The astrophysical odds of GW151216

Gregory Ashton\textsuperscript{1,2,*}, Eric Thrane\textsuperscript{1,2},
\textsuperscript{1}School of Physics and Astronomy, Monash University, Vic 3800, Australia,
\textsuperscript{2}OzGrav: The ARC Centre of Excellence for Gravitational Wave Discovery, Clayton VIC 3800, Australia

11 June 2020

ABSTRACT
The gravitational-wave candidate GW151216 is a proposed binary black hole event from the first observing run of Advanced LIGO–Virgo. Not identified as a bona fide signal by LIGO–Virgo, there is disagreement as to its authenticity, which is quantified by \( p_{\text{astro}} \), the probability that the event is astrophysical in origin. Previous estimates of \( p_{\text{astro}} \) from different groups range from 0.18 to 0.71, making it unclear whether this event should be included in population analyses, which typically require \( p_{\text{astro}} > 0.5 \). Whether GW151216 is an astrophysical signal or not has implications for the population properties of stellar-mass black holes and hence the evolution of massive stars. Using the astrophysical odds, a Bayesian method based on coherence, we find that \( p_{\text{astro}} = 0.03 \), suggesting that GW151216 is unlikely to be a genuine signal. We also analyse GW150914 (the first gravitational-wave detection) and GW151012 (initially considered to be an ambiguous detection) and find \( p_{\text{astro}} \) values of 1 and 0.997 respectively. We argue that the astrophysical odds presented here improve upon traditional methods for distinguishing signals from noise.

Key words: gravitational waves – black hole mergers

1 INTRODUCTION
Transient gravitational-wave-astronomy has opened a new window with which to study black holes and neutron stars. The LIGO (Aasi et al. 2015) and Virgo (Acernese et al. 2015) collaborations have now completed three observing runs and announced 13 binary coalescence signals (Abbott et al. 2019a, 2020; The LIGO Scientific Collaboration et al. 2020). The data collected by these observatories is public allowing independent groups to reaffirm observations and identify new candidates (Zackay et al. 2019; Nitz et al. 2019b; Venumadhav et al. 2019, 2020).

In addition to astrophysical signals, gravitational-wave detector data contains transient non-Gaussian noise artefacts, often referred to as glitches (Blackburn et al. 2008; Abbott et al. 2016a; Nuttall et al. 2015; Cabero et al. 2019; Powell 2018). Glitches degrade our ability to identify signals, i.e. the sensitivity of the detector; when the cause of the glitch is fully understood, the optimal solution is to remove the data containing the glitches which improves the sensitivity of the detector (Abbott et al. 2018a). However, the cause of many glitches is not understood and hence they cannot be removed from the data, but must be treated as part of the background noise of the detector. Traditional search methods (see, e.g. Cannon et al. (2013); Usman et al. (2016) and Capano et al. (2017) for a review of the methods) deal with this by estimating the background using bootstrap methods (Efron & Tibshirani 1993). Bootstrap methods are defined by the use of an empirical distribution to estimate a quantity of interest. Subsequently, candidates are assigned an astrophysical probability, \( p_{\text{astro}} \), based on the empirical output of the search pipeline; see Abbott et al. (2016d, 2019a) for details. For loud events such as the GW150914 Abbott et al. (2016c), the first observed binary black hole coalescence, \( p_{\text{astro}} \approx 1 \). Meanwhile, for marginal candidates, \( p_{\text{astro}} \in [0.5, 0.99] \). Different search pipelines produce different values of \( p_{\text{astro}} \) due to differing assumptions. For loud events, this is of little consequence, but as we will see, understanding these assumptions can be crucial for marginal candidates.

GW151216 was reported as a significant trigger in O1, the first LIGO observing run, with \( p_{\text{astro}} = 0.71 \) by Zackay et al. (2019). The event was not included in the first LIGO–Virgo gravitational-wave transient catalogue covering the O1 and O2 observing runs (Abbott et al. 2019a). The candidate was also identified in Nitz et al. (2020), but with \( p_{\text{astro}} = 0.18 \), less than the 0.5 threshold used to determine inclusion in the catalogue. In the original analysis, Zackay et al. noted the large effective spin of the candidate, which led to a range of implications; e.g. Piran & Piran (2020); Fragione & Kocsis (2020); De Luca et al. (2020). However, in a systematic study (Huang et al. 2020), it was shown that support for the effective spin was sensitive to the choice of prior. We summarise the various significance estimates for all events analysed in this work in Table 1.

