Astron. Astrophys. 221, 180–190 (1989)

ASTRONOMY
AND
ASTROPHYSICS

# A powerful test for weak periodic signals with unknown light curve shape in sparse data

**O.C. De Jager**[1], **J.W.H. Swanepoel**[2], **and B.C. Raubenheimer**[1]

[1] Department of Physics, PU for CHE, Potchefstroom 2520, South Africa
[2] Department of Statistics, PU for CHE, Potchefstroom 2520, South Africa

Received November 21, 1988; accepted March 2, 1989

**Summary.** A problem with most tests for periodicity is that they are powerful enough to detect only certain kinds of periodic shapes (or "light curves") in the case of weak signals. This causes a selection effect with the identification of weak periodic signals. Furthermore, the subjective choice of a test after inspection of the data can cause the identification of false sources. A new test for uniformity called the "$H$-test" is derived for which the probability distribution is an exponential function. This test is shown to have a very good power against most light curve shapes encountered in X- and $\gamma$-ray Astronomy and therefore makes the detection of sources with a larger variety of shapes possible. The use of the $H$-test is suggested if no a priori information about the light curve shape is available. It is also shown how the probability distribution of the test statistics changes when a periodicity search is conducted using very small steps in the period or frequency range. The flux sensitivity for various light curve shapes is also derived for a few tests and this flux is on average a minimum for the $H$-test.

**Key words:** tests for uniformity – data analysis – gamma rays – pulsars

## 1. Introduction

In X-ray and $\gamma$-ray astronomy one sometimes has to identify a periodicity in data dominated by counting statistics. In the pre-analysis one folds (or superimpose) the arrival times on the phase interval (0, 1) (or (0, $2\pi$)) using the appropriate parameters (e.g. period, period derivative etc.) of the periodic source under study. This interval represents one full period of rotation and in the absence of any periodicity, the folded events will be uniformly distributed if the source frequency is larger than the mean count rate (i.e. "sparse data") and one can test for periodicity by applying a *test for uniformity* on the circle.

Beran (1969) derived a complete class of tests for uniformity on the circle and some of the most important tests known to the Astrophysics community are special cases from his class (see Sect. 2): Two important special cases are Pearson's $\chi^2$-test with $K$ bins (see Leahy et al., 1983a,b), which seems to be the most popular and the $Z_m^2$-test (Buccheri et al., 1983) which involves the sum of the Fourier powers of the first $m$ harmonics. Although the latter test is already an improvement over Pearson's test (because the $Z_m^2$-test is bin-free), both suffer with respect to the choice of a smoothing parameter (the number of bins $K$ or harmonics $m$). In fact, it was shown by De Jager et al. (1985), Protheroe (1988) and Buccheri and De Jager (1989) that the capability or "power" of these tests to detect specific signal shapes are strongly dependent on the choice of the corresponding smoothing parameter if the signal is weak. In fact, to detect a broad peak on the light curve one should use small values for $K$ or $m$ but narrow and/or multiple peaks are best identified when using large values for $K$ or $m$. However, it is impossible to make the correct choice for the smoothing parameter if the light curve shape is unknown a priori. To avoid this problem Buccheri et al. (1983) and Protheroe (1988) suggested using $Z_2^2$ (i.e. the $Z_m^2$-test with $m = 2$) as a general omnibus test (i.e. a test which is very powerful to detect any light curve shape). The rational is the following: The test will be nearly as powerful as the Rayleigh test (Mardia, 1972) to detect a broad peak and it will also be more powerful than the Rayleigh test when considering narrower and/or double peaks. The Rayleigh test statistic itself is a special case of $Z_m^2$ which is obtained by taking $m = 1$. However, from the power study done in Sect. 4 it will become clear that the omnibus property of $Z_2^2$ is not as good as thought, the reason being that the smoothing parameter is fixed at $m = 2$ and no flexibility is allowed.

In the search for a good omnibus test for uniformity, one may ask why X-ray astronomy is so successful with Pearson's $\chi^2$-test with $K$ between 16 and 20 so that this test may serve as the desired omnibus test? The only reason for its success is that the signal strengths involved in X-ray Astronomy are "large" enough to warrant very significant detections when only moderate sample sizes are involved. However, Pearson's test enjoys no exclusive status in this regard – any other "consistent test" (which will be defined later) would do just as well in the case of strong signals. In the case of weak signals we observe that all tests are not equally powerful against a given light curve shape and each test known up to now perform good for a certain light curve shape, but loses sensitivity (relative to other tests) for other light curve shapes. This behaviour is better illustrated in Sect. 4. The selected omnibus test should have a good power relative to all other tests for any realistic light curve shape.

Another important condition for the success of any test is its "consistency" property: A test is consistent against a given light

---

*Send offprint requests to:* O.C. De Jager curve shape if the significance of the detection improves as more data are added under fixed conditions. The test is "inconsistent" if the latter property does not hold. Buccheri and Ögelman (1988) have shown that the "$y$-test" of Harwit et al. (1987) is inconsistent against a wide variety of light curve shapes and peak positions on the light curve. The Rayleigh test is also inconsistent against light curve shapes having no power in the fundamental harmonic and $Z_2^2$ is inconsistent against light curves with powerless fundamental and second harmonics. The selected omnibus test should be consistent against all realistic light curve shapes.

This paper is aimed at the construction of a consistent and powerful (against all realistic light curve shapes) test for uniformity. Its power should be comparable to that of $\chi^2$ with large $K$ and $Z_m^2$ with large $m$ for narrow and/or multiple peaked light curves and its power should also be comparable to that of the Rayleigh test for sinusoids. Thus, the test should be independent of a subjectively chosen smoothing parameter and it should also be simple to calculate, without requiring $\simeq n^2$ calculations (where $n$ is the number of events). We were able to construct such a new test for periodicity (the "$H$-test") which is aimed at the detection of periodic sources in sparse data for which the light curve shape is unknown a priori or time variable, while the signal may be weak. This test statistic involves the $Z_m^2$ statistic as its basis, but $m$ is chosen optimally, objectively and automatically using the experimental data as the only input. The probability distribution of the test statistic is easily expressed by an exponential function and the number of steps required to calculate the statistic $H$ is of the order of $40n$. The philosophy behind this test was briefly discussed by De Jager et al. (1988), who have called it the "$H_m$-test", and they used it for the identification of TeV $\gamma$-rays from PSR 1509-58. However, the more simplified interpretation of the distribution of $H$ (as presented in this paper) also leads to a general improvement in the power. Buccheri and De Jager (1989) also gave preliminary results about this new testing procedure. The power of the $H$-test is also compared with that of other tests for a wide range of possible light curve shapes and on the basis of these results it follows that the $H$-test sets a standard for X- and $\gamma$-ray Astronomy which can only be rivalled if one is certain (a priori) which light curve shape is to be expected, so that an appropriate test (e.g. a test from Beran's class) may be chosen.

If the source period is uncertain, one should fold the arrival times with a series of periods in a certain chosen range. It is acknowledged by researchers that the number of independent Fourier spacings (IFS) which fits into the period range equals the number of "independent trials" performed and should be taken into account when evaluating the significance. However, this is not the complete truth if one evaluates the test statistic more than once within an IFS. One has then to multiply the $p$-level (or "chance probability for uniformity" as known to the Astrophysics community) with a factor $r > 1$ which reduces the apparent significance (De Jager, 1987; De Jager et al., 1988 and Buccheri and De Jager, 1989). This factor for oversampling will also be derived for the $H$-test (see Sect. 5).

The power of a test may be used to calculate the minimum flux sensitivity attainable by any test for uniformity for given experimental parameters. This will be done in Sect. 6 for a number of tests including the $H$-test. Due to the latter's good overall power performance, the minimum flux attainable by the $H$-test is also relatively small for most light curve shapes which are physically acceptable.

