RESEARCH ARTICLE

Similarity of trajectories taking into account geographic context*

Maike Buchin[1], Somayeh Dodge[2], and Bettina Speckmann[3]

[1]Department of Mathematics, Ruhr-University Bochum, Germany
[2]Department of Geography and Environmental Studies, University of Colorado at Colorado Springs, USA
[3]Department of Mathematics and Computer Science, TU Eindhoven, The Netherlands

Abstract: The movements of animals, people, and vehicles are embedded in a geographic context. This context influences the movement and may cause the formation of certain behavioral responses. Thus, it is essential to include context parameters in the study of movement and the development of movement pattern analytics. Advances in sensor technologies and positioning devices provide valuable data not only of moving agents but also of the circumstances embedding the movement in space and time. Developing knowledge discovery methods to investigate the relation between movement and its surrounding context is a major challenge in movement analysis today. In this paper we show how to integrate geographic context into the similarity analysis of movement data. For this, we discuss models for geographic context of movement data. Based on this we develop simple but efficient context-aware similarity measures for movement trajectories, which combine a spatial and a contextual distance. These are based on well-known similarity measures for trajectories, such as the Hausdorff, Fréchet, or equal time distance. We validate our approach by applying these measures to movement data of hurricanes and albatross.

Keywords: movement data, geographic context, spatiotemporal analytics, similarity measures, trajectory analysis, modeling context, environmental factors, Fréchet distance, Hausdorff distance, equal time distance

---

*A preliminary version of this paper appeared in [10]. We carefully revised and extended the paper, in particular, we have added the concept of robustness and the second case study on albatross.

1   Introduction

Over the past years the availability of devices that can be used to track moving objects (e.g., GPS systems, smart phones, geo-sensors, surveillance cameras, RFID tags, etc.) has increased dramatically, leading to an explosive growth in movement data. Objects being tracked range from animals and humans (e.g., for behavioral studies), vehicles (for traffic modeling and prediction), to hurricanes, and sports players. Tracking an object gives rise to a sequence of time-ordered points in space, called a *trajectory*. Naturally the goal is not only to track objects but also to extract information and generate knowledge from the resulting data. Consequently recent years have seen a significant increase of interest in the development of analytical techniques and computational methods for the study of movement [28].

The movement of animals, people, and vehicles is embedded in a geographic context. In this study, the geographic context is identified as *the locational circumstances of a moving agent*, which includes the external factors connected to the underlying landscape (e.g., topography, land cover) or the surrounding environment (e.g., weather condition) in which the movement takes places. This context both enables and limits movement. For instance, cars are constrained to move on road networks and hurricanes cannot develop over cold ocean currents. Variations in geographic context parameters may also cause certain behavioral responses in moving agents. For instance, movement of soaring species is governed and controlled via variation in wind speed and uplift [16]. Previous movement ecology studies suggest that wind is a key factor in determining migration paths and short-term flight patterns of pelagic birds [16]. For instances, Galapagos albatrosses (*phoebastria irrorata*) make a clockwise movement pattern from/to the Galapagos Islands (where they nest) to forage at the productive coastal areas of Peru. During their transit flights, albatross outbound movements towards the Peruvian coast are hampered by head winds and hence their trajectories take a more northern route and a winding pattern, while the return trajectories follow a directional path facilitated by tail-wind assistance [16]. We will analyze this data in Section 6, see also Figure 10.

Context has been an important part of research in geography and GIScience. In fact, Tobler's first law of Geography is based on the similarity in the proximate context of geographic entities [40]. There has been a large body of literature in GIScience on studying and modeling the link between geographic phenomena and context through the classification and characterization of neighborhood space. This includes the use of cellular automaton models in the study of urban growth [14], agent-based models in behavioral studies [1,41], and the least-cost path analysis [18], to name but a few. In those studies, the context of neighboring space is used to characterize spatial or spatio-temporal phenomena. In this study, we use geographic context as another discriminator for trajectory analysis. That is, a movement trajectory can be identified not only based on how the agent moves but also based on the characteristics of the surrounding environment and the underlying landscape (i.e., geographic context). We therefore characterize movement similarity based on *the nature of space* that the object is moving through.

A fundamental analysis task on trajectories is similarity analysis. It answers the question: "How similar are the movements paths of two or more objects?" Similarity analysis can be the basis of other tasks in Geographic Knowledge Discovery processes, such as clustering, pattern recognition, simplification, or representation [31]. Also, it can be an analysis task by itself, for instance in hurricane analysis. Hurricanes are known to follow

similar paths, particularly when they form in close proximity of each other in space and time [15]. Therefore, when a new hurricane evolves, meteorologists use past hurricanes with a similar initial track for predicting the track of the developing hurricane, in particular the location of future recurvature points and landfalls. Hurricanes are also known to be strongly influenced by geographic context, most importantly the underlying land/sea structure, geographic latitude, surface temperature, and surface pressure [19]. Hurricanes whose tracks are very similar in shape may still be very different in their nature. Consider the hurricanes in Figure 1. Spatially, their trajectories are very similar. However, they differ in geographic context (land/sea). When a hurricane hits land, its energy source—the warm sea surface—is taken away, which will severely weaken it. This can cause two similar hurricane tracks to differ in their nature, e.g., exhibit different intensities as they evolve and therefore distractive impacts. Thus, it is crucial to distinguish these hurricanes. Recognizing the significance of similarity analysis in movement research, this study aims at integrating contextual parameters into the similarity analysis of movement trajectories.

Our goal in integrating context into the analysis is twofold: (1) to learn about the movement from the context (e.g., an animal heading towards a goal made a detour because of an obstacle), and (2) to learn about the context from the movement (if all tracks avoid an area, there is likely an obstacle there). We develop context-aware similarity measures for the first task: understanding movement based on context. These similarity measures allow to distinguish trajectories by their spatial component as well as their context, and hence enable the investigation of the relation between geographic context and movement similarities.

Contribution and organization. In Section 2 we first discuss the background of our research and related work. Then, in Section 3, we discuss models for geographic context that allows us to integrate context into the analysis of movement data. In Section 4, we develop simple but efficient context-aware algorithms for trajectory similarity analysis based on our model. These combine a spatial and contextual distance and extend well-known algorithms for trajectory similarity using the Fréchet distance or the equal-time distance. In Section 5, we extend our approach to make it robust to small changes in context. The context of two trajectories clearly plays a significant role in similarity analysis. However, to the best of our knowledge, we present *the first context-aware approach to trajectory similarity for movement not constrained to networks*. In Section 6 we validate our approach by applying it to tracks of hurricanes and albatross.

2 Background—movement analysis

The importance of temporal aspects of movement has attracted a range of studies in GI-Science and related disciplines, including investigations of space-time settings (i.e., space-time path, prism, and station) [26, 30], modeling moving objects and their collective dynamics [20, 22], development of new analytical methods for movement pattern discovery [28], exploratory data analysis [5, 42], and visual analytics techniques for movement [3].

A review of the literature reveals that the existing geographic knowledge discovery (GKD) techniques for movement data, including similarity search techniques, are mostly based on the geometric properties of the trajectories (i.e., the path of an object through space and time) and embedding contextual variables has been ignored [33]. This has been identified as pitfall of current methods for movement analysis [29, 37].
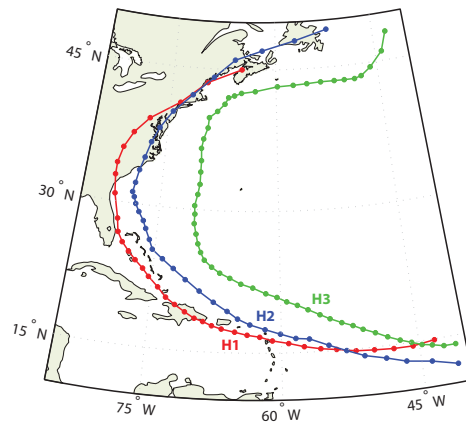
Figure 1: Trajectories of hurricanes distinguished by context: tropical storm Chris 1988 (red), hurricane Bertha 1996 (blue), and Edouard 1996 (green).

