This is a repository copy of *GERry: a code to optimise the hunt for the electromagnetic counter-parts to gravitational wave events*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/216510/

Version: Accepted Version

Proceedings Paper:
O'Neill, D.S., Lyman, J.D., Ackley, K. et al. (19 more authors) (2024) GERry: a code to optimise the hunt for the electromagnetic counter-parts to gravitational wave events. In: Benn, C.R., Chrysostomou, A. and Storrie-Lombardi, L.J., (eds.) Observatory Operations: Strategies, Processes, and Systems X. SPIE Astronomical Telescopes + Instrumentation 2024, 16-22 Jun 2024, Yokohama, Japan. SPIE , p. 46.

https://doi.org/10.1117/12.3018319

© 2024 The author(s). Except as otherwise noted, this author-accepted version of a paper published in Observatory Operations: Strategies, Processes, and Systems X proceedings is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

GERry: A Code to Optimise the Hunt for the Electromagnetic Counter-parts to Gravitational Wave Events.

David O'Neill[a], Joseph Lyman[a], Kendall Ackley[a], Danny Steeghs[a], Duncan Galloway[b], Vik Dhillon[c,d], Paul O'Brien[e], Gavin Ramsay[f], Kanthanakorn Noysena[g], Rubina Kotak[h], Rene Breton[i], Laura Nuttall[j], Enric Pallé[d], Don Pollacco[a], Krzysztof Ulaczyk[a], Martin Dyer[c], Felipe Jiménez-Ibarra[b], Tom Killestein[g], Amit Kumar[a], Lisa Kelsey[j], Ben Godson[a], Dan Jarvis[c], and the GOTO Collaboration.

[a]Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
[b]School of Physics & Astronomy, Monash University, Clayton VIC 3800, Australia
[c]Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK
[d]Instituto de Astrof ísica de Canarias, E-38205 La Laguna, Tenerife, Spain
[e]School of Physics & Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK
[f]Armagh Observatory & Planetarium, College Hill, Armagh, BT61 9DG, UK
[g]National Astronomical Research Institute of Thailand, 260 Moo 4, T. Donkaew, A. Maerim, Chiangmai, 50180 Thailand
[h]Department of Physics & Astronomy, University of Turku, Vesilinnantie 5, Turku, FI-20014, Finland
[i]Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK
[j]Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK

arXiv:2407.18642v2 [astro-ph.IM] 29 Jul 2024

Abstract. The search for the electromagnetic counterparts to gravitational wave (GW) events has been rapidly gathering pace in recent years thanks to the increasing number and capabilities of both gravitational wave detectors and wide field survey telescopes. Difficulties remain, however, in detecting these counterparts due to their inherent scarcity, faintness and rapidly evolving nature. To find these counterparts, it is important that one optimises the observing strategy for their recovery. This can be difficult due to the large number of potential variables at play. Such follow-up campaigns are also capable of detecting hundreds or potentially thousands of unrelated transients, particularly for GW events with poor localisation. Even if the observations are capable of detecting a counterpart, finding it among the numerous contaminants can prove challenging. Here we present the Gravitational wave Electromagnetic RecoveRY code (GERRY) to perform detailed analysis and survey-agnostic quantification of observing campaigns attempting to recover electromagnetic counterparts. GERRY considers the campaign's spatial, temporal and wavelength coverage, in addition to Galactic extinction and the expected counterpart light curve evolution from the GW 3D localisation volume. It returns quantified statistics that can be used to: determine the probability of having detected the counterpart, identify the most promising sources, and assess and refine strategy. Here we demonstrate the code to look at the performance and parameter space probed by current and upcoming wide-field surveys such as GOTO & VRO.

Keywords: Multi-messenger astronomy,strategy,survey.

*David O'Neill, David.O-Neill@Warwick.ac.uk

1 Introduction

Astrophysics has long relied almost solely on the electro-magnetic force to inform us about the Universe. The first detection of gravitational waves (GW) by LIGO in 2015 however, paved the way for a new era of routine multi-messenger astrophysics through the combination of traditional electromagnetic (EM) datasets complemented by GW information. The current array of GW inferometers are capable of detecting the mergers of stellar mass compact objects. While a combined EM and GW detection of binary black hole (BBH) mergers are not expected due to the expected lack of EM emission

from such an event, neutron star - black hole (NSBH) and, in particular, binary neutron star (BNS) mergers are capable of producing an EM signal. Indeed, short gamma-ray bursts (SGRB) originate from the powerful relativistic outflows produced in BNS[1] and possibly from NSBH merger events.[2]

While SGRBs are regularly detected, they are not the only EM emission released from BNS mergers. The decay of r-process elements in the merger ejecta can also produce longer-lasting but faint kilonovae (KNe). However, the number of detected KNe remains very small, with only a handful of photometric candidates, and two spectroscopically confirmed cases: the notable KN AT 2017gfo (see the review by Margutti & Chornock[3]) and a KN associated with the GRB 230307A.[4]

This small sample is not only driven by the inherent low rates of BNS mergers, but also due to the faint and rapidly evolving nature of KNe which may only remain detectable in the optical for $\leq$5 days,[5,6] compared with supernovae (SNe), which generally evolve over timescales of weeks to months.

The current array of survey telescopes such as ATLAS,[7] BlackGEM,[8] DES,[9] GOTO,[10,11] PanStarrs,[12] and ZTF[13,14] typically have single or multiple mounts with a combined field of view on the order of 10–100s deg^2 and provide high cadence data across much of the sky, whereas the upcoming 8.4m Vera C. Rubin Observatory[15] sacrifices field of view and cadence ($\sim$ 9.6 deg^2) for a much deeper imaging. Despite this current array of wide field instruments, KNe remain elusive due to their rapid, faint and rare nature.

The distance and sky-localisation of GW events are sensitive to the number of detectors online during the event. With still a low number of detectors, GW localisations are typically heterogeneous, and place challenging requirements on EM follow-up campaigns, which often need to not only cover large sky areas, but do so at high cadence and with good sensitivity. Due to the variety of GW localisation possible, coupled with different EM signatures expected, identifying a single strategy to optimise recovery probability EM counterparts is complex.

In this paper, I show how the Gravitational wave Electromagnetic RecoveRY code (GERRY) can help in these respects, by providing a method to rapidly provide an in depth statistical analysis of GW follow-up observations and returning the probability of detecting a KNe event that folds in observing footprints, temporal and wavelength coverage, Galactic extinction, and observing sensitivity in comparison to the expected counterpart temporal evolution. This can be used to determine whether a KN is likely to be among the detected sources as well as to test the effectiveness of the observing strategy for a given event. Typically such analysis is performed long after any KN would have faded. By providing this information rapidly while the KN is still potentially visible, it can allow users to adapt their strategies and determine whether the mobilisation of further follow-up resources is required.

2 Methods

2.1 Multi-Order Coverage skymaps

One of the most popular methods of handling sky projections in modern astronomical data analysis is the use of Hierarchical Equal Area isoLatitude Pixelization (HEALPix[16]). HEALPix allows the sky to be subdivided into pixels or tiles, the number of which is dependent on the desired resolution. This data format was adopted by LIGO for producing the probabilistic skymaps output by 3D localisation codes such as BAYESTAR and LALinference. While the HEALPix grid itself is 2D, the localisation

codes encode the probability contained within each tile as well as the estimated distance in Mpc, thus turning the 2D HEALpix skymap into a 3D map.

This format was further built on by utilising an multi-resolution grid in the form of Multi-Order Coverage skymaps (MOCs[17]). Originally developed to enable the rapid spatial/coverage queries such as intersections and unions on a HEALPix grid, it was incorporated into LIGO GW maps where credible regions/contours with higher probability densities are assigned higher HEALPix resolutions with the inverse being true for lower probability dense areas. The primary advantage of this format over the previous 'flat' resolution maps is they are less computationally intensive to both handle and analyse.