## 2. Tests for uniformity on the circle

Consider a set of arrival times $t_i$ ($i = 1, \ldots, n$). Assume firstly that the frequency parameters of the source are known so that one may obtain the phases $\theta_i$ by folding the $t_i$'s. Let also $\theta \in (0, 2\pi)$. To test for the presence of a periodic signal one should actually test for the absence of such a signal (i.e. the null hypothesis $H_0$), which is usually characterized by a test for uniformity. Let $f(\theta)$ be the unknown periodic density function (or light curve) of the folded arrival times. The null hypothesis may then be written as follows

$$H_0: f(\theta) = 1/2\pi \quad \text{with} \quad \theta \in (0, 2\pi) \tag{1}$$

In the presence of a periodic signal $f(\theta)$ will differ from Eq. (1) and may in general be written as follows under the alternative hypothesis

$$H_A: f(\theta) = p f_s(\theta) + (1-p)/2\pi \tag{2}$$

Here $p$ is the strength of the periodic signal and the source function $f_s(\theta)$ gives the relative radiation intensity as a function of $\theta$. To determine whether a periodic signal is present in the data, one should test the hypothesis $H_0: p = 0$ against $H_A: p > 0$ and to do this we use a measure of the "distance" between $f(\theta)$ and the uniform density $1/2\pi$. A good measure is given by the functional

$$\psi(f) = \int_0^{2\pi} (f(\theta) - 1/2\pi)^2 \, d\theta \tag{3}$$

Clearly, the hypothesis for uniformity should be rejected when $\psi(f)$ is too large. However, $\psi(f)$ is unknown (since $f(\theta)$ is unknown) and one therefore has to estimate this functional. This can be "crudely" done by replacing $f$ in $\psi(f)$ by a consistent estimator $\hat{f}_h$ and then use $\psi(\hat{f}_h)$ (or some equivalent form thereof) as a test statistic for uniformity (Beran, 1969): The statistical literature gives an extensive discussion of various choices of $\hat{f}_h$ and their properties under the heading "density estimators" (see De Jager et al., 1986, for their implementation to periodic analyses). These estimators are always characterized by some smoothing parameter $h$. In case of a histogram, $K = h$ gives the fixed corresponding number of bins for the histogram and Eq. (3) results within constants to the well known $\chi^2$ statistic of Pearson which is asymptotically $\chi^2$ distributed with $K-1$ degrees of freedom under $H_0$:

$$\chi^2_{K-1} = 2\pi n \psi(\hat{f}_K) = \sum_{j=1}^{K} (X_j - n/K)^2/(n/K) \tag{4}$$

Here $X_j$ is the number of events in the $j$'th bin. However, this statistic is dependent on the choice of the bin positions so that it is unfortunately not invariant under rotations. On the other hand, if we specify $\hat{f}_m$ as the Fourier Series Estimator (FSE) with $m$ harmonics ($m$ is now the smoothing parameter), i.e. [see e.g. Hart (1985) and De Jager et al. (1986)]

$$\hat{f}_m(\theta) = \left[1 + 2\sum_{k=1}^{m} (\hat{\alpha}_k \cos k\theta + \hat{\beta}_k \sin k\theta)\right] \Big/ 2\pi \tag{5}$$

where the empirical trigonometric moments are given by

$$\hat{\alpha}_k = (1/n)\sum_{i=1}^{n} \cos k\theta_i \quad \text{and} \quad \hat{\beta}_k = (1/n)\sum_{i=1}^{n} \sin k\theta_i \tag{6}$$

then Eq. (3) reduces within constants to the $Z_m^2$ statistic (Beran, 1969) which was introduced by Buccheri et al. (1983):

$$Z_m^2 = 2\pi n \psi(\hat{f}_m) = 2n \sum_{k=1}^{m} (\hat{\alpha}_k^2 + \hat{\beta}_k^2) \tag{7}$$

One of the advantages of this test above Pearson's $\chi^2$-test is that it is rotation invariant and the null distribution of $Z_m^2$ also tends to that of a $\chi^2$ distribution with $2m$ degrees of freedom as $n \to \infty$. One can now see that Pearson's $\chi^2$-test and the $Z_m^2$-test are "cousins" through the functional $\psi(\hat{f}_h)$, but with different parent density (or light curve) estimators. If $m = 1$ then $Z_1^2$ corresponds to the well known Rayleigh statistic.

A much wider concept of testing for uniformity was introduced by Beran (1969): He approached the problem from a viewpoint which is very useful for X- and $\gamma$-ray astronomy and considered alternative densities of the form given by Eq. (2). The approach taken was to assume that the signal strength $p$ is weak and unknown, but the signal density $f_s(\theta)$ was used to construct the test statistic

$$B = 2\pi n \int_0^{2\pi} \left[ (1/n) \sum_{i=1}^{n} f(\theta + \theta_i) - 1/2\pi \right]^2 d\theta$$

If $c_k$ $(k = 1, 2, \ldots)$ are the Fourier coefficients (or "characteristic function") of $f_s(\theta)$, then $B$ may be written in an even more convenient form

$$B = 2n \sum_{k=1}^{\infty} |c_k|^2 (\hat{\alpha}_k^2 + \hat{\beta}_k^2) \tag{8}$$

One of Beran's most important contributions was to show that this test statistic is "locally most powerful" for alternative densities of the form given by Eq. (2) if the signal strength is weak. This means that $B$ is the best (i.e. most powerful) test to detect weak signals if $f_s(\theta)$ is known a priori. Furthermore, no knowledge of $p$ is required to construct such a test. It is now of interest to see that $B$ equals the $Z_m^2$-test statistic if $|c_k| = 1$ for $1 \leq k \leq m$ and zero otherwise, but it is difficult to rewrite Pearson's $\chi^2$ statistic in terms of Eq. (8).

Watson's $U^2$-test for uniformity (Watson, 1961) was investigated by Quesenberry and Miller (1977) who have found this test to be a good omnibus test which can be used if the shape of the alternative is unknown a priori. This test is also derived from Beran's general class if $f_s(\theta)$ is taken as a triangular density function with a full width of $2\pi$. The resulting test statistic is easy to calculate

$$U^2 = 2n \sum_{k=1}^{\infty} (2\pi k)^{-2} (\hat{\alpha}_k^2 + \hat{\beta}_k^2) \tag{9}$$

Protheroe (1985) introduced a test for periodicity which is constructed using the distances between the phases for all pair combinations. He has shown that this test is very powerful for narrow peaks on a light curve (see e.g. Fig. 3). With this test a very narrow peak in the orbital phasogram from LMC X-4 was identified at PeV energies (Protheroe and Clay, 1985), It is unfortunate that the calculation of this test statistic requires a number of steps which is of the order of $n^2$ which will be quite time consuming in the case of VHE $\gamma$-ray astronomy where $n \gtrsim 10^3$. Furthermore, the critical values were not supplied by the author for $n > 200$ and these should be obtained from simulations. However, this test is very useful in UHE $\gamma$-ray astronomy where $n < 200$ (but not for long since new experiments with large collection areas have been built to give large counting rates).

From the power studies undertaken in Sect. 4 it will become clear that the $Z_m^2$-test is potentially a good test in the sense that the number of harmonics $m$ can be adjusted to detect both narrow (by using a large number of $m$) and broad (by using a low value of $m$) light curve shapes. However, the main problem here is that $f_s(\theta)$ is not known a priori so that $m$ cannot be chosen before inspection of the data. Thus, we have a problem with the choice of the smoothing parameter (the same problem applies to Pearson's $\chi^2$-test). Other tests like Watson's $U^2$- and Protheroe's tests are free of any smoothing parameter and they appear to be very powerful for certain light shapes, but perform still poorly (relative to other tests) against other possible shapes. In the following Section the $H$-test for uniformity is derived which proves to be a solution to both problems: The smoothing parameter involved is chosen in an objective and automatic way without requiring any knowledge a priori of $f_s(\theta)$ and it is nearly the most powerful test for most possible light curve shapes.

## 3. The $H$-test for uniformity: a suggested solution

The $H$-test uses the $Z_m^2$-statistic as basis, but we suggest a very simple solution for the choice of the smoothing parameter $m$: Specify $m$ as some suitable function of the data $\theta_i$ $(i = 1, 2, \ldots, n)$, using Hart's rule (Hart, 1985). His rule amounts to calculating that value of $m$, say $M$, which minimizes an estimator of the mean integrated squared-error (MISE) between the FSE (Eq. (5)) and the true unknown light curve $f(\theta)$:

$$\mathrm{MISE}(m) = E \int_0^{2\pi} [\hat{f}_m(\theta) - f(\theta)]^2 \, d\theta \tag{10}$$

This entails the determination of $M$ by the following procedure which is basically Hart's rule, but terms to the order $1/n$ were neglected to give a simplified procedure

$$\underset{1 \leq m < \infty}{\mathrm{maximum}} (Z_m^2 - 4m + 4) = Z_M^2 - 4M + 4$$

When $M$ is determined in this way, one has the following desired property:

$$\mathrm{MISE}(M) \lesssim \mathrm{MISE}(m), \quad m = 1, 2, \ldots$$

However, from a practical viewpoint it is impossible to search for $M$ through an infinite amount of harmonics and we suggest a truncation after 20 harmonics. This leads to our definition of the $H$ statistic:

$$H \equiv \underset{1 \leq m \leq 20}{\mathrm{maximum}} (Z_m^2 - 4m + 4) = Z_M^2 - 4M + 4 \tag{11}$$

In view of Eqs. (7) and (11) one can use the statistic $Z^2 \equiv Z_M^2$ (i.e. by using this optimal $M$ in the $Z_m^2$ statistic) as a test for uniformity. The latter statistic is now independent of any subjectively chosen smoothing parameter.

