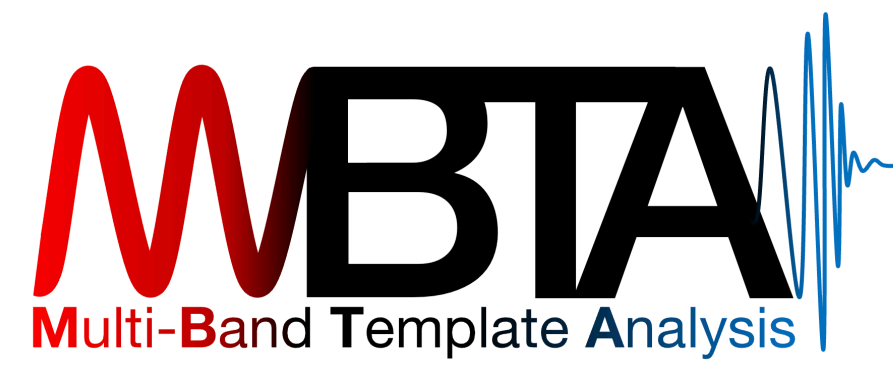




1506
UNIVERSITÀ
DEGLI STUDI
DI URBINO
CARLO BO



Istituto Nazionale di Fisica Nucleare
SEZIONE DI FIRENZE

Tecniche di Machine Learning per Analisi Dati di Onde Gravitazionali

IFAE 2024

Mobilia Lorenzo¹ e il gruppo di MBTA

03-04-2024

MBTA - Multi Band Template Analysis

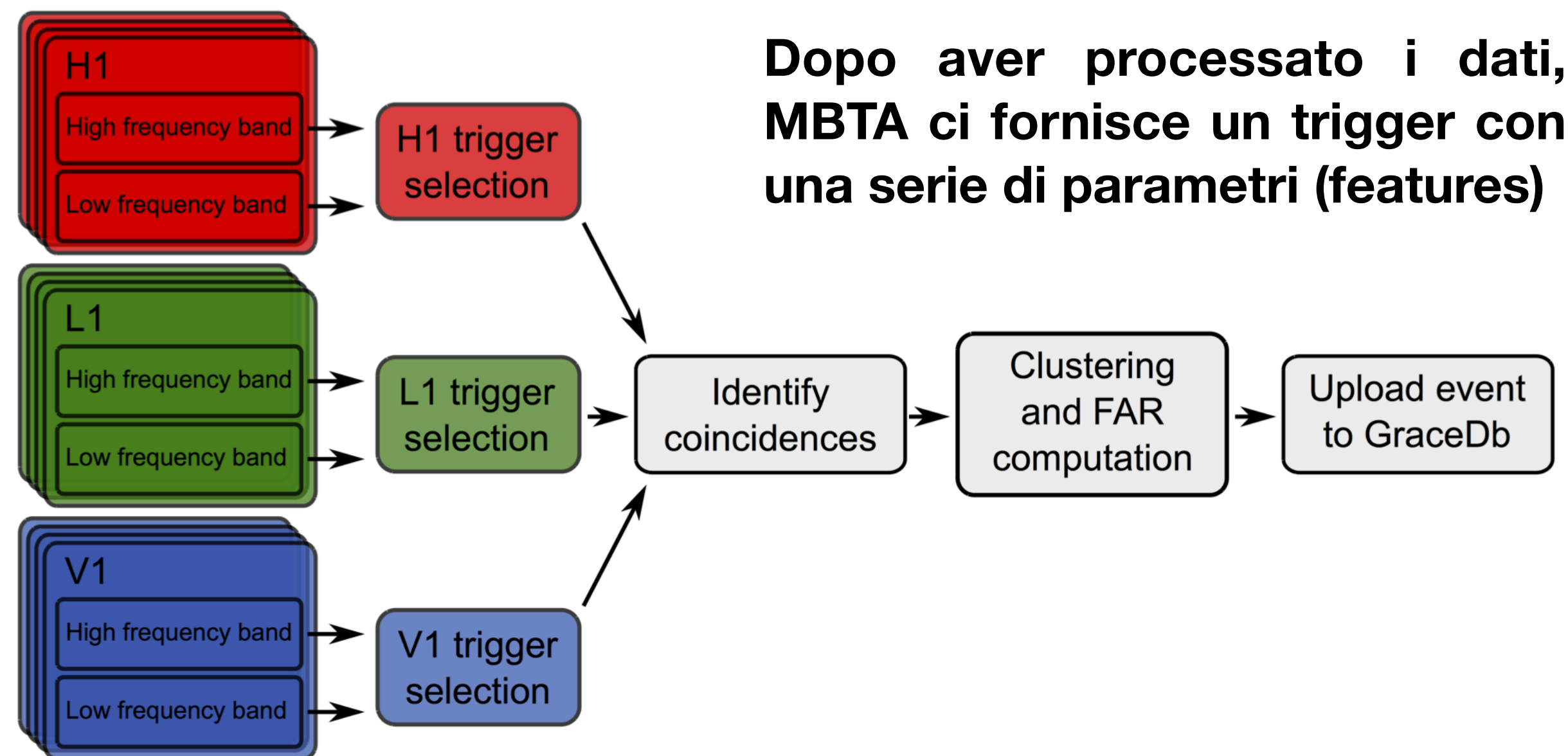
- Pipeline che utilizza la tecnica di Matched Filtering per la misurazione di Compact Binary Coalescence
- Match Filtering è effettuato utilizzando diverse bande di frequenza
- Attualmente partecipa alla campagna di acquisizione dati O4 nella collaborazione Ligo-Virgo-Kagra

ARTICOLI RILEVANTI:

<https://doi.org/10.1088/1361-6382/abe913>

<https://doi.org/10.1088/0264-9381/33/17/175012>

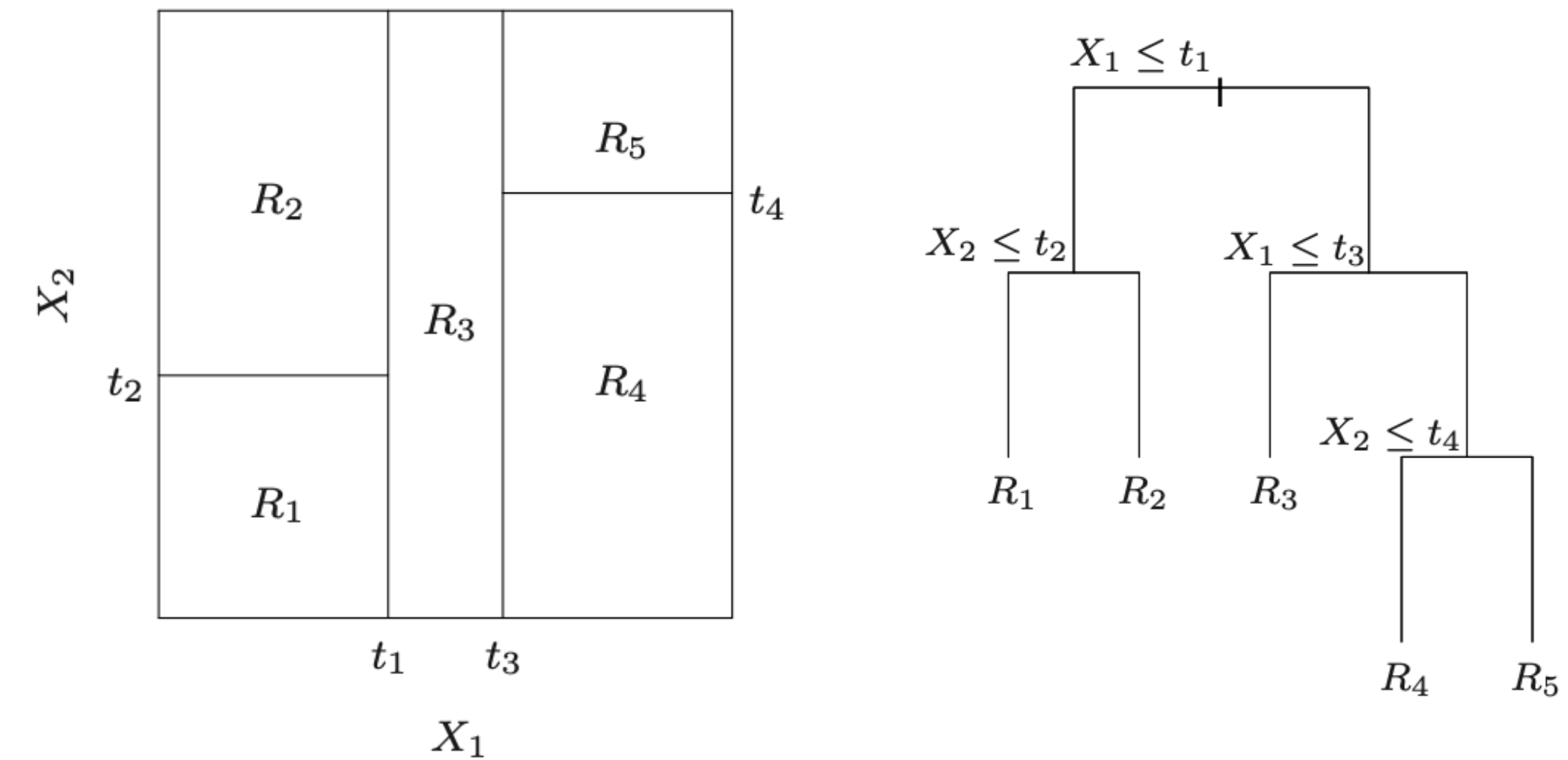
<https://doi.org/10.1088/1361-6382/ac482a>



Overview della pipeline MBTA

Algoritmo Random Forest

- Idea di base dei metodi ‘ad albero’: dividere lo spazio delle feature in una serie di rettangoli per avere un fit di un modello semplice in ognuno (tree method)
- Unire molti “deboli” classificatori (alberi) per ottenerne un più potente finale “committee”
- Random Forest: ridurre la varianza riducendo la correlazione tra gli alberi, ottenibile tramite selezione random delle variabili di input
- Idea: si può applicare questo strumento per incrementare la detection capability di una pipeline?
- RF per onde gravitazionali: “A classifier for gravitational-wave inspiral signals in non-ideal single-detector data” (T. Dal Canton, T. Dent, S. J. Kapadia, 2017).

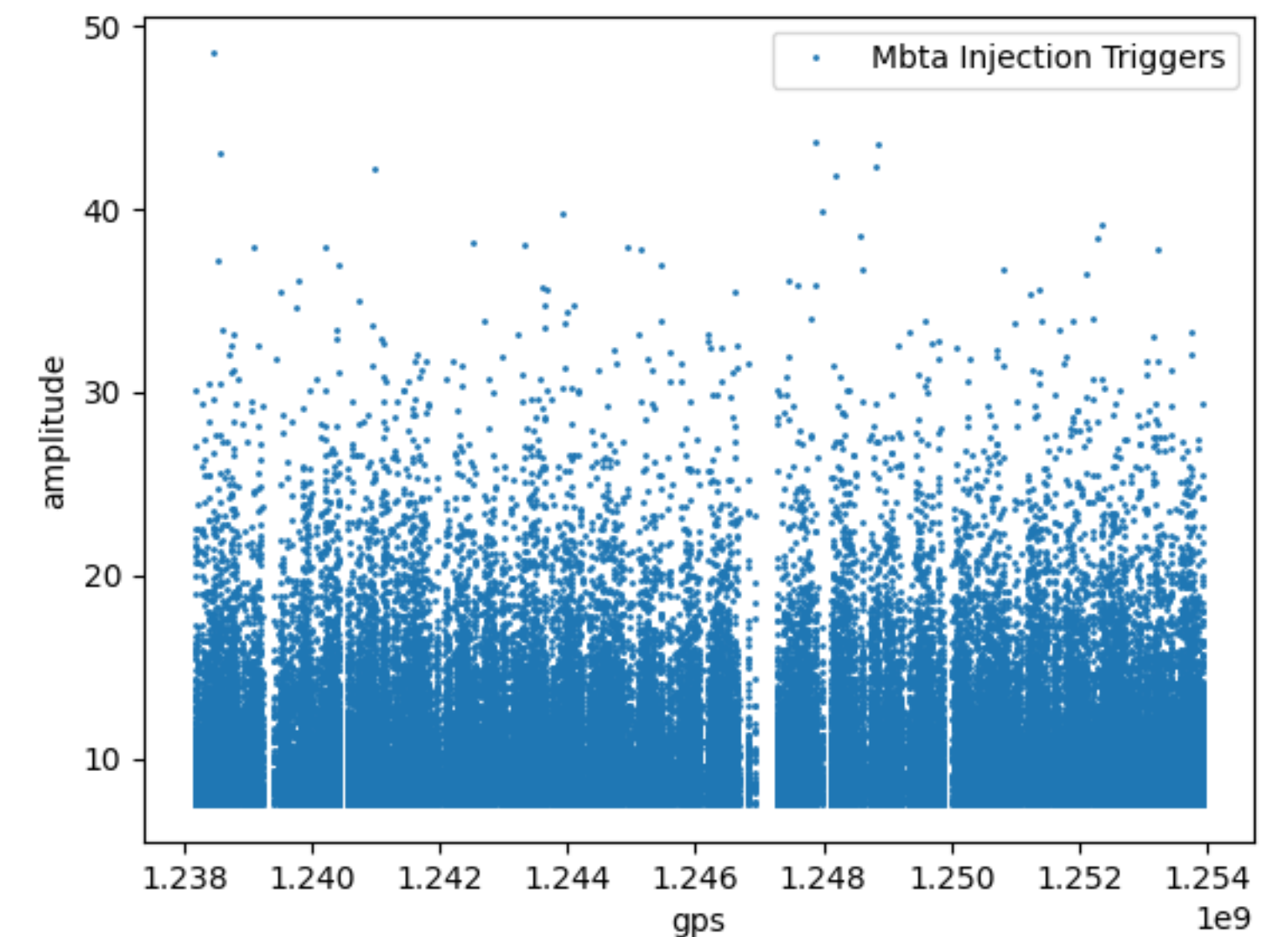
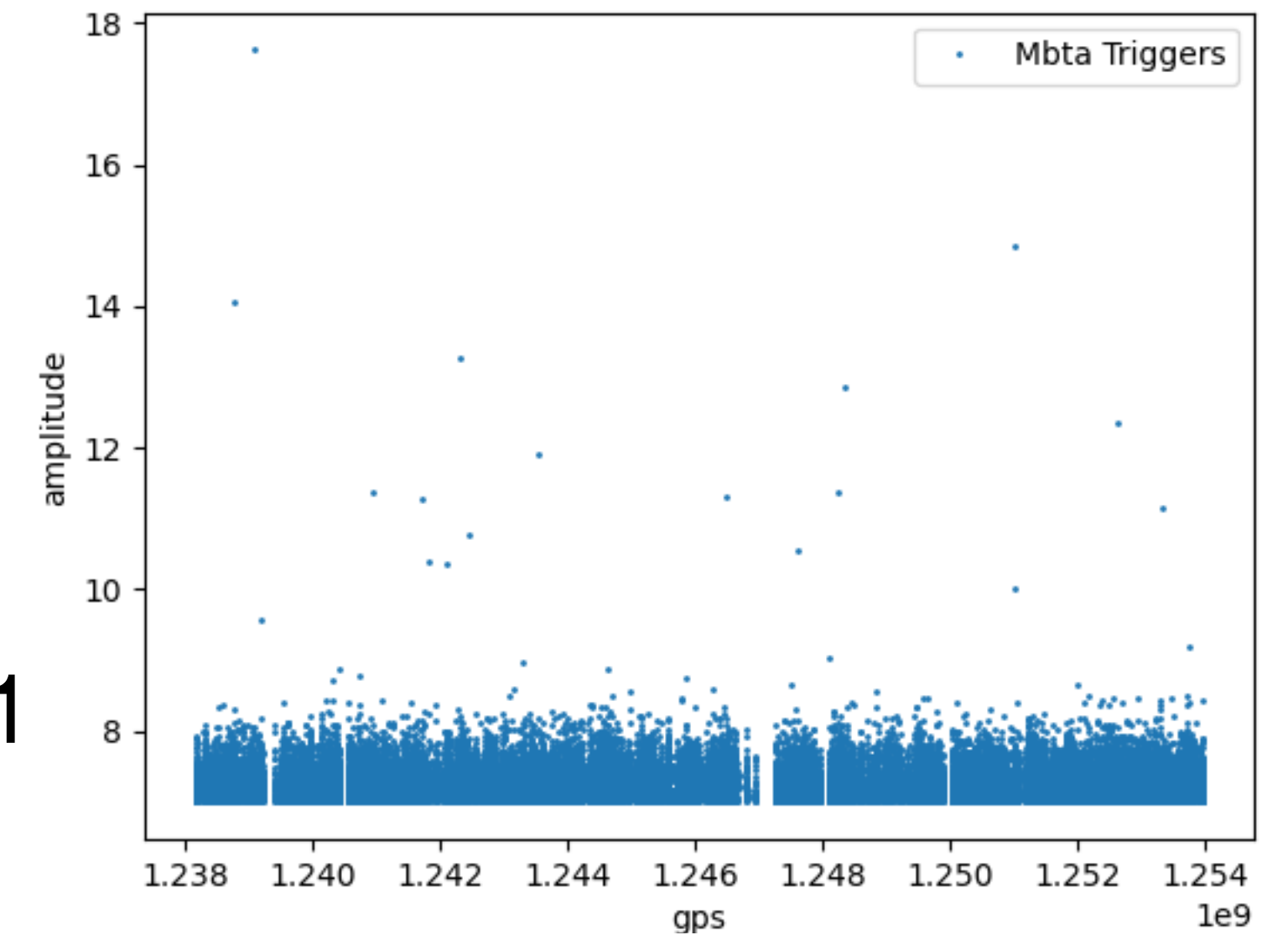


