CLUSTERING OF ATTRIBUTES ON REGULAR POINT LATTICES

WITH AN APPENDIX ILLUSTRATING THE APPLICATION OF THE PROCEDURE TO *SIREX* INFESTATIONS IN PLANTATIONS OF *PINUS RADIATA*

By A. M. W. VERHAGEN*

INTRODUCTION

Patterns of trees with a common attribute such as virus disease or high productivity for example, often suggest a tendency to clustering. The presence of clustering in the absence of known predisposing factors, and the absence of clustering in the presence of such factors require explanation and provide information useful for understanding and remedial action. However, chance effects may mask as well as suggest clustering so that gauging it merely by visual inspection of the patterns may be misleading. It is therefore desirable to develop an objective measure of clustering and study its statistical properties in the presence of random effects. In a context where an infected tree (a success) is suspected of contaminating adjoining neighbour trees, the average number of trees not affected (failures) adjoining a success suggests itself as a measure of clustering. In the case of trees on regular point lattices such as orchards and plantations, some sites may not carry trees, so that the absence or presence is not relevant. Sites on the lattice for which the attribute is irrelevant or cannot be ascertained (missing observation) will be called empty sites. Throughout this paper it will be convenient to refer to the occupied sites as successes or failures according to the presence or the absence of the attribute being considered, thus following the usual terminology of Bernoulli trials in probability theory.

The number of failures adjoining successes will be studied as a measure of clustering for general finite and infinite regular point lattices, with any prescribed pattern of empty sites, followed by applications to the square and the triangular lattices in two dimensions, and the time series case (one dimension).

SOME GENERAL THEORY

(i) *Preliminary analysis*

Finite lattices are often more realistic in application than infinite ones, so that when r successes are observed, the 'no replacement case', i.e. the random allocation of r successes to the non-empty sites of the lattice seems the appropriate null situation to be considered. Investigation of clustering on the finite lattice therefore involves combinatorial analysis. To facilitate this, it will be helpful to realize that the sites whose successes and failures are involved in the measure of clustering proposed, are the complement of the empty sites, and the sites which are isolated failures, i.e. failures without adjoining successes. Both these types of sites are defined without involving successes, so that combinatorial study of these will be simpler. Since the total number N of sites may be represented as

$$N = r+F+S \tag{1}$$

* Division of Mathematical Statistics, C.S.I.R.O., Melbourne.

where F is the number of failures adjoining the r successes, and S is the total number of sites which are either empty or isolated failures, it follows that for given r and N the required behaviour of F as a measure of clustering may be studied directly in terms of S with the aid of equation (1). It is proposed in this section to find general expressions for the mean and variance of S. Clustering, both positive and negative may then be tested, by comparing the observed value of S with a range of four standard deviations centred on the mean, constructed on the assumption that successes and failures occur independently on the occupied sites.

(ii) *Score variates* s_a *for the sites* (a) *of a lattice*

Consider any lattice, regular or otherwise in any dimension, consisting of N sites. Let each site (a), $(a = 1, \ldots, N)$ have a score variate s_a which is unity when the site is empty or an isolated failure, and zero otherwise. The total number S of empty sites and isolated failures may then be written

$$S = \sum_{a=1} s_a \tag{2}$$

and be called the grand score.

(iii) *Product moments of score variates*

The product moment of the score variates s_a and s_b depends on the relative positions of the sites (a) and (b) and the empty sites. To analyse this dependence more precisely define for each site (a) which is an isolated failure an index value α_a as the number of non-empty adjoining sites, including the site (a) and for each pair of isolated failures (a) and (b) an index value $\alpha_a^b = \alpha_b^a$ as the number of occupied sites adjacent to the numbers of the pair, including the pair. Also define a function ε_a which equals unity when the site (a) is empty and zero otherwise. The functions α_a and ε_a are determined by the position of the empty sites on the lattice. The number ε of empty sites on the lattice equals

$$\varepsilon = \sum_{a=1}^{N} \varepsilon_a \tag{3}$$

There are $\binom{N-r}{r}$ ways of allocating r successes to the $N-\varepsilon$ occupied sites, and of these a number $\binom{N-\varepsilon-\alpha_a}{r}$ will leave the site (a) an isolated failure if it is not empty. The score variate of a non-empty site (a) therefore has an expected value $\binom{N-\varepsilon-\alpha_a}{r}\Big/\binom{N-\varepsilon}{r}$

The expected value $E[s_a]$ of the score variate for any site (a) may therefore be written

$$E[s_a] = \varepsilon_a+(1-\varepsilon_a)\binom{N-\varepsilon-\alpha_a}{r}\Big/\binom{N-\varepsilon}{r} \tag{4}$$

Similarly among the $\binom{N-\varepsilon}{r}$ ways of allocating r successes to $(N-\varepsilon)$ occupied sites there are $\binom{N-\varepsilon-\alpha_a^b}{r}$ ways which leave both the sites (a) and (b) isolated failures when they are both non-empty. The expected value $\gamma_a^b \equiv E[s_a s_b]$ for any pair (a) and (b) is therefore

$$\gamma_a^b = E[s_a s_b] = \varepsilon_a\varepsilon_b+(1-\varepsilon_a)(1-\varepsilon_b)\binom{N-\varepsilon-\alpha_a^b}{r}\Big/\binom{N-\varepsilon}{r}+(1-\varepsilon_b)\varepsilon_a\binom{N-\varepsilon-\alpha_b}{r}\Big/\binom{N-\varepsilon}{r}$$
$$+(1-\varepsilon_a)\varepsilon_b\binom{N-\varepsilon-\alpha_a}{r}\Big/\binom{N-\varepsilon}{r} \tag{5}$$

When in equation (5) the site (a) is the same as the site (b) the product moment reduces to the second moment of s_a since $\alpha_{a_1}^{a} = \alpha_a$.

