



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Vrotsou, K., Janetzko, H., Navarra, C., Fuchs, G., Spretke, D., Mansmann, F., Andrienko, N. & Andrienko, G. (2015). SimpliFly: A Methodology for Simplification and Thematic Enhancement of Trajectories.. *IEEE Transactions on Visualization and Computer Graphics*, 21(1), pp. 107-121. doi: 10.1109/tvcg.2014.2337333

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/14164/>

**Link to published version:** <https://doi.org/10.1109/tvcg.2014.2337333>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

City Research Online:

<http://openaccess.city.ac.uk/>

[publications@city.ac.uk](mailto:publications@city.ac.uk)

---

# SimpliFly: A Methodology for Simplification and Thematic Enhancement of Trajectories

Katerina Vrotsou, Halldor Janetzko, Carlo Navarra, Georg Fuchs, David Spretke, Florian Mansmann, Natalia Andrienko, and Gennady Andrienko

**Abstract**—Movement datasets collected using today's advanced tracking devices consist of complex trajectories in terms of length, shape, and number of recorded positions. Multiple additional attributes characterizing the movement and its environment are often also included making the level of complexity even higher. Simplification of trajectories can improve the visibility of relevant information by reducing less relevant details while maintaining important movement patterns. We propose a systematic stepwise methodology for simplifying and thematically enhancing trajectories in order to support their visual analysis. The methodology is applied iteratively and is composed of: (a) a simplification step applied to reduce the morphological complexity of the trajectories, (b) a thematic enhancement step which aims at accentuating patterns of movement, and (c) the representation and interactive exploration of the results in order to make interpretations of the findings and further refinement to the simplification and enhancement process. We illustrate our methodology through an analysis example of two different types of tracks, aircraft and pedestrian movement.

**Index Terms**—Visual analysis, trajectories, simplification, thematic enhancement, clustering.



## 1 INTRODUCTION

TECHNOLOGICAL advances in storage capacity and data collection methods are leading to huge, multivariate movement datasets that are challenging to analyse. In such datasets a trajectory represents the movement of an object, and is defined by a sequence of recorded positions each comprised of at least its geographic coordinates and a timestamp. Trajectories collected through GPS and advanced tracking devices typically sample the movement with very high frequency. Apart from position and time, additional information may be collected or computed concerning the characteristics of the movement (speed, acceleration, course, turn rate, etc.), the characteristics of the moving objects (e.g. physical conditions), and the environment of the movement (e.g. weather).

As a result, trajectories become large in terms of the number of points defining them, and in addition the information available at each such point is multivariate and may be measured at a very high level of detail. Often the information load becomes overwhelming. Visually analysing such datasets is difficult due to: (a) perceptual limitations – tiny elements or small differences are hard to distinguish; (b) cognitive limitations – the user cannot keep the full information in mind and needs to focus on relevant items, which may be hard to find amongst all the details; and possibly even (c) performance issues – rendering times for many, complex trajectory elements may impede interactivity. Hence, simplification is needed to improve the visibility of relevant information by omitting

unnecessary details while maintaining the overall context and letting the user focus on the important data features.

A primary focus in the analysis of movement data is the identification and exploration of movement patterns. Movement patterns, in this context, are representations of behaviour [1] of single or groups of moving objects, or relations [2] within and between the objects themselves and/or the characteristics of their movement. As such, movement patterns are elusive to define and highly context dependent [3]. There has been increasing research work concerned with defining and categorizing movement patterns and appropriate methods for revealing such [1], [2], [3], [4], [5]. In this work we wish to achieve a flexible definition of movement patterns by allowing analysts to select characteristics of movement that are interesting to them in the context of their analysis. A pattern is then identified through similarity of these characteristics. We are using a visual analytics approach for revealing such patterns by combining data mining and visualization techniques.

The core contribution of this paper is the establishment of a systematic methodology for visually analysing trajectories based on simplification and thematic enhancement. Within this methodology we propose two new trajectory simplification methods. Our approach consists of three interconnected steps:

- 1) *Trajectory simplification* is applied to reduce the complexity of the trajectory structure. We consider three different task-dependent simplification types and introduce new methods for two of them (Section 3).
- 2) *Thematic enhancement* is applied to the simplified trajectories to accentuate patterns with respect to attributes characterizing the trajectory segments (Section 4).
- 3) *Representation and exploration* techniques are used to display and interact with the processed trajectories in two and three dimensions (2D and 3D resp.) facilitating their analysis and providing feedback for further refinement of the simplification and enhancement (Section 5).

- Katerina Vrotsou and Carlo Navarra are with Linköping University, Sweden, E-mail: [firstname.lastname@liu.se](mailto:firstname.lastname@liu.se).
- Halldor Janetzko, David Spretke and Florian Mansmann are with University of Konstanz, Germany, E-mail: [firstname.lastname@uni-konstanz.de](mailto:firstname.lastname@uni-konstanz.de).
- Georg Fuchs, Natalia Andrienko and Gennady Andrienko are with University of Bonn & Fraunhofer IAIS, Germany, E-mail: [firstname.lastname@iais.fraunhofer.de](mailto:firstname.lastname@iais.fraunhofer.de).

We aim to achieve a more flexible exploratory analysis situation in which the analyst has the opportunity to interact with complex information in a simplified way. In order to preserve relevant movement patterns, we want to simplify the trajectories not only based on their shapes and other geometric properties, but with respect to various attributes characterizing the movement; both structural and thematic.

Our paper presents a generic methodology, which can be implemented in various ways, rather than a particular system or toolkit. To keep to the focus of the paper, we avoid presenting the user interface specifics, performance metrics, and other implementation-specific details.

## 2 RELATED WORK

Research in the domain of movement data analysis is extensive and continuously increasing, summarizing examples include [2], [6], [7]. In the scope of this paper, we limit our discussion to research directly related to the three steps of our presented methodology: trajectory simplification, thematic enhancement and representation and exploration approaches.

### 2.1 Trajectory simplification

Research in the geographic community suggests a number of approaches to geometric simplification (e.g., [8], [9], [10]) and smoothing (e.g., [11]) of lines for generalizing angular lines (e.g., for canals or buildings), sinuous lines (e.g., for hydrography) and contour lines (e.g., for reliefs) in maps, especially when drawing these maps on different scales. Laube and Purves [12] studied the effects of point sampling for various scales based on high-resolution animal movement trajectories. While these methods only simplify the lines without distorting the space, there is a group of methods involving space distortion. LineDrive [13], for example, displays route maps in such a way that the areas with complex driving directions are enlarged whereas long highway sections are shrunk. Most of the remaining road network is not displayed. Line schematization uses predefined sets of angles for representing paths (cf. [14]).

Our research focuses on line simplification without space distortion. A very well known approach among geometry-based simplification is the Douglas-Peucker algorithm [8]. Despite its limitations [15] the approach is widely used and we, therefore, choose it a representative of this group of simplification methods. As an addition to such conventional geometric simplification, we suggest two new approaches, density-based and property-based simplification, so that the most suitable approach can be chosen depending on the data characteristics and analysis tasks.

### 2.2 Thematic enhancement

Thematic enhancement of trajectories is based on dividing trajectories into *episodes*. Dykes and Mountain [16] define movement episodes as fragments of trajectories where movement characteristics (speed, direction, sinuosity, etc.) are relatively constant whereas a significant change indicates the beginning of a new episode. Partitioning of trajectories into episodes is conventionally called trajectory segmentation.

Buchin et al. [17] discuss different spatio-temporal criteria for segmenting trajectories. A wide variety of segmentation methods has been developed to address such criteria while giving substantial attention to performance (e.g., [18], [19]). Trajectory segmentation is commonly done based solely on movement characteristics and does not consider other kinds of position-related attributes. There has been, however, an increasing amount of research concentrating on the importance of considering surrounding, contextual, and application dependent information in mobility analysis, thus promoting the analysis of semantically enriched trajectories instead of simply raw movement. This work is summarized in Parent et al. [5]. In our work, we approach such semantic enrichment by supporting the segmentation of trajectories with regard to attributes that are deemed relevant to the analysis goals. Hence, the notion of episodes is generalized to arbitrary position-related attributes.

The results of trajectory segmentation are used for further analysis of movement data. Laube et al. [20] search for similar episodes in trajectories of several moving objects for detecting particular patterns of relative movements. Dodge et al. [21] analyse characteristics of trajectory episodes to classify the movement modes. Sester et al. [22] segment trajectories into a graph reflecting typical behaviour and identify unusual behaviour based on it.

Our thematic enhancement approach uses clustering of episodes based on multiple attributes characterizing them in order to reveal movement patterns. Clustering of trajectory fragments has previously been done for different purposes. Lee et al. [23] introduce a partition-and-group clustering framework to enable the discovery of common sub-trajectories in multiple trajectories. A similar approach using incremental clustering is provided in [24]. For the classification of trajectories according to their properties, Lee et al. [25] propose a framework that generates a hierarchy of discriminative features. Our clustering differs from these methods in that it is applied to arbitrary attributes chosen according to the analysis goals and is used for supporting visual exploration of movement patterns.

Movement patterns can be extracted from trajectories using purely automatic methods, visual exploration approaches or combinations of both [2], [3], [5]. Among these, an advantage of visual exploration is that it does not require each kind of potentially interesting pattern to be formally and precisely defined and the search algorithm to be developed and implemented. Instead, interesting attributes characterising the movement can be selected and the behaviour of the data with respect to these ad-hoc queries can be observed. Thus, visual exploration can allow the analyst to discover new, previously unknown or unexpected kinds of movement patterns.

