Introduction to trajectory data mining

KIRSI VIRRANTAUS
AALTO UNIVERSITY
GIS-E4020
31.1.2019

Learning outcomes

1. To know the basic terminology in trajectory data mining

2. To get an idea how generic data mining methods are used for trajectory data – according to Aggarwal,C., (2015)

3. To know some cases from geospatial research

4. To get an idea about the topics for future reserach on semantic trajectories

Contents

1. Introductory part
   - Concepts, definitions and sources of trajectory data

2. **Regular data mining methods** in trajectory data mining (Aggarwal,C., 2015)

   -association rules in trajectories – frequent patterns and colocation

   -clustering of trajectories

   -classification of trajectories

3.Some **examples from geospatial research** (Demsar&Virrantaus, 2010; van Kreveld, early 2000)

   -**visual data mining** of trajectories

4.**Semantic trajectory** data mining (Mazimpaka,J., Timpf,S., 2016)

5.**Research topics** (Dodge,S., Weibel,R., Lautenshuetz,A-K; Parent,C. et al)

1. Trajectory data, definition

trajectory(en) = liikerata(su)

trajectory of a moving object is a continuous function $\tau(t)$ of time t such that given a time instant t it returns the position of the moving object

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

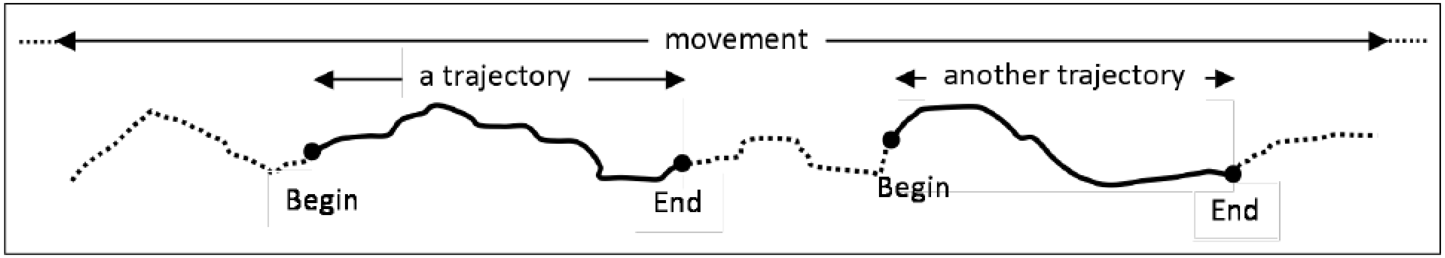


Figure 3 - Trajectories extracted from a movement track visualized as dotted line

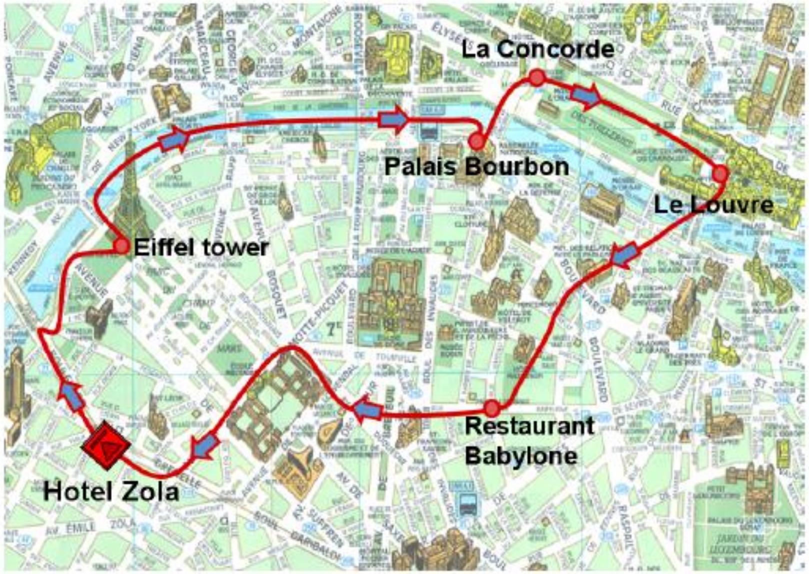


Figure 1 - 2D visualization of a one-day spatial trace left by a tourist visiting Paris – background map downloaded from Mappery.com, copyright unknown

