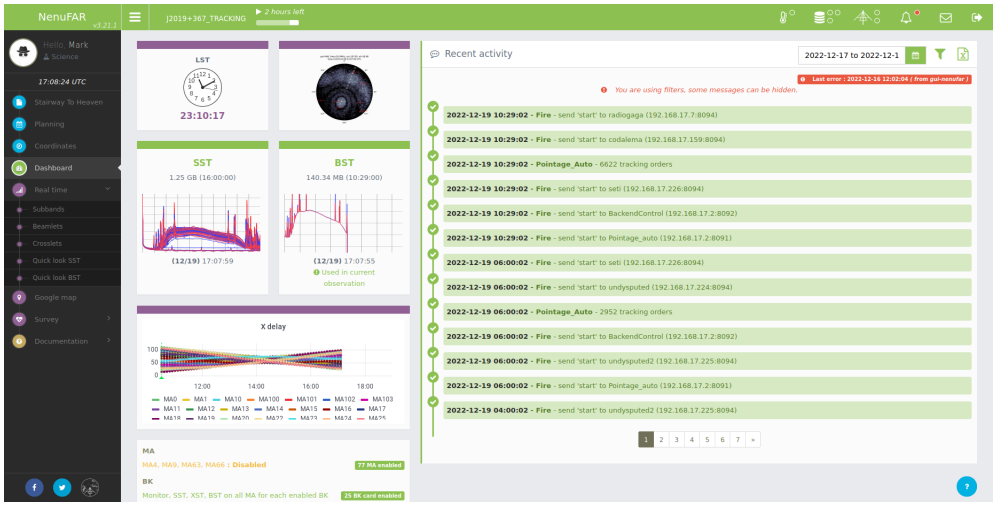


Figure 2.8: Graphical window allowing to control information about the telescope in real-time.

Résumé du chapitre :

NenuFAR est l'acronyme de *New Extension in Nançay Upgrading LOFAR*, et est un nouveau radio-télescope français situé sur le site de l'Observatoire Radioastronomique de Nançay. Il a été inauguré en octobre 2019, et était en phase scientifique initiale entre juillet 2019 et décembre 2021.

NenuFAR est un télescope à réseau phasé composé de 96 mini-réseaux (MR ci-après) de 19 antennes, distribués au sein d'un disque de 400 m de diamètre autour de la station française de LOFAR (LOw Frequency ARray). Les 96 MR sont répartis suivant une distribution radiale gaussienne, et suivant une distribution angulaire uniforme, permettant d'obtenir un faisceau global gaussien. Au début du relevé, seulement 56 MR étaient construits et répartis à l'intérieur d'une ellipse de 200 x 400 m. Courant 2022, 24 nouveaux MR ont été installés, et la construction des 16 MR restants est planifiée pour 2023. S'ajoute à ces premiers MR 4 MR distants (6 au final) situé à 3 km, et servant pour l'imagerie.

Chacun des MR est un ensemble de 19 antennes espacées de 5,5 m, et réparties en hexagone. Toutes les antennes ont le même alignement, mais les MR sont eux alignés suivant six angles différents. Les MR sont reliés entre eux par différentes lignes à retard, permettant de pointer différents endroits dans le ciel suivant une grille discrète. La combinaison des signaux de toutes les antennes permet de créer un faisceau analogique avec une largeur à mi-hauteur d'environ 8° à 85 MHz et 46° à 15 MHz. Ensuite, la combinaison des signaux de tous les MR se réalise de manière numérique, permettant un pointage continu, et également de créer un faisceau numérique de 0,5° à 85 MHz et 2,9° à 15 MHz.

Dû au pointage analogique sur une grille, NenuFAR peut observer efficacement des sources avec des élévations entre 20 et 90°. Les antennes ont été conçues de manière à pouvoir observer le ciel entre 10 et 85 MHz, correspondant respectivement à la fréquence de coupure ionosphérique et la bande FM.

Chaque antenne produit deux signaux, correspondant aux deux polarisations mesurées par l'antenne. Ces signaux sont collectés et numériser par le receveur *LaNewBA* (LOFAR super station Advanced NEW BAckend) pour former un signal à synthèse de faisceau de résolution fréquentielle 195,3125 kHz et de résolution temporelle 5,12 μs.

L'aire efficace de NenuFAR dépend à la fois de l'élévation et de la fréquence, donnant, pour un pointage au zénith, une aire efficace de 69 000 m^2 à 15 MHz et de 6 300 m^2à 85 MHz. Sur une heure d'observation, la sensibilité minimale de flux varie également de 130 mJy à 15 MHz à 9 mJy à 85 MHz, toujours pour un pointage au zénith.

Les données de LaNewBA sont ensuite traitées par deux machines nommées *UnDySpuTed* (Unified Dynamic Spectrum Pulsar and Time Domain receiver), permettant d'observer jusqu'à quatre positions différentes dans le ciel sur 37,5 MHz de largeur de bande.

Les données peuvent être traitées suivant quatre modes d'observation. Le mode *folded* où les données sont corrigées de la dispersion intra-canal, puis empiler à la période du pulsar. Le mode *single pulse*, également avec correction de la dispersion intra-canal mais sans empilement, permettant de voir les impulsions individuelles. Le mode *dynamic spectrum* permettant de générer des plans temps-fréquence avec une résolution fréquentielle accrue. Enfin, le dernier mode d'observation est le mode *waveform*, permettant de garder les données d'origine à 5,12 μs de résolution temporelle.

Afin de programmer les observations, une interface web dénommée *VCR* pour Virtual Control Room a été développée. Elle permet ainsi de voir la planification du télescope, de programmer des observations par l'import de fichier de configuration, et dispose d'outils de contrôle en temps réel ou non du télescope.

Part II

The NenuFAR pulsar blind survey

Chapter 3

Context and expectations

3.1 Overview of the history of the pulsar surveys

A survey is a program aiming to observe the sky in order to, either discover new sources or characterize the properties of a sample of known sources in a uniform way. In the context of the work presented in the first part of this thesis, it is a survey searching for pulsars in the radio wave range.

Depending on the specific goal of the survey, different strategies can be used to carry out the search. In the case of a search everywhere without restrictions, an all-sky survey can be performed. However, for some types of surveys, specific sky regions can be preferred to search for. In the case of pulsars, we know that their population is related to the population of massive stars, leading thus to realize sometimes (especially at very high frequencies) a survey of the galactic plane rather than an all-sky survey. Similarly, the progenitor stars of pulsars are massive stars, resulting in preferentially search in the massive star regions. Therefore, it exists of surveys of the galactic center, globular cluster surveys, or surveys of supernova remnants. On the opposite side, because of the important diffusive effects yielded by the ISM, some surveys observe the sky regions around the Galactic plane to perform a pulsar search deeper than a survey of the galaxy.

The first pulsar was discovered by the Mullard Observatory (Hewish et al. 1968) of the University of Cambridge. In the wake of this initial discovery, the sky was observed with the same telescope to realize the first Cambridge survey (Cole & Pilkington 1968; Pilkington et al. 1968). This first pulsar survey has permitted the detection of a substantial part of the first detected pulsars at the end of the sixties and the beginning of the seventies. At the same time, a northern hemisphere sky survey was carried out using the Green Bank telescope (Huguenin et al. 1968), and a southern hemisphere sky survey with the Molonglo Observatory (Large et al. 1968).

During the seventies, in the northern hemisphere, two other telescopes: Jodrell Bank (Davies & Large 1970) and Arecibo (Hulse & Taylor 1974), rapidly added several dozens of pulsars. In the southern hemisphere, a large sky survey: the second Molonglo survey, uniformly observed all the sky below +20°of declination. The discovery of 154 pulsars has permitted more than double the population of known pulsars (Manchester et al. 1978).

From the beginning of the nineties until the 2000s, some new important surveys are done, significantly increasing the number of known pulsars. In the northern hemisphere, the surveys performed using Green Bank and Arecibo were particularly prolific, especially the GBNCC (Green Bank North Celestial Cap survey) (McEwen et al. 2020) with 165 discoveries and the PALFA (Arecibo Pulsar survey using ALFA) (Lazarus et al. 2015; Parent et al. 2022) with 176 discoveries. In the southern hemisphere, after the two Molonglo surveys, several surveys are done using telescopes of the Parkes observatory, as the HTRU (High Time Resolution Universe Pulsar Survey) discovering 242 new pulsars (Keith et al. 2010). The Parkes telescope is the one that has discovered the most substantial number of pulsars, in particular with the 834 discoveries of the PMPS (Parkes multi-beam pulsar survey) (Manchester et al. 2001).

More recently, a new generation of radio telescopes, able to observe at lower frequencies, has been online for some years. Currently, the giant telescope FAST (Five-hundred-meter Aperture Spherical Radio Telescope) allows us to observe the sky in a very large bandwidth from 70 MHz to 3 GHz with high sensitivity. In only a few years, the different surveys performed with FAST at high frequencies

around 1.4 GHz (especially the GPPS), have discovered more than 200 pulsars (Han et al. 2021). Other low-frequency telescopes are based on the technique of the phased array, allowing one to observe lower frequencies while obtaining a larger collecting area. Using this type of telescope, the survey LOTAAS (Sanidas et al. 2019) has performed a low-frequency pulsar survey around 135 MHz with the telescope LOFAR (LOw Frequency ARray). The particular interest in the LOTAAS results is the significant, and larger than expected, fraction of the discoveries which are slow pulsars, i.e. with periods of several seconds. Moreover, the discovery of long-period pulsars, especially the discovery of J0250+5854 (Tan et al. 2018) which was the slowest pulsar at the time of the discovery, has indicated the possibility of an unknown population of slow pulsars located in the right part of the $P - \dot{P}$ plane.

Finally, another older phased array telescope: UTR-2, observing at a very low frequency, just above the ionospheric cutoff, between 8 and 33 MHz, has recently carried out a transient survey of the northern sky. With the detection of 40 known pulsars (Zakharenko et al. 2013; Vasylieva et al. 2014), this survey has confirmed the possibility of detecting pulsars at frequencies lower than those observed by NenuFAR.

3.2 Low-frequency observational difficulties

Except for the first discoveries, the majority of the known pulsars have been discovered at frequencies greater than 300 MHz. The advantage of these high-frequency observations is partially due to the diffusive effects produced by the ISM, which substantially increase towards low frequencies. The first one is dispersion, delaying the low frequencies relative to the high frequencies. In case of large dispersion, the pulse is spread, causing a decrease in the measured SNR, making it difficult the detection of weak pulsars. However, for low-frequency observations, the scattering of the pulse is more problematic. The amplitude of the pulse broadening increases rapidly, strongly limiting the observation of pulsars more distant than a few kpc. Moreover, the scattering screens generate scintillation at long timescales, potentially reducing the measured flux of a pulsar during an observation. At high frequencies, these ISM effects are less pronounced, making it easier the detection of unknown pulsars by a survey.

In addition to the diffusive effects caused by the ISM, another difficulty is the flux of the emission of the pulsars as a function of the observing frequency. Pulsars have a continuous broadband radio spectrum, with an increased flux towards lower frequencies until about 100 - 140 MHz. However, with a spectral slope of -2.6 (Lawson et al. 1987), the background noise due to the galactic noise increases more rapidly than the flux of the majority of the pulsars. Hence, for many pulsars, although the flux below 300 MHz is higher than above, the measured SNR is lower.

Finally, a last and important difficulty specifically concerns observations at frequencies below 100 MHz. For a large fraction of the known pulsars, a spectral turnover is seen in the range $100 - 140$ MHz (Bilous et al. 2020), meaning that the flux of the pulsars decreases towards lower frequencies (see Chapter 1.2.2). On the other side, the galactic noise continues increasing towards lower frequencies, leading to the dominant contribution to the background noise. At the frequencies of NenuFAR, the sky temperature thereby reaches several thousands of K, with a factor of 50 between "hot" regions located in the galactic plane, and "cold" regions (see Chapter 2.4). It consequently results in large variations of the amplitude of the background noise depending on the sky direction.

Because of these diverse difficulties at lower frequencies, the pulsars are easier to observe at high frequencies until their flux became too low above about a few GHz. Thus, except for the most intense known pulsars initially discovered at 81.5 MHz, almost all the surveys in history have chosen to search for pulsars at high frequencies generally around 1 or 2 GHz.

3.3 Expectations

3.3.1 Population synthesis and associated detections

To obtain an estimation of the possibilities of detection of a pulsar survey with NenuFAR, we have carried out simulations based on the Python software PsrPopPy. It is a program whose first part is the generation of a population of pulsars in the galaxy, followed by a second part simulating the operation of detection for each pulsar. This second part is performed for a series of historical surveys defined by the user.

In our case, we have started from the PMPS of Parkes to create the initial population of galactic pulsars. The series of surveys to simulate was then: PMPS, HTRU of Parkes, PALFA of Arecibo, GMRT survey, GBNCC of Grenn Bank, and LOTAAS of LOFAR. This last one was not present in the standard version of PsrPopPy and has been added. At each step, the program computes a flux and determines whether the pulsar is detected or not by the survey. The different surveys are done in chronological order. The LOTAAS survey is the most recent one and the closest to NenuFAR in frequency. Retrieve, as faithfully as possible the results of LOTAAS, in terms of detection or discoveries, was thus very useful to have a correct base for simulating the NenuFAR survey. Finally, the characteristics of a survey with NenuFAR have been determined and given as input to ultimately calculate the detection and discoveries.

However, the pitfall of PsrPopPy is that it is entirely based on the Parkes discoveries, i.e. on high-frequency discoveries. As a consequence, the computing of the flux of pulsars doesn't take the spectral turnover into account. It results from this lack of an overestimation of the flux below 100 MHz. To overcome this fundamental issue to properly simulate a survey with NenuFAR, I have implemented the turnover within PsrPopPy. The turnover is written based on the mathematical definition of the single turnover described in Bilous et al. (2020), i.e. a double power law of the form:

$$S(\nu) = \begin{cases} S_0 \cdot \left(\frac{\nu}{\nu_0}\right)^{-\alpha_{hi}} & | \ \nu > \nu_t \\ S_0 \cdot \left(\frac{\nu_t}{\nu_0}\right)^{-\alpha_{hi}} \cdot \left(\frac{\nu}{\nu_t}\right)^{\alpha_{lo}} & | \ \nu < \nu_t \end{cases} \tag{3.1}$$

With S_0 the reference flux at 1.4 GHz, ν_0 the reference frequency of 1 GHz, ν_t the frequency of the turnover, α_{hi} the high-frequency spectral index, and α_{lo} the low-frequency spectral index. A parameter has equally been added to define the fraction of the population showing a turnover. For each generated pulsar, the presence of a turnover or not is then randomly chosen according to the given fraction. Finally, in the case of a pulsar with turnover, the flux is calculated following another added parameter, which is a low-frequency spectral index defining relative to a statistical distribution.

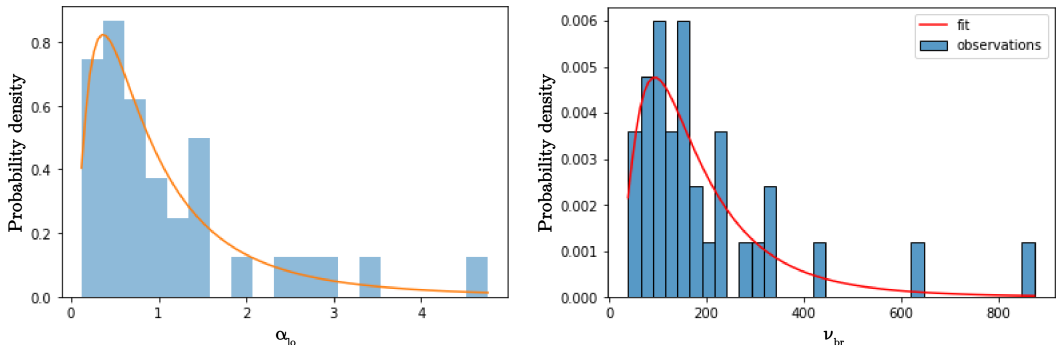


Figure 3.1: Distribution of the low-frequency spectral index in the left panel and turnover frequency in the right panel (R.Gros).

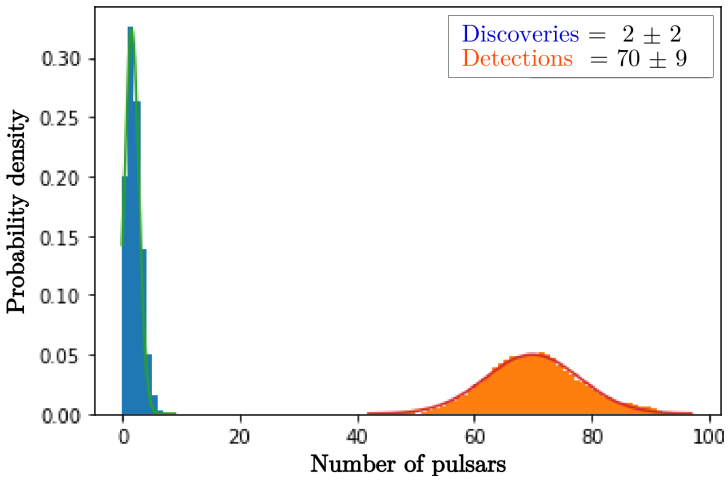


Figure 3.2: Distributions of the number of detections and discoveries for the 250 carried out simulations of PsrPopPy (R.Gros).

The initial population synthesis is randomly carried out following some distributions for the initial luminosity, period, and distance. Moreover, to determine if there is detection or not, PsrPopPy simulates scintillation. As a consequence, the results can significantly vary from one to another simulation. 250 simulations have therefore been performed in order to obtain average numbers of detection and discoveries.

The distribution of low-frequency spectral indices, presented in the left panel in Figure 3.1, has been determined using results of Bilous et al. (2020) to a log-normal distribution of mean 1.19 and standard deviation 1.27. Also, we have chosen that 50% of the population must show a turnover, and the turnover frequencies are equally defined based on those found by Bilous et al. (2020). They have been defined following a log-normal distribution of mean 195 MHz and standard deviation 155 MHz (see the right panel in Figure 3.1) Using these parameters of the simulation, the average number of detection is 70 ± 9 pulsars, and the number of discoveries is 2 ± 2 (see Figure 3.2).

The discoveries simulated by PsrPopPy are essentially due to synthetic pulsars without turnover, and then still having an increasing flux at the NenuFAR frequencies. For some discoveries of the 250 simulations, it may also be due to a very steep spectrum, allowing us to keep a sufficient flux below 100 MHz even with a turnover.

3.3.2 Widening of the pulsar emission cone

Simulations performed with the modified version of PsrPopPy give a low number of discoveries. However, as said previously, the synthetic population on which is based the simulation is generated relative to high-frequency surveys. Also, in order to explain the larger and larger number of slow pulsars discovered these last years, a theory based on the widening of the emission cone has been proposed (see Tan et al. (2018) for example).

According to the RFM, the lower the frequency the higher the emission height. It results from a higher altitude a wider cone for the pulsar beam. As a consequence, the emission cone has to be wider at the NenuFAR frequencies compared to those of the older surveys. That leads to the fact that the fraction of the sky covered by the emission cone is larger at low frequencies. The consequence is therefore that certain pulsars should be invisible at high frequencies and only visible at lower frequencies.

Moreover, the RFM has been modeled by adding two parameters to the frequency: the period P and the period derivative $\dot{P}$ of the pulsar. Using fit results of Kijak & Gil (2003), the altitude of the emission can be modeled as a function of these three parameters (see Equation 1.13 in Chapter 1.1.4). Combining with the equation of the radius of the cone (see Equation 1.12 in Chapter 1.1.4, the radius $\rho(\nu, P, \dot{P})$ may be modeled as:

$$\rho(\nu, P, \dot{P}) \;=\; \sqrt{\frac{9\pi}{2c} \cdot 163.10^6 \cdot \frac{\dot{P}^{0.07}}{\nu^{0.26} \,.\, P^{0.7}}} \;\approx\; 2.77 \cdot \frac{\dot{P}^{0.035}}{\nu^{0.13} \,.\, P^{0.35}} \tag{3.2}$$

With c the speed of light, and ν the observation frequency. Inserting the equation of the emission height, we obtain a radius of the cone as a function of the frequency, but also of the period and the period derivative. The fraction of sky $C(\eta, \rho)$ covered by the beam of the pulsar can be then geometrically calculated using the relation of Emmering & Chevalier (1989):

$$C(\eta,\rho) \;=\; 2\cdot\frac{1}{4\pi}\iint d\varphi\cdot d(\;\cos(\rho)\;) \;=\; \frac{1}{2\pi}\int_0^{2\pi} d\varphi\int_{\eta+\rho}^{\eta-\rho} d(\;\cos(\rho)\;) \;=\; \cos(\eta-\rho)-\cos(\eta+\rho) \tag{3.3}$$

η designates the angle between the rotation and magnetic axes. Using a sine distribution of η (used by Emmering & Chevalier (1989)), the probability of detection $f_{cross}(\rho)$ can be estimated by:

$$f_{cross}(\rho) \;=\; \int_0^{\pi/2} C(\eta,\rho)\cdot\sin(\eta)d\eta \;=\; 1-\cos(\rho)+\left(\frac{\pi}{2}-\rho\right)\cdot\sin(\rho) \tag{3.4}$$

Starting from the pulsars of the northern hemisphere of the ANTF catalog, I have divided them into five classes following their position in the $P-\dot{P}$ diagram. The five classes are represented by a color in Figure 3.3. The five classes are:

- millisecond pulsars (MSP);
- young pulsars such as the Crab pulsar (YP);
- normal pulsars (NP);
- slow pulsars featuring a period longer than 2 s (SP);
- high energy pulsars such as magnetars (HEP).

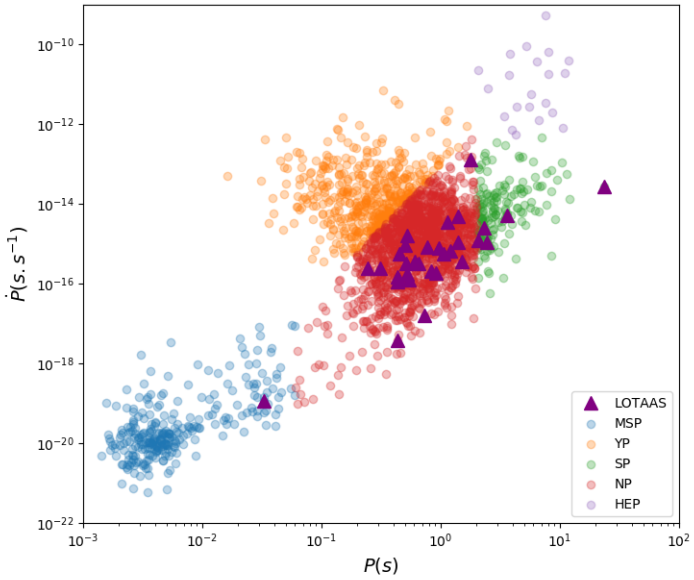


Figure 3.3: $P-\dot{P}$ diagram showing the classification of the pulsars of the ATNF catalog into 5 classes. The recent low-frequency discoveries of LOTAAS are marked in purple triangles.

For each class, a characteristic period and period derivative have been then determined by the average value of all the pulsars of the class. Based on the average characteristics, the probability has been calculated for a frequency of 300 MHz. The lower the frequency, the higher the probability of detection. Therefore, the frequency of 300 MHz has been chosen because corresponds to the usual minimum

frequency of the large majority of the older surveys (except the recent survey LOTAAS of LOFAR). Assuming that all the pulsars emitting towards Earth at the frequency $\nu = 300$ MHz have been discovered, the initial population can be simply retrieved using the relation:

$$N_{det}(\nu, P, \dot{P}) \;=\; f_{cross}(\nu, P, \dot{P}) \cdot N_{init} \;\Rightarrow\; N_{init} \;=\; \frac{N_{det}(\nu = 300MHz, P, \dot{P})}{f_{cross}(\nu = 300MHz, P, \dot{P})} \tag{3.5}$$

To obtain only high-frequency discoveries in the number of detected pulsars, the recent low-frequency discoveries of LOTAAS around 135 MHz have been removed from the set of discoveries. This set is the list of pulsars from version 1.68 of the ATNF catalog (Manchester et al. 2005). Computing the initial population for the five classes, it is interesting to look at the proportion of each class in the ATNF relative to the whole population and its initial estimated population in Table 3.1.

Table 3.1: Pulsar populations for each of the 5 classes. (1) Number of pulsars in the ATNF catalog. (2) Probability that the emission cone crosses the line of sight of Earth. (3) Computed number of pulsars in the initial population. (4) Proportion of the population of the class which is already discovered.

Class	$\mathbf{N}_{ATNF}$ (%) (1)	$\mathbf{f}_{cross}$ (2)	$\mathbf{N}_{init}$ (%) (3)	$\mathbf{N}_{ATNF}$ / $\mathbf{N}_{init}$ (%) (4)
MSP	180 (18)	78.7	226 (7)	80
YP	142 (14)	38.2	372 (12)	38
NP	596 (59)	25.5	2 156 (67)	28
SP	84 (8)	17.1	456 (14)	18
HEP	3 (0)	17.2	17 (1)	18

We can notice that MSPs represent 18% of the known pulsars, whereas only 7% of the initial population. It is exactly the opposite for the proportion of SPs of 14% over the whole initial pulsar population rather than 8% currently in the ATNF (without the six SP of LOTAAS). Moreover, about 80% of the MSPs would be already discovered, while merely 18% for the SPs and HEPs.

MSPs possess a small light cylinder and, according to Equation 3.2, a very wide cone. It results from this wide emission cone a high probability to detect them. However, in the context of a survey with NenuFAR, the scattering combined with their short period makes it difficult to detect them. Furthermore, because the light cylinder is very close to the surface, the emission of MSPs seems to not follow the RFM.

The YPs and HEPs are the two classes in the upper part of the $P-\dot{P}$ diagram. These pulsars show an extremely large variability of the emission and scattering. Moreover, especially in the case of HEPs which comprise extreme pulsars as magnetars, the standard model of the pulsars (and this model of course) is not necessarily proper. As a consequence, the probability of detection of these three types should be very low, leading to specifically focus on NPs and SPs. In addition, to support this statement, it should be noted that LOTAAS has merely discovered NPs, SPs, and a few MSPs (see purple triangles in Figure 3.3).

Because of their long period, SPs have a large light cylinder, leading that the emission at a given frequency is higher in altitude relative to the surface than MSPs for example. However, the radio emission is solely realized along the open field lines. Because of this large light cylinder, the open field lines come from a smaller region in the polar cap of the pulsar. It results from a smaller polar cap a tighter emission cone than shorter period pulsars, and consequently to a lower probability to cross the line of sight of Earth. Hence, decreasing the observing frequency is particularly useful to detect SPs and rather slow NPs, allowing for a widening of the cone.

The major problem for detecting pulsars at low frequencies is scattering. For each class, the distribution relative to the DM has been thereby calculated. A cutting factor f_{sca} has subsequently been determined by integrating the DM distribution until the DM (DM_{sca} in the equation) for which the average scatter broadening τ_s (computed using the relation of Bhat et al. (2004)) is greater than the average period of the class P:

$$f_{sca}(\nu) \;=\; \sum_{DM=0}^{DM_{sca}} p(DM) \;\mid\; \frac{1}{N_{\tau_s}} \cdot \sum_{DM=0}^{DM_{sca}} p(DM) \cdot \tau_s(\nu, DM) \;\geq\; \langle P \rangle \tag{3.6}$$

The scattering is averaged using the proportion $p(DM)$ (assumed to be equally the probability) of pulsars with such a DM. This cutting factor represents the fraction of pulsars with $DM < DM_{sca}$. An average number of discoveries N_{disc} may be estimated using the following relation:

$$N_{disc}(\nu, P, \dot{P}) \;=\; (\; N_{init} \cdot f_{cross}(\nu, P, \dot{P}) - N_{ATNF} \;) \cdot f_{sca}(\nu) \tag{3.7}$$

These numbers are only average numbers. Then, in order to evaluate an error, for each class, the standard deviation of the period $\sigma(P)$ of the pulsars is computed. The scattering factor is subsequently calculated again for the periods $P - \sigma(P)$ and $P + \sigma(P)$. Finally, a minimum and a maximum number N_{disc} are calculated to define the interval at 1σ.

In order to validate results, the consistency with LOTAAS discoveries has been tested. Table 3.2 presents thereby results at the frequencies of LOTAAS and NenuFAR. For the LOTAAS frequency, the scattering factor has been calculated for a maximum DM of 546 pc.cm^{-3} (Sanidas et al. 2019). For NenuFAR, the calculus has been done for a maximum DM of 100pc.cm^{-3} (dispersion and scattering for the NPBS will be discussed in more detail in Chapter 6).

Table 3.2: Scattering factors and estimated number of discoveries for the normal and slow pulsar classes. N_{LOTAAS} designates the real number of discoveries of the LOTAAS survey.

	LOTAAS			**NenuFAR**	
Class	$\mathbf{N}_{LOTAAS}$	$\mathbf{f}_{sca}$	$\mathbf{N}_{disc}$	$\mathbf{f}_{sca}$	$\mathbf{N}_{disc}$
NP	45	0.551	$46\ ^{+2}_{-6}$	0.307	$33\ ^{+0}_{-5}$
SP	6	0.668	$8\ ^{+1}_{-1}$	0.362	$6\ ^{+0}_{-0}$

For LOTAAS, the theory of the widening of the cone at low frequency returns a number of $46\ ^{+2}_{-6}$ NP consistent with the real number. Concerning the SPs, there is a small difference of one pulsar between the lower limit of the interval and the real number. Nevertheless, these two results show that the model is not entirely false.

For NenuFAR, the number of discoveries has been calculated comparing with the number of pulsars in the initially used ATNF catalog, i.e. added of the discoveries of LOTAAS N_{LOTAAS}. Finally, I have obtained $33\ ^{+0}_{-5}$ NPs and $6\ ^{+0}_{-0}$ SPs. This interval from 34 to 39 pulsars is much larger than the expectations of PsrPopPy. Nonetheless, it should be noted that, contrary to PsrPopPy, this model doesn't take the telescope sensitivity into account. Moreover, it also simulates scintillation. As a result, this result is rather an estimation of the maximum number of new pulsars geometrically detectable.

3.4 Aim of the present survey

Because of the difficulties involved by the low frequencies, almost all the searches for pulsars have been outside of the frequency range of NenuFAR between 10 and 85 MHz. Thus, a first obvious interest of a pulsar survey using NenuFAR is to reach a part of the spectrum relatively unknown. Moreover, except UTR-2 but which only observes in the lowest part of the bandwidth of NenuFAR, the previous telescopes below 100 MHz were less sensitive.

Because of the low sky temperature at high frequency, the sensitivity of surveys at high frequencies than 300 MHz features good sensitivity. Consequently, the pulsars visible below 100 MHz and invisible above 300 MHz have necessarily a steep spectrum. Furthermore, the recent pulsar survey of LOFAR LOTAAS, observing between 119 and 150 MHz has provided stronger constraints on the spectral indices allowing it to be invisible for LOFAR. LOTAAS has evaluated this sensitivity between 1 and 5 mJy. According to calculations of sensitivity in Chapter 2.5, the range for NenuFAR is from 0.27 to 3 mJy at 85 MHz. We can therefore determine the maximum spectral index γ to be detectable solely by NenuFAR.

$$\gamma \leq -\frac{\log(\, S_N/S_L \,)}{\log(\, \nu_N/\nu_L \,)} \tag{3.8}$$

Here S_N is the minimum flux required by NenuFAR, S_L is the minimum flux required by LOTAAS, ν_N is the frequency of NenuFAR, and ν_L represents the frequency of LOTAAS. Computed for the central frequency of LOTAAS of 135 MHz and 50 MHz for NenuFAR (almost the central frequency of NenuFAR), in the most favorable case for NenuFAR, the spectral index must be less than 2.93. In the worst case, γ decreases to -1.62.

The fact that LOTAAS could discover a significant number of pulsars leads to thinking there is a population of steep spectrum pulsars. However, such pulsars necessarily emit merely at a very low frequency. It results from this statement that this potential population cannot currently be known. As a consequence, the first interest is to reveal a part of this potential population.

Another interest in low-frequency pulsar surveys is created by the widening of the cone towards low frequencies. According to the RFM model, more pulsars have to be visible at low frequencies. Also, the calculations predict a substantial part of the short-period pulsars have been already discovered. However, for long-period pulsars, because of their narrower emission cone, it is required to observe as lower as possible in terms of frequency. As a result, whether the RFM is correct, a potential population of slow pulsars could be discovered.

Moreover, LOTAAS has discovered more slow pulsars than expected. Also, at least some of them have shown a steep spectrum. Increasing the slow pulsar population could be equally interesting for testing for a potential link between steep spectrum pulsars and slow pulsars.

The survey presented in the first part of this thesis is a survey of type all-sky. The final goal of this survey is to search for pulsars in the entire sky visible at the Nançay Observatory, i.e. above -23° of declination. The work presented here is actually the first stage of the survey, consisting of the observation of the entire sky above 39° of declination. Other stages will be carried out in the future in order to cover all the northern sky.

The first part of this thesis exposes the development of the *NenuFAR pulsar blind survey*, designated as NPBS hereafter. The first chapter explains the construction of the pointing grid allowing to observe efficiently the targeted sky, and the associated observing program is reported in the second chapter. Finally, the two last chapters present the processing pipeline uses to search for pulsar candidates, and an analysis method to sort candidates and select the most interesting ones for follow-up observations.

Résumé du chapitre :

Un relevé aveugle a pour but d'observer des parties du ciel dans le but de découvrir de nouvelles sources astrophysiques. Ils peuvent viser des parties spécifiques du ciel tel que le centre galactique, ou être au contraire des observations du ciel entier.

Une part majeure des pulsars découverts dans les années 60 et 70 ont été trouvés dans le cadre du premier relevé de Cambridge (Cole & Pilkington 1968; Pilkington et al. 1968). Durant ces premières années après la découverte du premier pulsar, plusieurs relevés du ciel ont été menés avec différents télescopes dans l'hémisphère nord : le télescope de Green Bank (Huguenin et al. 1968), le télescope de Jodrell Bank (Davies & Large 1970), et Arecibo (Hulse & Taylor 1974). Concernant le ciel de l'hémisphère sud, deux principaux relevés ont été réalisés à l'observatoire de Molonglo (Large et al. 1968; Manchester et al. 1978), et plus particulièrement le second relevé qui a doublé le nombre de pulsars connus à l'époque avec 154 découvertes.

Depuis le début des années 90, de nouveaux grands relevés du ciel ont été effectués : GBNCC (McEwen et al. 2020) avec Green Bank, PALFA (Lazarus et al. 2015) avec Arecibo, ou pour l'hémisphère sud le PMPS (Manchester et al. 2001) avec le télescope de Parkes. Ce dernier est d'ailleurs le relevé le plus prolifique avec pas moins de 834 découvertes. Plus récemment, la nouvelle génération de radio-télescopes réalise, grâce à une meilleure sensibilité et des fréquences observables différentes, de nouveaux relevés du ciel, tel que le GPPS (Han et al. 2021) avec FAST. Plusieurs relevés à plus basse fréquence sont également entrepris comme LOTAAS (Sanidas et al. 2019) avec LOFAR ou le relevé de UTR-2 (Zakharenko et al. 2013; Vasylieva et al. 2014) entre 10 et 30 MHz.

L'importante majorité des pulsars a été découverte à hautes fréquences. Cela s'explique par trois raisons majeures. La première est que le phénomène de "scattering" devient très important en amplitude, élargissant de manière très importante les impulsions qui deviennent alors indétectables. La seconde raison est que, bien que le flux des pulsars augmentent vers les basses fréquences, le flux du fond galactique augmente plus rapidement. Le rapport signal sur bruit diminue donc de plus en plus vers les basses fréquences. Enfin, dû au retournement du spectre vu pour de nombreux pulsars, le flux diminue drastiquement en dessous de 100 MHz. Le fond galactique devient alors prépondérant, menant à un facteur 50 de différence entre les régions chaudes et froides du ciel.

Le relevé présenté dans cette thèse est un relevé du ciel entier au-dessous de 100 MHz. Cette partie du manuscrit présente le développement du *NenuFAR pulsar blind survey*, où les deux premiers chapitres exposent la définition de la grille de pointage et la mise en œuvre du programme d'observation. Les deux derniers chapitres détaillent le traitement des données et l'analyse des candidats pulsars identifiés.

Chapter 4

Pointing grid

The first stage of the NPBS aimed to observe the entire sky above 39° in about two years, i.e. between 2020 and 2022. In order to observe the targeted sky, an observing program has been developed, based on a pointing grid defining the sky positions to observe.

4.1 Observing constraints

A pointing grid divides the sky into areas to observe corresponding to the size of the telescope beam. In a given time, it should allow for maximizing the coverage of the targeted sky. Hence, the definition of the grid must answer some observing constraints.

In practice, not all of the sky can be observed because of the limitations of the instrument and available time. Thus, some constraints must be taken into account, and a domain of definition has been chosen for the different parameters of the pointing grid.

4.1.1 Total observing time

The survey presented here is based on the results of the NenuFAR pulsar census. The census has observed the known pulsars and has detected about 180 pulsars. These results have confirmed the possibility of the detection of pulsars with NenuFAR. The NPBS is actually an extension of the census, where the search is now in blind mode.

For the census, the pulsars were directly targeted and were observed for a minimum duration of 20 min. In the context of a blind search, the positions of the pulsar and the pointing may be different. To obtain a sufficient flux to have a detection, the duration of an observation has been increased compared to the census to be set to 30 min.

Using hexagonal tiling, a first estimation of the number of pointings for the grid has been calculated, allowing an evaluation of the required total observing time. According to this estimation, the survey has been prepared for about 960 hours of observation. With NenuFAR, it is possible to observe up to four pointings at the same time (four digital beams). Combined with the duration of 30 min, the first constraint is that the number of pointings must be of about 7960 at the maximum.

For this initial stage of the NPBS, the choice was made to observe the northern polar cap, starting from 90° of declination and descending in declination. Depending on the beam size, the number of pointings will therefore limit the lowest declination attained by the first stage of the survey.

4.1.2 Beam shape

NenuFAR is a phased array telescope composed of mini-array spread over a certain area. Contrary to telescopes using dishes, the beam of NenuFAR is a function of the observed sky position. The beam of a phased array is based on the Fourier transform of the projected area on the sky. As a consequence, the shape and size of the beam are dependent on the azimuth and elevation of the source. The higher the source and the larger the size of the beam.

At the time of the definition of the pointing grid at the beginning of 2020, NenuFAR was not complete and was composed of 56 mini-arrays spread into an ellipse of about 200 x 400 m. As a result, the beam of NenuFAR had equally an elliptic shape. Moreover, the ellipse was slightly misaligned relative to the north-south and east-west axes, leading to an inclined elliptic beam.

The pointing grid aims is to define positions in order to cover the targeted sky, which is in the case of the survey the northern polar cap. However, the regular tiling of a sphere with inclined ellipses is very difficult and inefficient. To have an efficient tiling of the sphere, we have required a circular beam at the zenith. The first advantage is to have a rotation symmetry in the azimuth, leaving only an elevation dependency on the beam size. The second advantage is, because of the rotation symmetry, to have an aligned beam relative to the azimuth and elevation axes. As a consequence, we obtain a beam with a constant size in the azimuth, which is more elongated at lower elevations.

In addition, the gain of NenuFAR is also dependent on the projected area, leading to a maximum gain at the zenith. With the aim to search potentially weak pulsars, it has been chosen to maximize the sensitivity of NenuFAR by observing the pointings only during their meridian transit, i.e. at the time of the day when the source is at its highest elevation.

4.2 Selection of mini-arrays

4.2.1 Criteria of selection

Because of the ellipticity of the NenuFAR beam with 56 MA, it was required to select a sub-array composed of MAs allowing to obtain a circular beam at the zenith. To realize the selection, two criteria have been checked.

Theoretically, to have a circular beam, the shape of the sub-array on the ground must also be circular. However, the real shape of the beam is dependent on the positions of the different MAs in the sub-array, resulting in a beam which is not necessarily perfectly circular. Thus, the first criterion was obviously to check the ellipticity of the sub-array beam.

In the case of different sub-arrays showing a similar ellipticity, a second criterion was defined to choose the best sub-array, based on the maximization of the sensitivity. The second criterion is consequently the gain of the sub-array beam.

To find the most optimized sub-array, simulations of the digital beam for many diverse sub-arrays have been realized using the Python software *Nenupy*[1] developed by A.Loh (Loh & Girard 2020), and which is specially made to model NenuFAR beams and observations. *Nenupy* is a tool developed to model various aspects of NenuFAR, allowing, in particular, to model a beam (analog and digital) at a given position in horizontal coordinates (Alt-Az frame).

In order to obtain a circular beam, the tested sub-arrays have been built by selecting mini-arrays inside a disk defined by a centroid position and a given diameter. The digital beam of the sub-array was then simulated at the zenith, i.e. for an elevation of 90° and azimuth of 0°. Finally, the corresponding ellipticity and gain were computed.

According to the geometry of NenuFAR with 56 MA, which is an ellipse of 200 x 400 m, four diameters have been tested: 200, 205, 210, and 220 m. Concerning the centroid, its position has been moved on all the NenuFAR field with a step of 5 m. It results from the two criteria a map corresponding to all the tested sub-arrays. In all, 7 930 different sub-arrays have been tried.

4.2.2 Ellipticity map

To calculate the ellipticity, a digital beam was modeled for all azimuths from 0 to 360° with a step of 6°. For each azimuth, the gain is computed for elevations starting from 90 down to 88° with a step of 0.1°. The beam size has been determined for a gain attenuation, defined by $10log_{10}\left(\frac{G(\delta)}{G_{max}}\right)$ with $G(\delta)$ the gain at the elevation δ and G_{max} the gain at the center of the beam, arrives at $-3\ dB$. This value corresponds to a decrease of half of the gain compared to the gain at the center of the beam.

[1] https://nenupy.readthedocs.io/en/latest/index.html

For each tested sub-array, we obtain the beam size α as a function of the azimuth angle a. Also, in the case of a circular beam, the beam size must be equal to the average size whatever the azimuth. To estimate the degree of ellipticity of the beam generated by a sub-array of centroid position (x, y), an ellipticity estimator ϵ_{xy}, designed to obtain a value between 0 and 100, has been calculated based on the standard deviation of the beam size $\sigma(\ \alpha_{xy}(a)\ )$:

$$\epsilon_{xy} = \frac{100}{1 + 100 \cdot \sigma(\ \alpha_{xy}(a)\ )} \tag{4.1}$$

With $\sigma(\alpha_{xy}(a))$ the standard deviation of the beam sizes in all the azimuth directions compared to the average beam size $\overline{\alpha_{xy}}$:

$$\sigma(\ \alpha_{xy}(a)\ ) \ = \ \sum_{a=0°}^{180°} (\ \alpha_{xy}(a) - \overline{\alpha_{xy}}\ )^2 \tag{4.2}$$

For a circular beam, the standard deviation must be null, resulting in an ellipticity estimator of 100. On the opposite side, the more elliptic is the beam, the larger the standard deviation, leading to an ellipticity estimator which tends towards 0.

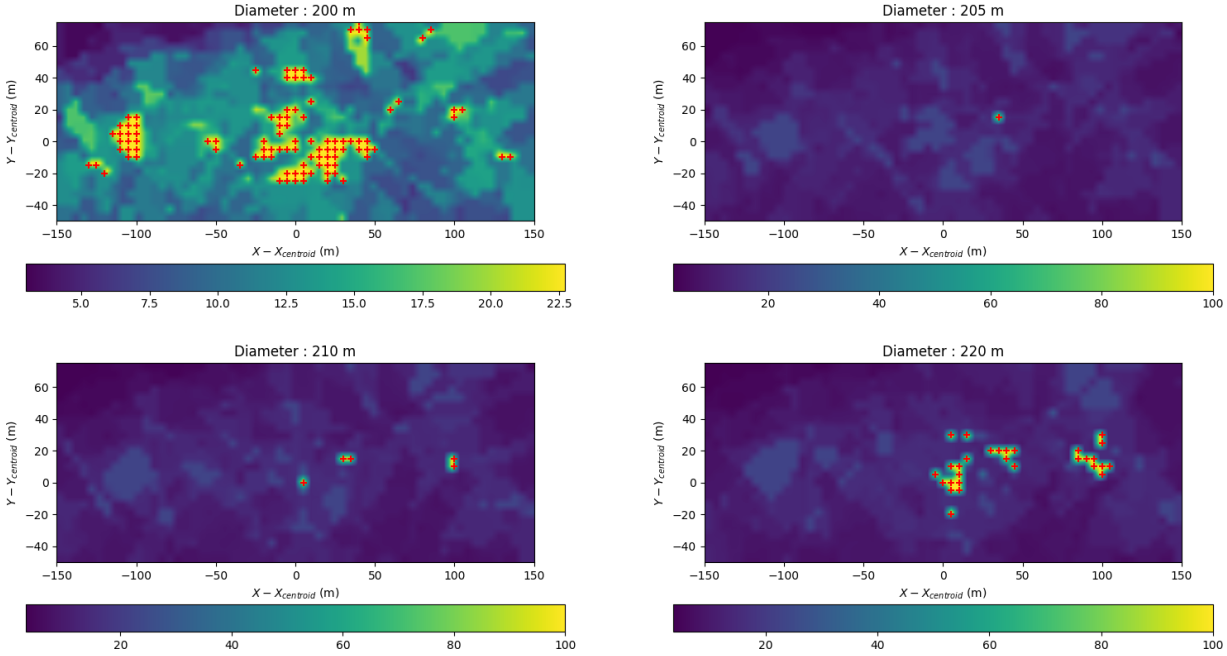


Figure 4.1: Maps of the ellipticity estimator for 4 fixed sub-array diameters: 200, 205, 210, and 220 m. Red crosses indicate the sub-arrays with the maximum value for the ellipticity estimator in the map.

Figure 4.1 shows the maps of the ellipticity estimator for the four tested sub-array diameters. For all the maps, each point corresponds to a centroid position (and then to a sub-array), and the color bar is ranged from blue for the most elliptic beams to yellow for the less elliptic beams. Finally, the best positions, corresponding to the maximum value of the ellipticity estimator ϵ_{xy} are marked with a red cross.

On the top left graph corresponding to a diameter of 200 m, there are a lot of yellow points and many red crosses in different parts of the map. However, the color bar is normalized and the maximum value is only 22.5, which is far from the value of 100 for a circular beam. This diameter doesn't allow us to find a subarray with a circular beam at the zenith and can be rejected.

In contrast, the three other maps present are essentially blue, with only some yellow points marked with red crosses. For the three corresponding diameters, these points correspond to a value of 100. As a result, it is possible to find sub-arrays with a circular beam at the zenith for the diameters 205, 210, and 220 m.

4.2.3 Gain map

For the second criterion of choice, the gain of the digital beam needs to be determined. In the context of a survey, an eventual pulsar is not necessarily located at the center of the beam. The gain value at the center of the beam is thereby not sufficient, and it needs to maximize the gain over the full digital beam.

For each tested sub-array, once the beam size is determined, the gain of the digital beam is evaluated as the sum of the gains of all the elevations inside this beam (i.e. with gain attenuation greater than -3 dB). However, in order to normalize the values of the different digital beams, the sum of the gains is divided by the number of elevations taken into account. It results in an average gain $\overline{G}_{xy}$, allowing it to normalize relatively to the size of the beam.

$$\overline{G}_{xy} = \frac{1}{N_\delta} \sum_{i=0}^{N_\delta} G_{xy}(\delta_i) \tag{4.3}$$

Here $G_{xy}(\delta_i)$ represent the gain of the beam at the elevation δ_i, and N_δ is the number of elevations with a gain attenuation greater than -3 dB corresponding to the size of the digital beam.

Similarly to the previous ellipticity maps, Figure 4.2 presents, for the four tested sub-array diameters, maps of the normalized average gain. But these 4 maps represent the gain of the digital beam G_{xy} for each centroid position (x, y). As well as the ellipticity maps, the lower gains are colored in blue, the higher in yellow, and the highest gain in the map is marked with a red cross.

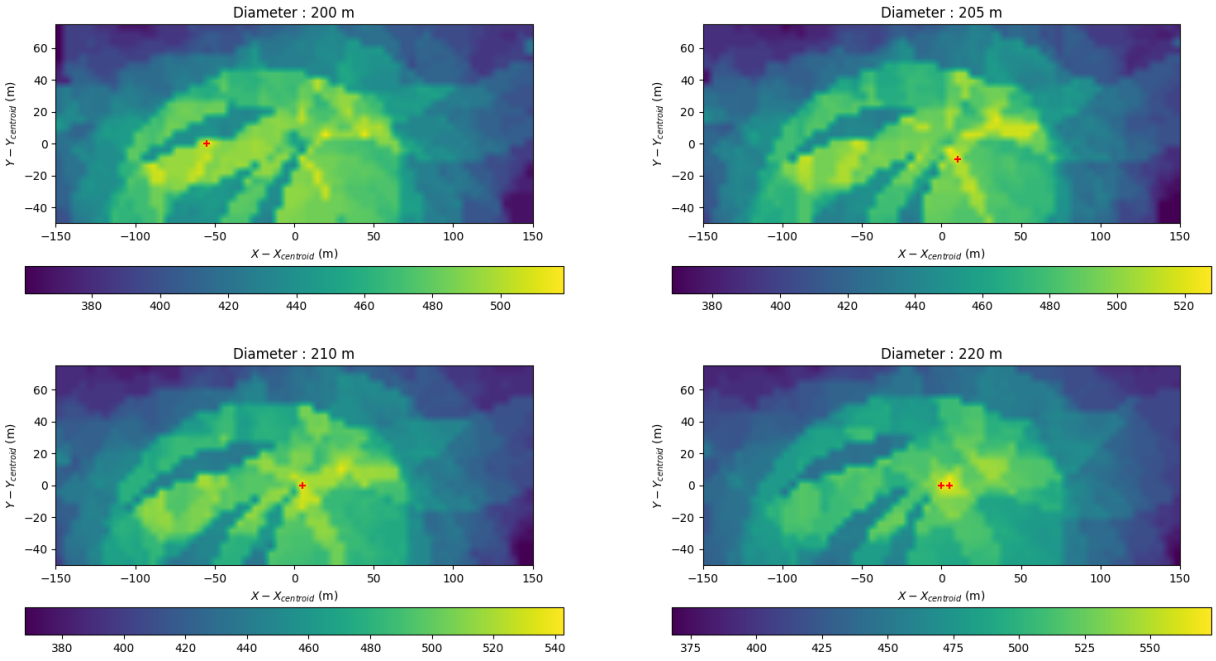


Figure 4.2: Maps of the normalized average gain for 4 fixed sub-array diameters: 200, 205, 210, and 220 m. Red crosses indicate the sub-arrays with the maximum gain on the map.

The three diameters 200, 205, and 210 meters show a unique point allowing to optimize the gain, and there are two positions for the diameter of 220 m. Concerning the 200 m diameter, it was already rejected because of the impossibility to find a sub-array with a circular beam at the zenith. For the three other diameters, the ellipticity maps are crossed with the corresponding gain maps, to check if it exists positions matching a circular beam with the best gain. In the case of the diameter 205 m, no matching sub-arrays are found, resulting in the rejection of this diameter. Finally, for the diameters of 210 and 220 m, there are three positions where the ellipticity and the gain match well.

4.2.4 Final mini-array selection

According to the two criteria, the optimized sub-array to use for the NPBS must be one of these three configurations. The centroid positions of the corresponding sub-arrays are $(x, y) = (0, -5)$ m for a diameter of 210 m, and $(x, y) = (0, 0)$ m and $(x, y) = (0, -5)$ m for a diameter of 220 m.

The two configurations for a diameter of 220 m are actually the same selection of MAs, leaving only two possible sub-array configurations. The two remaining sub-array configurations show close gain and a circular beam at the zenith. However, the larger the sub-array, the smaller the telescope beam. Furthermore, in order to maximize the coverage of the sky, the smaller diameter is, therefore, more interesting.

As a consequence, the MAs constituting the final sub-array used for the NPBS are those inside a disk of 210 m around the centroid position X = 5 m and Y = 0 m relative to the NenuFAR centroid (the centroid of the 56 mini-arrays online in 2019). Figure 4.3 shows the map of the MAs of NenuFAR relative to the centroid of the optimized sub-array. Red hexagons represent the 25 selected MAs for the optimized sub-array, and the black hexagons correspond to excluded MAs.

Finally, two MAs contained inside the disk of 210 m in diameter have been excluded because too polluted by *Radio Frequency Interference* (RFIs hereafter, see Chapter 5.2.1 for more details). These two MAs are located on the bottom of the disk, which is indeed close to the correlators of NenuFAR and LOFAR-FR606. Because of not perfect isolation of the air conditioning systems in a particular point of the Faraday cage, many broadband RFIs are emitted in the specific direction of these two MAs, resulting in unusable data.

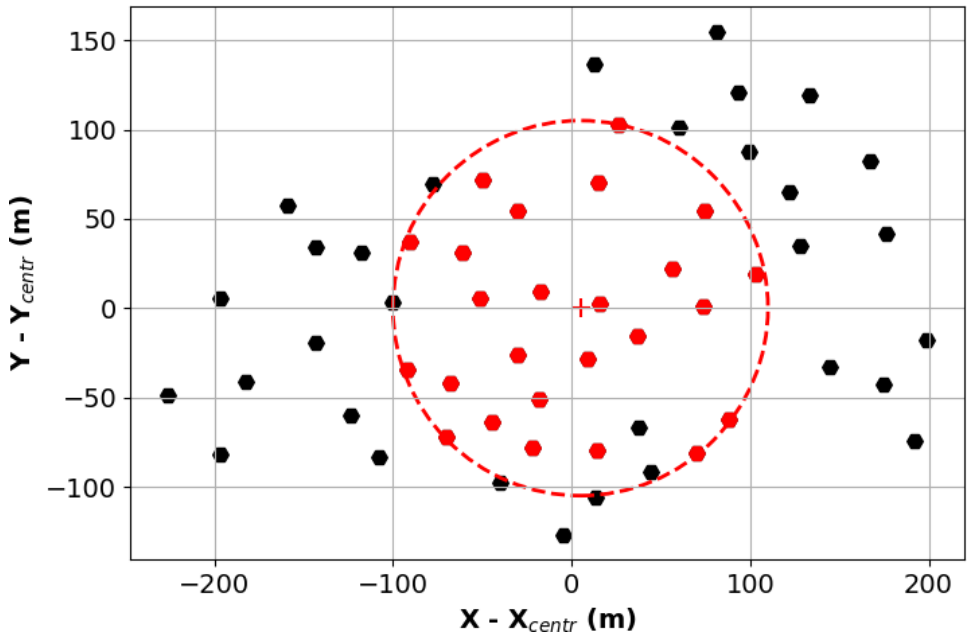


Figure 4.3: Map of the 56 MAs of NenuFAR in 2019 centered on the centroid of the 56 MA. The direction X corresponds to the East-West axis and the direction Y to the North-South axis. The red hexagons are the mini-arrays selected in the final sub-array of the NPBS (those inside a circle of 210 m of diameter in dashed red). The black hexagons are the excluded mini-arrays.

4.3 Determination of the observing parameters

4.3.1 Parameters

To realize the tiling of the sky, one needs to know the size of the telescope beam, which is the size of the digital beam in our case. The telescope beam is a Fourier transform of the sky on the ground plane. In the case of a circular beam, the telescope beam radius α can be estimated using the relation of the full-width half-maximum of a circular aperture:

$$\alpha = \arcsin\left(\frac{1}{2} \cdot \frac{1.02c}{\nu \cdot \eta \cdot D}\right) \tag{4.4}$$

Here, ν is the observing frequency, D is the diameter of the telescope, c is the speed of light, and η is a factor related to the effective area of NenuFAR depending on the elevation. The beam size is a function of the diameter, which is the diameter of the sub-array of 210 m. It thereby remains the observing frequency, which is the first observing parameter required to be determined.

In the more complex configuration, regular tiling of a sphere is possible with hexagons. The NenuFAR beam being elliptic (or at least circular), a regular tiling is impossible, resulting in unobserved gaps between the pointings. In order to reduce the unobserved sky, an overlap rate τ_{ov}, corresponding to the fraction of the beam radius in overlap with the neighbor beam, has been defined.

4.3.2 Criteria

Because of the constraint on the total available time (see Section 4.1.1), the number of poitings is limited to approximately 7 680. As a consequence, a high overlap rate means a decrease in the fraction of the sky observed. On the opposite, observed at lower frequency increases the size of the beam, leading to also an increase in the total observed sky. These two observing parameters are consequently related.

To determine these parameters, three criteria can be defined with the aim to maximize the observed sky. Pointings are defined from the north pole to the lowest reaching declination. The first criterion is then obviously the lowest declination of the grid.

Also, because of the non-regular tiling of the beams, it is important to check the coverage of the sky which will be observed by the grid. According to Equation (4.4), it results from the observing frequency dependency of the beam size a changing along the bandwidth. For sky coverage, two different criteria can be therefore defined. The first criterion is the average sky coverage along the bandwidth. For the second criterion, one can notice that the higher the frequency, the smaller the beam size. Hence, the second criterion is the sky coverage at the highest frequency.

4.3.3 Determination of the sky coverage

To determine the sky coverage of a pointing grid, the chosen method was to compare it with a regular grid. I have then defined a linear spherical gridding placed on the sky observed by the pointing grid (i.e. for declinations greater than the minimum declination of the tested pointing grid). The different tested pointing grids had several thousands of pointings. Hence, the sky grid for the comparison was constructed with about 4 million points, allowing a sufficient density of points around each pointing, while keeping a reasonable computing time.

For each point of the sky grid, the closest pointing of a given pointing grid, located at the angular separation $\vec{\rho}$, is searched. The beam associated with the pointing being elliptic, the size $\vec{\theta}$ is dependent on the direction. As a result, to determine whether the sky point is inside the beam, the beam size is projected in the direction between the pointing and the sky point $\hat{\rho}$. In the celestial frame (RA-DEC frame) of the pointing grid, a ratio $\mathcal{R}$ is then calculated between the projected beam size and the angular separation:

$$\mathcal{R} = \frac{\hat{\rho} \cdot \vec{\theta}}{\|\vec{\rho}\|} = \frac{\vec{\rho} \cdot \vec{\theta}}{\|\vec{\rho}\|^2} = \frac{\sqrt{(\rho_r \cdot \theta_r)^2 + (\rho_d \cdot \theta_d)^2}}{\rho_r^2 + \rho_d^2} \tag{4.5}$$

The indices r and d correspond to the components in right ascension and declination. In the case of $\mathcal{R} > 1$, it means that $\theta > \rho$, so then the sky point is inside the beam. Finally, the percentage of unobserved sky area A_{UO} can be estimated by the ratio between the points with $\mathcal{R} < 1$ and the total number of points in the sky grid N_{grid}:

$$\mathcal{A}_{UO} \;=\; 100 \cdot \frac{N_{grid}(\ \mathcal{R} < 1\ )}{N_{grid}} \tag{4.6}$$

The beam size θ is also dependent on the frequency, and the variation of the unobserved sky area relative to the frequency can be evaluated with the variation of the size relative to the frequency. Using Equation (4.4), the frequency dependence of θ compared to the central frequency ν_c is:

$$\sin(\ \theta(\nu)\ ) \;=\; \frac{1}{2}\frac{1.22\ c}{\nu \cdot \eta \cdot D} \;=\; \frac{1}{2}\frac{1.22\ c}{\nu_c \cdot \eta \cdot D} \cdot \left(\frac{\nu_c}{\nu}\right) = \sin(\ \theta(\nu_c)\ ) \cdot \left(\frac{\nu_c}{\nu}\right) \tag{4.7}$$

So, using the approximation of the little angles: $\sin(\ \theta(\nu)\ ) \approx \theta(\nu)$, the relation 4.5 for the ratio $\mathcal{R}$ can be calculated for the frequency ν:

$$\mathcal{R}(\nu) \;=\; \frac{\sqrt{(\rho_r \cdot \left(\frac{\nu_c}{\nu}\right) \cdot \theta_r)^2 + (\rho_d \cdot \left(\frac{\nu_c}{\nu}\right) \cdot \theta_d)^2}}{\rho_r^2 + \rho_d^2} \;=\; \left(\frac{\nu_c}{\nu}\right) \cdot \mathcal{R}(\nu_c) \tag{4.8}$$

Based on the unobserved sky at the central frequency, it is possible to evaluate the evolution of this unobserved sky along the bandwidth.

4.3.4 Final central frequency and overlap rate

NenuFAR can observe between 10 and 85 MHz. However, the frequencies below 25 MHz present a very high fraction of RFI, leading to be rejected in the context of the survey. On the opposite side, for frequencies above 80 MHz, the RFIs caused by the FM band are equally important. As a result, the observed bandwidth of 37.5 MHz should be contained between 25 and 80 MHz, resulting in a central frequency approximately in the range of $40 - 65$ MHz. Moreover, the sensitivity of NenuFAR is maximum around $55 - 60$ MHz. Consequently, the central frequency should be defined around these best frequencies.

Concerning the overlap rate, some preliminary tests have shown the unobserved sky area is too high for small values of overlap. On the opposite side, the very high overlap rates obviously don't allow to reach low declinations. As a consequence, the overlap rate should be defined in the range of $0.45 - 0.8$.

56 grids have been simulated for central frequencies between 50 and 62 MHz with a step of 2 MHz and an overlap rate in the range of 0.45 - 0.8 with a step of 0.05. The result of these simulations is reported in Figure 4.4, which is composed of eight panels for the different overlap rates. In each panel, the percentage of unobserved sky area has been computed for 21 frequencies linearly distributed. Each line corresponds to a central frequency, indicating the unobserved area from the lowest to the highest frequency along the bandwidth of 37.5 MHz. Also, the minimum declinations reached by the seven different grids are superposed in each graph with green diamonds.

Based on these results, the two criteria about the sky coverage (see Section 4.3.2) have been defined to a maximum unobserved area of 12.5% at the highest frequency, and a mean unobserved area of less than 2.5%. In Figure 4.4, the dashed black line represents this limit of 2.5% and allows us to see the frequencies with sufficient coverage.

In the graphs corresponding to the overlap rates below 0.6, we can see that, whatever the central frequency, the unobserved area is larger than 2.5% in a large part of the bandwidth. Moreover, the limit of 12.5% is also exceeded whatever the central frequency. On the opposite, for overlap values greater than 0.65, some of the grids exceed the 40° of minimum declination reached, and the majority of the tested grids exceed 35°. As a consequence, according to the defined criteria, it seems that the overlap rate may be chosen between 0.6 and 0.7.

A high overlap rate means redundantly observing more sky. Moreover, the lower the overlap, the lower the minimum declination reached by the grid. Hence, a low overlap close to 0.6 should be preferred. Furthermore, on the graph corresponding to an overlap rate of 0.6 in Figure 4.4, central frequencies below

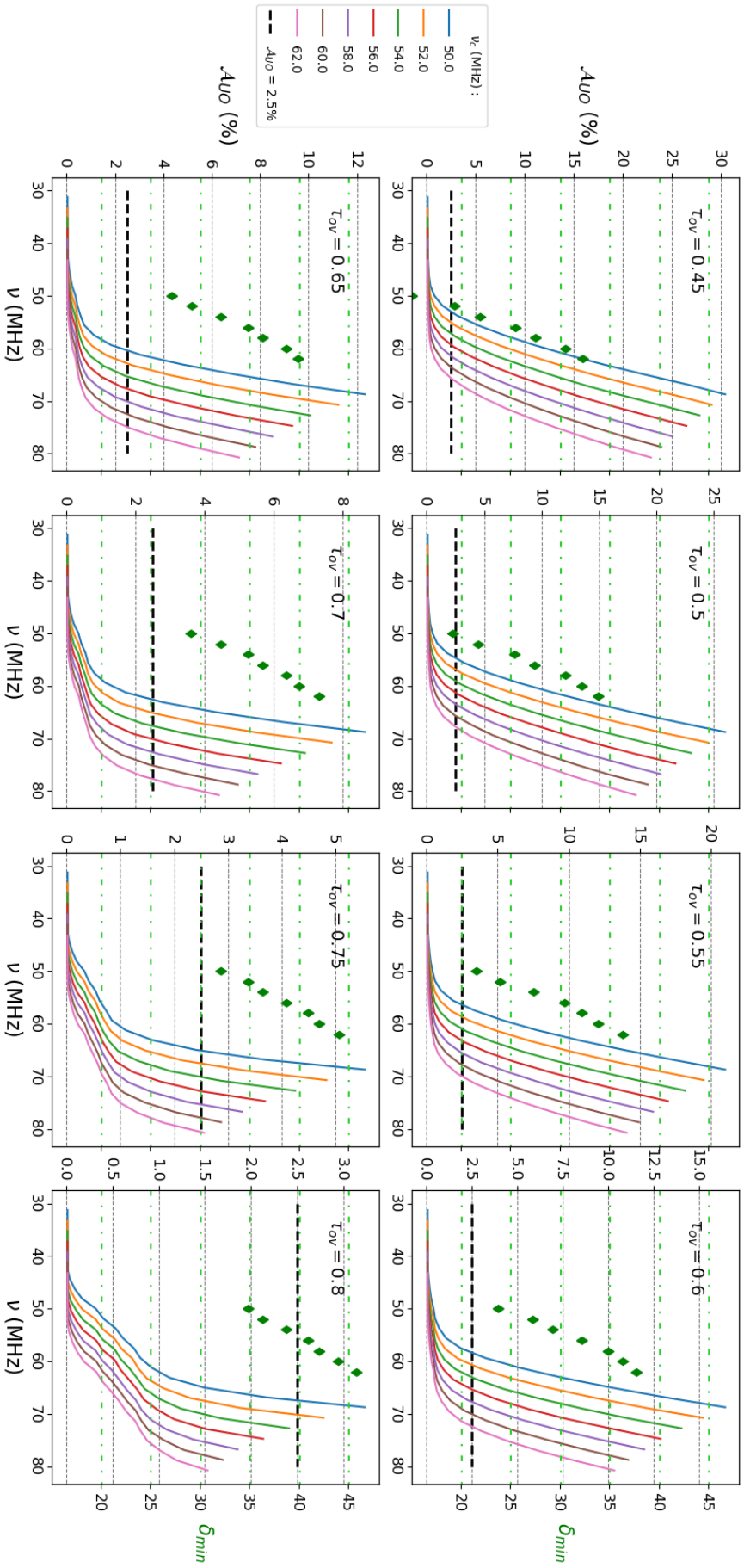


Figure 4.4: Evolution of the fraction of the unobserved area along the bandwidth for seven values of central frequencies ν_c and eight different values of overlap rate τ_{ov}. Each graph corresponds to a fixed overlap rate, and each line in the graph corresponds to a specific central frequency. The green diamonds correspond to the lowest declination (right y-axis and dashed grid lines) reached by the grid for each central frequency. The dashed black line represents the 2.5% limit for the fraction of unobserved area.

56 MHz rise above the limit of 12.5% of the maximum unobserved area. Moreover, those higher than 58 MHz give a minimum declination higher than 35°. As a consequence, a central frequency close to a value of 58 MHz should be preferred.

In the second step, a more precise analysis has been carried out in a more restricted range. Grids have been simulated with a better resolution of 0.25 MHz in frequency in the range [56 ; 60] MHz, and with an overlap rate step of 0.01 in the range [0.58 ; 0.65]. Typically, the smaller the unobserved sky area, the higher the minimum declination. As a result, the unobserved sky area is related to the minimum declination. To compare all the grids, a unique normalized indicator has been calculated taking into account these two characteristics. The indicator $\mathcal{N}_{var}$ is normalized relative to the grids with the same given central frequency, leading to obtaining an indicator as a function of the overlap rate.

$$\mathcal{N}(\tau_{ov})_{\nu_c} = \left(\frac{\delta(\tau_{ov})_{\nu_c} - \delta^{min}_{\nu_c}}{\delta^{max}_{\nu_c} - \delta^{min}_{\nu_c}} \right) + \left(\frac{\mathcal{A}_{UO}(\tau_{ov})_{\nu_c} - \mathcal{A}^{min}_{UO,\nu_c}}{\mathcal{A}^{max}_{UO,\nu_c} - \mathcal{A}^{min}_{UO,\nu_c}} \right) \tag{4.9}$$

The first term quantifies, for a given central frequency ν_c, the variation of the minimum declination δ compared to the lowest declination in all the different grids. The second term quantifies, still for a given central frequency ν_c, the variation of the unobserved sky area $\mathcal{A}_{UO}$ compared to the lowest unobserved sky area in all the different grids. As a result, the minimum value of $\mathcal{N}$ corresponds to the grid with the minimum combination of the two desired characteristics.

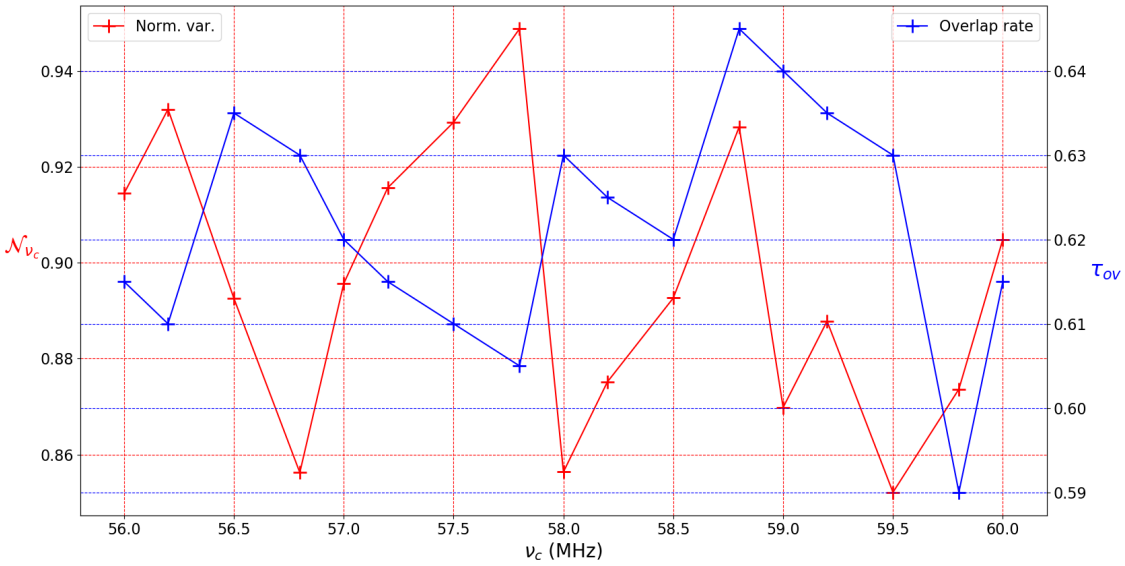


Figure 4.5: Minimal normalized variation $\mathcal{N}_{var}$ (red curve) relative to the frequency, and corresponding overlap rate τ_{ov} (blue curve).

Figure 4.5 shows a graph where, for each central frequency,the value of the minimal normalized variation $\mathcal{N}_{\nu_c}$ is indicated by the red line, and the corresponding overlap rate τ_{ov} is indicated by the blue line. In this graph, we can identify three points with a low value of $\mathcal{N}_{\nu_c}$ at frequencies of 56.75, 58, and 59.5 MHz. Also, for the three points, the corresponding overlap rate is 0.63. Thus, the optimal overlap rate τ_{ov} for the pointing grid of the NPBS has been set to 0.63.

For the choice of the frequency, the three minimum points present an equivalent value of $\mathcal{N}$. In the first analysis presented in Figure 4.4, compared to the neighbor central frequencies, the central frequency of 58 MHz appeared to represent a reasonable compromise between the maximum unobserved area and the minimum declination reached by the grid. Consequently, the central frequency ν_c of the final pointing grid used for the NPBS has been fixed to 58 MHz.

4.4 Determination of the pointing positions

With the different defined observing and grid parameters, the positions of the pointings can be determined to construct the grid. The beam size is dependent on the elevation of the pointing, leading to generating the grid using an iterative process rather than a direct tiling of the sky.

4.4.1 Distribution in declination

The shape of the beam of NenuFAR is not dependent on the declination but on the elevation. Thus, the distribution of the grid in declination must be computed in the telescope frame, i.e. in the Alt-Az frame.

The basic condition to select the MAs of the sub-array of the survey is to have a circular beam at the zenith (see Section 4.1.2). The starting point to determine the grid is consequently to calculate the radius of the digital beam at the zenith with Equation (4.4) of the full-width-half maximum of the circular aperture. At the zenith, we can consider the effective area of the sub-array is maximum, and in order to estimate the diameter, the parameter η has been set to 1.

$$\alpha_{zen} = \arcsin\left(\frac{1}{2} \cdot \frac{1.02\ c}{\nu \cdot D}\right) \tag{4.10}$$

Here α_{zen} is the beam diameter at the zenith, c the speed of light, ν the central frequency, and D the diameter of NenuFAR. Based on the radius of the digital beam at the zenith, the lower elevations where pointings must be placed can be determined starting from the elevation $h_0 = 90°$. For each elevation h_i, the elevation just below h_{i+1} is then placed at a distance of a diameter, subtracted to the overlap τ_{ov}:

$$h_{i+1} = h_i - (2 - \tau_{ov}) \cdot \alpha_i \ \mid \ h_i, h_{i+1} \ \in \ [\ 30° \ ; \ 90°\ ] \tag{4.11}$$

Here α_i designates the radius of the beam at the elevation h_i. The next iteration corresponds to a lower elevation, where the beam radius is larger. Consequently, the beam radius α_{i+1} at the elevation h_{i+1} is determined relative to the reference beam size at the zenith α_{zen}:

$$\alpha_{i+1} = \arcsin\left(\frac{\sin(\alpha_{zen})}{\sin(h_{i+1})}\right) \tag{4.12}$$

Because the projected area highly decreases for lower elevations, NenuFAR can observe efficiently above 20°. Also, in the context of the NPBS, we have been required to observe only elevations above 30°Å list of available elevations is finally computed until an elevation lower than 30° is reached.

For the pointing grid, we are interested in a tiling of the sky in the celestial frame, i.e. in RA-DEC coordinates. In the context of the survey, pointings are observed only during the transit at the meridian. In this configuration, each elevation h_i corresponds to two specific declinations $\delta(h_i)$. These declinations can be determined using the trigonometric relation of the frame transform:

$$\delta(h_i) \ = \ \arcsin\{\ \sin(l_{Nançay}) \cdot \sin(h_i) - \cos(l_{Nançay}) \cdot \cos(h_i) \cdot \cos(a)\ \} \tag{4.13}$$

With $l_{Nançay}$ the latitude of Nançay is equal to 47.376511 °, and a the azimuth. The observations are done at the transit at the meridian, corresponding to the two azimuths 0°and 180°. As a result, the lowest declination corresponds to the azimuth $a = 0°$, and the highest declination corresponds to $a = 180°$.

4.4.2 Distribution in right ascension

Once the list of declinations is realized, it needs to place the pointings in the right ascension to tile the sky. Due to the observation at a constant azimuth, the beam radius in the right ascension is constant and equal to the beam radius at the zenith α_{zen}. The perimeter of a parallel in a sphere decreases when the declination increases, resulting in a smaller number of pointings at high declination. For a declination $\delta = 0°$ and a given overlap τ_{ov}, the number of pointings is computed based on the reference number of pointings able to be defined in the circle of perimeter 2π:

$$N(\delta = 0°) = \frac{2\pi}{(2 - \tau_{ov}) \cdot \alpha_{zen}} \ \mid \ N(\delta = 0°) \ \in \ \mathbb{N} \tag{4.14}$$

For a given declination δ_i, the number of pointings $N(\delta_i)$ is then:

$$N(\delta_i) = \left\lceil \frac{2\pi}{(2-\tau_{ov})\cdot\alpha_{zen}} \cdot \cos(\delta_i) \right\rceil \quad | \quad N(\delta_i) \in \mathbb{N} \tag{4.15}$$

The pointings are placed with a linear spacing on the circle of declination δ_i:

$$\forall\, n \in [\,0; N(\delta_i)-1\,] \subset \mathbb{N}\,,\; RA_n^{\delta_i} = \frac{n}{N(\delta_i)} \cdot 2\pi \tag{4.16}$$

Ellipses don't allow a perfect tiling of a sphere, resulting in blanks inside ellipses. To increase the coverage, a shift $\Delta RA(\delta_i)$ depending on the declination is applied. Finally, for each declination δ_{i+1}, the first pointing in the right ascension is placed in the middle of the two firsts pointings of the above declination δ_i such as:

$$\Delta RA(\delta_{i+1}) = \Delta RA(\delta_i) + \frac{\pi}{N(\delta_i)} \tag{4.17}$$

Starting from the north pole, i.e. $\delta = 90°$, the pointings are distributed in right ascension by decreasing declinations. The definition of the grid is finally stopped when the total number of pointings exceeds 7960 (the number of hours planned to be observed, see Section 4.1.1).

4.5 Final pointing grid

Table 4.1 reminds the general parameters used to observe the sky, with observations of 30 minutes in a bandwidth of 37.5 MHz in the manner to have 4 simultaneous close pointings per observation. All observations are made during the transit of the sky area at the meridian, i.e. for an azimuth of 0° if observed towards the north or 180° if observed towards the south.

Table 4.1: General parameters used for observations and the definition of the pointing grid.

Parameter	**Value**
Azimuth	0° or 180°
Duration	30 min
Central frequency	58 MHz
Bandwidth	37.5 MHz
Number of beams	4 per obs.
Overlap rate	0.63
Number of MAs	25
Number of antennas	476
Diameter	210 m

The last three parameters of the table refer to the sub-array of NenuFAR used for the NPBS. It is composed of 25 mini-arrays of 19 antennas, corresponding to a total of 476 antennas spread on a disk of 210 m in diameter.

For observations made at a frequency of 58 MHz with a sub-array diameter of 210 m, we obtain, according to the relation of the beam radius (see Equation (4.4)), a digital beam of 1.44° representing a solid angle of 8.10^{-6} sr or 0.026 deg^2. Taking the overlap rate of 0.63 into account, the grid is then defined in 43 elevations (listed in Table 4.2 with the corresponding beam diameter) allowing to observe until an elevation of 42.97°.

Table 4.2: Computed elevations with the corresponding beam diameter in elevation α_{el}.

Elevation (°)	α_{el} (°)	**Elevation** (°)	α_{el} (°)
90.000	1.438	67.808	1.554
89.015	1.439	66.744	1.566
88.029	1.439	65.672	1.579
87.043	1.440	64.590	1.593
86.056	1.442	63.499	1.607
85.069	1.444	62.398	1.623
84.080	1.446	61.286	1.640
83.089	1.449	60.163	1.658
82.097	1.452	59.027	1.678
81.102	1.456	57.878	1.699
80.104	1.460	56.714	1.721
79.104	1.465	55.535	1.745
78.101	1.470	54.340	1.770
77.094	1.476	53.128	1.798
76.083	1.482	51.896	1.828
75.068	1.489	50.644	1.860
74.048	1.496	49.369	1.895
73.023	1.504	48.071	1.934
71.993	1.513	46.746	1.975
70.957	1.522	45.393	2.021
69.914	1.532	44.009	2.071
68.865	1.542		

In addition to the elevation dependence, the beam size is also a function of the frequency of observation. Table 4.3 presents the sizes of the beam for the lowest observation frequency, the central frequency, and the highest frequency. These sizes are calculated for the two extreme elevations: the zenith and the lowest reached elevation at 42.97° corresponding respectively to the smallest beam and the largest beam. The beam size is decomposed in size in RA and size in DEC, which are, for observation during the transit at the meridian, equivalent to the sizes in the azimuth and elevation. The two last columns show the average angular size of the elliptical beam and the corresponding digital beam solid angle. Finally, the three last lines present the same quantities for the average beam of the total pointing grid.

From this table, we can extract the information that the beam has a mean size of 2.07° and covers a mean solid angle of 0.028 deg^2. The beam diameter varies along the bandwidth and along the elevation of a factor of 2.4, from 1.54° at 76.75 MHz at the zenith to a size of 3.73° at 39.25 MHz for the lowest elevation. This implies a solid angle of the beam which varies from 0.012 deg^2 to 0.173 deg^2, representing an important variation of a factor of 14.4. As a consequence, it results in a large difference in the sky coverage between these two extrema.

After conversion from the telescope frame to the celestial frame using Equation (4.13) and distribution of the pointings in right ascension, the obtained final grid is composed of 7 692 pointings distributed over 51 declinations between 39.47° and 90° of declination. Table 4.4 lists the declinations with the number of pointings for each one. The grid is represented in Figure 4.6 on three different frames: the celestial frame in RA-DEC, the horizontal frame (Alt-Az) in azimuth-elevation, and the bottom plot presenting the stereographic projection of the RA-DEC grid on the celestial equator.

Figure 4.6: Final pointing grid showing the 7 692 digital pointings to observe. The bottom plot shows the stereographic projection of the RA-DEC grid on the equatorial plane. The top left plot shows the RA-DEC grid and the top right plot shows the same grid in the altitude-azimuth frame. Each color represents a particular declination.

