



# Null-stream analysis of Pulsar Timing Array data: localization of resolvable gravitational wave sources

Janna M Goldstein,<sup>1★</sup> John Veitch,<sup>1,2</sup> Alberto Sesana<sup>1</sup> and Alberto Vecchio<sup>1</sup>

<sup>1</sup>*School of Physics and Astronomy and Institute for Gravitational Wave Astronomy, University of Birmingham, Edgbaston B15 2TT, UK*

<sup>2</sup>*Institute for Gravitational Research, University of Glasgow, Glasgow G12 8QQ, UK*

Accepted 2018 April 4. Received 2018 January 25; in original form 2018 March 19

## ABSTRACT

Supermassive black hole binaries are expected to produce a gravitational wave (GW) signal in the nano-Hertz frequency band which may be detected by Pulsar Timing Arrays (PTAs) in the coming years. The signal is composed of both stochastic and individually resolvable components. Here, we develop a generic Bayesian method for the analysis of resolvable sources based on the construction of ‘null streams’ which cancel the part of the signal held in common for each pulsar (the Earth term). For an array of  $N$  pulsars there are  $N - 2$  independent null streams that cancel the GW signal from a particular sky location. This method is applied to the localization of quasi-circular binaries undergoing adiabatic inspiral. We carry out a systematic investigation of the scaling of the localization accuracy with signal strength and number of pulsars in the PTA. Additionally, we find that source sky localization with the International PTA data release one is vastly superior than what is achieved by its constituent regional PTAs.

**Key words:** black hole physics – gravitational waves – pulsars: general.

## 1 INTRODUCTION

The recent successes of the LIGO–Virgo collaboration (Abbott et al. 2016a,b, 2017) brought gravitational wave (GW) astronomy in the spotlight. Despite their great achievements, ground-based interferometers are only sensitive in a frequency range from about 10 to 1000 Hz and are thus suited for detection of stellar mass compact objects such as stellar mass black holes or neutron stars (Abadie et al. 2010). The GW spectrum, however, extends for several more decades in frequency (Colpi & Sesana 2017). In particular, the low-frequency band is expected to be dominated by GW signals coming from a class of much more massive astrophysical sources: supermassive black hole binaries (SMBHBs; Begelman, Blandford & Rees 1980).

The adiabatic inspiral of  $10^8$ – $10^{10}$   $M_\odot$  SMBHBs at cosmological distances generates loud GWs in the nHz-to-μHz frequency range (see e.g. Sesana, Vecchio & Colacino 2008), where ground based interferometers are completely deaf. Fortunately, Nature provided us with formidably stable natural clocks that might allow to hear such low-frequency waves in the foreseeable future: millisecond pulsars (MSPs; Lorimer 2008). Located at kpc distances within the Galaxy, MSPs behave like cosmic lighthouses sending periodic radio signals to the Earth. If a GW crosses the path of the radio photons, their null geodesic is modified, effectively resulting in a GW-induced redshift (Sazhin 1978). In practice radio pulses arrive

on Earth a little bit earlier or later than expected, an effect that can be measured if the time of arrivals (TOAs) of the radio pulses can be determined with enough precision. The TOAs of the most stable MSPs can be currently determined with an uncertainty of about 100 ns (Verbiest et al. 2016), an accuracy level approaching the expected delays induced by the most massive SMBHBs populating the Universe (Sesana, Vecchio & Volonteri 2009).

SMBHBs are expected to be common in the Universe, and PTAs will be mostly sensitive to the incoherent superposition of GWs coming from the large population of these cosmological sources (Rajagopal & Romani 1995; Jaffe & Backer 2003; Sesana et al. 2008). At the high-mass end and for sufficiently high frequencies, however, SMBHBs become sparser, and the loudest ones will likely be individually detectable as deterministic sources (Sesana et al. 2009). Consequently, several algorithms and pipelines have been assembled in recent years to detect and characterize both a stochastic GW background and individual deterministic sources (see Perrodin & Sesana 2017 for a recent review). In both cases, the challenge is to determine whether the data are better described by noise only or noise plus some GW signal. No GW detection has been reported thus far, and several pipelines have been used to produce upper limits on the strength of each type of source (Arzoumanian et al. 2014; Zhu et al. 2014; Lentati et al. 2015; Shannon et al. 2015; Taylor et al. 2015; Arzoumanian et al. 2016; Babak et al. 2016).

The problem of detecting a GW signal in PTA data is complicated by the variety of noise sources that can be either peculiar to each pulsars (e.g. spin modifications due to movements in the pulsar crust) or common in all observed systems (e.g. an error in the time

★ E-mail: [jgoldstein@star.sr.bham.ac.uk](mailto:jgoldstein@star.sr.bham.ac.uk)

standard used as a reference to measure TOAs). The latter are more insidious as they may introduce correlations in the residuals between the timing model and the observed times of arrival. The Hellings and Downs curve describes the cross-correlation due to an isotropic GW background between pairs of pulsars as a function of their angular separation on the sky (Hellings & Downs 1983). Common noise sources such as errors in the Solar system ephemeris or clock errors also produce cross-correlations between pulsars which may be confused with a GW signal (Tiburzi et al. 2016; Taylor et al. 2017).

Previous approaches to the detection problem have been developed which marginalize over these errors by including uncertainties in the timing model itself. However, these methods are still vulnerable to unmodelled systematics which may remain. For robust detection one would like to have an empirical estimate of the noise with which to compare the observed data. It is difficult to produce such an empirical background distribution due to the finite amount of PTA data available and the fact that the GW signal cannot be removed from the data. This issue has been addressed by the development of techniques to e.g. scramble the timing residuals so that the GW-induced cross-correlations are not present in the scrambled data realization (Taylor et al. 2017).

In this work, we investigate an alternative approach applicable to deterministic and individually resolvable signals. Through this method one cancels the GW signal exactly by exploiting redundancies in the data when the number of data streams exceeds the number of independent degrees of freedom (i.e. the number of polarizations) in the GW signal. Given independent data streams from  $N$  detectors and  $M$  GW polarizations, one can construct  $M$  GW polarization streams and  $N - M$  ‘null streams’, which have all GW power from a particular direction removed. This allows a discrimination between GWs and noise which can be used to construct a statistical model of the data. Assuming as working hypothesis that General Relativity holds, only the two tensor polarizations of the GW field are non-vanishing, thus allowing the construction of  $N - 2$  null streams.

The null-stream formalism is quite general and has been applied to analyses across the gravitational-wave spectrum. For networks of ground-based detectors, the method has been proposed to discriminate between signal and unmodelled noise (Wen & Schutz 2005; Ajith, Hewitson & Heng 2006; Chatterji et al. 2006; Rakhmanov 2006). In the context of the Laser Interferometer Space Antenna (LISA), an example of null stream is given by the Sagnac configuration of the detector (Amaro-Seoane et al. 2017), in which the interferometer channel are combined to cancel out GW signals, thus serving as detector calibrator to assess the instrumental noise level (Shaddock 2004). Recently, Zhu et al. (2015, 2016) adapted those techniques to PTAs, and investigated the benefits of using null streams to reconstruct the GW signal properties and quantify detection confidence in a frequentist framework.

Here, we develop a Bayesian PTA analysis using the null-stream formalism for an arbitrary deterministic GW source. We derive the associated likelihood function and use it for the recovery of the source properties. Although the null-stream formalism works in both the time and frequency domains given appropriate interpolation, for simplicity we consider the frequency domain analysis of simulated data containing a monochromatic GW source from an SMBHB (Hazboun & Larson 2016).

As a first application of the method, we investigate its performance in localizing resolvable SMBHBs. Sky localization is of paramount importance for PTA science, because it opens the possibility of identifying the source galactic host and of looking for

possible electromagnetic counterparts; consequently, it has been tackled by several authors in recent years (Sesana & Vecchio 2010; Lee et al. 2011; Babak & Sesana 2012; Ellis, Siemens & Creighton 2012; Taylor, Ellis & Gair 2014; Zhu et al. 2015, 2016; Wang & Mohanty 2017). For this specific problem, the null-stream technique is expected to produce equivalent results to other analysis methods (exploiting a linear transformation on the data). However, this application serves to lay out the formalism in the Bayesian framework. This will be used in future work to exploit the main advantage of null streams: by creating combinations of data that contain noise only, they are a powerful tool to discriminate signal from noise, thus allowing to tackle the issue of detection confidence, which is critical in PTA data analysis (Taylor et al. 2017). We perform a systematic investigation of the source sky localization as a function of signal-to-noise ratio (SNR) and number of pulsars in the array. We then consider the case of a detection with current PTAs, demonstrating the great benefits of combining regional PTA data under the aegis of International Pulsar Timing Array (IPTA).

The paper is organized as follows. In Section 2, we describe the null-stream construction and the Bayesian framework employed to extract source properties from the data. In Section 3, we investigate systematically the scaling of source sky localization capabilities as a function of the main PTA parameters and compare our results with previous work based on the Fisher Matrix approximation. In Section 4, we apply our techniques to current PTAs and demonstrate the benefit of the world-wide IPTA network. We summarize our main findings and discuss future prospects for expanding this work in Section 5.

## 2 METHOD

The basic idea behind the null-stream method is the following: data are obtained from  $N$  detectors that have a linear response to a GW signal. The two polarizations of the GW can be reconstructed from the detector output, which leaves the possibility to construct an additional  $N - 2$  independent data combinations. If the detector responses are known, these combinations can be made such that any GW signal is cancelled out, leaving only noise, hence the name null streams.

The following sections will explain in detail the construction of null streams for Pulsar Timing Array (PTA) data (Section 2.2) and our choice of a continuous wave signal (Section 2.3). Then, the use of null streams in the sky localization of a single source GW signal (Section 2.4) and the implementation for discrete data (Section 2.5) are discussed. First in Section 2.1, the formalism is set up in terms of the signal and the response of the PTA.

### 2.1 Response of a PTA

Assume we have a plane wave propagating in the direction  $\hat{\Omega}$ , with angular frequency  $\omega$ . A coordinate system can be chosen by using  $\hat{\Omega}$  and two additional orthonormal vectors:

$$\begin{aligned}\hat{\Omega} &= (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta) \\ \hat{m} &= (\sin \phi, -\cos \phi, 0) \\ \hat{n} &= (\cos \theta \cos \phi, \cos \theta \sin \phi, -\sin \theta).\end{aligned}\tag{1}$$

Here,  $\theta$  and  $\phi$  are the polar sky coordinates of the direction of propagation of the GW ( $-\hat{\Omega}$  points towards the source). The two orthogonal polarizations of the GW can be written in terms of

the  $+$  and  $\times$  transverse basis tensors:

$$\begin{aligned} e_{ij}^+(\hat{\Omega}) &= \hat{m}_i \hat{m}_j - \hat{n}_i \hat{n}_j \\ e_{ij}^\times(\hat{\Omega}) &= \hat{m}_i \hat{n}_j + \hat{n}_i \hat{m}_j. \end{aligned} \quad (2)$$

The metric perturbation due to the GW then is given by

$$h_{ij}(t) = h^+(t)e_{ij}^+ + h^\times(t)e_{ij}^\times, \quad (3)$$

where  $h^+(t)$  and  $h^\times(t)$  are the amplitudes of the two polarizations.

