Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **510,** 1597–1611 (2022) https://doi.org/10.1093/mnras/stab3434
Advance Access publication 2021 November 27

Removal and replacement of interference in tied-array radio pulsar observations using the spectral kurtosis estimator

M. Purver [1]★ C. G. Bassa,[1,2] I. Cognard,[3] G. H. Janssen,[1,2] R. Karuppusamy,[1,4] M. Kramer [1,4] K. J. Lee,[1,4,5] K. Liu,[1,3,4] J. W. McKee [6] D. Perrodin,[1,7] S. Sanidas,[1,8] R. Smits[1,2] and B. W. Stappers [1]

[1]*Jodrell Bank Centre for Astrophysics, The University of Manchester, Manchester M13 9PL, UK*
[2]*ASTRON (Netherlands Institute for Radio Astronomy), Postbus 2, NL-7990 AA Dwingeloo, the Netherlands*
[3]*Station de Radioastronomie de Nançay, Observatoire de Paris, F-18330 Nançay, France*
[4]*Max Planck Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany*
[5]*Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China*
[6]*Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada*
[7]*INAF – Osservatorio Astronomico di Cagliari, Via della Scienza 5, I-09047 Selargius (CA), Italy*
[8]*Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, NL-1098 XH Amsterdam, the Netherlands*

Accepted 2021 November 22. Received 2021 November 16; in original form 2020 July 1

ABSTRACT

We describe how to implement the spectral kurtosis method of interference removal (zapping) on a digitized signal of averaged power values. Spectral kurtosis is a hypothesis test, analogous to the t-test, with a null hypothesis that the amplitudes from which power is formed belong to a 'good' distribution – typically Gaussian with zero mean – where power values are zapped if the hypothesis is rejected at a specified confidence level. We derive signal-to-noise ratios (SNRs) as a function of amount of zapping for folded radio pulsar observations consisting of a sum of signals from multiple telescopes in independent radio-frequency interference environments, comparing four methods to compensate for lost data with coherent (tied-array) and incoherent summation. For coherently summed amplitudes, scaling amplitudes from non-zapped telescopes achieves a higher SNR than replacing zapped amplitudes with artificial noise. For incoherently summed power values, the highest SNR is given by scaling power from non-zapped telescopes to maintain a constant mean. We use spectral kurtosis to clean a tied-array radio pulsar observation by the Large European Array for Pulsars: the signal from one telescope is zapped with time and frequency resolutions of 6.25 ms and 0.16 MHz, removing interference, along with 0.27 per cent of 'good' data, giving an uncertainty of 0.25 μs in pulse time of arrival (TOA) for PSR J1022+1001. We use a single-telescope observation to demonstrate recovery of the pulse profile shape, with 0.6 per cent of data zapped and a reduction from 1.22 to 0.70 μs in TOA uncertainty.

Key words: methods: analytical – methods: data analysis – methods: numerical – methods: statistical – techniques: interferometric – pulsars: general.

1 INTRODUCTION

Terrestrial sources of radio-frequency interference hamper many astronomical radio observations, swamping the signal from outer space with intense and unwanted human-made radiation. Although it is usually impossible to subtract interference from useful information received at the same time and frequency, it is possible to remove or replace contaminated portions of data entirely, and so prevent them from affecting the integrated signal.

To remove interference, we wish to identify it using a fair and reliable method. One such method, a general signal processing technique termed 'spectral kurtosis' (Dwyer 1983), was first applied to radio astronomy by Nita et al. (2007), subsequently refined by Nita & Gary (2010a,b) and applied again by Gary, Liu & Nita (2010). It is a statistical method that attempts to separate 'Gaussian white noise' from everything else, on the assumption that the useful information resembles Gaussian white noise while the interference does not. It uses the signal power to make a binary decision about whether to remove data, and it can be applied with fine resolution in time and frequency. We refer to the quantity used to make this decision as 'the estimator'.

Pulsars are weak and rapidly varying astronomical radio sources that require observations with fine time resolution, so interference cannot easily be 'averaged out' with integration across time, frequency, or multiple telescopes. Spectral kurtosis has previously been used to remove interference from single-telescope pulsar observations from the Parkes, Lovell, and Green Bank Telescopes (van Straten 2013; Dolch et al. 2014; Lam 2016; Kar et al. 2019). We extend this approach to cover the case of a pulsar observation consisting of a sum of signals from an array of telescopes in independent interference environments, which results in a variable number of telescopes contributing to the observation at different times and frequencies. Since it is possible to compensate for contaminated data at one telescope using uncontaminated data from the rest of the array, we compare four methods of replacing the data that are lost from an array pulsar observation when interference is removed.

We apply one such method to a timing observation made as a part of the Large European Array for Pulsars (LEAP) project (first shown

★ E-mail: mark.purver@alumni.manchester.ac.uk

© 2021 The Author(s)
Published by Oxford University Press on behalf of Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

in section 4.5 Bassa et al. 2016). This array comprises five widely separated radio telescopes, whose simultaneous pulsar observations can be summed coherently to produce a tied array of equivalent sensitivity to a single telescope of 195 m in diameter. Coherent summation is not always possible, so LEAP sometimes acts as an incoherently summed array equivalent to a 130-m telescope. LEAP provides pulsar observations of greater sensitivity than any other existing steerable radio telescope, enabling precise timing of dynamical properties, such as a pulsar's rotational period and proper motion. The project aims to use high-precision timing to detect gravitational waves, and interference removal is necessary to maintain its accuracy. Spectral kurtosis can be applied independently to each telescope's data, excising interference while sacrificing as little useful information as possible.

In Section 2 of this paper, we fully describe the implementation of the spectral kurtosis method of Nita & Gary (2010b). In Section 3, we explain advantages and disadvantages of this method of interference detection. In Section 4, we explore the effects of general interference removal on folded pulse profiles produced using observations from an array of telescopes, and look at how adverse effects can be mitigated while maximizing signal-to-noise ratio (SNR). In Section 5, we summarize the application of spectral kurtosis to a LEAP observation. In Section 6, we conclude.

2 SPECTRAL KURTOSIS METHOD OF INTERFERENCE DETECTION

Each instance of the spectral kurtosis estimator is derived from a portion of the radio signal, and its value is a measure of the statistical properties of that portion. We use a hypothesis test to classify data as either 'good' or 'bad' based on the value of the estimator, where bad data are usually interference and good data are usually not. The null hypothesis is that data are good, because we know the values expected from good data. The classification provides evidence of whether a portion is *likely* to be good or bad in reality, but, since it examines a finite number of data, it cannot tell us whether the portion is definitely good or bad. We therefore remove portions that are *suspected* of being bad when they give estimator values that are outside specific limits, and we call this removal 'zapping'. We can control our level of suspicion, because the estimator allows us to define the fraction of good data that we are willing to zap mistakenly. In general, the more good data we are willing to lose, the more interference we will eliminate. If the amount of interference in an observation is small then we may zap more good data than bad, but the spoiling effect of even a small amount of dominant interference is usually sufficient to justify this.

In the following subsections, we explain how to calculate the estimator and its limits. We provide some example values for variables used in the calculations, as an aid to implementation of the method. Where numerical integration is required, we refer to routines within the GNU Scientific Library (GSL) for the C and C++ programming languages. All variables used have real values, with imaginary quantities shown explicitly using the imaginary unit i.

2.1 The radio signal

Although the estimator is a function of signal power, we begin with the amplitudes from which power is derived. The digitized signal initially consists of an evenly sampled time series of radio amplitudes, covering a fixed frequency bandwidth; we assume that the continuous signal has been limited to cover the same bandwidth before being sampled, so that its information is captured accurately (Shannon 1949). Each time sample records the amplitude in either one or two polarization components, and each amplitude may be either real or complex (in the case of complex amplitudes, the imaginary part is simply a phase-shifted version of the real part in which each sinusoidal wave making up the signal has been shifted by $\frac{\pi}{2}$ radians, capturing the same information as a real signal of twice the sampling rate). The amplitudes are drawn from a set containing a fixed number of discrete values, stored using a corresponding number of 'sampling bits', and we assume this set to be large enough to approximate a continuum of values for statistical purposes. Although the number of values does not change, the values themselves can be calibrated dynamically during an observation.

The frequency resolution of the signal can be improved, at the expense of time resolution, by performing discrete Fourier transforms (DFTs) on sequences of amplitudes; the more time samples are used in each DFT, the more the frequency resolution improves (pp. 260–262 Bracewell 2000). Although this is often referred to as 'moving from the time domain to the frequency domain', the values that come out of a DFT are still amplitudes, with each amplitude representing the signal at one time and one frequency. DFTs of consecutive sequences of amplitudes can therefore be used as a time series with multiple frequency channels, where each DFT contributes one time point to all channels. The process simply exchanges time resolution for frequency resolution, and it is worth noting that the DFT of a single value gives the value itself, i.e. the time-domain signal can be thought of as being the frequency-domain signal with one channel. The channelized amplitudes are generally complex, representing the magnitude and phase of the signal at each time and frequency. Given the same signal and equipment, a complex time series consisting of T samples and a real time series consisting of $2T$ samples produce almost the same amplitudes in channels 1 to $T-1$ of their respective DFTs (where T is a positive integer), as long as the frequencies in the signal are within the range of the channels. This similarity allows amplitudes from telescopes with real and complex sampling to be added together in the frequency domain (section 4.1 of Bassa et al. 2016). There are some differences in the way that continuous signal frequencies are distributed into discrete bins, which can be mitigated by applying different weights to each bin; the differences are larger in the lowest- and highest-frequency channels, and these channels are often not used because they do not approximate the original signal well even if weights are used (pp. 281, 288 Bracewell 2000). The DFT amplitudes in channels 0 and T from the real time series are themselves always real, so they cannot generally be compared to the amplitude in channel 0 from the complex time series, and the complex time series does not produce a channel T.

2.2 The statistical distribution of the signal

The identification of interference by the spectral kurtosis method is based on the probability density function (PDF) of a set of amplitudes, which gives the probabilities of a single amplitude taking any given value and which we refer to as the 'distribution' of the set. The number of amplitudes in the set can be chosen at will, and the time and frequency ranges covered by the set are the time and frequency resolutions of zapping.

Depending on the origin of the signal, the distribution can have different general forms (shapes), and other distinguishing characteristics, such as different mean values. We can use spectral kurtosis to test whether each measured set of amplitudes is well described by a particular distribution. In a 'good' signal of Gaussian white noise, the amplitudes (taking the real and imaginary parts as separate values if the signal is complex) are uncorrelated and have a Gaussian

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

distribution with a mean of 0 when collected over the time and frequency resolutions of zapping and across all polarizations; in a 'bad' signal, the distribution of amplitudes when collected over these ranges is non-Gaussian, contains correlated amplitudes and/or does not have a mean of 0. Although the portion of the signal from the pulsar might have a non-Gaussian distribution, it is typically too weak to substantially alter the total signal distribution, so a detectably bad signal is usually caused by interference. If a bad signal is caused by unpolarized interference, the distribution over time and frequency within a single polarization is bad, and two polarizations have the same distribution as one another; if a bad signal is caused by polarized interference, two polarization components are differently distributed. In all but pathological cases, the real and imaginary parts of a complex signal have the same distribution as one another.

2.3 The estimator

The estimator is an unbiased estimate of the scaled variance divided by the square of the mean for a set of samples of summed power. It is referred to as 'spectral' because it can be calculated separately for each frequency channel by using Fourier transforms, although it can also be calculated across the full bandwidth without leaving the time domain. It is called 'kurtosis' because the variance of signal power involves the mathematical fourth power of amplitude values, but each element in the calculation is more generally the square of a sum of squares rather than a simple fourth power.

