Chapter 1: Basic Structures and their Perceptual Relevance

The structures employed in this thesis, which constitute input to the visual system or to some computational device, are configurations of general interest to the study of human visual perception. Some fundamental patterns used in various explanations of visual perception are generated by transformations (and this, in turn, might provide some clue as to how they might be perceived by the visual system). In the absence of a clear taxonomy of structure employed as visual stimuli, this thesis explores an approach as it applies to the perception of structure generated by (1) the application of transformations to random arrays of points; (2) the application of transformations to regular arrays of points; and (3) the detection of structure in otherwise random arrays.

Glass patterns

A Glass pattern (Glass, 1969; Glass & Perez, 1973) is produced by taking a random array of dots, applying some transformation, and then superposing the transformed array onto the original. Exactly half the number of dots are mapped onto the other half by the transformation. Glass patterns are of special theoretical interest because, in order to perceive structure in such textures, it has previously seemed necessary to suppose the visual system must solve some form of massive *correspondence* problem (of working out which dots should be paired with which).[1] In this respect, the phenomenon resembles the integration of information from disparate images in stereopsis, the detection of mirror symmetry, and the perception of structure and motion from successive images in experiments employing random dot kinetograms. It is to be expected, therefore, that models devised to explain these various phenomena should be closely related.

As suggested by Figures 1.1(a) to 1.1(d), the human visual system is sensitive to Glass pattern structure generated by a variety of different transformations. These include translations in the vertical, horizontal, and diagonal directions (Wilson, Loffler, Wilkinson, & Thistlethwaite, 2001), as well as rotations, expansions from a centre, and spiral and hyperbolic transformations (Wilson & Wilkinson, 1998). In the case of dilations and rotations, human observers are sensitive to transformations by a constant amount as well as by amounts that increase with distance from the centre of dilation or rotation. Indeed, we may surmise that the visual system is sensitive to any structure generated by the uniform application of any suitably constrained transformation (or combination of transformations) to a well-populated random array of points.

---

[1] Cardinal and Kiper (2003) show that randomization of luminance of dots in Glass patterns had no effect on the pattern of their results. They found that neural mechanisms responsible for detection of Glass patterns are not restricted to luminance directions in color space, and have a broad tuning in color space. Moreover setting the luminance of dots equal to that of the background posed no problem for detection. Hence an argument based on the luminance distribution corresponding to short oriented lines in Glass patterns should not preclude the question of correspondence.

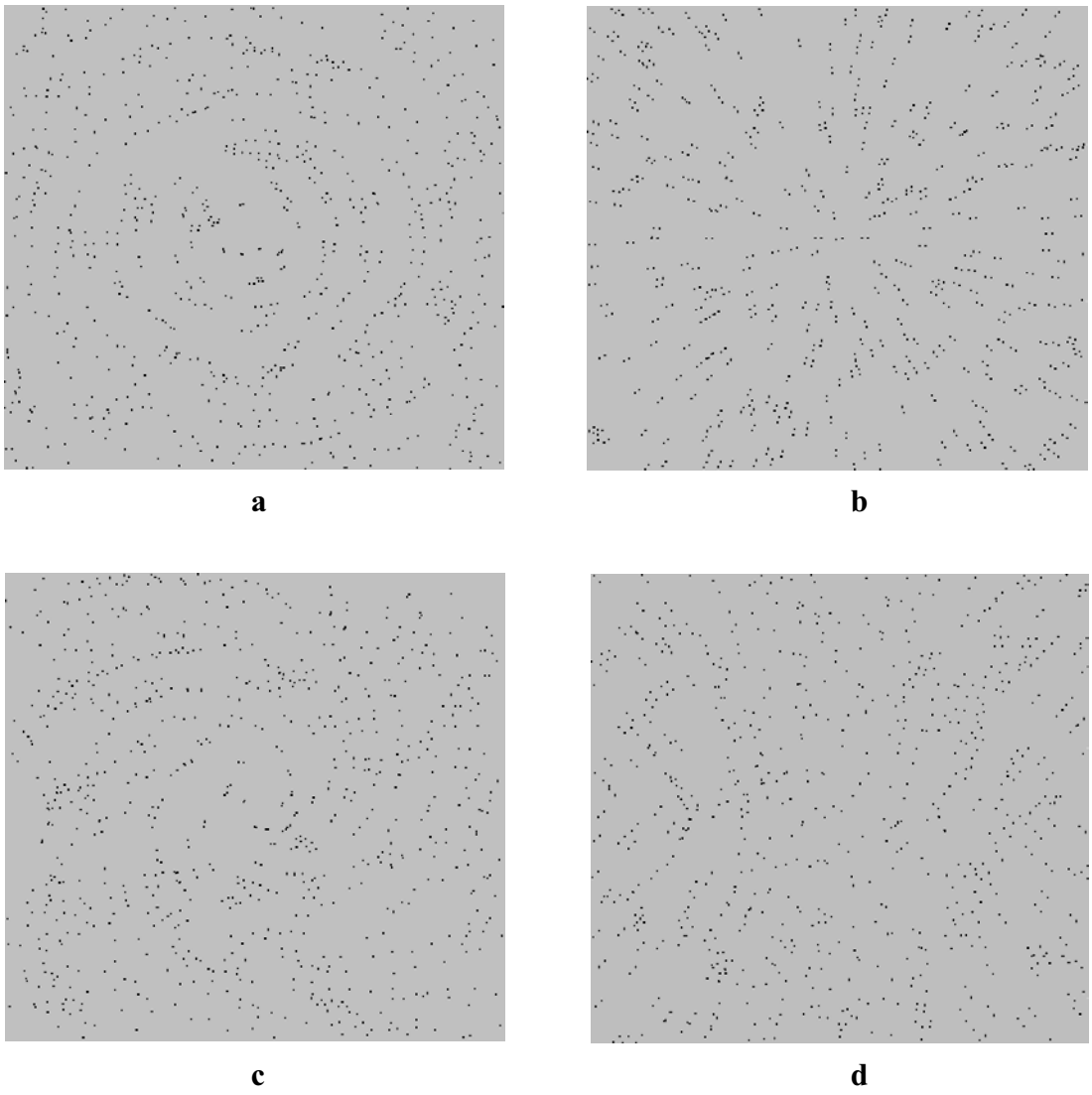


Figure 1.1: Glass patterns generated by means of (a) rotation, (b) dilation, (c) a screw or spiral transformation, and (d) a hyperbolic transformation. Each pattern contains 400 points.

Patterns with reflection (mirror) symmetry

The most common pattern with reflection symmetry employed by the thesis is symmetric about a central vertical line, as per Figure 1.2. (A good many forms, both natural and person-made, are correspondingly symmetric about a central vertical plane.) Such a pattern with reflection symmetry can be produced by taking a random array of dots, applying a depth rotation of 180^{o}, and then superposing the transformed array onto the original. It can be more simply produced by overlying an original set of dots with its horizontal distance complement (maximum possible x-coordinate distance minus actual x-coordinate distance, with y-coordinate distance unchanged). Again, exactly half the number of dots are mapped onto the other half by the transformation.

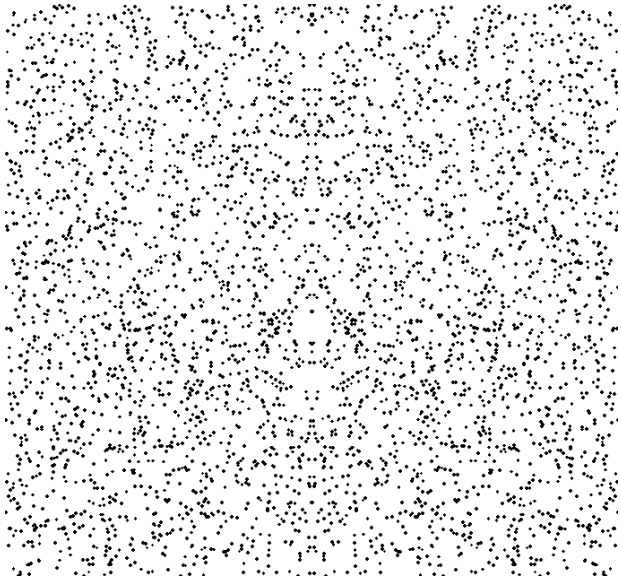


Figure 1.2: Pattern with reflection symmetry

Marroquin patterns

Marroquin (1976) investigated the perceptual effects that arise if we take a regular array (such as a square lattice of dots), apply some transformation, and then superpose the transformed array onto the original. So-called Marroquin patterns can be constructed by superposing a single transformed copy of the original array (or by superposing two, or more, different transformations on the original). Examples of a number of Marroquin patterns, generated using single rotational transformations, are shown in Figure 1.3.

Glass figures tend to evoke a clear perception of a single, stable structure, even when they are generated using combinations of two transformations (Prazdny, 1984). Although the structure suggests motion or 'static flow', the perception is of one single, consistent impression. In contrast, the perceptual effect of Marroquin patterns can be complex, with an unstable organization in which different symmetric structures compete and replace one another at random.[2] The experience is similar to that of viewing certain tiling patterns, both regular and aperiodic. It can also be similar to that of viewing a kaleidoscope.

Between them, Glass and Marroquin patterns exhaust a general partitioning of the class of structured arrays (into those where a transformation has been applied to a random or to a regular array). If the generative transformational approach provides a promising account of the perception of Glass figures, then we should ask if it provides a similarly good account of the perception of Marroquin patterns.

The question turns out to be not entirely straightforward, however, because the constrained nature of the original array in Marroquin patterns gives rise to more complex, competing, and unstable perceptions. It is therefore difficult to know how the perception of Marroquin patterns might best be measured. One possibility, explored by Wilson, Krupa, and Wilkinson (2000), is to examine the duration of time for which a particularly salient sub-structure is reported as 'visible', and to vary the position in the array at which such a sub-structure is likely to be seen. Although Wilson et al. report remarkably consistent results, three of the four observers tested were the authors themselves. Since a few minutes' inspection of Figure 1.3 will show a variety of alternative sub-structures that appear and

[2] This occurs at transformation magnitudes greater than those at which Glass pattern structure is generally discerned. Figure 1.3 (a) shows the effect of a transformation applied to a regular array that is typical in magnitude of that applied to a random array to produce a Glass pattern.

disappear spontaneously, it is far from certain that less committed and experienced observers would produce results of equal clarity and consistency.

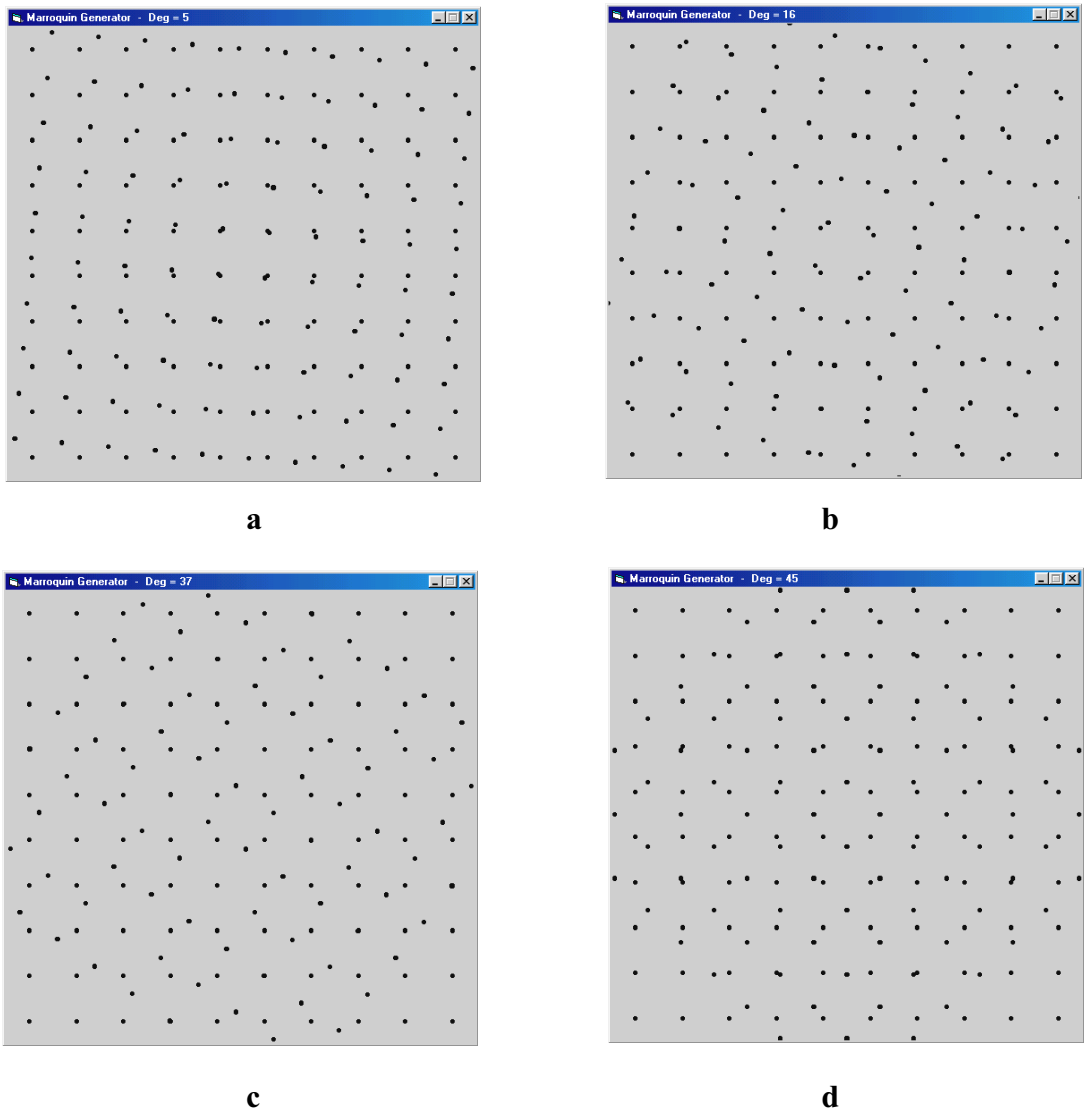


Figure 1.3: Examples of Marroquin patterns generated by taking a square lattice of dots, applying a rotational transformation, and then superposing the transformed array onto the original. Panel (a) shows a rotational offset of –5°, panel (b) –16°, panel (c) –37°, and panel (d) – 45°.

MacKay patterns

MacKay (1957a, 1957b, 1961) presented examples in which subjects fixate inspection patterns of alternating black and white stripes for a while, and then view a random test field. The elements in the test field then appear to have a streaming motion normal to the orientation of the stripes. This visual after-effect occurs with concentric, radial, and parallel, alternating patterns, to cite but a few. Figures 1.4(a) and (c) show examples of inspection stimuli, and Figure 1.4(b) shows an example of a test field, which is completely unstructured.

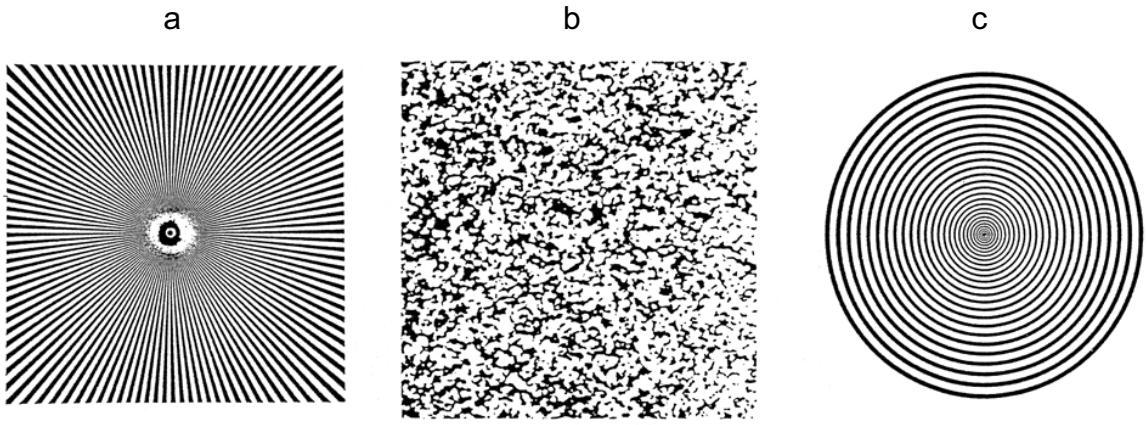


Figure 1.4: Examples of radial (a) and concentric (c) MacKay patterns, and an unstructured test field (b).

The after-effect can be explained by supposing that the streaming motion normal to the orientation of the stripes indicates continued activity of transforming units concerned with the perceptual organization of the inspection pattern. The perceptual organization of Figure 1.4(a), for example, is indicted by a rotating spoke stamped in increments. If the inspection pattern strongly activated units responsible for the complementary transformation—the one needed to form the pattern—this should be reflected in a bland test field. In the case of the radial MacKay pattern, rotations imposed on random dot kinematograms reflect the continued inertial activity of rotational transforming units.

Patterns with various degrees of image element dispersion

The human visual system is also sensitive to degrees of dispersion of image elements; see Preiss and Vickers (2005), Appendix A, for example. This can be shown by the use of patterns with varying degrees of clustering or regularity of image elements. Such patterns have elements that are either grouped together (mutually attracted) or spread apart (mutually repelled) more so than those of random distributions. Figure 1.5(a) shows an instance of a fixed number of elements in each cluster and Figure 1.5(b) shows an instance of a variable number of elements in each cluster (Poisson clustering). Figures 1.5(c) and 1.5(d) show instances of random and semi-regular patterns respectively.

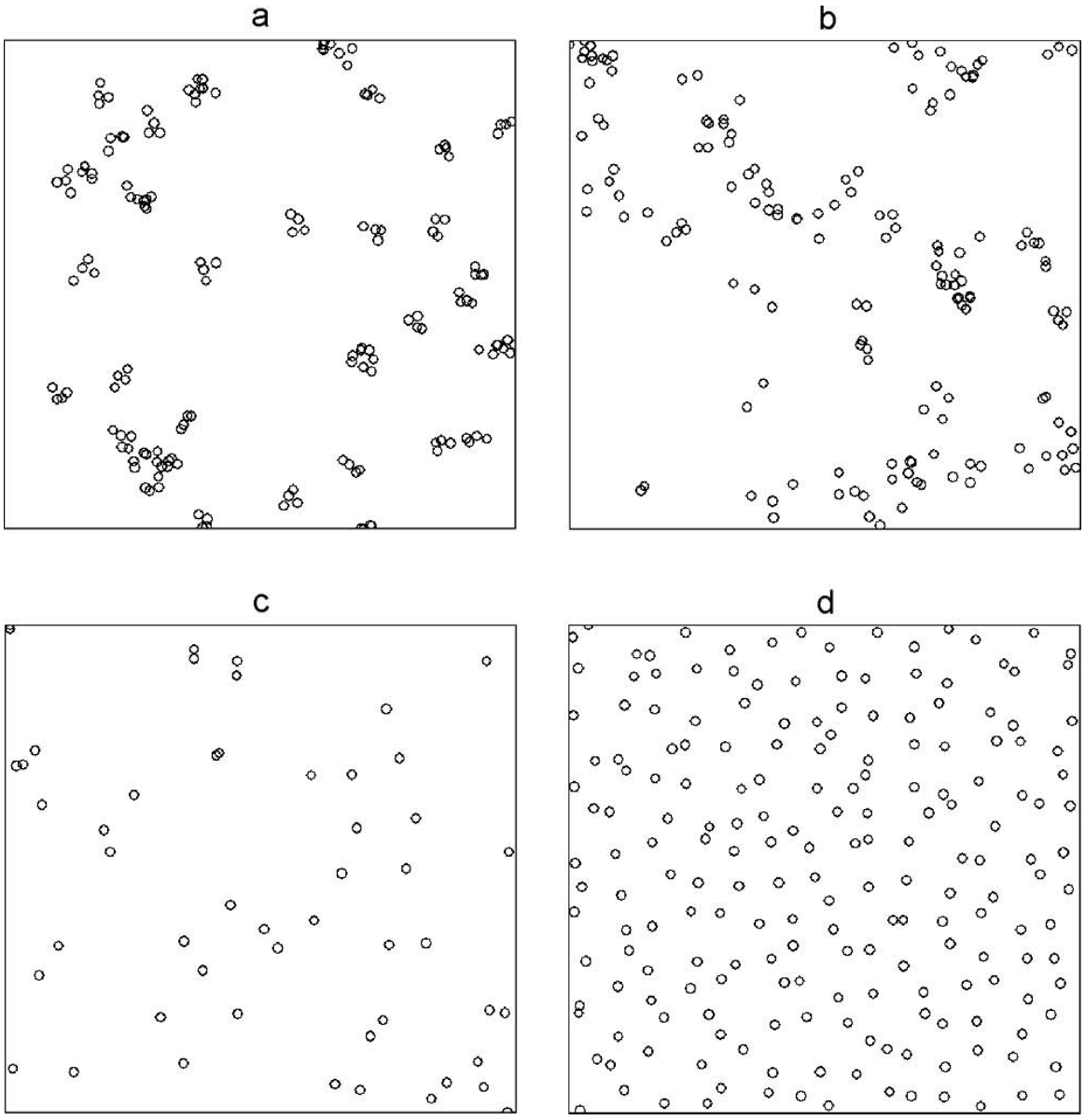


Figure 1.5: Examples of patterns with (a) fixed clustering, (b) Poisson clustering, (c) a random, and (d) a semi-regular distribution.

Travelling salesperson problem (TSP)

A typical formulation of the travelling salesperson problem involves travelling between cities (represented by points). Given a set of *N* randomly distributed cities and a cost incurred in moving between any two of them, devise a tour such that each city is visited once and the cost of the tour is minimum. This defines an 'open' tour: a 'closed' tour has the overriding provision that it finishes at the starting point. Figure 1.6(a) shows an open tour and Figure 1.6(b) shows a closed tour. A closed tour may be either the same as the corresponding open tour, but with the extra link, or it may be different to the corresponding open tour.

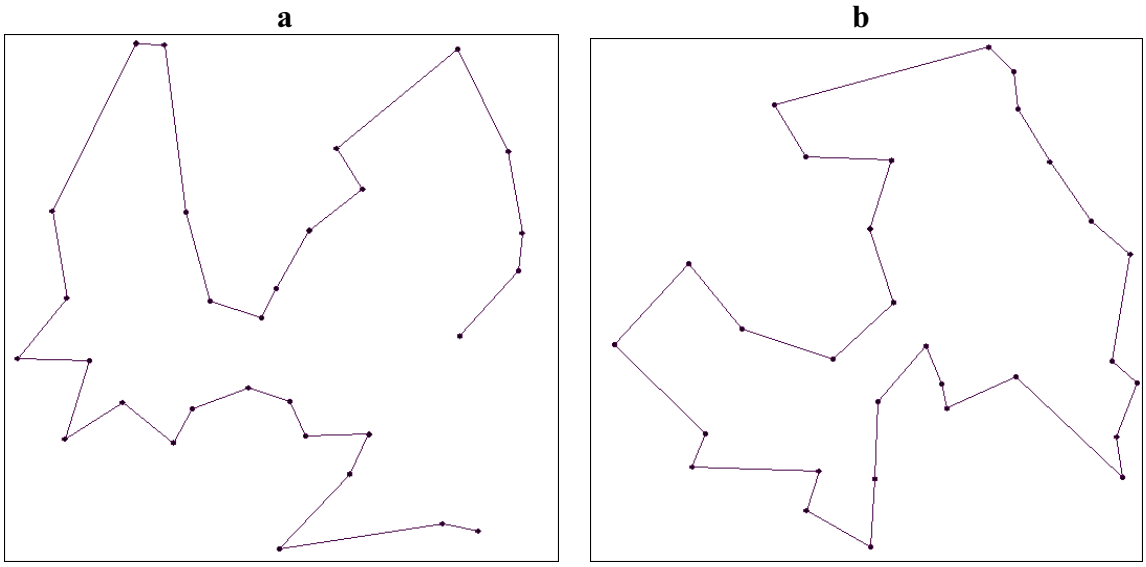


Figure 1.6: Open tour for 30 sites (a). Closed tour for 30 different sites (b).

The travelling salesperson problem belongs to a class of computationally intractable problems. Calculation of the optimal itinerary involves considering *N*! possible itineraries, which is not feasible for all but a small number of sites, and totally out of the question for 30 sites. The travelling salesperson problem is interesting because people often get quick, near optimal solutions to problems with numbers of sites that involve considerable computing time.

3-D objects and their 2-D projections

Retinal images are two-dimensional projections, usually from views of three-dimensional objects. Accordingly, a representation of these is addressed in the thesis. The example shown in Figure 1.7 shows a representation of a truncated pyramid, for which there is a dot (small circle) at each vertex. Adjoining lines do not figure in computations: their purpose is for figural clarification only.

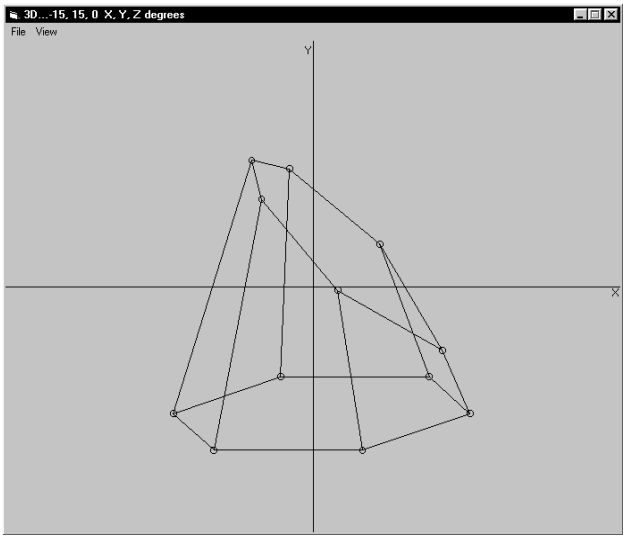


Figure 1.7: Truncated pyramid. (After Anton & Rorres, 1987.)

Fractal patterns

A fractal image is a geometric figure comprising patterns that repeat at increasingly smaller scales (see Oliver, 1992, for example). Repeating patterns exhibit self-similarity, which is a pivotal property of fractals. A figure has this property when a smaller structure is similar in shape to a larger structure, which in turn is similar to an even larger one, and so on.

Fractal curves are typical of those found in nature. The natural world can be modelled by fractal geometry, and any image can be produced by fractal geometry. A natural system is plausible from an evolutionary perspective, and by taking the cue from such a system it is possible to solve the inverse problem of finding the fractal encoding for a visual array.

This can be achieved by a system that uses a constrained sequence of transformations to work out how a fractal curve is generated and that is capable also of generating a copy of the curve. (The underlying idea is that what we experience as perceptual organization is the output of a generative process that applies multiple transformations to stimulus elements, thereby producing an output. The output can then be tuned and steered by matching it with the current visual input.)

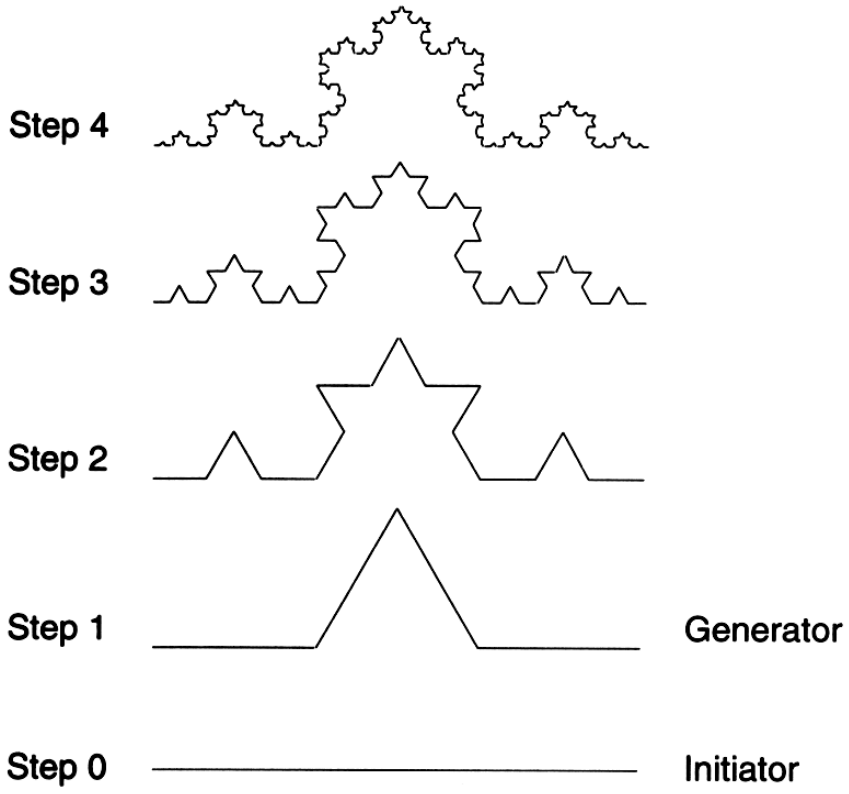


Figure 1.8: Successive steps in generating the Koch curve.

Generation of fractal stimuli

Generation of fractal stimuli may be illustrated with the help of a construction called the Koch curve. As shown in Figure 1.8, the Koch curve begins with a seed (initiator), which, in this case, is a simple straight line. Four copies are made of this line, each 1/3 of the length of the original. A new curve is then constructed by shifting and rotating the smaller copies. The process is then repeated on each of the resulting smaller straight line elements. At each

iteration, the number of copies is 4 and the contraction factor is 3^3. Each time the curve goes through the iteration process its length is multiplied by 4/3, with the result that its length after k iterations is $(4/3)^k$.

The nonlinear increase in length that occurs as the number of iterations is increased brings out the point that fractal curves and surfaces may be regarded as the output of nonlinear dynamical systems based on positive feedback. Each element of successive outputs is subjected to the same processes of contraction, multiplication and transformation. That is, a fractal object can be regarded as a temporal record of the operation of a dynamical process.

To explain why the generation of fractal images might have some relevance for understanding human visual perception, it is helpful to contrast a single-copy reduction machine (such as a conventional photocopier) with a multiple-copy reduction machine. In the single-copy version, a single reducing (similarity) transformation, applied repeatedly to a single copy, simply makes the copied image smaller and smaller. In contrast, under certain conditions, a set of transformations, applied to multiple reduced copies, generates a stable image. In this process, the nature of the initial elements is unimportant: what is important is the transformations. That is, all the information contained in the final image is effectively contained in the collage of transformations that is used to produce it. An important question, therefore, concerns the conditions under which such a process gives rise to a stable image.

These conditions are specified in the *contraction mapping principle*, a useful summary of which is given by Peitgen, Jürgens, and Saupe (1992a, pp. 263-277). According to the contraction mapping principle, the repeated iteration of a collage of contractive transformations, applied to a set of points, will give rise to a set of points that approaches closer and closer to an invariant set. Another way of saying this is that the limit image produced by the dynamical iterated function process is a stable attractor of the process.

Fractal objects generated for this thesis are represented by arrays of dots, as per Figure 1.9. Each dot is located at a vertex (end of a line) of the fractal object. (Hence the greater the number of iterations, or steps, involved in producing an object, the greater the number of dots required for its representation.)

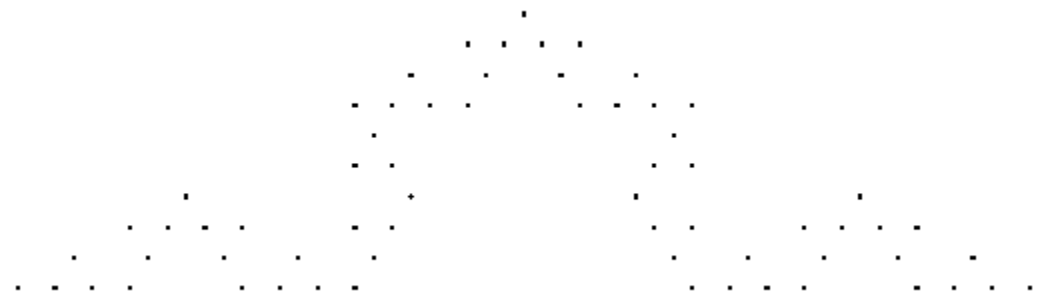


Figure 1.9: Koch curve, at step 3 in Figure 1.8, represented by points.

Regular structures embedded in noise

A task that is of practical importance in many contexts is the detection of a regular structure embedded in noise. The human visual system can perform such tasks at a level well beyond that of automatic detection programs. Indeed, this singular ability to detect figures embedded in distracting or noisy backgrounds is a facility that constitutes the focus of a number of neuropsychological tests of cognitive functioning. It is therefore of some interest to consider how the generative transformational approach might be applied to problems of this

[3] Sometimes ‘contraction factor’ is called ‘reduction factor’.

kind. Figure 1.10 shows an example of a regular array of image elements embedded in a random array of image elements, as used in this thesis.

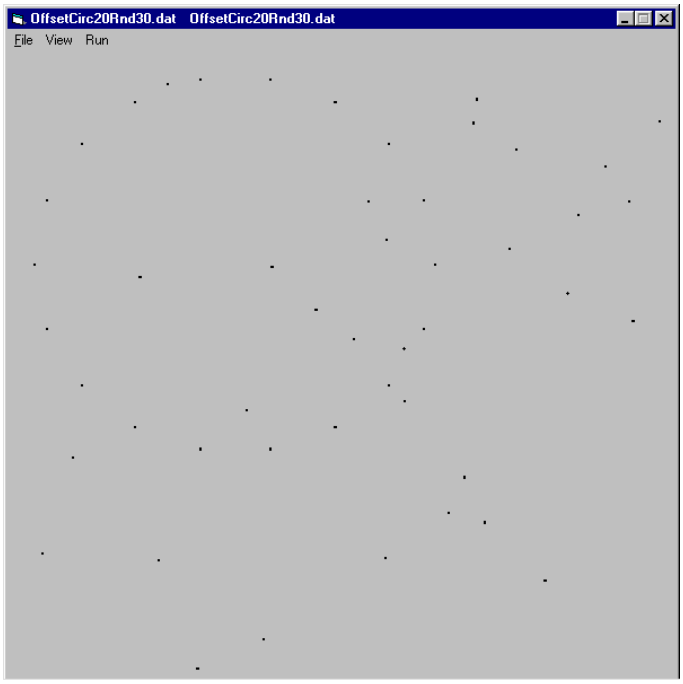


Figure 1.10: Example of a regular structure (circle in region of upper left quadrant) embedded in noise.

Representational momentum

Upon exposure to a sequence of systematic transformations of an object, such as displacements and/or rotations, there is a tendency to mentally continue the implied motion. Specifically, Freyd and Finke (1984, 1985) studied subjects' ability to recognize rectangles that had been subjected to a sequence of transformations (such as progressive planar rotations). A consequence of their findings was that subjects were most likely to recognize a previously seen rectangle when it was subsequently displayed in a position or orientation that corresponded to a continuation of the progressive sequence of transformations. By analogy with physical momentum, they dubbed this tendency 'representational momentum'. Representational momentum clearly motivates a transformational approach to visual perception.

Real and apparent motion

A powerful source of information about objects in our environment is provided by the systematic transformations of visual images that occur when either an observer or objects are in motion. For example, in a classic series of experiments, Wallach and O'Connell (1953) showed observers the projected shadows of wire-frame figures that were placed on a turntable between a point source of light and a translucent screen. When the turntable was stationary, the shadows were seen as two-dimensional patterns. However, as soon as the turntable was set in motion, the shadows were seen clearly and immediately as rigidly rotating three-dimensional objects.

Wallach and O'Connell's experiment investigated the perception of real motion, insofar as it involved the (continuous) transformation of an imaged object undergoing continuous movement. However, most studies of motion perception have been concerned with what is termed *apparent* motion (or *stroboscopic* motion when the frame rate is rapid, as

in cinema). Apparent motion is the term given to the perception of movement when successive, static stimuli are presented in discretely different positions and/or orientations. Because no ‘real’ continuous movement takes place, the perception of apparent motion is often referred to as ‘illusory’.

Temporal patterns

The production of temporal patterns in the thesis mainly involves successive, short period presentations of displays of dots, which have some consistent change of spatial layout (transformation) from one display to the next. Dot patterns for successive displays are often embedded in randomly distributed dots that differ from one display to the next. Such patterns are loosely dubbed ‘temporal’ because, in the extreme, there may be no spatial pattern in any one display, only a pattern that emerges over time, with successive displays.

Generally

A range of patterns is employed by the thesis; some not mentioned above because they will be more or less self-explanatory upon encounter. Beside spatial patterns, temporal patterns are considered. Besides two-dimensional patterns, three-dimensional patterns are considered. Besides Euclidean geometry, fractal geometry is considered. And many incidentals are also addressed along the way. The patterns were chosen because of their applicability to a range of questions relating to investigation of visual phenomena. Ultimately, structure in these quite different representations can be reconciled by just a handful of considerations relating to statistical regularity and symmetry in its various guises, in which perception of configuration becomes an optimising process. In short, it is proposed that analysis of distributions of distances and directions between image elements is sufficient for the detection of structure and motion, and can be used to select transformations that, when applied to image elements, generate a replica of the image. That is, information about relative positions of image elements can be used to select transformations that maximize generalized self-similarity within or between arrays.

Section 1: A Computational Approach to Low-level Grouping in Visual Perception

Chapter 2: Production of Structure by Joining Pairs of Points on a Surface

This section is chiefly concerned with developing some useful statistical approaches to questions in visual pattern detection, although it addresses related matters and other concerns in later chapters. First, however, an investigation of distance relationships between elements of visual stimuli is undertaken. And later on, a neural network approach to the perception of reflection symmetry is developed. Many observations established in Section 1 are assumed by accounts given in Section 2, hence the purpose of some accounts in section 1 need not be immediately evident.

Much of the statistical theory in Chapters 3, 4, and 5 is not original, but the implementation along with application to configurations of interest to the study of human visual perception is original. Much of the work in the other chapters of the section is original, and work anywhere that is not original is clearly referenced.

Brief summary of chapter

In accordance with the proximity principle of Gestalt psychology, a hierarchy of distance relationships ranging from nearest neighbour to Delaunay neighbour is discussed. The hierarchy is then put to work in an undertaking to shed some light on phenomena connected with a range of different visual stimuli. These include Glass patterns, the traveling salesperson problem, and MacKay patterns. Such application to disparate stimuli provides a test of versatility for methodology.

The work outlined in this chapter is original, with the exception of that indicated of others along with the graph theoretic structures.

Distance metric

Euclidean distance is an instance of the general concept of distance called the Minkowski p-metric. For two points located by *xy*-coordinates in the plane the generalized distance between them is

$$[(x_1 - x_2)^p + (y_1 - y_2)^p]^{1/p}, \qquad p = 1, 2, \ldots, \infty, \qquad (1)$$

and for Euclidean distance, p = 2. Many operations discussed in this thesis employ Euclidean distance in two dimensions.

Hierarchy of distance relationships

Nearest neighbours

As determined in this thesis, a nearest neighbour of an object belonging to a class of objects is another object of the same class at least Euclidean distance. Where objects are dots, they are often treated as (dimensionless) points. Generally, distance is measured between centres of objects. There are N nearest neighbours for any N objects. Figure 2.1 shows a plane surface containing $N = 30$ randomly located dots. The nearest neighbour distance for each dot is the distance from each dot to its nearest dot. Some of the nearest neighbour relationships are reciprocal. For example, the nearest neighbour of dot A is dot B and the nearest neighbour of dot B is dot A. Other nearest neighbour relationships are not symmetrical. For example, dot

C has dot B as its nearest neighbour, but dot B does not have dot C as its nearest neighbour. The former are sometimes called mutual or reflexive nearest neighbours.

The number of nearest neighbour edges, or links, for different dot configurations is some *variable* value less than *N,* since edges do not differentiate the two kinds of nearest neighbour distance. Note that a point may be the nearest neighbour to several points or to no points at all; but, of course, the point must *have* a nearest neighbour.

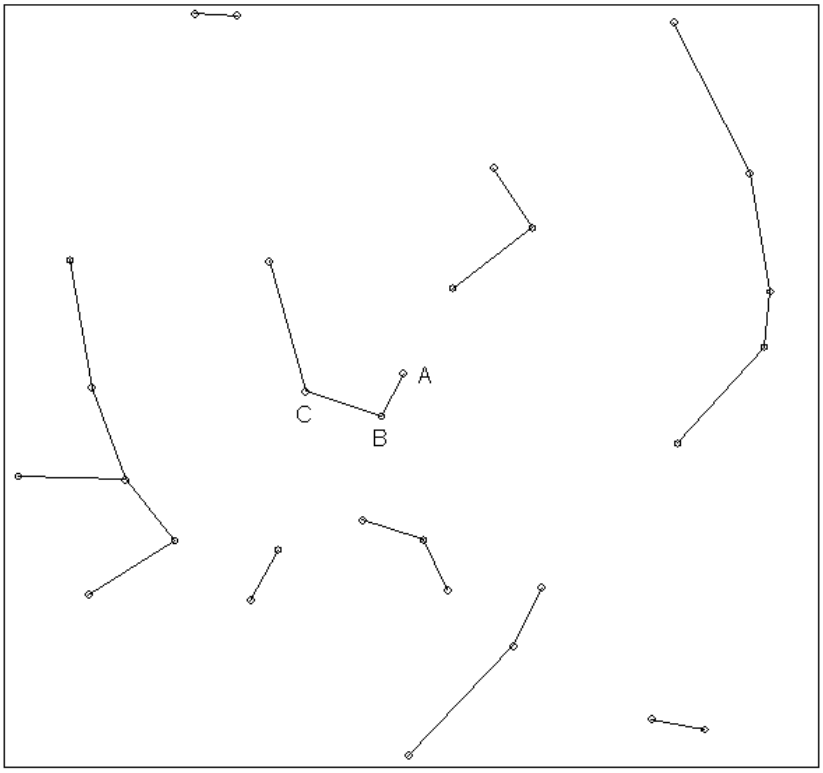


Figure 2.1: Nearest neighbour links.

Minimum spanning tree

A minimum spanning tree is a least cost network of edges that links all *N* dots in an array, and contains no loops (Aho, Hopcroft, & Ullman, 1987; Kingston, 1990). As determined in the thesis, 'cost' equates to 'Euclidean distance'. There are $N - 1$ minimum spanning tree edges for any *N* points. Figure 2.2 shows the minimum spanning tree for the same 30 dots used in Figure 2.1. Dots are linked such that a path can be traced from any dot to any other dot, and the total length of all the edges is minimal.

A minimum spanning tree can be constructed by sorting all the possible dot-to-dot distances into ascending order, and then by linking in order of increasing distance from the smallest. Edges that would form loops are discounted, and the process continues until $N - 1$ edges are completed.

A somewhat more elegant (and intuitive) construction of a minimum spanning tree is first to link all nearest neighbours. The linked pairs, now considered as elements, are then linked to their nearest neighbour elements by closest members of pairs. The process continues until there is just one element that links all dots. This method for finding the minimum spanning tree is a recursion upon nearest neighbours.

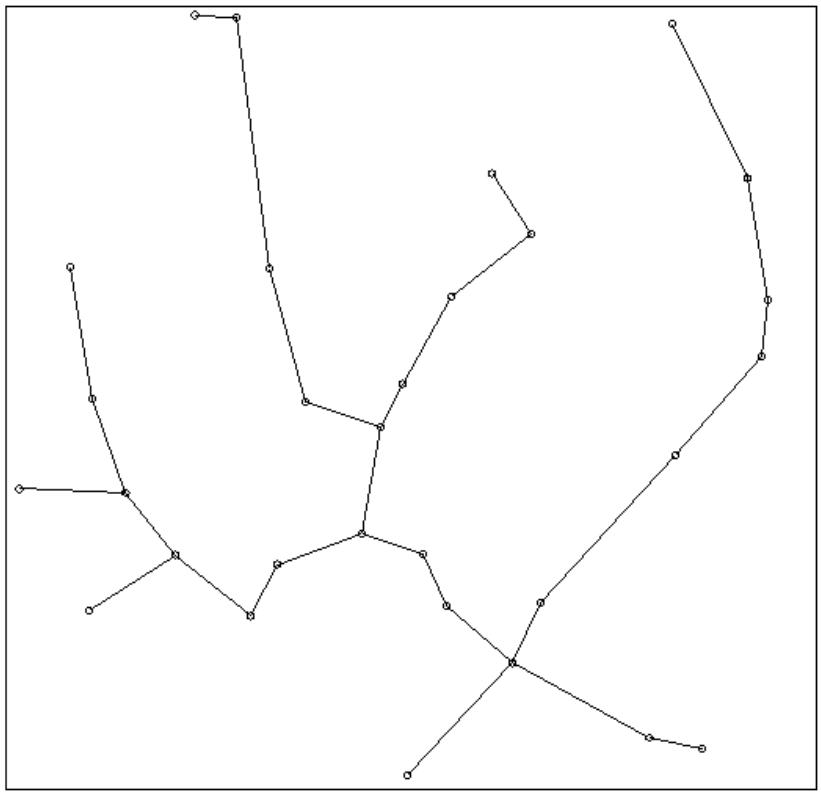


Figure 2.2: Minimum spanning tree.

Relative neighbourhood graph

Two dots in a collection of dots are relative neighbours if they are at least as close to each other as they are to any other dot (Boots, n.d.). That is to say, two dots are relative neighbours provided no other dot is closer to both of them than the distance between them. The number of relative neighbourhood edges for some *N* points is bound from below by the number of edges in the corresponding minimum spanning tree, and from above by the number of edges in the corresponding Gabriel Graph, which is defined below.

Consider the same 30 randomly located dots. Pairs are linked if no dots are located in the interior of the intersection of two circles, as per dots A and B in Figure 2.3. Each circle is centred on a member of the pair, and the common radius is equal to intra-pair distance. To take another example from all possible pairs, C and D are not linked because a dot is located in the interior of the intersection of their two circles. The relative neighbourhood graph includes the minimum spanning tree.

