

The NenuFAR Pulsar Blind Survey (NPBS): Pulsar search and pipeline optimization

Author:
Nour Harika
M2 Theoretical Physics and Modelisation

Supervisor:
Dr. Cherry Ng-Guiheneuf
Permanent researcher at LPC2E

Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, CNRS Orléans
Observatoire Radioastronomique de Nançay

02/09/2025

Master Internship Report



Figure 1: NenuFAR radio telescope array: over 2000 antennas, with every 19 forming one mini array.

Contents

| | |
|---|----|
| A Recognition note | 5 |
| 1 Introduction | 6 |
| 2 Pulsar Theory and Detection Principles | 6 |
| 3 NenuFAR Telescope and detection pipelines | 9 |
| 3.1 Why use NenuFAR for low frequency observations? | 9 |
| 3.2 Pulsar Observations | 10 |
| 3.3 The Detection Pipeline of NenuFAR | 10 |
| 4 NPBS Observing Strategy and Data Format | 11 |
| 5 The data format used for this study | 12 |
| 5.1 Task 1: ATNF vs NenuFAR PSR | 12 |
| 5.2 Interference | 13 |
| 5.3 Task 3: PSR search | 13 |
| 5.4 Task 4: rfifind PRESTO | 13 |
| 6 ATNF vs NenuFAR pulsar catalog | 13 |
| 7 Handling the Radio Frequency Interference | 18 |
| 8 Pulsar Search | 20 |
| 9 rfifind with PRESTO | 23 |
| 10 Preliminary Results and Findings | 24 |
| 11 Future work | 24 |
| 12 Conclusion | 25 |
| References | 27 |
| Appendix | 28 |
| A The LOFAR Telescope Network and the French Station | 28 |
| B Pulsar Keyword Definitions | 29 |
| C RFI base study | 29 |
| D rfifind output profile | 30 |
| E Codes created for the study | 31 |

List of Figures

| | | |
|----|--|----|
| 1 | NenuFAR radio telescope array: over 2000 antennas, with every 19 forming one mini array. | 1 |
| 2 | Pulsar image. <i>Figure by the National Radio Astronomy Observatory.</i> | 6 |
| 3 | P–Pdot diagram showing pulsar populations: young pulsars - Central Compact Objects (CCOs) - and magnetars (upper-left), normal pulsars (center), millisecond pulsars (lower-left), and old pulsars - Isolated Neutron Stars (INS) - near the “graveyard” (bottom-right). <i>Image by A.K.Harding.</i> | 7 |
| 4 | Pulsar profile by NenuFAR, the shape of the pulsar is a typical one. | 8 |
| 5 | Another Pulsar profile by NenuFAR, multiple peaks but still is a pulsar signal. | 8 |
| 6 | The BS Flowchart that explains the pipelines of NenuFAR data processing. <i>Flowchart by Cherru NG on github.</i> | 12 |
| 7 | Format of the mask zap chans.txt files, giving all the information about a pointing. | 13 |
| 8 | Population of both NenuFAR (cyan) and ATNF (violet) scattered based on Right Ascension vs declination. The galactic plane curvature is visible. | 14 |
| 9 | Population of both catalogs presented in galactic coordinates. | 14 |
| 10 | Population of PSR in 3D. | 15 |
| 11 | DM vs GB to show that PSR of ATNF are the closest to the Galactic Center | 15 |
| 12 | $P - \dot{P}$ diagram of both catalog, showing that NenuFAR mostly detects young and normal pulsars. | 16 |
| 13 | Period vs DM, NenuFAR can't detect more than 100 DM and 10 s. | 16 |
| 14 | Histogram of DM distribution of the pulsars. | 16 |
| 15 | Flux density, the lower the frequency the brighter the pulsar, green represents the highest flux and lowest frequency. | 16 |
| 16 | Pulse width distribution W50 of the pulsars. | 17 |
| 17 | Duty cycle pulsar vs Period, validating again the PSR population of NenuFAR. | 17 |
| 18 | CDF Pulsar Period, NenuFAR coincide with ATNF detections only around young/normal pulsars period (around 1 s and less than 10 s). | 18 |
| 19 | KStest values of NenuFAR and ATNF with CDF of DM. The P-value: 4.0018×10^{-76} , meaning NenuFAR is not sampling the same population as ATNF. | 18 |
| 20 | Statistical study of flags, emphasizing on the bad channels in gray. The gray area represents the flags that peak more than 40%. | 19 |
| 21 | Masked flagged at threshold 40%. | 20 |
| 22 | Four signal profiles of the $P = 140$ ms RFI. | 21 |
| 23 | Potential candidate of $P = 170$ ms. | 22 |
| 24 | Potential candidate of $P = 79.8$ ms. | 23 |
| 25 | International LOFAR locations. | 28 |
| 26 | Statistical study of flags, showing by a cumulative distribution the % of data affected when every channel is removed. We decided to go at threshold 40%. | 29 |
| 27 | Output file “...rfifind.ps” where colors are bad (red is periodic RFI, blue/-green are time-domain statistical issues. This file is the output of the rfifind by itself, without the -zapchan list of masking at 40 % threshold. | 30 |

28 This file is the output of the `rfifind` with `-zapchan` list of the 246 bad channels, the 40 % threshold. We notice a different distribution of the red lines, some of them are masked after the list, that were not before. Same with the blue and green line.

31

A Recognition note

Dear all,

I am truly grateful to have completed my internship at LPC2E, CNRS Orléans, with all expectations not only met but exceeded. Those who know me are aware of how deeply passionate I am about astrophysics, and this experience has reaffirmed my dream of pursuing a career as a researcher in the field.

Coming from a background more focused on theory and modeling, I initially feared being too distant from astrophysics to truly enter the domain. However, after this long and fruitful internship—my second in astrophysics—I feel more confident than ever to continue on this path. Even if it means spending countless hours examining data across many screens (and updating my glasses prescription!), I have discovered that this is where my passion lies. I am determined to keep this mindset, push boundaries, and ultimately work toward a PhD in astrophysics.

I would like to warmly thank Dr. Cherry Ng-Guiheneuf for her supervision and guidance. Entering the challenging world of pulsars and radio signals was not easy, but her support and expertise made all the difference. I would never have grasped the concept of "seeing the sky through antennas", a tiny bit different from optical telescopes, without this experience, especially working with thousands of them at once!

My thanks also go to the Astro team for their valuable input and discussions, which created a truly stimulating scientific environment. Everyone is conducting many interesting researches about different ways to study pulsars and transients.

Finally, I am grateful to my professors and our department's secretary for their continuous support and encouragement through challenges, setbacks and successes.

This journey has been a remarkable step forward, and I look forward to building on it in the future.

Sincerely,

Nour Harika

1 Introduction

Pulsars are the remnants of massive stars that have undergone supernova explosions, resulting in extremely dense, rapidly rotating neutron stars composed mainly of neutrons, from where it's name comes from. They emit beams of electromagnetic radiation from their magnetic poles. When these beams are aligned with the Earth during the star's rotation, they can be detected as periodic pulses, hence the name *pulsars*.

The NenuFAR Pulsar Blind Survey, referred to by NPBS, is a novel all-sky survey designed to search for pulsars in frequency range from 10 to 85 MHz, a frequency range previously unexplored by others for such surveys [3]. Prior to this, the lowest-frequency large-scale survey was LOTAAS, operating at a central frequency of 135 MHz [18]. NPBS is expected to discover between 8 and 20 new pulsars, many of which are anticipated to be atypical, occupying unusual regions of the period (P in s) and the period derivative (P dot in $s s^{-1}$) diagram (later referred by $P\dot{P}$).

The work conducted during this internship was organized into several key tasks, as illustrated in the workflow diagram. It is a continuation of Dr. Mark Brionne who worked on the NPBS. It began with a comparison between the ATNF Pulsar Catalogue [11] and unpublished detections from NenuFAR, to better understand the instrument's strengths and limitations. This was followed by an analysis of radio frequency interference (RFI) to enhance data quality and pulsar detection capabilities using the PRESTO software suite [17]. The final part of the work focused on identifying and evaluating promising pulsar candidates from the past three years of observations from the Nançay radio astronomy site.

2 Pulsar Theory and Detection Principles

Pulsars are rapidly rotating neutron stars, the compact remnants of massive stellar explosions: the supernovae. First discovered in 1967 by Jocelyn Bell Burnell and Antony Hewish [8], their highly periodic radio pulses were soon associated with strongly magnetized, dense stars composed almost entirely of neutrons [5][13]. These stars typically possess masses around $1.4 M_{\odot}$ (1.4 times the mass of our Sun) and radii on the order of 10-13 km [12]. Their rotational periods range from milliseconds to several seconds.

The surrounding plasma in the magnetosphere is accelerated by intense magnetic fields (typically 10^8 to 10^{14} G), giving rise to coherent radio emission along curved magnetic field lines [6].

As the emission is confined to a cone centered on the magnetic axis, and since the magnetic and rotational axes are generally misaligned, this beam sweeps across the sky like a lighthouse. When the Earth lies within

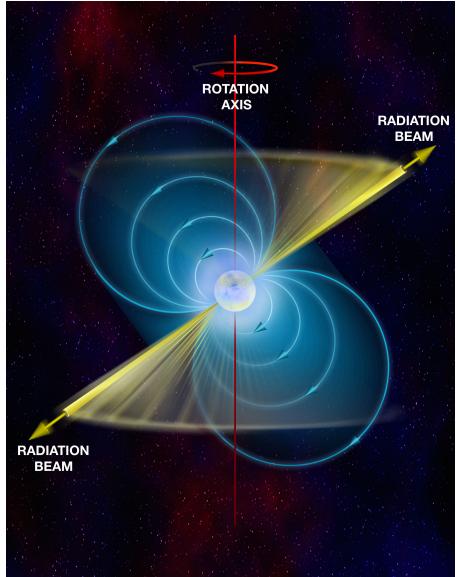


Figure 2: Pulsar image. Figure by the National Radio Astronomy Observatory.

this emission cone, a periodic radio signal can be detected by radio telescopes. This beaming effect explains the apparent pulsation observed in timing data. The Pulsars are known to be the clock of the cosmos due to their very precise periodicity.

Historically, the earliest pulsars were found in low-frequency surveys, including the Cambridge survey around 81.5 MHz [14]. However, most pulsar surveys have since operated at higher frequencies (300-1400 MHz), where radio interference is less severe and the ionospheric and interstellar effects are reduced. Notably, the LOFAR Tied-Array All-Sky Survey (LOTAAS) explored the low-frequency regime at a central frequency of 135 MHz [18], pushing the limits of low-frequency detection before the advent of the NenuFAR Pulsar Blind Survey, which now extends this frontier down to < 85 MHz.

A key diagnostic tool in pulsar astrophysics is the $P-\dot{P}$ diagram, where P (in s) is the rotation period and \dot{P} (in s/s) its time derivative, telling us how fast the pulsar is slowing down. This plot allows classification of pulsars based on their evolutionary stage, magnetic field strength, and energy loss rate. Young, energetic pulsars and millisecond pulsars occupy distinct regions of this diagram. From these quantities, the characteristic age τ_c

$$\tau_c = P/(2\dot{P}) \quad (1)$$

The surface dipolar magnetic field strength B of a pulsar, where the unit Gauss G of magnetic field ($1 G = 10 T^{-4}$) can be estimated:

$$B \approx 3.2 \times 10^{19} \sqrt{PP} \quad (2)$$

Based on the regions of the $P-\dot{P}$ diagram and spin down energy loss rate \dot{E} , we classify the currently known pulsars into five categories: the millisecond pulsars (MSP), the young pulsars, the normal or ordinary pulsar, high energy pulsar or magnetars and slow pulsars. In this work we are only interested in the last two population of pulsars.

Another essential feature is the pulsar's *pulse profile*, a phase-folded average of many pulses, which reveals the beam geometry and emission structure. The shape can vary widely between pulsars: some show single peaks, while others have multiple components or complex substructures. The profile's stability across time is one of the key reasons pulsars can be used as cosmic clocks.

Beyond their role in studying compact objects, pulsars serve as sensitive tools for probing the Universe on large scales. In particular, *Pulsar Timing Arrays* (PTAs) exploit the exceptional timing stability of millisecond pulsars to detect low-frequency gravitational waves from supermassive black hole binaries [1, 9]. Moreover, variations in pulse arrival times also reveal subtle

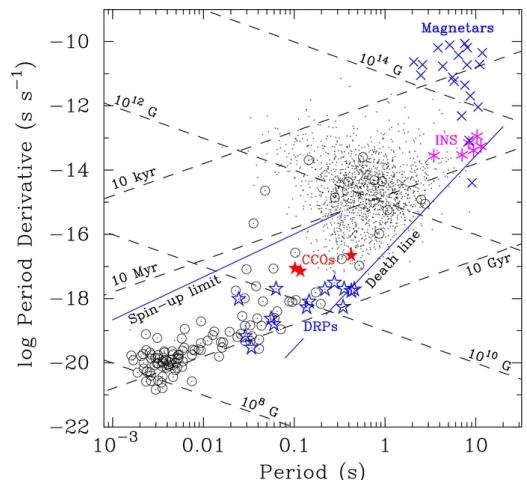


Figure 3: $P-\dot{P}$ diagram showing pulsar populations: young pulsars - Central Compact Objects (CCOs) - and magnetars (upper-left), normal pulsars (center), millisecond pulsars (lower-left), and old pulsars - Isolated Neutron Stars (INS) - near the “graveyard” (bottom-right).

Image by A.K.Harding.

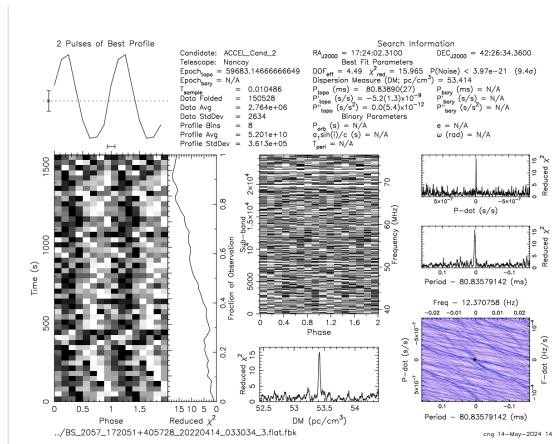


Figure 4: Pulsar profile by NenuFAR, the shape of the pulsar is a typical one.