In real world applications the movement of an organism is very much influenced by its internal (i.e., the focal individual) as well as external factors (i.e., the environment and underlying context) as suggested in the movement ecology paradigm by [35]. That is, geographic and environmental conditions can cause similar movement patterns and thus potentially be considered an important indicator for the identification of similarities in movement. Moreover, recent studies suggest that the collective dynamics of moving objects and their movement patterns can be also influenced by the interaction between different mobile agents and presence of factors such as competition, attraction, or avoidance [32]. Hence, it is essential to consider the external influential factors in movement analysis [17].

Integration of context in movement studies challenges the development of new methods for context-aware movement analysis [23]. There are many context-free approaches to measure the similarity of trajectories, e.g., [15] and references therein. An exception is the analysis of trajectories on road networks and subway systems [25, 38]. Here the known underlying network reduces the dimensionality of the problem and leads to more efficient algorithms and more meaningful results. However, for movement not constrained to a network, hardly any context-aware analysis algorithms exist. A notable exception is the work by [4] which uses an event based model as opposed to the geometric model we propose here. Besides geographic context, also temporal context can influence movement, e.g., people sleeping at night, and birds migrating in spring. Contrary to geographic context, time is a natural component of trajectories, and some similarity measures take time into account [9, 21, 34].

3   Modeling context

This section discusses models for geographic context that allow integration of context into the algorithmic analysis of movement data. We first present various types of geographic context that frequently occur and are relevant for movement data. For each type of context suitable data models proposed in GIScience [11] and their properties are described. In the

following section we will take one model, a polygonal subdivision, for integration into similarity analysis.

- **Network**. Some entities are constrained to move on a network, e.g., cars on roads, trains on tracks, boats on rivers, whereas other entities may be constrained to cross a network only at certain points, e.g., people on foot, or exhibit movement in an open space, e.g., bird movements.

  Model: labeled geometric graph.

- **Land cover**. The type of land cover influences for instance the speed of an object, e.g., a hurricane is faster on water than on land.

  Model: labeled polygonal subdivision.

- **Obstacles**. Some parts of geographic space are impassable for some entities, e.g., lakes for cars or pedestrians.

  Model: set of polygons.

- **Terrain**. The slope and altitude of a location influence movement, e.g., cyclists are faster downhill than uphill.

  Model: grid or TIN.

- **Ambient attributes**. Geographic or environmental attributes, such as weather conditions, orographic, and thermal uplifts.

  Model: point, grid, vector data, or annotated information.

- **Time**. Moment or duration in temporal space as indicators of temporal patterns, such as season, moon phase, time of day.

  Model: attribute.

- **Other agents** Presence of other agents can cause the emergence of certain movement patterns (e.g., attraction and competition among animals lead to particular behavioral patterns such as courtship or fighting, respectively).

  While other agents definitely influence movement, we will not discuss them further in this paper, since they are not a form of geographic context.

Obstacles may be part of a network or land cover. Obstacles and attributes can also be modeled as labeled polygonal subdivisions. Here, obstacles are modeled as a subdivision of obstacles and non-obstacles. Attributes are modeled as a subdivision into zones of equal attribute values. Also, several types of geographic context (e.g., land cover, properties of the terrain, attributes) can be treated as further attributes of a trajectory. That is, each point of the trajectory can be annotated by the geographic context value, e.g., type of land cover, slope, temperature. However, this will not reveal if two points are in the same zone of attribute values, i.e., the same region of land cover, or slope, or temperature.

Context may be *discrete* or *continuous*, i.e., it takes on discrete values, such as land cover, or continuous values, such as temperature. This distinction plays a role when comparing different contexts. However, trajectories are typically discrete themselves. Also, the influence of context depends on the scale at which it is considered [24]. We discuss this further in Section 5. Furthermore, context may be *dynamic* or *static*, i.e., it may change over time or not. A changing context needs to be taken into account when comparing trajectories that occurred at different times, i.e., by using appropriate context values for each trajectory.
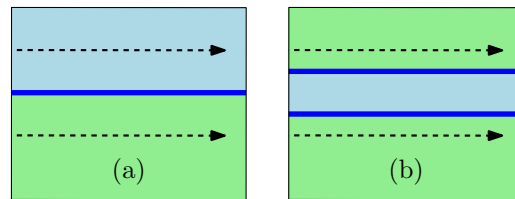
Figure 2: Trajectories (in dashed, black) distinguished by geographic context (boundaries shown in bold, blue).

Context influences movement in different ways. We can distinguish whether context *limits* or *enables* movement, and whether it does so *fully* or *partially*. For instance, a road network may limit movement *fully* or *partially*, e.g., a car will always stay on it, whereas a tractor may leave it to go on a field. When we know a context has full influence we can use it, for instance, to detect outliers. In our approach, we will not explicitly make use of the distinctions *limit/enable* and *partial/full*. This would be an interesting path for future work.

Our framework is based on the movement paradigm by [35]. A moving agent has an internal state (why move?), a navigation capacity (when and where to move?) and a motion capacity (how to move?), and it is influenced by external factors (the environmental context). These four components interact to produce the movement path. Although the movement paradigm by Nathan et al. [35] was originally introduced for movement ecology, we believe that a similar paradigm also applies to other domains. In particular, this is the case for hurricane movements. A hurricane only forms under a favorable climatic conditions in terms of wind speed, air pressure, and sea surface temperature. It cannot form on land, or on cold water. Usually, when a hurricane hits land, its movement direction and speed changes. Therefore, the internal state, navigation, and motion capacities of hurricanes are highly related to external factors (ambient attributes and geographic context).

4   Context-aware similarity measures

Our goal is to define a similarity measure that takes into account the geographic context. For this, we first ask: "How does geographic context influence the similarity of trajectories?" We claim that a fundamental influence of geographic context is that it may distinguish trajectories. For example, context distinguishes the hurricane tracks in Figure 1. Figure 2 shows the most basic situations that can occur: in (a) two entities are moving in areas of different context, e.g., one on water, the other on land. In (b) two entities are moving in areas of the same type of context (e.g., land), but are separated by a region of different context (e.g., a river). These trajectories may appear similar when the geographic context is *not* taken into account, but they differ when it is.

In our approach, we mainly consider geographic context that is modeled as a labeled polygonal subdivision. As discussed in the previous section, this may model land cover, obstacles, or attributes aggregated to zones. Thus it is an important model covering many types of geographic context. In particular, land cover is an essential context in animal ecology. Not covered by this model are networks and terrains. Nevertheless, the model can be adapted for terrains using classification (e.g., topographic contour lines). For network-constrained data, approaches of trajectory similarity exist [27, 39]. For attributes where an

aggregation to zones would not be meaningful, a weighted multi-dimensional approach can be taken (see Section 6).

Note that geographic context has further implications on similarity than distinguishing trajectories with different contexts. Geographic context may influence attributes of a movement paths (such as speed or sinuosity), which influence the similarity. For instance, a person typically walks slower on sand than on a road. Thus, two spatially close trajectories on sand/road may not be considered similar when using a speed-dependent similarity measure. Here, we address similarity measures that distinguish trajectories with differing contexts. We see this as the most fundamental and general influence of geographic context on trajectory similarity.

Problem statement. We are given two trajectories, and a labeled polygonal subdivision of the area in which the trajectories move. We want to define similarity measures for trajectories that take into account the context modeled by the subdivision.