### 2.3 Representation and exploration

The most common techniques for visual representation of discrete objects' movement in a spatial context are static and animated maps [27] and interactive space-time cubes [28], [29] with linear symbols representing trajectories. Movement in three-dimensional space, e.g., in the air or under water, are harder to visualize than movements in 2D. Ware et al. [30] represent a trajectory of a whale by a 3D ribbon in

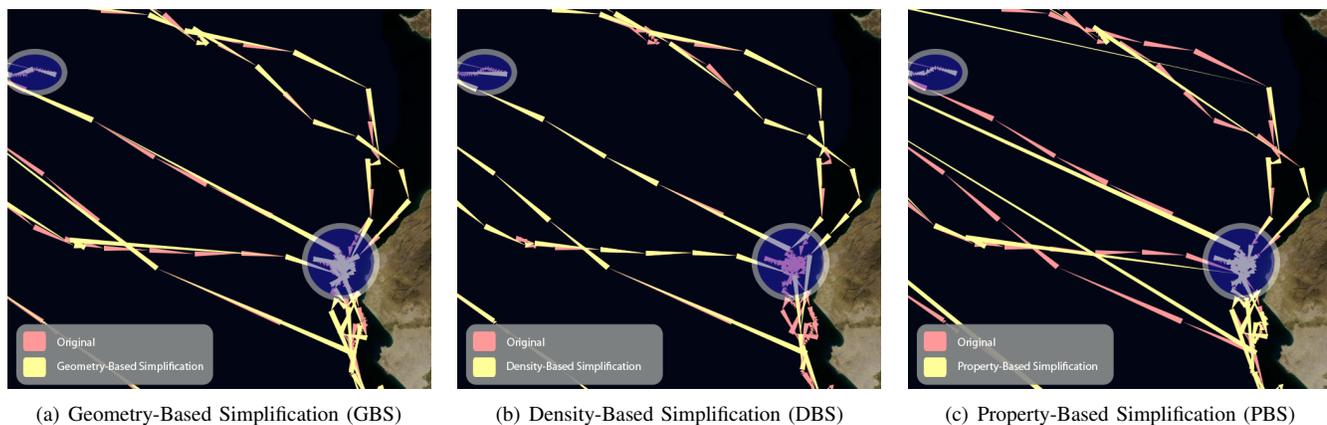


Fig. 1. Galapagos albatross trajectory simplified by three simplification approaches and represented using tapered segments [26]. (a) Geometry-Based Simplification (using the Douglas-Peucker algorithm with  $\epsilon = 0.05^\circ$ ) achieves a good overall fit to the original trajectory. (b) Density-Based Simplification (with  $\epsilon = 0.01^\circ$ ) precisely models the long migration segments and simplifies dense regions so that the path of the albatross can be nicely followed after entering the marked dense region on the right. (c) Property-Based Simplification (with upper bound  $k(N) = 50\%$ ) merges segments having similar attribute properties while preserves detail where segments display change in properties.

a perspective view with glyphs on its surface showing the direction of the movement. Hurter et al. [31] represent multiple trajectories of aircraft in horizontal or vertical 2D projections with animated transitions from one projection to another.

Since these techniques are not scalable to a large number of trajectories nor to complex trajectories, various aggregation-based methods are intensively used: temporal histograms, traffic density surfaces, accessibility surfaces [16], [32], [33], flow maps [34], [35], [36], [37], and spatially ordered treemaps [38]. A comprehensive review of the existing ways to aggregate movement data and techniques for visually displaying the results is provided in [2]. However, aggregation across multiple trajectories is not a suitable approach when the task is to explore movement patterns occurring *within* trajectories. We perform aggregation individually within each trajectory by uniting consecutive trajectory segments into episodes and deriving aggregate characteristics of the episodes from the original data.

There are relatively few methods for analyzing relations between moving objects and elements of the spatiotemporal context of the movement. The dynamics of the distances of moving objects to selected locations is computed and visualized in [39]. Lundblad et al. [40] attach data about weather conditions to positions of ships and display the data on interactive linked views. Fedak et. al. [41] represent movement of sea animals together with the context information such as sea surface temperature, shoreline, and ocean depth.

Our methodology allows the use of any visualization techniques suitable for representing individual trajectories, i.e., not involving aggregation across multiple trajectories. The geometries of trajectories can be represented by lines, ribbons, tubes, or sequences of glyphs in a two- or three-dimensional display. Thematic characteristics of trajectory episodes can be encoded by colours of the corresponding parts of these geometric representations or by superimposed symbols (glyphs).

### 3 TRAJECTORY SIMPLIFICATION

The initial step of our approach involves the morphological simplification of trajectories. Each trajectory is simplified separately and the result of such simplification is a trajectory that is comprised of fewer segments than the original one by aggregating consecutive data points defining these segments. The primary reason for this is to reduce the number of segments constituting the trajectories so as to bring forward their salient characteristics without unnecessary, potentially distracting detail; secondarily, these simplifications in principle reduce rendering and interaction response times, allowing an analyst to fluently interact with even larger datasets. Depending on the goals of analysis, we distinguish three types of simplification tasks:

- 3.1 *Geometry-based simplification (GBS)*. The main goal is to convey the overall shape of a trajectory. GBS only considers (geographic) vertex coordinates and disregards thematic attributes associated with trajectory segments.
- 3.2 *Density-based simplification (DBS)*. The goal is to show segment-related information with as much detail as possible while meeting limitations of screen resolution and perception. This means areas of high data point density (in view space) are simplified to yield fewer aggregate segments, whereas in less dense areas the original trajectory data points are represented directly. Generally, the higher the on-screen data point density the more simplification is needed. This type of simplification focuses on *where* a moving object has been - general locations visited, predominant movement direction, etc. - and results in an aggregated trajectory showing the overall route followed.
- 3.3 *Property-based simplification (PBS)*. The goal is to enable detection and exploration of relevant movement patterns in terms of spatial and thematic attributes. A trajectory is simplified by dividing it into homogeneous

parts in terms of attributes of interest. The approach includes a trade-off between degree of simplification and degree of internal homogeneity. This type of simplification focuses on the *how* a moving object has moved, and aims at simplifying without losing relevant multi-attribute semantic movement patterns.

The following Sections 3.1-3.3 describe approaches corresponding to each of the defined simplification tasks. Examples of these can be seen in Figure 1 where Galapagos albatross flight trajectories have been simplified.

### 3.1 Geometry-based simplification approach

Geometry-based simplification approaches aim at detecting the most significant points of a shape and preserving them while omitting less important points. A popular representative of this group of methods is the Douglas-Peucker algorithm [8] for line simplification. The algorithm first finds the farthest point from the straight line connecting the first and last point of the line that needs to be simplified. This point is used to split the line into two parts. Then the Douglas-Peucker algorithm is recursively applied to each part until there is no point whose distance to the straight connecting line is higher than a given threshold. There are variants of the Douglas-Peucker algorithm that take into account additional information and give more weight to some points than to others, e.g., time spent in a location. Points with low weights can be discarded by applying a user-defined threshold.

### 3.2 Density-based simplification approach

The purpose of density-based simplification (DBS) is to remove detail from trajectory portions that exceed screen resolution and/or perceptual limits: dense clusters of data points can, in principle, be replaced by a single cluster representative. This kind of simplification is, hence, an inherently viewpoint-dependent operation based on screen resolution and the current zoom level (in 2D) or virtual camera position (in 3D). Smooth interaction with the representation requires its application at interactive frame rates. The naïve approach of density-based clustering directly in screen space is thus hardly feasible as it implies re-calculating clusters for all visible trajectories based on changed node densities for every update of the viewpoint. We therefore propose a more efficient technique using real world/object coordinate-based densities.

The key idea here is to capture the relative densities of trajectory points in object space once in a clustering preprocess, and as the viewpoint changes these cluster results are transformed into screen space to obtain absolute (pixel-based) densities, which is computationally far less demanding. Figure 2 shows a schematic overview of this approach. Its individual steps are detailed in the following subsections.

#### 3.2.1 Clustering in data space

The primary notion of density-based clustering is that of so-called core objects that have at least  $MinPoints$  neighbouring points within a maximum neighbourhood distance threshold  $\epsilon$  according to a defined distance metric. Points that are in the neighbourhood of at least one core point are called density-reachable. Core points and density-reachable points constitute

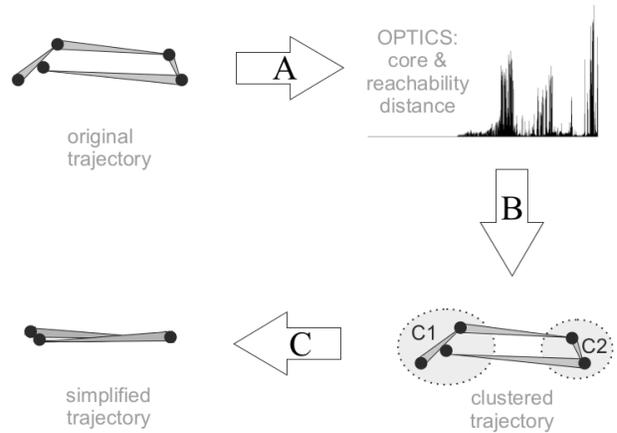


Fig. 2. Schematic depiction of the density-based simplification approach. The input trajectory points are ordered into a core-reachability distance plot using OPTICS (A). From this plot a cluster partition is extracted depending on current on-screen size (B). Finally the clustered data points are processed sequentially to generate a simplified trajectory from cluster representatives (C).

dense regions or clusters, whereas other points are considered noise not belonging to any cluster. Both  $MinPoints$  and  $\epsilon$  are user-selected input parameters to the clustering process [42].

OPTICS [43] is an extension to this general density-based approach that applies a specific sorting method to the input points prior to clustering. Beginning with an arbitrary core point, it first builds a core-reachability distance plot for a given value of  $MinPoints$  and a maximum distance value,  $d_{max}$ . From this plot cluster partitions of the point set can be extracted for different neighbourhood distance thresholds,  $\epsilon \leq d_{max}$ .

