

2.15 or Not 2.15? An Historical-Analytical Inquiry into the Nearest-Neighbor Statistic

Chris Philo¹  and Peter Philo²

¹School of Geographical and Earth Sciences, University of Glasgow, Glasgow, G12 8QQ, UK, ²Independent Scholar

Nearest-neighbor analysis (NNA)—a method for assessing the degree to which a spatial point pattern departs from randomness in the direction of being either clustered or regular—was imported into academic geography from an article published in 1954 by ecologists Clark and Evans. In its simplest form, concerned with distances to the first nearest neighbor, NNA hinged on the behavior of the “nearest-neighbour statistic,” R_n , supposed to vary between 0 (complete clustering) and 2.15 (complete regularity). NNA and the wider body of work on point pattern analysis quickly gained in sophistication, meaning that this simple test statistic only featured briefly in the frontline research literature of geographical analysis. Nonetheless, given its easy-to-grasp logic, it became and remains a staple of quantitative geography textbooks for undergraduate students and statistical methods training in school-level geography curricula. The purpose of this paper is to chart the history of NNA and its test statistic in academic geography, and to provide an analytical demonstration that the value of 2.15 has been mistakenly identified as an upper limit for the latter. Broader speculations are then offered about what may be learned from the history of how the discipline has handled this statistic.

Introduction: revisiting nearest-neighbor analysis (NNA) and its statistic

NNA occupies a minor if not insignificant place in the history of academic geography’s encounter with what has variously been badged as “the quantitative revolution,” “locational analysis,” or “spatial science.” In its simplest formulation NNA calculated a nearest-neighbor statistic or index, typically denoted R_n , supposedly lying between 0 and 2.1491, often rounded to 2.15. The purpose of this article will be to chart the *history* of NNA, and more particularly the treatment of its characteristic statistic, from initial engagement by “scholarly users” (circa 1950s–1960s) through to its deployment as a teaching tool by “educational users” (1970s onwards).¹ The distinction between these different cohorts of “users” is easy to detect in the literature and is central to the history that we are telling. Indeed, we emphasize how interest in NNA, and particularly its statistic, quickly waned for scholarly users but was rekindled by their pedagogic counterparts, for whom the figure of 2.15, the assumed maximum value of the statistic, acquired what might

Correspondence: Chris Philo, School of Geographical and Earth Sciences, University of Glasgow, Glasgow G12 8QQ, UK
e-mail: Christopher.Philos@glasgow.ac.uk

Submitted: 27 August 2020; Revised version accepted: 11 February 2021

doi: 10.1111/gean.12284

© 2021 The Authors. *Geographical Analysis* published by Wiley Periodicals LLC on behalf of The Ohio State University.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, **333** which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

be described as unquestioned, even “mythic,” status. Arising from empirical work by the lead author (Philo 2004, Chap. 5, pp. 374–7), combined with expertise of the second author (e.g., Philo 1969), this article also adds a brief *analysis* to demonstrate that claims about 2.15 as a maximum value are mistaken. Taken together, our twin-track historical-analytical inquiry allows speculations to be advanced about what might be learned from how academic geography has tackled NNA and its statistic.

A prefatory clarification is that two fundamentally different approaches for analyzing spatial distributions have arisen in academic geography and cognate disciplines, with both searching for either clustering of individual elements, their apparent pulling together, or dispersion, their apparent pushing apart (e.g., Moore 1954). One approach—called “quadrat analysis,” evoking botanists laying quadrats (square grids) over patches of vegetation, counting plants of different species within each quadrat—concentrates on the numbers of individuals occupying given portions or “sub-areas” of a study area. The observed density of individuals per sub-area can be compared with expected densities on the assumption of randomness, that individuals are scattered randomly across the area, or against other theoretically derived density distributions (reflecting different statistical families of distribution-generating models). A second approach relies on the spacing or distances between individuals within a study area. The observed distances, measured and then aggregated in some manner, can be compared, in the same vein as above, with expected distances on the assumption of randomness or against other theoretically derived distance parameters. The geographer Dacey (1965, p. 144) distinguished between “quadrat measures” and “spacing measures,” the latter “us[ing] distances between points,” while Dacey and Tung (1962, p. 84) stated that “[a]cceptable methods of identification of pattern types are the quadrat method and ... variations of the nearest neighbour method.”

NNA in academic geography: an outline history

Clark and Evans (1954)

A history of NNA in academic geography begins with the mid-1950s article by Clark and Evans (1954), published in the journal *Ecology*, specifying an innovative method for quantifying the spatial distribution of a given community of plants (or possibly animals) within a given area. As they argued, it is helpful for researchers to be able to measure—to quantify—whether such a distribution displays either clustering, wherein individual members appear to clump together into one or more clusters, or regularity, wherein individuals appear to be regularly or evenly spaced apart from one another. Put another way, the objective is to decide the extent to which a distribution departs from randomness—or no discernable pattern—in the direction of either clustering or regularity, since such departures suggest the operation of “underlying forces active in the formation of particular patterns” (Clark and Evans 1954, p. 446). Randomness was here defined as occurring when every individual or “point” in a distribution “has ... the same chance of occurring on any sub-area as any other point,” every sub-area (however exactly defined) within the overall area possesses an equal chance of “receiving a point,” and when “the placement of each point has not been influenced by that of any other point” (Clark and Evans 1954, p. 446). In situations where “equal chances” are not the case and where points are influencing the position of each other, then “underlying forces” will likely be at work inflecting the distribution toward either clustering or regularity.

Building on prior work by ecologists and others,² these ecologists proposed an investigation where “[t]he distance from an individual to its nearest neighbour, irrespective of direction,

provides the basis for this measure of spacing” (Clark and Evans 1954, p. 447). Their approach was simple: for a particular real-world distribution comprising a known number of individuals occupying a specified area of a known size, it entailed comparing an *observed* measure of “nearest-neighbourliness,” here the mean distance of all individuals from their respective nearest neighbors in that distribution, with an *expected* measure of that mean based on the assumption of randomness in the placement of the same numbers of individuals across the same area. “The ratio of the observed mean distance to the expected mean distance serves as a measure of departure from randomness,” explained Clark and Evans (1954, p. 447). They identified a range of values for this ratio, running from a lower limit of 0 (zero), the hypothetical case where all individuals occupy exactly the same location and the expected mean is 0,³ through to a claimed upper limit of 2.1491, where all individuals are evenly spaced as far apart as possible and where the observed mean reaches the maximum it can do so. A value of 1 obviously arises when the observed and expected means are identical. These three values were taken to signal total *clustering* (0), full *randomness* (1), and complete *regularity* (or *uniformity*) (2.1491). Any real-world distribution could be readily located on this spectrum, informed deductions made about whether it lent toward being clustered, random or regular, and objective comparisons made between it and other real-world distributions. In their own words, Clark and Evans (1954, p. 447) stated that R_n , their ratio, “has a limited range, with values indicative of perfectly uniform, random and completely aggregated patterns of distribution.”

There is more to be said about the Clark and Evans article, such as noting the application of their approach to “actual data,” pertaining to both an artificially contrived distribution and real-world distributions of grassland plants and forest trees. They also acknowledged various “problems of procedure”: many nearest neighbors in real-world distributions are “reflexive” or “pairwise,” leading the same nearest-neighbor distance to be measured twice, although including such “double distances” is statistically justified; the existence of boundary effects, arising from the arbitrary specification of study areas, means that what might otherwise be nearest-neighbor distances are missed because they include individuals lying out with the study area, a complication difficult to avoid; and the fact that in the method “individual components ... have been treated as dimensionless points” (Clark and Evans 1954, p. 450), usually justifiable because individuals are relatively so much smaller than the overall area. It might be added that the visualization here was “in terms of two-dimensional space, i.e. with reference to populations on plane surfaces” (Clark and Evans 1954, p. 446), and deployed square or rectangular “plots,” although in principle study areas could be any shape.

More relevant to what follows is the article’s appendix, where the authors derived a formula for the *expected* measure of the mean nearest-neighbor distance, deploying a Poisson exponential function, a visualization based on sectors of a circle, attention to probability distributions and calculus steps. Here the authors also derived what is supposedly the upper limit of R_n , 2.1491, occurring when individuals are maximally spaced, remarking that “[t]he mean distance between nearest neighbours is maximised in a hexagonal distribution, where each point has 6 equidistant nearest neighbours” (Clark and Evans 1954, p. 452 [also p. 447]). Graphical illustrations of such a distribution will be shown below, although the distribution here might more logically be described as *triangular* rather than hexagonal, a slip that has since led some geographers and others astray.⁴ Tellingly, Clark and Evans (1954, p. 447) also proposed that use of R_n should be complemented by a “test of significance” based on the difference between observed and expected nearest-neighbor distance means, divided by the standard deviation of the expected mean

distance, set against “the normal curve” (thus enabling a Z-test). This procedure arguably came to take precedence—for frontline geographical analysis, if not for geographical education—over any stand-alone interpretation of R_n itself.

NNA and geographical research

Following their 1954 article, the authors and others (Clark and Evans 1955, 1979; Clark 1956; also Thompson 1956) continued to investigate the potential for NNA in ecology, including attention to “higher-order” nearest neighbors (beyond immediate or 1st-order to 2nd-, 3rd or n^{th} -order nearest neighbors). Geographers deploying NNA at the research frontier started to appear in the late-1950s, with the most high-profile contribution straightforwardly following Clark and Evans (1954) being Getis (1964), studying the shifting locations of grocery stores in Lansing, Michigan. He calculated and compared R_n for these retail distributions across seven different years—by decade, from 1900 to 1960—disclosing “a consistent U-shaped trend” (Getis 1964, p. 395) whereby R_n was just above 1.0 in 1900, indicating an initially random pattern, before dipping down closer to 0.5 in 1920, suggesting clustering, and then gradually returning back toward 1.0 by 1960 (Fig. 1). He explicitly stated the possible range for R_n to lie between 0 and 2.1491 (Getis 1964, p. 395). Intriguingly, he began his article by lamenting that NNA “has been successfully used by plant ecologists, but with less success by geographers,” claiming that the early (unpublished) efforts to use NNA by the likes of Dacey and Berry led them to “criticise it for a number of reasons, the most prominent being insufficient rationale for objectively defining the study area” (Getis 1964, p. 391). As Getis realized, the same distribution of points viewed through different-sized “sampling windows,” including more or less of the points, may result in differing values of R_n . A particular difficulty then arises when a distribution under study is solely contained within a given area, such that progressively increasing the size of the sampling window beyond that area results in the distribution appearing increasingly clustered (with R_n values decreasing accordingly).⁵ Getis’s response was a guarded restatement of NNA’s usefulness in situations where the researcher could sensibly justify the prior specification of a study area, and when undertaking temporal comparisons of R_n ’s behavior over time in the same study area, as in his Lansing inquiry. Sherwood (1970) emulated Getis in this regard, studying premises occupied by grocers and solicitors-accountants in Shrewsbury, United Kingdom, for various dates, but allowing and justifying a progressive enlargement of the study area in tandem with the town’s own urban expansion.

