Mon. Not. R. astr. Soc. (1982) **199**, 683–705

The clustering of quasars from an objective-prism survey

Adrian Webster *Mullard Radio Astronomy Observatory, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE*

Received 1981 September 16; in original form 1981 July 23

Summary. The positions and redshifts of 108 quasars from the Cerro Tololo objective-prism survey are subjected to Fourier Power Spectrum Analysis in a search for clustering in their spatial distribution. It is found that, on the whole, these quasars are not clustered but are scattered in space independently at random. The sole exception is a group of four quasars at $z = 0.37$ which has a low probability of being a chance event and which, with a size of about 100 Mpc, may therefore be the largest known structure in the Universe. The conclusions disagree with Arp's analysis of this catalogue: his 'clouds of quasars' ejected by certain low-redshift galaxies, for example, are attributable to sensitivity variations among the different plates of the survey. It is shown that analysis of deeper surveys is likely to show up quasar clusters even at high redshift, and could therefore provide a useful new cosmological probe.

1 Introduction

The manner in which the quasars are distributed in space is of interest for a number of reasons. Foremost is the information the distribution contains on the large-scale arrangement of matter in the Universe since, on a popular hypothesis, the quasars lie at the great distances indicated by their redshifts and are thus scattered throughout a large volume; if they are grouped on scale sizes of say $0.5°$–$50°$, the corresponding clusters in space would have linear dimensions of order 30–5000 Mpc. Another reason is the possibility that there exist associations of quasars with different redshifts which have a very low probability of being chance superpositions of objects at different distances. Much of the evidence in support of the hypothesis of non-cosmological redshifts is of this kind, but is controversial because objective *a priori* statistical tests have not commonly been applied in the past.

For these reasons we analyse the clustering in a new sample of optical quasars, namely the Cerro Tololo Interamerican Observatory (CTIO) catalogue compiled by Osmer & Smith (1980) using a Schmidt camera equipped with an objective prism. This sample contains 108 quasars, and is therefore of a useful size for a statistical analysis; moreover it is particularly uniform since the quasars were found in a well-defined region by the application of one standard procedure. The principal statistical method employed here is Fourier Power Spectrum Analysis (PSA), which was originally developed to study the distribution of radio sources on the celestial sphere (Webster 1976a), but which is flexible enough to be applied to the quasar sample.

In a sense the distribution of quasars on the celestial sphere has been investigated already, for the distribution of the radio sources has been studied in some detail and quasars make up a substantial fraction of the radio source population. The conclusion of many of these studies (Pearson 1974; Webster 1976b, 1977; Seldner & Peebles 1978; Fanti, Lari & Olori 1978; Masson 1979) is that there is no clear structure apparent in the distribution of the radio sources: they are scattered at random on the celestial sphere, or nearly so, from which it may be inferred that there is no very marked structure in the large-scale distribution of material in the Universe. In order to improve the sensitivity of the search for structure, two courses are open: radio surveys of greater depth may be made which provide more sources per steradian for analysis; or catalogues of objects of known redshift may be analysed, since a clustering test is more sensitive when all three coordinates of the objects are known. The latter course is pursued here.

One study has already been published of the distribution of the quasars in the CTIO catalogue. Arp (1980) has investigated various aspects of the distribution by a method which involves no statistical test and is therefore quite different from that in the present paper. Some doubt has been attached to the validity of Arp's method in the past, largely because of its subjective nature, but there now exists an opportunity to check its accuracy because one of the results of its application to the CTIO catalogue is the discovery of marked clustering among the high-redshift quasars, and this may be treated straightforwardly by the PSA method. This comparison concerns directly only one aspect of one catalogue, but is of wider importance since agreement between the results will increase the confidence which may be attached to the subjective method in the other cases to which it has been applied.

2 Preliminaries

2.1 INSPECTION OF THE DATA

The survey is made up of contiguous fields, each about $5^\circ \times 5^\circ$ in size and centred on declination -40°. In order to avoid difficulties associated with galactic obscuration, we concern ourselves only with the high-latitude part of the survey, which extends from $21^h\,50^m$ to $04^h\,00^m$ and is compiled from 15 plates, 14 of which lie entirely within this range of RA. The region is highly elongated, being about 71° long in RA by only 5° in Dec. Fig. 1 is a plot of the positions of all the 108 catalogued quasars in this region and Fig. 2 is a plot of their redshift against RA.

Two aspects of Fig. 1 are apparent to the eye. The first is the lack of very strong clustering. There are no obvious groupings with memberships of for example 10 or more, and it may be anticipated that in the statistical analysis which follows we shall chiefly be concerned with establishing whether the quasars are weakly clustered or are not clustered at all. The second aspect is an apparent change in the surface density of quasars which occurs at RA $\sim 01^h\,30^m$: earlier than this the density seems systematically higher.

This latter effect is also seen in Fig. 2. The high-redshift quasars ($z > 1.8$) show a moderate decrease in density, and the decrease in the low-redshift objects is particularly marked: only one of the 20 quasars with $z < 1.8$ falls later than $01^h\,35^m$. A second phenomenon to be seen in Fig. 2 is that three of the seven quasars with redshifts smaller than unity have identical redshifts ($z = 0.37$) and lie within half a degree of each other in RA at $23^h\,20^m$. Together with a fourth at $z = 0.36$, 3.5° away, they form a tight group whose statistical significance it will be of interest to assess. A third effect is a marked variation of the number of quasars with redshift; this is attributable to the well-known observational artefact that the probability of

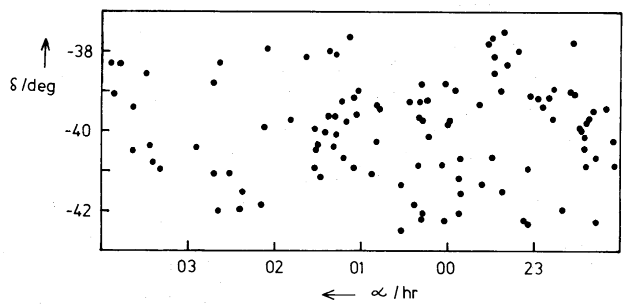

Figure 1. The distribution of the CTIO quasars in RA and Dec.

detecting a quasar by a spectroscopic method depends on whether its redshift is such that a strong emission line is brought into the instrumental passband. This variation has nothing to do with any true grouping of the quasars in the radial direction, so it is necessary to eliminate its effects from the statistical analysis.

2.2 THE LARGE-SCALE DENSITY VARIATION

The large-scale variation of the surface density of quasars between the early and the late parts of the survey is now investigated further, first by assessing the significance of the effect quantitatively.

The number of quasars in each of the 14 plates which are entirely contained within the survey (i.e. excluding plate number 16985) is listed by Osmer (1981). A boundary between the two parts of the survey was chosen, rather arbitrarily, to lie between plates 16467 and 16459 (at about $01^{\mathrm{h}}\,27^{\mathrm{m}}$). The eight plates earlier than this have a mean count of 9.75 quasars per plate, and the six later 4.67; the density differs therefore by just over a factor of 2. Is this difference statistically significant, or could the counts in the two sets of plates reasonably be considered to be two independent samples drawn from a single probability distribution? The choice of statistical test to answer this question is governed by our

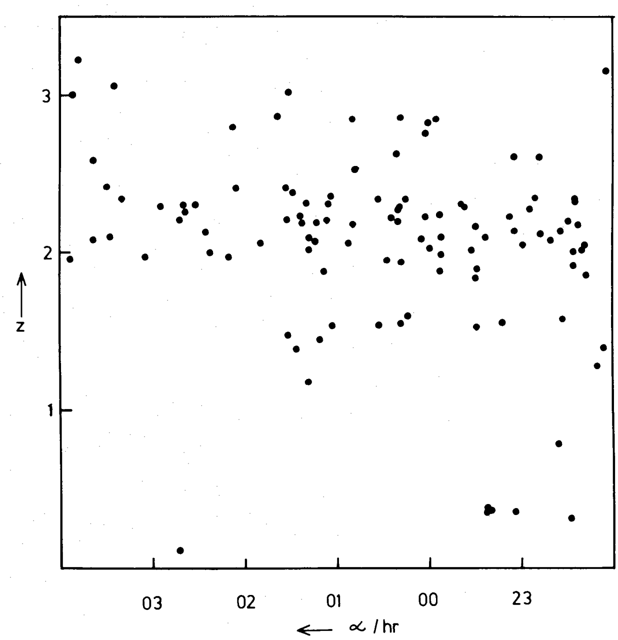

Figure 2. The distribution of the CTIO quasars in RA and redshift.

unwillingness to specify, even tentatively, the type of the underlying probability distribution. It would be possible, for example, to compare the two samples with a Poisson distribution, but this would beg the question of clustering, for only a population of independently and uniformly distributed quasars would obey Poisson statistics. A distribution-free procedure is required, so the Kolmogorov-Smirnov two-sample two-tailed test was chosen (Lindgren 1976). The result was found that the two parts of the quasar sample are significantly different at the 1 per cent confidence level, so there are good grounds for supposing that the quasar densities are truly different.

There are two possible reasons for such a difference:

(1) An intrinsic variation in the density of quasars in space. This would be most remarkable if true because the variation affects simultaneously quasars of all redshifts, and on the cosmological hypothesis this implies a correlation between quasars separated by distances of the order of the Hubble radius.

