42 Mathematical models for the reconstruction of prehistoric settlement patterns: Central German examples

Thomas Weber

42.1 BASIC ASSUMPTIONS

If we wish to analyse prehistoric settlement patterns via mathematical models, we have to presume that maps of recently found archaeological material represent the real prehistoric situation at least to a certain degree. There must not necessarily be a linear connection, but it is most important that the selection processes responsible for the survival of the finds operate in the same way on different groups with respect to their settlement behaviour. Using techniques of point pattern analysis as I shall show here, it is obvious that all the sites observed in the landscape can only be seen as points on the map — without any gradation of their importance or size.

42.2 METHODS

Two large groups of methods can be distinguished for the description of point patterns. The first one is based on cell counts — the whole map is divided into a number of cells usually quadrats of the same size. The numbers of items (find–spots) in these cells are counted. The result of this counting can be used to calculate the spatial autocorrelation coefficient, measuring the randomness of neighbourhood between quadrats more, less or not filled with points. They can also serve as a basis for establishing a frequency distribution showing the numbers of quadrats with 0, 1, 2, ..., *n* find–spots. Observed distributions may be compared with theoretical (e. g. Poisson, negative binomial, and Neyman) models adapted to the given conditions.

The second group of point pattern descriptive methods include distance measurements between the sites observed on the map. Nearest neighbour analysis based on the distance between each point and its nearest neighbour is the best known example. If we also use the distances to the second, third, etc. nearest neighbour we speak of a second, third order nearest neighbour analysis, up to an "all neighbour" analysis (Dacey 1963; Cziesla & Lindenbeck 1989).

Pair correlation functions represent the probabilities of finding certain distances between the points in the area measured on a continuous scale. Observed curves of these values can be compared with theoretically elaborated models.

Furthermore, triangulation and tessellation describing Thiessen polygons surrounding the sites may give us a third way to analyse archaeological distribution maps (Zimmermann 1991, pers. comm.).

There have been several attempts to study the spatial behaviour of prehistoric Central German populations using different techniques of analysis. The region between the Thuringian Forest, the Erz Mountains and the North German Lowlands (Figure 42.1) was a former part of the GDR, and therefore there were poor conditions for obtaining the necessary hard– and software. One of the three attempts, however, was undertaken by a foreign colleague.

42.3 EXAMPLES

42.3.1 Nearest Neighbour Analysis: Neolithic in the Lower Saale area

N. Starling investigated the distribution of the Neolithic cultures in a 4700 km² area surrounding the Lower Saale primarily using the material in the archives of the Halle Landesmuseum. He

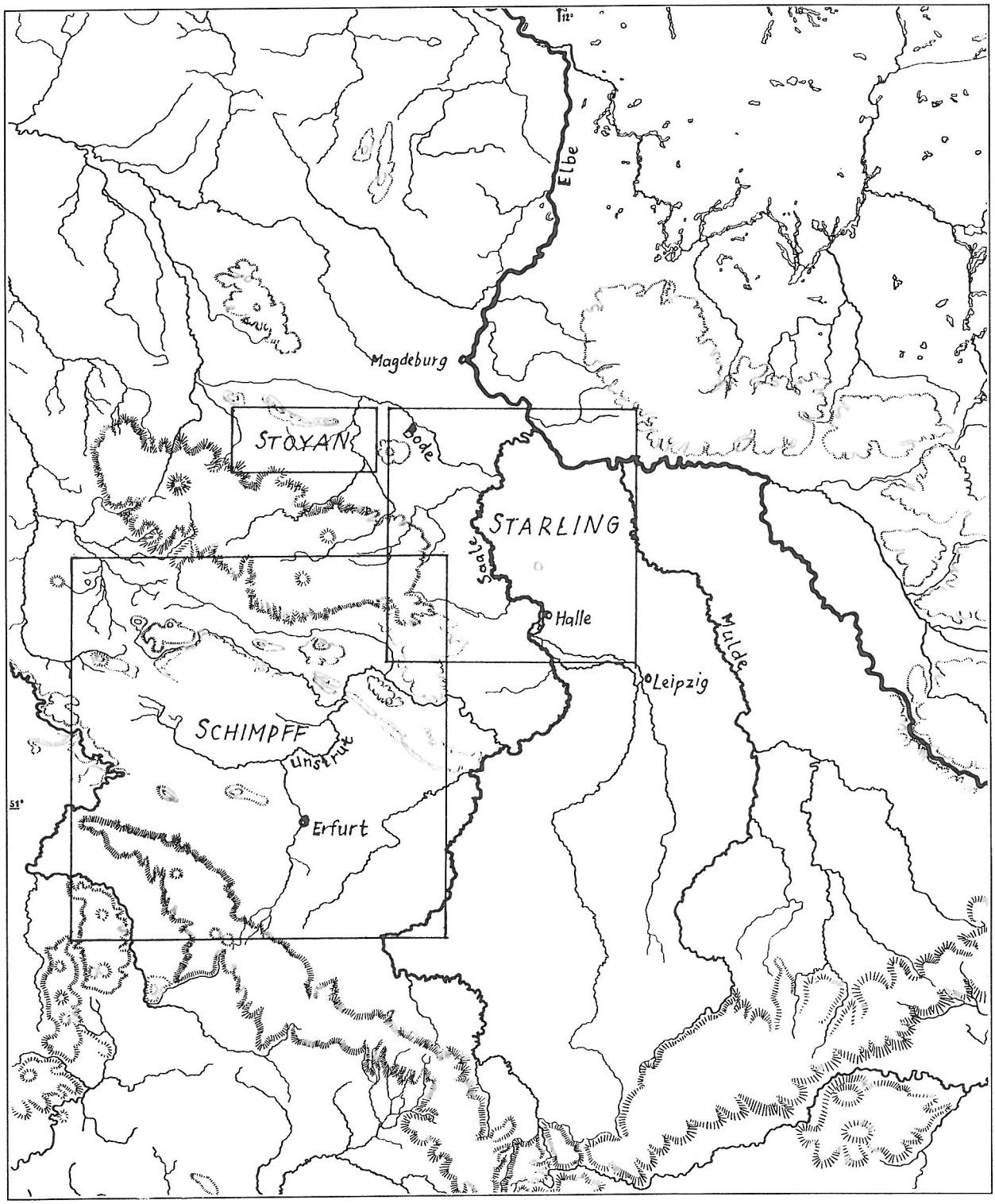

Figure 42.1: Map of Central Germany showing the working areas for Starling's study of Neolithic sites, Stoyan's investigation of medieval and modern settlements, and research on Merovingian land-use in the 6th and 7th century by Schimpff and Weber.

published his study in 1983 — especially the results of the spatial analysis (Starling 1983). The area is mostly covered by Weichselian loess with a black soil (Cernozem), giving high fertility and therefore used for agriculture from the early beginnings of the Central European Neolithic (6th millennium BC — Behrens & Rüster 1981). The plains are relatively dry with rainfall of about 500

mm per year so that the flat and rather small river valleys — Saale and Bode — and also the smaller rivulets are important for settlement behaviour. They do not form real boundaries. Only in the north–eastern corner of the map, the Elbe separates the region of sandy soils from the main area with its loess plains.

