Gravitational-wave parameter inference using Deep Learning

J.D. Álvares, J.A. Font, F. Freitas, O.G. Freitas, A. Morais, S. Nunes, A. Onofre, A. Torres-Forné

Ongoing MSc work under the supervision of A. Onofre and J.A. Font
Special thanks to F. Freitas

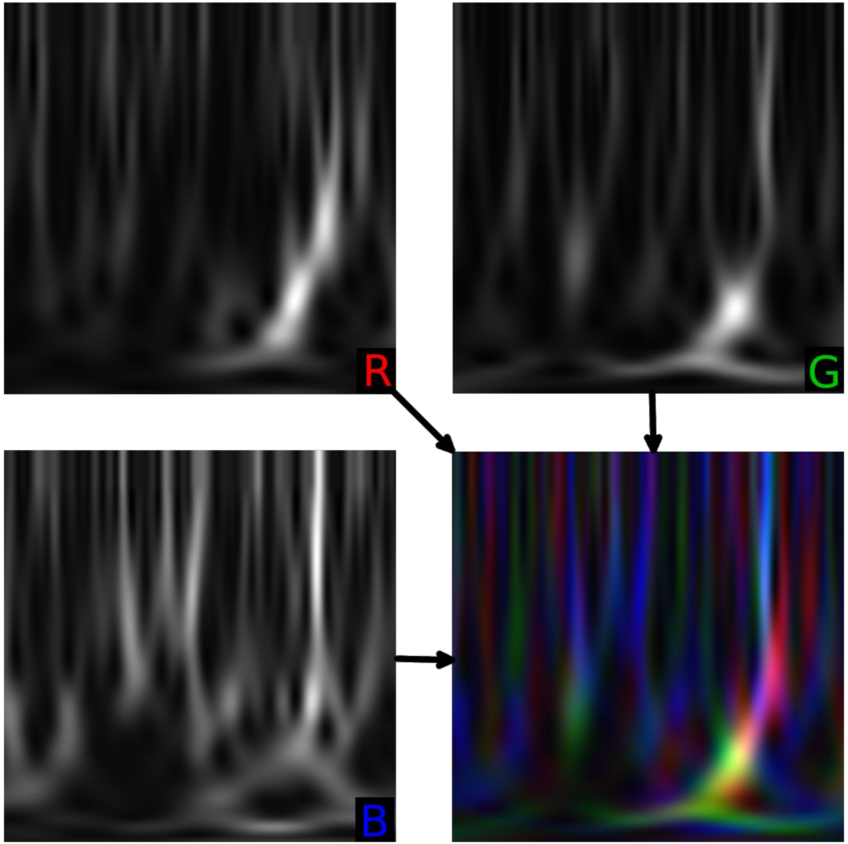
Goals

Using spectrogram data:

- Train a classifier capable of distinguishing between binary black hole signals and detector noise
- Train a network capable of performing parameter estimation on BBH signal data

Why spectrograms?

- “Chirp” signature is a notable characteristic of GW signals from binary mergers
- High information density
- Computer Vision is a widely studied field with a plethora of well-developed tools
- **We can use RGB spectrograms:**
  - Hanford data -> Red
  - Livingston data -> Green
  - Virgo data -> Blue

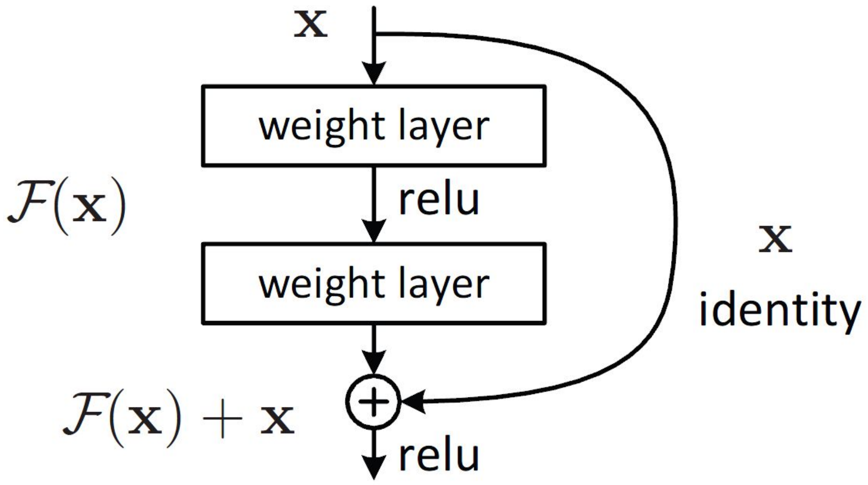


1 - Classifier

Building a network that can identify the presence of a GW signal through a spectrogram input

Choosing an architecture - Residual Networks

- Capable of state of the art results in standard classification tasks
- Good balance between performance and training time
- Avoids gradient vanishing/exploding during training
- Deeper versions of a residual network should provide, at least, equal performance to shallower versions



The CNN was implemented using fastai

www.fast.ai, arXiv:2002.04688 [cs.LG]

What we found

We created datasets for signals generated at 100, 400, 1000, 1500 and 2000 Mpc, using pyCBC to BBH signals with randomly sampled masses $(m_1, m_2) \in [5, 100]M_\odot$ (Used the SEOBNRv4_ROM approximant)

We trained a model on each of these datasets, then tested their performance on the other datasets.