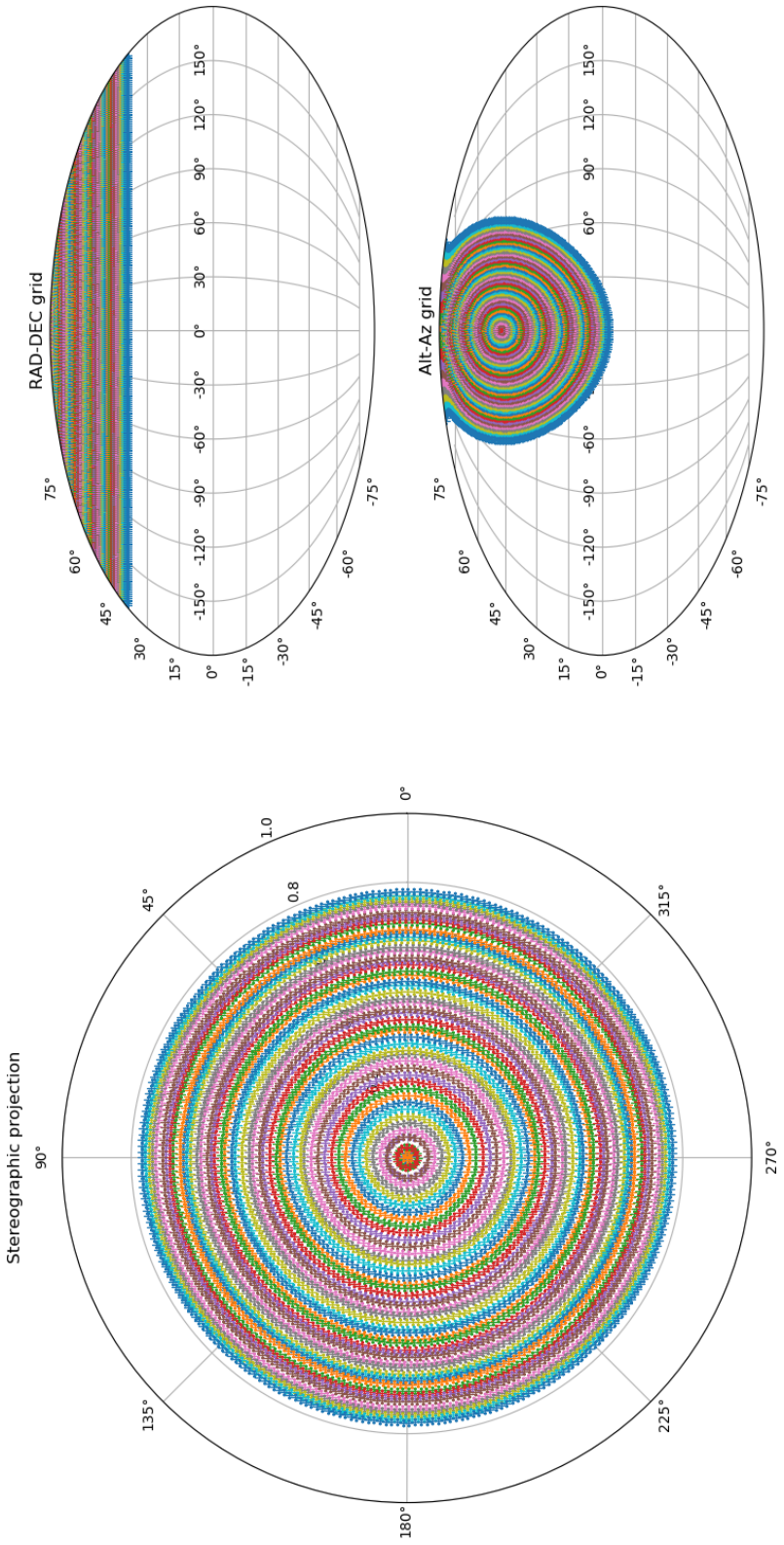


Table 4.3: Evolution in elevation and frequency of the beam size for the lowest, central, and highest frequencies, for the zenith and the minimum elevation. (1) Average beam diameter in right ascension. (2) Average beam diameter in declination. (3) Average angular beam diameter. (4) Corresponding solid angle of the beam.

Elevation (°)	**Frequency** (MHz)	(1) **RA** (°)	(2) **DEC** (°)	(3) **Avg. Diam.** (°)	(4) **Solid angle** (deg^2)
90	39.25	2.13	2.13	3.01	0.084
	58	1.44	1.44	2.03	0.026
	76.75	1.09	1.09	1.54	0.011
29.93	39.25	2.13	3.06	3.73	0.173
	58	1.44	2.07	2.52	0.054
	76.75	1.09	1.57	1.91	0.023
Average	39.25	2.13	2.21	3.06	0.090
	58	**1.44**	**1.49**	**2.07**	**0.028**
	76.75	1.09	1.13	1.57	0.012

From this grid, we can compute the angular distances between the closest pointings. Table 4.5 presents the minimum, the maximum, and the average angular distances in the right ascension and declination direction in the two first columns. Finally, the last column gives the angular separation on the sky.

We can see that, in the right ascension direction, the pointings are spaced by an average distance of 1.35°İn the declination direction, the pointings on two neighboring lines are spaced with an average distance of 1°. Finally, the angular separation goes from 0.10 to 0.99°. Furthermore, the mean value of 0.98° is close to the maximum separation, meaning the majority of the pointings are approximately spaced by the average distance. This is consistent with a mean beam diameter of about 2° and an overlap rate of 63% of the beam radius.

To conclude, the typical characteristics of the pointing grid used for the NPBS are resumed in Table 4.6. The overlap rate and the central frequency are chosen to have a sky covering close to 100% over the majority of the bandwidth. It results from the chosen values for these two parameters, a grid that should allow us to observe 98% of the sky above 38.74° of declination. Ultimately, even at the highest frequency of 76.75 MHz (where the beam size is the smallest), the fraction of unobserved sky will be less than 10% with a sky coverage of 90%.

Table 4.4: Computed declinations with the corresponding number of pointings for each declination.

Declination (°)	**Number of pointings**	**Declination** (°)	**Number of pointings**
89.362	5	63.326	165
89.294	5	62.306	170
88.016	13	61.291	176
88.002	13	60.280	182
86.729	21	59.273	187
86.633	22	58.270	193
85.477	29	57.270	198
84.245	37	56.272	203
83.033	45	55.277	209
81.838	52	54.285	214
80.659	60	53.294	219
79.496	67	52.305	224
78.347	74	51.318	229
77.211	81	50.331	234
76.087	88	49.345	239
74.975	95	48.359	243
73.874	102	47.374	248
72.783	109	46.389	253
71.702	115	45.403	257
70.630	122	44.417	261
69.565	128	43.431	266
68.509	134	42.443	270
67.460	141	41.454	274
66.417	147	40.463	278
65.381	153	39.471	283
64.351	159		

Table 4.5: Minimum, maximum, and average distances between the centers of the neighboring pointings. (1) Angular distances in right ascension. (2) Angular distances in declination. (3) Angular separation.

	(1) **RA** (°)	(2) **DEC** (°)	(3) **Ang. Sep.** (°)
Minimum	1.27	0.01	0.10
Maximum	1.42	1.28	0.99
Average	**1.35**	**1.00**	**0.98**

Table 4.6: Characteristics of the NPBS pointing grid.

Parameter name	**Parameter value**
Lowest freq.	39.25 MHz
Highest freq.	76.75 MHz
Zenith beam size	1.44°
Number of elevations	43
Min. elevation	42.97°
Number of DEC	51
Min. DEC	38.74°
Number of pointings	7 692
Min. total obs. time	961.5 h
Avg. beam angular diameter	2.07°
Avg. beam solid angle	0.028 deg^2
Avg. beam separation	0.98°
Avg. sky coverage	98.41 %
Sky coverage at the high. freq.	90%

Résumé du chapitre :

Le relevé aveugle a pour but d'observer le ciel au-dessus de 39° de déclinaison en à peu près deux ans. La première étape a donc été de définir une grille de pointage maximisant la couverture du ciel, tout en respectant deux contraintes. La première contrainte est le temps disponible. Basé sur un pavage hexagonal, le temps total pour effectuer le relevé avait été évaluer à 960 heures au total. La seconde contrainte est la modification de la forme du faisceau du télescope en fonction de la position de la source, empêchant ainsi un pavage régulier du ciel.

De manière à avoir un faisceau symétrique, permettant un pavage régulier, un sous-ensemble de MR a été sélectionné, utilisant des simulations du faisceau au zénith pour différents groupes de MR contenus à l'intérieur d'un disque positionné. 7 930 différents sous-ensembles ont été testés, correspondant à différentes positions centrales et à quatre diamètres différents : 200, 205, 210, et 220 m.

Pour chaque simulation, le degré d'ellipticité et le gain du télescope ont été calculés, obtenant à la fin des cartes d'ellipticité et de gain. Trois configurations permettaient d'obtenir un faisceau parfaitement circulaire au zénith, tout en ayant le gain maximum possible. La configuration finale choisie est un groupe de 25 MR répartis dans un disque de 210 m positionné avec des décalages sur les axes propres de NenuFAR (en 2019) de -5 m et 0 m par rapport au centre de gravité de NenuFAR.

Le faisceau du télescope en en fait la transformée de Fourier du ciel par rapport au plan du télescope. Ainsi donc, le diamètre du faisceau peut être estimé en première approximation par l'équation de la transformée de Fourier dans le cadre d'une ouverture circulaire. Le diamètre du faisceau en réalité dépendant de la fréquence d'observation, induisant par conséquent une variation de la taille du faisceau au cours de la largeur de bande.

Pour déterminer la fréquence centrale optimale, les deux critères de choix ont été la proportion de ciel observé et la déclinaison minimale atteinte. Pour optimiser la couverture de ciel, un paramètre de recouvrement, correspondant à la fraction de rayon qui recouvre le faisceau du pointage voisin.

56 grilles ont été simulées pour des fréquences centrales entre 40 et 65 MHz et des taux de recouvrement entre 0,4 et 0,8. Pour chacune, la proportion de ciel non observé a été calculée pour 21 fréquences réparties sur la largeur de bande. Trois configurations obtenaient des couvertures de ciel et des déclinaisons minimales similaires. Finalement, celle permettant le meilleur compromis est une fréquence centrale de 58 MHz avec un taux de recouvrement de 0,63.

La définition des positions des pointages de la grille se fait en partant d'un pointage circulaire au zénith, utilisant l'équation de l'ouverture circulaire. Les élévations suivantes sont ensuite calculées grâce au rayon précédemment déterminé. Le processus est ensuite itératif, avec le calcul du rayon correspondant à cette nouvelle élévation. Enfin, pour chaque élévation, les pointages sont ensuite distribués linéairement en azimut.

La grille de pointages finalement utilisée comporte 7 692 pointages, permettant d'observer le ciel, entre 39,25 et 76,75 MHz, au-dessus de 39° de déclinaison avec une proportion moyenne de ciel observé de 98,41 %.

Chapter 5

Observing program

5.1 Selection of the pointings to observe

The pointing grid defined in the previous chapter comprises 7 692 pointings. Only half of the bandwidth is used, leading to possibly observing up to four numerical beams, i.e. four sky pointings, at the same time. The following step was to solve the problem: which pointings among the 7 692 of the grid can be observed in a given time slot?

It results from the duration of each observation of 30 minutes a minimum time to observe all the targeted sky of the survey of 961.5 hours (about 40 complete days). This problem needs therefore to be solved for a total of almost 2 000 observations to program. I have created a script in *Python* called SELECTING_POINTINGS.PY, with the aim to automatically search for the best pointings to observe for a given time slot. The program searches among the 7 692 pointings that are yet unobserved and define those which are visible for the given date. In the second step, the program performs some choices to optimize the filling of the sky.

5.1.1 Selection of the visible pointings

To determine whether a pointing is visible or not, we have relied on a basic condition due to the fact that NenuFAR is a phased array telescope. Thus, this characteristic of NenuFAR occurs that its sensitivity is proportional to its projected area on the sky. Consequently, the sensitivity of NenuFAR is related to the elevation of the target in the sky. Logically, we want to observe each pointing at its maximum elevation. The equation of the elevation relative to the declination of the target is provided by the following relation:

$$h = \arcsin\left(\cos l_{Nançay} \cdot \cos\delta \cdot \cos A_H + \sin l_{Nançay} \cdot \sin\delta \right) \tag{5.1}$$

With h the elevation, $l_{Nançay}$ the latitude of Nançay, δ the declination of the pointing, and A_H the hour angle of the pointing. If we look at this equation, we can see the maximum elevation occurs for an hour angle A_H equal to 0 or 180°, corresponding to the moment when the target passes through the meridian. The basic condition is thus to observe each pointing during its meridian transit.

The observing time is divided into time slots of 30 min, where the middle of the time slot is defined as the reference time for the meridian transit. The search of the visible pointings is therefore to identify those with a meridian transit close to the reference time of the observation. This additionally allows for minimizing the variation of the elevation of the target during the observation, with a symmetric variation around the maximum elevation.

For each reference time, the pointing grid is placed in the frame of the telescope by transforming the coordinates of the grid from the celestial frame (RA-DEC) to the horizontal frame (Alt-Az). The position of each pointing relative to the meridian of Nançay is then estimated by looking at the azimuth of the pointing at the reference time. Azimuth shifts are determined relative to a reference azimuth, of 0°for pointings towards the south (i.e. a declination lower than the latitude of Nançay), and of 180°for those towards the north.

Finally, the pointings inside a sky band of more or less 2° around the meridian at the reference time of the observation are selected as good pointings able to be observed. This band corresponds to pointings with an angular sky separation in the azimuth direction less than 2° from the meridian at the corresponding elevation. This value of 2° allows maintaining a difference in time between the transit at the meridian and the reference time of fewer than 8 minutes. It permits strongly limits the variation of the elevation during the observation. Indeed, for the extreme case of a target passing through the zenith with a difference of 8 minutes compared with its meridian transit, we can calculate using Equation (5.1) that the variation of the elevation during the observation is of about 4° with a minimum elevation of 86°.

5.1.2 Optimization of the sky coverage

The basic condition to determine the visible pointings can give a large number of pointings, at least at the beginning of the observing program. Thus, we need to choose the best pointing, and so to define a good criterion of choice. This criterion must obey two constraints to be considered a good criterion. The first one is, in order to obtain the best able efficiency of observation, the maximization of the filling of the time slots. The second one is that, in order to avoid saturation of certain sky bands, it must give a homogeneous filling of the sky in the long term.

To have the best efficiency, it is necessary to observe a maximum of pointings for each allowed observing time. To do this, we have used a property of NenuFAR which can observe up to four sky areas at the same time. Indeed, we can define four digital beams which can observe different sky areas, with the condition that they must be contained inside the analog beam of NenuFAR, which has a diameter of about 8° at the central frequency of 56 MHz. This analog beam being created by physical delay lines, it is fixed and can point to only one sky area. The direction of the beam is changed every six minutes to follow the movement of the sky. During these six minutes, to follow the target, the digital beam transits inside the analog beam.

Consequently, to avoid having an significant loss in sensitivity, we need to prevent digital beams being at the edge of the analog beam. Thus, to optimize the filling of the time slot, three or four pointings are searched, with the constraint that they must be separated by less than 5°. With this value, we ensure all the digital pointings are closer than 2.5° relative to the mean coordinates used to define the analog beam. During the six minutes, at the maximum, the movement of the sky corresponds to an angular sky separation of 1.5°. It results that digital beams are, at each moment, close to the center of the analog beam.

This first constraint allows defining of some good groups of pointings which have the best possible efficiency for the given time slot. Now, among these groups of pointings, we need to know the one which will permit us to have the most efficient coverage of the sky in the long term. The aim is to homogeneously fill the grid, leaving a minimum of “holes" in the sky coverage.

To respond to these constraints, we can use a geometric property of the pointing grid. For an RA-DEC grid on the sphere, there is a tightening of the right ascension lines towards the poles, i.e. towards the high declinations. With the constraint to observe only during the meridian transit, the observing time is directly related to a range of visible right ascensions. On the other hand, pointings having a similar angular size on the sky whatever the declination, high declination pointings cover consequently a larger right ascension range than low declination pointings. As a consequence, at high declination, there are fewer pointings, which can be in addition seen longer than at lower declination.

This means there are more opportunities to observe the high declination pointings. The extreme case occurs for the pointings with a declination greater than 88°, which are permanently located closer than 2° from the meridian, and consequently which can be observed at any moment. We can conclude declination is an appropriate criterion of choice to define what group of pointings must be observed first. Thus, because of the available time to observe the pointings at low declination is shorter than at higher declinations, the second criterion is to choose the group of pointings with the lowest declination.

With this criterion and the presented constraints to manage the scheduling of the NPBS, I could define the process of selection. I have thereby written the algorithm of the script SELECTING_POINTINGS.PY to choose in an automatic way all the pointings to observe for any allocated time.

The process consists in selecting the pointings not already observed inside the visible band of 4°

around the meridian, and next, ordering these by declination. The second step is to search the three closest visible pointings of the pointing with the lowest found declination. This process is next iterated on each pointing, starting from the lowest found declination towards the highest declination, until a group of at least three pointings separated by less than 5° is found.

The last step of the algorithm is for the case where no remaining group is found. After a while, some right ascension regions are rather well filled, and there may eventually only remain some isolated pointing, the “holes" left in the sky cover. Ultimately, this final part searches the isolated visible pointings and tries to find, if it is possible, a neighbor at less than 5° to complete the filling of the grid.

5.2 Observing characterisitcs

5.2.1 Definition of the observing time

The searched pulsars being probably faint, the associated signals might easily be drowned in the ambient noise. For this reason, we need the best possible observing conditions to disentangle the pulsar and the background.

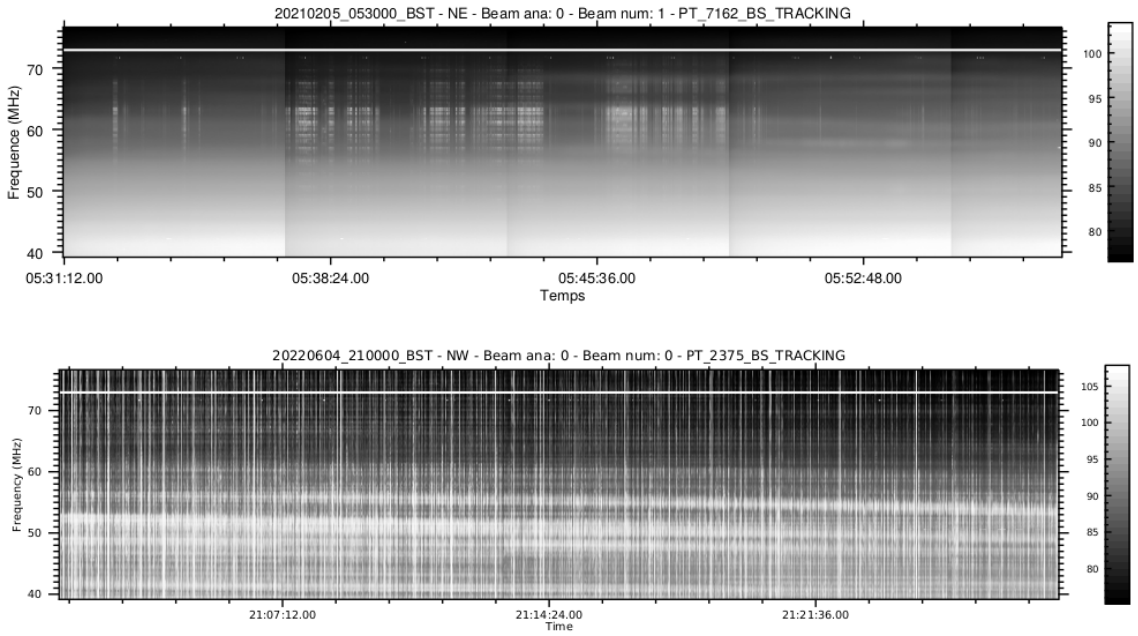
The radio-wave frequency range is naturally occupied by an important number of signals coming from a lot of different sources, e.g. the FM radio between 87.5 and 108 MHz. Moreover, there are an important number of technologies that emit electromagnetic waves in different frequencies and various directions, which are RFI. Some of them, such as FM radio, emit in a narrow band, and are precisely localized in frequency, while their intensity is approximately constant in time. Figure 5.1 shows two dynamic spectra, i.e. a graph of the measured signal in time and frequency, which presents different types of RFI. On both plots, we can notice a white horizontal line at about 73 MHz, which is an example of a narrowband RFI. On the opposite side, some RFI, such as those created with a mobile phone for example, are spread over many frequencies, but in a more or less short time. The top panel of Figure 5.1 shows an example of RFI localized in frequency between 57 and 65 MHz and localized in time. Also, there are natural sources of RFI, as shown in the bottom panel of Figure 5.1, which presents an observation during a thunderstorm.

The global signal can be seen as a sum of three major components: the mean background noise (which is typically white noise), the sum of the different RFI, and the pulses of the pulsar if there is one in the observation. Methods exist to find RFI and remove them. However, when a bin of data seeming to be an RFI is found, it is impossible to disentangle the signal from the RFI of an eventual pulsar signal. Consequently, to remove RFI, we need to zap entirely the bin of data, which may remove some signal of the pulsar at the same time.

The consequence is that the more RFI is present in the data, the more pulsar signal is lost. The search for pulsars is, primarily, essentially based on a periodicity search using the Fourier transform. Consequently, with fewer pulsations, the periodicity search will be less efficient.

To use the allocated time with the best efficiency, the NPBS requires observations during the quietest moment of the day, i.e. with the lowest level of RFI. The measure of the rate of RFI on the site of Nançay relatively of the hour has been monitored for some years using NenuFAR observations. These measures have shown the rate of RFI is, in a logical way, drastically lower during the night, especially between 21 and 6 h UTC (Bondonneau et al. 2021). Thus, all the observations of the blind survey were done in this time-lapse of 9 hours, divided into 18 slots of 30 minutes. Hence, it allows us to observe up to 72 sky pointings per night.

Figure 5.1: Dynamic spectra (time vs frequency graph), corresponding to two observations of the blind survey, showing different types of RFI, appearing in the white parts of the plot. Top: broadband RFI from 57 MHz to 65 MHz. Bottom: close thunderstorm in the sky visible by NenuFAR during the observation. One can see the typical signature, i.e. a lot of vertical lines, representing broadband signals, with in addition important intensity variations (large white patterns between 45 and 55 MHz) making the background inhomogeneous. On the two plots, there is also an example of narrowband RFI appearing at 73 MHz which is continuous in time.



5.2.2 Observing cadence

Using the time slot of 9 night hours, it is possible, if each night of observation is totally filled, to observe the whole sky above 39° in 106 nights. However, it is impossible to monopolize the telescope time each night for more than 3 months. Furthermore, observing with such a high cadence doesn't represent the best way to optimize the available observing time.

All the pointings are observed during their transit at the meridian at the same moment of the day between 21 and 6 h UTC. As a consequence, for consecutive nights, the visible sky is almost the same from one day to the next. There is thereby just a small shift of the observable sky band of 4° (see Section 5.1.1) due to the daily movement of the sky. Also, all the pointings have a diameter in the right ascension of about 1.4°. As a consequence, observations during consecutive nights result in the observation of almost the same pointings, and then to fill a single sky band.

All observations have a duration of 30 minutes. Consequently, in the time slot of nine hours, from the first meridian transit to the last one, the time-lapse is actually 8.5 hours. With a daily shift of about four minutes, the sky band located at the end of the night can be seen for 127.5 days. For a daily observation, this particular sky band will thereby be observed 127 times. With the choice made to observe from the lowest declination to the highest declination, we should approximately observe 127 different declinations in this sky band. However, the pointing grid is distributed over 51 declination values (see Chapter 4.5). Hence, all the pointings in this particular sky band will be done in about one and a half months.

The first consequence is that, after 51 days, the sky bands visible during the 3.5 first hours (51×4 min = 204 min) of the time slot will be unoccupied. The second consequence is that, due to the lower number of pointings at high declination (see Chapter 4.5), the polar cap will be rapidly observed. However, due to the longer time to be observable (see Section 5.1.2), these high declination pointings might be kept as "jokers" for the time slots (especially at the end of the observing program), where all the low declination pointings are already observed. To avoid this saturation, it is required to space in time the observing

nights. The time difference must allow to linearly fill the right ascensions in the first step and to linearly fill the declinations in the second step.

To estimate the optimal cadence, I have realized simulations of the scheduling of the blind survey for different cadences. The start of the observing program was set for the beginning of August 2020, with the aim to complete the observing program in about two years, corresponding to the end of this thesis. As a consequence, the maximum duration of two years determines the possible cadence. In order to observe the 7 692 pointings in 24 months with a maximum of 72 pointings a night, the delay between two consecutive observation nights must be less than 6.85 days.

Four different cadences have been simulated: 3, 4, 5, and 6 days between two nights of observation. For each cadence, a calendar has been defined, and a mock schedule has been created using the script which selects the pointings that can be observed (see Section 5.1). Because the planning of NenuFAR is set for a semester, simulations were done over a time-lapse of 6 months. The result allows us to see the trend of each cadence, and then choose the most adapted value.

Cadence	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
3 days	100%	103%	49%	46%	47%	52%
4 days	104%	104%	85%	67%	57%	65%
5 days	103%	108%	97%	88%	75%	77%
6 days	103%	110%	97%	97%	96%	90%

Table 5.1: Evolution of the filling of the allocated time slots over 6 months. The ratio of filling is calculated relative to the mean number of observations a month for the given cadence.

Table 5.1 presents the fraction of filling of time slots. This fraction corresponds to the number of hours used in a month relative to the average number of hours. The average number of hours corresponds to the total number of nights on the semester divided by 30.5 (an average number of days in a month) and multiplied by 9 hours (the duration of one time slot). Then, for months with 31 days, it is possible to have a fraction a little bit greater than 100% if almost all the allocated hours are occupied. We can see this effect in the two first months because the time slots are almost entirely filled. Also, the fraction of used time decreases rapidly for higher cadences, where just about half of the hours are used at the end of the semester. On the opposite side, for the low cadences, the filling stays more homogeneous, especially with a cadence of six days where 90% of the allocated hours are used after 6 months.

As we can observe up to four pointings in one observation, the number of hours with an observation is not the only thing that matters. To increase the efficiency of the use of the observing time, it needs to maximize the number of pointings per observation, trying to have up to four pointings for a maximum of observations. Furthermore, a too-low number of pointings per observation results in a slow coverage of the sky. Hence, it could be difficult to complete the whole program in the planned two years.

Table 5.2 presents the fraction of filling of the observations with pointings. This fraction corresponds to the number of pointings carried out in the month relative to an average number of possible pointings (or digital beams). The average number is calculatd as the average number of hours for a month, previously used for the calculations presented in Table 5.1, multiplied by four. For the same reason that for results of Table 5.1, percentages greater than 100% can occur.

We can see that the values in Table 5.2 approximately follow the same trend as the filling of time slots. That is to say the first months are well filled, but with a decrease in the last months of the semester, which is especially very important for the high cadences. Furthermore, the fractions in the two tables are close, meaning that the not used digital pointings are essentially the digital pointings of the not used time slots. The same values between both tables in the first months mean each observation is filled with four pointings. Also, for the last months, the values in Table 5.2 are slightly lower than those of Table 5.1, indicating a decrease in the filling of the observations.

Cadence	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
3 days	100%	101%	43%	42%	43%	47%
4 days	104%	104%	81%	61%	53%	60%
5 days	103%	108%	97%	82%	70%	71%
6 days	103%	109%	97%	97%	90%	84%

Table 5.2: Evolution of the filling of the possible numerical beams over 6 months. The ratio of filling is calculated relative to the mean number of possible numerical beams a month for the given cadence.

If we examine the global results of the simulations given in Table 5.3, the mean number of pointings by observation is close to four for all the different cadences, but with still a better filling for the cadence of six days.

Table 5.3 resumes the global result of the simulation at the end of the six months for each cadence. In terms of the use of observing time, the highest cadence of three days obviously allows one to make more observations with 740 observations, representing 34% more than for the lowest cadence of six days. However, we can notice that the efficiency of the use of the observing time is clearly lower for the cadence of three days rather than six days. In the case of the highest cadence, although there are more realized observations, 188 hours are finally unused. On the opposite side, with a cadence of six days, almost all the hours in the semester are used. This loss time is consistent with the expected behavior explained in Section 5.1.2, where certain sky bands are rapidly totally observed, and then saturated.

In terms of efficiency in observed pointings, for the cadence of three days, as there are more realized observations, there are obviously also more pointings that are observed. It results in a sky coverage 8% larger than for a cadence of six days. However, if we compare the covered sky relative to the used number of hours between these two cadences, we can remark that we cover just 30% of supplementary sky with 34% of observing time in addition. Thus, it results that all the observations are not filled with four pointings. It leads to a loss in efficiency with an average number of pointings done per observation decreasing from 3.92 in the case of six days to 3.79 in the case of three days.

These results conclude that for a high cadence of observation, the time window when pointings can be observed decreases very rapidly. Thus, in this case, to avoid wasting observing time, one has to plan precisely the decrease of the time window, especially at the end of the observing program where the window is short. Moreover, the filling of the grid is inhomogeneous, resulting to observe rapidly the

Cadence	(1) **# obs.**	(2) **Not used hours**	(3) **% of used time**	(4) **# pointings**	(5) **Obs. grid**	(6) **Avg. # pt.**
3 days	740	188	66 %	2804	36 %	3.79
4 days	665	81.5	80 %	2564	33 %	3.86
5 days	608	29	91 %	2361	31 %	3.88
6 days	551	3.5	99 %	2158	28 %	3.92

Table 5.3: Results of the simulations for the four tried cadences. (1) Number of observations carried out during the 6 months. (2) Total number of not used hours at the end of the 6 months. (3) Fraction of the total mock allocated time used. (4) Total number of pointings carried out during the 6 months. (5) Fraction of the grid which is observed at the end of the 6 months. (6) Average number of pointings of the realized observations.

“joker" pointings, i.e. the high declination pointings. The consequence is to have a scheduling less flexible. Furthermore, if the allocated time doesn't match exactly with the planned program, the probability to have holes in the sky coverage and/or time slots covering the same sky regions is higher. The last point is that, for a high cadence, even with a perfect match between the allocated time and the planned program, there is necessarily a loss in the efficiency of the use of the observing time.

Finally, the chosen cadence was the slowest, i.e. one series of observations every 6 days. With this observing cadence, we expect consequently to have linear progress of the observed grid, with better efficiency of the use of the observing time, and finally with more flexible scheduling.

5.2.3 Segmentation of the observing program

NenuFAR is a new telescope, which was officially inaugurated in October 2019. During one semester following the inauguration, the pressure on the telescope was relatively low, because of the fact that few projects were ready to use NenuFAR. Consequently, we expected to obtain easier observing time during this first semester. On the opposite side, because of the increasing pressure on the telescope, more difficulties were expected to obtain the wished time for the next semesters.

The necessary time to complete the blind survey is at least 961.5 hours (7 692 / 4 = 1 923 observations = 961.5 hours). That means that, even with the slowest cadence, the required time is about 250 hours per semester, with many constraints. Indeed, we need to block an entire night every 6 days, and this is during the entire semester. The required time is important, especially just for one project, and we didn't have the certainty to obtain this sufficient amount of hours to complete the program within the first two years. Moreover, because of bad observations or instrumental problems, some observations have to be repeated, resulting that the final real-time will be greater. Thus, it was essential to optimize the time allocated for the first semesters.

The size of a digital beam is actually defined as the loss in the gain of -3 dB (see Chapter 4.2.2), corresponding to a gain divided by a factor of 2. Consequently, with a digital beam, it is possible to observe a larger sky area, but with a large decrease in sensitivity beyond the determined size. Moreover, the positions of the pointings are placed in order to add an overlap of 63% of the beam radius between two neighbor pointings (see Chapter 4.3.4). Using this property, it is possible to have substantial coverage of the sky above 39° in less time. As a consequence, the observing program has been divided into two phases, where each phase consists to observe every second pointing. Thanks to this, all the sky can be observed in half the time, but with less sensitivity.

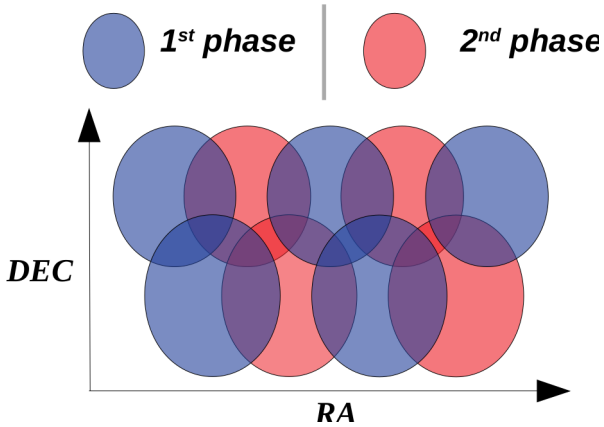


Figure 5.2: Schematic view of the organization of the pointing grid in two separated observation phases. Each ellipsis represents a digital beam of NenuFAR in the sky (in the celestial frame (RA-DEC)), corresponding to one pointing of the defined pointing grid. The first phase consisted of observing the blue beams (phase 1) during the first year of the observing program, and the red beams (phase 2) during the second year.

Also, this optimization was not just to fill the time slots, but also to try to rapidly cover the maximum amount of sky. To help us do that, we can use a constitutive element of the pointing grid. Indeed, the pointing grid was defined by adding an overlap between the neighbor pointings. In consequence, certain sky areas are redundantly observed.

Using this overlap, it is possible to avoid redundancies by dividing the whole program into two phases, consisting to observe every second pointing. Figure 5.2 shows a scheme of this segmentation, where the different pointings of the grid are drawn with an ellipsis representing the computed size of the numerical beam. Because the overlap rate was set to 63% of the beam radius, the sky coverage is still subtantial even for a single phase. The observing program was finally done in two successive phases: a first phase carried out during the first semesters to ensure relatively good coverage of all the sky above 39°; a second phase in the last semesters to complete the grid.

Figure 5.2 shows the principle of the observing program in two separated phases, where ellipses represent the determined area (i.e. corresponding to the gain attenuation of -3 dB). With this configuration, the first phase is interesting because it allows us to observe more than half of the sky with just half of the needed time.

However, the expected pulsars to search might be faint, making it difficult to detect it far from the center of the pointing. The second phase consequently allows observing the gaps between the pointings of phase 1 with good sensitivity. In addition, in the edges of the beams of phase 1, the sensitivity significantly decreases. With the used overlap, during phase 2, edges of phase 1 are observed a second time. Furthermore, if we combine the sensitivity loss with the effects of scintillation, it may be useful to observe these locations two times. Finally, the addition of phase two allows us to reach the aim of 98% of coverage of the sky above 39° with a gain attenuation better than -3 dB.

Moreover, with this configuration, the right ascensions to observe are less close. Then, in the case of consecutive nights of observation, the probability of having redundancies is lower. Also, the beam is indeed larger than the computed size with a first null of the gain located at approximately 2.5 radii. Thus, it will be possible to observe the entire targeted sky with only about 500 hours of observation in one year.

Concerning the second phase of the program, the utility is obviously to complete these “holes" of weak sensitivity left by phase one (the parts between the blue ellipses in Figure 5.2 where the sensitivity of phase one is very low). Then, this last phase firstly aims to detect pulsars outside of the beams of the first phase (i.e. with a gain attenuation lower than -3 dB). Although they are inside a beam of phase one, the power of the received signal can be faint because of the decreasing sensitivity of the beam. Thanks to the high overlap rate, a pulsar located on the edge of a phase one beam can indeed be closer to the center of a phase two beam. Thus, secondly, it may permit the detection of pulsars located on the edge of beams of phase one.

5.3 The observing program

5.3.1 Global summary

After some observation tests in March 2020 and July 2020, the observing program has been officially started in August 2020 and was planned for a duration of at least 2 years until August 2022. The observing times for NenuFAR are allocated by semesters which begin in June and November, leading observation during five semesters for the NPBS. The allocated times for the survey are for the different semesters:

- semester 1: 245 hours
- semester 2: 295 hours
- semester 3: 307 hours
- semester 4: 232 hours
- semester 5: 83 hours

In all, this represents 1 162 hours, distributed to the majority in nights of 9 hours between 21 h and 6 h UTC as asked. Taking into account that it is impossible to fill completely the allocated times due to the constraints previously explained, this corresponds to about 2 000 observations of 30 min to schedule, approximately 90 observations a month.

The algorithm of selection of the pointings has been coupled with a function to write automatically the configuration files used by the telescope to set up and carry out the observation. Furthermore, the whole algorithm is wrapped in a loop allowing to select iteratively, inside a contiguous time slot, the pointings to observe by incrementing half an hour the meridian transit time.

These 1 162 hours represent on average 5 nights of observation a month at a mean cadence of 6 days between two nights. Because of errors that can occur, the best is to schedule the survey night by night to have a permanent update of the remaining pointing grid to observe. However, for the scheduling of the observations with NenuFAR, there is a system of a broom wagon requiring to schedule at least 6 days before the day of the observation. There is then a delay time of at least one week, making the scheduling night-by-night less efficient.

Consequently, the program of iterative pointing selection and generation of configuration files was wrapped in another external loop. This second loop, based on a calendar with the dates and hour edges of the time slots, allows optimizing the time used for the scheduling, by generating all the configuration files for many days. Nevertheless, in order to have a relatively regular update on the remaining pointings, I chose to schedule observations on a monthly basis.

5.3.2 Phase 1 summary

The first observation of phase 1 was made the 6^{th} of August 2020, and the main observing program was carried out for 12 months until mid-July 2021 with 98.1% of the grid observed. To obtain observations of the entire sky above 39°, the remaining pointings have been observed in priority during the next months, with the latter observations done in August 2022. However, the program of phase 1 is not totally complete. There remain a group of 20 unlucky pointings, observed in June 2021 and rejected for too much RFI, then observed again in July 2021 and still rejected for too much RFI. They were observed once again in August 2022, but unlikely during a thunderstorm!

Figure 5.3 presents the evolution of the sky coverage month by month. The top panel shows an Aitoff representation corresponding to the progress of the pointings observed, where each color corresponds to a different month. The bottom panel shows in the first part the number of observations (blue bars) made in a month, and the number of bad observations rejected (red bars). The green line corresponds to the number of observed pointing (in good observations), with in the second part, the corresponding evolution of the fraction of the grid (for phase 1) which is done. Also, the characteristics of the observation of phase 1 are presented in Table 5.4.

In all, 1 209 observations have been done to observe 3 826 pointings, representing a filling of the allocated time of 94%. The number of realized observations has been approximately regular with between 70 and 140 observations a month, except in November 2020 and for the ending in July 2021.

Concerning the evolution of the number of pointings, we can see a decreasing trend, indicating a loss in efficiency (less than four digital beams per observation). This decrease can be explained by the fact that the unobserved sky area is getting smaller over time and further shifted towards the high declinations where the number of pointings is also decreasing. Thus, a large part of the last observations is done to fill the “holes" in the sky, i.e. for isolated pointings. Nevertheless, the average filling of the observations for phase 1 is 89%.

Finally, of these 1 209 observations, there were 115 bad observations, representing an error rate of 9.5%. An important contribution is especially provided by two months, representing more than half of the total: October 2020 with 29 bad observations and June 2021 with 28. The 29 bad observations of October are instrumental errors due to one night when the back-end machine UnDysPuTed did not start, and one night when the files were unreceived on the storage disk of the Nançay data center. Concerning the month of June 2021, the bad observations are due to thunderstorms at Nançay. The others are due to many broadband RFI many times during the observations, probably due to some lightning from distant thunderstorms.

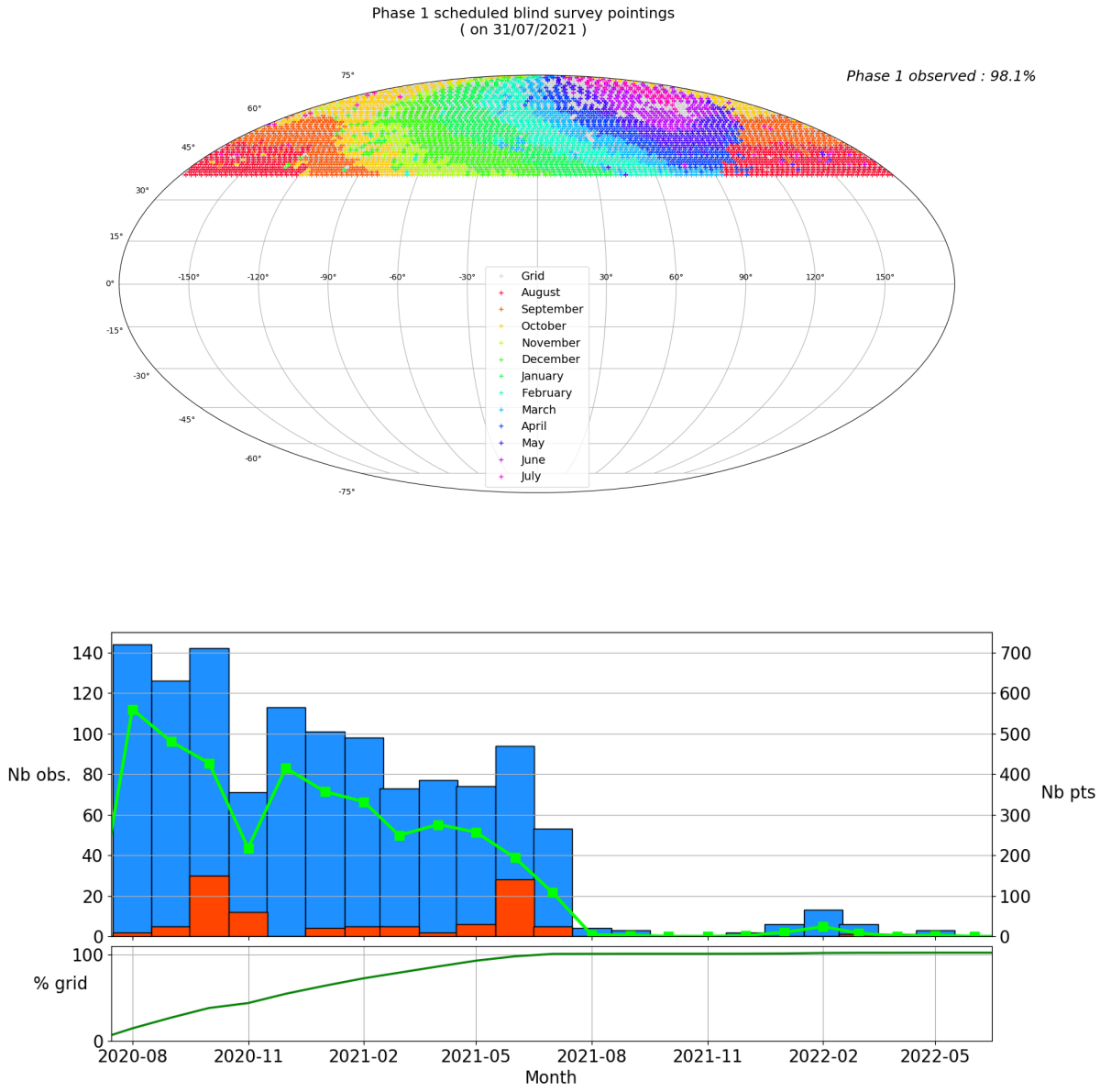


Figure 5.3: **Phase 1 observation progress** Top panel: evolution of phase 1 of the covered sky area month by month, where each color corresponds to one month. Bottom: histogram of the number of observations in blue bars, number of bad observations in red bars, and number of observed pointings in green. The fraction of the observed grid is represented below.

The second column of Table 5.5 indicates the number of bad observations for phase 1 as a function of the type of error. The fraction relative to the total number of bad observations is noticed in brackets. The three first types of errors are the presence of too many RFI in the dynamic spectrum of the observation, the identification of a thunderstorm, and lastly a problem related to the instrument, the back-end, or the storage. Also, we can remark that about a quarter of the bad observations are due to the rejection of the observation by the VCR of NenuFAR. This type of error was caused for circumpolar pulsars, which have a double transit at the meridian at their maximum elevation on the same day. In this case, the VCR was unable to determine the position of the meridian transit in the day. It was then unable to compute the elevation table for the tracking of the pulsar. This bug has since been solved.

5.3.3 Phase 2 summary

Phase 2 has begun just after the end of the main program of phase 1. The first observation was made the 7^{th} of July 2021, and the main program has been finished after 12 months of observations in mid-July 2022.

Figure 5.4 shows the same plots as Figure 5.3 with the evolution of the observed sky in the top panel, and the evolution of the number of observations, bad observations, pointings, and the progress of the filling of the grid for phase 2 this time. The corresponding characteristic of the observation of phase 2 are listed in Table 5.4.

At the 31^{th} of August 2022, 1 077 observations had been realized, allowing to observe 3 650 pointings. As we can see on the bottom panel of Figure 5.4, the number of observations carried out is fairly regular with between 70 and 100 observations a month, except for August 2021 where we have obtained an amount of time twice as asked, and November 2021 with an allocated time divided by two. Nevertheless, the filling of the allocated time is pretty good with an average filling of the time slots of about 97%.

Contrary to phase 1, the number of pointings done follows well the number of observations, meaning a good efficiency in the use of possible digital beams. Indeed, the filling of the observations is slightly better than for phase 1 with an average filling of 93% of the available beams.

In terms of bad observations, their number decreased by 23 relative to phase 1 with 92 observations lost. However, in proportion, the error rate is rather constant for the two phases with a rate of about 8.5%. A large part of the incurred errors in phase 2 occurred in July 2021 and June 2022, during the season of thunderstorms. In Table 5.5, we can see that almost 30% of the bad observations are due to a thunderstorm in the visible sky.

Also, as for phase 1, an important part, the majority of the errors indeed, occurred because of the problem of double meridian transit on the same day. The problem has been corrected at the beginning of 2022 by the team of NenuFAR in Nançay, and no new VCR rejection was noticed since.

Characteristic	**Phase 1**	**Phase 2**
Nb of observations	1209	1 077
Nb of pointings	3826	3 650
Alloc. hours filling	94%	97%
Obs. filling	89%	93%
Nb of bad obs.	115	92
Error rate	9.5%	8.5%
Observed grid rate	99.5%	94.9%
Remaining pointings	20	196

Table 5.4: Summarize of the different characteristics of the observing program for the two phases at the 31^{st} of August 2022.

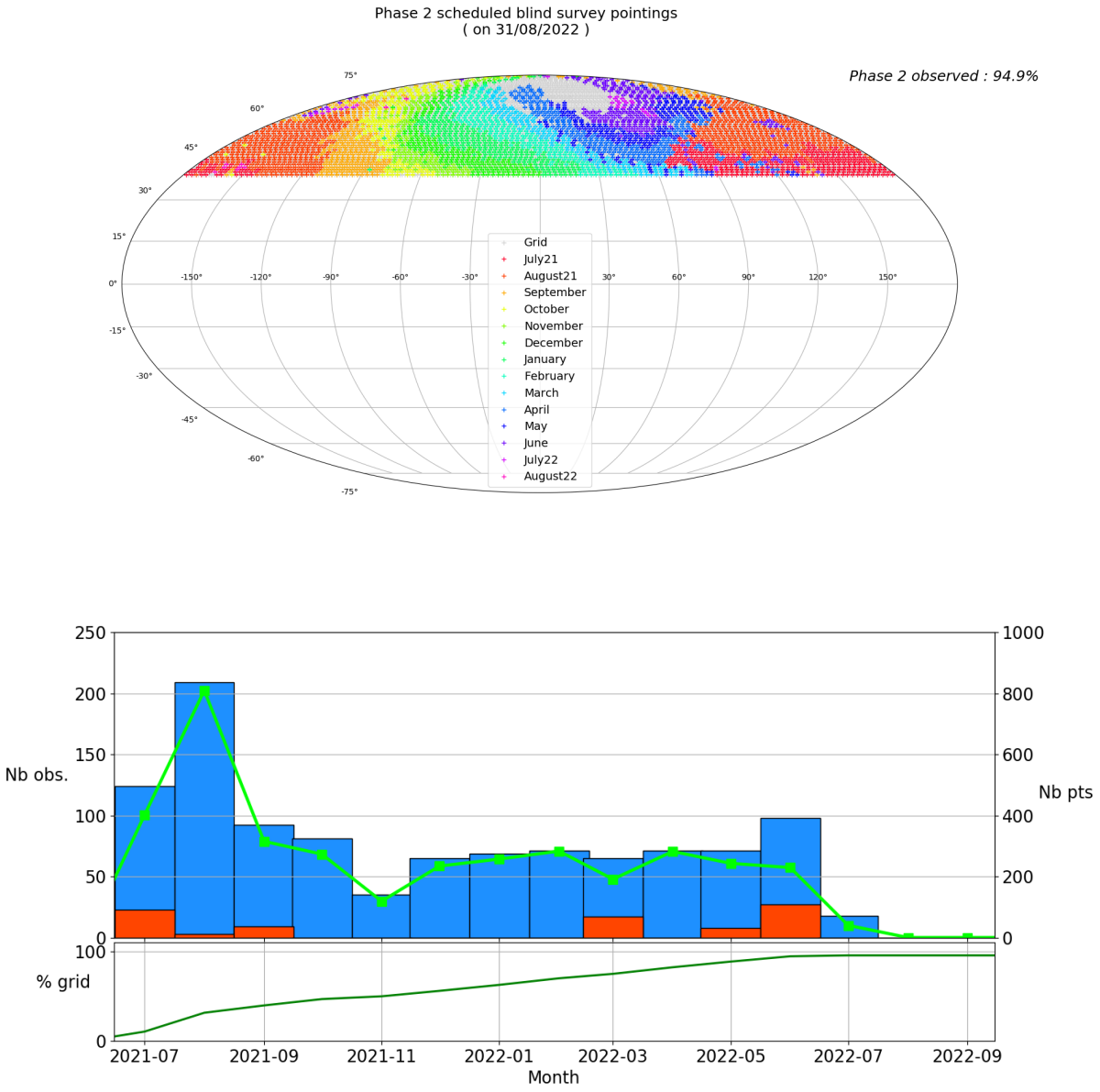


Figure 5.4: **Phase 2 observation progress** Top panel: evolution of phase 2 of the covered sky area month by month, where each color corresponds to one month. Bottom: histogram of the number of observations in blue bars, number of bad observations in red bars, and number of observed pointings in green. The fraction of the observed grid is represented below.

Table 5.5: Number of bad observations for the two phases of observations with the corresponding proportion relative to the different types of errors.

Type	**Phase 1**	**Phase 2**
Too many RFI	45 (42.86%)	16 (18.39%)
Thunderstorms	4 (3.81%)	26 (29.89%)
Instru. and oth.	29 (27.62%)	6 (6.90%)
VCR rejection	27 (25.71%)	39 (44.83%)
TOTAL	**105 (100.00%)**	**87 (100.00%)**

5.3.4 Planning of the remaining observations

At the 31^{st} of August 2022, 216 pointings remain to be observed to complete the observing program. Observational constraints being to observe the pointing only its transit at the meridian, and in between 21 h and 6 h UTC, these last pointings can only be observed during part of the year. Then, to plan the end of the observing program, we need to compute the time of year when the pointings have their meridian transit during the time slot 21 – 6 h UTC.

Of these 216 pointings, 20 pointings are from phase 1 and are comprised by pointings which observations have been rejected three times. These pointings are grouped and can be observed during summer 2023 between the end of April 2023 to the beginning of August 2023.

The majority of the remaining pointings concern thereby the finishing of phase 2. The resulting times of year show that we can divide into two parts the remaining pointings. The first one is composed of old pointings not observed during the first months of observation of phase 2 and can be thus observed before the end of the semester 2022. 40 pointings are in this case, and could be done in 3 nights of November: the 10^{th}, the 25^{h} and the 31^{st}. These nights have already been requested and obtained.

The second part is composed of pointings not observed during the spring of 2022. These pointings could not have been done due to three reasons: a cadence of observation a little bit too low during the fourth semester, one night at the end of May with a back-end problem, and finally two nights at the beginning of June with thunderstorms. This part of the remaining pointings is represented by the grey area in the sky view of Figure 5.4.

We can see that these pointings are grouped together, and furthermore at high declination, allowing to use of several digital beams, unlike isolated pointings. We can observe these pointings between mid-March 2023 to mid-April 2023, where all these 156 pointings can be done. Also, because of the fact they are grouped, it is possible to calculate the number of required nights. Taking account of the rate of filling of the allocated times and beams, we obtain 3 nights to complete phase 2, and thus the entire observing program of the NPBS.

Résumé du chapitre :

Pour pouvoir observer les 7 692 pointages dans les 960 heures planifiées, il est nécessaire de sélectionner les pointages permettant d'optimiser le temps d'observation. Afin d'obtenir la meilleure sensibilité, les pointages doivent observer avec une élévation maximale. En conséquence, les pointages sont observés durant leur passage au méridien, permettant d'avoir une relation directe entre l'élévation et la déclinaison.

Pour chaque plage horaire de 30 min, des groupes de quatre pointages sont sélectionné dans une bande de ciel de 2° autour de l'ascension droite correspondant au milieu de l'observation. De plus, dû au resserrement des lignes d'ascensions droites vers les hautes déclinaisons, celles-ci peuvent être observées sur de plus longues périodes de l'année. Le ciel est ainsi observé en démarrant par les plus basses déclinaisons possibles.

Afin de réduire au maximum la présence d'interférences radiofréquences (RFI for radio frequency interferences en anglais), les observations ont été réalisées uniquement de nuit sur la plage horaire 21 h – 6 h.

Pour observer l'ensemble des pointages, il est nécessaire d'avoir une cadence d'observation suffisamment grande. Cependant, une trop grande cadence entraîne l'observation des hautes déclinaisons très rapidement, laissant finalement des plages horaires peu utilisées vers la fin du programme d'observation.

Différents programmes d'observation avec quatre cadences régulières ont été simulés : une nuit d'observation tous les 3, 4, 5, et 6 jours. Pour définir la meilleure cadence, le taux de remplissage des plages horaires ainsi que le taux de remplissage moyen des faisceaux numériques disponibles ont été calculé sur six mois. Malgré une proportion de ciel observé plus faible au bout des six mois, il est apparu qu'utiliser une cadence de six jours permet de remplir presque intégralement les plages horaires disponibles tout en ayant le plus haut nombre moyen de pointages observés par observation effectuée.

Utilisant un taux de recouvrement assez important entre les pointages voisins, le programme d'observation de la grille a été divisé en deux phases observées successivement. La première phase observe un pointage sur deux, permettant d'observer la majorité du ciel en moitié moins de temps. La seconde phase observe les pointages restant, permettant d'obtenir la couverture du ciel voulue.

Les observations ont été faites sur cinq semestres pour un total d'environ 2 000 observations sur 1 162 heures. La phase 1 a débuté le 6 août 2020, et le programme principal s'est terminé mi-juillet 2021 avec 98 % des pointages de la phase 1 observés. Le reste des pointages ont été observés durant le reste de 2021 et durant 2022. Au 31 août 2022, 20 pointages restent à être observés, et sont planifiés pour le mois de mai 2023.

La phase 2 a commencé mi-juillet 2021, et au 31 août 2022, 95 % des pointages ont été observés. Les 226 pointages restants sont dus à de nombreuses nuits orageuses durant l'été 2021. Leur observation est planifiée pour le printemps 2023.

Chapter 6

Data processing

6.1 Method

6.1.1 Aim of the processing

The data employed for the blind survey are dynamic spectra. A dynamic spectrum is a two-dimensional type of data showing the measured flux relative to time and frequency of observation. The first step of the data processing is to prepare and clean this 2D raw data to obtain reduced scientific data which can be used for the pulsar search.

Once reduced data are obtained, we can carry out the search process to try to find the characteristic signal of a pulsar. To do this, the next parts of the processing process consist to apply a search method specific for pulsars, which uses the extreme periodicity of the pulsars. The search method is based on the Fourier transformation to isolate candidates presenting a periodic pulsation.

6.1.2 General process

The processing task is composed of several successive steps, where each step performs a specific action on the data allowing to detect a pulsar signal.

The first steps have for the objective to prepare the raw data for the real searching. This part is operated to remove the non-astrophysics signals. Furthermore, preparing the data is useful to reduce the flux variations in time or frequency, which are essentially due to the instrumentation and the scintillation. These effects reduce the signal to noise ratio of the searched signals.

The radio signal of the pulsars passes through the interstellar medium, which creates some diffusion effects on the measured signal. One of these is the dispersion of the signal relative to the frequency, producing a retardation of the signal as a function of the frequency. In consequence, on the 2D data, the expected signal is mixed and become hard to see. One needs to correct these dispersion effects to increase the signal to noise ratio of the signal to try to detect. The next steps of the processing are done to perform this work, and then facilitate the periodicity search.

Finally, on the dedispersed data, the last steps of the processing consist to search for periodicities on the obtained time series. Once periodic signals are found, we conclude the processing stage by selecting and characterizing the candidates which are possibly a pulsar. These best pulsar candidates are thus folded like a real pulsar signal to be checked during the post-processing stage.

6.1.3 Softwares

The search processing pipeline used for this survey is based on the software PRESTO for *PulsaR Exploration and Search TOolkit* (Ransom 2011)[1]. PRESTO is a suite of programs developed by S. Ransom, which is dedicated to the search of pulsars. The software is essentially written in C with some more recent parts in Python, allowing to be very efficient in the memory usage, even for the treatment of large data. Furthermore, PRESTO can be easily used, taking in input many various types of file formats as the standard formats in radio astronomy: PSRFITS and FILTERBANK.

PRESTO doesn't search for pulsars in the time domain as a fast folding algorithm does, but in the Fourier domain using FFT methods to identify periodicities in time series. The steps of the explained above processing pipeline follow the general process used by PRESTO to do the pulsar search.

PRESTO was chosen as base software due to this performance to find pulsars in the context of a survey. It was indeed used in the past for many pulsar surveys, including some prolific surveys as the PALFA realized with Arecibo, which has discovered 176 new pulsars according to the version 1.67 of the ATNF (Manchester et al. 2005)[2], or the low frequency survey LOTAAS of LOFAR (Sanidas et al. 2019) with 53 discoveries (always according to the ATNF).

The other parts of the pipeline were written in Python 2.7, in particular to optimize the processing time. The different PRESTO programs are encapsulated in multiprocessing structure, which are pretty efficient and simple to implement using the dedicated Python package. In the same way, some parts of PRESTO take many parameters in input. These parameters are not all the time constant, especially for the final folding of the candidates, and need to be correctly set.

Finally, two Python scripts have been created from scratch for the search processing pipeline of this survey: the flattening of the NenuFAR data and the computing of the dedispersion plan adapted to the constraints provided by the very low frequencies seen by NenuFAR.

6.2 Preprocessing

6.2.1 Data format

The data measured by NenuFAR are recorded using a specific observation mode called *dynamic spectrum* (Bondonneau et al. 2021), created by Ismaël Cognard (LPC2E), with the aim to transform the output raw data of the beamformer LaNewBA in a dynamic spectrum which can be used. These data are then recorded in a specific format SPECTRA which allows shaping the data on time and frequency as wished. The particular interest is to be able to increase the number of frequency channels above the standard value of 192 for 37.5 MHz of bandwidth. The increasing is done, after the actual observation, passing the time information towards frequency information, using the Fourier transform of time blocks of data in a frequency channel.

The SPECTRA files generated on *UnDySpuTed* are converted into smaller files in the format FILTERBANK. The advantage of this format is that it is a standard format in astronomy that can be sent in input data in the software PRESTO. Moreover, the FILTERBANK files are binary files very simple in their structure, composed of a header with the metadata of the observation, followed by a sequence of data values ordered as: a series of polarization, itself composed by a series of times, itself composed by a series of frequencies.

Furthermore, in the context of the survey, we will have a huge number of files to store and need to decrease as much as possible the size of each file. With this in mind, the data were written in the FILTERBANK 8-bit integers rather than in floats, in order to win a factor of about 4 in the global size. Also, the search is realized on the total flux, and then the polarization is useless. Consequently, the polarizations are summed, allowing to win an additional factor of 4. Thanks to these changes, the size of a file decreases from 56 GB from a starting SPECTRA file to a FILTERBANK file of 3.5 GB.

[1] https://github.com/scottransom/presto
[2] https://www.atnf.csiro.au/research/pulsar/psrcat/

6.2.2 Frequency and time resolution

One of the major problems for the observation of pulsars is the dispersion of pulses. This is especially the case at low frequencies where the differential time delay caused by the dispersion becomes very important, even inside a frequency channel.

For the inter-channel dispersion, an incoherent dedispersion method based on the shifting of time samples is used during the processing of the data. Note that for the intra-channel dispersion, a coherent method in the Fourier space is normally used to dedisperse the data (Hankins & Rickett 1975). This method needs to be done directly on the output raw data of the telescope, before the final channelization. However, in the case of a survey, the DM of the pulsar is unknown, making it impossible to correct the dispersion before the processing.

Starting from the dispersion relation for the signal, we can derive the relation between the intra-channel time delay Δt_{smear} and the frequency channel bandwidth $\Delta\nu$:

$$\Delta t_{smear} = 2\mathcal{D} \cdot \frac{DM \cdot \Delta\nu}{\nu^3} \tag{6.1}$$

With $\mathcal{D}$ the dispersion constant, DM the dispersion measure of the pulsar, and ν the center frequency of the channel. Using this relation, we can compute the intra-channel smearing for different DMs relative to the channel bandwidth. Figure 6.1 shows the smearing at the lowest frequency of the survey of 39 MHz.

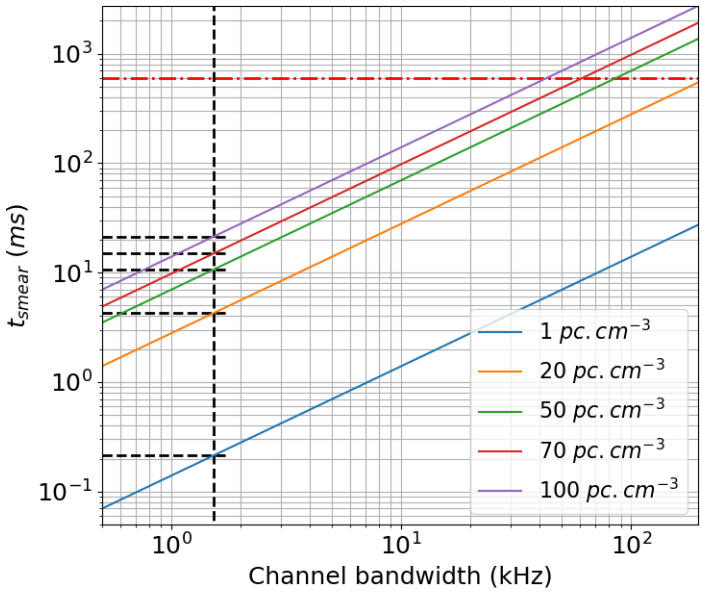


Figure 6.1: Evolution of the intra-channel dispersion at 39 MHz relative to the channel bandwidth for 5 different DMs. The black dashed lines show the frequency resolution of the survey of 1.52 kHz, and the red dashed line corresponds to the median of the period for the non-MSP pulsars.

In this graph, we can see the evolution of the smearing calculated for five different DMs: 1, 20, 50, 70, and 100 pc.cm^{-3}. The represented channel bandwidth range from 0.5 kHz corresponding to the total bandwidth of 37.5 MHz divided by 75 000 to 195.3125 kHz corresponding to the nominal frequency resolution (192 frequency channels).

We can see that for the nominal resolution, we have a smearing of 27 ms for a DM of 1 pc.cm^{-3}, and above 1 second for a DM of 50 pc.cm^{-3}. Moreover, the horizontal red dashed line positioned at 600 ms represents the median of the period of all the non-MSP pulsars of the ATNF catalog[3] (Manchester et al. 2005). Also, we can notice the smearing reaches this median period for a DM of just 20 pc.cm^{-3}.

[3] ATNF catalog version 1.67: http://www.atnf.csiro.au/research/pulsar/psrcat

Consequently, it is impossible to use the nominal frequency resolution, and we need to increase the frequency resolution. The choice was made to divide the standard number of frequency channels by a factor of 128 to reach 24 576 channels, corresponding to a bandwidth of 1.52 kHz. The black dashed lines in Figure 6.1 mark the different smearings for the chosen resolution. With 1.52 kHz, the intra-channel dispersion is largely decreased, with a smearing of about 4 ms for 20 pc.cm^{-3}and 20 ms for a DM of 100 pc.cm^{-3}.

Within the *dynamic spectrum* mode, we can increase the frequency resolution using a method based on the Fourier transform. Thus, the increase in the frequency resolution requires a decrease in the time resolution. Moreover, the nominal time sampling is 5.12 μs (the inverse of 195 kHz), and to avoid oversampling of the intra-channel dispersion, we can increase the time sampling.

With an intra-channel dispersion smearing of 20 ms for a DM of 100 pc.cm^{-3}(see dashed black lines in Figure 6.1), the time sampling can be largely increased. Consequently, in order to avoid oversampling at least for the low DMs, the time sampling for the observations of the survey has been set to 10.486 ms. According to Equation 6.1, the intra-channel dispersion smearing is smaller than the time sampling until a DM of 50 pc.cm^{-3}.

Furthermore, the large increases in the number of frequency channels considerably increase the size of the data. Also, this long-time sampling allows a significant decrease in the size of the files, preventing a too-long processing time.

6.3 Processing steps

6.3.1 Flattening of the data

NenuFAR is a phased array telescope whose tracking of the targeted sky area is done by moving the beam rather than directly the telescope as for a single dish telescope. Then, for NenuFAR, they are two beams: the analog and the digital, which need to be moved during the observation to track the intended pointing. Consequently, these tracking adjustments introduce regular variations in time, causing variations in the measured amplitude in the data (amplitudes are here integer values between 0 and 255 proportional to the voltage measured by NenuFAR).

A consequence is that the global data statistics are corrupted by these variations. Basically, the standard deviation is increased, and the median is moved relative to the real median of the data. Then, the SNR being calculated based on these statistics, the SNR of a potential pulse will be underestimated. Moreover, the detection of RFIs is based on the search for outlier parts of data, deduced using these statistics. Thus, these variations produce a double problem, and we need to correct them in order to try to improve the data statistics.

To develop and test the correction of the variations, two different types of data were used. The first was data from the blind survey which are dynamic spectra data characterized by a long-time sampling. The second was data coming from the observation of pulsars in globular clusters which are waveform data featuring a shorter time sampling. Figure 6.2 presents the time series integrated in frequency for two different observations: a blind survey observation of 30 min at the top, and a globular cluster observation of 1 hour at the bottom. In the two time series, we can clearly identify many periodic variations essentially towards the low amplitudes, corresponding to the tracking adjustments in question.

Characterization of the variations

On these two time series, the large variations which appear as vertical lines correspond to the tracking adjustment of the analogue beam. For tracking, the sky motion of this beam is carried out by applying a physical delay between the antennas and the mini-arrays. Thus, all the delays are not possible, and the beam follows a grid, which thereby generates a discrete pointing every 6 minutes.

Figure 6.2: Time series integrated in frequency showing the variations of the measured amplitude during observation for the blind survey in the top panel and for the globular cluster in the bottom panel.

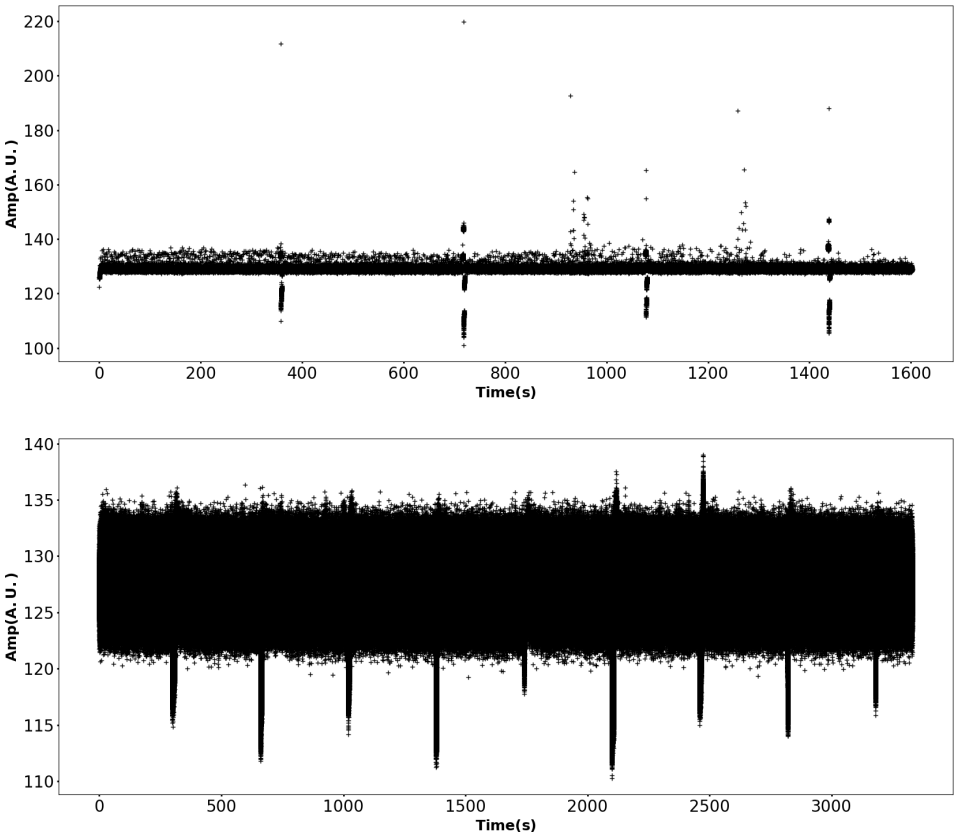


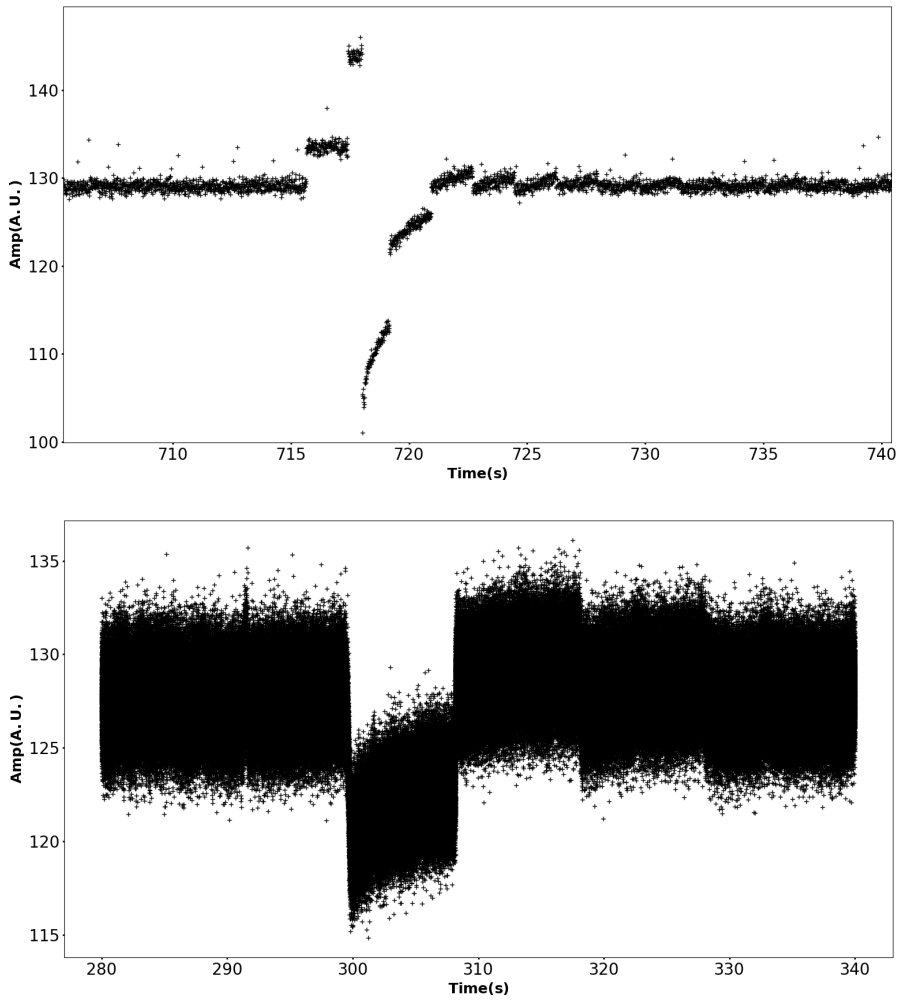
Figure 6.3 shows a zoom of the previous time series around one of the analogue beam tracking adjustments. We can see a clear statistically significant jump in the amplitude. Also, on the top panel, we can notice that the jump has a pseudo-logarithmic shape, consistent with the fact that for each tracking adjustment, the electronics must activate the new physical delays. This shape is then the electronic transitional regime of the telescope.

However, we can notice that the analogue jump doesn't follow exactly a logarithmic function, due to smaller and more frequent jumps in the amplitude. On the bottom panel, we can see that the time series after the jump in amplitude caused by the analogue tracking adjustment is segmented in blocks of data of 10 s.

Indeed, within the 6 minutes, the tracking is carried out by the steps of the digital beam, which are done numerically every 10 seconds. Then, these shorter jumps are due to the digital beam tracking adjustments which produces some little variations in the global time series. Furthermore, it appears that during the 10 s, the amplitude follows a slightly increasing trend, which creates a discontinuity at the new digital pointing.

Finally, a third type of variation can be found in the blind survey data. The data of the survey have a long time sampling, which has the advantage to average the time series, thus smoothing the digital tracking adjustments. This averaging also allows us to reduce the dispersion of the points of measure, and then to reveal smaller jumps of amplitude. Thanks to this, shorter jumps of about 1.2 seconds also appear in the time series. These have a smaller amplitude and can be identified on the top panel of Figure 6.3, especially inside the logarithmic increasing trend after the analogue tracking adjustment.

Figure 6.3: Time series showing one analogue tracking adjustment in the blind survey data in the top panel, and in the globular cluster data in the bottom panel.



Correcting method

To correct these different variations of the amplitude, one needs to subtract the trend of the time series from the amplitudes. The first obvious method to use is to divide the global time series into blocks of 6 mins and to fit with a logarithmic function the amplitudes. However, this logarithmic fit is a global fit over 6 mins, producing edge effects at the location of the discontinuities created by the shorter jumps. Certainly, the analogue discontinuities are pretty well reduced, but the other discontinuities due to the two other types of jumps are, conversely, increased. In consequence, a local fit is needed.

The best local method to absorb the different jumps could be to use a spline cubic fit, allowing an efficient local fit adjusted to the size of the block of data to fit. Unfortunately, with this method, a little error in the position of the knots of the function produces a bad fit of the block of data. It results from this bad fit some waves due to the fact that the points of the neighboring block corrupt the fit. This effect of waves is especially important during the phase of the rise of the amplitude just after the analog tracking adjustment. On the two panels of Figure 6.3, the shorter time variations are clearly visible at the beginning of the block of 6 mins. At this position, the shorter jumps are amplified by the global rising trend of the longer jump: the analogue tracking adjustment.

Thus, the spline cubic method is too much sensitive and complex to implement, to have a reliable flattening for the 7 692 pointings to process. An easier way, which is less precise but more robust, is to use a running average method.

Steps of the flattening method

To realize the flattening of the data of the blind survey, I have written a script in *Python 2.7*: FLAT_TIMES_SERIES.PY, to reduce and flatten the data before the RFI mitigation. Using the Python library FILTERBANK.PY of PRESTO created by S. Ransom, the script takes the raw filterbank file in input, and writes a new filterbank file with the flattened data. The script works in four steps, where the three first ones prepare the data to improve the global trend in order to facilitate the running average, and the last one performs the flattening of the data.

After normalization of the time series for each frequency to remove the bandpass variations, the first step search for the positions of the analogue jumps in the global time series integrated in frequency. The second step identifies the major RFIs in the frequency bandpass integrated in time, by searching the frequency channels with an amplitude larger than 3σ relative to the median of the bandpass. In this case, for each time series of the dynamic spectrum, the corresponding channel is set to the median of the time series. The third step is the same thing, but in the other dimension, i.e. the search of the major RFIs greater than 3σ in the global time series integrated in frequency. This first simple RFI mitigation, which removes the strongest RFIs, has two advantages: obtain better statistics for the running average and facilitate the work of the more precise RFI mitigation step (see Section 6.3.2).

Finally, once the interfering samples and channels are removed, we are able to compute a good smooth running average. Then, in the last step, for each frequency, the running average is determined by the convolution of the time series with a Gaussian window. Firstly, the Gaussian window allows to reduce the edge effects caused by the edges of the window. And secondly, a Gaussian function allows weighting the average, in order to smooth the impact of a discontinuity when we are close to the discontinuity.

Then, to obtain a good smoothing that doesn't increase too much the discontinuity for the close samples, it needs to have a-not-too large width of the Gaussian function. But, at the opposite side, too narrow Gaussian could smooth the time series and obscur the pulses of a potential pulsar. The standard deviation was set to 178 ms, representing a width of 34-time samples. Except for the very wide pulses, this is usually sufficiently large to preserve the pulses. Also, representing about 15% of the size of the shortest variations of 1.2 s, it is usually sufficiently small to follow the discontinuity pretty well. The other parameter is the window size, which was set to 1.845 s, corresponding to a deviation of 5σ relative to the center of the window.

Figure 6.4 shows the output plot of the flattening script FLAT_TIMES_SERIES.PY. This resulting plot summarizes the metadata of the file to flatten and shows six graphs allowing to compare the raw data before flattening on the left, with the data after flattening on the right. The plots in Figure 6.4 present the result of the script on a test observation of the pulsar B0355+54 with NenuFAR (the telescope line which indicates FAST is due to the fact that NenuFAR is untaken into account by TEMPO), realized in the globular cluster data format.

The top panels show the dynamic spectra of the data, where the flattened data are a little bit more homogeneous, particularly around 1 500 s where the dark vertical feature has disappeared. Moreover, some vertical patterns, unseen on the raw data, seem to appear around 75 - 80 MHz. These frequencies correspond to powerful known RFIs, which need to be removed during the RFI mitigation.

The middle panels represent the global time series integrated and averaged in frequency, where we can clearly see the analogue tracking adjustments on the left graph (exactly as in the right plot of Figure 6.2). After the flattening, it is true that the jumps each 6 min are still visible. However, statistically, they are largely reduced compared to the standard deviation of the time series.

Finally, the bottom panels plot the frequency bandpass integrated and averaged in time, where some strong RFIs are apparent around 75 - 80 MHz. After the flattening, the channels corresponding to these RFs are still clearly identified, but only because they are set to the median while the graph shows the mean. Moreover, these channels are generally entirely set to the median and are unimportant in the context of the search for pulses. However, they can strongly impact the RFI mitigation done in the next processing step.

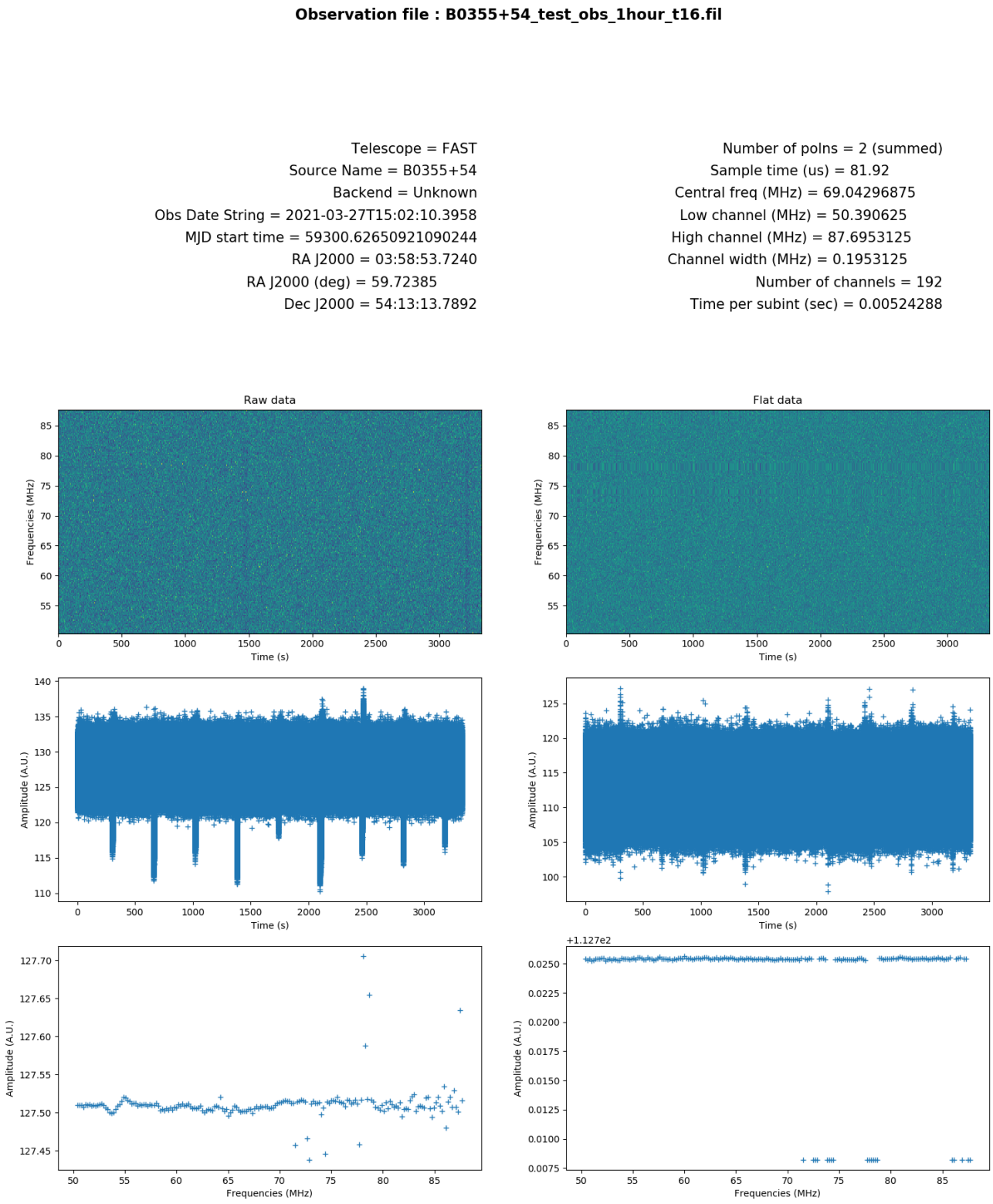


Figure 6.4: Output plot of the script FLAT_TIME_SERIES.PY showing the raw data on the left and the flattened data on the right. Top: dynamic spectrum, middle: time series integrated in frequency, bottom: frequency bandpass integrated in time.