(2) A variation of the overall sensitivity of the survey, as determined by such factors as the seeing, the telescope, the photographic plate and its processing, and the physiological and psychological influences in the recognition of quasar spectra on the plates when scanning by eye.

Fortunately there is a simple method available for assessing whether the second possibility applies. If the sensitivity varies, it is presumably the faint quasars whose likelihood of detection is most affected, and this leads us to expect a relative scarcity of faint quasars on the Late 'low-sensitivity' plates. Osmer & Smith (1980) present the magnitudes of all the quasars, as measured by individual photometry, so we simply tested for a difference in these magnitudes between the Early and Late quasars. Again the Kolmogorov-Smirnov test was applied because the true distribution of the magnitudes is unknown, and we asked whether the measured magnitudes of the 28 Late quasars could reasonably have been drawn from the same distribution as the 80 Early. This time the one-tailed test was applied because we are only interested in finding out if the Late quasars are brighter than the Early. It was found that indeed they are brighter, and the difference is significant at the 1 per cent level. This supports the idea that the difference in quasar surface density between the Early and Late plates is caused by a variation of the sensitivity.

There is other evidence to the same effect. I am indebted to Dr M. Smith firstly for pointing out that the atmospheric seeing is likely to be the major variable in a survey of this kind, and secondly for providing information from the observing log which supports the idea that seeing is indeed principally responsible for the difference between the Early and Late samples. Plates 17050 and 17051 in the Late sample were both taken on the same night of relatively poor seeing, which would account for the low counts of six and three quasars respectively, whereas the images on Plate 16467 in the Early sample were noted to be of particularly fine quality and the count is 15, the highest in the survey.

The fact that the difference in sensitivity occurs at certain RA is explained by Osmer's (1981) remark that seven of the plates, taken at the so-called middle epoch, seem less sensitive than the rest. These plates are principally located at the Late end of the survey.

In summary, it seems that the existence of plate-to-plate sensitivity variations is well established and that something is understood of their pattern. This information is of the greatest importance when analysing for clustering. First, it leads us to expect that there will be spurious effects in those PSA statistics whose wavelength is comparable with or greater than the diameter of a plate (5°). Second, it is possible to reduce their influence by choosing not to analyse the whole catalogue, but to concentrate instead on a limited range of RA within which the plate material is likely to be less inhomogeneous: accordingly, much of our

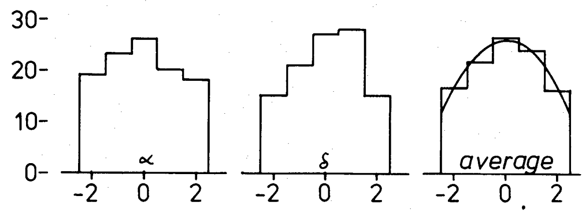

Figure 3. Histograms showing the average marginal distribution of the quasars on the photographic plates. (a) The distribution in RA. (b) The distribution in Dec. (c) The average of the histograms in (a) and (b); the curve is the best-fitting parabola. In these diagrams the bins are 1° wide, the horizontal scale indicates the offset from the plate centre in degrees, and the vertical scale gives the number of quasars per bin.

analysis is confined to the Early RA range. This restriction is not expected to eliminate the variations entirely, and it is likely that rms sensitivity variations of many tens of per cent remain between the plates. Third, not all of the previous clustering analyses of the survey have taken account of sensitivity variations, and one of them seems as a result to have led to erroneous conclusions (see Section 5.1.1).

2.3 SMALL-SCALE SENSITIVITY VARIATIONS

Another possible type of instrumental effect is a sensitivity variation across each plate. Two cases may be distinguished. In the first the variations are peculiar to each plate, and might be caused, for example, by irregularities in the emulsion; they are not easy to study because of the small number of quasars on any one plate. In the second the variations are systematic and are common to all the plates. They might be the result of instrumental effects such as vignetting and can be investigated by superimposing all the plates. Fig. 3(a) shows the distribution of the quasar count in equal strips of RA and Fig. 3(b) similarly in Dec, for all the plates which lie entirely within the survey region. There is a tendency in both histograms for fewer quasars to be found near the edges of the plates than in the centre, indicating that there does indeed exist an instrumental effect such as vignetting. To estimate its strength the RA and Dec strip-counts were averaged (Fig. 3c) and a parabola fitted. By assuming that this 1-D parabolic form reflects an underlying 2-D parabolic dependence of the surface density of quasars on the angular distance from the plate centre, it was found that the rms variation in the surface density attributable to this systematic effect on any of the plates is 0.30 ± 0.12 times the mean density. Fitting a Gaussian to the strip scans and repeating the calculation with a Gaussian radial dependence gave an rms of (0.29 ± 0.12) times the mean density.

The wavelengths in the PSA most likely to be affected are those corresponding to the plate size ($\sim 5°$) and somewhat smaller.

3 The power spectrum analysis

3.1 GENERAL CONSIDERATIONS

The method employed here follows closely that described in Webster (1976a). Only a brief summary of the principal concepts and of some further results is given here.

(1) It is useful throughout to compare the values of various statistics with what would be expected on the assumption that the quasars are distributed in space completely at random and independently of one another. This assumption that there is no particular structure in the quasar distribution is termed the 'null hypothesis' and is here abbreviated as $\mathscr{H}_0$. By adopting this assumption one can readily calculate the mean and standard deviation of each

of the PSA statistics, and since the standard deviation will be used often we introduce a special notation for it: if ξ is any of the statistics then $\sigma_0(\xi)$ is its standard deviation calculated on the basis of $\mathscr{H}_0$.

(2) For the radio sources only two coordinates were available, so only the one- and two-dimensional PSA statistics were needed. Now that the quasar survey is available, with three coordinates per object, a three-dimensional PSA is possible. The basic statistic in this case is $|r_{uvw}|^2$ where

$$r_{uvw} = \sum_{j=1}^{m} \exp i(ux_j + vy_j + wz_j). \tag{1}$$

Here u, v and w are integers, (x_j, y_j, z_j) the coordinates of the jth quasar out of a total of m quasars, and each component of the coordinates lies in the range $-\pi$ to $+\pi$. As for the two-dimensional statistics, by multiplying $|r_{uvw}|^2$ by $2/m$ a statistic I_{uvw} is obtained whose probability distribution on $\mathscr{H}_0$ has a particularly convenient form, and by averaging to give $Q'(k)$ or $Q(k)$ (where the wavenumber $k = 1/\lambda$, and λ is the wavelength) the power of the method is enhanced.

(3) For any number of dimensions, on $\mathscr{H}_0$ $\mathbb{E}(Q'(k)) = 1$ (where $\mathbb{E}$ denotes expectation) irrespective of the value of k. On the other hand, if the quasars are clustered then $\mathbb{E}(Q')$ is greater than unity for small values of k, and decreases towards unity as k is increased. The richness and size of the clusters may be estimated from the behaviour of Q': for small k, $(Q') = q$, where q is the average number of quasars per cluster; and the characteristic wavenumber at which $\mathbb{E}(Q')$ shows the decrease is $k \sim 1/L$, where L is the typical diameter of a cluster.

3.2 THE DISTRIBUTION OF QUASARS IN CELESTIAL COORDINATES

We first present the results of a 2-D PSA in which the redshift is more or less ignored and only the RAs and Decs of the quasars are taken as coordinates. An important reason for this choice of coordinates is that the results do not depend on any assumption about the nature of the redshift, and in particular on whether the redshift is correlated with distance. A hypothetical population of clusters containing quasars with substantially different non-cosmological redshifts might show up in this analysis, but could not if the redshifts were also taken as a coordinate.

The method of 2-D PSA treats the points as though they were distributed on a rectangle in a Euclidean plane, and the null hypothesis is correspondingly that the points are scattered uniformly, independently and at random in that space. The quasars, however, are distributed on the celestial sphere so a projection is needed on to the plane. The proper projection is an equal-area projection, since it will not introduce artificial density irregularities: if the points are uniformly distributed on the sphere their projection will also be uniformly distributed in the plane. In practice, for a survey of a narrow strip in Declination such as this, the simpler Mercator's projection is a sufficiently good approximation to an equal-area projection, and we adopt it here:

$$\begin{aligned} x_j &= 2\pi(\alpha_j - \alpha_{\min})/(\alpha_{\max} - \alpha_{\min}) - \pi, \\ y_j &= 2\pi(\delta_j - \delta_{\min})/(\delta_{\max} - \delta_{\min}) - \pi, \end{aligned} \tag{2}$$

where (α_j, δ_j) are the celestial coordinates of the jth quasar, $\alpha_{\min}$ and $\alpha_{\max}$ are the limits of the RA range ($21^{\mathrm{h}}\,50^{\mathrm{m}}$ and $01^{\mathrm{h}}\,31^{\mathrm{m}}.5$ respectively) and $\delta_{\min}$ and $\delta_{\max}$ of the Dec range ($-42^\circ.5$ and $-37^\circ.5$).