2.2 healpix-alchemy

GERRY is primarily built on HEALPIX-ALCHEMY,[18] a Python package that enables rapid querying of GW MOCs using SQLAlchemy in conjunction with a PostgreSQL 14 database. HEALPIX-ALCHEMY utilises the multi-range data type in the form of a range set in PostgreSQL 14, whereby each HEALPix tile at a given resolution is represented by a *rangeset* corresponding to the indices of the much smaller constituent tiles (sub-pixels) at the highest resolution at which the tile indices can be stored as 64bit integers. Given the sequential nature of these "sub-pixels", one HEALPix tile at any resolution can always be represented by a single range set containing the minimum and maximum indices of the sub-pixels, which will take the form, e.g., [576460752303423488, 577586652210266112). Therefore, the GW skymap can be broken down into an array of rangesets for each tile. These tiles are then stored in the database along with the tiles' probability density values, distance estimate and distance uncertainty. The same process is performed for the all image in the follow-up campaign, whereby the footprint of each image is mapped onto a multi-resolution HEALPix grid. The code employ MOC resolution of the constituent HEALPix tiles such that higher resolutions (smaller tiles) are at the edges and corners of the images and lower resolution tile (larger tiles) in the centre, in order to maintain the original image field of view as much as possible whilst reducing computational cost later. Now that both the probabilistic GW skymap and the images have been broken down into HEALPix tiles and stored in the database, their HEALPix rangesets can now be efficiently queried against each other. Given the sequential nature of the rangesets, if the range of one rangeset overlaps that of another then the tiles they represent also must overlap in sky position by an amount corresponding to this rangeset overlap. This effectively allows one to trivially find which image tiles overlap or contain GW skymap tiles, the size of the overlap, and the area and the probability contained within the image tiles (see Singer et al. 2022[18] their figure 4).

2.3 GERRY*: from 2D to 3D*

While HEALPIX-ALCHEMY allows the user to easily find the 2D properties of your follow-up campaign, GERRY extends this to 3D by utilising the encoded distance information contained within the skymaps in combination with the sensitivity of the field images. GERRY operates under a simple concept, if the distance to the GW event at a particular sky position is known from the skymap, one can use this alongside the expected temporal evolution of the counterpart to estimate the expected apparent magnitude at the time any image of this sky position was taken. By converting the encoded distance distribution to a brightness distribution, the image depths then provide a means to calculate the fraction

of this brightness distribution that was recovered, and thus the fraction of 3D probability recovered in each tile. These recovery fractions provide quantified values such as the total probability volume recovered (P_{vol}) or the probability of detecting the counterpart in at least once across the imaged sky ($P_{transient}$ or P_{tr}). In the following subsections we show in detail how these values are calculated.

2.4 General processing

2.4.1 Find image coverage and overlap

The first step is to identify skymap tiles that are overlapped or contained wholly within an image by querying their respective rangesets for overlaps. Each skymap tile and overlapping image tile combination is stored in a PANDAS[1] dataframe alongside information that will be used in processing: skymap probability density, area, and distance information, and the image tile observing time, filter, and depth (and uncertainty). Most skymap tiles are not contained entirely within a single image tile, therefore the area of the skymap tile covered by every partially overlapping image tile has to be calculated. For each skymap tile and overlapping-image tile combination, the overlap area is calculated using the size of the overlapping rangeset multiplied by the area of a single sub-pixel ($\sim 1.2\times10^{-14}$ deg^2). The probability recovered by a particular skymap-image tile overlap is then the probability density multiplied by the overlap area.

2.4.2 Generate counterpart magnitudes

The total recovered 3D probability can not be known unless we estimate the light curve evolution of the counterpart. The probabilities of finding a KN-like counterpart will be a strong function of time since the event, due to their rapid evolution. The potential diversity of GW EM counterparts has yet to be mapped given that for KNe, GW170817-KN2017gfo remains the only KN event with extensive light curve coverage across the EM spectrum. Until such a time we have a robust sample of KNe light curves, we rely on models. An input light curve file to GERRY contains the phase and expected absolute magnitude and magnitude uncertainty of the counterpart. A separate light curve is required for all the image filters. GERRY will calculate the counterpart phase at the time the image was taken (with respect to the GW event) and find the closest point in time in the given model light curve. Given that we are using model light curves, the density of photometric points can be arbitrarily increased in to improve temporal resolution. If an observed light curve is being used, we recommend interpolating between the points down to a time step on the order of $\sim$1–2 hours for rapidly evolving transients.

The apparent magnitude of the counterpart in a given skymap tile at a given time can then be found using the absolute magnitude and the embedded GW skymap distance and distance uncertainty information. Galactic foreground extinction is also calculated at the position of the skymap tile, (specifically the central coordinates of the tile) using the using the Planck collaboration dust map queried through the DUSTMAPS Python package[19] and applied to the estimated magnitude if the filter bandpass has been provided, this can also be turned off if desired. This process generates the expected magnitudes for the counterpart in all skymap-image tile overlaps.

[1] https://pandas.pydata.org/pandas-docs/stable/index.html

2.5 Polygons and Probabilities

The counterpart recovery fraction at a particular sky position is calculated by modelling the counterpart's probability distributed over magnitude space as a normal cumulative distribution function with mean and standard deviation determined by the estimated magnitude and uncertainty. The total probability recovered at this position of the sky, P_{int}, is then the value of the cumulative distribution at the achieved image depth. The uncertainties are determined by repeating this process for the upper and lower image depth uncertainty bounds. This adds to the dataframe of skymap-image tile overlaps the recovered probability values.

It is likely, however, that a skymap tile has been imaged multiple times, and it is imperative for the following calculations to know exactly what part of the skymap tiles each overlapping image tile physically covered. For example, two image tiles may overlap 50% of a particular skymap tile – this could mean 100% or 50% coverage of that skymap tile, depending how the image tiles are arranged on-sky, which naturally would affect recovery statistics. In order to facilitate this, all overlapping image HEALPix rangesets for a particular skymap tile are merged and split into unique non-overlapping rangesets or "polygons" for each overlapping image combination. This results in a list of unique skymap-image tile combinations to calculate the probability statistics such as P_{vol} and P_{tr} .

P_{vol} is the maximum 3D volume probed over the course of the campaign. For skymap tiles observed multiple times, only the most constraining observation (i.e the observation which recovered the most probability) is considered. It is the sum of the maximum probability achieved at each part of the imaged sky. A P_{vol} value of 100% indicates that all of 3D probability space for a given event (assuming a given set of light curve models) has been recovered. In essence it is the often-quoted P_{fov} – the sum of the 2D probabilities contained within the bounds of the images – extended to 3D space. Unlike P_{fov} however, P_{vol} more informatively folds in the limitations of the images' sensitivities in the face of an expected counterpart's evolution.

P_{tr} is the probability of detecting the desired counterpart *at least once* across the imaged sky. In cases where a region of sky has been imaged multiple times, P_{vol} only includes the probability contribution of only the most constraining tile. P_{tr} on the other hand combines the probability contribution from all overlapping image tiles. E.g. for a single skymap polygon i, the P_{tr} value would be:

$$P_{tr,poly_i} = P_{tot,poly_i}(1 - [(1 - P_{int_1,img_1}) \times (1 - P_{int_2,img_2}) \ldots \times (1 - P_{int_n,img_n})]) \tag{1}$$

Where $P_{tr,poly}$ is the probability of detecting the counterpart at least once within this polygon, $P_{tot,poly}$ is the total 2D probability contained within the polygon, (the skymap probability density multiplied by the area of the polygon), and P_{int_n,img_n} is the recovered probability of this polygon from the nth image tile. The final P_{tr} value is then simply the sum of all $P_{tr,poly}$ values for the total number of polygons N:

$$P_{tr} = \sum^{N} P_{tr,poly_i} \tag{2}$$

In cases where no images overlap in the follow-up campaign (and assuming a non-infinite image depth):

$$P_{fov} \geq P_{tr} = P_{vol} \tag{3}$$

For follow-up with overlapping images, the probability volume is being sampled more than once, therefore:

$$P_{fov} \geq P_{tr} \geq P_{vol} \tag{4}$$

Another output of GERRY is the P_{tr} skymap. This skymap contains P_{tr} density in place of the standard localisation probability density. Since P_{tr} is the probability of detecting the EM counterpart at any position of the sky, it can be used for a much more focused search for EM counterpart candidates. The P_{tr} density will naturally be negligible for areas of the sky that suffer from e.g high dust extinction, poor observing conditions, imaging 'too-late' after the event. By cross-matching detected sources with the P_{tr} skymap in place of the standard skymap, the search area can be properly prioritised for human-inspection of candidates, allowing for a much more rapid identification of any counterpart candidate.

3 Detailed Example: S190425z

Of LIGO's O3 observing run, S190425z was the only statistically significant BNS merger. At an estimated distance of $\sim$150 Mpc, it had the potential to produce a detectable, albeit faint, EM counterpart. However, localisation was poor ($\sim$9000 deg^2) due to only being detected in the Livingston (L1) and Virgo (V1) detectors, and much of the probability was spread across both northern and southern hemispheres.