6.3.2 RFI mitigation

In order to succeed in finding a pulsar, we need to reduce the data by removing the non-astrophysical signals present in the data, to try to keep only the signal of the pulsar. These other signals pollute the data, interfering with the pulsar signal, and sometimes obscure the less powerful pulsar signals. This second step of the processing pipeline consists then of searching these signals, the RFIs, and zapping them to be untaken into account for the next steps.

The signal of a pulsar is expected to be rather faint. Hence, the utility, and especially the importance, of the RFI zapping, lies in the fact that the RFIs are generally powerful compared to these signals of pulsars. Then, the first consequence of the presence of RFIs is that the mean, the median, and the standard deviation are increased. Also, the problem is that pulsar signals can be identified with their SNR, calculated based on these statistics. The presence of RFIs decreases the SNR of a pulsar signal which can be therefore missed by the search. The second consequence is related to the method used to find pulsars: the Fourier transform. In the case of a powerful RFI, even if it is not periodic or rare, it can increase the Fourier power of some bins in the Fourier transform. This RFI can be potentially considered as a good pulsar candidate rather than the real and maybe less powerful pulsar.

Because of these reasons, the objective of the RFI mitigation is to remove as much RFIs as possible, trying to tend to the ideal case where the data is only composed of the pulsar signal and white noise.

The RFIs can be identified by some characteristics and footprints in the statistics. The majority of the RFIs are non-astrophysical and are human-generated. These RFIs are then located on Earth or just around if due to satellites, and consequently don't undergo dispersion. The RFI search uses this specific characteristic relative to a pulsar signal, and the mitigation is thus carried out on the dispersed data at DM = 0 pc.cm^{-3}.

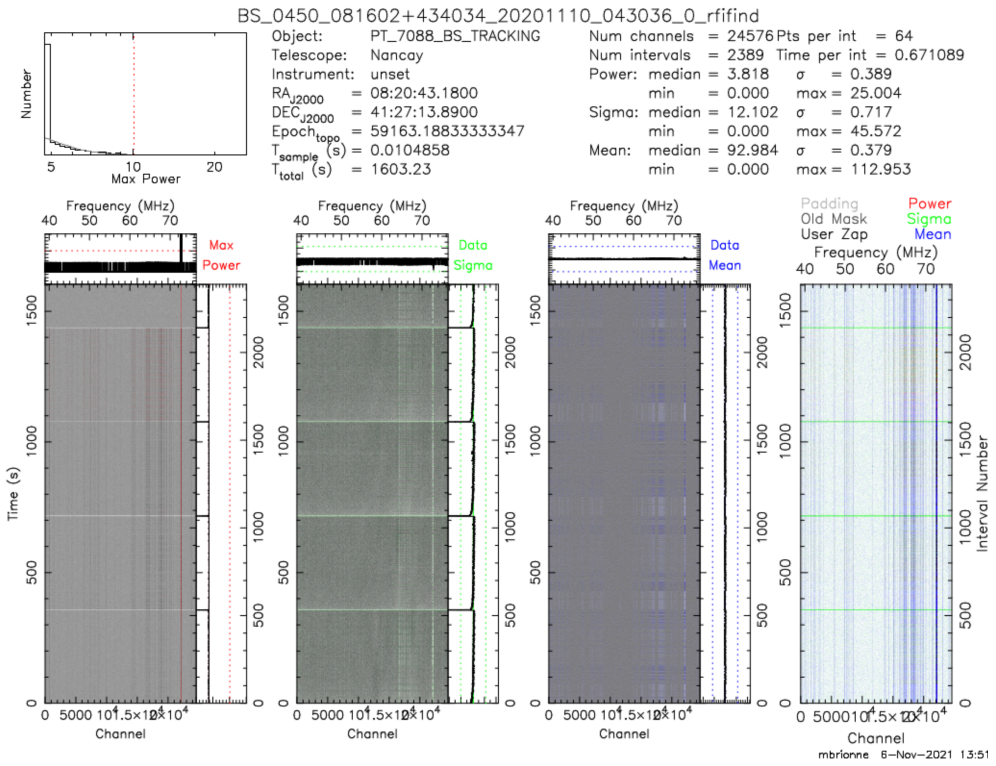


Figure 6.5: Output resulting plot of RFIFIND. Top left: distribution of the Fourier power in the observation, top right: information and statistics about the data. Bottom: dynamic spectra reporting different statistics about the chunks, where the zapped chunks are colored. From left to right: maximum Fourier power (red), standard deviation (green), mean (blue), and the last one presents the final global mask.

Moreover, the presence of RFI impacts the statistics relative to a clean expected signal. A frequency narrowband RFI which is continuous in time is characterized by a low standard deviation and a high median of the corresponding channel. If this RFI is intermittent in time, the median will still be high, but at the opposite of a continuous RFI, the standard deviation of the channel will be high. In the case of broadband RFI, whether it is continuous or not in frequency, the consequences are exactly the same, but in the corresponding time sub integration this time.

The RFI mitigation is realized using the program RFIFIND from PRESTO, which uses statistical methods to identify RFIs on different scales: the frequency channel, the time sub-integration, and the chunk of data in time and frequency.

For each chunk of data corresponding to the data in one time sub-integration and one frequency channel, the mean and the standard deviation are computed. The mean is then compared to the median of the means of the corresponding sub-integration and the median of the means of the corresponding channel. If the difference is larger than a given threshold, the chunk is marked as an RFI (blue parts of the dynamic spectrum in the third and fourth bottom plots in Figure 6.5). The same comparison is done for the standard deviation relative to the median of the standard deviations (green parts in the second and fourth plots). Finally, the maximum Fourier power of the chunk is computed and identified as an RFI in case of power greater than another specific threshold (red parts in the first and fourth plots).

The first parameter to set for the RFI mitigation is the size of a time sub-integration. The size of the block of data used to define statistics needs to be sufficiently large to obtain correct statistics, but not too large to lose the impact of a possible RFI signal. The FILTERBANK format doesn't have any sub-integration structure compared to the PSRFITS format. By default, the time sub-integration is fixed to 2 400 time samples. For high-frequency observations where the time sampling is generally around the tens of μs, the chunks are correct to do a good RFI mitigation. However, for our frequencies, we need to use a long-time sampling of about 10 ms, representing a chunk size of 24 s. This setting is excessively long to efficiently find an RFI of a few ms, with the drawback to lose 24 s of data each time an RFI is found. The value of the size of a chunk in PRESTO was set to 671 ms corresponding to 64 samples.

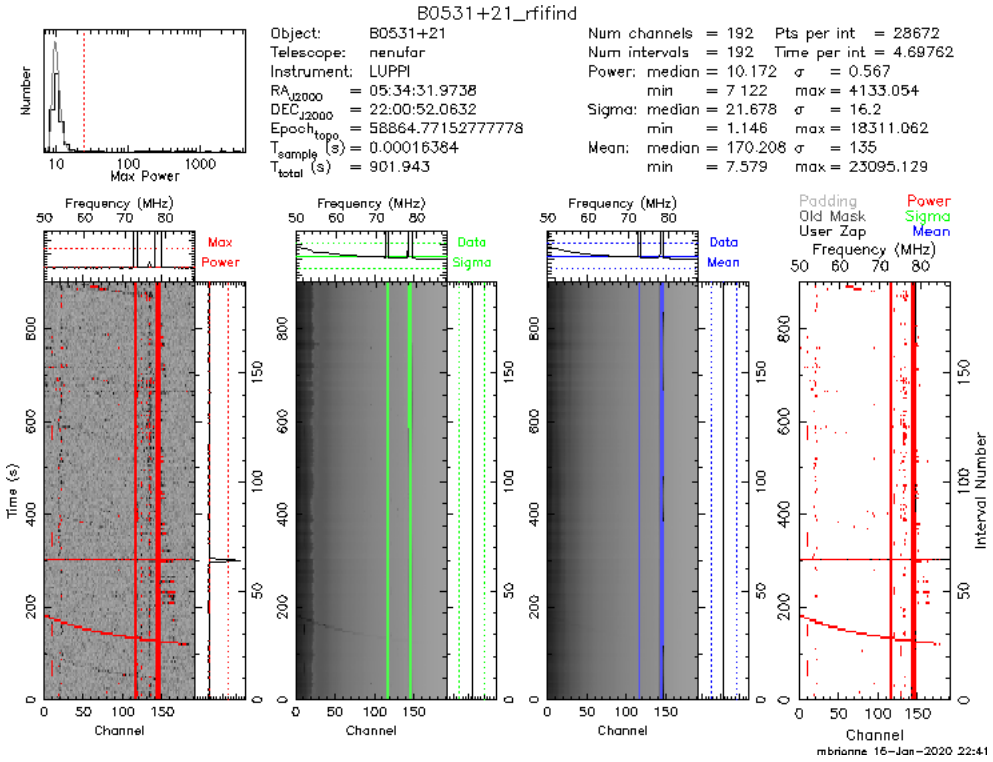


Figure 6.6: RFI mitigation of an observation of the Crab pulsar, where a strong pulse around 100 – 200 s is considered as an RFI due to a too-low Fourier power threshold.

The evaluation of the RFI is a function of two used thresholds: a first for the means and standard deviations, and a second for the maximum Fourier power. Thanks to the flattening carried out in the first stage, the amplitude variations are weak. This allows us to use reduced thresholds without the risk to remove some good parts of the data. Usually, the pulses of a pulsar are too weak to excercise a significant influence on the statistic of the data compared to the white noise. The first threshold was then set to 3σ, which permits the removal of the majority of the bad parts.

The second threshold concerns the third test, which is more sensitive to the presence of pulses. In the case of strong pulses, for a pulsar making giant pulses or a powerful RRAT for example, an overly lower threshold can identify some powerful pulses as RFIs. Figure 6.6 presents the example of an observation of the Crab pulsar with a very powerful pulse (not a giant pulse) flagged by the RFI mitigation. To avoid this, a high threshold of 10σ has been used. This test allows finding outlier bins in the dynamic spectra while the two first tests catch bad channels and bad sub-integrations.

A last parameter concerning the maximum proportion of RFIs is used by RFIFIND. Here, the standard parameter was kept. If more than 30% of a channel or sub-integration is marked as RFI, an overly important part of the channel or the sub-integration is lost. Performing a good periodicity search is therefore impossible, or at least, too corrupted. This one is then entirely considered as not sufficiently good to be used in the following stages and consequently entirely zapped.

A last type of RFI, which is especially important in the context of a pulsar survey, is birdies. The birdies are periodic or quasi-periodic non-astrophysical signals which are consequently, in the case of powerful birdies, prominent in the Fourier transform of the data.

As for the other RFIs, the birdies are generally terrestrials and then can be seen in the dispersed data at DM = 0 pc.cm^{-3}. Thus, to identify the birdies present in the data, we perform the Fourier transform of the time series integrated in frequency at DM = 0 pc.cm^{-3}. The Fourier bins showing a power greater than 10σ are qualified as birdies, and marked on a birdies list used by PRESTO for the next processing steps.

6.3.3 Dedispersion

The RFI mitigation allows bringing out the signal of the pulsar by removing the most RFIs. Nevertheless, the individual pulses are still weak, and their intensities can be too low to find the pulsar in the periodicity search. This one works on time series, then it is possible to increase the intensity of the pulses by integrating in frequency the dynamic spectrum, in order to increase the SNR of each individual pulse by a factor $\sqrt{N_{chan}}$, where N_{chan} is the number of channels.

However, to be able to sum the channels each other, it raises the problem of the dispersion of the pulses. Because of different retardation for each frequency, the summation of the channels will indeed mix the pulses on time. As a consequence, in order to correctly add the channels, the dispersion must be corrected. To do this, a method of incoherent dedispersion is used, consisting to shift the time series of a frequency channel by the number of time samples corresponding to the time delay generated by the dispersion. In the case where the DM is accurate, the dedispersion aligns the pulses of all the frequencies on the same sample index. In the final summed signal, the amplitudes of all the different pulses are therefore added to each other to increase the pulsar signal.

Standard dedispersion plan

The dedispersion step must thereby generate one time series for each tried DM. Consequently, a large number of dedispersion trials needs to be made, and a dedispersion plan (DD plan hereafter) should be created to optimize the computing time. To produce the DD plan, the used standard method is the *diagonal DM*, which takes into assumption a quasi-linear dispersion on the total bandwidth.

Thus, for this method, the DM step dDM corresponding to the minimum DM occurring a time delay of one time sample t_s is determined using the approximation of the dispersion relation of the pulsars. This DM step is then expressed by reversing the relation 6.1 but for the total bandwidth $N_{chan} \cdot \Delta\nu$ rather than for the channel bandwidth $\Delta\nu$.

$$dDM = \frac{1}{2\mathcal{D}} \cdot \frac{t_s \cdot \nu_c^3}{[N_{chan} \cdot \Delta\nu]} \tag{6.2}$$

With $\mathcal{D}$ the dispersion constant, and ν_c the central frequency. The diagonal DM DM_{diag} is then defined as corresponding to a diagonal time shifting of the pulse of $N_{chan} \cdot t_s$ on the dynamic spectrum.

$$DM_{diag} = \frac{1}{2\mathcal{D}} \cdot \frac{N_{chan} \cdot t_s \cdot \nu_c^3}{[N_{chan} \cdot \Delta\nu]} = \frac{1}{2\mathcal{D}} \cdot \frac{t_s \cdot \nu_c^3}{\Delta\nu} \tag{6.3}$$

In this method, the assumption of linearity expects that the smearing generated by the diagonal DM is one time sample for all the frequencies. Consequently, for the DMs greater than the diagonal DM, the time sampling is downsampled by a factor of two. A new diagonal DM is then determined with this new time sampling.

Using the diagonal DM method, the DM step for the survey is $5.91.10^{-3}$ pc.cm^{-3}. In consequence, to reach 100 pc.cm^{-3}, it needs 16 921 trial DMs for each pointing. The corresponding time is too large for the 7 692 pointings of the whole survey, and the DM step must be increased as much as possible.

Moreover, the first diagonal DM is evaluated at 145 pc.cm^{-3}, giving thus no downsampling for the DMs lower than 100 pc.cm^{-3}. The scattering being extremely important for higher DMs (see Figure 6.8), this survey doesn't expect to reach the region of high DMs. The DD plan can be summarized as a unique line and is hence not a real DD plan able to use.

Furthermore, for a DM of 145 pc.cm^{-3}, the dispersion time delay between two neighbor channels is at the lowest frequency of 31 ms, representing about 3 time samples, while the delay is just about 1 time sample (10.5 ms) for the central frequency, and less than the half (4.19 ms) at the highest frequency. The time sampling is therefore largely exceeded in the lowest frequency, and is far from the time sampling for the highest frequency. Contrary to the higher frequencies, at the NenuFAR frequencies, the ratio of the dispersion delays between the edges of the bandwidth is 7.4, when this ratio is, for example 2.8 for a survey between 1.2 and 1.7 GHz. As a consequence, designing the DD plan by the definition of the downsampling relative to the diagonal DM is not efficient for the frequency range of NenuFAR, and the standard DD plan must be adapted.

Adaption of the dedispersion plan

The dedispersion step was done using PREPSUBBAND from PRESTO, which takes as input the filterbank file, performs an incoherent dedispersion for all the given DMs, and finally generates one binary file containing the dedispersed time series for each DM. The particular interest of PREPSUBBAND is that it allows to carry out of dedispersion for a group of contiguous DMs rather than a single DM, reducing the computing time compared to a group of single dedispersion. Moreover, it also allows optimizing the dedispersion step, in addition to the downsampling, by dividing the dynamic spectrum into frequency subbands (hence its name), where a subband is a group of channels with a difference of dispersion smearing lower than one time sample.

An adapted DD plan was constructed based on the real dispersion smearing to compute the optimized parameters. To accelerate the computing time, the step was multi processed, allowing to divide and distribute groups of contiguous DMs on each processor. For each group, the parameters are calculated relative to the highest DM and at the lowest frequency, allowing also to obtain a finer determination of the best-optimized parameters.

The first parameter to compute is the downsampling n_s to use for the group of DMs. It is determined by the dispersion smearing in the lowest frequency channel.

$$n_s = \frac{\mathcal{D} \cdot DM}{t_s \cdot \nu_0^2} \cdot \left\{ 1 - \frac{1}{(1 + \Delta\nu/\nu_0)} \right\} \tag{6.4}$$

With ν_0 the lowest frequency of 38.966 MHz, and $\Delta\nu$ the channel bandwidth. Once the minimum downsampling was obtained, the number of channels n_c contained in a subband can be defined. Starting from the lowest frequency, the number of channels corresponds to the bandwidth for which the dispersion delay is lower than the time sampling $n_s \cdot t_s$.

$$n_c = \frac{\nu_0}{\Delta\nu} \cdot \left\{ \left(1 - \frac{n_s \cdot t_s \cdot \nu_0^2}{\mathcal{D} \cdot DM} \right)^{-\frac{1}{2}} - 1 \right\} \tag{6.5}$$

Figure 6.7 illustrates the needed time sampling (in blue) and subband frequency bandwidth (in red) relative to the DM for the observing parameter of the survey. The curve of the time resolution $n_s \cdot t_s$ is given by the equation 6.4, and the subband frequency bandwidth $n_c \cdot \Delta\nu$ which follows the equation 6.5. For the frequency resolution, the two curves correspond to the bandwidth for the cases of downsampling of one or two time samples.

Definition of the global parameters

Maximum DM At the frequencies of NenuFAR, the diffusive effects are very important. Also, due to the narrow frequency channel bandwidth, the intra-channel dispersion is negligible, and the major effect to take into account is the scattering (see Chapter 1.3.2). To construct the DD plan of the NPBS, the relation of Bhat et al. (2004) (recall here) has been used to estimate the scattering characteristic time τ_s.

$$\log \tau_s = -6.46 + 0.154 \log(DM) + 1.07 \log^2(DM) - 3.86 \log\left(\frac{\nu}{\nu_0}\right) \tag{6.6}$$

Here, ν_0 is a reference frequency. The coefficients of this empirical law were certainly determined based on higher frequency observations, and there is a large dispersion of the measured τ_s values. Nevertheless, we can use it to roughly estimate an average value of the broadening caused by the scattering. Using this relation, we can compute this average value of broadening τ_s relative to the DM for the two edge frequencies of the survey. Figure 6.8 shows this evolution, where the blue line corresponds to the lowest frequency of 39 MHz, and the orange line to the highest frequency of 76 MHz.

Indeed, the smearing time generated by the scattering is greater than the time sampling from 20 pc.cm^{-3} at the lowest frequency of 39 MHz, and for a DM of 42 pc.cm^{-3} at the highest one of 76 MHz. The smearing even goes over the second on the majority (and rapidly on the totality) of the bandwidth for the DMs higher than 100 pc.cm^{-3}. This consequently gave us a strong constraint on the maximum DM to reach to have an efficient search.

In Figure 6.8, the dashed red line shows the value of 600 ms, representing the median of the periods of the pulsars in the ATNF. The dashed black lines correspond to the DM of 70 pc.cm^{-3} and its associated smearing. From 70 pc.cm^{-3}, the mean smearing is greater than 600 ms, meaning an important probability that the scattering of the potential pulsar be greater than its period. As a consequence, beyond 70 pc.cm^{-3}, the majority of the pulsars have their pulse spread at a time greater than one period, giving impossible detection.

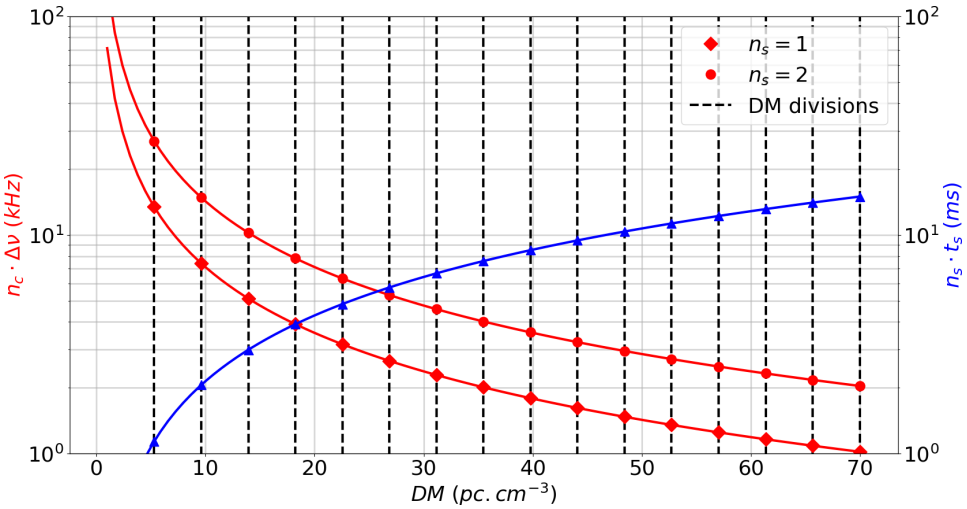
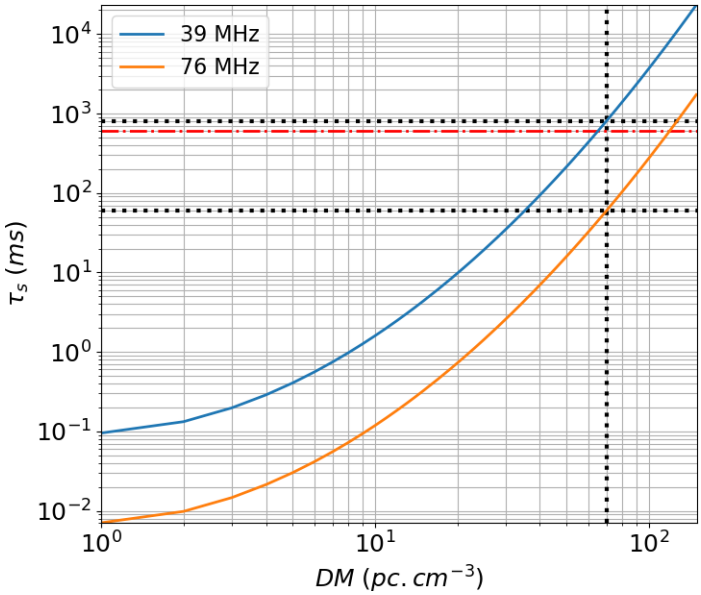


Figure 6.7: Subband frequency bandwidth $n_c \cdot \Delta\nu$ (in red) and time sampling $n_s \cdot t_s$ (in blue) needed by the NenuFAR pulsar blind survey relative to the DM. The red line with the lozenges (bottom line) corresponds to the frequency bandwidth required for a downsampling n_s of one time sample, and the line with the circles (top line) is for two time samples. The black dotted lines represent the DM segmentation of the DD plan used for the survey.

Figure 6.8: Evolution of the expected average scatter broadening as a function of the DM for the two edge frequencies of the survey. The black dashed lines show the maximum DM reached by the survey of 70 pc.cm^{-3}, and the red dashed line corresponds to the median of the period for the non-MSP pulsars.



This result was confirmed by the NenuFAR pulsar census, which has just detected two pulsars beyond 70 pc.cm^{-3}. The search beyond higher DMs will be thereby certainly poor in pulsar detection for a long time to dedisperse. The search for high DMs should probably be thereby inefficient. In addition, the computing time of the processing to reach a DM of 70 pc.cm^{-3} is already expected to be very long. As a consequence, the maximum DM to be searched has been set to 70 pc.cm^{-3}.

DM step The second starting parameter to define is the DM step. The minimum DM to obtain a dispersion delay of one time sample is 5.22×10^{-3} pc.cm^{-3}. But this DM step results in overmuch DMs to try, and then an overly long time to perform the dedispersion step. It is then required to increase it as much as possible. In December 2020, 96 test observations have been done in the real observing conditions, i.e. containing pointings without pulsars, and in some cases, pointings with a pulsar not necessarily in the center of the beam. To define the largest possible value of the DM step, the sensitivity of the search relative to the DM was tested trying to detect the weakest detectable pulsar in the test set: J2337+6151.

The processing pipeline was applied to the observation of J2337+6151 for various decreasing DM steps until the pulsar was found. Starting from 0.1 pc.cm^{-3}, the pulsar has been finally detected for a DM resolution of 0.01 pc.cm^{-3} in the third position of the candidates with a significance of 8.2σ. For this DM step, the pulsar was found on seven consecutive DMs. That means that for 0.02 pc.cm^{-3}, it should again appear in three consecutive DMs, which represents the limit to consider as a potential pulsar signal. The minimum DM step able seems to be 0.02 pc.cm^{-3}. However, including a margin of error because of the variability caused by the scintillation, the scattering of the pulsar, the position of the pulsar relative to the center of the beam... I have chosen to reduce the DM step, which has been set to 0.01 pc.cm^{-3}. Note that this is only a factor two above the recommended 0.005 pc.cm^{-3}. This implies trying a lot of DMs and consequently disposing of large computational power.

Dedispersion plan of the NPBS

With these defined global parameters, 7 000 DMs must be computed. The pulsar with the lowest known DM has a DM of 2.38 pc.cm^{-3}, resulting in the removal of DMs between 0 and 1. The dedispersion step was sent on 16 CPUs, and after removing this too-low DMs, allows dividing the DD plan into groups of 431 DMs. Table 6.1 presents the final DD plan, allowing to try 6 896 DMs, computed for the dedispersion step dDM equal to 0.01 pc.cm^{-3}. We can note a weak evolution of the downsampling n_s with one time sample up to 48.41 pc.cm^{-3}, and just two time samples until the end at 70 pc.cm^{-3}. On the opposite

Table 6.1: DD plan used for the dedispersion step of the NenuFAR pulsar blind survey. (1) Minimum DM of the group of DMs trials. (2) Used DM step. (3) Used downsampling. (4) Number of frequency subbands used to compute the dispersion shifting to perform. (5) Number of DMs to try in this group.

(1) $\mathbf{DM}_{min}$ (pc.cm^{-3})	(2) **dDM** (pc.cm^{-3})	(3) $\mathbf{n}_s$	(4) $\mathbf{N}_{sub}$	(5) $\mathbf{N}_{DM}$
1.00	0.01	1	3072	431
5.31	0.01	1	6144	431
9.62	0.01	1	8192	431
13.93	0.01	1	12288	431
18.24	0.01	1	12288	431
22.55	0.01	1	24576	431
26.86	0.01	1	24576	431
31.17	0.01	1	24576	431
35.48	0.01	1	24576	431
39.79	0.01	1	24576	431
44.10	0.01	1	24576	431
48.41	0.01	2	24576	431
52.72	0.01	2	24576	431
57.03	0.01	2	24576	431
61.34	0.01	2	24576	431
65.65	0.01	2	24576	431

side, the number of subbands N_{sub}, corresponding to the number of divisions of the total bandwidth, is large from the DM of 1 pc.cm^{-3} and evolves fastly with the DM.

These evolutions are due to our initial choice of time and frequency resolution. Modifying the downsampling afterward is effectively not really interesting because the major smearing is due to the scattering. Also, setting the downsampling relative to the scattering time scale yields a too-quick increase in the time sampling. Furthermore, the estimation of the scattering presented in Figure 6.8 is just a statistical mean value. Hence, with a too-long time sampling, there is a risk to lose weakly scattered pulsars. The scattering causes a broadening of the pulse independently of the used bandwidth. The subband bandwidths in the DD plan were thus, contrary to the downsampling, not chosen according to the scattering. The subband bandwidths for the survey have been indeed defined in order to neglect the intra-channel dispersion.

For the lowest group of DMs from 1 to 5 pc.cm^{-3}, the dispersion delay on the total bandwidth reaches from 2 s to 10 s. Thereby, it already needs 3 072 subbands of just eight channels, representing a bandwidth of 12.16 kHz. This bandwidth is 16 times smaller than the nominal frequency resolution of about 195 kHz, showing the chosen initial resolution is not overestimated. Moreover, the fast increase of the number of subbands with the DM needs 24 576 subbands from 22.55 pc.cm^{-3} corresponding actually to a subband of one channel. For this DM, the time delay is too much important, that the smearing in a bandwidth of just two channels is already greater than one sample. The dedispersion must thus be done channel by channel without any possible optimization until the DM of 48.41 pc.cm^{-3} where the time sampling can be increased.

6.3.4 Periodicity search

Search methods

At the end of the dedispersion step, we have 6 896 time series integrated in frequencies, which can be used to begin the real search. For this step, the search method uses the fundamental characteristic of the pulsars, which is the periodicity of their signal. The more the signal is well dedispersed more it looks like a Dirac comb. These periodic signals can be therefore found in the Fourier domain using the discrete Fourier transform (DFT).

Thereby, for each DM, the DFT of the time series is computed using a fast Fourier transform algorithm (FFT). The interest of the previous dedispersion step is that, the closer to the exact DM value narrower the integrated pulse. Moreover, the narrower the pulse greater the amplitude of the pulse. As a consequence, the Fourier power of a periodic pulsar signal must increase when the DM is closer to the exact value.

The periodicity search was carried out using REALFFT from PRESTO. It performs a real FFT, consisting to compute only the positive Fourier frequencies to avoid the redundancy of the symmetric negative Fourier frequencies. The output of REALFFT is a binary file containing the values of the FFT normalized in SNR, which can be used to identify the major frequencies.

Before searching for the periodic candidates, it needs to remove the false candidates which are the birdies. In the case of narrow-band birdies, they can actually appear in many DMs rather than just at DM 0. The list of birdies generated during the RFI mitigation step is used to zap these false powerful periodic candidates.

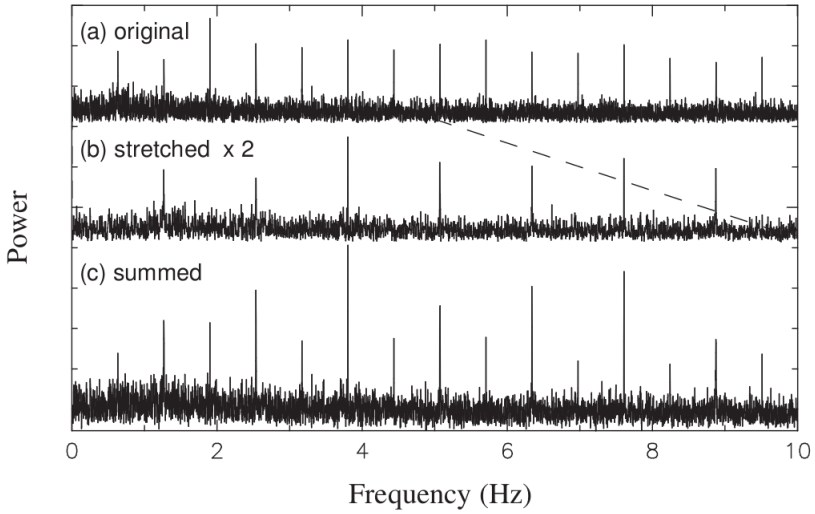
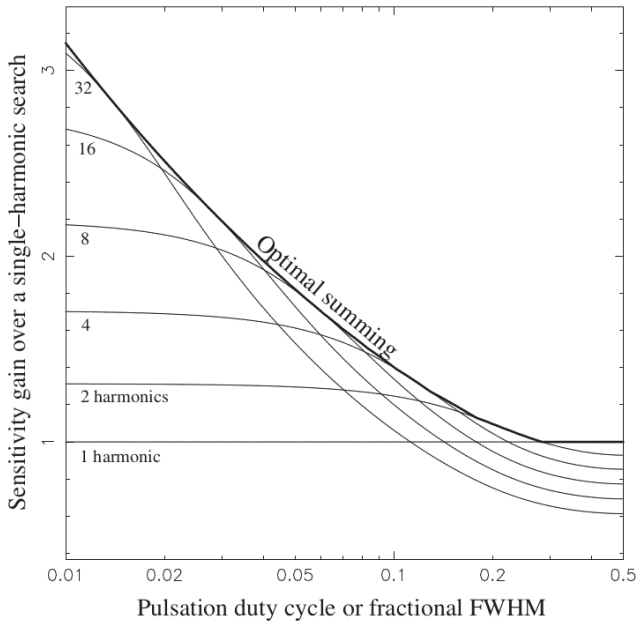


Figure 6.9: Illustration of the incoherent harmonic summing method where half of the initial Fourier series (top panel) is stretched (middle panel) to be finally summed with the initial one (bottom panel) (Lorimer & Kramer 2012).

The Fourier transform allows finding the fundamental period of a periodic signal, but it can be also found some harmonics of this signal. To detect more easily the weak Fourier signal of a pulsar, PRESTO uses this property using the *incoherent harmonic summing* method (Taylor & Huguenin 1969). The method, presented in Figure 6.9, consists to stretch half of the FFT series and to sum this one with the initial complete FFT series. The harmonics are summed each other, while the noise is increased by only a factor $\sqrt{2}$. The process can be iterated to increase significantly the SNR of the searched signal.

In the time domain, the pulses appear as a Dirac comb spaced by the period P, where each Dirac delta is indeed a top hat function with a certain width W corresponding to the width of the pulse. In the Fourier domain, the signal appears thus as a Dirac comb spaced of 1/P inside an envelope function described by a cardinal sinus of width 1/W. The number of harmonics able to be seen is equal to P/W, corresponding to the inverse of the duty cycle of the signal. Figure 6.10 shows the evolution relative to the duty cycle of the gain of the incoherent harmonic summing method compared to the standard search

Figure 6.10: Sensitivity ratio of the different number of summed harmonics relative to the sensitivity without harmonic summing (1 harmonic line) function of the duty cycle of the signal. (S. Ransom)



without. The number of iterations linearly increases the computing time, and it is then useless to sum more harmonics than W/P.

However, because of the strong scattering, the range of pulse width is very large. In consequence, between a weakly scattered long-period pulsar whose duty cycle is very small and a strongly scattered short-period pulsar whose duty cycle is almost 1, the possible duty cycles are spread over the full range, from 0 to 1. Combined with the expected weakness of the Fourier signal, we decided to set the maximum number of harmonics to sum to 32.

Searched periods

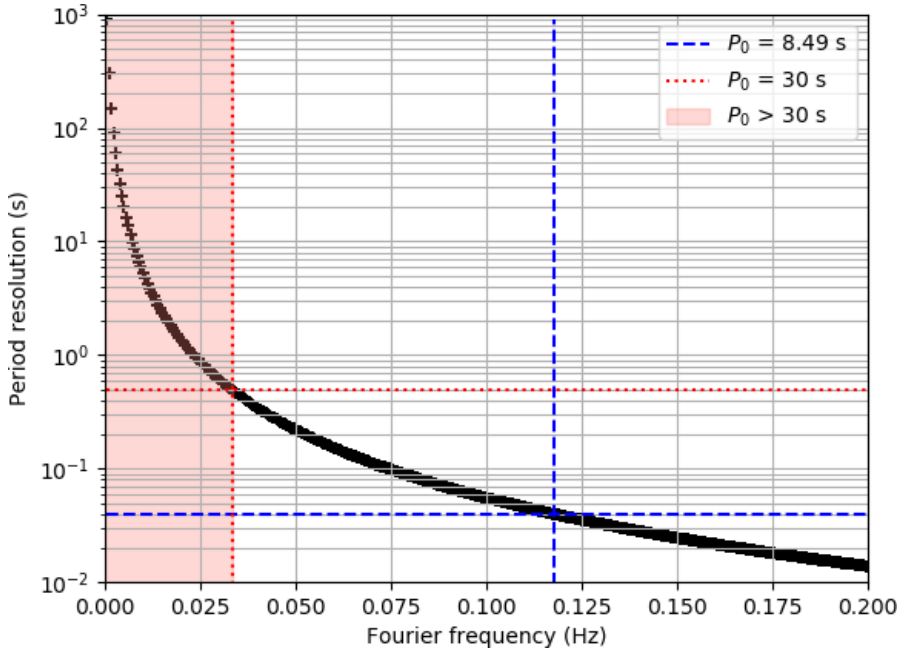
Periods able to be searched by the survey are determined by two constraints. The first one is the dispersion which defines the lowest available periods. The dispersion requires a high-time sampling, which is set to 10 ms, corresponding for the DFT to a Nyquist frequency of 50 Hz. The periods less than 20 ms are thus impossible to detect. Moreover, the average scattering is of about some dozens of ms from the DMs $\sim 30 - 40$ pc.cm^{-3}, giving the pulsars with a period less than $70 - 80$ ms generally very difficult to detect.

The second constraint concerns the highest periods and is due to the sensitivity of the DFT at low Fourier frequencies. The Fourier frequency resolution being linear, i.e. constant, there is a loss in the corresponding resolution in the period. The resolution in period δP_i of the DFT can be estimated by the following relation:

$$\delta P_i \;=\; \frac{1}{f_i} - \frac{1}{f_{i+1}} \;=\; \frac{N_s \cdot t_s}{i^2 + i} \tag{6.7}$$

With N_s the number of time samples, and i the bin of the Fourier frequency which is defined as $f_i = i/(N_s \cdot t_s)$. Figure 6.11 shows the evolution of the difference of the probed periods between two neighbor Fourier frequencies in the region of the low Fourier frequencies. The dashed blue line marks the Fourier frequency where the difference in period becomes larger than two time samples. The error in the period can consequently lead, for pulsars with a larger period, to a drift of more than one time sample by period. This frequency corresponds approximately to $1/(t_s \cdot \sqrt{N_s})$, which gives a period of 8.49 s for the survey.

Figure 6.11: Period resolution of the Fourier transform relative to the Fourier frequency. The blue dashed line corresponds to a period of 8.49 s, representing a period resolution of two time samples. The dotted red line identifies the maximum period of 30 s reached by the survey, and the pink area marks the excluded periods.



Furthermore, the red noise present in the low frequencies of the Fourier transform requires sufficiently powerful pulses to be correctly identified compared to this red noise. But, for pulsars featuring a long period, there are obviously fewer pulses during the observation, leading to a lower total flux than faster pulsars.

However, a pulsar with a long period can have a larger pulse width and therefore can absorb a drift in the period, i.e. a larger error in the period. In consequence, it has been decided to still consider a longer period than 8.49s, and the range of periods was finally set from 30 ms to 30 s. Indeed, for a period of 30s, the resolution in the period is equal to 491.80 ms (dotted red line in Figure 6.11), representing merely 1.6% of the period.

Identification of the candidates

The identification of the major frequencies is made using ACCELSEARCH, which is a program of PRESTO performing an acceleration search in the FFT files previously generated. FFT files being normalized in SNR, the possible candidates are identified by comparison with a given threshold.

The acceleration search is useful in the context of a search of binary pulsars. The orbital motion of the pulsar results in a drift of the pulse with time, spreading the profile. If not taken into account, this is also perceived as an apparent variation of the period of the pulsar, producing a drift in frequency within the Fourier transform proportionally to the acceleration of the motion along the line of sight.

However, the quasi totality of the binary systems is MSPs, whose vast part is not in the range of periods able to be reached by the survey. Moreover, the search in acceleration linearly increases the computing time with the amplitude of the scanned drift. Consequently, no acceleration search is done, corresponding to the search of only the signals without any drift of Fourier bins, leading to a search of only the single pulsars. The acceleration search will be eventually carried out in a larger future reprocessing of the data of the survey to include the binaries.

For the current search, ACCELSEARCH identifies the Fourier bins without any drift, which have a Fourier power greater than a given value normalized in SNR. At the frequencies of NenuFAR, i.e. of the survey, the strong scattering results in a broadening of the pulse, reducing thereby the Fourier power of the pulsar signal. Moreover, a large part of the pulsars observed below 100 MHz seems to present a spectral turnover, reducing the measured flux of the pulsar and consequently its Fourier power.

Therefore, many reasons can produce why the Fourier power of the signal of a pulsar is weak. To avoid missing an eventual weak and scattered pulsar, the threshold used to consider a signal in the FFT as a potential candidate has been set at a low value of 2σ.this step is indeed just a search of the period. The further steps will perform a finer search allowing us to see if the signal is really a pulsar signal.

6.3.5 Candidate sifting

The periodicity search step finds plenty of candidates for each DM, and we need to combine all the candidates to sift the good and the bad candidates. To perform this, the Python script ACCEL_SIFT.PY from the Python library of PRESTO is used.

In the case of a little error on the DM of the pulsar, its pulse is spread, but can still be detectable with less power. The signal can thereby appear in several DMs. The first pass of ACCEL_SIFT.PY identifies the candidates with the same period. These are finally merged into a unique candidate, keeping the information on the DM where it has been found.

However, the pulsar can't appear in many totally different DMs, but in contiguous DMs with an increasing Fourier power towards the good DM. Also, for each unique candidate, the DM value, which is supposed to be the closest to the exact value, is given by the DM with the highest SNR. An error of one DM step, i.e. 0.01 pc.cm^{-3}, yields a spread of about 20 ms representing $2-3$ time samples. Consequently, due to this sensitivity relative to the DM, a candidate is considered a potential pulsar if the number of consecutive DMs where it is detected is at least three.

In the resulting list of good candidates, as for the Fourier transform, it is possible that some candidates are indeed harmonics of the same candidate. To avoid keeping redundant candidates, the candidates are compared to each other in order to identify the major harmonics. The duplicate candidates (i.e. the harmonics) are removed, and the candidate with the fundamental is kept. For the weakest pulsars where a few harmonics are found, it effectively remains at the end only the fundamental. But for the most powerful, plenty of harmonics with uncommon period ratios are found and can't be easily identified. The strongest known pulsars are present in many candidates at the same DM, as B0809+74 for example, which appears 21 times in the final list of good candidates.

Finally, the two outputs of ACCEL_SIFT.PY are a plot of the remaining candidates in a DM – period plan illustrated by Figure 6.12, and an ASCII file listing these good candidates. They are ordered by the significance of the Fourier candidate, with much information such as the period, the DM, the Fourier power, and the bin position. The different DMs where the candidate was found are also marked, allowing us to eventually examine afterward the quality of a candidate by looking at the evolution of the SNR with the DM.

6.3.6 Folding

The previous steps of the processing give the period and the DM of the candidates. But it is insufficient to confirm it is a pulsar. To do this, an ultimate step consists to do the folding of the candidate. Once folded, it is possible to check if the candidate actually shows a correct profile and a structure in time and frequency compatible with a real pulsar.

This final step is realized using PREPFOLD from PRESTO, which prepares the data performing a folding. PREPFOLD takes as input the data extracted from the filterbank file, dedispersed at the DM of the candidate, and finally folds the dedispersed data at the period of the candidate. These folded data are used to generate an integrated profile, a time-phase diagram with the corresponding time evolution of the SNR, and a frequency-phase diagram. In addition to this useful work, the first interest of PREPFOLD is the possibility to resample the data in order to improve the SNR.

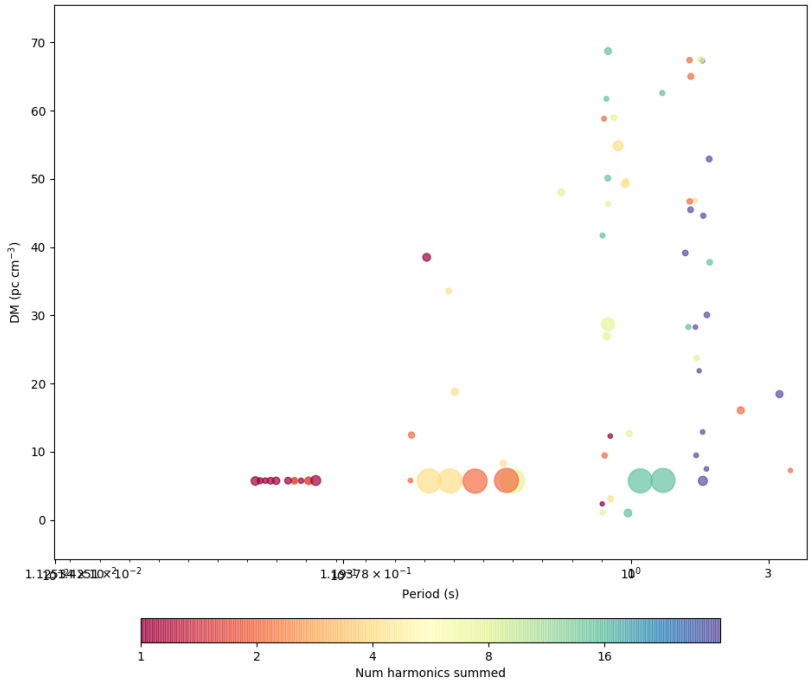


Figure 6.12: Output plot of ACCELSEARCH showing the detected candidates in the plane DM (y-axis) – period (x-axis). The diameter of circles is proportional to their significance. At about 8 pc.cm^{-3}, many candidates are aligned and correspond to several harmonics of the pulsar B0809+74.

Figure 6.13 shows an example of an output plot generated by PREPFOLD for a candidate corresponding to the pulsar B0809+74. On the top part, the integrated profile of the candidate is drawn, and the metadata of the candidate and the results of the folding and dedispersion are summarized. The bottom part can be divided into three subparts: the left one shows the evolution of the folded profile with time, the middle one shows the evolution with frequency and DM, and the right one shows the result of the search in the period and period derivative.

The second interest of PREPFOLD is the capacity to do a fine search in DM, period, and period derivative. These finer searches are done by computing the phases yield by the fine variations to perform a more accurate folding of the data. The processing time for the dedispersion being proportional to the number of DMs to try, the used DM resolution has been set to a relatively high value for our frequencies. Also, the characteristics of the data make it then possible to increase this resolution. Also, concerning the periods, the search using a folding technic allows compensating the loss of period resolution of the FFT for the candidate with a long period. Hence, this finer search allows for increasing the sensitivity in DM and period compared to the rough previous searches. Finally, in addition to the improvement of the properties of the candidate, the evolution of the SNR relative to the DM, period, and period derivative can also be checked.

For each found candidate, PREPFOLD extracts the period and the DM of the candidate from a list of candidates based on the output list of the sifting step. The folding, dedispersion, and computing of the different searches are then carried out using several resampling and searching parameters. There are indeed seven parameters to define: three for the resampling and four for the fine searches.

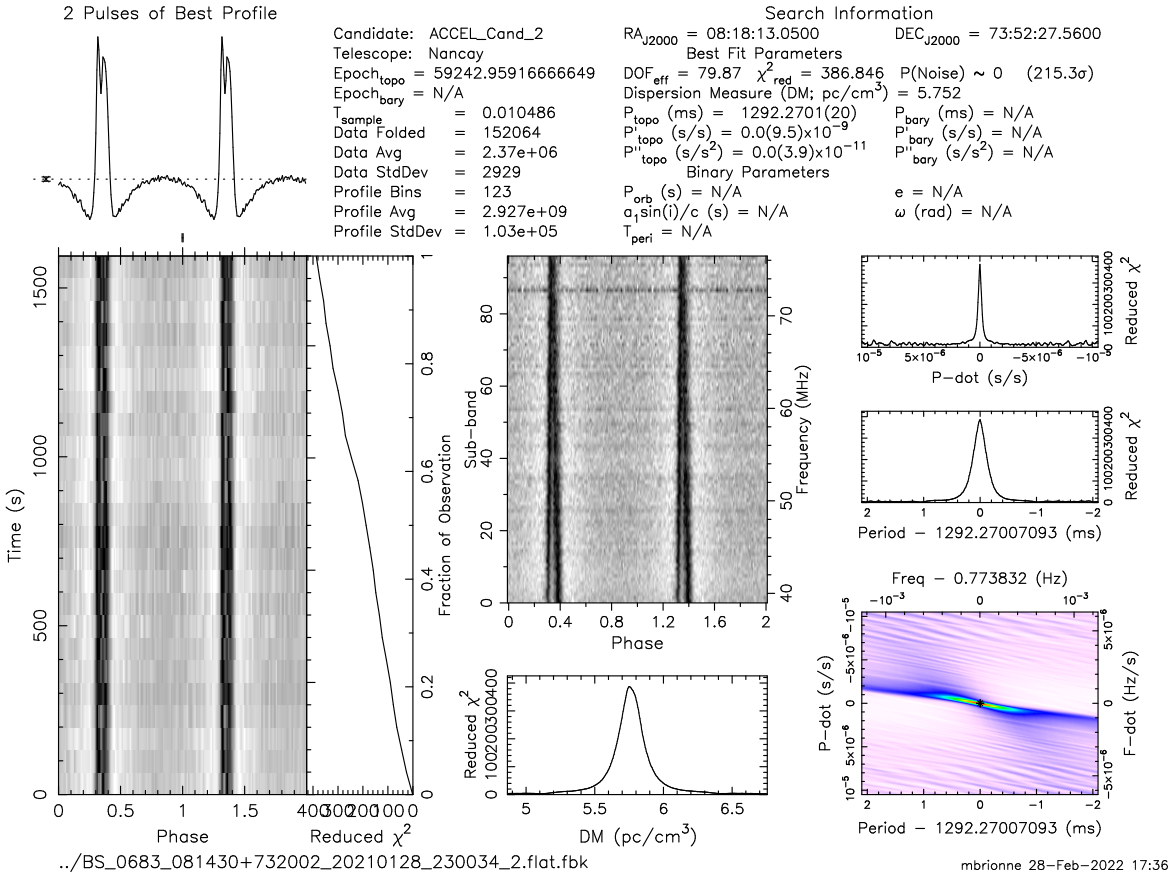


Figure 6.13: Output graph generated by PREPFOLD showing the 8 plots used to control the validity of the candidate in diverse ways.
Top: the left panel shows the integrated profile over two periods, and the right panel notices various information concerning the observation, the data of the initial observation, and finally the results related to the fine searches in DM, period, and period derivative.
Bottom left: time-phase diagram, equally drawn over two periods, with just on the right of the plan, the evolution of the χ^2 of the profile along the duration of the observation.
Bottom middle: frequency-phase diagram for the top plot (over two periods), and the fine search of the DM for the bottom plot.
Bottom right: the top plot represents the fine search in period derivative, the middle one concerns the fine search in the period, and finally the bottom and last one is the $P - \dot{P}$ plan computed based on the two previous plots.

Resampling parameters

Profile resolution The first parameter used by PREPFOLD is the number of bins p_l in the folded profile (corresponding to the option of PREPFOLD *-n*: number of bins in the profile). To obtain the best able time resolution, the number of bins is defined relative to the time sampling of the observation t_s. However, in the case of a candidate with a long period P_0, the number of bins can become important, resulting in a longer computing time for the fine searches. In consequence, the number of bins in the profile was defined relative to a limit of 128 bins.

$$p_l = \begin{cases} \dfrac{P_0}{t_s} + 0.5 & | \quad p_l \leq 128 \\ \dfrac{\nu_0^2 \cdot P_0}{5.10^{-3} \cdot \mathcal{D}} & | \quad p_l > 128 \end{cases} \tag{6.8}$$

Therefore, for the candidates with a period greater than 1.342 s, the size of the bin is defined as the time corresponding to the dispersion smearing produced by a DM variation of 5.10^{-3} pc.cm^{-3}at the lowest frequency ν_0.

Moreover, to be able to see if there is a pulse, it requires a minimum number of points to discriminate the peak from the baseline. We can consider that three points upon the baseline represent a reasonable number to build a peak. Also, at least three other points are required to establish the baseline. According to the fact that the spreading of the pulse due to the scattering can be important, the lower limit was then defined to a minimum number of eight bins in the profile. Furthermore, this limit has been equally determined in order to be able to perform a correct post-analysis of the candidate (see Chapter 7).

Time resolution The second resampling parameter is the number of parts N_p representing the number of divisions of the observation time T_{obs} (corresponding to the option of PREPFOLD *-npart*: number of sub-integrations to use, with a default value of 64). In the case of the profile, all the periods in the observation are folded. However, for the generation of the time-phase plan, it is required to keep a certain time resolution for the sub-integrations. The total time is thus divided into N_p time sub-integrations where $T_{obs}/(N_p \cdot P_0)$ periods are folded.

In the standard parameter of PREPFOLD, the number of parts is calculated for five periods. However, in the context of a low-frequency survey, we rather expect a weak flux for the majority of the pulsars. Consequently, the number of parts was defined in order to fold 50 periods by sub-integration.

$$N_p = \frac{T_{obs}}{50 P_0} \quad | \quad 8 \leq N_p \leq 32 \tag{6.9}$$

Two constraints on the number are still imposed to keep an appropriate number of parts, whether it is in terms of time computing or time resolution. Except for RRATs and pulsars showing a nulling effect, the pulsars emit continuously in time, and we can exploit this property to discriminate a pulsar with another signal. For the candidates with a period greater than 4.05 s, the number of parts is thereby limited to eight in order to have a sufficient time resolution to see the evolution of the signal of the candidate in time.

In the opposite case, the higher the number of parts longer the computing time. This is especially the case for the period and period derivative searches which carry out a number of operations proportional to the number of profiles (i.e. the number of parts). Thus, for the candidates with a period less than 675 ms, the number of parts is constrained at 48, letting already a reasonable time division, and furthermore allowing to fold more periods.

Frequency resolution The last resample parameter concerns the sampling uses for the dedispersion carried out in the initial data before performing the folding (corresponding to the option of PREPFOLD *-nsub*: number of sub-bands to use, with a default value of 128 for the blind survey data). In the previous dedispersion step, the obtained dedispersed time series was integrated in frequency with the aim to make an FFT of the signal and find periodicities. We need therefore to have the most able precise dedispersion to increase the SNR of an eventual pulsar signal in the FFT. The present folding step is an additional step to check many properties, whose behavior in frequency and where we know already the DM and the period.

Moreover, in light of the scattering, there is a significant probability that the folded pulse be spread, essentially in the lowest frequencies. In consequence, in order to avoid increasing too much the computing time, it is possible to sacrifice a little accuracy in the dedispersion by reducing the requirements on the size of the frequency subband.

The number of subbands to use n_{sub} is thereby computed starting from the same equation as for the dedispersion step (reversed for the number of subbands rather than the number of channels). However, the number of subbands is there calculated for the smearing of the quarter of the period of the candidate $P_0/4$ at the lowest frequency channel of bandwidth $\Delta\nu$.

$$n_{sub} = \frac{\Delta\nu}{\nu_0} \cdot \left\{ \left(1 - \frac{P_0 \cdot \nu_0^2}{4\,\mathcal{D} \cdot DM} \right)^{-\frac{1}{2}} - 1 \right\}^{-1} \tag{6.10}$$

The candidates with a long period and a small DM give a resulting number of subbands very small. Also, in order to keep a little bit of frequency resolution, the number of subbands is still limited to a minimum of 32. Furthermore, this type of candidate has a profile with a high number of bins, i.e. a better time resolution. Consequently, they can be more sensitive to a bad DM correction, and this limitation should allow for improving it.

The number of subbands is effectively reduced compared to the most accurate one used before. However, even for low DMs, it remains important because of the strength of the dispersion in the lowest frequencies. As shown in the paragraph 6.3.3, the dispersion smearing decreases approximately by a factor of three for the central frequency relative to the lowest frequency. That leads to an error at the central frequency of only $1/12^{th}$ of the period of the pulsar.

Fine search parameters

DM resolution After these three first parameters allowing to set the different resolutions of the dedispersed folded candidate, there are four other parameters to fix to adjust the DM, the period, and the period derivative.

The first one is related to the size of the DM step dDM to use for the search. In our case, this folding step is the occasion to do a DM search with the best DM resolution able to be reached at our frequencies. Then, the coefficient related to the DM step c_d (corresponding to the option of PREPFOLD *-dmstep*: minimum DM stepsize, with a default value of two) is set to one to have the minimum one, which is determined by PREPFOLD with the following relation:

$$dDM = c_d \cdot \frac{\nu^2 \cdot P_0}{\mathcal{D} \cdot p_l} \quad | \quad c_d = 1 \tag{6.11}$$

The computed DM step represents the DM producing a smearing of one bin in the profile of length p_l. With this parametrization, for candidates with a period less than 1.342 s, a DM resolution equal to 5.22×10^{-3} pc.cm^{-3}can be reached. According to the time sampling of the initial data of 10.486 ms, we obtain thereby the limit value possible.

DM range The next parameter is to define the DM range where to search the precise DM. PERPFOLD is taken as input parameter a factor f_d (corresponding to the option of PREPFOLD *-ndmfact*, with a default value of three) which is used to compute the number of DMs to try as $2f_d \cdot p_l + 1$. This factor can be then defined function of the wished DM range to reach.

$$\Delta DM = dDM \cdot (2f_d \cdot p_l + 1) \quad \Leftrightarrow \quad f_d = \frac{1}{2p_l} \cdot \left(\frac{\Delta DM \cdot \mathcal{D} \cdot p_l}{\nu_0^2 \cdot P_0} - 1 \right) \tag{6.12}$$

Because of the strong scattering yet, the error in DM can be a little bit important. It could therefore be useful to search not too close to the DM value found with the previous steps.

Furthermore, the DM fine search is not solely used to adjust the DM, but it is also very useful to control if the candidate has good behavior relative to a DM variation. Typically, without a strong scattering, the pulse has a Gaussian shape that is slightly deformed towards later times. Consequently, the SNR of the profile of the pulsar must approximately increase symmetrically around the best DM. In consequence, to provide a DM window with sufficient size to check the evolution in DM, the DM range ΔDM was set to a large value of 2 pc.cm^{-3}.

Period range The two last parameters concern the period search and are similar to the two DM parameters: a parameter for the frequency resolution and another for the period range to reach. As for the DM search, the period search is not merely used for the adjustment of the period of the candidate, but it is equally useful to check the behavior of the candidate relative to an error in the period. For a pulsar, an error in the folding period, or the period derivative, results in a linear drift in the rotational phase of the pulse in time. In the same way that for the evolution in DM, the evolution in the period and period derivative must be approximately symmetric with a maximum SNR at the best period and period derivative.

In the case of the DM, the pulse is already well folded in each subband, and then a little error misaligns the lowest frequencies. However, higher subbands are still aligned, and it needs a relatively important error to completely misalign the highest subbands. It is therefore needed to try a large range of DMs to obtain the evolution in DM in terms of SNR. On the contrary, the SNR of the profile is more sensitive to an error in period or period derivative, because this error infers directly in the folding, and then in the generation of the profile. Thus, during the folding, the error is propagated and multiplied by the number of periods to fold, which goes from multiplication by 60 for a pulsar with a period of 30 s to 60 000 for a pulsar with a period equal to the minimum searched period of 30 ms.

The period and period derivative searches are consequently more sensitive and need a range of tried values smaller than for the DM search. Moreover, the sensitivity to the period or period derivative error is logically proportional to the period of the pulsar. It thus needs a range which is also a function of the period of the candidate.

The number of tried periods is defined in PREPFOLD in the same way that for the DM, but with the period factor f_p (corresponding to the option of PREPFOLD *-npfact*, with a default value of two) which replaces the DM factor: $n_{per} = 2f_p \cdot p_l + 1$. The period range can be thereby expressed by the following relation:

$$\Delta P(P_0) = \frac{1}{f_{min}(P_0)} - \frac{1}{f_0} = \frac{1}{f_0 - \frac{(n_{per}(P_0)-1)}{2} \cdot \delta f} - \frac{1}{f_0} \tag{6.13}$$

Where f_{min} represents the lowest rotational frequency to reach, $f_0 = 1/P_0$ is the rotational frequency of the candidate, and δf is the rotational frequency resolution. This last one is defined as $1/(p_l \cdot T_{obs})$ in PREPFOLD, allowing to isolate and compute the f_p corresponding to the planned period range.

$$f_p = \frac{T_{obs}}{P_0} \cdot \left(1 + \frac{1}{r_p}\right)^{-1} \tag{6.14}$$

With $r_p = \Delta P(P_0)/P_0$ the fraction of period to use as the period range. Because of the sensitivity previously explained, this fraction is a constant value set to 0.002. Although this value is weak, it still represents, for a typical pulsar of 1 s, an error of 2 ms by the period. Accumulated over about 1 800 periods, it produces at the end a cumulative error of 3.6 s which is more than enough to completely lose the pulsar.

Period resolution The previous calculated period factor is determined for the maximum rotational frequency resolution δf. This factor is used to obtain the number of periods to try. The period and the period derivative produce a combined effect on the SNR of the folded profile. To evaluate their covariance, PREPFOLD builds a $P - \dot{P}$ plan, where each point is equal to the SNR of the integrated profile for a certain couple of values $(P, \dot{P})$.

The shape of the computed $P - \dot{P}$ plan is a square of n_{per} bins of side. The time to generate this $P - \dot{P}$ plan can become very long in case of a very high number of periods, and then also period derivative. To avoid consuming overmuch time in this work, the number of periods to try was limited to 1 001, representing a constraint on the period factor. The period factor f'_p is thus computed again relative to this limit:

$$2f_p \cdot p_l + 1 \leq 1001 \quad \Rightarrow \quad f'_p = \frac{500}{p_l} \tag{6.15}$$

To keep our period window, the rotational frequency δf is consequently adapted using the period step factor c_p taken by PREPFOLD (corresponding to the option of PREPFOLD *-pstep*, with a default value of two).

$$\delta f' = c_p \cdot \delta f = \frac{c_p}{p_l \cdot T} \Rightarrow c_p = \frac{T_{obs}}{P_0 \cdot f'_p} \cdot \left(1 + \frac{1}{r_p}\right)^{-1} \tag{6.16}$$

A last verification is done to avoid too many long period steps relative to the period of the candidate. A last condition is to require a period step of less than half of the period. In this case, the c_p is defined as $p_l/2$, and f_p is defined again taking this new rotational frequency resolution.

6.4 Implementation

6.4.1 Processing node

The processing of the survey data may take a very long computing time and use very high computing power. Three solutions have been proposed to perform the processing: use one of the processing nodes of the Nançay data center, or an exterior processing node, or develop a dedicated machine.

The processing of the current survey was estimated to be many months of computing between one and two dozen months following the computing power of the machine. Moreover, it is normally planned to extend the survey at the lower declination in order to cover at the end all of the northern sky. The entire time of processing is consequently of some years. The problem with the utilization of the processing nodes of the Nançay data center is there are shared machines. The computing power varies then so much relative to the number and the used resources of the tasks run by all the different users. In addition, it is unreasonable to use the entire power of the processing node for many months. Concerning another processing node, the problem is also the sharing, added to the fact that it is unsure to have the needed computing time for some years.

The second point is the storage of the almost 8 000 files with their resulting data of processing. The storage of the raw data is evaluated for the two observing phases to about 27 TB, and the survey needs thereby many dozens of TB of storage. Combined with the fact that these data are processed in another place than the Nançay data center, local storage seemed to be an easier choice to avoid a lot of big transfers.

In consequence, it was decided to build a machine entirely dedicated to the NenuFAR pulsar blind survey, working only on the processing of the data for the next years. Furthermore, the choice of a new dedicated machine allowed us to adapt the characteristics of the machine to the needs of the survey processing, such as high computing power and important storage.

The machine has been designed by Louis Bondonneau and Ismaël Cognard in 2020. Unlikely, because of the different lockdowns and the supply difficulties from Asia, the machine called *Baudroie*[4] has been installed in July 2021 only.

Baudroie is equipped of 2 processors AMD EPYC 7282 with 16 cores with double threads of 3.2 GHz, giving thus 64 high-performing CPUs in total. For the memory, two memory modules of 128 MB are installed, allowing to use up to 256 MB of RAM. Finally, concerning storage, six HDDs of 11 TB each are mounted on the machine allowing entirely saving data from this survey for eventual further reprocessing.

[4] The "baudroie" is the French name of the humpback anglerfish, which is a fish in the abyss emitting light by bioluminescence in the darkness to find prey. This name was chosen because our Baudroie searches our favorite prey, the pulsars, blindly as the real humpback anglerfish.

6.4.2 Optimization

Once the machine was installed in the server room of the LPC2E in Orléans, the processing pipeline could have been tested in real condition. Each step previously presented is included in a Python script where the PRESTO command (or the Python function for the flattening step) is encapsulated in a multi-processed structure, using the Python package MULTIPROCESSING. However, although the multiprocessing done was efficient, some bottlenecks remained, and the capacities of the machine were not totally used.

Without any optimization, except the multiprocessing done with Python for each step, the processing pipeline required 7 hours for one pointing, using all the available CPUs during a large part of the time. This processing time leads to 6 years and 2 months to compute the whole survey. So, to reduce this computing time to the maximum, the pipeline was benchmarked to find the least efficient parts and saturations which slow down the computing.

Flattening step

The first step of the processing pipeline realizes the flattening of the raw data before doing the search per se. The flattening needs to do many convolutions on the entire dynamic spectrum channel by channel. Furthermore, work in a multi-processed structure requires treating simultaneously many channels, leading to the use of important memory storage. A second limitation has been identified concerning the saturation of the reading and writing access, limiting the parallelized processes to just about three simultaneously.

A multi-processed structure greater than 3 processes at the same time is consequently useless. The problem was bypassed by launching on three CPUs some pointings in offbeat to screen the reading and writing of the data. However, the important storage asked occurs a saturation of the memory from three simultaneous pointings. Finally, for the flattening step, the pointings are sent two by two in the script limited to three processes in parallel for each.

The data of the dynamic spectrum to process are voluminous, and the control graph generated at the end took about 55% of the computing time. To significantly decrease the time of the flattening step, once the flattening step was completely validated, the generation of the control graphs has been removed. Without the plotting, a factor of 5.8 was globally won on this step, representing a total of more than 3 months for the entire survey.

RFI mitigation step

An evident bottleneck of the processing pipeline was noticed for the RFI mitigation step. The RFI mitigation cannot be indeed multi-processed, because of the fact that the RFI identification needs to cross different statistics on various scales. This step is then difficult to parallelize and is consequently done on just one CPU.

Although this step was fast (about 2% of the processing time), it seemed that, at first sight, the parallelization of the different RFI mitigation could gain a significant time on the whole survey. However, the RFI mitigation encounters the same problem as the flattening step, namely the saturation of the reading and writing of the data.

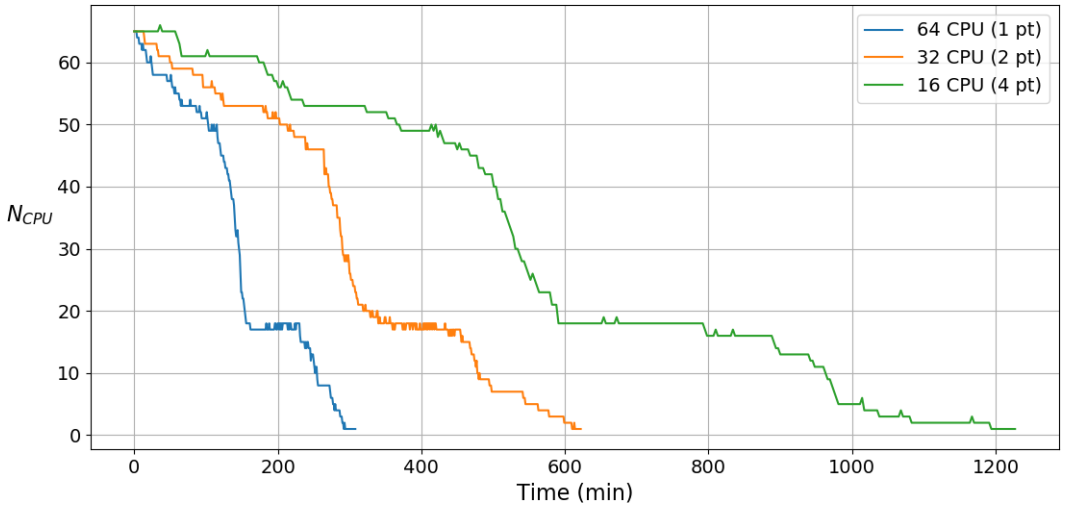
The tests made with different numbers of pointings simultaneously treated have shown that the reading/writing of Baudroie is saturated from 16 pointings in parallel. In consequence, for the RFI mitigation step, the pointings are sent 16 by 16 to the RFI mitigation program RFIFIND, allowing to win obviously a factor 16 on this step which represents about 8 weeks on the whole survey.

Dedispersion step

The dedispersion step represents by far the longest stage of the processing pipeline, with 5 hours per pointing, and then 71% of the total processing time. Furthermore, the benchmarking has shown an inhomogeneity of the different dedispersion done, leading to a substantial loss in efficiency for the carried out multiprocessing.

The evolution of the used resources has been done to estimate the loss in efficiency and the possibilities that can accept Baudroie. Figure 6.14 shows the evolution of the number of used CPUs over time for the three tests made: run 1 pointing on 64 CPUs, 2 simultaneous pointings on 32 CPUs each, and 4 simultaneous pointings on 16 CPUs each.

Figure 6.14: Time evolution of the number of CPUs used to carry out the dedispersion for three configurations: 1 pointing on 64 CPUs (blue line), 2 simultaneous pointings on 32 CPUs each (orange line), and 4 simultaneous pointings on 16 CPUs each (green line).



It appears clearly that about 1/4 of the dedispersions are fastly ready after just 1/3 of the total time whatever the looked test. Also, about 3/4 of the CPUs are freed after the middle of the total computing time for all the tests too. The dedispersion step is thereby very inefficient in the most straightforward way consisting to run the pointings in the row on all the available CPUs, revealing an important bottleneck.

Table 6.2 summarizes the results of the three treated tests with the number of used CPUs, the number of pointings simultaneously treat, the averaged used memory, the maximum and average number of used CPUs, the total time and the computing time relative to the number of pointings done. Whatever the configuration, the average number of used CPUs is all the time about 30 (46.9% of the available resources), meaning that the decreasing law of the number of used CPUs is linear with the number of pointings simultaneously launched. The used memory is also all the time about 165 GB, meaning that the allocated memory is proportional to the number of used CPUs and not to the number of files to treat.

One can notice that in the configuration with four simultaneous pointings, the total number of used CPUs falls below 48 CPUs after about one hour and a half. Each pointing being launched on 16 CPUs, it is, therefore, possible to run a new pointing. In order to fill as many as possible the CPUs, the Python script making the dedispersion has been set to work on 16 CPUs, and has been encapsulated in a Bash loop launching a new pointing as soon as 16 CPUs are freed.

This configuration of parallelism allows working up to six pointings at the same time. In addition, the use of the resources in CPUs is more efficient with an average number of used CPUs about 57 (89.1% of the available resources).

Folding step

The last step of the processing pipeline, the folding of the found candidates, is shorter than the dedispersion step. But the several dozen of candidates to fold still needs significant computing time, in particular on the global survey where the number of candidates is several hundreds of thousands.

In a similar way that for the dedispersion step, three different configurations were tested, checking the evolution of the use of the memory and CPUs. The use of the memory is constant whatever the configuration, meaning the use of memory is proportional to the number of processes like for the dedispersion step. The computing time is in contrast not linear following the number of used CPUs. Thereby, dividing by two the number of CPUs from 64 to 32 increases the time by 1.4, and by a factor of 3 for four times fewer CPUs.

Table 6.2: Utilization of the memory and the CPUs for three different configurations to carry out the dedispersion step of the processing pipeline. (1) Number of CPUs to multiprocess the dedispersion of the pointing. (2) Number of pointings treated at the same time. (3) Averaged used memory. (4) Averaged used number of CPUs. (5) Total computing time to process all the pointings. (6) Equivalent computing time for one pointing (Tot. time / N_{pt}).

(1) $\mathbf{N}_{CPU}$	(2) $\mathbf{N}_{pt}$	(3) **Avg. mem.** (GB)	(4) **Avg.** $\mathbf{N}_{proc}$	(5) **Tot. time** (min)	(6) **Time / obs** (min)
16	4	170	25	1197	300
32	2	164	30	610	310
64	1	159	32	292	292

Moreover, in the case of folding on simultaneous pointings rather than only one, the average number of used CPUs changes following the chosen pointings. This change is due, contrary to the dedispersion step where the dedispersion plan is the same for all the pointings, to a different number of candidates to fold. In addition, it can have large differences in the folding parameters between the candidates.

In consequence, the folding step for a pointing is slowed by certain candidates which are longer to compute because of their folding parameters. Running several pointings at the same time leads to an optimization of the CPU resources. Nevertheless, even with the best optimization, the inhomogeneity of the number of candidates to fold will result in an inhomogeneous computing time following the group of pointings to fold.

To maximize the CPU utilization and minimize as much as possible the computing time, the folding step is carried out by exploiting an identical method as the dedispersion step. Then, the Python script making the folding is multi-processed on 16 CPUs and is also encapsulated in a similar Bash loop, starting a new pointing as soon as 16 CPUs are freed.

6.4.3 Progress of the processing

Finished processing

The operation of the processing pipeline was tested on the 96 pointings observed during a full-scale test phase in December 2020. Once this test has been checked and validated, the processing of the survey data has been started. The processing of the data has begun on the 4^{th} of November 2021, starting with the data of phase 1 of the observing program.

The raw files generated by the back-end machine *UnDySpuTed* are sent to the storage of the Nançay data center ordered by year and months. The processing of the data is then actually also carried out month by month, leading to easier management of the data. Moreover, the two phases are consecutive rather than simultaneous, and the treatment month by month allows the treatrement of the two phases almost independently. At the 15^{th} August of 2021, 8 months of data are processed from November 2020 to June 2021, representing about 29.68% of all the pointings of the survey.

Table 6.3 presents a summary of the first eight processed months, showing the number of processed pointings with the corresponding computing time and the sizes of the needed storage before and after the processing. We can see that since November 2021, 2 283 pointings have been processed in 247 days, representing an averaged computing time of 2.38 hours to process one pointing of 30 minutes duration.

The initial raw files, i.e. the FILTERBANK files containing the initial dynamic spectrum, represents 8.7 Tb of data in all, and after the processing, 13.7 Tb of resulting data are created and are saved in the storage disks of the processing machine Baudroie. These resulting data are composed of different log

files, result plots, results tables, of the flattened FILTERBANK file, of all the PFD files containing the data of each found candidate, and finally of the files generated by ACCELSEARCH for the periodicity search. These last files are not light binary files but are kept in the case where we want to redo a folding of the candidates with different parameters. The whole storage for these eight months of data is thus 22.4 Tb.

Because of modifications in the conversion process of the SPECTRA files in FILTERBANK files, the three first months are required to be converted again. Then, contrary to the next months, these three months are not well ordered on the Nançay data center, and the new FILTERBANK files coexist with the old ones. Before removing of the original SPECTRA files and ordering the data to process, it is necessary to validate the changing of the files to treat. It was consequently decided to begin by the next months beginning with the data of November 2020.

The current presented data processing is finished since the 15^{th} of August 2022 and has been conducted until June 2021. The main part of phase 1 of the observing program was finished in mid-July 2021, allowing the treatment of approximately about 60% of phase 1. Since this date, the processing has been paused for two reasons. The first one is that the three remaining months (from August 2020 to October 2020) are composed of an important number of pointings of more than 400 for each month. Each of them requires about one and a half months to process the data. In consequence, that leads to a duration of more than four months, resulting in the completion of the processing of phase 1 after the writing of this thesis. The second reason is that pausing the processing allowed free Baudroie. The processing node may have been used for the post-processing of the candidates found in the eight processed months (see Chapter 7). Baudroie has been thereby used to efficiently perform this following analysis, permitting us to obtain some first final results for this manuscript.

Quantification of the efficiency

For the optimized steps, constituting the longest steps of the processing pipeline, the total computing time was measured for each processed month. Table 6.4 presents the different computing times in days for the very long steps: the flattening and the dedispersion, and in hours for the shorter steps: the RFI mitigation and the folding. These different computing times have been then compared to the theoretically required times without any optimization, in order to estimate the efficiency of the carried-out optimization. Table 6.5 shows for each of the four steps the fraction of the real computing time compared to the time without optimization.

The result is a constant ratio of the processing time in the different months for each step. Furthermore, one can notice that the most substantial gain is realized in the RFI mitigation, with a division of the

Table 6.3: Summary of the processing done at the 15^{th} of August 2021, presenting the number of processed pointings, the computing time to process, the averaged computing time per pointing, the size of the initial FILTERBANK files, and the size of the result files after processing. $^{(\star)}$ 2.38 hours don't represent the sum, but actually the mean of all the averaged computing times per pointing.

Month	**Nb. of pointings**	**Computing time** (days)	**Time / pointing** (hours)	**Raw file size** (Tb)	**Final size** (Tb)
11/20	213	23	2.49	0.8	1.2
12/20	413	44	2.07	1.4	2.4
01/21	355	38	2.45	1.6	2.1
02/21	330	36	2.48	1.2	2.0
03/21	249	27	2.54	0.9	1.5
04/21	272	29	2.56	1.0	1.7
05/21	257	28	2.57	0.9	1.6
06/21	194	22	1.85	1.0	1.2
Total	**2283**	**247**	**2.38** $^{(\star)}$	**8.7**	**13.7**

Month	**Flattening** (days)	**RFI mitigation** (hours)	**Dedispersion** (days)	**Folding** (hours)
11/2020	2	4	21	17.5
12/2020	4	8	39	45
01/2021	3	7.5	33	33.5
02/2021	2.5	7.5	32	34.5
03/2021	2.5	6.5	23	24
04/2021	2.5	6.5	25	28
05/2021	2	4	25	27
06/2021	1.5	3.5	19	24.5
Total	**20**	**47.5**	**217**	**234**

Table 6.4: Computing times of each optimized step for each of the 8 processed months of data.

Month	**Flattening** (%)	**RFI mitigation** (%)	**Dedispersion** (%)	**Folding** (%)
11/20	22.54	11.27	47.32	16.43
12/20	23.24	11.62	45.33	21.79
01/21	20.28	12.68	44.62	18.87
02/21	18.18	13.64	46.55	20.91
03/21	24.10	15.66	44.34	19.28
04/21	22.06	14.34	44.12	20.59
05/21	18.68	9.34	46.69	21.01
06/21	18.56	10.82	47.01	25.26
Avg. ratio	**21.02**	**12.48**	**45.62**	**20.50**

Table 6.5: Ratios of the real computing time relative to the expected time without optimization for each optimized step and for each of the eight processed months of data.

necessary time by a factor of 8. Concerning the flattening and folding steps, there are also rather efficient with a gain of almost 5. The last and longest step, consisting of the dedispersion, is less efficient than the three others. However, the optimization made nevertheless allows dividing by a factor of 2.2 the very long required time. Finally, these ratios of gain represent a saved time in days of:

- flattening: 75,1 days ;
- RFI mitigation: 13.9 days ;
- dedispersion: 258.6 days ;
- folding: 37.8 days.

These won days provide a global reduction of the processing time of 385.4 days, representing more than one year solely for these eight first months of data. The real processing time represents indeed just 40% of the initial expected time. It results from this substantial global gain that the carried-out optimizations are quite effective and not done in vain.

Further processing

At the 31^{st} of August 2022, 2 283 pointings are already processed. As said beforehand, a significant part of the remaining pointings of phase 1 to process are comprised in the three first months of observation. They can be processed in four months, leading to being done at the beginning of 2023.

Also, without issues such as thunderstorms, for example, the entire observing program is expected to be finished at the end of March 2023. According to the processing time of the first eight months, the 4 266 pointings corresponding to the observations from July 2021 to March 2023, could be processed in about 15 months.

As a consequence, the processing of all the observations of the NPBS will normally be completed approximately in May 2024. It results from this estimation of a total processing time approximately of two years and four months, largely lower than the initial required duration of 6 years.

6.4.4 Validation of the search pipeline

After the end of the processing of each month, a validation step was realized in order to verify the capacities of detection of the search pipeline. The control consists to perform a targeted search of the known pulsars in the candidate list. The known pulsars are indeed defined based on the list of the known pulsars of the ATNF catalog, which have been detected by the NenuFAR pulsar census (Bondonneau et al. in preparation). After the selection of the pulsars with a declination greater than 39°, the final list includes 54 pulsars potentially detectable by the survey. However, we have to keep in mind that the census was a standard search of known pulsars, observing the exact position of the pulsar, and performing standard processing with an intra-channel coherent dedispersion at the best-known DM and with folding using a precise ephemeride.

Knowing the coordinates of these pulsars, the search for it was carried out in the pointing data where it should be located. Once the corresponding pointing was found, the targeted search consists to select the candidates with a DM closer than 1% compared to the DM given by the ATNF. For each of these candidates, the ratios between the period of the pulsar and the period of the candidate are calculated to identify the eventual harmonics of the pulsar. In the eight months currently processed, representing approximately half of the sky, 19 known pulsars should be present in at least one of the processed pointings. 46 candidates have been found as corresponding to a harmonic of one of the known pulsars. Finally, these candidates result in the re-detection of 6 pulsars.

Table 6.6: List of the 6 known pulsars redetected in the context of the targeted search. (1) Expected SNR at the center of the pointing. (2) Angular separation between the position of the pulsar and the center of the closest pointing. (3) Expected SNR taking the sensitivity loss due to the angular separation into account. (4) Number of found candidates corresponding to a harmonic of the pulsar. (5) Significance computed by PREPFOLD for the best candidate.

