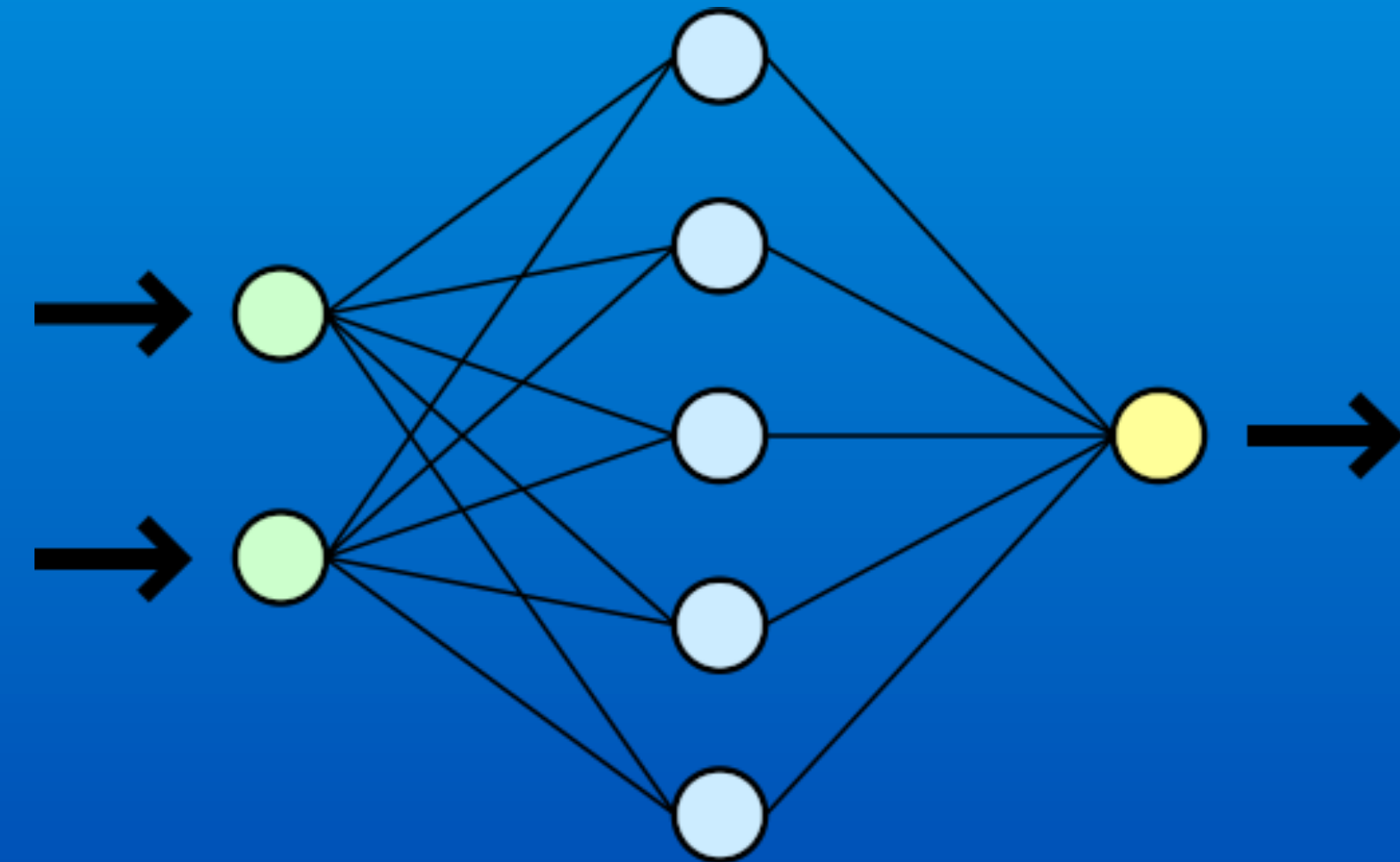
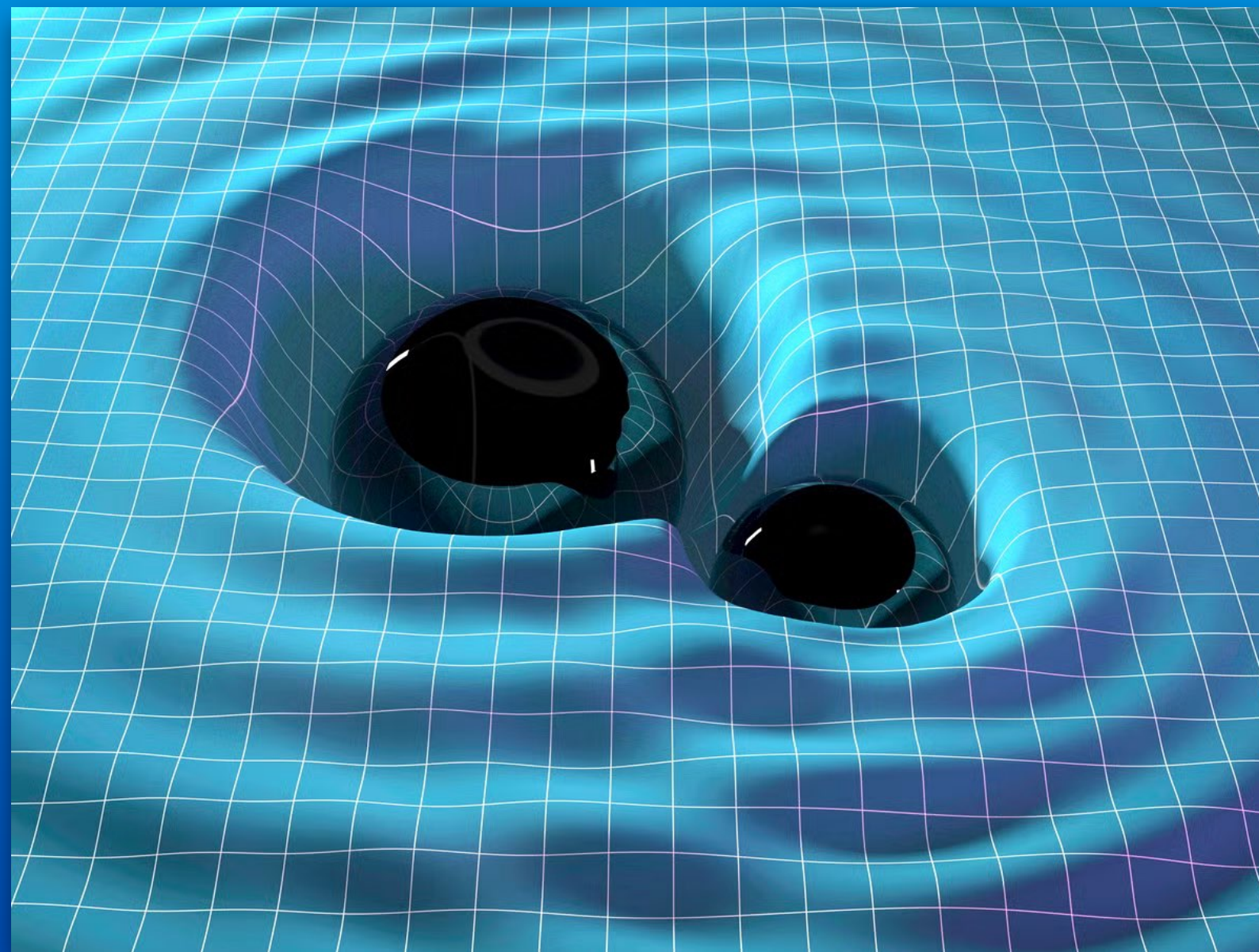


AresGW: Unveiling New Gravitational Wave Events with Machine Learning

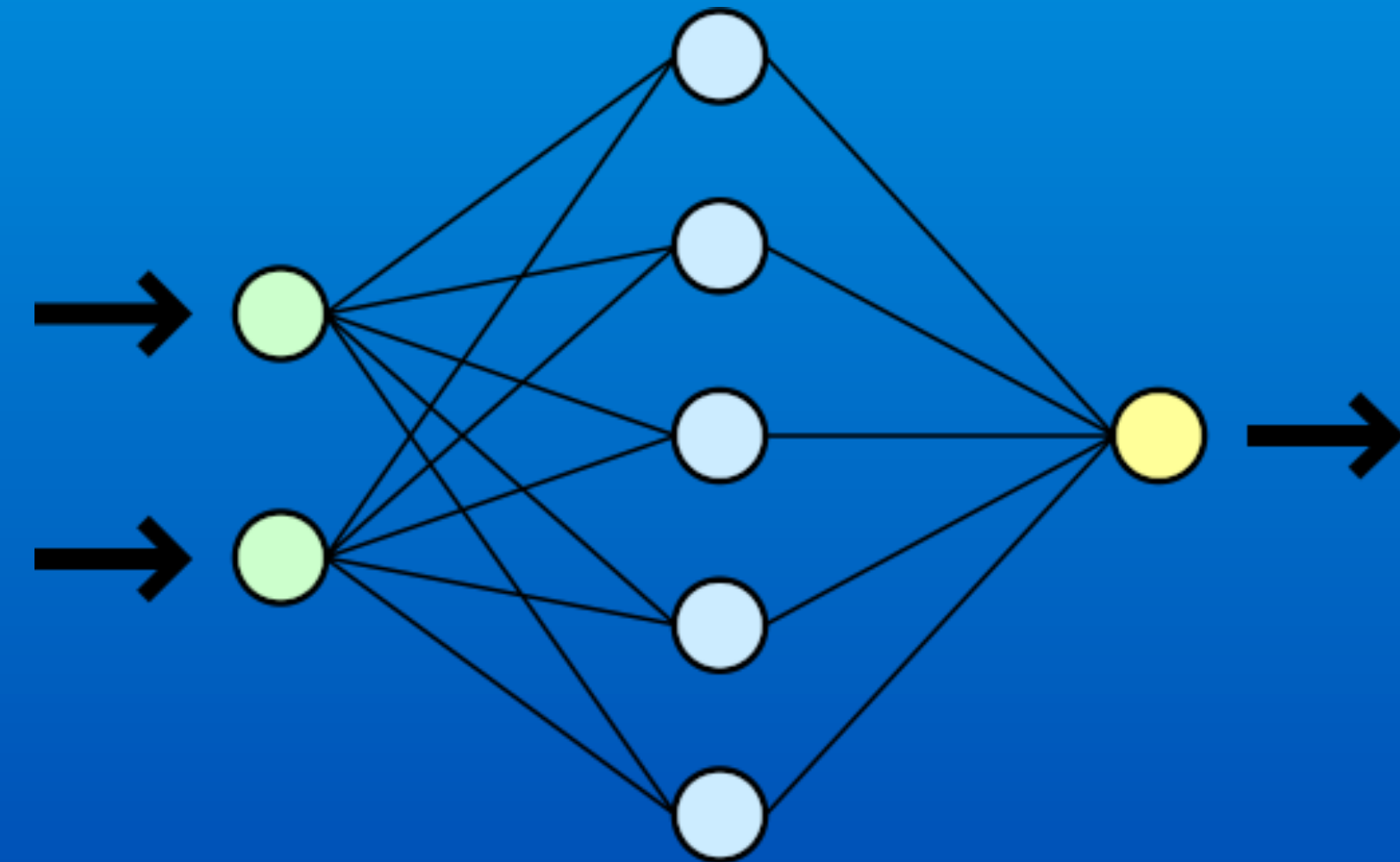
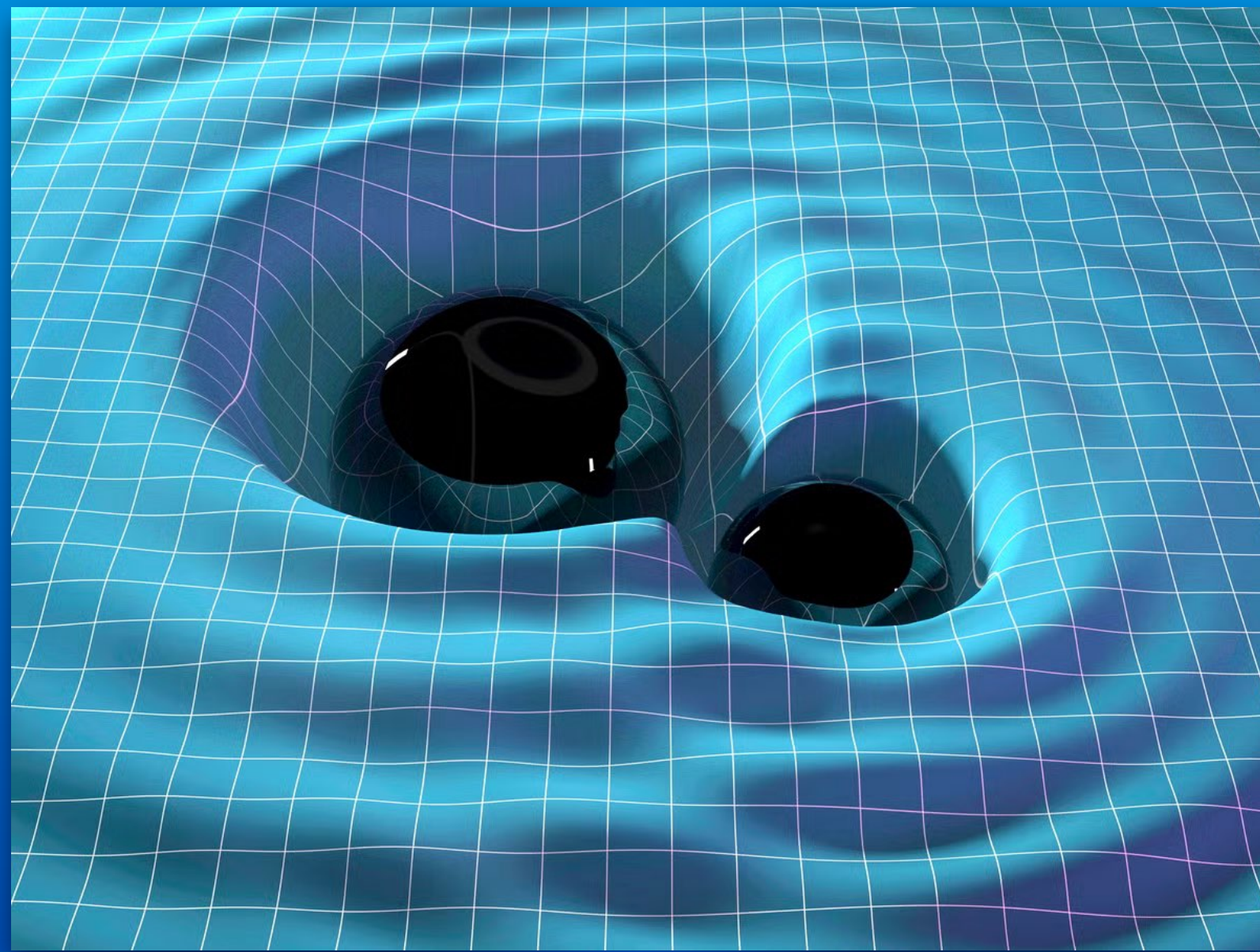
Aristotle University of Thessaloniki



Alexandra Eleni Koloniari

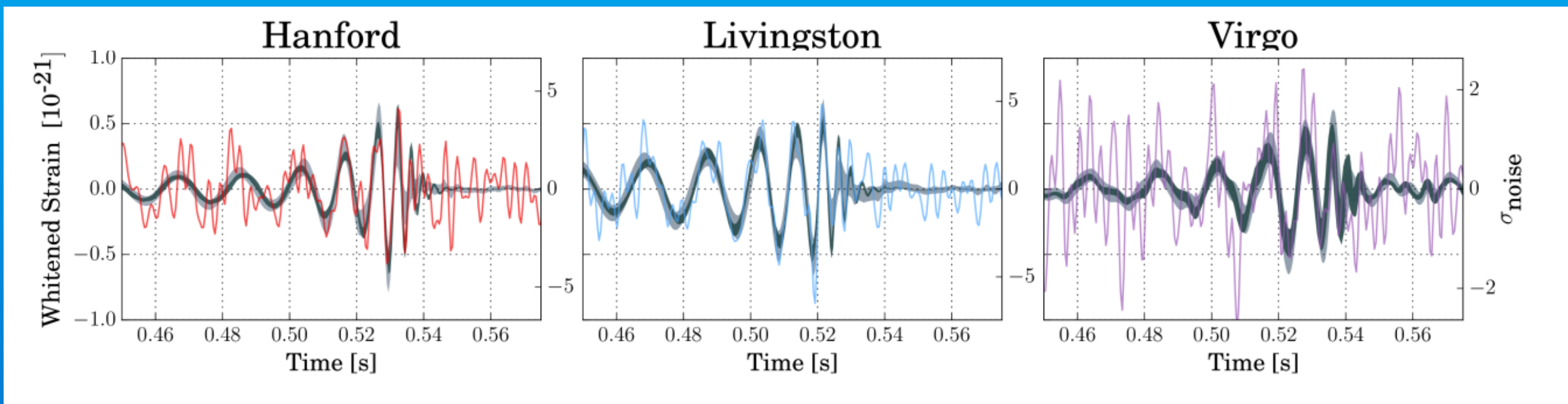
AresGW: Unveiling New Gravitational Wave Events with Machine Learning

Aristotle University of Thessaloniki



Alexandra Eleni Koloniari

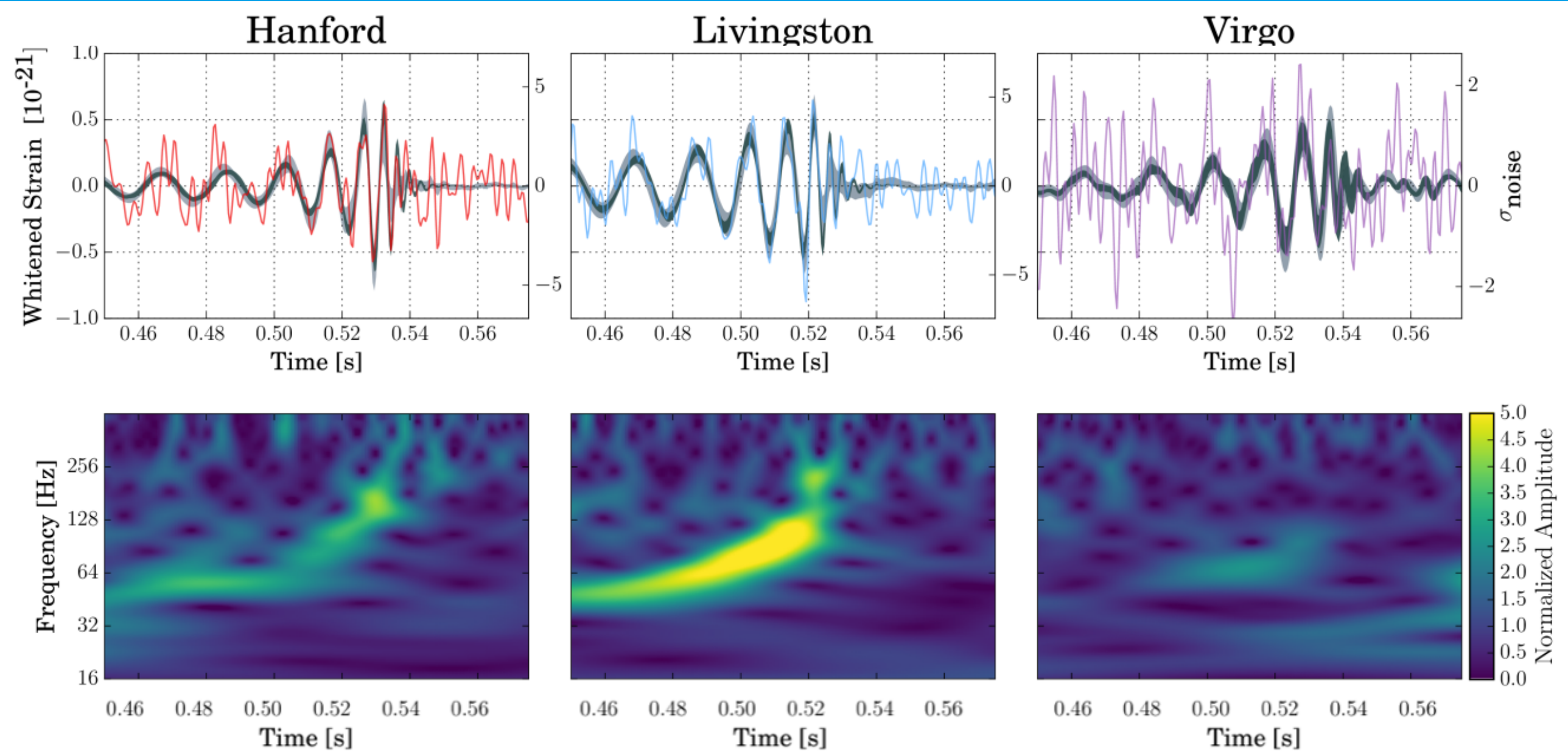
Different Representations of Gravitational Waves



Strain time series

GW strain data as a function of time

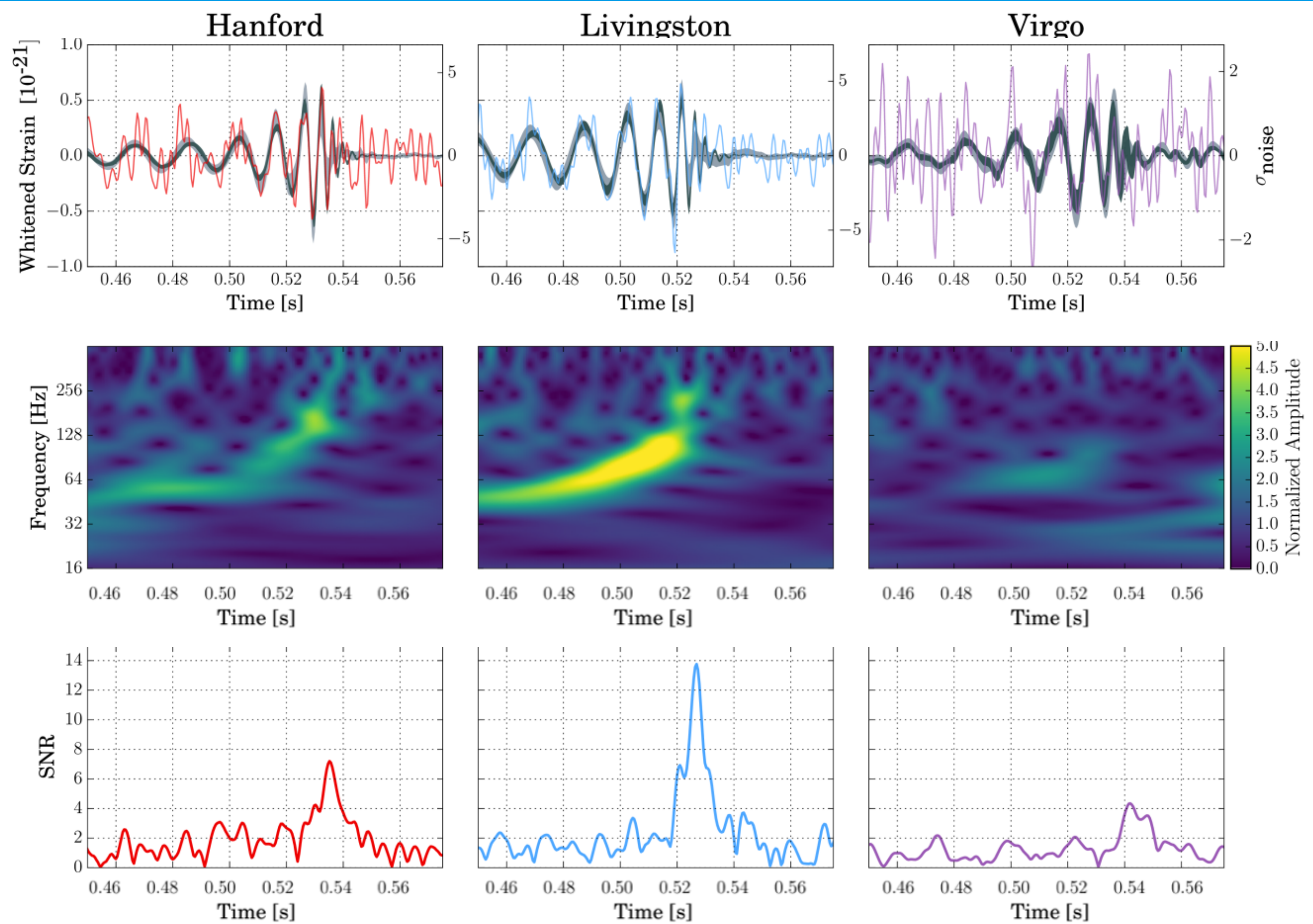
Different Representations of Gravitational Waves



