Empirically estimating the distribution of the loudest candidate from a gravitational-wave search

Rodrigo Tenorio,* Luana M. Modafferi, David Keitel, and Alicia M. Sintes

Departament de Física, Institut d’Aplicacions Computacionales i de Codis Comunitaris (IAC3), Universitat de les Illes Balears, and Institut d’Estudis Espacials de Catalunya (IEEC), Carretera de Valldemossa km 7.5, E-07122 Palma, Spain

(Dated: February 15, 2022)

Searches for gravitational-wave signals are often based on maximizing a detection statistic over a bank of waveform templates, covering a given parameter space with a variable level of correlation. Results are often evaluated using a noise-hypothesis test, where the background is characterized by the sampling distribution of the loudest template. In the context of continuous gravitational-wave searches, properly describing said distribution is an open problem: current approaches focus on a particular detection statistic and neglect template-bank correlations. We introduce a new approach using extreme value theory to describe the distribution of the loudest template’s detection statistic in an arbitrary template bank. Our new proposal automatically generalizes to a wider class of detection statistics, including (but not limited to) line-robust statistics and transient continuous-wave signal hypotheses, and improves the estimation of the expected maximum detection statistic at a negligible computing cost. The performance of our proposal is demonstrated on simulated data as well as by applying it to different kinds of (transient) continuous-wave searches using O2 Advanced LIGO data. We release an accompanying Python software package, distromax, implementing our new developments.

I. INTRODUCTION

The search for gravitational-wave (GW) signals can be formulated as a multi-hypothesis test between a background-noise hypothesis and a set of signal hypotheses, each asserting the presence of a signal with a specific set of parameters [1]. Actual search implementations, however, usually split this process into three stages: a detection stage, which simply assesses the presence of a feature in the datastream unlikely to be caused by noise (null-hypothesis test); a validation stage, in which candidates are sieved through a set of vetoes to discard any instrumental causes; and a parameter-estimation stage, in which a proper Bayesian hypothesis test is carried out to infer the actual parameters of any detected signal. This division is motivated by the increasing computing cost of each stage [2, 3], as a simple null-hypothesis test (usually assuming Gaussian noise) is orders of magnitude more affordable than a single parameter-estimation stage.

The standard detection stage consists in performing a finite number of detection statistic evaluations over the parameter-space region of interest, usually using matched-filtering against a bank of waveform templates [2, 4–11]. Loud templates, i.e. those scoring a high detection statistic, are deemed “signal candidates” and selected for the validation stage. The detection statistic can be usually interpreted as a Bayes factor, assessing the preference of the data for a particular signal hypothesis (represented by the template at hand) versus the background-noise hypothesis. Thus, the detection stage is a multi-hypothesis test in disguise in which parameter-space marginalization has been approximated to zeroth-order by maximization [12].

Loudest candidates from a template bank fall generally into one of two categories: The strongest excursions away from the background, such as an instrumental feature [13, 14] or a very clear GW signal (such as GW150914 [15]), are usually comparatively simple to deal with, as strong candidates tend to show characteristic signatures according to their cause. But weaker outliers that are in principle compatible with both a weak signal or an extreme event of the general noise background require a more careful analysis.

While much of the statistical framework used in this work is generally applicable, we mainly focus on the search for continuous gravitational-wave signals (CWs) [16], produced by long-standing quadrupolar deformations, such as in the case of non-axisymmetric spinning neutron stars (NS) [17]. From the point of view of the current generation of advanced detectors (Advanced LIGO [18], Advanced Virgo [19], and KAGRA [20]), they belong in the weak-signal regime, meaning they are expected to blend into the background distribution. Characterizing the expected distribution of extreme background candidates is, thus, a simple approach to identify interesting outliers in a search and quantify their significance.

Pioneering work on describing the distribution of the loudest candidate from a CW search using the $F$-statistic [21, 22] was presented in [23] and later extended in [24, 25]. Despite its wide applicability in the CW literature (see e.g. [26–32]), basic assumptions of the method make it insufficient for realistic template banks with a certain degree of correlation between neighbouring templates [33]. Latest developments on the subject used extreme value theory (EVT) to propose a suitable ansatz to circumvent the problems posed by template-
bank correlations [34]; but the concrete method requires re-evaluating full template banks many times (similar in spirit to that in [24]) and is thus computationally unsuitable for wide parameter-space searches.

This work proposes **distromax**, a new method to describe the distribution of the loudest candidate stemming from a generic GW search. The method generalizes with respect to previous approaches presented in [23, 25] in two main aspects. First, the method is robust to typical degrees of template-bank correlations arising either due to the overlap of nearby templates or mild non-Gaussianities in the data. Second, the method is applicable to a wider class of detection statistics, including other $\mathcal{F}$-statistic-based detection statistics such as line-robust statistics [35, 36] or transient CW search statistics [37], as well as detection statistics from other search approaches. An implementation of **distromax** is publicly available as a homonymous Python package [38].

The paper is structured as follows: Section II introduces basic data-analysis tools for CW searches and discusses the origin of parameter-space correlations. Section III describes the quantitative effect of parameter-space correlations on the distribution of the loudest candidate, comparing standard approaches in the field to extreme value theory results. Section IV introduces **distromax** to estimate the distribution of the loudest outlier of a search and discusses its basic phenomenology on synthetic data. In Section V, we apply **distromax** to the results of a search on O2 Advanced LIGO data for (transient) CW signals. Appendix A collects basic results in extreme value theory and provides further references for the interested reader. Appendix B proposes a simple method to deal with narrow-band noise disturbances, common in realistic CW searches. The robustness of **distromax** to the presence of weak CW signals is discussed in Appendix C.

### II. CONTINUOUS WAVE SEARCHES

In this section, we revisit the basics of CW searches to frame our discussion of **distromax**. Section II A reintroduces the $\mathcal{F}$-statistic and explicitly constructs its distribution under the noise hypothesis; Sec. II B uses the explicit construction to discuss the two possible origins of parameter-space correlations affecting a template bank; Sec. II C completes the analysis deriving the standard result for the distribution of the $\mathcal{F}$-statistic under the signal hypothesis.

The response of a ground-based GW detector to a passing CW or long-duration CW-like transient (tCW) is given by the linear combination of four linear filters [21, 37]

$$s(t; A, \lambda, T) = w(t; T) \sum_{\mu=0}^{3} A_\mu h_\mu(t; \lambda),$$

where $A$ represents the source’s amplitude parameters, namely GW amplitude $h_0$, inclination angle $i$, polarization angle $\psi$, and initial phase $\phi_0$, which can be combined into the so-called JKS decomposition $\{A_\mu, \mu = 0, 1, 2, 3\}$; and $\lambda$ describes the phase-evolution parameters, namely the GW frequency and spindown $\{f_0, f_1, f_2, \ldots\}$, the sky position $\vec{n}$, and possibly binary orbital parameters if the source orbits a companion. The time-dependent quadratures $h_\mu(t; \lambda)$ encompass the detector’s antenna pattern effects on the signal. The window function $w(t; T)$ is a time-dependent amplitude modulation parametrized by the transient parameters $T$ to account for tCW signals [37]. The standard CW signal model is recovered for $w(t; T) = 1 \forall t$.

Given a data stream $x$, the detection problem consists in deciding between the background noise hypothesis $H_N$, under which the data stream contains only Gaussian noise $x = n$, and the signal hypothesis $H_S$, according to which there is a (t)CW signal with a defined set of parameters $x = n + s(\lambda, A, T)$. Further hypotheses accounting for different non-Gaussian populations, such as narrow instrumental artifacts in the data [13], can be also included in the analysis [35, 36], although the usual approach is to apply post-processing veto strategies targeting specific types of disturbances [28, 39–41].

#### A. $\mathcal{F}$-statistic under the noise hypothesis

A basic tool to conduct CW searches is the $\mathcal{F}$-statistic, first introduced in [21, 22] as a maximum-likelihood estimator with respect to amplitude parameters $A$, and later re-introduced in a Bayesian context [12, 42–44]. The basic idea is to exploit the linear dependency of Eq. (1) on $A$ to analytically marginalize the matched-filtering likelihood using a suitable set of priors. The result can be readily expressed as a quadratic form [45]

$$2\mathcal{F}(\lambda) = \sum_{\mu, \nu=0}^{3} x_\mu(\lambda) \mathcal{M}^{-1}(\lambda) x_\nu(\lambda),$$

where $x_\mu$ are the projections of the data stream $x$ onto the four quadrature functions

$$x_\mu(\lambda) = \langle h_\mu(\lambda), x \rangle$$

and $\mathcal{M}^{-1}(\lambda)$ is the inverse Gram matrix associated to the four quadrature functions

$$\mathcal{M}_{\mu\nu}(\lambda) = \langle h_\mu(\lambda), h_\nu(\lambda) \rangle.$$

The functional scalar product [46]

$$\langle x, y \rangle = 4 \Re \int_0^\infty df \frac{x(f) y^*(f)}{S_n(f)}$$

accounts for the presence of correlated noise in the data stream through the single-sided power spectral density (PSD) $S_n$. Current implementations of Eq. (5) make use of the so called $\mathcal{F}$-statistic *atoms* [45], evaluated over individual Short Fourier Transforms (SFTs) of the data. These could be simply described as a set of complex-valued spectrograms (from now on *atomic spectrograms*)
containing both phase and amplitude information, whose proper combination results in an efficient computation of Eq. (2).

Under the noise hypothesis $\mathcal{H}_N$, the data stream is composed of zero-mean Gaussian noise and Eq. (3) implies the four projections $\{n_\mu(\lambda)\}$ are drawn from a 4-dimensional Gaussian distribution with covariance matrix $\mathcal{M}(\lambda)$. Hence,

$$\{n_\mu(\lambda)\} \sim \text{Gauss}(0, \mathcal{M}(\lambda)) \, ,$$

and $n_\mu(\lambda)$ values can be constructed as a linear combination of four zero-mean unit-variance Gaussian random variables

$$n_\mu(\lambda) = \sum_{\nu=0}^{3} L_{\mu\nu}(\lambda) g_\nu[\lambda] \, ,$$

where $g_\nu[\lambda] \sim \text{Gauss}(0,1)$ and $L$ is a $4 \times 4$ matrix such that $LL^T = \mathcal{M}$ (e.g. Cholesky decomposition). Here the square brackets indicate that Gaussian numbers are to be drawn independently for each template $\lambda$, but their distribution does not depend on $\lambda$; as opposed to round brackets, which represent deterministic relations.

