An approach to evaluating motion pattern detection techniques in spatio-temporal data

Patrick Laube*[a] and Ross S. Purves[b]

[a]Department of Geography, University of Zurich

Winterthurerstrasse 190, CH-8057 Zurich

Telephone: +41-44-635 51 31

Fax: +41-44-635 68 48

plaube@geo.unizh.ch

[b]Department of Geography, University of Zurich

Winterthurerstrasse 190, CH-8057 Zurich

Telephone: +41-44-635 65 31

Fax: +41-44-635 68 48

rsp@geo.unizh.ch

*Corresponding author

An approach to evalating motion pattern detection techniques in spatio-temporal data

Abstract. This paper presents a method to evaluate a geographic knowledge discovery approach for exploring the motion of point objects. The goal is to provide a means of considering the significance of motion patterns, described through their *interestingness*. We use Monte-Carlo simulations of constrained random walks to generate populations of synthetic lifelines, using the statistical properties of real observational data as constraints. Pattern occurrence in the synthetic data is then compared with observational data to assess the potential interestingness of the found patterns. We use motion data from wildlife biology and spatialisation in political science for the evaluation. The results of the numerical experiments show that the interestingness of found motion patterns is largely dependant on the configuration of the pattern matching process, which includes the pattern extent, the temporal granularity, and the classification schema used for the motion attributes azimuth and speed. The results of the numerical experiments allow interestingness to be attached only to some of the patterns found, – other patterns were suggested to be not interesting. The evaluation method helps in estimating useful configurations of the pattern detection process. This work emphasises the need to further investigate the statistical aspects of the problem under study in (geographic) knowledge discovery.

Keywords. geographic knowledge discovery, motion, lifelines, pattern detection, constrained random walk, Monte Carlo experiments.

1. Introduction

Location aware devices are becoming ubiquitous and will increase our capability to collect spatio-temporal motion data by many orders of magnitude. The ubiquity of such devices is reflected by the fact that in the summer of 2004 the Japanese Telecommunications Council declared that all mobile phones introduced in Japan after 2007 should have self-locating functionality. Studies of so-called moving point objects (MPO), incorporating information about changing positions of discrete objects in time and space, have been identified as a key emerging research area in GIScience (Miller 2003). By studying MPOs through time, individual geospatial lifelines can be derived from large datasets collected, for example, from people carrying GPS-enabled phones and PDAs (e.g. Dykes & Mountain, 2003), tracked animals in field studies (e.g. Wentz et al. 2003), or even tracked football players in sports scene analysis (e.g. Iwase & Saito, 2003).

It has been recognised that not only is spatial data special, but also the handling and analysing of spatio-temporal data, and above all motion data, requires the development of new concepts (Frank, 2001; Mark, 2003). Traditional analytical methods for spatial and spatio-temporal data were developed in an era when data collection was expensive and computational power was weak (Miller & Han, 2001). Miller and Han thus reason that "traditional spatial analytical techniques cannot easily discover new and unexpected patterns, trends and relationships that can be hidden deep within very large and diverse geographic datasets" (Miller & Han, 2001, p. 3).

The integration of knowledge from the field of GIScience about space-time together with the emerging field of Knowledge Discovery in Databases opens up the possibility for Geographic Knowledge Discovery (GKD). Applications which generate large volumes of spatio-temporal data, such as high-resolution (in time and space) satellite-based systems (Griffiths & Mather, 2000), and in our case the analysis of geospatial lifelines (Hornsby & Egenhofer, 2002; Mark, 1998) face multiple challenges in the storage and exploration of high-volume spatio-temporal datasets and thus present excellent cases for the application of GKD.

It has been recognized in the knowledge discovery in databases (KDD) literature that discovery systems can generate a glut of patterns, most of which are of no interest to the user (Silberschatz & Tuzhilin, 1996; Padmanabhan, 2004). Thus, it is recommended that data mining be carried out with regard to the statistical aspects of the problem (Fayyad et al., 1996).

In this paper work is presented which extends the *relative motion* (REMO) GKD approach developed to identify motion patterns in groups of MPOs (Laube & Imfeld, 2002, Laube et al., 2004; 2005). The REMO approach proposed so far (section 2.2) allows the user to search large volumes of data for instances of pre-defined motion patterns, which are constructed on the basis of existing knowledge about the motion of the objects under study. However, the user has no means by which to estimate the significance (for example, the uniqueness) of the extracted patterns. The number and the extent of patterns found may depend significantly on both the motion data and the parameterisation of the pattern detection process. Many more patterns may be identified in a space where motion is constrained, for example on a football pitch, than in the seemingly chaotic motion of children in a playground. A central question for GKD is therefore, how can we assess the significance of patterns extracted from such cases? The central issue of this paper is therefore to provide a means of considering the significance of REMO patterns.

Our approach focuses on the use of Monte-Carlo simulations to generate synthetic lifelines constrained by the statistical properties of real observational data. Pattern occurrence in the simulated data is then compared with observational data to assess the potential significance of the patterns. Finally, having identified potentially interesting patterns we return to the observational data to investigate whether these patterns have meaning in terms of the system under investigation.

The paper is structured as follows. Section 2 provides a literature overview on geographic knowledge discovery in general, the REMO GKD approach, pattern interestingness, and on different potential approaches to simulating geospatial lifelines. In Section 3 the central ideas of our evaluation approach are introduced. The data used in the study are introduced in Section 4. Section 5 describes the methodology for producing constrained random walks to produce synthetic lifelines, and their use in Monte-Carlo simulation for numerical experiments. The results of these experiments are presented in Section 6, and their meaning and application to the observational data is discussed in Section 7 before examining the general applicability of these results to the field of GKD.

2. Background

2.1. Geographic Knowledge Discovery

Knowledge discovery in databases (KDD) is a set of methods for identifying high-level knowledge from low-level data in the context of large datasets. As an interdisciplinary approach it integrates methods from machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge acquisition for experts systems, data visualisation, and high-performance computing (Fayyad et al., 1996). KDD moves beyond the traditional domain of statistics to accommodate data normally not amenable to statistical analysis. Descriptive statistics usually involve small and relatively noiseless numeric data scientifically sampled from a large population with a specific question in mind. In contrast, KDD is designed for data collected and stored in many scientific or enterprise databases which are potentially noisy, non-numeric, and incomplete (Miller & Han, 2001).

The KDD process is interactive and iterative, involving data selection, data cleaning and pre-processing, data reduction and projection, exploratory analysis and data mining, and result interpretation and evaluation. The central belief of KDD is that information is hidden in very large databases in the form of interesting patterns (Miller & Han, 2001). Data mining, that is the application of specific algorithms for extracting patterns from data, is thus just one component of the overall KDD process, (Fayyad et al., 1996). Classical data mining tasks are exploratory data analysis, descriptive modelling, predictive modelling, discovering patterns and rules, and retrieval by content (Hand et al., 2001).

Miller and Han (2001) identified unique needs and challenges for integrating KDD into GIScience. They argue that geographic data has unique properties that require special KDD and specifically data mining approaches. As examples of uniquely spatial properties they list the geographic measurement framework, spatial dependency and heterogeneity, the complexity of spatio-temporal objects and relationships, and diversity of involved data types. Hence, the need for specific geographic knowledge discovery and geographic data mining techniques is proposed as an important area for research. In recent years some explicit GKD approaches have been presented. One such example is the work of Ng (2001), which uses outlier detection (discovering patterns and rules) to investigate trajectories of video tracked people, as opposed to the retrieval by content approach of REMO.

2.2. The REMO approach

REMO GKD seeks to discover motion patterns in the lifelines created by a group of MPOs, moving without constraints in a featureless space. Suitable lifeline data consists of the trajectories of a set of MPOs each featuring a list of fixes (x,y,t). The approach is based on two key features (Figure 1):

- a transformation of the lifeline data (a) to an analysis matrix (b,c) featuring motion attributes (i.e. speed, change of speed or motion azimuth),
- followed by matching of formalized patterns on the matrix (d).

The geospatial lifelines in Figure 1 (a) represent the tracks of four GPS-collared deer. Deer O_1, moves with a constant motion azimuth of 45° during an interval from t_2 to t_5, i.e. four discrete time steps of length ∂t, and shows a *constancy* pattern. The four deer with the same motion azimuth of 45° at the time t_4 show a *concurrence* pattern. An MPO anticipating the motion of others shows a *trend-setter* pattern. Thus, a *trend-setter* pattern is constructed by searching for *constancy* in conjunction with *concurrence* (e.g. deer O_1 anticipates at t_2 the motion azimuth 45° that is reproduced by all other MPOs at time t_4).

The approach has utility in a number of fields, for example:

- urban planners might use such approaches to identify behavioural patterns which suggest how groups of individuals interact in relatively unconstrained spaces such as station concourses;

- wildlife biologists can investigate migration patterns, leading to more effective protection of, for example, animals at important points in their life cycles; and

- in soccer scene analysis managers may use such an approach to identify key playmakers and their followers.

(place Fig.1 about here)

The following references provide more detail about the REMO approach. Laube and Imfeld (2002) introduced the basic idea of relative motion analysis, Laube et al. (2005) introduced algorithms and formalisms to detect motion patterns. Finally, Laube et al. (2004) extended the initial family of relative motion patterns to include the absolute postitions of moving objects.

2.3. Pattern Interestingness

Finding interesting patterns in large databases requires that pattern discovery algorithms are developed, where the patterns found are considered to be an anomalous feature of the data, departing from what might be expected (Hand, 2004). The perennial problem of pattern discovery is to define what an anomalous pattern is within the context of a very large dataset. A feature of discovered patterns is considered by the KDD community to be their *interestingness* (Silberschatz & Tuzhilin, 1996, Padmanabhan, 2004) – that is to say do the patterns have some meaning to the user within the context of the KDD process. Within KDD criteria for interesting patterns are divided into those which are objective and subjective. Objective measures depend solely on the structure of the pattern and the underlying data. Subjective measures depend also on the class of users exploring the data, bearing in mind that patterns that are of interest for one user class, may be of no interest to another class (Silberschatz & Tuzhilin, 1996). Tuzhilin and Silberschatz (1996) identify two reasons why a pattern is interesting from a subjective point of view: *unexpectedness*, which indicates how surprising the pattern is to a user, and *actionability* which indicates whether the user can act on the pattern to his/her advantage. However, in very large datasets a pattern may be unexpected, but nonetheless not unusual. A key need in examining unexpectedness is therefore to quantify how often a supposedly unexpected pattern occurs within the dataset, and better, the likelihood of such an unexpected pattern occurring. In this paper we use interestingness to assess the potential significance of mined patterns in a GKD process.