4.1 Approaches

A trajectory in our setting has a spatio-temporal part (its position in time and space) as well as a contextual part (its position in the labeled polygonal subdivision). Generally, we see three approaches to context-aware movement similarity analysis: these two parts (i.e. trajectory and context) can be treated as:

(1) *equal* and similarity computed in multi-dimensional space,
(2) *independent* and similarity computed separately,
(3) *integrated* and similarity computed in an integrated way.

Next, we briefly discuss the first two approaches and compare all three approaches. We conclude that an integrated approach is most suitable and give a solution for this in Section 4.2.

For the equal approach, the context parts (position in a labeled polygonal subdivision) are mapped to numerical values for a (possibly weighted) multi-dimensional analysis. Note that typically no straightforward such mapping will exist. Here, the mapping (and possibly weighing) determines the relative weight of context vs. space and time.

For the independent approach, the trajectory is split into two: a (context-free) spatio-temporal trajectory, and a (pure) context trajectory. The context trajectory would be the sequence of labeled cells of the subdivision that the trajectory visits (and corresponding time stamps). For example, consider the (short) trajectory:

$$\langle (x_1, y_1, t_1, C_1), (x_2, y_2, t_2, C_1), (x_3, y_3, t_3, C_3) \rangle$$

Its spatio-temporal part is $\langle (x_{a1}, y_{a1}, t_{a1}), (x_{a2}, y_{a2}, t_{a2}), (x_{a3}, y_{a3}, t_{a3}) \rangle$ and its context part is $\langle C_1, C_1, C_3 \rangle$. Known similarity measures can be applied to the spatio-temporal trajectory and the context trajectory separately. This gives two distance values: a spatial distance and a context distance. These can then be combined using an additive (weighted sum) or multiplicative (weighted average) approach, or one distance can be used as filter for the other.

Comparison of approaches. We claim that the equal approach is not appropriate for two reasons. First, mapping context (given as labels of a polygonal subdivision) to numerical
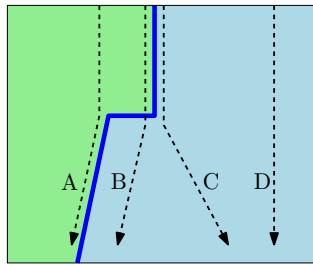
Figure 3: Trajectories (in dashed, black) over geographic context (boundaries shown in bold, blue).

values loses information. For instance, how to meaningfully map land cover types to numerical values? Second, space, time, and context are not equal. The independent approach applies only when location and geographic context are independent of each other, which they seldom are. Hence, we claim that an integrative approach should be chosen.

Consider, for instance, the (abstract) situation in Figure 3. Four trajectories A, B, C, D are shown over a subdivision of two cells. Trajectories A, B are closest spatially, but differ in context. Trajectories C, D are close with respect to context, but differ spatially. Trajectories B, C differ, when considering context and space separately. However, when considering space and context jointly, trajectories B, C are the most similar. Trajectories B, C are at first close spatially, but separated by context, then they are close in context, but with a larger spatial distance. As another example, consider two hurricanes with similar paths, but different points of landfall, due to one hurricane traveling along the coast before making landfall. These would be considered more similar under an independent than an integrated approach. In the following Section 4.2, we develop similarity measures in an integrative approach.

4.2   Integrated similarity measures

Now we show how to extend existing similarity measures to make them context-aware. For this, we integrate contextual and spatial distance. The main idea is to define the distance between two points as their spatial distance plus their context distance. Intuitively, this means it "costs" to cross context boundaries. Points with equal contexts, i.e., in the same cell of the subdivision, will get zero context cost. Thus, for equal context the distance becomes the spatial distance.

Note that adding costs only makes sense if the two costs have comparable scales. That is, we require to be able to combine spatial and context distance. If the spatial and context distance are incomparable, then an integrative approach, which outputs one distance value, seems infeasible. For convenience, we introduce a scaling parameter for the context distance, which we call the *context weight*. This allows us to first define a context distance, and then relate it to the spatial distance by setting the context weight. The value of the context weight will be determined by the application, and choosing it appropriately is critical for the analysis. In our experimental evaluation in Section 6 we discuss this in more detail. Specifically, we suggest to choose the context weight based on an interpretation of it, and running the analysis at different scales to evaluate the effect of the context weight.

Based on this notion of integrated point-to-point distance, we propose a framework consisting of three ingredients:

(1) a spatial distance, e.g., Euclidean distance,
(2) a context distance (see below), and
(3) a distance measure based on point-to-point distances, e.g., Hausdorff, Fréchet, or equal time distance.

Choosing all of these ingredients results in a context-aware similarity measure for trajectories. That is, our approach extends known distance measures (3) to make them context-aware, by adding a spatial distance (1) and a context distance (2). If all three ingredients are metrics, so is the resulting measure.

In this approach, we take into account that space, time, and context are not equal. We use time in the overall distance measure to determine the matching of points on the trajectories. The relative weight of space and context is determined by the scaling parameter. The framework also allows to integrate several context parameters, e.g., land cover and slope. For this, we would simply add several context costs.

Next, we first discuss different options for a context distance. Then, we discuss how to compute the resulting context-aware similarity measures.

In this study we consider three popular measures for trajectory similarity that are based on spatial point-to-point distances: comparing distances at equal times [9, 21, 34], the Fréchet distance [6–8], and the Hausdorff distance [12, 36].

4.2.1 Context distance

We propose to use a cost between cells of the subdivision as context distance. That is, similar context is determined not by the specific point in a cell but by the cell itself. An alternative would be to use a context distance between points. A possible disadvantage of a distance between cells is that this will ignore "islands," see Figure 4 (a). The two shown trajectories, though separated by a cell (e.g., island) of the subdivision, still lie in the same cell. Thus, their context distance will be zero. A context distance between points could, for instance, consider the context along (shortest) paths between the points and thus detect the island. In some applications, this may be more meaningful.

An advantage of a distance between cells, however, is that the resulting distance measure is a metric. This property is not necessarily maintained when choosing a context distance between points, see Figure 4 (b). Suppose as context distance between two points we add a cost for each cell boundary of the subdivision crossed by a shortest path between the points. Then, as Figure 4 (b) shows, this context distance does not fulfill the triangle inequality (the distance between the two outer trajectories will be less via the middle trajectory than directly). To remedy this, one could use *geodesic shortest paths* between points, i.e., allow paths to go around islands. This again, may lead to "jumping" over islands, that is, islands may increase the spatial distance but not the context distance.

Summarizing, we propose a distance between cells, because it gives an intuitive and sound definition. In particular, it handles the cases in Figure 2 and Figure 3.

Choices. For measuring similarity between cells of the subdivision we have two independent choices to make, resulting in four different context distances between cells. The choices are:
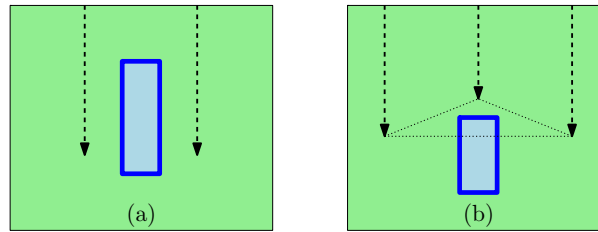
Figure 4: Trajectories (in dashed, black) over a geographic context modeled as polygonal subdivision (in bold, blue).

> **labels**: unit cost or cost dependent on label of the cell,
> **paths**: unit cost or cost of shortest path between cells.

The first choice refers to whether we assign a cost depending on the labels of the subdivision, or not. A unit cost means the cost between cells does not depend on the label. Alternatively, the cost may depend on the label. For instance, imagine the subdivision models land cover. Then we may choose to give a higher cost between *grass* and *water* than between *grass* and *wood*. If we choose a cost dependent on the labels, we still want to maintain the triangle inequality. That is, we choose costs $c(L_1, L_2)$ between labels L_1, L_2 such that for all three labels L_1, L_2, L_3 holds $c(L_1, L_3) \leq c(L_1, L_2) + c(L_2, L_3)$. A unit cost would assign an equal cost to all different labels (and zero to equal labels). This makes sense, when the relation between labels is not known.