Another approach is to apply the statistic $H$ in (11) to test for uniformity. $H$ can be viewed as a rescaled version of $Z^2$. Most of the discussion below will be focussed on this suggested test. Among others, it will be shown that the $H$-test is, under certain circumstances, more powerful than existing statistical tests in the literature. However, it will also become clear that for certain multimodal light curves (i.e. for three or more peaks) the $Z^2$-test performs better than existing tests (including the $H$-test) with regard to power. We therefore recommend applying the $H$-test in general, except in cases where one anticipates a light curve consisting of more than two peaks, in which case the $Z^2$-test is preferable.

The number $M$ obtained from Eq. (11) will be a good estimate of the true optimal number of harmonics (provided the latter is $\leq 20$, which will mostly be the case in practice). However, from a mathematical viewpoint it is difficult to justify the use of $H$ (in Eq. (11)) itself as a test statistic for uniformity. In fact, one can question this use of $H$ as follows: What choice of the constant $c$ in a more general formulation of Eq. (11), e.g. $Z^2 - cM + c$ (where $M$ is still the optimal $m$ which followed from the choice $c = 4$), will give the best omnibus test? The choice $c = 0$ results in the $Z^2$ statistic and its properties, but from power studies it followed that $c = 4$ (which implies Eq. (11) and thus the $H$ statistic) results in the best omnibus test. It can also be shown that the $H$ statistic always assumes positive values and reduces to the Rayleigh statistic if $M = 1$. In this case only the two probability distributions will differ.

We now turn to the estimation of the probability distribution of $H$ under uniformity from which the significance (quantified by the $p$-level) of a detection can be determined: We obtain an asymptotic (for large $n$) approximation for the distribution of $H$ without simulating the individual phases $\theta_i$ – only $Z_m^2$ deviates had to be simulated from $\chi^2$ distributions with 2m degrees of freedom: Some $10^8$ values of $H$ were simulated from Eq. (11) using the constraint of searching for $M$ within the first 20 harmonics. A single probability distribution was fitted to the $H$-values using a minimum $\chi^2$ goodness of fit procedure:

$$\mathrm{Prob}(H > h) = a\exp(-bh) \quad \text{if} \quad 0 < h \leq 23 \tag{12a}$$

$$\mathrm{Prob}(H > h) = c\exp(-dh + eh^2) \quad \text{if} \quad 23 < h < 50 \tag{12b}$$

where the constants from the fit are given by $a = 0.9999755$, $b = 0.39802(5)$, $c = 1.210597$, $d = 0.45901(4)$ and $e = 0.0022900(5)$. The numbers in brackets refer to the approximate errors applicable to the last digit of the corresponding number. The errors on $a$ and $c$ are dependent on those on $b$, $d$ and $e$. However, for $h \ll 50$, the errors involved are negligible when calculating $p$-levels from Eq. (12). Due to the finite number of simulations we cannot give any reliable parametric equation for $h$-values larger than 50, except to say that the $p$-level $\mathrm{Prob}(H > 50) \simeq 4\,10^{-8}$. It is also useful to know what the mean and standard deviation of $H$ will be in the case of uniformly distributed events on the circle: The values of these quantities are both equal to 2.51.

The result given above is applicable to sample sizes larger than $\simeq 100$, i.e. when the asymptotic distribution of $Z_m^2$ approaches that of the $\chi^2$ distribution with $2m$ degrees of freedom. In the case of small samples ($10 \leq n \leq 100$) we approximated the distribution of $H$ by an exponential function of the form

$$\mathrm{Prob}(H > h) \simeq w(h)\exp(-b_n h)$$

where the constant $b_n$ is given in Fig. 1 for various $n$ between 10 and 100. The function $w(h)$ is given in Fig. 2 and is weakly dependent on $n$, except when $n$ becomes as small as 10. It is clear that $w(h) \simeq 1$ for $h \lesssim 20$ but increases (for $15 \lesssim n \lesssim 100$) for larger values of $h$. Since we have defined "small sample sizes" by $n \leq 100$ and the asymptotic case by $n > 100$, one should still require continuity between the small sample size distribution and the asymptotic distribution (relation (12)) at $n \simeq 100$: If one uses Eq. (12) for the case $n = 100$, $p$-levels will be overestimated by a factor which equals $\exp(0.008h)$ for $h \lesssim 20$, which entails a maximum error of 20%. However, this error increases for $h \gtrsim 20$. For these small sample sizes we have restricted the search for $M$ (in Eq. (11)) to the interval $1 \leq m \leq n/5$ if $n \leq 100$, but for any larger $n$, the range for $M$ is restricted between the values 1 to 20 (as mentioned earlier). The maximum number of 20 is arbitrary and one must truncate the search routine somewhere or else the number of steps necessary for the calculation of the test statistic will approach infinity. Furthermore, by following the lines of Beran (1969), it can be shown that the $H$-test is "consistent" (as defined in the Introduction) if at least one $|c_k|$ is non-zero for $1 \leq k \leq 20$, where $c_k$ (for $k = 1, 2, \ldots$) are the Fourier coefficients of the signal density function $f_s(\theta)$ as given in Eq. (2). If a larger truncation value (instead of 20) is chosen, one will find that the constants in Eq. (12) will change slightly, but the critical values of the test statistic will not change significantly, so that the power of the test will remain nearly invariant. Finally, the total number of computational steps necessary to calculate $H$ from Eq. (11) is of the order of $40n$ and not $n^2$, which is desirable.

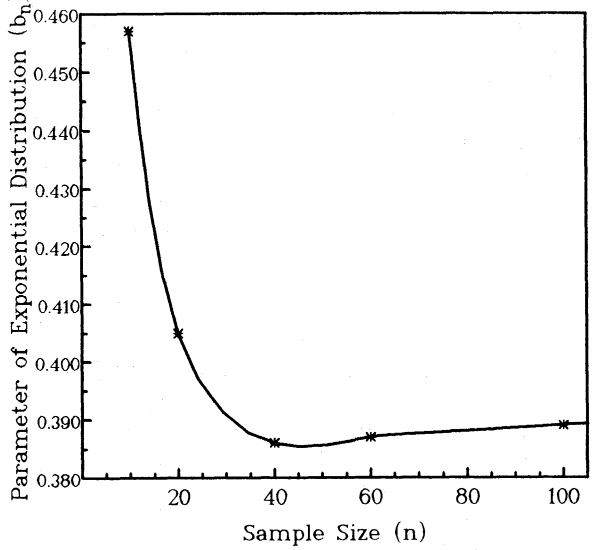


**Fig. 1.** The parameter $b_n$ for the probability distribution $P(H > h) = w(h)\exp(-b_n h)$ of the $H$-statistic in the case of small sample sizes ($10 \leq n \leq 100$). In the case of large $n$ we have the asymptotic result $b_n \to 0.398$

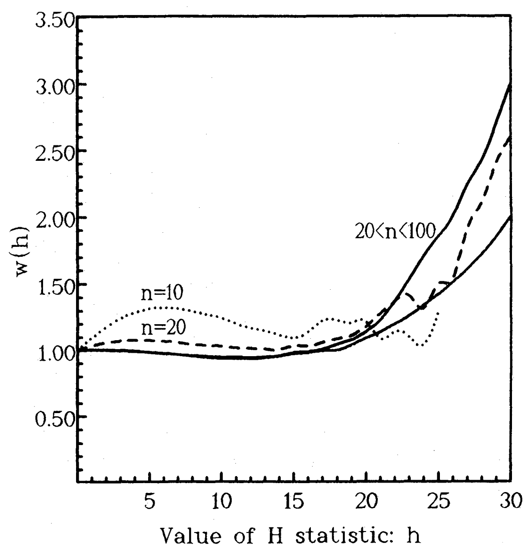


**Fig. 2.** The function $w(h)$ for the probability distribution $P(H > h) = w(h)\exp(-b_n h)$ of the $H$-test in the case of small sample sizes ($10 \leq n \leq 100$). A 95% confidence band is given by the solid lines for sample sizes larger than 20, but $\leq 100$

## 4. The power of the $H$-test

The efficiency of a test to reject the null hypothesis (given the latter is false) is quantified by its power: The power of a test (as seen in the context of this study) is the probability that a periodic source, with given characteristics, will be identified above a given detection threshold for a given sample of size $n$. In our case the source characteristics are specified by the signal strength and the signal density as given by Eq. (2). The "detection threshold" is specified by the probability $\alpha$ to make a Type I error. This means that $\alpha$ is the prescribed probability to falsely accept an indication of a pulsed signal from a source as being a real periodicity. The procedure to identify a periodic signal is to calculate the test statistic from the phases, calculate the $p$-level and reject $H_0$ if the $p$-level is less than $\alpha$. The choice of $\alpha$ is very subjective and depends on the situation. For a first detection of a source one should keep $\alpha$ quite low – say 0.001 and for confirmation purposes $\alpha = 0.05$ should be an upper limit.

