

Research Internship Report

Data Analytics for Improving Fast Radio Burst Detection Rates

Yashasvi Ghadale

2019PH10666

B.Tech, Engineering Physics
Indian Institute of Technology, Delhi

Prof. Shriharsh Tendulkar

Tata Institute of Fundamental Research, Mumbai
Supervisor

Prof. Suprit Singh

Indian Institute of Technology, Delhi
Internal Supervisor

Introduction

Abstract

Broadly, my work as a part of this project involved optimizing the transient detection pipeline and classification system for the Canadian Hydrogen Intensity Mapping Experiment's Fast Radio Burst (CHIME/FRB) project. The aim of my project was to analyze the sensitivity of the pipeline components to FRBs that are outside of the normal phase space — particularly very long timescales or highly scattered FRBs.

1.1 Fast Radio Bursts

Fast Radio Bursts (FRBs) are brief (few milliseconds long) bursts of radio waves coming from far beyond the Milky Way. The phenomenon was first reported in 2007 and as of mid-2017, roughly two dozen have been reported. Their origin is unknown. However, they are ubiquitous: current best estimates suggest these events are arriving at Earth roughly a thousand times per day over the entire sky at the fluence thresholds of typical radio telescopes.

1.2 Important FRB Characteristics for Detection

1. Dispersion Measure: The DM is defined by the integral of n_e along the line of sight and can be derived from pulsar and FRB detections based on the emission's time delay as a function of frequency
2. Width: FRBs have an associated intrinsic timescale, as emitted by the source. This timescale, known as the intrinsic width. The intrinsic width of FRBs and its distribution offers some useful constraining information for FRB models, and it plays an important role in the detectability of bursts.
3. Scattering: The interstellar scattering of radio waves, manifests itself as the broadening in time at low frequency of a pulse. A screen of inhomogeneous media between a pulsating radio source and the observer could potentially explain amplitude variations between pulses. Such a scattering screen also leads to temporal broadening.
4. Brightness: One of the simplest imaginable observable properties of FRBs is how bright a radio telescope's detection appears against the radio sky background and system noise. First, a dedispersed timeseries must be created by choosing the DM which maximises the pulse SNR and averaging the signal over the receiver bandwidth. Then the flux density of the pulse $S(t)$ must be integrated over to obtain a fluence*

*More characteristics are described in reference number 1

2.1 CHIME CHIME/FRB

CHIME is a novel radio telescope that has no moving parts. Originally conceived to map the most abundant element in the Universe-hydrogen-over a good fraction of the observable universe, this unusual telescope is optimized to have a high "mapping speed", which requires a large instantaneous field of view (~ 200 square degrees) and broad frequency coverage (400-800 MHz). The digitized signals collected by CHIME will be processed to form a 3-dimensional map of hydrogen density, which will be used to measure the expansion history of the Universe. At the same time, these signals can be combed for transient radio emission, making CHIME a unique telescope for discovering new Fast Radio Bursts and for monitoring many pulsars on a daily basis.

CHIME's large collecting area, wide bandwidth, and enormous field-of-view make it a superb detector of FRBs. As of mid-2020, CHIME has detected well over 1000 Fast Radio Burst sources. With it's high detection rate, the CHIME/FRB project has been able to study the population properties of FRBs, identify a large number of repeating sources, and distinguish between different FRB morphologies. Bright CHIME-discovered FRBs will be found in real-time and reported immediately to the worldwide astrophysical community for multi-wavelength follow-up.