Introducing these results into Eq. (2),

$$2F(\lambda) = \sum_{\mu=0}^{3} g_\mu[\lambda]^2 \, ,$$

we obtain $2F(\lambda)$ as the Euclidean norm of a 4-dimensional Gaussian vector. Consequently, the probability distribution associated to $2F$ under the noise hypothesis $\mathcal{H}_N$ for a fixed template $\lambda$ is given by a $\chi^2$ distribution with four degrees of freedom

$$p(2F|\mathcal{H}_N) = \chi^2_4(2F) \, .$$

### B. Template-bank correlations

The statistical properties of the right-hand side of Eq. (14) are independent of the specific phase-evolution template $\lambda$ under consideration. This suggests that evaluating $2F$ over a template bank using a single noise realization could, under suitable conditions, be equivalent to evaluating $2F$ for a single template over an ensemble of noise realizations.

Gaussian vectors $\{g_\nu[\lambda]\}$ are constructed from a noise stream as follows:

$$g_\nu[\lambda] = \sum_{\mu=0}^{3} L_{\nu\mu}^{-1}(\lambda) \langle h_\mu(\lambda), n \rangle \, .$$

The noise stream is projected onto four different deterministic functions of $\lambda$, $\{h_\mu(\lambda)\}$, and combined using a set of weights $L_{\nu\mu}^{-1}(\lambda)$, also dependent on $\lambda$. Such a projection is a weighted average of the atomic spectrogram bins visited by the frequency-evolution track associated to $\lambda$. Since the atomic spectrograms are constructed using finite time and frequency resolutions, the number of independent Gaussian vectors constructible out of them is equivalent to the number of templates with non-overlapping frequency tracks over the spectrograms (i.e. crossing different spectrogram bins). This result was stated in a simpler fashion in [23] by arguing that the typical number of bins in a narrow-banded atomic spectrogram is orders of magnitude smaller than the number of templates in a typical CW search crossing said spectrogram.

As discussed in [47], the average dissimilarity in frequency-evolution tracks of nearby parameter-space points is related to the fractional loss in detection statistic, usually referred to as mismatch [48]

$$m = 1 - \frac{2F(\lambda + \Delta \lambda)}{2F(\lambda)} \simeq \sum_{i,j} \Delta \lambda_i \Delta \lambda_j g_{ij} + O(\Delta \lambda^3) \, ,$$

where $\Delta \lambda_i$ represents an offset in an arbitrary parameter-space dimension and $g_{ij}$ is the parameter-space metric [48–53]. In the context of a grid-based CW search, the parameter-space metric can be employed to set up a template bank at a pre-specified maximum mismatch value [7, 54, 55]: the higher the mismatch, the coarser the template bank. An ensemble of templates with non-overlapping frequency-evolution tracks, then, corresponds to a coarse-enough template bank in the sense of large parameter-space mismatch.

In a real search, template banks tend to be set up using a moderate mismatch (e.g. $m \sim 0.2$) in order to produce dense-enough parameter-space coverings [7]. A first kind of template-bank correlation arises, then, as a result of the template-bank construction strategy. Latest developments on the subject [10, 53, 55, 56], however, suggest higher mismatch values ($m \sim 1$) could actually be compatible with a successful CW search, potentially suppressing the effect of these correlations.

A second kind of template-bank correlations, briefly discussed in [34], arises due to non-Gaussianities in the data (e.g. narrow instrumental features [13] or transient “pizza-slice” disturbances [29, 36, 57]). In this case, it is not a matter of re-using the same data on different templates; rather, a region of a priori independent spectrogram bins gets correlated due to the presence of a strong disturbance. As a result, non-overlapping templates crossing said correlated spectrogram region become correlated as well.

This same formalism applies to the search for tCWs, as the standard strategy in such cases is either to maximize or marginalize out any dependency on the transient parameters [37], obtaining in the end a detection statistic over an equivalent template bank to that of CW searches. Discussion on specific tCW detection statistics is postponed to Sec. V.

The presence of correlations in a template bank, thus, is a generic property of (t)CW searches, and their effects on any newly proposed method should be properly understood before attempting to interpret results on a real setup.
C. \( F \)-statistic under the signal hypothesis

We conclude this summary of standard CW search methods by considering the distribution of the \( F \)-statistic when there is a signal in the data. To derive its probability distribution under the signal hypothesis \( \mathcal{H}_S \), we simply apply \( n = x - s(\lambda, \mathcal{A}) \) and repeat the same reasoning up to Eq. (7), obtaining

\[
x^*(\lambda) = \sum_{\nu=0}^3 \nu L_{\nu}^{-1}(\lambda) (g_{\nu}[\lambda] + m_{\nu}(\lambda, \mathcal{A})) ,
\]

where

\[
m_{\nu}(\lambda, \mathcal{A}) = \sum_{\kappa=0}^3 L_{\kappa}^{-1}(\lambda) s_{\kappa}(\lambda, \mathcal{A})
\]

and, consequently, \( g_{\nu}[\lambda] + m_{\nu}(\lambda, \mathcal{A}) \) is a Gaussian random number with mean \( m_{\nu} \) and unit variance.

Introducing these results in Eq. (2),

\[
2F(\lambda) = \sum_{\nu=0}^3 (g_{\nu}[\lambda] + m_{\nu}(\lambda, \mathcal{A}))^2 ,
\]

the \( 2F \) under the signal hypothesis corresponds to the norm of a four-dimensional uncorrelated Gaussian vector with identity covariance matrix and mean vector equal to \( \{m_{\nu}(\lambda, \mathcal{A})\} \). The probability distribution is, as a result, a non-central chi-squared distribution with four degrees of freedom

\[
p(2F | \rho^2 \mathcal{H}_S) = \chi^2(2F; \rho^2) ,
\]

where the non-centrality parameter \( \rho^2 \) is defined as

\[
\rho^2 = \sum_{\nu=0}^3 m_{\nu}^2 = \sum_{\mu, \nu=0}^3 A_{\mu} M_{\mu \nu} A_{\nu} = \langle s, s \rangle.
\]

This quantity is referred to as the (squared) signal-to-noise ratio (SNR) in the literature. Concretely, \( \rho^2 \) is the maximum attainable SNR corresponding to the case where signal parameters are perfectly matched by a phase-evolution template [45].

III. LOUDEST CANDIDATES, THE “EFFECTIVE NUMBER OF TEMPLATES”, AND EXTREME VALUE THEORY

The standard problem of estimating the distribution of the loudest candidate in a search is posed as follows: Let \( \xi = \{\xi_i, i = 1, \ldots, \mathcal{N}\} \) be a set of detection statistic values obtained by evaluating a template bank with \( \mathcal{N} \) templates in a noise-only data stream. Let \( f \) be the probability distribution of such a detection statistic under the noise hypothesis. Describe the probability distribution of the loudest candidate \( \max_{i=1, \ldots, \mathcal{N}} \xi_i \).

We can easily construct said distribution using the joint cumulative density function (CDF) of the entire template bank

\[
P(\max_{i} \xi_i \leq \xi^*|\mathcal{N}) = P(\xi_1 \leq \xi^* \text{ and } \ldots \text{ and } \xi_\mathcal{N} \leq \xi^*) .
\]

For the case of an uncorrelated template bank, each template is independent and the joint CDF factors into the product of individual CDFs:

\[
P(\max_{i} \xi_i \leq \xi^*|\mathcal{N}) = \prod_{i=1}^\mathcal{N} P(\xi_i \leq \xi^*) = \left[ \int_{\xi^*} \text{d}\xi f(\xi) \right]^{\mathcal{N}} .
\]

Consequently, the probability density function associated to \( \xi^* = \max_{i=1, \ldots, \mathcal{N}} \xi_i \) is simply

\[
p(\xi^*|\mathcal{N}) = \mathcal{N} f(\xi^*) \left[ \int_{\xi^*} \text{d}\xi f(\xi) \right]^{\mathcal{N}-1} .
\]

Template-bank correlations imply that we sample fewer independent combinations of the data than with an uncorrelated bank of the same \( \mathcal{N} \). In other words, they reduce the “trials factor” of a search, diminishing the expected detection statistic of the loudest candidate in a similar fashion to evaluating a smaller template bank. Given a fixed false-alarm probability, neglecting template-bank correlations and naively using Eq. (19) would overestimate the corresponding threshold, potentially leading to missing interesting candidates.

Extensive analyses in [23, 58] concluded the effect of template-bank correlations on Eq. (19) could be reproduced to an acceptable level by adjusting \( \mathcal{N} \) to the “effective number of templates” \( \mathcal{N}' \) in the template bank at hand. Although in some cases an empirical estimate was possible [26], most applications obtained an effective number \( \mathcal{N}' \) via numerical fits to search results [24, 27, 28, 30–32]. Further studies on this topic [33], however, exposed a systematic discrepancy between the family of distributions spanned by Eq. (19) and the actual distributions obtained due to template-bank correlations.

An example of this discrepancy is illustrated in Fig. 1. We evaluated a template bank containing \( 2.23 \times 10^6 \) CW templates over frequency and spindown parameters \((f_0, f_1)\) with a realistic mismatch of \( m = 0.2 \) on 7 days of simulated Gaussian noise. We grouped the resulting \( 2F \)-statistic values into batches containing 223 templates each, from which the loudest \( 2F \)-statistic value was retrieved. The effect of parameter-space correlations was tested by either grouping templates within contiguous 5 mHz frequency bands or pooling an equivalent number of templates after randomly shuffling the results.

Shuffling the results before retrieving the loudest value tends to break any contribution from parameter-space correlations, as nearby templates are likely to end up in different batches. The resulting distribution can be properly fitted assuming an uncorrelated template bank. The apparent mismatch between the obtained effective
Estimating the loudest candidate’s distribution typically entails fitting an ansatz to a set of samples generated using a numerical procedure. As briefly
of a smaller template bank, then extrapolate such distribution to account for the template bank reduction. Dividing the initial template bank into smaller subsets makes multiple loudest candidates available to properly fit a distribution.

For the first step, one splits the dataset \( \xi \), containing \( \mathcal{N} \) (possibly correlated) values, into \( B \) batches, each of them with \( n = \mathcal{N}/B \) elements. This partition can be done such that each batch contains a similar subset of the overall population so that the per-batch maxima (batchmax samples) \( \{\xi_b^*\}, b = 1, \ldots, B \) are independent draws from the same unknown distribution: \( \xi_b^* \sim p_n \). If we choose a sufficiently high number of batches \( B \), the batchmax distribution \( p_n \) can be obtained by fitting a suitable ansatz to the data.

