Searching and characterizing compact binary coalescence signals:
challenges and solutions in real data

Greg Ashton

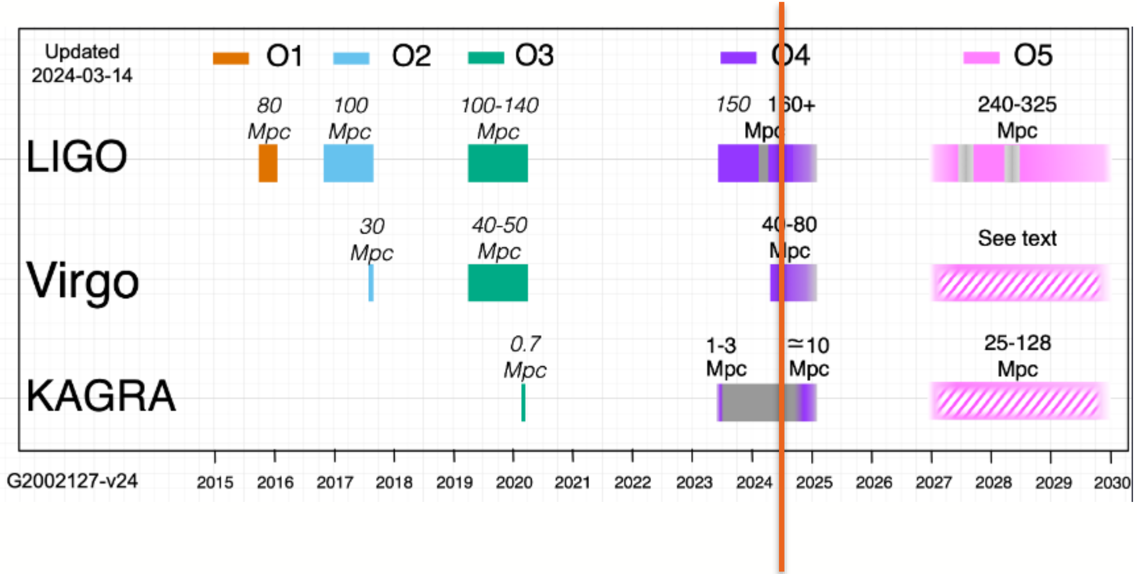
ROYAL HOLLOWAY UNIVERSITY OF LONDON

Background and motivation

Observing runs

O4b started on the 10th of April

We are here



observing.docs.ligo.org/plan/

Observations to date

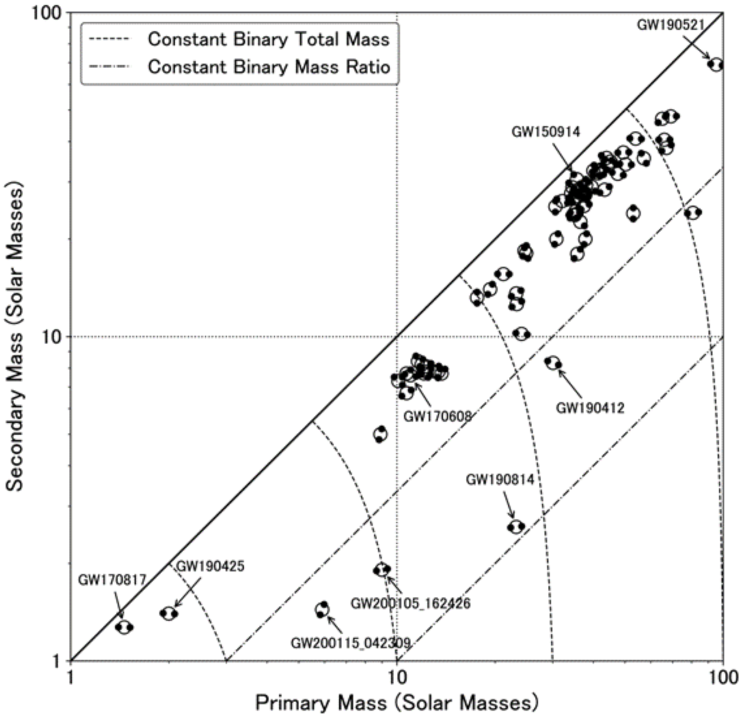
ROYAL HOLLOWAY UNIVERSITY OF LONDON

- O1-O3 produced nearly 100 observations
- All signals arise from Compact Binary Coalescence (CBC):
  - Binary black hole collisions
  - Binary neutron star collisions
  - Neutron star – black holes

Stand out highlights:

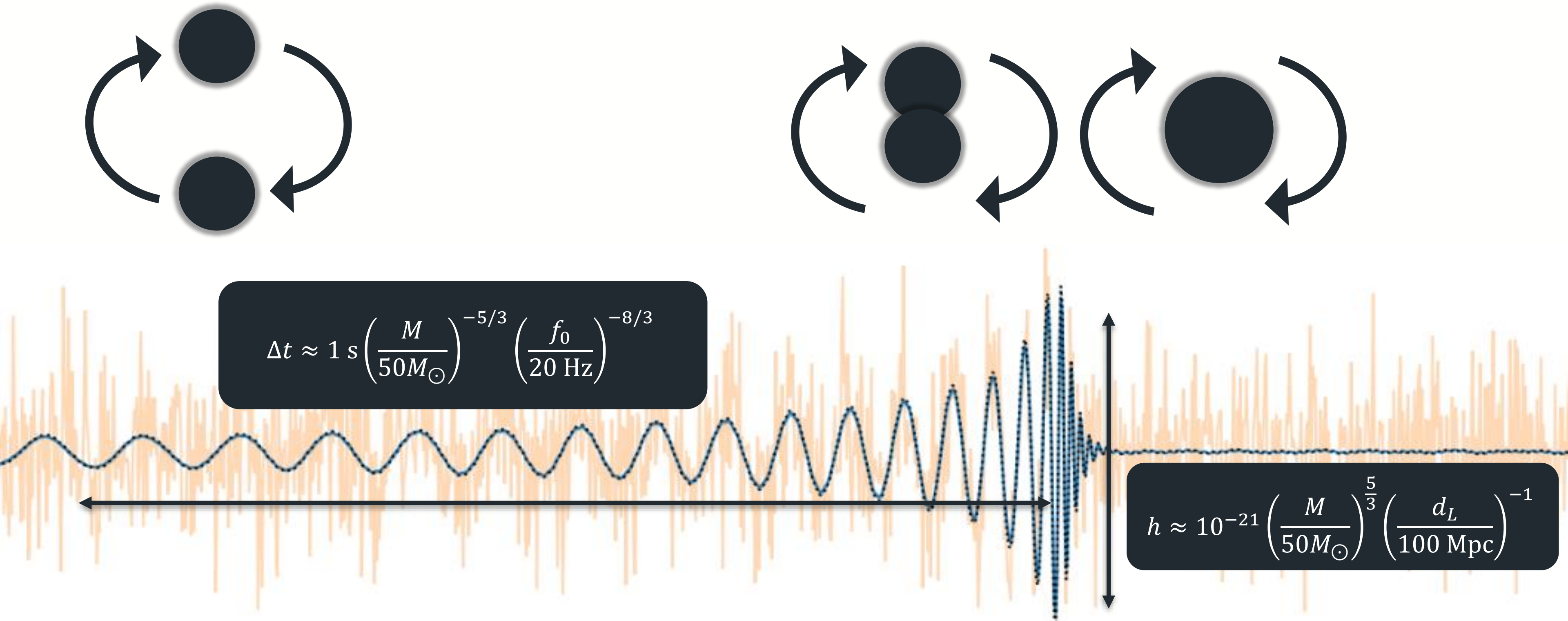
- GW1050914: the first observation
- GW170817: a multimessenger BNS

Credit: LIGO-Virgo-KAGRA Collaboration / IGFAE / Thomas Dent



Compact binary mergers (CBC)