2.2 FRB Detection Process at CHIME

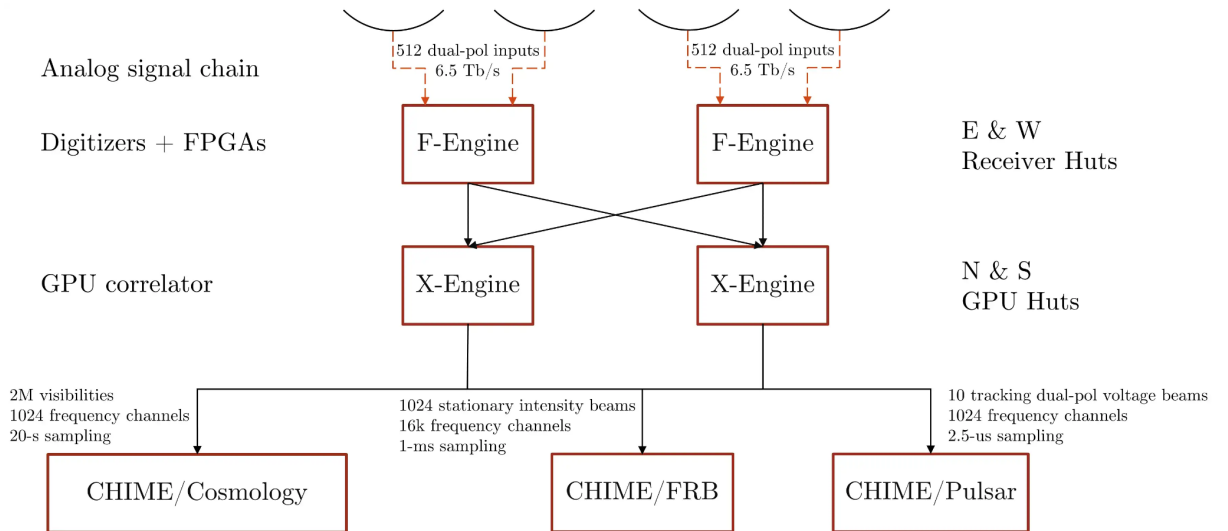


Figure 1: Schematic of the CHIME telescope signal path. The four cylinders (black arcs), the correlator (Fand X-Engines), and the backend science instruments are shown. The dashed orange segments depict analog signals carrying coaxial cables from the 256 feeds on each cylinder to the F-Engines in the corresponding East or West receiver huts beneath the cylinders. The black segments depict digital data carried through copper and fiber cables. Networking devices are not shown. The X-Engine is housed in two shipping containers (labeled North and South) adjacent to the cylinders. The HI intensity map making (CHIME/Cosmology) and CHIME/Pulsar backends are housed in a shielded room in the DRAO building, and the CHIME/FRB backend (hatched red) is in a third shipping container adjacent to the cylinders. Note that the total input data

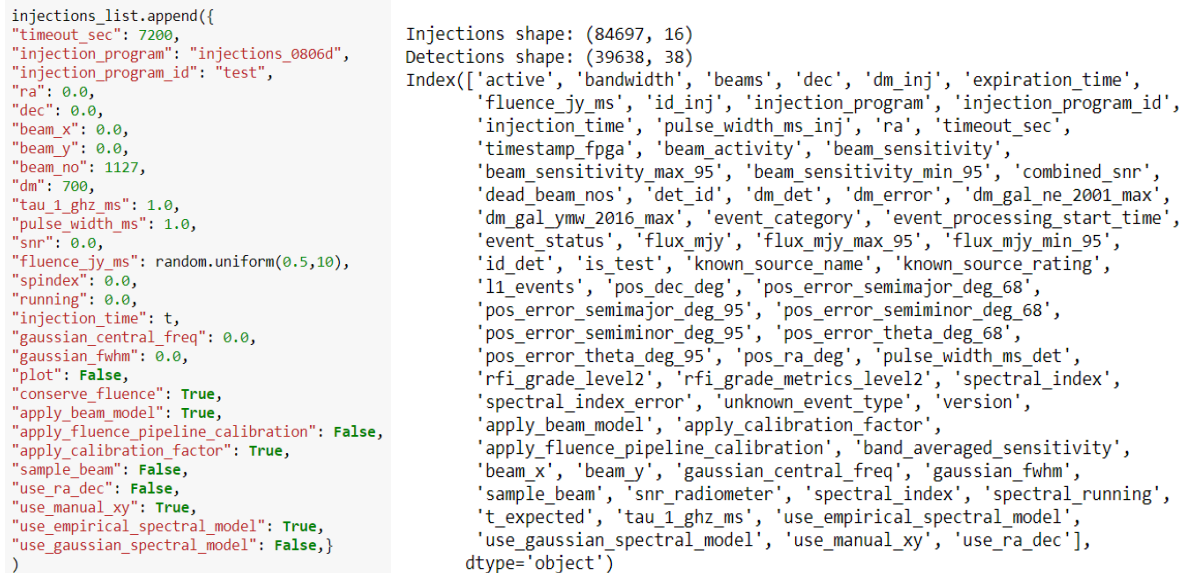
rate into the F-Engine is 13 Tb/s. The data rate into the CHIME/FRB backend is 142 Gb/s. Adapted from CHIME/FRB Collaboration et al (2018)