As the second step, the overall loudest value \( \xi^* \) is then

\[
\xi^* = \max_{b=1,\ldots,B} \xi_b^*,
\]

which corresponds to the loudest of \( B \) \( p_n \)-distributed random variables. This operation was already described in Eq. (19), which we here recast as an operator in the space of probability distributions for later convenience: given a probability distribution \( f \), \( \text{MaxProp}_B f \) corresponds to the distribution of the loudest candidate over a set of \( B \) independent samples of \( f \):

\[
\text{MaxProp}_B f(x) = B f(x) \left( \int_x^\infty d\xi^* f(x^*) \right)^{B-1}. \tag{22}
\]

The distribution of the overall loudest value \( \xi^* \) is then simply\(^2\)

\[
p(\xi^*|\mathcal{H}_N) = \text{MaxProp}_B p_n(\xi^*). \tag{23}
\]

The initial proposal in [25] described the batchmax distribution \( p_n \) using a Gaussian Kernel Density Estimation (KDE) over the set of batchmax samples \( \{\xi_b^*\} \). Eq. (23) was then implemented as a numerical integration. The final decision threshold was based on the support of the resulting distribution.

We find that the use of Gaussian KDEs introduces inaccuracies into the estimation of \( p(\xi^*|\mathcal{H}_N) \). The reason is twofold. First, KDEs are prone to overfitting histogram artifacts which arise due to finite sample sizes. This is illustrated in the upper panel of Fig. 3. As a result, the propagated distribution in this case displays an unintended bimodality, as shown in the lower panel of Fig. 3. Second, for the detection statistics we consider here, the batchmax distribution falls off exponentially (see Appendix A), at a much slower pace than a Gaussian tail. This tends to cause \( \text{MaxProp}_B \) to underestimate

---

2 Note that [23], inserting an empirical histogram as the \( f \) in Eq. (22) (to account for a bias due to implementation details in the \( F \)-statistic, see also Sec. V), constitutes an earlier application of this principle.
distribution \( p_n \). In the case of a data stream free of loud disturbances, this can be simply achieved by randomly shuffling the results of a search before grouping them into batches. (A discussion of the effects of shuffling data with loud disturbances is deferred to Secs. IV E and IVF and Appendix B.) The batchmax distribution then corresponds to that of the loudest candidate over \( n \) templates, which, as discussed in Sec. III, tends to a Gumbel distribution as \( n \to \infty \). Hence, we propose the following ansatz for the batchmax distribution:

\[
p_n(x) = \text{Gumbel}(x; \mu_n, \sigma_n) ,
\]

where \( \mu_n, \sigma_n \) are obtained by direct fit to the batchmax samples. This choice is similar to that of [87], which directly fitted an exponential tail (upper tail of a Gumbel distribution) to the batchmax distribution.

EVT distributions, such as Gumbel, are max-stable distributions: the distribution of the loudest outlier from a set of EVT distributions is itself an EVT distribution of the same kind, albeit with different parameter values. As a result, the MaxProp operator can be re-expressed in a closed form in terms of the location and scale parameters of the distribution. Concretely, it is straightforward to show that

\[
\text{MaxProp}_B \text{Gumbel}(x; \mu_n, \sigma_n) = \text{Gumbel}(x; \mu_s, \sigma_s)
\]

where

\[
\mu_s = \mu_n + \sigma_n \ln B ,
\]

\[
\sigma_s = \sigma_n .
\]

Thus, the target distribution is readily obtainable through a simple algebraic calculation after performing a fit to the batch-max samples:

\[
p(\xi^*|\mathcal{N}) = \text{Gumbel}(\xi^*; \mu_s, \sigma_s) .
\]

Summarizing, \textit{distromax} exploits the max-stability of the Gumbel distribution to estimate the distribution of the loudest candidate of a search, \( p(\xi^*|\mathcal{N}) \). To do so, search results are shuffled into \( B \) disjoint batches from which the loudest candidates are retrieved. These \( B \) batchmax candidates, by construction, can be interpreted as draws from an EVT distribution \( p_n \) whose parameters can be estimated using a standard maximum-likelihood fit such as \texttt{scipy.rv_continuous.fit} [77].

In broad terms, the batch size \( n \) determines how close the batchmax distribution is to an EVT one, whereas the number of batches \( B \) determines how sharply the parameters of \( p_n \) can be determined. Real searches usually contain a fixed number of templates \( \mathcal{N} = nB \), meaning a trade-off is required: On the one hand, choosing a large \( n \) (hence a small \( B \)) produces a small number of samples, each well consistent with an EVT distribution, but increases the variance of the \( p_n \) fit. On the other hand, a large \( B \) (hence a small \( n \)) produces a big number of samples drawn from a distribution which
has not fully converged to an EVT distribution, meaning the estimated parameters may be biased with respect to the actual distribution. We devote the following subsections to further discuss the role played by each of these parameters.

### C. The MaxProp operator

We now characterize the phenomenology of the MaxProp operator on a Gumbel distribution, which corresponds to the asymptotic distribution followed by standard (t)CW detection statistics.

Let us consider a template bank with \( \mathcal{N} = 10^6 \) templates. Given a batch size \( n \), we model a batchmax distribution as a Gumbel distribution with \( \sigma_n = 2 \) and \( \mu_n = \sigma_n \ln n \). This is equivalent to considering a detection statistic following a \( \chi^2_k \) distribution which has already converged to its corresponding EVT distribution, preventing finite sample-size effects from polluting the analysis. The ground truth distribution of the loudest candidate from the template bank then corresponds to the propagation of said Gumbel distribution over \( B = 10^B/n \) batches, i.e. a Gumbel distribution with \( \sigma = \sigma_n \) and \( \mu = \sigma \ln \mathcal{N} \).

Batch sizes \( B \in [1, 10^6] \) are analyzed by drawing \( B \) batchmax samples from the aforementioned batchmax distribution with \( n = 10^B/B \); \( \mu_n \) and \( \sigma_n \) are fitted using `scipy.stats.gumbel_r.fit` and propagated using Eq. \( (25) \). The resulting CDF is compared against the ground truth CDF, shown in Figure 4. The relative error in the estimated location and scale parameters is shown in Fig. 5.

As previously anticipated, a low number of batches \( B \lesssim 10^3 \) results in a greater dispersion of the estimated parameters. As the number of batches reach the \( 10^3 \lesssim B \lesssim 10^4 \) range, batchmax histograms become more robust and relative parameter deviations achieve sub-percent levels.

#### D. Characterizing the batchmax distribution

On the other hand, to test convergence of batchmax distributions, we take as an example the case of a \( \chi^2 \) distribution with 4 degrees of freedom (Gamma distribution with shape parameter \( k = 2 \) and scale parameter \( \theta = 2 \)), deferring to Appendix A other generic distributions and further references. For a \( \chi^2_k \) distribution, the limit \( n \to \infty \) corresponds to a Gumbel distribution [66], as shown in Fig. 6.

Given a set of random variables following a specific distribution \( f \), the convergence of the loudest draw towards an EVT distribution is driven by the behaviour of \( f \)’s tail (tail-equivalence [66]). More specifically, the role of the batch size \( n \) is related to how likely it is to draw a sample within the tail of the distribution: the higher the number of samples \( n \), the more likely it is to retrieve a value from the upper tail, hence the lower the dependency on other details of the distribution’s shape. Using a low batch size causes batchmax samples to be dominated by the bulk instead of the tail, keeping the resulting distribution from properly converging to an EVT distribution.

\( \chi^2 \) random variables are non-negative, as they are...
the sum of the squares of standard Gaussian random variables. Gumbel distributions, on the other hand, present a double exponential decay in their lower tail [Eq. (20)]. Consequently, as shown in the case $n = 1$ in Fig. 6, batchmax distributions with low $n$ tend not to follow a Gumbel distribution. As the number of samples $n$ increases, the effects due to the distribution’s bulk become milder and we find a better agreement to the expected distribution.

E. Parameter estimation accuracy and comparison to previous approaches

Finally, we present a more realistic set of results evaluating the $\mathcal{F}$-statistic over an actual template bank on 7 days of Gaussian noise using ComputeFstatistic_v2 [59]. The template bank is constructed using the gridType=8 option with maximum mismatch $m = 0.2$, for a fixed sky position $(0, 0)$ in equatorial coordinates around $f_0 = 50$ Hz and $f_1 = -10^{-8}$ Hz/s, containing $\mathcal{N} \approx 8 \times 10^6$ templates. A ground truth distribution is numerically constructed by evaluating this template bank on 900 realizations of Gaussian noise and retrieving the loudest $2\mathcal{F}$ value from each. distromax is then applied to the individual realizations in order to test its accuracy.

We compare two different batching approaches: batching contiguous frequency bins and shuffling the results into random batches. To produce comparable results, the shuffled batches contain the same number of templates $n$ as each contiguous batch. The motivation behind these two approaches is related to the potential presence of correlated outliers in real detector data: instrumental artifacts tend to affect relatively well-localized frequency bands [13]. Frequency-wise batching could thus prevent very loud outliers from polluting a high number of batches and overestimating the expected loudest outlier. Not shuffling the template bank, however, could require an increase in the batch size to obtain proper convergence to a Gumbel distribution, but then the reduced number of batches would imply an increase in the variance of the estimate.

Results are shown in Figs. 7, 8, and 9 in terms of the estimated mean and standard deviation from each method against those of the ground-truth distribution. We also compare to the original proposal of [25] by using a Gaussian KDE to approximate $p_\chi$ in Eq. (24).

We start by discussing the performance of the Gaussian KDE. The first significant feature is the lack in precision of the estimated parameters, which are over-dispersed regardless of the choice of $B$ and $n$. We also note that the bulk of these results tend to underestimate both the location and scale parameters of the Gumbel distribution with respect to the ground truth. This is related to the shape of the kernel function being used, as previously discussed in Sec. IV A: the tails of a Gaussian distribution fall off more rapidly than those of a Gumbel distribution,
FIG. 9. Comparison of methods to estimate the distribution of the loudest $F$-statistic outlier from a template bank using the same dataset as in Fig. 7. In this figure, the $N \approx 8 \times 10^6$ templates are grouped together by joining 100 consecutive frequency bins or shuffled into $B = 1801$ batches with a batch size of $n = 4419$.

FIG. 10. Summary of regimes in which distromax operates. Shaded regions represent combinations of $n$ and $B$ for which the distromax results suffer from low precision (high variance) due to a low number of batches (Sec. IVC) or low accuracy (high bias) due to a low batch size (Sec. IV D).

yielding a lower mean and standard deviation. The insensitivity of these results to the choice of $B$ and $n$ and to shuffling suggests that this particular KDE-based ansatz does not return a reliable estimate of the batchmax distribution.

distromax results, on the other hand, return a more consistent picture. Sample variance increases as the number of batches goes down both with and without shuffling. Complementarily, the bias in the estimated parameters reduces as $n$ increases ($B$ decreases), although this effect is only significant for the method without shuffling.

We observe a significant bias reduction by shuffling. Randomly shuffling samples results in more homogeneous batches with weaker inner correlations; as a result, batchmax samples are closer to the expected Gumbel distribution, improving the accuracy of the recovered parameters. Also, as previously anticipated, using bigger batches in the non-shuffling case does improve accuracy, although with a significant increment in variance due to the correspondingly lower number of batches.