PSR	**P_0** (s)	**DM** (pc.cm^{-3})	(1) **SNR_0**	(2) **Ang. sep.**	(3) **SNR_{exp}** (°)	(4) **Harm.**	(5) **Sigfc.**
J0454+5543	0.34073	14.590	23.7	0.71	12.3	1	16.6
J0700+6418	0.19567	8.774	20.0	0.17	19.2	1	14.1
J0814+7429	1.29224	5.751	204.0	0.65	118.3	21	215.3
J1115+5030	1.65644	9.186	11.7	0.21	11.1	9	57.6
J1509+5531	0.73968	19.619	181.8	0.43	143.4	13	131.1
J1813+4013	0.93109	41.557	10.0	0.76	4.8	1	6.9

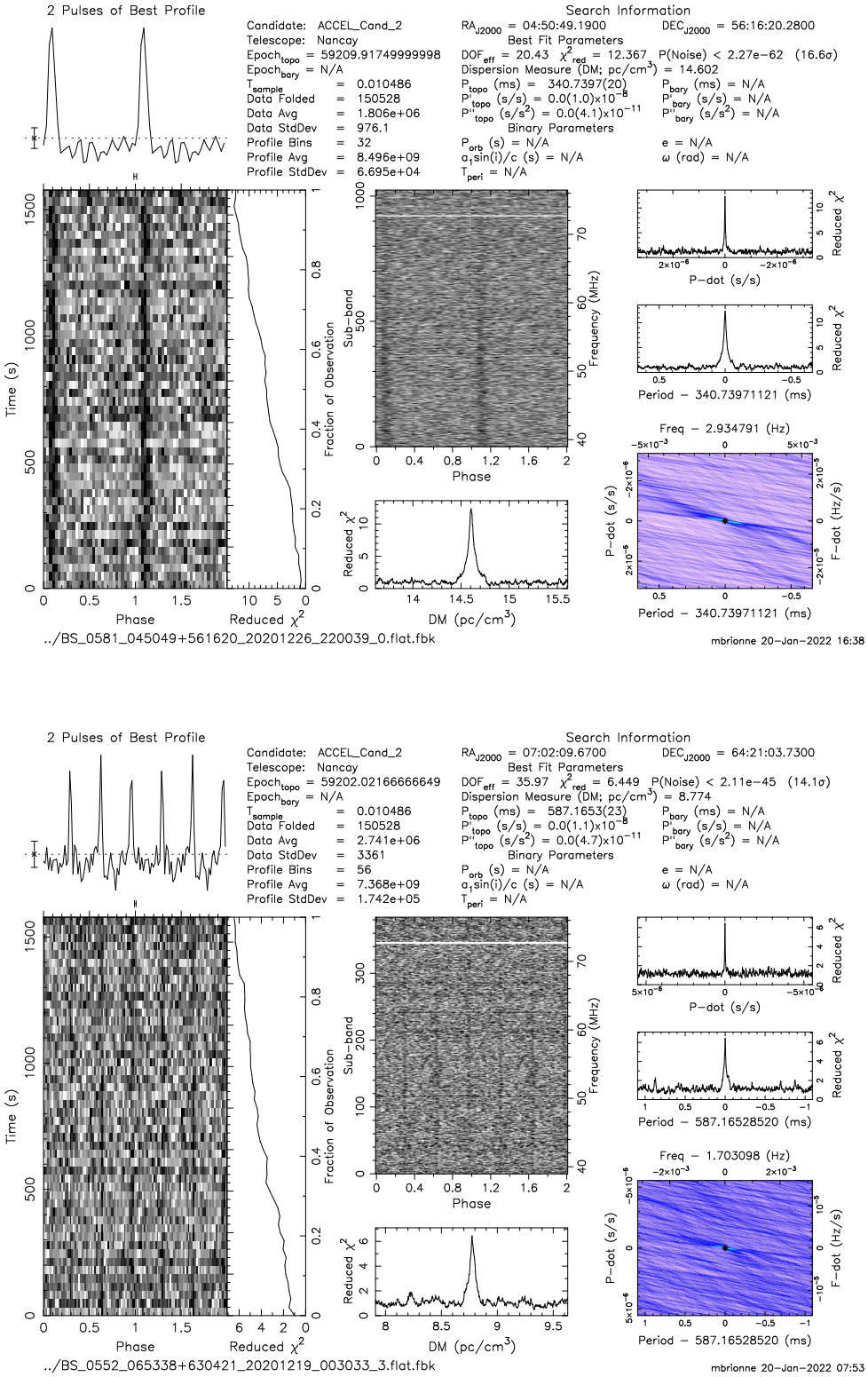
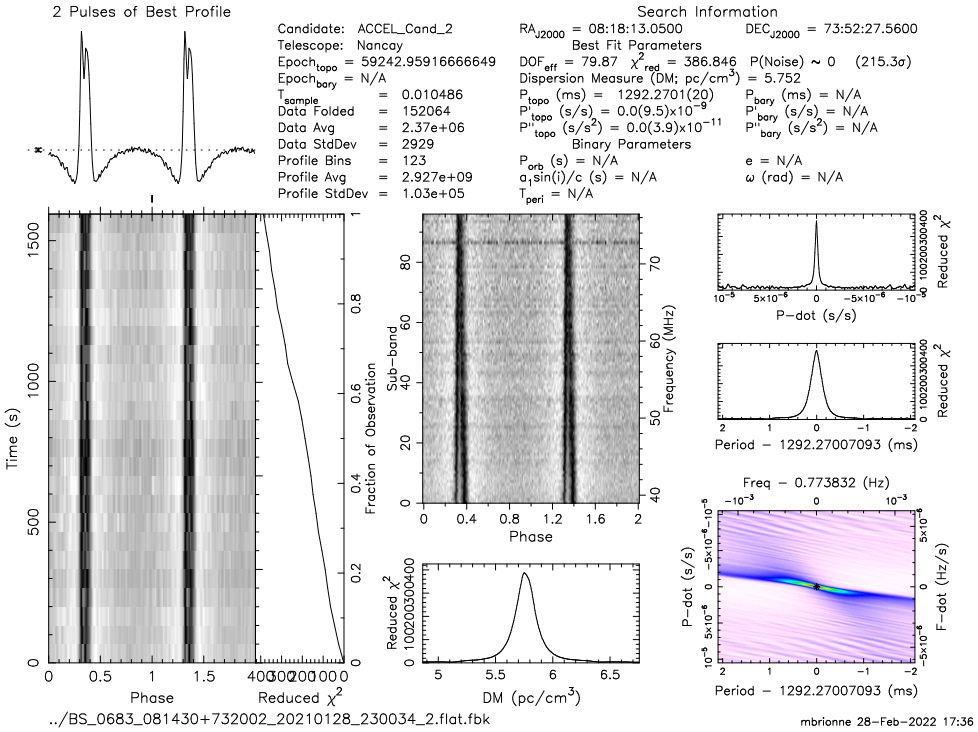


Figure 6.15: Result plots for the re-detections of J0454+5543 and J0700+6418.



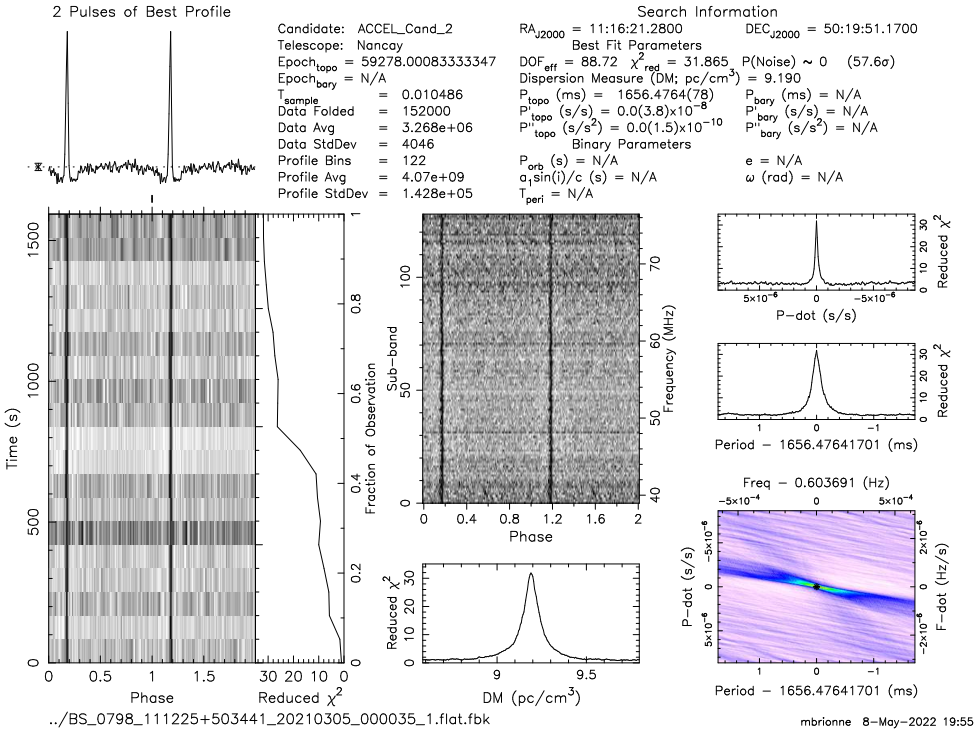
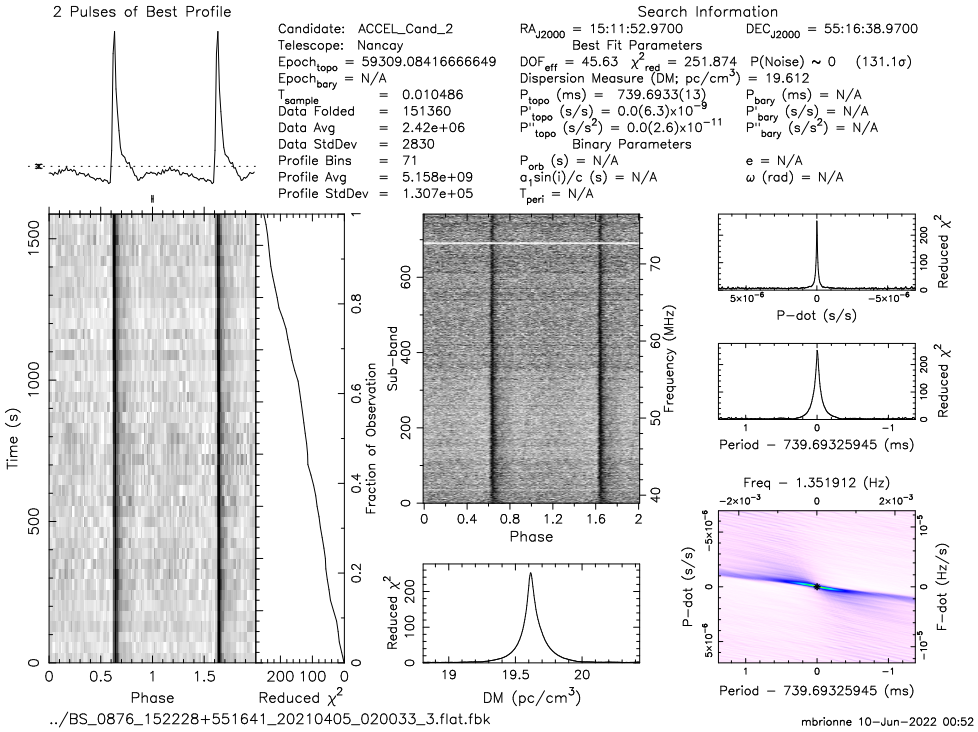


Figure 6.16: Result plots for the re-detections of J0814+7429 and J1115+5030.



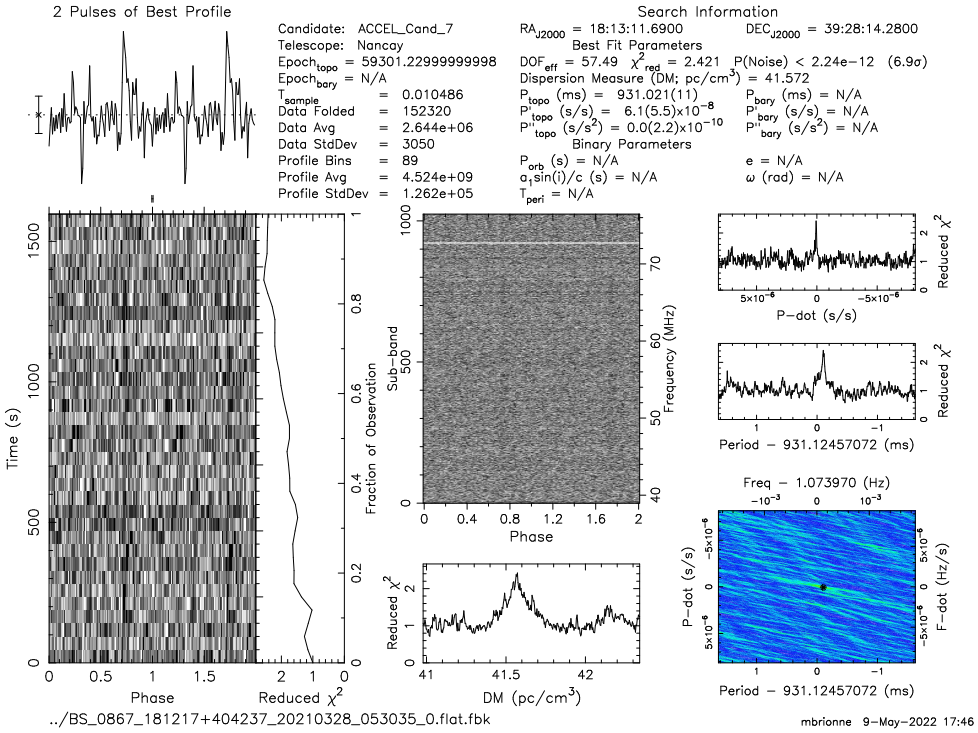


Figure 6.17: Result plots for the re-detections of J1509+5531 and J1813+4013.

Table 6.6 presents the list of the six known pulsars redetected in the processed data. The two last column shows the number of candidates corresponding to the pulsar and the significance determined by PREPFOLD for the best-found candidate. Also, the resulting plot generated by PREPFOLD for each one is presented in Figure 6.15,6.16 and 6.17. We can notice that the three most powerful pulsars are very clearly detected, with the detection of several harmonics of the pulsar. Furthermore, the fundamental harmonic is properly systematically found with the highest significance.

Concerning J0454+5543 and J0700+6418, only one candidate is proposed, but the detection is rather clear. The only faint detection is J1813+4013, on the bottom panel in Figure 6.17. For this pulsar, the pulse can be merely identified a little in the integrated profile, and very lightly in the frequency-phase diagram.

Baseline variations Although the detection of J0814+7429 (see Figure 6.16) is clear, it is noticed that the baseline is not flat. The same behavior can be equally seen for the pulsars J1115+5030 and J1509+5531. However, the bumps of the baseline around the pulse are largely smaller than for the J0814+7429. Finally, no bumps are seen for the three last pulsars, and the baseline is properly flat as expected.

By looking at the SNRs of the pulsars in Table 6.6, it appears that the three pulsars presenting bumps are also the three most intense, leading that it should have a relation between the presence of bumps and the SNR of the observed pulsar.

In fact, the initial time series measured by NenuFAR is not flat, requiring to perform a flattening step. This flattening step, described in Section 6.3.1, uses a running average to normalize the time series. As a result, in the case of single pulses sufficiently intense to bring out noise, the running average will necessarilyy increase around the position of the single pulse, with a maximum average at the exact position of the pulse. Moreover, the resulting plots of PREPFOLD show the profile, integrated in frequency and especially time. As a consequence, to obtain this final baseline over the period of the pulsar, it is required that a lot of intense single pulses are present in the observation.

To conclude about these bumps, the utilization of a running average, whatever the employed method, can introduce local variations of the baseline. However, this behavior occurs solely for intense pulsars detectable in single pulses. At the NenuFAR frequencies, that concerns a few pulsars, and the vast majority of the detections should be not impacted. Furthermore, the pulsars showing these bumps are sufficiently intense to be easily detected despite all. Besides, the three pulsars showing a non-flat baseline are in spite of the three best detections of the survey.

Binary detection The case of J0700+6418 is intriguing for two reasons. The first one is that the best-found candidate is not the fundamental harmonic, but the harmonic 3:1 (i.e. of a period three times longer than the period of the pulsar). As it can be noted on the bottom panel of the resulting plot in Figure 6.15, six peaks appear in the profile (for two periods, then three peaks by a period of the pulsar). Indeed, it appears that this pulsar is in a binary system of 24.68807286(2) hours of orbital period (Jones & Lyne 1988). Although the search pipeline doesn't perform any binary search, J0700+6418 is nevertheless detected.

A binary search was performed in order to check if the binarity of the pulsar interferes with the detection of the fundamental harmonic. J0700+6418 has been searched using an acceleration search over 10 Fourier frequency bins. It results from this binary search a similar detection of the pulsar. As for the standard search (i.e. without any acceleration), the fundamental harmonic of J0700+6418 is unfound. In this case, the acceleration search allows us to detect the pulsar in a harmonic of lower Fourier frequency: the harmonic 5:1 (five times the nominal period).

We can conclude the search pipeline can eventually detect binary pulsars, providing harmonics with a sufficient Fourier power to obtain a significance greater than 2σ (the minimum threshold asked for the periodicity search), and with a sufficiently long orbital period.

Period sensitivity The second reason is that, although it is the harmonic 3:1 of J0700+6418 at 587.17 ms which has been detected, the fundamental is well detected. J0700+6418 is, therefore, the fastest detected pulsar with a period of 195.67 ms. That means the search pipeline is able to detect a periodicity representing 20 time samples.

However, on the 46 candidates corresponding to a known pulsar, there are actually some candidates, corresponding to shorter period harmonics of the most powerful pulsars, with a period lower than 195 ms. The candidate with the shortest period is the harmonic 1:20 of J1509+5531, with a period of just 39.98 ms. This candidate is well detected with a significance of 11.6σ, i.e. better than the detection of J1813+4013. As a consequence, that leads to decreasing the minimum detectable period by the survey to about 40 ms.

Limit of detectability Table 6.7 presents the 13 pulsars not detected by the survey. To validate the search pipeline, it needs to evaluate if these pulsars would have been detected or not. To do this, the expected SNR was calculated based on the SNR measured by the NenuFAR pulsar census. Contrary to the survey, which has constant observing parameters, the census adapted the parameters relative to the observed pulsar. In consequence, based on the radiometer equation, the proportionality factor corresponding to the difference in observing parameters was determined. For each pulsar with a given flux and a given duty cycle, the SNR expected at the center of the digital beam, SNR_0 is estimated by:

$$SNR_0 = \sqrt{\frac{T_{sur} \cdot \Delta\nu_{sur}}{T_{cen} \cdot \Delta\nu_{cen}}} \cdot SNR_{cen} \tag{6.17}$$

With SNR_{cen} the SNR obtained by the NenuFAR pulsar census (Bondonneau et al. in preparation), T the observation durations, and $\Delta\nu$ the frequency bandwidth of observation (indicated by the index *cen* for the census and *sur* for the survey). If we neglect the scintillation, this expected SNR is the best SNR able to obtain and is indicated in Tables 6.6 and 6.7 in the columns SNR_0.

Table 6.7: List of the 13 undetected pulsars localized in the sky area of a processed pointing. SNR_0 designates the expected SNR for a beam pointed on the pulsar, and SNR_{exp} is the expected SNR corrected of the angular separation.

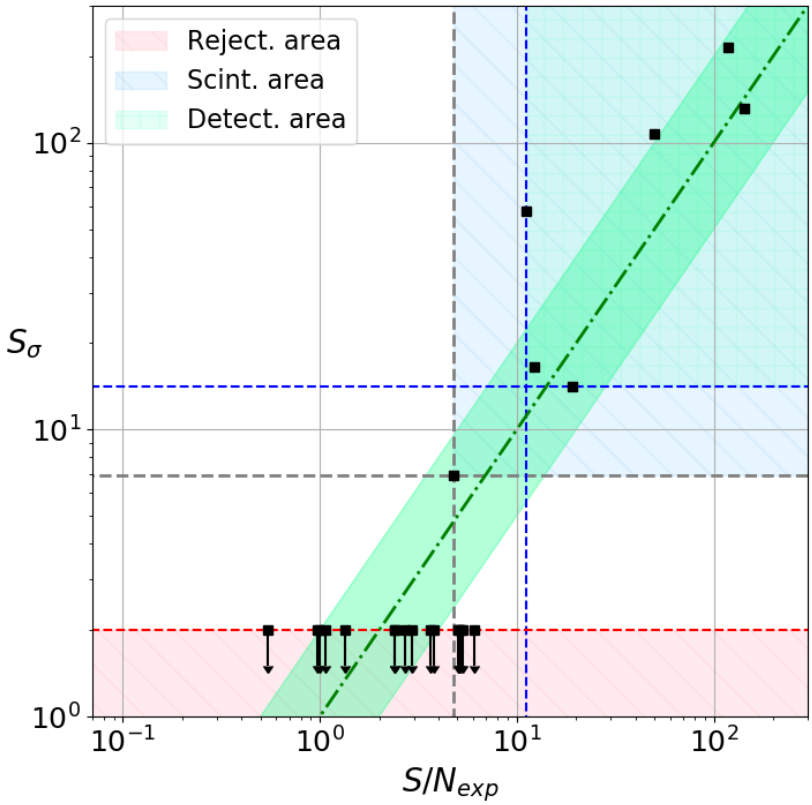
PSR	**P$_0$** (s)	**DM** (pc.cm^{-3})	**SNR$_0$**	**Ang. sep.** (°)	**SNR$_{exp}$**
J0518+5125	0.91251	39.240	1.2	0.347	1.1
J0653+8051	1.21444	33.319	1.1	0.726	0.5
J0740+6620	0.00289	14.962	0.1	0.796	0.1
J0742+4334	0.60619	36.260	11.0	0.773	5.1
J0815+4611	0.43424	11.270	10.2	0.892	3.6
J0921+6254	1.56799	13.154	9.5	0.670	5.3
J1059+6459	3.63117	18.500	5.7	0.756	2.7
J1242+39	1.31000	26.000	8.2	0.973	2.4
J1343+6634	1.39410	30.030	13.8	0.880	5.1
J1426+52	0.99580	25.370	5.5	0.228	5.2
J1518+4904	0.04093	11.611	1.4	0.512	1.0
J1628+4406	0.18118	7.330	8.5	0.793	3.8
J1821+4147	1.26186	40.673	1.9	0.503	1.4

However, this best SNR to expect is valid if the pulsar is located at the center of the numeric beam, as in the context of the census. The pointings of the survey are not placed relative to the pulsars, and consequently, it is required to take the angular separation (indicated in the columns *Ang. sep.*) between the position of the pulsar and the center of the pointing into account. For this reason, for each of the 19 pulsars, a corrected SNR was defined to estimate the loss in sensitivity of the digital beam relative to the angular separation. This loss has been evaluated using a Gaussian approximation of the digital beam.

$$SNR_{exp} = SNR_0 \cdot \exp\left(-\frac{\theta^2}{2\alpha^2}\right) \tag{6.18}$$

Where θ is the angular separation, and α is the angular beam radius. The corrected SNR (indicated in the columns SNR_{exp}) can strongly decrease the best expected SNR in the casea of a pulsar actually located in a pointing of phase 2 of the observing program.

Figure 6.18: Comparison between the significance of detection with the expected SNR for the 19 known pulsars present in the processed data of the NPBS. The light green area corresponds to the sure detectability area and the light blue area to the detectability dependent on the scintillation. The green diagonal contour marks slope 1 between the significance and the SNR. The red area represents the area of significances lower than 2σ, where a candidate is rejected.



One can notice that all the undetected pulsars listed in Table 6.7 have a very low expected SNR, with a maximum SNR of 5.3. It is then possible to estimate a limit of detectability of the search pipeline. Figure 6.18 represents the significance calculated by PREPFOLD compared to the expected SNR corrected relative to the angular separation. In this graph, the green area located on the top right part of the plot indicates the expected SNR and significances where we can assume it is sure to detect the pulsar.

However, the faintest detected pulsar, J1813+4013, has an expected SNR of only 4.8, which is less than the J0921+6254 with an SNR of 5.3. This difference in the significance of the detection is certainly due to the effect of the scintillation, which randomly increases or decreases the flux of the pulsar during a certain time lapse. The conclusion is we can't define a precise limit for the minimum detectable SNR. In Figure 6.18, this ambiguous area is colored in pale blue and corresponds to the expected SNR and significances where, because of the scintillation, it is unsure to detect the pulsar.

To conclude, we can assume the search pipeline of the NPBS can detect all the pulsars having an SNR greater or equal to 11.1 and might appear with a final significance greater than 14.10σ. Furthermore, pulsars with an SNR of at least 4.8 can potentially be detected with a final significance of 6.9 and might appear with a final significance greater than 14.10σ.

Résumé du chapitre :

Les observations sont réalisées en utilisant le mode *dynamic spectrum*. Les données obtenues sont ainsi donc des spectres dynamiques correspondant à des plans temps-fréquence. Le traitement des données doit nettoyer les données, puis appliquer une méthode de recherche de signaux pulsars basés sur la transformation de Fourier rapide (FFT pour fast Fourier transform). Le programme de traitement est basé sur la suite de programmes dédiée à la recherche de pulsar PRESTO (Ransom 2011).

Aux fréquences de NenuFAR, la dispersion génère d'important retards temporels. Aussi, l'étalement des impulsions dû au scattering devient prépondérant. Pour diminuer l'impact de ces effets, les données des observations sont réalisées avec une résolution fréquentielle accrue 1,52 kHz, et en contrepartie, avec une résolution temporelle réduite de 10 ms.

Toutes les six minutes, le faisceau analogique du télescope doit être repointé vers la nouvelle position dans le ciel. A ce moment, il se présente un saut d'amplitude dû au redémarrage de l'électronique des antennes. Afin d'améliorer les statistiques des données qui seront utilisées pour la recherche de RFI, les données ont été mise à plat en utilisant une moyenne glissante par convolution de la série temporelle avec une gaussienne.

La seconde étape consiste en la recherche de RFI, dont le but est de retirer un maximum de signal non-astrophysique. Cette étape est réalisée avec RFIFIND qui identifie les RFI par comparaison des valeurs de médiane, d'écart-type, et de puissance Fourier entre différents bloques de données. Les seuils d'identification ont été fixés de manière à garder un maximum de signaux astrophysiques transitoires puissants.

Dû à la dispersion des impulsions, il est nécessaire de dédisperser les données en utilisant un plan de dédispersion afin d'optimiser le temps de calcul. Cependant, la méthode usuelle de définition du plan de dédispersion n'étant pas adaptée aux fréquences de NenuFAR, un nouveau plan a été calculé en prenant en compte la dispersion réellement subie par les impulsions. De plus, le scattering peut également générer un étalement de l'impulsion de plusieurs centaines de milli-secondes dès des mesures de dispersion faibles, rendant difficile la détection de la majorité des pulsars. Finalement, les données ont été dispersées jusqu'à une mesure de dispersion de 70 $\mathrm{pc.cm^{-3}}$ avec un pas de 0,01 $\mathrm{pc.cm^{-3}}$.

Sur chaque série temporelle dédispersée, une recherche de signaux périodiques est réalisée en utilisant une FFT, en utilisant la sommation incohérente d'harmonique (Taylor & Huguenin 1969) jusqu'à 32 harmoniques. Finalement, les signaux de période contenue entre 30 ms et 30 s, apparaissant dans au moins trois séries temporelles contiguës, et ayant une amplitude Fourier supérieure à 2σ sont gardés comme candidats.

Pour chacun des candidats trouvés, les données originales mises à plat sont empilées à la période trouvée et dédispersée à la mesure de dispersion trouvée en utilisant PREPFOLD. Les données sont ré-échantillonnées de manière à optimiser le rapport signal sur bruit dans le but d'optimiser les recherches fines effectuées autour des valeurs trouvées par les étapes précédentes. Une recherche fine de la période et de la dérivée de la période est effectuée dans un intervalle représentant 0,2 % de la période initialement trouvée ; et une recherche fine de la mesure de dispersion est également effectuée dans un intervalle de 2 $\mathrm{pc.cm^{-3}}$.

Le traitement des données est effectué sur une machine dédiée. Afin de diminuer le temps de calcul, de multiples tests de temps et puissance de calcul ont été réalisés, permettant d'optimiser le traitement en minimisant les temps morts. Chacune des étapes précédemment exposées a été individuellement parallélisée en utilisant des multi-processing en Python, puis encapsulé dans une seconde parallélisation en bash.

Au 31 août 2022, huit mois de données, soit 2 283 pointages, ont été traités. En répartissant les deux parallélisations, tout en ajoutant des décalages dans les démarrages des traitements pour chaque étape, le traitement des données réalisées a été accéléré, représentant 40 % du temps de calcul initial.

Sur les données actuellement traitées, une phase de validation des observations et procédés mis en œuvre a été effectuée en cherchant les pulsars connus contenus dans les pointages traités, et précédemment détectés par le NenuFAR pulsar census (Bondonneau et al. in preparation). Dans les pointages ciblés, les caractéristiques des candidats ont été rapprochés des caractéristiques des pulsars connus. Finalement, 46 candidats correspondant à six pulsars connus ont été identifiés.

19 pulsars connus sont normalement présents dans les pointages traités. Cependant, après correction des différences de paramètres observationnels par rapport au census, et correction de l'écart de position entre le pulsar et le pointage, il est apparu que les rapports signal sur bruit des pulsars non détectés étaient bien effectivement les plus bas. Des limites préliminaires de capacité de détection du relevé ont ainsi pu être évaluées. En terme de rapport signal sur bruit, une limite minimale peut être estimée à une valeur de 4,8. Aussi, en terme de période, au vu des candidats correspondant à la détection d'un pulsar connu, une limite minimale peut être évaluée à environ 40 ms.

Chapter 7

Analysis of the blind survey candidates

7.1 Context and method

7.1.1 Context

All (or at least the majority) of the pulsars to be discovered by the NenuFAR pulsar blind survey are expected to be weak. In the data processing, the choice was consequently made to use low constraints in the candidate selection. For this reason, it is inappropriate to only use the Fourier power (the normalized coherent power found in the FFT) of a periodic candidate as selecting criterion. A pulsar with a weak flux will have a weak Fourier power, when a non-pulsar signal which is passed through the RFI mitigation, could appear with a high Fourier power. For example, before the flattening step was well parametrized, it sometimes could appear strong candidates, which were actually harmonics of the residuals of the analog jumps of NenuFAR.

The folding stage is thus important to go further than the simple Fourier selection (even followed by a good sifting stage). It allows testing eight different features of the signal of the candidate, in order to compare with the normal behavior of a pulsar signal. To sort all these candidates, it needs then to analyze the output results of PRESTO (see Figure 6.13 in Chapter 6.3.6).

Because of the low constraints, the processing pipeline finds an enormous number of candidates. For the whole blind survey, i.e. for the two phases of observation, about 500 000 candidates are expected. The majority of those are not pulsars. As a consequence, the blind survey needs an automatic program to identify the likely pulsar signals hidden in this huge amount of resulting data.

For this reason, I have developed a Python script applying a method to analyze the data used to create each output plot of PREPFOLD. This allows excluding most of the candidates which do not look like a pulsar signal.

7.1.2 Method

PREPFOLD generates eight different resulting plots (see Figure 6.13 in Chapter 6.3.6):

1. an integrated profile,
2. a time-phase plane,
3. a frequency-phase plane,
4. the evolution of the χ^2 of the profile as a function of time,
5. the evolution of the χ^2 of the profile as a function of DM,
6. the evolution of the χ^2 of the profile as a function of the period,
7. the evolution of the χ^2 of the profile as a function of the period derivative,
8. a $P - \dot{P}$ plane of the χ^2 of the integrated profile.

The method of analysis of a candidate consists of the analysis of data from each of the eight plots. For each plot, a Monte-Carlo simulation by global coverage of the parameter space (rather than by directed Markov chain) is realized to compare the data of the candidate with plenty of possible signals of pulsars.

The parameter space is populated using a random generation of each parameter of the model using the Python package STATS of SCIPY. The parameters are generated inside a range of typical values for the known pulsar population, following one of the three prior distributions:

- normal distribution (denoted as $\mathcal{N}$ in the different tables),
- half-normal distribution (which is a normal distribution but only for the values greater than the mean, and denoted as $\mathcal{HN}$),
- or uniform distribution (denoted as $\mathcal{U}$).

The data of the candidate are extracted from the PFD file generated by PREPFOLD using PREPFOLD.PY, which is the Python version of the corresponding C file. The extracted raw data are then processed to recreate each result plot. Some of these plots are created using preexisting functions implemented in the Python script. For those without existing functions, as for the computing of the $P - \dot{P}$ plane, based on the preexisting Python functions and the process performed in the C file, I have written and added the missing functions in PREPFOLD.PY.

To evaluate the similarity of the signal of the candidate with the signal of a pulsar, an estimator of the maximum likelihood ratio is employed. For each model, the χ^2 between the data $D(x, y)$ and the simulated model $M(x, y \mid \theta)$ is thereby calculated.

$$\chi^2(\theta) \;=\; \sum_{i}^{N_i} \sum_{j}^{N_j} \frac{(\; D(x_i, y_j) - M(x_i, y_j \mid \theta)\;)^2}{Var(D(x, y))} \tag{7.1}$$

With N_i and N_j respectively the numbers of bins in the directions x and y, and $\theta \;=\; \{\theta_1, \theta_2, \cdots \theta_i, \cdots, \theta_n\}$ is the set of parameters defining the model. Each of these parameters θ_i takes the values randomly drawn depending on the prior function of the parameter. The best model is then identified as those with the minimum χ^2, and the statistical test is determined as the ratio of the probability of the best model relative to the probability of the null model. The null model is defined as a signal without any pulsar signal, i.e. comprising only noise. It is determined by a similar Monte-Carlo simulation, where the model is just a flat signal corresponding to a constant baseline. The level of this baseline is defined following the same prior as for the pulsar model, and the best null model corresponds to the baseline with the minimum χ^2.

The probabilities relating to the obtained χ^2 values are calculated by the cumulative distribution function of the χ^2 law, representing the probability to have a χ^2 lower than the considered value (Leahy et al. 1983):

$$L_b \;=\; Pr(\; \chi^2 \;<\; \chi_b^2\;) \;=\; \int_0^{\chi_b^2} \gamma_k(\chi^2) \cdot d\chi^2 \tag{7.2}$$

$$L_0 \;=\; Pr(\; \chi^2 \;<\; \chi_0^2\;) \;=\; \int_0^{\chi_0^2} \gamma_k(\chi^2) \cdot d\chi^2 \tag{7.3}$$

Here γ_k is the probability density function of the chi-square law for k degrees of freedom. L_b represents the probability related to the best model of value χ_b^2, and L_0 the probability related to the null model of value χ_0^2. The statistical test can be then related with a standard normal variable (Li & Ma 1983), allowing to define the significance α:

$$\alpha \;=\; \sqrt{-2 \ln\left(\frac{L_b}{L_0}\right)} \;\mid\; L_b \;<\; L_0 \tag{7.4}$$

For each resulting plot of PREPFOLD, in the case of a model with a better χ^2 than the null model, the significance α is computed with Equation (7.4). In the opposite case, the significance is set to 0.

Finally, to define a good potential pulsar candidate, the signal of the candidate needs to show similarities with a pulsar signal in more than one of the eight result plots. A global significance Σ is therefore calculated by the linear combination of the significances previously computed for each plot.

$$\Sigma = \sum_{i=1}^{8} \zeta_i \cdot \alpha_i \tag{7.5}$$

Here, ζ_i is the weight coefficient related to the resulting plot i. These coefficients have been determined by training the analysis method on a set of observations done in real conditions during December 2020 (see Section 7.3.3).

7.1.3 Advantages and drawbacks of a Monte-Carlo method

The analysis presented here aims to estimate if a candidate can plausibly be a pulsar. Also, the combination of a Monte-Carlo simulation with a method of maximum likelihood ratio allows having a good similarity test, which is quite robust and allows obtaining a good classification of the candidates.

The parameter spaces of the models are relatively small in the number of dimensions, with between three and six parameters. Also, the possibilities of what should look like the signal of a pulsar provides relatively strong constraints. This allows it to properly cover the space with a reasonable computing time of some months for the whole survey. Consequently, the choice was made to use a Monte-Carlo simulation by global coverage of the parameter space rather than by a Markov chain. This latter is indeed especially interesting for large parameter spaces with an important number of parameters. With a Markov chain, the algorithm searches an extremum by random walk (even in the directional case), and as an optimization algorithm, can find just a local extremum. The advantage of the chosen method here is it allows finding the global extremum.

To classify a set of data, it is also possible to use other types of algorithms such as, for example, neural networks, machine learning, deep learning,... However, the first advantage of a Monte-Carlo simulation is it is easier to develop compared to the complex algorithms of neural networks and others. Moreover, a neural network is interesting in the context of a wide range of possibilities, and therefore without an underlying mathematical model. On the contrary, Monte-Carlo simulations are very efficient in the opposite context, where a precise mathematical model is known.

The main drawbacks of the Monte-Carlo simulations are the drawbacks of the Bayesian inference. The accuracy of the result of the simulation is highly dependent on the good definition of the model and priors, that is in terms of choice of the probability density function, domain of definition, or in the coverage of the parameter space.

Monte-Carlo simulations are mathematical methods, while neural networks (and associated methods) rather are empirical methods. As a consequence, in order to correctly constrain and fine-tune the algorithm, it needs to perform training allowing it to obtain efficient learning. The training set must be thereby sufficiently large and varied to cover all possible situations.

In our case, we don't possess a set of training observations that is sufficiently large and varied. On the other hand, we know the behavior, and consequently, the associated mathematical form of a pulsar signal for each resulting plot. Combining with a small parameter space, a Monte-Carlo simulation seems to be a better choice, that is in terms of complexity, required accuracy, and time computing.

7.1.4 Asymmetric pulsar profiles

Intrinsic asymmetry The used mathematical models are based on the shape of the profile of a pulsar. An ideal profile is a narrow Gaussian, but a real profile is more complex and shows asymmetries. The first one is due to the intrinsic asymmetry of the peak, which can be slightly deformed compared to a perfect Gaussian shape. However, compared to the two next presented asymmetries, this one is very weak, leading to a rather correct approximation by a Gaussian shape.

In addition to the intrinsic asymmetry of the peak, a lot of pulsars present a profile with several peaks. Each peak is different, in shape or intensity, resulting in an asymmetric shape of the envelope of the profile. However, all the pulses are close to each other, so that, the global shape of the profile can be approached by a Gaussian envelope. Furthermore, the time sampling of 10.486 ms doesn't allow distinguishing peaks closer to 20 ms (two samples). Therefore, in the majority of cases, multi-peak profiles will appear as a single peak profile featuring a light asymmetry. The envelope can be thus approached, as for the intrinsic asymmetry of the peak, by a Gaussian shape.

Asymmetry due to the scatter broadening The third reason and the main one at the frequencies of NenuFAR, is due to the scattering of the pulse. The scattering effect produces a multipath propagation (Scheuer 1968), where each different path arrives with a time delay compared to the initial wave. Unlike the two previous asymmetries which were intrinsic to the pulsar, this last one is caused by the ISM. The asymmetry generated by scatter broadening can be thus described by the convolution of the intrinsic pulse with the transfer function representing the scattering by screens in the ISM. The global shape of the measured profile is finally the intrinsic shape with an exponential tail towards the high rotational phases.

The scatter broadening modifies the profile shape compared to a pure Gaussian. The scatter broadening will increase the pulse width, leading to an increase in the duty cycle of the pulsar. The amplitude of the asymmetric variation, and then the possibility to fit the profile with a Gaussian, is thereby relative to the duty cycle of the pulsar. In the case of a low duty cycle, the asymmetry is low and a Gaussian is sufficient to fit the shape of the pulse.

At the other end, a very strong scattering can result in a duty cycle greater than one, i.e. resulting in a pulse width larger than the period of the pulsar. In this case, the pulse shape cannot be disentangled from a variation of the baseline. In any case, a large part of these pulsars are therefore outside of the possibility of detection of an FFT.

In the case of a large duty cycle of less than one, the asymmetry is important and can be potentially disentangled from a variation of the baseline. However, for a weak pulsar and due to the coarse time sampling of the survey, the profile of the pulsar will be smoothed, and finally misshapen. This type of pulsar is expected to appear in the same way that the observation of the MSP J0034-0534 (left panel of Figure 7.1), which has a large duty cycle due to the scattering. Hence, this type of pulsar can be properly fitted with a wide Gaussian. Also, for an intense pulsar (i.e. having a high radio flux density), the exponential tail clearly appears, as for B2217+47 in the right panel of Figure 7.1. Nevertheless, the analysis presented here doesn't aim to adjust the parameters of the profile but only to discriminate a potential pulsar signal from a noise signal. Thus, even in the case of an intense pulsar, it necessarily exists a Gaussian which will be largely better than the null model, and resulting in a high significance.

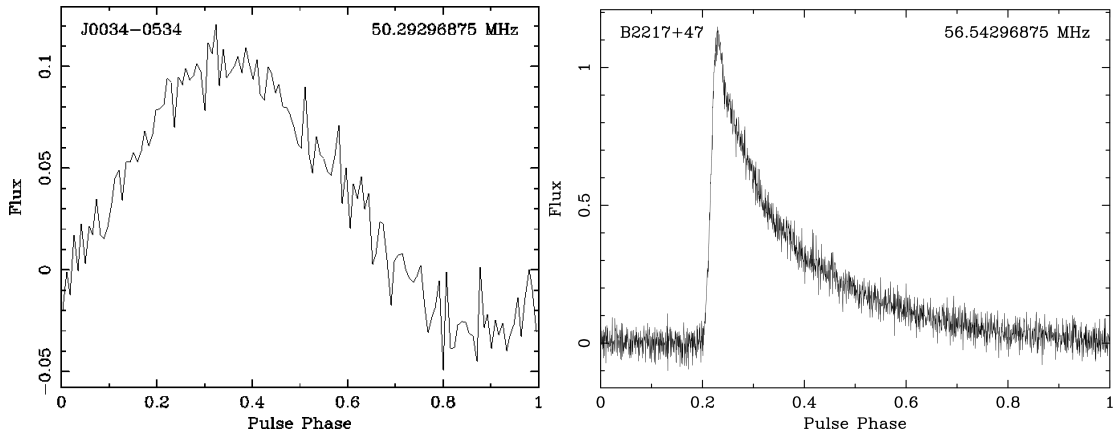


Figure 7.1: Left: smoothed profile of J0034-0534 showing a high duty cycle due to scatter broadening. Right: profile of B2217+47 showing a clear exponential tail due to scatter broadening in addition to the starting Gaussian peak.

The asymmetry caused by the scatter broadening could be problematic for certain pulsars which are in the grey area between the different cases explained above. For these pulsars, it can be difficult to find a Gaussian allowing to obtain sufficient significance to be identified as a potential pulsar signal. However, because of the coarse time sampling and the low-frequency constraints in terms of dispersion, we expect rather slow pulsars with a period of at least some hundreds of ms, and with a low DM of less than 70 $pc.cm^{-3}$. As a consequence, we especially expect pulsars with a low duty cycle, corresponding to a low scattering effect.

Efficiency One needs to keep in mind that this analysis does not aim for a precise description of the profile, but just to provide an estimation of: *"Does the candidate look more like a pulsar or noise?"* Moreover, this process shall be repeated for about 500 000 candidates. As a consequence, the process of analysis must be light and fast.

Include the asymmetry added by the scattering asked a more complex function called *exponentially modified Gaussian function* (EMG hereafter). This function is longer and more complex to evaluate than a simple Gaussian function (more details are given in section 7.2.1). Furthermore, this model involves an additional parameter allowing to describe the asymmetry. Consequently, the parameter space is larger, lengthening global computing too.

For the majority of the pulsars, using a Gaussian function as the based model will be sufficient to obtain a pulsar model better than the null model, and then obtain a correct significance. Using a Gaussian model yields a loss in significance compared to an EMG model. This loss is relevant only for pulsars similar to B2217+47 (see Figure 7.1), i.e. with a clear deviation relative to a Gaussian shape. Nevertheless, this type of pulsar could be missed only whether they have a period long enough to avoid seeing the exponential tail, and with a weaker flux generating a too-low significance. However, a too-low flux or period generates, for those which are detected, a smoothing effect as the MSP in the left plot in Figure 7.1. As a consequence, this type of pulsar should be rather rare.

The use of the EMG function as the base model for the profile requires a significant increase in computing time to correctly cover the parameter space with the same efficiency as for a Gaussian function. Thus, comparing the additional computing time and the low additional number of pulsars potentially detected by the EMG, it seems that the EMG function is a rather inefficient relative to the simple Gaussian function. As a result, for the analysis presented in this chapter, the EMG has been unused, and the Gaussian model has been preferred.

7.2 Models

7.2.1 Integrated profile

Model function The first control diagram of PREPFOLD is the integrated profile of the candidate. This profile results from the folding on p_l bins to the period of the candidate with integration in frequency and time sub-integrations. There are a large variety of profiles from one to several peaks, such as B1919+21 presenting five peaks. However, due to our coarse time sampling of 10.486 ms, a profile featuring multiple peaks can almost all the time be represented by a single smooth peak.

The profile of a pulsar can be precisely expressed by a model $f_1(\phi)$ with an EMG function, corresponding to the analytic form of the convolution between a normal distribution (the intrinsic pulse) and an exponential distribution (the scatter broadening).

$$f_1(\phi) \;=\; \frac{\lambda}{2} \cdot \exp\left\{ \frac{\lambda}{2} \cdot (2\mu + \lambda \cdot \sigma^2 - 2\phi) \right\} \cdot \Phi(\, \mu + \lambda \cdot \sigma^2 - \phi \,) \tag{7.6}$$

With ϕ the rotational phase, μ the normal distribution mean, σ the normal distribution standard deviation, and λ the exponential rate. The function Φ refers to the complementary error function expressed as:

$$\Phi(x) \;=\; \frac{2}{\sqrt{\pi}} \cdot \int_x^{\infty} e^{-t^2} \cdot dt \tag{7.7}$$

Table 7.1: Parameters used for the model of the integrated profile, with the explored parameter space and the prior distribution for each parameter. p_l represents the number of bins in the profile.

Parameter	**Parameter space**	**Prior**
A_1	$[\ 3\ ;\ +\infty\ [$	$\mathcal{N}(\ 5\ ,\ 5\ )$
σ_1	$\left[\ \frac{1}{p_l}\ ;\ 0.5\ \right]$	$\mathcal{HN}\left(\ \frac{1}{p_l}\ ,\ 0.3\ \right)$
$\phi_{0,1}$	$[\ 0\ ;\ 1\ ]$	$\mathcal{U}$
B_1	$[\ \text{-}1\ ,\ 1\ ]$	$\mathcal{N}(\ 0\ ,\ 1\ )$

As said previously, due to the longer computing time required by the EMG compared to the potential increase in detection (see Section 7.1.4), the profile model is then approximated with a simple normal law of the form:

$$f_1(\phi) \ = \ A_1 \cdot \exp\left\{ \ -\frac{(\phi - \phi_{0,1})^2}{2\sigma_1^2} \ \right\} + B_1 \tag{7.8}$$

Where A_1 is the amplitude of the profile, $\phi_{0,1}$ represents the rotational phase of the maximum of the profile, σ_1 is the width of the profile, and B_1 is the background signal (i.e. the baseline of the profile).

Parameter space A detectable pulsar must have a profile that stands out from the background noise. The profile is normalized relative to the standard deviation of the data, leading to obtaining amplitudes in SNR. Hence, to be considered as a potential pulsar, the amplitude A_1 must have a value of at least 3. This value corresponds, in terms of Gaussian probability, to the fact that 99.73% of the noise is statistically below this limit. For a signal with an amplitude less than 3, it could be a very weak pulsar, but we cannot be sure it is not just a fluctuation of the baseline. We thus require for such a signal that the significance is low.

To be considered as a possible real pulse profile of a pulsar, the profile must have more than one point above the background (i.e. with an amplitude above at least 1σ). In other words, we require that the minimum half-width of the Gaussian be of one profile sample, corresponding to a width σ_1 of $1/p_l$ in the rotational phase. For the maximum width, a profile with a width larger than half of the period can't be disentangled from a wide variation of the baseline. A pulsar with such a profile consequently can't

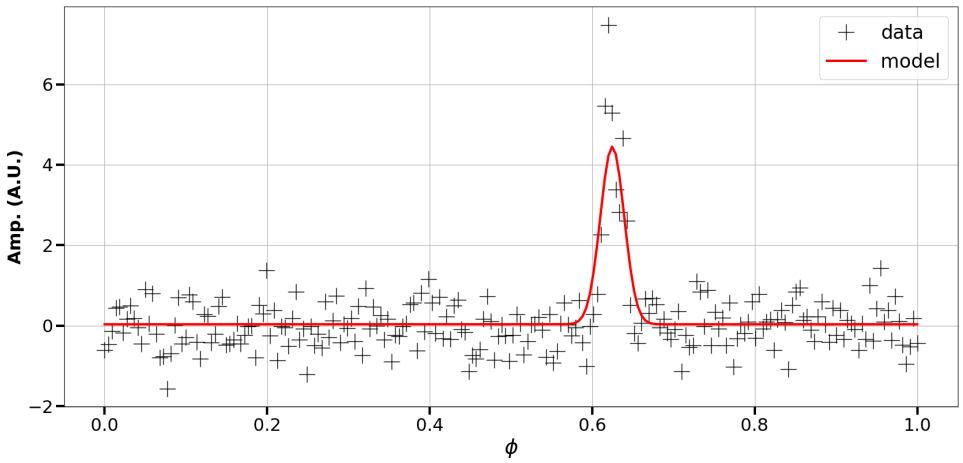


Figure 7.2: Result of the Monte-Carlo simulation for the integrated profile of J0323+3944. The initial data are in black crosses and the best model is shown by the red line.

be normally detected by the FFT. As a result, the upper limit for the width of the Gaussian is set to $\sigma_{1,max} = 0.5$.

Concerning the position of the profile peak $\phi_{0,1}$, it can be placed at any position in the entire profile. There are, therefore, no particular constraints on this parameter, which can take any value between 0 and 1.

The last parameter of the model is the level of the background B_1. It results from the normalization that the amplitude of the background signal B_1 must be close to 0. However, because of random fluctuations of normal white noise and the impact of the pulse, the level of the baseline can be slightly modified compared to the ideal case. As a consequence, the level of the background can be set from -1 to 1, i.e. in the range of weak fluctuations of the noise.

Priors We require a pulsar signal to be greater than 3 and is expected to be at least about 5 for a clear detection. The amplitude can considerably vary from the weakest to the strongest pulsars, leading to a wide prior function. As a consequence, the prior function has been set to a normal function of mean equal to 5 and a width of 5, allowing to reach SNR up to several dozens.

Because of our observing and search parameters, we expect to detect rather slow pulsars with a weak scatter broadening. This corresponds to integrated profiles with a rather small duty cycle, which are also the easiest periodicity candidates to find for the FFT. The prior function for the profile width has been therefore set to a half-normal distribution of mean $1/p_l$ and width 0.3, allowing to search for potential wide profiles and scattered pulsars.

Because there are no constraints on the position of the profile, the parameter $\phi_{0,1}$ is defined with a uniform distribution between 0 and 1.

Although there are no constraints on the value of the baseline within the defined parameter range, the normalized level of the baseline should preferentially be close to the ideal value of 0. As a consequence, for the prior of the parameter B_1, a normal distribution with a mean of 0 and a width of 1 has been preferred compared to a uniform distribution.

Table 7.1 summarizes the parameter spaces and prior functions for each parameter of the simulation. A result of the Monte-Carlo simulation of the integrated profile for a candidate of a test observation comprising the pulsar J0323+3944 is presented in Figure 7.2 where the best model found is superposed with the real data.

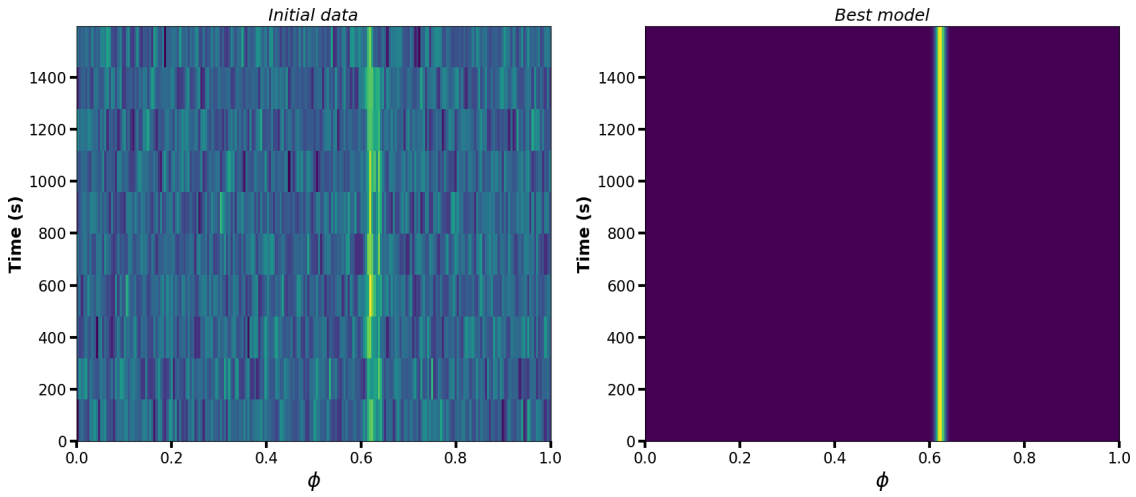
7.2.2 Time vs phase plane

Model function The second control diagram is the time-phase plane, showing N_p successive profiles integrated in frequency and folded to a sub-integration time T_{obs}/N_p. A time-continuous RFI will appear in all profile bins in all sub-integrations. On the opposite side, as for the profile, a pulsar will appear in only some bins, and, in addition, continuously for the duration of the observation.

Table 7.2: Parameters used for the model of the time-phase plane, with the explored parameter space and the prior distribution for each parameter. p_l represents the number of bins in the profile.

Parameter	**Parameter space**	**Prior**
A_2	$[\ 3\ ;\ +\infty\ [$	$\mathcal{N}(\ 5\ ,\ 15\ )$
σ_2	$\left[\ \frac{1}{p_l}\ ;\ 0.5\ \right]$	$\mathcal{HN}\left(\ \frac{1}{p_l}\ ,\ 0.3\right)$
$\phi_{0,2}$	$[\ 0\ ;\ 1\ ]$	$\mathcal{U}$
B_2	$[\ \text{-}1\ ,\ 1\ ]$	$\mathcal{N}(\ 0\ ,\ 1\ )$

Figure 7.3: Result of the Monte-Carlo simulation for the time-phase plane of J0323+3944. Left: initial data. Right: best model found.



The 2D model $f_2(\phi, t)$ is indeed an extension of the model of the profile $f_1(\phi)$ (see Equation 7.8), where a time dimension t is added. Except for RRATs and pulsars with nulling effect, a pulsar is normally rather constant over time. This time dependence is thus simply modeled with a boxcar function $\Pi(t)$.

$$f_2(\phi, t) \;=\; \Pi(t) \cdot f_1(\phi) \;=\; \Pi\left(\frac{t - T_{obs}/2}{T_{obs}/2} \right) \cdot \left[\frac{A_2}{N_p} \cdot \exp\left\{ -\frac{(\phi - \phi_{0,2})^2}{2\sigma_2^2} \right\} + B_2 \right] \tag{7.9}$$

With A_2, σ_2, $\phi_{0,2}$, and B_2 the corresponding parameters of the function f_1 for the function f_2. $T_{obs}/2$ represents the middle time of the observation, leading to defining the boxcar function between 0 and T_{obs}. The global amplitude of the profile A_2 needs to be divided by the number of sub-integrations N_p, in order to take into account the fact that the final amplitude of the profile is spread over all the time sub-integrations.

Parameter space Because of the high similarity between the functions f_1 and f_2, the parameter spaces are identical for all the parameters. Also, a pulsar signal being constant during the entire observation, the boxcar function doesn't add any parameter and is numerically just a unity function.

Priors Concerning the priors associated with the parameters σ_2, $\phi_{0,2}$, and B_2, they are also the same as for the integrated profile.

There is only a difference for the parameter A_2, in order to take account of the distribution of the intensity of the pulse in the time dimension.

The amplitude parameter A_2 is similar to the amplitude parameter of the model $f_1(\phi)$. However, for this model, the Gaussian term in rotational phases is related to one time sub-integration, leading that the amplitude being distributed in several time sub-integrations N_p. Moreover, the normalization is done sub-integration by sub-integration. As a consequence, it results that the amplitude parameter may have a wider range of values. The prior function is still a normal distribution of mean 5 but with a width of 15.

Table 7.2 summarizes the parameter spaces and prior functions for each parameter of the simulation. A result of the Monte-Carlo simulation of the time-phase plan for the pulsar J0323+3944 is presented in Figure 7.3 with the real plane in the left panel and the best model found in the right panel.

7.2.3 Time evolution of the χ^2

Model function The third control diagram is related to the second, and allows to control the evolution of the profile in time. As for the time vs phase plane, contrary to an RFI, we require the signal of a pulsar to be constant in time.

Moreover, except for pulsars producing occasional giant pulses, the flux of the pulsars is unsupposed to dramatically change in 30 minutes. The flux must be then relatively constant during the observation, resulting in a χ^2 of the profile in each time part expected to be equally constant. As a consequence, the cumulative χ^2 during the observation must increase linearly with time.

To check the behavior of the candidate compared to the expected linear evolution, the model of evolution is thereby a simple linear function $f_3(t)$:

$$f_3(t) \;=\; A_3 \cdot t + B_3 \tag{7.10}$$

With A_3 the time derivative of the χ^2 of the profile, t the time, and B_3 the starting χ^2 at the beginning of the observation. Numerically, the model is discrete as a function of the time part index, where each value of f_3 is computed for a time sub-integration. As a result, the amplitude parameter needs to be normalized relative to the number of time sub-integrations N_p.

Parameter space As for the amplitude of the two previous models, the amplitude must be at least 3 in SNR at the end of the observation. As a consequence, the lower limit for the parameter of increase of the χ^2 is defined as $3/N_p$.

Concerning the background parameter B_3, the reasoning is the same as for the previous background parameters. Actually, the evolution of the χ^2 is normalized by PREPFOLD relative to the first sub-integration, in order to have a starting χ^2 equal to 1. The parameter B_3 must thereby be very close to 1 and is defined in the range $0.5 - 1.5$.

Priors The time derivative of the amplitude A_3 is defined in the same way as the amplitude of the total amplitude A_1 with a similar normal distribution. The mean increase of the amplitude is set to 5 by time sub-integration, leading to a final χ^2 of 5. The width is equally set to $5/N_p$, allowing it to easily reach a final χ^2 of $20 - 30$.

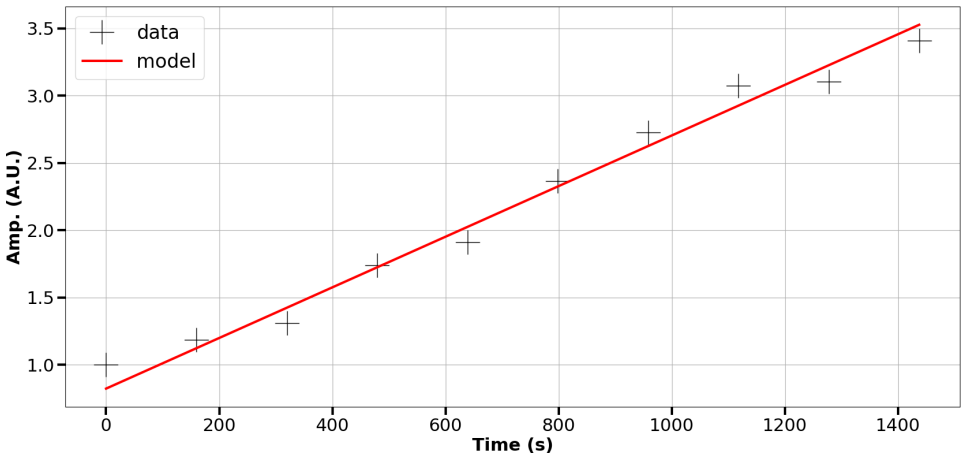
The background parameter B_3 is also defined in the same way that the parameter B_1, i.e. with a wide normal distribution promoting the central ideal value. The difference is the narrower parameter space and the central value of 1 rather than 0. The normal distribution is then set with a mean of 1 and a width of 0.5.

Table 7.3 summarizes the parameter spaces and prior functions for each parameter of the simulation. A result of the Monte-Carlo simulation of the evolution of the χ^2 over time for the pulsar J0323+3944 is presented in Figure 7.4 where the best model found is superposed with the real data.

Parameter	Parameter space	Prior
A_3	$\left[\frac{3}{N_p} \; ; \; +\infty \right[$	$\mathcal{N}\left(\frac{5}{N_p} \, , \, \frac{5}{N_p} \right)$
B_3	$[\, 0.5 \, , \, 1.5 \,]$	$\mathcal{N}(\, 1 \, , \, 0.5 \,)$

Table 7.3: Parameters used for the model of the time evolution of the χ^2, with the explored parameter space and the prior distribution for each parameter. N_p represents the number of time sub-integrations.

Figure 7.4: Result of the Monte-Carlo simulation for the evolution of the χ^2 over the time for J0323+3944. The initial data are in black crosses and the best model is shown by the red line.



7.2.4 Frequency vs phase plane

Model function The fourth diagram is similar to the time-phase diagram, i.e. it is an extension of the integrated profile $f_1(\phi)$. The time-phase diagram was an extension of the profile in phase ϕ in the time dimension. For this control plot, the profile is this time extended in the frequency dimension ν.

As for the second model, the function $f_4(\phi, \nu)$ is the model of profile $f_1(\phi)$ modulated by a frequency-dependent function $C(\nu)$. However, contrary to the time-phase diagram where the profile is supposed to be constant in time, the evolution in frequency of a pulsar is unconstant over the entire bandwidth. The flux of the pulsars depends indeed of the spectrum of the pulsar and on the bandpass of NenuFAR.

Even if it is unconstant, a pulsar is unsupposed to have a parsimonious emission in frequency, but must have a continuous emission on a certain bandwidth. In order to characterize the continuous evolution in frequency, the modulation function $C(\nu)$ is then defined as a second-order polynomial function with a maximum within the observed band.

$$f_4(\phi, \nu) = C(\nu) \cdot (f_1(\phi) - B_1) + B_4 \tag{7.11}$$

$$\Leftrightarrow \; f_4(\phi, \nu) = (\nu - a_4) \cdot (b_4 - \nu) \cdot \frac{A_4}{N_s} \cdot \exp\left\{ -\frac{(\phi - \phi_{0,4})^2}{2\sigma_4^2} \right\} + B_4 \tag{7.12}$$

With A_4, $\phi_{0,4}$, and σ_4 the parameters corresponding to the model $f_1(\phi)$. Because of the normalization in every frequency subband, the noise is unmodulated in frequency, and the corresponding parameter B_4 is placed on the exterior of the global function. As for the time-phase model, the amplitude is split into the frequency subbands. The global amplitude of the profile A_4 is then divided by the number of subbands N_s.

For the frequency modulation, the polynomial equation is expressed in factored form, where the parameters a_4 and b_4 define the frequencies where the pulse disappears. a_4 corresponds to the lowest frequency where the pulse is detected, and b_4 is the highest frequency. Physically, when the frequency is outside of the range [a_4 ; b_4], that means that there is no pulse, leaving only the background noise. Consequently, the function $C(\nu)$ is equal to 0 for frequencies lower than a_4 or greater than b_4.

Parameter space The parameters A_4, $\phi_{0,4}$, σ_4, and B_4 define the profile in phase in the same way that for the model f_1. The parameter spaces to use for these are consequently the same as for the model $f_1(\phi)$.

The frequency bandwidth observed in the NPBS, between 39 and 77 MHz, was chosen in a part of the total bandwidth where the sensitivity of NenuFAR is high, with a maximum of bandpass around 55 MHz. Hence, if a pulsar signal is detectable, it must preferentially appear around the center of the bandwidth of the survey. Moreover, a pulsar must appear in a sufficiently large bandwidth to be disentangled from a narrowband RFI. As a consequence, for the model, the pulse is required to appear at least in the central third of the bandwidth around the central frequency of the survey of 56 MHz. The lowest frequency

Table 7.4: Parameters used for the model of the frequency-phase plane, with the explored parameter space and the prior distribution for each parameter. p_l represents the number of bins in the profile, and $\Delta\nu$ is the frequency bandwidth of the observation (37.5 MHz).

Parameter	**Parameter space**	**Prior**
A_4	$[\ 3\ ;\ +\infty\ [$	$\mathcal{N}(\ 5\ ,\ 150\ )$
σ_4	$\left[\ \frac{1}{p_l}\ ;\ 0.5\ \right]$	$\mathcal{HN}\left(\ \frac{1}{p_l}\ ,\ 0.3\right)$
$\phi_{0,4}$	$[\ 0\ ;\ 1\ ]$	$\mathcal{U}$
B_4	$[\ \text{-}1\ ,\ 1\ ]$	$\mathcal{N}(\ 0\ ,\ 1\ )$
a_4	$\left]\ -\infty\ ;\ \nu_0 + \frac{1}{3}\Delta\nu\ \right]$	$\mathcal{N}\left(\ \nu_0 + \frac{\Delta\nu}{4}\ ,\ \Delta\nu\right)$
b_4	$\left[\ \nu_0 + \frac{2}{3}\Delta\nu\ ;\ \infty\ \right[$	$\mathcal{N}\left(\ \nu_0 + \frac{\Delta\nu}{4}\ ,\ \Delta\nu\right)$

where a pulse can be seen a_4 is thereby required to be lower than $\nu_0 + \Delta\nu/3$, with ν_0 the lowest observed frequency and Δnu the bandwidth. In the opposite case, the highest frequency where the pulse can be seen b_4 is required to be greater than $\nu_0 + 2\Delta\nu/3$.

Priors The model and the space parameters for the parameters A_4, $\phi_{0,4}$, σ_4, and B_4 being identical, the prior distributions are also the same as those used for the profile model.

There is a difference concerning the width of the distribution of the amplitude parameter A_4. As for the time-phase model, the amplitude is split across the different subbands, needing to increase the width of the distribution to obtain a proper fit in the case of intense pulsars. Furthermore, contrary to the time-phase model, the amplitudes of the profiles are unconstant over the subbands. The profiles located close to the central frequency consequently possess a higher amplitude than those on the edges. The maximum allowed value of the amplitude needs then to be substantially increased compared to the time-phase model. Finally, it has been set to a normal distribution of mean 5 and half-width 150.

Finally, the two last parameters are physically the same meaning and can be defined with a similar prior distribution. The values are defined according to a normal distribution where the mean values, corresponding to the most probable edges of a pulse, are set to 1/4 of the bandwidth for a_4 and 3/4 for b_4. Intense pulsars show an almost constant profile on the entire bandwidth, leading to defining the frequency edges following a wide distribution. The half-widths of the two prior distributions are then fixed to the bandwidth.

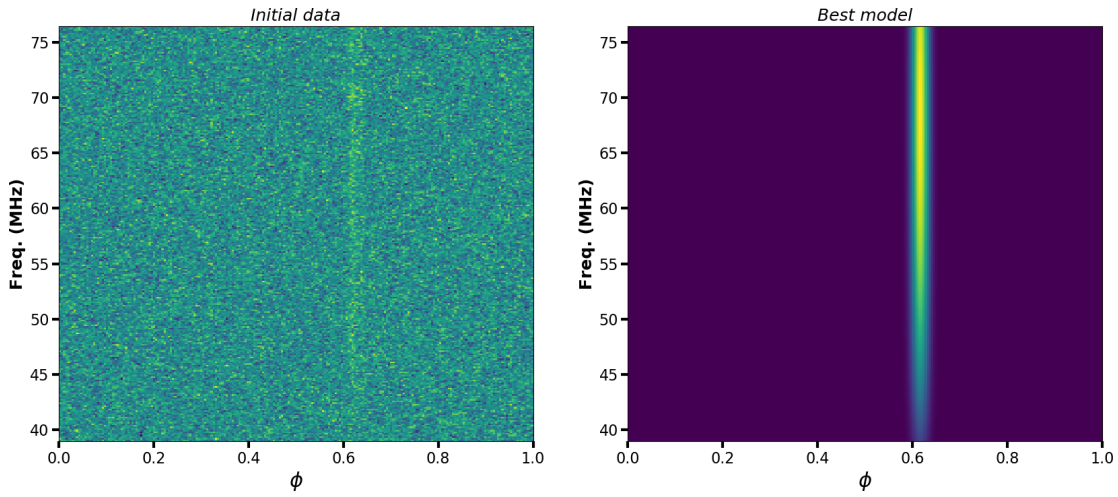
Table 7.4 summarizes the parameter spaces and prior functions for each parameter of the simulation. A result of the Monte-Carlo simulation of the frequency-phase plan for the pulsar J0323+3944 is presented in Figure 7.5 with the real plane in the left panel and the best model found in the right panel.

7.2.5 Evolution of the χ^2 relative to the DM

The four first control diagrams allow checking the behavior of the pulse in diverse ways. The four last ones are related to the fine searches and allow checking the behavior of the pulse for an error in DM, period, period derivative, or for a combined error in period and period derivative.

Model function The fifth diagram can be used to verify the evolution of the χ^2 of the profile relative to an error in DM. A typical narrowband RFI for example, appearing continuous over time does not show any variations of the χ^2 in DM. And for a broadband RFI on a certain time lapse, it will rapidly be dispersed for any DM different from 0 pc.cm^{-3}, resulting in a null impact in the χ^2.

Figure 7.5: Result of the Monte-Carlo simulation for the time-phase plane of J0323+3944. Left: initial data. Right: best model found.



For a profile of mean flux $\bar{P}$ with N_ϕ bins in rotational phase ϕ, the χ^2 is proportional to:

$$\chi^2 \;\propto\; \sum_{\phi}^{N_\phi}(\; P(\phi) - \bar{P}\;)^2 \;=\; \sum_{\phi}^{N_\phi} P(\phi)^2 - N_\phi \cdot \bar{P}^2 \tag{7.13}$$

Here P_ϕ represents the flux at the phase ϕ. In the case of bad DM, the flux is distributed on the neighbor phases, but the total flux is constant whatever the error in DM. Then, the mean flux $\bar{P}$ is equally constant whatever the error in DM. The χ^2 as a function of DM is then just given by the variation of the quadratic flux $P(\phi)^2$. It can be expressed as the maximum flux value P_0^2, corresponding to the reference quadratic flux for the case of an ideal dedispersion, more or less a fraction of this reference flux.

$$\chi^2(\delta DM) \;\propto\; \sum_{\phi}^{N_\phi}\{\; P_0(\phi) + \varphi(\delta DM, \phi)\cdot P_0(\phi)\;\}^2 - N_\phi \cdot \bar{P}^2 \tag{7.14}$$

Where $\varphi(\delta DM, \phi)$ is a function describing the fraction of flux added or removed due to the error in DM δDM in the phase ϕ. The sign of φ is positive for the phases different from the phase of good alignment ϕ_0, and negative for the phase ϕ_0 meaning the loss of flux. It allows dividing the quadratic flux variation into two different terms.

$$\chi^2(\delta DM) \;\propto\; P_0(\phi_0)^2 \cdot (\; 1 - |\varphi(\delta DM, \phi_0)|\;)^2 + \sum_{\phi\;\neq\;\phi_0}^{N_\phi - 1} P_0(\phi)^2 \cdot (\; 1 + |\varphi(\delta DM, \phi)|\;)^2 - N_\phi \cdot \bar{P}^2 \tag{7.15}$$

The total flux integrated over phases ϕ is conserved whatever the δDM. The lost flux in ϕ_0 is therefore equal to all the gained flux of the other phases. Using the fact that $\forall\, n > 1\;,\; x_i > 0\;;\; (\sum_i^n x_i)^2 \;>\; \sum_i^n x_i^2$, we can conclude that the χ^2 is globally decreasing, and can be then modeled with a similar global equation:

$$\chi^2(\delta DM) \;\sim\; \chi^2(0) \cdot (\; 1 - \Phi(\delta DM)\;)^2 - N_\phi \cdot \bar{P}^2 \tag{7.16}$$

Where $\Phi(\delta DM)$ is the global function representing the global loss in χ^2 due to the error in DM. In the frequency-phase diagram, the DM deviation results in quadratic variations of the pulse. The variation of flux in the profile integrated in frequency is consequently probably not linear. However, in the context of a fine search in DM, the DM deviations are small. The variation of the flux in each phase can thereby be approximated by a linear variation relative to the DM deviation. The function $\varphi(\delta DM, \phi)$ can be then evaluated as a linear function of δDM, and by extension the global function $\Phi(\delta DM)$ too.

$$\lim_{\delta DM \to 0} \varphi(\delta DM, \phi) \;\propto\; \delta DM \;\Rightarrow\; \Phi(\delta DM) \;\approx\; a_\Phi \cdot \delta DM \tag{7.17}$$

Table 7.5: Parameters used for the model of the evolution of the χ^2 as a function of DM, with the explored parameter space and the prior distribution for each parameter. dDM represents the DM step used for the fine search, ΔDM the DM window, and DM_b is the DM value found by the search pipeline.

Parameter	**Parameter space**	**Prior**
A_5	$[\ 3\ ;\ +\infty\ [$	$\mathcal{N}(\ 5\ ,\ 150\ )$
σ_5	$\left[\ 2\ dDM\ ;\ \frac{\Delta DM}{4}\ \right]$	$\mathcal{N}\left(\ \frac{\Delta DM}{8}\ ,\ \frac{\Delta DM}{4}\right)$
$DM_{0,5}$	$\left[\ DM_b - \frac{\Delta DM}{4}\ ;\ DM_b + \frac{\Delta DM}{4}\ \right]$	$\mathcal{N}\left(\ DM_b\ ,\ \frac{\Delta DM}{8}\right)$
B_5	$[\ 0\ ,\ 2\ ]$	$\mathcal{N}(\ 1\ ,\ 0.5\ )$

The model $f_5(DM)$ used for the evolution of the χ^2 relative to the DM is therefore a model with four parameters, and which is symmetric relative to the best DM.

$$f_5(DM)\ =\ A_5 \cdot \left(\ 1 - \frac{|DM - DM_{0,5}|}{\sigma_5}\ \right)^2 + B_5\ \ |\ \ f_5\ \geq\ B_5 \tag{7.18}$$

Where $DM_{0,5}$ is the DM allowing to maximize the χ^2 of the profile, A_5 represents the amplitude of the maximum χ^2 at the $DM_{0,5}$, σ_5 describes the width in δDM of the curve of the χ^2, and B_5 is the baseline value. The function is consequently limited at the minimum value B_5 for the edges of the DM window.

Parameter space The parameter A_5 represents the amplitude of the χ^2 of the profile compared to the minimum χ^2 normally corresponding to a noise signal. Because A_5 must correspond to the χ^2 of the most correctly dedispersed profile, it is required to have a significant evolution as a function of the DM. As a consequence, a correctly dedispersed pulsar signal must have an amplitude parameter greater or equal to 3 (as for the previous diagrams).

The parameter σ_5 represents the width of the "DM profile", and is then related to the speed of increase of the χ^2 when the DM error decreases. At NenuFAR frequencies, pulsars with a low-duty cycle are extremely sensitive to the DM and can be completely spread for a little error in DM smaller than a fraction of 1 pc.cm^{-3}. To be able to see the evolution, the minimum half-width of the increase of the χ^2 needs to be of at least 2 DM steps. In the opposite case of pulsars featuring a high-duty cycle, a little error in DM only causes a slight modification of the profile shape, resulting in a weak reduction of the χ^2. The DM window has been chosen to be large with a size of 2 pc.cm^{-3}. Also, to be considered as a probable pulsar signal, the half-width of the "DM profile" must not exceed 1/4 of the DM window.

The DM window is centered around the best DM value found by PREPFOLD. As a result, the parameter allowing the most correct dedispersion, $DM_{0,5}$, might be close to the center of the DM window. As a consequence, a value for $DM_{0,5}$ further than 1/4 of the center of the DM window is considered to be impossible for a real pulsar signal.

The last parameter B_5 corresponds to the χ^2 of the profile for a completely bad dedispersion. The evolution relative to the DM being normalized to 1, the parameter B_5 must be relatively close to 1, and is limited between 0 and 2.

Priors Because of the wide range of shapes and duty cycles that a pulsar can show, the sensitivity relative to the DM error can significantly vary. The range of the amplitude is therefore defined with a normal distribution of mean 5 (as for the other diagrams). However, after tests on the most intense pulsar comprised in the observation tests: B0809+74, it appeared that the value of the half-width required to be substantially increased to a value of 150.

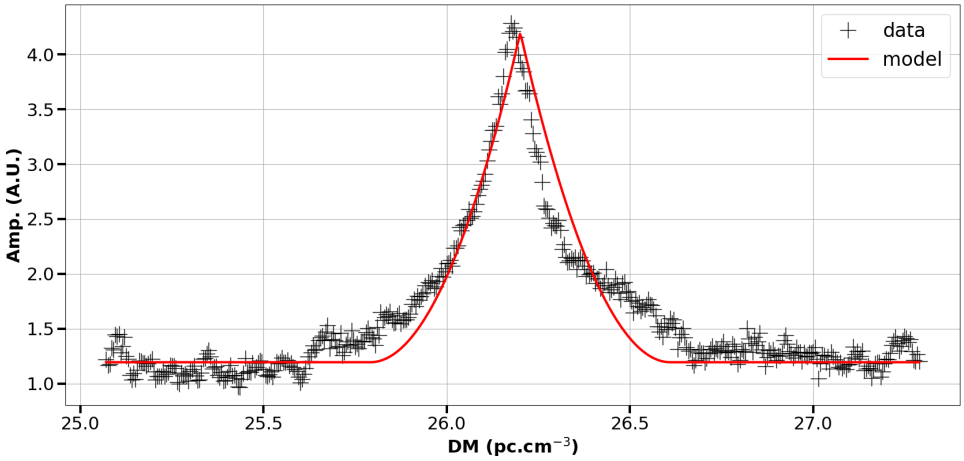


Figure 7.6: Result of the Monte-Carlo simulation for the evolution of the χ^2 relative to the DM for J0323+3944. The initial data are in black crosses and the best model is shown by the red line.

For the same reason, the width parameter σ_5 needs to reflect a large variety of DM sensitivities, and the prior distribution should not be too constraining. Consequently, the characteristic values of the distribution have been defined following a very wide normal distribution around a mean value of 1/8 of the DM window, with a large half-width of 1/4 of the DM window.

The best DM $DM_{0,5}$ is also defined with a normal distribution fixed around the DM value found by PREPFOLD. However, compared to the two previous parameters, $DM_{0,5}$ is more constrained. In order to stay close to this initial value, the distribution needs to be narrower with a half-width of 1/8 of the DM window.

Finally, the baseline parameter B_5 must be defined close to 1 in the large majority of the cases. It is then defined using a normal distribution of mean 1 and with a half-width of 0.5.

Table 7.5 summarizes the parameter spaces and prior functions for each parameter of the simulation. A result of the Monte-Carlo simulation of the evolution of the χ^2 as a function of the DM for the pulsar J0323+3944 is presented in Figure 7.6 where the best model found is superposed with the real data.

7.2.6 Evolution of the χ^2 relative to the period

Model function The next control diagram is similar to the previous one but is related to the evolution of the χ^2 of the profile relative to an error in the period rather than in DM. The model used is based on the same reasoning as that for the evolution relative to the DM. Moreover, in the case of the period, the shift due to the period error is really linear. The model $f_6(P)$ is a function of the period P in the same form:

$$f_6(P) \;=\; A_6 \cdot \left(1 - \frac{|P - P_{0,6}|}{\sigma_6} \right)^2 + B_6 \;\;|\;\; f_6 \;\geq\; B_6 \tag{7.19}$$

Here A_6 is the amplitude, σ_6 is the parameter of the width of the curve of the χ^2, $P_{0,6}$ is the period allowing to maximize the χ^2, and B_6 is the level of the baseline. As for the evolution relative to the DM, the function is limited to the χ^2 of the baseline.

Parameter space The equation for the model $f_6(P)$ is similar to the equation for the model $f_5(DM)$, involving the same parameters. The arguments given for the parameters of the evolution in DM are still valid for the evolution in the period. As a consequence, the parameter spaces used for the model $f_5(DM)$ are equally appropriate for the model $f_6(P)$.

Table 7.6: Parameters used for the model of the evolution of the χ^2 relative to the period P, with the explored parameter space and the prior distribution for each parameter. dP represents the period step, ΔP is the period window, and P_b is the period value found by PREPFOLD.

Parameter	Parameter space	Prior
A_6	$[\ 3\ ;\ +\infty\ [$	$\mathcal{N}(\ 5\ ,\ 300\ )$
σ_6	$\left[\ 2\ dP\ ;\ \frac{\Delta P}{4}\ \right]$	$\mathcal{N}\left(\ \frac{\Delta P}{8}\ ,\ \frac{\Delta P}{4}\right)$
$P_{0,6}$	$\left[\ P_b - \frac{\Delta P}{4}\ ;\ P_b + \frac{\Delta P}{4}\ \right]$	$\mathcal{N}\left(\ P_b\ ,\ \frac{\Delta P}{8}\right)$
B_6	$[\ 0\ ,\ 2\ ]$	$\mathcal{N}(\ 1\ ,\ 0.5\ )$

Priors Concerning the definition of the values to try for each parameter, the global behavior in the period being similar to that in DM, the prior distributions to use are also similar.

The diagram is normalized by PRESTO with the χ^2 value of the baseline, in order to have a baseline value of 1. Also, tests on the most intense pulsar, B0809+74, have revealed it is required to set a higher value of the amplitude. The sensitivity relative to the period error should be greater than for the DM error, causing larger variations of the profile. As a consequence, the half-width of the normal distribution used to define A_6 is consequently increased to 300, with a mean of 5.

Table 7.6 summarizes the parameter space and the prior function for each parameter used for the simulation. Figure 7.7 presents the result of the best model found by the simulation of the evolution of the χ^2 relative to the period for a candidate of a test observation of the pulsar J0323+3944.

Table 7.6 summarizes the parameter spaces and prior functions for each parameter of the simulation. A result of the Monte-Carlo simulation of the evolution of the χ^2 as a function of the period for the pulsar J0323+3944 is presented in Figure 7.7 where the best model found is superposed with the real data.

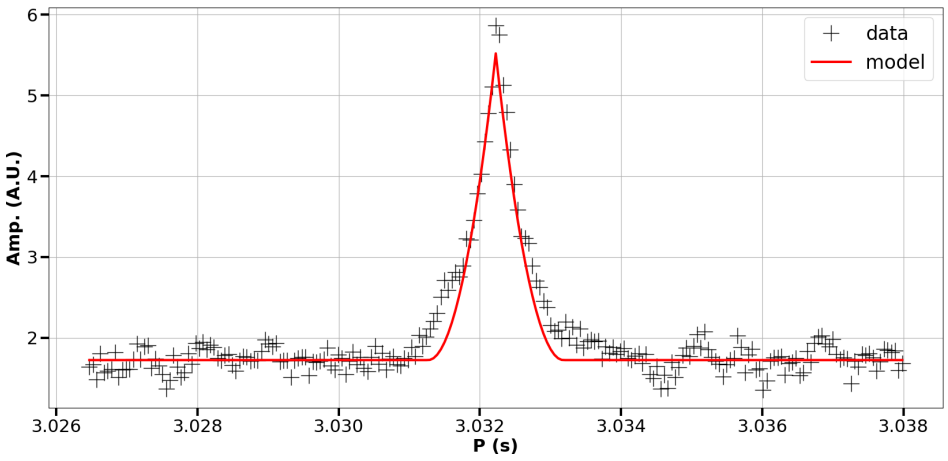


Figure 7.7: Result of the Monte-Carlo simulation for the evolution of the χ^2 relative to the period for J0323+3944. The initial data are in black crosses and the best model is shown by the red line.

