Mon. Not. R. astr. Soc. (1989) **238,** 319–348

The clustering of local maxima in random noise

P. Coles *Astronomy Centre, University of Sussex, Brighton BN1 9QH*

Accepted 1988 November 21. Received 1988 November 21; in original form 1988 October 18

Summary. A mixture of analytic and numerical techniques is used to study the clustering properties of local maxima of random noise. Technical complexities restrict us to the case of 1D noise, but the results obtained should give a reasonably accurate picture of the behaviour of cosmological density peaks in noise defined on a 3D domain. We give estimates of $\xi_{\mathrm{pk-pk}}(r)$, the two-point correlation function of local maxima, for both Gaussian and non-Gaussian noise and show that previous approximations are not accurate. Furthermore, we show that the strong dependence of $\xi_{\mathrm{pk-pk}}(r)$ on the shape of the underlying correlation function, $\xi(r)$, ensures that no simple approximations to $\xi_{\mathrm{pk-pk}}(r)$ are obtainable for general $\xi(r)$. We find that zero-crossings of $\xi_{\mathrm{pk-pk}}(r)$ do not, in general, coincide with those of $\xi(r)$. This poses a problem for the CDM model, in that the cluster–cluster correlation function is clearly positive at distances where we expect it to be negative if clusters are identified with peaks of a Gaussian random field. Using a log-normal field to model the density distribution obtained after non-linear evolution from Gaussian initial conditions, we find that a moderate amount of non-linear evolution, as expected on cluster scales, does not have a drastic effect on the bias achieved. We also study the distribution of nearest-neighbour distances for local maxima and find that, for high maxima, this distribution is very flat, leading to a scaling of the mean nearest-neighbour distance with sample size, similar to that observed by Einasto & Einasto.

1 Introduction

It has been known for some time that the two-point correlations of rich clusters are much larger than those of galaxies. The two-point correlation function for bright galaxies has the form

$$\xi_{\mathrm{gg}}(r) \simeq \left(\frac{r}{r_{0,g}}\right)^{-1.8},$$

* As usual, $h = H_0/100$ km s^{-1} Mpc^{-1}.

where $r_{0,g} \simeq 5\,h^{-1}$ Mpc (Groth & Peebles 1977; Davis & Peebles 1983). The first attempts to determine the two-point correlation function for rich clusters gave the same shape but a different amplitude:

$$\xi_{cc}(r) \simeq \left(\frac{r}{r_{0,c}}\right)^{-1.8}, \tag{1}$$

where $r_{0,c} \simeq 25\,h^{-1}$ Mpc (Bahcall & Soneira 1983; Klypin & Kopylov 1983; Postman, Geller & Huchra 1986). This result is less well established than $\xi_{gg}(r)$ for two reasons. Firstly, the Bahcall & Soneira (1983) sample is very small (only 104 clusters) and the statistical uncertainties are correspondingly large; a realistic analysis of the errors in determinations of $\xi_{cc}(r)$ has shown that a 95 per cent confidence interval for the value of $r_{0,c}$ is (12.8, 39.2) h^{-1} Mpc (Ling, Frenk & Barrow 1986). Secondly, there are considerable doubts as to the reliability of Abell's catalogue (Abell 1958); the visual method used to obtain the richness class of clusters can lead to large systematic errors (Lucey 1983). The second of these points has recently received much attention. Bahcall & Soneira (1983) and Bahcall, Soneira & Burgett (1986) noted that $\xi_{cc}(r)$ seems to be very elongated in the line-of-sight direction. Their explanation for this was the (inferred) existence of large cluster peculiar velocities (~ 2000 km s^{-1}). A careful analysis of a larger 3D cluster catalogue (Struble & Rood 1987) has revealed clear evidence for projection effects in the Abell catalogue (Sutherland 1988). When allowance is made for these effects, the 'best' cluster–cluster correlation function is

$$\xi_{cc}(r) \simeq \left(\frac{r}{14\,h^{-1}\,\mathrm{Mpc}}\right)^{-1.8}, \tag{2}$$

and there is no evidence for large cluster velocities or for significantly non-zero correlations at r greater than $\sim 30\,h^{-1}$ Mpc. Other workers have claimed that these effects are small in the sample of $R \geq 1$, $D \leq 4$ clusters that is the source of most current estimates of $\xi_{cc}(r)$ and, therefore, stick to the form (1) (Dekel, Blumenthal & Primack 1988).

This enhancement of $\xi_{cc}(r)$ compared to $\xi_{gg}(r)$, together with the observed tendency of the correlation length to increase with cluster richness (Bahcall & Soneira 1983; Bahcall *et al.* 1986), led Kaiser (1984) to suggest that rich clusters might be *biased* tracers of the mass distribution on cluster scales. In Kaiser's model the enhanced clustering of clusters is primarily a statistical effect that occurs because clusters form only at the high peaks of a smoothed random field of density perturbations. This idea was later discussed by Peacock & Heavens (1985) and Bardeen *et al.* (1986), who suggested its relevance to theories of biased galaxy formation (see also Davis *et al.* 1985; Couchman 1987a, b).

Unfortunately, calculations of the correlation properties of peaks (i.e. local maxima) of the density are rather difficult to perform rigorously. To avoid messy mathematics, Kaiser (1984) calculated the correlation function of points lying above a threshold rather than local maxima, an approach also used by Politzer & Wise (1984) and Jensen & Szalay (1986). One can think of this approximation as being equivalent to taking the correlation function of clusters of galaxies to be equal to the correlation function of those galaxies which lie in clusters. There are two problems with this approach. Firstly, as pointed out by Coles (1986), this approximation systematically overestimates $\xi_{cc}(r)$ for distances up to $r \simeq 20\,h^{-1}$ Mpc (see also Peacock & Heavens 1985; Bardeen *et al.* 1986). Secondly, all calculations based on the high-level region approximation predict that $\xi_{cc}(r) = 0$ whenever the underlying matter correlation function is zero. Monte-Carlo simulations of 3D noise by Otto, Politzer & Wise (1986a) reveal that maxima do not possess this property. Unfortunately, their analytic expressions for the correla-

tions of peaks are incorrect (Otto, Politzer & Wise 1986b) and they give no explanation for this effect. This point is particularly relevant to the cold-dark-matter (CDM) model, in which context biasing is usually discussed; the CDM correlation function goes to zero at $r \sim 18\,h^{-2}$ Mpc if $\Omega = 1$ (Otto *et al.* 1986a) and both forms of $\xi_{cc}(r)$, (1) and (2) are clearly non-zero at this distance.

These two problems motivate a study of the correlations of local maxima of random noise. Bardeen *et al.* (1986) obtained asymptotic expressions for the correlation function of peaks of Gaussian random noise, $\xi_{pk\text{-}pk}(r)$, as $r \to \infty$, but these cannot be accurate for all possible correlation functions. Further progress with peaks of 3D fields is horrendously difficult – see Cline *et al.* (1987). All is not lost, however. The reason for studying peaks rather than high regions is that peaks define a spatial point process. It is likely therefore that we could get at least a good qualitative understanding of the behaviour of peaks of random noise in 3D by studying 1D noise. In fact, Appendix A shows that this approximate method should be a quantitative, as well as qualitative, improvement on previous estimates.

As mentioned above, the form of $\xi_{pk\text{-}pk}(r)$ is important for theories of biased galaxy formation as well as for theories of the origin of rich clusters. In hierarchical models, clusters are relatively late-forming and therefore undergo little non-linear evolution. Studies of biased galaxy formation, however, must take into account the substantial amount of non-linear evolution which must have occurred on galaxy scales. Since the only analytic studies of biasing that are possible are predominantly linear, it would be unwise to use these analytic methods to study the galaxy distribution. Studies of bias in the galaxy distribution can only be performed using *N*-body experiments (Davis *et al.* 1985; Bardeen, Bond & Efstathiou 1987; Batuski, Melott & Burns 1987; White *et al.* 1987a, b). Furthermore, in the CDM model both galaxies and clusters are biased tracers of the mass distribution and one would have to handle biasing on two different scales simultaneously in order to study the model in detail. To avoid unnecessary complications, therefore, we shall look only at rich clusters for a concrete cosmological application for our results and we will not worry about how individual galaxies fit into the picture. Of course, *some* non-linear evolution must have occurred on cluster scales – we discuss the possible effects of this later in the paper.

The layout of this paper is as follows: in Section 2 we give some technical background concerning Gaussian random fields and previous estimates of $\xi_{cc}(r)$; in Section 3 we study $\xi_{pk\text{-}pk}(r)$ for 1D Gaussian noise; in Section 4 we perform a similar analysis of non-Gaussian noise in an attempt to allow for non-linear effects; in Section 5 we look at nearest-neighbour distributions for peaks as an independent test of the biasing hypothesis. Finally, in Section 6, we present and discuss the conclusions.

2 Technical background

2.1 GAUSSIAN RANDOM FIELDS

The density fields usually considered are Gaussian random fields, as predicted by most inflationary models. A Gaussian random field can be represented as a superposition of sinusoidal components with random phases:

$$\delta(\mathbf{x}) = \sum_k c(k) \cos[\mathbf{k}\cdot\mathbf{x} + \phi(\mathbf{k})], \tag{3}$$

where

$$\delta(\mathbf{x}) = [\rho(\mathbf{x}) - \langle\rho\rangle]/\langle\rho\rangle,$$

where $k = |\mathbf{k}|$, the $\phi(\mathbf{k})$ are independent and uniformly random on $[0, 2\pi]$ and $P(k)$, the power-spectrum of the random field, is proportional to $\langle |c(k)|^2 \rangle$.

An important consequence of the form (3) is that by the central limit theorem δ is normally distributed. Furthermore, the joint distributions of the values of δ at different points are multivariate Gaussians (as are the joint pdf's of δ and its spatial derivatives). This allows substantial progress to be made in determining the statistical properties of such fields.

The Fourier transform of $P(k)$ is called the correlation function:

$$\xi(r) \propto \int_0^\infty P(k)\, k^2 \left[\frac{\sin(kr)}{kr}\right] dk.$$

We shall be dealing mainly with 1D random noise in this paper. It is therefore convenient to define a 1D power-spectrum $Q(k)$

$$Q(k) = \frac{2}{\pi} \int_0^\infty \delta(r) \cos(kr)\, dr. \tag{4}$$

The moments of this power-spectrum are defined thus:

$$\sigma_i^2 = \int_0^\infty Q(k)\, k^{2i} dk. \tag{5}$$

Note that $\sigma_0^2 = \sigma^2 = \xi(0)$ is the *variance* of the field and $\sigma_1^2 = -\xi''(0)$ is the mean-square spatial derivative of the field. It is possible to define a coherence length, r_c, as follows:

$$r_c = \sqrt{\frac{\xi(0)}{-\xi''(0)}}. \tag{6}$$

It is worth remarking that it is necessary to smooth the random field with some sort of filter (usually a Gaussian) to select structures on the scale of interest (in this case, a cluster scale). The smoothing radius must be chosen to produce the correct number-density of objects (Coles 1986).

There are many sources for further technical details concerning Gaussian random fields (Rice 1945, reprinted in Wax 1954; Adler 1981; Vanmarcke 1983; Peacock & Heavens 1985; Bardeen *et al.* 1986;Couchman 1987a, b).

2.2 THE HIGH-LEVEL REGION APPROXIMATION

Kaiser (1984) sought to explain the enhanced correlations of Abell clusters as a primarily statistical effect that does not depend on the existence of substantial spectral power on cluster scales. He argued that, if clusters form at the high peaks above a sharp threshold $u\sigma$ of an initially Gaussian density field (smoothed on an appropriate scale: Section 2.1), then they exhibit enhanced two-point correlations over those of the underlying matter distribution.

Here we outline the statistical details. We calculate the two-point correlation function for *regions* of a Gaussian random field lying above the level $u\sigma$ where σ^2 is the variance of the field (which is assumed to have mean value zero). In dealing with finite regions exceeding a given threshold instead of the true peaks (i.e. local maxima) of the field one can avoid the complication of having to consider derivatives of the field. Let the field possess a covariance function $\xi(r)$. In the simplest version of this model the galaxies trace the mass so that $\xi(r) = \xi_{gg}(r)$. It is possible to calculate the correlations of such regions as follows:

if we choose a point at random, the probability that the field at that point takes a value exceeding the level $u\sigma$ is just

$$P_1 = \int_{u\sigma}^{\infty} P(y)\,dy, \tag{7}$$

where, for a Gaussian field,

$$P(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-y^2}{2\sigma^2}\right). \tag{8}$$

If we choose another point at a distance r from the first, the probability that both points lie above the threshold is

$$P_2 = \int_{u\sigma}^{\infty}\int_{u\sigma}^{\infty} P(y_1, y_2)\,dy_1\,dy_2. \tag{9}$$

From the Gaussian nature of the field it follows that (Section 2.1)

$$P(y_1, y_2) = \frac{1}{2\pi}\frac{1}{\sqrt{\xi^2(0) - \xi^2(r)}} \exp\left\{-\frac{\xi(0)y_1^2 + \xi(0)y_2^2 - 2\xi(r)y_1y_2}{2[\xi^2(0) - \xi^2(r)]}\right\}. \tag{10}$$

Hence, the two-point correlation function for the regions, $\xi_u(r)$, is

$$1 + \xi_u(r) = \frac{P_2}{P_1^2}. \tag{11}$$

The exact evaluation of the integrals involved in (11) is difficult. The original approximation by Kaiser (1984) for large u and small $\xi_u(r)$ was

$$\xi_u(r) \simeq \frac{u^2}{\sigma^2}\,\xi(r). \tag{12}$$

Judicious choice of u can therefore explain the enhanced two-point correlations of rich clusters (1) or (2). The model can also explain, at least qualitatively, the trend of increasing r_0 with cluster richness, if we identify richer and richer clusters with higher and higher thresholds.

