Presented at GISRUK 1997. Accepted for publication in Innovation 5, ed. S. Carver, Taylor and Francis, 1998.

Computational techniques for non-Euclidean planar spatial data applied to migrant flows

Mike Worboys, Keith Mason and Jenny Lingham

Departments of Computer Science and Geography
Keele University, Staffs ST5 5BG UK

tel: 01782 583078, fax: 01782 713082, email: michael@cs.keele.ac.uk

Abstract

This paper describes the development of an approach to spatial analysis in the setting of a non-Euclidean space. Almost all current GIS functionality assumes that constituent spatial entities in the system are referenced to the coordinatized Euclidean plane. However, there are many situations for which representation in a Euclidean space is not the most appropriate model, nor may even be derivable. Spatial interaction provides an example, since the interactions determine relationships between spatial entities that may not be representable in Euclidean space. The paper presents some background to interaction analysis and non-Euclidean techniques. The major contribution is to develop a method for treatment of non-Euclidean proximity relationships and apply it to migrant flow data.

1. Introduction

Almost all current GIS functionality assumes that constituent spatial entities in the system are referenced to the coordinatized Euclidean plane. That is, they all have coordinates positioning them with respect to a predetermined reference frame. Also, the usual Euclidean properties and relationships like distance and bearing between locations, lengths and angles of line segments, and area and perimeter of regions, may be calculated in the normal way. However, there are many situations for which representation in a Euclidean space is not the most appropriate model, nor may even be derivable. Examples include travel time spaces, qualitative distances and flow spaces.

This paper considers the potential for spatial analysis that takes us outside the geometry associated with Euclidean space. Topology is the key concept, and the paper explores the potential for using topological concepts, even when a metric or Euclidean space is not available. In particular, since many geospatial phenomena are best modelled in a finite context, we explore the potential of topologies on finite sets (so called *finite topologies*). We show that finite topologies provide an analysis of collections of spatial entities as hierarchies.

The methods are applied to flow spaces, resulting from linkages between geographically distinct areas brought about by 'flows' between origins and destinations, and exemplified by journeys-to-work, diffusion of disease or of innovations, population migrations, freight flows, international trade, patient referrals to hospital and shopping patterns. In seeking to represent or analyze such interactions the objective is to understand or to model and predict the flows, and a variety of techniques for such analysis and visualization are presented in this paper.

2. Background

2.1 Visualizing interactions: Flow mapping

Of all cartographic forms, flow mapping is one of the most problematical. Perhaps the simplest visualizations are those that deal only with the end product - choroplethic representations showing rates of inflow or outflow per unit area, without any indication of respective origins or destinations. Alternatively, symbolization of the interaction itself, the flow, may be used whenever visualization of dynamic linkages between places is required. Broadly, flow maps are constructed according to one of two basic styles: those which are concerned with qualitative movements and which symbolize linkages and directionality by way of simple unscaled lines and arrowheads, and those which are concerned as well with quantitative movements, where the width of the line is varied in proportion to the amount of flow recorded between places.

A number of innovative variations upon these solutions to mapping flows have also been reported, for the most part in avoidance of problems associated with line overlap and visual clutter. One widely used variant (see for example Dorling 1991) is that which symbolizes connectivity through use of flow lines of equal width, but also portraying quantity of movement by means of proportional symbols placed at line ends or at the centre of areas receiving or generating the flows.

