EXTRAPOLATION OF THE NUMBER-FLUX DENSITY RELATION OF RADIO STARS BY SCHEUER'S STATISTICAL METHOD

A. Hewish

(Received 1961 June 14)

Summary

An attempt is made to assess the usefulness of Scheuer's statistical method of analysing the records derived from a phase-switching receiver. A Monte Carlo technique is described which enables the probability distribution of the recorded deflections to be computed rapidly in EDSAC for any assumed distribution of radio sources. Some typical models are considered and it is concluded that, subject to a small enough receiver noise, the statistical method gives information about the source distribution at flux densities considerably lower than those at which the sources may be counted individually. The method is applied to some observations at 178 Mc/s using the method of aperture synthesis. It is shown that the log $N$–log $S$ relation cannot have a uniform slope of $-1.5$. A slope of $-1.8$ yields good agreement provided that the source density is suitably truncated at small flux densities.

1. Introduction

Records obtained during a survey of radio stars may be analysed in two different ways. The usual method is to extract the positions and flux densities of individual sources; in this case the limiting flux density reached in the survey is determined either by the sensitivity of the receiver or by confusion errors arising from the presence of more than one source in the reception pattern simultaneously. To reduce confusion errors to an acceptable level, several authors have suggested that it is necessary to limit the number of sources extracted to about one source in 20 beam areas.

A different analysis of the data may also be carried out in which no attempt is made to isolate individual sources, but which gives statistical information concerning their number and flux density. In this method the record is sampled at equal intervals of time without regard to the occurrence of particular sources. For the case of an interferometer the deflection $D$ at a given instant is defined as the modulus of the quasi-periodic record, while for a total power system it is simply the total deflection as shown in Fig. 1. When a sufficient number of samples has been obtained, the result may be presented as the distribution function $P(D)$ where $P(D) dD$ is the probability of a deflection lying in the range $D$ to $D + dD$. The shape of $P(D)$ depends, of course, on whether an interferometer or a total power system is used and in this paper only interferometric systems are considered.

The quantity $D$ has the dimensions of flux density $S$, and the physical meaning of $P(D)$ is clarified by considering the ideal case of an aerial which receives with uniform sensitivity over a prescribed beam area. In this case, for sufficiently large values of $D$ corresponding to sources of high flux density which are clearly resolved, we have $P(D) = P(S)$ where $P(S)$ is the chance that a source of flux density $S$ to $S + dS$ lies within one beam area; for smaller values of $D$ this equivalence no longer holds, since there is then a finite chance of several sources lying in the beam area at once, and $P(D) \to 0$ as $D \to 0$ while $P(S)$ may increase without limit. When the reception pattern has an arbitrary
profile $P(D)$ and $P(S)$ are similarly related but a constant of proportionality is involved.

The precise relation between $P(D)$ and $P(S)$ has been fully discussed by Scheuer (1957). He showed that it was not possible to derive $P(S)$ uniquely from $P(D)$: what may be done, however, is to assume some relation for $P(S)$ which enables $P(D)$ to be calculated theoretically and then compared with the observed curve. If the experimental $P(D)$ curve agrees with the theoretical one, then the assumed $P(S)$ is consistent with the data. A population of radio sources is usually defined by $N(S)$ where $N(S)$ is the total number of sources per steradian having flux densities greater than $S$; but

$$N(S) \propto \int_{S}^{\infty} P(S) \, dS$$

so that assumed models of $N(S)$ may be directly related to $P(S)$. By this process it is possible, in principle, to extrapolate the observed $N(S)$ relation towards low flux densities where the sources can no longer be counted with accuracy owing to confusion errors. The utility of the statistical method in any practical application will, however, depend on the sensitivity of $P(D)$ to variations in the assumed values of $N(S)$.