✓ A network can be trained so that large amplitude signals can be detected

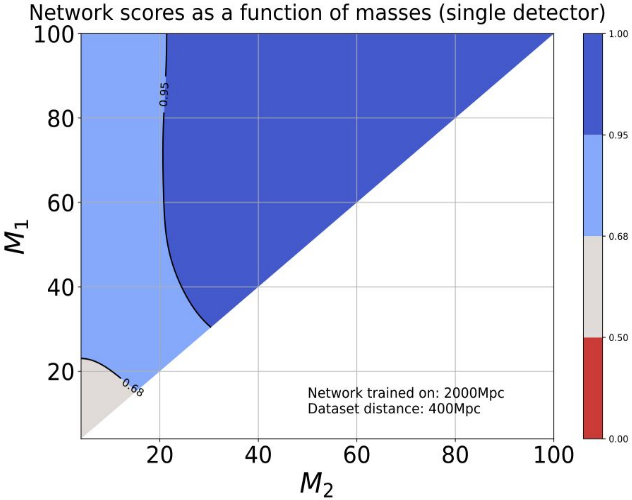
✓ A network can be trained so that small amplitude signals can be detected

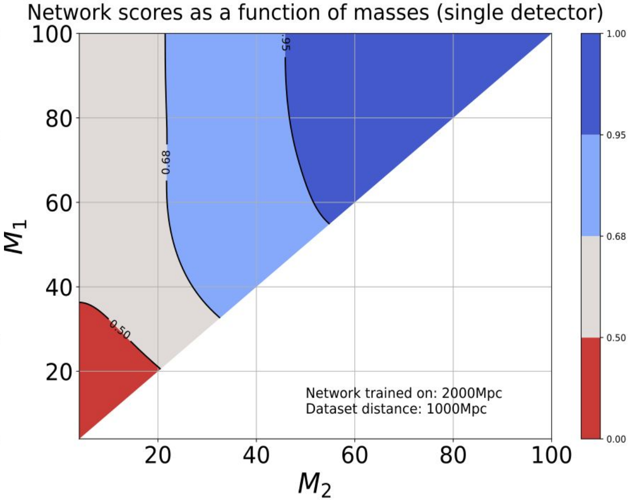
✓ A network trained on small amplitude signals is able to detect larger amplitude signals

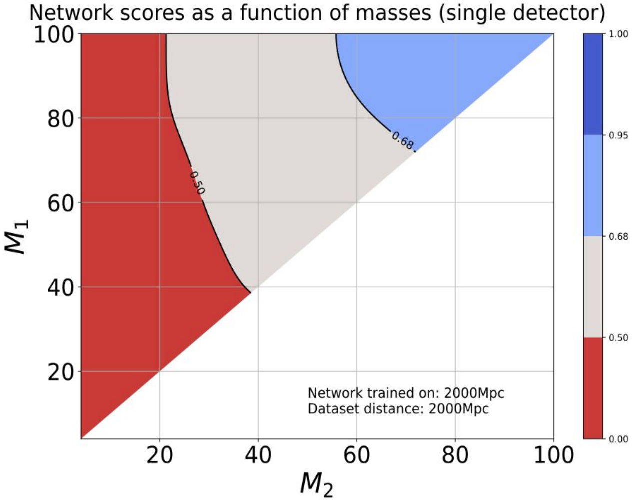
✗ A network trained on large amplitude signals is NOT able to detect smaller amplitude signals

Take home message: we can train on smaller amplitudes and retain the ability to detect large amplitude signals