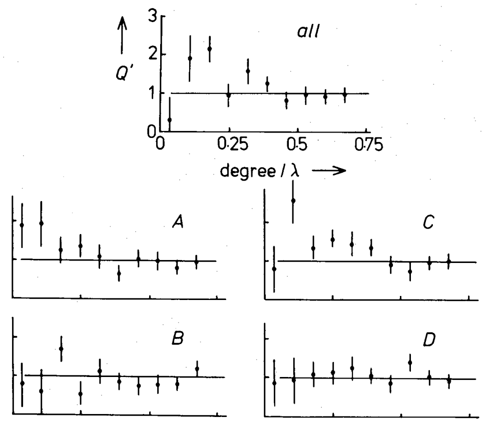

Figure 4. The run of Q' against $k = 1/\lambda$ for the two-dimensional distribution of the quasars in RA and Dec. The uppermost diagram is for the entire sample of quasars and the others are for the four redshift ranges listed in Table 1. The scales are the same in all five diagrams. The line $Q' = 1$, which gives the expected behaviour for unclustered quasars, is drawn in each diagram. The error bars are ± 1 standard deviation, calculated on the basis of the null hypothesis. Note that the standard deviation does not depend on the number of quasars in each sample (Webster 1976a).

The results of the PSA are shown in Fig. 4. The upper plot is for all the quasars, regardless of redshift, and beneath are four more diagrams, each for quasars in a particular redshift range. These four ranges of redshift are somewhat arbitrarily chosen and are set out in Table 1. We note immediately that there is no value of $Q'(k)$ greater than 3 in any of the diagrams: this may be taken directly as an upper limit to the average membership of any hypothetical clusters.

Considering the diagrams in Fig. 4 individually:

(all) For the quasars at all redshifts, the run of Q' against k is of the general form expected for a clustered population. The second and third points in particular are high, and thereafter Q' falls to the value unity. The characteristic scale of the effect is of the order $k = 0.2$ degree^{-1} and its significance may be gauged from the sum Q for all the terms in the first three intervals of k, which is $Q \pm \sigma_0(Q) = 1.76 \pm 0.25$. For comparison the value unity is expected on $\mathscr{H}_0$ in this and all other diagrams.

(A) The plot for the quasars of lowest redshift shows relatively high values of Q' for the first two points and lower values thereafter, which is as expected in the case of clustering. Over the first two intervals, $Q \pm \sigma_0(Q) = 1.91 \pm 0.41$, which differs from unity by just over $2\sigma_0$ so the effect is only of marginal significance.

Table 1. The redshift intervals, and the PSA statistics for clustering on small angular scales.

Redshift range	z_{min}	z_{max}	N quasars	$Q'(k)$, $0.35 < k \times$ degree < 0.7 (all $\pm\ \sigma_0 = 0.09$)	$Q'(k)$ $0.35 < k \times$ degree < 1.4 (all $\pm\ \sigma_0 = 0.05$)
All	0	3.3	83	0.97	0.96
A	0	1	6	0.87	0.94
B	1	1.8	13	0.90	0.97
C	1.8	2.5	53	1.01	0.91
D	2.5	3.3	11	1.04	1.02
(1)	(2)	(3)	(4)	(5)	(6)

(B) In this interval, by way of contrast, there is no clear sign of clustering. The third point does lie above unity by $2.2\sigma_0$ but the neighbouring points are low and the value of Q for the first three points is $Q \pm \sigma_0(Q) = 1.32 \pm 0.25$.

(C) The statistics in this interval of redshift again show the behaviour expected of a clustered population. The second point in particular is high ($Q' \pm \sigma_0(Q') = 2.50 \pm 0.58$) and the third, fourth, fifth and sixth points also lie systematically above the value unity. The values of $Q \pm \sigma_0(Q)$ for sums over the first two, four and six points respectively are 1.67 ± 0.41, 1.50 ± 0.19 and 1.43 ± 0.13, which differ from unity by typically $2\text{–}3\,\sigma_0$, and the characteristic wavenumber of the effect is roughly 0.2–0.4 degree^{-1}.

(D) Here there is no sign of clustering. All the values of Q' are consistent with unity and there is no increase in Q' at small values of k.

The most interesting of these results is the discovery of the signature of clustering in the full sample and in the redshift intervals A and C. Whether this may be attributed to clustering depends on whether any of the spurious effects which came to light in Section 2 could be responsible, and there are three ways of finding this out from the information available at this stage of the analysis (a fourth is discussed in Section 3.3):

(a) The angular scale. The angular scales of the plate-to-plate and the centre-to-edge sensitivity variations are not unique, but are of the general order of the plate size of $5°$. In the PSA, therefore, their characteristic wavenumber is expected to be of order $k \sim 0.2$ degree^{-1}, and it is seen in Fig. 4 that indeed the division between the large and small values of Q' lies at such a wavenumber. It is inferred that the angular scale of the effect is that expected from the sensitivity variations.

(b) The strength. The considerations in Sections 2.2 and 2.3 may be used to estimate the strength of the rms fluctuations in the surface density of quasars attributable to sensitivity variations. The variations in the surface density may also be calculated from the PSA results, since by Parseval's theorem the mean square deviation of the density is related to the sum of the squares of the coefficients in the Fourier series expansion of the density. By comparing these two estimates some idea may be obtained as to how much of the effect seen in the PSA results can be attributed to the known sensitivity variations. The quantifiable aspects of the sensitivity variations are the following. Two of the nine plates are mid-epoch plates, so we take the surface density on these two as equal to the mean on all the mid-epoch plates, specifically 4.0 quasars per plate. The density appropriate to the other seven was similarly taken as 10.125 quasars plate^{-1}. The centre-to-edge sensitivity variations were taken to be of radial parabolic form (Fig. 3c). Together, these sources of variation amount to an rms contribution equal to 0.44 times the mean surface density. Turning now to the PSA, we suppose that, corresponding to each statistic I_{uv} there is a Fourier term in the expansion of the underlying sensitivity variations whose amplitude is a_{uv}, so the probability of finding a quasar in a given element of area is no longer uniform but depends on position as

$$\rho(x,y) \propto 1 + \sum_{uv} a_{uv} \cos(ux + vy + \phi_{uv}) \tag{3}$$

where ϕ_{uv} is the phase. In this case the expectation value of I_{uv} on $\mathscr{H}_0$ is

$$\mathbb{E}\,(I_{uv}) = 2\,[1 + a_{uv}^2(m-1)/4] \tag{4}$$

giving

$$\mathbb{E}\,(Q') = 1 + \langle a^2 \rangle (m-1)/4, \tag{5}$$

where $\langle a^2 \rangle$ is the average value of a_{uv}^2 for the terms included in the sum for Q'. The value of a given I_{uv} for the survey may thus be taken as an estimator of a_{uv}^2. This Fourier term contributes $\frac{1}{2}a_{uv}^2$ to the mean square variation of ρ, so the normalized sum $Q'(k)$ of a number of the I_{uv} may be used as an estimator of $\delta\rho/\rho$, the rms variation of ρ normalized to the mean, contributed by the Fourier terms included. The estimate is:

$$(\delta\rho/\rho)^2 = (Q' - 1)\,\Sigma_\nu/(m-1), \tag{6}$$

where Σ_ν is the sum of the number of degrees of freedom (Webster 1976a). The value calculated in this way from the first six terms in the uppermost diagram in Fig. 4, which encompasses the entire 'clustering' effect seen there, is $\delta\rho/\rho = 0.76$. The known sensitivity variations give $\delta\rho/\rho = 0.44$ and therefore amount to rather more than half of the figure from the PSA. It follows that the remainder may in principle be attributed to true clustering. If so, then over the wavenumber range 0–0.4 degree^{-1} our estimate of the average number of quasars per cluster is $Q \pm \sigma_0(Q) = 1.25 \pm 0.13$. Some caution is required, however: $\delta\rho/\rho = 0.44$ is the sum of the sensitivity variations which were quantifiable, but there are likely to be others. In the first place, averaging the counts, which was necessary in order to calculate the rms variation, smoothes over any plate-to-plate sensitivity variations within each epoch. Secondly, large-scale effects within individual plates (due to emulsion irregularities for example) are not accounted for. Such is the uncertainty in the strength of these contributions that it is not possible to rule out the idea that the clustering-like effect is entirely attributable to the variation. It is inferred therefore from these considerations of the strength that the sensitivity variation is a major contributor and may conceivably account for the clustering effect entirely.

(c) The dependence on m. The possible clustering effect is seen with different strength in the different samples. These samples contain different numbers of quasars and it is of interest to see whether the strength of the effect shows the dependence on m given in equation (5). For this it is assumed that the underlying sensitivity variations do not differ between samples, so the value of any of the a_{uv} is the same for each sample. Accordingly, Q' was calculated for the first three points in each diagram (i.e. $k < 0.2$ degree^{-1}) and plotted against $(m-1)$ in Fig. 5. The points are seen to follow a linear relationship quite well, except possibly that for the redshift interval (A). It will be shown in Section 3.3 that the high value in this interval is attributable to an entirely different cause, namely the group of quasars at $z = 0.37$ mentioned in section 2.1, so it is proper to disregard the point in the diagram for range (A). It is then directly inferred that the dependence of Q' on $(m-1)$ is consistent with that expected if the entire effect is due to sensitivity variation. On its own, this result is not conclusive because certain types of clustering might similarly follow equation (5).

