Waveform consistency tests with cWB-2G

Trieste, 15 may 2024

Sophie

Outline

- waveform consistency test
- O4a preliminary results
- Three detectors network

- Back-up slides: systematic error from previous run

Waveform consistency test - goal

- Unmodeled searches can identify discrepancies between measured data and theoretical models

Coherent WaveBust unmodelled reconstruction of GW candidate events

VS

Compact binaries coalescence models. Used to infer physical parameters of the source

- Discrepancies might be due to noise artifacts, the influence of unknown binary parameters, missing physics in the waveform models, or deviations from General Relativity.

Waveform consistency tests - method

This test computes:

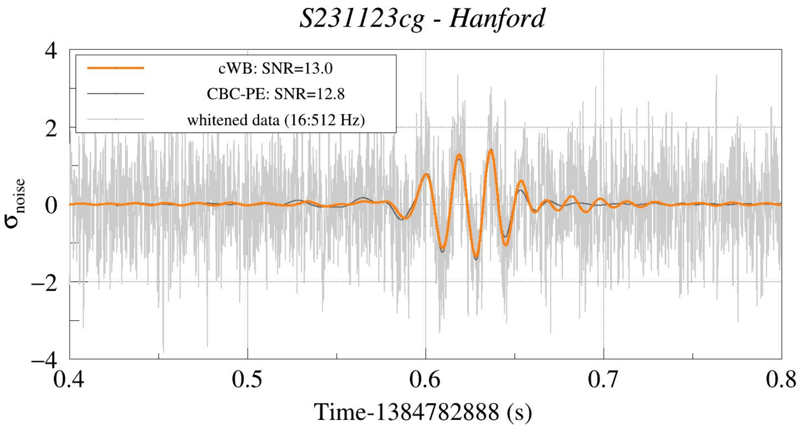
1. The match between the waveform of the event reconstructed by cWB and the maximum likelihood CBC-PE waveform.

$$\mathrm{Match}(x_{rec}(t), x_{inj}(t)) = \frac{(x_{rec}|x_{inj})}{\sqrt{(x_{rec}|x_{rec})}\sqrt{(x_{inj}|x_{inj})}}$$

$$(x_{rec}|x_{inj}) = \sum_{i}^{IFOs} \int_{t_{a,i}}^{t_{b,i}} x_{rec,i}(t) x_{inj,i}(t) dt$$

whitened waveforms

Match in [0,1]



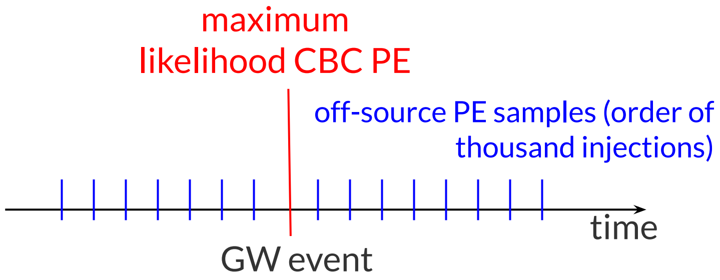
*CBC-PE : parameter estimates inferred using compact binary coalescence models

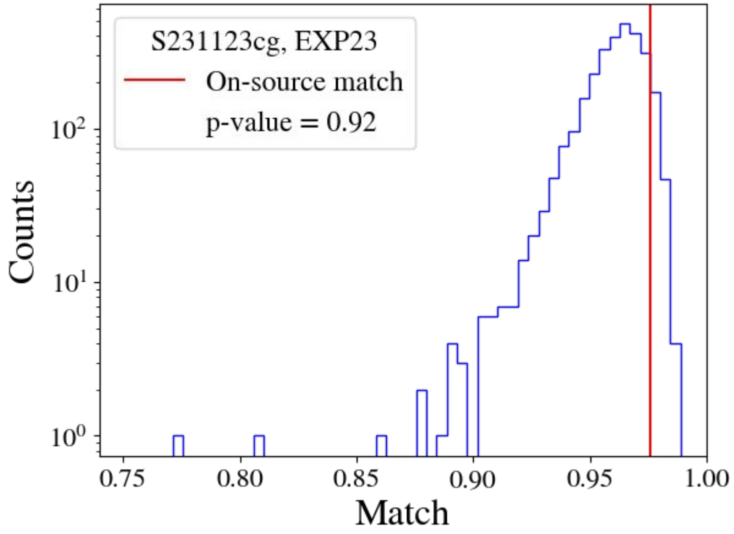
Salemi, F., et al. "Wider look at the gravitational-wave transients from GWTC-1 using an unmodeled reconstruction method." *PRD* (2019)

Waveform consistency tests - method

This test computes:

1. The match between the waveform of the event reconstructed by cWB and the maximum likelihood CBC-PE waveform.
2. The match distribution between injected CBC-PE samples and their cWB reconstruction.
3. p-value that quantifies if the discrepancy between the cWB waveform event and CBC-PE (point 1) is significant against the null hypothesis (point 2).



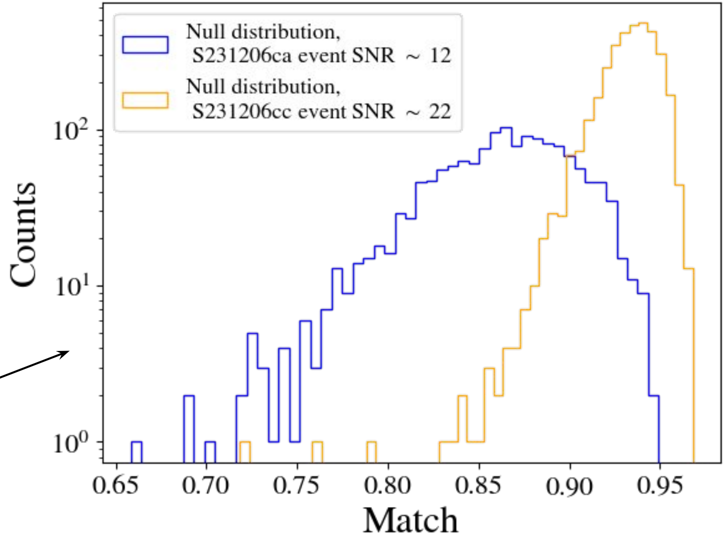


Salemi, F., et al. "Wider look at the gravitational-wave transients from GWTC-1 using an unmodeled reconstruction method." *PRD* (2019)

Waveform consistency tests - match vs SNR

- This procedure accounts both for detector noise, the uncertainty of cWB reconstruction, and the PE variability.

- The lower the SNR of the event under investigation the broader the null distribution



Null-distributions for S231206ca and S231206cc. Similar chirp mass (40.6 vs 36), but different SNR.

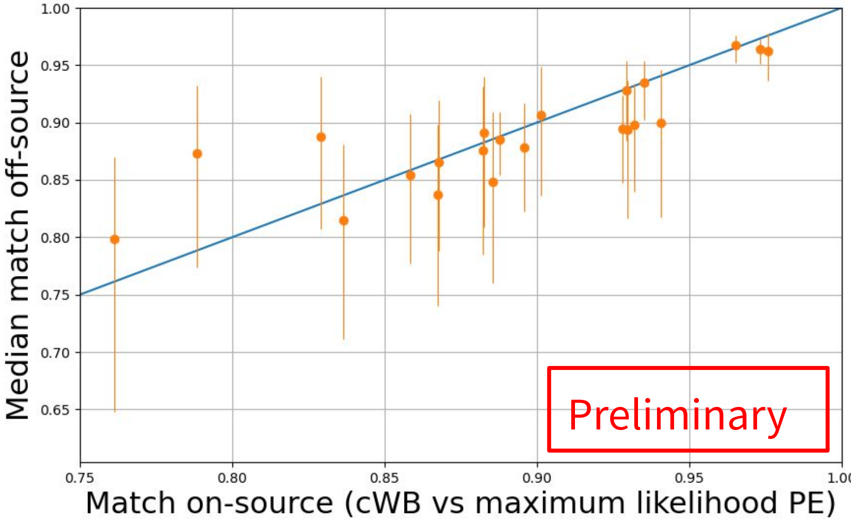
Follow-up of O4a GW events

- follow-up of interesting events wiki
- preliminary p-values distribution on a subset of GW events

