Proceedings

The First ICA Workshop on Geospatial Analysis and Modeling

8 July 2006, Vienna, Austria

Editors: Bin Jiang
          Daniel Z. Sui
About the cover

© MSN Maps & Directions

The location of the workshop is pinpointed with MSN Maps & Directions (mappoint.msn.com, or previously www.mapblast.com), with which we can easily make a query on the direction from a hotel you may stay to the workshop venue. Indeed, the online map engine has made the organization of the event much easier.
Preface

Current Geographic Information Systems (GIS) have maintained an increasing amount of geographic information (GI) collected from the real world. Nevertheless, our understanding of the GI for making sound decisions about real world problems is primitive. This gap between what GIS can provide (massive GI) and what is needed (understanding of GI) is created due to the lack of appropriate GIS models and tools, and makes real world problem solving problematic. This can be generally characterised as “information rich and knowledge poor”. That is, the amount of GI collected using GIS technologies involving remote sensing, global positioning systems and digital mapping has been growing rapidly, whereas geographic knowledge or useful information so to speak, for various decision makings in terms of real world problem solving, is rather limited. “Information rich and knowledge poor” has been becoming more and more obvious in the age of the Internet. For example, Google Earth (http://earth.google.com/) has integrated terabytes of satellite imagery, aerial photos and GIS data in its powerful servers, with which ordinary people using decent internet connection PC can zoom from space right down to street level, and easily pinpoint their individual houses. However, the Google service is still considered to be an information provider, rather than a knowledge supplier, because it lacks relevant tools for uncovering geographic knowledge, in various forms of patterns, structures, relationships, and rules.

Conventional maps provide a good way to illustrate patterns and structures of geographic phenomena and interactive visualization can further facilitate the knowledge discovery process. However, what cannot be ignored is that the underlying analytical and modelling methods can significantly contribute to the knowledge discovery process, in particular when the methods are integrated with various visual approaches. This is particularly true for geospatial information, which is essentially massive, complex, incomplete and uncertain. It is under this view that I came up with the idea of creating a new International Cartographic Association (ICA) commission on geospatial analysis and visualization (later on replaced by modeling to avoid a possible overlap with another commission). In the last International Cartographic Conference (ICC) held in La Coruña, Spain, the ICA executive committee officially approved my proposal to start an ICA working group on geospatial analysis and modelling. For this working group, we set the terms of reference as follows: (1) to network cartographers, geographic information scientists, and other researchers involved in geospatial analysis and modeling, (2) to facilitate the interaction and communication between the computational community and the cartographic community for creation of geographic knowledge; (3) to foster a new research community centered on visual geospatial analysis and modeling; (4) to organize ICC sessions and ICA workshops on geospatial analysis and modeling, and (5) to edit and publish geospatial analysis and modeling related reports, books, and special issues with scientific journals.

Collected with the proceedings is a set of papers selected from in total 24 full paper submissions following up our call for papers (http://www.hig.se/~bjg/ica/workshop/). The set of papers represents a perspective of current development on geospatial analysis and modeling. The workshop provides a forum for the researchers to discuss their work and exchange views on the current development. By the opportunity, I would like to thank all the authors for their quality contribution in making the event possible. I would also like to thank the ICA executive committee who entrusted me with the organization of this new working group. Special thanks go to Professor Wolfgang Kainz for providing essential logistic help for the workshop, my colleague Professor Daniel Z. Sui for his nice collaboration in the joint effort, and the Swedish Research Council Formas for providing partial financial support for the workshop.

Bin Jiang (Chair of the ICA working group on geospatial analysis and modeling)
## Contents

<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>iii</td>
</tr>
<tr>
<td>Bin Jiang</td>
<td></td>
</tr>
<tr>
<td>Movement beyond the snapshot - dynamic analysis of geospatial lifelines</td>
<td>1-20</td>
</tr>
<tr>
<td>Patrick Laube, Todd Dennis, Mike Walker, and Pip Forer,</td>
<td></td>
</tr>
<tr>
<td>Modular system of simulation-patterns for a spatial-process laboratory</td>
<td>21-34</td>
</tr>
<tr>
<td>Reinhard König and Christian Bauriedel</td>
<td></td>
</tr>
<tr>
<td>The dynamic geometry of geographical vector agents</td>
<td>35-46</td>
</tr>
<tr>
<td>Yasser Hammam, Antoni Moore, Peter Whigham, and Claire Freeman</td>
<td></td>
</tr>
<tr>
<td>Agent-based simulation for building fire emergency evacuation</td>
<td>47-57</td>
</tr>
<tr>
<td>L. Hajibabai, M. R. Delavar, M. R. Malek1 and A. U. Frank</td>
<td></td>
</tr>
<tr>
<td>How people visually explore geospatial data</td>
<td>58-72</td>
</tr>
<tr>
<td>Urška Demšar</td>
<td></td>
</tr>
<tr>
<td>Examining potential demand of public transit for commuting trips:</td>
<td>73-84</td>
</tr>
<tr>
<td>an integrative analysis combining GIS, statistics and data mining</td>
<td></td>
</tr>
<tr>
<td>Xiaobai Yao</td>
<td></td>
</tr>
<tr>
<td>A novel accessibility measure using travel time and space syntax indices</td>
<td>85-97</td>
</tr>
<tr>
<td>Seungjae Lee, Shinhae Lee, Byoungwook Lee, Chulmin Jun and Seungil Lee</td>
<td></td>
</tr>
<tr>
<td>The Implication of a City’s layout’s Visibility on Wayfinding Performance</td>
<td>98-109</td>
</tr>
<tr>
<td>Itzhak Omer and Ran Goldblatt</td>
<td></td>
</tr>
<tr>
<td>Effective Wayfinding Based on LBS Using Landmarks in Urban Environments</td>
<td>110-120</td>
</tr>
<tr>
<td>N. Samani, M. R. Delavar and M. R. Malek</td>
<td></td>
</tr>
<tr>
<td>The use of rough set in geographical information systems</td>
<td>121-130</td>
</tr>
<tr>
<td>Beniamino Murgante, Giuseppe Las Casas, and Anna Sansone</td>
<td></td>
</tr>
<tr>
<td>A topological interpretation for mass transit network connectivity</td>
<td>131-140</td>
</tr>
<tr>
<td>Seungjae Lee, Byoungwook Lee, Chulmin Jun, Seungil Lee, and Shinhae Lee</td>
<td></td>
</tr>
<tr>
<td>Implementation of a generic road-matching approach for the integration of postal data</td>
<td>141-154</td>
</tr>
<tr>
<td>Meng Zhang and Liqiu Meng</td>
<td></td>
</tr>
</tbody>
</table>
DMA: An algebra for multicriteria spatial modeling  
Salem Chakhar and Vincent Mousseau  

Spatial dependency modeling using spatial auto-regression  
Mete Celik, Baris M. Kazar, Shashi Shekhar, Daniel Boley, and David J. Lilja  

Performing topological correction in a GIS  
Stefano Bellesi and Eva Savina Malinverni
Movement Beyond the Snapshot
— Dynamic Analysis of Geospatial Lifelines

Patrick Laube¹, Todd Dennis², Mike Walker² & Pip Forer¹

¹School of Geography and Environmental Science
University of Auckland, Auckland, New Zealand
Phone: +64 9 373-7599 # 88202 Fax: +64 9 373-7434
Email: [p.laube, p.forer]@auckland.ac.nz
²School of Biological Science
University of Auckland, Auckland, New Zealand
Email: [t.dennis, m.walker]@auckland.ac.nz

ABSTRACT

GIScience is challenged by an unprecedented increase in the availability of tracking data related to human and animal movement, typically captured through location-aware portable devices such as GPS receivers. Capture of trajectory data at fine temporal and spatial granularities has allowed with the representation of detailed geospatial lifelines, opening new options for analysis. In this respect we propose a dynamic perspective to analysis which, in contrast to summary trajectory statistics on speed, motion azimuth or sinuosity, that refers to the variability of motion properties throughout the developing lifeline. Four specific lifeline context operators are identified in this paper: ‘instantaneous’, ‘interval’, ‘episodal’ and ‘total’. Using this framework, we discuss standardisations that integrate the extended set of motion descriptors within various temporal and spatial frames of reference and the proposed lifeline context operators and standardisations are illustrated using high resolution trajectory data obtained from homing pigeons carrying miniature global positioning devices.

Keywords. Geospatial lifelines, trajectories, movement, context operators, standardisation.

1 INTRODUCTION

Opportunities to trace individual movement have grown in tandem with the development of electronic transaction networks, location-aware devices and surveillance systems, all capable of tracking people (Mountain & Raper, 2001b), animals (Hulbert, 2001) or vehicles (Wolfson, Sistla, Chamberlain, & Yesha, 1999). Generically, this development has offered an opportunity to move ‘beyond the snapshot’ (Chrisman, 1998) with respect to our understanding of processes involving individual movement. Specifically, the recent arrival of devices capable of the low cost capture of high resolution locational data now allows the widespread construction of individually-based geospatial lifelines (Mark, 1998). Such individual lifelines presage a new era of movement analysis (Eagle & Pentland, 2005) in which scientists from various research fields previously
hampered by sparse and random movement observations can now be hard on the heels of their subjects as they move in space and time.

Yet while there is a growing commitment of resources to the large-scale recording of paths the analysis commonly conducted with trajectory data remains fairly limited in scope and sophistication (Wolfer, Madani, Valenti, & Lipp, 2001). In disciplines outside of geography which do not commonly use geospatial methods or theory this may be due to a lack of awareness and understanding of the power of spatial analysis and GISystems, and within geography GIScience’s’ fetish for the static may be a factor (Raper, 2002). What ever the cause, GIScience faces a challenge to develop sophisticated analytical tools that integrate geography’s spatial awareness with its long-term experience in processing large spatio-temporal data bases. This paper discusses opportunities and shortcomings of analysing lifeline data from a GIScience perspective, specifically in the situation where three spatial dimensions are involved and where movement is largely unfettered.

Clearly, the representation and analysis of geospatial lifelines challenge the GIScience community with respect to procedures for aggregation, generalisation, uncertainty and interpolation. Aggregation and generalisation of lifeline data can be considered to be important tools for coping with the voluminous outputs of movement-related agent-based simulations, for example with respect to emergency planning (Batty, Desyllas, & Duxbury, 2003) or transportation planning (Nagel, Esser, & Rickert, 2000). Uncertainty and generalisation of lifeline data is of interest for designers of location based services (LBS) analysing the lifelines of people tracked by location aware devices (Duckham, Kulik, & Worboys, 2003; Mountain & Raper, 2001a; Smyth, 2001). The more basic derivation of actual motion descriptors, such as speed, motion azimuth, or path sinuosity, also merits attention. Such descriptors build, the underlying basis for attempts to investigate the similarity of trajectories (Sinha & Mark, 2005), which is an important task in spatio-temporal data mining and geographic knowledge discovery (Miller & Han, 2001), and may well pay dividends in the fields of spatialisation (Skupin & Fabrikant, 2003), eye-movement analysis (Fabrikant, 2005) and the evolution of semantic relationships in cognitive spaces (Pike & Gahegan, 2003).

Lifeline data analysis is also relevant to a range of applied research fields outside of geography itself or in cognate areas, such as animal biology and biogeography. In behavioural ecology, the key factors in avian navigation are still not completely understood (Bonadonna et al., 2005; Wiltschko & Wiltschko, 2003), but Steiner et al. (2000) identify the analysis of the homing routes of racing pigeons as the optimal method for almost any study in this field. Advanced path analysis is furthermore considered to be a crucial obligation for the interpretation of behavioural experiments conducted with genetically modified animals, for example for water maze experiments with mice exploring spatial learning (Wolfer, Madani, Valenti, & Lipp, 2001). Similar analysis is also of increasing interest in agricultural science, contributing to the development, for example, of optimal grazing strategies for cattle with respect to livestock management (Ganskopp, 2001). Research involving video surveillance (Ng, 2001; Porikli, 2004; Shim & Chang, 2003) or sports scene analysis (Moore, Whigham, Holt, Aldridge, & Hodge, 2003) are further examples where disciplines exhibit deep interests in individual trajectories.

The distinctive feature of high resolution tracking data is that they allow the tracking of individuals along an actual movement path, leaving little need for interpretation between sparse observation points. Thus, at almost every instant along the lifeline we can robustly determine the individual’s current movement properties, such as speed, acceleration, motion azimuth, path sinuosity, as well as generate even more complex motion properties. Section 2 reviews work in this area, and elaborates the nature of the existing analytical tool set. In section 3.1 we propose lifeline context operators to derive motion descriptors along dense trajectories, and adopt a dynamic
analysis perspective for this. Section 3.2 rehearses a number of well known motion descriptors such as speed or motion azimuth, and introduces additional measures before discussing their computation as lifeline context operators. Section 3.3 explores the wide variety of analytical options that can be deployed by applying various standardisations for statistical analysis or using different aggregations of the tracking data. To illustrate these possibilities the trajectories of homing pigeons are identified as a research milieu (section 4), and data from this area are utilised to demonstrate and critique the techniques (Section 5). The paper concludes in section 6 with a brief review of prospects.

2 RELATED WORK

There is ample research on deriving overall descriptors of movement trajectories, such as time total, flying time, airline distance, net displacement, flight path, flight speed, bias from airline (e.g. Berger, Wagner, & Wolff, 1999; Steiner et al., 2000; Turchin, 1998; Wolfer, Madani, Valenti, & Lipp, 2001). In avian navigation research, where detailed tracking was impossible until recently, the recording of the vanishing bearing of homing pigeons was used as a ‘whole flight’ indicator, since this direction was believed to be closely related to their loft’s direction (Steiner et al., 2000; Wiltschko & Wiltschko, 2003). Further summary trajectory descriptors are the various measures describing the sinuosity or tortuosity of trajectories (Benhamou, 2004; Claussen, Finkler, & Smith, 1997).

Much less work has been done on describing movement with respect to instantaneous or ongoing characteristics of the trajectory, although the database community, especially people working on Moving Object Databases (MOD), have contributed with specific data models and query languages designed to return the state of a moving object at a given time, with ‘state’ referring to location, speed, or even direction (Sistla, Wolfson, Chamberlain, & Dao, 1997; Wolfson & Mena, 2004). Yet it is only recently that the need to move further beyond summary trajectory measures, and to dynamically explore the detailed information stored in individual lifelines, has been acknowledged in the behavioural ecology literature (Benhamou, 2004; Bonadonna et al., 2005). Benhamou (2004) for instance is exploring in-path variations of turning angles by using a window moving along the trajectory to derive the mean turning angle of the segments within that window. However, while a moving window is a straightforward device for deriving a range of motion properties along a trajectory, it fails to address the full complexity of dense spatio-temporal data. Mennis, Viger, and Tomlin (2005, p. 18) echo this situation in respect of the potential for additional approaches, noting that, it is very surprising that ‘despite the growing volume of research on spatio-temporal data models over the past dozen or so years, the extension of map algebra to the temporal dimension has been largely ignored by the spatio-temporal GIS research community.’ These authors propose ‘cube functions’ as an extension of map algebra for three dimensional spatio-temporal snapshot data (Mennis, Viger, & Tomlin, 2005) or even multidimensional data (Mennis, Leong, & Khanna, 2005), an approach also developed and critiqued in the context of movement and possible presence by Huisman (2006).

This paper seeks to explore the ground between the simple global analysis of movement paths and the more complex creation of spatio-temporal algebras by extending the dynamic analysis tools that combine aggregation, generalisation and information on sequence to generate new movement descriptors. In doing this we acknowledge both power and danger of using multiple temporal granularities and temporal zooming in the process of spatio-temporal knowledge representation. Theoretically recognised (Hornsby, 2001; Hornsby & Egenhofer, 2002), the actual algorithmic challenges of handling variable temporal granularities are sparsely addressed and the impact of different granularities on analyses is, as ever, a complex issue and under-researched.
Dutton’s (1999) work is one of a small number of exceptions by addressing line sinuosity as a way to investigate scale-specificity and characteristic points for line generalisation.

While more general approaches to derive procedural motion properties are also relatively few, a number of initiatives have encountered the issue of movement description. Dykes and Mountain refer to lifeline segments primarily for visualisation and exploratory analysis (2003), and propose the notion of episodes as well as a number of episode summaries and indices of movement, including absolute speed, direction, sinuosity and measurements of their variations. Mountain and Raper (2001a) also address summaries of point data histories for tailoring the information returned to a user from LBS. They also propose using rapid changes in speed and motion direction to identify breakpoints in lifelines, ultimately in order to sequence lifelines into episodes. Smyth (2001) presents knowledge discovery algorithms based on motion descriptors to design better LBS. These data mining algorithms attempt to assign predefined activities to segments of trajectories by analysing their measurable motion descriptors (such as speed, heading, acceleration). Brillinger et al. (2004) also address motion descriptors for exploratory purposes. They compute speed and predominant motion azimuth of moving elk and deer at various sub-lifeline granularities, analysing diurnal and seasonal motion patterns. Ng (2001) proposes an algorithms to detect outliers from a set of trajectories recorded with surveillance video footage. One of his algorithms detects outliers with respect to their motion speed characteristics. Finally, Laube et al. (2005) use motion descriptors in order to detect predefined motion patterns. They propose ‘detached attribute functions’, a technique to derive motion descriptors of lifelines at arbitrary given times irrespective of fix times.

Fundamental to all these ideas is aggregation and the simplified representation of the lifeline. Having defined a specific measure, the next step in the analytical process is to superimpose structure on the tracking data to comply with that measure’s needs, since each parameter may require different approaches to aggregating the recorded fixes (Wolfer, Madani, Valenti, & Lipp, 2001). The most fundamental analytical context for tracking data is the single fix, with its most important feature being the location in the embedding geography. Referring location to the fix’s encompassing zone, its proximity to a facility, or position on a linear feature is commonplace in mobile applications related to fleet management, vehicle guidance, prisoner monitoring or LBS.

The most obvious and embracing aggregation is the global one, grouping fixes into individual trajectories. Analytical tasks on individual trajectories range from visualisation (Kraak & Koussoulakou, 2004; Zhao, 2003) to the identification of behavioural episodes (Dykes & Mountain, 2003) to quite complex exploratory analysis approaches seeking to identify repetitive patterns (Imfeld, 2000). A standardised measure of the distance or time travelled in-path may furthermore be of interest in investigating the navigational strategy ‘path integration’ or ‘dead reckoning’ respectively, i.e. the integration of walking speed and angular variation along an outbound path in order to construct a homing vector (Merkle, Rost, & Alt).

The most complex level of aggregation is required for the analysis of the lifelines of an entire population of individuals. Occupancy maps (Steiner et al., 2000; Wolfer, Madani, Valenti, & Lipp, 2001) or continuous density surfaces (Dykes & Mountain, 2003; Kwan, 2000) give a summary overview of the recorded space-time activity, not only of populations of moving objects but also of individuals. Brillinger et al. (Brillinger, Preisler, Ager, & Kie, 2004) use a two-dimensional spatial frame of reference to explore a continuous field of movement properties, in their case the diurnal variation of movement azimuth of deer. Kwan et al. (2000) introduced another means of imposing structure to a population of lifelines. Their method for lifeline standardisation aligns all lifelines along a semantic axis, for example ‘work location-home location’ in order to uncover distinctive outliers with respect to the commuting pattern of the whole population. Finally, coming back to geography, the flow metaphor has been used to aggregate the movement of large numbers of tourists with respect to
two-dimensional and three-dimensional mapping and exploratory analysis (Forer, Chen, & Zhao, 2004; Forer & Simmons, 2000).

3 METHODS

This section introduces a set of analytical approaches for lifeline data. Mark (1998, p. 12) defines a geospatial lifeline as a “continuous set of positions occupied in space over some time period. Geospatial lifeline data consists of discrete space-time observations of a geospatial lifeline, describing an individual’s location in geographic space at regular or irregular intervals”. All methods introduced in this section apply for a data set consisting of n individuals, m fixes of the form \((x, y, z, t)\) per lifeline, and a total of \(p\) fixes per population. The individual trajectories need not be equal in length nor start or end at the same time. Assuming a sampling interval given from the tracking device and variable speeds of the moving objects, the step lengths between consecutive fixes are normally not of equal length. Figure 1 illustrates a typical geospatial lifeline connecting an origin \(O\) with a destination \(D\).

![Figure 1: A typical high-resolution geospatial lifeline.](image)

3.1 Lifeline Context Operators

Individual lifelines allow a dynamic perception of movement, providing the detailed data to investigate movement properties at arbitrary instantaneous times along given lifelines. Analysis of lifelines normally involves use of movement descriptors such as speed, acceleration, movement azimuth, and sinuosity. An analogy to spatial analysis may help to illustrate the analytical opportunities inherent in individual lifelines. Spatial variables may be investigated using local, focal, zonal, or global context operators (DeMers, 2002; Tomlin, 1990). Individual lifelines permit modelling of the movement descriptors described above as one-dimensional continuous streams, very much like two-dimensional continuous fields. Thus, in derivation of movement descriptor measures \(d\) along a lifeline, the two-dimensional context operators can be adopted for one-dimensional lifelines (figure 2)

- **instantaneous** (“local”). Derive \(d_i = d(t)\) at an infinitesimal instant in time.

- **interval** (“focal”). Use a moving interval (moving temporal window) to investigating a fixed length segment of the lifeline, computing \(d_{int} = d(t \pm \Delta t)\) respectively.

- **episodal** (“zonal”). Preliminary analysis may result in a partition of the lifeline in delimited episodes, each represented by a movement descriptor \(d_{eps} = d[t_{begin}, ..., t_{end}]\).

- **total** (“global”). Movement descriptors can be computed for whole trajectories as \(d_{tot} = d[t_0, ..., t_n]\). This is the traditional static perspective of lifelines.
Computing instantaneous movement descriptors often requires in practice the inclusion of at least two consecutive fixes or a short segment of the lifeline. Deriving speed and azimuth from given fixes requires at least two consecutive fixes, acceleration three. A measure such as sinuosity is intrinsically an interval measure, requiring a certain number of fixes or a segment of certain length to produce reasonable results. However, the concept of a smoothing moving window of an interval function moving along a lifeline may be adopted for the derivation of all movement descriptors listed above, for example, in the case of noisy or fragmentary data.

The length of the segment for interval analysis can be delimited in various ways, depending on the given tracking data and the research question. The interval can be given as a fixed time interval $d_{int} = d(t \pm \delta t)$. This procedure may include variable numbers of fixes if the fix sampling is irregular. The same is true if the length of the segment is held constant. If the computation of the movement descriptor at that time requires a fixed number of variables, the number of included fixes can be used to define the segment $d_{int} = d(\text{fix}(t) \pm i \text{ fixes})$ (figure 2).

The sampling times at which $d$ is computed, need not to be restricted to the sampling times of the fixes. Many analytical techniques require for instance equal step lengths (Benhamou, 2004), a feature that is hardly found in tracking data that often feature a constant sampling time. Thus, interval functions may be used to interpolate between known fixes (referred to as ‘re-discretisation’ in the biological literature (Claussen, Finkler, & Smith, 1997)) or to aggregate a set of fixes.

### 3.2 Computing Selected Lifeline Context Operators

Just as with spatial-context operators, the variability of possible lifeline-context operators seems to be unlimited. It is not the intention of this work to provide an all-inclusive overview of lifeline context operators. In contrast, this section discusses the potential of lifeline context operators, referring to both well known and new movement descriptors. Whereas some movement descriptors (e.g. speed, acceleration, and movement azimuth) can be
computed using instantaneous, interval or episodal operators, others intrinsically require at least interval operators (e.g., sinuosity).

**Location.** At a first glance it may come as a surprise to classify ‘location’ as a movement descriptor. However, keeping in mind that analysis of movement descriptors may be performed independently from the recorded fixes, lifeline-context operators may be used to interpolate a hypothetical location of an object at a time it had not been fixed in the first instance. Such an interpolation may use a simple interval average or an interval mean of $x$ and $y$ of the involved fixes. Furthermore, using distance-weighted averages it is possible to underline the importance of proximal over distal fixes. An episodal or even total representation of location may refer to the centroid of an episode or the entire trajectory.

**Speed and acceleration.** In kinematics the average velocity $v_{av}$ is defined to be the vector quantity equal to the displacement divided by the time interval (Sears, Zemansky, & Young, 1987). The instantaneous velocity of a moving object at a specific point in the path, or at a specific instant of time, is called the instantaneous velocity. This parameter is defined in magnitude and direction to be the limit approached by the average velocity as the two specified fixes near each other. However, most applications use the notion of speed as the scalar magnitude of the instantaneous velocity, computing it from the two closest fixes to query time $q$. 

$$v = \frac{\delta d}{dt}$$

whereas $\delta d$ represents the distance travelled between the two fixes and $dt$ refers to the elapsed time. Quite similarly, instantaneous acceleration has a scalar and a vector component. However, the simplest computation of instantaneous acceleration considers the change of speed of two consecutive steps $\delta v$, three consecutive fixes respectively.

$$a = \frac{\delta v}{dt}$$

Speed and acceleration are well suited for interval and episodal operators such as average, weighted average or median functions.

**Movement azimuth.** The simplest computation of an instantaneous movement azimuth computes the straight-line vector of two consecutive fixes. Interval, episodal and total azimuth operators must consider a given number of movement azimuth indications, originating from a set of successive steps. Movement azimuth can be conceptualized as a vector or a scalar. The use of the summary vector connecting the first and the last fix of an interval is a straightforward vector representation for azimuth. Directional statistics allow the computation of an average movement azimuth as a scalar (Mardia & Jupp, 2000). The azimuth summary vector and the azimuth mean scalar may substantially vary since the latter does not consider the lengths of the included steps (figure 3b).
**Figure 3:** Interval operator for movement azimuth. (a) Movement azimuth at query time \( q_t \) can be computed as the average direction of all involved step azimuths (\( \overrightarrow{az} \)), or as the azimuth of the summary vector of the interval (\( \overrightarrow{a'z} \)). (b) Whereas the first approach considers only directions, the latter considers also the lengths of the involved vectors, thus the direction of the summary vector of \( (p_1, p_5) \) is not equal \( (p_1, p_5') \).

**Sinuosity.** The terms *tortuosity*, *sinuosity*, *straightness*, and *path entropy* all refer to the degree of windiness of a trajectory. In this paper we list the most frequent concepts and refer to the literature for more detailed information. *Tortuosity* is normally calculated as the ratio represented by the greatest distance between any two points on the trajectory divided by the length of the path (Benhamou, 2004; Claussen, Finkler, & Smith, 1997). *Sinuosity* is normally computed from an equal step length rediscretized path, considering the standard deviation of the directional changes and the rediscretisation step length (Bovet & Benhamou, 1988; Claussen, Finkler, & Smith, 1997). The *straightness index* is given as the ratio of the beeline connector distance and the travelled path length (Benhamou, 2004; Weimerskirch et al., 2002). Other authors use a fractal dimension of trajectories (Bovet & Benhamou, 1988), or an indication for *path entropy* (Guilford, Roberts, Biro, & Rezek, 2004; Roberts, Guilford, Rezek, & Biro, 2004). Even though these sinuosity measures are mainly applied as total operators, in effect they are perfectly suited for interval and episodal operators.

**Navigational displacement.** The navigational displacement refers to the deviation of a trajectory to the direct beeline to the destination (see figure 4). We propose to evaluate the homing direction at every fix and use this azimuth to compute a dynamic measure of navigational displacement (figure 4). The navigational displacement can be assessed as a directed measure ranging from \(-\pi\) to \(\pi\), or as an undirected absolute value ranging form 0 to \(\pi\).

**Approaching rate.** The approaching rate is a measure that describes whether and how intensively a moving object approaches its destination \( D \). We propose an absolute and a standardised approaching rate. The *absolute approaching rate* expresses a speed at which the object approaches \( D \). It is given as the fraction of the approaching distance travelled towards \( D \) during an interval \( d(t \pm \delta t) \) over the temporal extent of the interval \( 2\delta t \). Since it is a speed indication, it is expressed in appropriate units (e.g., m/s). The *standardised approaching rate* expresses the fraction of the distance \( d_a \) actually travelled towards \( D \) and the total distance travelled during that interval \( \delta d \). This measure ranges from 1 (straight home), through 0 (perpendicular to home) to -1 (straight away from home).
First derivatives. A further opportunity we only start to explore is the investigation of first derivatives of movement descriptors. Rates of change of speed, azimuth or sinuosity may for instance be used for algorithms that automatically delimit different movement episodes in lifelines. Different behavioural episodes express different movement trajectories, and peaks in the first derivatives of the movement properties plots may help to delimit such episodes. In figure 6, for example, the final episode of fast and direct movement towards the end is indicated by a sudden drop in the trajectories’ rate of change of azimuth as well as a drop in sinuosity and an increase of speed.

Interval standard deviation of a movement descriptor. Finally, in certain cases one may want to investigate the distribution of a movement property within a moving interval. Using an interval function to derive the movement descriptor, the computation of its distribution introduces a second moving window, assigning, for example, the observed standard deviation of the movement descriptor under study to the central query time of the interval. Such a procedure results in a framed interval-interval measure, with the outer frame delimiting the interval used to compute the standard deviation and the inner frame delimiting the interval used to compute the movement descriptor.
Figure 7: The standard deviation of a movement descriptor computed as an interval operator at query time $q_t$. The rate of change of a movement descriptor can be investigated using a framed interval-interval measure, assessing the standard deviation in an external interval $I_e$ for a focal movement descriptor of a set of internal intervals $I_i$.

3.3 Lifeline Standardisation

Whereas mapping and visualisation of high-resolution geospatial lifelines provides useful first impressions of the richness of tracking data, for many scientific applications it is necessary to impose some degree of structure to the data before analysing it, in other words, to aggregate the movement descriptors in order to perform some type of statistical analysis. Lifelines extend both in space and time. Depending on the research question at hand, one may want to conduct analysis on movement descriptors using different frames of reference. A farmer may be interested in the spatial distribution of the grazing speed of cattle, an ethologist studying animal navigation, in contrast, may want to investigate sinuosity along standardised trajectories as a time-series analysis. Despite the differences in objectives, however, both analytical tasks require the computation of movement descriptors at given points in space and time, the only differences being the standardisation of the trajectory data.

The quantitative analysis and comparison of a set of $n$ trajectories bears a number of methodological challenges. Firstly, trajectories of tracked entities are rarely of equal length. Be it tracked animals, people or vehicles, moving in an unrestricted space intrinsically produce variable solutions with respect to trajectory lengths. Moreover, it is very likely that at least some of the trajectories feature variable starting or ending times, or both. Thus, analytical efforts to compare trajectories adopting a dynamic perspective, that is concurrently moving along a set of $n$ trajectories and comparing movement properties with respect to equal ‘in path times’ or ‘equal travel distances from start’, must first adopt a strategy of standardisation.

Different standardisations may aggregate along spatial (s) or temporal (t) dimensions, or combinations of the two (figure 8).

- **time series analysis (1t)**. Movement descriptors can be analysed in time series analysis, where a single temporal axis is the frame of reference (A).
- **path time (1t) or path distance (1s) analysis**. For some applications it may be relevant to investigate the variance of movement properties with respect to time or path already travelled in a trajectory (A and D), or vice versa, time to go or distance ahead in a trajectory (B and E).
- **equal duration (1t) or equal track (1s) wrapping**. Trajectories of unequal lengths can be ‘wrapped’ so that all trajectories correspond to a given duration or track (C and F). Such an approach is adequate if one wishes
to investigate, for example, whether a set of moving entities all increase speed in the last quarter of their journey.

- **distance analysis (1s).** Trajectories relating to specific locations in space (such as a nest, release site, origin or destination) can be analysed with respect to distance to or from that specific location (G). Such an approach permits statistical analysis of movement properties such as speed, azimuth or sinuosity in distance classes (for example, in box-and-whisker plots).

- **bipolar analysis (1s).** Other trajectories may be aligned along a bipolar set of two locations (H), such as origin-destination, release site-loft, work-home, and summer habitat-winter habitat. The two poles may or may not be fixed in space. Using semantic poles with variable locations, such as ‘home-work’ or variable ‘release sites’ and ‘loft sites’, a preliminary standardisation may align the trajectories for the bipolar analysis.

- **spatial variance analysis (2s).** Having a dense and potentially disperse cloud of fixes from either a single or a set of trajectories one can finally investigate two-dimensional variance of movement descriptors (I) (Guilford, Roberts, Biro, & Rezek, 2004).

**Figure 8:** Temporal and spatial standardisation approaches for lifeline data. The given lifelines have a temporal extent (top left) and a spatial extent (bottom left). Standardisations may be performed along one temporal dimension (t) and two spatial dimensions (s). A to I all represent a form of lifeline standardisation.

### 4 EXAMPLES

This section illustrates the methodology introduced above using data emerging from an interdisciplinary study in behavioural ecology, investigating the navigation of homing pigeons. Given a data set of a total of approximately \( m_n = 30000 \) fixes of \( n = 54 \) individual pigeon flight trajectories (Guilbert, Dennis, & Walker).

#### 4.1 Applying Lifeline Context Operators

**Example 1:** The sinuosity of the trajectories of homing pigeons is considered to express navigational uncertainty: the more variance in path direction, the greater the uncertainty (Roberts, Guilford, Rezek, & Biro, 2004). Following the biological hypothesis that the homing trajectory of a pigeon consists of different navigational episodes, dynamic analysis is preferred over a global approach (Benhamou, 2004; Claussen, Finkler, & Smith, 2004).
A first exploratory approach is to map the sinuosity patterns as they develop along the trajectories (figure 9).

Figure 9: Map of local sinuosity of 54 homing pigeon trajectories, the sampling rate is 1 second; the interval to derive the sinuosity is ±30 seconds.

Example 2: The second example illustrates the analysis of first derivatives of movement descriptors. The figure describes the rate of change of trajectory sinuosity of pigeon 22. For research of the navigational abilities and mechanisms of pigeons, it would be of great interest if one could quantitatively identify an event when an individual pigeon enters the familiar area near the loft (Burt, Holland, & Guilford, 1997). This event may be characterised by a sudden increase in speed and a parallel decrease of path sinuosity. In the mapped track in figure 10, the darker section near the loft end of the path which occurs before the animal heads directly NW may be such an event.
Figure 10: Rate of change of movement sinuosity of pigeon 22. The darker the fixes, the higher is the rate of change. The high rates of change around the last turn in the path may indicate the event, when the pigeon finally figured out where it was and started to fly straight home (light gray). Please note that the fixes close to the release site and the loft have been excluded to remove noise introduced by starting and landing behaviour.

4.2 Applying Lifeline Standardisations

Example 3: Standardisation of homing trajectories emerging from different homing experiments but equal ‘release site-loft configurations’ allows generation of large populations of comparable trajectories. With respect to analysis over temporal dimensions, harmonisation of all trajectories to so that they exhibit equal homing times may be a reasonable procedure. Figure 11 illustrates the navigational displacement of five such harmonised trajectories. Such a time-series analysis may help to identify different navigational states. Pigeon 22 for example expresses very distinct navigational ‘episodes’, illustrated as rather plateau-shaped sections of the plot, with intermittent events of sudden change of the direction (and therefore navigational displacement).
Example 4: Distance to loft is a very relevant reference for the analysis of flight trajectories of homing pigeons. In avian navigation research it may be of interest to quantify differences in the trajectory sinuosity with respect to distance to loft (figure 8G). Figure 12 illustrates this relation. From the figure it is very obvious that the sinuosity of the trajectories is high near the loft (0-500m, 500-1000m) and the release site (5500-6000m, 6000-6500m). This pattern may be explained by specific take-off, position determination and landing behaviours around the loft and the release site.

Example 5: Sinuosity can also be mapped as a two-dimensional spatial field. We used kriging with a rather coarse cell size (100m) and clipped the resulting surface with the convex hull of all fixes in order to minimise edge effects. We are well aware that the data present are not optimal for the interpolation of a continuous surface. However, interpreting with care, we may derive some spatial characteristics of the computed trajectory sinuosity surface. On the one hand we find low sinuosity corridors oriented along the direct beeline connector between

Figure 11: Navigational displacement of 5 pigeons, absolute values, standardised with respect to their arrival at the loft \((t = 0)\). The sampling rate is 5 second; the interval to derive the displacement as an interval measure is \(\pm 30\) seconds.

Figure 12: Trajectory sinuosity of 54 pigeons, aggregated in 13 concentric rings around the loft to express the relation between distance to loft and flight sinuosity.
release site and loft. On the other we see very distinct high sinuosity spots around the release site and the loft. This general pattern correspond with the findings of Guilford et al. (2004).

Figure 13: Trajectory sinuosity mapped spatially using density surfaces basing on kriging. The darker the grid cell, the higher the local sinuosity of this interpolated trend surface.

5 DISCUSSION

This paper has identified a framework for enriching the toolset available for analysing densely sampled lifelines, and has explored ways to implement a growing range of movement parameters within that broad framework. These have covered movement properties of lifelines already deployed in other research, such as speed, movement azimuth, turning angles, or sinuosity, as well as additional measures and standardisations. In section 3 we showed that there are, for many movement descriptors, more than one means of computation, including varying parameters such as interval lengths or weight factors. The choice of these naturally has impacts on the final analysis. For example the length of the analytical interval may be as influential as the dimension of a moving window in other focal operators: an important parameter for analysis but also a major influence as a smoothing filter. This observation is significant in movement research in that often little information is provided about lifeline data models and algorithms employed in computing movement descriptors in a particular study. In order to increase the transparency and the repeatability of analysis of movement trajectories, we suggest that researchers report more detail about how their lifeline descriptors are computed.

Our experience with pigeon flight data suggests that the selection of the algorithms used to compute lifeline context operators needs some care because not all algorithms are suitable for all data models or data-capture procedures. The interplay of the data of interest and the applied context operator algorithms may produce artefacts that are, once introduced, hard to recognize. One example of this is the impact of coarse sampling rates of position fixes on length estimates. They generally result in inaccurate representations of the actual path of a moving object, and ultimately underestimate the distances travelled (Estevez & Christman; Turchin, 1998), leading to an underestimation of derived speed parameters. As another example, the detection of directional change is very sensitive to variable sampling rates along a trajectory. The last curve towards D in figure 1 illustrates this problem. The moving object slows down making a curve. Having a fixed temporal sampling rate thus leads to a finer sampling in the curved area of the trajectory. This finer local sampling rate results in smaller directional changes occurring per unit time, and ultimately introduces an artefact because the curve itself is over sampled and over-partitioned. In that ‘high resolution’ or ‘dense’ lifelines nonetheless cover a range of sampling
frequencies it is also clear that the spatiotemporal resolution of the data comprising a movement trajectory limits the applicability of the various lifeline context operators.

Many authors also identify general difficulties in comparing lifelines of unequal length employing only total trajectory descriptors: Solutions sought include deploying fractal dimensions (Claussen, Finkler, & Smith, 1997; Dicke & Burrough, 1988) or path entropy (Guilford, Roberts, Biro, & Rezek, 2004; Roberts, Guilford, Rezek, & Biro, 2004) as measures which are scale independent and free from the effects of variable sampling rates. The very same idea has also been adopted in order to assess lifeline similarity. Porikli (2004) uses, for example, a trajectory distance metric based on Hidden Markov Models when assessing the similarity of the models representing the individual lifelines. These techniques are helpful but rely on total context (global) operators and thus exclude the detailed dynamic perspective of the line. In this paper, in contrast, we argue that applying standardisations such as wrapping or shifting to equalise the start and end times offers an alternative way to address the problem of unequal lifelines without excluding the dynamic view.

Some issues (such as the navigational displacement or the approaching rate) might be considered as specific to studies of animal navigation or movement. Although lifelines produced by GPS tracked-animals which unrestrictedly move in heterogeneous space may be quite different from humans acting in a normally constrained geography, most elements of the methods presented in this paper are of generic nature and thus of general interest for a wider GIScience audience.

6 CONCLUSIONS & OUTLOOK

Acknowledging the emerging opportunities for the mass analysis of individual movement data, many research fields interested in movement have recently showed a growing interest in dynamic spatio-temporal analysis methods for trajectory data. This eclectic set of disciplines includes geography, GIScience, data base research, animal behaviour research, surveillance and security analysts, transport analysts and market researchers. We have sought to adopt a GIScience perspective in this paper as a contribution the ongoing interdisciplinary discussion at SDH, with its long tradition of spatio-temporal data handling and analysis.

In this paper we have adopted the concept of spatial context operators associated with Tomlin’s map algebra to create a framework for the computation of descriptive measures of lifeline data. We have proposed instantaneous, interval, episodal, and total context operators applicable to a continuous stream of movement descriptors along a trajectory. We have illustrated this conceptual framework by applying it to some well known movement properties such as speed, movement azimuth, sinuosity and additionally propose some new movement descriptors which we believe show value. We have consequently proposed a set of standardisations to harmonise lifelines of differing length or chronology so as to allow consecutive statistical analysis.

This research, however, is at an early stage in which we are involved in a dialectic between data sets and the techniques for their analysis. This dialectic reveals several avenues for further research. The least problematic is broadening our experience by deploying different data sets from different contexts. Existing or on-stream data bases portraying terrestrial animal and human movement represent an immediate opportunity. In this paper we argued that the quantitative analysis of movement is very sensitive to the used data capture procedures, the data models representing the moving object and the algorithms which derive descriptive measures from the trajectories. Using the unlimited variety of lifeline context operators reveals that there is not just one natural way of computing for instance ‘speed’ or ‘movement direction’. Clearly there is room to examine the performance characteristics of various movement indicators relative to different algorithms or parameters for analysis. A related issue is undoubtedly that of granularity and scale in revealing patterns. High resolution lifeline data open
up nice opportunities for sensitivity experiments with systematically varied granularities of lifelines. For instance, assessing the similarity of lifelines basing on some measure of sinuosity or tortuosity, we will perform numerical experiments quantifying the sensitivity of the similarity measures on the used lifeline granularity.

These however are somewhat technical issues. Two more fundamental issues must dominate the ongoing agenda. One is an investigation of how lifeline properties relate to, and perhaps can draw definition from, the underlying (or surrounding) geography and the spatial processes that the individual is engaged in. This is not just a challenge in describing the nature of movement but an important foundation for gaining a greater knowledge of movement as a manifestation of process and structuration. In this respect such movement research is likely to be one of the major driving forces in the field of behavioural ecology research, where geography’s strong spatial awareness can empower behavioural ecology by linking movement data to their geographical context. The quantification of how moving objects, be they animals or people, react to their environment will ultimately help us to better understand their space-time use.

One important aspect of this movement/environment relationship is the degree to which the environment imposes highly differentiated spaces on the individual moving entity (for instance restricting movement to highways or gorges or differentially enabling individuals’ mobility). In the end movement can be represented in many spaces, and we might expect each one might offer different opportunities for the representation of movement.

The second major issue is to consider an extension of the context operators into the spatio-temporal realm. While the analogy with map algebra has proved useful in developing our operators to date, and is adequate in describing the operators presented here, our own operators currently follow a different path. Yet there may be a number of relative movement descriptors, involving two or more lifelines, which involve more complex approaches to spatio-temporal representation and something closer to a three dimensional map algebra to cope with their development. Whether this should be looked at using an essentially vector-based data model, or one structured on space-time voxels (Huisman, 2006), is unclear, but the investigation of these new opportunities beyond the snapshot beckon clearly.

ACKNOWLEDGEMENTS

Patrick Laube’s current work is funded by the Swiss National Science Foundation, grant no. PBZH2-110315. The authors wish to thank David O’Sullivan who provided valuable comments on the presented research and previous drafts of the manuscript.

REFERENCES


Forer, P., & Simmons, D. G. (2000). *Analysing and mapping tourist flows: Tourism, Recreation Research and Education Centre, Lincoln University, NZ.*


Guilbert, Dennis, T., & Walker, M. Unpublished data: University of Auckland.


Ng, R. T. (2001). Detecting outliers from large datasets. In H. J. Miller & J. Han (Eds.), *Geographic data mining and knowledge discovery* (pp. 218-235). London, UK: Taylor & Francis.


Modular System of Simulation-Patterns for a Spatial-Processes Laboratory

Reinhard KÖNIG and Christian BAURIEDEL
Dipl. Ing. Reinhard König, Faculty of Architecture, Chair of Stadtquartiersplanung und Entwerfen, TU Karlsruhe, koenig@entwurfsforschung.de, www.entwurfsforschung.de
Dipl. Ing. Christian Bauriedel, Faculty of Architecture, Chair computer science in architecture, Bauhaus-University Weimar, bauriedel@entwurfsforschung.de

Keywords: Urban Simulation, Cellular automata modeling, Agent-based modeling, Spatial-Processes, Pattern Language.

1. INTRODUCTION

Until now the only possible way to examine the theoretical basis of an urban planning project was to realize this project and explore the effects of such a ‘field experiment’ on its environment in real time. The problem is that at the evaluation of an urban project the results are mostly limited to qualitative conclusions. Long running analyses to produce quantitative statements are rarely. So a falsification of basic planning theories is very difficult, what can be identified as the reason for the existence of so many individual theories in urban planning. The causes for this can be found at the one hand in the methodological difficulties of such long running analyses and on the other hand in the lack of interest of the involved persons and their sceptical attitude on scientific methods at practical planning purposes in general.

In this paper we want to show an alternative way to the examination of particular planning theories by proposing a virtual laboratory for the exploration of spatial processes. For that geosimulation offers relatively new simulation techniques like cellular automata (CA) and multi agent systems (MAS), which allow the modeling of urban theory. By creating scenario models with the help of these components (CA and MAS), the effects of different planning can be observed in a laboratory environment over any period. The scientific findings are based on the interplay of theory and simulation. A scenario model is composed of subordinate simulation parts, in the following referred to as simulation pattern. These patterns are designed on different levels of complexity as regards the functionality and they can involve each other. Such a collection of simulation patterns could be the basis for a geographically object library.

2. URBAN GEOSIMULATION

The term 'geosimulation' comprises generative (or bottom-up) simulation techniques for the modeling of spatial systems. As techniques for the examination of geographical problems cellular automata (CA) and multi-agent systems (MAS) are used. In the context of spatial and urban planning the subarea of geosimulation is of interest which deals with the subject of urban phenomena.
For a detailed introduction in the field of urban geosimulation see Benenson, Torrens [2004, S. 2]. CA and MAS models offer a number of important innovations: They are disaggregated and interactions can be simulated between individuals and building units, the algorithms that they use can be derived directly from theoretical ideas of how cities work and they are often displayed as visual environments, so they are a lot easier to interpret.

There are a large number of separate models, all deal with special themes of urban processes. An actual collection of such models can be found for example in the book ‘Cities and Complexity’ [Batty 2004]. These models tend to be designed as scenario-exploring simulations instead of the traditional goals of simulation as a predictive exercise. Currently no approach for a fully spatially disaggregate land-use transportation and environment model exists [Wegener 2004].

For the development of comprehensive simulation models it is necessary to design each single model in a way that they can be combined with each other. Therefore the first step is to introduce a standardized convention of model-description on the basis of mathematical formalizations.

3. STANDARDISED FORMAT OF MODEL DESCRIPTION

We want to suggest a standardised format for model description suitable for all kinds of models, based on the idea of design patterns, first introduced by Christopher Alexander [1977] in ‘A Pattern Language’. According to Alexander, a single pattern should be described in three parts:  

- **Context** - under what conditions will this solution address this problem?
- **System of forces** - in many ways it is natural to think of this as the "problem" or "goal"
- **Solution** - a configuration that brings the forces into balance or solves the problems presented

Therefore a single entry in a pattern language should have a simple name, a concise description of the problem, a clear solution, and enough information to help the reader understand when this solution is the most appropriate one. It should also note which patterns must be considered beforehand, and which patterns it is natural to consider next. Alexander starts from the most complex main patterns in urban context and links to slave patterns for a more detailed elaboration of a problem. As the collection of simulation patterns shall serve to compound a scenario model, which ones content is not defined, it appears reasonable to change the logical configuration and begin with the description of the basic patterns and their components like special functions and processes. Then the possibility to combine different patterns can help to find answers for more complex questions.

It suggests itself to orientate oneself at the description of a design pattern of the ”Gang of Four”, which follows the following scheme1:

- **Pattern Name and Classification**: Every pattern should have a descriptive and unique name that helps in identifying and referring to it.

---

**Intent:** This section should describe the goal behind the pattern and the reason for using it. Also
Known As: A pattern could have more than one name.

**Motivation:** This section provides a scenario consisting of a problem and a context in which this pattern can be used. By relating the problem and the context, this section shows when this pattern is used.

**Applicability:** This section includes situations in which this pattern is usable. It represents the context part of the pattern.

**Structure:** A graphical representation of the pattern. Class diagrams and Interaction diagrams can be used for this purpose.

**Participants:** A listing of the classes and objects used in this pattern and their roles in the design.

**Collaboration:** Describes how classes and objects used in the pattern interact with each other.

**Consequences:** This section describes the results, side effects, and trade offs caused by using this pattern.

**Implementation:** This section describes the implementation of the pattern, and represents the solution part of the pattern. It provides the techniques used in implementing this pattern, and suggests ways for this implementation.

**Sample Code:** An illustration of how this pattern can be used in a programming language

**Known Uses:** This section includes examples of real usages of this pattern.

**Related Patterns:** This section includes other patterns that have some relation with this pattern, so that they can be used along with this pattern, or instead of this pattern. It also includes the differences this pattern has with similar patterns.

Generally the documentation of a design pattern should provide sufficient information about the problem which the sample treats, about the context of the application and about the suggested solution. Below our description derives from the one of the „Gang of Four“ and is adapted to the requirements of urban geosimulaion.

## 4. EXAMPLES

Our proposal for a documentation of simulation patterns shall be described with two examples, to show finally two experiments for a possible expansion and combination of these patterns.

For the model description tree layers are used: The formal mathematical representation of the most important processes, the textual description which is as generally understandable as possible and refers to analogies of the pattern, as well as the graphical figures, which illustrate the process once more. Every simulation shall be available in the internet as executable program. The structure therefore is organized as follows:

a) Name

b) Input, Interface (also used functions and procedures)

c) Mathematical Formalization

d) Textual Description

e) Figures
4.1 FIRST EXAMPLE: SEGREGATION-PATTERN

As first example we want to cite the well known simulation of a segregation process by location change, which goes back to Schelling [1969] and appears in many current publications about self-organizing systems. We assume the basic concepts of CA and MAS are known. The used example refers to Batty [2005, S. 51-57]:

**a) Name:** Segregation by change of location

**b) Input, Interface**
A given number \( M \) of mobile agents \( P^m, m = 1, 2, ..., M \) are spread randomly over a (40 x 40) cellular raster space (each cell is addressed by the letter \( i \) or \( j \)):

\[
P_j^m = \text{random}(j)
\]

There always some cells have to stay unoccupied, because the agents requires empty space where they can move to. At the initialisation each agent gets assigned one of two random chosen properties. This property means e.g. the preference of a person (agent) for wine or beer:

\[
P_i(0) = \varepsilon_i
\]

The corresponding agent is coloured either black or white.