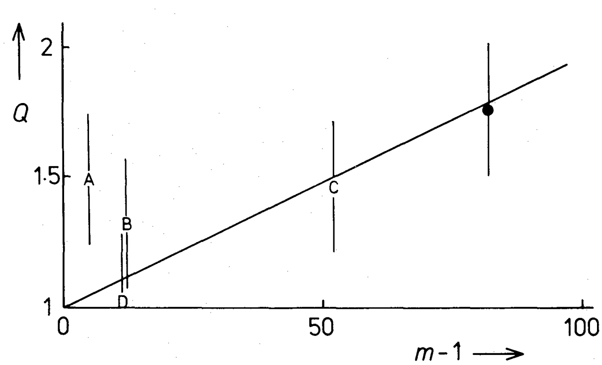

Figure 5. The dependence of Q on m, the number of quasars per sample, for each of the five samples shown in Fig. 4. Q is the value for $k < 0.2$ degree^{-1}.

The general result of these considerations is that the sensitivity variation contributes substantially to the clustering-like effects in Fig. 4, and that the uncertainty in the strength of the variation is such that it could account for the effect entirely.

We now turn our attention away from the large-scale effects and examine the strength of any small-scale clustering, since known coincidences on the sky, on scales smaller than one degree, of quasars of different redshift have led to the suggestion that the redshifts are not cosmological (e.g. Arp & Hazard 1980). Accordingly we discard the first five points in each of the diagrams in Fig. 4, for they relate to effects on large scales, and calculate the value of the statistic Q' for the remaining five (Table 1, column 5). The range is 0.35–0.7 degree^{-1}, so Q' gives an estimate of the strength of the clustering on scales smaller than $\lambda = 1.4°$. The PSA was also extended to smaller wavelengths to increase the statistical efficiency on smaller scales, and Q' for $0.35 \leqslant k \leqslant 1.42$ degree^{-1} is also shown in Table 1 (column 6); it quantifies the strength of clusters on scales $\lambda \lesssim 42$ arcmin. The pair of values for the entire sample is of particular interest in connection with the hypothesis of non-cosmological redshifts, since it is here that clusters containing quasars with widely different redshifts should show up: the values are $Q' \pm \sigma_0(Q') = 0.97 \pm 0.09$ and 0.96 ± 0.05 respectively, which lie within $1\,\sigma_0$ of unity, the expectation value for an unclustered population. There is thus no evidence from these results for clustering on small angular scales of quasars with different redshifts. For the PSA of quasars divided into the several redshift intervals, the results are similar: all lie within about $2\,\sigma_0$ of the value expected on $\mathscr{H}_0$ and there is no sign of clustering.

In order to see what limits this puts on the fraction of quasars in the population which can be members of clusters of small angular scale, it is necessary to introduce a result concerning the expectation value of the PSA statistics when clustering is present. If f_1 is the fraction of the quasars in the population which are individuals, not members of a cluster, and f_n is the fraction which are members of clusters of size n, then

$$\sum_{n=1}^{\infty} f_n = 1$$

and the expectation value of $Q'(k)$ for $k < 1/L$ is

$$q = \sum_{n=1}^{\infty} n f_n .$$

This result is approximate in that it ignores effects at the survey boundary, where a cluster may be divided, but is accurate enough when, as now, the clusters under investigation are small compared with the dimensions of the survey. If we suppose for a moment that there are no clusters with more than two members, $q = 1 + f_2$ and the value of Q' on small scales gives an estimate of f_2 directly. The values found above are entirely consistent with $f_2 = 0$ so there is no need to suppose there are any close pairs in the survey, but a limit to the number of pairs may be found by adding $2\sigma_0(Q')$ to Q' and calculating f_2: for the quasars at all redshifts this gives $f_2 < 0.15$ for $\lambda < 84$ arcmin and $f_2 < 0.06$ for $\lambda < 42$ arcmin. Repeating the calculation on the assumption that the population is made up only of single quasars and of triples gives 2σ upper limits on the same scales of $f_3 < 0.08$ and 0.03 respectively.

It is inferred that the distribution on the celestial sphere of the quasars at all redshifts in the CTIO catalogue contains no evidence for a statistically significant number of close pairs, triples, or multiples of any kind. The values of the power spectrum statistics are consistent with the hypothesis that on small angular scales the quasars are single, unclustered objects, and the 2σ upper limits on the fraction of quasars in the population which lie in pairs or triples are low, of the order 3–15 per cent. On small angular scales the quasars in this catalogue are mostly, and perhaps entirely, single objects.

The analysis therefore lends no support to the claim that quasars are associated with others of different redshift. No attention has been paid to any of the details adduced by supporters of this claim (such as accurate alignments in triples, or patterns in the redshifts with a low-redshift quasar lying between two high-redshift companions) because they are irrelevant: quasars originating in triples or ejected in pairs from an extant quasar will be detected as clustered whether or not the shapes are linear or the redshifts patterned. There is no sign of close triples in the catalogue so there is no sign of close collinear triples. If the quasars do generally originate with others of different redshift it seems necessary to suppose that, due to the incompleteness of the catalogue, no more than one of the quasars in each group is usually recorded; the presence or absence of this type of clustering in deeper and more complete samples therefore appears likely to decide the issue finally.

3.3 THE DISTRIBUTION IN RIGHT ASCENSION AND REDSHIFT

The second analysis described here is a 2-D PSA of the distribution of the quasars in RA and redshift, with no account taken of the Declination. An immediate problem with this choice of coordinates is what meaning to give to the wavelength of a Fourier component whose wave vector does not lie parallel to either coordinate axis, for the units of RA are not the same as those of redshift. This problem is surmounted by choosing a cosmological model (specifically with $H_0 = 75\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $q_0 = 0.05$) and opting to express the separation between any two quasars in terms of gigaparsecs (Gpc) at the present epoch; the wave-number is then conveniently expressed in Gpc^{-1} whatever the direction of the wave vector. A second problem arises from the large range of redshift, which implies that a given angle on the sky subtends a wide range of distances. As a result, it is not very meaningful to attempt a PSA of the whole sample: a given I_{uv} would not correspond to any very well-defined wave-length but instead to a blurred mixture of comoving scales. The PSA was therefore applied only to the four limited redshift intervals in Table 1, in each of which the blurring, although still present, is small. The final decision to be made before applying the PSA was what to do about the observational redshift dependence of the quasar density in the catalogue, brought on by the presence or otherwise of strong emission lines in the instrumental pass band. Here advantage was taken of the fact that this only affects those relatively few Fourier compon-ents whose wavevectors lie parallel to the redshift axis; such contaminated terms were simply discarded (*cf.* the treatment of the drift-scan effect in radio source catalogues, Webster 1976b).

The projection employed is

$$x_j = 2\pi\,\frac{(\alpha_j - \alpha_{\min})}{(\alpha_{\max} - \alpha_{\min})} - \pi, \quad y_j = 2\pi\,\frac{(z_j - z_{\min})}{(z_{\max} - z_{\min})} - \pi,$$

where $z_{\min}$ and $z_{\max}$ are the values from Table 1 for each redshift interval. The wavenumber associated with a given Fourier component is k where

$$(k/2\pi)^2 = u^2\Big/\left[(\alpha_{\max} - \alpha_{\min})\,\frac{\partial S}{\partial \alpha}\right]^2 + v^2\Big/\left[(z_{\max} - z_{\min})\,\frac{\partial S}{\partial z}\right]^2.$$

Here, S is the comoving distance and the partial differentials are evaluated at the mid-point of the field, i.e. at $\alpha = (\alpha_{\max} + \alpha_{\min})/2$, $z = (z_{\max} + z_{\min})/2$. The results are shown in Fig. 6 and are now discussed individually:

(A) In this low-redshift range the points are totally inconsistent with the null hypothesis. Every point is high, typically by $3\text{–}10\,\sigma_0$. The value of $Q \pm \sigma_0(Q)$ for the sum over all the

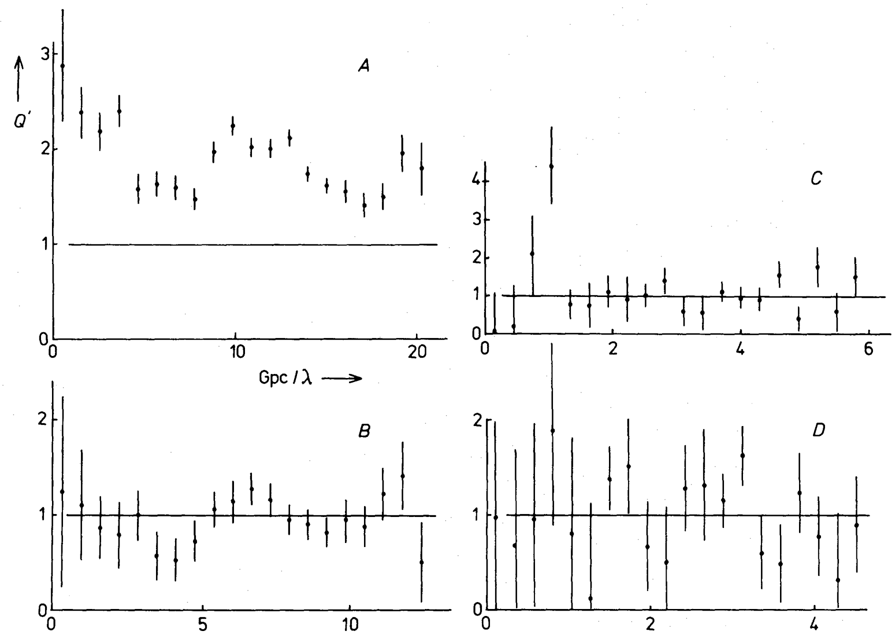

Figure 6. The results of a two-dimensional PSA in RA and redshift for the four ranges of redshift listed in Table 1. The error bars are again $\pm 1\sigma_0$.

points in the diagram is 1.831 ± 0.026, which differs from unity by $31.7\,\sigma_0$. It is not easy to discern a characteristic scale from the diagram as the points do not fall to unity in the wavenumber range presented.

