Python Plotting of Compressed GNU Radio Pulsar Data

Peter W East

Abstract

This article follows on from a recent Journal article on detailed analysis of compressed pulsar data recordings collected by GNU Radio [1]. A Python program has been developed to present the data analysis in a similar form to PRESTO [2]. As well as duplicating the PRESTO sub-plots with the signal-to-noise ratio (SNR) measuring statistic replacing Chi-square, some extra graphs are introduced to strengthen pulsar validation of low SNR acquisitions. The package features speeded-up data analysis, RFI mitigation and results presentation, together better facilitating pulsar recognition confidence.

In this article results for both strong and weak pulsar acquisitions are included.

Introduction

The technique involves compressing very long data records, as detected and averaged by the GNU Radio platform without losing any pulsar information so that many post-processing algorithms can be applied to rapidly explore detailed pulsar recognition aspects. This accelerated process produces a number of data files describing various pulsar parameters. The Python program presents the analyzed data graphically similar to PRESTO, so that parameter searches can be optimized, RFI blanked, SNR optimized and data confidently validated.

Python Data Plots v Pulsar Properties

Figure 1 shows an outline of the Python-generated data plots. Working from top to bottom, left to right, pulsar properties confirmed in the sub-plots are,

1. *Pulse Profiles* - Correct period and Pulse width.
2. *DM Search* - Correct Dispersion Measure, Extra-terrestrial source.
3. *Period/P-dot Search* - Peak at zero offsets, Accurate Period, Highly stable period.
4. *Band SNR* - Regular pulse train, Wide-band source, SNR frequency growth, Scintillation.
5. *Cumulative SNR* - Regular pulse train, Pulse SNR time growth, Scintillation, Antenna beamshape.
6. *Target Section SNR* - Scintillation, Continuous pulsed source.
7. *Compressed Data* - Noise-like data, Minimal RFI or spikes. Possible pulse visibility.
8. *Compressed Data Spectrum* - Minimal RFI, Possible spectral lines visible.
9. *2D Period P-dot* - Complementary period/P-dot properties, -45°central ridge, No ambiguity.
10. *Data Waterfall* - SNR growth, Continuous pulse train.
11. *Band Waterfall* - Wide-band source, SNR growth, Frequency scintillation.

