Analyses of overlapping gravitational wave signals using hierarchical subtraction and joint parameter estimation

Justin Janquart,[1,2]★ Tomasz Baka,[1,2] Anuradha Samajdar [ORCID],[3] Tim Dietrich[3,4] and Chris Van Den Broeck[1,2]

[1]*Nikhef - National Institute for Subatomic Physics, Science Park, NL-1098 XG Amsterdam, The Netherlands*
[2]*Institute for Gravitational and Subatomic Physics (GRASP), Department of Physics, Utrecht University, Princetonplein 1, NL-3584 CC Utrecht, The Netherlands*
[3]*Institut für Physik und Astronomie, Universität Potsdam, Haus 28, Karl-Liebknecht-Street 24/25, D-14476 Potsdam, Germany*
[4]*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, D-14476 Potsdam, Germany*

ABSTRACT

In the coming years, third-generation detectors such as Einstein Telescope and Cosmic Explorer will enter the network of ground-based gravitational-wave detectors. Their current design predicts a significantly improved sensitivity band with a lower minimum frequency than existing detectors. This, combined with the increased arm length, leads to two major effects: the detection of more signals and the detection of longer signals. Both will result in a large number of overlapping signals. It has been shown that such overlapping signals can lead to biases in the recovered parameters, which would adversely affect the science extracted from the observed binary merger signals. In this work, we analyse overlapping binary black hole coalescences with two methods to analyse multisignal observations: *hierarchical subtraction* and *joint parameter estimation*. We find that these methods enable a reliable parameter extraction in most cases and that joint parameter estimation is usually more precise but comes with higher computational costs.

Key words: gravitational waves.

1 INTRODUCTION

The observation of gravitational waves (GWs) originating from compact binary coalescences (CBCs) is now done routinely with the advanced LIGO (Collaboration 2015) and advanced Virgo (Acernese et al. 2015) detectors, with tens of detections reported in the O3 GW catalogue (Abbott et al. 2021a). The observations from this new information channel have had major impacts in fundamental physics (Abbott et al. 2021b), astrophysics (Abbott et al. 2021c), and cosmology (Abbott et al. 2021d). Moreover, the possibility to detect electromagnetic counterparts for binary neutron stars (BNSs), such as GW170817 (Abbott et al. 2017b), has opened the perspective to do new multimessenger studies. Further upgrades to the second-generation detectors, as well as the addition of new detectors such as KAGRA (Somiya 2012; Aso et al. 2013; Akutsu et al. 2019, 2020) and LIGO India (Iyer et al. 2011), will lead to the observation of numerous events in the coming years. Moreover, going from these second-generation (2G) detectors to third-generation (3G) detectors, such as Einstein Telescope (ET; Punturo et al. 2010; Hild et al. 2011) and Cosmic Explorer (CE; Reitze et al. 2019; Abbott et al. 2017a; Regimbau et al. 2017), will lead to a major jump in the detection rate as well as in the duration of the signal due to the combined global increase in sensitivity and the major enhancements for lower frequencies (Sathyaprakash et al. 2012). In turn, this will lead to a high probability of CBC signals overlapping in the 3G

detectors (Regimbau & Hughes 2009; Himemoto, Nishizawa & Taruya 2021; Relton & Raymond 2021; Samajdar et al. 2021; Pizzati et al. 2022). While the exact number of overlaps will depend on the merger rate density distribution for compact binary mergers, it is possible to have an idea of how often they will occur. Based on a catalogue of simulated mergers, Samajdar et al. (2021) find that between 750 and 20 000 s in a year will have at least two mergers happening, depending on the local merger rate. Using a Poisson detection rate and predicted merger rate densities, Pizzati et al. (2022) find that there will be between 40 and 300, 200 and 1500, and 170 and 5100 one-second chunks over the year in which at least two binary black hole (BBH) merger, two BNS mergers, one BBH and one BNS merger happen, respectively.

Previous works have studied the impact of such overlapping signals on data analysis when recovering one of the two signals and neglecting the presence of the other (Antonelli, Burke & Gair 2021; Himemoto et al. 2021; Relton & Raymond 2021; Samajdar et al. 2021; Pizzati et al. 2022). These works employed different techniques but all have the same basic conclusions: bias can occur in various scenarios, and is most likely when the signals merge close to each other. In Pizzati et al. (2022), using a Fisher matrix approach, the authors show that BNSs are less correlated in overlapping signals so that their bias becomes important only for close merger times (<0.1 s), while the correlation between BBHs is more important, meaning that the bias can happen for larger differences between the merger times. They then proceed to perform parameter estimation (PE) for overlapping BBHs in a LIGO–Virgo network varying some of the parameters, showing the appearance of biases for merger times

★ E-mail: j.janquart@uu.nl

close to each other. In Himemoto et al. (2021), the authors use Fisher matrices to study the bias that can occur in the parameters in both overlapping BBHs and BNSs, also finding that the bias becomes more important for short differences in the time of arrival. In Relton & Raymond (2021), the authors focus on a LIGO Voyager scenario, looking at the biases not only based on their difference in merger times, but also as a function of other parameters, such as the sky location. For overlapping BBHs, they find a more important bias for closer merger times. However, they show that the observed bias for a given difference in merger time can change substantially depending on the sky location of the two events. Moreover, two overlapping BBHs can be mistaken for one strongly precessing BBH. Furthermore, the authors suggest that for overlapping BNS and BBH signals no major bias will occur due to the different durations of the two signals. This is corroborated by the analyses done in Samajdar et al. (2021), where three scenarios are analysed: two overlapping BBH signals, two overlapping BNS signals, and the overlap of a BBH of varying masses with a BNS. In the latter case, the authors find that there is hardly any effect on the BNS PE, probably due to the difference in the number of cycles present in-band for this signal. However, in this scenario, the BBH can be affected by significant bias, especially when the BBH has high component masses. The bias mostly disappears when the merger times are separated by more than 2 s. For two overlapping BBHs, if the total masses and hence the durations are different, the PE is done relatively well. However, others (Himemoto et al. 2021; Relton & Raymond 2021; Pizzati et al. 2022) have shown that, if the two BBHs have similar source properties, biases can be present. Finally, for two overlapping BNSs, it appears that the signal with the highest signal-to-noise (SNR) is relatively well recovered in all cases, but not necessarily the quieter signal.

One problem that is not covered by previous studies is the effect of confusion noise on PE. Indeed, the high rate of the events and their duration will make for very few periods without signal in-band for the 3G detectors (Samajdar et al. 2021). As a consequence, it will be very difficult to estimate the noise present in the detectors, and additional biases can occur due to a mismodelling of the noise (Antonelli et al. 2021).[1] A demonstration of the effect of this noise on matched filtering and how the power-spectral density (PSD) could be computed are presented in Wu & Nitz (2022).

The presence of biases when signals merge close to each other and the relative occurrence of such scenarios based on the estimated rates shows that PE methods will have to be adapted to be suited for the 3G cases. Indeed, biases in the parameters estimated for the CBCs would impact any direct science case for the CBCs (such as measuring their mass distribution and rate, or testing general relativity; Hu & Veitch 2023) and also some indirectly related ones, such as the search for primordial backgrounds since this requires the subtraction of the foreground sources (Biscoveanu et al. 2020; Sachdev, Regimbau & Sathyaprakash 2020; Sharma & Harms 2020; Reali et al. 2022; Zhou et al. 2022a, b)

In this work, we look at two possible methods to analyse overlapping signals. One is *hierarchical subtraction* (HS), where we analyse one signal (typically the loudest), then subtract the maximum likelihood template before analysing the second one. However, if an important bias happens when analysing the first signal, the parameters of both events might be biased.[2] Optionally, one can also

perform a third run, subtracting the maximum likelihood parameters for the second event and re-analysing the dominant signal to reduce the bias in its recovery. Still, this is not guaranteed to lead to unbiased results. Therefore, we also implement a *joint parameter estimation* (JPE) framework, where the two signals are analysed at the same time to account for the entire model. In principle, this should be the most complete model one can use. Due to the high dimensionality of the parameter space, combined with the large duration of the signals in the 3G detectors, this framework is substantially slower than HS.[3] It would be close to impossible to follow the predicted rates for an ET and CE network using a simple JPE like the one used in this work. Such constraints could be alleviated by using recently developed techniques, like relative binning (Dai, Venumadhav & Zackay 2018; Zackay, Dai & Venumadhav 2018; Leslie, Dai & Pratten 2021), adaptive frequency resolution (Morisaki 2021), or machine learning (Dax et al. 2021; Williams, Veitch & Messenger 2021; Langendorff et al. 2022). One could also count on the development of more powerful computational methods, such as quantum computing (Gao et al. 2022) but it is difficult to have an idea of the state of such methods by the time the 3G detectors get online. Still, it is important to start preparing for the future of 3G detectors now, hence, it is important to start looking at the PE of overlapping signals to have the bases to build upon. Due to the limited computational resources, this work focuses on the PE of two overlapping BBH signals.