7.2.7 Evolution of the χ^2 relative to the period derivative

Model function A pulsar is a precise clock, with excellent accuracy in the period of the pulsation. However, because of the spin-down of the neutron star, there is a very light increase in the period over time. Therefore, in order to perfectly fold the data, one needs to take this period shift into account. This seventh diagram allows for obtaining the period derivative of the candidate. For a periodic signal which is not a pulsar, e.g. an electric signal, it is expected to show a perfect constancy over time, resulting in a period derivative strictly equal to 0. On the other side, a candidate which is a pulsar must have a clear peak in the χ^2 curve for a non-zero period derivative.

With this control plot, we can test the behavior of the profile of the candidate by adding a shift in the periodicity. Also, for the calculus of the χ^2 of a pulsar profile, this shift is globally equivalent to an error in the period. As a result, a model $f_7(\dot{P})$, identical to the period model, is employed.

$$f_7(\dot{P}) \;=\; A_7 \cdot \left(1 - \frac{|\dot{P} - \dot{P}_{0,7}|}{\sigma_7} \right)^2 + B_7 \;\;|\;\; f_7 \;\geq\; B_7 \tag{7.20}$$

With A_7 the parameter of amplitude, σ_7 the the width of the χ^2 curve, $\dot{P}_{0,7}$ the best period derivative to use for the folding, and B_7 the level of the baseline. As for the evolution relative to the DM, the function is limited to the χ^2 of the baseline.

Parameter space The model and the global behavior are the same between the evolution in the period and period derivative. The constraints on the function $f_7(\dot{P})$ are thereby identical to those on $f_6(P)$. As a consequence, the parameter spaces are the same.

Priors Finally, the expected behavior of the profile for a period derivative error is basically the same as for a period error. As a consequence, concerning the prior distributions to use, there are exactly the same as those for the evolution of the χ^2 in the period.

Table 7.7 summarizes the parameter spaces and prior functions for each parameter of the simulation. A result of the Monte-Carlo simulation of the evolution of the χ^2 as a function of the period derivative for the pulsar J0323+3944 is presented in Figure 7.8 where the best model found is superposed with the real data.

Parameter	Parameter space	Prior
A_7	$[\ 3\ ;\ +\infty\ [$	$\mathcal{N}(\ 5\ ,\ 300\ )$
σ_7	$\left[\ 2\ d\dot{P}\ ;\ \frac{\Delta\dot{P}}{4}\ \right]$	$\mathcal{N}\left(\ \frac{\Delta\dot{P}}{8}\ ,\ \frac{\Delta\dot{P}}{4}\ \right)$
$\dot{P}_{0,7}$	$\left[\ \dot{P}_b - \frac{\Delta\dot{P}}{4}\ ;\ \dot{P}_b + \frac{\Delta\dot{P}}{4}\ \right]$	$\mathcal{N}\left(\ \dot{P}_b\ ,\ \frac{\Delta\dot{P}}{8}\ \right)$
B_7	$[\ 0\ ,\ 2\ ]$	$\mathcal{N}(\ 1\ ,\ 0.5\ )$

Table 7.7: Parameters used for the model of the evolution of the χ^2 relative to the period derivative $\dot{P}$, with the explored parameter space and the prior distribution for each parameter. $d\dot{P}$ represents the period derivative step, $\Delta\dot{P}$ is the period derivative window, and $\dot{P}_b$ is the period derivative value found by PREPFOLD.

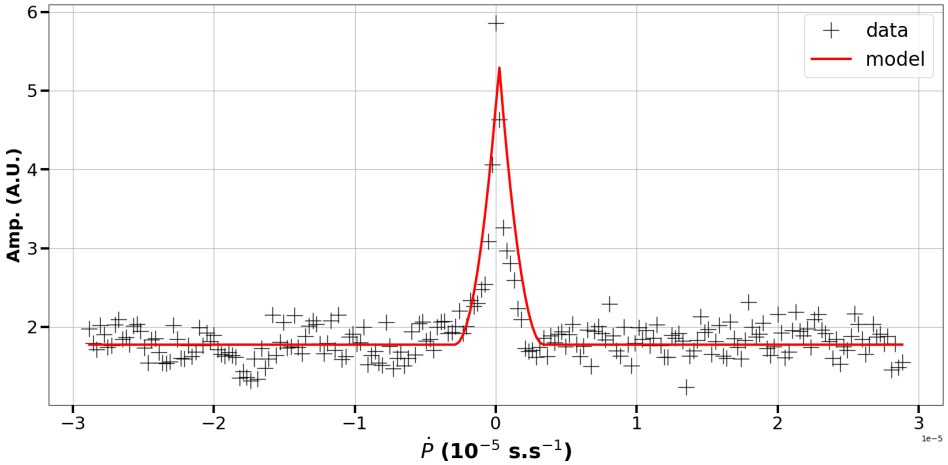


Figure 7.8: Result of the Monte-Carlo simulation for the evolution of the χ^2 relative to the period derivative for J0323+3944. The initial data are in black crosses and the best model is shown by the red line.

7.2.8 $P - \dot{P}$ plane

Model function The three previous control diagrams were one-dimensional, checking the behavior of the profile relative to the variation of only one variable. Contrary to the DM error which yields a shift of the profile in the frequency domain, an error in the period or period derivative both cause a shift in the time domain. These two variables thereby produce a similar effect on the profile, and the two effects can be combined or at the opposite can be compensated. To take the covariance between these two parameters, it needs indeed to perform two-dimensional computing of the χ^2, resulting in a $P - \dot{P}$ plane.

The evolution of the χ^2 relative to the period is computed around an initial value P_b and for a fixed value of the period derivative: the initial value $\dot{P}_b$ found by PREPFOLD. For the evolution of the χ^2 relative to the period derivative, it is the exact inverse with a search around the initial value $\dot{P}_b$ for a fixed value of the period P_b (those found by PREPFOLD). These two values are searched by PREPFOLD, by generating a $P - \dot{P}$ plane and extracting the maximum point of coordinates $(P_b, \dot{P}_b)$. The two previous one-dimensional plots are indeed created after the fact based on this plane.

The $P - \dot{P}$ plane allows seeing the behavior of the profile for any combination of the period and period derivative. The global structure within this more complete evolution diagram permits controlling whether the candidate signal is consistent with those of a pulsar. The majority of the RFIs have a null period derivative and must possess a structure only in the period. In the case of RFIs that are not periodic, no particular structure is expected to appear.

In the case of a pulsar signal, the pattern has to reveal the covariance of both parameters. For any couple $(P, \dot{P})$ where there is a cumulative effect, there is the anti-symmetric couple $(-P, \dot{P})$ or $(P, -\dot{P})$ yielding a compensation of the effects. As a result, for a pulsar, the $P - \dot{P}$ plane must show an asymmetric structure where the high χ^2 are located on a diagonal. After, on this diagonal, the χ^2 values are modulated by the deviation compared to the best couple of values.

The evolution of the χ^2 can be expressed in the same way that for the unidimensional versions. The difference is that the amplitude and the reference position for the maximum χ^2 are dependent on the second variable. Then, for a given $\dot{P}$, the evolution of the χ^2 relative the period P follows the equation:

$$\chi^2(P)|_{\dot{P}} = A(\dot{P}) \cdot \left(1 - \frac{|P - P_m(\dot{P})|}{\sigma_P} \right)^2 + B \tag{7.21}$$

Here $A(\dot{P})$ is the amplitude-dependent of the period derivative, σ_P the width of the evolution in the period, and B the noise parameter. $P_m(\dot{P})$ correspond to the period where the χ^2 is maximized for

the given $\dot{P}$. This maximum value of χ^2 must follow a diagonal corresponding to the couple of values $(P_m, \dot{P}_m)$ where the positive corrected effect is combined. In a $P - \dot{P}$ plane centered on the best values $(P_0, \dot{P}_0)$, P_m follows, therefore, a linear relation:

$$P_m \;=\; \kappa \cdot (\; \dot{P} - \dot{P}_0 \;) + P_0 \tag{7.22}$$

The equation of the χ^2 becomes a relation where the covariance between the period error δP and the period derivative error $\delta \dot{P}$ clearly appears.

$$\chi^2(P)|_{\dot{P}} \;=\; A(\dot{P}) \cdot \left(1 - \frac{|\delta P - \kappa \cdot \delta \dot{P}|}{\sigma_P} \right)^2 + B \tag{7.23}$$

The maximum χ^2 for a given $\dot{P}$ is obtained when $P \;=\; P_m$, and is equal to $A(\dot{P}) + B$. We retrieve a similar problem that for the evolution of the χ^2 relative to the period derivative in the model $f_7(\dot{P})$. The modulation function of the amplitude can be therefore expressed as:

$$A(\dot{P}) \;=\; A \cdot \left(1 - \frac{|\; \dot{P} - \dot{P}_0 \;|}{\sigma_{\dot{P}}} \right)^2 \tag{7.24}$$

Where A is the maximum amplitude reached for $P \;=\; P_0$ and $\dot{P} \;=\; \dot{P}_0$. The final model $f_8(P, \dot{P})$ used to simulate the $P - \dot{P}$ plane is finally given by the relation:

$$f_8(P, \dot{P}) \;=\; A_8 \cdot \left(1 - \frac{|\dot{P} - \dot{P}_0|}{a_8} \right)^2 \cdot \left(1 - \frac{\left| \; (P - P_0) + \dfrac{(\dot{P} - \dot{P}_0)}{\alpha_8} \; \right|}{b_8} \right)^2 + B_8 \tag{7.25}$$

Here A_8 represents the global amplitude, B_8 is the noise parameter, and a_8 and b_8 are the two parameters of the speed of the evolution of the χ^2. On the diverse tests carried out on known pulsars, it appeared that the sensitivity to the period error is greater than the sensitivity to the period derivative error. It means the numerical value is less than one: $|\; d(\delta P) \;/\; d(\delta \dot{P}) \;| \;=\; |\kappa| \;<\; 1$. Also, the constructive combination occurs obviously for errors opposed in sign, meaning that $\kappa \;<\; 0$. Thus, in place of κ, for the rest of the text, the parameter $\alpha_8 = -1/\kappa$ is preferred to define the model.

Table 7.8: Parameters used for the model of the evolution of the χ^2 in the $P - \dot{P}$ plane, with the explored parameter space and the prior distribution for each parameter. dP represents the period step, $d\dot{P}$ represents the period derivative step, ΔP the period window, $\Delta \dot{P}$ the period derivative window, and P_b and $\dot{P}_b$ are the period and period derivative values found by PREPFOLD.

Parameter	**Parameter space**	**Prior**
A_8	$[\; 3 \;;\; +\infty \;[$	$\mathcal{N}(\; 5 \;,\; 150 \;)$
a_8	$\left[\; 2 \; d\dot{P} \;;\; \dfrac{1.5 \Delta \dot{P}}{4} \; \right]$	$\mathcal{N}\left(\dfrac{\Delta \dot{P}}{8} \;,\; \dfrac{\Delta \dot{P}}{4} \right)$
b_8	$\left[\; 2 \; dP \;;\; \dfrac{1.5 \Delta P}{4} \; \right]$	$\mathcal{N}\left(\dfrac{\Delta P}{8} \;,\; \dfrac{\Delta P}{4} \right)$
α_8	$[\; 1.5 \;;\; +\infty \;[$	$\mathcal{N}(\; 3 \;,\; 3 \;)$
B_8	$[\; 0 \;,\; 2 \;]$	$\mathcal{N}(\; 1 \;,\; 0.5 \;)$

Reduction of the parameter space The parameters P_0 and $\dot{P}_0$ are normally two parameters to define, as made for all the other models. But adding these two parameters increase the parameter space to seven dimensions, which starts to become difficult to efficiently cover. Because of the high resolution of the fine searches in the period and period derivative, the $P - \dot{P}$ plane is indeed already the longest diagram to analyze.

However, because of the large number of candidates to analyze, the aim is to fastly realize the analysis. Also, the number of models to generate in order to correctly cover a parameter space of seven dimensions was excessively long. Moreover, the structure in the $P - \dot{P}$ plane is generally not very strong and rather diffuse. As a result, in the majority of the cases, finding clearly the parameters P_0 and $\dot{P}_0$ is difficult. A very long computing time is therefore required to obtain fine values of all the parameters.

As a consequence, the starting constraints for the choice of the method to do are no longer satisfied. Furthermore, the analysis of the $P - \dot{P}$ plane is more difficult than the other diagrams. The result is generally not clear enough compared to the other diagrams where a signal of a pulsar significantly appears. Given of the important probability to finally obtaining an inconclusive result compared to the other diagrams, it seems that it is inefficient to perform a precise and very long computation.

Finally, because of the enormous computing time inquired to obtain a sufficiently good and precise result, especially concerning the values of P_0 and $\dot{P}_0$, I have chosen to directly fix these two parameters at the maximum value of the data. Once these two values are defined, the parameter space is of five dimensions (similar to other diagrams). The simulations are then easier and faster, returning nevertheless a reasonable result.

Parameter space The model $f_8(P, \dot{P})$ used for the $P - \dot{P}$ plane is similar in its construction and its form to the one-dimensional models of the evolution of the χ^2 for $f_6(P)$ and $f_7(\dot{P})$. As a consequence, the parameter spaces for the equivalent parameters can be defined in the same way. The spaces for A_8 and B_8 are then the same that the spaces used for A_6 (or A_7) and B_6 (or B_7).

The two parameters related to the speed of the evolution a_8 and b_8 are defined following the space parameter of the width of the corresponding model. By construction, a_8 is related to the period derivative and is then defined as σ_7, and b_8 is related to the period and its definition is made as σ_6. However, for the previous model $f_6(P)$ and $f_7(\dot{P})$, one of the two parameters was fixed to the best value, i.e. without any error, and then the peak was relatively narrow. In the case of the $P - \dot{P}$ plane, the entire error is a function of the errors of the two parameters. It results from the associated covariance between the two errors a widening of the peak, essentially on the diagonal of the linear relation of slope κ. The space parameters of a_8 and b_8 must be therefore increased compared to the width with just one source of error. The upper limit of the spaces is set to 1.5 times the upper limit of the spaces for the model $f_6(P)$ and $f_7(\dot{P})$.

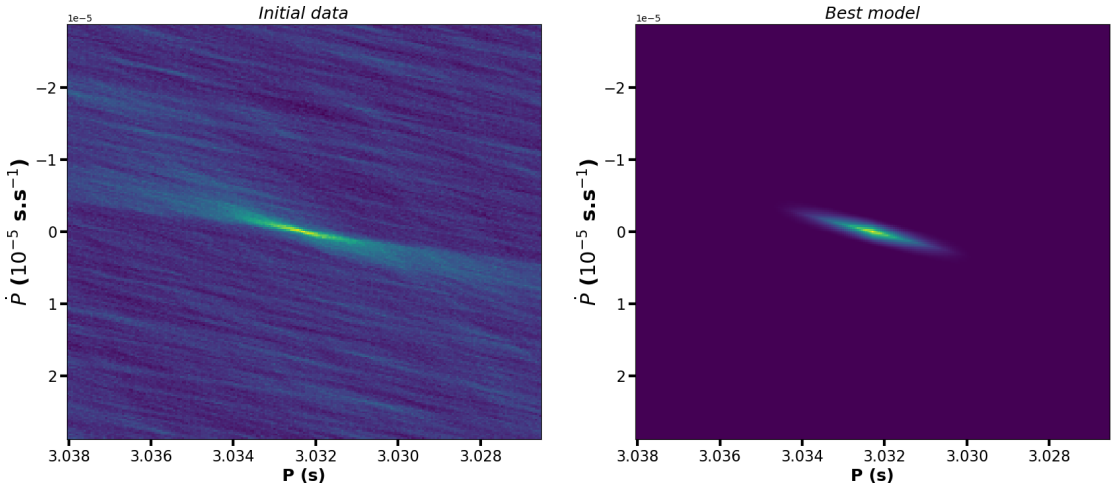


Figure 7.9: Result of the Monte-Carlo simulation for the $P - \dot{P}$ plane of J0323+3944. Left: initial data. Right: best model found.

The last parameter α_8 represents the ratio between the impact of the error in the period compared to the impact of the error in the period derivative on the χ^2 of the profile. The period derivative of a pulsar is lower than the period, with values around 10^{-15} $s.s^{-1}$. Then, the sensitivity of the χ^2 of the profile relative to the error in the period derivative is inevitably greater than for the error in the period. Consequently, the factor α_8 must be greater than 1 for all pulsars. This factor depends on the values of P and $\dot{P}$ of the pulsar, and on the shape of the profile. It is consequently very difficult to estimate an upper limit. The parameter space for α_8 has finally been defined between 1.5 and infinity.

Priors As for the parameter spaces, the prior distributions of A_8 and B_8 are identical to the prior distributions of the corresponding parameters for the two previous models $f_6(P)$ and $f_7(\dot{P})$.

The similarity of the width parameters a_8 and b_8 with the width of the previous models is obviously also used to define their prior distributions. a_8 is defined as the prior distribution of the width of the "period derivative profile" σ_7, and b_8 as the prior distribution of the width of the "period profile" σ_6.

Finally, for the only new parameter α_8, a pulsar can take a large range of values in the domain explained just above. Consequently, a normal distribution has been chosen, and the parameters were set relative to the results obtained during the test phase on different faint pulsars. According to the various carried-out tests, the standard behavior of the pulsars seems to follow a low-sloped diagonal. As a result, the normal prior distribution has been set to a mean of 3 with a standard deviation of 3, allowing for searching for pulsar signals around a such low-sloped diagonal.

Table 7.8 summarizes the parameter spaces and prior functions for each parameter of the simulation. A result of the Monte-Carlo simulation of the $P - \dot{P}$ plan for the pulsar J0323+3944 is presented in Figure 7.8 with the real plane in the left panel and the best model found in the right panel.

7.3 Adjustment of the significances

The different models used for the Monte-Carlo simulations have been parametrized in order to obtain a model significance that seems to be sufficiently correct compared to the tested signals. However, the obtained value is a function of a parameter external to the model. This parameter is the number of models to generate. It allows for modification of the coverage of the parameter space. Larger the number of models, the easier and more precise the determination of a significance close to the real best possible significance. However, a longer computing time will be required.

Moreover, the analysis of a candidate is actually done by the linear combination of the significances of the different models, allowing for obtaining a global significance (see equation 7.5). Nevertheless, just realize a simple average is not the best way to maximize the correctness of this combination. As a consequence, it is required to adjust as much as possible the weights of the models.

7.3.1 Number of models

A Monte-Carlo simulation tries to determine the best solution, which is in our case, the best adequacy with a signal of a pulsar. For this, it is required to sufficiently properly cover the parameter space, especially for this type of Monte-Carlo simulation without a directed Markov chain. The larger the number of models the better the significance. However, the function of the resulting significance is an asymptotic function converging more or less quickly towards the (at least the ablest!) correct significance, depending on the model and the candidate signal characteristics. Moreover, because of the considerable number of expected candidates of about 500 000, this analysis must be fast. As a consequence, these strong constraints require avoiding oversampling the parameter space of the simulations.

For each model, the number of models to generate has been evaluated by calculating the significance for 50 different numbers of models from 1 000 to 10 million. Beyond 10 million, in any case, the computing time becomes excessively long for the whole survey. This work was made on two pulsars sufficiently strong to be able to obtain an evident significance. A reasonable number of models has been finally estimated by looking at the convergence of the significance curve relative to the number of models.

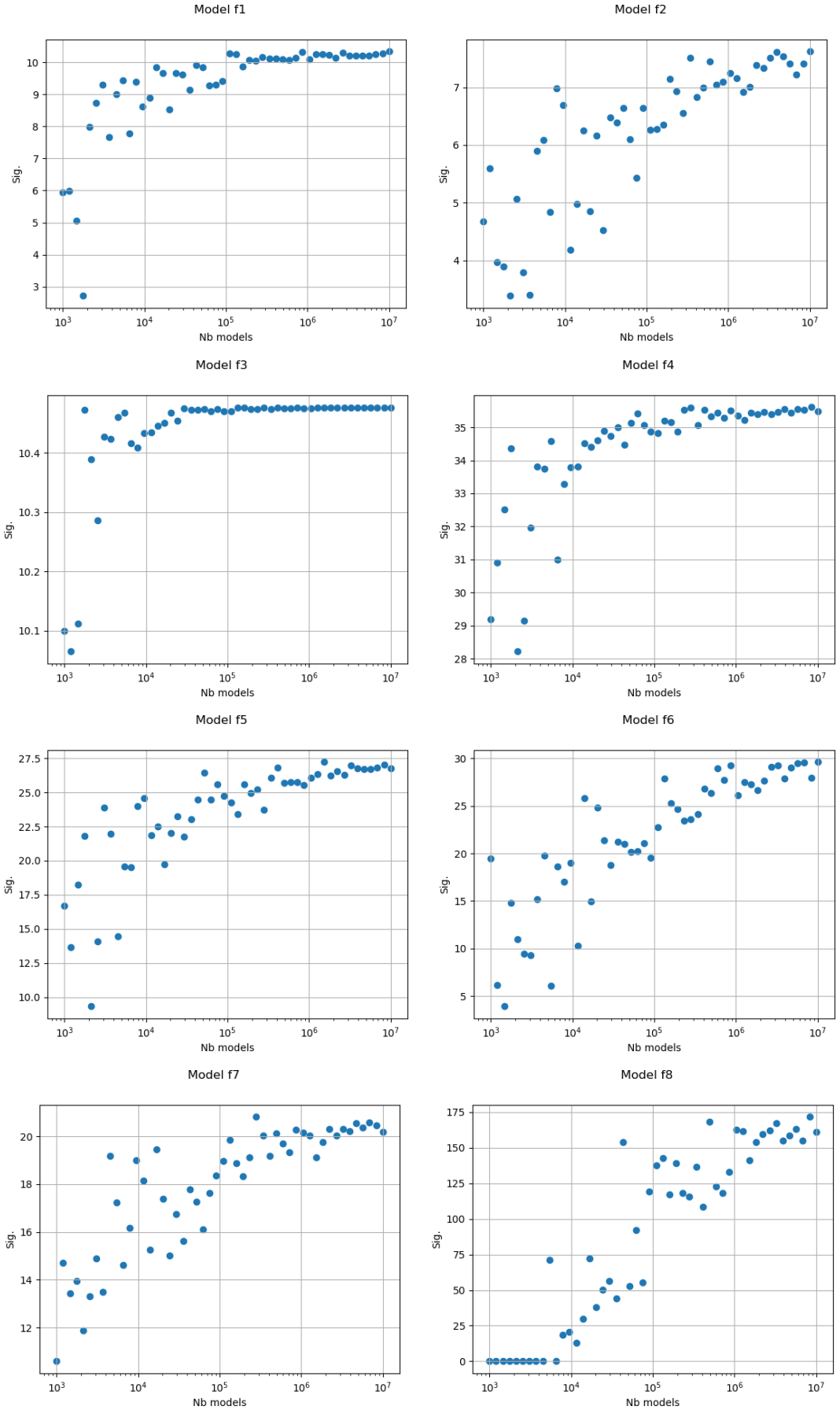


Figure 7.10: Significances relative to the number of models used for the Monte-Carlo simulation for each of the 8 models.

Table 7.9: Number of models to generate for the Monte-Carlo simulations. Each plot refers to one of the eight models.

Model	**Corresponding diagram**	**Nb. of models**
$f_1(\phi)$	Integ. profile	1.10^5
$f_2(\phi, t)$	Time-phase plane	5.10^5
$f_3(t)$	Time evolution	6.10^4
$f_4(\phi, \nu)$	Frequency-phase plane	1.10^5
$f_5(DM)$	DM evolution	5.10^5
$f_6(P)$	Period evolution	5.10^5
$f_7(\dot{P})$	Period derivative evolution	5.10^5
$f_8(P, \dot{P})$	$P - \dot{P}$ plane	1.10^5

Figure 7.10 shows the resulting curves of the obtained significances relative to the number of models used for the simulation. The aim is not really to obtain the best significance, but only a significance that is sufficient to be able to discriminate a potential pulsar from an RFI. By inspection of each diagram in Figure 7.10, a reasonable number of models to generate has been estimated taking the number where the function begins to converge. The convergence is considered as reached when the variations and the increasing trend become small. Choosing a number at the beginning of the convergence allows for obtaining a value of significance relatively close to the best one with a reasonable computing time.

The number of models finally chosen for each diagram is listed in Table 7.9. For models f_1, f_3, and f_4, convergence is rapidly obtained. As a result, the defined number of models should allow us to obtain a proper significance very close to the best one.

For models f_2, f_5, f_6, and f_7, convergence is not fully reached but seems to be approached after several millions of models. However, for these four diagrams, the computing time substantially increases above one million of models. Above 100 000 models, the significance only increases slowly, and the obtained significance can be then considered as sufficiently close. Although there should have at least one million of models to have a significance very close to the value at the convergence, the number of models has been set to a more reasonable value of 500 000 models. Such a value allows us to decrease the computing time of each at least by about two.

Finally, for the last model f_8, corresponding to the $P-\dot{P}$ plane, the number of models to use is harder to define. As explained previously, the signal in the $P - \dot{P}$ plane is very diffuse, and calculating a truth significance is more complicated than for the other models. Furthermore, the size of the plane is often larger than the other diagrams, leading to extending the computing time significantly. The constraints are consequently more important for this model, and the number must be reduced as far as possible. In the corresponding panel of Figure 7.10, there is a break in the function at 100 000. According to the expected accuracy of this model, the number of models of 100 000 has been considered as sufficient to obtain relatively proper discrimination between a pulsar and an RFI, while having a reasonable computing time.

7.3.2 Training set

Each model allows testing whether the data are compatible with a pulsar signal. But, adequacy can be correctly judged solely if there is adequacy in several of the control diagrams. As a consequence, we need to obtain a global significance by combining the result of each model to have a weighted average significance.

However, some diagrams are probably more accurate than others. For example, the $P - \dot{P}$ plane is more difficult to test than the profile. Some features of a pulsar signal can be more relevant than others, such as the frequency vs phase plane which seems to be a better way to check the signal than the DM evolution of the χ^2. The global significance must be constructed taking these differences of relevance between the significances of the models into account.

To determine the weights to apply, a training phase was carried out based on a set of observations, including some observations with a known pulsar and a large number of observations without known pulsars. The set was composed of 24 test observations (corresponding to 96 test pointings), realized during a specific observing phase test on the 26^{th} of December 2020. Of these 24 observations, 14 were chosen in order to have a pulsar at least in one of the four numerical pointings.

The 14 pulsars listed in Table 7.10 have been chosen based on the detections of the NenuFAR pulsar census. The set of pulsars was defined in order to try to have a large sample of types of pulsars, containing slow pulsars (with a period greater than one second), faster pulsars (but not MSP), pulsars with low DM (typically with DM $\lesssim$ 15 pc.cm^{-3}), and with high DM (typically with DM $\gtrsim$ 50 pc.cm^{-3}). The set was also designed aiming to test pulsars with various levels of expected SNR from J0242+62 which is normally almost undetectable to B1508+55 (J1509+5531 in Table 7.10) which is extremely intense.

The processing of the 96 pointings delivered a total of 6 170 candidates. The candidates have been then manually inspected one by one to classify each candidate into one of the three following classes: RFI, known pulsar, or potential pulsar. The first class obviously represents all the candidates featuring a signal without any characteristics of a pulsar signal, and then clearly identified as RFIs. The second class contains candidates whose signal corresponds to one of the fourteen known pulsars in Table 7.10. Finally, the last class concerns the signals which can appear as a pulsar signal, but which do not correspond to a known pulsar.

Table 7.11 shows the number of candidates classified in each of the three classes and the fraction of the class compared to the whole set. The vast majority of the candidates are clearly non-pulsar signals and are put in the RFI class. On the opposite, the number of candidates corresponding to a known pulsar is merely 1.17% of all the candidates. These fractions are due to the fact that the different thresholds used in the search processing pipeline are voluntarily set to low values in order to be sure to be able to detect even a faint pulsar.

Table 7.10: List of the known pulsars present in the test observations made in December 2020. The second, third, and fourth columns represent the period, period derivative, and DM of the pulsar in the ATNF catalog (Manchester et al. 2005). The star * means that the period derivative of J0242+62 is unknown. The last column indicates the SNR of the detection in the NenuFAR pulsar census (Bondonneau et al. in preparation).

PSR	**P** (s)	$\dot{\mathbf{P}}$ (s.s^{-1})	**DM** (pc.cm^{-3})	**SNR** census
B0052+51	2.115171148865 (4)	9.53764E-15 (4)	44.0127 (24)	6.0
J0242+62	0.592 (0)	*	3.92 (16)	99.4
B0320+39	3.032071956385 (10)	6.3562E-16 (7)	26.18975 (93)	66.1
B1322+83	0.670037418386 (3)	5.6630E-16 (3)	13.31624 (76)	17.6
B1508+55	0.739681922904 (4)	4.99821E-15 (4)	19.6191 (3)	1064.5
J1628+4406	0.18117848994715 (7)	1.9378E-17 (3)	7.32981 (2)	28.8
J1647+6608	1.59979837536 (3)	7.839E-15 (16)	22.55 (7)	24.4
J1821+4147	1.261857209134 (8)	1.7292E-15 (10)	40.673 (3)	6.0
B1839+56	1.6528618528869 (20)	1.494818E-15 (14)	26.77163 (17)	361.8
B1905+39	1.235757452781 (3)	5.40707E-16 (20)	30.966 (14)	7.2
B2022+50	0.372619054536 (3)	2.512052E-15 (17)	32.98817 (37)	13.5
B2154+40	1.525265633965 (12)	3.43256E-15 (7)	71.1239 (22)	22.7
J2208+4056	0.636957393614 (13)	5.2832E-15 (9)	11.837 (9)	114.7
B2334+61	0.495369868028 (5)	1.934498E-13 (8)	58.410 (15)	24.0

Table 7.11: Types of signals used to classify the candidates of the training set. RFI is for the signals which are clearly non-pulsar signals. POT is for the signals which potentially similar to a pulsar signal but which don't correspond to a known pulsar. PSR is for the signals corresponding to a known pulsar.

Class	**Number**	**Fraction**
RFI	5 689	92.2%
POT	409	6.63%
PSR	72	1.17%
Total	**6 170**	**100.00%**

The class POT in Table 7.11 refers to the ambiguous signals featuring a few, even sometimes more than a few, characteristics of a pulsar signal. 6.63% of the entire set are signals which can't be clearly identified as an RFI. In order to be able to detect a faint signal of a pulsar in the candidates of the survey, we have to make sure these ambiguous candidates are not ruled out, so that they can be examined in more detail thereafter.

With regard to this distribution, it is really valuable to have a good significance calculation to be able to reject swiftly the numerous candidates which clearly differ from a pulsar. Normally, with significances correctly adjusted, only about 5% of these test candidates (the candidates of the PSR class plus a significant part of those of the POT class) might appear with a global significance greater than 5σ, leading to a make a drastic reduction of the number of the candidates to examine.

7.3.3 Training stage

The goal of this training is to find a set of weights allowing us to obtain values of significance that permits us to efficiently discriminate the three classes. I have written a Python script implementing a method to search for a good set of parameters based on the training set of 6 170 candidates.

The method implemented is based on searching for the maximum value of a test function. Starting from an initial set of weights, the test function is determined for several neighbor sets of weights. The process is then iteratively done for the best found set of weights until that none neighbor set increases the value of the test function.

For each of the three classes, a test function is defined as the proportion of candidates with a significance Σ considered as good. The global test function F_{test} is then the mean of the three test functions:

$$F_{test} = \frac{1}{3}\left(\frac{N_{PSR}(\ \Sigma > \Sigma_{PSR}\ )}{N_{PSR}} + \frac{N_{POT}(\ \Sigma > \Sigma_{POT}\ )}{N_{POT}} + \frac{N_{RFI}(\ \Sigma < \Sigma_{RFI}\ )}{N_{RFI}}\right) \tag{7.26}$$

Here Σ_{PSR}, Σ_{POT}, and Σ_{RFI} are the thresholds used for each of the three classes. As for the model significance, the global significance Σ is also related to a statistical test with a standard variable, i.e. to a Gaussian probability.

For the candidates classified in the class PSR, we know they are pulsars. According to the Gaussian probability, it is required that they have a significance greater than 5σ, corresponding in terms of a normal distribution to a probability of 5.73×10^{-7} that the signal be a white noise fluctuation.

For the candidates classified in the class RFI, we are in the exact opposite case, where it is sure that these candidates are not pulsars. Consequently, the condition is equally opposed. Hence, it is required that these candidates obtained a global significance Σ less than 5σ.

For the last candidates classified in the class POT, the situation is more complex, According to the ATNF catalog, we know that they are not part of the known pulsars, but they are more or less similar to a pulsar signal. Thus, in terms of normal probability, these candidates need to correspond to a higher probability than the majority of the RFIs. The threshold for these candidates has consequently been set to 3σ, representing a Gaussian probability of $2.7.10^{-3}$. Some of these candidates are very similar to a pulsar signal, leading to do not fixed upper threshold.

Because of the number of dimensions of eight, the space to explore is vast, and the search method is therefore based on a directed path. At each step, the value of the test function is compared to the values of the test functions of the neighbor points. The next step of the path is given by the point with the maximum value of the test function. The path is stopped when no neighbor points show a larger value of the test function. However, this type of method can't be sure to find the global maximum rather than just a local maximum. In order to increase the number of paths in the space, the training was consequently performed with different starting points and different steps of search. 420 starting configurations were tried, and finally, two have given the same best result for the global test function, with in addition the same set of coefficients to use.

Model	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
ζ	0.30728	0.18576	0.09124	0.09124	0.09124	0.22177	0.00573	0.00573

Table 7.12: List of the 8 optimized coefficients ζ found by the training stage for each model.

Table 7.12 indicates the coefficients ζ obtained by the training stage. The coefficients are used to calculate the global significance and are then normalized in order that the sum of all is equal to 1. We can see the model which seems to give the most valuable pieces of evidence to discriminate a pulsar signal from an RFI is the integrated profile. Having a properly integrated profile is effectively mandatory to be evaluated as a potential pulsar, and the prominence of the model f_1 is consequently not really a surprise.

On the opposite side, the $P - \dot{P}$ plane and the evolution of the χ^2 relative to the period derivative are apparently not really significant to find a pulsar signal. Indeed, because of the high sensitivity of a signal relative to the variation in the period derivative, it is very difficult to realize a sufficiently good fit to discriminate a pulsar signal from an RFI. These diagrams are consequently actually supplementary pieces of evidence rather than key evidence, which is consistent with the result of the training.

The coefficients listed in Table 7.12 allow for obtaining a result for the global test function of 97.837%, representing to 6037 over 6170 candidates correctly classified. In the details of the three classes (summarized in Table 7.13), we can notice all the 72 candidates corresponding to a known pulsar obtain, as wished, a global significance greater than 5. Concerning the opposite case of the RFIs, the best configuration permits, in addition to the best total classification, to have equally the best proportion of good classification of the RFIs with 98.4%, letting only 91 bad candidates with a too-high significance.

The result for the last class of 95.11%, including the potential faint pulsar signals, is more nuanced. The best configuration doesn't return the best result for the potential pulsars. However, this class represents a grey area, where the signals are more difficult to classify compared to the two other types of signals. However, the configurations providing the best classification of the potential signals actually just increase slightly the proportion, with an increase of 3 candidates from 409 to 412.

Finally, applying these coefficients on the raw significances computed with the Monte-Carlo simulations, the initial number of candidates found by the search pipeline of 6 170 is decreased to 406 interesting candidates. These 406 candidates correspond to the 72 candidates of the class PSR, added of the candidates of the class POT with a global significance greater than 5σ. These interesting candidates are not all good candidates, and need to be examined in more details directly by eye to determine if a follow-up observation is required.

Table 7.13: Results of the classification by the training stage with proportions of test candidates and success conditions for each of the three classes. (1) Number of candidates initially classified in this class. (2) Condition to be considered as successful. (3) Percentage of successful candidates. (4) Number of successful candidates. (5) Number of unsuccessful candidates.

Class	(1) **Total**	(2) **Condition**	(3) **Success**	(4) **Good class.**	(5) **Bad class.**
PSR	72	$\Sigma \geq 5$	100%	72	0
POT	409	$\Sigma \geq 3$	95.11%	389	20
RFI	5689	$\Sigma < 5$	98.4%	5598	91
Global	**6170**		**97.837%**	**6037**	**133**

In conclusion, without the optimized coefficients obtained with the training, i.e. with $\zeta_i = 1 \; \forall \; i$, the number of candidates with a significance greater or equal than 5σ is 2 565, representing 41.6% of all the candidates. The use of the optimized coefficients allows decreasing this fraction of candidates to manually look to the 6.6% best one. The number of candidates to definitely exclude, i.e. with a significance less than 3σ, is also increased from 45.5% to 83.9%. To conclude, according to these test results, the training stage will permit: firstly to rule out the vast majority of the bad candidates with a successful classification closer to the real proportion of identified RFIs of 92.2% (see Table 7.11); secondly to hugely decrease the inquired time for the ultimate validation of the potential good candidates.

7.4 Results

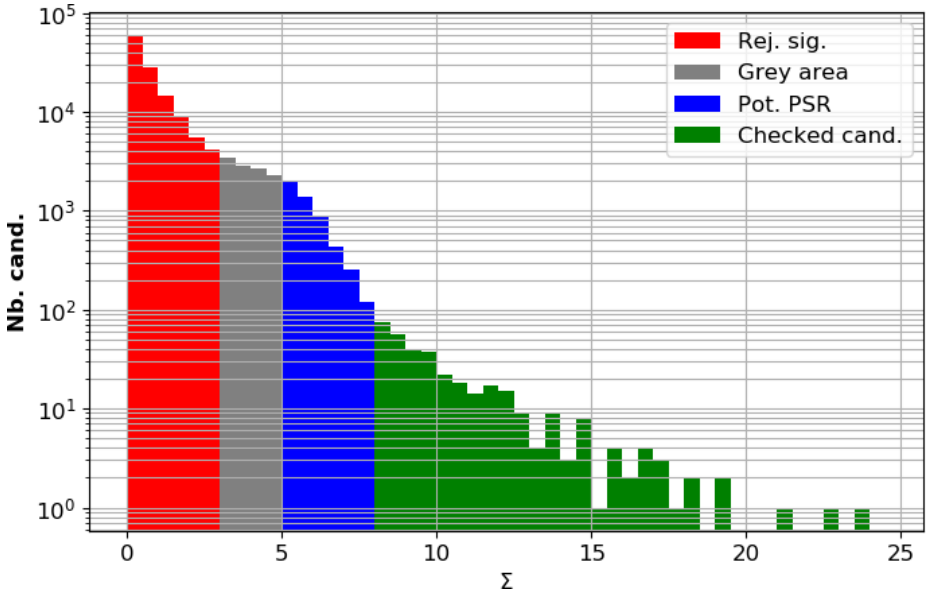
In the 8 months of observations processed, 137 200 candidates have been found by the search pipeline. The analysis of these candidates was carried out in three weeks using two processing nodes: one of the machines of the Nançay data center and Baudroie at LPC2E.

After classification using the coefficients indicated in Table 7.12, a very large majority of these candidates have been classified in the RFI class (i.e. with a global significance less than 3σ). In the details, 120 396 representing 87.6% of all the candidates are definitely excluded (red area in the histogram in Figure 7.11). The grey area (also in grey in Figure 7.11) between 3σ and 5σ includes 11 254 candidates (8.2% of all the candidates). Finally, 3.8% of the candidates have a significance greater than 5σ (in blue and green in Figure 7.11) and are kept as a potential pulsar, decreasing thereby the number of candidates to check from 137 200 to only 5 245.

The first control is to check against the list of the already known pulsars. The list of known pulsars is based on version 1.68 of the ATNF catalog and includes all the pulsars visible in Nançay, i.e. with a declination greater than -43°. For each candidate, the pulsars with a DM closer than 0.1 pc.cm^{-3} from the DM of the candidate are selected. The periods of these pulsars are then compared to the period of the candidate. Finally, we consider a good match between a pulsar and the candidate if the difference in the period is less than 1 ms.

The search in DM and period is imperfect, and the chosen range of difference is set to a large value in order to take account of the potential uncertainties. Concerning the DM, due to scatter broadening, the determination of the DM can be corrupted, especially for the highest DM. Thus, for the first selection in DM, a large window (for the frequencies of NenuFAR) is used to take account of an eventual large scattering variation at the moment of the observation.

Figure 7.11: Histogram of the global significances obtained for the 137 200 candidates of the survey. The red area represents the rejected candidates with $\Sigma < 3\sigma$. The grey area corresponds with the candidates with a global significance between 3 and 5 (the equivalent of the class POT in the training set). The candidates evaluated as a potential pulsar signal, i.e. with $\Sigma \geq 5\sigma$, are in blue and green, the green area representing those already manually checked.



Concerning the period, according to the precision reached with the pulsars, an error of 1 ms is important, especially for pulsars with a short period as MSP. However, due to our time sampling, we don't expect to find MSP, but slower pulsars with periods of at least about $80 - 100$ ms. Also, because of the FFT method used, the determination of the period of slow pulsars is less accurate. Consequently, for the majority of the expected pulsars, a deviation of 1 ms could be too constraining, allowing to increase the uncertainty on their period.

Also, the period derivative is unconsidered as a proper matching parameter. Because of a long time sampling of data, the search for the period derivative carried out by PREPFOLD is not very precise. As a result, the found period derivatives for candidates are generally large. That is equally the reason why the χ^2 curve of the period derivative is very narrow in most cases.

7.4.1 Re-detection of known pulsars

On the 5 245 candidates with a significance greater than 5σ, 40 matches with a known pulsar. These candidates have been manually inspected and validated. Nevertheless, of these, five candidates have coordinates very far from the coordinates of the pointing (from many degrees to some dozens of degrees). Because of the sidelobes of the telescope beam, it is possible to find a pulsar that is completely outside the main beam. However, according to their position, the corresponding pulsars are either below the horizon or at least at a very low elevation. As a consequence, the probability that these five candidates correspond to the matching known pulsar is either null or very weak. However, their signal looks like a pulsar signal and can be kept as interesting candidates for re-observation (see Section 7.4.2).

Table 7.14 lists the 35 candidates matching with a known pulsar with their period P and DM, ordered relative to their global significance Σ (see equation (7.5) and Table 7.12). The first column indicates the position in the list of all the candidates ordered relative to their global significance. We can notice that the best one in Table 7.14 is in second place rather than the first one. Indeed, some of the best candidates come from a unique observation where a large part of the frequencies have been zapped. The number one candidate is therefore, unfortunately, an artifact and doesn't look like a potential pulsar.

Table 7.14: List of the candidates with a global significance Σ greater than 5σ corresponding to a known pulsar. The first column Pos indicates their position relative to their global significance in the list of all the candidates.

Pos	**Cand. ID**	**P** (ms)	**DM** ($pc.cm^{-3}$)	**Σ**	**PSR**
2	C07000001	1292.27	5.75	22.877	J0814+7429
3	C08020001	1292.27	5.75	21.333	J0814+7429
4	C07020001	1292.29	5.76	19.427	J0814+7429
7	C54120001	538.44	43.50	18.147	J2219+4754
9	C35240001	739.69	19.61	17.370	J1509+5531
12	C33200001	739.69	19.61	16.544	J1509+5531
15	C10260001	1292.29	5.76	15.892	J0814+7429
16	C48160001	1656.48	9.19	15.793	J1115+5030
18	C33500001	1652.84	26.77	15.518	J1840+5640
32	C51680001	538.43	43.50	13.722	J2219+4754
33	C59160001	538.44	43.49	13.683	J2219+4754
34	C12700001	1292.23	5.75	13.600	J0814+7429
39	C31500001	1652.84	26.77	12.862	J1840+5640
43	C09100001	1292.32	5.75	12.651	J0814+7429
48	C31520001	1652.84	26.77	12.251	J1840+5640
50	C34360001	340.74	14.60	12.202	J0454+5543
52	C06060001	1292.30	5.74	12.184	J0814+7429
58	C04380001	1292.24	5.79	11.914	J0814+7429
96	C37340001	739.69	19.62	10.617	J1509+5531
103	C50540001	1656.48	9.21	10.415	J1115+5030
106	C09120001	1292.30	5.80	10.338	J0814+7429
153	C10280001	1292.22	5.79	9.423	J0814+7429
195	C32260001	714.53	26.76	8.853	J0332+5434
197	C08040001	1292.35	5.76	8.837	J0814+7429
227	C54140001	538.43	43.49	8.540	J2219+4754
266	C20460001	739.77	19.59	8.210	J1509+5531
281	C12800001	253.06	2.98	8.109	J0953+0755
405	C61720001	538.44	43.50	7.514	J2219+4754
770	C09020001	1292.22	5.81	6.835	J0814+7429
1327	C15540005	253.06	3.01	6.326	J0953+0755
2176	C20120007	253.05	2.96	5.897	J0953+0755
2441	C11400001	1292.33	5.74	5.801	J0814+7429
3089	C08000001	1292.23	5.78	5.569	J0814+7429
3965	C15520013	253.05	2.96	5.312	J0953+0755
4337	C05200001	1292.23	5.79	5.219	J0814+7429

The last column indicates the pulsar corresponding with the characteristics of the candidate. Some pulsars correspond actually to several candidates, i.e. they are detected in several pointings. Table 7.15 list the 8 known pulsars detected in the currently processed data, with their coordinates, period, and DM from the ATNF catalog. According to the NenuFAR pulsar census (Bondonneau et al. in preparation), J0814+7429 is the most intense known pulsar in the sky observed by the survey. The last column indicates the number of detections of the pulsar among the 35 candidates of Table 7.14. We can see J0814+7429 is detected 15 times, representing alone 35.71% of the 40 candidates.

The detection in 15 different pointings of J0814+7429 demonstrates the fact that the beam of a phased array is not merely limited to the gain attenuation limit of -3 dB, allowing the detection of an intense signal outside the beam. In Table 7.15, there is another interesting pulsar highlighting the properties and difficulties of the beam of a phased array. The survey has observed the sky above 39°. However, one of the eight pulsars is J0953+0755, with a declination of about 7°, i.e. 32° below the minimum declination reached by the survey. After checking the four corresponding candidates, it appears that the detections are very clear and unambiguous. The four pointings are indeed all at the correct right ascension, but at a declination of about 65°! These detections at about 60° of the good coordinates reveal the presence of the secondary lobes of the beam, which can possibly, therefore, provide disturbing signals within those of the primary lobe.

Comparing the known pulsars detected by the candidate analysis with those detected in the validation step of the processing pipeline (see Table 6.6 in Chapter 6.4.4), we can notice there are differences. Indeed, pulsars J0700+6418 and J1813+4013 are absent on the list in Table 7.15. For the case of J0700+6413, because of the fact that is a binary pulsar, it was the harmonic 3:1 which has been detected. However, for the moment, the harmonics of the known pulsars have been searched only for the candidates with a significance greater than 8σ. By searching in all the harmonics 3:1, it finally appears for a significance of 7.19, explaining the absence in Table 7.15.

In the case of J1813+4013, it was the second faintest pulsar detected. The resulting plot of the detection showed a very faint signal in all eight diagnostic plots. Its detection was possible only because, firstly, we have known that this pulsar might be in this pointing, and, secondly, due to the similarity between the faint obtained profiles with the profile of the NenuFAR pulsar census. As a consequence, it is not surprising that its global significance is low, even lower than 5σ. Finally, after a deeper search, it appears with a very low global significance of merely 2.64σ.

In the opposite case, there are four new detections compared to the list in the table of the re-detections in Chapter 6.4.4. One of the four is J0953+0755, which was obviously unsearched, because of its declination more than 30° below the minimum declination of the survey. For two of the four: J0332+5434 and J2219+4754, they are actually detected rather far from their sky position. In the previous chapter, the known pulsars were searched only at the closest pointing. However, for these two pulsars, the closest pointing of phase 1 is yet unprocessed, leading that the pulsar being consequently found in other pointings.

Concerning J0332+5434, in the results of the census, this pulsar is particularly strong and easily detectable. However, the candidate where the pulsar is found is located at 3.5° from the correct right ascension and at 3° from the correct declination. Considering an average beam diameter of about 1.5°, that means that the pulsar is found at a distance of approximately three beams from its real coordinates.

For J2219+4754, it is an intense pulsar, which is found more often than J0332+5434 with five detections. This is due to the fact that pointings relatively close to the closest pointing are processed. The detections are thereby done in pointings located at a distance slightly greater than 2° from the correct coordinates.

7.4.2 Candidates to re-observe

Once known pulsars were identified, the other 5 205 candidates with a global significance greater than 5σ require to be manually checked. This second control is made per level of significance starting from the greatest, which has the best probability of being a potential pulsar signal. At the 31^{st} October 2022, the candidates with $\Sigma \geq 8\sigma$ have been checked. Without the 27 known pulsars (from the previous list of 40), that is 271 candidates, representing the best 5.93% of all the candidates with $\Sigma \geq 5\sigma$.

Table 7.15: List of the 8 known pulsars re-detected by the survey at the 31^{st} October 2022. The characteristics (RA-DEC, P_0, and DM) are provided by the ATNF catalog. The last column indicates the number of candidates corresponding to the pulsar, corresponding also to the number of pointings where the pulsar is detected.

PSR	**RA**	**DEC**	**P** (ms)	**DM** (pc.cm^{-3})	**Nb. of corresp. cand.**
J0332+5434	03:32:59.4096 (1)	+54:34:43.329 (1)	714.519699726 (4)	26.7641 (1)	1
J0454+5543	04:54:07.7506 (1)	+55:43:41.437 (2)	340.729436235 (10)	14.59002 (15)	1
J0814+7429	08:14:59.50 (2)	+74:29:05.70 (11)	1 292.241446862 (3)	5.75066 (48)	15
J0953+0755	09:53:09.3097 (19)	+07:55:35.75 (8)	253.0651649482 (9)	2.96927 (8)	4
J1115+5030	11:15:38.400 (5)	+50:30:12.29 (5)	1 656.439759937 (3)	9.18634 (26)	2
J1509+5531	15:09:25.6298 (1)	+55:31:32.394 (2)	739.681922904 (4)	19.6191 (3)	4
J1840+5640	18:40:44.5372 (1)	+56:40:54.852 (1)	1 652.8618528869 (20)	26.77163 (17)	3
J2219+4754	22:19:48.139 (3)	+47:54:53.93 (3)	538.4688219194 (12)	43.4975 (5)	5

The list of the 40 candidates in Table 7.14 presents only candidates corresponding to the fundamental period of a known pulsar. However, principally in the case of intense pulsars, it is also possible to detect some harmonics of the period. As a consequence, each of these remaining candidates has been manually checked, then compared with all the pulsars of the ATNF until harmonics 1:20 and 20:1.

Finally, of these 271 candidates, 43 have been identified as harmonics of J0814+7429, J1115+5030, J1509+5531, and J1840+5640. Most of the others are false positives and can be rejected. Although no clear pulsar signal (other than from known pulsars) has been found, 10 candidates listed in Table 7.16, present nevertheless a signal compatible with a faint pulsar signal. One can notice that the selected candidates feature periods shorter than 300 ms, which is relatively short compared to the core target of the survey. Furthermore, although we expect to find essentially slow pulsars with small DMs, except one, all have a DM greater than 40 $pc.cm^{-3}$.

Due to the coarse time sampling which is used, it is unlikely that the candidates with a period less than 100 ms are effectively a pulsar, especially if combined with a high DM. The PREPFOLD resulting plots for the six candidates featuring a period greater than 100 ms, ranging from 112 to 253 ms, are shown in Figure 7.12, 7.13, and 7.14. The profiles of these candidates have approximately a Gaussian shape, with the first one which could present a scattering tail.

The resulting plots of the interesting candidates with a period shorter than 100 ms are presented in Figure 7.15 and 7.16. We can see the low time resolution makes to appear the profiles as triangles. As a consequence, it is difficult to clearly identify these interesting candidates as a potential pulsar signal. However, the detection with high reduced χ^2 in a precise DM (plot in the bottom center), in a precise period, and in a precise period derivative (the two plots on the top right) shows nevertheless a behavior similar to that expected for a pulsar signal.

The expectations were essentially about slow pulsars, i.e. with a period of at least several hundreds of ms. However, no slow candidates have been identified during this first pass of candidates with Σ greater than 8. Thus, an alternative expanded search was performed for all the candidates with a global significance Σ between 5 and 8 and featuring a period greater than 500 ms. 120 candidates have been thereby manually checked and two of them, presented in Figure 7.17, have been added to the list of interesting candidates to re-observe. Their characteristics are listed in the middle part of Table 7.16.

In addition to these 12 potential candidates, 5 candidates corresponding to unlikely detections of a known pulsar have been added to the list of interesting candidates to reobserve. Of the 40 candidates corresponding to a known pulsar, these five candidates feature a period and a DM similar to the characteristics of a known pulsar but are far away from the sky position of the pulsar. The resulting plots for these five candidates are presented in Figure 7.18, 7.19, and 7.20, and show signals potentially similar to faint pulsars.

Moreover, above all, there is only the first one in Figure 7.18 and the one in Figure 7.20, for which the corresponding pulsar was above the horizon. As a consequence, for those below the horizon, no telescope beam sidelobes can explain the detection, and they are real candidates for re-observation. For the two others, according to the NenuFAR pulsar census, the corresponding pulsars are faint, meaning that the probability that they can be detected by a beam sidelobe is extremely low. As a result, these two candidates can be kept as interesting candidates to re-observe.

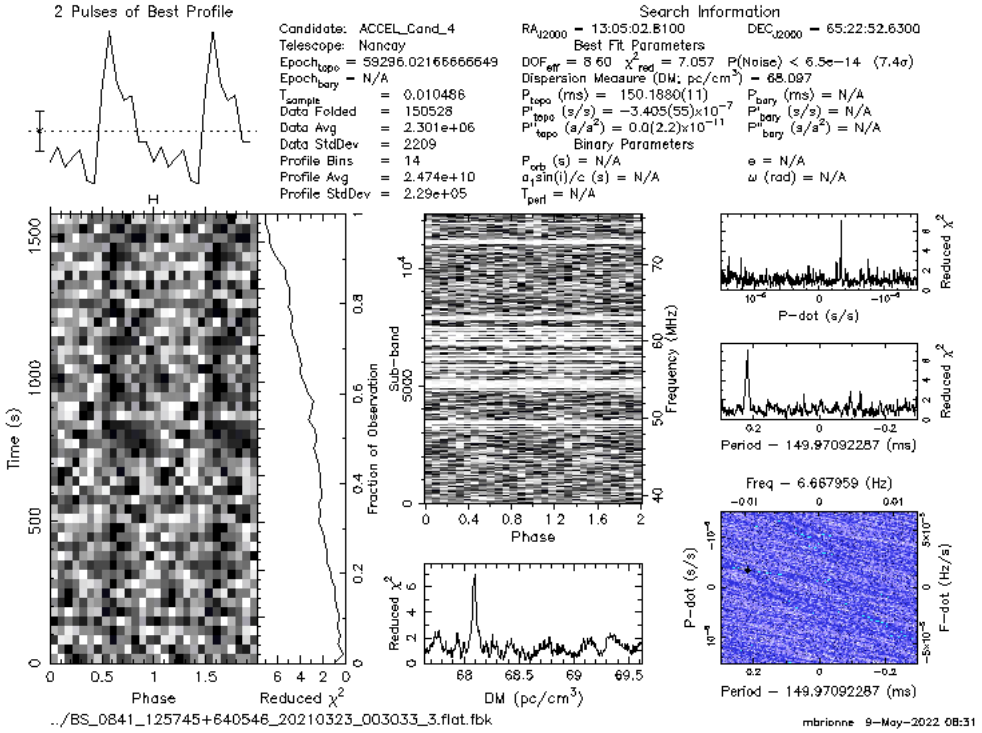
Ultimately, a list of 17 interesting candidates (see Table 7.16) is selected from the first eight months of processed data. These interesting candidates will be re-observed during 2023 in the same observing conditions as for the survey. In the case of a similar detection several times, the candidate will be defined as a discovery and observed with a standard pulsar setup to precisely determine its characteristics.

The remaining data will be processed in 2023. The following analysis of the candidates found by the processing pipeline should allow us to increase the list of potential pulsar signals. Currently, 32.9% of the pointings are processed and analyzed, leading to 17 interesting candidates. According to these first results, we can therefore estimate the final number of pulsar candidates might reach approximately 52.

Table 7.16: List of interesting candidates to re-observe. The first part corresponds with candidates (without any known pulsar matching) having a significance Σ greater than 8 which looks like a potentially weak pulsar signal. The second part corresponds with slow candidates (period greater than 500 ms) with Σ between 5 and 8 showing a potential pulsar signal. The third part indicates candidates matching with a known pulsar but with an uncertain or null probability to correspond to the matching pulsar.

Type	**ID cand**	**ID pt**	**RA**	**DEC**	**P** (ms)	**DM** ($pc.cm^{-3}$)	**Σ**
Potential candidate	C02460001	P0246	10h35m54.925s	+80d39m33.2169s	66.04	46.75	16.992862
	C23340001	P2334	08h24m59.857s	+62d18m23.0884s	99.71	67.53	11.075839
	C17200003	P1720	13h05m02.8113s	+65d22m52.6414s	150.19	68.10	10.323366
	C14240001	P1424	11h45m21.3242s	+67d27m35.3004s	79.39	47.54	10.137329
	C02460005	P0246	10h35m54.925s	+80d39m33.2169s	192.72	42.83	9.706041
	C51680009	P5168	22h32m59.1225s	+48d21m33.9413s	118.94	52.98	8.860400
	C18680001	P1868	12h19m08.5382s	+64d21m03.723s	87.19	39.07	8.806965
	C00800004	P0080	08h35m26.2434s	+85d28m37.0044s	112.34	56.77	8.781099
	C00400002	P0040	11h13m03.2967s	+86d43m43.094s	210.95	12.13	8.370859
	C20300008	P2030	12h46m33.5468s	+63d19m34.3584s	253.82	58.43	8.143770
Slow candidate	C30800007	P3080	09h58m59.3592s	+58d16m11.6611s	527.07	39.23	5.307475
	C35500005	P3550	18h11m01.2976s	+55d16m38.9662s	561.00	59.47	5.268478
Unlikely PSR	C71760006	P7176	16h02m25.5817s	+40d27m47.5565s	85.10	18.00	6.154808
	C25260006	P2526	10h45m20.3917s	+61d17m28.5112s	95.16	21.07	5.778986
	C60240018	P6024	08h34m03.5656s	+45d24m11.5147s	223.17	25.30	5.307813
	C26820014	P2682	07h24m15.3568s	+60d16m49.2766s	153.63	23.93	5.073638
	C21620003	P2162	07h58m33.5468s	+63d19m34.3584	492.40	62.95	5.028297

Figure 7.12: Candidates with a global significance greater than 8σ and a period greater than 100 ms showing a potential pulsar signal. Potential candidates C17200003 and C02460005.



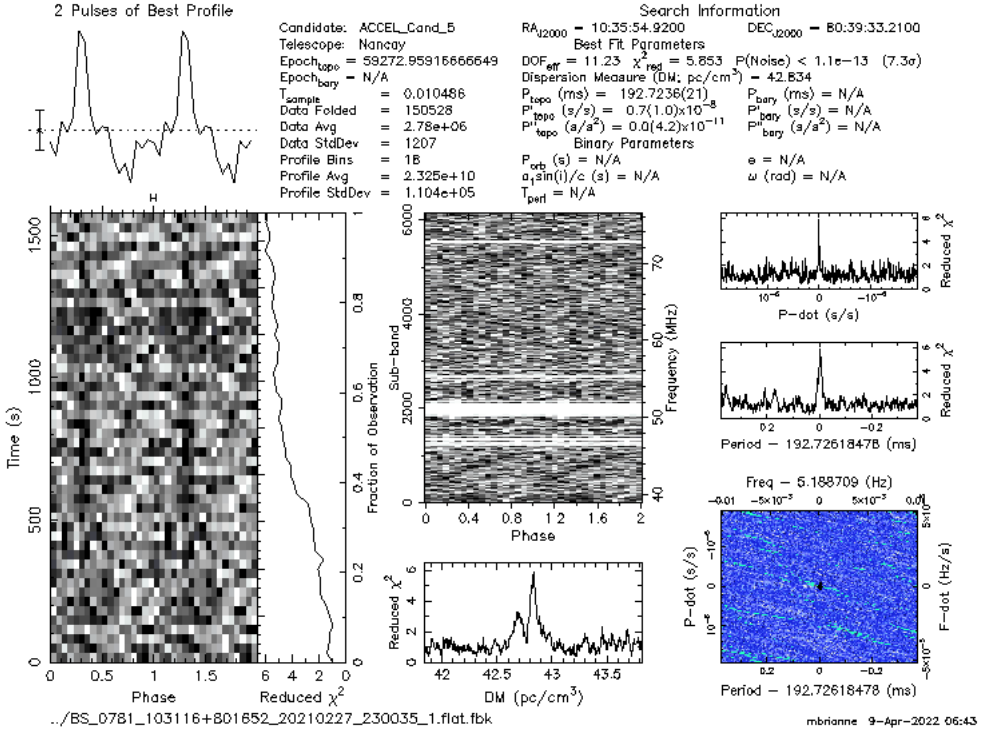
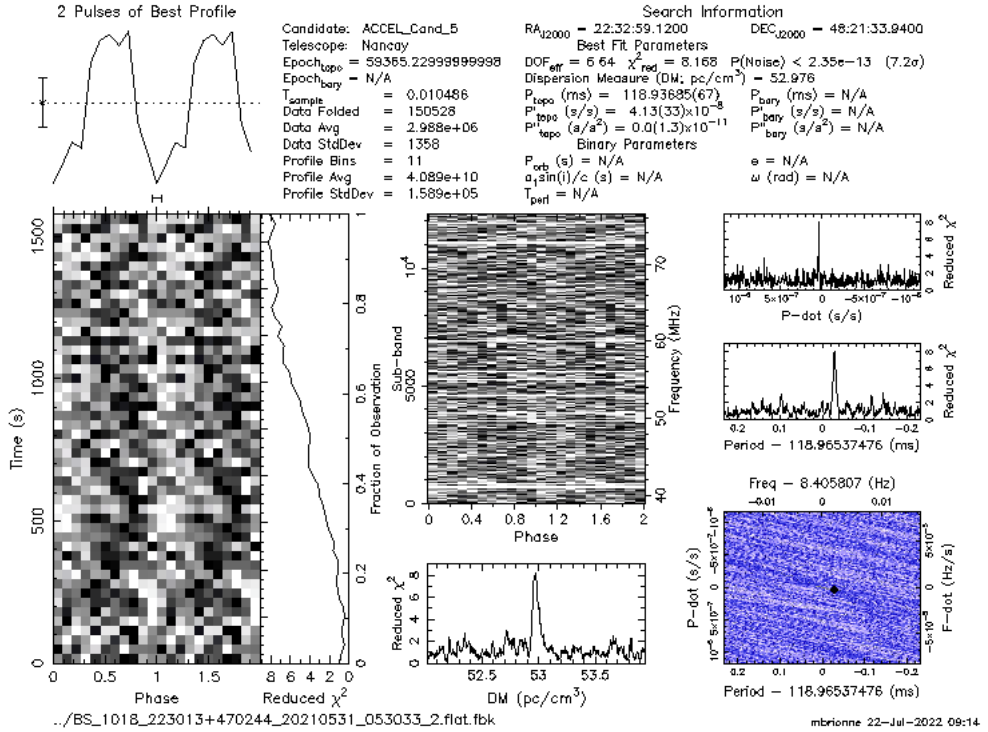


Figure 7.13: Candidates with a global significance greater than 8σ and a period greater than 100 ms showing a potential pulsar signal. Potential pulsar candidates C51680009 and C00800004.



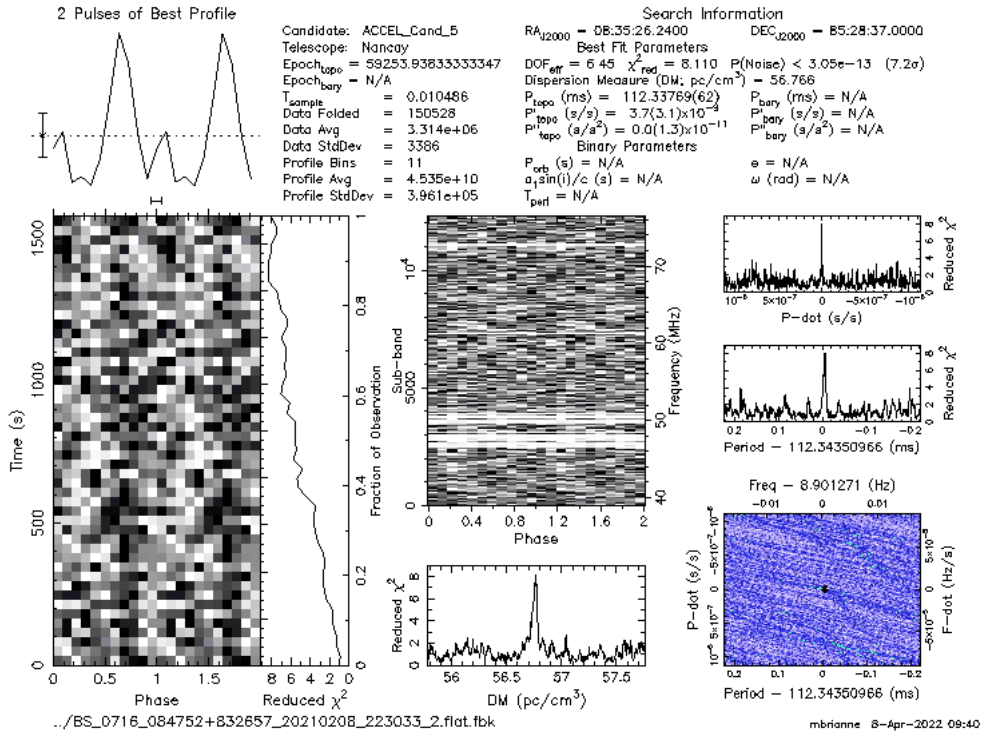
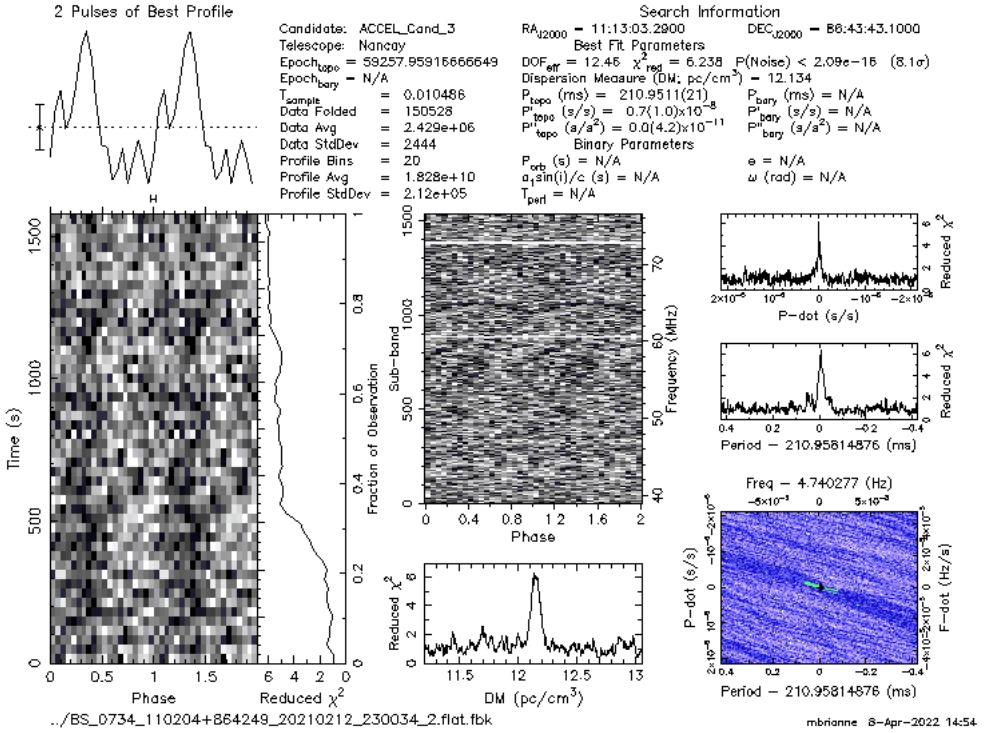


Figure 7.14: Candidates with a global significance greater than 8σ and a period greater than 100 ms showing a potential pulsar signal. Potential pulsar candidates C00400002 and C20300008.



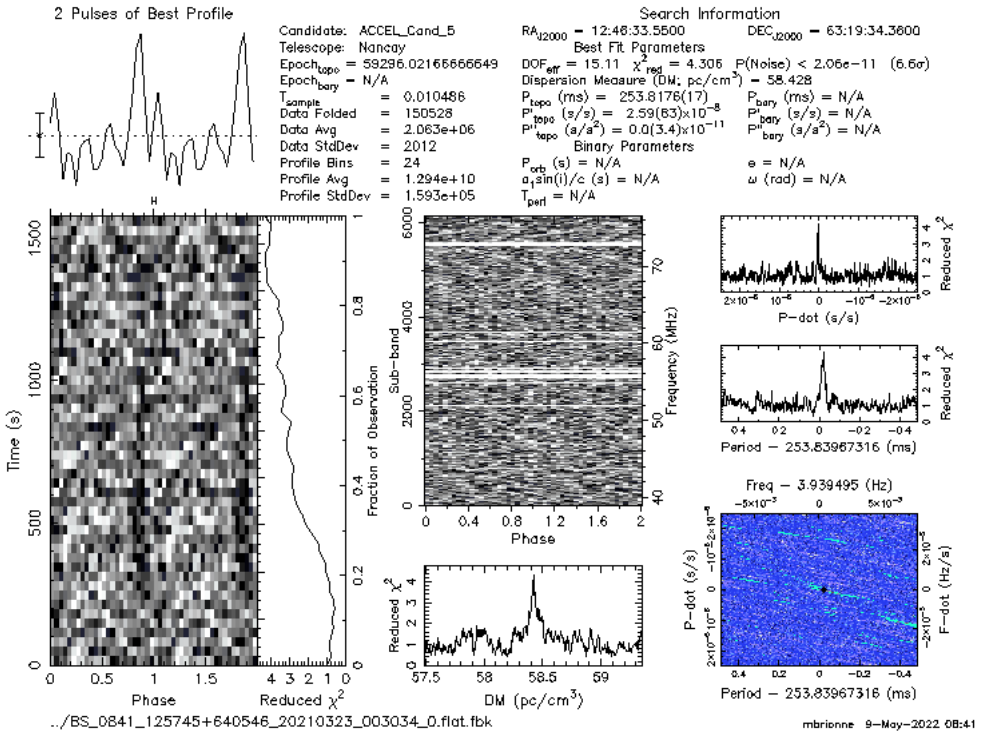
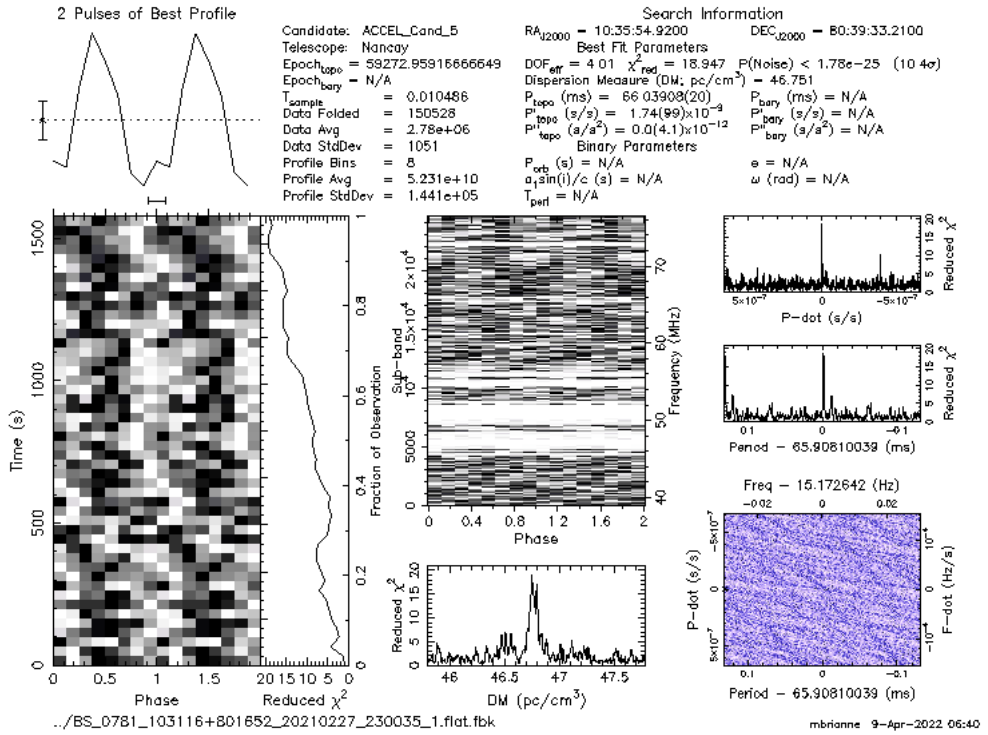


Figure 7.15: Candidates with a global significance greater than 8σ and a period less than 100 ms showing a potential pulsar signal. Potential pulsar candidates C02460001 and C23340001.



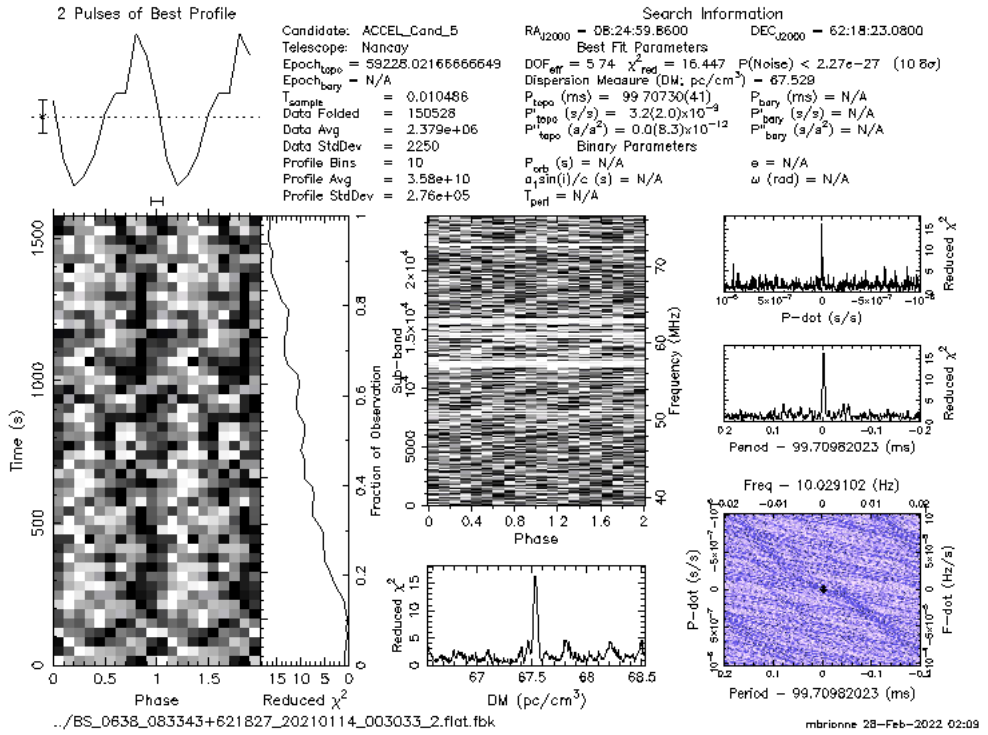
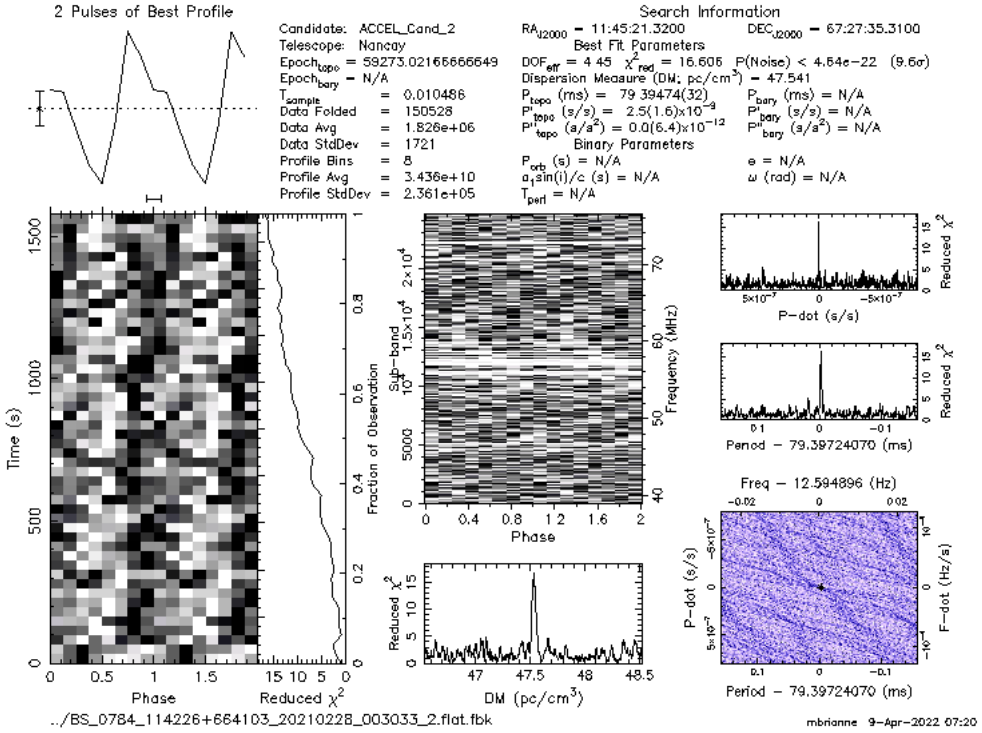


Figure 7.16: Candidates with a global significance greater than 8σ and a period less than 100 ms showing a potential pulsar signal. Potential pulsar candidates C14240001 and C18680001.



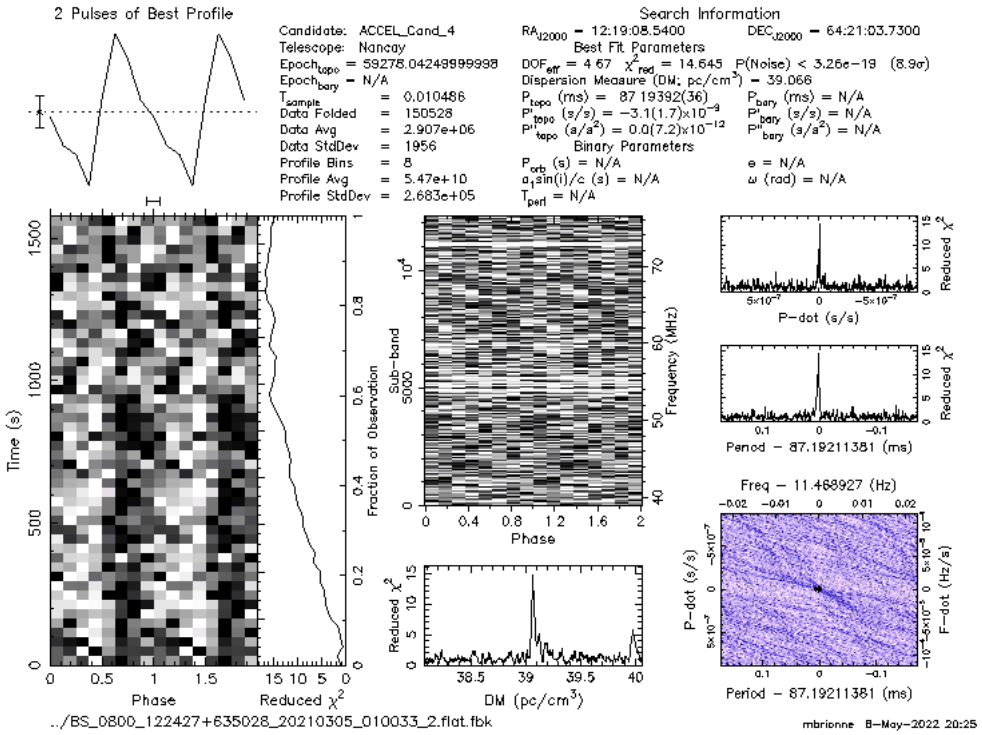
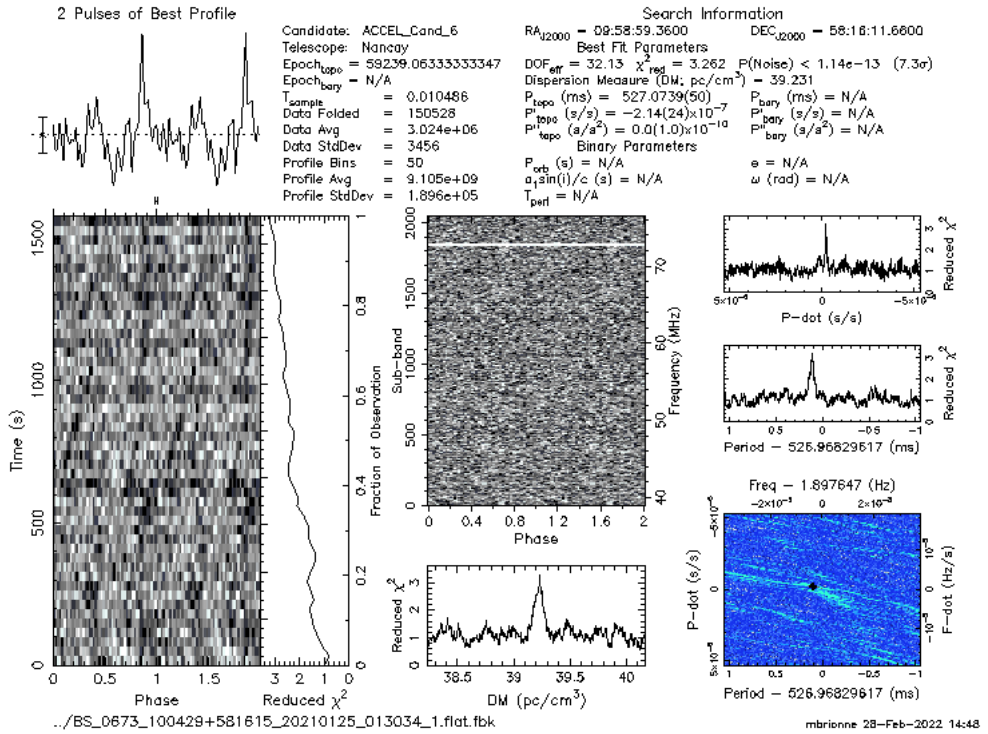


Figure 7.17: Slow candidates with a period greater than 500 ms and a global significance between 5 and 8σ showing a potential pulsar signal. Slow candidates C30800007 and C35500005.