**c) Mathematical Formalization**

The number of beer drinkers in the local (Moore) neighbourhood is noted formally with \( b_i(t) \) and the number of wine drinkers corresponding with \( w_i(t) \):

\[
n_i^b(t) = \sum_{k \in \Omega \text{ beer}} P_k(t) \quad \text{and} \quad n_i^w(t) = \sum_{k \in \Omega \text{ wine}} P_k(t)
\]

The conditions, if a agent changes its position or stays in place are as follows:

\[
\begin{align*}
\text{if } & \left( \frac{n_i^b(t)}{n_i^w(t) + n_i^b(t)} \right) < \varphi \\
\text{or } & \left( \frac{n_i^w(t)}{n_i^w(t) + n_i^b(t)} \right) < \varphi \\
\text{then } & P_i(t) \rightarrow P_i(t + 1), \quad \text{otherwise } P_i(t + 1) = P_i(t)
\end{align*}
\]

where \( \varphi \) is the percentage threshold of agents with the same properties in the local neighbourhood. If
the number of agents with the same properties falls below this value, the considered agent is not satisfied with its environment and moves away:

\[ P_k^m(t) = \text{move} \left\{ P_j^m(t-1), \text{random}(\Phi), \text{random}(d) \right\} \]

\( \Phi \) is the random heading (given as angle in degree) of the agent and \( d \) is the maximum distance an agent can move. For the program below we have chosen \( \Phi = 360^\circ \) and \( d = 10 \) cells.

d) **Textual Description**

If an individual does not find enough like-minded agents in the (3x3 Moore) neighbourhood, it is dissatisfied and moves to another random chosen place. In a simplified way, one can say birds of a feather flock together.

The results are surprisingly in so far as the percentage of neighbours who are satisfied in a steady state is considerably higher than the tolerance threshold. One can notice that the patterns are far more segregated than the threshold level might imply.

e) **Figures**

![Simulation Figures](image)

Figure e-1: The simulation structure consists of 40 x 40 = 1600 cells. The settings for the threshold \( \varphi \) and the number of agents \( M \) differs in each illustration. From left to right in the upper row: initial spread with \( M = 1000 \) agents, then steady states with

\( \varphi = 50, M = 1100 / \varphi = 30, M = 800, \) and in the lower row:

\( \varphi = 50, M = 1500 / \varphi = 74, M = 1000 / \varphi = 30, M = 1200. \)

f) **Program Parameters**

- The tolerance threshold \( \varphi \).
- The density dependence on the number \( M \) of agents existing in the system.

If the density is too high, the agents can not move to find new configurations with a satisfactory neighbourhood.
The problem is reminiscent of one of those sliding puzzles where you have to move the numbers or letters into a pattern and there is only one spare slot in which to make each movement“ [Batty, 2005, P.56]

g) Core Algorithm
I. check the sort of agents in the neighbourhood
II. compare the number of agents with same property with the threshold
III. move to a new location or remain
IV. begin again with I.

h) Further Elaboration
- Various properties with different tolerance thresholds for each agent [cf. Hegselmann, Flache, 1999].
- Elements like streets where agents can accumulate [cf. König, Bauriedel, 2004].
- Areas with different attractiveness for agents (e.g. a city centre).

i) Integration in superordinate Areas
- Preferred settlement locations of urban inhabitants.
- Trigger for changes of the place of residence.
- Formations of ghettos and mono-functional urban areas.

j) URL of the Program
http://www.entwurfsforschung.de/Strukturfor/delphi/delphiF.htm#Segregation02

k) References

4.2 Second Example: Economies of Scale-Pattern
The second example shows a program we have developed to represent the consequences of increasing returns to scale for the catchment areas of competing markets.

a) Name: Economies of Scale

b) Input, Interface
The global parameters production costs D, travel expenses F and the factor for returns to scale S. For the simulation we assume a homogenous distribution of potential customers over the cellular space. In our example we choose random positions for ten competing Markets M (assume they all sell the same product).
c) **Mathematical Formalization**

In the beginning the bid price $P$ is equal in all markets and the catchment areas result only from the distance $d$ to the next Market (looks like a raster voronoi-diagram), because in this case the relative Prices $N$ at a location $i$ depend on the travel expenses $F$ only.

$$N_i(t) = P_j(t) + F_i$$

The travel expenses $F$ results from the global parameter for the costs $K$ (e.g. prize of petrol for a distance unit) and the distance $d_{ij}$ between a customer (cell) and the respective market.

$$F_i = d_{ij} \cdot K$$

A customer is assigned to the catchment area $G$ of market $M_j$ where the relative price $N$ at location $i$ of the customer is the lowest.

$$G_i \leftarrow \min_j N_i$$

The bid price of a market $m$ at $t+1$ is calculated from the demand $\sum G_{i,m}$ (A size of the market’s catchment area), the global production costs $D$ and the factor for the returns to scale $S$:

$$P_{i}(t+1) = \frac{\sum_{i,m} G_{i,m}(t)}{\left( \sum_{i,m} G_{i,m}(t) \right)^S} \cdot D$$

Since the bid price $P$ is now different in each market (in the case of varying extensions of the catchment areas and increasing returns to scale), this effects on the calculation of the relative price $N_i$ and on the catchment area $G$.

d) **Textual Description**

The model shows, how increasing returns to scale effects on the catchment area of a market (or trader/manufacturer). At the production of a product particular costs arise. If a bigger amount of products can be sold (in the model this depends on the size of the different coloured catchment areas), the production costs of one product go down (economies of scale). As a result it is worth for a customer to cover a longer distance to buy a product at a market where a product is sold cheaper. On the other hand this causes the travel expenses to rise. A growing demand in a market leads to a further price-reduction. In the resulting competition for catchment areas some markets can assert one against others and monopolise locally. At the border of the catchment areas the travel expenses and the production costs in the market with the favourable price balance each other. The effect of increasing returns to scale increases with expensive products. The process is not simply reversible.
e) Figures

Figure e-2: From above left to below right the illustrations 1-3 show the expansion or reduction of the separated catchment areas of the particular markets by increasing the parameter for the returns to scale. After a supplier has putted all rivals out of the market he monopolised (yellow). Now it is very hard for new competing suppliers to produce at a lower price to enter the market once again, because in the beginning a supplier has just a small catchment area to underbid the monopolist (illustration below right).

f) Program Parameters

- production costs $D$, travel expenses $F$ and the factor for possible returns to scale $S$
- increasing the factor for returns to scale expanse or reduce the catchment areas of a market
- increasing the parameter for the travel expenses per distance unit reduces the effects of increasing returns to scale
- Expensive products (parameter production costs) strengthen the effects of increasing returns to scale

Core Algorithm

I. calculate the local prices $N$ for each cell $i$
II. define the assignment to a catchment area $G$ for each cell $i$
III. calculate the bid price $P$ in the markets $M_j$
IV. go to I. and start again

h) Further Elaboration

- Analyse in more detail the relationship between catchment areas and the factor for returns to scale
• Connect the Economies of Scale-Pattern with the Social Gravity-Pattern (flow of traffic) and the von Thünen-Pattern (both not documented in this paper – see internet reference)
• Include the effect of savings by run multiple errands through concentrating different markets at a location
• Expand the model so, that the markets can try to expand their catchment area by changing their locations. A market can try to find a better poison assuming that the other markets stay where they are or that the competing markets also try to move to a more promising location. Here we enter the area of game theory and can bring the model in analogy to questions like the prisoner’s dilemma (cooperate or act egoistic).

i) **Integration in superordinate Area**

land market

j) **URL of the Program**

http://www.entwurfsforschung.de/Strukturfor/delphi/delphiF.htm#Skalenertr

k) **References**


Niveauproduktionsfunktion, Skalenerträge

4.3 Experimentation

Based on the both presented simulation patterns, we want to show how they can be expanded and combined for further research. In a first experiment the Segregation pattern shall be extended by central places (markets) to which the agents want to be so near as possible. Thereby it comes to a competition for the most central places.

to a) **Name**: Segregation with Central Places

to b) **Input, Interface**

Segregation pattern

The initial conditions stay the same, only the definition of one central place (ore more places) at cell m was added.

to c) **Mathematical Formalization**

The rules for the segregation process remain the same. We just add a calculation for the agents to change their position, if there is a better (more central) location to occupy:

\[
\text{if } d_{i,m} > d_{j,m} \text{ then } P_i(t) \rightarrow P_i(t+1), \text{ otherwise } P_i(t+1) = P_i(t)
\]

If the distance \(d_{i,m}\) from cell \(i\) to the next central place at cell \(m\) is grater than the one at the compared location \(d_{i,m}\), then the corresponding agent changes its position from \(i\) to \(j\) (cf. hill climbing algorithm and potential field).

to d) **Textual Description**
Apart from the requirement for a minimum of similar agents in the neighbourhood, the agents compete for the most central places in the cellular space. As a result the segregation structure concentrates inside a circle around the central place (fig. e-3). In the case of several central places from a particular tolerance threshold only one sort of agents accumulate to a market (third illustration at fig. e-4). So we can detect a higher-ranking segregation to centers.

to e) Figures

Figure e-3: The simulation structure consists of 40 x 40 = 1600 cells and $M = 800$ agents in each version. At all six illustrations the central place is in the middle of the cellular field. The values for the segregation threshold vary at the illustrations from above left to below right: / $\phi = 0$ / $\phi = 0.125$ / $\phi = 0.25$ / $\phi = 0.375$ / $\phi = 0.5$ / $\phi = 0.625$. For the last value of $\phi$ there is hardly a stable structure. For values of $\phi = 0.75$ or greater, there are no stable structures at all and the agents move chaotically over the cellular field.

Figure e-4: The simulation structure consists of 40 x 40 = 1600 cells and $M = 1000$ agents in the four versions. Ten central places were distributed randomly over the field (red marked cells). The segregation threshold was chosen from left to right with:
$\phi = 0$ / $\phi = 0.25$ / $\phi = 0.5$ / $\phi = 0.75$. It is remarkable that at the third illustration from the right ($\phi = 0.5$), the markets are surrounded each with only one sort of agents.
to g) **Core Algorithm**

I. check the sort of agents in the neighbourhood  
II. compare the number of agents with same property with the threshold  
III. move to a new location or remain  
IV. if remain, check for a more central place  
V. if there is a better location, move to it else remain  
VI. begin again with I.

to h) **Further Elaboration**

- Introduction of agents with different priorities to occupy a location (e.g. by their economic potential)

to j) **URL of the Program**

[http://www.entwurfsforschung.de/Strukturfor/delphi/delphiF.htm#Segregation03](http://www.entwurfsforschung.de/Strukturfor/delphi/delphiF.htm#Segregation03)

The next experiment brings together the both simulation patterns Segregation and Economies of Scale. This enables the investigation of the agent’s site selection by the combined conditions of segregation and the search for a location as near to a market as possible. To avoid a repeat of the description of the both simulation patterns only the way how the patterns are brought together and the functions are added is described in the following.

to a) **Name**: Economies of Scale with Segregation  

to b) **Input, Interface**

Segregation pattern and Economies of Scale pattern  

The cellular space of the Economies of Scale pattern with the ten competing markets is initialized first. Afterwards the agents are distributed randomly over the field. The ten markets represent the central places for the agents now.

to c) **Mathematical Formalization**

As first step the bid prices $P_m$ of the markets $m$, the local prices $N_i$ at each location $i$ (for each cell) and the resulting catchment areas $\sum G_{i,m}$ were calculated. The demand $C$ of potential customers is no longer distributed equally over the cellular space now but is represented by the mobile agents. An agent $A_i$ is stored as customer to this market, on which catchment area it is located:

$$C_m = \sum_{i} A_i \in G_{i,m}$$

The bid price of a market at $t+1$ is calculated from the demand $C_m$ in a market, the global production costs $D$ and the factor for the returns to scale $S$:

$$P_i(t+1) = C_m / C_m^S \ast D$$

Then each agent tests its neighbourhood like it is described at the Segregation pattern. The search for the most advantageous place is no longer orientated on the distance to the centers but on the local prices $N_i$.

*if* $N_{im} > N_{jm}$  
*then* $A_i(t) \rightarrow A_i(t+1)$,  
*otherwise* $A_i(t+1) = A_i(t)$
The identifier for an agent was changed to $A$ to avoid confusion with the bid price $P$.

to d) **Textual Description**

By the combination of the both simulation patterns the competition of the separate markets for customers can be better investigated than with the *Economies of Scale* pattern only. The customers are not distributed equally now but they can choose their location on the basis of several criterions. As well as a location near the centre another criterion is the neighbourhood tolerance and the quality of a location (cell), defined by the catchment areas of the markets, which are related to the bid prices at the several markets. Altogether the customer’s mobility has a strengthening effect on the monopolizing as a result of increasing returns to scale (different rows in fig. e-5).

to e) **Figures**

![Figure e-5](image-url)

Figure e-5: The simulation structure consists of $40 \times 40 = 1600$ cells and $M = 800$ agents in each illustration. Ten central places were distributed randomly over the field (red marked cells). In the upper row the agents remain at their start position. Just the factor $S$ for the returns to scale has an effect on the catchment areas. In the middle row the behaviour of the agents is added to compete for the most central places. In the bottom row the segregation with the threshold $\varphi = 0.5$ was activated. The parameters for the separate illustrations from left to right were chosen as follows:

**upper row:**
- $S = 1, \varphi = 0$
- $S = 1.02, \varphi = 0$
- $S = 1.05, \varphi = 0$
- $S = 1.15, \varphi = 0$

**middle row:**
- $S = 1, \varphi = 0$
- $S = 1.02, \varphi = 0$
- $S = 1.05, \varphi = 0$
- $S = 1.08, \varphi = 0$

**bottom row:**
- $S = 1, \varphi = 0.5$
- $S = 1.02, \varphi = 0.5$
- $S = 1.05, \varphi = 0.5$
- $S = 1.08, \varphi = 0.5$

to g) **Core Algorithm**
I. calculate the local prices $N$ for each cell $i$
II. define the assignment to a catchment area $G$ for each cell $i$
III. calculate the bid price $P$ in the markets $M_j$
VII. check for each agent $A$ the sort of agents in its neighbourhood
VIII. compare the number of agents (with same property) with the threshold
IX. move agent to a new location or remain
X. if remain, the agent checks for a more central place
XI. move agent to the better location or remain
IV. go to I. and start again

to j) URL of the Program

http://www.entwurfsforschung.de/Strukturfor/delphi/delphiF.htm#Segregation04

We have experimented with some more Simulations that are available via the internet at
http://www.entwurfsforschung.de/Strukturfor/delphi/delphi.htm but the descriptions are still in German language. Currently we are working on the documentation and combinations of further simulation patterns. We hope to present the results on the ICA Workshop in July.

5. INTENTION OF THE PROPOSED LABORATORY

The indirect object of the presented efforts about the collection of simulation patterns is to establish a virtual planning laboratory. But this aim is in a distant future, so in an indirect step at least the material shall be collected and compared, which exists about spatial simulations until now. For this a standardized convention for model-description is necessary in such a way we have proposed it as simulation pattern above. On such a basis the programming of scenario models shall be simplified with the help of a geographically object library, like it is outlined under the next point. Such an expandable and free accessible library would be very valuable for further exploration about urban dynamics. In addition to that the library could be used as material source for appropriate lectures.

The direct object of a simulation pattern can be seen in the description of a solution for a particular class of questions. Further benefits results from the name of each pattern. This simplifies the discussion under the developers, because one can talk in an abstract way about a simulation structure and the pro and contra of a solution. Design patterns are first independent from a concrete programming language.

6. PROGRAMMING

As mentioned before, it shall be possible to use the simulation patterns as basis for the development and documentation of an object library for geographically based automata simulations. For this the object oriented programming (OOP) paradigm offers the possibility of the compilation of a modular and extendable library using ‘classes’ or ‘components’. Popular OOP languages are Java, VisualBasic.net, Delphi, C++ und C#. An interesting project in this context is “Obeus” [Benenson, Aronovich, Noam, 2005], which build on the Microsoft.NET Framework with a view to simplify the simulation programming by providing rudimentary classes for cells and agents. The properties and behaviour (methods) are adaptable by using the C# language. The .NET Framework makes it possible to combine classes, which was produced with different OOP languages.
7. CONCLUSION

- A modular system of simulation patterns would enhance the scientific basis of the examination of spatial processes.
- Creation of a standardized research method.
- Examination of theoretical models in a virtual laboratory with the help of simulations.
- Basis for an object library for geographically based automata simulations.
- Simulation patterns can be the basis for future development of user-friendly applications to support the planning practice.
- I am looking for people interested in cooperation for a simulation pattern collection.

REFERENCES

The programs the screenshots are taken from are available at:
http://www.entwurfsforschung.de/Strukturfor/delphi/delphi.htm


The Dynamic Geometry of Geographical Vector Agents

Yasser Hammam, Antoni Moore, Peter Whigham & Claire Freeman*

Spatial Information Research Centre
Information Science Department
(*Geography Department)
University of Otago, Dunedin, New Zealand
Phone: +64 3-479-8301 Fax: +64 3-479-8311
Email: {yhammam}, {amoore}, {pwhigham}@infoscience.otago.ac.nz

ABSTRACT

This paper introduces vector agents (VA), an attempt to impose a systematic framework on the geometric element of Torrens and Benenson’s Geographic Automata System (GAS). Both schemes use vector geometry as an antidote to the geographically unrealistic regular cell cellular automata (CA). The work reported here explores the properties of irregular and dynamic VAs in particular, a subclass of geometry not hitherto covered in detail in a spatial agent modelling context. This is surprising, given the amount of real-world phenomena that could be effectively expressed spatiotemporally through such a geometry. An exploration of the interplay between midpoint, edge and vertex displacement in evolving various geometries found that a wide range of real-world objects could potentially be modelled by the various sizes, change rates and complexities possessed by the simulated objects derived. It is proposed that vector agents offer the ability to explicitly control geometric form through the alteration of simple parameters. Furthermore, these specialised objects potentially have the ability to perform a range of generalisation and transformation operations on themselves to reach their goals; this is enabled by them having their own flexible vector geometry.

1.0 INTRODUCTION

This paper introduces a generic spatial modelling framework populated by objects called vector agents (VA). VAs are based on an irregular vector data structure, behave according to an agent-oriented paradigm and are predominantly driven by geometry. They have developed out of a need to create a systematic basis for the geometric components of Geographic Automata Systems (GAS), which were outlined by Torrens and Benenson (2005). As with GAS, it is proposed that the vector agent is more spatially realistic than Cellular Automata (CA). Although a central property of GAS is the use of irregular vector objects as automata, the geometry introduced in the GAS paper examples remained invariant throughout the time frame of a given simulation. Modelling objects with dynamic geometry is briefly referred to but not expanded upon. Recognising that geometry may have an important, even leading role in these systems, the vector agent is detailed here (GAS was devised out of a desire to let the “geography” lead the simulation; geometry is considered a subset of that, in its widest sense).

By way of precedent, the use of Cellular Automata (CA) for spatial modelling is well established with complex spatial phenomena such as city growth and differentiation (Batty, 2000, 2001; Torrens and O’Sullivan, 2000). However, the representation of space as a collection of regular square cells is regarded as a limiting assumption in a spatial simulation domain (Benenson and Torrens, 2004a, b). Wahle et al (2001) have designed and run CA so that it adopts the form of a real object, in their case a linear CA to represent a road network for modelling traffic flow. Initial research towards an irregular CA have used Voronoi diagrams for the tessellation of cells.
instead of the fixed and regular geometry of conventional CA (Shi and Pang, 2000). Figure 1 shows these fixed regular and irregular (real-world data) geometric structures. However, these rigid configurations have simplistic notions of neighbourhood, based on topology (White and Engelen, 2000). Various alternative approaches have been investigated, including Delaunay triangle links (Semboloni, 2000) and planar graphs (O'Sullivan, 2000, 2001) being used as the basis for cell neighbourhoods.

Since the introduction of the spatial agent (Rodrigues et al., 1998; Rodrigues, 1999; Sanders et al, 1997), the regular cell framework has been routinely used as the spatial basis for agent systems, specifically Multi Agent Systems (MAS) (Barros, 2003; Batty et al., 2003; Haklay et al., 2001). The agent is “…an autonomous decision-making system…” (Wooldridge, 1997) “…perceiving its environment through sensors and acting upon that environment through effectors.” (Russell and Norvig, 1995). Agents can roam around the cell structure, reacting to stimuli and in so doing making the CA a more mobile and flexible model (Torrens and Benenson, 2005). Seeking to escape the rigidity of the fixed cell, the term automata has recently been employed independently to describe autonomous objects in the spatial simulation domain that can be irregular and also non-fixed (Benenson and Torrens, 2004b). The prime example is the Object-Based Environment for Urban Simulation (OBEUS), a pioneering example of a GAS (Torrens and Benenson, 2005). This system is populated with Geographic Automata (GA), which, like CA has states and neighbourhoods and rules for the changing of state based on input from the neighbourhood. GA are distinguished by having the means to explicitly store its own geometry as well as rules for the movement of that geometry and rules for the alteration of neighbourhood effects. Two types of agent were distinguished – the fixed agent (e.g. land parcel) and the non-fixed agent (e.g. social “actors”) of no physical manifestation.

Objects with dynamic geometry are prevalent in the real world (e.g. natural habitats and urban areas) and are subject to irregular growth. However, these concepts have not been addressed by spatial automata and agents. In Torrens and Benenson’s paper (2005), the existence of objects that change their geometric form over time is briefly mentioned: “…there are instances in which georeferencing is dynamic for the geographic automata that represent infrastructure objects, for example when land parcel objects are sub-divided during simulation”. Moving towards this state of being are the aforementioned non-fixed objects, which point to a vector object but have no geometry in themselves (“indirect georeferencing”).

As a possible answer to the question “how can agent-based systems model these dynamic objects and their interaction (in terms of both state and geometry) in a more realistic fashion?”, generic vector agents are introduced (an early non-generic prototype of vector agents specific to von Thunen’s model of agricultural land use is outlined in Hammam et al., 2004). Moving forward from agents with fixed vector boundaries, this paper will focus on the process of generating non-deterministic vector agents through the application of irregular fractals. Section 2 will start with a discussion of the full scope of a vector agent. It will then detail the fractal algorithm as well as discuss the way in which vector boundary growth (e.g. midpoint displacement) can be controlled in form (spatial) and rate (temporal) to model a range of real-world spatio-temporal phenomena. Experimental results from several combinations of geometric control parameters (e.g. growth with 50% midpoint displacement, 50% vertex displacement) are presented, discussed and linked visually to real urban examples in Section 3. Finally, section 4 draws some conclusions and describes future directions for this ongoing research.

2.0 THE VECTOR AGENT

2.1 Broad properties of the vector agent

A vector agent is goal-oriented, adaptable (characteristics of what Luck et al 2003 call an autonomous agent), physically defined by a Euclidian geometry and able to change its own shape while interacting with other agents in its neighbourhood using a set of rules (though the latter role will not be tested in this paper). More specifically, vector agents can:

- Represent any discrete geographic phenomena entity through an irregular (or regular) vector data structure: the agent therefore has the advantage of being more realistic and flexible for representing real-world features, such as buildings, roads …etc., which is not facilitated through generalised square cells or Voronoi diagrams. The irregular vector agent will be explored in this paper.

- The irregular vector agent may move bodily and is either based on a real-world object or is born with a nondeterministic shape boundary which subsequently changes: this advances the capability of an agent to construct a algorithm-based shape with increasing structural complexity, rather than assign a new entity to an object with fixed boundary (Torrens and Benenson, 2005), or allocate itself in space with a static shape to just interact spatially with other agents (Rodrigues, 1999). This can be achieved using...
different operators for assigning and changing the object boundary such as midpoint displacement, edge displacement and vertex displacement. The vector-agent provides a flexible mechanism for meeting a certain threshold and satisfying the agent’s goal using more complex operations such as a generalisation technique (for achieving a desired fractal dimension), or split (to meet a certain size) – see section 2.2.

- The entity is abstracted so that it is able to define its own location in space: the agent can allocate itself randomly or via attraction and repulsion forces generated from the surrounding environment. This can overcome the limitations of a restricted neighbourhood exhibited in CA.

- Represent dynamic neighbourhoods in which both extent and nature of relations can change: the vector agent should have advanced interaction capabilities with variant topological relations. Representing dynamic neighbourhoods and rules of neighbourhood change are also a feature of GAS (Torrens and Benenson, 2005), but are not within the scope of this paper.

- Since the agent is an abstraction of a real-world entity in the simulation domain, the agent’s goals are consequently abstraction of the entity’s properties: the agent can therefore maintain its identity derived by that entity’s interactive behaviour in a spatial domain (not constrained by a regular geometry, such as with CA).

2.2 The processes of geometric change

Prior to considering the implementation of an irregular geometry agent class (section 3), it is necessary to know the specific characteristics of the point, line and particularly polygon geometry of a vector agent within the spatial simulation domain, and how they can change.

Having put forward an argument for dynamic geometry, a taxonomy of dynamic polygons emerges (reflected in the right-most part of Figure 1):

- the polygon that is invariant (Fig 1.1, 1.2, 1.3)
- the polygon that moves (or can perform other transformations on itself) but does not change its shape (Fig 1.3b)
- the polygon that changes its shape but does not move otherwise (e.g. translation) (Fig 1.4)
- the polygon that both changes its shape and can perform transformations on itself (Fig 1.4b)

![Figure 1: Various methods of discrete spatial modelling, arranged on a continuum from regular to irregular geometry, and static to dynamic models. The form of vector agent discussed in this paper is type 4.](image)

“Changing shape” is taken to mean the local, irregular or random changing of coordinates as a result of some stimuli. They can also include cartographic generalisation processes such as aggregation, reduction etc (a good overview of generalisation processes may be found in Jones, 1997). It is proposed that the generalisation process of fractal enhancement has the most scope for generating widely domain-applicable objects and will be detailed in the next section, as the basis for a stochastic vector agent generation method.
Transformation processes are global and include affine transformations such as translation (e.g. bodily moving a polygon), rotation, shearing etc. and analytical transforms occurring due to projection change.

2.3 Shape Construction and Evolution in a System Environment

In many cases real-world phenomena exhibits fractal characteristics, such as self-similarity (Peitgen et al., 1992). As well as the classic coastline example, urban patterns and landscape features, for example, follow a self-similar geometry to some degree. However, some man-made objects, such as buildings, have no apparent self-similarity, though they do in fact possess some statistical fractional sense when magnified. In general, real-world geographic entities that have fractal forms also have irregular boundaries, therefore an irregular fractal geometry. Regular fractal geometry is rarely, if ever, seen in the real world at the geographical scale, though there are a number of well-known micro-scale object examples that can be constructed with simple rules that induce a systematic physical form (e.g. the Von Koch snowflake, a fern frond). Therefore, methods for generating models of shapes with prescribed fractal dimension with exact self-similarity are not perceived as realistic models at the geographic scale, due to their lack of randomness (Peitgen et al., 1992). Something more than just the self-similarity fractional law for abstracting objects is needed to model irregularity in nature.

2.3.1 Midpoint displacement through Brownian Motion

A randomization process must be utilized to evolve more natural and realistic shapes. One such dynamic process is Brownian Motion (BM), which characterises the random movement of microscopic particles (Rucc, 1994). Pure BM can be defined as tracing out the total random walk path travelled by a point in the plane in appropriate units of time. For the irregular vector agent simulation, the randomly-parameterised alteration of vector geometry form occurs according to BM (e.g. in boundary growth and contraction). The most popular way to computationally model BM to alter line geometry is called random midpoint displacement (Kenkel and Walker, 1996). This method has been used in computer graphic simulation and 3D cartography to derive artificial terrains through irregular and stochastic fractals (i.e. natural surfaces are also self-similar) (Goodchild and Mark, 1987). Simple BM has been generalised so that the fractal dimension for various geographic phenomena can be used to derive artificial models of the same phenomena.

Considering a single line segment as an initiator, recursive subdivision by midpoint displacement occurs, with the displacement made perpendicular to the line. A generalised algorithm for simple displacement is given by the formula in Equation 1 (adapted from Laurini and Thompson, 1992):

\[ P_{\text{new}} = 0.5 (P_1 + P_2) + \mu \sigma_0 2^{\text{th}} \]

where \( P_1 \) and \( P_2 \) are the start and end points of the line segment being subdivided represented in vector form. The second group of terms (governing the amount of perpendicular displacement) includes \( \mu \), which is a random number from a Gaussian distribution; \( \sigma_0 \) is the standard deviation of that Gaussian curve (equal to 1; mean = 0), \( l \) is the level of recursivity, and \( h \) is the Hurst exponent specifying the roughness of an object (Laurini and Thompson, 1992; Voss, 1988). The fractal dimension \( D \) is equal to \( (2 - h) \). With this relation the fractal dimension \( D \) of regular Brownian motion (\( h = 0.5 \)) is 1.5. When \( h < 0.5 \) a shape is considered rough and when \( h > 0.5 \) the shape is assumed to be smooth.

Theoretically, to enable the creation of more varied irregular objects than through perpendicular midpoint displacement alone, the midpoint assumption may be relaxed, where displacement occurs relative to a point situated on the line but nearer to one end point than the other (expressed by a value \( r \), where \( r = 0 \) indicates that the base point for displacement is at \( P_1 \) and \( r = 1 \) indicating \( P_2 \)). Also, the requirement for perpendicular displacement may be waived, with displacement angle \( \alpha \) from the original line allowed to vary from \(-90^\circ \leq \alpha \leq 90^\circ \). This non-midpoint non-perpendicular displacement is shown in Figures 2c – f and expressed in Equation 2 and simplified in Equation 3.

\[ P_{\text{new}} = 0.5(P_1 + P_2) + (r - 0.5)(P_2 - P_1) + \mu \sigma_0 2^{\text{th}} \]

\[ P_{\text{new}} = (1 - r)P_1 + rP_2 + \mu \sigma_0 2^{\text{th}} \]

where \((r - 0.5)(P_2 - P_1)\) in Equation 2 is the adjustment along the line relative to the midpoint and the displacement \( \mu \sigma_0 2^{\text{th}} \) is on a bearing \( \alpha \) from the base point on the line (with the perpendicular at 0°).

2.3.2 Vertex and edge displacement
In summary, the Brownian Motion formula will be extended as a basis for constructing the irregular geometry in the vector agent model. However, modifications that allow the geometric shape more freedom to evolve stochastically have been made. Firstly, vertex displacement adjusts a single point by a random amount (without the parameters needed for fractal modelling) on a bearing $\beta$.

$$P = P + \mu \sigma_0$$  

(4)

Edge displacement applies equation 4 to two consecutive existing points, ensuring that the magnitude and direction of displacement is equal for both on a bearing $\gamma$. Edge and vertex displacement can be defined based on two connected points in a shape. Possible actions include: no action occurs for both points; one or other of the points is moved (vertex displacement); or both points are moved (edge displacement, but the new edge is not necessarily parallel to the old one).

All three algorithms (applied to randomly chosen points / pairs of points) will be illustrated in section 3.

3.0 MODEL IMPLEMENTATION AND EXPERIMENTAL RESULTS

This section describes a testing environment for the dynamic geometric framework discussed in section 2. The main aim here is to create a simulation, which involves a spatial environment with elements positioned in it.

3.1 Model Elements

The main focus of our simulation model at this stage is to demonstrate that a typical vector agent is capable of initiating its own shape, subsequently changing over time with a number of modification parameters using different growth and contraction operators with various probabilities. The simulation is therefore composed of the following:

- a continuous vector space (coordinate space) with predefined $x$, $y$ coordinate limits. This is a passive or static object that will never change its state.
- The shape class, a vector agent, which can search for unoccupied space, fulfilling its preference by interacting with other agents and enquiring about availability of empty space.
- The neighbourhood, a rule-based class that the shape agents use to establish the current nature of interaction with other agents. This is in the form of the topological relations addressed in section 3.2.
- The shape geometric behaviour class, containing a choice of three different algorithms to be implemented by the vector agent at each iteration of the simulation. Initially, the geometric form of the vector agent is a random point allocation. Then, a second point is placed at a random angle and distance from the first. New point displacement by Brownian Motion is used to initiate the third point and close the polygon. The three algorithms can be summarised as follows:
  - Splitting one of the shape edges with a new point generated by the Brownian Motion algorithm as described in Section 2.3.1. The point being displaced is assigned randomly along the edge – i.e. in equation 3, $r$ is the uniform random number controlling this, varying between 0 and 1 (Figure 2. e, f). The displacement of this point is allocated into its new position with a random displacement angle $\alpha$ made relative to the edge split location.
  - Moving a whole edge to new coordinates at a uniform random distance and angle (Figure 2. g, h).
  - Displacing one of the shape vertices by a uniform random distance and angle (Figure 2. i, j).

The latter two algorithms have been included to provide scope for producing shapes that are more akin to those found in the real world. Use of the BM algorithm alone would produce star-shaped objects, which represent only a subset of real-world objects (e.g. the results of radial development). Each of the three geometry change algorithms will be assigned probabilities of use, with different probabilistic combinations leading to various object characteristics. An object with an excessively concave boundary (i.e. the star shape) would be derived from new point displacement being given a probability of close to 1 (the new points get displaced but the existing points do not move). A more compact object would eventuate from a high probability value being assigned to vertex displacement (each existing point has the potential to be displaced). In a generic sense, the system must theoretically be able to deal with any phenomenon in the simulation domain. Therefore, the model was elaborated in such a way that the shape change algorithms are as simple and complete as possible to achieve such variability.
Figure 2: How a vector agent shape is born and evolves in the spatial simulation domain: (a) initialising by random point, (b) allocating second point by random displacement, (c, d) applying the random new point displacement and accomplishing closed polygon, (e, f) choosing any edge randomly and applying the new point displacement, (g, h) edge displacement, (i, j) vertex displacement.

3.2 Simulation Results

The simulation starts by initialising the desired number of agents. Every agent starts by exploring any available empty space and allocating itself randomly as a point (i.e. random \( x, y \)), then extending itself to a line. Subsequent points are generated based on the BM algorithm to initially achieve a closed geometry (polygonal shape) (Figure 2. a, b, c, d), subsequently increasing that polygon’s complexity. Thus, the vector agent starts to evolve, geometrically driven by one of the previous algorithmic operators in the shape behaviour class (Figure 3). Note that by default a vector agent is initiated with equal probability of applying each operator (\( p = 0.33 \) for each). Also present, but not tested in this paper, is the ability to modify these probabilities when a vector agent is exposed to relevant neighbourhood stimuli (other vector agents). Therefore an agent may be able to change its interactive behaviour based on perception of the environment.

Figure 3: Schematic representation of agent behaviour algorithm
Figure 4 shows a typical simulation run for 600 time steps. Four agents were initialised, with no restriction on shape size. It takes 3 time steps for the vector agent to achieve closed polygons, and is therefore capable of evolving a simple shape. As the simulation progresses, the agents individually conduct one of the algorithmic operators at each timestep. By time 400 a hole has formed in one of the largest polygons, mirroring many real-world geographic phenomena, such as cities. By increasing the shape complexity, this type of geometry form is likely to occur, arising from the intersection of some edges while deploying different operators. Since the displacement operators only operate on the shape’s outer boundary, the inner boundary remains constant, as can be seen in subsequent time steps.

As well as the pseudo-natural variation, systematic geometry can be observed (t50, where parallel edge displacement has manifested itself as straight lines and right angles in the case of one polygon). Twelve different combinations of algorithmic operators have been applied with an algorithm probability (p) being varied for each experiment. Fractal measurement for the different combinations demonstrates that vector agents can successfully produce a wide range of desired shapes according to application domain demand. For each combination, the simulation was run to 1000 time steps, enough simulation time for all significant geometric behaviour to be revealed. Figure 5 graphs the average fractal dimension (D) and agent shape size over the 1000 time steps for all algorithm probabilities.
Figure 5: Fractal dimension and the trends over time of shape size generated by applying different operations with different probability (p) as simulation result for first 1000 time steps
The first three experiments (Fig 5a - c) were run with each of the three algorithms in turn. Predictably, midpoint displacement varies closely but randomly around the input roughness $h = 0.5$ (fractal dimension $D = 1.5$), though it took nearly 200 iterations to reach this equilibrium state. The edge displacement algorithm quickly reaches a stable plateau of around $D = 1.65$, the ability to displace two points at a time clearly increasing the complexity of the shape. Vertex displacement also stabilises swiftly so that $D$ is approximately 1.43. The lower fractal dimension indicates the retention of a simpler polygon; unlike midpoint displacement, each point can potentially move, increasing the likelihood of a smoother boundary. The graphs for shape size reinforce these inferences, with a positive correlation between fractal dimension and end shape size (the area of the shape increases scope for boundary complexity). Therefore, we have three algorithms with their own distinct geometric properties.

Figures 5d – l show the graphs arising from various combinations of probabilities. In all cases, the inclusion of two or more algorithms has an additive effect on the fractal dimension, some reaching as high as $D = 1.8$. Unless midpoint displacement is mixed with one or other of the remaining algorithms in equal proportions (Figs 5 e and f), the change in $D$ follows quite a rough curve (though not as variable a local trend as midpoint displacement alone – fig 5a). Another phenomenon to highlight is that edge and vertex displacement have an unstable influence on each other (Figs 5 g and h) when compared to their sole results in Figs 5b and c. The shape growth rate with combined use of algorithms generally behaves predictably (e.g. in Fig 5g, combining edge and vertex displacement in equal proportions results in a middling curve, about halfway between the curves for edge displacement (fig 5b) and vertex displacement (fig 5c) alone).

In designing strategies for recreating real geographic phenomena, such nullifying effects (in both growth rate and variability of fractal dimension change) should be taken into account. Also of note is the marked effect on sizes derived from use of the edge displacement algorithm, which can easily dominate other objects. One mode of generic use could be to use midpoint displacement to adjust the shape to the desired level of complexity (fractal dimension), taking into account the possible additive effects of using other algorithms. Control in the temporal axis as well would entail the use of vertex and edge displacement to model slow and fast shape growth rates respectively. This scheme is described in Figure 6.

We can visually test our findings by considering the city as an example of a complex spatial phenomenon and comparing patterns from this domain with the vector agent’s dynamic geometry class. Thus the similarity between our model output with real-world data can be demonstrated (Figure 7, 8). Ostensibly, the shapes generated from the vector agent simulation may appear similar to CA model output, for example simulating an urban growth pattern (Figure 9). However, beneath the visual appearance of a contiguous irregularly-formed area, the CA output comprises a group of cells that has no overall identity or notion of the macro-object it appears to depict. The vector agent polygon has its own identity and as such is able to be manipulated and queried as a closer representation of a real-world object. This also makes the neighbourhood, states and rules possessed by the object conform more to their equivalents in reality (a similar argument is put forward by Torrens and Benenson, 2005). It is more meaningful to associate transition rules with recognisable phenomenon objects than abstract cells.

These initial comparisons and observations support some of the vector agent’s objectives, which are the ability to simulate irregular objects and manipulate their geometric boundary. These have resulted in generating different shapes that can easily be used to represent different types of complex phenomena, such as those found in the
evolving spatial structure of a city. Note, however, that the comparisons carried out here are not a formal proof of similarity, but a visual comparison that supports the notion that the underlying geometries are comparable. No entity states or transition rules linked to a specific geographic process have been used; this will be the next stage in model implementation.

Figure 7: Comparison between vector agent simulation output and land use parcels. (a) A sample of land use parcels for Swindon, south central England, with average fractal dimension 1.570 (source: Batty and Longley, 1994); (b) Vector agent output of average fractal dimension 1.581

Urban area for Dunedin city, New Zealand

Vector-agents

Figure 8: Comparison between vector-agents simulation output and mesh-blocks for urban area

Figure 9: Comparison between (a) simulated urban scenario of Sydney for urban development generated by CA (source: Liu, 2001) and (b) vector agent simulation output with non-restricted shape size.

4.0 CONCLUSION AND FUTURE RESEARCH DIRECTION

This paper has outlined a group of geometry-led agents called vector agents (VA). They have emerged out of a desire to impose a systematic framework to the “georeferencing convention” element of Geographic Automata Systems (GAS – Torrens and Benenson, 2005), a geographically-led system alternative to CA that uses Multi Agent Systems (MAS). The derivation of irregular and dynamic VAs has been explored, in particular the
interplay of midpoint displacement (normally used for fractal enhancement), edge displacement and vertex
displacement in evolving various geometries. The size and shape complexity ranges thus derived for vector agent
objects suggests possible modelling of real-world phenomena with a wide range of geometric characteristics
(visualy tested here on urban polygon objects).

The next stage is to incorporate the other elements of GAS (states, transition rules, neighbourhoods,
nearbourhood rules) and apply vector agents to simple scenarios in a real-world domain. Work is currently
underway on developing a model of Burgess’ concentric ring model of the city (Burgess, 1925), with a view to
eventually modelling a real city, building in other model elements on the way (e.g. sectors, multiple nuclei, the
compact city philosophy). The Burgess model adopts a Delaunay neighbourhood (Semboloni, 2000) for its
vector agents and adopts Egenhofer and Franzosa’s (1991) topological descriptions (e.g. disjoint, meets,
overlaps) to characterise the relationships between different agents.

The vector agent manifesto is that it offers the ability to explicitly control geometric form by the alteration of
simple parameters. The VA polygon can be treated as a single object with its own identity, which paves the way
for a whole range of generalisation and transformation operators to be applied on it. This cannot be effectively
achieved with a contiguous group of CA cells that happen to all have the same identity. Having said this, the use
of vector agents will dramatically increase the model complexity and therefore processing speed.

ACKNOWLEDGEMENTS

This work would not be possible without funding from Otago University Postgraduate Research Scholarship.

REFERENCES

Conference on Geocomputation, University of Southampton, UK.
Leung, Y. (Eds.) Geocomputational Modelling: Techniques and Applications, Advances in Spatial Science,
Batty, M., Desyllas, J. and Duxbury, E. (2003) The Discrete Dynamics of Small-Scale Spatial Events: Agent-
Based Models of Mobility in Carnivals and Street Parades, International Journal of Geographic Information
Benenson, I. and Torrens, P. M. (2004b) Geosimulation: Automata-Based Modelling of Urban Phenomena,
University of Chicago press, Chicago, pp. 47-62.
Association of American Geographers, 77, pp. 265-278.
Land-Use Modelling, Proceedings of the 16th Annual Colloquium of the Spatial Information Research Centre,
University of Otago, Dunedin, New Zealand.


AGENT-BASED SIMULATION FOR BUILDING FIRE EMERGENCY EVACUATION

Hajibabai, L\textsuperscript{1,}, M. R. Delavar\textsuperscript{1,}, M. R. Malek\textsuperscript{1} and A. U. Frank\textsuperscript{1,2}

1: Center of Excellence in Geomatics Engineering and Disaster Management, Dept. of Surveying and Geomatics Engineering, Engineering Faculty, University of Tehran, Tehran, Iran
lbabai@geomatics.ut.ac.ir
mdelavar@ut.ac.ir
malek@ncc.neda.net.ir

2: Institute of Geoinformation and Cartography, TU WIEN, Austria
frank@geoinfo.tuwien.ac.at

ABSTRACT

There are a number of events which motivates to discuss intelligent protection and efficient evacuation of people from buildings in fire emergencies. An agent-based strategy in a fire emergency is developed in this paper, which would allow the assessment of the achieved information for people in such critical situations. In this process, the agent would decide upon paths with the use of signs and topological specifications in decision points in order to find its way to escape from the building. The theoretical outset of the research is the observation that humans show distinct behavioral and cognitive preferences when dealing with stressful situations. This paper outlines the agent-based wayfinding strategy in disaster management due to fire emergency, delivering simulation results based on the methodology employed, and further discusses the results.

Geospatial Information Systems (GIS) incorporate a wide range of processes applied on geospatial data to support decision-making programs. In case of some disasters such as building fire emergencies, a number of other parameters like stress and temporal considerations will be added. In order to have optimum management of such events, which are mostly spatially-based, expert methods in GIS are applied. The goal of the research is an agent-based modelling in decision-making and navigation in an indoor environment in a fire emergency. Construction and inspection of mental representations of spatial environments and exploring these models in agent-based simulation have been discussed and the proposed computational model tested in an indoor complex building. Initial results of the simulation proved the reliability of the model.

Keywords: agent-based modelling, fire emergency, wayfinding, crisis management

1. INTRODUCTION

There have been frequent events which cause extensive life loss because the time needed for safe evacuation from a threatened building was not available [13]. In order to solve such problems and avoid occurrence of these
events, it is recommended to consider intelligent approaches such as agent-based wayfinding simulation in the stressful situations. So, there is a need to understand how the people, building, and environment react together. People need to engage in symbolic interaction to develop new emergent definitions of the situation. The real-life studies will proceed in complex settings like hospitals to provide a rich picture of wayfinding preferences and strategies used. Requirements must be based on the understanding that humans are information-processing adaptive agents that pursue goals aimed at protecting themselves, others and valued artifacts.

The main part of this research is the agent-based simulation model of a building evacuation in a fire emergency. This research outlines the cues and using them in navigation in an indoor environment in a fire emergency case. This paper presents a model of some aspects of the agent-based simulation of wayfinding in a building fire emergency, where appropriate elements of human perception and cognition are realized using knowledge gained by a cognizing spatial agent in the environment. The goal-driven spatial reasoning that leads to action begins with incomplete and imprecise knowledge derived from imperfect observations of the space. The physical complexity relates to the visual access and the spatial layout of the built environment. Such information can be achieved through the signs. The presence and uniformity of social control agents and the dissemination of evacuation directions may further modify the outcomes of the interactions.

In order to model the construction and inspection of mental representations of spatial environments and to explore these models computationally and to test the wayfinding information provided to people, the model is implemented in an indoor complex building. This paper presents the development of computational infrastructure for the simulation of escaping from fire propagation in multiple domains using agent platforms as an informal validation of the data-driven environment. To represent and simulate people’s processes of wayfinding, it is necessary to understand how people immediately make sense of spatial situations while performing a wayfinding task which will occur in a building occupant movement during fire emergencies. Some of the assumptions for rescue operation in these situations may fail and so we should use intelligent approaches in these contexts. The first step is the correct understanding of the nature of the intelligent agent. Then the scientific background of the agent-based wayfinding simulation is introduced in this paper.

2. HISTORY

Previous research on human wayfinding [10, 12], spatial perception and cognition, psychology, and philosophy has focused primarily on mental representations rather than processes of wayfinding. They focused on the development of computational models that simulate wayfinding in familiar environments. Route-planning tasks were solved by using a previously acquired mental representation of the environment. During wayfinding in an unfamiliar environment people cannot use such representations but have to rely on other sources to satisfy their information needs [17].

Past approaches based on physical systems representations in building evacuations do not serve well when integrating physical systems approaches with the goal-driven adaptive performance of people [13]. The present day multiplicity of models of emergency evacuation, each with their own strengths and weaknesses and without the appropriate methods of validation must be superseded by a government sponsored effort to create a uniform simulation platform that would combine what is good in existing models, provide proper validation tools, and encourage multi disciplinary collaboration to advance them. Other models available nowadays, a potentially misleading situation that needs to be corrected. In such context, research and theory in the social sciences can
have an important effect in grounding the models in realistic assumptions regarding social behavior in crisis situations, and such modeling in turn could enrich our understanding of collective behavior in crisis situations [13].

The goal of the previous work of the agent-based simulation of human behavior was to develop a computational theory of perceptual wayfinding. That theory uses an agent-based approach and can explain people’s wayfinding behavior in unfamiliar buildings. It is different from previous computational models for wayfinding, which were built to investigate how mental representations are created, stored, and used. These models assume that people become familiar with their environments over time and, therefore, acquire cognitive maps [6, 11]. In many situations though, people have to find their ways to novel destinations in unfamiliar environments [16]. The main result of this work is a practical tool that can be used to test the wayfinding information presented to people in an environment. The agent-based simulation framework allows analysis of the agent’s wayfinding process with respect to success or failure of reaching a goal. It discovers where and why wayfinding problems for the agent occur and what needs to be done to avoid them. This tool can help designers and architects to test and assess possible design alternatives prior to the construction of a building [16].

This paper focuses on the agent-based simulation of finding the exit ways in a fire building emergency which yields plausible results and can be used to explain people’s wayfinding behavior in crisis situations.

3. BASIC CONCEPTS OF THE RESEARCH

Several recent events have motivated discussions on how to best protect and safely evacuate building occupants during fire emergencies. In this modelling, we should not neglect the different behaviors of human beings in crisis situations. A concept for an expert system will be described that, through the use of agent technology, can recommend immediate actions that can be taken to mitigate the situation and prevent further deterioration, and can be used to aid the rescue workers and evacuees in rescue efforts and safe egress. Therefore, major concepts and scientific backgrounds for modelling wayfinding have been introduced below.

3.1. Agent

“Intelligent agent technology is a rapidly developing area of research. However, in reality, there is a truly heterogeneous body of work carried out under the ‘agent’ banner” [3]. Agent theory is a young scientific field without common paradigms. Different people from different domains working in the field have different understandings of the concepts. It is an important new way to conceptualize and implement software applications. It introduces the concept of collective intelligence and emergence of structures by interaction into the field of artificial intelligence [3]. In the field of social simulation it offers the promise to allow the modelling of autonomous individuals and interaction between them. Agent theory has the potential to stimulate and contribute to a broad variety of scientific fields. A truly autonomous intelligent agent should be able to operate successfully in a wide variety of environments, given sufficient time to adapt [18].

Neglecting the focus on particular methods and applications, the core element of agent theory is the agent concept [3]. According to the heterogeneity of the field, there is no common agreement about a definition of the term ‘agent’. “Anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors” can be an agent [18]. An agent is an entity that is capable of acting in its environment and capable of perceiving its environment (Figure 1). The environment provides percepts to the
agent, which the agent perceives through its sensors. A sensor is anything that can change the state of the agent in response to the change in the world [9]. Agents are environment dependent: once an agent leaves the environment to which it is adapted, it may no longer be considered as an agent [7].

![Figure 1. An agent embedded into its environment [3]](image)

This general definition allows the application of the agent concept to very different fields. Agent-based simulation provides a new solution to the simulation of social processes, because it allows including representations of individuals with individual capabilities and preferences into the model. An agent can be viewed as satisfying an ordered set of goals to achieve some overall objectives. The agent takes a sequence of actions in order to satisfy the next goal in the set. The effect of the change can be a new set of actions to achieve the same overall objective as before, or it may even result in a new overall objective if the original objective cannot be achieved anymore in the current environment.