TABLE 1.—SUMMARY OF NEAREST NEIGHBOR MEASURES^a

Year	N	ρ	\bar{r}_A	\bar{r}_B	R	σ_{r_B}	z
1900	20	0.000000594	696.29	648.50	1.074	75.82	-0.63
1910	33	0.000000899	354.87	527.31	0.673*	47.98	+3.59
1920	94	0.000000918	343.24	521.87	0.658*	28.14	+6.35
1930	124	0.000001097	368.33	477.37	0.772*	22.11	+4.93
1940	133	0.000000956	405.05	511.35	0.792*	23.18	+4.59
1950	117	0.000000797	470.86	560.10	0.841*	27.06	+3.30
1960	68	0.000000375	814.46	816.46	0.998	51.75	+0.04

^a Values in feet.

* Significant at the 0.01 level.

Source of data: Field data gathered by author for Lansing, Michigan, computations by author.

Figure 1. Summary of nearest-neighbor measures in Getis study; note behavior of R (R_n), the nearest-neighbor statistic (Getis 1964, p. 395).

Continuing his commentary on previous geographical studies, Getis (1964, p. 395) stated that “[t]here are a few unpublished papers in the geographical literature which demonstrate the empirical use of the technique,” adding that, “to the author’s knowledge, only Dacey has attempted an empirical analysis that has been published, and this work is applied to a special case and is, therefore, of limited value.” The latter reference was to Dacey (1960a; also Porter 1960) on “the spacing of river towns,” tackling reflexive and higher-order nearest-neighbor distances, measured by river and by air, between pairs of towns along the banks of the Mississippi. Such an attempt to apply NNA to a “linear” case was amplified subsequently by Pinder and Witherick (1973, 1975) and Selkirk and Neave (1984), the latter speaking of NNA for “one-dimensional distributions” that could include lines, curves, and circles. Oddly, Getis appeared ignorant of a second empirical study by Dacey (1962), wherein he adopted a Clark and Evans (1954) version of NNA to assess whether settlement distributions in South West Wisconsin, United States, conform to “Central Place Theory,” finding limited evidence for regularity of spacing for upper tiers (larger centers) in the settlement hierarchy. Jensen-Butler (1972) paralleled Dacey (1962) when studying a “central place system” in County Durham, United Kingdom, using R_n —and its hypothetical range “between 0 ... and 2.149” (Jensen-Butler 1972, p. 357)—in the search for tendencies toward regularity at different “hierarchical levels.” Nor did Getis register a study by King (1962) which offered a comparative analysis of urban settlement patterns across 20 U.S. regions. Using straight-line distances between all urban places and their nearest neighbors within each region when calculating regional R_n s, ranging from weakly clustered (0.70, Utah) to weakly regular (1.38, Missouri), King (1962) concluded that the overall picture tended toward randomness rather than patterning. Birch (1967), meanwhile, conducted a simple NNA of farmstead distributions in the U.S. Corn Belt, taking the method to remoter rural settings than the urban focus of these other inquiries.

Elsewhere, Dacey (1960b) wrote more theoretically, deploying a sophisticated grasp of probability theory and mathematical reasoning, elaborating on concepts of randomness (Dacey 1965) and disclosing perhaps surprising ways in which randomness and regularity are inter-related (Dacey 1964a). Dacey and Tung (1962, pp. 86–7) claimed that regularity, “when the mean area assigned to each point is at a maximum,” arises “[i]n an unbounded space ... when each point is at the vertex of at least six equilateral triangles; this produces a hexagonal arrangement of points” (Dacey and Tung 1962, pp. 86–7). They thereby extended the Clark and Evans (1954) statement about hexagonality, but also, by implication, suggested that assumptions about complete regularity (including the interpretation of 2.1491) only apply to the hypothetical situation of “unbounded space” (irrelevant to actual situations studied by geographers: see below). Elsewhere again, Dacey (1964b, p. 559) worried that geographers were too concerned with clustering, showing “little or no interest in more regular than random point patterns,” despite the supposedly regular spacing of settlements and services predicted by “Central Place Theory” (also Dacey 1962, 1963). A follow-on might have been greater focus on the behavior of R_n when regularity is seemingly present, as in our article, but Dacey shifted back from NNA to quadrat approaches comparing observed point density variations across a plane with predictions of such variability using probability laws generative of regularity. He explicitly contrasted the latter with “the negative binomial, Neyman Type A, and similar contagious probability distributions” (Dacey 1964, p. 559) then gaining purchase within the hallways of geographical analysis.

In his wide-ranging *Locational Analysis in Human Geography*, Haggett (1965, pp. 90–1, pp. 231–2) referenced the Dacey and King studies—noting the supposed maximum of R_n of 2.15 arising for “a uniform triangular lattice”—and then offered a fuller account of NNA—here

depicting the supposed upper-limit as arriving “where the points are as far as possible from each other and therefore a regular hexagonal distribution.” The confusion of the triangular and the hexagonal recurs here, but what also becomes clear is the extent to which NNA, and more specifically R_n , were subsumed into a much larger portfolio of approaches to the “description of relative location” (the title of Haggett 1965, Chap. 8). This relatively minor place of NNA/ R_n at the research frontier was reinforced by the encyclopedic second edition of the same volume, wherein “distance-based methods,” including NNA and variants, comprised only seven pages (Section 13.6) near the end of a 35-page chapter on “Point Pattern Analysis” (Chap. 13) dominated by different forms of “quadrat counts” (Haggett, Cliff, and Ord 1977, esp. pp. 439–45). R_n was only cited once, “defined on the interval [0, 1, 2.15],” with 2.15 said to characterize “a maximally spaced, perfectly regular pattern when the points will be located at the vertices of a hexagonal grid” (Haggett, Cliff, and Ord 1977, p. 439). Similarly, Getis and Boots (1978; also Boots and Getis 1988), in their survey of the geographical “study of points, lines and area patterns,” only devoted a few pages to “distance measures” (Section 2.2.2) in a larger chapter on “Point Pattern: Poisson Process” (Chap. 2), outlining NNA but emphasizing here the calculation of Z-scores (see above). No mention was made of R_n or 2.15. In their statistical package for point pattern analysis, Chen and Getis (1998) followed an initial step (“basic descriptive statistics”) with a second step facilitating NNA—calculating mean observed and expected distances to first nearest neighbor, the variance and Z-scores—and then a third step dealing with frequencies of observed and expected distances (to first, second and larger-order neighbors), taking into account possible boundary effects and using Monte Carlo testing (also Ingram 1978). Again, no place for R_n .

Bringing the history forward, noting contributions to the current journal, *Geographical Analysis*, the trajectory is clear. Where the derivation is still recognizably from “original” NNA, referencing Clark and Evans (1954), the main impetus becomes recasting the procedures involved, with scanty or no reference to R_n . Two examples are: Jones (1971, p. 367), who explored the behavior of R_n across different orders of neighbors for different sorts of random and regular lattice, one finding being that the statistic appears to “converge asymptotically to the value of 2.1491 quoted by Clark and Evans ... for the hexagonal case”; and Thomas (1977, p. 415), who proposed a different nearest-neighbor statistic, based not on Poisson but rather the geometric distribution and its principle of “equal likelihood,” although he still took 2.1491 as the maximum when comparing the Clark and Evans statistic with his alternative for analyzing regular distributions based on the “equilateral triangle.” Where authors notably depart from NNA, the situation is quite different. A few, uncoincidentally from engineering backgrounds, have deployed substantially reframed species of nearest-neighbor measures to probe an expansive “geometry” of features—“point-like, line-like, surface-like”; maybe networked and amenable to representation and analysis as “graphs” or “sub-graphs” (Pace and Zou 2000)—before addressing the implications for calculating test statistics, estimating distances to different orders of nearest-neighbor “facilities” under differing conditions, and then practically planning for facility opening or closure (e.g., Okabe and Yoshikawa 1989; Okabe and Yamada 2001; Miyagawa 2009, 2014).