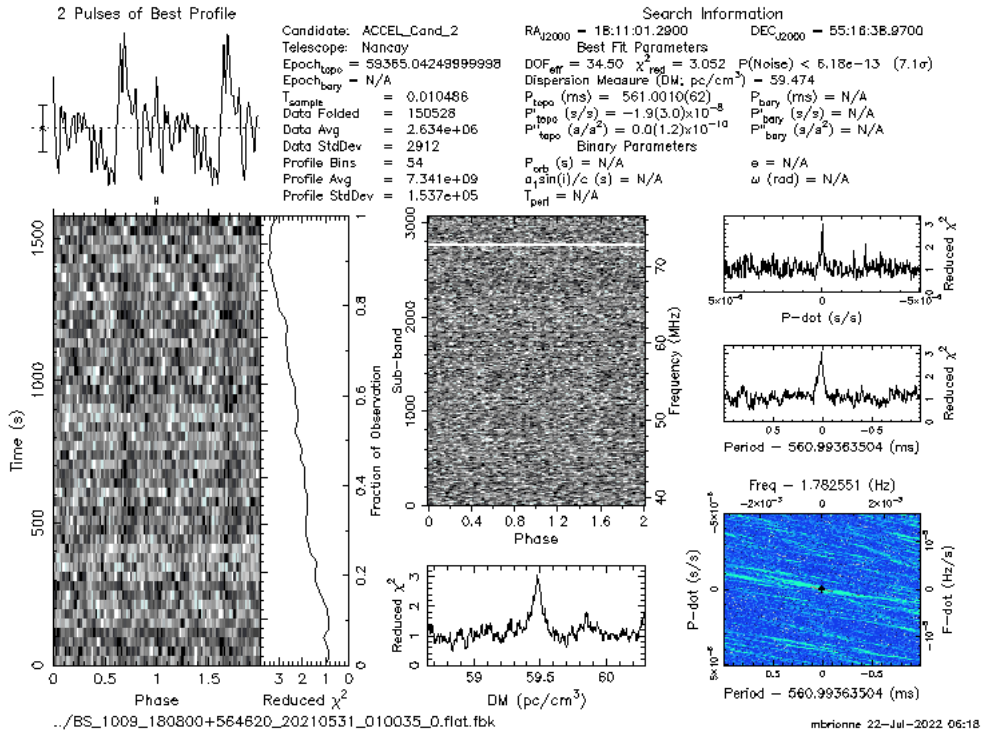
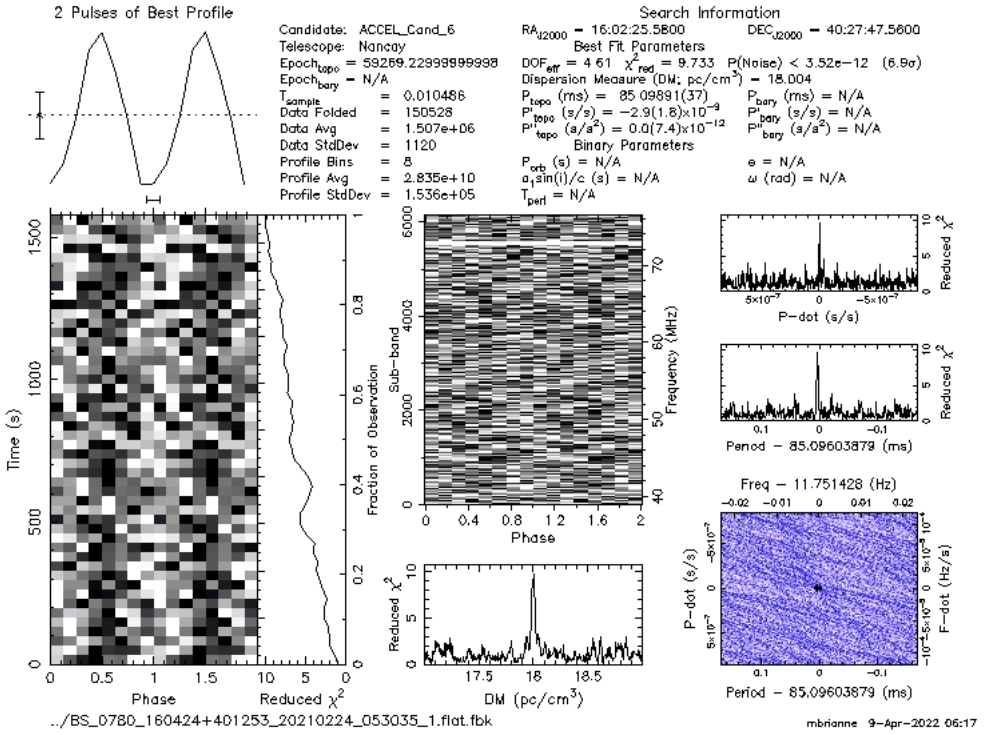


Figure 7.18: Candidates whose period and DM potentially correspond to a known pulsar, but whose coordinates are far from the survey pointing coordinates. Due to the low probability of effectively corresponding to the pulsar, they need to be re-observed. Unlikely pulsar matching C71760006 and C25260006.



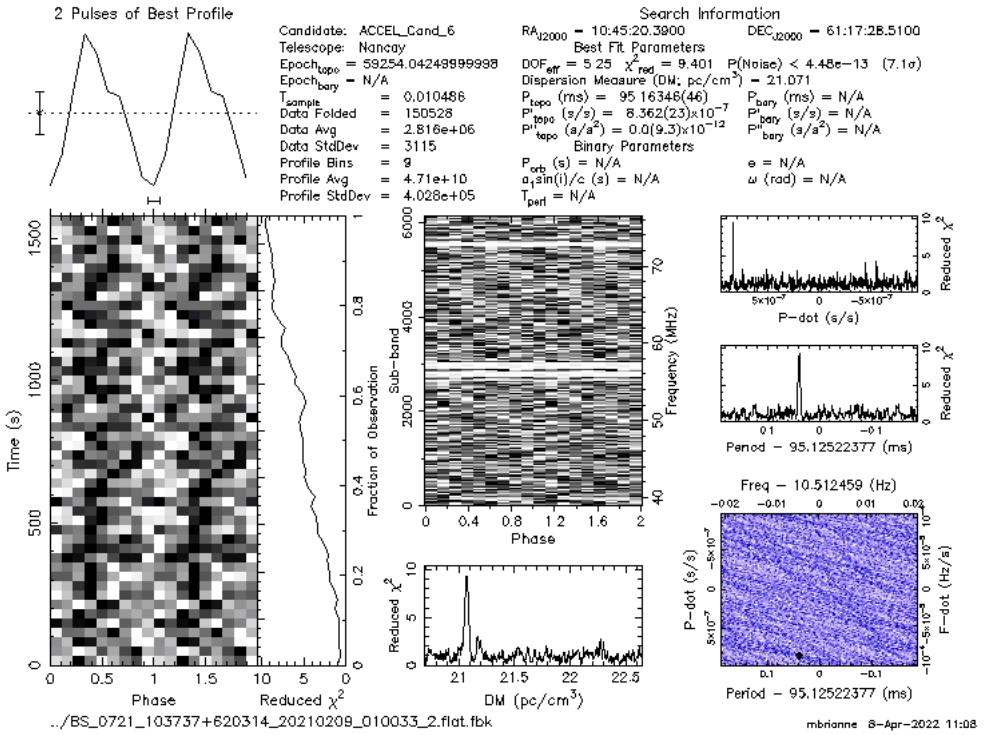
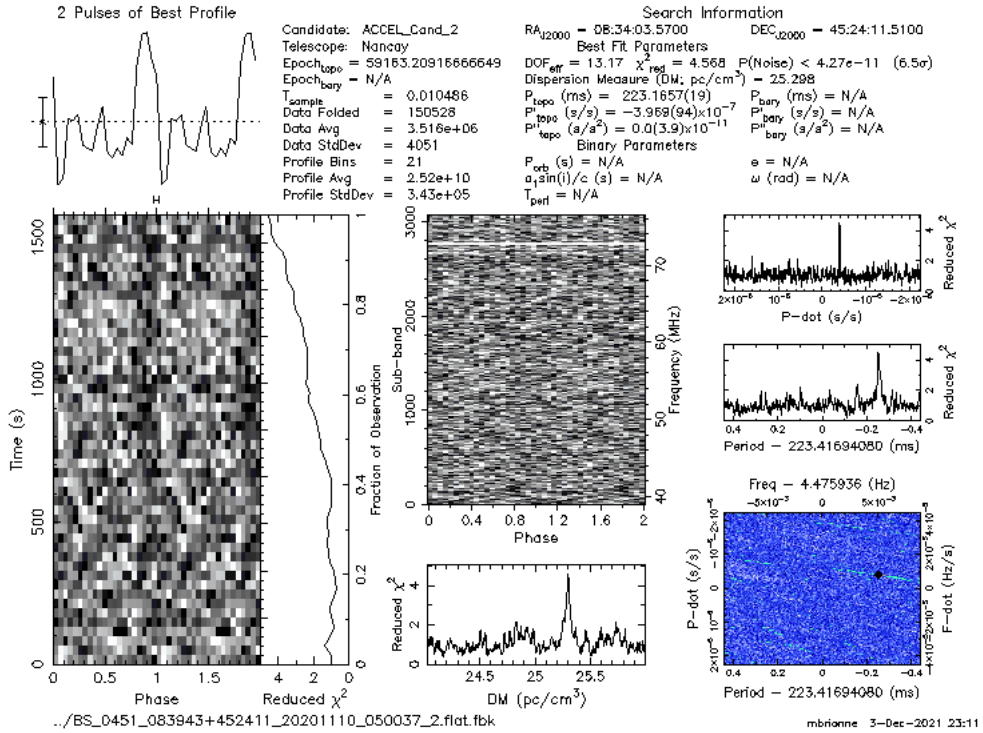


Figure 7.19: Candidates whose period and DM potentially correspond to a known pulsar, but whose coordinates are far from the survey pointing coordinates. Due to the low probability of effectively corresponding to the pulsar, they need to be re-observed. Unlikely pulsar matching C60240018 and C26820014.



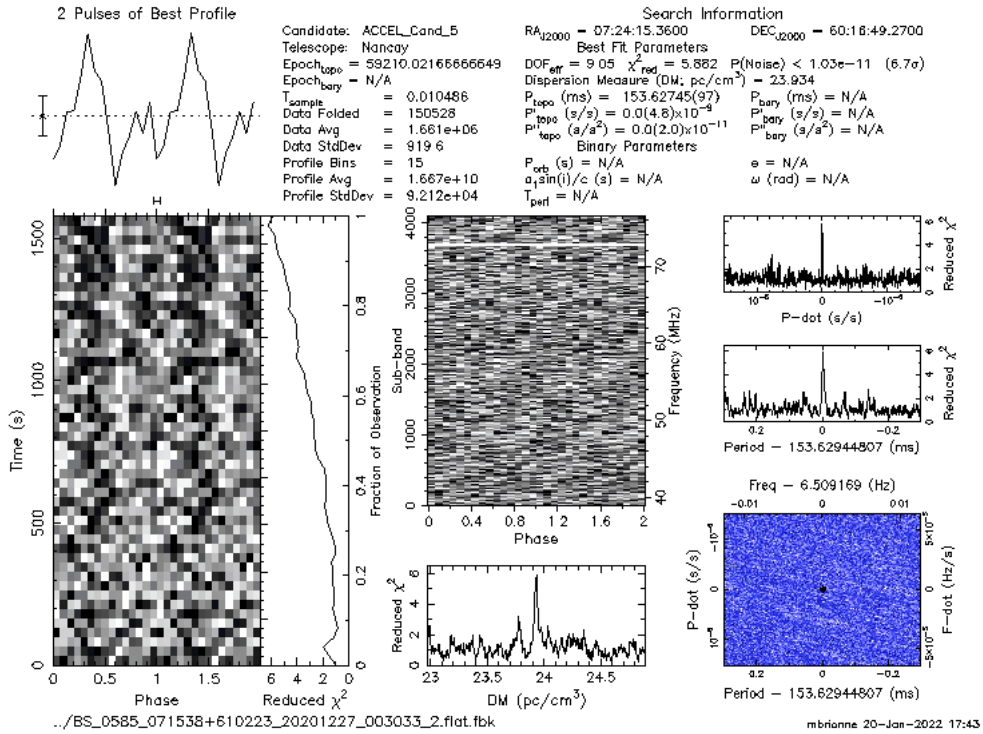
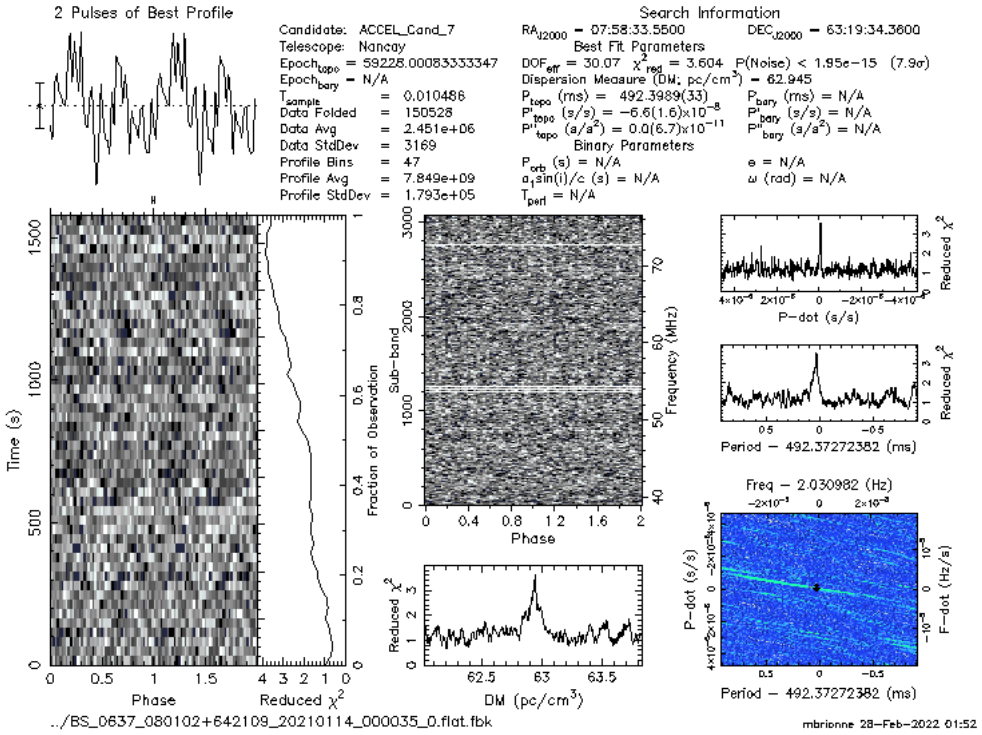


Figure 7.20: Candidates whose period and DM potentially correspond to a known pulsar, but whose coordinates are far from the survey pointing coordinates. Due to the low probability of effectively corresponding to the pulsar, they need to be re-observed. Unlikely pulsar matching C21620003.



Résumé du chapitre :

Dans le cadre de ce relevé, il est attendu notamment d'éventuellement trouver des pulsars disposant de plutôt faibles flux. La recherche de candidats pulsars effectuée génèrent ainsi, due aux seuils faibles employés, un nombre très important de candidats. Dans le but d'extraire les candidats les plus intéressants, j'ai mis au point un script Python permettant de trier les candidats en les comparant à des modèles de pulsar.

La méthode employée est basée une comparaison des données réelles générées par PREPFOLD avec des simulations Monte-Carlo. Pour chacun des huit diagrammes de contrôle d'un candidat créer par PREPFOLD, une significativité calculée comme étant le rapport de vraisemblance entre le meilleur modèle pulsar et un modèle nul est déterminé. Une significativité globale est finalement obtenue en effectuant une moyenne pondérée des différentes significativités individuelles.

Le but de cette première analyse est de classifier rapidement l'important nombre de candidats. Pour ce faire, la méthode utilisée doit être numériquement légère et rapide. La simulation Monte-Carlo a l'avantage d'être basée sur des modèles mathématiques, et peut donc être implémentée facilement pour une application rapide. Elle a ainsi été préférée à d'autres méthodes telles que l'apprentissage profond, qui nécessite pour sa part de posséder un suffisamment large et varié ensemble d'exemples pour effectuer un long apprentissage.

Toujours dans l'optique d'efficacité, le modèle de base choisi a été simplifié à un modèle purement gaussien ne prenant pas en compte l'étalement exponentiel du au phénomène de scattering. Comme notre résolution temporelle est basse et que les cibles privilégiées sont les pulsars lents ne subissant que peu le scattering, l'utilisation d'un modèle plus complexe tel qu'une gaussienne modifiée exponentiellement aurait un coût en temps de calcul très probablement bien supérieur à l'augmentation du nombre de découvertes finales.

Trois des diagrammes de contrôle sont le profil intégré, le plan temps-phase, et le plan fréquence-phase. Pour ces trois ci, le modèle est donc basé sur une fonction gaussienne avec quatre paramètres à ajuster : l'amplitude, la phase centrale du profil, la largeur du profil, et enfin l'amplitude du bruit. Pour les plans bidimensionnels, cette fonction de base est multipliée à une fonction constante en temps, et par une fonction de modulation de type polynôme du second degré en fréquence nécessitant l'ajustement de deux paramètres supplémentaires.

Un digramme relié au plan temps-phase mesure la linéarité de l'évolution du χ^2 du profil en fonction du temps. Ce digramme est lui comparer à un simple modèle linéaire.

Les quatre diagrammes restants sont reliés à l'évolution du χ^2 du profil intégré en temps et fréquence lorsque l'on fait varier légèrement une des caractéristiques du candidat. Ces diagrammes sont donc directement reliés aux recherches fines en période, dérivée de période, et mesure de dispersion effectuées par PREPFOLD. Dans le cadre de ces recherches, le modèle utilisé pour approcher cette évolution est une fonction quadratique centrée sur la meilleure valeur.

Finalement, le dernier digramme de contrôle est le plan $P - \dot{P}$, qui est la combinaison des recherches fines en période et dérivée de la période. En conséquence, le modèle utilisé ici est une multiplication de deux fonctions quadratique, dont une dépendant de la covariance entre l'erreur en période et l'erreur en dérivée de la période.

Afin de pouvoir déterminer des significativités globales suffisamment justes pour extraire les candidats pulsars intéressants, il est nécessaire d'évaluer de justes coefficients de pondération. Pour ce faire, une phase d'entraînement a été réalisée en utilisant un panel de 6 170 candidats, dont 72 correspondant à des pulsars connus. Les candidats ont été classés en trois catégories : RFI, signaux potentiels, et signaux pulsars. Les meilleurs coefficients correspondent à ceux permettant d'avoir le meilleur taux de bonne classification pour l'ensemble des trois catégories, qui est de 98 % pour les coefficients finalement retenus.

En sus des coefficients de pondération, un facteur important de réussite d'une simulation Monte-Carlo est la bonne couverture de l'espace des paramètres. Ainsi, le nombre de modèles à utiliser pour chacun des huit diagrammes de contrôle a été estimé par des tests de convergence de significativité sur deux candidats correspondant à deux pulsars connus.

Sur les huit premiers mois de données traitées, 137 200 candidats ont été trouvés. L'analyse de ces candidats a finalement permis d'exclure 87,6 % de ces candidats. Sur ceux restants, 5 245 candidats (soit 3,8 % des candidats initiaux) ont obtenu une significativité globale supérieure à 5 σ correspondant au seuil défini lors de la phase d'entraînement pour caractériser les signaux de pulsars connus.

Les caractéristiques de ces derniers candidats ont été comparées à celles de tous les pulsars connus. Au final, 40 candidats ont été identifiés à huit pulsars connus. Sur les candidats restants, seul ceux ayant une significativité globale supérieure à 8 σ ont été contrôlés manuellement. Dix candidats ont au final été sélectionner pour ré-observation entre décembre 2022 et mai 2023.

La cible principale du relevé étant les pulsars lents, tous les candidats ayant une période supérieure à 500 ms et une significativité entre 5 et 8σ ont également été vérifiés manuellement. Deux nouveaux candidats intéressants ont ainsi été ajoutés à la liste des candidats à ré-observer.

Pour finir, cinq candidats possèdent des caractéristiques similaires à cinq pulsars connus. Cependant, il est apparu que les positions des pulsars en question étaient très éloignées de celles des pointages des candidats. Certains pulsars étaient même sous l'horizon au moment de l'observation, et donc impossible à voir. De plus, par rapport aux résultats du census de NenuFAR, ces pulsars ne devraient pas être détectés. Dus au fait que ces candidats aient un comportement de potentiels pulsars dans plusieurs des diagrammes de contrôle, ces cinq candidats n'ont pas été exclus, et ont donc été rétrogradés de re-détection à candidats intéressant à ré-observer.

Chapter 8

Conclusion and further work

The NPBS is a pulsar survey observing between 39 and 76 MHz, i.e. at lower frequencies than the majority of the older surveys. According to expectations based on the standard model of the pulsars, a low-frequency survey shouldn't discover a number of pulsars comparable to higher-frequency surveys. Nevertheless, the observation of a particularly unknown part of the spectrum could provide discoveries of more exotic pulsars or related objects emitted especially at low frequencies. The project presented in the first part of this thesis was aimed to observe the entire sky above 39°, in order to discover new pulsars. With this in mind, 7 692 pointings of a solid angle about 0.28 deg^2 have been defined, allowing to cover approximately 98% of the targetted sky.

The observing program started in August 2020, with the goal to observe all the pointings before the end of this thesis, i.e. before the end of 2022. 1 156 hours of observation have been carried out, allowing to observe 98% of the pointing grid.

Using a processing pipeline adapted to the low-frequency constraints, a third of the data have been processed since December 2021. The 137 200 candidates found by the processing stage have also been analyzed with an automatic sorting algorithm. Finally, about 80% of the candidates have been excluded, and in the remaining candidates, 10 known pulsars have been identified (with the combination of the targetted search in Chapter 6 and the candidate analysis in Chapter 7), and 17 interesting candidates have been selected for re-observation.

To complete the observing program, the 176 remaining pointings are planned to be observed during spring 2023, resulting in the observation of the entire sky above 39° at the end of May 2023. In addition to the remaining pointings, the 17 selected candidates will be observed at least two times each between December 2022 and May 2023. After analysis of the data, additional observations will be done at the end of 2023.

The data processing is ongoing and is expected to be finished at the beginning of 2024. Based on the current number of candidates found on the first third of processed data, more than 400 000 candidates are expected for the whole survey. The analysis of the 250 000 – 300 000 new found candidates could be processed in about 2 or 3 months. Depending on the usable nodes, the analysis could be carried out in parallel or in a row, leading to obtaining the results for the whole survey between the beginning and the middle of 2024. With reobservations and processing of the eventual interesting candidates, the final result can be expected by the end of 2024.

The search method currently used and presented in this thesis is relatively simple, but fast and efficient to detect the majority of pulsars. However, some types of pulsars are hardly detectable with a search method based on the FFT. Further work for the NPBS is required in order to enlarge this initial search using other search methods.

The FFT is very efficient for continuous periodic signals. However, for pulsars showing a not continuous emission, such as pulsars featuring a nulling effect or the RRATs, the FFT is not an adapted search method. In the current search pipeline, a single pulse search is already performed for each pointing, but the results are unanalyzed. According to the NenuFAR census results, a few pulsars show detectable single pulses (about $20 - 30$ pulsars). Also, a single pulse analysis requires computing time to process the results of the single pulse search and time to inspect the results of the performed analysis. As a consequence, for the first stage of processing and analysis presented in this thesis, the single pulse analysis was unrealized. However, a single pulse analysis could be done using a dedicated program as the program RRATrap (Karako-Argaman et al. 2020) or the extended version Clustering (Josephy 2020).

Because of the substantial computing time required by the processing of the 7 692 pointings, the search was performed without acceleration search. Excepted for particular cases illustrated by the detection of J0700+6418, the standard search performed here shouldn't allow the detection of binary pulsars. Furthermore, the vast majority of binary pulsars are MSPs, which are undetectable by the NPBS. However, according to the ATNF catalog (v.1.68), there are, nevertheless, 61 non-MSPs binary pulsars (with a period greater than 30 ms). As a consequence, in the context of an extension of the carried-out search with new processing of the data, an acceleration search could be added.

According to the RFM, the altitude of emission is a function of the observing frequency, period, and period derivative of the pulsar. In particular, the lower the frequency, the highest the altitude of emission. It results from this higher emission a widening of the emission cone towards low frequencies. As a consequence, the probability that the beam of a pulsar crosses the line-of-sight of Earth is higher at a low frequency than at a higher frequency. In theory, by this geometric effect, a low-frequency survey could observe more pulsars. However, the short-period pulsars already have a wide cone at high frequency, leading that the widening lower. On the opposite side, although the long-period pulsars have a tighter cone than short-period pulsars, the effect of the widening with the frequency is larger. The consequence is the distribution of the low-frequency discoveries should be shifted towards long periods. Combined with the difficulties involved in observing at low frequencies (especially below 100 MHz), the slow and old pulsars are expected to be the core target of the NPBS.

This expectation has been confirmed by the recent discoveries of slow pulsars by the low-frequency survey LOTAAS performed using LOFAR. Indeed, the number of pulsars discovered by LOTAAS with periods of several seconds was greater than expected (Sanidas et al. 2019). In particular, LOTAAS has discovered a pulsar of 23.5 s: J0250+5854 (Tan et al. 2018), which was the slowest known pulsar at the time. Since J0250+5854, other very slow radio transients have been discovered, especially a pulsar with a period of 76 s: J0901-4046 (Caleb et al. 2022), and a supposed magnetar of 18 min: (Hurley-Walker et al. 2022), which was observed in X-ray with a radio counterpart. These recent discoveries suggest it exists a population of slow pulsars yet unobserved.

However, because of the loss in sensitivity of the FFT at long periods, the search method currently used is not adapted to efficiently find slow pulsars. In the context of a search of slow pulsars, the *fast folding algorithm* (FFA) is more adapted. Contrary to the FFT which used the properties of the Fourier transform, an FFA is based on the direct folding of the data to search a periodic signal. The FFA is a brute force method, requiring a large computing power and a long computing time. To start the processing of the NPBS data, adding an FFA to the standard FFT method would have considerably slowed down the processing, which is already very long. Some programs using the FFA exist to perform a pulsar search, such as the software RIPTIDE (Morello et al. 2020). Following the standard pulsar search, in the context of an extension of the search, a reprocessing of the NPBS data will have to be done by performing an FFA method.

The last chapter of this part has presented a first analysis of the candidates found by the search processing. The current analysis method aims to fastly classify the candidates, in order to sort and exclude highly unlikely candidates. This analysis method, based on Monte-Carlo simulations, permits reducing the number of candidates to manually verify. In this context, the method is currently relatively simple (to be swift) and could be improved to obtain a more precise and deeper classification.

Because of a time constraint, the number of pulsar models used to compare the data has been set to the minimum value allowing to obtain a reasonable significance. As a consequence, the first step of

improvement is to increase the number of models to obtain better coverage of the parameter spaces. A second step of improvement is to add scatter broadening using a *exponentially modified Gaussian* function rather than a simple Gaussian function as the base model for the pulse. Some improvements to the model could be included to detect the time and frequency variations of the emission, as for example a possible nulling effect. Nevertheless, these improvements require more substantial computing time, and the efficiency of the results compared to the cost has yet to be defined.

Finally, always in the context of the extension of the pulsar search, other types of analysis could be performed. The current method is based on the standard model of a pulse and thus may not be totally efficient to find pulsars slightly different from this basic assumption. Some analysis methods, based on deep learning, for example, could be added to the present analysis, conditionally to have a sufficient training stage.

Résumé du chapitre :

Cette partie de la thèse a présenté la mise au point d'un relevé de pulsars dans le ciel au-dessus de 39° de déclinaison. Pour ce faire, une grille de pointage de 7 692 pointages a été définie, permettant d'observer environ 98 % du ciel ciblé. Le programme d'observation à commencer en août 2020 et est au 31 août 2022, après 1 156 heures d'observation, terminé à 98 %.

Au 31 août 2022, huit mois de données correspondant à un tiers de l'ensemble du relevé ont été traités. Finalement, 137 200 candidats ont été trouvés, puis analyser en utilisant un algorithme de tri basé sur des simulations Monte-Carlo. 80 % des candidats ont été exclus, et 17 candidats intéressants ont été sélectionner pour être ré-observer.

Afin de terminer le programme d'observation, les 176 pointages restants seront observés durant le printemps 2023. Concernant les 17 candidats intéressants, ils seront eux observés chacun à deux reprises entre décembre 2022 et mai 2023.

Les deux tiers de données non traitées vont l'être courant 2023, et il est attendu que l'ensemble des données du relevé soit traité d'ici à début 2024. Par rapport au nombre de candidats trouvés dans les huit mois actuellement traités, le nombre de nouveaux candidats à analyser se situe ente 250 000 et 300 000, représentant un temps d'analyse de deux à trois mois.

La présente recherche de candidats était une recherche de signaux périodiques basé sur la FFT. Cette méthode est très efficace pour les signaux continus parfaitement périodique, et plus spécialement de période courte. Cependant, elle n'est pas adaptée à la recherche de signaux non continus tels que les phénomènes transitoires tels que les RRAT (rotative radio transient) (McLaughlin et al. 2006). Pour ce type de pulsars, une analyse basée sur la recherche d'impulsions individuelles pourrait être accomplie.

Pour diminuer le temps de calcul déjà très long, aucune recherche de pulsars binaires n'a été effectuée. Surtout, la très grande majorité de ces pulsars sont des pulsars milli-secondes impossibles à détecter due à la faible résolution temporelle. Cependant, dans le cadre d'une extension de l'analyse des données, une recherche en accélération pourrait être ajoutée.

La FFT est aussi une méthode peu adaptée pour les signaux de période relativement longue. Or, ces dernières années, plusieurs pulsars lents disposant de périodes allant de quelques secondes jusqu'à 18 minutes ont été découverts ; et en outre dans une proportion plus importante que prévue. Pour ce type de pulsars, des méthodes telles que les algorithmes d'empilement rapide (*fast folding algorithm*) sont très efficaces. Les pulsars ciblés comme cible potentielle du présent relevé, une analyse de ce genre devrait être effectuée sur les données du relevé.

Enfin, le dernier chapitre de cette partie portait sur la présentation d'une méthode d'analyse des candidats trouvés dans le but d'exclure les faux positifs trouvés lors du traitement des données. Cependant, cette méthode permet une analyse préliminaire dont le but est d'être légère et rapide, et est donc basée sur un modèle de pulsar plutôt standard. Ainsi donc, dans l'optique de détecter des pulsars différents de ceux-ci, une extension prenant en compte l'étalement dû au scattering, où des variations non continues en temps ou fréquence, pourraient être ajoutées.

Part III

Super-dispersive effects

Chapter 9

Context

The electromagnetic waves of pulsars need to travel into space to arrive on Earth. The interstellar space is not empty but is a low-density medium furthermore ionized. Consequently, since the discovery of the first pulsar, it was noticed that the passing of the electromagnetic wave traveling inside this plasma (the ISM) may undergo a diffusion.

This diffusion effect can be easily described with the first order of the Navier-Stokes and Euler equations, resulting in a dispersion law for the electromagnetic wave. The dispersion is a frequency-dependent relation, generating a difference in the arrival time between two frequencies, and appearing as a shift in frequency of the measured pulse (see the Dispersion section in Chapter 1.3.1 for more details).

This simple relation is, however, just a first-order approximation of the real dispersion, taking only the major effect produced by the plasma of the ISM into account. In consequence, some other effects, physical or geometrical, can occur and add an extra time delay to the pulse, creating therefore additional dispersion terms.

9.1 History

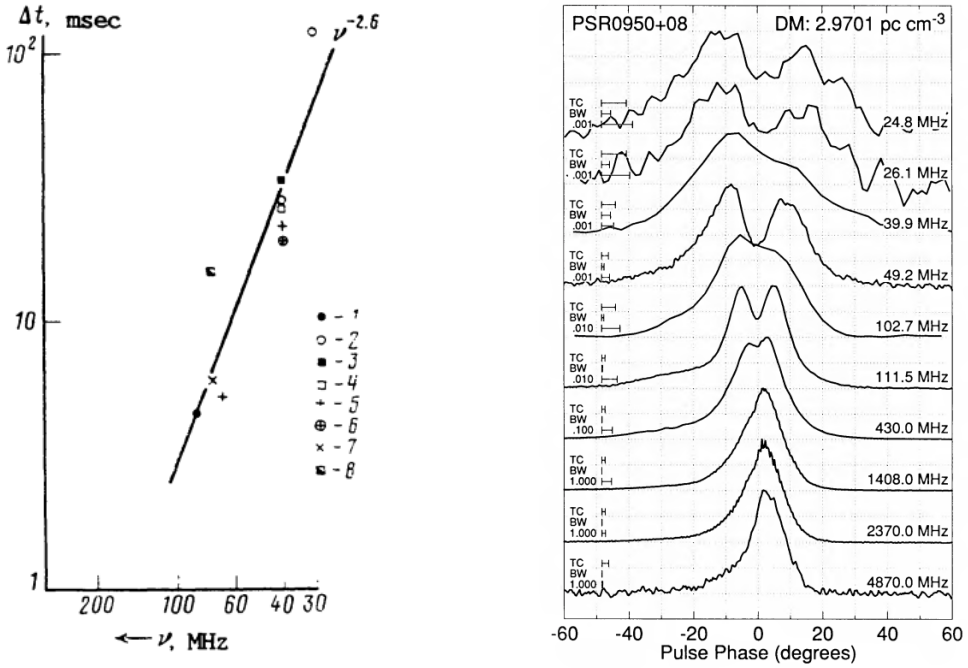
The first observation of a potential super dispersive effect was done in 1985 by Shitov & Malofeev, who compared the time delay of mean profiles of B0809+74 obtained at different frequencies of 30, 39, and 62 MHz with reference profiles at 102 and 400 MHz. An extra delay of 120 ms in the arrival times was measured and attributed to a geometric delay due to the twist of the magnetic field lines within the magnetosphere of the pulsar. In this theory, the magnetic field lines twist in the opposite sense relative to the spin of the neutron star. The emission cone is then deflected proportionally to the altitude. Also, according to the RFM model, the produced delay may be more important for the lower frequencies.

The next year, Kuzmin (1986) presented a study on the literature data from eight pulsars, showing also a frequency dependence of the DM for many of them. The extra delays measured between the data ranging from 30 MHz to 1.7 GHz were fitted using a law based on the twist magnetic field theory. In addition to the standard dispersion term in ν^{-2} (with ν the frequency), Kuzmin (1986) obtained a second term proportional to $\nu^{-2.6}$.

In 1988, Shitov et al. extended their first study to 20 pulsars. They studied these pulsars by comparison of the mean profiles between two or three frequencies from 30 and 102 MHz. On the set, 15 had a sufficiently accurate DM measurement, and 11 presented extra delay in agreement with the twisted magnetic field theory.

Hankins et al. (1991) raise the problem of the alignment of the profiles which defines the DM measurement. They observed two pulsars: B0950+08 and B1133+16, from 25 MHz to 5 GHz, and tried to determine the DM by the alignment of the average profiles in the first case, and of microstructure in the second case. They noticed that, on this very broad frequency range, the important frequency evolution of the shape of the profile makes it difficult the precise alignment below 100 MHz. Because of these variations, different positions of the fiducial point can be defined, leading to different measured DM. Finally, differences in the determined DM were observed between the two methods, and moreover, these differences depended on the used frequency interval.

Figure 9.1: Left: fit of the extra dispersive delays using a frequency-dependent power law in $\nu^{-2.6}$ (Kuzmin 1986). Right: average profiles at different frequencies of B0950+08 aligned for a DM of 2.9701 pc.cm^{-3} (Hankins et al. 1991).



In 2005, Ahuja et al. realized observations with the GMRT, where two frequencies were simultaneously observed. In the same way as Hankins et al. (1991), the DM was then measured by two methods: on the average profile, and on a time series of single pulses. For each observation, a single epoch DM was determined using a cross-correlation between the two frequencies. As a result, they have also shown DM differences relative to the method and the frequency bandwidth. In addition, they have seen a time variability of the DM. Two years after, in a follow-up paper, Ahuja et al. (2007) observed that, in the case of a correlation measurement, the shape and asymmetry of the profile modify the obtained DM. Finally, taking this effect into account, they measured frequency variations of the DM, which cannot be easily explained by twisting magnetic field or ISM refraction.

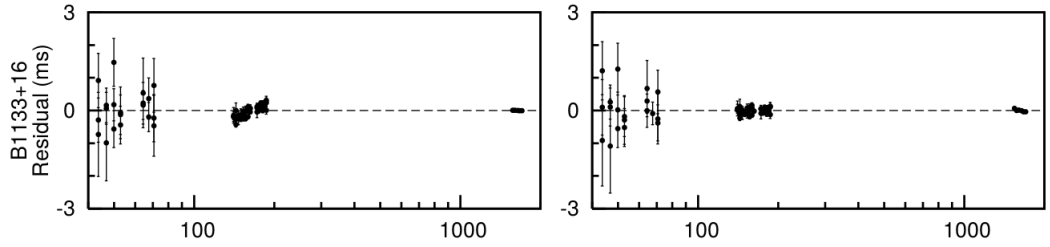
To precisely measure DMs, a standard method is to perform timing of the times of arrival (TOA) of pulses of the pulsar. The differences between the timing model and the TOA generate timing residuals which could be associated, among other things, with a DM variation between frequencies. Hassall et al. (2012) performed an improvement of the timing method by taking the profile variation into account. To follow the evolution of the shape of the profile, the timing was realized using a specific profile template for each frequency subband. These frequency-dependent templates allow a significant decrease in the timing residuals and upper limits in the timing residuals caused by higher-order dispersive terms have been determined.

The ISM is actually a turbulent medium (see Chapter 1.3.2), and Foster & Cordes (1990) examined in detail the effect of the electron density turbulences in the TOA of a simulation of B1937+21. They have shown the presence of refractive screens on the line of sight can produce different delays in the timing residuals, including a dispersion delay which participates to limit the precision of the pulsar timing.

These variations are then related to the scattering caused by the ISM. Also, using observation of B1737+13 over 36 weeks, Hemberger & Stinebring (2008) have shown a time variability of the timing residuals which is produced by the changes of the scattering over long timescales.

To improve the accuracy of timing models, Cordes & Shannon (2010) have added a dispersive perturbation term in their TOA model. This supplementary term is relative to the motion of the pulsar relative to the scattering screens, permitting it to take the associated variations observed by Hemberger & Stinebring (2008) into account. Furthermore, in another publication, Cordes et al. (2016) have shown

Figure 9.2: Comparison of the timing residuals obtained without (left panel) and with (right panel) frequency-dependent profile templates (Hassall et al. 2012).



that the multipath propagation yields by the scattering could result in differences in the timing residuals between two epochs or two frequencies. As a result, they can be interpreted as DM variations, and in the frequency case, appear as chromatic DM.can be interpreted as DM variations, and in the frequency case, appear as chromatic DM.

9.2 Major possible effects

Since almost the discovery of the pulsars, some reasons were invoked to explain an eventual deviation compared to the standard law of dispersion (Tanenbaum et al. 1968). The diverse causes can be either the ISM, the magnetosphere of the pulsar, or due to a problem with the measure of the DM.

9.2.1 Effects due to the ISM

Effects related to the scattering screens

The dispersion is due to the electromagnetic wave diffusion by the ISM. Also, this ISM is a turbulent medium with a spatial extension, generating, in addition, the scattering phenomena. Bhat et al. (2004) and (Cordes et al. 2016) have shown the scattering is frequency dependent, leading to potentially modifying the DM value. The DM variation is proportional to the *scattering measure* and doesn't follow exactly a law in ν^{-2} as the standard dispersion. Moreover, in the case of a variation of the electron density in the scattering screen, an effect of the refraction of the light from the different parts of the screen occurs. It results in additional time delays proportional to the gradient of the electron density.

The scattering measure and the gradient in electron density of the scattering screen are global values determined for the whole screen (or set of screens). The local variations are smoothed but can disturb the global DM measurement. time variations. The scattering shows long-time variations. As a result, in the case of measurements using large time scales, the smoothing effect produced by the utilization of global values becomes more significant.

Variations of the plasma properties

The turbulence in the scattering screens present in the ISM is a source of variations of the undergone dispersion. However, even the standard ISM without any turbulence is not a perfect cold plasma. To derive the standard law of dispersion ordinarily used, some properties of the plasma are neglected. A real plasma is not a cold plasma but features indeed a temperature giving thermal energy to the electrons of the plasma. The temperature adds a decreasing factor to the plasma frequency, generating a relative error in the DM proportional to the temperature T.

$$T > 0 \;\; ; \;\; \frac{\Delta DM}{DM} = -\frac{3}{2}\frac{k_B}{m_e c^2}\cdot T \approx -10^{-8}\cdot T \tag{9.1}$$

With k_B the Boltzmann constant, m_e the electron mass, and c the speed of light. For a typical warm ionized medium inside the Milky Way (such as HII regions), the temperature is of some thousands of K. Then, for a typical order of magnitude of 10^1 pc.cm^{-3}, the contribution (in absolute value) can be then greater than 10^{-4} pc.cm^{-3}. However, this contribution, which is already in the extreme case of a

hot region of the Milky Way, must also be weighted by the size of the warm area. The other ignored property concerns the presence of other types of elements that the electrons. This correction is also an adding factor to the plasma frequency which increases the dispersive delay. The relative error in DM is then proportional to the fraction of each element in the plasma:

$$\frac{\Delta DM}{DM} \;=\; \frac{m_e}{m_p} \cdot \sum_{Z>1} \frac{n_Z}{n_e} \cdot \frac{Z^2}{A} \;\approx\; 5.10^{-4} \cdot \sum_{Z>1} \frac{n_Z}{n_e} \cdot \frac{Z^2}{A_Z} \;\; \text{pc.cm}^{-3} \tag{9.2}$$

With Z the charge number, A_Z the corresponding mass number, m_p the proton mass, n_Z the density of the element, n_e the density of the electrons in the plasma. For a typical order of magnitude of DM 10^1 pc.cm^{-3}, the contribution of the helium, with a fraction of density n_Z / n_e of about $9.6.10^{-2}$, is thereby close to 10^{-3} pc.cm^{-3}.

These effects are additional terms of the standard law and are equally frequency-dependent in ν^{-2}, modifying then solely the slope coefficient of the law. The standard DM is defined based on an average of the electron density, which is considered relatively stable. These additional terms are related to some locations on the line of sight and are not necessarily constant. Moreover, it appears that these two effects have opposite signs and are consequently in competition. It could result in positive or negative DM variations, which are directly dependent on the timescales of the variations of the temperature and composition of the different parts of the line of sight. In the context of studies of the dispersion on large timescales, these effects can consequently eventually impact the result.

Lastly, these corrective terms can be added to the higher-order terms of a real plasma. The first one is due to the magnetic field of the plasma, resulting in a difference between the two states of polarization of the light. This effect is modeled by the rotation measure and a higher-order term in ν^{-3}. The other higher-order terms are due to the asymptotic development at the infinity used to define the standard law (cold plasma without magnetic field).

9.2.2 Propagation in the magnetosphere

Path difference due to the RFM

To try to explain eventual differences in the dispersion between frequencies, in addition to the effects produced by the ISM, some other reasons related to the magnetosphere of the pulsar were proposed. According to the RFM model, the different parts of the pulse, corresponding to different frequencies, come from different altitudes in the magnetosphere. Consequently, there is a path difference yielding a geometric time delay between the frequencies. Using the empirical values found by Kijak & Gil (2003), the difference of altitude, and then the time delay Δt_{path}, can be estimated between two frequencies $\nu_2 > \nu_1$:

$$\Delta t_{path}(\nu_1, \nu_2) \;=\; \frac{r_{em}(\nu_1) - r_{em}(\nu_2)}{c} \;=\; 4.10^5 \cdot \frac{P^{0.30} \cdot \dot{P}^{0.07}}{c} \cdot \left(\frac{1}{\nu_1^{0.26}} - \frac{1}{\nu_2^{0.26}} \right) \tag{9.3}$$

With P the period of the pulsar, and $\dot{P}$ its period derivative. For an extremely slow pulsar of 30 s of period and 10^{-13} s.s^{-1} of the period derivative, the time delay between a frequency $\nu_1 = 10$ MHz and a frequency $\nu_2 = 1$ GHz is about of 11.8 ms. According to the usual cold plasma dispersion relation, this extra time delay represents a DM variation of about 3.10^{-4} pc.cm^{-3}. For the frequency range of NenuFAR ($10 - 85$ MHz), the DM variation is decreased to reach approximately 2.10^{-4} pc.cm^{-3}.

The path difference is low, giving a small variation in DM, but also generates a difference in the magnetospheric dispersion undergone. Always according to the previous relation of the altitude of emission, the DM difference yielded by the magnetosphere of the pulsar can be estimated as:

$$\Delta DM(\nu_1, \nu_2) \;\approx\; n_{e,magn} \cdot c \cdot \Delta t_{path}(\nu_1, \nu_2) \tag{9.4}$$

With $n_{e,magn}$ the electron density of the pulsar magnetosphere. For an extreme case such as the slow pulsar took just before, the variation in DM is about $1.1.10^{-10} \cdot n_{e,magn}$ pc.cm^{-3} between 10 MHz and 1 GHz, and decrease to $7.0.10^{-11} \cdot n_{e,magn}$ pc.cm^{-3} for the frequency range of NenuFAR. The DM difference of this extra propagation is strongly dependent on the density of the magnetosphere of the

pulsar. For example, taking an order of magnitude of the density similar to the maximum density of the Earth ionosphere of about 10^6 cm^{-3}, the order of magnitude of the DM deviation for the NenuFAR frequency range is about 10^{-5} pc.cm^{-3}.

Also, because of the density of Earth ionosphere, we can notice that a DM variation is equally generated by it. However, ionosphere is very thin with a thickness of less than 1 000 km. As a consequence, the DM deviation is reduced compared to those of the pulsar magnetosphere. Nevertheless, in the case of strong solar storm, the associated solar wind can increase the density of electrons (and ions) in the Earth ionosphere, resulting therefore to an increase of the DM deviation during a certain timelapse. It is not excluded that a similar effect may occur in the pulsar magnetosphere due to an external wind of charged particules.

Contrary to the corrective terms, the advantage of these two effects related to the propagation in the pulsar magnetosphere is they are directly frequency dependent. However, the first part of the estimation of the order of magnitude is dependent of the correctness of the empirical coefficients of the RFM model, and the second part is strongly dependent on the unknown density of the pulsar magnetosphere. As a consequence, the DM variations caused by the difference of altitude emission could be too small to be detectable, even at the frequencies of NenuFAR.

Twisting of the magnetic field lines

The last magnetospheric effect was actually the first proposed by Shitov (1983): the twisting of the magnetic field lines. The radiative friction creates forces in the opposite direction of the rotation, resulting in a curvature of the magnetic field line. This twisting produces a retardation of the observed pulse (Shitov & Malofeev 1985) relative to the altitude of emission r.

$$t_{twist}(r) \;=\; \frac{P}{2\pi} \cdot \arctan\left(\left(\frac{r}{R_L} \right)^3 \cdot \sin^2(\delta\phi) \right) \tag{9.5}$$

Where $R_L \;=\; c \cdot P \;/\; 2\pi$ represents the radius of the light cylinder, and $\delta\phi$ is the angle between the rotation and the magnetic axis. According to the RFM, the altitude of emission is frequency-dependent and can be estimated with the same relation based on Kijak & Gil (2003) used before. Except for the pulsars with a period lower than 100 ms, the calculation shows the altitude of emission is small compared to the light cylinder, and the difference of extra time due to the twisting can be evaluated more simply with the following form:

$$\Delta t_{twist}(\nu_1, \nu_2) \;\approx\; 4\pi^2 \sin^2(\delta\phi) \cdot \left(\frac{r(\nu_1)^3 - r(\nu_2)^3}{c^3 \cdot P^2} \right) \tag{9.6}$$

As for the extra propagation only due to the RFM, the twisting of the magnetic field lines adds retardation because of the extra distance and an extra dispersion because of the extra distance in the plasma of the magnetosphere. At the NenuFAR frequencies, considering a typical pulsar of 1 s of period and 10^{-13} $s.s^{-1}$, and for the mean angle such as $\sin(\delta\phi) \;=\; 1/2\pi$, the retardation represents a DM deviation of about 4.10^{-9} pc.cm^{-3}, and the extra dispersion generated by the pulsar magnetosphere is about $2.10^{-15} \cdot n_{e,magn}$ pc.cm^{-3}.

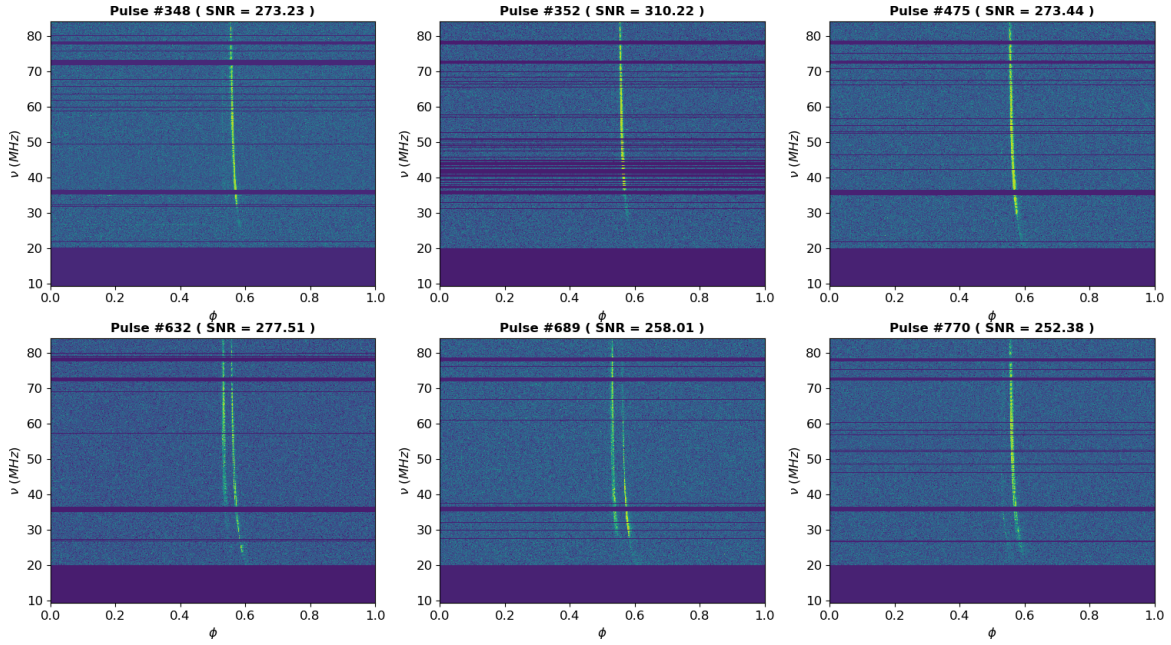
In consequence, the twisting of the magnetic field lines seems to be largely lower than the other effects. However, for the pulsar with a short period, the light cylinder is largely narrower. The curvature of the field lines is thus more important, leading to a stronger effect of the twisting. Nevertheless, this theory has been finally excluded to explain the observed DM variations.

9.2.3 The problem of the measure

Variations exterior to the pulse

These previous effects are intrinsic to the dispersion law and are then undergone by the pulse at each time and frequency. These are consequently independent of the chosen method to measure the DM of the pulsar. However, the final measure of the DM is a function of the dispersion law with these effects, and of the accuracy and biases of the measure in the observed pulse.

Figure 9.3: Single pulses of B1133+16 in a single observation of 30 minutes observed with NenuFAR. One can notice the time variability of the pulses within a unique observation is important either in the number of components or observable frequency bandwidth.



The first effect of modifying the DM presented above is due to the scattering. The presence of turbulence in the scattering screen (or the multiplicity of the screens) results in a global smoothed value. The local variations are thus erased, leading to an entanglement of the fine super dispersive effects with the scattering variations. In addition to the local variations of the scattering screen, Hemberger & Stinebring (2008) have shown the scattering is unconstant in time. These time variations are dependent on the line of sight and can vary enough to be detectable over some weeks (Hemberger & Stinebring 2008). As a consequence, in the case of a method using long timescales, the time variations of the scattering screens, whether local or global, can strongly impact the DM measurement.

Variations intrinsic to the pulse

The scattering and the other effects are exterior to the pulsar (caused by the plasma of the ISM or the magnetosphere) but also exist variations intrinsic to the pulse. Many pulsars show indeed important variations of their profile in frequency, permitting Cordes to elaborate the RFM. Also, Hankins et al. (1991) have shown the frequency variation of the profile makes difficult the choice of the position of the fiducial point. This problem was partially solved by Hassall et al. (2012) by the use of templates calculated on frequency subbands rather than in the total bandwidth. Although this method works well, it remains light frequency variations inside the subband which could continue to disturb a fine search of DM deviations. The resulting error in DM is especially important in the context of pulsars with an asymmetric profile (Ahuja et al. 2007), which are more difficult to precisely model.

In addition to the frequency variations of the pulse, there is also a time variability of the pulse. Figure 9.3 shows the frequency-phase plan for the six most intense pulses in a unique observation of B1133+16, observed during 30 minutes with NenuFAR in single pulse mode. One can notice the six pulses are different, sometimes presenting only one peak, sometimes presenting two peaks, and with different ratios of intensity between the two peaks. The consequence is that the basic assumption that pulses are constant in time is not correct at short time scales. A DM measurement performed with a unique pulse template on different pulses simultaneously results to lose fine information about the differences between the diverse pulses. It results from these local time variations the generation of errors in the measured DM.

Method of measure

All the effects mentioned are physical effects occurring before the measure. However, the inaccuracy of the obtained DM value can also be impacted just by the method carried out to measure this DM value.

The way to measure the DM is related to the problem noticed by Hankins et al. (1991) of the alignment, i.e. how to define the fiducial point allowing to determine the dispersion? Thereby, the comparison of the resulting DM between methods based on the average profile or the microstructure (Hankins et al. 1991) or with a method based on the correlation (Ahuja et al. 2005) have revealed differences. Also, the methods based on the definition of a template are facing the problem to obtain a template sufficiently precise, despite the numerous variations presented in this section.

Finally, the variations occur at various scales, in time or frequency. The timescales and the size of the frequency bandwidth used to determine the DM have, therefore, equally an impact on the obtained DM value. A method using large scales will generate an averaging effect, with the consequence to hide the fine effects.

9.3 Method of DM search used for this study

In the context of this work, I have tried taking the different effects observed in the diverse past studies into account. The research axis and the methods used to measure the DM have thus been chosen to be different that those done in the previous studies.

9.3.1 Timescales

An important source of error is given by the time variability of many effects occurring in the DM measurement. The used method must then try to avoid variations in time. However, the timescales are various followings the effect, which may be potentially short. As a consequence, it needs to carry out the measure of the DM on the shortest able timescale.

Moreover, the longer the timescale the smoother the fine dispersion effects. Also, when these fine effects occur in an order of magnitude much smaller than the mean standard dispersion effects, the averaging is harmful to distinguish DM deviations. Reducing the timescale as much as possible has thereby the other advantage to avoid any averaging effect.

As a consequence, the best possible timescale is the time-lapse of a single pulse. It allows obtaining in a way an "instantaneous DM", corresponding to the finest able measure of the DM. Although measuring a DM in a single pulse involves substantial difficulty due to the lack of flux, such a measurement is notably different from other studies about dispersion and related subjects. The typical way is to use the timing process with data folded over time sub-integrations precisely to obtain sufficient flux and accuracy.

One can notice that measurements in directly in single pulses don't completely remove the time variability. Although the time variability inside the width of the pulse, i.e. within the timelapse of the radio emission, is certainly too short to have a detectable variability of the DM, the pulse arrives in Earth at different times due to the dispersion. It still remains a time variability of the measured pulse due to scintillation and scattering variability. Also, another time variability is the Earth's rotation. However, although the delay time is substantial at the NenuFAR frequencies, the variation in the arrival time of the wave during an observation is extremely small. Hence, for an observation of one hour, the maximum DM variation for a pulsar such as B1508+55 with a DM of about 19.602 pc.cm^{-3} can be typically estimated to approximately 1×10^{-5} pc.cm^{-3}. As a consequence, compared to the other expected effects, the barycentric corrections of the single pulse can be neglected at the first step.

9.3.2 Frequency bandwidth

A large frequency bandwidth allows us to better distinguish the frequency-dependent DM deviations relative to the standard law of dispersion. However, as for the timescale, the larger the bandwidth the larger the frequency variations of the profile of the pulsar (see Chapter 1.2). In addition, to obtain a large frequency bandwidth, it is useful to use multi-telescope observations. However, if the measure is performed in single pulses, it raises the problem of synchronization, i.e. of the correct alignment, of the chosen pulse between the different telescopes.

The dispersion scaling is in ν^{-2} where ν represents the frequency of observation, leading to the lower the frequency the higher the dispersion delay of the pulse. It results in a better precision of the DM measurement at low frequencies compared to higher frequencies. In the case of NenuFAR, we are already at a very low frequency, close to the ionospheric cutoff at about $10 - 15$ MHz. Moreover, the able frequencies range from 10 to 85 MHz, allowing it to cover about 3 octaves. To obtain this at higher frequencies, that represent, for example, a range between 300 MHz and 2.4 GHz, requiring most of the time several telescopes.

NenuFAR alone allows for covering a large frequency span while avoiding introducing errors due to the phasing of different telescopes. As a consequence, the DM measurements have been done solely on NenuFAR observations, with a subdivision in frequency subbands.

9.3.3 Type of measurement

After the observing parameters, the measurement method of the DM must reduce as much as possible the other errors. The method will be applied on single pulses to avoid any averaging. Thus, it is required that the used method equally reduces at the minimum the averaging. Moreover, to keep sufficient precision, it is also required to be the least dependent on the shape of the pulse, either in frequency or in time.

To reduce the error which could erase the searched fine effects, it is crucial to minimize as much as possible the introduction of errors. As a result, the number of intermediate operations has to be reduced to the minimum required to obtain a DM value. The employed method ideally needs to be self-consistency.

Finally, in order to prevent a measure dependent on the observer, the used method must be based on a criterion independent of any choice of fiducial point. To avoid the thorny problem of the fiducial point noticed by some studies, the measurement has to be done relative to a criterion that is as much as possible objective, and which doesn't take any input parameter (no human choices!).

9.3.4 Measurement method used in this study

To conclude, all the measurements of DM done for this study have been realized on frequency subbands of single pulses from a unique observation made with NenuFAR. In order to obtain enough precision on the DM, one needs to select only intense pulsars, showing a sufficient flux to provide single pulses easily detectable.

The measurement is performed using a method without averaging, and without definition of a fiducial point to determine the dispersion of each part of the pulse. Initially, the first method developed was based on the Radon transform, which is a mathematical transform allowing one to determine the level of inclination of a structure relative to a fiducial direction. In the case of the dispersion of a pulse, the fiducial direction is the frequency axis, and the level of inclination decreases towards zero when the dispersion decreases. However, this method is purely geometric, and the decrease in the number of channels, resulting from the subdivision in subbands, yields an increase in the error in the determination of the angle of inclination. Although returning good results for an entire single pulse, it thus appeared that the uncertainty on the DM for a single subband was clearly insufficient because of the lack of frequency and time resolution.

The method based on the Radon transform has ultimately been dropped, and another method was developed. This second method is based on the study of the dispersion of the pulse in the Fourier space. With this one, the dispersion can be determined using an objective mathematical criterion. Moreover, the DM values for the subbands obtain a better uncertainty compared to the method based on the Radon transform. From this method, I have developed a DM search algorithm presented in the following chapter.

Résumé du chapitre :

L'impulsion radio du pulsar traverse le milieu interstellaire qui est un milieu ionisé de faible densité. Et depuis la découverte du premier pulsar, le phénomène de diffusion de l'onde électromagnétique induit par la traversée de ce plasma est connu. Cette diffusion peut être facilement décrite par les équations de Navier-Stokes et Euler au premier ordre pour obtenir une relation de dispersion de l'onde. Cette relation décrit l'apparition d'un délai temporel entre deux fréquences. Cependant, cette relation est déterminée dans le cadre simple du premier ordre d'un plasma froid homogène dans le milieu interstellaire. Ainsi, elle ne prend pas en compte certains effets physiques ou géométriques pouvant ajouter des délais supplémentaires à cette dispersion usuelle.

En 1985, (Shitov & Malofeev 1985) ont mesuré un délai supplémentaire de 120 ms entre plusieurs profils moyens du pulsar B0950+08 obtenus à différentes fréquences. Ils ont attribué ce délai supplémentaire à une torsion des lignes de champ magnétique induisant ainsi, dû au RFM, un délai de propagation entre les fréquences. Kuzmin (1986) présente une étude similaire étendue à plusieurs pulsars, et ajuste une relation en $\nu^{-2,6}$, où ν est la fréquence d'observation, pour cette "super-dispersion". Shitov et al. (1988) étendent l'étude à 20 pulsars, et avancent que onze de ces pulsars semblent montrer des résultats en accord avec la théorie de la torsion des lignes de champ magnétique.

En comparant des mesures de dispersion obtenues sur des profils intégrés et des microstructures, Hankins et al. (1991) trouvent des résultats différents, et identifient ainsi le problème du choix du point de référence pour la mesure du délai. Ils montrent que ce problème est d'autant plus important du fait de l'évolution en fréquence de la forme du profil des pulsars. Ahuja et al. (2005, 2007) réalisent une étude similaire en mesurant la dispersion avec deux méthodes différentes. Ils obtiennent également des résultats différents suivant la méthode ou les fréquences regardées, et aboutissent à des conclusions similaires. Finalement, Hassall et al. (2012) résout une partie du problème en développant une méthode de timing basé sur des modèles de profils dépendant de la fréquence pour la mesure de la dispersion. En utilisant ces modèles, il contraint de manière importante les limites supérieures des potentiels effets super-dispersifs.

En sus du problème de la mesure, Foster & Cordes (1990) ont montré que la présence d'écrans réfractifs sur la ligne de visée entraînait des délais supplémentaires limitant la précision du timing. De plus, Hemberger & Stinebring (2008) ont également montré que ces écrans produisant notamment le phénomène de scattering varie dans le temps, induisant de ce fait des variations temporelles de ces délais. Afin de limiter l'impact de ces écrans, Cordes & Shannon (2010) utilise des termes perturbatifs au sein de la procédure de timing mesurant la dispersion. Enfin, Cordes et al. (2016) montre que la propagation multi-chemins causée par les écrans réfractifs peut provoquer des différences pouvant être attribuées à des variations chromatiques de la mesure de dispersion.

Au cours des années, plusieurs effets ont été proposés pour expliquer d'éventuels effets super-dispersifs. Il y a tout d'abord des effets physiques dus au milieu interstellaire, le premier étant bien évidemment le phénomène de scattering dont l'impact en terme de mesure de dispersion est évalué de manière globale, générant ainsi des variations fines. Le second est la température du milieu traversé par l'onde, pouvant créer des variations de l'ordre de 10^{-4} pc.cm^{-3}, mais qui doivent être pondérées par la taille de la zone traversée. Enfin, le dernier est la présence d'ions au sein du plasma, qui pourrait éventuellement générer des variations également de l'ordre de 10^{-4} pc.cm^{-3}au maximum. Cependant, il est à noter que ces termes ont la même dépendance fréquentielle que la loi usuelle, et que de plus, ils sont localisés sur la ligne de visée et peuvent se compenser.

Les seconds effets possibles sont d'ordre géométrique. En se basant sur le RFM, il doit exister un délai de propagation entre les différentes fréquences de l'impulsion du pulsar. Entre les fréquences minimale et maximale de NenuFAR, ce délai peut être estimé à 2×10^{-4} pc.cm^{-3}. En outre, cette distance supplémentaire accroît également la dispersion subie par les plus hautes fréquences d'environ 7×10^{-11} multipliés par la densité électronique à l'intérieur de la magnétosphère du pulsar. L'autre terme géométrique est donné par la théorie initiale de la torsion des lignes de champ magnétique. L'accroissement de la distance à parcourir entre les différentes fréquences est cependant encore plus faible, avec une différence de chemin entraînant une variation de 4×10^{-9} pc.cm^{-3}et un ajout de dispersion de 2×10^{-15} multipliés par la densité électronique de la magnétosphère du pulsar.

En plus de ces effets physiques et géométriques, il y a également des problèmes liés à la mesure elle-même. Ainsi, les variations temporelles du scattering peuvent entraîner des différences dans les mesures finales suivant l'intervalle de temps utilisé. Les variations intrinsèques au pulsar, tel que la variation fréquentielle du profil, l'asymétrie de celui-ci, ou encore les différences de forme d'une impulsion à l'autre, peuvent également entraîner des différences dans la mesure. Et enfin, bien évidemment, la méthode utilisée peut modifier de manière importante le résultat d'une mesure.

Dans le contexte de l'étude présentée dans cette partie de la thèse, la méthode a été définie de manière à tenter d'éviter les différents écueils observés dans les précédentes études. Notamment, une importante source d'erreurs est donnée par les différentes variations temporelles. La présente étude se porte donc sur des mesures de dispersion "instantanées" réalisées sur des impulsions individuelles.

De plus, la détermination des délais supplémentaires est facilitée par l'utilisation de larges bandes de fréquences. Cependant, l'utilisation de plusieurs télescopes pose le problème de la synchronisation particulièrement important dans le cadre d'une étude sur des impulsions individuelles. En conséquence, l'étude n'utilise que des observations de NenuFAR, qui permet tout de même d'obtenir une bande de fréquence relative importante d'environ trois octaves.

Aussi, la méthode de mesure doit éviter autant que possible les effets de moyennage, et être le moins dépendant possible de la forme de l'impulsion, que cela soit en temps ou en fréquence. De plus, afin d'éviter d'introduire toute erreur pouvant noyer les effets recherchés, il est important de réduire au maximum le nombre d'intermédiaires, en essayant de réaliser une méthode aussi auto-cohérente que possible. Enfin, le point le plus important est d'appliquer une méthode basée sur un critère le plus objectif possible, afin de se soustraire à l'épineux problème du choix du point de référence pour la mesure effectuée.

La méthode présentée dans cette partie est ainsi donc basée sur une mesure sur des sous-bandes de fréquences d'une impulsion individuelle de pulsars puissants, en se basant sur un critère mathématique identifiant la dispersion dans l'espace de Fourier.

Chapter 10

Models and methods

10.1 Dynamic spectrum model

The dynamic spectrum $S(t,f)$ of a NenuFAR observation can be modeled by a series of five terms: the background white noise $B(t_0,f_0)$, the jumps of the analogical beam $J(t)_{f_0}$, the broadband RFIs $R_b(f)_{t_0}$), the narrow band RFIs $R_n(t)_{f_0}$), and the signal of the pulsar $P(t,f)$.

$$S(t,f) \;=\; B_{t_0 f_0} + J(t)_{f_0} + R_b(f)_{t_0} + R_n(t)_{f_0} + P(t,f) \tag{10.1}$$

Using the linearity of the Fourier transform (FT hereafter), the dynamic spectrum in the Fourier domain $\hat{S}(\omega,\nu)$ is then the sum of the Fourier transforms of each term.

$$\hat{S}(\omega,\nu) \;=\; \hat{B}_{t_0 f_0} + \hat{J}(\omega)_{f_0} + \hat{R}_b(\nu)_{t_0} + \hat{R}_n(\omega)_{f_0} + \hat{P}(\omega,\nu) \tag{10.2}$$

The FTs used are realized from the space domain defined by the time t and the frequency f to the Fourier space defined by the Fourier frequencies dual of the time ω and the Fourier frequencies dual from the frequencies ν. The FTs carried out for this work use an orthogonal normalization following the relation:

$$\begin{aligned} \mathcal{FT}:\; S(t,f) \;&\rightarrow\; \hat{S}(\omega,\nu) \\ t,f \;\in\; \mathbb{R}^2 \;&\rightarrow\; \omega,\nu \;\in\; \mathbb{C}^2 \\ \hat{S}(\omega,\nu) \;&=\; \int dt \int df \cdot S(t,f) \cdot e^{-i(\omega t + \nu f)} \end{aligned} \tag{10.3}$$

$\hat{S}(\omega,\nu)$ being a sum of five terms, it is possible to treat each term independently of the others. To each one, a model is defined, and the FT of this model can be directly calculated.

10.1.1 Model of noise

The white noise can be described as a function independent of the time and frequency. At each point (t_0,f_0), it can be defined with a value provided by a normal distribution $\mathcal{N}$ of mean $\langle S\rangle$ and standard deviation $\sigma(S)$:

$$B_{t_0 f_0} \;=\; \mathcal{N}(\;\langle S\rangle, \sigma(S)\;) \tag{10.4}$$

The FT of the background noise is then easily determined by the following equation:

$$\hat{B}_{t_0 f_0} = \mathcal{N}(\;\langle I\rangle, \sigma(I)\;) \cdot \int dt \cdot e^{-i\omega t} \cdot \int df \cdot e^{-i\nu f} \quad = \mathcal{N}(\;\langle I\rangle, \sigma(I)\;) \cdot 2\pi\delta(\omega) \cdot 2\pi\delta(\nu) \tag{10.5}$$

We obtain a Fourier signal totally located in the center of the 2D FT of the dynamic spectrum at $\omega = 0$ and $\nu = 0$. The FT of the dynamic spectrum can be easily normalized by setting the central point to 0.

10.1.2 Model of beam tracking adjustment

The second term of the dynamic spectrum is tracking adjustments of the analog beam of NenuFAR. The tracking adjustments are not really another type of signal, but are actually an intrinsic variation of the time series which can be modeled by a logarithmic function relative to the time:

$$J(t)_{f_0} \;=\; A_j \cdot \log(\; B_j \cdot t \;) \tag{10.6}$$

Where A_j represents the amplitude of the tracking adjustment, and B_j is the increasing coefficient of it. In an observation longer than six minutes, it is indeed a sum of tracking adjustments.

$$\hat{J}(\omega)_{f_0} \;=\; A_j \cdot \int dt \cdot \log(\; B_j \cdot t \;) \cdot e^{-i\omega t} \cdot \int df \cdot e^{-i\nu f} \;=\; \frac{A_j}{B_j} \cdot \left(-\frac{\pi}{|\omega|} - 2\pi \cdot \gamma \cdot \delta(\omega) \right) \cdot 2\pi\delta(\nu) \tag{10.7}$$

With $\gamma \approx 0.5772156649$ the Euler-Mascheroni constant, representing the limit of the difference between the harmonic series and the logarithm. The obtained FT is an inverse structure in ω located in a line in $\nu = 0$ with a peak in $\omega = 0$.

10.1.3 Model of RFI

The two following terms refer to the RFIs present in the dynamic spectrum. The RFIs are divided into two classes: the broadband RFIs in frequency and those which are narrowband. Basically, a broadband RFI is located at some precise times. In the other case, the narrowband RFIs are precisely located in frequency but have a certain time extension.

The two types are then described with the same function, where we just inverse the frequency and time dependencies. As a first approximation, a broadband RFI can be described as a boxcar function in frequency located in a specific time t_0 :

$$R_b(f)_{t_0} \;=\; A_b \cdot \Pi\left(\frac{f - f_c}{\Delta f} \right) \cdot \delta(t - t_0) \tag{10.8}$$

With A_b the amplitude of the RFI, Δf its bandwidth and f_c the central frequency of this bandwidth.

$$\hat{R}_b(\nu)_{t_0} \;=\; A_b \cdot \int dt \cdot \delta(t - t_0) \cdot e^{-i\omega t} \cdot \int df \cdot \Pi\left(\frac{f - f_c}{\Delta f} \right) \cdot e^{-i\nu f} \tag{10.9}$$

$$\Leftrightarrow\; \hat{R}_b(\nu)_{t_0} \;=\; A_b \cdot e^{-it_0\omega} \cdot e^{-if_c\nu} \cdot 2\Delta f \cdot \mathrm{sinc}\left(\frac{\Delta f \cdot \nu}{2\pi} \right) \tag{10.10}$$

In the same way, the narrowband RFI can be described as a boxcar function in time located in a specific frequency f_0 :

$$R_n(t)_{f_0} \;=\; A_n \cdot \Pi\left(\frac{t - t_c}{\Delta t} \right) \cdot \delta(f - f_0) \tag{10.11}$$

With A_n the amplitude of the RFI, Δt the time-lapse around the time t_c where the RFI is seen. The corresponding FT is consequently of a similar form to that of a broadband RFI:

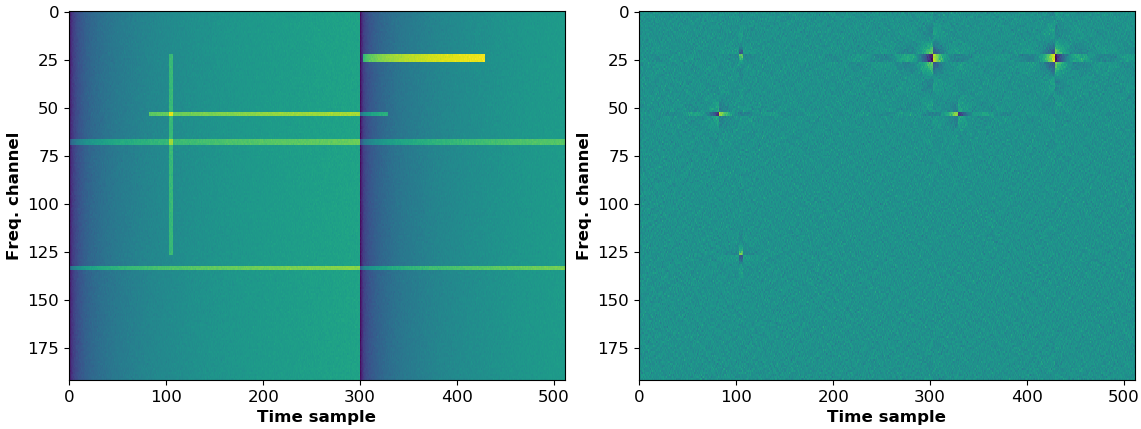
$$\hat{R}_n(t)_{f_0} \;=\; A_n \cdot \int df \cdot \delta(f - f_0) \cdot e^{-i\nu f} \cdot \int dt \cdot \Pi\left(\frac{t - t_c}{\Delta t} \right) \cdot e^{-i\omega t} \tag{10.12}$$

$$\Leftrightarrow\; \hat{R}_n(t)_{f_0} \;=\; A_n \cdot e^{-if_0\nu} \cdot e^{-it_c\nu} \cdot 2\Delta t \cdot \mathrm{sinc}\left(\frac{\Delta t \cdot \omega}{2\pi} \right) \tag{10.13}$$

Obviously, a real dynamic spectrum is a sum of different RFIs, and furthermore, a real RFI is not necessarily finite in frequency or time. A more precise and generic model for the RFIs is a sum of functions, where each one is a product of two boxcars: one for the frequency extension and one for the time extension. The general FT of the N_{RFI} RFIs present in the dynamic spectrum can ultimately be expressed as:

$$\hat{R}(\omega, \nu) \;=\; \sum_{k}^{N_{RFI}} A_k \cdot e^{-i\omega t_{0,k}} \cdot 2\Delta t_k \cdot \mathrm{sinc}\left(\frac{\Delta t_k \cdot \omega}{2\pi} \right) \cdot e^{-i\nu f_{0,k}} \cdot 2\Delta f_k \cdot \mathrm{sinc}\left(\frac{\Delta f_k \cdot \nu}{2\pi} \right) \tag{10.14}$$

Figure 10.1: Left: Simulation of a dynamic spectrum with five different RFIs patterns, and two tracking adjustments starting in the time samples 0 and 300. Right: reconstructed dynamic spectrum after applying a cross mask in the 2D Fourier transform of the dynamic spectrum.



Following this simple model, where the frequency part and the time part are totally independent, it is noticed that the FT of an RFI is located for the two directions within an envelope given by a cardinal sine. The power of the RFI is therefore concentrated in a cross around the directions $\omega = 0$ and $\nu = 0$, where the width of the cross depends on the width of the RFI in the time-frequency domain.

Using a cross mask in the Fourier domain could therefore remove a significant part of the RFIs. However, the real case gives an Airy function rather than a product of two cardinal sines. Hence, even with a large cross, it remains part of the power of the RFIs, leading to keeping in the time-frequency domain the edges of the RFIs. Figure 10.1 shows a simulated dynamic spectrum, where five RFIs of various types are placed in some locations: broadband, narrowband, finite, and infinite. Two simulated tracking adjustments are also placed, a first starting at the time sample 0, and a second at the time sample 300. The left panel represents the initial dynamic spectrum, and the right panel represents the reconstructed dynamic spectrum using a cross mask. We can notice the problem of the remaining edges of the finite RFI, and at the opposite the total disappearance of the infinite RFI and tracking adjustments.

10.1.4 Model of the pulsar pulse

The last component of the dynamic spectrum and the most interesting one is the pulse of the pulsar. The pulse can be easily described by a main function $D(t, f)$ modulated by a second function $C(f)$, representing the spectral distribution of the pulse. This second term takes the variation of the measured flux due to the intrinsic spectrum of the pulsar and the telescope bandpass into account. It also takes the scintillation into account, which is not necessarily constant in frequency.

$$P(t, f) = A_p \cdot C(f) \cdot D(t, f) \tag{10.15}$$

With A_p the amplitude of the pulse. The main term $D(t, f)$ represents the time distribution of the pulse, which is frequency dependent due to the dispersion. The distribution of the flux of a pulse in time can be modeled, at first approximation, with a Gaussian distribution for each frequency f around the time $t_r(f)$:

$$D(t, f) = \exp\left(-\frac{(\ t - t_r(f)\ )^2}{2\sigma_t^2}\right) \tag{10.16}$$

Where $t_r(f)$ corresponds to the time position of the maximum flux of the pulse at the frequency f. This time position is given by the dispersion law of the pulsars: $t_r(f) = \mathcal{D} \cdot DM \cdot f^{-2} - t_0$, with t_0 the reference time position of the pulse at the highest frequency.

The parameter σ_t represents the width of the pulse. For a real pulse, this parameter must normally take the intrinsic frequency variation of the width of the pulse into account, and equally the effects of the scattering. However, in terms of a DM search, the width variation in frequency doesn't impact the found DM value. The essential element is indeed only the variation of the pulse asymmetry in frequency. Therefore, the frequency dependence is important to take into account only in the context of a strong scattering modifying significantly the shape (more particularly the symmetry) of the pulse. In consequence, at first approximation, we can consider the power of the pulse is distributed following a constant Gaussian. This simpler model allows us to neglect the frequency dependence in order to have easier computing of the FT of the model.

The modulation function $C(f)$ being not time-dependent, the FT of the main function $D(t, f)$ can be calculated as a unidimensional FT in time, encapsulated in the FT in frequency. The Fourier transform of the model of the pulse is then given by the following relation:

$$\hat{P}(\omega, \nu) = A_p \cdot \int C(f) \cdot \left(\int dt \cdot D(t, f) \cdot e^{-i\omega t} \right) \cdot e^{-i\nu f} \cdot df \tag{10.17}$$

$$= A_p \cdot \int C(f) \cdot \left(e^{-it_r(f)} \cdot \sigma_t \cdot e^{-\frac{1}{2}\sigma_t^2\omega^2} \right) \cdot e^{-i\nu f} \cdot df \tag{10.18}$$

The reference time $t_r(f)$ at the frequency f is indeed the sum of the reference time at the highest frequency t_0 with a delay $\delta t(f)$: $t_r(f) = t_0 + \delta t(f)$, where $\delta t(f)$ corresponds to the frequency-dependent delay created by the dispersion. The final form of the FT of the pulse is then:

$$\hat{P}(\omega, \nu) = A_p \cdot \sigma_t \cdot e^{-\frac{1}{2}\sigma_t^2\omega^2} \cdot e^{-it_0\omega} \cdot \hat{C}(\omega, \nu) \tag{10.19}$$

Where $\hat{C}(\omega, \nu)$ represents the function that entirely contains the information about the frequency distribution of the pulse and the dispersion of the pulse in frequency:

$$\hat{C}(\omega, \nu) = \int df \cdot C(f) \cdot e^{-i(\ \delta t(f)\omega + \nu f\ )} \tag{10.20}$$

In the case of an ideally dedispersed pulse, i.e. $\delta t(f) = 0\ \forall\ f$, this function is then simpler and corresponds directly to the FT of the function $C(f)$. The dedispersed function is then a product of two terms: a first term dependent of ω, containing the time structure of the pulse; a second term $\hat{C}^d(\nu)$ solely dependent of ν containing the frequency structure of the pulse.