The second choice refers to whether we assign a cost depending on the distance of the cells in the subdivision, or not. A unit cost means the cost between cells does not depend on the distance. Alternatively, the cost may reflect the length of a shortest path between the cells in the subdivision. The former only distinguishes whether two agents are moving over the same type of context (e.g., land cover), while the latter would also distinguish between different cells of the same type, (e.g., same land cover but separated by an obstacle, as in Figure 2 (b)). For a cost based on path distance we consider the *dual graph* of the subdivision. That is, we consider the graph, where each cell C constitutes a vertex of the graph, and edges exists between neighboring cells. A shortest path then refers to a path of minimal cost, where the cost of each edge is determined by the first choice, that is, either a unit cost, or a cost depending on the labels of the cells. An example of this is shown in Figure 5: the left hand shows a subdivision in four cells C_1, C_2, C_3, C_4 with labels corresponding to land cover (in blue), and the dual graph of the subdivision (in black). The right hand shows the distance matrix between cells. For instance, the distance between C_2 and C_3 equals $c* = min(c1 + c3, c2 + c4)$ because there are two possible paths via C_1 or C_4, the "cheaper" of which is chosen.

4.2.2   Computation

The proposed context-aware similarity measures can be computed by extending algorithms for Hausdorff, Fréchet and equal time distance in three ways:

(1) computing the context distance matrix (if using path lengths between cells),
(2) locating points in the subdivision and (if necessary) refining trajectories,

(3) adding context costs when computing the distance measure.

The first two steps are pre-processing to the main algorithm in step 3.

Computing the context distance matrix. Theoretically this is known as the *all-pairs shortest paths* problem on a planar graph. The fastest known algorithms for this problem on planar graphs run in (sub)quadratic time. However, in practice, a slower, but simpler algorithm, e.g., the well known Floyd-Warshall algorithm with a cubic execution time may be preferred. In particular, this holds if the size of the subdivision is (much) smaller than the size of the trajectories, and the algorithm in step 3 dominates the execution time.

Locating points in the subdivision and refining trajectories. If we do not need to refine trajectories, then we only need to compute in which cells the vertices of the trajectories lie. For this we can use a standard point location data structure like a trapezoidal map. Computing this data structure takes $O(m \log m)$ preprocessing time. A point location query, that is reporting the cell of a given trajectory point, then takes $O(\log m)$ time for a subdivision of size m, with $O(m)$ space requirements. Thus, this takes $O(n \log m)$ time for a trajectory of size n.

If we need to refine the trajectories, we also need to find all intersections of the trajectory with subdivision boundaries. For this, we use known algorithms to preprocess the subdivision for *ray shooting queries*: given a point for which we know the location and a ray starting at that point, we want to know where this ray intersects the subdivision. We can do the preprocessing step in $O(m)$ time, where m is the size of the subdivision. The queries then take $O(\log m)$ time. We can locate the first vertex of the trajectory in $O(\log m)$ time using a point location data structure. Then we can find the intersections of the first trajectory edge and the location of the second vertex of the trajectory using a ray shooting query from the first vertex in the direction of the second vertex. If we intersect a cell boundary before reaching the second point, we continue from there. After reaching the second vertex, we process the remaining trajectory in the same way. The running time of this step is $O(h \log m)$ per shoot for h intersections. Thus, we need $O((n + h) \log m)$ time in total.

In practice, we expect that trajectories do not intersect the subdivision very often. In this case, also simpler strategies apply, for example as described in the next section.

Adding context costs when computing the distance measure. Algorithms for the Hausdorff and equal time distance are straightforward to extend by simply adding the context

Distance matrix:

	C_1	C_2	C_3	C_4
C_1	0	c_1	c_1	c_3
C_2	c_1	0	$c*$	c_2
C_3	c_1	$c*$	0	c_2
C_4	c_3	c_2	c_2	0

where $c_1 = c(\text{forest, grass})$, $c_2 = c(\text{grass, water})$, $c_3 = c(\text{forest, water})$ and $c* = \min(c_1 + c_3, c_2 + c_4)$
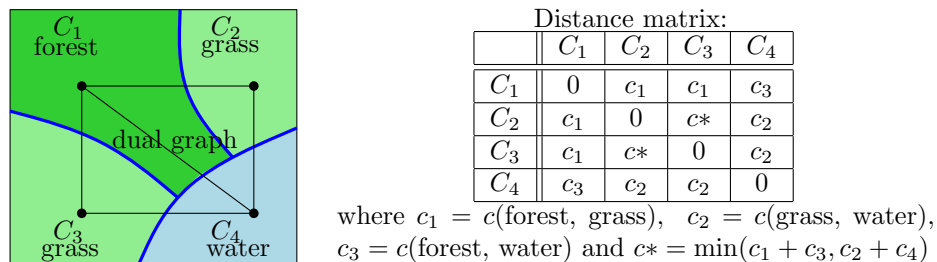
Figure 5: Example of context distance along shortest paths.

cost to the spatial cost. For the Fréchet distance, the decision algorithm based on the *free space diagram* [2] can be extended as follows. The description assumes familiarity with the original algorithm; we omit further details here but refer for these to [2, 7]. First, we need that the trajectories are segmented at context boundaries, as described in the preprocessing above. Then each trajectory edge lies completely in one subdivision cell. With this, each free space cell (corresponding to two trajectory edges) receives a constant extra context distance. To extend the algorithm, we simply add this in each cell. The execution time of this algorithm does not change: $O(n^2 \log n)$. Note that we have to refine the trajectories, thus their complexity may increase. However, we expect to not add more than a linear number of intersection points, which will not affect the asymptotic execution time. For computing the Fréchet distance, a set of critical values is searched, employing the decision algorithm in each step [2]. Critical values are distances between points on the trajectories. Here, we can again simply add the context distance.

4.2.3   Fréchet distance in weighted regions

In a related approach, the Fréchet distance is extended for weighted regions [13]. In this approach, the cost of a path is modeled as the weighted sum of length of the path in each weighted region. Our model of adding context costs when crossing context boundaries can be "simulated" by their model, as follows: give each context boundary a width ϵ (for some small $\epsilon > 0$) and weight $(c_i + 1)$. Give each cell the weight 1. Then a path of length ℓ crossing b boundaries has weight $\ell + \sum_{i+1}^{k} c_i$, which equals the length of the path plus context cost of the path (in our model). Thus, we could use their algorithms for our approach. However, their algorithms give approximate solutions and have much higher running times (more than $O(n^4)$).

5   Robustness to small changes in context

Our context-aware similarity measures combine a spatial distance and a context distance. Whereas the spatial distance changes continually, context may change abruptly (e.g., type of land cover). The similarity measures we have defined so far, take changes in context into account independent of their duration. This may lead to unwanted effects, in particular, when abrupt, but brief changes occur, e.g., due to noise in the data. In this section, we discuss how to make our similarity measures more robust to brief changes in context.

Let us first consider such changes. In some cases, a brief change in context may be significant, but often it is not. In the example of hurricanes, the following two scenarios occur, illustrated in Figure 6:

- a hurricane moves over an inlet after landfall; and
- a hurricane moves over the tip of a body of land.

Both cases have a significant impact on the context trajectory, although they are likely not relevant to the hurricane. In particular, a small change in the geographic path would alter the context distance significantly, e.g., if the hurricane made landfall just next to the inlet, or moved just under the tip of body of land.