These features are consistent with the basic phenomenology discussed in Secs. IVC and IV D. The different regimes in which distromax operates depending on $N$, $n$, and $B$ are summarized in Fig. 10. These values are extracted from the general behaviour of distromax observed throughout the tests performed in this section. As a general working principle, distromax requires at least $N \approx 10^6$ in order to return a cogent answer; for smaller template banks, on the other hand, off-sourcing as described in Sec. III requires little computational effort.

F. Discussion

The results above on simulated data demonstrate the overall performance of distromax under Gaussian-noise conditions. The concrete output of distromax is a simple
estimation on the probability of the loudest candidate of a search under the noise hypothesis \( p(\xi^*|H_N) \). However, the specific statement to be drawn from \( p(\xi^*|H_N) \), such as a threshold choice, is entirely dependent upon the scope of the analysis at hand. For the sake of completeness, we briefly review the assumptions on which the \texttt{distromax} method relies, as well as possible consequences of violating them.

First, the distribution of the detection statistic at hand must belong to the domain of attraction of the Gumbel distribution; roughly, this means its probability distribution should be unbounded and decay at a slower rate than a power-law \([66]\). Nonetheless, as discussed in Appendix A, this method could be easily adapted to detection statistics within a different domain of attraction; in such cases, Eqs. (26) and (27) would have to be adapted to the corresponding EVT distribution.

Second, the data at hand must be free of strong disturbances. As discussed in Sec. II B, loud disturbances in a data stream typically translate into parameter-space regions returning enhanced detection statistics with respect to a non-disturbed data stream. The width of the affected region will depend on the characteristics of the disturbance and the template bank, but in general there can be an extended set of templates with correlated response to the disturbance. Attempting to construct a batchmax distribution by shuffling the samples into different batches would result in a distribution shifted towards the right-hand side of the expected Gumbel distribution. Not using shuffling would suppress the effect of disturbances if the resulting associated population of templates was well localized in frequency; the resulting distribution, however, would be less accurate and in particular could still overestimate the Gumbel parameters, as discussed in Figs. 7, 8, and 9. The robustness of \texttt{distromax} results to the effect of mild disturbances on the data can also be tested by generating several sets of batchmax samples and comparing the location and scale parameters of the corresponding Gumbel distributions. The wider the distribution over shuffling realizations, the bigger the effect of noise disturbances.

Narrow spectral features (“lines”), in particular, are common noise disturbances affecting \((t)CW\) searches, with excess power typically concentrated within a few frequency bins \([13]\). \(CW\) signals themselves are another typical example of well-localized “disturbances”: should a (strong) \(CW\) signal be present in a datastream, a blind application of \texttt{distromax} could result in overestimation of the loudest candidate’s distribution, potentially flagging the \(CW\) signal itself as a background-noise fluctuation. The sensitivity of current interferometric detectors, however, makes \(CW\) searches to operate in the weak-signal regime (see Appendix C). As a result, \(CW\) signals are unlikely to actually affect the estimations provided by \texttt{distromax}.

Due to their crucial role in \(CW\) searches, we discuss in Appendix B a simple proposal to reduce the effect of narrow-band disturbances so that \texttt{distromax} results can still provide a cogent answer. Whenever possible, we recommend the application of informed veto strategies against instrumental artifacts (see \[16\] and references therein) \textit{before} attempting to process the results using \texttt{distromax}. The method discussed in Appendix B is just a complementary algorithm to prevent a specific type of strong disturbances from invalidating an analysis. The characterization and improvement of this or similar algorithms to deal with more generic disturbances is left for future work.

Third, in principle \texttt{distromax} assumes the complete set of detection statistic values from the full template bank is available. Several wide parameter-space searches, however, use \textit{toplists} \([78, 88, 89]\), meaning they only keep a small fraction of detection statistic values corresponding to the louder templates. The basic requirement is to use a toplist such that the tail of the distribution is properly represented. Falling short (i.e. not reaching the bulk of the distribution) could result in inaccurate fits to batchmax samples. Incidentally, the results discussed in Appendix B clarify the suitability of \texttt{distromax} to these searches.

Fourth, and closely related to the second point, template banks must not be too strongly correlated. EVT ensures \texttt{distromax} is robust to a certain degree of template-bank correlations; more specifically, answers provided by \texttt{distromax} will be cogent as long as the dominant contribution to batchmax samples comes from the tails of the involved distributions. As discussed in e.g. \([48]\), template bank setups using a small mismatch return highly correlated samples in the vicinity of local parameter-space maxima, adding additional features to the results distribution (see e.g. Fig. 2 in \[33\]). This may have similar effects to the presence of loud disturbances, as a significant fraction of the resulting batchmax samples would come from samples around a few local maxima and not be representative of the tail of the background distribution. Consequently, the corresponding batchmax distribution will not be fully converged to an EVT distribution and the final Gumbel parameter estimation will be affected. No simple amendment, other than using a higher mismatch \([10, 53]\), is currently available to obtain robust results with \texttt{distromax} in this situation.

This phenomenon was observed in \([90, 91]\), where \texttt{distromax} was applied to process the result of both a search for \(CW\) signals and a search for long-duration transient GWs from glitching pulsars. The \(CW\) search used a mismatch of \(m \simeq 0.02\) combined with a toplist, which reduced the effect of such a dense parameter-space converging on the batchmax distribution. Preliminary studies for the long-duration transient GW search using \(m = 0.02\) \textit{without a toplist}, on the other hand, revealed a poor performance of \texttt{distromax}; in that case, the solution was to increase the mismatch to \(m = 0.2\), using a similar setup as in \[32\].

As we will discuss during Sec. V, \texttt{distromax} is suitable to be applied in real-data searches with typical mismatch setups. Moreover, further detection statistics beyond \(2F\),
such as “line-robust” statistics [35, 36] or tCW search statistics [37], can be processed using the same method. This has the effect of improving the quality of distromax results, as these statistics are designed to diminish the effect of noise disturbances on (t)CW searches, providing a cleaner set of batchmax samples.

V. APPLICATION TO O2 DATA

As a demonstration, we apply distromax to data of the two advanced LIGO detectors [18] from the O2 observing run [92] in Hanford and Livingston. We study the statistics of results obtained in [32] for narrow-band searches targeting the Vela and Crab pulsars, which experienced glitches on 12th December 2016 and on 27th March 2017 respectively. The template bank for each detector was not constructed with a fixed in the transient parameter space. Hence, we can directly estimate thresholds from the samples allowing for end-to-end analysis using $B_{S/GL}$.

A. CW detection statistics

We first consider results for CWs of duration four months corresponding to the maximum observation time in [32]. We begin with the standard $2F$ as its distribution is well known. As we see from Fig. 11, the histogram of the full $2F$ results matches well with a standard $\chi^2$ distribution. One can also fit Eq. (19) treating $N$ as a free parameter, obtaining $N' \approx 1$. This is equivalent to considering each $2F$ sample as the trivial maximum of a single draw from a $\chi^2$ distribution. This $\chi^2$ distribution can then be propagated using the total number of templates $N_3$ as $B$ in Eq. (22), yielding an estimated distribution for the maximum of $2F$, which assumes that the template bank correlations are negligible. We then compare the resulting distribution with the one obtained by the distromax method in Fig. 11. The two resulting distributions for the maximum agree well.

We also apply the distromax method on a different statistic for the CW search, namely the line-robust statistic $B_{S/GL}$ [35]. This is a Bayes factor derived from the likelihood ratio between the signal hypothesis and the combined hypothesis noise hypothesis of Gaussian noise and lines. The lines are modelled based on the assumption that they look exactly line a signal, but are present in only one detector. The results are shown in Fig. 12. Since the underlying distribution of $B_{S/GL}$ is unknown, one cannot do the equivalent of fitting Eq. (19). Nevertheless, we can still apply the distromax method, for which the only constraint is that said distribution falls off faster than a power-law (for the case here discussed involving a Gumbel distribution). The details of the exact distribution are not needed. Indeed, the batchmax distribution is well-fitted by a Gumbel distribution. In previous studies using $B_{S/GL}$ on real data [24, 29–31, 94–97], it was used only to improve the robustness of the search against disturbances by using it as the toplist ranking statistic, but the final significance statements were made returning to $2F$ because no closed-form distribution was known for $B_{S/GL}$. Now with distromax we can directly estimate thresholds from the samples allowing for end-to-end analysis using $B_{S/GL}$.

B. tCW detection statistics

We now investigate the case of tCWs, which was the main focus of [32]. As briefly mentioned in Sec. II, such signals can be modelled as CWs modulated by a window function dependent on the transient parameters $T$, namely the start time of the transient signal $t_0$ and its duration $\tau$. For this analysis a rectangular window function was used. As a detection statistic for tCWs, the $F$-statistic at fixed $\lambda$ can be maximized over transient parameters [37], thus obtaining $2F_{\text{max}} = \max_T 2F$. (We use this notation instead of simply $\max 2F$ to avoid confusion with the maximum CW detection statistic $2F$ over a full template bank.) We expect local correlations to have a more severe impact when using this statistic because it can pick up short-duration non-Gaussianities or simple fluctuations that the CW $F$-statistic would not be susceptible to.

When the distromax package shuffles the dataset in the batchmax stage, several batches can be contaminated by the same noise fluctuation, and therefore the batchmax distribution reflects this contamination. The result is a more ragged distribution with peak-like features, as one can see in Fig. 13.

To estimate the distribution of the loudest candidate for this dataset, [32] made several simplifying assumptions. While there is no known distribution for $2F_{\text{max}}$ that could be directly inserted into Eq. (19), its value at each template $\lambda$ is the maximum of $2F$ values over the transient parameters $T$, which individually follow a $\chi^2$ distribution. However, there is a high degree of correlations in the transient parameter space. Hence, fitting Eq. (19) to the $2F_{\text{max}}$ samples, the result is an “effective number of transient templates” $N' \approx 55$ (compared to a nominal number of $N_T \approx 2 \times 10^6$.

---

3 As discussed in detail in Sec. 8.7.1 of [23], the result of any fit of Eq. (19) to $F$-statistic-based detection statistic samples cannot be directly interpreted as an “effective number of templates” even in the absence of template-bank correlations. This is due to a small upwards implementation bias in the $F$-statistic computation [45, 93]. For the $N' \approx 1$ fit to the $F$-statistic samples, this seems to approximately cancel with the effect of template bank correlations.
transient templates at each $\lambda$).

4 It was then assumed that $N_\lambda \times N'$ could be interpreted as an “effective number of templates” over the full, non-maximized, parameter space ($\lambda, T$). Consequently, the distribution of the overall loudest was obtained by propagating a $\chi_4^2$ distribution using $B = N_\lambda \times N'$ in Eq. (22).