2.4. Simulating Motion

Although the potential of large datasets containing many thousands of MPOs for use in GKD is clear, the actual availability of very large datasets has been limited in the past for two main reasons. Firstly, these datasets were simply not available for a combination of legal (e.g. privacy) and technical reasons (Pfoser & Theodoridis, 2003), and secondly research in KDD and GKD has proceeded on the basis of the expected increase in volume of data. Thus, considerable work has been done in generating synthetic motion data in a diverse range of subjects including ecology, transport modelling and database research. Importantly, the use of synthetic data allows the generation of arbitrary spatial and temporal granularities. A summary of key elements of this work is introduced below.

In population ecology synthetic motion data is widely used to study the distribution and abundance of organisms. See Turchin (1998) for an introduction to the measuring and modelling population redistribution in animals and plants. In transport modelling synthetic motion data emerges from traffic simulation for transportation planning (Nagel et al., 2000). In the spatio-temporal data base management (STDBMS) community synthetic data has been used generating semantics-based trajectories of MPOs for designing novel data types and access methods for spatio-temporal databases (Pfoser & Theodoridis, 2003) and for benchmarking of proposed database models (Brinkhoff, 2002; Saglio & Moreira, 2001).

A variety of different approaches exist to the simulation of motion data. These range from relatively simple random walk procedures, commonly used in ecology (Turchin, 1998) through to complex multi-agent simulations (Batty et al. 2003).Whatever the approach, there are two key features: firstly, the simulation of the MPOs themselves and secondly, the environment within which they can act, interact and react.

In our case, the environment has no influence on the MPOs other than constraining their movement within some bounding box. Our datasets consist solely of fixes, with no associated environmental attributes for a given time and place. Thus, we focus here on the representation of MPOs as simple point objects performing some kind of *random walk.* Within the random walk framework MPOs link moves, each featuring a duration, a speed and a direction. A move can be influenced by the direction of its previous move, local conditions or attraction towards some destination. Typically (but not necessarily), there is a stochastic element involved in selecting the next move (Turchin, 1998). For *correlated random walks* the direction of the current move does affect the direction of the next move. Step length and turning angle of the subsequent move are drawn from some stochastic frequency distribution, for instance in the case of direction with turning angles concentrated around zero. A *biased random walk* includes the influence of an absolute direction, i.e. a long distance attraction. The work presented in this paper adopted a *constrained random walk,* where constraints from the analysis of real observation data are used to set turning angle distributions and step sizes (Wentz et al., 2003).

Monte Carlo simulation is widely used in GIScience and other fields to quantify the effects of variables with some uncertainty (Heuvelink, 1999). In combination with constrained random walks and the associated frequency distributions, Monte Carlo techniques offer the possibility of generating multiple lifelines for use in experiments with synthetic data and comparison with observational data.

2.5 Aims of this paper

From this background we can summarise the key aims of this paper.

- How can we estimate the interestingness of motion patterns mined by the REMO GKD approach?

- Is the use of Monte Carlo simulated populations of synthetic motion data an appropriate strategy for evaluating a motion pattern detection approach?

- Do interesting patterns as defined through these techniques correspond with events which have meaning to domain experts?

3. Evaluating the REMO approach

Hand (2004) defined pattern discovery as the search for anomalous features of the data, departing from the expected. However, the challenge lies in defining the expected. If an unexpected pattern is also unique, then the user can reasonably examine its interrelationships and infer meaningfulness. If, on the other hand, an unexpected pattern occurs many times how can we first assess if it is not only unexpected, but unusual given the parameterisation of the data? The first step in assessing such unexpectedness is therefore to develop techniques to simulate lifelines which have similar statistical properties to the observed data. Our approach to simulating lifelines uses constraints derived from the objects under study to derive synthetic motion data of arbitrary spatial and temporal granularities.

Silberschatz and Tuzhilin (1996) propose unexpectedness as a measure of interestingness of patterns. They argue that patterns are interesting because they contradict our expectations, given by our system of beliefs. One option to capture the beliefs is to formulate a statistical hypothesis. The degree of belief is then defined as a significance level for which the statistic for a certain test is on the borderline of acceptance of the hypothesis. The main drawback of this statistical approach to indicate unexpectedness is that not every belief can be formulated as a testable hypothesis (Silberschatz & Tuzhilin, 1996). For this reason we replace the statistical hypothesis with Monte Carlo simulations, choosing the detour of numerical experiments to learn about “the expected” from the stochastic properties of the simulations.

Since our aim is to assess the interestingness of REMO patterns, we hypothesise that interesting patterns will have a higher ratio of occurrence to those created in simulated CRWs. Monte Carlo simulations are used to generate a population of simulated CRWs, which can be compared with the observational data. Our underlying assumption is that those patterns which appear to be outliers from the stochastic properties of the simulations are those which we can attach some initial meaning to, prior to further investigation by the user.

The technique used to generate lifelines is crucial. Synthetic lifelines created by a complete random walk would not be suitable in assessing the interestingness of motion patterns since within the complete random walk framework every step is completely independent of the previous step with respect to both direction and step length. Such lifelines do not represent the characteristic motion of either migrating animals or the Swiss districts. Behavioural ecologists have addressed this problems by using simulation techniques that consider the characteristics of MPOs and their motion behaviour (Bergmann et al., 2000; Byers, 2001; Turchin, 1998; Wentz et al., 2003).

For many simulation problems the use of stochastic frequency distributions known from the literature which characterise the objects and their motion are a suitable choice. However, if such general knowledge is not available, one possible solution is to derive frequency distributions directly from the observation data at hand. For example, Wentz et al. (2003) used two different techniques to create synthetic trajectories filling gaps in the incomplete lifelines of primate species for analysing home range and daily ranging patterns. They illustrated that the use of constrained random walk produced reasonable approximations of field observations and for certain species performed significantly better in analysis than simple linear interpolation between the known points in the trajectory.

4. Data

Two contrasting datasets are used in evaluating the REMO GKD approach. One of these consists of wildlife data, with a relatively small number of objects, but well understood behaviour. The other dataset are a classic example of spatialisation, where aspatial attributes are projected into a geographic space. This second dataset consists of many more objects and time steps than our wildlife data, and is a typical example of a dataset where GKD might be expected to reveal hitherto unseen patterns.

4.1. Porcupine Caribou

The Porcupine Caribou Herd Satellite Collar Project is a cooperative project that uses satellite radio collars to document seasonal range and migration patterns of the Porcupine Caribou Herd (PCH) in northern Yukon, Alaska and Northwest Territories (NWT). Details about the Porcupine Caribou Project including the original tracking data can be found under the following URL (http://www.taiga.net/satellite).

This data set has been selected because it has the ideal granularity suitable to analyse seasonal migration. Furthermore seasonal migration is a known behaviour within the PCH and well documented (Fancy et al., 1989; Fancy & Whitten, 1991; Griffith et al., 2002). A key test of the usefulness of the REMO system must be its ability to identify known patterns, such as this migration. Just as with any other wildlife field study, this observation data set does not provide a complete coverage of all individuals over the whole study period. Animals die, lose their GPS collar, or had their collars removed. Thus, our experiments focus on a subgroup of the herd, consisting of 10 individuals simultaneously tracked over almost two years, starting from March 2003 (see Figure 2a).

(Insert Fig. 2 about here)

4.2. Abstract Data Points

Frequent popular referendums in Switzerland (approx. 8-10 per year) allow researchers to make detailed inferences about value conflicts within society. Hermann and Leuthold (2001) developed an inductive approach to explore basic ideological conflicts in Switzerland. By performing factor analysis on referendum data at the district level of all 158 federal referendums held between 1981 and 1999, they hypothesised a structure of mentality, which was interpreted as being composed of three dimensions: political left vs. political right, liberal vs. conservative and ecological vs. technocratic (Hermann & Leuthold, 2003). In these two dimensional ideological spaces the 185 districts can be positioned at intervals of one year, from 1981 until 1999 and their movement through this hypothesised ideological space over a period of 20 years followed (see Figure 3a).

This data set has been chosen for two reasons. Firstly it features a large number of simultaneously tracked MPOs (n = 185). Secondly, the districts of a Canton (member states of the Swiss Federation) show a certain institutional and cultural similarity, potentially expressed in similar motion behaviour within the abstract space, – potentially mirrored in detectable patterns. The major limitation of this data set lies in its small number of time steps (t = 20).

(Insert Fig. 3 about here)

5. Methodolgy

5.1. Simulating Lifelines using Constrained Random Walk

To generate simulations with properties as close as possible to the observed motion phenomena, we use constraints derived from the observation data. The constraints are given as frequency distributions of step length and direction change (Figures 4–7). This establishes an empirical link between the simulated and real data, without which the utility of the random walk model is considered as being severely restricted (Turchin, 1998).

(place Fig. 4-7 about here)

For the construction of synthetic trajectories using a constrained random walk, we generate for every synthetic MPO its own step length and direction change array, for instance of size d = 1000, corresponding to the observed frequency distributions of the MPO. The construction of any new point P_{t+1} of the trajectory requires selection of a sample s_{t+1} from the step length distribution as well as a sample a_{t+1} from the direction change angle distribution, where the sample is selected at random from the distribution (see Fig. 8).

(place Fig. 8 about here)

A crucial element of constrained random walks are the initial conditions. In both case studies we use random starting locations for the artificial MPOs since the algorithms of the REMO GKD approach evaluated in this paper only consider relative, and not absolute, positions of MPOs. Differing approaches were used for the starting direction. For the PCH data random starting angles were applied since a long temporal interval, with observed distributions lying between –180 and +180 degrees exists (Fig. 4) For the moving districts, by contrast, the actual first step directions of the observation data were used to initialise the random walk model. This option was chosen because the starting angle has a large influence on the outcome the simulation under the given conditions of a short temporal interval and a narrow direction change distribution (Fig. 6).

5.2. Monte-Carlo Experiments

The design of the MC experiment is twofold: In a first step we generate a set of r data sets with synthetic lifelines of n MPOs (see Figure 9). In a second step we run the REMO pattern matching algorithms on the generated lifeline data systematically varying the properties of the pattern matching process (see Fig. 10).

Generating synthetic data. For a direct comparison of the observed data with the simulated trajectories we reproduced 100 synthetic groups as described above (see Figurers 2b and 3b).

- **PCH.** n = 10 MPOs, t = 96 weeks, ∂t = 2 weeks, r = 100 simulation runs.

- **Moving districts.** n = 185 MPOs, t = 20 years, ∂t = 1 year, r = 100 simulation runs.