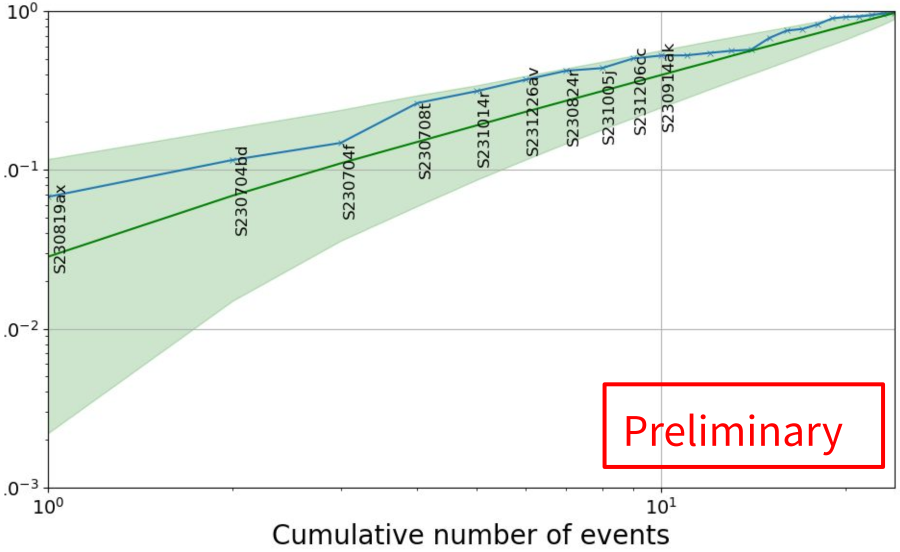
Recent presentations to burst group: DCC G2302408, G2302304,G2301377)

O4a events

Plots as in GWTC-3



Off-source versus on-source match values. The blue line indicates the null hypothesis, and the error bars indicate the symmetric 90% confidence interval.
No significant deviations



Distribution of p-values, and symmetric 90% interval about the median (green). **No significant deviations**

O4a events

Some open questions:

1. which GW events should be analysed? (O4a has about 80 candidate events. Should we apply a threshold in SNR? threshold in the variability of the off-source distribution? only ‘interesting’ events?)
2. do we have a systematic error (p-value often higher than expected)?
3. O4a catalog will have a different structure: likely this analysis will not be included there

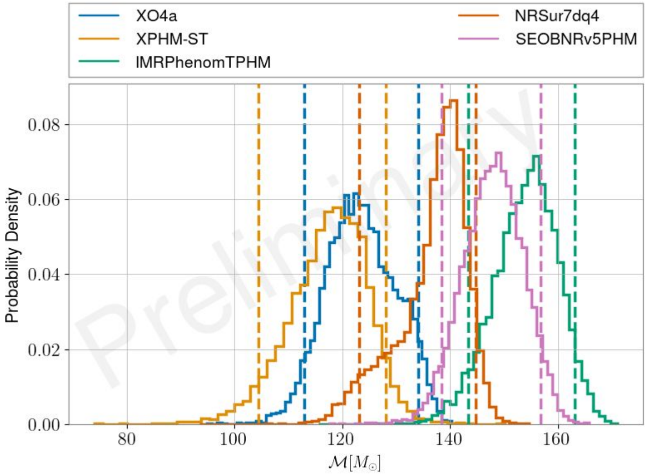
Interesting event - S231123cg

study science case team started two weeks ago git. scope: decide if this candidate event deserves a dedicated paper and understand the results of various analyses.

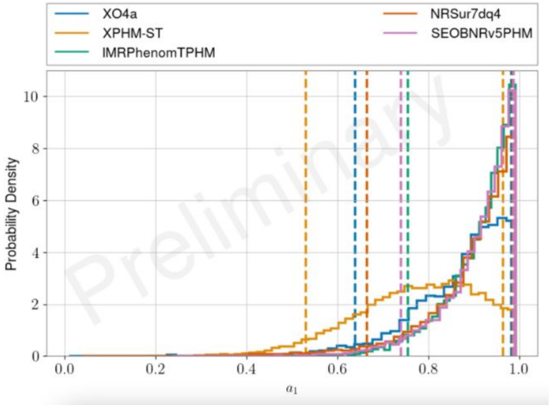
Brief summary:

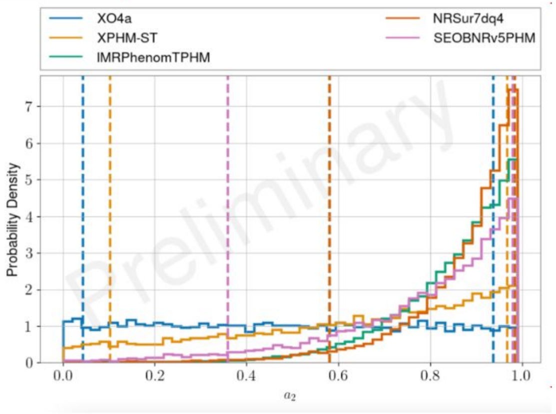
- **detection** : cWB 2G all-sky IFAR>490 years (saturated background link), cWB XP all-sky link IFAR = 481.8 yr , XP BBH IFAR >= 4581.1 year link
- **data quality** : no concern for detection, glitch [6,4]s before in H1 before than has been subtracted by BayesWave for PE link
- **parameter estimation**: **high mass, high spin**, inconsistencies between different waveform models (not solved) LVK slides
- **ringdown**: debate if there is evidence for multiple modes

Interesting event - S231123cg



Detector frame total mass differs across approximants.



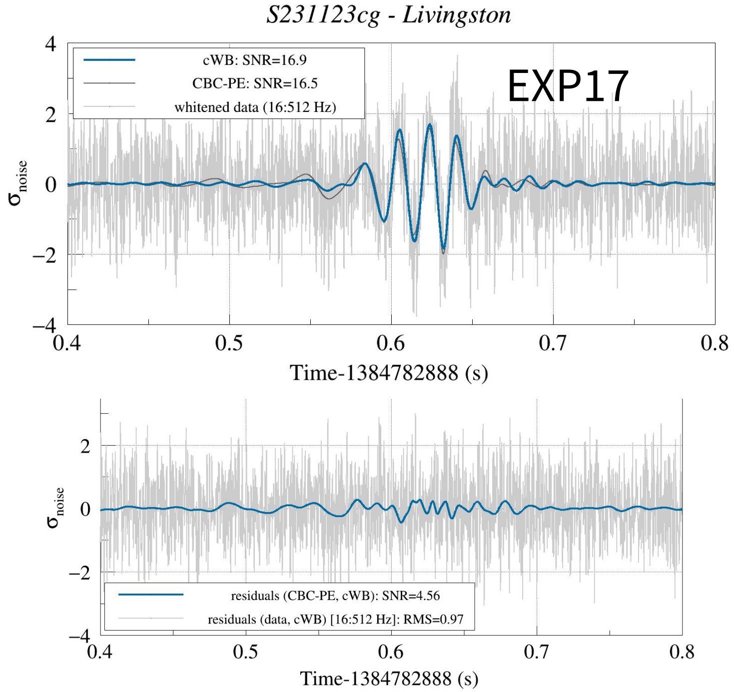


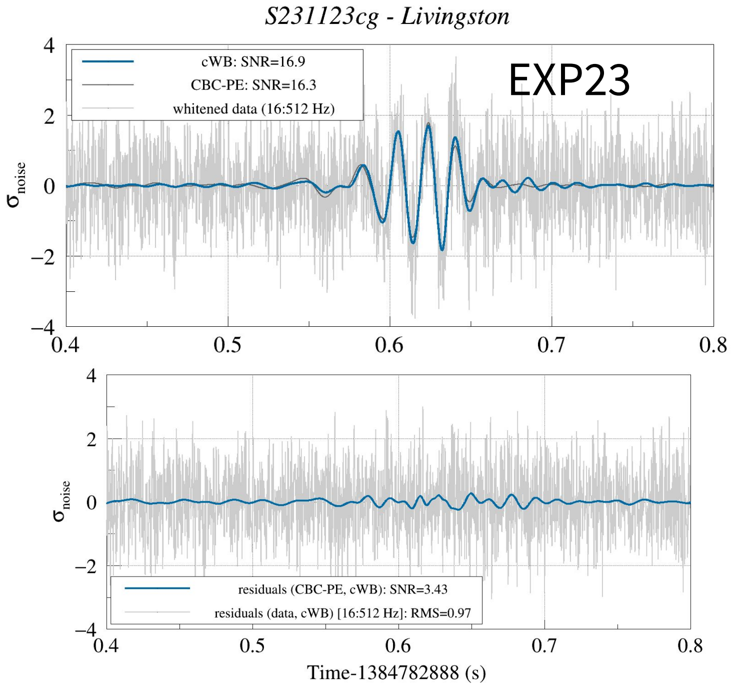
High individual spins preferred by most models. We see some differences in the measurements with the FD models

from [dcc]

