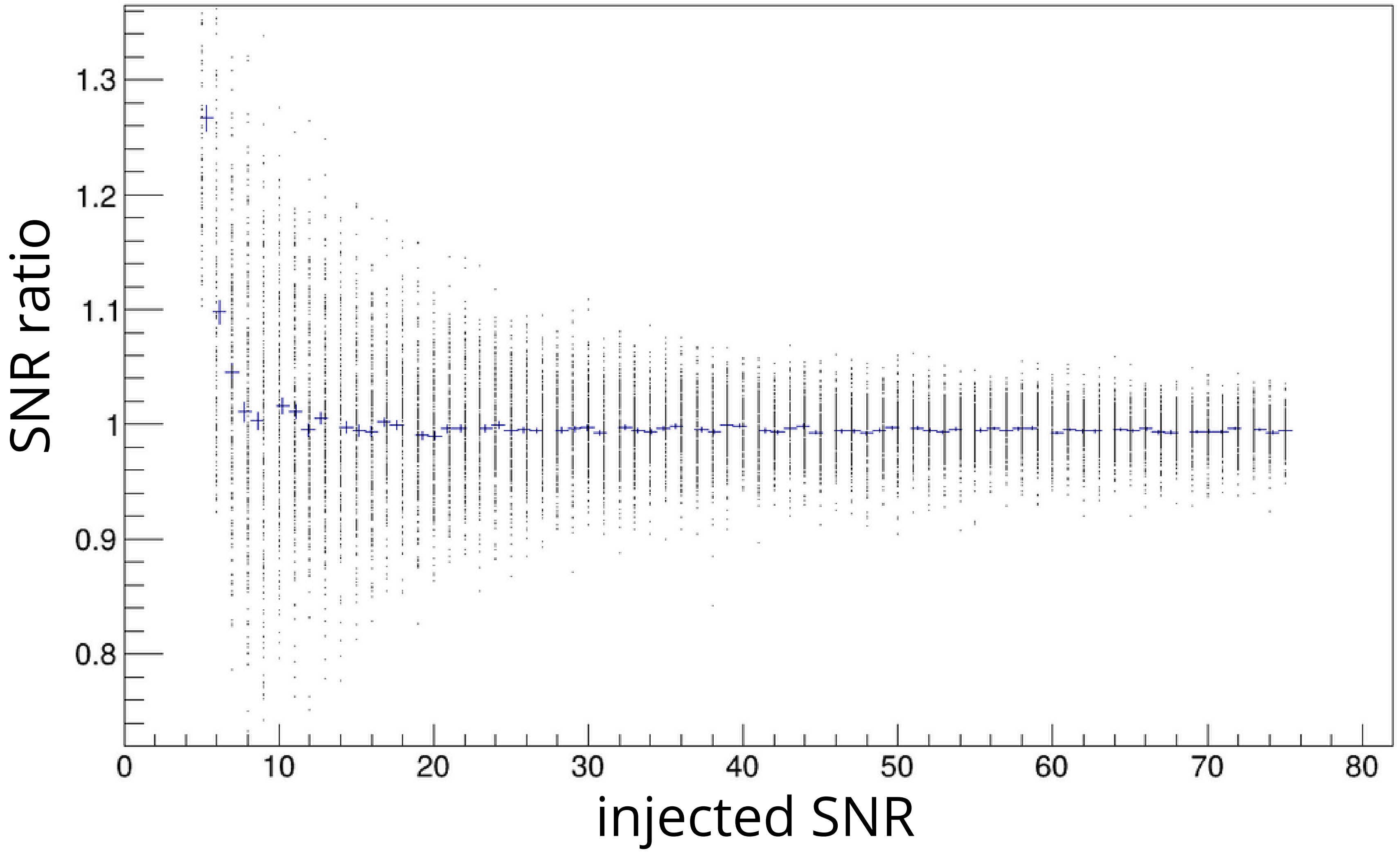
Fast bank testing with MBTA



Gaspard Joubert
IP2I Lyon

January 13th 2025

Context

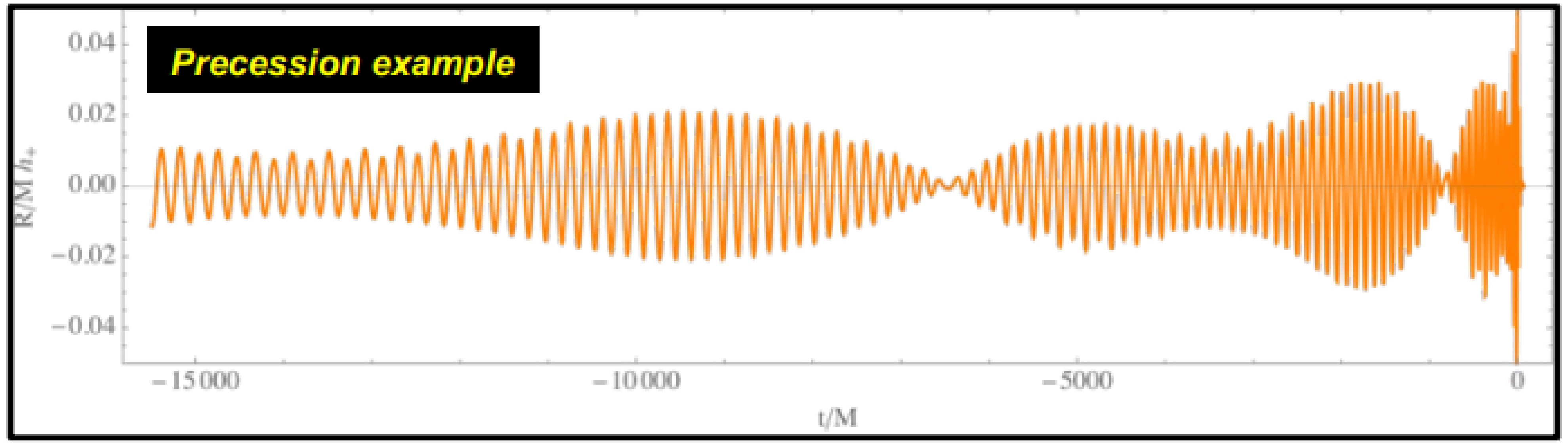
What do we want ?

- Following of the work presented in September during an MBTA meeting, check that the results of the tool are coherent with the MBTA results

- Study the performance of MBTA template bank

- Study the performance of MBTA for signals with precession

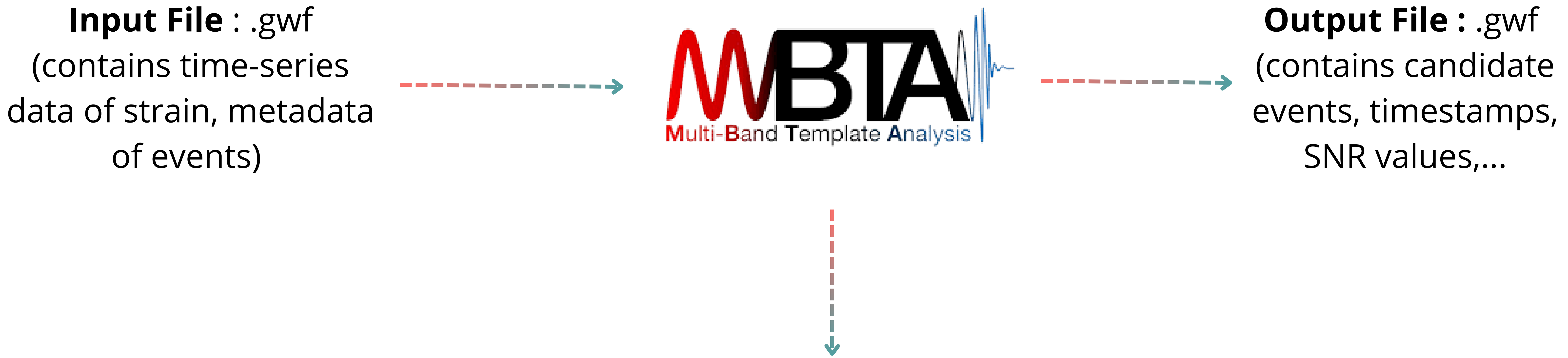
- Optimize Lorenzo Mobilia's code to create 3-band banks



O4a precessing injections fitting factor from MBTA template banks - S.Viret
https://wiki.ligo.org/pub/CBC/Searches/MBTA/2024/MBTA_100924.pdf

MBTA (Multi-Band Template Analysis)

How MBTA approximately work



Input File : .gwf
(contains time-series data of strain, metadata of events)

Output File : .gwf
(contains candidate events, timestamps, SNR values,...

data pre-processing, matched filtering, event triggers...

Lots of options :

MBTA_BANK_VT (Big bank containing all virtual template)

MBTA_BANK_RT* (Smaller banks containing fewer signal to speed up the process)

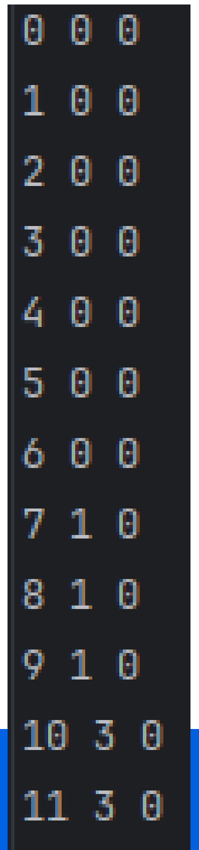
MBTA (Multi-Band Template Analysis)

Why MBTA is fast

~Half the SNR**2 is in the [24,80]Hz band and [80, 2048]Hz band
=> We make smaller bank in these bands

MBTA_BANK_VT : Big bank with 815555 templates
LF_bank : contains 54071 templates
HF_bank : contains 19168 templates
VT_LFHF_match : File matching LF et HF template to Virtual templates

VT_LFHF_match :



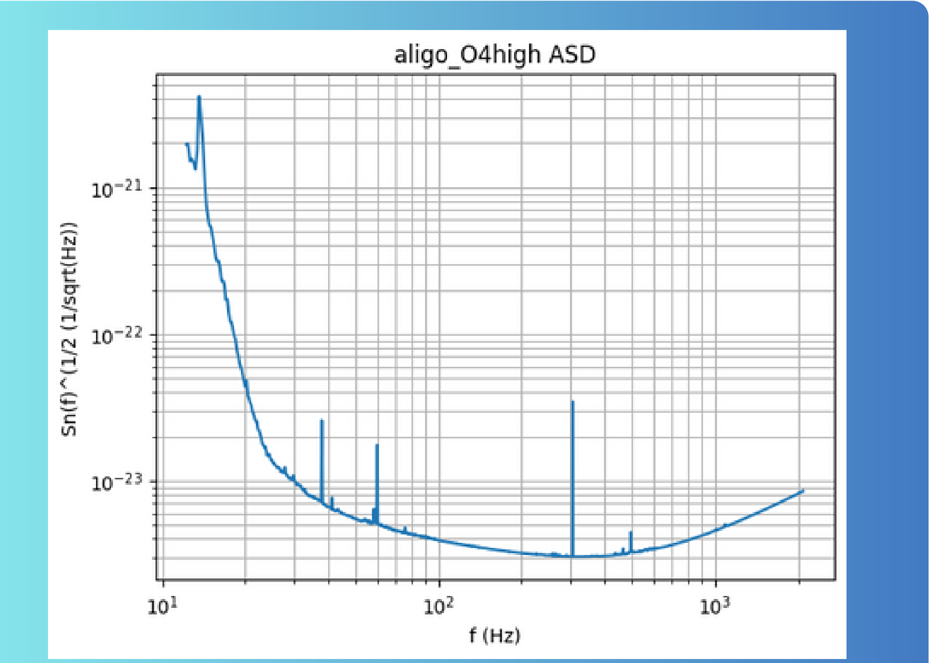
```
0 0 0
1 0 0
2 0 0
3 0 0
4 0 0
5 0 0
6 0 0
7 1 0
8 1 0
9 1 0
10 3 0
11 3 0
```

~~815555 templates between 24 and 2048 Hz~~
-> ~73000 smaller templates on 2 frequency bands