3.1 Injections

While CHIME/FRB’s design allows for the highest FRB detection rate of all current surveys, like any telescope, it is subject to selection biases affecting our interpretation of the observed population of bursts. These biases come primarily from the fact that CHIME’s beams are complicated, and from the RFI littering on-sky data. In order to undo the biases which are introduced in the real-time pipeline, CHIME/FRB has developed a service that injects model populations of synthetic FRBs into the live detection pipeline. By injecting pulses over a broad set of parameters and tracking the (non-)detection status of injections, it is possible to construct a probability distribution of event detection, $P(X)$ (where X is the “true” set of FRB parameters), in a large multi-variable parameter space. This probability distribution, known as the selection function, is independent of the model FRB population.

3.2 Creating Synthetic FRB Databases using Injections

Using the injection systems at CHIME, it is possible to inject a series of synthetic pulses with specific values of characteristic features such as right ascension, declination, dispersion measure, signal to noise ratio, fluence, etc. These are then passed through a duplicate detection pipeline at CHIME, resulting in detections, and are stored as datasets.



```

injections_list.append({
    "timeout_sec": 7200,
    "injection_program": "injections_0806d",
    "injection_program_id": "test",
    "ra": 0.0,
    "dec": 0.0,
    "beam_x": 0.0,
    "beam_y": 0.0,
    "beam_no": 1127,
    "dm": 700,
    "tau_1_ghz_ms": 1.0,
    "pulse_width_ms": 1.0,
    "snr": 0.0,
    "fluence_jy_ms": random.uniform(0.5,10),
    "spindex": 0.0,
    "running": 0.0,
    "injection_time": t,
    "gaussian_central_freq": 0.0,
    "gaussian_fwhm": 0.0,
    "plot": False,
    "conserve_fluence": True,
    "apply_beam_model": True,
    "apply_fluence_pipeline_calibration": False,
    "apply_calibration_factor": True,
    "sample_beam": False,
    "use_ra_dec": False,
    "use_manual_xy": True,
    "use_empirical_spectral_model": True,
    "use_gaussian_spectral_model": False,})

