Dirichlet process Gaussian-mixture model

Del Pozzo, Walter; Berry, Christopher; Ghosh, Archisman; Haines, Tom; Singer, Leo; Vecchio, Alberto

DOI:
10.1093/mnras/sty1485

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **479,** 601–614 (2018)
Advance Access publication 2018 June 6

doi:10.1093/mnras/sty1485

Dirichlet process Gaussian-mixture model: An application to localizing coalescing binary neutron stars with gravitational-wave observations

W. Del Pozzo,[1,2★] C. P. L. Berry,[2★] A. Ghosh,[3,4★] T. S. F. Haines,[5] L. P. Singer[6,7] and A. Vecchio[2]

[1]*Dipartimento di Fisica "Enrico Fermi", Università di Pisa, and INFN sezione di Pisa, Pisa I-56127, Italy*
[2]*Institute of Gravitational Wave Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom*
[3]*Nikhef, Science Park 105, NL-1098 XG Amsterdam, Netherlands*
[4]*International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore 560012, India*
[5]*Department of Computer Science, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom*
[6]*Astroparticle Physics Laboratory, NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, MD 20771, USA*
[7]*Joint Space-Science Institute, University of Maryland, College Park, MD 20742, USA*

Accepted 2018 June 1. Received 2018 May 28; in original form 2018 January 24

ABSTRACT

We reconstruct posterior distributions for the position (sky area and distance) of a simulated set of binary neutron star gravitational-waves signals observed with Advanced LIGO and Advanced Virgo. We use a Dirichlet process Gaussian-mixture model, a fully Bayesian non-parametric method that can be used to estimate probability density functions with a flexible set of assumptions. The ability to reliably reconstruct the source position is important for multimessenger astronomy, as recently demonstrated with GW170817. We show that for detector networks comparable to the early operation of Advanced LIGO and Advanced Virgo, typical localization volumes are $\sim 10^4$–10^5~Mpc^3 corresponding to $\sim 10^2$–10^3 potential host galaxies. The localization volume is a strong function of the network signal-to-noise ratio, scaling roughly $\propto \varrho_{\rm net}^{-6}$. Fractional localizations improve with the addition of further detectors to the network. Our Dirichlet process Gaussian-mixture model can be adopted for localizing events detected during future gravitational-wave observing runs and used to facilitate prompt multimessenger follow-up.

Key words: gravitational waves – methods: data analysis – methods: statistical – gamma-ray burst: general – stars: neutron.

1 INTRODUCTION

Bayesian inference is frequently used in astronomy as a means of combining new data with prior knowledge to construct a better model for our understanding of astronomical systems. Our state of knowledge about the values of a system's parameters is encoded in a probability distribution. An efficient and effective means of mapping a probability distribution is using a stochastic sampling algorithm, such as nested sampling (Skilling 2006) or Markov-chain Monte Carlo (Gregory 2005, chapter 12). These explore parameter space and, in doing so, return a set of samples randomly drawn from the desired probability distribution. These samples can be used to calculate summary statistics such as expectation values; however, for some applications it is desirable to have a smooth probability density function. This leaves the question of converting a discrete set of samples into a continuous probability density function.

The crudest means of reconstructing a probability density function is by creating a set of bins and counting the number of samples that fall in each. This is extremely difficult to do robustly: bins must be sufficiently small to resolve the features of the distribution (and avoid introducing artefacts from the quantization) but still large enough that they contain sufficient samples to provide a fair estimate of the underlying probability density at that location. It is almost impossible to do this using a single bin size; in practice, we must adapt to the shape of the distribution, which is not usually known beforehand.

In this paper, we explain an algorithm using Dirichlet processes (DPs) to build a Gaussian-mixture model (DPGMM) that can be used to build probability distributions from a set of samples. We specialize to the question of inferring the (three-dimensional) location of an astronomical system; however, the algorithm may be generalized for working with different parameter spaces. Our DPGMM can be used to efficiently combine the three-dimensional probabil-

★ E-mail: walter.delpozzo@unipi.it (WdelP); cplb@star.sr.bham.ac.uk (CPLB); archis@nikhef.nl (AG)

ity distribution with a galaxy catalogue to produce a list of most probable host galaxies.

This work originates from the field of gravitational-wave astronomy. The new generation of detectors began operation in 2015 September (Abbott et al. 2017a), with the first observing run (O1) of Advanced Laser Interferometer Gravitational-wave Observatory (aLIGO; Aasi et al. 2015a). This run yielded the first observations of binary black hole coalescences, GW150914 (Abbott et al. 2016c), GW151226 (Abbott et al. 2016e), and potentially LVT151012 (Abbott et al. 2016a,b). The second observing run (O2) began in 2016 November, with Advanced Virgo (AdV; Acernese et al. 2015) joining for the final month of 2017 August. The extension of the gravitational-wave detector network to include additional observatories improves the prospects for localizing the source on the sky (e.g. Singer et al. 2014; Abbott et al. 2017a; Gaebel & Veitch 2017). O2 saw further binary black hole detections, GW170104 (Abbott et al. 2017d), GW170608 (Abbott et al. 2017k), and GW170814 (Abbott et al. 2017e) as well as the first binary neutron star (BNS) detection, GW170817 (Abbott et al. 2017f). The complete results of O2 are yet to be announced.

Gravitational-wave observations do not pinpoint the source of transient signals, instead the source location is inferred probabilistically. The source location is of paramount importance for identifying a multimessenger counterpart: both for targeting follow-up observations and for establishing that a candidate counterpart is associated with the gravitational-wave source.[1] Extensive electromagnetic and neutrino follow-up has been performed for the binary black hole detections (e.g. Abbott et al. 2016f; Adrián-Martínez et al. 2016; Albert et al. 2017a) with no conclusive counterpart yet found. This is not surprising. BNSs are the more promising source for counterparts (e.g. Metzger & Berger 2012; Piran, Nakar & Rosswog 2013), and GW170817 was accompanied by detections across the electromagnetic spectrum (Abbott et al. 2017h). A short gamma-ray burst, GRB 170817A, was observed independently of the gravitational-wave localization (Goldstein et al. 2017; Savchenko et al. 2017), but the (three-dimensional) localization from gravitational-wave observations was crucial for identification of a kilonova counterpart (Arcavi et al. 2017a; Coulter et al. 2017; Lipunov et al. 2017; Soares-Santos et al. 2017; Tanvir et al. 2017; Valenti et al. 2017). Multimessenger observations give a range of insights, such as testing the speed of gravity (Abbott et al. 2017i); exploring the host environment and formation history of merging compact binaries (Abbott et al. 2017j; Blanchard et al. 2017; Im et al. 2017; Levan et al. 2017; Pan et al. 2017), and estimation of the Hubble constant (Abbott et al. 2017g; Guidorzi et al. 2017). The question of sky-localization potential for a realistic astrophysical population of BNS systems has been investigated in Singer et al. (2014) and Berry et al. (2015). For the early observing runs, localizations were typically of the order of hundreds of square degrees, making follow-up observations challenging. The probability of observing a counterpart can be enhanced using galaxy catalogues to pick out the most likely locations (Fan, Messenger & Heng 2014; Hanna, Mandel & Vousden 2014) including information on the distance of the source can significantly aid this process (Nissanke, Kasliwal & Georgieva 2013; Gehrels et al. 2016; Singer et al. 2016).

Even without observing a counterpart, inferring the (three-dimensional) location of gravitational-wave sources is useful. Comparing posterior distributions on location with galaxy catalogues makes it possible to assign a probability that a signal originated from a particular galaxy. Comparing the luminosity distance from the gravitational-wave observation with the redshift measurements for the galaxies gives a measure of the Hubble constant (Schutz 1986). Combining results from a few tens of observations from the advanced-detector network could measure the Hubble constant to an accuracy of ∼5 per cent at 95 per cent credibility (Del Pozzo 2012; Chen, Fishbach & Holz 2017). This is comparable to the existing constraints from the *Hubble Space Telescope* Key Project (Freedman et al. 2001) and inferior to current results from the *Planck* cosmic microwave background observations (Ade et al. 2016), the SH0ES Type Ia supernovae survey (Riess et al. 2016, 2018), or from the weak lensing measurements (combined with baryonic acoustic oscillation and Big Bang nucleosynthesis data) from the Dark Energy Survey (Abbott et al. 2017b). However, the gravitational-wave measurement is independent of the usual systematics, making it a valuable check.

While the primary purpose of this work is to document our implementation of a DPGMM for gravitational-wave source localization, and to demonstrate its effectiveness, the techniques described are of general applicability and could be of interest for a wide range of problems. We begin in Section2 with background material on DPs and the DPGMMs; those only interested in our results may skip this section. We apply the DPGMM to reconstruct the position posterior probabilities densities of a set of simulated BNS signals. We use the (well-studied) catalogue of results generated to model the expected early operation of the advanced-detector network presented in Singer et al. (2014) and Berry et al. (2015); this is described in Section 3. In Section 4, we present our results for the source localization. Our reconstructed three-dimensional posteriors indicate that BNSs could be localized to $\sim 10^4$–10^5 Mpc during the early runs of the advanced-detector era, assuming perfect detector calibration (cf. Singer et al. 2016). The introduction of more detectors will improve both two-dimensional and three-dimensional localization, and so the probability of successfully identifying multimessenger counterparts to the gravitational-wave signal.

2 USE OF DIRICHLET PROCESSES

2.1 Posterior distributions

In many fields of astronomy and astrophysics, one of the main challenges is to be able to accurately measure the physical parameters of interest and consequently make reliable statements about the systems that have been observed. Given a set of observations and a model, one must infer the values of the parameters. The dimensionality of parameter space is frequently large, necessitating the use of stochastic samplers for exploration (MacKay 2003, chapter 29). For making reliable inferences about compact binary coalescences (the inspiral and merger of neutron star–neutron star, neutron star–black hole, and black hole–black hole binaries), the LIGO Scientific and Virgo Collaborations (LVC) have devoted significant time and effort to develop LALINFERENCE (Veitch et al. 2015), a suite of programs that are part of the LVC Algorithm Library.[2] Other fields have equiv-

[1] It may be sufficient to associate a gravitational-wave signal with a gamma-ray burst by time coincidence alone since both are short lived (cf. Aasi et al. 2014a,b; Abbott et al. 2016h), but additional spatial coincidence gives greater confidence (cf. Abbott et al. 2017i).

[2] In addition to the stochastic sampling algorithms of LALINFERENCE, localization of BNSs can also be performed using BAYESTAR (Singer & Price 2016), a more expedient algorithm, which we do not consider here.

alent specialized codes, such as COSMOMC for cosmic microwave background (and other cosmological observations) analysis (Lewis & Bridle 2002) or TEMPONEST for pulsar timing (Lentati et al. 2013), or may use general samplers such as EMCEE (Foreman-Mackey et al. 2013). The output of any of these is a list of independent samples drawn from the posterior probability distribution of all relevant parameters. These samples can then be used to reconstruct information about the parameters of interest.