We apply OPTICS sorting, as a one-time preprocessing step, to generate a core-reachability distance plot for a trajectory’s data points (Figure 2A). The plot is built for a neighbourhood size  $MinPoints = 2$ . This allows extraction of clusters of only two trajectory points as the smallest possible simplification step.  $d_{max}$  is chosen as the length of the trajectory’s bounding box diagonal. Density clustering the set of trajectory data points with this neighbourhood distance threshold guarantees that for any one point all other points are within its neighbourhood and thus, all belong to a single cluster. Therefore, the highest degree of simplification reached by our DBS approach is to collapse the entire trajectory into a single point.

#### 3.2.2 Clustering result transformation

To arrive at an actual, screen resolution-dependent clustering of a trajectory’s data points, the core-reachability distance plot of that trajectory is evaluated for a specific distance value,  $\epsilon$ , according to the OPTICS algorithm [43]. This allows selection of an overall simplification level for the trajectory: higher values of  $\epsilon$  result in larger and fewer clusters being extracted from the plot (and thus, a more simplified trajectory representation), whereas smaller values of  $\epsilon$  generate smaller and more clusters retaining more of the original trajectory’s details. Note that the evaluation of the core-reachability distance plot has linear complexity with respect to the number of trajectory points.

In our approach,  $\epsilon$  is determined based on the width in pixels of the primitive used to represent trajectory segments (e.g. simple lines or triangles, cf. top-left Figure 2). For this, an inverse projection of the pixel diagonal length at the current zoom level into data coordinate space is performed. For 2D maps  $\epsilon$  is determined by finding the geographic distance covered by the diagonal of a pixel in the map representation multiplied by the primitive width in pixels, thus yielding  $\epsilon$  as the primitives' width in data coordinate space. This results in segments of the simplified trajectory that are at least as long as they are wide in screen space. Shorter segments would only add variations in the trajectory's path which are hard or even impossible to perceive, since the corresponding bends between segments would be masked by the resulting overplot.

### 3.2.3 Simplification

The final step is to derive a simplified trajectory representation from the view- and resolution-dependent data point clusters by replacing each cluster by its representative point, thus reducing the number of trajectory line segments (Figure 2C). In our approach we use the arithmetic mean point of a cluster.

However, a trajectory is a temporally ordered sequence of points, whereas data points are clustered only with respect to their spatial positions. This creates the risk of the temporal aspect being lost as points from temporally disjoint trajectory segments may be assigned to the same cluster (cf. cluster  $C_1$  in the lower-right of Figure 2). In order to avoid losing the temporality of the trajectories we address this risk explicitly by sequentially processing the trajectory data points from beginning to end. Consecutive points assigned the same cluster ID are aggregated to obtain their mean representative, and as the next ID in sequence is encountered, the current simplified trajectory segment is finalized. Thus, if a cluster ID is re-encountered later in a trajectory the corresponding points from the cluster are associated with a different segment. Figure 2C shows an example for this: points from cluster  $C_1$  are aggregated into two distinct vertices for the simplified trajectory.

Also note that singleton data points, or 'noise' in the selected cluster partition, are never merged with any other trajectory points, because they are sufficiently far from any other location at the current simplification level. On the one hand, there is no need in merging since these points do not contribute to clutter and overplotting; on the other hand, these point may convey important trajectory shape and object location information that should be preserved.

## 3.3 Property-based simplification approach

The property-based simplification approach (PBS) makes use of the regionalization method proposed in [44], [45] to perform structural simplification of trajectories.

Regionalization aims to aggregate a large set of spatial objects into a smaller set of spatially contiguous regions that share similar attributes by optimizing an objective function. Guo [44] proposed a set of six methods for performing regionalization under the name 'REDCAP' which stands for 'REgionalization with Dynamically Constrained Agglomerative Clustering and Partitioning'. Each of the methods is

applied, in a similar manner, in two steps: first, a spatially contiguous hierarchy of the set of spatial objects is retrieved by means of hierarchical clustering, and second, the hierarchy is partitioned in order to retrieve a set of aggregate regions. The six methods included in REDCAP are made up of three clustering methods: (1) single-, (2) average-, and (3) complete-linkage clustering (SLK, ALK, and CLK respectively), and two strategies for spatially constraining the clustering process: (1) first-order and (2) full-order constraining (see [44] for details on the methods).

Regionalization hence is a process of simplifying a set of connected spatial regions with respect to both the regions' structural/geographical properties as well as additional attributes characterizing them. Reasons for choosing this method are that: (1) an aggregation hierarchy instead of an absolute segmentation is produced which allows switching between simplification levels, (2) the spatial contiguity of the trajectory segments is inherently considered in the computation of the aggregation hierarchy, thus preserving the temporal aspect of movement, and (3) the user has the flexibility to choose the attributes and clustering method to be used.

### 3.3.1 REDCAP adaptation for trajectory regionalization

Our property-based simplification approach uses the four best performing methods included in REDCAP: Full-Order ALK, SLK, CLK, First-Order SLK, and also Ward's hierarchical clustering method [46].

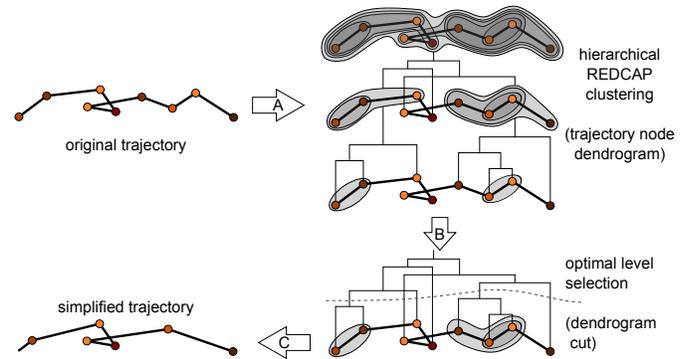


Fig. 3. Schematic depiction of the property-based simplification approach. Contiguity constrained hierarchical clustering is applied to the input trajectory (A) in order to retrieve a hierarchy of regionalization results. An optimal level is selected depending on segmentation quality (B). Finally, the trajectory is simplified by merging segments that are clustered together at the chosen level.

Each trajectory is considered as a unique set of spatial objects – the trajectory positions – and regionalization is applied to each trajectory separately. The process of simplifying a trajectory works as follows: The user interactively selects the attributes that should be considered in the regionalization as well as the type of clustering to be used, and then initializes the process. First, a contiguity matrix is created which sets connectivity constraints on the trajectory positions. Then, clustering is applied with respect to the selected attributes and contiguity constraints, creating a hierarchy of connections in

the form of a trajectory node dendrogram (Figure 3A). Note that contiguity constraints in particular enforce only adjacent trajectory points are ever merged into clusters, which ensures that the temporality of the trajectories is preserved.

A partitioning process is performed next that iteratively divides this hierarchy, from the top down, into a pre-set maximum number  $k$  of regions (movement episodes). At each iteration one hierarchy node is split into two sub-trees corresponding to spatially contiguous movement episodes, until  $k$  is reached (Figure 3A). This step therefore results in  $k$  aggregation levels for trajectory simplification. The partitioning is performed in such a manner that the heterogeneity of each region is minimized. This heterogeneity depends on (1) the values of the user-selected attributes characterizing the region, (2) the region size, (3) the standard deviation of each region, and (4) the preservation of the original data distribution [44].

### 3.3.2 Selection of $k$

Regionalization in REDCAP requires the user to set the maximum number of regions,  $k$ , to be retrieved for the set of spatial objects. Since we consider each trajectory as a separate set, each trajectory requires an individual  $k$  depending on its length (number of contained points,  $N$ ) and character. Therefore, allowing a user to set a single global value for  $k$  is not possible. We handle this requirement by allowing the user to instead specify a global upper bound of desired simplification. This bound is expressed as a percentage value relative to the trajectory’s length that determines the maximum number of episodes (regions) generated for each. The default value is  $k(N) = 50\%$ , or  $k = N/2$  regions. This can optionally be combined with a global absolute upper bound,  $k_{max}$ , so that for each trajectory,  $k = \min(k(N), k_{max})$ .

The result of the entire process is a hierarchy of  $k$  aggregation levels each of which contains a unique regionalization – in our case this corresponds to a trajectory segmentation into movement episodes. Each level consists of  $i = 1 \dots k$  regions that correspond to clusters of position records of the original trajectory. The computation of this hierarchy is a one-time preprocessing step and needs to be recomputed only if a new simplification with respect to a different set of thematic attributes is to be obtained.

### 3.3.3 Quality and aggregation level selection

After the  $k$  aggregation levels have been retrieved, a quality measure is computed for each level ( $k$ ), analogous to the heterogeneity measure used in the partitioning step. The level with the best quality is chosen as our simplification result (Figure 3B).

We suggest the optimal aggregation level as the one that minimizes the function:

$$\arg \min_{i \in K} \left( \frac{1}{|K_i|} \sum_{j \in C_i} \frac{n_j}{N} \sum_{a \in A} \text{var}(a) \right) \quad (1)$$

where  $K = \{K_i\}$  the set of retrieved regionalizations with  $|K_i| = i$  regions at each level  $i \in K$ ;  $C_i$  is the set of clusters that compose each level  $i$ , having  $n_j$  points in each cluster  $j \in C_i$ ;  $A$  is the set of attribute values characterizing each cluster of

points with  $\text{var}(a)$  the variance of each attribute; and  $N$  the length of the trajectory (total number of points).