In this paper an attempt is made to assess the usefulness of the statistical method. Analytical methods which allow $P(D)$ to be computed when $N(S)$ is given have been described by Scheuer (1957). The computations are laborious, however, and do not lend themselves readily to rapid comparison of different $N(S)$ models. A method has therefore been adopted in which $P(D)$ is computed by a Monte Carlo technique using the electronic computer EDSAC II. The basis of the method is to consider a typical beam area and to populate it, with appropriate statistical fluctuations, according to some assumed $N(S)$ model. The different contributions to the total deflection are then combined with random phase and the computation is repeated until $P(D)$ may be determined with sufficiently small statistical uncertainty. Using pseudo-random numbers generated in the computer it is possible to derive
$P(D)$ for a given model in a matter of minutes. The computational method
is given in Section 2, and in Section 3 the general utility of the statistical method
is discussed in the practical case where receiver noise must be taken into account.

Finally, in Section 4, the method is applied to results obtained at 178 Mc/s
in a survey using the principle of aperture synthesis. The radio observations
have been fully described in another paper (Scott, Ryle and Hewish 1961)
where it was explained how the observed $P(D)$ relation was automatically
derived from the computer as part of the routine synthesis programme.
A summary of the available data concerning the observed $N(S)$ relation
obtained during this and other surveys has also been given (Scott and Ryle
1961) and it has been shown that effects due to the angular diameter and
clustering of sources can only be relevant to a small fraction of the total number
(Leslie 1961). In the present paper the statistical method is used to extrapolate
the observed $N(S)$ relation to smaller flux densities. It is shown that the shape
of the observed $P(D)$ relation cannot be explained by any $N(S)$ model such that
$N(S) \propto S^{-1.5}$ for all values of $S$. Models with $N(S) \propto S^{-1.8}$ are in good agreement
with the observations, provided that the source density is appropriately
truncated at small flux densities.

2. Computation of $P(D)$ by a Monte Carlo method

For a phase-switching interferometer the value of $D$ at any given sampling
instant is the resultant of a large number of contributions of different amplitude
and phase arising from the sources lying within the beam area at that moment.
The essence of the Monte Carlo method is to consider a typical beam area, to
populate it randomly with sources whose mean distribution is defined by some
$N(S)$ model and to sum the different contributions with random phase. The
computation is repeated for a large number of hypothetical beam areas, the
population being varied randomly from sample to sample, until the theoretical
$P(D)$ may be derived with a reasonable statistical certainty. About 3000 samples
were usually taken which gave statistical errors comparable to those of the
observations. The computation will now be discussed in detail.

(i) The correction for the experimental reception pattern.—The basic compu-
tation was carried out for an assumed ideal reception pattern having uniform
sensitivity over the entire beam area. With a realistic reception pattern a source
of given flux density may give rise to a variety of deflections according to
its position in the beam. This effect could have been introduced into the
computation without difficulty by means of another parameter subject to random
variations, but it involved less computing time to adopt an alternative procedure
in which the $N(S)$ model was suitably modified as a preliminary step. Since
the sources lying within each interval of flux density $S$ to $S + dS$ will give
deflections of different magnitudes appropriate to a random sampling of the
reception pattern, the result obtained when using a graduated reception pattern
is exactly simulated by smoothing the $N(S)$ model. The smoothing process
is an exact convolution if the reception pattern is formulated as the probability
of obtaining a given deflection when a constant source is randomly positioned
within it. The essential randomness is then introduced in calculating the
random population of the beam area as discussed below. When calculating in
this way it is assumed that both the phase and amplitude of the quasi-periodic
record may be considered as independent random variables. This assumption

\*
is justified if, as in the present case, the interference pattern contains many cycles.

The equivalent areas of the experimental 178 Mc/s reception pattern, considered constant over small intervals of relative sensitivity, are shown in Fig. 2; these values were derived from the published reception pattern (Scott, Ryle and Hewish 1961) and include the effect of side lobes. The logarithmic intervals adopted are the same as those employed in the main computation discussed below.