This article is structured as follows: In Section 2, we explain the two methods applied to perform the analysis of the overlapping signals. In the next section, we explain the setup of the analyses, while in Section 4 we present the results of the analyses. Finally, Section 5 provides conclusions and outlook.

2 DESCRIPTION OF THE METHODS

When performing GW data analysis on CBC signals, our objective is to find the posterior probability density function of the binaries' parameters (θ): $p(\theta|d, \mathcal{H}_s)$, where d represents the data, and $\mathcal{H}_s$ is the hypothesis under which we work (e.g. there is a GW signal in the data). Using Bayes' theorem (Veitch & Vecchio 2010):

$$p(\theta|d, \mathcal{H}_s) = \frac{p(d|\theta, \mathcal{H}_s)p(\theta|\mathcal{H}_s)}{p(d|\mathcal{H}_s)}, \tag{1}$$

where $p(\theta|\mathcal{H}_s)$ is the prior on the binary parameters, and $p(d|\theta, \mathcal{H}_s)$ is the likelihood for the data given a set of parameters. Finally, $p(d|\mathcal{H}_s)$ is called the evidence and represents a normalization factor for the posterior probability density:

$$p(d|\mathcal{H}_s) = \int d\theta \, p(d|\theta, \mathcal{H}_s)p(\theta|\mathcal{H}_s) \,. \tag{2}$$

For GW inference, $d(t)$ is the output of the interferometers, which can be seen as made of a noise component $n(t)$ and, under the signal hypothesis, a GW component $h(t)$:

$$d(t) = n(t) + h(t) \,. \tag{3}$$

---

[1] We will not consider this in this work as we will only look at overlapping binary signals.

[2] In this work, the term 'bias' is used colloquially and denotes any changes in the recovered posteriors due to the presence of another signal.

[3] For our experiments, the two frameworks were run on the same cluster using 16 *Intel(R) Xeon(R) Gold 6152* CPUs. The average run time for JPE is 23.8 d, while for HS, the first run took an average of 6.3 d, the second run an average of 4.3 d, and the last run took an average run time of 6.1 d. So, on average, JPE takes seven more days to complete than HS if we perform the three runs. If one is satisfied with the two first runs, the difference between the two approaches goes up to about two weeks.

In our scenario, the GW component can consist of one or more signals. In the latter case, $h(t) = \sum_{i=1}^{N} h_i(t)$, where $h_i(t)$ is the representation of each individual GW signal, and N is the total number of GW signals present in the data stretch under consideration.

Assuming the noise to be Gaussian, the likelihood of having the data $d(t)$ given the presence of a GW component $h(t)$, is given by the proportionality

$$p(d|\theta, \mathcal{H}_s) \propto \exp\left[-\frac{1}{2}\langle d - h(\theta)|d - h(\theta)\rangle \right], \qquad (4)$$

where

$$\langle a|b\rangle = 4\Re \int_{f_{\min}}^{f_{\max}} \frac{\tilde{a}^*(f)b(f)}{S_n(f)} df . \qquad (5)$$

In this expression, $\tilde{a}(f)$ refers to the Fourier transform of the function $a(t)$, * is the complex conjugate, and $S_n(f)$ is the power spectral density (PSD). $f_{\min}$ and $f_{\max}$ represent, respectively, the lower and the upper frequency chosen for the analysis.[4]

2.1 Joint parameter estimation

When the noise component in equation (3) is made of multiple signals, the likelihood in equation (4) becomes

$$p(d|\theta, \mathcal{H}_s) \propto \exp\left[-\frac{1}{2}\left\langle d - \sum_{i=1}^{N} h(\theta_i)\middle| d - \sum_{i=1}^{N} h(\theta_i)\right\rangle \right], \qquad (6)$$

where we just expanded the expression for $h(t)$ compared to the previous expression. Here, N is the total number of signals, and θ_i represents the set of parameters describing the ith GW signal so that $\theta = \{\theta_1, \ldots, \theta_i, \ldots, \theta_N\}$ in this case.

Using equation (6) gives rise to the method of JPE, where we jointly look for $\{\theta_i\}_{i=1,\ldots,N}$. Since each θ_i is a set of 15 parameters for spinning BBHs, a set of 16 parameters for neutron-star black hole (NSBHs), and a set of 17 parameters for BNSs,[5] it means that, if we have X BBHs, Y NSBHs, and Z BNSs (so $N = X + Y + Z$), the parameter space has $15X + 16Y + 17Z$ parameters to explore. In the end, this means that as soon as we consider two signals, the parameter space grows to at least 30 dimensions, which is already challenging with our traditional methods, showing the difficulty to analyse several signals jointly.

In addition, we can assume that there is some uncertainty on the total number of signals N in the data. In this case, it is also possible to sample over N and the signal types, effectively allowing for any number of signals to be present in a given data stretch.

The problem of joint analysis of several signals has already been looked at in other contexts, such as the characterization of the nearly monochromatic signals coming from white dwarf binaries in LISA (Littenberg et al. 2020), BNSs in the Big Bang Observer (Cutler & Harms 2006), or supermassive black holes in

pulsar timing array searches (Petiteau et al. 2013). However, the different characteristics of the signal looked for in these various context makes the methods different from one case to the other.

In this work, we will only consider the possibility to have two signals in the data. So, we write the data as

$$d(A, B, t) = h_A(t) + h_B(t) + n(t), \qquad (7)$$

where we just denote the signals by A and B, without any importance on which signal is A and which signal is B.

In this case, the likelihood takes the particular form

$$p(d(A, B)|\theta, \mathcal{H}_s) \propto \exp\left[-\frac{1}{2}\left\langle d(A, B) - h_A(\theta_A) - h_B(\theta_B)\middle| \right.\right.$$
$$\left.\left. d(A, B) - h_A(\theta_A) - h_B(\theta_B)\right\rangle \right]. \qquad (8)$$

This is just the explicit form of equation (6) for two signals.

In principle, if the sources are of the same nature, the labels A and B are interchangeable during the sampling, making the likelihood symmetric in two events. In our algorithm, we do not impose any conditions on the parameters to break this symmetry. As a consequence, this is something that needs to be done in a post-processing step, as we need to assign drawn samples to the correct event. Sometimes, not accounting for this condition leads to bimodalities. In this work, we use a time ordering condition, taking the samples for event A to be those that arrive first in time, and the samples for event B to be those arriving second in time. In future work, this condition could directly be imposed in the algorithm by having a conditional prior such that the time of arrival of one event is always smaller than the other. We also note that the condition can be imposed on other parameters, such as the chirp mass for example.

2.2 Hierarchical subtraction

In HS, the idea is not to fit the two signals at once but to rather do a combination of individual signal analyses and subtraction of best-fitting parameters. Therefore, we start by running a single event PE analysis on the data $d(A, B)$ to get the characteristics of the dominant signal. If we label by A the loudest signal, we can denote the best-fitting parameters (typically the maximum likelihood parameters) $\hat{\theta}_A$, and the waveform corresponding to this signal

$$\hat{h}_A(t) = h(t, \hat{\theta}_A). \qquad (9)$$

Using this, we can get the data for signal B given signal A by subtracting the best-fitting template

$$d(B, r_A, t) = d(A, B, t) - \hat{h}_A(t), \qquad (10)$$

where r_A are the residuals of signal A due to the imperfect subtraction. This is an approximate data strain for the second event in the data since the maximum likelihood parameters used to model the first event are prone to errors, with errors coming from the modelling itself but also from the presence of the second event when characterizing the first one.

We can then analyse $d(B, r_A, t)$ to get the parameters for event B, leading to a posterior distribution for the two events. In principle, if the bias on the first recovery is not too important, then the posteriors on the second event should be fine. However, this approach is less robust than JPE, where we correctly account for the presence of several events.

[4]Usually, $f_{\min}$ is chosen to be the lowest frequency at which the noise is approximately stationary, and $f_{\max}$ is the Nyquist frequency.
[5]Typically, a BBH is described by two mass parameters, six spin parameters, a distance parameter, the inclination, two parameters for the sky location, the merger time, the phase of coalescence, and the polarization angle. Usually, for each neutron star present in the system, one also adds a tidal deformability, however, for BNSs the dimensionality could increase even further if higher-order tidal contributions, spin-induced quadrupole effects, or resonant effects are also taken into account.

In addition, formally, when parameter estimation is performed, one can also add the noise-related calibration parameters, which would further increase the dimensionality of the parameter space.