(place Fig. 9 about here)

REMO GKD with variable configurations. We performed REMO pattern detection experiments for the patterns *constancy, concurrence*, and *trendsetter* with respect to motion *azimuth* and *speed* for both case studies. Three nested loops control the experiments (see Fig 10).

- In the innermost loop we vary the pattern extension *p*. In the case of the pattern over time (e.g. constancy) *p* expresses the pattern length, that is, over how many time steps a pattern exists. *p* takes values from 2 time steps (i.e. 4 weeks with the PCH data, 2 years with the Swiss district data) to a maximum number representing the whole time period (t = 96 weeks, , t = 20 years). In the case of patterns across objects (e.g. concurrence), p stands for the number of involved MPOs, ranging from 2 to the total number of MPOs in the group (10 caribou, 185 districts respectively).

- This procedure we apply *r* times for all the *r* synthetic motion data sets in the next exterior loop. Thus r = 100 represents the number of Monte Carlo runs.

- Finally, in the outermost loop, we repeat the MC experiment for the predefined set of attribute classifications *c*. For the experiments involving motion azimuth we used c = [2, 4, 8, 16, 36, 360] attribute classes (i.e. two classes with 180° intervals, four classes with 90° intervals, eight classes with 45° intervals, …). For the experiments involving speed we used c = [2,4,8,16] classes.

(place Fig. 10 about here)

5.3. Assessing interestingness of REMO patterns

For every pattern matching run of the configuration (*c,r,p*) both the number of possible and the number of found patterns are calculated. The ratio of found to possible patterns is then used to provide a standardised means to indicate how many patterns were found per pattern matching run of configuration (*c,r,p*).

The maximum number of patterns that can be packed into a matrix is dependent on the number of attribute classes and the pattern extent. This maximum number decreases as pattern extent increases. The influence of the number of classes is less obvious, since it depends on the pattern type. For constancy the number of classes has no influence on the maximum number of possible patterns since the matrix could be packed with an alternation of attributes with only two classes. For concurrence the maximum number of patterns per column corresponds to the number of attribute classes, since ordering of the column is irrelevant in instantiating a pattern

All result plots will have the following structure (see for example Figure 12). The mean of the ratio of the number of patterns found for the Monte-Carlo simulations to the possible number of matches for this pattern length are plotted as whisker plots. The *x*–axis represents the extent *p* of the patterns, referred to as pattern *length* for patterns over time (e.g. constancy) and pattern *width* for pattern across objects (e.g. concurrence). The ratio on the *y*–axis represents the number of patterns found. The whisker plots for the simulation data (n = 100) feature mean, 25th and 75th percentile, as well as minimal and maximal values.

Following Tuzhilin and Silberschatz (1996) we define that patterns are worthy of investigation when we find more matches in the observed data than in the simulated data. We further suggest, that when the ratio of the number of patterns observed is an order of magnitude different from the simulated data some qualitative notion of interestingness can be extracted. For example, in Figure 12 (bottom) the ratio of found/possible has a value of about 0.2 for the observed data, and a mean of around

0.02 for the simulated data with a pattern width of five. We would suggest therefore that concurrence patterns of width five with 8 azimuth classes are worthy of investigation in this (PCH) dataset.

6. Results of the numerical experiments

In this paper we present results from experiments performed, with varying pattern extension and attribute granularity, in order to evaluate the interestingness of the following patterns: *constancy*, *concurrence* and *trendsetter* (Table 1).

Fig.	**Data**	**Motion property**	**pattern**	**attribute granularities**	**Pattern extent *p***
11	Porcupine Caribou Herd	azimuth	constancy	4 (top), 8 (bottom)	length 2-20 time steps
12	Porcupine Caribou Herd	azimuth	concurrence	4 (top), 8 (bottom)	width 2-10 individuals
13	Porcupine Caribou Herd	azimuth	trendsetter	4 (top), 8 (bottom)	3 and 4 steps anticipation, width 2-10 individuals
14	Porcupine Caribou Herd	speed	constancy	8	length 2-20 time steps
15	Porcupine Caribou Herd	speed	concurrence	8	width 2-10 individuals
16	Swiss political districts	azimuth	concurrence	4 (top), 8 (bottom)	width 2-65 districts

Table 1. Configurations of the numerical experiments

(place Fig. 11-16 about here)

The results we present in this section involve the motion properties *motion azimuth* and *speed*, focussing on azimuth. Although having performed experiments with a wide range of different attribute granularities, we selected for this paper the two most commonly chosen granularities for indicating directions, that is four and eight cardinal directions (corresponding to the geographical directions N, E, S, and W, as well as N, NE, E, SE, S, SW, W, NW, respectively).

A simple relationship between granularity and pattern length is present in all plots – in both observed and simulated data – finer attribute granularities have less simulated (and also less observed) patterns as pattern the extent increases. See for example Figure 12 to compare the number of *concurrence* patterns found with azimuth granularity of four and eight classes respectively. A further feature of changing granularity is that as granularity increases, in our data, we find more observed than simulated patterns as the pattern extent increases. For example, in Figure 12 (top), with 4 classes, more short patterns are present in the simulated data than the observed.

Due to the discrete nature of the REMO patterns as well as the analysis matrix, the number of maximally possible patterns per matrix does not decrease smoothly with increasing patterns extents. In contrast, the number of possible patterns per matrix expresses discrete jumps, reflected as artefacts in the result plots (see for example Figure 16 top). For example, with 4 azimuth classes and 185 districts, maximally 4 instances of a concurrence of width 46 are possible at a single time step (4*46=184<185). Increasing the concurrence width by one to 47, all of a sudden only 3 instances of concurrence can be packed in a time step (4*47=188>185). However, since simultaneously also the number of found patterns is expected to decrease for the same reason, the ratio *found/possible* remains a reliable measure to indicate the success of a pattern.

7. Discussion

In discussing the results shown in Section 6, we consider two different aspects of the study. We firstly seek to interpret the plots derived from comparison of the properties of the observed data with the simulations, without consideration of the context of the data. That is to say, we do not take note of whether our MPOs are caribou or Swiss political districts in this first discussion. Secondly, the results from the simulations are used to identify potentially interesting patterns within our data. We look for these patterns and then consider them in the light of the context of the data – for example, the spring migration of caribou, or the socio-political shifts in the society in the case of our political districts. Thirdly, we discuss the proposed method in the light of our findings.

7.1. Discussion of features of plots

Consider first the case of constancy illustrated in Figure 11 (top), with 4 classes, where we find more short patterns in the simulated data than the observed. This simple result leads us to conclude that we can attach no meaning to these short patterns in the observed data. By contrast, in Figure 12 (bottom), with 8 classes, almost all pattern width have more matches in the observed data than the simulated. Following our definition of interestingness these patterns are unexpected and therefore interesting.

Figure 11 shows an example where simulation shows that constancy patterns in the observed PCH data are not worthy of further investigation. Here, it is clear that for a granularity of 4 classes the number of simulated patterns is greater than or equal to the number of observed patterns. For patterns derived from 8 attribute classes, the situation is slightly less clear cut. A small ratio of long patterns occur in the observed data, but since the simulated data still suggest these are unusual (no long patterns being found in the simulated data), they may merit further examination.

The numerical experiments for concurrence affirm a straightforward hypothesis about the relation between attribute classification and pattern extent (pattern length/width). The number of MPOs divided by the number of attribute classes gives a lower threshold of pattern interestingness. With 185 MPOs and 4 classes (8 classes), concurrence patterns with a width of less than 46 (below 23) are expected simply by randomly sampling motion azimuths from the 4 classes (8 classes).

It is at a first glimpse rather surprising that below the threshold many less concurrence patterns are found in the real data that would be expected from the experiments (see Fig. 12 (top), Fig. 16). Assuming that the observed motion is indeed coordinated one can easily explain the lack of real patterns below the threshold.

Where coordinated motion does in fact exist, then wider concurrence patterns will be found. This in turn reduces the number of occurrences of narrower patterns, as seen in the observed data in Figures 12 (top) and 16. By contrast, in the simulated data where the angular distribution is evenly distributed across motion azimuth classes, narrow patterns are more likely to be found.

7.2. Analysis of results with respect to case studies

7.2.1. Porcupine Caribou Herd (PCH)

Figure 17 depicts the REMO matrix for the PCH trajectories with c=8 colour coded motion azimuth classes. The pattern matching results illustrated in Figures 18 und 19 are performed on this matrix.

(place Fig. 17 about here)

Fancy and Whitten (1989) refer to two short directed seasonal migration patterns in the PCH seasonal motion, a spring migration in May and a fall migration in September. The PCH “commutes” between calving areas near the Beaufort Sea in early summer and the winter areas in the South. Having in mind that the sampling rate of ∂t = 2 weeks is rather coarse compared to the duration of the whole seasonal migration, caribou moving straight on for longer temporal intervals (∂t = 4, 6, 8 weeks) cannot be expected in their trajectories.

From Figure 12 we derived a concurrence width of five as being a meaningful pattern. Thus, the results illustrated in Figure 18 showing concurrence patterns of at least width five can be assigned interestingness. All patterns appear in the migration seasons, where PCH heads Northwest (spring; light and dark green colours) and Southwest (late summer and early fall; orange, red colours), respectively.

(place Fig. 18 and 19 about here)

Bearing in mind that very long periods of straight motion are not expected, we investigated whether we could find trend-setting caribou, anticipating the seasonal migration before the rest of the herd. Indeed, as shown in Figure 19 (top), we find trend-setting individuals with at least five followers anticipating the spring and fall migration. For instance, the caribou Lynetta, anticipated in mid August 2004 the fall southward migration of more than 5 followers caribou in early September (see Figure 19, orange pattern in right part of plot). Again, we have chosen the number of followers that expressed the largest difference between observed and simulated in Figure 13, here p = 5 individuals.

Investigating the REMO matrix representing c = 8 speed classes, one would expect to find many concurrence patterns both in the fast migration season (blue) as well as in the slower sedentary seasons (red) (Figure 20). As shown in Figure 15. the pattern matching process results indeed in relatively high ratios of found/possible. However, since the variance of the ratio in the simulated data is rather high and the ratio found in the observation data is in the same range, we can’t assign interestingness to the found concurrence patterns. Also with constancy (Figure 14) the observed patterns lie in the rather wide range of the simulation. Again, with a coarse temporal sampling rate and many different activity periods in the caribou year (two migration seasons, scattering in the calving season, mid-August dispersal (Fancy, et al. 1989), we cannot expect to find significant speed constancy patterns.