10.2 Dispersion in the Fourier space

In the initial dynamic spectrum, i.e. before dedispersion, in the five terms of S, the only term with an FT dependent on the DM is the term P describing the pulse. However, if we perform dedispersion, the FT of the tracking adjustments and broadband RFIs will be also impacted, modifying the global structure of the FT of the dynamic spectrum.

Thereby, in the case of analysis at DM = 0 pc.cm^{-3}, the impact of the dispersion on the dynamic spectrum in the Fourier space can be studied only with the FT of the pulse.

$$\frac{\partial \hat{S}(DM)}{\partial DM} = \frac{\partial}{\partial DM}(\ \hat{B}_{t_0 f_0} + \hat{J}_{f_0} + \hat{R}_{b,t_0} + \hat{R}_{n,f_0} + \hat{P}(DM)\ ) = \frac{\partial \hat{P}(DM)}{\partial DM} \tag{10.21}$$

Furthermore, in the context of analysis for a DM different from 0, the tracking adjustments and RFIs have normally been removed (at least for a substantial part). We can then consider the majority of the power in the dynamic spectrum can be imputed to the pulse (and white noise for the background). As a consequence, in the following section, the dispersion in the Fourier space is studied only relative to the pulse.

10.2.1 Variations from the DM

The FT of the pulse expressed by the equation 10.19 indicates that in the Fourier space, the structure in ω is a Gaussian shape centered in $\omega = 0$, modulated by a sine function. This global shape is next modulated in the direction of ν by the shape of $\hat{C}$. As a result, the modulus of the FT of the pulse $hatP(\omega, \nu)$ is therefore a Gaussian in ω modulated in the direction ν by the modulus of the function $\hat{C}(\omega, \nu)$.

$$\|\hat{P}(\omega,\nu)\|^2 = A_p^2 \cdot \sigma_t^2 \cdot e^{-\sigma_t^2 \omega^2} \cdot \|\hat{C}(\omega,\nu)\|^2 \tag{10.22}$$

In the ideal case of a finite pulse without frequency structure, if the pulse is ideally dedispersed, the Gaussian in ω is contained in an envelope function whose shape is the square of a cardinal sine relative to ν. Also, from that there is a dispersion time delay $\delta t(f)$ different from 0, this envelope function is deformed in the ν direction and obtains, in addition, a certain structure in the ω direction.

In order to study the behavior relative to the DM, I have modeled a pulse for three different qualitative dispersions of the pulse: a high dispersion corresponding to dedispersion for a DM very far from the ideal DM, a medium dispersion, and a low dispersion close to the ideal alignment of the pulse. For each one, the 2D FT was computed, and the pattern produced by the FT of the pulse has been analyzed.

The function $\hat{C}(\omega, \nu)$ is actually the Fourier transform of the product $C(f) \cdot e^{-i\delta t(f)\omega}$, which is equivalent to a shift of the FT of $C(f)$:

$$\hat{C}(\omega,\nu) = \mathcal{FT}\{\, C(f) \cdot e^{-i\delta t(f)\omega} \,\} = \hat{C}(\, \nu - \nu_0(\omega) \,) \tag{10.23}$$

Where $\nu_0(\omega)$ is a reference frequency dependent on the dispersion. The global structure might therefore follow an oblique antisymmetric line relative to the axis $\omega = 0$ for a bad DM, and symmetric around 0 for the ideal DM. The norms of the FT of the dynamic spectrum for the three dispersions are presented in the different panels of Figure 10.2. For all, the global pattern in ω has a Gaussian shape and is inclined proportionally to the dispersion, which is exactly the expected behavior.

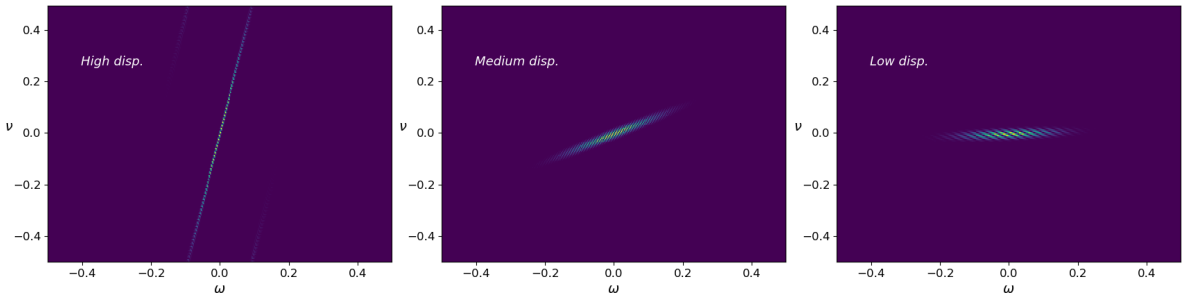


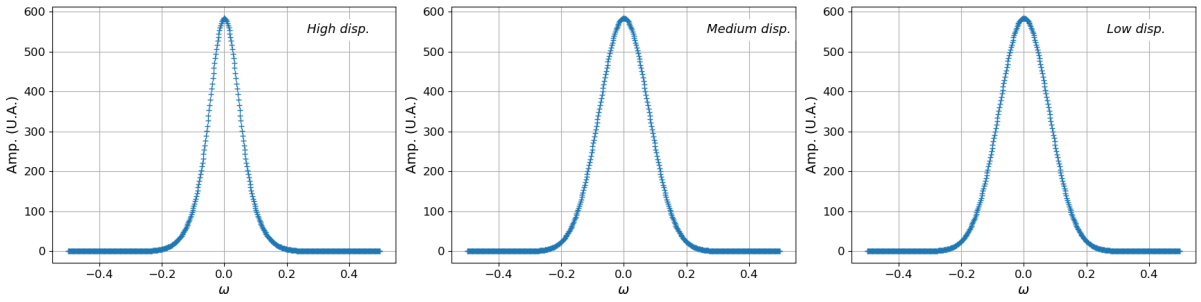
Figure 10.2: Fourier transform of the dispersed simulated dynamic spectra, where the central pattern (elongated structure in yellow and green) is the Fourier transform of the pulse. Left: for high dispersion of the pulse. Middle: medium dispersion. Right: low dispersion.

The pattern of the FT of the pulse can be examined in each direction, in order to identify a particular behavior dependent on the dispersion of the pulse. Hence, I have thereby marginalized the FT of the pulse relative to one direction to look at the distribution of the Fourier power in the other direction.

In the first case, the marginalization is made in the direction ν, to obtain the structure relative to ω. According to the equation 10.23, the function of modulation $\hat{C}(\omega, \nu)$ needs to be shifted in the direction ν relative to the DM. In the other direction ω, the global structure of the marginalized FT of the pulse must therefore not change relative to the dispersion.

Figure 10.3 shows the FT of the pulse marginalized in ν for the three different dispersions: from a very dispersed pulse in the left panel to a weak dispersed pulse in the right panel. The structure of the FT of the pulse in the direction ω is almost the same for the three dispersions. The amplitude and width are

Figure 10.3: Marginalization in the direction ν of the FT of the dispersed simulated dynamic spectra, showing the distribution of the Fourier power in the direction ω. Left: for high dispersion of the pulse. Middle: medium dispersion. Right: low dispersion.



constant, but maybe the width is a little bit larger in the weakly dispersed. This behavior is consistent with the equation 10.23, where the dispersion might exclusively create a shift in the direction ν without any changing in ω.

Contrary to the marginalization in ν, the structure of the FT of the pulse in the other direction, i.e. after marginalization in ω, shows an evident change relative to the dispersion of the pulse. Figure 10.4 presents the three diagrams of these marginalized FTs of the pulse in the direction ν. The first substantial and distinct difference is the width of the structure, considerably larger in the strong dispersed case than for the two other cases. The second property is the variation of the maximum amplitude, which hugely increases in the weakly dispersed case.

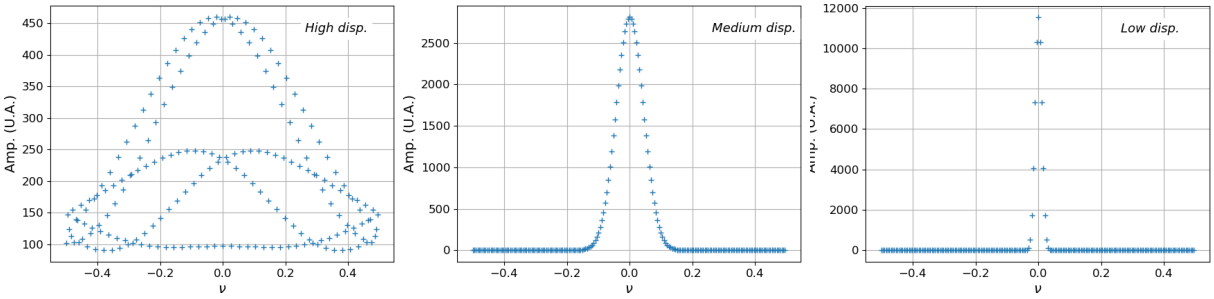


Figure 10.4: Marginalization in the direction ω of the FT of the dispersed simulated dynamic spectra, showing the distribution of the Fourier power in the direction ν. Left: for high dispersion of the pulse. Middle: medium dispersion. Right: low dispersion.

In consequence, there are no variations in ω relative to the dispersion. Furthermore, the lower the dispersion the wider the width. Also, the lower the dispersion the higher the amplitude of the FT of the pulse in the direction ν. This behavior is well consistent with the expected shift of the function $\hat{C}(\omega, \nu)$ in the ν direction for a dispersion different from 0.

In fact, for a pulse appearing in a large frequency bandwidth relative to the bandwidth of the dynamic spectrum, there is also a difference in the width of the FT of the pulse in the direction ω. However, it is only and totally due to the shape of the modulation function $\hat{C}(\omega, \nu)$. The larger the frequency bandwidth, the smaller the width of $\hat{C}(\omega, \nu)$ in the direction ν. Thus, for a pulse appearing in a large bandwidth, its FT is very narrow in ν, leading to a modification of the width of the FT in the direction ω.

Finally, in the direction ω, there is either no difference or a difference totally due to the modulation function in the other direction. In the other direction ν, there are always differences, whatever the shape of the pulse, which is totally defined by the level of dispersion undergone by the pulse. As a result, in light of these behaviors, it is possible to define a mathematical indicator allowing to quantify the level of dispersion.

10.2.2 Dispersion indicator

The shape of the pulse (and of the dynamic spectrum) is totally due to the impact of the dispersion on the modulation function $\hat{C}(\omega,\nu)$. Also, the level of dispersion can be estimated by looking at the shape of the norm of this modulation function relative to the dispersion. In order to easily evaluate the impact of the dispersion, the function $\hat{C}(\omega,\nu)$ can be discretized.

$$\hat{C}(\omega,\nu) \;=\; \int C(f)\cdot e^{-i\delta t(f)\omega}\cdot e^{-i\nu f}\cdot df \;\rightarrow\; \sum_{f} C(f)\cdot e^{-i(\;\delta t(f)\omega+\nu f\;)} \tag{10.24}$$

The norm $\|\hat{C}(\omega,\nu)\|^2$ can be finally determined by the following equation:

$$\|\hat{C}(\omega,\nu)\|^2 \;=\; \hat{C}(\omega,\nu)\cdot\hat{C}(\omega,\nu)^{\dagger} \;=\; \sum_{j,k} C(f_j)\cdot C(f_k)\cdot e^{-i(\;\Delta\delta t(f)_{j,k}\omega+\nu\Delta f_{j,k}\;)} \tag{10.25}$$

With $\Delta\delta t(f)_{j,k} \;=\; \delta t(f_k)-\delta t(f_j)$ the difference of time delay between the two frequencies f_k and f_j, and $\Delta f_{j,k} \;=\; f_k - f_j$ the corresponding frequency difference. The obtained sum can be divided into three parts depending on the frequency difference: a first one for $f=f'$, a second one for $f>f'$, and a third one for $f<f'$. We obtain, therefore, a first term for the same frequencies, and two symmetric terms where the global phases dependent of $\delta t(f)$ and f are opposite in sign. These two last terms can be rearranged into a unique term. The norm of the modulation function can be finally written in the simple form:

$$\|\hat{C}(\omega,\nu)\|^2 \;=\; \sum_{f_j=f_k} C(f_j)^2 + \sum_{f_j>f_k} C(f_j)\cdot C(f_k)\cdot 2\cos(\;\Delta\delta t(f)_{j,k}\omega+\nu\Delta f_{j,k}\;) \tag{10.26}$$

Figure 10.4 has shown that the pattern in the direction ν is centered around the value $\nu \;=\; 0$, and moreover, that the lower the dispersion the higher the maximum amplitude at $\nu \;=\; 0$. Based on these observations, we can define a reference value of the modulation function, corresponding to the case where there is no dispersion, i.e. $\delta t(f) \;=\; 0\;\forall\; f$, and evaluated at $\nu \;=\; 0$:

$$\|\hat{C}^d(\nu=0)\|^2 \;=\; \sum_{f_j=f_k} C(f_j)^2 + \sum_{f_j>f_k} C(f_j)\cdot C(f_k) \tag{10.27}$$

According to the fact that the cosine function is always lower than 1, we now have at $\nu \;=\; 0$ the following inequality dependent on the different dispersion time delays included in the term $\Delta\delta t(f)$:

$$\sum_{f_j>f_k} C(f_j)\cdot C(f_k)\cdot 2\cos(\;\Delta\delta t(f)_{j,k}\omega\;) \;\leq\; \sum_{f_j>f_k} C(f_j)\cdot C(f_k) \tag{10.28}$$

For the central frequential frequency $\nu \;=\; 0$, the rest of the equation 10.22 of the FT of the pulse is just an amplitude factor independent of the dispersion. The previous inequality leads into:

$$\|\hat{P}(\omega,0)\|^2 \;\leq\; \|\hat{P^d}(\omega,0)\|^2 \;\;\forall\;\omega \tag{10.29}$$

The other terms of the dynamic spectrum are, if there are unremoved, essentially contained in an area centered either in $\omega \;=\; 0$ or in $\nu \;=\; 0$. Then, the inequality can be extended taking into account the carried out marginalization.

$$\int \|\hat{S}_{DM}(\omega,0)\|^2\cdot d\omega \;\leq\; \int \|\hat{S}_{DM_0}(\omega,0)\|^2\cdot d\omega \;\;\forall\;\omega \tag{10.30}$$

Where $\hat{S}_{DM}$ designates the FT of the dynamic spectrum with a dispersion at a certain DM (more exactly a deviation relative to the ideal DM), and $\hat{S}_{DM_0}$ represents the reference dynamic spectrum without any dispersion (i.e. for DM_0 which is the ideal DM of the pulsar). We can finally conclude that to estimate the level of dedispersion, a good indicator $\Lambda(DM)$ is the marginalization relative to ω of the squared norm of the dynamic spectrum at the central Fourier frequency $\nu \;=\; 0$:

$$\boxed{\Lambda(DM) \;=\; \int \|\; S_{DM}(\omega,0)\;\|^2\cdot d\omega \;\leq\; \Lambda(DM_0)} \tag{10.31}$$

This dispersion indicator allows for quantifying the dispersion of a pulse. It is therefore possible to use it to search for the DM of a pulse. A search algorithm can be developed to compute the dispersion indicator in the 2D FT of the dynamic spectrum, in order to find the maximum value $\Lambda(DM_0)$.

10.3 DM search

10.3.1 Search algorithm

The pattern of the pulse in the FT of the dynamic spectrum follows a line that tends to be horizontal for the ideal DM. As we can see in Figure 10.4, the value of the dispersion indicator increases until the maximum value is obtained for this ideal DM. In consequence, we can assume that the function $\Lambda(DM)$ is a monotonic function before and after the maximum value. Using this property for the DM search, the DM space can be thus linearly divided into some values, where the dispersion indicator will be computed. The DM search window is then iteratively reduced around the best-found DM, corresponding to the maximum of the function $\Lambda(DM)$.

In the context of this study, the DM window was divided into 20 trial DMs values at each iterative step. For each step, the global algorithm is composed of four steps, iteratively carried out until the error on the dispersion indicator doesn't allow to clearly determine a maximum value. The four steps done at each iteration are:

- divide the DM range in 20 DMs linearly spaced and stored in the vector $\vec{DM}$;
- compute the dispersion indicator vector $\vec{\Lambda}$ for each of the 20 DM;
- extract the best DM, DM_b, corresponding to the maximum of the dispersion indicator vector: $DM_b = \vec{DM}[i] \mid \max\{\vec{\Lambda}\} = \vec{\Lambda}[i]$;
- fix the new DM range from $DM[i-1]$ to $DM[i+1]$.

The second step is itself decomposed into five sub-steps, allowing the calculation of the dispersion indicator for the given DM to try:

- dedisperse the initial dynamic spectrum with a trial DM ;
- produce the 2D FT of the dynamic spectrum $\hat{S}(\omega, \nu)$;
- compute the squared norm of the obtained FT of the dynamic spectrum $\|S(\omega, \nu)\|^2$;
- compute the marginalized function by summing over ω;
- store the value in $\nu = 0$ corresponding to the dispersion indicator $\Lambda(DM)$ in a vector.

To avoid as much as possible the propagation of errors in the computing of the dispersion indicator, the number of intermediate operations has been limited to the minimum. The dedispersion realized for each DM is therefore carried out directly on the initial dynamic spectrum, i.e. at DM = 0 pc.cm^{-3}. The only operation performed on this dynamic spectrum is the RFI mitigation, allowing for the zapping of the majority of the RFIs and the time samples where are located the analog beam tracking adjustments.

10.3.2 Dedispersion method

In the context of the study of super dispersive effects, we need to obtain a precise dedispersion. Thus, rather than perform a standard incoherent dedispersion shifting the bins in the time domain; the dedispersion of the dynamic spectrum has been made using a coherent dedispersion, correcting the phases in the Fourier domain.

If we place our point of view on the resulting image plane, the dispersed dynamic spectrum $S(t, f)$ can be written as the ideally-dedispersed model previously used $S^d(t, f)$ convolved by a transfer function $H(t, f)$. This one moves differently relative to the frequency of the pulse on the resulting image.

In the global dynamic spectrum, if the cleaning of the noisy data was performed, the RFIs and the tracking adjustments are normally almost entirely removed. Also, the white noise is not affected by the dispersion, and at first approximation, the transfer function impact only the pulse term. The resulting image can be then expressed by:

$$S(t, f) = S^d(t, f) * H(t, f) \approx P^d(t) * H(t, f) \tag{10.32}$$

The dispersion being solely dependent on the frequency, it is possible to simplify the problem by looking at the dispersion channel by channel. For each channel corresponding to the fixed frequency f, the problem is then a convolution of a time series $P_f^d(t)$ with the time-dependent transfer function $H_f(t)$. The inverse problem can be ultimately formulated in a simpler form by the following relation:

$$S_f(t) \;=\; P_f^d(t) * H_f(t - t_r^f) \tag{10.33}$$

Using the convolution theorem, the dedispersed pulse can then be retrieved by the division term by term:

$$\forall\, \omega \;\; \hat{S}_f(\omega) \;=\; \hat{P}_f^d(\omega) \cdot \hat{H}_f(\omega) \;\Leftrightarrow\; P_f^d(t) \;=\; \mathcal{FT}^{-1} \left\{ \; \frac{\hat{S}_f}{\hat{H}_f}(\omega) \; \right\} \tag{10.34}$$

The transfer function must locate the flux at the reference position for the given frequency t_r^f. It can be then modeled in the same way as the dispersive function $D(t)$ used for the pulse's model by a Gaussian model:

$$H_f(t - t_r^f) \;=\; \exp\left(-\frac{(t - t_r^f)^2}{2\sigma_t^2} \right) \tag{10.35}$$

Where t_r^f is given by the dispersion relation of the pulsars. The Gaussian model allows us to determine the form of the FT of the transfer function $\hat{H}_f(\omega)$:

$$\hat{H}_f(\omega) \;=\; \mathcal{FT}\{\; H_f(t - t_r^f) \;\}(\omega) \;=\; e^{-it_r^f\omega} \cdot \sigma_t \cdot e^{-\frac{1}{2}\sigma_t^2\omega^2} \tag{10.36}$$

In the context of the dedispersion, we need only to correct the phases of the signal, represented by the first term. The two other terms concern the amplitude and can be then ignored. Finally, the coherent dispersion is thereby realized in each frequency channel, by the computing of the dedispersed time series $P_f^d(t)$ following the equation:

$$\forall\, \omega \;\; P_f^d(t) \;=\; \mathcal{FT}^{-1} \left\{ \; \hat{S}_f(\omega) \cdot e^{+it_r^f\omega} \; \right\} \tag{10.37}$$

The final dedispersed dynamic spectrum $S^d(t, f)$ used for the computing of the dispersion indicator is then built by stacking these different time series in the frequency direction.

10.3.3 Uncertainty

To find the best DM, it needs to identify the maximum value of the dispersion indicator. The uncertainty on the DM is then directly related to the uncertainty in the definition of the best dispersion indicator at each step of the search algorithm.

In the context of the search in a dynamic spectrum where RFIs and the tracking adjustments are properly removed, each point of the FT of the dynamic spectrum $\hat{S}$ is only based on the FT of the pulsar and noise. The dispersion indicator can be then evaluated for each couple (ω, ν), in first approximation, as a part due to the pulse $\tilde{P}$ and a part representing the background noise $\tilde{B}$.

$$\|S(\omega,\nu)\|^2 \;=\; \|\hat{B}(\omega,\nu)\|^2 + (\; \|\hat{P}\|^2 + 2\; \hat{B} \cdot \hat{P} \;)(\omega,\nu) \;=\; \tilde{B}(\omega,\nu) + \tilde{P}(\omega,\nu) \tag{10.38}$$

Therefore, the associated error in each (ω, ν) is defined as the quadratic sum of the two errors. Theoretically, a perfect noise must be located in the center of the FT of the dynamic spectrum, and can be easily removed by setting this center point to 0. However, real noise is not perfect white noise, and the FT of the noise is extended. As a consequence, it remains a residual of the noise in the FT of the dynamic spectrum, limiting the definition of a maximum value for the pattern of the pulse.

This noise term is a function of the observation and can be measured directly in the FT of the dynamic spectrum. The pattern of the FT of the pulse is located on one of the two diagonals. Hence, the noise level can be evaluated in another part of the plan, where the pattern of the FT of the pulse is (almost) not present. The standard deviation of this part defines the uncertainty due to this Fourier noise.

$$(\Delta\Lambda)^2 \;=\; \sum_{\omega}\{\;(\;\Delta\tilde{B}(\omega,0)\;)^2+(\;\Delta\tilde{P}(\omega,0)\;)^2\;\} \;=\; N_\omega\cdot Var\{\;\tilde{B}(\omega,0)\;\}+\sum_{\omega}(\;\Delta\tilde{P}(\omega,0)\;)^2 \quad (10.39)$$

Where $Var\{\tilde{B}(\omega,\nu)\}$ represents the variance of the noise, and N_ω is the number of Fourier frequencies ω. The second source of uncertainty is the uncertainty on the FT of the pulse due to the errors in the different parameters of the pulse. According to the equation 10.19 of the FT of the pulse, three parameters that can eventually generate an error: the amplitude of the pulse A_p, the width of the pulse σ_t, and the frequency bandwidth of the pulse Δf occurring in the function $\hat{C}(\omega,0)$.

$$(\Delta\hat{P})^2(\omega,0) \;=\; \left|\frac{\partial\hat{P}}{\partial A_p}\right|^2\cdot(\Delta A_p)^2+\left|\frac{\partial\hat{P}}{\partial\sigma_t}\right|^2\cdot(\Delta\sigma_t)^2+\left|\frac{\partial\hat{P}}{\partial(\Delta f)}\right|^2\cdot(\Delta(\Delta f))^2 \quad (10.40)$$

In fact, none of these parameters is completely known, leading to the inability to calculate analytically the uncertainty. Furthermore, the DM and the frequency bandwidth of the pulse generate variations in the function $\hat{C}(\omega,\nu)$, whose exact form is unknown. Also, these variations are not necessarily linear, increasing then the difficulty of evaluation. That's why the search is realized by iterations rather than by direct resolution.

In addition, contrary to the noise, these uncertainties are dependent on the shape of the pulse and not on the observation. It is not therefore possible to evaluate this error directly by a measurement in the data. Because of these problems, these uncertainties might be evaluated using simulations in order to quantify the variations relative to the variation of each parameter.

For this preliminary study, only the uncertainties on the noise have consequently been estimated. For the moment, it is difficult to evaluate the importance of the error created by the lack of accuracy of the pulse shape compared to the error due to the noise. Nevertheless, the uncertainty defined here allows us to have a first rough estimation of the order of magnitude we can expect in terms of DM. However, for a deeper study, it is obviously required to realize a deeper and more accurate quantification of the uncertainties, implementing for example a Monte-Carlo simulation allowing to cover plenty of cases. Moreover, deeper simulations additionally allow for the identification of eventual covariances.

For the DM measurement presented in this work, the uncertainty on the dispersion indicator is determined based on the standard deviation computed on a noisy square. This standard deviation is then normalized to obtain the corresponding standard deviation on the line $\nu = 0$ used for the calculation of the final Λ.

$$\Delta\Lambda \;=\; \sqrt{\frac{N_\omega}{N_{noise}}}\cdot\sigma\{\;\hat{S}_{noise}\;\} \quad (10.41)$$

Where $\sigma\{\hat{S}_{noise}\}$ designates the standard deviation of the part of the FT of the dynamic spectrum with only a noise signal, and N_{noise} corresponds to the size of this noisy square. The uncertainty in the DM can be finally determined based on the range of the dispersion indicator satisfying the following inequation:

$$\Lambda(DM_b)-\Delta\Lambda(DM_b) \;<\; \Lambda(DM_i)+\Delta\Lambda(DM_i) \quad (10.42)$$

Where DM_b is the best DM corresponding to the greatest computed dispersion indicator $\Lambda(DM_b) = \max\{\;\vec{\Lambda}\;\}$. This inequation means the inability to clearly determine a maximum value of the dispersion indicator. While the two closest neighbors don't satisfy the previous inequation, the DM window can be tightened around this best-found DM. The two edges of the DM uncertainty are therefore defined as the first neighbors, DM_i, having an upper limit (at 1σ) of the dispersion indicator (i.e. added to the uncertainty) less than the lower limit of the dispersion indicator of the best DM.

10.3.4 Test to an integrated observation of B1919+21

In order to test the method of DM search explained in this chapter, the DM has been searched in some observations of some pulsars. To illustrate the result of the method, a DM search on observation realized with NenuFAR the 26^{th} of October 2021 of B1919+21 is presented here. B1919+21 is particularly interesting because it is weakly scattered, allowing to obtain a relatively clear DM value. Furthermore, its profile presents five peaks, allowing for testing if the presence of several peaks disturbs the DM search method.

The observation has been made in single pulse mode, for a duration of 1 hour and using the total available bandwidth of 75 MHz. Table 10.1 presents the characteristics of the observation used for the search. The observation was then divided into sub integrations of one period using DPSR, and folded to obtain a single dynamic spectrum with 1 024 phase bins and 384 channels.

Table 10.1: Characteristics of the observation of B1919+21 used for the DM search.

Characteristic	**Value**
PSR	B1919+21
MJD	59044
Duration	3285.0516 s
Folding period	1.3373021601895 s
Time resolution	1.3059 ms
Centre frequency	49.9023 MHz
Frequency resolution	195.3125 kHz

The DM search is carried out in a range of $\pm$ 0.1 pc.cm^{-3} around the reference DM from the ATNF catalog of 12.44399 pc.cm^{-3}. Figure 10.5 presents four graphs of the dispersion indicator Λ relative to the tried DMs. Each graph corresponds to an iteration of the algorithm at a fixed DM resolution. The top left panel shows thereby the first iteration of the algorithm for a DM resolution of 1.10^{-2} pc.cm^{-3}, the top right for 1.10^{-3} pc.cm^{-3}, the bottom left for 1.10^{-4} pc.cm^{-3}, and the bottom right for 1.10^{-5} pc.cm^{-3}. As expected, the variation of the dispersion indicator relative to the DM is monotonic in each part before and after the maximum. The algorithm has finally converged at the fourth iteration in a DM of 12.43890 $\pm 1.10^{-5}$ pc.cm^{-3}.

The obtained DM was then compared to the reference value of the ATNF catalog and to the best DM found by PDMP. PDMP is the standard method used to correct the DM of an observation with NenuFAR. Contrary to the method presented here, the process of PDMP works differently, aiming to obtain the best SNR of the integrated profile. Table 10.2 indicates the different SNRs and peak SNRs obtained for the three DMs. It appears that the SNR and peak SNR are largely increased by the correction done by PDMP. Also, although the DM search in Fourier space doesn't try to maximize the SNR, the obtained SNR and peak SNR are slightly increased too.

The observation has also been dedispersed for each of the three DMs, and the resulting integrated profile was computed. Figure 10.6 shows in the left column a zoom of the folded frequency-phase plan around the pulse, and the right column shows the corresponding integrated profiles for each DM. In addition of a better SNR, the pulse in the frequency-phase plan seems to be slightly better aligned for the DM found by the method in Fourier space. Also, in the integrated profile, the structures due to the multiple peaks seems to be better defined.

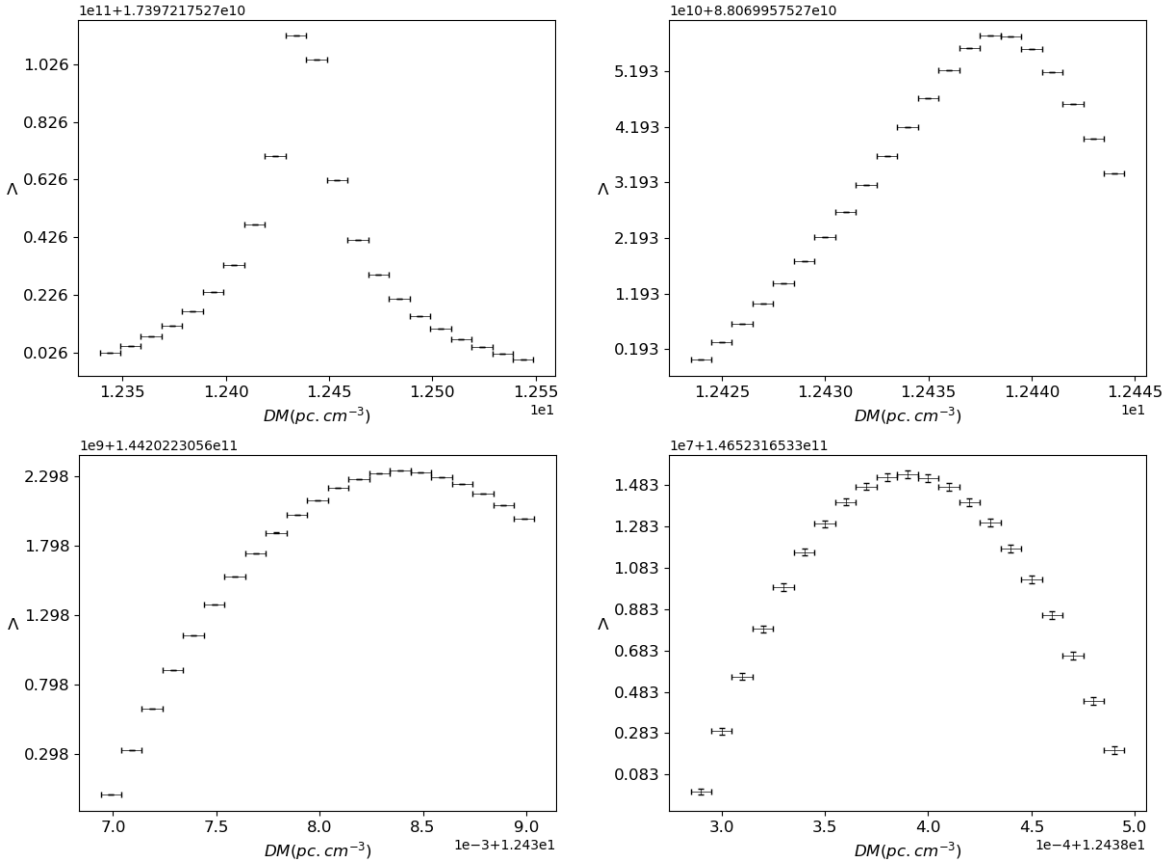
Table 10.2: Values of the best DM obtained with PDMP and the DM search in Fourier space, with their resulting SNR and peak SNR. The first line corresponds to the reference DM given in the ATNF catalog.

Method	**DM** ($pc.cm^{-3}$)	**ΔDM** ($pc.cm^{-3}$)	**SNR**	**Peak SNR**
Reference	12.44399	-	1716	368
PDMP	12.4398	$\pm 4.10^{-4}$	1901	454
Fourier search	12.43890	$\pm 1.10^{-5}$	1995	479

We can conclude that the DM search in Fourier space can be used to try to determine a DM value for a single pulse. Indeed, the method used by PDMP is totally based on the SNR as criterion for a good dedispersion. As a result, it works well in the cases where the SNR is high as this folded observation. However, in the context of a single pulse, the SNR is largely decreased, which is pretty bad for the method used by PDMP.

Furthermore, the precision reached by PDMP is less than for the Fourier search with an error in 10^{-4} $pc.cm^{-3}$.assume that the DM error obtained with the standard method will be really unsufficient to hope to distinguish a super dispersive effect.PDMP is completely based on the SNR, we can expect that the error will still increase for a determination on a single pulse. Furthermore, for this work, it is required to divide the single pulse in frequency subbands, leading to increase the error yet. According to the order of magnitude of the predicted various effects, we can reasonably assume that the DM error obtained with the standard method will be really unsufficient to hope to distinguish a potential super dispersive effect. The search with the dispersion indicator is normally less dependent of the SNR. As a consequence, we can expect that the better precision obtained with the folded observation doesn't decrease overmuch.

Figure 10.5: Iterative steps of the DM search for four successive decreasing values of DM resolution. Top left: 1.10^{-2} pc.cm^{-3}, top right: 1.10^{-3} pc.cm^{-3}, bottom left: 1.10^{-4} pc.cm^{-3}, and bottom right: 1.10^{-5} pc.cm^{-3}. Each step is computed around the best DM value found in the previous step for the DM resolution ten times greater.



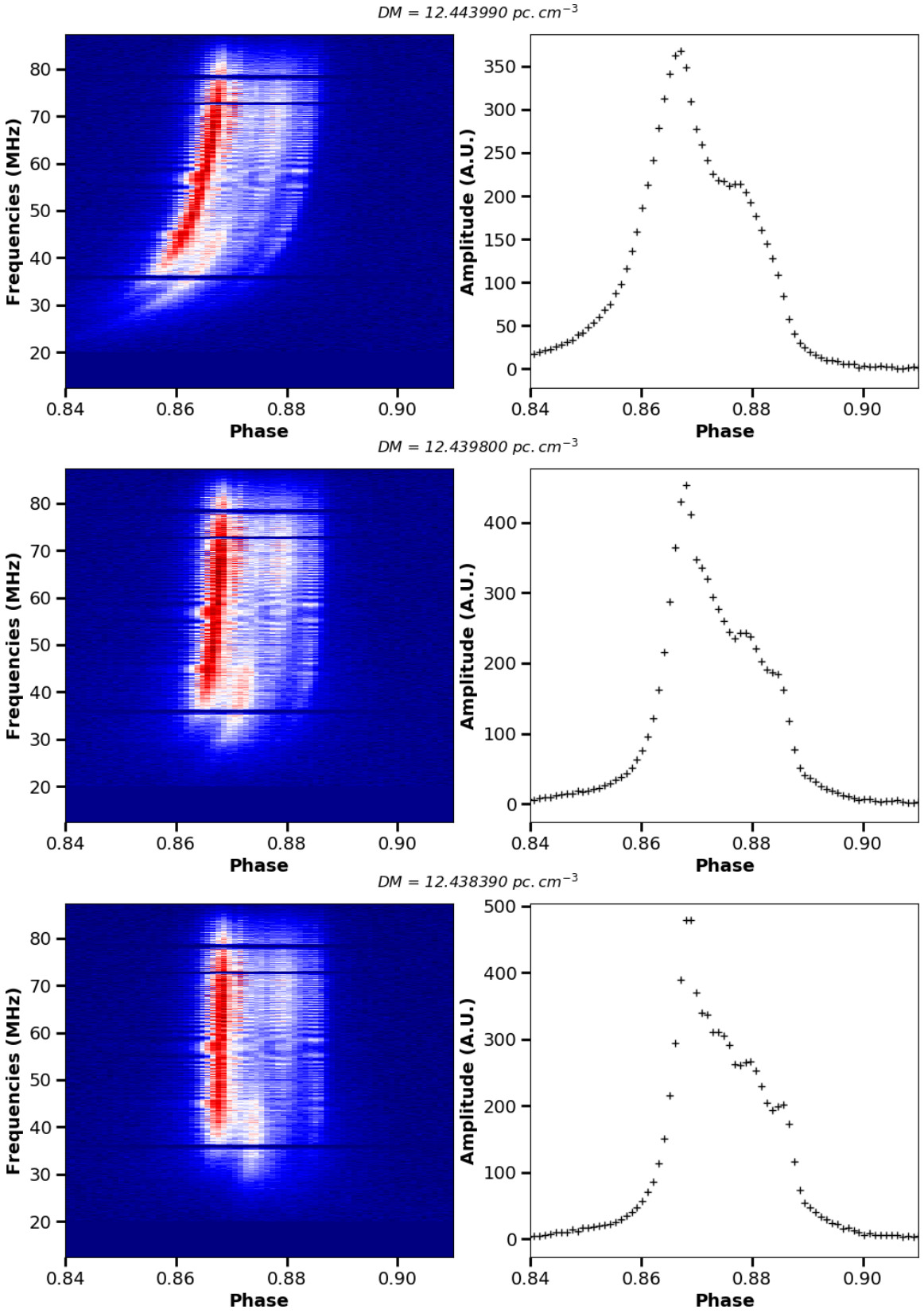


Figure 10.6: Comparison of the frequency-phase plane and corresponding integrated profiles of B1919+21 after dedispersion using three different DMs: the DM of the ATNF catalog and resulting from DM searches. All graphs are zoomed around the pulse in the phases 0.84 - 0.91. From top to bottom: the reference DM, the DM found by PDMP, and the DM found by the search in Fourier space.

Résumé du chapitre :

Un spectre dynamique peut être modélisé par la somme de cinq termes : le bruit de fond, les sauts d'amplitude de repointage du télescope, les RFI large bande, les RFI étroites en fréquence, et l'impulsion du pulsar. En utilisant la linéarité de la transformation de Fourier, la transformée de Fourier (TF) d'un spectre dynamique est donc la somme des TF individuelles de chacun de ces termes.

Le bruit peut être modélisé par une distribution normale réduite, apparaissant uniquement au centre de la TF du spectre dynamique. Les sauts d'amplitude suivent des fonctions logarithmes, les localisant au sein de la TF finale suivant l'axe des fréquences fréquentielles (la dimension duale de la dimension fréquentielle réelle). Les RFI peuvent-elles être modélisées par des fonctions portes, apparaissant finalement comme une somme de sinus cardinaux alignés sur les axes centraux principaux de la TF. En conséquence, en théorie un masque aligné sur les axes centraux principaux de la TF du spectre dynamique devrait permettre de corriger, au moins en partie, ces différents effets.

Pour modéliser une impulsion, il est nécessaire d'utiliser le produit de deux termes : un terme de modulation fréquentielle de la distribution du flux en fréquence, et un terme relatif à la distribution du flux en temps. Le terme de distribution en temps peut être défini par une fonction gaussienne, ayant un terme de position centrale proportionnel à la dispersion subie à cette fréquence. Dans le cadre d'une dispersion nulle, la TF de l'impulsion est le produit de deux termes indépendants. Et dans le cas inverse, une partie de l'information temporelle est gardée dans le terme de la TF de la distribution du flux en fréquence.

Les structures autres que celle de l'impulsion sont particulièrement visibles, au contraire de cette dernière, dans le contexte d'un spectre dynamique non dédispersé, et sont "noyées" dans le contexte d'un spectre dynamique dédispersé. De plus, généralement, ces premiers termes peuvent être corrigés efficacement par les méthodes standard. En première approximation, seul le terme de l'impulsion peut être considéré comme dépendant de la dispersion pour la recherche effectuée.

Théoriquement, la norme carrée de la TF de l'impulsion est donnée par la forme de la norme carrée de la TF du terme de distribution en fréquence, et contenue dans une enveloppe gaussienne dépendant des fréquences duales du temps. Des simulations de spectres dynamiques avec une impulsion plus ou moins dispersée ont confirmé les résultats attendus, à savoir une structure plus ou moins inclinée de la TF de l'impulsion. Aussi, dans la direction duale du temps, la TF de l'impulsion est bien toujours la même en forme et en amplitude. Et dans l'autre direction, la structure de la TF de l'impulsion devient de plus en plus étroite et forte en amplitude lorsque l'on s'approche de la bonne mesure de dispersion.

Partant de ces constats, un indicateur de dispersion a été déterminé comme étant, au niveau de la fréquence centrale dans la direction duale des fréquences, l'intégration de la norme carrée de la TF du spectre dynamique le long de la direction duale du temps. Cet indicateur étant maximal dans le cas d'une dispersion idéalement corrigée, la recherche de la meilleure mesure de dispersion peut s'effectuer par la maximisation de celui-ci.

Un programme de recherche de mesure de dispersion a été développé sur ce principe. Afin de minimiser les programmes intermédiaires, le spectre dynamique est dédispersé de manière cohérente directement au sein d'une première étape du programme. L'indicateur de dispersion est ensuite calculé pour vingt mesures de dispersion différentes linéairement espacées dans un intervalle de 0,2 $pc.cm^{-3}$. La recherche se fait ensuite de manière itérative autour du maximum trouvé jusqu'à ce que le maximum soit impossible à trouver du fait de l'incertitude.

Cette méthode de recherche (et le programme associé) a été testée sur une observation empilée du pulsar B1919+21, qui possède un flux important, et un profil comportant cinq pics. Après comparaison avec la méthode habituellement utilisée, la mesure de dispersion finalement trouvée par la méthode en espace de Fourier est légèrement différente, et semble donner un alignement des composantes du profil légèrement meilleur, tout en permettant de réduire l'incertitude associée d'un facteur vingt pour cette observation.

Chapter 11

Results of the DM measurements

11.1 Presentation of the study

The study of the deviations to the cold plasma dispersion law has been performed using observations carried out with NenuFAR. For the DM measurements, the choice was to realize the measure on single pulses and consequently, the data are from observations in single pulse mode (see Chapter 2.5). The constraint of single pulse measurement limits the selection of pulsars to those which are sufficiently powerful, to obtain at least some pulses clearly visible above the noise. Furthermore, in order to minimize the possible perturbations because of scattering, the selected pulsars must be weakly scattered.

11.1.1 Selected pulsars

According to the NenuFAR pulsar census, only 21 pulsars can be detected in single pulse mode with NenuFAR. 4 pulsars on this list have been finally selected for the measurements. The first selected pulsar is B1508+55, which is the third most intense pulsar of the 21. B1508+55 lends itself well to the measurement of DM because it has the advantage of a sharp and very weakly scattered profile. One can notice that its integrated profile presents an echo of the pulse close to this one caused by a scattering screen. Nevertheless, this echo features a low flux compared to the pulse. Also, it is undetectable in single pulses (see Figure 11.5 and 11.7 in Section), and is consequently not disturbing for this work. Furthermore, for this study, the DM measurements were made on two different observations of B1508+55. This allows us to see potential time variations of the potential deviations (then other than well-known long-timescale variations) on time scales larger than an observation.

The second chosen pulsar is B1133+16. It is the second most intense pulsar and shows a profile with a clear double peak structure. Moreover, it is a broadband profile without scatter broadening. B1133+16 is thus interesting for two reasons: firstly, its very sharp and intense profile is easily detectable; secondly, compared to B1508+55 which has a simple component profile, B1133+16 possesses two clearly distinct components.

The most powerful pulsar detectable in single pulses at NenuFAR frequencies is B1919+21. It is weakly scattered and reveals a complex internal structure with five very close peaks (contrary to the two distinct peaks of B1133+16). In addition, these five components appear at different bandwidths, making this pulsar particularly interesting for DM measurements along the bandwidth. As a consequence, B1919+21 was added to the set of pulsars to study.

The fourth and last pulsar is only the 13^{th} most powerful of the 21 pulsars detectable in single pulses. However, B0950+08 is particularly interesting due to its double peak profile, where each peak seems to have a different DM. In this work, the DM measurements are realized in a unique single pulse, independently of the others. Consequently, although it shows a drifting subpulse (see Chapter 1.2.3), that should not be a problem in our case.

Table 11.1: List of the selected pulsars, with the MJD date of the studied observation, the period and DM from the ATNF catalog.

PSR	MJD	P (s)	DM ($pc.cm^{-3}$)
B0950+08	58942	0.2530651649482 (9)	2.96927 (8)
B1133+16	58883	1.187913065936 (3)	4.84066 (34)
B1508+55	58887 , 58944	0.739681922904 (4)	19.6191 (3)
B1919+21	59044	1.3373021601895 (9)	12.44399 (63)

11.1.2 Number of subbands to use

To detect potential DM deviations, it is required to have several measurement points per single pulse. Thus, every single pulse must be divided into several frequency subbands, permitting measuring a DM value for each one.

The larger the number of measurement points, the easier the detection of a potential trend. Consequently, in the ideal case, we want the maximum number of subbands possible. However, increasing the number of subbands means decreasing the bandwidth over which the DM is measured. Also, the smaller the bandwidth, the smaller the flux of the pulse. Although the used method is not based on the maximization of the SNR, the determination of the maximum dispersion indicator requires a sufficient flux of the pulse to be discriminated from the background noise. If the dispersion indicator is dominated by noise, the dispersion of the values of the dispersion indicator increases, making it difficult to find the maximum value.

To obtain proper measurements, I needed to define a number of subbands not too small to evaluate a potential trend, but at the same time with a bandwidth of subband not too small to have a sufficient flux. To reveal fine variations, maybe non-linear, I have required a minimum of four measurement points. The bandpass of NenuFAR is relatively flat over the majority of the bandwidth. Nevertheless, there is a falling off of the bandpass below 25 MHz and above 80 MHz. In addition, because of a substantial fraction of RFIs in the low frequencies, a large part, and sometimes even the totality of the channels below 20 or 25 MHz is generally flagged. Moreover, the flux of pulsars is unconstant over the entire bandwidth. As a consequence, in order to have at least four points of measure in the pulse of the pulsar, I have performed measurements using a minimum number of subbands of six.

The diverse measurements carried out on the different pulsars have shown it is difficult to obtain four valid points. Indeed, for many pulses, the two lowest subbands and also the highest subband are flagged or don't contain sufficient flux. As a consequence, the number of measurement points is frequently of only three. The selected pulsars are extremely intense, and I have chosen to measure the DM on the most intense single pulses of the observation. We can thus assume the flux is sufficiently important to increase the number of subbands. As a consequence, I have increased the number of subbands to eight, allowing to win one or two points of measure per pulse.

Ultimately, I have tried increasing the number to 12 subbands. Unfortunately, many measured DM values seem aberrant. Even for B1133+16, which shows single pulses relatively constant over a large part of the bandwidth, the SNR in each of the 12 subbands is insufficient, resulting in measures polluted by noise.

For each observation, the measurements have been done three times, with six, eight, and twelve subbands. This triple measurement allows for identifying the aberrant points by comparison between the three measures. For six and eight subbands, the number of subbands is close, leading that the corresponding bandwidths are equally close. Without being equal, the DM values in these two cases might be consequently rather close.

The search for the DM is indeed a search for the maximum dispersion indicator. For a sufficiently intense pulse, the dispersion indicator can be clearly determined, resulting in a typical monotonic function relative to the DM, which must be peaked around the DM of the pulsar. In the case of a not sufficiently intense pulse, the noise in the dynamic spectrum is dominant, generating a random statistical dispersion of the values of the indicator. As a consequence, in the case of a pulse with low SNR, the algorithm may select a peak caused by noise rather than the peak caused by the pulse.

It results from comparisons between the three measures that a too-large difference in the obtained DM values means the measure is dominated by noise. For points presenting such a difference, the result can't be considered a sure DM value. In the results with eight subbands, the bad measurement has been therefore manually removed. In addition, a second control was done by checking by eye the frequency-time plane of the single pulse. In the case of an evident lack of flux or overmuch remaining RFIs in a subband, the measurement of the corresponding subband is also directly removed to avoid the problem of a noisy dominant measure.

11.1.3 Detailed process

Before realizing the DM search, the first step was to prepare initial data, which are simple dynamic spectra, in order to obtain a file with a series of time sub-integrations of one period of the pulsar for every single pulse. The initial single pulse data has been then folded at the period of the pulsar using DSPSR. The second step was to perform an RFI mitigation using the software NENUPLOT. NenuPlot is a program developed by L.Bondonneau, which is based on COAST GUARD of P.Lazarus.

In order to reduce the number of intermediate steps, the RFI mitigation could be done directly on the dynamic spectrum using PRESTO. The advantage would have been to don't use DSPSR. However, firstly, working on files with a sub-integration structure for every single pulse is easier. Secondly, NENUPLOT, such as COAST GUARD, write information about RFIs to flag directly in the weight matrice contained in the FITS file. Thus, contrary to PRESTO which writes RFIs information in three different binary files, NENUPLOT allows to manage more simply the RFI mask. Furthermore, COAST GUARD is an efficient and recognized program, and NENUPLOT is the standard program used for the RFI mitigation of the pulsar observations with NenuFAR.

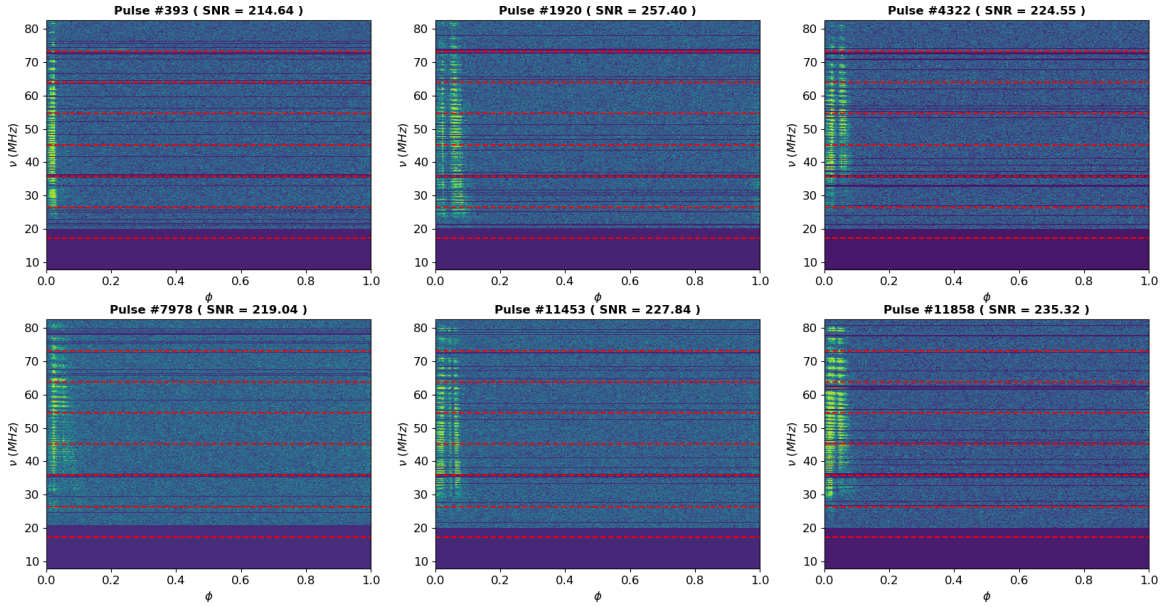


Figure 11.1: Dynamic spectra of the six most intense single pulses in the observation of B0950+08 at the MJD 58942. The limits of the subbands are represented by the dashed red lines.

The computing time is substantial, needing to select a rather limited subset of pulses. Finally, I have chosen to work on the six most intense pulses. For each pulse, the dynamic spectrum is divided into subbands. For each subband, the DM is then individually searched in a window of 0.1 pc.cm^{-3} around the reference value of the ATNF catalog. For pulsars showing the largest DM time variabilities, the DM value can change to some units of 10^{-2} pc.cm^{-3} between observations over some years (Donner et al. 2020). The DM window was therefore chosen in order to be sufficiently large to be sure to find a DM that is relatively far from this reference value. Moreover, before the search, a first dedispersion at the reference DM has been realized. For each pulse, the resulting frequency-phase plane could be manually checked to verify that the reference DM is close to the ideal value.

For each trial DM, the initial entire dynamic spectrum of the studied subband, i.e. at DM 0, is dedispersed for the DM to try. On the dedispersed dynamic spectrum, the sub-integration of the pulse to study is isolated, and the dispersion indicator and its associated error are then calculated. This process is done until the convergence of the algorithm is reached. All single pulses and subbands are finally iteratively processed.

11.2 DM measurements

For each observation, the first figure shows the dynamic spectra (frequency-phase plane) of the sub-integrations of the six major single pulses. The dynamic spectra are ordered in time and are dedispersed for the reference DM of the ATNF catalog. The eight subbands are also indicated by the dashed red lines, allowing us to see if the pulse is properly present with sufficient flux.

The second figure presents the measured DM values relative to the frequency for the six pulses, equally ordered in time. The values correspond to the valid measurements carried out in the observation divided into eight subbands. The uncertainty at 1σ is represented by the green contour, and the blue contour represents the uncertainty at 3σ. For each pulse, the mean DM of the six DM values is indicated by the dashed red line. Finally, a linear fit of the DM values, weighted with the uncertainty at 1σ, is drawn as the dark green line.

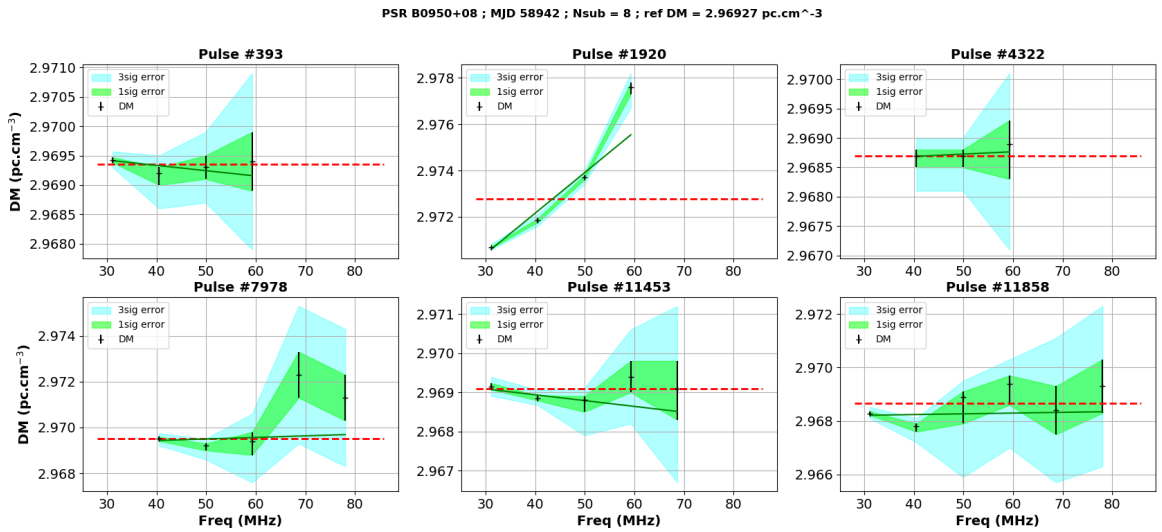


Figure 11.2: DM measurements of the different subbands for the six most intense pulses of the observation of B0950+08 at the MJD 58942. The 1σ error is defined by the green contour, and the 3σ error by the blue contour. The red dashed line corresponds to the mean DM value, and the dark green line to the weighted linear fit.

11.2.1 B0950+08

Compared to the other studied observations, the pulses of B0950+08 are rather wide. Moreover, pulses can be seen on a large bandwidth, ranging about from 30 to 80 MHz. The measurements have been therefore carried out from the third to the eighth subband.

On the six pulses, B0950+08 presents a time variability of the pulses. The first pulse (number 393) in Figure 11.1 features a single component, while four of the five others show two components. The exception is pulse 11 453 which seems to present a third narrow component between the two ordinary wider components.

In terms of frequency, the differences are small. Nevertheless, there is a variation of the lowest frequency where the pulsar is visible, with pulse 1 920 seen from about 25 MHz, and pulse 4 322 seen only from 30 to 35 MHz. Also, one can notice small-scale fluctuations of flux in frequency for every pulse. According to the scintillation measures done using LOFAR (Wu et al. 2022) and using a frequency dependence at the power of 4.4, the scintillation bandwidth at the NenuFAR frequencies typically ranges from about 2 Hz to 7 kHz. The scintillation is therefore almost entirely averaged within the channel bandwidth of 195 kHz. However, concerning B0950+08, a scattering screen close to Earth produces scintillation with a higher amplitude than most of the pulsars, increasing its scintillation bandwidth to approximately 100 kHz at 88.57 MHz (Smirnova & Shishov 2008). As a consequence, the observed fluctuations in Figure 11.1 could probably be attributed to the scintillation.

For pulse 393, the flux in frequencies above 60 MHz is insufficient to obtain consistent values between six and eight subbands, and these high frequencies are removed. Pulse 4 322 shows a sufficient flux located in a smaller bandwidth, and the measurements can be done only between about 35 and 65 MHz. Finally, the subband of pulse 7 978 around 30 MHz shows a lack of flux and is removed.

The DM values presented in Figure 11.2 are, for five of the six pulses, consistent with a constant DM (at least within the 3σ error), except for two points of pulse 11 858. For this last pulse, the two values at 30 and 40 MHz are slightly above the mean DM of the pulse.

Pulse 1 920 is very intriguing, revealing a clear decrease of the DM with decreasing frequencies. This trend is equally observed with the measurements using six subbands. As we can see in Figure 11.1, the pulse shows a first narrow component and a second wider component. Moreover, pulse 1 920 is the most intense of the observation and might therefore provide the most precise measurements. We can notice a curvature of the pulse is visible, at least of the second component, leading to a DM greater than the

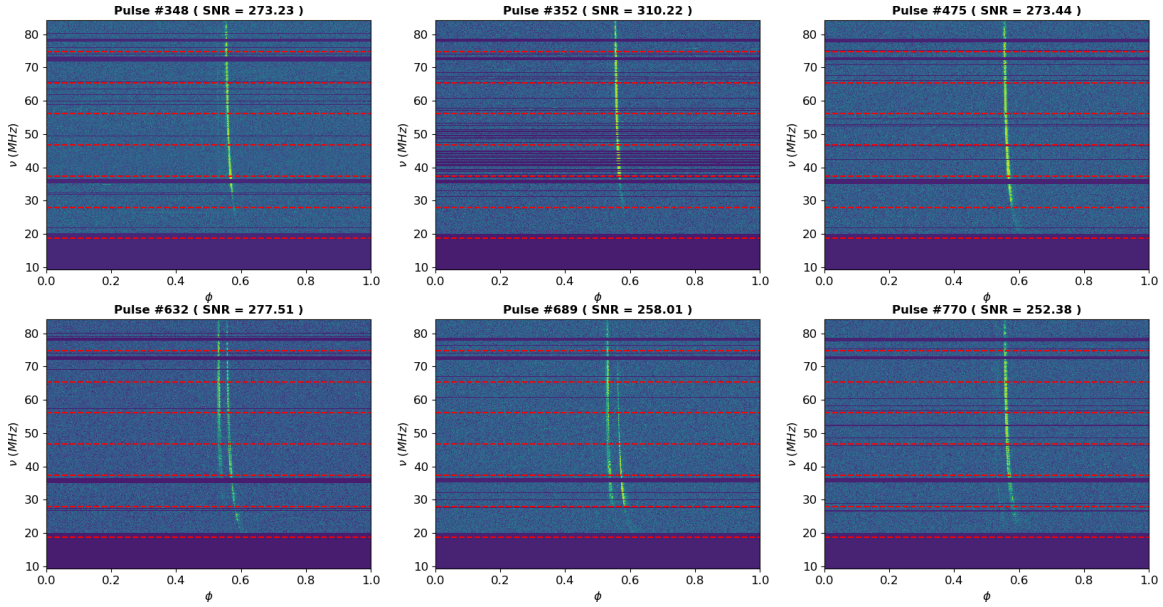


Figure 11.3: Dynamic spectra of the six most intense single pulses in the observation of B1133+16 at the MJD 58883.

reference DM value. The DM value of the subband around 30 MHz is $2.97067(^{+3}_{-5})$ pc.cm^{-3}and is indeed effectively greater than the reference DM of 2.96927 pc.cm^{-3}.

The DM measurements performed in this observation of B0950+08 show values rather consistent with a constant DM, with all the same slight variations compared to the computed mean DM. However, the most intense pulse of the observation presents a substantial difference with the mean DM, featuring a DM value that decreases with frequency.

11.2.2 B1133+16

B1133+16 is a pulsar presenting an integrated profile with two narrow components. However, the six most intense pulses in Figure 11.3 show an important time variability. Two of the six feature the double component structure, while the four others have just a unique component. In addition, the two pulses with a double-component structure have a frequency variability of the flux of each component. In terms of flux, for pulse 632, the second component is dominant, while it is the first component for pulse 689.

For the four pulses with a single component structure, the pulse is all the time composed of the second component of the integrated profile. There are all the same very faint appearance in the middle frequencies of the first component for pulse 348 and around 30 MHz for pulse 770.

In terms of frequency, the six pulses have an important flux on a larger bandwidth than observations of other pulsars of this study. Moreover, all the pulses have a flux relatively constant between 25 and 85 MHz. As a consequence, the measurements have been performed between the third and the eighth subband. Also, all the values measured with eight subbands are consistent with the values obtained with six subbands, and no measure points are thereby removed.

The DM measurements presented in Figure 11.4 are in the majority, consistent with a constant DM. However, it remains some measure points with slight variations around the mean DM value. These variations, located in pulses 348, 475, 632, and 770 occur at some precise frequencies and become consistent for most of them with the mean DM taking the 3σ uncertainty. The two exceptions concern the lowest subband of pulses 348 and 632, showing a DM value lower than the mean value.

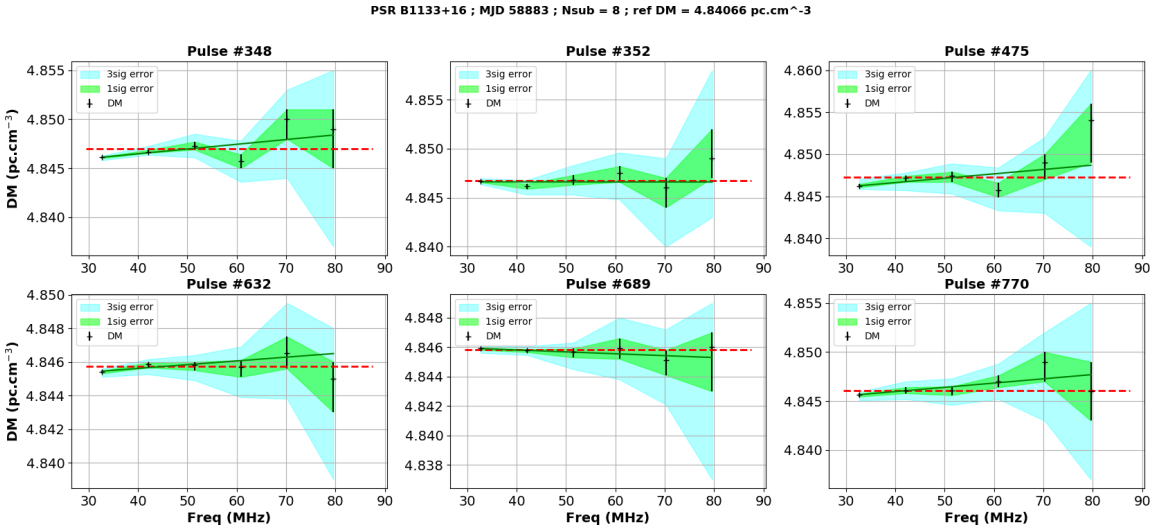


Figure 11.4: DM measurements of the different subbands for the six most intense pulses of the observation of B1133+16 at the MJD 58883.

11.2.3 B1508+55

Observation at MJD 58887

B1508+55 shows pulses with a single component, and the six major pulses in Figure 11.5 are quite similar. It is indeed a pulsar with two very close components, which can be seen by very faint variations between the left and the right of the width of each pulse.

Contrary to B1133+16, the pulses appear in a smaller bandwidth from about 40 to 80 MHz. For this pulsar, it is therefore required to measure on eight subbands, to try to obtain at least three points in the middle frequencies. In addition, the two high subbands of pulse 83 contained many flagged frequency channels. Finally, only three points are consistent between the measure with six and eight subbands. In

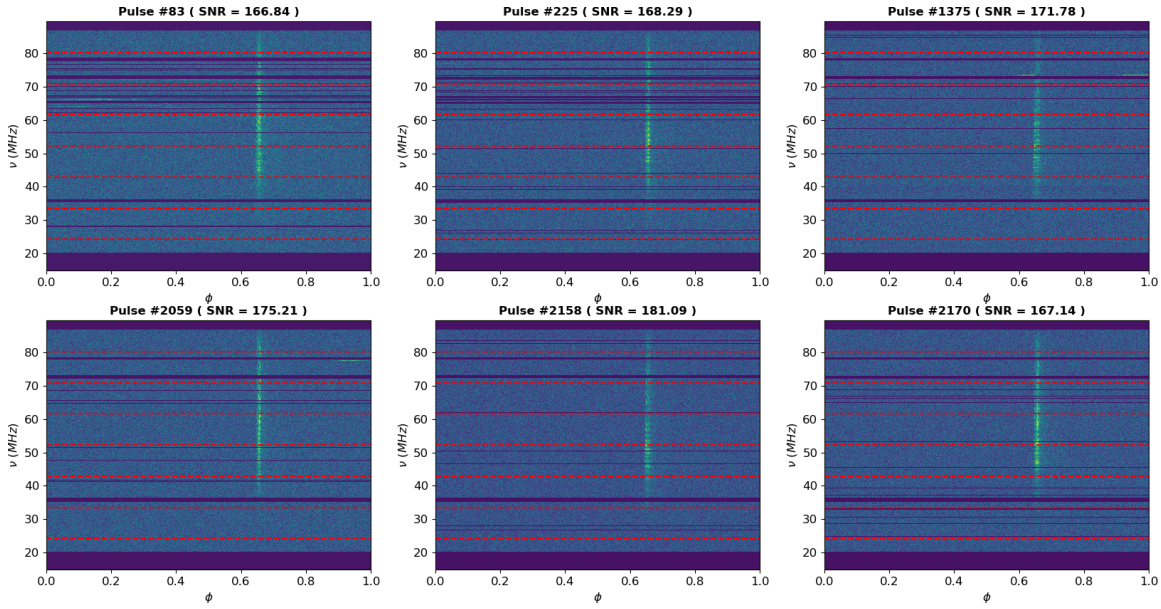


Figure 11.5: Dynamic spectra of the six most intense single pulses in the observation of B1508+55 at the MJD 58887.

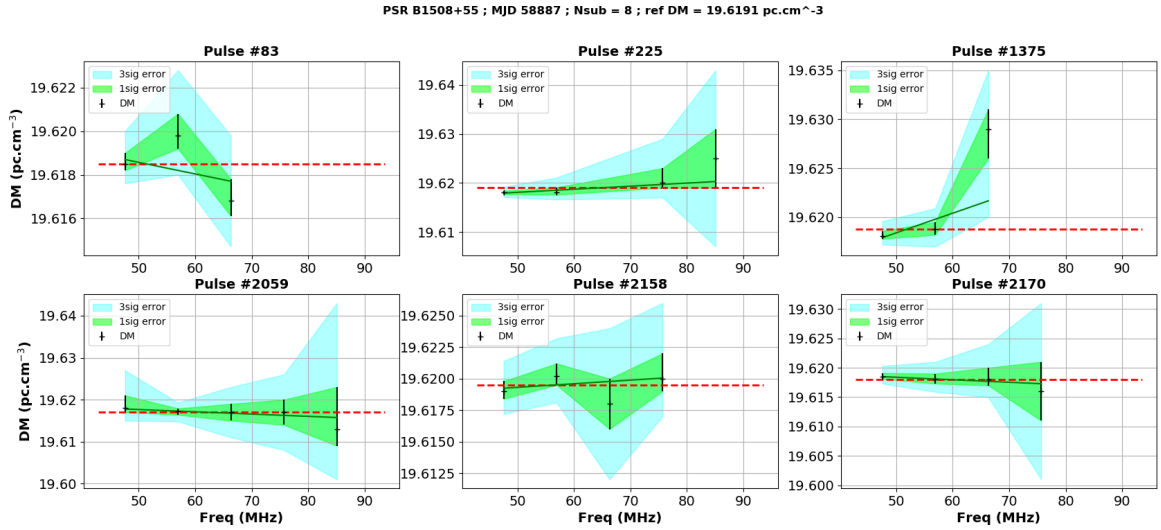


Figure 11.6: DM measurements of the different subbands for the six most intense pulses of the observation of B1508+55 at the MJD 58887.

high frequencies, pulse 1 375 shows a weak flux, with a remaining RFI located over the pulse, leading to the removal of these two highest subbands.

In Figure 11.6, one can see that pulses 83 and 1 375 present variations of the DM values around the mean DM, which are not consistent at 1σ with a constant DM. However, for pulse 83, these variations are largely consistent at 3σ. Concerning pulse 1 375, the subbands around 65 MHz are not consistent with a constant DM. For this point, it is noticed that the difference between DM values with six and eight subbands is not consistent with their 1σ uncertainty ranges. However, their 1σ ranges are very close and are then largely consistent within their 3σ uncertainty ranges. In addition, with six or eight subbands, the measurements reflect the same trend, with a DM value not consistent with a constant DM. Hence, this point has been all the same kept.

The four other pulses have DM values consistent with a constant DM. For pulse 225, we can notice a slightly decreasing trend. For the two highest subbands, the DM values are consistent with the constant DM only on the lower edge of the 1σ uncertainty range. At the opposite side of the bandwidth, the DM value at 57 MHz is also only consistent at 3σ with a constant DM.

Observation at MJD 58944

For the second observation of B1508+55, the dynamic spectra of the six major pulses in Figure 11.7 appear in the same bandwidth that for the previous observation, i.e. between 40 and 80 MHz. The first point to notice is an important variation of the global DM. This observation is 57 days after the previous one, and the mean DMs are significantly lower compared to the first one.

Another difference is this observation is cleaner, with especially fewer flagged channels above 70 MHz. As a consequence, compared to the first observation, it is possible to keep safe the two highest subbands. Hence, by comparing the DM values between six and eight subbands, the majority of points have been kept as valid measures, except for pulses 1021 and 1234.

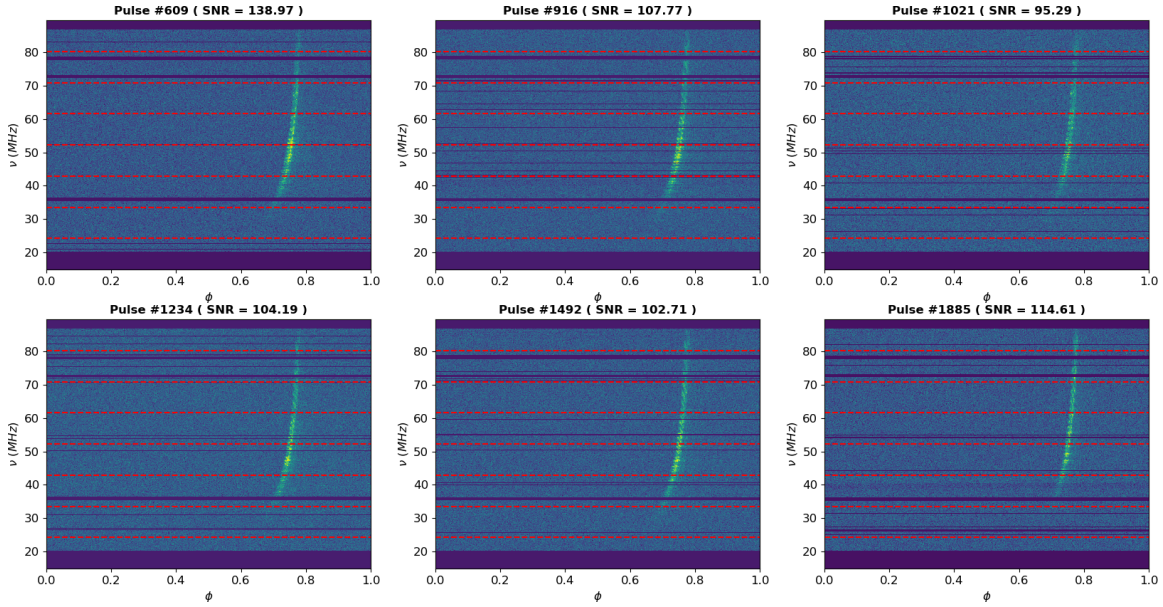


Figure 11.7: Dynamic spectra of the six most intense single pulses in the observation of B1508+55 at the MJD 58944.

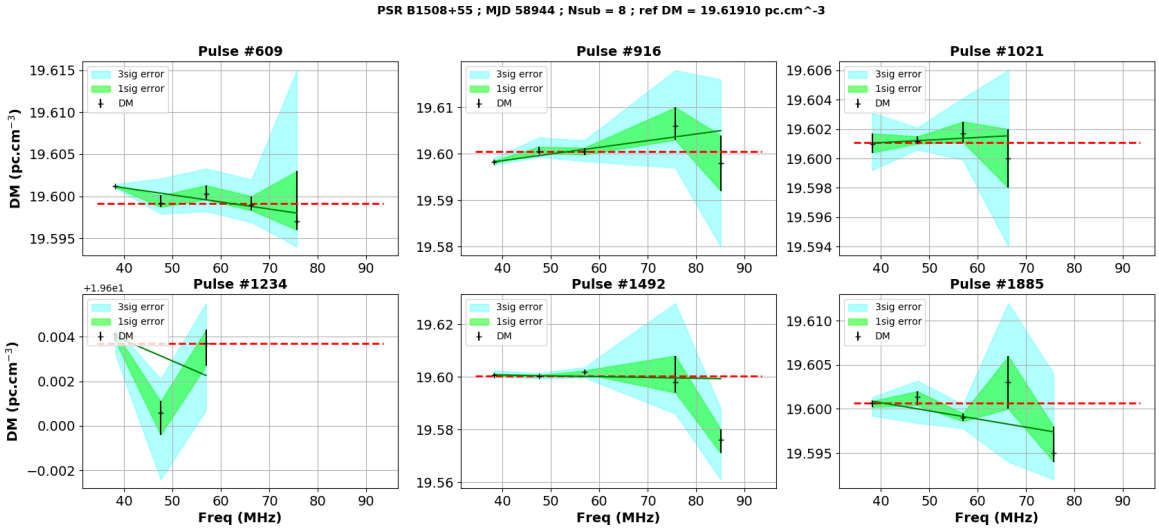


Figure 11.8: DM measurements of the different subbands for the six most intense pulses of the observation of B1508+55 at the MJD 58944. (The DM values in the graph of the pulse 1234 are variations around 19.6000 pc.cm^{-3}.)

Pulse 1021 is the only pulse where the two highest subbands have been removed. This is due to a weak flux of the pulse and a substantial part of the flagged subband. For pulse 1234, between six and eight subbands, it presents variations around the mean DM, but without any common trend. Moreover, the dispersion of the measures using eight subbands is larger than with six subbands and not really consistent with each other. As a consequence, only three points, corresponding to the middle band where the flux is maximum, are kept. However, the large dispersion of the values compared to the other pulse leads to thinking that the noise is relatively important or even dominant for this pulse.

In Figure 11.8, there are slight variations of the values around the mean DM, seeming to follow trends with a low slope. However, the majority of the points are consistent with a constant DM.

11.2.4 B1919+21

B1919+21 presents pulses narrower than the other studied pulsars. For every pulse in Figure 11.9, it is possible to identify several components. Also, the flux of the pulse appears in different bandwidths, seeming to be essentially located in high frequency between 60 and 80 MHz. Globally, below 40 MHz, there is no or very weak flux, resulting in five measurements for the best cases.

Pulse 1264 has the majority of its flux located between 65 and 80 MHz. Measurements with six and eight subbands return non-consistent results, leading to keeping only two points of measure having a similar DM.

Although there is flux in the high subbands, the corresponding measurements for pulse 739 are not consistent between six and eight subbands and have been removed.

For the other pulses, the measurement points in Figure 11.10 are almost all consistent at 1σ with a constant DM, except the lowest frequency point of pulses 1265 and 2054.

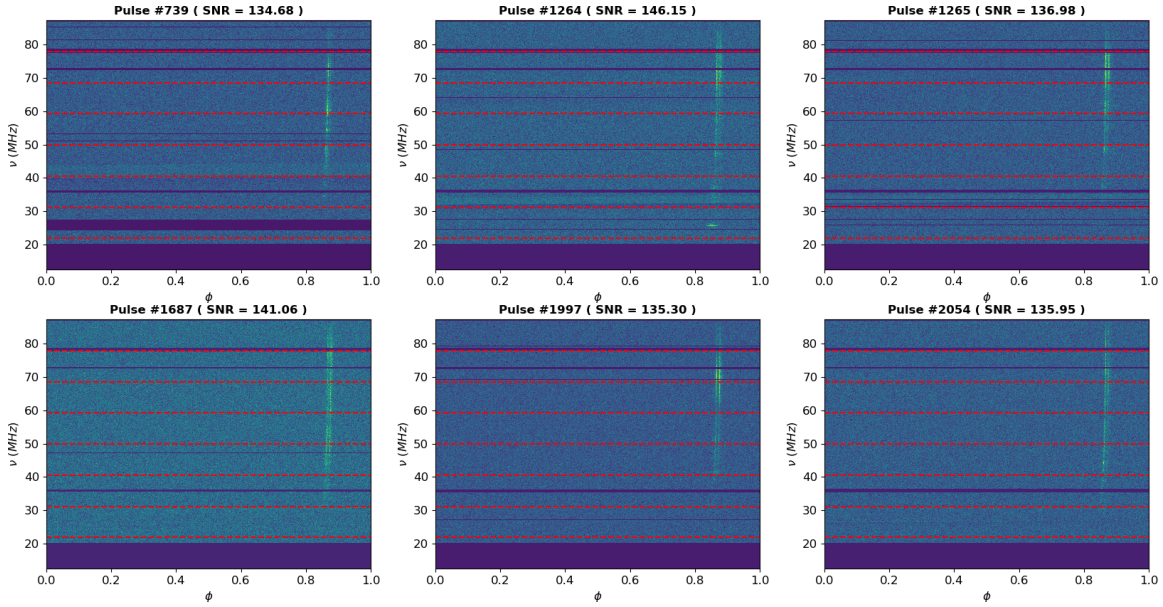


Figure 11.9: Dynamic spectra of the six most intense single pulses in the observation of B1919+21 at the MJD 59044.

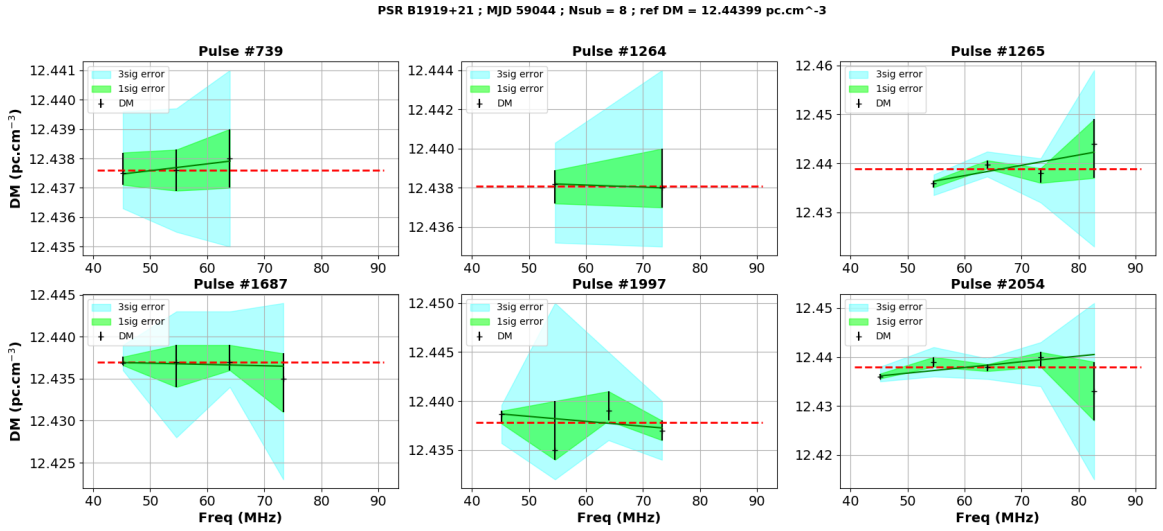


Figure 11.10: DM measurements of the different subbands for the six most intense pulses of the observation of B1919+21 at the MJD 59044.

11.3 Combined measurements

With the aim of trying to detect a potential global trend, the measured DM values have been compared for each pulsar. However, the DM values of the pulses are different for each pulse. As a result, these are the DM deviations that have been combined. Each deviation has been calculated relative to the mean DM of the corresponding single pulse. For each subband, the weighted average DM deviation (taking every individual uncertainty into account) and the simple average DM deviation are then determined. An average uncertainty is also computed as the simple mean of the uncertainties of the points for this frequency.

Figure 11.11 presents the result of the mean DM deviations for the five studied observations. The averaged deviations are indicated in dark squared, and the simple average deviations are in green triangles. The salmon contour represents the average uncertainty. A linear global trend and a second-order trend are subsequently fitted and drawn with the blue line for the first-order one and the red line for the second-order. The two fitted trends are weighted with the average uncertainty previously determined.