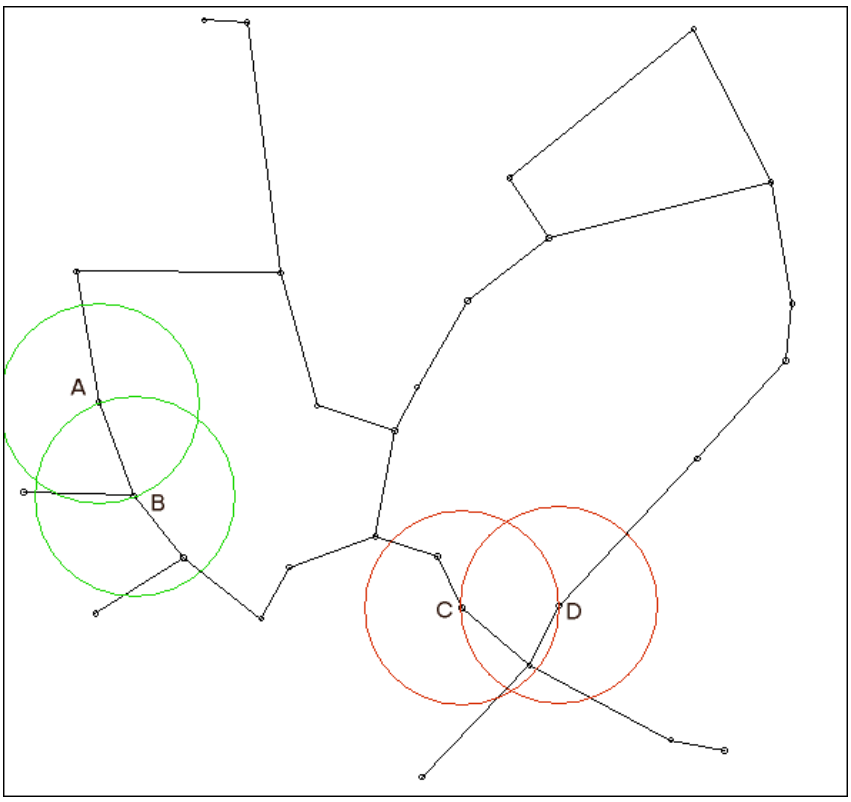


Figure 2.3: Relative neighbourhood graph. Pairs of dots are linked if no dots are located in the interior of the intersection of two circles, as per dots A and B. Dots C and D are not linked because a dot is located in the interior of the intersection of their two circles.

Convex hull

The convex hull of a collection of points is the smallest convex set that contains the points (O'Rourke, 1994). Consider the same 30 randomly located dots. Imagine an elastic band stretched beyond the bound of the dots, then placed on the form and allowed to shrink onto outlying dots. The set of such outer dots, as shown in Figure 2.4, identifies the convex hull.

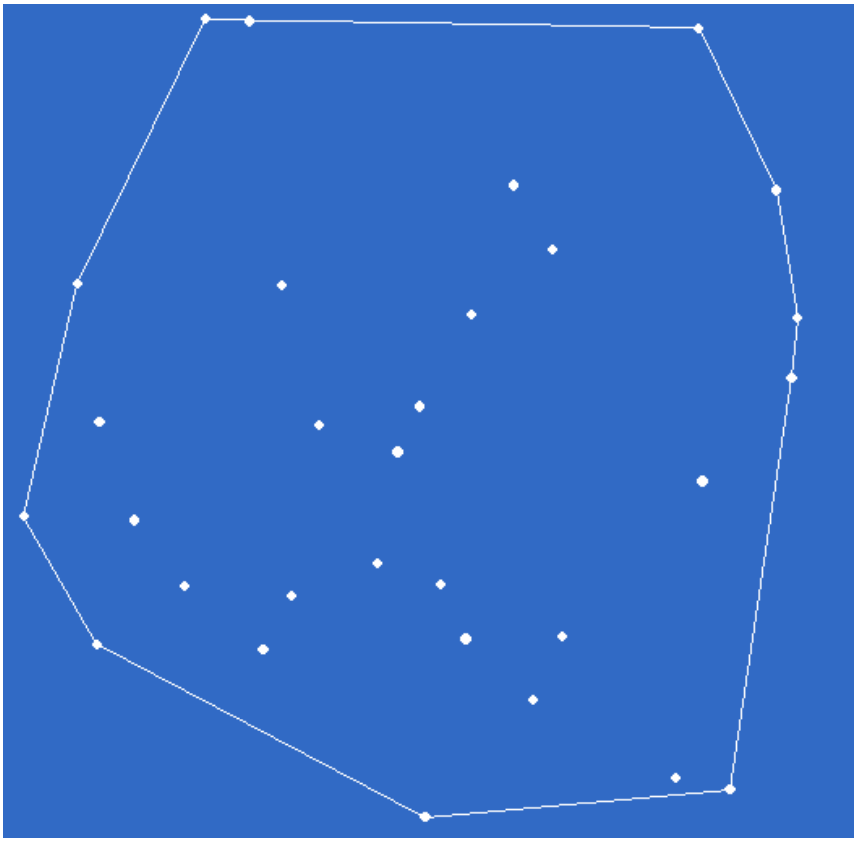


Figure 2.4: Those dots joined by the lines form the convex Hull for the 30 randomly distributed dots.

Voronoi partitioning

For the purpose of image segmentation, Voronoi partitioning is here considered in two dimensions only. For a collection of points in the plane, the Voronoi diagram partitions the plane such that each point is associated with the region closest to it (O'Rourke, 1994). The Voronoi diagram is an optimal device that embodies all information concerning proximity, or neighbourliness. By Euler's formula for polyhedra, which states that faces + vertices = edges + 2, a Voronoi diagram has at most $3N - 6$ edges. (See O'Rourke, 1994, pp. 118-121, for a discussion and proof.)

Considering the same 30 dots, there is a partition into areas *within* which any position is closest to some dot. Along the boundary segments of the areas any position is equidistant from two neighbouring dots, and at points of intersection of the segments—ends of segments—any position is equidistant from three neighbouring dots.

Imagine circles centred on respective points. Let the circles increase in radii at a uniform rate until their circumferences touch, and then intersect, in some succession. Straight lines that delineate the respective spans of intersection lengthen until they meet. Points interior to the convex hull become enclosed by the edges of polygons, while points on the convex hull become partially enclosed. (The latter are located on the diagram between divergent lines extending beyond the edges of the form. See Figure 2.5.) Once the last pair of lines is joined, the surface within and around the convex hull is completely partitioned.

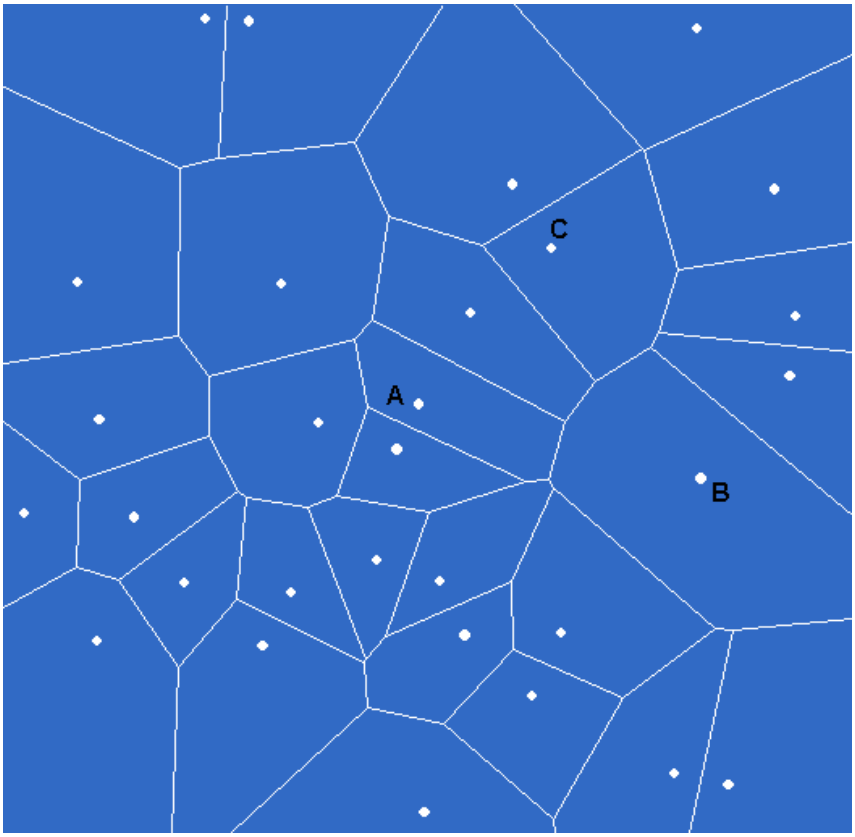


Figure 2.5: Voronoi diagram for the same 30 randomly distributed dots.

The dots shown in Figure 2.5 are said to be in 'general position' (Preparata & Shamos, 1985), which is the common situation, and for which points of intersection involving three boundary segments are equidistant from three neighbouring dots. For dots in general position, no three dots are collinear and no four dots are co-circular. Put another way: for dots in general position three or more dots are not collinear and four or more dots are not co-circular.

A situation in which three dots are collinear or four dots are co-circular is considered 'degenerate', but nevertheless valid. For degenerate situations, inconsequential but lengthy details must be added to statements and proofs of theorems concerning properties of Voronoi diagrams; hence utilization of points in general position by O'Rourke (1994), and Preparata and Shamos (1985), for such statements and proofs.

An example of a degenerate situation is that of rectangularly located dots. Points of intersection involve four boundary segments, equidistant from four dots. Hence analyses involving Voronoi related treatments may show meaningful differences between regularly placed elements and more randomly placed elements, which might be helpful in discriminating degrees of regularity from randomness.

The Voronoi neighbours, other than the nearest neighbour, of a point may be more distant than non-neighbours. For example, point A in Figure 2.5 is more distant from one of its neighbours at point B than it is from a non-neighbour at point C.

Delaunay triangulation

The Delaunay triangulation and Voronoi diagram are dual structures. Any Delaunay triangle edge connects two sites only if the sites' Voronoi regions have a common edge. Since Delaunay triangle edges have one-to-one correspondences with Voronoi polygon edges, there are at most $3N - 6$ Delaunay edges. A Delaunay triangulation contains the same information as the corresponding Voronoi diagram: one is simply a transformed version of the other.

Nonetheless, as demonstrated in Figures 2.6 and 2.7, what it means to be a neighbour is shown more clearly by Delaunay triangle edges. Neighbours of any dot are those dots directly linked to it. Delaunay triangulation embodies all information concerning adjacency.

Delaunay triangulation reflects the optimal quality of Voronoi partitioning. It is the angle optimal triangulation that maximizes minimum angles. In the two dimensional case, Delaunay triangulation is the maximal planar subdivision; which is the subdivision in the plane such that no edge connecting two points can be added without crossing at least one other edge, hence destroying planarity. (In order to *cross*—as opposed to *intersect*—an edge in the plane, another edge needs to traverse outside the plane, in the third dimension, thus destroying planarity.)

Consider the same 30 dots, along with their Voronoi partitioning, as shown in Figure 2.6. Observe that a neighbour of a dot is another dot with a common boundary. If a line is drawn from a dot to one of its neighbours, the line is either perpendicularly bisected by a Voronoi boundary or passes through two or more Voronoi boundaries, but is perpendicularly bisected by the extension of the common boundary. (Observe that the outer edges form the convex hull.) A dot can be either directly across the 'fence' from its neighbour or indirectly across the 'fence', which corresponds to two kinds of adjacency.

Numbers of Delaunay neighbours belonging to each dot for different dot configurations can have different distributions. For dots randomly distributed in the plane there are, on average, six neighbours to a dot and four of these are, on average, directly adjacent neighbours. (This can be derived from Boots, n.d., pp. 6 & 26, and was confirmed in the thesis by experimentation).

For direct neighbour edges, circles imposed upon the edges as diameters (hence passing through the dots at each end) do not include any other dot, while for indirect neighbour edges they do. Moreover in a triangle constituted of direct edges, any position is nearer to one or another of the triangle's vertex dots than to any other dot. This is not necessarily the case for a triangle constituted of a mix of edges or all indirect edges. A graph constituted of all direct neighbour edges is called a Gabriel Graph (Boots, n.d.). Moreover it is possible for the whole of a Delaunay triangulation to consist of direct neighbour edges (again with the upper limit of $3N - 6$ for number of such edges).

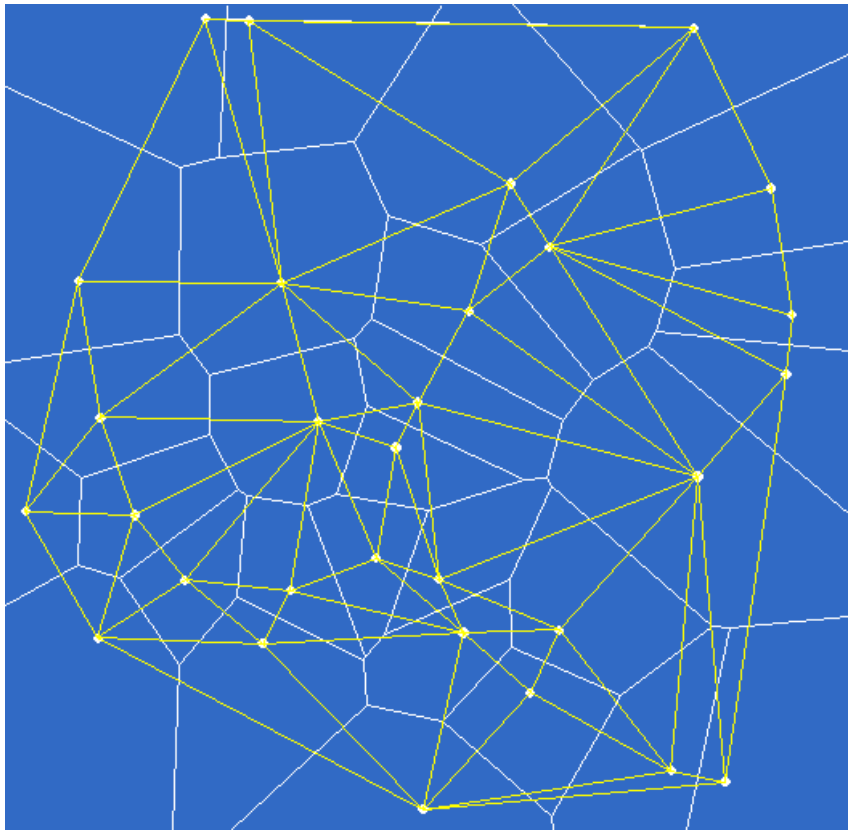


Figure 2.6: Voronoi diagram (white) with Delaunay triangulation (yellow) superposed, for the same 30 randomly distributed dots.

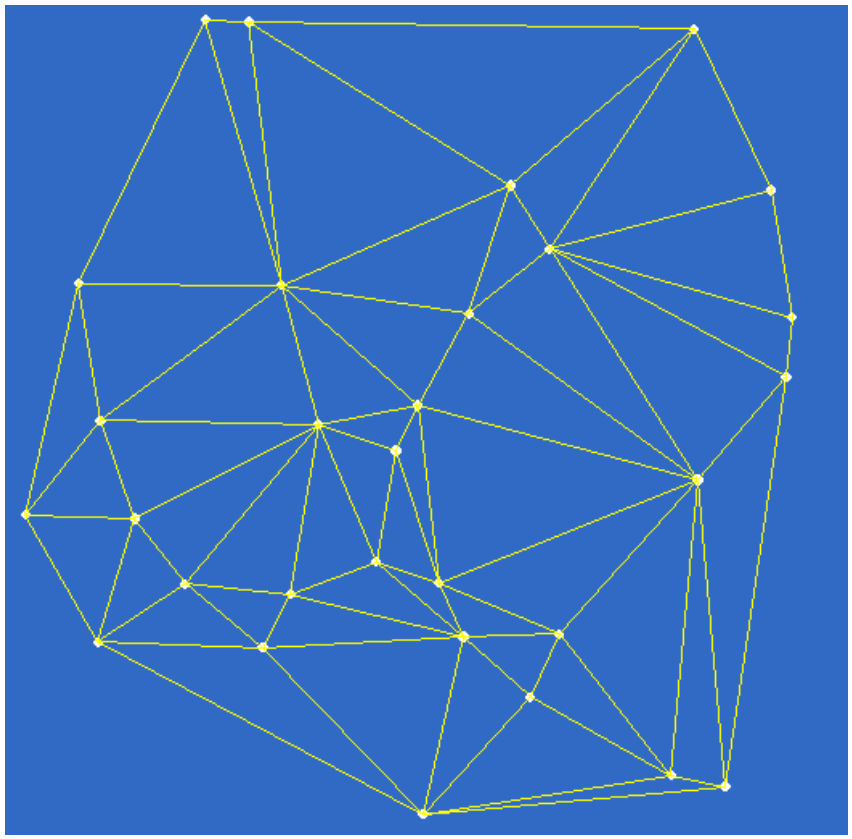


Figure 2.7: Delaunay triangulation for the same 30 randomly distributed dots.

The partitioning just described is for plane surfaces, and uses Euclidean distances. However, different situations could demand different metrics for which this kind of partitioning can be applicable. Choosing p = 1 in expression (1), page 12, might be appropriate where points of interest are located on a grid, such as in many cities for example. Distance between the points is then measured along the lines of the grid. Other situations in flat space can demand different numbers of coordinates while p = 2, and curved spaces can demand different numbers of coordinates and other values for p. Moreover Voronoi partitioning can be defined using other interval concepts such as those associated with time or cost measures (Boots, n.d.).

Comment on computational complexity

Nearest neighbour, and minimum spanning tree algorithms are traditionally of order N^2, and relative neighbourhood is of order N^3. To find the set of nearest neighbours for 100 dots, for example, the distance from each dot to every other dot needs to be calculated, which takes $100 \times 100 = 10{,}000$ operations.[1] Delaunay triangulation via an appropriate algorithm, Fortune's plane-sweep algorithm for example (O'Rourke, 1994), has order $N \log(N)$ which, for 100 dots, takes $100 \times 2 = 200$ operations. Hence all subsets can be computed in fewer operations by first performing Delaunay triangulation.

Hierarchy

Nearest neighbour edges are a subset of minimum spanning edges, which, in turn, are a subset of relative neighbourhood edges, which, in turn, are a subset of direct neighbour edges. Nearest neighbour edges are constituted of the shortest edge from within each set of Delaunay neighbour edges.

So far in the hierarchy of point-to-point linking relationships based on distance, the relevant linking factors are nearest neighbour, minimum spanning, relative neighbourhood, direct Delaunay and all Delaunay, both direct and indirect. These, of course, are all embedded in the superset containing all point-to-point edges. The set containing all point-to-point edges out to some cut-off distance, where the upper limit for the cut-off distance can include all possible point-to-point edges, will be addressed later on.

Applications to pattern detection and visual perception

Pattern structure

Structure in point patterns, as perceived by the human visual system, might be captured best by one or other of the relational hierarchy. Sometimes nearest neighbour or minimum spanning tree captures structure and Delaunay triangulation does not do so well, and other times the reverse is true. Quite often, in cases like these, the relative neighbourhood graph does well in either situation. It sits roughly in the middle of the hierarchy and spans the gap in number of edges from minimum spanning tree to Delaunay triangulation.

[1] There are actually $N^2 - N$ calculations, because the distance from any dot to itself is not calculated. However, this situation must still be tested by an operation.

Figure 2.8 shows an example of a pattern with hierarchical organization. Grouping is on the basis of dots first, then on the basis of rough crosses as units. These units, not dots, then group to form triangular units, and these units, not crosses, then group to form pairs of triangles.

Figure 2.9 shows structure captured by the neighbourhood hierarchy, excluding the relative neighbourhood graph. And Figure 2.10 shows structure captured by the relative neighbourhood graph.

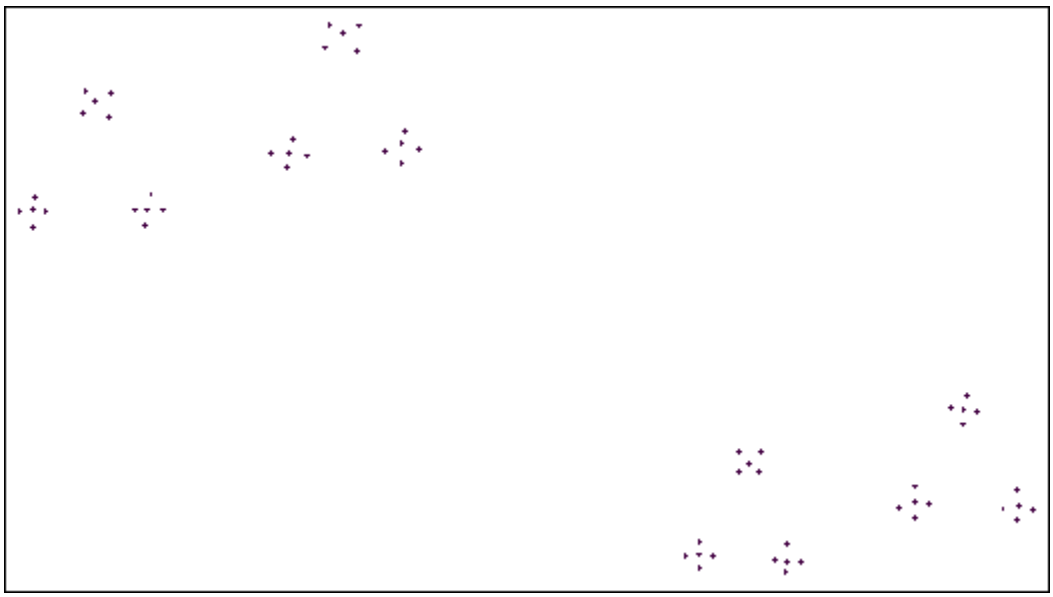


Figure 2.8: Example of hierarchical organization, in which there are groups of groups of groups. (Triangular formation after Skiena, "Minimum Spanning Tree", 1997.)

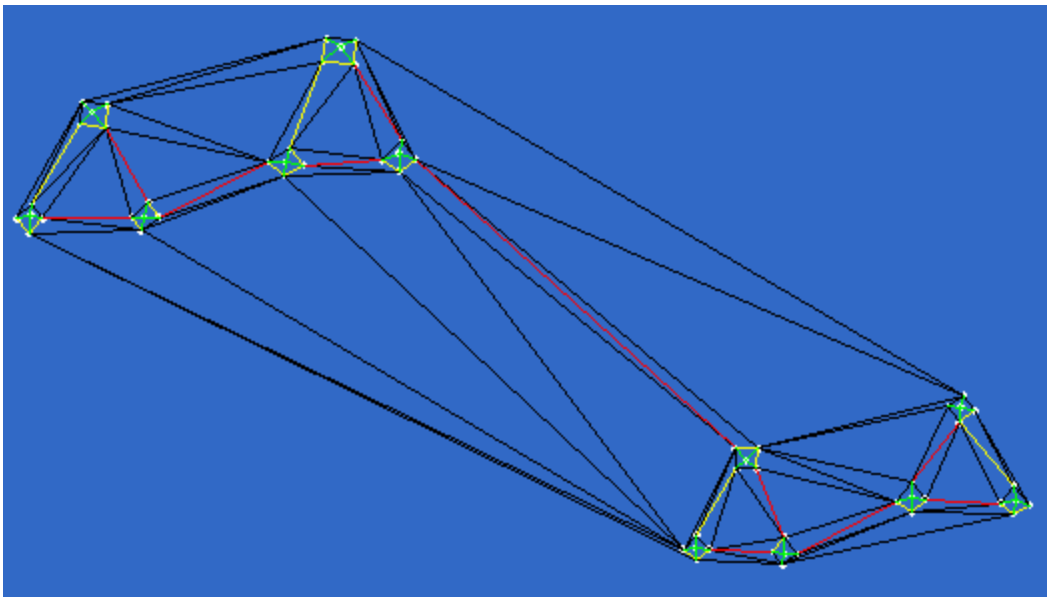


Figure 2.9: Structure captured by nearest neighbour edges (green), remaining minimum spanning tree (red), remaining direct Delaunay edges (yellow), and indirect Delaunay edges (black).

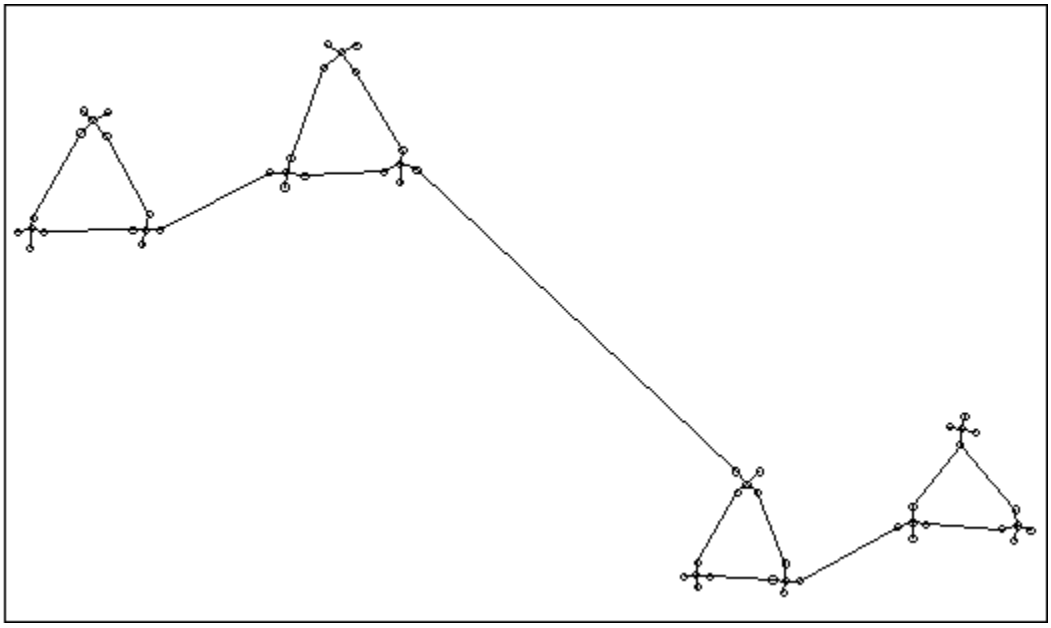


Figure 2.10: Structure captured by the relative neighbourhood graph.

Figure 2.9 shows that nearest neighbour relationships clearly capture the most basic structure identified by crosses. Some remaining Delaunay edges, both direct and indirect, capture the polygonal structure enclosing the crosses at this level. Then some remaining minimum spanning edges followed by remaining direct Delaunay edges capture the triangular structure.[2] The relative neighbourhood graph, shown in Figure 2.10, makes a reasonable compromise in capturing structure otherwise seen across the neighbourhood hierarchy.

Application of Delaunay neighbours to Glass patterns

Glass patterns (Glass, 1969; Glass & Perez, 1973) are generated by taking a uniformly random spatial distribution of dots and superposing a geometrically transformed copy. Resulting textures are seen as having clear structure, consisting of dot pairs locally orientated along the direction of the transformation used to generate them. A transformation cannot be too large, otherwise structure becomes degraded and ultimately disappears. Because perception of structure in Glass patterns is immediate, and does not require attentive scrutiny, they provide an important device in the study of pattern detection.

Dry, Vickers, Lee, and Hughes (2004) found that correlations between the frequencies of alternative responses by subjects to translation Glass patterns, oriented in one or the other of two diagonal orientations and dosed with various proportions of dots distributed randomly and uniformly (noise dots), showed that participants based their judgments on first and second nearest neighbour links that fell within a ±15° tolerance of the positive or negative diagonals.

For the following analysis of Glass patterns—the first of several throughout the thesis—Delaunay neighbours for each dot of a display are ordered in ascending distance, from which Delaunay neighbour loadings are calculated. ‘Delaunay neighbour loading’ is indicated as the distribution of numbers of correspondingly ranked Delaunay neighbours across the display summed over each dot, whose orientations fall within a constrained range of angles about transformation direction(s), and which are proportional to numbers of all correspondingly ranked Delaunay neighbours for the display (which is trivially the same as

[2] It was hoped that indirect Delaunay edges would capture the pair-wise triangular structure, but maybe there is not a strong case here; unless the total number of black edges within pairs compared with those between pairs is considered significant. Indeed the same might be said for red and yellow edges.

numbers of angles for all correspondingly ranked Delaunay neighbours). Put simply, Delaunay neighbour loading is the proportion of Delaunay neighbours (by rank) that lie in a limited range about transformation direction(s).

The analysis shows that transformation pairs belonging to relatively sparsely populated Glass patterns load close to 100%, in a constrained range of angles about transformation direction(s), upon nearest neighbours, and somewhat less on otherwise close neighbours. As Glass pattern density is increased, the loading peak progressively decreases and occurs progressively more on more distant neighbours. See Figures 2.11 to 2.16, for which the constrained range is $\pm 10^{o}$ about transformation direction(s). The magenta lines show Glass pattern loadings on Delaunay neighbours (alternatively called natural neighbours) and the black lines show corresponding loadings for the same number of noise dots.

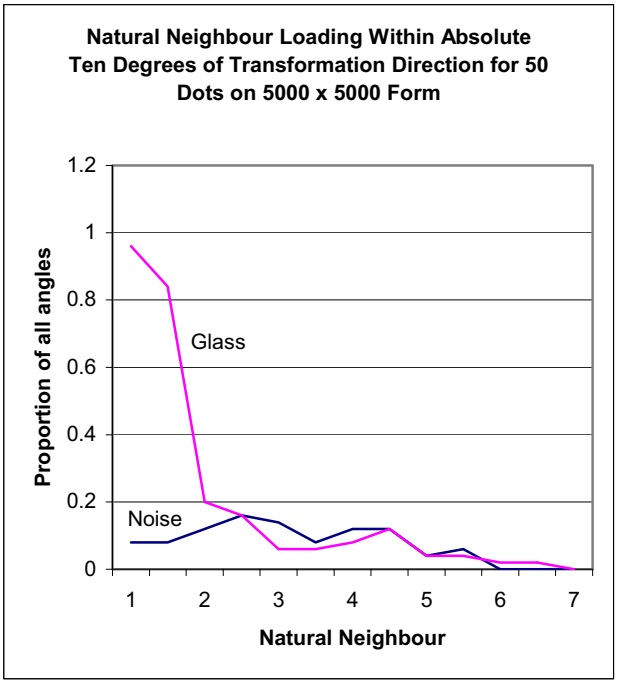


Figure 2.11: Natural neighbour loadings for a Glass pattern with 50 dots and the same number of noise dots. The graph shows the extents to which the different neighbours load on the transformation direction(s) $\pm 10^{o}$, belonging to the Glass pattern.

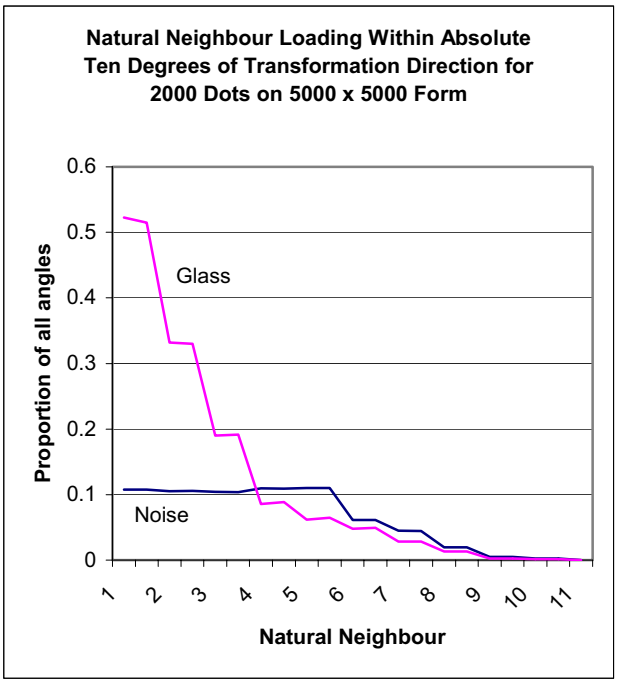


Figure 2.12: Natural neighbour loadings for a Glass pattern with 2,000 dots and the same number of noise dots. The graph shows the extents to which the different neighbours load on the transformation direction(s) ±10°, belonging to the Glass pattern.

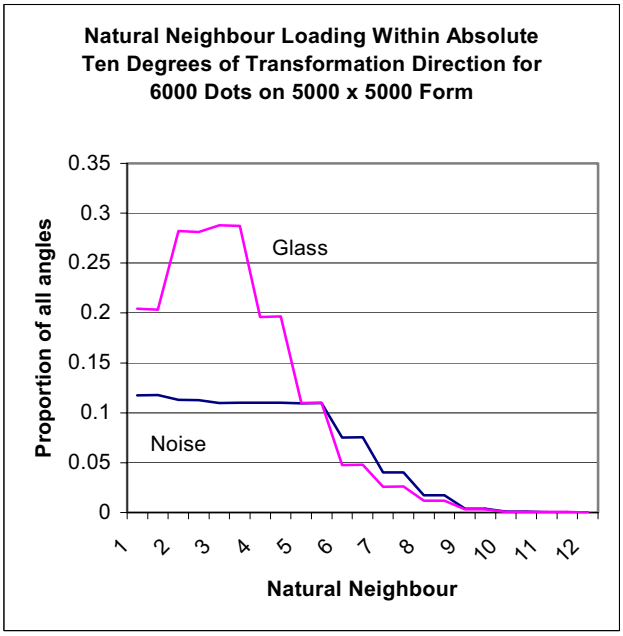


Figure 2.13: Natural neighbour loadings for a Glass pattern with 6,000 dots and the same number of noise dots. The graph shows the extents to which the different neighbours load on the transformation direction(s) ±10°, belonging to the Glass pattern.

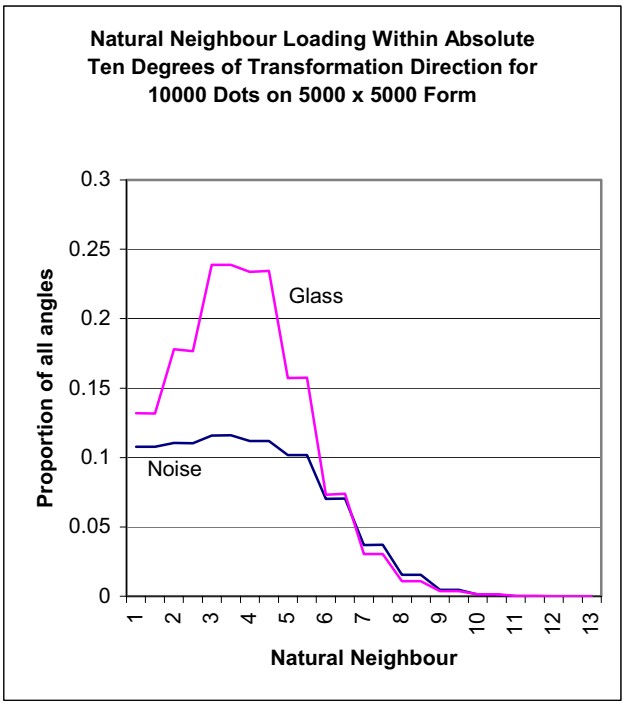


Figure 2.14: Natural neighbour loadings for a Glass pattern with 10,000 dots and the same number of noise dots. The graph shows the extents to which the different neighbours load on the transformation direction(s) $\pm 10^{o}$, belonging to the Glass pattern.

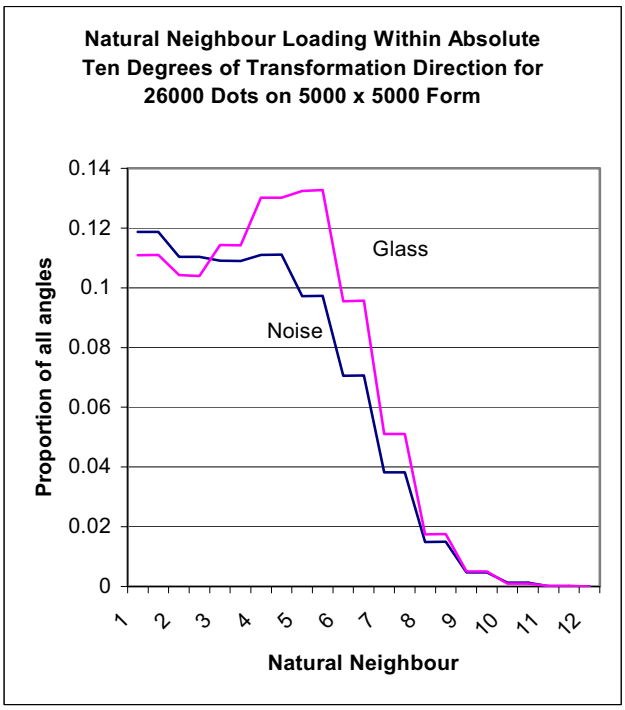


Figure 2.15: Natural neighbour loadings for a Glass pattern with 26,000 dots and the same number of noise dots. The graph shows the extents to which the different neighbours load on the transformation direction(s) $\pm 10^{o}$, belonging to the Glass pattern.

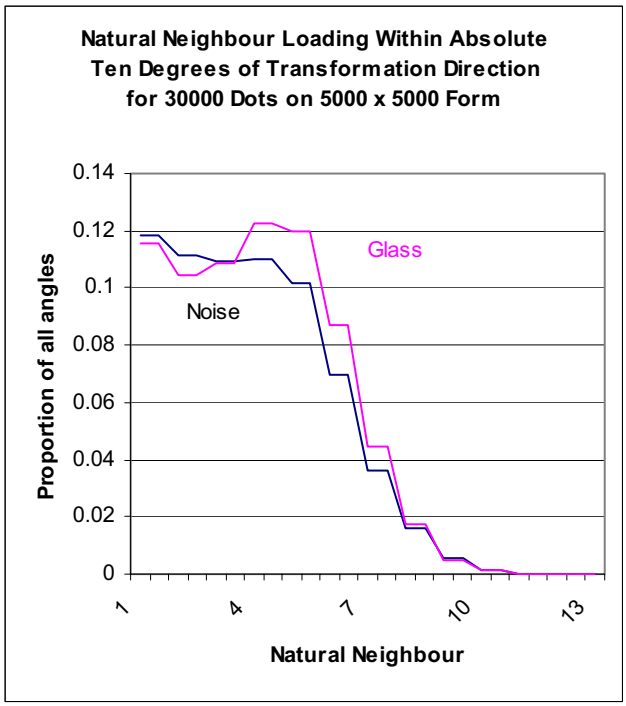


Figure 2.16: Natural neighbour loadings for a Glass pattern with 30,000 dots and the same number of noise dots. The graph shows the extents to which the different neighbours load on the transformation direction(s) $\pm 10^{o}$, belonging to the Glass pattern.

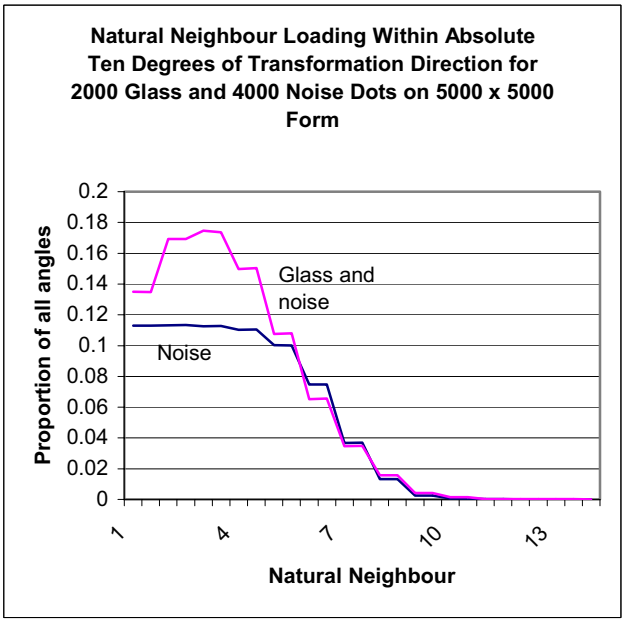


Figure 2.17: Natural neighbour loadings for a Glass pattern with 2,000 dots and 4,000 noise dots. The graph shows the extents to which the different neighbours load on the transformation direction(s) $\pm 10^{o}$, belonging to the Glass pattern.

The ranges of neighbours are ultimately exhausted as the loading peak descends towards the loading for noise, and this appears to be more or less coincident with transfiguring Glass effect.[3] That is, when the background colouration shows evidence of the transformation effect as the foreground, or pattern, colouration becomes fused. A concomitant reduction in loading peak is also seen in cases of Glass patterns dosed with various levels of noise. Figure 2.17 results from a Glass pattern of 2,000 dots dosed with 4,000 noise dots. Compare Figure 2.17 with Figure 2.12, the latter resulting from a Glass pattern of 2,000 dots dosed with no noise dots. The degree of degradation of the pattern dosed with noise is obvious.

Note the progressive *y*-axis scale reduction of the graphs, which indicates a much greater range of loading peak reduction than is apparent by just looking at the curves in sequence. Note also that the distribution for the number of Delaunay neighbours of each dot belonging to a random display peaks at six, and neighbours are effectively exhausted at around eleven. (The probability of getting 14 neighbours is estimated at just .000005.) The distribution, shown in Figure 2.18, is consistent with the range of values evident in the graphs for natural neighbour loadings.

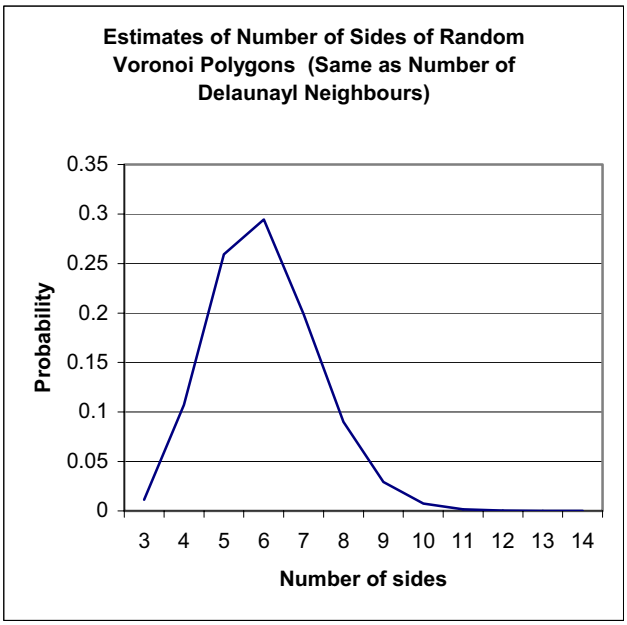


Figure 2.18: Estimates of number of sides of Voronoi polygons, hence Delaunay neighbour edges, resulting from uniformly random distributions of points. Data source: Hinde and Miles, 1980, Table III, p. 215.

As Glass pattern density is increased the classic transformation effect is replaced by streaking in orientations along transformation direction(s). Further increase sees a transfiguration. Only when the foreground becomes substantially fused does the transformation effect fall away, and then it still does not disappear entirely. However, the density ranges at which these effects show prominence are dependent on dot size in relation to form area.

[3] Transfiguring: altering, changing, changing by reversal, reversing, turning.

Distance to neighbours

The above analyses deal with dimensionless points, i.e. centres of dots. Making dots smaller so as to approximate dimensionless points simply increases the density ranges at which effects show prominence, and this occurs independent of any analysis outcome. Effects of increasing Glass pattern density are obvious from the graphs, but at high densities some neighbour loading remains.

A device that can be used—at least for now—is a dot size term: hence distance to neighbours needs to be taken from nearest point on circumference to nearest point on circumference of dots or, more generally, boundary to boundary of objects. This rids the analysis of all those (diminishing) distances that would otherwise maintain some directional neighbour relativity after fusion. Figure 2.19 shows boundary-to-boundary natural neighbour loadings for a Glass pattern with 30,000 dots of radius 12 screen units. Figures 2.20 and 2.21 show the Glass pattern and corresponding noise respectively.

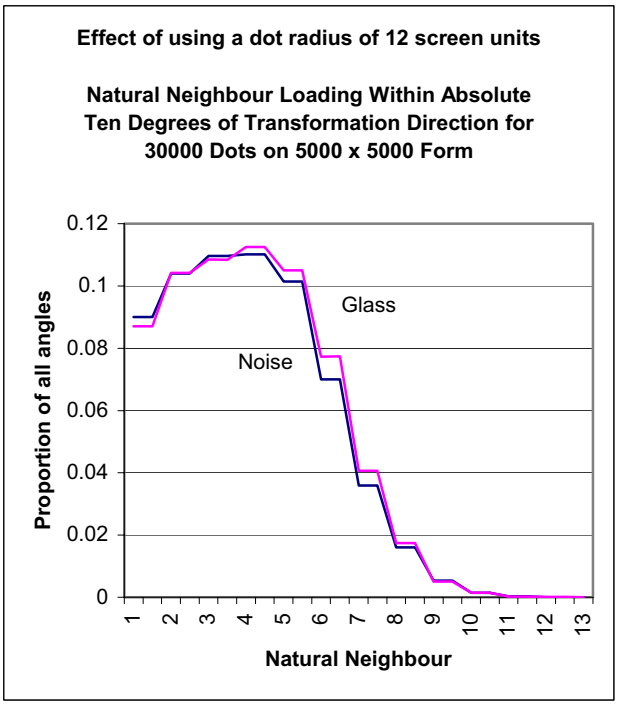


Figure 2.19: Natural neighbour loadings for a Glass pattern with 30,000 dots and the same number of noise dots. The graph shows the extents to which the different neighbours load on orientations along the transformation direction(s) $\pm 10^{o}$, belonging to the Glass pattern. Closest neighbours begin to null first.

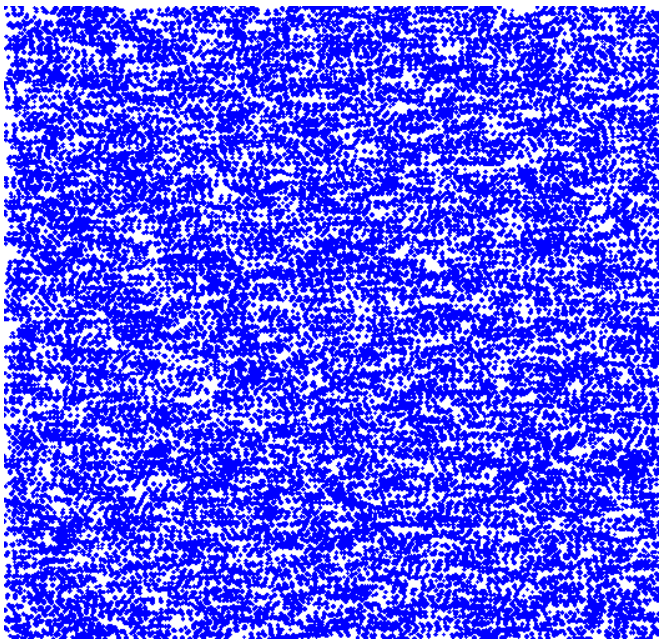


Figure 2.20: Horizontal Glass translation composed of 30,000 dots.