In Antonelli et al. (2021), the authors also suggest a way to correct the bias due to the individual characterization of two signals.[6] Once the two signals have been analysed separately, we can use the two best-fitting posteriors to evaluate the bias made in the model reconstruction for each signal. The estimated biases can then be applied as a correction factor to the best-fitting parameters. We could then redo the subtraction of each event and analyse it again, but now with a subtracted signal that is closer to the real one, reducing the possible bias in the recovered posterior. Though this method is attractive, it requires multiple PE runs, which are expensive in a 3G detector context, and the computation of the biases also requires solving a combination of Fisher matrices and numerical derivatives, making it a non-trivial operation.

3 SETUP OF THE ANALYSES

Due to the computational resources that would be required to analyse 3G signals, we focus on overlapping BBHs with masses in [30, 60] $M_\odot$.[7] We use a network of detectors made of one triangular ET with 10 km arm-lengths, and a CE detector located at the LIGO-Hanford position and with 40 km arm-lengths. We generate stationary Gaussian noise from the detectors' PSDs, where, for ET, we use the ET-D PSD (Punturo et al. 2010; Hild et al. 2011), and for CE, we use the projected PSD from Abbott et al. (2017a) and Reitze et al. (2019). We then inject two simulated BBH signals into the artificial noise. For this study, to limit the computational burden, we take a lower cutoff on the signals of 20 Hz. A representation of the waveforms obtained by the addition of two BBHs can be found in Fig. 1. One sees that the final signal has a non-trivial shape, illustrating the risk of biased posteriors when not accounting for the presence of two signals. In addition, one also sees that depending on the relative SNR of the signals, the observed deformation in the waveform is different. For the case where one signal has a significantly higher SNR than the other ($\sim \times 2$, top panel), the quieter signal will somewhat deform the signal, but the observed waveform will resemble mostly the loudest signal. On the other hand, for signals with close SNRs (bottom panel), we see that the deformation of the signal can be more complicated, without really having a dominant signal (except for the fraction of a second where one signal has merged and the other is still merging). Based on these observations, one can expect HS to be more effective when one signal clearly dominates over the other.

In our study, we simulate 55 such mergers. To produce these high-mass signals, we sample the component masses from the Power law + Peak distribution from Abbott et al. (2021a) but keep only the systems that fulfill the mass requirement. In addition, the events are sampled in redshift according to the merger rate density reconstructed from Oguri's fit (Oguri 2018). The sky location is drawn to be uniform on the sky, and the spin parameters are picked from an isotropic distribution. For overlapping signal events, the coalescence time of the first event is drawn from a uniform distribution spanning over an entire year, while the second event is placed 0.1 s later. An overview of the functions used to make the binaries and the priors used for the analysis is given in Table 1.

Since the SNR of the signals can reach hundreds to thousands in an ET and CE network, and such high values make the computation time
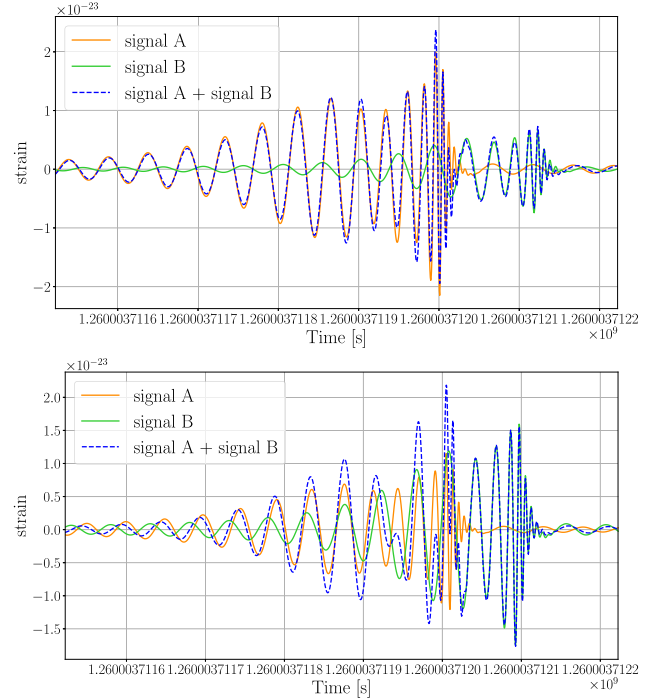

Figure 1. Top panel: Representation of two overlapping BBHs and the underlying signals for signals with different SNRs (SNR$_A$ = 46.1 and SNR$_B$ = 22.2). In this case, the sum of the two signals is mostly dominated by the loudest signal, while the effect of the quieter signal takes up only after the loudest signal has merged. Bottom panel: Representation of two overlapping BBHs and the underlying signals for signals with similar SNRs (SNR$_A$ = 34.3 and SNR$_B$ = 30.2). In this scenario, there is no signal dominating over the other for the entire duration of the event and the overlapping signals have more complicated features.

even longer, we decided to rescale the SNR to take values constrained between 8 and 50. This is done by adjusting the luminosity distance. However, since we expect the SNR ratio between the events to play a role, we try to keep this ratio as close as possible to the original one. So, if the loudest signal has a value above 50, we rescale it to take a value between 45 and 50 (this value is drawn randomly from a uniform distribution). Then, we rescale the quieter signal with the same factor. If this value is below 8, then we choose a new scaling factor to bring the SNR back between 8 and 13 (once more using a uniform distribution). Each system is then analysed once without additional noise, and once injected in Gaussian noise generated from the PSDs.

For the different runs, we choose fixed priors for the various parameters. The right column of Table 1 gives an overview of the priors used for the different parameters. In particular, we take a uniform prior on chirp mass ($\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}$) and mass ratio ($q = \frac{m_2}{m_1}$), with bounds of [4, 200] and [0.1, 1] $M_\odot$. We also take a uniform in comoving volume prior for the luminosity distances, with bounds going from 1 to 100 Gpc. These priors are adapted to cover any possible signal present in our set of data. The priors for the other parameters correspond to the usual priors taken for BBHs. When doing the JPE and HS runs, the priors are the same for the two events, and we do not add any conditions related to the signals (for example time ordering of the signals or enforcing one signal to be heavier than the other).

---

[6]In their paper, they also account for the possible confusion background due to the sum of all the mergers going on in the background.

[7]The higher masses are chosen to not have a signal with a too long duration while still enabling overlap for the difference in times of arrival used in this work.

Table 1. Overview of the functions used to generate the different parameters for the BBH population and the priors used for the PE recovery.

Parameter	Population generation	Prior
Primary component mass	Power law + Peak (Abbott et al. 2021a)	–
Mass ratio	Power law + Peak (Abbott et al. 2021a)	$\mathcal{U}(0.1, 1)$
Chirp mass	–	$\mathcal{U}(4, 200)$
Redshift	Oguri's fit (Oguri 2018) + rescaling	–
Luminosity distance	–	Uniform comoving volume [1, 100] Gpc
Spin amplitude 1	$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$
Spin amplitude 2	$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$
Tilt angle 1	Uniform in sine	Uniform in sine
Tilt angle 2	Uniform in sine	Uniform in sine
Spin vector azimuthal angle	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Precession angle about angular momentum	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Inclination angle	Uniform in sine	Uniform in sine
Wave polarization	$\mathcal{U}(0, \pi)$	$\mathcal{U}(0, \pi)$
Phase of coalescence	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Right ascension	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Declination	Uniform in cosine	Uniform in cosine
Time of coalescence	Uniform over a year (second precision)	$\mathcal{U}(t_{inj} - 0.1, t_{inj} + 0.1)$

We note here that an alternative approach is to use narrower priors informed by the results of low-latency searches (Regimbau et al. 2012; Relton et al. 2022), which very likely could only be applied for a couple of parameters such as the chirp mass and the coalescence time. In Relton et al. (2022), the authors show that matched filtering pipelines and unmodelled searches can pick up overlapping signals with reasonable accuracy. They also suggest some enhancements to make the pipelines even more suited for the challenge of overlapping signals detection. In addition, they show that for most of the overlapping signals, the error on the chirp mass is not much larger than for the non-overlapping case, even if for the occurrences where the signals are very close in time, the error increases. Therefore, using such searches to set narrower priors is a realistic alternative. However, currently, they also seem to contain risks, as an increased difference in the value recovered for some parameters can happen, and taking too narrow a prior could lead to the exclusion of the actual value from the prior. In the end, extra developments would be needed to make sure that using these results to narrow down the initial priors is viable.

To have a basis of comparison, we also do the PE of the individual signals. This is done by using the same priors as the one explained above but injecting only one of the two signals in the noise.

All the PE runs are performed using BILBY (Ashton et al. 2019) with the DYNESTY (Speagle 2020) sampler. For the JPE runs, we added our own adapted joint likelihood in the package to keep a consistent framework.