The fits to the $2F_{\text{max}}$ sample histograms approximately catch the peak of the distribution, but fail to correctly recover the overall shape. On the other hand, with distromax, the Gumbel fits to the batchmax samples are noticeably better aligned to both the peaks and the tails of the batchmax histograms than the $N'$ based fits are to the full samples histograms. The propagated distributions from both methods still overlap, but their differences are larger than in the CW case.

Again, we apply distromax also on an alternative detection statistic for the tCW search, namely the $B_{S/G}$ statistic, also derived in [37]. This statistic does not deal with the transient parameters by maximizing over the $T$ space, but rather marginalizes over it using a uniform prior. This results in less contamination from disturbances and noise fluctuations. Despite its known better detection efficiency [37], one reason why this detection statistic has not been used in [32] is that its distribution is not analytically known, hence no simple fit of Eq. (19) could be done. With the distromax method [38], this is no longer a problem. The results are shown in Fig. 14. Using the same data sets as before, the batchmax histograms are much smoother and better fit...
FIG. 13. Values of $2F_{\text{max}}$ obtained in the tCW analysis of [32] on LIGO O2 data after glitches in the Vela (left panel) and Crab (right panel) pulsars. The gray dotted line is the histogram of the samples obtained by the search. There is no known distribution for $2F_{\text{max}}$, but it is a maximum over $F$-statistics over the transient parameters which individually follow a $\chi^2_4$ distribution. Due to the high degree of correlations in the transient parameter space, in [32] the “effective number of transient templates” was obtained by fitting Eq. (19) to the $2F_{\text{max}}$ samples, obtaining $N' \approx 55$ (dotted blue line). The distribution of the overall loudest was then obtained by propagating the $\chi^2_4$ distribution using $B = N_\lambda \times N'$ in Eq. (22) (dash-dotted blue line). This corresponds to treating each sample as a batch with a single element. On the other hand using distromax we plot the batchmax histogram ($n = 1000$, black solid line) and we fit it with a Gumbel distribution (dashed red line). The propagated distribution is obtained by applying Eq. (25) with $B = N_\lambda/n$ (solid red line).

FIG. 14. Values of $\log_{10} B_{S/G}$ obtained in the same [32] tCW analysis as in Fig. 13. Vela results are on the left panel, Crab results are on the right panel. Here there is no direct fit of Eq. (19) because no closed-form expression for the distribution of $B_{S/G}$ is known, and it cannot be easily related to the original $F$-statistic, so only the search samples (dotted gray line), the batchmax histogram (black solid line), the Gumbel fit to it (dashed red line), and the propagated final distribution (solid red line) from distromax are shown.

by a Gumbel distribution than their $2F_{\text{max}}$ counterparts. This indicates that $B_{S/G}$ is a more robust detection statistic than $2F_{\text{max}}$ on real data. As for $B_{S/GL}$ for CWs, with distromax it can now also be used as an end-to-end detection statistic.

VI. CONCLUSION

We have introduced distromax, a new method to estimate the distribution of the loudest candidate in a gravitational-wave search. This method culminates a series of developments in the continuous gravitational-wave literature aimed at re-cycling wide parameter-space search results into a proxy distribution for the expectation over different background noise realizations. An implementation of the method is freely available as a Python package [38].

Our specific proposal uses max-stable distributions from extreme value theory to provide a generic approach, applicable to any detection statistic displaying a light-tailed distribution under the noise hypothesis (that is, unbounded and decaying faster than a power-law). This is in contrast with previous approaches based on the $F$-statistic, whose very specific assumptions prevented a successful generalization.

Although we have focused on the case of detection statistics with light-tailed distributions, as that is the
standard encountered in CW searches, extensions to other kinds of distributions are possible by using a different family of max-stable distributions.

We have demonstrated the general applicability of distromax using both synthetic Gaussian-noise data and the results of a real search on Advanced LIGO O2 data for (transient) continuous gravitational-wave signals from the Vela and Crab pulsars. Results show a significant improvement with respect to previous estimation methods due to the robustness of distromax to realistic template-bank correlations.

Additionally, the possibility of using further detection statistics suppressing the effect of lines ($B_{s/GL}$, $B_{s/GLAL}$) or transient instrumental artifacts ($B_{IS/G}$) presents two further advantages for (transient) continuous gravitational-wave searches: first, distromax allows us to process the results directly in terms of these more informative statistics; second, the built-in suppression of instrumental features in these statistics itself improves the convergence of batchmax samples to a max-stable distribution, improving the quality of the results provided by distromax. This last point also makes plots of the batchmax distribution a useful tool to diagnose the data quality of a specific frequency band using its deviation with respect to the expected max-stable distribution.

ACKNOWLEDGEMENTS

We gratefully thank Patrick M. Meyers for testing and discussing improvements to the distromax package, Karl Wette for sharing code for the KDE version of estimating loudest-candidate distributions and comments on the manuscript, John T. Whelan for valuable comments on the manuscript, and the LIGO–Virgo–KAGRA Continuous Waves working group for fruitful discussions. This work was supported by European Union FEDER funds, the Spanish Ministerio de Ciencia e Innovación, and the Spanish Agencia Estatal de Investigación grants PID2019-106416GB-I00/AEI/MCIN/10.13039/501100011033, RED2018-102661-T, RED2018-102573-E, Comunitat Autònoma de les Illes Balears through the Conselleria de Fons Europeus, Universitat i Cultura and the Direcció General de Política Universitaria i Recerca with funds from the Tourist Stay Tax Law ITS 2017-006 (PRD2018/24, PRD2020/11), Generalitat Valenciana (PROMETEO/2019/071), EU COST Actions CA18108, CA17137, CA16214, and CA16104. R. T. is supported by the Spanish Ministerio de Universidades (ref. FPU 18/00694). L. M. is supported by the Universitat de les Illes Balears. D. K. is supported by the Spanish Ministerio de Ciencia, Innovación y Universidades (ref. BEAGAL 18/00148) and cofinanced by the Universitat de les Illes Balears. This research has made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center (https://www.gw-openscience.org/), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459. This paper has been assigned document number LIGO-P2100277.

Appendix A: Basic results of extreme value theory

Let us consider a set of $n$ independent and identically distributed random variables $\{x_1, \ldots, x_n\}$ each following a probability distribution $f$. These variables can be identified with a detection statistic evaluated on a set of parameter-space templates, with $f$ corresponding to the detection statistic’s distribution under the noise hypothesis. We are interested in describing the probability distribution of the highest detection statistic value (usually referred to as the largest order statistic $[62] x^* = \max\{x_1, \ldots, x_n\}$, denoted as $f^*$, in order to evaluate the significance of outliers resulting from a CW search.

A first ansatz for $f^*$ can be constructed by considering the probability of drawing $x_i = x^*$ for a single $i \in [1, n]$ and $x_i < x^*$ for the remaining $n - 1$ values, taking into account all possible sortings:

$$
 f^*_n(x) = n \int_0^x f(x') \left(\int_0^{x'} f(x') \, dx''\right)^{n-2} \, dx'.
$$

(An alternative derivation of this results is presented in Sec. III.) This approach is sufficient if the probability distribution $f$ is well understood and different random variables $x_i$ are independent from one another so that the joint distribution factors into the product of individual distributions. In the case of searches on real data, however, parameter-space correlations cause $f^*_n$ to deviate from Eq. (A1) [33, 34, 58].

Extreme Value Theory (EVT) provides asymptotic closed forms for Eq. (A1) in the limit of $n \to \infty$

$$
 f^*_n \xrightarrow{n \to \infty} \text{GEV}(\gamma),
$$

where GEV refers to the generalized extreme value distribution and $\gamma \in \mathbb{R}$ is referred to as the extreme value
index [62, 63, 65, 66]. According to the specific properties of the random variables at hand, GEV distributions can be shifted and rescaled by location and scale parameters, $\mu_n$ and $\sigma_n$. The general dependency on the number of random variables being drawn $n$ is due to the increased chances of drawing an extreme value as the number of independent trials increases. This is usually referred to as the trials factor.

In a practical case, assuming $n$ so that the convergence is suitable for the application at hand, Eq. (A3) can be recast into a closed form

$$f^*_n(x) = \text{GEV}(\mu_n + x \sigma_n; \gamma). \quad (A3)$$

Typical prescriptions for $\mu_n$ and $\sigma_n$ [67, 68] tend to be valid for $n \gtrsim 10^4$. The GEV distribution has the specific property of being max-stable, meaning that the distribution of the maximum sample out of $n$ draws from a GEV distribution is again a GEV of the same kind (same $\gamma$).

The value of $\gamma$ depends on the right-hand tail behaviour of $f$, and determines the functional form of $f^*$ out of three possibilities. We follow the definitions given in [66] (where $\gamma$ is referred to as $\xi$): finite tails with power-law behaviour correspond to $\gamma < 0$ (Weibull distribution), light tails correspond to $\gamma = 0$ (Gumbel distribution), and power-law tails correspond to $\gamma > 0$ (Fréchet distribution). Fig. 15 illustrates

5 The scipy Python package [77] implements these three distributions under the stats module, although it uses a different
the convergence towards each of these families using paradigmatic probability distributions, namely a uniform distribution in \([0, 1]\), a standard exponential distribution, and a standard Cauchy distribution.

We focus our attention on the location and scale parameters, as they are relevant in terms of convergence speed. EVT imposes very loose conditions on them, so the choice of \(\mu_n\) and \(\sigma_n\) as functions of \(n\) is not unique for a given distribution, and the main difference across different choices is the speed with which the resulting distribution will approach the GEV one. We illustrate this using a Gaussian distribution, which is in the domain of attraction of the Gumbel distribution and is famous for being quite slow to converge. Figure 16 compares the prescription of location and scale parameters originally proposed by Hall [98] to the improvement proposed by Gasull [67].

However, in this paper we are mainly interested in \(\chi_k^2\) distributions, where \(k \in \mathbb{N}\) denotes the degrees of freedom of the distribution, as CW statistics are quite frequently constructed as the norm of a Gaussian vector and hence follow \(\chi_k^2\) distributions. A significant improvement over the classical literature was presented in [68], where closed expressions for \(\mu\) and \(\sigma\) for a generic \(\Gamma\) distribution were
obtained. Figure 17 shows the convergence of different \( \chi^2 \) distributions towards a Gumbel distribution.

We provide an implementation of the corresponding expressions discussed in [68] within the \texttt{distromax} Python package [38]: Although the \texttt{distromax} method itself does not use any of these results (since \( \mu \) and \( \sigma \) are estimated from the data), they can still be used to produce theoretical estimates.