One can determine the power of a test for a given $f(\theta)$ by counting the relative number of simulations for which the $p$-level is less than $\alpha$. For the purpose of showing the behaviour of the power, any value of $\alpha$ may be used – the final conclusions remain the same (see e.g. Protheroe, 1987). We have chosen 0.05 to obtain good statistics and therefore small errors on the simulated power values. A lower value would require more simulations to give similar errors. The power curves for the following tests were calculated: Rayleigh, $Z_2^2$, $Z_{10}^2$, Protheroe's statistic, Watson's $U^2$, $Z^2$ and $H$. We concentrated on weak signal strengths and the shape of the signal (i.e. $f_s(\theta)$) was modelled to resemble the shapes mostly observed in X- and $\gamma$-ray astronomy: (i) Firstly, a Gaussian single peak with duty cycle $\delta$ (which is the full width at half maximum). A convenient approximation for such a signal shape is the periodic von Mises density (Mardia, 1972) with concentration parameter $\kappa$ which is related to $\delta$ as follows: $\delta \simeq \cos^{-1}(1 - 0.693/\kappa)/\pi$. The use of the von Mises density is valid for $\kappa > 1$, but for $\delta = 0.5$ we use a sinusoid. (ii) The second alternative is double peaked Crab- and Vela-like light curves with variable relative peak intensities. (iii) The third type of alternative is evenly spaced multiple peaks of the same intensity: The density function $f(\theta) = a/(1 + b\cos(k\theta))$ which generates $k$ equally spaced peaks on the interval $(0, 2\pi)$ was used since its harmonic content is complex. The power is then investigated as a function of $k$ and the rationale behind such a study is the following: The suggested test for our purposes should have a high power for $k$ between one and about four, but for $k > 5$, the shape of the light curve becomes less realistic since such multiple peaks are very seldom (if ever) expected from periodic sources and the power of the test for such cases must actually be lower: High frequency (with large $k$) oscillations may be the cause of unwanted systematic effects due to the observation technique and it is actually desirable to have a test statistic with a low power for such cases, so that these type of systematic effects may be eliminated as possible sources. The sample size for each power study was taken as 300 (except when stated otherwise) with $\alpha = 0.05$.

### *4.1. A single peak embedded in background*

The power of the various tests are compared in Fig. 3 for various values of $\delta$ and for $p = 0.1$. It is clear that the $H$-test performs generally best of the tests considered, simply because its power is overall high while all other tests perform good for certain values of $\delta$ but unacceptably low for other values of $\delta$. The "power envelope" is also shown on the Figure and represents the best that any test for uniformity can do in the case of such weak signals. This curve corresponds to the power curve of the "locally most" (i.e. only for weak signals) powerful test which is derived from Beran's work for these alternatives and can in principle be used only when the duty cycle is known a priori. Observe the increase in power for all tests when $\delta$ decreases. The physical consequences of this will be discussed in Sects. 6 and 7. In Fig. 4 the power curves of the $H$-test and some other tests are drawn

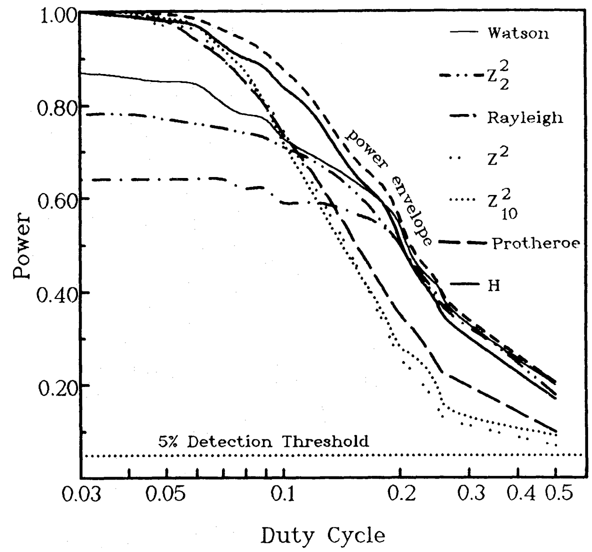


**Fig. 3.** The power curves of different tests for uniformity as a function of the duty cycle of a single-peaked Gaussian light curve with 10% signal strength and a total sample size of 300. The detection threshold was taken as 5%. The "power envelope" (the upper dashed curve) represents the best that any test can do in this case

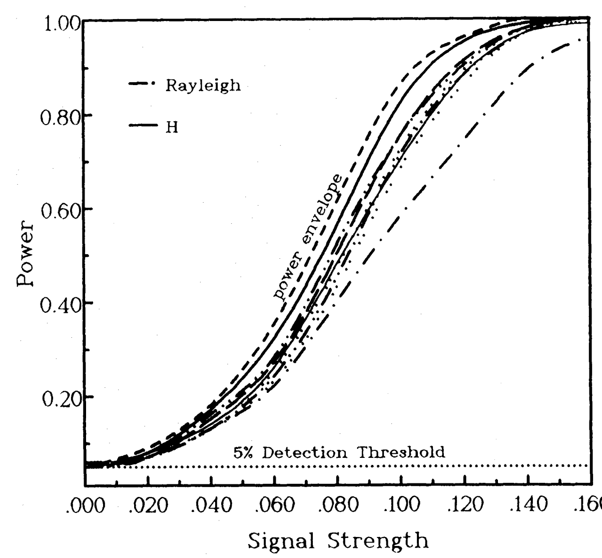


**Fig. 4.** The power curve of different tests for uniformity as a function of the signal strength for a single-peaked Gaussian light curve with 10% duty cycle. The sample size is 300 and the detection threshold is 5%. The line types are the same as in Fig. 3. The power envelope (upper dashed curve) is also indicated

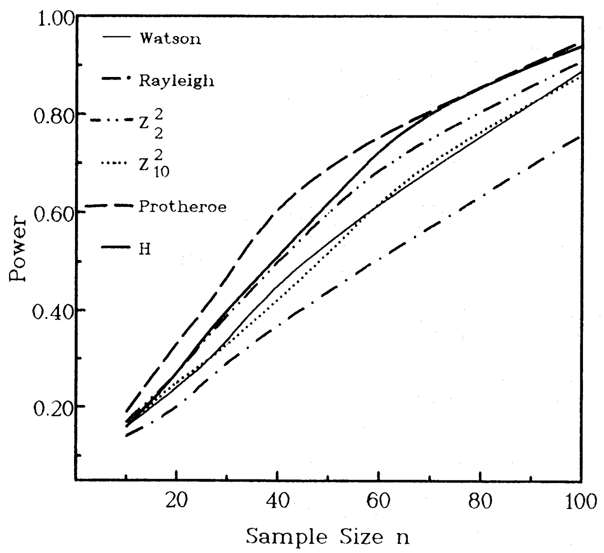


**Fig. 5.** The power curves of different tests as a function of the sample size *n*. A single-peaked Gaussian shape with a duty cycle of 10% and a signal strength of 20% was used. The detection threshold is 5%

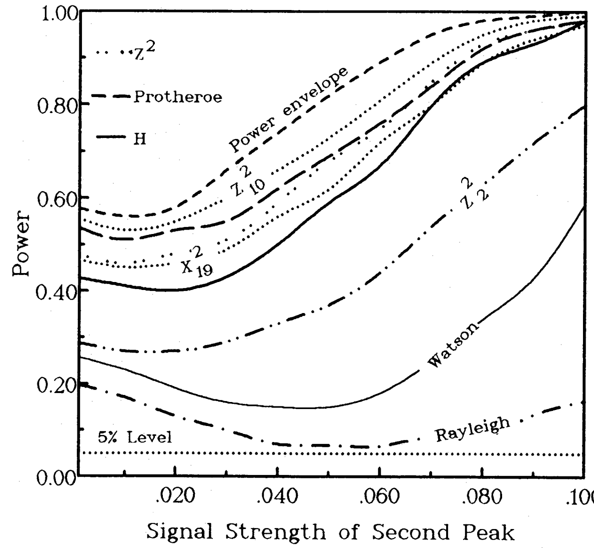


**Fig. 6.** The power curves of various tests for Crab- and Vela-like double-peaked light curves (with 3.7%- and 5.4% duty cycles respectively). The symbol "$\chi_{19}^2$" represents Pearson's $\chi^2$-test with 20 bins. The signal strength of the first peak is 5% and the power is drawn as a function of the strength of the second peak. The sample size was again taken as 300 with a 5% detection threshold. The power envelope (upper dashed curve) has the same meaning as in Figs. 3 and 4

as a function of the signal strength for a single peak with a duty cycle of 10%. The *H*-test is best here for all values of *p* and its power is always within a few percent below the power envelope. In Fig. 5 we have taken $\delta = 0, 1$ and $p = 0, 2$ but the total number of events *n* is allowed to vary between 10 and 100. Protheroe's test seems to be excellent for small sample sizes, but its power decreases relative to that of other tests (e.g. the *H*-test) as *n* increases. It is clear that the *H*-test remains fairly powerful relative to that of other tests when *n* decreases. Thus, it is also good for small sample sizes (at least down to $n = 10$).