Many other contributions to *Geographical Analysis* have tackled the wider universe of point-pattern analyses identified by Haggett, Cliff, and Ord (1977), deploying, inspecting, and refining the family of spatial statistics specified by Chen and Getis (1978), while others have adopting exploratory and computational methods—perhaps allied to Monte Carlo simulations—rather than “the step-by-step mathematical rigour” (Haining and Hudson 2009, p. 346) preferred by the likes of Dacey when advancing beyond basic NNA. More generally, a word-search through potentially relevant papers in *Geographical Analysis*—ones reviewing the history or presenting

the state-of-the-art with respect to point pattern analysis and (in certain respects *the* major shift *on* from such analysis) “spatial autocorrelation”⁶—returns a complete absence of references to NNA, R_n and 2.15. To be clear, then, what we are charting here is a waning of interest by scholarly users if not entirely in NNA *per se*, since it retains a limited presence as one ingredient of point pattern analysis, but certainly in R_n and 2.15. “My view,” remarks a reviewer of this paper, “is that research moved beyond NNA when studying point patterns because it is an incomplete measure; relying on nearest-neighbour differences instead of all inter-event distances,” thereby only capturing *some* elements of a point pattern and arguably not its most important ones. Methods for characterizing and probing point patterns “simply became more advanced and moved on,” in which context “NNA did not provide a suitable framework and was left behind (the same way in which quadrat analysis was left behind).”⁷

We suspect that researchers such as Getis, Dacey, and others themselves quickly realized that 2.15 was *not* the maximum value possible for the statistic,⁸ but, despite our searching, we have not found any readily accessible published statements to this effect. Indeed, that absence is one justification for our present paper, and especially its analytical component as offered below. Moreover, if such research-based statements had become available from the 1960s or more recently, we presume that the pedagogical users, addressed in our next section, would not have been so *certain* in their own pronouncements about NNA, R_n and the maximal status of 2.15. A plausible inference is that, for mathematically proficient geographers operating at the frontiers of spatial science, NNA indeed came to seem too simplistic—especially in its Clark and Evans (1954) *first* nearest-neighbor guise—but *also* unworthy of sustained reassessment or commentary, even about potential misunderstandings of 2.15. The research frontier swept forward, but without leaving behind a published trace of what the scholars involved may already have divined about limitations of the nearest-neighbor statistic.

NNA and geographical education

Given the relatively easy-to-grasp quality of basic NNA, it is perhaps unsurprising that it emerged as—and then, remained—a staple of work in geographical education for both undergraduate and school-level students. A much-cited article in this regard was Pinder and Witherick (1972), published in the journal *Geography*, an organ of the Geographical Association (GA), the United Kingdom’s prime body promoting geographical education (particularly in schools). These authors gave a clear recounting of NNA in its Clark and Evans (1954) guise, arguing that prior limited uptake reflected its obfuscation by front-line researchers:

This relatively slow progress may ... be attributed to the fact that successive authors have not expressed the nearest-neighbour formula in its simplest form. When written out in full, the formula can seem daunting to those with relatively little skill in mathematics, a circumstance that may well have deterred many potential users of the technique. This is unfortunate because, once the formula is written in elemental terms, it can be applied by any student capable of using square root tables. ... [F]amiliarity with the approach should provide pupils intending to read geography at university with a logical, even elegant, view of location pattern assessment. (Pinder and Witherick 1972, p. 277)

Carefully explaining the “nearest-neighbour formulae” in its simplest formulation, with different notation but consistent with how we present it below (3), Pinder and Witherick (1972, p. 280) discussed “the interpretation of R_n ” by relating how “all values lie along a continuous

scale extending from 0 (completely clustered) through 1.0 (random) to 2.15 (perfectly regular).” Referencing their “nearest-neighbour (R_n) value scale” (Fig. 5), they identified “the ultimate situation ... in which the points are spaced with perfect regularity. When this occurs, the points are arranged according to a lattice of equilateral triangles with each point equidistant from six other points. With this arrangement, R_n reaches its maximum value of 2.15” (Pinder and Witherick 1972, p. 280).

They also demonstrated how R_n behaves when the same distribution (here, 10 regularly spaced points) is set within progressively smaller study areas (all square and hence proportionally equivalent). Fig. 2 shows their graphing of this behavior, with R_n rising toward 2.15. At this point the authors added, but with no justification, that “it would be inadvisable to employ [NNA] on strongly elongated areas, or indeed on any essentially linear patterns” (Pinder and Witherick 1972, p. 285). Such a claim does chime with aspects of our own analysis below, while the authors were prompted in their own subsequent active research to investigate further NNA and “linear patterns” (Pinder and Witherick 1973, 1975). Moreover, they also reasoned that a zone of R_n values could be identified, decreasing its range with increasing numbers of points, that *all* plausibly characterize randomness in a given distribution. Fig. 3 shows their graphing of such a zone, together with results (the $R_{n,s}$) from the authors’ own empirical NNAs of tors on Bodmin Moor, schools in Southampton and hatters in Luton (at different dates). Crucially, it is *this* diagram that has been frequently reproduced in subsequent educational treatments of NNA, presumably because it neatly captures the apparent “essentials” for mobilizing NNA and interpreting R_n .

An article broadly similar in tenor and message appeared a year later, concluding that “in general the method of Clark and Evans is better than methods using distance to higher nearest neighbours, both in distinguishing clearly non-random patterns from random ones and in providing the opportunity of interpreting the rejection of the hypothesis [of no pattern]” (De Vos 1973, p. 316). Oddly, given the appendix to Clark and Evans (1954), De Vos (1973, p. 307) remarked that “[i]t can be shown (although I have not found the proof anywhere ...) that $R_{[n]}$ is maximal [here given as 2.14914] if the objects form a hexagonal pattern.” Arguably, these two articles signposted the way for an uptake of NNA, with its iconic R_n range from 0 to 2.15,

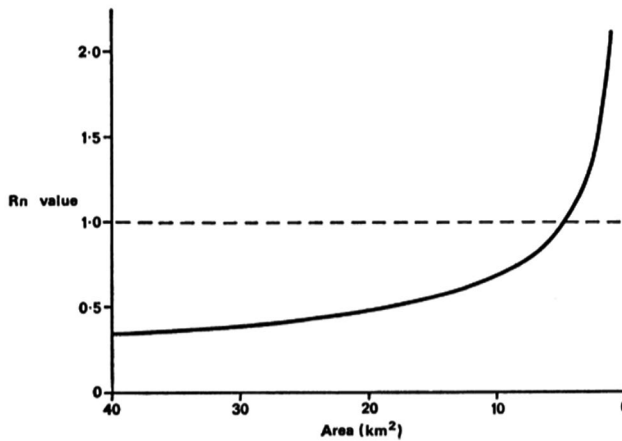


Figure 2. The behavior of R_n for the same distribution as affected by decreasing study area (Pinder and Witherick 1972, p. 284).

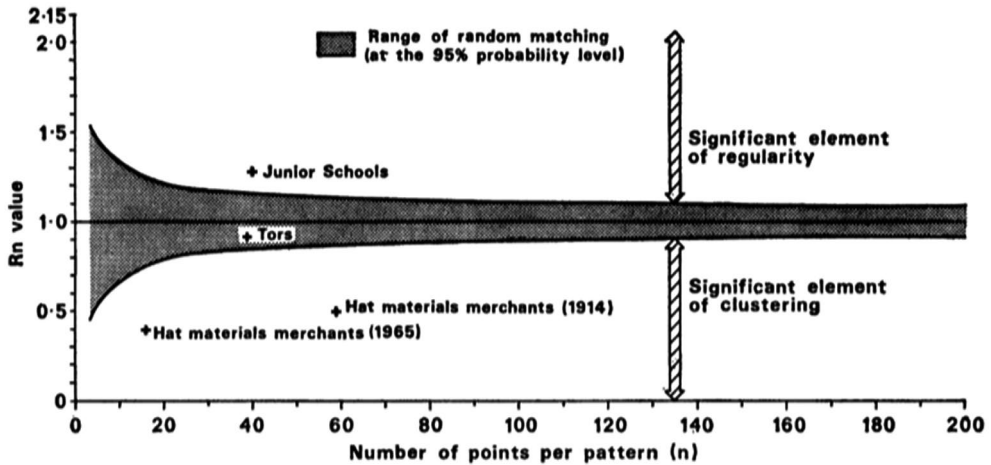


Figure 3. A graphical representation of how to interpret R_n as affected by increasing number of points (Pinder and Witherick 1972, p. 287).

in literature and resources for teaching quantification in geography to university and school-level students. Dawson (1975, p. 42) cited “[s]everal new publications on quantitative methods in geography [that] have appeared in the last year or two for senior-school pupils and undergraduates. They have all included [NNA] as a method for describing areal patterns.” Vincent (1976, p. 161) agreed that NNA and quadrat “techniques, purporting to measure pattern,” are now “firmly established in the undergraduate and school literature.” Selkirk and Neave (1984, p. 356) later confirmed that “[t]he technique is now sufficiently well-known for it to be included in a number of standard basic statistical textbooks for geographers and others.” Waugh’s popular textbook for A-level (or equivalent) students in the United Kingdom, *Geography: An Integrated Approach*, across numerous editions, started to reference NNA together with the statement “that a regular or perfectly uniform pattern of points would have a nearest neighbour statistic $R_n = 2.15$ which would mean that each point was equidistant from its neighbours” (Waugh 2009, p. 402). Additionally, it is easy to locate NNA instruction materials on current online platforms offering a basic introduction to NNA and positioning R_n and its alleged range as a simple tool for geographical analysis (in some cases obviously derivative from the Waugh textbook). In Table 1 we gather together examples of such claims from textbooks and online sources.