**Appendix B: Addressing disturbed data**

The intended output of \texttt{distromax} is an empirical estimation of the distribution of the loudest candidate produced by noise-only data in a CW search \( p(\xi^* | H_N) \). To do so, the basic assumption is that the output of a search \( \xi \) corresponds mostly to samples of the detection statistic from a single well-behaved distribution. In practice, this generally means Gaussian noise, plus only a small number of samples coming from another population such as a non-Gaussianity in the data or a CW signal. If the number of such samples is negligible compared to the number of batches used in \texttt{distromax}, so would be their effect in the batchmax distribution.

CW searches in real data, however, are populated by various kinds of noise disturbances. Concretely, a prominent type are narrow-band instrumental features (“lines”), which tend to concentrate their effect within a few frequency bins, but especially for higher-dimensional searches (several spin-down terms and/or all-sky searches) can still affect a large number of frequency bins, for which the ad hoc approach of excising or \textit{notching} the disturbed frequency band returns a similar result to a proper Bayesian analysis.

In this Appendix, however, we concentrate on the typical case of loud noise disturbances polluting a small number of frequency bins, for which the ad hoc approach of excising or \textit{notching} the disturbed frequency band returns a similar result to a proper Bayesian analysis. A full treatment of the mixture model problem is left for future work.

As previously discussed in Sec. IV, this approach is conceived to deal with a specific set of common noise disturbances so that \texttt{distromax} can be applied on a larger range of real-data results. However, users are encouraged to understand and curate their search results using standard CW vetoes (see [16] and references therein) before falling back to this specific notching algorithm.

Leveraging thresholding algorithms from the image-processing literature [100] we propose a simple algorithm capable of notching frequency bands containing prominent disturbances. Since we focus on noise disturbances within well-localized frequency bands, we attempt to flag their corresponding candidates focusing on the loudest detection statistic in each frequency bin \( f_0 \)

\[
\xi^*(f_0) = \max_k \xi(f_0, \lambda),
\]

where \( \lambda \) contains any other relevant parameter-space dimension. The resulting envelope can be thought of as a one-dimensional gray-scale image in which we are interested to discern the background from an object (the polluted band); the distinction is made by properly selecting a gray-value (detection-statistic) threshold such that object pixels (polluted-band samples) lie above it, leaving nothing but background below.

We illustrate the effects of our notching algorithm using a synthetic template bank containing a narrow-band disturbance. The template bank contains 1000 \( \times \) 1000 templates spanning the \( (f_0, f_1) \) parameter space over the [100,100.1] Hz frequency band. The corresponding \( 2F \) is drawn from a \( \chi^2 \) distribution for each template in the bank. We refer to this \( \chi^2 \)-drawn set of samples as the ground truth. An outlier is introduced by replacing samples in the [100.04, 100.05] Hz sub-band with an equal amount of samples drawn from a non-central \( \chi^2 \) distribution with non-centrality parameter \( \rho^2 = 25 \). The sample projection over the \( f_0 \) subspace is shown in Fig. 18.
We tested different thresholding techniques, including standard approaches such as the Otsu threshold [101], the minimum cross-entropy threshold [102, 103], and the minimum method threshold [104], using the implementations available in the skimage package [105]. We find the minimum method threshold skimage.filters.threshold_minimum performs best in our specific study, noting that the implementation of the notching procedure in [38] allows for a flexible selection of thresholding strategies.

Once an appropriate threshold $\xi_T$ has been established, we proceed to notch any frequency bin containing at least one sample above threshold. Specifically, we remove all the samples from the frequency bins $f_0$ where $\xi^*(f_0) > \xi_T$. This step can be applied multiple times in order to take care of multiple lines in a band with very different amplitudes, “shoulders” of broad lines, or features such as spectral leakage. The specific implementation provided in [38] implements a simple stopping criterion: notching iterations stop whenever the threshold $\xi_T$ falls below a pre-specified quantile of $\xi^*(f_0)$. The default value, which performs well for our specific example, stops whenever $\xi_T$ is lower than the top 20% values of $\xi^*(f_0)$.

The result of notching, as opposed to simply removing samples over threshold, is illustrated in Fig. 19, where the distribution being notched is shown as a blue histogram. Simply removing samples above threshold would be equivalent to cutting the tail of the histogram while leaving the bulk untouched. While such an approach would be relatively harmless in the case of a disturbance strong enough to be cleanly separated from the background, it is rendered ineffective in the case of a relatively mild disturbance, as the polluted-band samples tend to overpopulate the tail of the distribution itself. Completely notching the band, on the other hand (red histogram), properly deals with the overpopulation of outliers and returns a distribution consistent with an undisturbed background.

After notching disturbed bands, we can simply apply the shuffling and batching procedure described in Sec. IVB. The resulting batchmax distributions, including the unnotched and ground truth distributions, are shown in Fig. 20. Location and scale parameters of the best fitting Gumbel distributions are shown in the legend and compared to the ground truth distribution in Fig. 21.

In this example, the estimated parameters using notching show only 1% relative difference with respect to ground truth parameters, as opposed to the strong bias suffered by the unnotched estimates. Moreover, we remark the robustness of the method to a mild overnotching of non-polluted frequency bins: the convergence to a Gumbel distribution is mainly related to the properties of the “bulk of the tail”; trimming the most extreme events from the background distribution does not affect significantly the fitting of a Gumbel distribution. This is clearly seen in Fig. 19, where the notched distribution differs from the ground truth by a few samples. These samples correspond to the background samples over threshold in Fig. 18, which belong to the tail of the non-disturbed distribution. As briefly commented in Sec. IVF, this result justifies the extent up to which distromax is applicable to toplist-based searches: so long as the toplist reaches the bulk of the distribution, distromax should be capable of returning a cogent answer.

The application of more notching iterations than strictly required, however, could result in an underestimation of the Gumbel parameters due to the removal of too many samples in the tail of the distribution. This is particularly important for the scale parameter: underestimations around 5% are often obtained across several realizations of the example setup.
discussed here when using the notching procedure with the stopping criterion as described above. For a typical batch size of $B \approx 10^3 - 10^4$ (ln $B \approx 10$) and fiducial values of $\mu_* = 50$ and $\sigma_* = 2$ [Eq. (26)] (similar to the values encountered in a CW search using the $2F$ statistic), a 5% underestimation in $\sigma_*$ implies about 20% of underestimation in $\mu_*$. The main consequence of this is a shift of the resulting $p(\xi^*|H_N)$ towards lower values, potentially resulting in an increased number of candidates scoring over the specified threshold.

**Appendix C: Robustness against injections**

As discussed in Sec. IVF, (t)CW signals themselves can be considered as disturbances when trying to estimate a background distribution. We therefore need to test that the distromax method is robust to the presence of (t)CW signals and will not be biased upwards, which would lead to picking too high a threshold and missing those signal-related candidates. Such a bias is not entirely avoidable, but should ideally only appear for signals stronger than expected in any practical real-data search situation.

To test this, we re-use the upper limits injections in O2 data from the same analysis [32] as in Sec. V. Simulated signals of increasing amplitude $h_0$ were added to the original short Fourier Transform (SFT) data (“injections”). For each chosen $h_0$, there are 50 data sets with parameters $\{f_0, f_1, t_0\}$ uniformly distributed over their respective search ranges (see [32]) and the remaining amplitude parameters $\{\cos \psi, \phi_0\}$ randomized over their natural ranges. ComputeFstatistic_v2 [59] was used to reanalyze a small range around the injection point and we combine these results with the original detection statistic samples for the rest of the search band. We then apply distromax as in Sec. V, checking the resulting propagated distribution for the maximum of each detection statistic. We also test the notching procedure introduced in Appendix B by gradually increasing the num_iterations option, from 0 to 5. The means and standard deviations of the propagated distributions for the four statistics discussed in Sec. V are shown in Fig. 22, Fig. 23, Fig. 24 and Fig. 25.

As one increases the amplitude $h_0$ of the injected signal, an increasing number of templates will produce elevated values of the detection statistic. In Sec. IVF, we found that shuffling of batches in the batchmax step is generally preferred. However, with a strong signal present that affects multiple templates, the shuffled batchmax distribution will inevitably become contaminated, leading to an overestimation of the final distribution parameters. Such a trend is indeed visible in the results for all detection statistics, more clearly for the Vela analyses (left-side panels). However, the effect is small compared to the actual increase of the detection statistic at the templates with injections. For the highest amplitudes tested, the detection statistic can reach values $\sim 20$ times above the expected loudest background sample. Furthermore, the upwards shift in the estimated distribution is generally mild, with the mean shifting by less than one standard deviation of the original injection-free mean, at least as long as $h_0$ does not reach significantly above the 90% upper limits set in [32].

In addition, the notching feature can be useful in limiting the rise of the estimated means in the presence of signals with large $h_0$, by treating the templates with elevated detection statistic as disturbances for the purpose of background estimation and removing them before applying the batchmax procedure. Generally 1–2 iterations of notching have little influence on the estimated Gumbel distribution mean of the original data without injections while helping to reduce the rise of the estimated distribution mean with injection $h_0$. With more iterations of notching, results become more robust towards strong injections, while for the original data the means in some cases are estimated lower. This would always be conservative in the sense that one would retain more candidates for follow-up even when “over-notching” clean data, however it would lead to additional human and computing effort to follow up candidates that are clearly noise fluctuations.

Hence, as already discussed in Appendix B, we recommend using the notching feature only if required. If data is clean – i.e. no unusual features in the batchmax histograms – notching is not necessary, and more likely leads to underestimated distribution parameters.
which would correspond to an overly conservative threshold choice. If, on the other hand, data exhibits strong and numerous spikes in the statistic, the plain distromax method may lead to overestimated distribution parameters, and notching can be a useful tool in such situations. Regarding the potential presence of (t)CW signals in the data, distromax seems robust to these, with or even without notching, for the typical target signals of current (t)CW searches (not standing out far above the noise background); and if in doubt, notching can still help to provide more conservative thresholds.