(place Fig. 20 about here)

7.2.2. Swiss political districts

Investigating the motion direction of the Swiss political districts in Figure 16, we see that meaningful concurrence patterns appear to occur after a threshold length of 48 for four classes and for a threshold of 25 with eight classes, where significantly more patterns than predicted by the simulations are observed. Laube et al. (2005) suggested a concurrence of 45 districts as a hint for a political left-right divergence, ideologically separating the German and the Latin part of Switzerland in the 1980s and 1990s. Now, having numerical evaluation experiments at hand, we can indeed attach interestingness to that pattern (Figures 21 and 22). Investigating the Figure 16 bottom, we see that concurrence patterns with a width of 45 appear in the simulated pattern with a ratio of found/possible of almost 0, in the observed data with around 0.1 respectively, hence satisfying the required order of magnitude. In the same publication Laube et al. (2005) identified a concurrence pattern of width 18. Considering the results in Figure 16 we would not assign any interestingness to that pattern anymore and rather expect to find patterns with this extent purely by chance.

(place Fig. 21 and 22 about here)

7.3. Discussion of proposed method

One might first argue that using observation data to parameterise the simulation and thereafter comparing the simulation with actual observation data has an inherent danger, in that any such reasoning is circular. However, our aim is not to analyse the statistical properties of the static trajectories of MPOs, but rather to consider the unexpectedness of emerging motion patterns. Comparison of our data with a random sample would increase the likelihood of a pattern being unexpected (since any spatial or temporal autocorrelation in motion azimuth and step length would be removed) and unexpected patterns would be more likely. Thus, arguably, the use of correlated random walks based on the distribution of observational data is a stricter test of unexpectedness.

The selection of constraints for the numerical experiments is crucial. In our work, we have used all available MPOs to define constraints. The initial direction of movement was defined differently in the two datasets. For the caribou, where turning angles span ±180° the initial direction of movement was randomly assigned and had no influence on the results due to the large number of time steps. However, in the case of the Swiss political districts the use of a random starting angle, coupled with the narrow distribution of turning angles and the small number of timesteps resulting in motion with similar characteristics in terms of speed, but widely varying motion azimuths. By using observed motion azimuth this influence was reduced. A further step in measuring the potential influence of initial azimuth would be to randomly subdivide the data to derive both constraints and initial directions of movement for comparison with the remaining observed data.

As shown above, small samples of lifelines may represent very specific motion characteristics, such as in the case of the moving districts. For instance, if we simply observe the caribou during migration periods, the constraints given by the frequency distribution would describe for most MPOs a rather rapid and linear motion behaviour. Thus, in such data, motion azimuth concurrence could be expected and such patterns would be of no interest. In contrast, a sedentary period with randomly foraging MPOs would simply not express any pattern to assess interestingness. However, by using lifeline populations from constraints featuring both migratory and sedentary periods, we are able to identify instances of such patterns if they are actually present in the observational data. It will be important to revisit the methodology as larger samples of observational data, with more normal distributions become available.

Two datasets are used in this study. In the case of the caribou, only ten MPOs are present. In the case of the Swiss political districts a total of 185 MPOs exist, for a relatively short number (20) of time-steps. Initial research (Laube and Imfeld, 2002; Laube et al., 2005) sought to identify patterns in these datasets through visual inspection of patterns. However, though these datasets appear initially small, when the number of potential combinations of pattern, granularity, and motion type are considered very large volumes of data are created. Furthermore, Kwan (2000) argues convincingly that the user's ability to identify meaningful patterns is quickly swamped using visual exploration. Therefore, we here seek to marry techniques based on ideas for dealing with large datasets with commonly used visualisation approaches. The use of these techniques shows considerable promise, allowing patterns to be identified which were not apparent only through inspection, but will require further testing as large datasets become available.

8. Conclusions

8.1. Main contributions

In this paper we presented a method to assess qualitative data mining measures of interestingness in geographic knowledge discovery. The GKD approach under study is the REMO GKD, developed to detect motion patterns in the lifelines of moving point objects. We propose a procedure to estimate the interestingness of the motion patterns found in lifeline observation data. Therefore, we first generate populations of constrained random walks using Monte Carlo simulations. Secondly, we compare the success of the pattern matching process on observation data with that in the simulated data.

Previous publications discussing the REMO GKD approach suggested that there is a strong relation between the parameterisation of the pattern detection process and the number of patterns found (Laube & Imfeld, 2002; Laube et al., 2005). Having the numerical experiments of this work at hand, we have now evidence to reinforce this hypothesis.

We suggest that such an approach – testing the interestingness of patterns by performing numerical experiments to measure their unexpectedness is one which can be transferred to other techniques which seek to extract information from very large datasets. Furthermore, as was shown by our experiments, we would argue that caution must be attached to patterns that are presumed to be interesting only through visual inspection.

8.2. Implications for studies of moving point objects

The MC experiments with both the Caribou and the Swiss district data point out that performing the REMO GKD process a lower pattern length/width threshold must be adopted depending on the classification schema.

The evaluation method introduced in this work can be used to examine useful configurations of pattern matching sessions (i.e. attribute granularities, pattern lengths), which may not be obvious. The method helps, for example, in selecting suitable compromises between granularity and information.

In general, methods similar to our approach, can be used in (geographic) knowledge discovery to focus the search for meaningful patterns. Hence, not only the (putatively biased) knowledge of the potential user may be used to configure the GKD process, but also evidence from numerical experiments.

Motion pattern studies like REMO are crucial to better understand why 'spatial (and spatio-temporal) is special' with respect to knowledge discovery, – to learn more about the 'G' in GKD. Carried out carefully, GKD has a huge potential for applications in spatialising socio-economic processes or wildlife studies.

8.3. Outlook

This paper discusses a general methodological issue – how can we evaluate evaluate geographic knowledge discovery methods. A generalisable set of techniques, focussed on the notion of pattern interestingness is introduced, and applied to the REMO framework. To illustrate such an approach we selected three exemplary patterns from REMO and developed pattern interestingness experiments. However, to improve our confidence in this approach, we must advance the interestingness studies by extending the set of investigated relative motion patterns, including the whole family of motion patterns introduced in Laube & Imfeld (2002), and Laube et al. (2004, 2005). Furthermore, moving beyond the specific REMO framework, the concept of pattern interestingness should be included within similar data mining applications and may thus improve our confidence in the patterns discovered through knowledge discovery methods within GIScience.

Future work may include alternative approaches to the generation of lifelines to optimally approximate the MC simulations for the observed motion process. The use of Markov Chain transition matrices would allow inclusion of in-path autocorrelation, such as preferred turning sequences for certain intervals (Jones and Smith, 2001). The use of cellular automatons (Dijkstra et al, 2000) or autonomous software agents (Batty and Jiang, 1999) would additionally allow MPOs to interact with each other and with their environment, potentially leading to more realistic simulation of lifelines.

Acknowledgements

The authors wish to thank Michael Hermann and Heiri Leuthold, University of Zurich, for providing the Swiss districts data, and the Porcupine Caribou Technical Committee, the Porcupine Caribou Management Board and the Wildlife Management Advisory Council (North Slope) for providing the excellent caribou data. The authors would furthermore like to acknowledge invaluable input from Nikos Koutsias, Robert Weibel, and Stephan Imfeld, all University of Zurich.

References

Batty, M., & Jiang, B. (1999). Multi-agent simulation: new approaches to exploring space-time dynamics within GIS. Technical report, Centre for Advanced Spatial Analysis, University College London.

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 Geographical Information Science*, *17*(7), 673–697.

Bergman, C.M., Schaefer, J. A., & Luttich, S. N. (2000). Caribou movement as a correlated random walk. *Oecologia*, *123*(3), 364-374.

Brinkhoff, T. (2002). A Framework for Generating Network-Based Moving Objects, *GeoInformatica, 6*(2), 153-180.

Burstedde, C., Klauck, K., Schadschneider, A. & Zittartz, J. (2001). Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A, 295*(3-4), 507-525.

Byers, J.A. (2001). Correlated random walk equations of animal dispersal resolved by simulation. *Ecology, 82*(6), 1680-1690.

Dijkstra, J., Jessurun, A.J., &Timmermans, H.J.P. (2000). A Multi-Agent Cellular Automata System for Visualising Simulated Pedestrian Activity. In S. Bandini, and T. Worsch (Eds.), *Theoretical and Practical Issues on Cellular Automata, Proceedings of the Fourth International Conference on Cellular Automata for Research and Industry* (pp. 29-36). Berlin-Heidelberg: Springer.

Dykes, J.A. & Mountain, D.M. (2003). Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and application. *Computational Statistics and Data Analysis, 43*(4), 581-603.

Fancy, S.G. & Whitten, K.R. (1991). Selection of calving sites by Porcupine herd caribou. *Canadian Journal of Zoology, 69,* 1736-1743.

Fancy, S.G., Pank, L.F., Whitten, K.R. & Regelin, W.L. (1989). Seasonal movement of caribou in arctic Alaska as determined by satellite. *Canadian Journal of Zoology, 67,* 644-650.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 37–54.

Frank, A. U. (2001). Socio-Economic Units: Their Life and Motion. In Frank, A.U., Raper, J. & Cheylan, J.P. (Eds.), *Life and motion of socio-economic units, GISDATA 8* (pp. 21-34). London: Taylor & Francis.

Griffith, B., Douglas, D.C., Walsh, N.E., Yound, D.D., McCabe, T.R., Russell, D.E., White, R.G., Cameron, R.D. & Whitten, K.R. (2002). The Porcupine caribou herd. In Douglas, D.C., Reynolds, P.E. & Rhode, E.B. (Eds.), *Arctic Refuge coastal plain terrestrial wildlife research summaries* (pp. 8-37), U. S. Geological Survey, Biological Resources Division, Biological Science Report USGS/BRD/BSR-2002-0001.

Griffiths, G.H. & Mather, P.M. (2000). Remote Sensing and landscape ecology: landscape patterns and landscape change. *Int. J. Remote Sensing, 21*(13-14), 2537-2539.

Hand, D. J. (2004). Pattern Discovery. *Journal of Applied Statistics, 31*(8), 883-884

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining.* Cambridge, MA: MIT Press.

Hermann, M., & Leuthold, H., (2001). Weltanschauungen und ihre soziale Basis im Spiegel der Volksabstimmungen. *Swiss Political Science Review, 7,* 39–63.

Hermann, M., & Leuthold, H., (2003). *Atlas der politischen Landschaften. Ein weltanschauliches Portrait der Schweiz.* Zurich: vdf-Verlag.