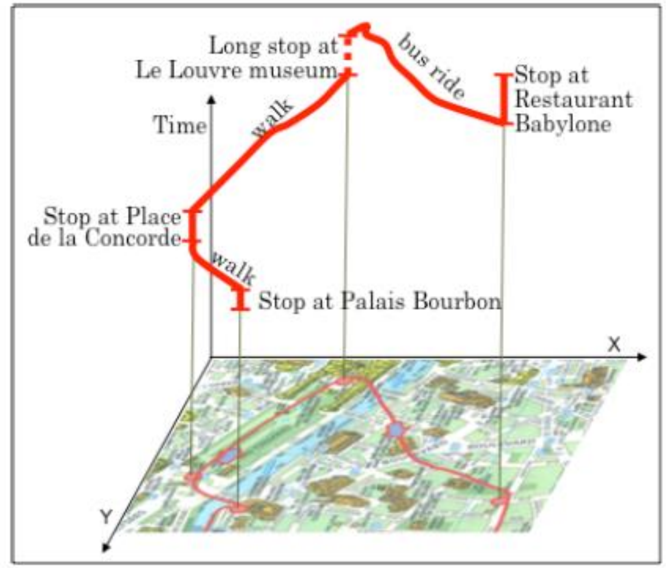


Figure 2 - A time-geography diagram showing part of the previous tourist track

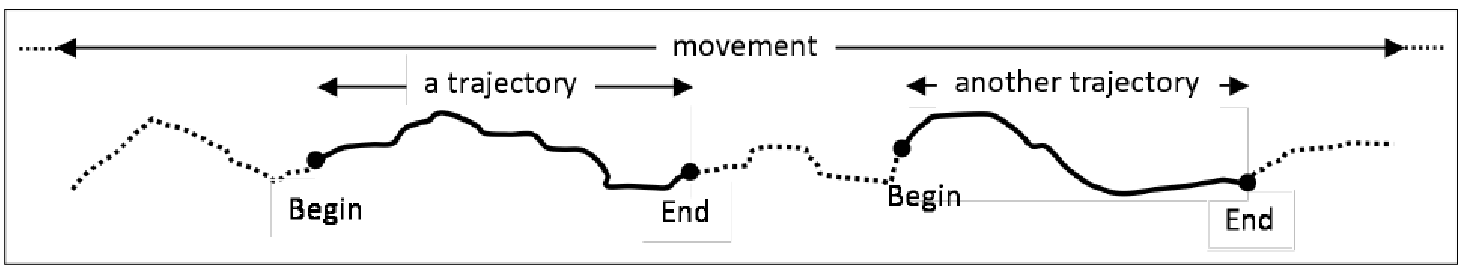


Figure 3 - Trajectories extracted from a movement track visualized as dotted line

Movement track in 2d and in 3d-presentation. **Trajectory** can be seen as the interesting part of the entire movement track.

(Parent et al.)

Terms used with similar meaning

Track (=jälki,su)

A mark or line of marks left by a person, animal or vehicle in passing

Sequence (= järjestys, peräkkäisyys, ketju, su)

A set of related events, movements or items that follow each other in a particular order

Correspondingly:

Trajectory data mining

Sequential data mining

Sources of trajectory data

Depending on the recording system, data is available in different original forms

Mobile data: temporally ordered sequencies of geographical coordinates; the moving object carries GPS tracking device (phone or special tracking device)

Social media location data: content found in Internet social media (content can include coordinates but also place names etc., that data can be **geocoded**); geocoding is a process of transforming physical address into location

Others:

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 trajectory data and applications