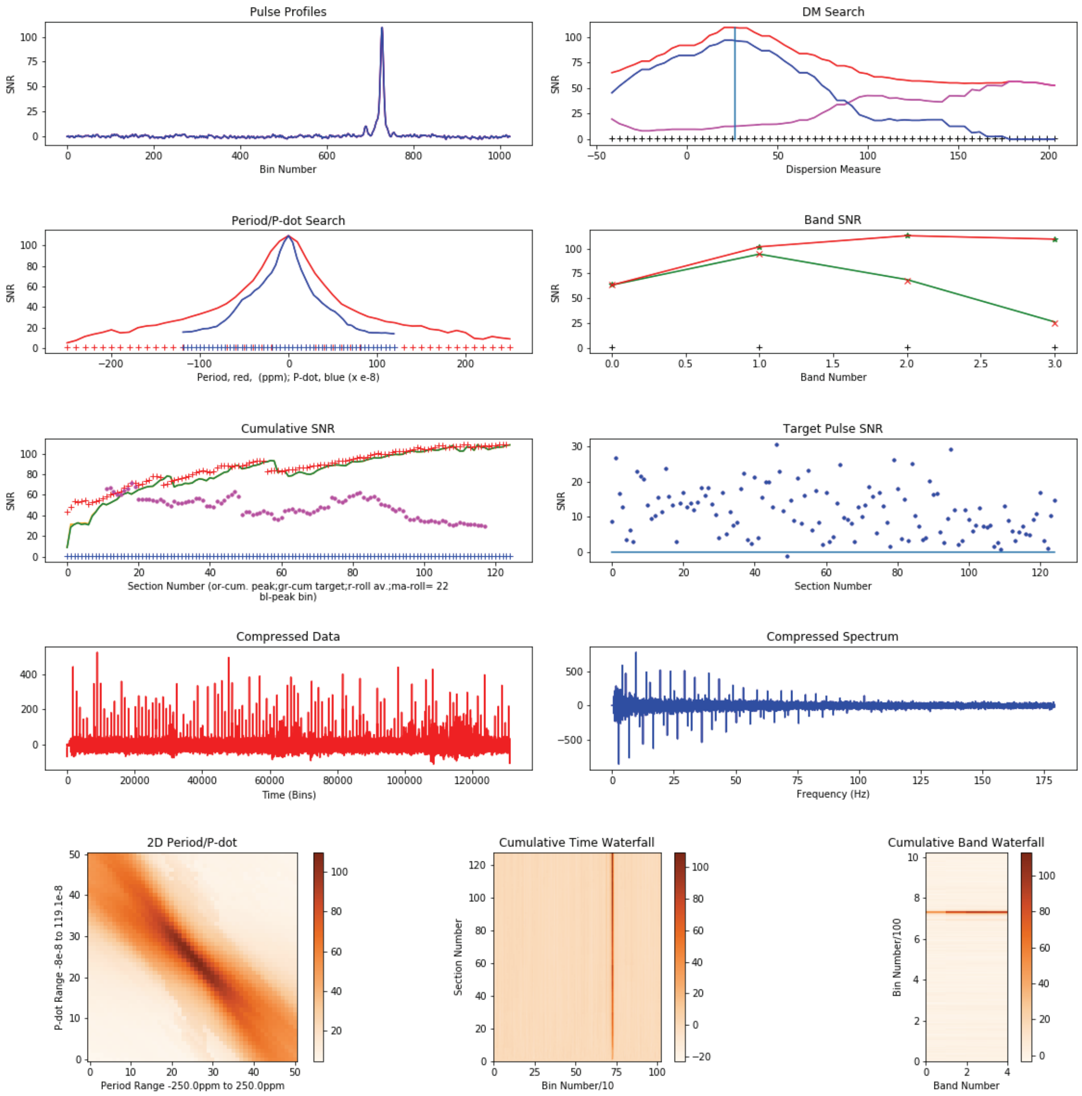
Figure 1 Python Data Presentation

New information introduced includes SNR/band and cumulative SNR in the 'Band SNR sub-plot, and the red and magenta rolling average curves in the Cumulative SNR sub-plot. In the latter, the magenta data with a selected fixed number rolling average indicates the stronger or less scintillating parts of the data record. The red dotted data monitors the peak SNR as a function of the rolling block average size, and identifies the optimum data folding region.

Example 1. Pulsar B0329+54 @ SNR = 109:1

syn_compress16 B0329_3_25_21_05min.bin 4 0.5 714.605858 128 1024 1.5 26.7 0.0 10 2 395 22

Input Data file: = B0329_3_25_21_05min.bin	No. of Output Fold Sections = 128 ; No. bins = 1024
Pulsar Data	No. of Output samples = 131072
Mode = 1 Period Range Multiplier = 10.00	Working file duration = 300 secs
Pulsar Period = 714.605858 ms	Number of pulsar periods = 420
Data clock= 0.50 ms	P-dot factor = 100
Pulsar Pulse Width = 6.50 ms	ppm adjustment = 0.00
Pulsar Dispersion Measure = 26.70	Max Pulse ppm drift = 0.00 ms
Input file bytes = 4800000	Compression Ratio = 3.28
RF Center = 395.0 MHz, RF Bandwidth = 2.0 MHz	Period Search Range = -250.00 ppm to 250.00 ppm
DM Band Delay = 7.192ms, No. Delay bins = 10.305	P-dot Search Range = -119.10 ppm/100 to 119.10 ppm/100
No. Data Samples = 2400000	Rolling Average Number = 22
No: FFT Bands = 4	Blanked Bands:
Input Data per FFT Band = 600000	Blanked Sections: 0