In this work, we study the significance of GW151216 using the astrophysical odds (Ashton et al. 2019a). This method is different from traditional methods in that it eschews bootstrap noise estimation. Instead, it directly models and fits for the population properties of glitches as they appear projected onto the parameter space of compact binary coalescence signals. By combining the notion of glitches as incoherent signals (Veitch & Vecchio 2010; Isi et al. 2018), and using contextual data to measure the population odds, we argue that the astrophysical odds presented here improve upon traditional
properties of glitches, the astrophysical odds can elevate the signi-
ficance of marginal candidates based on their coherence between
detectors and their properties in the context of typical glitches. The
odds is a Bayesian ratio of probabilities comparing a signal and
noise hypothesis complete with prior probability; it can be used to
directly weight posteriors in the context of a population analysis,
disposing of the need for arbitrary thresholds for inclusion (Gala-
dauge et al. 2019; Gaebel et al. 2019) and can be employed directly to
the analysis of multi-detector events using the framework laid out in
Ashton et al. (2018).

In order to give the results for GW151216 context, and to
validate our method, we also analyse two other binary black hole
systems, with the first and most significant signal in O1 and
GW151012, first reported as a “trigger” (Abbott et al. 2016b) and
subsequently upgraded in significance to a candidate (Abbott et al.
2019a; Nitz et al. 2019a). In the future, we expect more candidates to
be identified in the open data by independent pipelines (see, e.g.
Venumadhav et al. 2020). While we focus here on GW151216,
our broader goal is to establish a unified catalogue, sourced from
multiple groups, each event with a single, reliable value of \( p_{\text{astro}} \).
The \( p_{\text{astro}} \) in this unified catalogue will not depend on the search
pipeline used to first identify each trigger.

## 2 Method

Following Ashton et al. (2019a), we use a Bayesian framework to
calculate the astrophysical odds, \( O \). The odds answers the question:
what is the ratio of probability that a \( \Delta t = 0.2 \text{s} \)-duration data
segment \( d_{i} \) spans the coalescence time \( t \) of an astrophysical signal
versus the probability that it contains noise? The noise can be either
Gaussion or it can include glitch. The odds for a signal in data
segment \( d_{i} \) in a larger dataset \( d \) are

\[
O_{S_{i}/N_{i}}(d) \approx \frac{(\xi) \mathcal{L}(d_{i}\mid S_{i})}{\mathcal{L}(d_{i}\mid N_{i}, \Lambda_{N})} \, \mathcal{L}(\Lambda_{N} \mid d_{i=1}^{k}) \, d\Lambda_{N}.
\]

(1)

Here, \( \xi \) is the probability of a signal in \( d_{i} \) and its expected value
\( \langle \xi \rangle \) is the prior odds; we discuss more below. The term \( \mathcal{L}(d_{i}\mid S_{i}) \) is the
Bayesian evidence (marginal likelihood) for \( d_{i} \) given the signal
hypothesis. This is the likelihood function commonly used to
estimate the parameters of merging binaries, (Veitch et al. 2015).
Meanwhile, \( \mathcal{L}(d_{i}\mid N_{i}, \Lambda_{N}) \) is the likelihood of the data given
the noise hypothesis. The noise hypothesis is that the data contain
either Gaussian noise or non-Gaussian glitches, modelled by un-
correlated (between detectors) binary mergers (Veitch & Vecchio
2010; Ise et al. 2018; Ashton et al. 2019a). The noise likelihood is
marginized over \( \Lambda_{N} \), a set of hyper-parameters that describe the
distributions of glitch parameters.

Finally, \( \pi(\Lambda_{N} \mid d_{i=1}^{k}, I) \) is the noise parameter prior informed
by conditional data \( d_{i=1}^{k} \) and any other cogent information \( I \);
the importance of this will be made clear later on. We refer to this
distribution as the glitch population properties. Our present purpose
is to describe how we calculate \( O \) for the three events considered in
this paper, so we take Eq. 1 as given and refer readers to Ashton et al.
(2019a) for more information including a derivation of \( O \) and a
discussion of the motivation for our noise model.