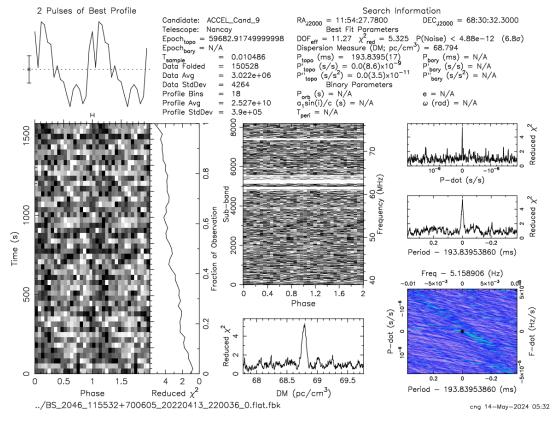


Figure 5: Another Pulsar profile by NenuFAR, multiple peaks but still is a pulsar signal.

changes caused by the interstellar medium (ISM), such as scattering, scintillation, or time-variable dispersion (See section 3.2), offering a unique probe of electron density fluctuations across different spatial and temporal scales [2]. Thus, pulsar timing provides valuable insights into both fundamental physics and the structure of the ISM. That is a whole vast field of study with pulsars.

Pulsars also offer a natural probe of the ISM, thanks to the dispersion of radio waves as they propagate through ionized gas. Lower-frequency components of the signal are delayed more than higher-frequency ones — a phenomenon quantified by the *dispersion measure* (DM), defined as:

$$DM = \int_0^d n_e(l) dl \quad (3)$$

corresponding to the integrated electron density on the line of sight between the pulsar and earth. where $n_e(l)$ is the electron density along the line of sight. This delay Δt between two frequencies ν_1 and ν_2 (in MHz), with a dispersion constant of $D = 4.1488064239(11) \times 10^3 \text{ MHz}^2 \cdot \text{s} \cdot \text{cm}^3 \cdot \text{pc}^{-1}$, is given by:

$$\Delta t \simeq D \cdot DM \left(\frac{1}{\nu_1^2} - \frac{1}{\nu_2^2} \right) \text{ (ms)}. \quad (4)$$

This effect must be corrected (de-dispersed) in data processing pipelines to recover the intrinsic pulse shape (see Section 3.1).

Ongoing surveys, especially at low radio frequencies, aim to expand the known pulsar population, particularly in the region of long-period or low-luminosity objects, which are often missed by traditional high-frequency searches. These efforts help constrain population models and test theories of neutron star evolution, magnetic field decay, and gravitational physics.

3 NenuFAR Telescope and detection pipelines

The **New Extension in Nançay Upgrading LOFAR**, is a very large low-frequency radio telescope in France, which is among the most powerful in the world in its operating between 10 MHz and 85 MHz with high sensitivity across its band [19]. This frequency range corresponds to the lowest spectral "window" that is observed from the surface of Earth.

The telescope consists of numerous dual-polarization dipoles grouped into hexagonal mini-arrays (MA); each MA contains 19 antennas, and the full core currently has dozens of MAs combined to form a large phased array capable of both beamforming and interferometry [19]. The telescope has a flexible modes of observation: by beamforming or imaging.

The main scientific researches done with NenuFAR are:

- Study of pulsars at low frequency
- Study and detection of exoplanets
- Detection of radio signal of the "Cosmic Dawn" (the beginning of the universe)

For further information on the LOFAR network, see Appendix A

3.1 Why use NenuFAR for low frequency observations?

Observing pulsars at low radio frequencies (below 100 MHz) offers unique scientific advantages. Pulsars typically have steep radio spectra, meaning their flux density increases at lower frequencies. As a result, many pulsars are brighter and more easily detectable in this band. Moreover, low frequencies allow for the detection of long-period and low-luminosity pulsars that may be missed at higher frequencies. These observations also enhance studies of pulse morphology and propagation effects such as interstellar scattering, which are more pronounced in this regime.

However, radio observations at low frequencies face several challenges that complicate the detection of pulsars. The interstellar medium introduces three major physical effects that affect the observed signal.

Firstly, the **dispersion measure** arises from the presence of free electrons in the ionized plasma of the ISM. It introduces frequency-dependent time delays, scaling with ν^{-2} according to equation 4 as discussed previously in section 2. This delay causes a stretching of the pulse across the frequency band, leading to a phenomenon called DM smearing. NenuFAR addresses this with high spectral resolution, dividing the 75 MHz band into 24576 channels of ~ 1.526 kHz. This effect can be corrected through a technique called **coherent dedispersion**, a direct correction on the raw signal voltage [7].

Secondly, **scatter broadening** occurs due to multi-path propagation caused by small-scale fluctuations in the ISM. The radio waves take slightly different paths to reach the observer, leading to temporal broadening of the pulse and a reduction in the signal-to-noise ratio (S/N). Unfortunately, this effect becomes more significant at lower frequencies and is not correctable (approximately $\tau_s \propto \nu^{-4}$). Thus, NenuFAR is most sensitive to nearby pulsars with low scattering environments.

Thirdly, **scintillation** arises from electron density irregularities in the ISM, which distort the pulsar wavefront and create an interference pattern in time and frequency—similar to stellar twinkling. It appears in two forms: the **Diffractive scintillation** is a rapid signal variations over short timescales (minutes–hours) and narrow frequency bands, caused by small-scale turbulence; and the **Refractive scintillation**, a slower variations over longer timescales (days–weeks) and broader frequency bands, due to large-scale inhomogeneities. Both types can impact pulsar detectability, especially at low frequencies. However, for nearby pulsars and wide fractional bandwidths, scintillation may average out across observations.

NenuFAR’s high sensitivity to the propagation effects, wide bandwidth, larger beams and fine frequency resolution, brighter pulsars at low frequency, coherent and incoherent dedispersion, flexible signal processing, make it particularly well-suited to enhance pulsar detection at low frequencies. Rather than avoiding these ISM effects, NenuFAR helps us study them more precisely, turning propagation effects into scientific tools.

You can see more details in M. Brionne’s paper [4]

3.2 Pulsar Observations

Pulsars emit beams of radio waves, sweeping across the sky as the it rotates. When these beams cross Earth’s line of sight, they produce periodic radio pulses that can be detected by ground-based telescopes. In the case of NenuFAR, which operates in the low-frequency range (10 to 85 MHz), these pulses appear as sharp, mountain-shaped peaks in the averaged pulse profile.

For pulsar observations, NenuFAR uses a custom-built pulsar backend, LUPPI (Low frequency Ultimate Pulsar Processing Instrumentation), which digitizes incoming signals at 200 MHz sampling rate, yielding time resolution around $5.12 \mu\text{s}$ [3]. These digital streams are channelized into 1024 sub-bands (195 kHz each), coherently beamformed with full polarization: Stokes I: Total intensity, Q: Linear polarization (horizontal vs vertical), U: Linear polarization (diagonal $\pm 45^\circ$), V: Circular polarization; enabling dynamic spectrum, folded profile, and single-pulse modes. The pipeline performs real-time coherent dedispersion on the known pulsar (see subsection 3.1), crucial at such low frequencies to preserve pulse integrity [3].

This setup allows NenuFAR to detect and precisely time known pulsars, as well as train to conduct blind searches for new ones, including millisecond and long-period pulsars. Its capabilities can lead to the discovery and characterization of numerous pulsars below 100 MHz, studies of fine pulse structure, and the monitoring of DM variations [3].

3.3 The Detection Pipeline of NenuFAR

The pulsar detection process at NenuFAR can be conceptually divided into two main stages: front-end data acquisition and back-end processing. This separation helps clarify the transformation of raw analog signals into scientifically usable pulsar candidates.

Front-End Acquisition (Upstream) NenuFAR consists of multiple mini-arrays of dual-polarization dipole antennas. These analog radio signals are first digitized at high time resolu-

tion using the LUPPI backend. A beamforming step then aligns the phases of the signals from different antennas, creating a coherent beam in a targeted direction on the sky.

After beamforming, the wideband signal is already channelized, that is a split into narrow frequency sub-bands using a polyphase filter bank. For NPBS, this results in 24756 channels of approximately 1.526 kHz bandwidth each across the 39–76 MHz frequency range. These channelized, high-time-resolution data form the input to the next stage of processing.

Back-End Processing (Downstream) It consist of an offline or real-time data processing using a dedicated pulsar search pipeline adapted from PRESTO for low-frequency observations.

The downstream processing of the NPBS data begins with RFI mitigation, where terrestrial sources of interference are statistically identified and flagged to minimize contamination in the data. Following this, dedispersion is applied to correct for the frequency-dependent delays induced by the interstellar medium

$$\Delta t \quad (\text{Eq4}),$$

quantified by the dispersion measure (DM). This involves trial dedispersions over a wide range of DM values to account for unknown source distances.

Next, a periodicity search is carried out using Fast Fourier Transform (FFT)-based methods to uncover periodic signals from potential pulsars. For unidentified sources, this search spans a broad set of trial periods and DMs, and includes acceleration searches to detect pulsars in binary systems. Finally, candidate visualization is performed for the most promising signals by generating diagnostic plots, including folded pulse profiles, signal-to-noise (S/N) versus DM curves, and time-frequency diagrams, which are manually inspected to confirm pulsar candidates.

This pipeline enables the detection of both known and new pulsars in a highly dispersed, noisy, and low-frequency regime, demonstrating NenuFAR’s capability for blind pulsar surveys.

However, in a blind search like the NPBS, the period and DM are initially unknown, so a coherent dedispersion cannot be applied. Instead of folding, the processing pipeline scans over a wide range of trial dispersion measures and trial spin periods to identify periodic signals. Promising candidates are visualized through diagnostic plots showing the pulse profile, S/N versus DM, and other properties for further inspection. See section 8 for the details.

4 NPBS Observing Strategy and Data Format

The survey aims to uncover new pulsars with steep radio spectra (spectral index $\alpha \sim < -3.2$) and very low turnover frequencies, below 85 MHz. A detection in this frequency regime would help constrain the population of low-frequency pulsars, while a non-detection would provide valuable information on the limits of their existence in this parameter space.

The current survey phase focuses on the sky above +39° declination, covering the frequency range from 39 MHz to 76 MHz. The survey targets a total of 7692 distinct sky pointings, each with a beam radius of approximately 1.5° and an integration time of 27 minutes. This strategy ensures a near-complete sky coverage of 98%.

Data acquisition began in August 2020 and ended in July 2022 after 2 phases of data acquisition. The survey was a total of 7664 pointings, 960 hours of limited time with 27 min per observation. The recorded time-series data are processed using a modified version of the PRESTO pulsar search pipeline, adapted for low-frequency observations. Due to the effects of scattering and limited time resolution, the search is restricted to pulsars with spin periods between 30 ms and 30 s, and dispersion measures (DM) between 1 and 70 pc cm⁻³.

To validate the setup and pipeline performance, 24 known pulsars have been re-observed in the processed dataset. Out of these, seven pulsars were successfully re-detected, with DMs ranging from 5 to 42 pc cm⁻³ and periods between 40 ms and 3.5 s. Most of these detections correspond to the brightest sources in the sample. Notably, one detection is attributed to a favorable scintillation effect. From the faintest detection, the minimum required signal-to-noise ratio was estimated to be $\sigma = 4.8$, corresponding to a flux density threshold of approximately 6.9 mJy in the coldest regions of the sky.

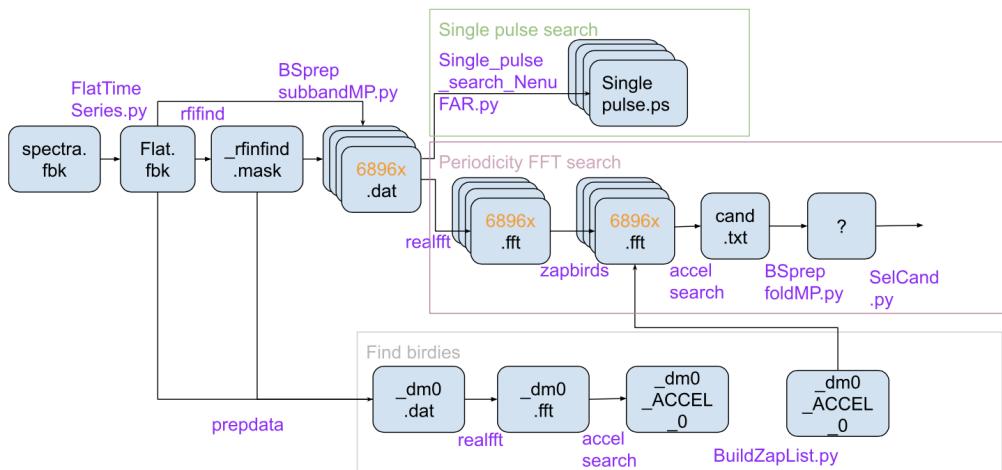


Figure 6: The BS Flowchart that explains the pipelines of NenuFAR data processing. *Flowchart by Cherru NG on github.*

5 The data format used for this study

In this section the focus will be on the data format used in my internship work.

5.1 Task 1: ATNF vs NenuFAR PSR

The preparation of the data for the comparison between the ATNF catalog [11] from the PSRQPY package [15], with 4343 already discovered pulsars, and the re-detected pulsars from NenuFAR, with 182 detected pulsars. This Python library contains all the data from the ATNF pulsar catalog in a query, where you can choose to use any of the entries. To complete the study, the team shared with me the unpublished detected pulsars from the NenuFAR telescope, then I cross-joined both data to fill the empty entries in the NenuFAR PSR (pulsar) data. See section 6 for the details of the task.

5.2 Interference

To detect the Bad Channels where there is the most RFI, I used some `.mask` and `mask_zap_chans.txt` files of flags, from already processed observation data with the `rfifind`, that are the output of this code. There was a total of 94 files to analyze. The file is in the format of **MASK_BS_ObsID_RAJD±DECJ_Date_Time**
_Pointing. More information about dealing with the RFIs in section 7

```
#output: /home/nour/Nancay/Rfifind-Masks/MASK_BS_2312_122512+744239_20230319_000034_2/mask_zap_chans_3.txt = filepath
#-----subdir_path-----file-----
#-----file_path-----
```

Figure 7: Format of the `mask_zap_chans.txt` files, giving all the information about a pointing.

