Introduction to Trajectory Data Mining

Kirsi Virrantaus

Aalto University

GIS-E4020

Learning outcomes

- To know the basic terminology in trajectory data mining
- To get an idea how generic data mining methods are used for trajectory data
- To know some cases from research
- To get an idea about the topics for future reserach

Contents

1. Introductory part
   - Concepts and definitions in trajectory data mining
   - How trajectories are created ?
   - How trajectory data can be used?

2.How to mine trajectory data

-generic methods and special methods for trajctories

3.Some examples

-from Marc van Kreveld and research by Demsar&Virrantaus

4. Semantic trajectory data mining and challenges

1. Trajectory data, definition

- trajectory(en) = liikerata(su)

- trajectory of a moving object is a continuous function *τ(t)* of time *t* such that given a time instant *t* it returns the position of the moving object

- in reality, the moving objects trajectory is recorded by a finite set of observations at discrete times $t_1, t_2, \ldots t_n$

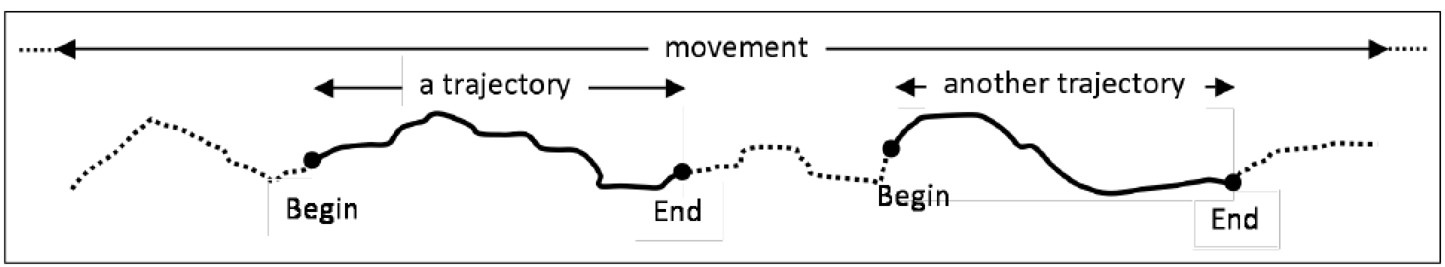


Figure 3 - Trajectories extracted from a movement track visualized as dotted line

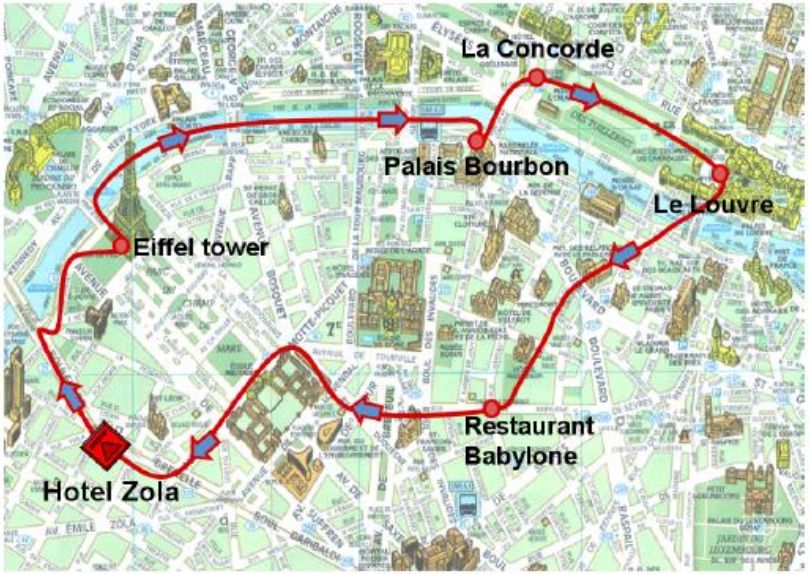


Figure 1 - 2D visualization of a one-day spatial trace left by a tourist visiting Paris – background map downloaded from Mappery.com, copyright unknown

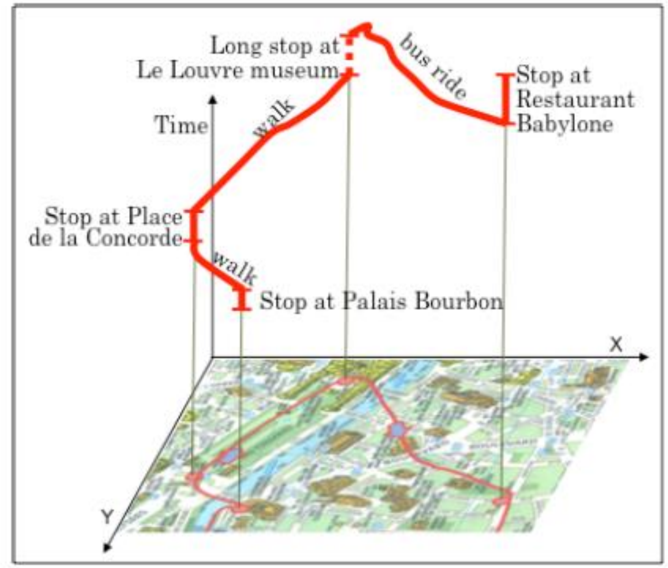


Figure 2 - A time-geography diagram showing part of the previous tourist track

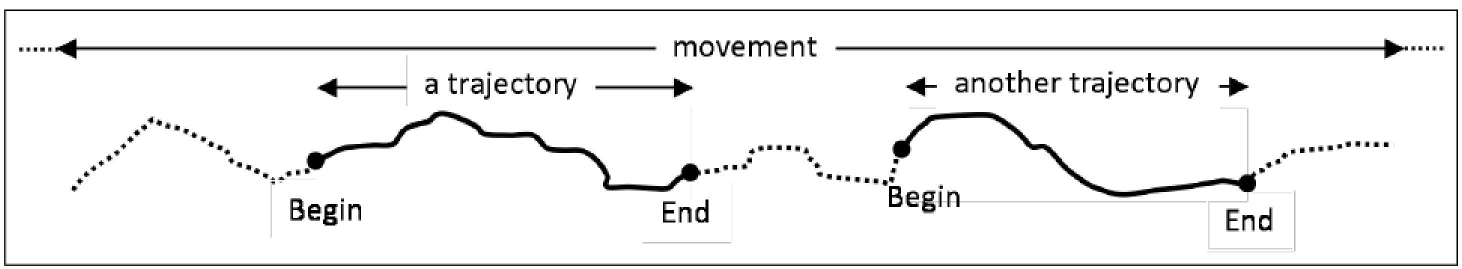


Figure 3 - Trajectories extracted from a movement track visualized as dotted line

Movement track in 2d and in 3d-presentation. **Trajectory** can be seen as the Interesting part of the entire movement track. (Parent et al.)

Some more concepts on trajectories

- **Raw trajectories**, collected data without context
- **Semantic trajectories**, added context
- **Holes**, missing data
- **Semantic gaps**, meaningful stop

- Cleaning the trajectory (removing noise)
- Interpolating of a trajectory (for example by splines)
- Synchronizing trajectories
- Map matching (accurate trajectories)

- Behaviour of trajectories (individual and collective)

- These concepts and semantic trajectory mining will be discussed more in the end of the lecture (Parent et al.)

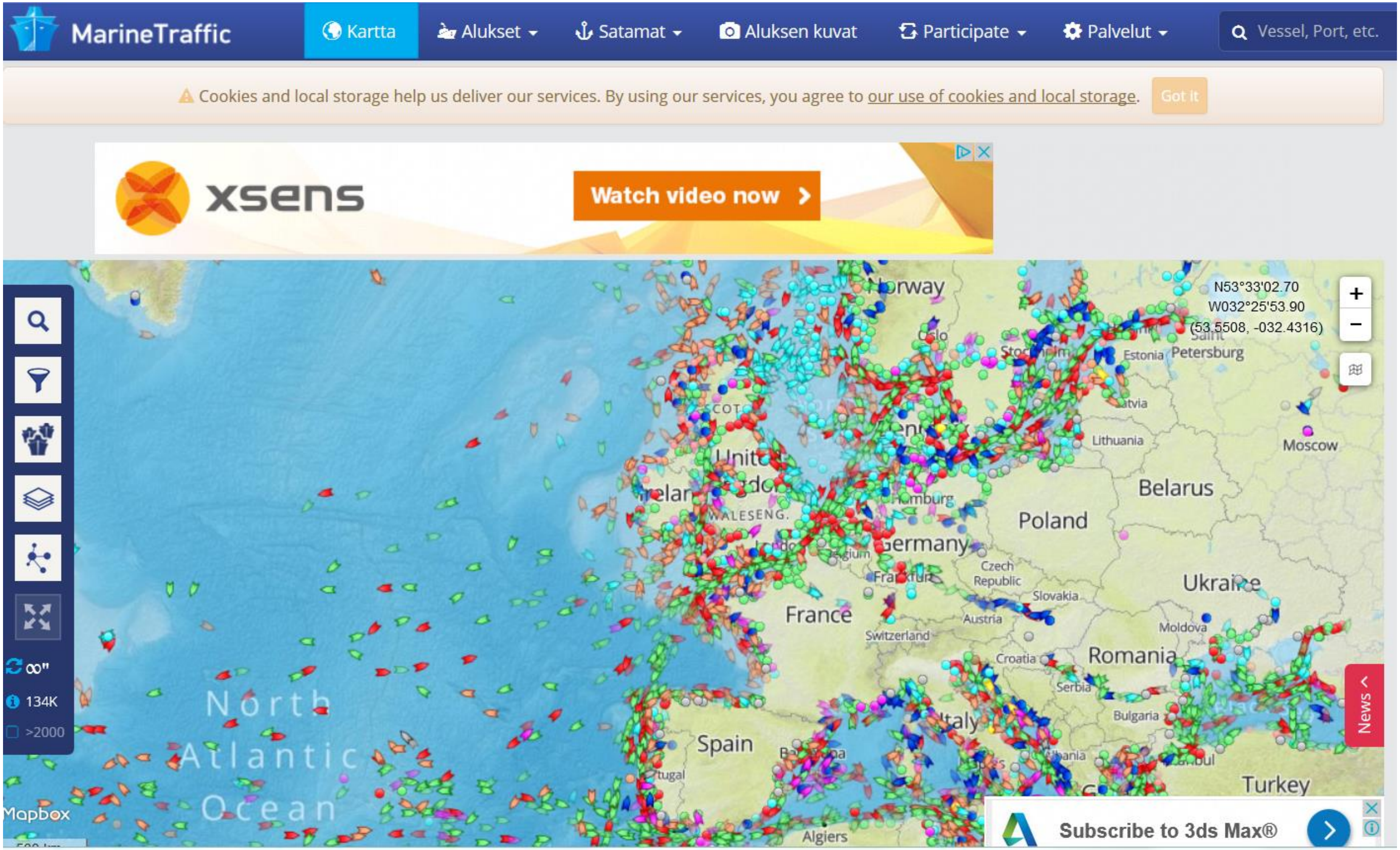
How trajectories are created ?

- Depending on the recording system, data is available in different original forms
- **GPS-data**: temporally ordered sequencies of geographical coordinates; the moving objects carries the device (GPS)
- **GMS-data**: temporally ordered sequences of identifiers of the cells in which the moving object passes (mobile phone)
- **Geo-social network based data**: content found in Internet social media (content can include coordinates but also place names etc. that can be **geocoded**)
- **RFID** (radio frequency identification): a sequence of identifiers of RFID readers through which the moving objects passed
- **Wi-Fi based data**: a sequence of identifiers of access points that communicated with the moving object (wireless network)

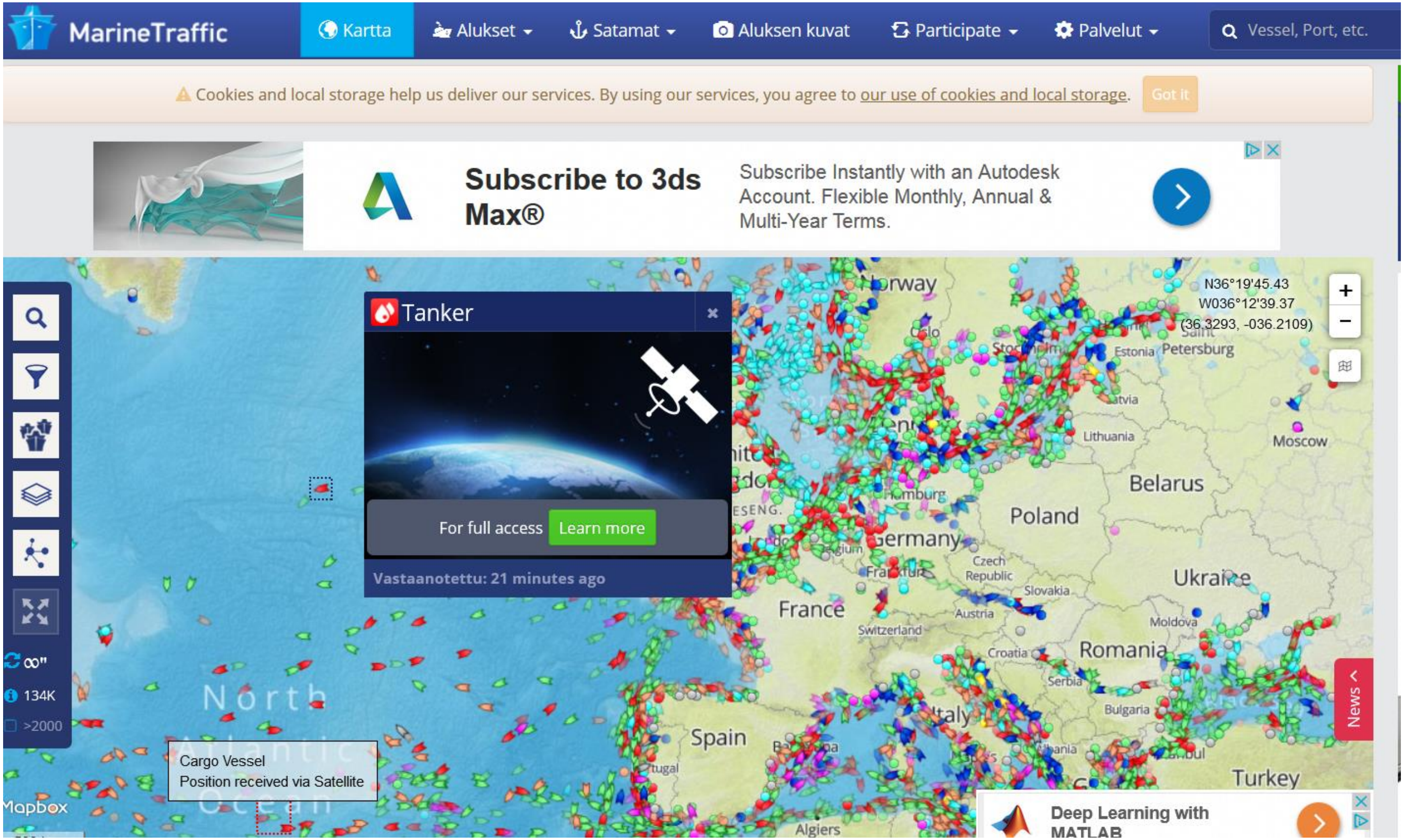
Examples of TD and applications

- Examples of trajectory data:
  - tracked animals (reindeers, moose,birds, ...), people (elderly people) , GSMs (e.g. for traffic purposes)
  - tracked vessels; real time tracking in “MarineTraffic.com”
  - http://www.marinetraffic.com/ais/fi/default.aspx
  - trajectories of tornadoes http://www.tornadohq.com/
  - sports scene analysis (players on a field)
- Trajectory data mining has many important, real-world applications driven by the real need
  - Homeland security (*e.g.*, border monitoring), situation pictures
  - Law enforcement (*e.g.*, video surveillance)
  - Weather forecast
  - Traffic control, marine traffic (VTS., Vessel Traffic Services)
  - Location-based services


Photo: C. Tonn



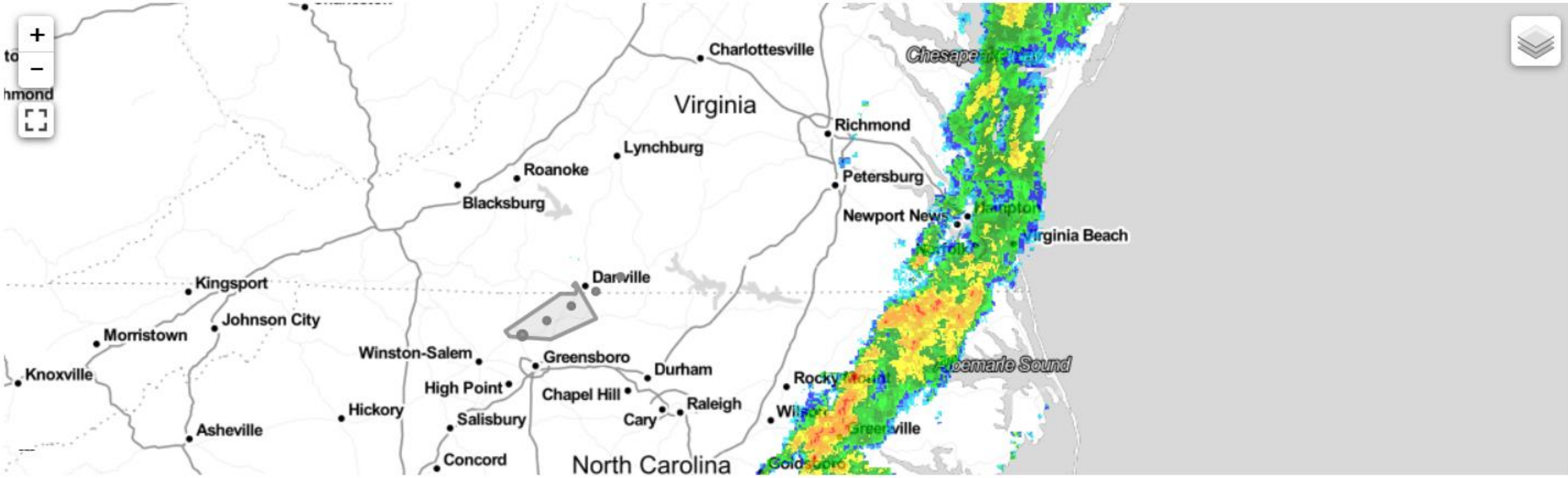
Vessel tracking by AIS (automatic identification system)



By clicking the symbol you can get the position information and many other details lof the vessel in question.

tornado tracker map

toggle radar



Tornado Warnings Today

If you want to see ALL severe weather warnings, go to the **severe weather map** page.