(iv) *The mean* E[S] *and the variance* VAR[S] *of the grand score* S

It follows from equation (2) that the mean $E[S]$ of S equals

$$E[S'] = \sum_a{}' E[s_a] = \varepsilon + \sum (1-\varepsilon_a) \binom{N-\varepsilon-\alpha_a}{r} \Big/ \binom{N-\varepsilon}{r} \tag{6}$$

and the second moment

$$\begin{aligned} E[S'^2] = {} & \varepsilon^2 + \sum_{a,b} (1-\varepsilon_a)(1-\varepsilon_b) \binom{N-\varepsilon-\alpha_a^b}{r} \binom{N-\varepsilon}{r} \\ & + \sum_{a,b} (1-\varepsilon_a)(\varepsilon_a) \binom{N-\varepsilon-\alpha_b}{r} \Big/ \binom{N-\varepsilon}{r} + \sum_{a,b} (1-\varepsilon_a)\varepsilon_b \binom{N-\varepsilon-\alpha_a}{r} \Big/ \binom{N-\varepsilon}{r} \end{aligned} \tag{7}$$

so that

$$\mathrm{VAR}\,[S'] = E[S'^2] - \{E[S']\}^2 \tag{8}$$

may be obtained by substitution of equations (6) and (7) in (8). The determination of $E[S]$ and VAR S from the general combinatorial solution derived here, may for any lattice characteristics in hand be readily automated with the aid of a digital computer. For regular lattices the number of possible values of α_a^b is small and so is the number of different Newtonian coefficients involving these, so that they can be supplied as input data. In practice the general solution can usually be further reduced by incorporating the special features and regularities of the lattice involved, by wrapping it around a torus by fusion of boundary sites. The remainder of the paper is devoted to reduced formulae for regular lattices.

REDUCED FORMULAE FOR THE SQUARE LATTICE (EIGHT ADJOINING SITES)

(i) *No empty sites* ($\varepsilon = 0$)

Consider the torus of $N = m \times n$ sites obtained by fusing the border rows and also the border columns of an $(m+1)(n+1)$ square lattice. Then the co-ordinates (i, j) may be used as labels for the $m \times n$ sites. Application of equation (4) gives

$$E[s_{ij}] = \binom{N-9}{r} \Big/ \binom{N}{r} \qquad (i = 1, \ldots, m, j = 1, \ldots, n) \tag{9}$$

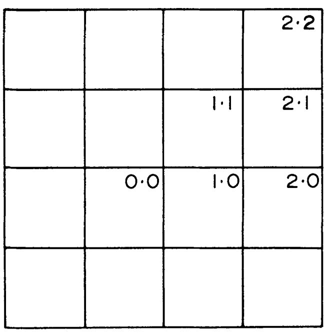

FIG. 1. Sites on the square lattice which share some of their eight neighbours with the typical site (0·0).

It is clear that the typical site (0·0) on the square lattice represented in Fig. 1 shares neighbours with twenty-four other sites, and that these sites are of five kinds, represented by the sites (1·0), (1·1), (2·0), (2·1) and (2·2) respectively 4, 4, 4, 8 and 4 times. Application of equation (5) gives

$$
\begin{aligned}
E[s_{00}s_{00}] &= \binom{N-9}{r}\Big/\binom{N}{r} & E[s_{00}s_{20}] &= \binom{N-15}{r}\Big/\binom{N}{r} \\
E[s_{00}s_{10}] &= \binom{N-12}{r}\Big/\binom{N}{r} & E[s_{00}s_{21}] &= \binom{N-16}{r}\Big/\binom{N}{r} \\
E[s_{00}s_{11}] &= \binom{N-14}{r}\Big/\binom{N}{r} & E[s_{00}s_{22}] &= \binom{N-17}{r}\Big/\binom{N}{r}
\end{aligned}
\tag{10}
$$

and

$$
E[s_{00}s_{ij}] = \binom{N-18}{r}\Big/\binom{N}{r} \tag{11}
$$

for sites (ij) not sharing neighbours with (0·0). Equations (6) and (7) then reduce to

$$
E[S'] = N\binom{N-9}{r}\ \binom{N}{r} \tag{12}
$$

$$
E[S^2] = N\left\{\binom{N-9}{r}+\binom{N-12}{r}+4\binom{N-14}{r}+4\binom{N-15}{r}+8\binom{N-16}{r} +4\binom{N-17}{r}+(N-25)\binom{N-18}{r}\right\}\Big/\binom{N}{r} \tag{13}
$$

provided the number m of rows and the number n of columns both exceed 3.

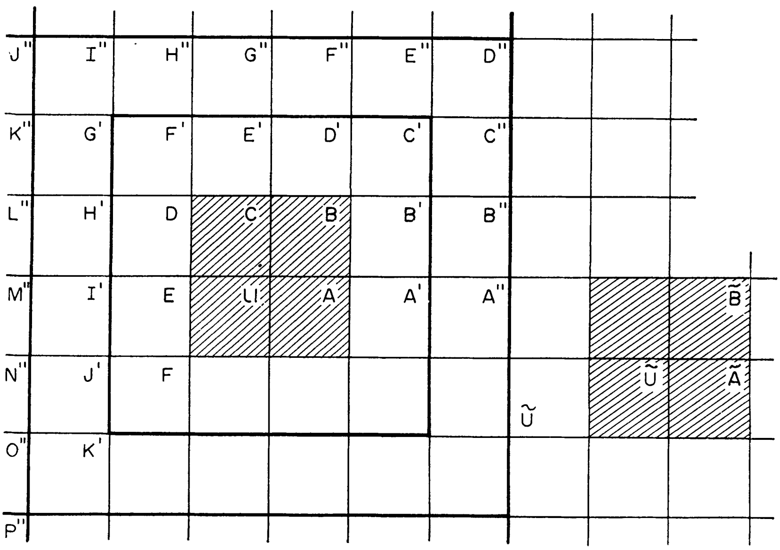

FIG. 2.

(ii) *Well-separated empty sites*

For any given empty site U one may define three concentric boundaries of sites as indicated in Fig. 2.

An empty site will be called 'well-separated' when no other empty sites are adjacent to the sites on these three boundaries. It will be assumed in this section that all empty sites are well separated or more precisely that at least one of the co-ordinates of any pair of empty sites differ by five units or more. Let $V_1, \ldots, V_\varepsilon$ represent the set of nine sites, on or within the first boundaries of each of the empty sites and let $\bar{V}$ be their complement containing $N-9\varepsilon$ sites. The expected value of the score variate s_{ij} is (cf. equation 4)

$$\begin{aligned} E[s_{i,j}] &= 1 && \text{when } (i,j) \text{ is empty in } V_1, \ldots, V_\varepsilon \\ &= \binom{N-\varepsilon-8}{r} \Big/ \binom{N-\varepsilon}{r} && \text{when } (i,j) \text{ is not empty in } V_1, \ldots, V_\varepsilon \\ &= \binom{N-\varepsilon-9}{r} \Big/ \binom{N-\varepsilon}{r} && \text{when } (i,j) \text{ is in } \bar{V} \end{aligned} \tag{14}$$

so that (cf. equation 6)

$$E[S'] = \varepsilon + 8\varepsilon\binom{N-\varepsilon-8}{r} \Big/ \binom{N-\varepsilon}{r} + (N-9\varepsilon)\binom{N-\varepsilon-9}{r} \Big/ \binom{N-\varepsilon}{r} \tag{15}$$