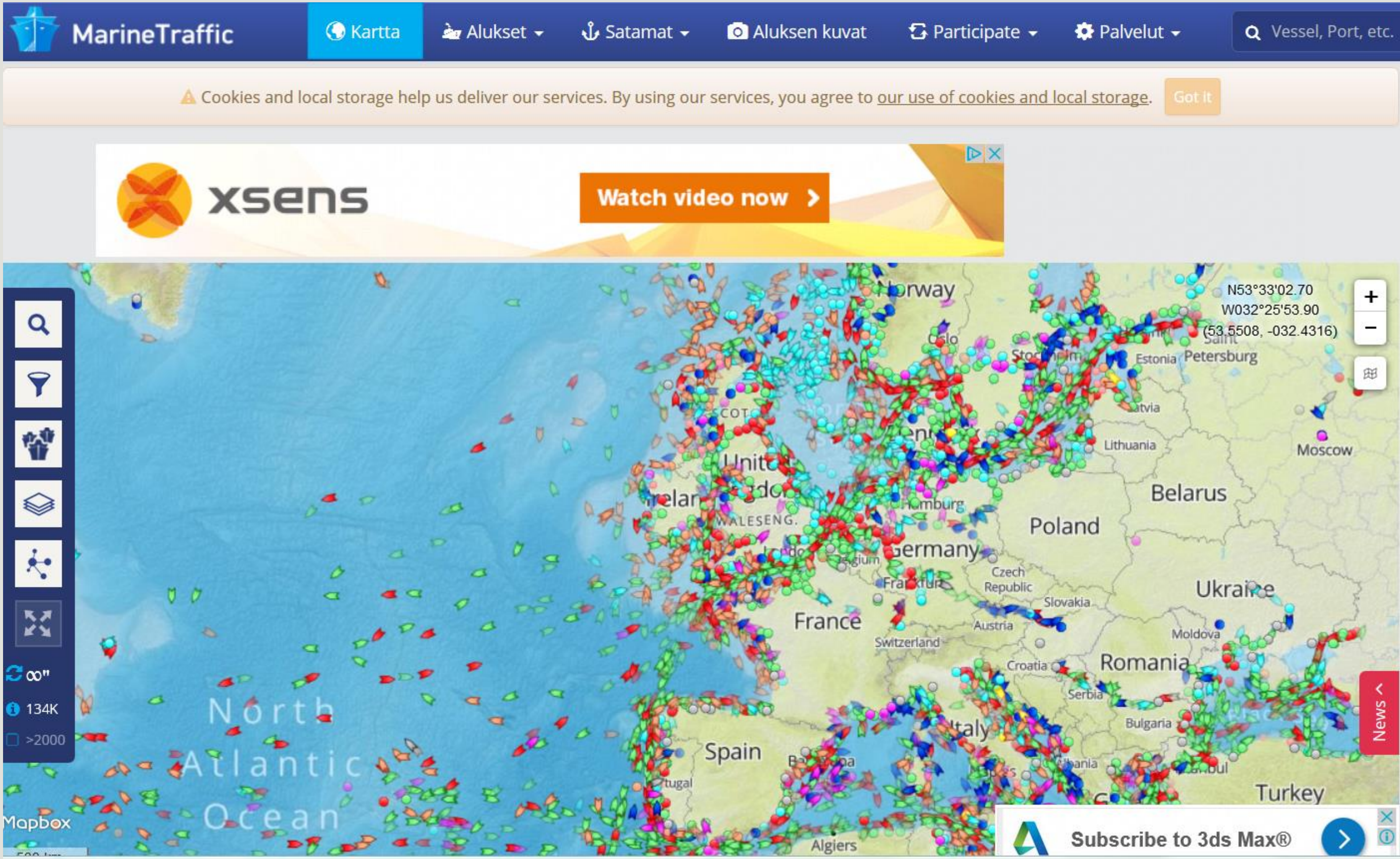
Examples of trajectory data:

- tracked animals (reindeers, moose,birds, ...), people (elderly people), mobile devices (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)

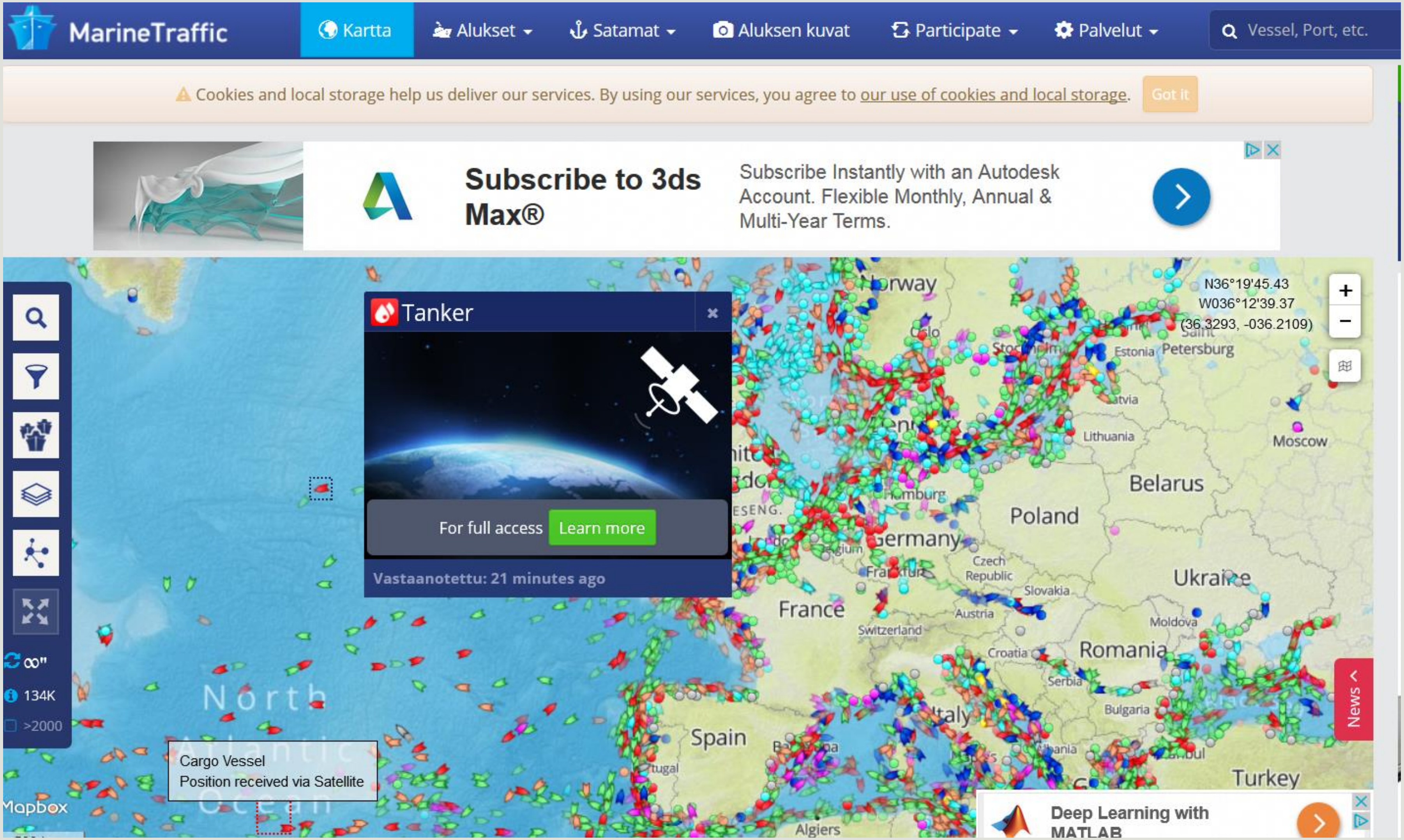
Trajectory data mining has many important real-world applications driven by the real need