![Graph showing beam area vs relative sensitivity](https://example.com/graph.png)

**Fig. 2.**—The experimental 178 Mc/s reception pattern expressed as increments of beam area for equal logarithmic intervals of relative sensitivity.

The result of convolving $N(S)$ with the probability distribution of the reception pattern may be denoted by $N(D)$ where $N(D)$ is the total number of deflections $> D$ which combine to give the net deflection appropriate to each beam area. It should be noticed that whereas $N(S)$ is expressed as the number of sources per steradian, $N(D)$ is the number of deflections per beam area. When $N(S)$ obeys a simple power law, e.g. $N(S) \propto S^{-1.5}$, $N(S)$ and $N(D)$ are related by a simple scale factor; but when $N(S)$ is terminated, or suffers a change of slope, the corresponding change in $N(D)$ will be smoothed by the convolution process. An example of this is shown in Fig. 3 where $N(D)$ is given for an abruptly terminated $N(S)$ model.

(ii) *Populating the sample beam area.*—For simplicity the full range of $D$ was partitioned into discrete intervals $\Delta D$ such that $\log_{10} \Delta D/D = 1/10$. Within a given partition the $\Delta N$ deflections were all assumed to have the same magnitude. The first stage of the main computation was to populate the partitions randomly according to an average distribution defined by some particular model $N(D)$. This process is equivalent to constructing a typical beam area containing a random distribution of sources of different flux density. The manner in which the required population was computed depended upon the value of $\Delta N$ and four different methods were employed.

(a) $\Delta N \leq 1$. This case corresponds to the small chance of obtaining a large deflection in one beam area. Here it was sufficient to generate a random number
in the range 0–1000 and to add a deflection if the random number fell inside a given range of $1000 \times \Delta N$ integers. Different selections of integers were chosen to represent the different partitions and one trial was allowed for each partition. This method will lead to errors, however, if there is a finite chance of obtaining more than one deflection in a given partition during a run of, say, 3000 sample beam areas, that is, when $3000 \Delta N^2 \sim 1$ or $\Delta N \sim 1.8 \times 10^{-2}$.

![Diagram](https://example.com/diagram.png)

**Fig. 3.** $N(D)$ derived by convolving the $N(S)$ model (c) depicted in Fig. 5 (b) with the reception pattern at expressed in Fig. 2.

(b) $10^{-2} < \Delta N < 5$. To achieve the correct population when a given partition may contain zero, or more than one deflection, a range of partitions was considered such that $\Sigma \Delta N \sim 10$. In a true random population $\Delta N$ will vary, from sample to sample, by an amount $\delta N$ such that $\delta N^2 \sim \Sigma \Delta N$ where $\delta N$ is approximately normally distributed. For each sample beam area a deviation $\delta N$ was generated by taking a random number and using it to choose one term in a table of random deviates. This determined the value $\Sigma \Delta N' = \Sigma \Delta N + \delta N$ for the sample and a further $\Sigma \Delta N'$ random numbers were then generated and used to distribute the $\Sigma \Delta N'$ deflections amongst the different partitions with a probability proportional to $\Delta N$.

(c) $5 < \Delta N < 50$. For small enough values of $D$ there is a negligible chance of obtaining zero deflections within any partition. In this case it was sufficient to compute a series of values $\Delta N' = \Delta N \pm \delta N$ in order to specify the random
population. One random number was therefore generated for each partition and used to assign a random deviation as in case (b).