For instance, if we look again at the hurricanes Erin and Katrina (see Figure 9, right). We would intuitively say that these have equal context with respect to land/sea. However, hurricane Katrina moved over an inlet after hitting land near New Orleans. Thus,

her context (on a fine scale) differs from that of Erin. This implies, that for a large context weight, hurricanes Katrina and Erin will be considered different by our context-aware similarity measures. However that Katrina moved over an inlet is insignificant for her path. Therefore, we would like our similarity measures to be more robust to such brief changes in context. That is, we would like to be able to consider context at a coarser scale. Note that this problem does not occur for spatial distance, which changes continuously.

There are several possibilities to take into account the duration of a change in context (and reduce the influence of brief changes) in our similarity measures. The duration can be taken into account in the annotation process of the context variable, as well as in the computation of context distance, or both. Generally, the idea is to aggregate the context in a neighborhood of a point. Next, we discuss several options of choosing a neighborhood, and of assigning an aggregated context value in a neighborhood. Some of these are more closely tied to the annotation process, and others more to the computation process.

For the neighborhood of a point we consider two options, as illustrated in Figure 7:

- a geographic neighborhood (e.g., a spatial disc around each point), and
- a neighborhood along the trajectory.

Using a geographic neighborhood needs to be incorporated in the annotation process. Using a neighborhood on the trajectory does not require further annotation. Both types of neighborhoods have their applications in different cases. Sometimes the context for a geographic neighborhood of the trajectory may be available, for instance in the case of land/sea for hurricanes. However, sometimes it may not, and one may have only an annotated trajectory in hand. For instance, this is the case for a hurricane annotated by internal wind speed or an animal trajectory annotated by activity type.

The choice of neighborhood size determines the scale we wish to consider. However, we need to take into account that the size of neighborhood, and the methods we propose, are influenced by the granularity of the data. In particular, choosing a neighborhood on the trajectory is restricted by the sampling granularity of the trajectory. Furthermore, a trajectory neighborhood should be defined in terms of space or temporal duration, rather than number of sampling points, which coincide only for regular sampling.

Given the context values in a neighborhood of a point, we propose three options of assigning an aggregated context value for the point, namely by choosing:

- the predominant context value,
- the weighted average context value, or
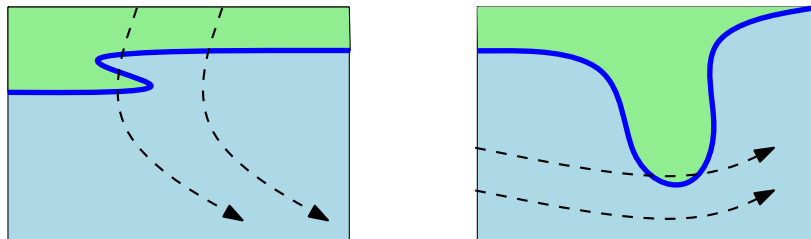- the minimized context cost.

Figure 6: Trajectories (in dashed, black) over a geographic context (in bold, blue).

for the point. For the predominant context value, we simply take the context value of the majority of points, possibly with a weighing of points (e.g., giving the actual point a higher weight). For the weighted average, we compute an average context value, where the weights correspond to the distribution of context type in a neighborhood. Note, that in the case of non-continuous context values (e.g., land cover in contrast to temperature), this leads to d-dimensional context values where d is the number of context classes. A predominant or weighted average context value can be assigned during the annotation process or as a preprocessing step (as well as during computation). A minimized context cost of two points is the minimum context distance over all pairs of context values in the neighborhoods of these points. Such a minimized context cost can only be chosen in the computation process.

An advantage of the minimized context cost is that it can (by definition) only decrease with growing neighborhood sizes. That is, the minimized context cost is monotone decreasing with respect to neighborhood sizes. Selecting the predominate or weighted averaging value do not follow this rule. That is, with growing neighborhood sizes, the context distance obtained from these values may increase or decrease. A weighted average context value is most suitable for continuous context values (e.g., temperature) that is sparse but relatively precise. For data with faraway outliers or abrupt changes in context, averaging is less applicable. Here, choosing a predominant context value is more robust.

For a minimized context cost, the resulting distance measure is no longer a metric (the triangle inequality is no longer fulfilled). Figure 8 shows examples, where this is the case, and also illustrates the difference between a geographic neighborhood and a neighborhood on the trajectory. Using either neighborhood type, the middle trajectory in the left figure will be considered similar (in context) to both the lower and the upper trajectory, whereas the upper and lower trajectories are clearly different (in context). In the right figure, this happens only using geographic neighborhoods, not trajectory neighborhoods.

Note that in all cases, the resulting context value depends on the size of the neighborhood we choose, as well as the granularity of the data. When choosing the predominant context value or minimizing the context cost, the size of neighborhood can be thought of as the scale at which we ignore changes in context. For an averaged context value, the neighborhood size can be thought of as "width" of smoothing. Again, we stress that reasonable neighborhood sizes heavily depend on the data granularity.

6 Evaluation on hurricane and albatross data

In the previous section, we proposed context-aware similarity measures for trajectories, which extend known measures. Furthermore, we discussed how to make these robust
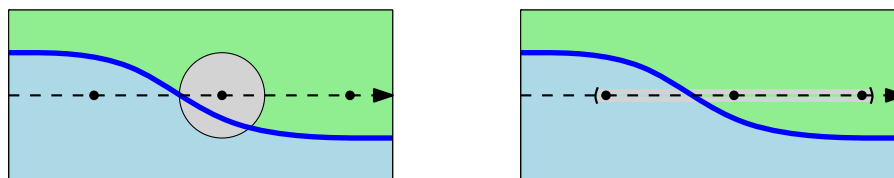

Figure 7: Trajectory neighborhoods: geographic (right) or along the trajectory (left).

to small changes in context. We implemented these measures for the Fréchet distance, and tested them on two data sets: hurricane and albatross tracking data. Our aim is to demonstrate the effect and usefulness of our approach. For this, we chose two different data sets: hurricanes with land cover (land/sea) and albatross with wind speed as context. For the hurricanes, the context is given in form of a (simple) labeled subdivision, and we use a distance between cells of this division as context distance. For the albatross, the context is given as continuous attribute to the data, and we us difference in wind speed as context distance. Next, we first describe the similarity measures used, the experiments on hurricane, and then on albatross data.

6.1 Similarity measures

We used context-aware Fréchet distance at different context weights. On the hurricane data we also tested several robustness extensions. We chose Fréchet distance as a distance measure that compares the shape of the tracks well. For the context distance of hurricanes, we chose a unit distance between land/sea (the only option between only two labels) and a shortest path distance between cells (here, paths had length at most two, with only one large sea cell). For the context distance of albatross, we used difference in wind speed. In both cases, we varied the value of the context weight at similar magnitudes as the spatial distance (see below). Thus, in the terminology of Section 4, for hurricanes we used the following ingredients: (1) Euclidean distance, (2) shortest path distance with unit costs between different labels, and different context weights, (3) Fréchet distance.

Context weight. Recall that the context weight is used to weigh the context distance, thus putting the spatial and context distance in relation. In particular a context weight of zero implies ignoring context. One can interpret the context weight as follows: Two hurricanes with spatial distance close to zero but differing context are considered as similar as two hurricanes with equal context and a spatial distance of the value of the context weight. In our experiments, we used the context weights 0, 300, and 500 for hurricanes, and 0, 10, and 20 for albatross. These values were chosen as representative values of similar magnitude as the spatial distances of the hurricanes (see Table 1 and Table 3).