By using this function, the resulting quality value of each aggregation level is a trade-off between different characteristics of the regionalization; since the value takes into account not only the attribute variation within each region but also the number of regions composing the level, the size of each region, and the total size of the trajectory. We thus avoid minimally simplified trajectories with many small regions achieving unnecessarily high quality scores.

### 3.3.4 Simplification

After selecting the appropriate aggregation level, the original trajectory is simplified by merging points belonging to the same region. The fact that spatial contiguity of the points is considered in the regionalization ensures that it is not possible to merge points located far apart in space nor time. As in DBS, each region’s representative point is set to the arithmetic mean of the merged point positions and is assigned the average value of their thematic attributes.

## 3.4 Method comparison and discussion

The three simplification approaches considered produce trajectories of different structure revealing different movement properties. These differences are illustrated in Figure 1 using a dataset of Galapagos albatross flight trajectories.

Geometry-based simplification represents strictly structural properties of movement without accounting for any additional attributes. We have, in this work, chosen the Douglas-Peucker algorithm as a popular representative of point-based algorithms for line simplification. Such algorithms, however, have been disputed by several authors (see e.g., [15], [47]) as being suitable only for minimal simplification and not for generalization of complex lines. In particular, as can be seen in Figure 1(a), the Douglas-Peucker algorithm (like other line simplification methods) does not give sufficient geometric simplification in the case of self-intersecting trajectories, for which the density-based simplification method can give better results. Since we are aiming for simplification that goes beyond a trajectory’s raw geometry, in the following we will concentrate on our proposed density-based and property-based simplification.

Density-based simplification (Figure 1(b)) is most appropriate for revealing *where* a moving object has been with the highest possible detail with respect to resolution, zoom level and information density. Hence, the approach addresses primarily screen and perceptual limitations in terms of how much detail is actually visible and needed. Property-based simplification (Figure 1(c)) focuses instead on *how* a moving object behaves with respect to both structural and thematic attributes. The approach thus primarily addresses cognitive limitations. To some extent perceptual limitations are also addressed since irrelevant differences (w.r.t. selected attributes) are hidden within the aggregated segments. Display resolution limitations are not of immediate concern, but can of course be addressed through property filtering to detect the positions of relevant patterns and zooming to see them in more detail.

Figure 4 shows results of the two new simplification methods proposed in this work (DBS and PBS) applied to the

same 3D trajectory of a flight track, juxtaposed in the same image. The two simplification results look dissimilar due to differences between the methods in merging of segments and averaging of their attribute values. The trajectory is displayed using a tapered segment representation [26] in which the size of each triangle reflects the size of each segment in order to reveal the most characteristic difference between the two methods. Using DBS, points that are closely located with respect to a specified threshold are replaced by a representative point; whereas using PBS merging of points instead depends on the characterizing attributes being considered in the simplification. Contrariwise, several segments having similar properties are merged by PBS while these are preserved during DBS. These characteristics can be seen in Figure 4 highlighted in yellow, to the left a single segment of the PBS simplified red trajectory corresponds to several of the blue one, while to the right of the figure a single segment of the DBS simplified blue trajectory corresponds to several of the red one. Such differences are also highlighted in Figure 1. Since movement patterns are defined by the segments' characteristics of movement (see Section 1), by merging or preserving segments depending on the similarity of their attributes ensures the preservation of such patterns during the simplification process.

Both proposed methods (DBS and PBS) allow a user to interactively steer the simplification process by appropriately setting parameters depending on task and interest. In DBS a user decides the simplification level by setting *MinPoints* and  $\epsilon$  as described in Section 3.2. In PBS a user interactively selects the trajectory attributes that should influence the simplification and the type of clustering to be used as described in Section 3.3.

Simplification of trajectories, by default, implies a trade off between the number of data points and the accuracy of the simplified representation. Moreover, while actual movement is continuous, data describing movement are discrete. In the proposed simplification methods we use the mean point of

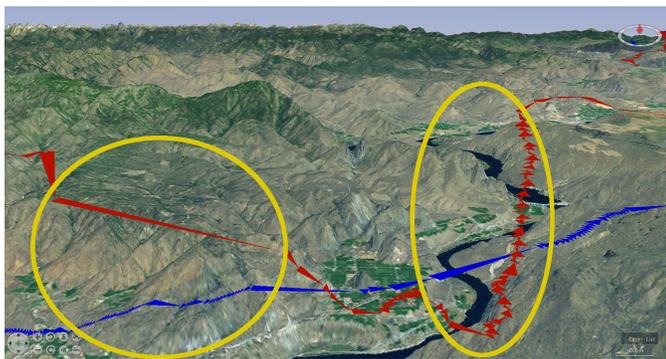


Fig. 4. Results of our two proposed simplification methods applied to the same trajectory and displayed in the same image using tapered segments: DBS (blue) and PBS (red). Differences between the methods are highlighted in yellow. Left: a single PBS trajectory segment of constant descend corresponds to many DBS segments of slightly varying course. Right: a single blue DBS segment corresponds to many PBS segments of an upwards spiral.

a cluster (in DBS) or adjacent segments (in PBS) as a suitable representative that considers all merged points in the merging process. This may change the actual location of the trajectory points and result in irregularly sampled data even if the original data were regularly sampled. As a consequence, artefacts can be introduced in the representation, such as the insertion of intersections of trajectories or the displacement of existing intersections' positions. Even though this can imply a spatial and temporal distortion between trajectories, their overall movement direction is retained and the character of the trajectories' thematic attributes is adequately preserved and represented. Therefore, since the goal of our methodology is to enable the visual analysis and exploration of movement patterns in the data and not to study their exact geographical displacement, we accept this compromise. Furthermore, if intersection (or, more generally, the neighbourhood of trajectories) is an important movement characteristic for the performed analysis, it is possible to compute and then use attribute(s) reflecting those relations within our proposed PBS method, as described in Section 3.3. Examples of such attributes may be the *distance to the Nth nearest trajectory* or the *number of neighbours in a given space-time window*; a full taxonomy is available in [48].

Finally, despite the loss of detail during the simplification, we are preserving the temporal aspect of the trajectories. When a cluster (i.e., an uninterrupted sequence) of trajectory segments is replaced by a single representative segment, the latter receives the time reference of the whole cluster. More specifically, the start and end time of the representative segment correspond to the start time of the first segment and the end time of the last segment of the cluster respectively. Hence, the time span and the duration of the representative segment coincide with the time span and duration of the cluster it represents. This means that the temporal information of each trajectory is coarsened but not distorted.

## 4 THEMATIC ENHANCEMENT

Understanding and analysing movement requires more than simply studying the geographic displacement of a moving object. Attributes giving additional information about the trajectory itself and/or the conditions at the visited positions are important complements in the analysis process [5]. Multiple such attributes may be available for a trajectory and each of them, separately or in concert, can offer insight into different aspects of the movement. Enabling the analyst to identify and accentuate these aspects in a flexible manner depending on task and available data is, therefore, crucial for understanding the behaviour of a moving object. We define *thematic enhancement* as this accentuation of movement attributes aiming to reveal movement behaviour. Thematic enhancement is applied to all trajectories in order to bring forward similarities in movement behaviour which correspond to movement patterns [1], [3], [4], [5]. We thus use this enhancement to highlight multi-attribute movement patterns existing in the trajectory data. Depending on which movement attributes are chosen to be accentuated, different types of patterns are revealed. The basic types of such patterns being, according to [1], similarity, dissimilarity, and arrangement patterns such

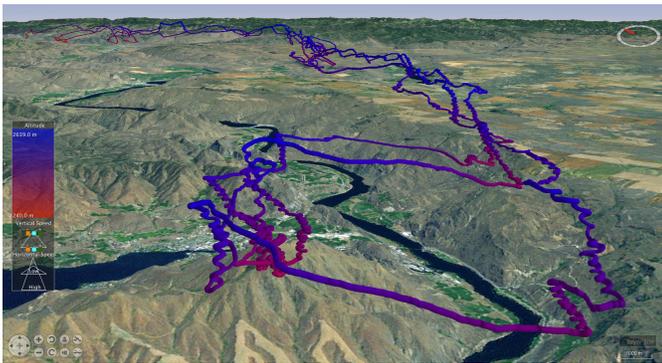


Fig. 5. Paragliding trajectories in Chelan Butte, WA, USA displayed using a tube representation with colour signifying altitude and thickness – speed.

as trend, concentration and alignment. We achieve thematic enhancement by allowing the user to choose and combine visual and computational methods as discussed below.

Directly mapping attribute values to the colour and/or shape parameters of the trajectory representation makes it possible to observe continuous changes in movement behaviour and can reveal this way simple, or primitive, patterns in the data [1], [4]. In Figure 5, for example, changing patterns with regard to altitude and velocity are revealed by mapping these attributes onto colour and tube diameter respectively. The number of attributes that can be represented in this way, however, is limited and additional attributes need to be encoded in extra visual cues such as glyphs attached to the trajectory representation. In Figure 7, for example, triangle glyphs are overlaid on the segments. Colour and shape of these can be used to encode extra attributes. In our case the glyph’s aspect ratio is proportional to segment size and its colour indicates the slope of the segment.