Equation 1 differs slightly from the expression in Ashton et al.
(2019a) because of two simplifying assumptions. First, we assume
that the prior signal probability \( \xi \) is independent of the glitch hyper-
parameters \( \Lambda_{N} \). Second we assume that the prior signal probability
\( \xi \ll 1 \) (as expected for astrophysical signals). This allows us to
factorise the prior-odds \( \pi_{S_{i}/N_{i}}(d_{i=1}^{k}, I) \) and approximate them by

\[
\pi_{S_{i}/N_{i}}(d_{i=1}^{k}, I) \approx \frac{\xi(\xi)}{(1-\xi)(\xi)} \, d\xi \approx \langle \xi \rangle,
\]

(2)

where \( \langle \xi \rangle \) is the expectation value of \( \xi \). We omitted \( I \) in Ashton et al.
(2019a) as all inferences were made from the contextual data alone.
In this work, we will make good use of cogent prior information
and hence re-introduce it in order to show where this information
is important. With this formalism out of the way, we turn our
attention to the evaluation of \( O \) using data from O1.

The first step is to define the contextual data. The contextual
data is drawn from a span of time near to the candidate of interest.
Ideally, one would like to include as much contextual data as pos-
sible, though, not so much that the detector performance is likely
to have changed. A comprehensive study of transient noise in O1
was performed by Abbott et al. (2016a). Using the single-detector
burst identification algorithm Omicron (Abbott et al. 2016a;
Chatterji et al. 2004), the rate of all glitches with signal to noise ra-
tio (SNR) > 5 during O1 was found to vary by epoch, but typically
was less than 0.5 s\(^{-1}\). However, louder glitches with SNR > 10
have a typical rate (excluding vetoed epochs) between 0.01 and
0.001 s\(^{-1}\). Given this rate, a coincident-observing 24 h period
will contain several thousand quiet glitches and a few hundred loud
 glitches: a sufficient number to estimate typical population proper-
ties. Thus, we use 24 h of contextual data, which is long enough
to provide adequate estimates of the population properties of glitches,
but short enough to control computational costs. We define \( d_{i=1}^{k} \)
to be the set of Omicron triggers in the contextual data.

The next step is to calculate the expectation value of \( \xi \). This is
often referred to as the “duty cycle” (Smith & Thrane 2018). It
is the expectation value for the fraction of segments containing the
coalescence time of a gravitational-wave signal. The duty cycle is
straightforwardly related to the local merger rate \( R \) and the average
time between mergers in the Universe \( t \): \( \xi \sim R \sim t^{-1} \).

By assuming a plausible cosmological model,\(^2\) Abbott et al.
(2018b) obtained \( t = (223 \pm 35) \text{Myr} \) based on a local merger rate of
\( R = 103.5 \pm 110 \text{Gpc}^{-3} \text{yr}^{-1} \). Since then, Abbott et al. (2019b)
updated the estimated local merger rate to be \( R = 53.2 \pm 5.5 \text{Gpc}^{-3} \text{yr}^{-1} \).
Combining these results, we obtain a point estimate of

\[
\langle \xi \rangle \approx 4.5 \times 10^{-4} \left( \frac{\Delta T}{0.2 \text{s}} \right) \left( \frac{R}{59 \text{Gpc}^{-3} \text{yr}^{-1}} \right).
\]

(3)

We approximate the posterior for merger rate as a log-normal dis-
tributions, centred on \( \langle \xi \rangle \), with shape parameters estimated by fitting
the 90% credible intervals given above. Using these fits, we Monte-
Carlo sample the distribution \( \pi(\xi \mid I) \) in Eq. 2 (we are dropping the
contextual data \( d_{i=1}^{k} \) in deference to the information \( I \) used above)
and find that \( \langle \xi \rangle = 7.4 \times 10^{-4} \). Thus, roughly one in 1/\( \langle \xi \rangle \approx 1400 \)
segments contains a coalescence time. For the odds to favour a signal
hypothesis, the astrophysical Bayes factor (i.e. all the terms in
Eq. 1 except \( \langle \xi \rangle \)) must be larger than these prior odds.