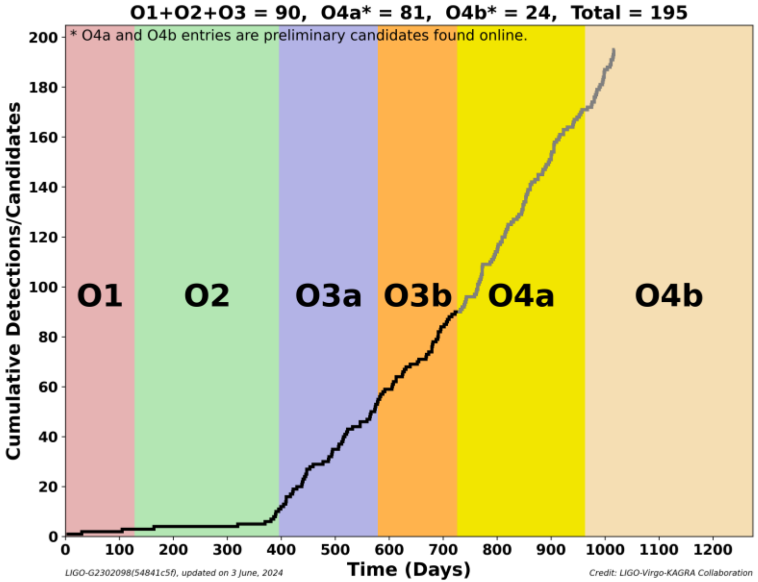
Inspiral **Merger** **Ringdown**



$$\Delta t \approx 1\ \mathrm{s} \left(\frac{M}{50 M_\odot}\right)^{-5/3} \left(\frac{f_0}{20\ \mathrm{Hz}}\right)^{-8/3}$$

$$h \approx 10^{-21} \left(\frac{M}{50 M_\odot}\right)^{\frac{5}{3}} \left(\frac{d_L}{100\ \mathrm{Mpc}}\right)^{-1}$$

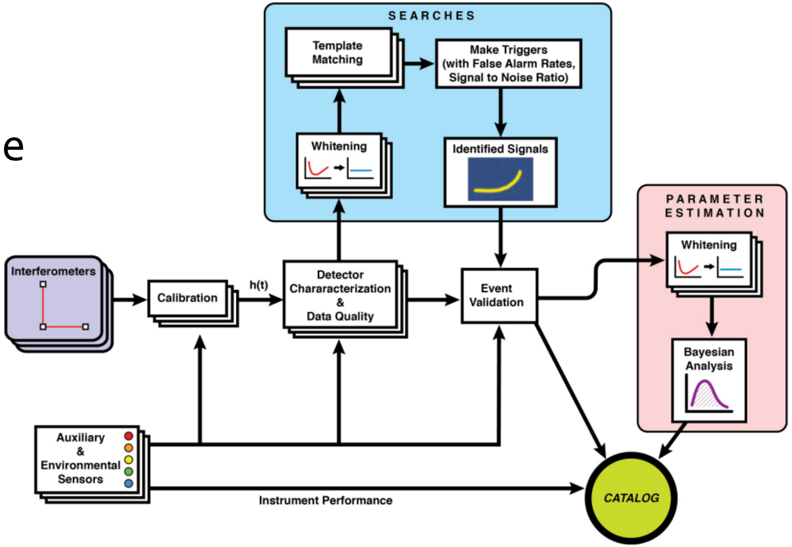
CBC triggers



- Total number of triggers in O4 now exceeds all previous runs
- No clear binary neutron star candidates..
- No multimessenger candidates..

Gravitational-wave data analysis

- Sam Finn (1992):
  - **Search:** decide if the data contains a signal
  - **Parameter Estimation:** assume the presence of a signal and measure its parameters
- LIGO-Virgo-KAGRA:
  - Calibration, Detector Characterisation
  - Search + Parameter Estimation
  - Population studies, Tests of General Relativity, Cosmology, Searching for lensed pairs, …



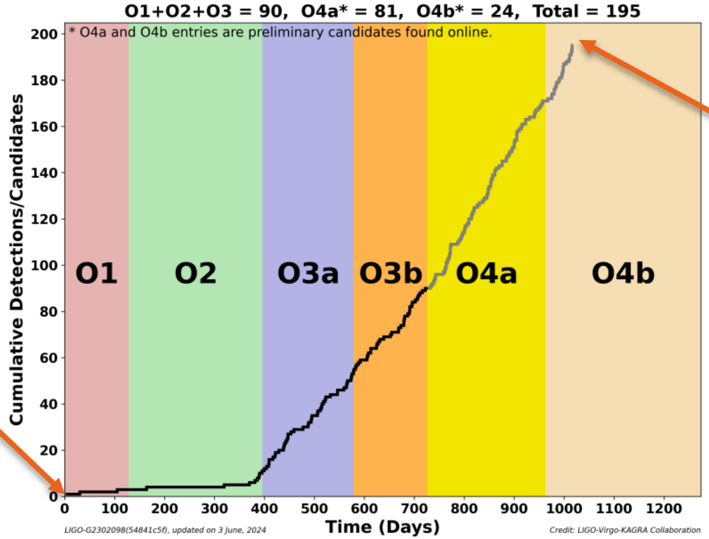
arXiv:1908.11170

Challenge: more signals!

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Challenge: more signals

200+ scientists per event

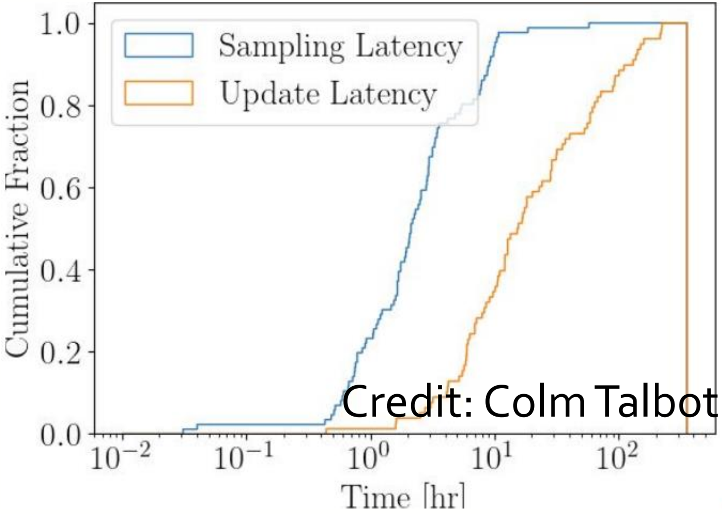


1 scientists per event

Challenge: more signals

Time is of the essence: low-latency and high-latency analyses

- Identify and address bottlenecks
- Build faster software:
  - More efficient
  - Leverage computational parallelisation
  - Improve validation/checking
- Automate everything



Credit: Colm Talbot

CBCFlow

ASIMOV

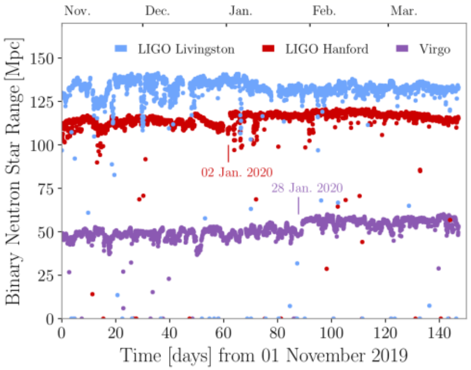
BILBY

Challenge: an asymmetric detector network

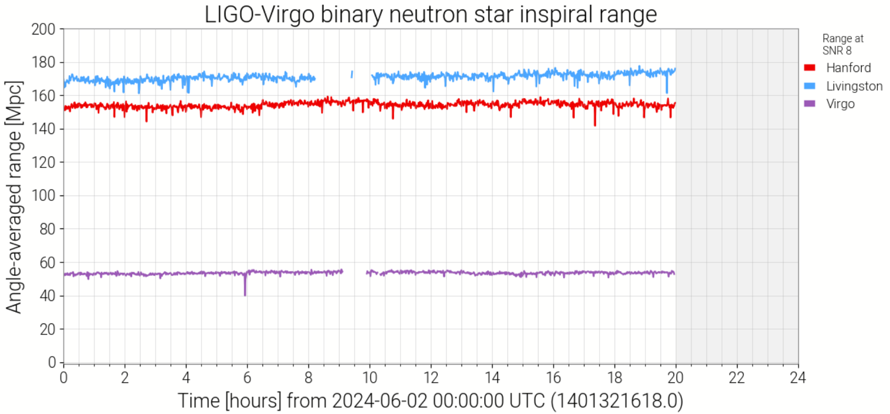
Challenge: an asymmetric detector network

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Virgo did not join O4a due to technical challenges
- Excellent duty cycles across LIGO-Virgo in O4b
- LIGO 3x more sensitive than Virgo
- KAGRA to join later in the run with ~10 Mpc

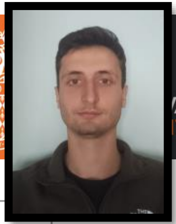


O3 (arXiv:2111.03606)



O4 (gwosc.org/detector_status/day/20240602/)