A GW propagating through the Galaxy affects the traveltime of radio emission travelling from a pulsar to the Earth. The resulting redshift in the pulse TOAs depends on the relative angle between the pulsar, in direction  $\hat{p}$ , and the GW propagation direction  $\hat{\Omega}$  (see e.g. Anholm et al. 2009; Lee et al. 2011).

$$z(t, \hat{\Omega}) = \frac{1}{2} \frac{\hat{p}^i \hat{p}^j}{1 + \hat{\Omega} \cdot \hat{p}} \Delta h_{ij}, \quad (4)$$

where  $\Delta h_{ij} = h_{ij}(t_E, \hat{\Omega}) - h_{ij}(t_p, \hat{\Omega})$  being  $h_{ij}(t_p)$  and  $h_{ij}(t_E)$  the metric perturbation at the pulsar at the time of pulse emission and at the Earth at the time of pulse detection, respectively. Only the Earth term adds up coherently in the analysis of multiple pulsars in the array. Therefore, for any burst-like signal with duration shorter than the traveltime of the pulses only the Earth term is relevant. For a continuous wave (e.g. from a SMBHB) on the other hand, the pulsar term is present, although its frequency may differ slightly from that of the Earth term as it samples different periods in the wave-train of the slow inspiral (Babak et al. 2016). Whether the pulsar and the Earth terms fall at different frequencies or not depend on the intrinsic properties of the GW source, the distance to the pulsar and the relative pulsar-source angular separation. Implementing realistic SMBHB population models and considering plausible developments of current PTAs, Rosado, Sesana & Gair (2015) found that either situation is possible, with comparable probability. To simplify the problem, as a first implementation, only the Earth term will be considered in our description of the signal.<sup>1</sup> This results in the following definition for the response functions  $F^+$  and  $F^\times$ :

$$\begin{aligned} z(t, \hat{\Omega}) &= \frac{1}{2} \frac{\hat{p}^i \hat{p}^j}{1 + \hat{\Omega} \cdot \hat{p}} (h^+(t)e_{ij}^+(\hat{\Omega}) + h^\times(t)e_{ij}^\times(\hat{\Omega})) \\ &\equiv F^+(\hat{\Omega})h^+(t) + F^\times(\hat{\Omega})h^\times(t). \end{aligned} \quad (5)$$

The observables for a PTA are not the redshifts, but the residuals  $r(t)$  obtained by taking the difference between the predicted and measured TOAs. The relation between the two is simply that the residuals are the integrated redshifts:

$$r(t) = \int_0^t z(\tau) d\tau. \quad (6)$$

Since the response functions are time independent, applying the integral to the previous equation (5) yields

$$r(t, \hat{\Omega}) = F^+(\hat{\Omega})H^+(t) + F^\times(\hat{\Omega})H^\times(t), \quad (7)$$

where  $H^+ = \int_0^t h^+(\tau) d\tau$  and similarly for  $H^\times$ . Our final calculations will be done in the frequency domain, for which we can substitute  $r \rightarrow \tilde{r}$  and  $H^+ \rightarrow \tilde{H}^+$ , and  $H^\times \rightarrow \tilde{H}^\times$ , since the Fourier transform is linear.

<sup>1</sup>The addition of the pulsar term is, on the other hand, necessary for non-evolving sources, and it might also help improving sky localization, as shown by Lee et al. (2011) and Zhu et al. (2016). We will address this case in future work.

Equation (7) can be written for each pulsar in the PTA, resulting in a collection of residuals  $\{r_i\}$ . Labelling the response functions  $F_i^+$  and  $F_i^\times$  for the pulsar in the direction  $\hat{p}_i$ , this can be combined into the matrix equation:

$$\mathbf{r} = \begin{pmatrix} F_1^+ & F_1^\times \\ F_2^+ & F_2^\times \\ \vdots & \vdots \\ F_N^+ & F_N^\times \end{pmatrix} \begin{pmatrix} H^+ \\ H^\times \end{pmatrix} \equiv \mathbf{F} \begin{pmatrix} H^+ \\ H^\times \end{pmatrix}, \quad (8)$$

where we have defined the response matrix  $\mathbf{F}$ .  $\mathbf{F}$  depends on the location of the GW source  $-\hat{\Omega}$ , but not on the parameters of the specific form of the GW polarizations, which makes the following null-stream construction general.

## 2.2 Null-stream construction

For a fixed direction  $\hat{\Omega}$ , the matrix  $\mathbf{F}$  defines a mapping from the two-dimensional space of GWs  $\mathbf{H} \equiv (H^+, H^\times) \in \mathbb{R}^{2,2}$  to the  $N$ -dimensional space of residuals from  $N$  pulsars in the array. The image of this mapping contains the residuals induced by a GW, but the measured response data  $\mathbf{d} \in \mathbb{D}$  is subject to additional noise  $\mathbf{n}$  such that  $\mathbf{d} = \mathbf{F}\mathbf{H} + \mathbf{n}$ . For  $N > 2$ , the space  $\mathbb{D}$  can be split into a two-dimensional subspace with the image of  $\mathbf{F}$ , and an orthogonal  $N - 2$ -dimensional subspace. This second subspace – the null space  $\mathbb{A}$  – is spanned by a set of  $N - 2$  independent null streams that are also independent of the GW.

There are different ways to choose the  $N - 2$  independent null streams from detectors' output  $\mathbf{d}$  [e.g. Zhu et al. (2015, 2016) use a different method than us]. However, it is possible to impose the more stringent requirement that the null streams are orthogonal. The method that we describe here has been adapted from works by Chatterji et al. (2006) and Rakhmanov (2006). In short, it's a construction of a set of orthogonal basis vectors for the null space  $\mathbb{A}$ .

We are interested in inverting the mapping  $\mathbf{F}$  from gravitational waves to a given set of residual data so that we may reconstruct the signal and find the null streams. The maximum likelihood solution for  $\hat{\mathbf{H}}$  is given by  $\hat{\mathbf{H}} = \mathbf{F}_{\text{MP}}^{-1} \mathbf{d}$ , where  $\mathbf{F}_{\text{MP}}^{-1} \equiv (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{F}^\top$  is the Moore–Penrose pseudo-inverse of  $\mathbf{F}$  (Rakhmanov 2006). This can be seen as a projection of the data on to the two-dimensional subspace of  $\mathbb{D}$  spanned by the column vectors  $\mathbf{F}^+$  and  $\mathbf{F}^\times$  of the response matrix  $\mathbf{F}$ . The null streams are found by projecting on to the orthogonal space  $\mathbb{A}$ , to ensure the null-streams contain no component of the signal. Say a basis for the null space is  $\{\hat{e}_i\}$ , with  $\hat{e}_i \cdot \mathbf{F}^+ = \hat{e}_i \cdot \mathbf{F}^\times = 0$  where  $i \in 1, \dots, N - 2$  labels the basis vectors. Then the matrix  $\mathbf{A}$  with rows  $\hat{e}_i^\top$  is the null-space projection matrix because  $\mathbf{A}\mathbf{F} = \mathbf{0}$  as per construction [where  $\mathbf{0}$  is a  $(N - 2) \times 2$  zero matrix]. The  $N - 2$  null streams can then be calculated as

$$\mathbf{A}\mathbf{d} = \mathbf{A}(\mathbf{F} \begin{pmatrix} H^+ \\ H^\times \end{pmatrix} + \mathbf{n}) = \boldsymbol{\eta} + \mathbf{A}\mathbf{n}. \quad (9)$$

Here, we define  $\boldsymbol{\eta}$  to be the vector of null streams which all equal zero ( $\eta_i = 0$ ).

To find the basis  $\{\hat{e}_i\}$ , consider the projection operators  $\mathbf{P} = \mathbf{F}\mathbf{F}_{\text{MP}}^{-1}$  and  $\mathbf{S} = \mathbf{I} - \mathbf{P}$ , where  $\mathbf{I}$  is the  $(N \times N)$  identity matrix (see also Rakhmanov 2006). The first projects on to the column space:

$$\mathbf{P}\mathbf{F} = \mathbf{F}\mathbf{F}_{\text{MP}}^{-1}\mathbf{F} = \mathbf{F}, \quad (10)$$

<sup>2</sup>We drop the dependence on time as the logic applies to any particular timestamp (or Fourier frequency bin when using  $\tilde{\mathbf{H}}$ ).

and the second on to the null space:

$$\mathbf{S}\mathbf{F} = (\mathbf{I} - \mathbf{P})\mathbf{F} = \mathbf{F} - \mathbf{F} = 0. \quad (11)$$

However,  $\mathbf{S}$  is an  $(N \times N)$  matrix, whereas the null space only has  $N - 2$  dimensions. A way to reduce  $\mathbf{S}$  to  $[(N - 2) \times N]$  is to use the QR decomposition, which yields  $\mathbf{S} = \mathbf{Q}\mathbf{R}$ . Then, if  $\mathbf{S}$  has rank  $r$ , the first  $r$  columns of  $\mathbf{Q}$  form an orthonormal basis for the column vectors of  $\mathbf{S}$ . Therefore, the first  $N - 2$  columns of  $\mathbf{Q}$  form the basis  $\{\hat{\mathbf{e}}_i\}$  that we were looking for.

Because both the reconstructed GW polarizations and the null streams are informative, the projectors  $\mathbf{F}_{\text{MP}}^{-1}$  and  $\mathbf{A}$  are combined in the square matrix  $\mathbf{M}$ . The total projection of the data with  $\mathbf{M}$  is

$$\begin{aligned} \mathbf{M}\mathbf{d} &= \begin{pmatrix} (\mathbf{F}_{\text{MP}}^{-1})_1 \\ (\mathbf{F}_{\text{MP}}^{-1})_2 \\ \hat{\mathbf{e}}_1^\top \\ \vdots \\ \hat{\mathbf{e}}_{N-2}^\top \end{pmatrix} \left( \mathbf{F} \begin{pmatrix} H^+ \\ H^\times \end{pmatrix} + \mathbf{n} \right) = \begin{pmatrix} H^+ \\ H^\times \\ \eta_1 \\ \vdots \\ \eta_{N-2} \end{pmatrix} + \mathbf{M}\mathbf{n} \\ &\equiv \mathbf{h} + \mathbf{M}\mathbf{n}, \end{aligned} \quad (12)$$

where we have defined  $\mathbf{h}$  as the combined vector of GW polarizations and null streams.