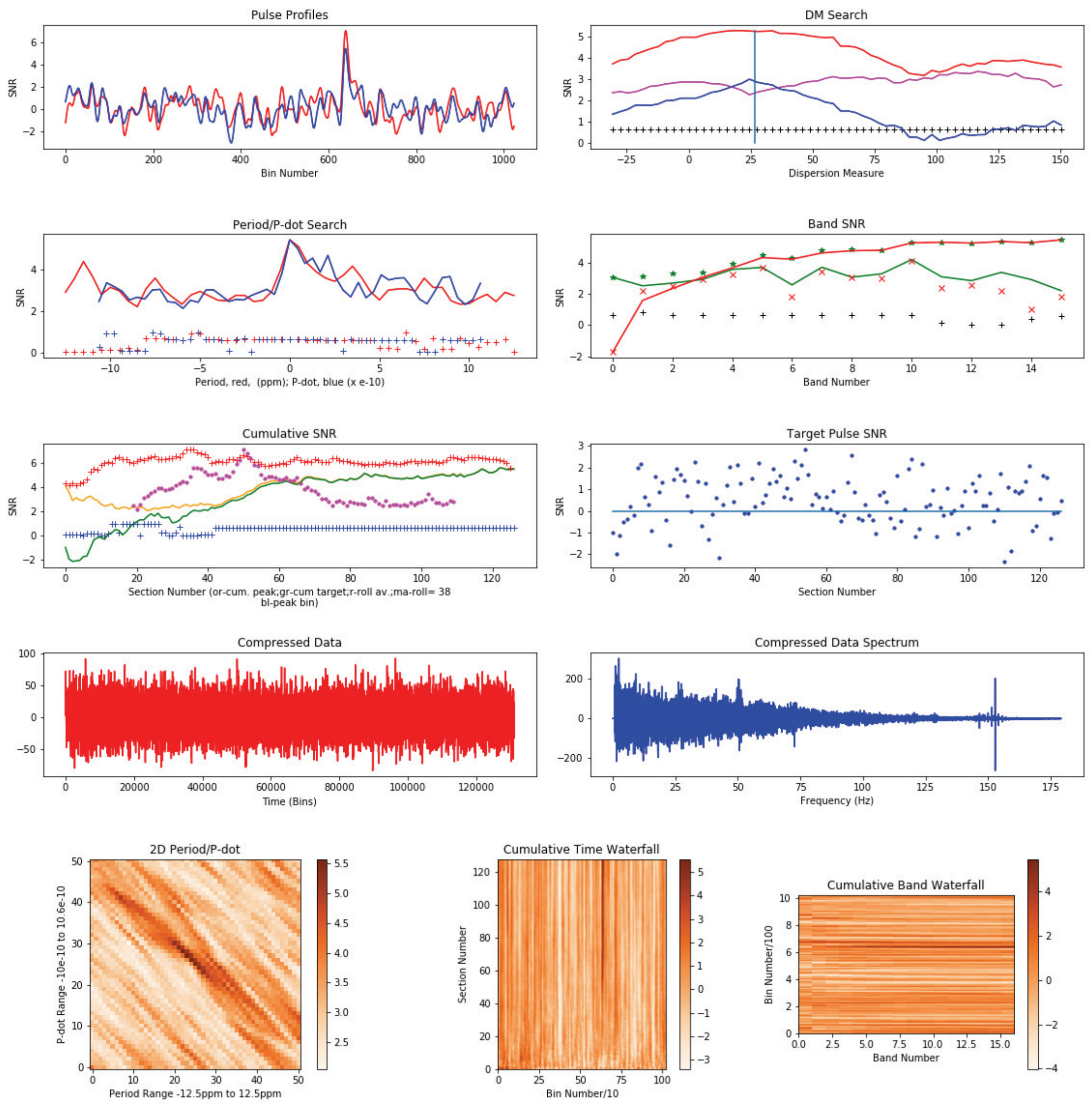
Notes:

1. Data provided by R. Ferranti collected with an RTL SDR and the Stanford Research Institute 46m dish
2. The first section, 0, was effectively blanked by reducing the data amplitude by a factor of 2.
3. Detailed descriptions of the sub-plots are given in Reference [1].
4. *syn_compress* command line detail is listed in the Appendix.