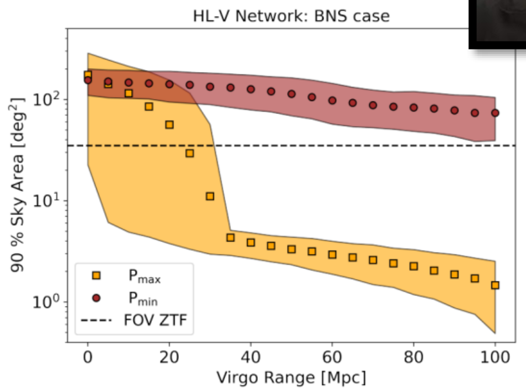
Work by Mattia Emma

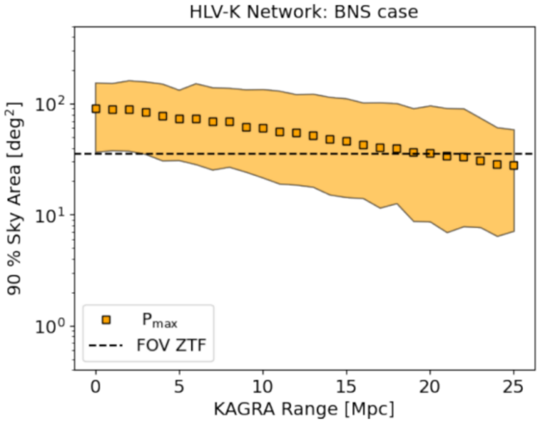


Challenge: an asymmetric detector network

Emma et al. (2024) study O4 sensitivities:

- For detection:
  - A third weaker IFO is not (on average) useful
- For parameter estimation:
  - A third detector is critical to improving **sky localisation**
  - When Virgo is within a factor of 6 of LIGO, it produces a significant improvement in the sky localisation
  - Though the KAGRA range is limited, it could provide a factor of a few reduction in sky area for an optimal sky location

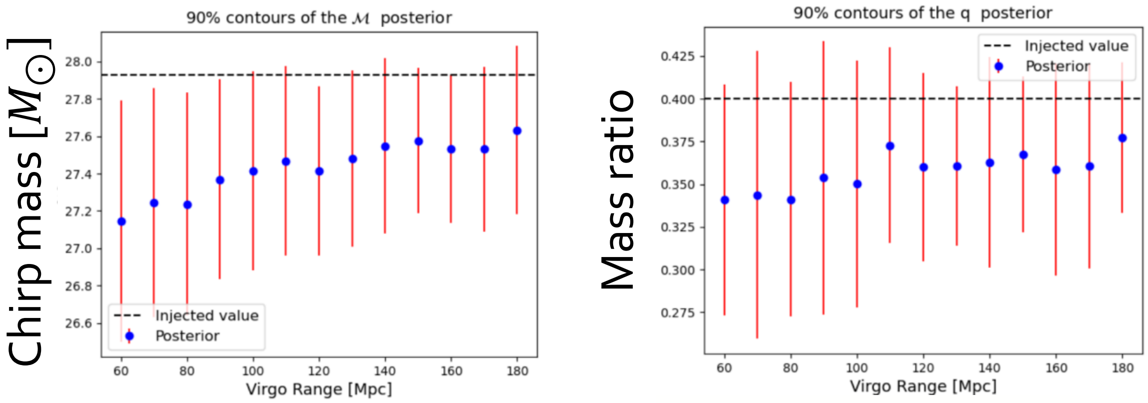




arXiv:2404.16949

Challenge: an asymmetric detector network

Extending the Virgo range closer to LIGO sensitivity would moderately improve intrinsic source parameter estimation

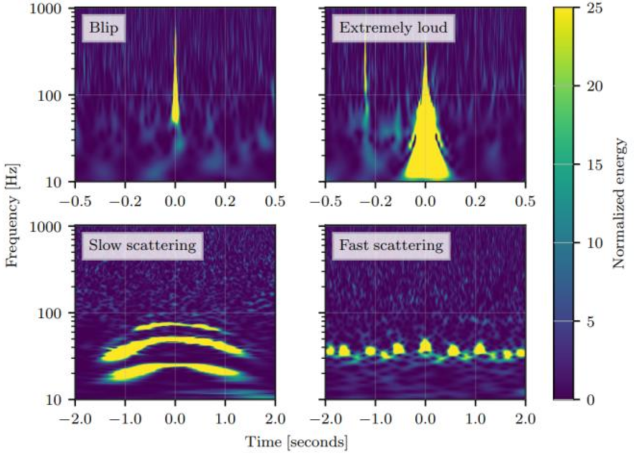


Challenge: real detectors are full of glitches

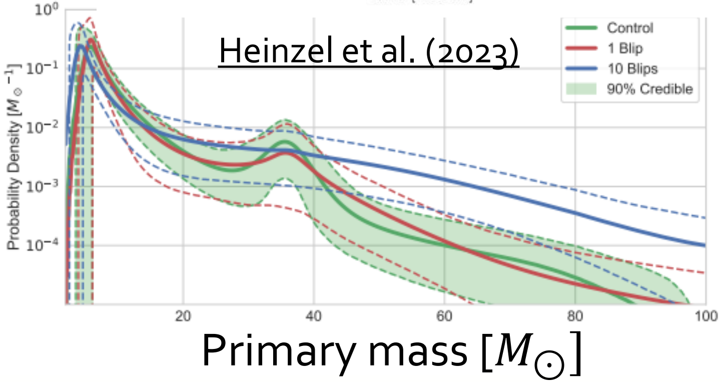
Challenge: real detector data is full of glitches

- **Glitches**: transient non-Gaussian noise
- One every few minutes
- Impact:
  - Reduce search sensitivity
  - Contaminate observed signals
  - Contaminate the population

Davis & Walker (2022)

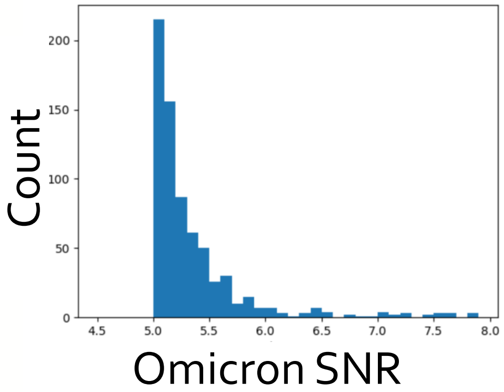


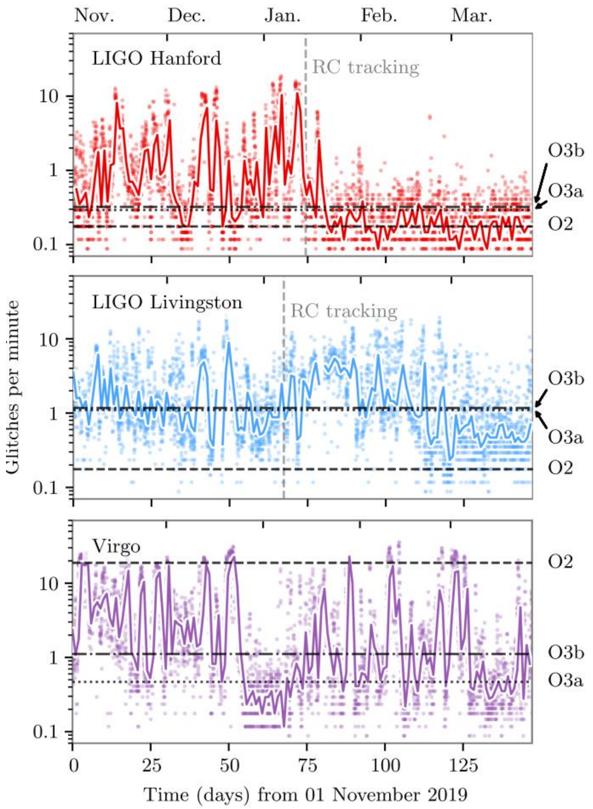
Heinzel et al. (2023)



Challenge: glitches

- How many LVK signals are contaminated by a glitch?
  - Glitches happen about **once per minute** (Omicron: $\rho > 6.5$)
  - Most signals last just a few seconds
  - $\sim$ **1/15 signals should be contaminated**..
- Note that the Omicron rate is saturated





Challenge: glitches

However, 20% of signals in GWTC-3 are "deglitched"

Deglitching:

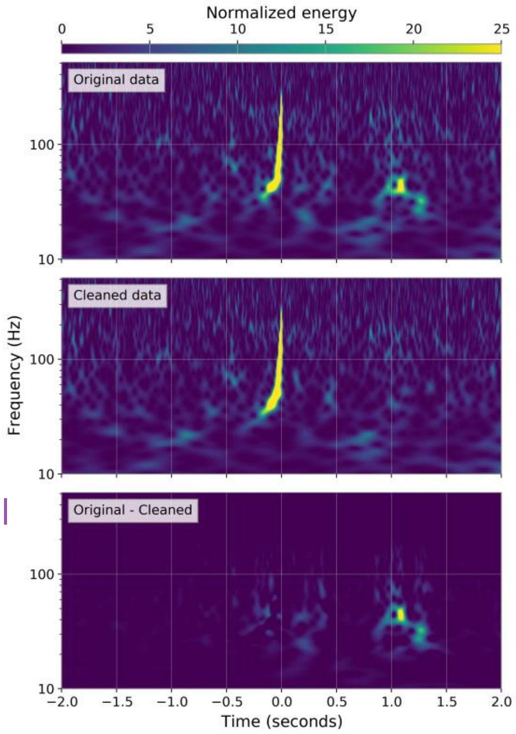
- Build a model for the glitch
- Subtract it from the data
- Analyse the "cleaned" data the standard Whittle likelihood:

$$\ln \mathcal{L}(\tilde{\boldsymbol{d}}|\theta, M) \propto -\frac{2}{T}\sum_j \frac{\left|\tilde{d}_j - \tilde{\mu}_j(\theta) - \tilde{g}_j(\vartheta)\right|^2}{P_j}$$

Fixed glitch model

Methods: BayesWave, gwsubtract, ML non-linear

Problem: deglitching ignores glitch model uncertainty



Alternative: Implicit modelling with Gaussian processes

Recall: the Whittle likelihood approximates the full Gaussian likelihood:

$$\ln \mathcal{L}(\boldsymbol{d}|\theta, M) = -\frac{1}{2}\boldsymbol{r}(\theta)^T \Sigma^{-1} \boldsymbol{r}(\theta) - \frac{1}{2}\ln\big((2\pi)^N |\Sigma|\big)$$

Where:

$\boldsymbol{r}(\theta) = \boldsymbol{d} - \boldsymbol{\mu}(\theta)$ is the time-domain residual

Σ is the noise covariance matrix

A Gaussian Process (GP) introduces a kernel with hyperparameters α to model the covariance:

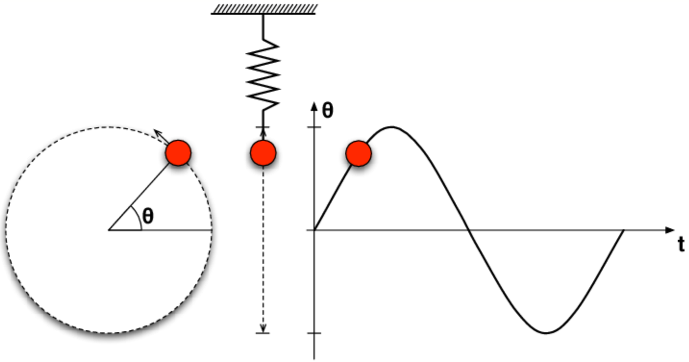
$$\Sigma \to \Sigma_{mn}(\alpha) = k(t_m, t_n; \alpha)$$

Then the GP likelihood is:

$$\ln \mathcal{L}(\boldsymbol{d}|\theta, \alpha, M) = -\frac{1}{2}\boldsymbol{r}(\theta)^T \Sigma(\alpha)^{-1} \boldsymbol{r}(\theta) - \frac{1}{2}\ln\big((2\pi)^N |\Sigma(\alpha)|\big)$$

GPBilby

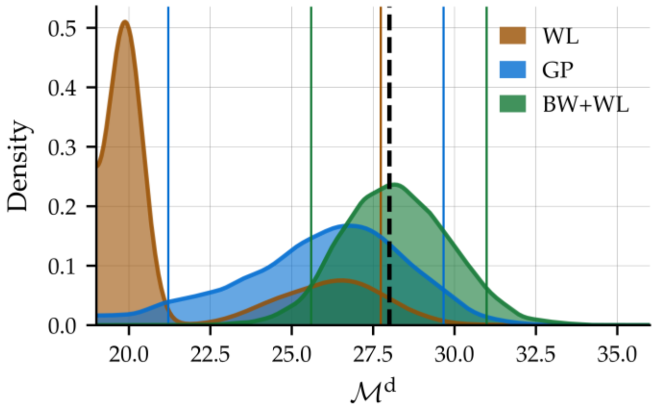
- Ashton et al. (2022): extension of the Bilby Bayesian inference library
- Uses the celerite GP package (Foreman-Mackay 2017)
- Construct kernels from mixture-model of simple harmonic oscillators
- Pre-whiten the data to avoid modelling the coloured noise

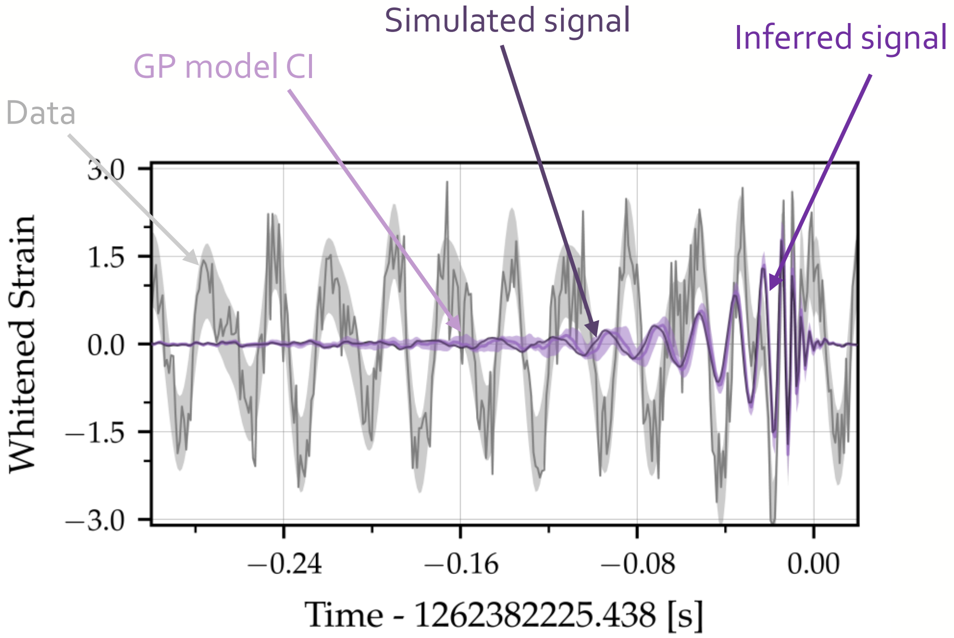


Egri-Nagy et al. (2010)

Validation of GPBilby

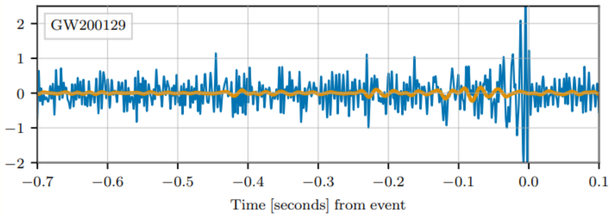
- Simulated signals contaminated by glitches
- WL: no de-glitching
- BW+WL with de-glitching
- GP: joint analysis

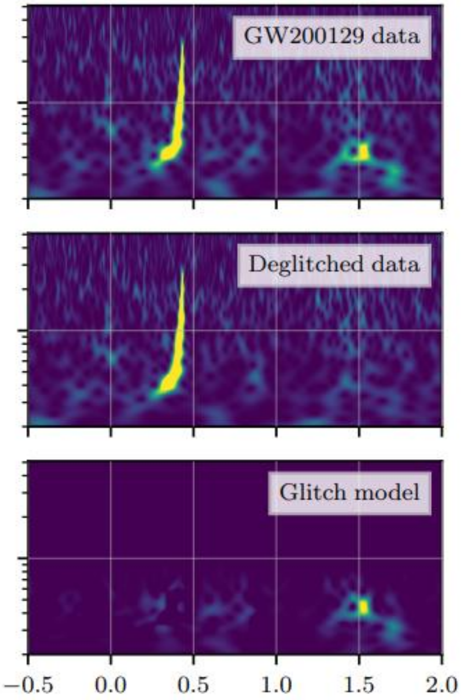




GW200129: glitch contamination

- GW200129: First claimed event with evidence for GR precession (Hannam et al. 2022) and measurement of large recoil kick (Varma et al. 2022)
- Event is glitch-contaminated, Payne et al 2022 conclude the evidence is dependent on the glitch model
- Looking directly at the glitch model Davis et al. (2022):



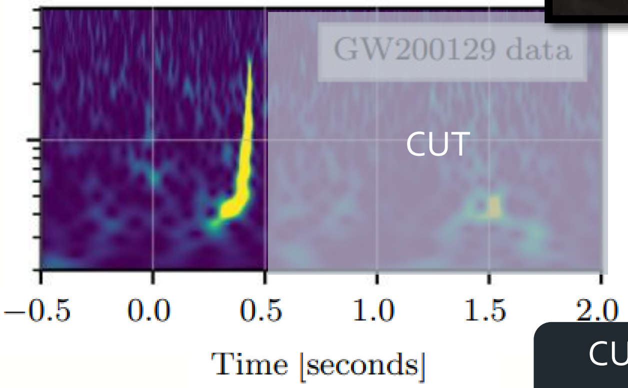


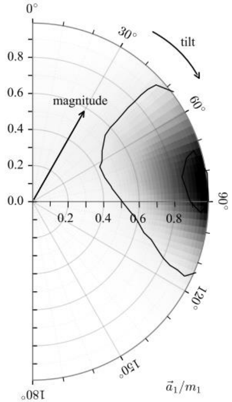
- There is a "wiggle" in the inspiral...