These conditions not only yielded a good basis for prehistoric agriculture and settlement but — of course — for modern tillage, too. Archaeological surface finds as a result of ploughing were and are very numerous during the last decades, and quite well organised rescue archaeology was able to find a larger number of remains. Comparing the numbers of find–spots for the different cultures discovered up to 1930 and between 1930 and 1980, one may establish the same relationships notwithstanding the clearly increased absolute numbers of sites. Therefore Starling seems to be right when he believes the relationships are representative for the area. The area itself may also be seen as representative of a larger region in the Central German Neolithic showing the characteristic traits of this landscape at that time.

The nearest–neighbour statistics, including distances between each site and its five nearest neighbours, showed interesting changes from the Linear Bandkeramik to the following cultural groups in this area. Points nearer to the border of the map than to their nearest neighbours have been excluded from the calculations to avoid boundary effects. As for the different — mostly pottery–defined — cultures numbers of sites vary between 33 (Salzmünde) and 266 (Corded Ware — Schönfeld) all the nearest neighbour values are significant at different levels (Figure 42.2; cf. Starling 1983, table 1–3). For the calculations Starling used the well–known formula

[1] $$R = \frac{D_{obs}}{D_{ran}}$$

[1a] where $$D_{ran} = \frac{1}{2\sqrt{p}}$$

and p = point density (Starling 1983:9). The result can vary between 0 for «a perfectly aggregated and or clustered distribution» and 2.1491 for «the maximum regularity ... in a uniform hexagonal pattern», with a value of 1 «for an ideal random pattern» (Whallon 1974:18).

Linear Bandkeramik has been found at 143 places included in the analysis. If we try to interpret the relatively high nearest– neighbour value

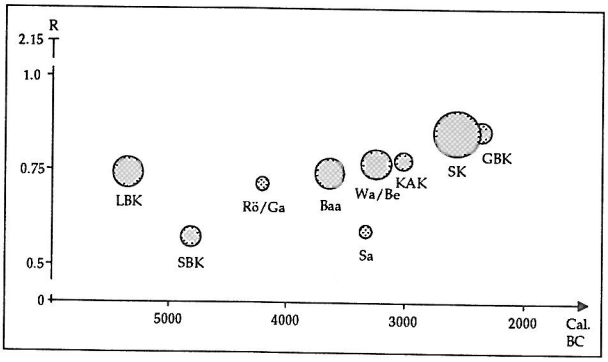

Figure 42.2: Nearest neighbour statistics for the Lower Saale Neolithic including distances up to the fifth nearest neighbour (after values given by Starling 1983). The (pottery–defined) cultures ordered by calibrated radiocarbon years (abscissa) show different nearest neighbour coefficients (ordinate). Different circle radii represent different numbers of sites.

of 0.739 we have to consider its duration of perhaps more than 500 years. Thus Stichbandkeramik sites have been found in quite a more concentrated pattern (0.592) while for Rössen/Gartersleben, the clustering decreased (0.712). High continuity percentages above 75% (Starling 1983:5, fig. 1) show an unbroken trend in the settlement behaviour of the Early Neolithic populations.

Middle Neolithic finds begin with Baalberge at a dispersed level of R = 0.734, are much more concentrated in Salzmünde (0.611) and have been discovered at a higher degree of dispersion (0.766) for the Walternienburg and Bernburg culture. The low continuity values (for Baalberge to the Early Neolithic at 33.3% — Starling 1983:6, fig.2) show the impressive settlement changes from the Early to the Middle Neolithic even in a limited loess–covered area.

The Late Neolithic material is divided in three cultural strata all showing very dispersed settlement patterns with tendencies to increase nearly to pure random values (Globular Amphora 0.777; Corded Ware/Schönfeld 0.838; Bell Beaker 0.845). Following Starling this trend represents a growing use of the landscape with a re–colonisation of abandoned areas and expansion into new ones caused by increased population. All these cultures, however, are primarily represented by grave–finds.

42.3.2 Deserted and surviving villages in the northern Harz Mountains' forelands: pair correlation functions

D. Stoyan (1986) analysed the distribution of modern and deserted villages (*Wüstungen*) in a 1000 km² region north at the Harz Mountains