Warning Start	Warning End	Phenomena	Warning Summary	Warning Counties	Warning States	Map
about 6 hours ago	**EXPIRED**	tornado	At 847 PM EDT, a severe thunderstorm capable of producing a tornado was located near Fort Littleton, moving northeast at 55 mph.	Franklin / Fulton / Huntingdon	PA	show me

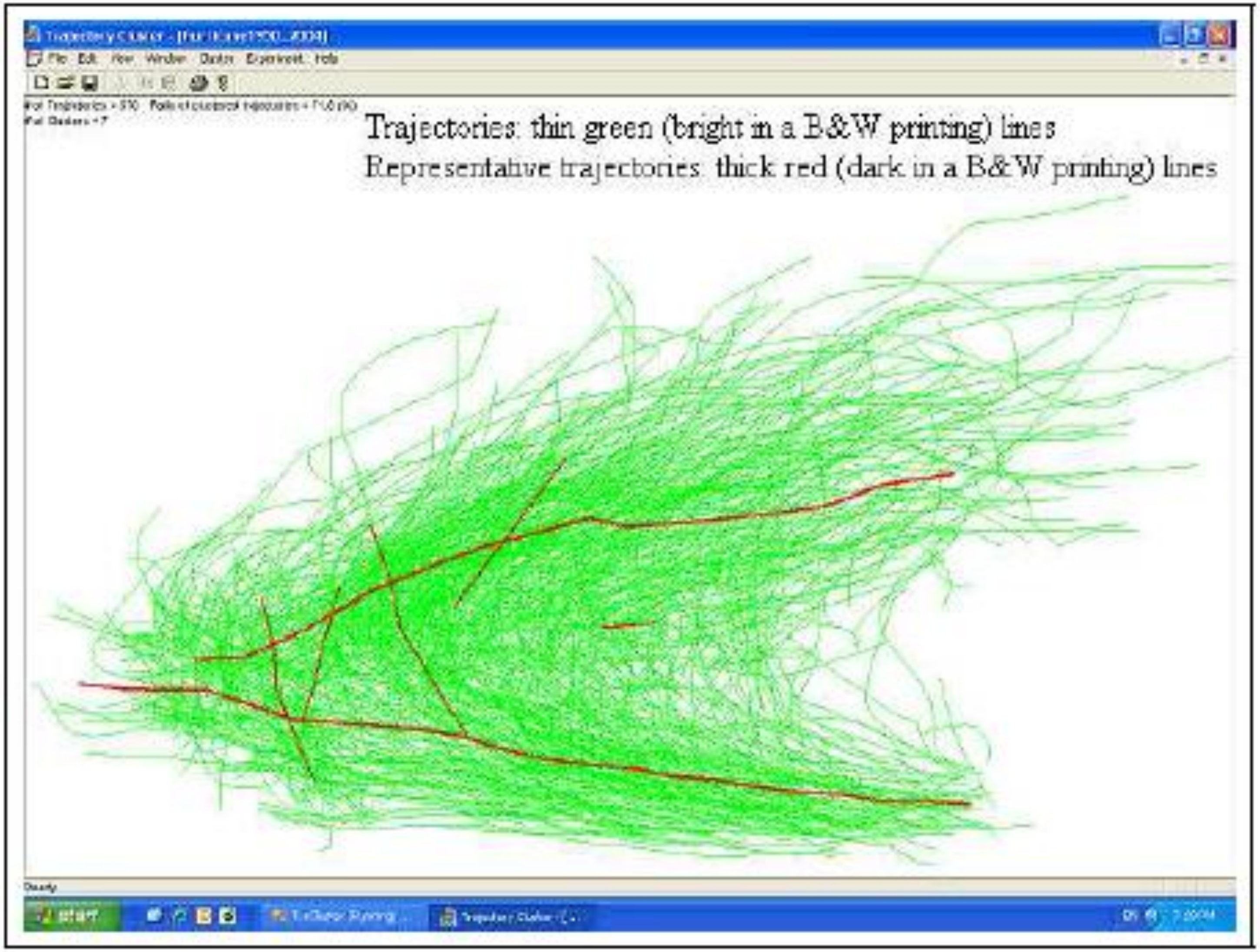


Figure 18: Clustering result for the hurricane data.

(Lee,Han&Whang)

2. How to mine trajectory data ?

- Trajectory data can be mined by using similar methods than other spatial data
  - Clustering
  - Classification

- There are also special algorithms/methods for trajectories
  - Frequent pattern mining
  - Group pattern mining
  - …

- Trajectory data mining process is **application driven**; depending on what we want to reveal from the sensed data
- The following intro is taken from Mazimpaka&Timpf-article (2016)

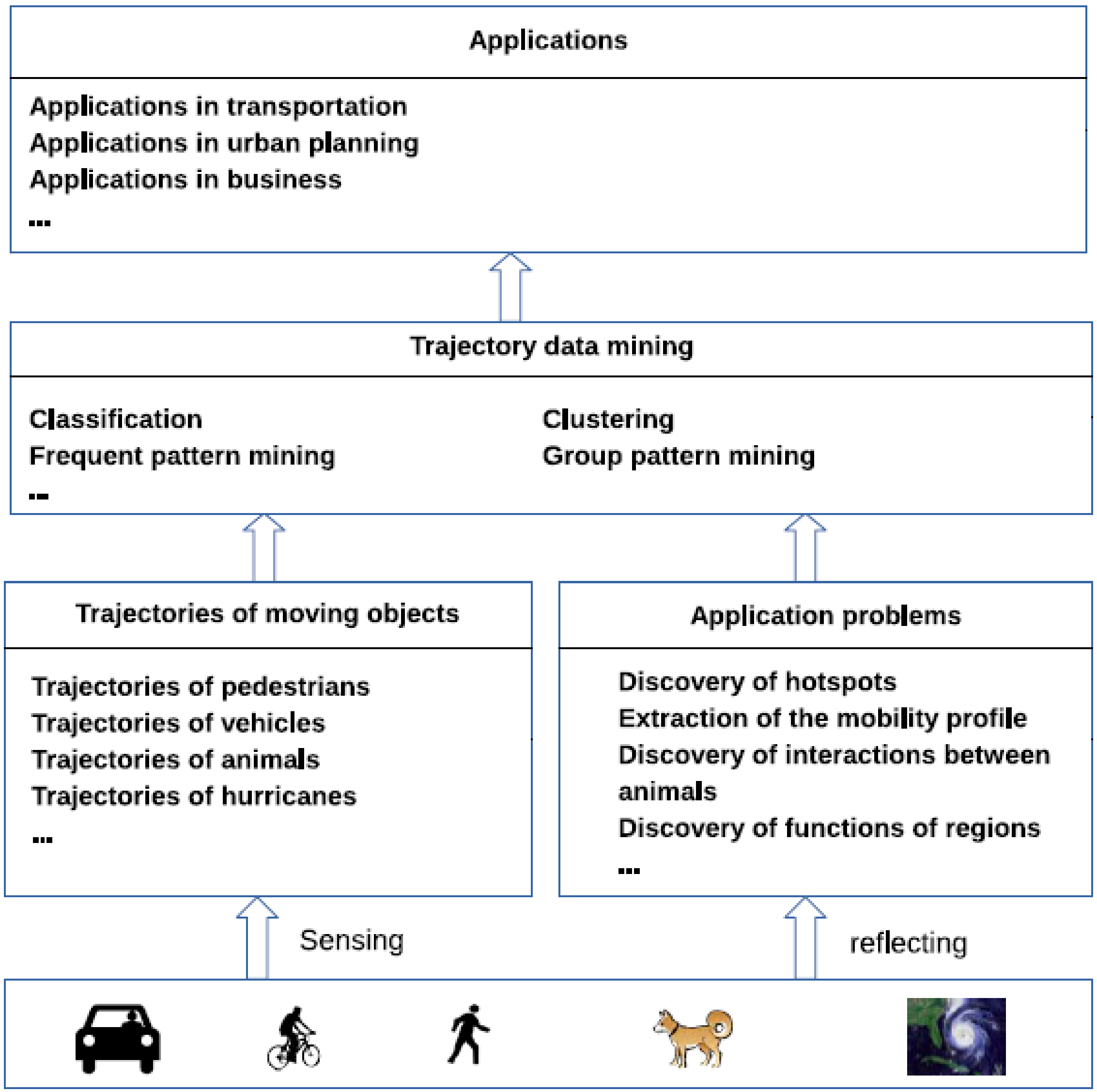


Figure 1: General framework of application-driven trajectory data mining.

(Mazimpaka & Timpf, 2016)

Primary mining methods

- **Clustering**
  - **based on similarity of trajectories** (raw trajectories)
  - by the **characteristics** of the trajectories (semantic trajectories)

- In most cases the input data set is **a set of trajectories** and each trajectory is analysed in turn

- Input can also be only **one trajectory** when some positions on it are analysed

- Trajectory analysis can be on either whole trajectory or parts of them

Similarity measure (for raw trajectories)

- For the **similarity measure** the distance between two trajectories is calculated
  - See Fig. 1

- With time shift
  - See Fig 3.

- With different duration
  - See Fig 4.

(Buchin et al.)

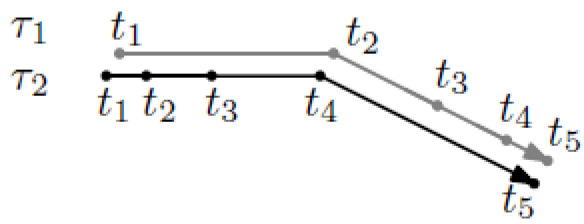


Figure 1: Trajectories with similar shape but different speeds at corresponding times.

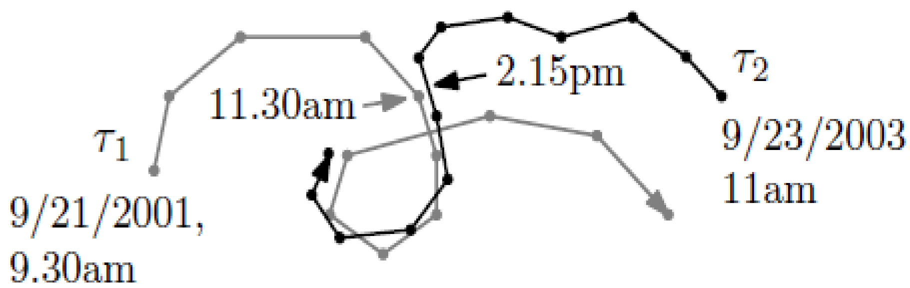


Figure 3: The middle part of trajectory τ_1 is similar to the last part of trajectory τ_2.

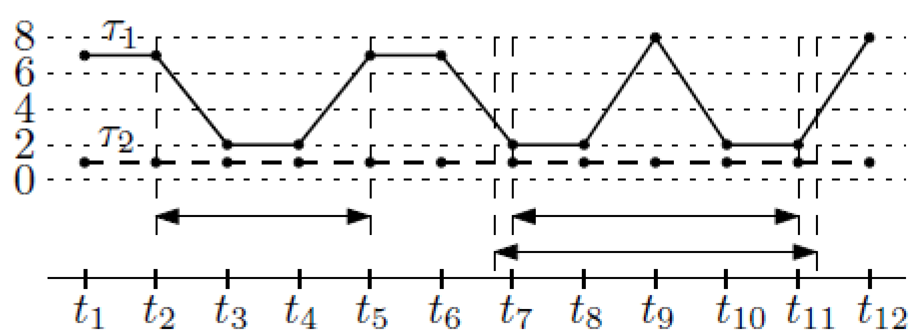


Figure 4: Subtrajectory similarity with non-fixed duration.

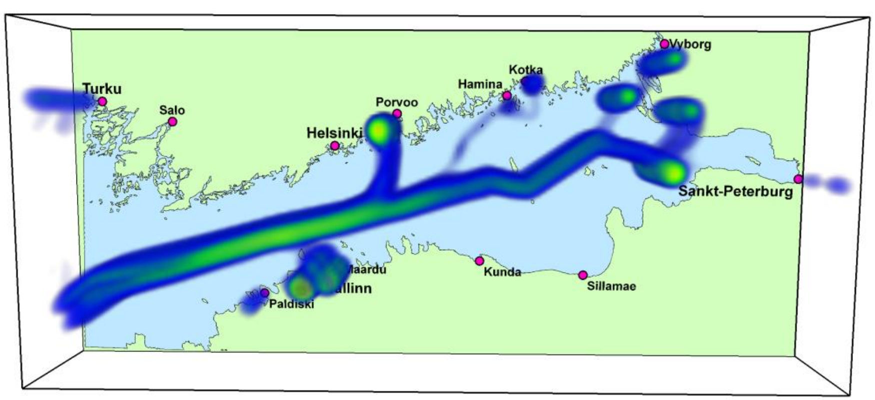
Trajectory similarity – variations of the problem

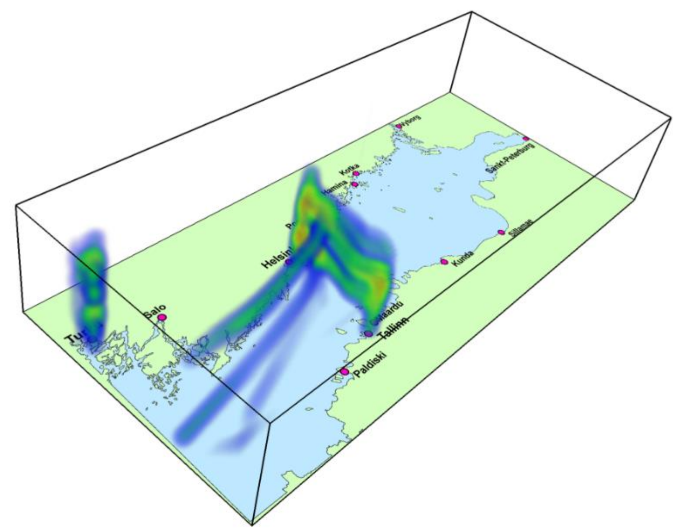
- Similarity analysis:
  - subtrajectories with smallest dissimilarity
- Variants
  - 1. Duration is fixed, starting time the same
  - 2. Duration is not fixed, but it must be not less than some minimum length, starting times are the same
  - 3. The duration is fixed, starting times may be different
  - 4. The duration is not fixed, starting times may be different

Examples: what you can find by clustering

- Most similar subtrajectories of exactly 3 hours in two animal trajectories recorded in a week.
- Most similar subtrajectories of two hurricanes for the duration of exactly four days, but the hurricanes occurred at different times.