Robustness. For the hurricane data we compared choosing the predominant context value and minimizing the context cost in a neighborhood along the trajectory. Since we had only relatively sparse trajectory neighborhoods available, we did not average context values. We used trajectory neighborhoods of 3, 5, and 7 trajectory points.
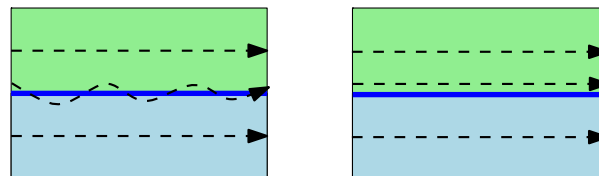

Figure 8: Trajectories (in dashed, black) over a geographic context (in bold, blue).
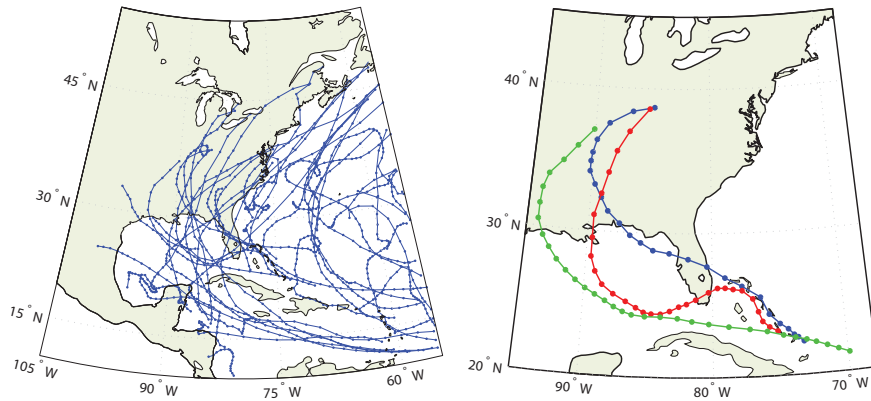
Figure 9: Left: Data set. Right: Hurricanes Erin 1995 (blue), Katrina 2005 (red), and Rita 2005 (green).

.

6.2   Hurricane data

Context.   For hurricanes, similarity is an interesting analysis task, which is for instance relevant for predicting hurricane paths (see Section 1). Hurricanes are known to be influenced by geographic context, in particular land/sea. Important geographic context factors for hurricanes can be distinguished as follows:

- **external factors**: temperature, barometric pressure, land/sea, topography
- **internal factors**: intensification, wind speed, move speed, diameter

In our tests we used land/sea as an important geographic context.

Data set.   We considered hurricanes in the North Atlantic Basin in the years 1995, 2004, and 2005. The data was obtained from NOAA National Hurricane Center[1]. The hurricanes are tracked every 6 hours (00:00, 06:00, 12:00, 18:00). The chosen years had predominant hurricane activities with 17 storms in 1995, 11 storms in 2004, and 20 storms in 2005, thus 48 in total. Furthermore, we used a geographic data set containing the coast lines for the polygonal subdivision into land/sea.

Preprocessing.   We cut the hurricanes at longitude 55° W at start and end, to ensure that entire hurricanes locate in a similar spatial region. Large differences in starts and ends would otherwise dominate the distance value. The data set is shown in Figure 9 (a). Next, we located and annotated trajectory points in the subdivision and computed intersection points of the trajectory with the coastlines. For this, since we have a sparse subdivision, we first split the coast line into constant size pieces. Then we build an R-tree of bounding boxes of these pieces, and query in this structure.

---

[1]http://www.nhc.noaa.gov

Similarity analysis.   We analyzed the similarity of the hurricane data using the context-aware similarity measure described above. That is, we computed the similarity of all pairs of hurricanes in this set, i.e., the complete distance matrix, using context-aware Fréchet distance at different context weights and with and without different robustness extensions. From this, we extracted the most similar pairs of hurricanes, and compared these for different measures.

First we computed the context-aware Fréchet distances at different context weights without robustness extensions, since the data did not contain noise or outliers. However, as we see below, we will need robustness to small but abrupt changes in context, that have no significant influence on the hurricanes' paths. We describe the results of these experiments below.

Varying context weights.   Table 1 shows the context-aware Fréchet distances for three different context weights (0km, 300km, 500km) of the nine spatially closest pairs of hurricanes. Besides the computed distance values (distance), for easier comparison the table also gives the ranks (rank) according to the order by these distance values. That is, for each context weight, we order hurricane pairs by increasing distances, and rank k means that a pair of hurricanes is the kth in this order. Thus, the rank gives an indication of relative distances. For instance, Chantal–Luis and Chantal–Irene move higher up in the order (have smaller rank) for larger context weight. This is because they have equal context (land), and therefore the context-aware distance is not affected by the context weight (see their distance values). Note that rank "-" indicates a rank lower than 10; we only show higher ranks to indicate high similarity of these hurricanes.

H1	H2	context	rank (0)	rank (300)	rank (500)	distance (0)	distance (300)	distance (500)
Harvey	Nate	water	1	1	1	248km	248km	248km
Alex	Gaston	water/land	2	4	-	315km	561km	761km
Chantal	Luis	water	3	2	2	489km	489km	489km
Erin	Katrina	water/land	4	5	-	482km	614km	814km
Alex	Ophelia	water/land	5	-	-	531km	808km	1008km
Erin	Rita	water/land	6	-	-	550km	850km	1050km
Chantal	Irene	water	7	3	3	551km	551km	551km
Katrina	Rita	water/land	8	-	-	572km	698km	898km
Allison	Wilma	water/land	9	-	-	615km	872km	1072km

Table 1: The context-aware Fréchet distances and ranks of the nine spatially closest pairs of hurricanes for three different context weights.

Half of the hurricanes in Table 1 go only over water. For these, context distance is zero, thus the summed difference does not increase. For our analysis, hurricanes crossing land are more interesting. We find that hurricanes with differing context, such as Hurricanes Rita and Katrina in Figure 9 (right), are distinguished well by our method when adding a context cost. However, hurricanes, which have changing, but similar context, such as Hurricanes Erin and Katrina in Figure 9 (right) remain more similar: they have the smallest distance and rank among all hurricanes with a water/land context. To see that this makes

sense, consider the track and intensity development of these two hurricanes. They first cross Florida before re-emerging in the gulf and making secondary landfall. One can see that they weaken over land, then re-intensify as they move back over the water. However, Katrina had more time to re-intensify over the warm waters of the Gulf of Mexico before making landfall again. In contrast, Hurricane Rita does not cross over Florida, making its first landfall between Louisiana and Texas. Hence, Rita had a very different development than Erin and Katrina.

However, note that Erin and Katrina still have an increasing distance with increasing context weight. This is due to a small change in context, which is only visible at a fine scale. Next, we show how this is handled by our robustness extensions.

Robustness. We tested our robustness extensions on the three hurricanes Erin, Rita, and Katrina, which exhibit different context similarities at different scales. At a fine scale, all three hurricanes have different (land/water) contexts, see Figure 9 (right).

- Erin: water – land (Florida) – water – land
- Katrina: water – land (Florida) – water – land – water (inlet) – land
- Rita: water – land

At a coarser scale, the inlet disappears. At an even coarser scale, also Florida may disappear; in particular, this will happen if we use trajectory neighborhoods. Intuitively, we would like to slightly decrease the scale, so that only the inlet disappears, and Erin and Katrina will be similar, but not Rita.

H1	H2	distance non-robust	distance 3-min	distance 5-min	distance 3-pred	distance 5-pred	distance 7-pred
Erin	Katrina	c+163 km	482 km	482 km	482 km	c+300 km	482 km
Erin	Rita	c+462 km	1302 km	549 km	c+462 km	c+462 km	633 km
Rita	Katrina	c+320 km	571 km	571 km	c+250 km	571 km	571 km

Table 2: Distances between the hurricanes Erin, Katrina, and Rita for a context weight $c \geq 900$. Here, k-min/pred refers to a minimized context cost/predominant context value in a neighborhood of size k.