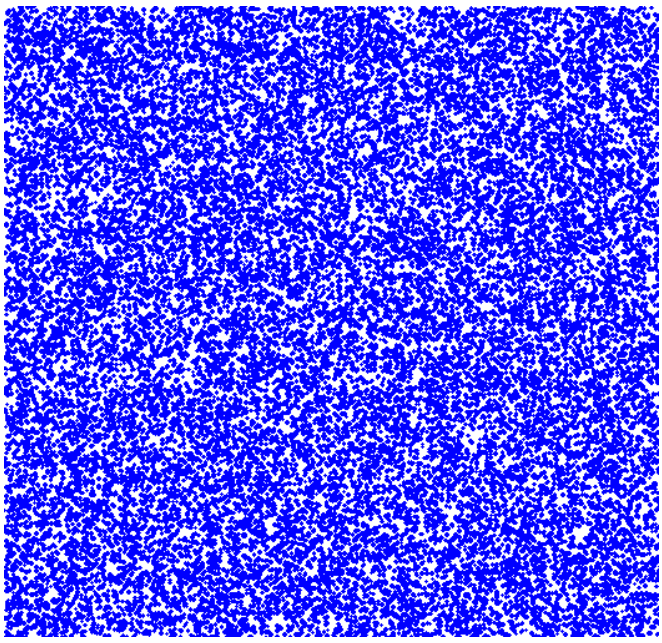


Figure 2.21: Noise composed of 30,000 dots.

Note that there is still some Glass effect evident in Figure 2.20, with Figure 2.19 showing some loading on neighbours six to ten.

As mentioned, a term for form area is also necessary to go with the term for dot size. Together, these give a foreground / background area ratio, which determines density ranges at which effects become prominent. The problem is, however, that adjacent or partially fused dots still have some degree of directional relation with one another (which diminishes with increased overlapping), and some Glass effect remains. A more sophisticated analysis might take into account degrees of fusion, or maybe the Voronoi diagram should be taken between separate entities, such that it changes upon dot contact. However, I have shown that the Glass effect persists at extremely high densities, whereupon explanation simply in terms of natural

neighbor loadings about transformation direction(s) is circumscribed. The question of Glass phenomena requires more work yet; hence it is developed in various shared contexts throughout the thesis, and is ultimately treated in a transformational context in Chapter 12 where it turns out that proximity links can still show structure without connecting transformation partners. Any utility of natural neighbor loadings about transformation direction(s) must then be subsumed in this more inclusive finding.

The travelling salesperson problem (TSP)

The travelling salesperson problem is a member of the class of computationally intractable 'NP-complete' problems (Lawler, 1985). It is a form of 'combinatorial optimisation' problem: the value of a function needs to be maximized or minimized. Owing to the enormous number of possible tour combinations for even a modest number of points, solutions for only a small number of points are realizable in any reasonable computing time. Hence computer scientists have devised heuristic procedures for near-optimal solutions.

The study of subject's solutions to travelling salesperson problems is important because they often get quick, near optimal solutions to problems for numbers of points that involve considerable computing time. And the travelling salesperson problem is relevant to the perception of pattern because quick, near optimal solutions are not haphazard. They appear to involve the perception of configuration, organization, or structure, in arrays. This has been shown convincingly by Vickers, Bovet, Lee, and Hughes (2003), who conclude that the perception of organization may be considered the perception of minimal forms of structure (p. 885).

So as to see how proximity measures might apply to the travelling salesperson problem, data from Vickers et al. (2003) were acquired, with permission, for 30 subjects. These data consisted of ordered indices to the site coordinates for tours made by each subject. Optimal solutions for tours were also provided. Each subject tackled the same series of 10 open ended and closed loop tours of 30 randomly located sites. Each of the 10 tours in the series was different, and the same series was used in random order for open and closed tours, except the latter were reflected about the 45^{o} diagonal. Half the group commenced with open tours and the other half commenced with closed tours.

For open tours, subjects were instructed to visit all sites; beginning on any site and finishing on a different site. For closed tours, they were instructed to visit all sites; beginning on any site and finishing on the same site. No instruction was given that solutions with crossed paths might be inefficient, and no subject reported having recognized closed tours as being transformed versions of open tours.

Analysis in terms of the Delaunay neighbour hierarchy

The data acquired from Vickers et al. (2003) for 16 of the subjects were reprocessed for the purpose of testing subject's adherence in their tours to edges of the Delaunay hierarchy. These subjects had the widest and most even possible spread of 'Percentage Above Optimal tour distance' (PAO) scores. ($\text{PAO}=\frac{\text{subject tour distance}}{\text{optimal tour distance}}\times 100$.) This spread facilitated interpretation of correlation statistics (for the obvious reason that tightly bunched data with spurious overlaps masks reciprocal relationships).

Results of reprocessing the data showed that travelling salesperson solutions by the 16 subjects loaded highly on Delaunay edges: average 97.62% of subjects links were Delaunay edges for open TSPs and 96.09% of subjects links were Delaunay edges for closed TSPs. Figure 2.22 depicts a hierarchy of edges, with Delaunay edges shown in yellow, red, and green, for an arbitrarily selected layout of 30 sites. Red and green edges belong to the

minimum spanning tree, of which the green edges belong to nearest neighbours. The offset white links show a closed tour generated by an arbitrarily selected subject.

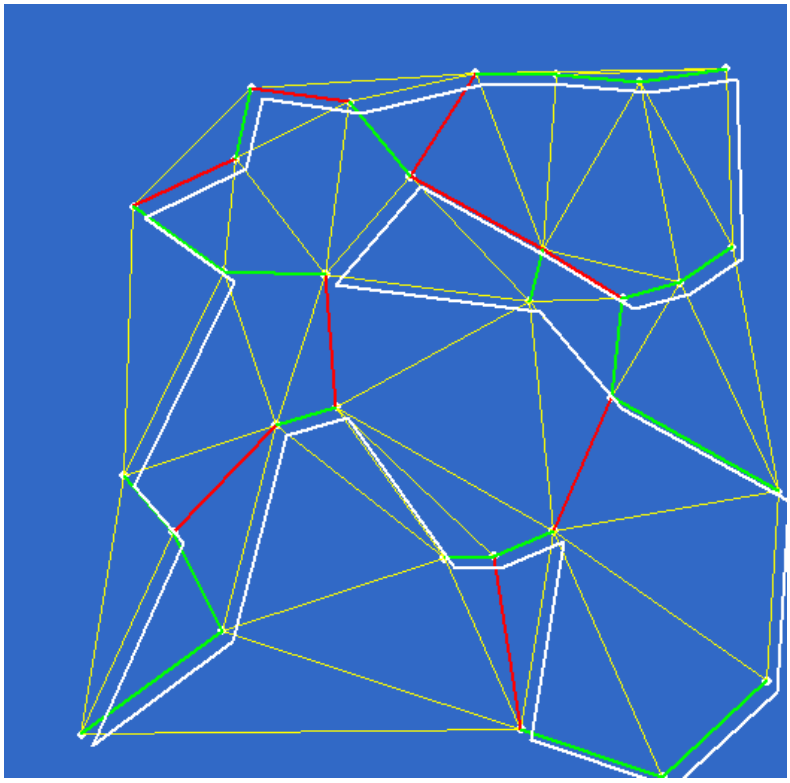


Figure 2.22: Delaunay diagram for 30 sites, with various subsets of the hierarchy colour coded (see preceding text). The offset white links show a closed tour generated by an arbitrarily selected subject.

As previously outlined, an average of two thirds of Delaunay edges for randomly distributed sites are 'direct' and one third are 'indirect'. For direct edges there was an average 89.09% loading by subjects for open TSPs (with the balance of 8.53% for indirect edges), and 86.64% loading by subjects for closed TSPs (with the balance of 9.5% for indirect edges). The small difference was because closed TSP solutions tended more towards a 'round tour' strategy, and wanted to load more heavily towards the convex hull, which has a higher proportion of indirect Delaunay edges. Loadings by subjects on nearest neighbour and minium spanning tree edges respectively were 80.02% and 71.95% for open TSPs, and 75.58% and 68.56% for closed TSPs.

The optimal open TSP solutions share 100% of their links with Delaunay edges, and the optimal closed TSP solutions share 98.33% of their links. Corresponding loadings for nearest neighbour and minium spanning tree edges are respectively 84.71% and 77.26% for optimal open TSPs, and 77.84% and 71.73% for optimal closed TSPs.

The distribution of PAO results for all 30 subjects was positively skewed (with tight bunching around the mode) whereas the distribution of PAO results for the 16 selected subjects was more uniform. Hence inclusion of all subjects in this particular analysis would have increased the loadings in favour of the hypothesis that subjects detect adjacency in the tradition of Delaunay.

The spread of subject's PAO scores facilitated some interesting correlations between the PAOs and proportions related to hierarchies of edges. Results are shown in Tables 2.1 to 2.4. Rather than do the correlations with proportions of the edge hierarchy used by subjects, they were done the other way around, i.e. with proportions not used. The longer the average

tour distance above optimal by a subject, the greater the proportion of nearest neighbour links, for example, not used in the subject's solutions. Any model for the TSPs clearly needs to yield statistically similar loadings to those indicated by the test subjects. Underlying this is the extremely heavy loading on Delaunay edges (Delaunay neighbours).

Table 2.1: Open tour correlations between PAO scores for 16 selected subjects and proportions related to hierarchies of edges.

	PAO	*Proportion of NN not used*	*Proportion of MS not used*	*Proportion of direct Delaunay not used*
Proportion of NN not used	0.84			
Proportion of MS not used	0.94	0.90		
Proportion of direct Delaunay not used	0.91	0.79	0.94	
Proportion of all Delaunay not used	0.78	0.66	0.77	0.86

Table 2.2: Open tour correlations between PAO scores for 6 selected subjects and proportions related to hierarchies of edges. PAOs came from the two lowest, the two highest, and two around the middle in the range for the 16 subjects.

	PAO	*Proportion of NN not used*	*Proportion of MS not used*	*Proportion of direct Delaunay not used*
Proportion of NN not used	0.90			
Proportion of MS not used	0.97	0.93		
Proportion of direct Delaunay not used	0.98	0.87	0.97	
Proportion of all Delaunay not used	0.85	0.79	0.87	0.92

Table 2.3: Closed tour correlations between PAO scores for 16 selected subjects and proportions related to hierarchies of edges.

	PAO	*Proportion of NN not used*	*Proportion of MS not used*	*Proportion of direct Delaunay not used*
Proportion of NN not used	0.82			
Proportion of MS not used	0.76	0.86		
Proportion of direct Delaunay not used	0.90	0.86	0.86	
Proportion of all Delaunay not used	0.82	0.76	0.75	0.68

Table 2.4: Closed tour correlations between PAO scores for 6 selected subjects and proportions related to hierarchies of edges. PAOs came from the two lowest, the two highest, and two around the middle in the range for the 16 subjects.

	PAO	*Proportion of NN not used*	*Proportion of MS not used*	*Proportion of direct Delaunay not used*
Proportion of NN not used	0.93			
Proportion of MS not used	0.76	0.89		
Proportion of direct Delaunay not used	0.88	0.94	0.84	
Proportion of all Delaunay not used	0.76	0.68	0.60	0.42

Very generally, minimum spanning tree and direct Delaunay edges figured more in open tours, and nearest neighbour and direct Delaunay edges figured more in closed tours. Minimum spanning links provide an efficient way of covering all bases in the case of open tours. However, for closed tours it seems that the requirement of finishing at starting points pressed subjects into employing somewhat more local links. In accord with Delaunay triangulation and the optimality of their solutions, subjects generally avoided crossing paths.

Application of neighbours to MacKay patterns

MacKay (1957a, 1957b, 1961) presented examples in which subjects fixate inspection patterns of alternating black and white stripes for a while, and then view a uniform test field. The elements in the test field then appear to have a streaming motion normal to the orientation of the stripes. Such visual after-effects occur with concentric, radial, and parallel, alternating patterns, for example.

Furthermore our observations show that rapid successive presentations of different static random dot patterns against a barred background induce an impression of dots streaming in directions normal to the bars, or in the case of curved bars normal to tangents of the bars. Again, backgrounds can be barred in any consistent manner: vertical, horizontal, herringbone, concentric, and radial all prove good examples.

At relatively low densities, the illusion can be explained in terms of nearest neighbours. The nearest neighbour of the majority of streaming dots is simply a point on a bar closest to a dot. Figure 2.23 shows one of a sequence of different random dot patterns against a concentric background. The whole range of Delaunay neighbours could be invoked to explain the effect as random dot densities become larger. On their own, however, nearest neighbours are sufficient to demonstrate the principle.

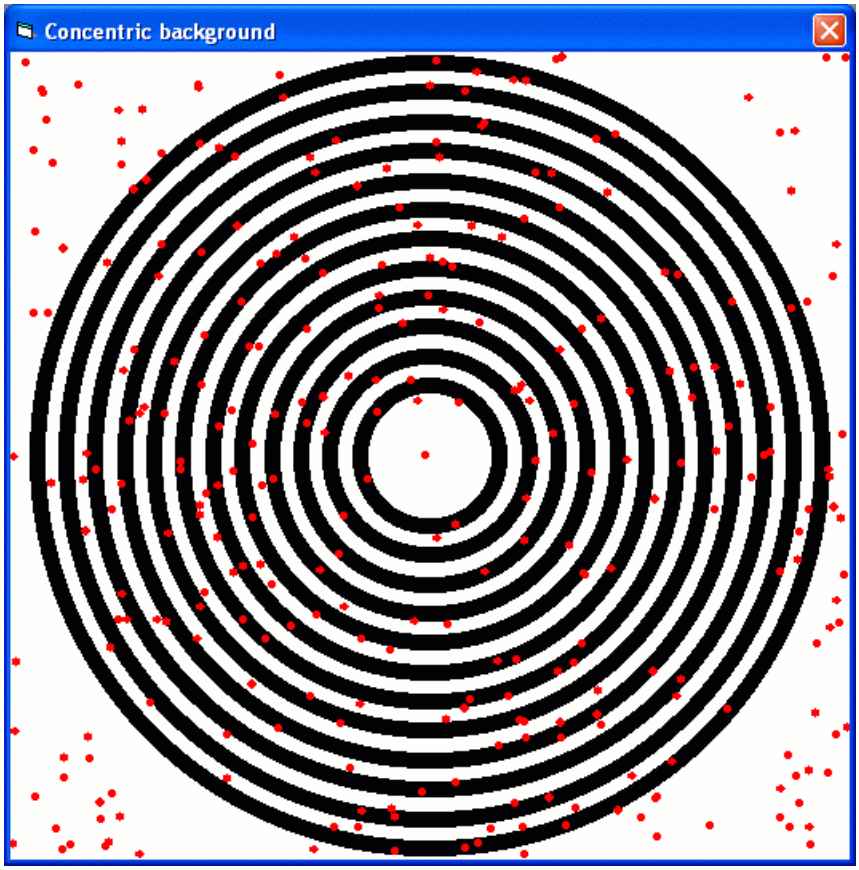


Figure 2.23: The nearest neighbour of the majority of dots is a point on a bar closest to a dot.

In an unpublished study, Vickers, Preiss, and Hughes (2003) showed the extent of motion seen in random dot kinematograms varied inversely with dot density and was closely related to the mean distance between nearest neighbours (see Appendix B). A variation on random dot kinematograms is to alternate every second random dot display with the same regular dot display. This was done for lines of dots on one occasion and concentric circles of dots on another occasion. First, a frame of independently generated random dots was presented, then a frame of lines of dots, then another frame of independently generated random dots, then the same frame of lines of dots, and so on. The procedure was then repeated, substituting concentric circles of dots for lines of dots.

In both cases the Mackay effect was evident, and this can be explained by sets if nearest neighbours linking dots from one frame to the next. (Nearest neighbours were not determined *within* a frame of dots but *between* consecutive frames, from one frame to the next.) See Figure 2.24, each part of which shows a set of nearest neighbour links from one frame to the next. (For explanatory purpose, consecutive frames are shown superposed.)

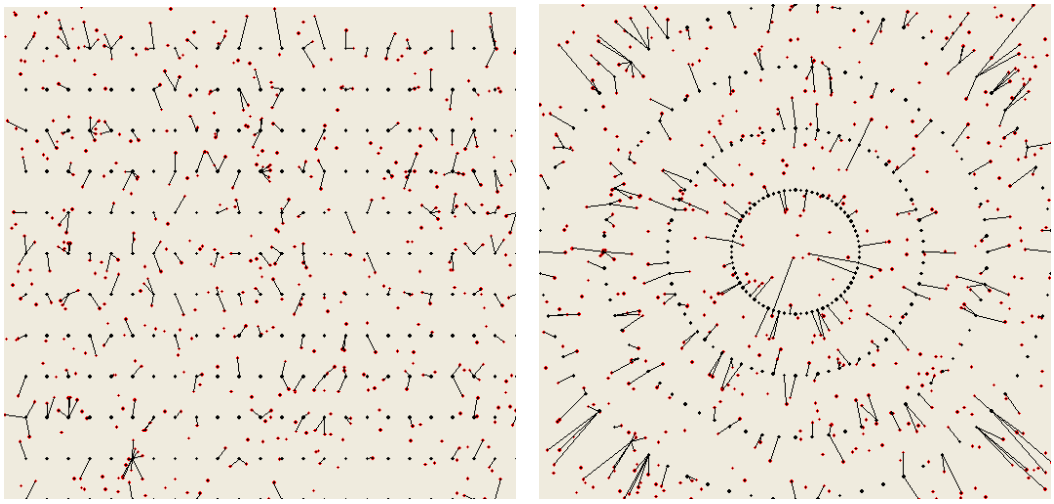


Figure 2.24: Independently generated random dots alternated with regular dots show the Mackay effect, which can be explained by nearest neighbours.

However, like the situation for Glass patterns, I have shown that the Mackay effect persists at extremely high densities, whereupon explanation simply in terms of proximity, again, breaks down. The MacKay visual after-effect requires more work, and is ultimately treated in a transformational context in Chapters 12 and 13.

Chapter 3: Nearest Neighbours

Barlow (1999) regarded statistical regularities in the environment as 'important for learning, memory, intelligence, inductive inference, and in fact for any area of cognitive science where an information-processing brain promotes survival by exploiting them' (p. 2). He referred to many researchers who have begun to show how the regular statistical properties of images are exploited by neural mechanisms. Chun (2000) showed that statistical regularities in the structure of images were encoded without explicit effort by observers. With regard to the obvious relevance of statistical properties of images, these next three chapters outline some statistical methods that can be used to discriminate characteristics of visual patterns and differences between such patterns.

Brief summary of chapter

Nearest neighbours are primary elements of the hierarchy of distance relationships. As such, statistical methods involving nearest neighbour analysis are discussed. These include nearest neighbours over an area, around the circumference of a circle, and along a line. After each theoretical exposition, the methodology is put to work in the analysis of visual patterns, including Glass, clustered, and regular patterns.

The statistical theory for nearest neighbours outlined in this chapter is not original. However, the implementation along with application to configurations of interest to the study of human visual perception is original.

Approach to nearest neighbours

Nearest neighbour distance is the elementary metric of point-to-point distance relationships. It heads the relationship hierarchy, and on its own can make important contributions to a surprising number of observed relationship phenomena. Where a nearest neighbour account is not sufficient on its own, the contribution from nearest neighbour distance remains important. The approach to nearest neighbours can be divided into two main considerations, and it is expedient to deal with each separately.

First, nearest neighbours play an important role in identifying the structure *within* a point pattern set. Nearest neighbour distance relationships within a point pattern set, whether summarized by mean nearest neighbour distance or by the entire frequency distribution of nearest neighbour distances, provide a measure for discriminating between three basic pattern attributes: clustered, random, and regular. Obviously the frequency distribution contains more information than the mean and, as such, can show degrees of different 'within' pattern attributes by extent of departure from the theoretical probability distribution for randomness. On the other hand, departure of the mean from the theoretical mean for randomness can show only degree of an overriding attribute.

Secondly, nearest neighbours play an important role identifying the relationship *between* two point pattern sets. Minimization of the largest nearest neighbour distance between two point pattern sets brings the sets into optimal proximity. In pattern sets of equal size reciprocal relationships are the same, hence such minimization can be reduced to minimization of the largest nearest neighbour distance *from* one pattern set *to* another, which brings the sets into optimal proximity. The pattern sets could be identical and one displaced from the other; or one could be some further affine transformation of the other, such as a dilation, shear, or rotation, or some combination thereof. Minimization of largest nearest neighbour distance steers transformations that bring two such patterns to identity, and along the way information about generating process for the pattern can be extracted. Additionally, change between successive (temporal) instances of a pattern set can be readily detected by examination of 'between' pattern nearest neighbour relationships.

Of course, a trivial situation exists for which the within and between pattern situations produce the same result. The distance relationship *between* elements of two superposed identical point patterns is effectively the same as the distance relationship among elements *within* of one or the other.

Poisson distribution and its relation to areal nearest neighbour distance

Before describing the nearest neighbour statistics involved within an areal point pattern set, an explanation of the Poisson distribution and its relation to areal nearest neighbour distance is expedient.

For a distribution of randomly located points, the probability of finding a point in an arbitrary area is the same for all points. In such a distribution, spacing is determined by chance. Boots and Getis (1988, pp. 13-14) describe a *homogeneous planar Poisson process* as a process in which points are generated in a study area subject to two conditions.

- Uniformity: Each location in the study area has an equal chance of receiving a point.
- Independence: The selection of a location for a point in no way influences the selection of locations for other points.

In other words the distribution of points is completely homogenous, and there is no interaction between points. This is known as complete spatial randomness (CSR), after Diggle (1983).

For a Poisson random variable, the probability that X is some value x is given by

$$\mathrm{P}(X = x) = \frac{u^x e^{-u}}{x!} \qquad x = 0, 1, \ldots, \qquad (1)$$

where u is the average number of occurrences in the specified interval. For the Poisson distribution, the mean equals u and the variance equals u.

By way of example, if the average number of events is 2.1 then the probability of getting 4 events is

$$\mathrm{P}(X = 4) = \frac{2.1^4 e^{-2.1}}{4!} = .0992 .$$

Now consider an area A, with a number of randomly located points N, then subregion a has, on average, $\frac{a}{A}N$ points. Otherwise stated, subregion a has, on average, $\frac{N}{A}a$ points, or λa points, where $\lambda = \frac{N}{A}$ is the average number of points per unit area, or density of points. The probability of finding r points in subregion a is

$$\mathrm{P}(R = r) = \frac{(\lambda a)^r e^{-\lambda a}}{r!},$$

following from direct substitution into Equation (1).

The probability of finding no point in subregion *a,* that is $r = 0$, is then

$$\mathrm{P}(0) = e^{-\lambda a}. \qquad (2)$$

The distribution of nearest neighbour distances under CSR depends on *N* and *A.* The probability of getting another point within distance *d* of some particular point is just the ratio of the corresponding subregion area, with radius *d,* to overall area

$$\frac{\pi d^2}{A}$$

and the number of points, on average, is

$$N \times \frac{\pi d^2}{A}$$

which is

$$\lambda \pi d^2$$

therefore, on the basis of equation (2)

$$\mathrm{P}(0) = e^{-\lambda \pi d^2}$$

This is a direct substitution into Equation (2), and is the probability of finding no points within distance *d* of some particular point.

It is also the probability that the distance, *D,* to its nearest neighbour is greater than *d.*

$$\mathrm{P}(D > d) = \mathrm{P}(0) = e^{-\lambda \pi d^2}$$

Hence

$$\mathrm{P}(D \le d) = 1 - \mathrm{P}(0) = 1 - e^{-\lambda \pi d^2} = F(d), \qquad (3)$$

which is the cumulative distribution function (cdf) for the nearest neighbour distance, *d.* (Note that the '≤' sign logically implies a cdf.) Figure 3.1 shows an example of a cdf of nearest neighbour distances for 200 points in an area of 25 million square units.

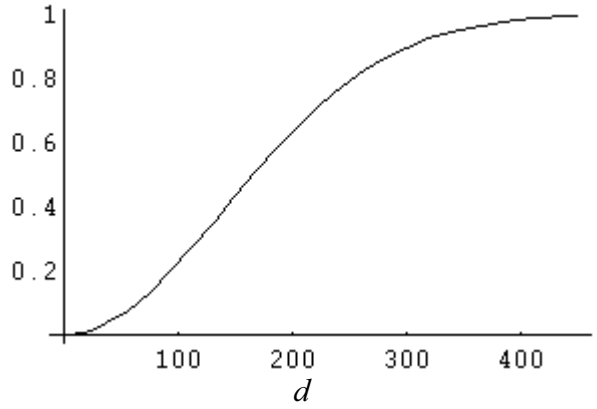


Figure 3.1: Poisson cumulative distribution function of nearest neighbour distances for 200 points in an area of 25 million square units.

Differentiating equation (3) gives

$$f(d) = 2\lambda\pi d e^{-\lambda\pi d^2}, \qquad (4)$$

which is the probability density function (pdf) for the nearest neighbour distance, *d.*
Note that the '=' sign logically implies a pdf. Figure 3.2 shows an example of a pdf of nearest neighbour distances for 200 points in an area of 25 million square units.[1]

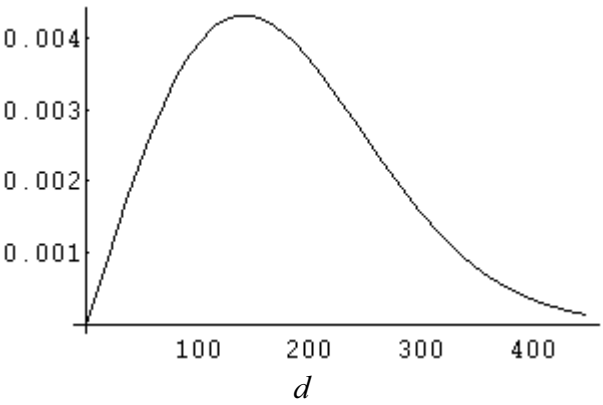


Figure 3.2: Poisson probability density function of nearest neighbour distances for 200 points in an area of 25 million square units.

[1] For a discrete distribution the pdf is the probability that the variate 'takes' the value *d*, i.e. P(*D* = *d*). For a continuous distribution the pdf is the probability that the variate 'has' the value *d.* The probability at a single point is effectively zero, hence it is often expressed in terms of an integral between two points.

$$\int_a^b f(d)\mathrm{d}d = \mathrm{P}(a < D < b)$$

In general terms, if *f(x)* is the pdf of a continuous random variable *X,* then *f(x)* × d*x* (a small positive quantity) is approximately the probability that *X* falls in the interval [*x*, *x* + d*x*]. So the integral of *f(x)* over an interval [*a, b*] is the probability that *a* < *X* < *b.*

Referring to equation (4), the mean, *u,* is the integral of the expression $d \times f(d)$ and the *variance* is the integral of the expression $(d-u)^2 \times f(d)$.

Nearest neighbour statistics for an areal point pattern set

The average nearest neighbour distance for an empirical distribution is just the total nearest neighbour distance divided by *N.* The simplest algorithm involves calculating the distances from each point to every other point. This results in *N* sets of *N* distances, and the minimum distance in each set yields a set of *N* nearest neighbour distances. Mean nearest neighbour distance is given by

$$\frac{1}{N}\sum_{1}^{N}\min_{j\neq i}\{u_{ij}\}$$

where u_{ij} is the distance from point *i* to point *j*. This is the observed mean for an empirical distribution of points, here denoted r_o. In a Euclidean *x, y* coordinate system $u_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$.

The mean nearest neighbour distance for CSR, which is the Poisson distribution of nearest neighbour distances, is given by

$$\frac{1}{2\sqrt{N/A}}, \text{ which is } \frac{1}{2\sqrt{\lambda}}, \text{ and is often given as } 0.5\sqrt{A/N}. \qquad (5)$$

This is the theoretical, or expected, mean, here denoted r_e, for a uniform random distribution of *N* points over an area *A.* Of course it must be taken for the same number of points distributed over the same area as those for the observed mean.

Dividing observed mean by expected mean gives the nearest neighbour statistic, here denoted *R,* which can be tested for significance (see Walford, 1995, for example).

$$R = \frac{r_o}{r_e}.$$

The nearest neighbour statistic is a descriptive device, which provides a quantitative summary of spacing between points. The more tightly points are *clustered,* the closer to 0 the value for *R* becomes; since average nearest neighbour distance decreases. The most tightly clustered situation is represented by all points superposed, for which the value of *R* equals 0. The closer to 1 the value for *R* becomes, the closer the points are to being *randomly* distributed. This follows from the way *R* is defined. The value of *R* equals 2.149 for perfectly uniformly spaced points; represented by a triangular lattice arrangement. Hence, the closer to 2.149 the value for *R* becomes, the more *uniformly spaced* are points.

Test for significance

The test for significance is a test to decide whether or not the difference between the observed mean, r_o, and the expected mean, r_e, can be reasonably attributed to chance. The significance of the difference between some *R* resulting from sample data and the *R* for CSR, which equals 1, can be examined by a *Z* test.

The null hypothesis states that *R* has a value other than 1 wholly due to sampling error: that is *R* is due to CSR. The object is to show that the probability of *R* being due to sampling error is small enough to accept the alternative hypothesis: that is *R* is due to some non-random, or patterning, mechanism. But before a *Z* test can be performed the standard deviation of the sampling distribution, or standard error, is required. The variance of the Poisson distribution of nearest neighbour distances is

$$\frac{4-\pi}{4(N/A)\pi N}, \qquad (6)$$

from which the standard error is

$$\sqrt{\frac{4-\pi}{4(N/A)\pi N}},$$

which is calculated to be

$$\frac{0.26136}{\sqrt{N(N/A)}}, \text{ or alternatively } \sqrt{0.06831\,A/N^2}, \qquad (7)$$

and is here denoted S_d.
The formula for the *Z* test is now

$$Z = \frac{r_o - r_e}{S_d},$$

which, by definition, has a normal distribution.

(Since each nearest neighbour distance for a random point pattern is distributed independently of all the others and follows the distribution of expression (4), the observed mean for a random point pattern is distributed approximately as the normal distribution, with mean as per expression (5) and variance as per expression (6)).

If the absolute value of *Z* exceeds the number of standard deviations for a chosen level of significance (often .01 or .05) then the null hypothesis that *R* has a value other than 1 wholly due to sampling error should be rejected in favour of some non-random mechanism.

Computational examples for nearest neighbour analysis of means

By way of example, nearest neighbour analyses of the mean are performed for the three point patterns shown in Figures 3.3, 3.4 and, further along, Figure 3.5.

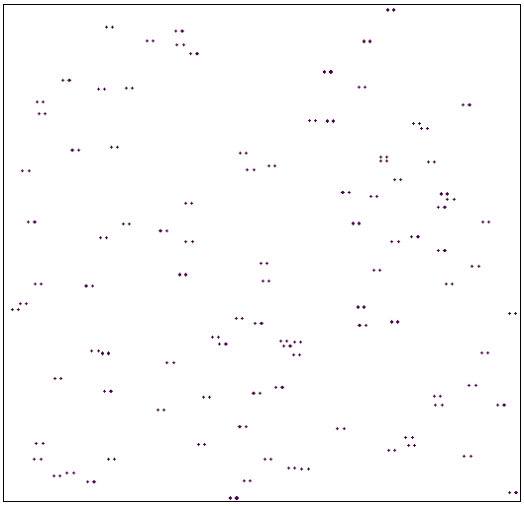


Figure 3.3: Glass translation consisting of 100 point pairs.

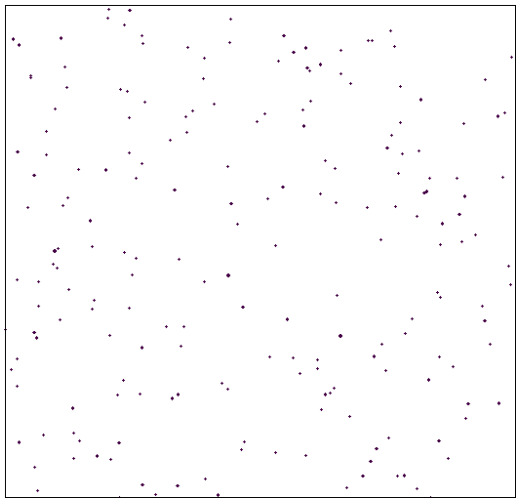


Figure 3.4: 200 randomly located points.

Figures 3.3 and 3.4 each have 200 points dispersed over equal screen areas, here scaled at 5,000 screen units × 5,000 screen units = 25 million square units. Formula (8) below, for the observed mean nearest neighbour distance, was implemented on a computer. By

$$r_o = \frac{1}{N}\sum_{1}^{N} \min_{j \neq i} \{u_{ij}\} \qquad (8)$$

the observed mean nearest neighbour distance for the Glass pattern, shown in Figure 3.3, is 59.249 units.

Because the figures have the same number of points dispersed over equal areas, the expected mean nearest neighbour distance from the one Poisson distribution is applicable to both.

$$r_e = \frac{1}{2\sqrt{N/A}} = \frac{1}{2\sqrt{200/25000000}} = 176.777 \text{ units}$$

The nearest neighbour statistic for Figure 3.3 is then

$$R = \frac{r_o}{r_e} = \frac{59.249}{176.777} = 0.335,$$

which is far from the nearest neighbour statistic of 1 expected for a random distribution.

Moving on to the test for significance: fist, the standard error is

$$\frac{0.26136}{\sqrt{N(N/A)}} = \frac{0.26136}{\sqrt{200(200/25000000)}} = 6.534$$

and

$$Z = \frac{r_o - r_e}{S_d} = \frac{59.249 - 176.777}{6.534} = -17.988,$$

which substantially exceeds the 1.96 standard deviations from tables of the normal distribution for the $\alpha = .05$ level of significance. Hence H_0 of CSR is rejected, and because the observed mean nearest neighbour distance is less than the expected mean, the points tend to clustering.

Confidence limits for r_e in the population are

$$r_e + Z_\alpha S_d \text{ and } r_e - Z_\alpha S_d.$$

If $\alpha = .05$, then for every 100 samples taken from a randomly distributed population it can be expected that the mean nearest neighbour distance for 95 of them will lie between the confidence limits.

Moving on to the random distribution shown in Figure 3.4, the observed mean nearest neighbour distance is 176.082 units. Since the expected mean is 176.777 units, $R = 0.996$, which is very close to the nearest neighbour statistic of 1 for CSR. $Z = 0.106$, which is much less than the 1.96 standard deviations from tables of the normal distribution for the $\alpha = .05$ level of significance. Hence H_0 of CSR cannot be rejected. For every 100 random samples bounded as per Figure 3.4 and drawn from the population from which the sample of Figure 3.4 was drawn, it can be expected that the mean nearest neighbour distance for 95 of them will lie between

176.777 + (1.96 × 6.534) = **189.584** units and 176.777 – (1.96 × 6.534) = **163.970** units.

Edge effect

For much of pattern detection in psychology, study areas are readily arranged to be rectangular or circular, and a square study area as a special rectangular instance is often used. These forms are addressed by several edge correction techniques (see Boots & Getis, n.d., for example).

For nearest neighbour analysis the edge effect is simple: some points in a study area that are proximate to the edges would have their nearest neighbours outside the edges, but they are paired with respective nearest points in the study area, which are *not* their nearest neighbours. Results from the (necessarily bounded) examples given above relate to infinite or unbounded CSR, therefore they are influenced somewhat by the edge effect. However the influence is marginal because of the reasonable number of points involved.

As the number of points increases, the ratio of number of inner points to number of peripheral points increases, which reduces the edge effect. Generally, the fewer the number of points the more important some kind of edge correction becomes. As a rough rule of thumb, edge correction should be applied when fewer than 200 points are involved but, ultimately, edge effect issues are matters for judgement and experience.

Buffer zone and toroidal edge correction techniques are left to pages 73 and 74 of the next chapter for discussion in fuller context. A further method of dealing with the edge effect, but restricted to nearest neighbour analysis, is to calculate mean nearest neighbour distance using only those nearest neighbour distances from points *i* to points *j* that are less than the distances between points *i* and a nearest boundary. Of course this reduces the number of nearest neighbour distances that contribute to a mean, which can be awkward if N is small; and because the method tends to retain small nearest neighbour distances it can introduce bias.

If assessment by a researcher indicates that points in a study area are dispersed in such a way that the great majority of nearest neighbour distances are legitimate, then edge correction may not be an issue. Only by applying calculations initially both with and without edge correction can a researcher get a feel for the necessity of its application. As far as the majority of studies underlying the content herein are concerned, the edge effect makes little difference. This can be appreciated by way of recalculation of the examples given above with a well-tried method of edge correction.

Donnelly (1978) showed that for N greater than 7 points in a study area that is not highly irregular, the adjusted expression (5)

$$\frac{1}{2\sqrt{N/A}} + (0.0514 + 0.041/\sqrt{N})\,B/N, \text{ which is } \frac{1}{2\sqrt{\lambda}} + (0.0514 + 0.041/\sqrt{N})\,B/N, \text{ and is often}$$

given as $0.5\sqrt{A/N} + (0.0514 + 0.041/\sqrt{N})\,B/N$, (9)

where B is the length of the study area boundary, approximates the mean nearest neighbour distance for the bounded Poisson distribution.

Expression (7) for the standard error of nearest neighbour distances for the bounded Poisson distribution becomes

$$\sqrt{0.06831\,A/N^2 + 0.037B\sqrt{A/N^5}}\,. \qquad (10)$$

See footnote[2].

The observed mean nearest neighbour distance for the Glass pattern shown in Figure 3.3 is, of course, still 59.249 units. Recalculating: the expected mean is now 182.207 units, and $R = 0.325$, which is still far from the nearest neighbour statistic of 1 for a random distribution. $Z = 17.525$, which substantially exceeds the 1.96 standard deviations from tables of the normal distribution for the $\alpha = .05$ level of significance. Hence H_0 of CSR is rejected, and because the observed mean nearest neighbour distance is less than the expected mean, the points tend to clustering.

The observed mean nearest neighbour distance for the random distribution shown in Figure 3.4 is 176.082 units, and the expected mean is 182.207 units. Hence $R = 0.966$, which is still close to the nearest neighbour statistic of 1 for a random distribution. $Z = 0.873$, which is less than the 1.96 standard deviations from tables of the normal distribution for the $\alpha = .05$ level of significance. Hence H_0 of CSR cannot be rejected.

The third example, Figure 3.5, shows a fairly regular pattern of 196 points in an area of 25 million square units.[3] It is here subjected to nearest neighbour analysis with edge correction; this time using alternate forms of the base expressions listed at (5) and (7).

---

[2] Since the modifications leading to expressions (9) and (10) were found by analysing simulated point patterns in study areas of assorted geometric shapes including rectangles, squares, circles, and ellipses, they should not be applied to irregularly shaped areas.

[3] The fixed form size throughout the examples eased the computer programming effort, but at the expense of a marginally more accurate value for R in this instance. The marginally more accurate value could have been realized by placing the boundaries as close as possible to the most outwardly located points.

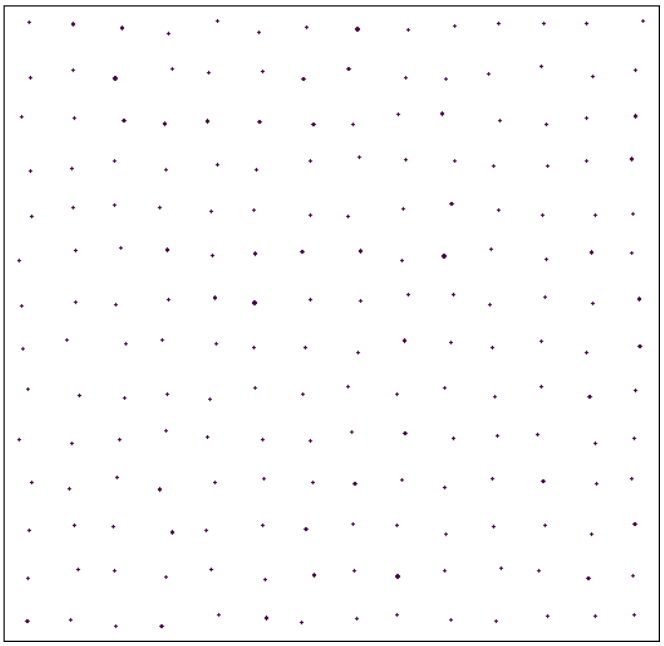


Figure 3.5: Fairly regular pattern.

The observed mean nearest neighbour distance is 315.383 units. The expected mean by expression (9)

$$0.5\sqrt{A/N} + (0.0514 + 0.041/\sqrt{N})\,B/N$$

is

$$0.5\sqrt{25000000/196} + (0.0514 + 0.041/\sqrt{196})\,20000/200 = 184.115 \text{ units},$$

therefore

$$R = \frac{r_o}{r_e} = \frac{315.383}{184.115} = 1.713,$$

which well exceeds the nearest neighbour statistic of 1 expected for a random distribution. The standard error of nearest neighbour distances for the bounded Poisson distribution by expression (10)

$$\sqrt{0.06831\,A/N^2 + 0.037B\sqrt{A/N^5}}$$

is

$$\sqrt{0.06831 \times 25000000/200^2 + 0.037 \times 20000\sqrt{25000000/200^5}} = 7.164,$$

and

$$Z = \frac{r_o - r_e}{S_d} = \frac{315.383 - 184.115}{7.164} = 18.332,$$

which substantially exceeds the 1.96 standard deviations from tables of the normal distribution for the $\alpha = .05$ level of significance. Hence H_0 of CSR is rejected, and because the observed mean nearest neighbour distance is greater than the expected mean, the points tend to regularity.

(Note that for tendency to clustering Z is negative; hence positive Z indicates tendency to regularity. But, of course, Z closer to zero indicates randomness: $1.96 \geq Z \geq -1.96$, for example, indicates randomness at the $\alpha = .05$ level of significance.)

Before moving on to refined nearest neighbour analysis, one last consideration warrants mention. If the maximum possible nearest neighbour statistic is 2.149 (R equals 2.1491, to five significant figures, for perfectly uniformly spaced points) then the maximum possible mean nearest neighbour distance between points is

$$\frac{R_{\max}}{R} \times R_o ,$$

which, for this example, is

$$\frac{2.149}{1.713} \times 315.383 = 395.656 \text{ units.}$$

When observing a display of dots showing nearest neighbour links, I found it useful to subjectively compare its mean nearest neighbour distance with the maximum possible spacing for the same number of dots in the same field. This provided me with a readily visualized form of feedback as to the statistical rating of the pattern in terms of straight distance comparisons.

Refined nearest neighbour analysis

From the point of view of distance relationships, considering all point-to-point distances provides maximum information. Important subsets of these—point-to-point distances out to some cut-off distance, indirect and direct Delaunay distances, relative neighbourhood distances, minimum spanning distances, and nearest neighbour distances, all result in loss of information. However, the objective is to choose optimal linking schemes: schemes that balance most information against a least number of distance links.

Conversion of data to some summary statistic, commonly to a single summary statistic, also results in loss of information. Maximum information is realized by comparing the distribution of all point-to-point distances for some point pattern with that for the equivalent theoretical distribution; if one exists, of course! And distributions for the optimal linking schemes produce a lot of information for relatively little outlay. A good return for outlay is the distribution for nearest neighbour distances; and this is the backbone of refined nearest neighbour analysis.

Before proceeding, it must be emphasised that *distance* is not the only measure for pattern detection that can be usefully treated by statistical methods. Least angle for Delaunay triangulation is another useful measure, for example. The mean and distribution of least angles for Delaunay triangles can have some advantages in some situations over corresponding distance statistics; and, of course, angles and distances can interact.

Refined nearest neighbour analysis involves comparing the distribution of some observed set of nearest neighbour distances, $F(d_i)$, with that of expected nearest neighbour distances for CSR, $P(d_i)$, (see Boots & Getis, n.d.). Observed nearest neighbour distances are ranked in ascending order and the proportions $F(d_i \leq r)$ are determined. The distances, r, usually coincide with nearest neighbour distances. Alternatively, they can be incremented by

some fixed amount. The proportions of observed nearest neighbour distances not greater than corresponding distances, *r,* are denoted *F(r),* and referred to as the *observed proportion.*

Corresponding proportions $P(d_i \leq r)$ of expected nearest neighbour distances for CSR are then calculated. The proportions of expected nearest neighbour distances not greater than corresponding distances, *r,* are denoted *P(r),* and referred to as the *expected proportion.* The equation for corresponding proportions of expected nearest neighbour distances is provided by the previously derived formula for equation (3), and is

$$P(r) = 1 - \mathrm{P}(0) = 1 - e^{-\lambda \pi r^2}. \qquad (11)$$

Diggle (1981) suggested the statistical expression

$$\max |F(r) - P(r)|,$$

denoted d_r, for comparing the two results. It signifies the largest absolute difference between the two cdfs for corresponding values of *r.* Since observed nearest neighbour distances for patterns are not typically independent, Diggle also suggested a Monte Carlo procedure for testing the significance of this difference.

For the Monte Carlo procedure, repeated equivalent Poisson simulations are invoked. The largest absolute difference between the observed proportion, as simulated, and the expected proportion, is calculated for each simulation: that is, a d_r is calculated for each simulation. If a d_r resulting from some empirical distribution then turns out to be one of the five largest when included with those resulting from 99 Poisson simulations, for example, then an H_0 of CSR can be rejected for $\alpha = .05$, and if it turns out to be the largest, then an H_0 of CSR can be rejected for $\alpha = .01$.

However, I choose to calculate upper and lower delimiting confidence values for the Poisson distribution by first calculating the standard error as per expression (7), and then multiplying by the desired $Z\alpha$. For the upper delimiting value the result is *added* to the *r* value for the largest absolute difference, which is then substituted for *r* in calculation by equation (11). For the lower delimiting value the result is *subtracted* from the *r* value for the largest absolute difference, which is again substituted for *r* in calculation by equation (11). For the purpose of refined nearest neighbour analysis, this is the only pair that need be calculated. However, the manner in which the cdf for an empirical pattern mirrors the pattern's noticeable (intrinsic) difference to that for CSR is important, and for this kind of evaluation each pair of delimiting values at each *r* is calculated for plotting along with the Poisson cdf.

More than just testing for statistical significance of d_r, deviation from CSR anywhere along a nearest neighbour distance cdf can give information about pattern type. A pattern of regularly spaced clusters, for example, can show more relatively short and long nearest neighbour distances than those shown for the corresponding CSR.

Sometimes statistical significance might be reached and other times not. What is important is how the cdf for an empirical pattern mirrors the pattern's noticeable difference to that of a random pattern. For example, see Figure 3.6, in which the magenta line shows the cdf for the Poisson distribution, and the dashed lines delineate the 95% confidence interval. The confidence interval is based upon measures associated with the random distribution, not measures associated with an empirical distribution that is being tested against such randomness. It provides a graphic reference concerning significance for empirical measures.

The small circles describe the cdf for the fairly regular pattern, Figure 3.5, given for the last nearest neighbour analysis of the mean.[4] The point of largest absolute difference between the two cdfs for corresponding values of *r* is marked with a small blue circle. This example clearly shows the dearth of short nearest neighbour distances and the surfeit of longer nearest neighbour distances, which gave rise to the larger than expected mean.