2.2 Analyzing interactions

Moving beyond visualization, spatial interaction models allow investigation of the determinants of flow spaces. Spatial interaction models usually operate on the premise that flows are likely to be in some way related to the sizes of the origin and destination areas, and in some way inversely related to the distance between them. By such means, patterns of flows can be modelled in terms of accessibility of destinations from origins, and concepts of demand and attractiveness - the ability of different areas to generate a flow - can be introduced into the analysis. One of the most common classes of spatial interaction model is the gravity model (Wilson, 1974), where flow is analogous to gravitational attraction and the generating and pull factors associated with origins and destinations is analogous to mass. Where the fitted model exactly reproduces the total observed flow such models are referred to as being 'doubly constrained' - separate parameters being allowed for each origin and destination. However, whilst the doubly constrained model provides useful insights into the role of distance in encouraging spatial interaction and provides information on the relative contributions of origins and destinations to the process, it provides little in the way of explanation for the observed flows. Consequently, a number of variations upon the basic model might be employed. The 'origin constrained' model attempts only to ensure that the predicted flows from an origin equal the actual flow, while the 'destination constrained' model is concerned that the reverse is true - that predicted flows into a destination are equal to those actually observed. The advantages of both adaptations is that they allow alternative configurations or additional origin or destination parameters to be incorporated into the model to predict total flows.

The overall patterns of flow associated with a particular place can be summarized through the use of migration fields. A migration field delineates the origins of immigrants to an area or the destinations of emigrants from that area. Typically, the fields will be defined as those areas contributing more than a specified number of migrants to the overall pattern of flow. A technique known as primary linkage analysis can be used to identify such fields (Haggett, Cliff and Frey 1977). In this, the hierarchical order of a region is measured by the number of people moving to that region and the largest outflow to a higher order region is used to determine the overall hierarchical structure.

2.3 New approaches to analysis and visualization using non-Euclidean spaces:

In this paper we investigate ways of representing and reasoning about interactions between spatial objects that are not necessarily referenced to the Euclidean plane. The work is motivated by the interest within the research community in the idea of 'geographic space', that may not be Euclidean (see for example the concept of ‘naive geography’ introduced by Egenhofer and Mark (1995)). Our view of the geographic space which surrounds us may be quite different from the representation of that space on a map; the distance relationship may not be symmetric, in that it may be quicker to go from A to B than it is to go from B to A, perhaps because of the terrain or the traffic conditions. We can think of ourselves as being immersed within geographic space, unable to view the whole of it at any one time. In order to explore it we must move around in it, both spatially and temporally. This space is thought of as being a horizontal and two dimensional. It emphasizes topological information, lacks the usual distance metric, and frequently lacks complete information.

Topology is the study of the properties of a space which are preserved when it is subjected to continuous transformations, for example rotations or stretches. For a general treatment of topology the reader is referred to Kuratowski (1996) or to Henle (1979). Digital topology has been used in fields such as computer vision, graphics and image processing for the study of the topological properties of image arrays. There is a substantial research literature on this subject, for instance Kong and Rosenfeld (1989). The work discussed here uses finite topologies, the topologies of finite sets (see for example Kovalevsky (1989)). As topology is essentially the study of continuity and limiting processes, it might be thought that finite topology would provide us with little insight. However, finite topologies do provide a hierarchical structure on sets of entities which is precisely the structure which we develop later in the paper. A start to this line of research is described by Worboys (1996) and defines a topology in a non-metric space, based on the construction of localities. Related work has been carried out by Egenhofer and Franzosa (1991) who present a framework for the definition of topological relations which is independent of the existence of a distance function.

Montello (1993) has discussed the importance of alternative geometries in the study of the large scale space in which we live, and which he refers to as environmental space. Montello notes that movement over time is usually required to explore this space, and that many authors have suggested that a Euclidean model is not the best one to represent it. Gahegan (1995) proposes qualitative spatial reasoning as a way of reasoning about objects which deals with topological relationships between them rather than their positions in space. It is suggested that the idea of proximity need not imply physical closeness, and could also depend on the context.

Another approach to the problem of dealing with objects in non-Euclidean space has been the use of multi-dimensional scaling (MDS). This is described by Gatrell (1983). MDS is used when we have a set of objects and wish to create a map representing the relationships between the objects. The input to a multi-dimensional scaling algorithm is a matrix of dissimilarities or distances. The output is a spatial configuration of the objects so that the distances between pairs of objects within the configuration match as closely as possible the input dissimilarities. The effect of the MDS is to distort the Euclidean plane. Gatrell also describes and gives examples of the use of Q-analysis, based on the work of Atkin (1978). Q-analysis is a technique that can reveal connectivities provided by relations on a set , or between sets, using concepts from algebraic topology.