The next step is to estimate the glitch population properties
\( \pi(\Lambda_{N} \mid d_{i=1}^{k}, I) \). We write the set of glitch hyper-parameters as
\( \Lambda_{N} \equiv \{ \xi_{\Lambda_{N}}, c_{\Lambda_{N}}, \Lambda_{N} \} \) where \( \xi_{\Lambda_{N}} \) and \( c_{\Lambda_{N}} \) are the prior proba-

\(^1\) The coalescence time is defined slightly differently for different waveforms, but
it is approximately synonymous with time of peak gravitational-wave am-
plitude. While a gravitational waveform can span several segments, the
coalescence time always falls in just one segment.

\(^2\) These error bars don’t include systematic uncertainty associated with the
astrophysical model, which might increase the uncertainty by a factor of \~2.

MNRS 000, 1–6 (0000)
bility for a glitch in the LIGO Hanford and Livingston detectors and \( \lambda_N \) is the remaining set of hyper-parameters describing the glitch population properties. Making the simplifying assumption that these are independent, we can write 
\[
\pi(\lambda_N|d_{\text{sk}}, I) = \pi(\xi^N_{\text{gl}}|d_{\text{sk}}, I)\pi(\xi^g_{\text{sk}}|d_{\text{sk}}, I)\pi(\lambda_N|d_{\text{sk}}, I).
\]
A computationally efficient means to infer \( \lambda_N \) is to use the SNR > 5 Omicron triggers present in the 24 h span of contextual data as a representative sample of glitches (we pre-filter this list to only include triggers with frequencies between 20 and 1000 Hz). By using only these Omicron triggers, we can save the time that would otherwise be spent analysing data segments consistent with Gaussian noise; they do not teach us about the properties of glitches. The inferred distribution of \( \xi^N_{\text{gl}} \) and \( \xi^g_{\text{sk}} \) given this contextual data is consistent with unity. This is not surprising since the Omicron pipeline is designed to identify non-Gaussian noise.

For calculations of the astrophysical odds, we approximate the distribution of \( \xi^N_{\text{gl}} \) and \( \xi^g_{\text{sk}} \) using a point estimates \( \xi_{\text{sk}}^N \) and \( \xi_{\text{sk}}^g \) given by the ratio of the number of Omicron triggers, for each detector, to the available data span. That is, we assume 
\[
\pi(\xi_{\text{sk}}^N|d_{\text{sk}}, I) = \delta(\xi_{\text{sk}}^N - \xi_{\text{sk}}^N).
\]

The values of these point estimates are reported in Table 1. To verify that these point estimates are appropriate, we additionally analyse an auxiliary set of conditional data: 1000 randomly selected times near to GW151216. This contextual data has too few glitches to give reasonable inferences about \( \lambda_N \), but gives a good measure of the glitch probability with medians and 90% credible intervals \( \xi_{\text{sk}}^N = 0.013^{+0.02}_{-0.01} \) and \( \xi_{\text{sk}}^g = 0.0034^{+0.01}_{-0.003} \). The Omicron rate estimates (Table 1) lie at the 90% and 96% percentiles for the Hanford and Livingston detectors respectively. We conclude that the Omicron triggers provide reliable point estimates, but that they are slightly conservative; by slightly overestimating \( \xi_{\text{sk}}^g \), there is a modest bias against the astrophysical hypothesis. In Sec. 5 we show that the results are robust to this conservative choice.

When writing out the prior previously, each term was conditional on both the contextual data as well as I. However, by using the Omicron triggers to infer \( \lambda_N \), but point estimates to infer \( \xi^N_{\text{gl}} \) and \( \xi^g_{\text{sk}} \), we see that we are calculating 
\[
\pi(\lambda_N|d_{\text{sk}}, I) = \pi(\xi^N_{\text{gl}}|I)\pi(\xi^g_{\text{sk}}|I)\pi(\lambda_N|d_{\text{sk}}, I).
\]

Having described details of our calculation, we now recap the procedure from start to finish. There are three steps. First, we identify a 24 h period of data passing the standard data-quality vetoes and absent of injected signals and the analysis segment itself. Second, we filter the available data against Omicron triggers to produce a list of contextual data segments known to contain glitches. Third, we analyse the lowest \( N \) of these triggers and estimate the glitch hyper-parameters \( \lambda_N \). In this step we vary \( N \) by a factor of two and check that the resulting glitch population posteriors are invariant: this demonstrates that we have captured the typical glitch population properties without analysing the entire available data set. Finally, we calculate the astrophysical odds, Eq. (1), using the distribution of hyper-parameters found in the second step, the prior odds \( (\xi^g) = 7.4 \times 10^{-4} \), and the point estimates \( \xi_{\text{sk}}^N \) and \( \xi_{\text{sk}}^g \).

