Analysis of Flight Variability: a Systematic Approach

Natalia Andrienko, Gennady Andrienko, Jose Manuel Cordero Garcia, and David Scarlatti

Abstract—In movement data analysis, there exists a problem of comparing multiple trajectories of moving objects to common or distinct reference trajectories. We introduce a general conceptual framework for comparative analysis of trajectories and an analytical procedure, which consists of (1) finding corresponding points in pairs of trajectories, (2) computation of pairwise difference measures, and (3) interactive visual analysis of the distributions of the differences with respect to space, time, set of moving objects, trajectory structures, and spatio-temporal context. We propose a combination of visualisation, interaction, and data transformation techniques supporting the analysis and demonstrate the use of our approach for solving a challenging problem from the aviation domain.

1 INTRODUCTION

Comparison of trajectories is a common task in movement analysis. Thus, there may be a need to compare actual trajectories of moving entities to expected ones (such as typical, predicted, or planned), or trajectories of movement under different conditions. Comparison tasks may involve multiple and even very many trajectories, and it may be necessary to compare each trajectory to a certain reference trajectory. There may be a common reference trajectory for a set of trajectories (e.g., a central trajectory of a cluster of similar trajectories [7, 13]), or each trajectory may have its individual reference. For example, flights in aviation are generally conducted according to previously created plans. For each flight, there is a planned trajectory and an actual trajectory, which may deviate from the plan. The comparison task involves detecting, measuring, and analysing the deviations of actual flights from the respective plans.

Trajectories can differ in various aspects: route geometry, times of trip start and end, times of reaching corresponding intermediate positions, and characteristics of the movement along the route, such as speed and acceleration. For 3D movement, relevant characteristics include the altitudes or depths of the positions along the route.

While a need may arise in detailed examination of specific differences between two trajectories (e.g., in a case of an incident or an abnormality), in applications dealing with large amounts of movement data comparison tasks mainly aim at revealing and understanding overall patterns of trajectory variation with respect to space, time, variety of moving objects, and different aspects of the movement context, such as weather and occurring events.

The research contribution of our paper consists of three major components. First, we present a conceptual framework for comparative analysis of trajectories. Second, we propose a generic analytical procedure for performing comparative analyses involving a large number of trajectories. The procedure consists of
1. pairwise point matching between trajectories,
2. creation of difference data by computing various difference measures for the matched pairs of points,
3. when appropriate, division of trajectories into structural parts based on essential distinctions, e.g., in the character, purpose, or context of the movement, and
4. analysis of the distribution of the difference data with respect to space, time, set of moving objects, trajectory structure, movement characteristics, and movement context.

Third, to support the analysis, we propose a combination of techniques including
• creation of artificial spaces according to analysis foci and visualisation of the distributions of difference data in these spaces,
• dynamic, filter-sensitive spatial and spatio-temporal aggregation of difference data in geographic and artificial spaces,
• techniques supporting comparison of aggregates that summarise different subsets of data.

We demonstrate the effectiveness of the analytical procedure and technique combination by applying them to a challenging problem requiring comparative analysis of real-world massive movement data from the aviation domain.

The remainder of the paper has the following structure. Based on an overview of related work (section 2), we present the conceptual foundations for comparative analyses of trajectories (section 3), introduce the analysis procedure and supporting techniques (section 4), apply them for aviation data analysis (section 5), and discuss the overall approach (section 6).
2 Relevant work

There exist many visualization [21] and visual analytics [6] methods targeting at transportation data and problems. However, these data and problems are so complex, varied, and multifaceted that the existing methods are not sufficient to meet the needs of the transportation domain. In the following subsections, we discuss three major components of the existing approaches: visualization, interactive filtering, and data transformations. Next, we describe specifics of the aviation domain and position our research against the state of the art.

2.1 Visualization of trajectories

Trajectories are traditionally visualized as lines on a map and in a space-time cube [31]. Trajectories of movement in 3D space require representation of the altitudes or depths, e.g., by varying the appearance of line segments or by using 3D space displays [20, 42]. Visual displays showing multiple trajectories often suffer from severe overplotting and visual clutter. To handle this problem, movement data can be represented in an aggregated way using discrete [3, 12] or continuous [35, 43] aggregation. Other approaches include edge bundling [22, 30] and schematic representation [28], which improve display readability but introduce distortions and undesired artefacts. Aggregated or schematic representations of trajectories hide a lot of details. More detailed exploration is possible for a small number of selected trajectories. Interactive selection that temporarily hides a part of data is often called filtering.

2.2 Interactive filtering of movement data

Movement data involve different aspects [5], including spatial positions and path geometries, positions in time, and various attributes characterizing the movement, such as the speed and direction, the moving objects, such as their weight or load, and the movement context, such as weather parameters or density of surrounding traffic. Many such attributes can be derived from the spatial positions of the trajectory points alone or in combination with data describing the spatial, temporal, or spatio-temporal context of these positions [10]. Subsets of movement data can be selected based on any of these aspects using interactive tools designed for spatial, temporal, and attribute-based filtering [5]. A prominent example of interactive filtering by direct manipulation is FromDaDy [29], where the user selects trajectories by drawing shapes in displays presenting various 2D projections of flight trajectories. Filtering operations can be applied sequentially to results of previous operations. Temporal filtering can be based on linear [1] or cyclic [24, 25] models of time or on selection of time intervals satisfying interactively specified query conditions [14].