Other workers have examined this model in even more detail. Politzer & Wise (1984) obtained the following expression [where $\xi_u(r)$ is not necessarily small but $u \gg 1$]

$$1 + \xi_u(r) \simeq \exp[u^2\xi(r)/\sigma^2]. \tag{13}$$

They also computed higher-order correlations in this approximation. Jensen & Szalay (1986) found an expression that gives $\xi_u(r)$ to arbitrary accuracy★ for any u:

$$\xi_u(r) \simeq \sum_{m=1}^{\infty} \frac{[w(r)]^m}{m!} A_m^2, \tag{14}$$

★This result appears to be a rediscovery of Pearson's (1901) *tetrachoric* expansion of the bivariate normal integral. It has been known by statisticians for many years that it converges so slowly to the result as to be almost useless (Gupta 1963).

where

$$A_m = \frac{2xH_{m-1}(x)2^{-m/2}}{\sqrt{\pi x \exp(x^2)\,\mathrm{erfc}(x)}}, \tag{15}$$

where $x = u/\sqrt{2}$; $w(r) = \xi(r)/\xi(0)$; $H_n(x)$ is a Hermite polynomial (Abramowitz & Stegun 1965).

Note that all these forms for the correlation function of regions above the level $u\sigma$ predict that $\xi_u(r) = 0$ whenever $\xi(r) = 0$ so that rich clusters are anticorrelated whenever the mass is anticorrelated.

3 Peaks in 1D Gaussian noise

3.1 Analytic studies of local maxima in 1D noise

The calculation of $\xi_{\mathrm{pk\text{-}pk}}(r)$ is rather similar to that of $\xi_u(r)$ (Section 2.2) except that we need to compute probabilities that points are local maxima above the level $u\sigma$. We denote the two relevant probabilities by P_1^* and P_2^* to replace the P_1 and P_2 in (11).

First consider P_1^*, the probability that a point chosen at random is a local maximum of the random field above the level $u\sigma$. This problem has been studied previously by Cartwright & Longuet-Higgins (1956). Without loss of generality, consider a Gaussian random process which has zero mean and unit variance. Denote this process $\xi(x)$.

If $\varepsilon(x)$ has a local maximum in the interval $(x, x+dx)$ then in this interval $\varepsilon'(x)$ must take a value in a range of width $\simeq |\varepsilon''(x)|\,dx$ about zero. The probability of this occurrence, and of $\varepsilon(x)$ simultaneously lying in the range $(\varepsilon, \varepsilon + d\varepsilon)$ is

$$\left[\int_{-\infty}^{0} P(\varepsilon, 0, \varepsilon'')|\varepsilon''|\,d\varepsilon''\right] dx,$$

where $P(\varepsilon, \varepsilon', \varepsilon'')$ is the joint pdf of $[\varepsilon(x), \varepsilon'(x), \varepsilon''(x)]$. For a Gaussian process this will be a multivariate Gaussian. The covariance matrix is

$$\mathbf{M}_{11} = \begin{pmatrix} 1 & 0 & -\sigma_1^2 \\ 0 & \sigma_1^2 & 0 \\ -\sigma_1^2 & 0 & \sigma_2^2 \end{pmatrix}, \tag{16}$$

which has determinant $\Delta = \sigma_1^2\sigma_2^2 - \sigma_1^6$ and the spectral moments, σ_i^2, are defined by (5). Now if $\boldsymbol{\varepsilon} = (\varepsilon, 0, \varepsilon')$ then

$$P(\boldsymbol{\varepsilon}) = \frac{1}{(2\pi)^{3/2}\Delta^{1/2}} \exp\left(-\frac{1}{2}\,\boldsymbol{\varepsilon}\cdot\mathbf{M}_{11}^{-1}\cdot\boldsymbol{\varepsilon}\right). \tag{17}$$

The probability per unit length that the point is a maximum above the level u is therefore

$$P_1^* = \int_u^{\infty}\int_{-\infty}^{0} P(\boldsymbol{\varepsilon})|\varepsilon''|\,d\varepsilon''\,d\varepsilon. \tag{18}$$

This expression can be integrated analytically in terms of special functions. First define $u_* = \Delta/\sigma_1^2\sigma_2^2$ and $\tilde{u} = u_*/\sqrt{1-u_*^2}$. Using the functions

$$P(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp(-t^2/2)\,dt \tag{19}$$

and

$$Q(x)=\frac{1}{\sqrt{2\pi}}\int_{x}^{\infty}\exp(-t^2/2)\,dt, \tag{20}$$

we find that

$$P_1^*=(\sqrt{1-m_*^2})\exp(-u^2/2)\,P\left(\frac{u}{\tilde{u}}\right)+Q\left(\frac{u}{u_*}\right) \tag{21}$$

(see also Cartwright & Longuet-Higgins 1956).

We now need P_2^*, the probability that two points separated by a distance r are both maxima above the level u. Assuming that the random process is stationary, this is equivalent to the probability that a point at $x=0$ and a point at $x=r$ are both local maxima above u. We need the joint pdf of $\boldsymbol{\varepsilon}=[\varepsilon(0),\varepsilon'(0),\varepsilon''(0),\varepsilon(r),\varepsilon'(r),\varepsilon''(r)]$. Again, the Gaussian nature of the process will ensure that this is a multivariate normal distribution. The 6×6 covariance matrix $\mathbf{M}$ is composed as follows:

$$\mathbf{M}=\begin{pmatrix}\mathbf{M}_{11} & \mathbf{M}_{12}\\ \mathbf{M}_{21} & \mathbf{M}_{22}\end{pmatrix}, \tag{22}$$

where $\mathbf{M}_{11}=\mathbf{M}_{22}$, which contains the correlations of field values and derivatives at one point, is given by (16). The off-diagonal components are rather more difficult to calculate. The general method is given by Rice (1945) (see Wax 1954). We illustrate it with a couple of elements. First consider $\langle\varepsilon(0)\,\varepsilon'(r)\rangle$. This is defined to be

$$\begin{aligned}\langle\varepsilon(0)\,\varepsilon'(r)\rangle &\equiv \lim_{T\to\infty}\frac{1}{T}\int_0^T\varepsilon'(x+r)\,\varepsilon(x)\,dx.\\ &=\xi'(r).\end{aligned} \tag{23}$$

Similarly we find that

$$\begin{aligned}\langle\varepsilon(r)\,\varepsilon'(0)\rangle &\equiv \lim_{T\to\infty}\frac{1}{T}\int_0^T\varepsilon'(x)\,\varepsilon(x+r)\,dx\\ &=\lim_{T\to\infty}-\frac{1}{T}\int_0^T\varepsilon'(x+r)\,\varepsilon(x)\,dx\\ &=-\xi'(r).\end{aligned} \tag{24}$$

Each component of $\mathbf{M}_{12}$ can be calculated in this way. The result is that

$$\mathbf{M}_{12}=\begin{pmatrix}\xi(r) & \xi'(r) & \xi''(r)\\ -\xi'(r) & -\xi''(r) & -\xi'''(r)\\ \xi''(r) & \xi'''(r) & \xi''''(r)\end{pmatrix}. \tag{25}$$

By symmetry, it is clear that $\mathbf{M}_{21}=\mathbf{M}_{12}^{T}$. Note also that if $r\to 0$, $\xi(r)\to 1$, $\xi'(r)\to\xi'''(r)\to 0$, $\xi''(r)\to-\sigma_1^2$ and $\xi''''(r)\to\sigma_2^2$ so that $\mathbf{M}_{12}\to\mathbf{M}_{11}$. The probability (per unit length^2) that two points separated by a distance r are both maxima above the level u is

$$P_2^*=\frac{1}{(2\pi)^3\|\mathbf{M}\|^{1/2}}\int_u^\infty\int_u^\infty\int_{-\infty}^0\int_{-\infty}^0|\varepsilon_3|\,|\varepsilon_6|\exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}\cdot\mathbf{M}^{-1}\cdot\boldsymbol{\varepsilon}\right)d\varepsilon_3 d\varepsilon_6 d\varepsilon_1 d\varepsilon_4, \tag{26}$$

where $\boldsymbol{\varepsilon}=\{\varepsilon_i\}=[\varepsilon(0),0,\varepsilon''(0),\varepsilon(r),0,\varepsilon''(r)]$ and $\|\mathbf{M}\|$ is the determinant of matrix (22) which will be a function of r.

The task now is to evaluate the integral (26) and hence evaluate $\xi_{\text{pk-pk}}$, using the fact that

$$1+\xi_{\text{pk-pk}}(r)=\frac{P_2^*}{(P_1^*)^2} \tag{27}$$

similarly to (11). It is difficult to perform such integrals analytically and, in any case, the result would be a series expansion similar to the form (15) which converges very slowly, especially for the large values of $\xi(r)$ that occur near $r=0$ (see Gupta 1963). Such a series expansion would tell us nothing that cannot be obtained from a direct numerical integration of (26). Unfortunately also, a numerical integration of (26) is rather CPU intensive because of the large number of integrand evaluations required when r is near zero.★ When r is close to zero, all the off-diagonal terms in the matrix $\mathbf{M}$ are comparable and the matrix tends to become singular and is consequently difficult to invert. Rather than use numerical integration to obtain the behaviour of $\xi_{\text{pk-pk}}(r)$ at all separations r, we therefore use a quicker method (Section 3.2) to determine a rough estimate of $\xi_{\text{pk-pk}}(r)$ at all separations and use the numerical integration to check the significance of the results [especially with regard to zero crossings of $\xi_{\text{pk-pk}}(r)$]. Integrations were carried out using standard NAG routines. The great advantage of using Monte-Carlo methods here is that we can extract very much more information from each simulation than just $\xi_{\text{pk-pk}}(r)$ (see Section 5).

However, before going on to show how to obtain rough estimates of $\xi_{\text{pk-pk}}$ (which are, in fact, rather accurate), it is worth pointing out a few details of the behaviour of $\xi_{\text{pk-pk}}$ that can be deduced just from the form of the integral (26). Firstly, two essentially separate effects give rise to the difference between $\xi_{\text{pk-pk}}$ and ξ. One is the 'thresholding' effect which occurs in the same way as detailed in Section 2.2 and the other is the fact that spatial derivatives of the process are themselves correlated which causes an enhancement in the probability of getting a local maximum near to a given local maximum. Secondly, it is relatively straightforward to expand the quadratic form in (26) to lowest order in r to see what the joint probability looks like at small distances. When this is done we find that $P_2^*\to\exp(-C/r^2)$ as $r\to 0$ (C is a constant) so that $\xi_{\text{pk-pk}}(r=0)=-1$, as expected. Thirdly, note that the integral depends on the derivatives up to fourth order of ξ and is therefore very dependent not just on the value of ξ but also on the detailed shape of $\xi(r)$. When r is close to zero, all the derivative terms in $\mathbf{M}$ may be the same order of magnitude. There is no hope of obtaining a simple fit of a single function of $\xi(r)$ to the behaviour at these distances. This has been born out by 2D numerical studies of the clustering of peaks (L. Appel, private communication) where the form of $\xi_{\text{pk-pk}}$ is found to have a very complicated dependence on the underlying power spectrum. Furthermore, note that to compute $\xi_{\text{pk-pk}}$ for cosmologically interesting fluctuation spectra, we would need to know all derivatives of $\xi(r)$ up to and including fourth order to the same accuracy. We do not even know

★ See Lumsden, Heavens & Peacock (1989) – they reduce the CPU time required by performing one of the four integrals in (26) analytically.

the form of $\xi(r)$ analytically in the CDM model and therefore the task of evaluating the other terms is made even more difficult. Couple this with the problem that we cannot really allow for non-linear effects and the problem seems rather intractable.

We can get a good qualitative understanding of what is going on, however, by using a couple of *model* functions for $\xi(r)$. Firstly, note that Kaiser (1984) chose a simple model which has $\xi(r) \propto 1/r^2$ at large r. This cannot be the form of $\xi(r)$ at all separations because of the divergent variance $\xi(0)$ but if we choose the form

$$\xi(r) = \left(1 + \frac{r^2}{2r_c^2}\right)^{-1}, \tag{28}$$

we should get a good idea of how closely $\xi_{\text{pk-pk}}$ resembles the predictions of Section 2.2.