For some applications, it is desirable to have a smooth estimate of the posterior probability density functions. For example, in our case, we will use the probability density functions to (i) calculate credible volumes to check and summarize our reconstructed localizations and (ii) correlate with galaxy catalogues to find the most probable host galaxies. The discrete nature of the samples can make computing the probability density function difficult. To address this problem, various techniques have been developed; the most common ones are histogramming and kernel density estimation (KDE). Both techniques can be effective when the shape of the posterior distribution function is simple or when the number of samples is large; however, when the number of samples is small, different choices of the bin size for histograms or of the kernel width for KDE can yield distorted results that depend on the actual choice of these parameters. Aware of these limitations, an alternative technique based on constructing a k-dimensional tree has been suggested for the estimation of credible regions in the two-dimensional sky plane (Sidery et al. 2014b).[3] This method successfully estimates the sky position, but since it must tile the region of interest with rectangular leaves, its applicability is still limited to simple distributions or large sample numbers. In this paper, we present a Bayesian non-parametric technique based on the DP that can be used on any set of posterior samples.

Our method is routinely used in different fields, e.g. in the context of unsupervised pattern recognition and non-parametric density estimation, but, to the best of the authors' knowledge, it is largely unknown to the astrophysical and gravitational-wave communities. A thorough introduction can be found in the compendium (Hjort et al. 2010); we give a short overview in this section. We begin by introducing the finite-dimensional version of the DP, which is the Dirichlet distribution (Section 2.2). We then describe the DP itself (Section 2.3), and how it can be used to reconstruct a probability density function using a Gaussian-mixture model (Section 2.4). Some specifics of our implementation of the DPGMM are described in Section 2.5.

2.2 The Dirichlet distribution

Consider a random experiment that can give a finite number of outcomes and imagine that we are only interested in registering the class of the outcome. For example, we may be interested in a coin toss where the outcome is either heads or tails, classifying a gravitational-wave source as a BNS, a neutron star–black hole or a binary black hole system, or registering the number of samples that fall inside a bin in order to construct a histogram. If we have k categories, after N samples, the likelihood of the observations is given by the multinomial distribution

$$p(n_1, \ldots, n_k | q_1, \ldots, q_k) = \frac{N!}{n_1! \ldots n_k!} \prod_{i=1}^{k} q_i^{n_i} \,, \,, \tag{1}$$

[3]This uses a two-step algorithm to ensure unbiased results (Berry & Gair 2013; Sidery et al. 2014a).

where n_i is the number of samples in the i-th category, so $N \equiv \sum_{i=1}^{k} n_i$, and q_i is the corresponding probability for a sample to be in that category. In a frequentist context, these probabilities can be estimated from the observed frequencies of each outcome, which becomes exact as N tends to infinity. However, there is nothing stopping us from applying Bayes theorem and asking: 'given the observed samples, how plausible are the inferred probabilities?' (Jaynes 2003, chapter 18). In other words, given the observed data, one can assign a probability distribution to the probabilities for each category.

To infer the probabilities $\boldsymbol{q} \equiv \{q_i\}$ given the observed counts $\boldsymbol{n} \equiv \{n_i\}$, we can use Bayes' theorem

$$p(\boldsymbol{q}|\boldsymbol{n}) = \frac{p(\boldsymbol{n}|\boldsymbol{q})p(\boldsymbol{q})}{\int \mathrm{d}\boldsymbol{q}\; p(\boldsymbol{n}|\boldsymbol{q})p(\boldsymbol{q})} \,, \,, \tag{2}$$

where $p(\boldsymbol{n}|\boldsymbol{q})$ is the likelihood defined in (1), and $p(\boldsymbol{q})$ is the prior distribution on the probabilities $\boldsymbol{q}$. To complete the inference, we only need to select an appropriate prior.

When we are interested in estimating the probability mass function from the observation of a discrete set of samples, a prior is required for the problem to be well posed. Without assigning a prior, estimating a probability density from a histogram can be, in some cases, troublesome. For instance, if one of the bins has been assigned no samples, the probability assigned to that particular bin will always be zero. Inclusion of a suitable prior circumvents this issue since it allows for a non-zero probability in each bin even without any observations (the role of the prior is to say that we expect that it is possible for a sample to be in each category). Therefore, we obtain sensible results from our inference, even when we have few samples.

A common choice for a prior in this situation is the Dirichlet distribution. As we will see, the Dirichlet distribution has several convenient properties that allow it to be tailored to match our prior expectations. One advantage of using the Dirichlet distribution is that it is *conjugate* to the multinomial distribution (Raiffa & Schlaifer 1961, chapter 3). This means that if we use a Dirichlet distribution as a prior with our multinomial likelihood, our posterior will also be a Dirichlet distribution (which can then be used as the prior for our next set of observations). This invariance under the inclusion of new data means that our inferences form a never-ending chain of Dirichlet distributions, which greatly simplifies computation and interpretation of results (Gelman et al. 2014, section 2.4).

The Dirichlet distribution is defined as

$$\mathrm{Dir}(\boldsymbol{q}|\boldsymbol{a}) = \frac{\Gamma(A)}{\prod_{i=1}^{k} \Gamma(a_i)} \prod_{i=1}^{k} q_i^{a_i - 1} \qquad \{a_i > 0\} \,, \,, \tag{3}$$

where Γ is the gamma function, $\boldsymbol{a} \equiv \{a_1, \ldots, a_k\}$ are the concentration parameters that control the shape of the distribution; $A \equiv \sum_{i=1}^{k} a_i$, and the probabilities $\boldsymbol{q}$ are normalized such that

$$\sum_{i=i}^{k} q_i = 1. \tag{4}$$

With a Dirichlet prior, the posterior distribution for the probabilities $\boldsymbol{q}$ given some data counts $\boldsymbol{n}$ is then

$$p(\boldsymbol{q}|\boldsymbol{n}) = \mathrm{Dir}(\boldsymbol{q}|\boldsymbol{a} + \boldsymbol{n}). \tag{5}$$

Hence, we can consider $\boldsymbol{a}$ as the set of prior counts for each category observed before our current observation set; since these are non-zero, we ensure that even when we have no samples in a bin, its probability is not zero. In general, for $\boldsymbol{q} \sim \mathrm{Dir}(\boldsymbol{a} + \boldsymbol{n})$, the ex-

pectation (mean) value of probability q_i is

$$\bar{q}_i = \frac{a_i + n_i}{A + N}\,; \tag{6}$$

thus, in the limit of $n_i \gg a_i$, such that the likelihood dominates over the prior, we recover the intuitive frequentist result n_i/N.

The Dirichlet distribution is a practical density estimator for discrete probability distributions. When we are beginning our inferences, we are typically starting from a state of ignorance: we do not prefer any one category over another and therefore must assign each equal probability. The corresponding uninformative choice of the Dirichlet distribution has (Gelman et al. 2014, section 3.4)[4]

$$a_i = 1. \tag{7}$$

Following collection of the samples, application of Bayes' theorem with this prior gives an expectation value

$$\bar{q}_i = \frac{n_i + 1}{N + k}. \tag{8}$$

For the case of two possible outcomes, this yields Laplace's rule of succession (Jaynes 2003, chapter 18; MacKay 2003, section 3.2). The modal value (maximum a posteriori estimate) for probability q_i is

$$\hat{q}_i = \frac{n_i}{N}\,,\,, \tag{9}$$

agreeing with the frequentist result.

Having established the properties of the Dirichlet distribution, we now consider its infinite-dimensional generalization, the DP.

2.3 The Dirichlet process

The DP was introduced in Ferguson (1973). It is a stochastic process that generalizes the Dirichlet distribution to infinite dimensions and can be used to set a prior on unknown distributions. While a draw from the Dirichlet distribution is a discrete distribution of finite length, a draw from the DP is a discrete distribution of infinite length. It is a probability distribution for other probability distributions; this additional freedom allows us to dispense with the need to specify bins. For a historical introduction to the DP and its properties, see Gupta & Richards (2001).

To define a DP, let us consider a probability distribution G over the parameter space Θ.[5] We use ϑ to denote an element or collection of elements of Θ, with $G(\vartheta)$ the corresponding probability (density). For G to be DP distributed, we require that for *any* set of partitions $\vartheta_1, \ldots, \vartheta_k$ of Θ (these could represent histogram bins), the vector $\boldsymbol{G} = (G(\vartheta_1), \ldots, G(\vartheta_k))$ is distributed according to a Dirichlet distribution. Introducing a base distribution H over Θ with $\boldsymbol{H} = (H(\vartheta_1), \ldots, H(\vartheta_k))$, and a (positive, real) concentration parameter a, we have that

$$\boldsymbol{G} \sim \mathrm{Dir}(a\boldsymbol{H})\,,\,, \tag{10}$$

and we say that G is DP distributed with base distribution (or base measure) H and concentration parameter a,

$$G \sim \mathrm{DP}(a, H). \tag{11}$$

[4]Setting the a_i to any constant will result in a uniform distribution. The choice of $a_i = 1$ has the attractive property corresponding to a prior weight of each bin having a single count. Using a larger value gives a stronger prior on the distribution being uniform, and more samples need to be collected before the inferred distribution will significantly deviate from this.

[5]For our application, Θ can be interpreted as the space of means and covariances that define our smoothing kernels (see Section 2.4).

Intuitively, H can be thought as the mean of the DP: distributions are drawn from around H such that the expectation value is $\bar{G}(\vartheta) = H(\vartheta)$. The concentration parameter a plays the role of the inverse variance of the DP, controlling how the samples are distributed across Θ: in the limit of $a \to 0$, the draws are all clustered at a single, random ϑ, while in the limit of $a \to \infty$ the draws follow exactly the base distribution (Gelman et al. 2014, section 23.2).[6] When a DP is used for inference, the concentration parameter controls the strength of the prior, with a larger value keeping us closer to our initial expectation of a distribution like H, in a similar way to how $\boldsymbol{a}$ sets the prior strength in a Dirichlet distribution (cf. Raiffa & Schlaifer 1961, section 3.3.4).

The DP has a similar conjugacy property to the Dirichlet distribution. Let us imagine that we have collected N observations $\zeta_i \sim G$, where i runs from 1 to N. If our prior is $G \sim \mathrm{DP}(a, H)$, then our posterior would be (Gelman et al. 2014, section 23.2)

$$G \sim \mathrm{DP}\left(a + N, \frac{a}{a+N}H(\vartheta) + \frac{1}{a+N}\sum_{i=1}^{N}\delta(\vartheta - \zeta_i)\right). \tag{12}$$

From this, we can obtain the posterior expectation of G, which is now our best prediction for future observations (Blei & Jordan 2006; Teh 2010)

$$\bar{G}(\vartheta) = \frac{a}{a+N}H(\vartheta) + \frac{1}{a+N}\sum_{i=1}^{N}\delta(\vartheta - \zeta_i). \tag{13}$$

The form is analogous to that in (6). We now need to know how to use the posterior DP.