(d) $\Delta N > 50$. When the number of deflections in any partition is sufficiently large, corresponding to many weak sources in the beam area simultaneously, there is no longer any necessity to compute the random walk summation by Monte Carlo methods. The contribution from any one partition will be a vector of random phase whose amplitude is governed, from sample to sample, by the well-known Rayleigh distribution function. For many of the $N(D)$ models a simple power law was adopted in the region of small $D$ and in this case the contribution due to the Rayleigh component was easily calculated. If $N(D) = AD^n$, where $A$ is constant, the mean square deflection $d(\bar{D}^2)$ due to components in the range $D$ to $D + dD$ is given by

$$d(\bar{D}^2) = nAD^{n+1} dD$$

so that

$$\bar{D}^2 = \int_D^\infty nAD^{n+1} dD$$

giving

$$D_{\text{RMS}} = \left( -\frac{nN_0}{n+2} \right)^{1/2} D_0.$$}

To compute the precise value of the deflection in each sample beam area a random number was generated and used to assign a deviation drawn at random from a table of deviations appropriate to a Rayleigh distribution having a standard deviation given by $D_{\text{RMS}} \sqrt{2}$.

For some of the models $N$ tended to a constant value for small values of $D$ and $D_{\text{RMS}}$ had to be derived by numerical integration.

(iii) The random walk summation.—Having populated a sample beam area it remains to add the component deflections with random phase. For each component a random phase angle was generated by choosing a unit vector from a table of vectors tabulated at $3^\circ$ intervals in the range $0 - 2\pi$, the choice being determined by a random number. Throughout the computation random numbers were generated as required by the multiplicative process $M_{n+1} = M \cdot M_n$ (modulo $2^{39}$) with $M = 5^{18}$ and $M_0 = 1$. Pseudo-random series of this nature are non-repetitive within a period of $2^{37}$ (Moshman 1954) and an initial test was carried out to check the randomness in a run of the required length. As an additional safeguard an arrangement was made whereby the calculation was halted if $M_n = M$; no such event occurred.

The resultant of the random walk summation for each sample beam area was accumulated as a histogram, together with the total population of each partition.

(iv) Receiver noise.—In any practical determination of $P(D)$, deflections arising from noise in the receiving system will be present in addition to those due to signals from the aerials. In the method of aperture synthesis the output of the receiver is recorded on punched tape, and the final integration, normally performed by the output filter of a receiver, is carried out by a convolution process in the computer itself. The equivalent filter is arranged to have a narrow frequency response centred on the mean frequency of the quasi-periodic signal and this ensures that noise fluctuations alone will produce a quasi-periodic output whose modulus is specified by a Rayleigh distribution. The deflection
due to receiver noise has a character which is thus exactly equivalent to case (ii(d)) and it may be treated in precisely the same manner. Putting $D_{\text{RMS}} = (D^2 + D_N^2)^{1/2}$, where $D_N^2$ is the mean square deflection due to noise alone, makes allowance for both quantities simultaneously.

(v) The validity of the Monte Carlo method.—Because a number of approximations have been used in the computation it is desirable to test the validity of the method by using it to reproduce well-known results before proceeding to unknown cases. Two such tests were applied. In the first a model was taken such that $N(D) \propto D^{-1.5}$ and the random population was assigned as described above. In the random walk summation, however, the individual deflections in all the partitions were given equal magnitudes. In this case $P(D)$ is the well-known Rayleigh distribution and it is seen from Fig. 4 (a) that the Monte Carlo result is in very good agreement.

**Fig. 4 (a).—**The continuous curve shows a Rayleigh distribution. The histogram indicates the same result obtained by the Monte Carlo method.

**Fig. 4 (b).—**$P(D)$ derived analytically by Scheuer for $N(S) \propto S^{-1.5}$ (continuous curve) compared with the histogram given by the Monte Carlo method.

© Royal Astronomical Society • Provided by the NASA Astrophysics Data System
As a second test the same model was taken and the computation carried out in the normal way. For this model, corresponding to a uniform spatial distribution of radio sources, \( P(D) \) has been computed analytically by Scheuer (1957) and both results are shown in Fig. 4 (b); the agreement is again seen to be close. It should be noted that “curve fitting” plays no part in the Monte Carlo method. A given model of \( N(S) \), and hence of \( N(D) \), defines the scale of \( D \) absolutely.