← Strain time series
GW strain data as a function of time

← Spectrograms
Time-frequency representation of the data

Different Representations of Gravitational Waves

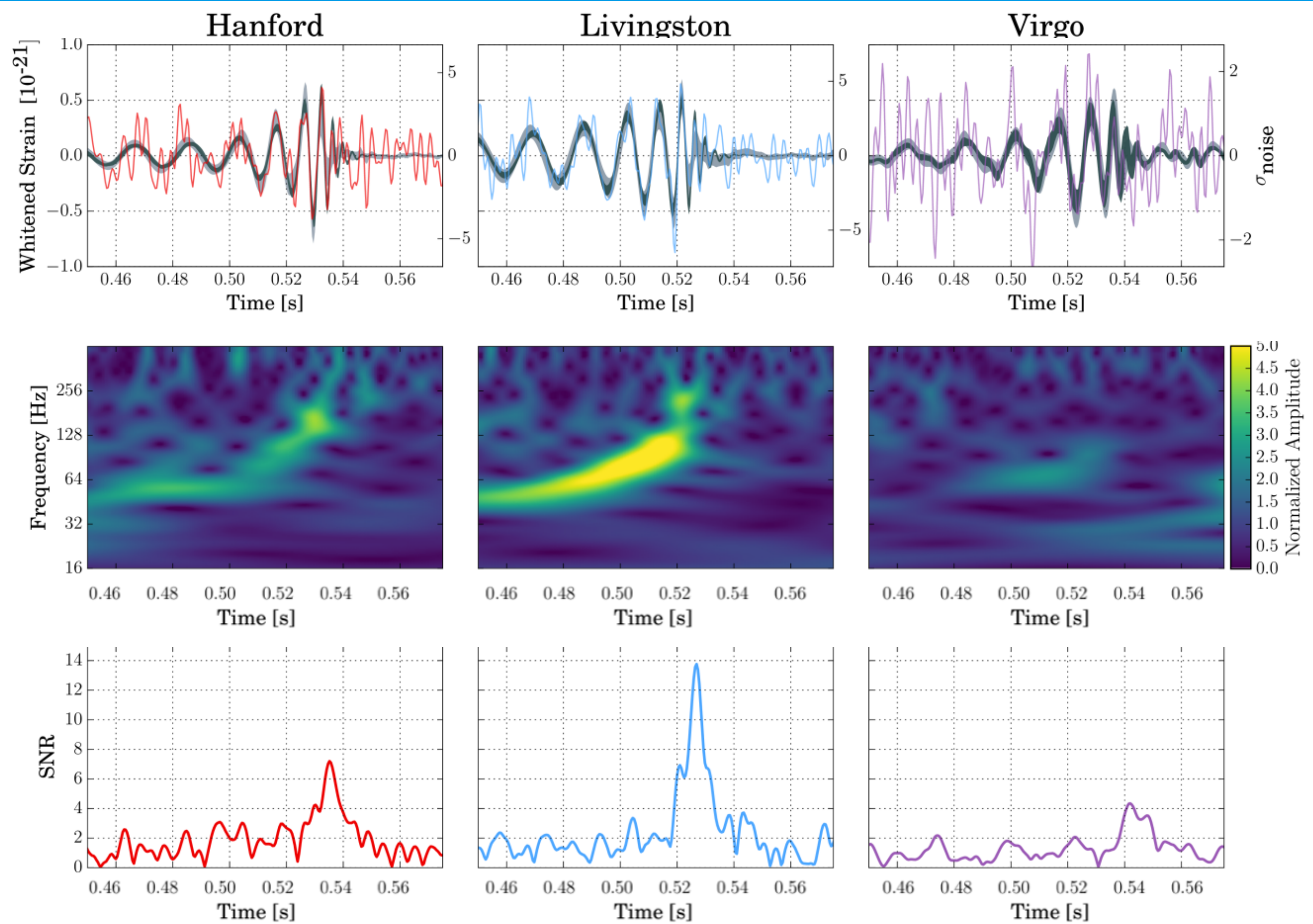


← Strain time series
GW strain data as a function of time

← Spectrograms
Time-frequency representation of the data

← Matched-filter SNR time series
This shows how well a template waveform matches the data over time

Different Representations of Gravitational Waves



← Strain time series
GW strain data as a function of time

← Spectrograms
Time-frequency representation of the data

← Matched-filter SNR time series
This shows how well a template waveform matches the data over time

Matched Filtering

It is a signal processing technique that compares a set of template waveforms to noisy data to detect signals with known morphologies.

Steps:

1. Fourier transform the data $\tilde{s}(f)$ and template $\tilde{h}(f)$
2. Compute the quantity:

$$(s | h) = 4 \Re \int_0^{+\infty} \frac{\tilde{s}(f) \tilde{h}^*(f)}{S_n(f)} df$$

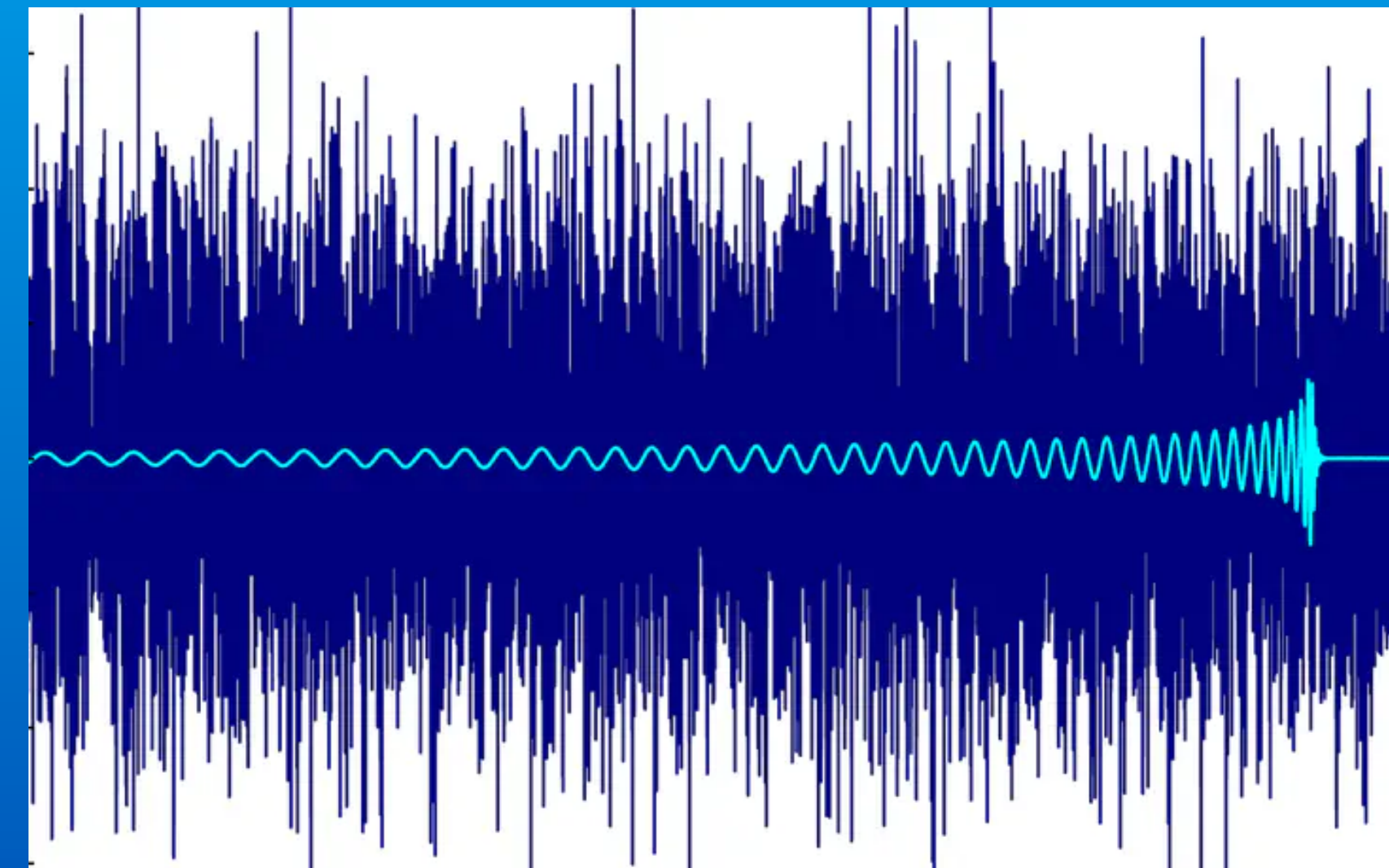
complex-conjugated template (pointing to $\tilde{h}^*(f)$)
noise power spectral density (pointing to $S_n(f)$)

3. Then the optimal SNR is:

$$\text{SNR}_{\text{opt}} = \sqrt{h | h}$$

Strengths of Matched Filtering

1. **Optimal for known signals in *Gaussian noise***
2. **Physically interpretable since it relies on *waveform templates***
3. **Well-established**
 - > Decades of development of GW template banks.
 - > Multiple detection pipelines (PyCBC, MBTA, GstLAL, IAS etc.)



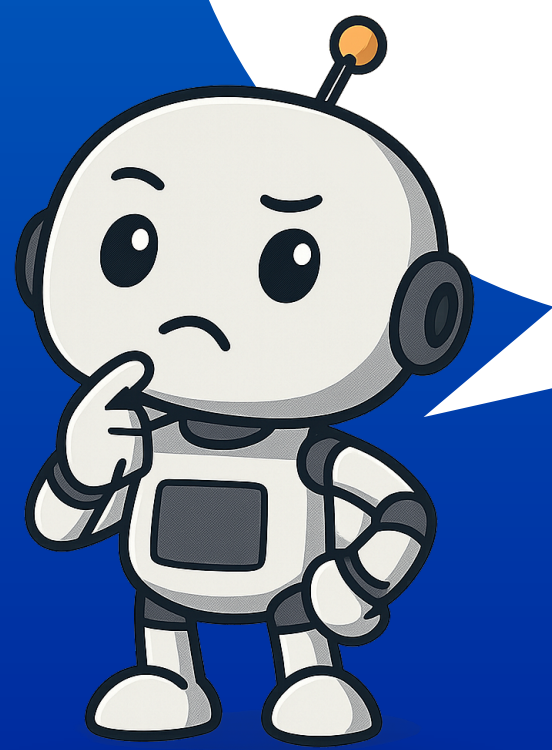
Limitations of Matched Filtering

1. Template dependence

-> Limited to signals similar to those covered by the waveform bank

2. Assumes Gaussian noise

-> Real detector noise is *non-Gaussian and non-stationary* (e.g., glitches), which reduces its effectiveness.



Limitations of Matched Filtering

1. Template dependence

-> Limited to signals similar to those covered by the waveform bank

2. Assumes Gaussian noise

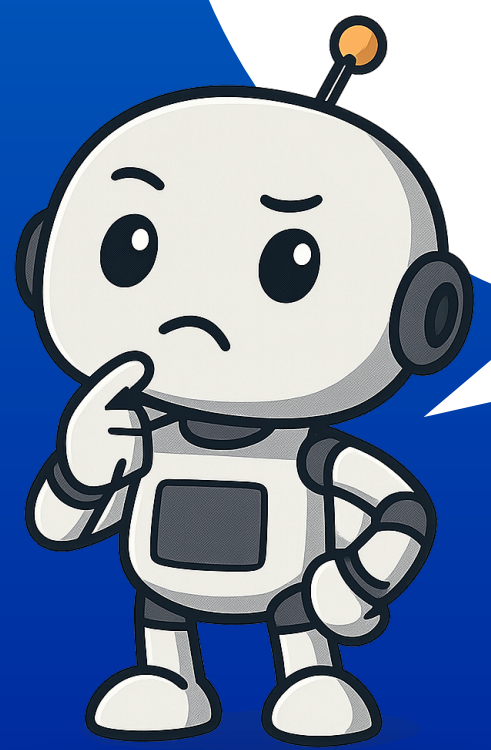
-> Real detector noise is *non-Gaussian and non-stationary* (e.g., glitches), which reduces its effectiveness.

3. Computationally expensive🔥

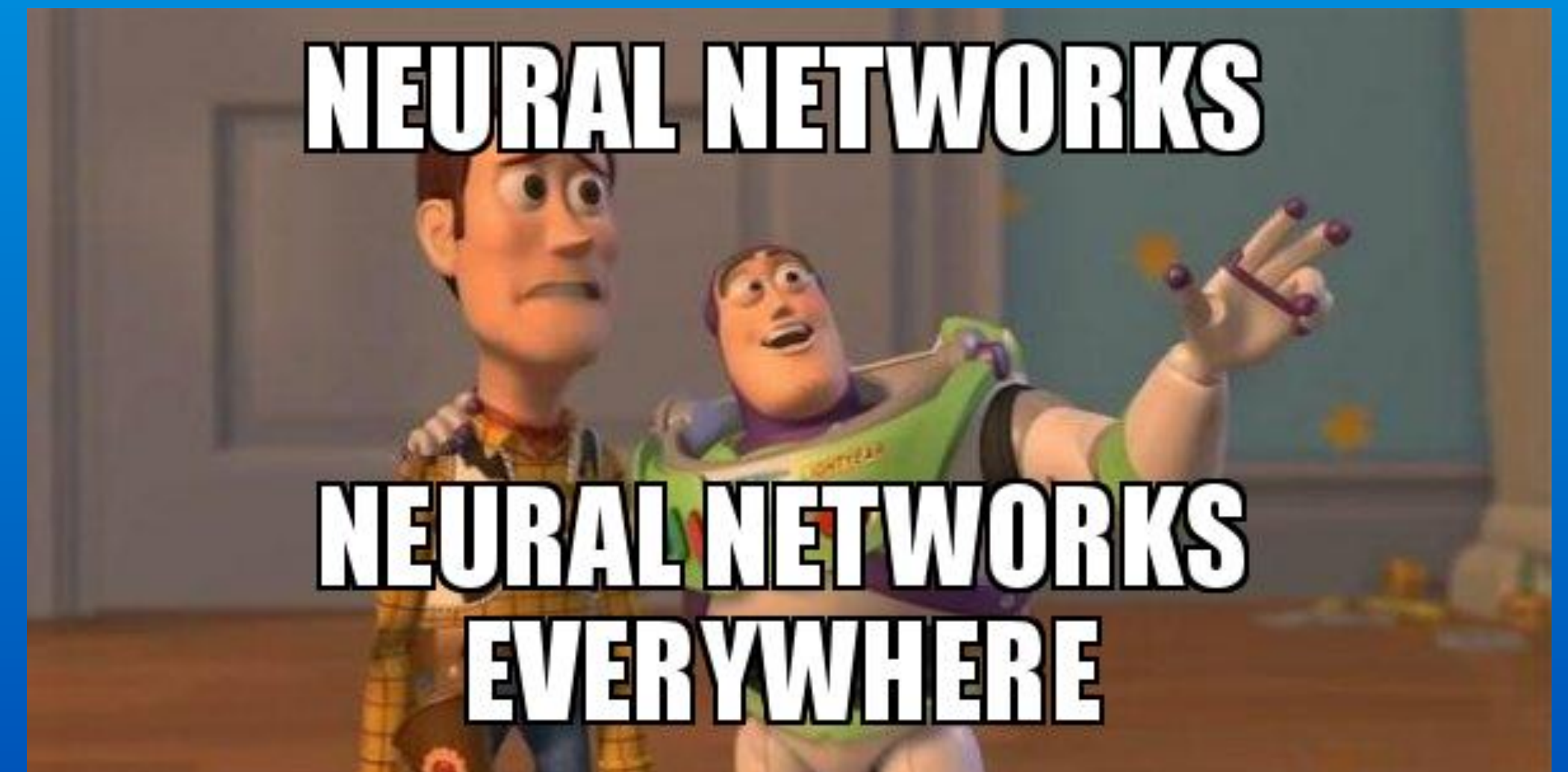
-> Comparing data against *millions of templates* is slow and resource-intensive, especially for long-duration or high-mass-ratio signals.

3. Limited by per-detector SNR thresholds🔥

-> Matched filtering often applies *fixed SNR thresholds* in each detector *separately* (e.g. ≥ 5.5), before coincidence checks.

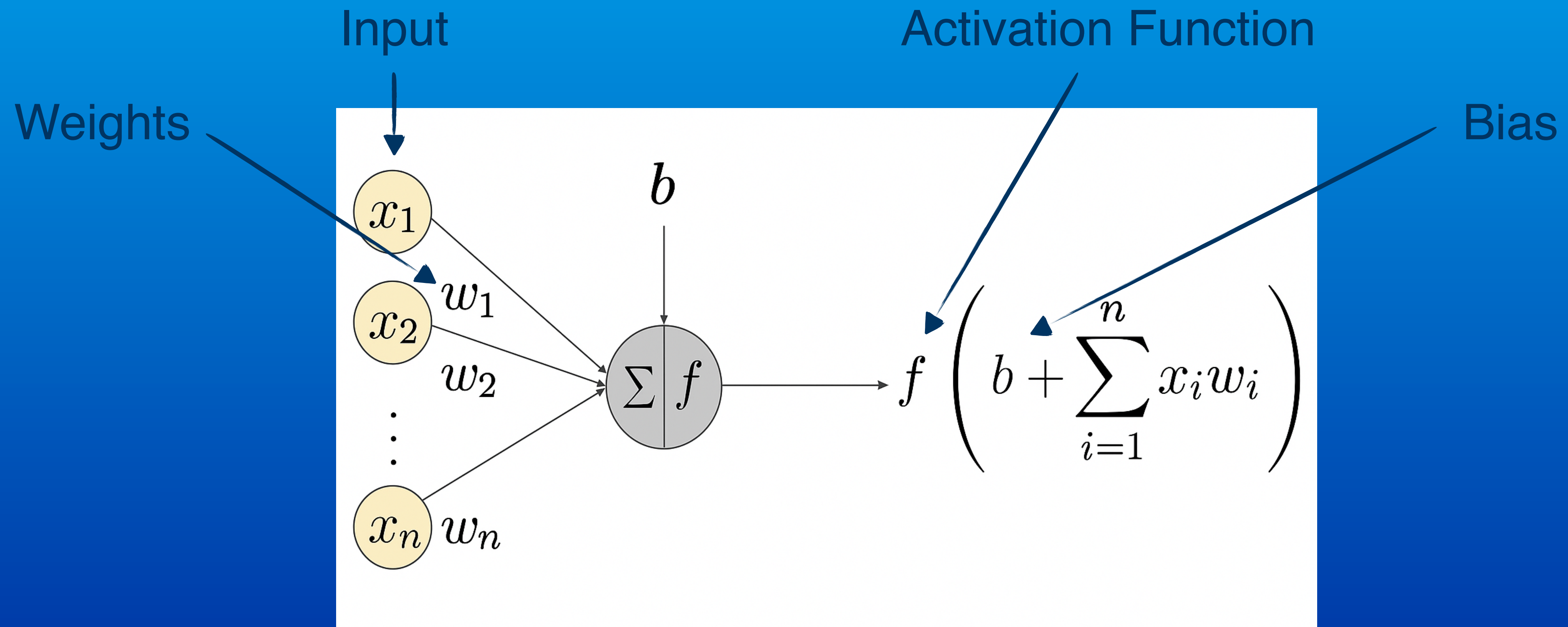


So is there anything else we can try?



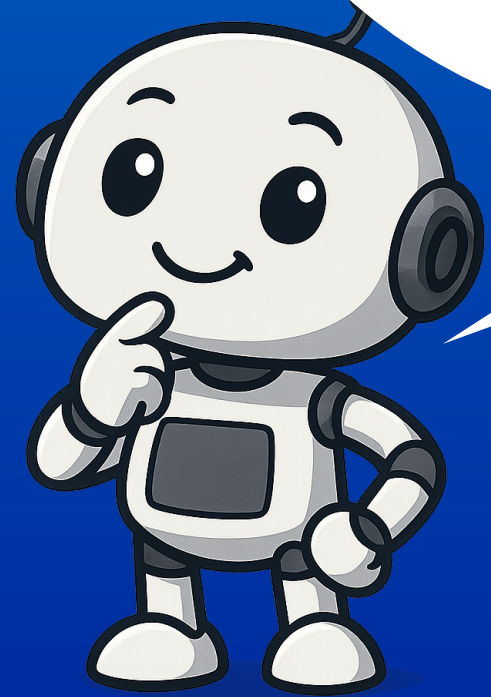
Neural Networks

They are *artificial intelligence* (AI) models inspired by *biological neurons and their connections* that learn patterns from *data* to make predictions or decisions.



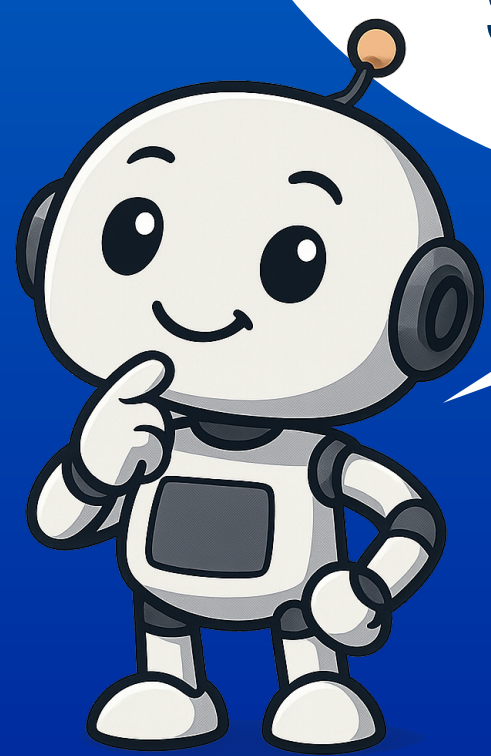
Strengths of Neural Networks

1. NNs sometimes can detect signals outside the training template space
2. Can recognize non-Gaussian noise *with proper training*



Strengths of Neural Networks

1. NNs sometimes can detect signals outside the training template space
2. Can recognize non-Gaussian noise *with proper training*
3. **Real-time detection speed** 🔥
 - > Once trained, NNs are *orders of magnitude faster* than matched filtering
4. **Adaptable to multi-detector data** 🔥
 - > NNs can detect GWs by jointly analyzing multi-detector data without requiring per-detector thresholds or coincidence tests.
 - > Can detect signals that are *consistently weak* in all detectors but still significant as a network SNR.



Limitations of Neural Networks

1. **Limited interpretability** – it's harder to understand why they make a detection.
2. **Training data dependency**
 - > Require large, diverse, and high-quality training sets (*Poor training = poor performance*)
3. *Might* **require retraining for new detector data**

A little bit of history...