### 2.3 Continuous wave signal

The null-stream method can be used with any assumption on the functional form of the GW polarizations  $h^+(t)$  and  $h^\times(t)$ . Here we specialize and assume the signal to take the form of a monochromatic continuous wave from a circular SMBHB. The monochromatic assumption is valid so long as the SMBHB is light enough such that any frequency evolution over the time-scale of the observation is negligible, which is generally true for relevant systems (Sesana & Vecchio 2010; Taylor et al. 2016). Both polarizations  $h^+$  and  $h^\times$  are related, via the angle  $\psi$ , to the GW signal emitted by the source:

$$h^+(t) = A^+(\lambda) \cos(2\psi) - A^\times(\lambda) \sin(2\psi) \quad (13)$$

$$h^\times(t) = A^+(\lambda) \sin(2\psi) + A^\times(\lambda) \cos(2\psi), \quad (14)$$

where

$$A^+ = A \frac{1}{2} (1 + \cos \iota^2) \cos(\omega_0 t + \phi) \quad (15)$$

$$A^\times = A \cos \iota \sin(\omega_0 t + \phi). \quad (16)$$

The frequency of the GW is  $f_0 = \omega_0/2\pi$  (which is twice the orbital frequency). For a chosen frequency and sky position, the remaining parameters of the source are the binary's orbital inclination  $\iota$ , the polarization angle  $\psi$ , the phase offset  $\phi$ , and the amplitude  $A$ , which we encapsulate in the parameter vector  $\lambda$ . The amplitude depends on the physical parameters of the SMBHB:

$$A = \frac{4\mathcal{M}(\pi f_0)^{2/3}}{D_l}, \quad (17)$$

where  $\mathcal{M}$  is the redshifted chirp mass of the binary,  $D_l$  the luminosity distance to the source and  $f_0$  the observed GW frequency (here  $G = c = 1$ ). However, in this work, we treat  $A$  as an overall scaling factor of the signal.

The form of the signal needs to be changed when considering the PTA residuals instead of the redshifts, as in equation (7). Applying

the time integral to equations (15) and (16) yields

$$A_{(t)}^+ \equiv \int_0^t A^+(\tau) d\tau = \frac{A}{2\omega_0} (1 + \cos \iota^2) \sin(\omega_0 t + \phi) \quad (18)$$

$$A_{(t)}^\times \equiv \int_0^t A^\times(\tau) d\tau = -\frac{A}{\omega_0} \cos \iota \cos(\omega_0 t + \phi), \quad (19)$$

where we disregard constants of the integration.

### 2.4 Localization

The predictable shape of the null streams (they contain only noise) can have many applications. For example, the null stream statistic should follow the statistic of the noise and can therefore be used to validate candidate GW signals and assess detection confidence, which we plan to investigate in the future. In this work, we use it to estimate the sky location of a GW source. Only when constructing the response function and the matrix  $\mathbf{M}$  using the correct sky location, do the signal components in the null streams cancel out. Thus, an estimate for the sky location is obtained by varying  $\hat{\mathbf{S}}$  until the null streams are closest to zero (and, consequently, the GW polarizations closely match the model).

To quantify this, consider the posterior distribution on the sky location, under the assumption  $\mathcal{H}_{\text{sig}}$  that a signal is present:

$$p(\hat{\mathbf{S}}|\mathbf{d}, \mathcal{H}_{\text{sig}}) = \frac{p(\hat{\mathbf{S}}|\mathcal{H}_{\text{sig}})p(\mathbf{d}|\hat{\mathbf{S}}, \mathcal{H}_{\text{sig}})}{p(\mathbf{d})}. \quad (20)$$

The prior on the sky location  $p(\hat{\mathbf{S}}|\mathcal{H}_{\text{sig}})$  is assumed to be flat. To calculate the likelihood, a model for the data is needed. In the presence of a signal and additive Gaussian noise, equation (12) describes what is needed:  $\mathbf{d} = \mathbf{M}^{-1}\mathbf{h} + \mathbf{n}$ . This naturally leads to the Gaussian log-likelihood function:

$$l = -\frac{1}{2} \left( (\mathbf{d} - \mathbf{M}^{-1}\mathbf{h})^\top \mathbf{\Gamma} (\mathbf{d} - \mathbf{M}^{-1}\mathbf{h}) \right) + \text{norm.} \quad (21)$$

where  $\mathbf{\Gamma}$  is the inverse of the covariance matrix appropriate for the expected noise of the detector. The normalization is not written explicitly, as the likelihoods are normalized numerically as a last step in the calculation. Equation (21) can be rewritten using  $\mathbf{l} = \mathbf{M}^{-1}\mathbf{M}$  to the following form:

$$l = -\frac{1}{2} \left( (\mathbf{M}\mathbf{d} - \mathbf{h})^\top ((\mathbf{M}^{-1})^\top \mathbf{\Gamma} \mathbf{M}^{-1}) (\mathbf{M}\mathbf{d} - \mathbf{h}) \right) + \text{norm.} \quad (22)$$

The  $\mathbf{M}^{-1}\mathbf{h}$  term in equation (21) depends on both the sky location  $\hat{\mathbf{S}}$  through  $\mathbf{M}$  and the GW model parameters  $\lambda$  through  $\mathbf{h}$ . In equation (22), these dependences are split up over the terms  $\mathbf{M}\mathbf{d}$  and  $\mathbf{h}$ , which simplifies calculations.

To obtain the likelihood  $p(\mathbf{d}|\hat{\mathbf{S}}, \mathcal{H}_{\text{sig}})$ , this  $l$  is marginalized over the GW parameters  $\lambda$ . By having split the dependence on  $\lambda$  from  $\mathbf{M}\mathbf{d}$ , this quantity has to be calculated only once for each sky location. For our choice of a continuous wave signal in Section 2.3, the marginalization is done with a combination of an analytical and numerical integration. A benefit to the particular method of null-stream construction used (Section 2.2) becomes apparent here. For a diagonal covariance matrix  $\mathbf{\Gamma}^{-1}$  of the detector noise, the transformed matrix  $(\mathbf{M}^{-1})^\top \mathbf{\Gamma} \mathbf{M}^{-1}$  is largely kept diagonal (except for the covariance between the GW polarization amplitudes in the first two entries of  $\mathbf{M}\mathbf{d}$  and of  $\mathbf{h}$ ), which can make the numerical computation more efficient.

To quantify how well a GW source is localized, we define  $\Omega_{90}$  as the fraction of the sky area containing 90 per cent of the likelihood.



This quantity can be expressed as a fraction of the sky or in square degrees (since the whole sky is  $4\pi \text{ sr} \approx 4.1 \times 10^4 \text{ deg}^2$ ).

## 2.5 Discrete data

One drawback of the null-stream construction is that it requires the ability to take linear combinations of the data at a particular time or frequency. In practice, PTA residuals are not observed at the same time for each pulsar, so an interpolation in time or frequency is required to use this method on real data. In the following, we make the simplifying assumption that we can work with Fourier transformed quantities  $\tilde{h}^+$  and  $\tilde{h}^\times$ .

Any PTA observations will be discrete in time, and so is our simulated data. As a simplification, the simulation has  $n$  data points evenly spaced in time, with cadence  $\Delta t$ . This allows for calculating the discrete Fourier transform efficiently with the Fast Fourier Transform algorithm (Cooley & Tukey 1965). In the case of unevenly sampled data, interpolation methods can be used to estimate residuals at evenly sampled timestamps, allowing the calculation of the Fourier transform. This was addressed, for example, by Zhu et al. (2015), who used linear interpolation between data points. Alternatively, Fourier coefficients for an arbitrary basis of frequencies can be directly estimated via a likelihood calculation for any type of data, as demonstrated in Lentati et al. (2013).

For the study of sky localization with the null-stream method, the assumption is made that a source has been detected at a known frequency  $f_0$ . Therefore, the likelihood calculation can be restricted to the Fourier component at this frequency. To speed up the calculation, the number of points and cadence is matched such that there is only one non-zero Fourier component. This is effected when  $\Delta t$  is a multiple of  $(nf_0)^{-1}$ , in which case bin number  $f_0/\Delta f = f_0 n \Delta t$  completely contains the signal power. In general, the power in a discrete Fourier transformed is spread over multiple bins and can still be recovered.

The model for the GW polarizations from equations (18) and (19) needs to be adapted to Fourier-transformed discrete data. The transform of the sine and cosine functions are delta functions, which yield a contribution at  $f = -f_0$  and  $f = f_0$  when integrated over frequency. The first can be disregarded since we only have positive frequencies. In the discrete transform, this power will end up spread over the bin corresponding to  $f_0$ , and so there is an additional factor  $1/\Delta f = T$ . The model then is

$$\tilde{A}_{(t)}^+(f_0) \approx T \frac{A}{4\omega_0} (1 + \cos^2 \iota) e^{i(3\pi/2 + \phi)} \quad (23)$$

$$\tilde{A}_{(t)}^\times(f_0) \approx T \frac{A}{2\omega_0} (\cos \iota) e^{i(\pi + \phi)}. \quad (24)$$

The full model is  $\mathbf{h} = (\tilde{H}^+, \tilde{H}^\times, 0, \dots, 0)$  (see equation 12). As such, the model can be written as  $\mathbf{h} = \mathbf{h}_0 e^{i\phi}$ , which means that the likelihood from equation (22) can be analytically marginalized over the phase  $\phi$  (from 0 to  $2\pi$ ) (e.g. Jaranowski & Krolak 2010). Without explicitly writing the normalization, the marginalized likelihood is given by

$$p(\mathbf{M}\mathbf{d}|\hat{\mathbf{S}}, \lambda, \psi, \mathcal{H}_{\text{sig}}) = \int_0^{2\pi} d\phi p(\phi|\mathcal{H}_{\text{sig}}) p(\mathbf{M}\mathbf{d}|\phi, \hat{\mathbf{S}}, \lambda, \psi, \mathcal{H}_{\text{sig}}) \\ \propto \exp\left(-\frac{1}{2}(|\mathbf{M}\mathbf{d}|^2 + |\mathbf{h}_0|^2)\right) I_0(|\mathbf{M}\mathbf{d} \cdot \mathbf{h}_0|). \quad (25)$$

Here, the dot product and norm is analogous to the product in equation (22), which is weighted by the transformed inverse covariance

matrix:  $a \cdot b = a^\top ((\mathbf{M}^{-1})^\top \mathbf{\Gamma} \mathbf{M}^{-1}) b$ . The last term in equation (25) is the modified Bessel function of the first kind  $I_0$ .

The other parameters of the SMBHB  $A$ ,  $\psi$ , and  $\iota$  cannot be marginalized analytically. To get the likelihood for a specific sky location,  $p(\mathbf{M}\mathbf{d}|\hat{\mathbf{S}}, \mathcal{H}_{\text{sig}})$ , the other parameters are marginalized numerically. This is a three-dimensional integral over prior ranges  $0-\pi$  for  $\psi$  and  $0-10^{-12}$  for  $A$ . The prior for the inclination is flat in  $\cos \iota$ , with a range  $-1$  to  $1$ .