4 RESULTS AND DISCUSSIONS

In this section, we show the results of the different approaches. We first compare the HS approach with single parameter estimation (SPE). In the latter, we only inject one of the two signals and perform PE on it. Then, we compare JPE with SPE and HS. Here, we focus on the results of the analyses performed with noise. The conclusion in the no-noise case are similar and can be found in Appendix A. For all the figures presented in this work, when plotting individual event results, we represent by a dot and label as 'loud' the loudest events in the pair, while we represent by a triangle and label as 'quiet' the quieter ones. We note that all our discussions and conclusions are based on our pool of overlapping signals, and should therefore be understood as valid within our experimental setup. More extended

studies are required to fully understand the behaviour of the different frameworks in various contexts.

4.1 Hierachical subtraction

We start with discussing the result of the analysis for the HS approach. In almost all of the HS runs, the first PE stage picked up the signal with higher SNR. Nevertheless, there were two cases where it instead picked up the quieter signal; these were instances where the SNRs of the injected signals were close to each other. In such a case, the signal picked first is not the same with and without noise. A representative example of the posteriors can be found in Fig. 2. While the widths of the distributions closely match the ones with the SPE runs, in most cases they show bias in the recovered parameters.

When comparing the HS and the SPE recovery of the parameters, one sees that the HS recovery is nearly always biased. However, this bias seems to be more pronounced when there is also noise in the injection. For example, in Fig. 2, one sees that the zero-noise HS recovery is close to the one for SPE with noise, while the recovery with noise is off. This shows that the mismodelling of the noise (due to an additional event) is present, as one could expect from previous works on biases in overlapping signals (Antonelli et al. 2021; Relton & Raymond 2021; Samajdar et al. 2021; Pizzati et al. 2022). However, this does not necessarily mean that all parameters are off, and the first signal's characteristics can be recovered.

In Fig. 3, we can see mismatches between the injected and recovered waveforms, comparing HS and SPE cases, with and without noise. The mismatch is defined as $1 - \bar{M}$, where $\bar{M}$ is the match between the waveforms, defined as

$$\bar{M} = \frac{\langle h_1|h_2\rangle}{\sqrt{\langle h_1|h_1\rangle \langle h_2|h_2\rangle}}, \tag{11}$$

where $\langle .|. \rangle$ is the noise weighted product defined in equation (5). The mismatch represents the dissimilarity between two waveforms. High values mean a major disagreement between the two waveforms, and smaller values mean that the waveforms agree well. In Fig. 3, we see that the average mismatch throughout the detections is always low, below 0.02. As expected, the presence of another signal leads to worse waveform recovery for HS compared to SPE (most points are below the diagonal). The worst recovery of the signals occurs for overlapping signals with similar SNRs. Note that the zero-noise
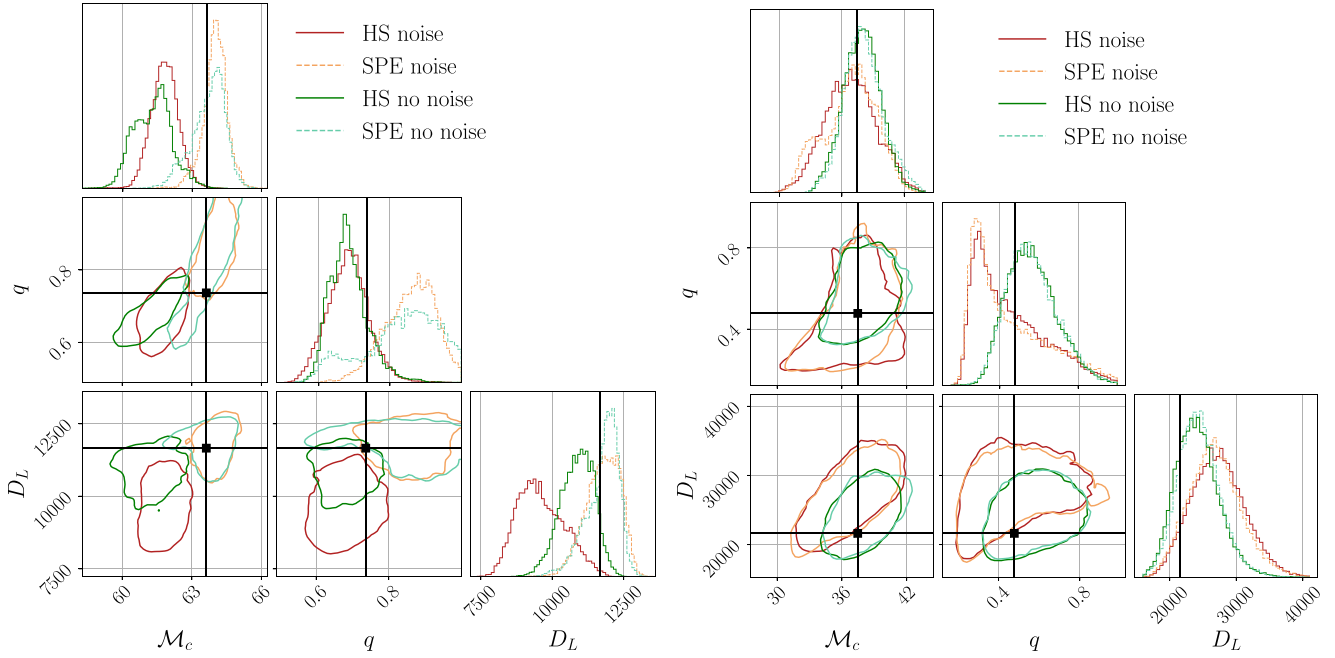
Figure 2. Comparison of the posteriors for the HS and the SPE methods with and without noise. We represent the chirp mass ($\mathcal{M}_c$), the mass ratio (q), and the luminosity distance (D_L). Left-hand panel: Case where HS is worse than SPE. HS posteriors are significantly shifted compared both to the injected value and SPE posteriors. Right-hand panel: Case where HS is close to SPE, with equivalent posteriors recovered in the two cases.
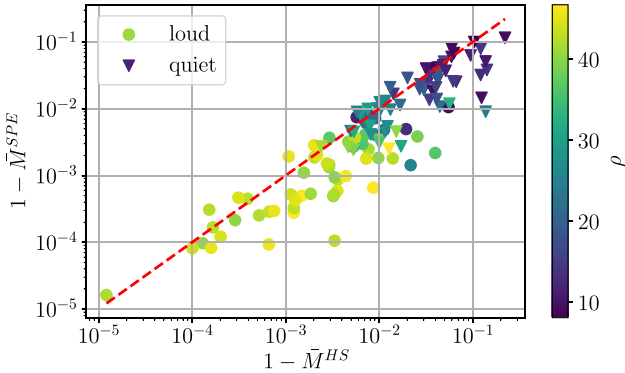

Figure 3. Mismatch for the SPE analysis versus the HS analysis, with noise. The red dashed line represents the diagonal where the mismatch is the same. HS is worse at signal recovery, as most of the points fall below the diagonal.

Figure 4. Representation of the offset of the recovered parameters, with noise for HS and SPE. Plotted is the difference between injected value and the median of the recovered value, normalized by the injected value. HS tends to give higher deviations than SPE.

case (see Appendix A) shows a clearer difference between SPE and HS recoveries. It is expected to have a larger difference in this case because the effect of the unmodelled signal is stronger when there is no noise since it is the only source of uncertainty in the signal.

By looking at the difference between the median of the recovered posterior and the injected value normalized by the injected value, we can quantify the offset in the recovery ($\Delta_{\mathcal{M}_c} = \frac{|\mathcal{M}_c^{\text{inj}} - \mathcal{M}_c^{\text{rec}}|}{\mathcal{M}_c^{\text{inj}}}$, where 'rec' stands for recovered and 'inj' stands for injected.). This is represented in Fig. 4 for the chirp mass recovery with noise and is representative of all the parameters. HS shows a higher offset in 71 per cent of the louder events and 51 per cent of the quieter ones. As expected, the deviation is larger for HS compared to SPE. For the first recovered signal, since we have an unmodelled signal present in the detectors, the noise properties are not properly modelled. Therefore, a larger offset is generally
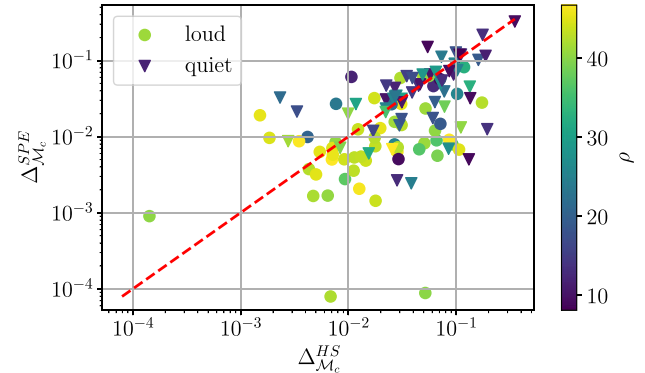
observed for the louder signal compared to the secondary one. However, when strong deviations are present for the first signal, it can reverberate on the second, also leading to worse recoveries for this event.