2.3 Transformation of time and space

An important tool for movement data analysis is data transformations [5], which include, among others, transformations of the spatial and temporal references. Transformations of time references include replacement of absolute times by relative positions within temporal cycles (annual, weekly, or daily) or with respect to the start and/or end times of trajectories [4]. Temporal references can also be replaced by chronological ordering, as in a trajectory wall display [40] and in a matrix of tram rides along a given route with rows corresponding to trips and columns to consecutive stops [37, 44]. Transformations of spatial references include replacing absolute spatial positions by relative positions within a group of jointly moving objects [13] and by positions in an abstract semantic space consisting of location categories rather than specific locations [15, 34].

Furthermore, artificial spaces can not only be obtained by transforming the spatial positions of trajectory points to another coordinate system but also constructed based on any attributes associated with trajectory points. This idea has not yet been applied to spatial trajectories, but it was used for representing changes of dynamic attributes of objects that do not necessarily move or are not spatial by nature. For example, the development of Swiss cantons was represented by trajectories in a space formed by two attributes [36]. A similar representation was used for stock market data [39]. An artificial space for representing changes and long-term evolution can also be constructed from multiple attributes using dimensionality reduction techniques [16, 41].

Artificial spaces can be treated similarly to usual (physical) spaces, which means that data transformed to artificial spaces can be visualised on maps using cartographic visualisation techniques [33]. One can also apply methods for spatial aggregation, e.g., build density maps [34].

2.4 Visual analytics for air traffic domain

Visual analytics approaches have been proposed for various specific problems in air traffic analysis, such as detection of holding patterns and problematic movement events [38], assessment of conflict probabilities regarding the traffic density [2], extraction and analysis of wind parameters [26], and detection of violations of rules imposed for decreasing noise in populated areas [19]. Sophisticated domain-specific analyses can be done by applying clustering to interactively selected relevant parts of trajectories [7]. Still, there are many analysis problems that have not yet been addressed in the visual analytics research. Due to the complexity and various specifics of the aviation domain, it is important to do research in collaboration with domain experts [7, 27].

2.5 Positioning of our work

While much research has been done on analysis of movement data, the problem of comparative analysis of trajectories has not been addressed yet. We systematically considered this problem and developed a general conceptual framework, where we defined the possible aspects in which trajectories can differ and figured out the objectives of analysing these differences for a large set of trajectories rather than for individual trajectories. Such comparative analyses aim at discovering and understanding patterns that exist among the differences with regard to their spatial, temporal, and spatio-temporal distributions, the distribution within the trajectories, and relationships to movement characteristics and spatio-temporal context.

Based on this conceptualisation of the problem, we developed a generic analytical procedure for comparative analysis of trajectories. Both the conceptualisation and the procedure are novel research contributions. They are complemented with a proposed suite of tools that can support the realisation of the approach. This includes a novel algorithm for matching trajectories with the aim to measure their differences. The other tools (data transformations, visualisations, and interaction techniques) were created by adapting and developing ideas from the previous research, namely, artificial spaces and dynamic aggregation.

3 Problem conceptualisation

3.1 Levels of analysis tasks

According to Bertin [17], tasks (questions) in data analysis can be differentiated according to the reading levels: elementary (referring to individual data items), intermediate (referring to groups of data items), and overall (referring to a whole data set). A unifying term synoptic was introduced [11] to refer to the intermediate and overall levels, both of which involve abstraction. Synoptic tasks deal with multiple data items, which are considered together, and require abstraction for deriving some general statements concerning all these items.

The distinction between elementary and synoptic levels is relevant, in particular, to comparison tasks [11]. In application to trajectories, elementary comparison tasks consist of detection and detailed examination of similarities and differences between individual trajectories. A smallest comparison task is pairwise comparison, i.e., comparison of two trajectories. An elementary comparison task involving more than two trajectories can be decomposed into several subtasks of pairwise comparison. From two trajectories, one can be treated as a reference to which the other trajectory is compared. For example, a planned trajectory can be a reference to an actual trajectory, or an actual trajectory can be a reference to a trajectory generated by a predictive model.

Synoptic comparison tasks involve joint consideration of similarities and differences of multiple trajectories from their reference trajectories; the latter may be either common or distinct. In analysing large collections of trajectories, synoptic tasks are of primary importance while elementary tasks need to be performed occasionally, in particular, for examination of trajectories with extreme differences from the reference trajectories.
3.2 Difference measures

Synoptic comparison tasks can be performed using computationally derived difference measures, which include distance in space, difference in time, and differences in movement characteristics, such as speed, direction, and, in case of 3D trajectories, vertical position (altitude or depth). Differences between the followed routes can be represented by the spatial distances between corresponding points of two trajectories and by the differences in the path lengths of corresponding parts of the trajectories. Temporal differences are measured as differences between the absolute and/or relative times of corresponding points. The relative times may be defined with respect to the trip starts and/or ends.

Differences in relative times are indicative of differences in movement speed, which can also be computed explicitly.

Derived difference measures can be organised in data records attached to components (i.e., points and segments) of trajectories. For uniformity, it can be assumed that difference data are attached to points, with data referring to segments being attached to the starting points of the segments. Each record consists of values of one or more difference measures. Since each point of a trajectory has its specific position in space (coordinates) and time (time stamp), difference data records become associated with these spatial and temporal positions, i.e., difference data are spatio-temporal by their nature.

Apart from local difference measures referring to points and segments, one can derive general difference measures referring to the whole trajectories. These measures include differences in the trip origins and destinations, start and end times, trip durations, path lengths, path curvatures, average speeds, and other general features of the trips.