The early stirrings of NNA in a geographical-educational context prompted a flurry of contributions to the journal *Area* during the mid-1970s, commencing with the piece from Dawson (1975) quoted above which prompted four “comments” published together in a mini-theme section subtitled “caution towards nearest neighest neighbours” (in order, Charlton 1976; Ebdon 1976; Sibley 1976; Vincent 1976) and then, a more analytical article by Haworth and Vincent (1976). The horizon for this “exchange” was indeed the parachuting down of NNA into the classroom, with contributors displaying skepticism about the extent to which the method can really disclose or interpret distributional “patterns,” as well as about the problematic inferential leaps required (of the novice geographer) from pattern back to (generating) process. R_n being maximized at 2.15 was universally accepted, and Haworth and Vincent (1976, p. 300) sought, using a graphical proof based on Thieson polygons, to “present a method to demonstrate that $R_{[n]}$ takes on its maximum value (2.14914) if the objects to be measured are placed in a regular

Table 1. Table of textbooks/online sources introducing NNA to students, together with example statements about R_n and 2.15

Textbooks/online sources	Statements about R_n and 2.15
Theakstone, W.H. and Harrison, C. (1970), <i>The Analysis of Geographical Data</i>	“... values of R_n range from zero, when all points are clustered at the same location, to 2.15, when the points have their maximum spacing, in which case they are distributed regularly in a hexagonal pattern” (p. 61)
Ebdon, D. (1977), <i>Statistics in Geography: A Practical Approach</i>	“... the nearest-neighbour index can have a value between 0.0, indicating a completely clustered pattern, and 2.15, indicating a completely dispersed pattern” (p. 125)
Hammond, R. and McCullagh, P. (1978), <i>Quantitative Techniques in Geography</i>	“... the Nearest Neighbour Index ranges from zero (indicating that all points are closely clustered) to 2.15 (indicating that all points are uniformly distributed throughout the area)” (p. 270)
McGrew Jr., J.C. and Monroe, C.B. (1993), <i>An Introduction to Statistical Problem-Solving in Geography</i>	“... the Nearest Neighbour Index $R_{[n]}$ has its maximum value of 2.149 for a perfectly dispersed arrangement” (p. 217)
Rogerson, P.A. (2004), <i>Statistical Methods for Geographers</i>	“... the Nearest Neighbour Statistic $R_{[n]}$ varies from zero (a value obtained when all points are in one location) to a theoretical maximum of about 2.14 (for a perfectly uniform or systematic pattern of points spread out on an infinitely large two-dimensional plane.”
Barcelona Field Studies Centre, Nearest Neighbour Analysis, last accessed 18/08/2020 at: http://geographyfieldwork.com/nearest_neighbour_analysis.htm	<p>“The nearest neighbour formula will produce a result between 0 and 2.15, where the following distribution patterns form a continuum:”</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <p>0</p>  <p>Clustered</p> </div> <div style="text-align: center;"> <p>1.00</p>  <p>Random</p> </div> <div style="text-align: center;"> <p>2.15</p>  <p>Regular</p> </div> </div>
Teachitgeography Resources You Can Trust, The 2014 World Cup—A Nearest Neighbour Analysis, last accessed 18/08/2020 at: https://www.teachitgeography.co.uk/resources/ks5/statistical-techniques/statistics/the-2014-world-cup-a-nearest-neighbour-analysis/23078	“The Nearest Neighbour Analysis will always generate a result between 0 and 2.15. Values of 2.15 indicate a regular pattern of the distribution ...”
IB Geography, last accessed 18/08/2020 at: https://www.geoib.com/nearest-neighbor-index.html	“The NNI measures the spatial distribution from 0 (clustered pattern) to 1 (randomly dispersed pattern) to 2.15 (regularly dispersed /uniform pattern).”
Royal Geographical Society with the Institute of British Geographers, 4i—A Guide to Nearest Neighbour Analysis, last accessed 18/08/2020 at: https://www.rgs.org/CMSPages/GetFile.aspx?nodeguid=153c11d5-2420-4e25-a972-03c91a774292&lang=en-GB	<p>Using the following scale, the researcher can then identify the extent of clustering exhibited by the category in question.</p> <div style="text-align: center;">  </div>

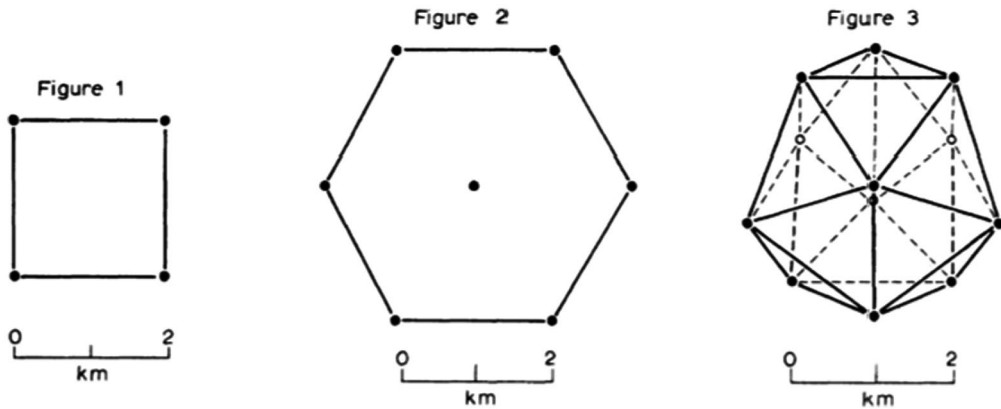


Figure 4. Ebdon's three figures for his mini-experiments with R_n (Ebdon 1976, p. 166).

close-packed hexagonal pattern," which, they elaborate, "is really a lattice of equilateral triangles and must not be confused with a hexagonal lattice where no points exist within each hexagon."

As a bridge to the analytical part of our article, attention can be drawn to the piece by Ebdon (1976, esp. p. 165), arguably the first geographer to worry in print about R_n exceeding 2.15, at the same time as reflecting on what happens when formulae specified on the basis of infinite points areas are applied "to relatively small numbers of points within finite areas." He conducted three mini-experiments with variants of a regular pattern (Fig. 4). The square figure gave $R_n = 4.0$, well above 2.15, leading him to exclaim that "this pattern is considerably more dispersed than it theoretically can be!" (Ebdon 1976, p. 166). The hexagonal figure gave $R_n = 3.2829$, also well above 2.15 (concurring with Philo 1997, p. 13).⁹ The "polyhedron" figure is intriguing, applying NNA to a three-dimensional surface—12 points located at the vertices of 20 equilateral triangles forming a polyhedron—in an attempt to avoid "the problem of dealing with a bounded area" (Ebdon 1976, p. 166). Here $R_n = 2.3534$, leading Ebdon (1976, p. 166) to conclude that "[e]ven with a larger number of points in an unbounded, though still finite area, the nearest-neighbour index exceeds its theoretical maximum value." The impression is that he regarded these findings as at yet unexplained curiosities, since in the remainder of his paper—running simulations of NNA on sets of four points randomly situated within a given area—he kept working with 2.15 as the "theoretical maximum" (and nothing of this complication appeared in his textbook a year later or in second edition: Ebdon 1977, pp. 124–6; Ebdon 1985, pp. 143–9). Intriguingly, explicitly citing Ebdon's curiosities, Pinder (1978, p. 383)¹⁰ admitted that "[s]mall point populations can give freak R_{\max} [maximum R_n] values," but then retreated to the orthodoxy: "it is accepted that a large population, fully dispersed with perfect regularity, will have R_{\max} equal to 2.14914." Nonetheless, there was still the hint here of gently challenging the hegemony of 2.15 and acknowledging the dangers of extrapolating from "theory" predicated on infinite spaces to the more finite world ultimately of most interest to geographers.

Analyzing NNA and querying 2.15

Let us now specify our preferred notation and forms for the relevant equations, differing a little from that found elsewhere in the literature. If n points are randomly distributed over an area A , then the random variable D_i represents the distance from point i to its nearest neighbor and

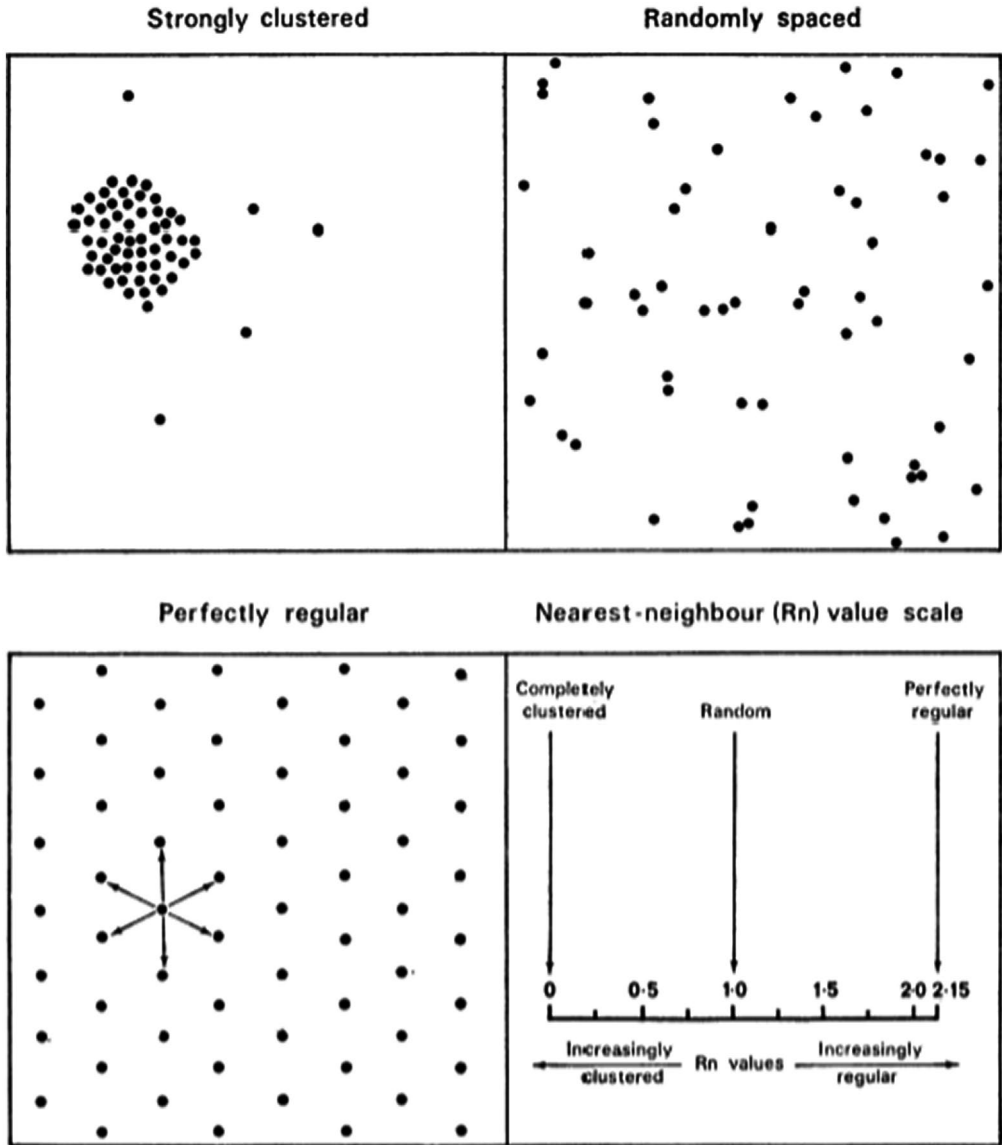


Figure 5. Example point patterns (we are most interested in ‘perfectly regular’) and diagram showing the supposed value range for R (R_n) (source: Pinder and Witherick 1972, p. 281).