Other important work which does not assume a Euclidean framework but attempts to describe directly spatial relationships such as contact and boundary includes the work of Clarke (1981, 1985), and later Cohn *et al.* (1993).on the Region Connection Calculus (RCC). Smith (1995) gives an account of mereotopology, which combines mereology (study of the part-whole relationship) and topological notions.

3 Migrant flows

3.1 Population Migration

In the UK, studies of migrant flows have received much recent attention, largely as a result of the role of internal migration in shaping local population distributions and structures, in circumstances of little or no natural population increase. Several important trends in population migration in the UK have been observed at different times and at a variety of geographical scales. Long-distance movements related to place-of-work have been particularly significant. Rural to urban and then, through counter-urbanization, urban to rural population movements have been instrumental in population mixing across the country.

Movements between Family Health Services Authorities (FHSA's) within England and Wales during the 1980's have been measured by the National Health Service Central Register (Rosenbaum and Bailey, 1991). Using patient re-registrations the biggest net gainers of population through migration, in terms of average annual gain, were found in the South West, the South East (excluding Greater London), and East Anglia. The principal losers were in Greater London. In broad terms, it was the non- metropolitan FHSA's which gained population during the 1980's, and the metropolitan FHSA's which lost. In Scotland, migration flows have been calculated from the Register of Sasines, which records every sale of inter alia private residential property (McCleery, Forbes and Forster, 1996). From this, short-distance migration is seen to dominate - a factor which is sometimes blurred when more coarse data has been employed elsewhere.

There are several reasons why population movements deserve close attention. Clearly, as an important component of population change, a proper understanding of migration is required to construct realistic population projections and to inform resource allocation decisions. But as well, the selective nature of migration can shape and re-shape the social and demographic characteristics of those areas losing or gaining populations by such means. This has implications for local authority revenue raising and service provision. Similarly, land-use planning policies and levels of investment in housing are influenced by migration. At the same time, population movements have important implications for epidemiological studies, not least in the spread of disease through increased contact made between susceptibles and infectives, but also through the combined affects of migration and disease latency as a confounding problem in trying to explain patterns of disease distribution. The explanation for the occurrence of a disease will have as much to do with past histories as it does with present sets of circumstances.

Given the significance of migration for shaping the geography of the United Kingdom, it is important that the range of techniques for visualizing and analyzing migrant flows is as extensive as possible. The method developed below introduces an approach to analysis that is less dependent upon the Euclidean characteristics of the spatial framework.

3.2 The Data

For present purposes the Special Migration Statistics (SMS) of the 1991 Census (Flowerdew and Green, 1993) will be used to establish migration flows between Family Health Service Authority (FHSA) areas. The latter have been selected for compatibility and comparability with the only other major dataset on migrant flows in this country, that developed from the National Health Service Central Register, which is also aggregated to FHSA's. In the SMS a migrant is defined as: 'a person with a different usual address one year ago to that at the time of the Census'. The SMS contain both simple origin-destination flow counts as well as more complex cross-classified counts. Clearly, the SMS can provide nothing more than a snapshot of migration patterns as they existed in the year prior to the Census. Nevertheless, they are sufficient to allow explorations of the ways in which non-Euclidean approaches may be brought to bear, and at the same time throw new light on spatial interactions as they exist through migrant flows in the UK. Figure 1 shows a representation of the dataset for the South of England, where an arrow is shown between region *A* and region *B* if not less than 20% of the total flow from *A* goes to *B*.