- 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



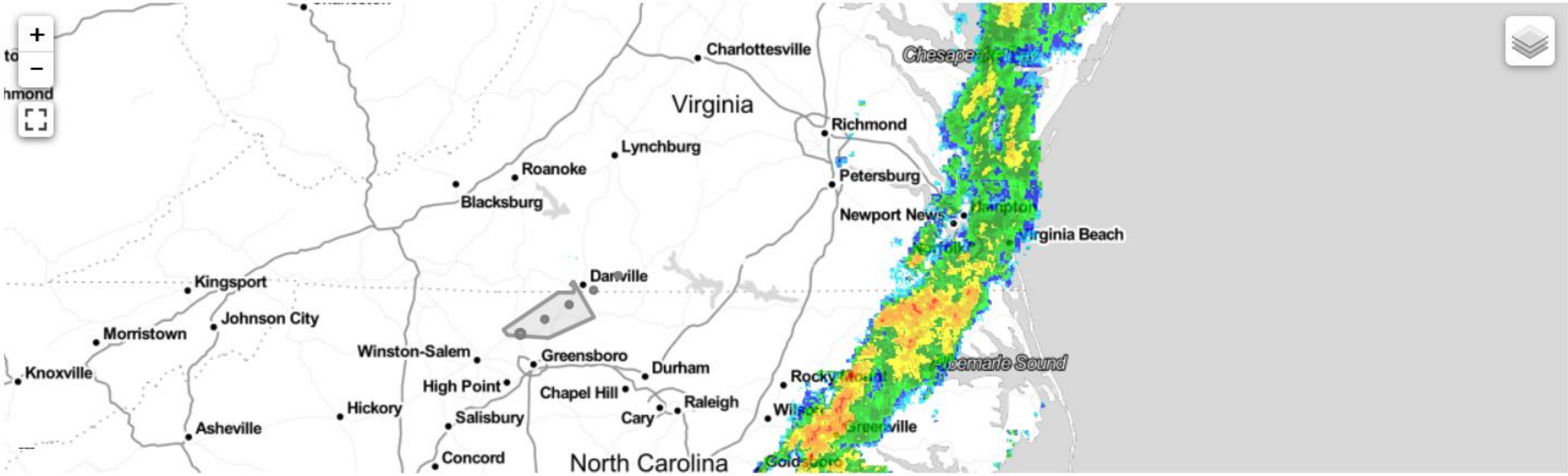
Vessel tracking by AIS (automatic identification system)