Method

Producing simulated data for MBTA

For Machine learning studies



We produce long frames of simulated noise+injections and simulated noise only

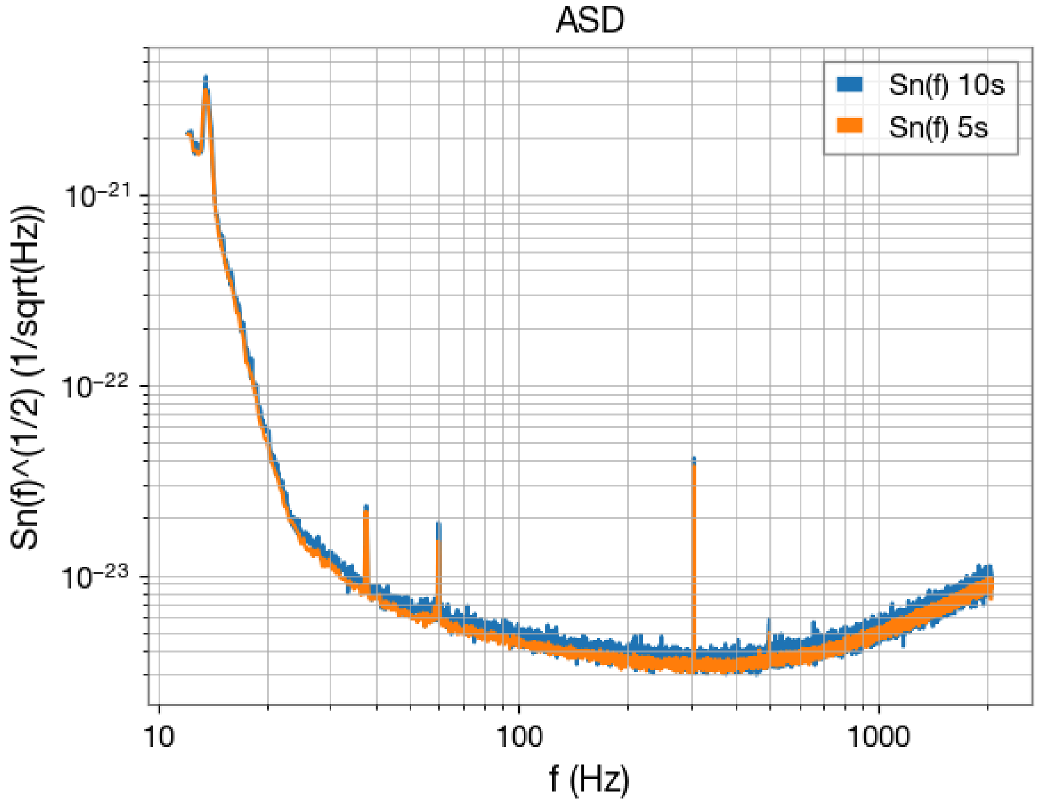
Use them in ML studies

Convert into a .gwf file

.gwf file with 2 channel :

- noise only
- signal + noise

Use MBTA to produce an ASD file



Run MBTA with this PSD file on the s+n channel up to the clusterization

How do we check ?

- Retrieve MBTA output
- Compare the SNR obtained to the injected

Results

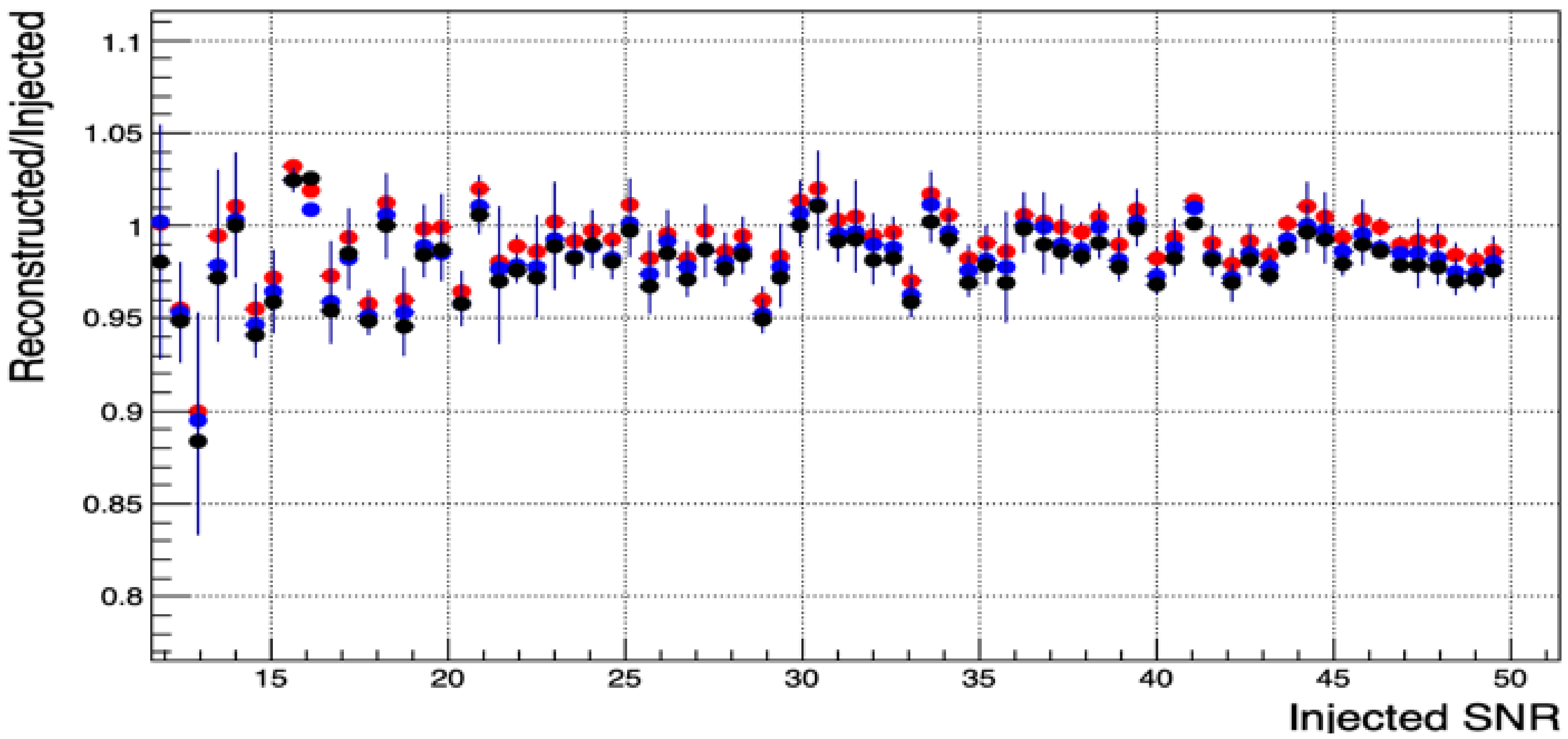
Step one : we inject 1 template and run MBTA with the same template using option MBTA_TEMPLATE_ALL_BANDS

- m1=m2=15 solar mass

- For 1 band, we obtain a fitting factor of 0.993

- For 2 and 3 bands, the fitting factor is a bit lower

=>The injected data is well reconstructed by MBTA
This validate our frame generator



Red : 1 frequency band [24, 2048]
Blue : 2 frequency bands [24, 80, 2048]
Black : 3 frequency bands [24, 55, 115, 2048]

Results

More complex case : a subset of the bank.

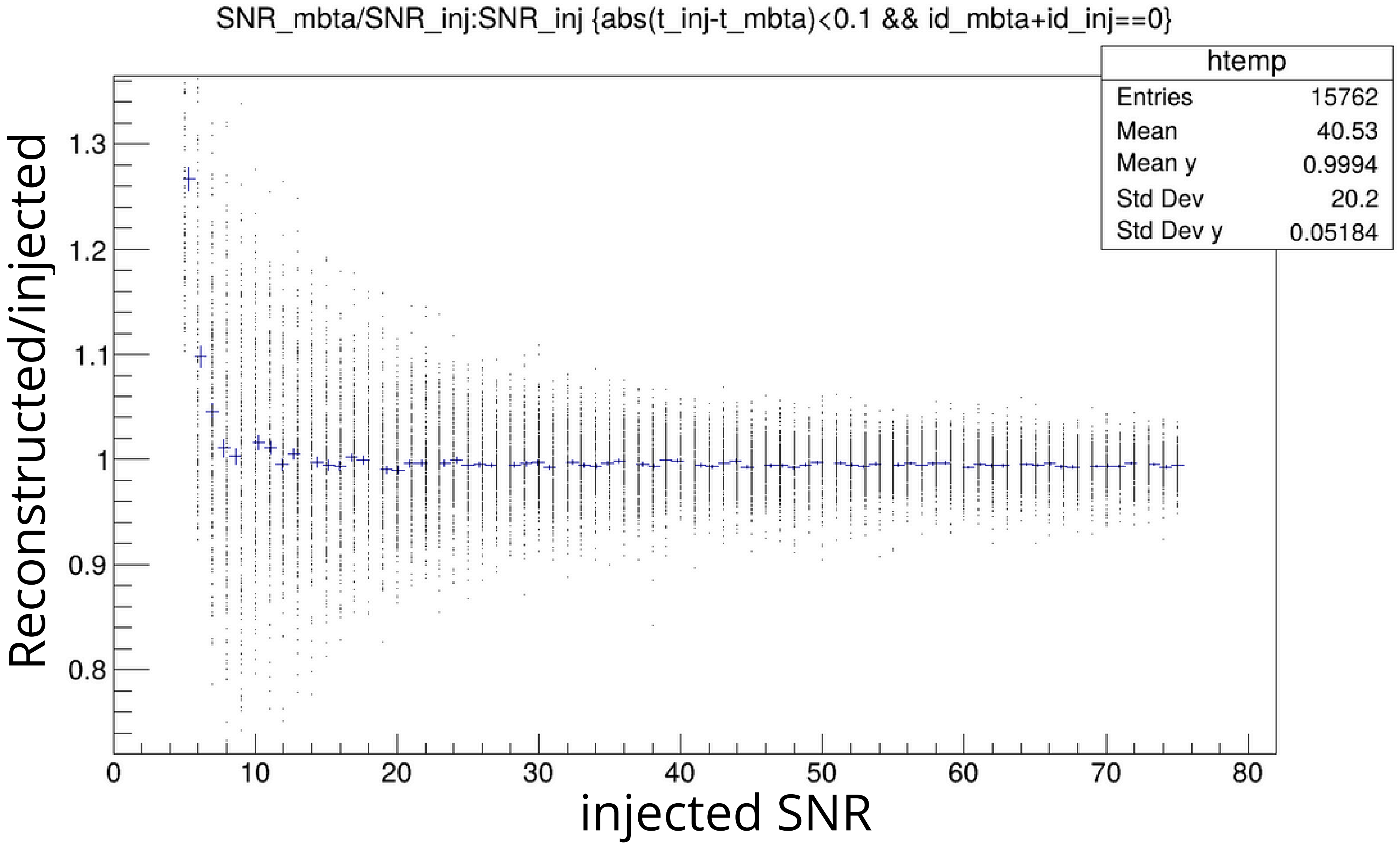
- A subset of 100 templates randomly selected from the bank “VT_bank_2bands”

- We recreate the necessary file for MBTA (LF_bank, HF_bank and VT_LFHF_match)

- Analyzed using 1 band [24, 2048]

=> The fitting factor is nearly 1

⚠ The detection treshold was set to 5, introducing a bias at low SNR.