5.3 Task 3: PSR search

To find new pulsars, I had to look into 300+ signal profile PDFs from years of observations, processed in binary files that takes a month to be completed for each run of one month data like 202202 (YYYYMM). Let's say the code name to give us the profile PDFs is `find_pulsars.py`, you run the code with `argparse` package in python to give an inline command, as »`python find_pulsars.py -month202202`, from which monthly observation signals you want the generated PDFs. All of that is done remotely on the terminal of the computer **baudroie**. I will explain further the code in Section 8.

5.4 Task 4: rfifind PRESTO

Working remotely on the terminal of **nancep7** in Nançay observatory, I have access to the PulsaR Exploration and Search TOolkit **PRESTO** [16], with which I can examine raw data, read files and search for RFIs. This tool helps in the whole study of the pulsar and candidates, from detecting the signals, search for specific candidate data, single pulse or periodic ect. It is a complete pipeline to follow on a candidate. I mainly worked on `-rfifind`, detecting many RFI in the signals. See Section 9 for more.

6 ATNF vs NenuFAR pulsar catalog

For this first task, the goal was to get familiar with pulsar terminologies, definitions and field of work. I got a good bibliography start with the "Handbook of Pulsar Astronomy" by M. Kramer [10], who was the supervisor of my internship supervisor. I was able to directly start the application because of my previous trainings in the course of Big data & Database, 3D modelisation and Machine Learning.

After crossing the two catalogs to get the main variables from the main ATNF catalog with the J2000 names (pulsar typical name format (J+RA+DEC like J0030+0451), I get a complete panda data frame of NenuFAR PSR to start my data cleaning, codes and visualization. The chosen entries to focus on are: Jnames **PSRJ**, Right ascension **RAJ**, Declination **DecJ**, period **P**, **DM**, **dotP**, frequency **F**, galactic coordinates galactic longitude **GL** and galactic latitude **GB**,

XX YY ZZ in X-Y-Z galactic coordinates, pulse width at 50 percent **W50**, and at 10% **W10**, mean flux densities at different frequencies **S400** for $f=400$ MHz. Look in Appendix B for each variable definition.

This task is mainly based on experimental learning, I was able to discover the pulsar theory and learn it from only plots and dataframes. Each visualization gave me an insight on how NenuFAR telescope works, while emphasizing on the advantage of using this low frequency telescope and its limitations. Therefore the comparison between ATNF PSRCAT - in violet - and NenuFAR PSR - in cyan - is important.

RAJD vs DECJD This sky map highlights the spatial distribution of known pulsars from the ATNF catalog compared with those observed by NenuFAR. The majority of pulsars lie along the Galactic disk, where stellar density is high. NenuFAR, being located in the Northern Hemisphere, is unable to observe declinations below approximately -25° , hence it lacks sensitivity to sources in the Southern sky. This geographic limitation naturally restricts its observable pulsar population, missing part of the Galactic plane.

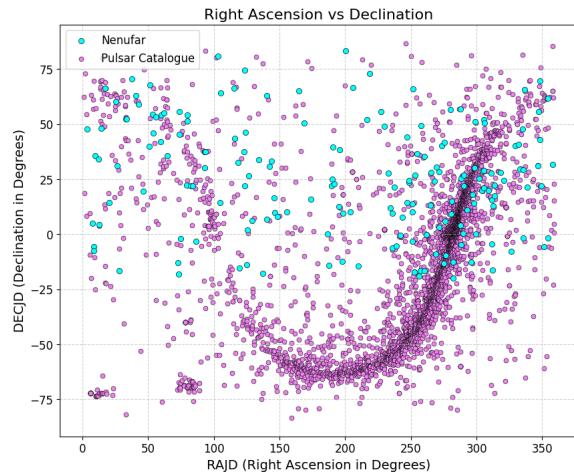


Figure 8: Population of both NenuFAR (cyan) and ATNF (violet) scattered based on Right Ascension vs declination. The galactic plane curvature is visible.

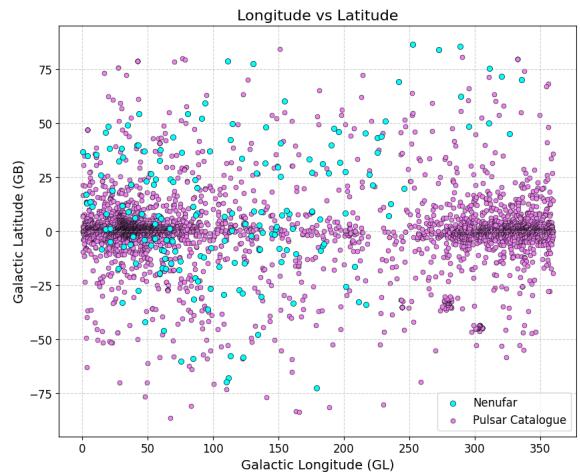


Figure 9: Population of both catalogs presented in galactic coordinates.

Galactic Longitude vs Latitude This diagram shows the pulsar distribution in galactic coordinates. The NenuFAR catalog exhibits difficulties of detection in Galactic longitude beyond $+250^\circ$. This corresponds to sky regions opposite to NenuFAR's location on Earth, from which signals are not easily detectable to the telescope. Additionally, the galactic latitude distribution further emphasizes NenuFAR's sensitivity to sources near the Galactic plane.

3D Galactic XYZ Coordinates The three-dimensional positions derived from distance estimates show that NenuFAR can only detect pulsars within a limited volume around the Sun. In contrast, the ATNF catalog spans a much larger region of the Galaxy. This reflects NenuFAR's sensitivity constraints and its focus on nearby, bright pulsars.

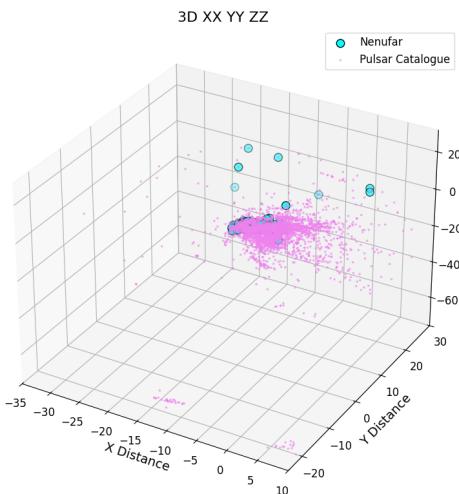


Figure 10: Population of PSR in 3D.

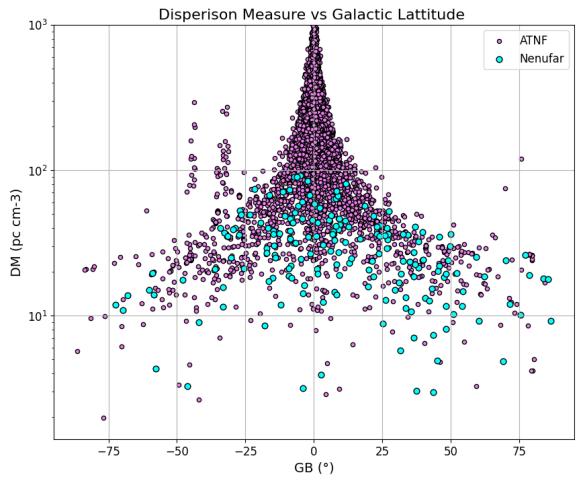


Figure 11: DM vs GB to show that PSR of ATNF are the closest to the Galactic Center.

Dispersion Measure vs Galactic Longitude In this plot, pulsars with higher DM tend to lie close to the Galactic plane. This is expected due to the increased presence of dense ionized interstellar material along the plane, which increases the DM. Compared to ATNF, NenuFAR’s detections are limited to DMs below ~ 100 pc cm⁻³. High electron turbulence in the ISM hides pulsars at NenuFAR frequencies, thus resulting in only sampling a nearby and less obscured population.

Period vs Period Derivative ($P-\dot{P}$) Diagram The $P-\dot{P}$ diagram is instrumental in classifying pulsar populations and studying their evolution. It shows distinct populations: normal pulsars, young pulsars with high \dot{P} (top right), and recycled millisecond pulsars (bottom left). NenuFAR’s limited coverage may reduce its sensitivity to extreme or exotic pulsars within this diagram. That is why the focus for NPBS is the population of **normal pulsars** and **slow pulsars**(See section 2 by M.Brionne [4]).

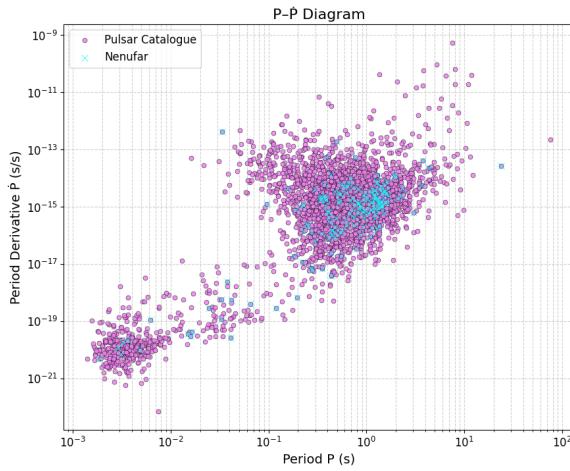


Figure 12: $P - \dot{P}$ diagram of both catalog, showing that NenuFAR mostly detects young and normal pulsars.

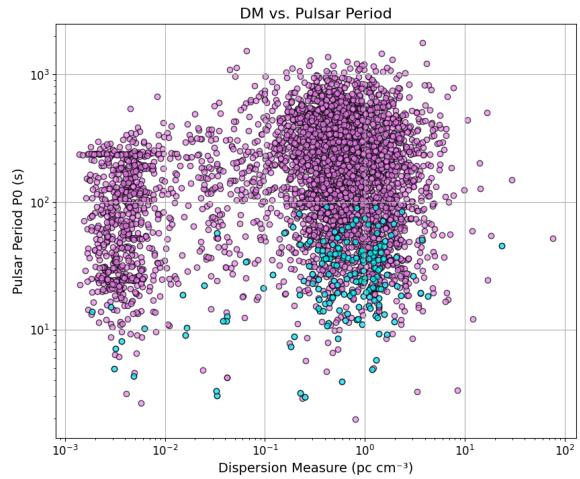


Figure 13: Period vs DM, NenuFAR can't detect more than 100 DM and 10 s.

Period vs Dispersion Measure This plot helps in understanding the interstellar medium's impact on pulsar detection. High DM values generally indicate either distant pulsars or sources behind dense ionized regions. NenuFAR detects pulsars mostly below 100 pc cm^{-3} and within a limited period range, in contrast to the broader range seen in ATNF. Older pulsars with longer periods may have lower DM values if they are located outside the Galactic plane.

Histogram of DM distribution The DM histogram reflects the distribution of pulsar distances and intervening medium. The ATNF catalog spans a wide DM range up to over 1750 pc cm^{-3} , while NenuFAR is limited to values around 100 pc cm^{-3} , reinforcing that NenuFAR primarily detects local, nearby pulsars.

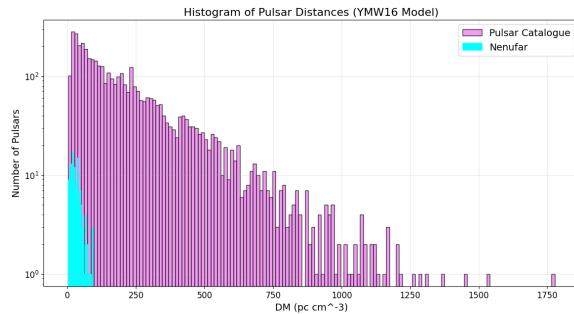


Figure 14: Histogram of DM distribution of the pulsars.

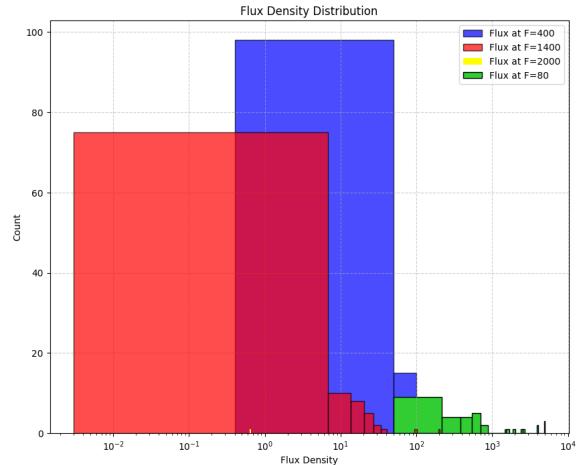


Figure 15: Flux density, the lower the frequency the brighter the pulsar, green represents the highest flux and lowest frequency.

Flux Density Flux density measurements available from NenuFAR indicate how bright pulsars appear at its low observing frequency. In general, pulsars are brighter at lower frequencies due to their negative spectral index. This makes NenuFAR well-suited for detecting pulsars with steep spectra, although detailed flux comparisons with ATNF are limited.

Pulse Width and Width vs Period Pulse width (W_{50}) tends to increase with pulsar period. NenuFAR observations confirm this trend but only within a narrow period range. While wider pulses are seen in slower pulsars, the duty cycle (fraction of the period during which the pulse is emitted) is inversely related to the period. This anti-correlation is consistent with theoretical expectations.

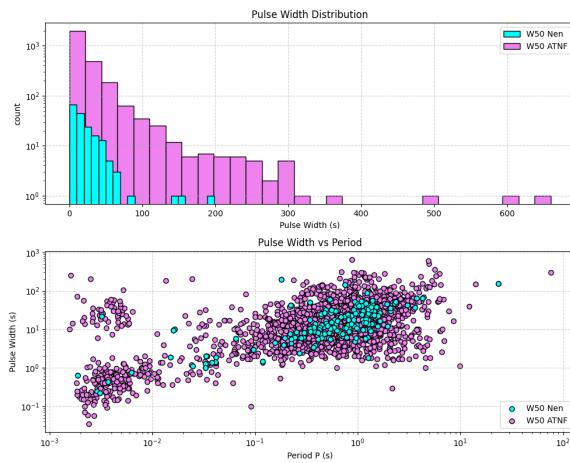


Figure 16: Pulse width distribution W50 of the pulsars.