We are also interested in zero-crossings of $\xi(r)$ and $\xi_{\text{pk-pk}}(r)$ and the above form of $\xi(r)$ never crosses zero. Bearing in mind that we want to compute derivatives of $\xi(r)$ as simply as possible, a good model to use is a Bessel function (Abramowitz & Stegun 1965) which has the advantage that one can calculate the derivatives from simple recursion relations (Appendix B).

$$\xi(r) = J_0(\lambda r). \tag{29}$$

3.2 MONTE-CARLO SIMULATIONS OF 1D NOISE

Although one can evaluate (26) numerically to reasonable accuracy using standard techniques, the CPU time taken to do this is rather large. It is much easier to evaluate an estimate of $\xi_{\text{pk-pk}}(r)$ using Monte-Carlo simulations and use the numerical integration to check the results.

A simple method exists for generating simulations of stationary 1D Gaussian noise with a specified covariance function $\xi(r)$. Note that the 1D power spectrum of such noise is just the Fourier cosine transform of the covariance function (4). A stationary Gaussian random process can therefore be constructed as a sum of waves with random phases, using

$$\varepsilon(x) = \sum_k c_k \cos(kx + \phi_k). \tag{30}$$

Such a construction is easily performed by generating amplitudes from a Gaussian distribution with variance* $\propto Q(k)$ and with phases drawn from a uniform distribution on $(0, 2\pi)$ using standard numerical techniques. A Fast Fourier Transform (FFT) of this random phase realization then produces $\hat{\varepsilon}(x_i)$, a discrete-step realization of $\varepsilon(x)$ with the desired covariance function.

We have performed simulations of Gaussian processes with $\xi(r)$ given by both (28) and (29). For the first form the required power spectrum is

$$Q(k) \sim \int_0^\infty \left(1 + \frac{x^2}{2r_c^2}\right)^{-1} \cos(kx)\, dx$$

$$\sim \exp(-kr_c\sqrt{2}). \tag{31}$$

The coherence length of this process is just r_c.

* The constant of proportionality just alters the variances of ε which we scale out of the problem by considering $u\sigma$ peaks. Without loss of generality, we can assume that $\sigma^2(\varepsilon) = 1$.

For the form (29) we get

$$Q(k) \sim \int_0^\infty J_0(\lambda x)\cos(kx)\,dx$$

$$\sim \begin{cases} (\lambda^2 - k^2)^{-1/2}; & k < \lambda \\ 0 \quad\quad\quad\quad ; & k \geq \lambda \end{cases} \tag{32}$$

The coherence length of this process is $\sqrt{2}/\lambda$. Both of these transformations can be found in Gradshteyn & Ryshik (1965).

For each of the realizations $\hat{\varepsilon}(x_i)$, we also evaluate $\hat{\xi}(r)$, the covariance function of the simulation to check for goodness-of-fit to the required form (28) or (29). Relatively few waves (~ 50) are needed for the form of $\hat{\xi}(r)$ to match the required form with acceptable accuracy and also for the distribution of $\hat{\varepsilon}(x_i)$ to be acceptably close to a Gaussian.

As mentioned above, this technique generates a discrete realization of $\varepsilon(x)$. All the simulations consist of 65 536 points. For each $\hat{\varepsilon}(x_i)$ we locate all maxima above various levels $u\sigma$ [i.e. those points x_i such that $\hat{\varepsilon}(x_{i-1}) < \hat{\varepsilon}(x_i) > \hat{\varepsilon}(x_{i+1})$ and $\hat{\varepsilon}(x_i) > u\sigma$]. The two-point correlation function of peaks at a separation r_i is defined in terms of the excess number of pairs of peaks (above that which would be expected in a random distribution) at a separation r_i. We calculate this by binning the number of pairs of peaks with a given separation to obtain $N_{\rm pk}(r_i)$. Then we use

$$1 + \hat{\xi}_{\rm pk\text{-}pk}(r_i) = \frac{N_{\rm pk\text{-}pk}(r_i)}{N_{\rm ra}(r_i)}, \tag{33}$$

where $N_{\rm ra}(r_i)$ is the number of pairs of peaks with separation r_i we would expect if the maxima were scattered at random along the length of the simulation. Clearly uncertainties in $N_{\rm ra}(r_i)$ can have a big effect on $\hat{\xi}_{\rm pk\text{-}pk}(r_i)$ so, to evaluate this accurately and also to remove any possible systematic effects due to the peak-finding algorithm, for each simulation we count the number of peaks above $u\sigma$ and generate a random distribution of that number of peaks. A large number of simulations (~ 150) of each process are generated and the resultant pair distributions $N_{\rm pk\text{-}pk}$ and $N_{\rm ra}$ combined to give one aggregate distribution. This method should give a good estimate of the errors in the final estimate of $\xi_{\rm pk\text{-}pk}(r_i)$.

Note that the prescription above for generating $\hat{\varepsilon}(x_i)$ produces a periodic realization of the process – the two ends of the simulation are identified. This allows the peak finding algorithm to 'wrap around' the edges of the simulations and therefore dispense with the need to consider edge effects which often occur in correlation analyses (Peebles 1980).

The final forms of $\xi_{\rm pk\text{-}pk}(r_i)$ thus obtained are shown in Fig. 1 for $\xi(r)$ given by (28) and Fig. 2 for $\xi(r)$ given by (29). Distances are given in computer units (cu) where 1 cu is the distance step $x_j - x_{j-1}$. The coherence lengths chosen for the two figures were 5 cu for the first and 7.07 cu for the second. The first graph (a) of each pair shows $\hat{\xi}_{\rm pk\text{-}pk}(r_i)$ for all maxima (i.e. the threshold $u \to -\infty$) and the second has $u = 2$. The wiggliness of the graphs should give a good indication of the likely statistical error in the estimates of $\xi_{\rm pk\text{-}pk}$. In addition to $\hat{\xi}_{\rm pk\text{-}pk}(r_i)$ the second graph of each pair shows the Politzer–Wise approximation (13).

Let us consider Fig. 1. We can see the 'repulsive core' predicted in Coles (1986) quite clearly for small separations in both (a) and (b). Note that in (a) the maxima are basically uncorrelated after we emerge from this core. In (b) however the threshold causes an amplification which the Politzer–Wise form clearly overestimates at small distances (due to the tail of the region-size

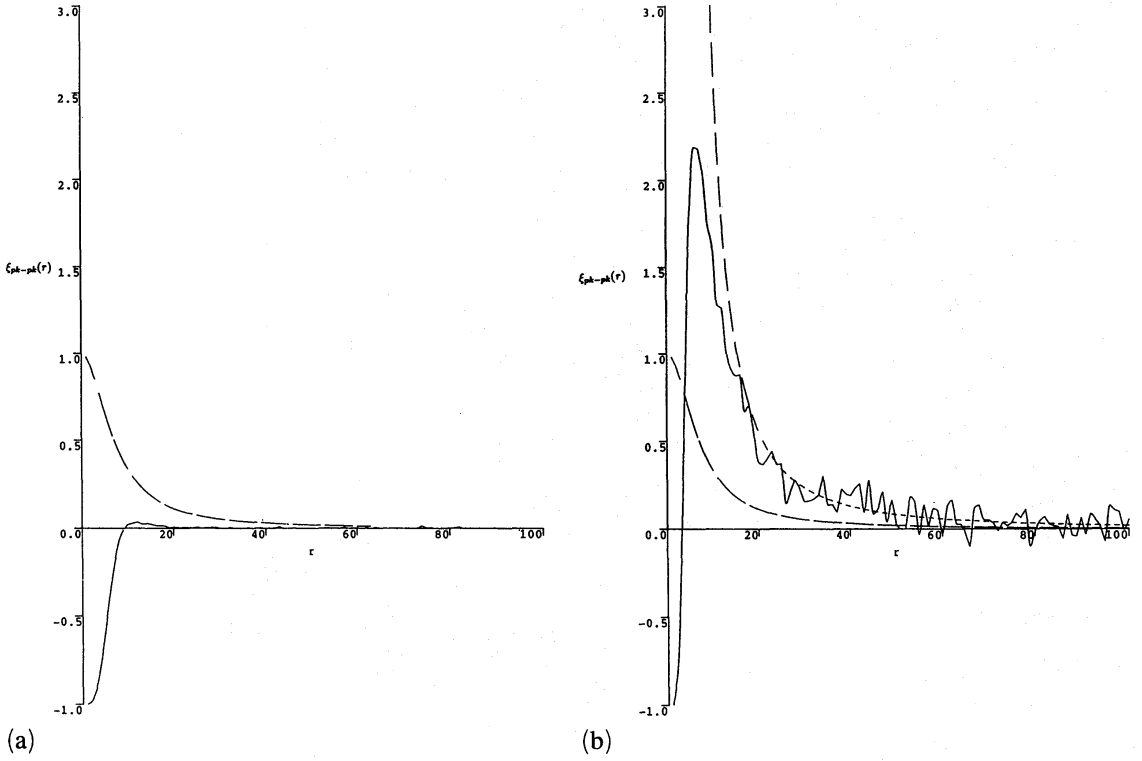

Figure 1. Estimates of the peak–peak correlation function for (a) all maxima and (b) maxima above 2σ for Gaussian noise with covariance function (28). The dashed line shows the underlying $\xi(r)$ and the dashed line [in (b) only] shows the Politzer–Wise approximation, $\xi_u(r)$.

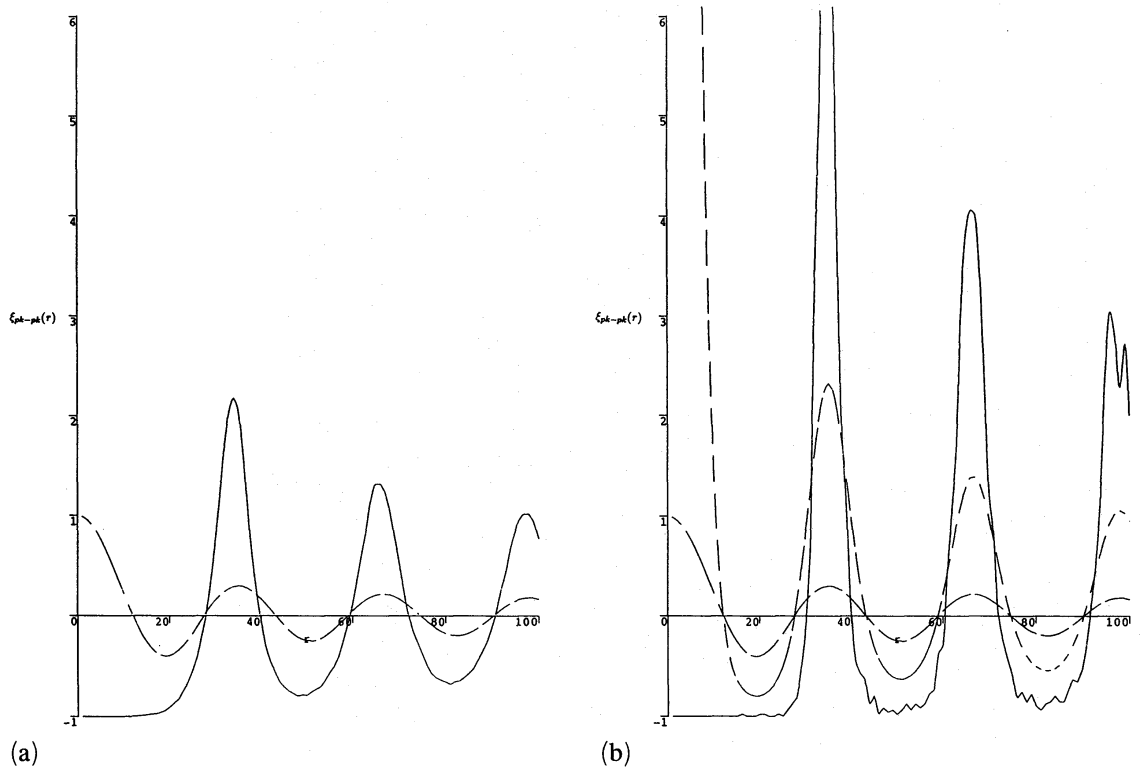

Figure 2. Estimates of the peak–peak correlation function for (a) all maxima and (b) maxima above 2σ for Gaussian noise with covariance function (29). The dashed line shows the underlying $\xi(r)$ and the dashed line [in (b) only] shows the Politzer–Wise approximation, $\xi_u(r)$.

distribution). At large distances, the Politzer–Wise form seems quite accurate. This shows that Kaiser's thresholding amplification is a real effect for maxima and that the whole phenomenon cannot be due to the overcounting effect described in Coles (1986). Results for different values of u show that higher thresholds produce a higher amplification.

Now consider Fig. 2. Again, the repulsive core is evident here but it seems to be much larger than the previous case. It is clear from (a) that, in this case, there are substantial enhancements of $\xi_{\rm pk\text{-}pk}$ over ξ even without a threshold. This effect is explained below. Note also that the zero-crossings of $\hat{\xi}_{\rm pk\text{-}pk}$ are *not* the same as those of ξ although the effect appears quite small even for this strongly oscillatory covariance function. The significance of this difference has been established by accurate numerical integration of (26) in the vicinity of the zero crossing predicted by the simulations and is quite clearly a genuine artifact that cannot be explained by statistical errors in the simulations. This confirms that the Otto *et al.* (1986a) effect is real. Note also that, although some enhancement is obtained without thresholding, inclusion of a threshold also causes an increase in the amplification. The Politzer–Wise form does not seem to be a very good fit at any distance in this model.