In order to avoid confusion of multiple attributes being displayed simultaneously and to be able to communicate complex movement patterns to the analyst we apply subsequent thematic clustering to the simplified trajectories. A feature vector comprising a user-selected subset of attributes allows the analyst to designate movement properties deemed relevant for analysis. The simplified trajectory segments are clustered into homogeneous movement *episodes* with respect to the attribute dimensions defined by this vector. For example, if the analyst is interested in exploring flock patterns then direction and number of neighbours may be appropriate segment attributes to explore, while if the interest is in finding similar angular patterns sinuosity and direction could be chosen instead. Qualitative colouring is then applied to the trajectory representation according to each segment’s episode classification. Groups of segments with similar properties are distinguished, thus revealing multi-attribute movement patterns present in the data (Figures 6(e), 7). For this thematic enhancement step, any combination of a partitioning clustering algorithm with a feature vector-based similarity measure may be chosen by the analyst. In particular, this allows integration of domain-specific similarity measures suitable for a particular analysis goal. For the purposes of this paper we have made experiments using

*K-means*, *K-medoids*, Expectation Maximization (EM) [42], and Self Organizing Map (SOM) [49]. We have also used Sammon’s projection [50] to classify trajectory segments into movement episodes. This is done by projecting all trajectory points onto 2D colour space and then applying tessellation to divide the space into regions reflecting the spatial density of the data points. The points contained in each obtained region constitute a cluster (episode class) [2].

This thematic enhancement process can be regarded as a further (cognitive) simplification of the trajectories, since focus is brought to episodes of similar movement patterns with respect to analysis of task-specific movement properties.

Thematic enhancement is interactively applied by the user to the simplified trajectories who selects arbitrary combinations of relevant attributes, similarity measure, and clustering algorithm. This makes the methodology appropriate for visual analysis of a wide range of datasets and analysis goals.

## 5 REPRESENTATION & EXPLORATION

We have implemented our methodology in Java and have used the NASA World Wind API [51] for our 3D visualization environment. The methodology is implemented within a larger visual analytics framework which includes both two- and three-dimensional representations for visually exploring trajectories. Suitability of one presentation form or the other is largely determined by the analysis task. If movement patterns are significantly influenced by the third dimension (e.g., paraglider flight tracks) a 3D representation is more appropriate, whereas a focus on visited places (e.g., migrative behaviour) benefits from a projective distortion-free, easier to navigate 2D representation. In both representation types several primitives are available to represent trajectory segments: tapered segments (Figure 1), tubes (Figure 5), ribbons (Figure 7), or lines (Figure 8). Each of these primitives is used to convey different aspects of trajectories.

Lines and ribbons offer a clean representation of the path covered and can be colour-coded according to attribute or classification values. Ribbons can, moreover, be overlaid by complementing glyphs encoding additional thematic attributes. We have overlaid the segments of our ribbon representations with triangle glyphs having an aspect ratio relative to the size of the segment and coloured according to the slope of the segment. Doing this, apart from allowing extra information to be displayed, also makes it possible to distinguish between the separate segments and gives indications of their size which provides valuable information regarding the results of the simplification process. In 3D representations, tubes can be used to encode one (continuous) attribute directly in the diameter, e.g. velocity in Figure 5. They also help to emphasise curved paths in 3D views better than flat ribbons or tapered segments. This becomes especially apparent when using such representations for displaying tracks with tightly curved movement as in Figures 4, 6(e). The tracks in these figures display an upward spiral movement which when simplified (described with fewer points) tends to appear as a zig-zag pattern. This is an artefact resulting from the chosen representation in combination with the simplification. However, since we are interested in distinguishing between

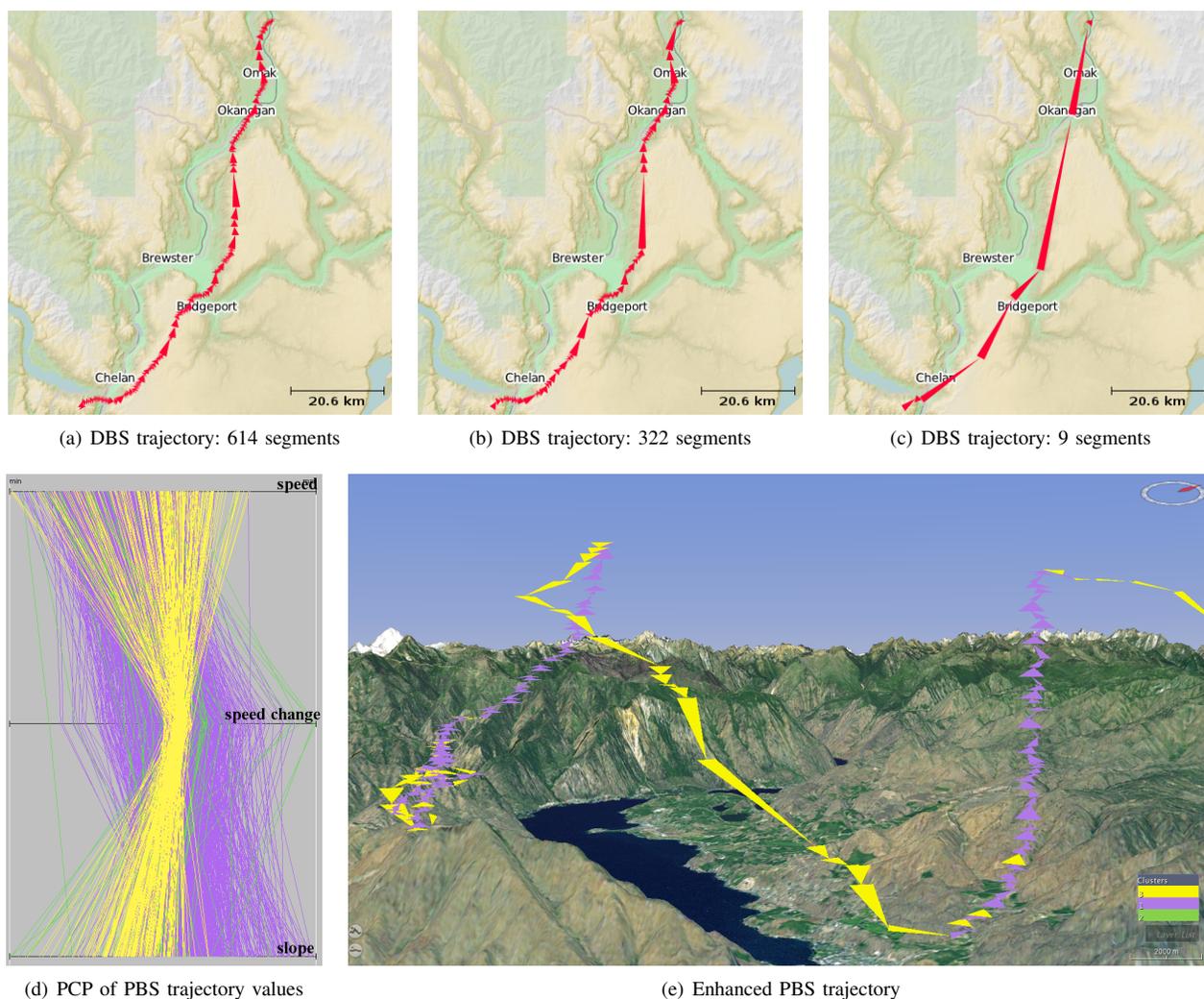


Fig. 6. Paragliding in Lake Chelan. (a-c) A glider’s trajectory simplified using DBS at various levels of detail resulting in: (a) 614, (b) 322, and (c) 9 segments. (d-e) The same trajectory simplified with PBS and thematically enhanced. (d) Parallel coordinates plot (PCP) displaying the attribute values of the PBS trajectory segments coloured by cluster. (e) Representation of the thematically enhanced PBS trajectory using tapered segments coloured by cluster membership.

the segments and accentuating the simplification results, we choose to accept these artefacts in this work for illustration purposes. Finally, using triangles or cones for representing each segment also accentuates the structural composition of a trajectory by clearly demarcating individual segments, and eases keeping track of movement direction within groups of trajectories. In our method, such visual demarcation helps to visually analyse the results of the simplification step. With all types of segment primitives, thematic enhancement of trajectories is provided through colour-coding the segments themselves, as well as by attached or embedded glyphs.

We have chosen to apply colouring of the trajectory segments in two ways. A continuous colour scale is used to encode a single thematic attribute (e.g., altitude in Figure 5); or qualitative colouring is employed to associate segments with clusters of attribute values. The latter therefore encodes multi-attribute movement patterns depending on the user-selected parameter settings for the enhancement step. Assignment of cluster colours is handled differently depending on whether

the generating algorithm produces cluster representatives. If it does not, such as the EM algorithm, unique random colours are assigned. In methods that do produce representative objects, such as *K-means*, *K-medoids* and SOM, colours are assigned by projecting the cluster representatives onto a 2D colour space using Sammon’s projection [50], as discussed in Section 4; it is well-suited for this purpose since similarity of colours corresponds to similarity of cluster representatives.

The framework within which our methodology is implemented provides supplemental views to represent detailed information about clusters and context. Such coordinated views are not explicitly coupled to our proposed methodology, but they provide an important complement to the visual analysis process that our methodology aims to achieve. For example, a parallel coordinates plot (PCP) is used in Figure 6(d) to explore attribute value distributions of the clustered objects; PCP and trajectory views are linked through a common colour scale. This can help the analyst to decide if the current thematic enhancement clustering is concise or requires further

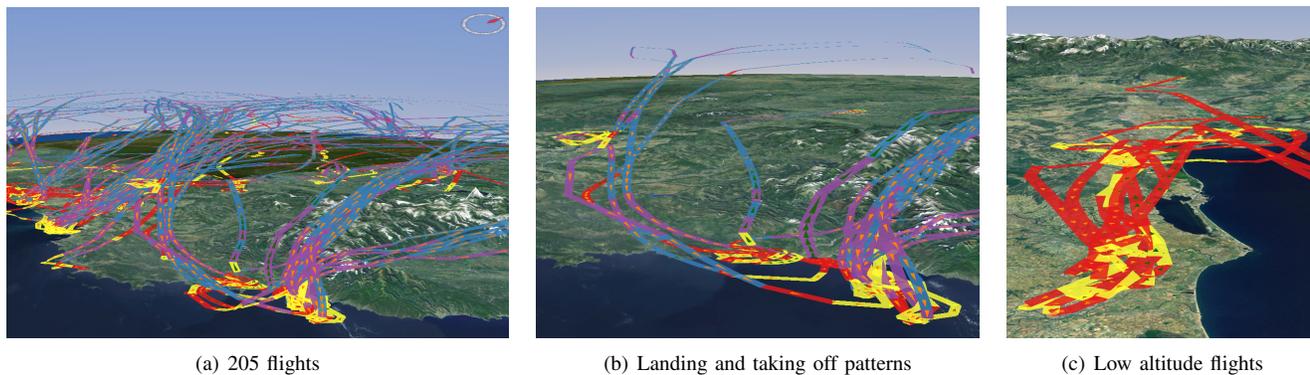


Fig. 7. Representations of flights over France simplified using PBS and thematically enhanced. (a) 205 flights are represented using ribbons. (b) Landing and take-off patterns characterized by turning movement displayed in yellow. (c) Low altitude trajectory segments are represented in red.

refinement. Other complementary view examples include histograms for showing point distributions within clusters, curves showing intra-cluster value frequency distributions, and space-time cubes to show spatio-temporal relations.