\bar{D}_i , also a random variable, represents the mean of all distances from each point to its nearest neighbor. Thus, d_i and \bar{d}_i are observed values of these random variables. D_i and \bar{D}_i can be shown to have probability distributions, and in an unpublished working paper (Philo 2006) the second author here—offering a fuller proof than that contained in Clark and Evans (1954, Appendix)—has proved that:

$$E[\bar{D}_i] = 0.5\sqrt{\frac{A}{n}} \tag{1}$$

As indicated, the nearest-neighbor statistic, R_n , entails a ratio between the *observed* and *expected* mean distances, which can be expressed as:

$$R_n = \frac{\bar{d}_i}{E[\bar{D}_i]} \quad (2)$$

Substituting $E[\bar{D}_i]$ from (1) and simplifying gives the general formula for R_n :¹¹

$$R_n = 2\bar{d}_i \sqrt{\frac{n}{A}} \quad (3)$$

Behavior of R_n : small experiments

One path we followed was to interrogate the regular pattern pictured in the Pinder and Witherick (1972) article, thereby in effect using the 1972 paper “against” its core claim about 2.15 as an unbreachable upper limit for R_n . The relevant diagram is reproduced in Fig. 5, our specific focus being the pattern described as “perfectly regular,” which, crucially for our later analysis, entails an equilateral-triangular distribution with each point (other than those at the margins) equidistant to six nearest neighbors and with the distribution filling the space available. We printed out hard-copy of this diagram and measured it carefully, although the numerical values given below derive from our print copy and not the reproduction here.

The original is imprecisely drawn and not all of the triangle sides are exactly the same length, but we measured the majority at 0.55 cm and so took \bar{d}_i to be 0.55 (every d_i here should of course be 0.55 cm). For our print copy, therefore, $n = 60$ (15 rows \times 4 points), $A = 19.36 \text{ cm}^2$ (4.4 \times 4.4 cm), and $\bar{d}_i = 0.55 \text{ cm}$. If the diagram had been accurately drawn, the columns with eight points would have stretched from the top to the bottom (“horizontal”) edges of the containing area, since $8 \times 0.55 \text{ cm} = 4.4 \text{ cm}$, with the top and bottom points laying *on* these edges rather than just below or above them (as in Fig. 5). Slotting the stated values for n , A and \bar{d}_i into the standard NNA calculation generated $R_n = 1.94$, not reaching 2.15 but on the Clark and Evans (1954) logic still indicative of regularity. We then experimented with bringing in the “vertical” edges of the area so that the left and right columns of points also now laid on these edges, and we calculating using simple geometry that the width of the area would have to be 3.29 cm, rendering $A = (3.29 \times 4.4 \text{ cm}) = 14.48 \text{ cm}^2$.¹² Using this new area value, we found $R_n = 2.23$, higher than 2.15. Thus, if the area surrounding this point distribution is reduced to the smallest possible, with points on the edges and none excluded, then 2.15 is exceeded. The implication is that R_n will be maximized where the edges of the enclosing area coincide with the peripheral points of the “perfectly regular” distribution in question, and also with the (usually invisible) outer gridlines of the equilateral-triangular lattice on which the points sit.¹³ Indeed, in our proof below, predicated on areas delimited by lattice gridlines, we show that in such cases R_n will *necessarily* exceed 2.15.

Moreover, in this instance we held the height of the area constant and only reduced the width, which also suggests that R_n may be sensitive to the ratio between width and length or “height.” To explore this possibility, we created a narrower study area, isolating a vertical strip of width 1.2 cm flush with the right edge of the area that partitioned off the two right-hand columns of points, leaving room on either side. Obviously, \bar{d}_i remained the same. Now, with $n = 15$ (1 column of 7 points + 1 column of 8 points) and $A = 5.28 \text{ cm}^2$ (1.2 \times 4.4 cm), $R_n = 1.86$. Next, the vertical strip was reduced to width 0.8 cm, still including the same two columns of points, but

leaving a gap between the strip and the right edge of the area (not of any analytical consequence). Now, with $A = 3.52 \text{ cm}^2$ ($0.8 \times 4.4 \text{ cm}$), $R_n = 2.27$, breaking the 2.15 barrier. Finally, the strip was reduced to width 0.47 cm, the narrowest possible for this distribution, with points now occupying the edges of the area. Here, with $A = 2.07 \text{ cm}^2$ ($0.47 \times 4.4 \text{ cm}$), $R_n = 2.97$, soundly exceeding 2.15. The implication, echoing our first experiment, is that R_n increases as the area around the same set of n points is narrowed (reducing width) but not shortened (maintaining height). In other words, R_n increases as the proportions of the rectangular study area around these n points change or, more accurately, as the ratio between the sides of the study area—specifically, its width to its “height”—decreases from unity. More plainly still, the suggestion is that R_n increases as the area becomes “elongated,” to use the term from Pinder and Witherick (1972, p. 285; also Pinder 1978, p. 384), a realization that chimes with what drops from our proof below.

Behavior of R_n : a worked proof (across infinite and finite space)

Consider n points of a distribution forming a regular triangular lattice, with each equilateral triangle having side s , of course equivalent to \bar{d}_i , and with area (given by standard formula) $\frac{\sqrt{3}}{4}s^2$. As shown in Fig. 6, each point occupies a diamond-shaped area, darkly or lightly shaded, which combines two triangles with area $\frac{\sqrt{3}}{2}s^2$. If the shaded pattern in Fig. 6 is extended to infinity, then the point density— n/A —becomes $\frac{2}{\sqrt{3}s^2}$. Recalling (3) but adding an asterisk (R_n^*) to specify that it relates here to a regular triangular lattice, and replacing \bar{d}_i with s :

$$R_n^* = 2s\sqrt{\frac{n}{A}} \tag{4}$$

Substituting the point density term into (4) gives:

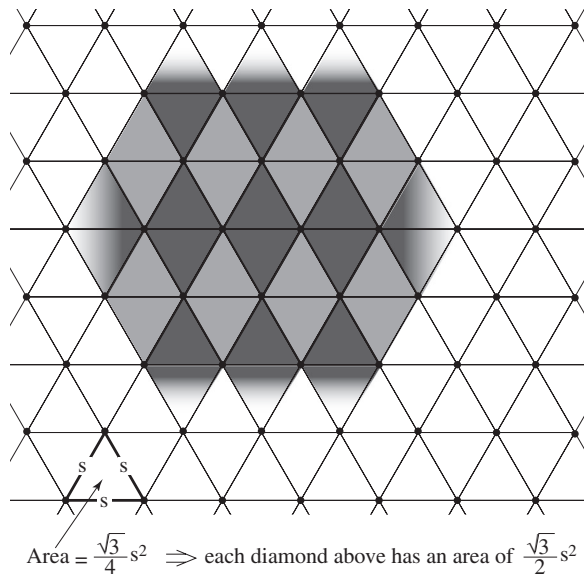


Figure 6. Regular triangular lattice with equilateral triangles and diamond-shaped areas (and note that hexagons are also easy to discern).

$$R_{\infty}^* = 2s \sqrt{\frac{2}{\sqrt{3} \cdot s^2}} = 2 \sqrt{\frac{2}{\sqrt{3}}} = 2.1491 \tag{5}$$

It is arguably for this reason that 2.15 is often quoted as the maximum value for R_n , but it does assume an *infinite* number of points distributed in a regular triangular lattice spread over an *infinite* area. Perhaps intuitively, researchers have assumed that under all conditions R_n necessarily must approach 2.15 “from below,” rendering 2.15 this assumed maximum, but the further question, as also explicitly asked by Ebdon (1976) and Pinder (1978), is whether this statement holds true for situations of *finite* n and A .

Consider n points of a distribution forming a regular triangular lattice, but now arranged in an approximately rectangular area of width w and height h , such as shown in Fig. 7, with triangular sides s , and with r being the ratio between w and h (and hence $w = rh$). There are b short lines (here $b = 3$) and $b + 1$ long lines ($b + 1 = 4$), k point in each long row ($k = 5$), and $k - 1$ points in each short row ($k - 1 = 4$), and therefore:

$$n = b(k - 1) + k(b + 1) = 2bk + k - b \text{ (here 32)} \tag{6}$$

The height of each triangle and the area of each triangle are, respectively, $\frac{\sqrt{3}}{2}s^2$ and $\frac{\sqrt{3}}{4}s^2$. The number of triangles is $(2k - 3) 2b$, and hence area A is:

$$A = (2k - 3) \frac{\sqrt{3}}{2} bs^2 \tag{7}$$

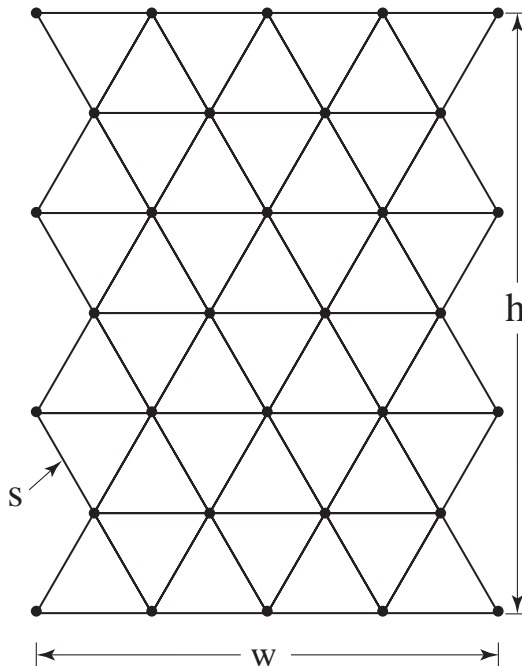


Figure 7. Example of the hypothetical distribution used in the following analysis. [Colour figure can be viewed at wileyonlinelibrary.com]