Finally, Fig. 5 shows how the widths of the 90 per cent confidence intervals of the posteriors for HS compare to SPE, normalized by the injected value ($\delta_{\mathcal{M}_c} = \frac{\sigma_{\mathcal{M}_c}}{\mathcal{M}_c^{\text{inj}}}$, where $\sigma_{\mathcal{M}_c}$ represents the width of the 90 per cent confidence interval). We observe that the widths of the distribution are consistent between the two, even though the recovery is biased.

It is of interest to compare how doing successive steps of PE affects the results. After reconstructing the quieter of the two signals, we subtract it from the initial data and do PE again. In principle, it should
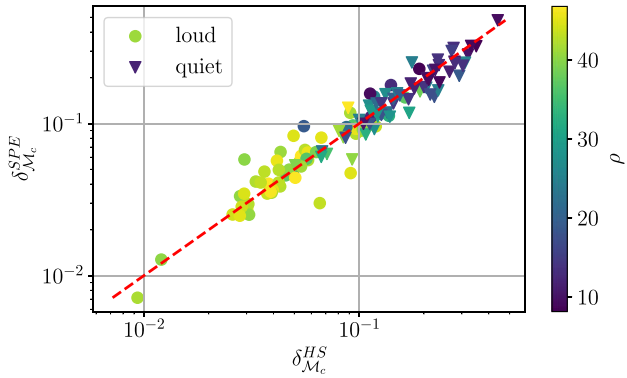
Figure 5. Comparison of the normalized width of the 90 per cent confidence interval for the chirp mass for the HS recovery and the SPE recovery. The width of the recovered distribution is largely unaffected by the presence of another signal.
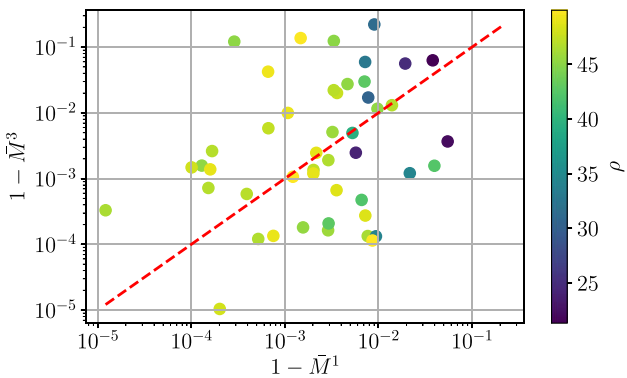

Figure 6. Comparison of the mismatch for the recovered waveform between the first (horizontal axis) and third (vertical axis) steps of HS. The mismatch after the third step is lowered in 62 per cent of the cases.

Figure 7. Comparison of the offset in recovered chirp mass between the 1st (horizontal axis) and the 3rd (vertical axis) HS steps. The recovered bias after the 3rd step is lowered in 62 per cent of the cases.

4.2 Joint parameter estimation

In what follows, we will discuss the results obtained from the analysis of the overlapping signals using JPE. We focus on the results obtained after having done the time ordering of the samples described in Section 2, as these are the samples where effectively one set of posteriors is matched with one signal, and the other set is matched with the other signal.

For the JPE recovery, more diverse scenarios are possible. We show three main cases in Fig. 8: one where the recovery is equivalent to the one from SPE, one where JPE has smaller bounds on the posteriors for one or more parameters, and one where JPE has trouble fitting the signal properly and biases can occur. While more in-depth studies are required to fully comprehend this behaviour, it could be originating from the mixed term of the two signals present in the likelihood when it was modified to account for multiple signals. In some cases, the narrower posteriors could be offset compared to the injected value, pushing the latter out of the 90 per cent confidence interval.

For all the events, we compare the mismatch between the maximum likelihood waveform for the event in the JPE scenario with the maximum likelihood for the event in the SPE and HS cases. This is represented in Fig. 9 for the noise cases. Independent of the presence of noise, we find that the mismatch is smaller for JPE than for HS but larger than for SPE. This is what one would expect since JPE accounts for the presence of the two signals and so should lead to smaller biases. However, fitting simultaneously two signals is more complex than analysing a single signal. Therefore, the SPE measures remain a better representation of the injected signal.

As before, we present the normalized distance between injection and recovery in Fig. 10 for the noise case. When comparing this offset for the JPE case against the SPE case, we find that 45 per cent of the events have a lower offset for JPE than for SPE. On the other hand, when we compare with HS, we find 65 per cent of the events with a lower offset for JPE. This confirms that JPE is better than HS to find the injected signals (when two signals are present at 0.1 s of each other in the data). This is indeed what one would expect, as JPE takes care of the mismodelling of the noise but leads to an increased complexity during the analysis.

For the spread in recovered posteriors, contrary to what one had for the HS approach, the normalized width of the 90 per cent confidence interval does not align itself on the diagonal. Indeed, since we have more varying scenarios, with larger or tighter posteriors in some cases, the spread can be significantly larger or tighter in the JPE case compared to the SPE scenario. In addition, since there is no

result in a better recovery of the louder signal than the original PE run.

Fig. 6 shows the mismatch for the recovered dominant waveform after the first and third HS run. The match after the third run is better in 62 per cent of the cases, compared with 50 per cent expected if the procedure had no effect at all. This small effect is also observed on the offset plot – see Fig. 7 – where the third HS step leads to better results in the same proportions. Even if the results get better for some events, it also leads to worse recoveries for other cases, and only a few of the other events have comparable results between the first and third HS steps. Therefore, it does not seem like simply applying successive HS steps converges to posteriors unaffected by the overlap. More sophisticated approaches appear to be needed, like an estimation of the deviation as suggested in Antonelli et al. (2021) and briefly explained in Section 2.2.

In the end, our runs for HS confirm what has been seen in previous research (Relton & Raymond 2021; Samajdar et al. 2021; Pizzati et al. 2022): doing PE for one signal neglecting the other can lead to significant biases when the two signals merge very close to each other. In addition, we have also shown that once the first signal is subtracted, analysing the second one with PE is less prone to deviations, even if the subtraction of the first event is not perfect.
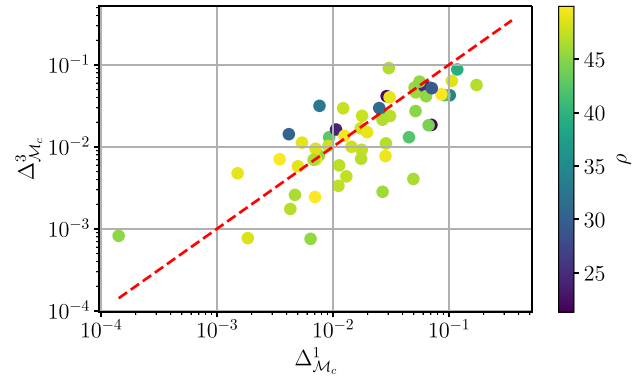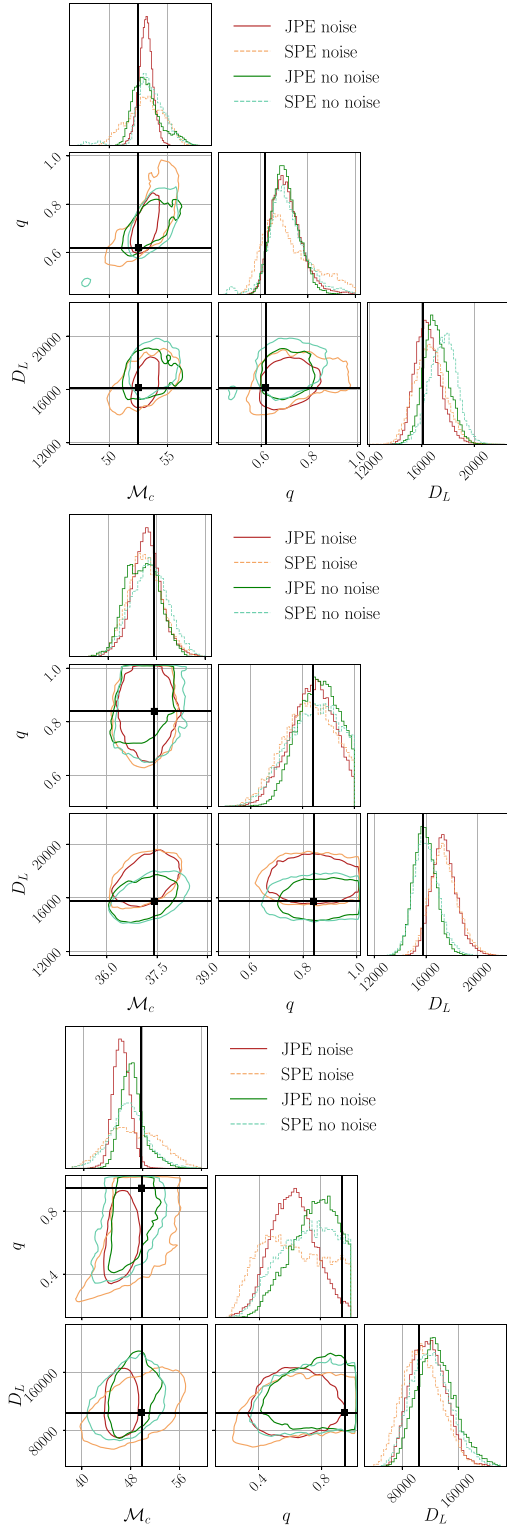
Figure 8. Comparison of the posteriors recovered for JPE and SPE with and without noise for different types of recovery. Top panel: the posteriors recovered with JPE for the noise case are narrower for the chirp mass and the mass ratio compared to the single PE case, while the posterior for the luminosity distance is narrower for JPE in the zero-noise case. Middle panel: JPE and SPE are very close to each other, with equivalent recovery in the two cases. Bottom panel: representation of a case where the JPE recovery is worse than for SPE. We get narrower posteriors, but the peak is shifted out of the 90 per cent confidence region.