## 3 RESULTS: SKY LOCALIZATION PERFORMANCE

To investigate the performance of our localization method, we ran a set of simulations in which a GW signal according to the SMBHB model (equations 15 and 16) is added to white noise. The likelihood as in equation (22) is then calculated over a grid of sky locations, and marginalized over the model parameters, to determine  $\Omega_{90}$ . This grid consists of 12 288 equal area pixels made using the HEALpix algorithm (Górski et al. 2005) via HEALPY.<sup>3</sup>

Simulations were carried out with a varying number of pulsars  $N$  in the PTA, and over a range of SNRs. For  $N$ , the values 3, 5, 10, 20, 30, 50, and 100 were chosen. To construct at least one null stream, 3 is the minimum number of pulsars needed, whereas 50 is about the number of pulsars in the combined data set of the current PTA observatories, the IPTA (Verbiest et al. 2016). However, IPTA pulsars are not all equally good timers and most of the information is carried by the  $\sim 10$  best ones. In this respect, an array with  $N = 50$  good timers is more comparable to what might be achieved in the future with the Square Kilometre Array (SKA; see Janssen et al. 2015). The range of SNR values used is 1–30. This is the cumulative SNR in the PTA, i.e. summed over the pulsars:

$$\text{SNR}^2 = \sum_{p=1}^N \sum_{i=0}^{n-1} \frac{r_{i,p}^2}{\sigma_p^2}. \quad (26)$$

Here,  $\{r_{i,p}\}$  is the time series of  $n$  residuals from pulsar  $p$ . The noise model consists of white noise in the residuals with rms  $\sigma_p$  for each pulsar. All  $\sigma_p$  are set to 100 ns.<sup>4</sup> To fix the SNR to a given value, the amplitude of the injected GW is adjusted accordingly.

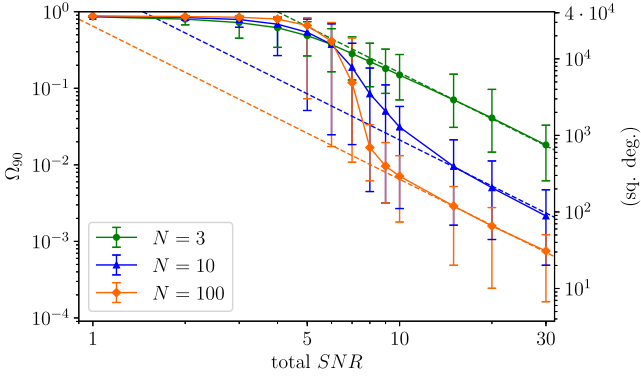
For each pair of  $N$  and SNR values, 10 simulations were performed injecting a GW source at  $\theta = \pi/2$ ,  $\phi = 0$ , with a frequency of 20 nHz, and pulsars at randomized locations (with a uniform prior over the sky). These random choices are seeded such that for a given  $N$ , for each SNR the same 10 PTA configurations are used. 300 data points were simulated with a cadence of  $10^6$  s, such that the data contain 6 full cycles of the GW signal.

### 3.1 Scaling with SNR at fixed $N$

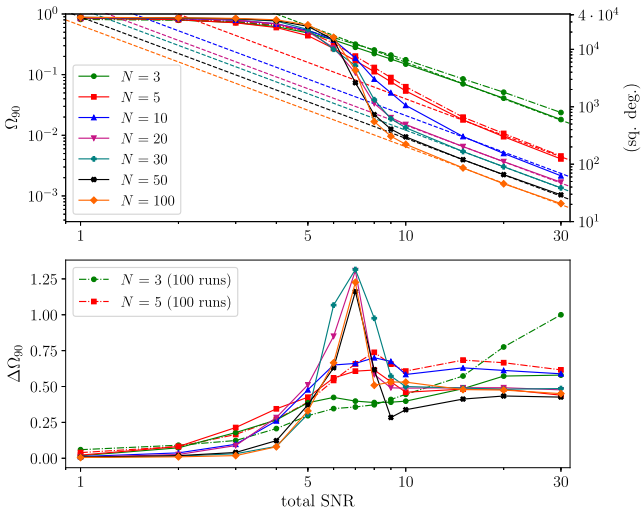
We investigate the sky localization as a function of the two main parameters identifying the detection, namely the SNR and the number of pulsars in the array. Here, we fix  $N$  and vary SNR, and in Section 3.2, we will fix SNR and vary  $N$ . In Fig. 1, the results for  $N = 3, 10$ , and 100 are shown with points referring to the mean  $\Omega_{90}$  and error bars spanning the range of results in the 10 simulations for each data point. At low SNR,  $\Omega_{90}$  is limited to 90 per cent of the sky, as there is no information gained from the data. For  $5 < \text{SNR} < 10$

<sup>3</sup>healpy.readthedocs.io

<sup>4</sup>A more sophisticated noise model could be used by taking the product  $r_i \Gamma_{ij} r_j$  with an inverse covariance matrix  $\mathbf{\Gamma}$ .



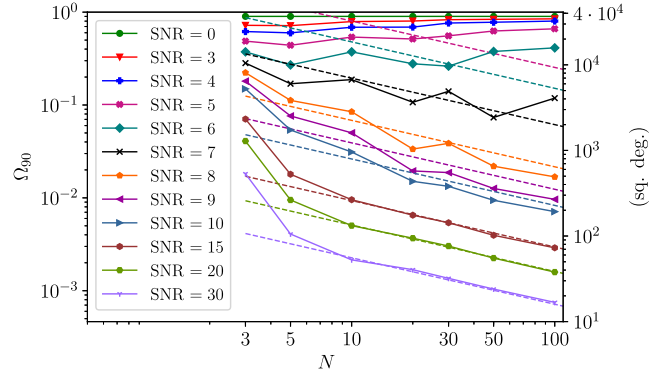
**Figure 1.** Fraction of the sky area containing 90 percent of the likelihood ( $\Omega_{90}$ ) versus total SNR. Data points show the mean value of 10 random realizations of a PTA with  $N = 3, 10$ , and  $100$  pulsars, with the errorbars showing the total span of results (from minimum to maximum). A power law  $\Omega_{90} \propto (\text{SNR})^{-2}$  is fitted to the last three points of each curve. For visibility, not all PTA sizes are plotted, but curves without errorbars are shown in Fig. 2.



**Figure 2.** Top panel: Fraction of the sky area containing 90 percent of the likelihood ( $\Omega_{90}$ ) versus total SNR, as in Fig. 1. Bottom panel: Standard deviation normalized by the mean in the 10 realizations used for each point in the top panel. Dashed lines represent values as obtained with a re-run of 100 realizations.

data become informative and the sky localization rapidly improves, eventually converging to a  $\Omega_{90} \propto (\text{SNR})^{-2}$  relation at high SNR. This has to be expected since at high SNR the likelihood surface can be approximated by a multivariate Gaussian around the true value of the source parameters (Vallisneri 2008). Parameter determination then follow the theoretical scaling  $\Delta\lambda \propto (\text{SNR})^{-1}$ . Sky localization is given by a combination of the two angle parameters  $\theta$  and  $\phi$  (or equivalently right ascension and declination); therefore, the scaling  $\Omega_{90} \propto (\text{SNR})^{-2}$  is recovered.

In the region around SNR from  $\sim 5$  to  $\sim 10$ , a transition occurs between the two regimes (from non-informative to informative data). In Fig. 2, the medians of the 10 runs for all values of  $N$  are plotted, along with the spread in  $\Omega_{90}$  in the bottom panel. The transition has a similar behaviour for all  $N$ , but the variance around the median value is much larger for large  $N$ . An explanation is that for low  $N$  the sky localization is still quite poor during the transition; regardless of the pulsar configuration,  $\Omega_{90}$  contains a significant fraction of the



**Figure 3.** Fraction of the sky area containing 90 percent of the likelihood ( $\Omega_{90}$ ) versus number of pulsars in the PTA. Data points show the mean value of 10 random realizations of a PTA with total SNR as indicated in the inset label. A power law  $\Omega_{90} \propto N^{-1/2}$  is fitted to each curve, ignoring the first two points (with  $N = 3$  and  $5$ ).

sky, as shown in Fig. A1 for the case  $N = 3$ . Conversely, for large  $N$  the information carried by the data in the transition region strongly depends on the specific pulsar location, as shown in Fig. A2 for the case  $N = 30$ . Here, we see that when some pulsars fall close to the source, its sky location is determined to high accuracy despite the low total SNR (e.g. bottom left-hand panel); on the other hand, when there are no pulsars located close to the line of sight to the source, sky localization is poor and  $\Omega_{90}$  can span as much as half of the sky (e.g. bottom right-hand panel).

### 3.2 Scaling with $N$ at fixed SNR

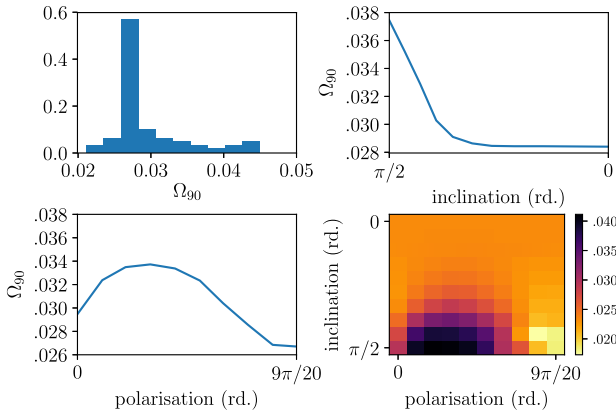
The median  $\Omega_{90}$  as a function of  $N$  is shown in Fig. 3 for all investigated SNR. For  $\text{SNR} \lesssim 6$  data are not informative and there is little dependence of the sky localization on  $N$ . As data become informative for  $\text{SNR} \gtrsim 7$ , sky localizations benefits from increasing  $N$ . In the range  $10 \leq N \leq 100$  the improvement in sky localization precision is well approximated by  $\Omega_{90} \propto N^{-1/2}$ , especially for the highest SNR investigated.

A possible explanation for this scaling behaviour can be given by the average (over random PTA realizations) angular distance of the closest pulsars to the source. For increasing  $N$ , the angular distance between the line of sight to the source and the closest pulsar scales with  $N^{-1/2}$  (for uniform randomly distributed pulsars). These closest pulsars contribute most to the sky localization (the antenna patterns are modulated on the smallest scales close to the pulsar).