- Typical trajectories of oil tankers and ferries at Baltic Sea (Demsar&Virrantaus)





Primary mining methods

- **Classification**
  - Objective: to find a rule to assign objects to pre-defined classes
  - Predifed classes exit and a set of objects in each class

- Trajectory classification algorithms first **train the classification model** by using training data set and then **apply the rule** to the data set to be classified

- Example: how to use classification

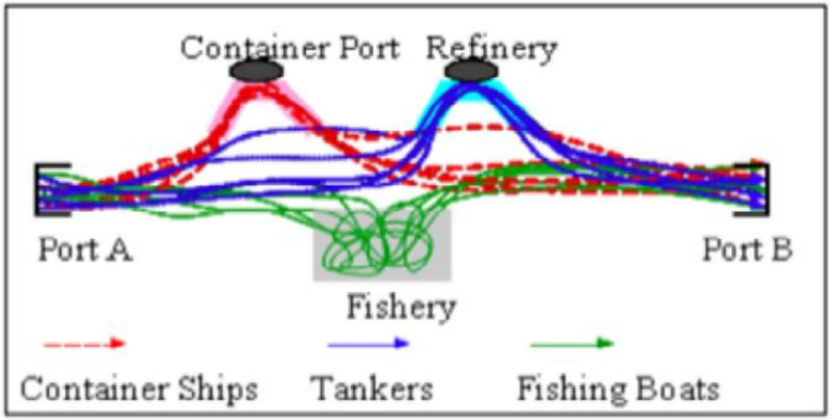


Figure 7 - Example of trajectory classification [Lee et al. 2008]

Primary methods		Secondary methods: Pattern mining	Secondary methods: Outlier detection	Secondary methods: Prediction
	Clustering	• Grouping periodically related locations [20, 85] • Grouping close trajectories [62, 129] • Extraction of places of interest for frequent pattern mining [49] • Aggregating close locations for sequential pattern mining [45]	• Grouping similar trajectories or sub-trajectories [77]	• Grouping similar users [137] • Grouping similar trips of a user [3, 41] • Grouping user's stay points for building trajectory patterns [136] • Grouping visited locations for building periodic patterns [61]
	Classification		• Categorization into normal and abnormal trajectories [82]	• Categorizing on-going trajectory to one of defined trajectory clusters [3], or trajectory patterns [104]

Table 1: Relationships between trajectory mining methods.

(Mazimpaka & Timpf, 2016)

Secondary mining methods

- The goal is to analyse the spatial, temporal por s-t arrangement of individual trajectories withing their categories or between categories

- Pattern mining
- Outlier detection
- Prediction

Pattern mining

- Aim: to **discover and describe the movement patterns** hidden in trajectories
- **Repetitive pattern** mining; periodic patterns by clustering
- **Frequent pattern mining**; analysis of popular routes
  - The later example by (Demsar&Virrantaus) belong to this group
  - First example by MK belong to this group
- Group pattern mining; analysis of groups of objects moving together
  - Second example by MK belong to this group

Secondary methods

- Outlier detection
  - Outliers often appear as a side product of for example clustering
    - The later example by (Demsar&Virrantaus) belong to this group

- Prediction
  - Aim: to guess the future location of the moving object based on existing trajectories
  - Methods based on Markov models

3. Examples, cases

- First example: Marc van Kreveld
  - one of the earliest developers of trajectory data mining methods; University of Utrecht,NL
  - visited our group in 2006
- Second example
  - A piece of research made by Urska Demsar and K.Virrantaus starting in 2010…
  - Analysis of AIS data

Marc van Kreveld

Professor; computational geometry and its application

Division Virtual Worlds
Department of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands

Phone: +31-30-253 4119
E-mail: **m.j.vankreveld** [curly symbol] **uu** [point] **nl**

- My publications according to DBLP
- My publications in Google Scholar Citations
- Geometric construction puzzles designed and made by me
- Composable art designed and made by me

Research interests: computational geometry, GIScience, graph drawing, puzzle games analysis and generation.

Urska Demsar

Geography & Sustainable Development - Lecturer in GeoInformatics

Postal address:
School of Geog & Geosciences
Sustainable Development
Irvine Building
St Andrews
United Kingdom

E-mail: **urska.demsar@st-andrews.ac.uk**

Direct phone: +44 (0)1334 463890

Research overview

My research area is Spatio-Temporal Visual Analytics. See **my website** for more info.

Early examples from Marc van Kreveld

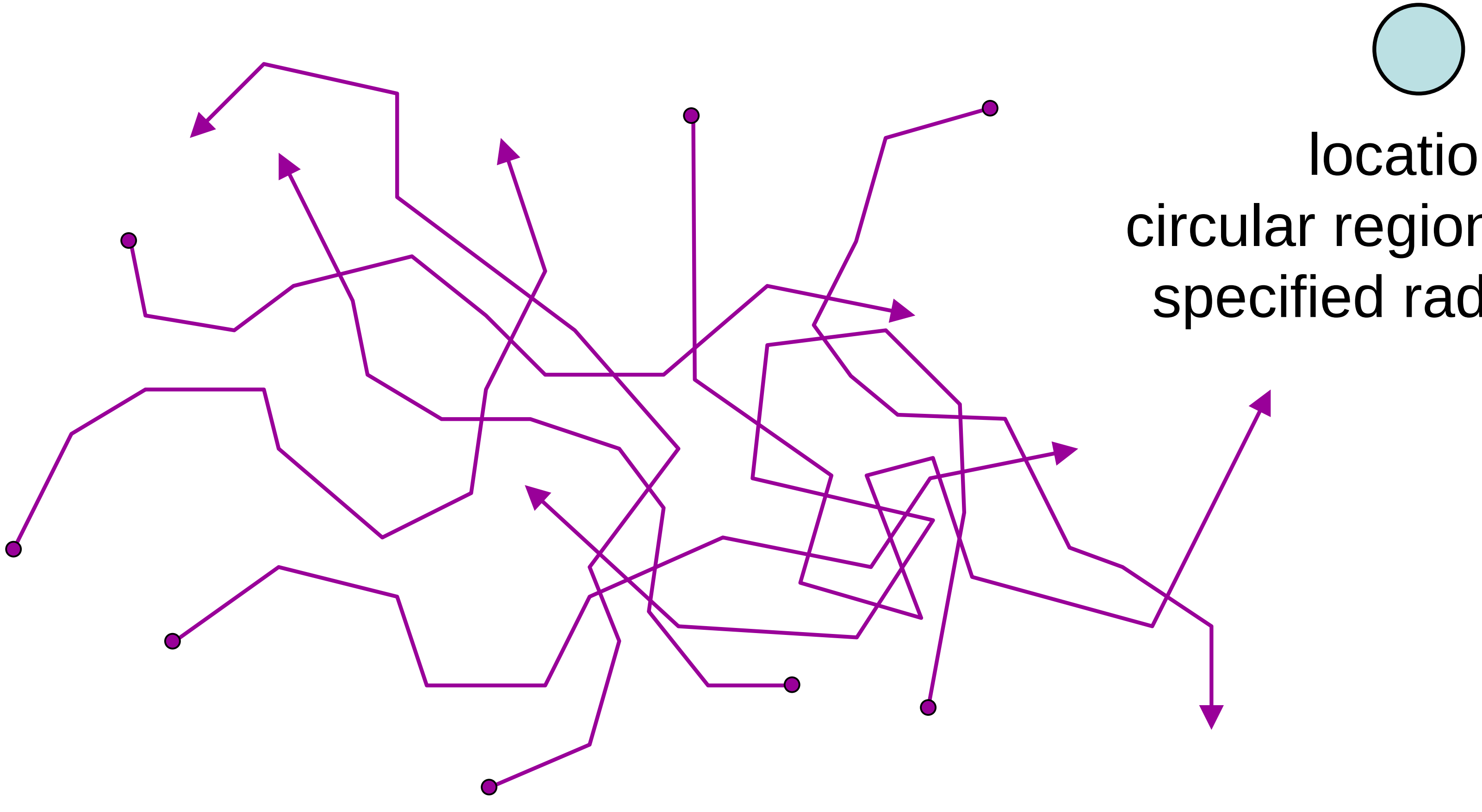
- The following 8 slides are from the presentation of Marc van Kreveld (MK)

- 1) What kind of questions can be answered by mining trajectory data?
  - Example and a simple algorithm:
    - Most visited locations ?

- 2) What kind of patterns we can identify in trajectory data?
  - More about this topic in the article by (Dodge et al., 2008)

1) Example pattern in trajectories (MK)

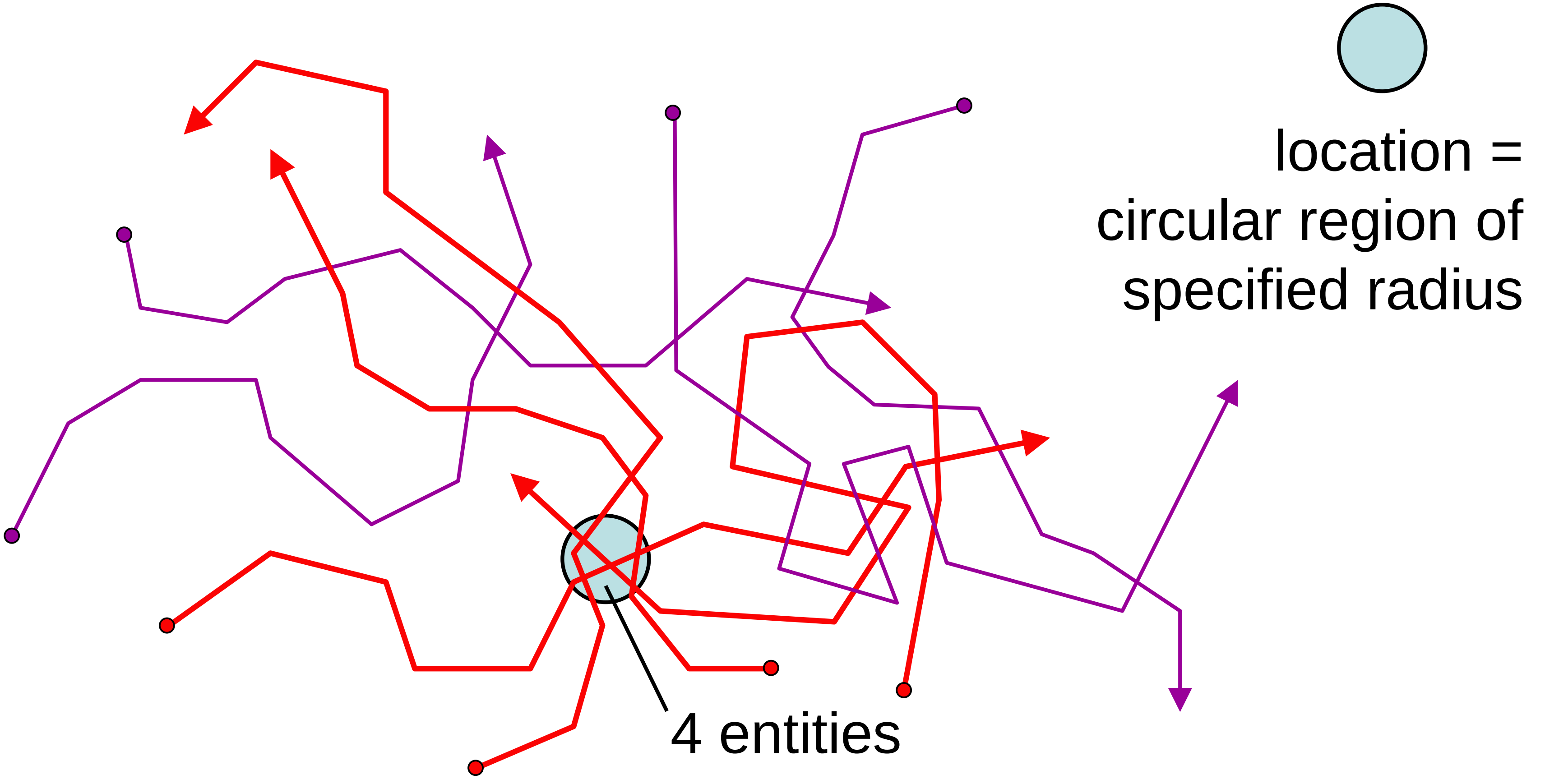
- What is the location visited by most entities? Useseimmin käyty paikka? (frequent visit)



location = circular region of specified radius

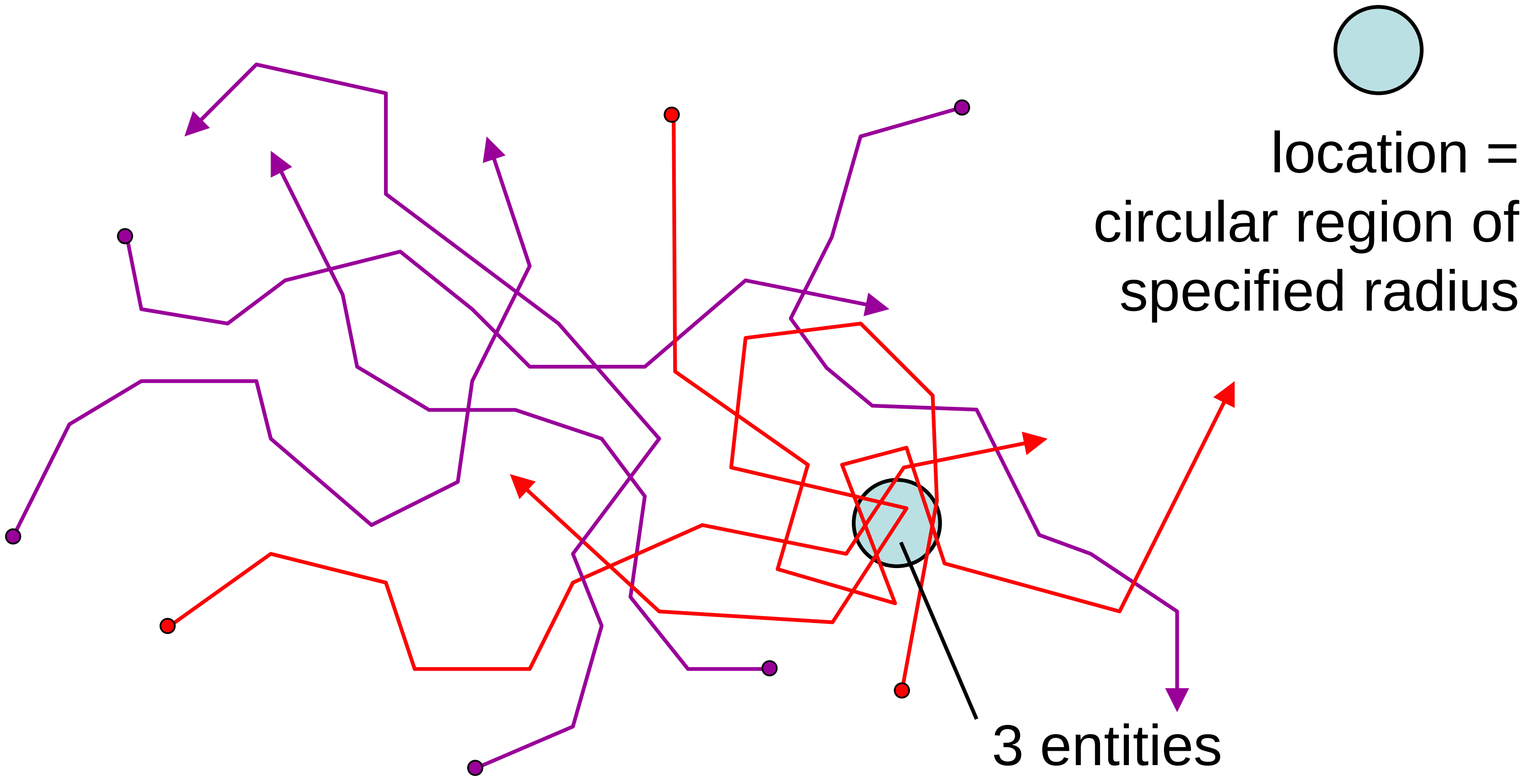
Example pattern in trajectories (MK)

- What is the location visited by most entities?



Example pattern in trajectories (MK)

- What is the location visited by most entities?