3. The value of the statistical method

To estimate the usefulness of the statistical method it is necessary to investigate the sensitivity of \( P(D) \) to changes in \( N(D) \). Only when such changes are greater than the statistical uncertainty inherent in a practical determination of \( P(D) \) can the method be usefully applied. Initially it is assumed that the receiver is ideal, so that noise components do not contribute to \( P(D) \). Two types of model have been considered, as shown in Figs. 5 (a) and 5 (b), such that \( N(S) \propto S^{-1.5} \) and \( N(S) \propto S^{-1.8} \); the latter model is more relevant to the actual distribution of sources. In both cases the effect of arbitrarily terminating \( N(S) \) at various values of \( S \) was investigated. \( P(D) \) corresponding to the different cases is plotted in Figs. 6 (a) and 6 (b); for slope \(-1.5 \) a significant change in \( P(D) \) can be detected when \( N(S) \) is terminated at \( S = 0.01 \), corresponding to the level at which there are 50 sources per beam area. For slope \(-1.8 \) there is a more pronounced change in \( P(D) \) when \( N(S) \) is terminated at \( S = 0.01 \), where there are 250 sources per beam area.

It is therefore clear that, in the absence of receiver noise, the statistical method gives information about \( N(S) \) at flux densities considerably lower
than those at which the sources may be counted individually. To investigate the effect of receiver noise the model shown in Fig. 5 (b) was taken and $P(D)$ computed for two different RMS noise levels. The results are shown in Figs. 7 (a) and 7 (b). While $P(D)$ is modified, as would be expected for sufficiently large noise levels, it is seen that the sensitivity of $P(D)$ to changes in $N(S)$ is not appreciably diminished provided that the maximum of $P(D)$, in the absence of receiver noise, occurs at a flux density somewhat higher than the RMS noise level. This result is to be expected in view of the Rayleigh-like character of $P(D)$. When the noise level is high enough to cause a significant displacement of the maximum of $P(D)$, but not so high that the shape of $P(D)$ is dominated by it, the receiver noise must necessarily be accurately known if the statistical method is to be applied without loss of precision.

The discussion of the preceding sections has been confined to interferometric systems. For a total power recording the situation is similar but the maximum of $P(D)$ is then located at a flux density near the mean d.c. level of the receiver output. While it is simple to compute $P(D)$ for a given $N(S)$ model by the

**Fig. 6 (a).—**$P(D)$ derived from the $N(S)$ models illustrated in Fig. 5 (a).

**Fig. 6 (b).—**$P(D)$ derived from the $N(S)$ models illustrated in Fig. 5 (b).
method described in Section 2, except that the contributions to $D$ are added directly and not with random phases, it is difficult in practice to define a flux density corresponding to the mean receiver output. The comparison of theoretical and observational $P(D)$ relations must then rely on the shape of the curves and the position of the maximum cannot be used as a criterion of agreement.

4. Application of the method to observations at 178 Mc/s

As described in detail in another paper (Scott, Ryle and Hewish 1961) the reduction of the observations made at 178 Mc/s using the method of aperture synthesis included the automatic computation of $P(D)$. In Fig. 8 is shown the experimental curve derived from a survey of $\sigma$-3 steradians ($\alpha = 20^h 40^m - 19^h 15^m$, $\delta = 40^\circ - 44^\circ$); the RMS noise level was determined in a separate experiment and found to be $0.085 \times 10^{-26}$ w.m$^{-2}$ (c/s)$^{-1}$. Counts of individual sources derived from this survey, and from "whole sky" surveys using different arrangements of the 178 Mc/s interferometer, gave the relation $N(S) \propto S^{-1.8}$ for sources having flux densities greater than $2 \times 10^{-26}$ w.m$^{-2}$ (c/s)$^{-1}$ (Ryle and Clarke 1961).

In attempts to account for the observed $P(D)$ curve in terms of appropriate models, attention was concentrated in the region of the maximum of $P(D)$, and models were sought which fitted this portion of the curve most closely.
For larger values of $D$ the statistical uncertainty is relatively greater since $P(D)$ is small; moreover, we have

$$\int P(D) \, dD \propto N(S)$$

(cf. Section 1) for large $D$ so that no information is gained by studying $P(D)$ in this region, where $N(S)$ is already known from the observed source counts.