Heuvelink, G. B. M. (1999). Propagation of error in spatial data modelling with GIS. In Longley, P., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (Eds.), Geographical Information Systems (pp. 207-217). New York, NY: John Wiley & Sons.

Hornsby, K., & Egenhofer, M. J. (2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence, 36*(1-2), 177-194.

Iwase, S. & Saito, H. (2003). Tracking Soccer Players Based on Homography among Multiple Views. In Ebrahimi, T & Sikora, T. (Eds.), *Visual Communications and Image Processing* (pp. 288-292), Proceedings of SPIE.

Jones, P. W., & Smith, P. (2001). *Stochastic Processes: An introduction.* London, UK: Arnold Publishers.

Kwan, M.-P. (2000). Interactive geovisualization of activity-travel patterns using three dimensional geographical information systems: a methodological exploration with a large data set. Transportation Research Part C, 8(1-6), 185-203.

Laube, P., & Imfeld, S. (2002). Analyzing relative motion within groups of trackable moving point objects. In Egenhofer, M. J. & Mark, D. M. (Eds.), *Geographic Information Science, Second International Conference, GIScience 2002, Boulder, CO, USA, September 2002*, *LNCS 2478* (pp. 132–144). Berlin-Heidelberg: Springer.

Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering Relative Motion Patterns in Groups of Moving Point Objects. *International Journal of Geographical Information Science.19*(6), 639–668.

Laube, P., Van Kreveld, M., & Imfeld, S. (2004). Finding REMO - detecting relative motion patterns in geospatial lifelines. In Fisher, P. F. (Ed.), *Developments in Spatial Data Handling, Proceedings of the 11th International Symposium on Spatial Data Handling* (pp. 201-214), Berlin-Heidelberg: Springer.

Mark, D. M. (1998). Geospatial Lifelines. *Dagstuhl Seminars No. 98471, Integrating Spatial and Temporal Databases.*

Mark, D. M. (2003). Geographic Information Science: Defining the Field. In Duckham, M., Goodchild, M. F. & Worboys, M.F. (Eds.), *Foundations of Geographic Information Science* (pp. 3-18). London: Taylor and Francis.

Miller, H. J. & Han, J. (2001). Geographic data mining and knowledge discovery: An overview. In Miller, H. J. & Han, J., (Eds.), *Geographic data mining and knowledge discovery* (pp. 3–32). London: Taylor and Francis.

Miller, H.J. (2003). What about people in geographic information science?, *Computers, Environment and Urban Systems*, *27*(5), 447–453.

Nagel, K., Esser, J., & Rickert, M. (2000). Large-scale traffic simulations for transportation planning. In Stauffer, D. (Ed.) *Annual Reviews of Computational Physics VII* (pp. 151-202), World Scientific Publishing Company.

Padmanabhan , B. (2004). The Interestingness Paradox in Pattern Discovery. *Journal of Applied Statistics*, *31*(8), 1019-1035.

Pfoser, D., & Theodoridis, Y. (2003). Generating semantics-based trajectories of moving objects. *Computers, Environment and Urban Systems, 27*(3), 243–263.

Saglio, J.-M., & Moreira, J. (2001). Oporto: A realistic Scenario Generator for Moving Objects, *GeoInformatica, 5*(1), 71-93.

Silberschatz, A., & Tuzhilin, A. (1996). What Makes Patterns Interesting in Knowledge Discovery Systems. *IEEE Transactions on Knowledge and Data Engineering*, *8*(6), 970-974.

Turchin, P. (1998). *Quantitative Analysis of Movement: Measuring and Modelling Population Redistribution in Animals and Plants*, Sunderland, Massachusetts: Sinauer Publishers.

Wentz, E. A., Campell, A. F., & Houston, R. (2003). A comparison of two methods to create tracks of moving objects: linear weighted distance and constrained random walk. *International Journal of Geographical Information Science, 17*(7), 623–645.

Wood, J. (2002). *Java Programming for Spatial Sciences*. London: Taylor and Francis.

Figure captions

Fig. 1. **The REMO GKD approach**. The geospatial lifelines of four MPOs (a) are used to derive at regular intervals the motion azimuth (b). In the REMO analysis matrix consisting of classified motion attribute values (c) generic motion patterns such as *constancy*, *concurrence* or *trend-setter* are matched (d).

Fig.2. **Observed and simulated Porcupine Caribou data.** (a) Tracks of 10 individuals of PCH between March 2003 and December 2004. (b) One sample of r = 100 simulated populations of constrained random walks. The simulated tracks feature the same frequency distributions for step length and turning angle as the observed tracks in (a). Please note that the observed paths overlap considerably and thus appear much more densely packed than the simulated tracks starting at random positions in space.

Fig. 3. **Observed and simulated Swiss District data.** (a) Tracks of 185 Swiss political districts moving from 1981 to 2000 between the ideological poles left vs. right and ecological vs. technocratic. (b) One sample of r = 100 simulated populations of constrained random walks. As in figure 2 also in this plot the simulated tracks start at random locations in space. Note that both simulated and observed tracks maintain a preferred turning direction (see for example the subset of 12 districts in Figure 7), expressing curly shapes.

Fig. 4. **Direction change frequency distribution of PCH sample.** The radar plot illustrates an overall directional persistence for moving straight on (0°).

Fig. 5. **Direction change frequency distribution of caribou individual Blixen.** Individual direction change distributions can considerably diverge from the overall frequency distribution in Figure 4.

Fig. 6. **Direction change frequency distribution of total Swiss political districts sample.** The direction changes of the moving districts are much more concentrated around 0 (moving straight on), and are thus not plotted in a radar plot.

Fig. 7. **Direction change frequency distribution of districts of Canton Zurich.** This subset of 12 districts illustrates the diversity in the direction change distribution among individual MPOs.

Fig. 8. **The next step.** For the construction of any step at t+1 two samples from the constraining distributions are needed. First, a step length s_{t+1} and second a direction change angle a_{t+1}

Fig. 9. **Monte Carlo simulations of constrained random walks.** The constraints' *step length distribution* (stepLenDist) and *direction change distribution* (dirChngeDist) for the random walk are first derived from an observation data set consisting of the trajectories of n MPOs. Every of the $r=100$ Monte Carlo runs generates thereafter a synthetic data set of the constrained random walks of n MPOs.

Fig. 10. **Numerical experiments.** The numerical experiments repeat the pattern matching process (pm) for all r synthetic data sets systematically varying the pattern extent p and the attribute classification c. For every pattern matching run the number of found patterns and the number of possible patterns is stored for subsequent statistical analysis.

Fig. 11. **Detecting motion azimuth constancy patterns in PCH.** The success of the pattern matching process is very similar for the observed and the simulated data, rapidly declining with increasing pattern length *p*.

Fig. 12. **Detecting motion azimuth concurrence patterns in PCH.** The obvious deviation of the found/possible ratio for the observed and the simulated data allows to assign interestingness to patterns. The greater the difference between the ratios for observed and simulated patterns, the higher the assigned interestingness.

Fig. 13. **Detecting motion azimuth trend-setter patterns in PCH.** Both plots are based on $c = 8$ motion attribute classes, the difference lies in the length of the required anticipation of the trendsetter, 3 and 4 time steps, respectively. Generally, the found/possible ratio of trend-setter lies below the results for constancy and concurrence, for both observed and simulated. This is not surprising, keeping in mind that the requirements for complex patterns such as trendsetter are much higher. However, observed lies considerably above simulated, allowing to assign interstingness to found trendsetter patterns.

Fig. 14. **Detecting speed constancy patterns in PCH.** The variance of the ratio found/possible for the simulated data is much larger than in most plots referring to motion azimuth. However, the observed value lies in the variance range and does therefore not allow to assign interestingness to found speed patterns.

Fig. 15. **Detecting speed concurrence patterns in PCH.** Just like constancy, also concurrence does not show considerable deviation between the observed and the simulated values.

Fig. 16. **Detecting motion direction concurrence in the Swiss political districts.** For both attribute granularities significantly more patterns are found than expected from the simulation, after a threshold given by the crossing of both graphs. The sharp jump in both graphs is an artefact caused by the computation of the ratio found/possible.

Fig. 17. **REMO matrix for PCH motion azimuth.** Rows represent caribou individuals, columns represent time steps ∂t two weeks. The $c = 8$ azimuth classes are colour coded, using a circular colour ramp ranging from green (N), blue (E), orange (S), to yellow (W).

Fig. 18. **Concurrence matches in PCH.** This pattern detection session searched for concurrence patterns including at least 5 caribou. Such patterns could be found in both northward spring migration and southward fall migration.

Fig. 19. **Trend-setter matches in PCH.** Trend-setting caribou individuals anticipating 3 time steps in advance the seasonal migration of at least 5 followers.

Fig. 20. **REMO matrix for PCH speed.** As a general overview the speed matrix illustrates two fast (blue) migration periods each season in 2003 and 2004, respectively, with intermitted slower sedentary periods.

Fig. 21. **REMO matrix for Swiss political districts motion azimuth.** Rows represent 185 districts, columns represent 20 years from 1981 until 2000. The $c = 8$ azimuth classes are colour coded, using a circular colour ramp ranging from green (N), blue (E), orange (S), to yellow (W).

Fig. 22. **Concurrence matches in Swiss political districts motion azimuth.** This pattern detection session searched for concurrence patterns including at least 45 concurrently moving districts. Two sequences of several concurrence patterns could be found during this pattern detection session.

Figures

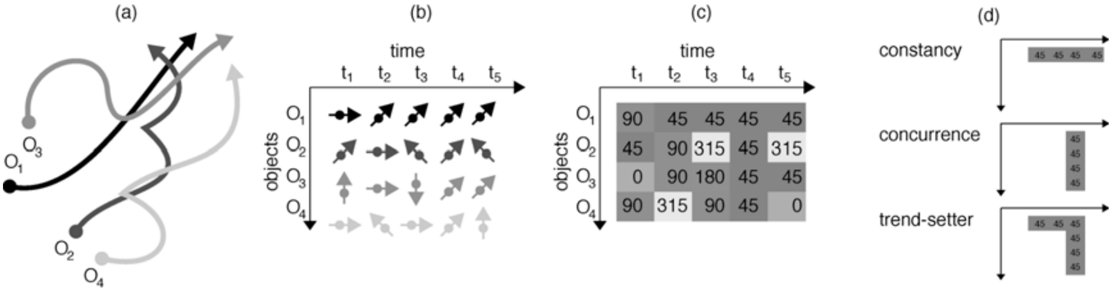


Fig. 1.