The key properties of agents are autonomy and the embedding of the agents into the environment. Autonomy means that the agent is capable of acting based on its percepts and individual experiences. An agent is autonomous if it acts without the intervention of humans or other systems [18]. This means that an agent has control over its internal state, and over its behavior. The agent’s behavior depends on its internal state and the built-in knowledge and on its own experience. There is common agreement in the multi-systems theory literature that the autonomy is a required property of agents. Everything beyond autonomy of an agent depends on the definition of an agent as a whole and its internal state and has to be specified according to the needs of an application. Cognitive agents have representation of the environment, they are able to draw up models of the environment and of the other agents and predict, explain and understand the changes in the environment and in the other agents. They have the ability to solve certain problems by themselves and reason on the basis of their representation of the world. Agents can change the location of objects in space by their actions [5].

Mobile agents have been proposed as a novel and useful paradigm for designing distributed applications. Mobile agent-based distributed applications are specially suited for mobile computing environments involving different types of devices because of better bandwidth conservation, support for disconnected operations, easier device/user-specific customization, etc. As a mobile agent migrates from machine to machine in a heterogeneous network, the environment in which it operates changes and it may encounter unexpected events like faults. The ability to adapt to dynamic environment and unexpected events is a key issue for mobile agents.

Agents can interact through explicit linguistic actions (communication) or by nonlinguistic (physical) actions modifying the world in which they act. Communication allows the agents to exchange information and to coordinate their activities. The fundamental requirement for agents to be able to interact on a high level is
communication. Communication is required to form a group out of the single agents. Communication protocols and languages are necessary to construct multi-agent systems [3]. Agents can communicate by exchanging information following two major strategies. Agents can directly exchange messages or exchange messages over a data repository shared by all agents of the system. Direct message exchange is called message passing. Indirect communication architectures are designated as blackboard architectures, because the common data repository for indirect communication is called a blackboard [3]. Most agent communication protocols are based on speech acts [3]. Speech act theory provides the foundation for high level communication. Speech acts can be represented and exchanged between agents with a message exchange architecture. The message exchange architecture is based on the classical theory of communication by Shannon and Weaver [3]. In this model the communication act consists of sending of some information from a sender to a receiver. On the sender side, the information is encoded using a language and decoded on the receiver side [3]. The agent-based model will be expressed in a formal computational language. The language must be expressible and understandable enough to allow a sophisticated representation of the agent framework [5].

3.3. Spatial Cognition
Simulating people’s wayfinding behavior in a cognitively plausible way requires the integration of structures for information perception and cognition in the underlying model [16]. The importance of knowledge about human spatial cognition for explaining and predicting people’s behavior in geographic space was stressed by a number of scientists [16]. Human spatial cognition is a part of the interdisciplinary and therefore wide-ranging research area of cognitive science [16]. In particular, cognitive science deals with the study of human intelligence in all of its forms, from perception and action to language and reasoning [16].

Spatial cognition concerns the study of knowledge and beliefs about spatial properties of objects and events in the world. How human beings deal with issues concerning relations in space, navigation and wayfinding is discussed by the term spatial cognition. Cognition is about knowledge: its acquisition, storage and retrieval, manipulation, and use by humans, and intelligent machines. Generally speaking, cognitive systems include sensation and perception, thinking, imagery, memory, learning, language, reasoning, and problem solving. In humans, cognitive structures and processes are part of the mind, which emerges from a brain and nervous system inside a body that exists in a social and physical world. Spatial properties include location, size, distance, direction, separation and connection, shape, pattern, and movement. Spatial abilities are cognitive functions that enable people to deal effectively with spatial relations, visual spatial tasks and orientation of objects in space. One aspect of these cognitive skills is spatial orientation, which is the ability to orient oneself in space relative to objects and events; and the awareness of self-location.

Spatial cognition is composed of several elements: landmarks, route maps and survey maps. Landmarks are identifiable environmental markers associated with specific geographic locations [4]. Route maps are sequences of instructions, often involving landmarks, that describe at a personal level how to get from one location to another [1]. Survey maps are similar to topological maps and describe the spatial layout of the environment as opposed to reflecting a specific navigational task [1].

People are able to imagine the physical demands of their response to the crisis and thus respond in terms of what they think they can do within the time and other considerations that they consider relevant as they formulate responses to the crisis. Thus, the elderly, the physically infirm, caretakers, women, the injured, will have a greater tendency to begin evacuating sooner than other categories of victims and will have a higher probability of becoming obstacles to the evacuation movement in constrained spaces. Human imagination, particularly how the
actual or perceived physical incapacity of the actors, and the extent to which the physical tasks of evacuating present important challenges to them, impact their timing of evacuation behavior.

Perceptions of danger are socially determined. Dangerous conditions by themselves are not always effective triggers for evacuation response, except perhaps in situation of mass behavior previously identified, in which the evacuation response is forced upon the person. Ambiguities and mixed messages and inaccurate interpretations of dangers often impact evacuation behavior, so that while it is true that the presence of inter subjectively verified and consistent signs of danger that are accurately perceived, such as smoking and loud sounds, facilitate the adoption of new behavior, this situation should not be assumed to be the normal state of affairs in simulation models.

3.2. Wayfinding
Wayfinding and route directions have developed into central research areas in cognitive science [8]. Finding one’s way in the environment, reaching a destination, or remembering the location of relevant objects are some of the elementary tasks of human activity. Wayfinding is a basic activity that people do throughout their entire lives as they navigate from one place to another. Many theories of spatial cognition have been developed to account for this behavior [14]. In order to represent and simulate people’s processes of wayfinding it is necessary to understand how people immediately make sense of spatial situations while performing a wayfinding task [15].

Wayfinding is a complex human activity involving moving along while evaluating alternatives and making decisions. It is defined as a spatial problem solving process with the three sub-processes decision-making, decision execution, and information processing [19]. Wayfinding is the cognitive element of navigation. It does not involve movement of any kind but only the tactical and strategic parts that guide movement. Wayfinding is not merely a planning stage that precedes motion. Wayfinding and motion are intimately tied together in a complex negotiation that is navigation. An essential part of wayfinding is the development and use of a cognitive map, also referred to as a mental map. When focusing on the mobility of humans, the ease of wayfinding within a building can be seen as an essential function of a building’s design [2, 12]. Motion is the motoric element of navigation. Navigation is the aggregate task of wayfinding and motion. It inherently must have both the cognitive element (wayfinding), and the motoric element (motion). Consequently, we use this term only when we mean to imply the aggregate task and not merely a part. Maneuvering is a subset of motion involving smaller movements that may not necessarily be a part of getting from "here" to "there" but rather adjusting the orientation of perspective, as in rotating the body, or sidestepping. This is an important distinction to make for the development of active transport interfaces for locomotion such as Gaiter.

“Spatial orientation and wayfinding subsume an ensemble of complex mental process. They allow people an idea of surrounding space, of their positions in that space, and they allow purposeful movement within that space” [12]. Wayfinding can be described as procession using spatial and environmental cues to find one’s way in a built or natural environment. The designer and client seeking to manipulate a function or use of a particular environment can also define wayfinding. Wayfinding should not be thought of as a “signage design”, but as a broader, more inclusive process. Signs of dangers such as smoke or fire impact both the decision to evacuate and the evacuation movement.
4. METHODOLOGY

Methodologies will be discussed for determining optimal and robust tactical and operational agent-based strategies for rapidly evacuating a large burning building or a building that has come under attack by enemy or a natural catastrophe [13]. These procedures explicitly consider the inherent dynamic and uncertain nature of circumstances requiring evacuation. Therefore, they give rise to robust agent-based evacuation plans with lower probability of failure than paths determined otherwise, enabling faster and more efficient evacuation of a building in the event of military attack, fire, natural disaster, or other circumstances warranting quick escape.

For developing a simulation method for human behavior in space, the background theory of the process should be in people’s real world experiences. It also considers the counter-response of evacuees whose path during egress is blocked by smoke accumulation near an exit. However, all the occupants of a certain decision point will initially traverse the same agent-specified path to an exit. Given a sequence of landmarks between the current position and a desired destination, the controlled object is programmed to execute the appropriate steps necessary for an agent to reach the goal point. The advent of performance based fire safety regulations and codes together with the needs for agent-based evacuation simulation models gives further impetus and sense of purpose for future endeavours.

This work focuses on properties of the environment as perceived and cognized by an agent. Much of the information people need to perform a task is in the world and the agent is perfectly tailored to make sense of this world. Formalizing the conceptual model for the cognizing agent allows to describe it more precisely than by using a verbal description and to create a practical tool for simulating the test case. The main parts of the model are an agent who tries to solve a wayfinding task in a crisis situation, objects within the built environment, information gained from the environment, and actions taken by the agent based on such information (Figure 2).

![Diagram](image)

**Figure 2.** The main parts of the agent-based model for wayfinding in the environment

The model simulates an agent in the complex building that has to perform the task of finding the exit ways in a fire emergency.

5. SIMULATION

Agent-based simulation for crisis management that incorporate social scientific processes concludes by pointing out the so far ignored insights that could be derived from a dual emphasis on the social psychology of the agent in the environment. It concludes with a number of predictions regarding the effects of social organizational variables on the timing of evacuation behavior.
This simulation takes the agent-based model and determines an optimal plan to evacuate the building in a "minimum" amount of time. The agent is provided a summary of results for the specified model, including total time periods. It is designed to produce results that take account of a fixed set of environmental features, assumed travel speeds, and an arrangement of varying levels of service. The consequence is that several sociological assumptions can be made but not articulated or translated into attributes or algorithms relating to the motion of humans. However, this relies upon viewing the movement of evacuees as a continuous flow, not as an aggregate of persons varying in physical abilities, individual dispositions and direction of movement. As a primarily agent-based model, the movement of individuals is established by a fixed set of motion rules. The model as a whole is comprised of some interacting sub models of movement, behavior, wayfinder (agent), and hazard. For instance, the hazard model will generate values that correspond to a particular configuration of threat across the simulated environment. The values of the variables associated with agent behavior, influence the calculations of the movement model.

Agent simulation for crisis management improves upon other simulation models that are concerned with numerical analyses of inputs or amounts of people and structures. This feature serves as an improvement to programs that only allow the agent to specify the occupants to follow the available paths considering the location of the fire or threat. The agent-based system for crisis management is grounded on empirical data taken from “real-world” experiments. If the agent sees an exit, it will proceed towards it, and if it receives any types of direction to leave, that will be carried out without fail.

This simulation is based on a series of decisions that occur concurrently during each cycle. The decision to begin moving in response to an announcement of an emergency is made on the basis of a changing global probability that takes into account the degree to which a visible threat is evident and the perceived number of agents that are and are not moving. The inclusion of a disposition such as fear or deference is not entirely without precedent. In general, the simulation model implements a keen awareness of the multiple social criteria that humans assess before deciding to evacuate, the need for clear information about the situation and exits in order to avoid extensive ambiguity, and the significant yet somewhat fragile nature of orderly movement in the face of a major threat. Agent-based simulation could clearly benefit from a set of specifications that generate more diverse and realistic physical environments.

In this simulation, the agent gains knowledge about the building indoor environment through visual perception of sign information at decision points in order to find its way to escape from the building. Neither the ability to learn nor a lasting cognitive-map-like representation of the environment is involved in deciding upon and taking an action. The agent’s decisions and actions are founded on wayfinding strategies and commonsense reasoning. Based on the knowledge in the world, the agent takes a sequence of actions until the wayfinding task is completed. Starting with imperfect observations of the space, the agent derives incomplete and imprecise knowledge and based on such knowledge takes actions which lead to further observations and knowledge, recursively to further actions until the agent’s goal is reached.

In this paper we present a formal model of the agent-based wayfinding process in a fire building emergency. The model integrated elements of people’s perception and cognition, therefore, focusing on how people make sense of their wayfinding environment (Figure 3).
Table 1 shows the required test data for the representation of some parts of the agent’s wayfinding environment in the building, their positions, the decision points which the agent is able to go, and also the incoming-direction values assigned to adjacent decision points from where the agent could have entered. They are based on the preferred directions of the cognizing agent.

<table>
<thead>
<tr>
<th>Position</th>
<th>Go-to</th>
<th>Direction</th>
<th>(Enter from, incoming Direction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>(1,4),(3,0),(4,1),(5,6)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>(2,4),(4,2)…</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6</td>
<td>(2,5),(3,6), …</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Test data of the building environment

Because of the dynamic nature of the fire, the propagation rule of the fire and smoke, that has a greater effect on the occupants than the fire itself, is also would be considered in order to model the agent’s escaping from the burning building. The stressful situation in a fire emergency also affects human behaviors in the wayfinding task. For example, they may select the longer or obstructed way. So, the stress factor would be considered in the simulation. Using the same wayfinding agent in two complex buildings and comparing the evacuation time from those buildings would determine the more efficient building design from the wayfinding point of view to facilitate crisis management in fire building emergencies.
The relative strengths and weaknesses of the models presented above point toward several key recommendations. From a social science perspective, the ideal simulation modeling approach should seek the development of sub-models that posit an active, “investigative” socially embedded agent that assesses the state of other persons and forms a definition of the situation in cooperation with others. Furthermore, these agent-centered calculations should be placed in an on-going interaction between the properties of a particular fire and other hazard and the physical surroundings in which the evacuation takes place.

6. CONCLUSIONS AND FUTURE DIRECTIONS

There have been various methods to facilitate rescue operations in building evacuation in fire emergencies. Agent-based simulation of building evacuation to find a less time-consuming exit way for escaping from the building is one of the proposed strategies. This model concentrates on people’s actual information needs during wayfinding and does not focus on learning a spatial environment. Its main principle is that all wayfinding information about the different destinations has to be presented to the agent at every decision point.

Simulation models of emergency evacuations can have enormous practical and scientific payoffs not only for the social sciences but also for other sciences such as engineering and public health. However, agent-based simulation models can realize their full potential as a tool for emergency planning and intervention only if they are inextricably linked to fieldwork and empirical investigations of emergency evacuations that would provide computer scientists and mathematicians with the appropriate parameters for social behavior. Thus, their future is multi-disciplinary, involving the expertise of computer scientists, engineers, fire scientists, social scientists, and emergency planners, among others.

Using this agent-based model in two different complex buildings with various exit ways in order to study the efficiency of the wayfinding task, will compare the escaping time duration of the agent from those indoor environments and determine which building design would be better in a crisis situation. Therefore, the result of this agent-based simulation can be used in improving building designs for efficient fire crisis management. The research demonstrates the minimum amount of knowledge in the world necessary for the agent to find its goal. The model could be extended by explicitly integrating other essential elements at various situations.

REFERENCES


How people visually explore geospatial data

Urška Demšar
Division of Geoinformatics
Dept. of Urban Planning and Environment
Royal Institute of Technology (KTH), Stockholm, Sweden
urska.demsar@infra.kth.se

ABSTRACT

This study presents a small exploratory usability experiment with the goal to observe how people visually explore geospatial data. The famous iris dataset from the pattern recognition field was put into geographical context for this experiment, in order to provide the participants with a dataset with easily observable spatial and other relationships. The participants were given free hand to explore this dataset with a visual data mining system in any way they liked. The protocols collected during the experiment with the thinking-aloud method were analysed with the aim to understand what types of hypotheses the participants formed, which visualisations they used to either derive, confirm or reject their hypotheses and what exploration strategies they adopted.

Keywords: exploratory geovisualisation, visual data mining, exploratory usability

1. INTRODUCTION

Human-computer interaction (HCI) is a discipline that explores the quality of interaction between the users and information systems. One of the basic requirements for developing a usable and useful information system is knowledge about users and how they use the system. This is the base of the user-centred design, which is a philosophy where the needs, wants, and limitations of the users of an information system are given attention at each stage of the design process (Preece et al. 2002).

Usability of an information system is defined as the ease with which people can employ it in order to achieve a particular goal and the extent to which the system supports them. The idea behind usability is that systems designed with their users' psychology and physiology in mind are easier to learn and more efficient and satisfying to use (Nielsen 1993).

Usability evaluation is the process of systematically collecting data on how users use a system for a particular task in a particular environment. It has three main goals: to assess the extent and accessibility of the system’s functionality, to identify any specific problems with the system and to assess users’ experience with the interaction. While the first two goals focus on evaluation of the functional capabilities of the system, the last one considers aspects such as how the system supports the users in typical tasks and how easy or difficult it is to learn (Dix et al. 2004).

One of the methodologies for usability evaluation is evaluation through user participation, sometimes also called usability testing. This methodology involves various approaches, such as controlled experiments, observational
methods, query methods (interviews and questionnaires) and methods that use physiological monitoring (eye tracking or heart-rate measurements) (Preece et al. 2002, Dix et al. 2004).

Usability issues in geovisualisation are not the same as those used in traditional human-computer interaction. Geovisualisation is concerned with cognitive issues, such as attention, perception, memory, problem solving, hypothesis generation and decision making. The key questions are whether geovisualisation systems are effective for spatial problem solving and if they support human intuition in the exploration of geospatial data. Traditional usability methods borrowed from human-computer interaction therefore need to be adapted accordingly (Fuhrmann et al. 2005, Swienty 2005).

Methods for usability assessment of geovisualisation systems can be divided into formal evaluation methods and exploratory usability methods. Formal methods are tests that measure users’ performance on carefully prepared tasks. Data collected in a formal evaluation consists of controlled measurements of the number of errors and time to complete the task. Data on user satisfaction is also typically collected in questionnaires and interviews. Exploratory usability reflects the nature of geovisualisation and is not concerned with users’ performance, but investigates how users use the tools. It produces descriptive data in the form of verbal protocols and observational notes (Preece et al. 2002, Griffin 2004, Fuhrmann et al. 2005).

Some recent examples of formal evaluations of geovisualisation tools include an evaluation of interactive tools for exploratory analysis of geographically referenced data (Andrienko et al. 2002) and an evaluation of a Public Participation GIS (PPGIS) for environmental planning (Haklay and Tobón 2003). Edsall (2003) presented a usability experiment for a geovisualisation system for health statistics. Ahonen-Rainio (2005) conducted formal user tests of metadata visualisations. Koua (2005) performed a formal usability evaluation of an exploration system that consisted of various geospatial visualisations and a Self-Organising Map. ESTAT - an exploratory geovisualisation toolkit for epidemiology was designed and evaluated by Robinson et al. (2005).

Formal usability evaluation is insufficient for testing the capability of geovisualisation tools for hypothesis generation and knowledge construction. Rather than conducting only controlled experiments, users also need to be engaged in free data exploration, where they are given free hand to do whatever they want with the tool. In this way formal and exploratory usability evaluation methods complement each other (Andrienko et al. 2002, Griffin 2004). However, not many such experiments have been found during the literature survey. One reason for this might be that they are more time-consuming to conduct, as the participants need training, in order to be able to use the tools independently during free exploration. In this sense it is useful to perform both types of evaluations in one experiment. A controlled experiment with pre-defined tasks can serve as training for independent free exploration. Some examples of recent evaluations with formal tasks and an exploratory usability session include testing of the coupling an aspatial visualisation system with a GIS system (Tobón 2002), an evaluation of data display devices for geographical models (Griffin 2004) and a low-cost usability evaluation where formal evaluation was combined with an exploratory usability experiment to evaluate application of visual data mining on emergency response data (Demšar 2006).

This study builds upon findings presented in Demšar (2006) and focuses on the exploratory usability of a visual data mining system. The system used in this study is based on GeoVISTA Studio (Gahegan et al. 2002, Takatsuka and Gahegan 2002), which is a collection of visual and computational data exploration components. The formal usability issues of the elements in this collection, such as identifying deficiencies in design and functionality of particular visualisations have been addressed by others (Edsall 2003, Robinson et al. 2005).
Rather, the goal of this study was to investigate how people visually explore geospatial data, what exploration strategies they adopt and how the visual data mining system supports them in the exploration and hypotheses forming process.

As the experiment was performed with limited resources regarding cost and testing personnel, it was designed with the goal to collect the maximum amount of information with the minimum possible means. Simple data capture techniques were used instead of expensive methods and the number of participants was limited according to the recommendations of the discount usability engineering approach (Nielsen 1994).

2. CASE STUDY: EXPLORING A GEOSPATIAL DATASET WITH VISUAL DATA MINING

A famous botanical dataset used in data mining and pattern recognition was put into a geographical context for this study. It was further extended with additional attributes in order to provide the participants with a dataset where clearly and easily observable spatial and other relationships existed. This section describes how the dataset was prepared and gives a brief introduction to the visual data mining system used for data exploration in the usability experiment.

2.1. The data

The original dataset consists of 150 samples of four measurements on three species of iris plants (Iris setosa, Iris versicolor and Iris virginica), made by E. Anderson and used by Fisher (1936) in his classic paper on discrimination analysis. The dataset contains 3 classes of 50 instances each, where each class refers to one species of iris plant. Originally the 100 plants of species Iris setosa and Iris versicolor grew together in the same colony, while the 50 Iris virginica flowers were taken from another natural colony. The four attributes measure the sepal length, the sepal width, the petal length and the petal width of each flower (Fisher 1936, Duda and Hart 1973).

Figure 1. A bivariate multiplot matrix of the original iris dataset.
The iris dataset is famous in data mining and pattern recognition and has been often used for evaluating clustering procedures. It is particularly suited for this purpose, because it is not too large and has the characteristics that one class (Iris setosa) is linearly separable from the other two, which are not linearly separable from each other (Witten and Frank 2000). This can be seen in fig. 1, which shows a bivariate multiplot matrix of the four original variables. The Iris setosa class (indicated in dark grey) is clearly separated from the other two classes (Iris versicolor is indicated in light grey and Iris virginica in black) in all of the scatterplots.

![Figure 1. Iris dataset multiplot matrix.](image)

**Figure 2.** Iris dataset put into geographical context – Iris virginica is separated from the other two classes.

For the purpose of this exploratory usability assessment, the dataset was put into geographical context by producing fake coordinates for each plant. These coordinates were computed as functions of the two out of the four original attributes. Several functions were tried, but finally a sum of sinus and cosinus was used, where the x coordinate was derived from petal length and the y coordinate from petal width. This assignment separated one flower species from the other two in the geographical space – a separation which also corresponded to how the flowers grew in two distinct colonies. The separated class was Iris virginica (fig. 2), which is one of the two linearly non-separable classes in the original four-dimensional attribute space. Such coordinate assignment provided the possibility to observe the differences in the separability in the attribute space and the geographical space.

The derived coordinates were linearly transformed to represent real world coordinates in a 2500x3500m area located near Stockholm. The location of the area had nothing to do with iris plants. It was chosen for two reasons. One was to make the dataset appear real to the participants of the test and the other was that the author had access to various maps of this area and could assign roughly realistic values to other attributes as described later in this section. In order to increase the dimensionality of the attribute space and provide a more complex environment with interesting features for visual exploration, six more attributes were added to the dataset.

Three of the new attributes were numerical and were chosen to represent three additional botanical characteristics of iris plants: stem length, leaf length and leaf width. Stem length was randomly chosen, leaf length was a function of stem length and sepal length, while leaf width was derived from sepal width and sepal length. The idea behind these derivations was to have a random, a semi-random and a dependent attribute in the dataset.

Three nominal attributes were also added to the dataset, representing bedrock, soil and landuse at the location of each plant. The values for these roughly corresponded with the study area (as observed on the geological, soil
and topographical maps of the area) and roughly fit the habitat conditions of the three iris species (SIGNA 2006). All the attributes of the dataset are described in table 1.

Table 1. The iris dataset

<table>
<thead>
<tr>
<th>Attribute name</th>
<th>Dependent upon</th>
<th>Range/values</th>
</tr>
</thead>
<tbody>
<tr>
<td>iris species</td>
<td>- (original attribute)</td>
<td>1 – iris setosa, 2 – iris versicolor, 3 – iris virginica</td>
</tr>
<tr>
<td>petal length</td>
<td>- (original attribute)</td>
<td>1.0 – 6.9 cm</td>
</tr>
<tr>
<td>petal width</td>
<td>- (original attribute)</td>
<td>0.1 – 2.5 cm</td>
</tr>
<tr>
<td>sepal length</td>
<td>- (original attribute)</td>
<td>4.3 – 7.9 cm</td>
</tr>
<tr>
<td>sepal width</td>
<td>- (original attribute)</td>
<td>2.0 – 4.4 cm</td>
</tr>
<tr>
<td>x coordinate</td>
<td>petal length</td>
<td>1601500 – 1605000</td>
</tr>
<tr>
<td>y coordinate</td>
<td>petal width</td>
<td>6583000 – 6585500</td>
</tr>
<tr>
<td>stem length</td>
<td>random value</td>
<td>10.7 – 99.3 cm</td>
</tr>
<tr>
<td>leaf length</td>
<td>stem length and sepal length</td>
<td>2.3 – 70.8 cm</td>
</tr>
<tr>
<td>leaf width</td>
<td>sepal width and sepal length</td>
<td>2.6 – 6.1 cm</td>
</tr>
<tr>
<td>bedrock</td>
<td>selected study area and habitat conditions for the three iris species</td>
<td>1 – granite, 2 – gneissgranodiorite, 3 – metagreywache</td>
</tr>
<tr>
<td>Soil</td>
<td>selected study area and habitat conditions for the three iris species</td>
<td>1 – clay, 2 – marsh, 3 – moraine, 4 - organic soil, 5 – peat, 6 - rocky soil, 7 – sand, 8 – sand block</td>
</tr>
<tr>
<td>landuse</td>
<td>selected study area and habitat conditions for the three iris species</td>
<td>1 – meadow, 2 – coniferous forest, 3 – deciduous forest, 4 – mixed forest, 5 – swamp, 6 – fields</td>
</tr>
</tbody>
</table>

2.2. The visual data mining system

The visual data mining system used in this study was built using GeoVISTA Studio, which is a java-based collection of various geographic and other visualisations and computational data mining methods for geoscientific data analysis and exploration (Gahegan et al. 2002, Takatsuka and Gahegan 2002).

The system consisted of the following visualisations: a parallel coordinates plot (PCP), a multiform bivariate matrix with scatterplots, spaceFill visualisations and histograms, and a bivariate geoMap, which was included to represent the spatial component of the data. Figure 3 shows the graphical user interface of the system. The decision to build the system in this way was based on the nature of the exploration task. Visualisations were chosen based on their ability to show the structure in the data, such as clusters and bivariate dependencies between the attributes. Visualisations were connected using interactive selection and brushing (simultaneous highlighting of the same data object in all visualisations). A brief description of each visualisation is given in the remainder of this section.
Figure 3. The visual data mining system with a geoMap, a multiform bivariate matrix and a parallel coordinates plot.

A parallel coordinates plot (PCP) maps the m-dimensional space onto the two display dimensions by using m equidistant vertical axes. Each axis represents one attribute. Axes are linearly scaled from the minimum to the maximum value of the corresponding attribute. Each data instance is displayed as a polygonal line intersecting each of the axes at the point which corresponds to the respective attribute value for this data instance. The PCP reveals a wide range of data characteristics, such as clusters in the data and functional dependencies (Inselberg 2002).

A multiform bivariate matrix is a generalisation of a scatterplot matrix and consists of bivariate and univariate visualisations. An element in the row $i$ and column $j$ in a multiform bivariate matrix is a scatterplot of the variables $i$ and $j$, if it is located above the diagonal, a spaceFill visualisation of the same two variables, if it is located below the diagonal and a histogram of variable $i$, if it is on the diagonal. Scatterplots and histograms are widely known and standardly used visualisations. A spaceFill however needs an introduction. It is a bivariate visualisation, where each data instance is represented by a grid square. The colour is assigned to each square according to one of the two display attributes. The second display attribute defines the order of the squares inside the rectangular display. Dependencies between the two variables are shown as visual patterns in the spaceFill (MacEachren et al. 2003). A spaceFill can be also used to visually estimate the strength of the relationship between the two displayed attributes (Demšar et al. 2006).

The spatial visualisation in the system was a geoMap from GeoVISTA Studio, which is a choropleth map, whose colour scheme is defined by a cross-tabulation of the two display attributes (Gahegan et al. 2002, Takatsuka and Gahegan 2002).
3. METHODOLOGY

In spite of some recent attempts to understand how people visually explore geospatial data (Tobón 2002, Griffin 2004, Demšar 2006), we do not know very much about it. The goal of this study was to investigate this by observing participants during a free exploration of the geospatial dataset with a visual data mining system. The interest was in what exploration strategies the participants adopted and how the system supported them in the exploration and hypotheses forming process. This section presents how the experiment was conducted.

3.1. The participants

The number of participants had to be restricted to the necessary minimum because of time and cost limitations and the fact that only one person was available to conduct the test and the analysis of collected data. In the end six participants took part in the experiment. All were students of the International Master Programme in Geoinformatics and Geodesy at the Royal Institute of Technology (KTH) in Stockholm. None of them was a native English speaker and they originally came from five different countries (Ghana, Russia, Slovenia, Spain and Sweden), but they all possessed a good knowledge of English. Three participants were female and three male. All of the participants had engineering background. Five of them had a GIS experience of 0-3 years and one of 3-10 years.

3.2. The experiment

The experiment was conducted at the testing place in an assigned office, where each participant separately performed the test under monitoring of the observer. The test was conducted on the same portable computer for everybody in order to ensure the same conditions for all the participants. It was conducted entirely in English. The procedure took from 1 to 1.5 hours per participant. Due to limited resources there was no monetary compensation offered and the participants took part in the test on a completely voluntary basis. They were thanked for collaboration and presented with a small gift for their participation at the end of the test. The test consisted of six steps:

- a brief introduction to the test,
- gathering background information,
- an introductory training, where the exploration dataset and the visual data mining system were presented to the participant,
- free data exploration and
- a post-test questionnaire.

The participants were first informed about the purpose of testing in the introduction to the test. They were informed that they could abandon the test at any time, should they not feel able to perform the required tasks. They were assured of their anonymity and notified that the data collected during the test session would be used solely for this study.

In the second step of the test the participants filled in a short questionnaire about their academic background, native language, GIS experience and potential colour blindness. Collecting data about a system whose functionality is largely based on visual representation of data requires exclusion of test persons with cognitive dysfunctions, including colour blindness. If not, irregularities in test participants’ colour perception affect the results (Swienty 2005). Any achromates among the participants would be excluded from the experiment.
The third step was the *introductionary training* about the dataset and the exploration system. In this step the participants independently worked through a script which introduced the data, each of the visualisations of the system and demonstrated how to interact with them. The training took about 45 min per participant. This step was necessary in order to ensure that the participants were completely familiar with the functionalities of the system, because the concept of visual data mining, the way the GeoVISTA-based tools function and some visualisations were new for everybody. The participants were allowed to ask questions during this step. They could keep the introductory script in order to come back to it for help during the actual exploration.

The main part of the experiment was *the free exploration*. During this step, the participants were given free hand to perform whatever exploration they wished in any way they wanted during a limited amount of time (15 min). They were not allowed to ask questions, but could check the introductionary script for help. They were asked to verbalise everything that they were thinking and trying to do and in this way externalise their thought processes so that the observer could keep a written protocol. This data capture method is called the Verbal protocol analysis (VPA) or the “thinking-aloud” method and is an efficient way for performing a low-cost usability experiment, because it produces a lot of information from a few participants (Jordan 1998). In this case a more relaxed variation of the thinking-aloud method was used, known as “cooperative evaluation”, which allows the observer to ask the user questions if his/her behaviour is unclear or if he/she stops talking (Dix et al. 2004). The allowed questions were of type: “What are you thinking now?” and “What are you trying to do?” This variation of the thinking-aloud method was chosen based on experience from a similar experiment (Demšar 2006) where it was difficult to get some participants to talk. Finally the participants had to fill in *a rating questionnaire* and give their opinion about particular visualisations and the system. The aim was to measure the perceived usefulness, ease of use and satisfaction with the system.

### 4. RESULTS

This section presents a summary of the participants’ opinions and the results of the analysis of the verbal protocols. The results are presented in descriptive form and without any statistical analysis, due to the limited amount of participants and the fact that most data were collected in the form of written protocols.

#### 4.1. Participants’ satisfaction and opinions

At the end of the test, the participants were asked to rate each of the visualisations (the map, the multiform bivariate matrix and the parallel coordinates plot) on a scale of 1 to 5. The numbers represented the following opinions: 1 = very poor, 2 = poor, 3 = fairly good, 4 = good, 5 = very good. The participants were assessing ease of use, understanding of the tool after the test and satisfaction with the tool for each of the visualisations. The average ratings for each of the three categories (ease of use, understanding and satisfaction) are presented in fig. 4.
The participants rated all the visualisations highly, around good on the average. They perceived the bivariate matrix as the easiest visualisation to use. The map was the easiest to understand, while the PCP and the matrix shared the first place on perceived satisfaction.

In the answers that the participants gave to the open-ended questions, the majority considered the PCP to be the most useful visualisation as well as the most difficult to learn. They suggested several other visualisations and elements that they would like to have in such a system: surface models, 3D models and relational tables.

4.2. Exploratory usability

The data source for exploratory usability assessment were the “thinking-aloud” protocols that the observer kept during the free exploration. The participants all used up the assigned exploration time and needed to be stopped at the end, except for one participant (P3), who finished the exploration before the end of assigned time, as the system proved to be too difficult to use and too overwhelming for this particular person.

The thinking-aloud protocols were analysed with the aim of understanding the hypothesis forming, data exploration and tool manipulation for each participant. Each protocol was scrutinised for hypotheses, which were numbered and analysed. They were categorised according to their source, i.e. if they originated in previous knowledge, were prompted by a pattern that the participant noticed in one of the visualisations or were refinements of previous hypotheses. Table 2 shows the total number of hypotheses that each participant made and the number of hypotheses according to each source type: background knowledge (BGK), observed pattern in some visualisation (VIS) or a refinement of a previous hypothesis (REF).

Table 2. The number and classification of hypotheses that participants made during free exploration.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Number of hypotheses</th>
<th>Total number of hypotheses</th>
<th>Number of hypotheses according to source type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>BGK</td>
</tr>
<tr>
<td>P1</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>P2</td>
<td>5</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>P3</td>
<td>3</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>P4</td>
<td>7</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>P5</td>
<td>9</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>P6</td>
<td>10</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

The visualisations that the participant used to derive/confirm/reject each hypothesis were also counted. The ballot matrix was for this purpose split up into the separate visualisations: scatterplot, spaceFills and histograms. The total and relative frequencies of each visualisation were calculated for each participant. The total frequency for visualisation i and participant j, \( f_T(i,j) \), equaled the number of hypotheses where the participant used this particular visualisation. The relative frequency for visualisation i and participant j, \( f_R(i,j) \), was calculated as
\[ f_{s(i,j)} = f_{T(i,j)}/N_j, \]

where \( N_j \) represented the total number of hypotheses of participant \( j \). Relative frequencies for each visualisation are shown in fig. 5.

Figure 5. Relative frequencies per visualisation type and participant (P1,…,P6).

The PCP had the highest relative frequencies, followed by the scatterplots and the map. Histograms were used by three participants, while only one participant used the spaceFills.

The way each participant thought and explored the data by manipulating the system while deriving/confirming/rejecting each hypothesis was mapped as a path on the internal model of the visualisation process, suggested by Tobón (2002) and presented in fig. 6. Tobón suggests that the visual exploration process occurs in systematic fashion, where users incrementally improve their understanding of the dataset as they loop through the outer circle of the exploration model. Occasionally, the users can go backward and forward between steps or across the model before proceeding to higher levels of abstraction.
In this study the participants could be divided into three groups according to the exploration strategies they adopted. The first group, consisting of participants P1 and P3, repeatedly went through the following steps: they formed a hypothesis, which was based completely on their background knowledge. They looked at visualisations trying to either confirm or reject their hypothesis and manipulated the graphics in order to investigate if they were right or not. They interpreted the result from one visualisation and continued browsing through other visualisations until they could interpret the data. Then, they abruptly broke off exploration, discarded the hypothesis and proceeded by forming another, completely unrelated hypothesis, which was once again based only on their background knowledge. Figure 7 shows their exploration strategy.

Figure 7. The exploration strategy of participants P1 and P3.

Participants P2 and P6 followed a broader exploration path in the model. They formed their hypotheses based on a pattern that they noticed in some visualisation. They also amended their ideas once they interpreted the data and used new observations in what they were trying to confirm/reject. However, once they confirmed/rejected a hypothesis, they did not proceed refining it, but went searching for some other visual pattern on the screen and started the process from the beginning. Their exploration strategy is shown in fig. 8.
The last group consisted of participants P4 and P5, who followed a similar exploration strategy as participants P2 and P6, except that they also refined some of their hypotheses and looped several times through the model. This is indicated by the additional path in fig. 9, which follows the exploration circle on the outside. Their hypotheses originated either in what they observed on the screen or in hypotheses that they previously made. Four participants (P2, P3, P4 and P6) performed a systematical exploration of some attribute during the exploration, i.e. they systematically selected each value of this attribute and scrutinised other visualisations for possible patterns. This way to explore never occurred to the other two participants, P1 and P5. However, participant P5 used brushing (simultaneous highlighting of the same data object in all visualisations) to compare values of different attributes for one data element. This never occurred to any of the other participants.
5. CONCLUSIONS AND FUTURE WORK

This study attempted to investigate how people explore geospatial data with visual data mining. As the extent of this study was limited, it was not possible to draw general conclusions. However, some observations about the participants’ exploration strategies and preferences in the use of various visualisations could still be made.

All participants were master students in Geoinformatics and as such familiar with geospatial data and GIS. Therefore it was expected that they should not have too many difficulties operating the system. The experiment showed that they were in principle able to understand the new concept of visually exploring geospatial data. However, an introductionary training was necessary, as the concept of visual data mining, unusual visualisations and interactive exploration was new for them.

The relaxed variation of the thinking-aloud method, the cooperative evaluation, significantly contributed to the quality of collected data. In a similar experiment (Demšar 2006) where a strict variation of the thinking-aloud was used (which did not allow the observer to say anything during the exploration), verbalising varied very much between the participants. There some of the participants fell completely silent at times, which made it difficult for the observer to keep the protocol. In this experiment, the observer was allowed to ask questions to the silent participants, which made keeping the protocols much easier.

One of the findings of this study is that in the free exploration the participants used a PCP more than they used any other visualisation, as indicated by the relative frequencies of use. On the other hand, judging from the questionnaire answers, the PCP was perceived as the most difficult visualisation to learn and understand. There is clearly a discrepancy between perceived learnability and actual learnability here.

The map and the scatterplots were in the shared second place regarding the frequency of use. It is not possible to say which of the two was more popular, as the sample of participants was not large enough. However, Griffin (2004), who evaluated a system with maps, scatterplots, time series graphs and model parameter graphs in a similar but much larger experiment, also noticed that in her case maps were used more than scatterplots and these two types of visualisations were used more than any other ones.

The relative frequencies also show that almost nobody used the spaceFills. The reason for this was perhaps that the participants were not familiar with the spaceFill concept before the test and therefore found this visualisation difficult to use. However, the PCP was also new to all the participants, so why one of the two new visualisations was extensively used and the other one not at all, is not clear.

In this experiment, the participants used three different exploration strategies. A similar grouping was found in another experiment (Demšar 2006), where the participants used two exploration strategies that correspond to the first and the last group in this study. No indication has been found that the way the participants explored the data had anything to do with their gender, nationality, academic background or GIS experience. In this case all participants shared a similar academic background and GIS experience, yet they still adopted different exploration strategies. Also, each exploration strategy was adopted by a group consisting of one male and one female participant.

The differences in the exploration strategies of the three groups of participants might indicate that investigating data visually with the aim to form a hypothesis is not as simple as it is sometimes said to be. A similar finding
was presented by Tobón (2002), who reasoned that although humans might be good at visual pattern recognition, they need to understand and decode information from the graphic display. If the visualisation that the person is looking at is too complex, it might be difficult for certain individuals to translate what they see into a meaningful hypothesis about the data.

An interesting topic to explore would be how the cognitive exploration process differs in different types of people and how it is related to previous experience. There seem to exist substantial inter-individual differences in forming exploration strategies. The question is what defines these and if and how they are related to users’ background and experience.

ACKNOWLEDGEMENTS
The author would like to thank the participants for their willingness to take part in the experiment and their enthusiasm. This study was supported by the Helge Ax:son Johnson Foundation.

REFERENCES

Inselberg A., 2002, Visualization and data mining of high-dimensional data. *Chemometrics and Intelligent Laboratory Systems*, 60:147-159.


Examining Potential Demand of Public Transit for Commuting Trips:
an integrative analysis combining GIS, statistics, and data mining

Xiaobai Yao
Department of Geography
University of Georgia, Room 204 GG Buidling
Athens, GA. USA 30605
Email: xyao@uga.edu

ABSTRACT
The benefits of public transit system have been increasingly recognized, particularly with the rapidly elevated gas prices and the persistent efforts on environment protection. However, insufficient coverage of public transit services has been a major barrier to many potential riders. To expand the coverage of public transit systems, the fundamental question is where the expansion should go. This study presents an integrative approach to finding answers to this question. The objective is to provide measures of potential demand of public transit service for commuting trips at a fine granularity for a region, and (2) to visualize the distribution of such demands in the region. It is expected that this kind of information will be very useful for local transportation planning authorities.

The presented approach combines GIS spatial data handling capability with statistical analysis, mathematical modeling, and data mining techniques. An empirical study of Atlanta, Georgia is carried out using this approach. CTPP data at fine zonal level (such as traffic analysis zones) and existing transit network data are prepared and preprocessed in GIS for subsequent analysis. Multiple regression is employed to identify predictive variables for the share of public transportation for work trips. Given the predictive variables and the associated weights (coefficients), two independent methods are developed to examine the potential demand. One method is a mathematical formulation of a Need Index. This index measures the relative magnitude of the potential need in each zone. The second method use a data mining technique, self-organizing maps, to explore clusters in the high-dimensional vector space of the predictive variables. The data mining method will then assign each zone to the closest cluster. The results of both methods are visualized in GIS. In the empirical study, high level of agreement is observed between the results from the two methods. This agreement speaks for the validity of both methods.

Keywords: Public transit system, Self-organizing map, Data Mining, GIS, Regression

INTRODUCTION
The benefits of public transit system have been increasingly recognized in the United States. Particularly with the rapidly elevated gas prices and the persistent efforts on environment protection, private car driving becomes increasingly costly to the individual citizens as well as to the society. Public transit system is naturally the most competitive alternative transportation mode for its high efficiency in cost and reduced emission due to much
higher passenger-ride ratio. However, public transit has not been popular in most American cities. Atlanta, for example, had public transit system coverage in only three counties out of its ten-county metropolitan region by year 2000. As a result, the insufficient coverage of public transit services has been a major barrier to many potential riders. Expanding public transportation coverage has never been such an important and urgent transportation planning strategy for many an American city. But the immediate question is where the expansion should go. This study presents an integrative approach to finding answers to this question. This approach integrates the respective strengths of geographic information systems (GIS), statistical analysis, mathematical modeling, and data mining in the efforts to measure potential demands of public transit. The ultimate goal is to provide methods and tools for decisions makers to identify areas with potential demand for public transit accessibility. An empirical study of Atlanta, Georgia is carried out using this approach.

There are two types of related research threads on public transit system in the literature. The first type of studies focuses on the relationship between public transit ridership and other related factors such as the characteristics of land use, socioeconomics, and the transit systems’ structure and level of services. Previous studies have typically applied regression analysis, GIS-based models (Azar and Ferreira 1994), and the popular multinomial logit (MNL) modal split model (Ben-Akiva and Leman 1985). Another thread of related research looks at the coverage of public transit systems. A critical factor in this coverage studies is the access cost (time or distance) that one has to overcome to get to a transit service stop (Murray 2005; Murray 2001). GIS has been a major modeling environment for this type of studies with implementation of location-allocation models in it (Wu and Murray 2005). Another side of the coverage issue is the theme of people’s accessibility to public transit services. Miller (1999) identifies three major theoretical approaches to measuring accessibility. The first focuses on space-time constraints (Hagerstrand 1975; Kwan 1998), the second approach is attraction-based and takes the spatial interaction form of measures. The third is the benefits-driven approach that is based on the maximization of either the user’s or the location’s benefit. A noteworthy variation is the measure of geometric accessibility that measures relative nearness of location (Jiang, Claramunt, and Batty 1999).

This study is explicitly related to the first of the abovementioned threads of studies, while the objective is to extend this type of investigations beyond the identification of predictive factors. This will be done with the provision of information on potential demand of public transit service and with the visualization of spatial distribution of such demands. The second type of abovementioned studies, which is on coverage and accessibility, should also shed lights on this current research. Although the research analyzes relationships and potential demands at a fine aggregate level (e.g. traffic analysis zones), it should be noted that the coverage of the accessibility to the transit service stops are not interrupted by the zone boundaries.

The next section presents an integrative analysis approach to investigating public transit demand. Each associated method is elaborated. Section 3 discusses an empirical study of public transit system in Atlanta. The paper is concluded with a summary of the integrative approach in Section 4.

**METHODOLOGY**

This study takes an integrative approach to examining the potential demand of public transportation. The approach takes advantages of the spatial analysis and visualization capability with GIS, the data modeling capability of statistical analysis for the identification of predictive variables, and the knowledge learning power of data mining techniques. The different types of techniques (and therefore software programs) are integrated
through a loose coupling strategy. Most information sharing and integration are done through data file import/export functions in the software or some scripts developed by the author for the same purpose.

**Identify Predictive Factors**

The first and the most fundamental question is which factors can be used to predict the share of public transit for commuting trips for a person or group of people residing in certain spatial area. The literature has primarily suggested land use characteristics and socioeconomic characteristics of the potential riders (Ortuzar and Willumsen 2001). Meanwhile many studies also find other significantly contributing variables such as the spacing between stops, centrality, interline transferability (Kuby, Barranda, and Upchurch 2004) and level of service of the transit system. While these variables all present their high relevance in one city or another, the significance of such relevance varies among different cities or areas. Thus I will first employ a multiple regression to identify the significant variables and relative contribution (coefficient) of each variable to predict the proportion or public transit ridership in the area.

$$R = \sum_{i=1}^{k} \beta_i v_i$$  \hspace{1cm} (1)

where R is the proportion of workers taking public transit as the primary mode, $v_i$’s are the identified independent variables, and k is the total number of these variables.

**Measure and Visualize Potential Demand**

The objective of this study is to provide information of potential demand of public transit at a fine aggregate level such as census blocks or traffic analysis zones (TAZ). Two methods are applied to provide such information based on the identified contributing variables. The first method is to derive a measure of the potential demand. This is a so-called Need Index (NI) as defined below.

Let’s first divide the right side of Equation (1) into two parts, one for the transit system related factors, and the other for all other factors. This will take us to Equation (2):

$$R = \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \alpha_i y_i$$  \hspace{1cm} (2)

where $y_i$’s are variables accounting for the network structure and level of service of transit systems, and $x_i$’s are the land use and socioeconomic variables or any other contributing variables that are not about the transit systems. The first part on the right side of Equation (2) is defined as $NI$. The last part on the right of Equation (2) is about the transit network characteristics. Let us use a variable $Net$ to denote it, as shown in Equations (3) and (4).

$$NI = \sum_{i=1}^{n} \beta_i x_i$$  \hspace{1cm} (3)
\[ \text{Net} = \sum_{i=1}^{m} \alpha_i y_i \]  

(4)

Therefore Equation (2) becomes

\[ R = NI + \text{Net} \]

or

\[ NI = R - \text{Net} \]  

(5)

The Need Index, NI, can be used to indicate the relative potential demand of public transit in a region. Provided the same network condition (ie.Net part is fixed), Equation (5) suggests that to achieve the the NI and R (proportion of public transit ridership) will increase or decrease in the same direction. Higher value of NI implies greater potential of public transit ridership for work trips. From a planner’s perspective, the areas of high NI but no or poor public transit coverage are the target areas for new or improved service.

The second method uses self-organizing map (SOM), a data mining algorithm, to identify clusters of the zones in the high-dimensional variable space. For each zone, we have a vector of the identified contributing variables, \(<x_1, x_2, \ldots, x_n>\). The \(x_i\)'s are the same variables as those in Equation (3), namely all the other contributing variables except those about the transit systems. These predicting variables each have a different degree of relevance (coefficients) to the public transit ridership. In addition, some variables may be entered with any unit of measurements, which will lead to different coefficients for the variable. To account for these, I modify the variable vector space into a weighted version as follow:

\[ <\beta_1 x_1, \beta_2 x_2, \ldots, \beta_n x_n> \]  

(6)

SOM (Kohonen 1995) is both a dimension-reduction method and a clustering method through unsupervised learning. The SOM algorithm maps high-dimensional data space into two-dimensional space, and at the same time it is a clustering method so that similar data samples tend to be mapped to the same or nearby neurons (Kohonen et al. 2000). Comparing to other dimension-reduction techniques such as Multi-dimensional Scaling, SOM can not only reduce dimensions but also display similarities, as nearby neurons (clusters) are similar than the dissimilar ones.

In the SOM algorithm, the inputs are the high dimensional vectors such as that expressed in (6). Each output node is also a vector of the same dimensions as that of the inputs. The output nodes are organized into an m by n 2-dimension output map. The size of the output map (m and n) is user-defined. The vector associated with each output node is called “weight” of that node. Initially, the weights of the output nodes are randomly generated. The SOM algorithm then presents each input vector to train the output nodes. At the end of the training, each output node will anchor a respective cluster. Each time an input vector is presented, the algorithm will compute the Euclidean distance \(d_j\) between the input vector and each output node \(j\).

\[ d_j = \sqrt{\sum_{i=1}^{N} (y_i(t) - w_{ij}(t))^2} \]  

(7)
where $y_i(t)$ is the value of $ith$ dimension of the input vector at time $t$, while $w_{ij}(t)$ is the value of the $ith$ dimension of the weight of the output node $j$ at time $t$. Here time means the count of iterations. One iteration is completed when all input vectors have been presented once to train the output weights.

The output node that gives the minimum distance $d$ is the winning node. Select the winning node $j^*$ and update weights to nodes $j^*$ and its neighbors using Equation 7.

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)(x_i(t+1) - w_{ij}(t))$$

(8)

where $\eta(t)$ is an error-adjusting coefficient ($0<\eta(t)<1$) that decreases over time. Usually multiple iterations are necessary until certain matching criteria are met.

Once the training process is completed, the $m^* n$ nodes in the output map represent the clusters. Any vector in the same multi-dimensional space can be presented to the established SOM and be assigned to the closest cluster.

These two methods are complementary to each other. The Need Index method gives a specific numeric measure for each zone, so that the potential demand of public transit of individual zones can be compared. It is also easy to apply a classification method to put these zones into categories according to the index. However, the Need Index is calibrated based on the assumption of linear relationship between the dependent and predictive variables. SOM avoids that assumption and identifies clusters in the multi-dimensional variable vector space. Since both methods allow the derivation of clusters/classes in the high-dimensional variable vector space, the results from both methods can be visualized and compared. A good agreement between the results from the two methods can greatly speak for the validity of the results, and vice versa.