Esempio di divisione dello spazio delle feature ottenuto da un singolo albero (credits: 1)

Dataset

- O3a Data: da 1 Aprile 2019 a 1 Ottobre 2019
- Double Coincidences (HL-Von)
- 83532 Injection (injection: trigger più 'rumoroso' nell'intervallo [-80/+40] ms intorno al tempo della injection [25662 BBH, 30581 BNS, 27285 NSBH])
- 43923 Triggers (Noise + Possibili eventi)
- Training 70% and test 30%

#	Triggers	INJ
Training	30643	58576
Test	13280	24956



Esempi di Triggers (Sopra) and Triggers+ Injection (Bottom)

Iperparametri e Features

- Iperparametri: #Trees: 3000, impurity criterion: 'gini', max #features: sqrt, min_samples_leaf: 12, max_depth: 24
- Features: considerate 5 feature fisiche e 11 statistiche
- Addestrato e testato un multiclassificatore utilizzando dati O3a -> output sono probabilità ps_{BBH} , ps_{BNS} , ps_{NSBH} e $ps = ps_{BBH} + ps_{BNS} + ps_{NSBH}$ (Coincidence Triggers) [pNoise = 1 - ps].
- Confronto tra ps ottenute dal RF e rwsnr (significanza statistica ottenuta dalla pipeline) tramite False Alarm Probability (α_s) and Number of Detections (N_d)

$$\bullet \alpha_s = \frac{1}{Nn} \sum_{i=1}^{Nn} \theta(\hat{p}_s^i - p_s)$$

$$\bullet N_d = \sum_{i=1}^{Ns} \theta(\hat{p}_s^i - p_s)$$

Features

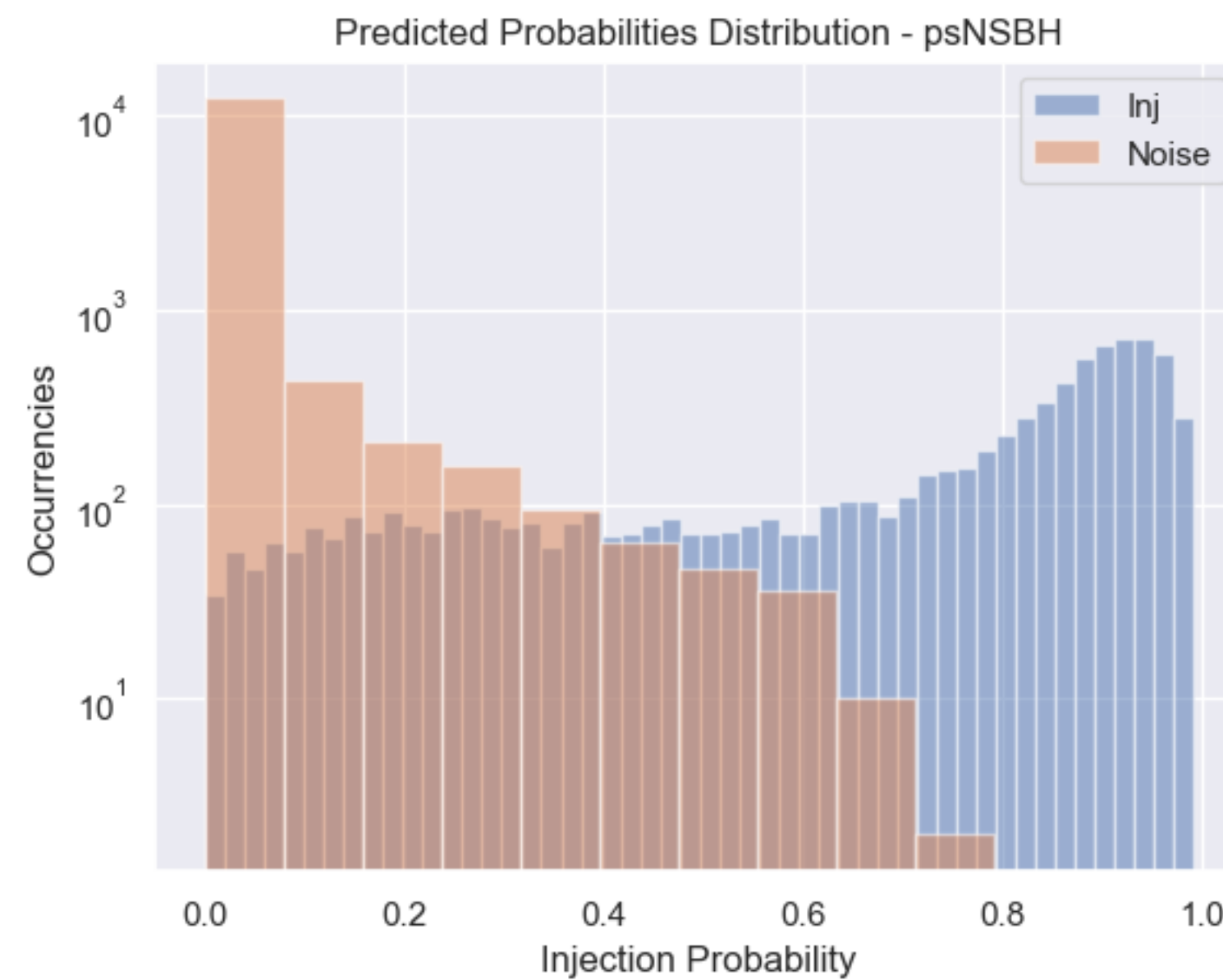
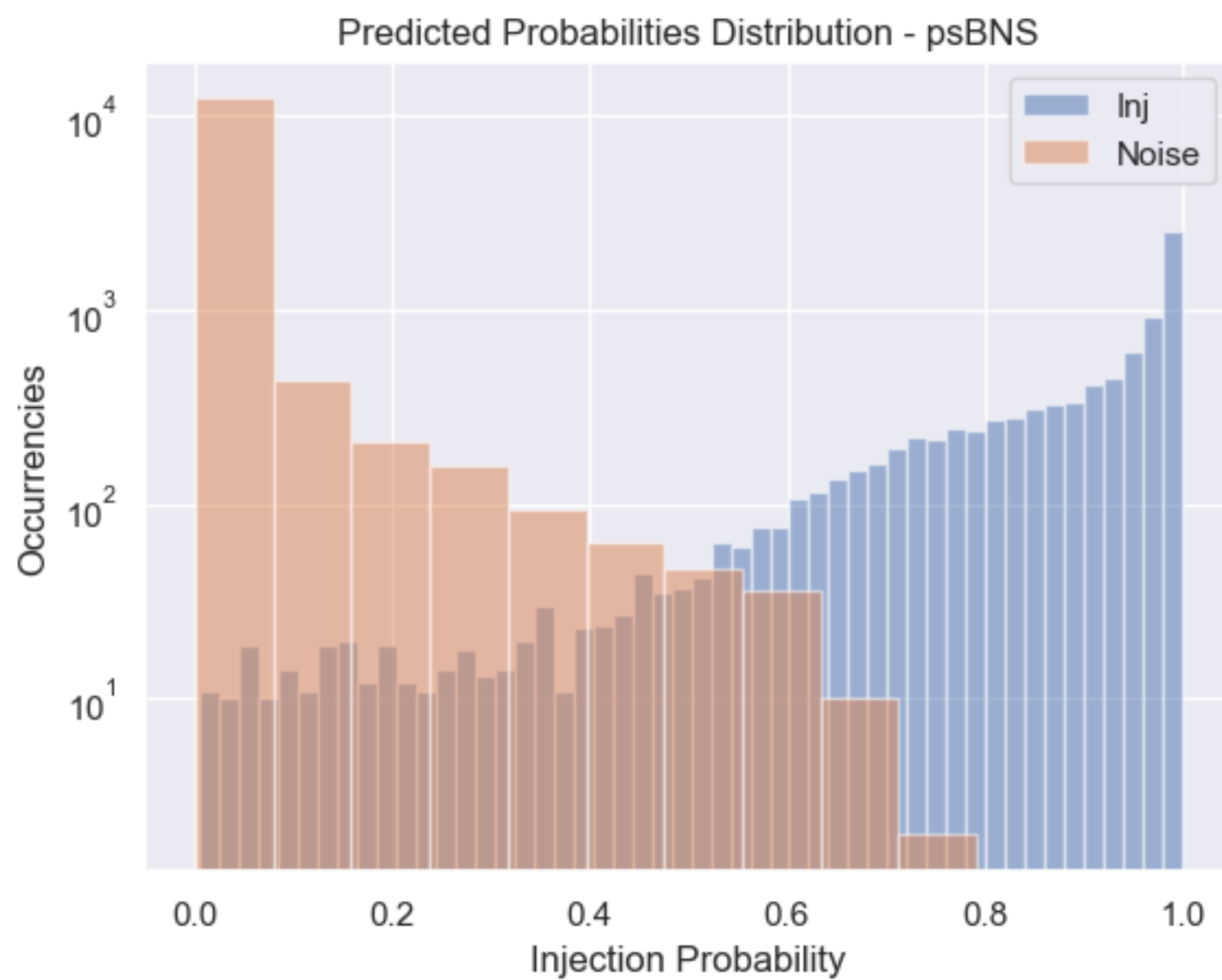
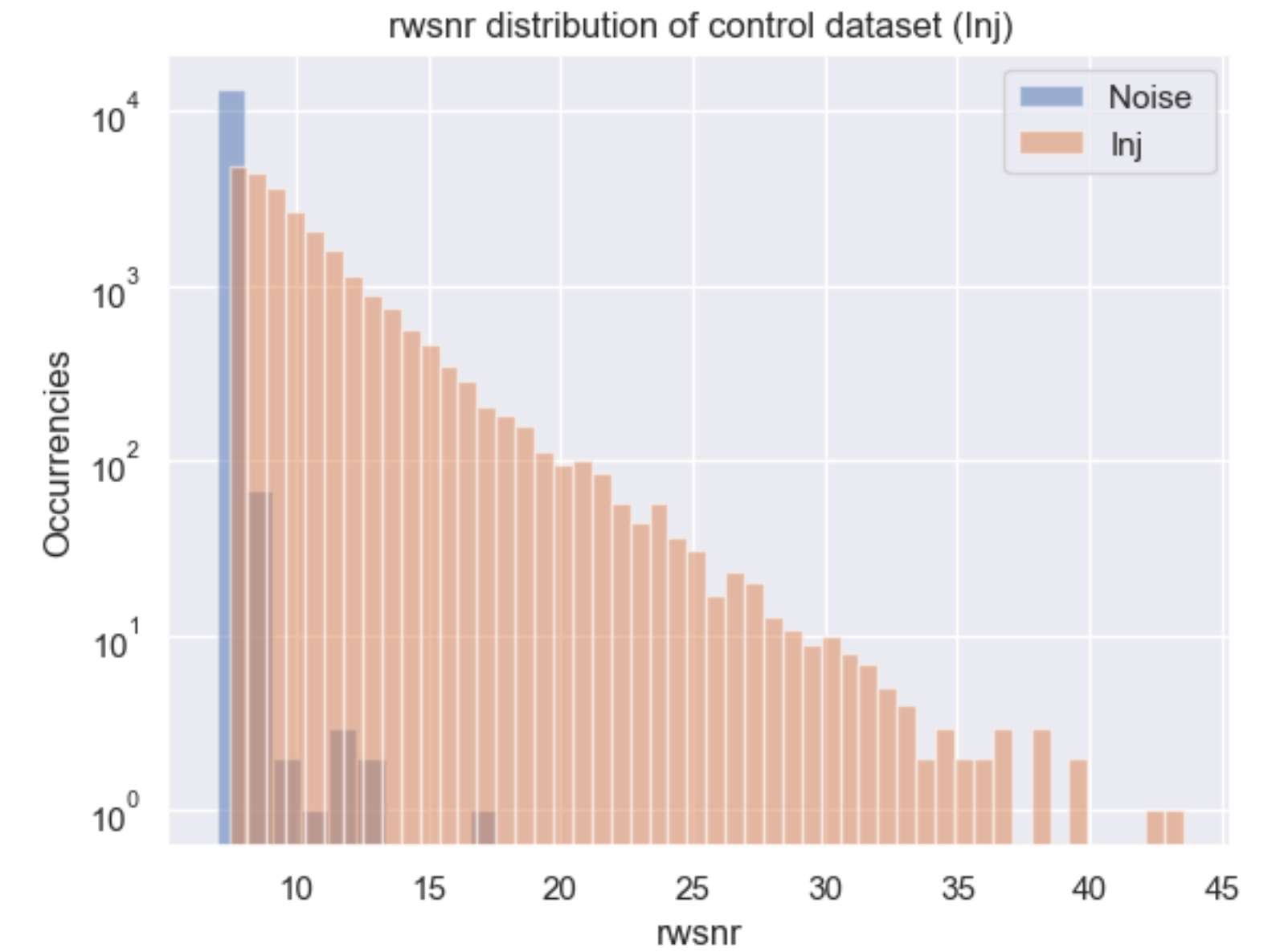
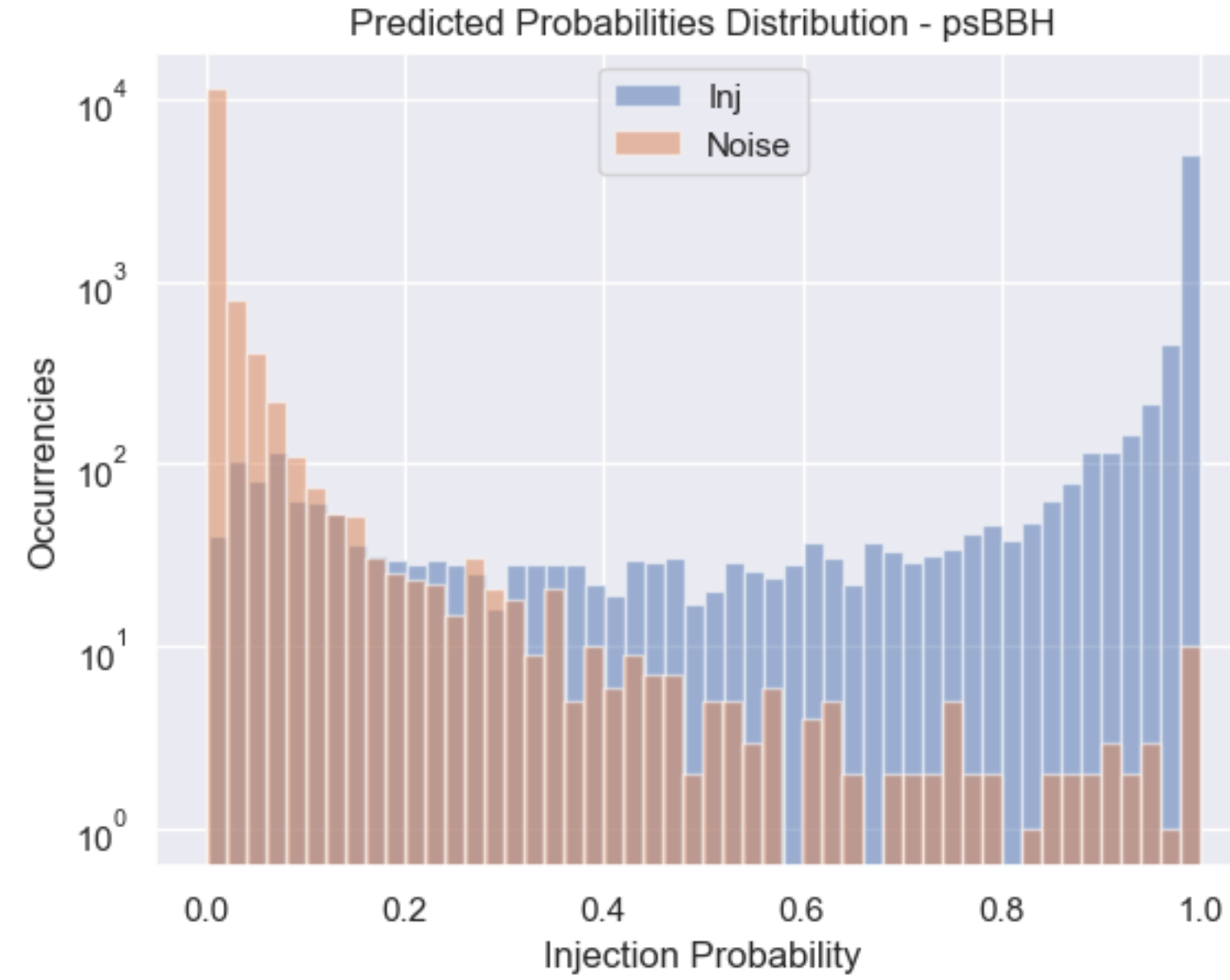
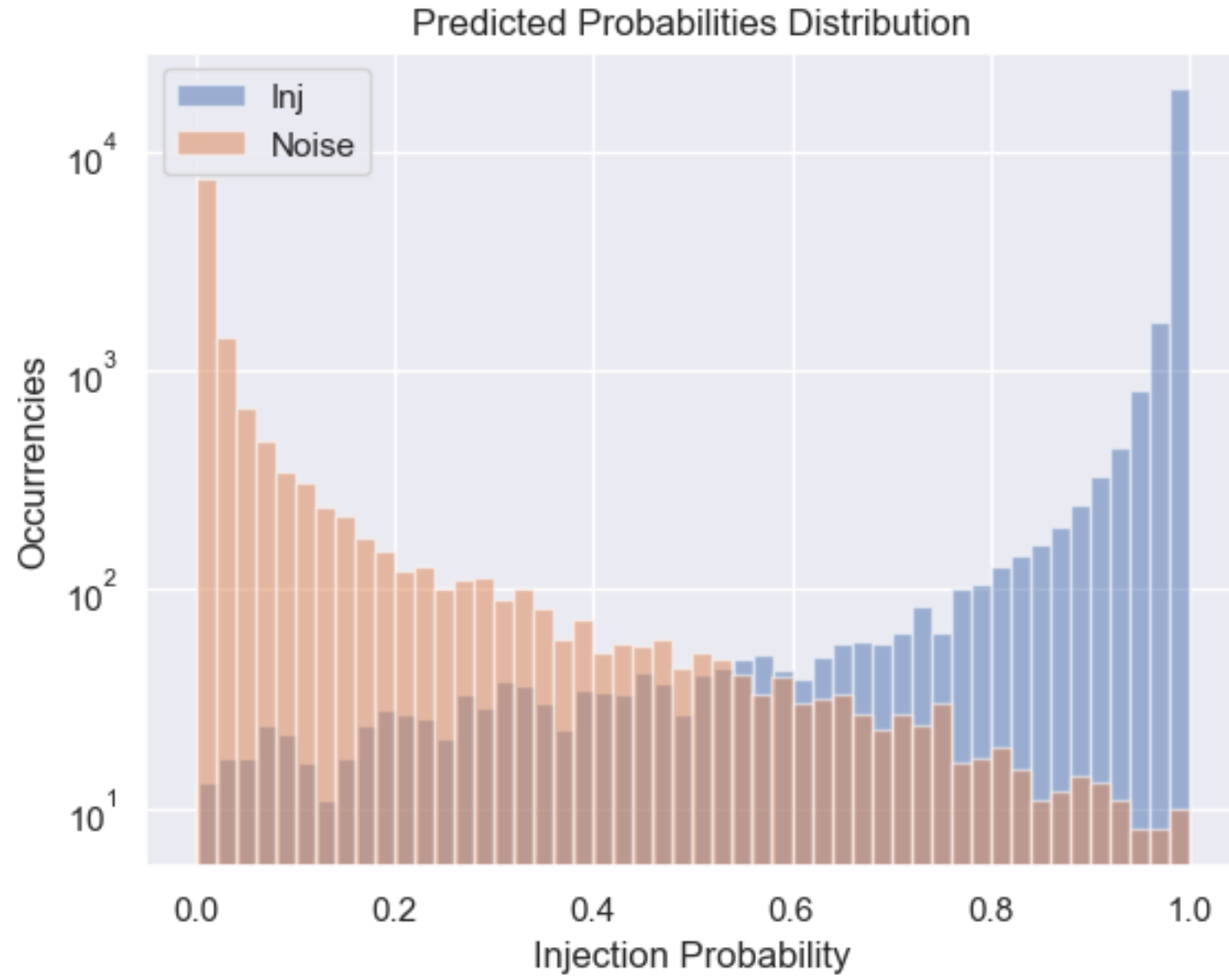
Fisiche