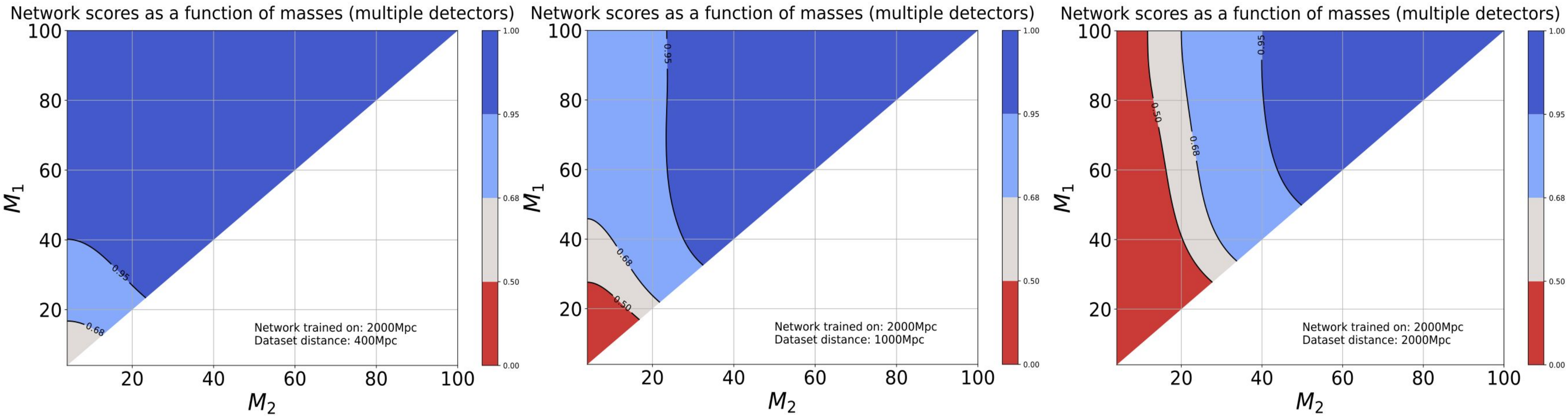
1 detector vs 3 detector performance



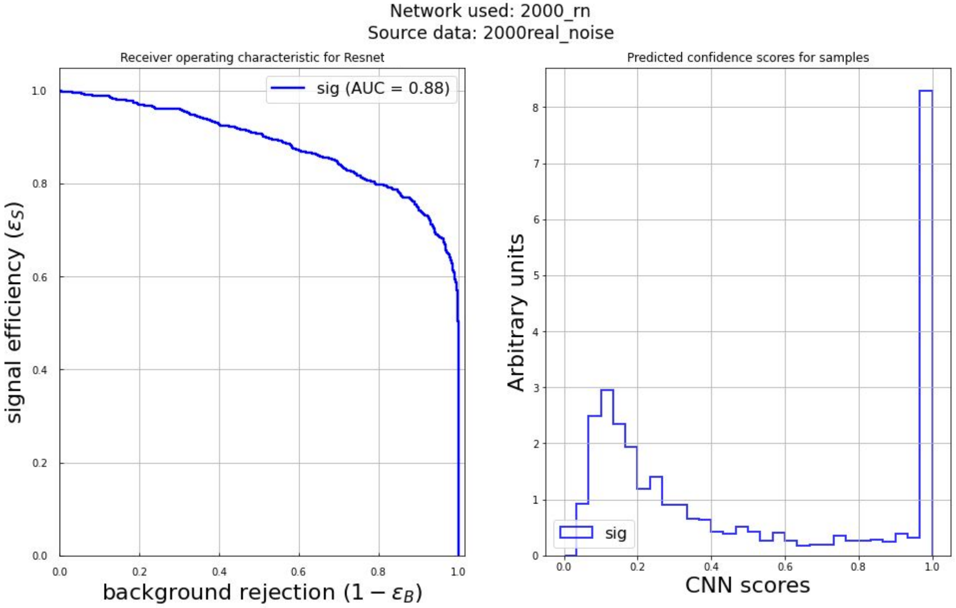




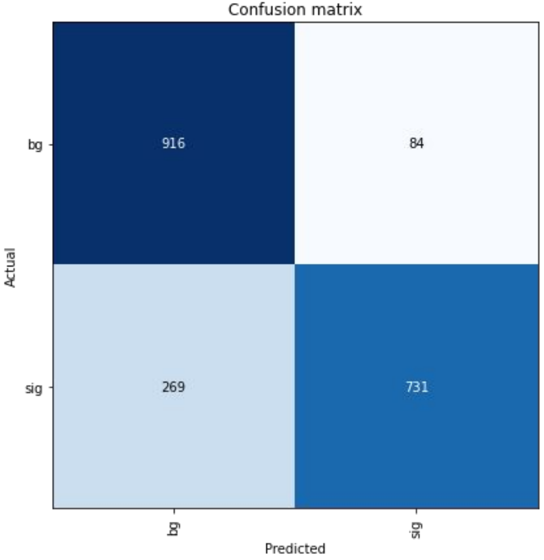
1 detector vs 3 detector performance



ResNet trained on 2Gpc data from 3 detectors



$$\varepsilon_B = \text{FPR} \approx 0.08$$



$$\text{PPV} = 0.9$$

$$\varepsilon_S = \text{TPR} \approx 0.73$$

$$\text{Acc} \approx 0.82$$

Performance of multiple-detector ResNet on O2 data

GWTC-1 Confident		GWTC-1 Marginal	
Event	Score	Event	Score
GW170814	1.00	MC151116	0.73
GW150914	1.00	MC161217	0.72
GW170823	1.00	MC170705	0.51
GW170104	1.00	MC170630	0.49
GW170729	0.99	MC170219	0.45
GW170809	0.97	MC161202	0.40
GW151012	0.96	MC170423	0.35
GW170608	0.92	MC170208	0.33
GW170818	0.88	MC170720	0.30
GW151226	0.87	MC151012A	0.26
-	-	MC151008	0.20
-	-	MC170405	0.14
-	-	MC170616	0.12
-	-	MC170412	0.09

Performance of multiple-detector ResNet on O3 data

GWTC-2			
Event	Score	Event	Score
GW190521	1.00	GW190708_232457	0.98
GW190602_175927	1.00	GW190909_114149	0.97
GW190424_180648	1.00	GW190514_065416	0.96
GW190620_030421	1.00	GW190814	0.95
GW190503_185404	1.00	GW190521_074359	0.95
GW190727_060333	1.00	GW190731_140936	0.92
GW190929_012149	1.00	GW190513_205428	0.92
GW190915_235702	1.00	GW190421_213856	0.87
GW190630_185205	1.00	GW190412	0.81
GW190519_153544	1.00	GW190728_064510	0.77
GW190706_222641	1.00	GW190719_215514	0.76
GW190413_134308	1.00	GW190803_022701	0.66
GW190701_203306	1.00	GW190930_133541	0.58
GW190517_055101	1.00	GW190828_065509	0.56
GW190408_181802	1.00	GW190924_021846	0.40
GW190910_112807	1.00	GW190707_093326	0.35
GW190828_063405	0.99	GW190720_000836	0.16
GW190413_052954	0.99	-	-
GW190512_180714	0.98	-	-
GW190527_092055	0.98	-	-

2 - Regression

Building a network that can utilize spectrogram data to perform parameter estimation on binary coalescence GW signals

Reasonable targets

- The linear GR approximation for the waveform of a compact binary coalescence in the detector frame is of the form:

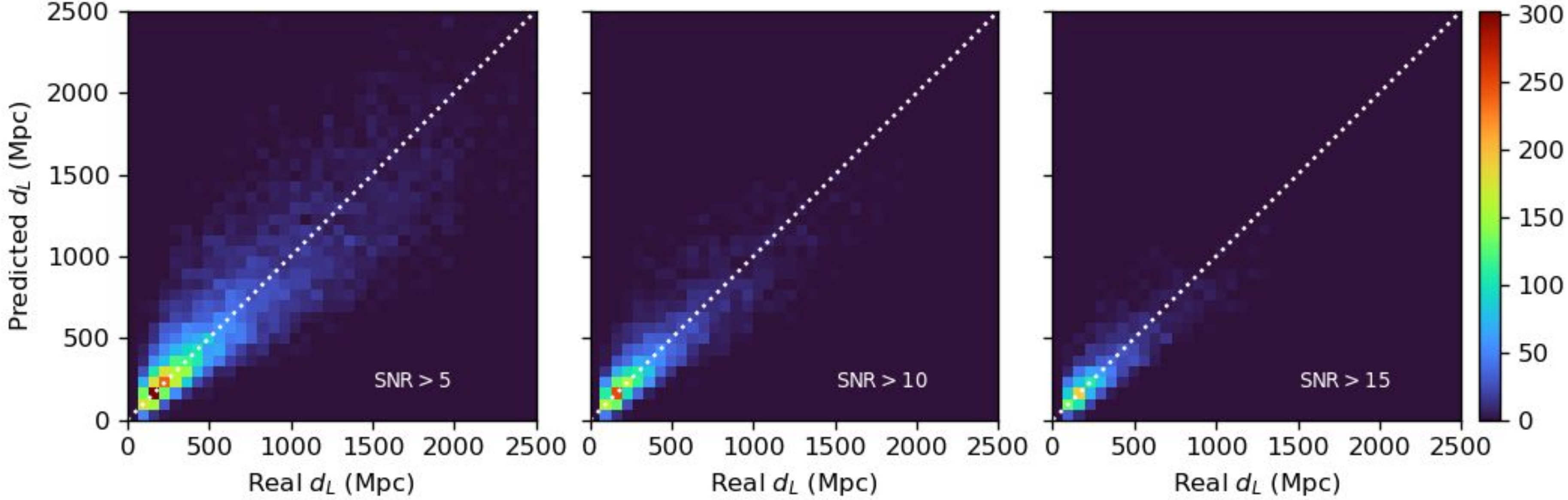
$$h_{\times,+} = \frac{\mathcal{M}^{5/3}}{d_L} F_{\times,+}(\theta, \phi) P(t)$$

  so $\mathcal{M}$, d_L are immediately good candidates for regression.
- We also include the effective inspiral spin χ_{eff}, as spin-orbit effects emerge at higher order terms
- The sky position can also be looked at (though this takes some extra care)