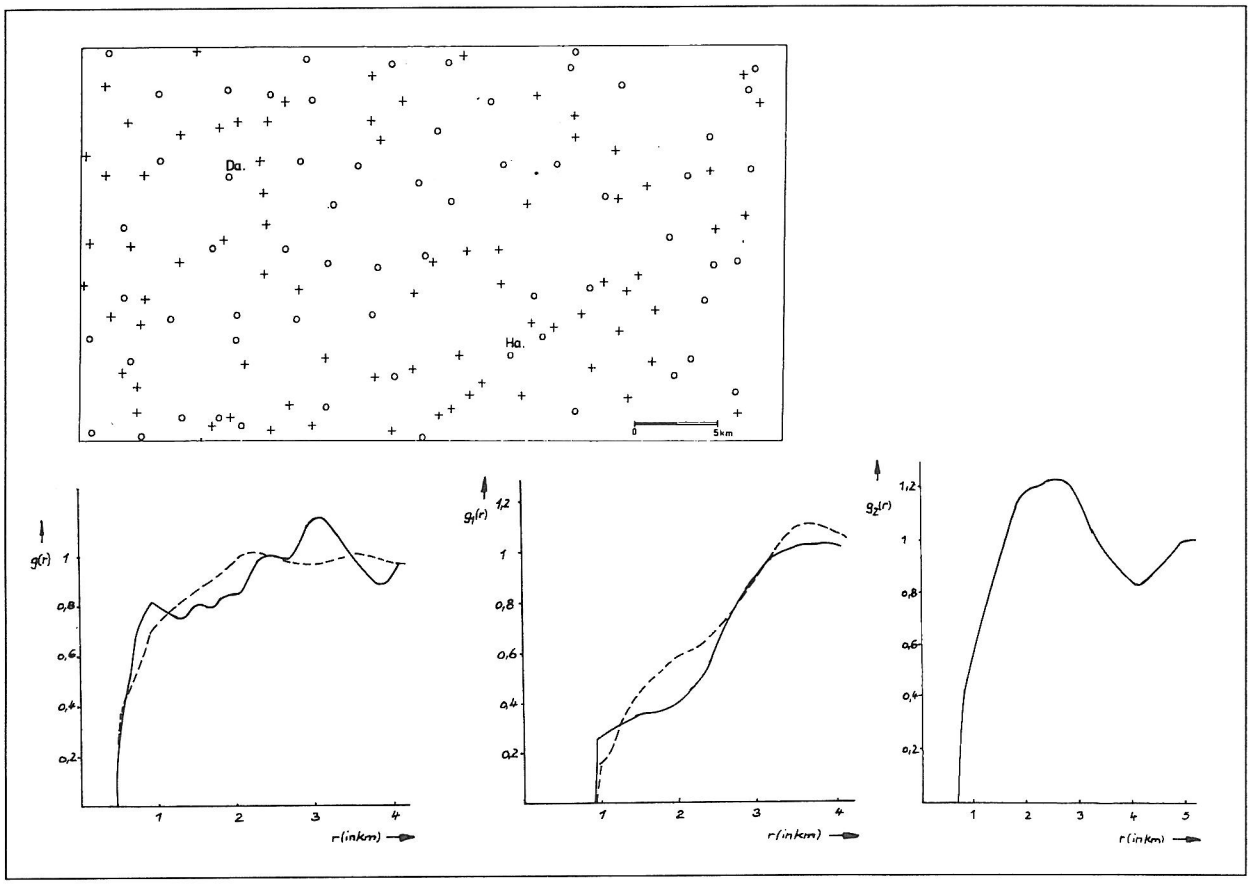

Figure 42.3: Map of Stoyan's 76 deserted (+) and 63 "surviving" medieval settlements in the region surrounding Halberstadt with pair correlation curves for all distances (left), distances between modern "surviving" (central), and deserted settlements (right). Broken lines indicate theoretical assumptions using a soft–core model. For further explanation, see the text.

near Halberstadt. Using pair correlation functions between geographical positions of the modern and the medieval settlements he tried to explain the late medieval settlement concentration process.

The *Wüstungen* are chiefly known from literary sources, especially field names that survived to the beginning of the 19th century and were collected by the *Historische Commission für die Provinz Sachsen* during its period of activity. Some of them could be verified by surface finds so that the chronology sometimes could be determined. From the find material, and from the onomastic and documentary viewpoint it may be concluded that all the recently extant ("surviving") settlements already existed during the middle ages. Stoyan investigated three populations in his study:

1) all medieval settlements containing the modern towns and villages as well;
2) presently extant ("surviving") places; and
3) deserted villages (*Wüstungen*) abandoned for different reasons mostly during the course of the middle ages.

Pair correlation functions describe the probabilities of finding certain distances between points (sites) measured on a continuous scale in this case in km. (The construction depends on an observed arithmetic mean λ (Lambda) showing the number of sites per km² so that λdF gives the probability of placing a point (site) in an infinitesimal small circle.) Pair correlations can be expressed by the formula

$$[2] \quad P(r) = g(r)\lambda dF_1 \lambda dF_2$$

where $g(r)$ represents the pair correlation function (Stoyan 1986:36–37). Its value tends to 1 for larger r's (independence of the observations), whereas lower values imply rarer cases and higher rather frequently found distances. These pair correlation functions are used especially in physics for the three–dimensional case (Stoyan & Mecke 1983:51). In our case, of course, only two dimensions are necessary.

Here three kinds of measurements may be made:

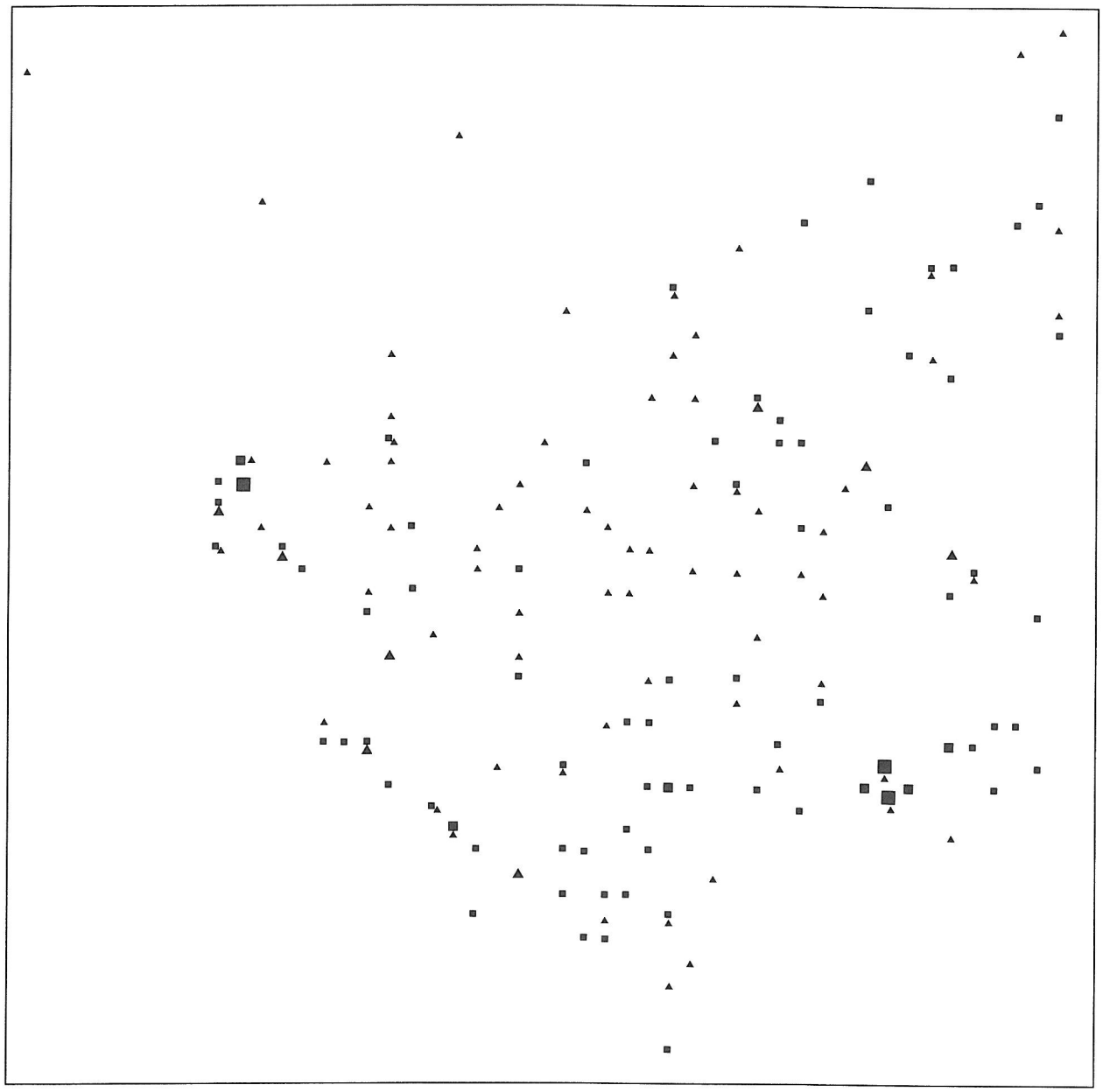

Figure 42.4: Distribution of 6th–century–sites (filled in circles) and 7th–century–sites (filled in triangles) in a 10 000 km² region including the Thuringian Basin. Since many of the find– spots cannot be mapped exactly only the 2 km–quadrats are indicated in which they have been found in varying numbers. The three size–categories of the symbols indicates 1, 2 and 3–4 finds respectively.

1) distances between all points (including modern settlements and deserted villages),
2) between the modern settlements only, and
3) between the deserted villages alone.

Figure 42.3 (upper part) shows the map containing modern (o) and deserted (+) settlements and the curves for the three pair correlation functions. All points (practically the picture for the middle ages) existed at a minimal distance of 0.4 km with quickly growing probability up to 0.9 km, reaching a value at 0.8 for $g(r)$. Between 0.9 and 2 km no important change could be found, while between 2 and 4 km a maximum appears with g(r) = 1.2 at 3 km. The modern settlements are much further apart, at least 0.9 km. Values up to 2 km, however, are rare, with $g(r)$ between 0.3 and 0.4. A slight increase follows with $g(r) = 1$ near 3 km. If we consider the pair correlation function only for the deserted villages, we can recognise the special conditions responsible for settlement decay. Small distances, of course, have been ob-

served very often but not only on the smallest level below 1 km. A clear maximum with $g(r)$ = 1.2 is formed for distances between 1.5 and 3 km (apparently corresponding to the low probabilities in the modern settlement behaviour even at these distances).

Stoyan tried to explain this picture using a theoretical model: the so–called soft–core process. For small distances, a strong repulsion between the points must be recognised, while this repulsion tends to decrease when the distances are larger. He simulated a point process beginning with a dense distribution of points on a map — a starting picture for modelling the medieval settlement. To explain the following development all the points have been supplied with a "strength value" responsible for their prosperity (as a result of geographical position, etc.). These values might change between 0 and 1 independently from the position of the sites (equal distribution). Later a process of "thinning" took place: twice — for the explanation of (i) the medieval and (ii) the modern situation. Only a limited number of (primary) points survived, dependent on their strength and/or their position on the map in the neighbourhood of "stronger" points. Under certain conditions (Stoyan 1986: 44–45), it was possible to find critical distances for the development of settlements — stronger points showed radii between 1.2 and 2.4 km (medieval situation) resp. 1.8 and 3.3 km (post–medieval abandonment process) in which serious difficulties existed for surviving of neighbouring villages while for the larger number of "weaker" settlements this value lay only near 0.4 resp. 0.9 km (Figure 42.3, lower part). The changed distances reflected the changed — increased — economic territories in modern times.

42.3.3 Merovingian settlements in the Thuringian Basin: Quadrat counts of site densities

The question of an intensified land utilisation (*Landesausbau*) in the Thuringian Basin during the Merovingian period has been investigated in V. Schimpff's (1987) study. The distribution of find–spots (mostly graves) in a 10000 km² area has been compared for the 6th and the 7th century using spatial autocorrelation coefficients and square counts in different grids. As Schimpff showed, the cemeteries were situated not far from the villages so that their distribution actually reflects the settlement distribution in a way quite more readily recognised than use of settlement material itself, not least because grave goods can be better dated. Since in this part of Thuringia a well–established tradition of surface find collecting and rescue archaeology has existed over a period of more than 100 years, the distribution of material ought to mirror the actual prehistoric pattern.

A larger part of the cemeteries has only been recorded under the name of the parish in which they have been found many years ago. The exact find–spots have not been recorded. Under these circumstances, a quadrat grid of 2 by 2 km seemed to be the highest possible level of precision for the mapping of the find–spots. If precise locations were known, a minimal distance between two burials of 200 m was accepted to form two separate items so that different parts of the same cemetery could be excluded.

The area is a more or less heterogeneous territory bordered by the Harz and Thuringian Forest Mountains and divided not only by smaller hill regions but also by the wet Unstrut lowlands. Therefore a pure random distribution of settlement points could be excluded *a priori* and an equal distribution as well. The sites were distributed inhomogeneously over the working area, for the two centuries in different patterns (Figure 42.4). The question remained as to the character and importance of this change.