Each summed power value, P_m, is assembled as a sum of squares of amplitude values that are either real, A_n, or complex, $A_n + iB_n$:

$$P_m = \sum_{n=1}^{2N} A_n^2 \tag{1}$$

or

$$P_m = \sum_{n=1}^{N} \left(A_n^2 + B_n^2\right) \tag{2}$$

In general, a set of power values has a non-zero mean, referred to as a 'baseline'. The baseline is usually subtracted from the values at some stage, but subtraction should not be done prior to calculation of the estimator. Our variable N is equivalent to the product Nd in Nita & Gary (2010b). N counts real and imaginary numbers and ranges in time, frequency and polarization without distinction, e.g. $N = 2$ could result from a sum of complex amplitudes at one time, one frequency, and two polarizations or from a sum of real amplitudes at two times, two frequencies, and one polarization. $2N$ is thus the total number of squared values that are summed to form each power value. For real amplitudes, $2N$ is a positive integer; for complex amplitudes, N is a positive integer. The time and frequency ranges are usually contiguous and evenly sampled, while the polarization range usually covers two orthogonal modes, although statistically these conditions are not necessary. When forming each power value, the use of one time and one frequency makes the estimator most sensitive to bad data (Nita & Gary 2010b), and the use of two polarizations allows it to be equally sensitive to polarized interference coming from different directions. The use of more than one time or frequency may be made in order to save data storage space or processing time, and power values can be added together to accomplish this without needing to know their constituent amplitudes. Spectral kurtosis can be extended to cases in which good data have a non-Gaussian amplitude distribution, resulting in non-integer values of $2N$ and a modification of equations (1) and (2), and this generalization has been used for two-bit data that do not approximate a continuous signal (Nita, Keimpema & Paragi 2019). Non-integer values of $2N$ could also be used if the amplitudes making up each summed power sample were distributed with different variances (e.g. if two polarization channels had different gain levels) – but we do not make use of this, preferring calibration to equalize variances prior to interference detection (section 4.3 of Bassa et al. 2016).

If formed from Gaussian-distributed amplitudes, the PDF of a set of M values of P_m is a gamma distribution. This gamma power distribution has the useful property that its variance and the square of its mean both scale linearly with the square of the variance of the Gaussian amplitude distribution, as long as the mean of the amplitude distribution is 0. We can therefore create a quantity that is independent of the amplitude variance of good data, which is the key motivation for defining the estimator (as in Nita & Gary 2010b) as

$$\hat{S} = \frac{(MN+1)V}{(M-1)\mu^2} = \frac{MN+1}{M-1}\left(\frac{M\sum_{m=1}^{M} P_m^2}{\left(\sum_{m=1}^{M} P_m\right)^2} - 1\right), \tag{3}$$

where μ is the mean of the set of summed power values, V is its variance and $\frac{MN+1}{M-1}$ is a scaling and unbiasing factor (note that factors, involving M and N can be brought inside the sums in equations (1), (2), and (3) to avoid the use of excessively large or small numbers during the calculations). The variance is scaled by the number of amplitudes contributing to each power value, so the factor would be N if it were not for the additional need to correct bias in an estimate of $\frac{V}{\mu^2}$ derived from a finite set. M is the total number of summed power samples that contribute to $\hat{S}$, and is therefore an integer; since variance is only meaningful for a set containing more than one value, we have the condition that $M \geq 2$. The set of M values usually covers a contiguous block of time and frequency: larger values of M coarsen the time and/or frequency resolutions of zapping but make the estimator more sensitive to bad data over the ranges of those resolutions (Nita & Gary 2010a), with Nita et al. (2007) advising that $M \geq 37$ is required to zap monochromatic interference. $\hat{S}$ represents an individual instance of the estimator, so we use S as the range of values that $\hat{S}$ can take; since the power values used to calculate $\hat{S}$ are always real, we have the condition that $S \geq 0$.

The underlying probability distribution of S can be revealed by making many measurements of $\hat{S}$. But the distribution for good data can also be approximated analytically, allowing it to be calculated more efficiently. The distribution depends on M and N, but, for Gaussian-distributed amplitudes with a mean of 0, it does not depend on the variance of those amplitudes. In other words, the estimator behaves in the same way for Gaussian amplitudes (good data) of any 'loudness', and can thus be used to distinguish them from most non-Gaussian amplitudes (bad data). The estimator shares this property with the t-statistic (Student 1908), and in fact a t-test could be used to classify good and bad data using amplitudes instead of power values. We have not undertaken an interference detection comparison between the estimator and the t-statistic, but have employed the estimator because it can be used on either averaged power values or amplitudes and because it can apply a consistent test to polarized signals regardless of the angle between the radio source and the receiver plane.

For good data, S has a mean of 1 for all allowed values of M and N; if we are to decide which amplitudes to accept as good without deriving the distribution of S empirically, we must calculate the distribution's shape as best we can by computing some of its higher moments as well.

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

2.4 The probability distribution of the estimator

In order to understand what values we expect the estimator to take when the null hypothesis is true, we determine the approximate cumulative distribution function (CDF), $P(S)$, that is produced by good $\hat{S}$ values. The CDF gives the fraction of good $\hat{S}$ values that are expected to fall at or below any level S, and its shape depends only on M and N.

A CDF is the integral of a PDF, $p(S)$, defined by

$$P(S) = \int_{S_{min}}^{S} p(s)\,\mathrm{d}s = 1 - \int_{S}^{S_{max}} p(s)\,\mathrm{d}s, \tag{4}$$

where s is simply a variable of integration and where the PDF and CDF are defined in the range $S_{min} \leq S \leq S_{max}$ (so $P(S_{min}) = 0$ and $P(S_{max}) = 1$). Gary et al. (2010) found that the required CDF is complicated to calculate, because it is the integral of a skewed PDF. Nita & Gary (2010a,b) showed that the CDF can be well approximated in most cases by numerically integrating a PDF called a Pearson distribution, which is defined by four parameters (given by equation (9) of Nita & Gary 2010b) and allows up to the first four of its statistical moments to be matched to those of the true PDF. The first parameter is the mean or first raw moment, which we set equal to 1. The second parameter is the variance or second central moment, and is given by

$$\mu_2 = \frac{2M^2N(N+1)}{(M-1)(MN+2)(MN+3)}. \tag{5}$$

Depending on the values of M and N, the PDF required may be a Pearson distribution of Type I, Type IV, or Type VI (Pearson 1895, 1901). To find out which Type, we use two parameters related to the third and fourth central moments (and therefore to skewness and kurtosis),

$$\beta_1 = \frac{8(MN+2)(MN+3)(MN(N+4)-5N-2)^2}{(M-1)(MN+4)^2(MN+5)^2N(N+1)} \tag{6}$$

and

$$\begin{aligned}\beta_2 = {} & \frac{3(MN+2)(MN+3)}{(M-1)(MN+4)(MN+5)(MN+6)(MN+7)} \\ & \times (M^3N^3(N+1) + M^2N^2(3N^2+68N+125) \\ & - MN(93N^2+245N+32) + 12(7N^2+4N+2)) \\ & \times \frac{1}{N(N+1)},\end{aligned} \tag{7}$$

to define:

$$\kappa = \frac{\beta_1(\beta_2+3)^2}{4(4\beta_2-3\beta_1)(2\beta_2-3\beta_1-6)} \tag{8}$$

(for any PDF, the quantities μ_2 and β_1 are non-negative and $\beta_2 \geq \beta_1 + 1$). We use Type I if $\kappa \leq 0$, Type IV if $0 < \kappa < 1$ and Type VI if $\kappa > 1$ (special cases of Type V if $\kappa = 1$ and Type III if $\kappa = \infty$ do not arise for the allowed values of M and N). Fig. 1 of Nita & Gary (2010b) shows which types correspond to different values of M and N, demonstrating that Type IV is likely to be used if $N \leq 13.5$, Type VI if $N \geq 14$ and Type I only if $M \leq 9$. When calculating the CDF for all three Types, we use an additional parameter (which is positive for the allowed values of M and N) giving the ratio of the third and second central moments:

$$\alpha_1 = \sqrt{\mu_2\beta_1} = \frac{4M(MN(N+4)-5N-2)}{(M-1)(MN+4)(MN+5)}. \tag{9}$$

For each Type, we calculate location-scale transformations of the PDF and the CDF, $p'(S')$ and $P'(S')$, where

$$P'(S') = \int_{S'_{min}}^{S'} p'(s)\,\mathrm{d}s = P(S) \tag{10}$$

when

$$S = aS' + \lambda \tag{11}$$

and where the values of a and λ are calculated differently for each Type (a is positive for any PDF). S' is used because $p'(S')$ is a simpler form of a Pearson distribution than $p(S)$. The transformation gives $p(S)$ the required moments: for example, the mean of $p'(S')$ should always be found to be $\frac{1-\lambda}{a}$, as this ensures that $p(S)$ has a mean of 1. In the following paragraphs, we explain how to calculate $P'(S')$, a, and λ for each Type (note that all square roots refer to the non-negative value).

2.4.1 *The CDF using Type I*

Type I corresponds to $\kappa \leq 0$, and applies only in some cases for which $M \leq 9$ (as long as $N \geq 0.5$, which covers all of its allowed values when calculating the estimator distribution caused by Gaussian white noise). Such small values of M give a PDF with a large variance, making the estimator so insensitive to interference that it might not be considered useful (Nita & Gary 2010a). Additionally, we have found that Type I does not provide a good approximation to the true PDF of S as generated using simulated random numbers, even though it does approximate the first four moments well. The problem is worst at small S, where the Type I PDF gives a substantial probability of $S < 0$, despite the fact that $\hat{S}$ cannot be negative. These negative values indicate that a Pearson distribution is inadequate when M is small, and that it is necessary to approximate more than four moments or to derive the PDF empirically in these cases. However, since we do not know the scope of applications for which spectral kurtosis will be employed, we include this part of the method for completeness.

Following a method equivalent to that of Kendall, Stuart & Ord (1994, pp. 217–220), we match the first four moments of the true PDF by using the parameters

$$c_0 = \mu_2(4\beta_2 - 3\beta_1), \tag{12}$$

$$c_1 = \alpha_1(\beta_2+3) \tag{13}$$

and

$$c_2 = 6 + 3\beta_1 - 2\beta_2 \tag{14}$$

to define:

$$c = \sqrt{c_1^2 + 4c_0c_2}. \tag{15}$$

We then use these with the further parameter

$$c_3 = 15 + 9\beta_1 - 7\beta_2 \tag{16}$$

to define:

$$n_1 = 2 + \frac{c_3}{c_2}\left(\frac{c_1}{c} - 1\right) \tag{17}$$

and

$$n_2 = 2 - \frac{c_3}{c_2}\left(\frac{c_1}{c} + 1\right) \tag{18}$$

(of these seven quantities, only c_3 can be negative for the allowed values of M and N in the Type I case). The transformed CDF can then be

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

calculated as

$$P'(S') = \frac{\int_0^{S'} s^{n_1-1}(1-s)^{n_2-1}\,\mathrm{d}s}{\int_0^1 s^{n_1-1}(1-s)^{n_2-1}\,\mathrm{d}s} = \frac{\mathrm{B}(S';n_1,n_2)}{\mathrm{B}(n_1,n_2)}, \tag{19}$$

where $\mathrm{B}(S';n_1,n_2)$ and $\mathrm{B}(n_1,n_2)$ are the incomplete and complete beta functions, respectively. This can be computed as the normalized incomplete beta function using GSL, or it can be found using numerical integration (pp. 9, 21 Bateman & Erdélyi 1953) as

$$P'(S') = \int_0^{S'} \exp((n_1-1)\ln s + (n_2-1)\ln(1-s) - \ln \mathrm{B}(n_1,n_2))\,\mathrm{d}s \tag{20}$$

with

$$\ln \mathrm{B}(n_1,n_2) = \int_0^\infty \left(\frac{(e^{-n_1 s} + e^{-n_2 s} - e^{-(n_1+n_2)s} - e^{-s})}{s(1-e^{-s})} - \frac{e^{-s}}{s} \right) \mathrm{d}s, \tag{21}$$

both of which approaches avoid the use of very large numbers during the calculation. The transformed CDF is defined in the range $0 \le S' \le 1$.