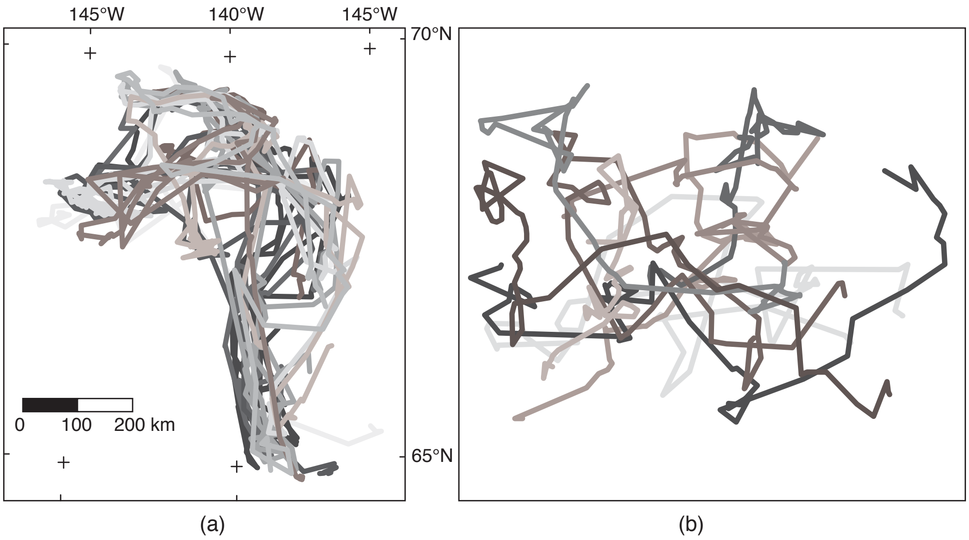


Fig. 2.

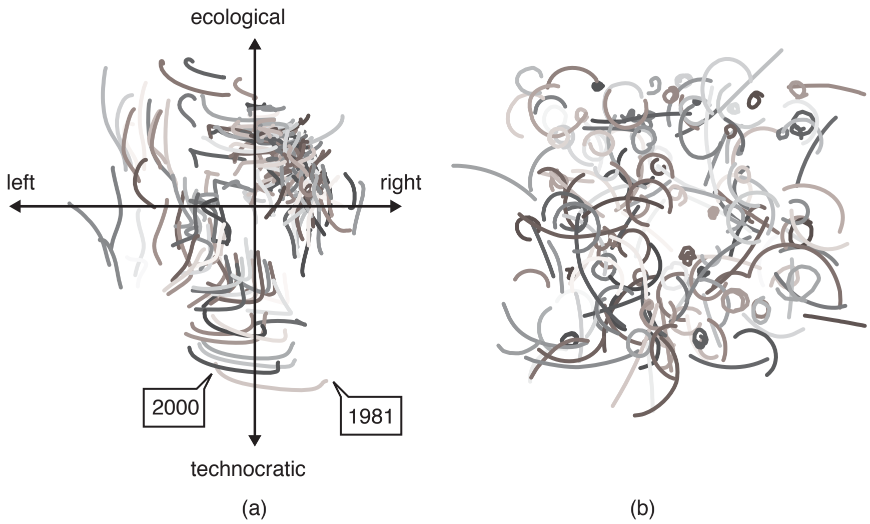


Fig. 3.

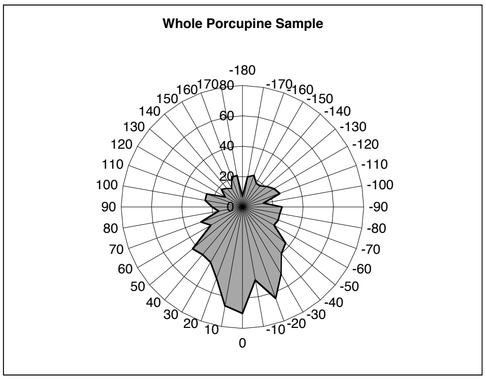


Fig. 4

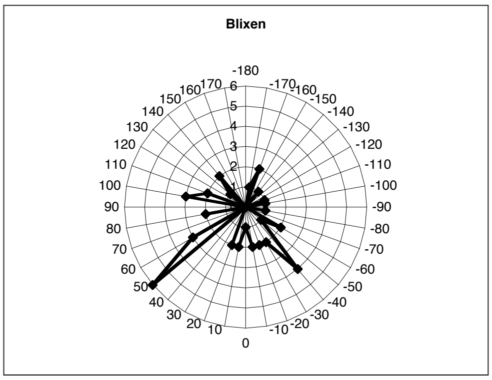


Fig. 5.

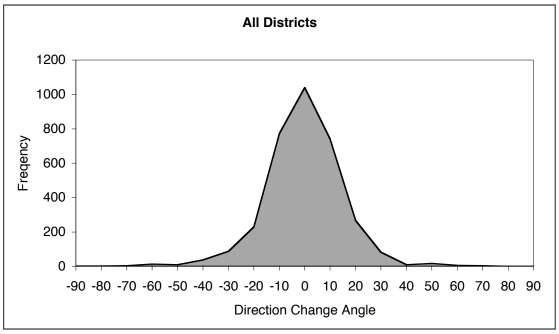


Fig. 6.

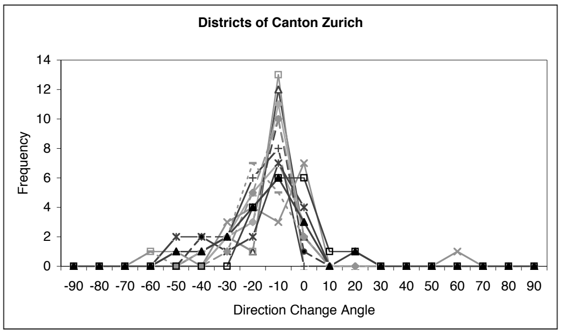


Fig. 7.