PHYSICAL REVIEW D

published 27 January 2023

First machine learning gravitational-wave search mock data challenge

Marlin B. Schäfer^{1,2}, Ondřej Zelenka^{3,4}, Alexander H. Nitz^{1,2}, He Wang⁵, Shichao Wu^{1,2}, Zong-Kuan Guo⁵,
Zhoujian Cao⁶, Zhixiang Ren⁷, Paraskevi Nousi⁸, Nikolaos Stergioulas⁹, Panagiotis Iosif^{10,9},
Alexandra E. Koloniari⁹, Anastasios Tefas⁸, Nikolaos Passalis⁸, Francesco Salemi^{11,12}, Gabriele Vedovato¹³,
Sergey Klimenko¹⁴, Tanmaya Mishra¹⁴, Bernd Brügmann^{3,4}, Elena Cuoco^{15,16,17}, E. A. Huerta^{18,19},
Chris Messenger²⁰ and Frank Ohme^{1,2}

Test Datasets

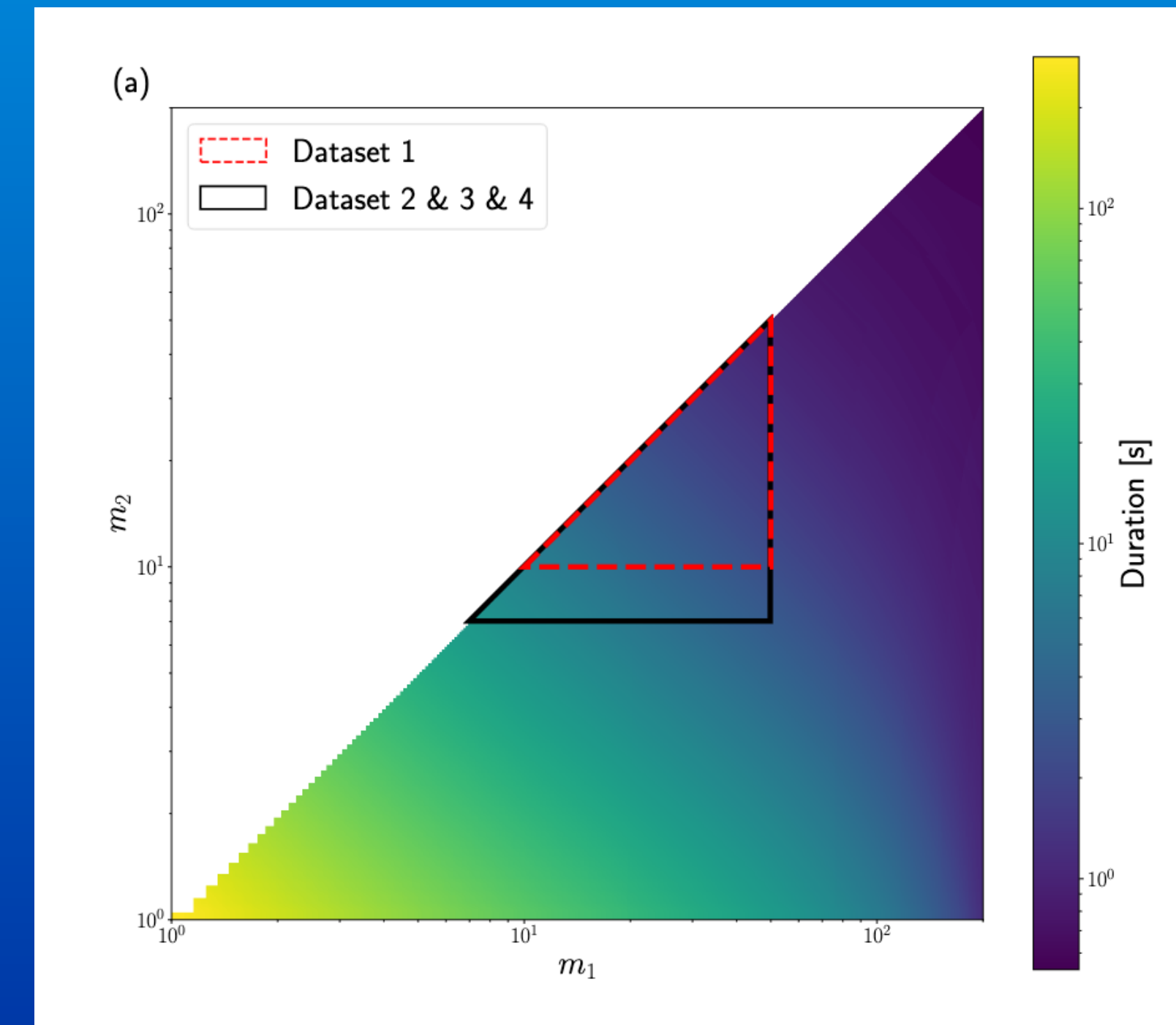
There were 4 *datasets*:

- 3 contained Gaussian noise with different PSD variations,
- 1 contained real O3a LIGO noise *cleaned of GWTC-2 events*,
- In all datasets, the positive samples contained injections generated using the *IMRPhenomXPHM* waveform model.

All files included 2 groups, “H1” and “L1”, representing data from the 2 LIGO detectors.

Parameter	Uniform distribution
Coalescence phase	$\Phi_0 \in (0, 2\pi)$
Polarization	$\Psi \in (0, 2\pi)$
Inclination	$\cos \iota \in (-1, 1)$
Declination	$\sin \theta \in (-1, 1)$
Right ascension	$\varphi \in (-\pi, \pi)$
Chirp-Distance	$d_c^2 \in (130^2, 350^2) \text{ Mpc}^2$

TABLE I. A summary of the distributions shared between all datasets from which parameters are drawn.



Results

Predecessor of AresGW model 1 and 2

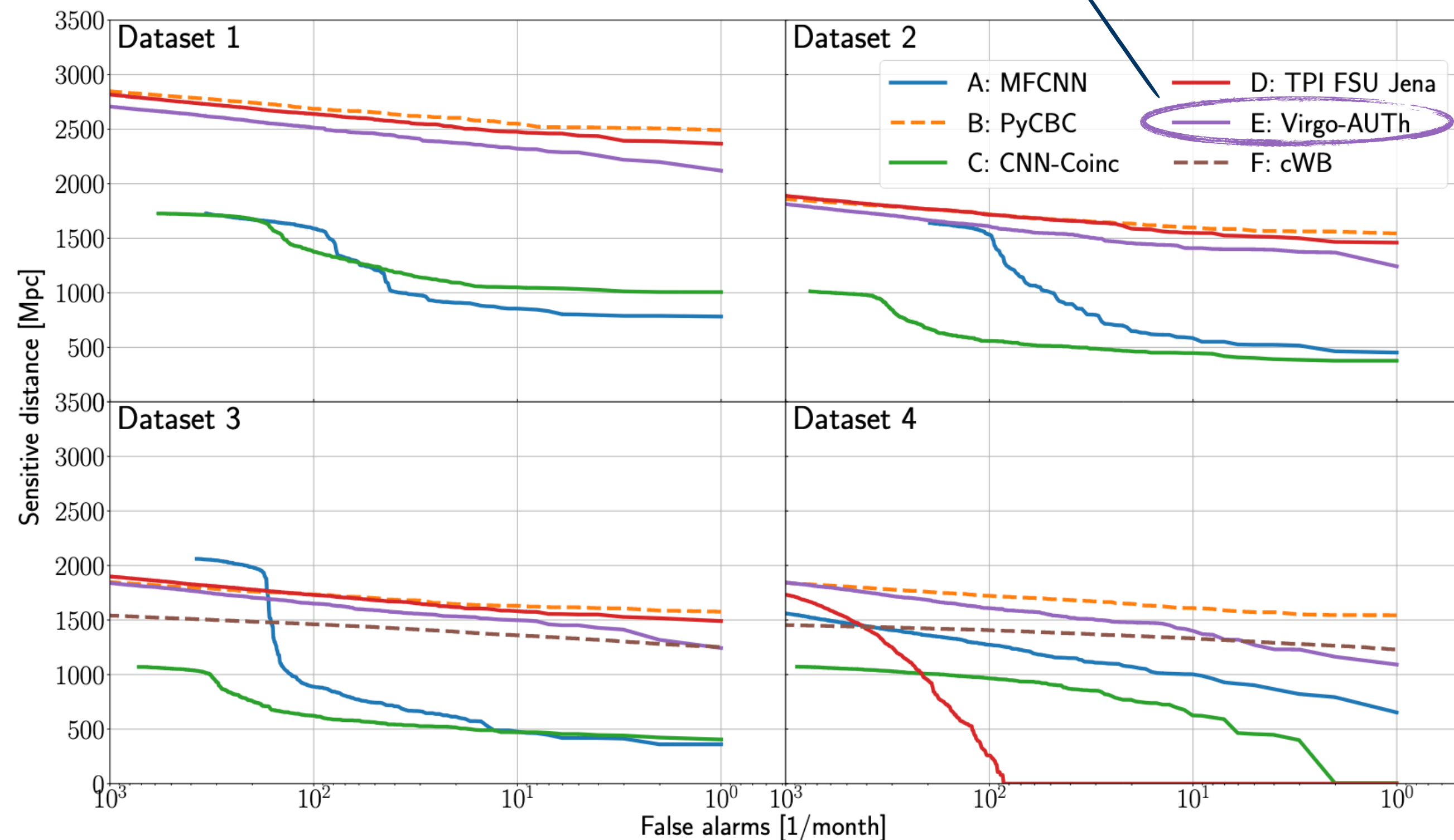


FIG. 2. The sensitive distances of all submissions and all four datasets as functions of the FAR. Submissions that made use of a machine learning algorithm at their core are shown with solid lines, others with dashed lines. The FAR was calculated on a background set that does not contain any injections.

Sensitive distance:

-> Represents the effective range within which a GW detection algorithm can detect sources at a given FAR.

-> Accounts for both *detection efficiency* and *source distribution*.

Can we reach traditional algorithms?

PHYSICAL REVIEW D

published 11 July 2023

Deep residual networks for gravitational wave detection

Paraskevi Nousi¹, Alexandra E. Koloniari², Nikolaos Passalis¹, Panagiotis Iosif²,
Nikolaos Stergioulas², and Anastasios Tefas¹

AresGW model 1

AresGW model 1 is a machine learning GW detection algorithm for BBHs.

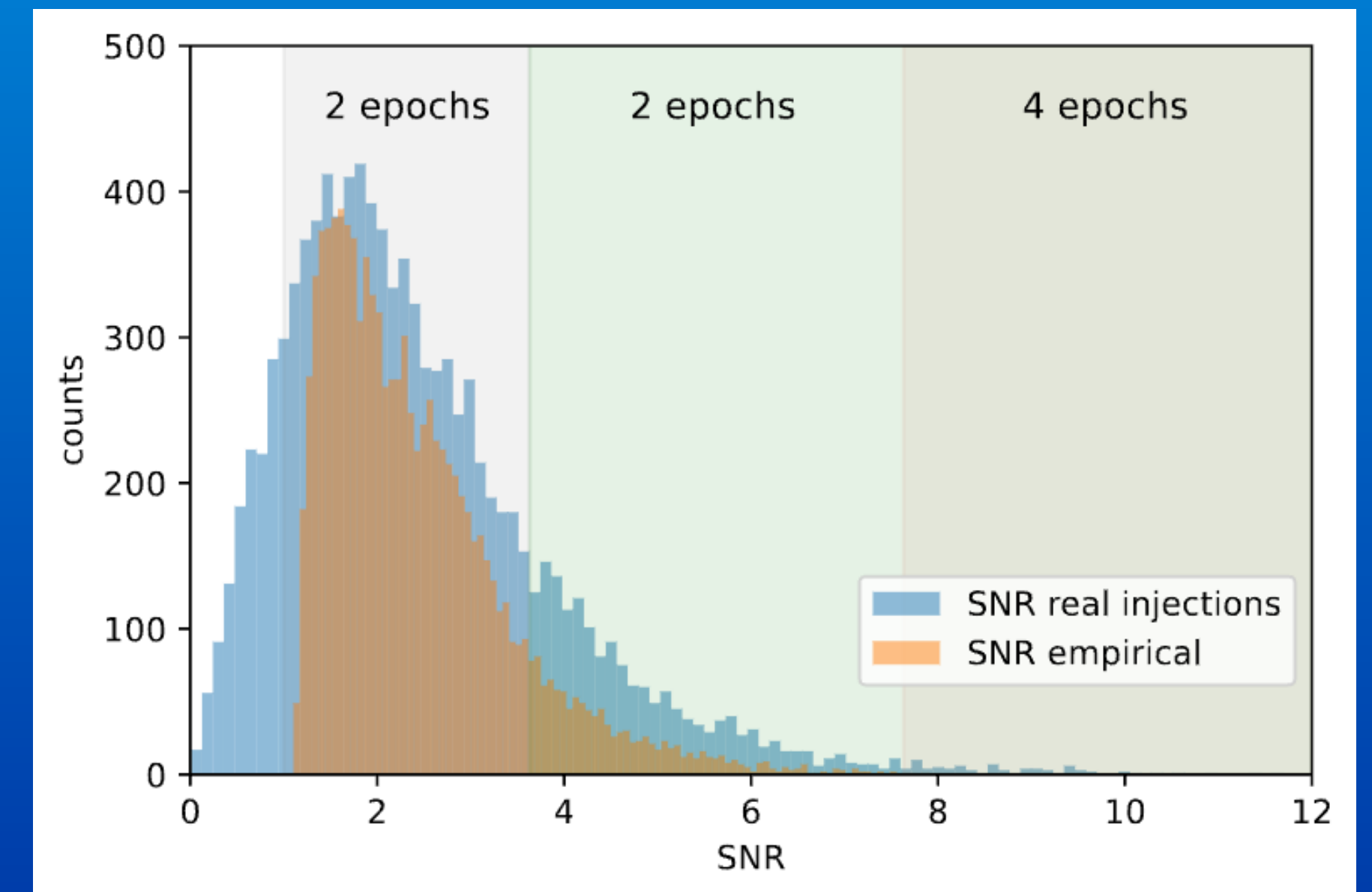
- 27 residual blocks
- 54 layers in total

*1.3M learnable
parameters*

Innovations

- Deep Adaptive Input Normalization (DAIN)
- Dynamic dataset augmentation
- Curriculum learning

SNR schedule



Training Dataset

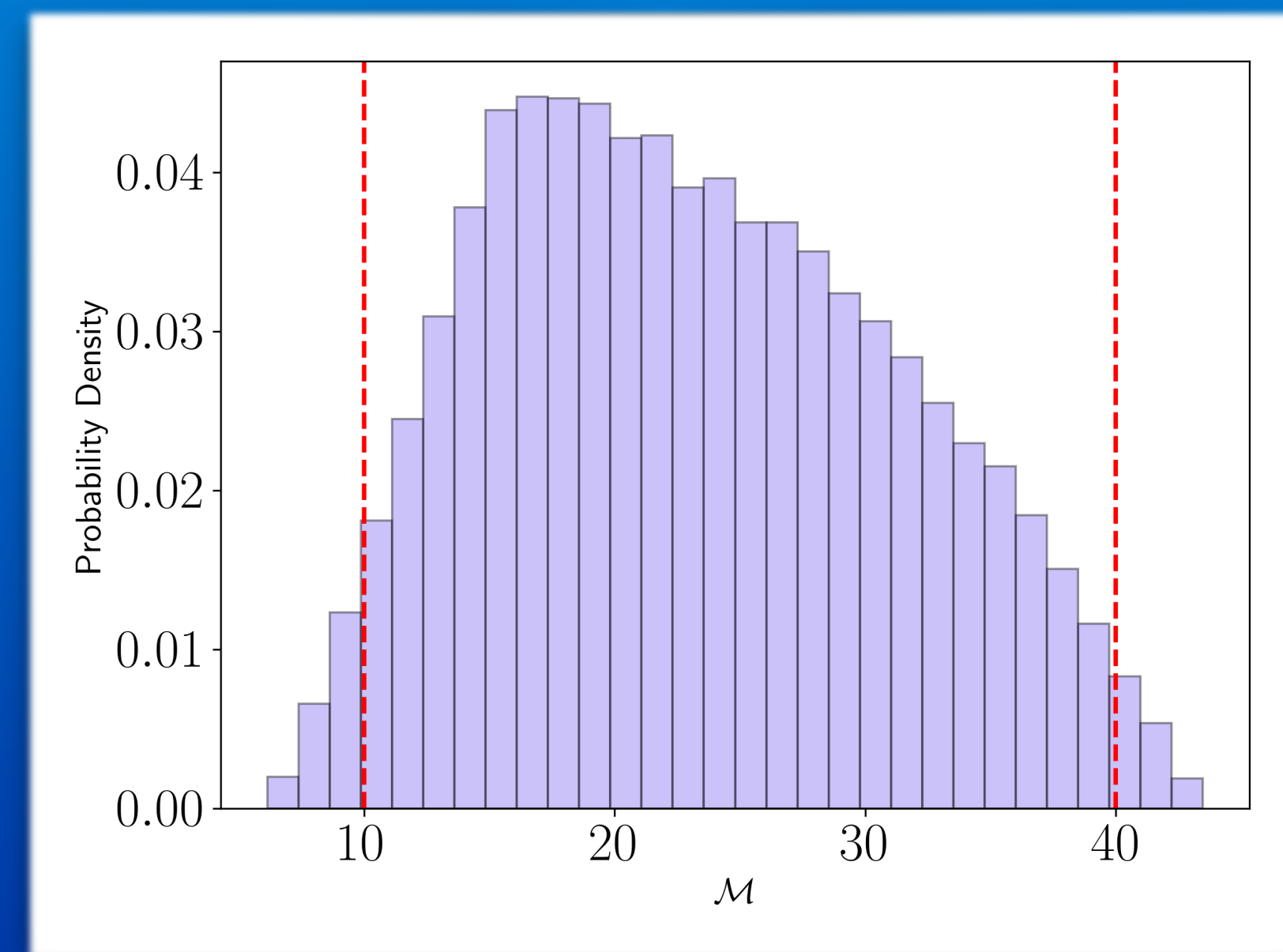
- Training Dataset Duration: 12 days
- Noise: Real data-quality noise of O3a from both LIGO detectors
- Waveform Model: IMRPhenomXPHM
- Mass Range: $7 M_{\odot} \leq m_{1,2} \leq 50 M_{\odot}$

Training Dataset

- Training Dataset Duration: 12 days
- Noise: Real data-quality noise of O3a from both LIGO detectors
- Waveform Model: IMRPhenomXPHM
- Mass Range: $7 M_{\odot} \leq m_{1,2} \leq 50 M_{\odot}$

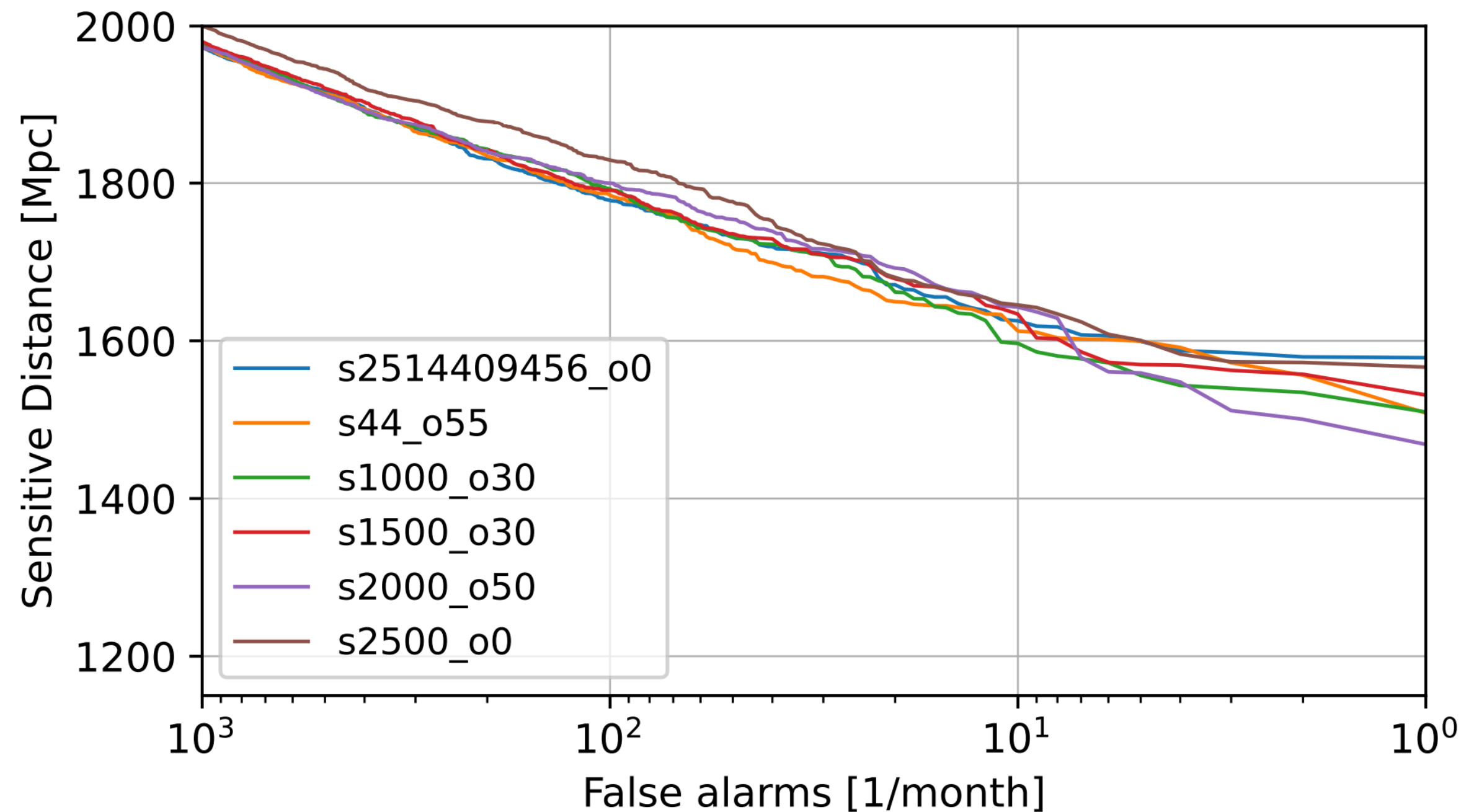
Effective Training Range:

- M_{chirp} effective range:
 - $\mathcal{M}_{\text{chirp}} \leq 10 M_{\odot}, p = 0.03$
 - $\mathcal{M}_{\text{chirp}} \geq 40 M_{\odot}, p = 0.02$
- } $10 M_{\odot} \leq \mathcal{M}_{\text{chirp}} \leq 40 M_{\odot}$
95 % CI