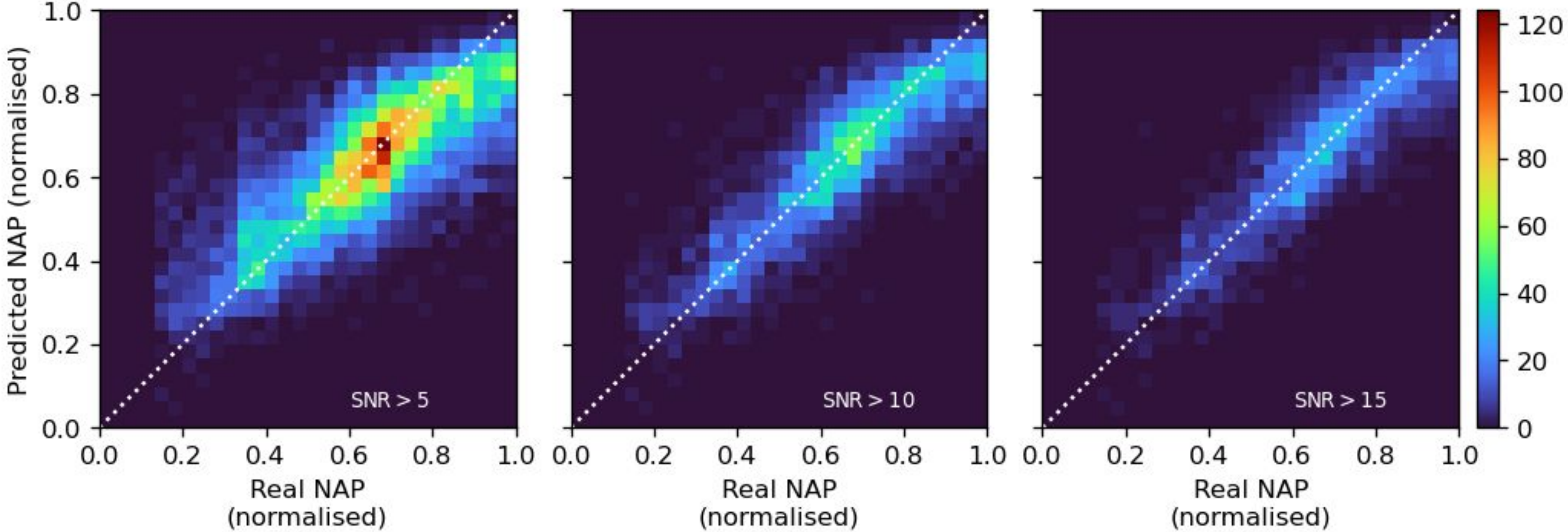
Deep Regression

- Base architecture: xResNet18
  - Mish activation
  - MaxBlurPool (improves shift invariance)
  - Dropout layers are used before pooling layers to simulate a gaussian process. This turns the network into a bayesian CNN, placing a distribution over the network's kernels.
- We use the SEOBNRv4HM_ROM approximant to generate our dataset
- Training set characterization:
  - m_1 and m_2 are randomly sampled between 5 and 100 solar masses
  - Distance is randomly sampled between 0.1 and 4 Gpc
  - The inclination is randomly sampled between 0 and π
  - Sky position and polarization are randomly sampled
  - Spin is sampled between [-1, 1] for each black hole, aligned with the orbital axis
  - We restrict the dataset elements to have an SNR>5
  - 31499 items
- Outputs: $[d_L, \mathrm{NAP}, \mathcal{M}, \chi_{\mathrm{eff}}]$