Example 2. Pulsar B0329+54 @ SNR = 5:1 - no RFI problems

syn_compress32 airntmon1.bin 16 2 714.492092 128 1024 6.5 26.7 -.15 0 .5 10 611 32

Pulsar B0329 - 25/10/2021	No: FFT Bands = 16
Mode = 0 Period Range Multiplier = 0.50	Input Data per FFT Band = 8400000
Pulsar Period = 714.491964 ms	No. of Output Fold Sections 128 ; No. Bins = 1024
Data clock= 2.00 ms	No. of Output samples = 131072
Pulsar Pulse Width = 6.50 ms	Working file duration = 16800 secs
Pulsar Dispersion Measure = 26.70	Number of pulsar periods = 23513
Input file bytes = 537600000	P-dot exponent = 10000
No. Data Bytes = 134400000	ppm adjustment = -0.15
RF Center = 611.0 MHz, RF Bandwidth = 10.0 MHz	Max Pulse ppm drift = -2.52 ms
DM Band Delay = 9.716ms, No. Delay bins = 13.924	Compression Ratio = 183.70
Blanked Bands:	Period Search Range = -12.50 ppm to 12.50 ppm
Blanked Sections: 0	P-dot Search Range = -10.63 ppm/10000 to 10.63 ppm/10000
No. Data Samples = 134400000	Rolling Average Number = 38



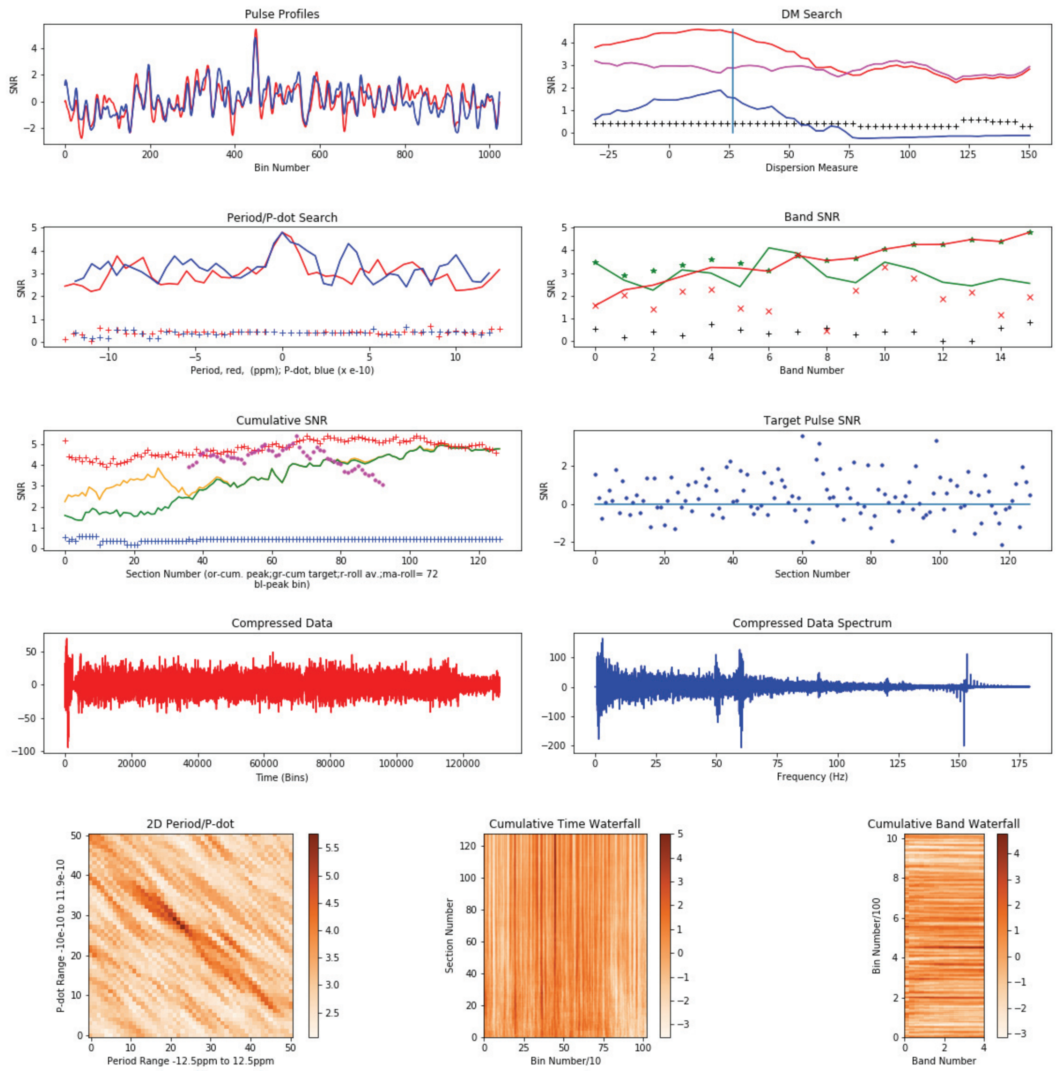
Notes:

1. Inspection of these data plots clearly underpins all the properties expected of a true pulsar signal acquisition.
2. No bands and just Section 0 was blanked by reducing the data amplitude by a factor of 2.
3. Search profiles/bin peak changes are predictable and these results are consistent with a pulse train.

Example 3. Pulsar B0329+54 @ SNR = 4.5:1 - Modest RFI attenuated

syn_compress32 airntsun1.bin 16 1 714.48825 128 1024 5.5 26.7 0 0 .5 10 611 72

Pulsar B0329 - 24/10/2021 Mode = 0 Period Range Multiplier = 0.50 Pulsar Period = 714.488250 ms Data clock= 1.00 ms Pulsar Pulse Width = 5.50 ms Pulsar Dispersion Measure = 26.70 Input file bytes = 960000000 No. Data Bytes = 240000000 RF Center = 611.0 MHz, RF Bandwidth = 10.0 MHz DM Band Delay = 9.716ms, No. Delay bins = 13.924 Blanked Bands: Blanked Sections 2 70 125 3 116 117 118 119 120 121 122 123 124 125 126 127 2 No. Data Samples = 240000000	No: FFT Bands = 16 Input Data per FFT Band = 15000000 No. of Output Fold Sections 128 ; No. Bins = 1024 No. of Output samples = 131072 Working file duration = 15000 secs Number of pulsar periods = 20994 P-dot exponent = 10000 ppm adjustment = 0.00 Max Pulse ppm drift = 0.00 ms Compression Ratio = 164.02 Period Search Range = -12.50 ppm to 12.50 ppm P-dot Search Range = -11.91 ppm/10000 to 11.91 ppm/10000 Rolling Average Number = 72to 12.50 ppm P-dot Search Range = -11.91 ppm/10000 to 11.91 ppm/10000 Rolling Average Number = 72



Notes:
1. Inspection of these data plots clearly underpins all the properties expected of a true pulsar signal acquisition.
2. Section with modest RFI attenuated by a factor of 2 per entry including a large spike in section 2.
3. Both low SNR examples indicate the antenna drift-scan profile in the rolling average plot.
4. The Cumulative SNR plot indicated that an improved DM Search result may arise by stopping at section 100.

Software

Example 1 data was recorded using the software: '*filterbank_4ch_2k*', offered by Hannes Fasching in Reference [3], based on work done by Andrea and Giorgio Dell'Immagine [4]. Operating under GNU Radio, the recorded data is in 2-byte binary form and is split into 4 sub-bands at a 2ksps clock rate. 5 minutes of recorded data was analyzed in '*syn_compress*' in 18 seconds, producing 20 parameter analysis files. The test mode to optimize period matching ran in under 10 seconds. Important numerical results are listed in the command terminal. The output plots allow identification of RFI affected bands or compressed data sections for setting up band and section blanking regions.

Example 2 and 3 data were recorded on two consecutive days using a variation of '*stupid_simple_pulsar*', designed by Marcus Leech [5]. In these examples the data is output as 32-bit binary in 16 frequency bands each at 500sps clock rate. In case 2, 4 hours 40 minutes of recording with significant data averaging was processed by '*syn_compress*' in under 25 seconds. The initial data fold of Example 2 indicated an SNR of 4.8:1 which rose to 7.9:1 after correct de-dispersion and optimum choice of the rolling average range and central section position (38,50). Matching the expected topocentric frequency on two consecutive days within the SDR clock and TEMPO accuracy is further evidence of valid pulsar detection.

The data compression and analysis software, '*syn_compress*' comprised two similar versions to cope with the two types of binary data and FFT band reversal (a filterbank file format version is planned). The data plotting program, '*pul_plot.py*' based on Python3 is run under the 'Spyder IDE'.