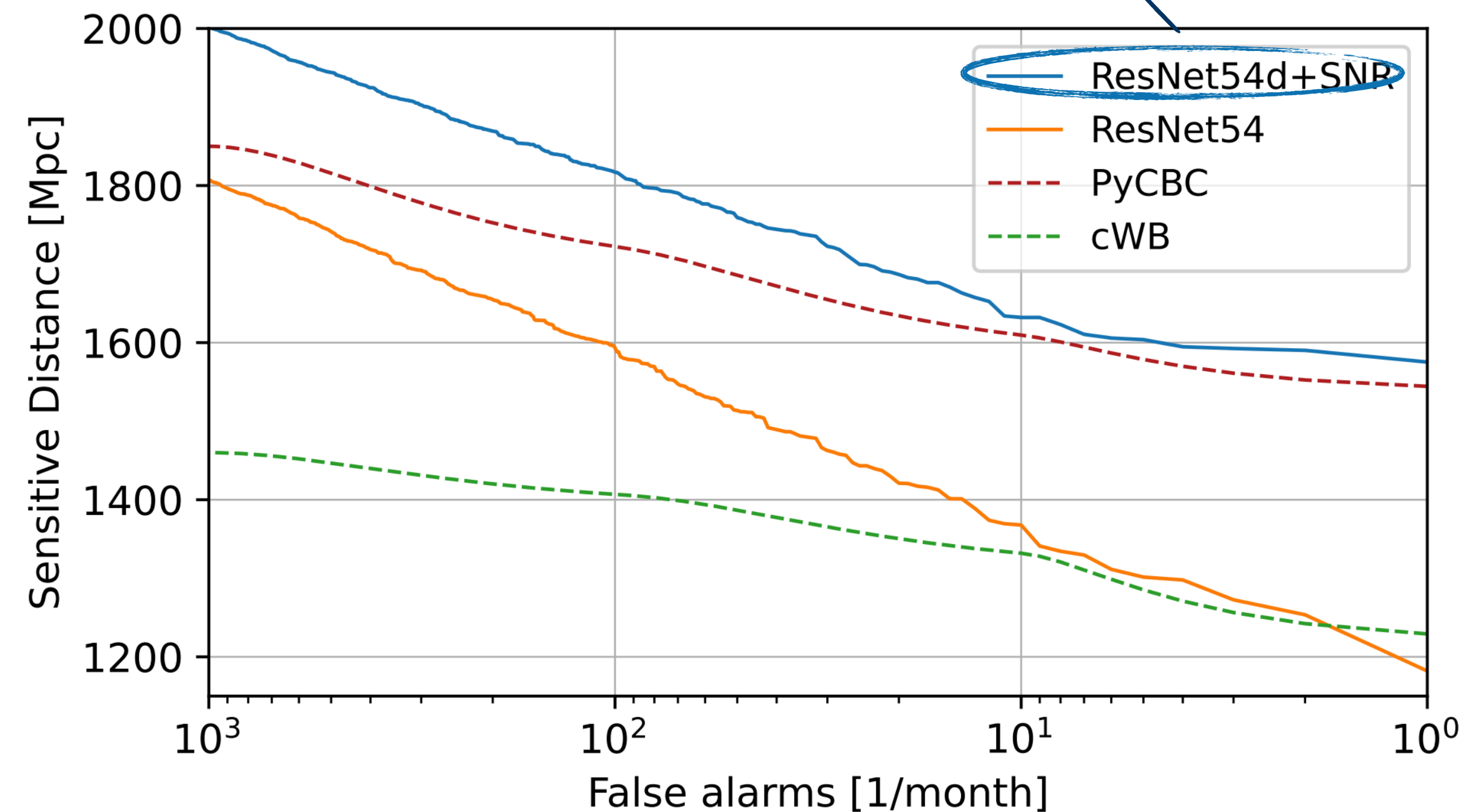


Results

Variance of different test datasets



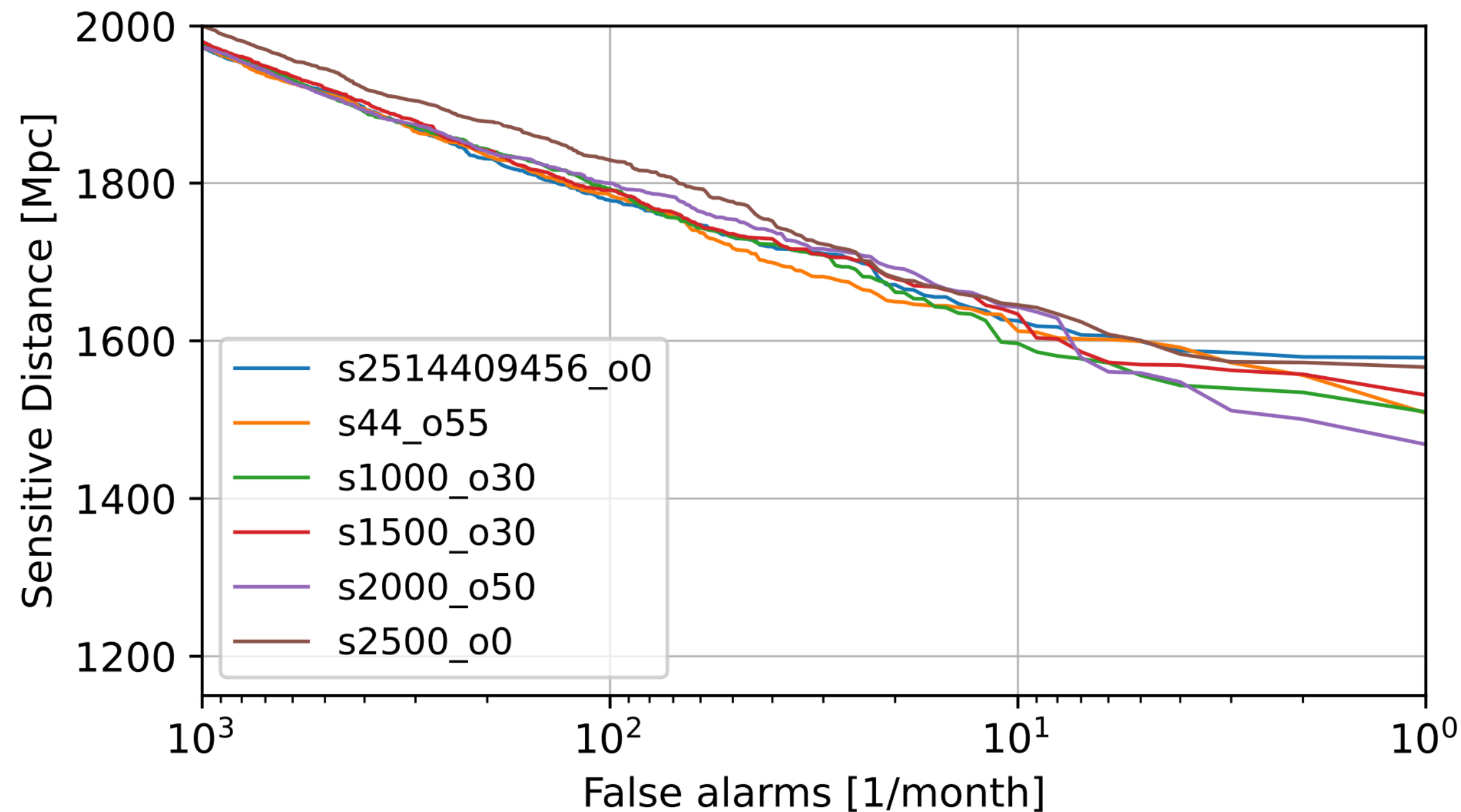
AresGW model 1 surpasses standard PyCBC in this set up



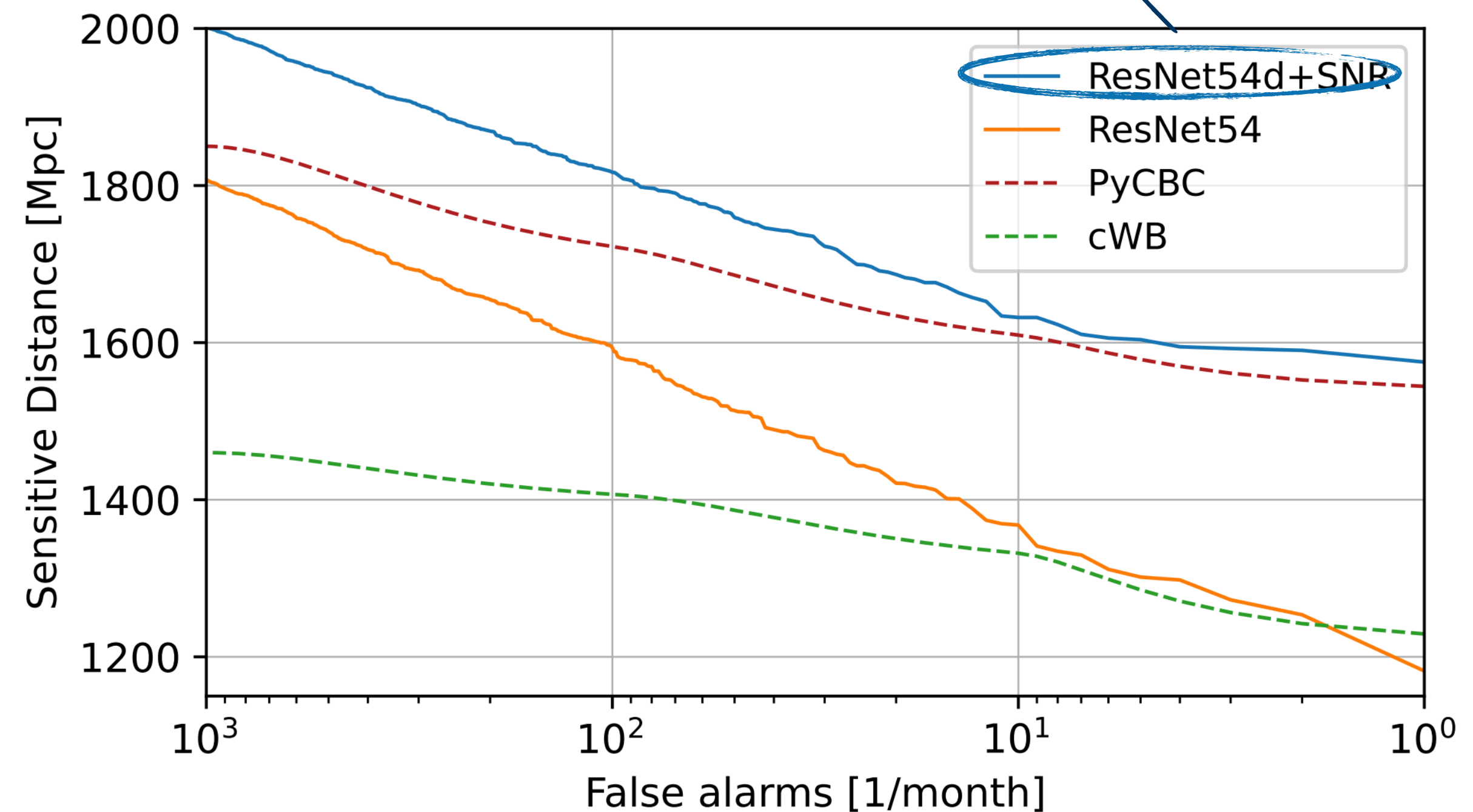
AresGW model 1: <https://github.com/vivinousi/gw-detection-deep-learning>

Results

Variance of different test datasets



AresGW model 1 surpasses standard PyCBC in this set up



AresGW model 1: <https://github.com/vivinousi/gw-detection-deep-learning>

How do sensitivity metrics fluctuate due to dataset variability?

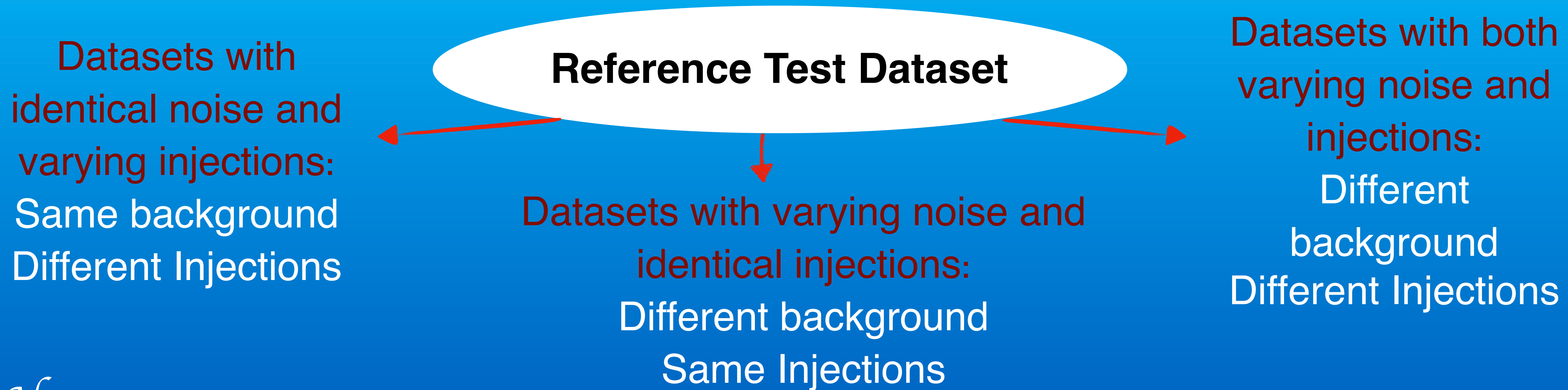
Robustness of Sensitivity Evaluations for Gravitational Wave Detection Algorithms

Alexandra E. Koloniari,¹ Lazaros Lazaridis,¹ Christos Paschalidis,¹ and Nikolaos Stergioulas¹

¹*Department of Physics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece*

(Dated: August 7, 2025)

One-month Test Datasets



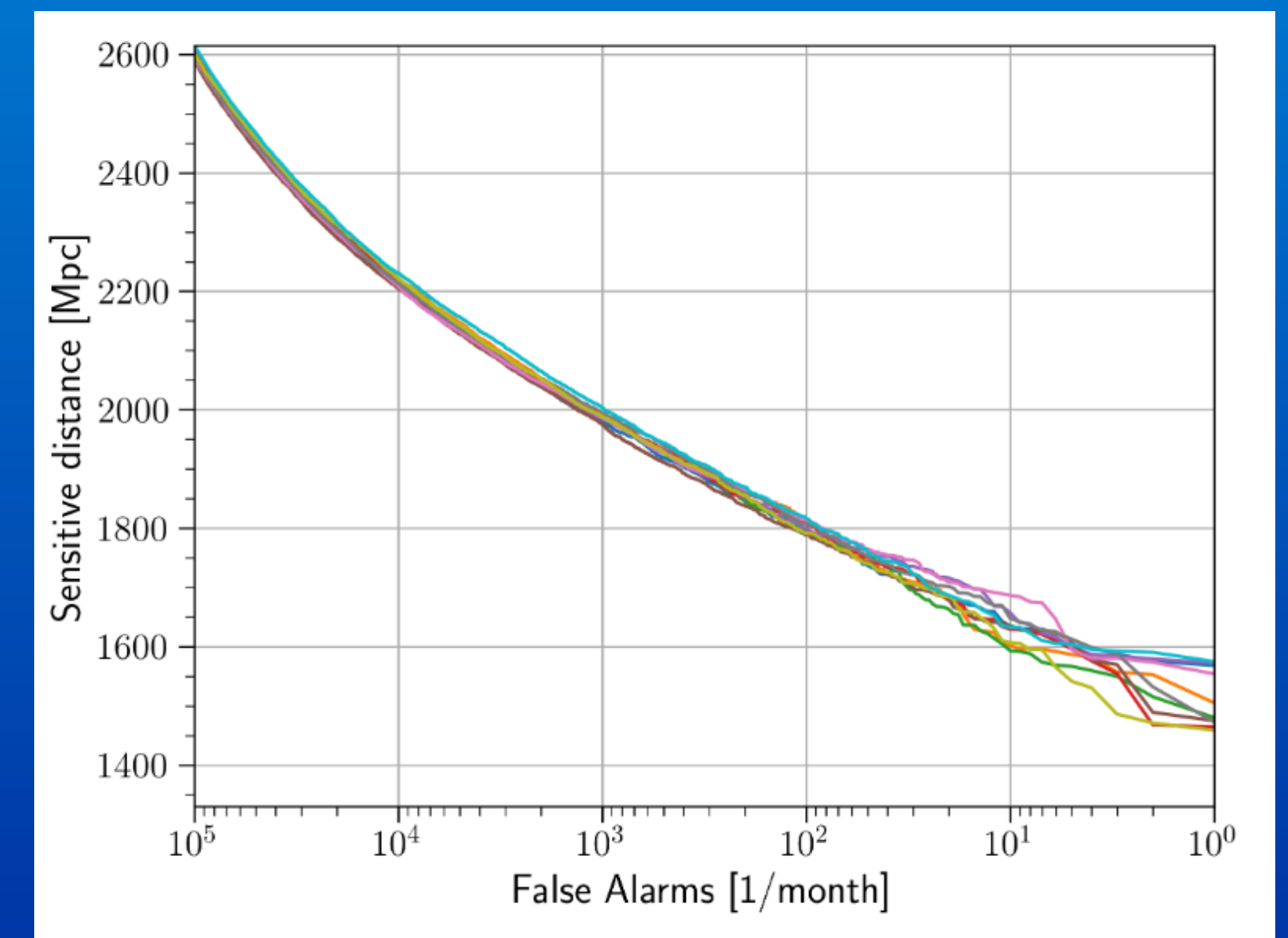
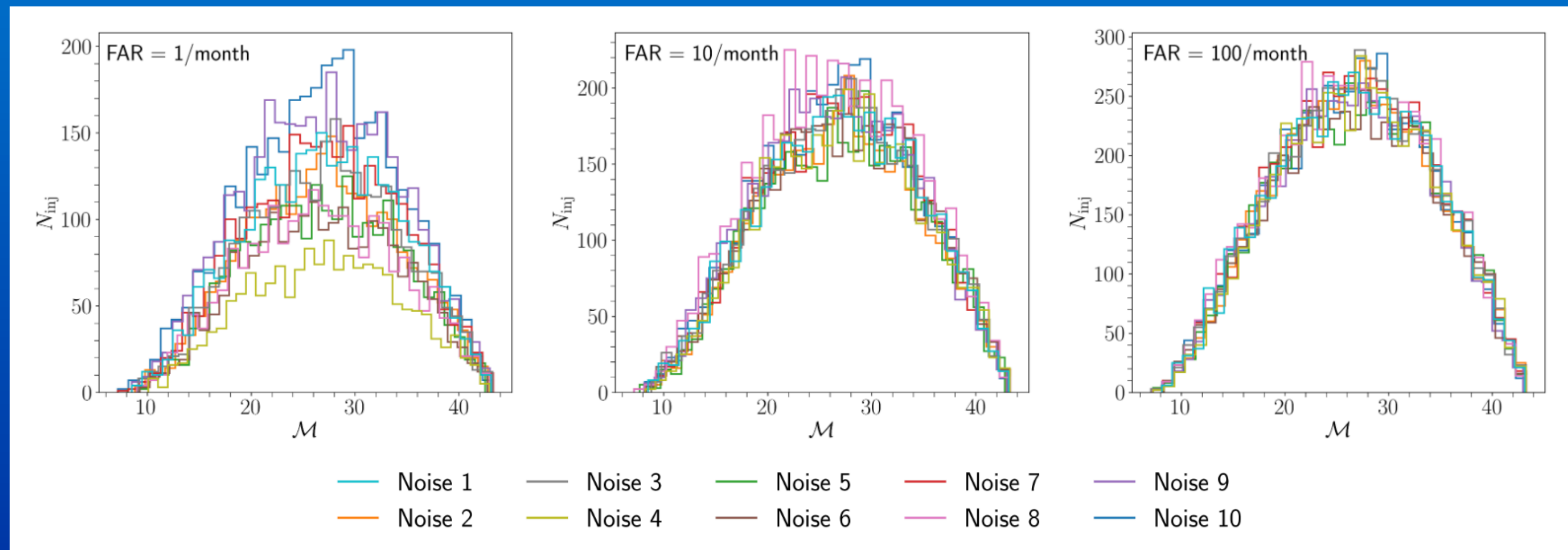
Notes:

- There are 28 test sets in total: $9 + 9 + 9 + 1$ reference set
- There are still 37 GW events from the AresGW, GWTC-2.1, IAS and OGC catalogs present in the data!

Test datasets: <https://gitlab.com/Alexandra1120/aresgw-variance>

Results

- $S_{1/\text{month}}$ is robust under single one-month evaluations.
- $N^F_{1/\text{month}}$ shows high variance when using a single one-month evaluation.
- $N^F_{100/\text{month}}$ is similarly robust with $S_{1/\text{month}}$, but this FAR is too high for credible detections.
- Including real GW events in noise biases N^F (and possibly other metrics).
- Robust evaluation requires: multiple datasets, real-event removal, uncertainty reporting, diverse metrics, and standardized protocols.



The Critical Test: Performance on Real Data

MACHINE
LEARNING
Science and Technology

PAPER

New gravitational wave discoveries enabled by machine learning

Alexandra E Koloniari^{1,*} , Evdokia C Koursoumpa¹ , Paraskevi Nousi² , Paraskevas Lampropoulos¹ , Nikolaos Passalis³ , Anastasios Tefas⁴  and Nikolaos Stergioulas¹ 

¹ Department of Physics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

² Swiss Data Science Center, ETH, Zürich, Switzerland

³ Department of Chemical Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

⁴ Department of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

* Author to whom any correspondence should be addressed.

E-mail: akolonia@auth.gr

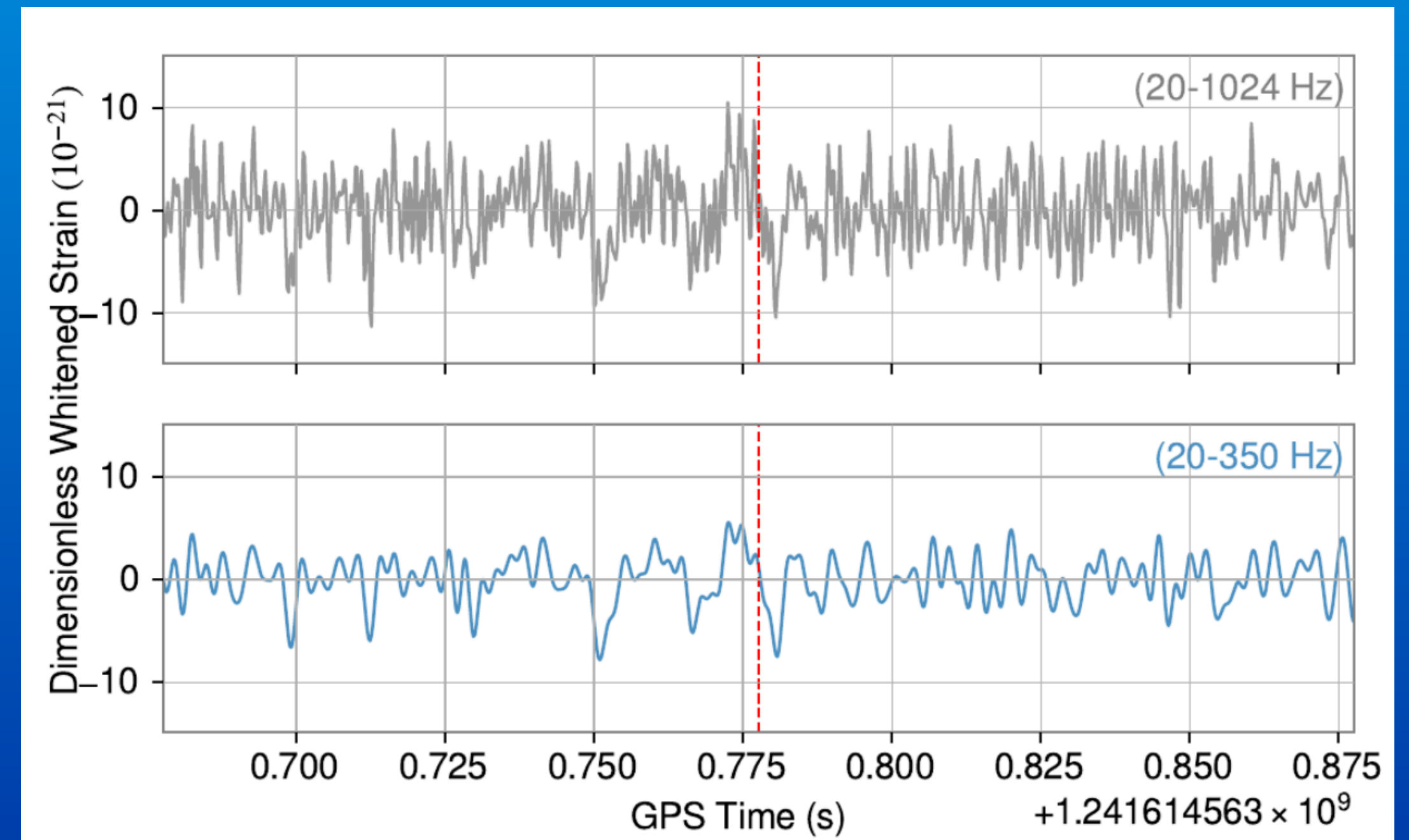
(Dated: January 28, 2025)