3.3 Context data

Since movement is much affected by the context [5, 9], it may be necessary to analyse the relationships between the computed differences and the context. Movement context consists of various spatial, temporal, and spatio-temporal objects, events, phenomena, and processes. Global context, which is common for all moving entities at a given time, can be distinguished from local context, which differs among locations or combinations of locations and times and, hence, is specific to each point in a trajectory. Aspects of local context that are relevant to the analysis goals can be represented by attributes attached to points of trajectories [10]. Thus, for a trajectory point, attributes can represent:

- presence or absence, number, or density of spatial objects or events in the spatial, temporal, or spatio-temporal neighbourhood of the point;
- attribute values or statistical summaries of attribute values associated with the spatial positions and/or times in the neighbourhood of the point.

Such local context attributes can be derived by joining trajectory data with data describing the relevant context based on the spatial and/or temporal references specified in the context data. Aspects of global context can be represented by time-dependent attributes characterising properties of time moments (e.g., day or night, week day or weekend), presence of events (such as a strike or a public holiday), and/or the overall traffic situation on the territory under study (e.g., low or intense traffic). Unlike local context attributes, global context attributes are not associated with specific trajectories or positions in trajectories.

3.4 Structural parts of trajectories

Trajectories may consist of heterogeneous parts differing in essential features. Thus, travelling people may use different transportation modes, wheeled vehicles use different road categories, movements of ball game players with the ball differ from running without the ball, and flight trajectories include take-off, ascent, cruise, descent, and landing. For valid analysis, it may be appropriate to distinguish structural parts of trajectories, or movement phases, based on essential differences in movement character, purpose, context, or other aspects pertinent to the analysis goals. Analysts may strive to study how differences relate to trajectory structures, or may separately consider differences for distinct structural parts, or may be interested in analysing only certain parts of trajectories.

3.5 Foci in synoptic comparison tasks

Synoptic comparison tasks aim at discovering and studying general patterns of differences with respect to the space, time, set of moving entities, global and local contexts, as well as with respect to the internal make-up of the trajectories, including the geometry of the followed path, life time from the start to the end, movement characteristics along the path, and, possibly, essentially different structural phases of the movement. Hence, there are varieties of synoptic comparison tasks focussing on one or several of the following aspects:

- spatial distribution of differences;
- temporal distribution of differences;
- statistical distribution of differences over a set of moving entities;
- internal distribution of differences within the trajectories in relation to their spatial and temporal features and structure;
- relationships of differences to movement characteristics;
- relationships of differences to local and global contexts.

As stated in section 3.1, synoptic tasks include overall and intermediate tasks. Overall comparison tasks are applied to the entire set of trajectories, and intermediate tasks are applied to subsets. Subsets may be selected based on space (e.g., trips within or crossing specific areas or trips with specific origins and/or destinations), time (e.g., trips within a chosen time period, or trips that occurred at certain times in a time cycle), categories or characteristics of moving objects (e.g., types of vehicles or flight operators), and/or properties of the trips, such as path length, duration, curvature, average or maximal speed, etc.

Intermediate tasks may involve comparison between subsets.

3.6 Comparison tasks in flight data analysis

In aviation, flights are generally conducted according to plans, which are created in advance and agreed between flight operators and flight managers. In reality, actual flights may deviate from the plans for a variety of reasons, such as weather or traffic conditions. Deviations from plans occur frequently and are not considered problematic. Still, better compliance to flight plans is desired, and specialists are interested in analysing deviations for revealing general patterns and interdependencies that may suggest directions to improving flight planning.

As mentioned in section 3.4, flight trajectories, either planned or actual, consist of several structural phases. In analysing flights and their differences from the plans, it is important to distinguish these phases and, possibly, select only what is relevant to specific analysis goals.

4 Approach

The proposed analytical procedure is schematically shown in Fig. 2. Comparative analysis of trajectories is based on computing difference measures (section 3.2), which are attached to the trajectories (general measures) and their points (local measures). The measures characterise pairwise differences between each trajectory and some reference trajectory. While obtaining general difference measures is straightforward, deriving local difference measures requires finding for each trajectory point the corresponding point in the reference trajectory, which is not trivial. An algorithm for point matching is proposed in section 4.1.1.

Examining differences between two trajectories (elementary comparison) is not too difficult. It can be supported by visualizing the trajectories on a map and/or in a space-time cube [31], whereas the temporal variation of their dynamic attributes can be shown in a time graph [32]. Previously derived difference measures can be explicitly encoded using suitable visual variables [23].
In our research, we are mostly concerned with synoptic comparison tasks (section 3.1), which focus on the distributions of difference data over space, time, set of moving objects, in relation to trajectory structure, and in relation to local and global contexts (section 3.5). We propose a combination of supporting techniques, which includes creation of artificial spaces according to analysis foci (section 4.2), dynamic, filter-sensitive aggregation of difference data (section 4.3), and operations enabling comparison of aggregates. These techniques are generic, i.e., applicable to trajectories of various kinds of moving objects. We tested their effectiveness by applying to data from the aviation domain. Analysis of flight trajectories requires distinguishing between flight phases. To perform our analysis properly, we developed a method for dividing a flight trajectory into phases (section 4.1.2). In general, division of trajectories into structural parts is done by defining and applying domain-specific rules based on values of relevant attributes (e.g. for travel mode detection [18]).