### 3 WAVEFORM MODELS, PRIORS, AND NOISE UNCERTAINTY

We use the aligned-spin waveform model IMRPhenomD (Husa et al. 2016; Khan et al. 2016) for the signal model and for the incoherent-between-detectors glitch model. In the future, it is desirable to extend this analysis to use more sophisticated waveforms including precession of the orbital plane and marginalization over systematic waveform uncertainties (Ashton & Khan 2020). However, we elect to use IMRPhenomD because it is fast and no published candidate events exhibit strong evidence of precession.

We use data from the Gravitational Wave Open Science Centre (The LIGO Scientific Collaboration et al. 2019) spanning 20 – 512 Hz. We estimate the noise properties, the power spectral density (PSD), from the median average of 31 non-overlapping 4 s periodograms using gwpy (Macleod et al. 2019; Macleod et al. 2020). The data used for estimating the PSD is off-source and immediately before the analysis segment in each instance. We do not include the effects of calibration uncertainty (Cahillane et al. 2017).

For signals, we use uniform priors in the chirp mass and mass ratio over the ranges \([13, 100]\) M\(_{\odot}\) and \([0, 125, 1]\) respectively; for the component spin prior we use the “c-prior” (see Eq. (A7) of Lange et al. (2018)) which places much of the prior support at small spins; this is equivalent to the aligned-spin prior (Config. B) used in Huang et al. (2020). For the remaining parameters we use standard priors (see Romero-Shaw et al. (2020)), which are informed by the astrophysical nature of expected signals. In the future, it is worth employing more realistic population models for mass and spin, though, this is outside our present scope; see Fishbach et al. (2020); Galaudage et al. (2019).

The informative prior distributions used for signals are not necessarily appropriate for the glitch model in which we project glitches into the compact binary coalescence signal parameter space. The astrophysical odds framework is designed to use knowledge about typical glitches by marginalizing over the contextual data. It does so by “recycling” posteriors obtained with an initial prior (see Appendix B of Ashton et al. (2019a)). This process is inefficient if the glitch posteriors strongly disagree with the initial prior. We find, in agreement with Davis et al. (2020), that glitches tend to have posterior support in regions of parameter space unusual for typical astrophysical signals, e.g., large negative spins and extreme mass ratios. To counter this inefficiency, we apply, a glitch prior uniform in the component spin \( x_1 \in [−1, 1] \) and \( x_2 \in [−1, 1] \). In testing, we find this improved the efficiency of the astrophysical odds in properly classifying glitches. One might worry that, by applying a different prior for glitches and signals, we are biasing the odds. However, posterior samples are ultimately recycled using hierarchical inference, and so these prior choices do not affect our results except to improve computational efficiency.

We also find that the astrophysically motivated comoving volumetric prior (Romero-Shaw et al. 2020) for luminosity distance can also decrease the efficiency of recycling as most glitches tend to occur around \( \sim 100 \) Mpc. To be clear, glitches have no physical distance; we refer here to the effective distance obtained by fitting glitches to binary merger waveforms. We therefore employ a uniform-in-luminosity distance prior for both signals and glitches, which ensures efficient recycling.