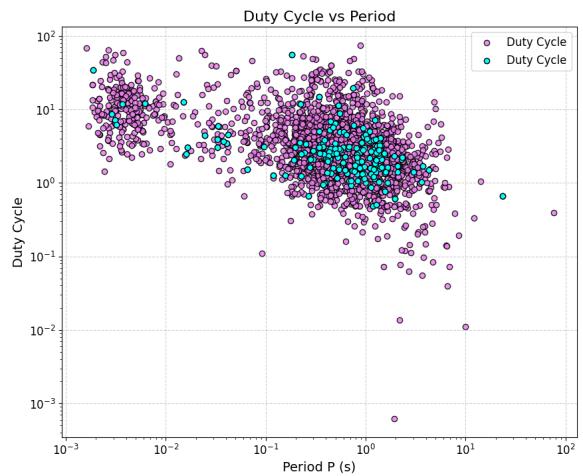


Figure 17: Duty cycle pulsar vs Period, validating again the PSR population of NenuFAR.

Duty Cycle vs Period The duty cycle, computed as $\frac{W_{50}}{P_0} \times 100$, indicates how long the pulse is active per rotation. Short-period pulsars often have higher duty cycles. NenuFAR’s data supports this trend, though it primarily includes nearby pulsars with moderate to low periods. Additionally, we include a histogram of the pulse widths at 50% intensity (**W50**) and 10% intensity (**W10**) to visualize their distributions across the observed pulsars. This helps identify typical width ranges and outliers, assess observational resolution limits, and better understand how pulse duration varies with pulsar type and geometry. Analyzing the pulse width distribution also supports the interpretation of duty cycle trends, since narrower pulses correspond to smaller duty cycles.

Cumulative Distribution Function (CDF) The CDF plots illustrate the cumulative fraction of pulsars below a given value of a parameter such as pulse width. These comparisons between NenuFAR and ATNF datasets reveal that NenuFAR pulsars are generally skewed toward lower values due to detection limitations and selection effects. It is only higher for the DM to show the precision of the detection of NenuFAR.

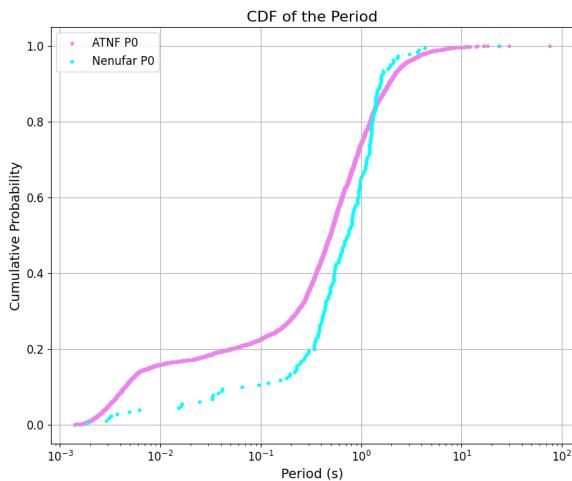


Figure 18: CDF Pulsar Period, NenuFAR coincide with ATNF detections only around young/normal pulsars period (around 1 s and less than 10 s).

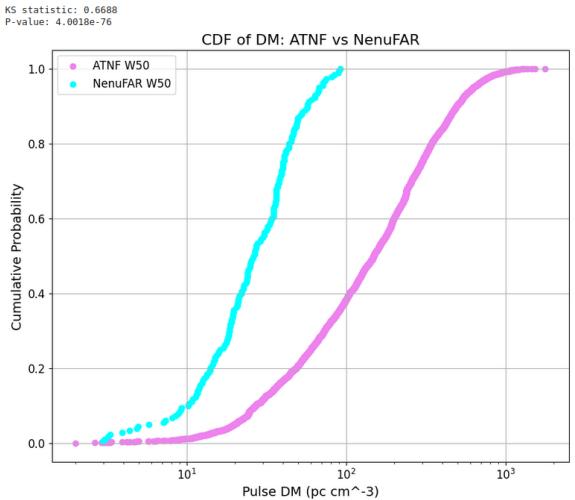


Figure 19: KS test values of NenuFAR and ATNF with CDF of DM. The P-value: 4.0018×10^{-76} , meaning NenuFAR is not sampling the same population as ATNF.

Kolmogorov–Smirnov (KS) Statistical Test To statistically assess whether NenuFAR and ATNF pulsar populations are drawn from the same distribution, a two-sample KS test is used. It quantifies the maximum difference between their CDFs and provides a p -value to evaluate significance. A small p -value would indicate that NenuFAR detects a statistically distinct subset of pulsars, shaped by its instrumental and observational constraints. Given that NenuFAR detects only a subset of all pulsars (due to its low-frequency range, sky coverage limits, sensitivity to closer/louder sources), we expect the KS test to return low p -values (< 0.05), indicating that NenuFAR and ATNF samples are drawn from different distributions for most parameters (especially DM, P_0 , flux, etc.).

7 Handling the Radio Frequency Interference

The second major task for improving pulsar detection was mitigating radio-frequency interference (RFI) from terrestrial or cosmic sources in order to clean the observational data. Using the PRESTO -rfifind tool, adapted for low-frequency searches, RFI was identified after initial preprocessing. This process generates `.mask` files, which form the basis of the present analysis.

The raw `mask_zap_chans.txt` files, stored in subdirectories named with the prefix MASK_BS followed by the observation ID, RA/DEC coordinates, date/time, and beam number (Figure 7), contain flagged or “bad” channels. I wrote a Python routine to scan all observation folders, check if the `mask_zap_chans.txt` file was empty or not, and then proceed to process the data.

The initial analysis focused on organizing the mask data and producing time-series visualizations to test whether RFI occurrence varied over time — for example, between day and night or across seasons. As expected for a radio astronomy site, a significant time-dependent trends were found. That is why the observations are done at night-time, from 11 PM to 9 AM next day, rather than the day-time. Nançay site is the “off-grid” location, which is largely free from

mobile or satellite service, though we did check for possible human-related interference during the day.

A second investigation was prompted by variations in compact flag counts between two months. I created a time-dependent plot for each pointing divided into four columns, each corresponding to one telescope beam, since NenuFAR has 4 beams. The telescope does not necessarily observe with all 4 beams at a time, so 1 pointing can have at least 2 beams and maximum 4. Since some RFI is always present, the absence of RFI in certain beam might indicate a technical problem; or just shows a part from an observation pointing with less RFI (it is more likely). Later, the specialists in Nançay did a check up on the beams to fix small issues and improve a pointing observation.

The main outcome of this stage was a statistical characterization of flagged channels. By aggregating counts from all `mask_zap_chans.txt` files and converting them to percentages, we identified the most frequently flagged channels — candidates for permanent exclusion. This list was generated for multiple thresholds (e.g., 70% → 231 channels, 60% → 236 channels, 50% → 241 channels, 40% → 246 channels), all representing less than 1% of total channels (See appendix C). Such a small fraction minimizes loss of useful data, since even an isolated good channel between two bad ones could contain valuable signal.

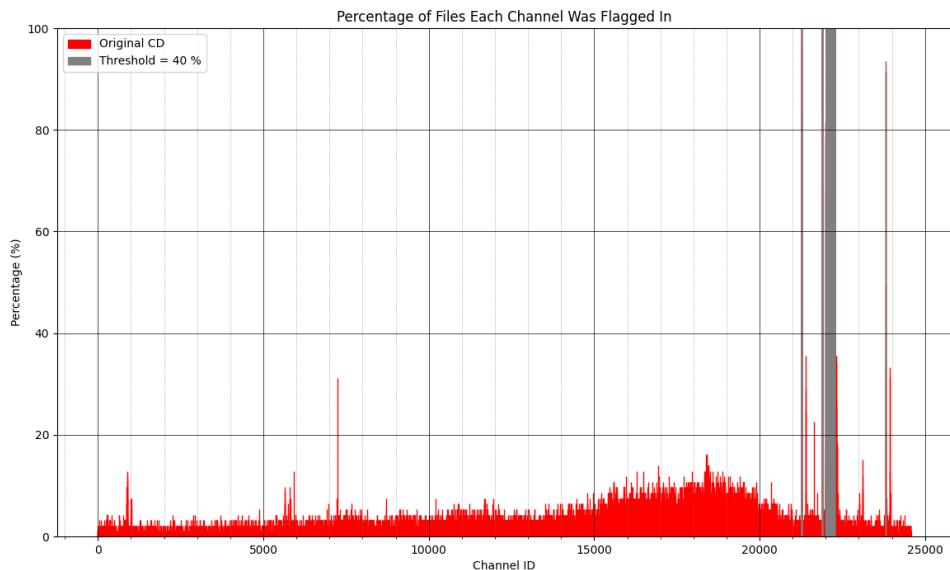


Figure 20: Statistical study of flags, emphasizing on the bad channels in gray. The gray area represents the flags that peak more than 40%.

We ultimately chose a conservative threshold of 40%, corresponding to 246 recurrently flagged channels, which will be excluded during the `-rfifind` stage (Section 9). This step should reduce persistent RFI and improve focus on genuine pulsar signals.

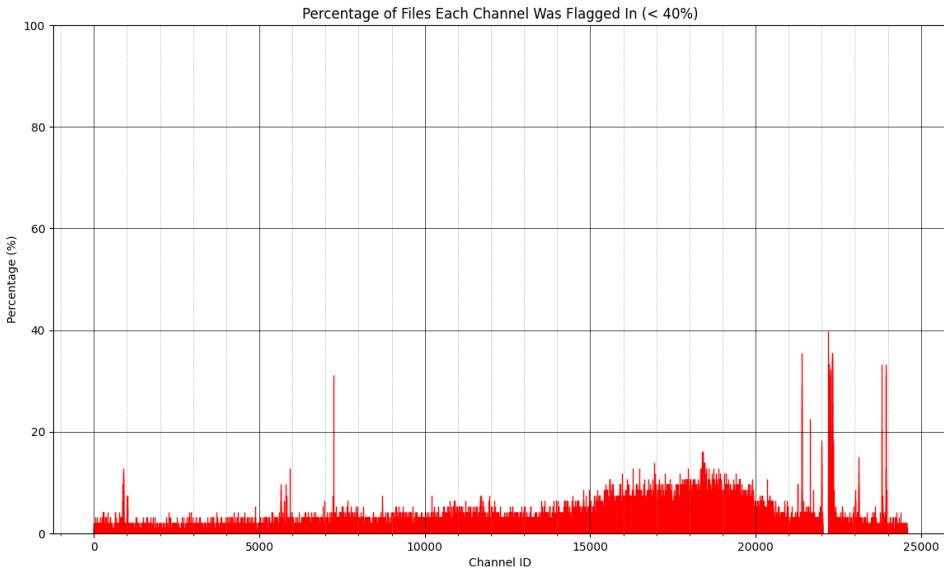


Figure 21: Masked flagged at threshold 40%.

8 Pulsar Search

The third part of my work was to jump into the pulsar search, working remotely on another Baudroie, a computer that holds an insanely large amount of data from blind radio observations. During this task, I learned to use the code developed by Mark Brionne, the previous doctoral student who worked on the NPBS, becoming familiar with both its advantages and limitations.

The goal of this task was to optimise Mark’s code and update it to filter more efficiently, perform better matching of observation signals with known pulsars, identify and remove RFI/Noise from the candidate (`Cand`) files, and finally inspect the remaining candidates to find a potential new pulsar in low frequency, as predicted by the NPBS simulations of M. Brionne [4].

In brief, the search pipeline processes all signal characteristics from a given month (specified via `argparse` commands), analyses over 30,000 observations, filters based on selection criteria for pulsar candidates, and outputs PDF files into two folders: one for Known PSRs and one for Candidates. The search for a new pulsar then involves manually inspecting each pulse profile in the Candidates folder to decide whether it is a true pulsar signal or RFI. Typically, these folders contain no more than 40–50 profiles, which indicates the code already performs well.

The main search parameters set by default in the pipeline are: periods up to 300 s, DM up to 300 pc cm^{-3} , harmonics up to the 20th order (an integer multiple or fraction of the pulsar’s true rotation frequency), and a signal-to-noise ratio that is neither too low ($\sigma < 3$) nor unrealistically high. All these ranges can be adjusted via `argparse`.

The updated version of the code, now referred to as “Nour’s code,” performs a double matching: first using Mark’s original matching method, and then a second pass that re-checks the Candidates folder against the ATNF pulsar catalog. The search is run directly from the terminal,

using filenames that contain all essential information, for example: BS_0716_084752+832657_20210208_223035_0_P163.16000_candDM47.840_ACCEL_0_ACCEL_Cand_2

The output of Nour's code includes a table of matched PSRs with RA/Dec, DM, period, and positional offsets to ensure the difference in sky position is reasonable (to avoid false matches from other pulsars' signals). Another table is produced for truly unknown candidates, listing their RA/Dec, period, and DM from the blind search. At the end, the user can choose whether to save the tables as csv files (y/n) and whether to remove missed PSRs from the candidate folder. This helps identify why the initial code failed to match them as known PSRs. The mismatch rate has never exceeded 7% across all candidate folders, which is acceptable but still worth reducing to avoid confusion during the search.

While inspecting candidate profiles from Mark's code, I noticed a recurring suspicious signal: a profile resembling a very good pulsar, but always with periods of either $P = 140.00059$ ms or $P = 139.99963$ ms, a very high S/N (10.5–17), a very close DM ($1\text{--}3.384 \text{ pc cm}^{-3}$), and no fixed location in the sky or time of occurrence. This clearly did not make physical sense. I generated time-dependent, beam-pointing plots, which confirmed the anomaly. Eventually, I classified it as RFI and decided to investigate it further.