```

```

Injections shape: (84697, 16)
Detections shape: (39638, 38)
Index(['active', 'bandwidth', 'beams', 'dec', 'dm_inj', 'expiration_time',
       'fluence_jy_ms', 'id_inj', 'injection_program', 'injection_program_id',
       'injection_time', 'pulse_width_ms_inj', 'ra', 'timeout_sec',
       'timestamp_fpga', 'beam_activity', 'beam_sensitivity',
       'beam_sensitivity_max_95', 'beam_sensitivity_min_95', 'combined_snr',
       'dead_beam_nos', 'det_id', 'dm_det', 'dm_error', 'dm_gal_ne_2001_max',
       'dm_gal_ymw_2016_max', 'event_category', 'event_processing_start_time',
       'event_status', 'flux_mjy', 'flux_mjy_max_95', 'flux_mjy_min_95',
       'id_det', 'is_test', 'known_source_name', 'known_source_rating',
       'l1_events', 'pos_dec_deg', 'pos_error_semimajor_deg_68',
       'pos_error_semimajor_deg_95', 'pos_error_semiminor_deg_68',
       'pos_error_semiminor_deg_95', 'pos_error_theta_deg_68',
       'pos_error_theta_deg_95', 'pos_ra_deg', 'pulse_width_ms_det',
       'rfi_grade_level2', 'rfi_grade_metrics_level2', 'spectral_index',
       'spectral_index_error', 'unknown_event_type', 'version',
       'apply_beam_model', 'apply_calibration_factor',
       'apply_fluence_pipeline_calibration', 'band_averaged_sensitivity',
       'beam_x', 'beam_y', 'gaussian_central_freq', 'gaussian_fwhm',
       'sample_beam', 'snr_radiometer', 'spectral_index', 'spectral_running',
       't_expected', 'tau_1_ghz_ms', 'use_empirical_spectral_model',
       'use_gaussian_spectral_model', 'use_manual_xy', 'use_ra_dec'],
      dtype='object')

```

Figure 2: On the left side is a python script used for injecting an artificial pulse with specified pulse parameters. On the right side is information displayed after running a successful injection program and passing it through the detection pipeline. It gives the total number of injections and detections along with a list of 71 parameters that are stored for each detected synthetic FRB.

4.1 Real-Time Injections Monitoring Dashboard Design

CHIME has a dedicated detections monitoring system for real FRBs signals picked up by the telescope. However, such a system has not been developed for monitoring detections coming from injected FRBs. As a sub-project, I also designed a real-time synthetic pulse injections monitoring dashboard including important visualizations required to assess the performance of the detection on an injection program. Some of which include-

- Histograms- Fluence, DM, L1/L2 grades
- Detection fraction as a function of fluence, width, and DMs
- number of injections, number of detections by the de dispersion algorithm
- number of detections passing rfi_grade_l1/ avg_l1_grade cutoff
- Number of detections passing rfi_grade_l2

5.0 Plotting Functions

5.1 Detection Fraction: plot_detection_fraction: Makes a detection fraction histogram of FRBs detected as a function of `x_parameter` at bonsai, L1, and L2. The filtered dataset is passed as a pandas data frame or a Numpy record array and `x_parameter` must be a valid column name data frame. The `x_range` can be specified as `[x_min, x_max]`. If `x_range` is None, then we use the min and max of the `x_parameter` in the dataset. `x_bins` specifies the number of bins to split the x-axis into.

Sample plot

```
: plot_detection_fraction(data, 5, 7, x_parameter = 'fluence_jy_ms', x_bins=10, x_range=[1,10000], x_log =True)
```