Our results for the robustness extensions on these hurricanes are summarized in Table 2. We see that the non-robust Fréchet distance (labeled "non-robust") distinguishes between all three hurricanes. Choosing a minimized context cost results in equal contexts and thus stable distances for all pairs of hurricanes already in a size 3 neighborhood (labeled 3-min in the table). Note that the large, but stable distance between Erin and Rita comes from the fact that minimization allows to create false "islands." Choosing the predominant context value in a neighborhood results in equal contexts and thus stable distances for all pairs of hurricanes only in a size 7 neighborhood (labeled 7-pred in the table). A predominant context value for neighborhoods of size 3 and 5 displays the non-monotonicity of the measure. A predominant context value for size 3 results in the intuitive similarity of Erin and Katrina, but not Rita.

The granularity of our data is relatively sparse compared to the scale at which we would like to ignore changes in context. For large neighborhood sizes (3 for a minimized context
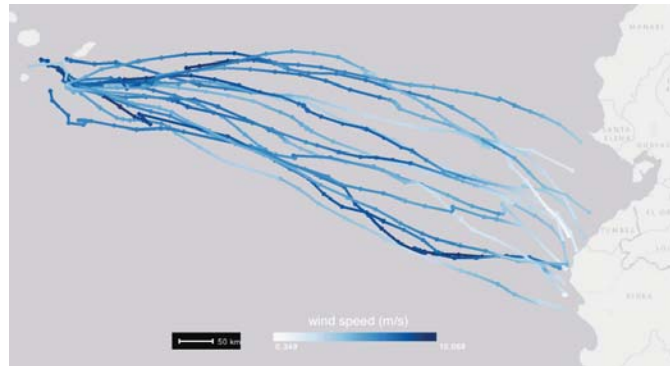
Figure 10: Nine Galapagos albatrosses flight between the Galapagos Islands (i.e., nesting site) and the Peruvian coast (i.e., foraging site). Tracks are annotated with wind speed (m/s): dark blue represents higher and light blue lower wind speed.

distance, 7 for predominant context values), we not only ignore the inlet but also Florida. This demonstrates that neighborhood sizes need to be chosen at the magnitude at which one wants to ignore changes in context. In particular, the granularity of the data restricts this magnitude.

6.3   Albatross data

In the second case study, we investigate the movement similarity of nine Galapagos Albatrosses tracked from June to September 2008 at sampling interval of 90 minutes. These albatrosses make extensive movements between the Galapagos Islands (i.e., nesting site) and the Peruvian coast (i.e., foraging site) [16]. The Movebank Env-DATA System (environmental data automated track annotation system) was used to link albatross movement tracks to wind datasets [16]. Using the Movebank Env-DATA service, the nine albatross tracks were first annotated with wind speed (m/s) computed from u- and v-wind components obtained from the NCEP Reanalysis 2 dataset[2] [16]. Figure 10 visualizes the nine tracks annotated with wind speed. As seen in the figure, albatrosses encounter a varied wind speed pattern and mostly are challenged by the wind along their outbound flights to the Peruvian coast [16, 42]. Thus here we study the effect of wind speed on the albatross outbound flight paths from the Galapagos to the coast.

Preprocessing and analysis   We (manually) segmented the tracks to obtain the flights from the Galapagos to the coast. This resulted in 16 flights in total. We computed the distance matrix for these flights using several context weights. For each scale we list the ten most similar pairs in Table 3 (flights are indexed 0 to 15).

Choosing context weight   We computed spatial distance in km and context distance (i.e., difference in wind speed) in m/s. Spatial distances for the ten most similar pairs of flights ranged between 35 and 54km (cf. Table 3). Wind speeds typically ranged between 0 and
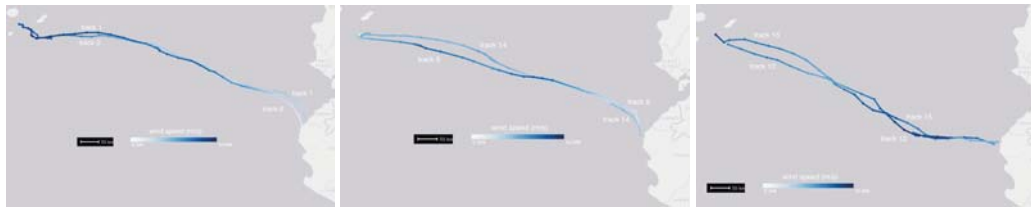
[2]http://www.esrl.noaa.gov

Figure 11: Flights of Albatross from the Galapagos to the coast. The left pair stays similar, the middle pair becomes less similar, and the right pair becomes more similar.

5m/s. Hence, we chose 10 and 20 as representative context values for a spatial and context distance of similar magnitude.

F1	F2	dist		F1	F2	dist		F1	F2	dist
9	12	35		9	12	59		13	14	96
9	15	38		9	15	68		0	1	98
6	14	39		13	14	69		9	12	102
0	1	42		0	1	70		9	15	106
13	14	42		4	10	72		12	15	110
4	5	49		12	15	76		5	10	110
4	10	49		0	14	79		7	9	116
0	14	50		6	14	82		6	14	127
12	15	53		5	10	84		4	11	128
6	13	54		4	5	91		0	14	132

Table 3: The ten pairs of flights with smallest context-aware Fréchet distances at context weights: 0 (left), 10 (middle), and 20 (right).

Results   Table 3 shows how the distances change with growing context weights. Some trajectory pairs (e.g., 9, 12; 9, 15; 0, 1; 13, 14) remain (relatively) similar for growing context weights. In contrast pairs 12, 15 and 5, 10 become relatively more similar; and pairs 4, 5 and 6, 14 become less similar. Figure 11 shows examples of these pairs.

7   Conclusion and future work

We propose to include geographic context in the analysis of movement data. Specifically, this study proposed context-aware similarity measures for trajectories. These measures extend known similarity measures, integrating a spatial and a context distance. The context distance is based on a subdivision modeling of the geographic context. The proposed measures were enhanced to be more robust to small changes in context that are not reflected in the spatial distance. The developed methods were tested on two different movement datasets (i.e., hurricane and albatross tracking data). The results suggests that our method is fast, simple, and effective. That is, it distinguishes trajectories by their spatial as well as contextual similarity.

We see several paths for future work. We would like to apply our ideas on integrating context into the analysis of movement in more knowledge discovery tasks such as trajectory clustering, simulation, movement pattern analysis. We plan to employ our method in further case studies on movement data, in particular taking into account geographic context in the form of different categories of land cover, habitat, and vegetation types. Context distances that take into account "islands" are an interesting open question as well, which we plan to consider further. Finally, we considered mainly polygonal subdivisions to model the geographic context. Although this covers many interesting types of geographic context, there are other types of context (e.g., continuous attributes, interactions with other agents) to be explored further.

Acknowledgments

References

[1] S. C. Ahearn, J. L. Smith, A. R. Joshi, and J. Ding. TIGMOD: An individual-based spatially explicit model for simulating tiger/human interaction in multiple use forests. *Ecological Modelling*, 140(1–2):81–97, 2001. doi:10.1016/S0304-3800(01)00258-7.

[2] H. Alt and M. Godau. Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry and Applications*, 5:75–91, 1995. doi:10.1142/S0218195995000064.

[3] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual Analytics of Movement*. Springer, 2013.

[4] G. Andrienko, N. Andrienko, and M. Heurich. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, 25:1347–1370, 2011. doi:10.1080/13658816.2011.556120.

[5] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data*. Springer, 2006.

[6] S. Brakatsoulas, D. Pfoser, R. Salas, and C. Wenk. On map-matching vehicle tracking data. In *Proc. 31st International Conference on Very Large Data Bases*, pages 853–864, 2005.