m1, m2, s1z, s2z, t_dur

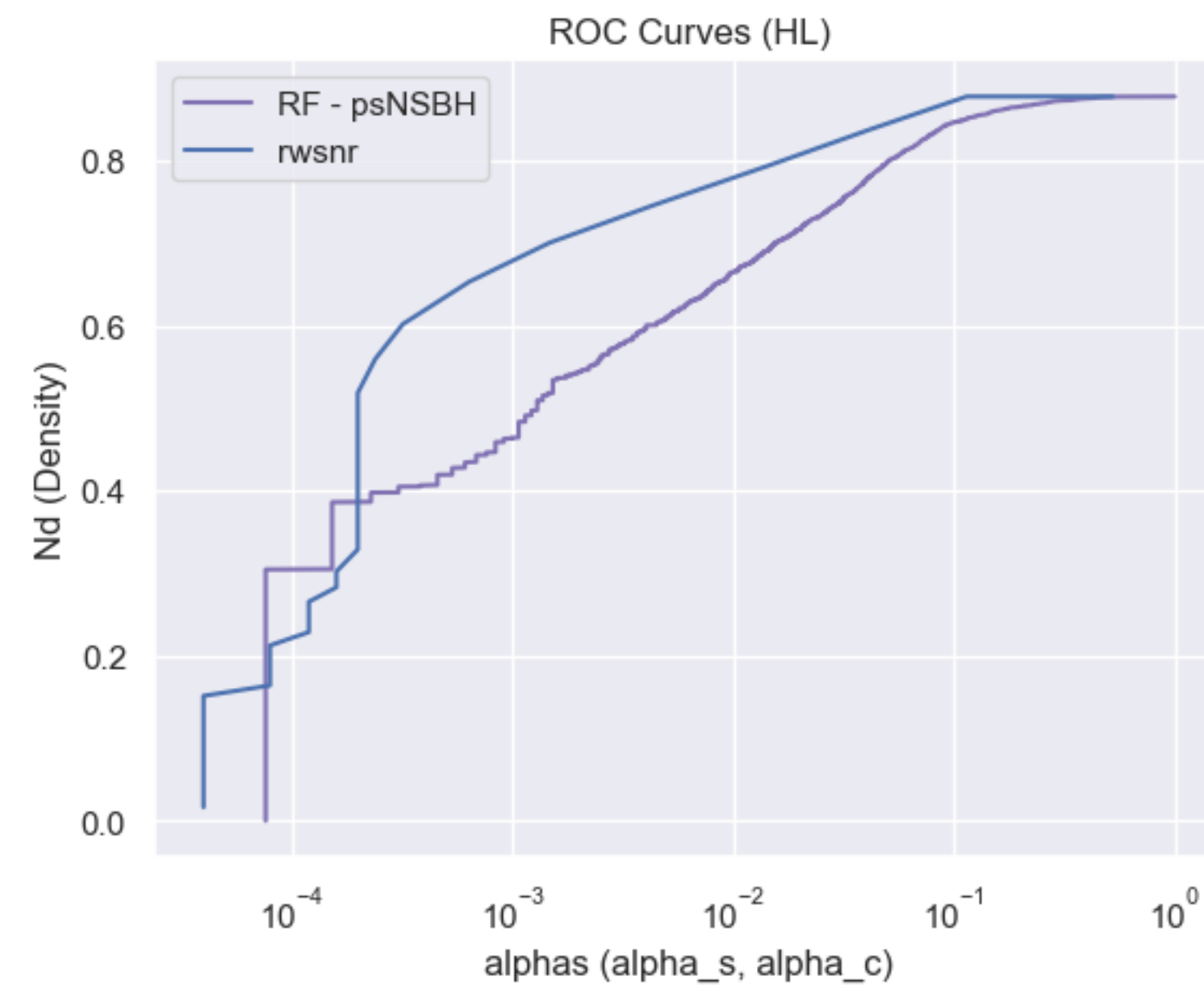
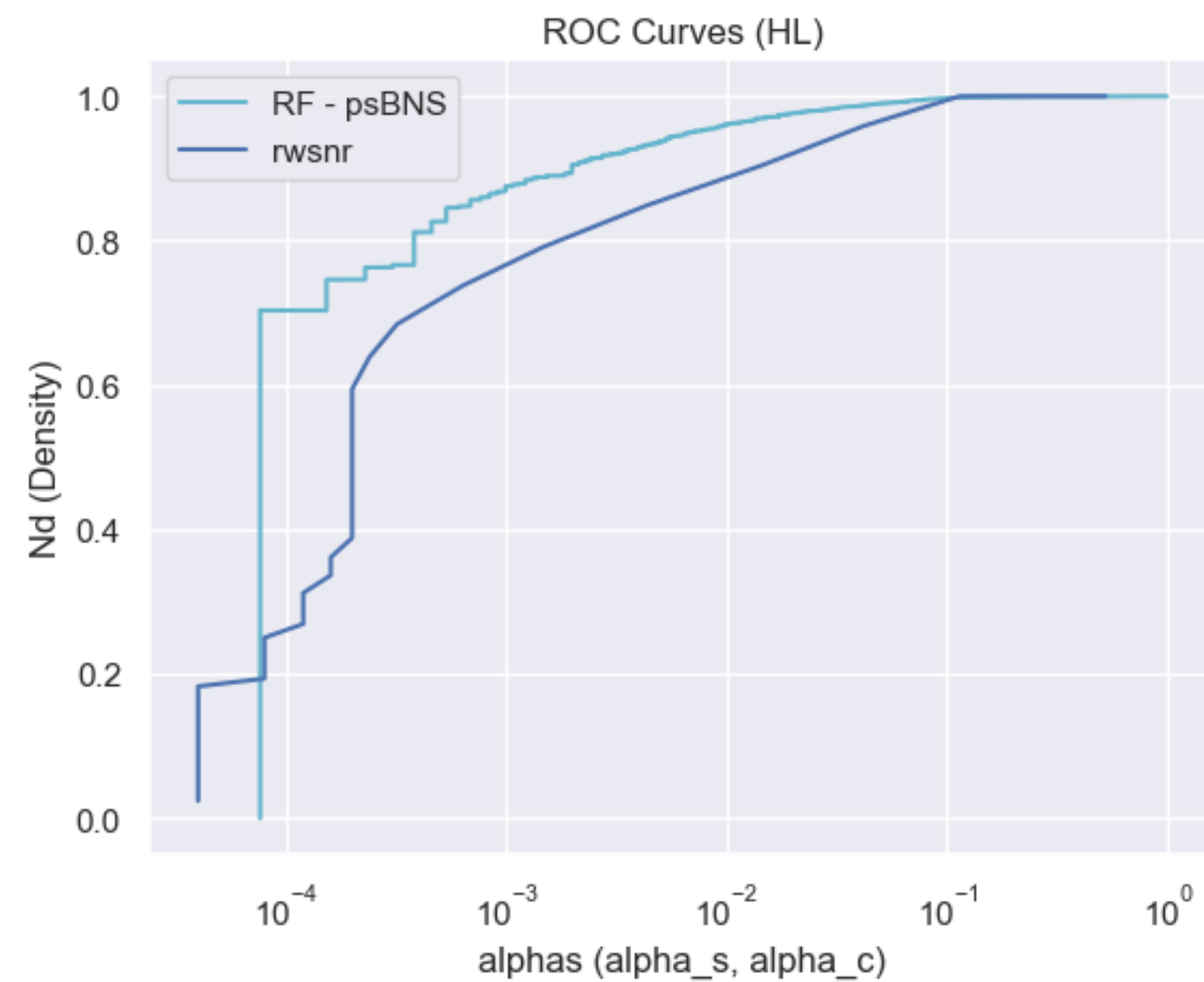
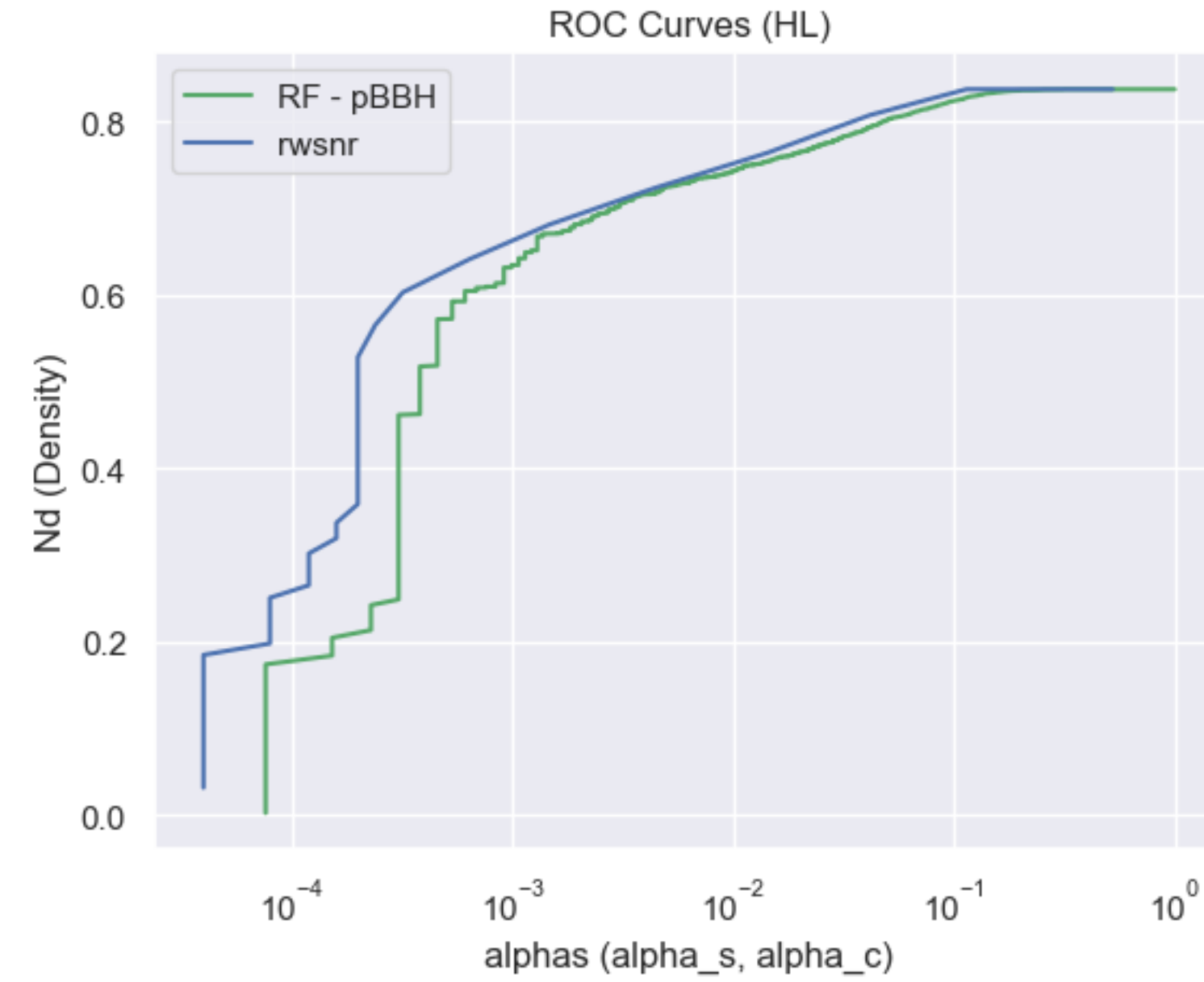
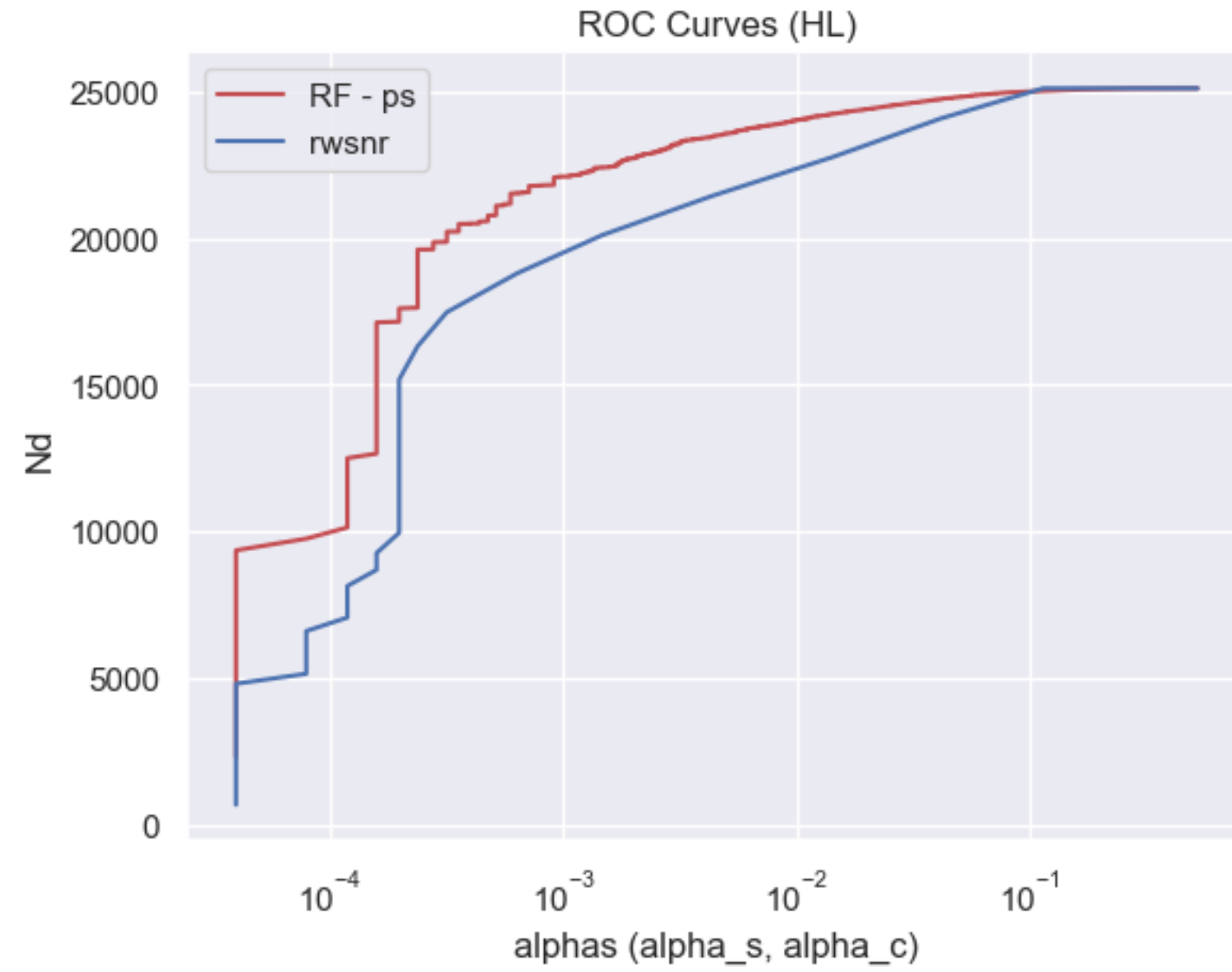
Statistiche