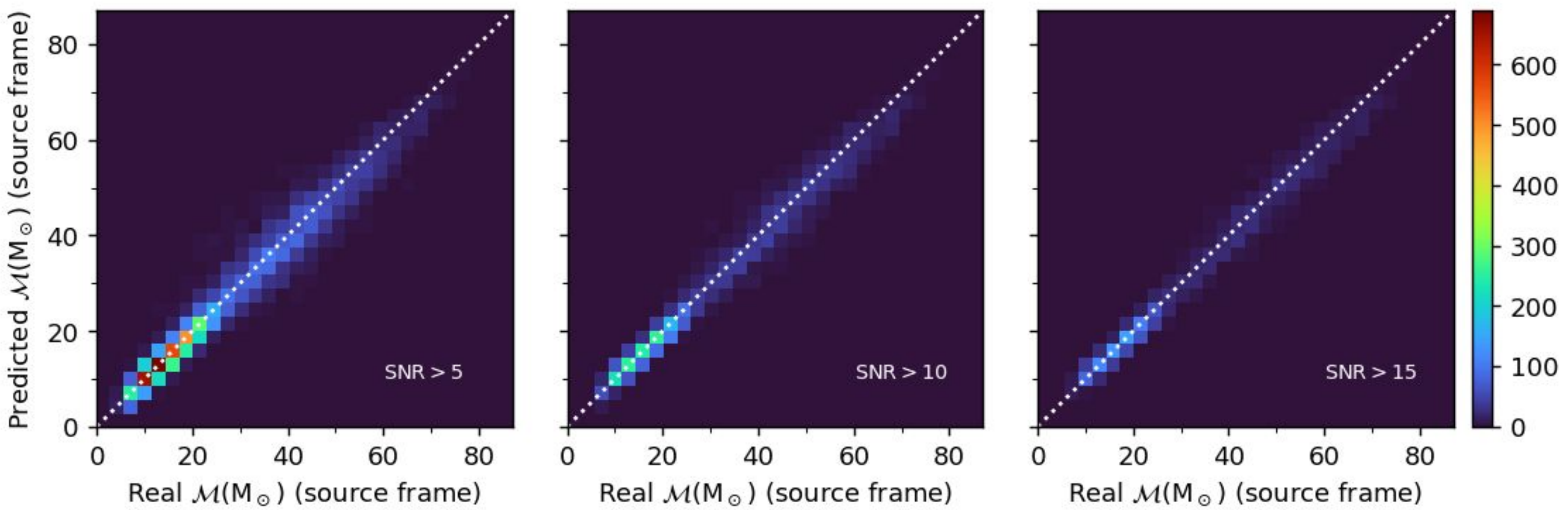
Regression mean results for d_L



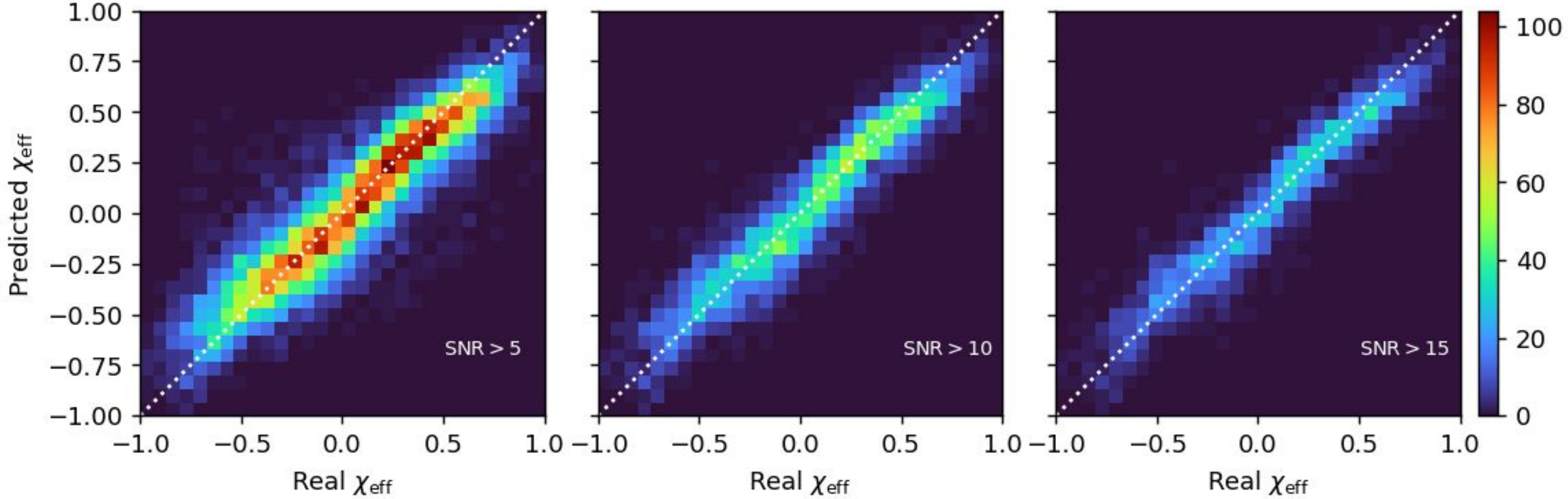
Regression mean results for NAP



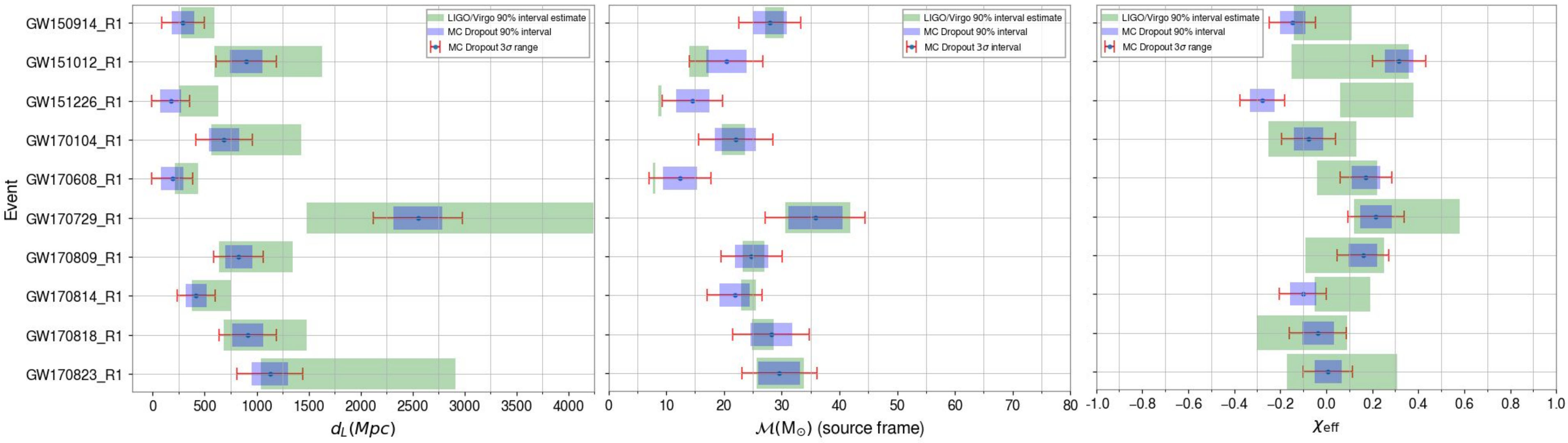
Regression mean results for $\mathcal{M}$