The transformation from S' to S can be made using equation (11) with

$$a = \frac{c}{c_2} \tag{22}$$

and

$$\lambda = 1 - \frac{1}{2c_2}\left(\frac{c_1 c_3}{2c_2 - c_3} + c\right). \tag{23}$$

The transformed PDF is a beta distribution, which has a mean of $\frac{1-\lambda}{a} = \frac{n_1}{n_1+n_2}$ as required. Example Type I values are: $M = 3$, $N = 4$, $a = 53.67$, $\lambda = -1.699$.

2.4.2 *The CDF using Type IV*

Type IV corresponds to $0 < \kappa < 1$. It applies in all cases for which $M \ge 240$ and $N \le 13.5$, and in some cases for which $14 \le M \le 239$ and $N \le 13.5$, so it is most commonly used when few amplitudes are used to make each power value. It approximates the true PDF of S well for all values of M and N to which it applies, working best when $M \gtrsim 25$.

Following Nita & Gary (2010a), we match the first four moments of the true PDF by using the parameter

$$r = \frac{6(\beta_2 - \beta_1 - 1)}{2\beta_2 - 3\beta_1 - 6} \tag{24}$$

to define:

$$u = 16(r-1) - \beta_1(r-2)^2. \tag{25}$$

We then use these to define:

$$w = r(r-2)\sqrt{\frac{\beta_1}{u}} \tag{26}$$

(these three quantities are all positive for the allowed values of M and N in the Type IV case). The transformed CDF can then be calculated as:

$$P'(S') = \frac{2^r \left|\Gamma\left(\frac{r+2+iw}{2}\right)\right|^2}{\pi\Gamma(r+1)} \int_{-\infty}^{S'} \frac{\exp(w\arctan s)}{(s^2+1)^{\frac{r+2}{2}}}\,\mathrm{d}s$$
$$= \int_{-\infty}^{S'} \exp\left(w\arctan s - \frac{(r+2)\ln(s^2+1)}{2} + r\ln 2 + 2\ln\left|\Gamma\left(\frac{r+2+iw}{2}\right)\right| - \ln\pi - \ln\Gamma(r+1)\right)\mathrm{d}s, \tag{27}$$

where Γ denotes the gamma function, i.e. $\Gamma(r+1) = \int_0^\infty s^r e^{-s}\,\mathrm{d}s$. The second form of equation (27) avoids the use of very large numbers during the calculation. Numerical integration is needed (unless potentially computationally expensive hypergeometric series are used), but the log-gamma functions can be computed using GSL, or they can be found using further numerical integration (p. 21 Bateman & Erdélyi 1953) as

$$\ln\Gamma(r+1) = \int_0^\infty \left(\frac{e^{-(r+1)s} - e^{-s}}{s(1-e^{-s})} + \frac{re^{-s}}{s} \right) \mathrm{d}s \tag{28}$$

and

$$\ln\left|\Gamma\left(\frac{r+2+iw}{2}\right)\right| = \Re\left[\ln\Gamma\left(\frac{r+2+iw}{2}\right)\right] = \int_0^\infty \left(\frac{\cos\left(\frac{ws}{2}\right)e^{-\frac{(r+2)s}{2}} - e^{-s}}{s(1-e^{-s})} + \frac{re^{-s}}{2s} \right)\mathrm{d}s. \tag{29}$$

The transformed CDF is defined in the range $-\infty \le S' \le \infty$.

The transformation from S' to S can be made using equation (11) with

$$a = \frac{\sqrt{\mu_2 u}}{4} \tag{30}$$

and

$$\lambda = 1 - \frac{\alpha_1(r-2)}{4} \tag{31}$$

(there is a typographical error in the definition of a in equation (57) of Nita & Gary 2010a, in which 6 should read 16). Example Type IV values are: $M = 1000$, $N = 2$, $a = 0.5008$, $\lambda = 0.5593$.

2.4.3 *The CDF using Type VI*

Type VI corresponds to $\kappa > 1$. It applies in all cases for which $M \ge 3$ and $N \ge 14$, and in some cases for which $3 \le M \le 239$ and $N \le 13.5$, so it is most commonly used when many amplitudes are used to make each power value. Like Type I, Type VI cannot identify interference fairly when M is small: it gives a substantial probability of $S < 0$ in some cases for which $M \lesssim 12$, even though $\hat{S}$ cannot be negative. In the more useful cases for which $M \gtrsim 25$, however, simulations with random numbers showed that Type VI provides a good approximation to the true PDF of S.

Following Nita & Gary (2010b), we match the first three moments (not four, as in the other cases) of the true PDF by using the parameter

$$h = 4 + \sqrt{\beta_1\left(\frac{1}{\mu_2} + 4\right) + 16} \tag{32}$$

to define:

$$\alpha = \frac{1}{\alpha_1}\left(\mu_2\left(h\left(\frac{1}{\alpha_1}\left(\frac{8\mu_2}{\alpha_1} - 1\right) + 1\right) + 4\right) + 1\right) - 1 \tag{33}$$

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

and

$$\beta = 3 + \frac{2h}{\beta_1}. \tag{34}$$

(these three quantities are all greater than 8 for the allowed values of M and N in the Type VI case). The transformed CDF can then be calculated as

$$P'(S') = \frac{\int_0^{S'} s^{\alpha-1}(1+s)^{-(\alpha+\beta)}\,\mathrm{d}s}{\int_0^1 s^{\alpha-1}(1-s)^{\beta-1}\,\mathrm{d}s} = \frac{\mathrm{B}\left(\frac{S'}{1+S'};\alpha,\beta\right)}{\mathrm{B}(\alpha,\beta)}, \tag{35}$$

where $\mathrm{B}(\frac{S'}{1+S'};\alpha,\beta)$ and $\mathrm{B}(\alpha,\beta)$ are the incomplete and complete beta functions, respectively. This can be computed as the normalized incomplete beta function using GSL (in which case the transformation can later be made directly from $\frac{S'}{1+S'}$ to S without explicitly calculating S'), or it can be found using numerical integration as

$$P'(S') = \int_0^{S'} \exp((\alpha-1)\ln s - (\alpha+\beta)\ln(1+s) - \ln \mathrm{B}(\alpha,\beta))\,\mathrm{d}s \tag{36}$$

with equation (21) for $\ln \mathrm{B}(\alpha,\beta)$, both of which approaches avoid the use of very large numbers during the calculation. The transformed CDF is defined in the range $0 \le S' \le \infty$.

The transformation from S' to S can be made using equation (11) with

$$a = 1 \tag{37}$$

and

$$\lambda = 1 - \frac{\alpha}{\beta - 1}. \tag{38}$$

The transformed PDF is a beta-prime distribution, which has a mean of $\frac{1-\lambda}{a} = \frac{\alpha}{\beta-1}$ as required. Example Type VI values are: $M = 600$, $N = 16$, $a = 1$, and $\lambda = -0.3393$.

2.5 The limits of the estimator

We now decide on the fraction of good data that we are willing to reject, $2f$, and use the transformed CDF to calculate lower and upper limits of S such that a fraction f of good $\hat{S}$ values are expected to fall below the lower limit, S_L, and a fraction f are expected to fall above the upper limit, S_U. The range between the limits is a confidence interval: we reject any portion of data that produces an estimator value outside the range, because we suspect those data of being bad. We wish to choose the smallest value of f that adequately removes interference, in order to keep as many good data as possible and avoid any substantial change to the overall amplitude distribution of an observation.

Since we are using $P'(S')$, we calculate transformed limits S'_L and S'_U that are related to S_L and S_U by equation (11). These must satisfy the condition:

$$P'(S'_L) = 1 - P'(S'_U) = f, \tag{39}$$

where $0 \le f \le 1$. The transformed PDF has a mean of $\frac{1-\lambda}{a}$ and a variance of $\frac{\mu_2}{a^2}$, so initial guesses for S'_L and S'_U can be found using a rough Gaussian approximation for the PDF:

$$S'_L \approx \frac{1 - \lambda - \eta\sqrt{\mu_2}}{a} \tag{40}$$

and

$$S'_U \approx \frac{1 - \lambda + \eta\sqrt{\mu_2}}{a}, \tag{41}$$

where η is a positive number such that

$$f = \frac{1}{2} - \frac{1}{\sqrt{\pi}} \int_0^{\frac{\eta}{\sqrt{2}}} \exp(-s^2)\,\mathrm{d}s = \frac{1 - \mathrm{erf}\left(\frac{\eta}{\sqrt{2}}\right)}{2} \tag{42}$$

and 'erf' denotes the error function. f can be calculated from η using the error function in GSL or using numerical integration; alternatively, η can be calculated from f using the inverse cumulative Gaussian distribution function in GSL. The use of η is not absolutely necessary, but it allows us to make reasonable initial guesses and to describe our non-Gaussian PDF using the familiar language of Gaussian distributions: if we choose $\eta = 3$, for example, then we can refer to S_L and S_U as 'three-sigma limits', meaning that they exclude the same fraction of good data from our PDF as limits that were three standard deviations from the mean would exclude from a Gaussian distribution (where standard deviation is the square root of variance).

After checking that the initial guess for a transformed limit falls within the range for which the relevant CDF is defined, we calculate the value of the CDF at that point. Since all CDFs increase monotonically, we make the guess a lower bound if the CDF falls below its target value or as an upper bound if the CDF falls above its target value. We then calculate the CDF at intervals in S' (moving in one direction by, for example, $\frac{\sqrt{\mu_2}}{a}$ at a time) until the CDF crosses its target value, giving us the other bound for the transformed limit. After this, we bisect the upper and lower bounds iteratively, and at each iteration we make the bisection point a new upper or lower bound according to the value of the CDF at that point. Once the upper and lower bounds are sufficiently close together, we take their bisection point as the transformed limit, and finally convert this to a limit on S using equation (11). We use our limits to zap portions of data that give $\hat{S} < S_L$ or $\hat{S} > S_U$.

For $\eta = 3$ ($f = 0.001350$), example values corresponding to those in Sections 2.4.1–2.4.3 are: $M = 3$, $N = 4$, $S_L = -1.492$, $S_U = 7.417$ (Type I); $M = 1000$, $N = 2$, $S_L = 0.8499$, $S_U = 1.1818$ (Type IV); $M = 600$, $N = 16$, $S_L = 0.8321$, $S_U = 1.1901$ (Type VI).

3 ADVANTAGES AND DISADVANTAGES OF THE ESTIMATOR

Spectral kurtosis is one of many methods designed to distinguish interference from useful data. While spectral kurtosis looks for non-Gaussianity in the distribution of amplitudes, most techniques flag outlying power values. 'Median absolute deviation', for example, distinguishes any portion of data whose power is substantially different to the portions around it, using a median-based variance estimate that is robust to outliers (Fridman 2009). The pulsar processing software PSRCHIVE can tackle narrowband interference by automatically zapping data in frequency channels that stand out from the median channel power, and can also remove impulsive interference by zapping parts of an average pulse profile that deviate from the expected shape (van Straten, Demorest & Oslowski 2012).

Other approaches to interference mitigation include: eliminating 'cyclostationary' signals that have periodic statistical properties (Ait-Allal et al. 2010); characterizing known sources of interference in a specific environment (Czech, Mishra & Inggs 2018); and removing signals that appear simultaneously to both a telescope and an adjacent reference receiver that is not pointing at the astronomical source (Briggs, Bell & Kesteven 2000).

Since no method of interference removal is perfect, we examine some of the particular features of the estimator.

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

3.1 Advantages of the estimator

The estimator is statistically unbiased, allowing us to choose fairly the fraction of good data that will be rejected. It is also simple in its assumption that the distributions of interference can usually be distinguished from the distribution of useful data: this is an unashamedly frequentist approach that requires no prior knowledge of the interference distributions. An examination by eye shows that spectral kurtosis is successful in zapping many different kinds of interference, with the loss of only a small fraction of good data.

As shown in Section 5, the estimator can be used to zap interference with fine time and frequency resolutions (e.g. 6.25 ms and 0.16 MHz), salvaging more useful data than other methods with coarser resolutions. Zapping can operate in either the time or frequency domain, allowing the time resolution to be improved at the expense of frequency resolution or vice versa. The ability of the estimator to detect interference generally improves as more power values are accumulated, since its variance decreases as M increases (Nita & Gary 2010a), but the best value of M also depends on the distribution and time-scale of the interference.