4.1 Algorithmic foundations

4.1.1 Matching points of two trajectories

The goal is to match points along the paths from the origin to the destination based on their spatial proximity, as shown in Fig. 3. The basic idea of the algorithm is schematically represented in Fig. 4. The distances \(d_1 - d_5\) are measured between two candidate points for matching \((p_2, q_2)\) and from these points to the preceding and following points of the other trajectory. If the minimum of the five distances \(d_{\text{min}} = \min[d_1, d_5]\) is the distance to a preceding or following point (i.e., not \(d_1\)), it is compared to the distance \(d_{\text{prev}}\) or \(d_{\text{next}}\), respectively. If \(d_{\text{min}}\) is smaller, the corresponding points are matched; in all other cases, \(p_2\) and \(q_2\) are matched. In the case when \(d_{\text{min}} = d_1\) or \(d_{\text{min}} = d_5\) and \(d_{\text{min}} < d_{\text{prev}}\) (Fig. 4, bottom left), the new match \((p_1, q_2)\) or \((p_2, q_1)\) cancels the previously made match \((p_1, q_1)\). When this happens, an attempt to match the point that lost its match to one of the previous points in the other trajectory is done.

The complete algorithm is represented by pseudocode as Algorithm 1, and examples of its results are shown in Fig. 3. Two trajectories are coloured in red and blue. The points that got matches are represented by dots of the same colours as the trajectories, and the points without matches are represented by pink and cyan dots. Yellow lines connect the matched points. These lines are only visible where the routes diverge. As can be seen, the algorithm handles quite well unequal numbers of available points in two trajectories.

4.1.2 Identification of flight phases

In analysing flight trajectories, it is important to distinguish three phases of each flight: ascent, cruise, and descent. The task is not trivial, because both ascent and descent can be (and most often are) step-wise, and the altitude (flight level) can also change during the cruise phase. It is not easy to determine whether a flat segment of a trajectory (i.e., where the altitude is constant) is already the cruise or still a step in the ascent phase, and whether a decrease of the altitude means the beginning of the descent or just a change of the cruise level. Flight phases can be quite easily distinguished visually from a representation of the vertical profile of a flight over time. However, it is much harder to define formal rules that could be automatically applied.

![Fig. 3. Examples of point matching.](image1)

![Fig. 4. Schematic representation of the basic idea of the point matching algorithm: \(p_1\) and \(q_1\) are previously matched points, \((p_2, q_2)\) is the current candidate pair, the double lines represent matches.](image2)

---

**Algorithm 1 Matching points of two trajectories**

1. Procedure `MATCHPOINTSOFPANDQ (P, Q, toleranceRatio)`
2. Input: \(P, Q\): trajectories
3. \(M ← < (1, 1) >\)
4. \((i, j) ← (2, 2)\)
5. While \(i ≤ P.length \land j ≤ Q.length\) do
6. \((i, j) ← NEXTCANDIDATEPAIR(M, P, i, j, toleranceRatio)\)
7. \((m, j) ← M.lastElement() \land (m, j):\) last matched points of \(P, Q\)
8. If \(\text{dist}(P, Q) < \text{min}(\text{dist}(P, Q), \text{dist}(P, Q_{m+1}), \text{dist}(P_{m+1}, Q), \text{dist}(P_{m+1}, Q_{m+1}))\)
9. \(M ← M - (m, j) + (i, j)\)
10. \(i ← i + 1\)
11. Else
12. If \(\text{dist}(P, Q) < \text{min}(\text{dist}(P, Q), \text{dist}(P_{m+1}, Q), \text{dist}(P_{m+1}, Q_{m+1}), \text{dist}(P_{m+1}, Q_{m+1})))\)
13. \(M ← M - (m, j) + (i, j)\)
14. \(j ← j + 1\)
15. Else
16. If \(\text{dist}(P, Q) < \text{min}(\text{dist}(P, Q), \text{dist}(P_{m+1}, Q), \text{dist}(P_{m+1}, Q_{m+1}))\)
17. \(j ← j + 1\)
18. Else
19. If \(\text{dist}(P, Q) < \text{min}(\text{dist}(P, Q), \text{dist}(P_{m+1}, Q), \text{dist}(P_{m+1}, Q_{m+1})))\)
20. \(i ← i + 1\)
21. Else
22. \(M ← M + (i, j)\)
23. \(i ← i + 1; j ← j + 1\)
24. End
25. End
26. End
27. Procedure `NEXTCANDIDATEPAIR (M, P, i, j, toleranceRatio)`
28. Input: \(M\): set of matching pairs; \(P, Q\): trajectories; \(i, j\): running indices over \(P\) and \(Q\)
29. \((m, j) ← M.lastElement()\)
30. \(dp ← \text{dist}(P, Q)\)
31. \(dq ← \text{dist}(Q, Q_{m+1})\)
32. If \(dp > dq\) then
33. \(j ← j + 1\)
34. \(dq ← \text{dist}(Q, Q_{m+1})\)
35. Else
36. \(j ← j + 1\)
37. \(dp ← \text{dist}(P, Q)\)
38. Return \((i, j)\)
The main idea of our approach is first to select roughly a time interval when the flight is very likely to be in its cruise phase. From this interval, we find the minimal altitude among the flat segments or peak points, which we treat as the minimal cruise altitude. We include peak points into consideration assuming that a very short flight may consist of ascent immediately followed by descent; in such a case, the cruise phase is represented by a single peak point. We then scan the entire trajectory and find the first and the last flat segments where the altitude is not lower than the minimal cruise altitude. These segments are treated as the beginning and end of the cruise phase. The trajectory parts before and after them are categorised as ascent and descent, respectively. The algorithm is presented by pseudocode as Algorithm 2, and an example of its results are shown in Fig. 5, where the altitude is not lower than the minimal cruise altitude. These flight trajectories between two cities are coloured in red (ascent), blue (cruise), and yellow (descent).