The first global comment is that, for most of the points, the average DM deviation is consistent with a DM deviation of 0, i.e. a constant DM. For observations of B1508+55 at MJD 58887 and B1919+21, all the points are even consistent with a constant DM. Concerning B1133+16, only the point at the lowest frequency is not consistent, with an average DM deviation slightly negative. For the observation of B1508+55 at MJD 58944, there is only the highest subband which shows a large negative average DM deviation caused by the pulse 1492 (see Figure 11.8). Nevertheless, the measures at 85 MHz for six, eight, and twelve subbands obtain similar values, leading to keeping this point as a valid measure. The last observation is B0950+08, where all the points are not consistent with a constant DM. This observation will be discussed later in more detail.

The second comment is that, although the majority of points are consistent with a constant DM (except for B0950+08), we can notice that a global linear trend of the average values different from a line at 0 pc.cm^{-3} can be determined. Also, except for the observation of B1508+55 at MJD 58944, the second-order fit deviates a little compared to the linear fit. Furthermore, this observation excepted, all the linear trends have a slope (designates as α in Table 11.2) similar in sign and absolute value.

All these linear trends are similar, but, according to the associated uncertainties, are also completely consistent with the constant DM. Thereby, maybe it will be interesting to expand the study in order to see if this linear trend can still be retrieved.

Table 11.2: Coefficients of the linear fits of the combined DM deviation relative to the frequency for the five studied observations, such as $\Delta DM = \alpha \cdot \nu + \beta$ with ν the frequency.

Observation	$\boldsymbol{\alpha}$ ($\times 10^{-5}$ pc.cm^{-3}.MHz^{-1})	$\boldsymbol{\beta}$ ($\times 10^{-3}$ pc.cm^{-3})
B0950+08	5 $\pm$2	-1.9 $\pm$0.5
B1133+16	2.3 $\pm$0.8	-1.0 $\pm$0.3
B1508+55 (MJD 58887)	4 $\pm$2	-2.0 $\pm$0.8
B1508+55 (MJD 58944)	-4 $\pm$6	2 $\pm$3
B1919+21	3 $\pm$3	-2 $\pm$2

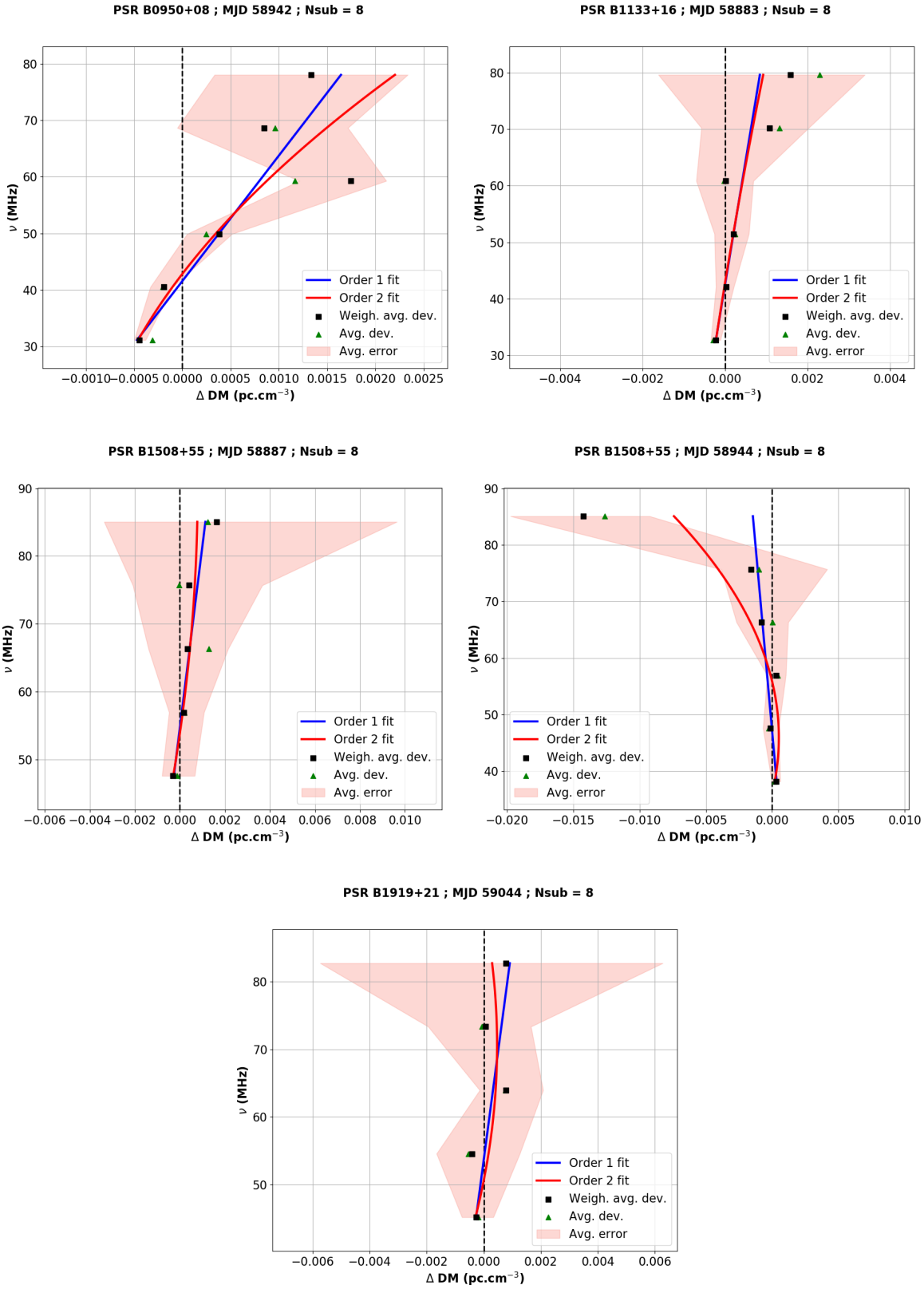


Figure 11.11: Combined measures of the DM deviations for the 5 observations divided into 8 subbands. The black squares correspond to the weighted average DM deviations, and the green triangles to the simple average DM deviation. The salmon contour represents the average uncertainty. The two lines show the weighted fit of the weighted DM deviations, with a linear fit in blue and an order two polynomial fit in red.

11.4 Specific case of B0950+08

11.4.1 Detail of pulse 1920

Measurements realized on B0950+08 are the only ones to show an important deviation relative to the constant DM. It is also the only pulsar seeming to present a better fit of the combined measure with a second-order law rather than a linear law. Furthermore, among all the studied pulses, pulse 1920 of B0950+08 (see Figure 11.2), is particular because it clearly doesn't follow the constant DM. Contrary to the majority of the other pulses with DM values largely different from the constant DM, this pulse was unremoved for different reasons.

The clear decreasing trend of the DM values towards low frequencies observed in Figure 11.2 with eight subbands is indeed equally seen with six and twelve subbands. Figure 11.12 shows the DM values measured for pulse 1920 using six, eight, and twelve subbands. The measurements at frequencies above 70 MHz have large uncertainties and don't show consistent values between the different used numbers of subbands. Finally by comparison, only points in agreement between the three measures are kept, and the higher frequency points are consequently removed in the graphs of Figure 11.12.

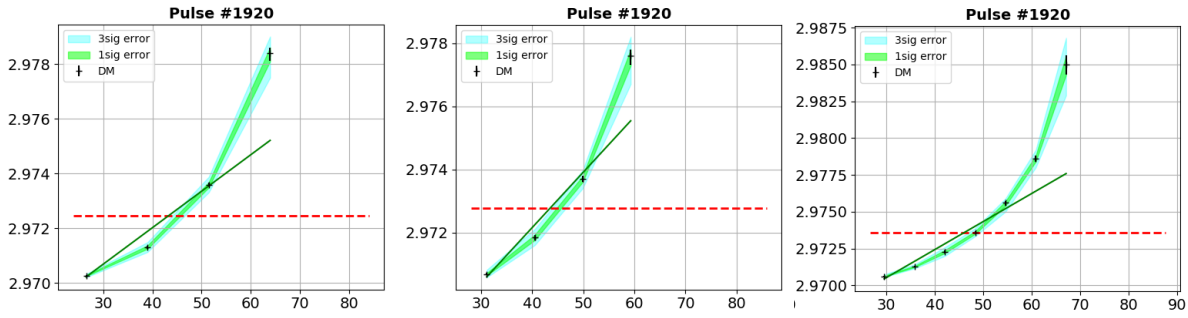


Figure 11.12: DM measurements for pulse 1920 for three different numbers of frequency subbands. Left: 6 subbands. Middle: 8 subbands. Right: 12 subbands.

For the three different configurations, the trend is remarkably similar, with very close DM values. Whatever the number of subbands used for the measure, we obtain the same results, with a parabolic shape, particularly visible with twelve subbands (right graph in Figure 11.12).

In the case of measurements dominated by a standard random noise, the obtained DM value in a subband is independent of the obtained values in the other subbands. We expect therefore to obtain random measures without continuity along the bandwidth. However, the DM values in Figure 11.12 seem to follow a continuous structure along a large part of the bandwidth, resulting in the rejection of noise as the source of the measured values. Here, the source must be a broadband structure, and due to the relatively small uncertainties, provide precise DM measures close to the DM value of the pulsar.

Moreover, the corresponding dynamic spectrum (see Figure 11.1) seems to be clean, without any remaining RFIs. Also, except for the pulse, no evident broadband pattern able to strongly impact the DM measurements can be seen.

Also, comparing with the other pulses of the observation, we can notice that the DM value obtained for pulse 1920 is greater than other pulses. The five other pulses have DM values between about 2.968 and 2.970 pc.cm^{-3}, while pulse 1920 obtains a minimum value of about 2.971 pc.cm^{-3}. However, in the dynamic spectrum of pulse 1920 in Figure 11.1, we can see that this pulse is visible at frequencies until about 25 MHz, which is lower than other pulses. Also, the pulse at the lowest frequencies seems to be slightly misaligned for the standard DM value of 2.96927 pc.cm^{-3}, with a curvature of the pulse towards the later times.

As a consequence, there are, a priori, no reasons to consider that the DM measurements of pulse 1920 are caused by another source than the pulsar, and thus to reject this pulse of the present study.

11.4.2 Combined measurement without pulse 1920

Pulse 1920 shows clear deviations to the constant DM, and the combined measure of B0950+08 in Figure 11.11 presents also a large DM deviation. Also, the sign of both deviations is similar. Then, I wanted to check if the deviation of the combined measure is essentially due to pulse 1920. The combined measure has therefore been computed with and without pulse 1920. The two resulting plots can be compared in Figure 11.13, where the left panel is the initial combined measure with pulse 1920 and the right panel is without it.

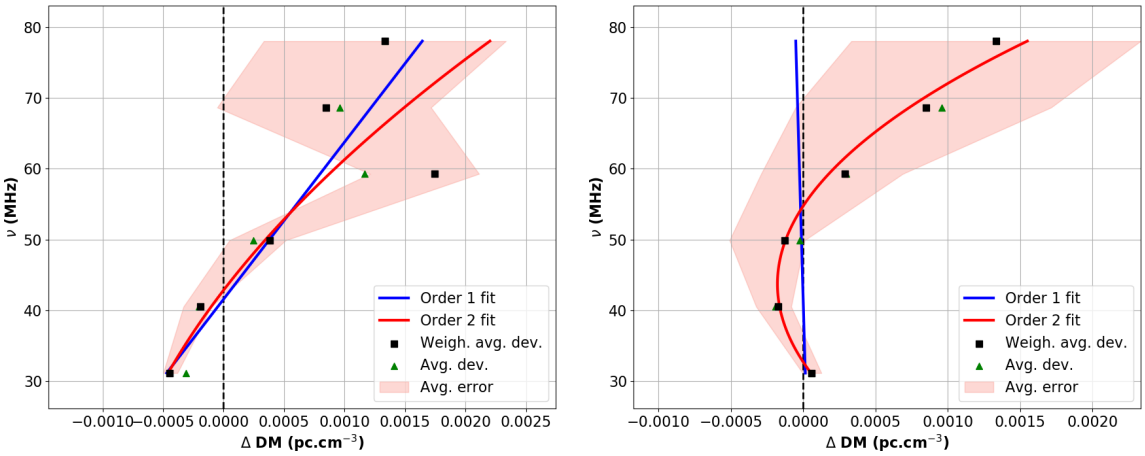


Figure 11.13: Comparison of the impact of pulse 1920 in the combined measure for B0950+08. Left: with pulse 1920 (same as in Figure 11.13). Right: without pulse 1920.

In the combined measures without pulse 1920, we can note that: firstly, other pulses equally deviate from the constant DM; secondly, the global trend equally doesn't match with a linear trend. Hence, with or without pulse 1920, the combined measure is better fitted with a second-order polynomial function. Moreover, without pulse 1920, the parabolic trend is even more remarkable.

Looking at measurements of each pulse in Figure 11.2, all pulses (except pulse 4322) effectively present DM values around $40 - 50$ MHz which are lower than at lower and higher frequencies. In fact, with its completely different behavior, pulse 1920 slightly compensates for DM deviations in the middle band and increases at low frequency. Ultimately, that leads to making the global trend more linear. As a result, the behavior of other pulses reveals a hollow in the DM values around $40 - 50$ MHz, consistent with the not linear global trend of B0950+08.

11.5 Discussion of results

11.5.1 Constant DM with small deviations

Globally, the measurements performed in this study are consistent with a constant DM. However, because they are carried out on subbands of single pulses, the uncertainties are important relative to the order of magnitude of potential super-dispersive effects. Consequently, we can notice there are no deviations to the standard dispersion law which may be visible at the usual uncertainties (for other types of DM measurements), even within single pulses. Nonetheless, in view of the uncertainties, this result was expected and is consistent with the expected order of magnitude of these effects.

Furthermore, for each observation, the measured DM values of the pulses are not identical along the bandwidth, resulting in a distribution of the DM deviations with a certain statistical dispersion. The combination of the DM deviations allows trying to identify eventual trends within this dispersion. According to the average uncertainties, the combined deviations of four of the five observations are consistent with a null deviation.

Nevertheless, one can notice that the weighted fits of the average DM deviations seem to follow linear trends. Moreover, for three over five observations, the linear trend is similar in absolute value and sign. Thereby, in order to improve the search, this behavior could be a clue of potential very weak deviation from the standard law, requiring a study with reduced uncertainties.

However, the observation of B0950+08 is different, with a significant deviation from the null deviation. In addition, contrary to other observations, without pulse 1920, the combined deviations unambiguously follow a parabolic trend. The second case of difference concerns the two observations of B1508+55, which show linear trends but are opposed in sign. As a consequence, although there would be a potential global trend, it exists of variations between pulsars and even between different observations of the same pulsar.

As a consequence, if we allow ourselves to consider them as possible deviations to a constant DM, we can then identify two elements. The first element is the deviations have a linear part, represented by the linear trends seen in the combined measures, and a non-linear part represented by the small statistical dispersion of the measures. The second element is that these deviations occur at different scales, from variations between observations to variations between pulses within a single observation.

11.5.2 Frequency drift in the observation of B0950+08

As described in Section 11.4, a completely different behavior have been seen in the observation of B0950+08. The combined measure presents a clear parabolic trend of the DM deviations compared to the mean DM, and pulse 1920 equally reveals a clear parabolic trend of the measured DMs. It results from these trends B0950+08 has a smooth frequency drift of its DM.

A similar behavior can be seen for some FRBs (fast radio bursts, Lorimer et al. 2007), whose the subbursts within the burst seem to have different DMs (Hessels et al. 2019). In the case of FRBs, the subbursts are slighlty delayed in time and shifted in frequency. Consequently, in a certain bandwidth, the alignement of the global burst is reach for different DM values, leading to an apparent DM drift in frequency. In addition to the time delay between subbursts, there are also real differences in the DM of subbursts, due to various location of the emission in the magnetosphere of the FRB. This frequency drift of the DM is called "sad trombone" effect.

Nevertheless, the alignement of microstructures drifted by the sad trombone effect produces an increase of the apparent DM towards the low frequencies. However, for B0950+08 here, whatever the combined measure or the values measured for pulse 1920, the DM values decrease towards low frequencies.

In the standard current theories to explain the sad trombone effect (see for example Wang et al. 2019; Rajabi et al. 2020)), the inverse effect, called "happy trombone", can also be generated. Moreover, such a drift have already been seen in some microstructure in the pulse of B0950+08. At extremely low-frequency using UTR-2, Ulyanov et al. (2016) have detected subpulse where the low frequency arrives earlier than the subpulse at higher frequency. To be consistent with the present observations, this behavior needs to be expanded from the frequencies of UTR-2 at about 20 MHz to those of NenuFAR up to about 80 MHz.

In additions, they have shown these microstructures could be aligned for two different DMs. These DMs were very close, differing of 1.10^{-3} pc.cm^{-3}, with the interesting particularity that low frequency microstructures had a lower DM than high frequency microstructure. This decrease of the DM towards low frequencies is then similar to the trend measured in this work.

Indeed, the trend of the combined measure is not monotonic, showing an increase below 40 MHz. Also, one can notice for every pulses in Figure 11.1 that the flux of the pulse completely disappears below 30 MHz. As a result, in the context of the alignement of several drifted microstructures, the lowest subband may comprise solely one microstructure. Thus, contrary to the subbands at higher frequencies, a not frequency-drifted DM could be obtained.

B0950+08 is the only pulsar of the four revealing such a behavior in DM. Also, for B0950+08, the scintillation observable in Figure 11.1 is particularly strong with a large scintillation bandwidth, leading to be one of few pulsars with frequency scintillation visible by NenuFAR. As a consequence, the possibility to detect these DM deviations may potentially attributed to the scattering screen close from Earth causing this important scintillation.

Furthermore, the frequency drift of the DM can be merely detected by the combined measure, except for pulse 1920. The drift in this pulse can be easily identified without combination of the measures. Therefore, in the case where the scattering screen would be the source of the drift, the specific behavior of pulse 1920 indicates a short timescale substantially shorter than the duration of the observation. Such a short timescale is equally consistent with a close and turbulent scattering screen.

11.5.3 More precise definition of the DM

The standard way to express the DM is based on the time of propagation of the electromagnetic wave from the pulsar to Earth (see Chapter 1.3.1). The linear form to define the time of propagation δt to reach Earth is:

$$\delta t(z_E, \nu) \;=\; t_0 + \frac{z_E}{c} + \frac{\mathcal{D}}{\nu^2} \cdot DM \;=\; t_0 + \frac{z_E}{c} + \frac{\mathcal{D}}{\nu^2} \cdot \int_0^{z_E} n_e(z) \cdot dz \tag{11.1}$$

With t_0 the starting time, $\mathcal{D}$ the dispersion constant, ν the frequency of the electromagnetic wave, $n_e(z)$ the electron density, and z the position in the line of sight. Here, z_E designates the distance between the pulsar and Earth.

However, if we consider the slight variations at short timescales of the electron density, it needs to add the time dimension to the electron density. One can define a density field $n_e(z,t)$ with two dimensions: the position in the line of sight z, and the time to cross this position t. In this context, the dispersion cannot be defined as a simple integration in space of the electron density. To evaluate the dispersion and then obtain the time of propagation, it needs to calculate the line integral of the density field over a path of propagation within this field.

$$\delta t(z_E, t_E, \nu) \;=\; t_0 + \frac{z_E}{c} + \frac{\mathcal{D}}{\nu^2} \cdot \int_0^{z_E} \int_{t_0}^{t_E} \{\; n_e(z,t) \circ \gamma(z,t,\nu) \;\} \cdot dz \cdot dt \tag{11.2}$$

Here, $\gamma(z,t,\nu)$ represents the path of propagation of the electromagnetic wave within the density field, and t_E designates the time to reach Earth. By this definition, a given position z is reached at the time t. However, this time is also a function of the dispersion undergone before, which is a function of the starting time t_0, the frequency ν, and of the previously crossed density field. As a consequence, the path γ is related to the frequency of the electromagnetic wave, resulting in the possibility to obtain slightly different times of propagation between different frequencies.

11.5.4 Frequency dependence of the DM measurement

The DM is not directly measured by the undergone dispersion but by the alignment of the times to obtain the same δt for two different frequencies. Also, in the standard context, the integral of the electron density is unique, i.e. independent of the frequency, and the DM can be determined using the usual relation. However, the result of the line integral of the density field, which we can designate as $\Phi(\nu)$, is dependent on the path. In the case of a single pulse, the starting time is the same, and the variations of the density field are then only frequency dependent. In the usual way, the determination of the DM between the frequencies ν_1 and ν_2 such as $\nu_1 \;<\; \nu_2$ gives the following relation:

$$\delta t(\nu_1) - \delta t(\nu_2) \;=\; \mathcal{D} \cdot DM \cdot (\; \nu_1^{-2} - \nu_2^{-2} \;) \;=\; \mathcal{D} \cdot \left(\; \frac{\Phi(\nu_1)}{\nu_1^2} - \frac{\Phi(\nu_2)}{\nu_2^2} \; \right) \tag{11.3}$$

The obtained DM is therefore dependent on the choice of the two frequencies used to perform the measurement:

$$DM \;=\; DM(\nu_1, \nu_2) \;=\; \frac{\Phi(\nu_1) \cdot \nu_1^{-2} \;-\; \Phi(\nu_2) \cdot \nu_2^{-2}}{\nu_1^{-2} \;-\; \nu_2^{-2}} \tag{11.4}$$

Based on this relation, we can define two types of measurement. In the case of the measurement relative to a fiducial frequency, one of the two frequencies is fixed, e.g. the frequency ν_2 as the highest frequency of the observation, resulting to obtain a “global" measurement of the DM. In this study, the measurement is carried out within a subband, independently of the other subbands. The two frequencies are consequently unfixed, leading to obtaining a “local" measurement, which is different from the “global" measurement.

11.5.5 Simulations of propagation

Based on Equation (11.2), it is possible to simulate the times of propagation caused by a density field with space and time variations. This time of propagation is computed by performing the numerical integration of the associated differential equation of propagation.

$$d\tau - d\zeta \cdot \left(1 + \frac{\lambda \cdot \rho(\zeta, \tau)}{\mu^2} \right) = 0 \tag{11.5}$$

Here, $\tau = c \cdot (\delta t - t_0)/z_E$ is the time variable, $\zeta = z/z_E$ is the space variable, $\rho(\zeta, \tau) = n_e(\zeta, \tau)/\langle n_e \rangle$ represents the variation of electron density relative to the mean density $\langle n_e \rangle$, $\mu = \nu/\nu_0$ corresponds to the ratio of frequency relative to the reference frequency ν_0, and $\lambda = c \cdot \mathcal{D} \cdot \langle n_e \rangle / \nu_0^2$ is a reference constant.

Figure 11.14 presents a space-time diagram representing the propagation of six different frequencies, from the highest frequency in purple to the lowest frequency in red. These frequencies correspond to a single pulse emitted at $\tau = 0$ and $\zeta = 0$, and arriving on Earth in $\zeta = 1$. The whole plane is filled with a density field in light grey. In this case, an over-density is added in a localized region, represented by the dark grey rectangle around $\zeta = 0.8$. The over-density impact here thus only three frequencies, which are in addition differently impacted.

Figure 11.15 presents the measurement associated with the DM deviation ΔDM relative to the frequency ratio μ for this case of propagation. These DM deviations are computed by relating the usual DM with the difference in the times of propagation between two frequencies, using the standard relation of dispersion. The mean DM is then subtracted to obtain a DM deviation.

$$\tau(\ \zeta = 1, \mu_1\ ) - \tau(\ \zeta = 1, \mu_2\ ) = \lambda \cdot \frac{DM}{DM_0} \cdot \left(\frac{1}{\mu_1^2} - \frac{1}{\mu_2^2} \right) \tag{11.6}$$

The result on the left is the “global DM", calculated relative to a reference frequency, fixed here to the highest frequency. The result on the right is the “local DM" calculated between the two closest frequencies. Of the two types of measurement, there are differences in the obtained DM deviations between high and low frequencies. The partial impact on some frequencies creates high DM deviations at middle frequencies, with finally a decrease of the deviations at the lowest frequency.

This behavior can be retrieved in some measurements of the observed single pulses of this study, as the pulses 348 and 475 of B1133+16 in Figure 11.4, where a sort of “hollow" in the measured DM occurs for frequencies around 60 MHz (thus opposite in sign with the simulation). Although the DM deviations of these pulses are small, it is possible to fit the measured DM deviations with a linear law, showing a slight deviation to a constant DM (but nevertheless consistent with the null deviation).

A second case with a different over-density is presented in Figure 11.16. Here, the over-density is localized later in time and is longer. The lowest frequency is then totally impacted, and the second lowest frequency (yellow line in Figure 11.16) is lower impacted.

In this case, the measurements of the DM deviations, shown in Figure 11.17, present a parabolic shape. This type of behavior is expected in the context of a super-dispersion as studied by Shitov & Malofeev (1985) and Kuzmin (1986), with a clear chromatic DM. As opposed to the previous case where the DM deviations are not monotonic in frequency, the deviations can be modeled by a monotonic power law reflecting directly the frequency dependence.

These behaviors can be related to the measurements of the observation of B0950+08. The measured DM for the pulse 1920 (see Figure 11.12) follows a parabolic shape. For the combined measure of B0950+08, especially the case without the pulse 1920 (see the right panel in Figure 11.13), the DM deviations are better fitted with a parabolic law similar to the parabolic fit in Figure 11.17 (with also an opposite sign).

As a consequence, with simple localized over-density or under-density, it seems that it is potentially possible to generate some variations of the DM relative to the frequency similar to those seen in this study, or in previous works. Nevertheless, these simulations are simple and need to be complexified and enlarged, in order to try to reproduce more diverse behaviors, and also with more precision.

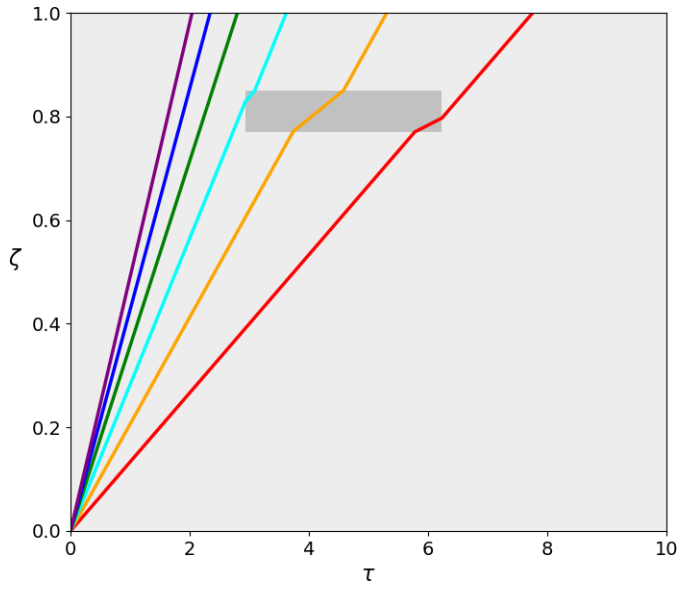


Figure 11.14: Propagation in space (ζ designates the relative position in the line of sight) and time (τ is the normalized time of propagation) of the electromagnetic wave for six frequencies from the highest in purple to the lowest in red. The light grey background represents the mean electron density field and the dark grey area represents a small localized over-density.

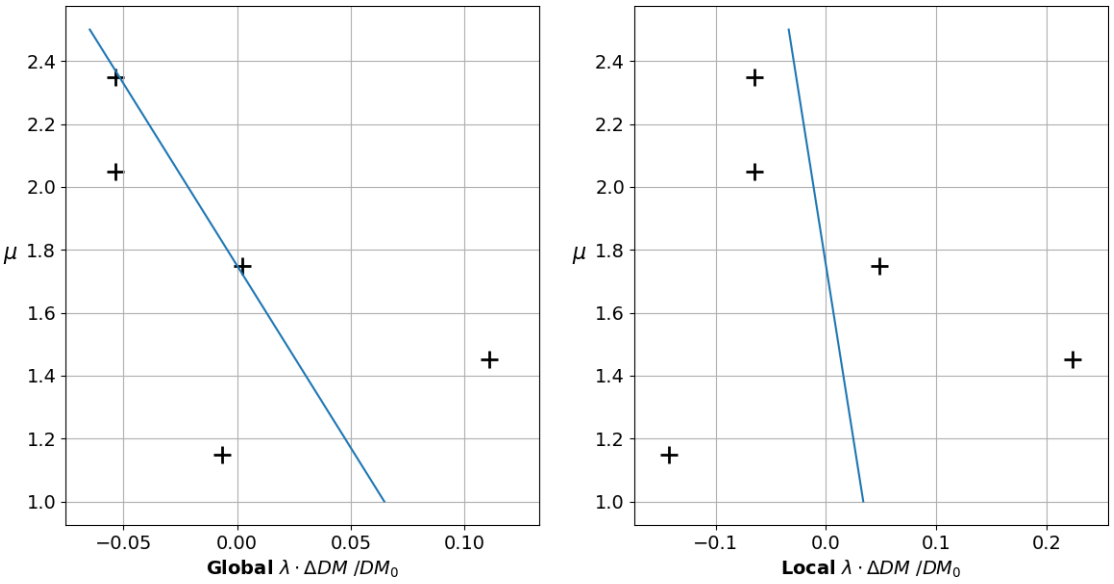


Figure 11.15: Normalized DM deviations in the case of a small localized over-density region. Left: “global" measurement of the DM deviation calculated relative to the highest frequency ratio μ. Right: “local" measurement of the DM deviation calculated between the two closest frequency ratios μ.

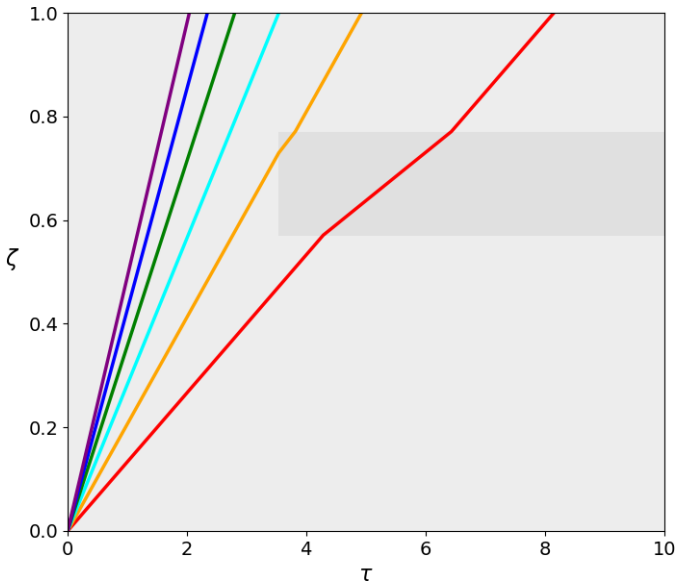


Figure 11.16: Propagation in space and time of the electromagnetic wave for 6 frequencies from the highest in purple to the lowest in red. The light grey background is the mean electron density field, and the dark grey area represents a later and wide over-density.

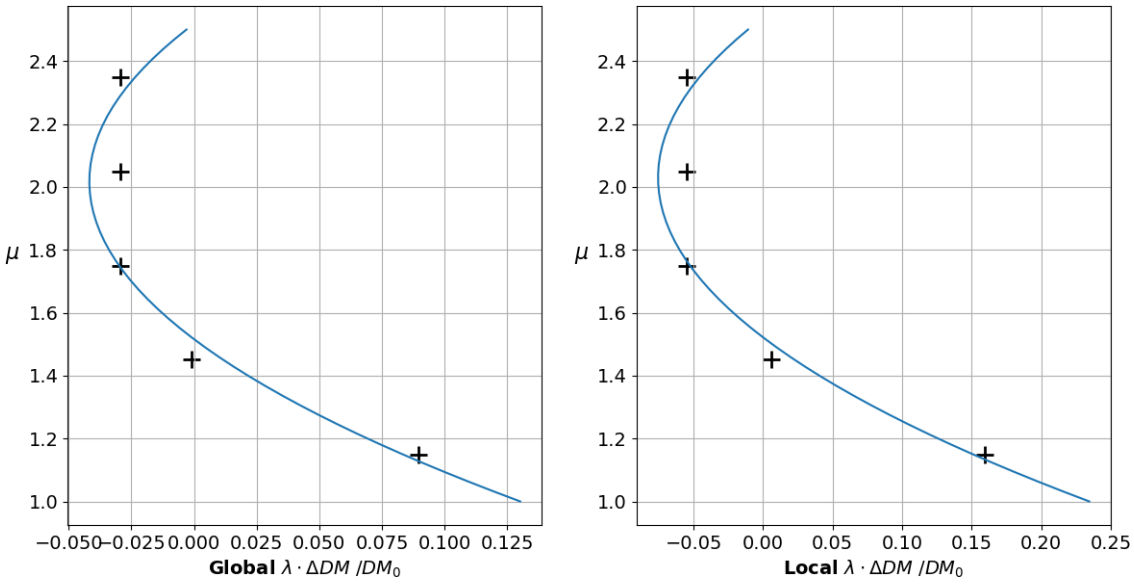


Figure 11.17: Normalized DM deviations in the case of a wide later over-density region. Left: “global" measurement of the DM deviation calculated relative to the highest frequency ratio μ. Right: “local" measurement of the DM deviation calculated between the two closest frequency ratios μ.

11.5.6 Second-order of the dispersion relation

Although the previous propositions are potentially more precise than the usual definition of the dispersion, there are, as for the usual dispersion, based on the order one of the dispersion relation of the propagation of an electromagnetic wave in a cold plasma. Moreover, if we allow ourselves to consider these small variations as possible, they have two characteristics which are: a linear part combined with a non-linear part, and a smaller amplitude than the usual DM variations, as those generated by the variations of the scattering screen (Hemberger & Stinebring 2008; Cordes et al. 2016). However, the second order of the dispersion relation permits to have these behaviors. If we take the second order of the field perturbations into account in the Navier-Stokes equation, we obtain the following vector equation:

$$\{ (n_0 + n_1) \cdot (-i)\omega + n_0 \cdot ik_i \cdot u_1^i \} \cdot u_1^j + n_0 \cdot (-i)\omega \cdot u_2^j = -\frac{e \cdot (n_0 + n_1)}{m_e} \cdot E_1^j - \frac{e \cdot n_0}{m_e} \cdot E_2^j \quad (11.7)$$

Here, n_0 is the mean density field and n_1 the perturbation of order one, u_1 and u_2 are the perturbations of the velocity field, $\omega = 2\pi\nu$ is the pulsation of the wave (with ν the associated frequency), k_i are the components of the wave vector, E_1 and E_2 are the perturbations of the electric field, e is the electron charge, and m_e is the electron mass. The Einstein summation convention is used for the scalar product $k_i \cdot u^i$.

If we neglect the perturbations of order two of the fields n_2 and E_2, we can determine the first order of the current, which is of the form:

$$\vec{j_1} = -e \cdot (n_0 + n_1) \cdot \vec{u}_1 = i\frac{e^2 \cdot (n_0 + n_1)}{m_e \cdot \omega} \cdot \left[1 - \frac{n_0}{(n_0 + n_1)} \cdot \frac{k_i \cdot u_1^i}{\omega} \right]^{-1} \cdot \vec{E}_1 \quad (11.8)$$

Based on this current, the associated dispersion relation through the second order for the transverse wave finally allows giving the propagation time:

$$t = \frac{1}{c} \int_0^{z_E} dz \cdot \sqrt[-1]{1 - \frac{e^2 \cdot (n_0 + n_1)}{\epsilon_0 \cdot m_e \cdot \omega^2} \cdot \left[1 - \frac{n_0}{(n_0 + n_1)} \cdot \frac{k_i \cdot u_1^i}{\omega} \right]^{-1}} \quad (11.9)$$

Here, c is the speed of light, and ϵ_0 is the dielectric vacuum permittivity. Using the usual asymptotic development at infinity, we finally obtain the second order of the dispersion relation that we can compare to the standard form. The first term is the standard dispersion term, which is a function of the mean density n_0. The second term is the dispersion generated by the perturbation of the density field n_1.

Also, in addition to the standard form in ω^{-2}, the difference is the non-linear term. This term is equally frequency dependent and is caused by the interaction between the velocity field and the wave. As a consequence, the dispersive delays might be slightly modified because of this corrective term:

$$\gamma(n_0, n_1, \omega) = \left[1 - \frac{n_0}{(n_0 + n_1)} \cdot \frac{k_i \cdot u_1^i}{\omega} \right]^{-1} \quad (11.10)$$

The time of propagation, calculated through the second order, is consequently the addition of two major terms to the minimum travel time of the wave:

$$t(\nu) = \frac{z_E}{c} + \mathcal{D} \cdot \int_0^{z_E} \frac{n_0}{\nu^2 \cdot \gamma(n_0, n_1, \nu)} \cdot dz + \mathcal{D} \cdot \int_0^{z_E} \frac{n_1}{\nu^2 \cdot \gamma(n_0, n_1, \nu)} \cdot dz \quad (11.11)$$

Globally, at large scales, the non-linear term should be close to 1, leading to a linear behavior of the dispersion. A combined measure, mixing different pulses, should have a global linear behavior. However, at shorter scales, the impact of the perturbation of the density field, represented by the second term, should increase. In the same way, the influence of the non-linear term should also increase, slightly modifying the frequency dependence. Moreover, the perturbation of the density field n_1 and the non-linear term are space and time-dependent, and vary at short scales, resulting in to match with the previous propositions. As a consequence, these two terms should locally influence the DM measurement.

Résumé du chapitre :

L'étude préliminaire présentée ici est basée sur cinq observations effectuées avec NenuFAR en mode d'observation *single pulse*. Sur les 21 pulsars détectables en mode single pulse, quatre des plus puissants ont été sélectionnés : B0950+08, B1133+16, B1508+55, et B1919+21. L'intérêt de ces pulsars est qu'ils ont un élargissement dû au scattering plutôt faible, et montre différents types de profils. Concernant, deux observations différentes de B1508+55 ont été analysées, dans le but de déceler d'éventuelles différences dans le temps.

Les dispersions ont été mesurées sur des sous-bandes de fréquences devant être suffisamment fines pour obtenir plusieurs points de mesures, tout en étant suffisamment large pour avoir suffisamment de flux. Les mesures ont ainsi été effectuées sur les six impulsions les plus puissantes, puis divisées en six, huit, et douze sous-bandes. Dû au manque de flux dans les plus basses et les plus hautes fréquences d'un certain nombre d'impulsions, il a été difficile d'obtenir un nombre de points suffisant avec six sous-bandes. De même, les mesures obtenues avec douze sous-bandes souffraient également d'un manque de flux. Ainsi donc, les résultats ont été pris sur les observations divisées en huit sous-bandes. Pour les trois mesures, nous sommes censés observer globalement la même tendance le long de la bande de fréquence. En conséquence, les deux autres mesures ont servi à valider la mesure avec huit sous-bandes.

Les données brutes ont été segmentées en sous-intégrations temporelles d'une période du pulsar en utilisant le programme DSPSR. La recherche de RFI a ensuite été réalisée en utilisant le programme NENUPLOT, habituellement utilisé pour nettoyer les observations NenuFAR, et qui est basé sur le programme reconnu COASTGUARD. Après division en sous-bandes, le spectre dynamique initial est dédispersé pour la mesure de dispersion à tester.

B0950+08 possède des composantes du profil larges, et visible sur une large bande de fréquences entre 30 et 80 MHz, et montre une variabilité temporelle de ses impulsions. Les mesures sont en accord avec une dispersion constante, exceptée pour l'impulsion numéro 1920.

B1133+16 possède lui un profil double pics fin, également détectable sur une large gamme de fréquences, et montre une très importante variabilité temporelle que cela soit en matière de nombre de pics visibles ou de bande de fréquences visibles. Les dispersions mesurées sont en accord avec une dispersion constante.

Les impulsions de B1508+55 sont elles très fines, et semblent correspondre à un profil à une seule composante. Entre les deux observations, il est à noter que la dispersion moyenne a significativement changé. Cependant, pour les deux observations, aucune claire déviation à la dispersion constante n'est établie.

Enfin, B1919+21 montre des impulsions dont le flux est essentiellement localisé en haut de bande, ne donnant au final que peu de points de mesures. Là encore, la dispersion constante semble correcte.

De manière à éventuellement mettre en évidence une tendance globale sous-jacente, pour chaque observation, les écarts de mesures de dispersion par rapport à la mesure de dispersion moyenne de l'impulsion ont été combinés fréquence par fréquence. Il est à noter que les valeurs combinées moyennes suivent des tendances linéaires légèrement différent de la dispersion constante. Cependant, excepter pour B0950+08, les mesures combinées avec incertitudes sont en accord avec une dispersion constante.

Les mesures de l'observation de B0950+08 dénotent particulièrement, et plus spécifiquement les mesures liées à l'impulsion numéro 1920. Cette observation semble montrer une très claire tendance parabolique des mesures de dispersion combinées en fonction de la fréquence. Cette même tendance parabolique s'observe également sur les mesures individuelles de l'impulsion numéro 1920. Néanmoins, même en l'absence de cette impulsion particulière, les mesures combinées font toujours état de cette tendance.

Globalement, les résultats de cette étude préliminaire ne montrent aucune déviation par rapport à une dispersion constante en fréquence. Néanmoins, si l'on s'autorise à considérer les petites déviations localisées et les légères tendances linéaires observées, deux éléments peuvent être identifiés concernant d'éventuels effets super-dispersifs. Le premier est le fait que ces déviations soient composées d'une partie linéaire expliquant la tendance globale, et d'une partie non linéaire générant à la dispersion des mesures individuelles. Le deuxième élément est que ces effets doivent apparaître sur de petites échelles de temps et de fréquences.

Partant de cela, nous pouvons noter que la relation de dispersion standard ne considère qu'une ligne de visée fixe avec une intégration directe du contenu en électrons sur celle-ci. Or, dans le cas de variations à courtes échelles, cette définition de la dispersion pourrait être précisée en utilisant plutôt une intégration curviligne prenant en compte le trajet, dans l'espace et dans le temps, de la propagation de l'onde. Dû au délai temporel entre les différentes fréquences, le trajet entre deux ondes de fréquences différentes serait différent, générant au final une mesure de la dispersion dépendante de la fréquence.

En réalisant des simulations simples modélisant un champ de densité dans l'espace et le temps, avec l'adjonction d'une zone de densité différente localisée à la fois dans l'espace et le temps, il est possible de recréer qualitativement les tendances linéaires et paraboliques observées dans les mesures effectuées. Il est aussi possible de montrer de fines variations dans la mesure de dispersion en fonction de la méthode de calcul de la dispersion entre deux fréquences.

Enfin, une autre chose à considérer est le fait que cette relation de dispersion est basée sur l'ordre un de la propagation de l'onde dans un plasma. Ainsi, concernant le fait que ces éventuelles déviations montrent une partie linéaire couplée à une partie non linéaire, si l'on ajoute le second ordre de la propagation de l'onde, il apparaît que celui-ci permet d'inclure un second terme linéaire dépendant des variations à courtes échelles de temps et d'espace de la densité électronique, et également un terme multiplicatif non linéaire de faible amplitude et dépendant de la fréquence. En conséquence, le second ordre de la relation de dispersion disposerait, tout du moins du point de vue qualitatif, des propriétés souhaitées.

Chapter 12

Conclusion and further work

In the second part of this thesis, the goal of the study was to try to identify super-dispersive effects, i.e. deviations to the cold plasma dispersion relation. Because the dispersion time delay increases towards lower frequencies, these effects might be essentially detectable at low frequencies and using a relatively large frequency bandwidth. In this context, the very low frequencies and the important relative bandwidth between 10 and 85 MHz of NenuFAR are thereby interesting.

Moreover, the super-dispersive effects represent deviations occurring at short scales, leading to fine effects compared to the usual dispersion time delay. Furthermore, averaging could cause the erasure of these potential fine deviations. As a consequence, to avoid any averaging, measurements have been performed on directly in the single pulses. Although that increase the difficulty to provide precise measures, the entire information about dispersion is saved.

Detecting super-dispersive effects requires precise DM measurements in multiple frequency subbands of a single pulse. A method for determining the DM has been developed, consisting of a DM search in the Fourier space. The DM search is based on the particular behavior of the Fourier transform of a dispersed pulse. Finally, by trying different DMs around a reference value, the DM value can be obtained by maximization of a dispersion indicator that I have described.

In the study presented here, five observations of four of the most intense pulsars have been selected. For each observation, the measurements have been performed on the six major single pulses of the observation. Finally, to obtain DM values for multiple frequencies, each pulse has been divided into six, eight, and twelve subbands. The consistency of the obtained DM values has been subsequently validated by comparing the results for the three different numbers of subbands.

The resulting measures have shown slight variations of the average DM between pulses within the same observation. Furthermore, inside single pulses, many slight deviations from the average DM have also been observed. Nevertheless, according to uncertainty at 1σ, the majority of the measured DM values are consistent with a DM constant in frequency.

In order to identify whether a trend exists, for each frequency, the DM deviations of the six pulses of the observation have been combined. For four of the five observations, a linear trend seems to appear with a slight slope relative to the null DM deviation. However, according to the average uncertainty at 1σ, the combined measures of all frequencies for these four observations are consistent with a null DM deviation.

The last one is an observation of the pulsar B0950+08. Although measurements in the single pulses seem to be approximately consistent with a constant DM, the combined measure presents a clear deviation to the null DM deviation. Furthermore, contrary to the four others, the trend of the combined measure is not linear but parabolic. Another point of interest in this observation of B0950+08 is the particular

behavior of the pulse number 1920. The measurements on this single pulse show important deviations to the mean DM, with a parabolic decrease of the DM values towards the low frequencies. The DM difference between the highest subband at about 60 MHz and the lowest subband at about 30 MHz is more than 6×10^{-3} pc.cm^{-3}.

B0950+08 is known to possess microstructure within its pulse with different DMs. Moreover, this parabolic trend, equivalent to a frequency-drift of the DM, is similar to the time-frequency drift of subbursts observed for some FRBs. The associated effect, called "sad trombone" is a possible cause of the DM deviations identified in this work. However, the decrease of DM values towards low frequencies seems to be not consistent with a bad alignement due to the sad trombone effect. A more complex cause, as the "happy trombone" or intrinsic differences between microstructure, seems to be required to describe the not monotonic trend seen in this observation.

In order to confirm if the specific behavior of B0950+08 is a characteristic of this pulsar, it is required to perform measurements in other observations of this pulsar. Furthermore, in the present study, only six pulses have been measured. Increasing the number of studied pulses in a unique observation will allow checking if the parabolic trend is an artifact created by these six pulses or not. Finally, the increase in the number of pulses may allow us to provide other single pulses similar to pulse 1920.

Moreover, in every observation, small and localized DM deviations not consistent with a constant DM may be detected. The majority of these points become nevertheless consistent with the 3σ uncertainty. A few points especially at low frequency remain all the same not consistent at 3σ. Currently, the uncertainty on the measured DM value is evaluated only as a function of the noise in the Fourier transform of the dynamic spectrum. To confirm these not consistent DM deviations are real deviations to a constant DM, the uncertainty due to the pulse shape and sampling parameters of the observation must be estimated.

Finally, the amplitudes of the uncertainties are, at least for the high frequencies, too large compared to the estimations of the order of magnitude of the expected super-dispersive effects. To answer this problem, the easiest way is to increase the SNR of the studied single pulses. Since the middle of 2022, the number of mini-arrays of NenuFAR has been increased to 80 (and might be increased to 96 in 2023), allowing to significantly increase the sensitivity of NenuFAR. Therefore, to obtain more precise DM measurements, it will be interesting to perform measurements on new single pulse observations realized with 80 mini-arrays (or even with 96 mini-arrays later).

For many models of super-dispersive effects, the expected deviations are continuous and monotonic, with a perfect chromatic DM following a simple power law. However, although there is a global consistency with a constant DM, if we allow ourselves to look at the slightly obtained DM deviations in this work, we can note there are no monotonic within a single pulse and monotonic for the combined measures. To explain both monotonic and non-monotonic variations, I propose a model based on two assumptions. The first one is to more precisely define the DM by adding the time dependence of the electron density on the line of sight, and consequently the time dependence of the measured DM. The second assumption is to take the second order of the dispersion relation into account.

Concerning the first point, I have performed some simulations of the propagation of the radio wave of a single pulse inside a simulated time-dependent ISM. These simulations are simple but allow for qualitatively retrieving the observed global behaviors. Nevertheless, more complex simulations of ISM are required to quantitatively match the model with observations. According to this model, there should be a temporal coherence of the DM deviations between pulses, with a shift in frequency of a particular DM deviation between neighbor single pulses. To confirm this expectation, measurements on several neighbor single pulses in a single observation are required. However, in the observations studied in the context of this thesis, for the majority of the single pulses, their flux was insufficient to obtain DM values, firstly with sufficient low uncertainties, and secondly in several contiguous pulses. Using NenuFAR observations with 80 mini-arrays could potentially allow the detection of a sufficient number of intense single pulses (in addition to better uncertainties).

Concerning the second point, the second order of the dispersion relation brings two modifications to the usual dispersion relation. The usual relation solely depends on the mean electron density on the line of sight, and then constant in time and space. The first modification is thereby the addition of a second linear term depending on the fluctuations (in time and space) of the electron density. The second modification is the addition, for each of these two terms, of a non-linear factor depending on the frequency.

In order to take this second order into account, one has to perform plasma simulations to estimate the amplitude of the non-linear term. Furthermore, they could allow us to evaluate the time and space scales where these non-linear effects could occur. Such a result could potentially allow us to obtain the impact on the frequency dependence of the dispersion relation.

The consequence of the first point is the creation of variations along the bandwidth and between single pulses. In the context of an analysis utilizing a combination of pulses such as for the timing process, these variations should introduce a statistical dispersion of the DM values within a certain timelapse, resulting in a slight modification of the TOA compared to a model with a fixed DM. The consequence of the second point is the non-linearities induced by the second order of the relation should also generate a statistical dispersion of the DM values, resulting equally in a statistical dispersion of the TOA of the pulses. As a result, obtaining upper limits in the DM deviations potentially caused by the two effects could permit obtaining an associated upper limit in the timing error. Finally, one can also note that the model presented here is a two-dimensional model using only one space dimension. An extension of this model is possible by adding the two transverse space directions in a four-dimensional model. The addition of the transverse directions could allow to include the spatial variations of the path of the electromagnetic wave in the ISM. Such a model could allow for taking the multipath propagation generating the scattering and scintillation phenomena into account.

Résumé du chapitre :

La seconde partie de ce manuscrit a pour but l'étude des effets super-dispersifs, correspondant à de fines déviations de la mesure de dispersion par rapport à relation de dispersion des plasmas froids habituellement utilisée. De par la dépendance fréquentielle de la dispersion, celle-ci a une amplitude qui augmente vers les basses fréquences et dans le cas de larges bandes de fréquences relatives. En ce sens, NenuFAR est le parfait instrument pour tenter d'éventuels effets super-dispersifs.

De plus, ces éventuels effets sont attendus pour être très fins, et ainsi apparaître à courtes échelles. Pour pallier à tout effet de moyennage, l'étude présentée ici porte sur des mesures "instantanées" sur des impulsions individuelles.

Les mesures ont ainsi été réalisées sur des sous-bandes de fréquences d'impulsions individuelles en utilisant une méthode développée pour cette étude. Celle-ci se base sur le comportement particulier d'une impulsion dispersée dans l'espace de Fourier, et réalise la recherche de la mesure de dispersion par la maximisation d'un indicateur de dispersion.

Les mesures ont été effectuées sur les six plus puissantes impulsions de cinq observations de quatre pulsars différents, en divisant l'impulsion en six, huit, et douze sous-bandes de fréquences.

Les mesures individuelles montrent de légères déviations par rapport aux dispersions moyennes des impulsions considérées. Néanmoins, la très grande majorité de ces mesures sont en cohérence du point de vue de l'incertitude à 1σ avec une dispersion constante en fréquence.

En combinant les déviations mesurées, il est apparu de légères tendances linéaires pour quatre des cinq observations. Cependant, là encore, les intervalles d'erreur à 1σ sont cohérents avec une dispersion constante.

L'observation de B0950+08 a cependant montré un comportement très spécifique. En particulier, l'impulsion numéro 1920 montre une variation de la mesure de dispersion d'environ 6×10^{-3} pc.cm^{-3} entre la plus basse et la plus haute sous-bande de fréquence, avec une évolution parabolique de la mesure. De plus, bien que les autres mesures individuelles soient cohérentes à 1σ avec la dispersion constante, la mesure combinée ne l'est pas, exposant une claire tendance parabolique.

Dans l'optique de vérifier le caractère spécifique de B0950+08 et de l'impulsion 1920, il serait intéressant d'analyser d'autres observations de ce pulsar, notamment en utilisant des observations de NenuFAR avec 80 MR. Aussi, il existe tout de même quelques points de mesure non cohérent avec une dispersion constante même à 3σ d'incertitude. Or, la méthode utilisée ici ne détermine l'incertitude que par rapport au bruit de fond présent dans la TF du spectre dynamique. En conséquence, afin de confirmer ou d'infirmer ces mesures divergentes, il faudrait quantifier de manière précise les incertitudes issues de l'échantillonnage des données. Un moyen aussi d'affiner ces intervalles d'incertitudes pourrait être de réaliser les mesures sur des impulsions ayant plus de flux, en utilisant des observations de NenuFAR avec 80 MR.

La plupart des modèles d'effets super-dispersifs produisent des variations monotones, suivant par conséquent des lois puissances simples en fonction de la fréquence. Or, si l'on s'autorise à considérer les légères déviations et tendances vues dans cette étude préliminaire, il apparaît qu'elles ne sont pas purement monotones.

Un modèle de propagation de l'onde avec des variations spatiales et temporelles de la densité électronique pourrait éventuellement permettre de recréer des effets super-dispersifs à courtes échelles. Un tel modèle devrait générer des déviations à la dispersion glissant en fréquence en fonction du temps, et pourrait ainsi donc être détecté en réalisant des mesures sur un ensemble d'impulsions contiguës.

En plus de la propagation en espace et temps de l'onde, l'ajout du second ordre de la relation de dispersion permettrait d'obtenir des déviations linéaires, donc monotones en fréquence, mais également non-linéaire grâce à un facteur multiplicatif non linéaire dépendant de la fréquence. Afin de contraindre quantitativement l'amplitude de ces deux termes sur la mesure de dispersion, il est nécessaire de réaliser des simulations précises, permettant de confirmer ou d'infirmer s'ils peuvent générer des effets suffisamment importants.

Pour finir, le modèle de propagation pourrait aussi être étendu à quatre dimensions afin de prendre en compte l'effet de propagation multi-chemins causant les phénomènes de scattering et scintillation.

Conclusion

The thesis aimed to present two works: a first work about the development of the first stage of the NenuFAR pulsar blind survey, and a second work about a preliminary study of the super-dispersive effects.

Concerning the survey, the sky has been observed in the frequency range from 39 to 76 MHz. This survey allows us to search for pulsars in a little-known part of the spectrum. That results in the expectations to potentially discover more exotic pulsars or related objects, such as pulsars featuring low-frequency steep spectra or clearly slower than the usual known population.

A pointing grid has been defined, allowing us to observe the northern polar cap above 39° with an average coverage of the sky of approximately 98% along the observed bandwidth. The observing program has been realized for the last two years and is finished at 98%. The last 2% will be observed in the first semester of 2023 to complete the observations of the first stage of the survey.

The processing pipeline of the data has been developed and is currently ongoing. The pipeline is an adaptation to the low-frequency constraints of a program dedicated to the search for pulsars at high frequency. A dedicated processing node has been specifically built for the survey. A third of the first stage data have been currently processed, resulting in the detection of more than 130 000 candidates. The remaining data will be processed during 2023, to complete the first stage processing of the survey for the beginning of 2024.

Finally, a method of analysis of the candidates has been developed in order to sort and exclude the candidates with a low probability to be a real pulsar. The analysis method is based on Monte-Carlo simulations, allowing to calculate a significance corresponding to a Gaussian probability of similarity with a signal of a pulsar. In the candidates corresponding to the currently processed data, the application of the method has allowed to exclude about 80% of the candidates. After inspection of the highest-ranked candidates, 17 have been considered a potential pulsar. These candidates are selected for at least two re-observations in 2023.

Concerning the second work presented in this thesis about the study of the super-dispersive effects, a few observations of NenuFAR has been analyzed. In order to prevent averaging effects, the chosen method was to perform measurements on single pulses only using NenuFAR. A specific method for determining the dispersion measure of frequency subbands of a single pulse has been developed. The used method is based on the estimation of the dispersion suffered by a pulse, using the particular pattern of this pulse in the Fourier space of the dynamic spectrum.

Using this measurement method, 30 pulses have been analyzed with eight measurement points. For the majority of the points, no conclusive evidence of deviations from the usual dispersion relation has been observed. However, some points, especially at low frequency, seem to present significant deviations. A particular observation of the pulsar B0950+08 presents an intriguing behavior compared to the other analyzed observations.

B0950+08 is known to have DM variations between microstructures at very low frequencies below 30 MHz. In addition, this parabolic trend could be similar to the time-frequency drift of subbursts, called "sad trombone", observed in some FRBs. Nevertheless, only six pulses from one observation have been analyzed. It is, therefore, necessary to expand the DM analyses of this pulsar at other observations of NenuFAR to conclude about this unique (for the moment) observed behavior for a pulsar with NenuFAR.

For the survey, according to the current number of candidates kept for re-observation, we can expect approximately 50 pulsar candidates for the whole survey. To be considered as a new pulsar detection, these candidates should be observed in the same conditions several times and obtain a similar result at each observation. Because of the recent discoveries of slow pulsars and pulsars featuring steep spectra, we can expect some of these interesting candidates could be new pulsars visible only at frequencies below 100 MHz.

For the moment, all the interesting candidates are faint. As a consequence, they have not allowed to provide a likely detection of an unknown pulsar. However, the sky above 39° represents only approximately 16% of the observable sky with NenuFAR. Moreover, merely a third of this sky has been processed. It is therefore difficult to conclude something about the populations of steep spectrum pulsars and slow pulsars.

Furthermore, the search method used in this work is not particularly efficient to detect slow pulsars. Other types of methods more adapted for the search of slow pulsars, or unconstant pulsars as RRATs, should be used. We can expect that could allow us to detect a larger number of pulsars than the periodicity search method applied here.

For the study of the super-dispersive effects, the number of analyzed pulses in this work is limited. However, although there are consistent at 1σ of 3σ for almost all, some slight deviations seem to be seen for particular points. Furthermore, the pulses analyzed in this work seem to show slight frequency-dependent linear trends of the DM. An increase of the number of pulsars, observations for each, and number of pulses in each observation could allow to confirm or not this trend.

Also, there is the case of the pulsar B0950+08, seeming to have an unusual behavior compared to the other pulsars of this study. These preliminary results equally require analyzing different observations of this pulsar. In addition, individual measurements of the single pulse components seem to be interesting to identify the source of the parabolic trend. Also, increasing the number of studied pulsars could permit us to potentially identify other pulsars with similar behavior. Finally, for the two points, in order to obtain conclusive results, it needs to reduce the uncertainties. The increase in the number of MAs of NenuFAR could allow us to obtain new observations with larger SNR, and therefore more precise measurements of the dispersion.

The small deviations which are here within the error at 1σ present nevertheless a dispersion of the deviations between frequencies, pulses, and pulsars. In order to improve the current model of the dispersion, it could be interesting to evaluate it in a more detailed way. Moreover, the deviations seem to don't necessarily follow a monotonic law relative to the frequency. To confirm the reality of these variations, the uncertainties of the measurement method used in this work must be more precisely evaluated. Nevertheless, I have proposed a model based on the actual relation taking the propagation in space and time of the electromagnetic wave into account. According to the first simple simulations carried out, it could qualitatively generate the different behaviors identified in this work. Furthermore, the second order of the plasma dispersion relation allows adding frequency-dependent non-linear terms and terms creating variations at smaller scales. Also, coupling these additional terms with the model of propagation could explain these small deviations. However, these two propositions require larger simulations than those done in this preliminary work to generate quantitatively the potential deviations to the usual dispersion relation.

Finally, two extensions of the present work could be developed. The first one is that the statistical dispersion of the DM deviations indicates a non-constancy at least over time of the DM at timescales shorter than an observation. Such DM deviations can induce variations in the TOA of the timing process. A precise measure of the DM deviations along the duration of one observation could allow for obtaining upper limits in the TOA error. The second one is that the model of propagation proposed here doesn't take the spatial propagations in the perpendicular directions into account. With the addition of the two other space dimensions, the model could potentially generate, in addition to the dispersion deviations, the scattering, and the scintillation undergone by a pulse. Such a model could potentially allow, at least for not too-complex cases, to describe the three effects with a unique model.

Résumé du chapitre :

Cette thèse ici présente avait pour but de présenter deux travaux : le premier portait sur le développement de la première étape du relevé aveugle de pulsars avec NenuFAR, et le second portait sur l'étude des effets super-dispersifs dans les observations de pulsations individuelles de NenuFAR.

Concernant le relevé aveugle, une grille de pointage a été construite permettant d'observer 98 % du ciel au-dessus de 39° de déclinaison. Au 31 août 2022, 98 % de la grille ont été observées, et les 2 % restants seront observés durant le printemps 2023.

Un processus de traitement des données adapté aux basses fréquences a été développé. Le traitement des données est actuellement en cours, et un tiers des observations ont déjà été traitées, trouvant environ 130 000 candidats. Le reste des observations sera traité durant 2023, et le traitement devrait être achevé au début de 2024.

Finalement, une méthode d'analyse des candidats basée sur des simulations Monte-Carlo des caractéristiques du candidat, a été développée dans l'optique de classifier le très grand nombre de candidats. Les candidats actuellement trouvés ont été analysés, permettant d'exclure environ 80 % de ceux-ci. 17 candidats intéressants ont été sélectionner pour ré-observation durant le printemps 2023.

Pour ce qui est du second projet, 30 impulsions individuelles de quelques observations de NenuFAR ont été analysées. Une méthode spécifique, basée sur l'étude du comportement dans l'espace de Fourier de l'impulsion, a été développée dans le but de mesurer précisément les mesures de dispersion de sous-bandes de fréquences d'une impulsion individuelle.

Dans la très grande majorité des mesures, aucune déviation significative à la relation de dispersion usuelle n'a été détectée. Cependant, quelques points de mesure localisés ne sont pas totalement en cohérence avec la loi standard, et le pulsar B0950+08 semble montrer un comportement différent.

Pour le relevé, compte tenu des candidats actuellement sélectionner comme intéressant, nous pouvons attendre environ 50 candidats pulsars pour l'ensemble du relevé. Pour être considéré comme une réelle détection, ils devront être observés plusieurs fois dans les mêmes conditions, et être trouvés à nouveau comme candidat. Due aux récentes découvertes de pulsars lents et de pulsars montrant des spectres très pentus, nous pouvons raisonnablement penser que certains de ces candidats pourront être des pulsars uniquement visibles au-dessous de 100 MHz.

De plus, les méthodes appliquées dans ce présent travail ne sont pas très adaptées pour la recherche de pulsars lents ou non continus dans le temps tels que les RRAT. L'utilisation de méthodes plus efficaces pourrait accroître le nombre de pulsars détectables.

Pour ce qui est de la recherche des effets super-dispersifs, le nombre d'impulsions sur lesquelles faire les mesures doit être augmenté. De plus, il est nécessaire de réaliser des observations avec plus de flux, afin de réduire les intervalles d'incertitudes pour mieux évaluer si les légères tendances observées tendent vers la dispersion constante ou non.

Dans le cas où les faibles déviations seraient confirmées, les modèles proposés de propagation dans l'espace et le temps, et la prise en compte du second ordre de la relation de dispersion permettraient éventuellement de modéliser ces comportements. Cependant, des simulations poussées permettant de quantifier précisément l'impact de ces modèles sur la mesure de dispersion doivent être réalisées.

Finalement, le modèle de propagation pourrait être étendue en incluant les deux autres dimensions spatiales de manière à tenter de modéliser la dispersion, le scattering, et la scintillation aux petites échelles.

Bibliography

Ahuja, A. L., Gupta, Y., Mitra, D., & Kembhavi, A. K. 2005, MNRAS, 357, 1013

Ahuja, A. L., Mitra, D., & Gupta, Y. 2007, MNRAS, 377, 677

Baade, W. 1942, ApJ, 96, 188

Baade, W. & Zwicky, F. 1934, Proceedings of the National Academy of Science, 20, 254

Backer, D. C. 1970a, Nature, 228, 1297

Backer, D. C. 1970b, Nature, 228, 42

Backer, D. C. 1976, ApJ, 209, 895

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

Bilous, A. V., Bondonneau, L., Kondratiev, V. I., et al. 2020, A&A, 635, A75

Bondonneau, L., Grießmeier, J. M., Theureau, G., et al. 2021, A&A, 652, A34

Bondonneau, L. et al. in preparation

Burgay, M., D'Amico, N., Possenti, A., et al. 2003, Nature, 426, 531

Caleb, M., Heywood, I., Rajwade, K., et al. 2022, Nature Astronomy, 6, 828

Cole, T. W. & Pilkington, J. D. H. 1968, Nature, 219, 574

Cordes, J. M. 1978, ApJ, 222, 1006

Cordes, J. M. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 278, Single-Dish Radio Astronomy: Techniques and Applications, ed. S. Stanimirovic, D. Altschuler, P. Goldsmith, & C. Salter, 227–250

Cordes, J. M., Bhat, N. D. R., Hankins, T. H., McLaughlin, M. A., & Kern, J. 2004, ApJ, 612, 375

Cordes, J. M. & Lazio, T. J. W. 2002, arXiv e-prints, astro

Cordes, J. M. & Shannon, R. M. 2010, arXiv e-prints, arXiv:1010.3785

Cordes, J. M., Shannon, R. M., & Stinebring, D. R. 2016, ApJ, 817, 16

Daugherty, J. K. & Harding, A. K. 1986, ApJ, 309, 362

Davies, J. G. & Large, M. I. 1970, MNRAS, 149, 301

Desvignes, G., Kramer, M., Lee, K., et al. 2019, Science, 365, 1013

Dewey, R. J., Taylor, J. H., Weisberg, J. M., & Stokes, G. H. 1985, ApJL, 294, L25

Donner, J. Y., Verbiest, J. P. W., Tiburzi, C., et al. 2020, A&A, 644, A153

Emmering, R. T. & Chevalier, R. A. 1989, ApJ, 345, 931

Faucher-Giguère, C.-A. & Kaspi, V. M. 2006, ApJ, 643, 332

Foster, R. S. & Cordes, J. M. 1990, ApJ, 364, 123

Gangadhara, R. T. & Gupta, Y. 2001, ApJ, 555, 31

Gil, J., Gronkowski, P., & Rudnicki, W. 1984, A&A, 132, 312

Glendenning, N. K. 1992, Phys. Rev. D, 46, 1274

Gold, T. 1968, Nature, 218, 731

Goldreich, P. & Julian, W. H. 1969, ApJ, 157, 869

Gupta, Y., Rickett, B. J., & Lyne, A. G. 1994, MNRAS, 269, 1035

Han, J. L., Wang, C., Wang, P. F., et al. 2021, Research in Astronomy and Astrophysics, 21, 107

Hankins, T. H., Izvekova, V. A., Malofeev, V. M., et al. 1991, ApJL, 373, L17

Hankins, T. H. & Rickett, B. J. 1975, Methods in Computational Physics, 14, 55

Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1995, Astronomy Data Image Library

Hassall, T. E., Stappers, B. W., Hessels, J. W. T., et al. 2012, A&A, 543, A66

Hemberger, D. A. & Stinebring, D. R. 2008, ApJL, 674, L37

Hessels, J. W. T., Spitler, L. G., Seymour, A. D., et al. 2019, ApJL, 876, L23

Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. 1968, Nature, 217, 709

Huguenin, G. R., Taylor, J. H., Goad, L. E., et al. 1968, Nature, 219, 576

Hulse, R. A. & Taylor, J. H. 1974, ApJL, 191, L59

Hurley-Walker, N., Zhang, X., Bahramian, A., et al. 2022, Nature, 601, 526

Jankowski, F., van Straten, W., Keane, E. F., et al. 2018, MNRAS, 473, 4436

Jenet, F., Finn, L. S., Lazio, J., et al. 2009, arXiv e-prints, arXiv:0909.1058

Jones, A. W. & Lyne, A. G. 1988, MNRAS, 232, 473

Josephy, A. 2020, Clustering: Code for clustering single pulse events, Astrophysics Source Code Library, record ascl:2011.018

Kalogera, V. & Baym, G. 1996, ApJL, 470, L61

Karako-Argaman, C., Kaspi, V. M., Lynch, R. S., et al. 2020, RRATtrap: Rotating Radio Transient identifier, Astrophysics Source Code Library, record ascl:2011.017

Kaspi, V. M. & Helfand, D. J. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 271, Neutron Stars in Supernova Remnants, ed. P. O. Slane & B. M. Gaensler, 3

Keith, M. J., Jameson, A., van Straten, W., et al. 2010, MNRAS, 409, 619

Kijak, J. & Gil, J. 2003, A&A, 397, 969

Kijak, J., Gupta, Y., & Krzeszowski, K. 2007, A&A, 462, 699

Komesaroff, M. M., Morris, D., & Cooke, D. J. 1970, Astrophys. Lett., 5, 37

Kulkarni, S. R. 2020, arXiv e-prints, arXiv:2007.02886

Kuzmin, A. D. 1986, Soviet Astronomy Letters, 12, 325

Landau, L. D. 1932, Phys. Zs. Sowjet, 1, 285

Large, M. I., Vaughan, A. E., & Wielebinski, R. 1968, Nature, 220, 753

Lawson, K. D., Mayer, C. J., Osborne, J. L., & Parkinson, M. L. 1987, MNRAS, 225, 307

Lazarus, P., Brazier, A., Hessels, J. W. T., et al. 2015, ApJ, 812, 81

Leahy, D. A., Darbro, W., Elsner, R. F., et al. 1983, ApJ, 266, 160

Li, T. P. & Ma, Y. Q. 1983, ApJ, 272, 317

Loh, A. & Girard, J. N. 2020, AlanLoh/nenupy: nenupy 1.1.0

Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, Science, 318, 777

Lorimer, D. R. & Kramer, M. 2012, Handbook of Pulsar Astronomy (Cambridge University Press, New York)

Lyne, A. G. & Graham-Smith, F. 2012, Pulsar astronomy (Cambridge University Press, New York)

Lyne, A. G. & Lorimer, D. R. 1994, Nature, 369, 127

Lyne, A. G. & Manchester, R. N. 1988, MNRAS, 234, 477

Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993

Manchester, R. N., Kramer, M., Stairs, I. H., et al. 2010, ApJ, 710, 1694

Manchester, R. N., Lyne, A. G., Camilo, F., et al. 2001, MNRAS, 328, 17

Manchester, R. N., Lyne, A. G., Taylor, J. H., et al. 1978, MNRAS, 185, 409

Maron, O., Kijak, J., Kramer, M., & Wielebinski, R. 2000, A&AS, 147, 195

McEwen, A. E., Spiewak, R., Swiggum, J. K., et al. 2020, ApJ, 892, 76

McLaughlin, M. A., Lyne, A. G., Lorimer, D. R., et al. 2006, Nature, 439, 817

Meltzer, D. W. & Thorne, K. S. 1966, ApJ, 145, 514

Miller, M. C., Lamb, F. K., Dittmann, A. J., et al. 2019, ApJL, 887, L24

Morello, V., Barr, E. D., Stappers, B. W., Keane, E. F., & Lyne, A. G. 2020, MNRAS, 497, 4654

Morris, D., Kramer, M., Thum, C., et al. 1997, A&A, 322, L17

Oster, L. & Sieber, W. 1976, ApJ, 210, 220

Pacini, F. 1967, Nature, 216, 567

Pacini, F. 1968, Nature, 219, 145

Parent, E., Sewalls, H., Freire, P. C. C., et al. 2022, ApJ, 924, 135

Piffl, T., Scannapieco, C., Binney, J., et al. 2014, A&A, 562, A91

Pilkington, J. D. H., Hewish, A., Bell, S. J., & Cole, T. W. 1968, Nature, 218, 126

Radhakrishnan, V. & Cooke, D. J. 1969, Astrophys. Lett., 3, 225

Rajabi, F., Chamma, M. A., Wyenberg, C. M., Mathews, A., & Houde, M. 2020, MNRAS, 498, 4936

Rankin, J. M. 1993, ApJ, 405, 285

Ransom, S. 2011, PRESTO: PulsaR Exploration and Search TOolkit

Richards, D. W. & Comella, J. M. 1969, Nature, 222, 551

Rickett, B. J. 1990, ARA&A, 28, 561

Ruderman, M. A. & Sutherland, P. G. 1975a, ApJ, 196, 51

Ruderman, M. A. & Sutherland, P. G. 1975b, ApJ, 196, 51

Sanidas, S., Cooper, S., Bassa, C. G., et al. 2019, A&A, 626, A104

Scheuer, P. A. G. 1968, Nature, 218, 920

Shapiro, S. L. & Teukolsky, S. A. 1983, Black holes, white dwarfs, and neutron stars : the physics of compact objects (John Wiley & Sons, Ltd)

Shitov, Y. P. 1983, Soviet Ast., 27, 314

Shitov, Y. P. & Malofeev, V. M. 1985, Soviet Astronomy Letters, 11, 39

Shitov, Y. P., Malofeev, V. M., & Izvekova, V. A. 1988, Soviet Astronomy Letters, 14, 181

Sieber, W. 1973, A&A, 28, 237

Sieber, W. & Oster, L. 1977, A&A, 61, 445

Smirnova, T. V. & Shishov, V. I. 2008, Astronomy Reports, 52, 736

Staelin, D. H. & Reifenstein, Edward C., I. 1968, Science, 162, 1481

Stappers, B. W., Kramer, M., Lyne, A. G., D'Amico, N., & Jessner, A. 2006, Chinese Journal of Astronomy and Astrophysics Supplement, 6, 298

Sturrock, P. A. 1971, ApJ, 164, 529

Tan, C. M., Bassa, C. G., Cooper, S., et al. 2018, ApJ, 866, 54

Tanenbaum, B. S., Zeissig, G. A., & Drake, F. D. 1968, Science, 160, 760

Taylor, J. H. & Huguenin, G. R. 1969, Nature, 221, 816

Ulyanov, O. M., Skoryk, A. O., Shevtsova, A. I., Plakhov, M. S., & Ulyanova, O. O. 2016, MNRAS, 455, 150

Urquhart, J. S., Figura, C. C., Moore, T. J. T., et al. 2014, MNRAS, 437, 1791

van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2

Vasylieva, I. Y., Zakharenko, V. V., Konovalenko, A. A., et al. 2014, Radio Phys. Radio Astron., 19, 197

Voisin, G., Cognard, I., Freire, P. C. C., et al. 2020, A&A, 638, A24

Wang, W., Zhang, B., Chen, X., & Xu, R. 2019, ApJL, 876, L15

Weisberg, J. M. & Taylor, J. H. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 302, Radio Pulsars, ed. M. Bailes, D. J. Nice, & S. E. Thorsett, 93

Wu, Z., Verbiest, J. P. W., Main, R. A., et al. 2022, A&A, 663, A116

Zakharenko, V. V., Vasylieva, I. Y., Konovalenko, A. A., et al. 2013, MNRAS, 431, 3624

Zarka, P. et al. in prep.

Mark BRIONNE

Relevé aveugle de pulsars avec NenuFAR et étude des effets super-dispersifs

Résumé :

Les pulsars sont des étoiles compactes ayant des vitesses de rotation rapides, présentant des périodes allant d'environ une milliseconde à plusieurs dizaines de secondes, et disposant d'un fort champ magnétique allant de 10^{12} à 10^{15} G. De par le ralentissement de la rotation et un non-alignement de l'axe de rotation et de l'axe magnétique, un effet phare est généré, menant à l'observation d'impulsions radio périodiques.

Le premier pulsar fut découvert en 1968, et plusieurs milliers d'autres ont été découverts depuis. Bien que les premiers pulsars furent découverts en dessous de 100 MHz, la majorité a été découverte à des fréquences supérieures à 300 MHz. NenuFAR est un nouveau radiotélescope inauguré en 2019, permettant d'observer de 10 à 85 MHz, avec une sensibilité supérieure aux précédents télescopes observant aux mêmes fréquences. En utilisant NenuFAR, un relevé à basses fréquences de pulsars de l'hémisphère nord a été initié, dans le but de trouver de nouveaux pulsars.
La première partie de cette thèse présente le développement du « NenuFAR pulsar blind survey ». Le travail de cette thèse présente la première étape de ce relevé, commencé en août 2020, ayant pour but d'observer le ciel au-dessus de 39° de déclinaison. Les deux premiers chapitres expliquent la création de la grille de pointage, et l'avancée du programme d'observation associé. Le chapitre suivant détaille les différentes étapes du traitement des données, qui est une adaptation aux contraintes des basses fréquences d'un processus de recherche pour les hautes fréquences, appelé PRESTO. Finalement, le dernier chapitre présente la méthode d'analyse des candidats pulsars trouvés par le processus de recherche. La méthode utilisée est basée sur des simulations Monte-Carlo, et permet de sélectionner les candidats les plus intéressants, listés à la fin de cette première partie de la thèse.

Les ondes électromagnétiques produisant les pulsations observées doivent traverser le milieu interstellaire pour aller du pulsar à la Terre. Ce milieu est un plasma, et la propagation de l'onde à travers celui-ci produit un effet de dispersion de l'onde. Depuis 1986, des effets super-dispersifs ont été proposés, correspondant à des déviations par rapport à la loi usuelle de dispersion dans un plasma froid. De manière à clairement détecter et caractériser ces déviations, plusieurs théories ont été testées depuis.
La loi de dispersion dans un plasma froid est dépendante de la fréquence, menant au fait que, plus la fréquence est basse et plus l'amplitude de la dispersion de l'onde est importante. Dans ce contexte, NenuFAR est par conséquent intéressant pour deux raisons. Premièrement, il permet d'obtenir des observations à basses fréquences entre la coupure ionosphérique et 85 MHz. Deuxièmement, avec un facteur huit entre la plus basse et la plus haute fréquence, il possède une importante largeur de bande relative, facilitant l'identification de possibles variations en fréquence.
De plus, les déviations cherchées devraient être de fines variations, c'est-à-dire montrant des amplitudes faibles comparées à celles de la dispersion usuelle, demandant ainsi des mesures précises. La seconde partie de la thèse présente les mesures de la dispersion réalisées sur des impulsions individuelles de pulsars puissants observés avec NenuFAR. Afin d'effectuer des mesures à de multiples fréquences sur une impulsion individuelle, une méthode de détermination de la mesure de dispersion a été développée, et est détaillée ici. Avec cette méthode, la dispersion est déterminée dans l'espace de Fourier du spectre dynamique de l'impulsion, permettant une mesure sur une fraction de l'impulsion individuelle. Finalement, les résultats de mesures sont présentés et discutés.

Mots clés : pulsars, relevé, radio-astronomie, basses fréquences, NenuFAR

Mark BRIONNE

NenuFAR pulsar blind survey and study of the super-dispersive effects

Abstract :

Pulsars are compact stars with high rotation velocity, featuring periods typically from about one millisecond to several dozens of seconds, and with a high magnetic field from 10^{12} to 10^{15} G. Due to a spin-down effect and a misalignment between the rotation and magnetic axes, a lighthouse effect is generated, resulting in the observation of periodic radio pulsations.

The first pulsar was discovered in 1968, and several thousands have been discovered since. Although the first one was discovered below 100 MHz, the majority has been discovered at frequencies higher than 300 MHz. NenuFAR is a new radio-telescope inaugurated in 2019, allowing to observe from 10 to 85 MHz, with a higher sensitivity than the previous telescopes observing at the same frequencies. Using NenuFAR, a low-frequency pulsar survey of the northern sky has been initiated, aiming to find new pulsars.

The first part of this thesis presents the development of the "NenuFAR pulsar blind survey". The work in this thesis presents the first stage of the survey, which has begun in August 2020, with the aim to observe the north polar cap above 39° of declination. The two first chapters explain the creation of the pointing grid and the progress of the associated observing program. The next chapter details the different steps of data processing, which is an adaptation of a high-frequency search pipeline called PRESTO to the low-frequency constraints. Finally, the last chapter presents the analysis method of the pulsar candidates found by the search pipeline. The used method is based on Monte-Carlo simulations and allows to select of the most interesting candidates, listed at the end of this first part of the thesis.

The electromagnetic waves producing the observed pulsations must travel from the pulsar through the interstellar medium to reach the Earth. This medium is plasma, and the propagation of the wave inside it yields a dispersion effect. Since 1986, some super-dispersive effects have been proposed, corresponding to deviations relative to the usual cold plasma dispersion relation. In order to clearly detect and characterize these deviations, different theories have been tested since.
The cold plasma dispersion relation is frequency dependent, resulting in the fact that, the lower the frequency the higher the amplitude of the dispersion of the wave. In this context, NenuFAR is therefore interesting for two reasons. First, it allows low-frequency observations between the ionospheric cutoff and 85 MHz. Secondly, with a factor of eight between the lowest and the highest frequencies, it has an important relative bandwidth, which facilitates the identification of possible frequency dependencies.
Moreover, the searched deviations might be small variations, i.e. showing a low amplitude compared to the usual dispersion, needing thereby precise measurements. The second part of this thesis presents measurements of the dispersion realized on single pulses of powerful pulsars observed with NenuFAR. In order to carry out multiple frequency measurements on single pulses, a method of determination of the dispersion measure has been specifically developed for this study and is detailed here. With this method, the dispersion is determined in the Fourier space of the dynamic spectrum of the pulse, allowing a measure in a fraction of a single pulse. Finally, the results of the measurements are presented and discussed.

Keywords : pulsars, survey, radio-astronomy, low-frequency, NenuFAR

Laboratoire de Physique et Chimie de l'Environnement et de l'Espace 3A Avenue. de la Recherche Scientifique 45071 ORLÉANS