AresGW model 2: New Enhancements

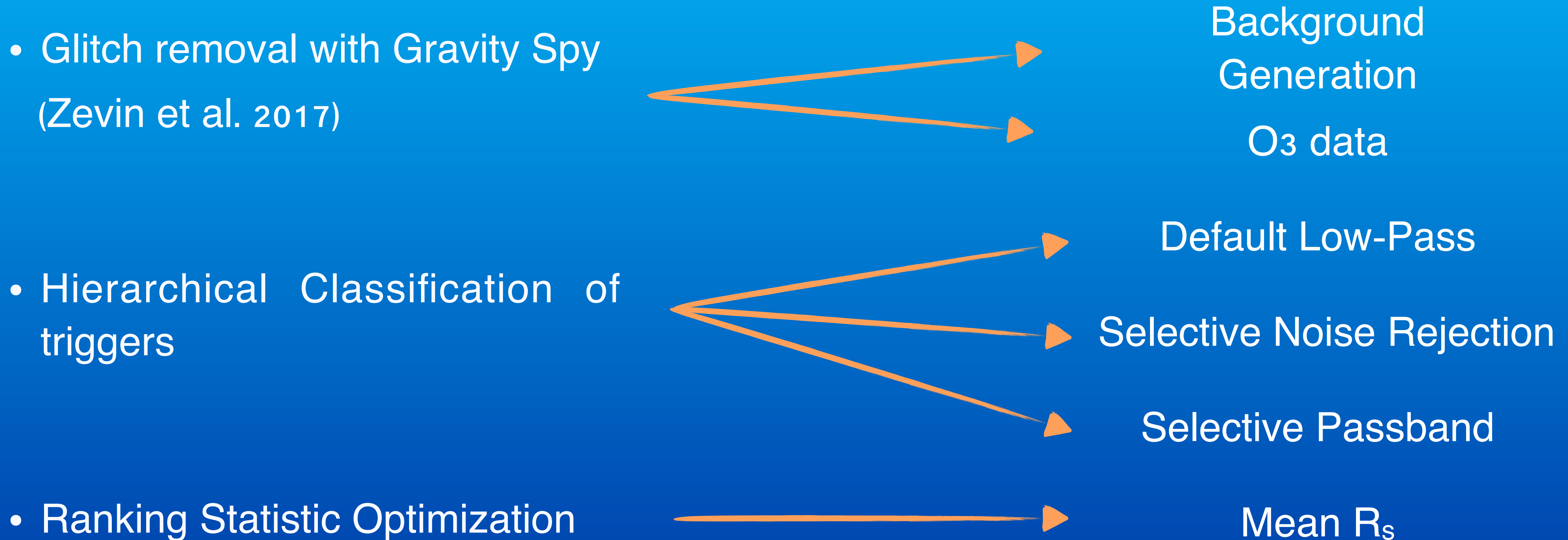
- Training Dataset Duration: 35 days for model 2 compared to 12 days for model 1
 - Double-precision floating point format (FP64) on only the final softmax layer
 - 350 Hz low-pass filter in both training and data analysis
- Only (on O3) data



$$\mathcal{R}_s = -\log_{10} (1 - \mathcal{R} + 10^{-16})$$



Data Analysis Methods



Data Analysis Methods

- Glitch removal with Gravity Spy
(Zevin et al. 2017)

Background
Generation
O3 data

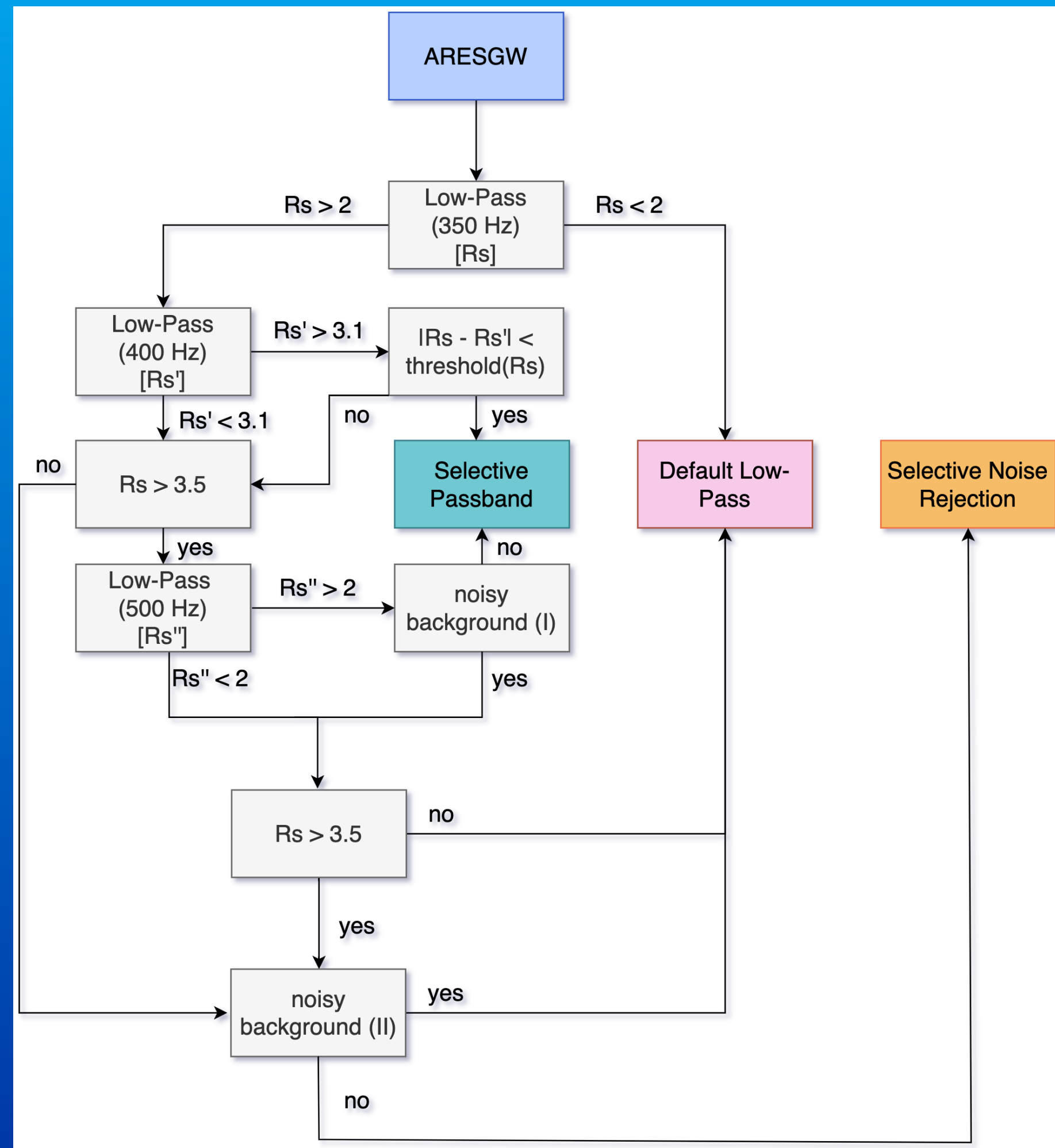
- Hierarchical Classification of triggers

Default Low-Pass
Selective Noise Rejection
Selective Passband

- Ranking Statistic Optimization

Mean R_s

Hierarchical Classification of Triggers



Default Low-Pass

Selective Noise Rejection

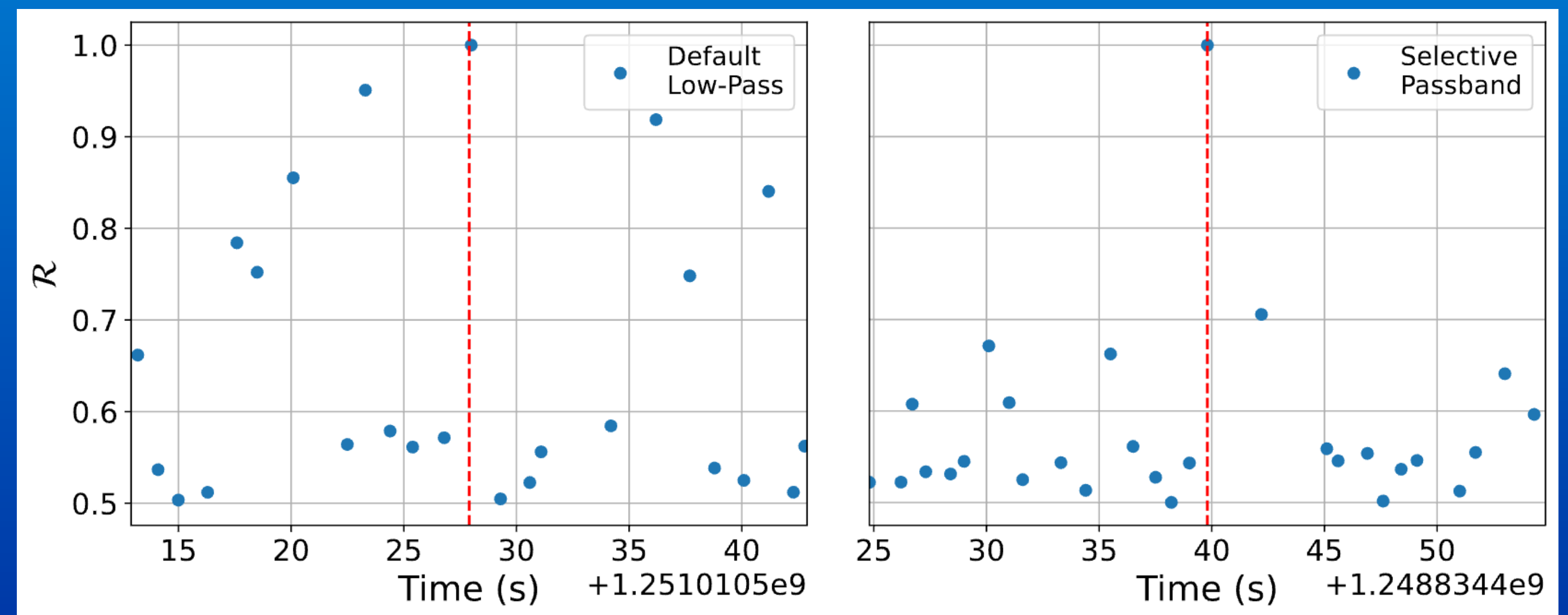
Selective Passband



Reduction of false alarms
by 61%

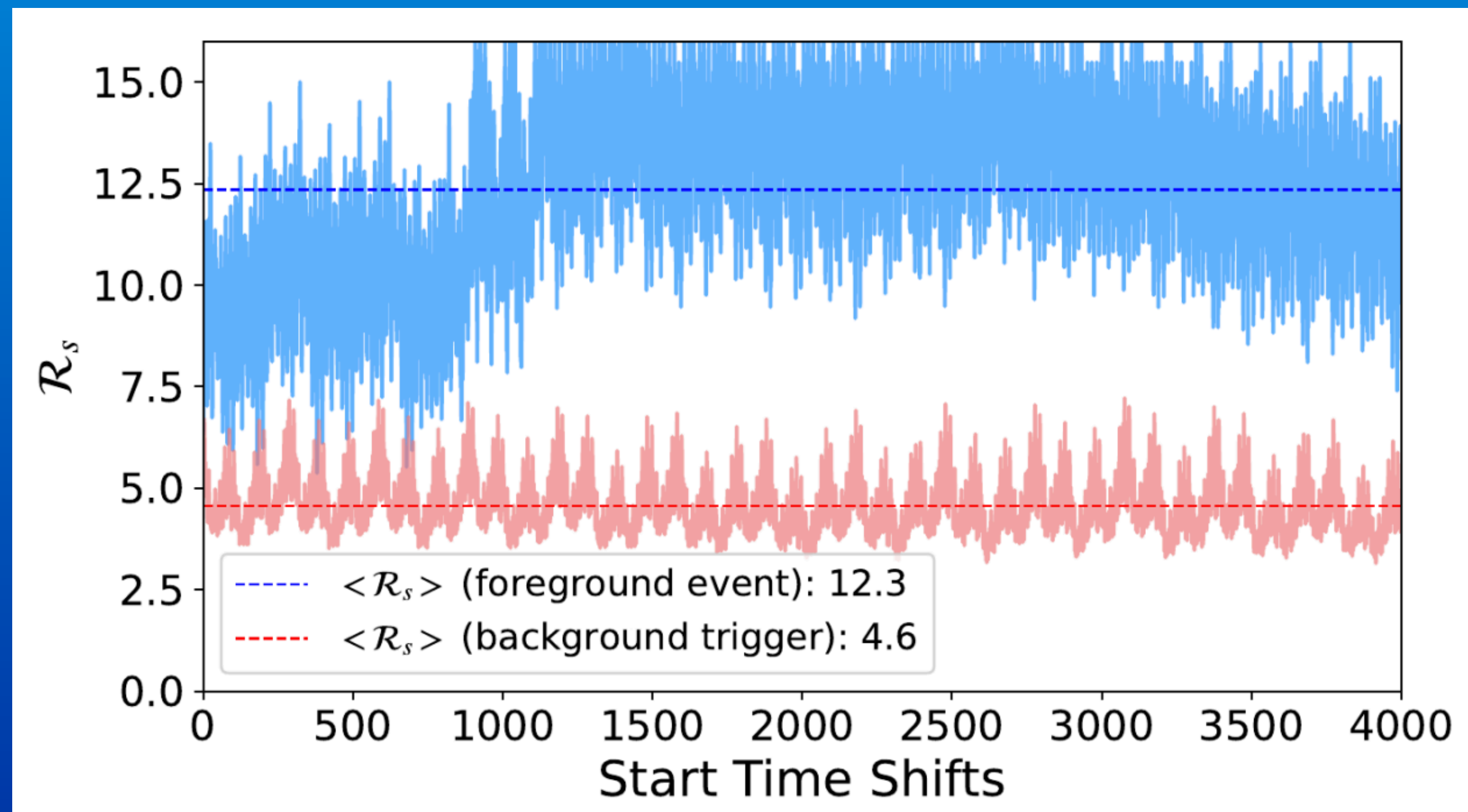


Reduction of false alarms
by 90%



Ranking Statistic Optimization

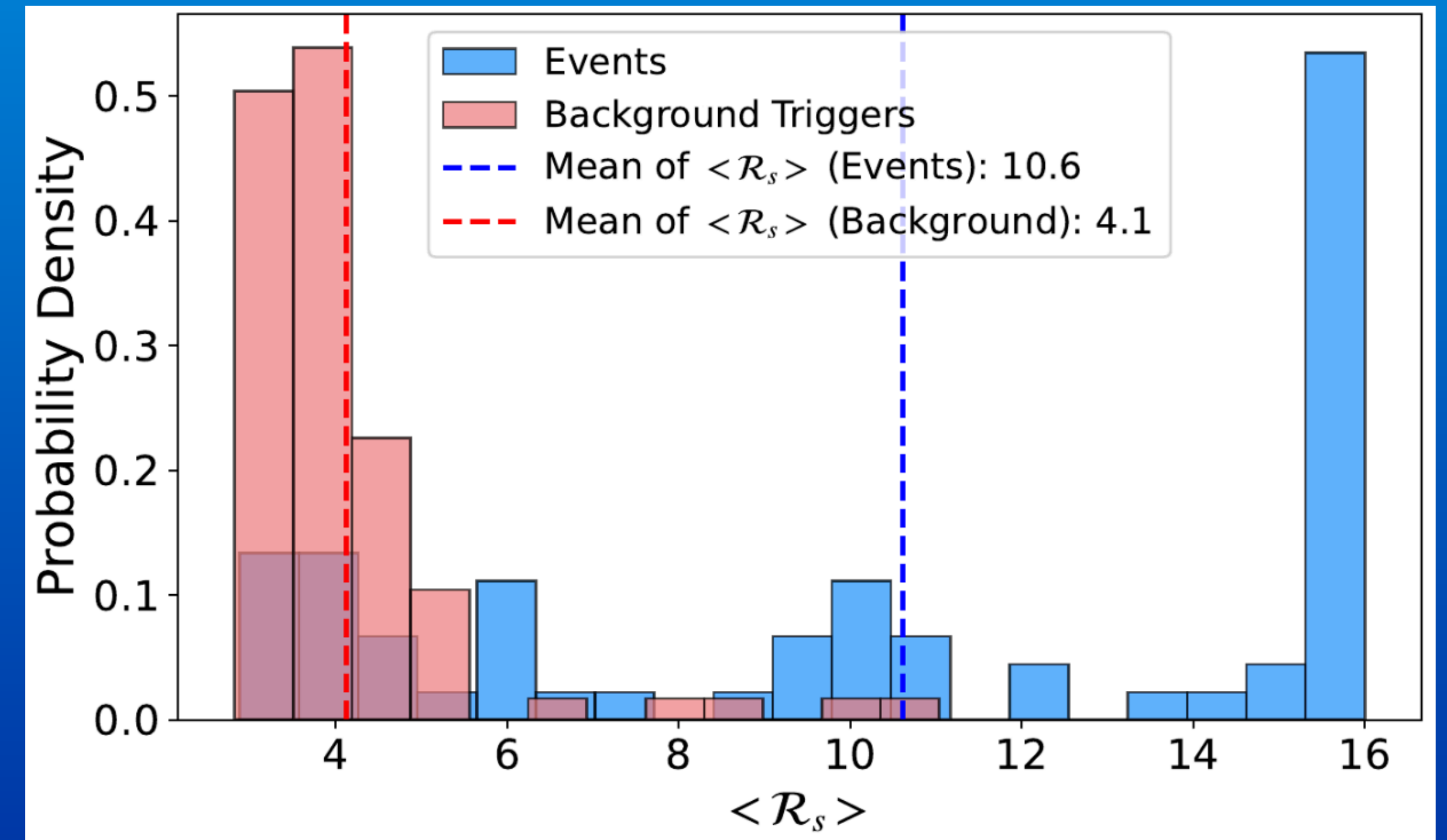
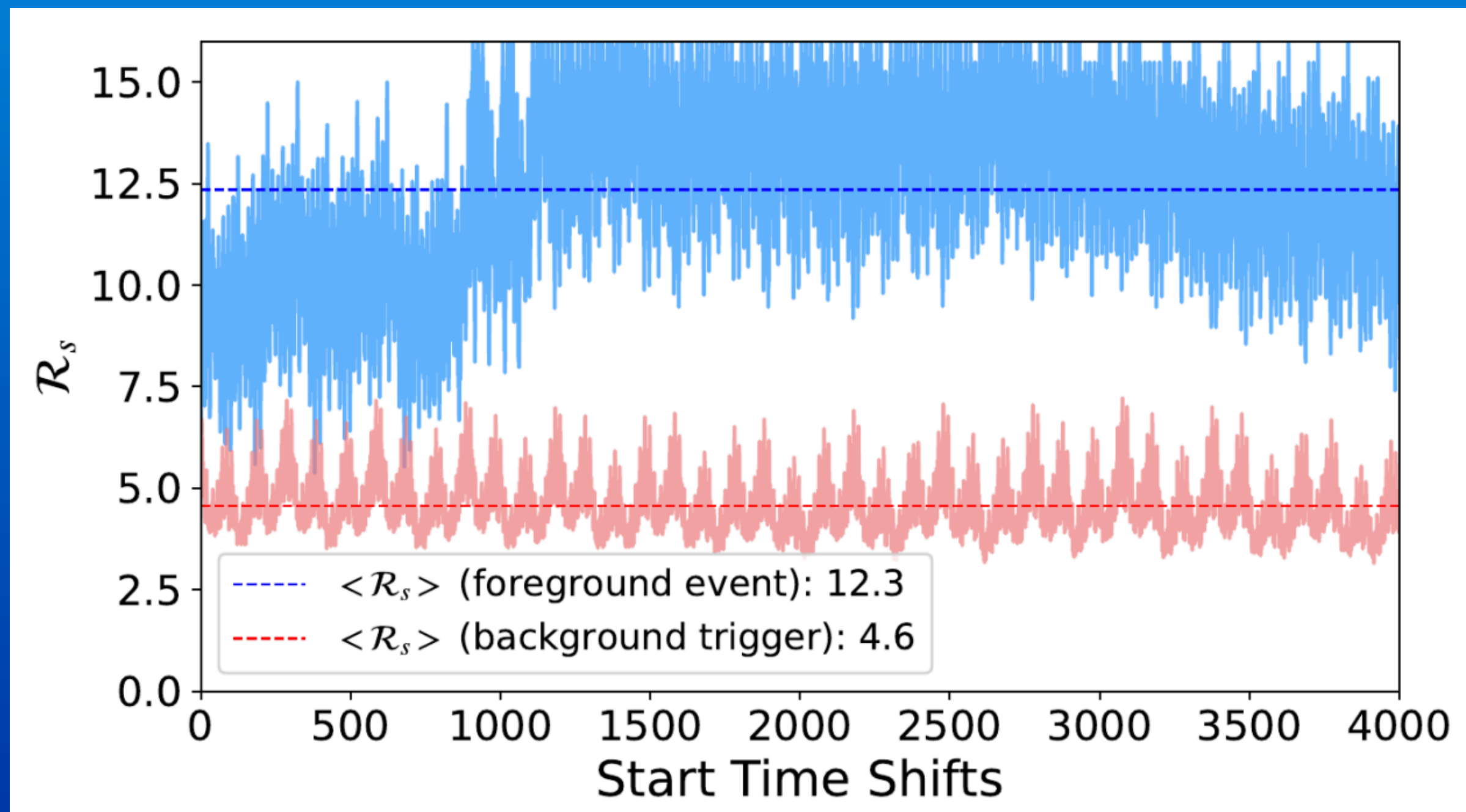
The light blue time series represents the \mathcal{R}_s with the shifted start time of an event, while the pink time series depicts a representative noise trigger