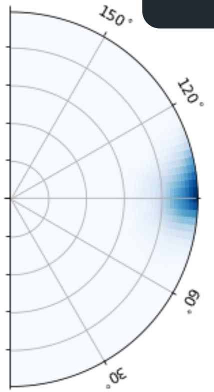
Work by Mattia Emma

GW200129: reanalysis

- Re-analyse the raw strain data with GPBilby
- Perform two analyses:
  - No-CUT: use the full data span
  - CUT: remove all data from 0.5s onwards (i.e. the visual glitch)
- Main takeaway: evidence of precession is robust to glitch treatment and choice of data
- Note: see also (arXiv:2311.09921) on improved non-linear glitch models from auxillary channels



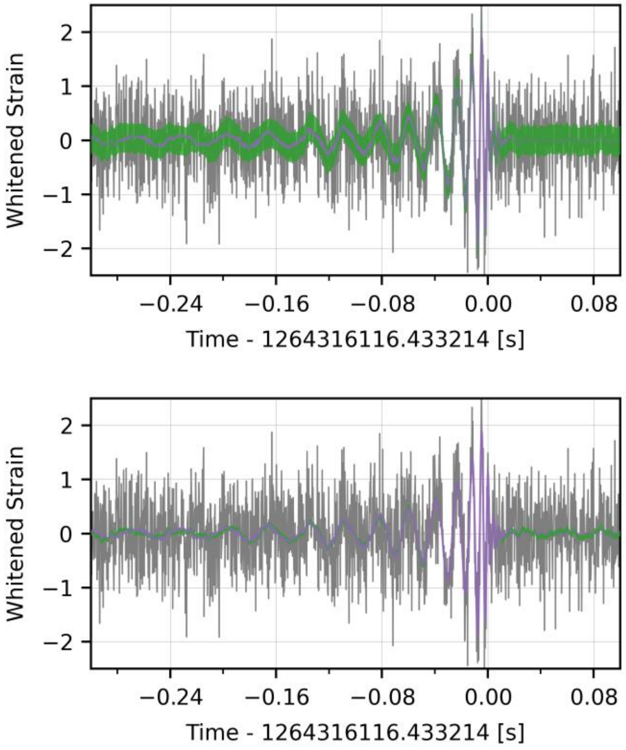




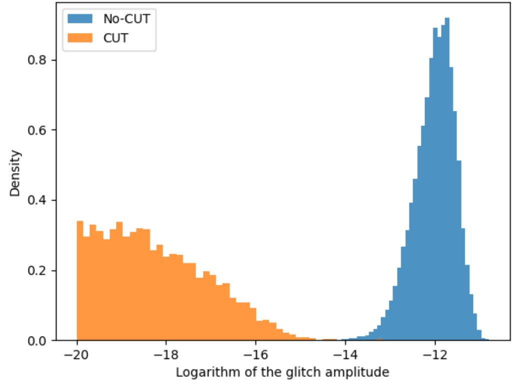
CUT spin disk

GW200129: confirmation

Whitened Data
Glitch model
Signal



The CUT analysis finds no glitch power and we do not observe any excess noise model in the inspiral

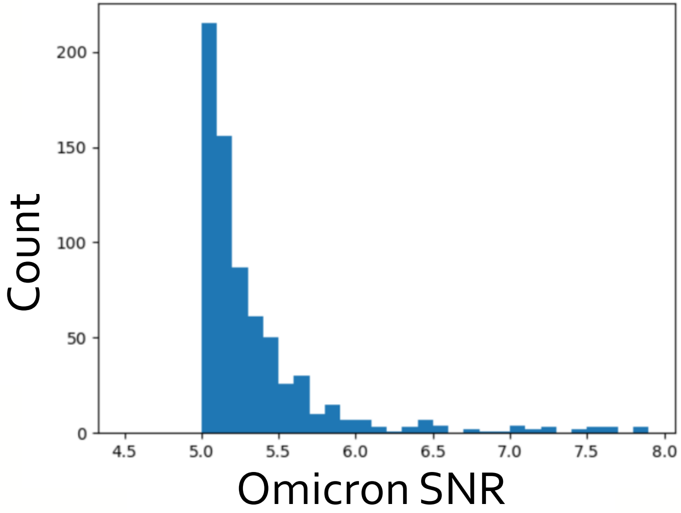


Challenge: glitches

A-priori probability of contamination ($\approx 1/15$) is mismatched with actual number of glitch-contaminated signals in GWTC-3 ($\approx 1/5$)

Either:

- Glitch-mitigation is overly conservative
- There are many sub-threshold glitches

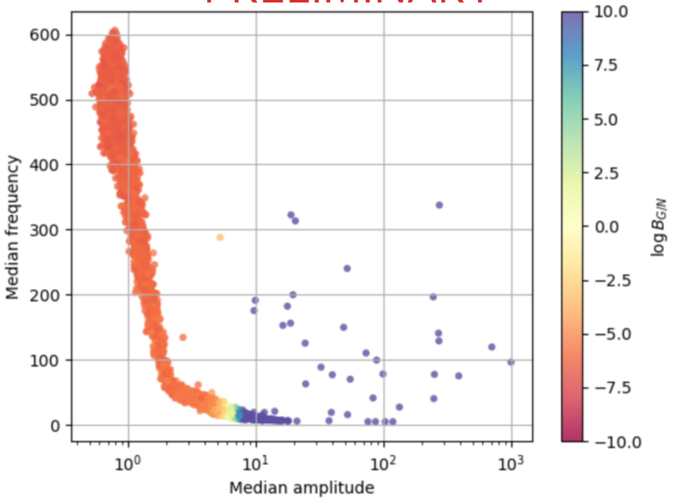


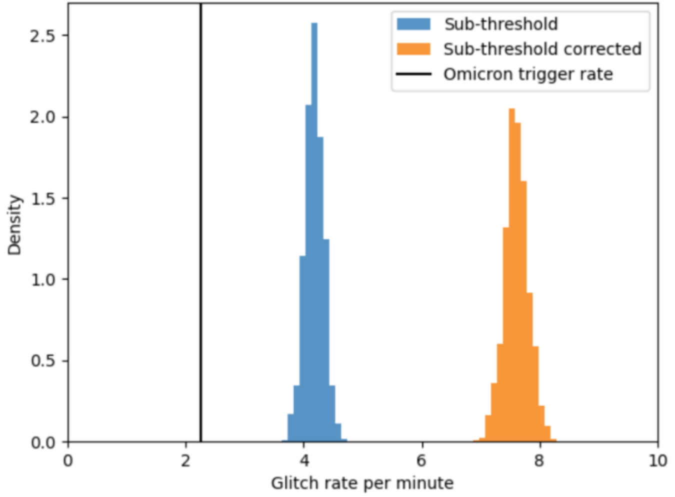
Challenge: glitches

Apply population-style analysis to observe the "population" of glitches (rate and properties)

- Using the "antiglitch" model (Bondarescu et al. 2023) to capture the salient features of a broad range of glitches
- Analyse 1 day of data so far
- We find that:
  - The rate of sub-threshold glitches can be a factor of 4 larger than the Omicron trigger rate
  - Glitch amplitude scales inversely
  - Frequency has distinct "modes"