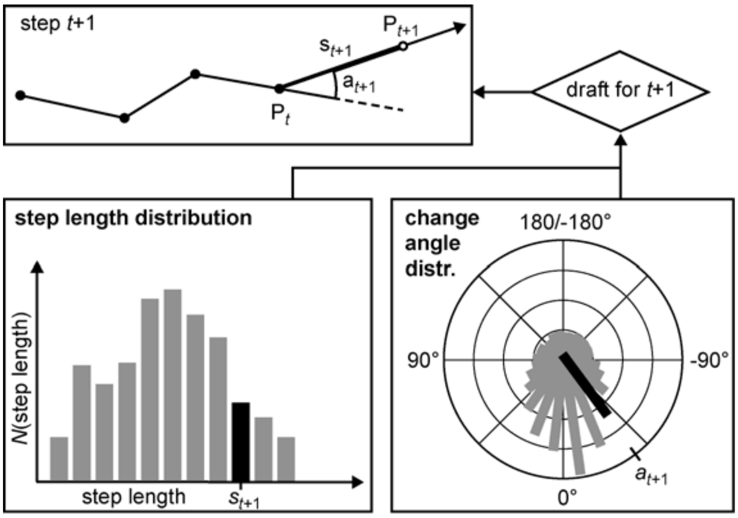


Fig. 8

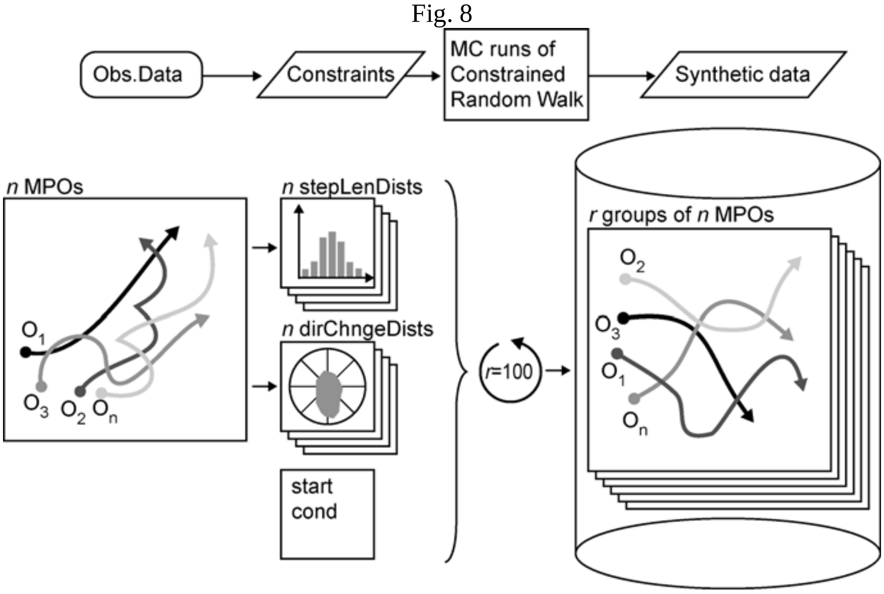


Fig. 9

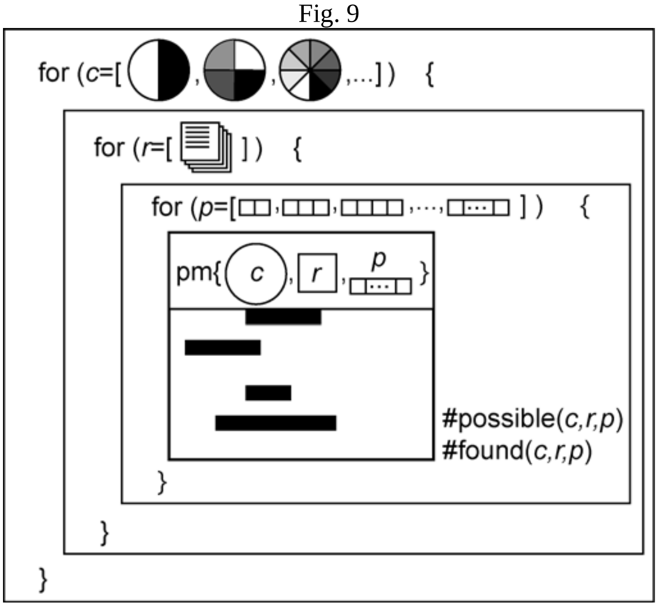


Fig. 10

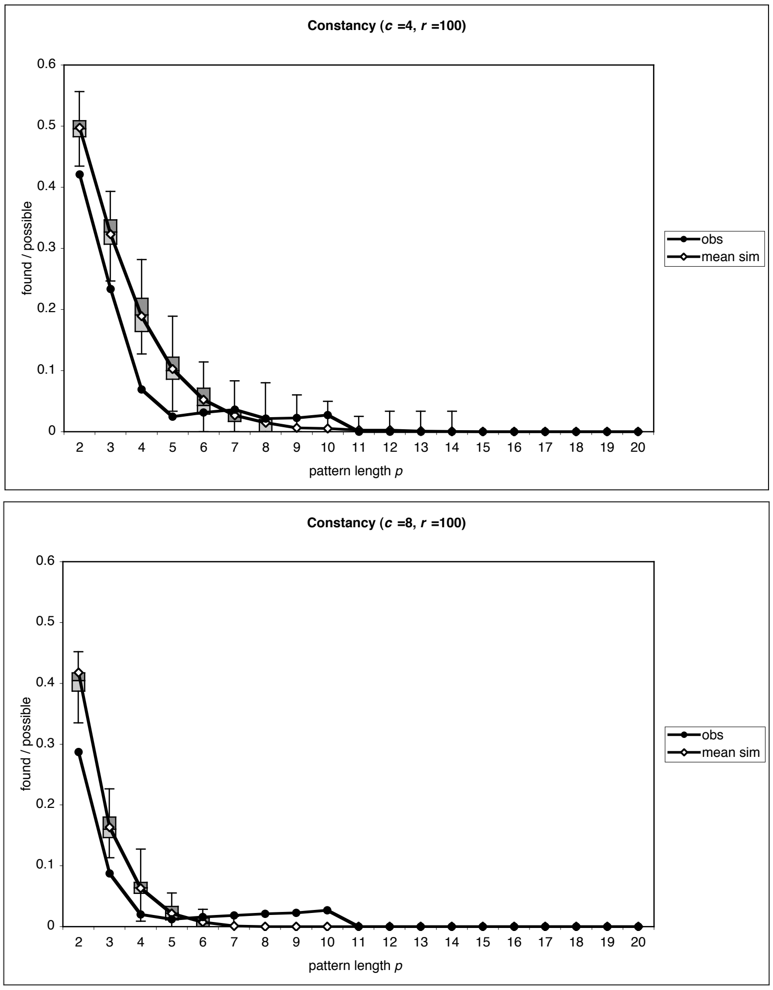

Constancy (c =4, r =100)
found / possible
0.6
0.5
0.4
0.3
0.2
0.1
0
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
pattern length p
obs
mean sim
Constancy (c =8, r =100)
found / possible
0.6
0.5
0.4
0.3
0.2
0.1
0
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
pattern length p
obs
mean sim

Fig. 11

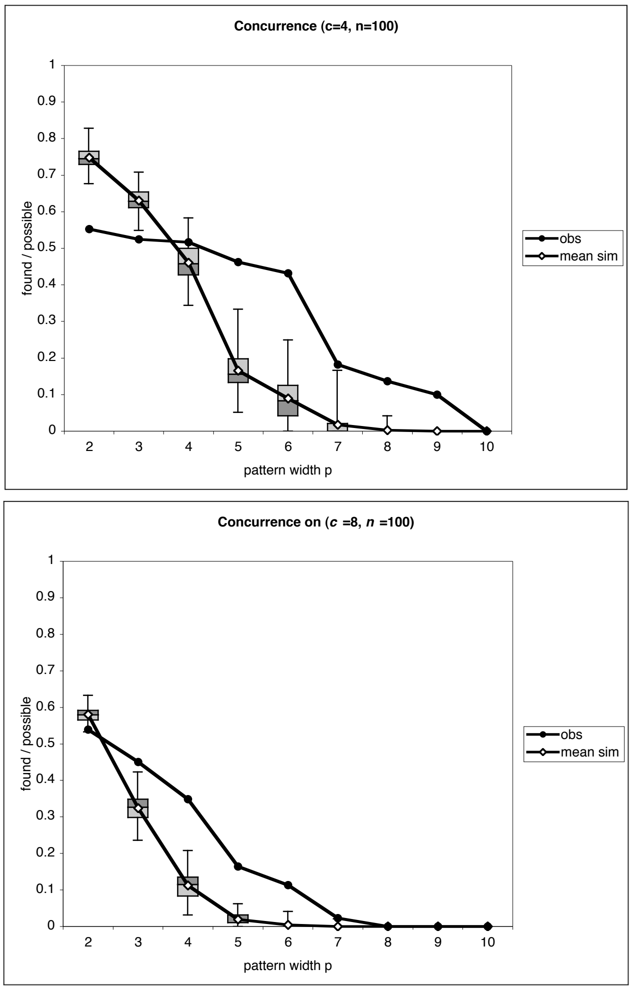
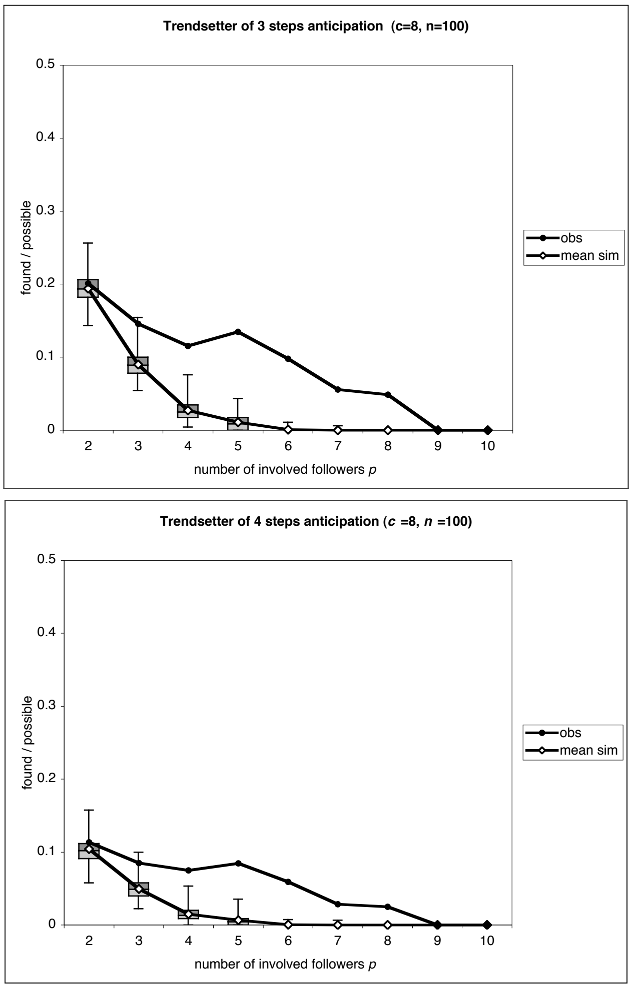

Concurrence (c=4, n=100)
found / possible
pattern width p
obs
mean sim
Concurrence on (c =8, n =100)
found / possible
pattern width p
obs
mean sim

Fig. 12



Trendsetter of 3 steps anticipation (c=8, n=100)
found / possible
0.5
0.4
0.3
0.2
0.1
0
2
3
4
5
6
7
8
9
10
number of involved followers p
obs
mean sim
Trendsetter of 4 steps anticipation (c =8, n =100)
found / possible
0.5
0.4
0.3
0.2
0.1
0
2
3
4
5
6
7
8
9
10
number of involved followers p
obs
mean sim

Fig. 13

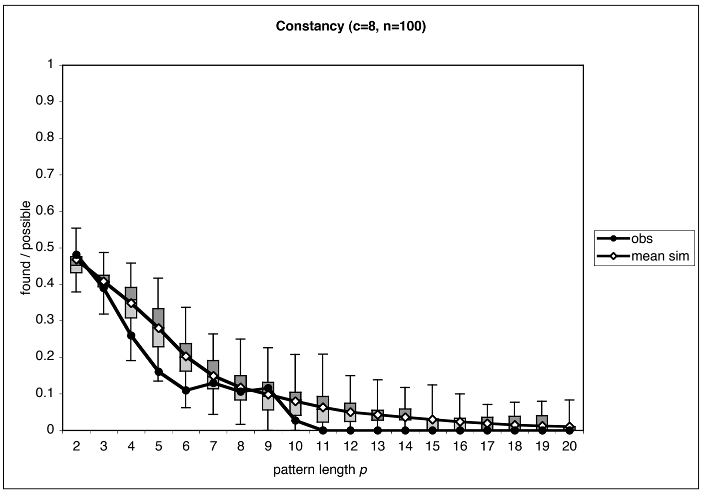

Constancy (c=8, n=100)
found / possible
pattern length p
obs
mean sim
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Fig. 14

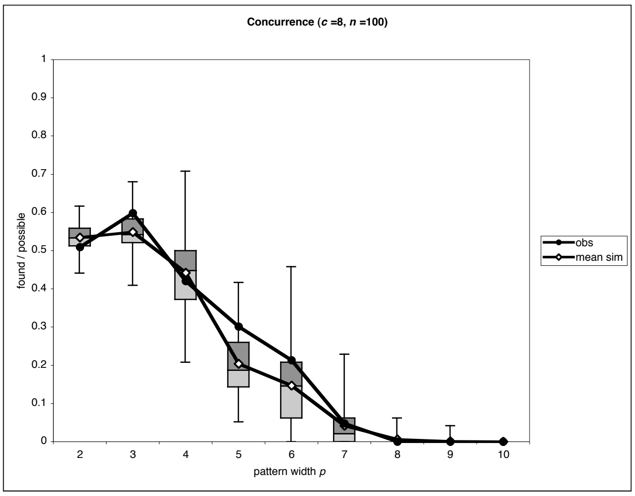

Concurrence (c =8, n =100)
found / possible
pattern width p
obs
mean sim
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2 3 4 5 6 7 8 9 10