In testing, we find that it is important to include uncertainty in our estimate of the PSD estimation. Failing to take this into account yields false-positive signals \((O > 1)\) in time-slice checks in which the H1 data is offset from L1 to destroy the coherence of real gravitational-wave signals in the data. The solution is to marginalise over uncertainty in the noise PSD as in Talbot & Thrane (2020); Banagiri et al. (2020). Using the median Student-t method from Talbot & Thrane (2020), the astrophysical odds calculated for the set of time-slice Omicron triggers behaves properly: all triggers result in an odds disfavouring an astrophysical interpretation (see Fig. 1). We conclude that marginalising over uncertainty in the PSD is necessary for a reliable odds, and so we apply this to all the results discussed below.
We present the astrophysical odds for the three events analysed. This will likely improve the ability of the method to a broader population and develop a more sophisticated glitch distribution. In future work, we will extend the analysis of glitches points does not significantly change the inferred glitch population. The conclusions are robust to this choice, we rerun the analysis of GW151216 using the Student-t likelihood.

Naively combining this Bayes factor (ignoring the effect of glitches) with a prior odds of \( \ln(\xi) = -7.2 \) the resulting odds is less than unity, providing evidence against an astrophysical origin. The subsequent BCR and astrophysical odds (which include the effect of this prior odds) make minor corrections, but retain the overall conclusion. As discussed in Section 3, we cannot neglect this marginalization if we want reliable odds. This underlines the importance of PSD estimation (for further discussion, see also Chatziioannou et al. (2019)). In the future, with improved methods for evaluating the uncertainty in the PSD (for example, building on the work of Biscoveanu et al. (2020) or developing a joint PSD and model method building on Littenberg & Cornish (2015)), we can reassess GW151216.

We now discuss the prior sensitivity of our results. The dominant prior choice is that of the \( \xi \)-distribution. In this work, we use an astrophysical prior based on the rate of binary black hole events in the O1 and O2 observing runs. The factorisation of the prior odds in Eq. (1) allows us to update the odds based on differing prior assumptions. In order to change the conclusions for GW151216, one would need to increase \( \xi \) by a factor of ~36 translating this into an updated merger rate, this would require a merger rate of \( K \sim 1600 \text{ Gpc}^{-1} \text{yr}^{-1} \), much larger than the current uncertainty on the merger rate (Abbott et al. 2019a). Similarly, a merger rate which would make GW151012 not of astrophysical origin (based on an updated prior odds) would also require a merger rate well outside of the current uncertainty. This demonstrates that our results are not sensitive to the choice of prior odds, given the current uncertainty.

Technically, the odds for GW151012 and GW150914 are biased because the data from these events is used to estimate the rate. However, we expect the error from this double-counting to be negligible. The other potential bias from our prior assumptions is the choice of point estimates for \( \xi_{\text{gw}} \) and \( \xi_{\text{gw}}^{\text{est}} \). To check how sensitive our results are to this choice, we rerun the analysis of GW151216 using \( \xi_{\text{gw}} = \xi_{\text{gw}}^{\text{est}} = 0 \) and find that \( \ln O = -3.5 \). This small shift from our calculated value confirms that our conclusion, that \( p_{\text{astro}} = 0.03 \), is robust to the choice of glitch hyper prior.

For GW150914 and GW151012, the astrophysical odds provide unequivocal evidence that these events are of astrophysical origin. Comparing the BCR and the \( O \) in Table 1 allows to assess the effect of the glitch hyper-model. For GW151012 and GW151216, only a small effect is observed, but for GW150914 the astrophysical odds is larger than the BCR by a factor of \( \sim 7 \). This demonstrates the ability of the astrophysical odds to increase our confidence in a signal based on how unlike the glitch population it is. We can also compare the BCR values derived in this work with that of Iset al. (2018). For GW150914 and GW151012, they find BCR values of 19.6 and 8.5 respectively; larger than the values found in this work (see Tab. 1). This difference is caused by an unknown combination...
of the choice of tuning parameters, the narrower source parameter priors, the use of a precessing waveform, or the marginalization over the PSD applied in this work. Given the significant impact of the marginalization over the PSD, we suspect this is likely to dominate, but we cannot determine this without further investigation.