PRELIMINARY





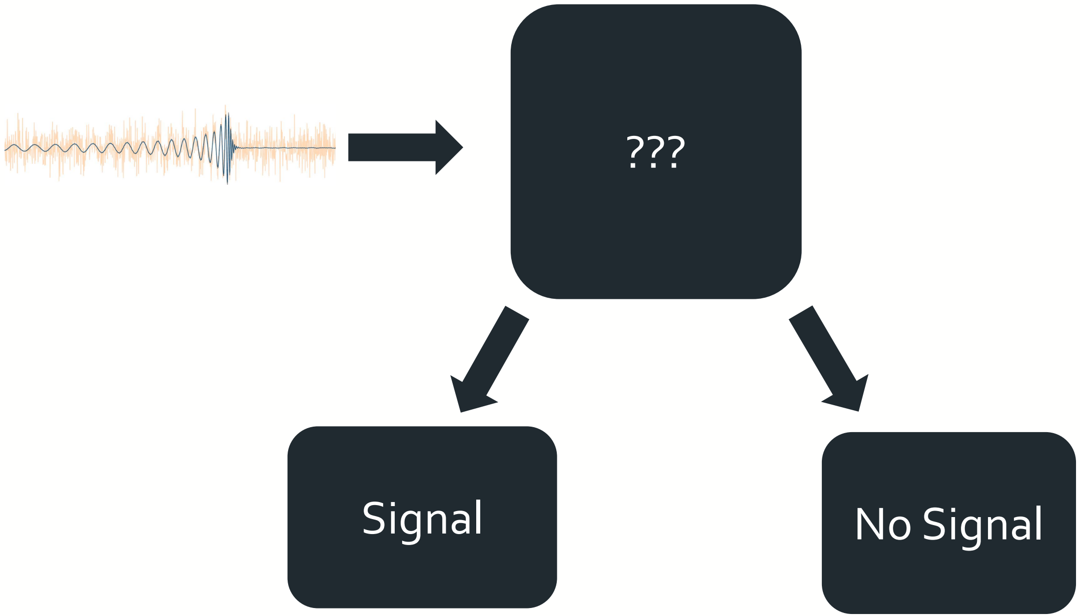
Conclusion on glitches

- More needs to be done to:
  - Understand and mitigate glitches
  - Analyse glitch-contaminated signals

Challenge: understanding search pipelines

Challenge: understanding search pipelines

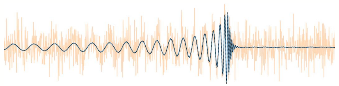
Search pipelines:



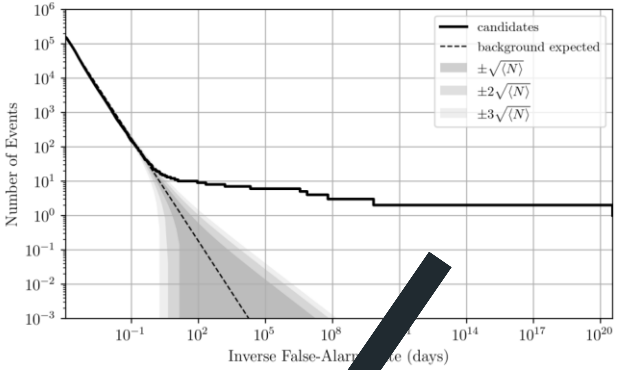
Challenge: understanding search pipelines

Ewing et al (2023)

Search pipelines:



???



Loudest candidate:

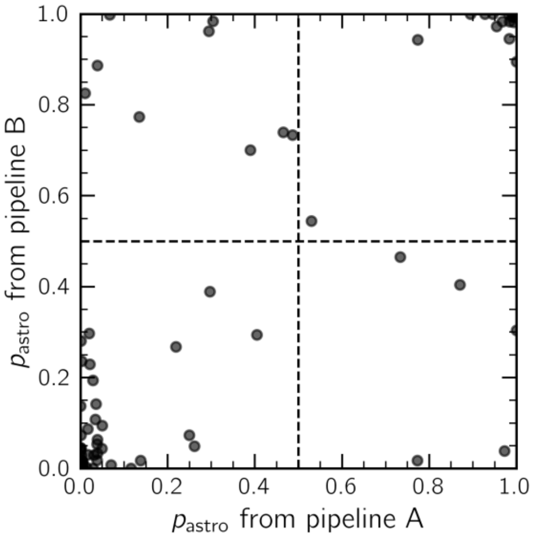
- FAR: 1/10 years
- p_{astro}: 90%

Context and decision procedure needed to interpret the candidate significance

Challenge: understanding search pipelines

Search pipelines don't agree:

- Different underlying assumptions
- Different models of the noise
- Intrinsic uncertainty on the significance



Non-experts often **do not** understand the differences between pipelines

Per-Pipeline Event Information

UID	Group	Pipeline	Search	gpstime	FAR (Hz)
G488676	CBC	CWB	BBH	1401177186.465	1.105e-06
G488669	CBC	MBTA	AllSky	1401177186.456	1.417e-08
G488673	CBC	gstlal	AllSky	1401177186.463	3.744e-12

Using Conformal Prediction to calibrate search pipelines

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Conformal prediction

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Use Conformal Prediction (CP) to estimate uncertainty for search pipelines

- CP: A technique developed in Machine Learning settings to augment a point-prediction with a **prediction interval**
- For example, take an image classifier:

Black-box Machine Learning classification algorithm

Squirrel (0.82)

Conformal prediction

AIM: Use Conformal Prediction (CP) to estimate uncertainty for search pipelines

- CP: A technique developed in Machine Learning settings to augment a point-prediction with a **prediction interval**
- For example, take an image classifier:

α is the user-defined error rate

Γ^{α} =[Squirrel, tree, fox]

Conformal Prediction

Black-box Machine Learning classification algorithm

Γ^{α} =[Squirrel, tree, fox]

Γ^{α} is the "prediction set"

What is a guarantee?

For some test data X, if it is **exchangeable** with the calibration data (of size n):

$$1 - \alpha \leq P\left(\hat{Y} \in \Gamma^{\alpha}\right) \leq 1 - \alpha + \frac{1}{n+1}$$

Where $\hat{Y}$ is the true label

If n is sufficiently large

$$P\left(\hat{Y} \in \Gamma^{\alpha}\right) \sim 1 - \alpha$$

How does CP work?

- We have images that come from K classes
- Determine an acceptable error rate α
- Define a conformity score

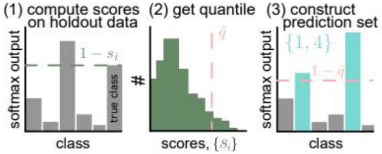
$$f(X) \in [0,1]^K$$

- Classification data $(X_1, Y_1) \dots (X_n, Y_n)$ where X is the data and Y the true class
- Iterate through classification data and calculate scores:

$$s_i = 1 - f(X_i)_{Y_i}$$

- Define $\hat{q}$ as (essentially) the $1-\alpha$ quantile of s_i
- Finally for new data X, construct the prediction set:

$$\Gamma^\alpha = \{y : A(X)_y \geq 1 - \hat{q}\}$$




```
# 1: get conformal scores. n = calib_Y.shape[0]
cal_smx = model(calib_X).softmax(dim=1).numpy()
cal_scores = 1-cal_smx[np.arange(n),cal_labels]
# 2: get adjusted quantile
q_level = np.ceil((n+1)*(1-alpha))/n
qhat = np.quantile(cal_scores, q_level, method='higher')
val_smx = model(val_X).softmax(dim=1).numpy()
prediction_sets = val_smx >= (1-qhat)  # 3: form prediction sets
```


arXiv:2107.07511

In human terms

Ask the algorithm, “is this a fox squirrel”?