Recording and analysis software versions for WINDOWS and LINUX are under development but can be made available for testing on request.

Conclusions

Using the SNR statistic for examining data recordings extends the scope of pulsar data analysis and verification to much lower peak SNR levels. The data compression approach usefully reduces the operator data investigation and optimization time to provide complementary results to PRESTO as presented by the Python graphics.

The new cumulative SNR plots for both multiple bands and sections confirms pulsar scintillating presence along the record, differentiates pulse trains from noise/RFI [1]. It also adds confidence in receiver drift-scan modes by confirming extra-terrestrial signal reception by noting that the median received signal amplitude generally follows the antenna beam pattern.

The most rewarding innovation is the rolling average scheme that exploits scintillation by identifying quite clearly, the optimum regions to fold data to produce the best SNR estimate.

Final Comments

Positively recognizing weak pulsar signals in noise and RFI is no easy task. This exercise has sought to identify and separate the various differentiating properties of pulsar pulse trains, noise and RFI and the Python plots in this article summarize most of them. There are more, as described in previous articles, the search profiles of dispersion measure, pulsar pulse train period and period rate are well-defined and matching these to the theory, also adds further confidence.

Spectral search and folding techniques on the other hand fall far short in detection sensitivity, compared to the standard fold algorithm.

Pulsar recognition prerequisites include system calibration against the Lorimer and Kramer pulsar radiometer equation, choosing a low RFI band at lower radio frequencies and, apart from pointing the antenna in the right place at the right time, ideally locking the SDR to a high stability reference oscillator [6]. A high accuracy SDR clock and using TEMPO to determine the expected topocentric period/frequency at the antenna location and observation time, conveniently narrows the search algorithms range to locate a possible pulsar target.

There are various simple checks that can be employed such as folding half records, double-period folds and half-band folds, but by far the most positive check is dispersion measure search indicating an extra-terrestrial source.

Notwithstanding the power of various amateur and professional tools, there is always a finite probability that limited data sets of even pure noise can result in false identifications of weak targets. However, with intelligent use of these tools there is a very high probability of correct identification, even for some very low SNR initial indications.

References

[1] PW East. Detailed Evaluation of Low SNR Pulsar Data Records, Journal of the Society of Amateur Radio Astronomers., March/April 2022.

[2] PRESTO Home, https://www.cv.nrao.edu/~sransom/presto/

[3] H Fasching (OE5JFL) Pulsars - How to Detect. https://qsl.net/oe5jfl/pulsar/detecting_pulsars.pdf

[4] G Dell'Immagine, A Dell'Immagine. Linux pulsar software for recording and analysis. https://github.com/gio54321/pulsar-distro-guide

[5] MD Leech. https://www.ccera.ca/papers/memo-12-a-pulsar-observing-capability-at-ccera/

[6] DR Lorimer, M Kramer, Handbook of Pulsar Astronomy, Cambridge University Press, 2005. p 265.

PW East May 2022

Appendix: Compression Software Command Parameters

The post processing software, *syn_compress.exe* runs in a 64-bit command terminal (*cmd.exe* for WINDOWS) *.*

The command calling terms are of the form;-

syn_compress <binfile> <No. FFT bins><data clock (ms))><period (ms)><No. sections><No. fold bins><pulse width (ms)><DM><ppm adjust><Test T/0><ppm range factor><RF bandwidth (MHz)><RF center (MHz)><Rolling average No.>

Text files Blankf.txt and Blanks.txt in the working directory contain space spaced band/section numbers that are reduced in amplitude by a factor of 2 for each entry and are required. Spikes and RFI corrupted bands/sections can be identified in the terminal print out.

Peter East, *pe@y1pwe.co.uk* is retired from a Defense Electronics career in radar and electronic warfare system design. He has authored a book on Microwave System Design Tools, is a member of the British Astronomical Association since the early '70s and joined SARA in 2013. He has had a lifelong interest in radio astronomy; presently active in amateur detection of pulsars using SDRs, and researching low SNR pulsar recognition and analysis. He has recently written another book, 'Galactic Hydrogen and Pulsars - an Amateurs Radio Astronomy' describing his work in Radio Astronomy.

He maintains an active RA website at *http://www.y1pwe.co.uk*