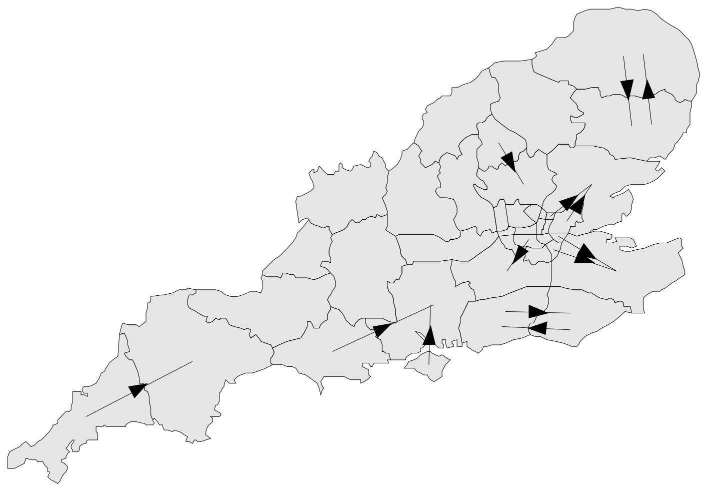

Figure1: Significant migrant flows shown for the South of England

4 The non-Euclidean space of localities

4.1 From flows to proximities

The technique will be explained with reference to the migration flow data described earlier. It should be emphasized that the method is quite general and may be applied to any proximity or accessibility relationship between spatial entities. The migration flow data consists of a 113 by 113 matrix $\mathbf{F} = [f_{ij}]$ of migratory flows, where f_{ij} is an integer representing the amount of flow from location *i* to location *j*, the locations being the 113 FHSA regions in England, Scotland and Wales. We convert the flows into normalized measures of proximity (or accessibility) p_{ij} by means of the functional relationship

$$\log_2(1 - p_{ij}) = - f_{ji} / \mu$$

where μ is the mean of all the flows in the matrix. For our example, $f_{12} = 1867$ and so $p_{21} = 0.97$.

Note that $0 \leq p < 1$, small flows result in near-zero proximity values, large flows result in values of proximity near to one, and a mean flow results in a proximity value of 0.5. The intuition is that the quantity of flow from *j* to *i* is an indication of the level of accessibility of *i* from *j* or the proximity of *i* to *j*. This measure of proximity is very different from the usual Euclidean distance, not necessarily being symmetric, for example. We are thus unable to proceed with the usual Euclidean analysis. The next step in our approach is to convert the proximity measure to a Boolean relation π with reference to a fixed cut-off parameter κ as follows:

$i\pi j$ if and only if $p_{ij} > \kappa$

In our example, we take κ=0.6, and as p_{21} = 0.97, it is true that $2\pi1$. We assume that $i\pi i$ for each location i, thus π is a reflexive relation. This assumption is strengthened by the large values observed for flows from a location to itself.

4.2 Localization

The next subsection will show that to derive a topology from the proximity relationship, it must be a partial order. A relation ρ is a partial order if it satisfies the following properties:

For all locations i,	$i\rho i$	REFLEXIVE
For all locations i, j, k	$i\rho j$ and $j\rho k$ implies $i\rho k$	TRANSITIVE
For all locations i, j	$i\rho j$ and $j\rho i$ implies $i = j$	ANTISYMMETRIC

We already have assumed that the proximity relation is reflexive. However, it is unlikely that any nearness type relation will be transitive. The antisymmetric rule holds if there do not exist distinct locations i and j for which i is proximal to j and j is proximal to i, again unlikely to hold.

We get the properties of a partial order from our proximity relation by a localization process now to be described. Define a new relation λ on the set of locations as follows. For all locations i, j, $i\lambda j$ if and only if any location which is proximal to i is proximal to j. That is:

$i\lambda j$ if and only if for all locations k, $k\pi i$ implies $k\pi j$

We read $i\lambda j$ as ‘location i is local to location j’. In our example ...

It is not hard to show that λ is both reflexive and transitive. It may also be made antisymmetric by factoring out (coalescing) all locations for i, j which $i\lambda j$ and $i\lambda j$. So relation λ provides a partial order on the set of location, thus providing a hierarchy of locations.

The partial order for the migratory flow data is shown in figure 2, where a directed segment from i to j represents the relation $i\lambda j$. In figure 2, the visualization relies on a mix of Euclidean and non-Euclidean spaces, the embedding of locations being Euclidean but the partial order being non-Euclidean.