From Fig. 7, $w = (k - 1)s$ and $h = \sqrt{3}bs$, and, as $w = rh$, it can be shown that:

$$k = \sqrt{3}br + 1 \tag{8}$$

Substituting k from (8) in (6) gives:

$$n = 2\sqrt{3}rb^2 + (\sqrt{3}r + 1)b + 1 \quad \text{or} \quad 2\sqrt{3}rb^2 + (\sqrt{3}r + 1)b + (1 - n) = 0 \tag{9}$$

Therefore, letting $Q = 3r^2 + 2\sqrt{3}r(4n - 3) + 1$ in (10):

$$b = \frac{- (\sqrt{3}r + 1) \sqrt{(\sqrt{3}r + 1)^2 - 8\sqrt{3}r(1 - n)}}{4\sqrt{3}r}$$

$$b = \frac{\sqrt{3r^2 + 2\sqrt{3}r(4n - 3) + 1} - (\sqrt{3}r + 1)}{4\sqrt{3}r}$$

$$b = \frac{\sqrt{Q} - (\sqrt{3}r + 1)}{4\sqrt{3}r} \tag{10}$$

Substituting k from (8) in (7) gives:

$$A = \frac{(2\sqrt{3}b^2r - b) \sqrt{3}s^2}{2} \tag{11}$$

which becomes:

$$A = f(r, n) \times s^2 \tag{12}$$

See Appendix for steps from (11) to (12) and definition of the function $f(r, n)$. Thus, from (4):

$$R_n^* = 2s \sqrt{\frac{n}{A}} = 2s \sqrt{\frac{n}{f(r, n)}} = 2 \sqrt{\frac{n}{f(r, n)}} \tag{13}$$

Crucially, this result shows that the value of R_n^* is independent of A , depending solely on r and n . Since r is the ratio of width and height, this finding establishes why R_n^* was affected in our experiment above when reducing the width while retaining the height of the area around the same set of points (with n remaining constant). It also implies grounds for the Pinder and Witherick (1972) caution about the possible influence of “elongation” on the performance of R_n .

Furthermore, replacing $f(r, n)$ in (13) with its definition from our Appendix:

Table 2. Table showing how R_n^* varies with r (ratio of width to height) and n (number of points): note that a *lot* of points are required before R_n^* begins to approach 2.15

		r			
R_n^*		0.2	0.5	0.8	1.0
n	50	2.712	2.562	2.529	2.521
	150	2.437	2.365	2.350	2.346
	300	2.344	2.297	2.286	2.284
	500	2.296	2.262	2.254	2.252
	1,000	2.251	2.227	2.222	2.221
	4,000	2.199	2.188	2.185	2.185

$$R_n^* = 2 \sqrt{\frac{8rn}{3r^2 + 2 + \sqrt{3}r(4n-1) - (\sqrt{3}r+2) \sqrt{3r^2 + 2\sqrt{3}r(4n-3) + 1}}} \tag{14}$$

$$R_n^* = 2 \sqrt{\frac{8r}{\frac{3r^2}{n} + \frac{2}{n} + \sqrt{3}r\left(4n - \frac{1}{n}\right) - (\sqrt{3}+2) \sqrt{\frac{3r^2}{n} + 2\sqrt{3}r\left(\frac{4n}{n} - \frac{3}{n^2}\right) + \frac{1}{n^2}}}} \tag{14}$$

Thus, for any value of r :

$$\lim_{n \rightarrow \infty} R_n^* = 2 \sqrt{\frac{8r}{\sqrt{3} \cdot r \cdot 4}} = 2 \sqrt{\frac{2}{\sqrt{3}}} = 2.149 = R_\infty^* \tag{15}$$

While Clark and Evans (1954) arrived at this value, 2.149, by considering an infinite number of points over an infinite area, result (15) confirms that this value can be approached in a finite area. Table 2 shows how the value of R_n^* —remember that this is only R_n for the specific but crucial case of a regular hexagonal pattern (equilateral-triangular lattice), not R_n for any possible distribution—varies with respect to r and n , demonstrating two things: first, perhaps counter-intuitively, that R_n^* approaches 2.15 “from above,” not from below, descending toward 2.15 as r and n increase; and second, that R_n^* for a constant n increases as r decreases (as the ratio of width to height decreases), as can be seen by tracking values of R_n^* from right to left across any row in the table.¹⁴

Conclusion

This paper is a contribution to work reconstructing, contextualizing, and sympathetically critiquing the history of academic geography under the press of “the quantitative revolution,” “locational analysis,” or “spatial science,” notably as pursued in the compelling writings of Barnes (e.g., 2001a, 2001b, 2003, 2004, 2013). Perhaps surprisingly, there are few engagements with the histories of specific methods, quantitative or qualitative, although indications are contained in textbooks such as Cloke et al. (2004, esp. Chap. 1 and opening pages of most chapters). Our inquiry here offers such an engagement, examining the history of NNA in academic geography,

noting a bifurcation in the research-frontier and more educational uses of NNA whereby the latter have remained much more energized by the simple, singular nearest-neighbor statistic R_n and its supposed range from 0 to 2.15. Such an obsession by geographers with a particular measure and number, perhaps one acquiring almost a “magical” status in certain fields, has been noticed before, as when Whatmore and Landström (2012) explored the background to Manning’s N , a “roughness coefficient,” beloved by fluvial geomorphologists.

Our inquiry has also added to the historical survey by undertaking its own analytical work, including a small empirical experiment and a mathematical proof where the behavior of R_n was investigated for points on a regular, equilateral-triangular lattice in situations of both infinite (bounded) and finite (bounded) space. The chief purpose was to demonstrate conclusively—not as anomaly or “freak” result contradicting the “theory” (cf. Ebdon 1976; Pinder 1978)—that in these situations R_n can, and indeed necessarily *will*, exceed the envisaged maximum of 2.15. Countless statements ever since Clark and Evans (1954) about R_n having an upper threshold of 2.15 are hence shown to be mistaken, a provocative finding to hold up to the history that we have narrated. Additionally, our analysis confirms the Pinder and Witherick (1972) suspicion about “elongated” study areas affecting the performance of R_n .

Finally, we speculate that our conjoint historical-analytical inquiry potentially raises bigger questions about how “the quantitative revolution” has diffused into academic geography. Amazing energy and enthusiasm accompanied the initial stirrings of this new departure for the discipline, almost inevitably leading to an extremely steep “learning curve” as the scholars involved familiarized themselves with quantitative techniques imported from elsewhere, such as NNA deriving from ecology, *and* then developed their own distinctive takes on these techniques. Given a rapid deepening of engagement with the mathematical intricacies of an explicitly *spatial* science, it is perhaps unsurprising that scholars—notably someone like Dacey—quickly shifted far beyond the likes of basic NNA, maybe being aware of their limitations but not necessarily seeing any need for (or having any interest in) spelling out these limitations chapter-and-verse. Conversely, those simpler techniques deemed as student-friendly—amenable to students doing calculations and interpreting statistical outputs—arguably began appealing to geographical educators, thereby sliding into the heart of quantitative geography pedagogy at school and undergraduate levels. That NNA persisted for so long in the textbooks, as one reviewer comments, “is itself perhaps a sign of this disconnection” between geographical educators and the research frontier, the mathematical difficulty of more sophisticated spatial point pattern analyses, as elaborated by specialist statisticians, being judged far beyond what could be handled in the geography classroom. Another reviewer wonders if the “simple truths” of such techniques, such as the basic NNA statistical specification of 2.15, might acquire a peculiarly protected, uncritically appraised, status. Thus, “in order for students to learn mathematical precepts in a discipline like geography, quick and dirty nostrums need be served up,” even if educators themselves become aware of underlying limitations, maybe errors, accompanying these precepts. Our surmise in this case is merely that a disjuncture arose between scholarly and pedagogic users of NNA and its statistic, partly due to the very different needs and ambitions that each constituency held (and still holds) with respect to spatial-analytical methods.¹⁵ We can offer no obvious solution to this bifurcated situation, except to hope that our contribution can serve to engage, enlighten and possibly also entertain those who are still actively deploying NNA, R_n , and 2.15, in whatever capacity.

Acknowledgements

We thank Mike Shand for drawing Figs. 6 and 7; we thank Gesa Helms and Julia Laphorn for assistance with reading, respectively, Hertz (1909) and the Einstein-Hertz correspondence (Note 1). Both authors acknowledge the love and encouragement of Pamela Philo (1931-2017). We thank three reviewers and the editor for their extremely helpful criticisms and generous comments.