Visual representation and interactive exploration of the results produced at all steps of our methodology are of paramount importance. They offer feedback and facilitate progressive refinement of intermediate results by creating a mutually reinforcing cycle of user-steered trajectory simplification, thematic enhancement, and subsequent visual inspection of the results. The need for refinement can arise at different steps of the proposed visual analytics loop.

Simplification of trajectories is performed as a one-time preprocessing step for a chosen selection of parameters, as described in Section 3. Visual inspection of the enhanced results may give indications to the analyst of an inadequate choice of simplification parameters leading to a second iteration of the simplification. Furthermore, inspection of the identified movement patterns may give rise to several iterations of enhancement followed by visual inspection in order to explore different aspects of the underlying movement behaviour. This iterative process is performed by the analyst within our framework using standard widget interfaces that allow them to steer all steps of the methodology.

## 6 VISUAL ANALYSIS EXAMPLES

We illustrate the functionality and potential of our methodology for visually analysing movement patterns using three example datasets.

### 6.1 Paragliding in Lake Chelan

We have considered a small paragliding dataset collected on 3<sup>rd</sup> June 2011 in Chelan Butte, WA, USA. The data consists of three flight trajectories composed of 12,786 recorded positions and having average length of 163 km. Figure 5 shows the trajectories represented as tubes with colour depicting altitude and diameter of the tube depicting speed. The trajectories are long and the movement displays complex sinusoidal patterns which are repetitive. These patterns represent the paragliders' climbing motion which occurs when they find thermal uplifts.

For the purpose of illustrating the methodology we will concentrate on a single trajectory composed of 4,950 points.

We first want to study where the paraglider has been in order to get an overview of the traversed path and surrounding area. We therefore apply density-based simplification with varying parameters and view the resulting trajectory on a map using a 2D tapered segment representation. Figures 6(a-c) show the trajectory simplified with three levels of detail; composed of 614, 322, and 9 segments respectively. These views reveal the large size of the territory covered and information about the route followed; such as direction (north-east) and length.

In order to observe how the glider has moved and explore the movement patterns in terms of multiple attributes we apply property-based simplification to the trajectory. We have used Full-Order CLK with respect to two attributes: course in degrees and slope, which resulted in a simplified trajectory consisting of 1047 points. We then enhance the simplified trajectory by clustering and colouring trajectory segments with respect to thematic attributes: speed, change in speed, and slope. We first use Sammon's method, as described in Section 4, to inspect the distribution of the segments' attribute values and retrieve 3 clusters. Since in paragliding there are 3 basic movement states: glide, climb at thermals, and dives, retrieving 3 clusters seems well justified so we then go ahead and cluster our data into three clusters using the EM algorithm.

The results of the thematically enhanced PBS trajectory can be seen in Figures 6(d-e). To the right (Figure 6(e)) the simplified trajectory is displayed using the tapered segment representation with its segments coloured with respect to their respective cluster membership. To the left (Figure 6(d)) the segment attribute values are shown using parallel coordinates and coloured by cluster in order to closer inspect the distribution of the values and character of the clusters. The PBS trajectory reveals three distinctive movement patterns of the paragliding flight. The yellow cluster represents glides, purple upward spiral movement (climbs) reflecting thermal lifts, and green downward spiral movement (dives) appearing primarily at the end of the flight (Figure 6(d-e)).

The simplification and enhancement results reflect and confirm the distinctions made about the goals of the two simplifi-

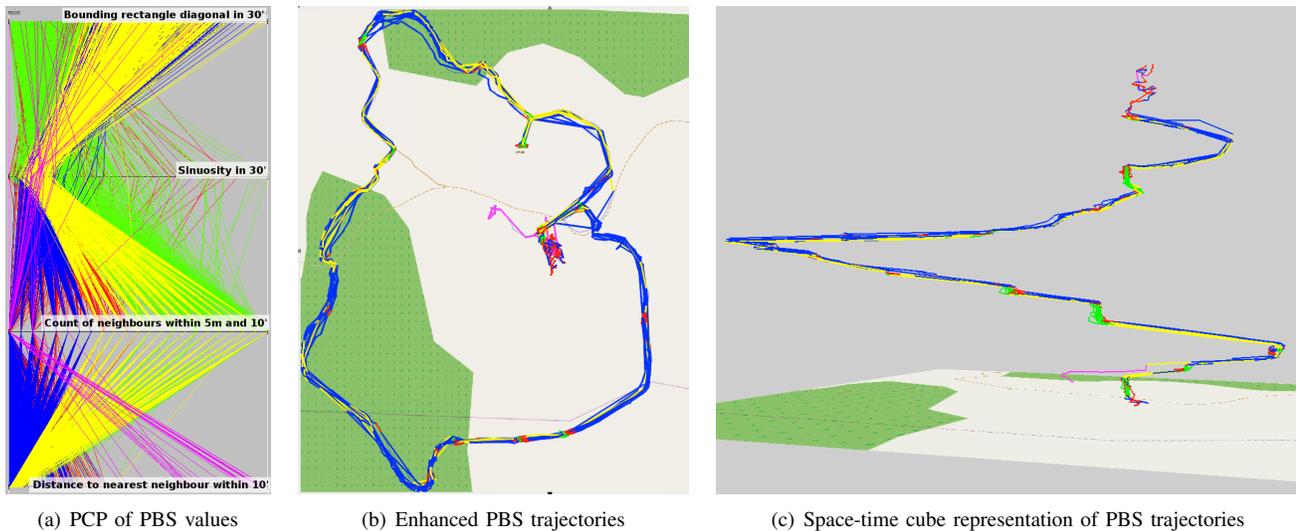


Fig. 8. Results of property-based simplification (PBS) and thematic enhancement of trajectories of a group of traveling scientists on a field trip in Estonia. a) Segment attribute values are displayed in a parallel coordinates plot. b) Simplified trajectories are displayed as lines on a 2D map. c) Simplified trajectories are displayed in a space-time cube representation. In all representations the trajectory segments are coloured according to their cluster membership.

cation methods. Being concerned with the overall whereabouts of a moving object density-based simplification conveys good overviews of the followed path and does not preserve the details of the movement. Property-based simplification instead merges homogeneous parts of the trajectory and retains detail where there is change in movement characteristics, which makes it better suited for identification of movement patterns.

## 6.2 Flights over France

In a second example we consider two larger datasets composed of flights over France collected using radar. The smaller dataset includes 205 trajectories consisting of 44,281 recorded positions, and the larger one includes 2,025 trajectories consisting of 214,083 recorded positions. Being collected through radar the data include trajectories of various aircraft ranging from large commercial airliners to light aircraft and helicopters.

We simplify the data with PBS using WARD hierarchical clustering with respect to speed, acceleration, and vertical speed. The resulting simplified datasets are composed of 8,939 and 56,475 trajectory points respectively. We then enhance the trajectories by clustering the simplified segments into four clusters using EM with respect to speed, acceleration, difference in relative speed, and turn in degrees; colouring them accordingly. The enhancement results of the smaller dataset can be seen in Figure 7. We observe that high altitude movement is clustered together and represented in blue, and similarly low altitude segments are grouped and shown in red. The attribute 'altitude' has not been used for clustering. The fact that movements on high and low altitudes have been separated by the clustering means that the movement properties at these altitudes are different. Purple segments are characterized by upward or downward slope and yellow ones reflect turns and curved movement. These four clusters make it possible to separate between commercial/charter flights

(Figure 7(a)) flying across the country, and low flying aircraft such as the ones in Figure 7(c). Furthermore, landing and take-off patterns are revealed by tracing the purple segments as in figure 7(b). Also holding patterns (i.e. aircraft orbiting the airport waiting for their time to land) can be identified as (yellow) curved movement before proceeding with the actual landing, as can be seen to the left of Figure 7(b).

## 6.3 Traveling scientists in Estonia

Our third example is concerned with pedestrian movement of scientists attending a field trip. The dataset used consists of 13 trajectories including a total of 13,141 recorded positions.

We simplify the trajectories with PBS using Full-Order CLK with respect to attributes direction, speed, and bounding rectangle diagonal in 30 seconds, which is a value showing whether a person is standing or moving within a confined area or moving forward covering a larger distance. The simplification results in a dataset of 3,324 trajectory points (25.3% compression ratio). We then enhance the simplified trajectories by applying *K-means* clustering to the trajectory segments with respect to attributes: bounding rectangle diagonal in 30 seconds, sinuosity in 30 seconds, count of neighbours within 5 meters and 10 seconds, distance to nearest neighbour within a 10 second time window. We retrieve 5 clusters.