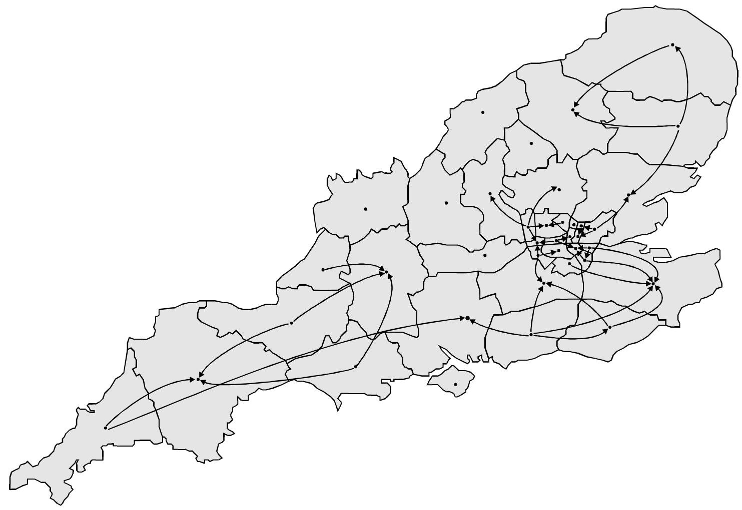

Figure 2: Locality relationship shown for the South of England

4.3 Finite topologies

Many fundamental relationships between spatial entities are describable in topological terms. A spatial relationship is *topological* if is invariant under continuous deformations of space. The topology that has been of almost universal concern in geospatial data handling is the so-called 'usual topology' of the Euclidean plane, derived from the 'as-the-crow-flies' distance function. But many other topologies exist, in all of which such concepts as interior, boundary and connectivity are defined and may have useful interpretations. In the general setting, a topological space is a set with a distinguished collection of subsets (called the *open sets*). These open sets must obey the following three conditions.

T1 The empty set and the entire space are open sets.

T2 The union of any family of open sets is an open set.

T3 The intersection of any finite family of open sets is an open set.

Further, if the topological space satisfies the further property that for any two distinct points in the space, there is an open set which encloses one but not the other, then the space is called *separable*.

The topologies that we use in this work are finite, in the sense that the underlying set is finite. It might be argued that restricting consideration to finite topologies misses many important aspects of topology, including continuity and limit processes. However, many models of real world data and processes are finite, and we contend that for finite topologies useful properties remain. Condition T3 implies that the intersection of all collections of open sets in a finite topology is open. For any location, consider all open sets that have it as a member, then the intersection of these open sets will be open, and is the minimal open set to which it belongs. Minimal open sets do not necessarily exist in the infinite case.

The key result for this work is that every separable, finite topology may be conceived as a partially ordered set. And conversely, any finite partially ordered set may be structured as a separable, finite topology. The proof of this result relies heavily on the concept of minimal open sets and is not hard, but outside the scope of this paper, and may be found in (Worboys 1997).

Returning to our flow analysis, at the end of the last subsection we showed that the locality relation λ results in a partial order on the set of FHSAs, and we have just shown that this leads to a topology. The remaining questions are how can this topology be visualized, and how does it help us in analysis of migration between FHSAs.

4.4 Visualization of finite topologies

This sub-section presents two separate visualizations of a finite topology (and therefore a finite partially-ordered set). The first approach is using the notion of a topological neighbourhood of a location. In the finite case, the minimal neighbourhoods of locations exist and correspond to the minimal open sets. Intuitively, these neighbourhoods may be thought of as regions of influence of their locations. In the case of proximity neighbourhoods (that is, neighbourhoods arising from the localization of proximity relations, the neighbourhoods may be thought of localities). Formal details of their properties may be found in (Worboys 1997). Figure 3 shows the neighbourhoods in the South of England in the case of the FHSA data.