S231123cg - waveform consistency

- Several PE runs, **no** significant discrepancy found with latest PE using different waveforms models (IMRPhenomXPHM EXP23, NRSur7dq4 EXP17) wiki





S231123cg

can we provide more info / strength to the detection?

1) increase background?
2) morphological comparison with loudest background triggers? (now only by eyes)
3) to be done: comparison with final PE

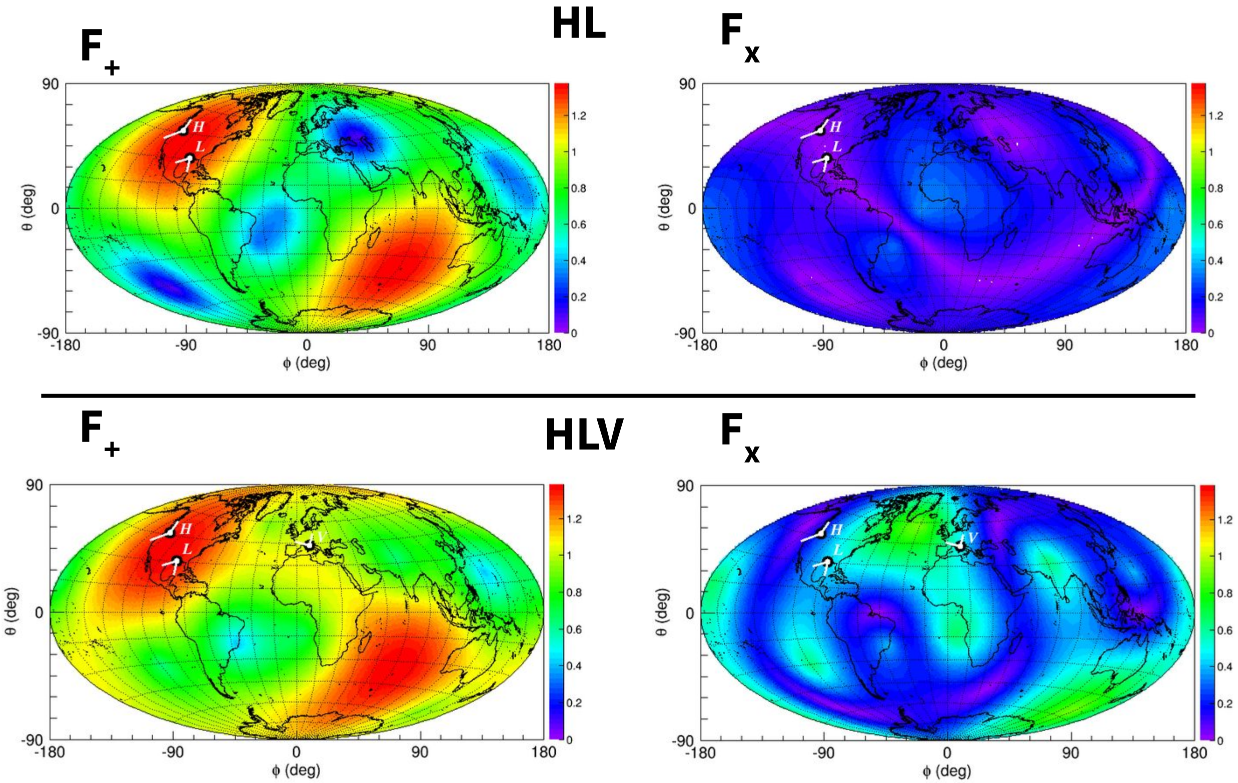
Waveform consistency

three detectors network (LIGO+Virgo)

HL and HLV antenna patterns

LIGO detectors are coaligned, and so they (almost) sense only one GW polarization (F_+)

The addition of Virgo improves the sky coverage and the response to the second polarization (F_x)



Three detectors network

cWB uses likelihood regulators to reject the reconstruction of the GW component NOT observed by the LIGO aligned detectors (F_x) . These regulators successfully reduce the false alarm rate of the HL coherent analysis.

To make full use of a third, not-aligned detectors, the likelihood regulators should be released.

In GWTC3 we report the waveform consistency test using HL network,

does the waveform reconstruction improve using HLV network?

Waveform consistency HL vs HLV

Two examples:

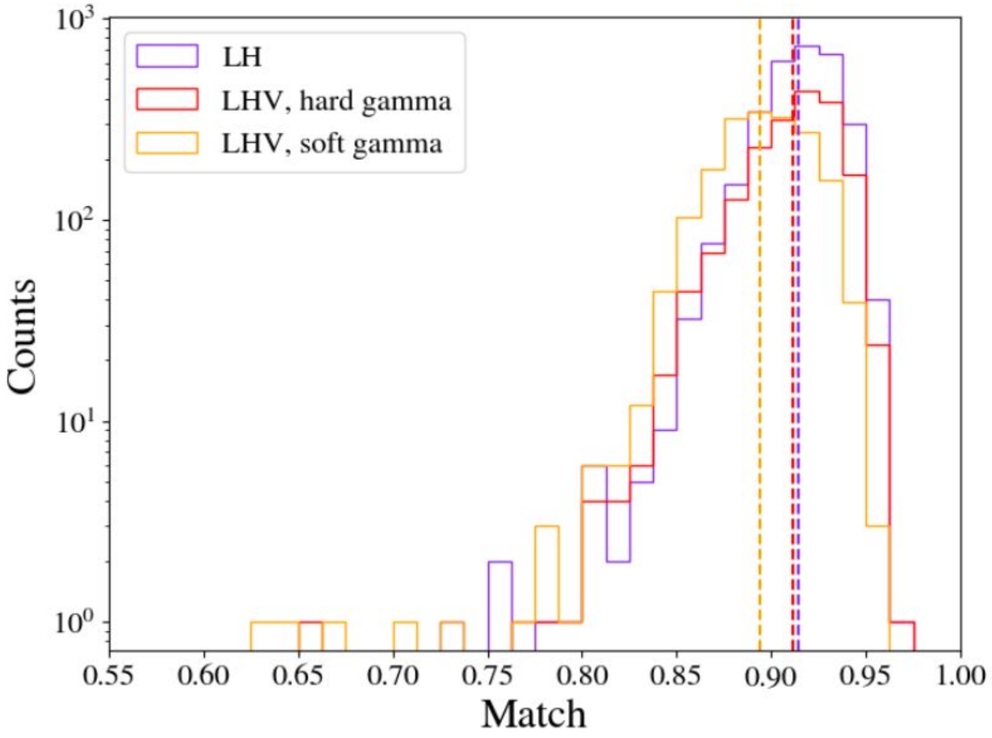
- **GW200224ca** (vanilla BBH, $m_1 = 40$ $m_2 = 32$, cWB SNR=20)
- **GW200311bg** (vanilla BBH, $m_1 = 34$, $m_2 = 27$, cWB SNR=17)(in GWTC3 catalog 200311_115853)

To evaluate cWB goodness of reconstruction, we inject the PE samples of these two BBH events off-source, and we compute the match between the injected waveforms and cWB reconstructions