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

tornado tracker map

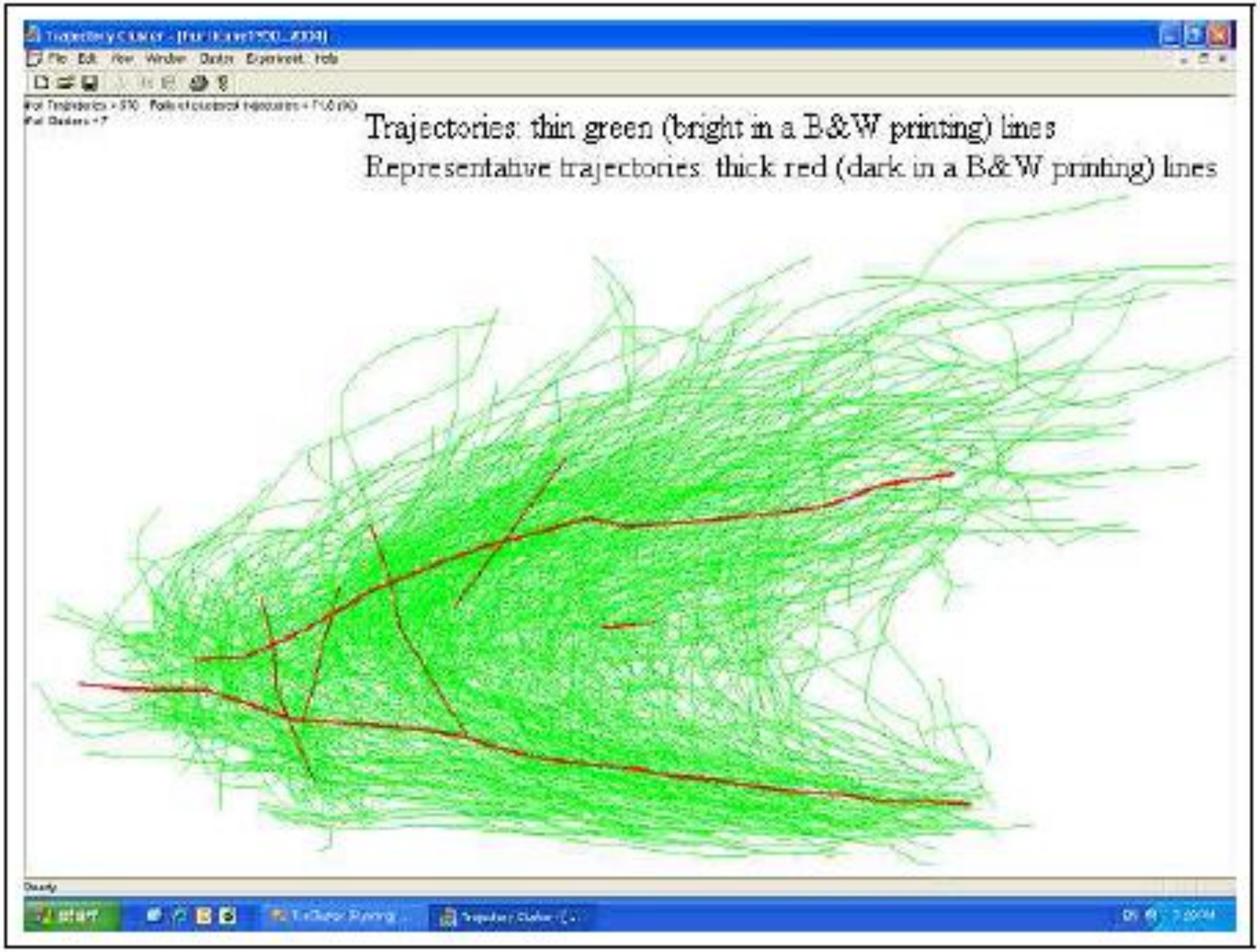
toggle radar



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



(Lee,Han&Whang)

Figure 18: Clustering result for the hurricane data.

Published research material on spatiotemporal trajectories

Trajectory data and moving data management and mining has been in the recent research focus

COST MOVE (Knowledge discovery from moving objects) project has been active

The project has studied:

- Trajectory data management issues
- Trajectory data mining (algorithms, similarity concept, basic forms of trajectories)
- Visual methods and cognition

The project has ended but a lot of good material is available on

http://www.move-cost.info/index.html

2. Regular data mining methods for trajectories

Trajectory data can be mined by directly using similar methods than other spatial or non-spatial data

Regular methods are modified so that they can manage **sequential data.**

- **Association rules** – frequent pattern mining, colocation
- **Clustering** – k-means, k-medioid, DBSCAN
- **Classification** – machine learning methods

Specialities of trajectory data:

Trajectories are **sequential** data; order between discrete steps

Trajectories are **spatiotemporal data,** a special form of **time series** data

Trajectories are a form of **multivariate data**, even **semantic trajectories**

2.1 Association rules for trajectory data mining

Association rules try to find **frequent itemsets**; for example the shopping basket example

Apriori algorithm is one popular algorithm used

In trajectory data mining we are looking for patterns inside the trajectories, **subtrajectories that occur often**