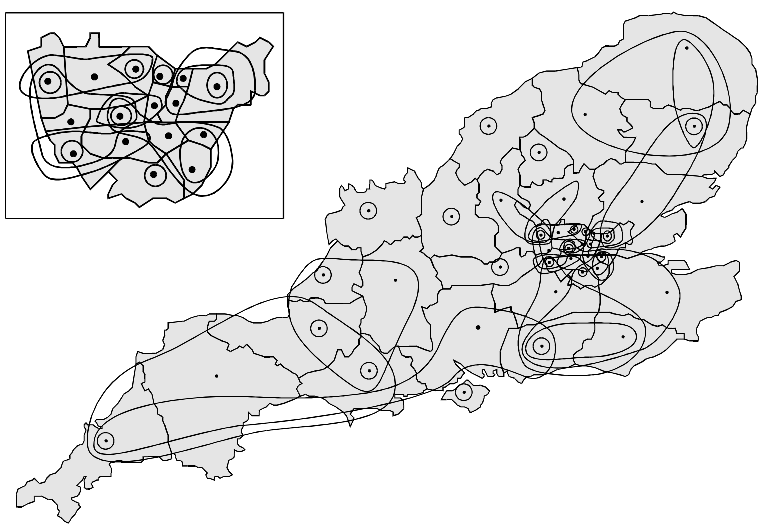

Figure 3: Neighbourhoods shown for the South of England with inset showing expanded view of London

These neighbourhoods can be combined into maximal connected components of the space. The resulting zoning provides some insight into the partition of the space into localities, each of which is a relatively self-contained region of spatial interaction (in this case through migration). Figure 4 shows the partition of the South of England.

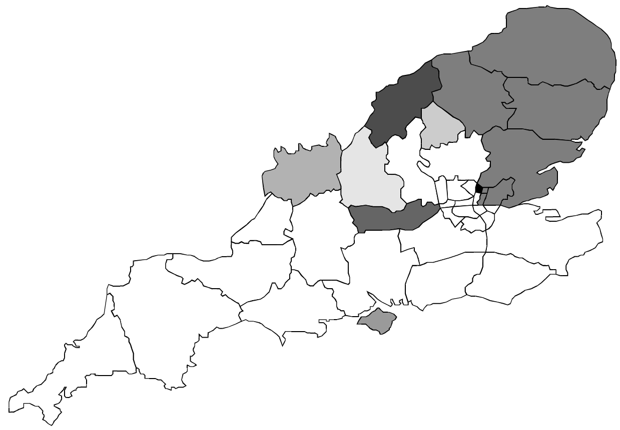

Figure 4: Partitioning localities for the South of England

The second approach relies on the result (again not proved here see [Worboys 1997] for details) that the hierarchy relating to a finite topology may be visualized as a cellular complex. A *cell* can be thought of as a face of a *n*-dimensional polytope: a point in 0-dimensions, straight line segment in 1-dimension, polygon in 2-dimensions, polyhedron in 3-dimensions, and so on. A cellular complex is a collection of cells, with the property that any face of a cell in the complex must also be in the complex. A hierarchy relating to a finite topology may be visualized as a cellular complex, with locations down the hierarchy represented as lower-dimensional faces of locations higher in the hierarchy. Figure 5 shows part of the cellular complex that represents the FHSA example.

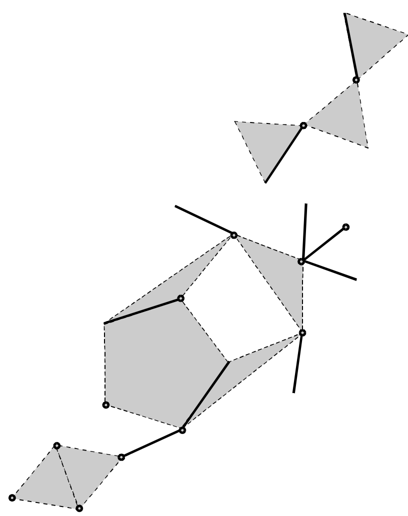

Figure 5: The FHSA migratory flow data for the South visualized as a cellular complex

References

ATKIN, R.H., 1978. Time as a pattern on a multi-dimensional structure. *Journal of Social Biol. Struct,* **1**: 281-295.