A certain level of clustering could be recognised for the find–spots «in an intuitive manner». Our — the calculations have been done by myself — first attempt tried to exclude these clusters using spatial autocorrelation for gauging grid–size: Following Deiters (1974:55) Geary's contiguity ratio was calculated:

$$[3] \quad c = (N-1)\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\delta_{ij}(x_i - x_j)^2}{4A\sum_{i=1}^{N}(x_i - \bar{x})}, \quad i \neq j$$

where N = number of quadrats (i,j = 1,2,...,N); x_i, x_j = numbers of points in the quadrats

$$[3a] \quad \bar{x} = \frac{1}{N}\sum_i x_i;$$

δ_{ij} = "weights": δ_{ij} = 1 when both are connected side by side, else δ = 0; A = number of all connections between the quadrats. Its value can vary between <1 (positive autocorrelation, clustering) and >1 (negative autocorrelation, regularity) whereas the value 1 describes a real random distribution of points on the map and is "desirable" in our case. The difference between any one value calculated from our observations and a spatial autocorrelation at 1 can be verified using the Z–test. Null hypothesis: No significant difference exists be-

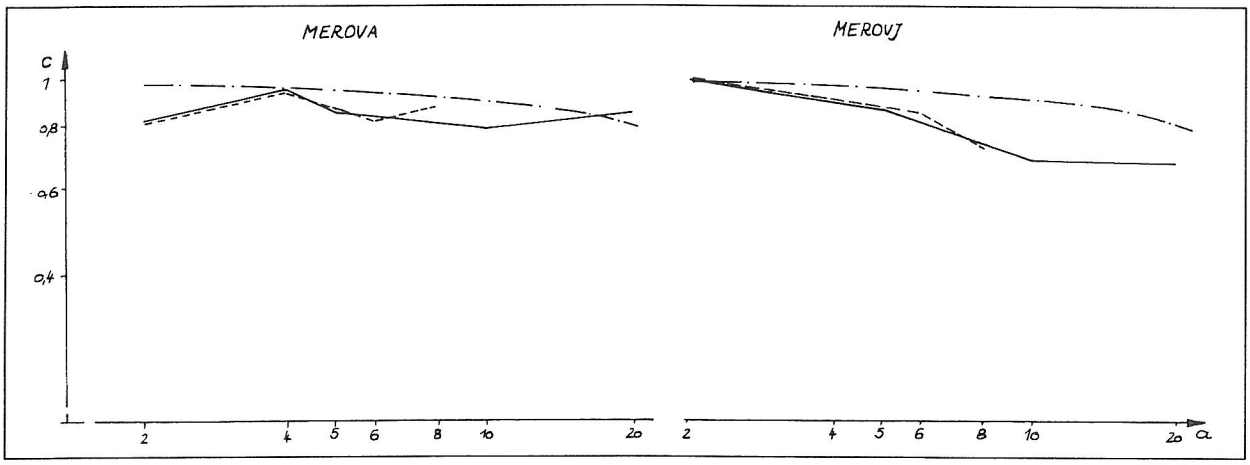

Figure 42.5: Spatial autocorrelation coefficients (c) for different grids of Early (MEROVA) and Late Merovingian sites (MEROVJ) in the Thuringian Basin. Observed values for the whole area (unbroken) for the map excluding the border quadrates to exclude "boundary effects" (broken) and limits for insignificant departure from random value c=1 with α=0.05 (broken–pointed line). The abscissa indicates the quadrate–side length for the different grids (logathmically scaled).

tween the expected — no spatial auto–correlation in the material — and the observed value. Under the assumption of normal distribution we assign the moments of sample distribution for *c*:

[4] $$M_1 = E(c) = 1$$

[5] $$M_2 = V(c) = \frac{(2A+D)(N-1)-2A^2}{(N+1)A^2}$$

[5a] where $D = \frac{1}{2}\sum_i x_i$

[6] $$\sigma_c = \sqrt{V(c)}$$

We obtain

[7] $$Z = \frac{E(c)-c}{\sigma_c} = \frac{1-c}{\sigma_c}$$

using the standard normal distribution so that we can prove the null hypothesis H(0): $c = 1$ that must be rejected when $|Z| > Z(\alpha)$. (Here we subtract the observed from the expected value for agreement of sign with the direction of the spatial autocorrelation.)

Figure 42.6 shows that these "ideal" conditions for further quadrat investigations have been fulfilled only one time: for the 2 km squares of the later period. In this case, however, as mostly for the smaller squares, the theoretical assumptions for the different models do not differ significantly for rather limited numbers of find–spots as we have here. Whenever we try to use quadrat counts to adapt any theoretical distribution to the observations given in our example, we also have to include the wider grids as always giving "clustered" distributions (from the point of spatial autocorrelation).

Table 42.1 shows the values observed with different grids for both distributions together with estimated values for three theoretical models:

42.3.3.1 Poisson distribution

This model is based on E. Weber (1972:134). The arithmetic mean calculated from the observed frequencies of cells with 0, 1, 2 ... points serve as estimation value for Lambda. Therefore we get

[8] $$H_0 = ne^{-\lambda}$$

and further

[9] $$H_i = \frac{\lambda}{i} H_{i-1}$$

The occurrence of a Poisson distribution reflects a total random pattern of settlements without any trend to clustering or to regularity. It can hardly be expected under the conditions of natural diversity of the landscape that influenced the possibilities of its use.