Apriori algorithm- type problem solving can be used

Not looking for frequent itemsets but **frequent patterns or colocations of trajectories**

2.1.1 Frequent patterns - subtrajectories

Sequential pattern mining is a non-spatial data mining method that can be applied to sequence data bases

- Sequence database consists of **ordered elements** or events
- Example can be consumer buying things, location visite by bus; (in these examples location is not of interest)
- The essential thing is the **sequence**; for example ABC is not the same as ACB, but ABC can be the same as AEBDC
- Simple algorithm for sequential pattern mining is Apriori –algorithm based mining methods

Sequential pattern mining can be applied to spatio-temporal data = trajectory data sets

For finding **frequent patterns in trajectories**

Example on sequence database compared to traditional transaction db

A *transaction database*

TID	itemsets
10	a, b, d
20	a, c, d
30	a, d, e
40	b, e, f

A *sequence database*

SID	sequences
10	<a(abc)(ac)d(cf)>
20	<(ad)c(bc)(ae)>
30	<(ef)(ab)(df)cb>
40	<eg(af)cbc>

In regular transaction database each itemset is made of single item
In sequence database sequences are made of subsets of items that occur in a specified order

Example: a consumer first visits shop1 and buys a, then he goes to shop2 and buys a,b,c; then he continues to shop 3 and buys a,c … and so on

Frequent sequential path mining

Determining **frequent sequential paths** in trajectory data can be seen as the simplest pattern mining method (Aggarwal., 2015)

- the trajectory data is converted into a 1-dimensional discrete sequence, after that any sequential pattern mining algorithm can be applied
- **Grid based discretization** –method is used
- Association rule –based mining algorithms can be applied; however the sequence is an additional constraint

Grid based discretization

1. convert multidimensional trajectory to 1-dimensional discrete sequence
   - For example by using **grid based discretization** (in geoinformatics we call this vector-to-raster –conversion)
   - Trajectory becomes a sequence of grid cells identified by matrix elements
   - Called as **spatial tile transformation**
   - The time stamp can be added to the sequence (spatiotemporal tile transformation)

2. apply any sequential pattern mining method

Drawback of this approach is the lost accuracy (granularity level)

Exercise

Draw a trajectory

Exercise discretization

- Create a grid of suitable resolution
- create the trajectory into a **sequence database** by using the grid element ids

Now it is easy to understand intuitively how Apriori algorithm is applied to find frequent patterns (subtrajectories, paths)

2.1.2 Colocation patterns

The goal is to **discover "social connections" between trajectories** of different individuals

Individuals sometimes frequently appear in the same places at same time – they might be somehow related to each other

1. Grid discretization: for each grid cell and time interval pair a list of identifiers (individuals or trajectories) is determined

2. The previous transformation causes unordered sets of ids for each cells in which at least two individuals have visited
   - This produces a vertical representation of the sequence dataset

3. Any frequent pattern mining method can be now applied; the result of frequent patterns correspond now to the frequent sets of **colocated individuals**

Exercise

Draw two (or more) trajectories

Create a grid of suitable resolution

Create sequence database according to the previous introduction

This simple exercise helps you to understand **how trajectories can be transformed into transactions in space and time**

The task will become more complicated when the trajectories are semantical, they have descriptive attributes

2.2 Trajectory clustering

There are two types of trajectory clustering methods:

1. Methods based on convential clustering methods
    - Clustering is based on **similarity of trajectories** (raw trajectories)
    - When the **distance function** has been defined any clustering method can be applied

2. Methods based on discretized trajectories and **pattern mining** methods applied

Also lots of so-called **ad hoc methods** in which the **characteristics** of the trajectories (semantic trajectories)

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

Similarity of (raw) trajectories

Clustering algorithms are based on distance measure

Distance between trajectories can be simple but also very complicated task

In the simplest case the distance can be managed by creating buffers around trajectories and intersecting the buffers; in this case we have only raw trajectory and we do not take into account time; clusters are made at constant distance (buffer, Kernel)

Fig1 (on the next slide) shows simple case when trajectories have same shape and same starting and end point

Fig 2 shows a more complicated situation

This goes towards very complex variations when space time and attributes are considered.

(Buchin et al.)

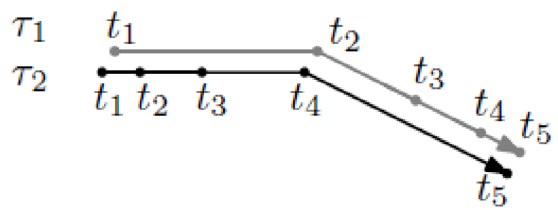


Figure 1: Trajectories with similar shape but different speeds at corresponding times.