H_snr, L_snr, chi^2_H, chi^2_L, nEvents, L_ER, H_ER, L_phase, H_phase, L_effDist, H_effDist

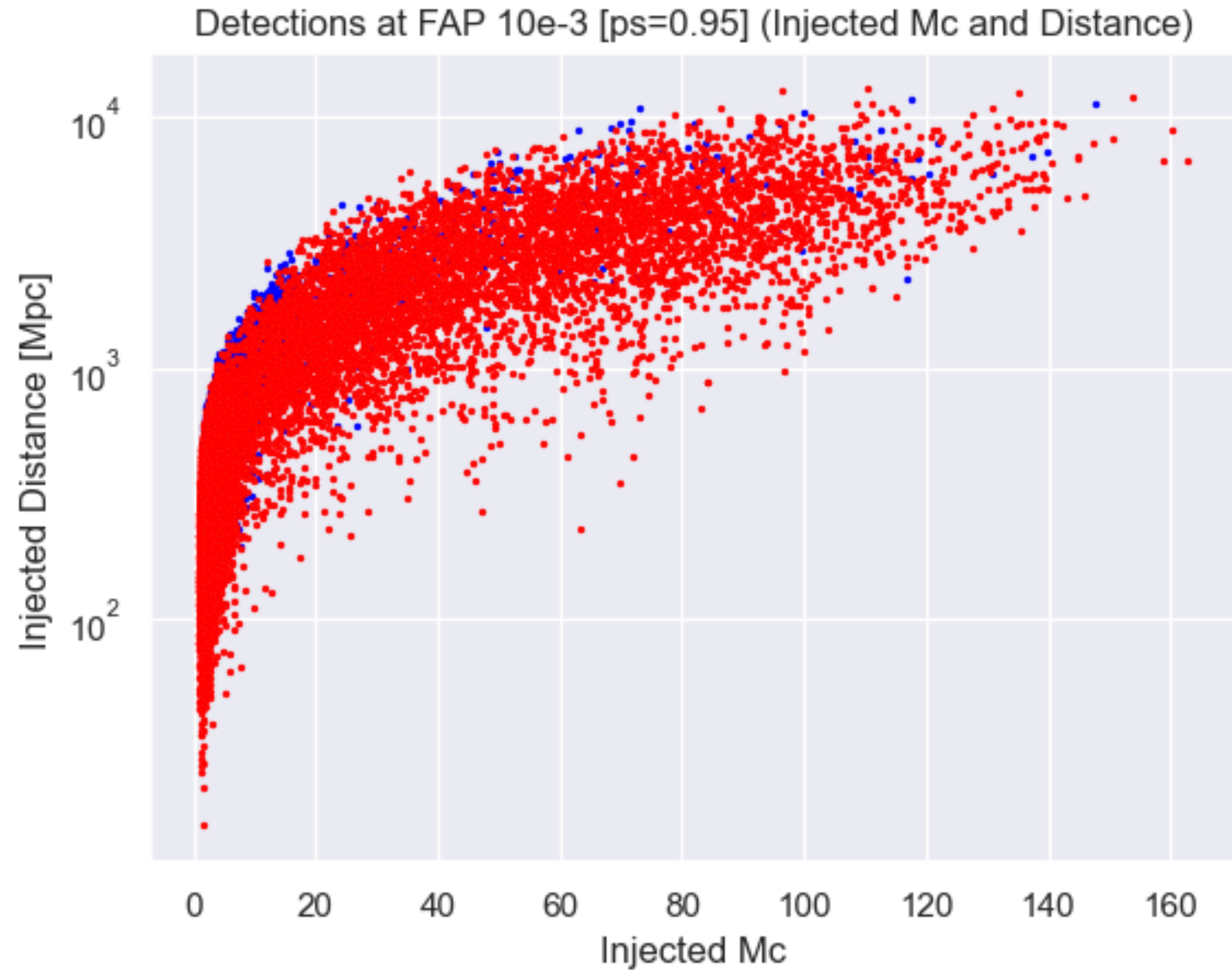
Risultati - Probabilità



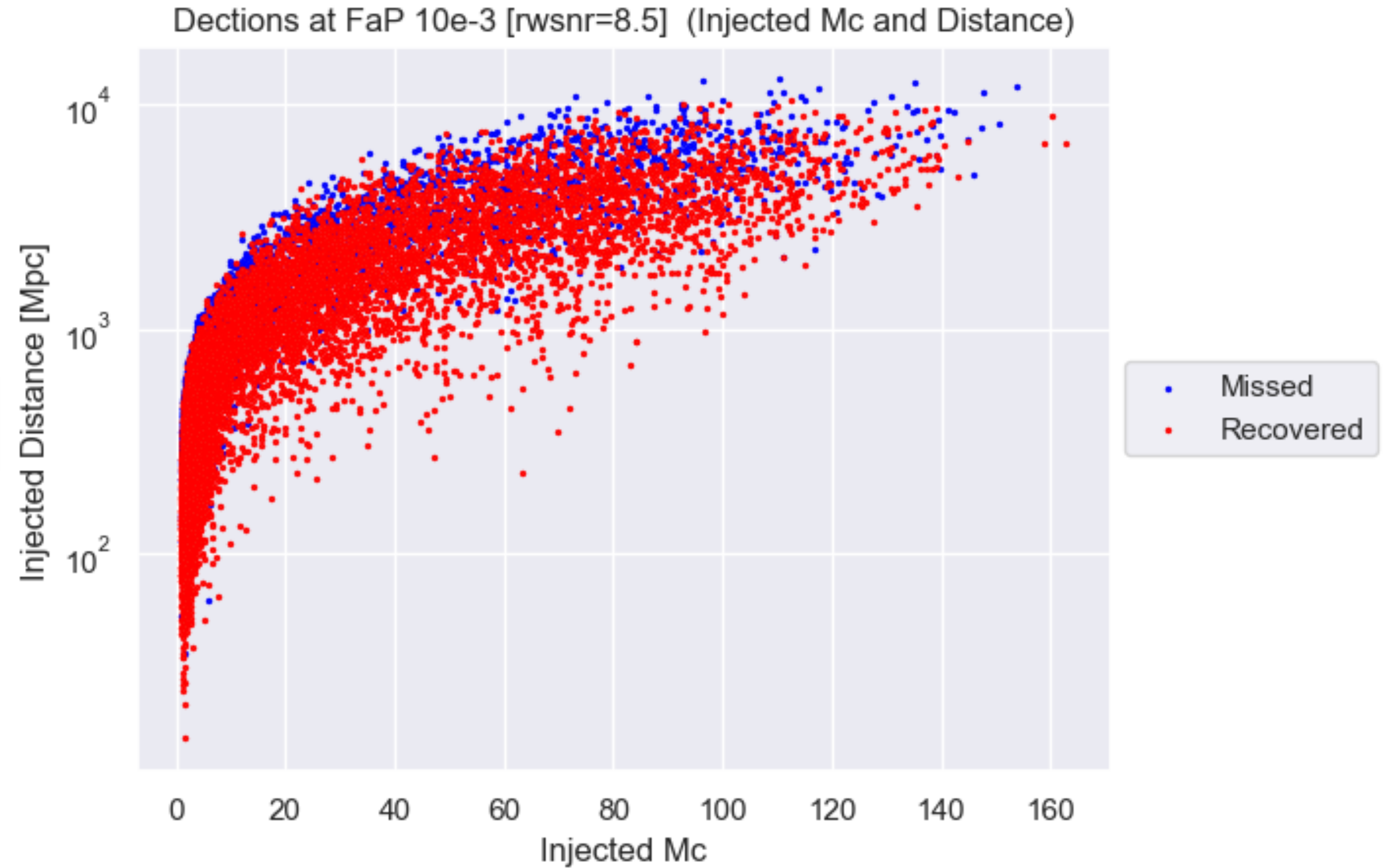
Risultati- Nd vs False Alarm Probability