It is possible to zap different types of interference by using the estimator with different values of M on the same data, either independently or jointly in a 'multiscale' approach (Gary et al. 2010), the latter also allowing transient signals to be detected and classified using data with as few as two sampling bits (Nita 2016; Nita & Gary 2016; Nita et al. 2019). It is also possible to make the estimator more sensitive to randomly changing signals and less sensitive to smoothly varying ones by normalizing the power at each frequency and time by the total power across the bandwidth at the same time, before using the normalized values to produce the estimator (Nita et al. 2007).

The estimator can be effective on power values that have been averaged together, so it can be used retrospectively after data have been compressed in this way, although its performance deteriorates as more values are used in each average.

3.2 Disadvantages of the estimator

The frequentist approach that makes spectral kurtosis simple is also its fundamental limitation. Ideally, we would zap data using a Bayesian method in which we had prior knowledge of the amount and the distributions of the interference. Without these things, we cannot give an accurate probability that any particular portion of data is bad. But prior knowledge is not generally available to us, since the interference environment often changes on time-scales shorter than the duration of our radio observations. We therefore choose to use spectral kurtosis based on our long-term experience and suspicions about interference, but must acknowledge that the technique will be more successful in some situations than others.

Inevitably, the estimator will sometimes cause us to zap useful data or fail to zap interference. This can happen in two ways. First, data can be mislabelled as good or bad. We can control the fraction of good data that will be rejected (the type I error rate), but we cannot predetermine the fraction of bad data that will be accepted (the type II error rate), since it depends on the similarity between the distributions of estimator values from bad data and the distribution from good data. Correlated Gaussian noise might be labelled as good, for example, and Nita et al. (2007) found that periodic interference with a duty cycle of 40–60 per cent can closely mimic a good data distribution. Secondly, useful data can have a 'bad' distribution or interference can have a 'good' distribution. The former case could occur if single pulses from a pulsar had an SNR close to or greater than 1 in a single frequency channel; the latter case could result from interference that had uncorrelated Gaussian-distributed amplitudes whose variance remained approximately constant over the zapping time-scale and bandwidth. Zapping the pulsar, in particular, could alter its apparent profile shape (see Section 4.1). To combat these problems, we should use frequency channels that are too fine for single pulses to be seen above the observational noise, and we may need to employ other methods of interference removal as well as, or instead of, spectral kurtosis.

The statistical nature of spectral kurtosis makes it less suitable for data that have been averaged over many power values. As the number of power values averaged together (N) increases, the estimator loses its ability to distinguish between different gamma power distributions and therefore to detect certain types of interference (Nita & Gary 2010b). The method works best on power values sampled near the Nyquist rate, requiring substantial data storage space and computational power when zapping.

4 EFFECTS OF INTERFERENCE REMOVAL ON PULSE PROFILES

Regardless of the method used to identify interference, zapping alters our data. We look at situations where this may cause a problem for pulsar observations, and examine four methods of compensating for the lost data in array observations with telescopes in independent interference environments.

4.1 Situations in which zapping may alter profile shape

When making simultaneous radio observations using multiple telescopes in order to increase the SNR, the signals from each telescope may be summed coherently, using amplitudes with phase information, or incoherently, using power without phase information. When portions of the signal at each telescope are independently zapped prior to summation, the final signal has a variable number of contributing telescopes as a function of time and frequency. This can pose a problem for pulsar observations in particular, because of the importance of profile shape.

After summation over the available telescopes, a pulsar observation is typically summed as a function of the pulsar's rotational phase in a process called 'folding'. This gives the 'average pulse', or profile. Samples are summed incoherently in a number of phase 'bins', giving a profile with phase and frequency resolution, to which time resolution is added by repeating the process (pp. 165–166 Lorimer & Kramer 2005). Profiles give a higher SNR than individual pulses, which allows more accurate timing of the pulsar. Timing accuracy also relies on the profile shape remaining highly stable over time, and so every care is taken to avoid altering it instrumentally. Profile shape change could be caused by portions of a strong pulsar signal being zapped as interference, but, even when zapping is unrelated to pulsar emission, we must still understand its effect on the profile.

To quantify the profile change caused by zapping, we begin with the amplitude signal at a single telescope with no interference (the signal may be real or complex and in the time or frequency domain). We neglect variation between individual pulse measurement, which is caused by phenomena such as pulse jitter and interstellar scintillation (see e.g. Lorimer & Kramer 2005, fig. 1.1 and pp. 8, 92, 202), and assume that all pulses are identical monochromatic waves whose magnitude can be described as a function of rotational phase only. The signal power consists of a source (pulsar) contribution and a gamma-distributed random noise contribution, where the noise

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

comes from Gaussian-distributed random amplitudes. In a single time sample within a single frequency and polarization channel, before summation over N or $2N$ values (see Section 2.3), the source power alone would have a mean of $h(\theta)$ and a variance of 0 at a single phase value, θ, while the random noise power alone would have a mean of g and a variance of g^2 (where $h(\theta)$ and g are both non-negative). Because the source and noise have already been added together as amplitudes, within which the two contributions were independent, the summed power has a mean of $N(h(\theta) + g)$ and a variance of $N(2h(\theta)g + g^2)$. When many samples of summed power are added together in folding, the central limit theorem dictates that the folded power follows an approximately Gaussian distribution. Assuming that g does not change with time, the variance is approximately constant if $h(\theta) \ll g$, which is usually the case since $h(\theta)$ and g come from individual power samples and pulsar radiation at the Earth is weak. However, we can see that there is some phase-dependence to the variance, which could cause measured profile shape to deviate from the true profile shape given by the function $h(\theta)$ if the noise is non-Gaussian. Furthermore, zapping would cause variations in the power level that might manifest as additional phase-dependent noise or as an inconstant power baseline, where the baseline without zapping has a constant value of Ng.

4.2 Equalizing zapped data in array observations with independent zapping at each telescope

After summing a signal coherently over multiple telescopes, we may wish to return the resulting amplitudes to the time domain via an inverse Fourier transform. We can avoid artefacts from the process by minimizing the noise level changes that can be produced by zapping, which also produces a more constant power baseline in both coherently and incoherently summed signals. This minimization is achieved by equalizing the amplitude or power variances of summed samples that have different amounts of zapping, either by scaling up samples that have fewer contributing telescopes or by adding artificial noise to them. If we are not returning the signal to the time domain, we can instead equalize the power mean by subtracting different baseline values from summed samples that have different amounts of zapping. Below, we compare the typical SNR of the summed signal after using these three processes and after no equalization.

If we assume that a fraction, q, of samples are zapped independently at each of L telescopes, then the number of telescopes contributing to each summed sample is drawn from a binomial probability distribution. When many samples are zapped independently and added together in folding, and q is not very close to either 0 or 1, the central limit theorem again shows that the folded signal power follows an approximately Gaussian distribution. The zapping introduces a characteristic mean, $\mu(x)$, and variance, $\sigma^2(x)$, given by:

$$\mu(x) = \frac{1}{L^x} \sum_{l=1}^{L} l^x b_l \tag{43}$$

and

$$\sigma^2(x) = \mu(2x) - \mu(x)^2, \tag{44}$$

where

$$b_l = \frac{L!\, q^{L-l}(1-q)^l}{l!(L-l)!} \tag{45}$$

and where $0 \leq q \leq 1$ and x is a positive number. Since b_l is the probability of l telescopes contributing to a sample, where l is an integer in the range $0 \leq l \leq L$, we find that $\sum_{l=0}^{L} b_l = 1$. The characteristic mean and variance are normalized by L so that $0 \leq \mu(x) \leq 1$ and $0 \leq \sigma^2(x) \leq 0.25$ (if no zapping occurs, we have $q = 0$, $\mu(x) = 1$ and $\sigma^2(x) = 0$). The value of x depends on the method of variance equalization used and whether the summation of telescopes is coherent or incoherent: where l telescopes contribute to a summed sample, the mean source contribution after equalization is proportional to l^x. Four useful identities are:

$$\mu(1) = 1 - q, \tag{46}$$

$$\sigma^2(1) = \frac{q(1-q)}{L}, \tag{47}$$

$$\sigma^2(2) = \frac{4q(1-q)^3}{L} + \frac{(10q-4)q(1-q)^2}{L^2} + \frac{q(1-q) - 6q^2(1-q)^2}{L^3} \tag{48}$$

and

$$\mu(3) = (1-q)^3 + \frac{3q(1-q)^2}{L} + \frac{(3q-2)(1-q) + 2(1-q)^3}{L^2}, \tag{49}$$

with $\mu(2)$ and $\mu(4)$ found by placing these identities into equation (44).

In the following paragraphs, we provide expressions for the mean and variance of the zapped, equalized, and folded signal using multiple telescopes, employing these relations for the mean (expectation), E, and variance, Var, of a set of values, $\{X\}$:

$$\mathrm{E}\{X\} = \sum_{l=0}^{L} \mathrm{E}\{X_l\} b_l \tag{50}$$

and

$$\begin{aligned}\mathrm{Var}\{X\} &= \sum_{l=0}^{L} \mathrm{E}\left\{X_l^2\right\} b_l - (\mathrm{E}\{X\})^2 \\ &= \sum_{l=0}^{L} (\mathrm{Var}\{X_l\} + (\mathrm{E}\{X_l\})^2) b_l - (\mathrm{E}\{X\})^2,\end{aligned} \tag{51}$$

where $\{X\}$ is the union of L subsets of values, $\{X_l\}$, and where the numbers of members in the subsets follow a binomial probability distribution. Here, $\{X\}$ represents a set of signal values after summation across multiple telescopes. The values may be the magnitudes of amplitudes, or they may be power values. $\{X_l\}$ represents the subset of these values that are produced using l telescopes. The validity of the expressions has been verified using simulations of Gaussian-distributed random numbers representing signal amplitudes. We look at coherent and incoherent summation, and at four methods of dealing with zapping: variance equalization by scaling, variance equalization by addition of artificial noise, mean equalization, and no equalization at all. Mean values represent the pulse profile, and additive terms in their equations that do not depend on $h(\theta)$ are baselines that can be subtracted after folding. Variance values represent profile noise, and we give precedence to the terms that are largest in a typical pulsar observation, with the sum of all terms below the first line of each variance equation becoming equal to 0 if there is no zapping. The profile SNR, a measure of the quality of an observation, is given by:

$$\mathrm{SNR} = \frac{\mu'(\theta)}{\sigma(\theta)}, \tag{52}$$

where $\mu'(\theta)$ is the baseline-subtracted mean of the folded pulse profile and $\sigma(\theta)$ is the positive square root of its variance. Both of

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

these are functions of the pulsar's rotational phase, but the latter is only weakly dependent when $h(\theta) \ll g$. We assume that all interference has been zapped, that the pulsar signal itself is too weak to cause zapping, that the time series at all telescopes are perfectly aligned and that the signals have been scaled so that all telescopes record equal power variance. We also assume that the same fraction of samples are zapped at each telescope and that all telescopes record the same mean power. If these last two assumptions were not correct, the forms of the equations would still be valid, but $\mu(x)$ and $\sigma^2(x)$ would have to change into sums of all possible combinations of telescopes, L^x would have to change into weighted sums of contributing telescopes and $h(\theta)$ and g would have to change into weighted average values for the source and noise contributions in a single power sample at a single telescope.