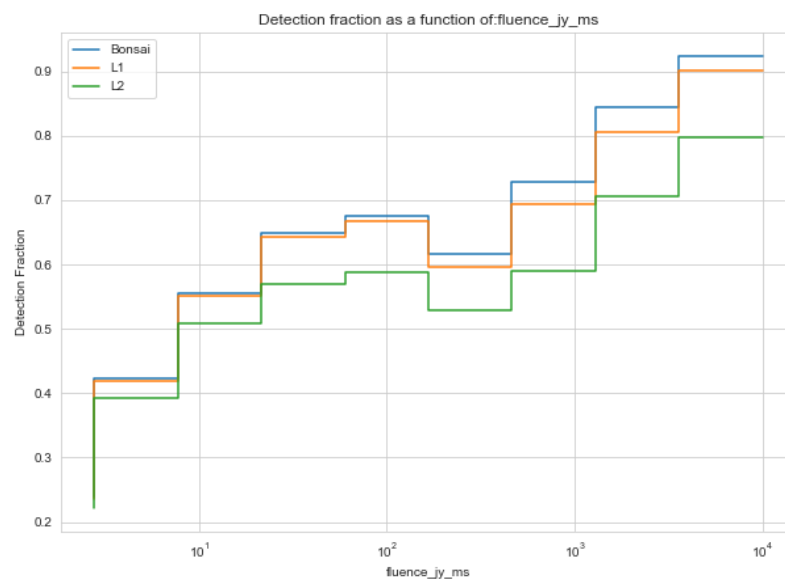


Figure 3: A sample detection fraction vs injected fluence plot generated using the plotting function developed

5.2 SNR-Fluence Ratio: `plot_snr_fluence_ratio`:

Makes an SNR-Fluence ratio vs `x_parameter` scatter plot FRBs detected at bonsai, L1, and L2. The filtered dataset is passed as a pandas data frame or a Numpy record array and `x_parameter` must be a valid column name data frame. The `x_range` can be specified as `[x_min, x_max]`. If `x_range` is None, then we use the min and max of the `x_parameter` in the dataset.

Input parameters:

`filtered_dataset`: dataset to be considered for plotting

`l1_cutoff`: cutoff average RFI L1 score; default: 5

`l2_cutoff`: cutoff average RFI L2 score; default: 7

`x_parameter`: observable plotted on the x-axis; default: "fluence_jy_ms"

`x_range`: specifies the range of the x-parameter to be plotted; default: full range

`x_bins`: specifies number of bins to be used for calculating the detection fraction; default: 10

`x_log`: (True/False), applies a log scale on the x-axis; default: False

Sample plot

```
plot_snr_fluence_ratio(data, 5, 7, 'dm_inj', x_log=False)
```

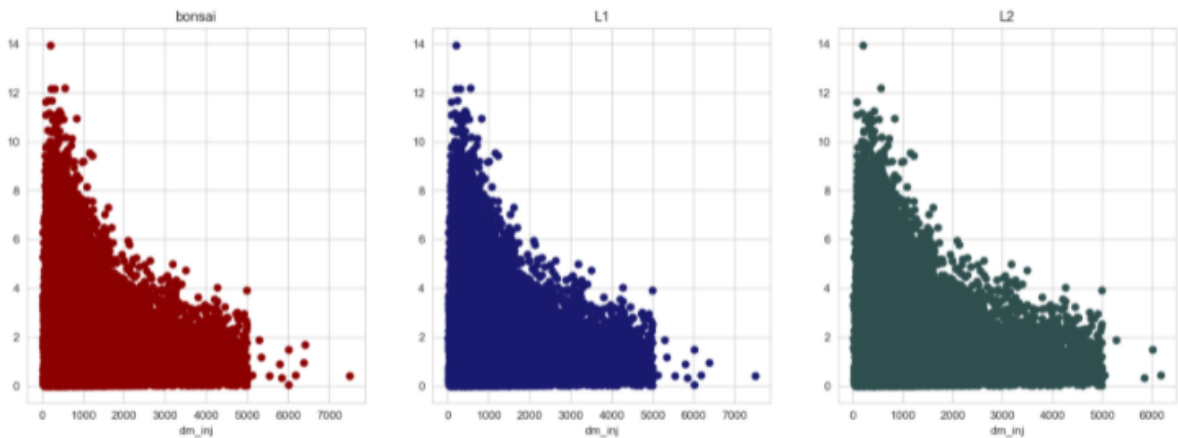


Figure 4: On the x-axis is the SNR Fluence ratio, and the y-axis represents DM (Dispersion Measure) values.