Risultati- Efficienza nella Detection



• Missed
• Recovered

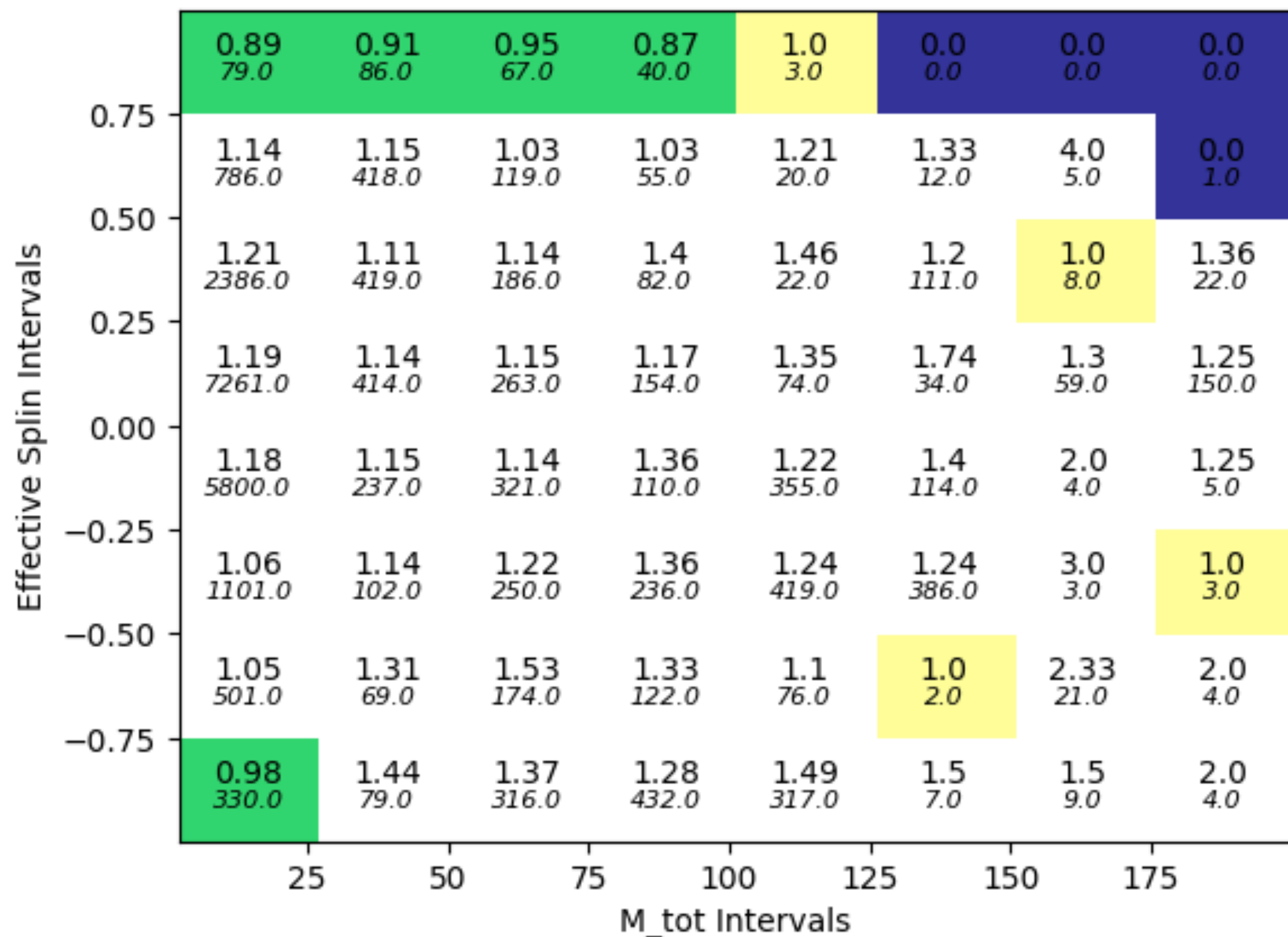


• Missed
• Recovered

Detection RF 21419
Detection rwSNR Cut 18062

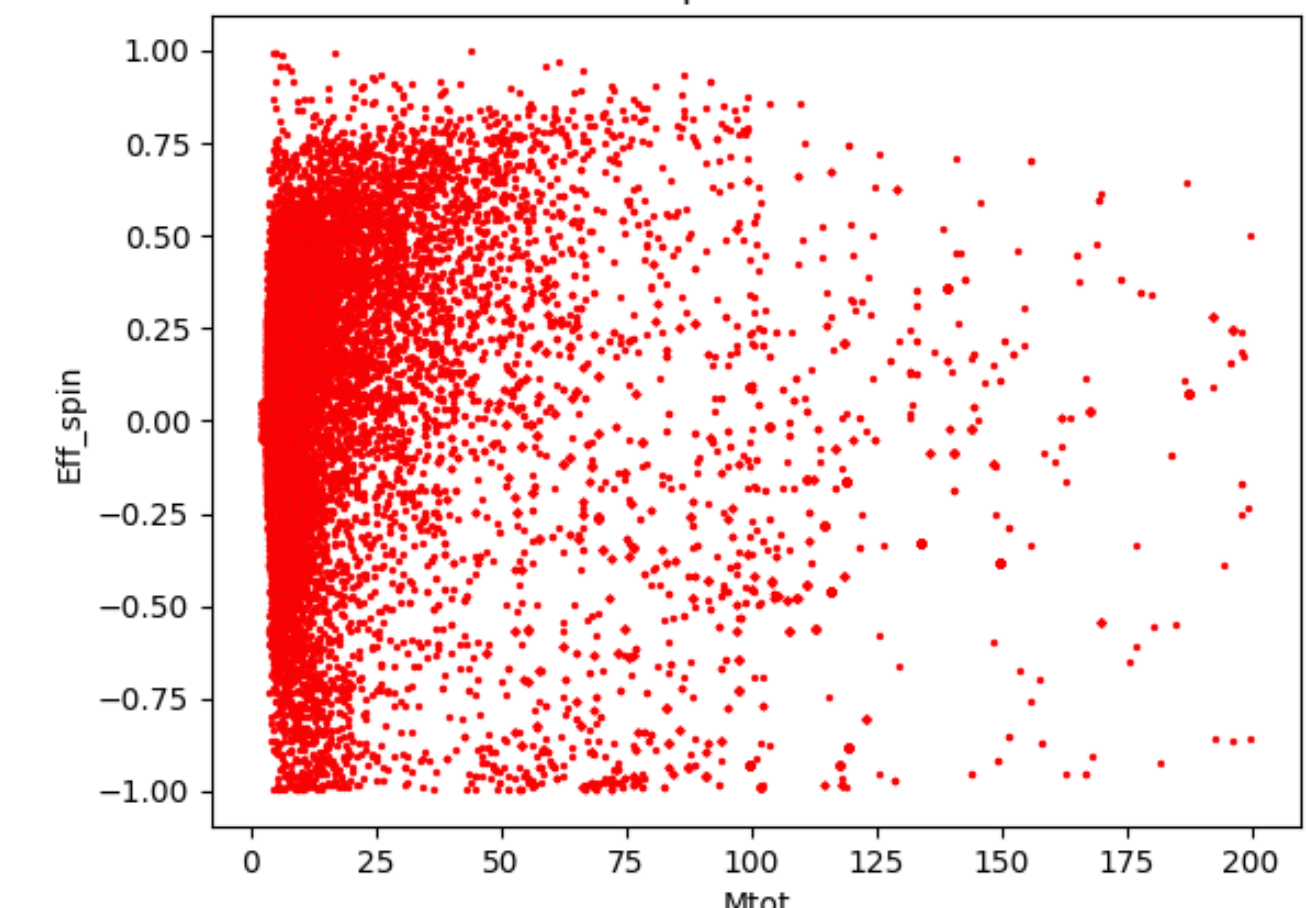
Risultati - Efficienza nella Detection - Parametri Fisici

Ratio between #events-RF and #events-rwsnr

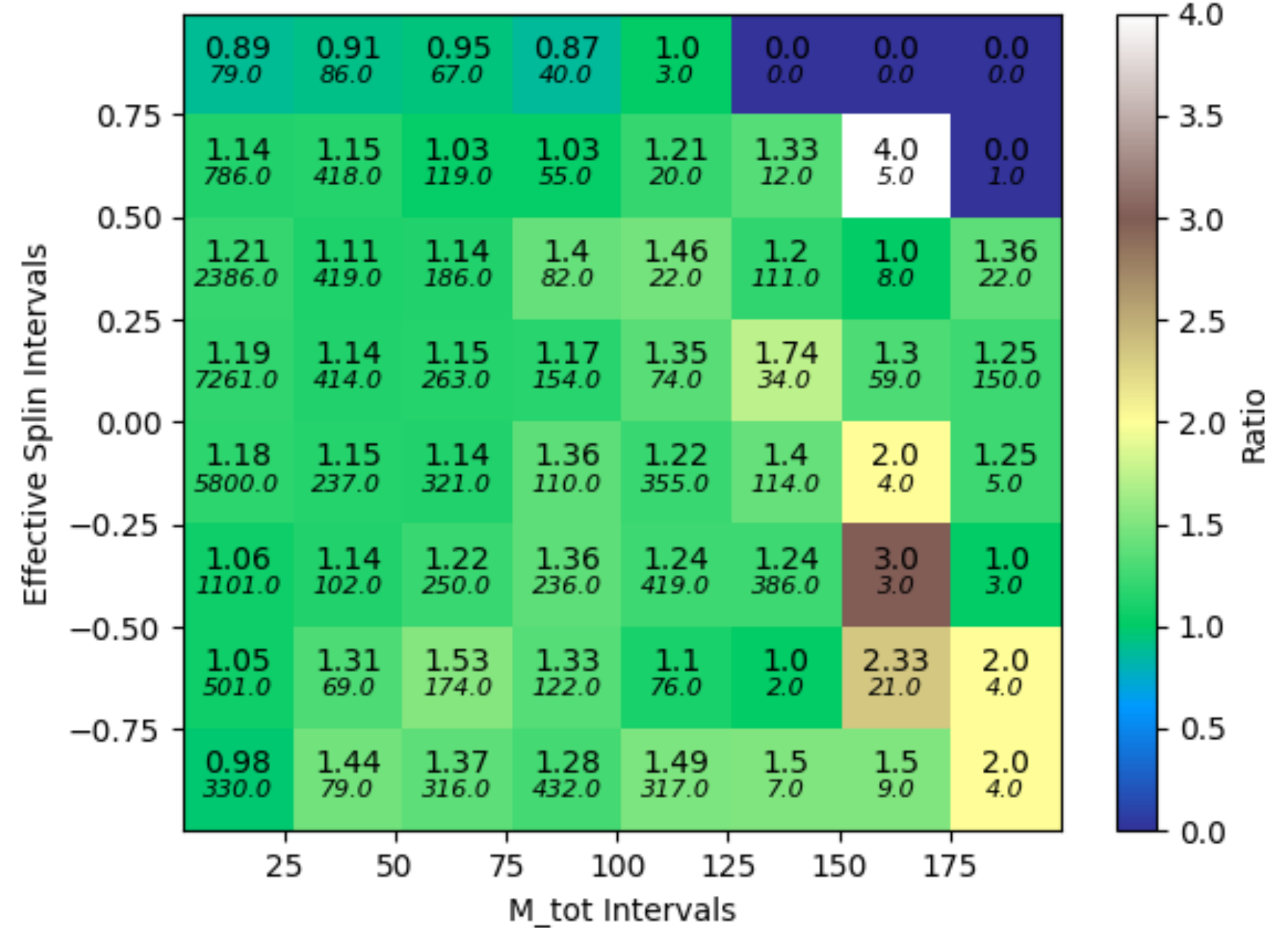


- #(*rwsnr*) = 0 and #(*ps-RF*) = 0
- ratio < 1
- Ratio = 1
- Ratio > 1

Eff spin and Mtot



between #events-RF and #events-rwsnr



In ogni 'box' (piccolo): $\frac{\#RF}{\#Control} \Big|_{\bar{if}\bar{ar}}$, $F\bar{a}P = 1 \times 10^{-3}$, sotto $\#Inj$

Conclusioni e prospettive

- Il Random Forest Algorithm mostra una migliore efficienza nel separare segnale da rumore
- Risultati preliminari della ricerca - ulteriori studi sulla ps sono richiesti
- Ottimizzare la classificazione (nuove features -architettura differente)
- Sviluppare uno strumento per l'analisi online

Grazie!
Domande?

Backup

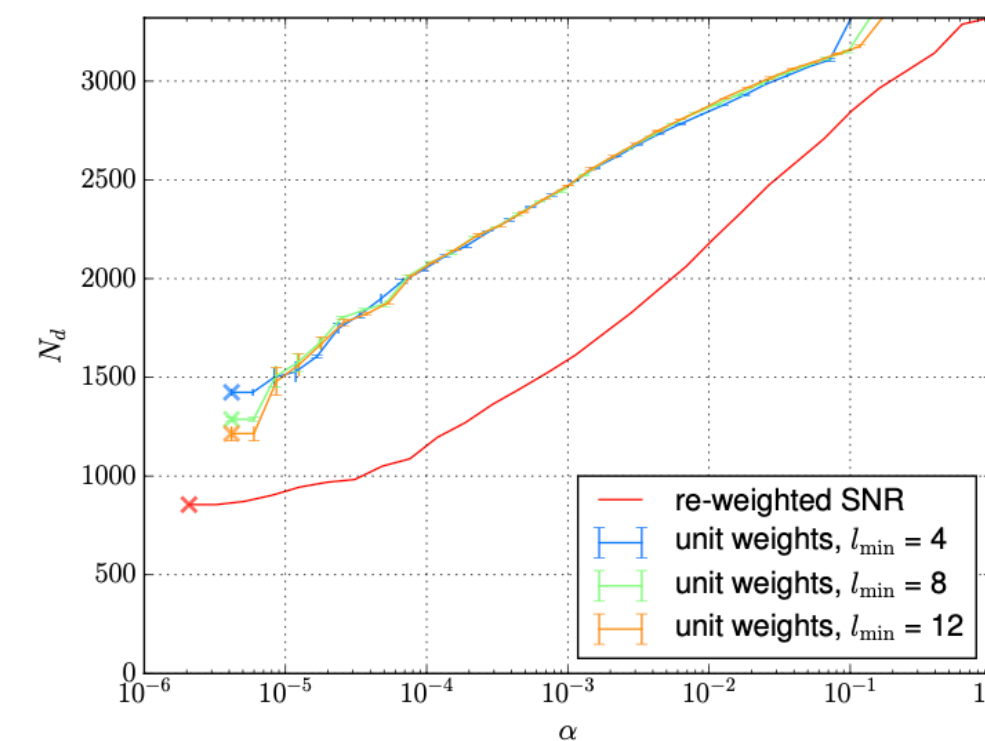
Random Forest for Gravitational Waves Search sposta in backup citi nella slide prima e metti questa in backup

- Articolo: “A classifier for gravitational-wave inspiral signals in non-ideal single-detector data” (T. Dal Canton, T. Dent, S. J. Kapadia, 2017).
- Considerato un set di GW Triggers (noise, injections, entrambi da pipeline cbc e omicron) con differenti feature per addestrare un RF.
- Il RF, dopo l’addestramento, assegna ad ogni evento una probabilità \hat{p} di essere o segnale \hat{p}_s o noise \hat{p}_n
- Definita una significanza statistica (False Alarm Probability e Number of Detections) e confrontata con la rwSNR (SNR, χ^2).

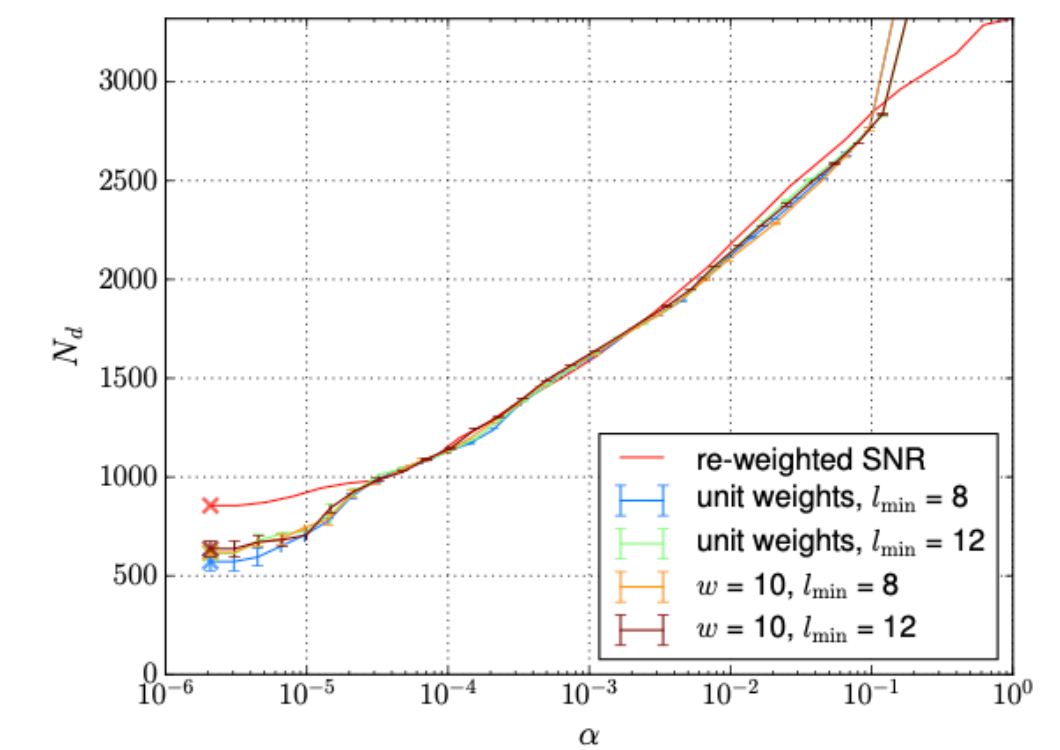
$$\alpha_s = \frac{1}{Nn} \sum_{i=1}^{Nn} \theta(\hat{p}_s^i - p_s)$$

$$N_d = \sum_{i=1}^{N_s} \theta(\hat{p}_s^i - p_s)$$

Risultati e confronto tra RF e rwSNR



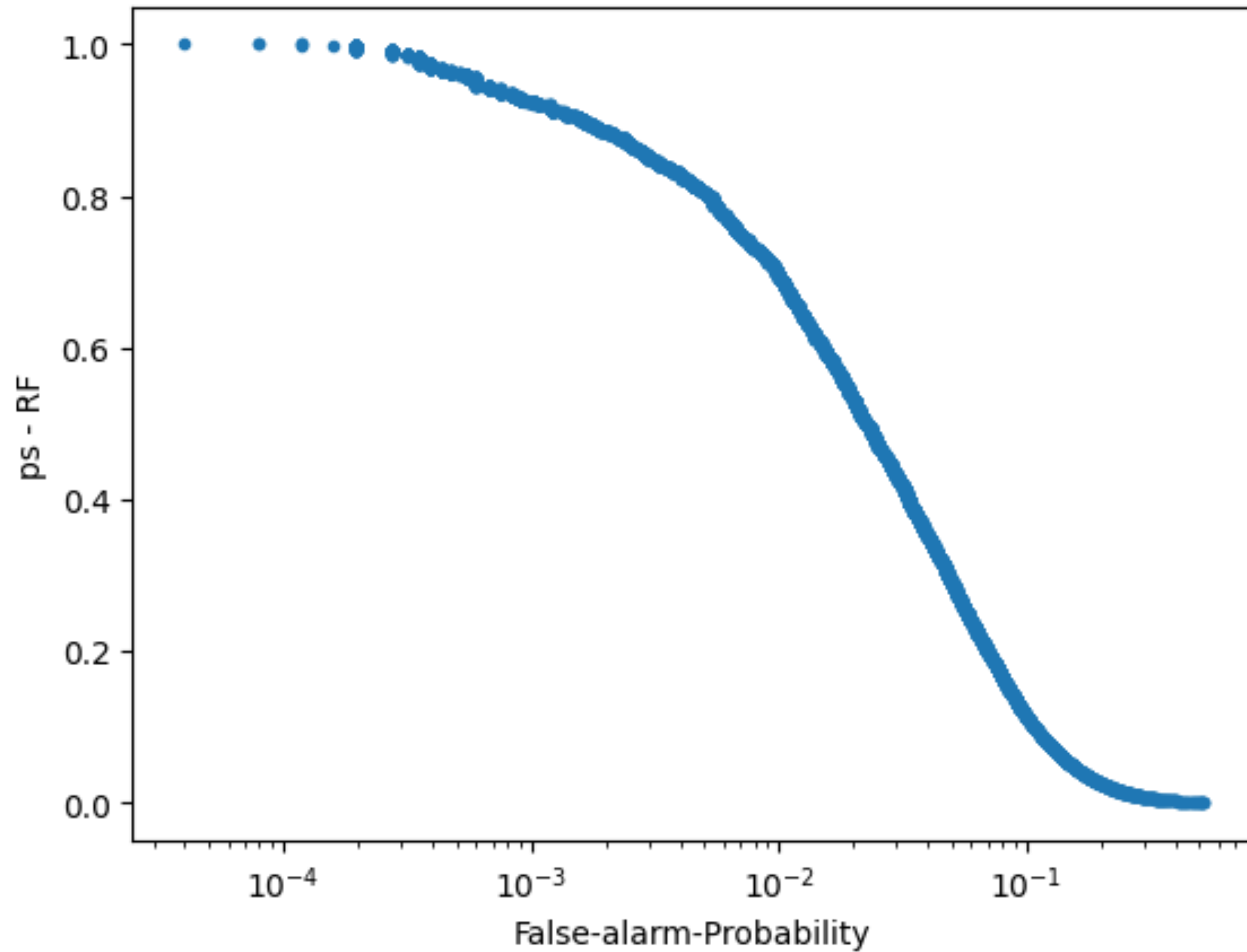
N_d as function of α , several feature considered