2) Trajectory pattern types (MK)

- Flock (lauma)
  - Several entities are withing a circular region and they move in the same direction
- Leadership (johto)
  - One entity is heading a flock in a specified direction
- Convergence (lähentyminen)
  - Several entities pass the same circular region
- Encounter (kohtaaminen)
  - Several entities are simultaneously inside the same circular region (assuming they keep their speed and direction)

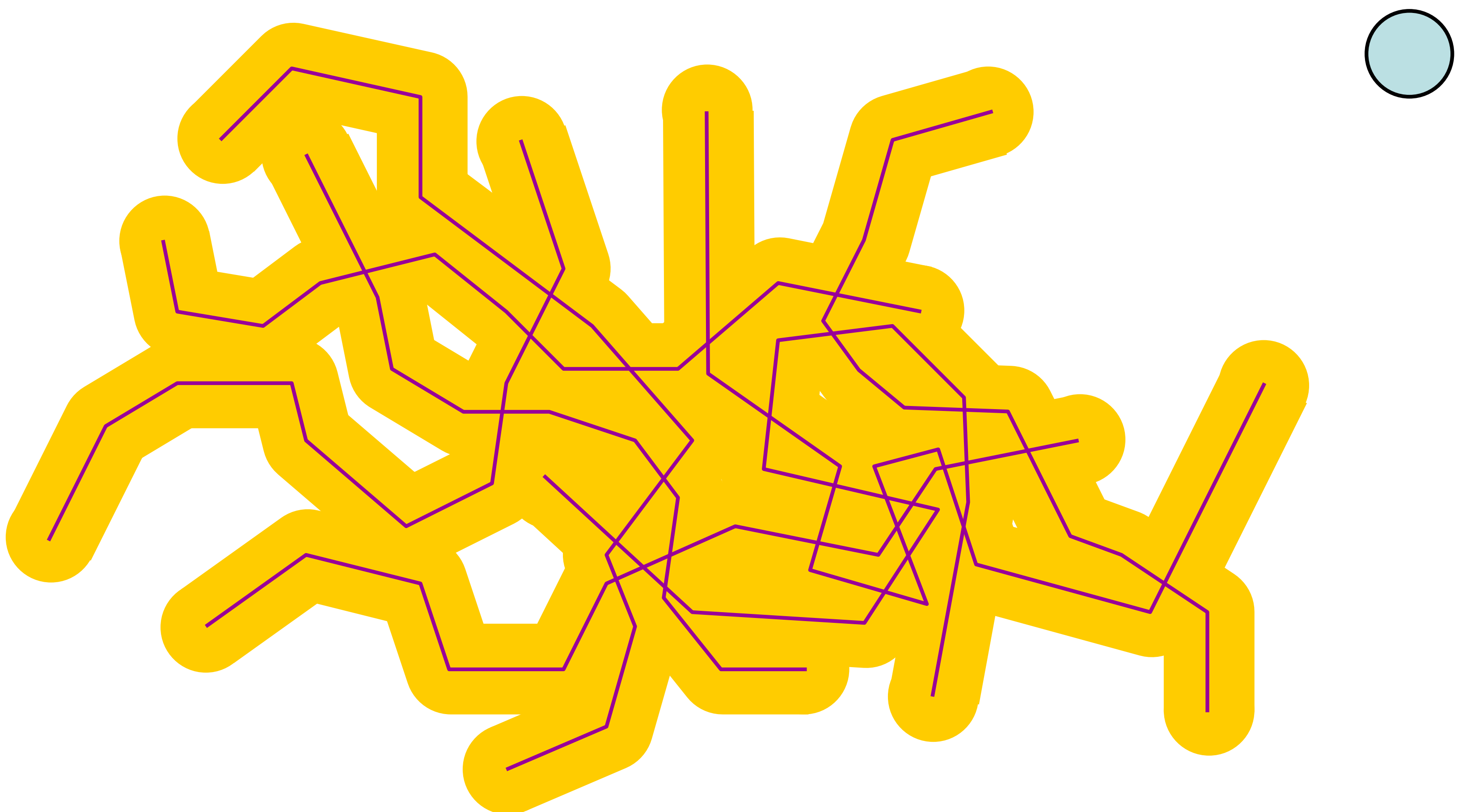
Example: Finding a flock

- **Flock**: near positions of (sub)trajectories for some subset of the entities during some time (liikeratojen osajoukko; kesto)
  - clustering-type pattern
  - different definitions are used

- Given: radius r, subset size m, and duration T, a **flock** is a subset of size $\geq m$ that is inside a (moving) circle of radius r for a duration $\geq T$

Example: Finding a flock

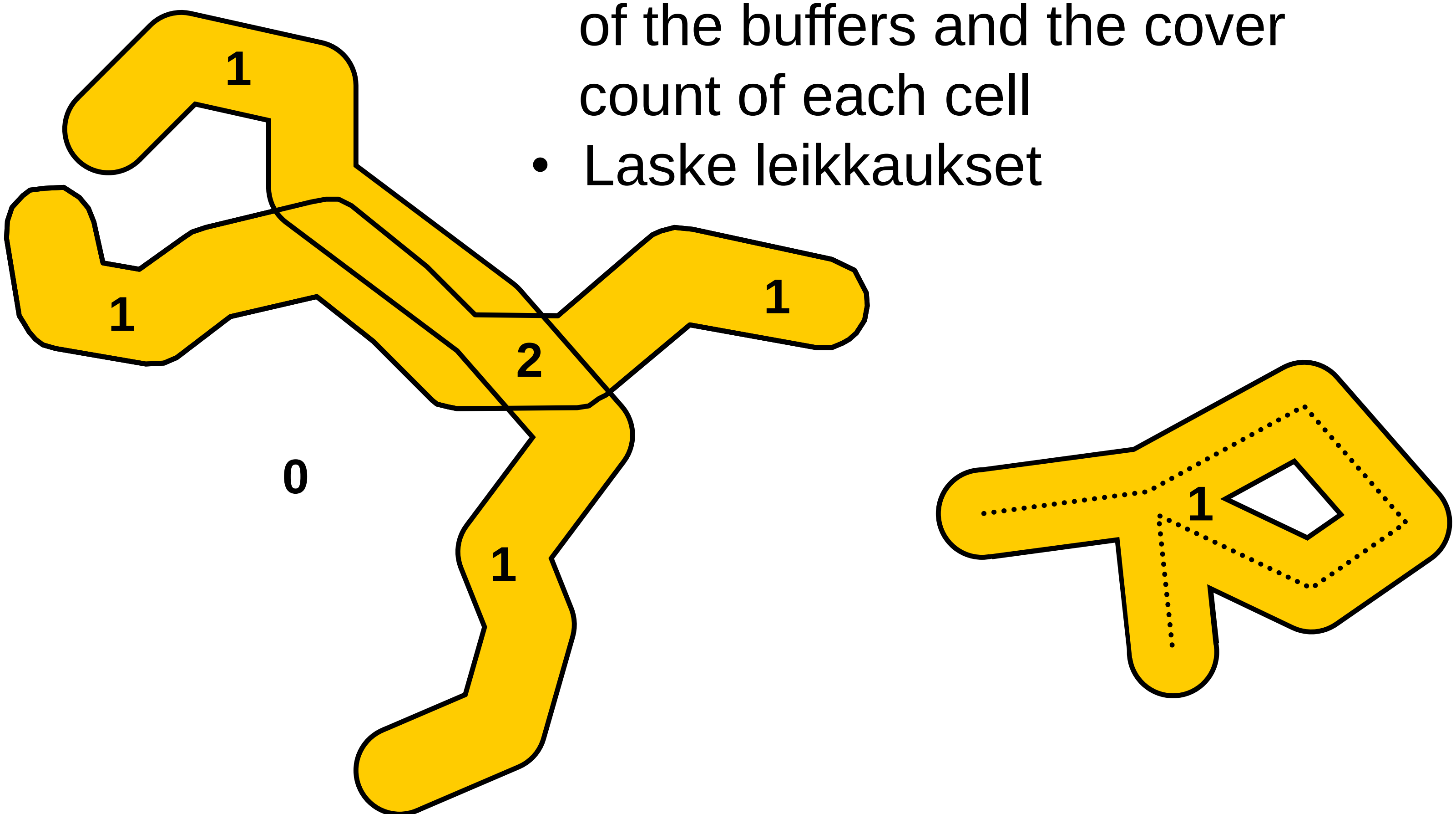
- Compute buffer of each trajectory
- Lasketaan puskureita käyttämällä



Example: Finding a flock

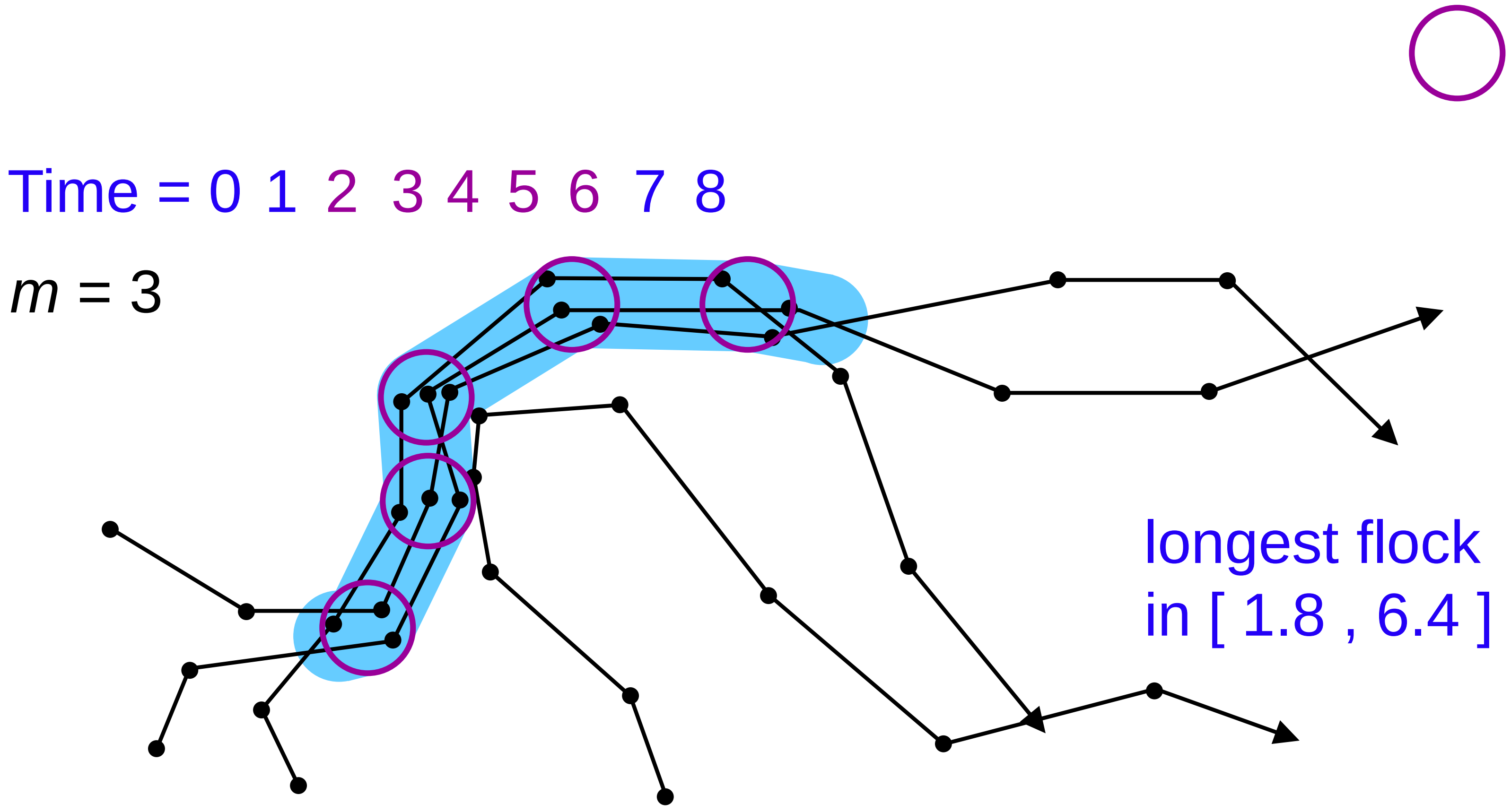
- Compute buffer of each trajectory
- Laske puskuri kaikille

- Compute the arrangement of the buffers and the cover count of each cell
- Laske leikkaukset

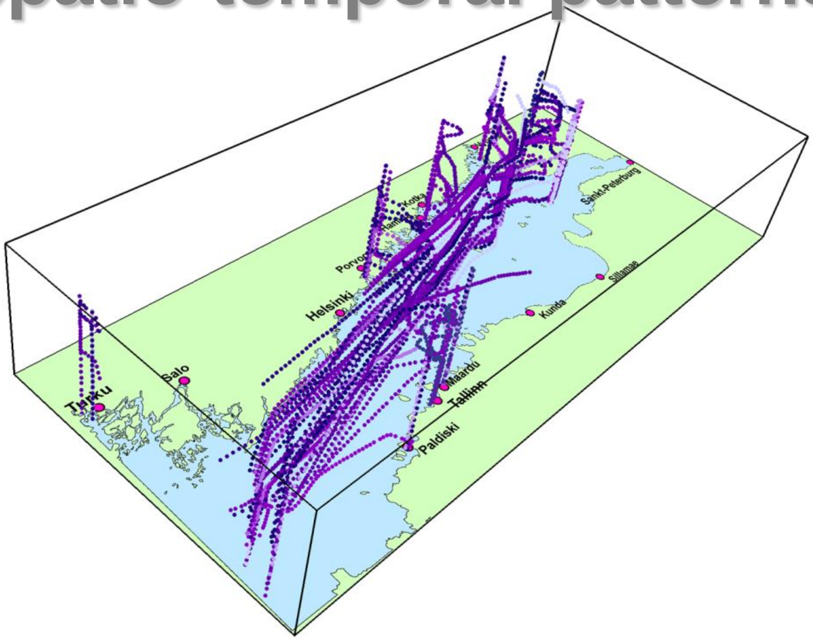


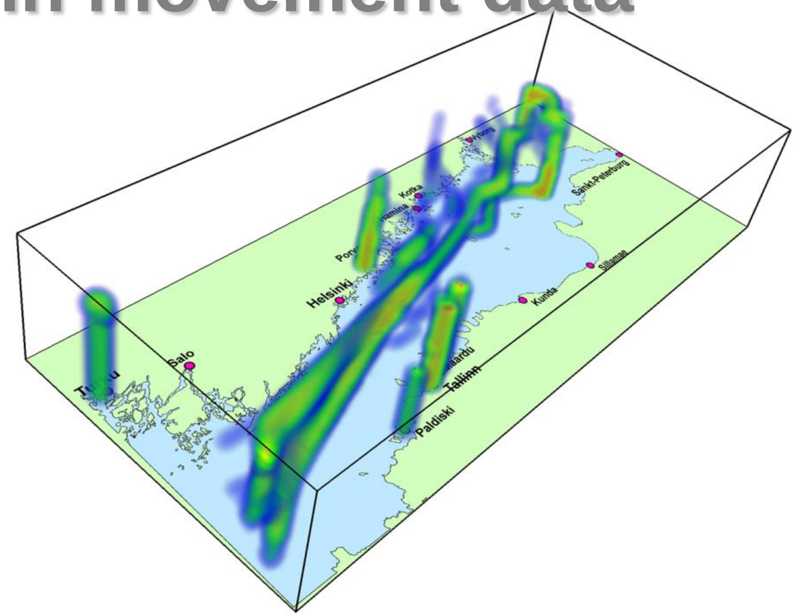
Example: Finding longest flock

- **Longest flock:** given a radius *r* and subset size *m*, determine the longest time interval for which *m* entities were within each other's proximity (circle radius *r*)



Space-time density of trajectories: exploring spatio-temporal patterns in movement data





Dr Urška Demšar,
National Centre for Geocomputation, NUI Maynooth, Ireland

Prof Kirsi Virrantaus,
Department of Surveying, Aalto University, Finland

National Centre for Geocomputation
An tIonad Náisiúnta Geoiríomhaireachta

NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

Aalto University
School of Science and Technology

Problem:
Space-time density of real data –
vessel trajectories in the Gulf of Finland

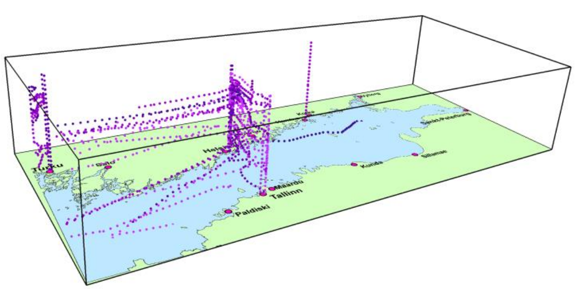
Two data sets:

- -one day AIS (the 26th of January, 2008)
- -one month AIS data (January 2008)

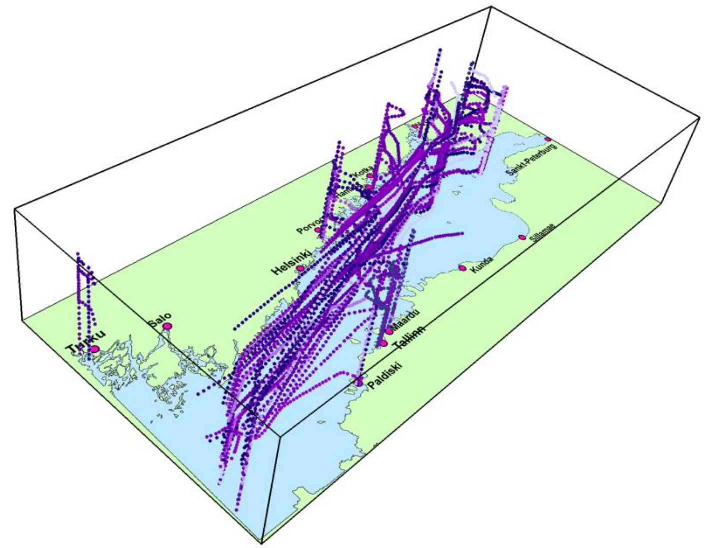
AIS = Automatic Identification System based on GPS positioning

Densities for 2 vessel types were produced by using trajectory data of:

- -passenger ships (ferries)
- -tankers



-trajectories



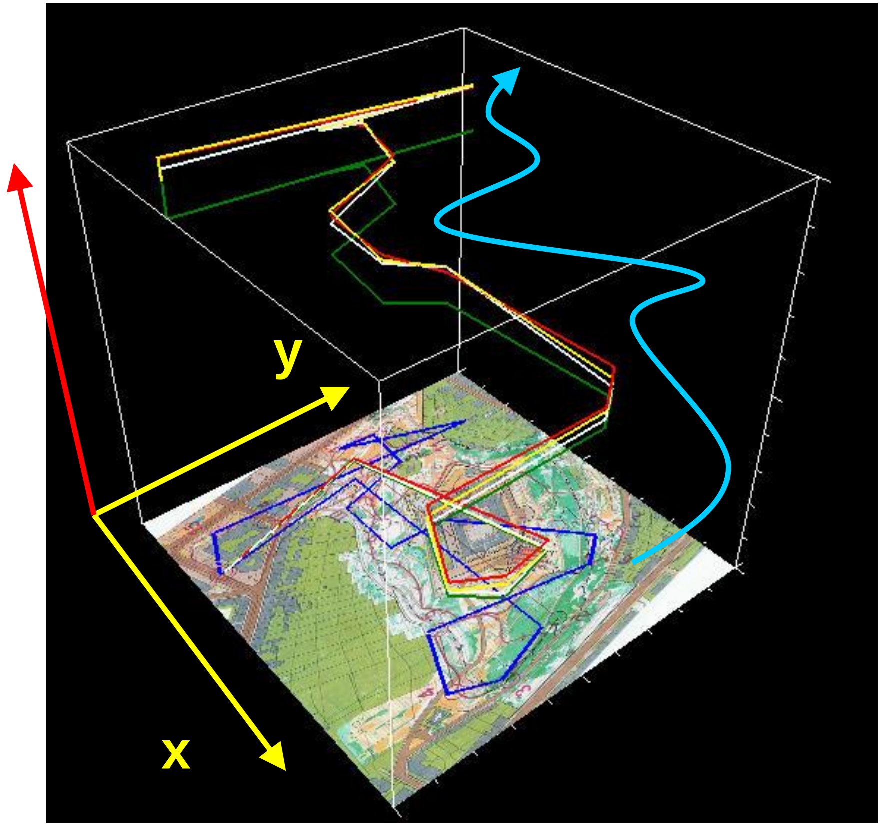
Method:
Time in 3D – Space-Time Cube and its variations

Space-time cube – developed in time geography to show people's movements through both geographic time and space. Applications: GPS traces of people/animals/objects – ships, vehicles, airplanes, etc.