6.1 Detection Fraction Plots and Findings

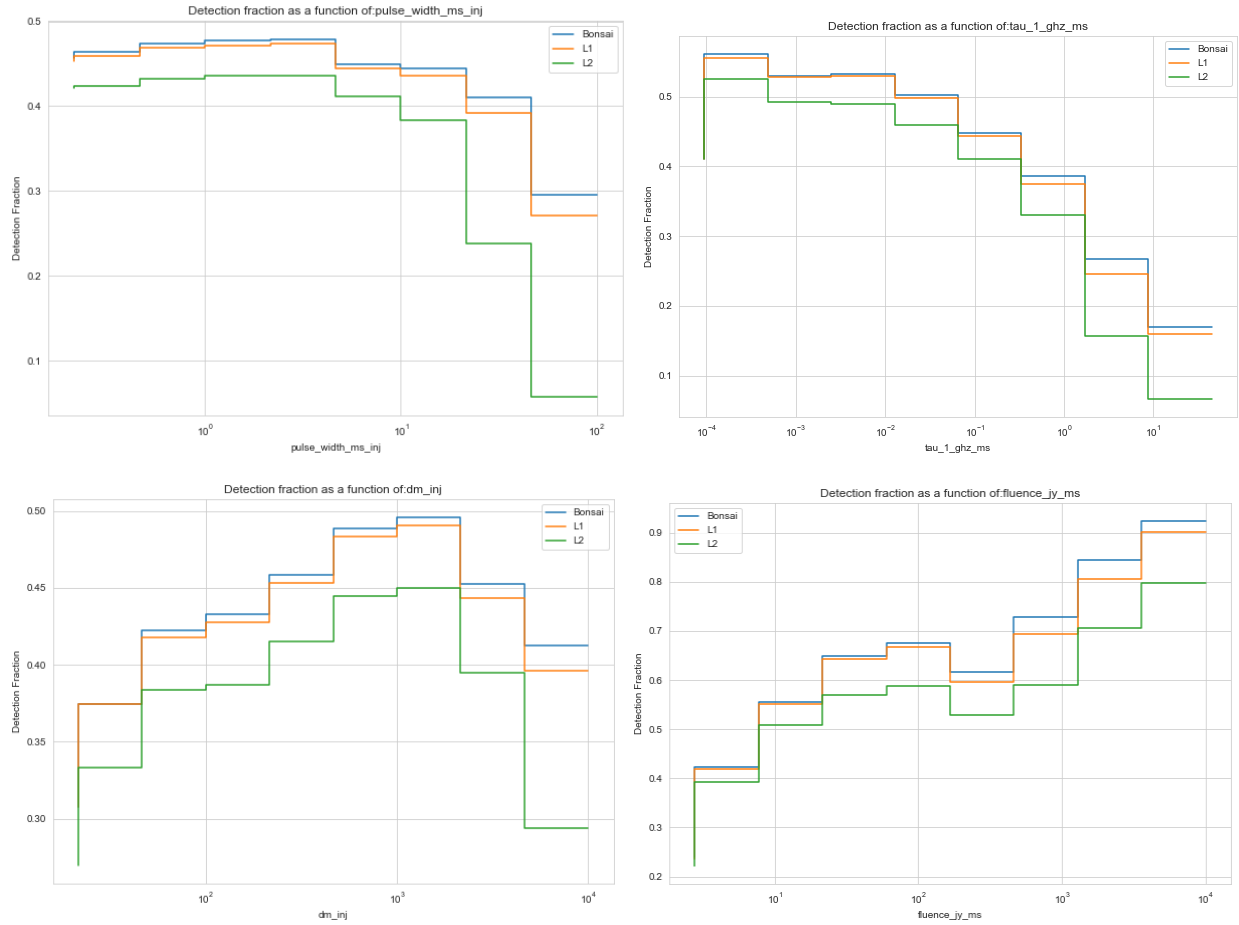


Figure 5: Detection fraction as a function of pulse width, scattering time, DM, and fluence respectively; generated using the plotting functions described earlier.