At the time, a GOTO prototype system (GOTO-4) was in operation,[10,20] it consisted of a single mount at Roque de los Muchachos observatory in La Palma, and was equipped with four 0.4m telescopes with a combined field-of-view of $\sim$20 deg^2 . GOTO observations of the skymap spanned approximately 3 days from 2019-04-25 20:40 to 2019-04-28 05:15 (+0.5 to 2.8 days post trigger) and consists of 1948 L-band (400–700 nm) images covering 2155 deg^2 . Figure 1 shows the outputs of GERRY of this observing campaign, including the skymap, observed regions, and spatially resolved reddening, image depth, and recovered probability information. The total 2D probability covered was P_{fov} = 32% owing to much of the probability being in-accessible to the GOTO-4 prototype. The light curve used was generated with MOSFIT[21] a python package typically used to fit semi-analytical light curves to photometry points for parameter estimation. Here, we use it in the 'generative' mode, where MOSFiT draws samples from a given list of parameter priors and generates model light curves in the desired filters. For S190425z, we use the input parameters detailed in Smartt et al. (2023).[22] Like AT 2017gfo, the expected counterpart similarly peaks and fades rapidly within 24 hours, albeit $\sim$0.7-1.0 mag fainter. Using this light curve we find the total probability volume probed by the GOTO prototype was P_{vol} =$1.61^{+1.28}_{-0.83}$% with P_{tr} =$3.01^{+1.86}_{-1.36}$%, the coverage and depth is comparable to that achieved by other surveys and serves to highlight the challenges of such follow-up, particularly when large areas of the skymap are inaccessible.

Figure 2 shows the expected light curve evolution along with the image depths achieved over the course of the campaign. While images were taken at the expected KN peak, they did not probe the required depth to recover the expected counterpart. This is primarily due to the distance of the source as the majority of the imaged probability suffered little dust reddening. The vast majority of the P_{vol} recovered on the first night when the KN was at peak, with the subsequent rapid fading rapidly diminishing the recovered probability on subsequent nights.

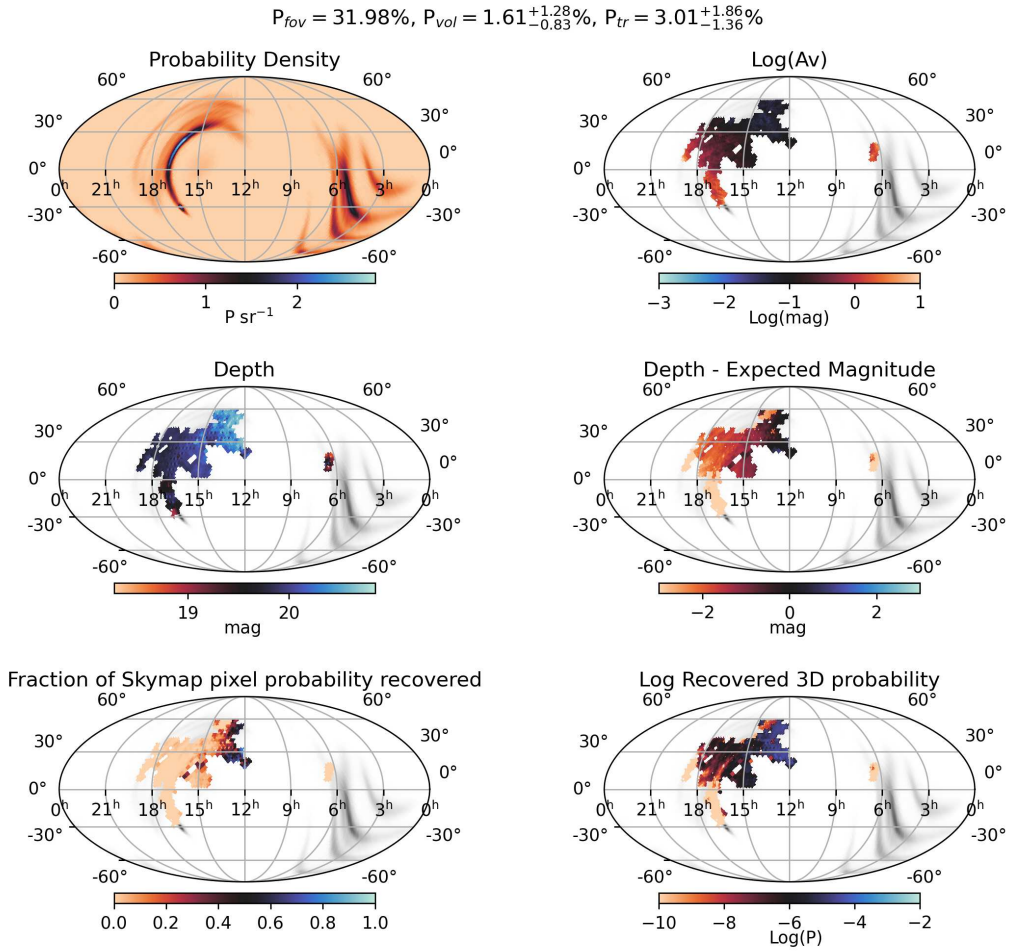

Fig 1 Skymap coverage information output by GERRY. **Top left:** The standard skymap probability density information. **Top right:** Line of sight extinction of all imaged areas of the sky. **Mid left:** The average depth achieved at each part of the imaged sky. **Mid right**: The difference between the depth and expected counterpart magnitude. **Bottom left:** The ratio of the recovered 3D tile probability $P_{int,tile}$ to the total probability contained within the tile $P_{tot,tile}$. **Bottom right:** The maximum 3D probability recovered in each tile.

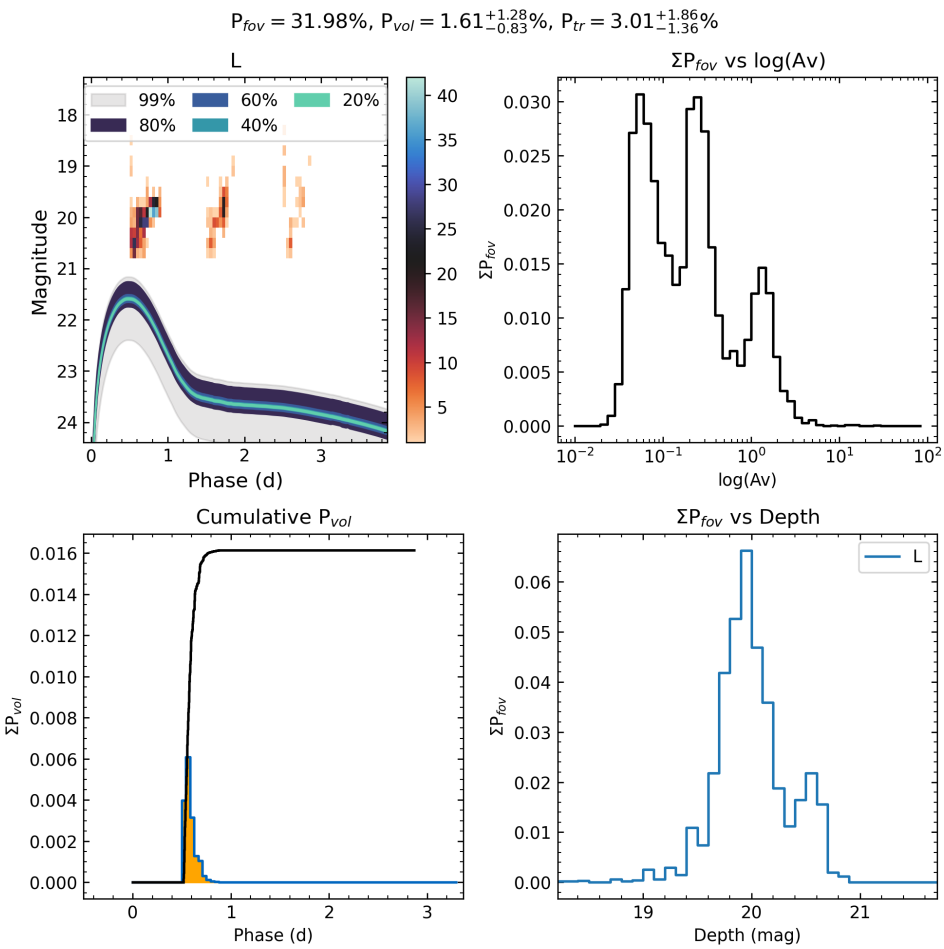

Fig 2 Plot showing the output from GERRY detailing the results of the codes analysis of the follow-up of S190425z. **Top left:** Plot showing the expected light curve distribution and evolution, along with 2D histogram data showing the density of images binned into observation times and depths achieved. The light curve distribution is generated by placing the model light curves in all imaged skymap tiles and calculating the apparent magnitude based on the distance estimate contained within each skymap tiles. **Top right:** How the imaged 2D sky probability was distributed across milky way reddening. Most of the skymap suffered little extinction ($A_V < 0.2$ mag). However some images covered the galactic plane resulting in very high extinction values. **Bottom left:** The probability volume (P_{vol}) evolution over the course of the campaign. In this case, the overwhelming majority of the recovered volume was done so in the first night of the campaign while a KN would've been at peak. The fading of the KN in the subsequent nights results in little to no new volume recovered. **Bottom right:** The imaged 2D sky probability distribution over image depth.