Notes

- 1 The helpful terms “scholarly users” and “pedagogic users” are suggested by a reviewer.
- 2 A curiosity is that Clark and Evans (1954, p. 451) note an article by Hertz (1909), “virtually unknown to biologists” and suffering from “inaccessibility,” presumably due to being written in German but maybe also the recency of WWII. This article considers the distribution of particles in a gas, hence conducting a three-dimensional analysis: it never speaks of a nearest-neighbor statistic or indeed of 2.15. Hertz was a philosopher and mathematician, specializing in optics and electromagnetic theory, working at the Mathematical Institute in Göttingen, Germany. His dissertation adviser was Hilbert, a leading mathematician of the early-1900s noted for his “axiomatising” of physics. Einstein visited Göttingen, and the intellectual rigor and “oral culture” of the Institute supposedly influenced the early development of his general theory of relativity. Hertz, with his interest in applying the calculus of probabilities to the motions of molecules through space, may have been part of that influence, and the pair corresponded. Hertz, with a Jewish background, was badly affected by the rise of Nazism, being dismissed from his post in 1933 but escaping to Prague, Paris, and then London with the assistance of the Social Democratic Party of Germany (SPD). For reasons not entirely clear, in one letter (1915) Einstein accused Hertz of “political cowardice” and a lack of “civil courage.” Here, then, NNA holds tenuous connections to arguably the most significant scientific and geopolitical events of the last century. Our sources here include: Howard and Norton (1993), MacLane (1995), Rowe (2004), also Einstein (1998, Letters #108, 111, 125–8).
- 3 One reviewer helpfully remarks that “the lower limit (0) is an uninteresting case both theoretically and practically. In point process theories points are not allowed to occupy the exact same location; in practice, why would anyone deploy a statistic to confirm the non-variation?”
- 4 *Different* regular lattices, made up of equilateral triangles, squares, hexagons, and more, all generate *different* possible values and ranges for R_n . As is clear from Fig. 6, a triangular lattice can easily have a hexagonal pattern detected in or superimposed upon it. Some geographers have nonetheless confused themselves by thinking in terms of a *genuinely* “hexagonal lattice ... as, for example, in the erroneous diagram presented by Theakstone and Harrison (1970, p. 61)” (Haworth and Vincent 1976, p. 300). Indeed, “Theakstone and Harrison present ‘a hexagonal pattern’ ... in which each point is equidistant from only three other points” (Pinder and Witherick 1972, p. 288). It is tempting to suggest that some geographers, seduced by Central Place Theory, have ended up seeing hexagons everywhere, whereas in NNA the key distribution, the one maximizing mean nearest-neighbor distances, is actually equilateral-triangular.
- 5 One reviewer distinguishes between “window sampling” and “the small world model.” In the former, the researcher samples R_n values for different parts of a point distribution encased by different “sample windows,” potentially of different sizes, adding that “as the window size increases, the area increases, but also reveals more points in the process.” There would hence be no necessary increase or decrease in R_n . In the latter, the “small world” comprises a given set of points beyond which there are *no* more points, and here increasing the study area inevitably correlates with decreases in R_n .
- 6 Cliff and Ord (2009, p. 354), reflecting on their turn to “spatial autocorrelation,” cited the claim from Ripley (1990, pp. 55–6) that “point processes *per se* are becoming less important.” They clarified that, whereas “research about spatial processes before 1969 had dealt almost exclusively with either spatial point processes or with spatial processes ... recorded at point locations,” Dacey’s innovation had been to recenter “the focus squarely on data aggregated over a region.” One reviewer caveats that “K-functions are still being actively used when the focus is on whether the underlying process is clustered or dispersed.”

- 7 This reviewer adds that “[t]he methodological development applicable to spatial point pattern analysis has largely moved outside of geography (and geographical analysis) and back to statistics” (e.g., Møller and Waggepetersen 2004).
- 8 An initial reviewer claimed that “[t]he fact that 2.15 is not the true upper limit for the statistic is well known,” but provided no supporting references.
- 9 If connecting lines are drawn from the boundary points to the one at the center of Ebdon’s “hexagon,” it is clear that the case here could also be described as six equilateral triangles, reflecting that conflation of hexagons and triangles discussed in Note 4. Ebdon’s R_n value (3.28) for this case is in the same ballpark as the R_n value (3.30) stated for six equilateral triangles in Note 14, the minor difference being due to rounding “errors” in the differing calculations undertaken by us and by Ebdon.
- 10 The papers by Ebdon (1976) and Pinder (1978) prompted follow-ups (in order, Boots 1979; Pinder 1979; Ebdon 1980). Ebdon (1980, p. 154) dubbed them “the slow-moving saga of nearest-neighbour analysis in the pages of *Area*.” Perhaps with more prescience than intended, the last paper here was followed by an italicized statement: “*This discussion is now closed – editor.*”
- 11 It would obviously be easy to plug conjectured combinations of values for n , A and \bar{d}_i into equation (3) that would produce R_n s above 2.15, but there is no guarantee that these combinations can ever arise in “real” distributions.
- 12 Inspecting the Pinder and Witherick (1972) diagram, it is clear that the width as pictured is equivalent to the “heights” of seven equilateral triangles (on the invisible lattice underlying the points) plus a little extra on either side. The height of a single triangle of base 0.55 cm can readily be calculated as 0.47 cm, and hence the minimum width of the area inclusive of all points and removing the surplus on either side = $(7 \times 0.47 =) 3.29$ cm.
- 13 Actually, there can never be *complete* overlap of a straight-line enclosing area and the lattice gridlines for a regular, equilateral-triangular lattice, since one pair of peripheral gridlines must always “zig-zag,” as is obvious from Fig. 7. This complication requires careful handling in our proof.
- 14 It might be noticed that the R_n value, 2.23, generated for the minimum-sized area enclosing all the 60 points in the Pinder and Witherick (1972) diagram, is only approached in Table 2 when n starts to exceed 1000. This difference likely arises because the *rectangular* area in our experiment is inevitably slightly larger than the area enclosed by an equilateral-*triangular* lattice based on 60 points, since the former will include the sub-areas *in* the “zigzags” (equivalent to the unenclosed half-diamonds on either side of Fig.) which are excluded from the latter. By the logic of how R_n is calculated, a larger A will lead to a smaller multiplier under the square root sign, and hence a smaller value for R_n . For the record, for the “toy” example of a single equilateral triangle with sides 1 unit long ($n = 3$; $A = 0.43$ cm²; $\bar{d}_i = 1.0$ cm), $R_n = 5.28$; for that of 2 such triangles ($n = 4$; $A = 0.86$), $R_n = 4.32$; for 3 ($n = 5$, $A = 1.29$), $R_n = 3.94$; for 4 ($n = 6$; $A = 1.72$), $R_n = 3.74$; for 6 ($n = 7$; $A = 2.88$), $R_n = 3.30$; for 9 ($n = 10$; $A = 3.87$), $R_n = 3.22$; for 16 ($n = 15$; $A = 6.88$), $R_n = 2.96$. Once over 16, R_n slips under 3.0, but, as is apparent from Table 2, will then only *very* gradually dip down toward 2.15 as n increases.
- 15 There has arguably been a closely related “theoretical”/“practical” divide in play here, suggested by how Pinder (1979, p. 210) responded to Boots (1979): “He relies on the complex theoretical argument advanced by Dacey with supporting evidence provided by reference to very inaccessible work undertaken by Matern and Persson. My inclination is to take the results of Ebdon’s practical investigation”

APPENDIX

$$b = \frac{\sqrt{Q} - (\sqrt{3} \cdot r + 1)}{4\sqrt{3}r} \tag{10}$$

$$A = \frac{(2\sqrt{3}b^2r - b) \sqrt{3} \cdot s^2}{2} \tag{11}$$

Substituting b from (10) in (11) gives:

$$\begin{aligned}
 A &= \frac{\sqrt{3}}{2} \left[2\sqrt{3r} \left\{ \frac{\sqrt{Q} - (\sqrt{3} \cdot r + 1)}{4\sqrt{3r}} \right\} - \frac{\sqrt{Q} - (\sqrt{3} \cdot r + 1)}{4\sqrt{3r}} \right] s^2 \\
 A &= \frac{\sqrt{3}}{2} \left[\frac{Q - 2(\sqrt{3r+1}) \cdot \sqrt{Q} + (\sqrt{3r+1})^2}{8\sqrt{3r}} - \frac{\sqrt{Q} - (\sqrt{3r+1})}{4\sqrt{3r}} \right] s^2 \\
 A &= \frac{1}{16r} \left[Q - 2(\sqrt{3r+1}) \sqrt{Q} + \sqrt{3r^2 + 2\sqrt{3r} + 1} - 2\sqrt{Q} + 2(\sqrt{3r+1}) \right] s^2 \\
 A &= \frac{1}{16r} \left[Q - 2(\sqrt{3r+2}) \sqrt{Q} + 3r^2 + 4\sqrt{3r} + 3 \right] s^2 \\
 A &= \frac{1}{16r} \left[6r^2 + 4 + 2\sqrt{3r}(4n-1) - 2(\sqrt{3r+2}) \sqrt{Q} \right] s^2 \\
 A &= \frac{1}{8r} \left[3r^2 + 2 + \sqrt{3} \cdot r(4n-1) - (\sqrt{3} \cdot r + 2) \sqrt{3r^2 + 2\sqrt{3r}(4n-3) + 1} \right] s^2
 \end{aligned}$$

Therefore:

$$A = f(r, n) \times s^2 \quad (12)$$

Which defines the function $f(r, n)$ as:

$$\frac{1}{8r} \left[3r^2 + 2 + \sqrt{3r}(4n-1) - (\sqrt{3r} + 2) \sqrt{3r^2 + 2\sqrt{3r}(4n-3) + 1} \right]$$

References

- Barnes, T. J. (2001a). "Lives Lived, and Lives Told: Biographies of Geography's Quantitative Revolution." *Environment and Planning D: Society and Space* 19, 409–29.
- Barnes, T. J. (2001b). "Rethorising Economic Geography: From the Quantitative Revolution to the 'Cultural Turn'." *Annals of the Association of American Geographers* 91, 546–65.
- Barnes, T. J. (2003). "The Place of Locational Analysis: A Selective and Interpretive History." *Progress in Human Geography* 27, 69–95.
- Barnes, T. J. (2004). "Placing Ideas: *Genius Loci*, Heterotopia and Geography's 'Quantitative Revolution'." *Progress in Human Geography* 29, 565–95.
- Barnes, T. J. (2013). "Folder 5, Box 92." *Social and Cultural Geography* 14, 784–91.
- Birch, B. P. (1967). "The Measurement of Dispersed Patterns of Settlement." *Tijdschrift voor Economische en Sociale Geografie*, 58, 68–75.
- Boots, B. N. (1979). "Underestimation in Nearest-Neighbour Analysis." *Area* 11, 208–10.
- Boots, B. N., and A. Getis. (1988). *Point Pattern Analysis*. Newbury Parks, CA: Sage.
- Charlon, B. (1976). "On Botanists' Boundaries and Buttercups." *Area* 8, 170–1.
- Chen, D.-M., and A. Getis. (1998). *Point Pattern Analysis*. Available online at: ceadsev1.nku.edu/longa/geomed/ppa/doc/html/ppa.html [last accessed 29/08/2007].
- Clark, P. J. (1956). "Grouping in Spatial Distributions." *Science* 1956, 373–4.
- Clark, P. J., and F. C. Evans. (1954). "Distance to Nearest Neighbour as a Measure of Spatial Relationships in Populations." *Ecology* 35, 445–53.