CLARKE, B.L.,1981. A calculus of individuals based on 'connection'. *Notre Dame J. Formal Logic,* **22**: 204-218.

CLARKE, B.L., 1985. Individuals and points. *Notre Dame J. Formal Logic,* **26***:* 61-75.

COHN, A.G., RANDELL, D.A., CUI, Z. AND BENNETT, B., 1993. Qualitative Spatial Reasoning and Representation. In Carrete, P. and Singh, M.G. (eds.), *Qualitative Reasoning and Decision Technologies,* CIMNE, Barcelona.

DORLING, D., 1995. Visualizing the 1991 census. In Openshaw, S. (ed.) *Census Users Handbook.* Cambridge: GeoInformation International.

EGENHOFER, M.J. AND FRANZOSA, R.D., 1991. Point-set topological spatial relations. *Int. J. Geographical Information Systems,* **5**(2): 161-174.

EGENHOFER, M.J. AND MARK, D.M., 1995. Naive geography. Technical report 95-8, National Centre for Geographic Information and Analysis.

FLOWERDEW, R. and GREEN, A., 1993. Migration, transport and workplace statistics from the 1991 Census. *The 1991 Census Users Guide* (ed. Dale, A. and Marsh, C.), HMSO.

GAHEGAN, M., 1995. Proximity operators for qualitative spatial reasoning. In Frank, A.U. and Campari, I. (eds.) *Spatial Information Theory, Proceedings of COSIT,95, Lecture Notes in Computer Science*. Springer-Verlag, Berlin, Germany, pp.31-44.

GATRELL, A.C., 1983. *Distance and space: A geographical perspective.* Clarendon Press, Oxford.

HAGGETT, P., CLIFF, A.D. and FREY, A., 1977. *Locational Analysis in Human Geography.* London: Edward Arnold.

HAYES, P., 1978. The Naive Physics Manifesto. In Mitchie, D. (ed.) *Expert Systems in the Microelectronic Age.* Edinburgh, Scotland: Edinburgh University Press, pp. 242-270.

HENLE, M., 1979 *A Combinatorial Introduction to Topology*. San Francisco: Freeman.

KONG, T.Y., AND ROSENFELD, A., Digital Topology: Introduction and Survey. *Computer Vision, Graphics and Image Processing,* **48**: 357-393.

KOVALEVSKY, V.A., 1989. Finite topology applied to image analysis. *Computer Vision, Graphics and Image Processing,* **46**: 141-161.

KURATOWSKI, 1996. *Topology.* New York: Academic Press.

McCLEERY, A., FORBES, J. and FORSTER, E., 1996. Deciding to move home: A preliminary analysis of household migration behaviour in Scotland. *Scottish Geographical Magazine,* **112**(3): 158-168.

MONTELLO, D.R., 1992. The geometry of environmental knowledge. In Frank, A.U., Campari, I. and Formentini, U. (eds.), *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Lecture Notes in Computer Science 639.* Springer-Verlag, Berlin, Germany, pp. 136-152.

ROSENBAUM, M. and BAILEY, J., 1991. Movement within England and Wales during the 1980's, as measured by the NHS Central Register. *Population Trends* **65**: 24-34.

SMITH, B., 1996. Mereotopology: a theory of parts and boundaries. Submitted to *Data & Knowledge Engineering.*

WILSON, A.G., 1974. *Urban and Regional Models in Geography and Planning.* London: Wiley.

WORBOYS, M.F., 1996. Metrics and Topologies for Geographic Space, in *Advances in Geographic Information Systems Research II: Proceedings of the International Symposium on Spatial Data Handling, Delft*, Kraak, M.J. and Molenaar, M. (eds.), International Geographical Union, pp. 7A.1-7A.11, 1996.

WORBOYS, M.F., 1997. Using finite topologies to extract hierarchies from data referenced to non-Euclidean spaces. Submitted for publication.