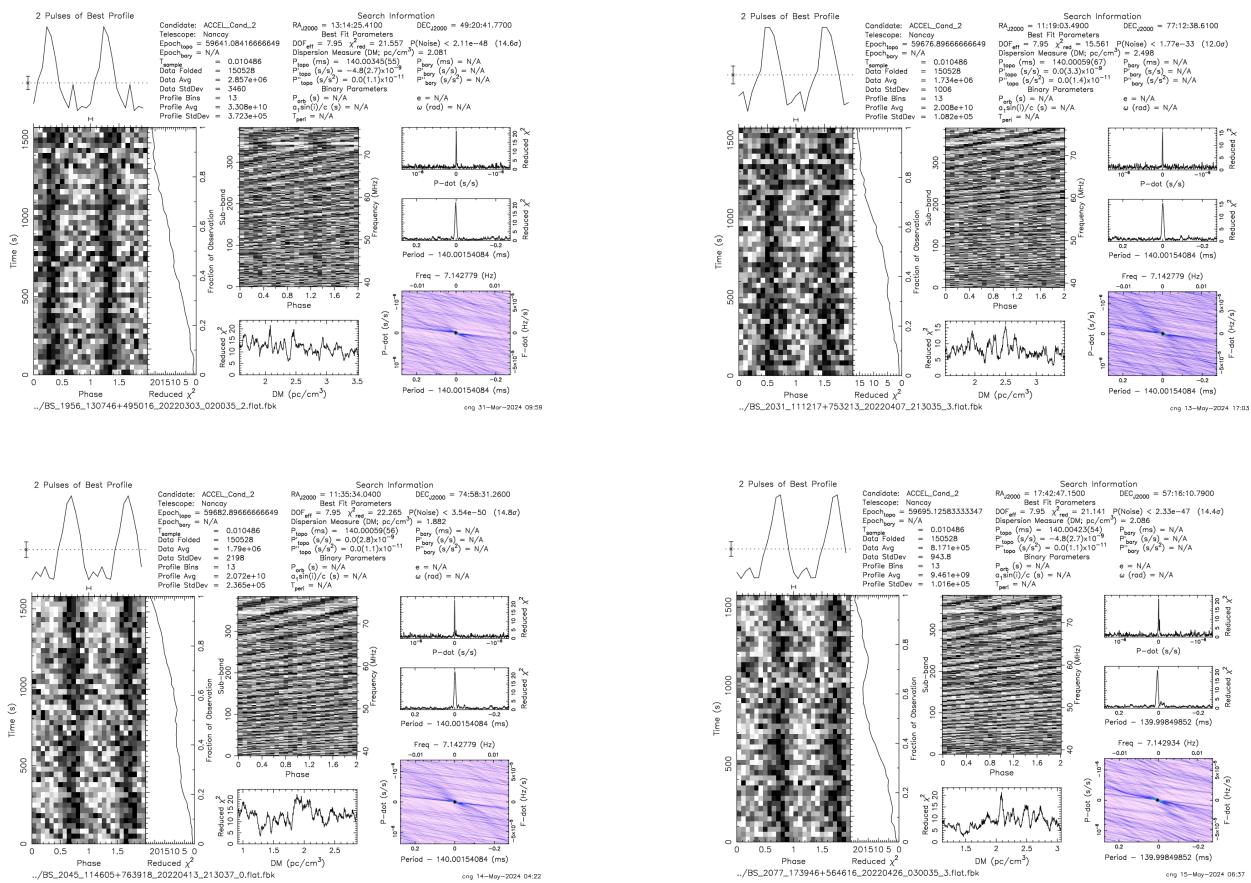


Figure 22: Four signal profiles of the $P = 140$ ms RFI.

After presenting these results in a NenuFAR TeleCon meeting, it was agreed that this signal was indeed RFI—most likely from a human-made source or a passing satellite. Specialists at the Nançay site are currently investigating. As a preventive measure, Nour's code now removes all $P = 140$ ms signals before processing. With these original and new filters combined, the pipeline produces cleaner data, speeds up processing, and reduces mismatches, false positives and RFI contamination.

Returning to the pulsar hunt, I examined more than 250 plots from about four years of observations, manually classifying each as either a potential candidate or not. I identified two interesting pulsars from the blind survey candidate profiles:

- BS_2256_192234+462319_20220701_003036_0_DM32.68_P170.01_DM32.66_ACCEL_0_ACCEL_Cand_10
- BS_1834_115346+450923_20220120_033036_0_P79.87000_candDM47.220_ACCEL_0_ACCEL_Cand_9

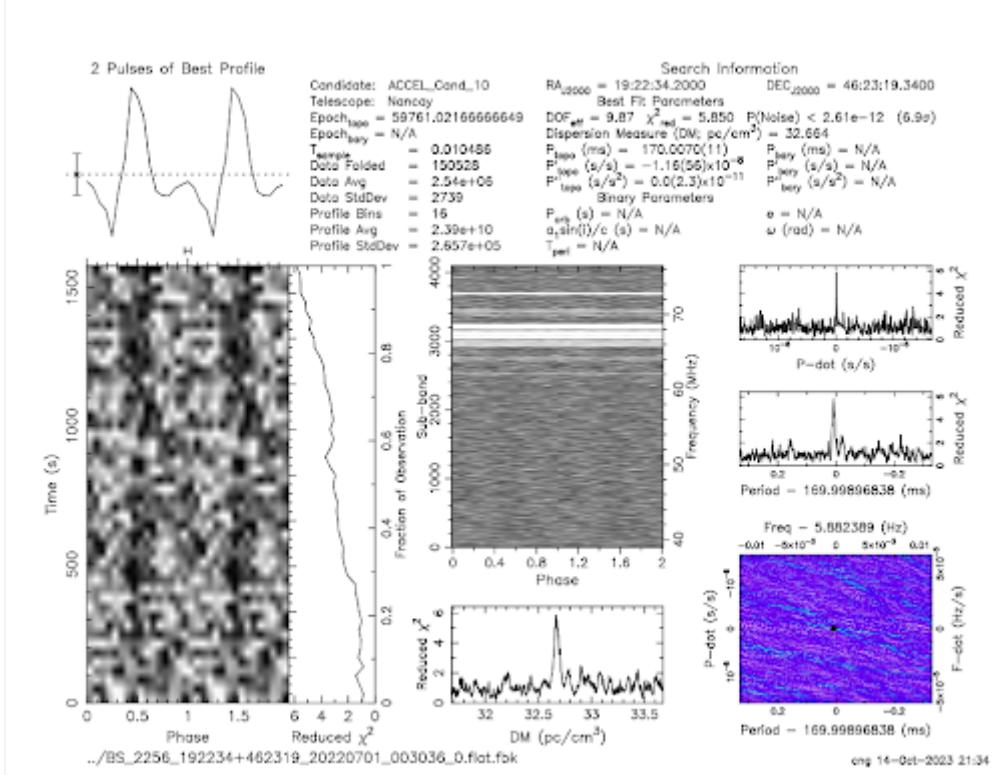


Figure 23: Potential candidate of $P = 170$ ms.

For the first one, the period $P = 170.0070$ ms is unusually precise, which may be suspicious—but it does not necessarily rule it out as a good candidate.

For the second one, the period, the DM, the P – \dot{P} diagram look all good but the signal looks a bit contaminated so it is hard to decide whether it is a good candidate as well.

We will need further follow-up observations, for all the potential candidates, to validate all our promising candidates and confirm any new pulsar discoveries.

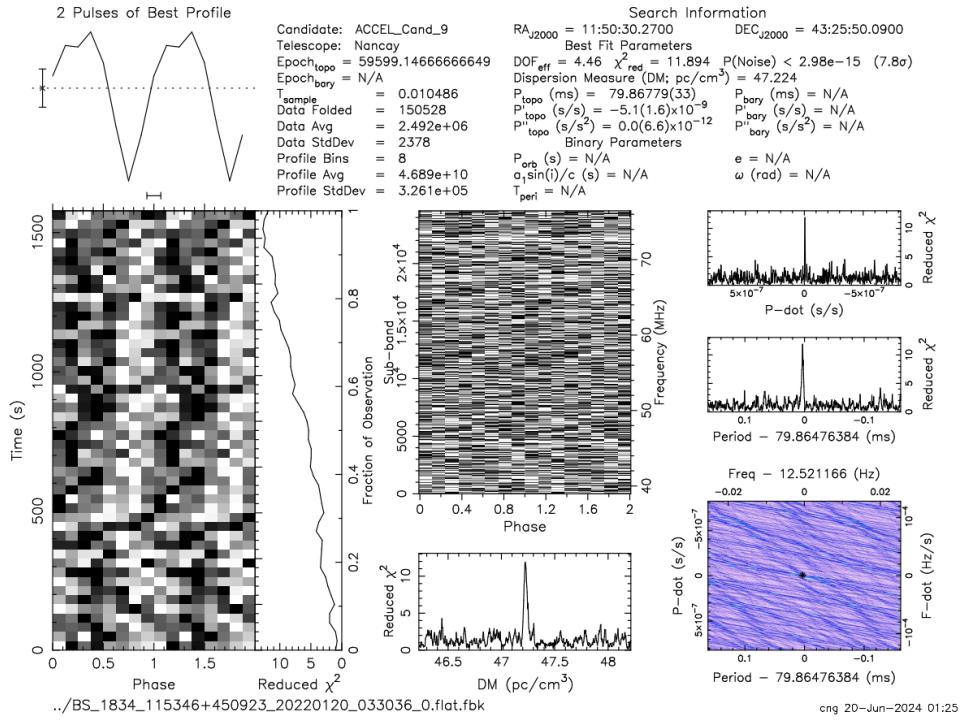


Figure 24: Potential candidate of $P = 79.8$ ms.

9 rfifind with PRESTO

PRESTO¹, developed by Scott M. Ransom, is a widely used software suite for pulsar data analysis, particularly tailored for searching and processing large volumes of radio observations. It offers a wide range of tools for dedispersion, Fourier-domain searches, acceleration searches, and RFI mitigation. Among these tools, `rfifind` is a crucial step to identify and remove RFI before any pulsar search pipeline.

The command `rfifind` scans the data for spectral and temporal anomalies characteristic of RFI and produces several output files. Two important options are:

- `-mask`: creates an `.mask` file that flags time-frequency intervals containing RFI. When applied during later processing, these flagged data are ignored in the analysis, but remain in the file for reference. This preserves most of the usable data.
- `-zapchan`: permanently removes specific frequency channels from the dataset, replacing their values so they do not contribute to any further analysis. This is more aggressive but ensures that persistent, narrow-band RFI does not leak into the pulsar search.

While `-mask` is ideal for broad, temporary interference, `-zapchan` is often preferable when dealing with stable, long-lived RFI sources contaminating fixed frequency channels, as it entirely eliminates their effect on the Fourier spectrum.

¹Pulsar Exploration and Search TOolkIt

Now, the list of the 246 channels from the RFI task in Section 7 is added to the `-zapchan` making option. The `rfifind` identifies by itself many RFIs, but the masking of 40% of the data is complementary, not necessary, to have better and cleaner data. For instance, that usually works best when the observation is quite contaminated.

The issue is that when the observation is not pretty much contaminated from the beginning, the `-zapchan` list will still remove the "bad channels" where in that case they might not have been identified as flags. SO we have to be careful on deciding whether to apply definitively the list or not, because it is important to stay conservative on the data.

Assuming the observation is contaminated, I did a statistical comparison to check how effective my input is. Logging in Nancep7 in Nançay, I ran `rfifind` at first with the usual commands, a second time with only my commands, and a third using both.

One can say there is an improvement but the issue lies on the beginning, about how contaminated or how much RFI was before removing. The resulting output file is in the Appendix D.

10 Preliminary Results and Findings

The preliminary results of the first task are presented here, with each paragraph referring to a specific parameter of comparison as outlined in Section 6. The analysis confirms that NenuFAR, while not detecting pulsars extremely deep due to its Galactic location limitations, excels in precision for (DM) measurements and achieves detections at lower signal-to-noise ratios (S/N) compared to other telescopes. This highlights its higher sensitivity in the low-frequency range, especially for sources that may remain undetected by higher-frequency instruments.

In terms of candidate selection, the pipeline now benefits from improved RFI mitigation. A new spectral profile of RFI was produced based on what was removed during processing. The RFI removal threshold was set at 40%, acknowledging that `rfifind` already performs a robust cleaning step. The resulting list of flagged channels is automatically added to the search pipeline for consistency across different observation sets. In addition, the persistent $P = 140$ ms signal—previously identified as spurious noise—has been entirely excluded from the search to prevent false positives.

Overall, the dataset is now cleaner, leading to faster and more reliable candidate selection. This not only reduces manual inspection time but also increases the chances of identifying genuine pulsar signals. With these improvements, future observations with NenuFAR can be relaunched with a more targeted strategy, focusing on the most promising regions of the parameter space and minimizing time lost to RFI or known noise artifacts.

11 Future work

The main objective of my master's internship is to search for a new pulsar at low frequencies, thereby testing and validating the predictions of the NPBS. The goal is to demonstrate that not only is it still possible to discover unknown pulsars in this frequency range, but also that

low-frequency observations remain highly relevant for pulsar astrophysics.

The NenuFAR pulsar census article, whose data I have been working with, is expected to be published soon. This dataset forms the foundation of the training on known pulsar, to improve the study of the data of the blind search, and offers a unique opportunity to explore the low-frequency sky with unprecedented sensitivity.

With the recent improvements in the pipeline and a reduced number of false candidates after RFI cleaning, we can now strategically reschedule blind survey observations with NenuFAR. Using the known locations and additional information obtained from candidate queries, follow-up observations can be planned with greater precision, focusing efforts on the most promising targets and maximizing the scientific return.

There is still some check ups to do and statistical studies before deciding on generalizing the work for all the past and future observations.

And finally after making a list of the best candidates, the next steps are to re-observe again in the specific sky location (RA and Dec) and DM, to detect again the signal and judge if it is a pulsar or not.

12 Conclusion

This internship has been a deep dive into the world of low-frequency pulsar hunting, from understanding the theoretical expectations of the NPBS to building and refining practical tools capable of handling the massive datasets from NenuFAR. Starting with the adaptation and optimization of Mark Brionne's original code, I developed an improved pipeline that not only better filters RFI and mismatched sources but also streamlines candidate classification, resulting in cleaner datasets and fewer false positives.