Results

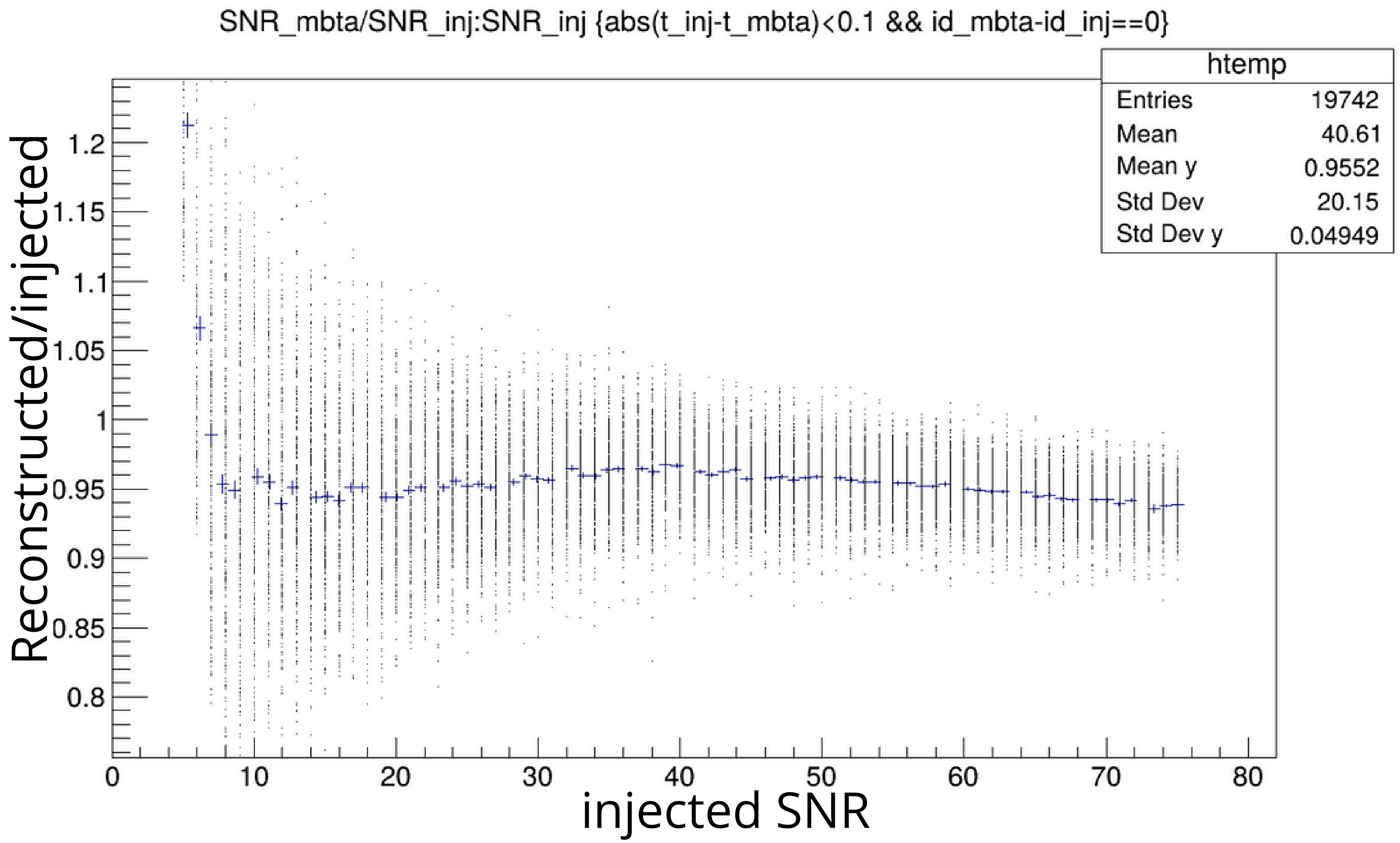
More complex case : a subset of the bank

- Same process as before

- Analyzed using 2 bands [24, 80, 2048]

=> The fitting factor is a bit lower ~0.95
Under study

⚠ Same bias at low SNR



VIRGO Quick results signals with precession

Testing on signals with precession

- We make frames using signals randomly selected from the O4a simulated injections with X_p>0.5 and M_ratio >0.3

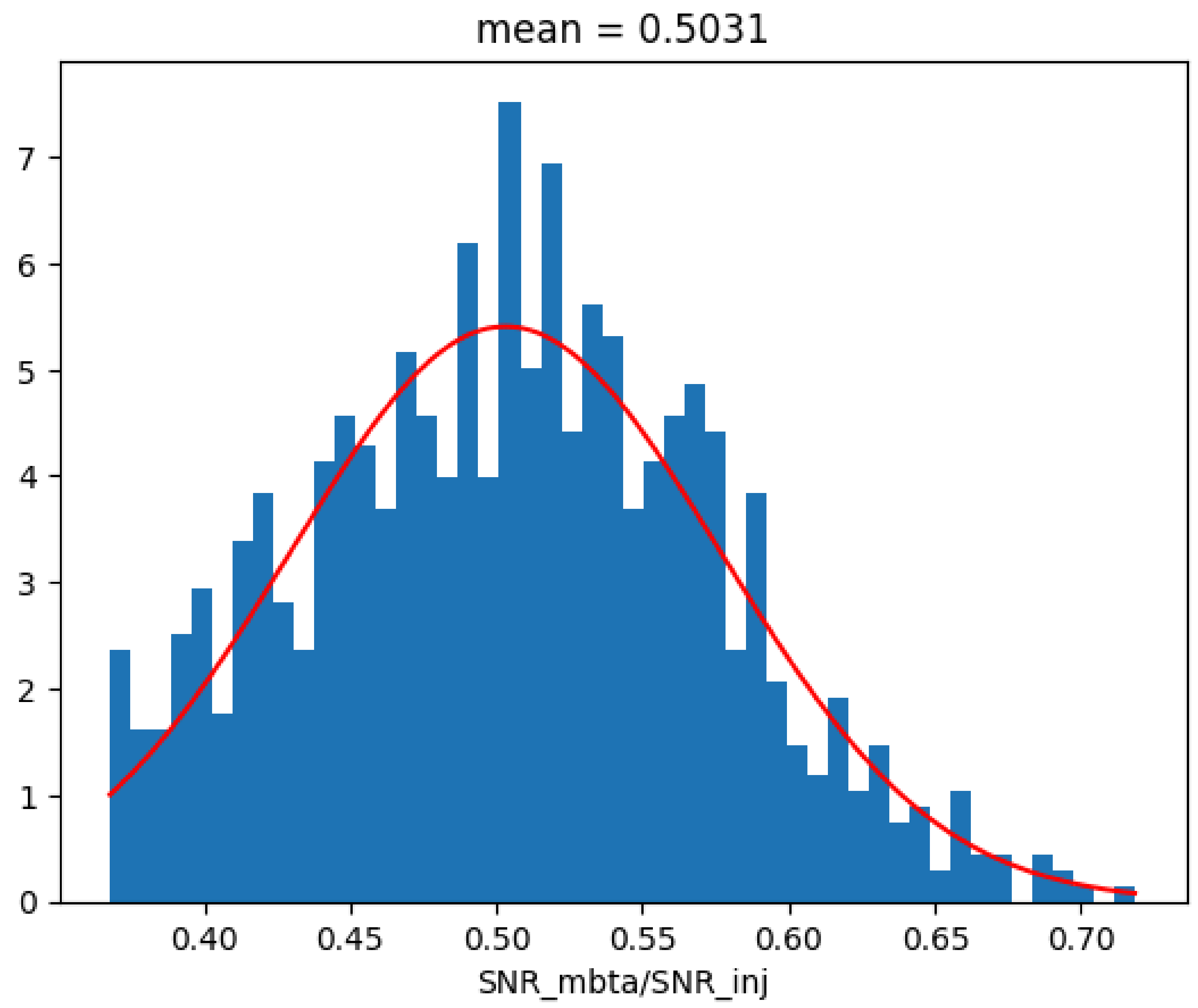
- all SNR=15

- We analyze them using the "VT_bank_1band" containing 9381 templates.

- Analyzed using 1 band [24, 2048]

- 95% of signals detected

- Preliminary results, work in progress



Trying do create sub-banks for tri-band

Optimizing PyCBC match function

- During MBTA F2F :
RT banks production by Lorenzo Mobilia :

- Bank production + verification with “rt-creation-mbta”
  - Divide the whole bank “VT_bank_2bands” into overlapping batches
  - In each batch, loop on each indices to calculate the fitting factor between templates of the batch using PyCBC math function
  - If ff>0.98, the associated template is removed

RT banks + association build in < O(12h)
=> We try to optimize it

- ~~PyCBC~~ -> Homemade function
We developped a match function with Numpy and Scipy

Speed at least x3/x4 can be more depending of the platform used(Lorenzo will verify the true value)

Values differences between our method and PyCBC are negligable

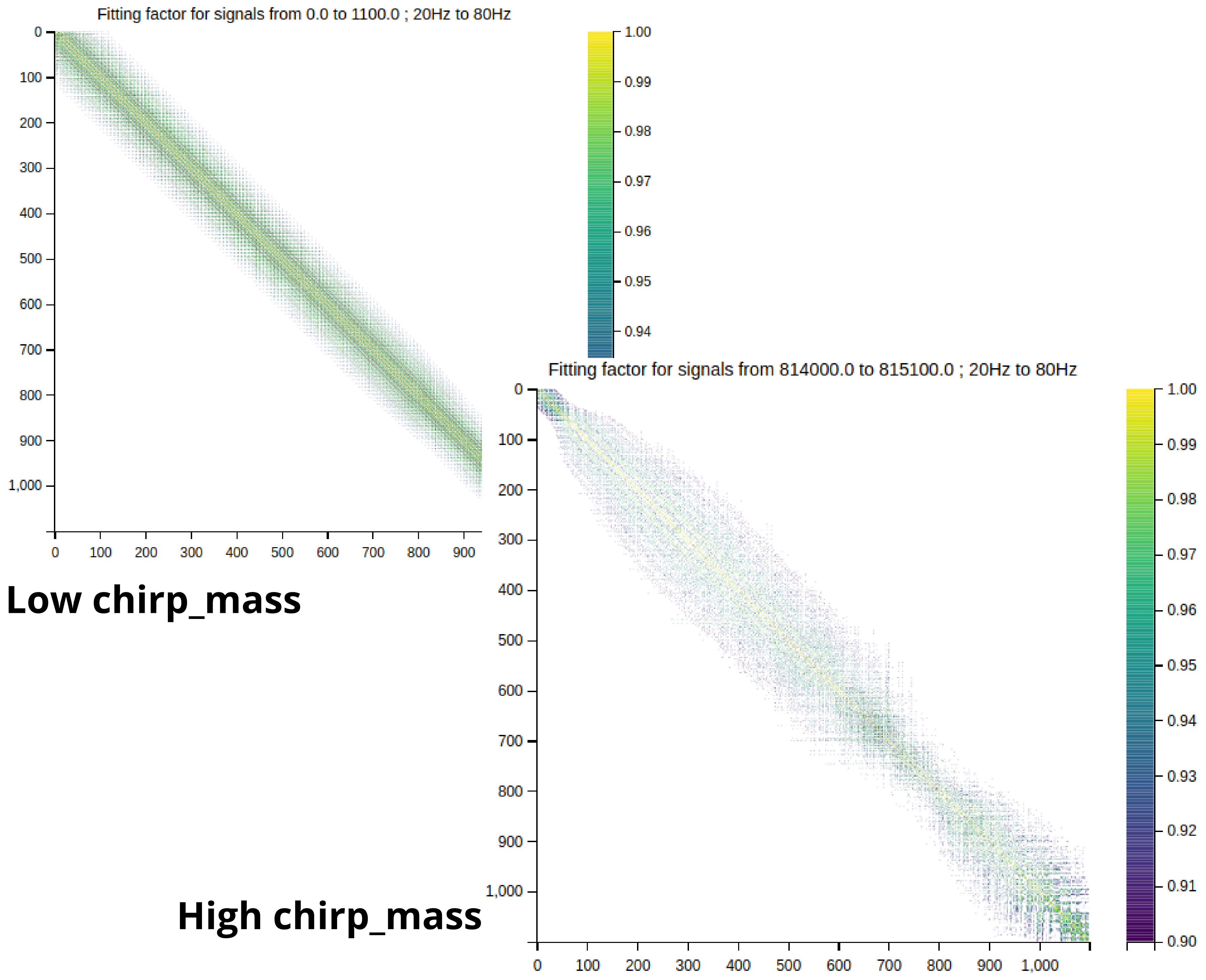
For 100000 fitting factor computation
On the CC : Pycbc match time : 298 seconds
Custom match custom : 13 seconds
} Ratio -> 22

On my laptop : Pycbc match time : 665 seconds
Custom match custom : 4.6 seconds
} Ratio -> 141