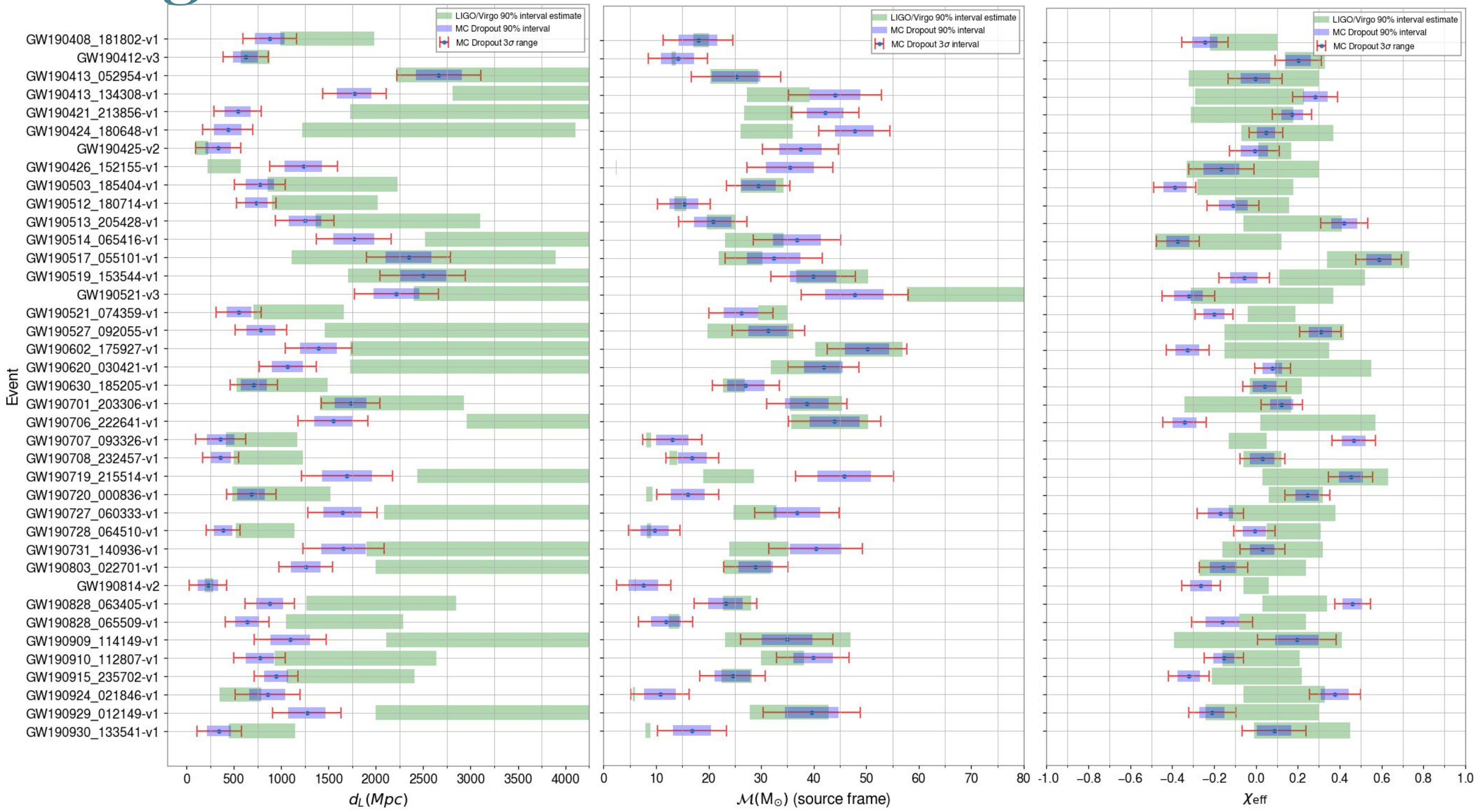
Regression mean results for χ_{eff}



Taking a look at current detections: GWTC-1

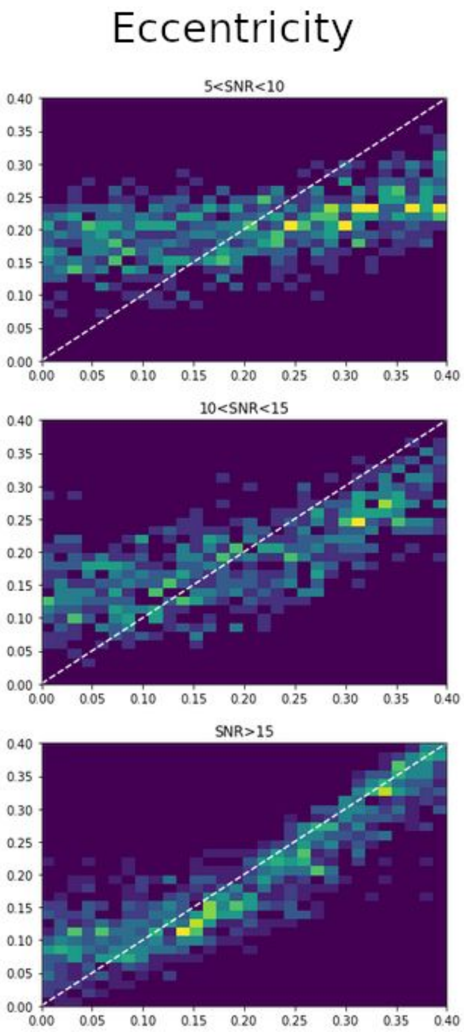


Taking a look at current detections: GWTC-2



Future work: Extending the parameter space

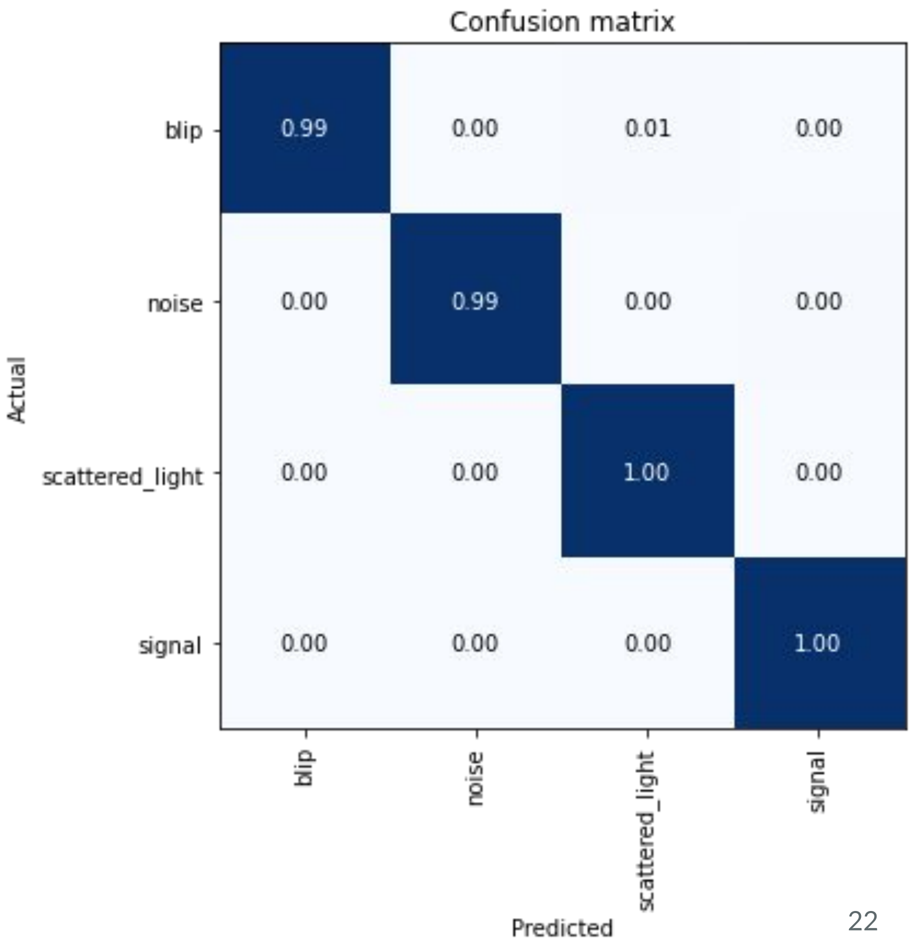
- Early tests with the eccentric TaylorF2e approximant show these methods can resolve eccentricity for total mass $M<20M_{\odot}$.
- Further tests with approximants that allow a larger parameter space are in the works.
- Effects such as precession could also be looked into.



Future work: GW detection

- We tried a different, shallower architecture (ShuffleNet)
- To apply a classifier over arbitrary LIGO/Virgo data, including glitches should contribute to robustness
- To minimise false detections, score averaging + higher thresholds can be used

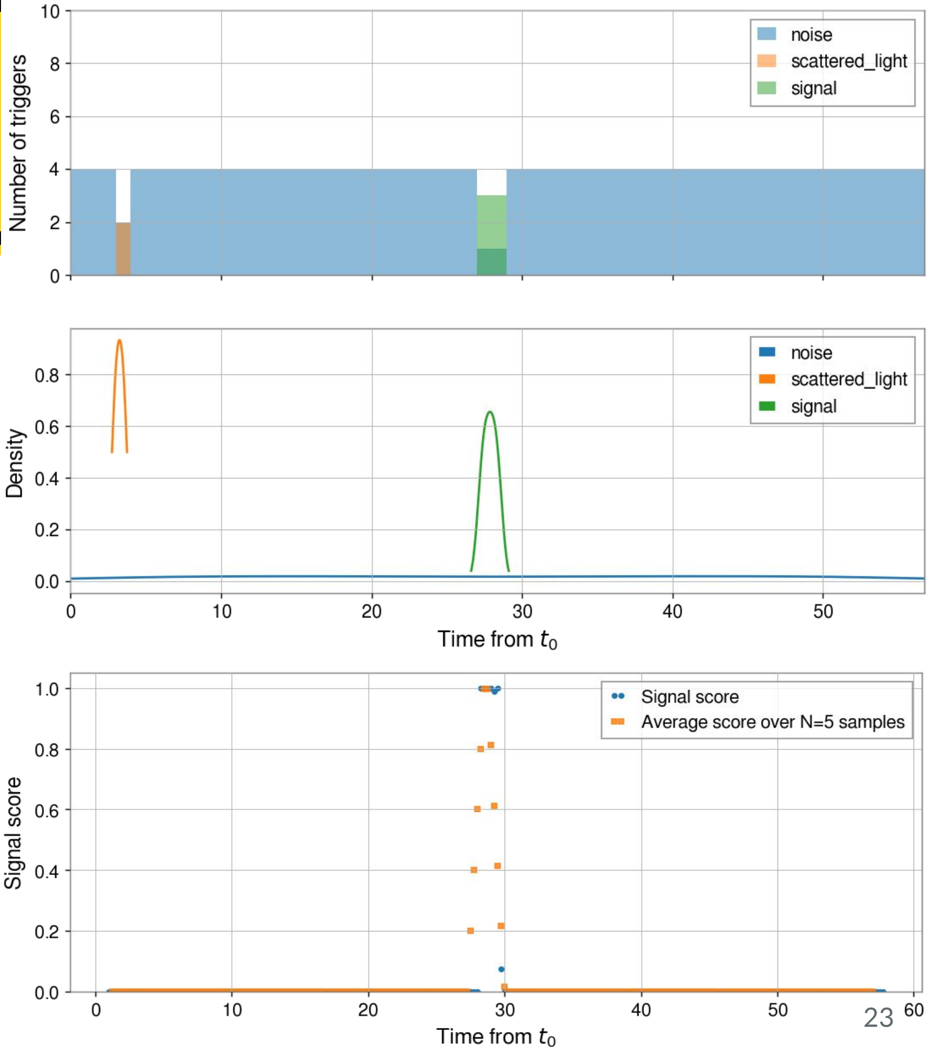
- Example on the right: a 60s segment centered around GW190929
- From the 49 known BBH signals, 28 meet the detection conditions under this setup
- Further optimization is likely possible



Future work: GW detection

- We tried a different, shallower architecture (ShuffleNet)
- To apply a classifier over arbitrary LIGO/Virgo data, including glitches should contribute to robustness
- To minimise false detections, score averaging + higher thresholds can be used

- Example on the right: a 60s segment centered around GW190929
- From the 49 known BBH signals, 28 meet the detection conditions under this setup
- Further optimization is likely possible



Thank you for your attention

For a more in-depth discussion, check out our paper at https://doi.org/10.1088/1361-6382/ac0455

GWG
VNIVERSITAT DE VALÈNCIA
CF-UM-UP
Universidade do Minho
Gr@v
universidade de aveiro
theoria poiesis praxis