The combination of updated filtering criteria, targeted removal of persistent noise sources (such as the $P = 140\text{ ms}$ signal), and the integration of `rfifind` masks and channel zapping has significantly improved the efficiency and reliability of the search. The result is a more manageable candidate list, enabling precise and strategic follow-up observations of promising blind survey detections.

While the search has not yet yielded a confirmed new pulsar, several intriguing candidates have emerged that warrant additional follow-up with NenuFAR and complementary telescopes. This work also highlights the ongoing relevance of low-frequency observations, demonstrating that even in a sky already well-mapped at higher frequencies, there is still room for new discoveries.

In the bigger picture, the improved pipeline and RFI strategies developed here will remain useful beyond this project, contributing to the upcoming NenuFAR pulsar census and helping future searches operate more efficiently. The road to discovering new pulsars is never a straight line, but every iteration sharpens the tools, narrows the search, and brings us closer to hearing the faint, steady heartbeat of a new neutron star.

References

- [1] Gabriella Agazie, Akash Anumarlapudi, Anne M. Archibald, Paul T. Baker, Bence Bécsy, Laura Blecha, Alexander Bonilla, Adam Brazier, Paul R. Brook, Sarah Burke-Spolaor, Rand Burnette, Robin Case, J. Andrew Casey-Clyde, Maria Charisi, Shami Chatterjee, Katerina Chatzioannou, Belinda D. Cheeseboro, Siyuan Chen, Tyler Cohen, James M. Cordes, Neil J. Cornish, Fronefield Crawford, H. Thankful Cromartie, Kathryn Crowter, Curt J. Cutler, Daniel J. D’Orazio, Megan E. DeCesar, Dallas DeGan, Paul B. Demorest, Heling Deng, Timothy Dolch, Brendan Drachler, Elizabeth C. Ferrara, William Fiore, Emmanuel Fonseca, Gabriel E. Freedman, Emiko Gardiner, Nate Garver-Daniels, Peter A. Gentile, Kyle A. Gersbach, Joseph Glaser, Deborah C. Good, Kayhan Gültekin, Jeffrey S. Hazboun, Sophie Hourihane, Kristina Islo, Ross J. Jennings, Aaron Johnson, Megan L. Jones, Andrew R. Kaiser, David L. Kaplan, Luke Zoltan Kelley, Matthew Kerr, Joey S. Key, Nima Laal, Michael T. Lam, William G. Lamb, T. Joseph W. Lazio, Natalia Lewandowska, Tyson B. Littenberg, Tingting Liu, Jing Luo, Ryan S. Lynch, Chung-Pei Ma, Dustin R. Madison, Alexander McEwen, James W. McKee, Maura A. McLaughlin, Natasha McMann, Bradley W. Meyers, Patrick M. Meyers, Chiara M. F. Mingarelli, Andrea Mitridate, Priyamvada Natarajan, Cherry Ng, David J. Nice, Stella Koch Ocker, Ken D. Olum, Timothy T. Pennucci, Benetge B. P. Perera, Polina Petrov, Nihan S. Pol, Henri A. Radovan, Scott M. Ransom, Paul S. Ray, Joseph D. Romano, Jessie C. Runnoe, Shashwat C. Sardesai, Ann Schmiedekamp, Carl Schmiedekamp, Kai Schmitz, Levi Schult, Brent J. Shapiro-Albert, Xavier Siemens, Joseph Simon, Magdalena S. Siwek, Ingrid H. Stairs, Daniel R. Stinebring, Kevin Stovall, Jerry P. Sun, Abhimanyu Susobhanan, Joseph K. Swiggum, Jacob Taylor, Stephen R. Taylor, Jacob E. Turner, Caner Unal, Michele Vallisneri, Sarah J. Vigeland, Jeremy M. Wachter, Haley M. Wahl, Qiaohong Wang, Caitlin A. Witt, David Wright, and Olivia Young. The NANOGrav 15-year Data Set: Constraints on Supermassive Black Hole Binaries from the Gravitational Wave Background. *The Astrophysical Journal Letters*, 952(2):L37, August 2023. arXiv:2306.16220 [astro-ph].
- [2] J. Antoniadis, P. Arumugam, S. Arumugam, S. Babak, M. Bagchi, A.-S. Bak Nielsen, C. G. Bassa, A. Bathula, A. Berthureau, M. Bonetti, E. Bortolas, P. R. Brook, M. Burgay, R. N. Caballero, A. Chalumeau, D. J. Champion, S. Chanlaridis, S. Chen, I. Cognard, S. Dandapat, D. Deb, S. Desai, G. Desvignes, N. Dhanda-Batra, C. Dwivedi, M. Falxa, I. Ferranti, R. D. Ferdman, A. Franchini, J. R. Gair, B. Goncharov, A. Gopakumar, E. Graikou, J.-M. Grießmeier, L. Guillemot, Y. J. Guo, Y. Gupta, S. Hisano, H. Hu, F. Iraci, D. Izquierdo-Villalba, J. Jang, J. Jawor, G. H. Janssen, A. Jessner, B. C. Joshi, F. Kareem, R. Karuppusamy, E. F. Keane, M. J. Keith, D. Kharbanda, T. Kikunaga, N. Kolhe, M. Kramer, M. A. Krishnakumar, K. Lackeos, K. J. Lee, K. Liu, Y. Liu, A. G. Lyne, J. W. McKee, Y. Maan, R. A. Main, S. Manzini, M. B. Mickaliger, I. C. Nițu, K. Nobleson, A. K. Paladi, A. Parthasarathy, B. B. P. Perera, D. Perrodin, A. Petiteau, N. K. Porayko, A. Possenti, T. Prabu, H. Quelquejay Leclere, P. Rana, A. Samajdar, S. A. Sanidas, A. Sesana, G. Shaifullah, J. Singha, L. Speri, R. Spiewak, A. Srivastava, B. W. Stappers, M. Surnis, S. C. Susarla, A. Susobhanan, K. Takahashi, P. Tarafdar, G. Theureau, C. Tiburzi, E. van der Wateren, A. Vecchio, V. Venkatraman Krishnan, J. P. W. Verbiest, J. Wang, L. Wang, and Z. Wu. The second data release from the European Pulsar Timing Array - V. Search for continuous gravitational wave signals. *Astronomy & Astrophysics*, 690:A118, October 2024. Publisher: EDP Sciences.
- [3] L. Bondonneau, J.-M. Grießmeier, G. Theureau, I. Cognard, M. Brionne, V. Kondratiev, A. Bilous, J. W. McKee, P. Zarka, C. Viou, L. Guillemot, S. Chen, R. Main, M. Pilia, A. Possenti, M. Serylak, G. Shaifullah, C. Tiburzi, J. P. W. Verbiest, Z. Wu, O. Wucknitz, S. Yerin, C. Briand, B. Cecconi, S. Corbel, R. Dallier, J. N. Girard, A. Loh, L. Martin, M. Tagger, and C. Tasse. Pulsars with NenuFAR: Backend and pipelines. *Astronomy & Astrophysics*, 652:A34, August 2021. Publisher: EDP Sciences.
- [4] M. Brionne, J.-M. Grießmeier, I. Cognard, G. Theureau, L. Bondonneau, R. Gros, A. Loh, R. A. Main, I. P. Kravtsov, V. Zakharenko, V. I. Kondratiev, J. W. McKee, A. Possenti, C. Tiburzi, B. Cecconi, S. Corbel, J. N. Girard, and P. Zarka. The NenuFAR Pulsar Blind Survey (NPBS): I. Survey overview, expectations, and first redetections. *Astronomy & Astrophysics*, 693:A96, January 2025. Publisher: EDP Sciences.
- [5] T. Gold. Rotating Neutron Stars as the Origin of the Pulsating Radio Sources. *Nature*, 218:731–732, May 1968. ADS Bibcode: 1968Natur.218..731G.
- [6] Peter Goldreich and William H. Julian. Pulsar Electrodynamics. *The Astrophysical Journal*, 157:869, August 1969. Publisher: IOP ADS Bibcode: 1969ApJ...157..869G.

- [7] T. H. Hankins and B. J. Rickett. Pulsar signal processing. *Methods in Computational Physics*, 14:55–129, January 1975.
- [8] A. Hewish, S. J. Bell, J. D. H. Pilkington, P. F. Scott, and R. A. Collins. Observation of a Rapidly Pulsating Radio Source. *Nature*, 217:709–713, February 1968. ADS Bibcode: 1968Natur.217..709H.
- [9] G. Hobbs, A. Archibald, Z. Arzoumanian, D. Backer, M. Bailes, N. D. R. Bhat, M. Burgay, S. Burke-Spolaor, D. Champion, I. Cognard, W. Coles, J. Cordes, P. Demorest, G. Desvignes, R. D. Ferdman, L. Finn, P. Freire, M. Gonzalez, J. Hessels, A. Hotan, G. Janssen, F. Jenet, A. Jessner, C. Jordan, V. Kaspi, M. Kramer, V. Kondratiev, J. Lazio, K. Lazaridis, K. J. Lee, Y. Levin, A. Lommen, D. Lorimer, R. Lynch, A. Lyne, R. Manchester, M. McLaughlin, D. Nice, S. Osłowski, M. Pilia, A. Possenti, M. Purver, S. Ransom, J. Reynolds, S. Sanidas, J. Sarkissian, A. Sesana, R. Shannon, X. Siemens, I. Stairs, B. Stappers, D. Stinebring, G. Theureau, R. van Haasteren, W. van Straten, J. P. W. Verbiest, D. R. B. Yardley, and X. P. You. The International Pulsar Timing Array project: using pulsars as a gravitational wave detector. *Classical and Quantum Gravity*, 27:084013, April 2010. Publisher: IOP ADS Bibcode: 2010CQGra..27h4013H.
- [10] D. R. Lorimer and M. Kramer. *Handbook of Pulsar Astronomy*, volume 4. 2004.
- [11] R. N. Manchester, G. B. Hobbs, A. Teoh, and M. Hobbs. The ATNF Pulsar Catalogue. *The Astronomical Journal*, 129(4):1993–2006, April 2005. arXiv:astro-ph/0412641.
- [12] M. C. Miller, F. K. Lamb, A. J. Dittmann, S. Bogdanov, Z. Arzoumanian, K. C. Gendreau, S. Guillot, A. K. Harding, W. C. G. Ho, J. M. Lattimer, R. M. Ludlam, S. Mahmoodifar, S. M. Morsink, P. S. Ray, T. E. Strohmayer, K. S. Wood, T. Enoto, R. Foster, T. Okajima, G. Prigozhin, and Y. Soong. PSR J0030+0451 Mass and Radius from NICER Data and Implications for the Properties of Neutron Star Matter. *The Astrophysical Journal Letters*, 887(1):L24, December 2019. arXiv:1912.05705 [astro-ph].
- [13] F. Pacini. Rotating Neutron Stars, Pulsars and Supernova Remnants. *Nature*, 219:145–146, July 1968. ADS Bibcode: 1968Natur.219..145P.
- [14] J. D. H. Pilkington and J. F. Scott. A survey of radio sources between declinations 20° and 40°. *Memoirs of the Royal Astronomical Society*, 69:183, January 1965. ADS Bibcode: 1965MmRAS..69..183P.
- [15] M. Pitkin. psrqpy: a python interface for querying the ATNF pulsar catalogue. *Journal of Open Source Software*, 3(22):538, February 2018.
- [16] Scott Ransom. PRESTO: Pulsar Exploration and Search TOolkit. Astrophysics Source Code Library, record ascl:1107.017, July 2011.
- [17] Scott M. Ransom. New search techniques for binary pulsars, 2001. original-date: 2009-09-29T21:35:23Z.
- [18] S. Sanidas, S. Cooper, C. G. Bassa, J. W. T. Hessels, V. I. Kondratiev, D. Michilli, B. W. Stappers, C. M. Tan, J. van Leeuwen, L. Cerrigone, R. A. Fallows, M. Iacobelli, E. Orrú, R. F. Pizzo, A. Shulevski, M. C. Toribio, S. ter Veen, P. Zucca, L. Bondonneau, J.-M. Grießmeier, A. Karastergiou, M. Kramer, and C. Sobey. The LOFAR Tied-Array All-Sky Survey (LOTAAS): Survey overview and initial pulsar discoveries. *Astronomy & Astrophysics*, 626:A104, June 2019. Publisher: EDP Sciences.
- [19] Philippe Zarka, Michel Tagger, Laurent Denis, J. Girard, Alexander Konovalenko, Marcellin Atemkeng, Morel Arnaud, Sylvain Azarian, M. Barsuglia, A. Bonafede, Frédéric Boone, Albert Bosma, Remy Boyer, M. Branchesi, Carine Briand, Baptiste Cecconi, Sébastien Celestin, Didier Charrrier, E. Chassande-Mottin, and Vyacheslav Zakharenko. Nenufar: Instrument description and science case. pages 1–6, 04 2015.

Appendix

A The LOFAR Telescope Network and the French Station

The LOFAR (LOw Frequency ARray) telescope is a distributed radio interferometer operating in the 10–240 MHz range. It consists of stations across several European countries, including the Netherlands (core), Germany, Poland, Sweden, the UK, and France. Each station is equipped with arrays of Low Band Antennas (LBA) and High Band Antennas (HBA), and the network is coordinated through the central processing hub in the Netherlands.

LOFAR has enabled groundbreaking discoveries in various fields, such as:

- The detection of large-scale cosmic-ray air showers via radio signals.
- High-resolution imaging of the solar corona and heliosphere.
- Monitoring transient radio sources and pulsars.
- Deep surveys of the high-redshift universe in the LOFAR Two-metre Sky Survey (LoTSS).

The French LOFAR station, located at the Nançay Radio Observatory, is known as FR606. It contributes both to international observations and standalone experiments. It played a pivotal role in:

- Studying long-period pulsars and millisecond pulsars at low frequencies.
- Contributing to international baselines for Very Long Baseline Interferometry (VLBI) with LOFAR.
- Developing technical expertise and backends for advanced processing, such as coherent beamforming and FRB search pipelines.