To visualise our results for the three candidates and various realisations of a background, in Fig. 1, we show the evolution of candidates through three stages of Bayesian significance estimates. Individual candidates are labeled by their ID. In blue, are the Omicron triggers identified for each of the three epochs around each event; we show these together as no differences in behaviour per-epoch were found. All the significance estimates use evidence obtained by marginalizing over the uncertainty in the PSD. For the Omicron trigger candidates, we see two distinct clusters: those with $\ln(B_{\text{coh,inc}}) \sim 0$ and those with $\ln(B_{\text{coh,inc}}) < -1$. These can be understood as a cluster of candidates where the data is reasonably Gaussian in both detectors (thus tricking the coherent Bayes factor which only compares signal evidence against glitch evidence) and a cluster of candidates with a strong glitch in one detector resulting in a Bayes factor favouring the glitch hypothesis. When subsequently analysed with the BCR metric (Izi et al. 2018), the Gaussian cluster is weighted down because the BCR includes Gaussian noise in its alternative hypothesis. Finally, when applying the glitch hyper-prior a small correction is applied based on the likeness of the candidates to the glitch population. For the candidates initially in the $\ln(B_{\text{coh,inc}}) < -1$ cluster, this results in a modest down-weighting: i.e. the odds having marginalized over the glitch population are slightly better at distinguishing glitches. In future work, we expect that a more detailed glitch model will yield further improvement in the ability of the odds to distinguish glitches.

In pink, we also show the evolution of a set of Omicron triggers analysed with a time-slide. That is, we take the set of triggers and apply a 1 s shift between the Hanford and Livingston data. This ensures that the set of triggers do not contain coherent astrophysical signals. The figure demonstrates that the three Bayesian significance estimates perform equivalently for the Omicron triggers under a time-slide as they do without.

6 CONCLUSION

We find that the marginal gravitational wave candidate GW151216 is not of astrophysical origin, $p_{\text{astro}} = 0.03$. Our $p_{\text{astro}}$ estimate is smaller than that of the original detection claim $p_{\text{astro}} = 0.71$ (Zackay et al. 2019), or the PyCBC analysis $p_{\text{astro}} = 0.18$ (Nitz et al. 2020). Taken together with (Huang et al. 2020), we urge the community to use caution when considering the astrophysical implications of this event. We also analyse GW150914, the loudest signal in the first advanced-LIGO observing run, and GW151012, a candidate first marked as marginal, but subsequently upgraded. We find overwhelming support that these are astrophysical signals.

This work lays out the framework for applying the astrophysical odds (Ashton et al. 2019a) to a growing catalogue of gravitational-wave transients. In doing so, we seek to provide a single $p_{\text{astro}}$ for candidate events from multiple groups. Our results do not rely on the output of a search pipeline, and it is easy to see the assumptions that go into our calculations. It is also straightforward to update our significance estimates to keep pace with advances in noise modelling. Unlike traditional search methods, it does not use bootstrap realisations of the noise, but models the noise as incoherent-between-detector signals. In future work, we anticipate a number of improvements including: adding additional alternative models, for example, sine-Gaussians; improved waveforms; improved methods of estimating the noise PSD; and the addition of calibration uncertainty.
REFERENCES