The results of the thematically enhanced PBS trajectories can be seen in Figure 8. In Figure 8(b) the PBS simplified trajectories are displayed as lines on a 2D map with line segments coloured with respect to their cluster membership, while in Figure 8(c) they are shown within a space-time cube (STC) representation. The trajectory segment attribute values are displayed in Figure 8(a) using a parallel coordinates representation, coloured according to cluster membership. This way the distribution of the segment values and the characteristics of the movement in each of the clusters are revealed.

Each of the clusters corresponds well to a characteristic movement pattern. Blue represents forward (large bounding rectangle diagonal), straight (low sinuosity) movement of small (low count of neighbours), dense (small distance to nearest neighbour) groups. Yellow represents similar movement but in larger groups. Green represents groups hovering around probably to observe something in the scenery or take pictures etc. Red represents stops in the group and pink appears to be straight movement away from the group, this in fact corresponds to an individual having forgotten their camera in the car and going back to get it.

## 7 QUALITATIVE USER SURVEY

In order to assess the functionality and results of the proposed methodology we performed a small qualitative user survey. We conducted a questionnaire survey with the help of the “traveling scientists” of the Estonian field trip example described in Section 6.3. The goal of the user survey was not to make a formal evaluation but rather to examine whether the participants of the field trip could recognize and label their movement behaviour during the field trip. Thus, our aim was to investigate whether applying our methodology would preserve the movement patterns of the travellers. We have used the results of the PBS approach in the user survey since it is the most appropriate when analysing ‘how’ people have moved.

Nine field trip participants took part in the survey, one woman and eight men, all of which were geoinformatics experts. Each participant received 4 images via email and was asked to answer two questions concerning their movement using these images and their memories of the field trip. The images they received were: (1) a 2D line representation of the original and (2) PBS simplified trajectories, (3) a thematically enhanced 2D line representation of the simplified trajectories (Figure 8(b)), and (4) a space time cube representation of the thematically enhanced simplified trajectories (Figure 8(c)). The two latter thematically enhanced images consisted of trajectory segments clustered according to several group movement attributes reflecting compactness of group movement, sinuosity and speed, as described in Section 6.3. The participants were asked to:

- 1) estimate what the compression ratio of the simplified trajectories was, choosing between: (a) less than 15%, (b) 15% to 30%, (c) 30% to 60%, or (d) more than 60%,
- 2) study the thematically enhanced representations and provide their interpretation of the five highlighted clusters, i.e. the movement characteristics of each cluster.

Concerning the assessment of the compression ratio, participants either answered correctly or they estimated that the reduction was less than it actually was. This is positive since it indicates that the simplified trajectories representations adequately resemble the original ones.

The interpretations of the clusters varied between the participants in terms of response content. Three participants did not adequately respond to the survey. From the remaining 6 participants, all of them characterized green cluster segments as having low speed, looking around, or being stops at points of interest. This interpretation follows closely our definition

of the cluster in Section 6.3 as ‘*representing groups hovering around*’. The movement properties of the clusters can be seen in Figure 8. The red cluster was also described as stationary behaviour, stops or as having high compactness and low movement. In fact several commented on the similarity of the red and green segments. The interpretations of this cluster also are close to our definition of red segments as ‘*stops*’. The blue cluster segments were identified as being the main group, with high speed, and constant movement which agrees with our interpretation of them. Also, participants commented on the yellow and blue segments being similar which they in fact are. Their difference is mostly in the *count of neighbours*, as seen in Figure 8(a), which is of course not visible in the static images (Figures 8(b),(c)) presented to the participants. Finally, all participants labelled the pink cluster segments as outliers which we also identified as such in our classification.

Overall, the results we received involved meaningful interpretations of the segments’ cluster membership. This indicates that the simplification process did not distort the underlying movement characteristics of the data, since these were identifiable when highlighted on static image captions of the representations. Furthermore, even though we did not make a similar survey concerning our other examples, we had experts, two experienced paragliders and an air traffic controller, inspect our enhanced visualizations and confirmed that they reflect the important stages of flights and produce an accurate overall image of flight activities.

## 8 DISCUSSION AND CONCLUSIONS

We have proposed a systematic visual analytics methodology that aims to support the visual analysis of movement by the simplification and thematic enhancement of trajectories.

Simplification is achieved by reducing the number of trajectory segments that need to be presented to the user. Such simplification improves the perception and cognition of the presented information in the following ways. First, it reduces the perceptual and cognitive load on the user (the user needs to process fewer visual items representing the segments) [52]. Second, it reduces the display clutter and, hence, its negative impact on perception [53], [54]. Third, it removes task-irrelevant details and thereby helps the user to focus on task-relevant information, which facilitates cognition [52], [54].

We have proposed two new simplification methods: density-based and property-based simplification (DBS and PBS respectively). We have furthermore discussed how simplified trajectories can be enhanced by clustering segments with respect to thematic attributes and representing them accordingly. Finally, we have illustrated by example how our methodology can be applied in order to explore movement datasets of different character and identify patterns within them. Using our approach we have detected expected patterns, such as paraglider climbs at thermals, landings and take-offs of aircraft, and discovered less obvious ones, such as holding patterns before landing, and characteristic and irregular group movement behaviour. The methodology has been primarily developed for three dimensional trajectories, such as flights (hence *SimpliFly*), but is not limited to them, as demonstrated by the human pedestrian movement example.

The two simplification methods that form the basis of our methodology result in trajectories with diverging properties and distinctive differences. These become apparent at all steps of the process. DBS is better suited for revealing *where* a moving object has been while PBS is better suited for revealing *how* an object has behaved during its movement. These differences continue to be apparent in the thematic enhancement resulting from clustering segments with respect to a selection of characteristic attributes. In PBS thematic attributes are considered already in the simplification itself. This leads to better preservation of the movement characteristics of interest and thus patterns of movement, as defined in sec 1, since only similar/dissimilar segments are merged during the proposed simplification methodology. DBS instead mainly focuses on the structural simplification and optimal use of screen space at the cost of accompanying characteristics of movement being toned down. DBS is therefore more appropriate for extracting and displaying the overall, rough pattern of movement while PBS is better for a more detailed study of movement patterns.

The combination of computations and representation techniques we propose facilitates flexible exploration during which all contributing factors are steered and tailored by the analyst who decides the simplification method and level of simplification, and the attributes that should be considered in both the property-based simplification and the thematic enhancement.

## ACKNOWLEDGMENTS

The Galapagos albatross dataset was provided by Sebastian M. Cruz from the Max Planck Institute for Ornithology through movebank.org. The paragliding dataset was downloaded from the open source server Leonardo (Paragliding Forum) [55]. We thank Christophe Hurter for providing access to the France flights dataset and the Mobile Tartu 2012 participants for providing data and participating in the user survey.