The integration of LOFAR-FR606 with the NenuFAR system allows for complementary observations, especially in beamformed modes, enabling richer time-domain studies of pulsars and transients at ultra-low frequencies.

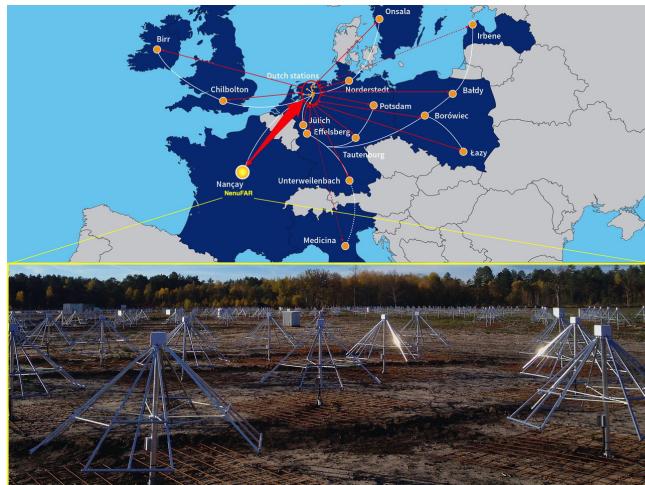


Figure 25: International LOFAR locations.

B Pulsar Keyword Definitions

1. **PSRJ**: Standard pulsar name based on J2000 coordinates (e.g., J0030+0451).
2. **RAJ**: Right Ascension in J2000 epoch, gives sky position (celestial longitude).
3. **DecJ**: Declination in J2000 epoch, gives sky position (celestial latitude).
4. **P**: Pulsar rotation period in seconds.
5. **DM**: Dispersion Measure, total electron content along the line of sight (pc/cm^3).
6. \dot{P} : Derivative of the period, indicates how fast the pulsar is slowing down.
7. **F**: Rotational frequency, the inverse of the period (Hz).
8. **GL**: Galactic longitude, pulsar's location in the Milky Way's coordinate system.
9. **GB**: Galactic latitude, same as above but vertical positioning.
10. **XX YY ZZ**: Cartesian coordinates (in parsecs or kpc) in the Galactic X-Y-Z frame.
11. **W50**: Pulse width at 50% of the peak intensity (degrees or milliseconds), refers to the duration of the radio signal received from a pulsar during each rotation.
12. **W10**: Pulse width at 10% of the peak intensity.
13. **S400**: Mean flux density at 400 MHz, in mJy (milliJansky), a measure of brightness.

C RFI base study

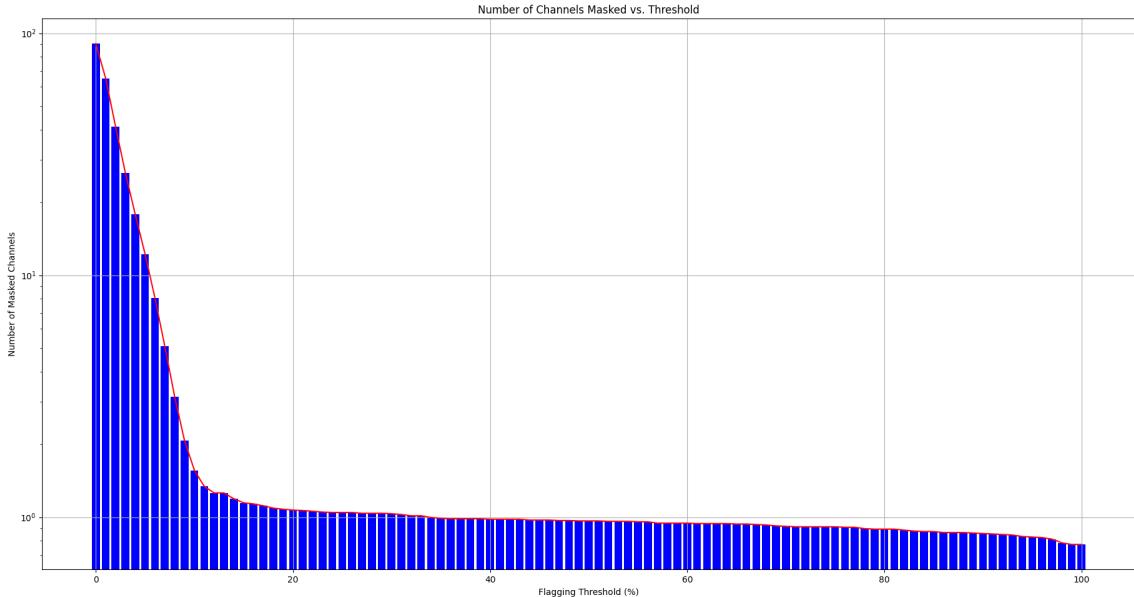


Figure 26: Statistical study of flags, showing by a cumulative distribution the % of data affected when every channel is removed. We decided to go at threshold 40%.

D rfifind output profile

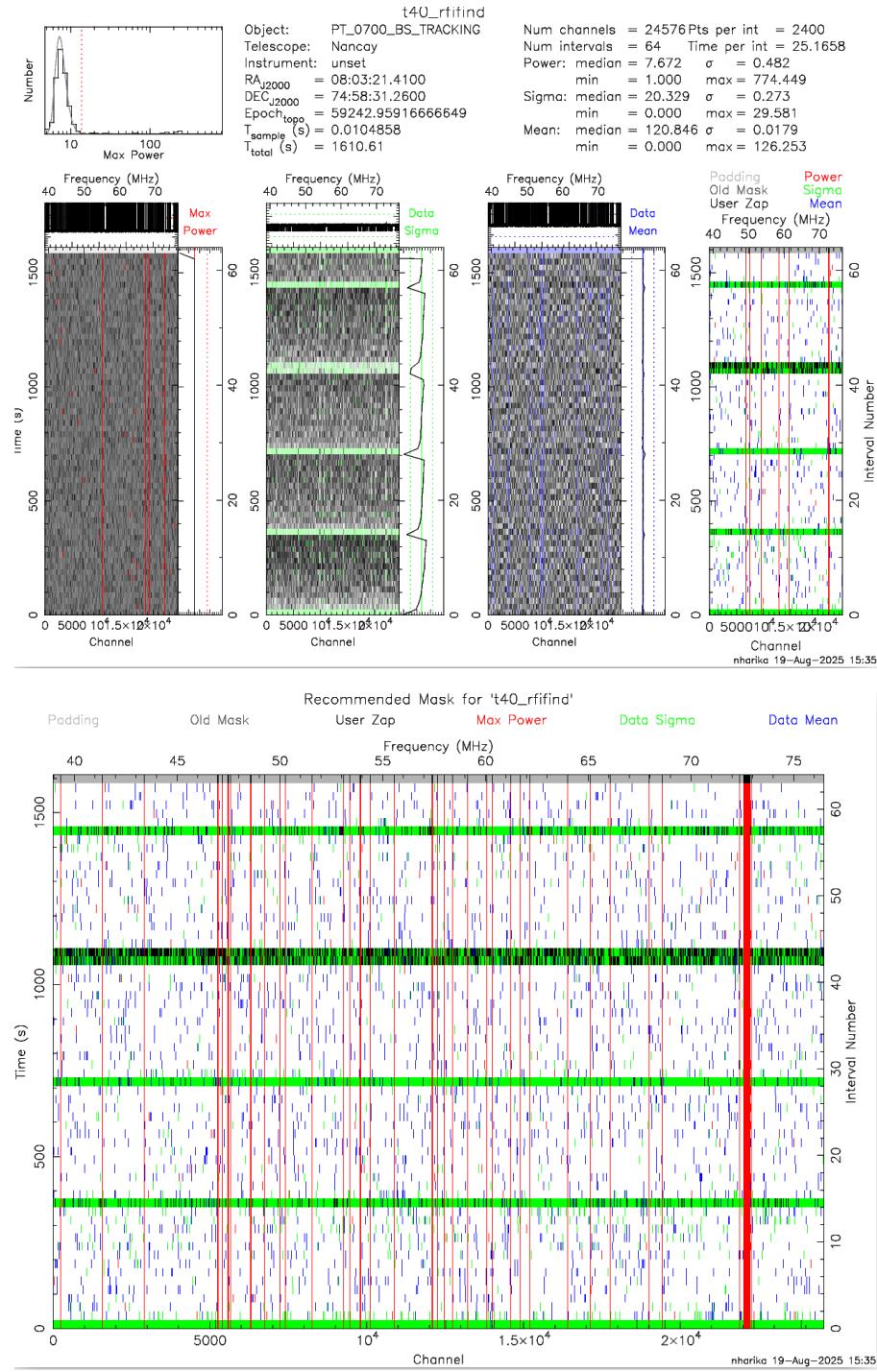


Figure 27: Output file “...rfifind.ps” where colors are bad (red is periodic RFI, blue/green are time-domain statistical issues. This file is the output of the rfifind by itself, without the -zapchan list of masking at 40 % threshold.

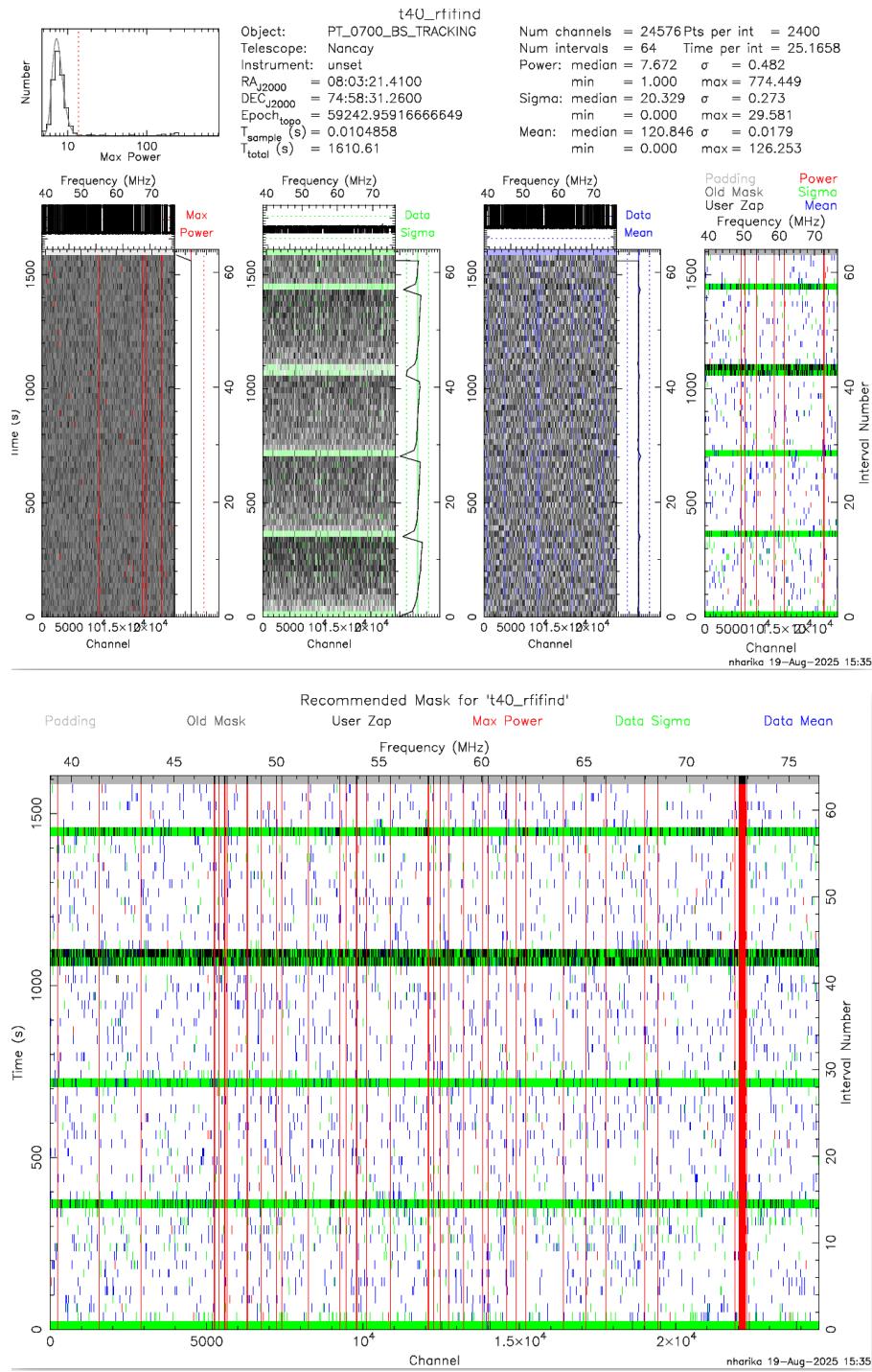


Figure 28: This file is the output of the `rfifind` with `-zapchan` list of the 246 bad channels, the 40 % threshold. We notice a different distribution of the red lines, some of them are masked after the list, that were not before. Same with the blue and green line.