GW200224ca

off-source match distribution (as a measure of the goodness of the waveform reconstruction)

- LH: highest match mean
- LHV hard regulator: similar matches as LH
- LHV soft regulator: lower mean and has some very low matches (<0.75)



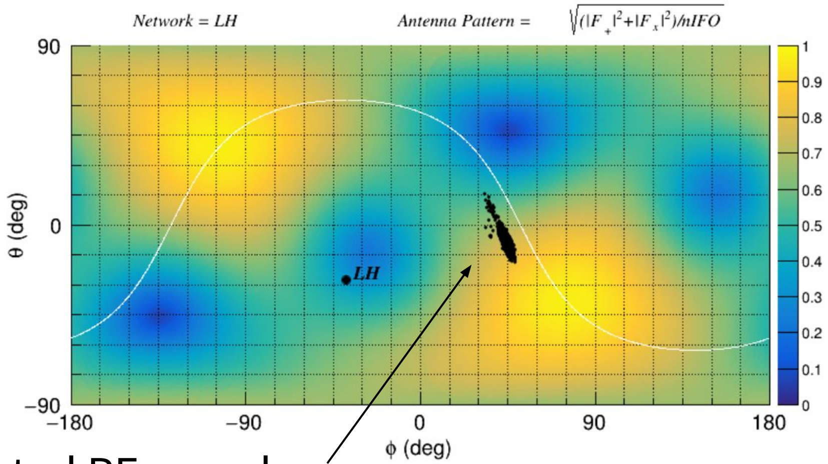
GW200224ca - HL

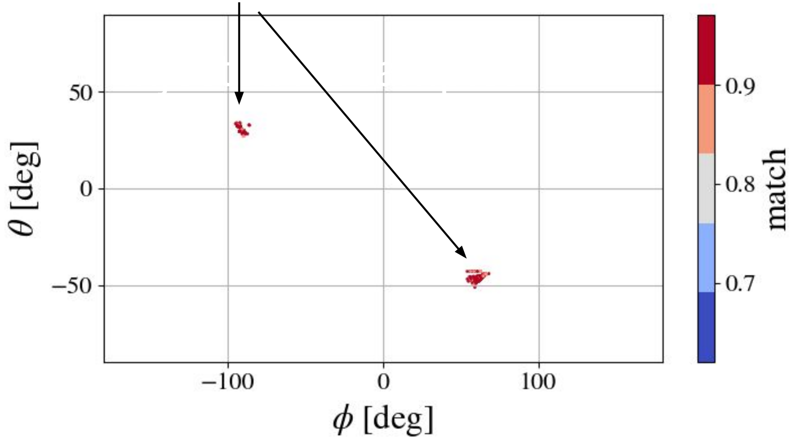
cWB SNR = 20 (sSNR L1 13.4, H1 13.4),

LH waveform consistency match on-source = 0.93, off-source $0.917^{+0.026}_{-0.040}$

ced link, LH off-source report

reconstructed localizations



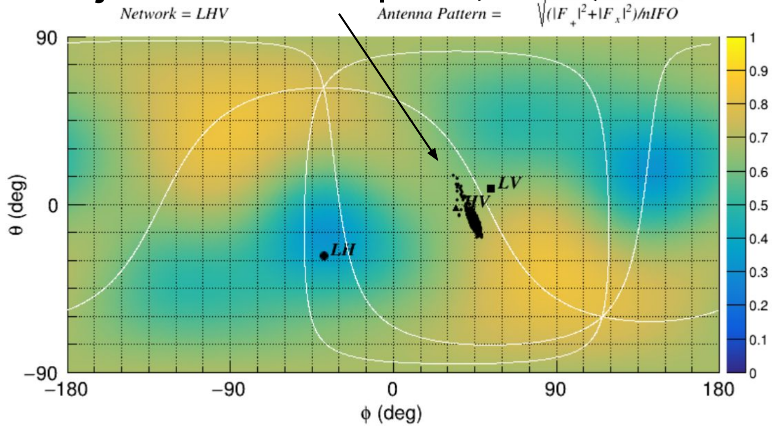


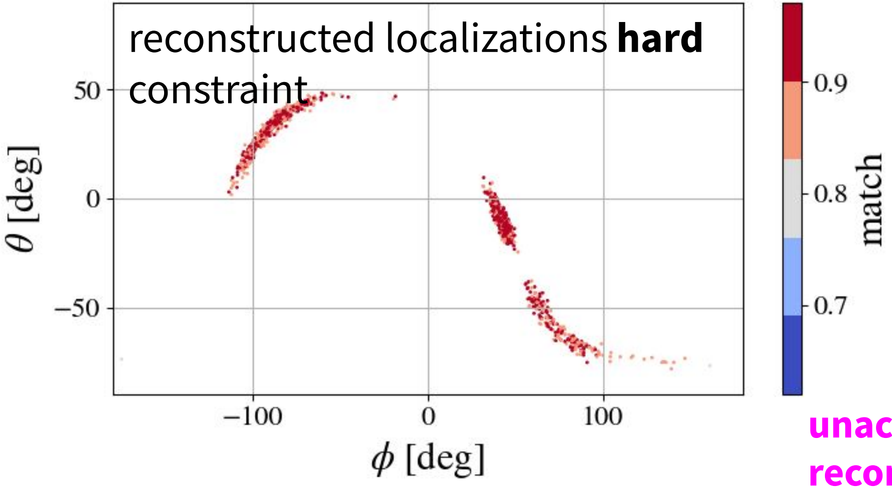
injected PE samples

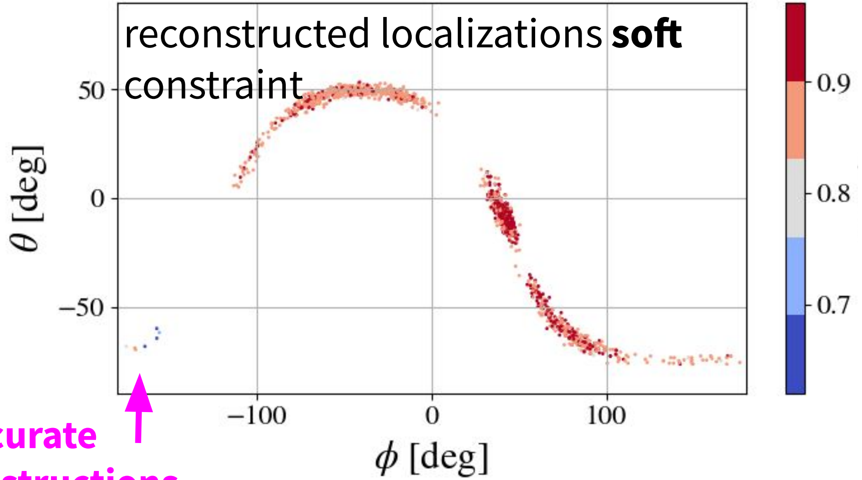
GW200224ca - HLV

- more uniform antenna pattern
- HLV hard $0.915^{+0.028}_{-0.051}$, link
- HLV soft $0.896^{+0.036}_{-0.044\,5}$ link
- sSNR in Virgo .1 with hard, 7.7 with soft

injected PE samples (as HL)