Samples from a DP are a weighted sum of point probability masses, and they can be constructed in several ways (such as the Blackwell–MacQueen urn scheme, Chinese restaurant process, or stick-breaking construction), each emphasizing a different property of the DP (Teh 2010). We use the stick-breaking construction, where a sample from a DP $G \sim \mathrm{DP}(a, H)$ can be represented as (Sethuraman 1994)

$$G(\vartheta) = \sum_{i=1}^{\infty} w_i \delta(\vartheta - \zeta_i)\,,\,, \tag{14}$$

where

$$w_j = \beta_j \prod_{i=1}^{j-1}(1 - \beta_i)\,,\,, \tag{15}$$

$$\beta_j \sim \mathrm{Beta}(1, a)\,,\,, \tag{16}$$

$$\zeta_i \sim H. \tag{17}$$

Here, the beta distribution is

$$\mathrm{Beta}(\beta|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\beta^{a-1}(1-\beta)^{b-1}\,; \tag{18}$$

it is the binomial specialization of the Dirichlet distribution. For brevity, we can combine (15) and (16) and denote $\boldsymbol{w} \equiv \{w_j\}$ as being constructed following the Griffiths–Engen–McCloskey (GEM) distribution (Pitman 2006, chapter 3):

$$\boldsymbol{w} \sim \mathrm{GEM}(a). \tag{19}$$

[6]In (10), the Dirichlet distribution only depends on the product $a\boldsymbol{H}$, but the potential degeneracy between the magnitude of a and $\boldsymbol{H}$ is broken by requiring that H is normalized to unity.

Since a sample $G(\vartheta)$ from a DP can be interpreted as a collection of point probability masses, it is a discrete distribution; $G(\vartheta)$ has no density but is instead atomic. Consequently, samples from a DP cannot be used directly to describe continuous distributions. Nevertheless, DPs are commonly used for non-parametric density estimation using draws from a DP to define a set of kernel functions (Lo 1984; Escobar & West 1995). We use a Gaussian-mixture model to reconstruct our inferred probability distribution as described in the next section.

2.4 The Gaussian-mixture model

To build a continuous probability density function from our DP draws, we use a mixture of smoothing kernel functions. Let us introduce $\mathcal{K}(\xi|\vartheta)$ as the family of kernel functions indexed by ϑ. Using our DP-distributed G, we can build a non-parametric probability density for ξ according to (Gelman et al. 2014, section 23.3)

$$p(\xi) = \int \mathrm{d}\vartheta\, \mathcal{K}(\xi|\vartheta) G(\vartheta). \tag{20}$$

This can be turned into a sum, an infinite mixture of kernels, using (14).

The common choice for the kernel function is a multivariate Gaussian

$$\mathcal{K}(\xi|\vartheta) \equiv \mathcal{N}(\xi|\mu, S^{-1})\,,\,, \tag{21}$$

where μ is the (multidimensional) mean and S is the precision matrix (the inverse of the covariance matrix). This choice defines the DPGMM; we describe the distribution for ξ as being made up of an infinite mixture of Gaussian clusters, each with their own mean and covariance. The mean and precision matrix are learned from the data when fitting the DP model.

To define the DP for μ and S, we must specify a base distribution. It is common practice to use conjugate priors for these applications, to exploit their useful properties. Different choices are possible (Görür & Rasmussen 2010) but at the price of losing the conjugacy property and therefore complicating the analysis substantially. The conjugate prior of a multivariate Gaussian distribution with unknown mean and precision matrix is the normal–Wishart distribution (cf. Escobar & West 1995)

$$\mathcal{NW}(\mu, S|\mu_0, \rho, \Lambda, \nu) = \mathcal{N}\left(\mu\middle|\mu_0, (\rho S)^{-1}\right) \mathcal{W}(S|\Lambda, \nu). \tag{22}$$

Here, the Wishart distribution with ν degrees of freedom is

$$\mathcal{W}(S|\Lambda, \nu) = \frac{|\Lambda|^{-\nu/2}}{2^{\nu m/2}\pi^{m(m-1)/4}} \left[\prod_{i=1}^{m} \Gamma\left(\frac{\nu+1-i}{2}\right)\right]^{-1} \times |S|^{-(\nu+m+1)/2} \exp\left[-\frac{1}{2}\,\mathrm{tr}(\Lambda^{-1}S)\right]\,,\,, \tag{23}$$

where S and Λ are positive-definite $m \times m$ matrices, and the expectation value is $\bar{S} = \nu\Lambda$. The normal–Wishart distribution introduces hyperparameters μ_0 (the expected value of the mean), ρ (a scale factor), Λ (a prior for the precision matrix), and ν (the number of degrees of freedom); these are common to all mixture components, expressing the belief that component parameters should be members of a single family. We choose the parameters of the normal–Wishart distribution to be the mean and precision of the observed samples, the scale factor to be equal to the requested resolution (see Section 2.5 for further details), and the number of degrees of freedom to be equal to the dimensionality of the problem plus two (this ensures that the distribution is well conditioned).

Due to its conjugacy to the multivariate Gaussian, choosing $\mathcal{NW}(\mu, S|\mu_0, \rho, \Lambda, \nu)$ as the base distribution for the DP, it is possible to marginalize out analytically the multivariate Gaussian parameters and obtain the non-parametric density estimate as a mixture of multivariate Student-t distributions.[7]

In addition to the base distribution, we also need a concentration parameter for our DP. This too can be updated from the data, but we must specify a prior distribution for it. We use a gamma distribution (Escobar & West 1995), specifically $a \sim \mathrm{Gamma}(1,1)$. The gamma distribution is given by

$$\mathrm{Gamma}(a|b, c) = \frac{c^b}{\Gamma(b)} x^{b-1} \exp(-cx)\,; \tag{24}$$

it is the univariate specialization of the Wishart distribution. It is especially convenient as it is conjugate to the beta distribution used in (16) (Blei & Jordan 2006). The prior expectation is $\bar{a} = 1$ (cf. Gelman et al. 2014, section 23.3).

Combining everything together, the prior DPGMM is assembled as

$$a \sim \mathrm{Gamma}(1, 1)\,,\,, \tag{25}$$

$$\boldsymbol{w} \sim \mathrm{GEM}(a)\,,\,, \tag{26}$$

$$\mu_i,\ S_i \sim \mathcal{NW}(\mu, S|\mu_0, \rho, \Lambda, \nu)\,,\,, \tag{27}$$

$$\xi \sim \sum_{i=1}^{\infty} w_i \mathcal{N}(\mu_i, S_i^{-1}). \tag{28}$$

We first calculate hyperparameters (concentration and base distirbution) to specify our DP; this determines parameters that describe a mixture of Gaussian kernels, and the sum of this mixture gives the distribution of the observed parameters ξ (in Section 2.5, we describe how ξ is a set of three-dimensional position coordinates). Given a set of data (particular realizations of ξ), we now have to solve the inverse problem to find its posterior probability density.

DPGMMs can be explored using Gibbs sampling (Neal 2000; Rasmussen 2000); however, we use the variational algorithm introduced in Blei & Jordan (2006) with the capping method described in Kurihara, Welling & Vlassis (2007). We use the publicly available implementation developed by one of the authors (previous applications include background subtraction; Haines & Xiang 2012, 2014).[8] Our choice of implementation allows the number of components in the DPGMM to grow without limit until the best-fitting model is found; this finite number of components is then used as our estimate for the posertior probability density. The multivariate normal mean vector and covariance matrix are set by maximizing the likelihood of the observed data vector ξ, given the number of components to which data have been assigned, see equation (17) in Görür & Rasmussen (2010).

2.5 Implementation for gravitational-wave data

We are interested in reconstructing posterior probability densities from a set of samples as calculated by a stochastic sampling algorithm (Veitch et al. 2015). To do so, we have adopted the algorithm presented in the previous subsection, specialized to the problem of

[7]The normal distribution is a limiting case of the Student-t distribution.
[8]The DPGMM module is available from github.com/thaines/helit/.

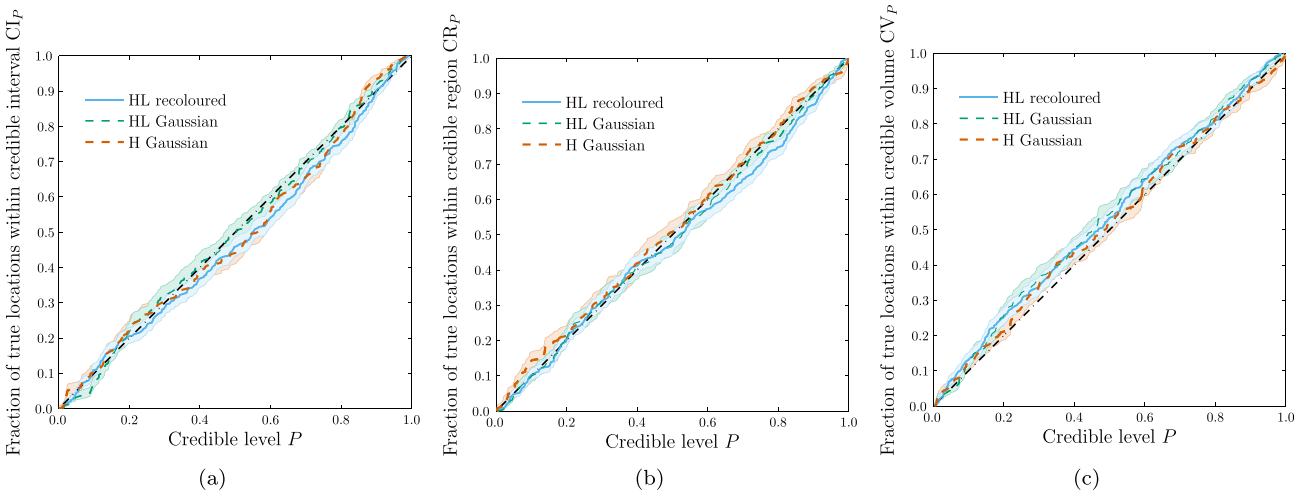

Figure 1. Fraction of true locations found within a (a) distance credible interval, (b) sky-area credible region, or (c) credible volume as a function of encompassed posterior probability. Results with simulated Gaussian noise are indicated by the dashed lines, results using recoloured S6 noise are indicated by the solid line, and the expected distribution is indicated by the dot–dashed diagonal line. The 68 per cent confidence interval for the cumulative distribution is enclosed by the shaded regions, this accounts for sampling errors and is estimated from a beta distribution (Cameron 2011).

Table 1. Comparison of sky-localization areas produced using the DPGMM to those produced using KDE. The mean and standard deviation of the log ratio for the 50 per cent credible region $CR_{0.5}$, the 90 per cent credible region $CR_{0.9}$, and the searched area A_* are listed for each method.