[1] E. T. Jaynes, *Probability Theory: The Logic of Science*, edited by G. L. Bretthorst (Cambridge University Press, 2003).
[2] B. J. Owen, Search templates for gravitational waves from inspiraling binaries: Choice of template spacing, *Phys. Rev. D* 53, 6749 (1996), arXiv:gr-qc/9511032.
[3] P. R. Brady and T. Creighton, Searching for periodic sources with LIGO. 2. Hierarchical searches, *Phys. Rev. D* 61, 082001 (2000), arXiv:gr-qc/9812014.
[4] B. S. Sathyaprakash and S. V. Dhurandhar, Choice of filters for the detection of gravitational waves from coalescing binaries, *Phys. Rev. D* 44, 3819 (1991).
[5] B. J. Owen and B. S. Sathyaprakash, Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement, *Phys. Rev. D* 60, 022002 (1999), arXiv:gr-qc/9808076.
[6] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries, *Phys. Rev. D* 85, 122006 (2012), arXiv:gr-qc/0509116.
[7] R. Prix, Template-based searches for gravitational waves: Efficient lattice covering of flat parameter spaces, *Class. Quant. Grav.* 24, S481 (2007), arXiv:0707.0428 [gr-qc].
[8] I. W. Harry, B. Allen, and B. S. Sathyaprakash, A Stochastic template placement algorithm for gravitational wave data analysis, *Phys. Rev. D* 80, 104014 (2009), arXiv:0908.2090 [gr-qc].
[9] C. Messenger, R. Prix, and M. A. Papa, Random template banks and relaxed lattice coverings, *Phys. Rev. D* 79, 104017 (2009), arXiv:0809.5223 [gr-qc].
[10] B. Allen, Optimal template banks, *Phys. Rev. D* 104, 042005 (2021), arXiv:2102.11254 [astro-ph.IM].
[11] K. J. Wagner, J. T. Whelan, J. K. Wofford, and K. Wette, Template Lattices for a Cross-Correlation Search for Gravitational Waves from Scorpius X-1, arXiv e-print (2021), arXiv:2106.10142 [gr-qc].
FIG. 23. Results of testing distromax on $\log_{10} B_{S/GL}$ values from simulated signal injections for the Vela (left) and Crab (right) O2 search parameter spaces, with all details as in Fig. 22.

FIG. 24. Results of testing distromax on $2F_{\text{max}}$ values from simulated signal injections for the Vela (left) and Crab (right) O2 search parameter spaces, with all details as in Fig. 22.

[12] R. Prix and B. Krishnan, Targeted search for continuous gravitational waves: Bayesian versus maximum-likelihood statistics, Class. Quant. Grav. 26, 204013 (2009), arXiv:0907.2569 [gr-qc].

[13] P. B. Covas et al., Identification and mitigation of narrow spectral artifacts that degrade searches for persistent gravitational waves in the first two observing runs of Advanced LIGO, Phys. Rev. D 97, 082002 (2018), arXiv:1801.07204 [astro-ph.IM].

[14] D. Davis, L. V. White, and P. R. Saulson, Utilizing aLIGO Glitch Classifications to Validate Gravitational-Wave Candidates, Class. Quant. Grav. 37, 145001 (2020), arXiv:2002.09429 [gr-qc].

[15] B. P. Abbott et al. (LIGO Scientific, Virgo),
Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. 116, 061102 (2016), arXiv:1602.03837 [gr-qc].
[16] R. Tenorio, D. Keitel, and A. M. Sintes, Search Methods for Continuous Gravitational-Wave Signals from Unknown Sources in the Advanced-Detector Era, Universe 7, 474 (2021), arXiv:2111.12575 [gr-qc].
[17] M. Sieniawska and M. Bejger, Continuous gravitational waves from neutron stars: current status and prospects, Universe 5, 217 (2019), arXiv:1909.12600 [astro-ph.HE].
[18] J. Aasi et al., Advanced LIGO, Class. Quant. Grav. 32, 074001 (2015), arXiv:1411.4547 [gr-qc].
[19] F. Acernese et al., Advanced Virgo: a second-generation interferometric gravitational wave detector, Class. Quant. Grav. 32, 024001 (2014), arXiv:1408.3978 [gr-qc].
[20] T. Akutsu et al. (KAGRA), KAGRA: 2.5 Generation Interferometric Gravitational Wave Detector, Nat. Astron. 3, 35 (2019), arXiv:1811.08079 [gr-qc].
[21] P. Jaranowski, A. Królak, and B. F. Schutz, Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection, Phys. Rev. D 58, 063001 (1998), arXiv:gr-qc/9804014.
[22] C. Cutler and B. F. Schutz, Generalized \( F \)-statistic: Multiple detectors and multiple gravitational wave pulsars, Phys. Rev. D 72, 063006 (2005), arXiv:gr-qc/0504011.
[23] K. W. Wette, Gravitational waves from accreting neutron stars and Cassiopeia A, Ph.D. thesis, Australian Natl. U., Canberra (2009).
[24] M. A. Papa, J. Ming, E. V. Gotthelf, B. Allen, R. Prix, V. Dergachev, H.-B. Egggenstein, A. Singh, and S. J. Zhu, Search for Continuous Gravitational Waves from the Central Compact Objects in Supernova Remnants Cassiopeia A, Vela Jr., and G347.3–0.5, Astrophys. J. 897, 22 (2020), arXiv:2005.06544 [astro-ph.HE].
[25] K. Wette, L. Dunn, P. Clearwater, and A. Melatos, Deep exploration for continuous gravitational waves at 171–172 Hz in LIGO second observing run data, Phys. Rev. D 103, 083020 (2021), arXiv:2103.12976 [gr-qc].
[26] J. Abadie et al. (LIGO Scientific), First search for gravitational waves from the youngest known neutron star, Astrophys. J. 722, 1504 (2010), arXiv:1006.2535 [gr-qc].
[27] J. Aasi et al. (LIGO Scientific, VIRGO), Directed search for continuous gravitational waves from the Galactic center, Phys. Rev. D 88, 102002 (2013), arXiv:1309.6221 [gr-qc].
[28] B. Behnke, M. A. Papa, and R. Prix, Postprocessing methods used in the search for continuous gravitational-wave signals from the galactic center, Phys. Rev. D 91, 064007 (2015), arXiv:1410.5997 [gr-qc].
[29] B. P. Abbott et al. (LIGO Scientific, Virgo), Results of the deepest all-sky survey for continuous gravitational waves on LIGO S6 data running on the Einstein@Home volunteer distributed computing project, Phys. Rev. D 94, 102002 (2016), arXiv:1606.09619 [gr-qc].
[30] S. J. Zhu, M. A. Papa, H.-B. Egggenstein, R. Prix, K. Wette, B. Allen, O. Bock, D. Keitel, B. Krishnan, B. Machenschalk, M. Shaltev, and X. Siemens, An Einstein@home search for continuous gravitational waves from Cassiopeia A, Phys. Rev. D 94, 082008 (2016), arXiv:1608.07589 [gr-qc].
[31] B. P. Abbott et al. (LIGO Scientific, Virgo), First low-frequency Einstein@Home all-sky search for continuous gravitational waves in Advanced LIGO data, Phys. Rev. D 96, 122004 (2017), arXiv:1707.02669 [gr-qc].
[32] D. Keitel, G. Woan, M. Pitkin, C. Schumacher, B. Pearlstone, K. Riles, A. G. Lyne, J. Palfreyman, B. Stappers, and P. Weltevrede, First search for long-duration transient gravitational waves after glitches in the Vela and Crab pulsars, Phys. Rev. D 100, 064058
L. Sancho de la Jordana and A. M. Sintes, A
R. Prix, Search for continuous gravitational waves: J. J. Bero and J. T. Whelan, An Analytic
S. J. Zhu, M. A. Papa, and S. Walsh, New veto for P. Leaci, Methods to filter out spurious disturbances
R. Tenorio, L. M. Modafferi, D. Keitel, and A. M. Sintes, Application C. Dreissigacker, R. Prix, and K. Wette, Fast and Accurate Sensitivity Estimation for Continuous-Gravitational-Wave Searches, Phys. Rev. D 98, 084058 (2018), arXiv:1808.02459 [gr-qc].
R. Tenorio, D. Keitel, and A. M. Sintes, Application of a hierarchical MCMC follow-up to Advanced LIGO continuous gravitational-wave candidates, Phys. Rev. D 104, 084012 (2021), arXiv:2105.13860 [gr-qc].
D. Keitel, R. Prix, M. A. Papa, P. Leaci, and M. Siddiqi, Search for continuous gravitational waves: Improving robustness versus instrumental artifacts, Phys. Rev. D 89, 064023 (2014), arXiv:1311.5738 [gr-qc].
D. Keitel, Robust semicoherent searches for continuous gravitational waves with noise and signal models including hours to days long transients, Phys. Rev. D 93, 084024 (2016), arXiv:1509.02398 [gr-qc].
R. Prix, S. Giampanis, and C. Messenger, Search method for long-duration gravitational-wave transients from neutron stars, Phys. Rev. D 84, 023007 (2011), arXiv:1104.1704 [gr-qc].
R. Tenorio, L. M. Modafferi, D. Keitel, and A. M. Sintes, distromax: empirically estimating the distribution of the loudest candidate from a gravitational-wave search, https://github.com/Rodrigo-Tenorio/distromax (2021).
L. Sancho de la Jordana and A. M. Sintes, A $\chi^2$ veto for continuous wave searches, Class. Quant. Grav. 25, 184014 (2008), arXiv:0804.1007 [gr-qc].
P. Leaci, Methods to filter out spurious disturbances in continuous-wave searches from gravitational-wave detectors, Phys. Scr. 90, 125001 (2015).
S. J. Zhu, M. A. Papa, and S. Walsh, New veto for continuous gravitational wave searches, Phys. Rev. D 96, 124007 (2017), arXiv:1707.05268 [gr-qc].
J. T. Whelan, R. Prix, C. J. Cutler, and J. L. Willis, New Coordinates for the Amplitude Parameter Space of Continuous Gravitational Waves, Class. Quant. Grav. 31, 065002 (2014), arXiv:1311.0065 [gr-qc].
S. Dhurandhar, B. Krishman, and J. L. Willis, Marginalizing the likelihood function for modeled gravitational wave searches, arXiv e-print (2017), arXiv:1707.08163 [gr-qc].
J. J. Bero and J. T. Whelan, An Analytic Approximation to the Bayesian Detection Statistic for Continuous Gravitational Waves, Class. Quant. Grav. 36, 015013 (2019), [Erratum: Class.Quant.Grav. 36, 049601 (2019)], arXiv:1808.05453 [gr-qc].
Prix, Reinhard, The F-statistic and its implementation in ComputeFStatistic GPLv2, https://dcc.ligo.org/LIGO-T0900149/public (2018).
L. S. Finn, Detection, measurement and gravitational radiation, Phys. Rev. D 46, 5236 (1992), arXiv:gr-qc/9209010.
R. Tenorio, D. Keitel, and A. M. Sintes, Time-frequency track distance for comparing continuous gravitational wave signals, Phys. Rev. D 103, 064053 (2021), arXiv:2012.05752 [gr-qc].
R. Prix, Search for continuous gravitational waves: Metric of the multi-detector F-statistic, Phys. Rev. D 75, 023004 (2007), [Erratum: Phys. Rev. D 75, 069901(E) (2007)], arXiv:gr-qc/0606088.
H. J. Pletsch, Parameter-space metric of semicoherent searches for continuous gravitational waves, Phys. Rev. D 82, 042002 (2010), arXiv:1005.0395 [gr-qc].
K. Wette and R. Prix, Flat parameter-space metric for all-sky searches for gravitational-wave pulsars, Phys. Rev. D 88, 123005 (2013), arXiv:1310.5587 [gr-qc].
K. Wette, Parameter-space metric for all-sky semicoherent searches for gravitational-wave pulsars, Phys. Rev. D 92, 082003 (2015), arXiv:1508.02372 [gr-qc].
P. Leaci and R. Prix, Directed searches for continuous gravitational waves from binary systems: parameter-space metrics and optimal Scorpius X-1 sensitivity, Phys. Rev. D 91, 102003 (2015), arXiv:1502.00914 [gr-qc].
B. Allen, Spherical ansatz for parameter-space metrics, Phys. Rev. D 100, 124004 (2019), arXiv:1906.01352 [gr-qc].
K. Wette, Lattice template placement for coherent all-sky searches for gravitational-wave pulsars, Phys. Rev. D 90, 122010 (2014), arXiv:1410.6882 [gr-qc].
B. Allen and A. A. Shoom, Template banks based on Zn and An* lattices, Phys. Rev. D 104, 122007 (2021), arXiv:2102.11631 [astro-ph.IM].
K. Wette, Empirically extending the range of validity of parameter-space metrics for all-sky searches for gravitational-wave pulsars, Phys. Rev. D 94, 122002 (2016), arXiv:1607.00241 [gr-qc].
O. Piccinni, Mitigation of transient disturbances in wide parameter space searches for continuous gravitational wave signals, Master’s thesis, Università di Roma La Sapienza (2014).
K. Wette, Estimating the sensitivity of wide-parameter-space searches for gravitational-wave pulsars, Phys. Rev. D 85, 042003 (2012), arXiv:1111.5650 [gr-qc].
LIGO Scientific Collaboration, LIGO Algorithm Library - LALSuite, free software (GPL) (2018).
C. Rover, C. Messenger, and R. Prix, Bayesian versus frequentist upper limits, in PHYSTAT 2011 (CERN, Geneva, 2011) pp. 158–163, arXiv:1103.2987 [physics.data-an].
S. Suvorova, P. Clearwater, A. Melatos, L. Sun, W. Moran, and R. J. Evans, Hidden Markov model tracking of continuous gravitational waves from a binary neutron star with wandering spin. II. Binary orbital phase tracking, Phys. Rev. D 96, 102006 (2017), arXiv:1710.07092 [astro-ph.IM].
M. Leadbetter, G. Lindgren, and H. Rootzen, Extremes and Related Properties of Random Sequences and Processes, Springer Series in Statistics (Springer New York, 1983).
S. Coles, An introduction to statistical modeling of extreme values, Springer Series in Statistics (Springer-Verlag, London, 2001).
J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, S. Coles, An introduction to statistical modeling of extreme values, Springer Series in Operations Research - Springer Series in Probability and Statistics (Wiley, 2004).
L. de Haan and A. Ferreira, Extreme Value Theory: An Introduction, Springer Series in Operations Research and Financial Engineering (Springer New York, 2006).
P. Embrechts, C. Klüppelberg, and T. Mikosch, Modelling Extremal Events: for Insurance and Finance (Springer Berlin Heidelberg, 2013).
A. Gasull, M. Jolis, and F. Utzet, On the norming constants for normal maxima, J. Math. Anal. Appl. 422, 376 (2015).
A. Gasull, J. López-Salcedo, and F. Utzet, Maxima