**CASE STUDY**

The study area is the metropolitan of Atlanta in Georgia, a major employment center in southeast United States. This study concerns the 10-county metropolitan area in which regional planning and intergovernmental coordination is managed by the Atlanta Regional Commission (ARC). Atlanta has experienced rapid population growth and the associated problem of traffic congestion. In response to this, the new long-range regional transportation plan, named “Mobility 2030”, has been released in which the regional transit aims to play a major role in alleviating congestion and improving air quality.

The ten counties and the public transit systems in the year of 2000 and 2005 are displayed in Figure 1. From the figure, noticeable expansion of the transit system can be found over the last past five years. This study will examine the relationship of public transit ridership for commuting trips, and aim to predict potential demand of public transit. The hope is that the measures and visualization of the spatial distribution of such potential demand could provide useful information in future transit system expansions.
Data and Preliminary Data Processing

The following two types of data sources are used for the empirical study.

1. The United States Census for Transportation Planning Package (CTPP) 2000 data.
2. The transit lines and transit stops GIS dataset provided by the Atlanta Regional Commission.

Both datasets are examined and pre-processed in ArcGIS.

The TCPP 2000 census data provides population, socioeconomic, and transportation related data aggregated at county, block group, traffic analysis zone (TAZ) or other place levels. I use the data at TAZ level because it has finer granularity than the level of census block. In addition, TAZs are divided according to transportation related criteria. There are totally 1593 TAZs in the 10-county study area. About 180 of them are excluded during the data pre-process because data for some variables are empty for the TAZ or because inconsistency is found with the data about of TAZs. As a result, 1417 TAZs are retained for the study. For each TAZ, I extracted the variables including population, number of households, number of jobs, number of workers, number of workers taking public transit as the primary transportation mode, numbers of workers who are in each of the three income levels respectively, and number of workers who have 0, 1, and 2 or more vehicles in the household. These variables are then processed in ArcGIS into percentages or density type of measurements.

The transit systems, including both transit lines and stop/station location, are provided in shapefile format. In data pre-processing using ArcGIS, I first calculate the number of transit stops/stations within each TAZ. This measure is further divided by the area of the TAZ, which converts the measure into stop-density.
Statistical Analysis

A multiple regression is conducted to identify the predictive factors for public transit ridership for commuting trips in Atlanta. Aggregated at the TAZ level, these variables are extracted or calibrated from CTPP 2000 data as mentioned above. The dependent variable is the percentage of workers using public transit as their primary transportation mode for work trips. Three types of independent variables are entered in the analysis.

1. Land-use characteristics
   - Population density
   - Average number of workers per household
   - Employment rate (percentage of people who have jobs)
   - Job density (total jobs in the TAZ divided by the area of the TAZ)
   - Percentage of home workers

2. Socioeconomic characteristics
   - Income: Percentages of workers in three income status (below poverty line, between 100-150% of the poverty line income, above 150% of the poverty line income)
   - Car ownership: Percentages of workers having 0, 1, 2+ vehicles in respective households

3. Network structure
   - Density of bus stops in the TAZ (number of bus stops divided by the area of the TAZ)
   - Density of rail stations in TAZ (number of rail stations divided by the area of the TAZ).

It should be noted that I converted all variables into rate, percentages, and density. It is done so that variations in size of the TAZs would not matter. In the preliminary test, two variables (the 150% income category and the 2+ vehicles in a household) are excluded in the regression because of identified collinearity problem with them. Table 1 shows the regression result after taking care of the collinearity issue. The table shows that the share of public transit is significantly related to all factors except the average number of workers per household, the percentage of home-workers, and rail density (grayed out in the table).

Table 1. Regression Results for the prediction of share of public transit for work trips

<table>
<thead>
<tr>
<th>Predictive Variables</th>
<th>(Unstandardized) Coefficients</th>
<th>Sig.</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Tolerance</td>
</tr>
<tr>
<td>(Constant)</td>
<td>1.334</td>
<td>.824</td>
<td>.106</td>
</tr>
<tr>
<td>Percentage of home workers</td>
<td>.008</td>
<td>.034</td>
<td>.816</td>
</tr>
<tr>
<td>Percentage of workers below poverty line (x1)</td>
<td>.074</td>
<td>.019</td>
<td>.000</td>
</tr>
<tr>
<td>Percentage of workers with income from 100% to 150% of poverty line (x2)</td>
<td>.103</td>
<td>.026</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>.421</td>
<td>.017</td>
<td>.000</td>
</tr>
</tbody>
</table>
At the significant level of 99%, eight variables are identified as predictive factors for the share of public transit in Atlanta’s commuting trips. They are the percentage of workers below poverty line, percentage of workers with income from 100% to 150% of poverty line, percentage of worker with 0 vehicle in the household, percentage of worker with 1 vehicle in the household, employment rate, population density, job density, and the density of bus stops in the zone. The R square of the model is 0.699, which tells us that about 70 percent of the variation in public transit shares for work trips can be explained by the model.

### Potential Demand Analysis and Visualization

Two types of demand analysis are carried out and compared through visualization. The need index of each TAZ in Atlanta is calculated following Equation (3).

\[
NI(i) = 0.074x_1 + 0.103x_2 + 0.421x_3 + 0.033x_4 - 0.045x_5 + 0.036x_6 - 0.026x_7
\]

where \(x_1\) through \(x_7\) are the variables as defined in Table 1, the \(NI(i)\) is the need index measure for the \(i\)th TAZ in Atlanta ARC counties. The computed measures are visualized in Figure 2(a). In this figure, the measure are classified into nine categories using Jenk’s Natural Break classification method (Jenks 1963). Although it would perceptually more appropriate to visualize four to six categories, Figure 2(a) displays a nine-category spatial distribution of the need index to make it more comparable to the nine-cluster view that is generated from a data mining approach.

The need index is based on the linear relationship assumption that is rooted in the linear regression. A second method is then applied without such an assumption. This second approach is the SOM algorithm. The result of
SOM is shown in Figure 2(b). The results generated from the two methods seem to agree to each other to a great extent.

(a). Result from the Need Index method  
(b). Result from the SOM method

Figure 2. Spatial Distribution of Potential Demand of Public Transit for commuting trips

The SOM method is calibrated from the seven out of the eight identified predictive variables. The size of the output map is arbitrarily defined as a 3 by 3 grid map. Figure 3(d) is the spatial configuration of the output nodes. The network variable, bus stop density, is excluded because we want to see the potential of demand, not the actual demand that is heavily influenced by the actual transit system itself. Coefficients in Table 1 are used to generate the weighted vector space according to (6) in Section 2. Constrained by the limited number of dimensions that can be visually depicted, Figure 3(a) displays the distribution of the 7-D input vectors in a 3-D (the first 3 dimensions of the 7-D) space. During training period with the 1417 input vectors (each corresponds to a TAZ), the output nodes are updated according to Equation (7) and (8). At the end of the training, nearby output nodes in Figure 3(4) will be more similar to each other than farther away pairs. The final output nodes (7D) are shown in a 3-D space in Figure 3(b). Figure 3(c) plots the input and output in the same 3-D space.
Figure 3. Input and Output nodes in the SOM method

The high level of agreement between the results from two methods reinforces the validity of both methods. The spatial pattern of potential demand revealed in Figure 2 can be examined against the current transit provision as shown in Figure 1. To make the examination easier, Figure 4 visualizes all the necessary information in a layered 3-D perspective view. The figure displays the potential demand distribution in 2000 (bottom layer), the 2000 transit system (middle layer), and the added transit lines between 2000 and 2005 (top layer). It is obvious that the expansion of transit systems in Atlanta from 2000 to 2005 have addressed many areas that have higher potential demand and were not served in 2000. However, some patches that have high potential demand are identified as still not being served in current system. These areas are in the northern most TAZs and some in the south and southwest regions. These areas could be the priority zones for future transit expansion in the ARC region.
The transit lines added between the years of 2000 and 2005

The transit system in the year of 2000

Potential demand distribution in the year of 2000

Figure 4. A layered 3-D perspective view for the examination of potential demands and current coverage of public transit system in Atlanta

CONCLUSIONS

This study investigates an integrative approach to exploring the relative magnitude and spatial distribution of potential demands of public transit service. GIS spatial data handling capability is combined with statistical analysis, mathematical modeling, and data mining techniques. In this approach, the integration of GIS and other spatial data analysis and modeling techniques are carried out through loose coupling. GIS serves as the central data center where all relevant data and attribute data are integrated and pre-processed for further data handing from various sources. The processed data are then imported into other data analysis programs through existing program functions or through scripting.

The study once again proves that multiple regression analysis is very useful in finding predictive variables for the share of public transit in people’s commuting trip mode choices. Eight variables of land use, socioeconomic, and transit system structure characteristics are identified in the statistical analysis. The regression model explains 70 percent of the variance in the share of public transit for all commuting trips.

Two methods are proposed to investigate the potential demand of public transit at an aggregate level (TAZ in this study). One is a mathematical modeling method, while the other is a data mining approach. The mathematical method provides a so-called Need Index to measure the relative magnitude of potential demand. The advantage of this method is the availability of numeric measure for each zone, while the disadvantage is the assumption of the form of the relationships. Self-organizing map is the employed data mining method which reduces high-dimensional vector in the predictive variable space into a 2-D space. The advantage of this method is that it is free of any prior assumption of either the data distribution or the form of the relationship among the variables. In addition, the similarity and/or dissimilarity between any two clusters can be qualitatively estimated according to the relative distance between the two clusters (output nodes) in the SOM output map.

The
The disadvantage of SOM method is that an overall ranking of the clusters is hard to achieve without extra information. From the case study, the results of both methods agree to each other to a great extent. This agreement speaks highly of the validity of modeling results of the two methods. It is concluded that an integrative approach to data exploration and visualization is very beneficial in this type of study. The benefits are due not only the combinational advantages from a variety of techniques but also to the mutual testing / confirmation possibility for results from each method.

REFERENCES


A Novel Accessibility Measure using Travel Time and Space Syntax

Indices

Seungjae Lee, Byoungwook Lee, Chulmin Jun and Seungil Lee
The University of Seoul

Shinhae Lee
Seoul Development Institute
E-mail: sjlee@uos.ac.kr

ABSTRACT

Accessibility generally means the degree of availability for accessing to active places and areas. In transportation fields it also means the degree that traffic modes access to certain areas. In other words, any area with high accessibility has many traffic modes that run to other areas from it. Seoul city has reorganized its bus systems as an hierarchical system since 1st, July, 2004. The Seoul bus reorganization is evaluated for an improved function of corridor, accessibility, efficiency and comfort of the bus. The Metropolis of Seoul changed public transportation convenient and rapid by rearrangement of bus lines with median bus lines, which is the priority policy of public transportation. Policy planners and policy makers have subjectively evaluated public transportation so far because there was no quantitative analysis tool.

The paper has therefore developed a methodology of calculating accessibility in bus network by using a space syntax technique. The purpose is an effective analysis of accessibility improvement before and after Seoul bus reorganization. Space syntax methodology has an advantage to calculate easily network connectivity and accessibility. However, it also has a weakness not to consider the distance and travel time for calculating accessibility. As a result, we have developed a hybrid of the space syntax methodology and the travel time measures. This method has been applied to evaluate the accessibility before and after the hierarchical bus systems introduced in Kangnam area of Seoul.

Keywords: public transport, space syntax, accessibility, hierarchical bus system

INTRODUCTION

Accessibility generally means the degree of availability for accessing to active places and areas. In transportation fields, it also means the degree that traffic modes access to certain areas. In other words, any area with high accessibility has many traffic modes that run to other areas from it. Traffic modes are to be classified into private transportation mode and public transportation mode. As the quality of life increases, demand of private transportation increases. It has caused a big social problem that leads to an increase in travel times and social costs because of the congestion in limited spaces. The latest traffic policies are changing the focus from the
private transportation to public transportation in order to use the limited spaces more efficiently. Public transportation is not only the most important way to solve chronic congestion and environment problems in urban traffic but also one of the fundamental civic rights because the management and level of public transportation service affect the city-dweller’s living. The executive organ of policy is aiming at priority of public transportation that guarantees services on time, comfortable riding and safety. The Metropolis of Seoul reorganized from the existing complicated and inconvenient bus system to efficient mass bus system on the first of July in 2004. Figure 1 describes theoretical background of the Seoul bus reorganization.

![Figure 1. Background of feeder line and corridor line](image)

Hierarchical line has a transfer between feeder line and corridor line. Hence, the average speed of the previous line is predominant over the hierarchical line up to the critical point, but as hierarchical line fulfills its function, travel time of hierarchical line is finally shorter than the one of the previous line. The Seoul bus reorganization is evaluated for an improved function of corridor, accessibility, efficiency and comfort of the bus. The Metropolis of Seoul changed public transportation convenient and rapid by rearrangement of bus lines with median bus lines, which is the priority policy of public transportation. Policy planners and policy makers have subjectively evaluated public transportation so far because there was no quantitative analysis tool.

The paper develops a methodology of calculating accessibility in bus network by using Space Syntax. The purpose is an effective analysis of accessibility improvement before and after Seoul bus reorganization. In section 2, existing accessibility studies are briefly reviewed. In section 3, the new accessibility measures using space syntax and travel time indices are introduced. In section 4 a proto-type example of calculation using the new index is presented and then in section 5, before and after the hierarchical bus systems introduced is evaluated in sub-areas of the Seoul network. In section 6, a brief summary of this paper is shown.

**EXISTING STUDIES**

**Definition of Accessibility**
As stated above, the general meaning of accessibility is the degree that signifies how easy one can access for any purpose to any place by any modes in any time. It also can be the degree of the ability to access to CBD, some major points or areas. The paper defined accessibility as access between bus stops. It can also be used for same meaning as serviceability of public transportation modes. The review of defining accessibility is described below.

**Table 1. Defining Accessibility**

<table>
<thead>
<tr>
<th>Author</th>
<th>Definition</th>
<th>Key factors in accessibility</th>
</tr>
</thead>
</table>
| Litman, 2003a | Accessibility refers to the ability to reach desired goods, services, activities and destinations (together called opportunities). Accessibility depends on mobility, mobility substitutes and opportunities as follows:  
  • Mobility - provided by walking, cycling, public transport, car sharing, taxi, cars, and other modes. All else being equal, an increase in the speed, service quality or affordability of a mode will improve access by that mode.  
  • Mobility substitutes - telecommunications and delivery services. These can provide access to some types of goods and activities, particularly those involving information.  
  • Land uses - the geographic distribution of activities and destinations. The dispersion of common destinations increases the amount of mobility needed to access goods, services and activities, reducing accessibility. When real estate experts say “location, location, location” they mean “accessibility, accessibility, accessibility”.  
  • Other factors - information availability, affordability, convenience and comfort, security and prestige. || “All else being equal” – Responses to mobility changes can mean that things are usually not “equal” and often complex behavioural responses need to be considered.  
The need to consider information, availability, comfort, security, prestige, speed, modes available, telecommunications, land uses and all potential activities emphasis that it is important not to confuse detailed components of accessibility with the definition “the ability to reach goods services, activities and destinations”. |
<p>| David Simmonds Consultancy et al (1998) | A way of measuring the ease with which a particular category of persons can reach a defined set of destinations, from a given origin (origin accessibility), or the ease with which a given destination (destination accessibility) can be reached by a particular set of potential individuals. | Different people have different levels of accessibility so definitions relate to a “particular category of persons” |
| SEU Report, 2003 | The ability of people being able to get to key services at reasonable cost, in reasonable time and with reasonable ease. | Key words “ability” and “reasonable” need to be defined. |
| DfT, 2001 | Difficulties with boarding and alighting vehicles, carrying items, confusion over use and staff attitudes. | Physical access to vehicles and transport providers understanding their customers are important. |</p>
<table>
<thead>
<tr>
<th>Source</th>
<th>Definition</th>
<th>Accessibility Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geurs et al, 2001</td>
<td>The extent to which the land-use transport system enables (groups of) individuals or goods to reach activities or destinations by means of a (combination of) transport mode(s).</td>
<td>Accessibility is best considered by a combined view of modes.</td>
</tr>
<tr>
<td>Ross, 2000</td>
<td>The ease of reaching some destination, and may include real or perceived costs in terms of time or money, distance travelled, level of comfort, availability and reliability of public transport, or any combination of these.</td>
<td>Real and perceived barriers are important.</td>
</tr>
<tr>
<td>Gray, 1989</td>
<td>A measure of the relative access of an area or zone to population, employment, opportunities, and community services</td>
<td>Comparative accessibility can be as important as absolute levels.</td>
</tr>
<tr>
<td>Handy, 2004</td>
<td>The ability to get what you need, ideally with a choice of destinations and using a choice of modes</td>
<td>“Need” and “ideally with a choice” are important concepts</td>
</tr>
</tbody>
</table>


### Existing accessibility index

Existing accessibility indexes are generally used for accessibility evaluation of the city. They can be classified into three measures, which are simple measure, opportunity measure, and value measure (see Halden, D., McGuigan, D., Nisbet, A. and McKinnon A., 2000)

Firstly, simple measure has simple methods to calculate and select variations. Various time and distance thresholds have been adopted within these measures. The accessibility measure for a location \( i \) is calculated as the sum of the opportunities available locations \( j \) within the defined threshold. However everyone has same opportunity regardless of travel distance and travel time.

\[
A_i = \sum O_j \delta_{ij}
\]  

Where,

\( O_j \) – accessible opportunity of destination \( j \)

\( \delta_{ij} \) – if there is a accessible opportunity, then 1 and if not, then 0

Secondly, opportunity measure regards travel distance and travel time as important factors and considers them by each access opportunity that the more travel distance and travel time, the less access opportunity.

\[
A_i = \sum O_j \exp(-\lambda t_{ij})
\]  

Where,

\( O_j \) – accessible opportunity of destination \( j \)
Lastly, value measure adds consideration of weight by each opportunity to second measure. Accessibility is calculated with the weight and general time between origin and destination as follows.

\[
A_i = \frac{1}{\lambda} \ln \frac{\sum \exp(\lambda t_{ij})O_j}{\sum O_j}
\]

(3)

Where,

\[O_j\] – accessible opportunity of destination j

\[\lambda\] – parameter

\[t_{ij}\] – travel time between i and j

DEVELOPMENT OF ACCESSIBILITY INDEX USING SPACE SYNTAX AND TRAVEL TIME

Space syntax

Space syntax was developed in 1980 by Professor Hillier of University College London in England. Space syntax is a method to analyze and describe physical structure of space. We can reinterpret space in a mathematical logic and computer performance. Module of space syntax is classified into axial analysis and convex analysis.

1) Axial Analysis
   - It represents physical structure of space by axial line, which connects the maximum points viewed by human eyes and analyses how axial lines connect to each other.

2) Convex Analysis
   - It represents physical structure of space by convex space and analyses how convex space connect each other.

This paper chose the axial analysis to evaluate Seoul bus reorganization because the traffic network can be described by node and link.
One network is described by one axial line in Space syntax. There are six networks that can be represented to six axial lines in the figure 2. It can be described again on the graph of space conception as follows.

We can directly move from link 1 to link 2 or link 3 in this space (figure 2). It comes under step 1 in space syntax. We can move from link 1 to link 4 or link 5 via link 3. It comes under step 2 that has one turn. We can move from link 1 to link 6 via link 3 and link 5. It comes under step 2 with two turns. Figure 2 describes the concept of steps in space syntax.
There is a major concept of a depth in space syntax. The depth can be determined by counting the number of nodes and steps (or turns) between certain two nodes. Equation of the depth is as follows.

\[ TD_i = \sum_{s=1}^{m} s \times N_s \]  \hspace{1cm} (4)

Where,

- \( TD_i \): the total depth of node \( i \)
- \( s \): the step from node \( i \)
- \( m \): the maximum number of steps extended from node \( i \)
- \( N_s \): the number of nodes at step \( s \)

Axial analysis including the depth is calculated in a very simple manner. That is why axial analysis can be easily applied on complex network and accessibility degree can be found through it. While using axial analysis on the traffic network, there are a few things to consider. In traffic network, especially bus network is described by bus stops and links. While space syntax is described by the axial line viewed from human eyes. The space syntax value is different depending on the location of the step, which means transfer in bus network.

\[ SS : 9 \quad \text{SS} : 7 \]

Figure 4. Space syntax value depending on the location of transfer

The number of the transfer is more sensitive than the location of the transfer in a traffic network. So this paper changes from the axial based space syntax to the link based space syntax value by performing an inner product operation.

**Inner Product Operation**

An inner product operation is a theory in which two different vectors decide a certain vector value. In order to get the inner product value by using the space syntax, we assume that one axial line is affected by other axial lines to cross in space syntax.

\[ \begin{align*}
\mathbf{a}_1 & \quad \mathbf{a}_2 \\
\cdots & \quad \cdots \\
\mathbf{a}_9 \\
\end{align*} \]

Figure 5. Example of network for Inner Product
Every axial line in figure 5, which is from $a_1$ to $an$, has each of space syntax value. The inner product value of the link between two axial lines is calculated with the value of the axial line and its weight as follows.

$$IP_{a_i-a_{i+1}} = \frac{\sqrt{a_i^2 + a_{i+1}^2}}{\sqrt{\sum_{k=1}^{n-1} a_k^2 \times W_i}} \times a_i$$

(5)

Where,

- $IP_{a_i-a_{i+1}}$ – Inner Product between $i$ & $i+1$
- $a_i$ – Value of axial line $i$
- $W_i$ – Weighting factor
- $n$ – Number of total crossing line

Weight factor ($W_i$) means the value when the average of each link value ($a_2$ to $a_2$) is same as the value of axial line 1 ($a_1$).

$$W_i = \frac{\sum_{k=2}^{n-1} \sqrt{a_k^2 + a_{k+1}^2}}{\sqrt{\sum_{k=1}^{n-1} a_k^2 \times (n-1)}}$$

(6)

$$a_i = \frac{\sum_{i=2}^{n-1} IP(a_i-a_{i+1})}{n-1}$$

(7)

Accessibility is finally evaluated with the space syntax value of each link and travel time between nodes.

$$Sb_{i,j} = \alpha(IP_{i,j}) + \beta(TTi_{i,j})$$

(8)

Where,

- $Sb_{i,j}$ = serviceability from $i$ to $j$
- $IP_{i,j}$ = Space Syntax value using Inner Product from $i$ to $j$
- $TTi_{i,j}$ = Travel Time from $i$ to $j$
α value and β value are calculated while considering the weight of the inner product value, travel time, and a bi-criteria method. The bi-criteria is a method in which two results of variations, which have different units, were merged into one result with their correlation. This paper’s process to evaluate accessibility of network is as follows.

**PROTO-TYPE NETWORK EXAMPLE**

Proto-type network for examination of the developed methodology is as follows.

There are nine bus stops in the toy network and bus stop numbers 2, 3, 4, 5, 7 and 8 are for transfer. Space Syntax values and Inner Product values of each bus line are as follows.
Accessibility can be evaluated in two viewpoints, which are accessibility of certain bus stop in total network and accessibility between two certain bus stops. Accessibility of certain bus stop in the total network is calculated by summation of accessibility values from it to all other bus stops. Accessibility between two certain bus stops is used for the evaluation between major places. We assume that weight of $\alpha$ value is same as one of $\beta$ value so we use 1 for $\alpha$ value and $\beta$ value. We also assume that travel speed of link is 20km/h to produce the travel time. The accessibility of every bus stop in total network is as follows.

Table 2. Accessibility represented with Inner Product and travel time

<table>
<thead>
<tr>
<th>Point to others</th>
<th>Inner Product</th>
<th>Travel time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop.1</td>
<td>22.23</td>
<td>48.15</td>
</tr>
<tr>
<td>Stop.2</td>
<td>15.23</td>
<td>33.50</td>
</tr>
<tr>
<td>Stop.3</td>
<td>14.23</td>
<td>37.72</td>
</tr>
<tr>
<td>Stop.4</td>
<td>17.23</td>
<td>37.94</td>
</tr>
<tr>
<td>Stop.5</td>
<td>18.01</td>
<td>43.50</td>
</tr>
<tr>
<td>Stop.6</td>
<td>20.88</td>
<td>52.65</td>
</tr>
<tr>
<td>Stop.7</td>
<td>20.04</td>
<td>41.83</td>
</tr>
<tr>
<td>Stop.8</td>
<td>21.99</td>
<td>49.94</td>
</tr>
<tr>
<td>Stop.9</td>
<td>26.34</td>
<td>61.83</td>
</tr>
</tbody>
</table>
The result is that bus stop number 2 has the lowest Space syntax value and travel time, which means that bus stop number 2 has the biggest accessibility from all other bus stops in this network.

**APPLICATION ON SEOUL BUS NETWORK**

In order to evaluate the effectiveness of Seoul bus reorganization, this paper computes accessibility before and after Seoul bus reorganization by using the developed methodology. The spatial area is Kangnam in Seoul, Korea. The bus travel speeds are 20km/h in Kangnam arterials on which median bus lines are operating, and 15km/h in the rest of the roads in Kangnam.
We compare accessibilities of a few major places from Jamsil before and after Seoul bus reorganization. The chosen major places are the Express Bus Terminal Station, Kangnam station, Seoul National Education University and Apgujeong. The result is as follows.

<table>
<thead>
<tr>
<th>Index</th>
<th>Inner Product</th>
<th>Travel Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamsil</td>
<td>15.74</td>
<td>35.71</td>
</tr>
<tr>
<td>Kangnam St.</td>
<td>16.93</td>
<td>37.48</td>
</tr>
<tr>
<td>National Univ. of Ed.</td>
<td>19.31</td>
<td>36.28</td>
</tr>
<tr>
<td>Apgujeong</td>
<td>26.97</td>
<td>96.78</td>
</tr>
</tbody>
</table>

**CONCLUSION**

This paper developed the methodology to calculate accessibility by using Space Syntax and travel time. We assume \( \alpha \) and \( \beta \) of the inner product value and travel time as an unity. We can understand that the lesser the inner product value and travel time, the better the accessibility. The developed methodology leads to the evaluation of the improved bus service and new traffic modes introduction effect. However, further studies are necessary in order to overcome the complexity of the calculation process.
REFERENCES


The Implication of a City’s layout’s Visibility on Wayfinding Performance

Itzhak Omer and Ran Goldblatt
Department of Geography and Human Environment, Tel Aviv University
Tel Aviv 69978, Israel
Tel.: +973(3)6406841. Fax:+972(3)6406243
omery@post.tau.ac.il; ran@eslab.tau.ac.il

ABSTRACT
We investigated the effect of different visual conditions in a simulated urban layout on acquisition of spatial knowledge. An experiment was carried out by means of a 3D virtual environment to test two properties of visual integration: overlapping between visual fields and the topological depth between these visual fields (length of visual chain). The results show that a high degree of visual fields’ overlapping (more common visible elements) between an origin and a target element helps people to construct (ordinal) procedural spatial knowledge, especially when the overlapping is a direct one (with one visual chain step) or with a short topological depth. It was also found that visual overlapping contributes to the construction of topological configurational knowledge.

1. INTRODUCTION
A city’s layout's visibility is the visual conditions of the urban environment such as the number and structure of perspective points or the scope of visual fields and their integration. Generally speaking, these visual conditions have the potential to effect the quality of spatial knowledge acquired by human subjects and their ability to construct a cognitive spatial representation of a given environment. Since human orientation partly depends on our ability to remember sets of visual landmarks and imagine their spatial relationship, visual conditions play a key role in the construction of a cognitive spatial representation. Moreover, Cornell et al. (1994) suggested a theoretical model for investigating wayfinding performance and learning process of an environment through direct experience (Navigation-based learning). According to this model, wayfinding is not merely affected by visual properties; visual properties function as building blocks, i.e. wayfinding is based on ordered recognition of familiar vistas or views along a route. In this respect, Oman et al. (2000) point out that “human keep track of their orientation and location by a normally effortless and reliable sensory integration process, even when visual cues are momentarily absent. Nonetheless, people occasionally need to reorient themselves when they view a familiar environment from an unfamiliar direction…. in such situations, the ability to imagine the spatial structure of an environment from a different direction is presumably important” (2002, p.355-356). A similar idea was suggested by Kuipers and Levitt (1988) who argued that integrating vistas by the observer may also serve as building block for a construction of configurational knowledge of a large-scale space from various observations i.e. the observer infers spatial structure from perceptions.
Previous studies have already found that cognitive spatial representation is influenced by the number of visual perspectives and the extension of their visual fields (Janzen et al., 2001), and by the integration between these visual fields. One of the main contributions to the understanding of how visual integration of spatial structure effects wayfinding performance is given by the research approach of space syntax (Penn, 2003). The important property from the perspective of this approach is the intelligibility of an environment, which is defined as the degree of correlation between the local integration of a spatial unit or its connectivity (the number of spatial units directly connected to a given spatial unit) and between its global integration (the average topological depth of the spatial unit from all other spatial units within a given system). Several researches have shown that a more intelligible environment helps people find their way more easily (Hillier et al., 1993, Bafna, 2003; Penn, 2003). Other researches have found that a high degree of global integration (short topological depth of a given units from all other visual units) contributes to the imagability of urban objects or street segments (Shokouhi, 2003). In this respect it should be noted that a highly imageable or legible city could positively effect peoples ability to orientate themselves and find their way in the city (Lynch, 1960). This understanding has inspired new approaches in the investigation of the way people use their visual senses when navigating (Kuipers, 1978) as well as their behavior with respect to the visual fields (Turner and Penn, 2002).

However, while the concept of visual integration has been examined with focus on the property of topological depth between spatial units, less attention has been paid to the property of the overlapping between visual fields. The specific aim of the study presented in this paper is to examine how the overlapping between visual fields and the topological depth of this visual overlapping (length of visual chain) affects the ability of people to acquire procedural spatial knowledge through direct experience (primary learning). For that purpose we had conducted experiments by using a virtual environment constructed specifically for that aim.

In the next section, we present a methodology for evaluating the influence of direct and indirect visual overlapping (overlapping between visual fields of salient landmarks) on the ability of people to acquire spatial knowledge during wayfinding tasks. In the third section, the objective quantative and subjective results of this examination is described. The conclusions and their significance are discussed in the concluding section.

2. METHODOLOGICAL FRAMEWORK

While virtual environments lack a certain degree of validity regarding the examination of the difference between cognitive mapping in real and in virtual environments, they do offer a tool that allows a control of the environments’ layout and a detailed documentation of the participants behavior (Williams et al., 1999). In the current research it was important to control the number and the structure of the landmarks that could be seen from any reached viewpoint and to examine the behavior of the participants during various wayfinding tasks. We were especially interested in the participant’s ability to choose the shortest path between any two given landmarks, as an indication for a procedural spatial knowledge.

For this aim, a 3D Virtual Environment (VE) of an imaginary small scale urban area of about 0.25 Square Km was built using Skyline® 4.6 software (see Fig. 1). As illustrated in figure 2 seven urban landmark elements with unique typical textures were used in the experiment: city square, clinic, restaurant, historic building, school, commercial center and a tree. The VE was designed with respect to two visual properties: overlapping between visible fields of landmarks (the number of common visible elements from any examined pairs of landmarks) and the length of the visual chain of those pairs (the topological depth). For that purpose, a different
number of landmark elements were visible from each of these elements, e.g. the city square was identified by the largest number of visible elements (four elements: clinic, restaurant, school and a tree), while the commercial center allowed only one element to be visible from it (the tree).

Fig. 1: Snapshots of the 3D Virtual Environment (VE) of the imaginary small scale urban area. built using Skyline® 4.6 software

Fig. 2: The seven landmark elements with unique typical textures that were used in the experiment: city square, clinic, restaurant, historic building, school, commercial center and a tree.

The experiment
14 participants (7 males and 7 females), at an average age of 29.8 (std. 4.5) took part in the experiment. All participants were graduate and undergraduate students at the department of Geography and Human Environment of Tel-Aviv University. The experiments were conducted on a 19” desktop monitor at the Environmental Simulation Laboratory. All participants were explained that after a learning phase on an
imagined 3D virtual city, they will be asked to perform a few wayfinding tasks and to be able to navigate from one landmark element to another. The experiment includes two phases: a learning phase and a wayfinding task phase.

**The learning phase:**
The participants were introduced to the VE through three rounds of 10 minutes recorded tour around the imaginary city. Each tour simulates a movement in an average speed of 14 kilometers per hour from a human’s eye perspective view (height of 2m). The participants were asked to carefully pay attention to the exact location of each landmark element, so that at the end they will know how to reach one from another. An important condition was taken into consideration when designing the route of the tour (see figure 3): the route should never go through the shortest path between any two examined landmark elements. This allowed us to later validate whether a procedural spatial knowledge was developed.

![Legend](image)

- 1- Clinic
- 2- Commercial center
- 3- Tree
- 4- School

Fig. 3: The path thru which the virtual city was learned.

Whenever a landmark element was reached during the tour, a two 360˚ round turns were presented to the participants so they will be familiar with all elements visible around the reached element. During the first round of the learning tour, the researcher verbally told them which landmark elements they are observing (the name of the element they have reached as well as the names of the visible elements they were observing). During the second and third learning rounds, the participants observed the recorded learning tour by themselves and the researcher did not verbally mention which elements they were observing. The participants were asked after each round to estimate on a scale of 1 to 5 whether they feel they are familiar with the city’s layout and with the relative position of the elements. After the participants have observed the three rounds of the recorded tour, the wayfinding task phase has begun.

**Wayfinding task phase:**
Five wayfinding tasks were examined (see table 1). In each task the participants were given the following instruction: “you are currently facing landmark element X. please guide the researcher how to reach landmark Y using the shortest path”. This method was chosen in order to prevent differences between participants in using the input device which could influence navigation performance. The starting point at each origin element was always the point from which all visible elements could be observed. The five tasks were presented to the participants in a random order.
As mentioned above, each task can be characterized by two properties: the number of common visible elements and the length of the visual chain between each examined pair of the landmark elements i.e. by a different number of visible elements that the origin and the target landmark have in common and by a different number of visual steps needed to be taken when reaching the target.

As illustrated in table 1, tasks 1, 2 and 3 (pairs ST, RS and HS) were designed to examine the property of the number of common visible elements. While these pairs are characterized by one visual length step they differ in the number of common visible elements between them. On the other hand, the two remained pairs (pairs 4 and 5: HC and CC) have no common visible elements; they are characterized by a visual length of 2-steps and 3-steps (through one common element along the undirected chain). That is, although these pairs do not share any common visible elements, they do share a transitive connection.

<table>
<thead>
<tr>
<th>Task</th>
<th>Origin</th>
<th>Target</th>
<th>Common visible elements</th>
<th>Length of the visual chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>(ST) School</td>
<td>Tree</td>
<td>1</td>
<td>1 (square)</td>
</tr>
<tr>
<td>2.</td>
<td>(RS) Restaurant</td>
<td>School</td>
<td>2</td>
<td>1 (historic building or square)</td>
</tr>
<tr>
<td>3.</td>
<td>(HS) Historic building</td>
<td>Square</td>
<td>3</td>
<td>1 (school or restaurant)</td>
</tr>
<tr>
<td>4.</td>
<td>(HC) Historic building</td>
<td>Commercial center</td>
<td>0</td>
<td>3 (restaurant and square and tree)</td>
</tr>
<tr>
<td>5.</td>
<td>(CC) Clinic</td>
<td>Commercial center</td>
<td>0</td>
<td>2 (square and tree)</td>
</tr>
</tbody>
</table>

Table 1: The five wayfinding tasks with the two visualization properties: the number of common visible elements from any two examined pairs of landmark elements and the length of the visual chain of those pairs

Methods of analysis
In order to investigate the implications of visualization on wayfinding performance, a movement tracking method was used during the experiment by recording the coordinates (X and Y) of each individual wayfinding task path, using the Skyline® Terra explorer 4.6 interface. Then, in order to obtain the trajectory patterns of each task, the real-time-log data was converted to GIS layers and visualized as polylines using ArcGIS 8.2
environment. Statistical analysis of these trajectory patterns was performed using SPSS 12.0 software. In addition, a detailed questionnaire was used for analyzing subjective differences between each wayfinding task. At the end of each task, participants were asked to estimate the difficulty of the tasks’ performance (by giving a score between 1 and 5). Then, they were asked to explain how they reached the target and what helped or prevented them understand the relative position between the target and the origin points.

4. RESULTS

According to the hypothesis of the research, it was expected that the more there are common visible elements between an origin and a target element, or alternatively, the shorter the visual chain between them (i.e. the less number of steps in a visual chain) - the easier it will be to reach the target landmark element and thus, it will be easier for the participants to complete the task. We estimated ease of task completion using two criterions: the first was to compare the shortest route between each pair of elements to those routes actually taken by participants (see appendix A) and the second, to analyze the participant’s subjective estimation: ease of task and wayfinding strategies.

The length of the visual chain

Comparison between task 4 (with 3-steps visual chain) and task 5 (with 2-steps visual chain) shows that the length of the visual chain indeed influences the tasks’ completion rate. Namely, as we expected, the shorter the visual chain between the origin and the target elements - the easier it is to complete the task and reach the target element.

When comparing both, the trajectory patterns (fig. 4a) and the mean length of the routes taken by the participants (fig. 4b) with those of the shortest available routes, it was found that in task 4 (HC), which requires 3 steps to be taken between the origin and the target elements, less participants successfully completed the task (following the shortest route) in comparison to task 5 (CC), which requires only 2 steps to be taken in order to successfully complete the task; While only 57% of the participants successfully completed task HC with a ratio of 1.62 between the mean length of the taken routes and the length of the shortest available ones, 75% of them successfully completed task CC, with a ratio of 1.29 between the mean length of the taken routes and the shortest ones.

Fig. 4a
The tasks completion success rate according to the trajectory patterns analysis (4a) and the ration between the mean routes length and the shortest routes length (4b) of task HC (with 3-steps visual chain between the historic building and the commercial center) and task CC (with 2-steps visual chain between the clinic and the commercial center)

Overlapping between visible fields
The second visual property, overlapping between visible fields of landmarks (the number of common visible elements between the origin and the target elements), was examined in wayfinding tasks 1 (ST), 2 (RS) and 3 (HS). Fig. 5 illustrates the task completion success rate (using the shortest route) in each of these tasks which differ in the number of common visible elements between them.

As we expected, task HS (historic building → square), in which its origin and target elements share three common visible elements, resulted in the best wayfinding performance; That is, 93% of all participants completed the task following the shortest available route between the elements. Since this path is also characterized by only 1 visual chain step, it may illustrate the integrative affect of the relatively high degree of overlapping (number of common visible elements) and the short visual chain length between the origin and the target elements; those two properties may assist the construction of an ordinal procedural knowledge. However, a comparison between tasks 1 (ST) and 2 (RS) do not match our initial hypothesis. Namely, in task 1 (with only 1 common element) the completion success rate was higher than in task 2 (with 2 common elements); 71% of the participants successfully completed task 1, while only 57% of them complete task 2. In order to understand these results, we later use the questioner in which the participants` explained how they completed the task and what guided them during the tasks’ completion.

Same results as illustrated above are also visible when comparing between the mean length of the routes taken by the participants and the length of the shortest available ones (see Fig. 6). While this ratio was 1.59 in task 2 (RS: restaurant → school), it was only 1.14 in task 1 (ST: school → tree) and 1.07 in task 3 (HS: historic building → school).

1 In cases where participants chose to follow a neighbor route to the shortest available one, it was also considered as a ‘success’, as long as no other element was visible along the path besides the one expected to act as a link to the target element.
Subjective evaluation

From the self-report verbal analysis we may assess the effect of the two investigated visualization properties on the participants wayfinding strategies and on their subjective experience when completing the tasks. It may provide us complementary information to the quantitative results presented above. In addition, the argument that wayfinding is not merely affected by visual properties but is also based on an ordered recognition of familiar vistas or views along a route can also be learned through the participant’s direct experience as illustrated in a verbal form. Participants explained us that when they were asked to get from one point to another, they used landmark elements as “visible steps” which helped them progress along the route till arriving to the target point. Furthermore, as we expected, the more visible elements there were from a given element, the more this element was dominant and functioned as an anchor during the wayfinding task. The square was the element with the largest number of visible elements from it (four visible elements); Participants declared that in each task they first wanted to find the square, and from there reach the target element. Table 2 demonstrates the use of views as wayfinding strategy when completing tasks HC and CC using the participants’ self-report.

| Task 4: HC (historic building  ➔  commercial center) | Task 5: CC (clinic  ➔  commercial center) |
“I remembered I have to get to the tree, and I knew that from there the commercial center will be visible.”
“from the historic building you can’t see the square…I remembered school, square, tree…I just needed a few station along the way till arriving…connect between buildings which I can see from the square..”

“initially I forgot I can`t see the commercial center from the square…but then I remembered that I can see the center from the tree…so I went to the tree..”

“from the beginning I already knew that O.K., I have to get to the tree!”

“I knew that when standing near the tree I can see the center…so I knew I first need to reach the square and from there observe the tree…”

“I remembered that from the clinic you can see the square, and from there it’s obvious that the tree is visible..”

“initially I thought this task will be the hardest as during the learning phase I first saw the center and only at the end I saw the clinic…but I decided I can use the tree through the square.. when I saw the tree I felt more confident!”

Table 2: the use of views as wayfinding strategy when completing tasks HC and CC using the participants’ self-report verbal analysis.

Generally speaking, from analyzing this documentation we may learn that indeed the participants were able to build an ordinal procedural spatial knowledge based on sequence of views, a kind of cognitive representation that enables a successful completion of wayfinding tasks using shortcuts and not only use the initially learned routes.

As illustrated in table 4, and similarly to the objective examination demonstrated above, participants estimated task 4 (HC- with 3 visual steps chain) as the hardest task of all the five given to them, with an average score of 2.8 (when 1 is the easiest the 5 is the hardest). For all other tasks, the average score ranged around 2.4-2.5. Note, that when only one visual chain step was needed for completing the task and when the square acted as a ‘visual connecting’ element (in task ST: School → Tree), the participants felt that the task was the easiest one for them (with an average score of 2.2).

<table>
<thead>
<tr>
<th>Task</th>
<th>Average Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1: (ST) School → Tree</td>
<td>2.2</td>
</tr>
<tr>
<td>Task 2: (RS) Restaurant → School</td>
<td>2.4</td>
</tr>
<tr>
<td>Task 3: (HS) Historic building → Square</td>
<td>2.5</td>
</tr>
<tr>
<td>Task 4: (HC) Historic building → Commercial center</td>
<td>2.8</td>
</tr>
<tr>
<td>Task 5: (CC) Clinic → Commercial center</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Table 4: Participants estimation regarding the tasks’ completion ease, where’s 1 is the easiest the 5 is the hardest.
The subjective evaluation can also be used to determine the level of spatial knowledge achieved during the experiment. At the end of the experiment, participants were requested to estimate the level in which they felt the city is “easy or hard for learning and understanding”. We separated between two categories: ease of understanding relative directions between the seven landmark elements and ease of being able to reach one element from another. As expected and as can be seen in Fig. 7, while the average score for being able to understand the relative directions between the elements was 2.6 (when 1 is a poor understanding and 5 is the best understanding), the average score for being able to reach one element from another was 3.9.

*Fig. 7:* Participants estimation for the city as easy or hard for learning and understanding relative directions between the seven landmark elements and reaching one element from another. Where’s 1 is a poor understanding and 5 is the best understanding.

Thus, as expected, the ordinal procedural spatial knowledge is only partly transferred into configurational knowledge. However, knowledge of geographic space does not have to be perfect in order to be configurational; various states of partial configurational knowledge may exists. The 'lowest' form of configurational knowledge might just show connections between objects ('topology'), and indeed Kuipers (1978) and some other authors have identified 'topological knowledge' as a kind of stage intermediate between 'procedural' and 'configurational'.

In order to investigate this issue, we asked eight of the participants to mark after the learning phase on an A4 sheet of paper the exact location of the city’s elements. For providing them reference points, we marked on the paper the layout of the boundary and the exact location of the tree, the square and the starting point (see Fig. 8). Then, using GIS software, we calculated for each participant the mean distance error of the elements’ location on the map in relation to their real location. Examination of the relation between the performance of the participants, as an indication of procedural knowledge, was found to be correlated ($R^2=0.69$) with the accuracy of the landmark locations in the maps drawn by the participants at the beginning of the wayfinding task phase.
5. CONCLUSIONS

According to the hypothesis of the research, a high degree of overlapping between visual fields (more common visible elements) of an origin and a target element helps people to construct (ordinal) procedural spatial knowledge. Moreover, when this overlapping is a direct one, namely, they are characterized by only 1 visual chain step, the participants achieved the best performance in the wayfinding tasks. It was also found, that visual overlapping and a short topological depth between them contribute to the construction of topological configurational knowledge. The results of the research encourage the use of complementarily methods to the graph theory based methods such as space syntax, for analyzing the visual structure of urban environment which focus on the overlapping between visual fields and the topological depth between them. A multidimensional topological analysis for a structural analysis of geographic systems is an example for such kind of method (Jiang and Omer, 2006).

REFERENCES


**APPENDIX A: A COMPARISON BETWEEN THE SHORTEST ROUTE BETWEEN EACH PAIR OF ELEMENTS TO THOSE ROUTES ACTUALLY TAKEN BY PARTICIPANTS**

<table>
<thead>
<tr>
<th>Restaurant → school</th>
<th>Taken route</th>
<th>Shortest route</th>
</tr>
</thead>
<tbody>
<tr>
<td>School → tree</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historic building → square</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historic building → commercial center</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinic → commercial center</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Effective Wayfinding Based on LBS Using Landmarks in Urban Environments

Samani, N., M. R. Delavar and M. R. Malek

Center of Excellence in Geomatics Engineering and Disaster Management, Dept. of Surveying and Geomatics Engineering, Engineering Faculty, University of Tehran, Tehran, Iran
najmehsamany@yahoo.com, mdelavar@ut.ac.ir, malek@ncc.neda.net.ir

ABSTRACT
Wayfinding in urban environments either pedestrians or drivers is one of the most important tasks of mobile GIS systems. Navigation services communicate optimal routes to users by providing sequences of instructions for these routes. Each single instruction guides the navigator from one decision point to the next. The instructions are based on geometric data from the street network, which is typically the only dataset available. Currently user applies navigation systems which provide instructions in the form of maps, pictograms and spoken language. However, they are so far not able to support landmark-based navigation, as the most natural navigation concept for humans which also play an important role for upcoming personal navigation systems. The goal of this paper is to solve landmark-based wayfinding problem in an urban area using LBS. In this context, the developments of systems capable of wayfinding based on landmarks to facilitate navigation are evaluated. This research utilizes the advantages of landmarks in wayfinding to provide a new means for guide different users in a novel urban environment via LBS. These types of wayfinding services can be used in any context based on the selected landmarks. The goal is to improve existing navigation services by concepts closer to the users understanding, adaptive for individual users, with flexibility for different tasks.

Keywords: LBS – Wayfinding- landmark- Mobile GIS

1. INTRODUCTION
Location Based Services (LBS) is a relatively new, growing technology field that focuses on providing information via mobile and field units based on individual spatial positions. Mobile geospatial information system (GIS) solutions and LBS, which extend beyond today’s use, will play a major role in the future. Especially tourists, fire fighters and field workers will benefit from getting localized information on mobile devices anywhere anytime. Such systems rely on a range of mobile hardware units that require different interaction styles compared to their desktop-based counterparts [5].

Navigating in urban environments either pedestrian or drivers is one of the important tasks of Mobile GIS systems [12]. Navigation services communicate optimal routes to users by providing sequences of instructions for these
routes. Each single instruction guides the wayfinder from one decision point to the next. The instructions are based on geometric data from the street network, which is typically the only dataset available. Enriching such wayfinding instructions with landmarks has been proved [13].

Mobile navigation systems are important assets for travelers visiting foreign environments, as they provide instructions on how to find the way to get to a chosen destination. Research has shown that finding ones way in a foreign environment is primarily based on cues in the environment [4]. It is envisaged that various landmarks and key features of neighborhood environments might be provided as spatial cues via LBS. Currently user applies navigation systems which provide instructions in the form of maps, pictograms and spoken language. However, they are so far not able to support landmark-based navigation, as the most natural navigation concept for humans also plays an important role for upcoming personal navigation systems. Particularly, nowadays, cities are becoming ever-increasingly complex requiring citizens to become ever more sophisticated in navigating urban spaces by foot, public transport or in private vehicles, so conventional navigations can not response successfully and adapting LBS with landmark-based wayfinding is more effective. In order to provide such navigation, the first step is to identify appropriate landmarks and display them on LBS to navigate user.

Using landmark-based wayfinding in GIS has been recently investigated. Human wayfinding researchers work on the processes that happen when people orient themselves and navigate through space. Theories try to explain how people find their ways in the physical world, what people need to find their ways, how they communicate directions and how people's verbal and visual abilities influence wayfinding.

This paper addresses the question of solving landmark-based wayfinding problem in an urban area using LBS. In this context, the developments of systems capable of wayfinding based on landmarks to facilitate navigator is evaluated. This research utilizes the advantages of landmarks in wayfinding to provide a new means to guide different users in a new urban environment via LBS. These types of wayfinding services can be used in any context based on selected landmarks. The goal is to improve existing navigation services by concepts closer to the human user, adaptive for individual users, with flexibility for different tasks.

Section 2 gives an overview of human wayfinding and highlights the importance of landmarks for navigation. It further describes the components of wayfinding instructions and how they are used in a navigation service. In section 3 we present the related works that carried out in LBS and applications in urban activities. Section 4 explains the usage of LBS on wayfinding task and the essence of landmark display on it. In the rest of this section, the challenges of cartographic restrictions are evaluated. A Simulation of landmark-based wayfinding using LBS in section 5 is used to demonstrate the proposed method. The final section gives conclusions and directions for future work.

2. RELATED WORKS

Cultural/tourist mobile guides are becoming common aids to combine information transfer with a guidance service. Mobile guides for pedestrians provide their users with location specific information, e.g. based on GPS coordinate [8].

Previous work on sensor-based information systems was predominantly conducted for mobile guiding systems [5]. The CYBERGUIDE system was one of the first that used location aware information to help tourists. The indoor component relied on infrared beacons broadcasting a unique ID that was used to display an arrow on a map.
whenever the user entered a new room. Additionally, the user's orientation was estimated from her/his actual walking, direction and the topology of the building. The outdoor system instead used GPS to determine the user's position and to display it on a map [2]. Both systems operated independently from each other and could not be combined. GUIDE is a location-aware multimedia tourist guide developed for the City of Lancaster. The system provides location based information based on a radio cell infrastructure [5].