4.2 Artificial spaces

Figure 5, bottom, shows an example of a constructed artificial space. The general idea is to choose two numeric attributes associated with trajectory points and treat their value ranges as spatial dimensions, which can be arranged as Cartesian or polar coordinate systems. The latter is suitable when one of the attributes has a cyclic value range. Examples are spatial direction (Fig. 1), time of the day (Fig. 6), and relative time within a week. Points of trajectories are represented by points in the constructed space according to their attribute values, and these new points are connected in the chronological order. The result is a trajectory in the artificial space. The artificial space with the trajectories can be represented in a map display analogously to geographical space. There is no background map, but it is possible to create a map layer with labelled grid lines or axes. A benefit of constructing maps of artificial spaces rather than implementing other types of display to represent dynamic attributes is that all map-based visualisation and interaction techniques can be uniformly applied to any map, either with geographic or artificial space. Thus, in Fig. 5 (bottom), a qualitative attribute of trajectory segments (flight phase) is represented on a map by colour coding, and Fig. 1 demonstrates the possibilities to build a density map and to set a "spatial" filtering window in an artificial space in the same way as it is done in a usual map. This uniformity is very convenient for analysis.

4.3 Dynamic aggregation and filtering

Aggregation is a common approach when it is necessary to deal with large amounts of data. Dynamic aggregation means that data aggregates are automatically updated in response to data filtering to include only the data satisfying the current filters. This, in turn, triggers automatic updating of all displays showing aggregated information.

Since our analytical framework involves intensive use of maps based on real or artificial spaces, we are primarily interested in using spatial aggregation methods in all such spaces. There are two types of spatial aggregation: continuous and discrete. Continuous aggregation represents data as a surface in which the value of some attribute gradually changes from position to position. This is realised through building

---

**Algorithm 2 Identification of flight phases**

1: **procedure** IDENTIFYFLIPTHASES \( TA, \text{tolerance} \)
2: input: \( TA = \{ (t_1, a_1), (t_2, a_2), \ldots, (t_N, a_N) \} \) \( \text{time-referenced flight levels (altitudes) of trajectory points} \)
3: input: \( \text{tolerance} \) \( \triangleright \) minor altitude changes to be ignored
4: \( \triangleright \) Classifying points into peak or flat Vs. ascent or descent
5: for \( i \leftarrow 1, N \) do \( \triangleright \) for all points of the trajectory
6: if \( (i \leq 1 \wedge a_i \geq a_{i-1} - \text{tolerance}) \lor (i = N \wedge a_i \geq a_{i-1} - \text{tolerance}) \lor (a_i \geq a_{i-1} \text{ tolerance}) \lor (a_i \geq a_{i+1} - \text{tolerance}) \) then
7: \( \triangleright \) PeakOrFlat\(_i \leftarrow 1 \)
8: \( \triangleright \) Peak or Flat
9: if \( N - t_i \geq 3 \text{hours} \) then
10: \( T_{\text{start}} \leftarrow t_1 + 45 \text{min}; T_{\text{end}} \leftarrow t_N - 75 \text{min} \)
11: else
12: \( T_{\text{start}} \leftarrow t_1 + 0.3 \ast (t_N - t_1); T_{\text{end}} \leftarrow t_N - 0.4 \ast (t_N - t_1) \)
13: \( \min \text{CruiseAltitude} \leftarrow \min(a_i) \) over [\( T_{\text{start}}, T_{\text{end}} \)]
14: for \( i \leftarrow 1, N \) do
15: if \( a_i \geq \min \text{CruiseAltitude} \) then
16: \( \text{atCruiseLevel}\_i \leftarrow 1 \)
17: else
18: \( \text{atCruiseLevel}\_i \leftarrow 0 \)
19: \( \text{atCruise}\_i \leftarrow \min(a_i) \) over [\( T_{\text{start}}, T_{\text{end}} \)]
20: **return** \( T_{\text{start}}, T_{\text{end}} \) \( \triangleright \) Division of trajectory into phases
21: \( \triangleright \) Ascent or Descent
22: \( \triangleright \) Empirically defined periods of the likely cruise phase
23: \( \triangleright \) Time of the day (local traffic intensity)
24: \( \triangleright \) Movement direction (Fig. 1)
25: \( \triangleright \) Spatial distance to the trip start or end
26: \( \triangleright \) Local context, as the traffic intensity around the points (Fig. 6)
27: \( \triangleright \) Basis of point matching results (section 4.1.1)
a raster, i.e., a grid with very small cells containing attribute values, which represent summaries, such as density, obtained from original data. Typically, creation of a raster involves spatial smoothing: each point from original data contributes not only to the value in the cell containing it but also to the values in the neighbouring cells within a chosen radius so that the contribution decreases with increasing the distance. The distance decay is regulated by a chosen kernel function, e.g., linear or quadratic. Continuous aggregation is demonstrated in Figs. 1 and 6, where sets of trajectories are represented by density surfaces built in artificial spaces. Examples of rasters in the geographic space can be seen in Figs. 7 and 8. Rasters are visualised by encoding the values by colours. The rendering typically involves value interpolation between neighbouring cells to make the surface appear smoother.