4.2.1 *Equalization of coherent observations*

With coherent summation, $\{X_l\}$ represents a set of magnitudes of amplitudes summed over l telescopes. Power values are the squared magnitudes of these summed amplitudes. The set of power values has a measurable mean, $\mathrm{E}\{X_l^2\}$, which is $(l^2\overline{h} + lg)$ if the amplitudes are complex or $(l^2\overline{h} + lg)/2$ if they are real, where $\overline{h}$ is the time-average of $h(\theta)$ over all phase values. This quantity should be measured using a sufficient number of samples to make a stable estimate. Each summed amplitude is associated with its own value of l after zapping, and we need to measure $\mathrm{E}\{X_l^2\}$ using all combinations of telescopes if $\overline{h}$ makes any substantial contribution, since the $\overline{h}$ and g terms vary with different powers of l and we cannot measure the two independently. If g dominates, however, then we can calculate $\mathrm{E}\{X_l^2\}$ using a separate measurement from each telescope. Ideally, it should be measured and used separately in each frequency and polarization channel, but an average can be used if it is similar in each channel. We assume that $\overline{h}$ remains constant, as does $h(\theta)$ at each phase value, but any measurable variation in these parameters could be included in g, which would then acquire phase dependence, and the following equations for coherent and incoherent observations would remain valid. In principle, we could also replace $\overline{h}$ with $h(\theta)$ in these equations if we were able to measure the time average of $h(\theta)$ separately at each phase value, but this would require longer, folded variance measurements.

To equalize variances by scaling, we multiply each summed amplitude by $\sqrt{(L^2\overline{h} + Lg)/(l^2\overline{h} + lg)}$, except when $l = 0$. We replace any sample that has been zapped at all telescopes with a random complex number in which the real and imaginary parts are drawn independently from a Gaussian distribution of mean 0 and variance $(L^2\overline{h} + Lg)/2$ if the amplitudes are complex, or with a random real number drawn from the same distribution if the amplitudes are real. Summed power is then formed from the equalized amplitudes (as in equation (1) or (2)), and W consecutive power samples are summed into each profile phase bin. The summation continues over F pulses, yielding a folded profile with a phase-dependent mean, $\mu_{C,S}(\theta)$, baseline-subtracted mean, $\mu'_{C,S}(\theta)$, and variance, $\sigma^2_{C,S}(\theta)$, given by using X^2 in place of X in equations (50) and (51):

$$\mu_{C,S}(\theta) = FWN(L^2\overline{h} + Lg)\sum_{l=0}^{L}\frac{(lh(\theta) + g)b_l}{l\overline{h} + g}, \tag{53}$$

$$\mu'_{C,S}(\theta) = FWN(L^2\overline{h} + Lg)\sum_{l=0}^{L}\frac{lh(\theta)b_l}{l\overline{h} + g}, \tag{54}$$

and

$$\begin{aligned}\sigma^2_{C,S}(\theta) = {} & FWN(Lg + L^2\overline{h})^2\sum_{l=0}^{L}\frac{(g^2 + 2lh(\theta)g)b_l}{(g + l\overline{h})^2} \\ & + F^zW^yN^2(Lg + L^2\overline{h})^2\sum_{l=0}^{L}\frac{(g + lh(\theta))^2b_l}{(g + l\overline{h})^2} \\ & - F^zW^yN^2(Lg + L^2\overline{h})^2\left(\sum_{l=0}^{L}\frac{(g + lh(\theta))b_l}{g + l\overline{h}}\right)^2,\end{aligned} \tag{55}$$

where subscripts C and S indicate coherent summation and equalization by scaling, respectively, and where $1 \le y \le 2$ and $1 \le z \le 2$. If $L\overline{h} < g$, as is usually the case, Taylor series allow the profile mean, baseline-subtracted mean and, variance to be expressed as

$$\begin{aligned}\mu_{C,S}(\theta) = {} & FWNL^2((\mu(1) + T_0)h(\theta) + (1 - \mu(1) - T_0)\overline{h}) \\ & + FWNLg,\end{aligned} \tag{56}$$

$$\mu'_{C,S}(\theta) = FWNL^2(\mu(1) + T_0)h(\theta) \tag{57}$$

and

$$\begin{aligned}\sigma^2_{C,S}(\theta) = {} & FWN(L^2g^2 + 2L^3\mu(1)h(\theta)g + 2L^3(1 - \mu(1))\overline{h}g) \\ & + 2FWNL^4(2\mu(1) - 2\mu(2) - T_1)\overline{h}h(\theta) \\ & + FWNL^4(1 + 3\mu(2) - 4\mu(1) + T_2)\overline{h}^2 \\ & + F^zW^yN^2L^4\sigma^2(1)(h(\theta) - \overline{h})^2 \\ & + 2F^zW^yN^2L^4(\mu(1) - \mu(2))(\overline{h}h(\theta) - \overline{h}^2) \\ & - 2F^zW^yN^2L^4(\mu(1)T_0 + T_0^2)(h(\theta) - \overline{h})^2 \\ & + F^zW^yN^2L^4(2T_1 - T_2)h(\theta)^2 \\ & - 2F^zW^yN^2L^4(T_0 + T_1)\overline{h}h(\theta) \\ & + F^zW^yN^2L^4(2T_0 + T_2)\overline{h}^2 \\ & - 2F^zW^yN^2L^3T_0(h(\theta) - \overline{h})g,\end{aligned} \tag{58}$$

where

$$T_0 = \sum_{k=1}^{\infty}\frac{L^k(-\overline{h})^k}{g^k}(\mu(k + 1) - \mu(k)), \tag{59}$$

$$\begin{aligned}T_1 = {} & \sum_{k=1}^{\infty}\frac{L^k(-\overline{h})^k}{g^k}(k\mu(k) - 2(k + 1)\mu(k + 1) \\ & + (k + 2)\mu(k + 2))\end{aligned} \tag{60}$$

and

$$\begin{aligned}T_2 = {} & \sum_{k=1}^{\infty}\frac{L^k(-\overline{h})^k}{g^k}((k + 1)\mu(k) - 2(k + 2)\mu(k + 1) \\ & + (k + 3)\mu(k + 2)),\end{aligned} \tag{61}$$

and the terms in the sums become smaller as k increases.

Since groups of M summed power samples are zapped together, the values of y and z depend on whether groups of samples being folded into each profile bin are zapped independently, and their ranges are $1 \le y \le 2$ and $1 \le z \le 2$. If $M \gg W$, then $y \simeq 2$ (most consecutive samples entering a bin are zapped together). The value of y decreases when M decreases towards W, but with the 'resonances' at which $y = 2$ if zapping and binning are synchronized (e.g. if $M = W$ and the M samples that are considered for zapping are the same set as the W samples that are summed into a bin). When M decreases below W, y decreases until $y \simeq 1$ when $M \ll W$ (many consecutive groups of samples entering a bin are zapped independently), although this limit is reached very slowly unless M is small. A baseband observation

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

of a pulsar with a period of 10 ms using 0.16-MHz-wide frequency channels and 1024 bins across the profile, with $N = 2$ and $M = 1000$, would give $y \simeq 2$, but a similar observation of a pulsar with a period of 10 s would give y a lower value (though not close to 1 unless M were made smaller). If $M \leq WD$, then $z = 1$, where D is the number of bins across the pulse profile and so WD is the number of summed power samples across a single pulse (all groups of samples entering a bin from different pulses are zapped independently). If $M > WD$ but $M \ll FWD$, then $z \simeq 1$, where FWD is the number of samples across an entire profile (many groups of samples entering a bin from different pulses are zapped independently). When M increases above WD and towards FWD, z increases until $z \simeq 2$ when $M \gg FWD$ (most groups of samples entering a bin from different pulses are zapped together). This last situation is undesirable as entire profiles would be zapped together, and $z \simeq 1$ is typical for a pulsar observation.

To equalize variances by adding artificial noise instead of scaling, we add a random number to each summed amplitude. If the amplitudes are complex, the random number is complex and has real and imaginary parts drawn independently from a Gaussian distribution of mean 0 and variance $((L^2 - l^2)\overline{h} + (L - l)g)/2$; if the amplitudes are real, the random number is real and is drawn from the same distribution. A folded profile, formed from summed power as in the previous paragraph, then has a mean, $\mu_{C,A}(\theta)$, baseline-subtracted mean, $\mu'_{C,A}(\theta)$, and variance, $\sigma^2_{C,A}(\theta)$, given by

$$\mu_{C,A}(\theta) = FWN(L^2\mu(2)h(\theta) + L^2(1 - \mu(2))\overline{h} + Lg), \tag{62}$$

$$\mu'_{C,A}(\theta) = FWNL^2\mu(2)h(\theta) \tag{63}$$

and

$$\begin{aligned}\sigma^2_{C,A}(\theta) = {} & FWN(L^2g^2 + 2L^3\mu(2)h(\theta)g + 2L^3(1 - \mu(2))\overline{h}g) \\ & + 2FWNL^4(\mu(2) - \mu(4))\overline{h}h(\theta) \\ & + FWNL^4(1 + \mu(4) - 2\mu(2))\overline{h}^2 \\ & + F^zW^yN^2L^4\sigma^2(2)(h(\theta) - \overline{h})^2,\end{aligned} \tag{64}$$

where subscript A indicates equalization by addition of artificial noise.

If we do not wish to return our data to the time domain as coherently summed amplitudes, we may choose not to apply any variance equalization but still to apply mean equalization to the summed power samples in order to obtain a higher SNR. In this case, we add $N((L^2 - l^2)\overline{h} + (L - l)g)$ to each summed power sample, as this quantity involves the measurable mean of a set of summed power samples using l telescopes. A folded profile, formed from summed power as in the previous paragraphs, then has a mean, $\mu_{C,M}(\theta)$, baseline-subtracted mean, $\mu'_{C,M}(\theta)$, and variance, $\sigma^2_{C,M}(\theta)$, given by

$$\mu_{C,M}(\theta) = FWN(L^2\mu(2)h(\theta) + L^2(1 - \mu(2))\overline{h} + Lg), \tag{65}$$

$$\mu'_{C,M}(\theta) = FWNL^2\mu(2)h(\theta) \tag{66}$$

and

$$\begin{aligned}\sigma^2_{C,M}(\theta) = {} & FWN(L^2\mu(2)g^2 + 2L^3\mu(3)h(\theta)g) \\ & + F^zW^yN^2L^4\sigma^2(2)(h(\theta) - \overline{h})^2,\end{aligned} \tag{67}$$

where subscript M indicates mean equalization.

If there has been very little zapping, we may choose not to use any equalization to mitigate its effects, at the cost of a slightly lower SNR. A folded profile, formed from summed power as in the previous paragraphs, then has a mean, $\mu_C(\theta)$, baseline-subtracted mean, $\mu'_C(\theta)$, and variance, $\sigma^2_C(\theta)$, given by

$$\mu_C(\theta) = FWN(L^2\mu(2)h(\theta) + L\mu(1)g), \tag{68}$$

$$\mu'_C(\theta) = FWNL^2\mu(2)h(\theta) \tag{69}$$

and

$$\begin{aligned}\sigma^2_C(\theta) = {} & FWN(L^2\mu(2)g^2 + 2L^3\mu(3)h(\theta)g) \\ & + F^zW^yN^2L^2\sigma^2(1)g^2 \\ & + 2F^zW^yN^2L^3(\mu(3) - \mu(1)\mu(2))h(\theta)g \\ & + F^zW^yN^2L^4\sigma^2(2)h(\theta)^2.\end{aligned} \tag{70}$$

4.2.2 *Equalization of incoherent observations*

Incoherent summation generally produces a lower profile SNR than coherent summation, but may be necessary if amplitudes cannot be stored or if the alignment of time series between telescopes cannot be made accurate enough to guarantee coherence. With incoherent summation, $\{X_l\}$ represents a set of power values summed over l telescopes. This set has a measurable mean, $\mathrm{E}\{X_l\}$, of $Nl(\overline{h} + g)$ and a measurable variance, $\mathrm{Var}\{X_l\}$, of $Nl(2\overline{h}g + g^2)$. These quantities should be measured using a sufficient number of samples to make stable estimates. Each summed power value is associated with its own value of l after zapping, and we can calculate the means and variances of the summed power values for all combinations of telescopes using separate measurements from each telescope, since each quantity varies with a single power of l.