The derivation of $E[S^2]$ when there are ε well-separated empty sites is carried out by counting the changes in the index value α for all pairs and then modifying formula (13) accordingly (cf. Appendix). Thus for well-separated empty sites;

$$\begin{aligned} E[S'^2] = \varepsilon'^2 + \Bigg[& (N-9\varepsilon-18\varepsilon^2+2\varepsilon N)\binom{N-9-\varepsilon}{r} + 40\varepsilon\binom{N-11-\varepsilon}{r} \\ & + (4N-48\varepsilon)\binom{N-12-\varepsilon}{r} + 48\varepsilon\binom{N-13-\varepsilon}{r} \\ & + (4N-4\varepsilon)\binom{N-14-\varepsilon}{r} + (4N+36\varepsilon)\binom{N-15-\varepsilon}{r} \\ & + (8N-100\varepsilon+63\varepsilon^2)\binom{N-16-\varepsilon}{r} + (4N-340\varepsilon-144\varepsilon^2+16\varepsilon N)\binom{N-17-\varepsilon}{r} \\ & + (N-25+369\varepsilon+81\varepsilon^2-18\varepsilon N)\binom{N-18-\varepsilon}{r} \Bigg] \Big/ \binom{N-\varepsilon}{r} \end{aligned} \tag{16}$$

provided the number m of rows and the number n of columns both exceed 3.

(iii) *Few empty sites and large lattices*

The ratio $\binom{N-K}{r} \Big/ \binom{N}{r}$ of Newtonian coefficients may be written $(N-K)!(N-r)!/N!(N-r-K)!$ and this reduces to

$$\frac{(N-K)!(N-r)(N-r-1)\cdots(N-r-K+1)(N-r-K)!}{N(N-1)\cdots(N-K+1)(N-K)!(N-r-K)!}$$

so that

$$= \frac{\binom{N-K}{r}}{\binom{N}{r}} = \frac{(N-r)(N-r-1)\cdots(N-r-8)}{N(N-1)\cdots(N-8)} \to \left(1-\frac{r}{N}\right)^K \tag{17}$$

as $r \to \infty$. Thus when the number of empty sites is small and the lattice large, equations (6) and (7) reduce (asymptotically) to

$$E[S] = Nq^9 \tag{18}$$

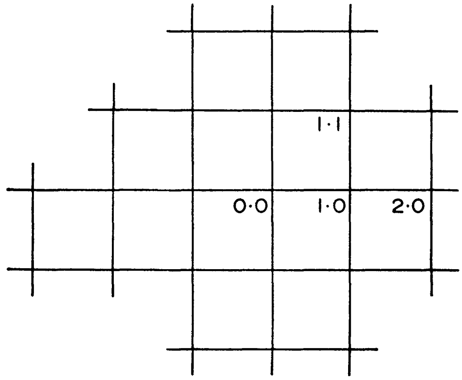

FIG. 3. Sites on the square lattice which share some of their four neighbours with the typical site (0·0).

$$E[S^2] = N(q^9+4q^{12}+4q^{14}+4q^{15}+8q^{16}+4q^{17}+(N-25)q^{18}) \tag{19}$$

where $q = (1-r/N)$, is usually called the probability of failure.

REDUCED FORMULAE FOR THE SQUARE LATTICE (FOUR ADJOINING SITES)

(i) *No empty sites* ($\varepsilon = 0$)

When on the $m \times n$ torus the diagonal sites are not taken as adjoining, each site has only four adjoining sites, equation (4) then becomes

$$E[s_{ij}] = \binom{N-5}{r}\bigg/\binom{N}{r} \qquad (i = 1, \ldots, m, j = 1, \ldots, n) \tag{20}$$

The typical site (0·0) on the square lattice in Fig. 1 shares neighbours with twelve other sites, which are of three types represented respectively by the (1·0), (2·0) and (1·1) four times each. Application of equation (5) gives

$$E[s_{00}s_{00}] = \binom{N-5}{r}\bigg/\binom{N}{r}, \qquad E[s_{00}, s_{20}] = \binom{N-9}{r}\bigg/\binom{N}{r} \tag{21}$$
$$E[s_{00}s_{01}] = \binom{N-8}{r}\bigg/\binom{N}{r}, \qquad E[s_{01}, s_{11}] = \binom{N-8}{r}\bigg/\binom{N}{r}$$

and

$$E[s_{00}s_{ij}] = \binom{N-10}{r}\bigg/\binom{N}{r} \tag{22}$$

for sites (i,j) not sharing neighbours with (0·0). Equations (6) and (7) reduce to

$$E[S] = N\binom{N-5}{r}\bigg/\binom{N}{r} \tag{23}$$

$$E[S^2] = N\left[\binom{N-5}{r}+8\binom{N-8}{r}+4\binom{N-9}{r}+(N-13)\binom{N-10}{r}\right]\bigg/\binom{N}{r} \tag{24}$$

(ii) *Well-separated empty sites*

When the sites on three boundaries constructed around the empty site U are not adjacent to other empty sites, as in Fig. (4) the site U is called a well-separated empty site.

Let all empty sites present on the torus lattice be well separated, i.e. the sum of the absolute differences between coordinates of any two empty sites exceeds five. Let $V_1, \ldots, V_\varepsilon$ represent the five sites on or within the first boundary of each of the empty sites and let $\bar{V}$ be their complement containing $N-5\varepsilon$ sites. The expected value of the score variate s_{ij} is (cf. equation 4).

$$\begin{aligned} E[s_{ij}] &= 1 && \text{when } (i,j) \text{ is empty in one of } V_1, \ldots, V_\varepsilon \\ &= \binom{N-\varepsilon-4}{r} && \text{when } (i,j) \text{ is not empty in one } V_1, \ldots, V_\varepsilon \\ &= \frac{\binom{N-\varepsilon-5}{r}}{\binom{N-\varepsilon}{r}} && \text{when } (i,j) \text{ is in } \bar{V} \text{ so that (cf. equation 6)} \end{aligned} \tag{25}$$

$$E[S] = \varepsilon+4\varepsilon\binom{N-\varepsilon-4}{r}\bigg/\binom{N-\varepsilon}{r}+(N-5\varepsilon)\binom{N-\varepsilon-5}{r}\bigg/\binom{N-\varepsilon}{r} \tag{26}$$