Log ratio	HL recoloured Mean	Standard deviation	HL Gaussian Mean	Standard deviation	HLV Gaussian Mean	Standard deviation
$\mathcal{R}_{CR_{0.5}}$	0.007	0.129	0.017	0.120	0.058	0.197
$\mathcal{R}_{CR_{0.9}}$	0.047	0.135	0.045	0.134	0.072	0.192
$\mathcal{R}_{A_*}$	0.066	0.361	0.095	0.376	0.020	0.495

estimating the posterior probability density for the distance D, right ascension α, and declination δ.[9]

Since the DPGMM is not designed to deal with periodic coordinates, we perform our analysis in Cartesian coordinates; we transform $\{D, \alpha, \delta\}$ into $\{x, y, z\}$ following the transformation:

$$x = D\cos(\alpha)\cos(\delta)\ ,\ , \tag{29}$$

$$y = D\sin(\alpha)\cos(\delta)\ ,\ , \tag{30}$$

$$z = D\sin(\delta). \tag{31}$$

It is these Cartesian-space samples that define our observations ξ, and we use their mean and inverse covariance to specify the hyperparameters of the normal–Wishart distribution (22). We apply the variational method of Blei & Jordan (2006), as described in Section 2.4, to obtain the probability density $p(x, y, z|\boldsymbol{w}, \mu, S^{-1})$. We transform back into $\{D, \alpha, \delta\}$-space using the Jacobian of the coordinate transformation:

$$p(D, \alpha, \delta|\boldsymbol{w}, \mu, S^{-1}) = p(x, y, z|\boldsymbol{w}, \mu, S^{-1}) \left\| \frac{\partial(x, y, z)}{\partial(D, \alpha, \delta)} \right\|\ ,\ , \tag{32}$$

where

$$\left\| \frac{\partial(x, y, z)}{\partial(D, \alpha, \delta)} \right\| = D^2\cos(\delta). \tag{33}$$

[9]We neglect the effects of cosmology and so do not distinguish between different distances; the farthest source we consider is at a (luminosity) distance of 269 Mpc, which corresponds to a redshift of less than 0.07 assuming standard cosmology (Ade et al. 2016).

We then obtain the non-parametric posterior density estimate by marginalizing away μ and S analytically, because of the choice of conjugate priors.

Once we have obtained $p(D, \alpha, \delta)$, we can use it for making statements about the probable location. For example, we can compute credible volumes by evaluating the model over a three-dimensional grid spanning the whole volume under consideration. By default, we use a uniform $\{D, \alpha, \delta\}$ grid that is $50 \times 1440 \times 720$. This is by far the most computationally expensive step in our analysis, taking on the order of ∼1 hr.[10] Possibilities for optimizing this, such as using an adaptive grid, will be investigated in the future. Once the density function has been evaluated over the grid, we sort each of the grid points according to their probability, compute the cumulative distribution and then find the set of points having a probability equal to the requested credible level. Two-dimensional posterior distributions for sky position, as well as one-dimensional posterior distributions for distance, are then obtained by numerical marginalization of the original three-dimensional distribution. Credible regions and intervals in the lower dimensional spaces are obtained in same way as their three-dimensional counterparts. As we explain in Section 4.4, we can also use $p(D, \alpha, \delta)$ directly, without computing credible volumes, together with galaxy catalogues to produce a list of most probable source galaxies.

[10]Across all data sets, the median run time is 2900 s and the central 90 per cent range is 20–4340 s using eight CPU cores.

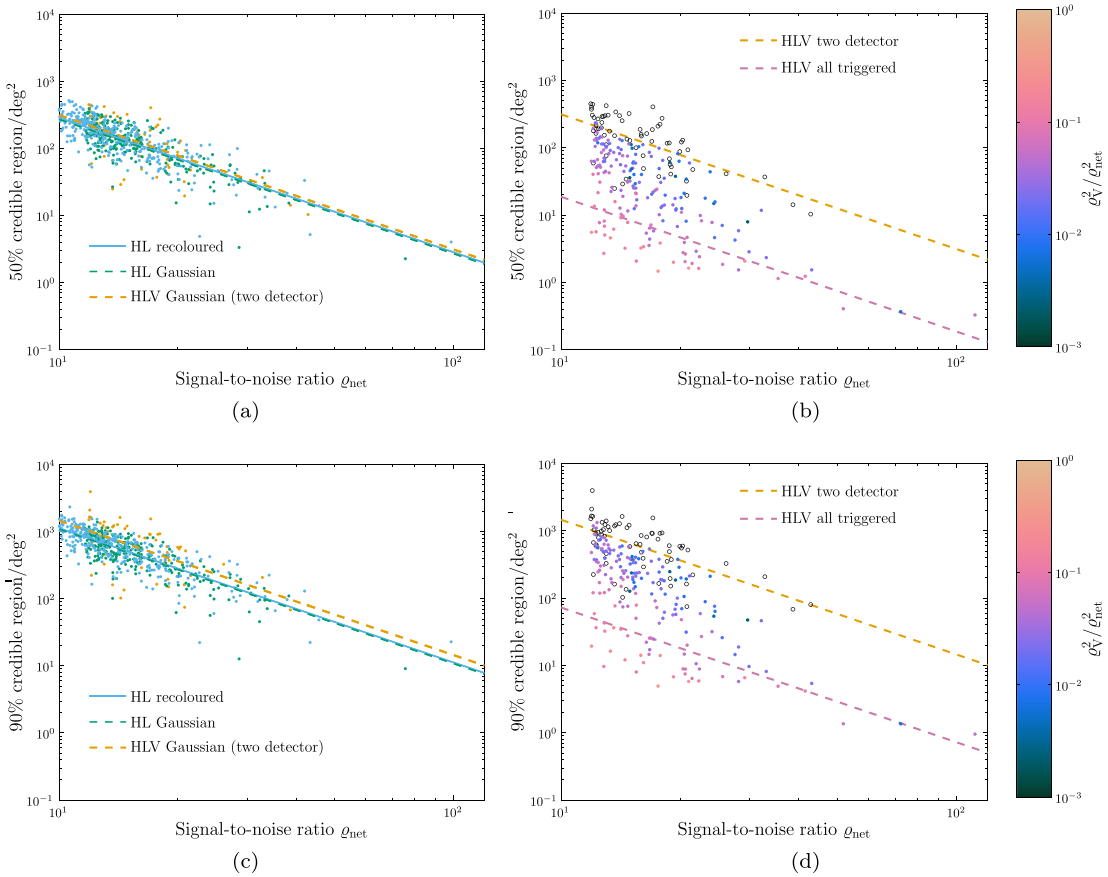

Figure 2. Sky-localization areas as a function of SNR $\varrho_{\rm net}$. The left column shows two-detector results, and the right column shows all HLV scenario results; the top row shows the 50 per cent credible region $\mathrm{CR}_{0.5}$, and the bottom row shows the 90 per cent credible region $\mathrm{CR}_{0.9}$. Individual results are indicated by points, and we include fiducial best-fitting lines assuming that the area $A \propto \varrho_{\rm net}^{-2}$. The left column shows both HL sets of results and the HLV results where only two detectors are operational, each has its own best-fitting line. The HLV two-detector results are also shown in the right column, indicated by the open points, the three-detector results are colour-coded by the fraction of the SNR (squared) from AdV; the lines are fits to the two-detector network runs and those three-detector network runs loud enough to trigger in all detectors.

3 SIMULATION

To demonstrate the effectiveness of the DPGMM at estimating probability density functions, we consider the problem of reconstructing the posterior distribution for the position of a (simulated) BNS coalescence, as would be observed in the early advanced gravitational-wave detector era (similar to during O1 and O2). The (three-dimensional) position is an illustrative test case since it gives an indication of how the technique functions in multiple dimensions, while still being easy to visualize. However, our main motivation for considering the position is the desire to be able to reliably reconstruct the location of a gravitational-wave source following a detection for the purposes of electromagnetic or neutrino follow-up (e.g. Adrián-Martínez et al. 2016; Abbott et al. 2017a, 2016f; Abbott et al. 2017h; Albert et al. 2017a; Albert et al. 2017b).

We use the data presented in Singer et al. (2014) and Berry et al. (2015). These consider two observing scenarios in anticipation of the early operation of the advanced detector network. The first scenario considers the two-detector network of LIGO Hanford and LIGO Livingston, with sensitivities similar to what was expected for O1; the second considers the three-detector network including AdV, with sensitivities similar to what was expected in O2; we refer to these scenarios as HL and HLV, respectively.[11] Singer et al. (2014) simulated 2 months of observations from each scenario, while Berry et al. (2015) only considered the HL scenario but used more realistic noise, including non-Gaussianity as seen in the sixth science (S6) run of initial LIGO (Aasi et al. 2015b). The detectors' duty cycles are assumed to be 80 per cent (cf. Abbott et al. 2017a), such that in the HLV scenario there are three-detector observations for 51.2 per cent of the time and two-detector observations for 38.4 per cent of the time. The assumed HL sensitivity was slightly less than actually achieved in O1, the assumed BNS detection range was ∼55 Mpc compared with the achieved range of ∼70 Mpc (Abbott et al. 2016g); conversely, the assumed HLV sensitivity was better than achieved for the majority of O2 (Abbott et al. 2017d,e). However, these data sets provide a qualitative illustration of what can be achieved during the early observing runs of the aLIGO–AdV network.

We refer to the Singer et al. (2014) results as HL Gaussian and HLV Gaussian, since the detector noise is Gaussian, and the Berry et al. (2015) results as HL recoloured because the noise is recoloured

[11] The HL and HLV scenarios are the 2015 and 2016 scenarios of Singer et al. (2014), respectively.