E Codes created for the study

1 Codes

The first code used in RFI studies is the cumulative count:

```
1  # makes 2 plots, and gives a list of the channels above the desired % (useful
2  ↪  for rfifind in PRESTO)
3  ##### OG CODE#####
4
5  import os
6  import numpy as np
7  import matplotlib.pyplot as plt
8  from matplotlib.ticker import AutoMinorLocator
9
10 # Path to the ALL_MASK directory
11 base_path = '/home/nour/Nancay/Rfifind-Masks/ALL_MASK' # change directory path
12 ↪  to your own path
13
14 ## Observation data usually is in teh following format
15 ## Big directory ( ALL_MASK ) > Subdirectories ( example:
16 ↪  MASK_BS_2083_174137+693355_20220504_023035_0 ) > file.txt
17 ↪  (mask_zap_chans_0.txt) ##
18
19
20 # Dictionary to count occurrences of each flagged channel
21 channel_counts = {}
22
23 # Counter for number of successfully read mask files
24 valid_file_count = 0
25
26 # Loop through subdirectories and collect data
27 for root, dirs, files in os.walk(base_path):
28     for file in files:
29         if file.startswith('mask_zap_chans') and file.endswith('.txt'):
30             file_path = os.path.join(root, file)
31             try:
32                 data = np.loadtxt(file_path, dtype=int)
33                 # Handle scalar values
34                 if data.size == 0:
35                     continue
36                 elif data.ndim == 0:
37                     data = [int(data)]
38
39                 valid_file_count += 1
40
41                 for chan in data:
42                     channel_counts[chan] = channel_counts.get(chan, 0) + 1
43
44             except Exception as e:
45                 print(f"Skipping file {file_path} due to error: {e}")
46
47 # Prepare cumulative data
48 all_channels = sorted(channel_counts.keys())
49 counts = np.array([channel_counts[chan] for chan in all_channels])
50 percentages = (counts / valid_file_count) * 100
51
52 # Threshold reporting
53 thresh_60 = [chan for chan, perc in zip(all_channels, percentages) if perc >=
54 ↪  60]
```

```

51 thresh_50 = [chan for chan, perc in zip(all_channels, percentages) if perc >=
52   ↪ 50]
53 thresh_40 = [chan for chan, perc in zip(all_channels, percentages) if perc >=
54   ↪ 40]
55
56 # Overlay thresholded bars in grey (for 40%)
57 thresh_60_counts = [channel_counts[chan] for chan in thresh_40]
58 thresh_60_percentages = [percentages[all_channels.index(chan)] for chan in
59   ↪ thresh_40]
60
61
62 # Plot 1: Cumulative count histogram
63 fig1, ax1 = plt.subplots(figsize=(14,8))
64 ax1.bar(all_channels, counts, color='darkblue', edgecolor= 'blue',
65   ↪ label='Original CD')
66 ax1.bar(thresh_40, thresh_60_counts, color='grey', edgecolor='grey',
67   ↪ label='Threshold = 40 %')
68 ax1.set_ylim(0,100)
69 ax1.set_xlabel('Channel ID')
70 ax1.set_ylabel('Count')
71 ax1.set_title('Cumulative Count of Flagged Channels')
72 ax1.grid(which='major', linestyle='-', linewidth=0.5, color='black')
73 ax1.xaxis.set_minor_locator(AutoMinorLocator(5))
74 ax1.grid(which='minor', linestyle=':', linewidth=0.5, color='gray')
75 ax1.legend()
76
77 # Plot 2: Cumulative percentage histogram
78 fig2, ax2 = plt.subplots(figsize=(14,8))
79 ax2.bar(all_channels, percentages, color='red', edgecolor= 'red',
80   ↪ label='Original CD')
81 ax2.bar(thresh_40, thresh_60_percentages, color='grey',
82   ↪ edgecolor='grey',label='Threshold = 40 %')
83 ax2.set_ylim(0,100)
84 ax2.set_xlabel('Channel ID')
85 ax2.set_ylabel('Percentage (%)')
86 ax2.set_title('Percentage of Files Each Channel Was Flagged In')
87 ax2.grid(which='major', linestyle='-', linewidth=0.5, color='black')
88 ax2.xaxis.set_minor_locator(AutoMinorLocator(5))
89 ax2.grid(which='minor', linestyle=':', linewidth=0.5, color='gray')
90 ax2.legend()
91
92 print(f'\nTotal valid mask files read: {valid_file_count}')
93 print(f'Channels flagged in 60% of files ({len(thresh_60)}'
94   ↪ channels):\n{thresh_60}')
95 print(f'Channels flagged in 50% of files ({len(thresh_50)}'
96   ↪ channels):\n{thresh_50}')
97 print(f'Channels flagged in 40% of files ({len(thresh_40)}'
98   ↪ channels):\n{thresh_40}')
```

The second code is to mask at the threshold at 40%:

```
1 import os
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from matplotlib.ticker import AutoMinorLocator
5
6 # Path to the ALL_MASK directory
7 base_path = '/home/nour/Nancay/Rfifind-Masks/ALL_MASK'
8
9 def load_channel_counts(base_path):
10     """
11         Load and count channel flags from all mask files under base_path.
12         Returns: channel_counts dict and total number of valid files.
13     """
14     channel_counts = {}
15     valid_file_count = 0
16
17     for root, dirs, files in os.walk(base_path):
18         for file in files:
19             if file.startswith('mask_zap_chans') and file.endswith('.txt'):
20                 file_path = os.path.join(root, file)
21                 try:
22                     data = np.loadtxt(file_path, dtype=int)
23                     if data.size == 0:
24                         continue
25                     elif data.ndim == 0:
26                         data = [int(data)]
27
28                     valid_file_count += 1
29                     for chan in data:
30                         channel_counts[chan] = channel_counts.get(chan, 0) + 1
31
32                 except Exception as e:
33                     print(f"Skipping file {file_path} due to error: {e}")
34
35     return channel_counts, valid_file_count
36
37 def plot_filtered_histograms(channel_counts, valid_file_count,
38                             threshold_percent=60):
39     """
40         Filters channels based on the threshold percentage, then plots
41         cumulative count and percentage histograms excluding them.
42     """
43     all_channels = sorted(channel_counts.keys())
44     counts = np.array([channel_counts[chan] for chan in all_channels])
45     percentages = (counts / valid_file_count) * 100
46
47     # Find channels to exclude
48     excluded_channels = [chan for chan, perc in zip(all_channels, percentages)
49                           if perc >= threshold_percent]
50
51     # Filter out excluded channels
52     filtered_channels = [chan for chan in all_channels if chan not in
53                           excluded_channels]
54     filtered_counts = np.array([channel_counts[chan] for chan in
55                               filtered_channels])
56     filtered_percentages = (filtered_counts / valid_file_count) * 100
```

```

54
55     # Plot 1: Cumulative count histogram (filtered)
56     fig1, ax1 = plt.subplots(figsize=(14,8))
57     ax1.bar(filtered_channels, filtered_counts, color='darkblue',
58             edgecolor='blue')
59     ax1.set_ylim(0,100)
60     ax1.set_xlabel('Channel ID')
61     ax1.set_ylabel('Count')
62     ax1.set_title(f'Cumulative Count (Channels < {threshold_percent}%)')
63     ax1.grid(which='major', linestyle='-', linewidth=0.5, color='black')
64     ax1.set_yscale('log')
65     ax1.xaxis.set_minor_locator(AutoMinorLocator(5))
66     ax1.grid(which='minor', linestyle=':', linewidth=0.5, color='gray')

67     # Plot 2: Cumulative percentage histogram (filtered)
68     fig2, ax2 = plt.subplots(figsize=(14,8))
69     ax2.bar(filtered_channels, filtered_percentages, color='red',
70             edgecolor='red')
71     ax2.set_ylim(0,100)
72     ax2.set_xlabel('Channel ID')
73     ax2.set_ylabel('Percentage (%)')
74     ax2.set_title(f'Percentage of Files Each Channel Was Flagged In (<
75                   {threshold_percent}%)')
76     ax2.set_yscale('log')
77     ax2.xaxis.set_minor_locator(AutoMinorLocator(5))
78     ax2.grid(which='minor', linestyle=':', linewidth=0.5, color='gray')

79     plt.show()

80
81     print(f'\nTotal valid mask files read: {valid_file_count}')
82     print(f'Channels excluded ({threshold_percent}% of
83           files):\n{excluded_channels}')
84     return excluded_channels

85     # --- Usage example ---
86     channel_counts, valid_file_count = load_channel_counts(base_path)

87
88     # Set threshold percent you want to filter (e.g., 60%)
89     excluded = plot_filtered_histograms(channel_counts, valid_file_count,
90                                         threshold_percent=40)

```

The third code represents the pulsar search pipeline and the output is a filtered out .csv file with all the potential candidates:

```

1      # PSR_ATNF_ADVANCED_match.py
2      # Advanced way to search, saves at the end on choice the csv,
3      # tracks the errors % of matched psr in the whole folder.
4      #
5      # Run with:
6      #   python3 PSR_ATNF_ADVANCED_match.py -year 2022
7
8      import os
9      import re
10     import argparse
11     import pandas as pd
12     import numpy as np
13     from astropy.coordinates import SkyCoord
14     import astropy.units as u
15     from psrpy import QueryATNF
16
17     # ----- FUNCTION DEFINITIONS -----
18
19     def parse_filename(filename):
20         """Extract period, DM, RA, DEC from filename."""
21         period, dm, ra_ex, dec_ex = None, None, None, None
22         match_pd = re.search(r'_P(\d+\.\d+)_candDM(\d+\.\d+)', filename)
23         match_coords = re.search(r'BS_\d+_((\d{6})\+?(\d{6}))_',
24                                  filename)
25
26         if match_pd:
27             period = float(match_pd.group(1))
28             dm = float(match_pd.group(2))
29
30         if match_coords:
31             ra_str = match_coords.group(1)
32             dec_str = match_coords.group(2)
33             ra_ex = f'{ra_str[:2]}h{ra_str[2:4]}m{ra_str[4:]}s'
34             dec_ex = f'{dec_str[:2]}d{dec_str[2:4]}m{dec_str[4:]}s'
35
36         return period, dm, ra_ex, dec_ex
37
38     def convert_atnf_ra_dec(raj, decj):
39         """Convert ATNF RAJ (hours) and DECJ (degrees) into sexagesimal strings."""
40         try:
41             ra_deg = float(raj) * 15.0 # Convert hours to degrees
42             dec_deg = float(decj)
43             coord = SkyCoord(ra=ra_deg * u.deg, dec=dec_deg * u.deg, frame='icrs')
44             ra_str, dec_str = coord.to_string('hmsdms').split()
45             return ra_str, dec_str
46         except Exception:
47             return None, None
48
49
50     def find_atnf_matches(period, dm, tolerance_ms=1.0, tolerance_dm=1.0):
51         """Find ATNF matches within tolerance on period (ms) and DM."""
52         return catalogue[
53             (np.abs(catalogue['PO'] * 1000 - period) < tolerance_ms) &
54             (np.abs(catalogue['DM']) - dm) < tolerance_dm)
55
56

```

```

57
58     def calculate_beam_offset(ra1, dec1, ra2, dec2):
59         """Angular distance in degrees between two sky positions."""
60         try:
61             coord1 = SkyCoord(ra1, dec1, frame='icrs')
62             coord2 = SkyCoord(ra2, dec2, frame='icrs')
63             return coord1.separation(coord2).degree
64         except Exception:
65             return np.nan
66
67 # ----- MAIN SCRIPT -----
68
69 def main():
70     # ---- ARGPARSE ----
71     parser = argparse.ArgumentParser(description="Match pulsar candidates to
72     ↪ known ATNF entries.")
73     parser.add_argument("-year", required=True, help="Year directory to analyze
74     ↪ (e.g. 2022)")
75     args = parser.parse_args()
76     year = args.year
77
78     # ---- DIRECTORY ----
79     main_dir = f'/home/nour/Nancay/Pulsar_findings/final_plots_tar/{year}'
80
81     bs_filenames = []
82     for subdir in os.listdir(main_dir):
83         full_path = os.path.join(main_dir, subdir)
84         if os.path.isdir(full_path) and subdir.startswith("Plots_CAND_"):
85             for fname in os.listdir(full_path):
86                 if fname.startswith('BS_') and fname.endswith('.pdf'):
87                     bs_filenames.append(os.path.join(full_path, fname))
88
89     # ---- QUERY ATNF ----
90     query = QueryATNF(params=['JNAME', 'PO', 'DM', 'RAJ', 'DECJ', 'GL', 'GB'])
91     global catalogue
92     catalogue = query.table.to_pandas()
93
94     matched_data = []
95     unmatched_data = []
96
97     # ---- LOOP FILES ----
98     for full_path in bs_filenames:
99         filename = os.path.basename(full_path)
100        period, dm, ra_ex, dec_ex = parse_filename(filename)
101
102        if period is None or dm is None or ra_ex is None or dec_ex is None:
103            continue
104
105        matches = find_atnf_matches(period, dm)
106
107        if not matches.empty:
108            for _, row in matches.iterrows():
109                atnf_ra_str, atnf_dec_str = convert_atnf_ra_dec(row['RAJ'],
110                ↪ row['DECJ'])
111                beam_offset = calculate_beam_offset(ra_ex, dec_ex, atnf_ra_str,
112                ↪ atnf_dec_str)
113                matched_data.append({
114                    'filename': filename,
115                    'JNAME': row['JNAME'],
116

```

```

112             'Period (ms)': period,
113             'DM': dm,
114             'RAJ': row['RAJ'],
115             'DECJ': row['DECJ'],
116             'GL': row['GL'],
117             'GB': row['GB'],
118             'beam_offset (deg)': round(beam_offset, 4)
119         })
120     else:
121         unmatched_data.append({
122             'filename': filename,
123             'Period (ms)': period,
124             'DM': dm,
125             'candidate_RA': ra_ex,
126             'candidate_DEC': dec_ex,
127             'RAJ': None,
128             'DECJ': None
129         })
130
131     # ---- DATAFRAMES ----
132     matched_df = pd.DataFrame(matched_data)
133     unmatched_df = pd.DataFrame(unmatched_data)
134
135     # ---- OUTPUT ----
136     print("\n===== MATCHED PULSARS =====")
137     print(matched_df.to_string(index=False) if not matched_df.empty else "None
138           found.")
139
140     print("\n===== UNMATCHED CANDIDATES =====")
141     print(unmatched_df.to_string(index=False) if not unmatched_df.empty else
142           "None found.")
143
144     # ---- STATS ----
145     total = len(bs_filenames)
146     matched = matched_df['filename'].nunique() if not matched_df.empty else 0
147     percent = (matched / total) * 100 if total > 0 else 0
148
149     print(f"\n Total BS files: {total}")
150     print(f" Matched pulsars: {matched}")
151     print(f" Unmatched pulsars: {total-matched}")
152     print(f" Match rate: {percent:.2f}%")
153
154     # ---- SAVE ----
155     save = input("\nDo you want to save the tables to CSV? (y/n):
156           ").strip().lower()
157     if save == 'y':
158         matched_df.to_csv(f"matched_pulsars_{year}.csv", index=False)
159         unmatched_df.to_csv(f"unmatched_candidates_{year}.csv", index=False)
160         print(f"CSV files saved: matched_pulsars_{year}.csv and
161           unmatched_candidates_{year}.csv")
162
163     if __name__ == "__main__":
164         main()

```