Aasi J., et al., 2015, Class. Quantum Gravity, 32, 074001
Abbott B. P., et al., 2016a, Class. Quantum Gravity, 33, 134001
Abbott B. P., et al., 2016b, Phys. Rev. D, 93, 122003
Abbott B. P., et al., 2016c, Phys. Rev. Lett., 116, 061102
Abbott B. P., et al., 2016d, ApJ, 833, L1
Abbott B. P., et al., 2018a, Classical and Quantum Gravity, 35, 065010
Abbott B. P., et al., 2018b, Phys. Rev. Lett., 120, 091101
Abbott B. P., et al., 2019a, Phys. Rev. X, 9, 031040
Abbott B. P., et al., 2019b, ApJ, 882, L24
Abbott B. P., et al., 2020, ApJ, 892, L3
Acernese F., et al., 2015, Class. Quantum Gravity, 32, 024001
Ashion G., Khan S., 2020, Phys. Rev. D, 101, 064037
Ashion G., et al., 2018, ApJ, 860, 6
Ashion G., Thrane E., Smith R. J. E., 2019a, Phys. Rev. D, 100, 123018
Ashion G., et al., 2019b, ApJS, 241, 27
Bangari S., Coughlin M. W., Clark J., Lasky P. D., Bizouard M. A., Talbot C., Thrane E., Mandic V., 2020, MNRAS, 492, 4945
Biscoveanu S., Haster C.-J., Vitale S., Davies J., 2020, arXiv e-prints, p. arXiv:2004.05149
Blackburn L., et al., 2008, Class. Quantum Gravity, 25, 184004
Cabero M., et al., 2019, Class. Quantum Gravity, 36, 155010
Cahillane C., et al., 2017, Phys. Rev. D, 96, 102001
Cannon K., Hanna C., Keppel D., 2013, Phys. Rev. D, 88, 024025
Capano C., Dent T., Hanna C., Hendry M., Messenger C., Hu Y. M., Veitch J., 2017, Phys. Rev. D, 96, 082002
Chatterji S., Blackburn L., Martin G., Katsavounidis E., 2004, Class. Quantum Gravity, 21, S1809
Chatziioannou K., Haster C.-J., Littenberg T. B., Farr W. M., Ghonge S., Millhouse M., Clark J. A., Cornish N., 2019, Phys. Rev. D, 100, 104004
Davis D., White L. V., Saulson P. R., 2020, arXiv e-prints, p. arXiv:2002.09429
De Luca V., Franciolini G., Pani P., Riotto A., 2020, arXiv e-prints, p. arXiv:2003.02778
Effron B., Tibshirani R., 1993, An Introduction to the Bootstrap. Chapman & Hall, London, UK
Fishbach M., Farr W. M., Holz D. E., 2020, ApJ, 891, L31
Fragione G., Kocsis B., 2020, MNRAS, 493, 3920
Gaebel S. M., Veitch J., Dent T., Farr W. M., 2019, MNRAS, 484, 4008
Galadage S., Talbot C., Thrane E., 2019, arXiv e-prints, p. arXiv:1912.09708
Huang Y., et al., 2020, arXiv e-prints, p. arXiv:2003.04513
Husa S., et al., 2016, Phys. Rev. D, 93, 044006
Isi M., et al., 2018, Phys. Rev. D, 98, 042007
Khan S., et al., 2016, Phys. Rev. D, 93, 044007
Lange J., O’Shaughnessy R., Rizzo M., 2018, arXiv e-prints, p. arXiv:1805.10457
Littenberg T. B., Cornish N. J., 2015, Phys. Rev. D, 91, 084034
Macleod D., et al., 2019, GWpy: Python package for studying data from gravitational-wave detectors (ascl:1912.016)
Macleod D., et al., 2020, gwpy/gwpy: 1.0.1, doi:10.5281/zenodo.3598469, https://doi.org/10.5281/zenodo.3598469
Nitz A. H., Dent T., Dal Canton T., Fairhurst S., Brown D. A., 2017, ApJ, 849, 118
Nitz A. H., Capano C., Nielsen A. B., Reyes S., White L. V., Brown D. A., Krishnan B., 2019a, ApJ, 872, 195
Nitz A. H., Nielsen A. B., Capano C. D., 2019b, ApJ, 876, L4
Nitz A. H., et al., 2020, ApJ, 891, 64
Powell J., 2018, Class. Quantum Gravity, 35, 155017
Romero-Shaw I. M., et al., 2020, arXiv e-prints, p. arXiv:2006.00714
Smith R., Thrane E., 2018, Phys. Rev. X, 8, 021019
Speagle J. S., 2020, MNRAS, 493, 3132
Talbot C., Thrane E., 2020, Gravitational-wave astronomy with an uncertain noise power spectral density, in prep.
The LIGO Scientific Collaboration the Virgo Collaboration et al., 2019, arXiv e-prints, p. arXiv:1912.11716
The LIGO Scientific Collaboration the Virgo Collaboration et al., 2020, arXiv e-prints, p. arXiv:2004.08342
Usman S. A., et al., 2016, Classical and Quantum Gravity, 33, 215004
Veitch J., Vecchio A., 2010, Phys. Rev. D, 81, 062003
Veitch J., et al., 2015, Phys. Rev. D, 91, 042003
Venumadhav T., et al., 2019, Phys. Rev. D, 100, 023011
Venumadhav T., Zackay B., Roguet J., Dai L., Zaldarriaga M., 2020, Phys. Rev. D, 101, 083030
Zackay B., Venumadhav T., Dai L., Roguet J., Zaldarriaga M., 2019, Phys. Rev. D, 100, 023007

This paper has been typeset from a TeX/LaTeX file prepared by the author.