[7] K. Buchin, M. Buchin, and J. Gudmundsson. Constrained free space diagrams: A tool for trajectory analysis. *International Journal of Geographical Information Science*, 24:1101–1125, 2010. doi:10.1080/13658810903569598.

[8] K. Buchin, M. Buchin, J. Gudmundsson, M. Löffler, and J. Luo. Detecting commuting patterns by clustering subtrajectories. *International Journal of Computational Geometry and Applications*, 21(3):253–282, 2011. doi:10.1142/S0218195911003652.

[9] K. Buchin, M. Buchin, M. van Kreveld, and J. Luo. Finding long and similar parts of trajectories. *Computational Geometry: Theory and Applications*, 44(9):465–476, 2011. doi:10.1016/j.comgeo.2011.05.004.

[10] M. Buchin, S. Dodge, and B. Speckmann. Context-aware similarity of trajectories. In *Proc. 7th International Conference on Geographic Information Science (GIScience)*, volume 7478 of *Lecture Notes in Computer Science*, pages 43–56, 2012.

[11] P. A. Burrough, R. McDonnell, P. A. Burrough, and R. McDonnell. *Principles of Geographical Information Systems*. Oxford University Press, Oxford, 1998.

[12] J. Chen, R. Wang, L. Liu, and J. Song. Clustering of trajectories based on Hausdorff distance. In *Proc. International Conference on Electronics, Communications and Control*, pages 1940–1944, 2011. doi:10.1109/ICECC.2011.6066483.

[13] Y. K. Cheung and O. Daescu. Fréchet distance problems in weighted regions. In *Proc. 20th International Symposium on Algorithms and Computation*, number 5878 in Lecture Notes in Computer Science, pages 97–111, 2009. doi:10.1007/978-3-642-10631-6_12.

[14] K. C. Clarke and L. J. Gaydos. Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, 12(7):699–714, 1998. doi:10.1080/136588198241617.

[15] S. Dodge. *Exploring Movement Using Similarity Analysis*. PhD thesis, University of Zurich, 2011.

[16] S. Dodge, G. Bohrer, R. Weinzierl, S. C. Davidson, R. Kays, D. Douglas, S. Cruz, J. Han, D. Brandes, and M. Wikelski. The environmental-data automated track annotation (Env-DATA) system: Linking animal tracks with environmental data. *Movement Ecology*, 1(1):3, 2013. doi:10.1186/2051-3933-1-3.

[17] S. Dodge, R. Weibel, and A.-K. Lautenschütz. Towards a taxonomy of movement patterns. *Information Visualization*, 7:240–252, 2008. doi:10.1057/palgrave.ivs.9500182.

[18] D. H. Douglas. Least-cost path in GIS using an accumulated cost surface and slopelines. *Cartographica*, 31(3):37–51, 1994. doi:10.3138/D327-0323-2JUT-016M.

[19] J. Elsner and A. Kara. *Hurricanes of the North Atlantic: Climate and society*. Oxford University Press, 1999.

[20] M. Erwig, R. H. Güting, M. Schneider, and M. Vazirgiannis. Spatio-temporal data types: An approach to modeling and querying moving objects in databases. *GeoInformatica*, 296:269–296, 1999.

[21] E. Frentzos, K. Gratsias, and Y. Theodoridis. Index-based most similar trajectory search. In *Proc. 23rd International Conference on Data Engineering*, pages 816–825, 2007. doi:10.1109/ICDE.2007.367927.

[22] A. Galton. Dynamic collectives and their collective dynamics. In *Spatial Information Theory*, volume 3693 of *Lecture Notes in Computer Science*, pages 300–315. Springer, 2005. doi:10.1007/11556114_19.

[23] C. Gschwend and P. Laube. Challenges of context-aware movement analysis—Lessons learned about crucial data requirements and pre-processing. In *Proc. 20th GIS Research UK Annual Conference (GISRUK)*, pages 241–246, 2012.

[24] C. Gschwend and P. Laube. Scale-effects in relating movement to geographic context. In *Proc. 8th International Conference on Geographic Information Science*, pages 51–56, 2014.

[25] Z. Gui, Y. Xiang, and Y. Li. Linear reference based similarity analysis between trajectories on road networks. *Journal of Computational Information Systems*, 22(8):9481–9489, 2012.

[26] K. Hornsby and M. Egenhofer. Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, 36(1):177–194, 2002.

[27] J.-R. Hwang, H.-Y. Kang, and K.-J. Li. Searching for similar trajectories on road networks using spatio-temporal similarity. In *Proc. Advances in Databases and Information Systems*, volume 4152 of *Lecture Notes in Computer Science*, pages 282–295, 2006. doi:10.1007/11827252_22.

[28] P. Laube. *Computational Movement Analysis*. Springer Briefs in Computer Science. Springer, 2014. doi:10.1007/978-3-319-10268-9.

[29] P. Laube and R. Purves. How fast is a cow? Cross-scale analysis of movement data. *Transactions in GIS*, 15(3):401–418, 2011.

[30] H. J. Miller. A measurement theory for time geography. *Geographical Analysis*, 4(2):63–45, 2005. doi:10.1111/j.1538-4632.2005.00575.x.

[31] H. J. Miller and J. Han, editors. *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, 2nd edition, 2009.

[32] J. A. Miller. Using spatially explicit simulated data to analyze animal interactions: A case study with brown hyenas in northern Botswana. *Transactions in GIS*, 16(3):271–291, 2012. doi:10.1111/j.1467-9671.2012.01323.x.

[33] D. Mountain. The dimensions of context and its role in mobile information retrieval. *SIGSPATIAL Special*, 3:71–77, 2011. doi:10.1145/2047296.2047311.

[34] M. Nanni and D. Pedreschi. Time-focused clustering of trajectories of moving objects. *Journal of Intelligent Information Systems*, 27:267–289, 2006. doi:10.1007/s10844-006-9953-7.

[35] R. Nathan, W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse. A movement ecology paradigm for unifying organismal movement research. *Proceedings National Academy of Sciences of the United States of America*, 105(49):19052–19059, 2008. doi:10.1073/pnas.0800375105.

[36] S. Nutanong, E. H. Jacox, and H. Samet. An incremental Hausdorff distance calculation algorithm. *Proc. VLDB Endowment*, 4(8):506–517, 2011. doi:10.14778/2002974.2002978.

[37] R. S. Purves, P. Laube, M. Buchin, and B. Speckmann. Moving beyond the point: An agenda for research in movement analysis with real data. *Computers, Environment and Urban Systems*, 47:1–4, 2014. doi:10.1016/j.compenvurbsys.2014.06.003.

[38] S. Sankararaman, P. K. Agarwal, T. Mølhave, and A. P. Boedihardjo. Computing similarity between a pair of trajectories. *Computing Research Repository (CoRR)*, abs/1303.1585, 2013. http://arxiv.org/abs/1303.1585.

[39] E. Tiakas, A. Papadopoulos, A. Nanopoulos, Y. Manolopoulos, D. Stojanovic, and S. Djordjevic-Kajan. Searching for similar trajectories in spatial networks. *Journal of Systems and Software*, 82(5):772–788, 2009. doi:10.1016/j.jss.2008.11.832.

[40] W. Tobler. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2):234–240, 1970. doi:10.2307/143141.

[41] P. Torrens, A. Nara, X. Li, H. Zhu, W. Griffin, and S. Brown. An extensible simulation environment and movement metrics for testing walking behavior in agent-based models. *Computers, Environment and Urban Systems*, 36(1):1–17, 2012. doi:0.1016/j.compenvurbsys.2011.07.005.

[42] G. Xavier and S. Dodge. An exploratory visualization tool for mapping the relationships between animal movement and the environment. In *Proceedings of the 2nd ACM SIGSPATIAL Workshop on MapInteraction*, pages 36–42, 2014. doi:10.1145/2677068.2677071.