Trying do create sub-banks

Fitting factor between bank templates

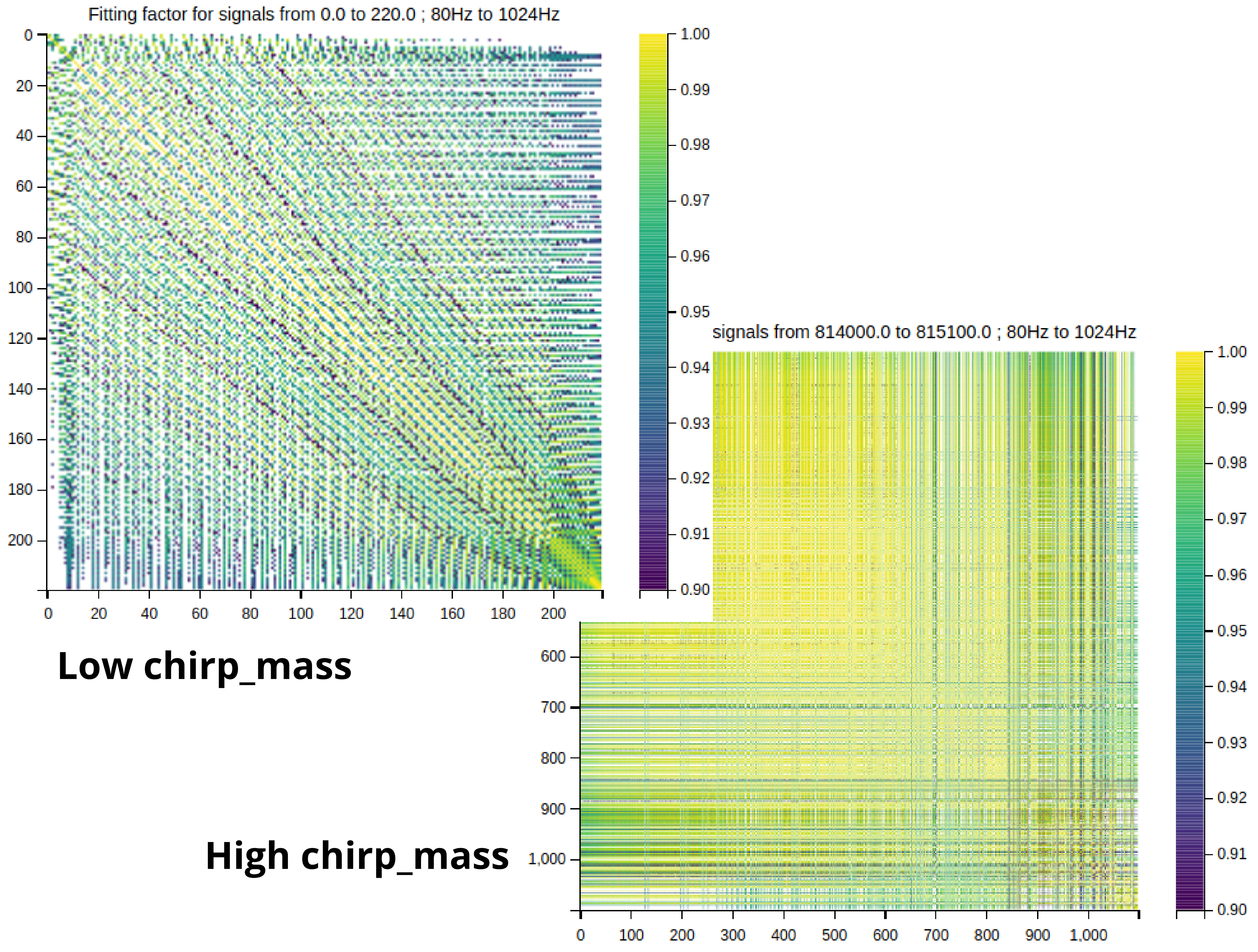
LF [20Hz, 80Hz] (only = ff>0.9)



Fitting factor for signals from 0.0 to 1100.0 ; 20Hz to 80Hz

Fitting factor for signals from 814000.0 to 815100.0 ; 20Hz to 80Hz

HF [80Hz, 1024Hz] (only ff>0.9)



Fitting factor for signals from 0.0 to 220.0 ; 80Hz to 1024Hz

signals from 814000.0 to 815100.0 ; 80Hz to 1024Hz

Conclusion

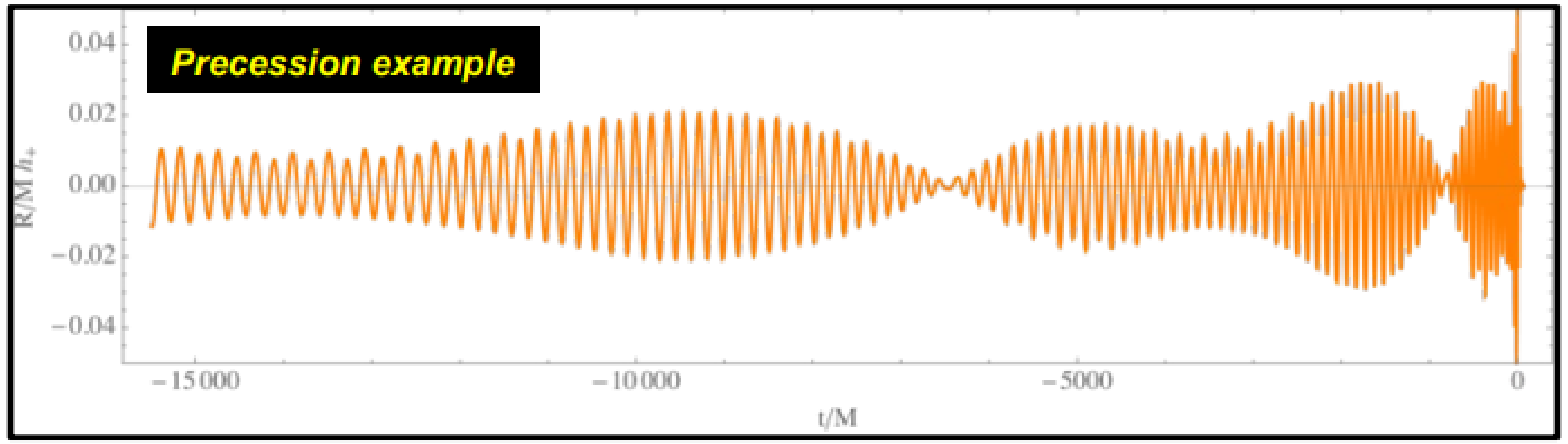
- We have a functional tool capable of testing template banks quickly

- A slight decrease in performance is observed for multiple bands or certain templates

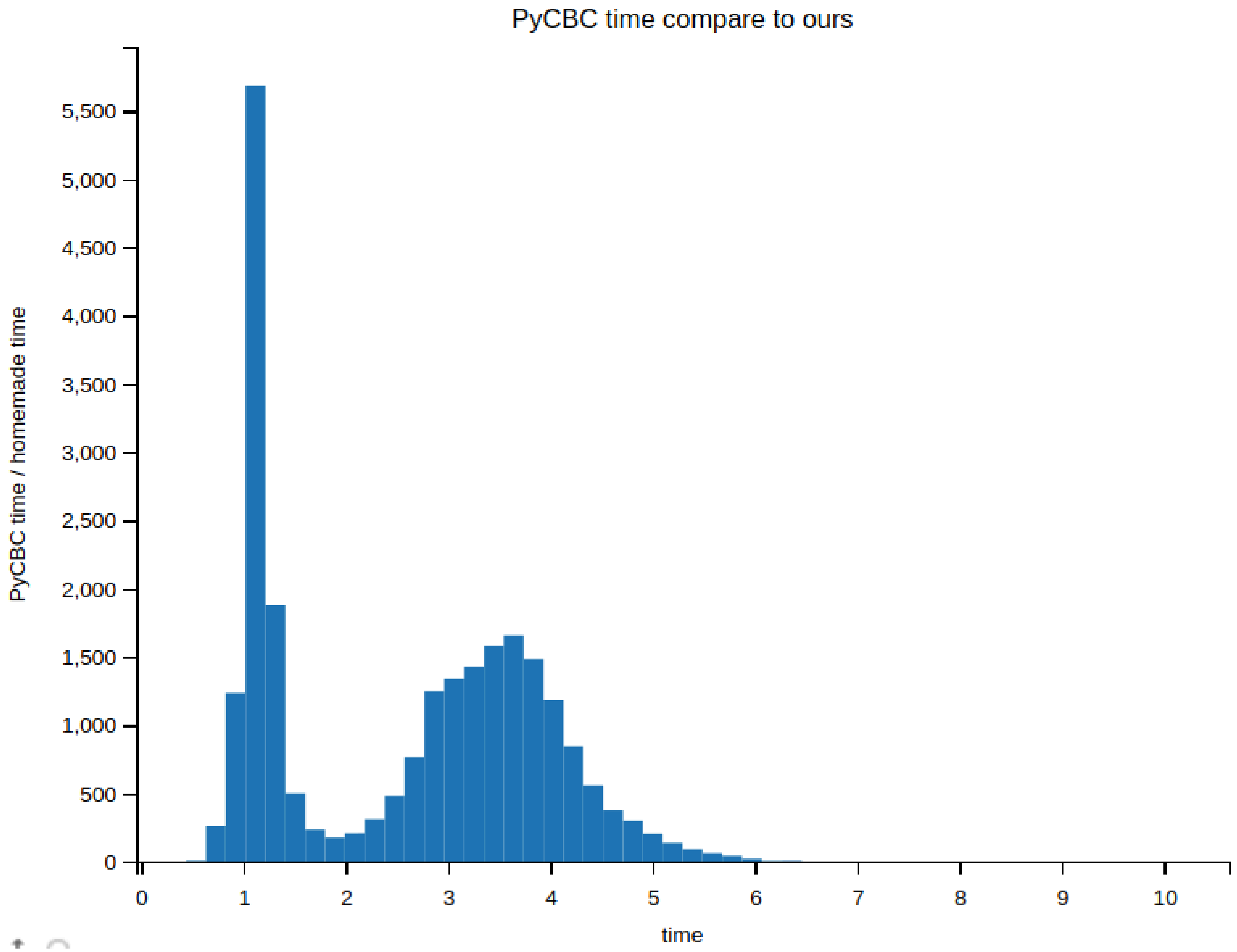
-We have now a framework to compare MBTA performance to the ML networks we are developping

- We can continue test banks for precessing signals

- We continue to optmize Lorenzo algorithm to make a 3-band bank



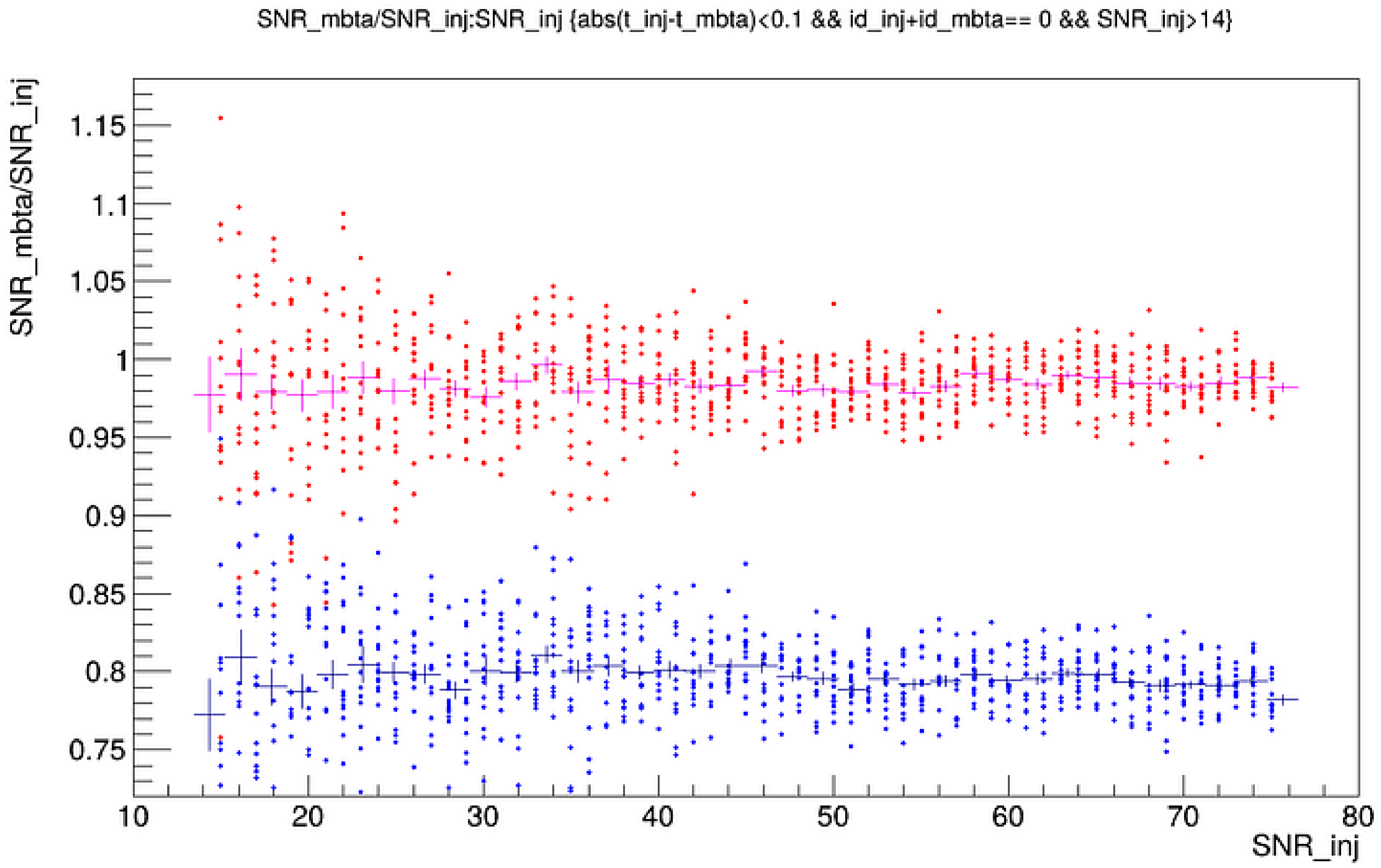
Thank you for your attention :)