3.3 DISCUSSION

All of these results can be explained quite simply when one recalls the argument given above that the difference between $\xi_{\rm pk\text{-}pk}(r)$ and $\xi(r)$ is due to two essentially separate effects, the thresholding effect (which is accounted for in the Kaiser model: Section 2.2) and an additional effect which is due to the correlations of the derivatives of the process. The thresholding effect relies upon the underlying $\xi(r)$ being non-zero (Section 2.2) whereas the derivative correlations depend on derivatives of $\xi(r)$ and are therefore not necessarily zero when ξ itself is zero.

Now consider the results for ξ given by (28). Because $\xi(r)$ is falling monotonically, the spatial derivatives of $\xi(r)$ are also falling and they fall more quickly than ξ itself. In (a), when we emerge from the repulsive core all of the derivative correlations have essentially disappeared. There is, therefore, no contribution to $\xi_{\rm pk\text{-}pk}$ from such correlations. But there is no thresholding amplification either so the net $\xi_{\rm pk\text{-}pk}(r)$ is close to zero. When the threshold is applied (b), an enhancement is achieved. This is reasonably well fitted by the Politzer–Wise formula because only the thresholding is playing a role in the enhancement.

On the other hand (29) is qualitatively different. When we emerge from the repulsive core in this case, the derivative correlations are of the same order as ξ itself (because ξ oscillates in r). These correlations can therefore produce an enhancement even when no threshold is applied. When the threshold is applied, the amplification is increased by the same mechanism as above. Note also that when a zero-crossing of ξ occurs, the derivative correlations are not necessarily zero and so can produce an enhancement of $\xi_{\rm pk\text{-}pk}(r)$ even when the thresholding enhancement is zero (as it must be whenever ξ itself is zero). This therefore explains the Otto *et al.* (1986a) result and also the results above (Section 3.2).

Incidentally, some authors (e.g. Peacock & Heavens 1985) have suggested that by considering all maxima (without the additional thresholding effect) one might achieve substantial biasing. The reason behind this is that $p_{\rm max}(\delta)$, the distribution of maxima of the overdensity field, can be very different from $p(\delta)$, which is the underlying Gaussian distribution of unconditioned field values. Clearly, if the mass correlations have the form (28) then this idea cannot work as no enhancement of peak–peak correlations is observed. In fact, the peaks are less clustered than the mass in this case. To make a viable biased model with (28) for the underlying correlation function, one clearly needs an additional threshold.

Note also that the 'repulsive core' seems much larger for the simulations in Fig. 2. We saw in Coles (1986) that the size of excursion regions was of order the coherence length which is slightly larger ($\sim$40 per cent) than in Fig. 1. This is not enough to explain the much larger

domain of anti-correlation, however. The real cause of this effect can be deduced by looking at the form of $\xi(r)$ and noting that it is negative at values of r greater than the coherence length. When we emerge from the repulsive core, $\xi_{\mathrm{pk-pk}}(r)$ is amplified by the factors discussed above but the underlying $\xi(r)$ is negative and so $\xi_{\mathrm{pk-pk}}(r)$ becomes even more negative. Clearly, $\xi_{\mathrm{pk-pk}}(r)$ cannot be < -1 so it 'saturates' until the underlying $\xi(r)$ starts to increase again. This effect makes the core look bigger than it actually is.

As a further point it is worth remarking that the 'repulsive core' which shows up so clearly in these simulations (and also in Peacock & Heavens 1985; Lumsden *et al.* 1989) seems to pose rather drastic problems for the Kaiser model as it is clearly not observed in the cluster–cluster correlation function, $\xi_{\mathrm{cc}}(r)$ (1). This is not really much of a problem, however, because we expect non-linear effects at small separations to produce rather many close pairs of clusters quite rapidly and thus eradicate this core. This occurs because the size of the core is roughly the size of a cluster before it collapses and after it has collapsed it is very much smaller. Nearby clusters can thus approach to within the original repulsive core.

The biasing model for rich clusters is usually discussed in the framework of the CDM cosmogony so it is interesting to make a few comments about this specific scenario. We have not used the correct $\xi(r)$ for CDM so anything we say will be approximate, but the results of this section do seem to pose serious problems for this model. The smoothed CDM covariance function looks like $\sim r^{-2}$ until we reach the scale set by the horizon size at matter–radiation equivalence $\sim 18\,h^{-2}$ Mpc (for $\Omega_0 = 1$), so we should get a reasonable fit to the behaviour of biased CDM if we take the form (28) for intermediate distances with a sudden switch to a 'steep' form like (29) at a distance of order the zero-crossing. The first obvious problem is with the zero-crossing of $\xi_{\mathrm{pk-pk}}(r)$ which moves inwards by $-5\,h^{-2}$ (estimated) Mpc compared to $\xi(r)$, our rough estimate agreeing with the result of Otto *et al.* (1986a). As discussed in Section 1, the cluster–cluster correlation function seems to be positive out to at least $30\,h^{-1}$ Mpc. The zero crossing therefore suggests that $h \sim 0.4$–0.5 or that $\Omega_0 < 1$ (or both). There is also a problem with the amplitude before we get to this point. The normalization of the CDM spectrum is uncertain: Bardeen *et al.* (1986) choose a value roughly twice that of White *et al.* (1987a, b). By scaling our numerical estimates according to these normalizations we find that neither normalization can reproduce the form (1), although the Bardeen *et al.* (1986) normalization can reproduce (2) (within the errors of our simulations). This is the preferred choice in any case because it is the only one that guarantees that one has the correct number density of clusters, although there may be better physical arguments for other normalizations, based on collapse time considerations. According to these approximate arguments it would seem that, for CDM to reproduce the observed cluster–cluster correlations, we require at least one (preferably more) of the following to be true:

(i) The Bardeen *et al.* (1986) normalization of the CDM spectrum is the correct one.
(ii) The Hubble parameter is rather small ($h \sim 0.4$–0.5).
(iii) $\Omega_0 < 1$.
(iv) The cluster–cluster correlation length is $\sim 14\,h^{-1}$ Mpc rather than $25\,h^{-1}$ Mpc.

Although these conclusions are based on approximate arguments, they are in agreement with those obtained by detailed 3D simulations of CDM fluctuations performed independently by Lumsden *et al.* (1989).

4 Clustering of maxima of non-Gaussian fields

4.1 INTRODUCTION

In the previous section, we assumed that clusters form at peaks of a Gaussian random field. If the initial perturbations were Gaussian, then this is equivalent to assuming that little non-linear

evolution has occurred on cluster scales. The standard practice is to assume that fluctuations can be described by the linear equations whenever the covariance function of the matter distribution $\xi(r) \leq 1$, or that the rms perturbation is less than, or of order, unity. However, it is clear that whenever $\xi(0) \sim 1$, some non-linear evolution must have occurred. This is obvious because the Gaussian distribution is defined on the interval $(-\infty, +\infty)$ and if the variance is of order the mean, a Gaussian distribution of overdensities would imply the existence of regions of negative density. The distribution must therefore develop a positive skew as the fluctuations become non-linear (Peebles 1980; Fry 1986; Grinstein & Wise 1986; Goroff *et al.* 1987). In biasing calculations we are interested in the extreme positive 'tail' of the overdensity distribution and it seems at first sight that a skewed distribution might manifest very different biased correlations. Unfortunately, it is very difficult to treat non-linear evolution of pertubations analytically but *N*-body experiments seem to suggest that a biased non-linear distribution seems to have roughly the same properties that one would expect from linear theory. These results seem rather counter-intuitive, given the discussion above, so it is important to attempt some analytical study to see if we can understand the results of the simulations. Fry's (1986) study of non-linear biasing indeed suggests that biasing is a 'robust' phenomenon (i.e. small perturbations away from the Gaussian have little effect on the qualitative behaviour of the biased distribution). In this section, we shall investigate this in a different way to see if Fry's conclusions stand up to more detailed study.

4.2 SECOND-ORDER PERTURBATION THEORY

One approach to the non-linear evolution of perturbations is via second order perturbation theory, as outlined in Peebles (1980). This approach is very approximate and it is important to acknowledge the limitations of the technique at the outset and admit that little quantitative accuracy is expected in the forthcoming calculations. The main problem with the perturbative approach is that, when evolution develops to the stage where second-order corrections are important, we also expect third-, fourth- and even higher-order terms to appear very quickly. The range of validity of a second-order expansion is therefore very small. Another problem is that the shape of the correlation function is not preserved in higher order perturbation theory, thus introducing further uncertainty into the calculations (Juszkiewicz, Sonoda & Barrow 1983). Here we shall ignore this and compare Gaussian and non-Gaussian distributions possessing the same covariance function.

Peebles (1980) discusses the second-order evolution of initially Gaussian perturbations. Let the overdensity field be denoted $\delta(\mathbf{x})$, where $\langle\delta\rangle = 0$. To second order he shows that

$$\langle\delta^3\rangle = \frac{34}{7}\,\xi(0)^2, \tag{34}$$

where $\xi(0) = \langle\delta^2\rangle$, the variance of the field and $\langle\delta^3\rangle$ would be zero for a Gaussian distribution (or any other symmetrical distribution) with zero mean. We can define a coefficient of skewness for a random variable x:

$$\gamma_1 \equiv \left(\frac{\mu_3^2}{\mu_2^3}\right)^{1/2} \tag{35}$$

(Kendall & Stuart 1977). The μ_n are the n-th order central moments. This coefficient measures the departure from symmetry: $\gamma_1 = 0$ for a distribution symmetrical about its mean. According to (35), the skewness of the second-order perturbations is

$$\gamma_1 = \frac{34}{7}\sqrt{\xi(0)}. \tag{36}$$

Now, as mentioned above, it is common practice to assume that the linear calculations are valid up until the time when $\xi(0) \sim 1$, or when by (36), $\gamma_1 \sim 5$. Unfortunately, we do not know what the full distribution of overdensities is at this stage but we do know that the distribution of densities must be bounded on the left at zero and it must be positively skewed with a coefficient of order 5. The approach we shall use here is to model the non-linear perturbations with a non-Gaussian distribution possessing these two properties in order to see what the effect is on the amount and nature of bias produced.

The log-normal is an obvious distribution to use as a model because of its simple relation to a Gaussian random field (Coles & Barrow 1987) and the fact that observations seem to imply that the large-scale galaxy distribution has a similar form (Hubble 1934 was the first to notice this). Using a distribution obtained via a transformation of a Gaussian seems to make sense, because high peaks in the underlying Gaussian field will then map to high peaks in the transformed field. Intuitively, we would expect clusters to form approximately where peaks are in the primordial field but the detailed statistics of the overdensity field must be perturbed away from the Gaussian by the time the clusters collapse. The log-normal random field should model this situation quite well. For the log-normal field, derived from a Gaussian with zero mean and unit variance, we find that $\mu_2 = e(e-1)$; $\mu_3 = e^{3/2}(e-1)^2(e+2)$ so that $\gamma_1 \simeq 6$. Bearing in mind all the uncertainties in this analysis, this is close enough to the value above for our use. One bad point about this distribution is that it is not uniquely determined by any finite number of moments (Kendall & Stuart 1977) so that we cannot claim that our results are in any sense unique. This entire analysis is not going to be quantitatively accurate in any case so we will not worry about these complications.

Another distribution we shall look at is χ^2_n. Unfortunately, the only distribution we can make any progress with is the simplest one which has $n=1$. This does, however, have the largest skewness of any χ^2_n distribution★ ($\gamma_1 \sim 3$) which is of the right order of magnitude. Note that this distribution does not have the attractive feature of the log-normal mentioned above. Both high and low regions of the Gaussian field x map to high regions of the field $y = x^2$. Another unrealistic feature of this model is that the pdf is infinite at $y=0$ and therefore the model contains a large number of 'void' regions. It seems more likely, bearing in mind Fry's (1986) analysis, that the pdf should $\to 0$ as $y \to 0$, meaning that there are no totally 'void' regions. It is clear that, although γ_1 is less for this model, the χ^2_1 case displays a much greater departure from Gaussian behaviour then the log-normal. For these reasons we shall not place too much emphasis on the results for this field, although if the assumption that cluster sites are near the primordial Gaussian matter peaks is not true, the results become more interesting.

We now consider biasing of these two non-Gaussian distributions, firstly in the Politzer & Wise (1984) high level region approximation (as discussed in Section 2.2) and secondly by deriving $\xi_{\mathrm{pk-pk}}(r)$ for the log-normal case as we did in Section 3 for the Gaussian.