Fig. 15

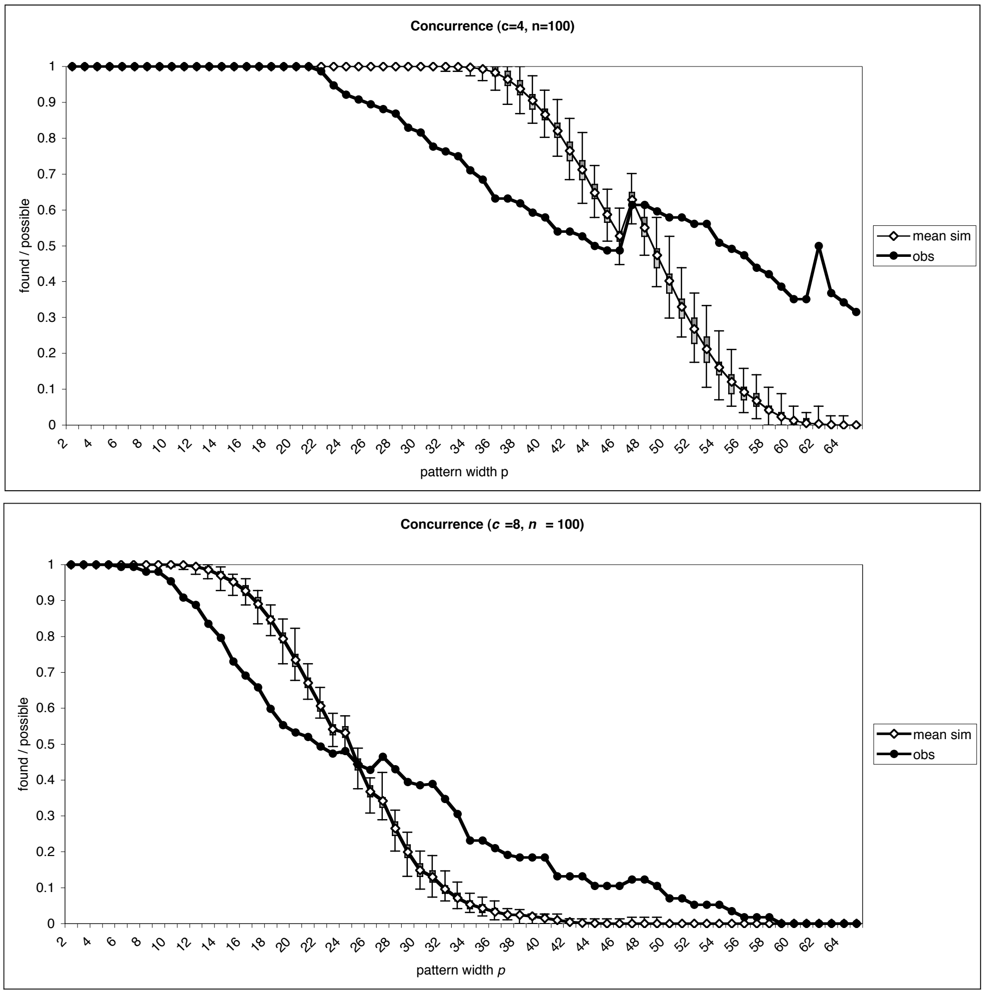

Concurrence (c=4, n=100)
found / possible
pattern width p
mean sim
obs
Concurrence (c =8, n = 100)
found / possible
pattern width p
mean sim
obs

Fig. 16

Fig. 17

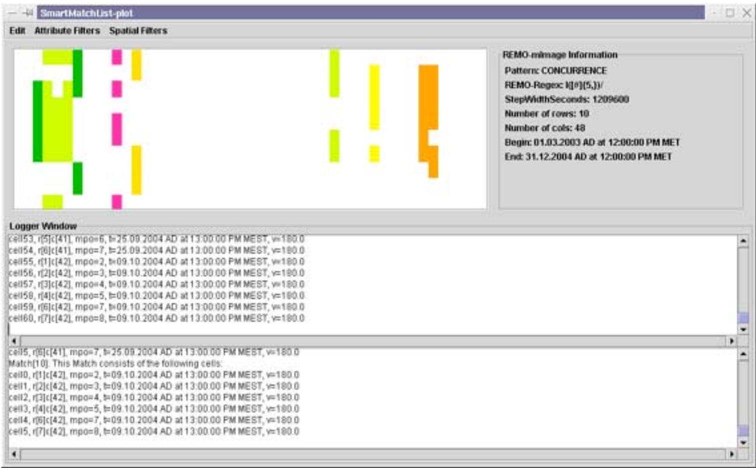


Fig. 18

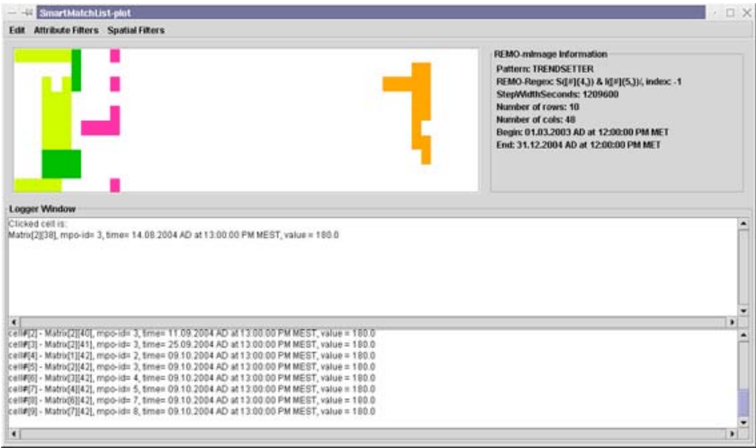


Fig. 19

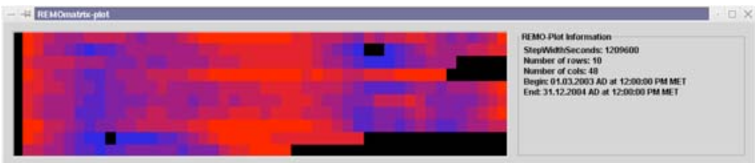


Fig. 20

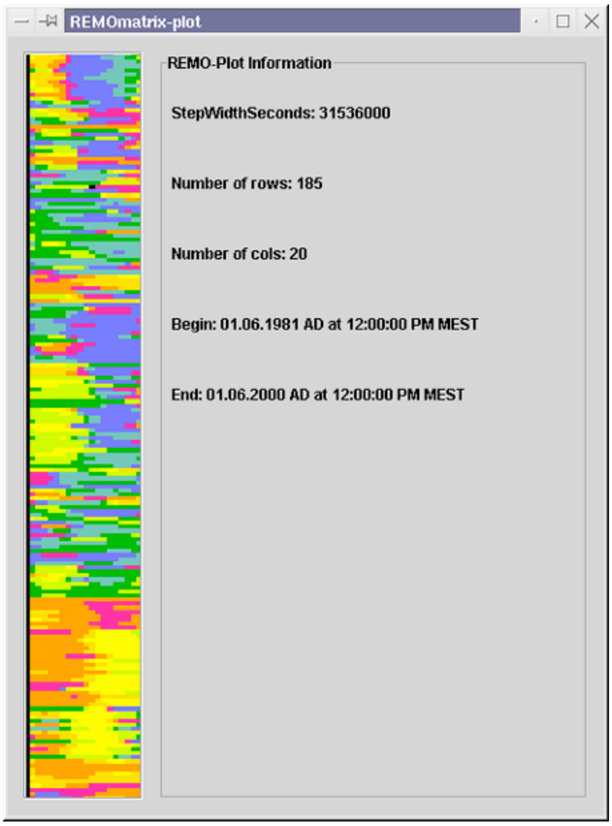

REMOmatrix-plot
REMO-Plot Information
StepWidthSeconds: 31536000
Number of rows: 185
Number of cols: 20
Begin: 01.06.1981 AD at 12:00:00 PM MEST
End: 01.06.2000 AD at 12:00:00 PM MEST

Fig. 21

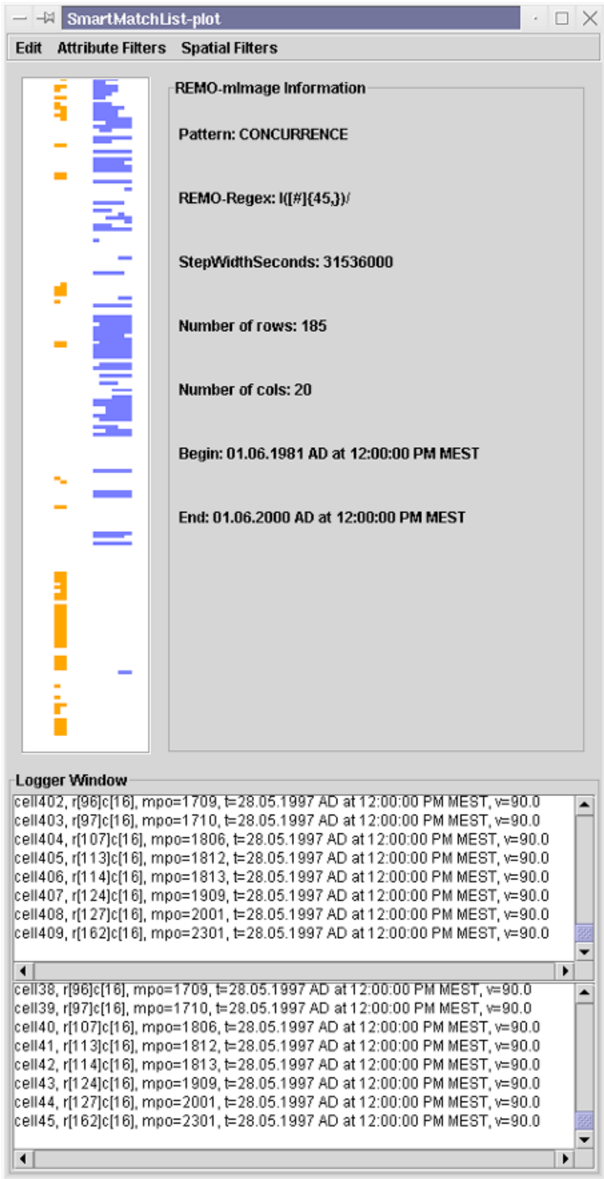

SmartMatchList-plot
Edit
Attribute Filters
Spatial Filters
REMO-mImage Information
Pattern: CONCURRENCE
REMO-Regex: l([#]{45,})/
StepWidthSeconds: 31536000
Number of rows: 185
Number of cols: 20
Begin: 01.06.1981 AD at 12:00:00 PM MEST
End: 01.06.2000 AD at 12:00:00 PM MEST
Logger Window
cell402, r[96]c[16], mpo=1709, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell403, r[97]c[16], mpo=1710, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell404, r[107]c[16], mpo=1806, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell405, r[113]c[16], mpo=1812, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell406, r[114]c[16], mpo=1813, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell407, r[124]c[16], mpo=1909, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell408, r[127]c[16], mpo=2001, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell409, r[162]c[16], mpo=2301, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell38, r[96]c[16], mpo=1709, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell39, r[97]c[16], mpo=1710, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell40, r[107]c[16], mpo=1806, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell41, r[113]c[16], mpo=1812, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell42, r[114]c[16], mpo=1813, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell43, r[124]c[16], mpo=1909, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell44, r[127]c[16], mpo=2001, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0
cell45, r[162]c[16], mpo=2301, t=28.05.1997 AD at 12:00:00 PM MEST, v=90.0

Fig. 22