Discrete spatial aggregation is based on dividing the space into regular or irregular compartments, which are usually larger than in a raster. For the compartments, various data summaries are derived, typically without involvement of spatial smoothing. Sufficiently large compartment sizes allow visual representation of the attached summaries not only by colour coding but also by symbols, diagrams, or glyphs, which can be applied to several summary attributes, e.g., the mean and the standard deviation. Discrete aggregation of movement data can produce not only place-based aggregates summarising visits of places (i.e., spatial compartments) by moving objects but also link-based aggregates summarising movements between the places [5, 12]. The latter are visualised on flow maps [33] using flow symbols (e.g., as in Fig. 8) with the widths proportional to aggregate values, such as move counts.

Although it is possible technically to generate link-based aggregates in an artificial space, this may not be a valid idea. Links in real space are associated with meaningful spatial directions, while directions in an artificial space may be meaningless; hence, a flow map in an artificial space can hardly be interpretable and useful for analysis.

Both continuous and discrete aggregates can be dynamic, i.e., sensitive to data filtering. However, this responsiveness is not always desirable. Thus, it may be necessary to compare aggregates obtained from a subset of data to overall aggregates or to aggregates of another data subset. Hence, there should be a mechanism to suspend dynamic re-aggregation for some of the derived datasets (map layers) while the others preserve their dynamic behaviour. This enables three modes of comparison [23]: by juxtaposition, as in Fig. 8, superposition, as in Fig. 13, or explicit encoding of differences or ratios, as in Fig. 6. The latter example, in particular, demonstrates that actual flights tend to deviate from their planned routes to spaces with low traffic intensity. The raster of density differences shows high increase of the computed trajectory density for higher traffic intensities.

Different types of interactive filtering applicable to spatio-temporal data have been described elsewhere [5, 14]. Among them, the filter type called ‘time mask’ [14] is suitable for supporting analysis with regard to the global context (section 3.3). The filter selects data from time intervals satisfying specified query conditions. The latter can be based on attributes characterising the global context.

5 CASE STUDY

The approach was validated in a case study performed collaboratively by data analysts and aviation domain experts. The team analysed a dataset describing 32,736 flights over the territory of Spain in the period from April 7 till April 14, 2016 (8 days). For each flight, its planned and actual trajectories are available. The map in Fig. 7, left, represents the planned and actual trajectories by red and blue lines, respectively. It is easy to note that the routes of the actual flights often deviate from the planned routes. The analysts applied Algorithm 1 to match points of the actual and planned trajectories and computed various difference measures, including the differences in the lengths of the path segments between matched points that spatially coincide with 1 km tolerance.

In the following, all illustrations representing subsets of data in real and artificial spaces (section 4.2) were obtained by means of interactive filtering and dynamic aggregation (section 4.3). In particular, segments of trajectories were filtered based on the aforementioned difference measure ‘difference of path lengths between coinciding points’. Apart from the techniques presented in section 4, the study involved other visual analytics techniques, in particular, time series clustering [8].

5.1 Problem statement

The domain experts wish to investigate the deviations of actual trajectories from the planned routes such that the lengths of the corresponding path segments significantly differed. Fig. 7, right, shows the densities of the actual flight segments that were at least 2.5 km longer or shorter than planned. The highest densities occurred around the major airports, but the domain experts are not interested in these deviations. They knew that path changes in airport vicinities are typically caused by changes of the take-off and landing directions, which, in turn, correspond to changes of the wind. What is really interesting to the experts is the path length changes during the cruise phase of the flights. Therefore, the analysts applied Algorithm 2 to both the flight plans and actual flights and selected only the cruise phase by means of filtering. There were 29,343 flights whose cruise phase took place at least partly over the territory of Spain. Path reductions by 2.5 km or more occurred in 10,695 of them (36.4%), and path increases by at least 2.5 km occurred in 3,139 flights (10.7%). Hence, path reductions, i.e., straighter path segments than it was planned, occurred quite frequently. Occurrences of path reductions and extensions in a trajectory are not mutually exclusive; there were 725 flights (2.5%) involving both.

For the domain experts, more interesting are the cases when aircraft fly more directly than it was planned. They explain that, when negotiating their plans with the flight management services, airlines usually ask for straighter paths but in most cases do not get such a permission and have to plan longer paths. However, during the actual flights, pilots are often allowed and even commanded by flight controllers to fly more directly. On the one hand, this increases the flight efficiency; on the other hand, it also entails serious drawbacks. When longer paths are planned, more fuel has to be taken on board, which increases the aircraft weight. Flying with extra fuel is not just useless but expensive and environmentally negative. Besides, unplanned path shortening can decrease the flight duration making the aircraft come too early to its destination airport, where it may have to wait in the air for a permission to land. Therefore, the main improvement desired by airlines is the
flight plan compliance with straighter path segments put in the plans. To find out if there is potential for this improvement, the aviation experts wish to uncover patterns (regularities or dependencies) in the path reductions, which would show where and/or when and/or under what conditions shorter paths could be allowed at the planning stage.

5.2 Seeking spatial patterns

The density and flow maps in Fig. 8, left, summarize all segments of the flight plans belonging to the cruise phase. In the flow map, the widths of the flow symbols are proportional to the flight counts. The maps on the right represent only the parts of the plans that were shortened in the real flights by at least 2.5 km. The overall spatial patterns differ quite much. It is seen that path reductions occurred mostly on certain routes, moreover, in particular parts of these routes. Furthermore, it appears that there may be directional patterns of path reductions.