![Graph showing $P(D)$ derived from observations at 178 Mc/s.](image)

**Fig. 8.—** $P(D)$ derived from observations at 178 Mc/s.

As a first attempt to account for the observations, models were taken such that $N(S) \propto S^{-1.5}$. The three models are illustrated in Fig. 9(a) and the corresponding $P(D)$ curves are given in Fig. 9(b). It is immediately clear that none of the models is satisfactory. From the trend of the curves it can be seen that the maxima are either too high (model (c)) or occur at too large a flux density (models (b) and (a)). Even without evidence provided by individual source counts it may be concluded that no model having $N(S) \propto S^{-1.5}$ can give a satisfactory explanation of the observed $N(S)$ relation.

In a second attempt to find a suitable model the observed $N(S)$ relation was extrapolated at constant slope ($-1.8$) and successively truncated at a series of flux densities. These models were discussed in a different context in the previous section and are shown in Fig. 5(b). The corresponding $P(D)$ curves are shown in Fig. 7(a) and it was found that model (e) fitted the observed $P(D)$ within the statistical error. For this model the truncation occurs at a flux density of $0.35 \times 10^{-26} \text{ w.m}^{-2} (\text{c/s})^{-1}$ corresponding to the level at which one source occurs, on the average, in two beam areas.

It was also possible to account for the observations by extrapolating from a flux density near the limit of the observed number counts with a slope of $-1.5$ and truncating at a flux density of $0.18 \times 10^{-26} \text{ w.m}^{-2} (\text{c/s})^{-1}$. Models of this type and the corresponding $P(D)$ curves are indicated in Figs. 10(a) and 10(b).
Fig. 9 (a).—Three models such that $N(S) \propto S^{-1.5}$ used in an attempt to explain the observed $P(D)$. $N(S)$ obtained directly from source counts at 178 Mc/s is denoted by the thick line.

Fig. 9 (b).—$P(D)$ computed from the models shown in Fig. 9 (a). The observed $P(D)$ is also shown (thick line).
Fig. 10 (a).—N(S) models in which the observed source counts are extrapolated with a slope of $-1.5$.

Fig. 10 (b).—P(D) computed for the models shown in Fig. 10 (a). The observed P(D) is shown (thick line).
Fig. 11 (a).—Models in which the observed source counts are extrapolated with a slope of $-1.0$.

Fig. 11 (b).—$P(D)$ computed for the models shown in Fig. 11 (a).

Since a sudden truncation corresponds to an unrealistic physical situation another series of models was taken in which the $N(S)$ relation was extrapolated to zero flux density with a slope of $-1.0$. This series is shown in Figs. 11 (a) and 11 (b) and it was again possible to derive a curve in reasonable agreement with the observations.
The three models which fitted the observations most closely are shown together in Fig. 12 and it appears that any satisfactory model must lie within the boundaries indicated. It is interesting to notice that the number of sources begins to depart appreciably from the expectations of a linear extrapolation of the observed source count at a flux density in the range \(0.1\) to \(1.0 \times 10^{-26}\) w.m\(^{-2}(\text{c/s})^{-1}\). In view of the strong evidence favouring an extragalactic origin of the radio sources (Ryle and Clarke 1961) this decrease is presumably a cosmological effect. The discontinuities in the curves will, of course, be smoothed by any spread in absolute luminosity. The magnitude of this effect may be estimated with reference to the smoothing action of the reception pattern already discussed in Section 2(i). Fig. 3 indicates the rounding-off which would occur for a sudden cut-off in the source distribution if the spread in absolute luminosity were of the order of 100 : 1.

Mullard Radio Astronomy Observatory,
Cavendish Laboratory,
Cambridge:
1961 May.

References