### 4.2. *Crab- and Vela-like double peaks*

Double peaked light curves are often expected to occur in nature and examples of this are the gamma-ray light curves of the Crab- and Vela pulsars. However, the relative peak intensities can vary. In this case we have taken the well known gamma-ray light curve of the Vela pulsar (Kanbach et al., 1980). The two peaks were approximated by Gaussian densities with duty cycles 3.7 and 5.4% respectively. The phase separation was taken to be a fraction 0.42 of the pulsar period and the signal strength of the first pulse was taken as 5%. In Fig. 6 the power curve of the tests are drawn as a function of the signal strength of the second pulse. In this way we have investigated the behaviour of various tests as a function of the relative peak intensities: The Rayleigh- and Watson's $U^2$-tests perform poorly for all values of the signal strength. Slightly better is the performance of $Z_2^2$, but its power does not approach unity so fast (for increasing signal strength) as the power curves of $H$, $Z^2$, $Z_{10}^2$ and Protheroe's statistic. These latter four tests appear to be good for such light curves, although *H* is worst amongst them. This case serves as a good example where significant improvement over the *H*-test is possible.

### 4.3. *Multimodal light curves*

From the previous study it seemed as if the *H*-test is not so successful against very narrow double peaks as $Z_{10}^2$, $Z^2$ and Protheroe's statistic, as it was against single peaks. The purpose of the following power study is twofold: (1) To investigate the powers for slightly broader double peaks and (2) as a function of the number of peaks on the light curve: To do this we have used the previously mentioned density function which generates multiple peaks. In Fig. 7 the single peak ($k = 1$) is broad with a duty cycle of 30% and Watson's $U^2$-, the Rayleigh-, $Z_2^2$- and the *H*-tests perform well as expected. However, the double peaks ($k = 2$) now have duty cycles of 15% and we see a behaviour which differs from Figure 6 (simply because the peaks are now

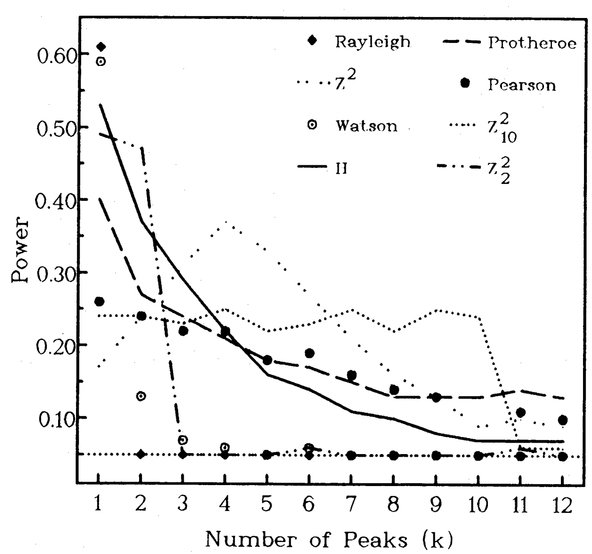


**Fig. 7.** The power curves of various tests for the density function $f(\theta) = 0.312/(1 + 0.2\cos(k\theta))$ as a function of the number of peaks *k* on the light curve. The label "Pearson" corresponds to Pearson's $\chi^2$-test with 20 bins. The sample size was 300 with a detection threshold of 0.05

broader than in Fig. 6): Of the four tests ($H, Z_{10}^2, Z^2$ and Protheroe's statistic), the $H$-test is best, which supports its omnibus property, but the $Z_2^2$-test performs better. However, the power of the latter test was much lower than that of the $H$-test for Crab- and Vela-like light curves. When more peaks are added to the light curve, the power of the $H$-test drops but the power of $Z^2$ increases and reaches a maximum for about four to five peaks. The latter is already an improvement relative to the $H$-test if more than three peaks are anticipated, and for narrower peak widths, the improvement already becomes evident for fewer peaks (as it was shown in Fig. 6). Furthermore, observe the low but steady power of the $Z_{10}^2$-test as more peaks are added. Pearson's $\chi^2$-test with 20 bins is equivalent (with respect to power) to the $Z_{10}^2$-test for less than five peaks and the decrease in power of $\chi^2$ (relative to $Z_{10}^2$) is due to the effect of binning. However, from a physical viewpoint a test should not be able to identify light curves with more than say five peaks and such multimodal light curves may be considered as noise. The Rayleigh-, Watson's $U^2$- and $Z_2^2$-tests cannot detect high frequency noise and of all the other tests considered, the $H$-test is best to discriminate against such unwanted data structures.

We can derive the following conclusions from these power studies: All tests (except $H$) are good for either narrow or broad peaks, but not for both. Any suggested omnibus test must be good for both cases and its power curve must be close to the power envelope for all physically acceptable light curve shapes. For single peaks with duty cycles $\gtrsim 6\%$ the power curve of the $H$-test is close to the power envelope which means that no significant improvement is possible for such cases. For sinusoids the power envelope is defined by the Rayleigh test, although the power of the $H$-test is not far below that of the Rayleigh test. In fact, this difference can be understood in terms of the probability distributions: For a sinusoid one will most likely get $M = 1$ as the optimal number of harmonics for the $H$-statistic. In this case the values of the test statistics of the Rayleigh and $H$ will be equal – say this value is $R$. The $p$-levels for the Rayleigh- and $H$- statistics will then be $\exp(-0.5R)$ and $\simeq \exp(-0.4R)$ respectively for large sample sizes and the $p$-level for the Rayleigh-test will then be a factor $\exp(0.1R)$ smaller than the corresponding value for the $H$-test. From other power studies it also became clear that the $Z^2$-, $Z_{10}^2$ and Protheroe's-tests become more powerful than $H$ if the duty cycle(s) decreases below 6%. If the number of peaks on the light curve increases above three, the $Z^2$-test becomes best – this is independent of the duty cycles involved. This illustrates that one can select the $Z^2$-test if three or more peaks are expected, but if more specific information is available, a very powerful test from Beran's class can be constructed. The reason why we cannot suggest $Z^2$ as an omnibus test is because of its poor power performance for broader ($\delta \gtrsim 0.1$) single- and double-peaked light curves. However, without any a priori knowledge, it is clear that the $H$-test is best to use since its power performance is relatively good for physically acceptable light curve shapes.

## 5. Oversampling due to scanning in period

Consider the arrival times $t_1, t_2, \ldots, t_n$ where the total observation time is given by $T = t_n - t_1$ and one has to search for a periodicity within the frequency interval $[\nu_1, \nu_2]$. The number of independent Fourier spacings (IFS) which have to be searched is given by

$$x = T(\nu_2 - \nu_1) \tag{13}$$

Consider the case where a certain range in period is search for evidence of periodicity using a test statistic $S$ and let $x$ be the number of IFS covered. If the step length between successive period values is given by $\Delta x = 1$, one can calculate the $p$-level $P(S > s)$ for a detection with $S = s$ (i.e. $s$ is the largest value obtained from $x$ values of $S$) from the distribution of the largest ordered statistic of $S$ from $x$ trials:

$$P(S > s) = 1 - (1 - P'(S > s))^x \simeq xP'(S > s) \quad \text{if} \quad P' \ll 1 \tag{14}$$

where $P'(S > s)$ is the $p$-level calculated from the known distribution function of $S$, which is applicable for testing at a fixed period. Equation (14) will be incorrect if $\Delta x \ll 1$. In the latter case the number of effective independent trials is actually larger than $x$. De Jager (1987), De Jager et al. (1988) and Buccheri and De Jager (1989) approached this problem by estimating that factor $r_x(s)$ with which $P(S > s)$ in Eq. (14) is underestimated. This was done through simulations since the analytical approach is difficult.