Movement of an object shown as **trajectory** in space and time:

- 3D points (x1,y1,t1), (x2,y2,t2), ..., (xn,yn,tn),
- points linked with straight linear lines (linear segments)

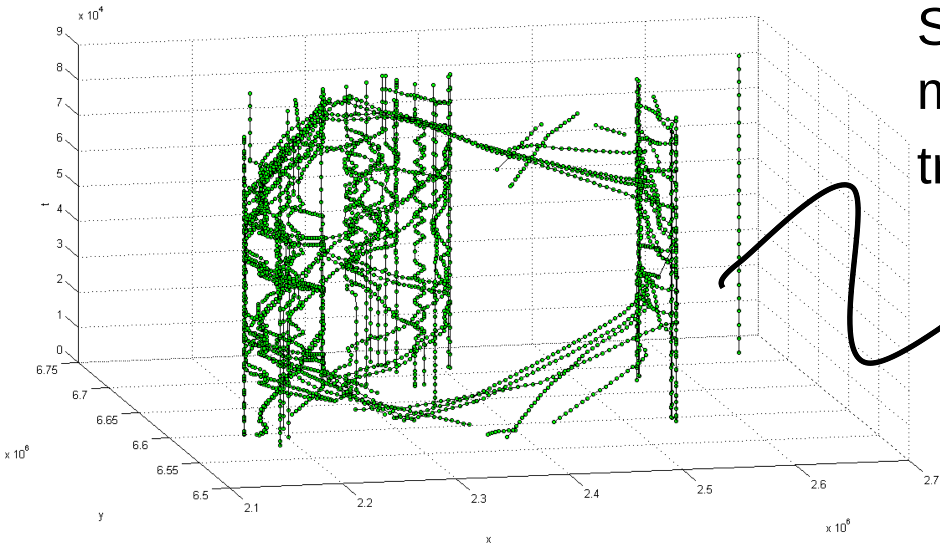
z axis –
time t



trajectories

x-y plane =
geographic
map

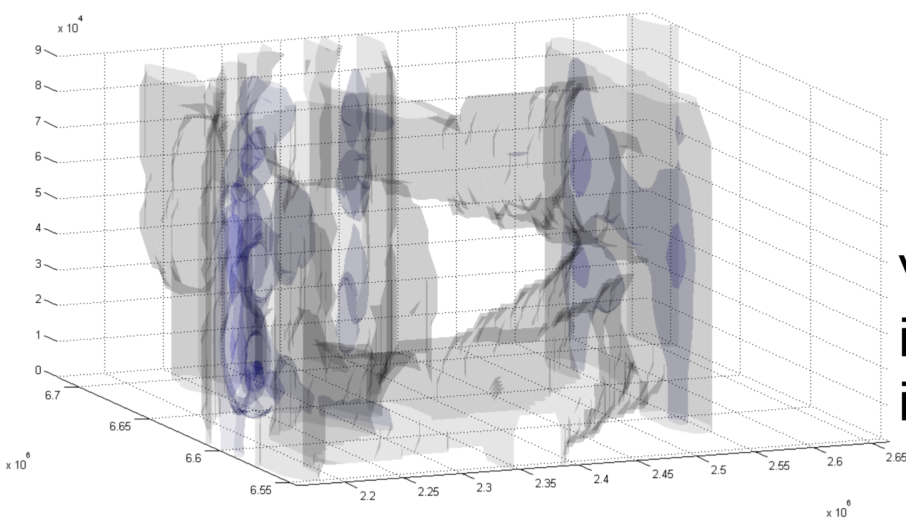
Analysis of the problem:
What is the problem – clutter and overprinting in space-time cube



Space-time cube becomes messy when there are too many trajectories (clutter & overprinting)

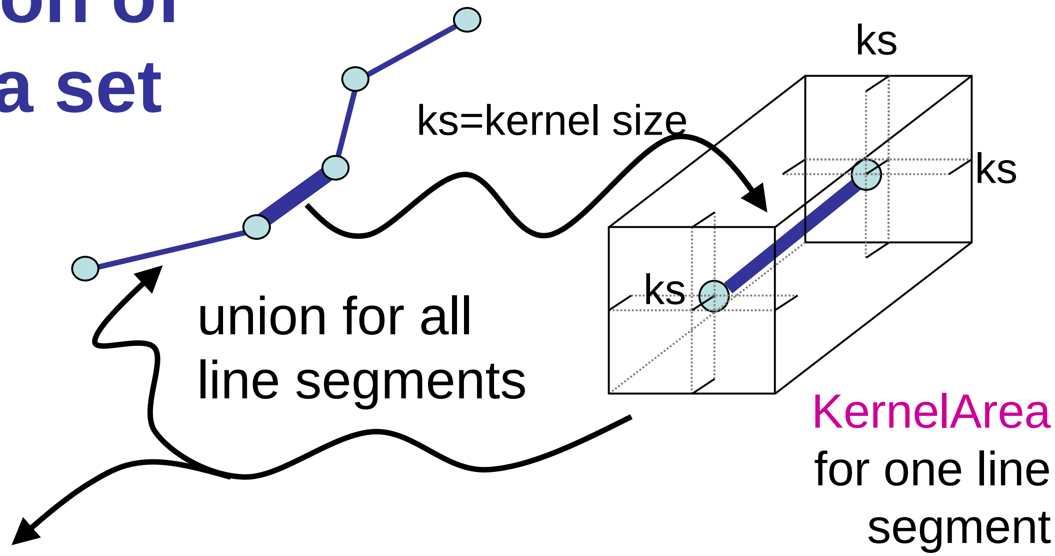
Vessel trajectories for 58 passenger ships

Solution:
Space-time density of trajectories

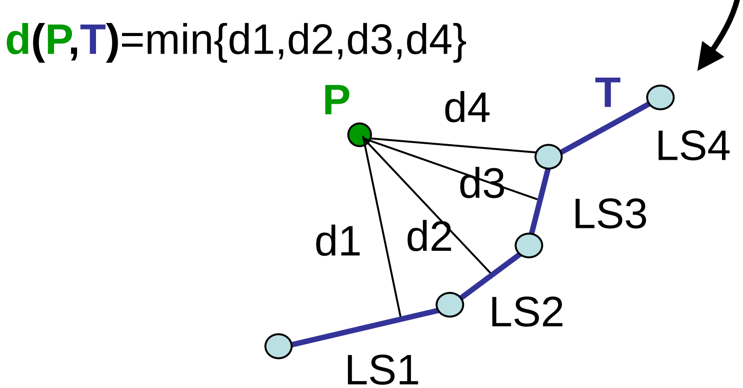


Volume visualisation of importance of each trajectory in space and time

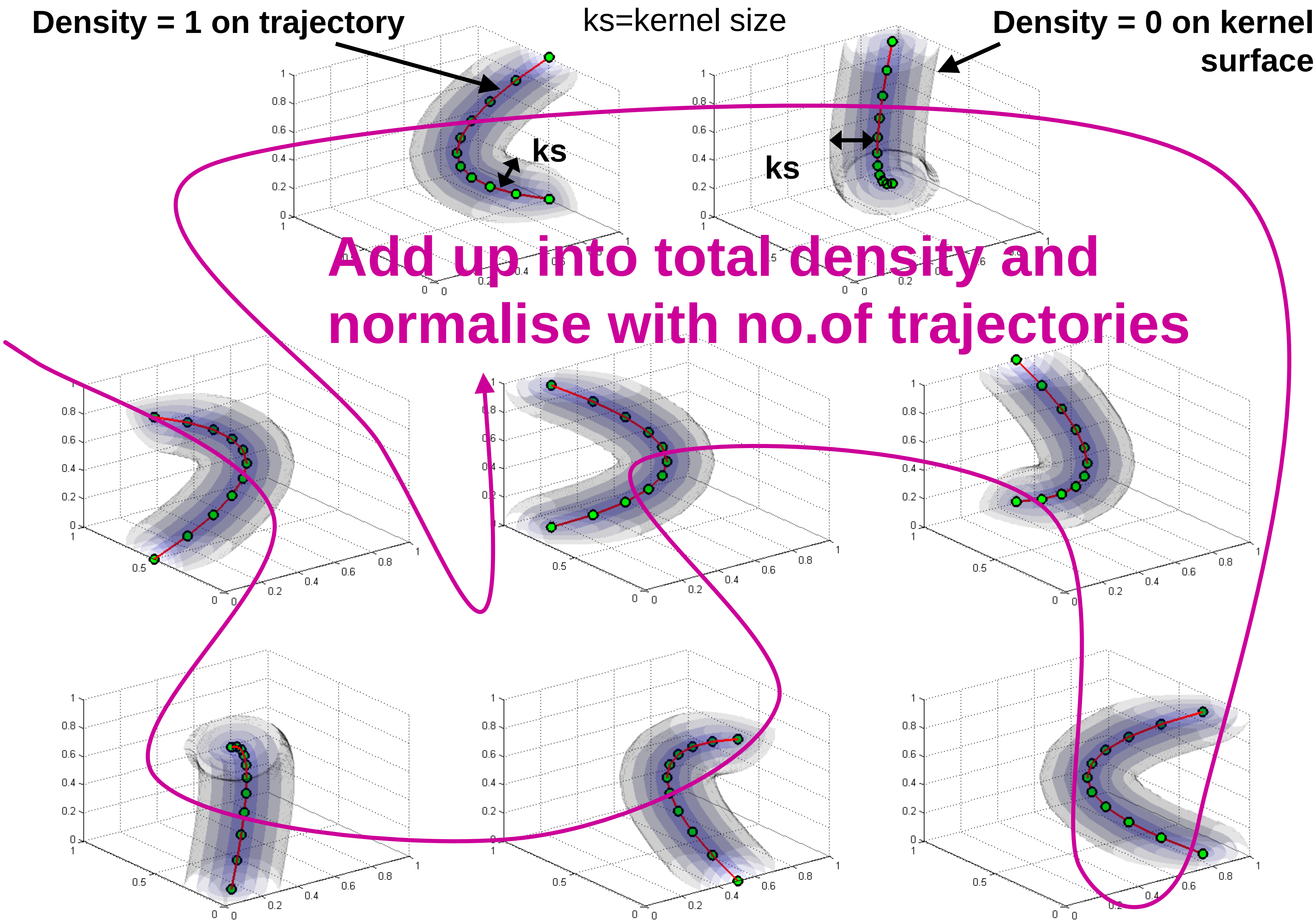
Algorithm for calculation of space-time density of a set of trajectories



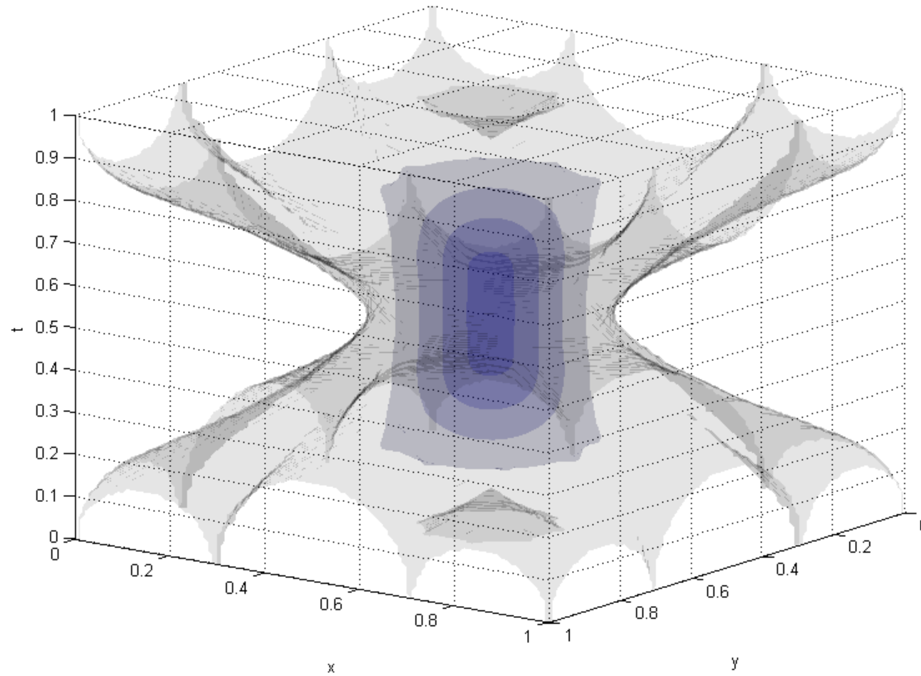

```
TotalDensity = 0;

for each trajectory T

    TrajectoryDensity = 0;

    calculate KernelArea around the trajectory;

    for each voxel P in the KernelArea
        calculate DistanceToTrajectory T;
        TrajectoryDensity = normalised DistanceToTrajectory
        with kernel size;
    end

    TotalDensity = TotalDensity +
                   TrajectoryDensity;

end

Normalise TotalDensity with number of trajectories;
```



Densities around each of eight simulated trajectories

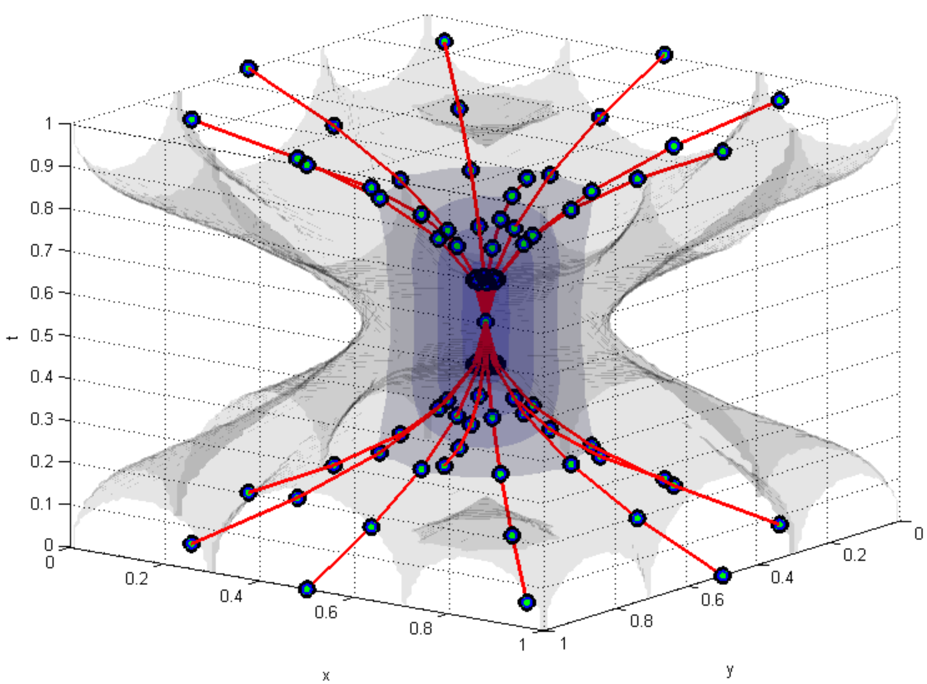


Total density of eight simulated trajectories



Shown with isosurfaces at 0.8, 0.6, 0.4, 0.2 and 0.