The MOBIS system is an electronic guide based on a PDA that provides information on the exhibits to a visitor of a museum. The PDA receives its position from infrared beacons distributed in the environment and uses this position as a pointer to a specific content that is stored in a database on the PDA [2]. The HIPS system uses sub-notebooks, which supports a broader range of media content than the PDA used for MOBIS. HIPS takes into account the absolute position, as well as the distance to objects in the exhibition and uses a radio back-channel for downloading information [5]. The MARS system is an augmented reality system that provides information for the buildings on the Columbia University Campus. More recently an additional indoor component has been developed that assists the user also in indoor navigational tasks [2]. The DEEP MAP project (carried out at the European Media Lab) focuses on a mobile tourist information guide that brings together results from natural language and intelligent graphics generation. This allows a multimodal user interface to offer the user a variety of information on the city of Heidelberg. Since GPS is used to determine the position DEEP MAP is at the moment restricted to outdoor scenarios [20]. The NEXUS system aims to provide a general framework for mobile and location aware computing. Another drawback of the aforementioned systems is their inability to adapt to the restricted cognitive resources of the user. Especially the use of mobile devices implies the need for adaptive presentation generation and interaction, given the fact that the user might be distracted or under time pressure [2]. However, very few systematic studies have been carried out on this aspect, particularly individual preferences with regard to cognitive abilities, spatial awareness, prior knowledge of the location, map reading ability and so on.

3. LANDMARK-BASED WAYFINDING

Wayfinding i.e. getting from some origins to a destination is one of the prime everyday problems human encounter [10]. It has been defined as purposeful, directed, and motivated movement from an origin to a specific distant destination, which can not be directly perceived by the traveler [1]. Such behavior involves interactions between the traveler and the environment. Human wayfinding takes place in large-scale spaces. Such spaces cannot be perceived from a single viewpoint; therefore, people have to navigate through large-scale spaces to experience them. People use various spatial, cognitive, and behavioral abilities to find their ways. These abilities are a necessary prerequisite to use environmental information or representations of spatial knowledge about the environment [13]. Landmarks act as primary features in the wayfinding task, and work as nodes for organizing other spatial information into a layout. They may be noticed or remembered because of dominance of visible form, peculiarity of shape or structure, or because of socio-cultural significance. Paths are static linear structures in an environment; they may be streets, footpaths, pavements, canals, rivers or railway tracks. Landmarks play an important role when humans navigate through foreign environments. For example, trying to find the way is much easier if the navigator can rely on a description of the route based on well-recognizable objects in the environment, instead of navigating solely on the basis of street names and metric directions. Landmark-based navigation applies knowledge about prominent objects in the environment to guide travelers through unknown areas [4].

Among the different meanings of landmark, it is an object or structure that marks a locality and is used as a point of reference. The concept is bound to the prominence or distinctiveness of a feature in a large-scale environment or landscape. Thus the landmark saliency of a feature does not depend on its individual attributes but on the
distinction to attributes of close features, being a landmark is a relative property. Studies show that landmarks are selected for route directions preferably at decision points. Another study has shown that mapped routes enriched with landmarks at decision points lead to better guidance, or less wayfinding errors, than routes without landmarks. Furthermore, different methods of landmark presentations were equally effective [13].

Lynch [9] defines landmarks as external points of reference—points that are not part of a route like the nodes in a travel network. He characterizes the quality of a landmark by its singularity, where singularity is bound to a clear form, contrast to the background, and a prominent location. The principal factor is the figure-background contrast. The contrast can be produced by any property, such as uniqueness in form or function in the local or global neighborhood [13]. Landmarks are essential parts of wayfinding directions and any communication about space.

Having identified a local landmark at the decision point, an instruction can be created following any grammar, for instance the grammar defined below [13].

```
(']’ denoting required elements, ‘{}’ optional elements, ‘UPPER CASE’ language elements, and ‘lowercase’ variables):
[AT landmarki] + [TURN LEFT | RIGHT | MOVE STRAIGHT] + {ONTO street name} +
{(PASSING | CROSSING) landmarkj}0…n +
[U N T I L  landmarkk].
With     i≠j≠k.
```

In this context the following advantages of wayfinding instructions are presented:

- Create a tourist/visitor friendly environment
- Facilitate ease of movement
- Provide distinct and recognizable signs
- Send clear and direct messages
- Display graphically consistent design
- Compliant with applicable regulations

Today’s car navigation systems provide driving instructions in the form of maps, pictograms, and spoken language. However, they are so far not able to support landmark-based navigation, as the most natural navigation concept is for humans and also plays an important role for upcoming personal navigation systems [3].

However, the original concept of delivering the instructions has not changed very much. Still, spoken language instructions use a relatively small set of commands (like “turn right now”), which only refer to properties of the street network. This is not optimal, since i) features of the street network typically are not visible from a greater distance due to the low driver position and small observing angle, and ii) the most natural form of navigation for humans is the navigation by landmarks, i.e. the provision of a number of recognizable and memorizable views along the route. Obviously, the introduction of buildings as landmarks together with corresponding spoken instructions (such as “turn right after the tower”) would be a step towards a more natural navigation. As we argue below, this would be well integrable into today’s navigation systems (Figure 1) as it would not imply a major modification of systems and data structures [3].
Therefore, showing landmarks on routes and giving wayfinding instructions based on them are possible and leads to better navigation.

4. USING LBS ON WAYFINDING

This section introduces the method to facilitate navigation systems by utilizing wayfinder map on LBS. According to the abilities of LBS and their restrictions, the use of landmark-based wayfinding would be a more effective solution.

4.1. The concept of LBS

LBS are a new industry at the core of which is GIS and spatial databases. With increasing mobility of individuals, the anticipated availability of broadband communications for mobile devices and growing volumes of location specific information available in databases, there will inevitably be an increase in demand for services providing location related information to people on the move [7]. LBSs consist of a broad range of services that incorporate location information with contextual data to provide a value-added experience to users on the web or wireless devices. In contrast to the passive fixed Internet, users in the mobile environment are demanding personalized, localized, and timely access to content and real-time services. Targeted data, combined with location determination technology, is essential to create personalized value to an end-user's mobile experience. This allows wireless carriers and portals to significantly increase the value of services to subscribers while opening up new revenue opportunities [6].

There are important similarities and differences between LBS technology and geospatial information systems (GISs). For beginners, much of the underlying mapping, spatial indexing, spatial operators, geocoding and routing technology that are used to deliver LBSs originate from the GIS industry. However, what makes LBS technology different is that it is a service deployed on a foundation of information technology (IT) and wireless technology. Another major difference is that LBS impose significant technology and services capabilities that exceed the general requirements of static GIS uses, namely [6]:

- **High Performance**: Delivers sub-second queries required for Internet and wireless.
- **Scalable**: Supports thousands of concurrent users and terabytes of data.
- **Reliable**: Capable of delivering up to 99.9999 up-times.
- **Current**: Supports the delivery of real-time, dynamic information.
- **Mobile**: Available from any device (wireless and wired) and from any location.
• **Open**: Supports common standards and protocols, HTTP, WAP, Wireless Markup Language (WML), Extensible Markup Language (XML) and Multimedia Markup Language (MML).

• **Secure**: Leverages the underlying database locking and security services.

• **Interoperable**: Integrated with e-BUSINESS APPLICATIONS (Customer Relationship Management, Billing, and Personalization) and wireless positioning gateways [6].

LBS can be defined as to the provision of geospatially-orientated data and information services to users across mobile telecommunication networks. LBS can be seen as the convergence of new information and communication technologies (NICT) such as mobile telecommunication system, location aware technologies and handheld devices with the Internet, GIS and spatial databases (Figure 2) [7].

![Figure 2: Convergence of technologies to create LBS [7]](image)

### 4.2. LBS and navigation

It is commonly understood that LBS applications must be able to integrate mapping, routing, searching and address location functionality with user-specified content. Also location information should be provided in a range of formats such as graphical, textual and voice. Concerns are focused on the capability and the diversity of devices to deliver and access location information. There are also discussions based upon the effectiveness and user friendliness of the way to deliver location-based information, such as “speak and listen”, maps and text. Given that the demand for real-time, fast-changing information is increasing, people will require more relevant, timely information in various forms. In the navigation LBS applications, instructions are given to people by means of maps, spoken word and text. It is envisaged that various landmarks, point of interest (POI) and key features of neighborhood environments might be provided as spatial cues via LBS [7].

Over the last few decades, most urban landscapes have been utterly transformed. Cities are becoming ever-increasingly complex requiring citizens to become ever more sophisticated in navigating urban spaces by foot, public transport or in private vehicles. The scale and pace of change in urban areas through expansion, infill, re-development and changing use continues to increase. Together this creates further uncertainties in the way that we structure and navigate through urban space. Thus urban wayfinding is an increasingly complicated task. From a GIS perspective, spatial databases and fast response to spatial queries lie at the heart of LBS [7].

LBS are services for mobile users that take the current position of the user into account when performing their task. Here is a subset of possible applications of LBS also relevant to urban users [16]:

1. "Speak and listen": Maps and text.
2. "Speak": Various landmarks, POI and key features.
3. "Listen": Maps, spoken word and text.

Figure 2: Convergence of technologies to create LBS [7]
• Traffic Information, e.g., there is a traffic queue ahead, turn right on the A3.
• Emergency Services, e.g., “help, I'm having a heart attack!”
• Roadside Emergency, e.g., “help, my car has broken down!”
• Law Enforcement, e.g., “what is the speed limit on this road where I am at?”
• Classified Advertising, e.g., “where are nearby yard-sales featuring antiques?”
• Object visualization, e.g., “where is the historic parcel boundary?”
• Underground Object Visualization, e.g., “where is the water main?”
• Public Safety Vehicle Management, e.g., “who is closest to that emergency?”
• Location-Based Billing, e.g., free calls on your phone, in a particular location
• Leisure Information, e.g., “How do we get to the Jazz Club tonight from here?”
• Road Service Information, e.g., “Where is the nearest petrol station?”
• Directions, e.g., “I’m lost, where is nearest Metro station?”
• Vehicle Navigation, e.g., “how do I get back to the Interstate from here?”
• Vehicle Theft Detection, e.g., “my car has been stolen, where is it?”
• Child Tracking, e.g., “tell me if my child strays beyond the neighborhood.”
• These examples take location as the only instance of the broader issue of “context awareness”

4.3. Giving instructions using LBS rely on landmark-based wayfinding

There are two different kinds of route directions to convey the navigational information to the user: either in terms of a description (verbal instructions) or by means of a depiction (route map). The structure and semantic content of both is equal, they consist of landmarks, orientation and actions [3]. Using landmarks is important, because they serve multiple purposes in wayfinding: they help to organize space, because they are reference points in the environment and they support the navigation by identifying choice points, where a navigational decision has to be made [11].

Mobile navigation systems are considered as one of the possible breakthrough technologies for broad band wireless internet access. Although quite sophisticated systems already exist for different transportation means, e.g. car navigation systems and more recently also pedestrian navigation systems, these systems only function in a
single well defined context. They are designed either to be operated by a driver in a car, or a by pedestrian in an outdoor scenario [2].

Various studies have been discussed, that navigation instructions should not only consist of street names and directions but also have to be improved by additional indications of landmarks. Nearly all answers included additional descriptions – landmarks – beside the routing instructions for better. The user feels more comfortable finding his way when supported by additional information like landmarks. Landmarks can help pedestrians in navigational problems and serve as decision support for turnarounds or as confirmation for a decision.

4.4. Challenges of cartographic limitation of Using LBS to support landmark-based wayfinding

Telecommunication infrastructure (mobile network), positioning methods, mobile input and output devices and multimedia cartographic information systems are prerequisites for developing applications, which incorporate the user’s position as a variable of an information system. Imparting spatial information within such a system, normally cartographic presentation forms are involved. Thus, the resulting system can be called map-based LBS. The possibilities of transmitting spatial information in context of a determined position by various presentation forms are primarily restricted by the limitations of the used mobile device [10]. Mobile computing environment has certain features that impose restrictions. The properties of mobile networks are: (relatively) low bandwidth, strong bandwidth variability, long latency, unpredictable disconnections and communication autonomy. The properties of mobile terminals are: small and low-resolution displays, limited input capabilities, limited computing power, limited power and small memory size. The practical conditions, when and where the mobile devices are used, bring also additional restrictions. The mobile users are typically in very unstable environment in varying conditions, where their cognitive capacity is demanded for other tasks as well. All these restrictions have to be taken very carefully into account when designing LBS. Some of the implied requirements are as follows [10]:

- Not very intensive use of mobile network and minimal high of transmitted data.
- Possibility to offline operation.
- User interface should be very simple and user friendly and the amount of presented Information content limited and well specified.

And LBS would provide the following facilities for users [10]:

- General considerations: Make sure the vehicle is on the right track using minimum cartographic data used
- Automatic scrolling: Automatic adaptation of the presented map section to the position of the user defined and has been found indispensable.
- Egocentric map view: Most of users preferred a “track up” oriented map that means the map is always adapted to the user’s direction of move, only some of them preferred “north up” orientation.
- Supported change of scale: Abilities of device to zoom and pan map can be useful. The change can be done abruptly, step by step or animated [10].

It has shown that the chosen presentation form and the effort of supporting cartographic methods is decisive for the acceptance of pedestrian navigation systems.
5. SIMULATION OF LANDMARK-BASED WAYFINDING USING LBS

This section demonstrates the applicability and usefulness of the presented approach by showing a simulation study.

5.1. Required data:
For the automatic selection of context-dependent landmarks for navigation, a number of data sources need to be used as follows:

• Digital city maps, such as the multipurpose maps of area. City maps provide boundaries and classifications of the built areas.
  • Navigation graphs for the actual means of travel. Navigation graphs are needed for route selection algorithms as well as for the route-specific classification into possible and real decision points.
  • Rectified, geo-referenced images of façades of each single building located at elements of the navigation graph.
  • Accessible databases such as yellow pages and GIS databases.

5.2. Simulation:
This part introduces the method implemented to assess the landmark-based wayfinding instructions via LBS. In this case we expect to see on the screen of navigation systems; routes, landmarks at decision points and instructions which orient navigator in urban environments.

First, we identified user-context; individuals encountered to new urban networks and drive on car, and measure the saliency of potential landmarks to extract them automatically. Second, using these extracted landmarks; wayfinder map, which consist of network routes and landmarks at decision points have been created. (Figure 4). So instruction and algorithms for wayfinding based on this structure is defined.

Figure 4: Landmark-based wayfinding instructions on LBS
The achieved results have been shown that implementation of the landmark-based wayfinding instruction on LBS follows several advantages:

- The volume of data would be reduced, so data transfer would be expedited.
- This system might be act more user-friendly.
- Wayfinding based on landmarks are more natural to human wayfinder and navigation using this system is more convenient.
- Using wayfinder map (routes, generalized areas and selected landmarks situated on decision point) is more understandable.

6. CONCLUSION AND FUTURE TRENDS

This paper has focused on landmark-based wayfinding using LBS. Navigating in urban environments is becoming an important issue. As described above, current navigation systems guide wayfinder based on coordinates and geometric dataset of network. However, it has been shown that wayfinding by means of available landmarks; particularly on decision points are more effective.

In this research a simulation of landmark-based wayfinding on LBS is implemented and the results notify that the method have more advantages over conventional approach, some of them are reduction of data volume and so data transfer time, increase the user abilities to use this navigation system since it is based on natural human wayfinding.

Future directions of this research are as follows:

- Selection and use of context-based landmark on LBS to be adapted for different modes of travel (e.g., pedestrian, bicycle, car) and user groups (e.g., tourist, business traveler).
- The method needs to be implemented and applied to larger datasets in order to analyze performance and computational cost.
- Different times (e.g., day or night-time) may require different landmarks therefore the integration of temporal constraints into route instructions is necessary.

REFERENCES

7. Li, C., 2002: Spatial ability, urban wayfinding and location-based services: a review and first results. Centre for Advanced Spatial Analysis, University College London, 1-19 Torrington Place, London WC1E 6BT, United Kingdom.
The use of Rough Set in Geographical Information Systems

Beniamino Murgante, Giuseppe Las Casas, and Anna Sansone

Università della Basilicata – via dell’Ateneo Lucano 10 – 85100 Potenza (Italy)
Tel.: +39(0)971-205125, Fax: +39(0)971-205185, E-mail: murgante@unibas.it

Abstract

The distinction among urban, peri-urban and rural areas inside a territory represents a classical example of uncertainty in land classification. The transition among the three classes is not completely clear and can be described by the Sorites Paradox, taking into account residential buildings and settlements. Peri-urban fringe can be considered as a transition zone between urban and rural areas, as an area with its own intrinsic organic rules, as a built area without formal organisation or as an abandoned rural area contiguous to urban centres. In any case, concepts as density of buildings, services and infrastructures or the degree of rural, residential and industrial activities, will lead to uncertainty in defining classes, due to the uncertainty in combining some properties. One of the methods which can be employed is the rough set theory, which represents a different mathematical approach to uncertainty capturing the indiscernibility. The definition of a set is connected to information knowledge and perception about phenomena. Several phenomena can be classified only in the context of information available about them. Two different phenomena can be indiscernible in some contexts and classified in the same way (Pawlak 1983). Rough set approach to data analysis hinges on the basis of two basic concepts: the lower approximation, which considers all the elements that doubtlessly belong to the class and the upper approximation, which includes all the elements that possibly belong to the class. Furthermore, the rough set theory takes into account only properties which are independent. This approach has been tested in the case of study of the Province of Potenza, in Southern Italy. This area is particularly suitable to the application of this theory, because it includes 100 municipalities with different number of inhabitants, quantity of services and distance from the main road infrastructures.

1. INTRODUCTION

In recent times, the expression “periurban area” has been frequently used in planning documents. This expression, despite its large use, has no clear and unambiguous definition, which can allow the exact edge of these zones. The major reasons of this deficiency can be identified in the complexity of the phenomenon to be analyzed and in the huge variety of territorial contexts in which it can occur. The expression “periurban fringe” evokes different meanings, following the various disciplinary approaches to the problem. For economists, periurban areas represent a disorganized city expansion, that involves an enormous increase of costs in services management and infrastructures maintenance. For ecologists, periurban fringes are zones in which the natural resources have been irreparably damaged because of the large concentration of economic activities. Agronomists take into account the loss of productivity, due to the increasing presence of buildings.

Urban planners have two different approaches: the first one considers the phenomenon from a theoretical point of view, in comparison with the consolidated concepts of city and rural area, the other one takes into account the increase of the economic value of real estate due to the transformation. The phenomenon is so complex to analyze, that even local laws denote a certain superficiality, lacking of precise indications. Although local laws correctly face the problem, giving different definitions and respecting different territorial peculiarities, on the other hand they consider only the proximity to urban areas. It is obvious that contiguity condition alone is not enough to define such a complex phenomenon. A clear definition of periurban fringe can be achieved considering this zone as an area with its own intrinsic organic rules, like the urban and rural ones and not as a transition zone from urban to rural areas. An accurate rule, defining the periurban fringe, can be composed by the combination of other rules. For instance, proximity to urban areas, contiguity to road network, presence of
utilities and urban services, density of population higher than in rural areas can generate a set of inclusive rules and if some of these rules are satisfied, the area can be included in the periurban fringe. In the same way, exclusive rules considering archaeological sites, heritage areas, environmental preservation areas, deep slope terrains, landslide areas, erosion areas can be achieved. If even one of this rules is satisfied, the area cannot be included in the periurban fringe.

The aim of this paper is to define more precise rules in order to describe the periurban phenomenon, using techniques of spatial statistic, such as point pattern analysis. This approach has been tested in the case of study of Potenza Province, in Southern Italy. This area is particularly suitable to the application of this method, because it includes 100 municipalities with different number of inhabitants, quantity of services and distance from the main road infrastructures.

2. TECHNIQUES OF SPATIAL ANALYSIS AND ROUGH SET THEORY

Usually, a spatial phenomenon can be analyzed in terms of first and second order properties. First order properties describe geographical events in terms of density or intensity, while second order properties analyze the relationships among spatial phenomena in terms of distance. In this study two techniques of spatial analysis have been used: kernel density and nearest neighbour distance. These two approaches are call effects of first and second order and they consider respectively the amount of events observed per unit area and the distance among them. All geographic themes achieved by means of these two techniques are combined with the rough set theory in order to define in “indiscernible” way the periurban fringe.

2.1 Kernel density

The kernel density is a technique of point pattern analysis that generates a grid from a point vector source. The numerical attributes associated to vector data can be considered as a weight increasing the intensity of the phenomenon. Bailey and Gatrell (1995) have developed a set of spatial analysis techniques applied in the field of spread of epidemics. These techniques are based on Waldo Tobler (1970) first law of geography: “Everything is related to everything else, but near things are more related than distant things”. Compared to classical statistical approaches, it is important to locate data, considering the events as spatial occurrences of the considered phenomenon. Each event is located in the space in an unambiguous way by its coordinates (x, y).

An event is a function of its position and attributes which characterize it and quantify the intensity:

\[ L_i = (x_i, y_i, A_1, A_2, \ldots, A_n) \]

While the simple density function considers the number of events for each element of the regular grid that composes the study region R, the kernel density takes into account a mobile three-dimensional surface, which weighs the events according to their distance from the point from which the intensity is estimated (Gatrell et al., 1996).

The density (intensity) of the distribution in point L can be defined by the following equation:

\[ \lambda(L) = \sum_{i=1}^{n} \frac{1}{\tau^2} k \left( \frac{L - L_i}{\tau} \right) \]

where, \( \lambda(L) \) is the intensity of point distribution, quantified at point L; \( L_i \) is the event number i, \( k() \) represents kernel function and \( \tau \) bandwidth. \( \tau \) can be defined as the radius of a circle generated from the intersection between the surface and the plan containing the study region R. Two major factors influence the results: dimensions of the grid and bandwidth (Batty et al. 2003). The bandwidth produces a three-dimensional surface, more or less corresponding to the phenomenon, allowing to analyze its distribution at different scales. The choice of bandwidth remarkably influences the surface of estimated density. If bandwidth is high, kernel density is
closer to the values of simple density. With a narrow bandwidth, the surface will capture local events, with density close to zero for the elements of the grid located far from each event. The right bandwidth can be determined by estimating the phenomenon and, if it is important, by highlighting peaks of distribution or smooth spatial variation.

2.2 Nearest neighbour distance

The nearest neighbour distance belongs to the point pattern analysis family and it can be considered as an alternative to the methods based on density. It is the most common technique in order to describe the second order properties of events distribution, measuring the distance.

The nearest neighbour distance for an event in a point distribution is the distance from a point to the nearest source. The source can be in vector or grid format and it represents the event to analyze. The distance between the events is normally defined by the following function:

\[ d(L_i, L_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \]  

Equation 3

If \( d_{\text{min}}(L_i) \) is nearest neighbour distance for an \( L_i \) event, it is possible to consider the mean nearest neighbour distance defined by Clark and Evans (1954) as:

\[ \overline{d}_{\text{min}} = \frac{\sum_{i=1}^{n} d_{\text{min}}(L_i)}{n} \]  

Equation 4

The result of the nearest neighbour distance is in raster format and it considers the distance among the cells barycentre. Also in this case, it is important to opportune estimate some factors, like the maximum distance within which one takes measures and the dimensions of the cells that compose the grid.

2.3 Rough set theory

Spatial informations can be classified following two criteria: the nature of data and the information reliabilities. The peri-urban phenomenon is a case with an high level of uncertainty, with ill defined data for the whole area and not much clear rules. An ontology of uncertainty in spatial information is reported in figure 1.

![Figure 1 An ontology of uncertainty in spatial information (adapted from Las Casas and Murgante, 2004).](image-url)
The attention will be focused on poorly defined data. Some data are said to be vague if their meaning is not clear, whereas ambiguous data can have at least two specific meanings. For instance, ambiguity of meanings leads to a discordance in data classification due to a different perception of the phenomenon. Vagueness, which is contrary to the Boolean concept, widely diffuse in the representation of geographical information, can be addressed by multi-valued logic and applications of fuzzy set theory. Inaccuracy produces uncertainty in the case of low quality of data, due to a high percentage of errors.

Rough set theory (Pawlak, 1982) is certainly less famous than the others outlined in figure 1, and it classifies elements in indiscernible way, using the available information. In other words the elements of the universe can be seen in a certain way in the context of an available information about them. The direct consequence is that two different elements can be indiscernible in some circumstances and in other contexts they can belong to different classes.

This methodology is based on the concepts of indiscernibility relation, upper and lower approximation and accuracy of the approximation.

In an Information System IS=(U,A), U is a set of defined objects \( U = \{X_1, X_2, \ldots, X_n\} \) and A is a set of attributes \( A = \{a_1, a_2, \ldots, a_n\} \). U and A are non empty sets. For each attribute \( a \in A \), we associate a set of values \( V_a \) called domain of \( a \). It is possible to define an attribute in order to classify all cases. A Decision System DS=(U,A∪d) is an information system in which a decision attribute, d, influences the classification. If we consider a set of attributes \( B \subset A \) it is possible to define the following indiscernibility relation \( \text{Ind}(B) \): two elements \( (X_i,X_j) \in \text{Ind}(B) \) are indiscernible by the set of attributes B in A if \( b(X_i)=b(X_j) \) for each \( b \in B \). The equivalence class of \( \text{Ind}(B) \) \( B \) is called elementary set. Lower and upper approximation (figure 2) are defined respectively as the elements contained with certainty in the set and as the objects which probably belong to the set. The difference between upper and lower approximation defines the boundary of set \( X \).

\[
\begin{align*}
LX &= \{x_i \in U \mid x_i \in X \} & \text{Lower approximation} \\
UX &= \{x_i \in U \mid x_i \in X \} \cup X & \text{Upper approximation} \\
BX &= UX - LX & \text{Boundary}
\end{align*}
\]

\[\text{Figure 2 Boundary, Upper and Lower Approximation of a set } X.\]

The accuracy is defined as the ratio of the cardinality of the lower and the upper approximation:

\[\mu_B(X) = \frac{\text{card}(LX)}{\text{card}(UX)}\]

The result must be included between 0 and 1. An interesting application of this methodology is represented in the field of geographical information. The topological relationships are defined by the well known 9-intersection
model \citep{Egenhofer1991}. Afterward Clementini and Di Felice \citeyearpar{1996} developed an extension of this model called broad boundaries model. A broad boundary consists of an inner and an outer edge increasing the previous 8 cases to 44. Also, in rough set theory a topological model has been developed \citep{Wang2002}. This model is based on the concepts of positive region (Lower approximation), boundary region and negative region (Universe without Upper approximation).

3. THE CASE STUDY

Potenza Province is located in Southern Apennine, is characterised by low settlement density and has a population of 400,000 inhabitants, distributed on 650,000 hectares. The main town of the province is Potenza with 70,000 inhabitants, while the demographic consistency of the other municipalities can be classified in three groups. Twelve towns count more or less 12,000 inhabitants, twenty municipalities have a population around 5,000 inhabitants, the population of the remaining 68 municipalities varies from 700 to 2,000 inhabitants.

One of the phenomena more frequent in western European countries is the spreading of scattered settlements. Generally this phenomenon is not likely to be found in territories with low density of population. Small Municipalities, in spite of depopulation, present a huge diffusion of scattered buildings due to the abandonment of old town centres. In bigger towns, diffusion of settlements in rural areas is determined by the high cost of flats in urban areas. An accurate analysis of the phenomenon has been carried out using all the potentialities of Geographical Information Systems. All the polygons which represent the buildings have been converted in points. All these data have been updated and integrated using digital orthophotos. These analyses have highlighted that the phenomenon begins immediately outside the areas planned for future settlements by masterplans.

The cartographic update has allowed to estimate the effects of such diffused phenomenon on environment. The conversion procedure in vector points has been useful in order to use kernel density function. This function has had multiple uses, from the epidemics localization \citep{Gatrell1996} to studies on the spreading of city services \citep{Borruso2004}, while has not been used enough in the field of territorial planning. The point theme of buildings has been compared with the census zone. In urban areas each census zone corresponds to one building, while in rural areas a census zone coincides with a small rural centre; another census zone contains scattered buildings outside the small rural centres. Considering a homogenous typology of buildings inside each census zone, it has been possible to associate the number of dwelling-houses to each building. This attribute of the point theme has been used as a weight in kernel function, in order to understand the impact of the number of dwelling-houses on a particular area. In the case study, it has been used a bandwidth value of 400 m and a grid cell dimension of 10 m.

An interesting result has been obtained comparing the map achieved with kernel density in 2004 to those from 1987. It was possible to modify the point theme from orthophotos taken in two different dates. This comparison has allowed to characterize more precisely zones where the phenomenon of settlement dispersion has had a huge growth. After the localization of areas where the phenomenon is more considerable, it is important to understand the factors which could increase it. Considering the relationship between scattered buildings and areas in landslide and in high slope, it can be noticed, that these factors do not discourage settlement growth. Rural buildings, in fact, have been indiscriminately made on landslides and on high slopes. Comparing the theme achieved with the kernel density with other themes (fig. 3) it is possible to give further interpretations of the phenomenon of settlement dispersion. Scattered settlements can be located as a crown surrounding the urban area or along the main line of road network.
The greater increase of settlement dispersion has occurred in zones mostly situated on the mountains, while in zones with an intensive agricultural production scattered buildings are considered as a threat. In the study case the kernel density has been considered according to the following classes:

- it is reasonable to classify a region as rural if the presence of houses is less than one per hectare;
- from 1 to 5 houses per hectare, it is possible to define the periurban class;
- urban features are predominant beyond 5 houses per hectare.

Depth of contiguity zone, for each centre, has been localized using a shape index for the boundary of the urban area. The shape index is the ratio between the perimeter of the urban area and the perimeter of the circle that inscribes it. It is obvious that such an index can assume values greater than one. The more the value is greater than one, the more the shape of the settlement will be jagged, long and narrow.

A good level of compactness corresponds to a shape index comprised between 1 and 1.6, a medium level to values between 1.61 and 2.4, a low level to an index greater than 2.4. In the following table, the 100 municipalities of Potenza Province are grouped in three classes according to compactness rate.

<table>
<thead>
<tr>
<th>Compactness</th>
<th>Index value</th>
<th>Number of Centres</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good compactness</td>
<td>1-1.6</td>
<td>12</td>
</tr>
<tr>
<td>Medium compactness</td>
<td>1.61-2.4</td>
<td>68</td>
</tr>
<tr>
<td>Poor compactness</td>
<td>2.41-4.81</td>
<td>20</td>
</tr>
</tbody>
</table>

*Table 1 Centre classes based on compactness rate.*
Table 1 highlights the low level of compactness of urban areas of Potenza province: settlement dispersion is spread along the main roads without following morphological features. Two criteria have been considered for the contiguity: the first is the ratio between area and perimeter of the urban region, the second is the ratio between area and perimeter of the circle inscribing the urban region. Proximity to road network has been determined with the straight line distance, assigning to each cell a distance value. The phenomenon of settlement dispersion has been observed in various municipalities with different dimensions and this study indicates that new buildings are completely within a distance of 200 m from the road networks. This distance has been included among inclusive rules. The following exclusive rules have been considered: area included within a distance of 150 m from rivers and streams, slope higher than 35%, Nature 2000 sites, hydro-geological risk zones.

4. RESULTS AND FINAL DISCUSSION

Combining the previous rules with Boolean operators and map algebra theory, three different periurban areas have been identified:

- the first edge has been obtained considering all previous rules (inclusive-exclusive) and the first contiguity, which takes into account the ratio between area and perimeter of the urban region;
- the second boundary has been achieved considering all previous rules and the second contiguity, which considers area and perimeter of the circle inscribing the urban region
- the third zone does not take into account any contiguity.

As expected, results are rather different according to the type of periurban area. In most municipalities, the smallest area is achieved considering the first case, mentioned above. The biggest region is yield without considering any contiguity rule. This trend is completely reversed in the case of Potenza municipality. The periurban fringe obtained without taking into account any contiguity rule is the smallest one, because the kernel function captures a low density of buildings in these zones. This result implies that areas close to the urban region are represented by settlements with at most 2 dwelling-houses for buildings. In confirmation of this hypothesis the relationship between the two periurban fringes and the two contiguity rules hold the same sequence. However, the greatest part of cases follows the order "first contiguity rule - second contiguity rule - without any contiguity rule". Considering the morphology of settlement system of Potenza Province, constituted in most of cases by urban areas located on the spire of Apennine and scattered buildings concentrated near the

![Figure 4 Scheme of the procedure for the location of Peri-urban fringe.](image-url)
road network along the valleys, the contiguity rule does not entirely capture the phenomenon leading to a wrong interpretation.

Considering Avigliano municipality (fig.5), the use of contiguity rules implies the exclusion of north-western and south-eastern settlements. From figure 5 it is easy to understand that the transition from the first to the second contiguity rule does not cause a large increase of extension. In figure 5, without any contiguity rule, a huge increase of the periurban fringe is yielded, which better represents the actual situation. In all cases, the more reliable region is the third one which is achieved without any contiguity rule, because it identifies the areas in which new transformations are more likely to occur.

\[\text{Figure 5 Avigliano, middle size municipality, the squared hatch adopts the first contiguity, the hatch with lines uses the second contiguity rule and dotted hatch does not consider the contiguity.}\]

Rough Set theory has been applied for the classification of different geographic layers. In the case study the entire provincial territory represents the universe \([U]\), the inclusive rules and exclusive rules are the attributes \([A]\) (table 2) and the objects (the cells of the grid) have been classified following the rules previously defined. The indiscernibility relations have been calculated in \([U]\) considering the attributes, obtaining the subset \([X]\). Therefore, all the cells which satisfy at the same time the exclusive and the inclusive rules are assumed as a subset \([X]\), contained in \([U]\). The decision system in this case is composed by the set \(X\), the set of attributes where the decision variable is the nearest neighbor distance.

<table>
<thead>
<tr>
<th>Attributes and decision variable</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B=) Nature 2000 sites</td>
<td>1 different from Nature 2000 sites</td>
</tr>
<tr>
<td>(Fi=) Hydrography</td>
<td>2 if the distance is included between 150 m. and 800 m.</td>
</tr>
<tr>
<td></td>
<td>3 if the distance is bigger than 800 m.</td>
</tr>
<tr>
<td>(Fr=) Hydro-geological risk zones</td>
<td>1 different from classes R3 ed R4</td>
</tr>
<tr>
<td>(P=) Slope</td>
<td>1 if the slope is less than 23,6%</td>
</tr>
<tr>
<td></td>
<td>2 if the slope is included between 23,6% and 35%</td>
</tr>
<tr>
<td>(D=) Density</td>
<td>2 if the density is included between 1 and 5 inhabitants per hectare;</td>
</tr>
<tr>
<td>(V=) Road Network</td>
<td>1 if the distance is less than 200 m.;</td>
</tr>
<tr>
<td>(C1=) Real Contiguity</td>
<td>1</td>
</tr>
<tr>
<td>(C2=) Ideal Contiguity</td>
<td>1</td>
</tr>
<tr>
<td>(Nei=) Nearest neighbor distance</td>
<td>1 if the minimum distance less than 100m</td>
</tr>
</tbody>
</table>
Table 2 Attributes and decision variable for rough classification.

The set \( X \) has been obtained according to the decision variable (NEI) in order to achieve lower and upper approximation of the set \( X \). Three classifications have been done. Two ones take into account the first and the second contiguity rule respectively, the third one doesn’t consider any contiguity rule. Lower and Upper approximation, and the accuracy for Nei equal to 1 have been computed for each case. Table 3 summarizes the results achieved in each classification method, taking into account the cells number included in periurban fringe.

<table>
<thead>
<tr>
<th></th>
<th>Map Algebra</th>
<th>Rough set</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>Accuracy</td>
<td></td>
</tr>
<tr>
<td>Without Contiguity</td>
<td>1721768</td>
<td>72857</td>
<td>928856</td>
<td>7.84%</td>
</tr>
<tr>
<td>Real Contiguity</td>
<td>766864</td>
<td>138056</td>
<td>435856</td>
<td>31.67%</td>
</tr>
<tr>
<td>Ideal Contiguity</td>
<td>1208171</td>
<td>44214</td>
<td>523862</td>
<td>8.44%</td>
</tr>
</tbody>
</table>

Table 3 Comparison among the cells number included in periurban fringe for each classification method.

More interesting is the comparison among the results achieved with Map Algebra technique and the Rough Set method (fig.6). The Rough Set approach allows a better interpretation of periurban phenomenon. This method
doesn’t produce a sharp boundary which unambiguously defines the periurban edge, but it allows to locale in an indiscernible way the cells with the same attribute values taken into account for the classification.

Despite widely held that rough set evaluations are used to achieve more soft constraints, our results show that the classification with Map Algebra produces a peri-urban fringe with a larger extension, while the rough classifications generate a narrower fringe. Between the two rough classifications the first one, with real contiguity, produces an acceptable accuracy and a better interpretation of the phenomenon.

REFERENCES
Batty M., Besussi E., Maat K., Harts J. (2003), “Representing Multifunctional Cities: Density and Diversity in Space and Time” CASA Working Papers, ISSN: 1467-1298
A Topological Interpretation for Mass Transit Network Connectivity
Chulmin Jun, Seungjae Lee, Hye Young Kim and Seungil Lee
The University of Seoul,
Dongdaemoon-Ku, Seoul, Korea
E-mail: {cmjun, sjlee, mhw3n, silee}@uos.ac.kr

ABSTRACT
The City of Seoul recently reformed the entire bus system as part of an intensified effort to relieve traffic congestion by encouraging mass transit means. Although the reformed system receives relatively positive reactions, it is still viewed to need improvements in route organization. Evaluation and organization of mass transit system becomes complex as the number of transportation modes increases. This paper presents an alternative method to assess the connectivity of mass transit system based on the topological structure of the routes network. It shows how to compute the connectivity of current transport routes configuration of multi-modal network and describe the results using the GIS. This study employs the hierarchical computation process of space syntax theory. The main methodological issue starts from the fact that the more transfers take place, the deeper the connectivity becomes making that area evaluated as less advantageous as for mass transit accessibility. By computing the connectivity of each bus or subway station with all others in a city, we can quantify the difference in the serviceability of city areas based on the mass transit. In computing the paths of origin-destination of routes, it employs the genetic algorithm. The process is illustrated using a network data of Seoul City built in a GIS.

Keywords: mass transit, transfer, space syntax, genetic algorithm, GIS

1. INTRODUCTION
Traffic congestion in the City of Seoul has lead to public-oriented transportation policy that encourages use of mass transit instead of privately-owned cars and recently the city government reformed the bus system on a large scale. However, there are some criticisms that the reformed system still partly shows over- or under-supply throughout the city area. Differences in accessibility to other areas from bus stops or subway stations cause differences in time, expenses and metal burden of users who travels the same distances. Current limitations in public transport planning call for robust methodology to assess the accessibility or serviceability of the transport routes.

Space syntax is the technique that has been used to derive the connectivity of urban or architectural spaces (Hillier 1996). The theory has primarily been applied in the research areas that seek to find the movement of human beings among indoor spaces or pedestrian paths and it has helped to compute the connectivity of the network of built environment quantitatively based on the topological structure of spatial links. The theory sees that spatial structure or layout has great impact on human social activities and displacement (Hillier 1984, Jiang 1999). Its primary principle is to model a spatial structure as a set of axial lines and compute spatial indices to derive the relationships between different parts of urban or indoor spaces. The resulting index is expressed as the integration of that space which is the degree to which that space is integrated and connected with other spaces in the defined area. Although transportation network problems are not included in typical applications of space syntax, we see the analogy between space syntax’s spatial integration and mass transit links in that both are based on the hierarchical transitions between spaces.

This paper proposes an alternative method to evaluate connectivity of mass transport network based on its topological structure and the computational principle from space syntax. The primary idea of the paper is that we consider a transfer of vehicles as a connection node that links two different routes, and the more transfers take place, the deeper the connectivity becomes making that area evaluated as less advantageous as for public transport accessibility. We suggest an algorithm to show how geometric accessibility based on the connectivity of transport routes, rather than their physical distance, can be computed. Also, in order to calculate the optimal path from an origin to a destination, we used genetic algorithm. Transport network data including routes and stops were built in a GIS and relational database and the algorithm was programmed in C# language. The bus and subway network data set in Seoul are used to illustrate the proposed algorithm.
### 2. AXIAL LINE-BASED NETWORK CONFIGURATION

Human movement is frequently described in an abstracted form using its topology. Topological description allows researchers to focus on the structural relationship among units of movement while disregarding the details of phenomena. For example, pedestrian movement can be described using network of simple lines without considering the details such as sizes of forms, number of people and speed of movement. Such network configuration is also referred to as graph, which is a way to represent a network by a set of vertices and a set of edges that connects pairs of vertices. Figure 1 illustrates how meandering streets can be mapped to a graph. Following space syntax principle, spaces are first broadly perceived as discrete components, for instance, linear lines, and then are combined forming a continuous network. These lines are called ‘axial lines’ in space syntax.

![Figure 1. Axial lines of a street network](image)

When spaces are mapped to a graph, the hierarchical relationship of component units is obviously captured. All this is best illustrated using Figure 1. Line 2 is accessible from line 1 by one turn, whereas line 4 is accessed by two turns. In other words, the relationship of 2 and 3 is called symmetrical with respect to 1 whereas the relationship of 4 to 1 is asymmetrical. In the literature of space syntax theory, this relationship is described through a variable called depth (Bafna 2003). If one were to represent each component with a node and a turn with a link connecting their respective nodes, one could then describe the hierarchy from each node as shown in Figure 2. Figure 2 shows the hierarchy from node (or street) 1.

![Figure 2. Hierarchical representation of street ‘1’ from Figure 1](image)

Depth of one node from another can be directly measured by counting the number of steps (or turns) between two nodes. The greater the depth of two nodes, the greater the hierarchical difference between them. The depth of a node (or a street) is defined by the number of nodes distant from a given number of steps to that node. If we take the example of Figure 2, the depth of node 1 for immediate neighbors (eg. step 1 nodes) is 2 since there are two nodes that can be accessed by one turn. On the other hand, the depth of node 1 in 2 steps distance is 4 since there are two nodes that can be reached by two turns, that is, 2 (nodes) × 2 (steps). Thus, the total depth from a node to all other nodes can be measured by summing the product of the level of step and the number of nodes in that step as given by:
\[ TD_i = \sum_{s=1}^{m} s \times N_s \]  

(1)

, where

TD\(\_\)i : the total depth of node \(i\)

s : the step from node \(i\)

m : the maximum number of steps extended from node \(i\)

Ns : the number of nodes at step \(s\)

The mean depth then is given by the total depth divided by \(k - 1\), where \(k\) is the total number of nodes in the graph (Hillier 1996). This means the average depth of a particular node. Figure 3 shows extreme cases from node 1 in a network of same number of nodes. One case contains a node that extend to the maximum number of steps, which is \(k - 1\) with the rest of nodes, one in each of intervening steps (case b in Figure 3), and the other contains only neighboring nodes to node a (case a in Figure 3).

![Figure 3. Symmetrical(a) and asymmetrical(b) layouts of streets](image)

In the case (a), completely symmetrical structure, in Figure 3, MD is computed as follows,

\[ MD = \frac{k - 1}{k - 1} = 1 \]

whereas MD of the case (b) is computed by

\[ MD = \frac{1 + 2 + \ldots + (k - 1)}{k - 1} = \frac{(k - 1)k / 2}{k - 1} = \frac{k}{2} \]

Here, \(1 \leq MD \leq \frac{k}{2}\) is derived. Therefore, MD is normalized as follows:

\[ 0 \leq \frac{2(MD - 1)}{k - 2} \leq 1 \]  

(2)

Now, using the normalized depth (ND), the depth from a node in a graph can be represented by a number ranging from 0 to 1. Using ND values makes it possible to compare depths of nodes from graphs with different number of nodes.
3. APPLYING TO MASS TRANSIT PROBLEM

The hierarchical description in space syntax summarized in the previous section primarily targets the abstraction of free movement of people mostly in built spaces rather than the movements in the mass transit which take place along fixed routes. However, we can derive the similarity of these two problems. Hierarchical framework of space syntax focuses on turns of spaces which are the basis for computing the depth of a certain space to others, while the mass transit problem generally entails transfers between vehicles. Thus, we can map turns of spaces to transfers between transportation means. In the hierarchical network description, the deeper the depth from a space to others, the more relatively difficult it is to move from that space to others. On the other hand, in mass transit, cost generally increases as the number of transfers between different modes increases. In this case, the cost can be either total fares or time taken in transfers, or it can even be seen as the mental burden that a traveler feels when he or she moves to or waits for the next vehicle in transfer areas. If we map the components using nodes and links in the previous section to a mass transit network, a node (or a street in Figure 1) can be seen as a stop, regardless of bus or subway, whereas a link between two nodes can be mapped to a transfer between two vehicles. This relationship is illustrated in Figure 4.

![Figure 4. Mass transit network](image)

If a person moves from stop 1 to stop 2, 3 or 4, he or she does not need to transfer because these stops are on the same route, subway line 1. However, if the traveler wants to go to stop 5 or stop 6 from stop 1, he or she has to transfer in the area A since stop 1 and stop 5 are not on the same route. Similarly, if the origin is stop 1 and the destination is stop 7 or 8, one transfer is needed in area B and, if the destination is stop 9, he or she has to transfer two times, one in area B and then in area C. One transfer from a transportation mode to another is the ‘spatial transfer’ which becomes one depth between spaces. If this network were that of pedestrian streets, when a person were to move from point 6 to point 7, he or she needs two turns, which means these points are two depths away from each other. However, in case of using pre-laid transport routes, the existence of a route that connects two points is first taken into account in computing the depths. Therefore, no transfer is needed in case of moving from stop 7 to 6 because there is a direct line connecting these two points.
The relationship of stops via routes is described in Figure 5. The connectivity from each stop to all others is hierarchically mapped to a graph. The procedure for generating a graph is iterative, starting with a stop and then progressively identifying the next neighboring stops until the entire stops are covered. The procedure first identifies the stops that are directly accessible from an origin using one route (step 1), then among these stops finds those stops that are belonged to transfer areas. Then, it looks for the routes that share these stops in the transfer areas. Next, it finds those stops that can be reached using the identified routes (step 2). It again looks for the stops belonged to the transfer areas. It continues iteratively in this manner.

Here, we can assume symmetrical and asymmetrical cases as we did in the previous section (Figure 6). The first case is when all stops are accessed from an origin via only one route. On the contrary, the other case is when all stops and routes are laid such that every stop is accessed by different means from the previous means. The former one is the completely symmetrical case, and the latter is completely asymmetrical one. Eqs. (1) and (2), which were defined for space connectivity, hold true for public transport network. We can then compute TD, MD and ND as shown in Table 1.

Figure 5. Hierarchical representation of network connectivity

Figure 6. Symmetry and asymmetry of the route connectivity
Note that the reciprocal of ND is also calculated. These values help intuitive interpretation about the relationship between the graph and the accessibility. That is, higher values of nodes indicate that the stop is less deep on an average from all other stops and, in other words, shows better accessibility to other stops on an average. As one may easily expect, stop 3 shows the highest accessibility, 14.0, followed by 9.33 of stop 2. Note that stop 7 ranks in the third. It’s because stop 7 has the routes that pass transfer areas, which makes it possible for a traveler to go to any other stops from stop 7 by only one transfer.

4. INTEGRATING INTO GIS

1) Building GIS Data

In order to apply the procedure proposed here to the mass transit network problem, we should consider utilizing GIS capabilities due to complexity and size of the network. However, current GIS data of the City of Seoul available to us as of now have limitations in reflecting the information necessary for the computation process suggested in this study. Mass transit networks of Seoul are composed of different modes such as metro bus, short-distance connection bus, airport limousine and subway and each of these modes is built in separate GIS data set. Also, a typical GIS data set is composed of geographical feature data and the table data, each record of which reflects its corresponding geographic feature. Thus, currently used GIS data structure alone can not capture the complex relationship of characteristics in mass transit.

To make the GIS network data of different modes usable, we first needed to organize them into a spatially integrated data set, where coordinate system and topology are matched against each other. Then, we added some attribute data tables that contain information about relationships between GIS network data features. The relationship among streets, routes, stops and transfer areas can be abstracted into an entity-relationship model in a relational database as shown in Figure 7. The procedure begins by defining a relation for each entity. One street section may contain more than one bus or subway route. Therefore the relationship between streets and routes is one to many (1:N). On the other hand, the relationship between routes and stops is many to many (M:N) because a route may include many stops and a stop may be shared by more than one route. We also need information about transfer areas for the computation process. Since one transfer area may contain more than one stop, the relationship of these two entities is one to many (1:N). The entities included in the E-R model are classified into two types; one is those that are built as the attribute tables of GIS data and the other is separately built tables. The former case includes Street, Route, and Stop entities and the latter Transfer Area. Also, the intersection entity generated from M:N relationship is also has to be created as a separate table. Since the data are constructed using such different formats, we cannot use macro languages provided as a subsidiary function of a proprietary GIS package. Thus, we used C# language to implement the proposed algorithm. We first created a test version using artificial networks similar to the one illustrated in Figure 4 and then extended it to the real networks of the Seoul City as illustrated in the following section.

<table>
<thead>
<tr>
<th>Stop No.</th>
<th>TD</th>
<th>MD</th>
<th>ND</th>
<th>ND⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>1.750</td>
<td>0.214</td>
<td>4.67</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>1.375</td>
<td>0.107</td>
<td>9.33</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>1.250</td>
<td>0.071</td>
<td>14.00</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>1.750</td>
<td>0.214</td>
<td>4.67</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>2.125</td>
<td>0.321</td>
<td>3.11</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>1.625</td>
<td>0.179</td>
<td>5.60</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>1.500</td>
<td>0.143</td>
<td>7.00</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>1.750</td>
<td>0.214</td>
<td>4.67</td>
</tr>
<tr>
<td>9</td>
<td>21</td>
<td>2.625</td>
<td>0.464</td>
<td>2.15</td>
</tr>
</tbody>
</table>
2) Generating a Path using GA

Computing the depth of a stop or a station in a target area entails finding paths from that stop to all others in the area, each of which being the minimum-cost path. In this study, a minimum-cost path is defined as the one that has the minimum number of transfers between the origin and the destination.

Finding an optimal path in a mass transit network is much more complex than in the network of privately-owned cars. It is because the network is composed of multi-modes of vehicles and also has time-constraints at transfer areas. If a vehicle has a list of pre-specified departure times and transfers to other modes take place, the departure time is constrained for each available mode and its departure schedule, and comparison among these different departure times needs to be performed in order to explore the minimum time path (Desrochers et al. 1988).

Genetic algorithms aim at such complex problems. The study used the GA-based approach in finding the minimum-cost paths. Although we avoid detailed description about GA because there are many books or papers that introduce GA principles, we summarize the major part that shows how GA was modified to fit the mass transit problem.