(B) Here there is no sign of clustering.

(C) The third and fourth points are high, suggesting clustering, and thereafter the points are scattered about unity. As it happens, the first four points contain only 1-D components whose wave vectors lie along the direction of Right Ascension so they have already been interpreted. They contain no redshift information and add nothing to the analysis in Section 3.2.

(D) The points are scattered about the value unity and there is no hint of clustering.

The large effect in redshift range (A) is due entirely to the presence of the tight group of quasars at $z \sim 0.37$ noted in Section 2. A full discussion of this interesting group is deferred until Section 4.

In the other ranges of redshift there is no clear clustering phenomenon revealed in this analysis apart from the reappearance of the large-scale effect found previously in (C). The new results in (C) do, however, provide an extra check on the sensitivity variations. At any small wavenumber all the components in the $\alpha - \delta$ PSA should be affected by the sensitivity variation, but in the $\alpha - z$ PSA only those whose wave vectors lie parallel to the RA axis, since the variations have no dependence on redshift. If, on the other hand, the effect is attributable to true clustering of the quasars in space, both sets of statistics should show the same signature of clustering. We therefore compare the low-wavenumber values of Q to see if they are similar or dissimilar. We ignore the waves with wavevectors parallel to the RA axis since they are common to both sets of results and cannot contribute to any difference, and in the $\alpha - z$ results we also discard the waves parallel to the redshift axis as before. In a suitable range of wavenumbers (0–3 Gpc^{-1}) we find $Q \pm \sigma_0(Q) = 1.52 \pm 0.24$ for the $\alpha - \delta$ results and 0.94 ± 0.18 for the $\alpha - z$; the former value is appreciably greater than unity and the latter is easily consistent with unity. The values therefore differ in the sense expected if sensitivity variation is the cause, in agreement with the three previous results. Unlike the previous results the new one depends on the assumption that the redshifts indicate distance.

On small scales there is no sign of clustering in any of the three redshift intervals with $z > 1$. It would be possible to set limits on the average membership of any small clusters from these results by the method used in Section 3.2, but this is deferred until the next section when the present results are combined with those from a 3-D PSA in order to increase the statistical weight.

3.4 A THREE-DIMENSIONAL PSA: SMALL–SCALE CLUSTERING IN THREE DIMENSIONS

A 3-D PSA was carried out to large wavenumbers in order to investigate small-scale clustering of the quasars. In a 3-D analysis the number of Fourier components with wavenumbers between k and $k + dk$ increases as k^2, which is faster than for the lower-dimensional analysis and gives a greater statistical weight.

The projection used is

$$x_j = 2\pi \frac{\alpha_j - \alpha_{\min}}{\alpha_{\max} - \alpha_{\min}} - \pi,$$

$$y_j = 2\pi \frac{\delta_j - \delta_{\min}}{\delta_{\max} - \delta_{\min}} - \pi,$$

$$z_j = 2\pi \frac{z_j - z_{\min}}{z_{\max} - z_{\min}} - \pi.$$

Consistency with our earlier notation has entailed two uses for the symbol z, as the third Cartesian coordinate and as the redshift. Fortunately the uses are well separated and no confusion need result, for the coordinate only appears on the left of the equations and the redshift only on the right.

The results for redshift ranges (C) and (D) are shown in Fig. 7. These are of particular interest in view of the possible clustering on scales of order 100 Mpc found by others in them (see Section 5). In preparing these results, no terms I_{uvw} were excluded from the sums; the redshift modulation and the sensitivity variations were instead allowed for by discarding

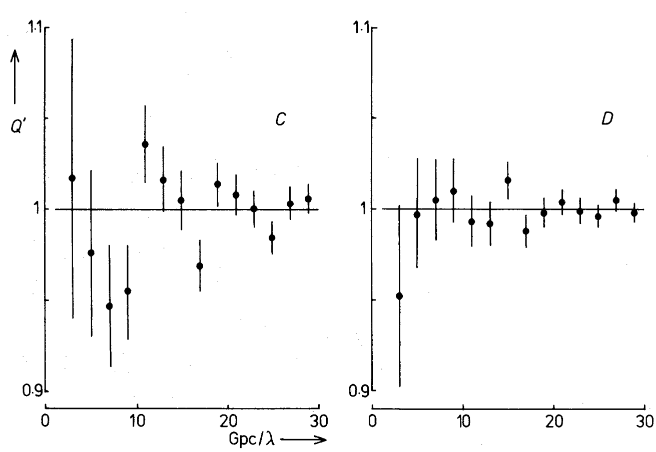

Figure 7. The results of a three-dimensional PSA in RA, Dec and redshift for ranges (C) and (D). The scales are the same in both diagrams and the error bars are $\pm 1\,\sigma_0$.

the first point (for $k = 0.2\,\mathrm{Gpc}^{-1}$), since their contribution is at small wavenumbers in this diagram.

The results in both cases are consistent with the null hypothesis. The values of Q' are scattered about the value unity and there is no tendency for larger values to be found at low values of k. There is no evidence here for clustering on any scale in the range $500 > \lambda/\mathrm{Mpc} > 30$. Ranges (A) and (B) were analysed over $0 < k < 10\,\mathrm{Gpc}^{-1}$ but are not of so much interest and are not presented here. (A) shows the group at $z = 0.37$ and nothing else of interest, whilst (B) is consistent with the null hypothesis.

4 The group of quasars at $z = 0.37$

In this section we turn our attention to the only promising grouping that the PSA has revealed, namely the collection of quasars near $z = 0.37$ and $12^{\mathrm{h}}\,20^{\mathrm{m}}$ which may contain as many as four objects.

4.1 HOW MANY MEMBERS?

Oort, Arp & de Ruiter (1981) include only the two closest members of this group (Q2319-383 and Q2321-375) in their table of potentially associated quasars, but mention in a footnote that the other two (Q2303-391 and Q2322-414) almost certainly also belong. Arp (1980) notes the three quasars with similar RA as being associated but makes no mention of the fourth (Q2303-391). There is thus some uncertainty as to how many quasars the group actually contains.

Here we attempt to cast light on this issue by a statistical method. We consider memberships of two and four, and assess which would be less likely to occur by chance if the quasars are set in place by a uniform Poisson process: this is then supposed to be the more likely to have been produced by a clustering process. We do not consider the possibility that the membership may be three because if Q2322-414 is included with the close pair there is little reason to exclude Q2303-391: the latter is only 2 Mpc further away than the former from the centroid of the pair (81 Mpc as against 79 Mpc) on our cosmological model, where as usual the distances are quoted at the present epoch on the assumption of comotion.

There are six quasars in the range $0 < z < 1$, $21^{\mathrm{h}}\,50^{\mathrm{m}} \leqslant \alpha \leqslant 01^{\mathrm{h}}\,31^{\mathrm{m}}.5$, so to assess the possible cluster membership of two the probability is calculated that, of six points scattered uniformly in a volume V, the closest pair should be separated by a distance no greater than some value r (when $r^3 \ll V$): this probability is

$$\binom{6}{2} 4\pi r^3/3\,V.$$

The separation of the closest pair in the group is 19.4 Mpc so we take this for the value of r, but the value of V to use is not so obvious. If all the quasars in the solid angle Ω out to a redshift of 1 had been found and catalogued, then the volume to adopt would simply be the comoving volume subtended by Ω out to $z = 1$; the probability for a pair on this assumption is shown in Table 2, column 5, line 1. There is, however, good reason to suppose that this is not the correct volume to adopt, because the number of quasars in a small redshift interval would increase rapidly with z owing to the increase in the volume contained in the interval and to the cosmical evolution of the quasar population. The actual sample of quasars shows no such effect. Five of the six have redshifts between 0.3 and 0.4 and only one has a higher redshift. Presumably there are selection effects which counteract the expected increase in the number with redshift and which therefore reduce the effective volume in which quasars

Table 2. The probabilities of the low-redshift group being a chance pair or a chance quartet.