42.3.3.2 Negative binomial distribution

The series is given by the formula

Grid size	N points ↓	Observed	Poisson ↓	Binomial 1	Binomial 2 ↓	Neyman
2x2 km	0	2422	2409,67	2407,71		2425,62
-	1	69	88,68	76,38		12,23
-	2	6	1,63	12,51		12,23
-	3	0	0,02	2,60		2,11
-	4	2	0,00	0,60		0,32
4x4 km	0	560	539,45	529,39		576,33
-	1	54	79,41	57,90		24,14
-	2	5	5,84	20,37		13,30
-	3	4	0,64	8,82		6,52
-	4	0	0,01	4,17		2,88
-	5	0	0,00	2,07		1,16
-	6	0	0,00	1,07		0,44
-	7	1	0,00	0,56		0,16
-	8	0	0,00	0,30		0,05
-	9	1	0,00	0,16		0,02
5x5 km	0	337	317,81	319,52	337,01	348,00
-	1	48	73,10	49,77	44,17	28,01
-	2	10	8,41	17,39	12,39	13,97
-	3	2	1,63	7,20	4,10	6,16
-	4	1	0,04	3,21	1,46	2,46
-	5	0	0,00	1,50	0,54	0,91
-	6	1	0,00	0,72	0,21	0,32
-	7	0	0,00	0,35	0,08	0,11
-	8	1	0,00	0,17	0,00	0,04
10x10 km	0	60	39,85	59,44	60,00	66,28
-	1	21	36,66	18,38	18,74	10,94
-	2	7	16,87	9,12	9,11	8,17
-	3	4	5,17	5,09	4,96	5,67
-	4	3	1,19	3,00	2,85	3,67
-	5	3	0,22	1,83	1,68	2,25
-	6	0	0,04	1,13	1,01	1,32
-	7	0	0,00	0,71	0,62	0,76
-	8	1	0,00	0,45	0,38	0,42
-	9	0	0,00	0,29	0,24	0,23
-	10	1	0,00	0,19	0,15	0,13
20x20 km	0	10	0,63	10,66	10,00	11,14
-	1	2	2,32	4,06	3,54	1,54
-	2	3	4,27	2,50	2,22	1,48
-	3	0	5,24	1,74	1,60	1,41
-	4	3	4,82	1,27	1,23	1,32
-	5	0	3,55	0,96	0,97	1,22
-	6	3	2,18	0,75	0,79	1,10
-	7	0	1,14	0,59	0,65	0,98
-	8	0	0,53	0,46	0,55	0,85
-	9	0	0,22	0,37	0,46	0,73
-	10	1	0,08	0,30	0,39	0,61
-	11	0	0,03	0,24	0,33	0,51
-	12	1	0,01	0,20	0,29	0,42
-	13	0	0,00	0,16	0,25	0,34
-	14	1	0,00	0,13	0,21	0,27
-	15	0	0,00	0,11	0,18	0,22
-	16	0	0,00	0,08	0,16	0,18

Grid size	N points ↓	Observed	Poisson ↓	Binomial 1	Binomial 2 ↓	Neyman
-	17	0	0,00	0,07	0,14	0,14
-	18	1	0,00	0,06	0,12	0,11
2x2 km	0	2416	2409,67	2424,55	2415,27	2415,51
-	1	76	88,68	69,71	78,13	77,46
-	2	8	1,63	5,22	5,99	6,57
4x4 km	0	549	539,45	558,90		547,79
-	1	61	79,41	56,56		64,43
-	2	14	5,84	8,04		11,02
-	3	1	0,29	1,25		1,54
5x5 km	0	329	317,81	342,37	328,99	326,39
-	1	51	73,10	47,71	55,10	58,38
-	2	19	8,41	8,12	12,10	12,56
-	3	1	0,64	1,46	2,87	2,24
10x10 km	0	61	39,85	67,37	61,00	58,38
-	1	14	36,66	18,44	18,02	16,36
-	2	6	16,87	7,54	8,78	12,02
-	3	13	5,17	3,42	4,84	6,79
-	4	5	1,19	1,63	2,82	3,43
-	5	0	0,22	0,80	1,70	1,65
-	6	0	0,04	0,40	1,04	0,77
-	7	1	0,00	0,20	0,65	0,34
20x20 km	0	7	0,63	7,36		8,18
-	1	3	2,32	4,54		2,29
-	2	2	4,27	3,20		2,21
-	3	6	5,24	2,34		2,07
-	4	0	4,82	1,76		1,87
-	5	0	3,55	1,33		1,63
-	6	1	2,18	1,01		1,39
-	7	1	1,14	0,78		1,15
-	8	2	0,53	0,60		0,94
-	9	9	0,22	0,46		0,75
-	10	1	0,08	0,36		0,59
-	11	1	0,03	0,28		0,46
-	12	1	0,01	0,22		0,36
-	13	1	0,00	0,17		0,28
-	14	2	0,00	0,13		0,21

Table 42.1: 6th–(left column) and 7th–century–sites (right column) in Thuringia analysed with different grids with quadrat sizes of 2x2, 4x4, 5x5, 10x10, and 20x20 km. The observed values are shown and compared with the fitted theoretical distributions: Poisson; negative binomial estimated by maximum likelihood method (Binomial 1) and by a "shortened" technique — where possible (binomial 2); Neyman (type A for the 7th century sites; type C for the 6th century sites).

$$[10]\quad P_x=\binom{n+x-1}{x}p^x q^{-(x+n)}=\frac{(n+x-1)!\left(\frac{p}{q}\right)^x}{x!(n-1)!q^n}$$

where P_x is the probability to obtain a quadrat with x individuals; k (=np)is the exponent of the distribution to be estimated; p = the probability of getting a "positive event" (here: occurrence of a find–spot in any one quadrat). As E. Weber (1972:144) has shown in the case of "over–dispersion", with a variance larger than the arithmetic mean, it should become negative; and q = the complementary event to p therefore will be >1.

When the sample size is n and there are a_x occurrences for the x individuals the expectation for a_x can be expressed as

$$[11]\quad E(a_x)=m_x=n\frac{(k+x-1)!p^x}{x!(k-1)!(1+p)^{k+x}}$$

The exponent k can be estimated in different ways:

(a) using a maximum likelihood technique (E. Weber 1972:220–221)

The number of observations larger than a(x) for each x is marked as A(x):

$$[12]\quad A_x=a_{x+1}+a_{x+2}+\cdots$$

The maximum likelihood estimation for k is given when

$$[12a]\quad \sum_x\frac{A_x}{k+x}-n\ln\left(1+\frac{\bar{x}}{k}\right)=0$$

This requirement has been met using a small program in Basic.

(b) a "shortened" technique based on certain assumptions (E. Weber 1972:219–220, 224). This variant could be used for arithmetic means <10 if the following inequality was fulfilled:

$$[13]\quad (m+0.17)(P_0-0.32)>0.2$$

Sometimes this was the case, and it was remarkable that the estimated values for cell frequencies often fitted better to the observations than the results of the maximum likelihood method (cf. Figure 42.6).

Under these circumstances the frequency of cells without points can be used for the estimation

$$[14]\quad R_2\log\left(x+\frac{\bar{x}}{R_2}\right)-\log\left(\frac{n}{z_0}\right)=0$$

where

$$[14a]\quad \left(1+\frac{\bar{x}}{R_2}\right)=q$$

I have also done this iteratively using a small Basic program.