Without uncertainty:

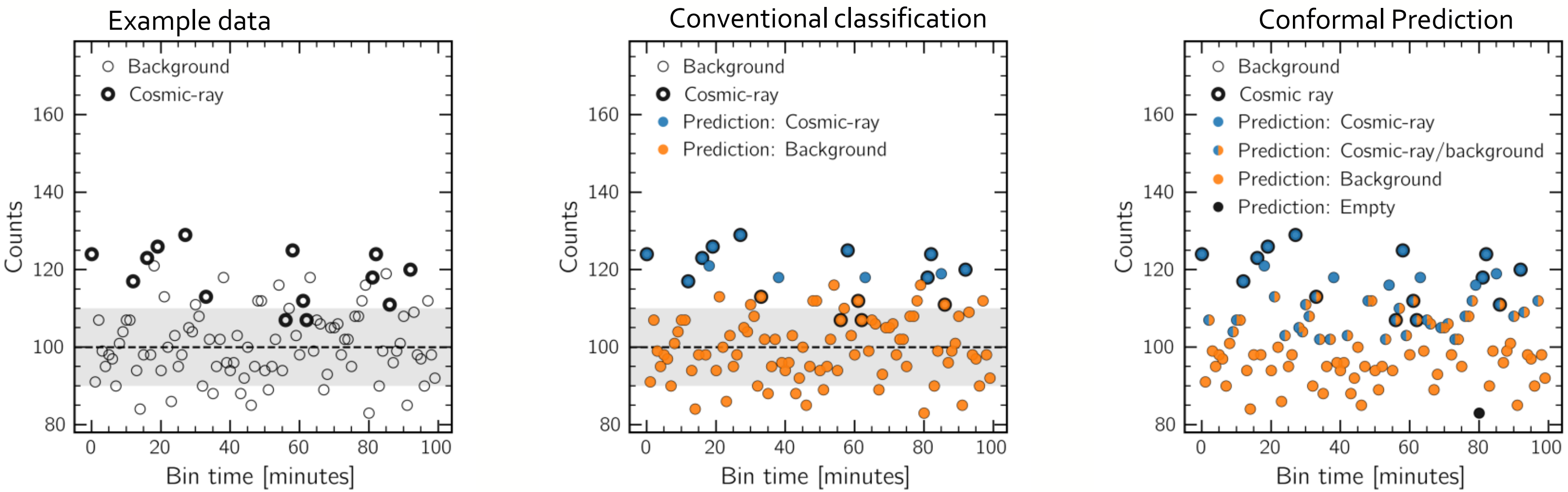
- A binary answer “yes/no”

With uncertainty:

- Yes, that is the only thing it could be!
- Maybe, but it also looks like a gray fox, a bucket, and a rain barrel
- (Not pictured): No, this doesn’t look like a fox squirrel at all

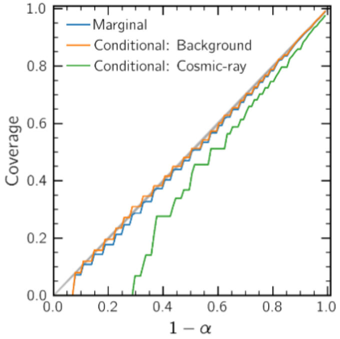
Calibrating search algorithms with conformal prediction

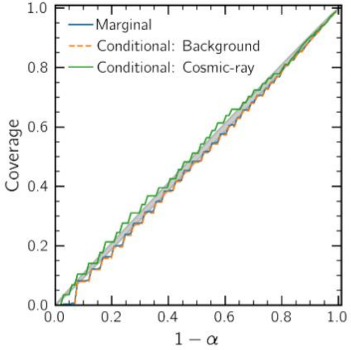
- Ashton et al. (2024) consider application of CP to binary search classification
  - Provides uncertainty estimate to single-pipeline predictions
  - Conformal prediction is self-calibrating
  - Can be applied to any classification algorithm with needing to understand internal details
  - Provides a framework to combine multiple pipelines together improving overall sensitivity



Checking everything works

- **Coverage**: the fraction of events for which the true label is in the prediction set
- Validate performance on a test set
- For standard algorithm, conditional labels are not guaranteed
- Need Mondrian Conformal Prediction (increases calibration error due to smaller n):





Application to Gravity Spy

True label ($\hat{y}$)	Gravity Spy prediction	CP prediction set
Blip	Blip	[Blip]
Koi Fish	Koi Fish	[Blip, Koi Fish]
Tomte	Koi Fish	[Blip, Koi Fish, Tomte]

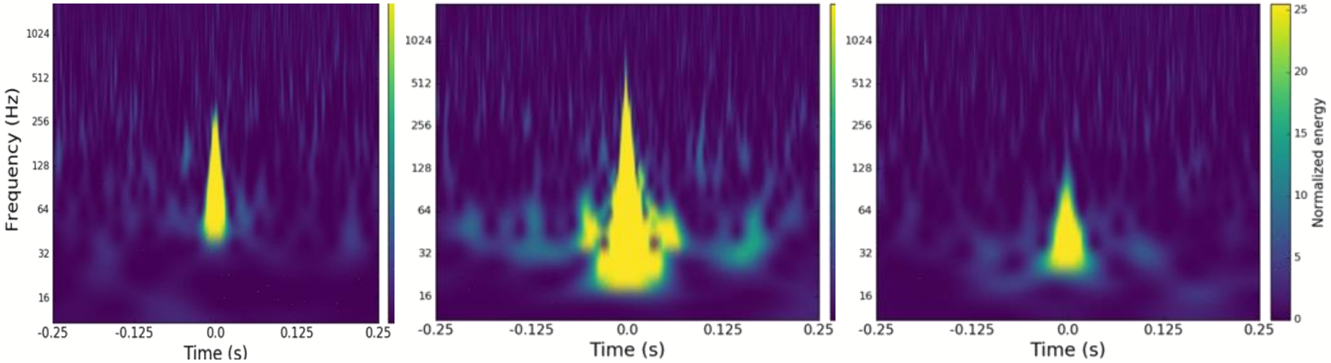


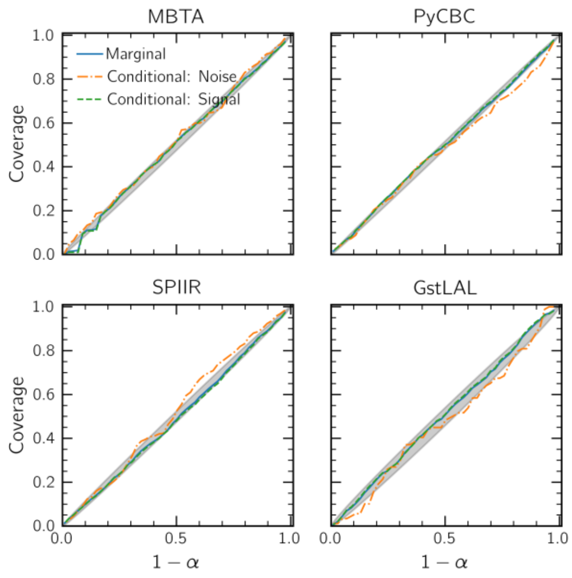
Figure 1: Reference glitch plots, left to right: 'Blip', 'Koi_Fish', 'Tomte'. From Gravity Spy (Zevin et al, 2017)

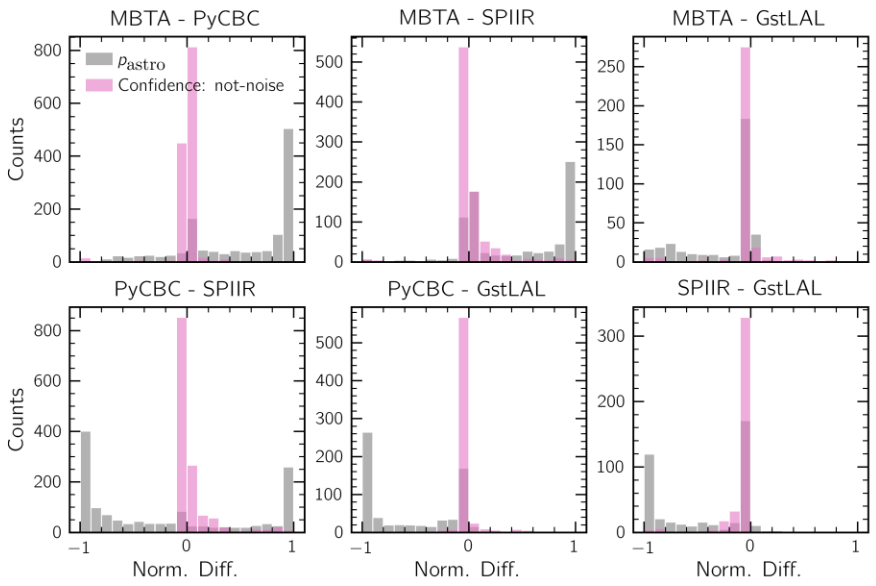
Work by Ann Malz

Application to gravitational-waves: MDC

Using Mock Data Challenge (MDC) data and the FAR as the conformity score

$$f(X) \rightarrow f(\mathrm{FAR})$$



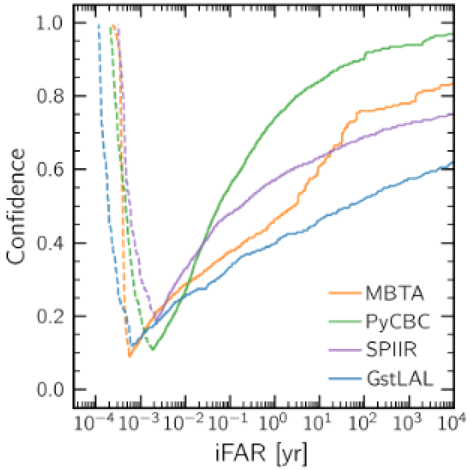


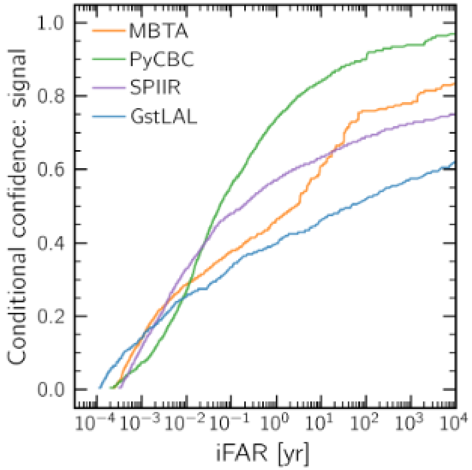
The **confidence** is used to assess significance