To investigate this, the analysts transform the planned trajectories to an artificial space with a polar coordinate system in which the angle represents the movement direction and the radius corresponds to the spatial distance from the beginning of the cruise phase or from the boundary of the studied territory if the cruise phase began outside of it. In Fig. 1, the trajectories are represented by lines drawn with 1% opacity (A) and by a density map (B). Both maps clearly show the prevalence of the north-eastern and south-western movement directions. After selecting only those segments that were substituted by shorter paths in the real flights, the density map changes its appearance as shown in Fig. 1C. There is a hot spot of very high density at the angle 225 and distance 0-15 km. The density range mapped onto the colour shades in Fig. 1C has been interactively reduced to make high densities beyond the hot spot more prominent. As a result, the density map has a hole in the position of the hot spot, but now it is easy to see that the densities of the path-reduced segments are generally higher in the directions north - north-east - east-north-east than in the opposite directions, except for the hot spot and two wings originating from it.

To see what trajectories correspond to the hot spot, the analysts create a filtering window enclosing it (shown in an inset in the lower right corner of Fig. 1C). The filter selects all trajectories that cross the window in the artificial space. A geographic map display shows the selected trajectories in the geographic space (Fig. 1D). They are drawn with 5% opacity; hence, brighter lines correspond to multiple overlapping trajectories. The map shows that many trajectories come from the south-west of France, but there are also bunches of trajectories whose cruise phase started elsewhere, e.g., south from Madrid.

The spatial pattern exploration has revealed the existence of particular traffic lines and directions where more path length reductions have occurred than elsewhere. The next task is to find out when and/or under what conditions this happened.

5.3 Seeking spatio-temporal patterns

To investigate the temporal aspect of the path reductions, the analysts use the discrete aggregation of the planned trajectories into flows in a generalized traffic network, as shown in Fig. 8. For the links of this network, they compute time series of flow magnitudes, i.e., flight counts per time unit. Specifically, they take time intervals of 1 hour length with 30 minutes shift, i.e., the intervals overlap by 30 minutes for temporal smoothing and diminishing boundary effects. The time series are computed for all flight segments and separately for those which were reduced in the actual flights. The histograms below the map (rows: days, columns: hourly intervals with step 30 minutes) represent the average per hour counts of all flights (left) and reduced segments (right) going through the links of the clusters 3-7.
occurred quite regularly; the time series of the reduced paths look similar to the complete time series. It appears that some proportion of the flights using these links can be safely allowed to plan more direct paths. Cluster 6 (light blue) has the highest regularity of reductions among the remaining clusters, which may also entail some potential for optimizing flight planning, although the flight counts here are much lower than in cluster 5. Cluster 4 (orange) could be of high interest from the perspective of planning optimization, as it involves many flights, but its temporal pattern of path reductions has low regularity. The same applies to the other clusters. There were some days and/or hours when more path reductions occurred than in the other days and/or hours. The experts suppose that some air traffic controllers may be more inclined than others to straighten paths in the flights under their control.

The experts want to verify their observations using a less aggregated representation than in Fig. 9. The analysts compute time series of path reductions for cells of a regular grid with square cells of the size 10x10 km and visualize the result in a space-time cube, as in Fig. 10, where the temporal dimension is oriented upwards. In this view, the values 3 or higher are represented by proportional sizes of circle symbols. Concentrations of circles signify spatio-temporal clusters of path reductions. The experts rotate the cube and observe where reductions were practised regularly and where they occurred more occasionally. The space-time cube offers them a more refined view of the spatio-temporal patterns of path reductions. They conclude that the patterns can mostly be considered individual-independent with a high level of confidence, making optimisation in strategic planning feasible.

The 2D time histograms also allow the experts to check if the path reduction patterns could be related to the traffic intensity, as it can happen that some flights are requested to deviate from their planned routes for decreasing the traffic intensity in the parts of the air space where it is expected to be high. The experts could not find such a relationship in their earlier studies with other datasets and other methods, but they wanted to check this result using visual analytics techniques. This can be done by comparing the histograms on the left and on the right of Fig. 9. The histograms on the left show the traffic intensities represented by the flight counts per time unit. The existence of a relationship would manifest through high path reduction counts in the times of high traffic and very low values in the remaining times, which is not the case. For clusters 5 and 6, the reduction counts look proportional to the flight counts; for the other clusters, the patterns of path reductions are less regular and thus dissimilar to those of flight counts.

5.4 Seeking relationships to temporal deviations

The domain experts want to see how the deviations from the planned routes are related to flight delays computed as the differences between the time till the actual flight end and the time till the planned flight end. Positive differences mean delays of the actual flights and negative differences mean the opposite. It could be hypothesized that air traffic controllers may tend to allow straighter paths to retarding flights to help them reduce the delays. For revealing the relationships and checking this hypothesis, the analysts build an artificial space where the X-axis represents time and the Y-axis represents the delays. After converting the actual trajectories to this coordinate system, the analysts create a density map (Fig. 11A). Generally, the densities are symmetric with respect to the line Y=0, which means that delays and advances are mostly balanced. The density map in Fig. 11B summarises the length-reduced parts of the actual trajectories. The density distribution pattern also looks quite symmetric with respect to Y=0, but there are no regions of very high densities around this line, as in the overall density distribution. This means that the flights involving path reductions are less frequently on time but more often deviate from the planned times. The symmetry of the density distribution signifies that retarding flights do not use shorter paths more often than others.