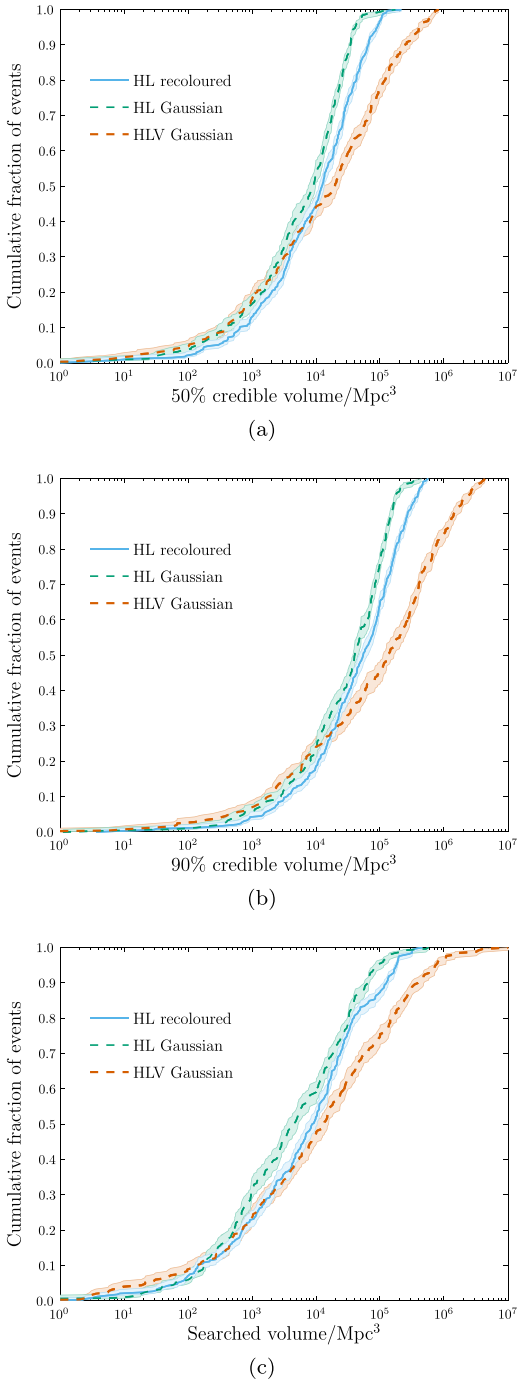

Figure 3. Cumulative fractions of events with localization volumes smaller than the abscissa value. The top panel shows the 50 per cent credible volume $CV_{0.5}$, the middle shows the 90 per cent credible volume $CV_{0.9}$, and the bottom shows the searched volume V_*. The 68 per cent confidence interval for the cumulative distribution is enclosed by the shaded regions; this does not include the inherent uncertainty in the volume estimates.

S6 noise.[12] Both share same catalogue of sources, an astrophysically motivated population of BNSs. The neutron star masses were chosen to be uniformly distributed between $1.2M_\odot$ and $1.6M_\odot$; the sources were distributed uniformly in co-moving volume and on the polarization–inclination two-sphere, and each neutron star was given a randomly oriented spin with a uniformly distributed magnitude up to a maximum $\chi_{\max} = 0.05$;[13] these ranges cover the observed population of BNSs (e.g. Mandel & O'Shaughnessy 2009; Özel et al. 2012; Kiziltan et al. 2013; Abbott et al. 2017f). Further details about the simulation can be found in Singer et al. (2014).

The simulated data were treated as real signals would be, first being passed through the GSTLAL detection pipeline (Cannon et al. 2012). On account of the difference in noise, slightly different detection criteria were used in Singer et al. (2014) and Berry et al. (2015), the former using a cut in the network signal-to-noise ratio (SNR) of $\varrho_{\rm net} = 12$ and the latter using a false-alarm rate (FAR) threshold of $10^{-2}{\rm yr}^{-1}$. Although, broadly consistent, this difference results in the inclusion of additional low-SNR ($\varrho_{\rm net} \approx 10$–12) events for the FAR-only cut.

Parameter-estimation codes are run on detections (Abbott et al. 2016a,d, 2017f), and we use the posterior samples generated by LALINFERENCE (Veitch et al. 2015). This analysis, for expediency, did not include the effects of the spins of the neutron stars; this does not influence our results, as spins do not impact the inferred localization when they are small as for our BNSs (Farr et al. 2016). The results also do not include the effects of uncertainty in the detector calibration. Initial results from aLIGO had 10 per cent uncertainty in amplitude and 10 deg uncertainty in phase (Abbott et al. 2017c), and this increased uncertainty in sky localization by a factor of ∼3–4 for GW150914 (Abbott et al. 2016d); however, the accuracy of calibration had been improved by the end of the run, such that its effects only increased the uncertainty in GW150914's sky localization by a factor of ∼1.3–1.5, and made negligible difference for the localization of LVT151012, GW151226, or GW170104 (Abbott et al. 2016a, 2017d).

Sky-localization accuracy and the distance estimation have been considered previously, and the three-dimensional localization remains an active area of research. Prospects for improving electromagnetic follow-up using a low-latency three-dimensional localization are discussed in Singer et al. (2016). The approach outlined in Singer et al. (2016) was used during O2 to provide prompt localizations using the BAYESTAR algorithm (Singer & Price 2016). It approximates the posterior distribution along a line of sight using an ansatz distribution, which assumes that the likelihood is Gaussian (cf. Cutler & Flanagan 1994). The resulting probability distributions can be efficiently communicated as a list of moments for pixels describing different lines of sight. At higher latencies, three-dimensional localizations were provided in O2 using the posterior samples from LALINFERENCE. These were post-processed using a clustering KDE algorithm, which is an updated version of the code used to construct the two-dimensional localizations in Singer et al. (2014) and Berry et al. (2015).[14] This code performs the KDE in Cartesian coordinates. The resulting distribution is then simplified, so that the results can be communicated using same summary statistics as

[12]The recolouring process consists of first whitening the noise (removing the colour), removing initial LIGO's frequency dependence, and then passing the noise through a filter (reintroducing colour) so that, on average, it has the aLIGO spectral density. This ensures the noise contains realistic non-stationary and non-Gaussian features, although these are not identical to those in the advanced detectors.

[13]The dimensionless spin magnitude is $\chi_{\max} = c|\boldsymbol{S}|/Gm^2$, where $|\boldsymbol{S}|$ is the modulus of the star's spin angular momentum vector and m is its mass. The limit $\chi_{\max} = 0.05$ matches that assumed for the low-spin prior used in the analysis of GW170817 (Abbott et al. 2017f).

[14]The KDE clustering algorithm, and accompanying documentation is available from github.com/farr/skyarea.

Table 2. Fractions of events with localization volumes smaller than a given size, and the fraction of searched volumes that contain fewer than the given number of galaxies in the GLADE catalogue (Dálya et al. 2018). A dash (–) is used for fractions less than 0.01.

Volume or no. of galaxies		HL recoloured	HL Gaussian	Two-detector HLV Gaussian	Three-detector HLV Gaussian	All HLV Gaussian
$\frac{\mathrm{CV}_{0.5}}{\mathrm{Mpc}^3} \leq$	10	0.01	–	–	0.02	0.02
	10^2	0.02	0.04	0.01	0.06	0.05
	10^3	0.13	0.17	0.06	0.22	0.18
	10^4	0.45	0.54	0.31	0.49	0.44
	10^5	0.97	0.99	0.66	0.82	0.78
	10^6	1.00	1.00	1.00	1.00	1.00
$\frac{\mathrm{CV}_{0.9}}{\mathrm{Mpc}^3} \leq$	10	–	–	–	–	–
	10^2	0.01	0.01	–	0.04	0.03
	10^3	0.04	0.06	0.01	0.09	0.07
	10^4	0.19	0.23	0.14	0.27	0.24
	10^5	0.65	0.77	0.31	0.50	0.45
	10^6	1.00	1.00	0.78	0.87	0.84
$\frac{V_*}{\mathrm{Mpc}^3} \leq$	10	0.02	0.01	–	0.05	0.04
	10^2	0.07	0.06	0.05	0.10	0.09
	10^3	0.23	0.32	0.13	0.27	0.24
	10^4	0.52	0.59	0.36	0.52	0.47
	10^5	0.87	0.95	0.70	0.77	0.75
	10^6	1.00	1.00	0.98	0.97	0.97
$n_*^{\mathrm{G}} \leq$	1	0.03	0.03	0.02	0.08	0.07
	10	0.20	0.24	0.15	0.26	0.23
	10^2	0.51	0.57	0.30	0.52	0.46
	10^3	0.88	0.95	0.73	0.80	0.78
	10^4	1.00	1.00	1.00	0.98	0.98

Table 3. Median localization volumes constructed using the DPGMM, and the median number of galaxies in the GLADE catalogue (Dálya et al. 2018) within these volumes.

Median	HL recoloured	HL Gaussian	Two-detector HLV Gaussian	Three-detector HLV Gaussian	All HLV Gaussian
$\mathrm{CV}_{0.5}/\mathrm{Mpc}^3$	1.2×10^4	8.9×10^3	5.2×10^4	1.3×10^4	2.0×10^4
$\mathrm{CV}_{0.9}/\mathrm{Mpc}^3$	5.4×10^4	4.0×10^4	2.9×10^5	1.0×10^5	1.3×10^5
V_*/Mpc^3	8.7×10^3	4.4×10^3	2.9×10^4	9.1×10^3	1.3×10^4
$n_{0.5}^{\mathrm{G}}$	1.3×10^2	8.1×10^1	3.5×10^2	9.2×10^1	1.5×10^2
$n_{0.9}^{\mathrm{G}}$	5.9×10^2	4.4×10^2	2.2×10^3	7.6×10^2	1.1×10^3
n_*^{G}	9.5×10^1	5.6×10^1	2.7×10^2	8.6×10^1	1.2×10^2

for the Singer et al. (2016) ansatz, giving a probability distribution for each line of sight. Our DPGMM is an alternative method for post-processing to produce three-dimensional localizations; next, we show that it is effective, and a comparison of techniques for gravitational-wave source localization is left for future work.

4 RESULTS

In this section, we describe our findings for the localization of BNSs. We begin by verifying that our reconstructed posteriors are well calibrated (Section 4.1). Then, we describe results for the (two-dimensional) sky-area analysis, before concluding with the full three-dimensional position results. A discussion of the implication of our results for multimessenger astronomy is given in Section 5.