To equalize variances by scaling, we add $N(L - l)(\overline{h} + g)$ to each summed power sample and then multiply the result by $\sqrt{L/l}$, except when $l = 0$. We replace any sample that has been zapped at all telescopes with a random real number drawn from a gamma distribution of mean $NL(\overline{h} + g)$ (shifted from $NL\sqrt{2\overline{h}g + g^2}$) and variance $NL(2\overline{h}g + g^2)$. A folded profile, formed as in Section 4.2.1, then has a mean, $\mu_{I,S}(\theta)$, baseline-subtracted mean, $\mu'_{I,S}(\theta)$, and variance, $\sigma^2_{I,S}(\theta)$, given by using equations (50) and (51):

$$\mu_{I,S}(\theta) = FWNL\left(\mu\left(\frac{1}{2}\right)h(\theta) + \left(1 - \mu\left(\frac{1}{2}\right)\right)\overline{h} + g\right), \tag{71}$$

$$\mu'_{I,S}(\theta) = FWNL\mu\left(\frac{1}{2}\right)h(\theta) \tag{72}$$

and

$$\begin{aligned}\sigma^2_{I,S}(\theta) = {} & FWNL(g^2 + 2(1 - q^L)h(\theta)g + 2q^L\overline{h}g) \\ & + F^zW^yN^2L^2\sigma^2\left(\frac{1}{2}\right)(h(\theta) - \overline{h})^2,\end{aligned} \tag{73}$$

where subscript I indicates incoherent summation.

To equalize variances by adding artificial noise instead of scaling, we add a random real number to each summed power sample. This number is drawn from a gamma distribution of mean $N(L - l)(\overline{h} + g)$ (shifted from $N(L - l)\sqrt{2\overline{h}g + g^2}$) and variance $N(L - l)(2\overline{h}g + g^2)$. A folded profile, formed as in Section 4.2.1, then has a mean, $\mu_{I,A}(\theta)$, baseline-subtracted mean, $\mu'_{I,A}(\theta)$, and variance, $\sigma^2_{I,A}(\theta)$, given by

$$\mu_{I,A}(\theta) = FWNL(\mu(1)h(\theta) + (1 - \mu(1))\overline{h} + g), \tag{74}$$

$$\mu'_{I,A}(\theta) = FWNL\mu(1)h(\theta) \tag{75}$$

and

$$\begin{aligned}\sigma^2_{I,A}(\theta) = {} & FWNL(g^2 + 2\mu(1)h(\theta)g + 2(1 - \mu(1))\overline{h}g) \\ & + F^zW^yN^2L^2\sigma^2(1)(h(\theta) - \overline{h})^2.\end{aligned} \tag{76}$$

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

We cannot return our incoherently summed power samples to meaningful amplitudes, and so we may choose not to apply any variance equalization but still to apply mean equalization in order to obtain a higher SNR. In this case, we add $N(L-l)(\overline{h}+g)$ to each summed power sample. A folded profile, formed as in Section 4.2.1, then has a mean, $\mu_{I,M}(\theta)$, baseline-subtracted mean, $\mu'_{I,M}(\theta)$, and variance, $\sigma^2_{I,M}(\theta)$, given by

$$\mu_{I,M}(\theta) = FWNL(\mu(1)h(\theta) + (1-\mu(1))\overline{h} + g), \tag{77}$$

$$\mu'_{I,M}(\theta) = FWNL\mu(1)h(\theta) \tag{78}$$

and

$$\sigma^2_{I,M}(\theta) = FWNL\mu(1)(g^2 + 2h(\theta)g) + F^zW^yN^2L^2\sigma^2(1)(h(\theta)-\overline{h})^2. \tag{79}$$

If there has been very little zapping, we may choose not to use any equalization to mitigate its effects, at the cost of a slightly lower SNR. A folded profile, formed as in Section 4.2.1, then has a mean, $\mu_I(\theta)$, baseline-subtracted mean, $\mu'_I(\theta)$, and variance, $\sigma^2_I(\theta)$, given by

$$\mu_I(\theta) = FWNL\mu(1)(h(\theta)+g), \tag{80}$$

$$\mu'_I(\theta) = FWNL\mu(1)h(\theta) \tag{81}$$

and

$$\sigma^2_I(\theta) = FWNL\mu(1)(g^2 + 2h(\theta)g) + F^zW^yN^2L^2\sigma^2(1)(g+h(\theta))^2. \tag{82}$$

4.3 Comparison of equalization methods for zapped data

The choice of equalization method may depend on the fraction of data that are zapped, which can vary with time, frequency, and telescope environment, and the choice may also depend on whether data are summed coherently or incoherently. In most cases, mean equalization and variance equalization by scaling produce a higher profile SNR than variance equalization by addition of artificial noise and no equalization. This is because artificial noise introduces additional variance without increasing the mean, as does a failure to equalize the power baseline.

We focus on the regime of $\overline{h} \ll h(\theta) \ll g$, which is typical of pulsar observations at phase values where the pulse is visible. The first inequality comes about because a normal pulsar gives no emission for the majority of its period, so the pulse itself is usually well above the average emission strength. The second inequality occurs because these are quantities in a single sample at a single telescope, in which noise usually dominates over source contribution (where noise does not dominate, coherent summation loses its SNR advantage over incoherent summation even if there is no zapping). Figs 1–3 use $h = 0.01$, $\overline{h} = 0.0001$, $g = 1$, $N = 2$, $W = 2$, $y = 2$, $F = 1000$, and $z = 1$ (representing an observation of a pulsar with a period of 10 ms) and show the relationships between profile SNR ($\mu'(\theta)/\sigma(\theta)$) and fraction of summed power samples zapped at each telescope (q) for 2, 5, and 100 identical telescopes in independent interference environments (this example also applies to identical groups of telescopes in which each group is zapped together and each group is in an independent interference environment; see Taylor et al. 2019 and Nita & Hellbourg 2020 for calculations of the estimator using multiple receivers in the same environment). Coherent and incoherent summation are shown, with all four methods of equalization.

Mean equalization usually gives the highest SNR, although the difference between mean equalization and variance equalization by

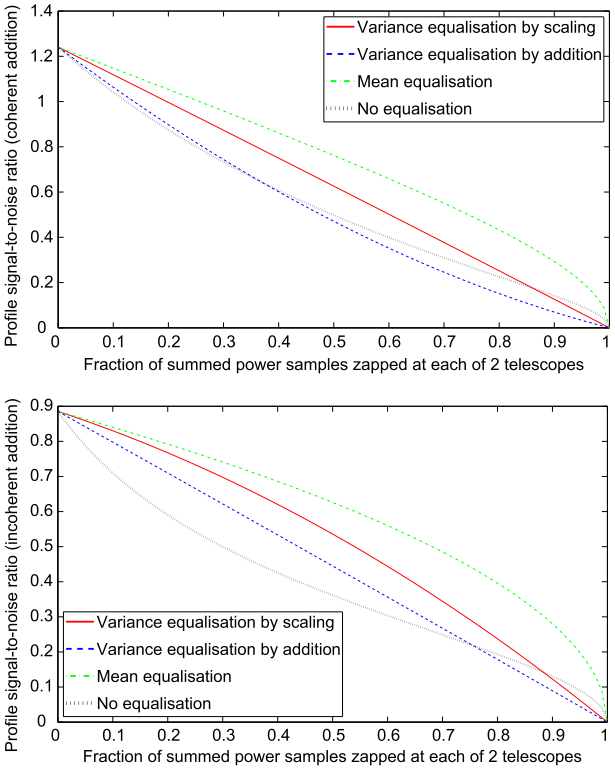


Figure 1. Profile SNR as a function of the fraction of summed power samples zapped due to interference, using coherent (top) and incoherent (bottom) summation of signals from two identical telescopes in independent interference environments, with four different methods of equalization (see text for other parameter values).

scaling decreases as more telescopes are added, and the second method is very slightly better for coherent summation of 100 telescopes with a small or moderate amount of zapping. The scaling method may be preferred for coherent summation in particular, because it avoids artefacts when returning amplitudes to the time domain (see Section 4.2), and the penalty in SNR is small unless there is a large amount of zapping. Mean equalization can fall behind if W or N increase sufficiently to cause the signal to make a substantial contribution to variance.

Variance equalization by addition of artificial noise lags some way behind the other active methods (except in the case of a single telescope, for which it is identical to variance equalization by scaling), but it gives a better SNR than the passive method of no equalization when summing two telescopes with a small amount of zapping, as well as preventing artefacts when returning coherently summed amplitudes to the time domain. With a larger number of telescopes, it only retains this SNR advantage for incoherent summation. It can be applied at each telescope without reference to the others, so it is computationally simpler than the scaling method and may be preferred if equalization needs to be done before the individual signals are combined – for example, interference-free signals from widely spaced telescopes may be needed in order to synchronize their summation, after which the artificial noise could be replaced by a different equalization method.

The method of no equalization gives a relatively poor SNR for incoherent summation, but its performance for coherent summation improves as the number of telescopes increases. It is the simplest method, because it does not require replacement values to be computed, and may be preferred when there is very little zapping, or

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

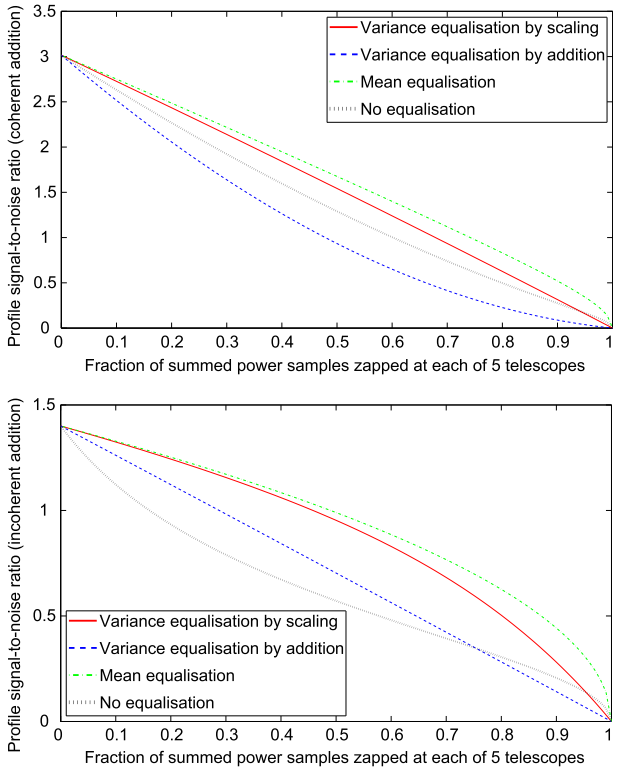


Figure 2. Profile SNR as a function of the fraction of summed power samples zapped due to interference, using coherent (top) and incoherent (bottom) summation of signals from five identical telescopes in independent interference environments, with four different methods of equalization (see text for other parameter values).

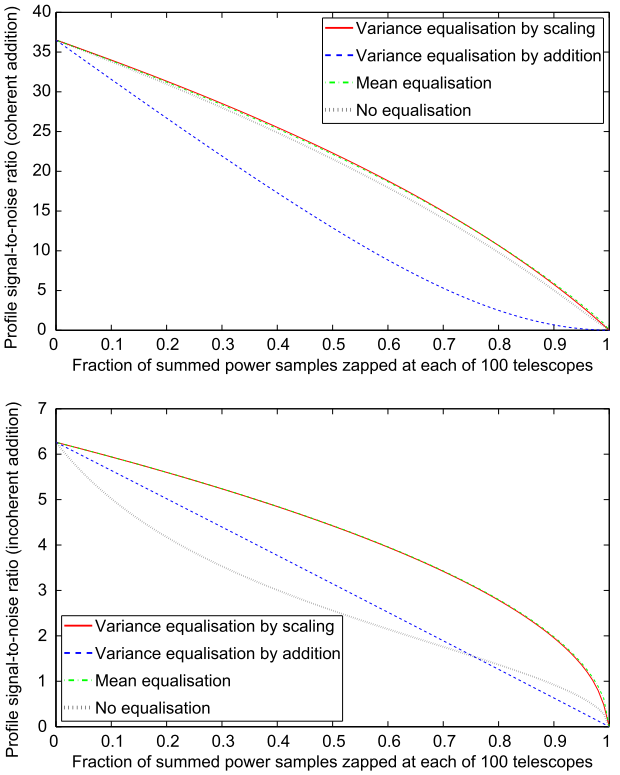


Figure 3. Profile SNR as a function of the fraction of summed power samples zapped due to interference, using coherent (top) and incoherent (bottom) summation of signals from 100 identical telescopes in independent interference environments, with four different methods of equalization (see text for other parameter values).

in the coherent summation of a large number of telescopes. However, its unequal power baseline can cause its SNR to decrease below that of all other methods when W or N increases, particularly for incoherent summation (e.g. when $W = 2000$ and $y = 1.5$, representing an observation of a pulsar with a period of 10 s).