The first question which arises before attempting to determine this factor is the required step length $\Delta x$ in independent Fourier spacings: The step length should not be too large, since one should be able to sample at least one "large" value of the test statistic $S$ near its maximum. The latter will on average correspond to the true period. If one requires $P'$ $(S > s)$ to be less than say 0.001, it will be found that $\Delta x$ should decrease as the duty cycle of the light curve increases for a fixed signal strength and sample size. This is because the fraction of an IFS (around the true unknown period) decreases in which $p$-levels less than 0.001 may be found. Similarly, for a fixed light curve shape one will also find this same behaviour if the signal strength is reduced. We therefore suggest that one should use at least 20 steps per IFS for weak signals. This will give a better probability for the detection of weak signals than the case where only (say) three to five steps are selected.

In the following simulation study we estimated the factor of underestimation due to a search with 20 steps per IFS and using the following test statistics: Rayleigh, $Z_2^2$, $Z_{10}^2$ and $H$. This factor was simulated from 200 Poissonian distributed arrival times and the search frequencies were taken to be much larger than the count rate and in this way we have restricted ourselves to sparse data. Since these calculations are time consuming, we could do only about 40 000 repetitions for the search for the maximum $s$ within twenty steps per IFS. This procedure was also repeated for more than one IFS (while keeping $\Delta x = 0.05$ IFS fixed). The factor of underestimation ($r_x(s)$) for $x$ IFS and a value $s$, is estimated by taking the ratio of the number of simulations for which the maximum value of the test statistic found (after searching within $x$ IFS with a step length of $\Delta x = 0.05$) exceeded $s$, and $P(S > s)$ calculated from Eq. (14). Thus, the true $p$-level which should be used is given by

$$P(S > s) = r_x(s)[1 - (1 - P'(S > s))^x] \tag{15}$$

In Fig. 8a–d plots are shown of $r_x(s)$ versus $s$ (for the various choices of $S$ as mentioned above) for different values of $x$. From Fig. 8a it appears as if $r_x(h)$ is linear with $h$ for $x = 1$ and $h \lesssim 20$, so that $r_1(h) \simeq 1 + 0.45h$. From Eq. (15) it follows that an approximate parametric expression for the true $p$-level, after a search within one IFS was conducted using the $H$-test, is given by the following equation (provided that $h \lesssim 20$):

$$P(H > h) \simeq (1 + 0.45h)\exp(-0.398h)$$

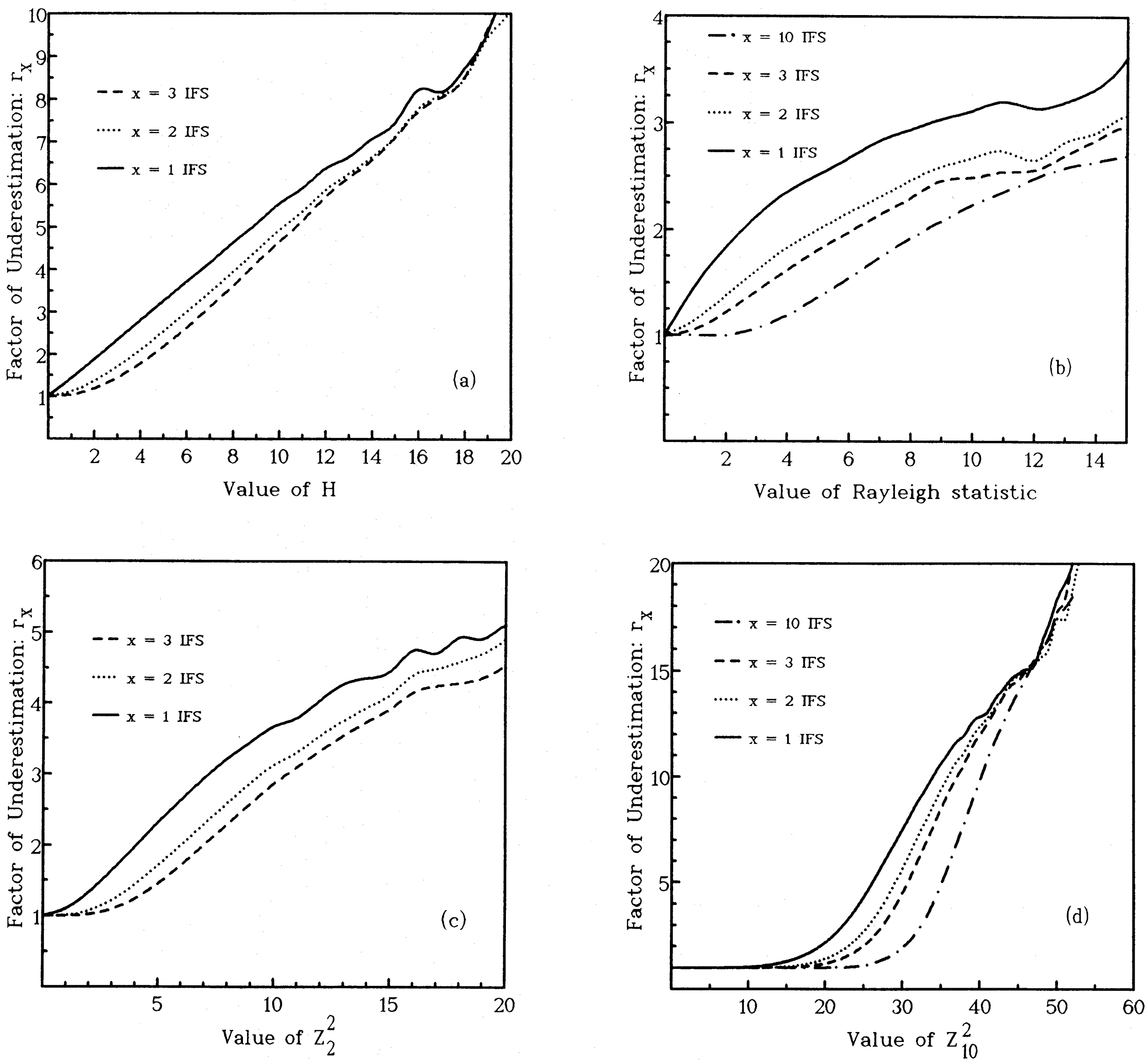


**Fig. 8a–d.** The "factor of underestimation" ($r_x$) as a function of the value of the following test statistics: **a** H, **b** Rayleigh, **c** $Z_2^2$ and **d** $Z_{10}^2$. This is given for a search within $x$ independent Fourier spacings (IFS), but with a fine scanning of 20 steps per IFS

To understand these results better, the reader is referred to the work of Buccheri and De Jager (1989): They investigated the degree of correlation (or dependence) between $S(x)$ and $S(x+\Delta x)$. They have shown that the degree of correlation increases as $\Delta x$ is reduced. In general one will find that $r_x(s)=1$ for all $s$ if $\Delta x = 1$, but this factor of underestimation increases if $\Delta x$ is reduced, but should converge to a constant value if $\Delta x \to 0$. The latter case corresponds to a convergence of the correlation coefficient to one so that $s(x) \simeq s(x+\Delta x)$. Furthermore, they also found (for a fixed $\Delta x$) that $Z_{10}^2(x)$ is much less correlated with $Z_{10}^2(x+\Delta x)$ than is evident for the Rayleigh statistic and the lesser this degree of correlation, the larger will $r_x(s)$ be for the corresponding test statistic. This is clear when one observes Fig. 8b–d. An alternative explanation for this is the following: As higher harmonics are added, the number of effective IFS (and thus "trials") applicable to the $m$'th harmonic is $m$ times $x$ and to account for this, $r_x(s)$ should increase.

From these graphs it also becomes evident that $r_x(s)$ decreases for a fixed $s$ if $x$ increases and it is important to know how fast $r_x(s)$ converges to unity as $x \to \infty$, but to answer this question would need many days of calculation on a super computer and for sufficiently large $x$, Eq. (14) would be correct to use. However, for the values for $x$ considered here and for $P'(S>s) \lesssim 10^{-3}$, it seems as if $r_x(S)$ becomes independent of $x$ for $Z_{10}^2$ and $H$, but not for the Rayleigh- and $Z_2^2$-statistics. This aspect was not clear from the studies presented by De Jager (1987) and more detailed studies are necessary.