We report values for the credible regions and volumes as well as the area or volume that would be searched (with a greedy algorithm) before discovering the true location (cf. Sidery et al. 2014b). The credible region CR_P is the smallest sky area that encompasses a total posterior probability P:

$$\mathrm{CR}_P = \min\left\{A : \int_A \mathrm{d}\mathbf{\Omega}\, p(\mathbf{\Omega}) = P\right\}, \tag{34}$$

where $p(\mathbf{\Omega})$ is the posterior probability density over sky position $\mathbf{\Omega} = \{\alpha, \delta\}$, and A is the sky area integrated over. The credible volume CV_P is the three-dimensional equivalent including distance too. We also use the distance credible interval CI_P, which we define to be the central (equal-tailed) interval that contains probability P (Aasi et al. 2013). The searched area A_* is the size of the smallest credible region that includes the true location; the searched volume V_* is the smallest credible volume that does same. The sizes of credible regions and volumes indicate the precision of our parameter estimates, whereas the searched areas and volumes fold in the accuracy too.[15]

[15]For electromagnetic follow-up, the searched area would be the minimal area of the sky that a telescope would need to cover, starting from the most probable point, before imaging the true location. However, it may

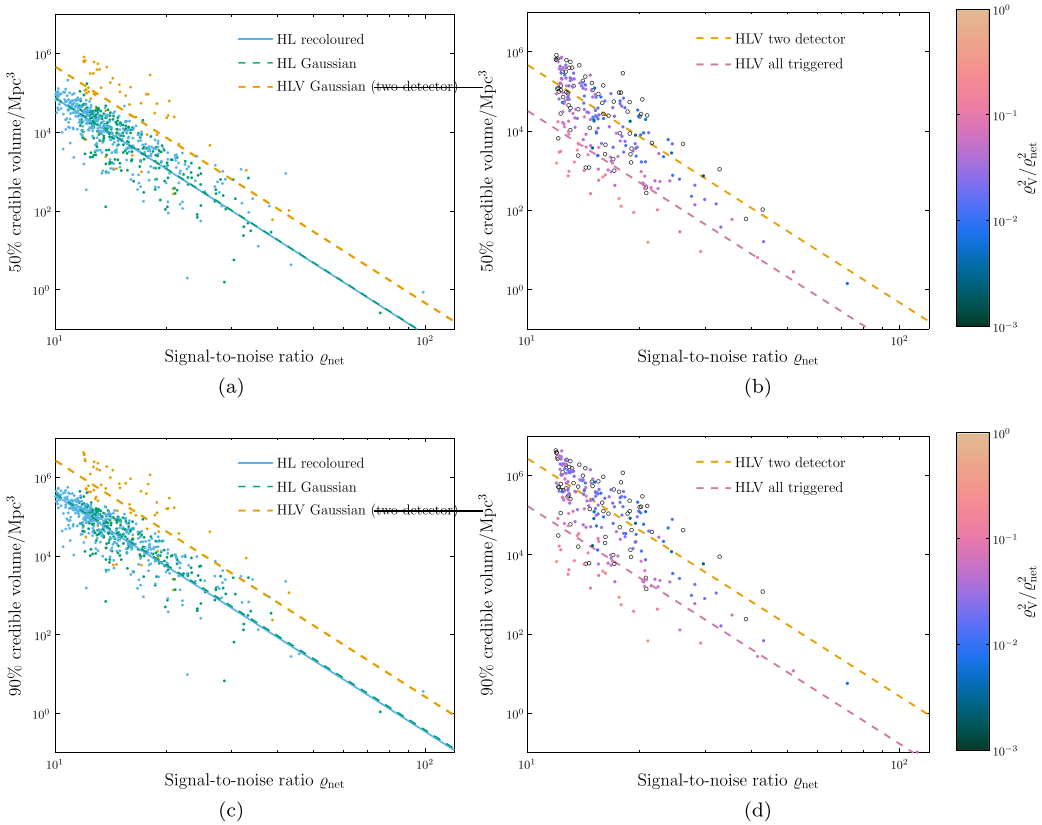

Figure 4. Localization volumes as a function of SNR $\varrho_{\rm net}$. The left column shows two-detector results, and the right column shows all HLV scenario results; the top row shows the 50 per cent credible volume $\mathrm{CV}_{0.5}$, and the bottom row shows the 90 per cent credible volume $\mathrm{CV}_{0.9}$. Individual results are indicated by points, and we include fiducial best-fitting lines assuming that the volume $V \propto \varrho_{\rm net}^{-6}$. The left column shows both HL sets of results and the HLV results where only two detectors are operation, each has its own best-fitting line. The HLV two-detector results are also shown in the right column, indicated by the open points, the three-detector results are colour-coded by the fraction of the SNR (squared) from AdV; the lines are fits to the two-detector network runs and those three-detector network runs loud enough to trigger in all detectors.

4.1 Calibration

To verify the self-consistency of results, we calculate the fraction of events that are located within the credible region or volume at a given probability. We expect that a proportion P is found within CI_P, CR_P, or CV_P (Cook, Gelman & Rubin 2006). A difference could arise if our prior does not match the injected distribution, but that should not be an issue here.[16] Fig. 1 shows the fraction of events found within a given CI_P, CR_P, and CV_P as a function of P; shown are results for three data sets, the HL Gaussian and HLV Gaussian results from Singer et al. (2014), and the HL recoloured results from Berry et al. (2015).

Since the one-dimensional distance and two-dimensional sky position probability distributions are constructed by marginalizing the three-dimensional position probability distribution, the CI_P, CR_P, and CV_P results are not independent. Using a Kolmogorov–Smirnov (KS) test (DeGroot 1975, section 9.5) to compare the expected and recovered distributions yields p-values of 0.09, 0.72, and 0.21 for the HL recoloured, HL Gaussian, and HLV Gaussian distances; 0.15, 0.15, and 0.62 for the HL recoloured, HL Gaussian, and HLV Gaussian sky areas, and 0.83, 0.94, and 0.58 for the HL recoloured, HL Gaussian, and HLV Gaussian volumes, respectively. None of the distributions show any significant deviations away from the expected results. The posteriors appear to be well calibrated .

4.2 Comparison with kernel density estimation

As a further consistency check, we can compare sky area results generated using the DPGMM to those from KDE as used in Singer et al. (2014) and Berry et al. (2015). This allows us to verify that both methods agree on an event-by-event basis. To summarize the variation in sky areas computed in different analyses, we use the log ratio (Grover et al. 2014; Farr et al. 2016)

$$\mathcal{R}_A = \log_{10}\left(\frac{A^{\rm DP}}{A^{\rm KDE}}\right), \tag{35}$$

where $A^{\rm DP}$ is a credible region or the searched area as determined by the DPGMM and $A^{\rm KDE}$ is same quantity from the KDE. The log ratio is zero when both agree.

We find there is a scatter in the log ratio around zero, as summarized in Table 1. The DPGMM results are more conservative on av-

not be possible to immediately identify a transient as the counterpart to a gravitational-wave signal; therefore, a larger area may be covered in practice to avoid false identifications. Additionally, the need to tile with a finite field-of-view telescope can further increase the actual area searched.

[16]Our priors do agree with the injected distributions, and the posterior distributions have been previously verified for sky area and distance (but not volume) in Berry et al. (2015).

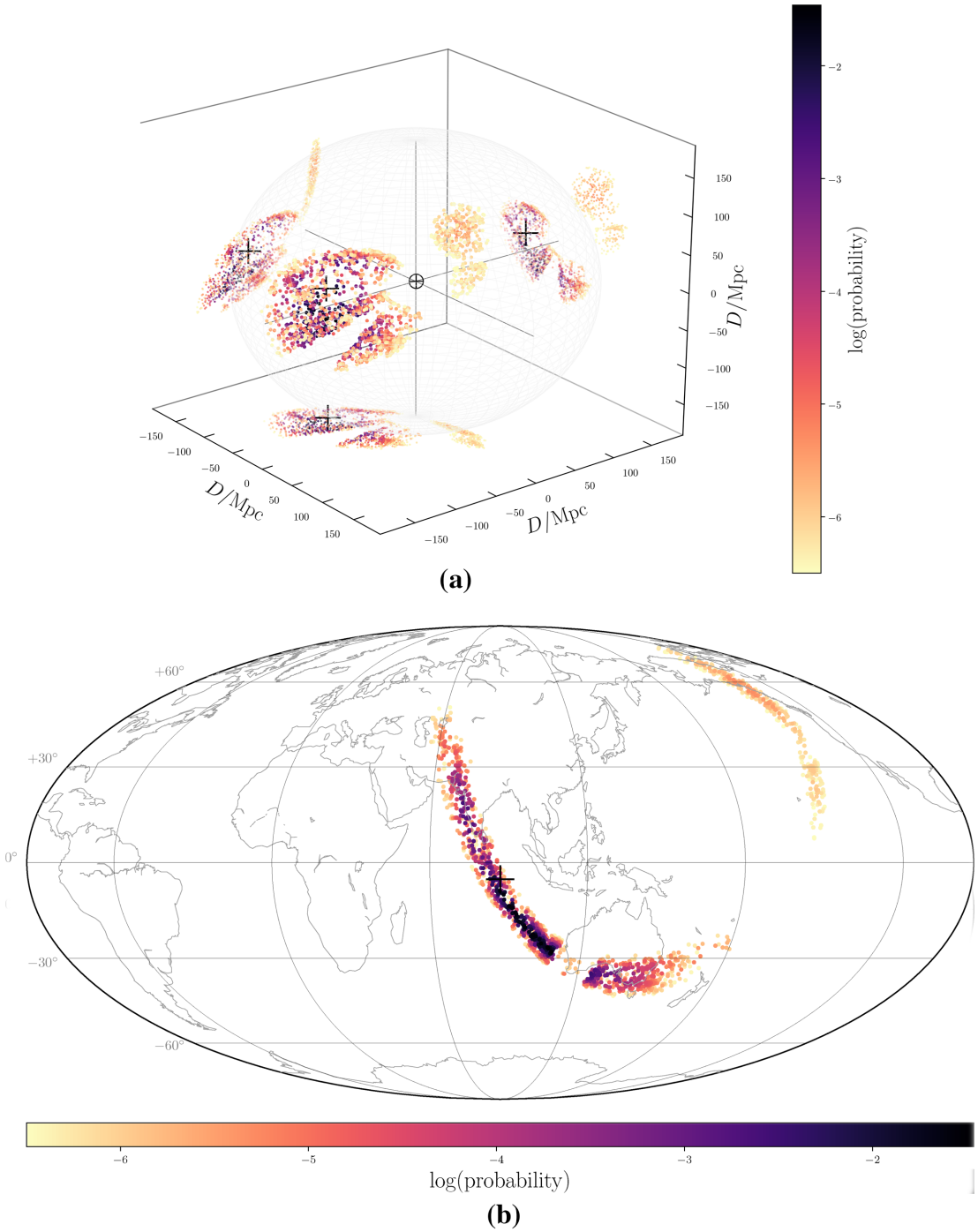

Figure 5. Example posterior distribution for possible galaxy hosts from the GLADE catalogue (Dálya et al. 2018). We show the full three-dimensional scatter plot (top) and its projection on to the plane of the sky (bottom). In both panels, galaxies are colour-coded according to the (log) posterior probability of being the host of source, and we show galaxies with the 90 per cent credible volume. In the three-dimensional plot, we show the projections along the axes directions to aid in recognizing the three- dimensional shape of the posteriors. The black crosses indicate the true source location. The gap in the larger branch of the distribution is due to the incompleteness of the catalogue in the direction of the plane of the Milky Way.

erage, being $\sim 10^{0.05} \simeq 1.1$ times larger than the KDE results. There is the largest difference in the HLV Gaussian results. This may be a consequence of these runs having a low number of (independent) posterior samples: the median number of posterior samples is 1000, whereas the median number is 8600 for both of the HL sets. Using a smaller set of posterior samples leads to less-accurate estimates for the sky localization. The sky localization areas from the two approaches agree within the typical uncertainty of $\sim$10 per cent.

We do not expect perfect agreement between the approaches since the DPGMM builds a three-dimensional probability distribution and projects this down to calculate sky areas, whereas the KDE directly computes sky areas. We expect the KDE to perform better since it is

especially designed to compute two-dimensional credible regions, and this is the case.