Same colour for all isosurfaces, different transparency.



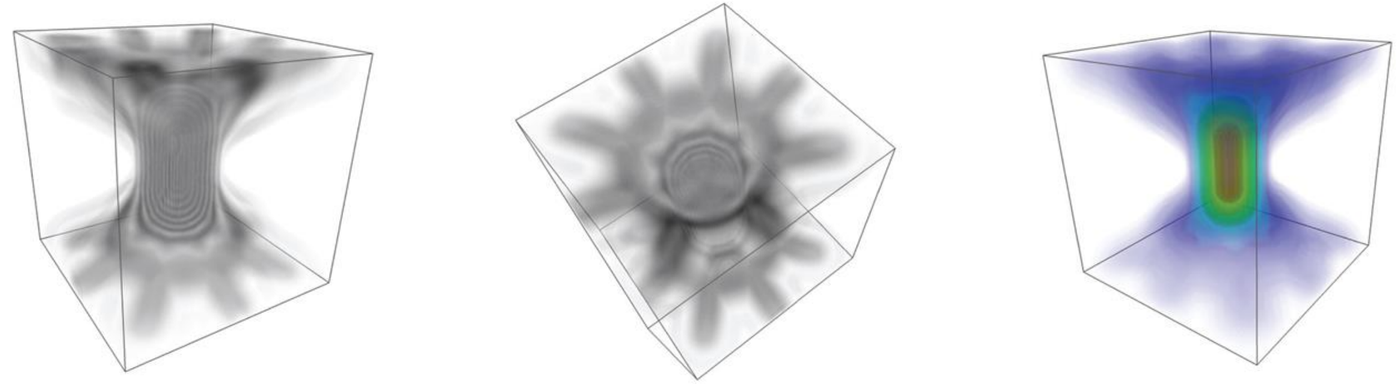
Density + superimposed trajectories

Visualisation possibilities – volume visualisation

3 methods:
Direct volume rendering
Isosurfaces
Clipping planes

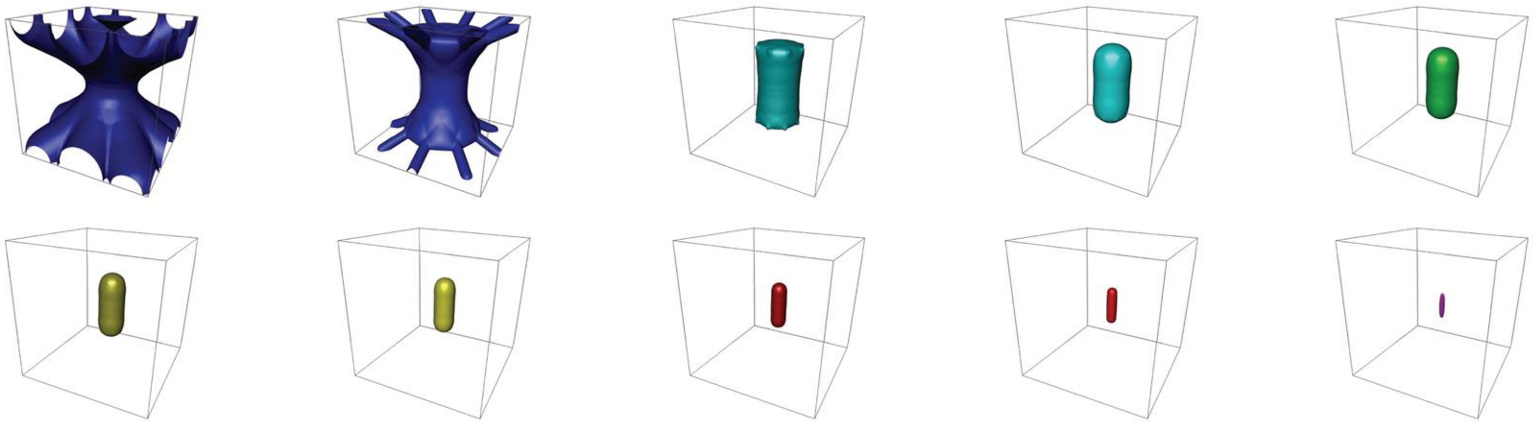
Direct volume rendering

-Transparency according to the scalar field values



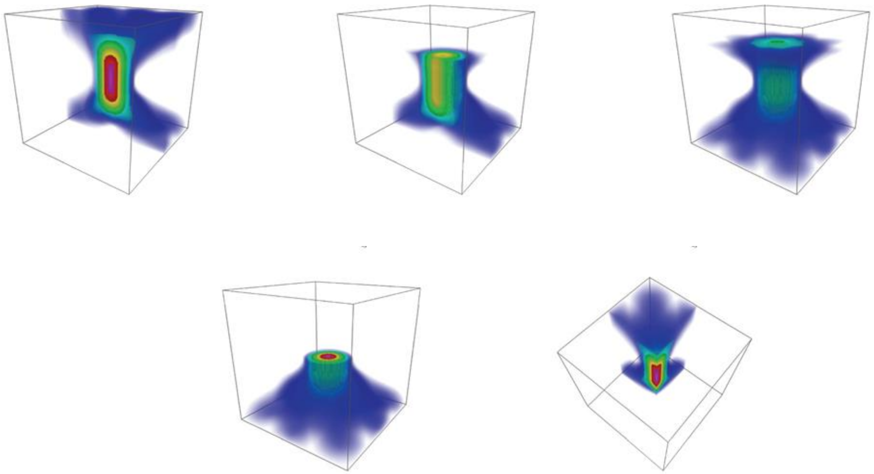
Isosurfaces

-2d surfaces that share the same scalar value; distribution of one particular value



Clipping planes

-cross section of a volume



Space-time density of real data – vessel trajectories in the Gulf of Finland

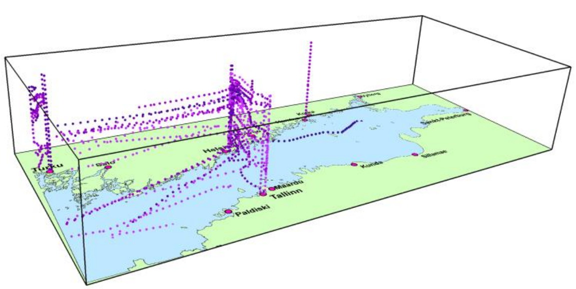
Two data sets:

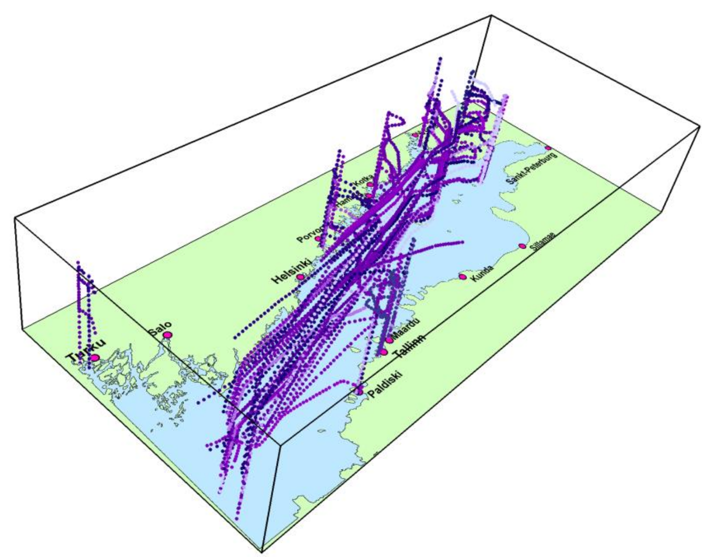
- -one day AIS (the 26th of January, 2008)
- -one month AIS data (January 2008)

AIS = Automatic Identification System based on GPS positioning

Densities for 2 vessel types were produced by using trajectory data of:

- -passenger ships (ferries)
- -tankers



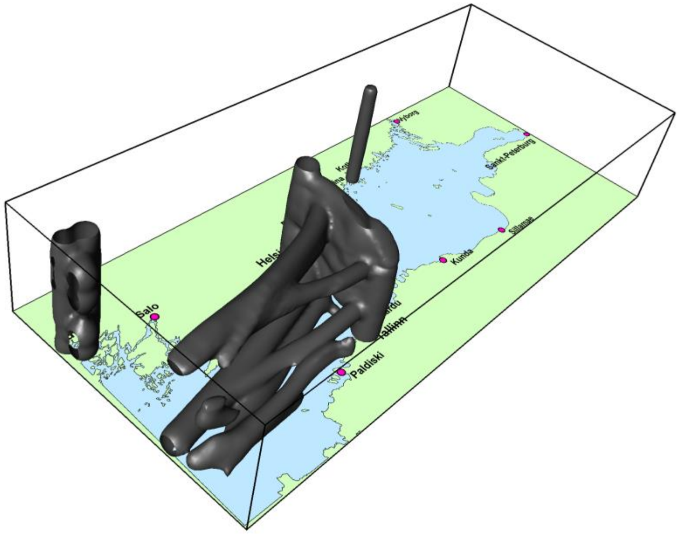


-trajectories

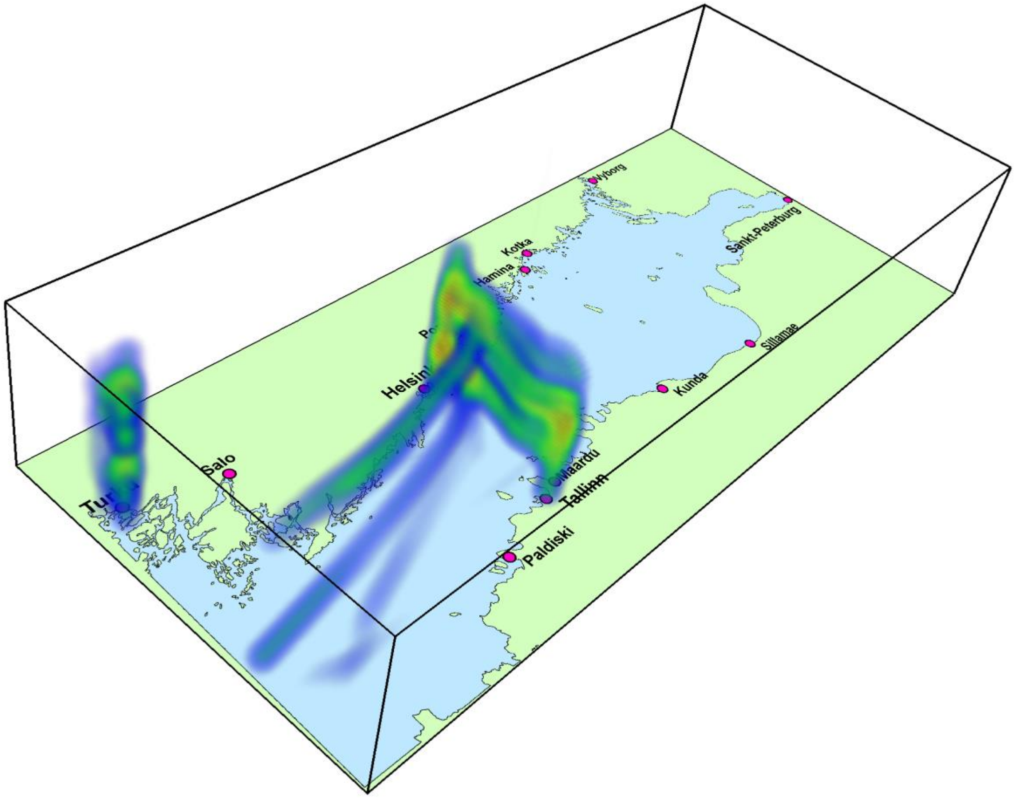
Visualisation possibilities and spatio-temporal patterns

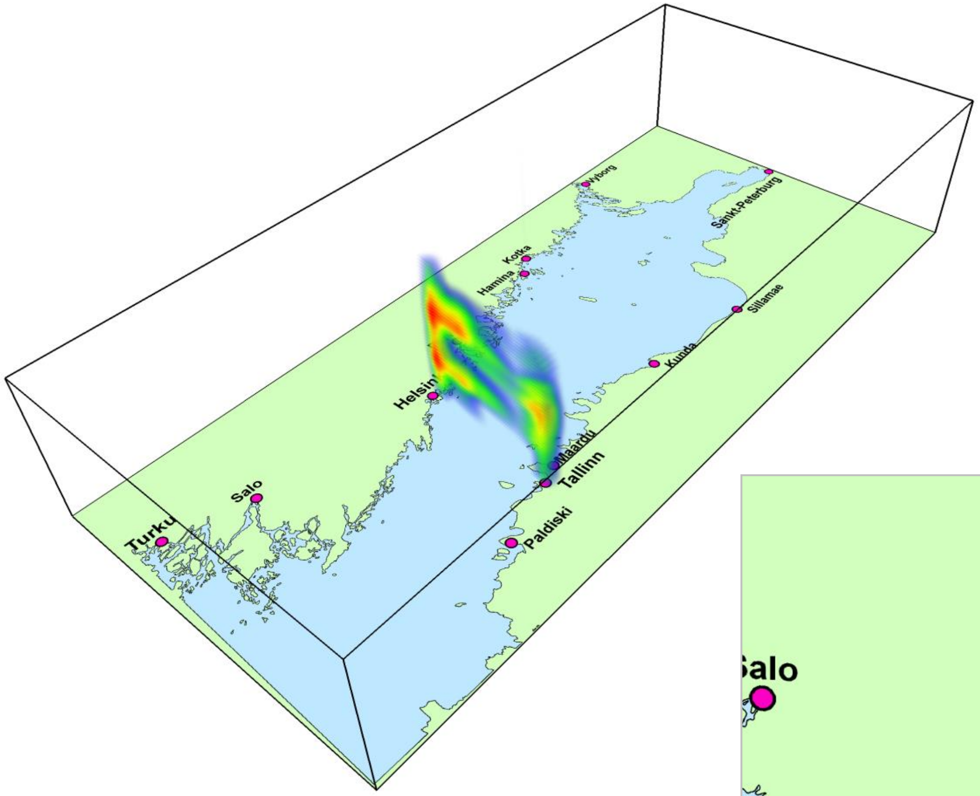
Same three possibilities as for simulated data