Ranking Statistic Optimization

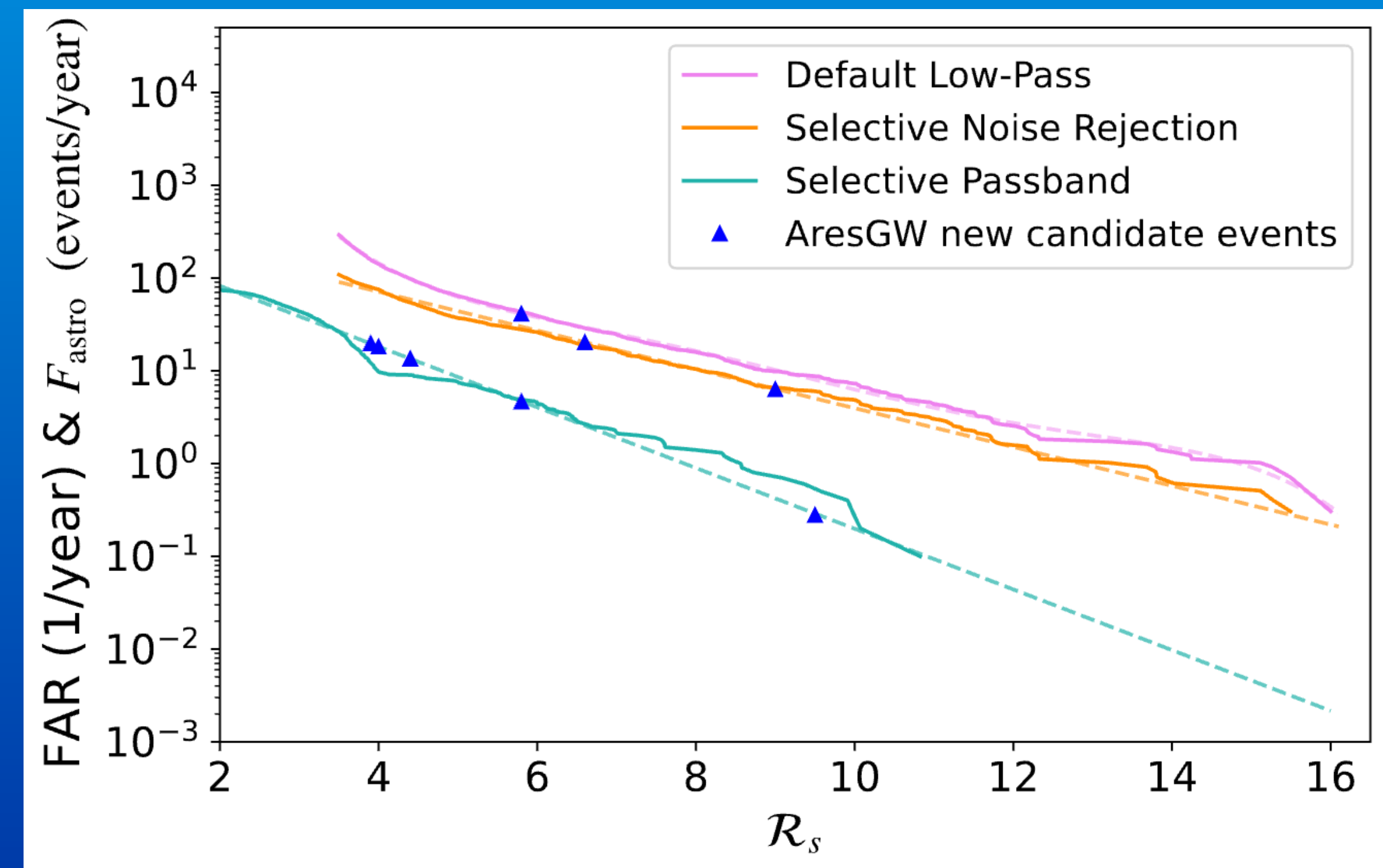
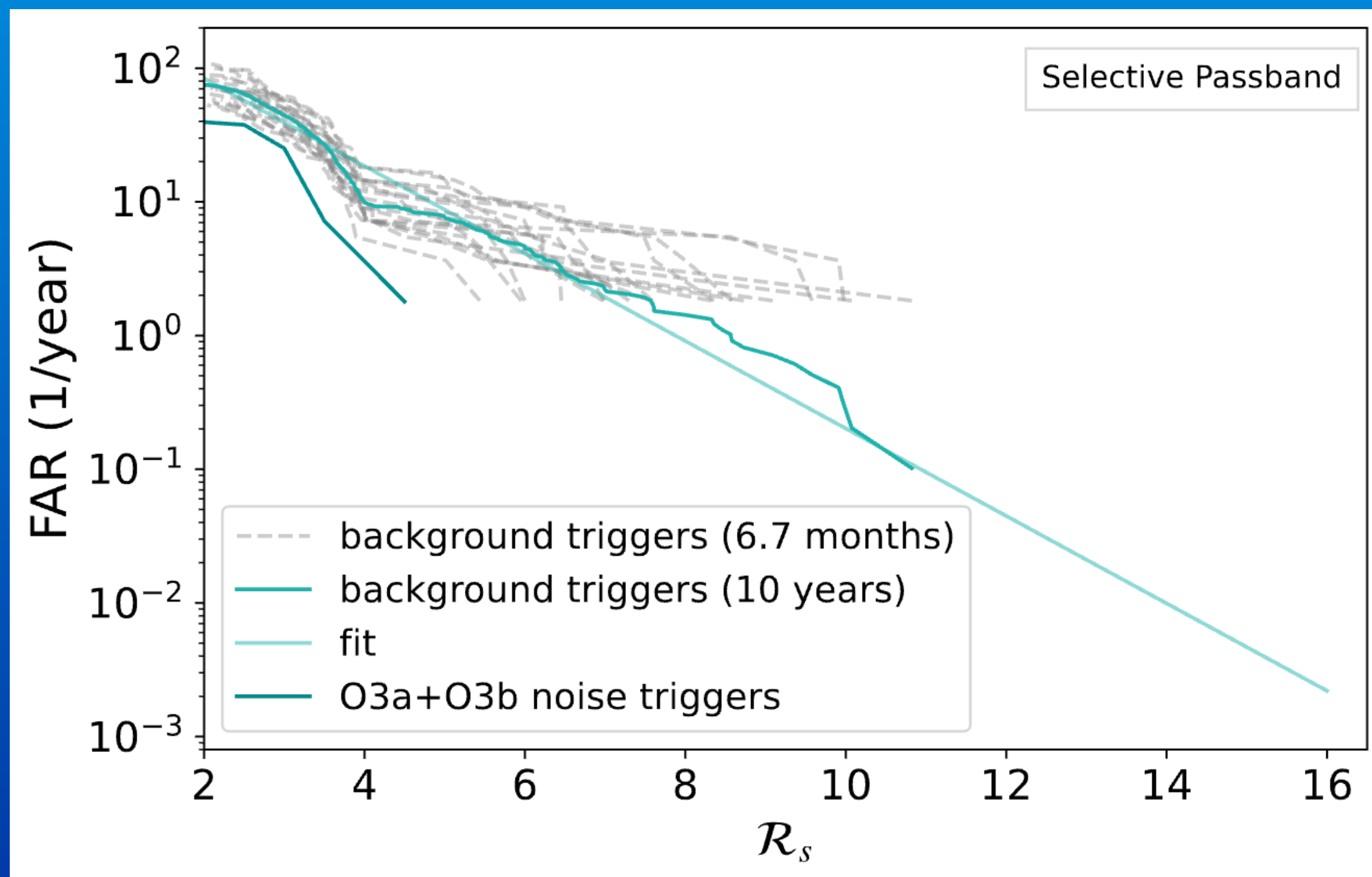
The light blue time series represents the \mathcal{R}_s with the shifted start time of an event, while the pink time series depicts a representative noise trigger

Histograms of the $\langle \mathcal{R}_s \rangle$ for the foreground events (light blue) and background triggers (pink)



Background Statistics (FAR)

81 days of data-quality O3a noise → Glitch removal → Different Time Shifts → Hierarchical Classification of Triggers



New Candidate Events (I)

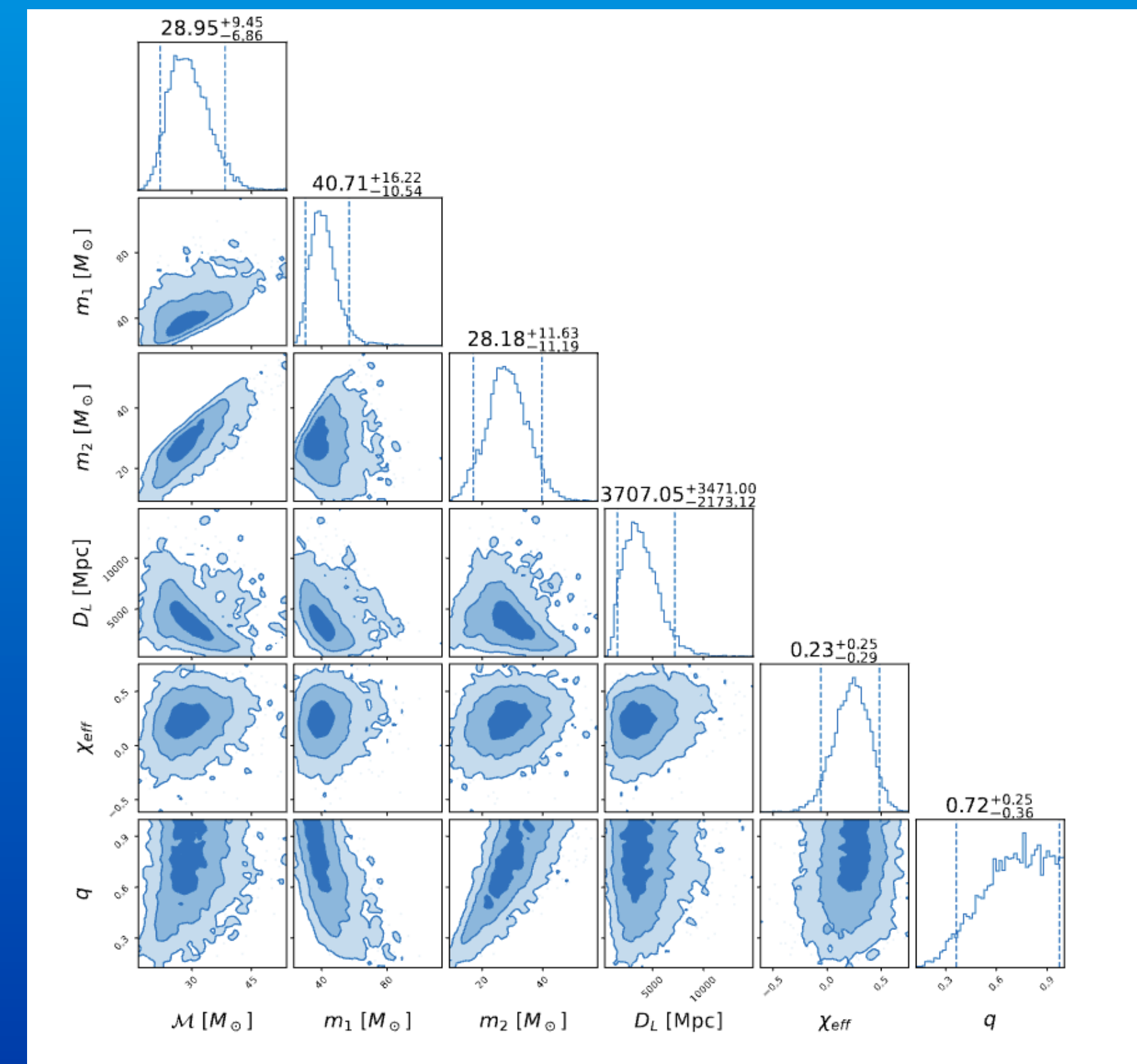
- Consistency tests

Arrival
time < 10 ms

χ^2 test

$$\hat{\rho} = \rho \times \begin{cases} 1 & \text{if } \chi_r^2 \leq \nu, \\ \left[\frac{1}{2} + \frac{1}{2} (\chi_r^2 / \nu)^3 \right]^{-1/6} & \text{if } \chi_r^2 > \nu, \end{cases}$$

- PE with Bayesian inference library (Bilby-Ashton et al. 2019)



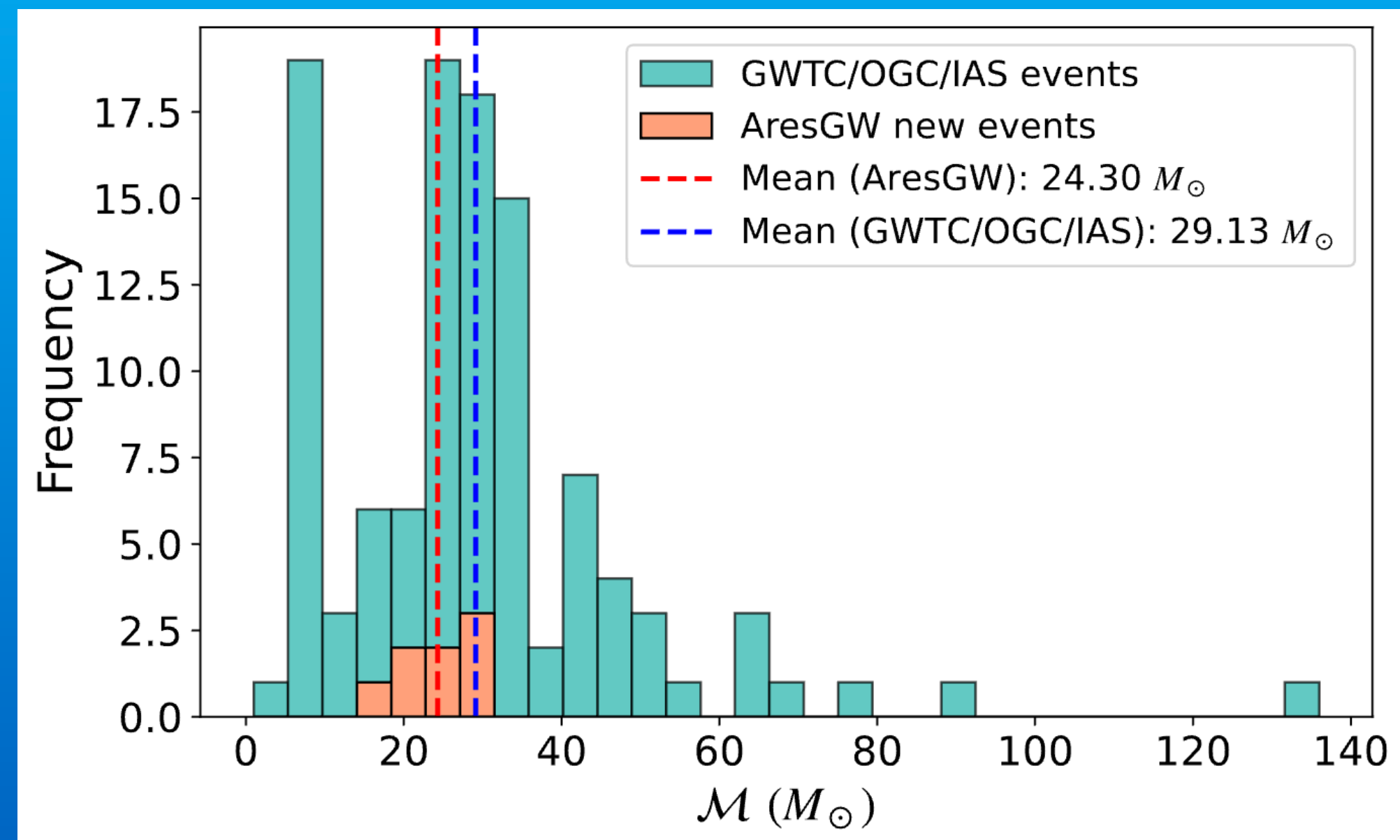
New Candidate Events (II)

Cumulative astrophysical probability of 5.94!

#	Event Name	GPS Time (s)	p_{astro}	FAR (1/yr)	$\langle \mathcal{R}_s \rangle$	Time delay (s)	χ_L^2	χ_H^2	Class
1	GW190511_125545	1241614563.77	1.00	0.27	9.54	0.0027	1.16	1.46	Selective Passband
2	GW190614_134749	1244555287.93	0.99	4.6	5.80	0.0012	0.65	0.80	Selective Passband
3	GW190607_083827	1243931925.99	0.99	6.5	8.95	0.0056	0.90	0.48	Selective Noise Rejection
4	GW190904_104631	1251629209.01	0.72	14	4.35	0.0002	0.38	0.71	Selective Passband
5	GW190523_085933	1242637191.44	0.68	20	6.60	0.0054	0.75	1.39	Selective Noise Rejection
6	GW200208_211609	1265231787.68	0.55	18	4.0	0.0063	0.69	0.98	Selective Passband
7	GW190705_164632	1246380410.88	0.51	49	5.82	0.0103	1.05	0.98	Default Low-Pass*
8	GW190426_082124	1240302101.93	0.50	20	3.91	0.0007	1.48	0.53	Selective Passband

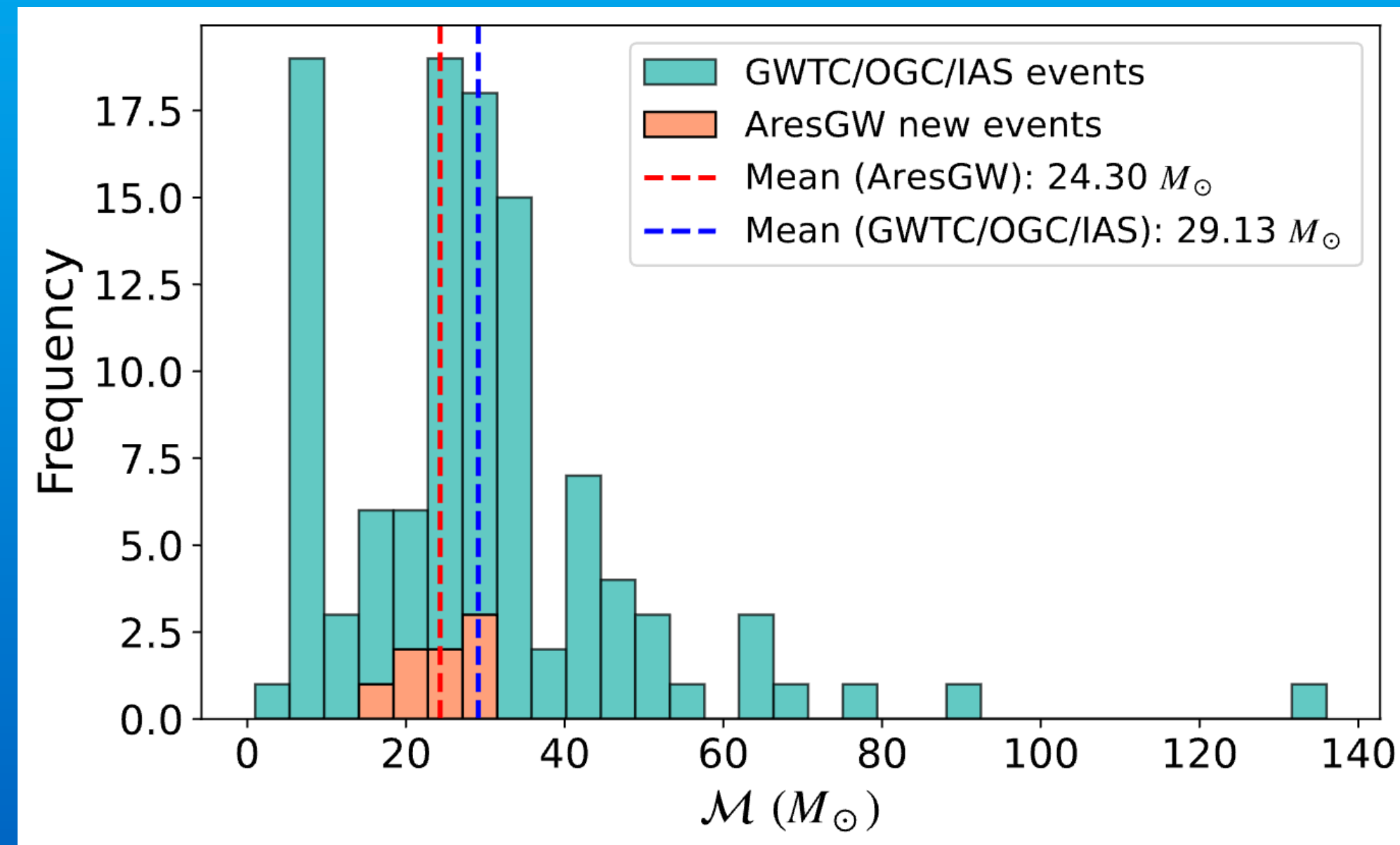
All of our eight new GW detections were
subsequently verified by an independent parameter estimation study
(Beyond GWTC-3 - Williams 2025)

Population properties of new candidate events



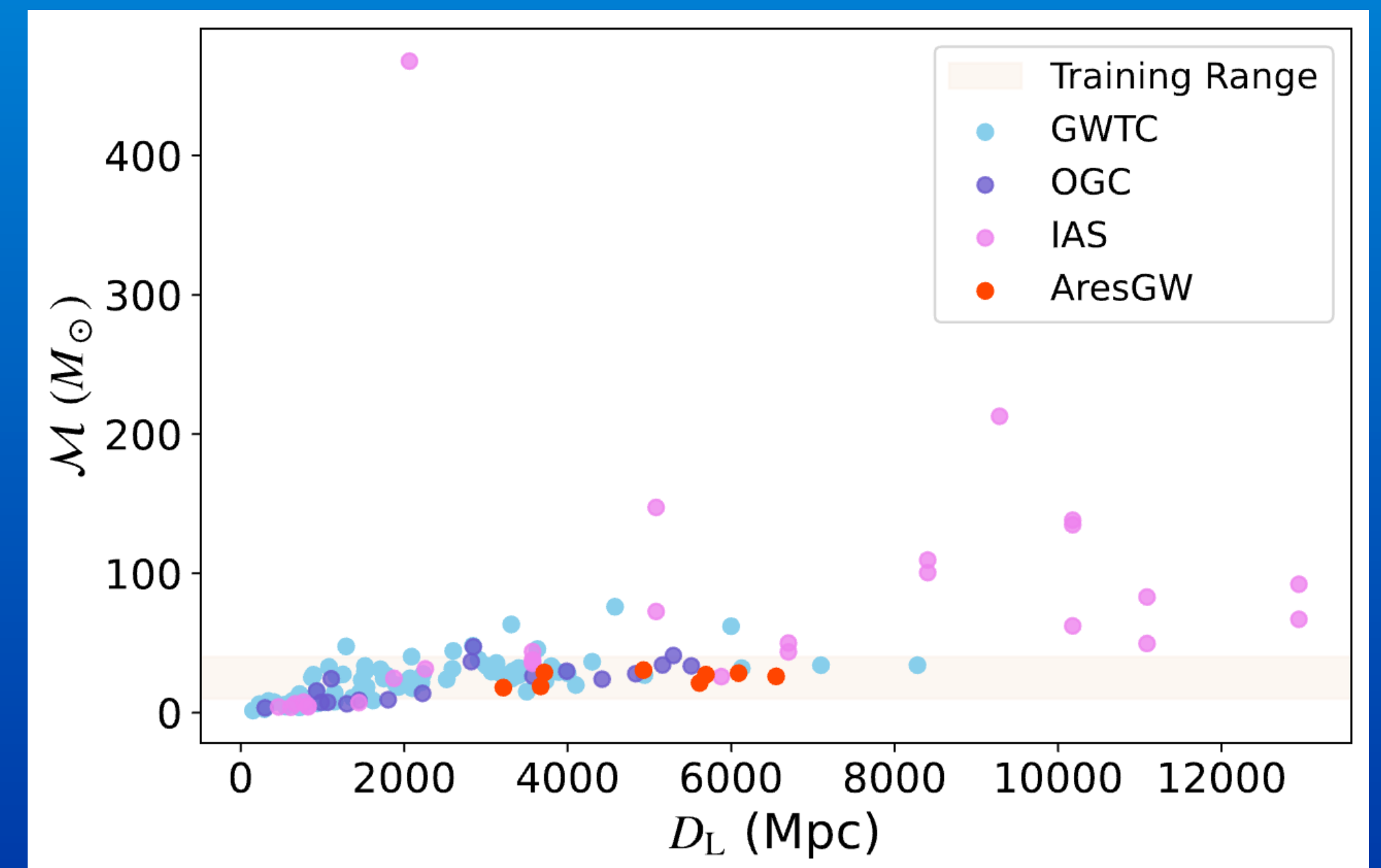
Our distribution aligns with that from other catalogs

Population properties of new candidate events

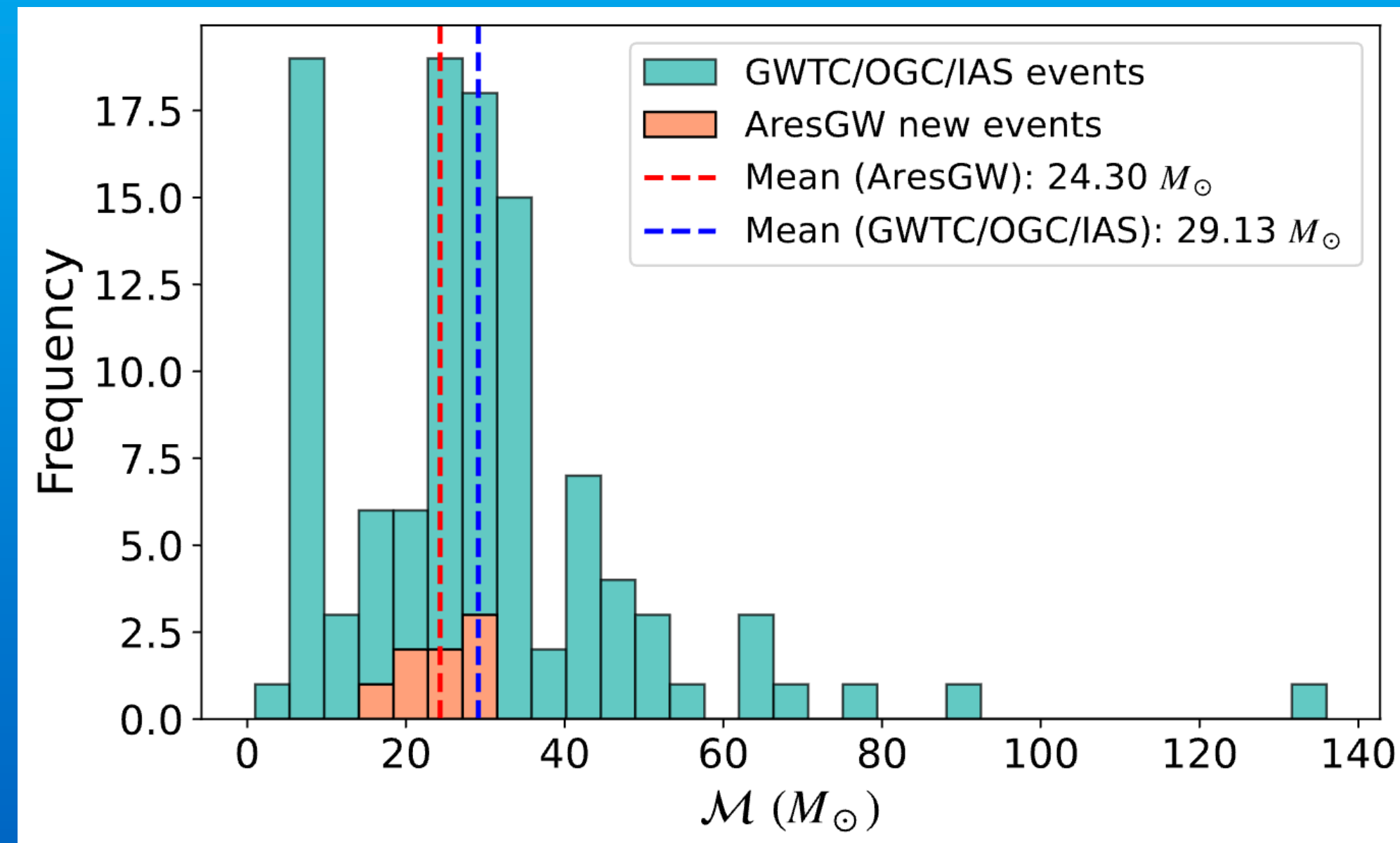


Our distribution aligns with that from other catalogs

Our new events tend to exhibit higher luminosity distances compared to the majority of the previously published confirmed events