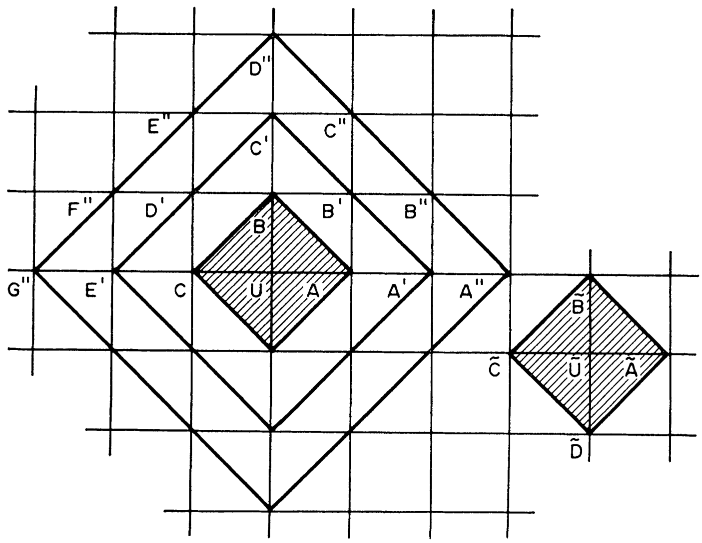

FIG. 4.

The derivation of $E[S^2]$ when there are ε empty sites all well separated, involves counting the changes in the index value α of all the pairs when the given pattern of empty sites is introduced. This gives

$$\begin{aligned} E[S^2] = \varepsilon^2+\bigg[&(8\varepsilon+4\varepsilon^2)\binom{N-4-\varepsilon}{r}+(N-5\varepsilon-10\varepsilon^2+2\varepsilon N)\binom{N-5-\varepsilon}{r} \\ &+48\varepsilon\binom{N-7-\varepsilon}{r}+(8N-52\varepsilon+16\varepsilon^2)\binom{N-8-\varepsilon}{r} \\ &+(4N-100\varepsilon-40\varepsilon^2+8\varepsilon N)\binom{N-9-\varepsilon}{r} \end{aligned}$$

$$+(N-13+101\varepsilon+29\varepsilon^2-10\varepsilon N)\binom{N-10-\varepsilon}{r}\Big]\Big/\binom{N-\varepsilon}{r} \tag{27}$$

for 'well-separated' empties on a torus lattice in which both m and n exceed 3.

(iii) *Few empty sites and large lattices*

Using equation (17) for the case of few empty sites and large lattices, the first and second moment, cf. equations (6) and (7), reduced to

$$E[S] = Nq^5 \tag{28}$$

$$E[S^2] = N(q^5+8q^8+4q^9+(N-13)q^{10}) \tag{29}$$

where $q = r/N$.

REDUCED FORMULAE FOR THE TRIANGULAR LATTICE (SIX ADJOINING SITES)

(i) *No empty sites* ($\varepsilon = 0$)

Another layout of practical interest in plantation and orchards is the triangular lattice illustrated in Fig. 5.

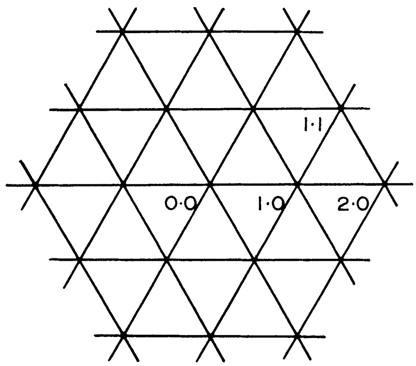

FIG. 5. Sites on the triangular lattice which share some of their six neighbours with the typical site (0·0).

For the triangular lattice, torus wrapped, it follows

$$E[s_{ij}] = \binom{N-7}{r}\Big/\binom{N}{r} \tag{30}$$

The typical site (0·0), Fig. 5 on the triangular lattice shares neighbours with eighteen other sites, and these are of three kinds represented by the sites (1·0), (2·0), (1·1) six times each. Application of equation (5) gives

$$E[s_{00}s_{00}] = \binom{N-7}{r}\Big/\binom{N}{r}, \quad E[s_{00}s_{20}] = \binom{N-13}{r}\Big/\binom{N}{r}$$

$$E[s_{00}s_{10}] = \binom{N-10}{r}\Big/\binom{N}{r}, \quad E[s_{00}s_{11}] = \binom{N-12}{r}\Big/\binom{N}{r} \tag{31}$$

and

$$E[s_{00}s_{ij}] = \binom{N-14}{r} \Big/ \binom{N}{r} \tag{32}$$

for sites (i, j) not sharing neighbours with (0·0). Equations (6) and (7) reduce to

$$E[S] = N\binom{N-7}{r} \Big/ \binom{N}{r} \tag{33}$$

$$E[S^2] = N\left[\binom{N-7}{r}+6\binom{N-10}{r}+6\binom{N-12}{r}\right.$$
$$\left.+6\binom{N-13}{r}+(N-19)\binom{N-14}{r}\right] \Big/ \binom{N}{r} \tag{34}$$

provided m and n both exceed 3.

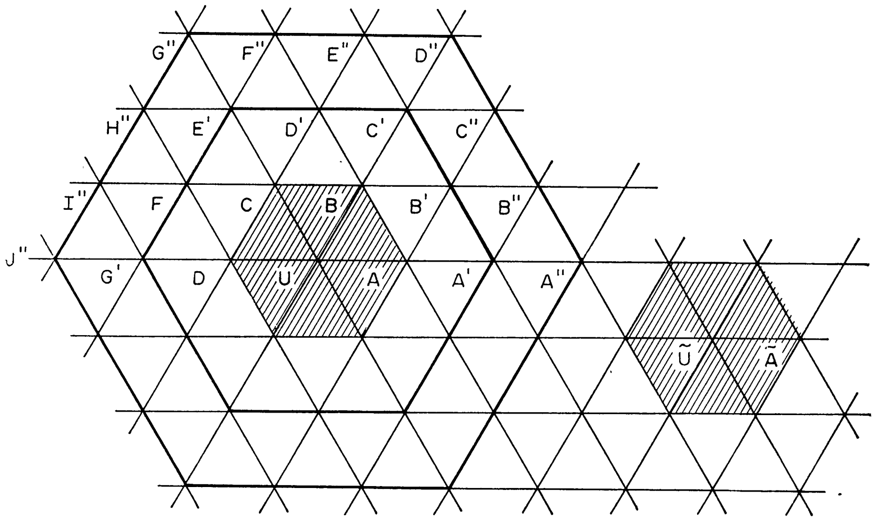

FIG. 6.