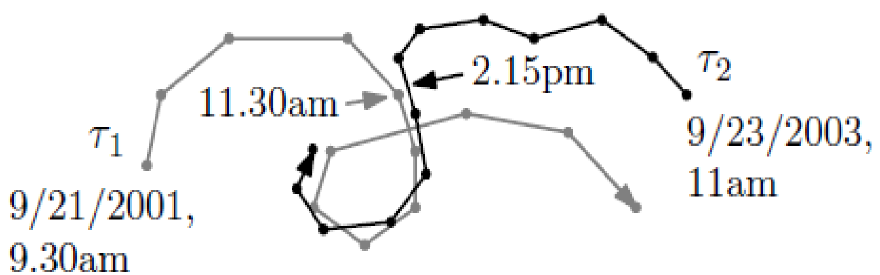


Figure 3: The middle part of trajectory τ_1 is similar to the last part of trajectory τ_2.

Trajectory clustering methods

When the trajectory **similarity calculation method** has been decided any regular clustering method can be used

- Hierarchical clustering
- Partitional clustering – Kmeans, Kmedoid

Special problem with trajectory clustering is: how to define **the centroid** (medoid, center point) of the clusters

In agglomerative clustering “close enough” trajectories are merged, some criteria must be decided also for that

Example: TRACLUS algorithm, clustering by DBSCAN, similarity calculated between subtrajectories

2.3 Trajectory classification

In classification the objective is to find a rule to assign objects to **pre-defined classes**

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

Similar types of methods as for clustering are used

- Distance based methods
- Sequence based methods

Trajectories are often semantic that makes

a new aspect to the algorithms

Example on classification of vessel trajectories.

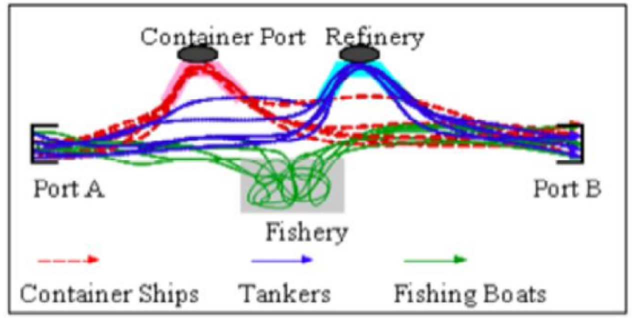


Figure 7 - Example of trajectory classification [Lee et al. 2008]

2.4 Outlier trajectories

Outliers are trajectories that differ significantly from the other data

Outliers are complementary concept to that of clusters; when clustering tries to determine groups of data, outliers are individual data that are different from the remaining data

Other terms for outliers: **abnormalities, discordants, anomalies**

In data mining scenarios outliers have numerous applications:

- Data cleaning – when outliers are noise (errors)
- Identifying unusual patterns – when outliers are correct data but show some unusual/unwanted pattern

Both distance based and sequence based methods can be used

3. Examples, cases

First example: A piece of research made by Urska Demsar and K.Virrantaus starting in 2010…

- Analysis of AIS data by using **visual clustering for pattern finding outlier detection and**

and example: Marc van Kreveld

- one of the earliest developers of **trajectory data pattern mining methods**; University of Utrecht,NL
- visited our group in 2006

Urska Demsar

Overview | Research publications | Research data | Activities and awards | Funded projects

Urska Demsar

School of Geography & Sustainable Development - Lecturer in GeoInformatics

Postal address:
School of Geog & Sustainable Devt
Irvine Building
St Andrews

United Kingdom

E-mail: ud2@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.