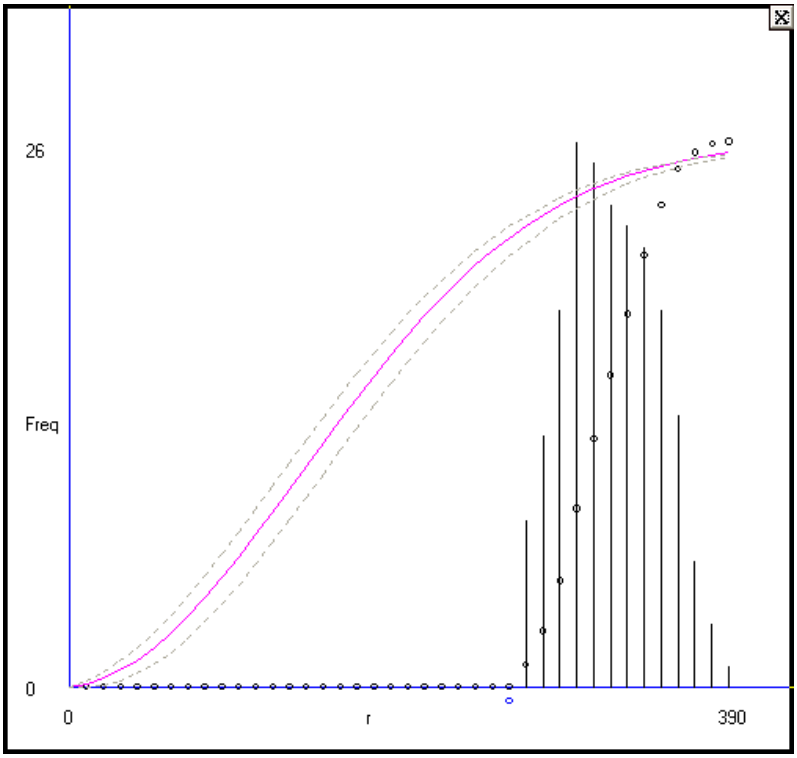


Figure 3.6: Distribution functions for the fairly regular example shown in Figure 3.5.

Table 3.1 below, frames the refined nearest neighbour analysis. Of course some columns of values necessarily correspond to values that gave rise to the graphs of Figure 3.6. Table 3.1, part A refers to the analysis without edge correction and part B refers to the same analysis with edge correction. For each part, *dr,* the point of largest absolute difference between the two cdfs for corresponding values of *r,* is highlighted in red. This difference, upward of the small blue circle in Figure 3.6, far exceeds the confidence limit, and because *F(r)* is less than *P(r)* a regular pattern is indicated. Remember, *F(r)* is the proportion of observed nearest neighbour distances not greater than the corresponding *r,* and given that it is less than expected at the point of maximum difference then the pattern is regular. While table 3.1, part A is fairly self-explanatory, part B, with edge correction, needs explanation. However, it will be noticed forthwith that edge correction, again, does not change the outcome.

[4] Ideally the two cdfs should coincide at the final data point, at least; but the scaling precision that produced the display is slightly imperfect. Additionally the vertical lines show a deliberately over-scaled histogram, for better viewing definition, outlining the pdf of the fairly regular pattern.

Table 3.1:

A. Refined nearest neighbour analysis without edge correction.

B. Refined nearest neighbour analysis with edge correction

Distance r (screen units)	Number of points for which $d_i \leq r$	*Observed proportion F(r)*	*Expected Proportion P(r)*	$\|F(r)–P(r)\|$	*Distance r* (screen units)	Number of points for which $d_i \leq r$	Number of points for which $u_i < r < d_i$	*Observed proportion F(r)*	*Expected Proportion P(r)*	$\|F(r)–P(r)\|$
10	0	0	0.005023	0.005023	10	0	0	0	0.005023	0.005023
20	0	0	0.014996	0.014996	20	0	0	0	0.014996	0.014996
30	0	0	0.029773	0.029773	30	0	0	0	0.029773	0.029773
40	0	0	0.049138	0.049138	40	0	0	0	0.049138	0.049138
50	0	0	0.072814	0.072814	50	0	0	0	0.072814	0.072814
60	0	0	0.100465	0.100465	60	0	0	0	0.100465	0.100465
70	0	0	0.131708	0.131708	70	0	0	0	0.131708	0.131708
80	0	0	0.16612	0.16612	80	0	0	0	0.16612	0.16612
90	0	0	0.203245	0.203245	90	0	0	0	0.203245	0.203245
100	0	0	0.24261	0.24261	100	0	0	0	0.24261	0.24261
110	0	0	0.283729	0.283729	110	0	0	0	0.283729	0.283729
120	0	0	0.326115	0.326115	120	0	0	0	0.326115	0.326115
130	0	0	0.369291	0.369291	130	0	0	0	0.369291	0.369291
140	0	0	0.412796	0.412796	140	0	0	0	0.412796	0.412796
150	0	0	0.456196	0.456196	150	0	0	0	0.456196	0.456196
160	0	0	0.499086	0.499086	160	0	0	0	0.499086	0.499086
170	0	0	0.5411	0.5411	170	0	0	0	0.5411	0.5411
180	0	0	0.58191	0.58191	180	0	0	0	0.58191	0.58191
190	0	0	0.621236	0.621236	190	0	0	0	0.621236	0.621236
200	0	0	0.65884	0.65884	200	0	0	0	0.65884	0.65884
210	0	0	0.694532	0.694532	210	0	0	0	0.694532	0.694532
220	0	0	0.728166	0.728166	220	0	0	0	0.728166	0.728166
230	0	0	0.759639	0.759639	230	0	0	0	0.759639	0.759639
240	0	0	0.788891	0.788891	240	0	0	0	0.788891	0.788891
250	0	0	0.815898	0.815898	250	0	0	0	0.815898	0.815898
260	0	0	0.840669	0.840669	260	0	0	0	0.840669	0.840669
270	8	0.040816	0.863245	0.822429	270	5	3	0.025907	0.863245	0.837338
280	20	0.102041	0.883691	0.78165	280	16	4	0.083333	0.883691	0.800358
290	38	0.193878	0.902094	0.708216	290	33	5	0.172775	0.902094	0.729319
300	64	0.326531	0.918556	0.592025	300	55	9	0.294118	0.918556	0.624438
310	89	0.454082	0.933194	0.479112	310	75	14	0.412088	0.933194	0.521106
320	112	0.571429	0.946132	0.374703	320	94	18	0.52809	0.946132	0.418042
330	134	0.683673	0.957501	0.273828	330	110	24	0.639535	0.957501	0.317966
340	155	0.790816	0.967433	0.176617	340	120	35	0.745342	0.967433	0.222091
350	173	0.882653	0.976058	0.093405	350	130	43	0.849673	0.976058	0.126385
360	186	0.94898	0.983507	0.034527	360	136	50	0.931507	0.983507	0.052
370	192	0.979592	0.989903	0.010311	370	141	51	0.972414	0.989903	0.017489
380	195	0.994898	0.995364	0.000466	380	144	51	0.993103	0.995364	0.002261
390	196	1	1	0	390	144	52	1	1	0

Table 3.1, part B is the same as Part A except that it has an extra column headed $u_i < r < d_i$, where u_i is the distance of each point i to the nearest part of the study area boundary. If a point is closer to the boundary than it is to another point, then it is discounted for the particular distance, *r*. The column heading states 'number of points for which the distance *r* is greater than the distance from a point *i* to the boundary while it is less than the distance from the same point *i* to its nearest neighbour'. For each *r* the accumulation of such numbers of points is subtracted from *N*. Hence each entry for *F(r),* the proportion of observed nearest

neighbour distances that are not greater than each *r,* is determined on the basis of some accumulation of reduction in *N.* (For more information regarding nearest neighbour analysis, refined and otherwise, see Boots & Getis, 1988).

Before moving onto nearest neighbour analysis in one dimension, it is important from a psychological perspective to appreciate how the cdf for an empirical pattern can mirror the pattern's noticeable difference to that of the Poisson distribution. Statistical significance need not come into the argument, other than to perhaps provide some kind of opportune support.

As a pattern becomes more clustered, its average nearest neighbour distance becomes smaller relative to that of the equivalent random situation. As a pattern becomes more regular, its average nearest neighbour distance becomes larger relative to that of the equivalent random situation. Additionally, the more clustered a pattern, the more limited the variability of its nearest neighbour distances. Likewise, the more regular a pattern, the more limited the variability of its nearest neighbour distances. With this in mind, the points shown in Figure 3.7 reflect an attempt to include a range of features: clustering, randomness, and regularity, in the hope that these will be adequately reflected in the distributions shown in Figure 3.9. Nearest neighbour links are depicted in Figure 3.8, and it must be noted that some of these are reflexive links, which are counted twice in the distributions.

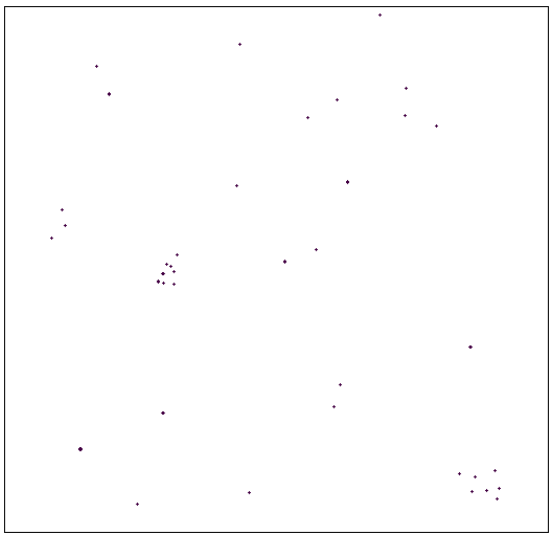


Figure 3.7: Pattern with a range of features: clustering, randomness, and regularity.

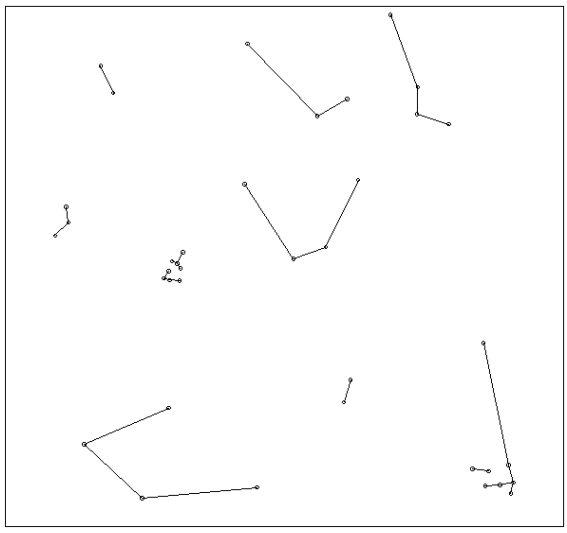


Figure 3.8: Nearest neighbour links for the pattern with a range of features.

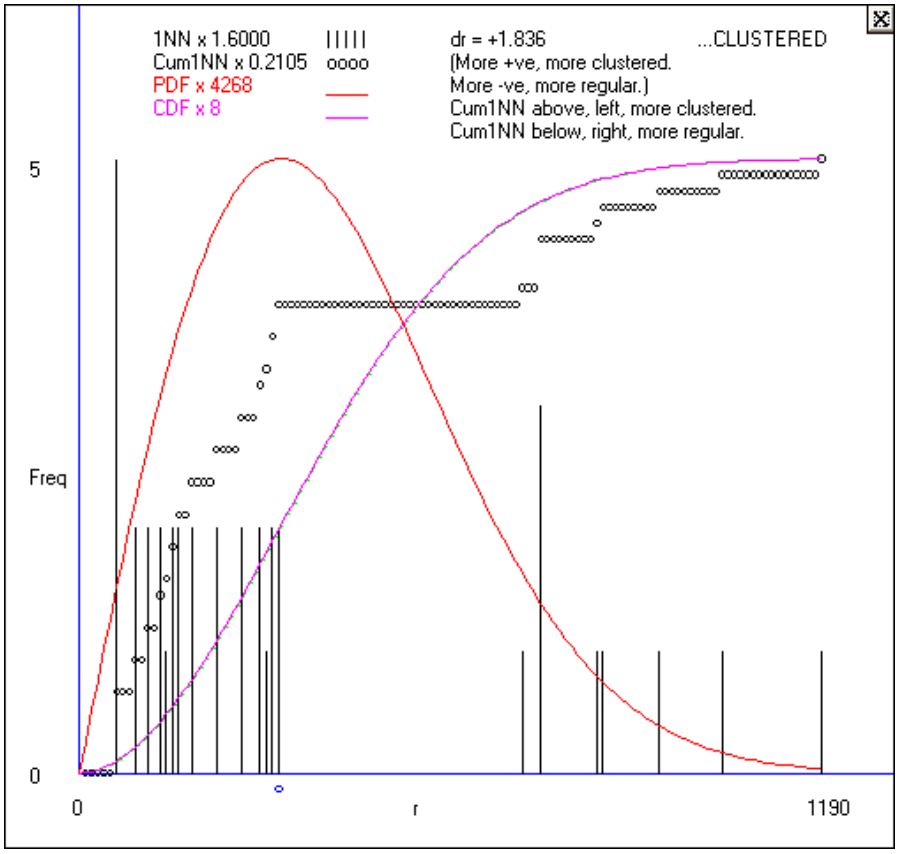


Figure 3.9: Refined nearest neighbour analysis plots for the pattern with a range of features. See text for analysis.

Table 3.2 is a computer readout for the nearest neighbour analysis of the mean, calculated as per the previous section.

Table 3.2: Nearest neighbour analysis of the mean.

Observed mean nearest neighbour distance	334.141
Standard deviation	310.137
Max possible nearest neighbour distance	937.237
Expected mean nearest neighbour distance	436.107
Nearest neighbour statistic	0.766
Standard error of expected mean	39.977
Z	2.551
Observed mean nearest neighbour distance < expected mean. Verdict at $\alpha = .05$:	Clustered

Figure 3.9 shows everything: theoretical (Poisson) and empirical cdfs and pdfs. The pdfs are over-scaled with respect to the cdf s for sake of better viewing definition. Although the one statistical verdict, clustered at $\alpha = .05$, is given by both nearest neighbour and refined nearest neighbour analysis, $Z = 2.551$ and $dr = 1.836$ respectively indicate that the pattern is not far removed from randomness. Therefore an examination of the full distributions is required to understand the analysis properly.

The nearest neighbour distances for which clustering is evident are shown by the small circles above the theoretical cdf, and those for which regularity is evident are shown by the

small circles below the cdf. Looking at the pdfs: for the empirical pdf an elevation of short, clustering, distances, shown by the left group of vertical lines, is evident and a generally lesser elevation of longer, regular, distances, shown by the right group of vertical lines, is evident. However, the overall shape of the empirical pdf has a meagre correspondence with that of the theoretical pdf. Looking at the cdfs: the average behaviour of the empirical distribution can be seen as not far from random. Nonetheless the overriding verdict is for clustering; basically because of the overall elevation corresponding to clustering in the empirical pdf. This is the kind of approach needed, along with the statistics, in order to realize the utility of refined nearest neighbour analysis.

Other neighbour analysis

Equivalent analyses can be usefully performed on second, third…out to fifth or sixth nearest neighbours, but such neighbours can lose information carrying effectiveness in some relationship to their ranking.

An extension of the reasoning stated for the derivation of equation (4) can be applied to show that

$$f(d) = \frac{2(\lambda\pi)^k r^{(2k-1)} e^{-\lambda\pi r^2}}{(k-1)!}$$

for the expected k^{th} nearest neighbour pdf. The expected k^{th} mean nearest neighbour distance is given by

$$\frac{(2k)!k}{(2^k k!)^2\sqrt{d}},$$

here denoted r_{ke}, and the k^{th} nearest neighbour statistic, here denoted R_k, can be tested for significance.

$$R_k = \frac{r_{ko}}{r_{ke}}.$$

Of course R_1 for the nearest neighbour statistic, given earlier on page 40 as R, is the situation for which $k = 1$ in these formulae.

The constants given in table 3.3, simplify calculations and, in particular, the variance values, last column, facilitate calculation of standard error.

Table 3.3: Values of constants for k^{th} nearest neighbour means and variances. (After Thompson, 1956).

Order of neighbours	Mean $= \gamma_1\sqrt{A/N}$ γ_1	Variance $= \gamma_2 A/N^2$ γ_2
1	0.5000	.0683
2	0.7500	.0741
3	0.9375	.0760
4	1.0937	.0770
5	1.2305	.0775
6	1.3535	.0778

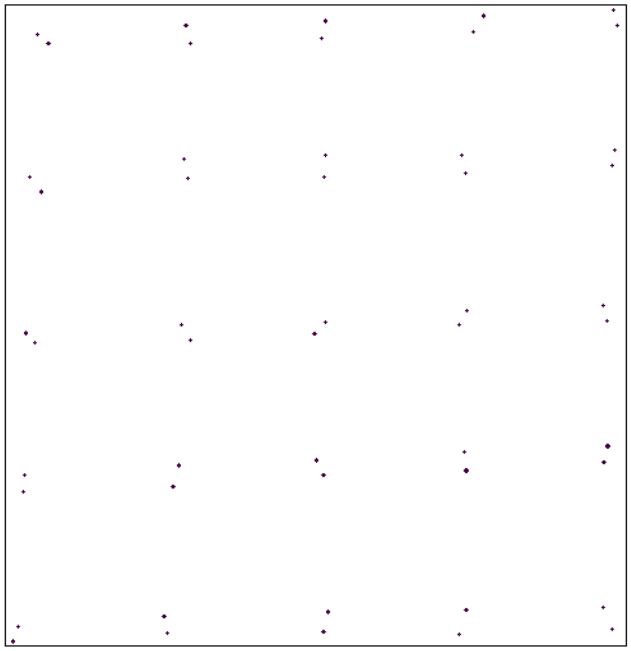


Figure 3.10: Pattern formulated to expose limitation of nearest neighbour analysis of means.

As already mooted, nearest neighbour analysis has its limitations. Patterns can be formulated for which it provides insufficient information. For example, such analysis applied to the pattern depicted in Figure 3.10 indicates clustering, which is constituted of the spacing within pairs. It does not indicate the regularity, which is constituted of the spacing between pairs. In this instance, application of the analysis using second nearest neighbour reveals the regularity.

Limitations can be experienced more under the demands of some disciplines than others. Probability distributions for point patterns, other than the probability distribution for points dispersed randomly, can be difficult to formulate on a theoretical basis. Although non-random patterns can be ranked according to degree of clustering through to regularity, not much can be inferred about generating processes.

However, generating processes for patterns of interest to psychology are often well defined. The way a pattern changes over changing viewpoint is defined by the set of affine transformations on the pattern, for example. A Glass pattern is formed by some transformation over randomly dispersed points; to give another example. In other words, theoretical distributions could be formulated for some patterns of interest to psychology, upon which they could be compared for quantitative difference.

Nearest neighbour distance relationships in one dimension

An account of nearest neighbour distances in one dimension completes the possibilities of nearest neighbour distance relationships pertinent to a two-dimensional form such as the retina. Because the nearest neighbour of any point in a one-dimensional pattern is just one of two possible neighbours, an exact binomial model rather than an approximate Poisson model is used. Before describing the nearest neighbour statistics involved within a one-dimensional point pattern set, a flavour of the mathematical reasoning involved in derivation is here given for just one situation: a closed curve of *unit* length along which points are randomly distributed. (Selkirk & Neave, 1984, p. 357, indicate that the situation for open

curves, including straight lines, involves a more complex application of the ideas used for closed curves, including circles.)

For any point on the curve, consider an interval centred on the point. The interval can be anywhere from zero length to the length of the curve. Say that the length of the interval is $2s$; then s is not greater than half the length of the curve. Given N points, the probability that no other point lies in the interval is $(1 - 2s)^{N-1}$. To see this, if there is just one other point somewhere on the curve then the probability that no other point lies in the interval is

$$(1 - 2s)^1 = 1 - 2s,$$

which is just the proportion of the entire curve that lies outside the interval. Let s equal a quarter the length of the curve, say, then the interval is half the length of the curve. Hence there is half a chance that no other point lies in the interval. If there are just two other points somewhere on the curve, then the probability that no other point lies in the interval is

$$(1 - .5)^2 = .25.$$

Doubling the number of other points has halved the chance that no other point lies in the interval; and so on.

Now let S be the distance from the point to the nearer of its two neighbours. The probability that $S \le s$ is equal to the probability that there is at least one other point in the interval. Hence

$$\mathrm{P}(S \le s) = 1 - (1 - 2s)^{N-1} = F(s), \qquad (12)$$

which is the cdf for S. Differentiating the cdf gives

$$f(s) = 2(N-1)(1-2s)^{N-2} \qquad (0 \le s \le 1/2),$$

which is the pdf for S.

Note that substituting 0 for s in (12) results in $\mathrm{P}(S \le s) = 0$ and substituting ½ for s in (12) results in $\mathrm{P}(S \le s) = 1$. This is embodied in the equation

$$\int_0^{1/2} f(s)\mathrm{d}s = [F(s)]_0^{1/2} = 1,$$

which shows that $f(s)$ is, indeed, a pdf.

The expected value of S is then the integral from 0 to ½ of the expression $s \times f(s)$. This is embodied in the equation

$$\int_0^{1/2} sf(s)\mathrm{d}s = 1/2N .$$

The same follows for each of the N points: each has an expected nearest neighbour distance of $1/2N$. Hence the expected sum of nearest neighbour distances equals $N \times 1/2N = ½$. Note that if the closed curve is not of unit length then the expected sum of nearest neighbour distances equals half the length of the closed curve. And, of course, the expected mean of nearest neighbour distances equals (½ × the length) / N. Note also that the result for the sum is

independent of *N.* Analyses with equivalent outcomes can be performed by statistics on the sum instead of the mean if preferred (Selkirk & Neave, 1984, pp. 356-362). The following provides a flavour of methodology involving both the sum and the mean.

Analysis of nearest neighbours in one dimension

Analysis of nearest neighbours in one dimension applies to points distributed along any line or open curve, or around a circle or closed curve. Because an open curve is bound only by its ends and a closed curve has no bound, edge correction is not the issue that it is for areal analyses. The sum of nearest neighbour distances for an empirical distribution is, of course,

$$\sum_{1}^{N} \min_{j \neq i} \{u_{ij}\} \quad ,$$

where u_{ij} is the distance from point i to point j. And the average nearest neighbour distance is, again, just the total nearest neighbour distance divided by N

$$\frac{1}{N} \sum_{1}^{N} \min_{j \neq i} \{u_{ij}\} ,$$

where u_{ij} is the distance from point i to point j. These are the observed sum and mean for an empirical distribution of points along an open or closed curve, and are again denoted r_o.

Analysis of nearest neighbour sum around a closed curve

Although this thesis often concerns mean nearest neighbour distance, in this instance using the sum of nearest neighbour distances can save a mathematical step with the same end result. As indicated above, the expected sum of nearest neighbour distances for CSR around a closed curve equals half the length of the curve. That is

$$\frac{W}{2},$$

where W is the length of the curve. This is again denoted r_e.

The two extremes for observed sums are those for which all points are superposed and all points are equally spaced along the length of the curve. For the former the nearest neighbour sum equals zero and for the latter the nearest neighbour sum equals the length of the curve. All other observed sums fall between these extremes. The standard error for the expected sum of nearest neighbour distances is given by

$$\sqrt{\frac{W^2}{6(N+1)}},$$

again denoted S_d.

These are used in the Z test.

$$Z = \frac{r_o - r_e}{S_d} .$$

If analysis by statistics on a normalized length for a closed curve is preferred, then dividing the formula for standard error by W is appropriate.

For a Z statistic that has an approximately normal distribution, N should be greater than 20, say, otherwise Z should not be used. Instead, the exact tables published in Selkirk and Neave (1984) should be consulted.[5]

The simple example given in Figure 3.11, conveniently for points dispersed around a circle, is here subjected to nearest neighbour analysis of the sum. The green links are nearest neighbour links.

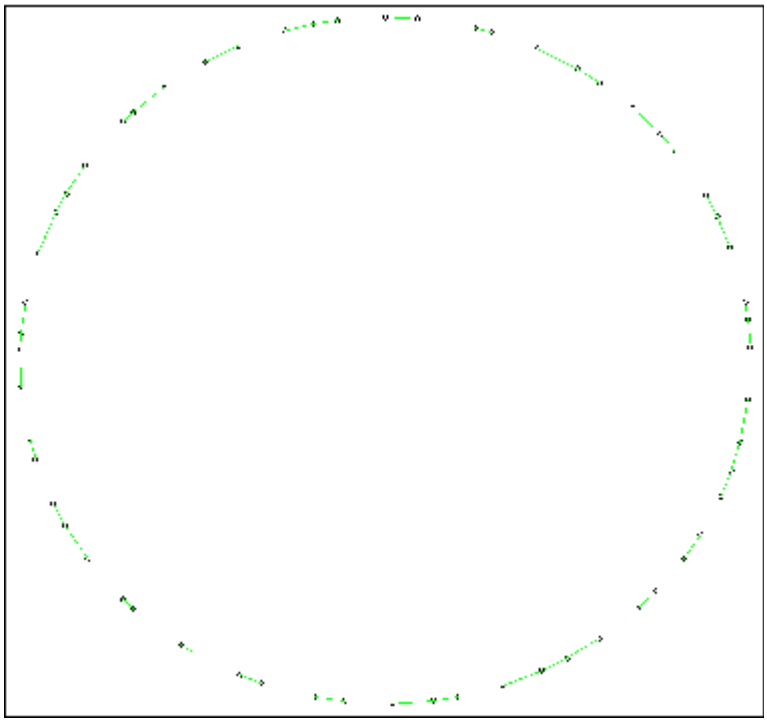


Figure 3.11: Points dispersed around a circle.

By

$$r_o = \sum_{1}^{N} \min_{j \neq i} \{u_{ij}\},$$

the observed sum of nearest neighbour distances for the pattern, shown in Figure 3.11, is 25,303 screen units. This would be normalized upon dividing by the circumference of the

[5] Instances of random point distributions containing some given $N > 20$, say, show a normal distribution about the mean $\frac{W}{2}$ for sum of nearest neighbour distances. However, progressively larger N results in decreasingly narrower normal distributions, such that as N goes to infinity then $\frac{W}{2}$ is effectively attained for the sum of nearest neighbour distances of any random point distribution. For progressively smaller N, this approximation to precision is in no way guaranteed. The distribution about the mean $\frac{W}{2}$ for sum of nearest neighbour distances becomes increasingly broader. Hence N needs to figure in the formula for standard error, even though the expected sum of nearest neighbour distances is independent of N.

circle in screen units. However, when points are arranged in a circular manner a device can be used to simplify matters: radii are imagined drawn from the centre of the circle to respective dots. The sum of nearest neighbour angles subtended at the centre of the circle, divided by 2π if working in radians or by 360 if working in degrees, produces the same result. The expected sum of nearest neighbour angles for CSR is π radians or 180 degrees, which, when divided by 2π radians or 360 degrees, correspondingly equals .500. (The same method can be also used if statistics on directions from a location are required.)

For the example, the observed sum of nearest neighbour angles is .725; which is also the sum of nearest neighbour distances on the normalized circle. By

$$\sqrt{\frac{1}{6(N+1)}}$$

the standard error is

$$\sqrt{\frac{1}{6(60+1)}} = .052$$

and

$$Z = \frac{r_o - r_e}{S_d} = \frac{.725 - .500}{.052} = 4.310,$$

which exceeds the 1.96 standard deviations from tables of the normal distribution for the $\alpha =$ .05 level of significance. Hence H_0 of CSR is rejected, and because the observed mean nearest neighbour distance is greater than the expected mean, the points tend to regularity.

The nearest neighbour statistic

$$R = \frac{r_o}{r_e} = \frac{.725}{.500} = 1.451,$$

where $R = 0$ is maximum clustering, $R = 1$ is random , and $R = 2$ is maximum regularity.

Analysis of nearest neighbour mean along an open curve

For analysis of nearest neighbours along an open curve, the expected sum or mean varies for different numbers of points. Apart from this, methodology is the same as before. For an open curve, two conditions are distinguished by the formulae: points located at both ends of a curve and no point located at either end of a curve. For the former the expected mean nearest neighbour distance for CSR is given by

$$\frac{W(N+2)}{2N(N-1)},$$

where W is the length of the curve. This is again denoted r_e.

The standard error for the expected mean nearest neighbour distance is given by

$$\sqrt{\frac{W^2(2N^2+7N-36)}{12N^3(N-1)^2}},$$

again denoted, S_d.

These are used in the Z test.

$$Z = \frac{r_o - r_e}{S_d}.$$

In the case of no point located at either end of a curve the expected mean nearest neighbour distance for CSR is given by

$$\frac{W(N+2)}{2N(N+1)},$$

where W is the length of the curve. This is again denoted r_e.

The standard error for the expected mean nearest neighbour distance is given by

$$\sqrt{\frac{W^2(2N^2+17N+12)}{12N^2(N+1)^2(N+2)}},$$

again denoted, S_d.

These are used in the Z test.

$$Z = \frac{r_o - r_e}{S_d}.$$

If analyses by statistics on the sum is preferred for an open curve, then multiplying the formulae by N is appropriate. This undoes the mean. If analysis by statistics on a normalized length is preferred for an open curve, then dividing the formulae by W is appropriate.

For a Z statistic that has an approximately normal distribution, N should be greater than 20, say, in both cases, otherwise Z should not be used. Instead, the exact tables published in Selkirk and Neave (1984) should be consulted.

The simple example given in Figure 3.12, conveniently for points dispersed along a straight line, which is 153,460 screen units long, is here subjected to nearest neighbour analysis of the mean in one dimension. The green links are nearest neighbour links.

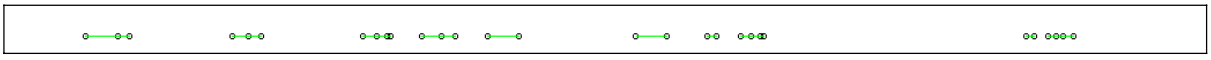


Figure 3.12: Points dispersed along a straight line.

By

$$r_o = \sum_{1}^{N} \min_{j \neq i} \{u_{ij}\},$$

the observed mean nearest neighbour distance for the pattern, shown in Figure 3.12, is 159.320 units.

By

$$\frac{W(N+2)}{2N(N+1)},$$

the expected mean nearest neighbour distance is

$$r_e = \frac{15360(30+2)}{2\times 30(30+1)} = 264.258 \text{ units.}$$

Moving on with the test for significance. Fist, by

$$\sqrt{\frac{W^2(2N^2+17N+12)}{12N^2(N+1)^2(N+2)}},$$

the standard error is

$$\sqrt{\frac{15360^2(2\times 30^2+17\times 30+12)}{12\times 30^2(30+1)^2(30+2)}} = 40.614,$$

and

$$Z = \frac{r_o - r_e}{S_d} = \frac{159.320 - 264.258}{40.614} = -2.584,$$

which exceeds the 1.96 standard deviations from tables of the normal distribution for the α = .05 level of significance. Hence H_0 of CSR is rejected, and because the observed mean nearest neighbour distance is less than the expected mean, the points tend to clustering.

The nearest neighbour statistic

$$R = \frac{r_o}{r_e} = \frac{159.320}{264.258} = 0.603,$$

where R = 0 is maximum clustering, R = 1 is random , and R = 2 is maximum regularity.

Nearest neighbour study

We have used categories of nearest neighbour analyses discussed in this chapter to model human perceptual judgment. Subject's perception of a graded range of point patterns, ranging from tightly clustered to highly regular, show that the human visual system is sensitive to degrees of dispersion of image elements in a way that closely parallels the results of nearest neighbour analysis.

Our study (Preiss & Vickers, 2005, detailed in Appendix A) employed stimuli consisting of dots distributed over areas, around circles, and along lines, each of various densities. An eleven-point scale was arranged as eleven squares abutted horizontally beneath the stimuli, and numbered consecutively from –5, through 0, to +5. From left to right, the five squares to the left of centre represented most to least clustered and the five squares to the right of centre represented least to most regular. The square at the centre represented randomness. Participants were 8 males and 8 females, aged from 19 to 52 years, drawn opportunistically

from the university student population. They were asked to respond with their subjective judgments of degree in the range from tightly clustered to highly regular.

For each stimulus, the value of the objective mean nearest neighbour distance, r_o, and of the nearest neighbour statistic, $R = r_o/r_e$, was calculated. It might be most simply supposed that participants judge the degree of clustering or regularity by estimating the objective mean nearest neighbour distance, r_o. Yet, although there were reasonably strong correlations in the case of all three pattern types (average Pearson r = .89), scattergrams of participant's linearly rescaled ratings against r_o in each case showed substantial deviations from linearity.

In contrast, Figure 4 shows the linearly rescaled mean ratings, produced by participants, for the areal, circular, and linear patterns, respectively, plotted against values of R. For all three pattern types, the relation between the ratings and R is well described by a straight line, with an intercept close to zero and a slope approaching unity. For the areal, circular, and linear pattern types, r^2 values were .97, .97 and .96, respectively. As illustrated in Figure 4d, data for all three pattern types are well described by a single linear function with an intercept of 0.12 and a slope of 0.92 (r^2=.96). It is clear that the theoretical values of R account for virtually all (at least 96%) of the variance in the empirical ratings.

In contrast, when differences between the linearly rescaled mean ratings and values of R were compared with the number of dots in each pattern, there was no significant correlation for any of the three pattern types. That is, dot density made no significant additional contribution to predicting response ratings. This reinforces the earlier finding that ratings were not satisfactorily accounted for by mean nearest neighbour distance alone.

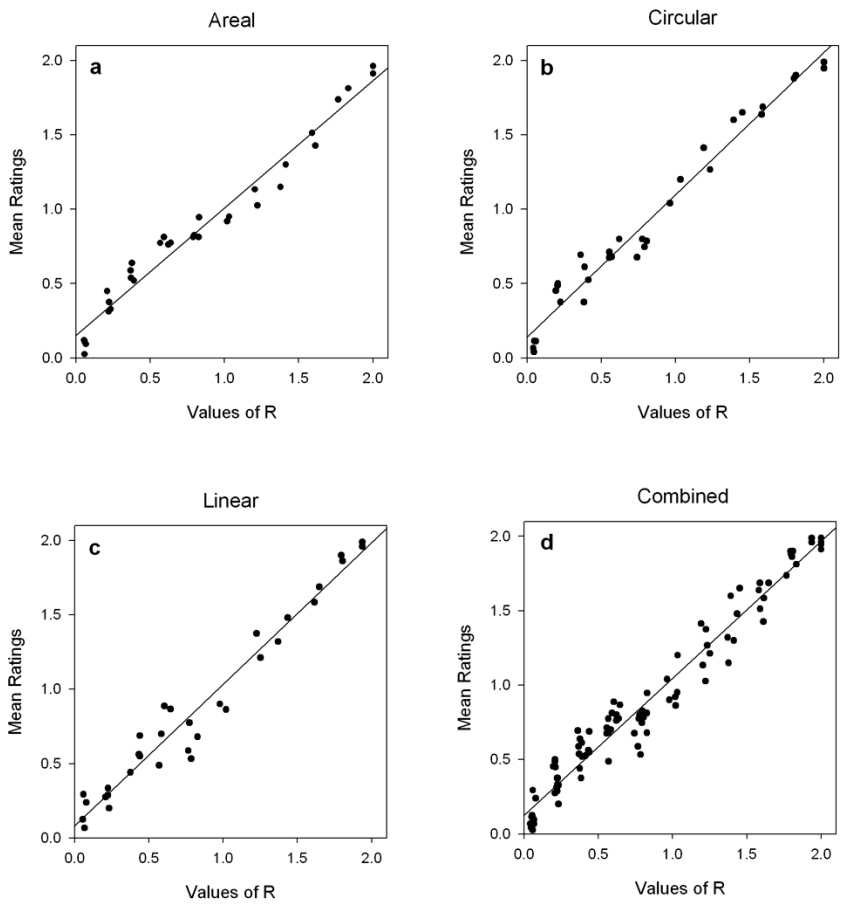


Figure 3.13: Mean subjective ratings of the degree to which patterns were perceived as clustered, random, or regular, plotted against objective values of R. From (a) to (d), respectively, the figures show the results for areal, circular, and linear patterns, and those for all three pattern types combined.

Nearest neighbour analysis and the perception of clustering, regularity, and randomness

The finding that participant's ratings of the degree of clustering, regularity, or randomness in dot patterns is not determined solely by the mean nearest neighbour distance is to be expected. For example, imagine a pattern, in which dots are located at the vertices of an (unseen) square grid, and in which the side of each grid element is one cm. We would be unlikely to judge such a pattern as more (or less) regular than a similar pattern, in which the side of each element is two cm, despite the fact that the mean nearest neighbour distance for the first pattern is half that for the second. That is, judgments of regularity are based on the structure of a pattern rather than the density of its elements. This has the advantage that such judgments can be expected to remain invariant (within broad limits) under the uniform dilation or contraction of a pattern (as well as under rotation and translation).

The finding that dot density makes no contribution to categorising the patterns as clustered or regular suggests that it would not be possible to explain the results in terms of a set of distances that exceed, or fall below, some critical absolute size. Such an explanation is, in any event, unlikely because by comparison with values of *R* for nearest neighbour distances, which in the study range from 0.05 up to 2, equivalent values for all interdot distances show minimal variation; despite extreme changes in the structure of the patterns. Equivalent values for all interdot distances are invariant with respect to the density of random dot patterns and, as illustrated in Table 1, are fairly insensitive to constraints in such patterns. Hence there seems to be little prospect for attempts to account for the operation of a Gestalt principle of proximity by means of neural structures (like those of Maloney, Mitchison, and Barlow, 1987, Wilson, Wilkinson, and Asaad, 1997, or Wilson and Wilkinson, 1998) that are sensitive to dot pairs (or dipoles) that are (anatomically) defined in terms of absolute distance.

	***R* values for nearest neighbour distances**			**Equivalent values for all interdot distances**		
	Areal	**Circular**	**Linear**	**Areal**	**Circular**	**Linear**
Most clustered	0.06	0.05	0.07	0.92	0.86	0.94
Most regular	2	2	1.94	1.02	1.01	1.15

Table 3.4: *R* values for nearest neighbour distances at most clustered and most regular employed in this experiment, compared to equivalent values at most clustered and most regular for all interdot distances.

The results are consistent with the view that participants are capable of doing exactly what they are asked to do in this experiment (i.e., assess the degree to which patterns are clustered, random, or regular; with respect to completely random patterns). Moreover, their judgments accurately reflect a statistical summary of the relational structure in each pattern. This suggests that participants not only have information about the mean nearest neighbour distance within a pattern, but that they can also calculate (or have access to) some quantity that represents the mean expected nearest neighbour distance for a random pattern with a similar number of dots. In other words, we can frame an explanation of the proximity principle at the computational level distinguished by Marr (1982). At this level, participants appear to make their judgments in a way consistent with the calculation $R = r_o/r_e$.

Chapter 4: Interpretation of Ripley's *K*-function in Spatial Point Pattern Analysis

Brief summary of chapter

Nearest neighbour analysis uses the *mean* of nearest neighbour distances in a test of significance for pattern type: it is a first order statistic. Spatial point pattern analysis, based on Ripley's *K,* uses the *variance* of point-to-point distances in analysis of pattern type: it is a second order statistic. (See Ripley, 1976, 1977.)

After the essential theoretical elements of Ripley's *K* are outlined, it is put to work in analysis of visual patterns, including Glass, clustered and regular, and Marroquin patterns. This involves graphing counts (or more specifically, sums) of pattern elements located in the ranges of incrementally increasing distance 'lags' from each pattern element. The range of inquiry can, but usually does not, cover all point-to-point distances. Edge correction techniques are also discussed, and a mathematical exposition on 'weighted edge correction' is given.

Ripley's *K* establishes useful distance limits in pattern detection, beyond which further associations of pattern elements fall below some level of significance. Needless to say, it is particularly suitable for setting distance limits in pattern detection methods that include association by distance as a parameter. Rather than suffer the overhead of dealing with all point-to-point distances, only distances out to some limit are considered. Like a good deal of early work in the thesis, the justification for including Ripley's *K* will not become wholly apparent until development of our transformational approach.

The distribution theory and statistical methodology outlined in this chapter are not original; however, their implementation along with application to configurations of interest to the study of human visual perception is original. Unlike the other computer programs used in the thesis, which were written from first principles, or written to reflect and develop our own theoretical ideas, this implementation closely follows Haase (1995).

All point-to-point distances

The geometry relevant to any pattern discriminating measure based on distance is contained somewhere within all point-to-point links. These are fundamental to distance relationships in point patterns. For completeness, all point-to-point distances are here briefly mentioned before interpretation of Ripley's *K*-function, which uses the variance of point-to-point distances.

The pdfs for all point-to-point distances for two situations involving complete spatial randomness (CSR) are shown in Figures 4.1 and 4.2.

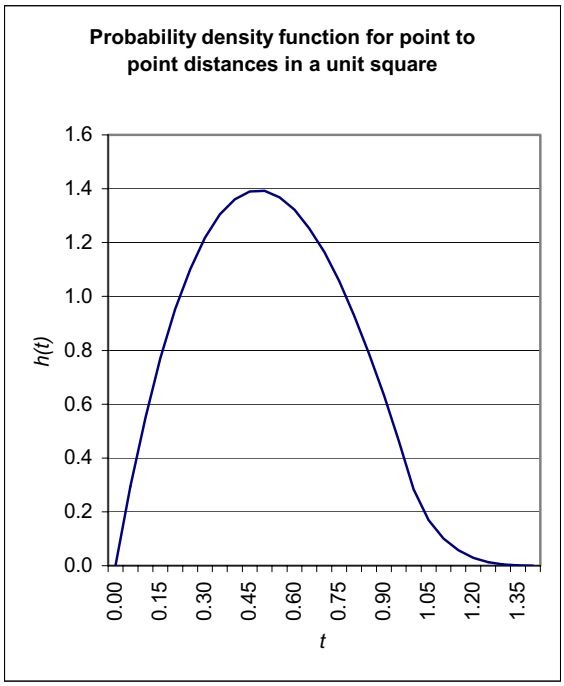


Figure 4.1: Pdf for all point-to-point distances for CSR in a square of unit side (unit square).

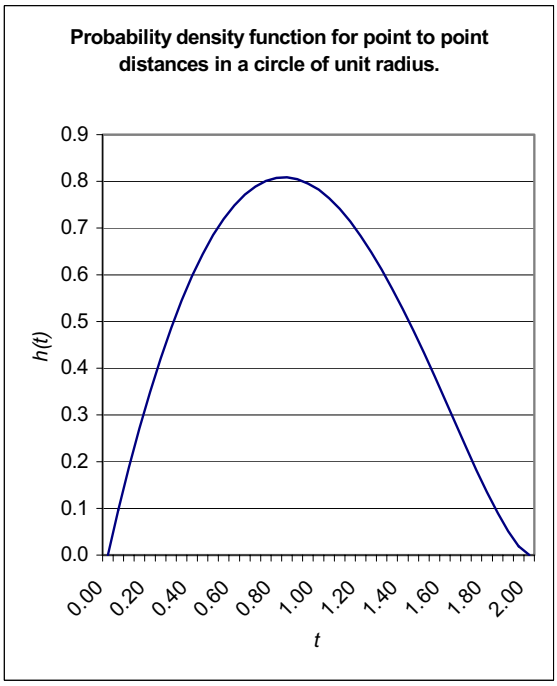


Figure 4.2: Pdf for all point-to-point distances for CSR in a circle of unit radius (unit circle).

A distribution of N points over some area has $\frac{1}{2}N(N-1)$ inter-point distances. The theoretical distribution of distances, T, between members of pairs of points for CSR depends on the size and shape of the area only. Formulae are here given for the cdfs of T for a square of unit side

$$H(t) = \begin{cases} \pi t^2 - \frac{8t^3}{3} + \frac{t^4}{2} & 0 \prec t \le 1 \\ \frac{1}{3} - 2t^2 - \frac{t^4}{2} + \frac{4(t^2-1)(2t^2+1)}{3+2t^2\sin^{-1}(2t^{-2}-1)} & 1 \prec t \le \sqrt{2} \end{cases}$$

and a circle of unit radius

$$H(t) = 1 + \pi^{-1}\left\{2(t^2-1)\cos^{-1}(t/2) - t(1+t^2/2)\sqrt{1-t^2/4}\right\} \qquad (0 \le t \le 2$$

(See Bartlett, 1964; Diggle, 1983).

Differentiating these gives the pdfs shown in Figures 4.1 and 4.2. Note that the formulae for point-to-point distances contain only references to distance. They are independent of N. The means and standard deviations of the distributions are independent of numbers of points. In other words the distributions of inter-point distances remain invariant with changes in point density. This is *not* the case for near neighbour distances.

For a square or circular area, the mean and standard deviation of point-to-point distances reduce as a function of reducing area (because pairings at greater distance intervals generally cannot be as far apart). The mean and standard deviation of near neighbour distances reduce as a function of increasing density (points per unit area). For a given number

of points they reduce with reducing area, but for a given area they also reduce with increasing number of points.

Density

For N points distributed in an area, A, the density, here denoted λ, equals N/A. In Figure 4.3, for example, there are 10 points in an area of 11 units × 9.5 units = 104.5 square units: $\lambda = 10/104.5 = 0.096$. The density is 0.096 of a point per square unit, which, since points are discrete, does not seem to make much sense. Put another way, this says that there is one point to 1/0.096 = 10.450 square units, on average. Had there been 200 points, for example, in Figure 4.3, the density would have been 200/104.5 = 1.914 points per square unit.

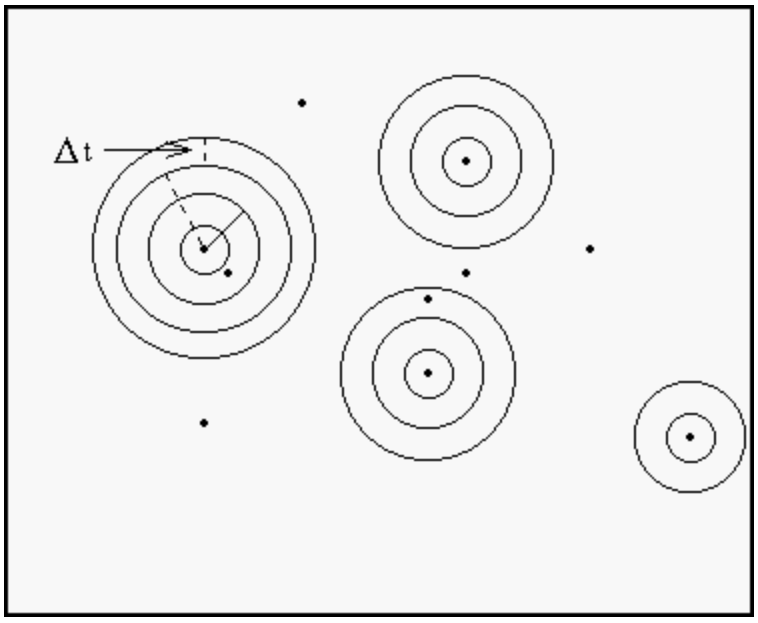


Figure 4.3: Ten points in an area 11 units × 9.5 units. The circles show distance increments from just four of the points, by way of example.