PyCBC time compare to ours
PyCBC time / homemade time
5,500
5,000
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
0
1
2
3
4
5
6
7
8
9
10
time

- Same process as before but only for template 815532

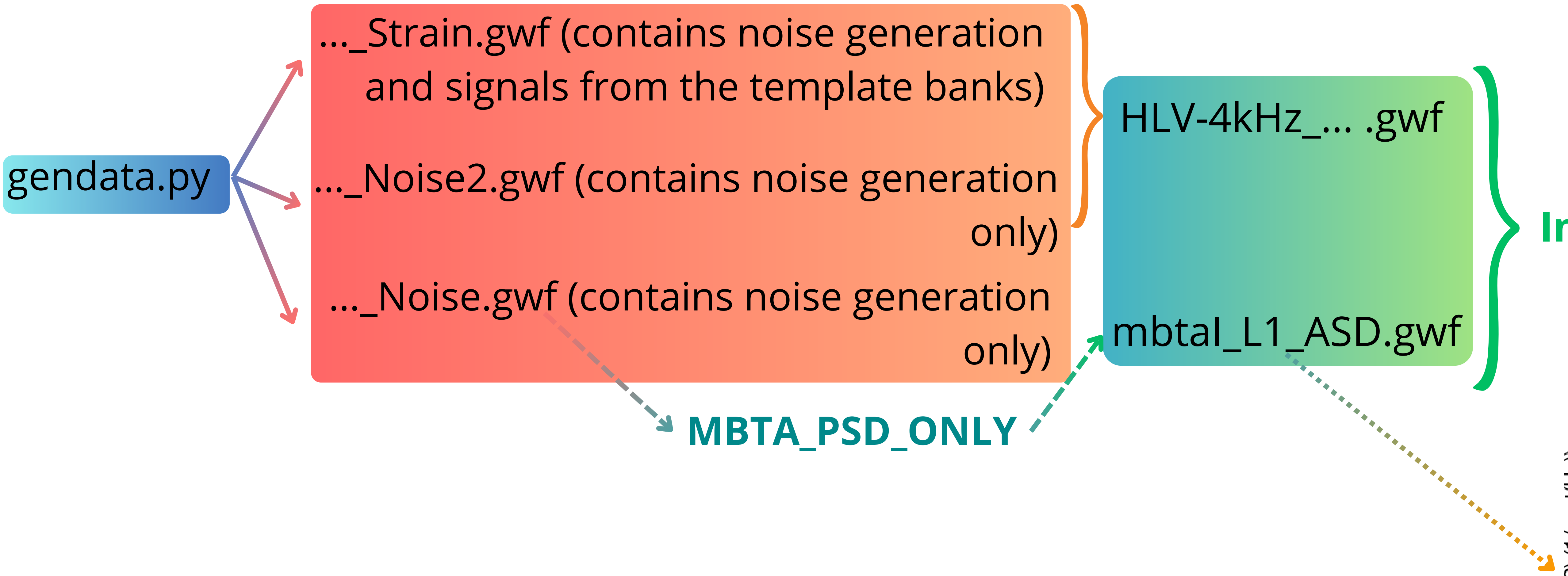
- Analyzed using 1 bands [24, 2048] in Red

- Analyzed using 2 bands [24, 80, 2048] in Blue



Method

Producing simulated data for MBTA



gendata.py

…_Strain.gwf (contains noise generation and signals from the template banks)

…_Noise2.gwf (contains noise generation only)

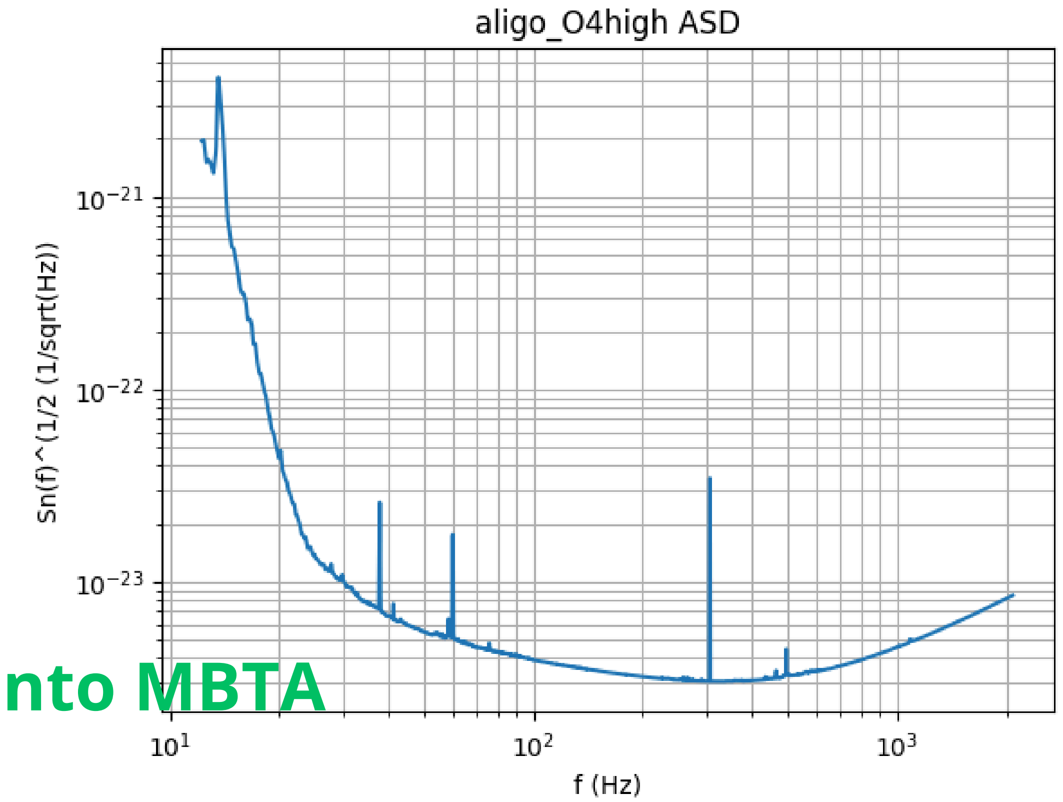
…_Noise.gwf (contains noise generation only)