4.3 Measurement uncertainty

4.3.1 Sky area

Having established that the DPGMM produces sensible results, we now present results for measurement accuracies. We begin by looking at sky-localization, as a final consistency check. The sky-localization precision depends upon the SNR, scaling as $\varrho_{\rm net}^{-2}$ (Fairhurst 2009; Berry et al. 2015). We check this relationship in Fig. 2, where we plot credible regions versus SNR for the two-detector and three-detector networks. Unlike previous analyses in Berry et al. (2015) and Farr et al. (2016), we do not use the SNR reported by the detection pipeline, but the SNR as determined by the maximum of the likelihood, $\mathcal{L} \sim \exp(-\varrho^2/2)$, found by LALINFERENCE. This is necessary as we consider events for HLV where there is no trigger (which requires a single-detector SNR of 4), and hence, no contribution to the GSTLAL's network SNR, from AdV, which is less sensitive than the aLIGO instruments. With a two-detector network, the scaling with SNR changes little between the HL and HLV scenarios (or when considering different combinations of two detectors for HLV); there is slightly worse performance for HLV as a result of a decrease in frequency bandwidth at a given SNR (Singer et al. 2014). In the HLV scenario, the big change comes from the introduction of a third detector. The improvement from the third detector is continuous (Abbott et al. 2017a), ranging from providing negligible additional information to a reduction in sky area (at a given network SNR) by a factor of ∼16; this is heuristically illustrated by the fraction of the SNR from AdV $\varrho_{\rm V}/\varrho_{\rm net}$, indicated by the colour-coding in Fig. 2 (b) and Fig. 2 (d).

4.3.2 Volume

Finally, we consider the full three-dimensional localization. The cumulative distributions of localization volumes, as constructed from our DPGMM, are shown in Fig. 3. Statistics summarizing these distributions are given in Table 2 and Table 3. The three sets of results are similar; the volumes for the HL recoloured results are slightly larger than the HL Gaussian results on account of the additional low-SNR events, and the HLV Gaussian results are *larger* still as the increased detector sensitivity allows us to detect sources at a greater distance.[17]

The three-dimensional localization also depends upon the SNR. The uncertainty in the three-dimensional location can be estimated as

$$\Delta V \sim D^2 \Delta D \Delta A, \tag{36}$$

where ΔD and ΔA are the uncertainty on the distance and sky location, respectively. The distance is inversely proportional to the signal amplitude (keeping all other parameters fixed) and hence $D \propto \varrho_{\rm net}^{-1}$; from a Fisher-matrix analysis, we expect that the fractional error in the distance is inversely proportional to the SNR $\Delta D/D \propto \varrho_{\rm net}^{-1}$ (Cutler & Flanagan 1994), and we have seen that $\Delta A \propto \varrho_{\rm net}^{-2}$ (Fig. 2). Combining these, we expect that

$$\Delta V \propto \frac{1}{\varrho_{\rm net}^6}. \tag{37}$$

The credible volumes versus SNR are plotted in Fig. 4 for the two-detector and three-detector networks. The trends are roughly as expected; there is significant scatter because the SNR also depends upon other source properties such as the binary inclination and the sky position relative to the detectors. We see that, although on average the HLV scenario localization is worse than in the HL scenario, when we only consider events with significant SNR in all three detectors, the localization is better than in HL (cf. Veitch et al. 2012). Adding a third detector in the HLV scenario can improve localization by (on average) a factor of ∼15.

4.4 Applications for electromagnetic follow-up

Gravitational-wave sky localizations can be large (e.g. Abbott et al. 2016f), making the prompt search for an electromagnetic counterpart difficult. The extra information inherent in a three-dimensional localization can help optimize this search. For example, astronomers could choose to prioritize areas of the sky where the source is more probable to be close by and hence appear brighter, or adjust exposure times such that times are longer where the distance is probably larger and shorter where the distance is probably smaller. A significant improvement is potentially possible by looking for counterparts that are coincident with galaxies, as opposed to searching blindly (e.g. Nissanke et al. 2013; Hanna et al. 2014; Blackburn et al. 2015; Gehrels et al. 2016; Singer et al. 2016), and this strategy was followed by several teams searching for counterparts to GW170817 using the three-dimensional localization provided by the LVC (Abbott et al. 2017h).

Using our DPGMM, it is simple to correlate our three-dimensional posterior probability distributions with galaxy catalogues to produce a list of most probable galaxies. This only takes a few minutes to calculate; since we do not have to evaluate the DPGMM on a grid, it is *quicker* than producing credible volumes. We use the Galaxy List for the Advanced Detector Era (GLADE) catalogue (Dálya et al. 2016, 2018).[18] This is constructed from the Gravitational Wave Galaxy Catalogue (White, Daw & Dhillon 2011), the Two Micron All-Sky Survey Extended Source Catalogue (Skrutskie et al. 2006), the Two Micron All-Sky Survey Photometric Redshift catalogue (Bilicki et al. 2014), and HyperLeda catalogue (Makarov et al. 2014); it contains ∼2000 000 galaxies and is estimated to be complete at 73 Mpc and 53 per cent complete at 300 Mpc.

As an example of the end data product of our analysis, Fig. 5 shows a DPGMM localization correlated with galaxies from the GLADE catalogue (Dálya et al. 2018). The full three-dimensional posterior distribution is shown in the top panel, and its projection on to the plane of the sky is shown in the bottom panel. These posterior distributions show the characteristic shapes of localizations; they are not simple blobs but can form disjoint regions (described as jacaranda seeds in Singer et al. 2016). From the two panels, we can see the benefit of the additional information gained by considering the three-dimensional localization, instead of only a two-dimensional localization; the probable distance range is not same for all lines of sight.

[17] The median true distances of detections are 50.1 Mpc, 47.8 Mpc, and 97.0 Mpc for the HL recoloured, HL Gaussian, and HLV Gaussian sets, respectively.

[18] Available from aquarius.elte.hu/glade/.

The most probable galaxies provide a starting point for a counterpart search. Further refinements could be made, such as factoring in the stellar mass of the galaxies (cf. Nuttall & Sutton 2010), potentially using luminosity as a mass proxy (e.g. Fan et al. 2014; Hanna et al. 2014; Arcavi et al. 2017b).

In Table 3, we include the number of galaxies included in the GLADE catalogue within the credible volumes $CV_{0.5}$ and $CV_{0.9}$, and the searched volume V_*: $n^G_{0.5}$, $n^G_{0.9}$, and n^G_*, respectively.[19] These are lower limits on the true number of galaxies, but provide estimates for the number of galaxies that would be searched using the catalogue and following a greedy algorithm weighting the galaxies by probability from the three-dimensional localization. In Table 2, we give numbers quantifying the distribution of n^G_*. The number of catalogue galaxies in the localization volumes is approximately consistent with a density of one galaxy per 100 Mpc^3.

5 CONCLUSIONS

We have explained how DPGMMs can be used for post-processing of parameter-estimation studies. This technique will be useful for a variety of inference problems within astrophysics. We have applied our approach to an example from gravitational-wave astronomy, reconstructing the three-dimensional location of a BNS using results from LALINFERENCE.

The era of gravitational-wave astronomy is here, and we need to understand how to extract the maximum amount of information from signals. Localization of BNS sources is important for multi-messenger astronomy as it allows for cross-referencing with galaxy catalogues. This is beneficial when searching for an electromagnetic counterpart (Nissanke et al. 2013; Hanna et al. 2014; Gehrels et al. 2016; Singer et al. 2016), as for GW170817 (Abbott et al. 2017h) but is still useful when none is found, for example for measurements of the Hubble constant (Schutz 1986; Del Pozzo 2012; Chen et al. 2017). The DPGMM three-dimensional localizations can be be used to find the most probable source galaxies within a matter of minutes of the LALINFERENCE analysis finishing, making it useful for prompt multimessenger follow-up activities.

We constructed localization volumes for a catalogue of BNS signals appropriate for the early operation of the advanced-detector era (Singer et al. 2014; Berry et al. 2015; Farr et al. 2016). We have verified that the three-dimensional localizations are well calibrated (cf. Cook et al. 2006; Sidery et al. 2014b) and have confirmed that when distance is marginalized out, these volumes reduce to sky areas that are consistent with two-dimensional KDE results. Our credible volumes have the expected proportionality with SNR, scaling roughly $\propto \varrho_{\rm net}^{-6}$.

Our results show that localizations for detections during early observing runs would be $\sim 10^4$–$10^5 Mpc^3$, corresponding to $\sim 10^2$–10^3 potential host galaxies within the GLADE catalogue (Dálya et al. 2018). Approximately half of events have searched volumes that contain 10^2 galaxies or fewer, and a few per cent of events have searched volumes that contain a single galaxy. Since our results do not include the effects of calibration uncertainty, they would be lower bounds for any actual detections: for the (O1-like) HL recoloured data set, we find that the median 90 per cent credible volume is $5 \times 10^4 Mpc^3$ and for the HL Gaussian data set it is $4 \times 10^4 Mpc^3$; moving ahead to the (O2-like) HLV scenario, the median 90 per cent credible volume is $1 \times 10^5 Mpc^3$ for the Gaussian data set. Greater sensitivity of the detectors means that we can detect signals from a greater distance and hence are sensitive to sources in a larger volume. However, localization does improve as further detectors are added to the network: the median 90 per cent credible volume in the HLV scenario for a two-detector network is $3 \times 10^5 Mpc^3$ but for a three-detector network it is $1 \times 10^5 Mpc^3$. The localization improves rapidly as the SNR of the signal increases, and the best localization occurs when there is significant SNR from each of the three detectors. Addition of further detectors, such as KAGRA (Aso et al. 2013) or the proposed LIGO-India detector (Unnikrishnan 2013; Abbott et al. 2017a), could further improve localization and the prospects of identifying a counterpart.

[19]Since the original set of simulated signals were drawn uniformly in volume, rather than from a galaxy catalogue, we cannot identify a true host galaxy that must be imaged to find the source.

ACKNOWLEDGEMENTS

The authors are grateful for useful suggestions from the CBC group of the LIGO Scientific and Virgo Collaborations; WDP thanks Neil Cornish, Tjonnie Li, Trevor Sidery, and John Veitch for early suggestions and discussions. We thank Will Farr for discussions on localization algorithms, and thank Ilya Mandel, Jonathan Gair, Hannah Middleton, Ewan Cameron, and the anonymous referee for comments on the manuscript. We thank the other authors of Singer et al. (2014) and Berry et al. (2015) for sharing the data for this work. We also would like to thank contributors of GLADE for making the catalogue publicly available, and especially Gergely Dálya for help with its documentation and use. This work was supported in part by Leverhulme Trust research project grant and in part by the Science and Technology Facilities Council. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation and operates under cooperative agreement PHY-0757058. This work used computing resources of the LIGO Data Grid including the Atlas computing cluster at the Albert Einstein Institute, Hannover; the LIGO computing clusters at Caltech, and the facilities of the Advanced Research Computing @ Cardiff (ARCCA) Cluster at Cardiff University. We are grateful for computational resources provided by the Leonard E Parker Center for Gravitation, Cosmology, and Astrophysics at University of Wisconsin-Milwaukee. Some results were produced using the post-processing tools of the `plotutils` library at github.com/farr/plotutils and `skyarea` library at github.com/farr/skyarea. The Dirichlet process Gaussian-mixture Model is included as a module available from github.com/thaines/helit/ and our implementation for three-dimensional localization is available from github.com/wdpozzo/3d_volume. We thank GW150914, GW170104, and GW170817 for delaying the completion of this work.