If points are Poisson (randomly) distributed, the expected, or theoretical, value for the number of points in a circle with radius t is $\lambda\pi t^2$. To borrow from the example given above for density, if the circle just happened to have a radius, t, equal to 1.824 units, for which its area, πt^2, equals 10.450 square units, then $\lambda\pi t^2 = 0.096 \times 10.450 = 1$. The circle is expected to have one point.

Ripley's *K* and the Poisson distribution

Ripley's K is a function of incrementally increasing distance lags, represented by the variable t. For the Poisson distribution $K(t) = \pi t^2$. A plot of $\sqrt{K(t)}$, which equals $\sqrt{\pi t^2}$, against t is linear. (The square-root transformation was suggested by Besag in the discussion of Ripley, 1977). The expression $\sqrt{\pi t^2} = \sqrt{\pi} \times t$ is a constant multiplied by t, which, plotted against t, is linear, with Y-intercept $\sqrt{\pi}$. Dividing by $\sqrt{\pi}$ gives $\sqrt{K(t)/\pi}$, for which the Y-intercept is zero. Subtracting t gives $\sqrt{K(t)/\pi} - t$, for which the plot is zero for all t. The mathematics provides a theoretical reference of zero for all t along the x-axis. This is a normalization related to the theoretical number of randomly distributed points in an area, by which distributions of estimates can be compared. (See Ripley, 1976, 1981)

Calculating distances in a plane

Positions of points in a plane are usually defined by rectangular coordinates *(x, y),* for which the distance between two points is $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. Distance between any point, *i,* and any other point, *j,* denoted $u_{ij,}$ is then $\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. Distances are calculated for all combinations where $i \neq j$, then sorted into ascending order and saved for further reference. (Note that distance from point *i* to point *j* is equal to distance from point *j* to point *i.* Inter-point distances are calculated twice because edge correction can prescribe different weightings to each; about which more later.)

Algorithm using Ripley's *K*

Distance *t* is set to an appropriate, small, initial value, and the number of inter-point distances not greater than *t* are counted from the sorted list. Distance *t* is then incremented by the amount of its initial value. Again, the number of inter-point distances not greater than *t* are counted from the sorted list, and so on, for equal increments of *t* that take it out to some value yet to be considered.

The process results in an array of count values: one value for each distance taken by *t*. Distance resolution is determined by the researcher's choice of distance increments for *t;* with greater resolution—along with greater processing overhead—for smaller increments.

Distance *t* can be considered a radius to a circle centred on each point in turn, from which point the distances to the other points are measured. As *t* is incremented, so the circle gets larger. Again, see Figure 4.3 above. For each other point in the circle or on the circumference, the counter for the particular radius is incremented by one.

The procedure effectively produces a cumulative distribution function constituted of estimates at intervals of Δt. In this sense it corresponds to refined nearest neighbour analysis. The cumulative distribution of estimates is effectively compared for significance with corresponding points on the theoretical cumulative distribution for random points, but Ripley's *K* includes the normalization described above.

Estimator for *K(t)*

Denote a value $I_t(u_{ij}),$ for which I_t is the count belonging to a value of *t* for the distance comparisons $u_{ij} \leq t$. The expression for the mean count of inter-point distances not greater than the value of *t* is

$$\frac{1}{N} \sum\sum_{i \neq j} I_t(u_{ij}), \qquad (1)$$

where *N* is the number of points in a study area. Here, the double summation means that for each point *i* there is a summation operation on every other point *j.* (For each member of the range of *i* the whole range of *j* not equal to *i* is invoked.) Expression (1) gives the average number of points over the areas of circles of radius *t.* It is the empirical equivalent of the Poisson-related function $\lambda \pi t^2$, hence it is necessary to divide by λ to get the estimator $\hat{K}(t)$. Dividing by λ means multiplying $\frac{1}{N}$ in expression (1) by $\frac{A}{N}$, where *A* is the study area. Inclusion of a weighting factor, $w_{ij,}$ which corrects for edge effects (discussed later), results in the estimator

$$\hat{K}(t) = \frac{1}{N^2} A \sum\sum_{i \neq j} \frac{1}{w_{ij}} I_t(u_{ij}) . \qquad (2)$$

$\hat{K}(t)$ is an approximately unbiased estimator for $K(t)$, and was given by Ripley (1976, 1981). (See, also, Diggle, 1983; Haase, 1995, 2000.)

Spatial point pattern analysis (SPPA) plot

To appreciate the relationship of $\hat{K}(t)$ with $K(t)$ consider the manner in which $\lambda \pi t^2$ increases as a function of *t.* The symbol λ is a constant peculiar to each study area, and so is π a constant, therefore the manner in which the number of points increases over increasing areas for the Poisson distribution is shown by the squared function. Yet for an empirical situation the manner in which the number of points increases over increasing areas need not correspond to this squared function. By comparison, it can be excessive at small *t* and deficient at larger *t*; to cite just one example.

For Ripley's *K* the normalizing expression $\sqrt{K(t)/\pi} - t$ is graphed against *t,* to give the SPPA plot. This kind of plot, as opposed to simply plotting *K(t)* against *t,* discards what is irrelevant and has better scaling resolution. If points are Poisson distributed, the resulting plot is everywhere zero. However, for a pattern of points the *K(t*) in the expression could be greater than or less than that for Poisson distributed points, hence the plot against the corresponding *t* would be greater than or less than zero.

Confidence intervals

The Monte Carlo procedure is recommended for finding a confidence interval. For a 99% confidence interval, the lowest and highest values of *K(t)* for each *t* from 99 Poisson simulations are used. On this basis, a 95% confidence interval requires 19 Poisson simulations for each *t.* If Monte Carlo is not practicable, Ripley suggests that approximate 1% and 5% significance points for populations are $1.68 \times \sqrt{A}/N$ and $1.42 \times \sqrt{A}/N$ respectively. For samples $1.68 \times \sqrt{A}/(N-1)$ and $1.42 \times \sqrt{A}/(N-1)$ should be used. Confidence intervals are based upon measures associated with the random distribution, not measures associated with an empirical distribution that is being tested against such randomness. They provide a graphic reference concerning significance for empirical measures.

Examples

For a distribution that is uniformly random, $\sqrt{K(t)/\pi} - t$ plotted against *t* is expected to fall within the confidence interval about zero. Figure 4.4 shows 400 such randomly distributed points and Figure 4.5 shows the corresponding SPPA plot. In all the examples, $\sqrt{K(t)/\pi} - t$ plotted against *t* is shown in red, and the upper and lower confidence interval delimiters are shown in green and blue respectively.

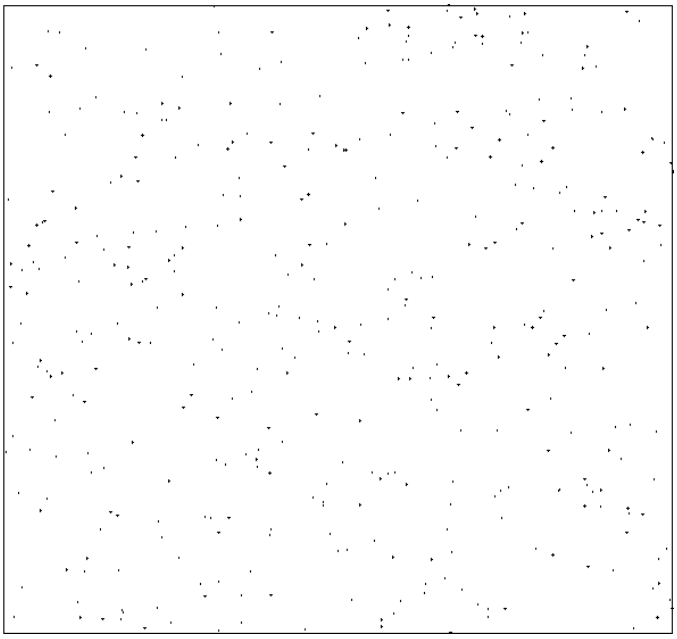


Figure 4.4: Random point pattern, from which Figure 4.5 was derived.

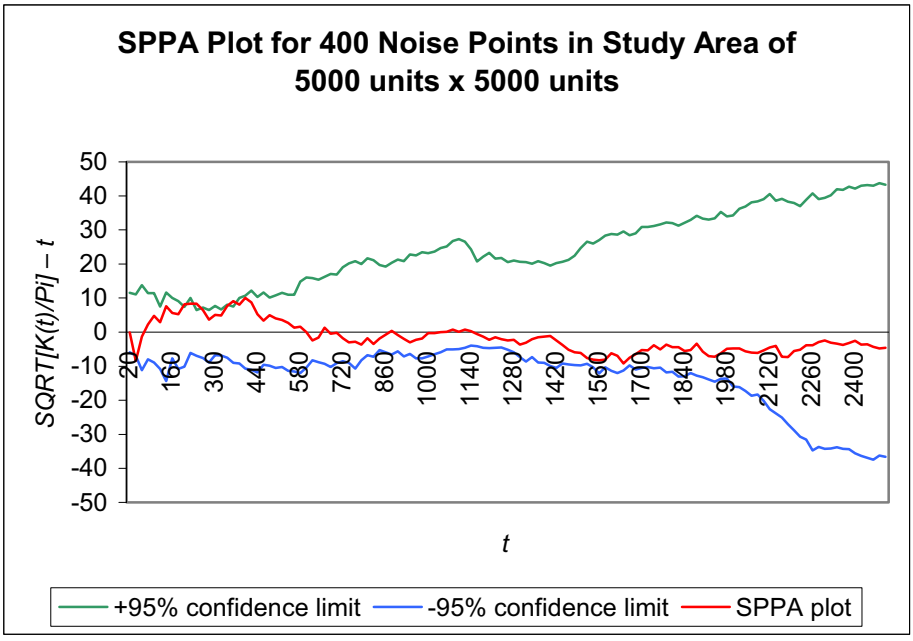


Figure 4.5: SPPA plot (shown in red) for the 400 randomly distributed points shown in Figure 4.4. Plotting was performed for distance lags out to half the length of a side of Figure 4.4. In keeping with the expectation that the plot falls within the confidence interval (delimited by the green and blue plots) no structure is evident.

For a distribution in which there is clustering, i.e. an excess of small distances, $\sqrt{K(t)/\pi} - t$ is high (above the positive confidence limit) for small *t,* and for a distribution in which there is regularity, i.e. a deficit of small distances, $\sqrt{K(t)/\pi} - t$ is low (below the negative confidence limit) for small *t.* See Figures 4.6 and 4.8, along with their counterparts, Figures 4.7 and 4.9.

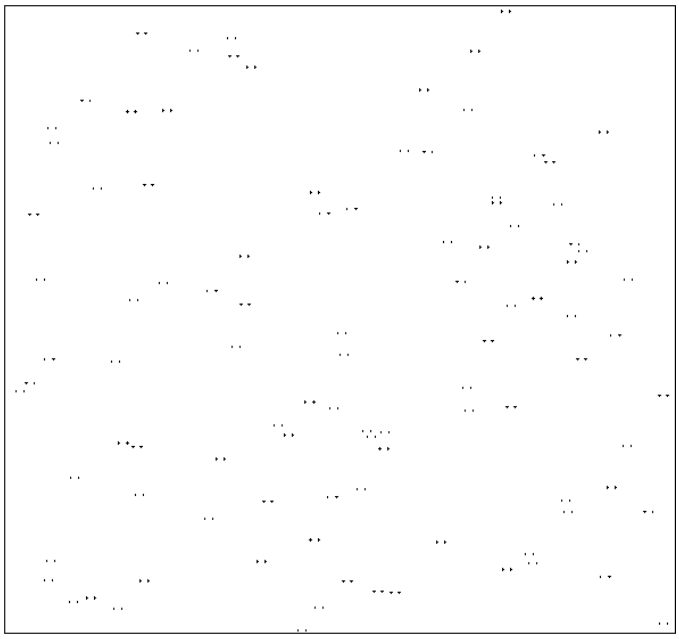


Figure 4.6: Horizontally translated Glass pattern, from which Figure 4.7 was derived. (The Glass pattern is that shown in Figure 3.3, for which a nearest neighbour analysis of the mean was performed.)

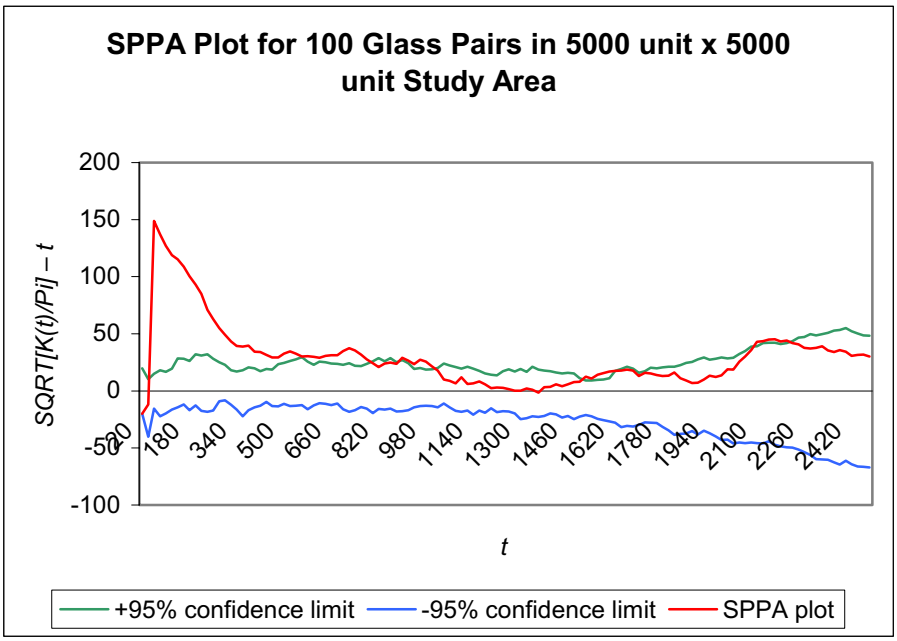


Figure 4.7: SPPA plot for the100 horizontally translated Glass pairs shown in Figure 4.6. Clustering is evidenced at relatively small distance lags by the part of the red plot that is well above the upper confidence delimiter. The clustering corresponding to displacement of elements of Glass pairs is at the distance lag indicated by the sharp peak.

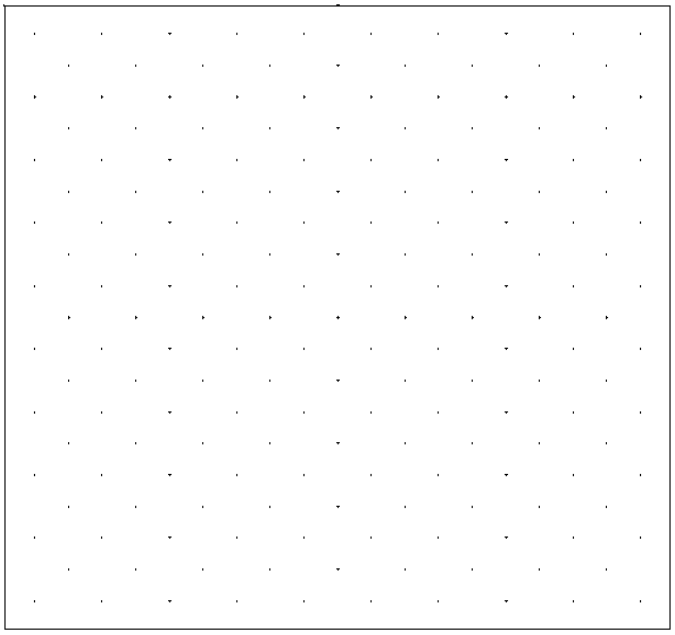


Figure 4.8: The most regular possible pattern, from which Figure 4.9 was derived.

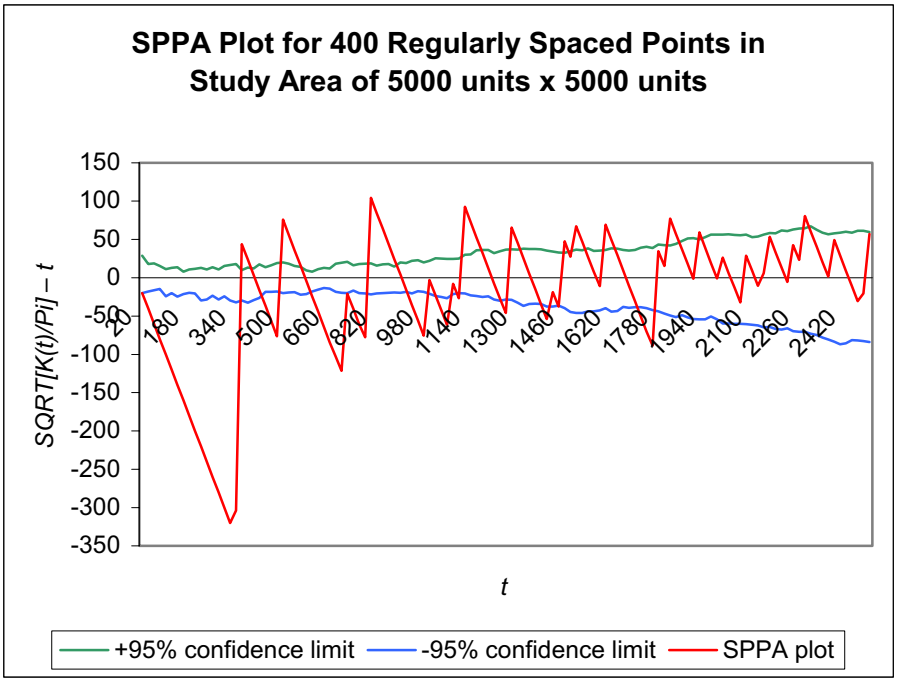


Figure 4.9: SPPA plot for the most regular possible pattern, represented by the triangular lattice arrangement shown in Figure 4.8. Note the repeating tendency over the distance lags. Regularity is pronounced at relatively small distance lags related to the least inter-point distances, and then repeats with diminishing tendency at increasing distance. The alternations indicate more points than expected at some regular distances followed closely by less points than expected at intervening regular distances.

Figure 4.10 exhibits clustering and regularity, hence the corresponding SPPA plot shown in Figure 4.11 shows clustering at small *t* and regularity at greater *t*.

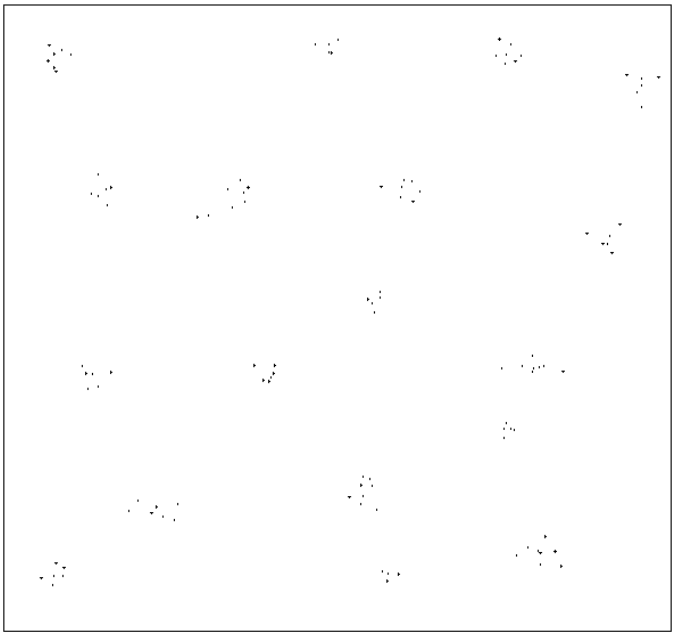


Figure 4.10: Composite pattern, with clustering at small distances and regularity at greater distances, from which Figure 4.11 was derived.

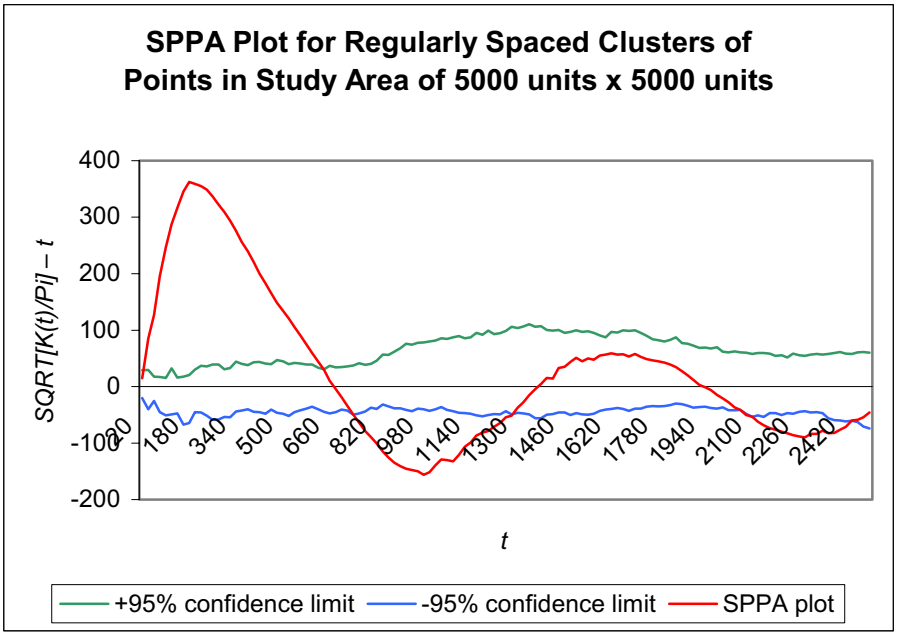


Figure 4.11: SPPA plot for the regularly spaced clusters shown in Figure 4.10. The plot shows clustering at small *t* and regularity at greater *t*. Note the repeating tendency. The regularity and clustering deviations indicate more points than expected at lags up to about an eight the length of a side of figure 4.10, followed by less points than expected at lags up to about a quarter the length of a side. The second clustering deviation is not significant. However, along with the following regularity deviation the clustering, regularity alternation that can be seen in Figure 4.10 over increasing distance is reflected.

Marroquin patterns are related to Glass patterns, only they are generated by transformations on structured arrays; such as a lattice of dots. Additionally the transformations are typically moved through any magnitude from relatively small to relatively large. See Figures 4.12 and

4.14 for Marroquin patterns produced by 16.5° and 37° of rotational offset (in which one square array of dots remained stationary while its identical other was rotated around the common centroid). A hallmark of Marroquin patterns is the alternating tendency, at increasing *t,* between clustering and regularity. This is shown in Figures 4.13 and 4.15 for the two Marroquin patterns. (In the next chapter the alternation will be seen with other kinds of measurement for the same two patterns.)

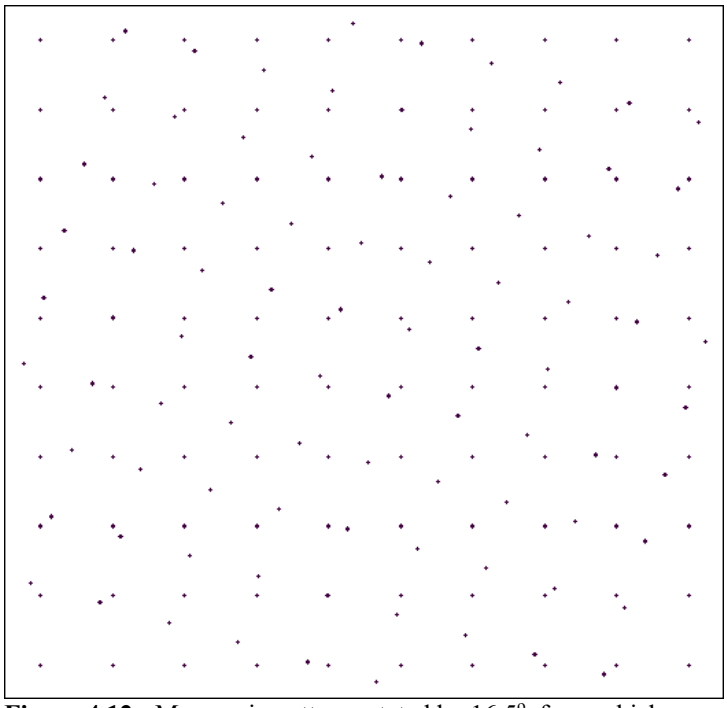


Figure 4.12: Marroquin pattern rotated by 16.5°, from which Figure 4.13 was derived.

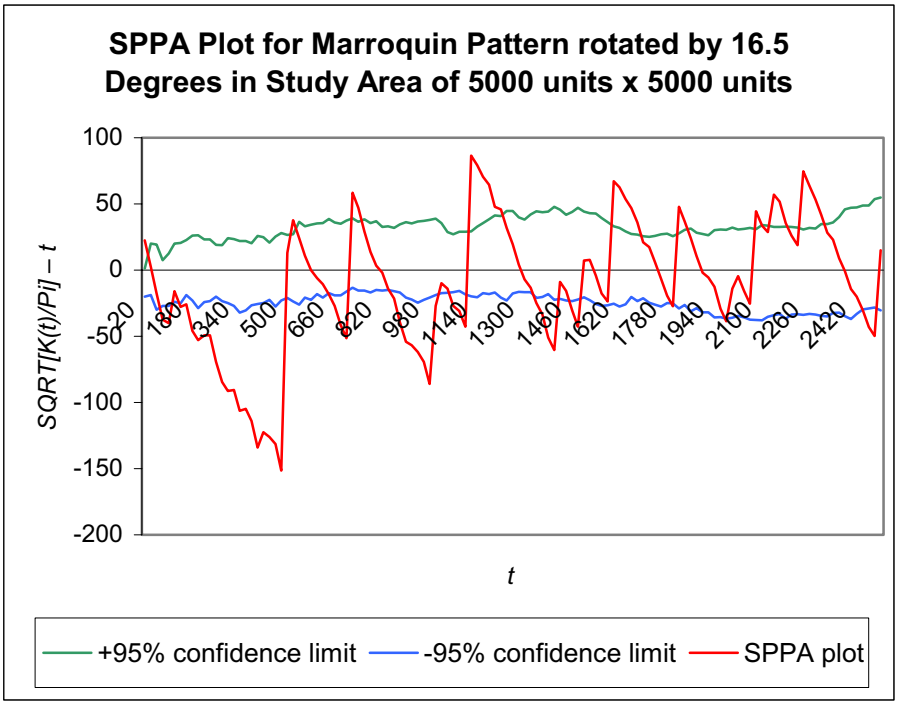


Figure 4.13: SPPA plot for the Marroquin pattern rotated by 16.5°, shown in Figure 4.12. Note the alternating tendency. The regularly spaced regularity and clustering of the pattern is evident in the plot.

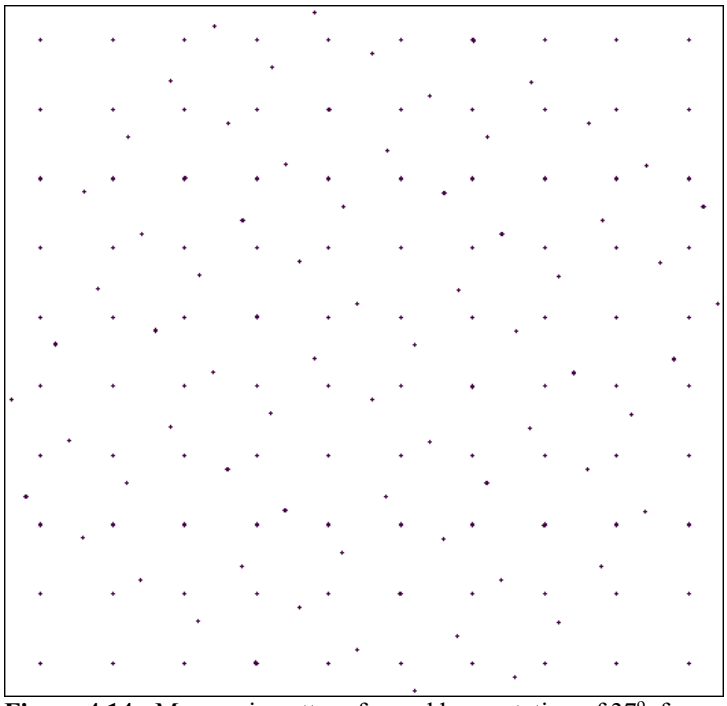


Figure 4.14: Marroquin pattern formed by a rotation of 37°, from which Figure 4.15 was derived.

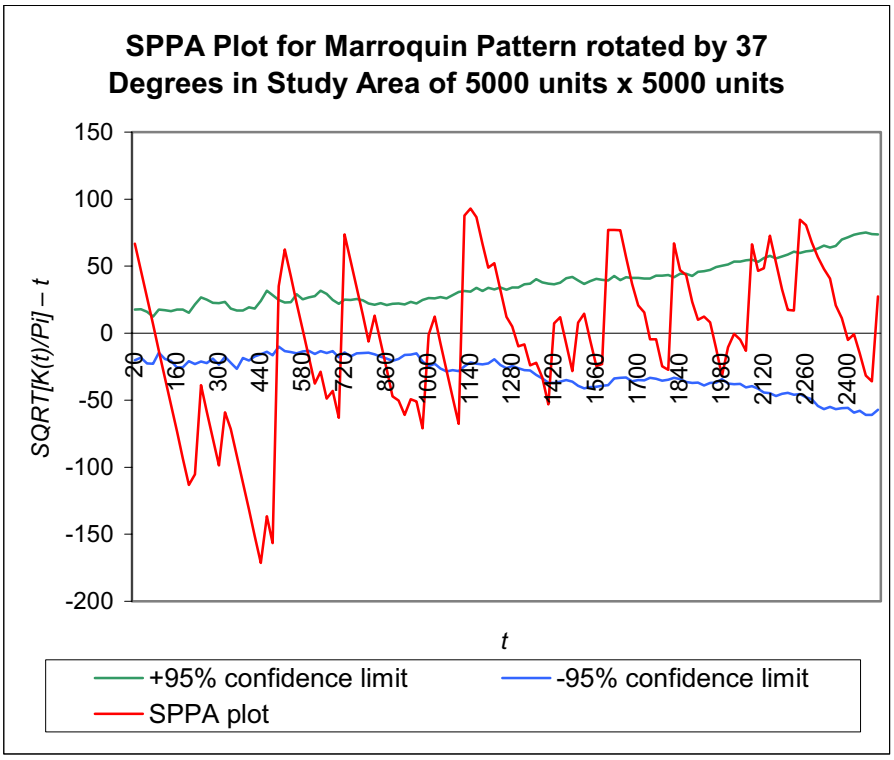


Figure 4.15: SPPA plot for the Marroquin pattern formed by a rotation of 37°, shown in Figure 4.14. Note the alternating tendency. Although rotation offset is considerably more than that for the previous Marroquin pattern, the plot is similar.

Ripley's *K* has the advantage over nearest neighbour analysis in that it provides information all the way out to a terminal distance for *t*. All point-to-point distances out to a terminal distance for *t* are used. For the two Marroquin patterns, clusterings and regularities

are repeated on increasing distance scales. This is seen by the repeating patterns of the graphs; albeit effects reduce as distances increase. This is partly due to a screening effect. Take Figure 4.11, for example. The longer distance regularity, equating to a deficit count at larger *t,* is somewhat diluted by the contribution of the shorter distance excess count for clustering. (The algorithm might be amended for some applications by discarding the cumulative distribution principle and considering the way counts change over 'bands' of distances. Of course, it then assumes the probability density principle.)

Edge correction in general

SPPA realistically applies to rectangular or circular study areas. For an ideal situation the study area needs to be embedded within a larger area, so that it is well buffered on all sides by statistically similar detail. See Figure 4.16.

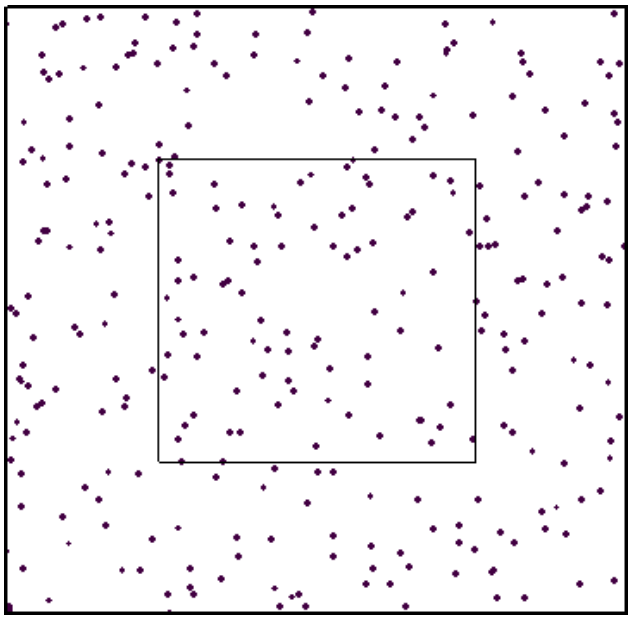


Figure 4.16: The inner area is the study area, and the outer area is the buffer zone.

For nearest neighbour analysis it is easy to appreciate the edge effect. Some points in the study area that are proximate upon the edges have their nearest neighbours outside the edges, in the buffer zone. Without a buffer zone these particular points in the study area would have been paired with respective nearest points in the study area, which would *not* have been their nearest neighbours.

For point-to-point distances, a concomitant consideration arises. Any distance from a first point in a study area to a second point in the study area that is greater than the distance from the first point to the study area boundary is problematic without a buffer zone: part of the spatial neighbourhood of the first point lies outside the study area.

A buffer zone needs to be at least equal in width to the largest value of *t*. If all point-to-point distances are to be processed, then, in an extreme case where each point of a pair of points lies in opposite corners of a study area, the buffer zone needs a width at least equal to the length of the diagonal of the study area. Since most patterns supply increasingly limited information as *t* increases, then the range of *t* (to be addressed later) is usually limited. Broadly, data for an area up to four times as big as the study area need to be recorded. This generally means that a buffer zone for Ripley's *K* needs to be appreciably bigger than a buffer zone for a typical nearest neighbour analysis.

A buffer zone is not always practicable; especially one suitable for Ripley's *K.* First it requires considerably more data recording than that for the area of interest, and second it needs the same distributional characteristics as that for the area of interest. It may be that the area of interest has a limited number of samples, and that further sampling on all sides is not an option.

Toroidal edge correction uses replicate data points across opposite edges and corners of a study area, as if the surface was rolled into one cylinder or another in which opposite edges or corners abut. (See Ripley, 1979, 1981.) This is acceptable for random distributions, but can otherwise cause errors. For example, well separated clusters that are on opposite sides of a study area come out close together, which biases the point-to-point distribution. Toroidal edge correction can give biased results for non-random patterns.

Other edge correction schemes attempt to weight calculations in such a way as to compensate for the lack of a buffer zone. A weighted edge correction scheme was described in the last chapter for nearest neighbour analysis. In that scheme, mean nearest neighbour distance was calculated using only those nearest neighbour distances from points *i* to points *j* that were less than the distances between the points *i* and their nearest boundaries. (The weighting coefficients were implicitly one or zero.) For Ripley's *K*, this method would result in penalising the distance count more with increasing *r*, which would be untenable.

Weighted edge correction

Two of the issues that have not yet been addressed are the terminal distance for *t* and the weighting factor w_{ij}. There is a connection between these, and it is convenient to deal first with weighting. A weighted edge correction scheme identified by Getis and Franklin (1987), and modified by Haase (1995, 1996), is now described for SPPA. While requiring that data for a study area alone be recorded, it assumes that the study area is bound by a larger region with the same point density and distribution.

Two basic situations arise for rectangularly bound point-to-point distances. See Figure 4.17.

1. The distance from point *i* to point *j* is greater than the distance between point *i* and the nearest boundary.

2. The distance from point *i* to point *j* is greater than either distance between point *i* and the two nearest boundaries.

For the second situation, two divisions exist.

1. The distance between point *i* and the nearest corner *is not* greater than the distance from point *i* to point *j.*

2. The distance between point *i* and the nearest corner *is* greater than the distance from point *i* to point *j*.

These situations will be addressed presently.

Although the distance from a point *i* to a point *j* is the same as the distance from point *j* to point *i,* a weighting asymmetry may apply. Part of the spatial neighbourhood at inter-point distance *ij* for point *i* can lie outside a study area while the whole of the spatial neighbourhood for point *j* can lie inside the study area, or some other part lie outside. This is why inter-point

distances are calculated twice. Counts for one or both may need modification by way of weighting.

The following three (corrected) equations are given in Haase (1995, 1996), and apply to Figure 4.17, which is a copy of his Figure 4, cases A, B, and C respectively. Weighting is based upon the proportion of the circumference of the circle, with radius u_{ij}, centred on point *i* and passing through point *j,* that lies within the study area boundaries.

- $w_{ij} = 1 - \cos^{-1}(e / u_{ij}) / \pi$
- $w_{ij} = 1 - [\cos^{-1}(e_1 / u_{ij}) + \cos^{-1}(e_2 / u_{ij}) + \pi / 2] / 2\pi$
- $w_{ij} = 1 - [2\cos^{-1}(e_1 / u_{ij}) + 2\cos^{-1}((e_2 / u_{ij})] / 2\pi$

For case A in Figure 4.17, the distance, u_{ij}, between point *i* and point *j* is greater than the distance, *e,* between point *i* and the nearest boundary. For case B, the distance between point *i* and point *j* is greater than either distance, e_1 *and* e_2, between point *i* and the two nearest boundaries. The same applies to case C, but the distance between point *i* and the nearest corner is greater than the distance between point *i* and point *j*.

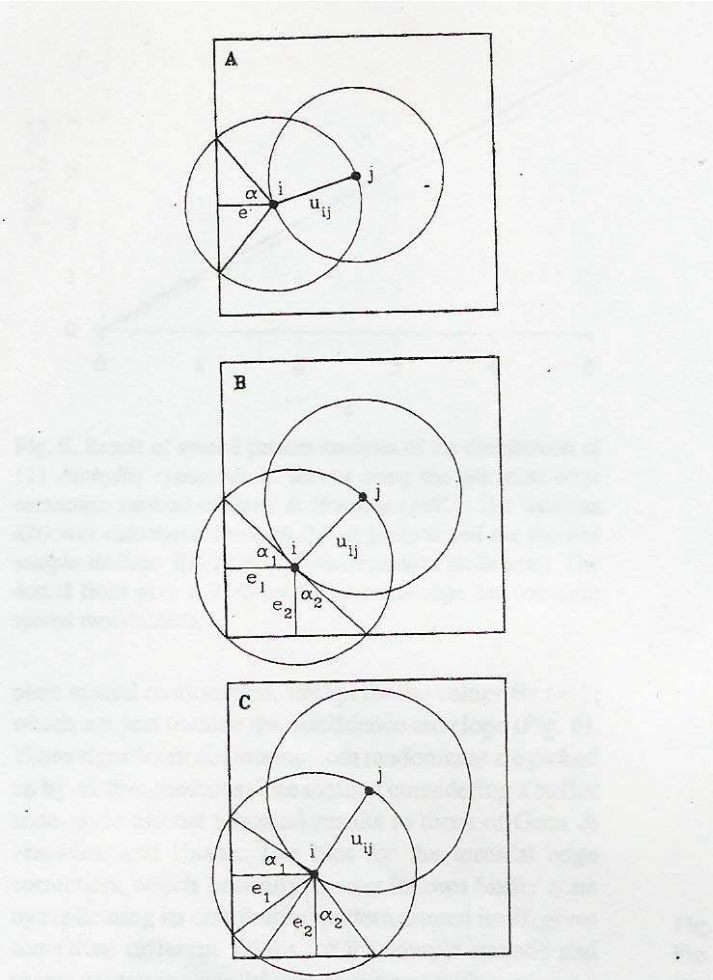


Fig. 4. Examples of three different cases to be considered for the weighted edge correction. The calculation of the parameter $I_t(u_{ij})$ in Ripley's estimator for $K(t)$ requires weighting by the proportion of the circumference of the circle with radius u_{ij} centred on point i and passing through j which lies within the plot. A, case 1 of the weighted edge correction, Eq. (3) has to be applied; B, case 2 of the weighted edge correction, equation (4) applies; C, case 3 (Haase, this paper) of the weighted edge correction, Eq. (5) applies.

Figure 4.17: Copy of Figure 4 from Haase(1995).

Unfortunately, cases B and C from the figure do not appear general. They appear to be special cases in which e_1 and e_2 are the same length. Moreover case B shows the circumference of the circle centred on point *i* grazing the corner of the study area. That is, the distance between point *i* and point *j* is the same as the distance between point *i* and the nearest corner. However, for case B the relevant part of the circumference of the circle centred on point *i* would be normally out beyond the corner.

With regard to case C, Haase (1995) emphasises his special instance of case B. He states:

> If both distances between point *i* and the two nearest boundaries are smaller than the distance between point *i* and the nearest corner…

> As an additional computational step it is necessary to calculate the distances of point *i* to all four corners of the plot. If all these distances are smaller than the distances between *i* and both nearest boundaries, e1 and e2…(p. 578)

But any two distances between point *i* and the two nearest boundaries, which correspond to the sides of a rectangle, must be shorter than their diagonal, which corresponds to the distance between point *i* and the nearest corner of the study area! The equations given in Haase (1995, 1996) still apply, however.

In the interest of generality, modified figures of Haase's (1995) cases B and C are given below, with a copy of case A included for completeness. See Figures 4.18 to 4.20, the latter two of which show unequal distances, e_1 and e_2, between point *i* and the two nearest boundaries. Figure 4.19 also shows the circumference of the circle centred on point *i* located out beyond the corner of the rectangular boundary. The geometric analysis deals with the more general situation, which includes the situation for Haase's cases B and C in Figure 4.17.

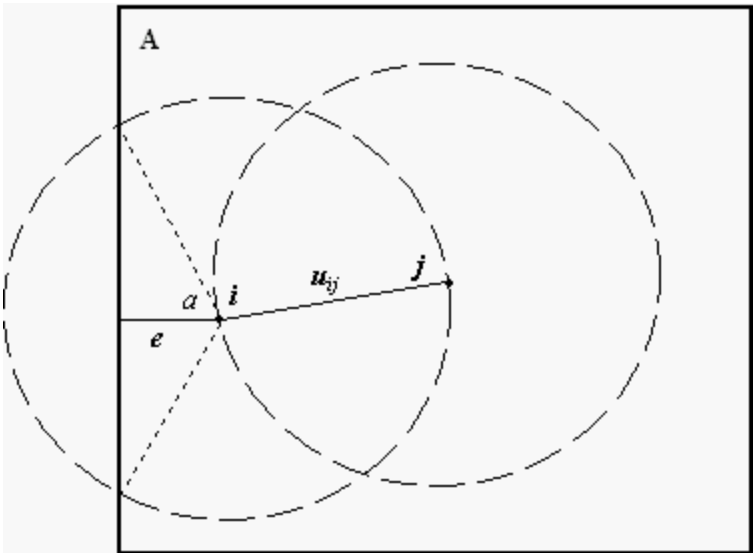


Figure 4.18: The distance, u_{ij}, from point *i* is greater than the distance between point *i* and the nearest boundary.

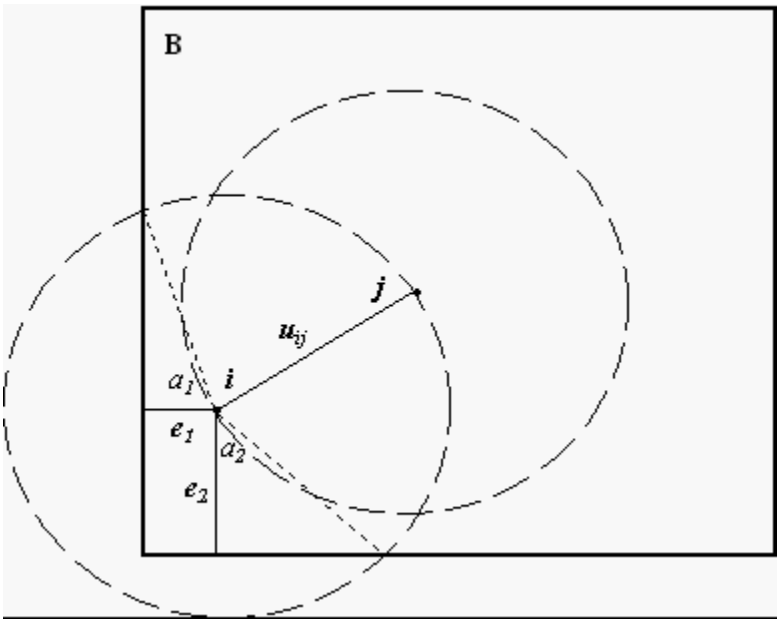


Figure 4.19: The distance, $u_{ij,}$ from point *i* is greater than either distance between point *i* and the two nearest boundaries.

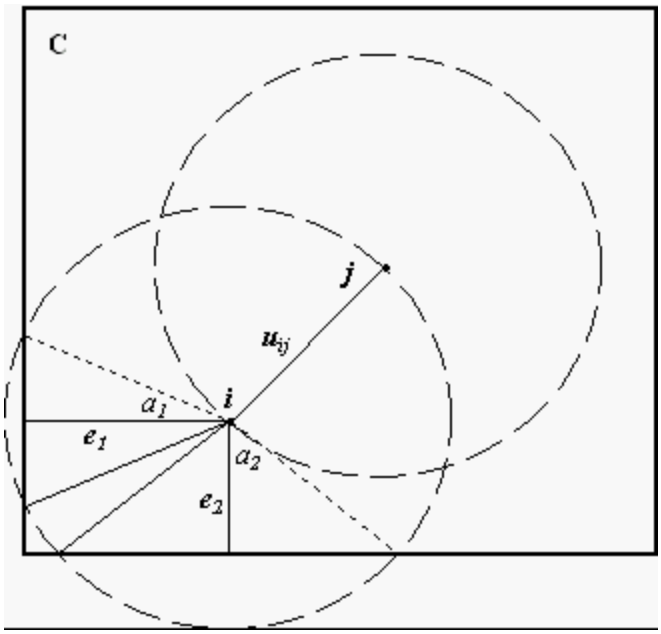


Figure 4.20: The distance, $u_{ij,}$ from point *i* is greater than either distance between point *i* and the two nearest boundaries. Additionally the distance between point *i* and the nearest corner is greater than the distance u_{ij}.

Geometric analysis

The idea of the geometric analyses is to find the proportion of the circumference of the circle that lies inside the study area. This can be accomplished by first finding the proportion that lies outside the study area. Referring to Figure 4.18 for the present, the proportion of the circumference indicated by the arc of the minor sector, within the dotted radii (same length as u_{ij}), is just $2\alpha/2\pi$ if α is in radians and $2\alpha/360$ if α is in degrees. My explanation uses radians.