Model for the z-distribution	z_0	z_1	V Gpc^{-3}	QSO density × Gpc^3	Pair probability	Quartet probability $r = 50$ Mpc	$r = 100$ Mpc	$\theta = 2^\circ.25$	$\theta = 4^\circ.5$
Unmodified square-law	0	1	0.47	13	0.001	4×10^{-9}	2×10^{-6}	1×10^{-8}	6×10^{-7}
Square-law with exponential cut-off	0.11	0.60	0.14	43	0.004	1×10^{-7}	7×10^{-5}	9×10^{-8}	6×10^{-6}
Truncated square-law	0.28	0.40	0.032	188	0.016	1×10^{-5}	7×10^{-3}	5×10^{-6}	4×10^{-4}

can be found. As a first attempt at estimating this effective volume, it was assumed that the probability of finding a quasar at a redshift z is $p(z)dz = 0.5\beta^3 z^2 \exp(-\beta z)dz$. This shows an increase proportional to z^2 at low z, as expected from the rate of increase of the comoving volume, and then a gradual cutoff due to selection effects which we have arbitrarily taken to be exponential. The parameter β was estimated from the six sample values of z by a maximum likelihood method, and z_0 and z_1 (the values of z at which the probability takes half its peak value) were taken as bounding the volume V (Table 2, line 2). Another estimate was obtained in an identical way except that the probability was taken proportional to z^2 out to $z = 0.4$ and zero thereafter (line 3). On account of the sharp cutoff this is an extreme case and yields an effective volume bordering on the unreasonably small, but, together with the similarly large volume for the case in line 1, it should bracket the likely range of V. The probability density with the exponential cutoff, whilst arbitrary, is the most likely of the three to yield a reasonable value for V. The pair probability is seen to be of order 10^{-2}–10^{-3}.

To investigate the possibility that the group is a quartet, it was supposed that three of the six quasars had been scattered in the volume V in no special configuration. The probability was then calculated that three more quasars distributed by a uniform Poisson process should find themselves within a small element of volume ΔV centred on one of the first three: this probability is $3(\Delta V/V)^3$. Column 6 of Table 2 shows this probability calculated with $V = 4\pi R^3/3$, where $R = 50$ Mpc is a typical separation of the members of the quartet. Since the probability depends on a high power of R it is shown again in column 7 for $R = 100$ Mpc, which is a little greater than the greatest actual separation. The probability in column 8 was calculated by separating the angular and redshift variables, and therefore takes special account of the smallness of the spread in z: $\Delta V = (\pi\theta^2/\Omega)\, 2\Delta z/(z_1 - z_0)$, where $\theta = 2.25°$ is a typical angular separation of members of the cluster and $\Delta z = 0.01$ is the greatest difference in z between members. In column 9 the same calculation is repeated with $\theta = 4.5°$, which encompasses the widest members.

All the probabilities for the quartet in any line of Table 2 are smaller than those for the pair, usually by a factor of 10^{-2}–10^{-4}. Only in the bottom line are the probabilities comparable, but the sharpness of the assumed z-distribution for this case is not entirely realistic. It is inferred therefore on the basis of these numbers that the statistical case for considering the cluster a quartet is stronger than that for a pair.

4.2 IS THE CLUSTER A CHANCE FLUCTUATION?

The probabilities in Table 2 are now taken as the basis of a discussion of whether the quartet is a real structure, produced by a non-uniform mechanism, or is a fluke generated by a uniform mechanism. For a uniform process and the most reasonable choice of effective volume, the probability lies in the range 10^{-4}–10^{-7} and would seem to be so small that a real structure is indicated. Before this inference may be drawn, however, a number of complications must be investigated.

The first is that the cluster may be an artefact of observational selection caused by the known variation in the sensitivity of the survey (Section 2). Could it be that the plate on which the cluster lies was particularly good for spotting quasars, perhaps on account of exceptional atmospheric seeing? The cluster does indeed lie in a part of the survey covered by good plates (not from the middle epoch). Any consequential increase in the probability of detection is offset, however, by the fact that three of them (Q2319-383, Q2321-375 and Q2322-414) lie very near the edge of the plate, in fact in the region of overlap with the next. Given the strength of the 'vignetting' effect in Section 2 it does not, therefore, appear that the quasars lie in a particularly favourable place for detection. Moreover, there is no increase in the density of quasars at other redshifts in this region, and it is not easy to think of an observational effect which would increase the detectability of low-redshift quasars without simultaneously increasing that of the high-redshift quasars.

Another possibility is that the spread in z is artificially small for observational reasons. The greatest difference between the redshifts of the members is 0.01, which is small when the redshifts are only quoted to two decimal places, and obliges us to consider rounding error. It does not seem though that this can account for the low probability of the quartet, since the probabilities in columns 6 and 7 of Table 2 are based on values of R determined principally by the angular spread of the cluster and not by the spread in redshift, yet they are still interestingly small. Only if the true spread in z before rounding were of the order of 0.04 or greater would the size of the cluster in the radial direction require a substantial increase in the value of R and so lead to a significant increase in the probability, and it is doubtful that the rounding error could be so great. One more possibility may also be ruled out: there is no strong wavelength-dependent structure in the absorption coefficient of the Earth's atmosphere near 3830 Å which could modulate the probability of detecting Mg II λ2798 Å to give a sharp peak at $z = 0.37$.

A completely different consideration is the *a posteriori* way in which the probabilities were calculated; certain parameter values were adopted after studying the observations rather than before, and caution is therefore required. The actual detection of the cluster was *a priori* in that the measurements were fed to a general-purpose computer program which was not specially adapted to them in any way and which revealed the signature of clustering in redshift range (A). The parameters with *a posteriori* values are:

(1) R. The choice of value has a great influence on the probability because it is raised to a high power. It was for this reason that two different values were tried.

(2) θ. As for R.

(3) V. The effective volume was estimated from the sample redshift distribution, which itself might be strongly weighted to peak at ~ 0.37 by the chance presence of the quartet. By reducing the effective volume, this peaking results in our probability estimates being higher than they would otherwise be and any correction would give even lower values.

(4) The cluster membership. This was taken as four because the probability calculated *a posteriori* was found to be lower than for two. Some small correction to the probability, perhaps by multiplication by ~ 3, is therefore necessary to take account of other possible values of the membership which *a priori* would have been considered equally interesting.

Yet another complication has to do with the division of the catalogue into redshift ranges, for without this the group would probably not have been noticed in the PSA: its signal would have been diluted by the noise from the large number of unclustered quasars at higher redshifts. This again seems to require some correction to the probabilities, since the greater the number of ways in which a random distribution of points is looked at, the greater is the chance of something odd emerging. Nevertheless, this correction does not seem to be large, firstly because the division amounts to a few extra ways only, and secondly because the division was entirely *a priori*, being forced upon us by the necessity of reducing the blurring of the comoving wavelength. In no sense did we keep looking at the catalogue in different ways, stopping when finally something odd turned up.

Despite these complications the probabilities in Table 2 are still interestingly low. Observational effects do not seem to be very important and the effects of the *a posteriori* choice of parameter values are not so large as to affect the probability crucially. It is therefore reasonable to take the view that the statistical arguments support the idea that the grouping represents a real structure which is worth investigating further. The statistical arguments would have been simpler and stronger if more than one group had been found in the survey, so a similar group of three low-redshift quasars noted by Oort *et al.* (1981) at $z = 0.06$ is of some interest; it was unfortunately found in a less systematic way, which makes its reality even harder to assess quantitatively and limits the support it lends to the above view.

4.3 WHAT IS IT?

If the quartet is a real grouping, what kind of structure does it trace? It has a claim to be the largest known structure in the Universe, so the question is of considerable interest. The only known structures of comparable size are the superclusters, so it is possible that it is a particularly large specimen of that type. Indeed, Oort *et al.* (1981) have already made out a case for the occurrence of quasars in superclusters. Arp (1980) on the other hand, noting that the large angular scale of the $z = 0.37$ group is similar to that of foreground clumps of galaxies, infers that it lies closer to the Earth than the redshift would ordinarily be taken to indicate and is therefore much smaller than a supercluster. This argument is not compelling, however: Oort (1981) maps out the galaxies in a number of superclusters whose sizes are not far short of the 100 Mpc or so of the $z = 0.37$ group, and finds a wide dispersion in the sizes. There is thus no obvious reason to suppose that there cannot exist specimens as large as 100 Mpc. Indeed, if superclusters with dimensions as large as this do exist, then looking for quasar groups might have been anticipated in advance as the best way of revealing them for the first time. A volume large enough to contain a number of quasars is more likely to contain examples of rare, large superclusters than the smaller volume accessible to galaxy studies, so on these grounds alone one might reasonably expect the quasars preferentially to trace the large end of the supercluster size distribution. Furthermore, large superclusters might be expected to contain more quasars on the average than small, which would contribute to the same effect. The idea that the group is a supercluster therefore seems to be perfectly reasonable.

Another possibility which may be mooted is that the group is an example of some structure larger than a supercluster, and is perhaps one of the first known specimens of the next largest animal in the hierarchical zoo (a superduper cluster? a hypercluster?). Little can be said about this hypothesis on the basis of the present data, but an analysis of the distribution of faint galaxies in the field would be of great interest.