We have seen that the factor $r_x(S)$ increases drastically if one adds higher harmonics to a test statistic like $Z_m^2$. For example: Consider the hypothetical case where $P'(S>s)=0.005$ was obtained after a search within one IFS with any test statistic $S$. From Fig. 8 it follows that the true $p$-levels will be equal to 0.016, 0.021, 0.064 and 0.035 for the Rayleigh- $Z_2^2$-, $Z_{10}^2$- and $H$-tests respectively. Thus, the probability to detect a source will therefore decrease if more harmonics (or bins) are added to the $Z_m^2$ (or Pearson's) statistic. It is thus worthwhile to reinvestigate the power of different tests if one has to search within say one IFS: In Figs. 9 and 10, some of the power studies of respectively Figs. 3 and 6 are repeated, but the factor of underestimation has been included for each case. One can see that the power of all tests have dropped. The Rayleigh statistic is least affected while the $Z_{10}^2$ is mostly affected. The superiority of the Rayleigh test is now quite clear in the case of sinusoids (i.e. duty cycle = 0.5). If one is certain that the light curve shape will be broad, then one should resort to the Rayleigh- or $Z_2^2$-tests if a search within one or more IFS is to be made. However, the true light curve shape is not always known on an a priori basis and it is clear that the $H$-test performs well in all other cases – for narrower shapes and Vela-like double peaks: In Figure 10 one can see that the power of $H$ equals the power of $Z_{10}^2$ and is better than that of any other test considered.

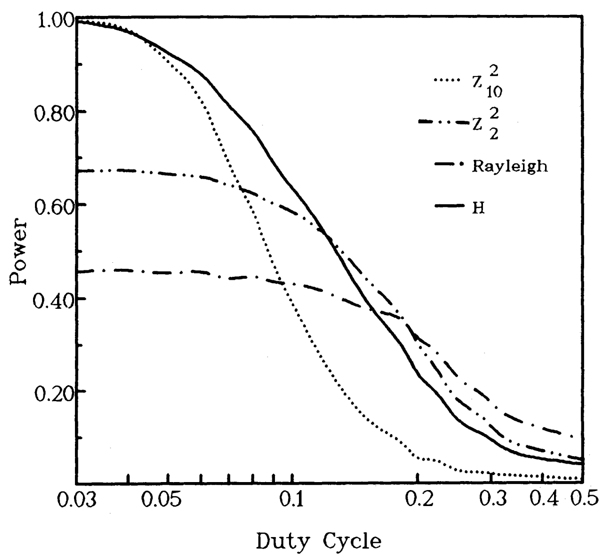


**Fig. 9.** The power of various tests versus duty cycle of a single peak (the parameters of Fig. 3 are applicable here), but the "factor of underestimation" due to a search within one IFS was taken into account

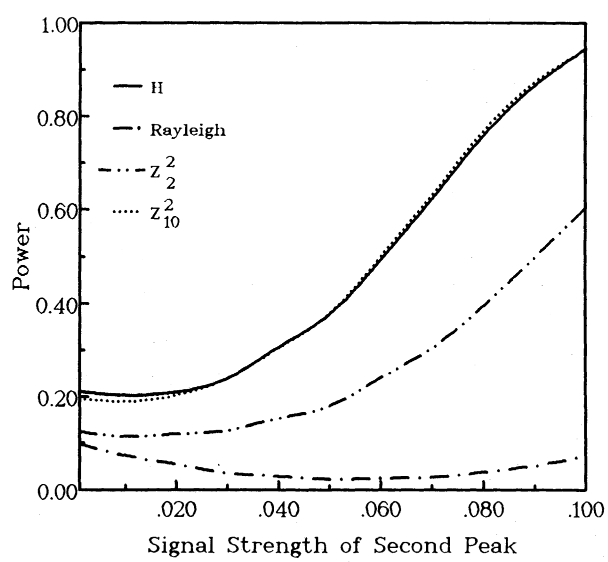


**Fig. 10.** The power versus signal strength of the second peak of a Vela-like light curve (the parameters of Fig. 6 are applicable here), but the factor of underestimation due to a search within one IFS was taken into account

## 6. Flux sensitivities for various tests

It is usually important to know what is the minimum flux that may be attained with a given test for uniformity. Buccheri et al. (1987) gave the approximate sensitivity level (in terms of photon counts) for the Rayleigh-test and a given detection threshold after a number of independent trials were made before arriving at a positive detection. We calculated the exact average minimum flux sensitivity for various tests from the power curves of the various tests. This is done by calculating the fundamental scaling parameter $p\sqrt{n}$, which is applicable to many tests for uniformity [Leahy et al. (1983a,b), De Jager (1987) and Protheroe (1988)]. This parameter is related to the physical parameters of the experiment as follows:

$$p\sqrt{n} = \bar{F}A(T/\bar{\lambda})^{1/2} \tag{16}$$

where $\bar{F}$ is the average flux of radiation during the observed time $T$ for which the average count rate was $\bar{\lambda}$ which corresponds to an effective collection area of $A$. We can now define the "minimum flux sensitivity" as the average flux necessary to detect a source with a given light curve shape with a $p$-level less than (say) 0.05 during (say) 99% of the times when the source is observed under similar conditions. This flux sensitivity is obtained by setting the power of the test under consideration equal to 0.99 for a Type I error rate of 0.05 and solving for $p\sqrt{n}$ by means of simulations.

We have done this for the $H$-, Rayleigh-, $Z_2^2$- and $Z_{10}^2$-tests (the behaviour of the latter is expected to be similar to the behaviour of Pearson's $\chi^2$-test with $\simeq 20$ bins). The flux calculations were done for a single peak on the light curve and the testing is performed at a fixed period (i.e. only a single trial). The results from the simulations are shown in Fig. 11 and one can see that stronger signals are necessary to detect broader peaks than narrower ones. The minimum flux sensitivity is lowest on average for the $H$-test which is to be expected on the basis of the power studies in Fig. 3. This result indicates that the $H$-test can detect on average the weakest sources if no a priori information about the light curve shape is available.

The "flux envelope" is derived from Beran's theorem (see Sect. 2) and gives the *absolute minimum average flux level attainable by any method and no average flux value can be identified below this envelope at the given detection threshold and confidence level given.* The superiority of the $H$-test for single peaks is now evident when it is observed that its minimum flux is at most 10% larger than the limiting flux envelope whereas the Rayleigh- and Pearson's $\chi^2$-tests may have minimum flux levels which are respectively 200% and 50% larger than the flux envelope.

## 7. Selection effects for weak periodic sources

From the power studies of Sect. 4 and the flux sensitivities given in Fig. 11, it is clear that sources with narrow radiation beams are more easily detectable than sources with wider beams. The implications of this are the following: If one observes a sample of periodic sources from a population of sources for which the

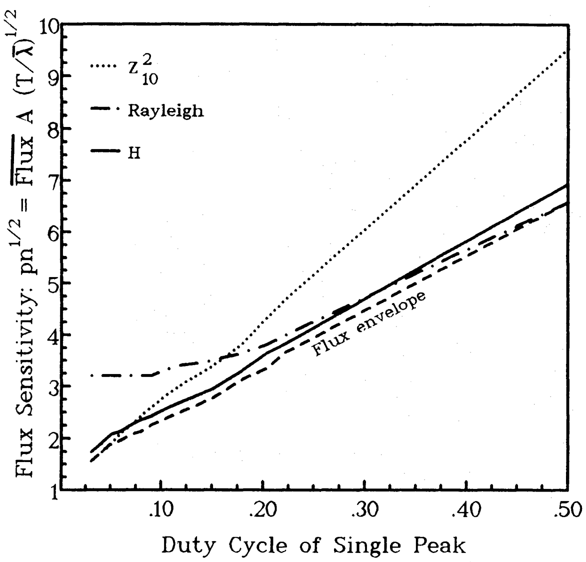


**Fig. 11.** The flux sensitivity versus duty cycle of a single-peaked light curve for various tests for uniformity. The "flux sensitivity" is defined as the average flux, times collection area, times $(T/\bar{\lambda})^{1/2}$ (where $T$ is the observation time and $\bar{\lambda}$ the corresponding average counting rate), necessary to give a positive detection with a $p$-level less than 5% per observation, but the latter should have a success rate of 99% of the times when the source is observed

observed flux at Earth (which is assumed to be relatively low) and the beam widths of radiation are uncorrelated, one would mainly expect the detection of *narrow-peaked light curves*. Consequently, any sample of period sources which were identified at a level which is barely above the detection threshold, will be biased against sources with wide beams. For example, from Fig. 11 it follows that any experiment which uses only the Rayleigh test gains a factor of two in sensitivity, while those using only $Z^2_{10}$ or $\chi^2$ (with $\simeq 20$ bins) will gain a factor of $\simeq 6$ in sensitivity if observations are switched from broad-beam to narrow-beam sources. It is also clear that this bias will be different from group to group since each group has its own preference with respect to the use of tests. However, it is important to note that all groups will detect more narrow- than broad-beam sources if the above-mentioned conditions hold.