(ii) *Well-separated empty sites*

When the sites on three boundaries constructed around the empty site U are not adjacent to other empty sites, as in Fig. (6) the site U is called a well-separated empty site. When all the ε empty sites are well separated let $V_1, \ldots, V_\varepsilon$ represent the seven sites on or within the first boundary of each of the empty sites and let $\bar{V}$ be their complement containing $N-7\varepsilon$ sites. The expected value of the score variate s_{ij} is (cf. equation 4)

$$E[s_{ij}] = 1 \qquad \text{when } (i, j) \text{ is empty in one of } V_1, \ldots, V_\varepsilon$$
$$= \binom{N-\varepsilon-6}{r} \qquad \text{when } (i, j) \text{ is not empty in one of } V_1, \ldots, V_\varepsilon$$
$$= \binom{N-\varepsilon-7}{r} \Big/ \binom{N}{r} \qquad \text{when } (i, j) \text{ is in } \bar{V} \tag{35}$$

so that (cf. equation 6)

$$E[S] = \varepsilon+6\varepsilon\binom{N-\varepsilon-6}{r} \Big/ \binom{N-\varepsilon}{r}+(N-7\varepsilon)\binom{N-\varepsilon-7}{r} \Big/ \binom{N-\varepsilon}{r} \tag{36}$$

The derivation of $E[S^2]$ when there are ε empty sites, all well separated, involves counting changes in the index value α of all the pairs. This is carried out in the appendix. Modification of equation (34) accordingly then gives

$$\begin{aligned} E[S^2] = \varepsilon^2 + \Bigg[& (6\varepsilon+12\varepsilon^2)\binom{N-6-\varepsilon}{r} + (N-7\varepsilon-14\varepsilon^2+2\varepsilon N)\binom{N-7-\varepsilon}{r} \\ & +48\varepsilon\binom{N-9-\varepsilon}{r} + (6N-60\varepsilon)\binom{N-10-\varepsilon}{r} + 36\varepsilon\binom{N-11-\varepsilon}{r} \\ & +(6N-54\varepsilon+36\varepsilon^2)\binom{N-12-\varepsilon}{r} + (6N-126\varepsilon-84\varepsilon^2+12\varepsilon N)\binom{N-13-\varepsilon}{r} \\ & +(N-19+157\varepsilon+49\varepsilon^2-14\varepsilon N)\binom{N-14-\varepsilon}{r}\Bigg] \Bigg/ \binom{N-\varepsilon}{r} \end{aligned} \qquad (37)$$

for well-separated empty sites on an $m \times n$ triangular torus lattice with both m and n exceeding 3.

(iii) *Few empty sites and large lattices*

Using equation (17) for the case of few empty sites and large lattices, the first and second moment, cf. equations (6) and (7) reduce to

$$E[S] = Nq^7 \qquad (38)$$

$$E[S^2] = N[q^7+6q^{10}+6q^{12}+6q^{13}+(N-19)q^{14}] \qquad (39)$$

where $q = r/N$.

REDUCED FORMULAE FOR THE LINE LATTICE

The line lattice with empty sites better known in the context of time series with missing observations also merits derivation of reduced formulae for special cases.

(i) *No empty sites* ($\varepsilon = 0$)

For the sites (i) of the line lattice, torus wrapped,

$$E[s_i] = \binom{N-3}{r} \Bigg/ \binom{N}{r} \qquad (40)$$

The typical site (0·0) on the line lattice (Fig. 7) shares neighbours with five other sites

FIG. 7.

and these are of two types represented by the sites (1) and (2) twice each. Application of equation (5) gives

$$E[s_0 s_0] = \binom{N-3}{r} \Bigg/ \binom{N}{r}$$

$$E[s_0 s_1] = \binom{N-4}{r} \Bigg/ \binom{N}{r}$$

$$E[s_0 s_2] = \binom{N-5}{r} \Bigg/ \binom{N}{r} \qquad (41)$$

Equations (6) and (7) reduce to

$$E[S] = N\binom{N-3}{r}\Big/\binom{N}{r} \tag{42}$$

$$E[S^2] = N\left[\binom{N-3}{r}+2\binom{N-4}{r}+2\binom{N-5}{r}+(N-5)\binom{N-6}{r}\right]\Big/\binom{N}{r} \tag{43}$$

(ii) *Well-separated empty sites*

When the first three neighbour sites of an empty site U are not adjacent to other empty sites, as in Fig. 8, the site is called a well-separated empty site.

FIG. 8.

When all ε empty sites are well separated, i.e. at least five units apart, let $V_1, \ldots, V_\varepsilon$ represent corresponding empty sites and the two closest neighbours and let $\bar{V}$ be the complement of $V_1, \ldots, V_\varepsilon$ containing $N-3\varepsilon$ sites. The expected value of the score variate s_i is (cf. equation 4)

$$\begin{aligned} E[s_i] &= 1 && \text{when } (i) \text{ is empty in one of } V_1, \ldots, V_\varepsilon \\ &= \binom{N-\varepsilon-2}{r}\Big/\binom{N}{r} && \text{when } (i) \text{ is not empty in one of } V_1, \ldots, V_\varepsilon \\ &= \binom{N-\varepsilon-3}{r} && \text{when } i \text{ is in } \bar{V} \end{aligned} \tag{44}$$

so that (cf. equation 6)

$$E[S] = \varepsilon+2\varepsilon\binom{N-\varepsilon-2}{r}\Big/\binom{N}{r}+(N-3\varepsilon)\binom{N-\varepsilon-3}{r}\Big/\binom{N-\varepsilon}{r} \tag{45}$$

The derivation of $E[S^2]$ when there are ε empty sites, all well separated, is obtained by counting changes in the index value of all pairs once the pattern of empty sites is introduced (cf. Appendix) changing equation (43) accordingly gives

$$\begin{aligned} E[S^2] = \varepsilon^2+\Bigg[&4\varepsilon\binom{N-2-\varepsilon}{r}+(N+\varepsilon-4\varepsilon^2+2\varepsilon N)\binom{N-3-\varepsilon}{r} \\ &+(2N-6\varepsilon+4\varepsilon^2)\binom{N-4-\varepsilon}{r}+(2N-26\varepsilon-12\varepsilon^2+4\varepsilon N)\binom{N-5-\varepsilon}{r} \\ &+(N-5+27\varepsilon+11\varepsilon^2-6\varepsilon N)\binom{N-3-\varepsilon}{r}\Bigg]\Big/\binom{N-\varepsilon}{r} \end{aligned} \tag{44}$$

(iii) *Few empty sites and large lattices*

For large lattices the use of equation (17) reduces the first and second moment (cf. equations 42 and 43)

$$E[S] = Nq^3 \tag{45}$$

$$E[S^2] = N[q^3+2q^4+2q^5+(N-5)q^6] \tag{46}$$

where $r = r/N$.