B. Krishnan, A. M. Sintes, M. A. Papa, B. F. Schutz, A. Miller, P. Astone, S. D’Antonio, S. Frasca, P. Virtanen, R. Gommers, T. E. Oliphant, D. Keitel, R. Tenorio, G. Ashton, and R. Prix, Pyfstat:

M. Isi, S. Mastrogiovanni, M. Pitkin, and O. J. Piccinni, Establishing the significance of continuous gravitational-wave detections from known pulsars, Phys. Rev. D 102, 023006 (2020), arXiv:2006.06907 [astro-ph.HE].

G. Ashton and R. Prix, Hierarchical multistage MCMC follow-up of continuous gravitational wave candidates, Phys. Rev. D 97, 103020 (2018), arXiv:1802.05450 [astro-ph.IM].

D. Keitel, R. Tenorio, G. Ashton, and R. Prix, Pyfstat: a python package for continuous gravitational-wave data analysis, J. Open Source Softw. 6, 3000 (2021).

P. Virtanen, R. Gommer, T. E. Oliphant, M. Haberland, T. Reddy, et al., SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nat. Methods 17, 261 (2020).

B. Krishnan, A. M. Sintes, M. A. Papa, B. F. Schutz, S. Frasca, and C. Palomba, Hough transform search for continuous gravitational waves, Phys. Rev. D 70, 082001 (2004), arXiv:gr-qc/0407001.

P. Astone, A. Colla, S. D’Antonio, S. Frasca, and C. Palomba, Method for all-sky searches of continuous gravitational wave signals using the frequency-Hough transform, Phys. Rev. D 90, 042002 (2014), arXiv:1407.8333 [astro-ph.IM].

A. Miller, P. Astone, S. D’Antonio, S. Frasca, G. Intini, I. La Rosa, P. Leaci, S. Mastrogiovanni, F. Muciaccia, C. Palomba, O. J. Piccinni, A. Singhal, and B. F. Whiting, Method to search for long duration gravitational wave transients from isolated neutron stars using the generalized frequency-Hough transform, Phys. Rev. D 98, 102004 (2018), arXiv:1810.09784 [astro-ph.IM].

M. Oliver, D. Keitel, and A. M. Sintes, Adaptive transient Hough method for long-duration gravitational wave transients, Phys. Rev. D 99, 104067 (2019), arXiv:1901.01820 [gr-qc].

P. B. Covas and A. M. Sintes, New method to search for continuous gravitational waves from unknown neutron stars in binary systems, Phys. Rev. D 99, 124019 (2019), arXiv:1904.04873 [astro-ph.IM].

J. T. Whelan, S. Sundaresan, Y. Zhang, and P. Peiris, Model-Based Cross-Correlation Search for Gravitational Waves from Scorpius X-1, Phys. Rev. D 91, 102005 (2015), arXiv:1504.05890 [gr-qc].

V. Dergachev, On blind searches for noise dominated signals: a loosely coherent approach, Class. Quant. Grav. 27, 205017 (2010), arXiv:1003.2178 [gr-qc].

E. Goetz and K. Riles, An all-sky search algorithm for continuous gravitational waves from spinning neutron stars in binary systems, Class. Quant. Grav. 28, 215006 (2011), arXiv:1103.1301 [gr-qc].

V. Dergachev and M. A. Papa, Sensitivity improvements in the search for periodic gravitational waves using O1 LIGO data, Phys. Rev. Lett. 123, 101101 (2019), arXiv:1902.05530 [gr-qc].

R. Abbott et al. (LIGO Scientific, VIRGO, KAGRA), Search for continuous gravitational waves from 20 accreting millisecond X-ray pulsars in O3 LIGO data, Phys. Rev. D 105, 022002 (2022), arXiv:2109.09255 [astro-ph.HE].

H. J. Pletsch and B. Allen, Exploiting global correlations to detect continuous gravitational waves, Phys. Rev. Lett. 103, 181102 (2009), arXiv:0906.0023 [gr-qc].

K. Wette, S. Walsh, R. Prix, and M. A. Papa, Implementing a semicoherent search for continuous gravitational waves using optimally-constructed template banks, Phys. Rev. D 97, 123016 (2018), arXiv:1804.03392 [astro-ph.IM].

R. Abbott et al. (LIGO Scientific, VIRGO, KAGRA), Narrowband searches for continuous and long-duration transient gravitational waves from known pulsars in the LIGO-Virgo third observing run, arXiv e-print (2021), arXiv:2112.10990 [gr-qc].

L. M. Modafferi, J. Moragues, and D. Keitel (LIGO Scientific, VIRGO, KAGRA), Search setup for long-duration transient gravitational waves from pulsars during LIGO-Virgo third observing run (2022) arXiv:2201.08785 [gr-qc].

R. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Open data from the first and second observing runs of advanced ligo and advanced virgo, SoftwareX 13, 100658 (2021), arXiv:1912.11716 [gr-qc].

Prix, Reinhard, F-statistic bias due to noise-estimator, https://dcc.ligo.org/LIGO-T1100551/public (2006).

B. Steltner, M. A. Papa, H. B. Eggenstein, B. Allen, V. Dergachev, R. Prix, B. Machenschalk, S. Walsh, S. J. Zhu, and S. Kwang, Einstein@Home All-sky Search for Continuous Gravitational Waves in LIGO O2 Public Data, Astrophys. J. 909, 79 (2021), arXiv:2009.12260 [astro-ph.HE].

J. Ming, M. A. Papa, A. Singh, H. B. Eggenstein, S. J. Zhu, V. Dergachev, Y. Hu, R. Prix, B. Machenschalk, C. Beer, O. Behnke, and B. Allen, Results from an Einstein@Home search for continuous gravitational waves from Cassiopeia A, Vela Jr. and G347.3, Phys. Rev. D 100, 024063 (2019), arXiv:1903.09119 [gr-qc].

M. A. Papa, H. B. Eggenstein, S. Walsh, I. DiPalma, B. Allen, P. Astone, O. Bock, T. D. Creighton, D. Keitel, B. Machenschalk, R. Prix, X. Siemens,
A. Singh, S. J. Zhu, and B. F. Schutz, Hierarchical follow-up of subthreshold candidates of an all-sky Einstein@Home search for continuous gravitational waves on LIGO sixth science run data, Phys. Rev. D 94, 122006 (2016), arXiv:1608.08928 [astro-ph.IM].

[97] J. Ming, M. A. Papa, H.-B. Eggenstein, B. Machenschalk, B. Steltner, R. Prix, et al., Results from an einstein@home search for continuous gravitational waves from g347.3 at low frequencies in LIGO o2 data, Astrophys. J. 925, 8 (2022), arXiv:2108.02808 [gr-qc].

[98] P. Hall, On the rate of convergence of normal extremes, Journal of Applied Probability 16, 433 (1979).

[99] M. F. Bennett, A. Melatos, A. Delaigle, and P. Hall, Reanalysis of $F$-statistic gravitational-wave searches with the higher criticism statistic, Astrophys. J. 766, 99 (2013), arXiv:1302.2635 [astro-ph.SR].

[100] M. Sezgin and B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, J. Electron. Imaging 13, 146 (2004).

[101] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9, 62 (1979).

[102] C. Li and C. Lee, Minimum cross entropy thresholding, Pattern Recognit. 26, 617 (1993).

[103] C. Li and P. Tam, An iterative algorithm for minimum cross entropy thresholding, Pattern Recognit. Lett. 19, 771 (1998).

[104] J. M. S. Prewitt and M. L. Mendelsohn, The analysis of cell images, Ann. N. Y. Acad. Sci. 128, 1035 (1966).

[105] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, scikit-image: image processing in Python, PeerJ 2, e453 (2014).