- Clark, P. J., and F. C. Evans. (1955). "On Some Aspects of Spatial Patterns in Biological Populations." *Science* 1955, 397–8.
- Clark, P. J., and F. C. Evans. (1979). "Generalisation of a Nearest-Neighbour Measure of Dispersion for Use in K Dimensions." *Ecology* 60, 316–7.
- Cliff, A. D., and J. K. Ord. (2009). "What Were We Thinking?" *Geographical Analysis* 41, 351–63.
- Cloke, P., I. Cook, P. Crang, M. Goodwin, J. Painter, and C. Philo. (2004). *Practising Human Geography: The Construction and Interpretation of Geographical Knowledge*. London: Sage.
- Dacey, M. F. (1960). "A Note on the Derivation of Nearest-Neighbour Distance." *Journal of Regional Science* 2, 81–7.
- Dacey, M. F. (1960). "The Spacing of River Towns." *Annals of the Association of American Geographers* 50, 59–61.
- Dacey, M. F. (1962). "Analysis of Central Place and Point Patterns by a Nearest-Neighbour Method." *Lund Studies in Geography, Series B, Human Geography* 24, 55–77.
- Dacey, M. F. (1963). "Order-Neighbour Statistic for a Class of Random Patterns in Multi-Interval Space." *Annals of the Association of American Geographers* 53, 505–15.
- Dacey, M. F. (1964a). "Two-Dimensional Point Patterns: A Review and an Interpretation." *Papers in Regional Science* 13, 41–55.
- Dacey, M. F. (1964b). "Modified Poisson Probability Law for Point Pattern More Regular than Random." *Annals of the Association of American Geographers* 54, 555–65.
- Dacey, M. F. (1965). "Order Distance in an Inhomogeneous Point Pattern." *Canadian Geographer* 9, 144–53.
- Dacey, M. F., and T.-H. Tung. (1962). "The Identification of Randomness in Point Patterns." *Journal of Regional Science* 4, 83–96.
- Dawson, A. H. (1975). "Are Geographers Indulging in a Landscape Lottery?" *Area* 7, 42–5.
- De Vos, S. (1973). "The Use of Nearest-Neighbour Methods." *Tijdschrift voor Economische en Sociale Geografie* 64, 307–19.
- Ebdon, D. (1976). "On the Underestimation Inherent in Commonly-Used Formulae." *Area* 8, 165–9.
- Ebdon, D. (1977). *Statistics in Geography: A Practical Approach*. London: Wiley-Blackwell.
- Ebdon, D. (1980). "Nearest-Neighbour Analysis in the Square." *Area* 12, 154–9.
- Ebdon, D. (1985). *Statistics in Geography*, 2nd ed. London: Basil Blackwell.
- Einstein, A. (1998). Vol.8: *The Berlin Years: Correspondence, 1914–1918* (edited by R. Schulmann, A. J. Kox, M. Janssen, and J. Illy). Princeton, NJ: Princeton University Press.
- Getis, A. (1964). "Temporal Land-Use Pattern Analysis with the Use of Nearest-Neighbour and Quadrat Methods." *Annals of the Association of American Geographers* 54, 391–9.
- Getis, A., and B. N. Boots. (1978). *Spatial Processes: An Approach to the Study of Point, Line and Area Patterns*. Cambridge, UK: Cambridge University Press.
- Haggett, P. (1965). *Locational Analysis in Human Geography*, 1st ed. London: Edward Arnold.
- Haggett, P., A. D. Cliff, and K. Ord. (1977). *Locational Analysis in Human Geography*, 2nd ed. London: Edward Arnold.
- Haining, R. P., and J. C. Hudson. (2009). "Encomium: Michael Francis Dacey: A Tribute." *Geographical Analysis* 41, 346–50.
- Hammond, R., and P. McCullagh. (1978). *Quantitative Techniques in Geography*. Oxford: Oxford University Press.
- Haworth, J., and P. Vincent. (1976). "Maximising the Nearest-Neighbour Statistic." *Area* 8, 299–302.
- Hertz, P. (1909). "Über den geigeneitigen durchschnittlichen Abstand von Punkten die mit bekannter mittlerer Dichte im Raume angerordnet sind." *Mahematische Annalen* 67, 387–98.
- Howard, D., and J. D. Norton. (1993). "Out of the Labyrinth? Einstein, Hertz and the Göttingen Answer to the Hole Argument." In *The Attraction of Gravitation: New Studies in the General History of Relativity*, 30–6. edited by J. Earman, M. Janssen and J. D. Norton. Boston, MA: Birkhäuser.
- Jensen-Butler, C. (1972). "Nearest-Neighbour Analysis of a Central Place System." *Tijdschrift voor Economische en Sociale Geografie* 63, 353–59.
- Jones, A. (1971). "An Order Neighbor Approach to Random Disturbances on Regular Point Lattices." *Geographical Analysis* 3, 361–78.

- King, L. J. (1962). "A Quantitative Expression of the Pattern of Urban Settlements in Selected Areas of the United States." *Tijdschrift voor Economische en Sociale Geografie* 53, 1–7.
- MacLane, S. (1995). "Mathematics at Göttingen Under the Nazis." *Notices of the American Mathematical Society* 42, 1134–8.
- McGrew Jr., J. C., and C. B. Monroe. (1993). *An Introduction to Statistical Problem-Solving in Geography*. Long Grove, IL: Waveland Press.
- Miyagawa, M. (2009). "Order Distance in Regular Point Patterns." *Geographical Analysis* 41, 252–62.
- Miyagawa, M. (2014). "Distribution of the Sum of Distances to the First and Second Nearest Facilities." *Geographical Analysis* 46, 321–33.
- Møller, J., and R. P. Waagepetersen. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton, FL: Chapman and Hall/CRC.
- Moore, P. G. (1954). "Spacing in Plant Populations." *Ecology* 35, 222–7.
- Okabe, A., and I. Yamada. (2001). "The K-Function Method on a Network and its Computational Implementation." *Geographical Analysis* 33, 271–90.
- Okabe, A., and T. Yoshikawa. (1989). "The Multi Nearest Neighbor Distance Method for Analyzing the Compound Effect of Infrastructural Elements on the Distribution of Activity Points." *Geographical Analysis* 21, 216–35.
- Pace, R. K., and D. Zou. (2000). "Closed-Form Maximum Likelihood Estimates of Nearest-Neighbour Spatial Dependencies." *Geographical Analysis* 32, 154–72.
- Philo, C. (2004). *A Geographical History of Institutional Provision for the Insane from Medieval Times to the 1860s in England and Wales: 'The Space Reserved for Insanity'*. Lewiston: Edwin Mellen Press.
- Philo, P. W. (1969). "An Algorithm to Evaluate the Logarithm of a Number to Base 2, in Binary Form." *Radio and Electrical Engineer* 36 (1), 49–50.
- Philo, P. W. (1997). *Probability Distributions of Distances to Nearest-Neighbours*, Unpublished Typescript (Available from Authors).
- Philo, P. W. (2006). *Two Point 1.5 or Not Two Point 1.5*, Unpublished Typescript (Available from Authors).
- Pinder, D. A., and M. E. Witherick. (1972). "The Principles, Practice and Pitfalls of Nearest-Neighbour Analysis." *Geography*, 57, 277–88.
- Pinder, D. A., and M. E. Witherick. (1973). "Nearest-Neighbour Analysis of Linear Point Patterns." *Tijdschrift voor Economische en Sociale Geografie*, 64, 160–3.
- Pinder, D. A. (1978). "Correcting Underestimation in Nearest-Neighbour Analysis." *Area* 10, 379–85.
- Pinder, D. A. (1979). "Reply to Boots." *Area* 11, 210–1.
- Pinder, D. A., and M. E. Witherick. (1975). "A Modification of Nearest-Neighbour Analysis for Use in Linear Situations." *Geography* 60, 16–23.
- Porter, P. W. (1960). "Earnest and the Orophagions: A Fable for the Instruction of Young Geographers." *Annals of the Association of American Geographers* 50, 297–9.
- Ripley, B. (1990). "Discussion: 'Statistical Methods for Point Pattern Data', by J. Keith Ord." In *Spatial Statistics: Past, Present and Future*, 55–7, edited by D. Griffith. Ann Arbor, MI: Institute of Mathematical Geography.
- Rogerson, P. A. (2004). *Statistical Methods for Geographers*. London: Sage.
- Rowe, D. E. (2004). "Making Mathematics in an Oral Culture: Göttingen in the Era of Klein and Hilbert." *Science in Context* 17, 85–129.
- Selkirk, K. E., and H. R. Neave. (1984). "Nearest-Neighbour Analysis of One-Dimensional Distribution of Points." *Tijdschrift voor Economische en Sociale Geografie* 75, 356–61.
- Sherwood, K. B. (1970). "Some Applications of the Nearest-Neighbour Technique to the Study of the Movement of Intra-Urban Functions." *Tijdschrift voor Economische en Sociale Geografie* 61, 41–8.
- Sibley, D. (1976). "On Pattern and Dispersion." *Area* 8, 163–5.
- Theakstone, W., and C. Harrison. (1970). *The Analysis of Geographical Data*. London: Heinemann Educational.
- Thomas, R. W. (1977). "Nearest-Neighbor Distances in an Infinite Entropic Point Pattern." *Geographical Analysis* 9, 409–18.
- Thompson, H. R. (1956). "Distribution of Distance to Nth Nearest Neighbour in a Population of Randomly Distributed Individuals." *Ecology* 37, 391–4.
- Vincent, P. (1976). "The General Case: How Not to Measure Spatial Patterns." *Area* 8, 161–3.

Geographical Analysis

- Whatmore, S. J., and C. Landström. (2012). "Manning's N —Putting Roughness to Work." In *How Well do Facts Travel? The Dissemination of Reliable Knowledge*, 111–35, edited by P. Howlett and M. S. Morgan. Cambridge, UK: Cambridge University Press.
- Waugh, D. (2009). *Geography: An Integrated Approach*. Oxford: Oxford University Press.