DISCUSSION

The literature on clustering in two dimensions is meagre (cf. Wold 1965). The square lattice with four adjoining sites, has been investigated by Finney (1947), Moran (1947, 1948), and Freeman (1953), using the number of joins between successes and adjoining failures as a measure of clustering. The average number of failures adjoining a success, used as a measure in the present paper, and conceived initially in a context of contamination of neighbours, is more sensitive in that it distinguishes between the pattern $+--+$ and the tighter pattern $+-+$, whereas the number of success–failure joins equals 2 in each case ($+$ represents success, $-$failure). Clustering on the line lattice has been studied in terms of the number of success–failure joins (Wishart & Hirschfeld 1937), in terms of the average number of right-hand failures adjoining a success (Besson 1924), and in terms of length of runs in Bernoulli trials (Feller 1950; Brooks & Carruthers 1953).

By sub-dividing the square lattice into regularly packed square sub-lattices of nine sites each, the total number of isolated failures may be written as a sum of nine terms, each being itself a sum of independent score variates of points, in similar positions, on the sub-lattices. The distribution of each of these terms tends to normality for large lattices (Central limit theorem). The sum of these normal variates is normal, so that the distribution of the number of isolated sites tends to normality, and the proposed range of four standard deviations centred on the mean has in the limit fiducial probability close to 95%. Similar reasoning applies to all regular point lattices discussed in this paper.

It is of theoretical interest to note that in the absence of specific alternatives to randomness, attempts to compute the so-called power of a test for clustering are inappropriate, because the general class of alternatives cannot be parametrized. The purpose of the significance test in hand is to probe randomness as a piece of information for tracking down possible mechanisms causing the pattern (e.g. spreading of virus diseases by insect vectors).

SUMMARY

The clustering of attributes (successes) on the sites of regular point lattices, e.g. virus disease on trees in orchards, is tested here by comparing the observed number of failures adjoining the successes with a range of four standard deviations centred on the mean, assuming random incidence. The required mean and variance are expressed in terms of the size and type of the lattice, the number of times the attribute is present, and the pattern of 'empty' sites to which the attribute is not relevant (e.g. treeless sites). With a view to immediate practical application in common situations the general result is worked out explicitly for the square, the triangular and the linear lattice (time series) each for the case where there are (i) no empty sites, (ii) few empty sites on large lattices, (iii) any number of 'well-separated' empty sites.

REFERENCES

Besson, L. (1924). On the probability of rain. Mon. Weath. Rev. Washington D.C. 52, p. 308 (abridged translation of article in *C.r. hebd. Séanc. Acad. Sci., Paris.* **178,** 1743).

Brooks, C. E. P. & Carruthers, N. (1953). *Handbook of Statistical Methods in Meteorology.* Her Majesty's Stationery Office, London.

Feller, W. (1950). *An Introduction to Probability Theory and its Application.* John Wiley & Sons, London.

Finney, D. J. (1947). The significance of association in a square point lattice. *J. R. statist. Soc.* B, **9,** 99–103.

Freeman, J. H. (1953). Spread of diseases in a rectangular plantation with vacancies. *Biometrika*, **40**, 287–96.
Moran, P. A. P. (1947). Random association on a lattice. *Proc. Camb. phil. Soc.* **43**, 321–8.
Moran, P. A. P. (1948). The interpretation of statistical maps. *J. R. statist. Soc.* B, **10**, 243–51.
Wishart, J. & Hirschfeld, H. O. (1937). (cf. Moran 1947). *J. Lond. math. Soc.* **11.**
Wold, H. O. (1965). *Bibliography on Time series and Stochastic Processes.* Oliver & Boyd, Edinburgh and London.

(*Received September* 1970; *revision received* 27 *March* 1971)

APPENDIX I

Well-separated empty sites

When a torus contains empty sites, the product moment of the score variates for some pairs of sites is different from what it would be in the absence of empty sites. because the index value α has changed. To find the net gain for specified values of α one may for each empty site (i) say define a set V_i of sites, containing the adjoining sites and the site itself and let $\bar{V}$ stand for the complement of the sets $V_1, \ldots, V_\varepsilon$. When ε is the number of well-separated empty sites present, the pairs of sites for which there is a change in the index value α may be classified into ε categories V_iV_i, $\varepsilon(\varepsilon-1)/2$ categories V_iV_j and ε categories of the type $V_i\bar{V}$. Tables 1, 3, 5 and 7 record the typical pairs and their frequencies within each category, together with these index value α when the presence of empty sites is ignored and α_2 when it is taken into account, respectively, for the square lattice (8 adjoining sites), square lattice (four adjoining sites) the triangular lattice (six adjoining sites) and the line lattice (two adjoining sites). Tables 2, 4, 6 and 8, derived from 1, 3, 5 and 7 respectively record the net gain in the number of pairs having specified index value α. The net gains are used to modify the formula for the second moments from the case of no-empty sites, to the case of well-separated empty sites.

APPENDIX II

Application of the analysis to the assessment of clustering of Sirex *infestations in plantations of* Pinus radiata

The work undertaken in this paper arose from discussions with Professor L. Stubbs* on data being collected for an investigation of the spreading of a disease of peach, caused by the combined action of two viruses, where the mechanism of spread has not been determined. I gratefully acknowledge his insight and subsequent interest. I am also greatly indebted to Mr A. Rudra,† who has used the theory developed in this paper to test his data on *Sirex* infestation for clustering and made the results available for publication here. The circles on the 25×36 square lattice in Fig. 9 represent sixty-four *Sirex*-infested trees in a *Pinus radiata* plantation in the Yan Yean catchment, (Melbourne M.B.W.). By calling infested trees failures, the number of isolated failures can be used directly as a test statistic with the probability of failure computed as $64/540 = 0{\cdot}119$.