4.3 THE POLITZER–WISE APPROXIMATION

4.3.1 Log-normal fluctuations

If $X(\mathbf{r})$ is a standard Gaussian random field (i.e. if it has zero mean and unit variance) then $Y(\mathbf{r}) = \exp[X(\mathbf{r})]$ is log-normally distributed (Coles & Barrow 1987). To use Y as a model for the matter distribution, we must first ensure that $Y(\mathbf{r})$ possesses the correct covariance

★The skewness of χ^2_n distributions is $2^{3/2}/\sqrt{n}$.

function. If $X(\mathbf{r})$ has a correlation function $w_x(r)$ (assumed isotropic) then the correlation function of Y is

$$w_y(r)=\frac{\exp[w_x(r)]}{e-1}, \tag{37}$$

so that in order to obtain a correlation function $w_y(r)$ we need

$$w_x(r)=\log[1+(e-1)\,w_y(r)]. \tag{38}$$

For simplicity, we shall use the form (28) for our matter correlation function, although we note that non-linear evolution will not preserve the shape of the correlation function. To compare the results with the Gaussian case we need to use

$$1+\xi_u(r)=\frac{Q_2}{Q_1^2} \tag{39}$$

(*cf.* equation 11). This involves the integral

$$Q_2=\int_u^\infty\int_u^\infty Q(y_1,y_2)\,dy_1\,dy_2, \tag{40}$$

where $Q(y_1,y_2)$ is the joint pdf of $y_1=y(\mathbf{r}_1)$ and $y_2=y(\mathbf{r}_2)$. This can be obtained from $P(x_1,x_2)$, the joint pdf of the underlying Gaussian variables $x_1=x(\mathbf{r}_1)$ and $x_2=x(\mathbf{r}_2)$ via

$$Q(y_1,y_2)=P[y_1(x_1),y_2(x_2)]\,\|J\|, \tag{41}$$

where $\|J\|$ is the Jacobian of the transformation $(x_1,x_2)\mapsto(y_1,y_2)$. The distribution $P(x_1,x_2)$ is just a bivariate Gaussian (10). The integral for Q_2 can be transformed back into an integral over x_1 and x_2 by putting $y_i=\log(x_i)$; $(i=1,2)$, the only differences are that the limits of integration change and that w_x must be given by (38) and is not equal to the correlation function of the mass. The resulting integral is

$$Q_2=\int_{\log u}^\infty\int_{\log u}^\infty P(x_1,x_2;w_x)\,dx_1\,dx_2. \tag{42}$$

A similar manipulation gives $Q_1=\int_{\log u}^\infty P(x)\,dx$. Now the expression can be approximated by the original Politzer & Wise (1984) expression (13):

$$\xi_u(r)\simeq\exp[(\log u)^2 w_x(r)]-1. \tag{43}$$

Now we need to choose the level u in an appropriate way. It makes sense to compare the Gaussian case which has a $\nu\sigma$ threshold level with a similarly defined threshold for the non-Gaussian model. So we choose

$$u=\mu_y+\nu\sigma_y, \tag{44}$$

where μ_y and σ_y^2 are the mean and variance, respectively, of the non-Gaussian distribution. For the log-normal we have $\mu_y=\sqrt{e}$ and $\sigma_y=\sqrt{e(e-1)}$ so that the resulting Politzer–Wise expression is

$$\xi_y(r)\simeq\exp\{\log[\sqrt{e}+\nu\sqrt{e(e-1)}]^2\log[1+(e-1)\,w_y(r)]\}-1, \tag{45}$$

where we have substituted the correct form for the correlation function of the non-Gaussian field.

4.3.2 χ^2 of order unity

To check that the results for the log-normal case are reasonably consistent with those obtained from another non-Gaussian distribution, we use the field

$$Y(\mathbf{r}) = X^2(\mathbf{r}). \tag{46}$$

This produces a field which has the χ^2 distribution of order unity. Again, the analysis is considerably simplified by the fact that Y is a function of a single Gaussian field. In this case we have $\mu_y = 1$ and $\sigma_y^2 = 2$ so we define

$$u = 1 + \nu\sqrt{2}. \tag{47}$$

We also need the fact that $w_y(r) = w_x^2(r)$ which defines the correlation function of Y in terms of that of X. The situation in this case is rather more complicated than (4.3.1) because both high and low regions of X map to high regions of Y. Since Y is above the level u if X is either below $-\sqrt{u}$ or above $+\sqrt{u}$, the joint pdf Q_2 must be constructed as follows:

$$Q_2 = A + B + C,$$

where

$$A = \int_{\sqrt{u}}^{\infty}\int_{\sqrt{u}}^{\infty} P(x_1, x_2; w_x)\, dx_1\, dx_2$$

$$B = \int_{-\infty}^{-\sqrt{u}}\int_{-\infty}^{-\sqrt{u}} P(x_1, x_2; w_x)\, dx_1\, dx_2$$

$$C = 2\int_{-\infty}^{-\sqrt{u}}\int_{\sqrt{u}}^{\infty} P(x_1, x_2; w_x)\, dx_1\, dx_2\,. \tag{48}$$

The symmetry of the Gaussian about zero means that the first two terms above are equal, $A = B$, but the final term is different. Now the only term in $P(x_1, x_2; w_x)$ that depends on the *sign* of x is the term involving $w_x x_1 x_2$ so that if we make the substitution $x_2 \mapsto -x_2$ we can change the sign of w_x and the integral becomes

$$C = 2\int_{\sqrt{u}}^{\infty}\int_{\sqrt{u}}^{\infty} P(x_1, x_2; -w_x)\, dx_1\, dx_2\,,$$

using this and the fact that $Q_1 = 2\int_{\sqrt{u}}^{\infty} P(x)\, dx$ we find that the Politzer–Wise approximation (13) for the expression for Q_2/Q_1^2 is

$$1 + \xi_\nu(r) = \tfrac{1}{2}\{\exp[(1+\nu\sqrt{2})\, w_x(r)] + \exp[-(1+\nu\sqrt{2})\, w_x(r)]\} = \cosh[(1+\nu\sqrt{2})\sqrt{w_y(r)}], \tag{49}$$

where we have defined ν by (47) and incorporated $w_x = \sqrt{w_y}$. Notice how the second part of the expression (49) is small for large ν and large w_x (i.e. small r). This is what we would expect, because if the field X is above a high threshold at some point it becomes extremely unlikely that, at a point less than a coherence length away, the field will be below a very low (negative) threshold.

The behaviour of these two functional forms is shown in Fig. 3 for the cases $\nu = 2,3$. The Gaussian case is plotted for comparison. The forms are clearly rather similar although both

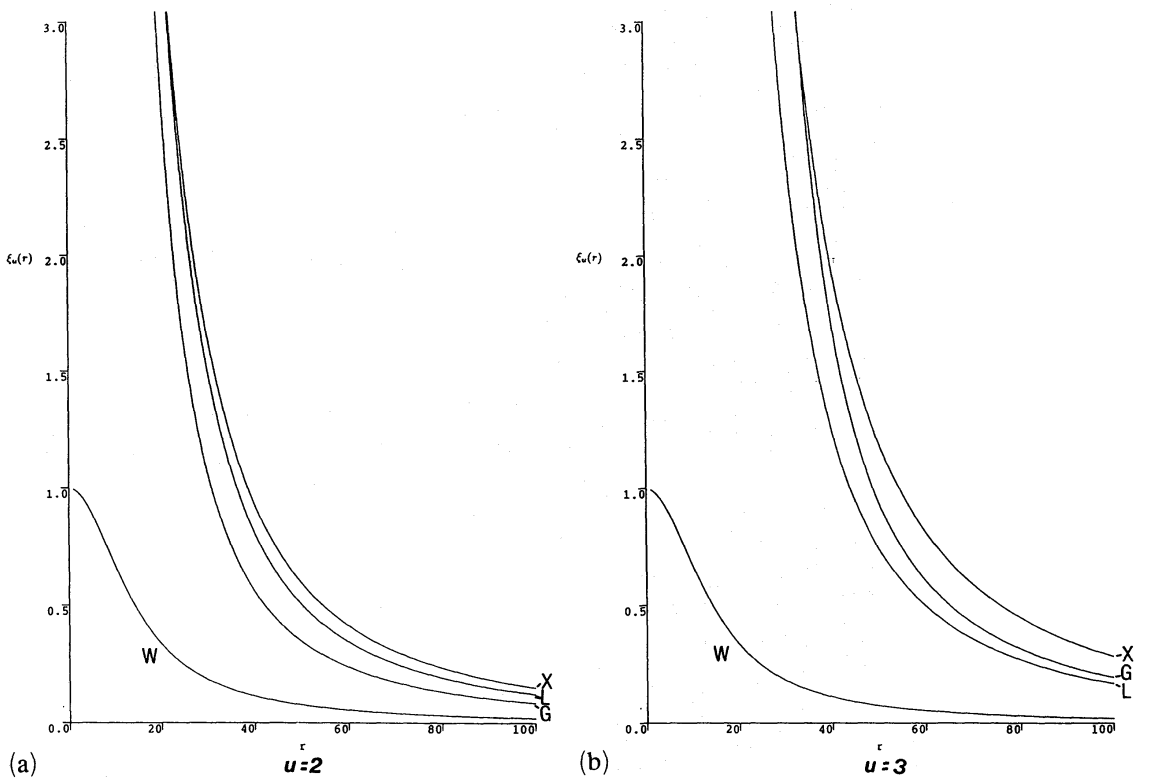

Figure 3. Behaviour of the Politzer–Wise approximation, $\xi_u(r)$, for log-normal (L), χ_1^2 (X) and Gaussian (G) random fields. The underlying correlation function (W) is shown for comparison. Plots are shown for $u=2$ and $u=3$.

non-Gaussian cases show a larger $\xi_\nu(r)$ than the Gaussian case. Remember, however, that Section 3 showed how we can only trust the Politzer–Wise approximation at very large distances so we must also look at the true peak–peak correlations to see whether this effect is real or merely an artifact of the systematic errors inherent in the Politzer–Wise approach. Note finally that both of these cases will have $\xi_\nu(r)=0$ whenever $\xi(r)=0$. The correlation function chosen for the graphs has the form (28).

4.4 PEAK–PEAK CORRELATIONS FOR THE LOG-NORMAL MODEL

We saw in Section 3 that it is rather difficult to deal with the peak–peak correlations analytically, even for Gaussian models (which are the simplest conceivable case). We had to resort to approximate 1D methods to investigate the qualitative behaviour analytically. In the Gaussian case the integrals involved had to be evaluated numerically. The non-Gaussian calculations are even more cumbersome but not qualitatively different so most of the details are omitted from this section. The easiest way to study the behaviour of peaks on log-normal noise is via Monte-Carlo simulations so, as we did in the Gaussian case, we use the direct evaluation of the integrals merely to check the consistency of the simulations. Throughout this section we shall work in 1D, for the same reason as we gave in Section 3.

The analytic expression for $\xi_{\mathrm{pk-pk}}(r)$ is extremely messy, so here we just outline the derivation. The integrals we require can be obtained by transformations of equations (18) and (26) to similar integrals over different variables, using the fact that the log-normal process is generated by an underlying Gaussian process. If the Gaussian process is $\varepsilon(x)$ then our log-normal process is $y(x)=\exp[\varepsilon(x)]$. The basic function we need is the joint pdf of the variables

$[y(0), y'(0), y''(0), y(r), y'(r), y''(r)]$. This set of variables is denoted $\mathbf{y} = \{y_i\}$. If the pdf of these six variables is $U_6(\mathbf{y})$ then

$$U_6(\mathbf{y}) = P(\boldsymbol{\varepsilon}) \| J \|, \tag{50}$$

where $P(\boldsymbol{\varepsilon})$ is a six-variate Gaussian, as described in Section 3. The quantity $\| J \|$ is the Jacobian of the transformation of variables

$$\begin{aligned} y(x) &= \exp[\varepsilon(x)] \\ y'(x) &= \varepsilon'(x) \exp[\varepsilon(x)] \\ y''(x) &= \varepsilon''(x) \exp[\varepsilon(x)] + [\varepsilon'(x)]^2 \exp[\varepsilon(x)] \end{aligned} \tag{51}$$

(and a similar set of three equations relates $y(0)$ to $\varepsilon(0)$, etc.). Note also that the covariance matrix involved in $P(\boldsymbol{\varepsilon})$ must be constructed using the covariance function of the underlying Gaussian process, $\xi^{(\varepsilon)}$. We shall use the form (28) henceforth and this means that

$$\xi^{(\varepsilon)}(r) = \log[1 + (e-1)(1 + r^2/2r_c^2)] \tag{52}$$

if $y(x)$ is to possess the correct covariance function. The resulting expression for U_6 is horrendous so we shall omit the details. The probability of obtaining two maxima above threshold separated by a distance r, denoted U_2^* is given by an integral similar to (26):

$$U_2^* = \int_u^\infty \int_u^\infty \int_{-\infty}^0 \int_{-\infty}^0 U_6(\mathbf{y}) |y_3| \| y_6 | dy_3 dy_6 dy_1 dy_4 \tag{53}$$

and we specify that $y_2 = y_5 = 0$ to select points with zero gradient and thus pick out maxima. This expression must be evaluated numerically. It is straightforward to evaluate U_1^* as an integral similar to (18) which can be evaluated numerically without much difficulty. The peak–peak correlation function is then given by

$$1 + \xi_{\text{pk-pk}}(r) = \frac{U_2^*}{(U_2^*)^2}. \tag{54}$$

As we did in the Gaussian case, we use numerical evaluations of these integrals to check the accuracy of the results obtained from Monte-Carlo simulations, which give a reasonably precise picture of the behaviour at a small cost in terms of CPU. The procedure used is to generate a Gaussian process as described in Section 3 which possesses the covariance function (28). This requires a power spectrum.

$$\begin{aligned} Q(k) &\sim \int_0^\infty \{\log[1 + (e-1)(1 + x^2/2r_c^2)^{-1}]\} \cos(kx)\, dx \\ &\sim \frac{1}{k} [\exp(-kr_c\sqrt{2}) - \exp(-kr_c\sqrt{e+1})]. \end{aligned} \tag{55}$$

As in Section 3, a random phase realization of this power-spectrum is generated and an FFT used to produce a discrete realization of the Gaussian process, $\hat{\varepsilon}(x_i)$. Each element of this process is then mapped using $\hat{y}(x_i) = \exp[\hat{\varepsilon}(x_i)]$ to form a log-normal process possessing the required covariance function (28). The form of $\xi_{\text{pk-pk}}(r)$ is extracted from these simulations in the same way as in the Gaussian case (Section 3). The results obtained from an ensemble of

such simulations are shown in Fig. 4. Results are given for $\nu = 2$ and $\nu = 3$, using (44) to define the threshold levels.