As shown in Figure 8, in GA, a global search process is performed on a certain population of chromosomes by gradually updating the population. Search processes are conditioned by two objectives: exploiting the best solutions and exploring the search space (Michalewicz 1994). The process for creating the first population is called the initialization. The updating processes of the population, the creation of successive generations, are done using so-called the genetic-operators: crossover and mutation. These genetic operators alter the composition of children of parent chromosomes. The search process is continued until it reaches the maximum number of generations while searching for “better” solutions that are evaluated by the “fitness” function. Therefore, the fitness function along with some parameters such as population size and probabilities of applying genetic operators are required in advance. Figure 9 shows a simple example of a network where different types of vehicles are present. In nodes such as 3 or 7, transfer does not happen. But the rest nodes allow the traveler to transfer to another mode.
**Representation:** In this study, a chromosome is represented by linking the stops from the source to the destination. If the source is Node 1 and the destination is the Node 9, a chromosome is an array of nodes that include Node 1 at the first position and Node 9 at the last.

**Initialization:** The initial population of chromosomes is created according to the preset population size. All nodes for each chromosome are initialized randomly as the following manner;

\[
\begin{align*}
C_1 &= (1, 2, 8, 9) \\
C_2 &= (1, 4, 5, 6, 9) \\
C_3 &= (1, 2, 5, 6, 7, 8, 9) \\
\ldots
\end{align*}
\]

**Evaluation:** The evaluation function or the fitness function plays the role of the environment, rating potential solutions in terms of their fitness. Evaluation function \( \text{eval} \) for node vectors \( C \) can be set as the total time taken in the path. For this study that considers the connectivity of paths, we used the number of transfers taken from the origin to the destination instead of the total time as follows;

\[
\text{eval}(C) = \text{gene_transfers}(x),
\]

**Selection:** Selection is a preparatory process that is needed for updating the current population. In order to preserve good chromosomes, some of them are reproduced in the next generation instead of participating in the mutation or crossover. This way, we can prevent those elite chromosomes from being deleted in the process. Selection process also includes the process that selects the parent chromosomes for crossover or mutation, which is described in the following section.

**Genetic Operators:** Some members in the initial population undergo alteration by means of two genetic operators: crossover and mutation. Crossover combines the features of two parent chromosomes to form two similar children by swapping corresponding segment of the parents. For example, if the parents are \( C_2 \) and \( C_3 \), then a common node (e.g. Node 5) can be selected and the portions of chromosomes after this node are crossed generating new children:

\[
\begin{align*}
C_2 &= (1, 4, 5, 6, 9) \quad \Rightarrow \quad C'_2 = (1, 4, 5, 6, 7, 8, 9) \\
C_3 &= (1, 2, 5, 6, 7, 8, 9) \quad \Rightarrow \quad C'_3 = (1, 2, 5, 6, 9)
\end{align*}
\]

Mutation arbitrarily alters the positions of one or more genes. In the transportation example, just exchanging a certain gene can generate a chromosome having disconnected link of nodes. Thus, we can modify the mutation process to fit this problem. If a certain gene is selected as the target of mutation, it can be thought of the temporary origin and then a portion of chromosome is created that reaches the destination. Assume \( C_3 \) has been selected and third gene, Node 5 has been selected as the mutation. Then, Node 5 becomes the temporary origin yielding a chromosome from this node to Node 9. After the mutation, new \( C_2 \) can be created as

\[
\begin{align*}
C_2 &= (1, 4, 5, 2, 8, 7, 6, 9) \quad \Rightarrow \quad C'_2 = (1, 4, 5, 2, 8, 7, 6, 9)
\end{align*}
\]

As seen from this, mutation can either increase or decrease the value of selected chromosomes.

Figure 9. An example of a multi-modal network
3) Case study

In order to show the feasibility of the method proposed here, we chose a CBD area called ‘Kangnam District’ of Seoul that shows high complexity in the network structure. Since the area contains a variety of modes such as subway and different types of bus lines, it was viewed suitable for the study.

The bus stops that are located near each other are grouped into a transfer area as shown in Figure 10, and the Integration value(ND) was computed for each of these transfer areas.

Figure 10. Transfer areas in the test area

Figure 11 shows the computed Integration(ND) for the study area. The stops grouped in a transfer have the same I-values. For the purpose of visual recognition, we divided the resulting I-values into 5 classes in the same color scheme. The darker the color, the higher the I-value becomes meaning the stop is located in more connected area with other locations.

Figure 11. Integration(ND) values for stops in the test area

5. CONCLUDING REMARKS

This paper presented an alternative method to assess the connectivity of mass transit network by defining the network relationship onto a graph of hierarchical structure. We derived an analogy between the concept of depths in space syntax and the degree of connectivity of network of transport routes. We interpreted the spatial transitions between spaces in space syntax as the transfers between transportation modes. We developed an algorithm to automate the computing process using GIS data. In order to generate optimal paths during the process, we employed the Genetic Algorithm. We constructed a relational database model to capture the
relationship between GIS data features and added attribute tables. By applying the proposed method to a CBD area, we could quantify the differences in the connectivity of city areas using the space syntax values.

We plan to expand this algorithm to more real situations such that it can incorporate origin-destination movements and real distances in the paths. In that case, the term ‘the topological connectivity’ used here can be modified to more general term such as ‘the accessibility’.

ACKNOWLEDGEMENT

This study was supported by the 2005 University Research Fund of the University of Seoul.

REFERENCES

Implementation of a generic road-matching approach for the integration of postal data

Meng Zhang, Liqiu Meng

Department of Cartography, Technical University of Munich
Arcisstr. 21, 80333 Munich, Germany
(meng.zhang@bv.tum.de, meng@bv.tum.de)

ABSTRACT

A generic matching approach for road networks is proposed in this paper and implemented under ArcGIS 9.0. By means of an innovative matching algorithm termed as “Unsymmetrical Buffer Growing”, the proposed approach yields high matching performance indicated by an average matching rate of 96.1% on large data sets of road networks. Meanwhile, our matching system has gained some self-learning abilities. Furthermore, a measurement of matching certainty on the basis of geometrical and topological similarity is introduced, which supports the human operator to judge the quality of matching results and easily detect possible matching errors.

Keywords: data integration, matching, road networks, Unsymmetrical Buffer Growing, matching certainty

1 INTRODUCTION

1.1 Motivation

The project “Integration of postal data” which is a co-operation between the Department of Cartography, Technical University of Munich, and German Federal Agency for Cartography and Geodesy requires that the road layer of Basis Digital Landscape Model (Basis DLM) be enriched with geo-referenced house numbers of post addresses. The available post addresses from German Federal Post Office were manually collected by postmen on the basis of road geometries from TeleAtlas Corp. Since the road database of TeleAtlas reveals a different geometric / semantic accuracy from that of Basis DLM, the attempt of a direct integration of postal data in Basis DLM is doomed to fail. Therefore, we divide the enrichment process into two main stages. The first stage is dedicated to the matching of road objects between Basis DLM dataset as reference and TeleAtlas dataset as target. In the second stage, a projection based on the rubber sheet principle is established for each pair of matched road lines. Thus, the discrete locations of house numbers as well as any other arbitrary point along a TeleAtlas road line can find its corresponding position along the homologous line in Basis DLM. As “rubber sheet” projection has already become a standard function available in actual ArcGIS version, our work is focused on the task of road matching between Basis DLM and TeleAtlas.

1.2 Test data

As indicated in the section 1.1, three datasets are investigated in our approach:
(a) Road layer of Basic Digital Landscape Model (Basis DLM)

The Basis DLM is from German mapping agencies. Its road layer is composed of geometries and general-purposed attributes of road lines (middle axes) which were captured through map digitization in combination of semiautomatic object extraction from imagery data. The data structure is defined in accordance with the terms of the Official Topographic Cartographic Information System (ATKIS). However, the road attributes are not completely covered with values, especially the street names which are essential clues for the matching are only sporadically available.

(b) Road objects from TeleAtlas Corp.

This dataset contains geometries and navigation-oriented attributes of road lines (middle axes) which were captured through map digitization, GPS-supported field measurement and dynamic supervision of traffic information. The data structure is defined by TeleAtlas which is one of the most important data suppliers for car navigation systems.

(c) Post addresses from German Federal Post Office

The individual house numbers are stored as discrete points distributed along the two road sides. The location of each individual house number was estimated or interpolated by the postman based on the beginning and terminating house number delimiting each street segment and the known house-numbering rules. For this reason, it may deviate more or less from the true location of the corresponding house entrance, but the topological relationship is preserved.

1.3 Related work

Geo-data matching has been a topic of intensive research since a decade or so and it is getting more and more complex with the increasing availability of diverse geospatial databases (Devogle et al. 1996; Dunkars 2003; Ochieng et al. 2003; Anders & Bobrich 2004; Meng & Töllner 2004).

In (Walter 1996) a geometrical matching approach for datasets GDF and ATKIS is proposed. To achieve the correct matching result, the combination of different methods, such as “Buffer Growing”, Angle-, Length- and shape- Comparison, are built and optimized in his work. Volz (2006) has extended the work by applying an iterative approach. He initializes the matching process by identifying seeds which show a high likelihood of correspondence. A combined edge and node matching algorithm is then used to detect 1:1 correspondences. In case no 1:1 match could be found, an enhanced edge matching approach is triggered to recognize 1:2 matches. The whole process runs in multiple iterations with stepwise relaxed constraints.

Mantel & Lipeck (2004) develop a multistage matching procedure for automatic updating the objects in a multiple representation database. The developed procedure is composed of semantic classification, an algorithm to compute possible match, rule-based selection and the interactive improvement of the matching result.

Zhang, Shi & Meng (2005) find out, that an accurate match can be hardly reached if only shape and location are compared. As shown in figure 1, in terms of shape and location, the polyline $I\rightarrow N-2$ is identified as the perfect match for the reference polyline $O\rightarrow O'$. In reality, however, the correct polyline should be $3\rightarrow N-1$ if topological relationship is considered. Therefore, they have included the topologic matching in their approach.
2 THE STRATEGIC MATCHING APPROACH

Matching rate and matching certainty are both significant criteria reflecting the quality of a matching approach. In most cases, high matching certainty is more important than high matching rate. E.g. a matching with 99% certainty and 70% matching rate is obviously more useful than one with 99% matching rate but 70% certainty. Hence the principle of our generic matching approach sounds as follows: striving for a nearly perfect match for the majority of line objects, finding a hypothetic match for as many remaining line objects as possible, keeping the computing speed as high as possible.

Figure 2 depicts the key components of our generic matching approach. First, the matching region is divided into several smaller areas. Then in the scope of every smaller matching area, the corresponding road objects between
two different datasets will be matched to each other. This process goes through three blocks: data preprocessing, matching of the road network and utilization of “Unsymmetrical Buffer Growing”. The preprocessing block aims at reducing the noise of irrelevant details, describing and improving the topology in the datasets to be matched. In the matching block, only geometric and topologic information can be utilized due to missing values of essential semantic attributes in many matching cases. The block “utilization of Unsymmetrical Buffer Growing” is driven by the analysis of matching results after the first iteration. If a systematic geometric deviation exists within the scope of the current matching area and its value is large enough (e.g. > 5m), an iterative process of “Unsymmetrical Buffer Growing” will be triggered, which leads to the reduction of matching errors.

2.1 **Division of the matching region**

In the proposed matching approach, “ODBC” is adopted to connect the database. The larger the databases to be matched, the slower the searching operation will be. To overcome this drawback, we divide a large matching region (e.g. > 20 km$^2$) into several smaller areas which define limited searching scopes. Such a division also enables the utilization of the new matching method - “Unsymmetrical Buffer Growing”, which is explained in section 2.4.

2.2 **Data preprocessing**

The process of data preparation is characterized by three procedures: reducing the noise or irrelevant details, describing topology with node-geometry and improving the topological differences. The preprocessed data are stored in a temporary file and used for the matching. However, the matching results are stored with their original geometries and attributes.

2.2.1 **Reducing the noise or irrelevant details**

By comparing the number of object classes, the overall distribution density, the relative distribution density of each object class, and the number of attributes existing in the two datasets, i.e. the reference dataset and the target dataset for the matching, it is possible to exclude those classes and attributes that exist only in one of the datasets. Further, geometries, topological relationships and semantic details of the dataset at higher resolution are simplified until its level of details approaches that of the other dataset.

2.2.2 **Describing topology with node-geometry**

Topology acts a very important role in the matching process. Zhang, Shi & Meng (2005) applied the variable “TopoR_Node” - the valence of a certain point, to roughly describe the topology (see Table.1).

<table>
<thead>
<tr>
<th>Example</th>
<th>A</th>
<th>A</th>
<th>A</th>
<th>A</th>
<th>A</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>“TopoR_Node” of node A</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Tab.1 The values of TopoR_Node (Zhang et al. 2005)

Unfortunately, such a rough description is not yet adequate for the topologic matching. As shown in figure 3, the nodes “P₀” and “P₁” have the same “TopoR_node” but different topology. Identifying them as a matched pair would cause a topological confusion.
2.2.2 Enriching the topological attributes

In order to enrich the topological attributes, four further variables $\text{Typ}_{\text{TopoR}=3}$, $\text{Winkel}_{\text{TopoR}=3}$, $\text{Typ}_{\text{TopoR}=4}$ and $\text{Winkel}_{\text{TopoR}=4}$ are defined, by means of which the topology of the nodes which have their "TopoR Node" equal to three or four can be represented more accurately (see Table 2 and Table 3).

![Diagram of topology](image)

**Fig.3** An example of confusing topology

In order to enrich the topological attributes, four further variables $\text{Typ}_{\text{TopoR}=3}$, $\text{Winkel}_{\text{TopoR}=3}$, $\text{Typ}_{\text{TopoR}=4}$ and $\text{Winkel}_{\text{TopoR}=4}$ are defined, by means of which the topology of the nodes which have their "TopoR Node" equal to three or four can be represented more accurately (see Table 2 and Table 3).

<table>
<thead>
<tr>
<th>Classification</th>
<th>$\text{Typ}_{\text{TopoR}=3}$</th>
<th>$\text{Winkel}_{\text{TopoR}=3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{A}$</td>
<td>$0$</td>
<td>$\text{The direction of } \overrightarrow{AA'}$</td>
</tr>
<tr>
<td>$\text{A}$</td>
<td>$1$</td>
<td>$\epsilon [0, 360^\circ]$</td>
</tr>
<tr>
<td>$\text{A}$</td>
<td>$2$</td>
<td>$\text{The direction of } \overrightarrow{AA'}$</td>
</tr>
</tbody>
</table>

| The nodes whose "TopoR Node" is equal to 3. |

<table>
<thead>
<tr>
<th>Classification</th>
<th>$\text{Typ}_{\text{TopoR}=4}$</th>
<th>$\text{Winkel}_{\text{TopoR}=4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{A}$</td>
<td>$0$</td>
<td>$\text{The direction of } \overrightarrow{AA'}$</td>
</tr>
<tr>
<td>$\text{A}$</td>
<td>$1$</td>
<td>$\epsilon [0, 360^\circ]$</td>
</tr>
</tbody>
</table>

| The nodes whose "TopoR Node" is equal to 4. |

2.2.3 Reduction of the topological differences

Though the noise and irrelevant details are reduced in the step 2.2.1, there are still topological differences between the datasets to be matched. Sometimes a node in one dataset may correspond to two or more adjacent...
nodes in the other. In order to build the unequivocal connection between the different datasets, the ambiguous nodes in the second dataset should be aggregated to a central point, as in figure 4.

Fig. 4  Further reduction of topological differences through node aggregation

2.3  Matching of road networks

For line matching, Buffer Growing (BG) is an efficient algorithm when the same real world objects from different datasets are similar in shape, location and topology (Mantel & Lipeck 2004, Walter 1996). The proposed matching process starts with the comparison of geometrical and shape similarity between each line object from the reference dataset (Basis DLM) and the candidate objects from the target dataset (TeleAtlas). The following steps are involved in the overall matching process.

2.3.1  Instantiation of the reference polyline

A sequence of line objects sharing certain characteristics from the reference dataset is chained to form a polyline and acts as the matching reference (cf. polyline $O\rightarrow O'$ in figure 5).

Fig. 5  Buffers around the individual segments of a reference polyline

2.3.2  Initialization of the scope of buffer

The buffer-growing principle requires the initialization of two parameters, “Buffer-P” and “Buffer-V”, which represent the scope of the buffer illustrated in figure 5. Proper parameter settings may benefit the matching result and on the other hand excessive large or small ones will probably lead to false or inefficient matching. The initial values of Buffer-P and Buffer-V are empirically determined.
2.3.3 Identification of possible matching candidates

A buffer is built around every segment of the matching reference. Then all the line objects from the target dataset that entirely fall inside these buffers are selected. These line objects can also be chained according to certain rules to generate one or more polylines, which are regarded as possible matching candidates, e.g. polyline $A \rightarrow G$ and $H \rightarrow K$ shown in figure 5. In other words, this step may result in the identification of an M-N-match ($M \geq 1$ and $N \geq 1$).

2.3.4 Exclusion of incorrect candidates

The set of possible matching candidates contains not only correct match, but some wrong suggestions. In order to exclude the false candidates, various measurements of angle, length, chord and location can be applied as follows:

1. $\Delta \beta = |\beta_{\text{reference}} - \beta_{\text{candidate}}| < T_\beta$ (e.g. 15°)

   $\Delta \beta$: Angle difference between reference and candidate, where $\beta$ is the angle between the straight line connecting the two end points and the x-axis.

2. $\Delta L = |L_{\text{reference}} - L_{\text{candidate}}| < T_L$ (e.g. 20m)

   $\Delta L$: Length difference between matching reference and candidate.

3. $\Delta C = |d_{\text{reference}} - d_{\text{candidate}}| < T_C$ (e.g. 12m)

   $\Delta C$: Maximal chord difference between matching reference and candidate.

4. $\frac{S_{\text{reference}} - S_{\text{candidate}}}{L_{\text{reference}} + L_{\text{candidate}}} < T_S$ (e.g. 10m)

   $S_{\text{reference}}, \text{ candidate}$: Area of the polygon enclosed by the reference and candidate.

where, $T_\beta, T_L, T_\delta$ and $T_s$ are thresholds. If necessary, other criteria can be also defined for specific data types.

The target polylines fitting for all of these criteria will be confirmed as the promising matching candidates, while others will be rejected.

2.3.5 Adjustment of the buffer parameter

On occasions, there may be no candidates fully fitting all the criteria, although the correct match does exist. One reason that can explain this phenomenon is traced back to the possibly too small buffer setting. On the other hand if the buffer is set too large in the beginning, many wrong candidates would be identified which may lead to erroneous chaining of line objects falling within the buffer. To overcome this problem, the scope of buffer should be stepwise enlarged until the proper candidate is detected or the buffer size is large enough.

2.3.6 Exactness inspection of the matching candidates

In some cases, more than one candidate would pass the exclusion criteria in Step 2.3.4 and each candidate is likely represented by a polyline that is longer than the reference, as the buffer around each line segment is always longer than the segment itself. In other words, the found candidate(s) are still not necessarily the exact match. In order to get the exact matching result, further information is needed. As shown in expression [1], the
variable “Matching_SumW” is defined as the measurement of the geometrical and topological similarity between the reference and its matching candidate. With “Matching_SumW” it is possible to value the matching accuracy.

\[
\text{Matching}_{-}\text{SumW} = F\left(\text{Geo}_{-}\text{Similarity}, \text{Topo}_{-}\text{Similarity}\right) \quad \ldots [1]
\]

Where,

\[
\text{Geo}_{-}\text{Similarity} = f(\Delta \beta, \Delta L, \Delta C, \overline{S}, a_{FS,FS}, a_{ES,ES}) \quad \text{Topo}_{-}\text{Similarity} = g(T_{Sr}, T_{Sr}, T_{Ec}, T_{Ec})
\]

\(\Delta \beta, \Delta L, \Delta C, \overline{S}\): see step 2.3.4

\(a_{FS,FS}\): Angle between the first line segments of the matching reference and candidate

\(a_{ES,ES}\): Angle between the last line segments of the matching reference and candidate

\(T_{Sr}\): represents the topology of the starting point of the matching reference

\(T_{Ec}\): represents the topology of the starting point of the matching candidate

\(T_{Sr}\): represents the topology of the terminating point of the matching reference

\(T_{Ec}\): represents the topology of the terminating point of the matching candidate

The variable “Matching_SumW” can be scaled to a number between 0 and 1, with 0 indicating an entirely wrong match, and 1 a perfect match. The procedure of exactness inspection is illustrated in figure 6.

Fig.6  Exactness inspection of matching results
2.3.7 Establishment of the links between the matched pair

In order to interactively visualize the matching results and transfer attribute information between the target dataset and the reference dataset in ArcGIS, it is necessary to establish the links between the corresponding road objects. The links can be generated with the following steps:

a) Find out the corresponding nodes with TopoR_Node \( \geq 3 \) between two datasets;
b) Split the polyline into smaller sections with such nodes as terminating points;
c) By means of interpolation, match the turning points in every divided section.

The identified corresponding coordinates are stored in a database and act as the link file for further processes in ArcGIS.

2.4 Unsymmetrical Buffer Growing

In the traditional BG-process, the buffer is symmetrically built around each line segment. However, if there is a systematic geometric drift in the matching area, such a symmetrical buffer may result in low matching rate and high uncertainty because the proper matching candidate (e.g. polyline \( C \rightarrow D \) in figure 7-a) may likely lie outside the buffer whilst some wrong candidates (e.g. polyline \( E \rightarrow F \) in figure 7-a) happen to fall inside.

To overcome the drawback of symmetrical BG, we introduced an iterative learning component in our algorithm. After each round of iteration a statistic analysis is automatically triggered. The mean \((\bar{x}, \bar{y})\) and variance \((\delta_x, \delta_y)\) of the position discrepancies between the matched pairs are used to indicate the possible systematic geometric drift \((G_{\text{dev}}\_x, G_{\text{dev}}\_y)\) in the current matching area:

\[
G_{\text{dev}}\_x = f(\bar{x}, \delta_x) \quad G_{\text{dev}}\_y = f(\bar{y}, \delta_y) \quad \ldots \quad [2]
\]

where, 
\[
\bar{x} = \frac{\sum_{i=1}^{n} l_{i,\text{ex}} - l_{i,\text{en}}}{n} \quad \bar{y} = \frac{\sum_{i=1}^{n} l_{i,\text{ey}} - l_{i,\text{en}}}{n} \quad \delta_x = \frac{\sum_{i=1}^{n} (l_{i,\text{en}} - l_{i,\text{en}})^2}{(n-1)} \quad \delta_y = \frac{\sum_{i=1}^{n} (l_{i,\text{en}} - l_{i,\text{en}})^2}{(n-1)}
\]

\(l_{i,\text{ex}}, l_{i,\text{en}}\): coordinates of the starting point of the link \(i\)

\(l_{i,\text{ey}}, l_{i,\text{en}}\): coordinates of the terminating point of link \(i\)

\(n\): total number of links in the current matching area

![Fig.7 Unsymmetrical Buffer Growing](image)
Depending on the values of $G_{dev_x}$ and $G_{dev_y}$, the position of buffer will be possibly so adjusted that a non-symmetrical buffer zone around line segments is adopted in the subsequent matching iterations (see example in figure 7-b). Such an iterative process is named as “Unsymmetrical Buffer Growing”.

Systematic geometric drifts often exist in some small local areas but seldom in large ones. This is another reason that we partition the whole matching region at first (see section 2.1). Unsymmetrical Buffer Growing as the extension of BG proves a promising approach in our test.

2.5 Conflict solution

If two or more line objects from the reference dataset correspond to a single object from the target dataset as shown in figure 8-a, they are said to be in conflict because the matching usually means a one-to-one relation. This problem can be solved by calculating the values of matching certainty (see 3.1). The line object with the largest matching certainty is regarded as the best match, whilst all the others will then be rejected (see figure 8-b).

![Fig.8](image)

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black: reference dataset</td>
<td>Red: target dataset</td>
</tr>
</tbody>
</table>

Fig.8 Conflict solution based on the Selection among the conflicted match

3 MATCHING RESULTS

3.1 Matching rate and certainty

To elucidate the performance of our matching approach, three test areas from Hessian, Germany are selected (see figure 9).
As shown in Tab.4, there are totally 4828 turning points of TeleAtlas in these three areas. 4641 points of them are correctly linked to the Basis DLM. 40 points are considered as poorly linked based on visual comparison. 147 points cannot be matched. The average matching rate reaches 96.1%; the rate of poor matching is smaller than 1.0%.

<table>
<thead>
<tr>
<th></th>
<th>Area (a)</th>
<th>Area (b)</th>
<th>Area (c)</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of line objects</td>
<td>Basis DLM</td>
<td>311</td>
<td>670</td>
<td>362</td>
</tr>
<tr>
<td></td>
<td>TeleAtlas</td>
<td>319</td>
<td>733</td>
<td>459</td>
</tr>
<tr>
<td>Number of points</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>along road lines of</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TeleAtlas</td>
<td>804</td>
<td>2413</td>
<td>1611</td>
<td>28</td>
</tr>
<tr>
<td>Number of matched</td>
<td>Correct match</td>
<td>766 (95.3%)</td>
<td>2292 (95.0%)</td>
<td>1583 (98.3)</td>
</tr>
<tr>
<td>points</td>
<td>Poor match</td>
<td>8 (1.0%)</td>
<td>16 (0.7%)</td>
<td>16 (1.0%)</td>
</tr>
<tr>
<td>Number of points with</td>
<td>30 (3.7%)</td>
<td>105(4.3%)</td>
<td>12 (0.7%)</td>
<td>147 (3.1%)</td>
</tr>
<tr>
<td>no match</td>
<td>26 seconds</td>
<td>65 seconds</td>
<td>42 seconds</td>
<td>133 seconds</td>
</tr>
</tbody>
</table>

Tab.4 A statistic overview of the matching results

3.2 Assessment of the matching quality

A few matching errors are inevitable in most matching approaches. The process to detect the errors is quite time consuming and labor intensive, because erroneous links need to be analysed one by one. Even if a match is visually correct, it is still coupled with a degree of uncertainty. In our approach, we assess the matching quality by classifying the matching results into various certainty levels. Based upon the variable “Matching_SumW”, the measurement of matching certainty is defined as follows: For conflicting match, \( \text{Matching\_Certainty} = \text{Min} \{ \text{Matching\_SumW}, 0.20 \} \); Otherwise, \( \text{Matching\_Certainty} = \text{Matching\_SumW} \).

Using our matching approach in an area of 200 km², ca.12000 objects in Basis DLM are matched to their corresponding objects in Teleatlas, among which more than 100 matching errors, i.e. falsely matched object pairs are manually detected. We conducted an uncertainty analysis on the overall matched pairs and the falsely matched pairs respectively. The distribution of “\( \text{Matching\_Certainty} \)” as well as its classification is illustrated in figure 10.
As shown in Figure 10, among the falsely matched objects, more than 90% have a “Matching_Certainty” of lower than 0.2, whilst nearly none of them reveals a “Matching_Certainty” larger than 0.75, which indicates:

(a) A wrong match can be associated with a very small “matching_certainty” (<0.2);
(b) A match with a “Matching_Certainty” of over 0.75 can be confirmed as a correct match.

According to this rule, we suggest to classify the matching results into three certainty levels (see Table 5).

<table>
<thead>
<tr>
<th>Class of the matching certainty</th>
<th>Perfect</th>
<th>good</th>
<th>possible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching_Certainty</td>
<td>[0.75, 1.00]</td>
<td>(0.20, 0.75)</td>
<td>[0.00, 0.20]</td>
</tr>
</tbody>
</table>

Among the ca.12000 matched pairs as mentioned above, there is no error detected among objects with a matching certainty ≥ 0.75. The error rate is about 0.05% among objects with a matching certainty between 0.20 and 0.75. Among the remaining objects with a matching certainty ≤ 0.20, there are about 7.0% errors.

### 3.3 Integration of postal data

The matching results, which are represented by some links between Basis DLM and TeleAtlas, can be fed into Arc GIS and distinguished by different colours according to their certainty levels. This allows the operator to easily detect and refine the matching errors, because most of the matching errors (>90%) are concealed in the certainty level “possible”. With the availability of these links, the house numbers of post addresses in this area can be automatically transferred from TeleAtlas to Basis-DLM in Arc GIS as shown in Figure 11.
4 CONCLUSION AND FUTURE WORK

In this paper we have proposed a matching approach for road networks, which works in Arc GIS 9.0 and can be applied for the matching for the linear street data as well as the transfer of information bound to one of the datasets to another. The introduction of “Unsymmetrical Buffer Growing” has added some generic nature to our approach which indeed has revealed high matching rate and certainty. Another contribution of this paper is the definition of the matching certainty as well as the classification of the matching results, which enables a more comfortable interaction.

The following elaboration work is still necessary:

- **Refining the current matching approach**

The matching performance of the proposed approach depends very much on the appropriate buffer setting. However the default setting may not be always proper to all cases. Even if a loop of enlarging the buffer size is applied in the proposed generic matching model, it is still not clear how to start from a convenient initial setting. A possible solution would be the combination of a default setting with an interactive user interface that allows the operator to visually estimate the buffer size based on the apparent matching pairs from the actual datasets. In addition, existing semantic information needs to be utilized as far as possible. Moreover, the weights of threshold parameters in our algorithms are determined empirically, therefore, they may not precisely represent the relative importance of each threshold. One possible way to solve this problem is to develop another learning module, for instance, based on the Artificial Neural Network so that the weights can be automatically determined after sufficient training cycles with examples.
Dealing with the special matching cases

Although the proposed matching approach can be applied to match the road networks and reveals high matching rate and certainty, for some special cases, e.g. looping crosses, parallel lines, etc., this approach still shows a negative performance. We attempted to improve it by adjusting some criteria or parameters in our current matching approach. However, it was not successful because the various types of special matching cases have quite different characteristics and it is almost impossible to commendably match all of them with the same criteria or algorithm. One possible way to solve this problem is to classify the special matching cases into several groups according to some rules, and for each group develop a proper matching approach separately. Thus, a desirable matching result can be expected.

ACKNOWLEDGEMENTS

The research described in this paper is funded by German Federal Agency for Cartography and Geodesy. We would like to thank Dr. Joachim Bobrich for his valuable advice and support.

REFERENCES

Zhang, M., Shi, W. & Meng, L., 2005: A generic matching algorithm for line networks of different resolutions, Workshop of ICA Commission on Generalization and Multiple Representation Computing Faculty of A Coruña University - Campus de Elviña.
DMA: An algebra for multicriteria spatial modeling

Salem Chakhar and Vincent Mousseau
Lamsade – University of Paris Dauphine,
Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, Paris, France.
E-Mail: {chakhar, mousseau}@lamsade.dauphine.fr

ABSTRACT

Map algebra is a powerful tool to implement cartographic modeling. Map algebra is now well recognized and most of its functionalities are supported by most of major geographical information systems products. Several extensions have been proposed to the original map algebra through the incorporation of new operators, the support of complex mathematical expressing and formulation, the support of other data types, the support of spatial dynamics modeling, the support of temporal dimension, and the support of visual modeling. On the other hand, spatial problems are inherently of multicriteria nature. However, neither the original map algebra nor its different extensions are able to support multicriteria aspects of spatial problems. This is due essentially to the absence of convenient operators permitting to support multicriteria modeling. The objective of this paper is to propose a new algebra especially devoted to multicriteria spatial modeling.

Keywords: Map algebra, Cartographic modeling, Multicriteria spatial modeling, GIS, Decision map.

1. INTRODUCTION

Map algebra is a powerful tool to implement cartographic modeling. It is introduced in the early 1980s by Tomlin (Tomlin, 1983) and then extended and refined in a book (Tomlin, 1990). The operands of map algebra are single-factor map layers. It uses spatial operators to generate new map layers as the result. Map algebra is originally raster data-oriented framework. However, several extensions have been proposed to the original map algebra. They concern the addition of new operators, the support of complex mathematical expressing and formulation, the support of other data types, the support of spatial dynamics modeling, the support of temporal dimension, and the support of visual modeling. Tomlin’s map algebra is now well recognized and most of its functionalities are now supported by most of major geographical information systems (e.g. Microstation raster module, Arc/Info-GRID, GRASS mapcalc, the the GRID Analyst of Integraph's Modular GIS Environment, Idrisi, etc.).

On the other hand, spatial decision problems are inherently of multicriteria nature, where several, often conflicting, evaluation criteria should be taken into account for evaluating different potential alternatives. However, neither the original map algebra nor its different extensions are able to support multicriteria aspects of spatial problems. This is due essentially to the absence of convenient operators permitting to support multicriteria modeling. The objective of this paper is to propose a new algebra, called decision map algebra (or DMA), especially devoted to multicriteria spatial modeling.
This paper goes as follows. Section 2 briefly introduces Tomlin’s map algebra. Section 3 deals with multicriteria analysis and modeling. Section 4 introduces the concept of decision map, the central ingredient of our algebra. Section 5 details the proposed algebra. Section 6 presents the ongoing implementation of DMA. Section 7 concludes the paper.

2. MAP ALGEBRA

Map algebra is a convenient framework for spatial analysis and modeling. The structure of map algebra consists of a set of map layers, primitives operations on and between map layers, and sequences of these operations (Berry, 1993). A map layer is an element of a cartographic model, defined as a collection of map layers each of which represents the spatial distribution of a particular attribute over a common study area. The elementary unit of a map layer is the location, usually represented as a cell in a grid space. Each location on a map layer is associated with a numerical value representing the value of the corresponding attribute at that location.

The map algebra operations are used to transform map layers, location by location, into new map layers in order to extract information useful to the user (Takeyama and Couclelis, 1997). Any complex manipulation of map layers is then represented as an algebraic composition of such operations. Thus, the structure of map algebra is similar to that traditional algebra in the sense that map layer are treated as if they where numbers in an equation (Berry, 1986).

The operations are classified into (Tomlin, 1990; Takeyama and Couclelis, 1997): (i) Local operations that compute a new value for every location as a function of the values of one or more existing values with that location; (ii) Zonal operations that compute a new value for each location as function of the values from a specified layer that are associated not just with that location itself but with all locations that occur within its zone on another specified layer; (iii) Incremental operations that characterize each location as an increment of one-, two- or three-dimensional cartographic form. The size and shape of these increments are inferred from the value(s) of each location relative to those of its adjacent neighbors on one or more specified layers; and (iv) Focal operations that compute each location's new value as a function of the existing values, distances, and/or directions of neighboring (but not necessarily adjacent) locations on a specified map layer.

Recent implementations of map algebra in GIS include global (per-layer) operations (Menon et al., 1992). These global operations are used, among others, for the generation of Euclidean distance and weighted cost distances maps, shortest path maps, nearest neighbor allocation maps, for the grouping of zones into connected, etc. In these extensions as in the original map algebra, statements and operations are normally expressed in an English-like syntax stimulating. Consider for instance the LocalRating operator that is used to characterize locations in terms of values from two or more layers. The following example is used to generate a new layer entitled OpenDevelopment on which Vegetation zones one, two and three (HardWoods, SoftWoods, and MixedWoods) are set to a value of six, while each location in zone zero (Open-Land) is set to that location’s value on the exiting Development layer (Tomlin 1990, p. 73):

\[
\text{OpenDevelopment} = \text{LocalRating of Vegetation with Development for 0 with 6 for 1...3}
\]
Enhanced spatial modeling often needs a sequence of basic map algebra operations. This sequence of operations is called a procedure. While map algebra is a well-organized and simple framework for cartographic modeling, it has several limitations (e.g. essentially raster-oriented framework, not adequate for dynamic modeling, does not support multicriteria spatial modeling). To overcome these limitations, several extensions have been proposed. They concern (i) the addition of new operators (e.g. Caldwell 2000); (ii) the support of complex mathematical expressing and formulation (e.g. Takeyama and Couclelis, 1997); (iii) the support of other data types (e.g. Armstrong and Densham, 1996; Lin, 1998; Corripio, 2003); (iv) the support of spatial dynamics modeling (e.g. Takeyama and Couclelis, 1997); (v) the support of temporal dimension (e.g. Mennis and Viger, 2005a, 2005b); (vi) the support of visual modeling (e.g. Egnhofer and Bruns, 1994; Pullar, 2004). However, all these extensions are not sufficient to support multicriteria spatial modeling.

3. MULTICRITERIA ANALYSIS AND MODELING

Multicriteria analysis (see, for e.g., Roy, 1996; Belton and Stewart, 2002), is a family of OR/MS tools that have experienced very successful applications in different domains since the 1960s. It has been advised in spatial context to overcome the limitations of GIS in spatial analysis and modeling. Indeed, GIS is a powerful tool of acquisition, management and analysis of spatially-referenced data but it is a limited tool in analysis and modeling as it is remarked by several authors (e.g. Carver, 1991; Laaribi, 2000; Malczewski, 1999; Chakhar and Martel, 2003, 2004). This is due essentially to its lack in more powerful analytical tools enabling it to deal with spatial problems, where several parties, often with conflicting objectives, are involved in the decision-making process and different, often contradictory, evaluation criteria should be considered. Naturally, multicriteria analysis is the most adequate tool to fill this gap.

It is generally assumed in multicriteria analysis that the decision maker (DM) has to choose among several possibilities, called alternatives. The set of alternatives, denoted \( A \), is the collection of all alternatives. Selecting an alternative among this set depends on many characteristics, often contradictory, called criteria. Accordingly, the decision maker will generally have to be content with a compromising solution.

The multicriteria problems are commonly categorized as continuous or discrete, depending on the domain of alternatives (Zanakis et al., 1998). Hwang and Yoon (1981) classify them as (i) multiple attribute decision-making (MADM), and (ii) multiple objective decision-making (MODM). According to Zanakis et al. (1998), the former deals with discrete, usually limited, number of pre-specified alternatives. The latter deals with variable decision values to be determined in a continuous or integer domain of infinite or large number of choices. In the rest of this paper, we focalize only on MADM methods.

The general schema of MADM methods is shown in Figure 1. The first requirement of nearly all MADM techniques is a performance table containing the evaluations or criteria scores of a set of alternatives on the basis of a set of criteria. Criteria are factors on which alternatives are evaluated and compared. More formally, a criterion is a function \( g_j \), defined on \( A \), taking its values in a totally ordered set, and representing the DM's preferences according to some points of view (Vincze, 1992). The evaluation of an alternative \( a \) according to criterion \( g_j \) is written \( g_j(a) \). Typically, in most of multicriteria problems, the DM considers that one criterion is more important than another. This relative importance is usually expressed in the form of weights, denoted \( w_j \), which are assigned to different criteria. In addition to weights, criteria are often associated with different thresholds:
The **indifference threshold**, generally denoted $q$, represents the largest difference preserving an indifference between two alternatives $a$ and $b$ in respect to a criterion $g_j$.

The **preference threshold**, generally denoted $p$, represents the smallest difference compatible with a preference in favor of $a$ in respect to criterion $g_j$.

---

**Figure 1.** The MADM general model

We remark that weights and thresholds are often called preference parameters. The next step consists in the aggregation of the different criteria scores using a specific aggregation procedure and taking into account the decision maker preferences. The aggregation of criteria scores permits the decision maker to make comparison between the different alternatives on the basis of these scores. Aggregation procedures are somehow the identities of the multicriteria methods. In MADM, they are usually categorized into two great families: (i) utility function-based family, and (ii) outranking relation-based family (see Vincke, 1992).

The first family is essentially of Anglo-Saxon inspiration. Its basic principle is that the DM looks to maximize an utility function $U(x) = U(g_1(x), g_2(x), \ldots, g_m(x))$ that aggregate the partial evaluations (i.e. in respect to each criterion) of each alternative into a global one. It is important to mention that this family does not recognize the incomparability situations (i.e. the DM can compare any two alternatives) and that the indifference is transitive (i.e. if an alternative $a$ is equivalent to another alternative $b$ and $b$ is equivalent to a third alternative $c$; than alternative $a$ is necessarily equivalent to $c$). The weighted sum is an example of this family. The second family, which recognizes the incomparability situations and does not exclude the intransitivity of preference relations, is usually considered as of European inspiration. In contrast with the first family, here the aggregation methods are said to be partial. Indeed, criteria are aggregated into partial binary relation $S$, such that $a S b$ means that “$a$ is at least as good as $b$”. The binary relation $S$ is called outranking relation. The most known method in this family is ELECTRE (see Roy, 1996).

The uncertainty and the fuzziness generally associated with any decision situation require a sensitivity analysis enabling the decision maker to test the consistency of a given decision or its variation in response to any modification in the input data and/or in the decision maker preferences.

The aim of any decision model is to help the decision maker take decisions. The final recommendation in multicriteria analysis may take different forms, according to the manner in which a problem is stated. Roy (1996) identifies three types of results corresponding to three main ways for stating a problem: (i) choice: selecting a
restricted set of alternatives; (ii) sorting: assigning alternatives to different pre-defined categories; and (iii) ranking: classifying alternatives from best to worst with eventually equal positions.

Finally, we mention that in spatial context, alternatives and evaluation criteria are associated with geographical entities and relationships between entities and therefore can be represented in the form of maps. Alternatives and criteria maps are generated using standard map algebra operations. In GIS-based applications, criteria generation process is often modeled in terms of flowcharts, which are intuitive and simple modeling environment, especially for users with limited knowledge.

4. CONCEPT OF DECISION MAP

In this section, we introduce the concept of decision map, which is the central ingredient of our algebra. First, we note that a full description of the concept of decision map and its generation process and uses are detailed in Chakhar et al (2005). Physically, a decision-map is a special kind of the map layer were the decision space is treated as a discrete surface composed of a finite number of polygonal homogenous spatial units obtained by applying a multicriteria sorting method. More formally, a decision map \( M \) is defined as the set \( \{ (u, f(u)) : u \in U \} \), where \( U \) is a set of homogenous spatial units and \( f \) is a function defined as follows:

\[
f: U \rightarrow E \\
u \rightarrow f(u) = \Phi [g_1(u), \ldots, g_m(u)],
\]

where \( E \) is an ordinal (or cardinal) scale, \( \Phi \) is a multicriteria sorting model and \( g_j(u) \) is the performance of spatial unit \( u \) in respect to criterion \( g_j \). Accordingly, a decision map summarizes the preferential information of the decision maker relatively to a set of conflicting evaluation criteria into an ordinal or cardinal information.

The first step for the construction of the decision map consists in the generation of criteria maps. Each criterion map represents a specific theme and is composed of set of homogenous spatial units; each of which is associated with one evaluation, \( g_j(u) \). Then, these criteria maps need to be superposed to obtain an intermediate map, which is composed of a new set of spatial units resulting from the intersection of the boundaries of the spatial units of the different criteria maps. Each spatial unit is characterized with a vector of \( m \) evaluation relative to the \( m \) criteria maps. Formally, to each spatial unit \( u \), we associate the vector \( [g_1(u), g_2(u), \ldots, g_m(u)] \).

Description of territory as a set of units is not new in land management (Joerin and Musy, 2000). In classical cartographic modeling, these units are often defined through census tracts or administrative and political boundaries. In this paper as in Joerin and Musy (2000), the initial subdivision of territory into homogenous units (Joerin uses the term zone instead of unit) should respect the spatial natural (forest, body of water) and human (highways, parks, buildings) boundaries.

To generate a final decision map, we should first use an aggregation mechanism to aggregate the vector associated with each spatial unit \( u \) in the intermediate map into one global evaluation. Mathematically, we write:

\[
g(u) = \Phi [g_j(u)]_{j \in F}.
\]

\( F \) is the criteria family and \( \Phi \) is defined as follows:

\[
\Phi: E^m \rightarrow E \\
[g_1(u), g_2(u), \ldots, g_m(u)] \rightarrow g(u)
\]
The aggregation mechanism $\Phi$ permits to assign each unit to one or several predefined categories on the scale $E$. In this paper, the multicriteria sorting model used is ELECTRE TRI (see Appendix B for an overview and Yu (1992) for a complete description of ELECTRE TRI method).

The major merit of a decision map is that it permits to use outranking-based aggregation models in spatial modeling problems. In fact, most of multicriteria analysis based-GIS use methods based on utility function-like aggregation methods (e.g. Carver, 1991; Jankowski, 1995). Theses aggregation models still dominate today and only few works (e.g. Martin et al., 2000; Joerin and Musy, 2000, 2001) use outranking-based aggregation models (see Roy, 1996). However, these last ones are more suitable in spatial context since they permit “to consider both objective and subjective criteria and require fewer amount of information from the decision maker” (Malczewski, 1999). In addition, they do not impose the transitivity of indifference and tolerate the incomparability situations. But the major (technical) drawback of outranking methods is that they are not suitable for problems implying a great or infinite number of alternatives since they require pairwise comparison across all criteria (Chakhar et al., 2005).

Subdividing the study area into homogenous spatial units permits to reduce significantly the number of potential alternatives to be evaluated and leads to “a manageable set of alternatives” (Hall et al, 1992; Wang, 1994; Joerin and Musy, 2001) and outranking methods can be applied quite easily. Indeed, in spatial multicriteria modeling, potential alternatives are represented through one of three atomic spatial entities, namely point, line or polygon (Malczewski, 1999; Chakhar and Mousseau 2004). Therefore, in a facility location problem, potential alternatives take the form of points representing different potential sites; in a linear infrastructure planning problem (e.g. highway construction), potential alternatives take the form of lines representing different possible routes; and in the problem of identification and planning of a new industrial zone, potential alternatives are assimilated to a set of polygons representing different candidate zones. In raster-based GIS, these three alternatives are defined as a single pixel, collection of linearly adjacent pixels, and a set of contiguous pixels, respectively.

By using the decision map concept, punctual, linear and polygonal decision alternatives can be defined, respectively, as an individual homogenous spatial unit, a collection of linear adjacent homogenous spatial units, and a collection of contiguous homogenous spatial units, respectively. These ways of modeling are illustrated in Table 1.

It is important to remark that in many real-world applications, one may be called to represent alternatives with a combination of two or more atomic entities (Malczewski, 1999; Chakhar and Mousseau 2003, 2004). In schools partitioning problem, for instance, decision alternatives can be assimilated to a combination of points and polygons where points represent schools and polygons represent zones to serve. A set of "point-point" composed alternatives may represent potential paths in a shortest path identification problem. In this paper, we consider only atomic decision alternatives. We just mention that in Chakhar and Mousseau (2004) we have proposed several other examples of spatial problems implying composed alternatives. We have also introduced a solution that may used to handle some of these problems.
<table>
<thead>
<tr>
<th>Problem</th>
<th>Representation in GIS</th>
<th>Proposed modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location problems</td>
<td>Point</td>
<td>one spatial unit</td>
</tr>
<tr>
<td>Linear infrastructure problems (e.g. highways)</td>
<td>Line</td>
<td>a series of linearly adjacent spatial units</td>
</tr>
<tr>
<td>Land used and management problems</td>
<td>Polygon</td>
<td>one or several contiguous spatial units</td>
</tr>
</tbody>
</table>

Table 1. Modeling spatial decision alternatives

To conclude this section, we compare our decision map concept to maps generated through conventional cartographic modeling:

- Cartographic modeling is essentially an automatic procedure with no or less (generally a priori) interaction with the decision maker. Our decision map generation process (see Chakhar et al 2005) is largely controlled by the decision maker;

- Maps produced through cartographic modeling are essentially presentation-oriented ones and they are roughly used for an effective spatial decision-aid activity. Our decision maps are decision-oriented ones. In addition, decision map is a generic tool that may serve to the generation of potential alternatives and can be extended to support collaborative and communicative spatial decision making (as it is detailed in Chakhar et al., 2005);

- Decision map permits to explicitly represent spatial preferences of the decision maker. In classic cartography modeling, preferences of the decision maker are often reduced to a tabular representation, often with no explicit relation to their spatial locations;

- Aggregation is performed in early steps in the cartographic modeling process, which may lead to a substantially lost of the preferential information. In our approach, aggregation is performed in latter steps in the generation process;

- Aggregation in our approach is based on outranking relations, which we think more appropriate to deal with spatial decision making. Especially, this permits the integration of both qualitative and qualitative criteria in problem modeling.

5. PROPOSED DECISION MAP ALGEBRA

To support multicriteria modeling, we introduce in this section a new algebra that we call decision map algebra (or DMA). It is important to note that DMA does not substitute Tomlin’s map algebra: DMA is an extension of Tomlin’s map algebra that is especially devoted to multicriteria spatial modeling. Further, we mention that in practice we need to combine both of them for a complete and efficient modeling. Moreover, several operators of DMA use the ones of map algebra. DMA contains three sorts of operands: (i) geographic objects; (ii) geographic maps; and (iii) some other operands required to multicriteria modeling. In addition to the standard operators of map algebra, DMA contains several new ones devoted to multicriteria modeling.
5.1 Primitives and definitions
This section defines the terms used in the specification of DMA. DMA contains several data types. The relationships among these data types are depicted in Figure 2. In this figure, the symbols ‘s’ and ‘agg’ signify the specialization/generalization and the aggregation relationships, often used in object oriented modeling; the double brakes symbol ‘{ }’ means ‘a collection of’; the ‘xor’ is the XOR binary operator. Geographic objects and maps are represented by oval or rectangular shapes with a solid line. The other shapes (with a dashed line) represent descriptive attributes and multicriteria concepts that are introduced for the purpose of DMA.

The basic data type in DMA (and also in most of map algebra-like languages) is map-layer. A map-layer is the most elementary data that contains the map image and other documentary items for the map-layer including the global geographic reference system, rSystem; the map scale, mScale; etc. There are two types of map-layer data type: raster and vector. The map image of a raster map-layer is composed of pixels. Each element of a aster map-layer is called gPixel. The map image of a vector map-layer is a collection of three types of geographical objects:

- **gPoint**: is an individual, one-dimensional repressing a punctual entity in real-world.
- **gLine**: is a two-dimensional object representing a linear entity in real-world.
- **gPolygon**: is a three-dimensional object representing a polygonal entity in real-world.

In DMA, a map-layer is either of raster or vector nature. This is ensured with the ‘xor’ symbol in Figure 2. For the purpose of DMA, a collection of gPixel, gPoint, gLine and gPolygon are denoted xSet, pSet, lSet and ySet, respectively. A gPixel can be associated with one value which may be any arbitrary data type representing a natural or an artificial attribute. A gPoint, gLine or gPolygon can be associated with several values representing a set of descriptive attributes that apply to these objects as a whole. The gPixel elements of a raster map-layer are fully identified with their XY coordinates on the map image. The gPoint, gLine and gPolygon associated with a vector-based map-layer need to be uniquely identified.