significant difference for this quantity between HS and SPE, the relation between JPE and HS is the same as between JPE and SPE. The increased discrepancy between the two approaches is represented in Fig. 11. Nevertheless, the posteriors are evenly distributed above and below the diagonal representation, showing that, on average, the width of the posteriors is not significantly bigger in one method or the other.

Based on our JPE results, we see that, using an adapted likelihood, we obtain better results than by analysing both signals sequentially. However, this comes with the drawback that the computational time and the complexity of the problem are increased, making the approach less stable as can be seen by the wider variations in the widths of the posteriors. On the other hand, in some cases, JPE leads to narrower posteriors compared to SPE. This could be due to the inner product from the mixed term of the two signals following the introduction of multiple signals in the likelihood. However, a more in-depth study should be performed on a larger set of events to understand this behaviour. This is left for future work.

5 CONCLUSIONS

In this work, we have presented two ways to perform PE for two overlapping BBH signals: HS and JPE. We have implemented them and compared them to the usual single-parameter estimation. Because of limited computational resources, we focus on high mass and medium SNR events; major adaptions to the parameter estimation framework are needed to deal with a larger variety of signals.

For the HS method, we analyse the first signal, then subtract it (more precisely, the maximum likelihood signal) from the data before analysing the second signal still left in the data. On the other hand, JPE consists of analysing the two signals simultaneously. We note that since the likelihood is symmetric in the two signals when they have the same nature, we need to add a post-processing step to have samples corresponding to each event. For this purpose, we order the samples in time.

We have applied both methods to a population of BBH mergers to show the feasibility of the two approaches and their respective drawbacks. For HS, as already mentioned in the literature, we showed that analysing one signal while neglecting the other can lead to offsets in the recovered posteriors because of the erroneous noise representation. However, even when some deviations occur in the first signal, it does not necessarily impact the recovery of the second signal. We have also shown that there is no significant broadening of the posteriors compared to usual single-parameter estimation approaches. Therefore, it shows that HS suffers from biases due to overlapping signals for signals measured at very close merger times, which can lead to an incorrect inference of the parameters.

For JPE, we have shown that the offset in recovered posteriors is smaller than for HS while remaining higher than for SPE. This is understood as JPE having a correct noise representation for the two signals, which HS lacks. On the other hand, solving the likelihood for JPE means we need to explore a 30-dimensional parameter space, making the analysis more complex and computationally challenging than SPE. However, the recovered width of the posteriors is, on average, approximately the same as for SPE. But while the average is the same, we are confronted with posteriors that can be narrower or broader in JPE compared to SPE. This behaviour may originate from the cross-term between the two signals entering the likelihood when adapting it to account for multiple signals. More extended studies are needed to understand this effect.
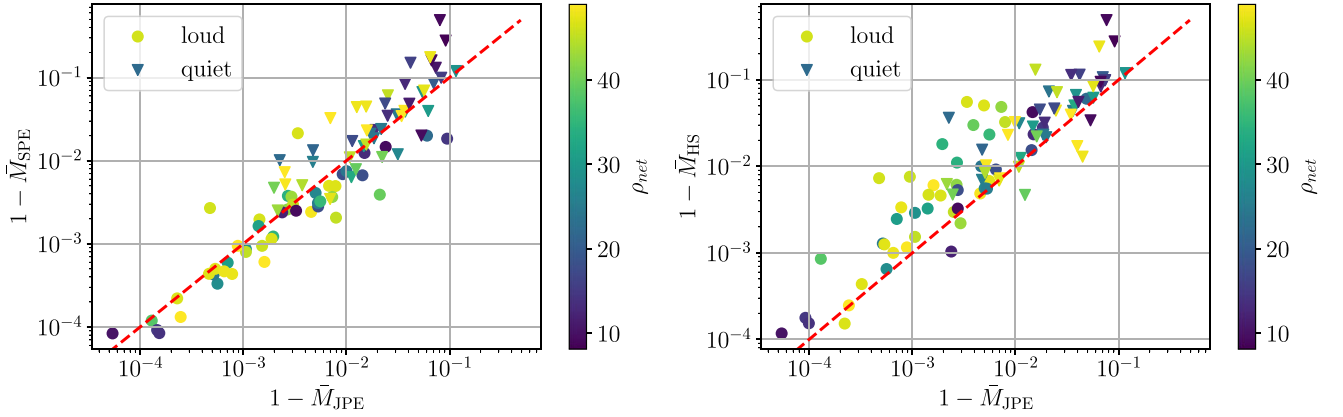
Figure 9. Left-hand panel: comparison of the mismatch for the JPE and the SPE cases with noise. Right-hand panel: comparison of the mismatch for JPE and HS with noise. We see that overall, the mismatch is higher for JPE than for SPE, while it is lower than that for the HS case. This is expected since JPE accounts for the two events in the data, which is better than neglecting one but more complicated than having only one signal present in the data and fitting that signal.
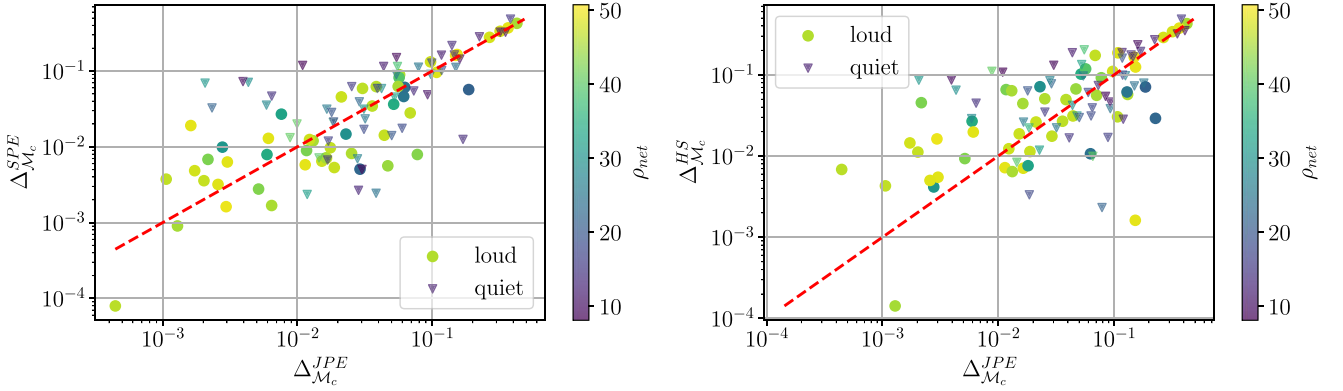

Figure 10. Left-hand panel: comparison of the offset of the recovered posterior for the chirp mass for JPE and SPE method. Right-hand panel: comparison of the offset of the recovered posteriors for the chirp mass for JPE and HS. The two plots indicate that the offset is reduced for JPE compared to HS, due to the better modelling of the noise, while it is still better in the SPE case, where the noise is well modelled and the problem at hand has a reduced complexity.
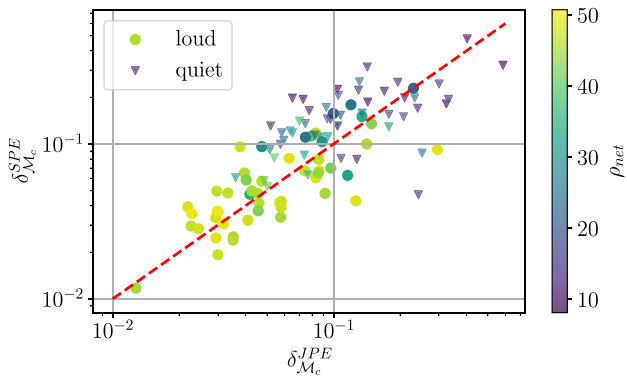
Figure 11. Comparison between the normalized width of the 90 per cent confidence interval for JPE and SPE for the noise case. Since the spreads are very close for SPE and HS, the same relation is valid for the JPE and HS comparison. We see that there are larger differences here between the two approaches but that globally, one is not better than the other as we have about 50 per cent of the events above the diagonal and the same proportion below.