4.5 DISCUSSION AND CONCLUSIONS

We can see immediately that the form of Fig. 4 is very similar to its Gaussian counterpart, Fig. 1. This confirms in a qualitative way that Fry's (1986) conclusion is correct. Note that if we choose $\nu = 2$ for the log-normal model, the Politzer–Wise approximation is larger than the corresponding Gaussian case. However, the peak–peak covariance function is virtually identical in the two cases (it has slightly lower amplitude at smaller distances and slightly greater amplitude at large distances). This confirms that there should be very little difference in the form of $\xi_{\mathrm{pk-pk}}(r)$ when non-linear effects are taken into account. This also confirms the comments made on this subject by Bardeen *et al.* (1986) and Lumsden *et al.* (1989).

One might object to our choice of threshold for the non-Gaussian case. The choice we have made seems to be the most logical bearing in mind Kaiser's original reasoning. An alternative choice would be to pick ν in such a way that the same *fraction* of field points lie above the threshold for both the Gaussian and non-Gaussian models. The threshold required for the log-normal model to obtain this sort of equivalence with a threshold of $\nu = 2$ for the Gaussian case would be $\nu \simeq 2.7$. A glance at Fig. 4(b) shows, that even in the case when $\nu = 3$, the difference between the Gaussian and log-normal models is small. The value of $\xi_{\mathrm{pk-pk}}$ for the log-normal model is greater than that in the Gaussian case by only 10–20 per cent.

Note that the Politzer–Wise expression is again not well obeyed by the true peak–peak correlations, although it does become accurate for very large r. In fact, the discrepancy

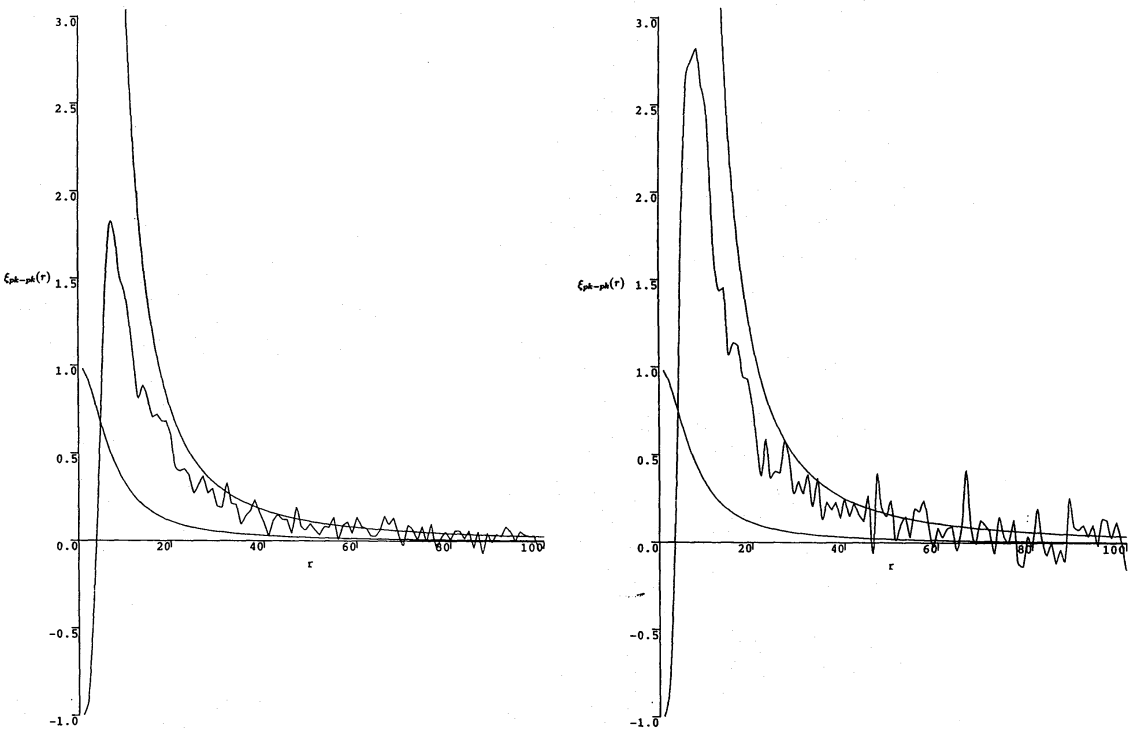

Figure 4. Estimates of the peak–peak correlation function for peaks above (a) $\nu = 2$ (b) $\nu = 3$ for log-normal noise with covariance function (28). The underlying $\xi(r)$ and the Politzer–Wise approximation, $\xi_\nu(r)$ are also shown.

between the high-level region approximation and the true peak–peak correlation is greater for the non-Gaussian models than for the Gaussian case. This suggests that one should treat with great caution attempts to study highly non-Gaussian fields using the Politzer–Wise approach (e.g. Matarrese, Lucchin & Bonometto 1986).

Our basic conclusion, therefore, is that if we use a plausible non-Gaussian model to mimic the degree of non-linear evolution we expect the matter distribution to have undergone on cluster scales, the resulting peak–peak correlations are roughly the same as one would predict using strictly linear theory. If one chooses a model which departs drastically from Gaussian statistics (such as χ^2_1: Section 4.3.2), however, one can obtain rather greater peak–peak correlations. Indeed it is important to point out that the non-Gaussian cases studied here by no means demonstrate the full variety of biasing behaviour one could expect to find in arbitrarily non-Gaussian fields. One could easily imagine highly non-Gaussian fields which exhibit markedly different biased statistics to those shown here. As a final point, it would be interesting to conjecture that the *stability* of high peak correlations to small perturbations away from the Gaussian is somehow related to the asymptotic theory of statistical extremes; a similar conjecture is made in Coles & Barrow (1987) concerning the areas of hotspots on the microwave sky (see also Catelan, Lucchin & Matarrese 1988).

5 Nearest-neighbour distributions and void sizes

So far in this study of the clustering of density maxima, we have concentrated on the two-point correlation function of peaks, $\xi_{\rm pk-pk}(r)$. This function contains only very limited information about the spatial distribution, however, and it is therefore interesting to look at other statistical characteristics in the hope of providing as many independent tests of the biasing hypothesis as possible.

The most obvious next step would be to look at the higher order correlations of peaks. Unfortunately, however, the n-point correlation functions for peaks are rather difficult to obtain for $n>2$, even by Monte-Carlo methods and even for 1D noise. Politzer & Wise (1984) and Jensen & Szalay (1986) have discussed higher order correlation functions using the high-level region approximation discussed in Section 2.2. We have already seen, however, that this approximation is inaccurate in the cosmologically interesting part of the parameter space (Section 3) and one should therefore be rather wary of placing too much reliance on these results. In addition to these theoretical difficulties, it is also true that higher order correlations are very difficult to determine with any accuracy from limited samples, such as those rich cluster samples available at the present time.

What we need is a statistic we can extract easily from a sample and which can also be studied in a theoretical context. One of the most simple such statistics is the nearest-neighbour distribution, $N_{\rm nn}(r)$, which is just the distribution of distances from each cluster to its nearest neighbour. The mean nearest-neighbour distance will be denoted $D_{\rm nn}$:

$$D_{\rm nn}=\sum_r rN_{\rm nn}(r)/\sum_r N_{\rm nn}(r).$$

Nearest-neighbour distributions have been discussed extensively in the past (Bogart & Wagoner 1973; Turner & Gott 1975; Fall *et al.* 1976; White 1979; Bahcall & Soneira 1981) and the distribution of nearest-neighbour *angles* was given for the projected cluster positions by Bahcall & Soneira (1983), although the small size of the latter sample makes the distributions rather noisy.

It is difficult to obtain an analytic expression for $N_{\rm nn}(r)$ except for very simple clustering models, but it is much easier to extract $N_{\rm nn}$ from a Monte-Carlo simulation than it is to extract

higher order correlation functions. In the discussion here, as previously, we shall not attempt to derive the quantitative 3D behaviour of $N_{nn}(r)$ but we use 1D simulations to see what the qualitative behaviour of this distribution is for peaks of a Gaussian field. Note that $N_{nn}^{(1D)}$ will not be even approximately the same shape as $N_{nn}^{(3D)}$ even if the 1D maximum constraint were to select 3D peaks; the difference can be illustrated by the Peebles (1980) Poisson cluster model. In this model, galaxies are distributed at random in space, with a number density given by some continuous density function $\rho(\mathbf{r})$. Not all spatial distributions can be represented in this way-among those that cannot would be hard spheres or 'peaks' as described above (Peebles 1980, p. 148). In 3D, the nearest-neighbour distribution will be given by

$$N_{nn}^{(3D)}(r) \sim \rho(r) r^2 \exp\left[-4\pi \int_0^r \rho(r) r^2 dr \right], \tag{56}$$

assuming the distribution is symmetric and using $r = |\mathbf{r}|$. The corresponding nearest-neighbour-along-a-line distribution is

$$N_{nn}^{(1D)}(x) \sim \rho(x) \exp\left[-\int_0^x \rho(x) dx \right]. \tag{57}$$

Notwithstanding these important reservations, we can use the 1D nearest-neighbour distribution to see whether this statistic seems to contain any useful information about the clustering of maxima. In addition, this approach is relevant to the 'void diameter' distribution discussed below (Einasto & Einasto 1988).

Fig. 5 shows the distribution of nearest-neighbour distances for peaks above levels $\nu = 0,1,2,3$ (a,b,c,d, respectively). These were obtained from the simulations of 1D Gaussian noise possessing the covariance function (28) discussed in Section 3. The striking fact to notice is that, as we increase the threshold ν, a long 'tail' in $N_{nn}(r)$ develops, extending to many times the coherence length (five computer distance units in Fig. 5). This long tail indicates that, despite being highly clustered, many of the high peaks are still rather isolated, reflecting the fact that they are very rare. This tail is also longer than one would expect to see in the 3D nearest-neighbour distribution, as one can see from (60) and (61); the exponential term falls much more quickly in the 3D case. The tails resemble the unclustered (Poisson) case very closely [$N_{nn}(x) \sim \exp(-\langle n \rangle x)$, where $\langle n \rangle \simeq \sigma_1 \exp(-u^2/2)/2\pi$ from Coles & Barrow 1987], so that these tails are not due to any correlation properties, just to the scarcity of points. Note also that, although the graphs are rather noisy for $\nu > 2$, the distributions are much better determined than $\xi_{pk-pk}(r)$; Fig. 1. This suggests that the 3D nearest-neighbour distributions might provide very useful statistical discriminants and that these should be studied further, especially because information is contained in $N_{nn}(r)$ about both the mean number density and $\xi(r)$. One requires $\langle n \rangle$ to estimate $\xi(r)$ which makes it difficult to assign realistic errors to determinations of $\xi(r)$ when $\langle n \rangle$ and $\xi(r)$ are both estimated from the same galaxy sample.

Fig. 6 shows what happens when we calculate the mean nearest-neighbour distance D_{nn} from subsets of each simulation of length L (in computer units). Notice that, for small ν, the mean is well-defined at relatively small L and does not increase with subset size but when $\nu = 3$ the flat tail of the distribution ensures that $D_{nn} \propto L$ out to more than 100 coherence lengths.*

This scaling phenomenon is interesting in its similarity to the scaling law obeyed by the void statistic studied by Einasto & Einasto (1988). Various statistics have been suggested to

* Note that if r is uniformly random on $(0, L)$ then $\langle r \rangle = L/2$ so that r increases linearly with L.