All four methods of equalization produce profile means that are linear functions of $h(\theta)$, so they do not make systematic changes to profile shape as long as the distribution of all profile noise is close to Gaussian. Even Gaussian profile noise with a phase-dependent variance produces Gaussian noise of constant variance in each complex-valued bin of a profile's DFT, which means that the standard method of frequency-domain pulsar timing should not produce unwanted correlations between timing residuals (often called 'red noise') when using equalized signals (Taylor 1992). The main danger to profile shape, other than zapping of the pulsar contribution itself, is non-Gaussian noise. We rely on the accumulation of data into each profile phase bin to make the noise approximately Gaussian, in line with the central limit theorem (even though the equations for profile mean and variance above do not depend on the noise distribution). Substantial non-Gaussianity could arise if q were very close to 1, or if q were very close to 0 when $h \gtrsim g$. It could also occur if the duration and bandwidth of a typical burst of interference were not much less than the duration and bandwidth over which the profile was folded, as there would then be few independent instances of zapping within each profile, and so the binomial distribution of zapping might not resemble a Gaussian shape. Similarly, it could happen if the time and frequency resolutions of zapping were not much less than the folding duration and bandwidth. As a rough guide, fewer than 100 independent instances of zapping within a profile may be too few

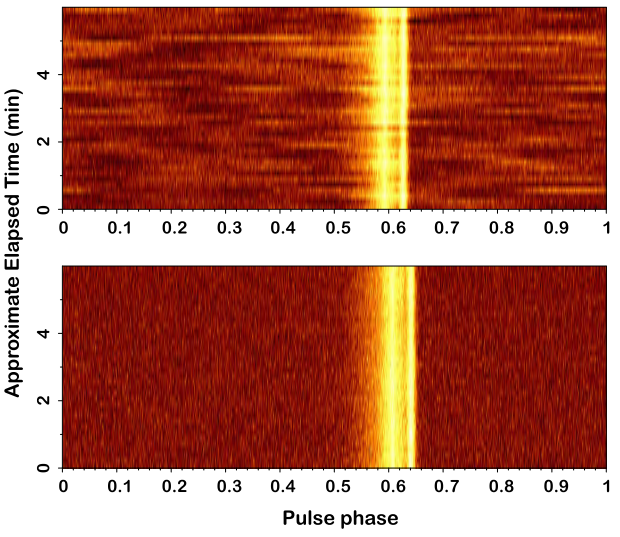


Figure 4. The pulse profile of PSR J1022+1001, with brightness representing power, during 6 min of a LEAP observation with a bandwidth of 128 MHz, both without (top) and with (bottom) interference removal using spectral kurtosis. The observation is a coherent summation of signals from the Jodrell Bank, Effelsberg, Nançay, and Westerbork radio telescopes, with spectral kurtosis applied to Nançay using $M = 1000$, $N = 2$, and $\eta = 3$, and rejected data replaced by artificial Gaussian noise. This high-resolution zapping makes the observation usable while sacrificing only a small fraction of data. Reproduced from Bassa et al. (2016, fig. 6).

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

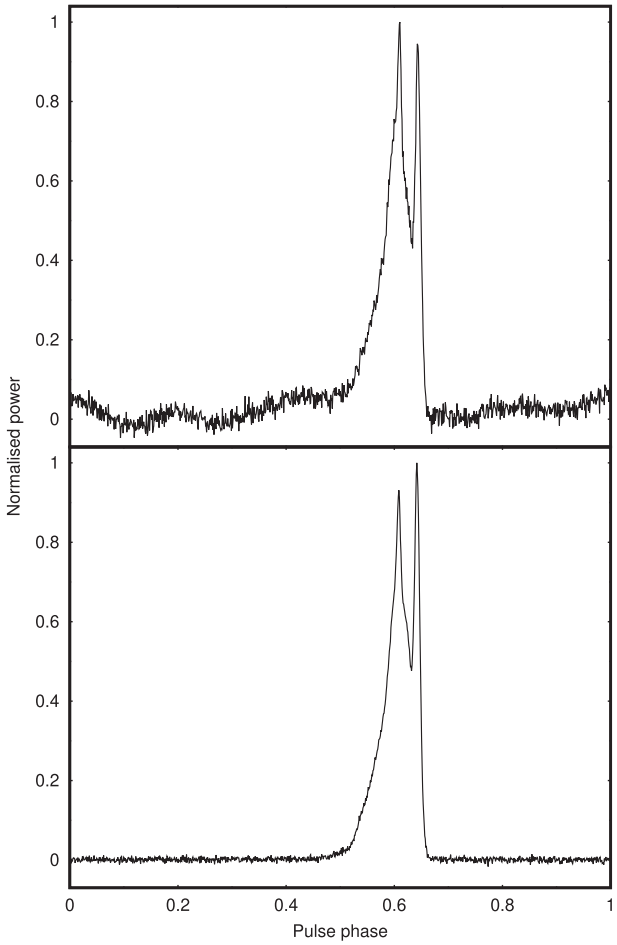


Figure 5. The pulse profile of PSR J1022+1001, folded over 30 min of a LEAP observation with a bandwidth of 96 MHz, both without (top) and with (bottom) interference removal using spectral kurtosis. The top panel shows the contribution from the Nançay radio telescope only, while the bottom panel shows the coherent summation of signals from the Jodrell Bank, Effelsberg, Nançay, and Westerbork radio telescopes, with spectral kurtosis applied to Nançay using $M = 1000$, $N = 2$, and $\eta = 3$, and rejected data replaced by artificial Gaussian noise. Zapping restores the profile shape and flattens the power baseline, allowing the pulse arrival time to be measured with a high estimated accuracy of 0.25 μs. This compares to 1.35 μs for Nançay alone and without zapping.

when $q = 0.1$. With or without zapping, it is worth noting that profile noise may be substantially skewed and non-Gaussian if few (less than about 50) amplitude samples contribute to the power in each profile bin.

5 INTERFERENCE REMOVAL FROM PULSAR OBSERVATIONS

Spectral kurtosis has been employed successfully by the LEAP project, which makes astronomical observations of pulsars using up to five radio telescopes simultaneously (Bassa et al. 2016; Smits et al. 2017). The aim of the project is to measure the times of arrival of pulses with sufficient accuracy to detect variation that is characteristic of the influence of gravitational waves, thereby measuring the strength of a background of low-frequency waves that is believed to permeate the Solar System from distant sources such as binary supermassive black holes (Hellings & Downs 1983). The

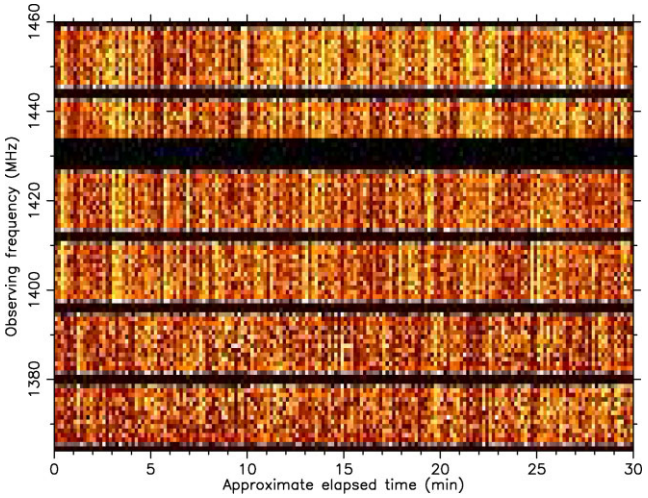


Figure 6. The dynamic spectrum of the Nançay observation shown in Fig. 5, with brightness representing power, before interference removal. Spectral kurtosis is applied before the pulsar observation is folded, and can zap the persistent broadband interference seen here.

signal from each telescope is converted to the baseband frequency range and sampled at the Nyquist rate to enable coherent summation (pp. 117–120 Lorimer & Kramer 2005), allowing spectral kurtosis to be used effectively alongside a simpler method that zaps portions of the signal whose power deviates greatly from an expected value or from the power of neighbouring portions (section 4.5 of Bassa et al. 2016). Each telescope's signal is recorded digitally using eight sampling bits, calibrated for polarization accuracy and then zapped if necessary, before the stored signals are summed with their amplitudes calibrated to maximize the SNR of the observation. Pulse profiles, showing the average radio emission from a pulsar as it rotates, can then be produced and timed.

Fig. 4, reproduced from Bassa et al. (2016), shows the improvement in the pulse profile of PSR J1022+1001 achieved by zapping a signal from the Nançay radio telescope using spectral kurtosis and replacing the rejected data with artificial Gaussian noise. The 6-min segment of this LEAP observation used four telescopes and covered a frequency range of 1332–1460 MHz. Each measurement of the estimator used 1000 power values averaged over two complex polarization channels ($M = 1000$ and $N = 2$), giving zapping resolutions of 6.25 ms and 0.16 MHz. The estimator thresholds were set using $\eta = 3$, meaning that 0.27 per cent of good data from Nançay were zapped. Through the application of spectral kurtosis, little data were lost and an observation that was riddled with interference became suitable for high-precision pulsar timing.

Fig. 5 shows the improvement in the pulse profile shape of PSR J1022+1001 achieved by zapping a signal from the Nançay radio telescope using the same parameters and resolutions as those above, while Fig. 6 reveals the persistent broadband interference that was removed. This 30-min LEAP observation took place on 2013 July 27, used four telescopes and covered a frequency range of 1364–1460 MHz. Following the application of spectral kurtosis, the PSRCHIVE tool PAT was used to align the LEAP profile with a template profile of high S/N using the Fourier phase gradient between them (Taylor 1992). This gave an estimated uncertainty of only 0.25 μs in the pulse arrival time associated with the zapped tied-array profile, compared to 1.35 μs for the non-zapped Nançay profile alone.

A further LEAP observation of PSR J1022+1001 demonstrates that spectral kurtosis recovers the unique profile shape that is critical

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

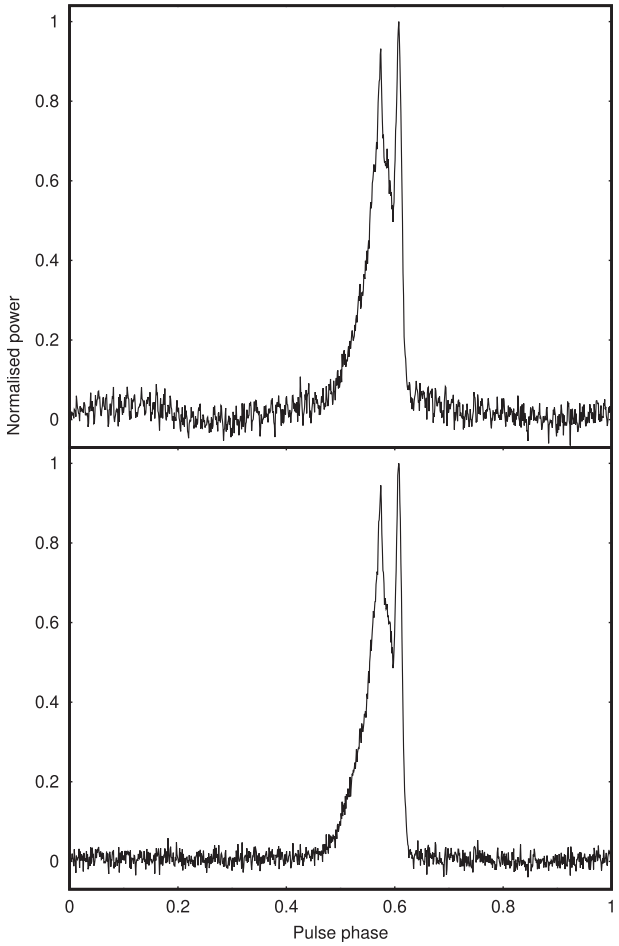


Figure 7. The pulse profile of PSR J1022+1001, folded over 60 min of a LEAP observation with a bandwidth of 128 MHz, both without (top) and with (bottom) interference removal using spectral kurtosis. Both panels show the contribution from the Nançay radio telescope only, with spectral kurtosis applied using $M = 1000$, $N = 2$, and $\eta = 3$, and rejected data replaced by artificial Gaussian noise. While zapping of the folded observation is ineffective at removing this interference, spectral kurtosis improves the estimated accuracy of the pulse arrival time measured at Nançay from 1.22 to 0.70 μs.