## REFERENCES

- [1] N. Andrienko and G. Andrienko, *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Verlag, 2006.
- [2] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel, *Visual Analytics of Movement*. Springer Verlag, 2013.
- [3] P. Laube, "Progress in movement pattern analysis," in *Behaviour Monitoring and Interpretation - Smart Environments*, 2009, pp. 43–71.
- [4] S. Dodge, R. Weibel, and A.-K. Lautenschütz, "Towards a taxonomy of movement patterns," *Inform Vis*, vol. 7, no. 3-4, pp. 240–252, 2008.
- [5] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis et al., "Semantic trajectories modeling and analysis," *ACM Comput Surv*, vol. 45, no. 4, p. 42, 2013.
- [6] F. Giannotti and D. Pedreschi, *Mobility, data mining, and privacy: geographic knowledge discovery*. Springer, 2008.
- [7] R. Güting and M. Schneider, *Moving objects databases*. Morgan Kaufmann Publisher, 2005.
- [8] D. Douglas and T. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Cartographica*, vol. 10, no. 2, pp. 112–122, 1973.
- [9] M. Visvalingam and J. D. Whyatt, "Line generalisation by repeated elimination of points," *Cartogr J*, vol. 30, no. 1, pp. 46–51, 1993.
- [10] G. Liu, M. Iwai, and K. Sezaki, "A Method for Online Trajectory Simplification by Enclosed Area Metric," in *ICMU*, 2012, pp. 40–47.
- [11] D. Burghardt, "Controlled line smoothing by snakes," *Geoinformatica*, vol. 9, no. 3, pp. 237–252, 2005.
- [12] P. Laube and R. Purves, "Cross-scale movement trajectory analysis," in *GIS Research UK 18th Annual Conference*, 2010, pp. 103–107.
- [13] M. Agrawala and C. Stolte, "Rendering effective route maps: improving usability through generalization," in *ACM SIGGRAPH*, 2001, pp. 241–249.
- [14] A. Gemsa, M. Noellenburg, T. Pajor, and I. Rutter, "On d-regular schematization of embedded paths," in *SOFSEM: Theory and Practice of Computer Science*. Springer, 2011, vol. 6543, pp. 260–271.
- [15] M. Visvalingam and J. D. Whyatt, "The Douglas-Peucker Algorithm for Line Simplification: Re-evaluation through Visualization," *Comput Graph Forum*, vol. 9, no. 3, pp. 213–225, Sep. 1990.
- [16] J. Dykes and D. Mountain, "Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications," *Comput Stat Data An*, vol. 43, no. 4, pp. 581–603, 2003.
- [17] M. Buchin, A. Driemel, M. J. van Kreveld, and V. Sacristan, "An algorithmic framework for segmenting trajectories based on spatio-temporal criteria," in *GIS*, 2010, pp. 202–211.
- [18] A. Anagnostopoulos, M. Vlachos, M. Hadjieleftheriou, E. Keogh, and P. Yu, "Global distance-based segmentation of trajectories," in *ACM SIGKDD*, 2006, pp. 34–43.
- [19] C. Panagiotakis, N. Pelekis, I. Kopanakis, E. Ramasso, and Y. Theodoridis, "Segmentation and sampling of moving object trajectories based on representativeness," *IEEE T Knowl Data En*, vol. 24, no. 7, pp. 1328–1343, 2012.
- [20] P. Laube, S. Imfeld, and R. Weibel, "Discovering relative motion patterns in groups of moving point objects," *Int J Geogr Inf Sci*, vol. 19, no. 6, pp. 639–668, 2005.
- [21] S. Dodge, R. Weibel, and E. Forootan, "Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects," *Comput Environ Urban*, vol. 33, no. 6, pp. 419–434, 2009.
- [22] M. Sester, U. Feuerhake, C. Kuntzsch, and L. Zhang, "Revealing underlying structure and behaviour from movement data," *Künstl Intell*, vol. 26, pp. 223–231, 2012.
- [23] J. Lee, J. Han, and K. Whang, "Trajectory clustering: a partition-and-group framework," in *Int'l Conference on Management of Data*. ACM, 2007, pp. 593–604.
- [24] Z. Li, J. Lee, X. Li, and J. Han, "Incremental clustering for trajectories," in *Database Systems for Advanced Applications*, 2010, pp. 32–46.
- [25] J. Lee, J. Han, and X. Li, "Trajectory Outlier Detection: A Partition-and-Detect Framework," in *IEEE Int'l Conf on Data Engineering*, 2008, pp. 140–149.
- [26] D. Holten and J. J. van Wijk, "A user study on visualizing directed edges in graphs," in *SIGCHI Conf on Human Factors in Computing Systems*, 2009, pp. 2299 – 2308.
- [27] I. Vasiliev, "Mapping Time," *Cartographica*, vol. 34, 1997.
- [28] M. J. Kraak, "The Space-Time Cube Revisited from a Geovisualization Perspective," in *Int'l Cartographic Conf*, vol. 1988–1995, Durban, South-Africa, 2003.
- [29] M.-P. Kwan, "Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set," *Transport Res C-Emer*, vol. 8, no. 1-6, pp. 185 – 203, 2000.
- [30] C. Ware, R. Arsenault, M. Plumlee, and D. Wiley, "Visualizing the underwater behavior of humpback whales," *IEEE Comput Graph*, pp. 14–18, 2006.
- [31] C. Hurter, B. Tissoires, and S. Conversy, "Fromdady: Spreading aircraft trajectories across views to support iterative queries," *IEEE T Vis Comput Gr*, vol. 15, no. 6, pp. 1017–1024, 2009.
- [32] P. Forer and O. Huisman, "Space, Time and Sequencing: Substitution at the Physical/Virtual Interface," in *Information, Place, and Cyberspace: Issues in Accessibility*. Springer Verlag, 2000, p. 73.
- [33] N. Willems, H. Van De Wetering, and J. Van Wijk, "Visualization of vessel movements," *Comput Graph Forum*, vol. 28, no. 3, pp. 959–966, 2009.
- [34] D. Guo, "Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data," *IEEE T Vis Comput Gr*, vol. 15, no. 6, pp. 1041–1048, October 2009.
- [35] D. Phan, L. Xiao, R. Yeh, P. Hanrahan, and T. Winograd, "Flow Map Layout," in *IEEE Symp on Information Visualization*, 2005, pp. 219–224.
- [36] W. Tobler, "Experiments in migration mapping by computer," *Am Cartographer*, vol. 14, no. 2, pp. 155–163, 1987.
- [37] K. Buchin, B. Speckmann, and K. Verbeek, "Flow Map Layout via Spiral Trees," *IEEE T Vis Comput Gr*, vol. 17, no. 12, pp. 2536–44, 2011.
- [38] J. Wood and J. Dykes, "Spatially Ordered Treemaps," *IEEE T Vis Comput Gr*, vol. 14, no. 6, pp. 1348–1355, 2008.
- [39] T. Crnovrsanin, C. Muelder, C. Correa, and K. Ma, "Proximity-based visualization of movement trace data," in *IEEE Symp on Visual Analytics Science and Technology*. IEEE, 2009, pp. 11–18.

[40] P. Lundblad, O. Eurenus, and T. Heldring, "Interactive visualization of weather and ship data," in *Int'l Conf Information Visualisation*, Barcelona, Spain, 2009, pp. 379–386.

[41] M. A. Fedak, P. Povell, and B. J. McConnell, "Mamvis: A marine mammal behaviour visualization system," *J Visual Comp Animat*, vol. 7, no. 3, pp. 141–147, 1996.

[42] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2006.

[43] M. Ankerst, M. Breunig, H. Kriegel, and J. Sander, "Optics: Ordering points to identify the clustering structure," in *ACM SIGMOD Record*, vol. 28(2). ACM, 1999, pp. 49–60.

[44] D. Guo, "Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP)," *Int J Geogr Inf Sci*, vol. 22, no. 7, pp. 801–823, Jul. 2008.

[45] D. Guo and H. Wang, "Automatic Region Building for Spatial Analysis," *Trans GIS*, vol. 15, pp. 29–45, Jul. 2011.

[46] J. Ward, "Hierarchical Grouping to Optimize an Objective Function," *J Am Stat Assoc*, vol. 58, no. 301, pp. 236–244, 1963.

[47] M. Visvalingam and J. Whyatt, "Cartographic algorithms: Problems of implementation and evaluation and the impact of digitising errors," *Comput Graph Forum*, vol. 10, no. 3, pp. 225–235, 1991.

[48] G. Andrienko, N. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel, "Scalable analysis of movement data for extracting and exploring significant places." *IEEE T Vis Comput Gr*, vol. 19, no. 7, pp. 1078–94, 2013.

[49] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, 2001.

[50] J. W. J. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Trans on Computers*, vol. 18, pp. 401–409, 1969.

[51] "Nasa world wind api," <http://worldwind.arc.nasa.gov>.

[52] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," *Learn Instr*, vol. 4, no. 4, pp. 295 – 312, 1994.

[53] D. M. Levi, "Crowding—an essential bottleneck for object recognition: a mini-review." *Vision Res*, vol. 48, no. 5, pp. 635–54, Feb. 2008.

[54] D. Whitney and D. M. Levi, "Visual crowding: a fundamental limit on conscious perception and object recognition." *Trends Cogn Sci*, vol. 15, no. 4, pp. 160–8, Apr. 2011.

[55] "Paragliding forum," <http://www.paraglidingforum.com/leonardo/>.



**David Spretke** received his Bachelor's degree (2007) and the Master's degree (2010) in Information Engineering from the University of Konstanz (Germany). Since his graduation he is a research assistant in the Data Analysis and Visualization Group at the University of Konstanz. His main research fields include Visual Analytics, Data Mining and the analysis of spatiotemporal events and movement of objects in geographic space.



**Florian Mansmann** is a post-doctoral researcher at the University of Konstanz, Germany where he leads research projects on network security and geoinformatics. He holds a B.Sc. degree from the University of Konstanz, a M.Sc. degree from Vrije Universiteit Brussels, Belgium, and a Ph.D. degree in computer science (2008) from the University of Konstanz. His research interests include geoinformatics, network security, real-time analysis, information visualization, knowledge discovery and visual analytics.



**Georg Fuchs** obtained his MSc in computer science in 2002 and his PhD in computer science in 2011, both from Rostock University. Since 2012 Dr. Georg Fuchs is affiliated with the Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS working in the field of visual analytics, with a focus on spatio-temporal data analysis. His research interests include information visualization, geovisualization, visual analytics, interactive knowledge discovery and data mining, and computer graphics.



**Katerina Vrotsou** is a post-doctoral research associate at Linköping University, Sweden where she works with interactive visual analysis of multidimensional event-based data. She received her MSc in Media Technology and Engineering in 2005 and her PhD in Visualization and Interaction in 2010 from Linköping University. Her research interests include information visualization, visual analytics, data mining, and interactive knowledge discovery.



**Natalia Andrienko** has been working at GMD, now Fraunhofer IAIS, since 1997. Since 2007, she is a lead scientist responsible for the visual analytics research. Since 2013 she is professor (part-time) at City University Londond. She co-authored the monographs *Exploratory Analysis of Spatial and Temporal Data* (Springer, 2006) and *Visual Analytics of Movement* (2013), over 60 peer-reviewed journal papers, over 20 book chapters and more than 100 conference papers. She received best paper awards at AGILE 2006 and IEEE VAST 2011 and 2012 conferences, best poster awards at AGILE 2007 and ACM GIS 2011, and VAST challenge award 2008.



**Halldor Janetzko** is a research associate in the Data Analysis and Visualization Group at the University of Konstanz. He received his B.Sc. degree in 2008 and his M.Sc. degree in 2010 from the University of Konstanz. Since his graduation he focuses on research in the fields of Visual Analytics, Data Mining and overlap-free representations of spatiotemporal movement data.



**Gennady Andrienko** is a lead scientist responsible for the visual analytics research at the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS) and professor (part-time) at City University London. He co-authored the monographs *Exploratory Analysis of Spatial and Temporal Data* (Springer, 2006) and *Visual Analytics of Movement* (2013), 60+ peer-reviewed journal papers, 20+ book chapters and more than 100 conference papers. Since 2007, Gennady Andrienko is chairing the ICA Commission on GeoVisualization. He co-organized scientific events on visual analytics, geovisualization and visual data mining, and co-edited 10 special issues of journals.

**Carlo Navarra** is a research engineer at Linköping University, Sweden. He received his B.Sc. degree in computer science from the University of Perugia, Italy and his M.Sc. in Advanced Computer Graphics from Linköping University, Sweden. His work is focused on the interactive visualization of spatiotemporal movement, and climate change data.