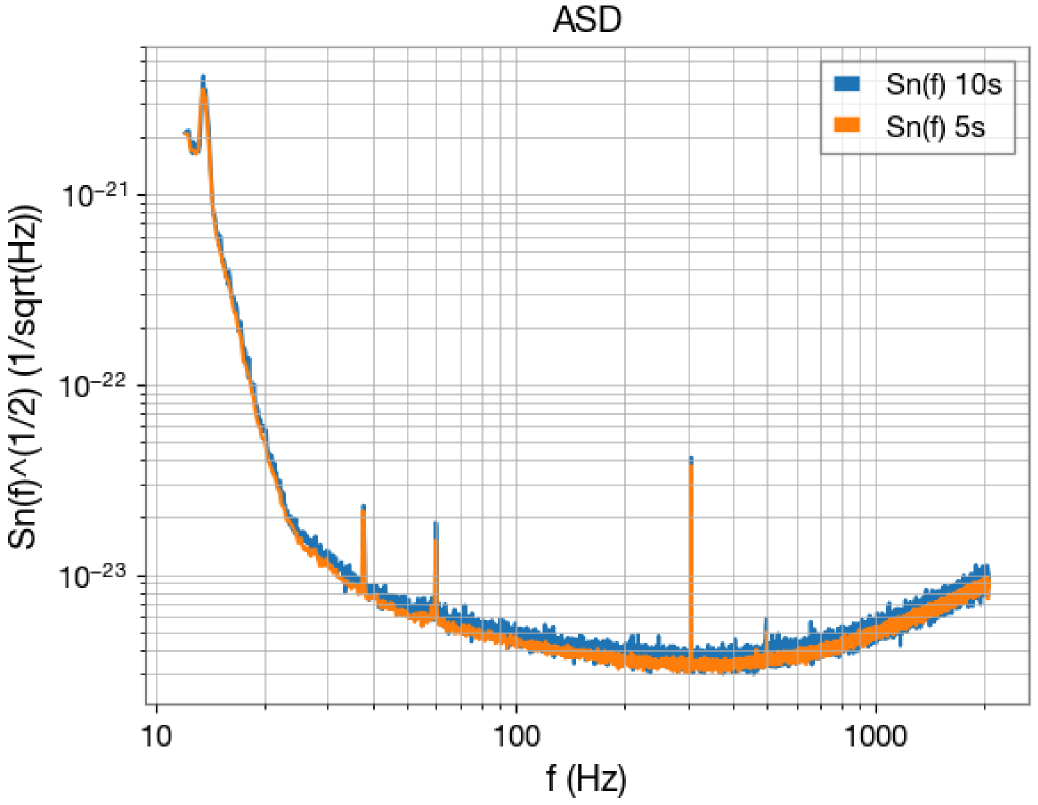
MBTA_PSD_ONLY

HLV-4kHz_… .gwf

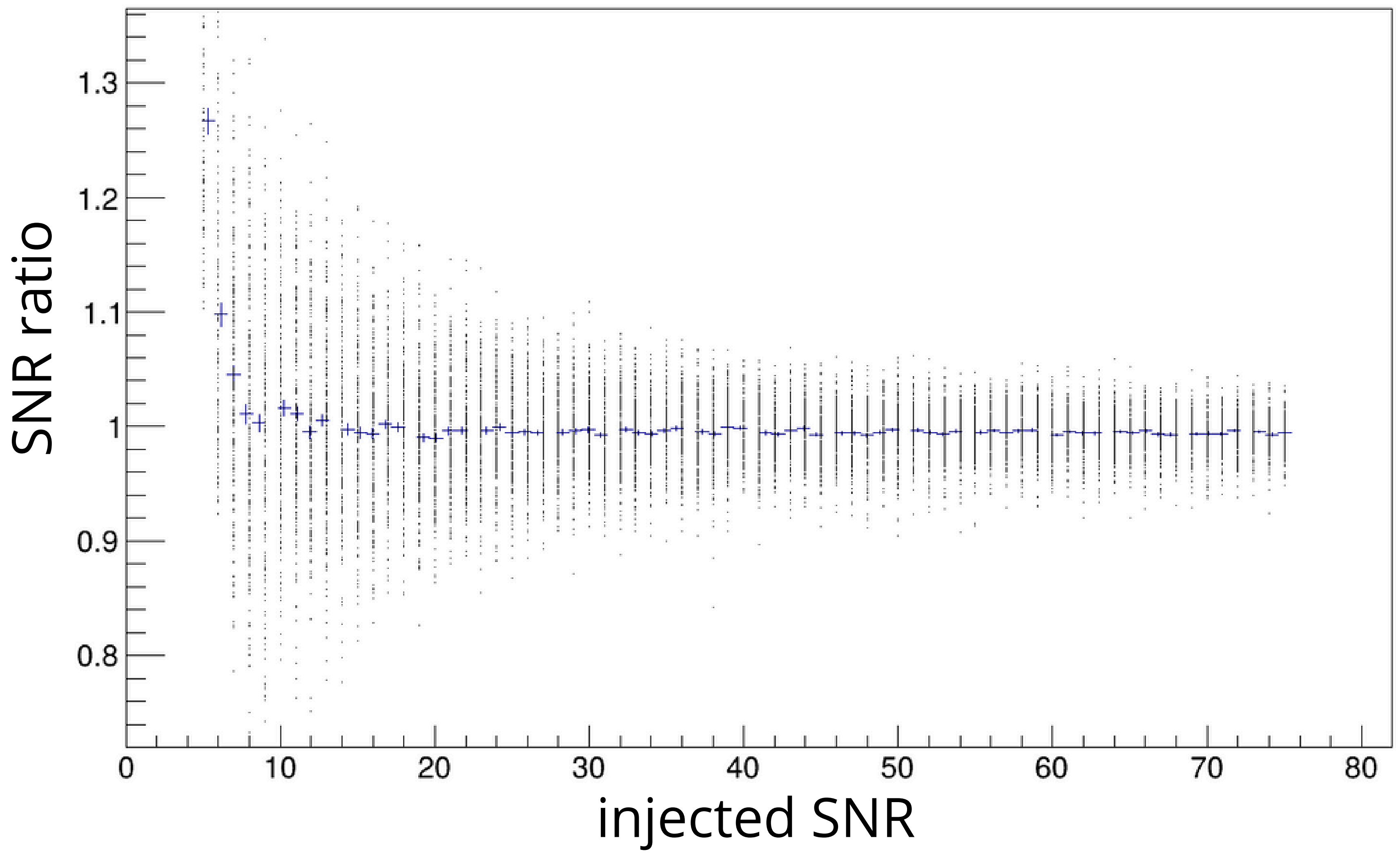
mbtaI_L1_ASD.gwf

Into MBTA





Fast bank testing with MBTA



Gaspard Joubert
IP2I Lyon

Décember 17th 2024

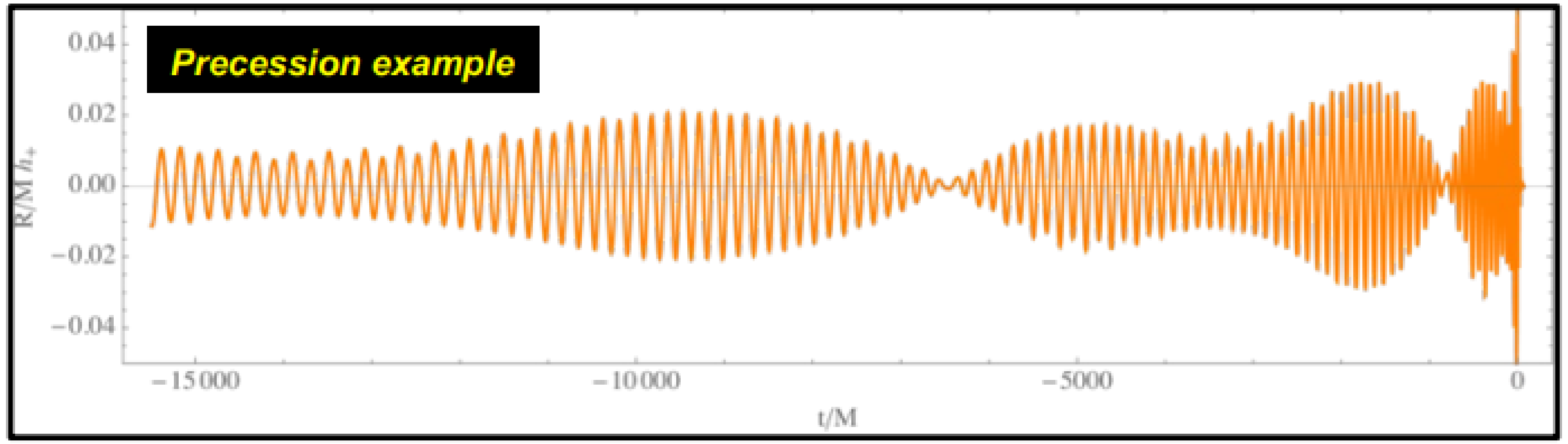
Context

What do we want ?

- Following of the work presented in September, check that the results of the tool are coherent with the MBTA results

- Study the performance of MBTA template bank

- Study the performance of MBTA template bank

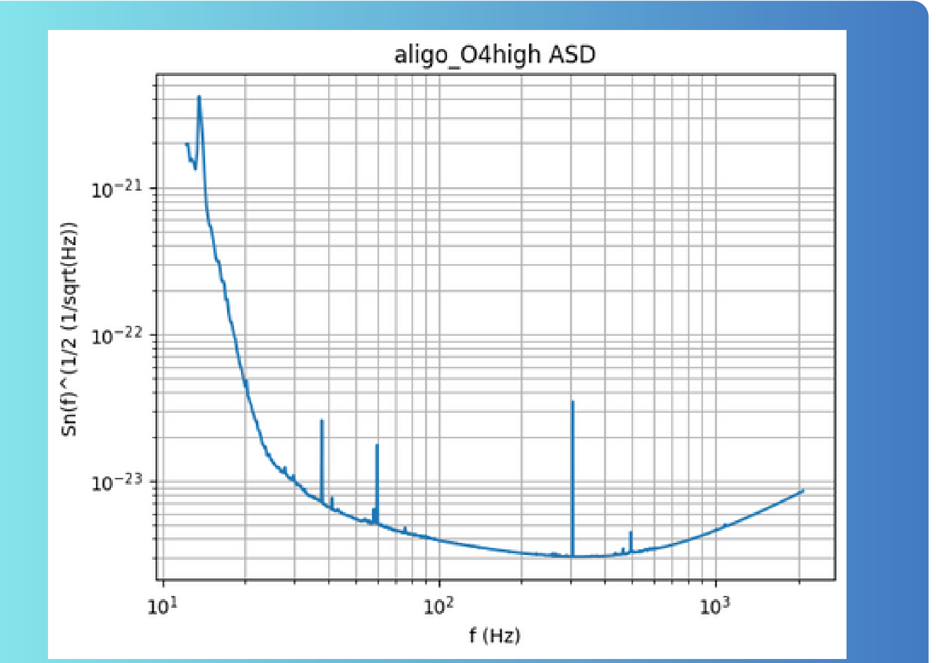


O4a precessing injections fitting factor from MBTA template banks - S.Viret
https://wiki.ligo.org/pub/CBC/Searches/MBTA/2024/MBTA_100924.pdf

Method

Producing simulated data for MBTA

For Machine learning studies



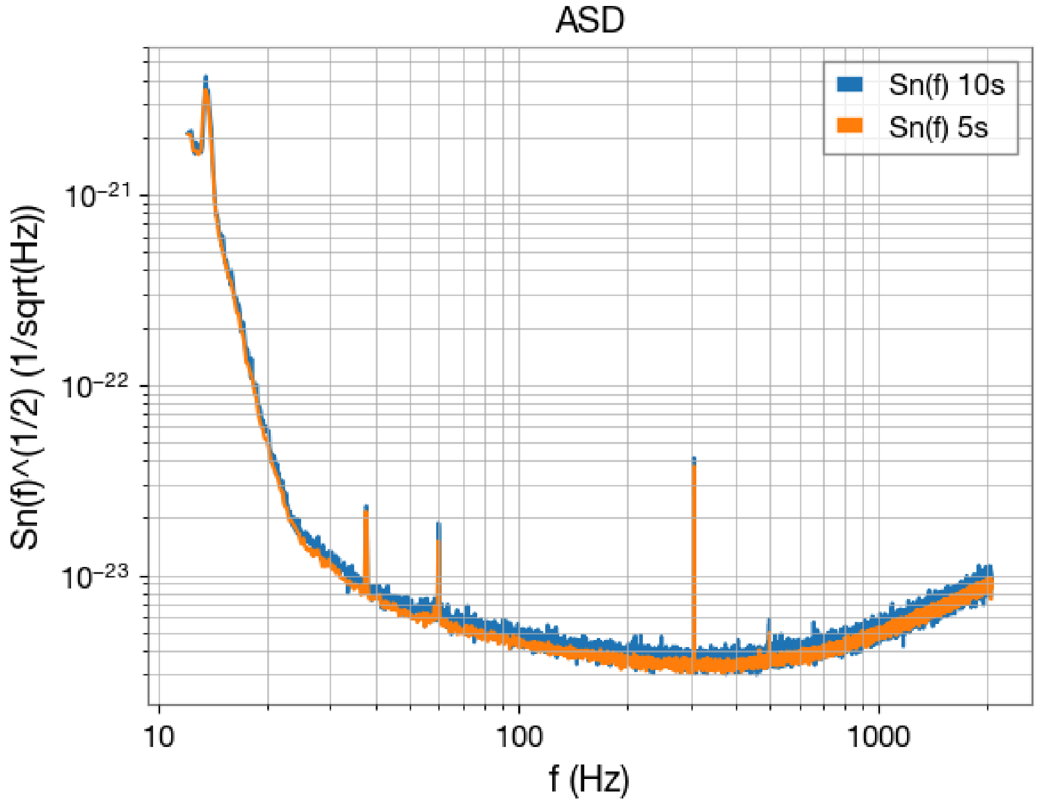
We produce long frames of simulated noise+injections and simulated noise only

Use them in ML studies

Convert into a .gwf file

.gwf file with 2 channel :
- noise only
- signal + noise

Use MBTA to produce an ASD file



Run MBTA with this PSD file on the s+n channel up to the clusterization

How do we check ?

- Retrieve MBTA output
- Compare the SNR obtained to the injected

Results

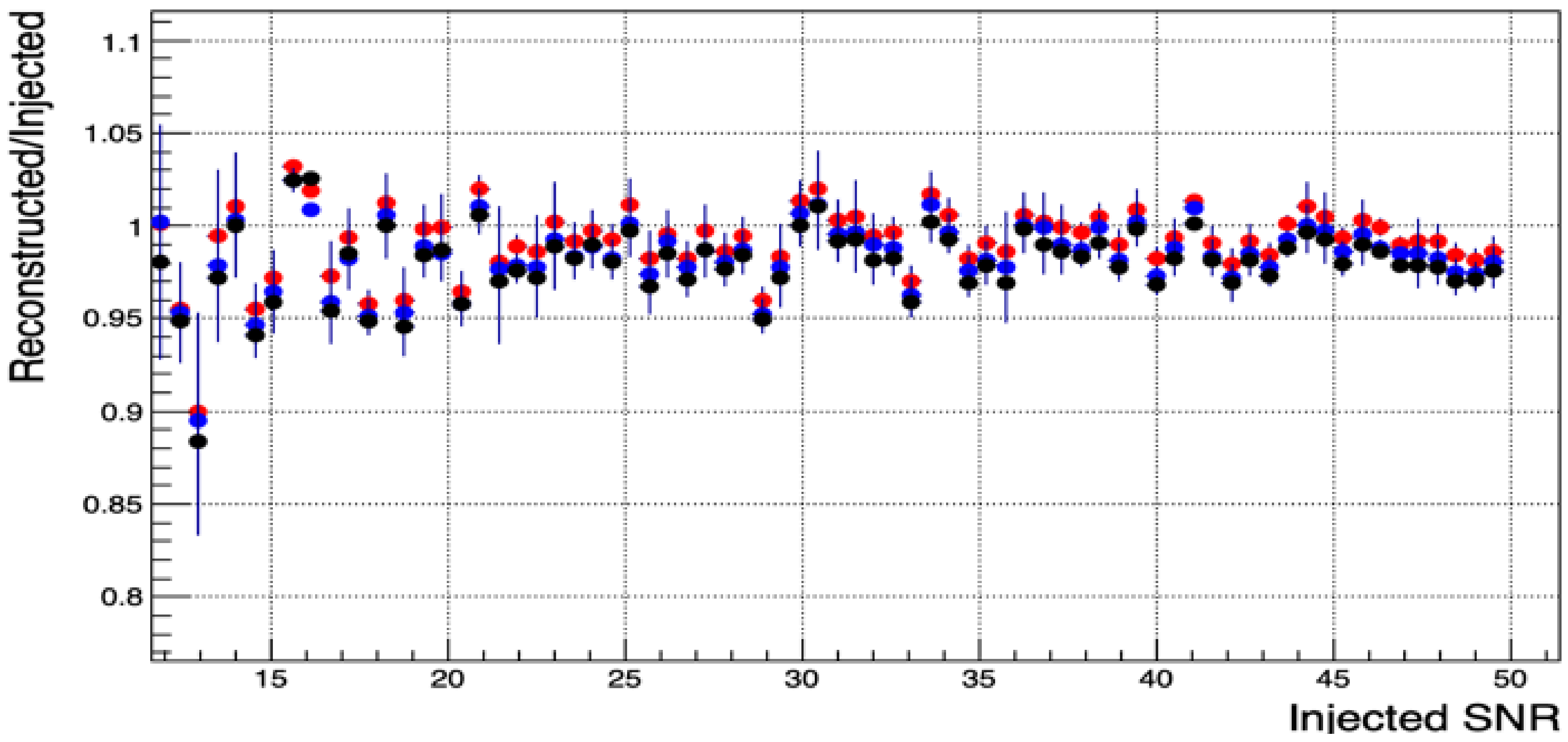
Step one : we inject 1 template and run MBTA with the same template using option MBTA_TEMPLATE_ALL_BANDS

- m1=m2=15 solar mass

- For 1 band, we obtain a fitting factor of 0.993

- For 2 and 3 bands, the fitting factor is a bit lower

=>The injected data is well reconstructed by MBTA
This validate our frame generator



Red : 1 frequency band [24, 2048]
Blue : 2 frequency bands [24, 80, 2048]
Black : 3 frequency bands [24, 55, 115, 2048]

Results

More complex case : a subset of the bank.