Angle $\alpha = \arccos(e/u_{ij})$, which is $\cos^{-1}(e/u_{ij})$, therefore the proportion of the circumference indicated by the arc of the minor sector equals $\frac{2\cos^{-1}(e/u_{ij})}{2\pi}$, which simplifies to $\frac{\cos^{-1}(e/u_{ij})}{\pi}$. Hence the proportion of the circumference of the circle that lies inside the study area is $1-\frac{\cos^{-1}(e/u_{ij})}{\pi}$.

A similar argument applies to Figure 4.19, but on this occasion $\pi/2$ radians needs to be added to the sum of α_1 and α_2, which gives all of the angle for the sector with the arc that lies outside the study area. This, then, is $\cos^{-1}(e_1/u_{ij})+\cos^{-1}(e_2/u_{ij})+\pi/2$, and $\frac{\cos^{-1}(e_1/u_{ij})+\cos^{-1}(e_2/u_{ij})+\pi/2}{2\pi}$ gives the proportion of the circumference that lies outside the study area. Hence the proportion of the circumference that lies inside the study area is $1-\frac{\cos^{-1}(e_1/u_{ij})+\cos^{-1}(e_2/u_{ij})+\pi/2}{2\pi}$.

For Figure 4.20, two sectors with arcs that lie outside the study area are involved. The proportion of the circumference of the circle that lies inside the study area is $1-\frac{2\cos^{-1}(e_1/u_{ij})+2\cos^{-1}(e_2/u_{ij})}{2\pi}$, which simplifies to $1-\frac{\cos^{-1}(e_1/u_{ij})+\cos^{-1}(e_2/u_{ij})}{\pi}$.

This expression can be used in cases where the sum of α_1 and α_2 is not greater than $\pi/2$ radians, but would be normally used where the sum of α_1 and α_2 is less than $\pi/2$ radians. At $\pi/2$ radians it gives the same result as the expression for Figure 4.19.

The sense of weighting factor, *w*

The sense of *w* can now be appreciated. It is based upon the assumption that the region around a study area has a point density and distribution similar to nearby regions in the study area. If a point is to be some (radial) distance from another point, then, again, the probability of it being in the study area is equal to the proportion to the circumference in the study area of a circle centred on that other point.

Suppose the proportion of the spatial neighbourhood of some point *i* (at distance $u_{ij} \leq t$) is 1. A value is registered for which the weighting does not increase the count for *K(t)*. Furthermore suppose that *t* is small and that many of the proportions are 1. The sum of the weighted elements does not increase the count for *K(t)* by much.

Upon incrementing *t,* the sum for the $u_{ij} \leq$ the previous *t* remains unchanged. But suppose the proportion of the spatial neighbourhood of some point *i* (now at a distance $u_{ij} >$ previous *t* and $\leq$ current *t*) is .8, i.e. $w_{ij} = .8$, or $4/5$. Although the point *j* at distance *u* from the point *i* is in the study area, it had a prior probability of .8 for being in the study area.

Accordingly the contribution of the point *j* to the count is increased by a factor of $1/.8$, or $5/4$. Those points found further out from the point *i,* but with distance u_{ij} > previous *t* and ≤ current *t,* have their respective contributions to the count increased by more, and those points found further in, also with distance u_{ij} > previous *t* and ≤ current *t,* have their respective contributions to the count increased by less; and maybe some not at all. This compensates the count in a statistical way for the deficit of points due to the variously abbreviated spatial neighbourhoods for the u_{ij}.

Note that in the tradition of a cumulative distribution function, counts for *K(t)* at greater values of *t* include all counts for lesser values of *t.* To recap: weighting each '1' registered for $u_{ij} \leq t$, increases its value inversely as the proportion of the circumference of the circle that lies within the study area (hence elements are summed instead of counted). Additionally, a closer point $j \leq t$ from point *i* is less likely to have its '1' altered by weighting. The more distant a point $j \leq t$ from point *i,* the more likely it is to have its '1' increased, and the more likely it is that the increase is greater.

Now it is easier to see why smaller increments for *t* give finer detail on spatial variability. It equates to sample increments chosen for a cumulative distribution function so as not to miss relevant density changes.

Methodology

Haase (1995) indicates the counts are summed (upon dividing each by its w_{ij}) after checking all point-to-point distances $\leq t$ for each distance *t.* Note that the methodology described herein uses list locations for w_{ij} (weightings) assigned to corresponding sorted list locations for u_{ij} (distances). This allows a weighted summation; which is restricted to just the inter-point distances not greater than *t*. To test $u_{ij} \leq t$ for every u_{ij} would be inefficient. And since summation for any *t* includes all summations for smaller *t,* then, in practice, these are not recalculated either. The current summation simply equals the previous (running) summation plus the summation for the current interval. Lastly, it is worth remembering that summation is over each and every point, out to other points not further than distance *t.*

Terminal value for *t*

Since methods of edge correction can relate to terminal value for *t* it is necessary to address this issue. Haase (1995, p. 577) reports that the edge correction handled by the three equations above gives approximately unbiased values for *t* up to $\sqrt{2}/2 \times$ the length of the side of a square study area, and up to $1/2$ the length of the shorter side of any other rectangular study area. This is because larger values for *t* mean that circles centred on points near the middle of the study area may intersect three boundaries.

In connection with patterns of vegetation, Ripley suggests $\sqrt{A/2}$ as the upper limit for *t,* where *A* is the area of the study region. This might be the case for many patterns, even where a maximum buffer zone could be employed, because most patterns typically supply increasingly limited information as *t* increases.

Chapter 5: Density Dependent Versus Density Free Analysis

Brief summary of chapter

Nearest neighbour analyses are density dependent statistical methods for differentiating point pattern types. Alternatively, statistical operations on distributions of least angles of Delaunay triangles afford an effective, density free method for differentiating point patterns.[1] Nearest neighbour distributions for various visual patterns are compared with corresponding distributions of smallest angles for respective Delaunay triangles. Edge-to-edge and edge-to-point ratios involving Delaunay triangles are also examined for ability to differentiate point patterns. The patterns include Glass, clustered and regular, and Marroquin patterns, as well as patterns with reflection symmetry.

Reflection symmetry, delineated by the Voronoi diagram and involving medial axes, is examined with a view to the way that it might form an 'inductive bias' for visual perception.

The distribution theory for least angles outlined in this chapter is not original. However, the implementation and application, along with that of nearest neighbours, involving configurations of interest to the study of human visual perception, are original. Observations regarding proportions and ratios of Delaunay edges are also original. Medial axes, of themselves, are not original; however, their implementation and application in this chapter are original.

Discriminating point patterns

Again, nearest neighbour analyses are density dependent statistical methods for differentiating point pattern types, ranging from degrees of clustering, through randomness, through degrees of regularity. Choice of study area boundaries affects outcomes. Density—number of points per unit area—is related to choice of boundaries for the area calculation. Statistical operations on distributions of least angles of Delaunay triangles provide a density free method for differentiating point patterns.

Furthermore the ratio formed by the number of direct Delaunay edges to the number of all Delaunay edges, appears to differentiate point patterns. My investigation shows that different point patterns can be formulated to give any number of Delaunay neighbours within sets, on top of which ratios can range from a lower bound of approximately .4 to an upper bound of 1.

Distribution of least angles

Figures 5.1 and 5.2 show the pdf and cdf for least angles of Delaunay triangles for complete spatial randomness (CSR). The least angle of a Delaunay triangle is here designated α. The marginal density of least angle was derived by Mardia, Edwards, and Puri (1977).

Since the least angle of any triangle cannot be greater than 60^{o}, or $\pi/3$ radians, which is approximately 1.05 radians, then this constitutes the upper limit of the scale for α. The expression for the probability density function of α is

$$\frac{2[\cos 2\alpha - \cos 4\alpha + (\pi - 3\alpha)\sin 2\alpha]}{\pi} \qquad 0 \prec \alpha \leq \pi/3,$$

and is denoted $f(\alpha)$.

---

[1] Distribution of least angles: a frequency count, in the form of a histogram, for the smallest angle in each Delaunay triangle.

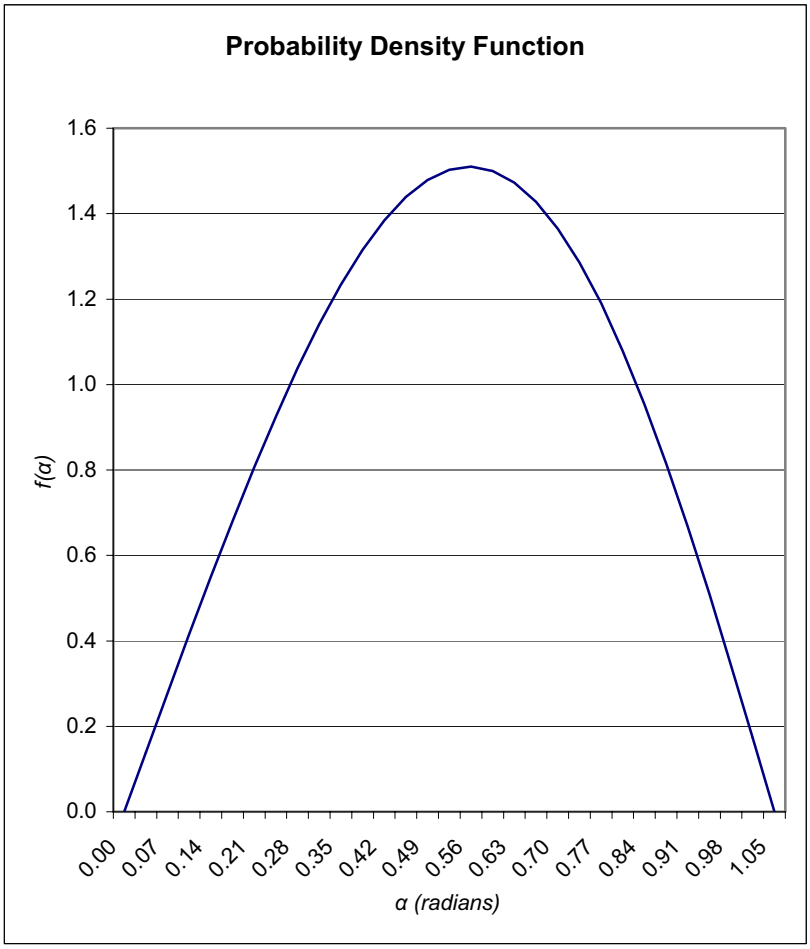


Figure 5.1: Pdf for least angles of Delaunay triangles for CSR.

The indefinite integral of the formula for the pdf gives the formula for the cdf. The expression for the continuous cumulative distribution function of α is then

$$1+\frac{2[-\frac{1}{2}(\pi-3\alpha)\cos 2\alpha-\frac{1}{4}\sin 2\alpha-\frac{1}{4}\sin 4\alpha]}{\pi} \qquad 0 \prec \alpha \leq \pi/3,$$

and is denoted $F(\alpha)$.

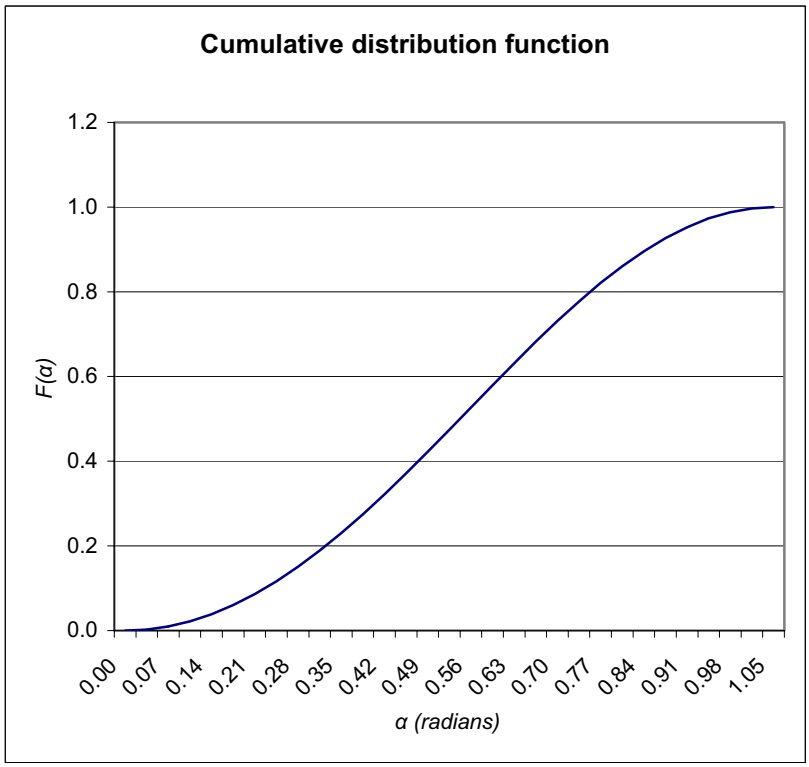


Figure 5.2: Cdf for least angles of Delaunay triangles for CSR.

The important feature of these formulae is that they have no terms for number of points or for area. Analyses based on this methodology are density free: the problem of choosing a minimal regular boundary for a set of points evaporates. This can have significant benefits, especially where there are large differences between local point densities. If an area needs to be made significantly larger to include just a few more points, the effectiveness of nearest neighbour analyses is depleted, for example.

For nearest neighbour distances, the means and standard deviations change with density, but for least angles this is a void point. There is but one mean for least angles belonging to CSR, and one standard deviation. The mean is the integral of $\alpha \times f(\alpha)$ for $0 < \alpha \leq \pi/3$, which equals 0.537 radians (or $\frac{27}{16\pi}$ radians). The variance is the integral of $(\alpha - mean)^2 \times f(\alpha)$ for $0 < \alpha \leq \pi/3$, which equals 0.052 radians (or $\frac{\sqrt{3} \times 27}{\pi \times 32} - \frac{1}{8} - (\frac{27}{16\pi})^2$ radians), and the square root gives the standard deviation as 0.227 radians. And, of course, the standard deviation associated with CSR is the standard error needed for statistical testing; as per nearest neighbour analysis. (The use in hypothesis testing follows previous methodology for nearest neighbours.)

Lastly, edge effects can be avoided by not counting angles from triangles for which vertices are common to vertices of the convex hull. My analyses with and without these angles show that edge effects are minimal for the examples employed herein. Consequently, edge effects are not considered for the following.

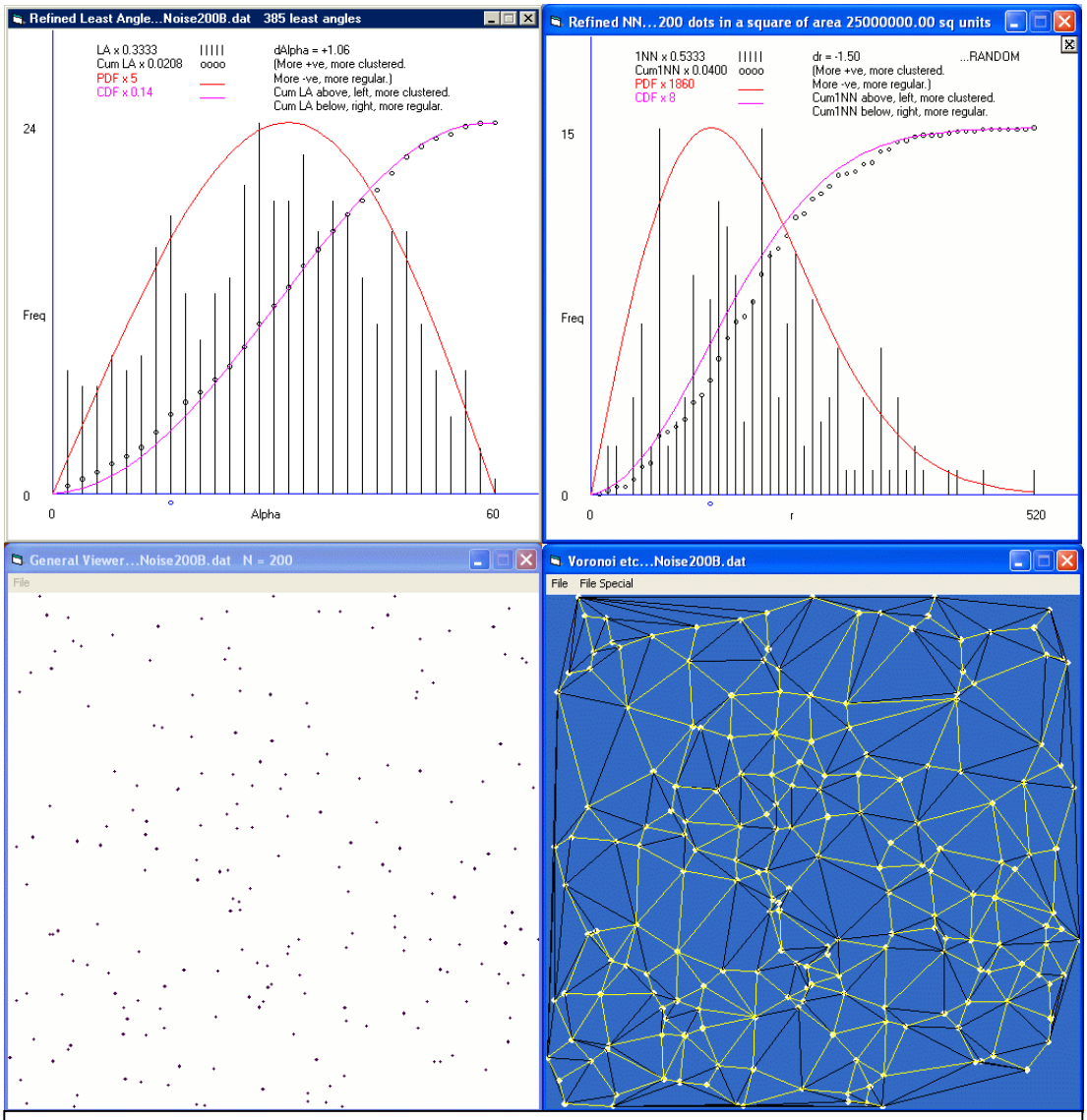
Implementation of least angle and other measures

The point of interest is the way in which the distribution functions correspond to the actual patterns, and, further, how least angle methodology overcomes some problems and ambiguities associated with nearest neighbour methodology. To this end, distribution

functions for both methods, representing a variety of point patterns, are displayed side by side in Figures 5.3 to 5.12.

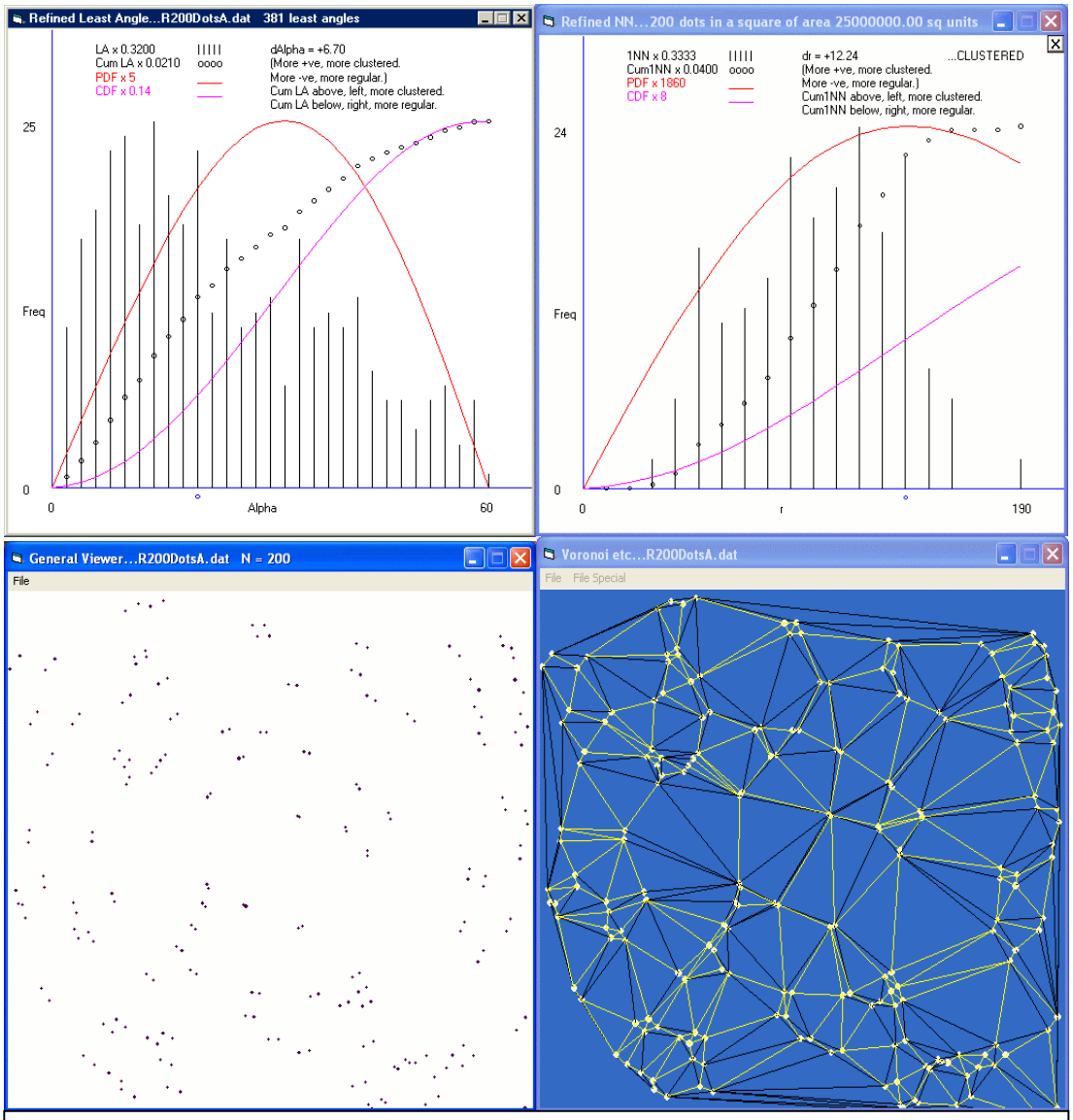
Some examples show agreement and others show grades of difference. The upper left form in each case belongs to least angle (given in degrees on this occasion), and the upper right form belongs to nearest neighbour. Each shows theoretical (Poisson) and empirical cdfs and pdfs. The pdfs are over-scaled with respect to the cdf s for sake of better viewing definition. Small circles show empirical cdfs and vertical lines empirical pdfs. Clustering is evident by the small circles above the theoretical cdf, and regularity is evident by the small circles below the cdf. The lower left form in each case shows the source pattern, and the lower right form shows the corresponding Delaunay triangulation. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges.

Keep in mind that internal Delaunay edges are doubly drawn, and edges on the convex hull are singly drawn. Keep in mind, also, that sometimes quite a few edges can be of equal length. The text box shows three ratios. The first ratio is formed by the number of *different* length *direct* Delaunay edges divided by the total number of *different* length Delaunay edges, both direct and indirect. For this ratio, all edges are treated as single links. The second ratio is formed by the number of *direct* Delaunay edges divided by the total number of Delaunay edges, both direct and indirect. For this ratio, all edges are treated as actually drawn. And the third ratio is formed by the total number of edges, treated as single links, divided by the number of points. These ratios were included in the investigation in the hope that they might discriminate between random displays and patterned displays, as well as between different types of patterns.



Proportion of different length direct Delaunay edges to total different length Delaunay edges = **0.622**
Proportion of direct Delaunay edge count to total Delaunay edge count = **0.627**
Edge to point ratio = **3.555**

Figure 5.3: Lower left: random point set. Upper left: least angle distribution functions for random point set. Upper right: nearest neighbour distribution functions for random point set. Lower right: Delaunay triangulation of random point set. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges. Least angle distributions show good agreement with nearest neighbour distributions for the random point set.



Proportion of different length direct Delaunay edges to total different length Delaunay edges = **0.630**
Proportion of direct Delaunay edge count to total Delaunay edge count = **0.619**
Edge to point ratio = **3.375**

Figure 5.4: Lower left: rotational Glass pattern. Upper left: least angle distribution functions for rotational Glass pattern. Upper right: nearest neighbour distribution functions for rotational Glass pattern. Lower right: Delaunay triangulation of rotational Glass pattern. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges. Least angle distributions agree with nearest neighbour distributions on clustering for the rotational Glass pattern, but the least angle distributions reflect the disposition of the points more accurately.

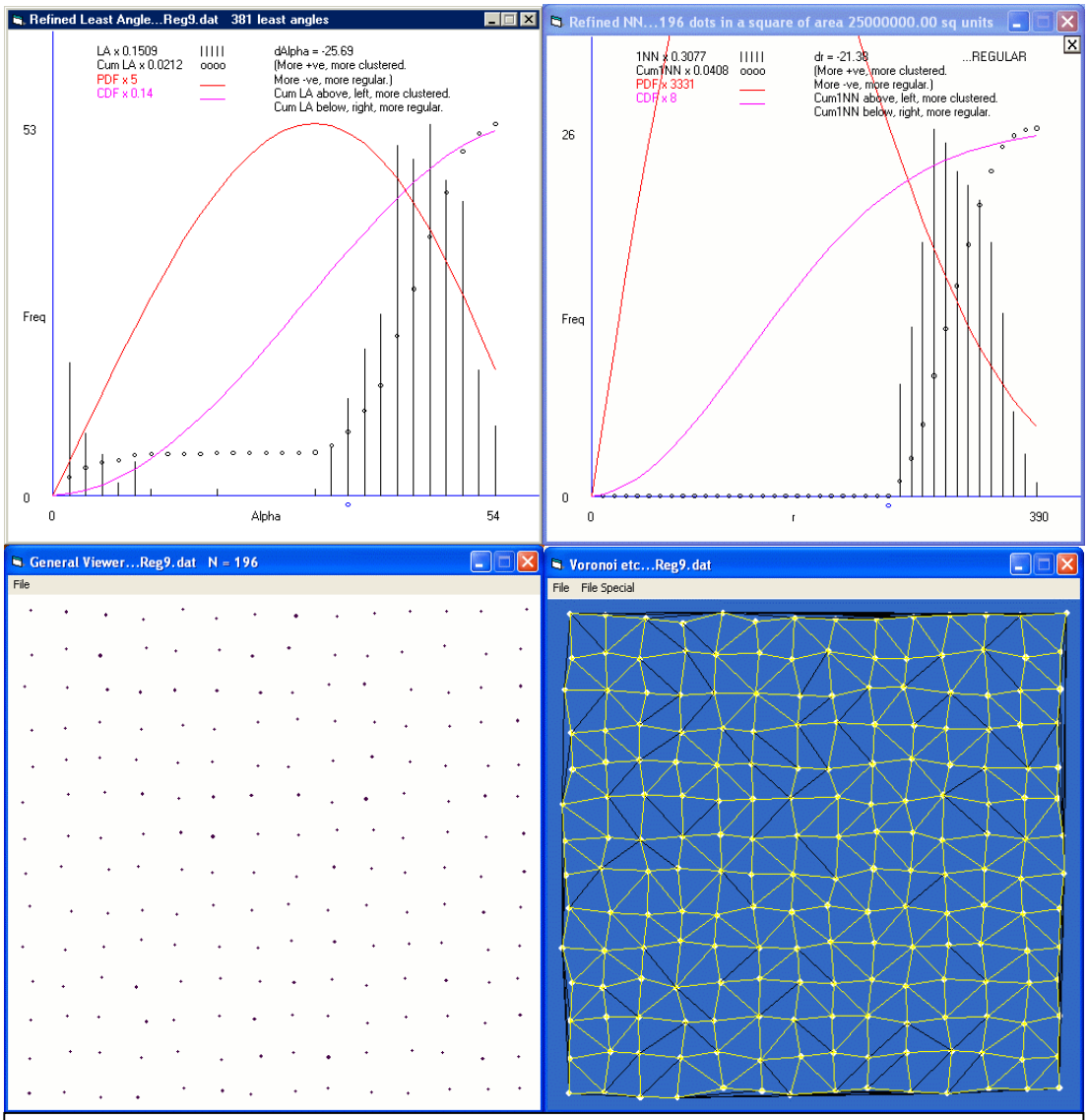
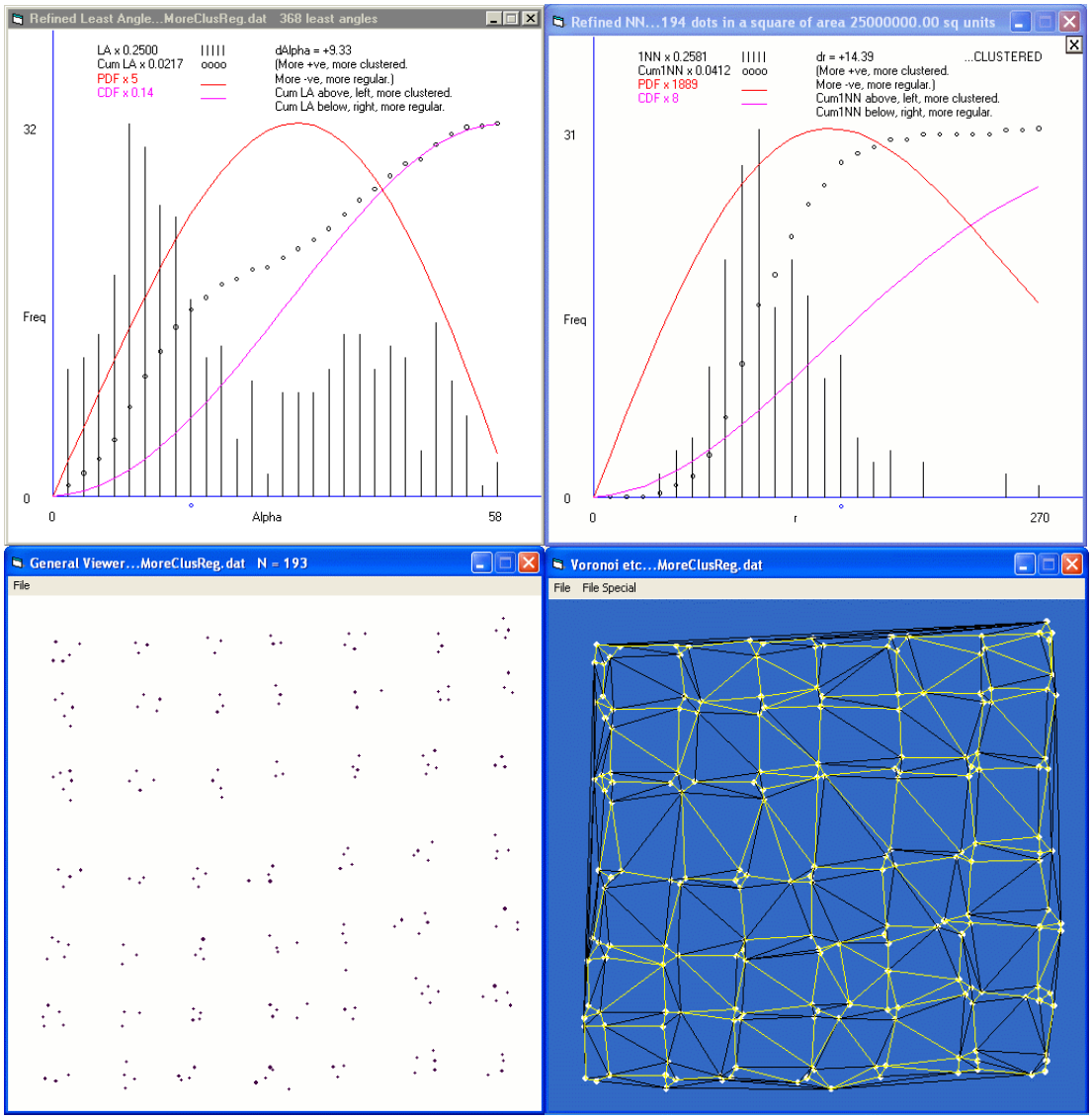
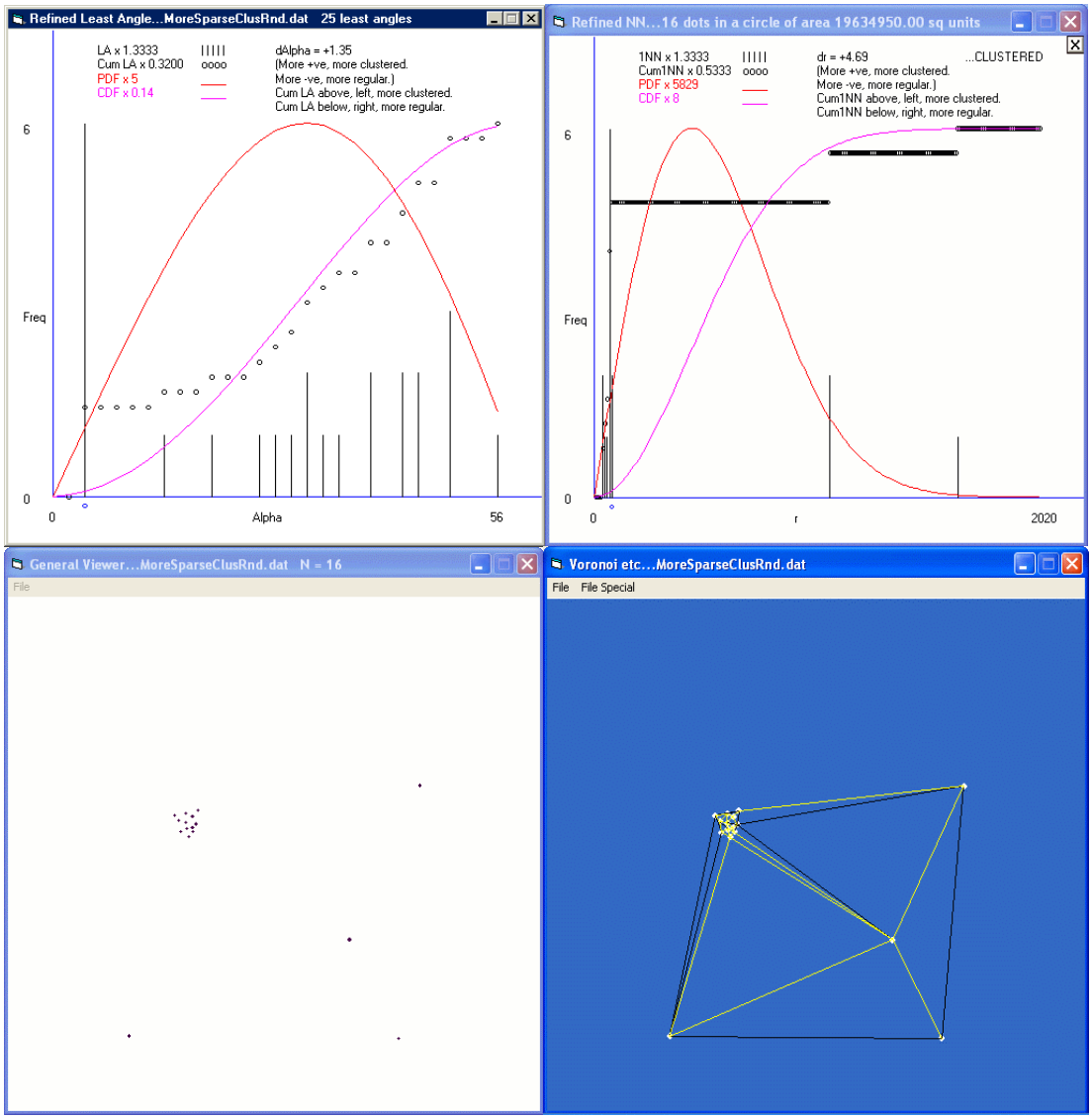


Figure 5.5: Lower left: fairly regular pattern. Upper left: least angle distribution functions for fairly regular pattern. Upper right: nearest neighbour distribution functions for fairly regular pattern. Glass pattern. Lower right: Delaunay triangulation of fairly regular pattern. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges. Least angle distributions agree with nearest neighbour distributions in a general sense on regularity for this fairly regular pattern.



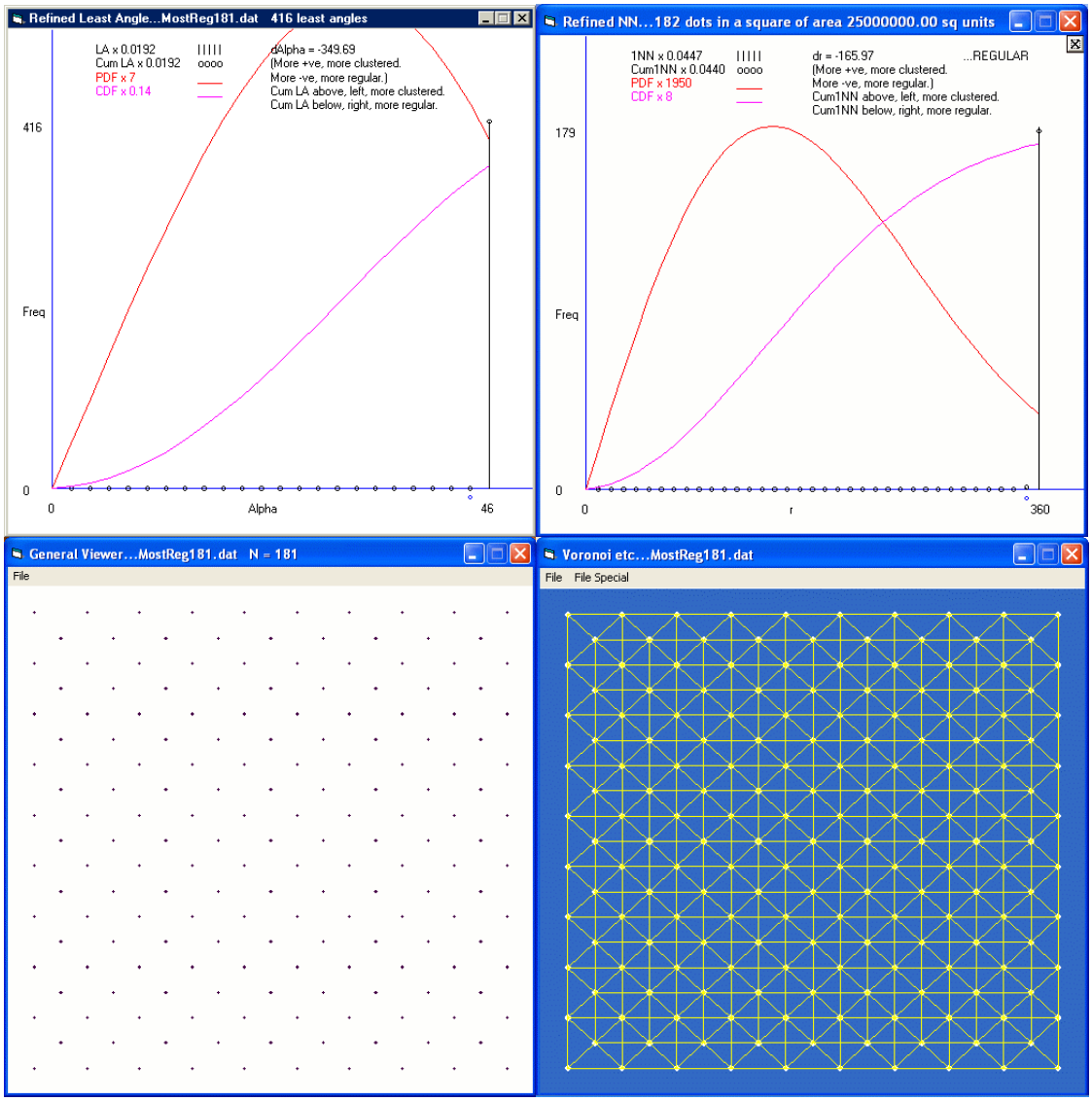
Proportion of different length direct Delaunay edges to total different length Delaunay edges = **0.592**
Proportion of direct Delaunay edge count to total Delaunay edge count = **0.594**
Edge to point ratio = **3.269**

Figure 5.6: Lower left: regularly spaced clusters. Upper left: least angle distribution functions for regularly spaced clusters. Upper right: nearest neighbour distribution functions for regularly spaced clusters. Lower right: Delaunay triangulation of regularly spaced clusters. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges. The pattern shows regularly spaced clusters, and demonstrates a situation for which least angle methodology has the advantage. The distributions for least angles nicely reveal the relative proportions of clustering and regularity, particularly with a dearth of middle angles for the empirical pdf. The local maximum for regularity in the nearest neighbour empirical pdf—above small blue circle—is not as obvious, and the nearest neighbour empirical cdf only shows clustering.



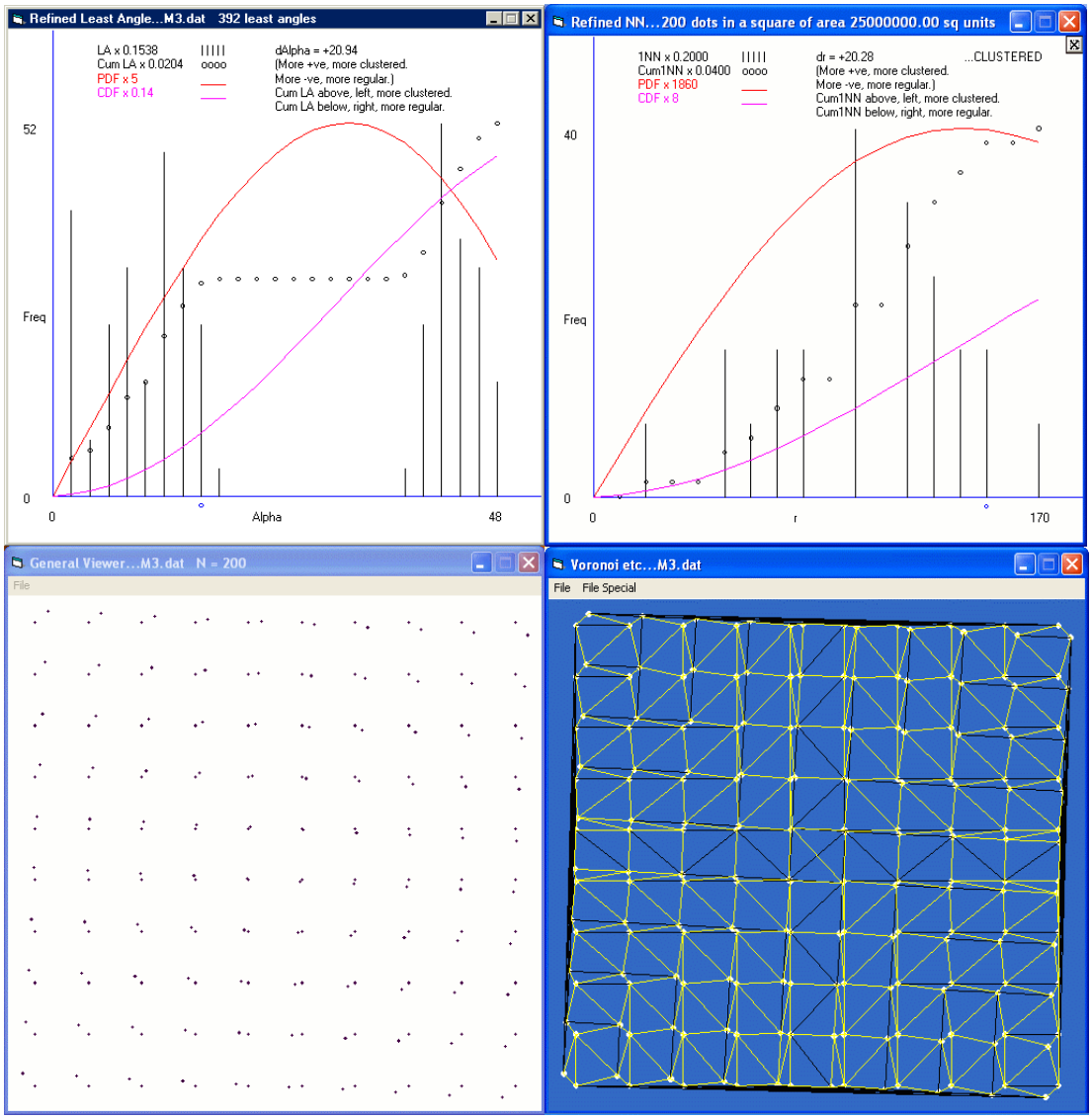
Proportion of different length direct Delaunay edges to total different length Delaunay edges = **0.675**
Proportion of direct Delaunay edge count to total Delaunay edge count = **0.707**
Edge to point ratio = **3.000**

Figure 5.7: Lower left: pattern with a mixture of attributes. Upper left: least angle distribution functions for pattern with a mixture of attributes. Upper right: nearest neighbour distribution functions for pattern with a mixture of attributes. Lower right: Delaunay triangulation of pattern with a mixture of attributes. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges. Least angle distributions clearly deal with the whole story: relative degrees of clustering, randomness, and regularity. The (density dependent) nearest neighbour distributions are not nearly as good at depicting the story under these conditions.