The expected number $E[S]$ of isolated failures is then found by application of equation (28) as $E[S] = 5409^5 = 0{\cdot}013$, whereas the observed numbers of isolated failures equals 4. If this difference is to be explained on a chance basis, it should not exceed a few standard

* Present address: Professor L. Stubbs, School of Agriculture, University of Melbourne.
† Present address: Mr A. Rudra, Forestry Department, University of Melbourne.

deviations. However, the standard deviations of the score S under random incidence is given by $\sqrt{E[S^2]-(E[S])^2}$, and when computed with the aid of equations (28) and (29) gives 0.114. The visual impression of clustering obtained by inspection of the pattern in Fig. 9 is therefore strongly confirmed statistically.

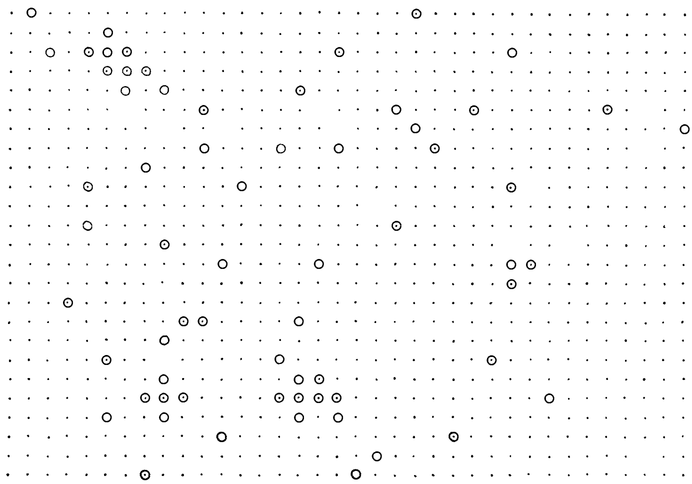

FIG 9. Circles indicating sixty-four *Sirex*-infested trees in a 25 × 36 square lattice plantation of *Pinus radiata*.

Table 1. *Typical pairs in the categories* (V_iV_i), (V_iV_j) *and* ($V_i\bar{V}$) *their frequencies and index values* α_1 *and* α_2, *without and with empty sites for the square lattice* (*eight adjoining sites*)

Category	Typical pair	α_1	α_2	Frequency	Typical pair	α_1	α_2	Frequency	Typical pair	α_1	α_2	Frequency
(V_iV_i), (ε times)	UU	9	–	–	AB	12	11	8	BB	9	8	4
	UA	12	8	4	AC	14	13	8	BC	12	11	8
	UB	14	8	4	AD	16	15	8	BD	15	14	8
	AU	12	8	4	AE	15	14	4	BE	16	15	8
	AA	9	8	4	BU	14	8	4	BF	17	16	4
(V_iV_j), ($\varepsilon(\varepsilon-1)/2$ times)	UŨ	18	–	2	AŨ	18	8	8	BŨ	18	8	8
	UÃ	18	8	8	AÃ	18	16	32	BÃ	18	16	32
	UB̃	18	8	8	AB̃	18	16	32	BB̃	18	16	32
(V_iV_i), (ε times)	UA′	15	9	8	AB″	16	15	16	BH	18	17	16
	UB	16	9	16	AC″	17	16	16	BI′	18	17	16
	UC′	17	9	8	AD″	18	17	16	BJ′	18	17	16
	UA″	18	9	8	AE″	18	17	16	BK′	18	17	8
	UB″	18	9	16	AF″	18	17	16	BD″	17	16	8
	UC″	18	9	16	AG″	18	17	16	BE″	16	15	16
	UD″	18	9	8	AH″	18	17	16	BF″	15	14	16
	Ux″	18	9	($2N-182-80$)	AI″	18	17	16	BG″	16	15	16
	AA′	12	11	8	AJ″	18	17	16	BH″	17	16	16
	AB′	14	13	16	AK″	18	17	16	BI″	18	17	16
	AC′	16	15	16	AL″	18	17	16	BJ″	18	17	16
	AD′	15	14	16	AM″	18	17	8	BK″	18	17	16
	AE′	16	15	16	Ax	18	17	($8N-72\varepsilon-320$)	BL″	18	17	16
	AF′	17	16	16	BC′	14	13	8	BM″	18	17	16
	AG′	18	17	16	BD′	12	11	16	BN″	18	17	16
	AH′	18	17	16	BE′	14	13	16	BO″	18	17	16
	AI′	18	17	8	BF′	17	16	16	BP″	18	17	8
	AA″	15	14	8	BG′	18	17	16	Bx	18	17	($8N-72\varepsilon-320$)

Table 2. *The number of changes from, and to, specified index values α and the number of net gains Δα on the square lattice (eight adjoining sites)*

Index value α	8	9	10	11	12	13
Changes from α	0	9ε	0	0	48ε	0
Changes to α	$8\varepsilon+16\varepsilon^2$	$-18\varepsilon^2+2\varepsilon N$	0	40ε	0	48ε
Net gain $\Delta\alpha$	$8\varepsilon+16\varepsilon^2$	$-9\varepsilon-18\varepsilon^2+2\varepsilon N$	0	40ε	-48ε	48ε
Index value α	14	15	16	17	18	
Changes from α	56ε	60ε	112ε	84ε	$-369\varepsilon-81\varepsilon^2+18\varepsilon N$	
Changes to α	52ε	96ε	$12\varepsilon+64\varepsilon^2$	$-256\varepsilon-144\varepsilon^2+16\varepsilon N$		
Net gain $\Delta\alpha$	-4ε	36ε	$-100\varepsilon+64\varepsilon^2$	$-340\varepsilon-144\varepsilon^2+16\varepsilon N$	$+369\varepsilon+81\varepsilon^2-18\varepsilon N$	

Table 3. *Typical pairs in the categories* (V_iV_i), (V_iV_j) *and* ($V_i\bar{V}$) *their frequencies and index values* α_1 *and* α_2 *without and with empty sites, for the square lattice (four adjoining sites)*