We define three new subclasses of map-layer data type: alternative-map, criterion-map and decision-map (see Figure 2). An alternative-map is a special kind of map-layer which, in addition to the general information associated with any map-layer, contains a collection of specific objects called alternatives. It is important to mention that alternatives are normally generated through decision-maps as explained in section 4. But the alternative-map data type is included in DMA to deal with spatial problems for which the use of a decision-map is difficult or not possible.

There are basically three types of alternatives: pAlter, lAlter or yAlter representing punctual, linear or polygonal decision alternatives, respectively. A collection of pAlter, lAlter and yAlter are denoted pAlters, lAlters, and yAlters, respectively. When it is necessary, the generic terms anAlter and sAlters will be used to denote an alternative and a set of alternatives, regardless to their nature. Each alternative is characterized with a vector of values relative to a set of criteria or attributes and applies to these objects as a whole. It is important to mention that for a given spatial decision problem, each alternative-map contains only one kind of alternatives; each of which is uniquely identified.

A criterion-map is a specific, mono-valued, map-layer where each spatial object (i.e. gPixel, gPoint, gLine or gPolygon) is characterized by one value representing the evaluation of this element in
respect to a given criterion. A criterion-map is different from a simple map-layer in the sense that it models some subjective information. That is, two persons may give different values for the same spatial object, along with their study perspectives.

Figure 2. Data types relationships

In multicriteria analysis, each criterion is often associated with a weight, a direction of optimization and several preference parameters (see section 3). To take into account these information, we associate to each criterion-map the following data types:

- **cWeight**: An importance degree reflecting the power of the criterion during the comparison of alternatives.
- **cDirection**: The optimization direction of the criterion which may be maximization or minimization.
- **qThreshold** and **pThreshold**: corresponding to the indifference and preference thresholds introduced in section 3;
- **vThreshold**: represents the smallest difference between the performance of two alternatives incompatible with the outranking assertion as explained in Appendix B.

For coherence reasons, we need to have: $0 \leq qThreshold \leq pThreshold \leq vThreshold$. A collection of evaluation criterion-maps are called **cFamily** (for criteria family) in DMA.

As mentioned earlier, a decision-map is a special kind of map-layer where the decision space is treated as a discrete surface composed of a finite number of polygonal homogenous spatial units. Each element of a decision-map is of type **sUnit** (for spatial unit). A collection of spatial units is denoted by **sSet**.
As explained earlier, to operationalize a decision-map, we need to assign all of its spatial units to a set of predefined categories through the multicriteria sorting method ELECTRE TRI. We denote by $u\text{Class}$ the class to which a spatial unit is assigned.

To support multicriteria modeling, we add a new data type, called $sd\text{-model}$ (for spatial decision model). The $sd\text{-model}$ is an aggregation of one decision-maps (or an alternative-map) and at least two criterion-maps data types.

Finally, we mention that several other data types (as decision table, aggregation operator, preference structure and choice function) are needed to formalize our algebra. These additional data types are not included in Figure 2 but they will be introduced progressively hereafter.

### 5.2 Formal specification of DMA

To specify our DMA, we adopt the algebraic specification method of Guttag and Horning (1978). The algebraic specification methods are mainly used in software engineering to describe the behavior of complex systems. An algebraic specification consists of three parts (Guttag and Horning; 1978; Dorenbeck and Egenhofer, 1991): (i) a set of sets including the data type to define and the types needed to define its properties; (ii) a set of operations defined on the operands. Each operation is defined by its name, the Cartesian product of the inputs sorts and the sort of the result; and (iii) a set of axioms (or equations) that describe the behavior of operations.

In the following, we discuss the formal specification of some data types. The specifications of the other data types are provided in Appendix A.

#### 5.2.1 Specification of $s\text{Unit}$ data type

The most elementary data type is the spatial unit, denoted $s\text{Unit}$. The formal specification of $s\text{Unit}$ data type in shown in Figure 3. The first operator associated with $s\text{Unit}$ is ASSIGN which permits to set the performance of a $s\text{Unit}$ in respect to a given criterion provided as parameter. The ASSIGN is a hidden function (Chan and White, 1987), i.e. it is not part of the algebra and serves its purpose in the specification only. The operator ADJACENT tests if two spatial units are adjacent or not. GET-EDGES and GET-VERTICES permit to extract, respectively, the set of edges and vertices for a given spatial unit. The specifications of these three spatial manipulation operators are similar to the ones associated with $g\text{Polygon}$ data type provided in Appendix A. The SCORE operator returns the performance of a spatial unit in respect to a given criterion. This operator is quite straightforward and its specification is not detailed in Figure 3.

Data type $s\text{Unit}$ is associated with a set of operators devoted to implement ELECTRE TRI model. The operators $PCONCORDANCE$, $CONCORDANCE$, $PDISCORDANCE$ and $DISCORDANCE$ implement Equations 1, 2, 3 and 4 in Appendix B, respectively. The $PCONCORDANCE$ and $PDISCORDANCE$ operators take in input two values and a criterion and generate a real that indicates the partial concordance and discordance indices, respectively. The two values correspond to the scores of the spatial unit and the profile to be compared in respective to the criterion under consideration.

To implement the $PCONCORDANCE$ and $PDISCORDANCE$ operators, we need to define the concept of decision table, denoted $d\text{Table}$. A decision table is a method to specify formally the behavior of operations, particularly those which can be described by a series of rules. It consists of two parts (Dorenbeck and Egenhofer, 1991): (i) a set of conditions which have to be satisfied simultaneously; and (ii) the corresponding actions to be taken upon
conditions. Decision tables are most naturally presented in the form of a table with the set of conditions being put into the upper half of the table and the corresponding set of actions underneath. The specification of \( dTable \) is provided in Appendix A. Each \( dTable \) has two operators: \( \text{CRETATE} \) and \( \text{ACTION} \). The specification of the \( \text{CRETATE} \) operator is \( \text{DEFFERED} \) (Meyer, 1988; Dorenbeck and Egenhofer, 1991), because it depends upon the particular context. The general structure of decision tables associated with \( \text{PCONCORDANCE} \) and \( \text{PDISCORDANCE} \) are shown in Table 2 and Table 3, respectively.

<table>
<thead>
<tr>
<th>v1 - v2</th>
<th>≥ p</th>
<th>≤ q</th>
<th>&lt; p ∧ q</th>
<th>ACTION</th>
<th>0</th>
<th>1</th>
<th>( [p - v1 + v2]/[p - q] )</th>
</tr>
</thead>
</table>

**Table 2.** Decision table associated with \( \text{PCONCORDANCE} \) operator

<table>
<thead>
<tr>
<th>v2 - v1</th>
<th>≥ p</th>
<th>≤ v</th>
<th>&lt; v ∧ p</th>
<th>ACTION</th>
<th>0</th>
<th>1</th>
<th>( [v - v1 + v2]/[v - p] )</th>
</tr>
</thead>
</table>

**Table 3.** Decision table associated with \( \text{PDISCORDANCE} \) operator

---

**Type:** sUnit  
**set:** map-layer, aCriterion, cFamily, sUnit, aProfile, pSet, lSet, dTable, real, value, Boolean

**Syntax:**
- ASSIGN \( sUnit \times aCriterion \times value \rightarrow sUnit \)
- ADJACENT \( sUnit \times sUnit \rightarrow \text{boolean} \)
- GET-VERTECES \( sUnit \rightarrow pSet \)
- GET-EDGES \( sUnit \rightarrow lSet \)
- SCORE \( sUnit \times aCriterion \rightarrow \text{real} \)
- PCONCORDANCE \( value \times value \times aCriterion \rightarrow \text{real} \)
- CONCORDANCE \( sUnit \times aProfile \times cFamily \rightarrow \text{real} \)
- PDISCORDANCE \( value \times value \times aCriterion \rightarrow \text{real} \)
- DISCORDANCE \( sUnit \times aProfile \times cFamily \rightarrow \text{real} \)
- SIGMA \( sUnit \times aProfile \times cFamily \rightarrow \text{real} \)
- OUTRANK \( sUnit \times aProfile \times cFamily \times value \rightarrow \text{boolean} \)

**Axioms:**

\( u: sUnit; h:aProfile; f:cFamily; g:aCriterion; t1,t2,t3:dTable; v:value \)

\[
\text{PCONCORDANCE}(u,h,g) = t1.\text{action} (\text{SCORE}(u,g),\text{SCORE}(h,g),g)
\]

\[
\text{CONCORDANCE}(u,h,f) = [\Sigma_{g \in \varepsilon} \text{PDISCORDANCE}(u,h,g) \times g.cWeight] / \]
\[
\sum_{g \in f} g \cdot cWeight
\]

\[
PDISCORDANCE(u,h,g) = t2.\text{action}(\text{SCORE}(u,g), \text{SCORE}(h,g), g)
\]

\[
\text{DISCORDANCE}(u,h,f) = \Pi_{g \in f} (\text{PDISCORDANCE}(u,h,g) > \text{CONORDANCE}(u,h,g)) / (1 - \text{CONORDANCE}(u,h,g))
\]

\[
\text{SIGMA}(u,h,f) = \text{CONCORDANCE}(u,h,f) \times \text{DISCORDANCE}(u,h,f)
\]

\[
\text{OUTRANK}(u,h,f,v) = t3.\text{action}(u,h,\text{SIGMA}(u,h,f), \text{SIGMA}(h,u,f), v)
\]

---

**Figure 3.** Formal specification of sUnit data type

In Tables 2 and 3, the parameters \( q, p \) and \( v \) should be mapped to \( q\text{Threshold}, p\text{Threshold}, \) and \( v\text{Threshold} \) attributes; and \( v1 \) and \( v2 \) correspond to \( \text{SCORE}(h,g) \) and \( \text{SCORE}(u,g) \), respectively. All of them are provided as parameters for \( \text{ACTION} \) operator.

The \( \text{CONCORDANCE} \) and \( \text{DISCORDANCE} \) operators take in input a spatial unit, a profile and a family of criteria; and generates a real value in \([0,1]\) corresponding to the global concordance and discordance indices. The \( cWeight \) attribute used in the specification of \( \text{CONCORDANCE} \) operator correspond to criterion weight (see Figure 2). The \( \text{SIGMA} \) operator permits to compute the credibility indices as in Equation 5 in Appendix B.

The operator \( \text{OUTRANK} \) permits to get the preference situation. As mentioned in Appendix B, there are four disjunctive possible situations that hold when comparing a spatial unit \( u \) to a profile \( h \): \( aIh, aPh, hPa \) or \( aRh \). The operator \( \text{OUTRANK} \) uses the concept of decision table. The decision table associated with \( \text{OUTRANK} \) operator is shown in Table 4. Note that \( u \) and \( h \) in Table 4 correspond to the spatial unit and the profile to be compared, respectively; and the values \( v1 \) and \( v2 \) in Table 4 correspond to \( \text{SIGMA}(u,h,f) \) and \( \text{SIGMA}(h,u,f) \), respectively. It is easy to see the four preference situations mentioned in Appendix B in the underneath part of Table 4. In addition to \( u, h, \text{SIGMA}(u,h,f) \) and \( \text{SIGMA}(u,h) \), the \( \text{ACTION} \) operator takes the value of the cutting level \( \lambda \) (see Appendix B), and returns the corresponding decision.

| \( v1 \) | \( \geq \lambda \) | \( \geq \lambda \) | \( < \lambda \) | \( < \lambda \) |
| \( v2 \) | \( \geq \lambda \) | \( < \lambda \) | \( \geq \lambda \) | \( < \lambda \) |
| **ACTION** | \( uIh \) | \( uPh \) | \( hPu \) | \( uRh \) |

**Table 4.** Decision table associated with \( \text{OUTRANK} \) operator

### 5.2.2 Specification of decision-map data type

Figure 4 specifies the decision-map data type. As it is shown in this figure, the syntax part of decision-map data type contains four operators. The \( \text{MAKE} \) operator creates a decision map as the intersection of a set of criterion-maps. The result of this intersection is an initial decision-map. Each spatial unit of this initial decision-map is associated with a set of values relative to different criteria.
**Type:** decision-map

**set:** map-layer, criterion-map, sUnit, cFamily, aOperator, aProfile, sProfiles

**Syntax:**

MAKE  criterion-map x...x criterion-map  \rightarrow  decision-map

CLASSIFY  decision-map x cFamily x sProfiles  \rightarrow  decision-map

GROUP  decision-map x cFamily x aOperator  \rightarrow  decision-map

MERGE  decision-map x cFamily x sUnit x...x sUnit x aOperator  \rightarrow  decision-map

**Axioms:**

d: decision-map; c1,...,cm,g: criterion-map; f;cFamily; v: value; h:aProfile; b:sProfiles

MAKE(c1,...,cm)  
=INTERSECT(c1,...,cm)

CLASSIFY(d,b,f)  
=  \forall (u) (u \in d)  
\forall (h) (u \in b)  
if OUTRANK(u,h,f,cLevel) then u.uClass \leftarrow h+1

MERGE(d,u1,u2,op,f)  
=\text{u.make}(d,\{\text{GET-VERTECES}(u1) \cup \text{GET-VERTECES}(u2)\} \setminus  
\{\text{GET-VERTECES}(u1) \cap \text{GET-VERTECES}(u2)\})  
\forall (g)(gef)\{\text{ASSIGN}(u,g,op.combine(\text{SCORE}(u1,g),\text{SCORE}(u2,g)))\}

GROUP(d,op)  
=  \forall (u1)\forall (u2)(u1 \in d)(u2 \in d)  
\land (u1 \neq u2)  
[if ADJACENT(u1,u2)  
\land u1.uClass = u2.uClass  \text{ then MERGE}(d,u1,u2,op,f)]

---

**Figure 4.** Formal specification of decision-map data type

The operator CLASSIFY is the implementation of the pessimistic assignment procedure of ELECTRE TRI (see Appendix B). It permits to assign each spatial unit to a predefined set of categories defined in terms of their profiles. As it is shown in Appendix B, the pessimistic assignment procedure with ELECTRE TRI compares each alternative \( u \) (here alternatives are the spatial units) to all profiles starting from the best to the worst. The spatial unit \( u \) is assigned to first class for which \( u \) outranks its lower limit.

The MERGE operator simply groups two or more adjacent spatial units. It uses the MAKE operator, inherited from gPolygon data type, to create a new spatial unit. The evaluations of the new spatial unit in respect to all criteria are obtained by aggregating, using the aOperator, the initial evaluations of original spatial units. The specification of the MERGE operator is shown for two spatial units. The generalization to more than two spatial units is straightforward.
The specification of the aOperator is defined in Appendix A. A aOperator must provide operation to combine a series of values. Since a large set of aggregation operators are available, the specification of its CRESTATE operator is DEFFERED (Meyer, 1988; Dorenbeck and Egenhofer, 1991), because it depends upon the particular aggregation operator.

The GROUP function takes a decision-map and an aggregation operator, aOperator; and generates a new decision-map by merging all adjacent spatial units that are assigned to the same class. It is simply the generalization of the MERGE operator to the entire decision-map.

5.2.3 Specification of criterion-map data type
Data type criterion-map specified in Figure 5 has two operators. The operator MAKE takes as input a predefined map algebra procedure, that is, a sequence of map algebra operations that takes in input one or more map-layers; and generates a criterion map.

The operator SET permits to set a preference parameter. The aParameter parameter may take the values of weight, direction, indifference, preference, or veto.

Within the multicriteria methods based on utility-function aggregation operator, the term criterion is a generic one that includes objectives and attributes. An objective is a statement about the desired future which is made operational by assigning to it one or more attributes describing a geographical entity or the relationship between geographical entities (Malczewski, 1999). The specification of criterion-map data type presented above does not include these aspects of criteria modeling but they will be included in the future. This will enhance the applicability of DMA.

<table>
<thead>
<tr>
<th>Type: criterion-map</th>
</tr>
</thead>
<tbody>
<tr>
<td>set: map-layer, criterion-map, cWeight, cDirection, qThreshold, pThreshold, vThreshold, aParameter, value, procedure</td>
</tr>
</tbody>
</table>

**Syntax:**

MAKE  procedure → criterion-map
SET  criterion-map x aParameter x value → criterion-map

**Axioms:**

\[ \text{SET}(c,a,v) = c.a \leftarrow v \]

**Figure 5.** Formal specification of criterion-map data type

5.2.4 Specification of sd-model data type
As mentioned earlier, the sd-model data type is an aggregation of a decision-map and at least two criterion-maps. It is important to remember that the concept of decision-map may not apply for some
spatial problems. In this case, we may use an alternative-map. In the specification below, we suppose that a decision-map is in use. The specification with an alternative-map is not included in this paper.

The specification of sd-model data type is provided in Figure 6. It contains different multicriteria modeling operators. The P-ALTERNATIVE operator permits to generate a set of punctual alternatives. It takes a decision-map and returns a set of spatial units verifying the constraints ensured by the expression <aCriterion><bOperator><value>; where aCriterion represents an evaluation criterion and bOperator is a binary operator. The L-ALTERNATIVE operator permits to generate a set of linear alternatives. It takes a decision-map and two spatial units representing the start and the end points. It generates a set of corridors relating start and end spatial units.

In multicriteria modeling, we may need to eliminate from consideration some alternatives that present some undesirable aspects. These restrictions imposed, by the nature or by human beings, on alternatives are called constraints. Constraints may also be called admissibility criteria since they represent criteria that must be fully verified by any alternative; true criteria represent conditions that need to be satisfied at maximum (Laaribi, 2000). The constraints dichotomize a set of alternatives under consideration into two categories: acceptable (or feasible) and unacceptable (or unfeasible). The CONSTRAINT operator associated with the sd-model implements the concept of constraint. It takes a set of alternatives and a constraint similar to the one used with P-ALTERNATIVE operator; and returns all the alternatives verifying the constrain. When several constraints are required, they may be modeled sequentially; each of which takes the result of the previous constraint as an input.

The alternatives generated by L-ALTERNATIVE operator are composed of a set of spatial units, each of which is associated with a set of partial evaluations. The EVALUATE takes a linear alternative, a family of criteria and an aggregation operator, aOperator, and returns the new partial evaluations that apply to the alternative as a whole. We note that the spatial units $u_1, \ldots, u_r$ in the specification of EVALUATE operator are the individual spatial units composing the alternative. Intuitively, the EVALUATE operator is not needed for punctual alternatives (generated by the P-ALTERNATIVE operator).

The SCORE operator is a hidden one. It takes an alternative and a criterion and returns the partial performance of that alternative in respect to the criterion. The P-VECTOR returns the performances of an alternative in respect to a family of criteria. The result is stored in a conventional ADT list. This function uses the INSERT operator associated with the ADT list (called aList in Appendix A).

The PAYOFF operator takes a set of alternatives and a family of criteria and returns the performance matrix. In the formal specification of this operator, the result matrix is defined as a list of performance vectors, i.e., each line is simply the performance vector of an alternative.

The DOMINATE is an implementation of the dominance relation $\Delta$ used in multicriteria modeling. The dominance relation is defined for two alternatives $a$ and $b$; and a family of criteria $F$ as follows:

$$a \Delta b \iff g_j(a) \geq g_j(b); j \in F,$$

with at least one strict inequality. Within DMA, DOMINATE operator takes a set of alternatives of the same type and returns all the non-dominated ones in respect to a family of criteria, cFamily.
In real-world problems, criteria scores can be quantitative or qualitative and can be expressed according to different measurement scales (ordinal, interval, and ratio). However, several multicriteria problems require that all criteria evaluations are expressed on the same scale. The NORMALIZE operator is introduced to re-scale, when it is necessary, the different criteria scores between 0 and 1. It takes in input a set of alternatives of the same type, a family of criteria and a normalization procedure, denoted nProcedure. The specification of a nProcedure (see Appendix A) is similar to that of aOperator. Since different methods of normalization are available (see Table 5 for some examples), the CREATATE operator of a nProcedure is a DEFFERED one.

<table>
<thead>
<tr>
<th>#</th>
<th>Rescaled value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$g(a) / \max_i g(a_i)$</td>
</tr>
<tr>
<td>2</td>
<td>$[g(a) - \min_i g(a_i)] / [\max_i g(a_i) - \min_i g(a_i)]$</td>
</tr>
<tr>
<td>3</td>
<td>$g(a) / \sum_i g(a_i)$</td>
</tr>
<tr>
<td>4</td>
<td>$g(a) / \sqrt{[\sum_i g(a_i)^2]}$</td>
</tr>
</tbody>
</table>

Table 5. Some normalization techniques ($g(a)$ is the initial evaluation)

The operator AGGREGATE uses a multicriteria aggregation procedure to aggregate all the partial evaluations into a global one. The aggregation procedures are denoted aProcedure (see Appendix A for the specification of aProcedure data type). The CHOICE, SORT and RANGE operators correspond to the three types of recommendations in multicriteria analysis that have mentioned in Section 3. The definition of these three operators needs the introduction of a new concept: preference structure. A preference structure pStructure permits to the decision maker to articulate his/her preferences when comparing two alternatives. In multicriteria analysis, a pStructure is often operationalized through a criteria function. Thus, a pStructure may be associated with several thresholds. Here, we adopt the most general way that considers the presence of two preference thresholds: an indifference threshold, called qThreshold, and a preference threshold, called pThreshold. A such pStructure permits to model three situations when comparing two alternatives a and b:

- $aPb \iff g(a) > g(b) + pThreshold$
- $aQb \iff g(b) + pThreshold \geq g(a) > g(g) + qThreshold$
- $aIb \iff g(b) + Threshold \geq g(a) \text{ and } g(a) + qThreshold \geq g(b)$

The $P$, $Q$ and $I$ symbols are the binary relations of strict preference, weak preference and indifference, often used in preference modeling. Any pStructure can be fully characterized in terms of its characteristic function. For example, we may associate to the preference structure above a characteristic function $S$ defined as:

$$aSb \iff aPb \lor aQb \lor aIb$$
The formal specification of pStructure data type is provided in Appendix A. It includes three operators, P, Q, and I, implementing the three situations mentioned above; and an operator, S, implementing the characteristic function.

The CHOICE operator may be implemented in terms of a choice function defined on a pStructure. A choice function \( C \) is a function that associates to a set \( B \) a subset \( C(B) \) of \( B \). For example, we may associate to pStructure defined above, the following choice function:

\[
C(B) = \{ a \in B : aSb, \forall b \in B \}.
\]

Note that other functions may also apply as for example:

\[
C(B) = \{ a \in B : \neg \exists b \in B : bPa \}.
\]

As it is shown in Figure 6, the CHOICE operator is defined in terms of the above choice function. It takes in input a set of alternatives, a preference structure and a family of criteria; and returns the alternatives that verify the choice function \( S \) associated with the preference structure.

The RANGE and SORT operators are defined by using the CHOICE operator. The RANGE operator establishes a partial pre-order on the set of alternatives \( A \). A pre-order consists of (i) a set of equivalence classes and (ii) an order relation upon these classes. Thus, it can be implanted as a series of CHOICE operators; each is of the following form:

\[
C(B) = \{ a \in B : aPb \lor aIb, \forall b \in B \}.
\]

In each step \( i \), RANGE returns the most preferred alternatives from the set \( B^i \) where \( B^i = B \setminus B^{i-1} \); and \( B^1 = B \). In the first step, the choice function is applied to all the alternatives in the set \( B \). The next steps use the set generated in the previous step minus the selected alternatives as input. As it is shown in Figure 6, the result of the RANGE operator is a list of ordered set of equivalence classes. The INSERT and GET operators are those of aList data type. They are used to insert and to get the alternatives for a given equivalence classe (identified by its position \( i \) in the list).

Compared to CHOICE and RANGE operators, the SORT operator has an important characteristic: the two first ones compare each alternative to all the other ones while the third one compares each alternative to a set of \( p \) profiles defining a set of \( p+1 \) predefined categories. The implementation of the SORT operator is similar to operator CLASSIFY associated with decision-map data type. The only difference is related to the fact that CLASSIFY uses the preference structure associated with ELECTRE TRI method and works on decision-map, while SORT is a more general one and can be used to implement other multicriteria sorting methods.

---

**Type:** sd-model

**set:** map-layer, decision-map, criterion-map, sUnit, uSet, cFamily, aAlter, pAlters, lAlters, sAlters, aProcedure, nProcedure, aOperator

**Syntax:**

- P-ALTERNATIVE decision-map x aCriterion x bOperator x value \( \rightarrow \) pAlters
- L-ALTERNATIVE decision-map x sUnit x sUnit \( \rightarrow \) lAlters
- EVALUATE aAlter x cFamily x aOperator \( \rightarrow \) aAlter
- SCORE aAlter x aCriterion \( \rightarrow \) real
- P-VECTOR aAlter x cFamily \( \rightarrow \) aList
PAYOFF aSet x cFamily → pTable
CONSTRAINT sAlters x aCriterion x bOperator x value → sAlters
DOMINATE sAlters x cFamily → uSet
NORMALIZE sAlters x cFamily x nProcedure → decision-map
AGGREGATE decision-map x cFamily x aProcedure → decision-map
CHOICE aSet x pStructure x cFunction → aSet
SORT aSet x sCategories x sProfiles → aSet
RANGE aSet x rdirection → aSet

Axioms:
m, r: decision-map; u: sUnit; f: cFamily; a: aProcedure; n: nProcedure; s: aSet; b: pStructure; c: cFunction; x: aAlter; y: sAlters; v, l, m: aList

P-ALTERNATIVE (d, g, op, v)
={u : u ∈ d ∧ SCORES(u, g) op v }

L-ALTERNATIVE (d, s, e)
={u_i : u_i ∈ d ∧ u_i = s ∧ u_n = e ∧ ADJACENT(u_i, u_{i+1})}

CONSTRAINT (y, g, op, v)
={x : x ∈ y ∧ SCORES(x, g) op v }

EVALUATE (x, f, op)
=∀ (g)(g ∈ f)[ASSIGN(u, g, op.combine(SCORE(u_1, g), ..., SCORE(u_r, g)))]

P-VECTOR (x, f)
= i ← 1
  ∀ (g)(g ∈ f)
  [insert(v, SCORE(x, g), i)
   i ← i+1
  ]

Figure 6. Formal specification of sd-model data type (continued in the next page)
NORMALIZE\((y,f,n)\)
\[
= \forall (x) \forall (g) \ (x \in y) \ (g \in f) \\
\text{SCORE}(x,g) \leftarrow n.\text{combine}(\text{SCORE}(x_1,g),\ldots,\text{SCORE}(x_r,g))
\]

AGGREGATE\((x,f,o)\)
\[
= o.\text{combine}(\text{P-VECTOR}(x,f))
\]

CHOICE\((s,p,f)\)
\[
= \{ a \in s : p.S(a,b,g) \ \forall \ b \in s \ \forall \ g \in f \}
\]

RANGE\((s,p,f)\)
\[
= i \rightarrow 1 \\
\text{While } s<>\emptyset \\
\quad [\text{insert}(l,\text{CHOICE}(s,p,f),i) \\
\quad \quad i \rightarrow i+1 \\
\quad \quad s \rightarrow s \setminus \text{GET}(i-1,l) \\
\quad ]
\]

SORT\((s,b,p,f)\)
\[
= \forall (x) \ (x\in s) \\
\quad [\forall (h) \ (u\in b) \\
\quad \quad \text{if } p.S(a,h,g) \text{ then } \text{insert}(l,x,h+1) \\
\quad ]
\]

---

**Figure 6.** Formal specification of \(sd\)-model data type (continued)

### 6. IMPLEMENTING DMA

The implementation of DMA is ongoing. DMA is being implementing through C++ on ArcGIS 9.1. Each data type in DMA is defined as a class and the operations associated with it are defined as methods for these classes. Figure 7 illustrates the generic definition some data types. The other data types are defined in similar way. Each of these classes contains, in addition to the methods corresponding to the operations of the data type that they implement, two specific methods representing the constructor and destructor of the class. These two methods are not shown. Figure 7 shows also the implementation of DISTANCE and SET operators with gPoint and criterion-map data types. The piece of code in Figure 8 shows a didactic example illustrating the use of some data types from DMA. The objective of this example is to select a corridor for some linear infrastructures. This example may apply in problems like the construction of highways, pipelines, etc. As mentioned earlier, within a decision-map, a corridor is modeled as a sequence of linearly adjacent spatial units linking two spatial units representing the start and end points. They are denoted \(s\) and \(e\) in Figure 8.

First, the example generates an initial decision-map, called \(r\), by intersecting three criterion-maps \(c_1\), \(c_2\) and \(c_3\); which we suppose that they have been already created. Then, the example uses the CLASSIFY operator to generate a final decision-map. To apply this operator and for the purpose of this example, cFamily and sProfiles are simply defined as one-dimensional and two-dimensional array, respectively.
class gPoint{
    //...
    public:
        gPoint MAKE(double, double);
        boolean ISEQUAL(gPoint, gPoint);
        double DISTANCE(gPoint, gPoint);
        double X(gPoint);
        double Y(gPoint);
    //...
};

double gPoint::DISTANCE(gPoint p, gPoint q) {
}

class map_layer{
    private:
        string rSystem;
        double mScale;
    //...
    public:
        map_layer MAKE(string name);
        gPoint PUT_P(map_layer, double, double);
        gLine PUT_L(map_layer, gPoint, gPoint);
        gPolygon PUT_Y(map_layer, pSet);
        map_layer INTERSECT(map_layer, map_layer);
    //...
}

class criterion_map{
    private:
        float cWeight;
        float qThreshold;
        float pThreshold;
        float vThreshold;
    //...
    public:
        criterion_map SET(char, float);
    //...
};
criterion_map criterion_map::SET(char type, float value){
if (type='q') this->qThreshold=value;
if (type='p') this->pThreshold=value;
if (type='v') this->vThreshold=value;
}

Figure 7. Generic definitions of some data types in C++

main()
{
    decision_map r;
criterion_map c1, c2, c3;
r.intersect(c1,c2,c3);
    string cFamily[3];
cFamily =(c1,c2,c3);
double sProfiles[4][3];
    //...
r.classify(r,b,f);
sUnit s, e;
    //...
    uSet corridors;
corrirods=r.L-ALTERNATIVE(r,s,e);
}

Figure 8. A didactic example illustrating some operations from DMA

7. CONCLUSION

Map algebra is a powerful tool to implement cartographic modeling. Several extensions to the original map algebra are available in literature. However, neither the original map algebra nor its different extensions are able to support multicriteria aspects of spatial problems. This is due essentially to the absence of convenient operators permitting to support multicriteria modeling. In this paper, we have proposed a new algebra, called DMA, especially devoted to multicriteria spatial modeling. As other map algebras, the proposed DMA has several merits: (i) it is a rigorous mathematical modeling framework, (ii) well adapted to object-oriented implementation, which is a natural way to deal with geographic objects and phenomena; and (iii) its independent from the way data are internally stored. Additionally, our DMA works for both vector and raster data representation (most of proposed algebra works only on raster data representation). More importantly, DMA is a powerful environment for multicriteria spatial modeling. In addition, using DMA needs a limited knowledge of multicriteria modeling. Currently, the implementation of DMA is ongoing. DMA is being implemented on ArcGIS 9.1 of ESRI using Visual C++. A part some common operators, the major part of DMA operators are especially devoted to outranking-based multicriteria modeling methods. The addition of
other operators devoted to utility function-based multicriteria modeling methods is under study. In future time, we envisage the development of visual and script-based versions of DMA.

ACKNOWLEDGEMENT

The first author would like to thank Rim Kalai for the constructive discussions concerning an earlier version of this paper.

REFERENCES


APPENDIX A: DMA DATA TYPES SPECIFICATIONS

In this Appendix we provide the description of DMA data types. To make the paper short, the axiom part is provided for only some operators.

Type: gPoint

set: gPoint, real, boolean
Syntax:
MAKE real x real → gPoint
ISEQUAL gPoint x gPoint → boolean
DISTANCE gPoint x gPoint → real
X gpoint → real
Y gpoint → real

Axioms:
i, j: real; p, q: gPoint
X(MAKE {i, j})
= i
Y(MAKE {i, j})
= j

Figure A.1. Formal specification of gPoint data type

Type: gLine
set: gPoint, gLine, real

Syntax:
MAKE gPoint x gPoint → gPoint
START gLine → gPoint
END gLine → gPoint
LENGTH gLine → real

Axioms:
p, q: gPoint
START (MAKE {p, q})
= p
END (MAKE {p, q})
= q

Figure A.2. Formal specification of gLine data type

Type: gPolygon
set: gPoint, gLine, gPolygon, real, boolean

Syntax:
MAKE gPoint x ... x gPoint → gPolygon
AREA gPolygon → real
CENTROID gPolygon → gPoint

179
CONTAINS  gPolygon x gPoint → boolean
INTERSECTS  gPolygon x gLine  → boolean

**Axioms:**
pl,…,pn, x, r: gPoint; l: gLine; v: real

AREA(MAKE(pl,…,pn))
   = _area
CENTROID(MAKE(pl,…,pn))
   = _gPoint
CONTAINS(MAKE(pl,…,pn), r)
   =if (∀(p) in {pl,…, pn} X(r) ≤ X(p) and Y(r) ≤ Y(p)) then T

INTERSECTS(MAKE (pl,…,pn), l)
   = CONTAINS( MAKE(pl,…,pn), SART(l))  or
      CONTAINS( MAKE(pl,…,pn), END(l))

___________________________________________________________________________

**Figure A.3. Formal specification of gPolygon data type**

---

**Type:** map-layer

**set:** map-layer, gPoint, gLine, gPolygon, rSystem, mScale, real

**Syntax:**

MAKE name  →  map-layer
PUT map-layer x real x real  →  gPoint
PUT map-layer x gPoint x gPoint  →  gLine
PUT map-layer x gPoint x … x gPoint  →  gPolygon
INTERSECT map-layer x … x map-layer  →  map-layer

**Axioms:**
m: map-layer; p,p1,p2, q1,…,qn: gPoint; l: gLine; y: gPolygon; v1, v2: real

PUT{m,v1,v2}
   = p.MAKE(v1,v2)

PUT{m,p1,p2}
   = l.MAKE(p1,p2)

PUT{m,v1,v2}
   = p.MAKE(v1,v2)

PUT{m,q1,…,qn}
   = y.MAKE(q1,…,qn)
**Figure A.4. Formal specification of map-layer data type**

**Type:** aOperator  
**set:** aOperator, value  

**Syntax:**  
CREATE DEFERRED → aOperator  
COMBINE value x value x ⋅⋅⋅x value → value  

**Figure A.5. Formal specification of aOperator data type**

**Type:** aProcedure  
**set:** aProcedure, value  

**Syntax:**  
CREATE DEFERRED → aProcedure  
COMBINE value x value x ⋅⋅⋅x value → value  

**Figure A.6. Formal specification of aProcedure data type**

**Type:** nProcedure  
**set:** nProcedure, value  

**Syntax:**  
CREATE DEFERRED → nProcedure  
COMBINE value x value x ⋅⋅⋅x value → value  

**Figure A.7. Formal specification of nProcedure data type**

**Type:** dTable  
**set:** dTable, bRelation, value  

**Syntax:**  
CREATE DEFERRED → dTable  
ACTION value x value x value → bRelation  

**Figure A.8. Formal specification of dTable data type**
Type: aList

set: aList, value, aPosition

Syntax:

aList

INSERT aList x value x aPosition → aList
LOCALIZE value x aList → aPosition
GET aPosition x aList → value
FIRST aList → aPosition
RAZ aList → aList
REMOVE aPosition x aList → aList
NEXT aPosition x aList → aPosition
BEFORE aPosition x aList → aPosition

Figure A.9. Formal specification of aList data type

Type: pStructure

set: pStructure, aAlter, sAlter, aCriterion, bRelation, value

syntax:

P aAlter x aAlter x aCriterion → bRelation
I aAlter x aAlter x aCriterion → bRelation
R aAlter x aAlter x aCriterion → bRelation
S aAlter x aAlter x aCriterion → bRelation

Axioms:

a,b: aAlter; g: aCriterion; p: pStructure, s: sAlter

P(a,b,g)
= if SCORE(a,g) > SCORE(b,g) + g.pThreshold then aPb else ¬(aPb)

Q(a,b,g)
= if [SCORE(b,g) + g.pThreshold ≥ SCORE(a,g)] ∧
  [SCORE(a,g) > SCORE(b,g) + g.qThreshold] then aQb else ¬(aQb)

I(a,b,g)
= if [SCORE(b,g) + g.qThreshold ≥ SCORE(a,g)] ∧
  [SCORE(a,g) + g.qThreshold ≥ SCORE(b,g)] then aIb else ¬(aIb)

S(a,b,g)
= if P(a,b,g) ∨ I(a,b,g) then aSb else ¬(aSb)

Figure 10. Formal specification of pStructure data type
APPENDIX B: OVERVIEW OF ELECTRE TRI METHOD

We give here a very brief overview of the ELECTRE TRI method. First, we mention that this overview is reproduced from Mousseau (2005). A complete description can be found in Roy and Bouyssou (1993). ELECTRE TRI is a multicriteria sorting method used to assign alternatives to predefined ordered categories. The assignment of an alternative \(a\) results from the comparison of \(a\) with the profiles defining the limits of the categories. Let \(A\) denote the set of alternatives to be assigned and let \(K=\{1,2,\ldots,n\}\) be the set of indices of the alternatives. Let \(F\) denote the set of the indices of the criteria \(g_1, g_2, \ldots, g_m\) \((F=\{1,2,\ldots,m\})\), \(k_j\) the importance coefficient of the criterion \(g_j\), \(B\) the set of indices of the profiles defining \(p+1\) categories \((B=\{1,2,\ldots,p\})\), \(b_h\) being the upper limit of category \(C_h\) and the lower limit of category \(C_{h+1}\), \(h=1,2,\ldots,p\). Each profile \(b_h\) is characterized by its performances \(g_j(b_h)\) and its thresholds \(p_j(b_h)\) (preference thresholds), \(q_j(b_h)\) (indifference thresholds) and \(v_j(b_h)\) (veto thresholds). In what follows, we will assume, without any loss of generality, that preferences increase with the value on each criterion and that \(\sum_{j \in F} k_j = 1\).

Further on, we use \(a \rightarrow C_h\) to denote that the alternative \(a\) is assigned to the category \(C_h\). ELECTRE TRI builds a fuzzy outranking relation \(S\) whose meaning is "at least as good as". Preferences on each criterion are defined through pseudo-criteria (see Roy and Vincke (1984) for details on this double thresholds preference representation). The threshold \(q_j(b_h)\) represents the largest difference \(g_j(a) - g_j(b_h)\) preserving an indifference between \(a\) and \(b_h\) in respect to criterion \(g_j\). The threshold \(p_j(b_h)\) represents the smallest difference \(g_j(a) - g_j(b_h)\) compatible with a preference in favor of \(a\) in respect to criterion \(g_j\). Thus, the limits of categories are defined in terms of profiles \(b_h, h \in B\); each one is delimited by two imprecision zones (See Figure B.1).

![Figure B.1. Defining of categories in terms of profiles](image)

To validate the proposition \(aSB_h\) \((b_hSa, \text{resp.})\), two conditions must hold:

1) **Concordance:** An outranking \(aSB_h\) \((b_hSa, \text{resp.})\) is accepted only if a "sufficient" majority of criteria are in favor of this proposition.

2) **Non-discordance:** When the concordance holds, none of the minority of criteria shows an "important" opposition to \(aSB_h\) \((b_hSa, \text{resp.})\).

Beside the intra-criterion preferential information, represented by the indifference and preference thresholds, \(q_j(b_h)\) and \(p_j(b_h)\), the construction of \(S\) also makes use of two types of inter-criterion preferential information:
i) the set of weight-importance coefficients \( \{k_j, j \in F\} \) is used in the concordance test when computing the relative importance of the coalitions of criteria being in favor of the assertion \( aSb_h \) (resp.)

\( \) ii) the set of veto thresholds \( \{v_j(b_h), j \in F, h \in B\} \) is used in the discordance test; \( v_j(b_h) \) represents the smallest difference \( g_j(b_h) - g_j(a) \) incompatible with the assertion \( aSb_h \) (resp.).

As the assignment of alternatives to categories does not result directly from the relation \( S \), an exploitation phase is necessary; it requires the relation \( S \) to be "defuzzified" using a so-called \( \lambda \)-cut: the assertion \( aSb_h \) (resp.) is considered to be valid if the credibility index of the fuzzy outranking relation is greater than a "cutting level" \( \lambda \) such that \( \lambda \in [0.5, 1] \). This \( \lambda \)-cut determines the preference situation between \( a \) and \( b_h \).

ELECTRE TRI constructs an indices \( \sigma(a,b_h) \in [0,1] \) (\( \sigma(a,b_h) \), resp.) representing the credibility of the proposition \( aSb_h \) (\( b_hSa \), resp.) respectively, \( \forall a \in A , \forall h \in B \). The proposition \( aSb_h \) (\( b_hSa \), resp.) holds if \( \sigma(a,b_h) \geq \lambda \) (\( \sigma(b_h,a) \geq \lambda \), resp.). The indices \( \sigma(a,b_h) \) is defined as follows (the values of \( \sigma(a,b_h) \) is defined in similar way):

1. Compute partial concordance indices \( S_j(a,b_h) \), \( \forall j \in F \):

\[
S_j(a,b_h) = \begin{cases} 
0, & \text{if } g_j(b_h) - g_j(a) \geq p_j(b_h) \\
1, & \text{if } g_j(b_h) - g_j(a) \leq q_j(b_h) \\
[p_j(b_h)-g_j(b_h)+g_j(a)]/[p_j(b_h)-q_j(b_h)], & \text{otherwise}
\end{cases}
\]

2. Compute global concordance indice \( S(a,b_h) \):

\[
S(a,b_h) = \sum_{j \in F} k_j S_j(a,b_h)
\]

3. Compute partial discordance indices \( \text{ND}_j(a,b_h) \), \( \forall j \in F \):

\[
\text{ND}_j(a,b_h) = \begin{cases} 
0, & \text{if } g_j(a) \leq g_j(b_h) + p_j(b_h) \\
1, & \text{if } g_j(a) > g_j(b_h) + v_j(b_h) \\
[v_j(b_h)-g_j(a)+g_j(b_h)]/[v_j(b_h)-p_j(b_h)], & \text{otherwise}
\end{cases}
\]

4. Compute the global discordance indice \( \text{ND}(a,b_h) \):

\[
\text{ND}(a,b_h) = \prod_{j \in F} \{[1-\text{ND}_j(a,b_h)]/[1-S(a,b_h)]\}
\]

With \( F' = \{j \in F : \text{ND}_j(a,b_h) > S(a,b_h)\} \)

5. Compute credibility indice \( \sigma(a,b_h) \):

\[
\sigma(a,b_h) = S(a,b_h) * \text{ND}(a,b_h)
\]

The values of \( \sigma(a,b_h) \), \( \sigma(b_h,a) \) and \( \lambda \) determine the situation of preference concerning \( a \) and \( b_h \):

- \( \sigma(a,b_h) \geq \lambda \) and \( \sigma(b_h,a) \geq \lambda \) \( \Rightarrow aSb_h \) and \( b_hSa \Rightarrow aIb_h \)
- \( \sigma(a,b_h) \geq \lambda \) and \( \sigma(b_h,a) < \lambda \) \( \Rightarrow aSb_h \) and \( \neg(b_hSa) \Rightarrow aPb_h \)
- \( \sigma(a,b_h) < \lambda \) and \( \sigma(b_h,a) \geq \lambda \) \( \Rightarrow \neg(aSb_h) \) and \( b_hSa \Rightarrow b_hPa \)
- \( \sigma(a,b_h) < \lambda \) and \( \sigma(b_h,a) < \lambda \) \( \Rightarrow \neg(aSb_h) \) and \( \neg(b_hSa) \Rightarrow aRb_h \)
Two assignment procedures are available: optimistic and pessimistic. Their role being to analyze the way in which an alternative \( a \) compares to the profiles so as to determine the category to which \( a \) should be assigned. The result of these two assignment procedures differs when the alternative \( a \) is incomparable with at least one profile \( b_i \)

i) Pessimistic procedure:
   a) Compare \( a \) successively to \( b_i; \ i=p, p-1, \ldots 0 \).
   b) Let \( b_i \) the first profile such that \( aSB_i \), then assign \( a \) to category \( C_{i+1} \) (\( a \rightarrow C_{i+1} \)).

ii) Optimistic procedure:
   a) Compare \( a \) successively to \( b_i; \ i=1, 2, \ldots p \).
   b) Let \( b_i \) the first profile such that \( b_iSa \), then assign \( a \) to category \( C_i \) (\( a \rightarrow C_i \))
Spatial Dependency Modeling Using Spatial Auto-Regression

Mete Celik1, Baris M. Kazar2, Shashi Shekhar1, Daniel Boley1, David J. Lilja3

ABSTRACT
Parameter estimation of the spatial auto-regression model (SAR) is important because we can model the spatial dependency, i.e., spatial autocorrelation present in the geo-spatial data. SAR is a popular data mining technique used in many geo-spatial application domains such as regional economics, ecology, environmental management, public safety, public health, transportation, and business. However, it is computationally expensive because of the need to compute the logarithm of the determinant of a large matrix due to Maximum Likelihood Theory (ML). Current approaches are computationally expensive, memory-intensive and not scalable. In this paper, we propose a new ML-based approximate SAR model solution based on the Gauss-Lanczos algorithm and compare the proposed solution with two other ML-based approximate SAR model solutions, namely Taylor's series, and Chebyshev polynomials. We also algebraically ranked these methods. Experiments showed that the proposed algorithm gives better results than the related approaches when the data is strongly correlated and problem size is large.

Keywords: Spatial Auto-Regression Model, Spatial Dependency Modeling, Spatial Autocorrelation, Maximum Likelihood Theory, Gauss-Lanczos Method.

INTRODUCTION
Extracting useful and interesting patterns from massive geo-spatial datasets is important for many application domains, including regional economics, ecology, environmental management, public safety, public health, transportation, and business [3, 15, 17]. Many classical data mining algorithms, such as linear regression, assume that the learning samples are independently and identically distributed (i.i.d.). This assumption is violated in the case of spatial data due to spatial autocorrelation [15] and in such cases classical linear regression yields a weak model with not only low prediction accuracy [17] but also residual error exhibiting spatial dependence. Modeling spatial dependencies improves overall classification and prediction accuracies. The Spatial auto-regression (SAR) model is a generalization of linear regression to handle these concerns.

However, estimation of the SAR model parameters is computationally very expensive because of the need to compute the logarithm of the determinant (log-det) of a large matrix. For example, it can take an hour of

---

1 Computer Science Department, University of Minnesota, MN, USA, {mcelik, shekhar, boley}@cs.umn.edu
2 Oracle Corporation, USA, baris.kazar@oracle.com
3 Electrical and Computer Engineering Department, University of Minnesota, MN, USA, lilja@ece.umn.edu
computation for a spatial dataset with 10K observation points on a single IBM Regatta processor using a 1.3GHz pSeries 690 Power4 architecture with 3.2 GB memory. This has limited the use of SAR to small problem sizes, despite its promise to improve classification and prediction accuracy.

ML-based SAR model solutions [1, 5] can be classified into exact [6, 8, 11-13] and approximate solutions [7, 10, 16], based on how they compute certain compute-intensive terms (log-det term) in the SAR solution procedure. Exact solutions suffer from high computational complexities and memory requirements due to the computation of all the eigenvalues of a large matrix. Approximate SAR model solutions try to approach the computationally complex term of the SAR model by reducing the computation time and providing computationally feasible and scalable SAR model solutions. This study covers only ML-based approximate SAR model solutions. However, we will also include exact solution in our experiments for comparison purposes.

In this paper, we propose a new ML-based approximate SAR solution, and compare and algebraically rank approximate ML-based SAR model solutions. In contrast to the related approximate SAR model solutions, our algorithm provides better approximation when the data is strongly correlated (i.e., spatial dependency is high) and problem size gets high. The key idea of the proposed algorithm is to find only the some of the eigenvalues of a large matrix, instead of finding all the eigenvalues, by reducing the size of large matrix dramatically using Gauss-Lanczos (GL) algorithm [2]. Because of this property of GL algorithm, we can save huge computation costs, especially when the matrix size is quite large.

The paper compares the proposed algorithm with two related approximate approaches and the exact solution procedure. Then, we algebraically rank them to determine which method gives better approximations in what conditions. Experimental results show that the proposed algorithm saves computation time for the large problem sizes. Experiments also showed that it gives better results when the dataset is strongly correlated (i.e., spatial dependency is high).

PROBLEM STATEMENT

In this study, we rank the algebraic errors of three ML-based approximate SAR model solutions [7, 10, 16]. Given a spatial framework, observations on a dependent variable, a set of explanatory variables, and neighborhood relationship among the spatial data, SAR parameter estimation based on Maximum Likelihood theory [1, 5] aims to find the optimum SAR model parameters by minimizing the likelihood function of the SAR model solution. The problem is formally defined as follows.

Given:

- A spatial framework \( S \) consisting of sites \( \{s_1, ..., s_n\} \) for the underlying spatial graph \( G \).
- A collection of explanatory functions \( f_{x_k} : S \rightarrow R^k, k = 1, ..., K \). \( R^k \) is the range of possible values for explanatory functions.
- A dependent function \( f_y : R \rightarrow R^y \)
- A family \( F \) (i.e., \( y = \rho Wy + x \beta + \varepsilon \)) of learning model functions mapping \( R^1 \times ... \times R^K \rightarrow R^y \).
- A neighborhood relationship \( R \) on the spatial framework

Find:

- The SAR model parameters \( \rho \) and the regression coefficient \( \beta \).

Objective:
• Algebraic error ranking of approximate SAR model solutions.

Constraints:
• $S$ is a multi-dimensional Euclidean Space,
• The values of the explanatory variables $x$ and the dependent function (observed variable) $y$ may not be independent with respect to those of nearby spatial sites, i.e., spatial autocorrelation exists.
• The domain of $x$ and $y$ are real numbers.
• The SAR parameter $\rho$ varies in the range $[0,1)$,
• The error is normally distributed with unit standard deviation and zero mean, i.e., $\varepsilon \sim N(0, \sigma^2 I)$ IID
• The neighborhood matrix $W$ exhibits sparsity.

Basic Concepts
The SAR model, also known in the literature as the spatial lag model or mixed regressive model $[1, 4, 5]$, is an extension of the linear regression model (equation (1)).

\[ y = \rho Wy + x\beta + \varepsilon \]  (1)

The data structures of the SAR model can be seen in Figure 1. Here $\rho$ is the spatial autocorrelation parameter, $y$ is an $n$-by-$1$ vector of observations on the dependent variable, $x$ is an $n$-by-$k$ matrix of observations on the explanatory variable, $W$ is the $n$-by-$n$ neighborhood matrix that accounts for the spatial relationships (dependencies) among the spatial data, $\beta$ is a $k$-by-$1$ vector of regression coefficients, and $\varepsilon$ is an $n$-by-$1$ vector of unobservable error. The spatial autocorrelation term $\rho Wy$ is added to the linear regression model in order to model the strength of the spatial dependencies among the elements of the dependent variable $y$. Moran’s Index $[5]$ can be used to see whether there is significant spatial dependency in the given dataset.