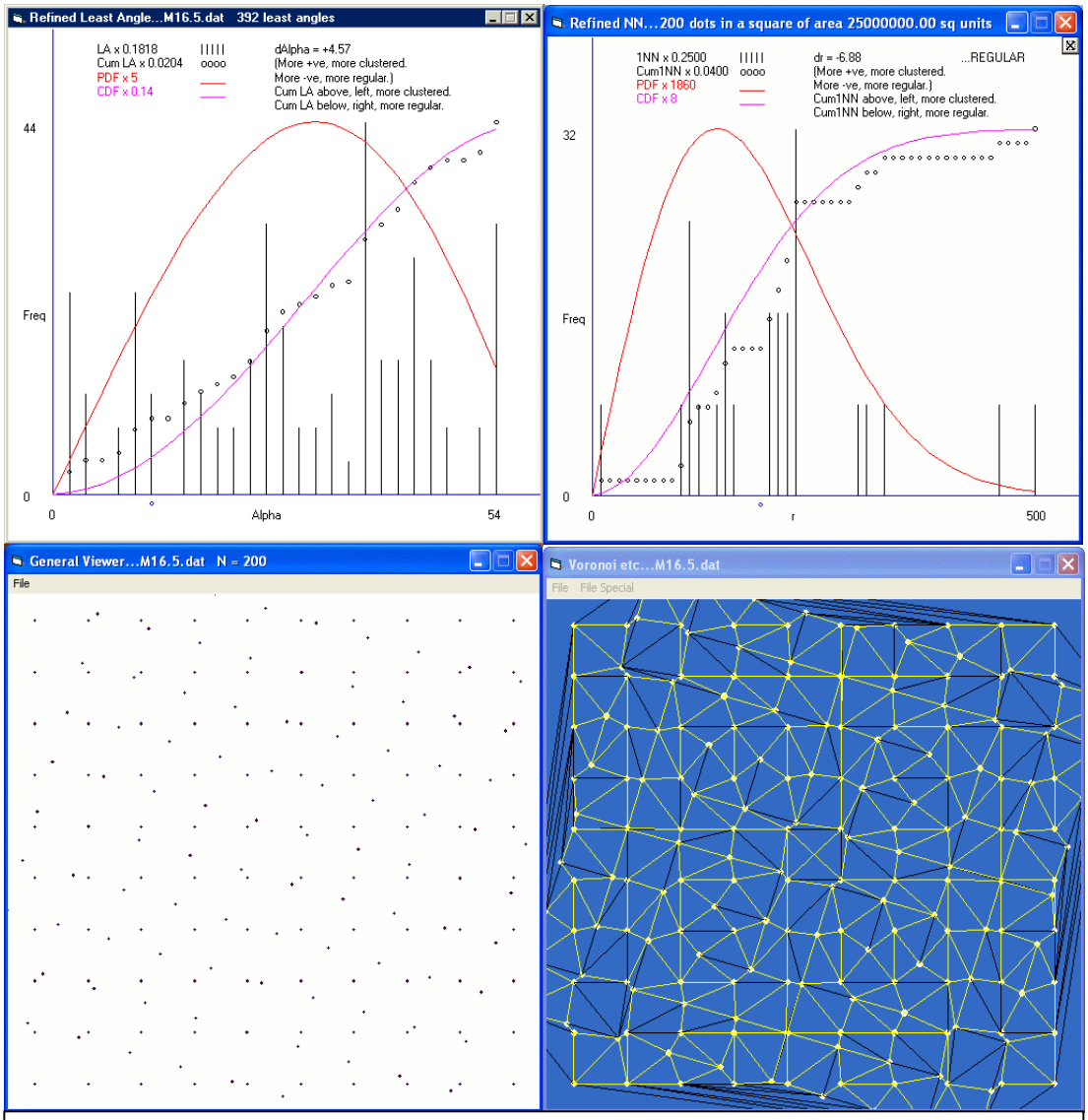
Proportion of different length direct Delaunay edges to total different length Delaunay edges = **1.000**
Proportion of direct Delaunay edge count to total Delaunay edge count = **1.000**
Edge to point ratio = **5.204**

Figure 5.8: Lower left: most regular pattern. Upper left: least angle distribution functions for most regular pattern. Upper right: nearest neighbour distribution functions for most regular pattern. Lower right: Delaunay triangulation of most regular pattern. Yellow edges are direct neighbour edges. There are no indirect neighbour edges in this instance. Least angle distributions agree with nearest neighbour distributions on perfect regularity for the most regular possible pattern. Note that for the most regular pattern there is essentially zero frequency for all but one value for least angles and one value for nearest neighbour distances.



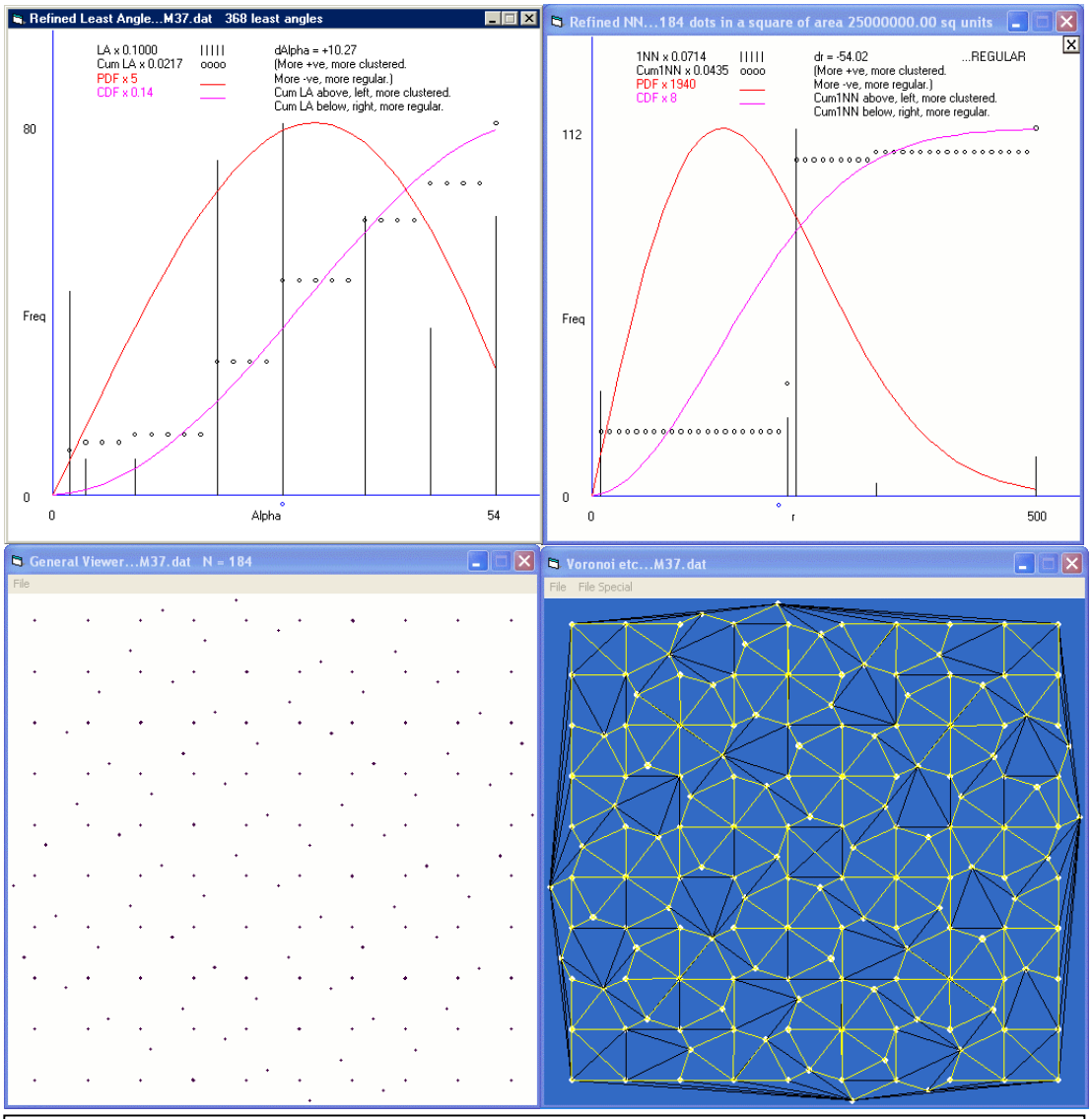
Proportion of different length direct Delaunay edges to total different length Delaunay edges = **0.872**
Proportion of direct Delaunay edge count to total Delaunay edge count = **0.656**
Edge to point ratio = **4.030**

Figure 5.9: Lower left: Marroquin pattern with 3° of rotation. (The pattern was generated by rotation, about the original, of points located at the intersections of an invisible square grid.) Upper left: least angle distribution functions for Marroquin pattern. Upper right: nearest neighbour distribution functions for Marroquin pattern. Lower right: Delaunay triangulation of Marroquin pattern. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges. Least angle distributions deal much better than nearest neighbour distributions with proportions of clustering and regularity in Marroquin patterns.



Proportion of different length direct Delaunay edges to total different length Delaunay edges = **0.799**
Proportion of direct Delaunay edge count to total Delaunay edge count = **0.736**
Edge to point ratio = **4.385**

Figure 5.10: Lower left: Marroquin pattern with 16.5° of rotation. Upper left: least angle distribution functions for Marroquin pattern. Upper right: nearest neighbour distribution functions for Marroquin pattern. Lower right: Delaunay triangulation of Marroquin pattern. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges. A hallmark of Marroquin patterns is the alternating tendency between clustering and regularity over increasing distance scales, represented perhaps a little more accurately here by the least angle distributions.



Proportion of different length direct Delaunay edges to total different length Delaunay edges = **0.756**
Proportion of direct Delaunay edge count to total Delaunay edge count = **0.668**
Edge to point ratio = **4.766**

Figure 5.11: Lower left: Marroquin pattern with 37° of rotation. Upper left: least angle distribution functions for Marroquin pattern. Upper right: nearest neighbour distribution functions for Marroquin pattern. Lower right: Delaunay triangulation of Marroquin pattern. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges. Least angle distributions show more clearly the disposition of points, and deal more effectively with the alternation.

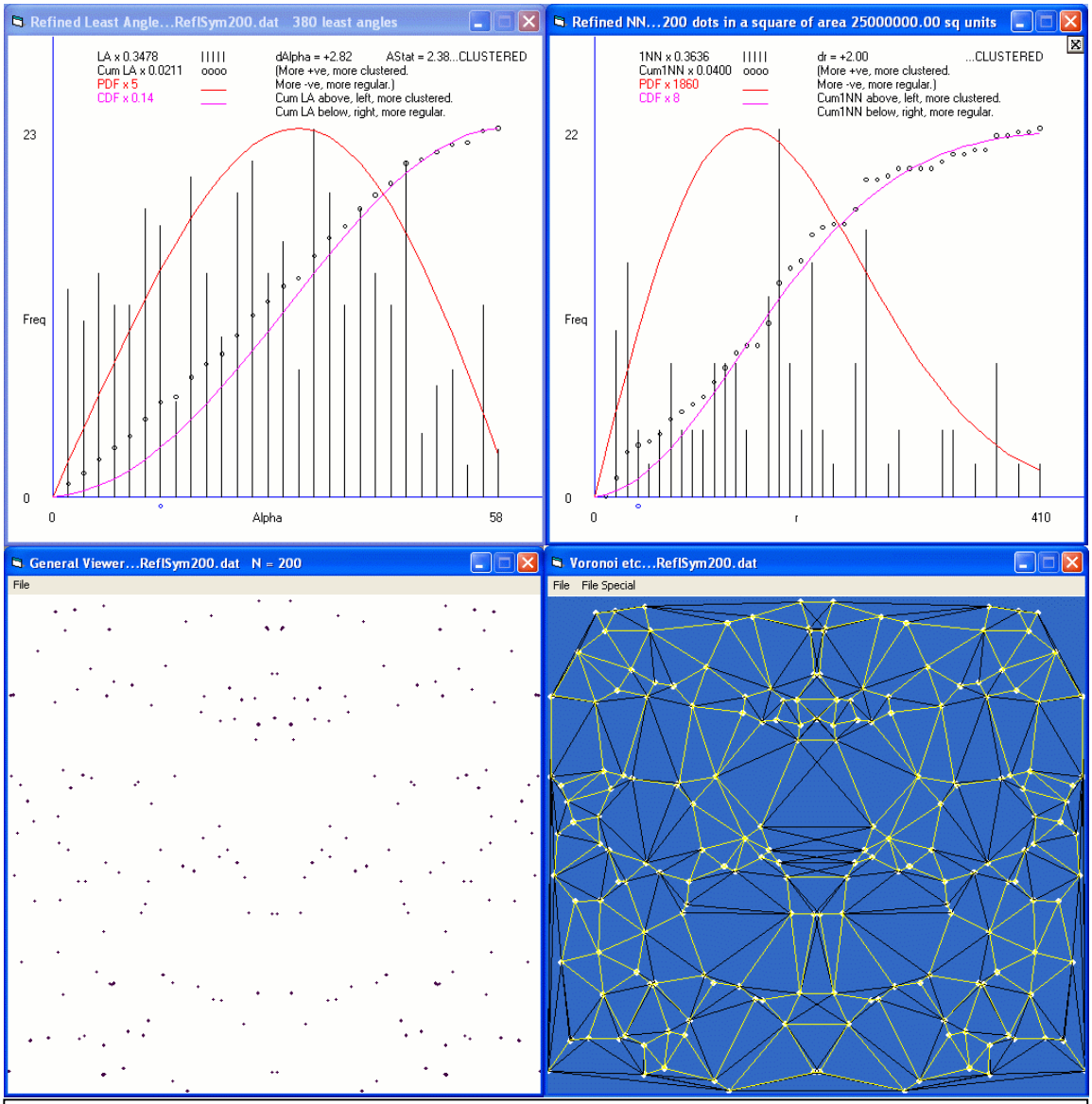


Figure 5.12: Lower left: Pattern with reflection symmetry. Upper left: least angle distribution functions for pattern with reflection symmetry. Upper right: nearest neighbour distribution functions for pattern with reflection symmetry. Lower right: Delaunay triangulation of pattern with reflection symmetry. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges. Neither least angle nor nearest neighbour deals well with reflection symmetry. Compare the distributions for these 200 points with the distributions for the 200 random points shown first, in Figure 5.3. The difference between the distributions is marginal at best.

By way of example, the alternating tendency between clustering and regularity over increasing distance scales for Marroquin patterns was seen in Chapter 3 using Ripley's *K*. Note that this tendency is also seen in least angle and nearest neighbour cdfs. Agreement between methods, inasmuch as what they are designed to reveal, exists for the other patterns as well, which is to be expected.

Generally, least angles have proved to discriminate pattern classes somewhat better than nearest neighbours. However, there is little difference between the two methods for random displays and reflection symmetry patterns, and there is no difference between the two

methods for a most regular pattern. (This must also apply to the degenerate situation. That is, a most clustered pattern; a pattern in which all points are superposed.)

Briefly, there are two ways to consider pair-wise edges in point patterns. One is to treat them as the theory suggests. If some edges are drawn more than once owing to over-drawing by a procedure, then give them the weight they deserve. Another is to treat them all as unweighted: after all, any link is singular to the eye. Each has its advantages.

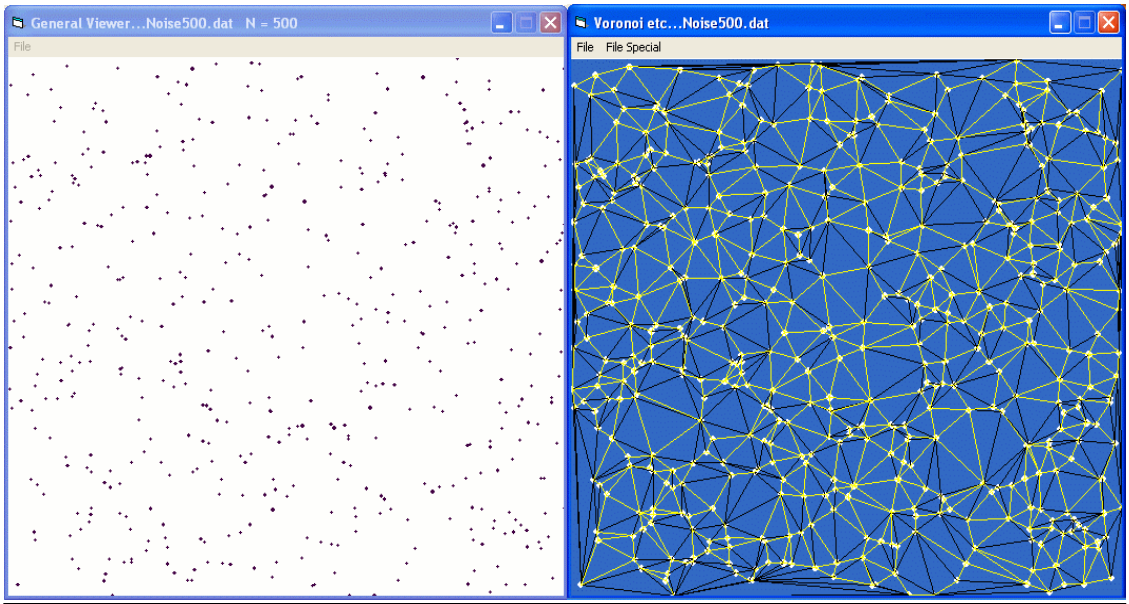
The pictures speak for themselves: unweighted and weighted treatments, different length connections and common length connections, applied to diverse patterns often discriminate them from random point sets and from one another quite well. And, of course, the ratios could be tested for statistical significance against that of a theoretical random point set.

Generally, if using ratios derived from a random display as the point of departure for patterns, then more regular patterns are commonly associated with greater ratios. The Marroquin patterns, for example, show this, and ratios derived from the most regular pattern (Figure 5.8) could be used as yet another point of departure. Notwithstanding this, the ratios do not discriminate Glass patterns from random displays, nor do they discriminate reflection symmetry patterns from random displays.

Before proceeding with the problem of reflection symmetry, it was shown in Chapter 3, pages 59 to 61, that participant's judgments accurately reflect a nearest neighbour statistical summary of relational structure in patterns. Owing to accordance between the density dependent and density free method, this may be interpreted in the more general sense of Barlow (1999), who emphasized the importance of statistical regularities in the environment and referred to many researchers who have begun to show how the regular statistical properties of images are exploited by neural mechanisms.

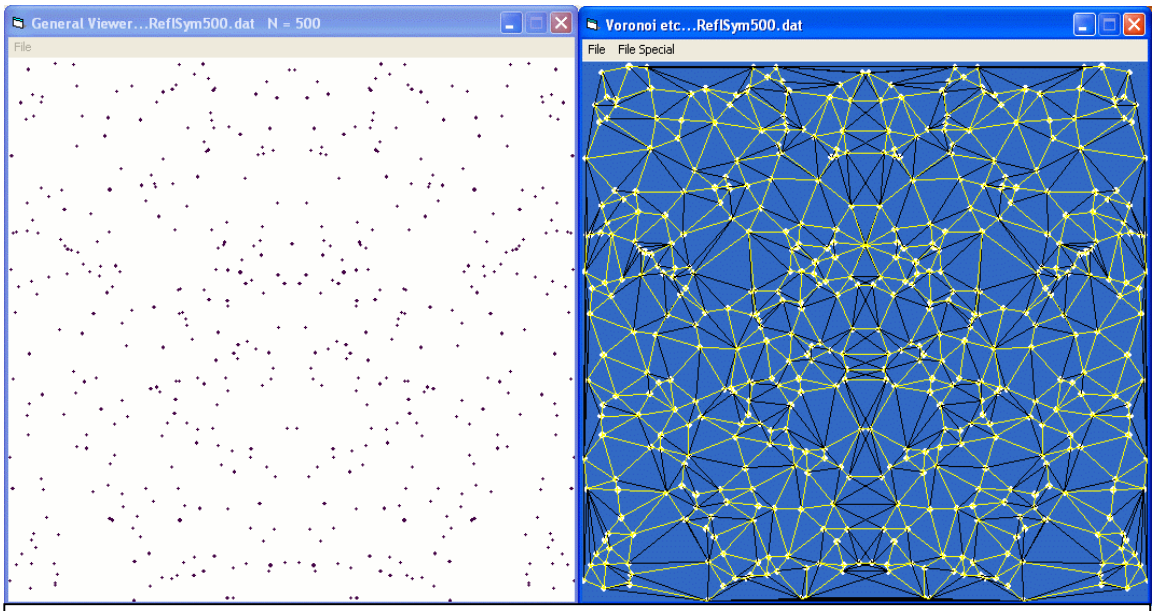
The problem of reflection symmetry

The measures outlined do not differentiate reflection symmetry, which is apparent from scrutiny of Figure 5.12. This point is reinforced by Figures 5.13 and 5.14, the first of which is composed of 500 randomly distributed points and the second of which is composed of 500 reflection symmetry points. (A common kind of reflection symmetry point pattern can be formed by taking a copy of a random point set, flipping it, and then placing it back on top of the original.)



Proportion of different length direct Delaunay edges to total different length Delaunay edges = **0.624**
Proportion of direct Delaunay edge count to total Delaunay edge count = **0.626**
Edge to point ratio = **3.694**

Figure 5.13: Left: random point set consisting of 500 points. Right: Delaunay triangulation of random point set. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges.



Proportion of different length direct Delaunay edges to total different length Delaunay edges = **0.630**
Proportion of direct Delaunay edge count to total Delaunay edge count = **0.632**
Edge to point ratio = **3.808**

Figure 5.14: Left: reflection symmetry pattern consisting of 500 points. Right: Delaunay triangulation of reflection symmetry. Yellow edges are direct neighbour edges and black edges are indirect neighbour edges.

Reflection symmetry and the Voronoi diagram

Why does reflection symmetry appear to refuse evaluation? The answer may lie in the construct that reflection symmetry is *fundamental* to descriptions of all point distributions; including, of course, random point sets. It is hidden in every pattern: it evades attention by its

obviousness. The proposed primacy of reflection symmetry is now addressed over the remainder of the current chapter and the whole of the next chapter.

Consider Figure 5.15. Reflection symmetry is obvious from the Voronoi diagram shown in Figure 5.15(a), which results from two points. It is also obvious from the Voronoi-Delaunay diagram shown in Figure 5.15(b), which results from four points.

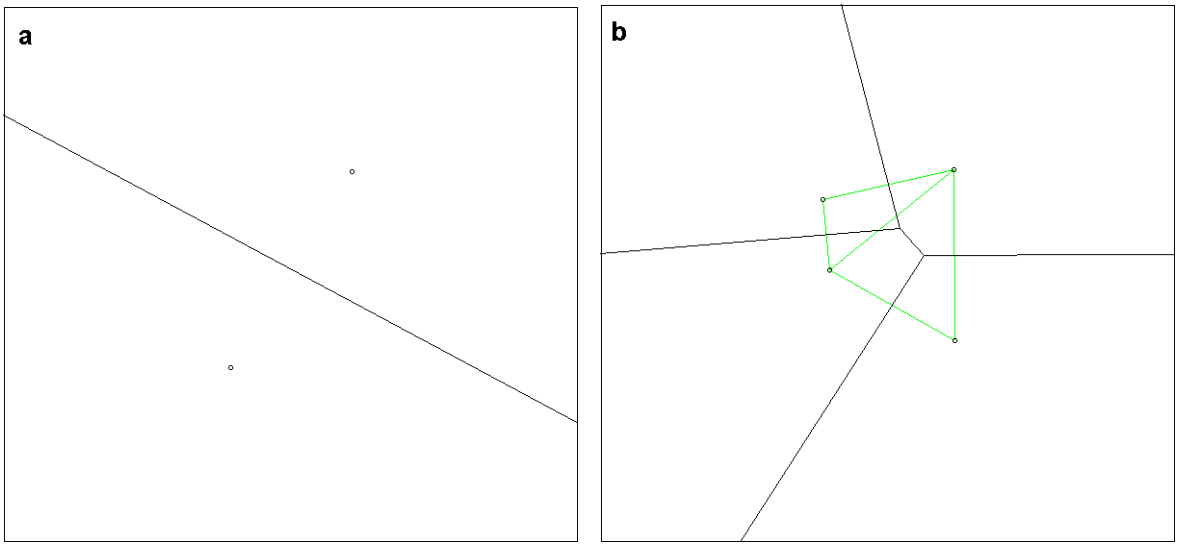


Figure 5.15: Voronoi diagram of two points (a). Voronoi diagram of four points (b).

The Voronoi diagram is a tessellation based on reflection symmetries! In the two dimensional sense it is a collection of *lines* of reflection, and these as medial axes. If the shape, or outline, of a figure is taken as the criterion, then included in the lines of reflection is the medial axis belonging to the outline. If some detail is included within the outline, then included in the lines of reflection are the medial axes belonging to the detail. See Figures 5.16 to 5.18, in which the medial axes are depicted, roughly, by the thicker white lines. Indeed, medial axes form a defined subset of the Voronoi diagram.

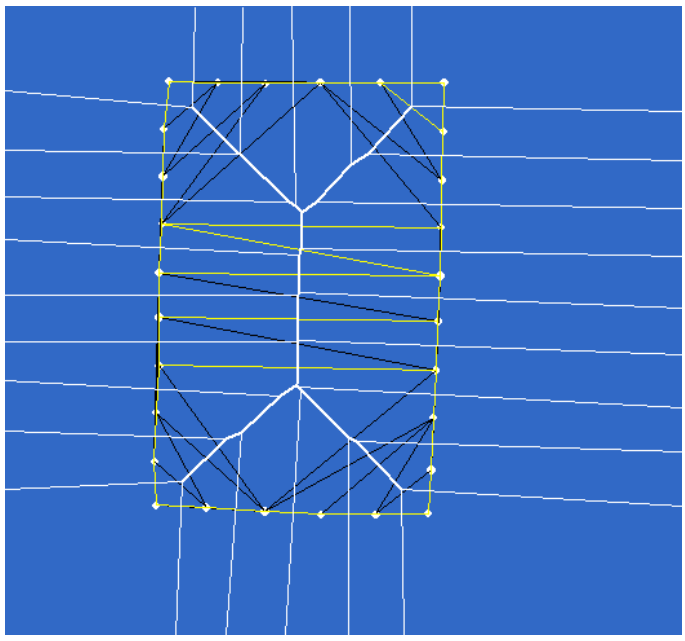


Figure 5.16: Medial axis (bold white) resulting from approximately regularly spaced sampling of hand-drawn rectangle.

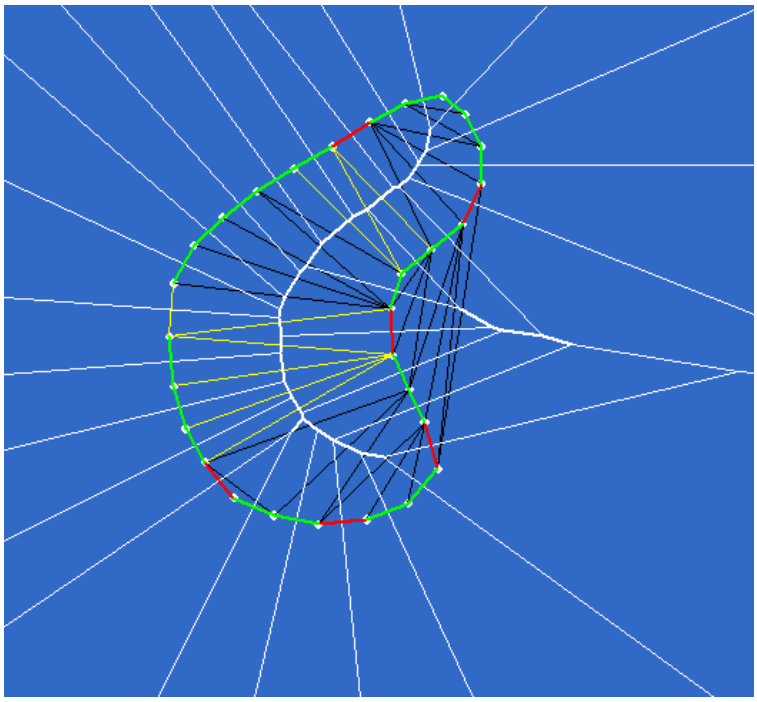


Figure 5.17: Medial axis (internal bold white) resulting from approximately regularly spaced sampling of hand-drawn, arbitrary shape.

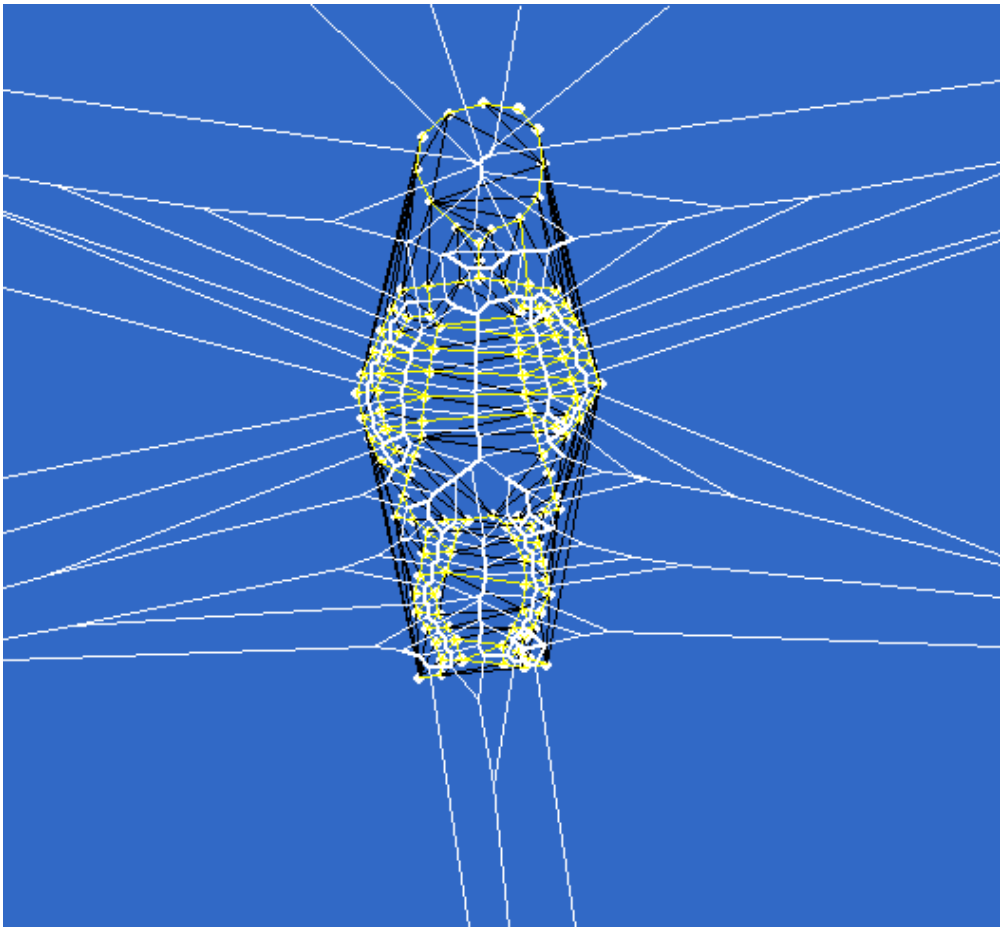


Figure 5.18: Medial axes (internal bold white) resulting from approximately regularly spaced sampling of hand-drawn girl in a dress.

A point to note, taking Figure 5.16 as the example, is that if the outline was continuous (a particular infinite set of points) then the medial axis would extend all the way to the corners; and would be reinforced everywhere between vertices. Note that the medial axis also says something about the shape of the object. It describes shape on a reflection proximity basis. It is not just a simple, single axis of reflection; except under the most basic of circumstances as per Figure 5.15(a).

For interest sake, Figure 5.17 shows nearest neighbours (green), the remaining minimum spanning tree (red), the remaining direct Delaunay edges (yellow), and the indirect Delaney edges (black), which delineate the outline of the shape described by the medial axis. It is not difficult to appreciate that a lot of direct Delaunay edges belong to the minimum spanning tree. Figure 5.18 depicts a stylised girl in a dress. Note how the medial axes pertain to reflection symmetries, elongations, and orientations of figural components. Furthermore the rotational symmetry of the stylised head is obvious from the thinner white lines. Figures 5.17 and 5.18 also show some external thicker white lines, which highlight the axes of external concavities. These kinds of underpinning to figural geometry characterize some essential features of the shapes, and may contribute to recognition of spatial relationships among components.

Inductive bias to visual perception

It is not difficult to find studies suggesting the importance of reflection symmetry as a visual primitive (see, for example, Freyd & Tversky, 1984; Kovacs & Julesz, 1994; Locher & Nodine, 1987). These studies emphasize the effortless detection of symmetry as a preattentive function. They indicate that the global impression fostered by symmetry configures serial scrutiny to exploit redundancy among local features. (Hence departure from symmetry results in more diverse scrutiny.) What these studies show might be interpreted as an *inductive bias* for visual perception. And this is more compelling if, as Leyton (1992) depicts, axes of symmetry, with degrees of curvature, can be generally identified in visual stimuli of any kind.

So the proposal is that reflection symmetry defies classification by ratios and any such method that might differentiate point sets because it lurks in all arrangements of points, or features. It underlies, or forms an inductive bias for, visual perception. We are predisposed to proceed from the specific to the general: from one part or other to the whole, and an important instrument in the realization of the whole is that of reflection symmetry. It provides a geometry at all relevant spatial scales, which serves the perception of organization. This geometry delineates structure in the form of relations among visual features that determine which features are perceived as belonging together. And the assumption of symmetry in the wider sense allows us to form a view of the whole from some limited viewpoint. We have a predisposition to induce the whole from one part or other.

The inevitable hypothesis is that perceptual sensitivity to reflection symmetry is inherent in the manner by which features of salience automatically associate with their areal regions as per the Voronoi diagram. And since Delaunay triangulation is the dual of the Voronoi diagram, it gives as much information, but in a different way. The Voronoi diagram pertains to what region belongs to a feature and Delaunay triangulation pertains to which features go with a feature. In this sense, we might question how such a mechanism can be implemented by networks of neurons, but first we need to question how detection of reflection symmetry is handled by something akin to networks of neurons; namely artificial neural networks.

Chapter 6: Reflection Symmetry and Neural Networks

Brief summary of chapter

This chapter shows how partitioning by reflection symmetry can be handled by neural networks. A simple configuration of artificial neural network with fixed symmetrical weights (corresponding to fixed synaptic strengths) is examined for ability to discriminate symmetry classes from asymmetry classes, as well as provide graded responses for degrees of asymmetry in pattern sets. The configuration is robust in symmetry detection with respect to logarithmic and other operations on weights.

Neural networks for which weights require training are examined for characteristics while, and upon, learning to discriminate symmetry classes from asymmetry classes. Weights develop in a symmetrical way, and the networks are robust in symmetry detection.

A strategy for getting a neural network to tessellate the field as per the Voronoi diagram is discussed with a view to the way that the network might form an inductive bias, based on reflection symmetry, for visual perception.

The work outlined in this chapter, with the exception of that indicated of other researchers along with basic neural network methodology, is original. The implementation of the neural networks is also original. (I wrote and proved the software from first principles.) Voronoi, Delaunay, and medial axes concepts, of themselves, are not original; however, their implementation and application in this chapter are original.

Introduction

Symmetry plays a remarkable role in perception problems. Peaks of brain activity are measured in correspondence with visual patterns showing symmetries (Di Gesù & Valenti, 1996; Norcia, Candy, Pettet, Vildavsky, & Tyler, 2002). Different symmetric dot patterns alternated every half a second with different random dot patterns of the same density show sequences of visual evoked potentials the same as those of random-random sequences up to about 0.2 second, after which response difference between symmetric and random patterns is considerable. Hence symmetry in dot patterns appears to be extracted after an initial response phase, which is indifferent to pattern configuration. Results are consistent with the hypothesis of Lee, Mumford, Romero, and Lamme (1998), and Tyler and Baseler (1998); both cited by Norcia et al. They agree that symmetry appears to be extracted by processing in the extrastriate cortex.

How might a symmetry mechanism be implemented, in a robust manner, in a biological neural system? To this end a detailed examination of various artificial neural networks, implemented in a collection of computer programs, was undertaken; even if somewhat naively. They were tailored from descriptions given in general neural network literature. The following account is not intended to provide details of neural network operation. Rather it assumes some elementary familiarity, such as could be gained from Aleksander and Morton (1991), Dayhoff (1990), Haykin (1994), Muller, Reinhardt, and Strickland (1995), or Sejnowski, Kienker, and Hinton (1986). Because little detail is provided on principles of operation of the different artificial neural networks, it is strongly recommended that the concerned reader consult a basic reference. For those not so concerned, the following account should be intuitively clear.

Symmetric networks

Connections within feedback, or recurrent, networks can be either symmetric or asymmetric. In a symmetric network the weights between connected units in opposite directions are equal; in asymmetric networks they are not. (If every pair of processing units has a connection in each direction, then a network is fully interconnected.) Symmetry has

proven important in determining whether activation converges to a single stable state or not. Activation in symmetric recurrent networks inevitably converges to a single stable state, whereas in asymmetric networks activation need not converge (Hopfield, 1982).

Error back-propagation on a hard learning task

The task is to detect whether or not a pattern of input points is symmetric about the middle. Let the number of points be even in this case: then 11011011 is symmetric and 11101010 is not, for example. (If each 1, otherwise each 0, represents a point, then areal point patterns can be represented by blocks of these.) For the arrangement depicted in Figure 6.1, weights between input units and hidden units are arranged symmetrically, –1, +1, +2, –2, –4, +4…$+2^{(n-1)}$, $-2^{(n-1)}$, $-2^{(n-1)}$, $+2^{(n-1)}$…+4, –4, –2, +2, +1, –1.

Input patterns are multiplied by the weights, the results of which are summed at the hidden units. (Outputs of hidden units are also multiplied by weights, the results of which are summed at the output unit.) Whether or not a pattern activates a unit depends on whether or not its bias, or threshold, is surmounted by a summed product. All symmetric patterns activate the negatively biased hidden units with values of 0. Hence their outputs have values of 0 and the positively biased output unit is switched on. Asymmetric patterns surmount the negative threshold; hence one of the hidden units switches to an output value of 1, say, and the output unit is switched off. Significantly, inputs of any bit width require only two hidden units.

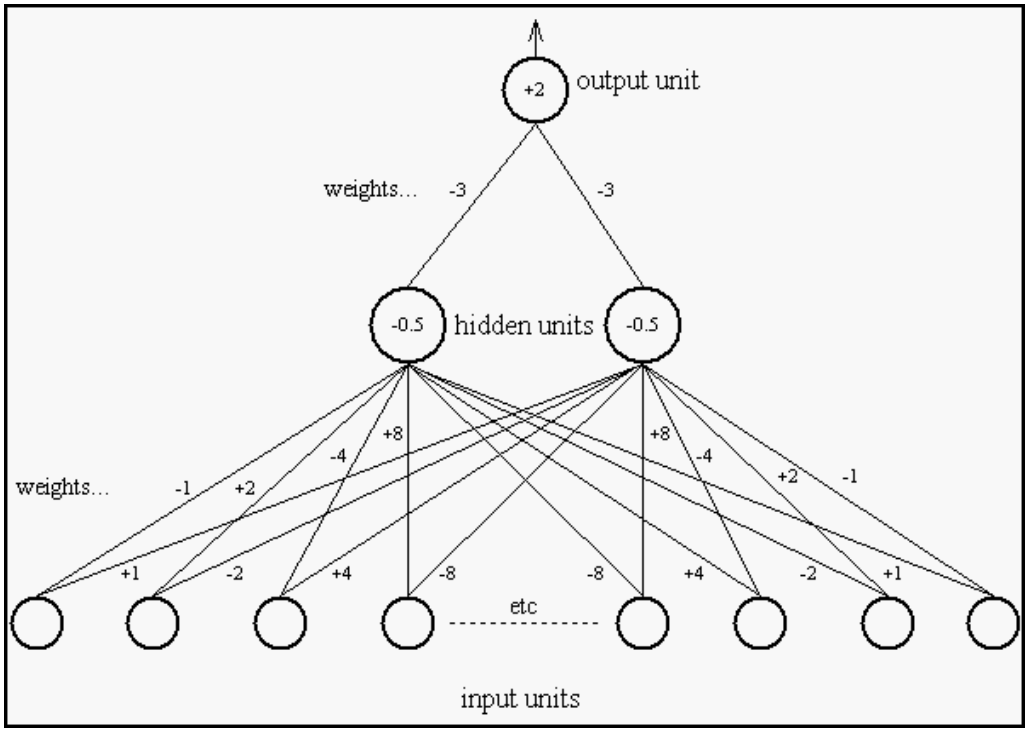


Figure 6.1: A network that detects symmetry. Only two hidden units are required for any number of input units (after Aleksander & Morton, 1991).

In this kind of connectionist modelling of mirror-symmetry detection, solution weights from inputs to hidden units are arranged symmetrically about appropriate axes. It could be argued that a large number of points (or features) require a large range of weight values (like folding and refolding a piece of paper: a thickness that extends from the earth to the moon requires just a moderate number of doublings). However the logarithm of the weights works equally as well.

Implementation

The networks shown in Figures 6.2 to 6.5 include a mechanism for giving graded responses to degrees of asymmetry in pattern sets, and this will be dealt with presently. Meanwhile Figure 6.2 shows an effective arrangement of weights, for a 16 bit wide input, scaled down to fit the form. Black squares represent 'hard-wired' negative weights proportionally by area and white squares represent hard-wired positive weights proportionally by area. The first column of weights are input, or first layer, weights, and represent those incident upon the first hidden unit. The second column of weights are also input, or first layer, weights, and represent those incident upon the second hidden unit. The other weights are second layer weights, and represent those incident upon the output unit.

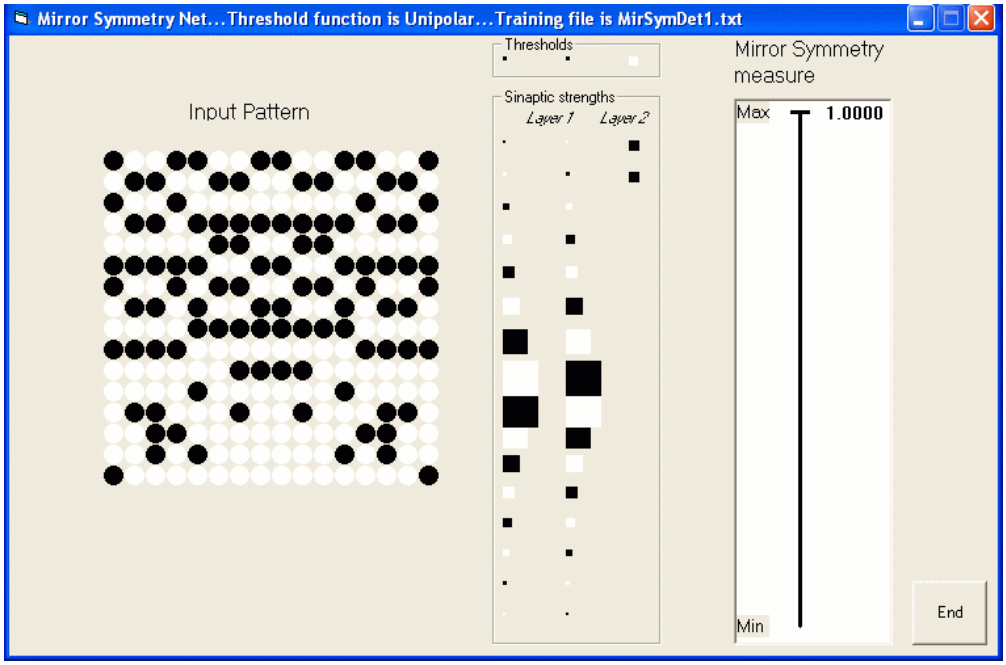


Figure 6.2: Weights (synaptic strengths) by area and polarity. (The symmetry measure, in response to the input pattern, is explained later on in the text.)

Figure 6.3 shows the logarithm of these, which works equally as well.

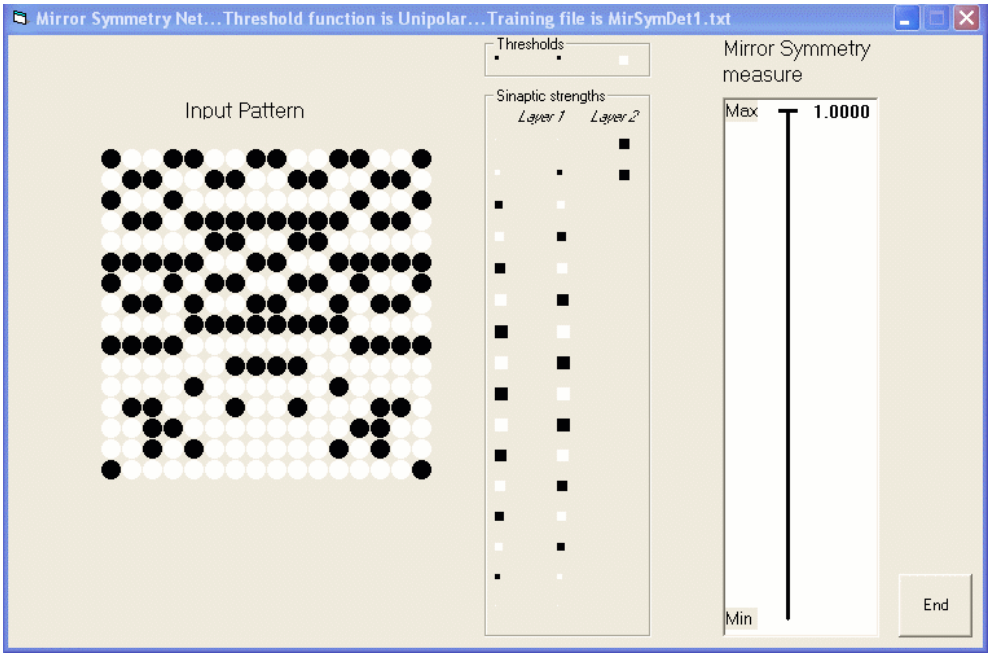


Figure 6.3: Logarithm of weights (synaptic strengths) by area and polarity. (The symmetry measure, in response to the input pattern, is explained later on in the text.)

Scaling makes no difference to detection outcome. Furthermore logarithms taken from any contiguous sequence of doublings makes no difference to detection outcome: see Figure 6.4.

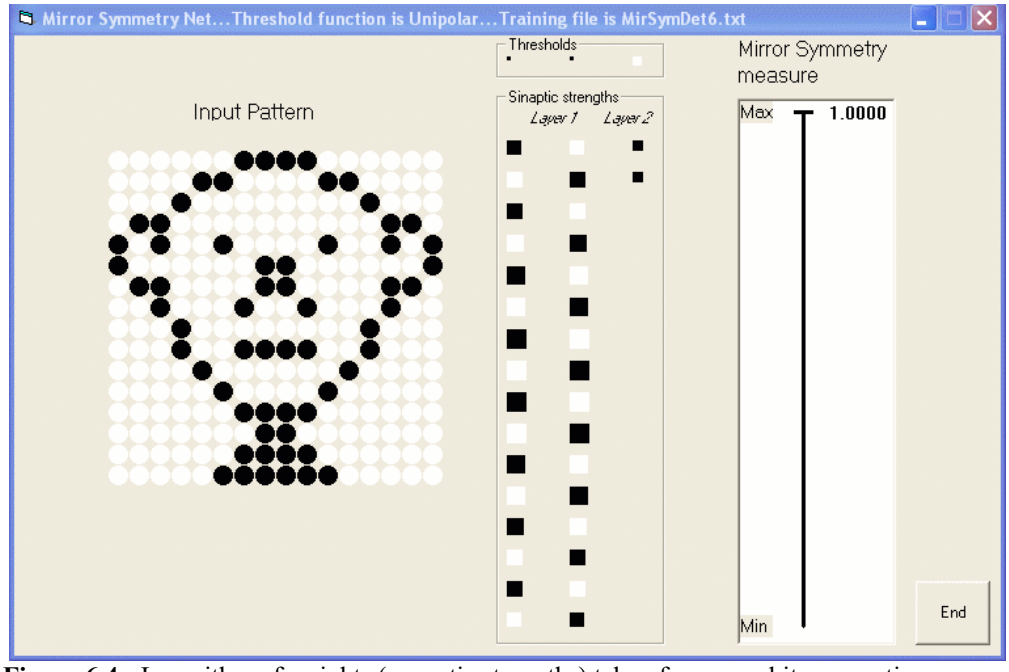


Figure 6.4: Logarithm of weights (synaptic strengths) taken from an arbitrary contiguous sequence of doublings. (The symmetry measure, in response to the input pattern, is explained later on in the text.)