Category	Typical pair	α_1	α_2	Frequency	Typical pair	α_1	α_2	Frequency	Typical pair	α_1	α_2	Frequency
(V_iV_i),	UU	5	–	1	AU	8	4	4	AB	8	7	8
(ε times)	UA	8	4	4	AA	5	4	4	AC	9	8	4
(V_iV_j), ($\varepsilon(\varepsilon-1)/2$ times)	U$\tilde{\text{U}}$	10	–	2	U$\tilde{\text{A}}$	10	4	8	A$\tilde{\text{A}}$	10	8	32
$V(_i\bar{V})$,	UA′	9	5	2	UE″	10	5	4	AA″	9	8	8
(ε times)	UB′	8	5	4	UF″	10	5	4	AB″	8	7	16
	UC′	9	5	4	UG″	10	5	2	AC″	9	8	16
	UD′	8	5	4	Ux	10	5	$(2N-10\varepsilon-40)$	AD″	10	9	16
	UE′	9	5	2	AA′	8	7	8	AE″	10	9	16
	UA″	10	5	2	AB′	8	7	16	AF″	10	9	16
	UB″	10	5	4	AC′	10	9	16	AG″	10	9	8
	UC″	10	5	4	AD′	10	9	16	Ax	10	9	$(8N-40\varepsilon-160)$
	UD″	10	5	4	AE′	10	9	8				

Table 4. *The number of changes from, and to, specified index values α and the number of net gains Δα, on the square lattice (four adjoining sites)*

Index value α	4	5	6	7
Changes from α	0	5ε	0	0
Changes to α	$8\varepsilon+4\varepsilon^2$	$-10\varepsilon^2+2\varepsilon N$	0	48ε
Net gain $\Delta\alpha$	$8\varepsilon+4\varepsilon^2$	$-5\varepsilon-10\varepsilon^2+2\varepsilon N$	0	48ε
Index value α	8	9	10	
Changes from α	64ε	36ε	$-101\varepsilon-29\varepsilon^2+10\varepsilon N$	
Changes to α	$12\varepsilon+16\varepsilon^2$	$-64\varepsilon-40\varepsilon^2+8\varepsilon N$	0	
Net gain $\Delta\alpha$	$-52\varepsilon+16\varepsilon^2$	$-100\varepsilon-40\varepsilon^2+8\varepsilon N$	$101\varepsilon+29\varepsilon^2-10\varepsilon N$	

Table 5. *Typical pairs in the categories* (V_iV_i), (V_iV_j) *and* $(V_i\bar{V})$ *their frequencies and index values* α_1 *and* α_2 *without and with empty sites, for the triangular lattice (six adjoining sites)*

Category	Typical pair	α_1	α_2	Frequency	Typical pair	α_1	α_2	Frequency	Typical pair	α_1	α_2	Frequency
(V_iV_i),	UU	7	–	1	AA	7	6	6	AD	13	12	6
(ε times)	UA	10	6	6	AB	10	9	12				
	AU	10	6	6	AC	12	11	12				
(V_iV_j),	UŨ	14	–	2	AŨ	14	6	12				
($\varepsilon(\varepsilon-1)/2$ times)	UÃ	14	6	12	AÃ	14	12	72				
$(V_i\bar{V})$,	UA	13	7	2	UF″	14	7	11	AG′	14	13	12
(ε times)	UB′	12	7	4	UG″	14	7	4	AA″	14	13	12
	UC′	13	7	4	UH″	14	7	4	AB″	14	13	24
	UD′	12	7	4	UI″	14	7	4	AC″	14	13	24
	UE′	13	7	4	UJ″	14	7	2	AD″	14	13	24
	UF′	12	7	4	Ux	14	7	$(2N-14\varepsilon-60)$	AE″	14	13	24
	UG′	13	7	2	AA	10	9	12	AF″	14	13	24
	UA″	14	7	2	AB′	10	9	24	AG″	14	13	24
	UB″	14	7	4	AC′	12	11	24	AH″	14	13	24
	UC″	14	7	4	AD′	13	12	24	AI″	14	13	24
	UD″	14	7	4	AE′	14	13	24	AJ″	14	13	12
	UE″	14	7	4	AF′	14	13	24	Ax	14	13	$(12N-4\varepsilon-360)$

Table 6. *The number of changes from, and to, specified index values α and the number of net gains Δα on the triangular lattice (six adjoining sites)*

Index value α	6	7	8	9	10
Changes from α	0	7	0	0	60ε
Changes to α	$6\varepsilon+12\varepsilon^2$	$-14\varepsilon^2+2\varepsilon N$	0	48ε	0
Net gain $\Delta\alpha$	$6\varepsilon+12\varepsilon^2$	$-7\varepsilon-14\varepsilon^2+2\varepsilon N$	0	48ε	-60ε
Index value α	11	12	13	14	
Changes from α	0	48	42ε	$-157\varepsilon-49\varepsilon^2+14\varepsilon N$	
Changes to α	36ε	$-6\varepsilon+36\varepsilon^2$	$-84\varepsilon-84\varepsilon^2+12\varepsilon N$	0	
Net gain $\Delta\alpha$	36ε	$-54+36^2$	$-126\varepsilon-84\varepsilon^2+12\varepsilon N$	$157\varepsilon+49\varepsilon^2-14\varepsilon N$	

Table 7. *Typical pairs in the categories* (V_iV_i), (V_iV_j) *and* ($V_i\bar{V}$) *their frequencies and index values* α_1 *and* α_2 *without and with empty sites for the line lattice (two adjoining sites)*

Category	Typical pair	α_1	α_2	Frequency	Typical pair	α_1	α_2	Frequency	Typical pair	α_1	α_2	Frequency
(V_iV_i),	UU	3	–	1	AU	4	2	2				
(ε times,)	UA	4	2	2	AB	5	4	2				
(V_iV_j), ($\varepsilon(\varepsilon-1)/2$ times)	UU	6	–	2	AŨ	6	3	4	AÃ	6	4	8
$(V_i\bar{V})$,	UA′	5	3	4	Ux	6	3	$(2N-6\varepsilon-8)$	AA″	5	4	4
(ε times,)	UA″	6	3	4	AA′	4	3	4	Ax	6	5	$(4N-12\varepsilon-16)$

Table 8. *The number of changes from, and to, specified index value α and the number of net gains Δα for the line lattice (two adjoining sites)*

Index value α	2	3	4	5	6
Changes from α	0	ε	8ε	10ε	$-27\varepsilon-11\varepsilon^2+6\varepsilon N$
Changes to α	4ε	$2\varepsilon-4\varepsilon^2+2\varepsilon N$	$2\varepsilon+4\varepsilon^2$	$-16\varepsilon-12\varepsilon^2+4\varepsilon N$	0
Net gain $\Delta\alpha$	4ε	$\varepsilon-4\varepsilon^2+2\varepsilon N$	$-6\varepsilon+4\varepsilon^2$	$-26\varepsilon-12\varepsilon^2+4\varepsilon N$	$27\varepsilon+11\varepsilon^2-6\varepsilon N$