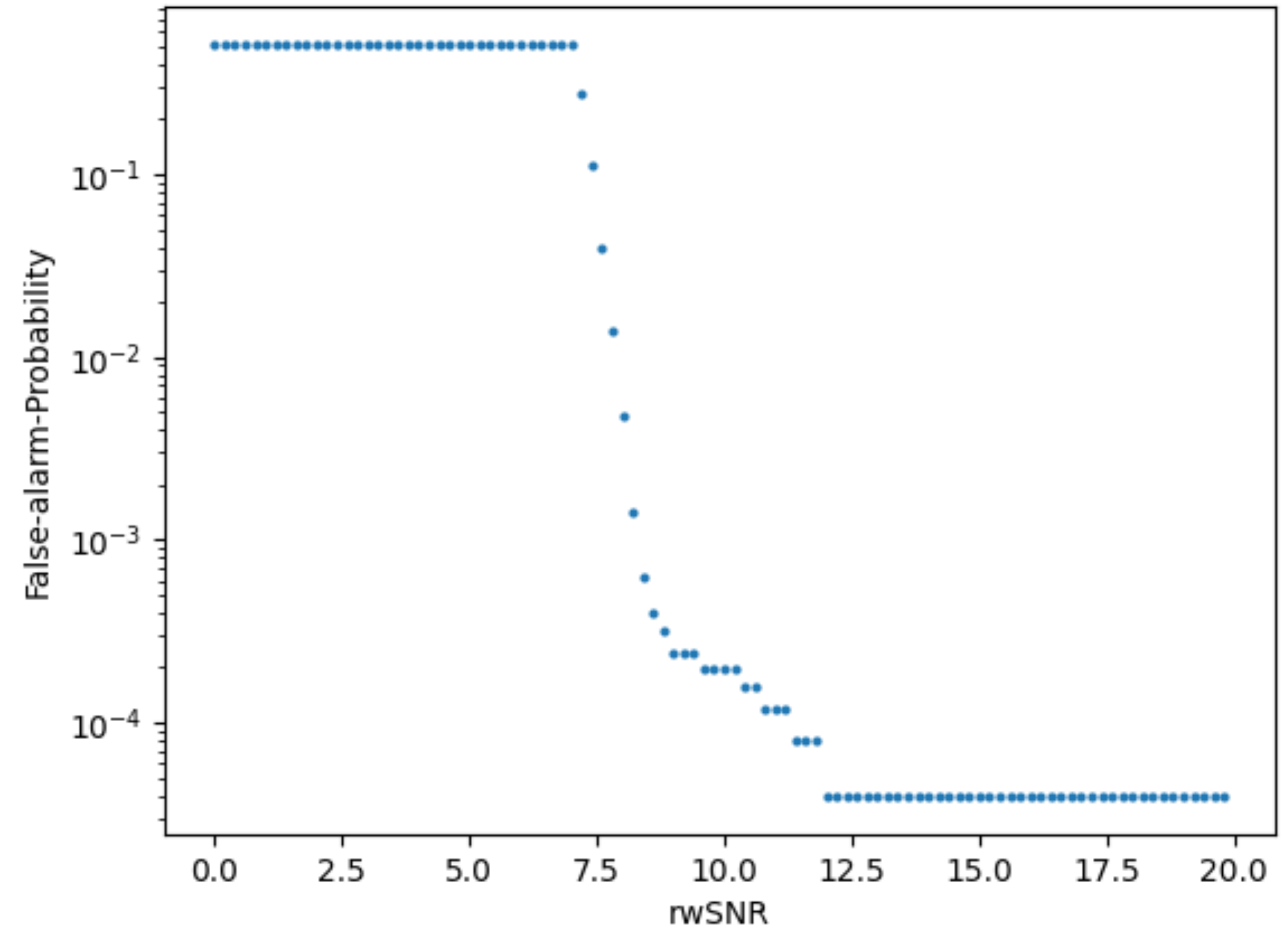
N_d as function of α , considered only snr and χ^2