Isosurfaces

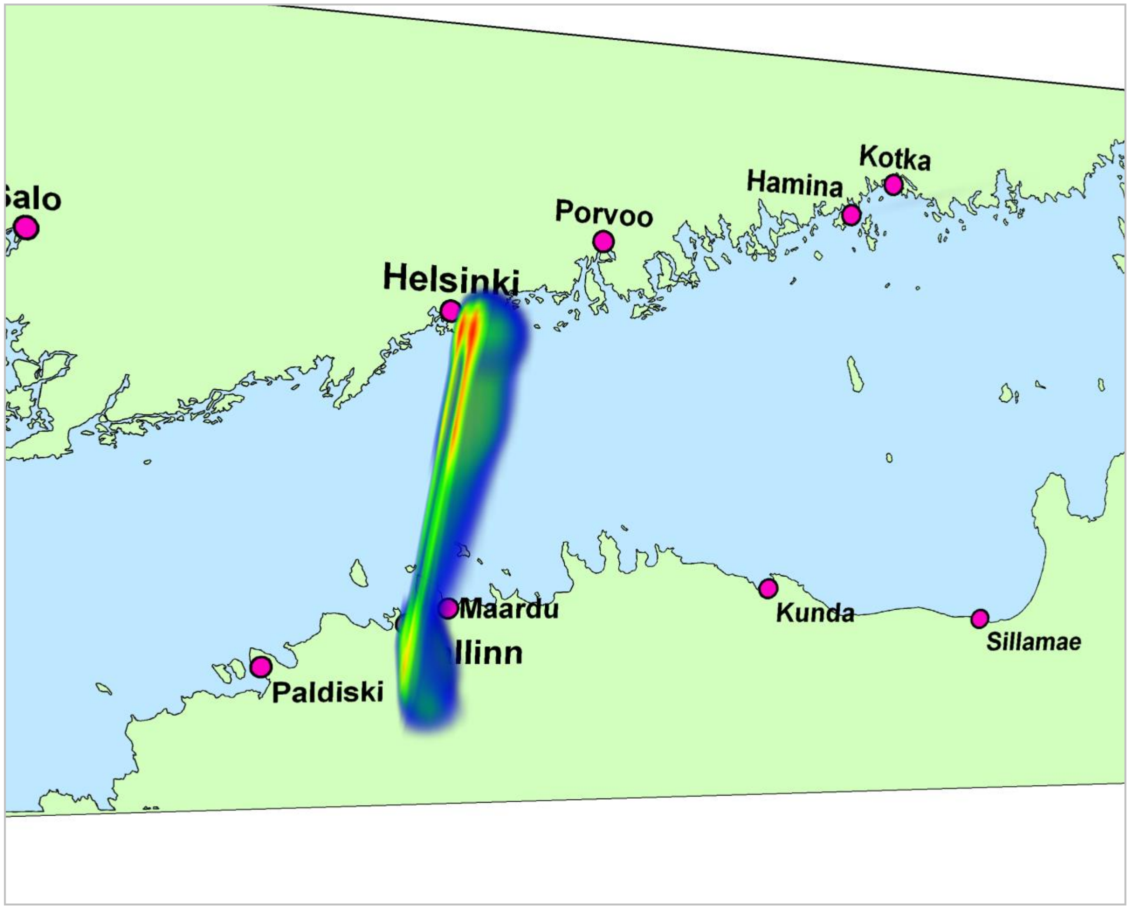


Direct volume rendering

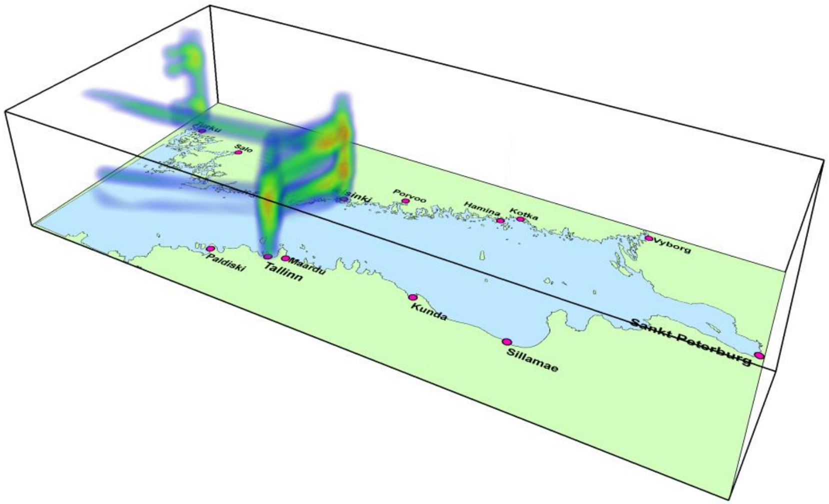


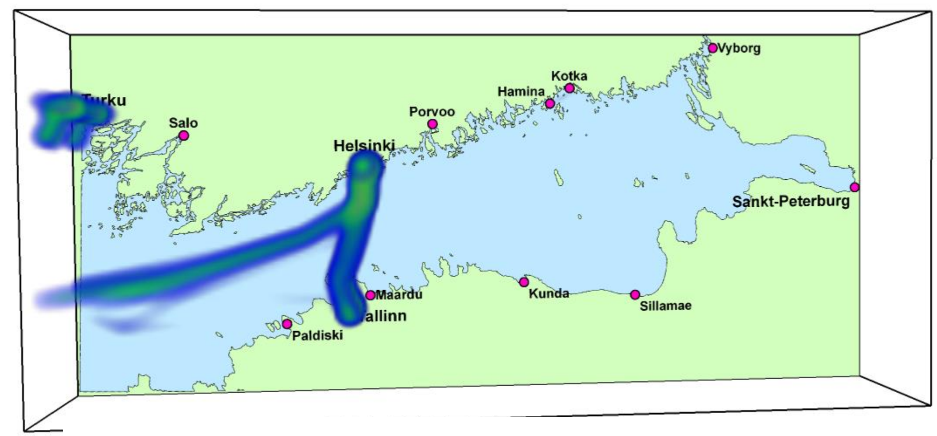


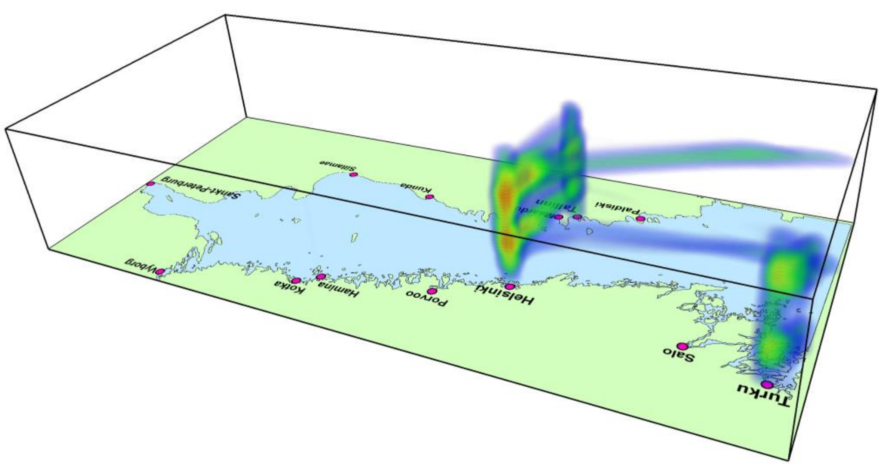
Clipping planes

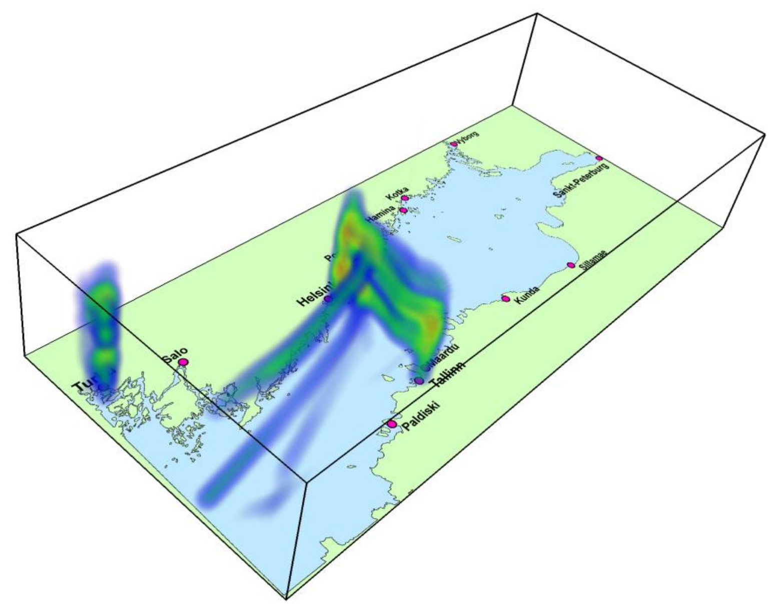


Results – passenger ships – monthly density

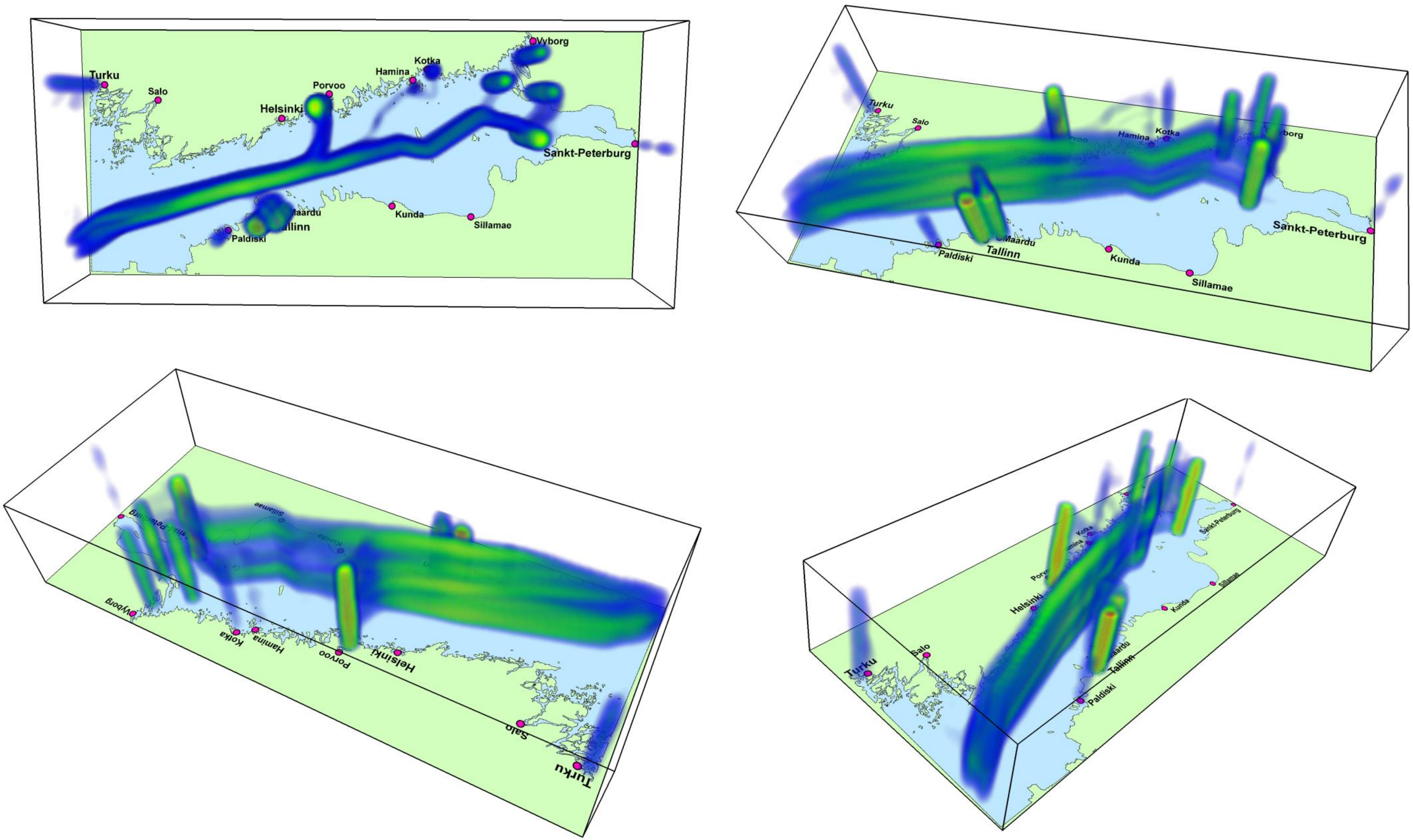




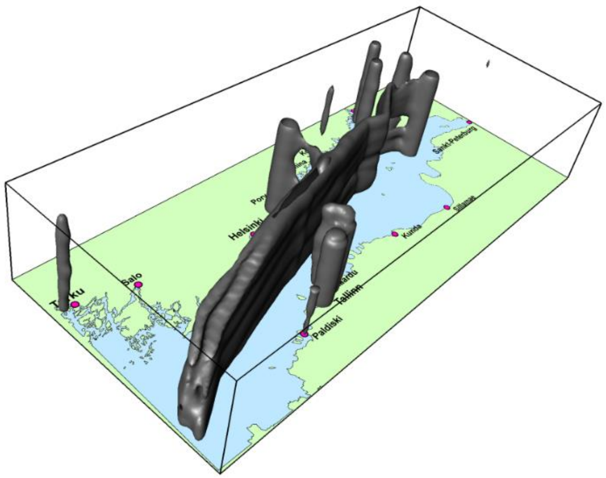


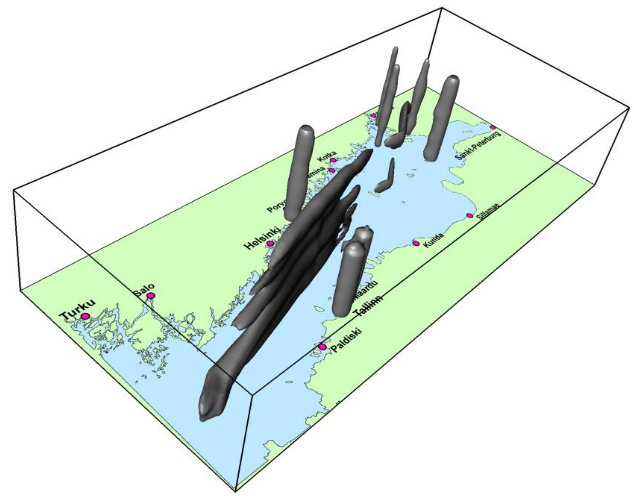


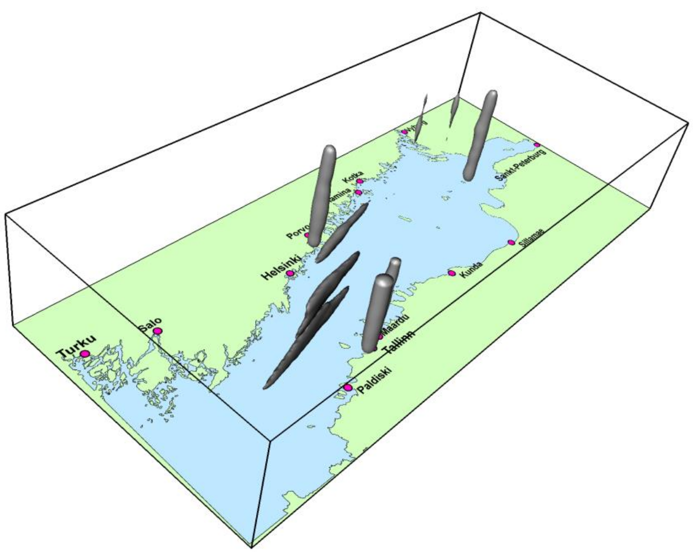
Results – tankers – monthly density – DVR

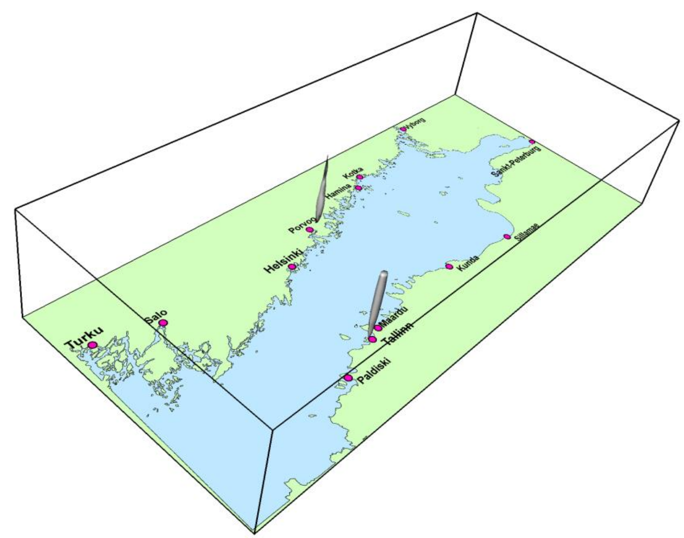


Results – tankers – monthly density - isosurfaces

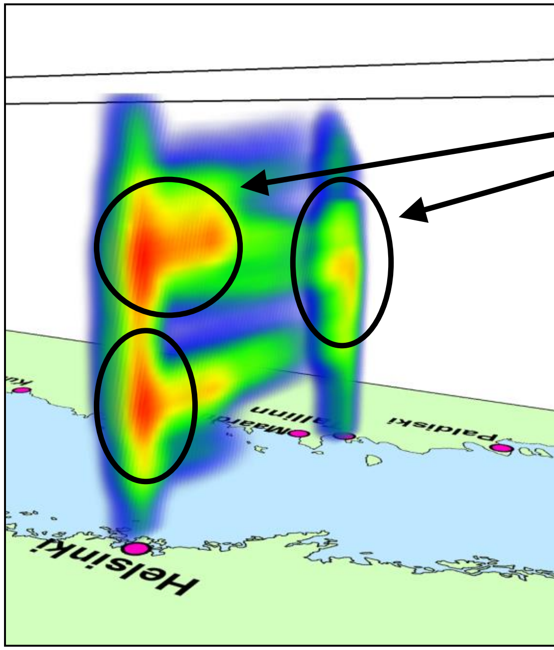






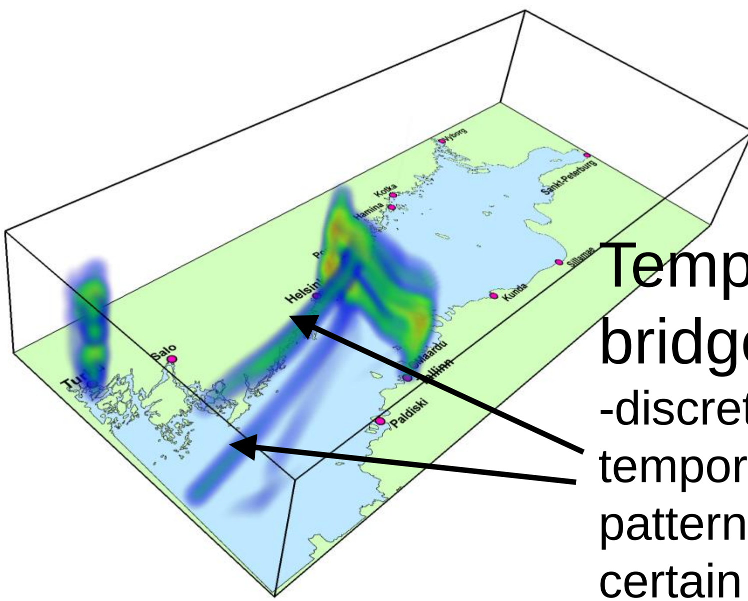


Spatio-temporal patterns to look for

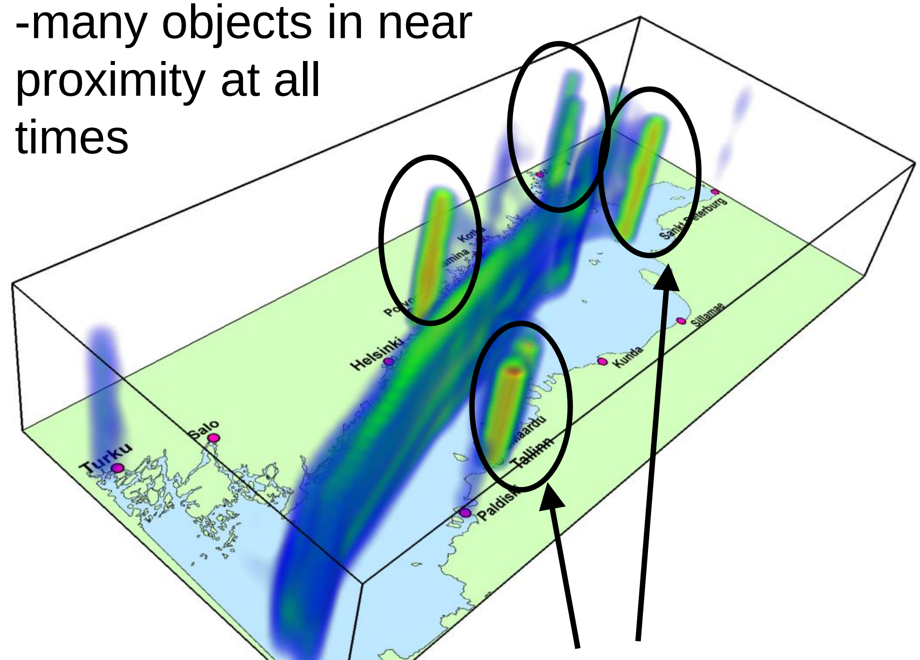


Spatio-temporal **hotspots**
-convergence of trajectories
in both space and time

-many objects in near
proximity at all
times



Temporal
bridges
-discrete
temporal
pattern at
certain times



Temporal towers

4. Semantic trajectories – future of trajectory data mining

- Paper by Parent et al., ”Semantic trajectories modeling and analysis”
- The paper introduces concepts of
  - Raw trajectories