This conclusion is however non-trivial and would need to be tested with  $N > 100$ . First, sky localization depends on the complex interplay of the antenna beam patterns of all the pulsars contributing to the array. Secondly, if the total SNR is held fix, not only the distance to the closest pulsar scales with  $N^{-1/2}$  but also the SNR contributed by each individual pulsar decreases, so that the  $\Omega_{90} \propto N^{-1/2}$  is not obvious. In any case, our systematic study indicates that for foreseeable future detections (involving a realistic number of pulsars up to 100 and SNR in the range of 6–30)  $\Omega_{90} \propto N^{-1/2}$  provides a good empirical fit to the sky localization scaling.

### 3.3 Dependence on source orientation

So far we considered optimally oriented sources, i.e. face on systems for which the two wave polarizations equally contribute to the signal (cf. equations 15 and 16), resulting in a circularly polarized wave.



**Figure 4.** Distribution of sky localizations ( $\Omega_{90}$ ) obtained with varying inclinations and polarization angles of the source. Top left: Normalized histogram of  $\Omega_{90}$  of all 100 runs. Top right: Distribution of  $\Omega_{90}$  for varying inclination. Bottom left: Distribution of  $\Omega_{90}$  for varying polarization angle. Bottom right: All  $\Omega_{90}$  shown as a colour plot for the grid of polarizations and inclinations used.

In this case, the signal does not depend on the polarization angle, which only adds a contribution to the initial phase offset. Although the sky localization scalings obtained in the previous sections are expected to hold for any source inclination and polarization, the normalization of  $\Omega_{90}$  might depend on those quantities. To assess this dependence, we proceed as follows. We fix a PTA of 10 pulsars and a source location in the sky. We then perform 100 simulations picking the source parameters from a  $10 \times 10$  uniform grid in polarization (chosen from  $\psi = 0 - 9\pi/20$ ) and inclinations (chosen from  $\cos(\iota) = 1 - 0$ ). For this particular experiment, we used noiseless data.

The bottom right-hand panel of Fig. 4 shows the obtained value of  $\Omega_{90}$  on the aforementioned grid, whereas the bottom left-hand and top right-hand panels show  $\Omega_{90}$  averaged over inclination and polarization, respectively. First, there is essentially no dependence of  $\Omega_{90}$  on  $\iota$  and  $\psi$  so long as the former is smaller than  $\approx \pi/3$ . This includes about 50 per cent of all binaries, assuming a uniform source orientation on the sphere. Secondly, the average sky localization degrades for  $\iota > \pi/3$ . However, compared to the reference value of  $\Omega_{90} = 0.028$  for the face-on case, the worst  $\iota - \psi$  combination results in  $\Omega_{90} = 0.046$ , which is a factor of 1.6 worse. The average sky localization of all the orientations with  $\iota > \pi/3$  is only a factor of 1.2 worse than the face-on case. We therefore conclude that the sky localization figures presented in Sections 3.1 and 3.2 are a fair representation of PTA capabilities for general SMBHBs.

### 3.4 Comparison with previous results

Our results can be compared to previous studies dealing with systematic investigation of sky localization accuracy as a function of number of pulsars and/or SNR (Sesana & Vecchio 2010; Taylor et al. 2016). Sesana & Vecchio (2010) investigated sky localization of individual GW sources with PTAs using the Fisher Matrix formalism. Their main result is shown in their fig. 7, where the sky localization  $\Delta\Omega$  is plotted against SNR and number of pulsars. Although results are overall compatible, there are several differences that are worth highlighting.

First, since they employ the Fisher Matrix formalism, Sesana & Vecchio find a perfect  $\text{SNR}^{-2}$  scaling down to  $\text{SNR} = 5$ . Our more realistic approach shows that this scaling kicks in only for  $\text{SNR}$

$\gtrsim 10$ , whereas for lower values, sky localization performances are much poorer. For example, at  $\text{SNR} = 5$  PTAs have essentially no source localization power and even at  $\text{SNR} = 7$ , typical performances are a factor of  $\approx 3$  worse than the  $\text{SNR}^{-2}$  extrapolation. This is particularly relevant since the signal builds up slowly with time, which means that the first confident single source PTA detection will necessarily have low SNR. PTAs will therefore have limited capabilities to pin down the source parameters in the early stages of detection.

Secondly, Sesana & Vecchio found that the  $N^{-1/2}$  scaling does not hold in general. Their fig. 7 shows that the sky localization improvement flattens out for  $N > 100$ , even though an  $N^{-1/2}$  line might provide a reasonable fit in the  $10 \leq N \leq 100$  range investigated in this work. It is likely that a saturation point is reached when the average contribution to the SNR of the closest pulsars of order 1. At that point, the signal added in each pulsar (if we keep the total SNR fixed) will be below the typical noise level, and no information about the source sky localization can be gained.

Thirdly, the normalization of the sky localization performance is different. For  $N = 100$  and  $\text{SNR} = 10$ , Sesana & Vecchio find a median  $\Delta\Omega \approx 40 \text{ deg}^2$ , to be compared to our value of about  $200 \text{ deg}^2$ . This is partly due to the different definition of  $\Delta\Omega$ , which in their study is the region of the sky with probability  $e^{-1} \approx 0.63$  of hosting the source. For a multivariate Gaussian likelihood surface, this area is a factor of 2.3 smaller than that enclosing the 90 per cent probability that we use. The 90 per cent probability region of Sesana & Vecchio is therefore  $\approx 100 \text{ deg}^2$ , which is only a factor of 2 smaller than what we find. Fisher Matrix calculation, however, provide a lower limit to the sky localization accuracy. Even for  $N = 100$  and  $\text{SNR} = 10$ , we find that the likelihood function is highly non-Gaussian, resulting in a slightly worse localization performance compared to the theoretical limit.

Taylor et al. (2016) constructed a Bayesian pipeline for detection and parameter estimation of *eccentric* binaries and carried out a systematic investigation of parameter errors as a function of SNR. Although the addition of eccentricity increases the complexity of the problem, we do not expect this parameter to couple with the source sky localization, and the results should be comparable with those of our analysis.

The relevant result for comparison is reported in their fig. 9, which shows  $\Delta\Omega$  as a function of SNR for a PTA of 18 pulsars with the properties of those used for the NANOGrav 9-yr GW upper limit (Arzoumanian et al. 2016). The trend of  $\Delta\Omega$  with SNR is very similar to what we found, showing an initial ‘transition phase’ up to about  $\text{SNR} \approx 8$ , then settling into the  $\text{SNR}^{-2}$  behaviour predicted in the strong signal limit. The overall normalization of the curve is also comparable. At  $\text{SNR} = 20$ , they find a 95 per cent probability region ( $\Omega_{95}$ ) of  $\approx 500 \text{ deg}^2$ , which is a factor of a few worse than the  $\Omega_{90}$  shown in our Fig. 1 for 10 and 20 pulsars, but comparable to the five pulsar case. This is likely due to the fact that the 18 pulsars they use are not randomly distributed in the sky and have different noise rms, therefore only the few best contribute significantly to the sky localization. Overall, we deem our results to be in agreement with those of Taylor et al. (2016).

## 4 IMPLICATIONS FOR CURRENT PTAS

The null-stream formalism developed in this work can be used to assess sky localization capabilities of current PTAs. In the previous section, we demonstrated the beneficial effect on sky localization of higher SNR and larger number of pulsars in the array. The obvious way to increase SNR and number of pulsars is to combine individual



**Table 1.** For each PTA, we list the number of pulsars  $N$ , the average number of TOAs per pulsar  $\bar{n}$ , and the average data set length  $\bar{T}$ . Also listed are the performances of each PTA for a face-on source with  $A = 10^{-13.5}$  averaged over position in the sky: SNR, sky localization  $\Omega_{90}$ , and improvement factors  $\mathcal{R}$  and  $\bar{\mathcal{R}}$  of IPTA compared to regional PTAs (see the text for details).

PTA	$N$	$\bar{n}$	$\bar{T}$ (yr)	SNR	$\Omega_{90}$ (deg <sup>2</sup> )	$\mathcal{R}$	$\bar{\mathcal{R}}$
EPTA	42	592	12.7	19.4	4492	22.0	3.3
PPTA	20	186	6.3	21.8	949	5.0	2.2
NANOGrav	17	50	4.8	8.0	14172	102.2	18.8
IPTA	49	1401	11.1	28.7	167	–	–

PTA data sets under the umbrella of IPTA. In this section, we therefore focus on the potential gain of IPTA for individual source localization.

With the aforementioned goal in mind, we need to compare the capabilities of an IPTA data set to those of the individual PTA data that went into its production. The only official IPTA data release to date is IPTA DR1, presented in Verbiest et al. (2016). We therefore use

(i) EPTA Data Release 1, presented by Desvignes et al. (2016), consisting of 42 MSPs monitored with radio telescopes at Effelsberg, Jodrell Bank, Nancay, and Westerbork;

(ii) the extended PPTA Data Release 1, presented by Manchester et al. (2013), consisting of 20 MSPs monitored with the Parkes radio telescope;

(iii) NANOGrav 5-yr data set, presented by Demorest et al. (2013), consisting of 17 MSPs, monitored with the Arecibo and Green bank radio telescopes;

(iv) IPTA DR1, presented by Verbiest et al. (2016), consisting of the combination of the three aforementioned data sets, for a total of 49 MSPs.

Several MSPs are monitored by multiple regional PTAs, and so the number of MSPs in IPTA does not correspond to the sum of those in the regional PTAs. However, by combining multiple data sets, IPTA features more high-quality pulsars than the regional PTAs. We also stress that we considered the regional PTA data releases that were used to build IPTA DR1, which is the meaningful thing to do since our scope is to assess the benefit of combining PTA data.

The current implementation of our technique allows to use different rms residuals and arbitrary sky location for each individual pulsar, but is only applicable to evenly sampled data spanning the same observation time. We therefore need to modify the PTA data accordingly, while keeping their properties as close as possible to the originals. For each PTA we therefore compute an average data set length  $\bar{T}$  as

$$\bar{T} = \frac{1}{N} \sum_{p=1}^N T_p, \quad (27)$$

where the index  $p$  runs over all pulsars in the array and  $T_p$  is the data set length of the  $p$ th pulsar. Likewise, we compute an average number of TOAs per pulsar  $\bar{n}$  as

$$\bar{n} = \frac{1}{N} \sum_{p=1}^N n_p, \quad (28)$$

where  $n_p$  is the number of TOAs of the  $p$ th pulsar in the array. We then round  $\bar{n}$  to the next integer. The values of  $\bar{T}$  and  $\bar{n}$  for each PTA are given in Table 1. Individual pulsar residual rms values are used as reported in Desvignes et al. (2016) (their table 1 under rms) for EPTA, in Manchester et al. (2013) (their table 7 under rms res.) for PPTA, in Demorest et al. (2013) (their table 2 under rms)

for NANOGrav, and in Verbiest et al. (2016) (their table 4 under residual rms) for IPTA.

Now that we have specified the properties of the PTAs, we conduct our experiment by considering a face-on circular SMBHB producing a monochromatic GW signal with frequency  $f = 20$  nHz and amplitude  $A = 10^{-13.5}$ , well within the reach of all PTAs. We place the source in turn at 48 different points in a grid over the sky and use the null-stream technique described in Section 2 to compute the  $\Omega_{90}$  sky localization.