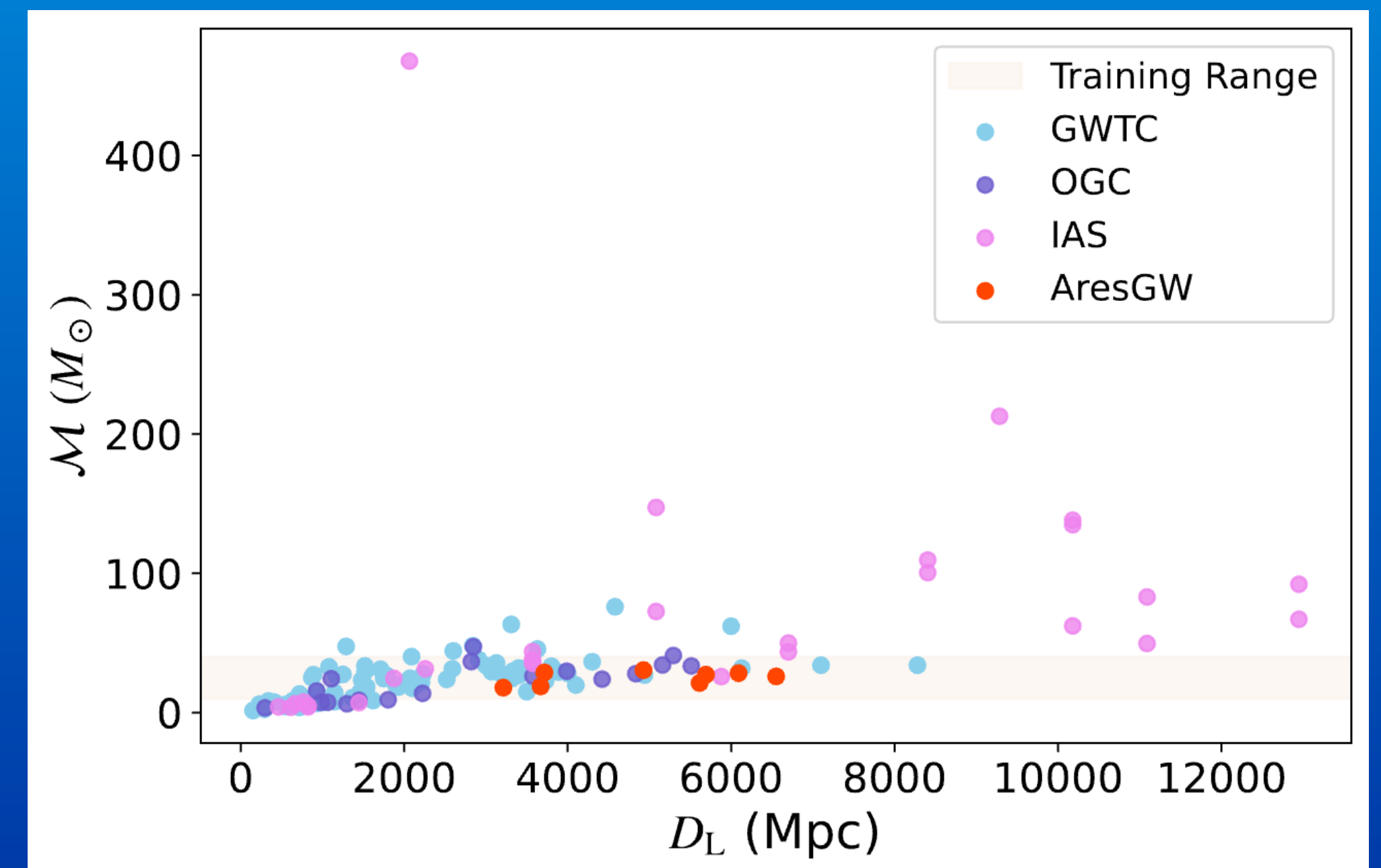


Population properties of new candidate events



Our distribution aligns with that from other catalogs

Our new events tend to exhibit higher luminosity distances compared to the majority of the previously published confirmed events



Known Events

43 published gravitational wave events (GWTC / OGC / IAS) *within our effective training range:*

AresGW model 2:

34/43 were confirmed with $p_{\text{astro_AresGW}} > 0.5$

9/43 candidate events were reported with 

$p_{\text{astro_AresGW}} < 0.5$

$p_{\text{astro_IAS}} = 0.63, 0.56$

$p_{\text{astro_GWTC}} = 0.54$

$p_{\text{astro_OGC}} = 0.5$

Known Events

43 published gravitational wave events (GWTC / OGC / IAS) *within our effective training range:*

AresGW model 2:

34/43 were confirmed with $p_{\text{astro_AresGW}} > 0.5$

9/43 candidate events were reported with 

$p_{\text{astro_AresGW}} < 0.5$

$p_{\text{astro_IAS}} = 0.63, 0.56$

$p_{\text{astro_GWTC}} = 0.54$

$p_{\text{astro_OGC}} = 0.5$

55 published gravitational wave events (GWTC / OGC / IAS) *outside of our effective training range:*

AresGW model 2:

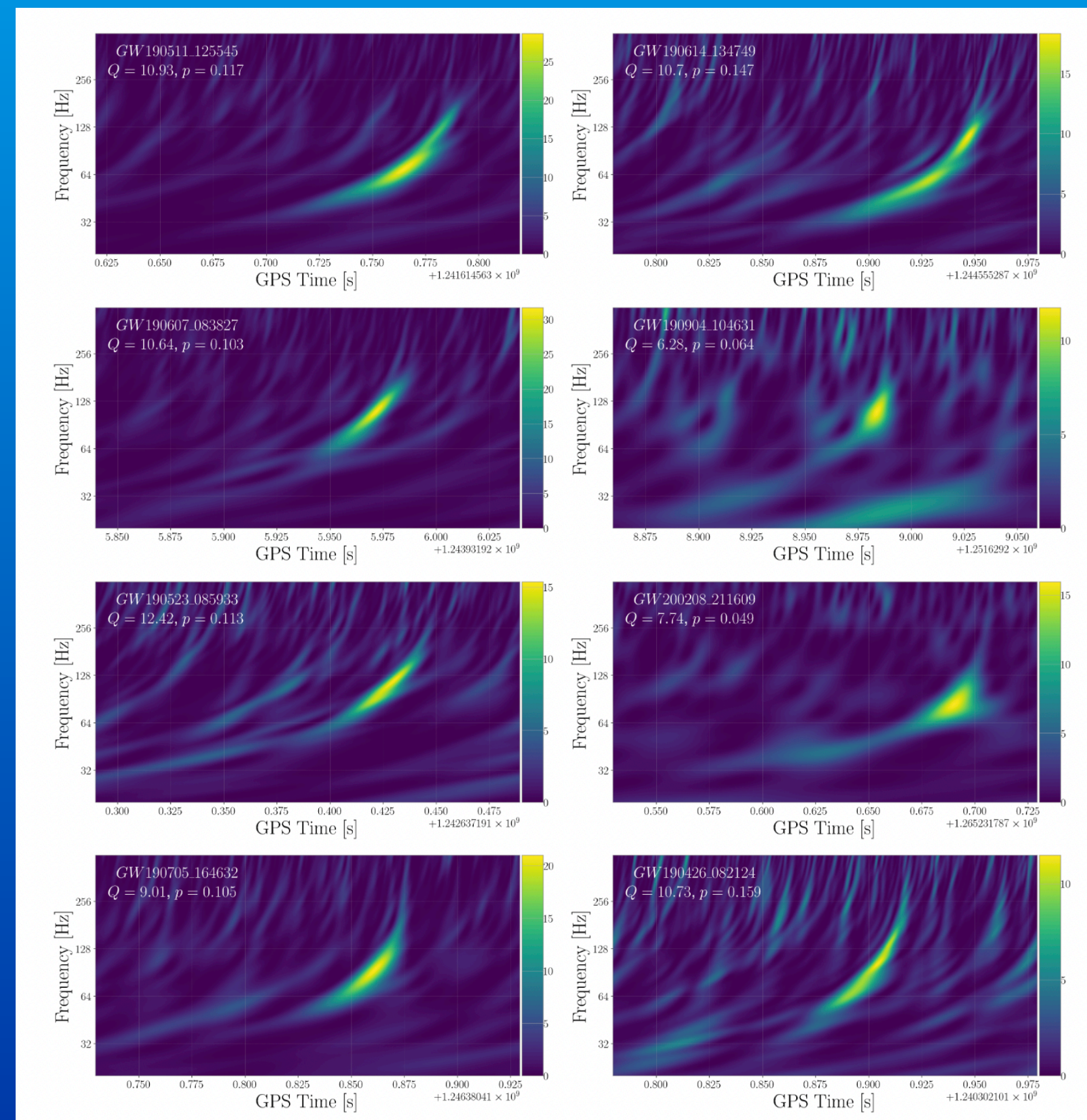
10/55 were confirmed with $p_{\text{astro_AresGW}} > 0.5$

Conclusions

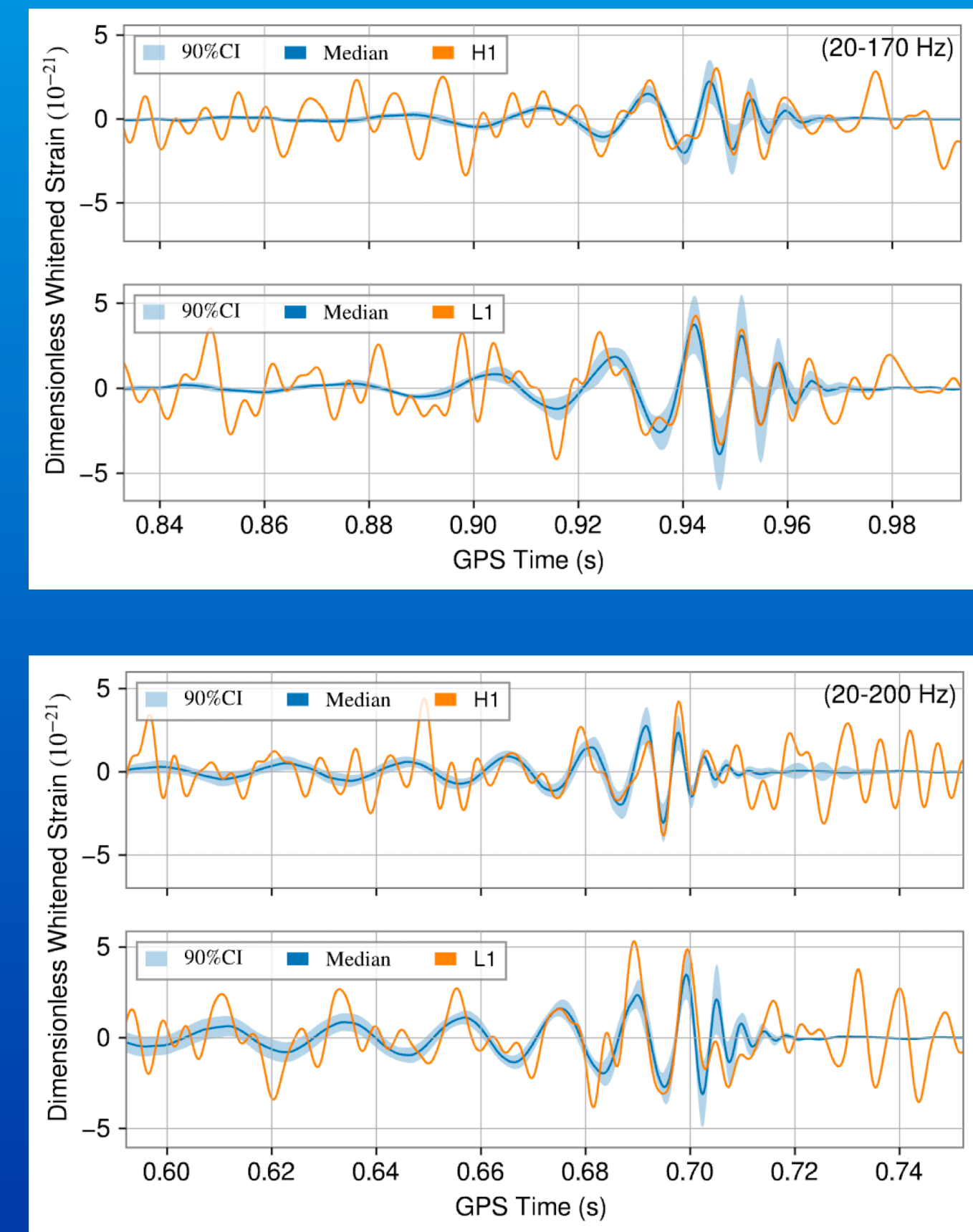
AresGW model 2:

- Detected 34/43 events within its effective training range
- Detected 10/55 events outside of its effective training range
- ◆ *Detected 8 new gravitational wave events!*

Qp plots of all 8 new events:



Time series examples for 2 new candidate events:



Future Directions

For AresGW:

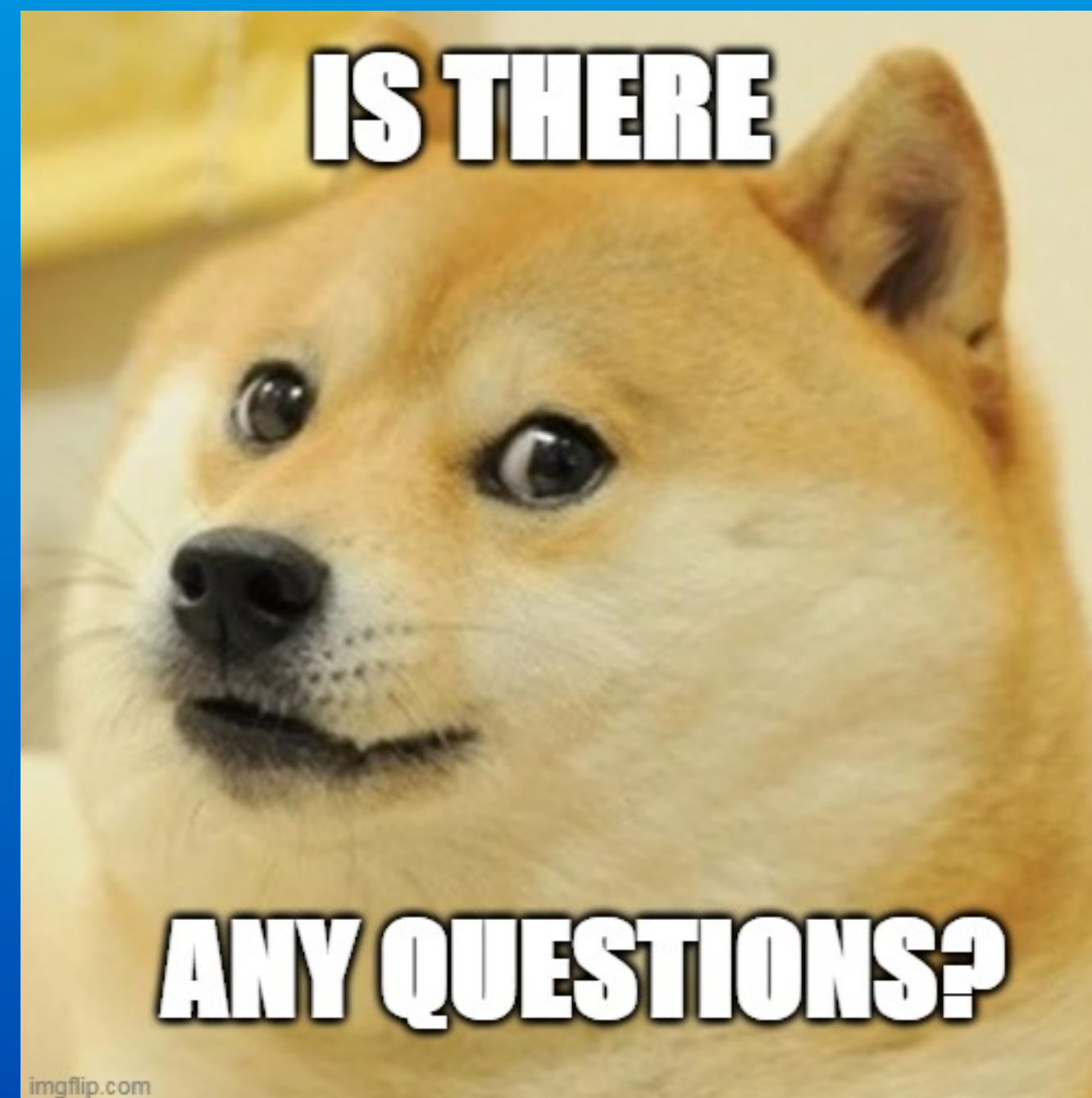
- Analyze O1, O2 and O4 data (with and without retraining)
- Try different combination of detectors (single, double, and triple detector setups)
- Extend AresGW for the offline detection of BBH mergers in other mass ranges, BNS mergers etc.

Broader Outlook for GW

Detection with ML:

- *Build shared datasets and metrics for fair comparisons*
- Combine matched filtering and ML for hybrid pipelines

THANK YOU!



Sensitive distance

The formula we use to estimate the sensitive volume $V(F)$ of a search algorithm is:

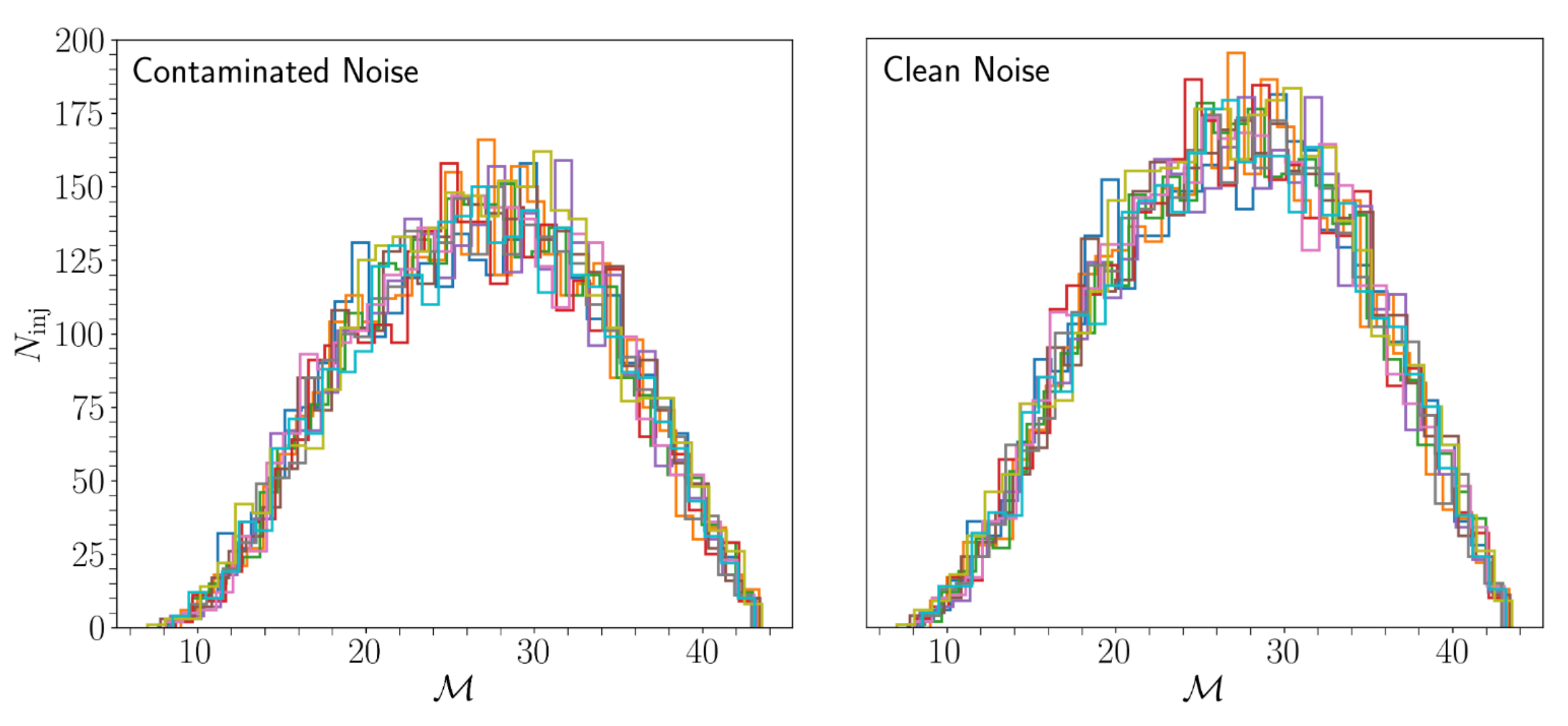
$$V(\mathcal{F}) = \frac{V(d_{\max})}{N} \sum_{i=1}^{N_{\text{inj}, \mathcal{F}}} \left(\frac{M_{c,i}}{M_{\max}} \right)^{5/2}$$

where $M_{c,i}$ is the chirp mass of the i -th found injection with FAR F and d_{\max} , $M_{c,\max}$ are the maximum injection distance and chirp mass, respectively, from the set of signals injected into the data.