The negative binomial distribution may obviously be used in many cases to understand settlement patterns. Perhaps the model shows the geographical inhomogeneity in a given area — differences of its sources, relief, rivers, soil types, flora and fauna patterns, etc. — and can often be interpreted as reflection of the "basic situation" when sites were "randomly" founded but under "variable circumstances". In that sense, not the proof for a negative binomial distributed settlement is the most interesting point in a study about archaeological sites in a given landscape, but rather the change of the parameters from period to period or even the possible adaptation of another kind of mathematical model for a later or preceding phase.

42.3.3.3 Neyman–type–A,B, or C–distribution (E. Weber 1972:154–157)

This «compound Poisson or apparent contagion process simply implies random inhomogeneity in the density of the population» (Hodder & Orton 1976:87) but may surely be seen here as a result of a "contagious process" starting from a negative binomial true clustered pattern by settlement filiation in which the new founded villages preferably originated in the neighbourhood of former existing sites.

The Neyman distribution can be adapted to an observed series using its arithmetic mean and variance. For the three types, there are different algorithms characterised by the critical parameter m (E. Weber 1972:155). Its value is given at 0 (Neyman A), 1(B), or 2(C). Special parameters are calculated by

$$[15]\quad m_2=\frac{(m+2)(s^2-\bar{x})}{2\bar{x}}$$

$$[15a]\quad m_1=\frac{(m+1)\bar{x}}{m_2}$$

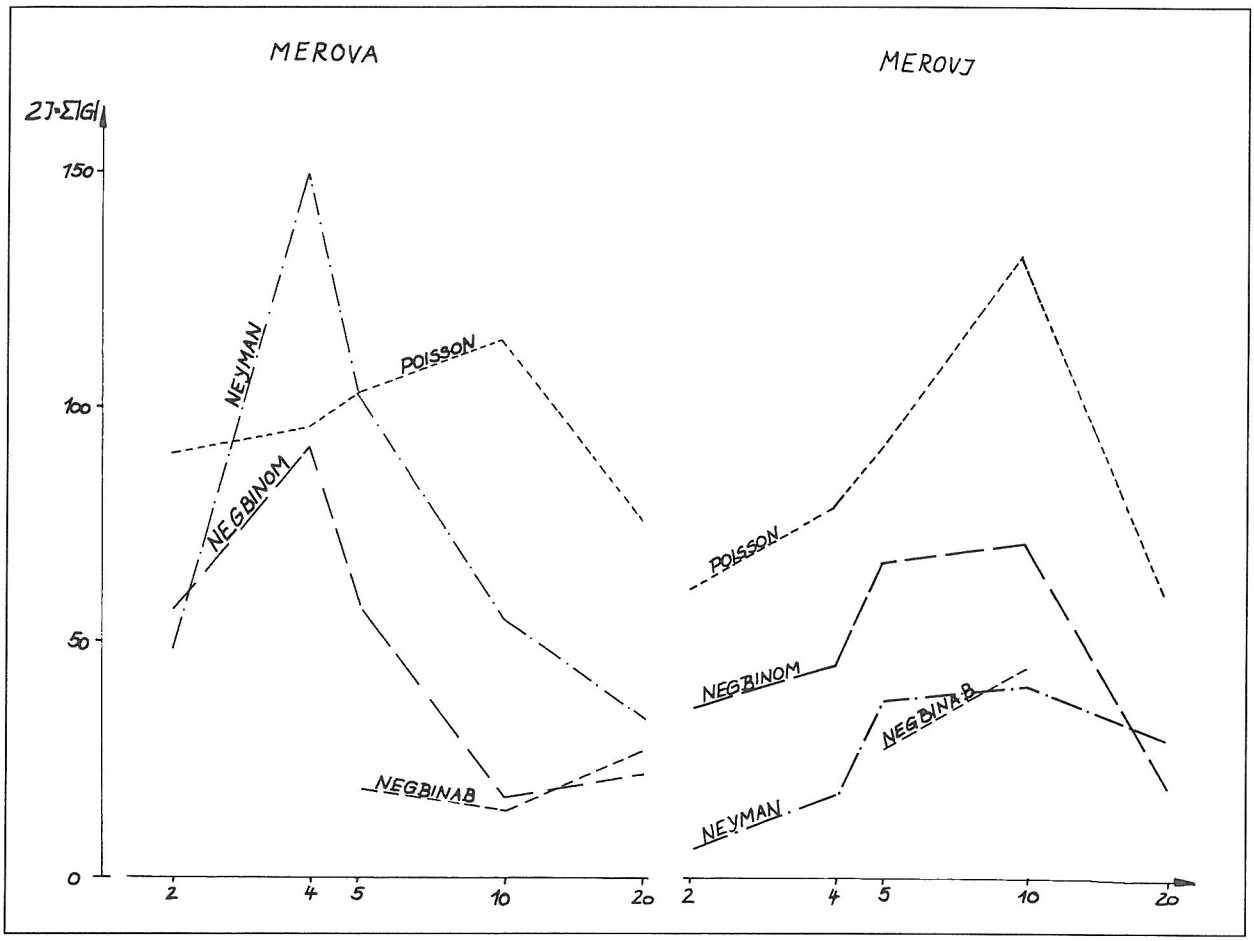

Figure 42.6: G (2I) values for the differences between observed and adapted theoretical distributions of Early (MEROVA) and Late Merovingian sites (MEROVJ). The abscissa indicates the side length of the quadrate for the different grids (logarithmically scaled). For further explanation, see the text.

The probabilities are calculated with

[16] $$P\{X=k+1\}=\frac{m_1 m_2 e^{-m_2}}{k+1}\sum_{t=0}^{k}\frac{m_2^t}{t!}P\{X=k-t\}$$

where X = number of events (points per quadrat).
For X = 0 we get

[16a] $$P_0=P\{X=0\}=e^{-m_1(1-e^{-m_2})}$$

for X = 1

[16b] $$P_1=P\{X=1\}=\frac{m_1 m_2 e^{-m_2}}{1}P_0$$

for X = 2

[16c] $$P_2=P\{X=2\}=\frac{m_1 m_2 e^{-m_2}}{2}\left[P_1+m_2 P_0\right]$$

etc.