Results are presented in Fig. 5, where contours have been interpolated over the grid. First, the uneven pulsar distribution in the sky results in a very source position-dependent sky localization precision. This is particularly true for EPTA and NANOGrav that have localizing power mostly in the left side of the map, where all the best pulsars are concentrated, but also for PPTA and IPTA to a lesser extent.

Secondly, the sky localization performance differs between PTAs. Due to the limited number of good pulsars and of the short data span, the NANOGrav 5-yr data set performs poorly. EPTA and PPTA, on the other hand, have comparable capabilities, even though the latter performs better in the right half of the map. The IPTA data set gives the best localization overall. The GW source can be localized to better than 20 deg<sup>2</sup> over a region of the sky of about 3500 deg<sup>2</sup>, whereas a comparable precision is achieved only by EPTA, on a smaller region of <1000 deg<sup>2</sup>. On the opposite end, IPTA can locate the source to better than 500 deg<sup>2</sup> regardless of its sky location and to better than 200 deg<sup>2</sup> over two-thirds of the sky. For comparison, PPTA can locate the source to better than 500 deg<sup>2</sup> in about half of the sky, and in some regions, localization is worse than 2000 deg<sup>2</sup>. On average, IPTA can localize the source within 167 deg<sup>2</sup>, whereas EPTA, PPTA, and NANOGrav can localize the source within 4492, 949, and 14172 deg<sup>2</sup>, respectively (cf. Table 1).

We can then define a relative improvement factor of IPTA sky localization with respect to regional PTAs as a function of the source location  $\hat{\Omega}$  as

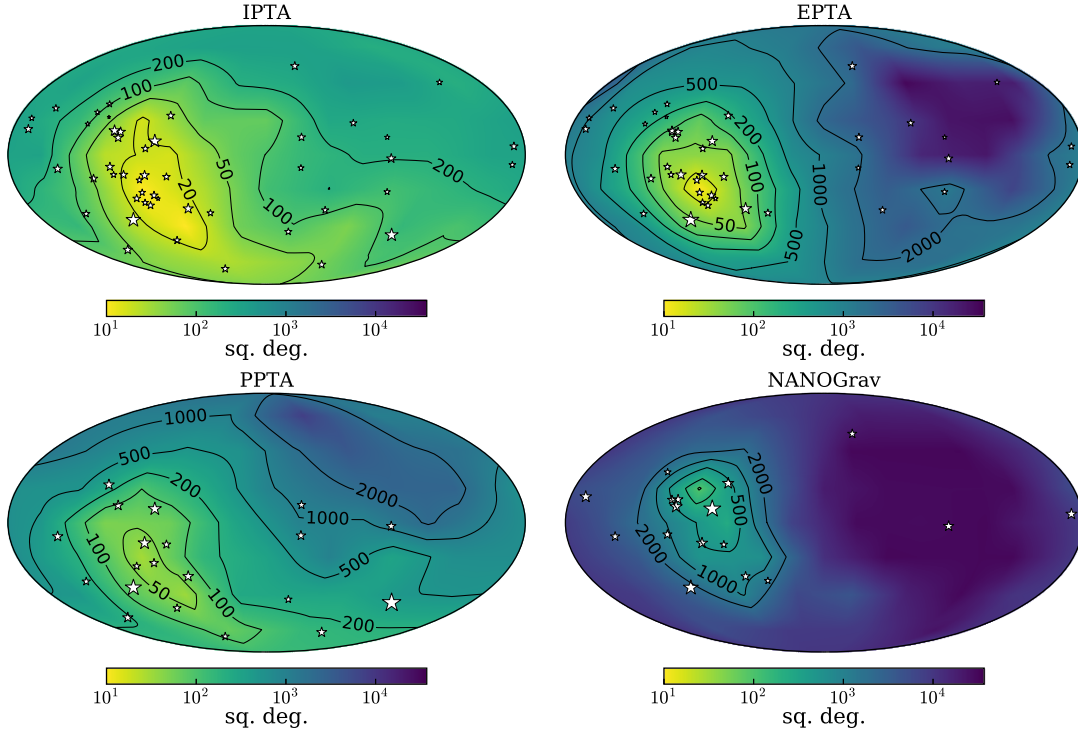
$$\mathcal{R}_X(\hat{\Omega}) = \frac{\Omega_{90, \text{IPTA}}(\hat{\Omega})}{\Omega_{90, X}(\hat{\Omega})}, \quad (29)$$

where  $X$  stands for EPTA, PPTA, or NANOGrav. This relative improvement is shown in Fig. 6. Compared to the best regional data set (PPTA), sky localization improves by more than a factor of 2 virtually everywhere in the sky, and up to a factor of 10 in some regions, confirming the superior performance of IPTA.

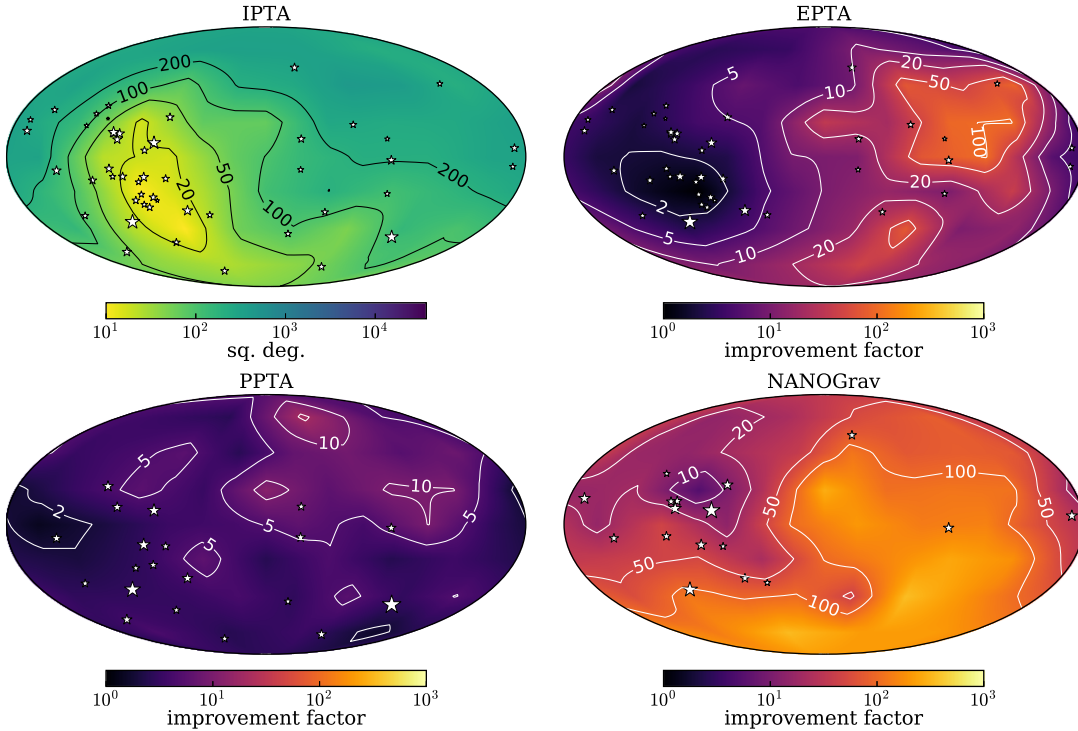
As shown in Section 3.1, the sky localization naturally improves as  $\text{SNR}^{-2}$ , but also (although to a lesser extent) as more pulsars are added to the array, even when keeping the total SNR fixed. We therefore investigate whether the benefits of the combined IPTA data sets go beyond the expected SNR scaling. We define the sky dependent ‘SNR gain normalized’ improvement factor  $\bar{\mathcal{R}}_X(\hat{\Omega})$  as

$$\bar{\mathcal{R}}_X(\hat{\Omega}) = \mathcal{R}_X(\hat{\Omega}) \times \left( \frac{\text{SNR}_X}{\text{SNR}_{\text{IPTA}}} \right)^2. \quad (30)$$

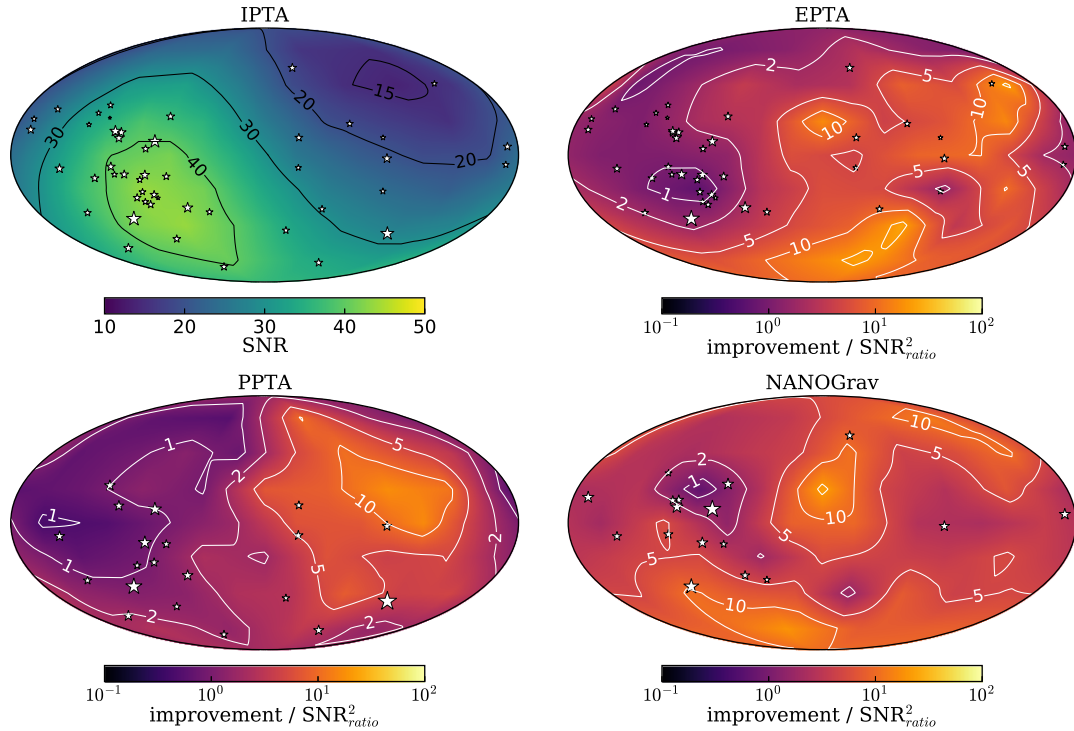




**Figure 5.**  $\Omega_{90}$  for a GW source as a function of sky location for IPTA, EPTA, PPTA, and NANOGrav. The simulated PTAs are approximates of the IPTA Data Release 1 and its constituent data sets: EPTA DR1, PPTA DR1, and the NANOGrav 5-yr data set (see the text for details). The maps are interpolated from 48 pixels for which a value of  $\Omega_{90}$  was obtained by placing a GW source with a fixed amplitude ( $1.0 \times 10^{-13.5}$ ) in the middle of that pixel. Contours are plotted at  $\Omega_{90} = 20, 50, 100, 200, 500$ , and  $1000 \text{ deg}^2$ .