Lastly a reversed sequence of weights works equally as well: see Figure 6.5.

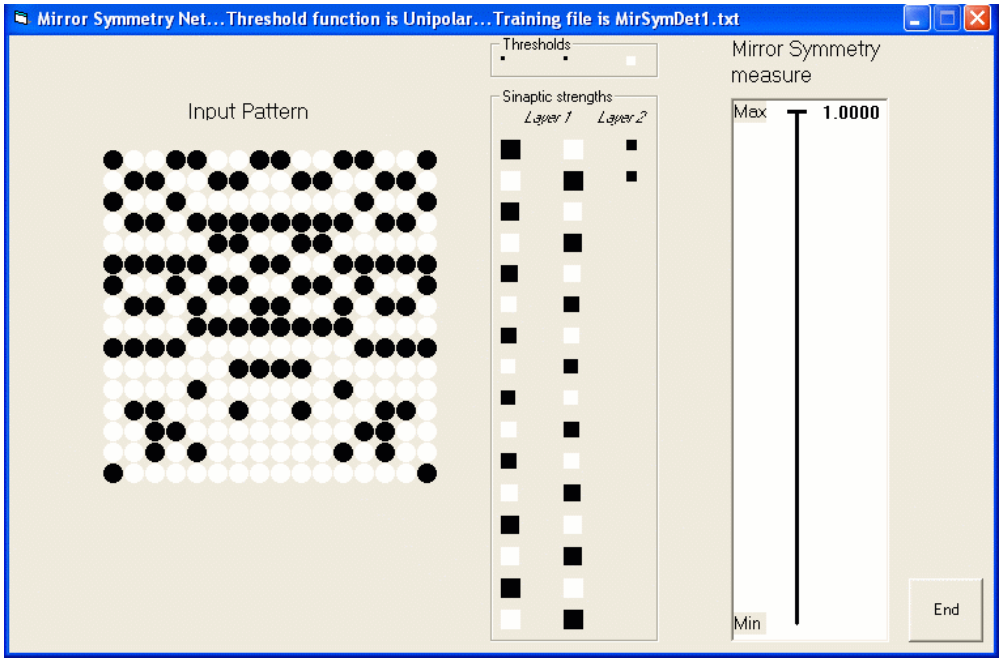


Figure 6.5: Logarithm of weights (synaptic strengths) taken from an arbitrary contiguous reversed sequence of doublings. (The symmetry measure, in response to the input pattern, is explained later on in the text.)

And this is closest to what happened when an error back-propagation network, based upon the multilayer perceptron, and with half the width of input (8 bits instead of 16 bits), was trained to discriminate symmetric patterns. See the otherwise raw weights in Figure 6.6.

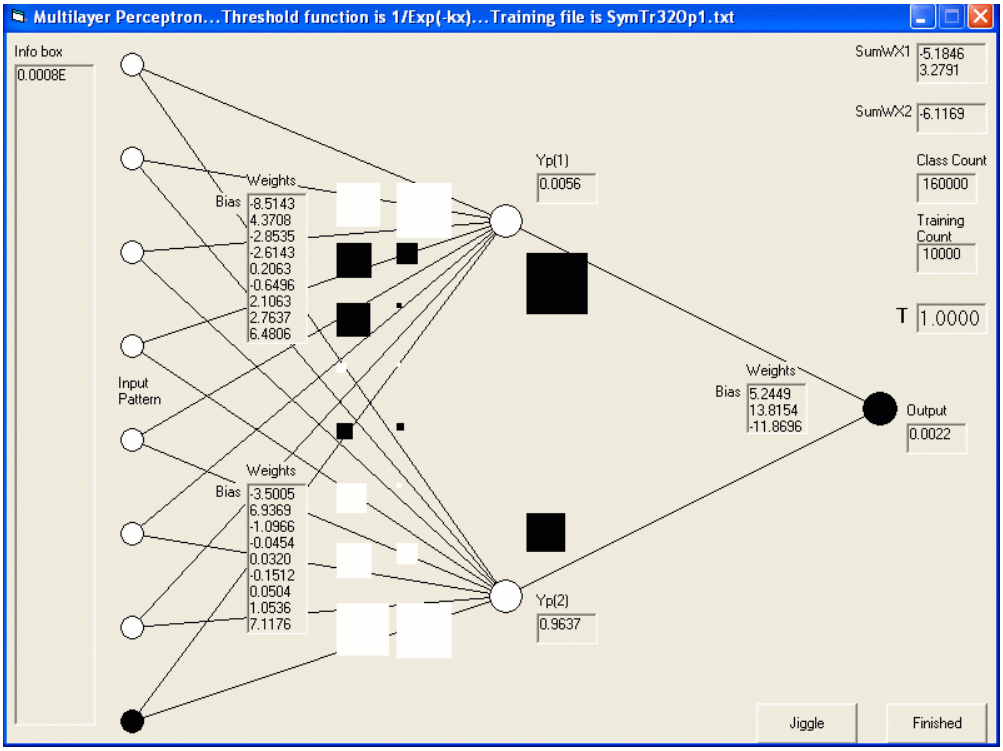


Figure 6.6: Weights by area and polarity.

Extensive training of various networks suggests a vast number of readily assumed weight organizations with some form of symmetry, of which any selection detects mirror symmetry. The arrangements are simple and robust.

Error back-propagation through a neural network, against the direction of synaptic connections, may be biologically implausible. However the supervised method of learning, through successive error reduction between output and target values, is comparatively simple in computational terms, and has proved helpful in neurobiological studies. The function of interneurons found in a leech, for example, have been understood by comparison with a computer simulated network trained by error back-propagation (Lockery, Fang, & Sejnowski, 1990; Lockery, Wittenberg, Kristan, & Cottrell, 1989).

Extensions

The networks shown in Figures 6.2 to 6.5 include a mechanism for giving graded responses to degrees of asymmetry in pattern sets. The simplest model counts the number of asymmetries in an input pattern, line by line. Strength of asymmetry in a line is judged by how many points are not symmetrically paired, and how far apart corresponding locations are. See Figures 6.7 and 6.8. If a line has perfect symmetry, then it contributes 1/(number of lines) to the graded response. If it has asymmetry then it provides a weakened contribution, related to strength of asymmetry, to the graded response.

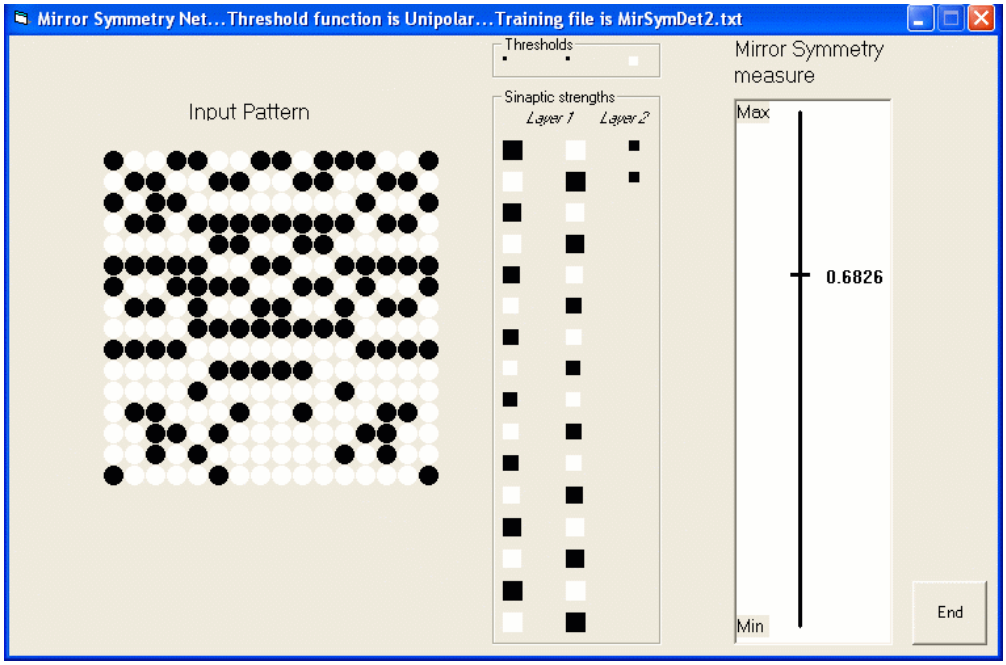


Figure 6.7: Pattern with partial asymmetry.

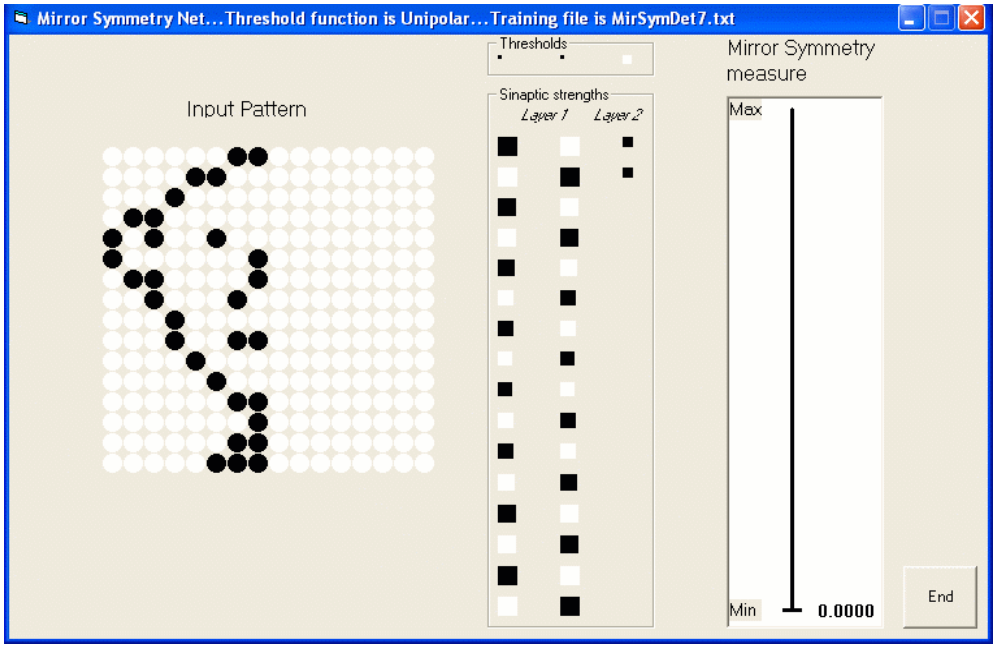


Figure 6.8: Pattern with complete asymmetry.

But, by way of caveat, consider the following point. The network depicted in Figure 6.6 does not train to a global minimum error if the number of symmetric inputs is too few relative to the number of asymmetric inputs. For example, with all eight bit combinations—16 symmetric inputs and 240 asymmetric inputs—the network repeatedly gets caught in a local minimum despite an otherwise effective simulated annealing regime. But there is no such training problem with 16 symmetric inputs and 48 asymmetric inputs, for example. However if symmetric weights are hard-wired, rather than trained, then the network discriminates any symmetry from any number of asymmetries. Or even if weights are partially hard-wired (seeded) for symmetry before training, the improvement is significant. Additionally the network under training does not necessarily produce commensurate, or smoothly varying, symmetric weights. (Commensurate symmetric weights are produced if the number of symmetric inputs is around, or over, half the total number of inputs.)

One further caveat is necessary. With some abbreviated set of input patterns employed in training, containing more than around half that are asymmetric, the asymmetric inputs, included in the training set with the symmetric inputs, need to be selected at random from all the possible asymmetric combinations. Too many asymmetric inputs with some chance class characteristic, or consistency, cause the network to accommodate by perturbation of weight symmetry, and slow the learning process.

Given a fair training regime, first layer weights are symmetric in magnitude only, or symmetric in polarity and magnitude. For the former, second layer weights have different magnitudes, and for the latter they have the same magnitudes. Hence the system is compensated by considerations related to symmetry. And, of course, polarity for second layer weights also figures in the balance. See Figures 6.9 and 6.10.

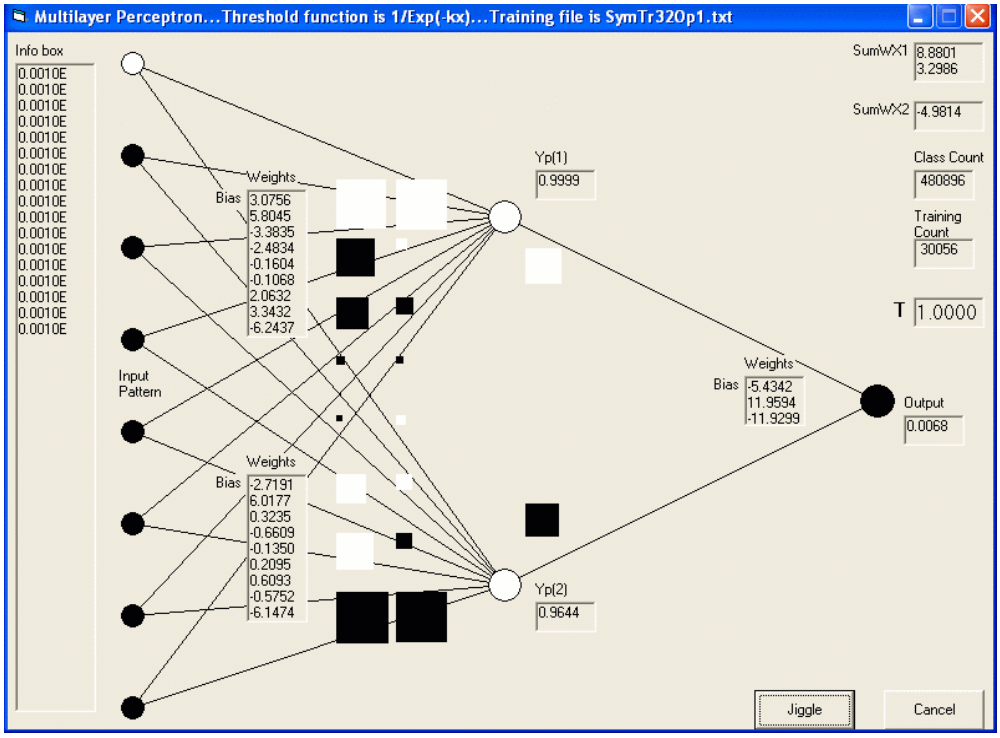


Figure 6.9: Weights by area and polarity.

Lastly networks with more hidden units and/or more layers of hidden units train successfully with relatively fewer symmetric inputs; just 16 symmetric inputs among 256 inputs, for example. The more complex the network, the greater its training capacity. See Figure 6.10.

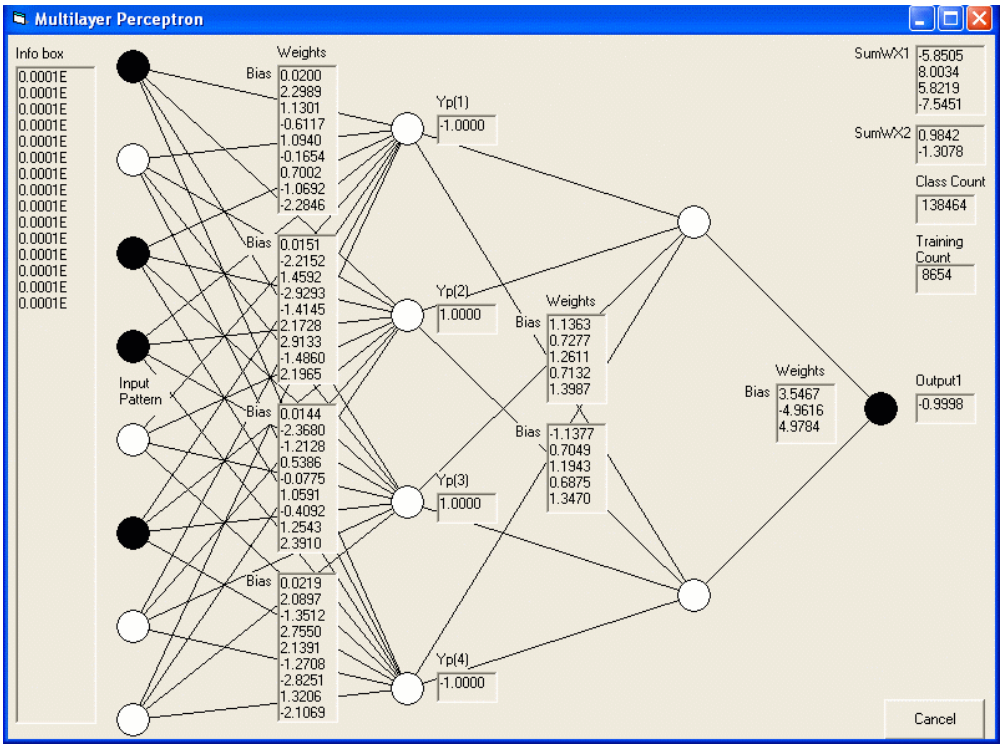


Figure 6.10: More complex network with more hidden units and more layers.

Despite there being just 16 symmetric input patterns amongst 256 input patterns, consisting of all possible eight bit combinations, the weights train symmetrically. Even so, inclusion of all

248 asymmetric patterns delivers a kind of symmetry inherent in the whole. See Figure 6.11 for a graphic representation of the trained weights. So as not to bias network training, the order within each set of training patterns was randomised for all networks.

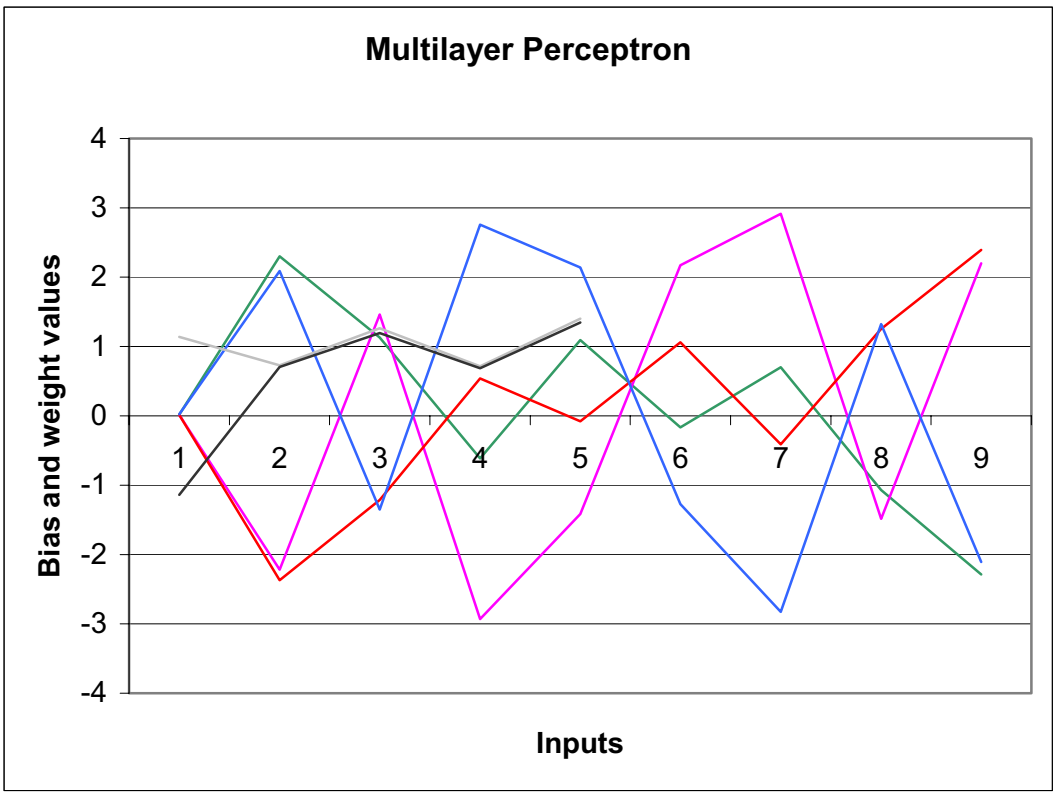


Figure 6.11: Trained weights for network shown in Figure 6.10. Green and red vertices at second layer inputs two to nine represent first layer weights incident upon second layer units one and three respectively. Magenta and blue vertices at second layer inputs two to nine represent first layer weights incident upon second layer units two and four respectively. Grey and black vertices at third layer inputs two to five represent second layer weights incident upon third layer units one and two respectively. (All colours at inputs one represent corresponding bias, or threshold, values for second and third layer units.)

In any event the idea is not to involve the reader in the technicalities of neural networks, but simply to engender an appreciation of the fact of simple and robust symmetric weights; robust enough that perturbing final weight configurations in a limited way appears not to interfere with performance.

Interestingly, other classes into which the training patterns can be segregated, odd or even parity for example, do not necessarily train as readily as that for symmetry. The gradient descent to the region of global minimum error between output and target values for symmetric patterns can be sudden and dramatic, especially for the more complex networks. The untrained weights typically pulse a few times by way of omen, and then quite suddenly the output error reduces considerably as the weights take on the appearance of their final configuration. All further training runs just serve to garnish the situation: to reduce the error to some small arbitrary amount.

Other networks

By way of summary for some other styles of networks investigated—Boltzmann, unsupervised learning by competition among processing units, unsupervised learning by

discovery of cluster structure, and some with step, as opposed to sigmoid style, activation functions—mirror symmetry was found generally not to be too difficult for them to learn.[1] With competitive learning, for example, each processing unit distinguishes a class. No target values are required for training: just a strategy for selecting the processing unit with the largest response to a pattern, and a method of reinforcing that response. Using a competitive learning network with two processing units, the weights shown in Figure 6.12, resulting from 14, eight bit symmetric patterns mixed with 14, eight bit asymmetric patterns, were established rapidly and repeatedly. Once again, note the symmetry of the weights for whichever processing unit learned almost all of the symmetric patterns.

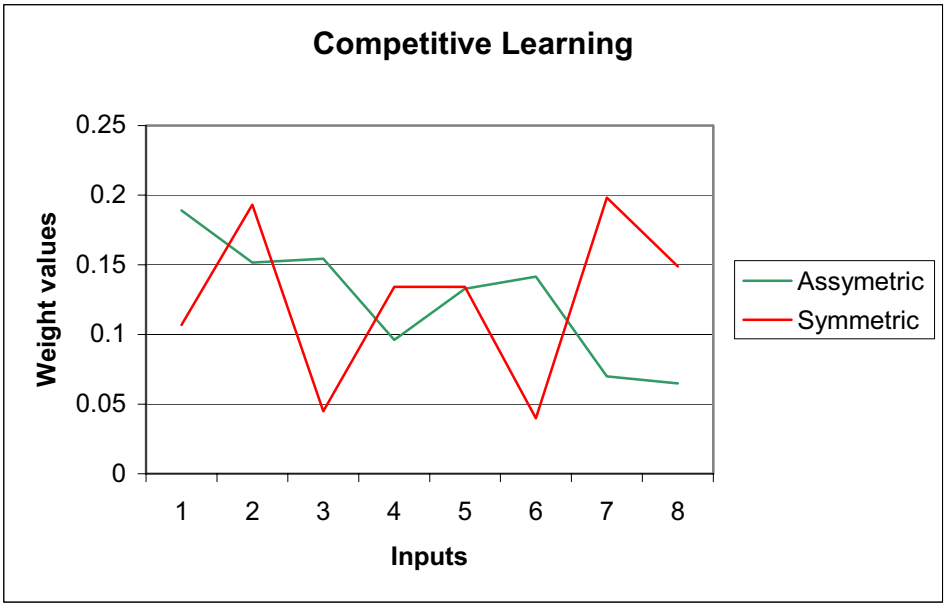


Figure 6.12: Green vertices represent weights for the processing unit that learnt the asymmetric patterns and red vertices represent weights for the processing unit that learnt the symmetric patterns.

Lastly, joining discrete values in the preceding two graphs by lines might seem inappropriate, but line graphs indicate symmetries and asymmetries more clearly than bar graphs. Just the vertices of the graphs correspond to weight values and no credence should be given to what lies between.

Neural networks and inductive bias

Earlier claims in this thesis amount to symmetry being psychologically fundamental, and so giving an inductive bias to perceptual organization. This perspective is reconciled with the neural network simulations by now showing that a neural network can implement that inductive bias.

Voronoi network

A hypothesis at the end of the last chapter was that perceptual sensitivity to reflection symmetry is inherent in the manner by which features of salience automatically associate with their areal regions. Hence it is fair to consider a neural network, based upon reflection

[1] A Boltzmann network has symmetric connections between units. It is a stochastic, recurrent network that employs simulated annealing.

symmetry, that produces Voronoi tessellations of the field. The properties of such a network might be of interest to the relevant neurophysiology.

To get a network with discrete inputs to draw Voronoi cells, the areal field first needs to be discretized by an appropriate tessellation, or tiling, such that distance in any direction is approximated by the number of tiles crossed along the path. A regular tessellation comprises congruent regular polygons. A regular polygon has three or more equal sides and angles, and there are just three regular polygons that tessellate in the (Euclidean) plane: triangles, squares and hexagons. See Figure 6.13.

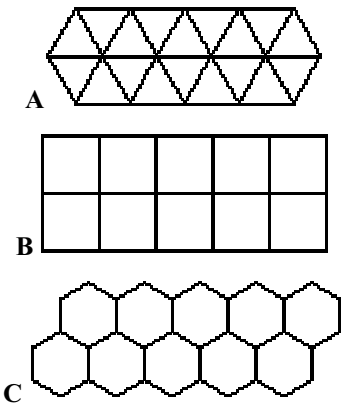


Figure 6.13: Tessellations of **A:** triangles, **B:** squares, **C:** Hexagons (after Alejandre, 1994–2003).

(Since a hexagon can be made up of six equilateral triangles with a common vertex, at the centre, then, fundamentally, there are just two regular polygons that tessellate in the plane.)

A tessellation of squares is suitable for symmetry detection, but is unsuitable for distance estimation. To cite the worst case: squares crossed at 45^{o} span approximately 1.414 times the distance than the same number of squares crossed at the vertical or horizontal. A tessellation of hexagons provides an acceptable approximation for equal distance in all directions, and suitably small hexagons provide fine discretization. The area of a hexagon needs to be small enough so that the likelihood of it overlying more that one point is remote; see Figure 6.14.

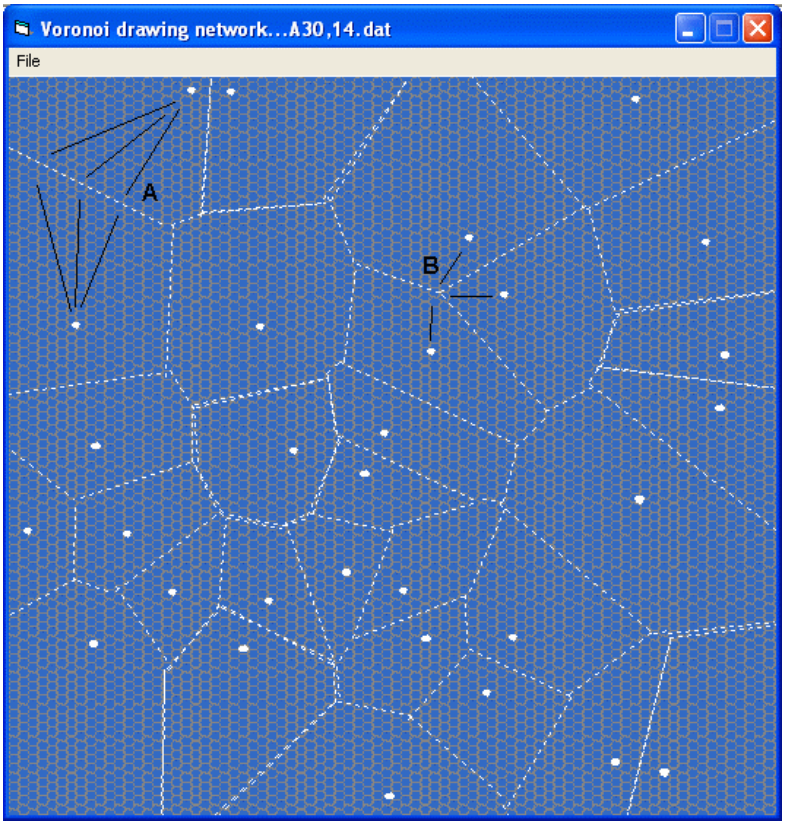


Figure 6.14: Output of the Voronoi drawing network is shown in white. The black lines at **A** show some binary symmetries that delineate the Voronoi segment. The black lines at **B** show a ternary symmetry that marks the point of intersection of Voronoi segments. (Some of the common segments are slightly disjoint due to the granularity of the distance measuring mechanism.)

The Voronoi drawing network detects binary symmetries on the basis of empty hexagons and hexagons with a point: 0s and 1s, say, respectively. Detection occurs when the input field contains two 1s, with an equal number of zeros (empty hexagons) to each 1 (occupied hexagon); or, in the minimal case, no zeros to each 1. In drawing a Voronoi cell segment, the network must be able to also detect ternary—and for degenerate situations, quaternary—symmetries among the binary bits so as to terminate the segment and initiate the next binary symmetry detection. (Ternary, and quaternary, symmetry detections are performed by appropriate multiple binary symmetry detections.) In the search for such positional symmetry, degrees of asymmetry, as previously outlined, are minimized via gradient descent.

Left at this, Voronoi type cells resulting from all possible half distances between point-pairs would be eventually constructed in an overlapping manner. But only the 'neighbourly' half distances are required.[2] Hence the mirror symmetry with the minimum number of zeros to each of a pair of 1s is the relevant initiating symmetry. (This is guaranteed to be a Delaunay edge because it is a nearest neighbour edge.) The number of zeros to the ones is then incremented until the ternary symmetry is encountered, upon which the next binary symmetry detection includes the one from the initiating symmetry. (The equidistant, other, one is simply recorded at this stage.) This occurs until a cell is complete, or the edge of

[2] As it turns out, half distance points between all possible point pairs provide a useful device for pattern discrimination, which is addressed in Section 2 of the thesis.

the form is encountered. Then another symmetry, involving the (next) hitherto recorded, but unprocessed, symmetry, is processed similarly, and so on, until all relevant symmetries have been processed.

Left at this, problems arise in exceptional situations: if more than two points at a time are co-linear across a form, for example. This can be overcome by having a variable input width to the network; and so on for other problems (which I refrain from mentioning, given the thrust of the chapter). While demonstrating feasibility through some choice of methodology, the more important consideration concerns the characteristics of such a network in terms of requirements for neurophysiology. They are:

- At least two complementary processing neurons per lineal input
- Logarithmic synaptic strengths
- Asymmetry minimising mechanism

In the case of just two processing neurons some 'hard-wiring', or seeding, of synaptic strengths might be expected. This is not necessary for slightly more complex neural arrangements. If no hard-wiring is evident then the environment needs to include a variety of asymmetry classes in order for the symmetries to train properly: either that, or a large number of symmetries relative to asymmetries. The prospect of logarithmic synaptic strengths is consistent with the logarithmic responses of eyes, ears, and other sense organs to stimuli (Berne & Levy, 1993; Somjen, 1972, for example). And any asymmetry minimising mechanism is consistent with the ubiquitous fact of energy minimisation found in nature (Barton, 1997; Nordholm, 1997, for example).

Neurophysiologic correlates

Among the expositions made by Cook (2003) in a paper on spatial regularity among retinal neurons, two are relevant to the current problems. The first elucidates the subject of his paper:

> Regular arrays of retinal neurons often extend dendrites in a competitive, territorial manner that minimizes the overlap of their dendritic fields and causes them to tessellate, 'tiling' the retina like the individual pieces…of a ceramic mosaic. (p. 463)

A tessellation of the field is a prerequisite for the above-described Voronoi drawing network, which forms an interesting coincidence with Cook's observation. The second exposition relates to mammalian ganglion cell mosaics, which appear to come in inner and outer stratified pairs:

> The dominance of mammalian alpha and beta cells in early studies of mosaics, starting with those that first showed the importance of dendritic stratification… has created a general impression that mosaic-forming neurons are typically organized into pairs of related 'subtypes' that differ only in their stratification and in giving opposite-sign physiological responses to visual stimuli…
>
> Thus, it is plausible that some of the order in these paired mosaics is created by a process that encourages neighboring multistratified cells to remodel their dendrites into opposite-sign pairs, and that each complementary pair of mosaics may have evolved by the developmental remolding of a single ancestral mosaic whose neurons remained bistratified or multistratified throughout life. (p. 472)

The above-described symmetry detecting networks with two hidden units have precisely this kind of complementary configuration; which forms an even more interesting coincidence with Cook's observation.[3]

Levels of medial axes

At this stage, another matter requires attention. It stems from the observation that peaks of brain activity are measured in correspondence with visual patterns showing symmetries (Di Gesù & Valenti, 1996; Oka, van Tonder, & Ejima, 2001). But the claim is that symmetry informs the detection process for any pattern. That medial axes are obtained—via some equivalent to the Voronoi process—from a salient subset of surface descriptors, particularly those delineating the peripheral, is evident from the approach. Much interior detail results in many submedial delineations that preclude the medial belonging to the peripheral. Nonetheless some of the symmetries that show peaks in brain activity are augmented by a lot of interior detail: symmetries that are immediately evident in the sense of all medials due to salient subsets of surface descriptors. Hence level of brain activity could be related to 'nesting levels' of symmetries.

Returning to the Voronoi network, the untenable situation is noted for Voronoi cells eventually resulting, in an overlapping manner, from all possible half distances between point pairs. (Point pairs, of course, represent the most basic symmetries, followed by triplets of points and, in degenerate cases, quadruplets of points.) The number of operations required to produce such overlapping cells would be prohibitive in any practical situation. However, salient subsets of pairings (which, of course, also account for triplets and quadruplets) result in medial axes and submedial delineations of import.

If Voronoi tessellation is performed for a subset constituting a peripheral (or maybe a subset constituting a convex hull) separately, for example, as well as for a whole set, and then superposed, the result is a Voronoi tessellation on a Voronoi tessellation. This provides two levels of detail. Figure 6.15 is a sampling representation of a whole set (without superposition) that shows nearest neighbour, minimum spanning tree, remaining direct Delaunay triangulation, and indirect Delaunay triangulation in different colours, along with the Voronoi tessellation. (Note: just the left half of the face was sampled and the sampling points mirrored to the right before processing, hence the right is not fitted exactly over the pictures in Figures 6.16 and 6.17.) Figure 6.16, shows Voronoi tessellation performed for the peripheral, and Figure 6.17 shows the superposition. Minimum spanning trees are shown in different colours, along with corresponding Voronoi tessellations. In any such representation, edges may be sampled discretely at roughly equal nearest neighbour intervals that may be scaled differently for different edges. The representation allows much freedom in overall sampling scale and point location.

[3] This should not be confused with the on- and off-centre activated centre-surround responses of bipolar and ganglion cells in the retina, which indicate a *difference* in light intensities that remains constant with change in illumination. These do not come in complementary pairs in any way to which Cook refers.

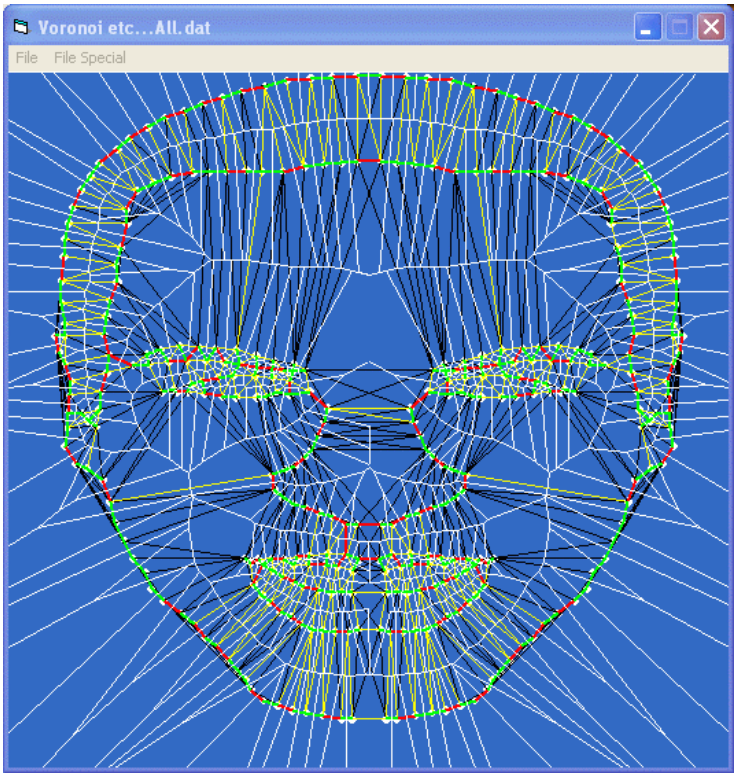


Figure 6.15: Sampling representation of whole face set. The dots are approximately regularly spaced sampling points, of which the coordinates are the only input to the computer program. All lines are output after processing. The Voronoi diagram is shown in white. Nearest neighbour edges are shown in green; remaining minimum spanning tree is shown in red; remaining direct Delaunay edges are shown in yellow; and indirect Delaunay edges are shown in black.

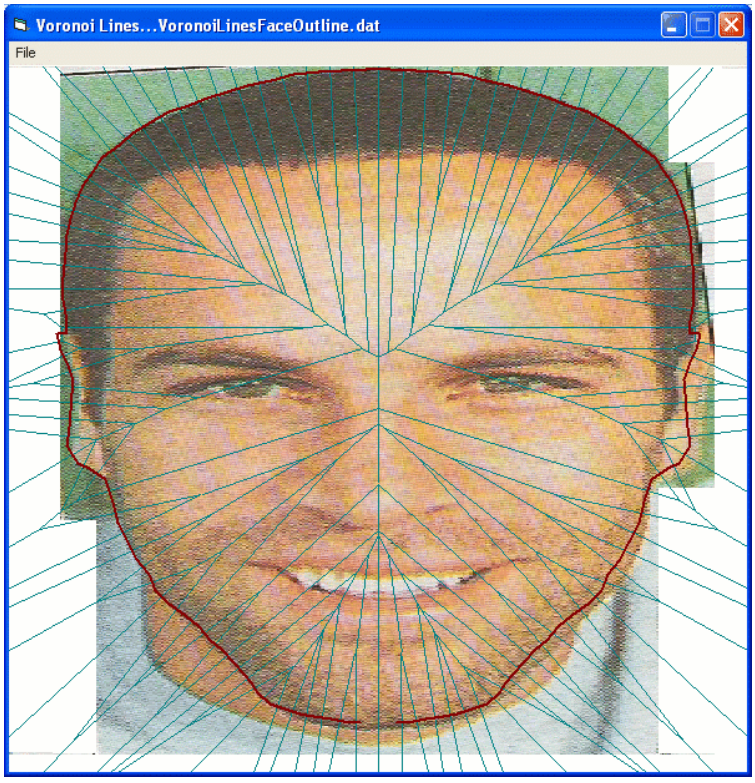


Figure 6.16: Voronoi diagram due to peripheral face set is shown by thin lines and minimum spanning tree is shown by the heavy outline.

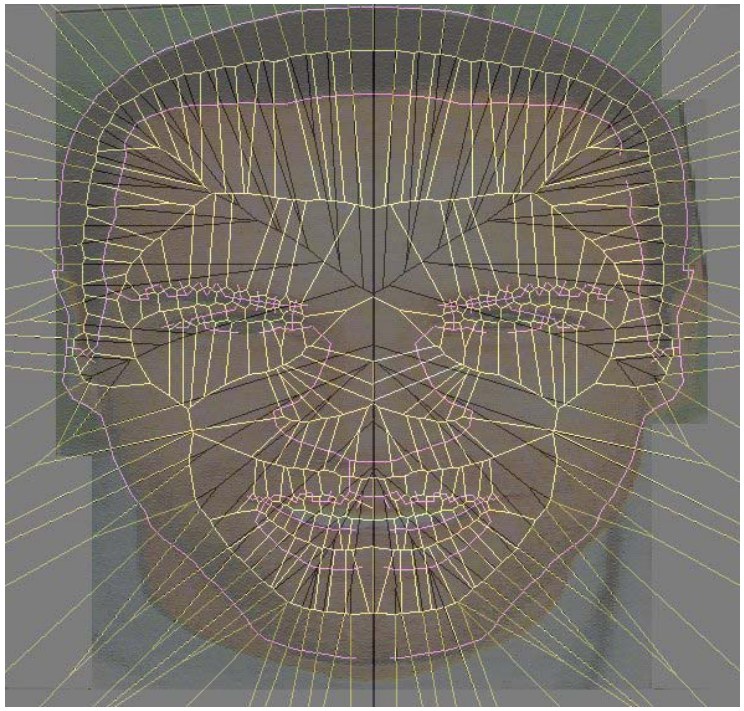


Figure 6.17: Superposition, with Voronoi diagram due to peripheral face set shown in black and Voronoi diagram due to whole face set shown in yellow. Minimum spanning trees are shown in magenta.

Just two superpositions of Voronoi tessellations, by way of example, may seem somewhat arbitrary. Groups of salient elements may be processed separately or together and then superposed, and could result in several superpositions of Voronoi tessellations

Finally the relative neighbourhood graph is given in Figure 6.18. Note that features do not have a gap in their outlines as per the minimum spanning tree. Nor are there as many edges as per Delaunay triangulation. Perhaps these edges may have a higher degree of pertinence among those of the Delaunay triangulation. Of course, all the measures described in the hierarchy are implicit in Delaunay triangulation. Of fundamental import for optimal computing methodology for these measures is the philosophy of the Voronoi tessellation approach.

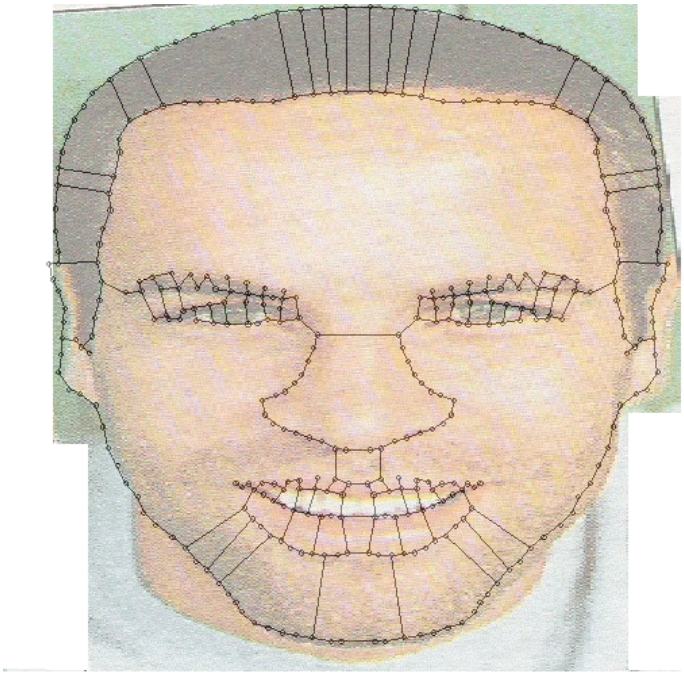


Figure 6.18: Relative neighbourhood graph of whole face set.

Application to human visual perception

According to Blum (1973), inventor of the medial axis transform, the medial axis appears to be of great help in the recognition of particular shapes. Somewhere in the visual processing mechanism, recognition of spatial relationships among components with hierarchical groupings needs to be established, not the least important being adjacent component relationships. These relationships are inherent in Delaunay triangulation and the Voronoi diagram, also from which medial axes can be obtained.

Medial and submedial axes relate to degrees of elongation or symmetry of a shape. They provide a compact description related to spatial arrangement, orientation, and size, which are shown in neurophysiological studies to be associated with the functional role of the primary visual cortex. See Lee, Mumford, Romero, and Lamme (1998), for example. According to Marr (1982) ‘shape’ is the geometry of an object’s physical surface. Such geometry is underpinned by axes that are determined by salient geometrical characteristics of shape. Medial and submedial axes provide an object-based coordinate system, based on a hierarchy of component axes, for surrounding spatial arrangements.

The 2D medial and submedial axes obtained from the projection of a 3D body onto an image plane often have very similar forms to the projection of the medial and submedial axes of the 3D body onto the image plane. The axes articulate how components are put together,

and such relationships among components are relatively invariant to change in viewpoint, as well as to bending and moving states of articulated shapes. And medial axes form the basis for reconstruction of a 3D body by the medial axis transform, which has been hypothesized to have an analogue in visual processing. Again, see Lee, Mumford, Romero, and Lamme (1998).

The effect of reflection symmetry on serial scrutiny has been shown by eye movement, or eye tracking, devices, which give a record of visual scanning behaviour by subjects presented with various visual stimuli. Locher and Nodine (1973), for example, observed that the eye movements of subjects who examined a randomized sequence of shapes were concentrated on one side of the axis of symmetry for symmetrical shapes. However, their eye movements did not exhibit such one sided scanning patterns for asymmetrical shapes. In a later, more refined, study they found that eye fixations were concentrated along a perceived axis of symmetry (Locher & Nodine, 1987). This approach might be advanced by a study that looks for scanning behaviour and fixation points consistent with the notion of submedial axes. However, because of visual span, which is broadened when expectation is satisfied and narrowed when further scrutiny is required, a multitude of minor submedials would not normally be expected to invoke saccades. Location of attention cannot be observed like that of gaze. Yet, there might be saccades associated with more dominant submedials; between the middle and peripheral of a figure, for example.

Concluding remarks

The problem of salience with regard to surface descriptors is a matter at hand for cognitive psychology. Artificial neural networks that employ competition through lateral inhibition are shown to simulate the property of the retina that highlights differences. This thesis investigates salience primarily based upon considerations of proximity; and any edge, graded change, or regularity, as might be detected by such a network, constitutes salience. These, approximated by some contiguity of points, permit the Voronoi process to produce symmetry descriptors in the form of relevant medials. The larger picture for salience, of course, includes considerations of where attention is directed, by what demands, and how narrow or broad the focus is: all associated with executive functions.

Having demonstrated some possibilities, it might be reasonable to implicate operations of a neural character in much of what proceeds from the application of Voronoi, Delaunay explanations of previous chapters. Neural operations that produce a Voronoi-like partitioning of the field, which results in a geometry at all relevant spatial scales, could well serve the perception of organization inasmuch as the proximity principle is concerned. Such partitioning facilitates delineation of structure in the form of relations among visual features that determine which features are perceived as belonging together.

Clearly not all can be researched in an experimental manner by one researcher given a limited time. Hence the theoretical orientation contained herein is necessarily accompanied by limited experimental support. The primary intent is to propose a theory, based upon somewhat novel methodology, for psychological consideration.