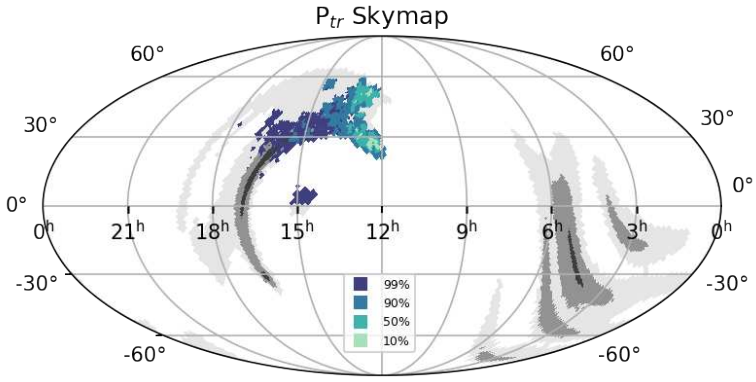

Fig 3 The P_{tr} skymap for the GOTO4 follow-up campaign of S190425z. While the follow-up covered 2155 deg^2 , the 90% credible area for a counterpart is ∼720 deg^2 .

Of the P_{tr} recovered (3%), much of it was localised in the northern region of the sky where higher image depths were achieved close to the expected KN peak (figure 3. The viable area of the candidate is reduced from 2155 deg^2 to 720 deg^2 , and candidates can be ranked based on this statistic when presented to humans for verification, which would facilitate faster identification.

4 Optimisation of follow-up strategy

LIGO's current O4 observing period is define by two periods. The first, 'O4a' started on the 24th May 2023 and ended on the 16th January 2024. The first period consisted primarily of the LIGO Livingston (L1) and Hanford (H1) detectors both operating with a BNS sensitivity range ∼120-160 Mpc with Kagra (K1) contributing for a period of 1 month with a sensitivity of ∼1 Mpc. The second observing sub-period, 'O4b' started on the 10th April 2024 and is currently underway as of the time of writing with the planned end date of February 2025 and consists of both the previous aforementioned detectors along the Virgo (V1) which is operating with a sensitivity of 55-60 Mpc.

The typical localisation area of a given gravitational wave relies heavily on triangulation of signals, and so depends strongly on the number of detectors online at any given time (see 4). Events detected by a single detector are typically poorly constrained and localisation maps consist of a dipole with the localisation area typically comparable to half the sky (∼21,000 deg^2). Two detector events are the most dynamic with localisation's ranging from small 'spots' on the order of 100s to 1000s deg^2 to larger ring-like or 'banana' shaped profiles stretching across both hemispheres with localisation areas of 1000s to 10,000s deg^2 depending on the SNR of the GW. Which strategy (or strategies) to employ is a critical consideration if trying to detect any EM counterpart to these events as the effectiveness of a strategy could change considerably depending on the localisation skymap and the estimated distance to the event.

Given that the primary objective of such follow-up is to detect EM counterparts, P_{tr} is an effective, quantitative measure of a given strategies performance assuming a model light curve. In this section we will demonstrate how using GERRY in tandem with both simulated GW events and follow-up campaigns allows for an effective comparison of the effectiveness of observing strategies. We will

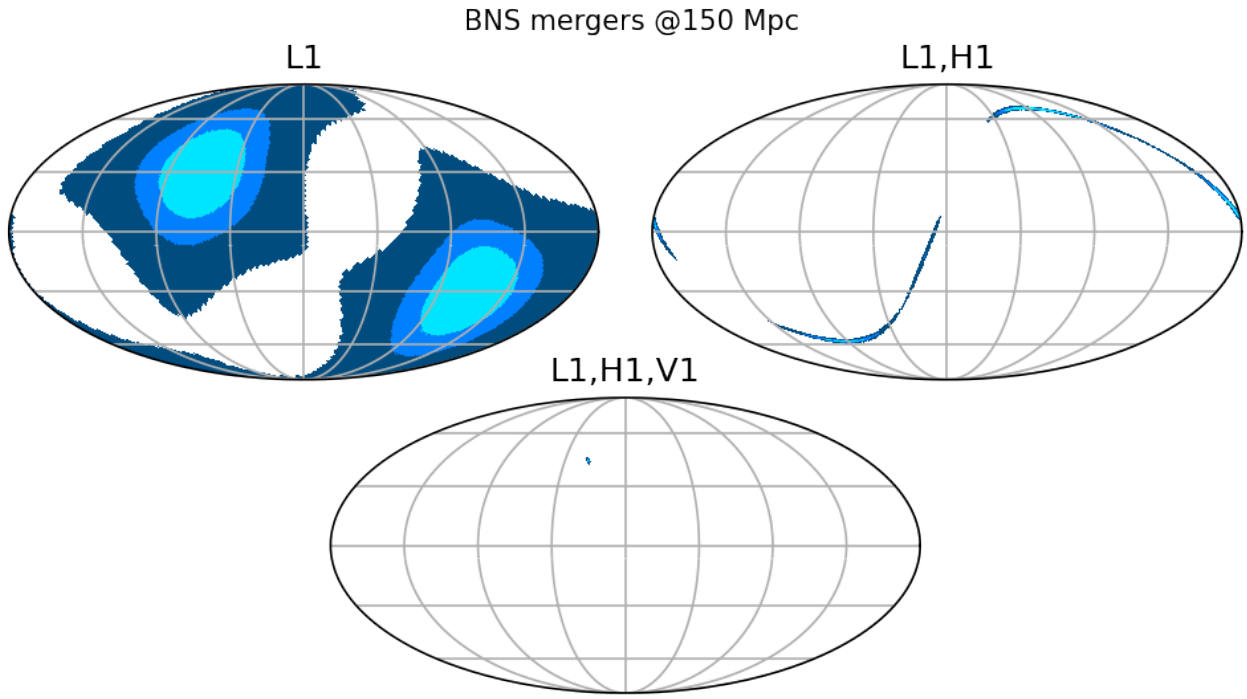

Fig 4 Plot showing the 90%, 50% and 25% localisation areas on sky for different simulated BNS mergers at ∼150 Mpc using a single detector (L1), both LIGO detectors (L1 and H1) and the LIGO detectors with Virgo (L1, H1 and V1). These were generated using the methods outlined in section 4.1.

also explore how a typical wide-field sky-survey like GOTO compares with the upcoming Vera C. Rubin Observatory (VRO).

4.1 Simulation setup

The GW events were simulated using the LIGO-SKYMAP[2] python package. One, two and three detector events were simulated up to distances of 300 Mpc in order to explore as much of the localisation area and distance parameter space as possible. The basic recipe for producing these events follow the instructions presented in the LIGO-SKYMAP documentation[3], with only minor changes. The simulations were set to trigger over a 1 year period from January 2024 over a uniform distance distribution up to 300 Mpc. The uniform distance (rather than volume) distribution was chosen so that closer distances ($\leq$ 100 Mpc) were being adequately sampled. In order to facilitate sampling over a wider range of localisation areas, the net GW SNR required for a 'trigger' from the detectors was lowered from the default value of 12.0 to 5.0. We limit the neutron star mass range to 1.1–2.0 $M_{\odot}$ in line with estimates from P. Landry & S. Jocelyn (2021).[23] For each detector combination, a detection by all of the detectors is required. GW detector sensitivities were not scaled and so use the default values. One detector events only use L1, two detector events use L1 and H1 and finally three detector events use L1, H1 and V1; ∼73 one-detector, ∼270 two-detector, and ∼103 three-detector events were simulated.

[2]https://pypi.org/project/ligo.skymap/
[3]https://lscsoft.docs.ligo.org/ligo.skymap/quickstart/bayestar-injections.html

The follow-up simulations were performed using a modified version GOTO-TILE (M. Dyer[4]), a package used by GOTO to define the follow-up schedule of a skymap. GOTO-TILE takes basic telescope information such as field-of-view, location, number of mounts and an pointing/exposure time to simulate the follow-up of a probabilistic skymap over a set duration. A fixed grid of tiles is generated on sky according to the field of view of the telescope. When used with a skymap, the code identifies which tiles cover the associated 95% probability contour. At a time step corresponding to the typical pointing time, tile visibility is calculated based on twilight and altitude constraints. From the remaining visible tiles, those which have been imaged least are prioritised. For tiles with the same number of observations, the contained probability and airmass are calculated and a tiebreaker is calculated by weighting the airmass using the contained probability.