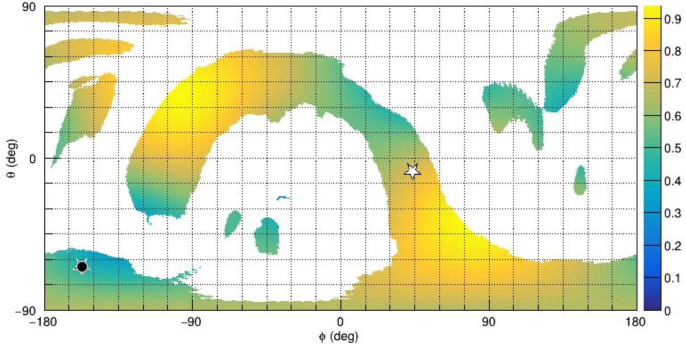


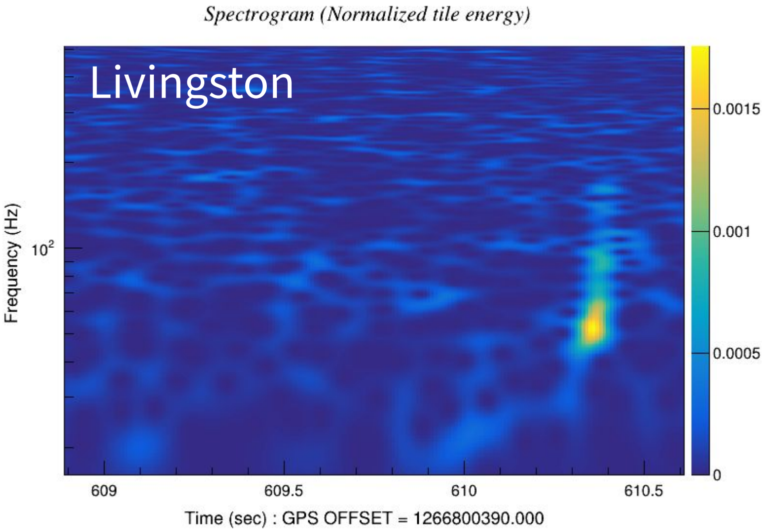


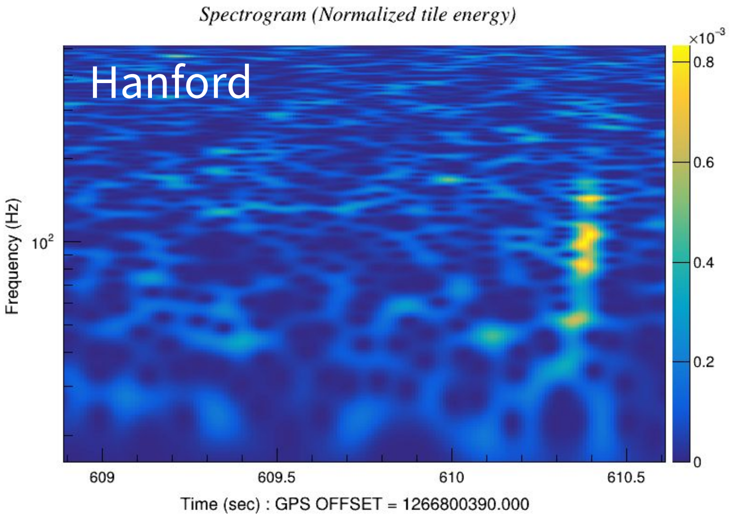
unaccurate reconstructions

GW200224ca - HLV lowest match

Spectrogram of the worst match using soft regulator. There is a low frequency glitch in Virgo ced.

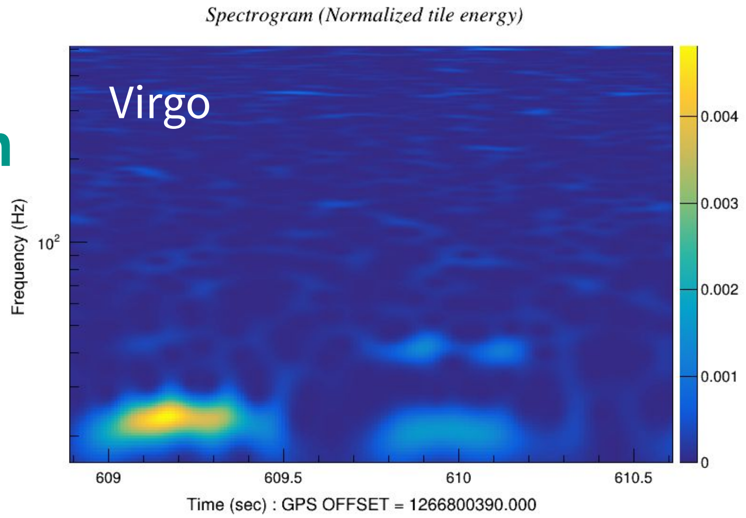




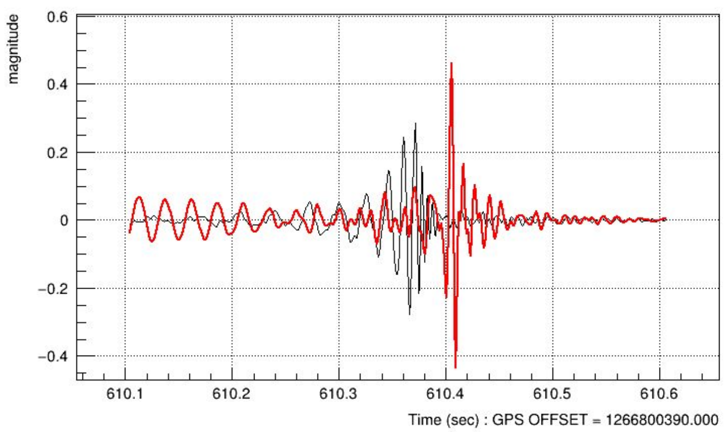


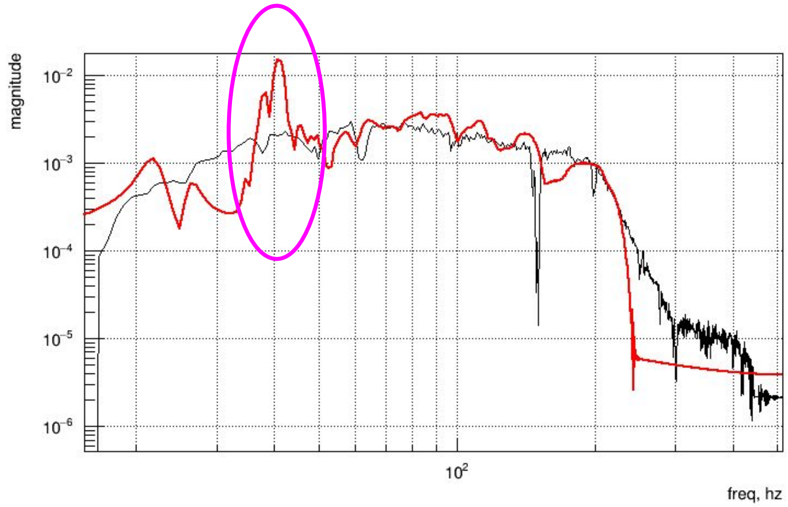
GW200224ca - HLV lowest match

Spectrogram of the worst match using soft regulator. There is a low frequency glitch in Virgo ced



Virgo reconstruction in time and frequency domain.
Black:injected, red cWB reconstruction





GW200224ca - HLV lowest match

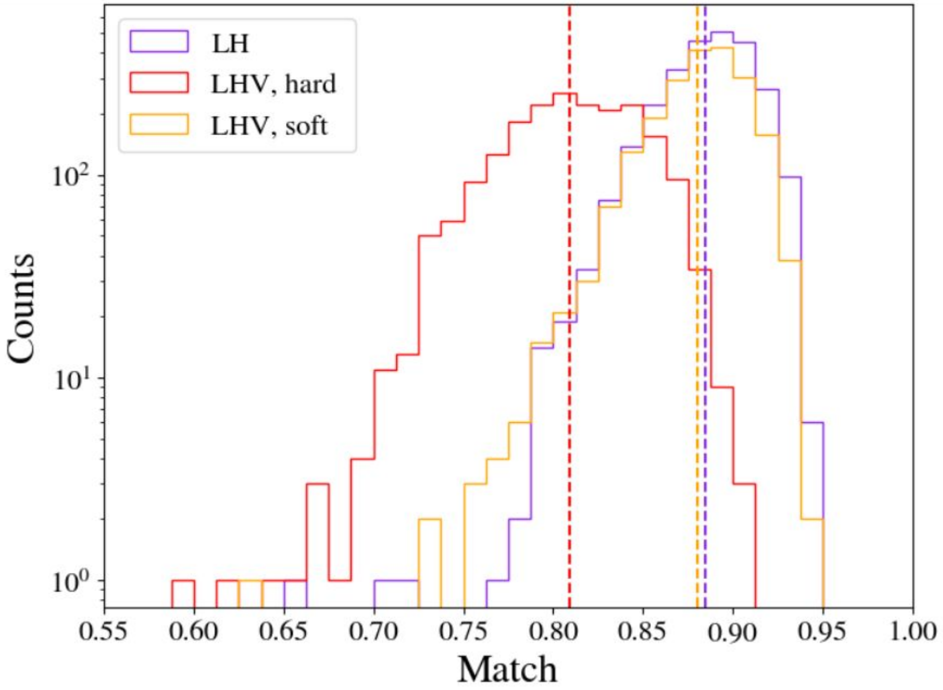
From the study of this event, we found that:

- the distribution of off-source matches does not improve using HLV network (similar or lower mean)
- the HLV network weakens the statistical power of the waveform consistency test giving few very inaccurate waveform reconstructions

GW200311bg

off-source match distribution (as a measure of the goodness of the waveform reconstruction)

- LH: highest match mean
- LHV hard regulator: worst matches
- LHV soft regulator: slightly worse than HL, better than hard regulator

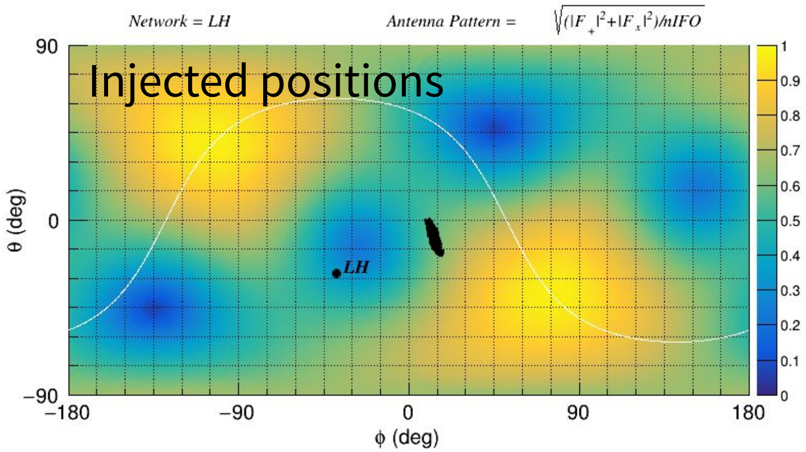


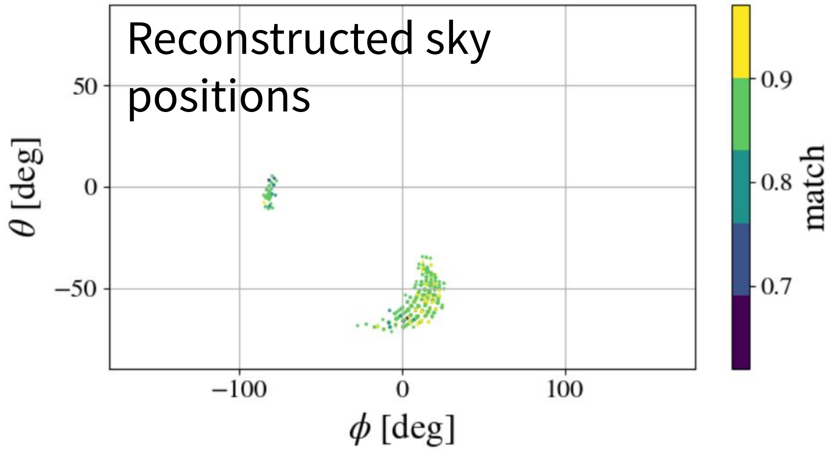
GW200311bg - LH network

cWB SNR =17 (L1 10.7, H1 12),

off-source= $0.888^{+0.035}_{-0.052}$

ced link , off-source report

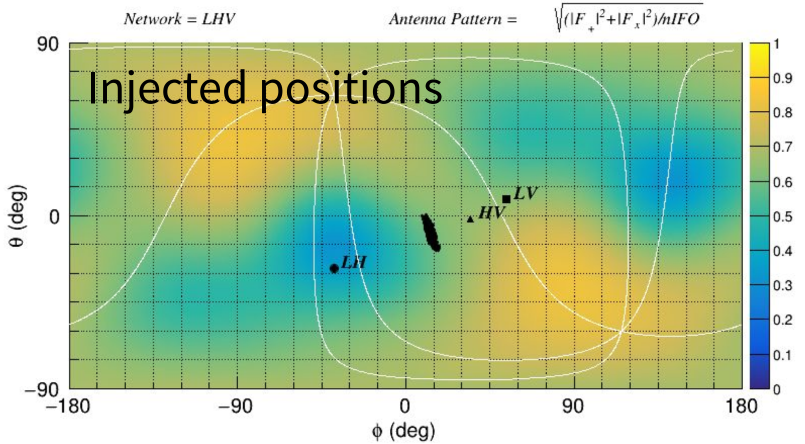


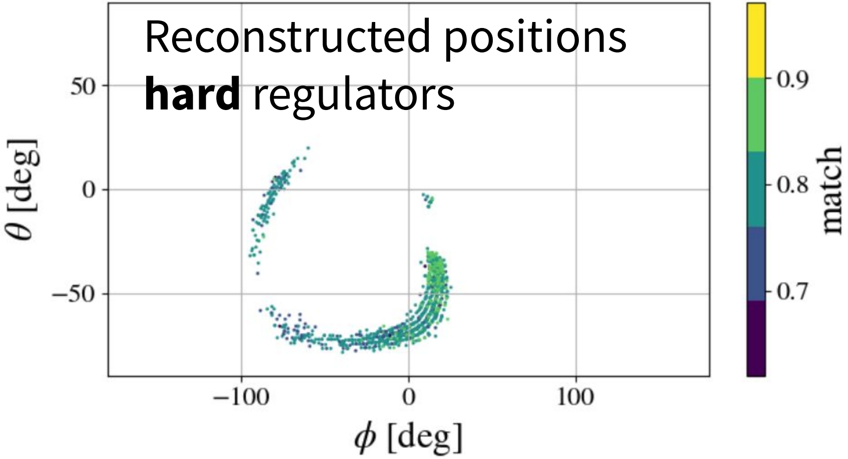


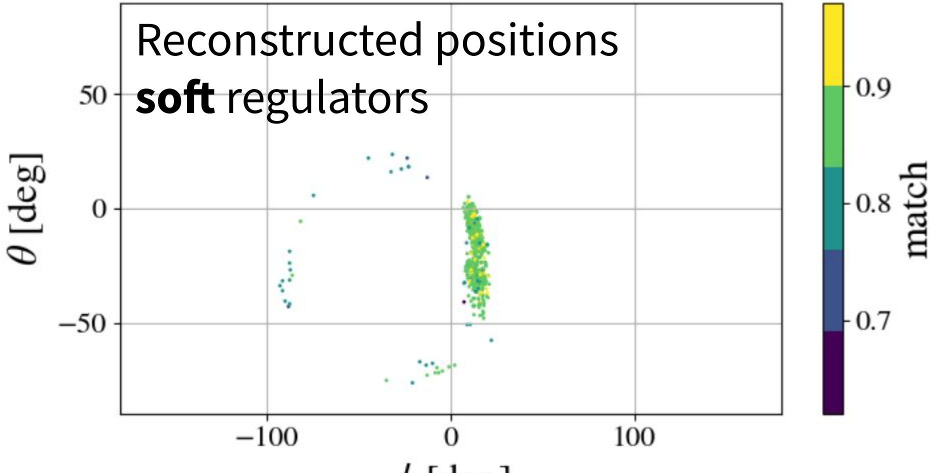
GW200311bg - HLV network

off source matches:

- hard: $0.811^{+0.056}_{-0.07}$
- soft: $0.884^{+0.034}_{-0.054}$







Conclusions

- Three detectors network: we study two GW events from O3. we found that the HLV network does not improve the statistical power of waveform consistency test.

- cWB-2G is performing the waveform consistency test on O4a events. No significant discrepancy between cWB reconstructed waveform and PE samples has been observed. There are open questions.