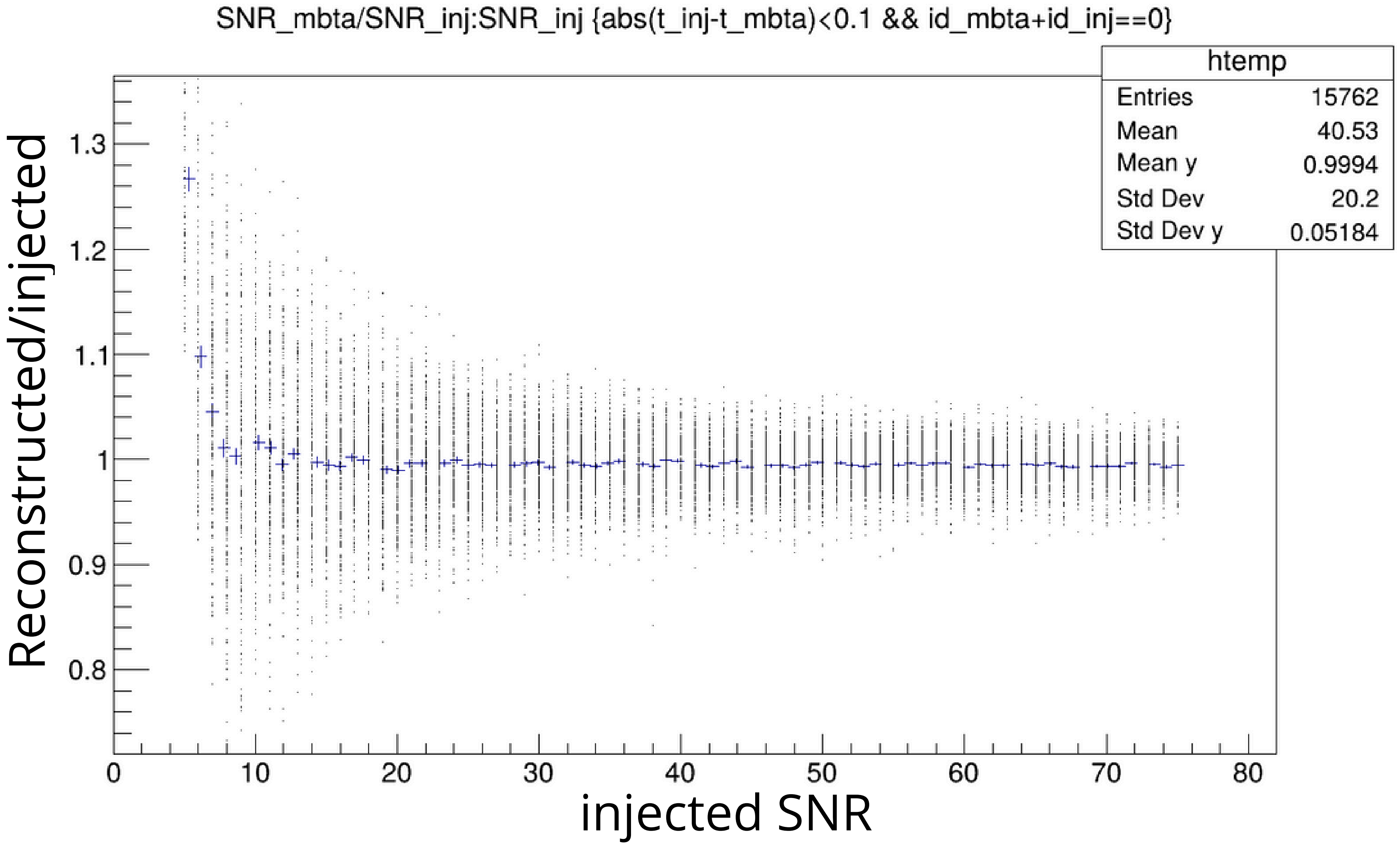
- A subset of 100 templates randomly selected from the bank “VT_bank_2bands”

- We recreate the necessary file for MBTA (LF_bank, HF_bank and VT_LFHF_match)

- Analyzed using 1 band [24, 2048]

=> The fitting factor is nearly 1

⚠ The detection treshold was set to 5, introducing a bias at low SNR.



Results

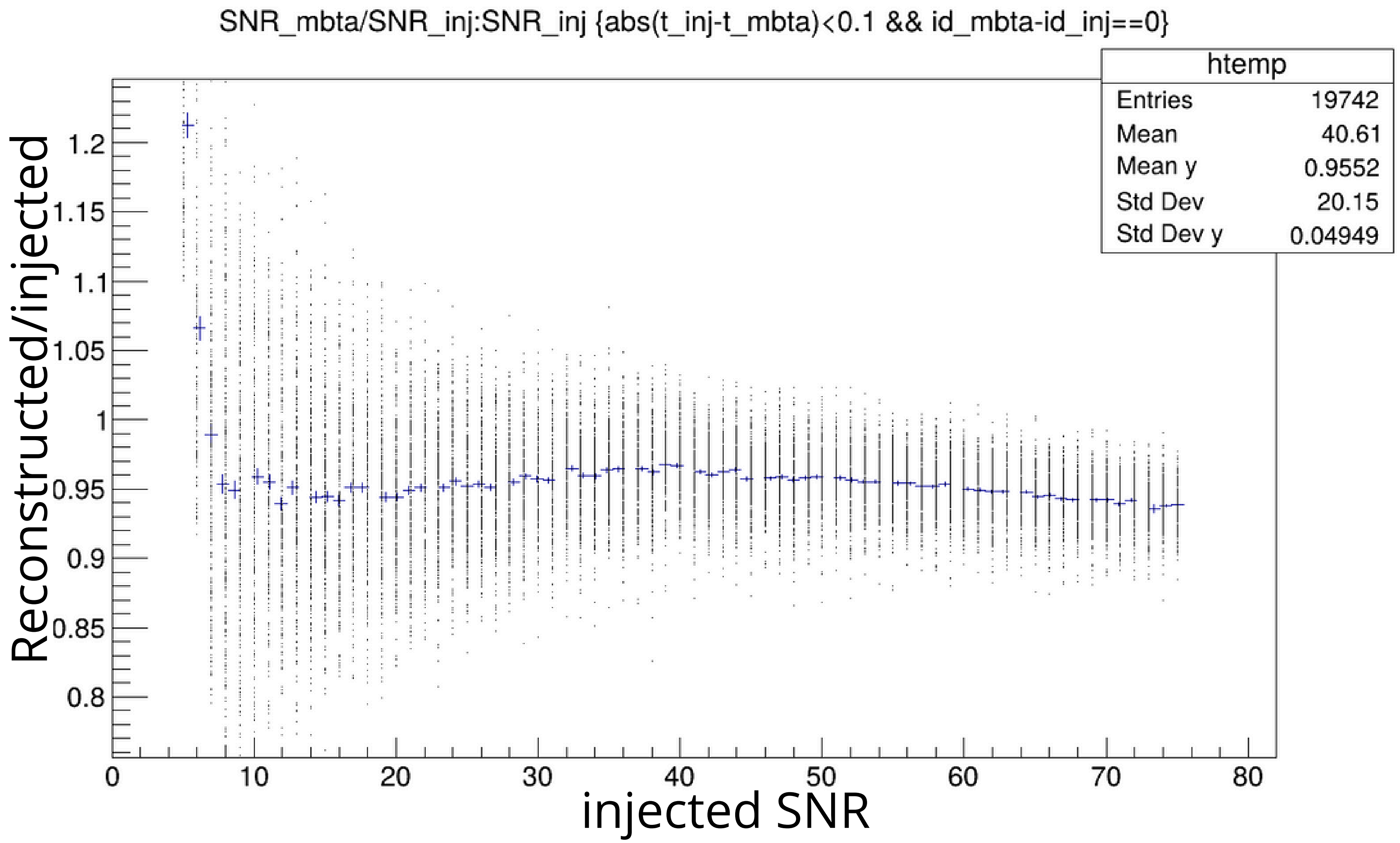
More complex case : a subset of the bank

- Same process as before

- Analyzed using 2 bands [24, 80, 2048]

=> The fitting factor is a bit lower ~0.95
Under study

⚠ Same bias at low SNR



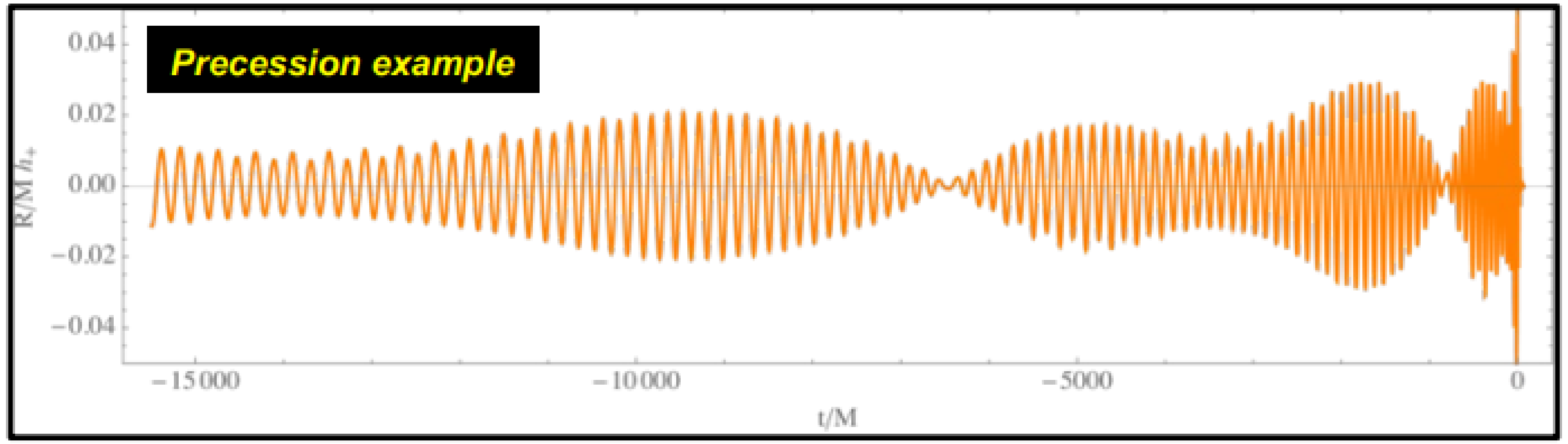
Conclusion

- We have a functional tool capable of testing template banks quickly

- A slight decrease in performance is observed for multiple bands or certain templates (Is it expected ?)