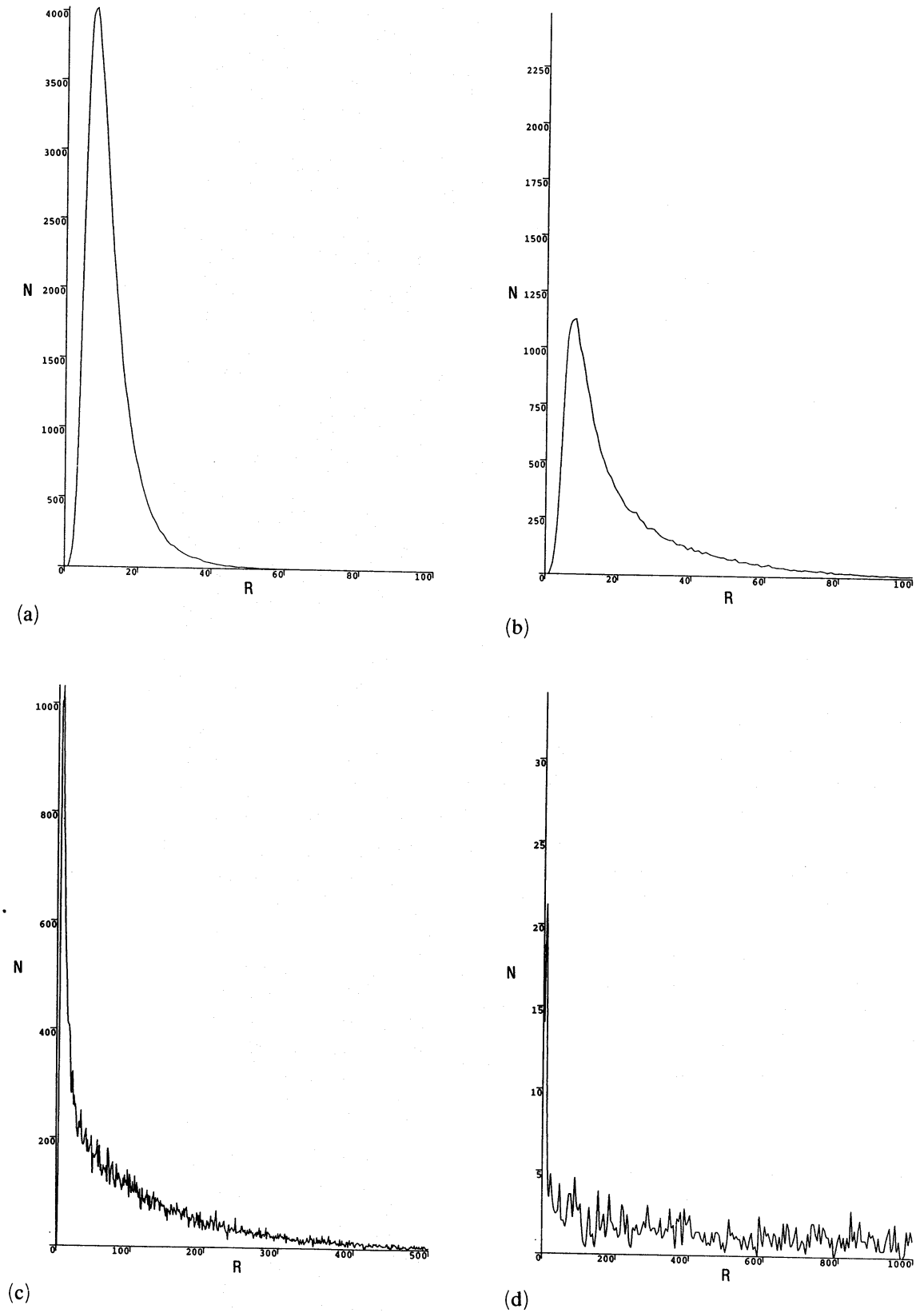

Figure 5. Distribution of nearest-neighbour distances for peaks above thresholds $\nu = 0, 1, 2, 3$ (a, b, c, d).

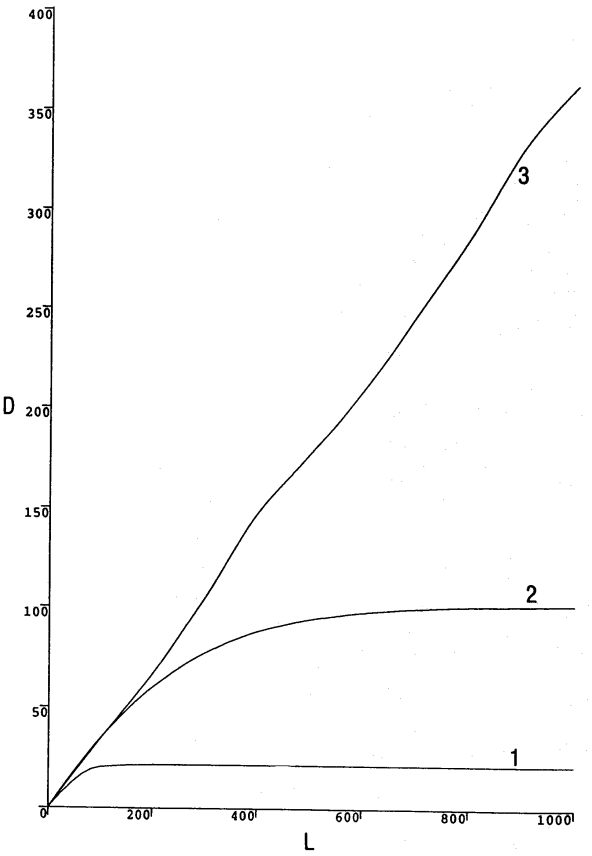

Figure 6. The scaling of D_{nn} with L for peaks above thresholds $\nu = 1, 2, 3$ for fixed coherence length.

characterize the tendency of the galaxy distribution to contain large void volumes having no (or very few) galaxies. One of the problems with these statistics is that it is very difficult to define the *shape* of voids in a systematic way. Einasto & Einasto dispense with this difficulty by advocating a 1D void size statistic. They take a cubic galaxy sample of side L, divide it into small 'cells' of side l and count the number of galaxies in each cell. Note that this is equivalent to *smoothing* the galaxy distribution on a scale $\sim l$. Next, they move along each 'row' of cells (each row containing L/l cells), isolate local maxima in the galaxy counts and calculate the distance between each number-count peak and the next such peak along the row. The mean of these distances is termed the mean void diameter.† If $k = L/l$ then $3k^2$ rows are sampled.

This statistic is slightly different to our mean nearest-neighbour statistic because the 'next peak along the row' might not be the 'nearest-neighbour' (which might be 'the previous peak along the row'). To see what effect this slightly different algorithm has, it is helpful to consider the Poisson Cluster Model again. The nearest-neighbour distribution in this case is easily seen to be

$$N_{nn}(x) \propto \rho(x) \exp\left[-\int_{-x}^{+x} \rho(x)\, dx \right]$$

$$\propto \rho(x) \exp\left[-2\int_{0}^{x} \rho(x)\, dx \right], \qquad (58)$$

†Note the similarity between this approach and that used in Coles (1986) and Coles & Barrow (1987) to estimate the mean sizes of regions above a threshold level.

if $\rho(x)$ is symmetric and the Einasto & Einasto void size distribution is just given by (57) so that for this simple model

$$D_{nn} = \tfrac{1}{2}\langle x_{EE}\rangle. \tag{59}$$

This simple relation does not hold for arbitrary spatial distributions of points but, for sensible models, the mean nearest-neighbour distance and the mean 'void diameter' should behave in a qualitatively similar way. We have extracted this statistic from the simulations in a similar way to the nearest-neighbour statistics and confirmed that the shape of the distributions are very similar. The mean values of x_{nn} and x_{EE} are compared in Table 1 for the case where $L = 1000$; they roughly obey (59) for smallish thresholds but the mean values are comparable for $\nu = 3$, again indicating that there are a sizeable number of isolated peaks in the latter case.

Einasto & Einasto take various galaxy samples with different scale lengths L and, keeping k fixed (i.e. altering l in proportion to L), they find that $D_{EE} = \langle x_{EE}\rangle \propto L$ for L up to $\sim 250\,h^{-1}$ Mpc. This is clearly not the same situation as that for which we obtained the scaling of L in Fig. 6 because the Einasto method involves smoothing the galaxy samples on different scales, thus each smoothed sample has a different *coherence length*. If we consider an underlying pattern that is roughly self-similar such as Gaussian fluctuations with $\xi(r) \sim r^{-\gamma}$, however, we expect the nearest-neighbour distribution to have the same shape on all scales. The form (28) used for these simulations is roughly self-similar on scales $r > r_c$ so if we identify the peaks in galaxy number-counts with peaks above a reasonably high level in an underlying density distribution possessing the covariance function (28) then there would seem to be no problem in reproducing, at least qualitatively, the observed scaling with L of the mean 'void diameter'.

This scaling with L does seem to imply that the galaxy distribution is self-similar to rather large scales. Einasto & Einasto argue that, if the observed self-similarity of D_{EE} is to be reproduced, the pattern must be either *random* or a *fractal*. The distribution $N_{EE}(x)$ looks sufficiently non-Poissonian for them to exclude the former possibility. The continuation of self-similarity to these large distances has interesting implications for the CDM model. The (unsmoothed) covariance function of the CDM fluctuations is approximately self-similar until we reach the zero crossing which is picked out by the scale of the horizon at matter-radiation equivalence $\sim 13\,h^{-2}$ Mpc if $\Omega_0 = 1$ (Sahni & Starobinsky 1984). We expect D_{EE} to scale with L until the appropriate l reaches this value. At $250\,h^{-1}$ Mpc, the value of l they use is $\sim 16\,h^{-1}$ Mpc which suggests that a small h could just about reconcile the observed scaling with the flat CDM model. However, it is by no means clear that the scaling they observe actually does imply a self-similar underlying distribution. We found a scaling of D_{nn} with L that was basically caused by the fact that high peaks are very rare and one therefore has to sample a very large volume in order to determine D_{nn} accurately. In the Einasto & Einasto analysis, this effect would also play a part if the catalogues they use contain a sufficiently small number of points that any self-similarity (or lack of it) would be hidden in the sampling uncertainties. For a small catalogue the scaling Einasto & Einasto observe might well all be due to the 'Poisson Tail' described above, so the argument that their scaling implies that the pattern must either be *random* or a *fractal* is not correct-it could be effectively a mixture of the two, with apparent

Table 1. Mean nearest-neighbour distances and 'void diameters' for $L = 1000$.

ν	0	1	2	3
D_{nn}	11.9	21.7	101.4	363.1
D_{EE}	18.6	43.0	204.6	400.3

randomness caused by the limited catalogue size. Before drawing any definite conclusions about the implications of this analysis for the CDM model it is therefore necessary to model these effects in detail, using 3D simulations of the galaxy samples studied by Einasto & Einasto (including the effects of noise) to investigate exactly where the break in scale-invariance should be observed.

6. Conclusions

The main conclusions of this work are as follows:

(i) The zero-crossings of $\xi_{\mathrm{pk-pk}}(r)$ do not, in general, coincide with those of $\xi(r)$; for cosmologically interesting correlation functions the first zero-crossing of $\xi_{\mathrm{pk-pk}}(r)$ should occur at smaller r than the first zero-crossing of $\xi(r)$.

(ii) The amplitude of the biased correlation function $\xi_{\mathrm{pk-pk}}(r)$ is smaller than previous estimates suggested (Kaiser 1984; Politzer & Wise 1984; Bardeen *et al.* 1986; Jensen & Szalay 1986) and these estimates only become accurate at very large r, when $\xi(r)$ is very small.

(iii) Non-linear evolution should have little effect on the form of bias achieved.

The first two effects cast some doubt on the ability of the CDM model to reproduce the observed cluster–cluster correlation function. Unfortunately this issue is clouded by uncertainties in both the observed $\xi_{\mathrm{cc}}(r)$ and the theoretically favoured normalization of the CDM spectrum, neither of which are known to better than a factor of 2, even ignoring the usual uncertainties in Ω_0 and h. It does seem, however, that one requires a high normalization *and* low $r_{0,\mathrm{c}}$ to reconcile CDM with observations. This conclusion is based on approximate arguments and it is not possible categorically to rule out CDM on the basis of this work, although our results agree with those obtained by other methods (Otto *et al.* 1986a; White *et al.* 1987a,b; Lumsden *et al.* 1989). More accurate determinations of $\xi_{\mathrm{cc}}(r)$ and N-body simulations with sufficient dynamical range to incorporate both galaxies and clusters would be required to suppy an unequivocal test of the CDM model. In the meantime, before these requirements can be achieved, it is worth looking at independent measures of the clustering behaviour of peaks. In Section 5, we looked briefly at one such measure, the nearest-neighbour distribution which might well provide a good test of the biasing hypothesis. We also mentioned that previous estimates of the properties of the higher-order-n-point correlation functions of biased distributions might also be inaccurate (Politzer & Wise 1984; Jensen & Szalay 1986) and this is certainly worth checking. In future work, we intend to look at the three-point correlation function of local maxima to see how closely it resembles the above predictions.

Acknowledgments

I thank John Barrow, Bernard Jones, Lone Appel, John Peacock and Alan Heavens for discussions. This work was supported by an SERC postgraduate studentship.