Through these detection fraction plots, we found that:

- For injections with pulse width values between 20 and 40 ms, the DF drops from about 0.4 at bonsai and L1 to 0.2-0.25 at L2
- Between 40 and 100 ms, DF drops from about 0.3 at bonsai and L1 to 0.05 at L2
- An increase in DFs is seen as DM values increase up till fluence value of 2000 pc/cm³
- For pulses with DM values between 2000 and 4500 pc/cm³, the DF drops from about 0.45 at bonsai and L1 to 0.40 at L2
- Between 4500 and 10000 pc/cm³, DF drops from about 0.42 at bonsai and L1 to 0.30 at L2

6.2 SNR-Fluence Ratio Plots

Following are the scatterplots generated using the `plot_snr_fluence_ratio` function described earlier.

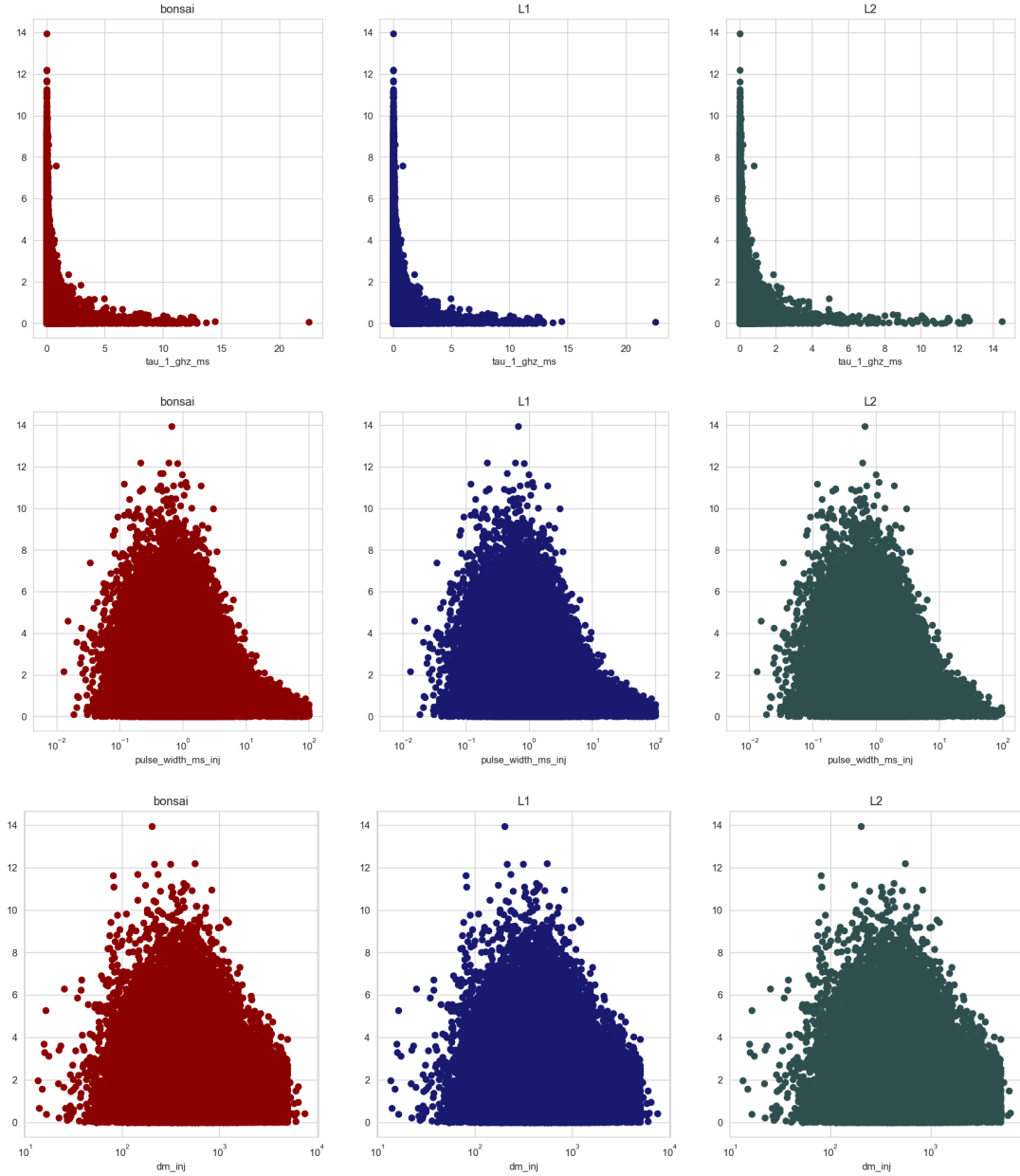


Figure 6: On the x-axis is the SNR-fluence ratio plotted as a function of the x_parameters- scattering time, pulse width, and DMs respectively.

7.0 Findings and Results

Important Findings

A. Detection Fraction vs Pulse Width



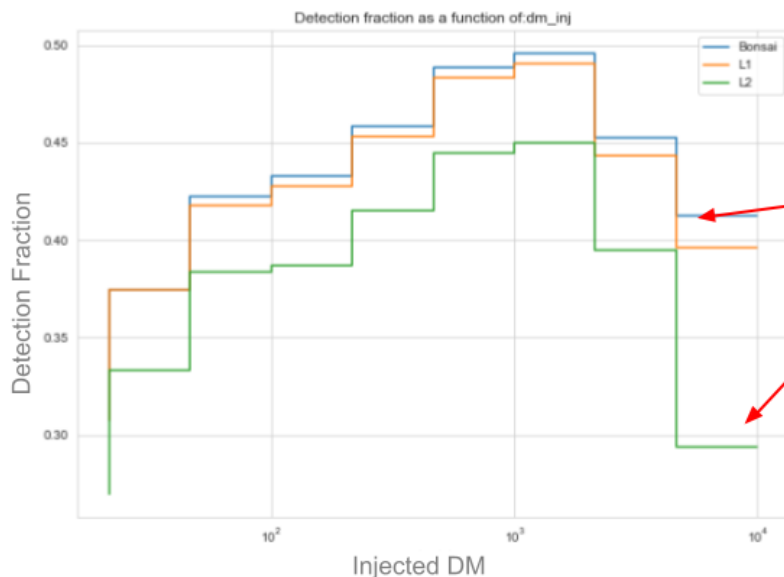
- Detected as a trigger by bonsai
- Identified as astrophysical by L1
- Identified as astrophysical by L2

Injections dataset used: from August