GOTO is now at design specification, with four mounts each having a field of view of $\sim$44 deg^2: two mounts at Roque de los Muchachos Observatory in La Palma and a further two at Siding Spring Observatory in Australia. All simualated observations were taken in the wide L-band filter. We explore two strategies, using shallow 4×90 s exposure sets per pointing to maximise 2D coverage and a deeper 12×60 s exposure set to maximise recovered volume per pointing. A 60 s overhead is assumed at each pointing to simulate slewing, rapid focusing and readout procedures giving a total of 420 s and 780 s for each strategy respectively. The image depths at each pointing were informed by archival GOTO image data. Images from the GOTO database for each of the above exposure times were queried at different airmasses. A curve was fit to airmass and average depth values which was then used to estimate the appropriate depth for each simulated image. In the case of the deeper 12×60 s strategy this results in a depth of 20.6 and 20.26 mag at airmass = 1 and 2, respectively. The shallower 4×90 s strategy had depths of 20.12 and 19.45 mag at corresponding airmasses.

For VRO, we used a single mount with 9.6 deg^2 field of view located at Cerro Pachón in Chile. Images were simulated in the 5 loaded filters (g,r,i,z,y), each with 30 s exposures. While VRO images are somewhat circular or hexagonal in nature, we simply assume a square field of view with the same total field of view area. We use the overheads detailed in Jones, R. L. et al. (2020)[5]. Along with four filter change times of 120s each with an additional 3 seconds for each readout and 30s for slew time for each pointing results in a total pointing time of 675s. We also simulate a three day follow-up duration, although this is unlikely in reality as targeted GW follow-up is not a primary science goal of VRO. However, it allows for the assessment of what is theoretically achievable from an 8 m class survey telescope. The strategy for VRO follow-up of GW events is not yet finalised, and so pointing strategy we use here is representative (an alternative could save filter change time at the expense of less colour information, for example). Given VRO has not yet achieved first light, we have no real data to sufficiently model the depth/airmass relation therefore we simply use the values for single exposures in each filter detailed here[6]. This is likely going to result in an over-estimate as it neglects the effects of sub-optimal observing conditions, therefore we will leave much more detailed modelling to other studies.

In order to estimate light curve properties and evolution for a typical KNe we again used MOSFIT.[21] We used the listed input priors found in table 1 of M. Nicholl et al. (2021),[5] with the exception

[4]https://github.com/GOTO-OBS/goto-tile

[5]Survey Strategy and Cadence Choices for the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST)

[6]https://www.lsst.org/scientists/keynumbers

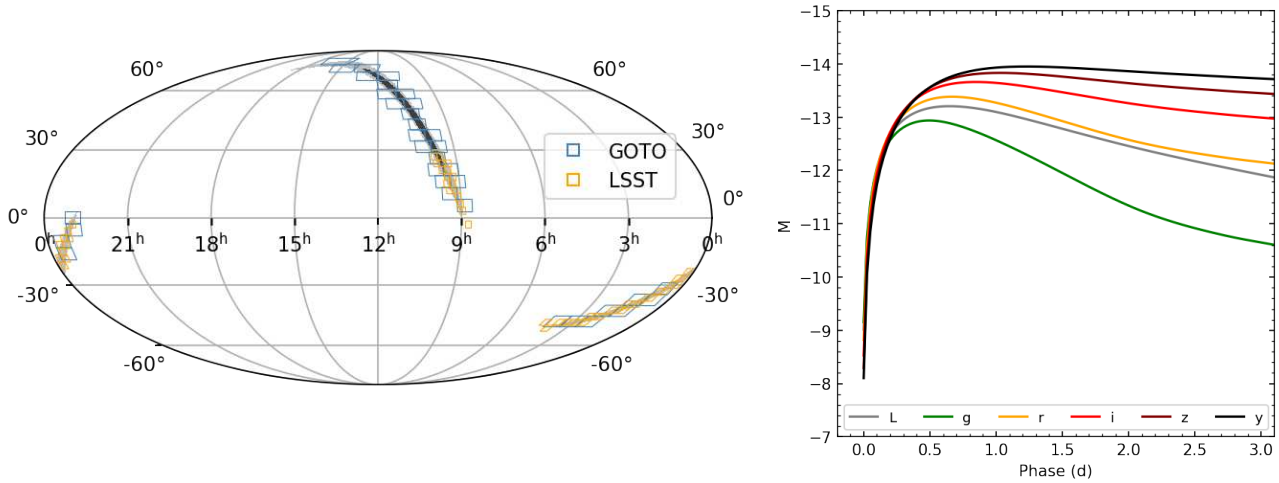

Fig 5 Left: Simulated localisation skymap and observational follow-up campaign of a two-detector GW event using GOTO (blue) and VRO (orange). **Right:** The fiducial KN lightcurve in GOTO L-band as well as the filters used in the VRO simulations.

of limiting chirp mass M and mass ratio q to correspond to the aforementioned neutron star mass range of 1.1–2.0 $M_{\odot}$. 10,000 light curves were generated and and the average absolute magnitude light curve was calculated. For all of the following simulations, unless otherwise stated, we disabled the effects of reddening as we primarily want to assess the effects of localisation area and distance on the instruments and strategies rather than Galactic effects.

An example of a simulated follow-up campaign and the light curves used can be seen in figure 5. The light curves peak at $M = -13$ in L and g, almost two magnitudes fainter than KN 2017gfo. How representative KN 2017gfo is of the KN population is poorly constrained, although it appears well withing the range of luminosities seen for other candidate KNe, if somewhat on the luminous side.[24] There may also be shortcomings of the simulated light curves, but we are primarily using them for our strategy-informing analysis where the observing campaign is the variable factor.

4.2 Results: GOTO strategy

The results of the simulations and the GERRY analysis are shown in Figure 6. Rather unsurprisingly, increasing the number of detectors results on average higher P_{tr} values as it becomes easier to completely cover the skymap. For the one detector scenarios, the faster 4×90 s strategy outperforms the 12×60 s strategy for BNS events less than 30 Mpc assuming an 'average' KN light curve. However, the deeper strategy starts to rapidly outperform the shallower one at larger distances as the primary limiting factor changes from 2D sky coverage to depth. For AT 2017gfo-like light curves, this threshold is increased to $\sim$100 Mpc

For two and three detector events, the size of the 90% credible area only has a minor effect as shown in Figure 7. GOTO's field of view is large enough that it is able to fully image skymaps with localisation areas in the order of 10–100s deg^2 with relative ease. Similarly to the one detector scenario, for the 'average' KN light curve the deeper strategy is preferable for GW events at $\geq$30Mpc. If KNe are more similar to AT 2017gfo we can see that both the shallower and deeper strategies