-We have now a framework to compare MBTA performance to the ML networks we are developping

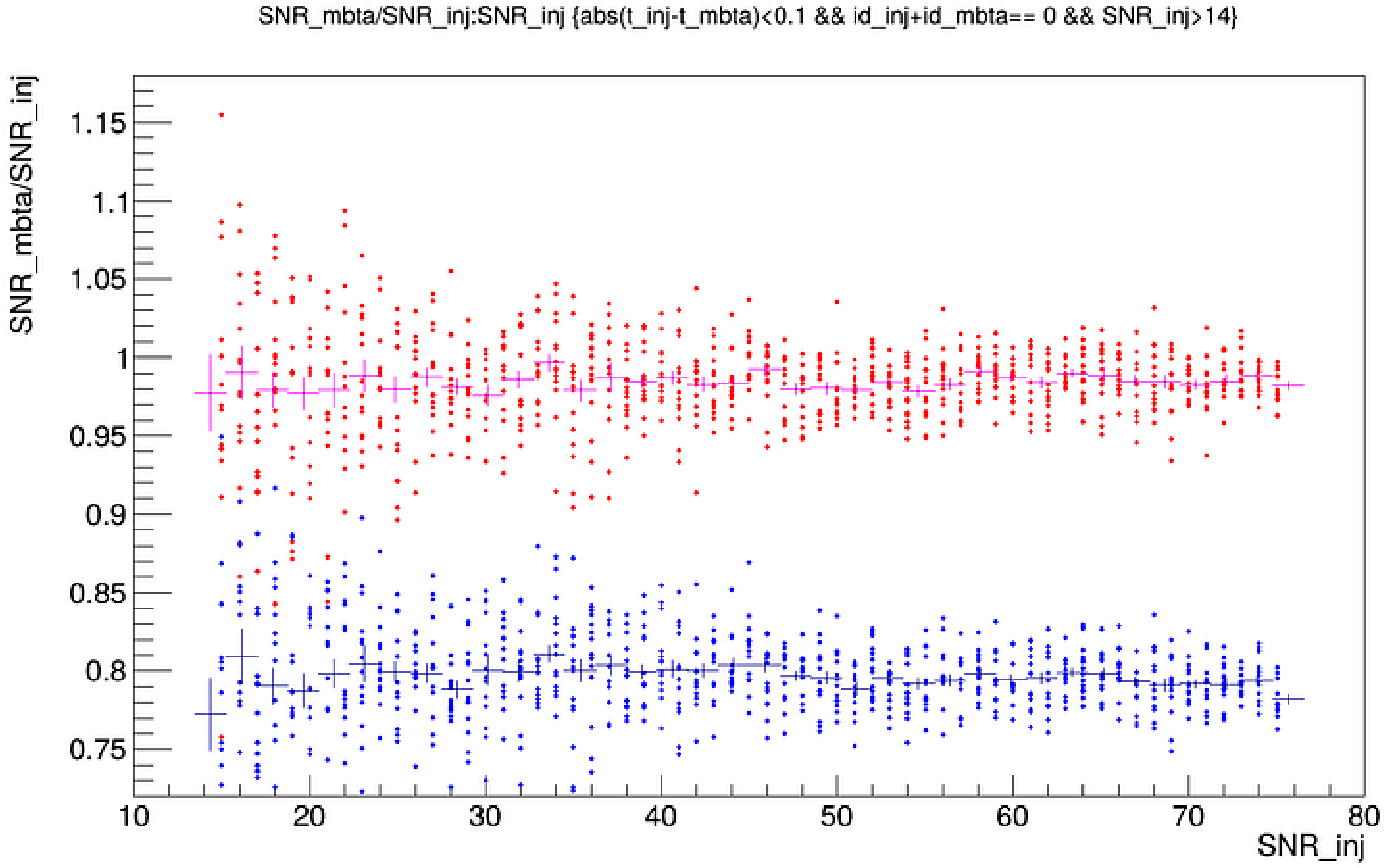
- We can test banks for precessing signals



- Same process as before but only for template 815532

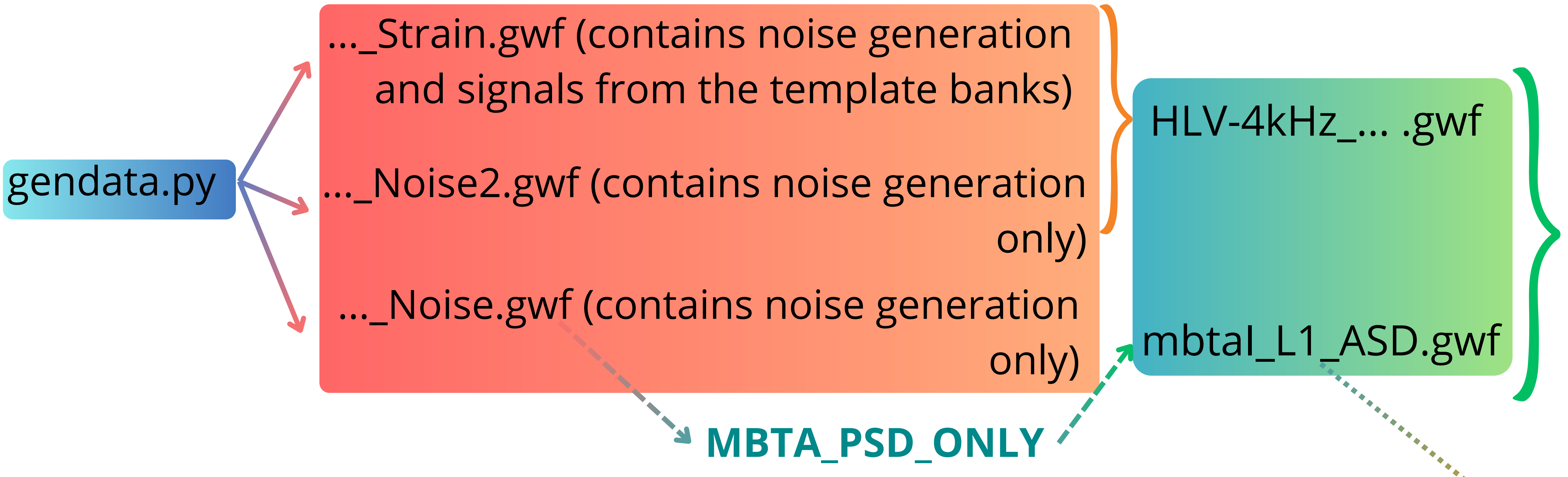
- Analyzed using 1 bands [24, 2048] in Red

- Analyzed using 2 bands [24, 80, 2048] in Blue



Method

Producing simulated data for MBTA



gendata.py

…_Strain.gwf (contains noise generation and signals from the template banks)

…_Noise2.gwf (contains noise generation only)

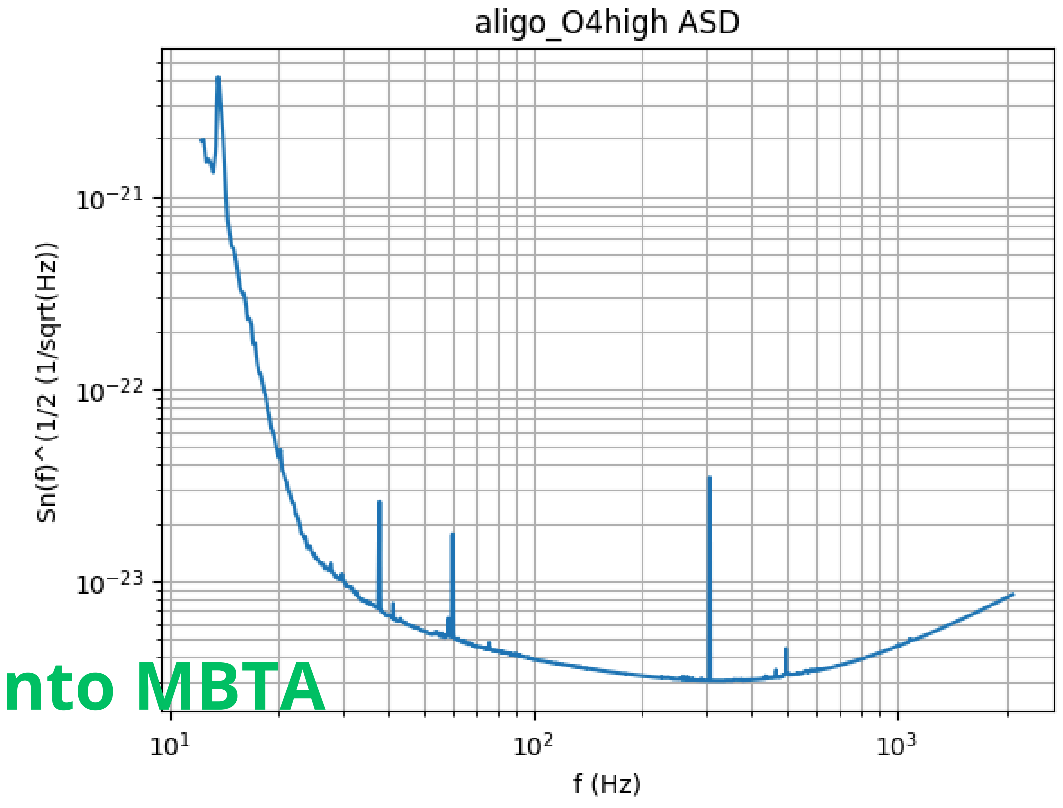
…_Noise.gwf (contains noise generation only)

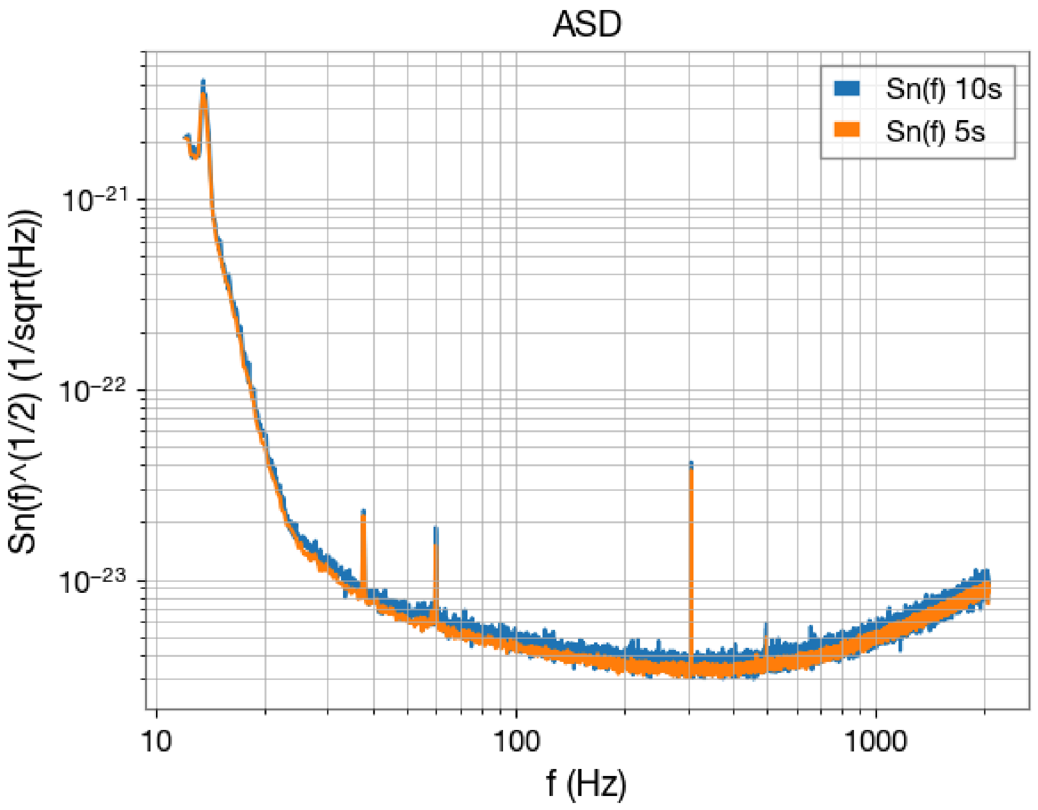
MBTA_PSD_ONLY

HLV-4kHz_… .gwf

mbtaI_L1_ASD.gwf

Into MBTA


aligo_O4high ASD
Sn(f)^(1/2 (1/sqrt(Hz))
10^{-21}
10^{-22}
10^{-23}
10^1
10^2
10^3
f (Hz)


ASD
Sn(f) 10s
Sn(f) 5s
Sn(f)^(1/2) (1/sqrt(Hz))
10^{-21}
10^{-22}
10^{-23}
f (Hz)