**Figure 6.** The top left-hand panel is the same as in Fig. 5, whereas the remaining panels show the improvement factor  $\mathcal{R}_X(\hat{\Omega})$  of IPTA compared to EPTA, PPTA, and NANOGrav (see definition in the main text). The maps are interpolated from 48 pixels, as in Fig. 5. Contours are plotted at  $\mathcal{R}_X(\hat{\Omega}) = 2, 5, 10, 20, 50$ , and  $100$ .



**Figure 7.** The top left-hand panel shows the source SNR as a function of sky location in the IPTA. The remaining plot show the ‘SNR gain normalized’ improvement factor  $\bar{\mathcal{R}}_X(\hat{\Omega})$  of IPTA compared to EPTA, PPTA, and NANOGrav (see definition in the main text). The maps are interpolated from 48 pixels, as in Fig. 5. Contours are plotted at a value of  $\bar{\mathcal{R}}_X(\hat{\Omega}) = 1, 2, 5, 10$ , and 20.

By normalizing  $\mathcal{R}$  with the square of the SNR ratios,  $\bar{\mathcal{R}}$  quantifies the improvement brought by the better IPTA sky coverage. Results are shown in Fig. 7 and highlight that IPTA benefits indeed go beyond the source SNR increment.  $\bar{\mathcal{R}}$  is larger than unity in most of the sky for all regional PTAs. (Exceptions are a fourth, a sixteenth and a forty eighth of the sky for PPTA, EPTA, and NANOGrav, respectively. These are the areas where sky localization is best for the regional PTAs). In all cases, gain factors of up to 10 are found in parts of the sky, where the beneficial effect of better sky coverage of IPTA is maximized. Averaged over the sky, we have  $\bar{\mathcal{R}} = 3.3, 2.2, 18.8$  for EPTA, PPTA, and NANOGrav, respectively, certifying the benefits of IPTA data combination. We remark that the great improvements compared to NANOGrav are simply because only their five year data set was included in IPTA DR1. An IPTA DR2, including the 9-yr NANOGrav data set is currently under construction; this will allow to verify the benefits of IPTA when combining three data sets of comparable quality.

We caution that these results have been obtained by using an average timestamp for all pulsars of each specific array. In practice, PTA data are not evenly sampled and the timespan of observations varies from pulsar to pulsar. We expect, however, that considering more realistic PTAs would only have a minor impact on our conclusions. Here, we consider typical resolvable sources at a frequency of several tens of nHz. So long as the cadence of observations is much shorter than the GW period, the assumption of evenly sampled data should not really matter. Furthermore, although the cadence and timespan of individual pulsars are different, they usually lie within a factor of 2 of the average values that we assumed in Table 1, again suggesting that by using the actual timestamp of each pulsar we should reach similar conclusions. None the less, it is important to verify these expectations by employing an algorithm that can handle

the complexity of more realistic data sets, an extension that we plan to explore in future work.

## 5 CONCLUSIONS

In this paper, we introduced a general mathematical description for the construction of null streams in response of an individual GW source. This method is general, works both in the time and frequency domain and can be applied to any deterministic waveform. We then provided a Bayesian framework to extract the GW source parameters by exploring the likelihood given by the comparison of the constructed null streams and theoretical model. As proof of concept, we applied the method to the special case of a monochromatic GW source generated by a circular SMBHB, considering the Earth term only in the PTA response function. We used this set-up to carry out a systematic investigation of PTA sky localization capabilities as a function of the array parameters using the sky region containing 90 percent of the source location likelihood distribution ( $\Omega_{90}$ ) as figure of merit.

We found that for  $\text{SNR} \gtrsim 10$ ,  $\Omega_{90}$  scales as  $\text{SNR}^{-2}$ , as expected from theoretical arguments in the high-SNR limit. However, we find that at low SNR, this scaling breaks down, and the source cannot be well localized. A transition between the two regimes is found for  $5 \lesssim \text{SNR} \lesssim 10$ , in which the  $\Omega_{90}$  improvement is much steeper than the theoretical scaling.  $\Omega_{90}$  is also found to scale as the inverse square root of the number of pulsars in the array  $N^{-1/2}$ , at least for  $10 < N < 100$  and  $\text{SNR} \gtrsim 8$ . As a reference point, the median  $\Omega_{90}$  for a GW source observed with  $\text{SNR} = 10$  in an array of 100 equal MSPs randomly distributed in the sky is about  $200 \text{ deg}^2$ . These results are generally consistent with previous findings based on Fisher Matrix calculation, although there are significant differences in the  $5 < \text{SNR} < 10$  transition region and in the  $\Omega_{90}$  normalization.

We then used our formalism to investigate the sky localization capabilities of regional PTAs compared to IPTA. We found that the benefits of combining data in the IPTA framework go beyond the mere gain in SNR due to the accumulation of a larger amount of data. When normalized by SNR gain, IPTA is found to perform a factor between  $\sim 2$  and  $\sim 20$  better than regional PTAs. This is because combining PTA data provides a better sky coverage and increases the number of high-quality pulsars that contribute informative data to the detection. These findings demonstrate that combining regional data under the IPTA umbrella maximizes the scientific potential of PTAs as GW detectors.

The framework we applied in this study can be improved in several ways and extended to study a number of problems relevant to PTA data analysis. In particular, our current implementation requires that data are taken at simultaneous times in all pulsars if it were to be applied to a time-domain analysis. One of our primary future goals is to develop an implementation that can handle arbitrary data sets, with unevenly sampled data, gaps, and different time spans, thus allowing the assembly of a pipeline that can be applied to real data.

We also considered only the Earth term of the GW signal which may or may not be appropriate for the loudest SMBHBs, as shown in Rosado et al. (2015). If the frequency of the pulsar and Earth term cannot be separated, then, while the Earth terms may still be cancelled by the null-stream method, there will remain a contribution to the power from the pulsar terms. This could be treated as an excess unmodelled (noise) power, or may be modelled explicitly by introducing an additional amplitude and phase term for each pulsar. Efficient methods exist to either maximize or marginalize the calculation over these parameters, as shown for example by Zhu et al. (2016) and Taylor et al. (2014), and this is another avenue we wish to explore.

Of great interest is also the expansion of the formalism to treat the cases of multiple deterministic sources and stochastic GW backgrounds. Besides the determination of source parameters, the null-stream formalism provides a powerful tool to validate candidate GW signals and assess detection significance, a possibility that we want to explore in the context of Bayesian model selection.

## ACKNOWLEDGEMENTS

AS is supported by a University Research Fellow of the Royal Society, and acknowledge the continuous support of colleagues in the EPTA. We thank A. Possenti and S. Taylor for useful comments. JV and AV were supported by UK Science and Technology Facilities Council (STFC) grant ST/K005014/1 and ST/N000633/1, respectively. The methods for this work are implemented using the PYTHON programming language,<sup>5</sup> and make extensive use of the NUMPY/SCIPY library (Jones et al. 2001; Walt, Colbert & Varoquaux 2011). Some runs were performed with the University of Birmingham BlueBEAR HPC cluster.

## REFERENCES

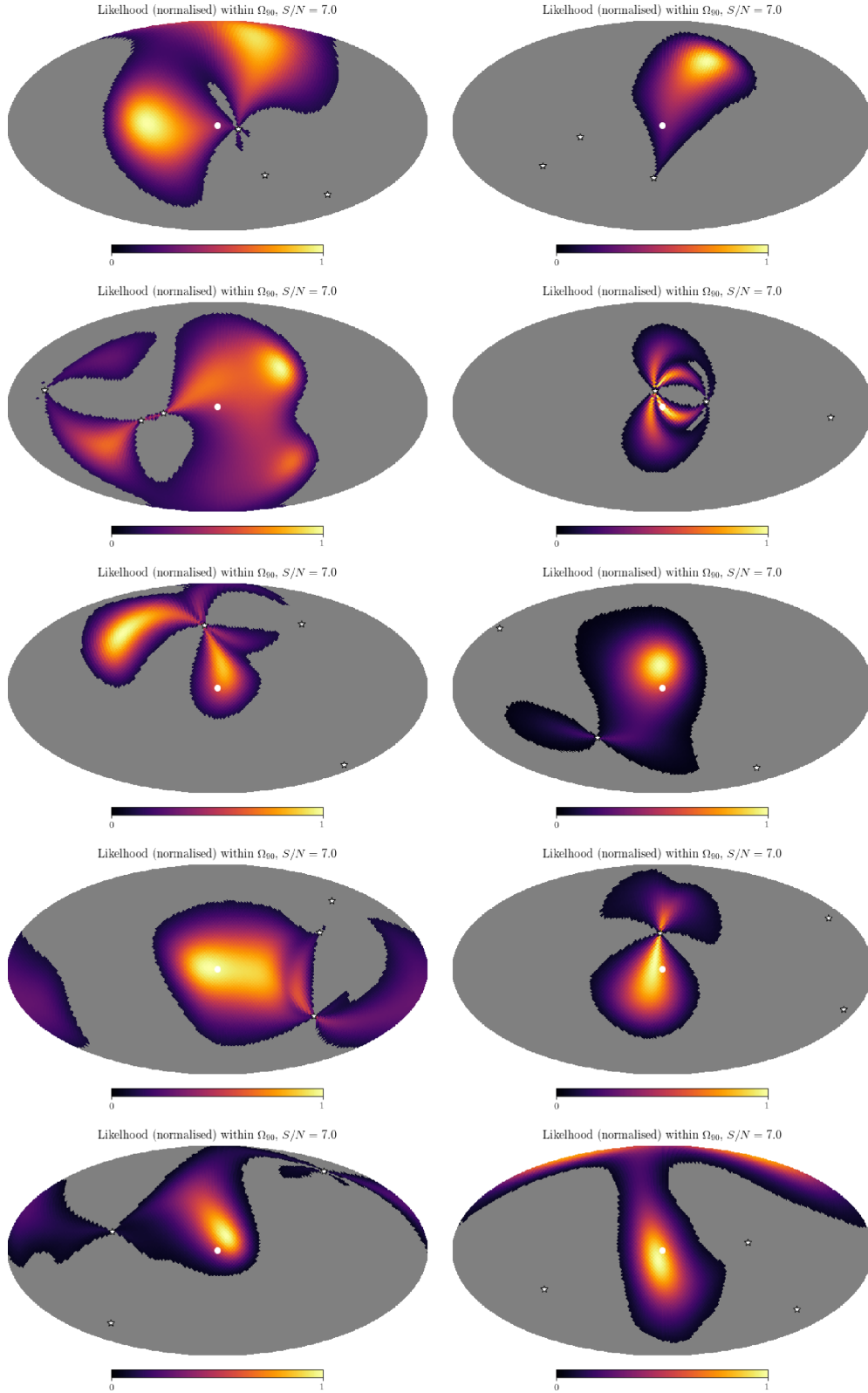
- Abadie J. et al., 2010, *Class. Quantum Gravity*, 27, 173001  
 Abbott B. P. et al., 2016a, *Phys. Rev. Lett.*, 116, 241103  
 Abbott B. P. et al., 2016b, *Phys. Rev. Lett.*, 116, 061102  
 Abbott B. P. et al., 2017, *Phys. Rev. Lett.*, 118, 221101  
 Ajith P., Hewitson M., Heng I. S., 2006, *Class. Quantum Gravity*, 23, S741