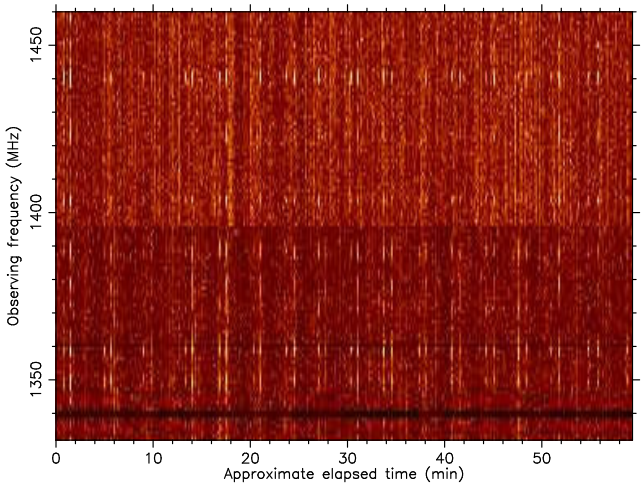


Figure 8. The dynamic spectrum of the Nançay observation shown in Fig. 7, with brightness representing power, before interference removal. The persistent broadband interference seen here cannot be separated from the pulsar signal using a folded observation, but can be zapped by spectral kurtosis using fine time and frequency resolutions simultaneously.

to high-precision pulsar timing. Fig. 7 shows the profile of PSR J1022+1001 produced from a 60-min observation made on 2021 May 15 by the Nançay radio telescope over a frequency range of 1332–1460 MHz, before and after zapping using the same parameters and resolutions as those above. Fig. 8 shows that the interference in the observation was persistent and broadband, and zapping of the folded observation using the PSRCHIVE tool PAZ had little effect. Spectral kurtosis zapped around 0.6 per cent of the data, and reduced the estimated timing uncertainty of the Nançay profile from 1.22 to 0.70 μs according to the PAT tool. Fig. 9 shows the residual profile produced by subtracting this Nançay observation from the four-telescope LEAP observation shown in Fig. 5, again using PAT, with the lack of residual structure indicating that spectral kurtosis did not alter the shape of the pulse profile.

6 CONCLUSIONS

This paper provides a recipe for the implementation of the spectral kurtosis method from start to finish, allowing signal interference to be zapped from real or complex time series data stored as either amplitudes or power (Section 2). The frequentist nature of spectral kurtosis makes it effective without prior knowledge of the interference that will be encountered, so it is widely applicable rather than being ideal in specific situations (Section 3). We have shown its success in enabling an accurate radio-frequency array observation of a pulsar in the presence of interference local to one telescope, allowing signals from multiple widely spaced telescopes to be combined with only a very small loss of usable information and without any apparent detriment to the shape of the pulse profile (Section 5). The preservation of the unique profile signature of each pulsar is crucial for precise timing of its rotation, and the timing information from the cleaned observations is being used in a long-term project to detect gravitational waves.

When zapping data that contain a rapidly varying signal such as pulsar emission, it is important that the estimator does not recognize the signal amplitudes as non-Gaussian, as the spectral kurtosis procedure would then remove the information of interest. Observers should therefore ensure that the time and frequency resolutions of an observation are too fine to allow single pulses to be detected (Section 3). In order to maintain a Gaussian noise distribution, the time and frequency resolutions of zapping should be much less than the duration and bandwidth of an observation or folded pulse profile, so that there are many independent opportunities for zapping within the observation (Section 4).

The quality of an observation made using an array of telescopes in independent interference environments is improved by compensating for zapped data, regardless of the zapping technique used, and the methods of compensation are applicable to any widely spaced array (Section 4). The highest SNR is usually obtained by mean equalization: equalizing the mean of the summed power so that its baseline level remains constant over time. Mean equalization is the most appropriate method for an incoherently summed signal. However, if the signal is coherently summed and its amplitudes are stored so that its time and frequency resolutions can be adjusted later, it is better to apply variance equalization by scaling: equalizing the variance of the summed amplitudes over time, which also results in a constant power baseline and avoids unwanted artefacts

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023

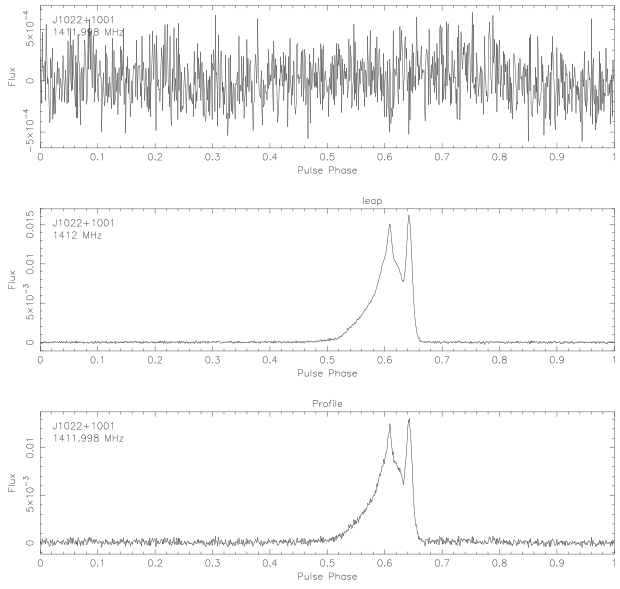


Figure 9. The pulse profile of PSR J1022+1001, with a residual profile (top) showing the difference between the four-telescope LEAP observation from Fig. 5 (middle) and the Nançay observation from Fig. 7 (bottom) after interference removal using spectral kurtosis. The Nançay observation has been reduced to the same frequency range as the earlier LEAP observation, and its profile has been scaled and shifted when producing the residual profile. Interference removal from the Nançay telescope does not appear to have changed the pulse profile shape, judging by the lack of structure in the residuals between these two observations that were made 8-yr apart.

if the amplitudes are transformed to different time and frequency resolutions. The SNR after variance equalization by scaling is usually slightly less than the SNR after mean equalization, but the difference is small when there is either a small-to-moderate amount of zapping or a large number of telescopes. Variance equalization by scaling is the most appropriate method when the signal may be summed coherently or incoherently in different parts of an observation. The alternative method of variance equalization by addition of artificial noise may be needed to allow signals from multiple telescopes to be synchronized, and can then be replaced with another method to improve SNR.

ACKNOWLEDGEMENTS

The European Pulsar Timing Array (EPTA) is a collaboration of European institutes working towards the direct detection of low-frequency gravitational waves, for which it has implemented the Large European Array for Pulsars (LEAP). The authors acknowledge the support of colleagues in the EPTA, and MP expresses gratitude for the patience of co-authors and family while the writing of this paper was completed. The work reported here has been funded by the European Research Council Advanced Grant 'LEAP', grant agreement ID 227947 (Principal Investigator: M. Kramer). KL and MK are supported by the European Research Council Synergy Grant 'BlackHoleCam', grant agreement ID 610058 (Principal Investigator: M. Kramer). KJL acknowledges support from the National Basic Research Program of China, 973 Program, 2015CB857101 and NSFC 11373011. The *Effelsberg Radio Telescope* is operated by the Max-Planck-Institut für Radioastronomie in Germany. The *Westerbork Synthesis Radio Telescope* is operated by the Netherlands Institute for Radio Astronomy (ASTRON) with support from the Netherlands Organisation for Scientific Research (NWO). The Nançay Radio Observatory is operated by the Paris Observatory and associated with the Centre National de la Recherche Scientifique in France. Pulsar research at the Jodrell Bank Centre for Astrophysics, and the observations using the *Lovell Telescope*, are supported by a consolidated grant from the Science and Technology Facilities Council (STFC) in the UK. The *Sardinia Radio Telescope* is operated by the Istituto Nazionale di Astrofisica (INAF) in Italy, and was undergoing its astronomical validation phase when the observations used in this paper were made.

DATA AVAILABILITY

The data and software programmes underlying this article will be shared on a reasonable request to the corresponding author.

REFERENCES

Ait-Allal D., Weber R., Dumez-Viou C., Cognard I., Theureau G., 2010, in Proc. 18th European Signal Processing Conference. p. 1841
Bassa C. G. et al., 2016, MNRAS, 456, 2196
Bateman H., Erdélyi A., 1953, Higher Transcendental Functions, 1st edn. McGraw-Hill, New York
Bracewell R. N., 2000, The Fourier Transform and Its Applications, 3rd edn., Circuits and Systems. McGraw-Hill, New York
Briggs F. H., Bell J. F., Kesteven M. J., 2000, AJ, 120, 3351
Czech D., Mishra A., Inggs M., 2018, Rad. Sci., 53, 656
Dolch T. et al., 2014, ApJ, 794, 21
Dwyer R., 1983, in Proc. ICASSP '83. IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 8. p. 607
Fridman P., 2009, A&A, 502, 401
Gary D. E., Liu Z., Nita G. M., 2010, PASP, 122, 560
Hellings R. W., Downs G. S., 1983, ApJ, 265, L39
Kar A., Kuske A., Hawkins L., Prestage R., Smith E., 2019, in American Astronomical Society Meeting Abstracts, Vol. 233. p. 152.01
Kendall M. G., Stuart A., Ord J. K., 1994, Kendall's Advanced Theory of Statistics, 6th edn. Distribution Theory, Edward Arnold, London
Lam M. T., 2016, PhD thesis, Cornell Univ.
Lorimer D. R., Kramer M., 2005, Handbook of Pulsar Astronomy, Cambridge Observing Handbooks for Research Astronomers. Cambridge Univ. Press, Cambridge
Nita G. M., 2016, MNRAS, 458, 2530
Nita G. M., Gary D. E., 2010a, PASP, 122, 595
Nita G. M., Gary D. E., 2010b, MNRAS, 406, L60
Nita G. M., Gary D. E., 2016, J. Geophys. Res. Space Phys., 121, 7353
Nita G. M., Hellbourg G., 2020, in Proc. 2020 XXXIIIrd General Assembly and Scientific Symposium of the International Union of Radio Science. p. 1
Nita G. M., Gary D. E., Liu Z., Hurford G. J., White S. M., 2007, PASP, 119, 805
Nita G. M., Keimpema A., Paragi Z., 2019, J. Astron. Instrum., 8, 1940008
Pearson K., 1895, Phil. Trans. R. Soc., 186, 343
Pearson K., 1901, Phil. Trans. R. Soc., 197, 443
Shannon C. E., 1949, Proc. Inst. Radio Eng., 37, 10
Smits R. et al., 2017, Astron. Comput., 19, 66
Student, 1908, Biometrika, 6, 1
Taylor J. H., 1992, Phil. Trans. R. Soc., 341, 117
Taylor J., Denman N., Bandura K., Berger P., Masui K., Renard A., Tretyakov I., Vanderlinde K., 2019, J. Astron. Instrument., 8, 1940004
van Straten W., 2013, ApJS, 204, 13
van Straten W., Demorest P., Oslowski S., 2012, Astron. Res. Technol., 9, 237

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

Downloaded from https://academic.oup.com/mnras/article/510/2/1597/6445038 by INIST-CNRS INSU user on 23 March 2023