- S231123cg: could we provide more in depth analyses?

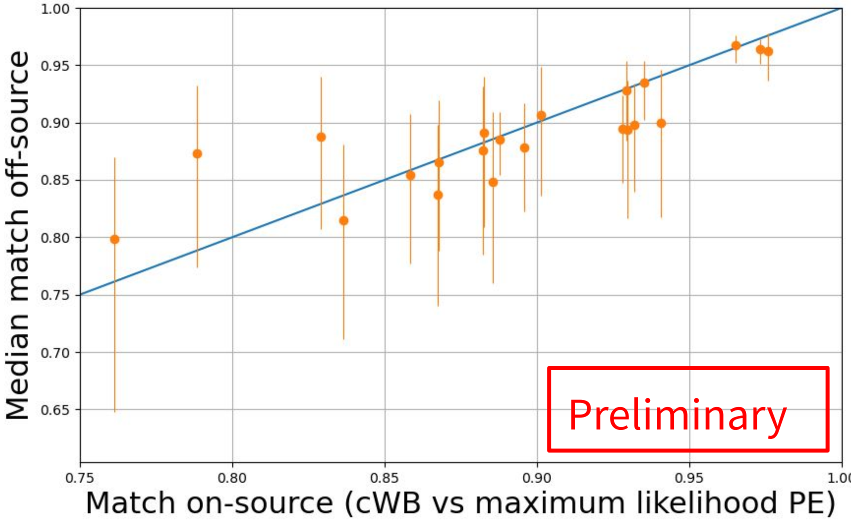
Thanks!

Extra slides

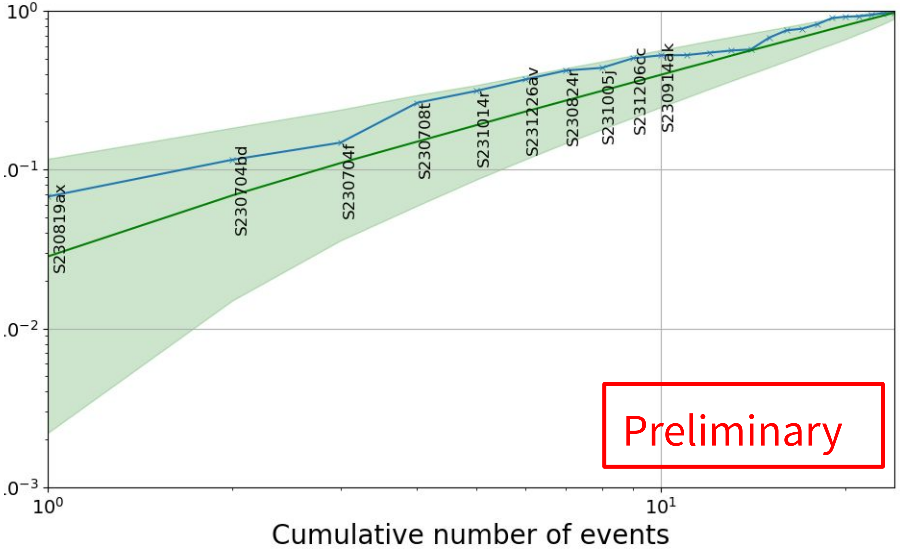
Waveform consistency test
Systematic error

(preliminary) set of O4a events

Plots as in GWTC-3



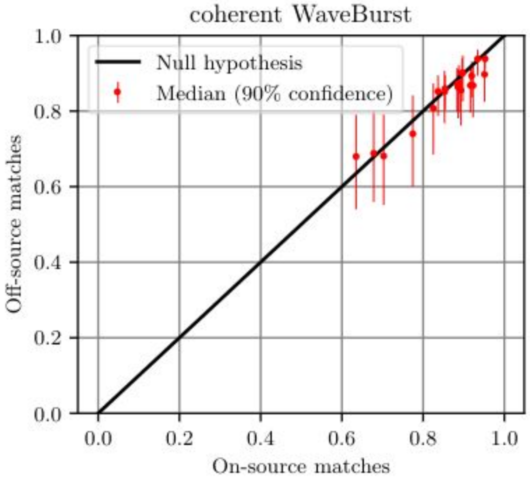
Off-source versus on-source match values. The blue line indicates the null hypothesis, and the error bars indicate the symmetric 90% confidence interval.
No significant deviations

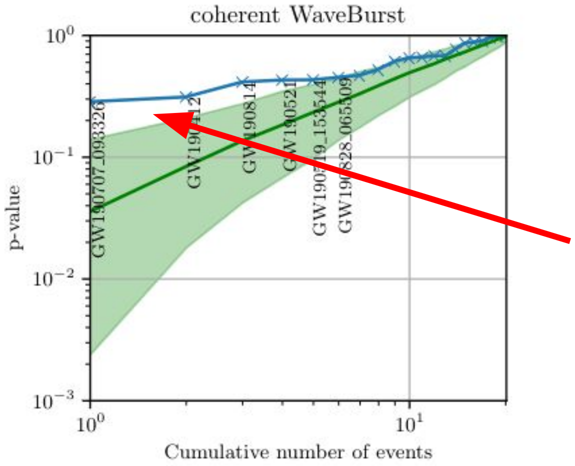


Distribution of p-values, and symmetric 90% interval about the median (green). **No significant deviations**