Overall, our results indicate that common techniques for a joint likelihood approach are not yet at their best, and several options to make the sampling more efficient could be possible for future work. For example, one could impose the time ordering (or chirp mass hierarchy) directly during the sampling, by imposing that the arrival time of one event is smaller than that of another event. This would prevent the sampler from confusing the two events and enable them to converge more easily. Another possibility could be to use narrower priors motivated by the output of the search pipelines. This is possible for the chirp mass and the time of arrival but still contains risk as the search pipelines themselves can provide inaccurate point estimates for some critical parameters (Relton et al. 2022).

One of the major issues with the methods suggested here is the computational time required, as the data analysis takes up to a few months for overlapping BBH mergers with lower component masses. This would make it extremely hard, if not impossible, to keep up with the detection rate of the 3G-detector network. However, methods exist to speedup traditional parameter estimation methods, such as relative binning (Dai et al. 2018; Zackay et al. 2018; Leslie et al. 2021) or adaptive frequency resolution (Morisaki 2021). These methods could be adapted to overlapping signals in future work to reduce the computational time, enabling one to analyse other types of systems and to use a lower minimum frequency to get closer to the real

3G scenario. A totally different approach that could help in the analysis of such signals in the future is machine learning, where major progress has been made in the parameter inference for single compact binary colescences (Dax et al. 2021; Kolmus et al. 2022). In the future, one could think of adapting these methods to overlapping signals. In parallel to this work, a proof-of-concept study applying machine-learning techniques to overlapping signals has shown that it is possible to extract posteriors using normalizing flows with a reasonable precision (Langendorff et al. 2022).

To this end, we believe that this work makes a first step towards the analysis of overlapping CBC signals, which will be crucial to analyse GW data in the 3G detector era, as overlaps will become quite common.

DATA AVAILABILITY

The data underlying this article will be shared in reasonable request to the corresponding authors.

REFERENCES

Abbott B. P., et al., 2017a, Class. Quantum Gravity, 34, 044001
Abbott B. P., et al., 2017b, ApJ, 848, L12
Abbott R., et al., 2021a, preprint(arXiv2111.03606)
Abbott R., et al., 2021b, preprint(arXiv2112.06861)
Abbott R., et al., 2021c, preprint(arXiv2111.03634)
Abbott R., et al., 2021d, preprint(arXiv2111.03604)
Acernese F., et al., 2015, Class. Quantum Gravity, 32, 024001
Akutsu T., et al., 2019, Nature Astron., 3, 35
Akutsu T., et al., 2020, Prog. Theor. Exp. Phys., 2021, 24
Ashton G., et al., 2019, ApJS, 241, 27
Aso Y., Michimura Y., Somiya K., Ando M., Miyakawa O., Sekiguchi T., Tatsumi D., Yamamoto H., 2013, Phys. Rev. D, 88, 043007
Antonelli A., Burke O., Gair J. R., 2021, MNRAS, 507, 5069
Biscoveanu S., Talbot C., Thrane E., Smith R., 2020, Phys. Rev. Lett., 125, 241101
Collaboration T. L. S., 2015, Class. Quantum Gravity, 32, 074001
Cutler C., Harms J., 2006, Phys. Rev. D, 73, 042001
Dai L., Venumadhav T., Zackay B., 2018, preprint(arXiv1806.08793)
Dax M., Green S. R., Gair J., Macke J.H., Buonanno A., Schölkopf B., 2021, Phys. Rev. Lett., 127, 241103
Gao S., Hayes F., Croke S., Messenger C., Veitch J., 2022, Phys. Rev. Res., 4, 023006
Hild S., et al., 2011, Class. Quant. Grav., 28, 094013
Himemoto Y., Nishizawa A., Taruya A., 2021, Phys. Rev. D, 104, 044010
Hu Q., Veitch J., 2023, ApJ, 945, 10
Iyer B., Souradeep T., Unnikrishnan C., Dhurandhar S., Raja S., Sengupta A., 2011, LIGO-India Tech. rep., https://dcc.ligo.org/LIGO-M1100296/public

Kolmus A., Baltus G., Janquart J., van Laarhoven T., Caudill S., Heskes T., 2022, Phys. Rev. D, 106, 023032
Littenberg T., Cornish N., Lackeos K., Robson T., 2020, Phys. Rev. D, 101, 123021
Leslie N., Dai L., Pratten G., 2021, Phys. Rev. D, 104, 123030
Langendorff J., Kolmus A., Janquart J., Van Den Broeck C., 2022, Phys. Rev. Lett., 130, 171402
Morisaki S., 2021, Phys. Rev. D, 104, 044062
Oguri M., 2018, MNRAS, 480, 3842
Petiteau A., Babak S., Sesana A., de Araújo M., 2013, Phys. Rev. D, 87, 064036
Pizzati E., Sachdev S., Gupta A., Sathyaprakash B., 2022, Phys. Rev. D, 105, 104016
Punturo M., et al., 2010, Class. Quantum Gravity, 27, 194002
Regimbau T., Hughes S. A., 2009, Phys. Rev. D, 79, 062002
Regimbau T., et al., 2012, Phys. Rev. D, 86, 122001
Regimbau T., Evans M., Christensen N., Katsavounidis E., Sathyaprakash B., Vitale S., 2017, Phys. Rev. Lett., 118, 151105
Reali L., Antonelli A., Cotesta R., Borhanian S., M. Çalışkan, Berti E., Sathyaprakash B.S., 2022, preprint(arXiv2209.13452)
Reitze D., et al., 2019, BAAS, 51, 035
Relton P., Raymond V., 2021, Phys. Rev. D, 104, 084039
Relton P., et al., 2022, Phys. Rev. D, 106, 104045
Sachdev S., Regimbau T., Sathyaprakash B. S., 2020, Phys. Rev. D, 102, 024051
Samajdar A., Janquart J., Van Den Broeck C., Dietrich T., 2021, Phys. Rev. D, 104, 044003
Sathyaprakash B., et al., 2012, Class. Quant. Grav., 29, 124013
Sharma A., Harms J., 2020, Phys. Rev. D, 102, 063009
Somiya K., 2012, Class. Quant. Grav., 29, 124007
Speagle J. S., 2020, MNRAS, 493, 3132
Veitch J., Vecchio A., 2010, Phys. Rev., D81, 062003
Williams M. J., Veitch J., Messenger C., 2021, Phys. Rev. D, 103, 103006
Wu S., Nitz A. H., 2022, Phys. Rev. D, 107, 063022
Zackay B., Dai L., Venumadhav T., 2018, preprint(arXiv1806.08792)
Zhou B., Reali L., Berti E., 2022a, preprint(arXiv2209.01310)
Zhou B., Reali L., Berti E., M. Çalışkan, Creque-Sarbinowski C., Kamionkowski M., Sathyaprakash B.S., 2022b, preprint(arXiv2209.01221)

APPENDIX A: ZERO-NOISE RESULTS

In this section, we show results for the posteriors obtained when the injections are analysed without noise. The conclusions drawn from these plots are the same as in the noise case, which suggests that our
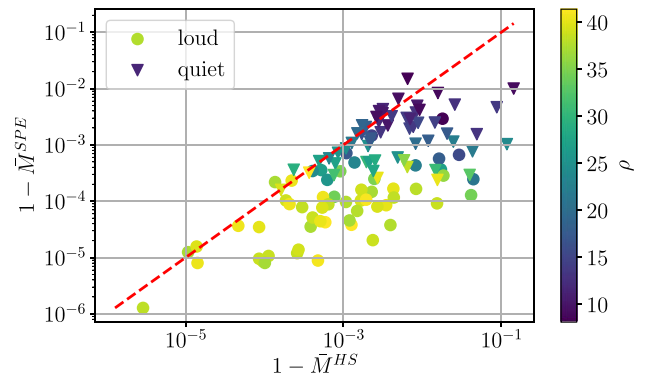

Figure A1. Mismatch for the SPE analysis versus HS analysis without noise. The red dashed line represents the diagonal where the mismatch is the same. The difference in waveform recovery between SPE and HS is more pronounced than in the case with noise, and it is clear that the recovery is worsened when using the HS approach.
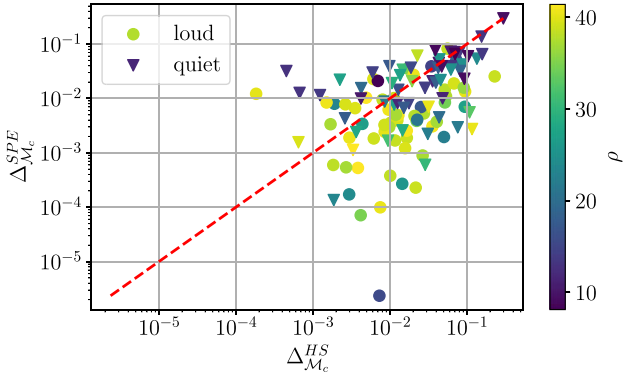
Figure A2. Representation of the offset of the recovered chirp mass without noise. Plotted is the difference between injected value and the median of the recovered value, normalized by the injected value for HS and SPE. One sees that offsets are more important for HS.

findings are robust, due to sampling effects, and not induced by the random noise added to the data.