REFERENCES

Aasi J. et al., 2013, Phys. Rev., 88, 062001
Aasi J. et al., 2014a, Phys. Rev., 89, 122004
Aasi J. et al., 2014b, Phys. Rev. Lett., 113, 011102
Aasi J. et al., 2015a, Class. Quantum Gravity, 32, 074001
Aasi J. et al., 2015b, Class. Quantum Gravity, 32, 115012
Abbott B. P. et al., 2016a, Phys. Rev., 6, 041015
Abbott B. P. et al., 2016b, Phys. Rev., 93, 122003
Abbott B. et al., 2016c, Phys. Rev. Lett., 116, 061102
Abbott B. P. et al., 2016d, Phys. Rev. Lett., 116, 241102
Abbott B. P. et al., 2016e, Phys. Rev. Lett., 116, 241103
Abbott B. P. et al., 2016f, ApJ, 826, L13
Abbott B. P. et al., 2016g, ApJ, 832, L21
Abbott B. P. et al., 2016h, ApJ, 841, 89

Abbott B. P. et al., 2017a, Living Rev. Relativ. , 21, 3
Abbott T. M. C. et al., 2017b, preprint (arXiv:1711.00403)
Abbott B. P. et al., 2017c, Phys. Rev. , 95, 062003
Abbott B. P. et al., 2017d, Phys. Rev. Lett., 118, 221101
Abbott B. P. et al., 2017e, Phys. Rev. Lett., 119, 141101
Abbott B. P. et al., 2017f, Phys. Rev. Lett., 119, 161101
Abbott B. P. et al., 2017g, Nature, 551, 85
Abbott B. P. et al., 2017h, ApJ, 848, L12
Abbott B. P. et al., 2017i, ApJ, 848, L13
Abbott B. P. et al., 2017j, ApJ, 850, L40
Abbott B. P. et al., 2017k, ApJ, 851, L35
Acernese F. et al., 2015, Class. Quantum Gravity, 32, 024001
Ade P. A. R. et al., 2016, A&A, 594, A13
Adrián-Martínez S. et al., 2016, Phys. Rev. , 93, 122010
Albert A. et al., 2017a, Phys. Rev. , 96, 022005
Albert A. et al., 2017b, ApJ, 850, L35
Arcavi I. et al., 2017a, Nature, 551, 64
Arcavi I. et al., 2017b, ApJ, 848, L33
Aso Y., Michimura Y., Somiya K., Ando M., Miyakawa O., Sekiguchi T., Tatsumi D., Yamamoto H., 2013, Phys. Rev. , 88, 043007
Berry C. P. L. et al., 2015, ApJ, 804, 114
Berry C. P. L., Gair J. R., 2013, MNRAS, 433, 3572
Bilicki M., Jarrett T. H., Peacock J. A., Cluver M. E., Steward L., 2014, ApJS, 210, 9
Blackburn L., Briggs M. S., Camp J., Christensen N., Connaughton V., Jenke P., Remillard R. A., Veitch J., 2015, ApJS, 217, 8
Blanchard P. K. et al., 2017, ApJ, 848, L22
Blei D. M., Jordan M. I., 2006, Bayesian Analysis, 1, 121
Cameron E., 2011, PASA, 28, 128
Cannon K. et al., 2012, ApJ, 748, 136
Chen H.-Y., Fishbach M., Holz D. E., 2017, preprint (arXiv:1712.06531)
Cook S. R., Gelman A., Rubin D. B., 2006, J. Comput. Graph. Stat., 15, 675
Coulter D. A. et al., 2017, Science, 358, eaap9811
Cutler C., Flanagan É. E., 1994, Phys. Rev. , 49, 2658
Dálya G. et al., 2018, preprint (arXiv:1804.05709)
Dálya G., Frei Z., Galgóczi G., Raffai P., de Souza R. S., 2016, Technical report, An Extended List of Galaxies for Gravitational-Wave Searches in the Advanced Detector Era. Eötvös Univ., Budapest, Hungary. Available at: http://aquarius.elte.hu/glade/GLADE_Documentation_1.3.pdf
DeGroot M. H., 1975, Probability and Statistics, Addison-Wesley, Boston
Del Pozzo W., 2012, Phys. Rev. , 86, 043011
Escobar M. D., West M., 1995, J. Am. Stat. Assoc., 90, 577
Fairhurst S., 2009, New J. Phys., 11, 123006
Fan X., Messenger C., Heng I. S., 2014, ApJ, 795, 22
Farr B. et al., 2016, ApJ, 825, 116
Ferguson T. S., 1973, Ann. Stat., 1, 209
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Freedman W. L. et al., 2001, ApJ, 553, 47
Gaebel S. M., Veitch J., 2017, Class. Quantum Gravity, 34, 174003
Gehrels N., Cannizzo J. K., Kanner J., Kasliwal M. M., Nissanke S., Singer L. P., 2016, ApJ, 820, 136
Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B., 2014, Bayesian Data Analysis, Texts in Statistical Science. 3rd edn, CRC Press, London
Goldstein A. et al., 2017, ApJ, 848, L14
Görür D., Rasmussen C. E., 2010, J. Comput. Sci. Technol., 25, 653
Gregory P. C., 2005, Bayesian Logical Data Analysis for the Physical Sciences. Cambridge Univ. Press, Cambridge
Grover K., Fairhurst S., Farr B. F., Mandel I., Rodriguez C., Sidery T., Vecchio A., 2014, Phys. Rev. , 89, 042004
Guidorzi C. et al., 2017, ApJ, 851, L36
Gupta R. D., Richards D. S. P., 2001, Int. Stat. Rev., 69, 433
Haines T. S. F., Xiang T., 2012, in Fitzgibbon A., Lazebnik S., Perona P., Sato Y., Schmid C., eds, Lecture Notes in Computer Science, Vol. 7575, Computer Vision–ECCV 2012. Springer, Berlin, p. 99
Haines T. S., Xiang T., 2014, IEEE Trans. Pattern Anal. Mach. Intell., 36, 670
Hanna C., Mandel I., Vousden W., 2014, ApJ, 784, 8
Hjort N. L. et al., 2010, Bayesian Nonparametrics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge Univ. Press, Cambridge
Im M. et al., 2017, ApJ, 849, L16
Jaynes E. T., 2003, Probability Theory: The Logic of Science. Cambridge Univ. Press, Cambridge
Kiziltan B., Kottas A., De Yoreo M., Thorsett S. E., 2013, ApJ, 778, 66
Kurihara K., Welling M., Vlassis N. A., 2007, in Schölkopf B., Platt J., Hoffman T., eds, Advances in Neural Information Processing Systems 19. MIT Press, Cambridge, MA, p. 761
Lentati L., Alexander P., Hobson M. P., Feroz F., Haasteren R. V., Lee K. J., Shannon M., Avenue J. J. T., 2013, MNRAS, 437, 3004
Levan A. J. et al., 2017, ApJ, 848, L28
Lewis A., Bridle S., 2002, Phys. Rev. , 66, 103511
Lipunov V. M. et al., 2017, ApJ, 850, L1
Lo A. Y., 1984, Ann. Stat., 12, 351
MacKay D. J. C., 2003, Information Theory, Inference and Learning Algorithms. Cambridge Univ. Press, Cambridge
Makarov D., Prugniel P., Terekhova N., Courtois H., Vauglin I., 2014, A&A, 570, A13
Mandel I., O'Shaughnessy R., 2009, Class. Quantum Gravity, 27, 114007
Metzger B. D., Berger E., 2012, ApJ, 746, 48
Neal R. M., 2000, J. Comput. Graph. Stat., 9, 249
Nissanke S., Kasliwal M., Georgieva A., 2013, ApJ, 767, 124
Nuttall L. K., Sutton P. J., 2010, Phys. Rev., 82, 102002
Özel F., Psaltis D., Narayan R., Santos Villarreal A., 2012, ApJ, 757, 55
Pan Y.-C. et al., 2017, ApJ, 848, L30
Piran T., Nakar E., Rosswog S., 2013, MNRAS, 430, 2121
Pitman J., 2006, Lecture Notes in Mathematics, Vol. 1875, Combinatorial Stochastic Processes. Springer, Berlin
Raiffa H., Schlaifer R., 1961, Applied Statistical Decision Theory. Studies in Managerial Economics, Graduate School of Business Administration. Harvard Univ., Boston
Rasmussen C. E., 2000, in Solla S., Leen T., Müller K., eds, Advances in Neural Information Processing Systems 12. MIT Press, Cambridge, MA, p. 554
Riess A. G. et al., 2016, ApJ, 826, 56
Riess A. G. et al., 2018, ApJ, 855, 136
Savchenko V. et al., 2017, ApJ, L15
Schutz B. F., 1986, Nature, 323, 310
Sethuraman J., 1994, Stat. Sin., 4, 639
Sidery T. et al., 2014b, Phys. Rev. , 89, 084060
Sidery T., Farr W., Gair J., Mandel I., 2014a, Technical Report LIGO 1400054, Having Confidence in Your Confidence: Parameter Estimation Post-processing. Available at: https://dcc.ligo.org/LIGO-P1400054/public. https://dcc.ligo.org/LIGO-P1400054/public
Singer L. P. et al., 2014, ApJ, 795, 105
Singer L. P. et al., 2016, ApJ, 829, L15
Singer L. P., Price L. R., 2016, Phys. Rev. , 93, 024013
Skilling J., 2006, Bayesian Analysis, 1, 833
Skrutskie M. F. et al., 2006, AJ, 131, 1163
Soares-Santos M. et al., 2017, ApJ, 848, L16
Tanvir N. R. et al., 2017, ApJ, 848, L27
Teh Y. W., 2010, in Sammut C., Webb G. I., eds, Encyclopedia of Machine Learnin. Springer, New York, p. 280
Unnikrishnan C. S., 2013, Int. J. Mod. Phys. , 22, 1341010
Valenti S. et al., 2017, ApJ, 848, L24
Veitch J. et al., 2012, Phys. Rev. , 85, 104045
Veitch J. et al., 2015, Phys. Rev., 91, 25
White D. J., Daw E. J., Dhillon V. S., 2011, Class. Quantum Gravity, 28, 085016

This paper has been typeset from a TEX/LATEX file prepared by the author.