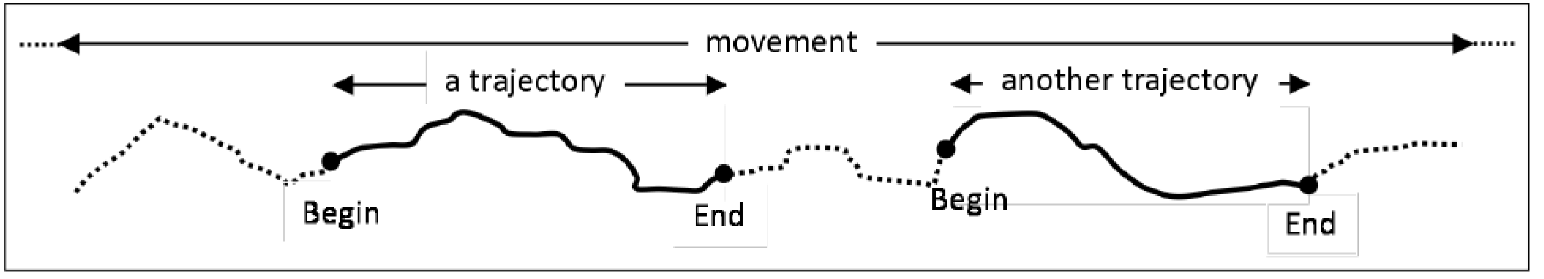


Figure 3 - Trajectories extracted from a movement track visualized as dotted line

Definition: A ***raw trajectory*** is a trajectory extracted from a raw movement track and containing only raw data for its Begin-End interval. It is defined as a tuple:

(trajectoryID, movingObjectID, trace: LISTOF position(instant, point, δ))

where δ denotes a possibly empty list of additional raw data (e.g. speed, direction). □

- Semantic trajectories
  - Knowledge added to raw trajectories from contextual data repositories = semantic enrichment = raw data is completed with annotations
  - Stops, moves and episodes are recognized

Definition: A ***semantic trajectory*** is a trajectory that has been enhanced with annotations and/or one or several complementary segmentations. It is defined as a tuple (a full explanation is given in [Spaccapietra and Parent 2011]):

(trajectoryID, movingObjectID, trajectoryAnnotations,
trace: LISTOF position (instant, point, δ, positionAnnotations),
semanticGaps: LISTOF gap (t_1, t_2),
segmentations: SETOF segmentation (segmentationID,
episodes: LISTOF episode (t_3, t_4, definingAnnotation, episodeAnnotations))) □

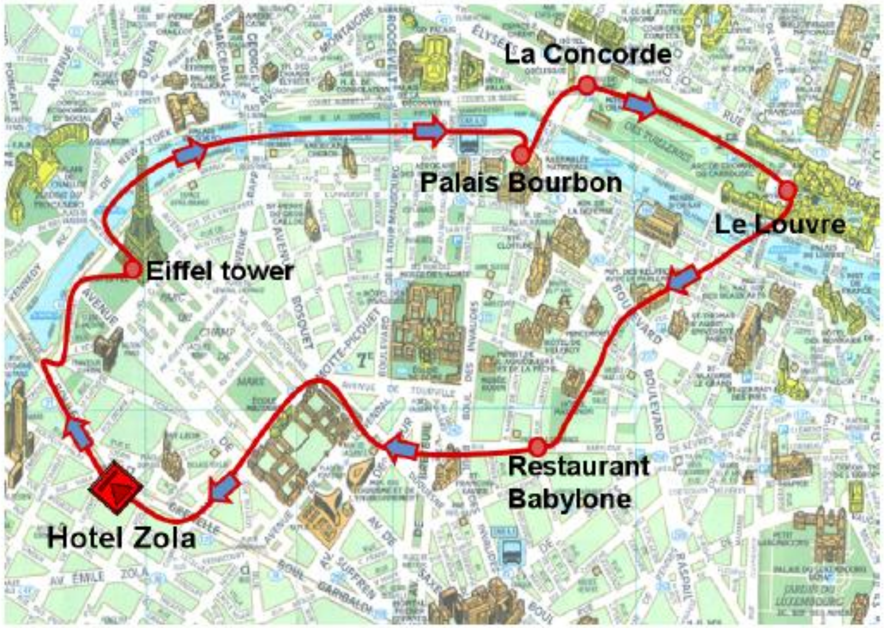


Figure 1 - 2D visualization of a one-day spatial trace left by a tourist visiting Paris – background map downloaded from Mappery.com, copyright unknown

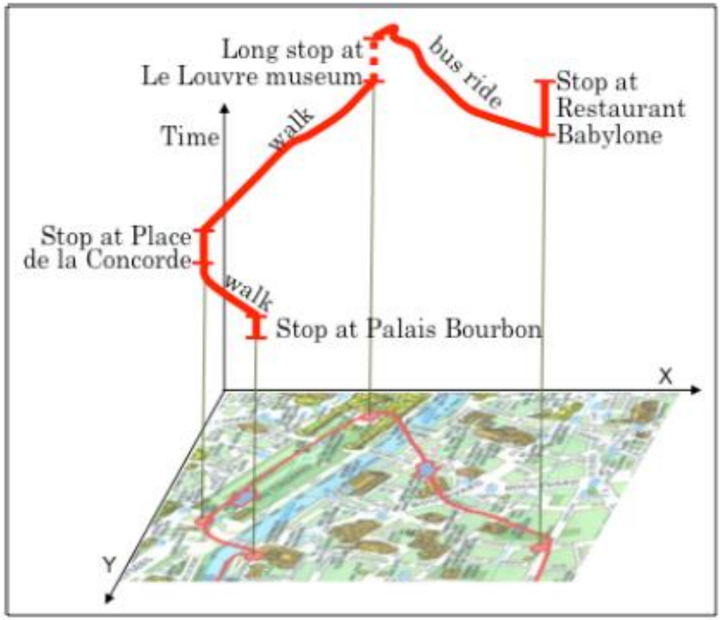


Figure 2 - A time-geography diagram showing part of the previous tourist track

Semantic trajectory

Processing of trajectories – many challenging topics

- Trajectory cleaning
- Map matching

- Semantic enrichment
- Using clustering and classification

- Discovering behaviours

- Privacy issues

Cleaning and map matching

- Trajectories are due to errors in same way that any other positioning (systematic and random errorr occur)
- Random errors can be cleaned by statistical methods and for example by spline interpolation

- Map matching means that trajectories are processed to fit to some existing network, like roads or streets (Juote,A., MSc thesis , 2016) by using for example distance and topology based methods or some probability theory based methods (Hidden Markov Model)

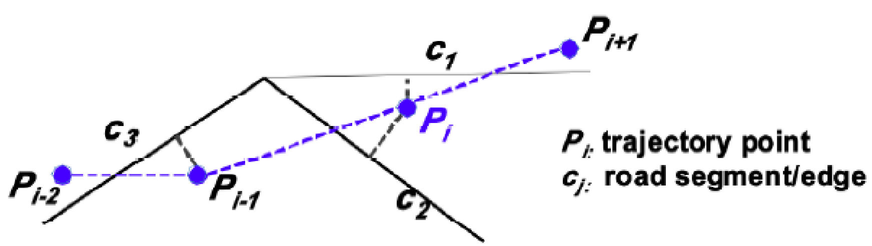


Figure 4 - The map-matching algorithm by [Brakatsoulas et al 2005]

Compression, segmentation, annotation

- Compression means that the amount of recorded data is decreased
- In segmentation some meaningful parts of the trajectory are identified, episodes
- Episode identification requires sematic context and process called annotation is related to this
- Annotation means that that some episodes get a ”label”
- Episode identification starts by identification of stops in the trajectories and adding semantics; entire trajectories can also be annotated

Data mining – identifying behaviours

- After preprocessing described the data is then mined by using regular methods like clustering and classification

- Data mining methods can reveal behaviours

- Some researchers have tried to make taxonomies out of the generic and application orieted behaviours

Taxonomy of trajectories

- Paper by Dodge et al. on taxonomy of trajectories

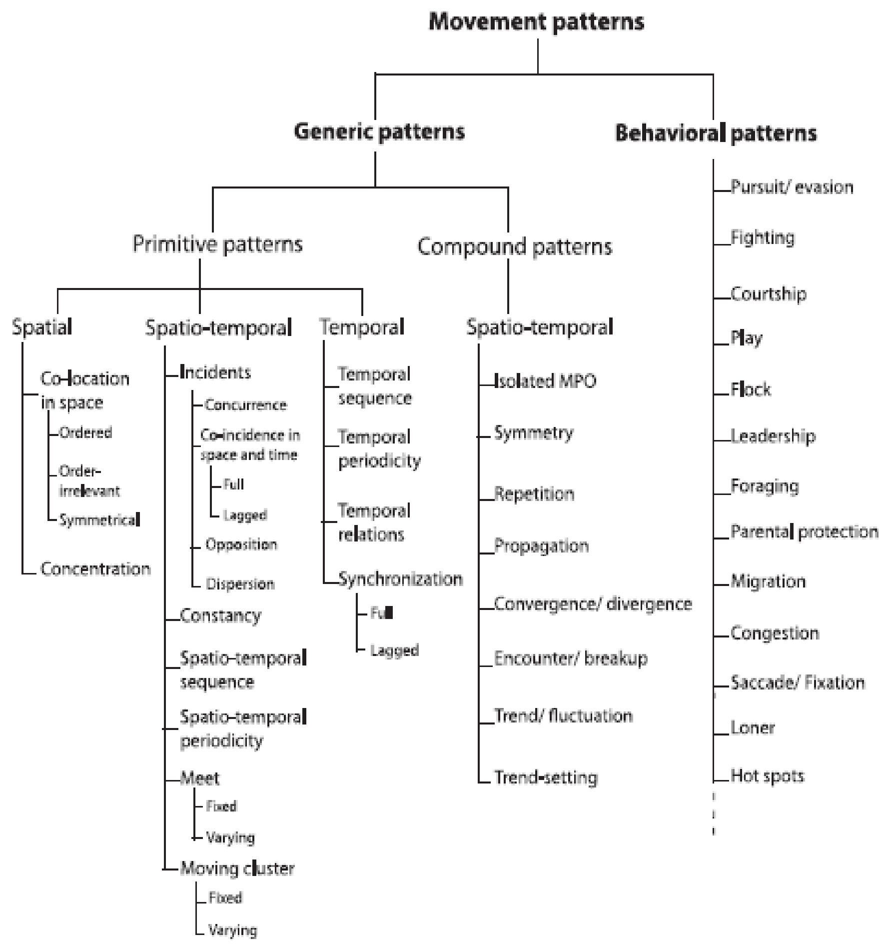


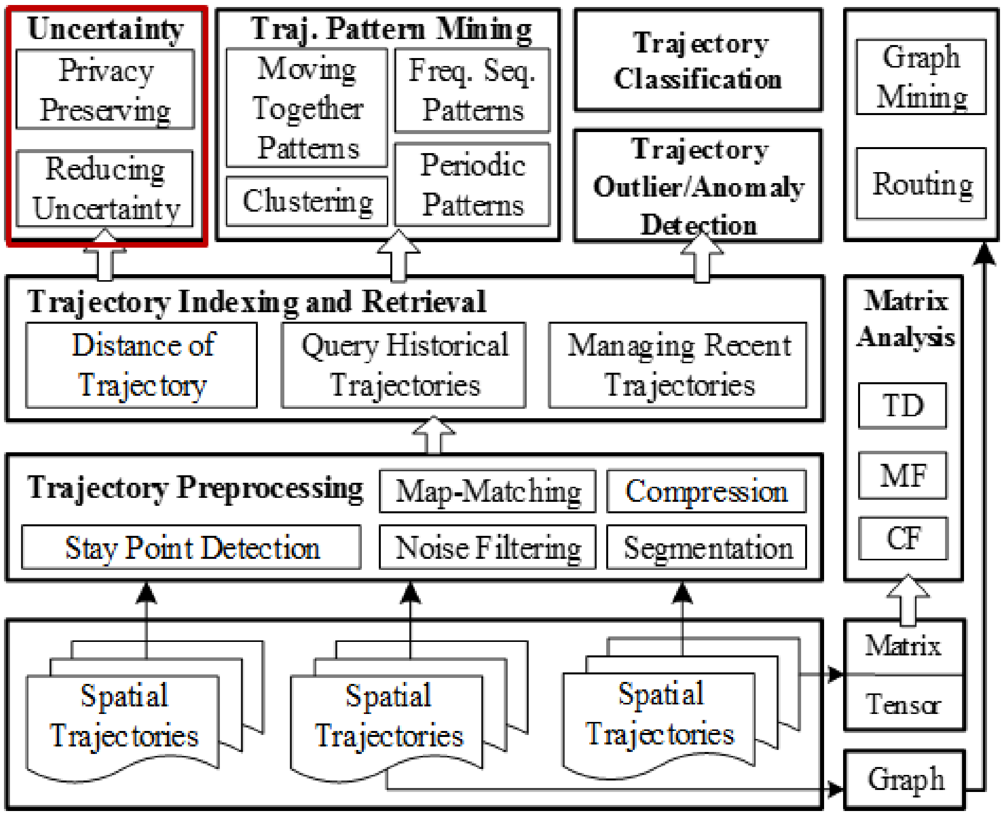
Figure 1: Classification of movement patterns

Geocoding of data without coordinates

- an interesting problem for research
- Analysis of data sets from social media, for example tweets including place names
- How to geocode that information into spatial data and even trajectories?
- Use of data mining and artificial intelligence
- MSc thesis ongoing (H. Grannabba)

Research probems on trajectory data

Yu Zheng. [Trajectory Data Mining: An Overview](). ACM Transactions on Intelligent Systems and Technology. 2015, vol. 6, issue 3.



Liikeratatiedon epävarmuus
Liikeratatiedon louhinta
-klusterointi
-luokittelu
-poikkeamat
Verkkodatan louhinta

Erityiskysymyksiä:
-karttasovitus
-tiivistys
-segmentointi

Literature/References

- Buchin,K., Buchin,M., Kreveld,van M., Luo,J, Finding long and similar parts of trajectories, https://pdfs.semanticscholar.org/796e/7bb3852cdfc73175829e8d5384eb033bee1c.pdf
- Mazimpaka,J., Timpf,S., Trajectory data mining: A review of methods and applications. 2016.
- Dodge,S., Weibel,R., Lautenshuetz,A-K, Towards a taxonomy of movement patterns.
- Parent,C. et al., Semantic trajectories modelling and analysis

- COST/MOVE (Knowledge discovery from moving objects) http://www.move-cost.info/index.html