Ps e rwsnr cut

ps vs False Alarm Probability

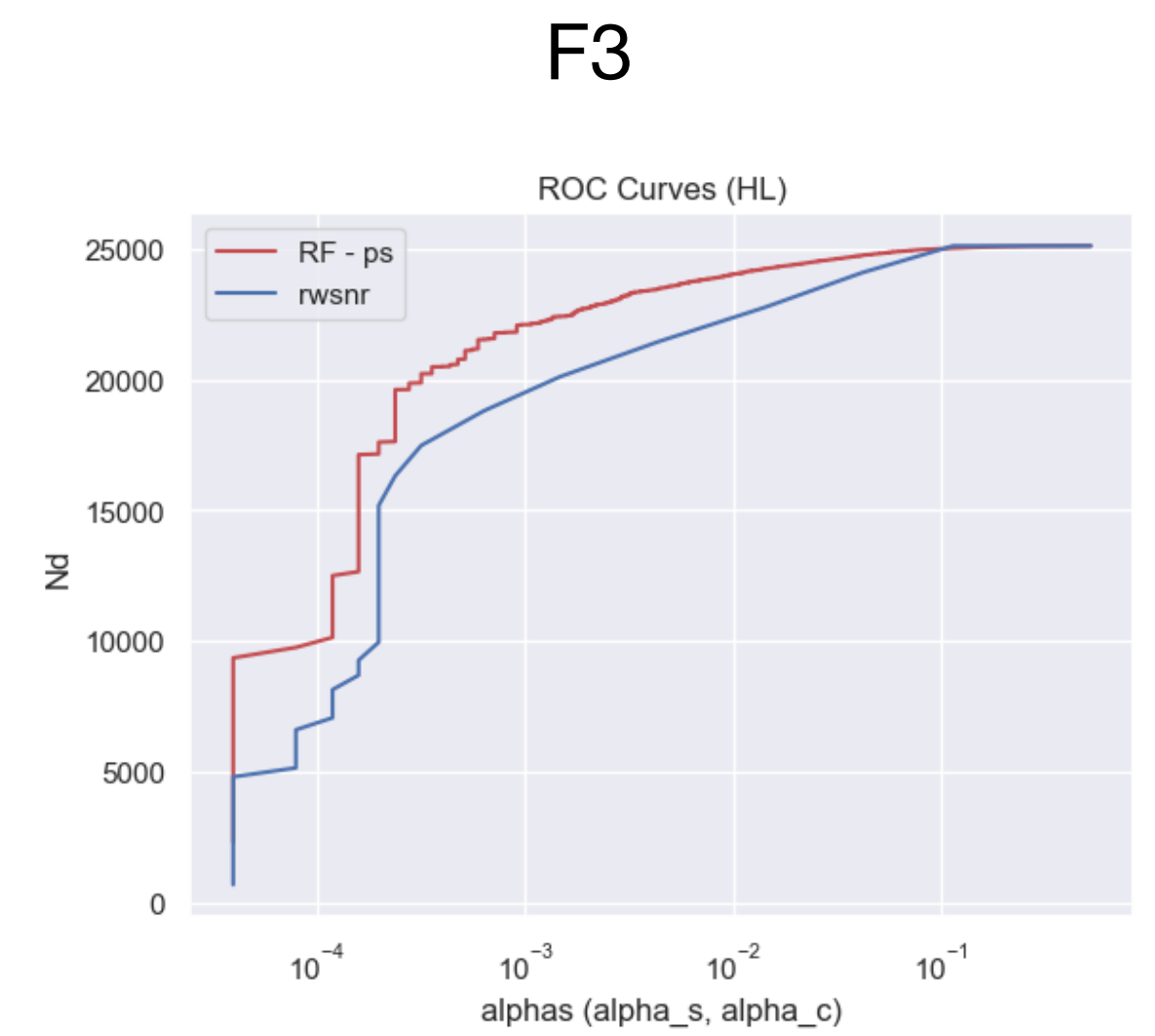
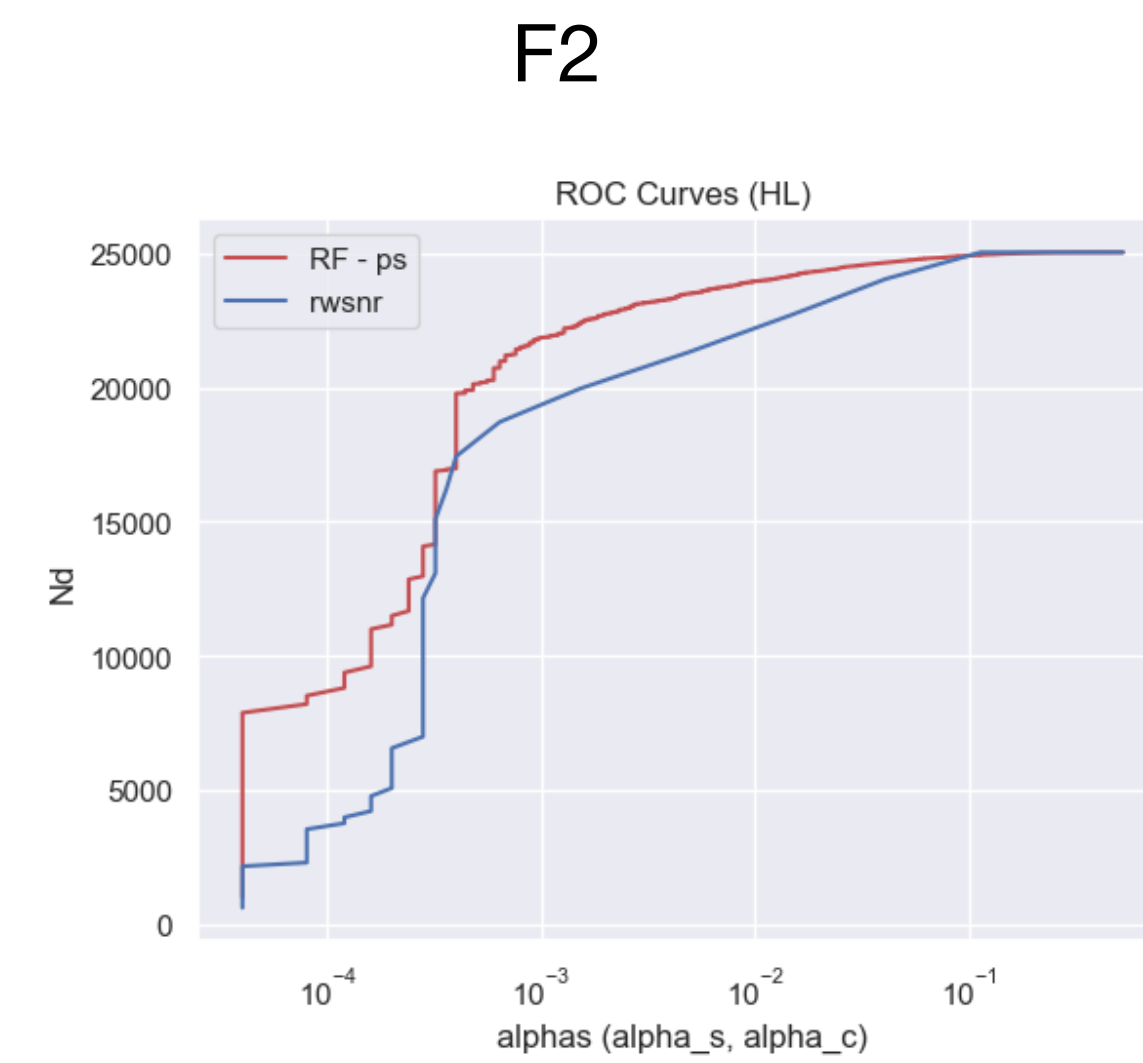
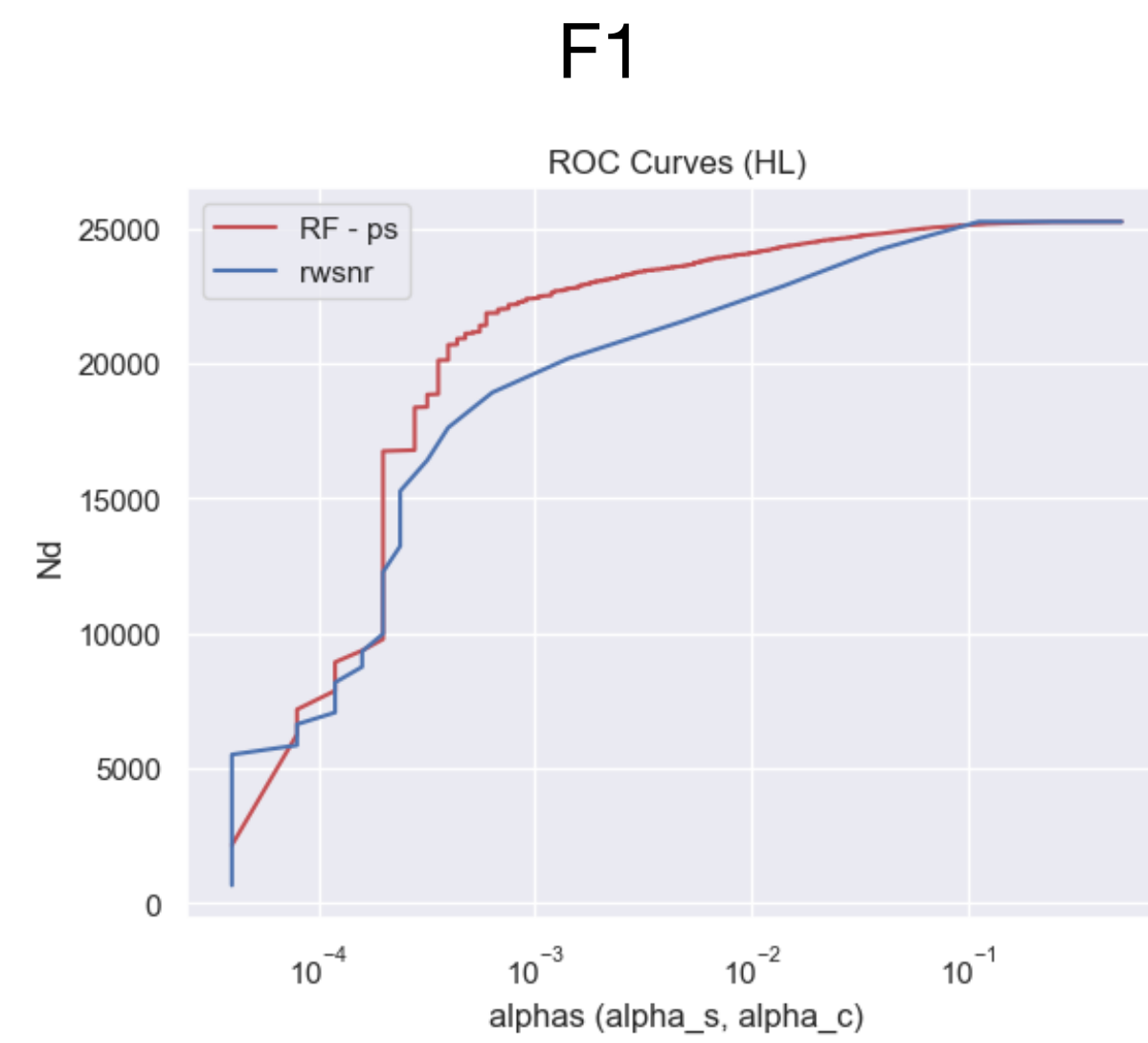
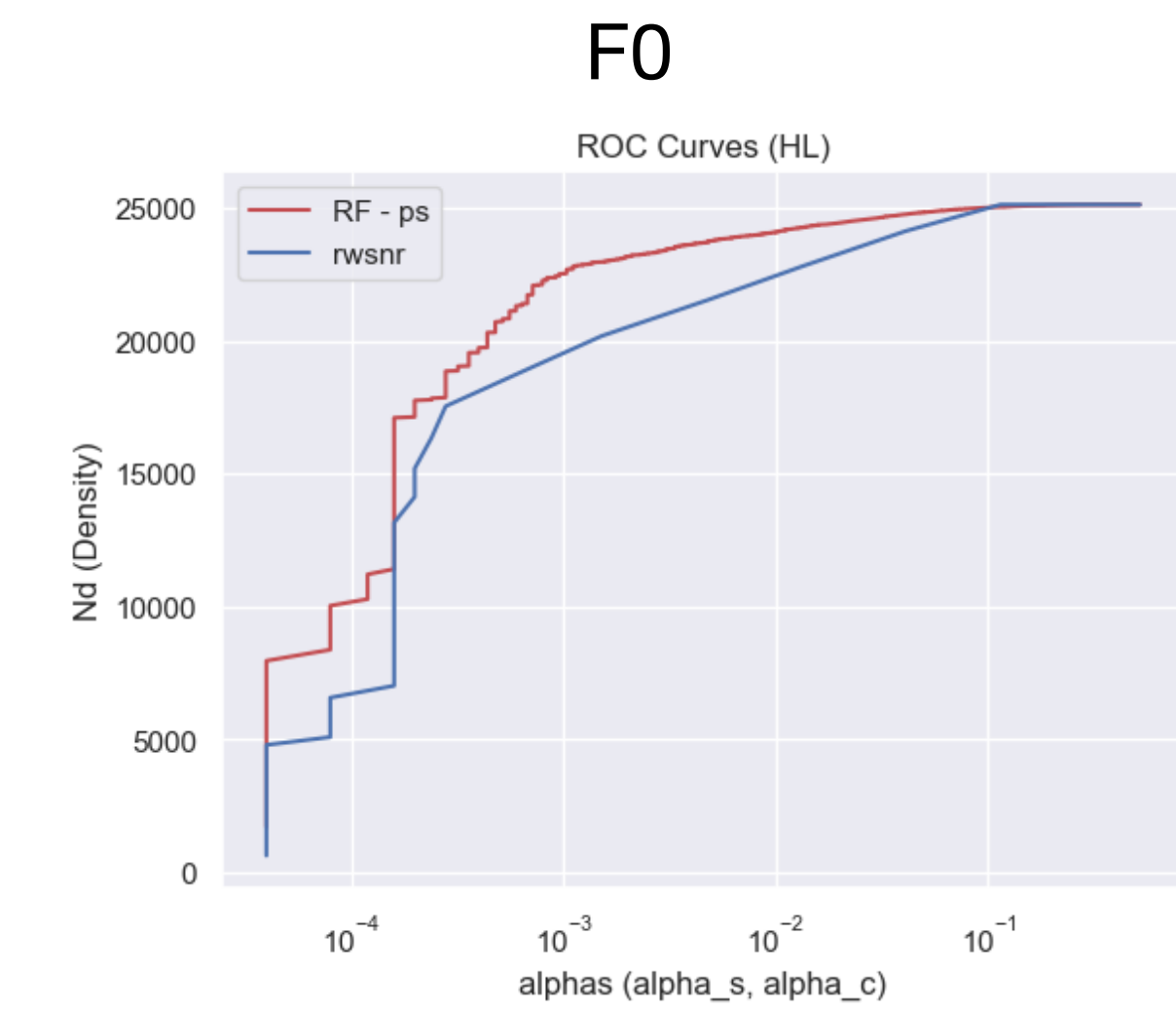
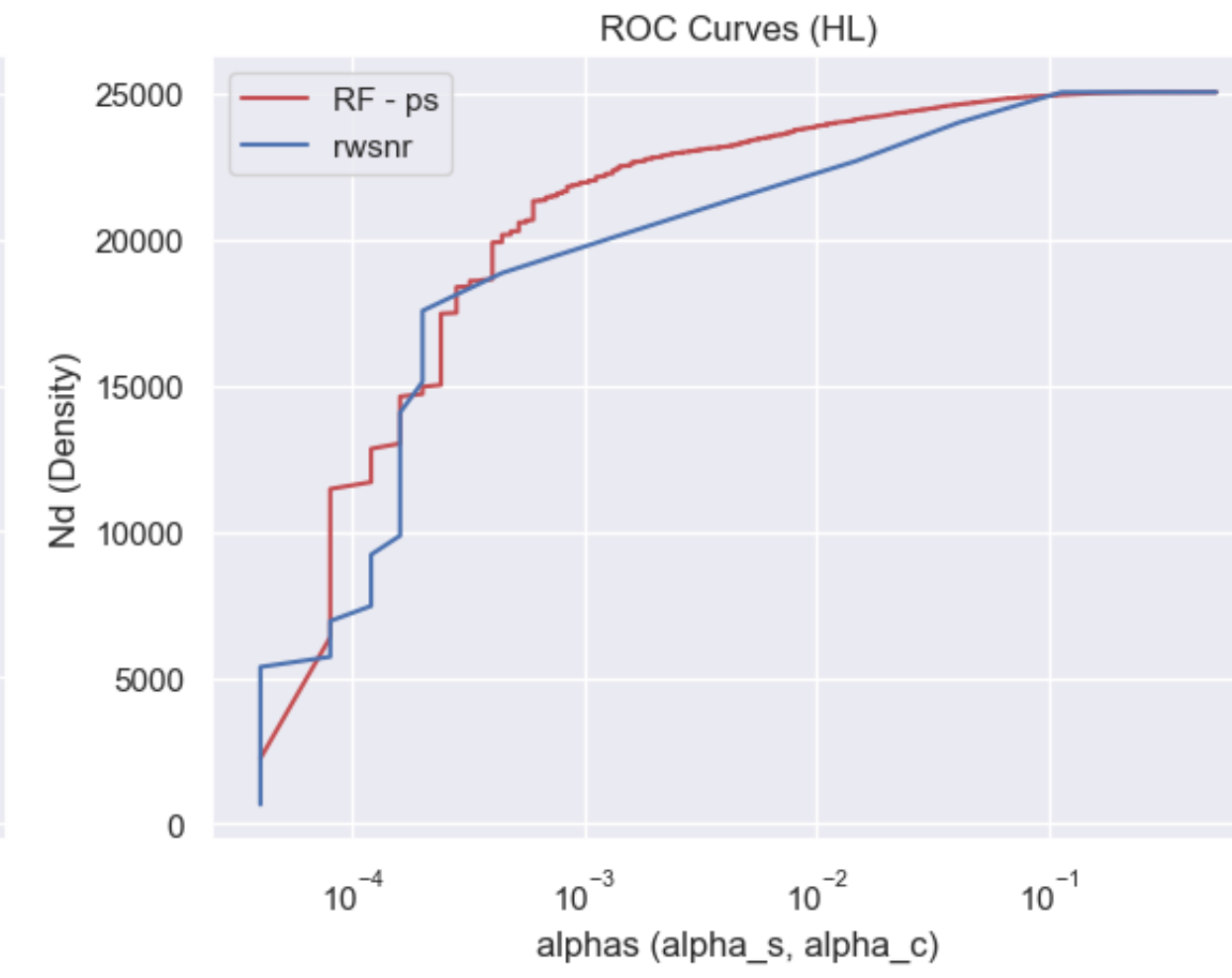
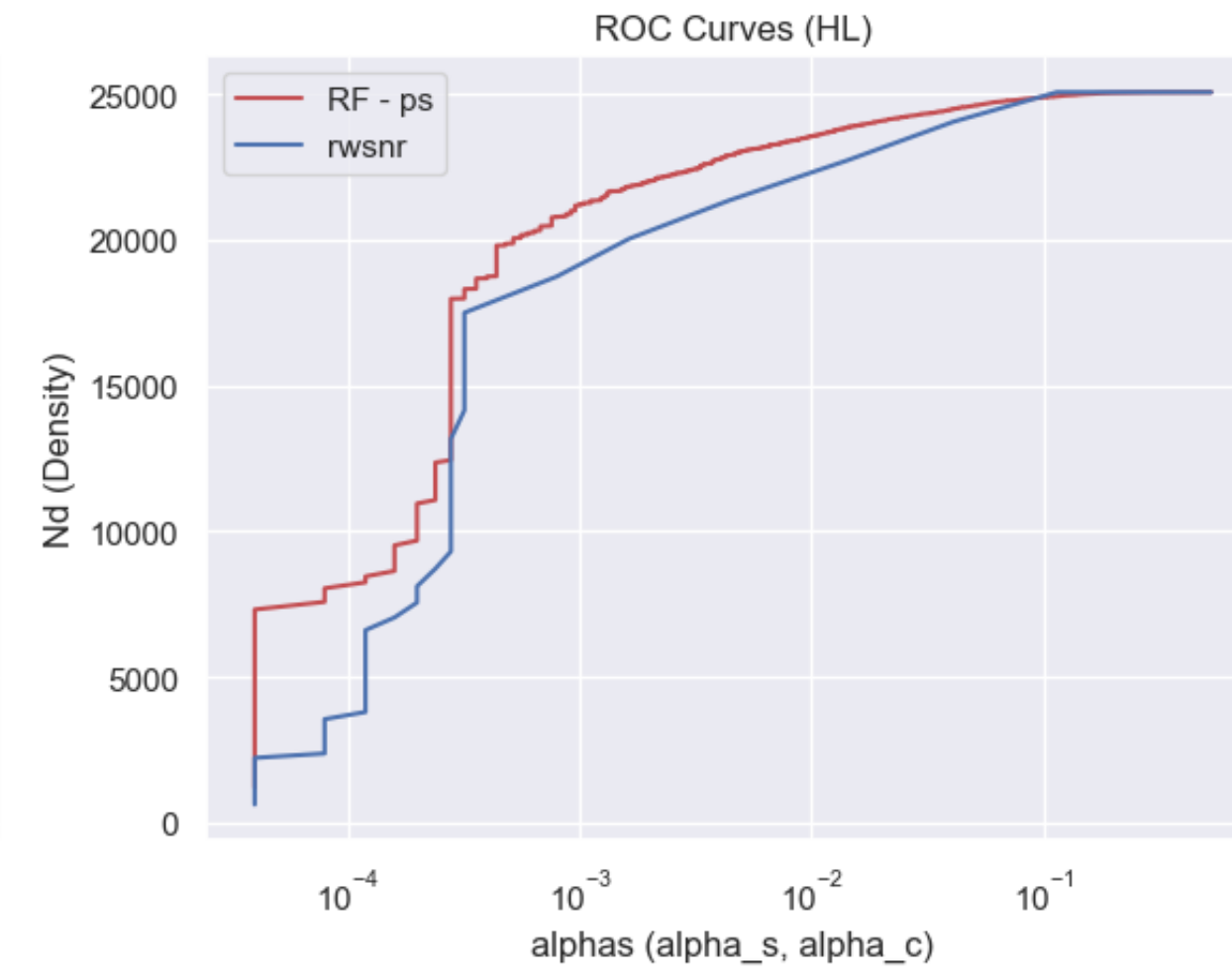
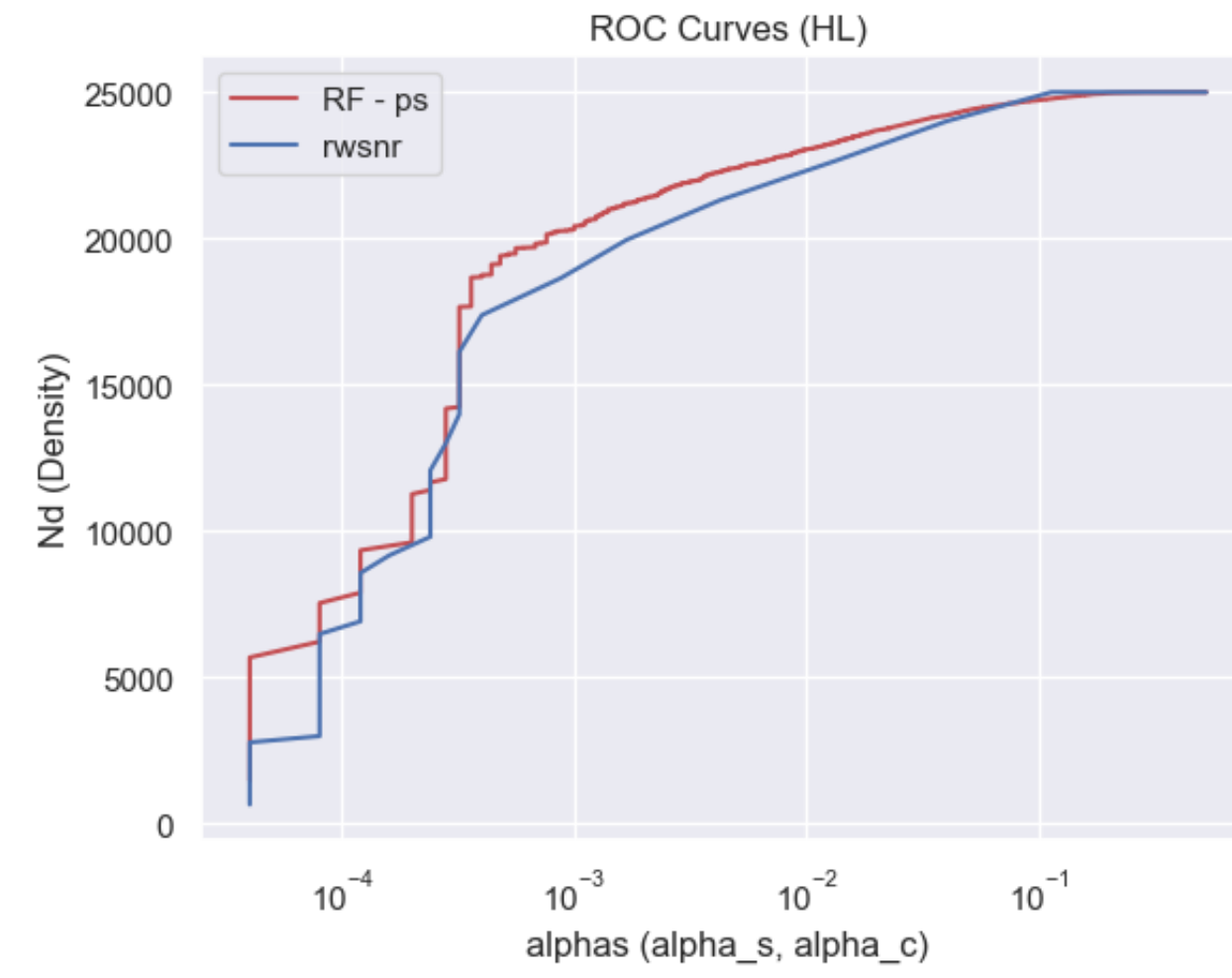
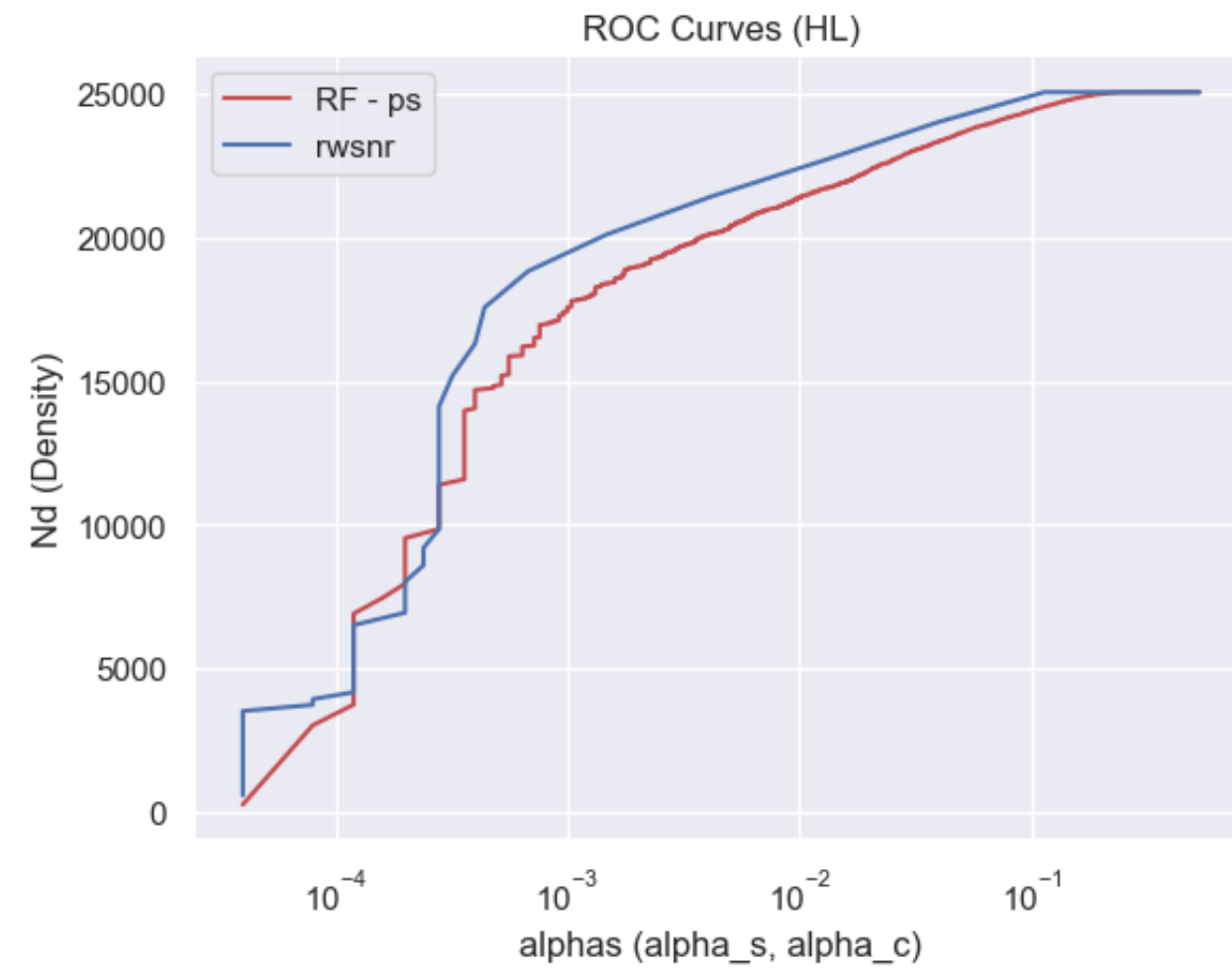