Bayesian CNN using mc-dropout

Y. Gal and Z. Ghahramani, ‘Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference’, *arXiv:1506.02158 [cs, stat]*, Jan. 2016

- Dropout is applied at validation and test time in addition to during training
- Multiple forward passes of a single input provide a distribution whose mean value tends to result in lower discrepancies with the real value
- Allows for the estimation of the model’s epistemic uncertainty

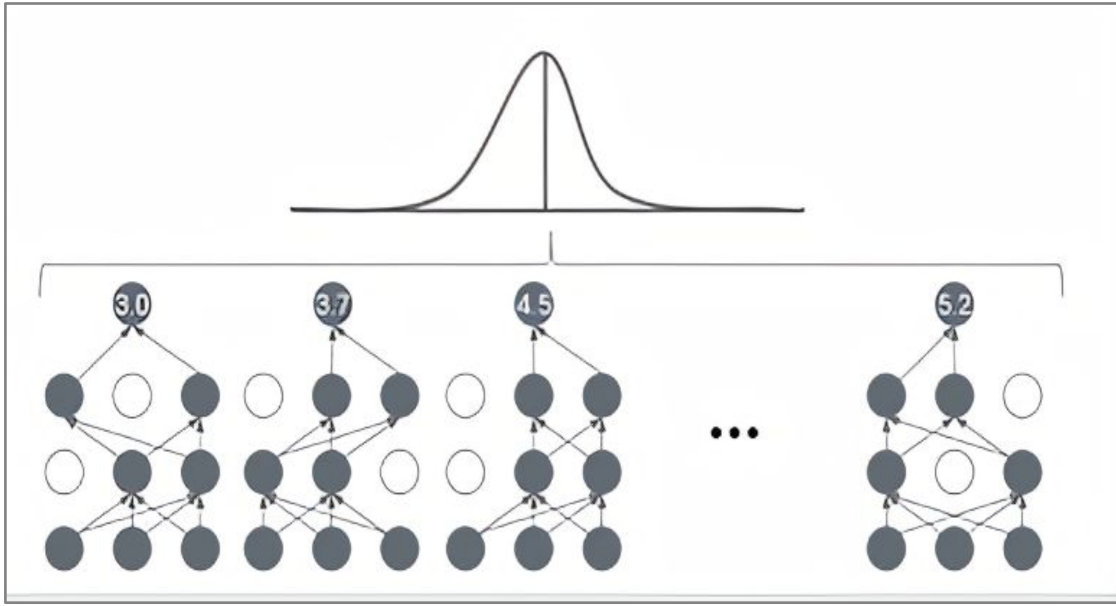
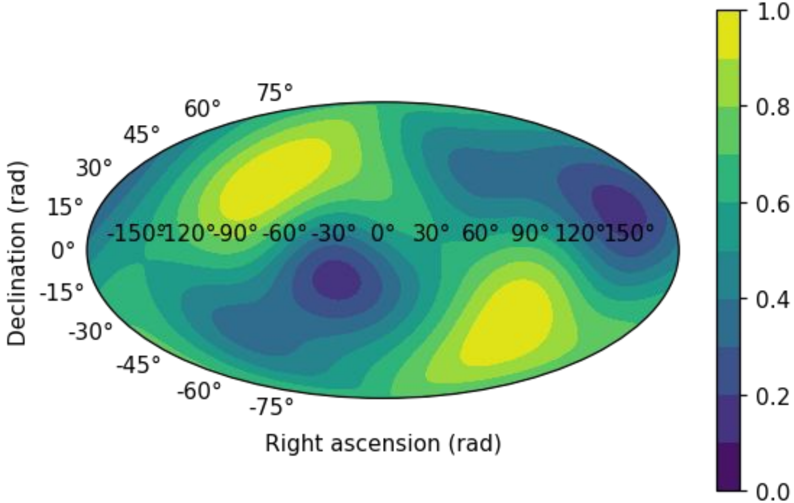


Image credits: Daniel Huynh

The Network Antenna Power (NAP)

- The sky position of an event cannot be fully determined with only 3 detectors, there being 2 points that are equally likely to be the origin of a certain event.
- This degeneracy between physically indistinguishable events can keep the network from converging
- The NAP can serve as a proxy observable from which we can deduce the sky position, given a certain event and the corresponding GPS time



Network Antenna Power as a function of sky position for (t mod 24h)=9742

Dataset information

Classification		
Parameters	Train size	Validation size
Single detector: $(m_1, m_2) \sim U(5, 100)\ M_\odot$, $d_L = [100, 300, 1000, 1500, 2000]$ Mpc, $\iota = \frac{\pi}{2}$, approximant: `SEOBNRv4_ROM`	4000 images 560×560 pixels 8-bit gray scale	1000 images 560×560 pixels 8-bit gray scale
Total images	20000	5000
Multiple detector: $(m_1, m_2) \sim U(5, 100)\ M_\odot$, $d_L = [100, 300, 1000, 1500, 2000]$ Mpc, $\iota = \frac{\pi}{2}$, approximant: `SEOBNRv4_ROM`	4000 images 560×560 pixels 8-bit RGB	1000 images 560×560 pixels 8-bit RGB
Total images	20000	5000

Regression		
Parameters	Train size	Validation size
Multiple detector: $(m_1, m_2) \sim U(5, 100)\ M_\odot$, $d_L \sim U(100, 2500)$ Mpc, $\iota \sim U(0, \pi)$, $spin \sim U(-1, 1)$, approximant: `SEOBNRv4HM_ROM`	12769 images 224×224 pixels 8-bit RGB	3192 images 224×224 pixels 8-bit RGB
approximant: `IMRPhenomPv2`	10338 images 224×224 pixels 8-bit RGB	2584 images 224×224 pixels 8-bit RGB
approximant: `IMRPhenomD`	10625 images 224×224 pixels 8-bit RGB	2656 images 224×224 pixels 8-bit RGB
Total images	43009	14689
Extra dataset		
Multiple detector: $(m_1, m_2) \sim U(5, 35)\ M_\odot$, $d_L \sim U(100, 2500)$ Mpc, $\iota \sim U(0, \pi)$, $spin \sim U(-1, 1)$, approximant: `SEOBNRv4HM_ROM`	15538 images 224×224 pixels 8-bit RGB	