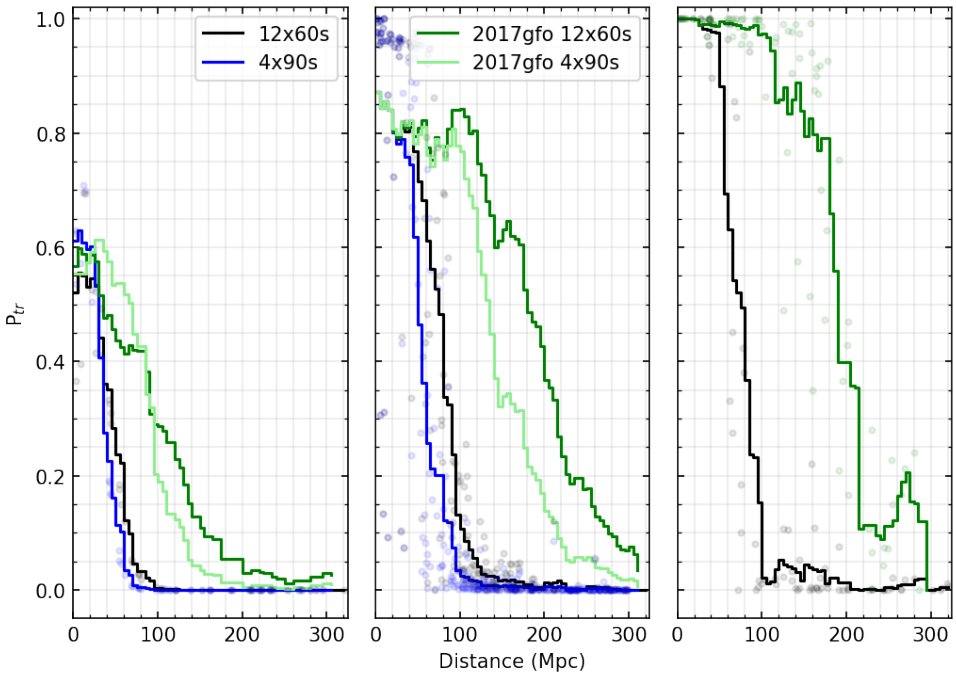

Fig 6 GOTO P_{tr} recovery over distance for (left to right) one, two and three detector scenarios. Shown here is a rolling average (lines) of individual simulations (points) using 25 Mpc bins in 5 Mpc intervals to minimise fluctuations due to distance bin choices. Black lines show the results of the deeper 12×60 s strategy and blue the shallower, but faster 4×90 s strategy. As a comparison, we also show the affect of using a model light curve of KN 2017gfo instead of the 'average' KN light curve. For 3 detector events we do not simulate the faster strategy as the very small localisation areas (typically 0.1–10deg^2) are already covered by a single GOTO pointing and so, offers no advantages. The unusual peak at 280 Mpc is the result of skymaps containing pixels with closer distance estimates that deviate from the typical distance distribution, resulting in higher P_{tr} values than expected at larger distances.

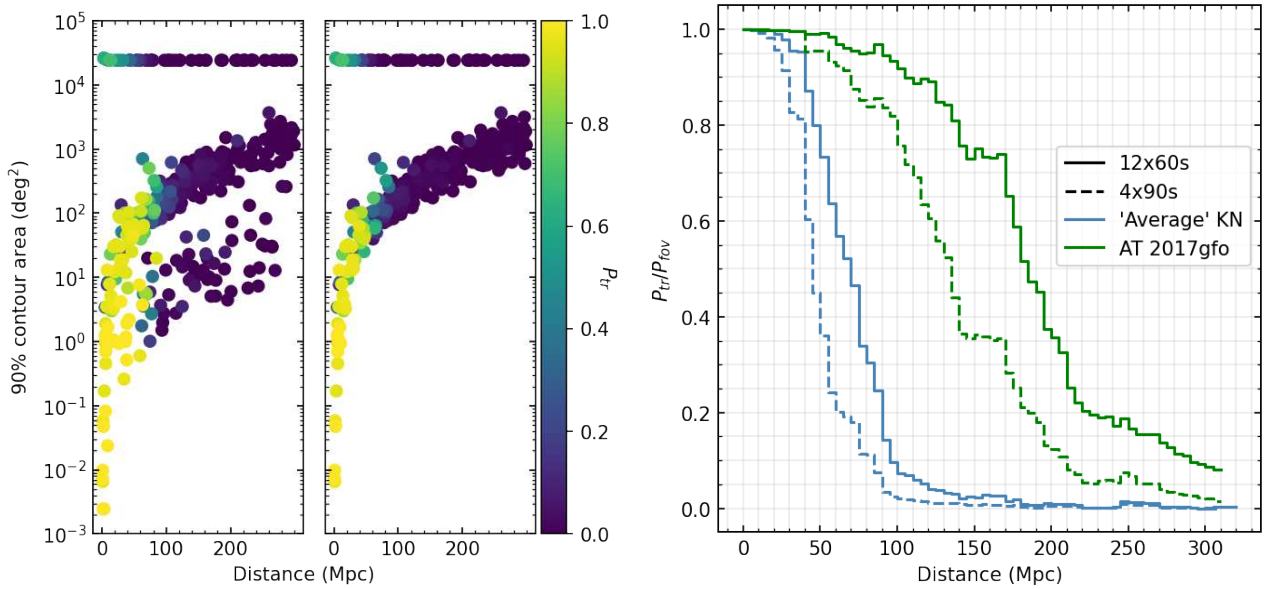

Fig 7 Left: Plot showing the size of the 90% credible area vs distance with P_{tr} mapped to the point colours for the deeper 12×60 s strategy and the shallower 4×90 s strategy. **Right:**Normalised P_{tr} values for each of the strategies showing only the effects of distance on P_{tr} recovery.

perform similarly up to ∼80 Mpc. We do not test the shallower strategy on three-detector events as their localisation is significantly less than the field of view of a single GOTO pointing, therefore it will not offer any increase in cadence or skymap coverage. In the three detector scenario, an unusual peak is visible at 280 Mpc. Typically GW skymap pixels have gaussian-like distance distributions. The simulated skymaps at times can display more irregular or varied distributions such as bimodal distributions. The P_{tr} peak is caused by skymaps with a strong 'closer' distance peak that results in higher P_{tr} values being recovered for distant skymaps. This is much less prominent in the 'average' light curve results as it is much fainter and is unlikely to be recovered even with the closer distances. From this analysis we can conclude that the shallower strategy is preferable for events ≤30 Mpc as it offers superior cadence for similar performance values while the deeper strategy should be employed for more distant events. To isolate the effects of GW distance/depth for each strategy, we can combine the results for all detector combinations for both strategies and normalise the average P_{tr} value using P_{fov} , which is shown in the right panel of Figure 7. Here we can more clearly see the true effective distance horizons for each strategy.

Given that P_{tr} represents the probability of detecting a KNe, we can convolve the curves in Figures 6 with the expected volumetric rate to obtain estimates for the total number of detected KNe over a given period.

Assuming the following: (1) localisations from the two-detector scenario (<1000 deg^2 out to 300 Mpc), (2) the BNS merger rate is 80–810 Gpc^{3}yr^{-1} as reported by R. Abbott (2021e),[25] (3) the more pessimistic 'average' KN scenario is representative of the KN population, and (4) there is no downtime for both GW interferometers and GOTO, we estimate that GOTO could recover 1 to 11 KNe per year using the 12x60s strategy. If KNe are more similar to AT 2017gfo, the recovery rate

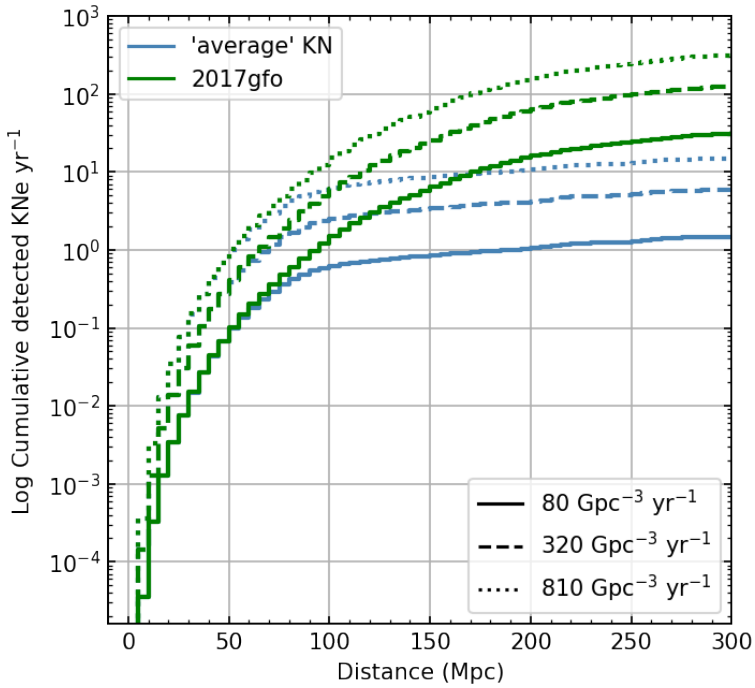

Fig 8 The estimated rates of detected KNe for GOTO assuming a AT 2017gfo-like light curve as well as the more 'pessimistic' average KN light curve using the 12x60s strategy.

increases by a factor of 10. Given the significant uncertainty in the inferred BNS merger rates, we plot the estimates and bounds provided in in Figure 8. This should be however, treated as a hard upper limit as both GW detectors as well as telescopes are taken offline regularly for maintenance. The GW detector BNS horizon also fluctuates depending on number and combination of GW detectors that are online. Telescopes also have to contend with observing time losses due to weather.

4.3 Results: VRO

The results for VRO are shown in Figure 9. It is unlikely VRO will trigger on GW events with large sky localisations ($\gtrsim$100s deg^2) due to the comparatively modest field of view and it not being core science of the observatory. We therefore limit the simulated events to two- and three-detector skymaps with localisation areas $\leq$150 deg^2 . The results show significant scatter in comparison to GOTO. This is almost solely driven by the skymap visibility from the VR, and we note here we are only showing results for events that have at least some part of their 95% localisation areas visible to VRO. Overall the results show that suffers little to no drop-off as a function of distance as VRO is able to reach the image sensitivity required for distant GW events, even for this faint 'average' KN.