The log-likelihood function (i.e., the logarithm of the ML function) to be optimized for the $\rho$ parameter is given in equation 2. The function contains two parts, such as log-det term and SSE term.

\[
\min_{\rho\in \mathbb{C}} \sum_{i=1}^{n} \frac{2}{n} \ln |I - \rho W| + \ln((I - \rho W)y)'(I - x(x'x)^{-1}x')y - \sum_{i=1}^{n} (1 - x(x'x)^{-1}x')y )^2 (I - \rho W)y
\]  (2)

The log-likelihood function optimized is using nonlinear optimization techniques, such as, golden section search, to find the best estimate for the SAR model parameters. Rather than optimizing for both SAR parameters $\rho$ and $\beta$, it is faster and easier to optimize one unknown (i.e., $\rho$) since both parameters are dependent on each other.
ML-BASED APPROXIMATE SAR MODEL SOLUTIONS

The exact SAR model solutions suffer high computational complexity and memory requirements even in parallel form [6]. These limitations have led us to investigate approximate solutions for SAR model parameter estimation with the main objective of scaling the SAR model for large spatial data analysis problems. We inspected two approximate SAR model solutions Taylor’s series expansion and Chebyshev coefficients, and developed a new approximate SAR model solution based on the Gauss-Lanczos algorithm. Then we compared all the approximate SAR model solutions.

Approximation by Taylor’s Series Expansion

[9] suggests an approximation of the log-det of a matrix by means of the traces of the powers of the neighborhood matrix, \( W \) (equation 3). It basically finds the trace of the matrix logarithm, which is equal to the log-det of the matrix. In this approach, the Taylor’s series expansion is used to approximate the \( \sum_{i=1}^{n} \ln(1 - \rho \lambda_i) \) where \( \lambda_i \) represents the \( i^{th} \) eigenvalue that lies in the interval \([-1,+1]\) and \( \rho \) is the scalar parameter from the interval \((-1,+1)\). The term \( \sum_{i=1}^{n} \ln(1 - \rho \lambda_i) \) can be expanded as \( \sum_{i=1}^{n} (\rho \lambda_i)^k / k \) provided that \( |\rho \lambda_i| < 1 \), which will hold for all \( i \) if \( |\rho| < 1 \). Equation 3, which states the approximation used for the log-det term of the log-likelihood function, is obtained using the relationship between the eigenvalues and the trace of a matrix, i.e.,

\[
\sum_{i=1}^{n} \lambda_i^k = tr(W^k).
\]

The approximation comes into the picture when we sum up to a finite value, \( r \), instead of infinity. Therefore, equation 3 is relatively much faster because it eliminates the need to calculate the compute-intensive eigenvalue estimation when computing the log-det term (Figure 2).

Figure 2. The system diagram of the Taylor's series approximation for the SAR model solution.

Approximation by Chebyshev Polynomials

This approach uses the symmetric equivalent of the neighborhood matrix \( W \) (i.e., \( \tilde{W} \)). The eigenvalues of the symmetric matrix \( \tilde{W} \) are the same as those of the neighborhood matrix \( W \). The lemma 3.1 leads to a very efficient and accurate approximation of the log-det term of the log-likelihood function shown in equation 2.
**Lemma 3.1:** The Chebyshev solution tries to approximate the log-det of $(I-\rho W)$ involving a symmetric neighborhood matrix $\hat{W}$ as in equation 4, which is the relationship of the Chebyshev polynomial to the log-det of $(I-\rho W)$ matrix. The three terms are enough for approximating the log-det term with an accuracy of 0.03% [7].

\[
\ln |I - \rho \hat{W}| = \ln |I - \rho \hat{W}| = \sum_{j=1}^{n} c_j(\rho)tr(T_{j+1}(\hat{W})) - \frac{1}{2} c(\rho) \quad (4)
\]

**Proof:** The proof of this equality is available in [14]. \(\square\)

The value of "q" is 3, which is the highest degree of the Chebyshev polynomials. Therefore, only $T_0(\hat{W})$, $T_1(\hat{W})$, and $T_2(\hat{W})$ have to be computed where:

\[
T_{k+1}(\hat{W}) = 2\hat{W}T_{k-1}(\hat{W}) - T_{k-1}(\hat{W}) \quad (5)
\]

The Chebyshev polynomial coefficients $c_j(\rho)$ are given in equation 6.

\[
c_j(\rho) = \frac{2}{q+1} \sum_{i=1}^{q+1} \ln(1 - \rho \cos(\pi(j-1)(k-1)/(q+1))) \cos(\pi(k-1)/(q+1)) \quad (6)
\]

In Figure 3, the ML function is determined by computing the maximum of the sum of the log-det of a large matrix and the SSE term. The SAR parameter $\rho$ that achieves this maximum value is the desired value that makes the classification most accurate. The parameter "q" is the degree of the Chebyshev polynomial, which is used to approximate the log-det term. The pseudocode of the Chebyshev polynomial approximation is presented in Figure 3. Lemma 3.2 reduces the computational complexity of the Chebyshev polynomial from $O(n^3)$ to approximately $O(n^2)$.

![Figure 3. The system diagram of Chebyshev polynomial approximation for the SAR model solution](image-url)
Lemma 3.2: For regular grid-based nearest-neighbor symmetric neighborhood matrices, the relationship shown in equation 6 holds.

\[ \text{tr}(\mathbf{W}^2) = \sum_{i} \sum_{j} w_{ij}^2 \] where the \((i,j)^{th}\) element of \(\mathbf{W}\) is \(w_{ij}\).

Proof: The equality property given in equation 6 follows from the symmetry property of the symmetrized neighborhood matrix. In other words, this is valid for all symmetric matrices. The trace operator sums the diagonal elements of the square of the symmetric matrix \(\mathbf{W}\). This is the equivalent of saying that the trace operator first multiplies and adds the \(i^{th}\) column with the \(i^{th}\) row of the symmetric matrix, where the \(i^{th}\) column and the \(i^{th}\) row of the matrix are the \(\mathbf{W}\) entries in a symmetric matrix. □

In the pseudocode of the Chebyshev approximation (Figure 4 (a)), the powers of the \(\mathbf{W}\) matrices, whose traces are to be computed, go up to 2. The parameter "q" is the degree of the Chebyshev polynomial which is used to approximate the term \(\ln|\mathbf{I}-\rho\mathbf{W}|\). The ML function is computed by calculating the maximum of the log-likelihood functions (i.e. the log-det term and the \(\text{SSE} \) term).

Algorithm 3.2: Gauss-Lanczos(\(\mathbf{A}, x\))

\[x_0 \leftarrow \mathbf{0}\]
\[x_{-1} \leftarrow \mathbf{0}\]
\[x_0 \leftarrow \mathbf{0}\]
\[x \leftarrow x_0∥x∥_2\]

for \(j \leftarrow 1 \) to \(r\)

\[a_j \leftarrow x^T_j \mathbf{A} x_{j-1}\]
\[v_j \leftarrow \mathbf{A} x_j - a_j x_{j-1} - v_{j-1} x_{j-2}\]
\[v_j \leftarrow ∥v_j∥_2\]
\[x_j \leftarrow v_j / v_j\]

do
\[w_j \leftarrow \text{entry of eigenvectors of } \mathbf{T}_r\]
\[I_r = \sum_j w_j^T w_j\]
if \(v_j^T = 0 \) or \(|I_r - I_{r-1}| < \varepsilon \) then break;

return \((I_r)\)

(a) Pseudocode of Chebyshev Algorithm

Algorithm 3.1: Chebyshev (\(\mathbf{W}, \mathbf{u}, p, q = \text{rmaxs} \))

\[\text{rmaxs} \leftarrow 3\]
\[\text{rmaxs} \leftarrow [1, 2, 3]\]

for \(k \leftarrow 1 \) to \(\text{rmaxs}\)
do
\[\mathbf{w}_k \leftarrow \text{conjr}((\text{real}(\text{imag})(k)-0.5))/\text{rmaxs}\]
for \(j \leftarrow 1 \) to \(\text{rmaxs}\)

\[\text{temp} \leftarrow (2/\text{rmaxs})(\ln(1-\rho x_k))\]
\[\text{temp} \leftarrow \text{temp} - x_k \ln(1-\rho x_k)\]

if \(\text{temp} < 0\) then break;

return \(\text{rmaxs}\)

(b) Pseudocode of Gauss-Lanczos Algorithm

Figure 4: The pseudocodes of: (a) Chebyshev and (b) Gauss-Lanczos algorithms

A New Approximation Based on Gauss-Lanczos

We developed a new ML-based approximate SAR model solution based on the Gauss-Lanczos algorithm (Figure 5). [2] suggests the GL method to approximate the eigenvalue problem of \(\ln|\mathbf{I}-\rho\mathbf{W}|\) (Figure 4(b)). First, the problem is transformed to quadratic form \(u^T f(\mathbf{A}) u\) (in our case \(\mathbf{A}\) equals the symmetric positive definite matrix \((\ln|\mathbf{I}-\rho\mathbf{W}|))\), where \(\mathbf{A}, u,\) and \(f\) represent a matrix, a vector, and a function, respectively. In our case, the function \(f\) represents the logarithm of a matrix. Then, the quadratic form is converted to a Riemann-Stieltjes integral problem (detailed information can be found in [2]). To approximate the integral, gauss-type quadrature rules are applied using the Lanczos procedure (equation 7).
\[
\ln |I - \rho W| = \text{tr}(\ln(I - \rho W)) \approx \frac{1}{m} \sum_{m} I_{r}^{(i)}
\]  
\text{(7)}

In equation 7, the quadrante formula is represented by \(I_{r}\), which is approximated by the GL method. The parameter \(m\) represents the number of runs of the GL method. To find a satisfactory estimation of the quantity of trace function \(\text{tr}\), the GL algorithm is applied \(m\) times and the average of the \(I_{r}\), 's are taken. The GL algorithm (Figure 5) takes two inputs, a real \(n\)-by-\(n\) symmetric positive definite matrix \(A\) and a real \(n\)-by-1 vector with \(x^{T}x = 1\). First, in the "for" loop (Figure 4(b)), GL computes \(r\)-by-\(r\) symmetric tri-diagonal matrix \(T\), until a convergence criterion (\(\gamma = 0\) or \(|I_{r} - I_{r-1}| < \zeta |I_{r}|\)) is satisfied, or GL computes \(r\) times, which can be specified by the user by transforming the \(A\) matrix to the quadrature form, where \(r \ll n\). Then, GL computes eigenvalues \(\lambda_{k}\) and first elements \(w_{k}\) of eigenvectors of matrix \(T\). Finally, \(L\) is calculated, as given in equation 8.

\[
I_{r} = \sum_{i=1}^{r} w_{i} f(\lambda_{i})
\]  
\text{(8)}

**Figure 5. Gauss-Lanczos approximation method**

**ERROR RANKING**

This section formulates the relative error ranking of the approximations to log-det and hence the effect on the estimation of the parameter \(\rho\).

Using the log-likelihood function \(\ell(\theta | y)\) given in equation 2, we can write \(\rho\) as a function of the \(\ell(\theta | y)\), such that \(\rho = f^{-1}(\ell(\theta | y))\). Thus the change (error) in the log-likelihood due to the approximation is reflected into the estimation of the parameter \(\rho\) as follows where the operator \(\Delta\) denotes the difference between the exact (i.e., true) and the approximated values:

\[
\Delta \rho = \frac{df^{-1}(\ell(\theta | y))}{d\ell(\theta | y)} \Delta \ell(\theta | y)
\]  
\text{(9)}

The quantity \(\Delta \rho\) is the error in \(\rho\) obtained from the approximate method. The quantity \(\Delta \ell(\theta | y)\) is the error in the log-likelihood function from the approximate method, which we can compute algebraically.

The derivation part will be the same for the different approximations since the initial \(\rho\) parameter is fixed and other variables are the same for each approximate solution. The error in the \(\rho\) parameter can be estimated by multiplying the error in the log-det by a derivative term.

Since we assume that we have the same \(\text{SSE}\) term for all SAR model solutions, we do not approximate it (i.e., \(\Delta \text{SSE} = 0\)). The term \(\Delta \ell(\theta | y)\) corresponds directly to the error in the log-det approximation i.e., \(\Delta \ln |I - \rho W|\).
**EXPERIMENTAL DESIGN AND SYSTEM SETUP**

In the experiments synthetic datasets were generated for different problem sizes, such as $n=400, 1600, 2500$ and for different spatial auto-regression parameters. We took 4-neighbors (i.e., North, South, East, and West neighbors) (Appendix I) of the interested cell (location) and all experiments were run on the same platform. All the experiments were carried out using the same common experimental setup summarized in Table 1.

<table>
<thead>
<tr>
<th>Factor Name</th>
<th>Parameter Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Size ($n$)</td>
<td>400, 1600, 2500 observation points</td>
</tr>
<tr>
<td>Neighborhood Structure</td>
<td>2-D with 4-neighbors</td>
</tr>
<tr>
<td>Candidates</td>
<td>Exact Approach (Eigenvalue Computation Based)</td>
</tr>
<tr>
<td></td>
<td>Taylor's Series Approximation</td>
</tr>
<tr>
<td></td>
<td>Chebyshev Polynomial Approximation</td>
</tr>
<tr>
<td></td>
<td>Gauss-Lanczos Approximation</td>
</tr>
<tr>
<td>Dataset</td>
<td>Synthetic Dataset for $\rho=0.1, 0.2, \ldots, 0.9$</td>
</tr>
<tr>
<td>SAR Parameter $\rho$</td>
<td>[0,1)</td>
</tr>
<tr>
<td>Programming Language</td>
<td>Matlab</td>
</tr>
</tbody>
</table>

**RESULTS AND DISCUSSION**

ML-based solutions of the SAR require computing the log-det of a large matrix $(I-\rho W)$, which is computationally expensive. Approximate SAR model solutions try to approximate the log-det of a large matrix by reducing computation cost of this term. It is observed that exact SAR model solution takes approximately 2 orders of magnitude of more time than approximate solutions. In this study, we algebraically ranked ML-based approximate SAR model solutions.

In the experiments we tried to identify the behavior of the candidate algorithms for different problem sizes and for different spatial autocorrelation values (thus different spatial dependencies). Exact and approximated results for the log-det term of the SAR model are given in Figure 6 and 7. The results of the GL approximation are the average of several runs. We generated synthetic datasets for different $\rho$ parameters. Figure 6 shows the approximation results for the log-det term of the SAR model of the candidate methods. We ran the experiments for three different problem sizes, such as 400, 1600, and 2500. It is observed that Taylor’s series approximation gives upper and lower bounds of the approximation log-det term of the SAR model for all problem sizes. Chebyshev approximation gives the optimum results when the spatial autocorrelation parameter $\rho$ is close to zero for all problem sizes. In contrast, for all problem sizes, GL approximation gives better results than Chebyshev approximation when the autocorrelation is high such as spatial autocorrelation parameter is close to 1. This behavior of the GL approximation can be explained by the fact that many cancellations occur while the GL calculates the logarithms of all the eigenvalues of matrix $T$, when the spatial autocorrelation low ($\rho$ is close to zero).
Figure 6. Exact and approximate values of log-det of SAR model. GL gives better approximation while spatial autocorrelation increases.

Figure 7 gives the difference in the accuracy of the results by approximation methods where the difference in the accuracy is defined by the absolute relative error defined in equation 10. It is observed that absolute relative error (% accuracy) of Taylor’s series and Chebyshev approximations increase while $\rho$ parameter is increasing but Chebyshev approximation gives better results than Taylor’s series approximation. In contrast, absolute relative error of GL algorithm decreases while $\rho$ parameter is increasing. We can conclude that GL algorithm gives more accurate results than the other methods. Therefore, GL is better than the other candidate solutions when the spatial autocorrelation is high ($\rho$ is close to 1).

\[
100 \times abs\left\{ \frac{\left(\frac{-2}{n}\right) \ln |I - \rho W|_{\text{exact}} - \left(\frac{-2}{n}\right) \ln |I - \rho W|_{\text{approximate}}}{\left(\frac{-2}{n}\right) \ln |I - \rho W|_{\text{exact}}} \right\} \tag{10}
\]
The computational cost of the Chebyshev approximation is $O(n^3)$. Using lemma 3.2 the cost of the Chebyshev approximation can be reduced to approximately $O(n^2)$. In contrast, the cost of the GL approximation is $2mrO(n^2)$, which includes $2mr$ matrix-vector multiplications of the rank-$n$ matrix. Thus, GL is slightly more expensive than Chebyshev and Taylor's series approximations. In the GL procedure, $m$ represents the number of iterations. In our experiments, $m$ was fixed (i.e., $m=400$) for each problem size. If the problem size is large enough, the effect of $m$ will be less in the computation cost. In GL, $r$ represents the size of tri-diagonal symmetric matrix $T$ where $r<<n$. The size of the $T$ matrix changes during the GL procedure according to various the problem sizes and $\rho$ parameters. In our experiments the value of $r$ varies between 5 and 8 where $r<<n$ for problem sizes 400, 1600, 2500. The effect of the $r$ parameter will also be less for the larger problem sizes. Results show that GL approximation is one of the candidate solutions for large problem sizes, especially when the spatial autocorrelation is high, and $m$ and $r$ parameters are smaller than the problem size.

It is also observed that the quality of the results of the GL algorithm depends on the number of iterations, as discussed before, and the initial Lanczos vector which is selected randomly. In our experiments, the initial Lanczos vector is selected as a discrete random vector where values of components are either -1 or 1 with the probability of 0.5. Finally, it is also observed that increasing the number of iterations can decrease the effect of the random number generator. However, increasing the number of iterations may lead to the increase in the computation cost of the GL approximation.

**CONCLUSION AND FUTURE WORK**

In this study we algebraically compared three approximate solution procedures of the SAR model and explained which method is better in what conditions and proposed a new Maximum Likelihood Theory based approximate SAR model solution based on Gauss-Lanczos algorithm. The key idea of the proposed algorithm is to find only some of the eigenvalues of a large matrix, instead of finding all the eigenvalues, by reducing the size of large matrix dramatically using Gauss-Lanczos algorithm [2]. In the experiments, Chebyshev polynomial approximation provides better approximation when the spatial autocorrelation is low. Gauss-Lanczos approximation gives better approximation than the other methods when the problem size is large and spatial autocorrelation is high. Our future work will examine how to parallelize the Gauss-Lanczos algorithm to decrease computation cost.

![Figure 7. % absolute relative errors of approximation methods defined in equation 10. % absolute error of GL decreases when spatial autocorrelation is high.](image)
ACKNOWLEDGEMENTS

The authors would like to thank the University of Minnesota Digital Technology Center and Minnesota Supercomputing Institute for use of their computing resources. The authors are grateful to the members of the Spatial Database Group, ARCTiC Labs Group, Professor Vladimir Cherkassky, Dr. Birali Runesha and Dr. Shuxia Zhang at the Minnesota Supercomputing Institute for valuable discussions. The authors also thank Kim Koffolt for helping improve the readability of this paper.

REFERENCES

APPENDIX I: FORMATION OF NEIGHBORHOOD MATRIX W

The neighborhood matrices used by the SAR model are the neighborhood relationships on one-dimensional regular and irregular grid spaces with two neighbors and two-dimensional regular or irregular grid space with "s" neighbors, where "s" is four, eight, sixteen, twenty-four and so on neighbors. The rows of the neighborhood matrix W sum to 1, which means that W is row-standardized i.e., row-normalized, row-stochastic, or Markov matrix (Figure 8(b)). A non-zero entry in the jth column of the ith row indicates that the jth observation will be used to adjust the prediction of the ith row where i is not equal to j. Thus, the ML theory estimated SAR model solutions used in our study accept neighborhood matrices from both regular and irregular grid spaces, which is a very important feature.

In Figure 8, we illustrate the formation of the neighborhood matrix on a 4-by-4 regular grid space. As noted before, modeling spatial dependency improves the overall classification (prediction) accuracy. Spatial dependency can be defined by the relationships among spatially adjacent pixels in a small neighborhood within a spatial framework that is a regular or irregular grid space. For the four-neighborhood case, the neighbors of the (i,j)th pixel of the regular grid are defined as below.

\[
\text{neighbors}(i, j) = \begin{cases} 
(i-1, i) & 2 \geq i \geq 1 \geq j \geq q & \text{North} \\
(i, j+1) & 1 \geq i \geq 1 \geq j \geq q-1 & \text{East} \\
(i+1, j) & 2 \geq i \geq 1 \geq j \geq q & \text{North} \\
(i, j-1) & 1 \geq i \geq 1 \geq j \geq q & \text{West}
\end{cases}
\]

To form row-normalized neighborhood matrix W, a non-row-normalized neighborhood matrix C and diagonal matrix D are used, such that \(W = D^{-1}C\). is formed by putting "1"s for neighborhoods of (i,j)th pixel of the spatial framework and by putting zeros for the rest of the entries. Values of D matrix can be formed as \(d_{ij} = \sum_{\alpha=1}^{n} c_{ij} \). In other words, W matrix is formed by dividing non-zero elements of C by corresponding diagonal element of D. Figure 8(a) illustrates the spatial framework and Figure 8(b) shows W matrix for the problem size 16.
PERFORMING TOPOLOGICAL CORRECTION IN A GIS

Stefano Bellesi*, Eva Savina Malinverni **

*Posizione di Funzione Informazioni Territoriali e Beni Paesaggistici - Regione Marche - Ancona (Italy)
stefano.bellesi@regione.marche.it

** Dardus – Faculty of Engineering - Università Politecnica delle Marche - Ancona (Italy)
e.s.malinverni@univpm.it

ABSTRACT

In the first part of the paper we analyze different inconsistency issues in spatial databases associated with the geometric representation of real objects coming from numerical cartography. In the numerical cartography the redaction and the representation of the real objects are organized in informative layers where classes and codes are identified. The aim of this research concerns the construction, representation and generalization, in the GIS, of different map scales by standard procedure, preserving the correspondences and the data structures. This issue involves many different types of data structures and data manipulations. We define the data models suitable to describe the physical reality in the GIS (the topological and the vector model) and the spatial relationships to perform a correct use of a spatial database. The topic of this paper is to show how topological relations can be used to detect and to correct errors in GIS data sets. In the second part, in fact, we take into account the procedure to correct the topological inconsistency in the maps at different scale, moreover after the map generalization. To make more effectively the research we implemented a new tool in a commercial software, called Topology, applying to different kind of spatial databases.

The supported normative follows different national and international documents for the standardization of the spatial database in the GIS: the Specific Techniques for the Topographical Database (DB) adopted by the Italian organization Intesa Stato-Regioni and Enti Locali in 1996 in conformity with the Standard ISO/TC 211; the documentation of the Open GIS Consortium; the Specific Techniques of the Series 25 produced by the Italian Geographic Military Institute (IGMI) and the Standard CEN TC 278.

Keyword: numerical cartography, GIS, vector model, topological inconsistency, correction tools.

INTRODUCTION

In spatial database the content follows some interpretations and laws of real geometry. This interpretation induces to many different types of data structures and data manipulations. The spatial databases deal with objects which have a position in the space as well as with spatial relationships among the objects. In fact the spatial relationships play an important role in the spatial database, since they are usually the base for a correct use of the queries. However the spatial relations, typically derived from positional information, are often imprecise and they can produce some results in conflict with the geometric representation of the real objects.

The Geographical Information Systems, based on the numerical cartography, offers two typologies of data structures: the topological and the vector model. These two data models, to be effective, must have the possibility to put inside them the description of the objects: the attributes. These three types of information (geometry, topology and attribute) are implemented in a GIS by means of some specific physical models, based on relational data structures, typically for an architecture hardware and software type client/server.

For the purpose of dialogue and to transfer the data among the different systems some formats have been defined. The topic regards the conversion of the vector format from CAD to GIS. In fact the vector data are built point by point using coordinates (geometry) to form complex elements (lines, arcs, polygons) by means a graph (network) that defines the connections of them (topology). The numerical map elements, which take the name of geometric primitives, compose the geometry of the topology. The vector data must always be accompanied by topological information and therefore the conversion between the two formats has to follow particular procedures. In comparison to the performances of the geometric representation in numerical cartography (like in a CAD) it is required in a GIS the maintenance and the management of all the mutual spatial relationships among
the different elements of the objects (the connection, the nearness and the inclusion) too. So it is necessary to reconstruct the topology of the object and to define formal rules for a correct use of geometric primitives to calculate successively some attributes of the same object (perimeter, surface…) (Rodriguez A., 2005).

THE IMPORTANCE OF THE MODEL DATA STRUCTURES FROM NUMERICAL CARTOGRAPHY TO GIS

To represent and to manage the spatial information by means a GIS, it is necessary to define the data models suitable to describe the physical reality. These models should be able to accept inside them every object of the physical world and should be sufficiently elastic to support all the combinations present in the reality. There are different models of spatial information which perform the geometric representation of spatial objects. The vector model is the most natural computer graphic translation of real objects, but the most unsuitable performances of it are the redundancies and the incongruities. The vector model has efficient algorithms for detecting properties of spatial objects (e.g., overlapping, intersection and spatial inclusion). In this model, the information in an n-dimensional space are represented by using m-dimensional geometric primitives. The common types of primitives used in this model are: points, arcs and polygons.

The other model is the topological one. The topological model repeats the traditional drawing in cartography. In fact to realize a topological structure it needs to divide the entities according to the primitives. The topological model realizes the data manipulation like in the life. This type of data manipulation involves concepts such as adjacency, connectivity and containment. A topological model can be understood as a planar network with the following primitives: points, nodes, arcs, polygons and regions. Topology is a mathematical procedure that determines the spatial properties and relationships between the objects as the connection among lines, the direction of a line, the nearness (contiguity) of areas and the definition of area. To make clear the difference between vector and topological model, look at in Figure 1 and Figure 2 how the same spatial objects are represented by the two models. The difference is that the topological model handles explicitly common boundaries and adjacency among polygons.

The data models of spatial information may be more or less efficient to determine spatial relations. Spatial relations such as adjacency and containment, however, do not require the knowledge of the absolute position of the data and they are efficiently handled with the topological model. Topological and other spatial relations are very important and are usually implicitly represented. Spatial relations are typically derived by data manipulation such that the check of the topological inconsistency involves not only the control of the stored data in the database, but also the check of the results of the data manipulation.

Figure 1 - The vector model data and some inconsistencies
TOPOLOGICAL INCONSISTENCY OF THE DATABASE

We told above that topological relations are of great importance for the GIS data sets consistency. Topological relations are based on the shape of the objects. A lot of errors contained in the GIS come from a lack of knowledge about topological relations among geographical objects stored in the database (van der Poorten et al., 2002). Consequently, the topology can help to find errors in the GIS data sets, and it can help to correct them. The topic of this paper is to present how topological relations can be used to detect and to correct errors in GIS data sets. A complete topological error correction approach required three different tasks: the definition of the errors, the check of the database and the correction of the errors in it. Since an error is defined as a forbidden topological relation between the objects, the way to correct it will be to create the correct topological relation between these objects (Ubeda et al., 1997 (a)).

An inconsistency arises when the integrity constraints are violated. The constraints must be taken into account when a database is updated, and the semantic and the quality of data must be preserved. Topological constraints consider in a particular way the geometrical properties and the spatial relations. They may be associated with structural considerations or topological conditions. The satisfaction of topological constraints ensures that some computational-geometry algorithms can be successfully executed: for example, the boundary of a region must be defined by a closed arc in order to calculate the area of a region. In general, an inconsistency associated with a primary error violates the basic principle of location or the attribute of uniqueness. Another spatial inconsistency, related to a secondary error, is referred to a contradiction between the stored data and the structural constraints of the own geometric primitives.

Other different kinds of spatial errors in GIS data sets can be defined. For example the structuring errors, like non polygon closure and self-intersecting lines coming from different data structures. The structuring errors depend on data model in a GIS. Topological consistency can be treated at the low level of data structure, counting nodes and arcs to assure that an object’s topology is complete. This strategy accounts for changes in the object geometry, but it does not assure consistency of the relations between objects. For example, it does not handle consistency of the topological changes which may happen when, at a coarse representation, several parts become a single object or when holes of objects disappear (Ubeda et al., 1997 (b)).

TOPOLOGICAL TEST OF THE DATABASE

To be able to manage the data correctly in a GIS one must take into account all the basic concepts of the topology, which suggests the necessity to perform a series of spatial relationships between geometric primitives. Every time a map is acquired it is necessary to verify that any geometric entity satisfies the topological relationships. In fact while in the phase of drawing some inconsistency, as the missed closure of two lines or the presence of double nodes, are irrelevant, at the computer stage these inaccuracies are not acceptable because they can bring to contradictory results.

The process to do is simple. If the datum is already topologically consistent it is possible to create, without mistakes, the topological structure before to verify the correctness of the data from a geometric point of view and successively to create the topological relationships. Some topological controls are applicable to mono-dimensional elements. The most frequent problems which can happen are: double and redundant vertices, overlapped arcs, double and tangent arcs (Figure 3). To test the presence of redundant vertices a control distance must be used: the minimum distance between two vertices. In this way all the overlapped vertices and all those
separated in respect of this length are removed. To verify the co-linearity between vertices too, a distance of tolerance must be inserted with respect to three consecutive nodes linking three segments. To recognize every arc considered overlapped a buffer is traced around every linear element. This verifies that any other arc is entirely contained in this area of control.

![Fig 3 - Some types of topological incongruities](image)

The errors of missed closure can generally be divided in two typologies: two nodes not closed; a node and an arc not closed. This topological check, based on the node control, finds that in a closed arc the external vertices are overlapped to the adjacent arcs, so can happen that more nodes have the same coordinates. If it doesn't occur, the arc is not perfectly closed. If inside the buffer there is another node, the arc is closed simply moving its starting or ending vertex. When an arc intersects another linear elements, the error is individualized and removed moving the node in the point of intersection and cutting the other arcs in the same position. The missed tangency between two arcs occurs when the vertex of an arc is close to another arc. All the corrected arcs can be broken in the point of tangency.

Other two errors regard the bi-dimensional elements: the overlapped areas and the areas containing holes inside. In this case it must be tested the adjacency of every region inside the map and the existence of the overlapping. If this happens every overlapping is underlined and all these areas are cut and turn into new adjacent regions among them (Figure 4). The last test checks the existence of the holes inside every area. This task is performed using a threshold value. All the regions with an area inferior to this value are considered holes and are underlined (Figure 5).

![Fig 4 - Cut of two overlapped areas](image)  ![Fig 5 - Filled hole in a area](image)

**A SET OF CORRECTION TOOLS IMPLEMENTED IN A COMMERCIAL SOFTWARE**

Some commercial GIS software have some evident gaps performing a topological investigation. These software do not deal with topological relations or consider only few relations so that the correction tools do not always work in an automatic way but sometimes require the support of a human operator. In this paper we present some tools for a right procedure to correct topological errors in the GIS. The visual interface for end-users is realized as a standalone software tool designed using MapBasic 7.0 integrated in the software MapInfo 7.5. The implemented simple code is able to personalize and to perform some topological corrections, for numerical cartography, applying different transformations, any time that a topological constraint is not verified.
commands are necessary for checking the topological structure correctness of any cartographic representation. The software is organized in two parts: one checks the mono-dimensional entities (points, lines, arcs) and the other verifies the bi-dimensional ones (polygons).

It must be underlined that it has been necessary to make a distinction among different primitives and to perform for each of them special commands and dedicated procedures. So the software is organized in many subroutines which perform specific tasks (Figure 6). To facilitate the use of every procedure a specific menu has been created, called Topology, inserted in the menu toolbar of MapInfo (Figure 7).

![Figure 6 - The subroutines implemented in the software Topology](image)

![Figure 7 - The menu Topology in MapInfo (Italian language)](image)

**AN APPLICATION TO A CADASTRAL MAP**

The first application presented concerns the topological control and correction of a cadastral map related to Genga (Ancona – Italy) at the scale 1:2000 (Figure 8).

![Figure 8 - The cadastral map](image)
As first step it is necessary to prepare the data. All the arcs of the map are organized in two tables: AAT (Arc Attribute Table) and NAT (Node Attribute Table). The Arc Attribute Table organizes all the linear entities, while the Node Attribute Table contains the respective starting and ending nodes. These tables are the first and important task to perform the topological correction tools in the better way. In this way the geometric information are unified in a table and the topological properties can be verified in an easy and fast way (Figure 9).

Figure 9 - AAT, NAT tables and the correspondence in the map

The corrections regard the relative position among the objects. The goal of this correction is the change of the topological relations between two objects without changing the relations with the other objects of the database. Such corrections can be used to adjust the borders of two closed regions, a line and the border of a region, or two lines. The length of the line or the area of the region doesn’t change. Another topological change between two objects is the split of one in two different sub-objects. The only condition is that the two new topological relations were different from the previous one. This kind of correction is useful to keep the planarity of a map.

At first in this application we checked the existence of double or redundant nodes. The following step we controlled the double arcs. A specific dialogue box in the software informs that the arcs are completely or only partially overlapped. At the end of this performance a window notifies about the errors individualized and, in positive case, how many arcs have been erased. In the tested map the software found three overlapped entities, then manually deleted, in such way to verify the correctness of the list of errors. After having decreased the number of arcs it was necessary to check their perfect closure in correspondence of the nodes. Continuing with the topological controls some not connected arcs were individualized and corrected but not in automatic way. Also two lines crossing each other are a forbidden topological relation. A way to correct such error is to create an intersection point and to split the line in two parts. After these corrections the geometry of the representation doesn't lose out substantial changes. The control distance used in every verification has been selected equal to 0.4 meter, based on the graphic error multiplied by the scale of the map (1:2000). The distance used for the co-linearity test has been set equal to 0.1 meter (Figure 10).

Figure 10 – The dialogue windows ask the values to perform the check
The use of the program has underlined two problems during the correction phase: the first one regards the opened arcs without intersection. They are not automatically removed, but the operator, examining the map, decides to make some changes. In this case the segments have been manually modified and then the same procedure has been iterated. The second problem takes place in correspondence of some nodes where the control area value has been too small to perform the closure of themselves. Also in this case we have performed a manual correction. The last control on the arcs regards the missed tangency. It didn’t produce any correction in the map. Over the first phase of checking, the table Polygon Attribute Table (PAT) has been created (Figure 11). It contains all the areas in the map, identified by the coordinates of its own centroid.

![Figure 11 - PAT table and the correspondence in the map](image)

Then the overlapping of the new generated areas has been verified. During this control none inaccuracy has been individualized. This is possible because the bi-dimensional entities directly depend on the mono-dimensional ones. If these entities don’t contain topological errors also the areas result correct.

**AN APPLICATION RELATED TO TECHNICAL MAPS AT DIFFERENT SCALE AFTER THE GENERALIZATION.**

This other application introduces the problem of the map generalization and the correspondence among different scale in the GIS. Multiple representations considers the same information but with a different representation or level of detail to perceive spatial reality in different ways. This introduces redundancy in the database and generates consistency problems throughout update operations (Zhou et al., 2001). The process called map generalization changes the representation of scale-dependent phenomena, eliminating excessive detail and simplifying the appearance and density of objects to improve the analysis capabilities (Davis et al., 1999). An important requirement of multi-scale spatial databases is that topological consistency is maintained both within individual features and between co-displayed features for all scales at which they may be retrieved (Laurini, 1998).

The consequent management of the topological structure and the corrections could be performed with the same implemented software adding appropriate changes. The study made on a sample area by four technical maps at the scale 1:10.000 of the Marche Region (DB10, year 1999) in comparing with the new cartographic product of IGMI at the scale 1:25.000 (DB25, year 1993). In particular the analyzed layer was the “Transports”, in detail roads and railways. The goal was to derive from a correct data structure in the map at the scale 1:10.000 a new valid geometric and topological data structure suitable to the scale 1:25.000. The procedure is based on different steps to limit the numbers of the processes to be performed manually, which can often produce errors. The work concerned: the realization of a graphic and symbolic representation; the definition of the characteristics of the geometric primitives in the GIS; the test of the coincidence between the symbolic representation and the same geometric primitives; the comparison between the DB10 with the DB25 data structures after the generalization; the planning for the GIS updating of other new layers, by a symbolic representation, derivable from the geometric primitives; the respect of the logical, spatial and topological relationships during the generalization of maps at different scale.

The road network in the GIS is represented by arcs and nodes. The Table 1 shows the list of all the features in the map distinct for code, description, geometric primitive associated and compared with the cartographic representation. Some remarks can be made:
• the cartographic representation in both scale is symbolic;
• every entity is represented by points or lines except the railway stations where the buildings are represented by the primitive area;
• the railway is a linear representation with some symbols inserted in correspondence of the bridges and the tunnels;
• the roads have anymore a symbolic representation.

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Geometry</th>
<th>Map Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>040101</td>
<td>One-track railway</td>
<td>Polyline</td>
<td>Polyline</td>
</tr>
<tr>
<td>040103</td>
<td>Shunting track</td>
<td>Polyline</td>
<td>Polyline</td>
</tr>
<tr>
<td>041101</td>
<td>Railway tunnel</td>
<td>Polyline</td>
<td>Polyline</td>
</tr>
<tr>
<td>041102</td>
<td>Railway bridge</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
<tr>
<td>041105</td>
<td>Railway crossing</td>
<td>Point</td>
<td>Symbolical</td>
</tr>
<tr>
<td>042101</td>
<td>Railway Station</td>
<td>Polygon</td>
<td>Polygon</td>
</tr>
<tr>
<td>050101</td>
<td>Motorway</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
<tr>
<td>050103</td>
<td>Paved road</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
<tr>
<td>050104</td>
<td>Under construction road</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
<tr>
<td>050105</td>
<td>Unpaved road</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
<tr>
<td>050106</td>
<td>Entrance to or exit of buildings</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
<tr>
<td>050301</td>
<td>Mule-track</td>
<td>Polyline</td>
<td>Polyline</td>
</tr>
<tr>
<td>050302</td>
<td>Path</td>
<td>Polyline</td>
<td>Polyline</td>
</tr>
<tr>
<td>051101</td>
<td>Motorway bridge</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
<tr>
<td>051201</td>
<td>Road bridge</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
<tr>
<td>051701</td>
<td>Tunnel road</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>Code</th>
<th>Definition</th>
<th>Geometry</th>
<th>Map Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAN010</td>
<td>L201</td>
<td>One-track railway</td>
<td>Polyline</td>
<td>Polyline</td>
</tr>
<tr>
<td>LAN010</td>
<td>L201A</td>
<td>Track inside railway station area</td>
<td>Polyline</td>
<td>Polyline</td>
</tr>
<tr>
<td>LAQ130</td>
<td>L219</td>
<td>Railway tunnel</td>
<td>Polyline</td>
<td>Polyline</td>
</tr>
<tr>
<td>LAQ040</td>
<td>L355A</td>
<td>Iron railway bridge, l&gt;25 m</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
<tr>
<td>LAQ040</td>
<td>L352A</td>
<td>Reinforced concrete railway bridge, l&gt;25 m</td>
<td>Point</td>
<td>Symbolical</td>
</tr>
<tr>
<td>PAQ040</td>
<td>P352A</td>
<td>Reinforced concrete railway bridge, l&lt;25 m</td>
<td>Point</td>
<td>Symbolical</td>
</tr>
<tr>
<td>PAQ062</td>
<td>P212B</td>
<td>Unguarded level crossing</td>
<td>Point</td>
<td>Symbolical</td>
</tr>
<tr>
<td>AQQ12</td>
<td>C222</td>
<td>Railway Station</td>
<td>Polygon</td>
<td>Polygon</td>
</tr>
<tr>
<td>LAP030</td>
<td>L302</td>
<td>A2-road of 4 lanes</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
<tr>
<td>LAP030</td>
<td>L303A</td>
<td>A3-paved road of 2 or 3 lanes, w=7 m</td>
<td>Polyline</td>
<td>Symbolical</td>
</tr>
</tbody>
</table>
LAP030 L303B A4-unpaved road of 2 or 3 lanes, w=7 m Polyline Symbolical
LAP030 L304A A5- paved road of 1 or 2 lanes, w=3,5÷7 m Polyline Symbolical
LAP030 L304B A6-unpaved road of 1 lane, w=3,5÷7 m Polyline Symbolical
LAP030 L305A A7- unpaved road of 1 lane, w=2,5÷3,5 m Polyline Symbolical
LAP030 L311 Secondary importance road into town Polyline Symbolical
LAP030 L315 A5/A6-under construction road Polyline Symbolical
LAP020 L301B Slip road Polyline Symbolical
LAP050 L306 B2-cart road Polyline Polyline
LAP050 L307 B3-mule track Polyline Polyline
LAP050 L308 B4-path Polyline Polyline
LAQ130 L316A Tunnel road Polyline Polyline
LAQ040 L351A Reinforced concrete A2 road bridge, l>25 m Polyline Symbolical
LAQ040 L351C Reinforced concrete A3/A4 road bridge, l>25 m Polyline Symbolical
LAQ040 L351E Reinforced concrete A5/A6 road bridge, l>25 m Polyline Symbolical
PAQ040 P351A Reinforced concrete A2 road bridge, l<25 m Point Symbolical
PAQ040 P351C Reinforced concrete A3/A4 road bridge, l<25 m Point Symbolical
PAQ040 P351E Reinforced concrete A5/A6 road bridge, l<25 m Point Symbolical
PAQ040 P356C A5/A6 road bridge in curve Point Symbolical
PAQ040 P356D A7 road bridge in curve Point Symbolical

Table 1 – DB10 and DB25 feature classes

To construct the symbology correctly and to test the arcs a buffer value could be used containing the half true value of the symbol (Table 2).

<table>
<thead>
<tr>
<th>Code</th>
<th>Symbolical values in mm</th>
<th>Buffer values in m</th>
</tr>
</thead>
<tbody>
<tr>
<td>L355A</td>
<td>0,50</td>
<td>6,25</td>
</tr>
<tr>
<td>L352A</td>
<td>0,50</td>
<td>6,25</td>
</tr>
<tr>
<td>L302</td>
<td>0,70</td>
<td>8,75</td>
</tr>
<tr>
<td>L303A</td>
<td>0,55</td>
<td>6,9</td>
</tr>
<tr>
<td>L303B</td>
<td>0,55</td>
<td>6,9</td>
</tr>
<tr>
<td>L304A</td>
<td>0,45</td>
<td>5,65</td>
</tr>
<tr>
<td>L304B</td>
<td>0,45</td>
<td>5,65</td>
</tr>
<tr>
<td>L305A</td>
<td>0,40</td>
<td>5</td>
</tr>
<tr>
<td>L311</td>
<td>0,35</td>
<td>4,85</td>
</tr>
<tr>
<td>L315</td>
<td>0,45</td>
<td>5,65</td>
</tr>
<tr>
<td>L301B</td>
<td>0,50</td>
<td>5</td>
</tr>
<tr>
<td>L351A</td>
<td>0,70</td>
<td>8,75</td>
</tr>
<tr>
<td>L351C</td>
<td>0,50</td>
<td>6,25</td>
</tr>
<tr>
<td>L351E</td>
<td>0,40</td>
<td>5</td>
</tr>
</tbody>
</table>
Table 2 – DB25 feature class “buffer” values

In the DB25 there are a lot of classes more than the DB10. Furthermore the roads are classified, in both maps, for the type of flooring and for the value of width and the bridges use the point as primitive, when the length is less than 25 meters. To assign to every road element a new class it was necessary to acquire some information on the ground or by orthoimages also to draw correctly the width of the roads and to overlay the Regional Environmental Database (ISTAT 2001) containing the polygons of the urban area. To simplify the task and to make it automatically it could be useful to align the two DB. At the end of this process all the classes of the bridges have been inserted in the database and automatically the new features from the DB10 have been generated. This step changes the starting network scheme with the risk to lose the topological consistency of the data structure. In the Figure 12 three possible cases of alteration are showed. The case (a) deletes a non connected arc. In the map at the scale 1:25,000 the small traffic access can be skipped. The case (b) regards the bridges and consists in the detection of two new dangle nodes which are automatically corrected by means a value of tolerance of 13 meters. The case (c) simplify the intersections. At the end of the analysis of the cases b and c it is possible to perform a merge correcting in automatic way also the case a to produce a final database, without to introduce topological errors.

Figure 12 - Example of changes of the representation scheme passing from DB10 to DB25

Complying the graphic representation of the generalized map at the scale 1:25000 a change of the scheme of the road network could be developed. The creation of a buffer equals an half value of the symbolic width for all the objects, increased of 2.5 meters (graphic error multiplied by the scale of the map), allows to perform a first check to individualize the overlapped areas or the holes inside the areas. The correction of these errors modifies the network but solves the problem of not correct geometric neighboring between the object in the graphic representation. At this point the generalization has been completed and the construction of the layers related to the railway, bridges and roads follows. The limits of the buffer must be represented and for every layer “area” (LAP020_A, LAP030_A, LAP040_A, LAQ040_A, PAQ040_A) the correspondent linear element (LAP020_L, LAP030_L, LAP040_L, LAQ040_L, PAQ040_L) must be built and operated the editing on it. Another code "NO PLOT", not visible in phase of printing, must be assigned (Figure 13).
The creation of the layers “area”, in both the maps, allows to reduce some procedures of editing which could be necessary in the traditional cartography, as the intersection with the contour lines or the layer “Hydrography” and every entity area of the layer “Vegetation” (Figure 14).

Figure 14 - Example of symbolic representation

There are two topological rules to keep: the connection between the linear elements and the coincidence of the nodes with the end vertices of all the arcs. On the contrary only a spatial rule is given: the coincidence of the railway network with the road one in correspondence of the level crossing. The produced network DB25, after the generalization, verifies the rules above listed. It is possible to complete the classification of all the other nodes of the network using the COD value (the number of convergent arcs in a node) as it is described in the scheme of the Table 3.

Figure 14 - Example of roads and railways layers representation
<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>COD</th>
<th>Derived DB25 class</th>
<th>Updating process</th>
</tr>
</thead>
<tbody>
<tr>
<td>G219</td>
<td>Railway tunnel</td>
<td>2</td>
<td>L219</td>
<td>by intersects</td>
</tr>
<tr>
<td>G355A</td>
<td>Iron railway bridge, l&gt;25 m</td>
<td>2</td>
<td>L355A</td>
<td>by intersects</td>
</tr>
<tr>
<td>G352A</td>
<td>Reinforced concrete railway bridge, l&gt;25 m</td>
<td>2</td>
<td>L352A</td>
<td>by intersects</td>
</tr>
<tr>
<td>P352A</td>
<td>Reinforced concrete railway bridge, l&lt;25 m</td>
<td>2</td>
<td>P352A</td>
<td>by join</td>
</tr>
<tr>
<td>P212B</td>
<td>Unguarded level crossing</td>
<td>4</td>
<td>P212B</td>
<td>by join</td>
</tr>
<tr>
<td>G222</td>
<td>Railway Station</td>
<td>2</td>
<td>C222</td>
<td>manual</td>
</tr>
<tr>
<td>G201</td>
<td>Bifurcation</td>
<td>&gt;=3</td>
<td>Feat. “LAN010”</td>
<td>by intersects</td>
</tr>
<tr>
<td>G202</td>
<td>Ending/starting track</td>
<td>1</td>
<td>Feat. “LAN010”</td>
<td>by intersects</td>
</tr>
<tr>
<td>G 351A</td>
<td>Reinforced concrete A2 road bridge, l&gt;25 m</td>
<td>2</td>
<td>L351A</td>
<td>by intersects</td>
</tr>
<tr>
<td>G 351C</td>
<td>Reinforced concrete A3/A4 road bridge, l&gt;25 m</td>
<td>2</td>
<td>L351C</td>
<td>by intersects</td>
</tr>
<tr>
<td>G 351E</td>
<td>Reinforced concrete A5/A6 road bridge, l&gt;25 m</td>
<td>2</td>
<td>L351E</td>
<td>by intersects</td>
</tr>
<tr>
<td>P351A</td>
<td>Reinforced concrete A2 road bridge, l&lt;25 m</td>
<td>2</td>
<td>P351A</td>
<td>by join</td>
</tr>
<tr>
<td>P351C</td>
<td>Reinforced concrete A3/A4 road bridge, l&lt;25 m</td>
<td>2</td>
<td>P351C</td>
<td>by join</td>
</tr>
<tr>
<td>P351E</td>
<td>Reinforced concrete A5/A6 road bridge, l&lt;25 m</td>
<td>2</td>
<td>P351E</td>
<td>by join</td>
</tr>
<tr>
<td>P356C</td>
<td>A5/A6 road bridge in curve</td>
<td>2</td>
<td>P356C</td>
<td>by join</td>
</tr>
<tr>
<td>P356D</td>
<td>A7 road bridge in curve</td>
<td>2</td>
<td>P356D</td>
<td>by join</td>
</tr>
<tr>
<td>G300</td>
<td>Intersection</td>
<td>&gt;=3</td>
<td>Feat. “LAP0%0”</td>
<td>by intersects</td>
</tr>
<tr>
<td>G301</td>
<td>Change code DB25 features “LAP%0”</td>
<td>2</td>
<td>Feat. “LAP0%0”</td>
<td>like last updating, by query where code is null</td>
</tr>
</tbody>
</table>

Table 3 – DB25 Updating Node feature class

The software implemented in Mapinfo 7.0, called Topology and illustrated in the preview application, has different limits if it is employed in this case. Particularly MapInfo doesn’t allow to create a buffer type Flat, suitable for the cartographic representation of the roads and it doesn’t perform a merge that maintains unchanged the topological structure of the network scheme. Moreover it doesn’t check the graphic representations of the lines and the symbols in millimeters. So the software Topology should add other commands and allow to export the file with the attributes x, y, COD (where x is the East coordinate and y the North coordinate). After this processing, however possible by a simple query SQL select too, the software could be suitable to be used to verify and correct again the maps.

CONCLUSION

The problem of topological consistency has been considered with regard multiple representations. We presented some procedures for maintaining consistency between these representations. The increase of automatic procedures needs to standardize entity definitions with their spatial and logical relationships. The maintenance of
topological consistency between multiple representations in a database requires line and polygon generalization procedures that can be guaranteed simplifications topologically consistent. The transition must take advantage of the semantic relationship that exists among the various representations of the same object. So it’s necessary, to fix for each entity in every different scale of representation the type of representation (symbolical or real measure), the metric tolerance (graphic error, minimum surface) and the possible characteristic of the object (shape). A further topic of future research is the development of incremental update procedures to maintain large databases of topologically consistent multi-scale data.

REFERENCES