- Amaro-Seoane P. et al., 2017, eprint [arxiv.org/abs/1702.00786](https://arxiv.org/abs/1702.00786)  
 Anholm M., Ballmer S., Creighton J. D. E., Price L. R., Siemens X., 2009, *Phys. Rev. D*, 79, 084030  
 Arzoumanian Z. et al., 2014, *ApJ*, 794, 141  
 Arzoumanian Z. et al., 2016, *ApJ*, 821, 13  
 Babak S., Sesana A., 2012, *Phys. Rev. D*, 85, 044034  
 Babak S. et al., 2016, *MNRAS*, 455, 1665  
 Begelman M. C., Blandford R. D., Rees M. J., 1980, *Nature*, 287, 307  
 Chatterji S., Lazzarini A., Stein L., Sutton P. J., Searle A., Tinto M., 2006, *Phys. Rev. D*, 74, 082005  
 Chatterji S., Lazzarini A., Stein L., Sutton P. J., Searle A., Tinto M., 2006, *Phys. Rev. D*, 74, 082005  
 Colpi M., Sesana A., 2017, in Augar G., Plagnol E., eds, *Gravitational Wave Sources in the Era of Multi-Band Gravitational Wave Astronomy*. World Scientific Publishing Co., p. 43  
 Cooley J. W., Tukey J. W., 1965, *Math. Comput.*, 19, 297  
 Demorest P. B. et al., 2013, *ApJ*, 762, 94  
 Desvignes G. et al., 2016, *MNRAS*, 458, 3341  
 Ellis J. A., Siemens X., Creighton J. D. E., 2012, *ApJ*, 756, 175  
 Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759  
 Hazboun J. S., Larson S. L., 2016, eprint [arxiv.org/abs/1607.03459](https://arxiv.org/abs/1607.03459)  
 Hellings R. W., Downs G. S., 1983, *ApJ*, 265, L39  
 Jaffe A. H., Backer D. C., 2003, *ApJ*, 583, 616  
 Janssen G. et al., 2015, *Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, Giardini Naxos, Italy, p. 37  
 Jaranowski P., Krolak A., 2010, *Class. Quantum Gravity*, 27, 194015  
 Jones E., Oliphant T., Peterson P. et al., 2001, *SciPy: Open Source Scientific Tools for Python*. [Online; accessed April 6, 2018]  
 Lee K. J., Wex N., Kramer M., Stappers B. W., Bassa C. G., Janssen G. H., Karuppusamy R., Smits R., 2011, *MNRAS*, 414, 3251  
 Lentati L., Alexander P., Hobson M. P., Taylor S., Gair J., Balan S. T., van Haasteren R., 2013, *Phys. Rev. D*, 87, 104021  
 Lentati L. et al., 2015, *MNRAS*, 453, 2576  
 Lorimer D. R., 2008, *Living Rev. Relativ.*, 11,  
 Manchester R. N. et al., 2013, *PASA*, 30, e017  
 Perrodin D., Sesana A., 2017, *ArXiv e-prints*, 1709.02816  
 Rajagopal M., Romani R. W., 1995, *ApJ*, 446, 543  
 Rakhmanov M., 2006, *Class. Quantum Gravity*, 23, S673  
 Rosado P. A., Sesana A., Gair J., 2015, *MNRAS*, 451, 2417  
 Sazhin M. V., 1978, *Sov. Astron.*, 22, 36  
 Sesana A., Vecchio A., 2010, *Phys. Rev. D*, 81, 104008  
 Sesana A., Vecchio A., Colacino C. N., 2008, *MNRAS*, 390, 192  
 Sesana A., Vecchio A., Volonteri M., 2009, *MNRAS*, 394, 2255  
 Shaddock D. A., 2004, *Phys. Rev. D*, 69, 022001  
 Shannon R. M. et al., 2015, *Science*, 349, 1522  
 Taylor S., Ellis J., Gair J., 2014, *Phys. Rev. D*, 90, 104028  
 Taylor S. R., Huerta E. A., Gair J. R., McWilliams S. T., 2016, *ApJ*, 817, 70  
 Taylor S. R., Lentati L., Babak S., Brem P., Gair J. R., Sesana A., Vecchio A., 2017, *Phys. Rev. D*, 95, 042002  
 Taylor S. R. et al., 2015, *Phys. Rev. Lett.*, 115, 041101  
 Tiburzi C. et al., 2016, *MNRAS*, 455, 4339  
 Vallisneri M., 2008, *Phys. Rev. D*, 77, 042001  
 Verbiest J. P. W. et al., 2016, *MNRAS*, 458, 1267  
 Walt S. v. d., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, 13, 22  
 Wang Y., Mohanty S. D., 2017, *Phys. Rev. Lett.*, 118, 151104  
 Wen L., Schutz B. F., 2005, *Class. Quantum Gravity*, 22, S1321  
 Zhu X.-J., Wen L., Xiong J., Xu Y., Wang Y., Mohanty S. D., Hobbs G., Manchester R. N., 2016, *MNRAS*, 461, 1317  
 Zhu X.-J. et al., 2014, *MNRAS*, 444, 3709  
 Zhu X.-J. et al., 2015, *MNRAS*, 449, 1650

## APPENDIX: SELECTED SKY MAPS

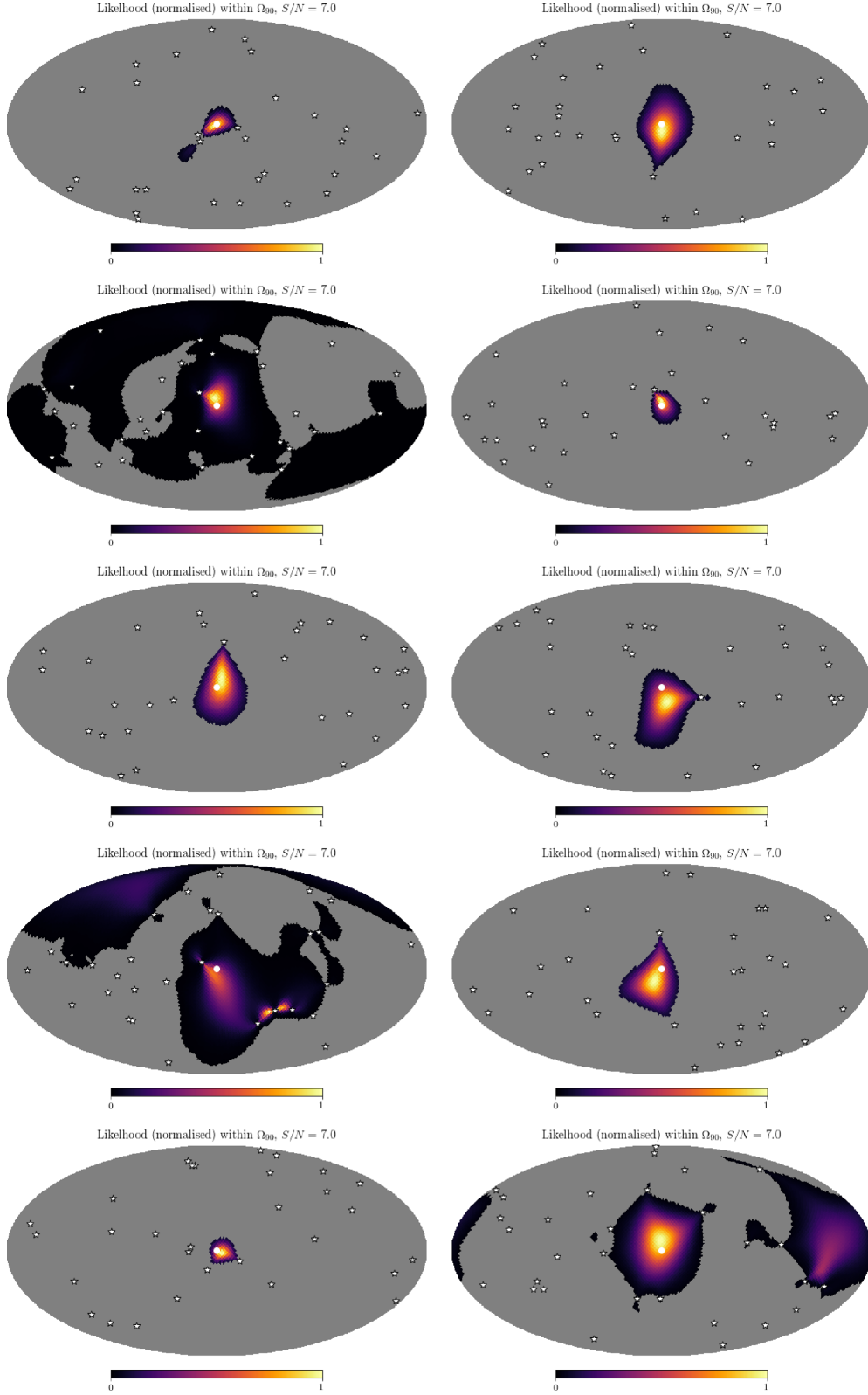
This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.

<sup>5</sup>[www.python.org](https://www.python.org)



**Figure A1.** Sky maps of 10 different PTA configurations with three pulsars, at a total SNR of 7. The injected source is always located in the middle of the map and indicated with a circle marker. The positions of the pulsars are marked with stars. Pixels not contributing to  $\Omega_{90}$  are masked in grey.  $\Omega_{90}$  ranges from 0.143 to 0.469 ( $\Delta\Omega_{90} = 0.563$  dex).





**Figure A2.** Sky maps of 10 different PTA configurations with 30 pulsars, at a total SNR of 7. The injected source is always located in the middle of the map and indicated with a circle marker. The positions of the pulsars are marked with stars. Pixels not contributing to  $\Omega_{90}$  are masked in grey.  $\Omega_{90}$  ranges from 0.0083 to 0.240 ( $\Delta\Omega_{90} = 1.84$  dex).