It is possible to examine what the 'observed' light curve of the counterpart would have looked like during a given follow-up campaign at any imaged sky position. Using the *"processing_dfs = True"* flag, we can use the output *imout* csv file which contains the UNIQ skymap tile ids and the counterpart magnitudes contained within. By converting an RA and DEC to UNIQ tile id (see here[7]). one can

[7]https://rtd.igwn.org/projects/userguide/en/latest/tutorial/multiorder_skymaps.html

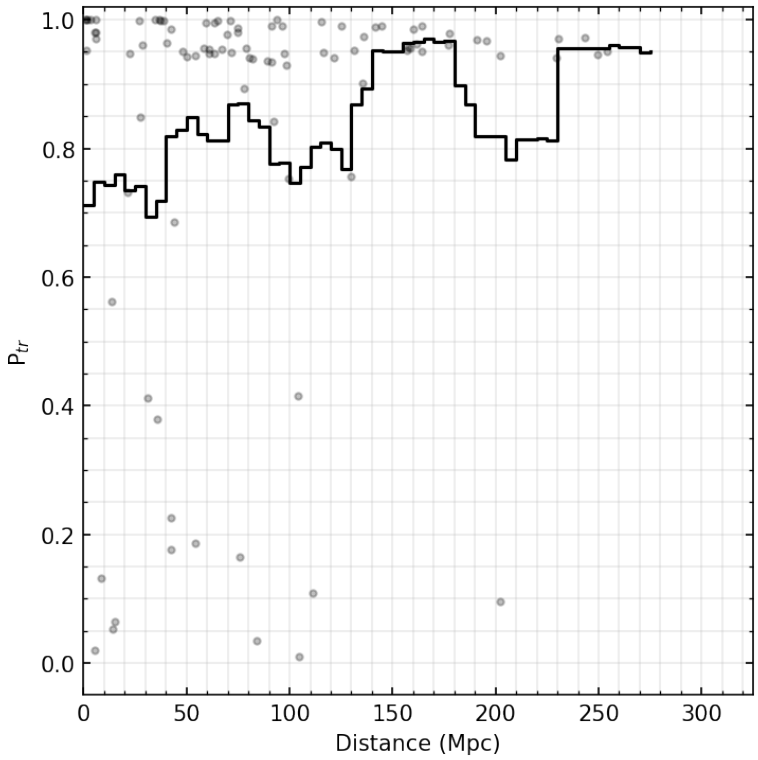

Fig 9 VRO follow-up P_{tr} recovery for a three-day follow-up versus source distance for two- and three-detector GW events with localisations $\leq$150 deg^2 . The pessimistic 'average' KN light curve model has been used.

filter the csv file and plot the respective 'phase' and 'mag' columns. An example is shown in figure 10 for both GOTO and VRO follow-up of the same skymap, at the same sky position.

4.4 VRO strategy

As mentioned VRO will likely perform more modest GW follow-up, involving a limited amount of time spent on a sub-set of opportune events, rather than continuously observing for nights uninterrupted post-trigger. We examine one of the strategies outlined in I. Andreoni et al. (2022),[26] specifically the 'minimal strategy': one coverage of the 90% localisation area with 30s exposures in $g + i$, followed by a second night of 180s exposures. Here we assume depths $m_g = 26.0$ and $m_i = 25.0$ using their figures 1 and 2. This strategy will be employed for skymaps with localisation areas between 20–100deg^2 . We select skymap simulation 301 with a localisation area of 59 deg^2 at a distance of 64 Mpc. The 'unrestrained' campaign resulted in equivalent P_{fov} and P_{tr} values of 94.7%, since almost the entirety of the visible probability volume was recovered (under which conditions P_{fov} = P_{tr}). We find that the minimal strategy outlined above returns almost identical P_{fov} , P_{vol} and P_{tr} values demonstrating it is more than capable of detecting a KN considerably more conservative survey strategy. As noted by I Andreoni et. al, the expected time investment using their *preferred* strategy for 20–100deg^2 localisations is 5–6 hours. Using GERRY we find that promising P_{tr} values ($\geq$90%) can be achieved for a skymap with localisation area 187deg^2 at a distance of $\sim$143 Mpc by using the *minimal* strategy above, even reduced to 30s exposures across both nights totalling 5.4 hours, potentially

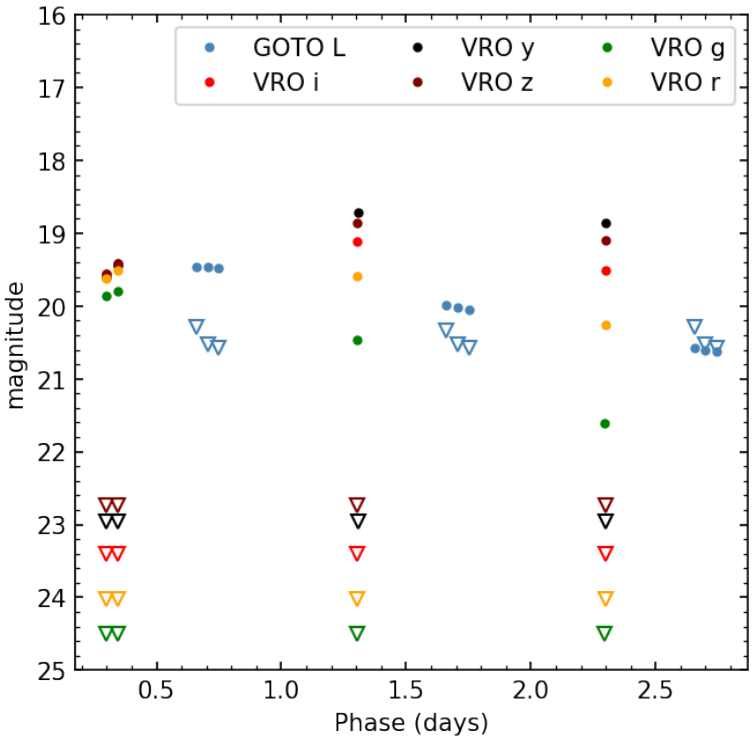

Fig 10 An example of the expected 'observed' light curves from the GW and follow-up simulations of a two detector GW event (446) at 63 Mpc assuming an 'average' KN. Also shown are the simulated image depths (open triangles).

increasing the number of trigger-able events. While a single night of 30s is capable of producing high P_{tr} values, having multiple epochs of detections to map rise/decline rates is important in separating a counterpart such as a KN from other, more common transients such as supernovae.

4.5 Combining community efforts in follow-up

GERRY is agnostic to the origin of the images it analyses. It can effectively and correctly combine image data from multiple follow-up campaigns into a single bulk analysis. As an example, shown in figure 11 is the follow-up of simulation 245, a BNS merger at 69 Mpc. Individually GOTO and VRO achieved P_{tr} values of 86.14% and 93.5% respectively. Combined, this value increased to 97%. GOTO and VRO are only two of an increasing number of wide-field survey instruments triggering on GW events. By folding in not only the efforts of GOTO and VRO, but also of ATLAS, BlackGEM, DECam, Pan-STARRS, ZTF, and others, GERRY can provide an quantitative and holistic description of the EM community's effort in follow-up up these rare event. The outputs can be used to perform a post-mortem and identify gaps in our combined capabilities and strategies. Such constraints, in the face of further non-detection follow-up efforts, will eventually allow for an assessment of EM model compatibility, helping to further our understanding of the nature and diversity of these events.

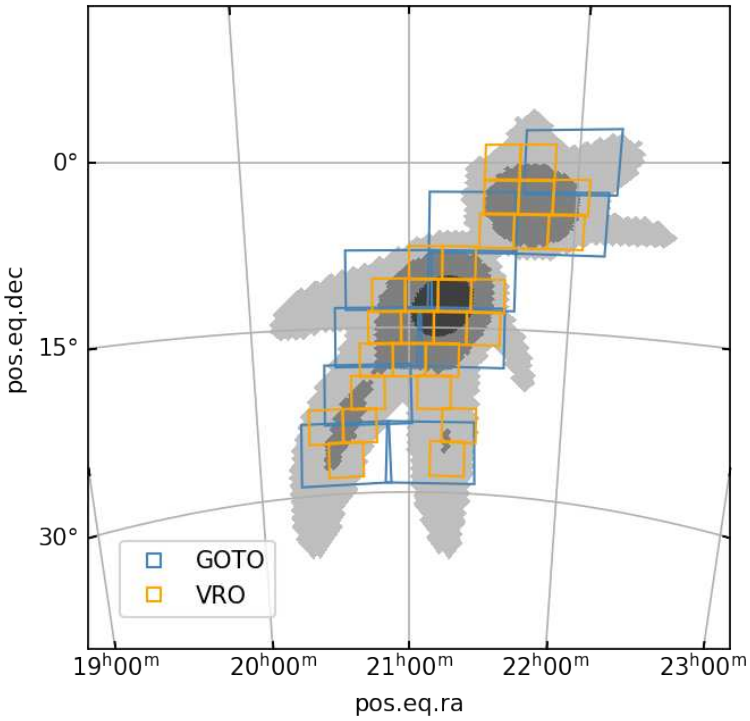

Fig 11 The footprints of the simulated GOTO and VRO observations of simulation 245.