The shape of the structure outlined by the four quasars is somewhat peculiar. In the first place the group is not centrally condensed. None of the quasars lies particularly close to the centroid: their distances from that point are 22, 41, 58 and 59 Mpc. Secondly the group is flat and lies nearly in the plane of the sky. The greatest thickness in redshift corresponds to a comoving thickness of 27 Mpc, in contrast with a greatest angular width corresponding to 95 Mpc. This flatness fits with the idea that it is a supercluster since the known examples are often extended rather than spherical (e.g. Oort 1981), but it could have other causes. Rounding error in the redshift measurements might in principle contribute, although we have already seen that this could not easily have reduced the apparent depth from as large a value as 100 Mpc. Another possibility is that the cluster is in fact spherical and that the quasars are not expanding with the substrate but are moving more slowly, which would give rise to a smaller dispersion in redshift than if they were comoving. This explanation argues against a low-density cosmological model, in which gravitational forces would be unimportant on large scales at such a late epoch. The present evidence is too meagre for any such cosmological inference to be drawn, but it is worth seeing whether the relatively low dispersion in z persists as a systematic property of other quasar groups as and when they are discovered.

Throughout this discussion it has been assumed that there are no more than four quasars in the group, but it could well be larger than this and extend beyond the survey boundary.

The width of the survey in Declination is small and the quartet extends right across it: indeed Q2321-375 lies on the nominal boundary. The discussion might therefore need to be modified if other quasars with redshift ~ 0.37 were to be found near the quartet but just outside the survey boundary.

5 Discussion

5.1 SUBJECTIVE *VERSUS* OBJECTIVE METHODS

The clustering of the quasars in the CTIO catalogue has been investigated by Arp (1980), using a method which is completely subjective. His results and ours do not agree satisfactorily, and here we discuss two major areas of difference.

5.1.1 Clouds of quasars 5° wide

Arp notes that the region of the survey between about 01^h and $01^h\,30^m$ shows a higher surface density of quasars than neighbouring regions, and interprets this in terms of a cloud of quasars thrown out by the low-redshift galaxy NGC 300. We note, however, that the width of the cloud is about 5°, the same as the diameter of the survey plates, and that the cloud is in fact coextensive with plate number 16467, centred on $01^h\,14^m\,14^s$. This is moreover the very plate which Dr M. Smith informed us showed images of particularly fine quality so it is not surprising that the quasar count is high, for purely observational reasons. There is therefore no evidence of a real quasar cloud in this region and no need to suppose that NGC 300 has thrown out any quasars. Similar remarks apply to the concentration of quasars which Arp notes just to the east of NGC 55: it is again coextensive with one of the better plates of the survey (16466, centred on $00^h\,22^m\,07^s$) so again there are grounds for doubting that NGC 55 threw out these quasars.

5.1.2 The high-redshift quasars

Arp also claims that the quasars with $z > 2.5$ show a distribution on the sky whose 'clumpiness is immediately apparent' and that on looking at the redshifts of the quasars in each group 'we find further evidence that the group cannot be accidental'. The membership of the groups is stated to be of order 5, and the average widths in RA and redshift of the specific examples given are 6°.3 and 0.22 respectively.

Our PSA Of the same redshift range (D) has by contrast found no evidence of clustering: the α–δ, α–z and 3-D results are all consistent with the null hypothesis. It was decided to investigate this discrepancy in more detail by repeating the α–z PSA with two changes which would make the results more directly comparable with Arp's. First, the whole RA coverage was analysed and not just our restricted range. Second, in order to prevent the intrusion of any prejudice that the redshift may indicate distance, the ratio of the scales on the redshift and RA axes was not set by adopting a cosmological model. Instead it was set by choosing the value for which the clusters described by Arp are on the average circular rather than elliptical and would therefore show up most clearly: $dz/d\alpha = 0.22/6°.3$. The results are shown as the filled circles in Fig. 8(a) but are little different from before. They agree well with the null hypothesis and reveal no hint of the signature of clustering.

Could it be that the eye can, and PSA cannot, distinguish between a set of unclustered points and a set clustered in the manner suggested? Whilst the opinion has been advanced that PSA is an efficient statistical method (Webster 1976a), there is no theoretical proof of its power and it is necessary to check this possibility empirically. A Monte-Carlo method was

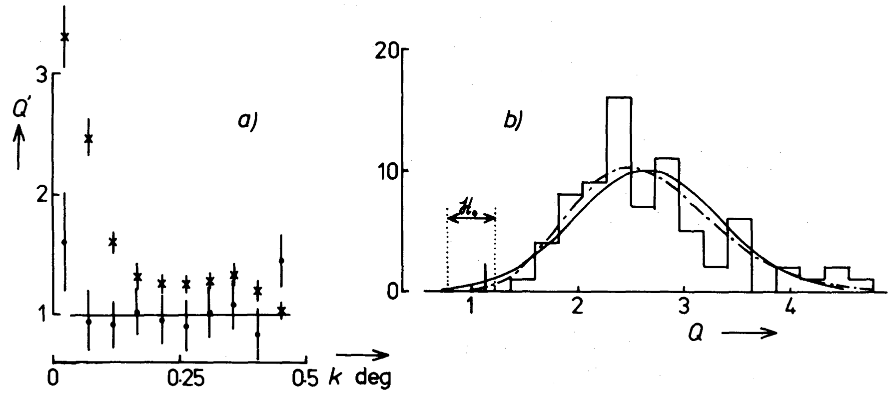

Figure 8. (a) The results of a two-dimensional PSA in RA and redshift for the CTIO quasars (solid points, error bars $\pm 1\sigma_0$) and for a Monte-Carlo simulation of a clustered population (crosses). The crosses give the mean of 75 independent simulations and their error bars are $\pm 1\sigma$ in the mean, as determined by the dispersion among the simulations. (b) the value of Q ($k = 0.095$ degree^{-1}) for the quasar sample (indicated by the short vertical line at $Q = 1.2$) and for the 75 Monte-Carlo simulations of a clustered population (indicated by the histogram). The vertical dotted lines are at $\pm 1\sigma_0$ about the value of $Q = 1$ expected on the null hypothesis. The solid curve is the Gaussian distribution which best fits the histogram and the broken curve similarly the gamma distribution.

devised in which artificial constellations of clustered quasars were generated in the computer with the properties outlined by Arp and were then subjected to PSA in exactly the same way as the real quasars. In this method a cluster centre was chosen at random by drawing the RA from a uniform pseudo-random number distribution whose width was rather larger than that of the survey: this extra width is to allow for the probability that some of the quasars lying within the survey might be members of clusters whose centres lie without. The redshift was similarly drawn. The position of the first quasar relative to this centre was then established by drawing its RA offset from a normal distribution with mean zero and width (=2 standard deviations) of 6°.3 and its redshift offset from a similar normal distribution of width 0.22. This was repeated for the other four members of the cluster, then a new cluster centre was drawn as before and the process repeated until the number of quasars lying within the survey area was the same as for the real quasars, $m = 17$. This artificial constellation of quasars was subjected to PSA in exactly the same way as for the real quasars, and the whole Monte-Carlo procedure was repeated 75 times; the average of the results is shown in Fig. 8(a) (crosses). The signature of clustering is seen very clearly, and it is inferred that the PSA method usually has little difficulty in distinguishing $\mathscr{H}_0$ from the type of clustering claimed by Arp. Furthermore there is evidently nothing marginal or doubtful about the result for the actual quasars: the points lie close to what is expected on $\mathscr{H}_0$ but nowhere near the Monte-Carlo results.

In order to assess how infrequently the Monte-Carlo process generates a constellation whose PSA resembles that of the quasar sample, a particular PSA statistic Q was chosen (Webster 1976a). Q is defined as a multiple of the sum of a number of terms $|r_{uv}|^2$ over a wavenumber range $0 < k \leqslant k_{\max}$, where we are free to set $k_{\max}$ as we wish. We chose to set it at the value of k (0.095 degree^{-1}) for which $(\hat{Q}(k) - 1)/\hat{\sigma}(Q(k))$ is a maximum, where $\hat{Q}(k)$ is the average value of Q from the 75 Monte-Carlo runs and $\hat{\sigma}$ the standard deviation of the 75 sample values of $Q(k)$. This is the value of k at which the Monte-Carlo results differ from 1, the expectation on $\mathscr{H}_0$, by the greatest number of standard deviations and is thus the value of k at which the PSA discriminates best between the clustered and unclustered processes. Choosing $k_{\max}$ from the Monte-Carlo results also has the benefit that it is objective and is not based on the results in hand from the actual quasar sample. Fig. 8(b) shows the statistic $Q(k_{\max})$, singly for the sample of actual quasars and as a histogram for the 75 Monte-Carlo results. The $\pm 1\sigma_0$ error range for $\mathscr{H}_0$ is marked, and it is seen that the two processes give well-separated values of Q: the histogram for the clustered population lies well to the right of the $\pm 1\sigma_0$ range for $\mathscr{H}_0$, so Q discriminates well between the processes. Moreover, the value for the actual quasar sample lies within the range for $\mathscr{H}_0$ but outside the range spanned by the histogram. It is directly concluded that the chance of the coordinates of the quasars being a realization of a clustering process of the kind described by Arp is no greater than about one in 75. A better figure may be obtained by fitting a theoretical distribution to the histogram and integrating to answer the question: what is p, the probability of the clustering process yielding a value of Q as small as the value for the quasars or smaller? The form of the exact theoretical distribution is unknown so two approximations have been tried: a normal and a gamma distribution, both fitted by choosing their parameters to give the same mean and standard deviation as the histogram. The true distribution is expected to be closer to the gamma than to the normal, but to have a greater positive skewness than the gamma by virtue not only of the simulation having clusters with a range of effective numbers of members (due to the survey boundary dividing some clusters), but also of the lack of total statistical independence amongst the $|r_{uv}|^2$ (Webster 1976a). For the normal and the gamma distribution $p = 0.009$ and 0.002 respectively. In both cases the histogram is more skew than the theoretical curves, by 3.2 and 2.0 times the standard error in the third moment expected for a sample of 75 independent drawings from the normal and gamma distributions respectively. In principle this procedure could be refined by finding a theoretical distribution whose third moment could be fitted exactly to the histogram in addition to the first and second, but in practice there is no very convenient distribution available, and in any case we have the result we need. As the fit of the theoretical curve is improved by increasing its skewness in going from the normal to the gamma, the area p under the left-hand tail of the distribution decreases. This trend is expected to continue to any better approximation than the gamma, so the result for the gamma gives us a firm upper limit to the probability whose value we seek: $p \leqslant 0.002$.