A1 Hierarchical subtraction

In this section, we present the complementary zero-noise result for the HS method.

Fig. A1 represent the mismatch for HS versus SPE for the zero-noise case. As for the noise case, HS leads to higher mismatches, meaning that the recovered (maximum likelihood) parameters are a worse representation of the injected signals.

Fig. A2 represent the offset for the chirp mass for HS versus SPE without noise. Here HS shows higher offset for 74 per cent of the louder events and 57 per cent of the quieter ones. There is no significant difference compared with the injections into noise, and HS is still prone to more deviations.

Fig. A3 represents the normalized width of the posteriors for HS versus SPE. Similarly to the analysis with noise, the widths of the distributions are very close to each other.
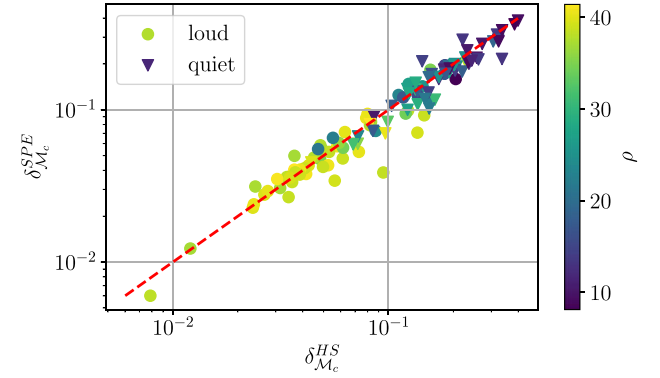

Figure A3. Comparison of the normalized width of the 90 per cent confidence interval for the chirp mass for the HS recovery and the SPE recovery, with zero-noise. The width of the recovered distribution is largely unaffected by the presence of another signal.

A2 Joint parameter estimation

In this section, we present the complementary zero-noise results for the JPE method.

Fig. A4 represent the mismatch between the JPE and SPE methods (left-hand panel) and the JPE and the HS methods (right-hand panel). This also shows that the JPE method leads to a better representation of the data than the HS method, but that the increased complexity of the problem leads to a decrease in the accuracy of the recovery. Still, the mismatch values are relatively low in the two cases.

Fig. A5 represent the comparison between the offsets between JPE and SPE, and JPE and HS for the zero-noise case. Here, one also has a larger offset for the JPE than for SPE (39 per cent of the events have smaller offset for JPE), and a larger offset for HS as for JPE (57 per cent of the events have a smaller offset for JPE).
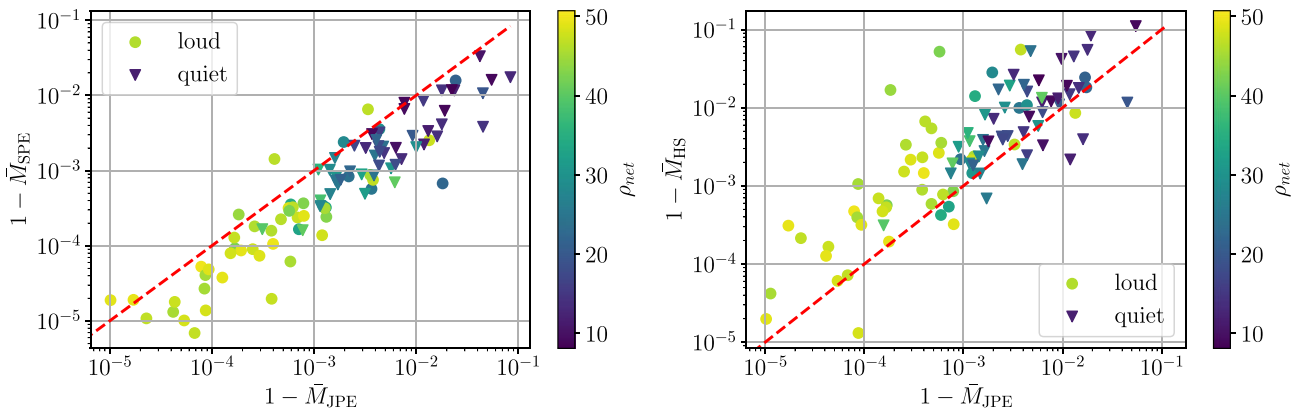

Figure A4. Left-hand panel: comparison of the mismatch for the recovered events for the two events for JPE and SPE without noise. Right-hand panel: comparison of the mismatch for JPE and HS without noise. Overall, the mismatch is higher for JPE than for SPE, while it is lower than for the HS approach. This is expected since JPE accounts for the two events in the data, which is better than neglecting one but more complicated than having only one signal present in the data and fitting that signal.
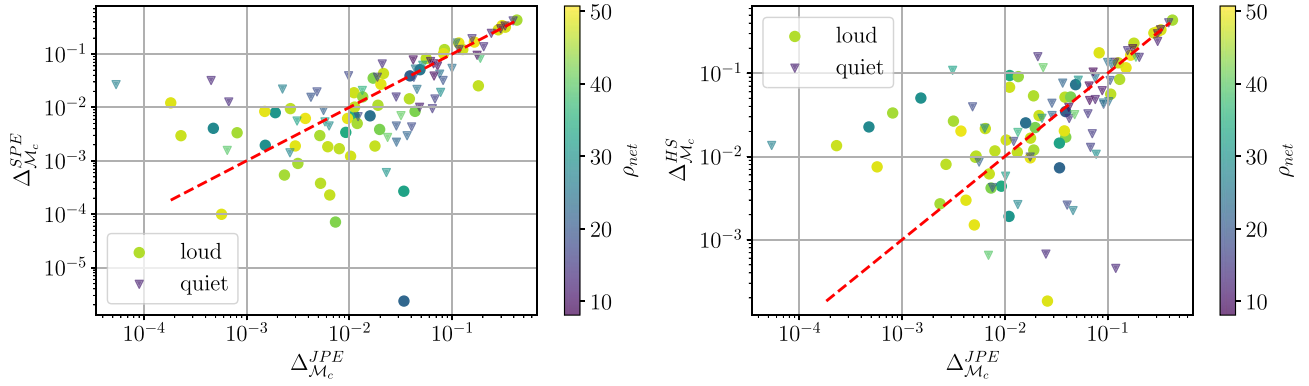
Figure A5. Left-hand panel: comparison of the offsets of the recovered posterior for the chirp mass for JPE and SPE method in the zero-noise case. Right-hand panel: comparison of the offsets of the recovered posteriors for the chirp mass for JPE and HS in the zero-noise case. The two plots indicate that the offset is lower for JPE than for HS, due to the better modelling of the noise, while it is still better in the SPE case, where the noise is well modelled and the problem at hand has a reduced complexity.
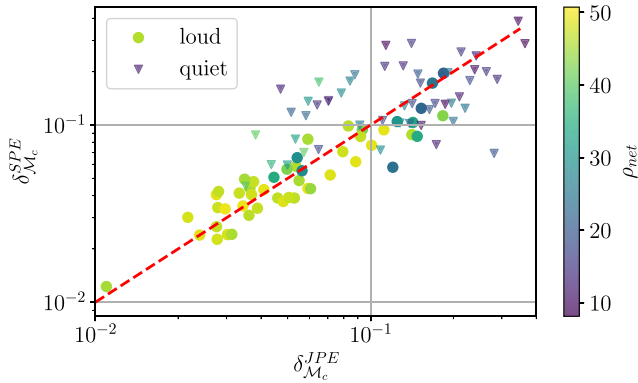

Fig. A6 represents the normalized width of the posteriors for JPE versus SPE (which is comparable to JPE versus HS since HS versus SPE has widths aligning along the diagonal). There is more variance in this plot than for SPE vs HS. This is because the JPE method is significantly different from SPE, and we have a bigger variety of posteriors. Indeed, for JPE, we sometimes get broader posteriors but also tighter ones, depending on the characteristic of the two signals present in the data. However, the points are evenly distributed on the two sides of the diagonal, showing that, on average, none of the methods has widened posteriors compared to the other.

Figure A6. Comparison between the normalized width of the 90 per cent confidence interval for JPE and SPE, in absence of noise. Since the spreads are very close for SPE and HS, the same relation is established for the JPE and HS comparison. There are larger differences between the two approaches but, globally, one is not better than the other as the events are approximately evenly distributed around the diagonal.

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