5 Conclusion

Multi-messenger astronomy is still a relatively new and rapidly growing frontier in time domain astronomy. In the case of finding the EM counterparts to BNS mergers, many challenges remain despite increasing number and sophistication of our instrumentation. In this paper we have shown how GERRY can be used to help observers overcome some of these challenges. Starting from the initial steps of developing an observing strategy, we have shown how, through the use of GW and follow-up simulations, one can optimise observing strategy in key parameter spaces. We have also shown how the code can be used on real GW events to assess the follow-up campaign's performance, identify whether a counterpart was likely to be detected, as well as narrowing down the possible sky positions of the detected counterpart. GERRY is a simple but effective high-level product that can play a key part in informing both collaborations internally and the community at large about past or ongoing follow-up efforts. Particularly, it can inform on the effectiveness of our combined strategies if integrated in tandem with other public tools such as the GW Treasure Map.[27] By analysing the combined efforts of many surveys using GERRY, one can begin placing tighter constraints on counterpart emission than is possible from each individual survey's follow-up campaign, from which future models could be more securely anchored. The development of the code continues with the aims of improving and streamlining the user experience.

6 Availability

All of the GERRY results files and simulation data used to make the plots will be hosted on the Github. The public release of GERRY itself is planned via Github, and is available via reasonable request to the corresponding author.

Acknowledgments

DO and JL acknowledge support from a UK Research and Innovation Fellowship(MR/T020784/1). The Gravitational-wave Optical Transient Observer (GOTO) project acknowledges the support of the Monash– Warwick Alliance; the University of Warwick; Monash University; the University of Sheffield; the University of Leicester; Armagh Observatory & Planetarium; the National Astronomical Research Institute of Thailand (NARIT); the Instituto de Astrof ísica de Canarias (IAC); the University of Portsmouth; the University of Turku; the University of Manchester and the UK Science and Technology Facilities Council (grant numbers ST/T007184/1, ST/T003103/1, and ST/T000406/1). DO acknowledges The Seventh Workshop on Robotic Autonomous Observatories as part of the Revista Mexicana de Astronomia y Astrofisica Conference Series (2024), which a preliminary version of this work will appear.

References

1 B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, "Multi-messenger Observations of a Binary Neutron Star Merger," *ApJ* **848**, L12 (2017).

2 B. P. Gompertz, M. Nicholl, J. C. Smith, *et al.*, "A multimessenger model for neutron star-black hole mergers," *MNRAS* **526**, 4585–4598 (2023).

3 R. Margutti and R. Chornock, "First Multimessenger Observations of a Neutron Star Merger," *ARA&A* **59**, 155–202 (2021).

4 A. J. Levan, B. P. Gompertz, O. S. Salafia, *et al.*, "Heavy-element production in a compact object merger observed by JWST," *Nature* **626**, 737–741 (2024).

5 M. Nicholl, B. Margalit, P. Schmidt, *et al.*, "Tight multimessenger constraints on the neutron star equation of state from GW170817 and a forward model for kilonova light-curve synthesis," *MNRAS* **505**, 3016–3032 (2021).

6 K. Kawaguchi, M. Shibata, and M. Tanaka, "Diversity of Kilonova Light Curves," *ApJ* **889**, 171 (2020).

7 K. W. Smith, S. J. Smartt, D. R. Young, *et al.*, "Design and Operation of the ATLAS Transient Science Server," *PASP* **132**, 085002 (2020).

8 S. Bloemen, P. Groot, P. Woudt, *et al.*, "MeerLICHT and BlackGEM: custom-built telescopes to detect faint optical transients," in *Ground-based and Airborne Telescopes VI*, H. J. Hall, R. Gilmozzi, and H. K. Marshall, Eds., *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **9906**, 990664 (2016).

9 Dark Energy Survey Collaboration, T. Abbott, F. B. Abdalla, *et al.*, "The Dark Energy Survey: more than dark energy - an overview," *MNRAS* **460**, 1270–1299 (2016).

10 D. Steeghs, D. K. Galloway, K. Ackley, *et al.*, "The Gravitational-wave Optical Transient Observer (GOTO): prototype performance and prospects for transient science," *MNRAS* **511**, 2405–2422 (2022).

11 M. J. Dyer, K. Ackley, J. Lyman, *et al.*, "The Gravitational-wave Optical Transient Observer (GOTO)," in *Ground-based and Airborne Telescopes IX*, H. K. Marshall, J. Spyromilio, and T. Usuda, Eds., *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **12182**, 121821Y (2022).

12 K. C. Chambers, E. A. Magnier, N. Metcalfe, *et al.*, "The Pan-STARRS1 Surveys," *arXiv e-prints* , arXiv:1612.05560 (2016).

13 E. C. Bellm, S. R. Kulkarni, M. J. Graham, *et al.*, "The Zwicky Transient Facility: System Overview, Performance, and First Results," *PASP* **131**, 018002 (2019).

14 M. J. Graham, S. R. Kulkarni, E. C. Bellm, *et al.*, "The Zwicky Transient Facility: Science Objectives," *PASP* **131**, 078001 (2019).

15 Ž. Ivezić, S. M. Kahn, J. A. Tyson, *et al.*, "LSST: From Science Drivers to Reference Design and Anticipated Data Products," *ApJ* **873**, 111 (2019).

16 K. M. Górski, E. Hivon, A. J. Banday, *et al.*, "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere," *ApJ* **622**, 759–771 (2005).

17 P. Fernique, T. Boch, T. Donaldson, *et al.*, "MOC - HEALPix Multi-Order Coverage map Version 1.0." IVOA Recommendation 02 June 2014 (2014).

18 L. P. Singer, B. Parazin, M. W. Coughlin, *et al.*, "HEALPix Alchemy: Fast All-Sky Geometry and Image Arithmetic in a Relational Database for Multimessenger Astronomy Brokers," *AJ* **163**, 209 (2022).

19 G. Green, "dustmaps: A Python interface for maps of interstellar dust," *The Journal of Open Source Software* **3**, 695 (2018).

20 B. P. Gompertz, R. Cutter, D. Steeghs, *et al.*, "Searching for electromagnetic counterparts to gravitational-wave merger events with the prototype Gravitational-Wave Optical Transient Observer (GOTO-4)," *MNRAS* **497**, 726–738 (2020).

21 J. Guillochon, M. Nicholl, V. A. Villar, *et al.*, "MOSFiT: Modular Open Source Fitter for Transients," *ApJS* **236**, 6 (2018).

22 S. J. Smartt, M. Nicholl, S. Srivastav, *et al.*, "GW190425: Pan-STARRS and ATLAS coverage of the skymap and limits on optical emission associated with FRB 20190425A," *MNRAS* **528**, 2299–2307 (2024).

23 P. Landry and J. S. Read, "The Mass Distribution of Neutron Stars in Gravitational-wave Binaries," *ApJ* **921**, L25 (2021).

24 B. P. Gompertz, A. J. Levan, N. R. Tanvir, *et al.*, "The Diversity of Kilonova Emission in Short Gamma-Ray Bursts," *ApJ* **860**, 62 (2018).

25 R. Abbott, T. D. Abbott, S. Abraham, *et al.*, "Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog," *ApJ* **913**, L7 (2021).

26 I. Andreoni, R. Margutti, O. S. Salafia, *et al.*, "Target-of-opportunity Observations of Gravitational-wave Events with Vera C. Rubin Observatory," *ApJS* **260**, 18 (2022).

27 S. D. Wyatt, A. Tohuvavohu, I. Arcavi, *et al.*, "The Gravitational Wave Treasure Map: A Tool to Coordinate, Visualize, and Assess the Electromagnetic Follow-up of Gravitational-wave Events," *ApJ* **894**, 127 (2020).