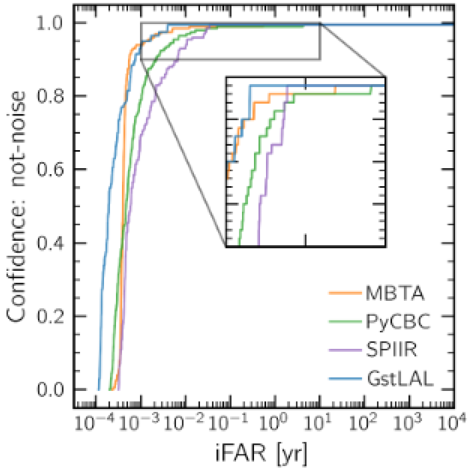
Definition 1 *The confidence is the value of α such that the size of Γ^α changes from 1 to 2 (i.e. the point where we go from the single to the double label).*

Definition 2 *The conditional confidence in label y is the minimum value of α such that $y \in \Gamma^\alpha$.*

Definition 3 *The not-noise confidence is the minimum $1 - \alpha$ such that the noise label is not included in Γ^α.*







Future potential for CP

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Enables automatic-calibration of the FAR
- Can regulate differences between pipelines
- Future potential to combine pipelines

$$f\,(\mathrm{FAR}) \rightarrow f\,(\mathrm{FAR}_A, \mathrm{FAR}_B, \ldots)$$

  - Currently we apply a simple combination algorithm
  - Enables leveraging pipeline strengths to improve overall performance
  - Provides opportunity to model and infer pipeline behaviour

Thank you for listening!

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Can you include glitches in your population?

- Choose a threshold that includes glitches
- Model the glitches
- Heinzel et al. (2023)

Likelihood for standard hierarchical population inference

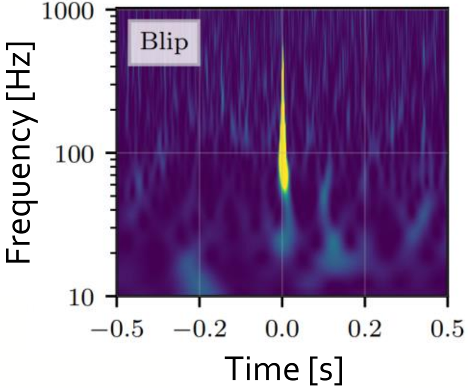
$$\mathcal{L}(\{d_i\}|\Lambda) \propto \prod_{i=1}^{N_{\text{events}}} \frac{\int d\theta \mathcal{L}(d_i|\theta) p_A(\theta|\Lambda)}{\alpha(\Lambda)}$$

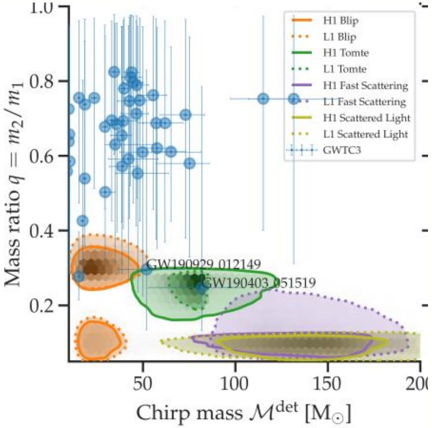
Likelihood for glitch-robust hierarchical population inference

$$\prod_{i=1}^{N_{\text{events}}} \frac{\eta \int d\theta \mathcal{L}(d_i|\theta) p_A(\theta|\Lambda_A) + (1-\eta) \int d\psi \mathcal{L}(d_i|\psi) p_G(\psi|\Lambda_G)}{\eta \alpha_A(\Lambda_A) + (1-\eta)\alpha_G(\Lambda_G)},$$

Developing a glitch model

- Initial study uses the blip population from Ashton et al. (2022)
- Characterized four glitch classes by their **astrophysical population imprint**
- Blip glitches:
  - Chirp mass of ~ 25
  - Bimodal mass ratio's, $q \sim 0.1$ and 0.3

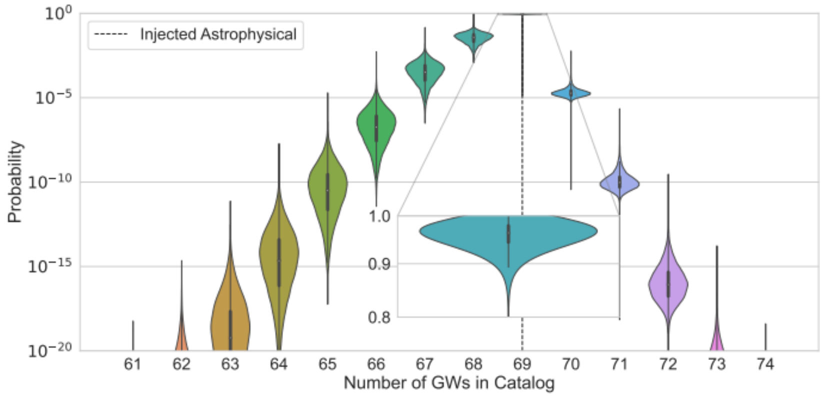


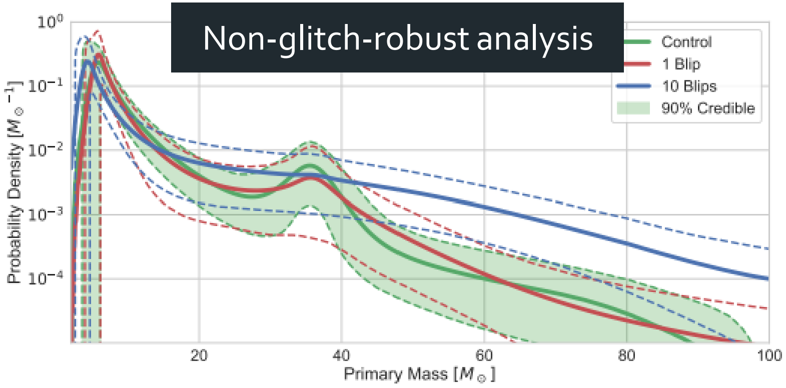


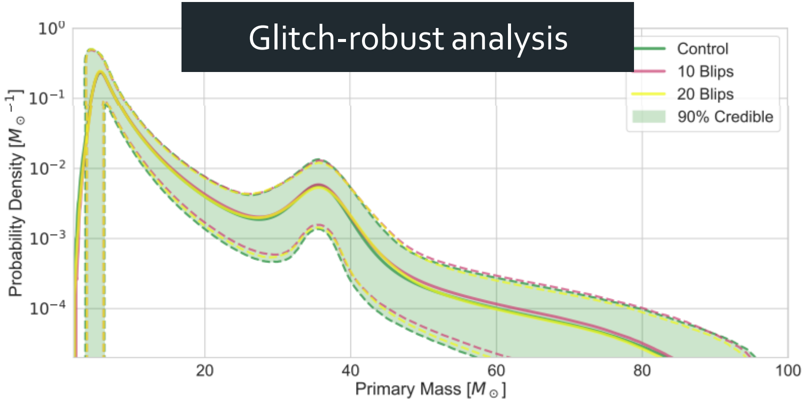
Glitch-robust population modelling

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Verify performance by contaminating the GWTC-3 catalogue with known blip glitches
- Can extract population properties, e.g. the number of GWs in the catalogue.
- Robust inference even when contaminated







Glitch-robust population modelling

ROYAL HOLLOWAY UNIVERSITY OF LONDON

- Still in development:
  - More complete glitch population models needed
  - Improved glitch modelling (i.e., using physical glitch models)
- Will enable population studies to dig down into the noise and extract features
- Quietest signals are at the largest redshift