Performance Evaluation on Datasets without Contamination by Astrophysical Signals

Datasets with identical clean noise and varying injections:

FAR = 1/month:



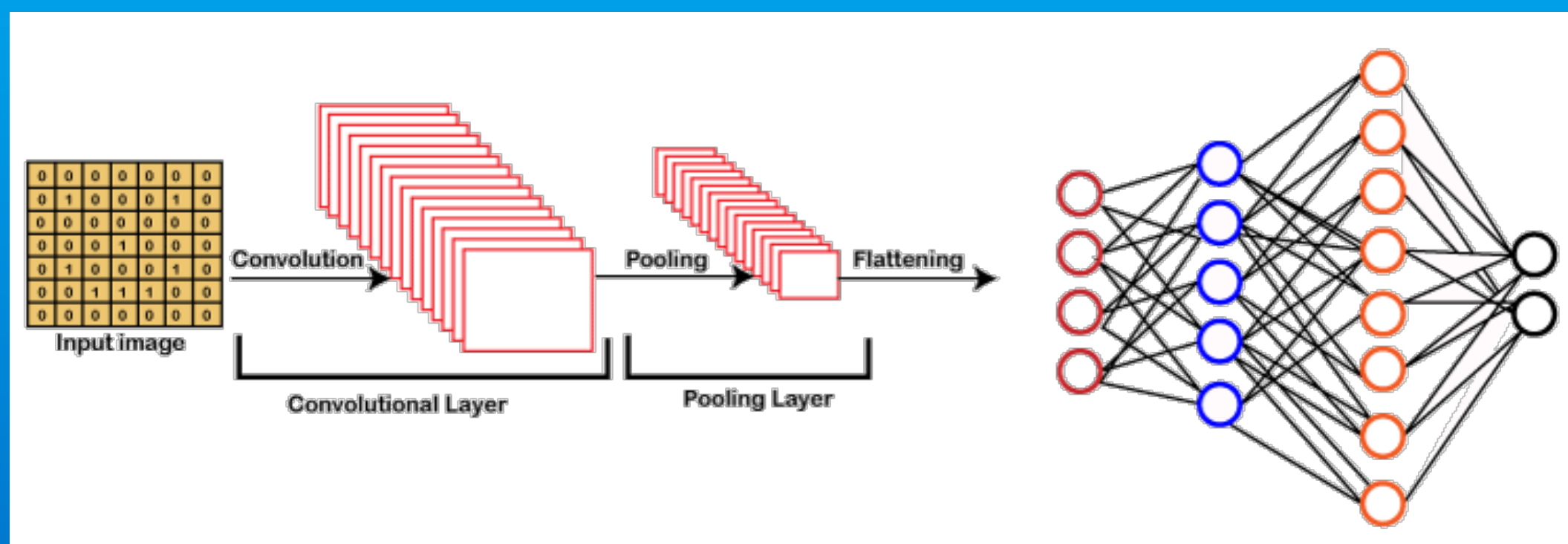
#	Injection Seed	N^F 1/month	S (Mpc) 1/month
1	2514409456	3475	1590.60
2	12019	3504	1600.20
3	10209	3518	1572.03
4	9801	3494	1587.90
5	6291	3433	1586.44
6	555	3458	1600.90
7	291	3474	1598.34
8	93	3395	1587.06
9	32	3506	1587.90
10	9	3454	1587.50
μ (mean)		3471 ± 27	1589.9 ± 6.1
σ (std. dev.)		$\in [25.9, 68.7]$	$\in [5.9, 15.5]$
$\sigma_N / \bar{\mu}_N$		$\in [0/7\%, 2.0\%]$	$\in [0.4\%, 1.0\%]$

After removing real GW events,
 $N^F_{1/month}$ increased by $\sim 19.2\%$!

After removing real GW events,
 $S_{1/month}$ increased by only $\sim 1.0\%$.

Popular Neural Network Types

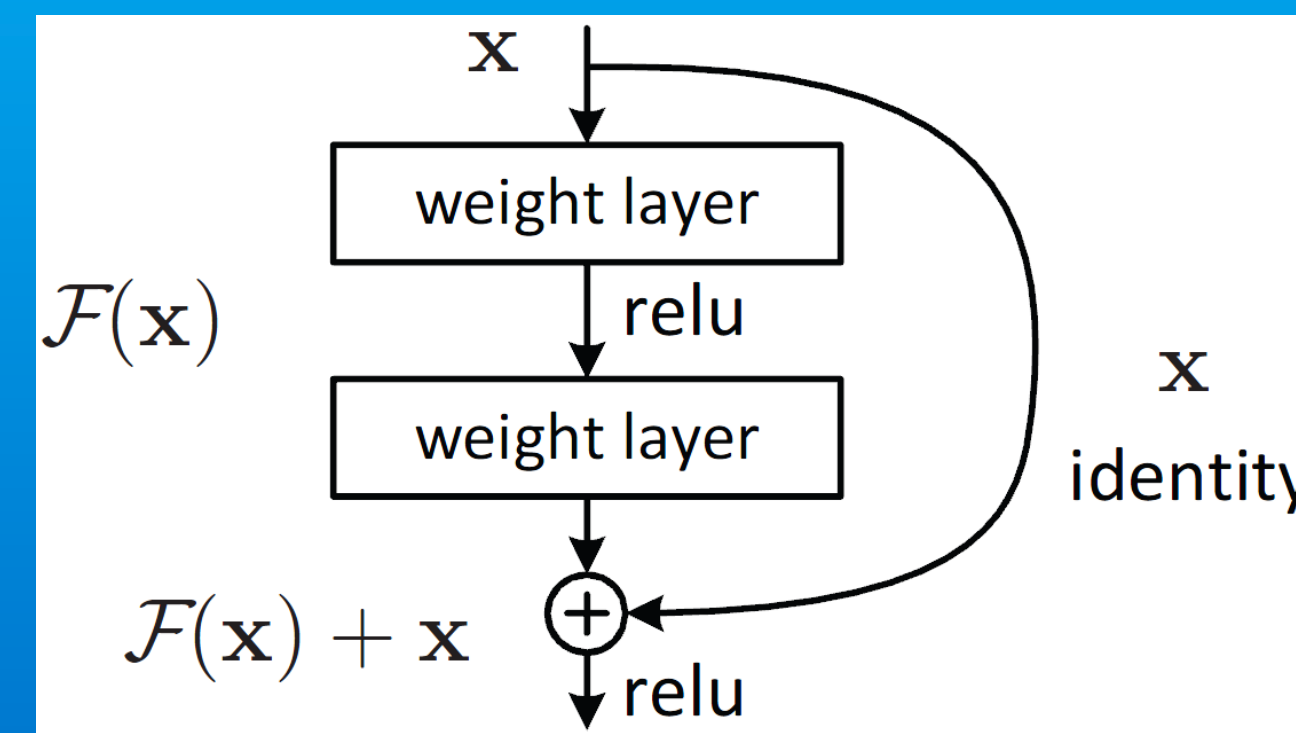
CNN (Convolutional Neural Network)



What it is: A NN designed specifically to process grid-like data, such as images.

Key idea: Instead of connecting every input pixel to every neuron, CNNs use convolutional layers that scan small regions (filters/kernels) of the image.

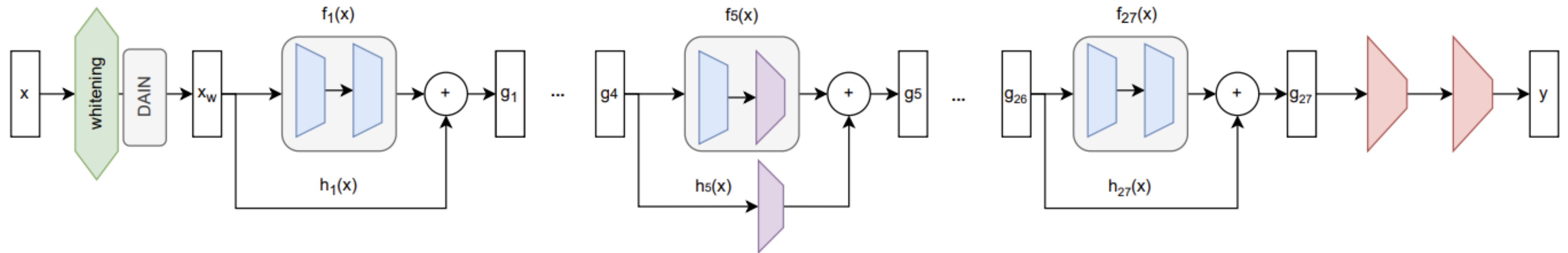
ResNet (Residual Network)



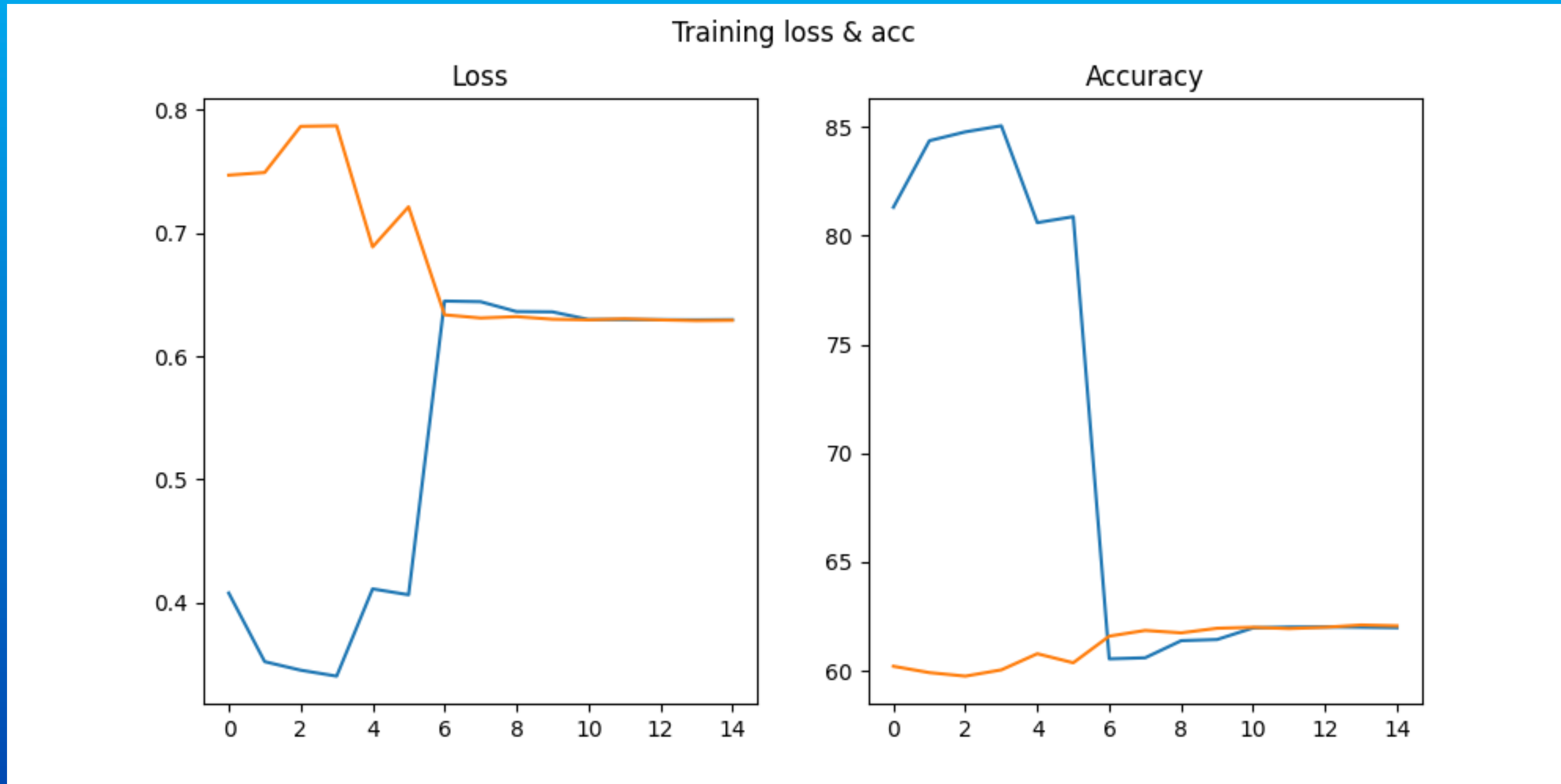
What it is: A special kind of CNN that adds *shortcut connections* (also called residual connections).

Key idea: ResNets use skip connections (residual blocks) to avoid vanishing gradients and enable training of very deep networks.

AresGW Architecture (ResNet 54)



Training Loss and Accuracy



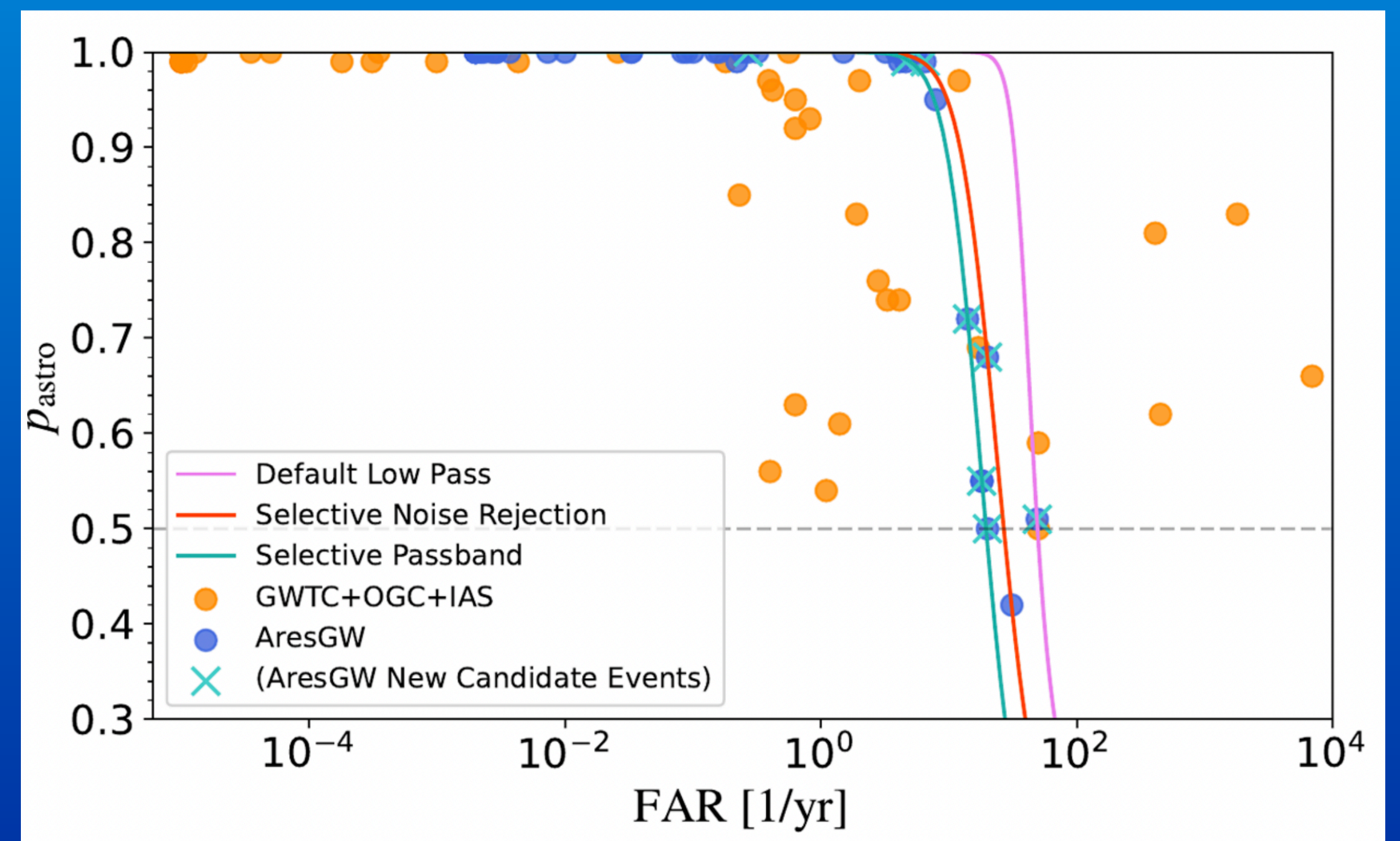
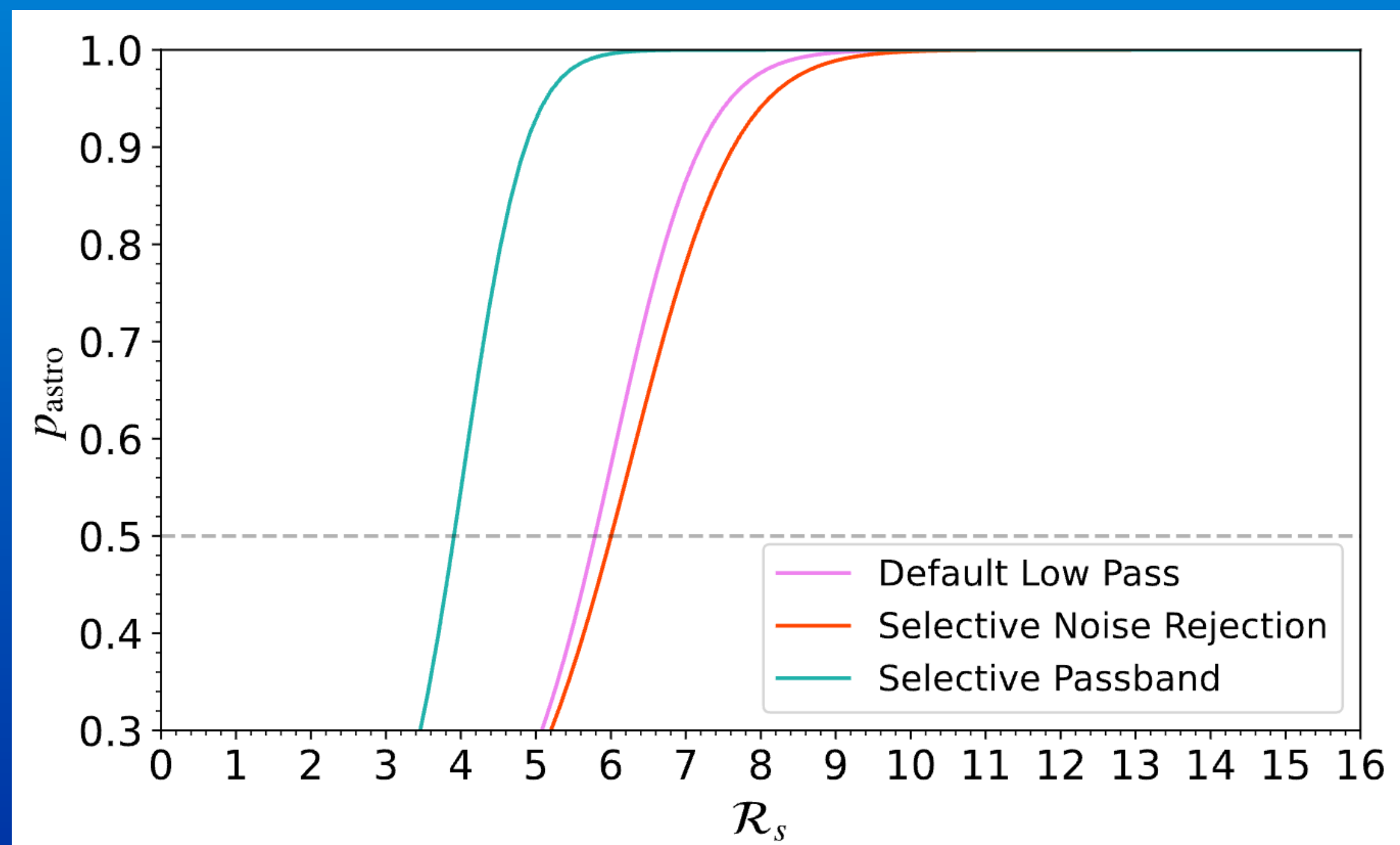
Astrophysical Probability (p_{astro})

$$b(\langle \mathcal{R}_s \rangle) = \frac{dB}{d\langle \mathcal{R}_s \rangle}$$

$$f(\langle \mathcal{R}_s \rangle) = \frac{dF}{d\langle \mathcal{R}_s \rangle}$$



$$p_{\text{astro}} = \frac{f(\langle \mathcal{R}_s \rangle)}{b(\langle \mathcal{R}_s \rangle) + f(\langle \mathcal{R}_s \rangle)}$$



Parameter Estimation

#	Event Name	\mathcal{M} (M_{\odot})	q	m_1 (M_{\odot})	m_2 (M_{\odot})	D_L (Mpc)	χ_{eff}	SNR (H1)	SNR (L1)	SNR $\hat{\rho}$ (network)
1	GW190511_125545	$28.95^{+9.45}_{-6.86}$	$0.72^{+0.25}_{-0.36}$	$40.7^{+16.2}_{-10.5}$	$28.2^{+11.6}_{-11.2}$	3707^{+3471}_{-2173}	$0.23^{+0.25}_{-0.29}$	2.29	7.34	7.29
2	GW190614_134749	$25.97^{+16.59}_{-6.20}$	$0.70^{+0.27}_{-0.36}$	$37.0^{+31.8}_{-10.7}$	$25.2^{+15.2}_{-9.7}$	6551^{+9562}_{-3558}	$0.05^{+0.34}_{-0.34}$	3.51	6.08	7.02
3	GW190607_083827	$30.48^{+7.21}_{-4.68}$	$0.78^{+0.19}_{-0.29}$	$40.5^{+12.0}_{-7.6}$	$31.0^{+9.1}_{-8.2}$	4928^{+2725}_{-2435}	$0.01^{+0.26}_{-0.30}$	4.04	7.29	8.33
4	GW190904_104631	$21.24^{+5.76}_{-4.40}$	$0.64^{+0.31}_{-0.33}$	$31.3^{+14.5}_{-8.5}$	$19.7^{+7.1}_{-7.2}$	5614^{+4441}_{-2864}	$0.05^{+0.30}_{-0.37}$	4.50	4.88	6.64
5	GW190523_085933	$23.82^{+10.24}_{-7.95}$	$0.49^{+0.45}_{-0.32}$	$41.7^{+19.3}_{-15.5}$	$19.4^{+14.6}_{-10.5}$	6091^{+6613}_{-3702}	$0.42^{+0.31}_{-0.45}$	3.48	5.14	6.02
6	GW200208_211609	$18.83^{+4.68}_{-3.18}$	$0.69^{+0.28}_{-0.40}$	$26.9^{+14.6}_{-6.3}$	$18.0^{+6.4}_{-6.9}$	3669^{+3413}_{-1985}	$0.01^{+0.37}_{-0.37}$	4.75	6.22	7.83
7	GW190705_164632	$27.21^{+7.34}_{-5.24}$	$0.52^{+0.41}_{-0.32}$	$44.7^{+24.8}_{-12.8}$	$23.0^{+11.7}_{-9.8}$	5692^{+4030}_{-2863}	$0.29^{+0.26}_{-0.34}$	4.42	6.88	8.11
8	GW190426_082124	$17.93^{+4.12}_{-3.42}$	$0.45^{+0.45}_{-0.28}$	$31.5^{+22.5}_{-11.3}$	$13.8^{+6.9}_{-5.2}$	3213^{+4555}_{-1573}	$-0.01^{+0.39}_{-0.50}$	5.15	4.46	6.41

Performance of AresGW model 2 in detecting BBH events in L+V or H+V data / O1 and O2 data

Even though AresGW was not trained on Virgo data, it generalizes well, when the Virgo detector is used in place of Livingston or Hanford:

#	Event Name	Catalog	Detectors	$\langle \mathcal{R}_s \rangle$	\mathcal{M} (M_\odot)	m_1 (M_\odot)	m_2 (M_\odot)
1	GW191216_213338	GWTC	HV	11.5	8.33	12.1	7.7
2	GW190630_185205	GWTC	LV	8.2	25.1	35.1	24.0
3	GW200112_155838	GWTC	LV	7.2	27.4	35.6	28.3
4	GW190708_232457	GWTC	LV	5.6	13.1	19.8	11.6
5	GW190620_030421	GWTC	LV	3.2	38.1	58.0	35.0
6	GW200302_015811	GWTC	HV	2.5	23.4	37.8	20.0
7	GW190925_232845	GWTC	HV	2.3	15.6	20.8	15.5
8	GW190910_112807	GWTC	LV	1.4	33.5	43.8	34.2

Even though AresGW was not trained on O1 and O2 data, it identifies 6/8 events in its effective training range with its greater $\langle \mathcal{R}_s \rangle$ value:

#	Event Name	Catalog	$\langle \mathcal{R}_s \rangle$	\mathcal{M} (M_\odot)	m_1 (M_\odot)	m_2 (M_\odot)
1	GW170104	GWTC	≥ 16.0	21.1	28.7	20.8
2	GW170729	GWTC	≥ 16.0	34.6	54.7	30.2
3	GW170809	GWTC	≥ 16.0	24.8	34.1	24.2
4	GW170814	GWTC	≥ 16.0	24.1	30.9	24.9
5	GW170823	GWTC	≥ 16.0	28.6	38.3	29.0
6	GW150914	GWTC	≥ 16.0	27.9	34.6	30.0
7	GW170818	GWTC	15.4	26.8	34.8	27.6
8	GW151012	GWTC	2.5	15.6	24.8	13.6
9	GW151226	GWTC	1.3	8.9	14.2	7.5