References

Abell, G. O., 1958. *Astrophys. J. Suppl.*, **3,** 211.
Abramowitz, M. & Slegun, T. A., 1965. *Handbook of Mathematical Functions,* Dover, New York.
Adler, R. J., 1981. *The Geometry of Random Fields,* John Wiley, New York.
Bahcall, J. M. & Soneira, R. M., 1981. *Astrophys. J.*, **246,** 122.
Bahcall, N. A. & Soneira, R. M., 1983. *Astrophys. J.*, **270,** 20.
Bahcall, N. A., Soneira, R. M. & Burgett, W., 1986. *Astrophys. J.*, **311,** 15.
Bardeen, J. M., Bond, J. R. & Efstathiou, G., 1987. *Astrophys. J.*, **321,** 28.
Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S., 1986. *Astrophys. J.*, **304,** 15.

Batuski, D. J., Melott, A. L. & Burns, J. O., 1987. *Astrophys. J.*, **322,** 48.
Bogart, R. S. & Wagoner, R. V., 1973. *Astrophys. J.*, **181,** 609.
Cartwright, D. E. & Longuet-Higgins, M. S., 1956. *Proc. R. Soc. London A,* **237,** 212.
Catelan, P., Lucchin, F. & Matarrese, S., 1988. *Phys. Rev. Lett.*, **61,** 273.
Cline, J. M., Politzer, H. D., Rey, S.-J. & Wise, M. B., 1987. *Communs math. Phys.*, **112,** 217.
Coles, P., 1986. *Mon. Not. R. astr. Soc.*, **222,** 9P.
Coles, P. & Barrow, J. D., 1987. *Mon. Not. R. astr. Soc.*, **228,** 407.
Couchman, H. M. P., 1987a. *Mon. Not. R. astr. Soc.*, **225,** 777.
Couchman, H. M. P., 1987b. *Mon. Not. R. astr. Soc.*, **225,** 795.
Davis, M. & Peebles, P. J. E., 1983. *Astrophys. J.*, **267,** 465.
Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M., 1985. *Astrophys. J.*, **292,** 371.
Dekel, A., Blumenthal, G. R. & Primack, J. R., 1988. Preprint.
Einasto, J. & Einasto, M., 1987. *Mon. Not. R. astr. Soc.*, **226,** 543.
Fall, S. M., Geller, M. J., Jones, B. J. T. & White, S. D. M., 1976. *Astrophys. J.*, **205,** L121.
Fry, J. N., 1986. *Astrophys. J.*, **308,** L71.
Goroff, M. H., Grinstein, B., Rey, S.-J. & Wise, M. B., 1986. *Astrophys. J.*, **311,** 6.
Gradshteyn, I. S. & Ryzhik, I. M., 1965. *Tables of Integrals, Series and Products,* Academic Press, New York.
Grinstein, B. & Wise, M. B., 1986. *Astrophys. J.*, **310,** 9.
Groth, E. J. & Peebles, P. J. E., 1977. *Astrophys. J.*, **217,** 385.
Gupta, S. S., 1963. *Ann. math. Statist.*, **34,** 829.
Hubble, E., 1934. *Astrophys. J.*, **79,** 8.
Jensen, L. G. & Szalay, A. S., 1986. *Astrophys. J.*, **305,** L5.
Juszkiewicz, R., Sonoda, D. H. & Barrow, J. D., 1984. *Mon. Not. R. astr. Soc.*, **209,** 139.
Kaiser, N., 1984. *Astrophys. J.*, **284,** L9.
Kendall, M. & Stuart, A., 1977. *The Advanced Theory of Statistics,* Vol. 1, Charles Griffin, London.
Klypin, A. A. & Kopylov, A. I., 1983. *Soviet. Astr. Lett.*, **9,** 75.
Ling, E. N., Frenk, C. S. & Barrow, J. D., 1986. *Mon. Not. R. astr. Soc.*, **223,** 21P.
Lucey, J., 1983. *Mon. Not. R. astr. Soc.*, **204,** 33.
Lumsden, S. L., Heavens, A. F. & Peacock, J. A., 1989. *Mon. Not. R. astr. Soc.*, **238,** 293.
Matarrese, S., Lucchin, F. & Bonometto, S. A., 1986. *Astrophys. J.*, **310,** L21.
Otto, S., Politzer, H. D. & Wise, M. B., 1986a. *Phys. Rev. Lett.*, **56,** 1978.
Otto, S., Politzer, H. D. & Wise, M. B., 1986b. *Phys. Rev. Lett.*, **56,** 2772.
Peacock, J. A. & Heavens, A. F., 1985. *Mon. Not. R. astr. Soc.*, **217,** 805.
Pearson, K., 1901. *Phil. Trans. R. Soc. Lond. A,* **195,** 1.
Peebles, P. J. E., 1980. *The Large Scale Structure of the Universe,* Princeton University Press, Princeton.
Politzer, H. D. & Wise, M. B., 1984. *Astrophys. J.*, **285,** L1.
Postman, M., Geller, M. J. & Huchra, J. P., 1986. *Astr. J.*, **91,** 1267.
Rice, S. O., 1945. *Bell. System. Tech. J.*, **24,** 46.
Sahni, V. & Starobinski, A. S., 1984. Preprint.
Struble, M. F. & Rood, H. J., 1987. *Astrophys. J. Suppl.*, **63,** 543.
Sutherland, W., 1988. *Mon. Not. R. astr. Soc.*, **234,** 159.
Turner, E. L. & Gott, J. R., 1975. *Astrophys. J.*, **197,** L89.
Vanmarcke, E. H., 1983. *Random Fields: Analysis and Synthesis,* MIT Press, Cambridge, Massachusetts.
Wax, N. (ed.), 1954. *Selected Papers on Noise and Stochastic Processes,* Dover, New York.
White, S. D. M., 1979. *Mon. Not. R. astr. Soc.*, **186,** 145.
White, S. D. M., Davis, M., Efstathiou, G. & Frenk, C. S., 1987a. *Nature,* **330,** 451.
White, S. D. M., Frenk, C. S., Davis, M. & Efstathiou, G., 1987b. *Astrophys. J.*, **313,** 505.

Appendix A: $\xi^{(1D)}_{pk-pk}(x)$ as an approximation to $\xi^{(3D)}_{pk-pk}(r)$

In Section 1 we argued that the study of peaks of stochastic processes in 1D should give us a good qualitative guide to the behaviour of peaks of 3D random fields. In this Appendix we show that, at least for large thresholds, the results should also be in reasonable quantitative agreement.

To see this, consider a 3D random field $\varepsilon(\mathbf{x})$ where $\mathbf{x}=\{x_i\}$, $(i=1,3)$ and suppose that our 1D process in Section 3 is a slice taken through this field [i.e. that $\varepsilon(x)=\varepsilon(x_1)$, say]. To study

the peaks of the 3D field we need to consider the joint distributions of $\varepsilon(\mathbf{x})$, $\partial\varepsilon(x)/\partial x_i = \varepsilon_i (i=1, 3)$ and $\partial^2\varepsilon(x)/\partial x_i \partial x_j = \varepsilon_{ij}$ $(i,j=1,3)$. We denote first derivatives as ε_i and second derivatives by ε_{ij}.

An important property of multivariate Gaussian distributions is that the conditional pdf of the first m of the variables [expressed as the vector $\mathbf{x}_1 = (x_1, \ldots, x_m)$] given the values of the remaining variables $x_2 = (x_{m+1}, \ldots, x_n)$ is also a multivariate Gaussian with mean vector

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \tag{A1}$$

and covariance matrix

$$\mathbf{V}_{1|2} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}, \tag{A2}$$

where we have used the notation $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = [(\mu_1, \ldots, \mu_m), (\mu_{m+1}, \ldots, \mu_n)]$ and

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}.$$

Choosing, as we did in Section 3, that $\varepsilon_1 = 0$ and $\varepsilon_{11} < 0$ specifies a peak of the 1D 'slice' but is not enough to specify a peak of the 3D field. To specify the latter, we would have to place similar conditions on all the other first and second derivatives. But consider, for example, the joint pdf of ε, ε_{11}, ε_{22}. Conditioning on some value of ε and some value of ε_{11} has an effect on the distribution of ε_{22} because these variates are not independent.* The joint pdf of $\mathbf{y} = (\varepsilon, \varepsilon_{11}, \varepsilon_{22})$ is a Gaussian with covariance matrix

$$\mathbf{V} = \begin{pmatrix} 1 & -\sigma_1^2 & -\sigma_1^2 \\ -\sigma_1^2 & \sigma_2^2 & \sigma_2^2/3 \\ -\sigma_1^2 & \sigma_2^2/3 & \sigma_2^2 \end{pmatrix}.$$

See, e.g. Peacock & Heavens (1985), Bardeen *et al.* (1986) and Couchman (1987a, b).

Using the above prescription for conditional distributions of Gaussian variates we can find the *conditional* pdf of ε_{22} for specified values of ε and ε_{11}. These conditional distributions are all Gaussian with some mean μ and variance σ^2 which we denote as $N(\mu; \sigma^2)$.

Firstly, the distribution of ε_{22} given that $\varepsilon = \varepsilon^*$ is found to be

$$P(\varepsilon_{22} | \varepsilon = \varepsilon^*) = N(-\sigma_1^2\varepsilon^*; \sigma_2^2 - \sigma_1^4).$$

Now the unconditional pdf of ε_{22} is $N(0; \sigma_2^2)$ so that picking a positive value of ε^* decreases both the mean value and the variance of ε_{zz}. Large values of ε^* will mean that the second deriative components are much more likely to be negative and therefore the point is more likely to be a local maximum then any typical point.

Next consider the conditional distribution of ε_{22} given that ε_{11} has some value, say ε_{11}^*. By the same method as above we find that

$$P(\varepsilon_{22} | \varepsilon_{11} = \mathrm{e}_{11}^*) = N(\varepsilon_{11}/3; 8\sigma_2^2/9),$$

so we can see that conditioning on a negative value of ε_{11}^* moves the mean downwards and also decreases the variance (although only by a small amount in this case – about 11 per cent).

When these two effects are added together, it is clear that thresholding ε and simultaneously selecting 1D maxima ($\varepsilon_{11} < 0$) will also tend to select 3D maxima because large positive values of ε and large negative values of ε_{11} shift the distribution of the other second derivative components downwards and reduce its dispersion. This combination of effects should be

* All these variables *are* independent of ε_1 so that conditioning on $\varepsilon_1 = 0$ has no effect on the distribution of ε_{22}.

enough to ensure that the 1D calculations should be in reasonable quantitative agreement with the 3D calculations, at least for very high thresholds; in particular the calculation should converge to the exact result much more quickly with increasing threshold than any calculations based on high-level regions.

Appendix B: Recursion relations and derivatives for Bessel functions

We start with the form (29) for $\xi(r)$ which is

$$\xi(r)=J_0(kr),$$

where $J_0(x)$ is a Bessel function of the first kind (Abramowitz & Stegun 1965). To avoid writing too many k's we set $k=1$ (it is easy to put the k's back at the end). We need to consider $J_0^{(n)}(x)$, the nth derivative with respect to x, for all $n\leq 4$. We find that it is possible to express all these functions in terms of x, J_0 and J_1 and the latter two functions are easy to compute numerically using standard NAG routines.

First note the standard results (Abramowitz & Stegun 1965) that

$$\xi'(x)=J_0'(x)=-J_1(x)$$

and

$$J_{n-1}(x)-J_{n+1}(x)=\frac{2n}{x}J_n(x)$$

and also

$$\left(\frac{1}{x}\frac{d}{dx}\right)^k[x^nJ_n(x)]=x^{n-k}J_{n-k}(x).$$

So that

$$\begin{aligned}J_0''(x)&=-J_1'(x)\\&=-\tfrac{1}{2}[J_0(x)-J_2(x)]\end{aligned}$$

and

$$J_2(x)=\frac{2}{x}J_1(x)-J_0(x) \tag{B1}$$

together lead to

$$\xi''(x)=J_0''(x)=\frac{1}{x}J_1(x)-J_0(x).$$

Similarly, we find that

$$J_0'''(x)=\tfrac{3}{4}J_1(x)-\tfrac{1}{4}J_3(x)$$

and

$$J_3(x)=\frac{4}{x}\left[\frac{2}{x}J_1(x)-J_0(x)\right]-J_1(x),$$

so that

$$\xi'''(x) = J_0'''(x) = \left(1 - \frac{2}{x^2}\right) J_1(x) + \frac{1}{x} J_0(x).$$

Finally, we find

$$J_0''''(x) = \tfrac{3}{8} J_0(x) - \tfrac{1}{2} J_2(x) + \tfrac{1}{8} J_4(x).$$

Using the fact that

$$J_4(x) = \left(\frac{48}{x^3} - \frac{8}{x}\right) J_1(x) + \left(1 - \frac{24}{x^2}\right)$$

and using (B1) for $J_2(x)$, we get

$$\xi''''(x) = J_0''''(x) = \left(1 - \frac{3}{x^2}\right) J_0(x) + \left(\frac{6}{x^3} - \frac{2}{x}\right) J_1(x).$$