Injections: 84,000
Total detections: 35,000

For injections with pulse width values between 20 and 40 ms, the DF drops from about 0.4 at bonsai and L1 to 0.2-0.25 at L2

Between 40 and 100 ms, DF drops from about 0.3 at bonsai and L1 to 0.05 at L2

B. Detection Fraction vs Injected DM



- An increase in DFs is seen as DM values increase up till Fluence value of 2000 pc/cm³
- For pulses with DM values between 2000 and 4500 pc/cm³, the DF drops from about 0.45 at bonsai and L1 to 0.40 at L2
- Between 4500 and 10000 pc/cm³, DF drops from about 0.42 at bonsai and L1 to 0.30 at L2

We are missing high DM and high Pulse Width FRBs that L2 classifies as RFI!

Appendix

1. MS thesis- Marcus Merryfield:
<https://escholarship.mcgill.ca/concern/theses/1n79h894t>
2. <https://chime-experiment.ca/en>
3. Code used on GitHub:
<https://github.com/yashasvighadale/Data-Analytics-for-FRBs/blob/main/codefile.py>
4. The detailed report in ipython format:
<https://github.com/yashasvighadale/Data-Analytics-for-FRBs/blob/main/Report-Copy1.ipynb>
5. The CHIME fast radio burst project: system overview (Amiri, Bandura, Berger et al)