PSA, it is concluded, discriminates efficiently and reliably between $\mathscr{H}_0$ and the clustering model. Furthermore, the distribution in RA and redshift of the high-redshift quasars is easily consistent with $\mathscr{H}_0$ but is most unlikely to have been generated by a clustering process of the kind based on Arp's description. Our results and Arp's therefore appear to be contradictory, but we have been unable to find any error in the statistics or the computer programming.

It may be that the reason for the difference is to be found in Arp's method. Assessing clustering by eye is likely to be difficult, for two principal reasons. First, the eye has a strong tendency to invent structure where none is present, *vide* Schiaparelli's Martian canals. Second, the uncontrolled application of any kind of subjective method provides too little protection against subconscious bias, particularly when the subject has a strong interest in one type of outcome. These difficulties are well known and numerous types of control have been devised to improve the reliability of subjective studies. In the present context, for example, computer-generated random constellations of quasars could be compared with the real constellation to give some reference. The control would be improved if the subject were unaware which is which, artificial or real sky. Better still would be to employ a naive subject who knows neither what the hypotheses are, nor whether any of them is favoured by an assessment that a particular cloud of points does or does not show clustering. In the absence of any mention to the contrary (Arp 1980) we infer that none of these elementary precautions was taken when assessing the CTIO catalogue for clustering, and that no control of any other type was exercised either. We may therefore attribute the disagreement with the PSA results to the unreliability of the uncontrolled subjective method.

5.2 OTHER STUDIES

While this paper was in preparation, two preprints were received containing other types of statistical study of the catalogue. The first is that of Osmer (1981) who has applied three tests to the distribution of the quasars in three-dimensional space: binning, nearest-neighbour, and correlation analysis. By and large his conclusions agree well with ours: evidence of sensitivity variations is found but no evidence of clustering. The only discrepancy is that no notice is taken of the $z = 0.37$ group, which stands out rather clearly in our PSA. Osmer notes that his tests are not all as powerful as PSA, which may account for this in part, but principally the difference is due to our considering separate redshift ranges while he does not and the signal from the one solitary group is swamped in the noise of the many unclustered quasars at higher redshift.

The second preprint is by Oort *et al.* (1981) who conduct a search for close pairs in this and other quasar samples, and estimate their likelihood from the mean density of quasars in the catalogue. The $z = 0.37$ group is found in this way, in agreement with our result. However, one 'probable' and five 'possible' pairs with separations ~ 100 Mpc are found in (C), where we found no evidence of close pairs. One possible contribution to this difference is that no account was taken of the variations of survey sensitivity when assessing the pairs, and the 'probable' pair does lie on a good plate with its centroid only ~ 1° away from the plate centre. The concentration of detected quasars towards the centre of the plates increases substantially the probability of finding a random pair there over what is calculated on the basis of the average quasar density, so this pair may not be quite so interesting after all. As for the 'possibles', we may take the 3-D PSA for (C) (Fig. 7) to set an upper limit to f_2, the fraction of quasars in (C) which are members of non-random pairs with separations $\lesssim 100$ Mpc. Excluding terms with $k < 2$ Gpc, which are affected by the sensitivity variations, $Q(10\,\mathrm{Gpc}^{-1}) \pm \sigma_0(Q) = 0.959 \pm 0.018$, so a $3\sigma_0$ upper limit is $f_2 \leqslant 0.13$, which may be compared with the expected 0.19 if all the pairs listed by Oort *et al.* were to be real. The two sets of results are therefore not in good agreement if all the 'possibles' are counted, but if several of them are discounted there is no great discrepancy.

Another interesting comparison is with the central idea of Oort *et al.* that quasars are located in galaxy superclusters. Their Poisson model of the occurrence of quasars in superclusters may be used to predict the value of q (the average number of quasars per group) in redshift range (C) for comparison with our findings. From Oort's (1981) parameters converted to $H_0 = 75\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, we calculate a typical supercluster density of order $10^3\,\mathrm{Gpc}^{-3}$, which gives a mean of $\mu = 0.037$ known quasars per supercluster in range (C). On the Poisson model this gives $q = \mu/(1 - \exp(-\mu)) = 1.019$ for the expected value of $Q(10\,\mathrm{Gpc}^{-1})$, which differs from unity by about one standard deviation of Q. The detection of quasar superclusters in (C) is therefore expected to be marginal and our lack of evidence for clustering may not be taken as evidence against the supercluster hypothesis.

5.3 FUTURE POSSIBILITIES

The future prospects for the systematic study of quasar clustering seem excellent. The following results suggest that a survey with a moderate increase in sensitivity over the CTIO survey may reveal clusters at high redshifts:

(1) The apparent existence of the group at $z = 0.37$.

(2) The supercluster model of Oort *et al.* (1981) which predicts clustering at high redshifts which is marginal in the CTIO survey but which should come to light as the sensitivity is improved.

(3) The existence of known close quasar pairs (found by a miscellany of methods which effectively probe deeper than the CTIO Schmidt) which *a posteriori* seem too frequent to be chance occurrences (e.g. Oort *et al.* 1981; Hazard, Arp & Morton 1979; Burbidge *et al.* 1980; Margon, Chanan & Downes 1981).

Systematic objective prism surveys with much better sensitivity are already in progress which yield a quasar density ~ 10 times that of the CTIO survey, and with this increase in signal it would be surprising if distant clusters were not to be found.

The importance of gathering a large, uniform and deep quasar sample and analysing it properly is hard to overstate:

(1) The distribution of matter on large scales is worth knowing *per se*.

(2) The development of structure with cosmic epoch may be revealed by looking at different redshift intervals. This might yield important clues to the origin of the structure.

(3) The rate of development of the density contrast depends on q_0, amongst other factors, and may yield an estimate of this important cosmological parameter. Its value also enters into the purely geometrical considerations of the relative angular sizes of similar clusters at different redshift, so two independent estimates may be possible.

Available methods for q_0 are generally plagued by two great difficulties: first, the possible cosmological evolution of the property under investigation (such as the candle power of a standard class of galaxy) and, second, selection effects such as the biassed detection of the brighter galaxies of any particular class with increasing distance. By looking at quasar clusters it may in principle be possible to sidestep or reduce these difficulties. The evolution of the density contrast hardly constitutes a difficulty if the evolution is the very property under study; moreover it may be well understood anyway, if it behaves as one of the simple types of linear perturbation in a Friedman cosmology. The selection effects would also be unimportant if all quasars, whatever their luminosity, trace out in the same way the variations in the same underlying universal matter density.

Acknowledgments

It is a pleasure to thank Dr M. G. Smith for helpful advice and information, and Dr M. Perryman for several interesting discussions.

References

Arp, H., 1980. *Astrophys. J.,* **239,** 463.

Arp, H. & Hazard, C., 1980. *Astrophys. J.,* **240,** 726.

Burbidge, E. M., Junkkarinen, V. T., Koski, A. T., Smith, H. E. & Hoag, A. A., 1980. *Astrophys. J.,* **242,** L55.

Fanti, C., Lari, C. & Olori, M. C., 1978. *Astr. Astrophys.,* **67,** 175.

Hazard, C., Arp, H. C. & Morton, D. C., 1979. *Nature,* **282,** 271.

Lindgren, B. W., 1976. *Statistical Theory,* Collier MacMillan, London.

Margon, B., Chanan, G. A. & Downes, R. A., 1981. *Nature,* **290,** 480.

Masson, C. R., 1979. *Mon. Not. R. astr. Soc.,* **188,** 261.

Oort, J. H., 1981. *Astr. Astrophys.,* **94,** 359.

Oort, J. H., Arp, H. & de Ruiter, H., 1981. *Astr. Astrophys.,* **95,** 7.

Osmer, P. S., 1981. *Astrophys. J.,* **247,** 762.

Osmer, P. S. & Smith, M. G., 1980. *Astrophys. J. Suppl. Ser.,* **42,** 333.

Pearson, T. J., 1974. *Mon. Not. R. astr. Soc.,* **166,** 249.

Seldner, M. & Peebles, P. J. E., 1978. *Astrophys. J.,* **225,** 7.

Webster, A. S., 1976a. *Mon. Not. R. astr. Soc.,* **175,** 61.

Webster, A. S., 1976b. *Mon. Not. R. astr. Soc.,* **175,** 71.

Webster, A. S., 1977. *Mon. Not. R. astr. Soc.,* **179,** 511.