It is well known that all positive detections of very- and ultra-high energy gamma ray sources were made at flux levels which are barely above the flux sensitivity of present-day ground-based telescopes (see Weekes, 1988 for a review). De Jager (1987) estimated $p\sqrt{n}$ for many of the positive detections in this field and those values roughly follow the $p\sqrt{n}$ versus duty cycle curves of Fig. 11. In fact, the narrow-peaked shapes were mainly detected with Pearson's $\chi^2$-test and a few times with Protheroe's statistic, while sources with broad peaks were mostly identified with the Rayleigh-test. Furthermore, the sample of narrow-peaked sources appears to be underpopulated which may indicate that the population of broad-beam emitters is in the majority, or that the beamwidth of radiation and corresponding flux are correlated in some way. Confirmation of the latter from more population studies may have important implications for the study of high energy sources. Another interpretation (which may be less popular) is to assume that some of those detections may be due to statistical fluctuations in the background noise.

## 8. Conclusions

In this paper it was shown that all the discussed classical tests for uniformity suffer in the sense that each one is very powerful against certain light curve shapes, but less powerful against other shapes so that none of those tests are fit to serve as an omnibus test for uniformity if no knowledge about the expected light curve shape is available a priori. The concentrated effort in this paper was towards weak signals and the need existing to find a test for uniformity which has a high power against most realistic light curve shapes.

It became clear that most tests have the undesired property that they are dependent on a smoothing parameter (e.g. the number of bins in Pearson's $\chi^2$-test or the number of harmonics with the $Z^2_m$ harmonic summing method). If the smoothing parameter (i.e. number of bins or harmonics) is small, the corresponding test will be very powerful against broad peaks and with a large smoothing parameter, narrow peaks will have priority in detection. In fact, from Beran's work it follows that the best choice of $m$ for $Z^2_m$ and a peak with a duty cycle $\delta$, is $m \simeq 1/(1.7\delta)$. However, $\delta$ is unknown a priori and one cannot afford to estimate the smoothing parameter from the data in a subjective way. We followed a direction by finding the optimal smoothing parameter (i.e. $M = m$) which describes the true unknown light curve shape best. This $M$ is estimated from the data in an automatic and objective way and we used this smoothing parameter for the $Z^2_m$-test. This led (as a first step) to the $Z^2$-test which is free of the subjectively chosen $m$. Thus, $M$ will automatically assume small values for broad peaks, but larger values for narrow or multiple peaks, which is desired. The $Z^2$-test proved to be powerful for most light curve shapes considered, except that the power performance was very poor for broad single- and double-peaked shapes. This disqualifies the $Z^2$-test as an omnibus test for X- and $\gamma$-ray astronomy.

The derivation of the "$H$-test" was actually a step further – the $H$-statistic itself is the numerical procedure which is used to estimate that optimal $M$. From the power studies it followed that the $H$-test is nearly the most powerful of all tests considered for a wide variety of physically acceptable light curve shapes, which identifies this test as a good omnibus test. However, there is still a constraint on this claimed superiority: For sinusoids the Rayleigh-test is better and for peak width(s) less than 6% the $Z^2_{10}$-, $Z^2$- and Protheroe's-tests performs better.

Another attractive feature of the $H$-test is the fact that it is relatively easy to calculate. In fact, it requires $\simeq 40n$ steps to calculate simply because the optimal $M$ is searched within the first 20 harmonics if the sample size is larger than 100. For smaller sample sizes the search interval is restricted to within the first $n/5$ harmonics. The probability distribution was derived for $n \to \infty$ and an exponential function resulted which is easy to use. This formula (Eq. (12)) is valid down to the $4\,10^{-8}$ level. For smaller $n$ (between 10 and 100) it was also shown that the probability distribution follows an exponential function, but with a slightly different form. In this small sample case it was also shown how the probability distribution deviates from the exponential function if the $p$-level is less than $10^{-4}$.

If a search for periodicity is to be conducted by calculating the value of the test statistic at a set of period (or frequency) values for which the corresponding spacings are much less than one independent Fourier spacing (IFS), it would be insufficient to take the total number of IFS as the effective number of trials – the latter will generally be larger than merely the number of IFS. In Sect. 5 it was shown how this extra "factor of underestimation" may be taken into account. This factor was estimated for various tests, which was shown to increase if the contribution from the higher harmonics is added to a test statistic (e.g. if $m$ or the number of bins is increased in the $Z_m^2$- or Pearson's $\chi^2$-tests respectively). For example: Consider the case where a $p$-level of 0.005 was obtained with both the Rayleigh and $Z_{10}^2$-tests after a search within one IFS was conducted. If the $p$-levels are adjusted with this factor due to the search, they will change to 0.016 and 0.064 respectively and when these effects are ignored, one can expect the detection of many more false sources than expected. The factor underestimation for the $H$-test is larger than the corresponding factor of the Rayleigh test, but less drastic than the effect for $Z_{10}^2$.

The power curve of a test can be used to calculate the minimum flux attainable by a test for uniformity for certain given experimental parameters. This was done for a few well known tests and the flux sensitivity for the $H$-test proved to be on average lower than the corresponding sensitivities of other tests. In the case of a single-peaked light curves it is important to realise that the absolute minimum flux value attainable by any method (no matter how powerful) cannot be more than 10% smaller than the minimum flux levels attainable by the $H$-test (but given duty cycles larger than 1%).

Since all tests have a higher power for narrow-peaked light curve shapes than for broader peaks, there should be a bias against the detection of broad-beam sources and if the source flux and beamwidth of radiation are uncorrelated for the total population of periodic sources, one should end with a source catalogue which consists mainly of narrow-peaked light curves. This behaviour is not seen in Very- and Ultra-High Energy Gamma Ray Astronomy, which may either imply a correlation between flux and beamwidth and/or some of the detections are false due to statistical fluctuations in the noise. Noise fluctuations identified with the Rayleigh test result in sinusoidal light curve shapes which appears to be physically acceptable and one cannot see the signature of noise in them. However, with the $H$-test one will usually be able to see such a signature if $M$ is large, since narrow-peaked statistical fluctuations identified with the $H$-test do not look physically acceptable. Apart from this, we have seen that the $H$-test also supresses high frequency noise (see Fig. 7), which may help to avoid systematic effects.

Finally, the suggested $H$-test proves to be a good omnibus test for uniformity which can be used if no knowledge of the light curve is available. There is however still space for improvement on the $H$-test for very narrow and very broad light curve shapes.

*Acknowledgements.* The authors wish to thank Dr. Lino Buccheri for many useful discussions, suggestions and encouragement. We also wish to thank Computer Services of the Potchefstroom University for the countless hours of computer time made available for these time-consuming calculations.

## References

Beran, R.J.: 1969, *Ann. Math. Stat.* **40**, 1196
Buccheri, R., et al.: 1983, *Astron. Astrophys.* **128**, 245
Buccheri, R., Ozel, M.E., Sacco, B.: 1983, *Astron. Astrophys.* **175**, 353
Buccheri, R., Ogelman, H.: 1988, *Nature* **331**, 309
Buccheri, R., De Jager, O.C.: 1989, *NATO A.S.I. Workshop on Timing Neutron Stars*, eds. H. Ögelman, E.P.J. van den Heuvel, Kluwer, Dordrecht, p. 95
De Jager, O.C., et al.: 1985, *Proc. 19th ICRC* (La Jolla) **3**, 481
De Jager, O.C., Swanepoel, J.W.H., Raubenheimer, B.C.: 1986, *Astron. Astrophys.* **170**, 187
De Jager, O.C.: 1987, Ph.D. thesis, Potchefstroom University (unpublished)
De Jager, O.C., et al.: 1988, *Astrophys. J.* **329**, 831
Hart, J.D.: 1985, *J. Statist. Comput. Simul.* **21**, 95
Harwit, M., Biermann, P.L., Meyer, H. Wasserman, I.M.: 1987, *Nature* **328**, 503
Kanbach, G., et al.: 1980, *Astron. Astrophys.* **90**, 163
Leahy, D.A., et al.: 1983a *Astrophys. J.* **266**, 160
Leahy, D.A., Elsner, R.F., Weisskopf, M.C.: 1983b, *Astrophys. J.* **272**, 256
Lehman, E.L.: 1959, *Testing Statistical Hypotheses*, Wiley, New York
Mardia, K.V.: 1972, *Statistics of Directional data*, Academic Press, New York
Protheroe, R.J., Clay, R.W., Gerhardy, P.R.: 1984, *Astrophys. J. Letters* **280**, L47
Protheroe, R.J., Clay, R.W.: 1985, *Nature* **315**, 205
Protheroe, R.J.: 1985, *Proc. 19th ICRC* (La Jolla) **3**, 485
Protheroe, R.J.: 1988, *Proc. Astron. Soc. Australia* **7**, 167
Quesenberry, C.P., Miller, F.L. Jr.: 1977, *J. Statist. Comput. Simul.* **5**, 169
Watson, G.S.: 1961, *Biometrika* **48**, 109
Weekes, T.C.: 1988, *Physics Reports* **160**, 1