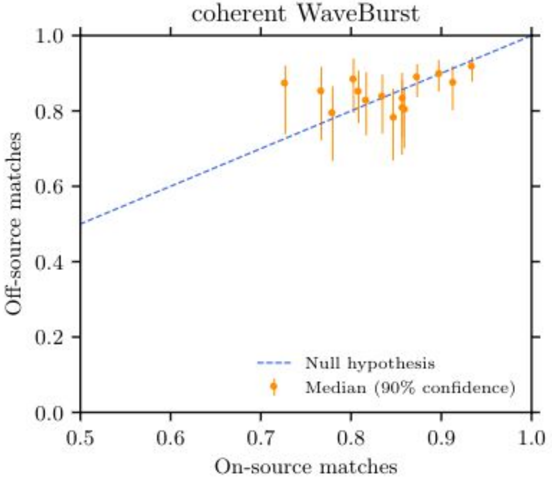
From previous runs

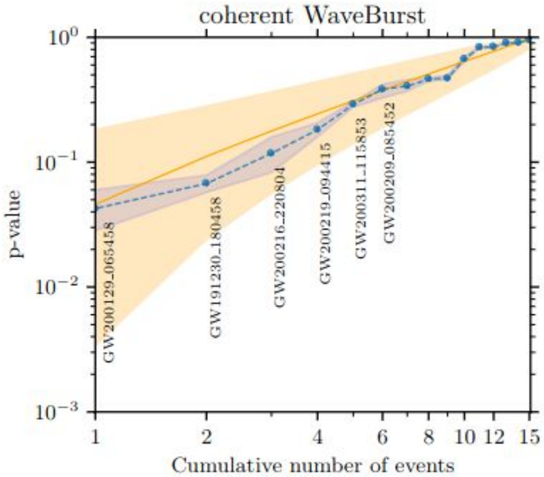
GWTC2



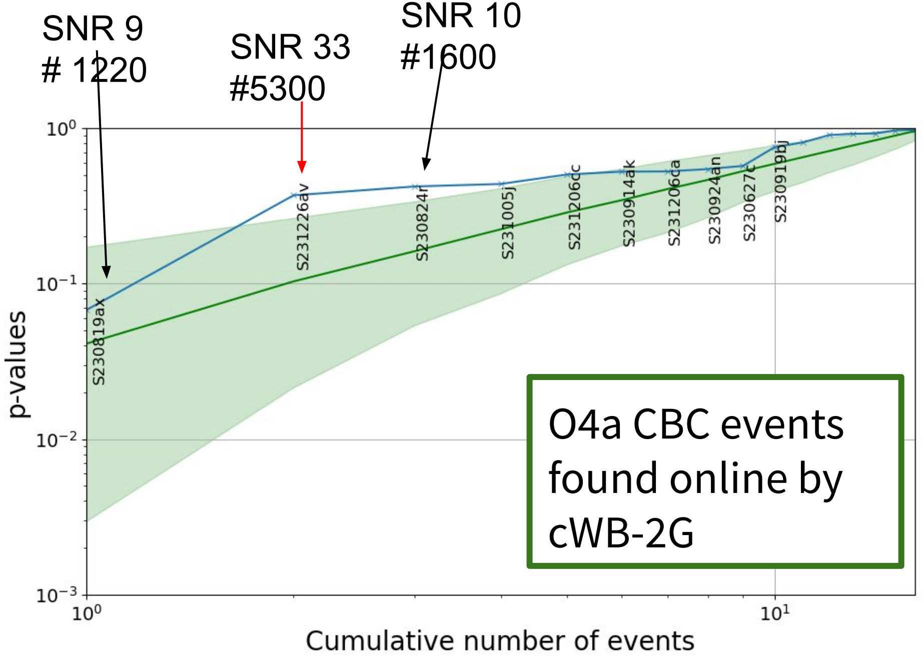


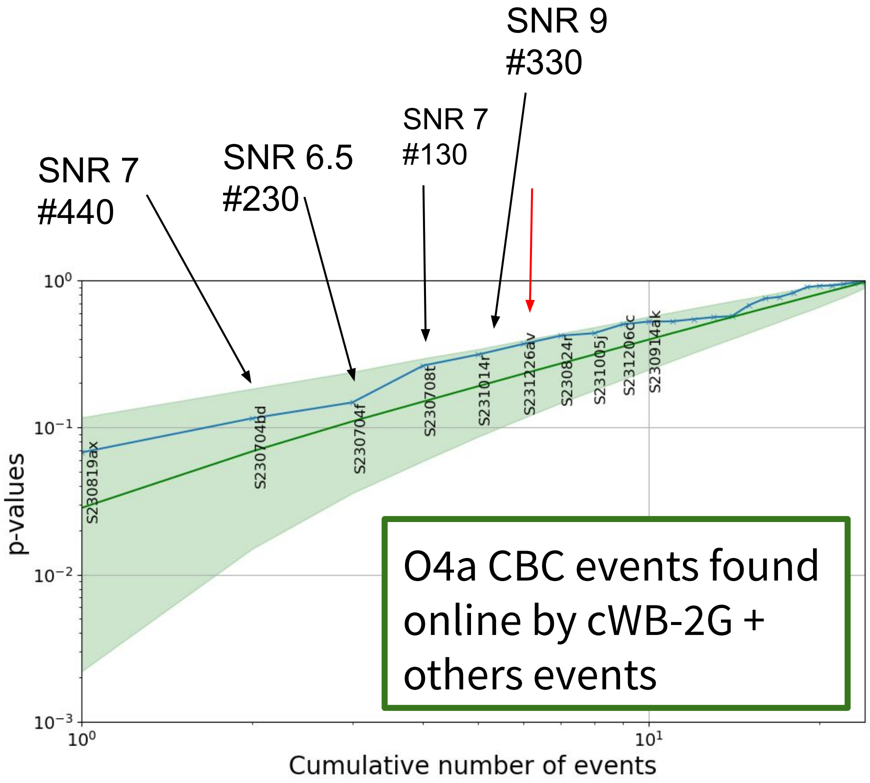
GWTC3





Systematic error - O4a





- Low-SNR events have a statistical uncertainty > systematic uncertainty
- Adding more low-SNR events, 'hide' our systematic error?

Which events should we analyse? Should we set a threshold on SNR?

Systematic error

Tests already performed by cWB-2G (no effect on the p-values distributions):

- off-source injections in a smaller/larger data segment
- PSD variability < variability PE samples
- cWB selection thresholds/post-processing

Systematic error

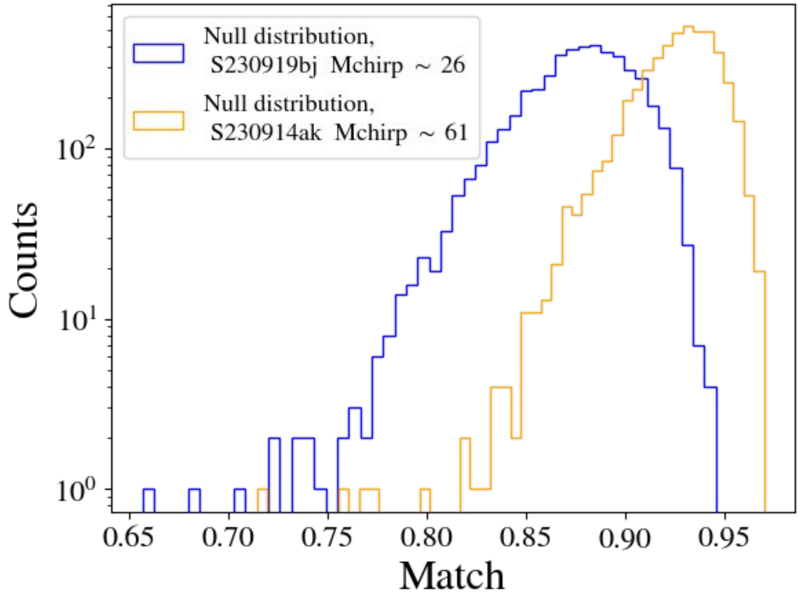
From Edoardo’s presentation(dcc):

- **Problem:** Since the p-values tend to be larger than expected for the null hypothesis, this seems to indicate that our null-hypothesis histogram is too wide, i.e., we are overestimating the variance of the null hypothesis.

- **Possible cause:** the on-source and off-source experiments are not the same. We implicitly assume that the samples in the PE distribution are all very similar, and it is a computationally economical replacement for the ideal procedure, where we should produce a new maximum-likelihood estimate for each injection.

Waveform consistency tests - match vs chirp mass

- Comparison between null distributions for two events with similar SNR, but different chirp masses

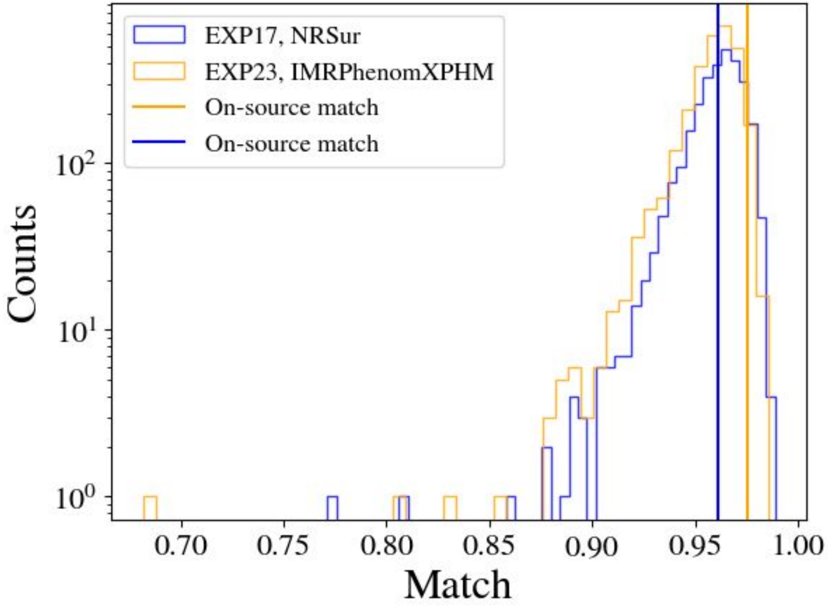


Null-distribution for S230919bj and S230914ak. Similar SNR (17 vs 16.9), but different chirp mass.

S231123cg - waveform consistency

Several PE runs, **no** significant discrepancy found with latest PE using different waveforms models (IMRPhenomXPHM, NRSur7dq4)

- EXP17 (NRSur7dq4): match 0.9611, p-value = 0.51
- EXP23 (IMRPhenomXPHM): match 0.976, p-value = 0.92



On-source match (vertical line) and null-distribution considering PE samples obtained with two different waveforms

comparison injected directions between two events