rwSNR vs False Alarm Probability



$$\alpha_s = \frac{1}{Nn} \sum_{i=1}^{Nn} \theta(\hat{p}_s^i - p_s)$$

Efficienza Detection F0->F7



F0

F1

F2

F3

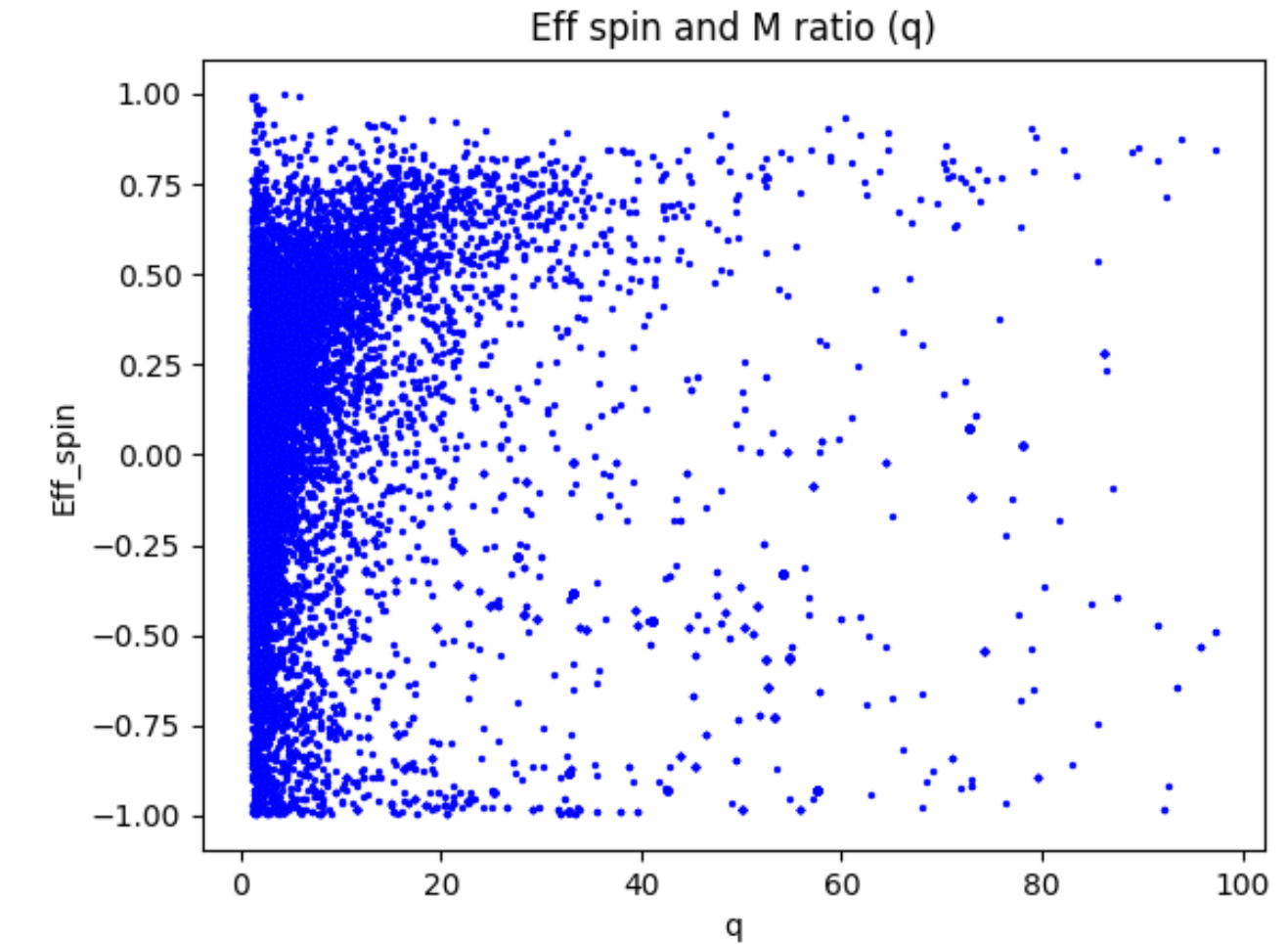
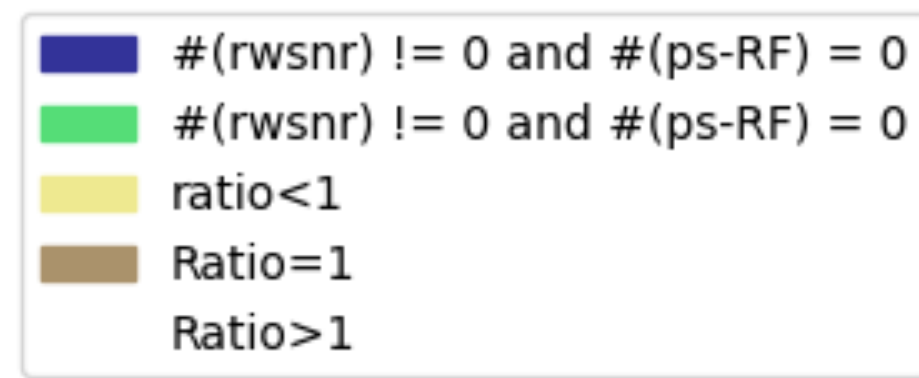
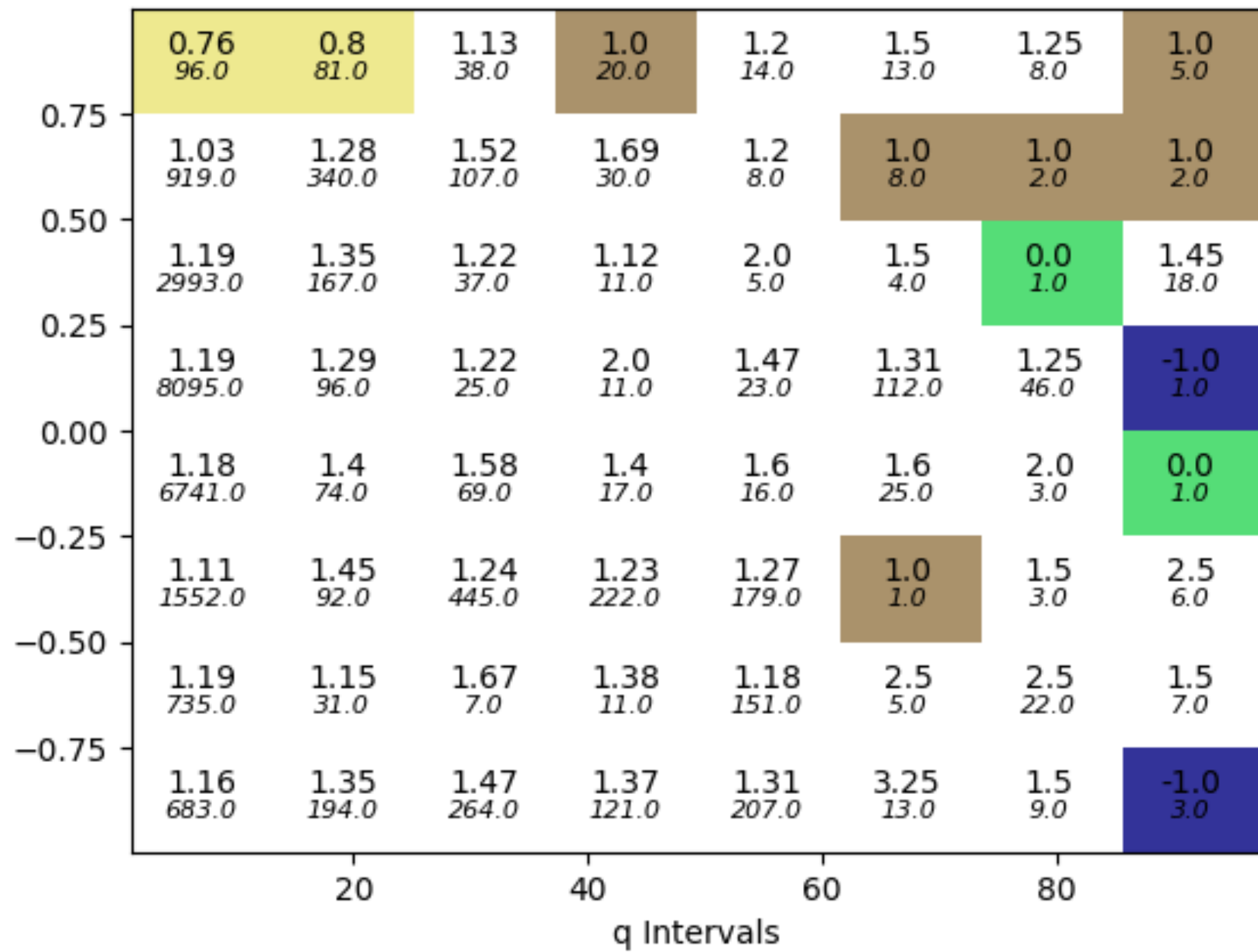
F4

F5

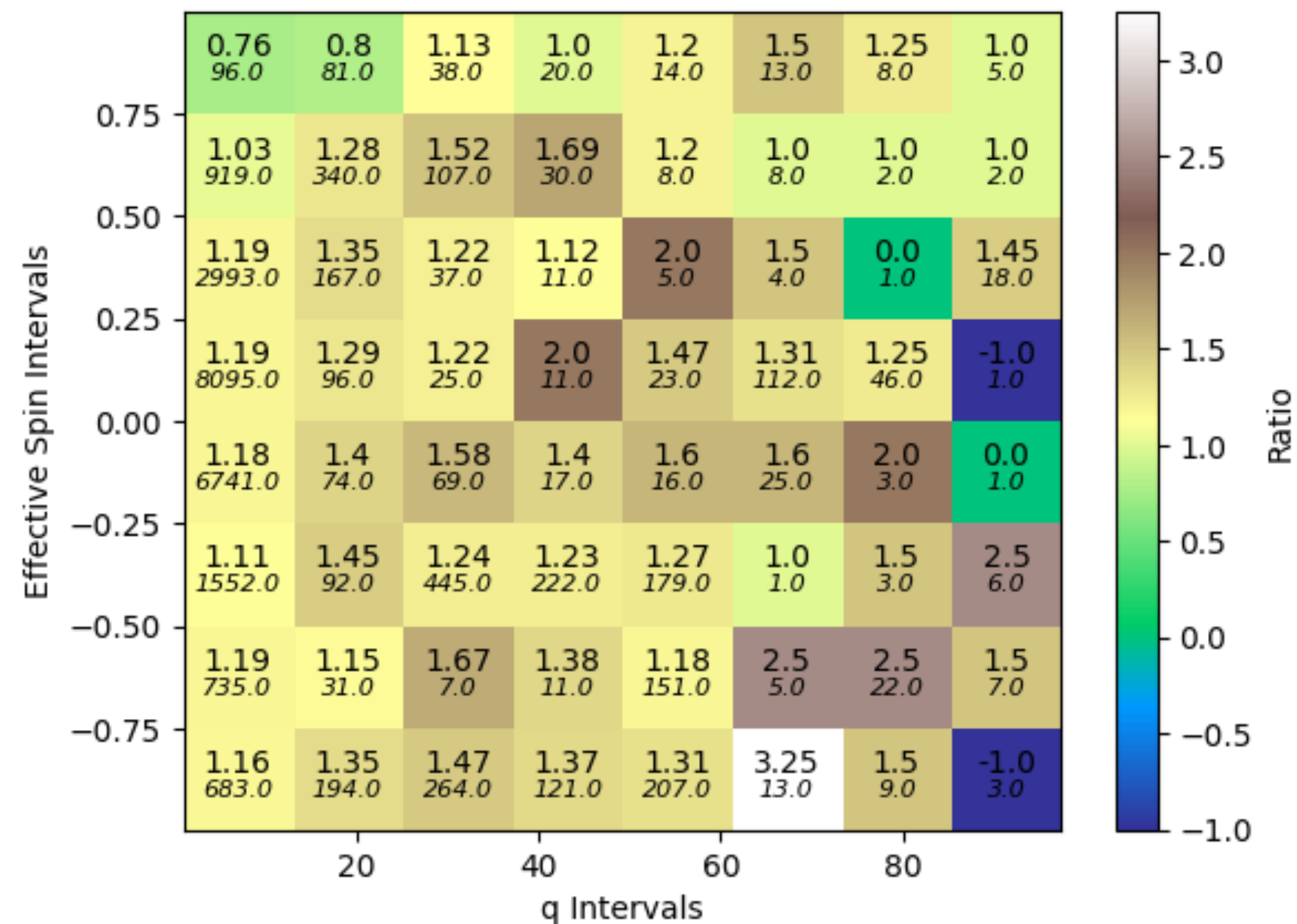
F6

Risultati - Efficienza nella Detection -Parametri fisici

between #events-RF and #events-rwsnr



between #events-RF and #events-rwsnr



In ogni 'box', in alto $\frac{\#RF}{\#Control} \Big|_{if\bar{a}r}$, $F\bar{a}P = 1 \times 10^{-3}$, in basso $\#Inj$