The frequencies are calculated by multiplication with the number of observations (archaeological sites) for each probability in the series:
We get for X = 0

[17a] $$H_0=nP_0$$

for X = 1

[17b] $$H_1=m_1 m_2 e^{-m_2}H_0$$

for X = 2

[17c] $$H_2=\frac{m_1 m_2 e^{-m_2}}{2}\left[H_1+m_2 H_0\right]$$

etc.
The calculations have been carried out in accordance with [16 and [17] using the values 0, 1, 2, ... for *k* (E. Weber 1972:150).

Real distinction between "true" and "apparent" contagion has been established for the investigation of negative binomial distributions by fusion of neighboured cells and new calculation of the p and k values for predicted frequency distributions. In the case of a "true contagious process" with real clusters the parameters for the original and new quadrats are related by the formulae

[18] $$k_s = sk_1$$

and

[18a] $$p_s = p_1$$

whereas, if generated by a 'spurious contagion' situation, the relationships (Hodder & Orton 1976:88) would be

[19] $$k_s = k_1$$

and

[19a] $$p_s = \frac{1}{1+s\left(\frac{1-p_1}{p_1}\right)}$$

When solution (1) represents a total random configuration of point structure, and solution (2) can be understood as a strictly clustered pattern, solution (3) may be interpreted as spatial realisation of a "contagious process" combining a "clumped" picture with a tendency of dispersed clusters. Therefore its degree of point agglomeration decreases — compared with the negative binomial model (Deiters 1974:49).

The results of the comparisons between observed and expected distributions can be tested using the Chi–square method (Deiters 1974:53–54). As it is evident in our case, the assumptions for the expected values can often be fulfilled only when class frequencies are summarised so that the differences between the models tend to be insignificant. Under these conditions no objective decision is possible as to which distribution optimally represents the observed frequency pattern.

Thus the attempt has been undertaken to compare expected and observed values using the *G*–test (Log–likelihood–chi–square) (Sokal & Rohlf 1981; E. Weber 1980) using the well–known formula

[20] $$G = 2\sum_{i}^{a} f_i \left| \ln\left(\frac{f_i}{\hat{f}_i}\right) \right|$$

which «can be seen as the sum of independent contributions of departures from expectation weighted by the frequency of the particular class» (Sokal & Rohlf 1981:698). To use both directions of departures — too small and too large values — the absolute values have been used for sum calculations. All expected frequencies >0.5 (rounded 1) were included. Frequency classes with expectations <0.5 were added. Of course, sometimes we observe sequences that do not fulfil all the demands for a serious use of the test (e. g. too many classes with small expected frequencies), and the degrees of freedom varied depending on the different additions of frequency classes for different theoretical distributions.

Therefore the results given in Figure 42.6 only show quite rough measurements of differences between observed and adapted theoretical distributions expressed by the 2I (or G) values. Generally, the Poisson model shows the largest differences between expectations and the observations (for all the grids in both periods with the exception of the 4 km–grid in the earlier period where the Neyman distribution yielded the largest deviation). Thus the patterns on the map seriously differ from a purely random situation, and the smaller G–values imply rather clustered settlement pattern for the 6th and for the 7th century.

Interestingly, an evident distinction for the both periods could be observed. Whereas the early Merovingian sites follow the negative binomial distribution (especially the model established with the "shortened" estimation technique) for the 7th century, the Neyman distribution gives the best adaptation to the observed frequencies. The results for most of the grids seem to indicate that the supposed process of colonisation in the Thuringian Basin (*Landesausbau*) really took place in the transition from the 6th to the 7th century.

42.4 FURTHER POSSIBILITIES OF SPATIAL STUDIES IN CENTRAL GERMAN SETTLEMENT BEHAVIOUR

The archaeological material from the Middle Elbe/Saale region is excellent for comparative settlement studies. During most prehistoric periods, the fertile soils afforded the best opportunities for different populations. The intensive agriculture of the last few centuries partly destroyed the sources, but it created — together with an elaborate tradition of surface find collecting and rescue archaeology — the framework for the discoveries of many sites in a larger territory than any which could be studied under economic

pressure (archaeological prospecting for open-casts, mines, etc.) in the last decades using more sophisticated site-localization techniques. Thus further investigations may follow to answer the questions for changing settlement behaviour of our ancestors in this region. Different ideas, different methods, and different attempts are welcomed.

References

Behrens, H. & B. Rüster
1981 Kalibrierte C14-Daten für das Neolithikum des Mittelelbe-Saale-Gebietes. *Archäologisches Korrespondenzblatt* 11:189–193.

Cziesla, E. & J. Lindenbeck
1989 All-neighbour-analysis: Eine Methode zur zonalen Darstellung der Zentralität räumlicher Verteilungen. *Archäologische Informationen* 12:237–240.

Dacey, M. F.
1963 Order neighbor statistics for a class of random patterns in multidimensional space. *Annals of the Association of American Geographers* 53: 505–515.

Deiters, J.
1978 *Zur empirischen Überprüfbarkeit der Theorie zentraler Orte.* Fallstudie Westerwald. Bonn.

Hodder, I & Cl. Orton
1976 *Spatial analysis in archaeology.* Cambridge — London — New York — Melbourne.

Schimpff, V.
1987 *Beiträge zur Besiedlungsgeschichte des Thüringer Beckens in der Merowingerzeit.* Unpubl. dipl.-work Martin-Luther-Univ. Halle-Wittenberg.

Sokal, R. R. & F. J. Rohlf
1981 *Biometry. The Principles and Practice of statistics in Biological Research.* 2nd ed. New York.

Starling, N. J.
1983 Neolithic settlement patterns in Central Germany. *Oxford Journal of Archaeology* 2:1–11.

Stoyan, D.
1986 Statistische Untersuchung der Verteilung von Wüstungen im Nordharzgebiet. *Urgeschichte und Heimatforschung* 23:33–46.

Stoyan, D. & J. Mecke
1983 *Stochastische Geometrie.* Berlin.

Weber, E.
1972 *Grundriß der biologischen Statistik.* 7. Aufl. Jena.

Weber, E.
1980 *Grundriß der biologischen Statistik.* 8. Aufl. Jena.

Whallon, R.
1974 Spatial analysis of occupation floors II: The application of nearest neigbor analysis. *American Antiquity* 39:16–34.

Author's address
Thomas Weber
Landesamt für archäologische Denkmalpflege
Sachsen-Anhalt
Richard-Wagner-Str. 9/10
D-(O-)4020 Halle/Saale
Tel +37 34 53 76 21
Fax +37 34 52 97 13