To verify this observation, the analysts produce a raster with average differences of the path lengths in the actual and planned trajectories. In Fig. 11C, the differences are represented using a diverging colour scale with shades of blue for negative differences (shorter segments in actual flights) and shades of brown for positive differences (longer segments in actual flights). The visible prevalence of negative values above the line Y=0 may mean that paths in retarding flights tend to be straightened to a higher degree than in other flights.

In a similar manner, the team investigates the existence of a relationship between the path straightening and the tendency to longer or shorter flight durations than planned. They consider the differences between the times till the respective flight ends in the actual and planned

Fig. 10. The space-time cube shows time series of path reductions computed for a fine regular division of the territory. The hourly counts 3 or higher are represented by proportional circle sizes.

Fig. 11. A-C: An artificial space with the X-axis representing time and Y-axis representing the delays of the actual flights compared to the plans. A: density of all segments of the actual trajectories; B: density of their shortened segments; C: the average differences of the lengths of the actual and planned segments. D: Analogous to C, but the Y-axis represents the differences in the time till the flight end in the actual and planned trajectories.

Fig. 12. Flow maps obtained by aggregation of the flight plan segments such that the corresponding actual segments take longer (left) and shorter (right) time by at least 5 minutes.
6.5 Comparing path reductions between airlines

The domain experts would also like to investigate whether different airlines are equally treated by air traffic controllers with regard to path straightening, or some of them have higher chances to get straighter paths than the others. The proportions of the number of path-reduced segments to the number of all segments differ significantly among the airlines; thus, for 8 major airlines, the proportions range from 10.6% to 33.5%. However, these differences are not indicative of unequal treatment, because different airlines have different routes, and it was observed earlier that path reductions occur on some links of the traffic network more often than on others. Conclusions concerning equal or unequal treatment can only be made based on comparing relative frequencies of path reductions for different airlines on the same links. Such comparisons can be done using flow maps, as in Fig. 13, where the maps represent aggregated data from the flight plans of four airlines. The dark purple flow symbols represent the aggregates obtained from the whole trajectories (cruise phase), and the yellow flow symbols drawn on top represent the aggregates from the shortened segments. The maximal absolute values of the flow magnitudes, which are represented by the maximal widths of the flow symbols, differ among the maps. This supports comparison of the relative frequencies.

Fig. 13. Comparison of path reductions for different airlines. Purple: aggregate flows from the complete trajectories (cruise phase) of each airline. Yellow: aggregates from the path-reduced parts of the trajectories.

Generally, the flow maps in Fig. 13 show that on the same links different airlines have approximately equal frequencies of straighter paths relative to the total frequencies of flying through these links. The large differences in the proportions of path reductions are due to differences in the routes served by the airlines. Hence, it can be concluded that different airlines receive equal treatment from the air traffic controllers.

5.6 Overall conclusion from the study

The study allowed the domain experts to gain valuable insights into the functioning of the air traffic management (ATM) system. Their questions regarding flight plan compliance were answered. They saw possibilities for optimising planning on certain important routes and links of the traffic network, found that route changes do not depend on personal preferences of traffic controllers, and observed that different airlines receive equal treatment. They confirmed their previous finding concerning the independence of the path reductions of the traffic intensity and learned that path reductions are also not used massively as an instrument to decrease flight delays. These findings basically show that path shortening does not happen due to special reasons, but its purpose may be convenience of traffic controlling and better distribution of the staff workload. Hence, shorter paths can be put in flight plans more often than now.

6 DISCUSSION AND CONCLUSION

What concerns the case study, the domain experts were very much satisfied by their collaboration with the analysts, who showed them the power of visual analytics techniques. The experts believe that the capability of visual analytics to reveal spatio-temporal patterns has a great potential for reducing inefficiencies in the ATM system by learning the actual behaviour of the system and being able to understand it better. Results as shown in this paper are thus of great interest not just scientifically but also from the operational perspective. The experts emphasise the importance of the problem addressed in the case study and the need for better flight planning and plan compliance, especially for reducing negative impacts on the environment. The study yielded relevant findings, which, of course, require further testing on other datasets. Nevertheless, the experts acknowledged that it showed

- the capability of visual analytics to detect patterns that can be applied for global benefit of the system,
- the ability to provide a better insight into the ATM system and thus support better decision making in strategic/pre-tactical phase,
- the usability of the results achieved, providing direct benefit in the operational environment.

Although our research on comparative analysis of trajectories was motivated by practical needs existing in the aviation domain, we strove to obtain results that could have a broader area of applicability. Based on our previous experiences in analysing diverse kinds of movement data from different domains, we considered comparative analysis as a general problem and developed a generic conceptual framework, which centres around the concept of pairwise difference measures. Accordingly, we developed a method to derive these measures through pairwise point matching between trajectories. The framework defines synoptic comparison tasks as seeking patterns in distributions of the differences with respect to space, time, trajectory structure, movement characteristics, and spatio-temporal context. Accordingly, we composed a suite of visual analytics techniques that can support exploration of these distributions and discovery of existing patterns.

Although this paper presents our approach in application to air traffic, we also tested it using different data examples. Except for dividing flight trajectories into phases, all other techniques and the whole analytical procedure presented in section 4 are domain-independent. They can be applied to comparative analysis problems in any domain where differences between trajectories following similar routes need to be quantified and analysed. The description of our case study provides an example that facilitates understanding of the general procedure and shows how to use the proposed techniques. Following this example, similar analyses can be done in different application areas.