Marc van Kreveld

Professor; computational geometry and its application

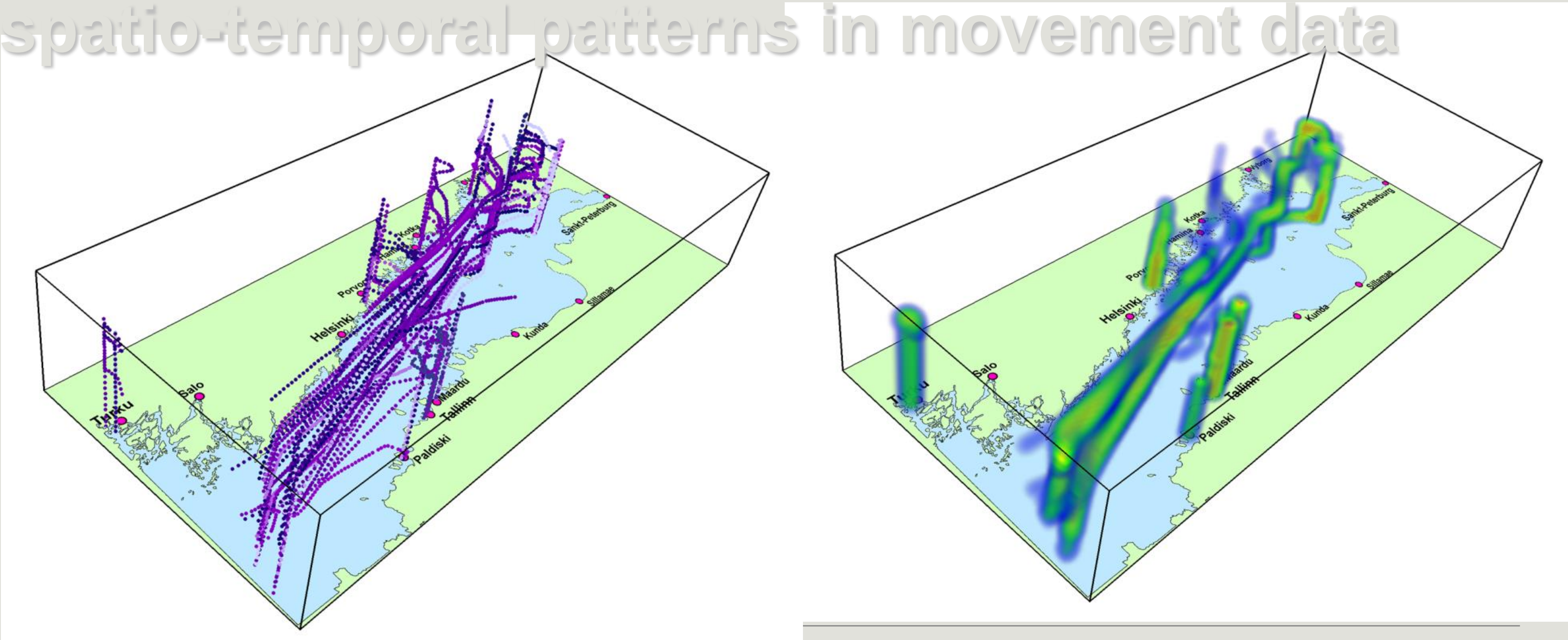
Division Virtual Worlds
Department of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands

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.

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

NCG
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)

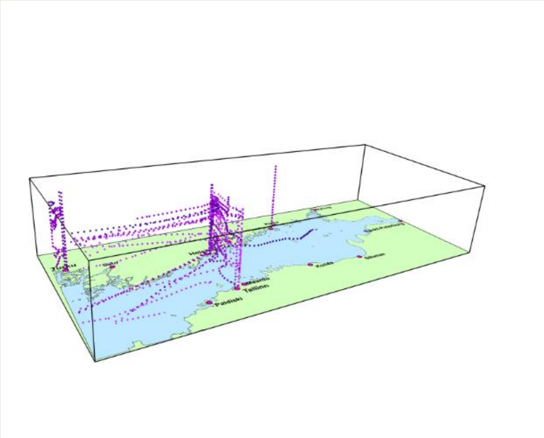
-one month AIS data (January 2008)

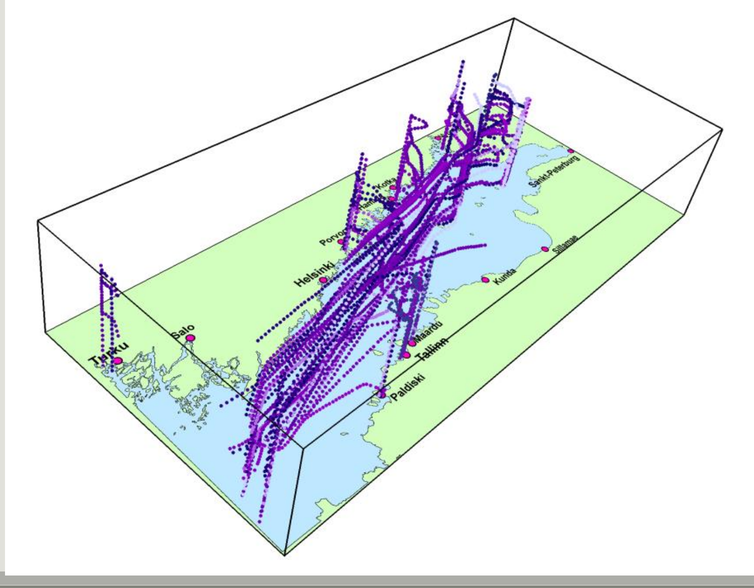
AIS = Automatic Identification System based on GPS positioning

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

-passenger ships (ferries)

-tankers





-trajectories

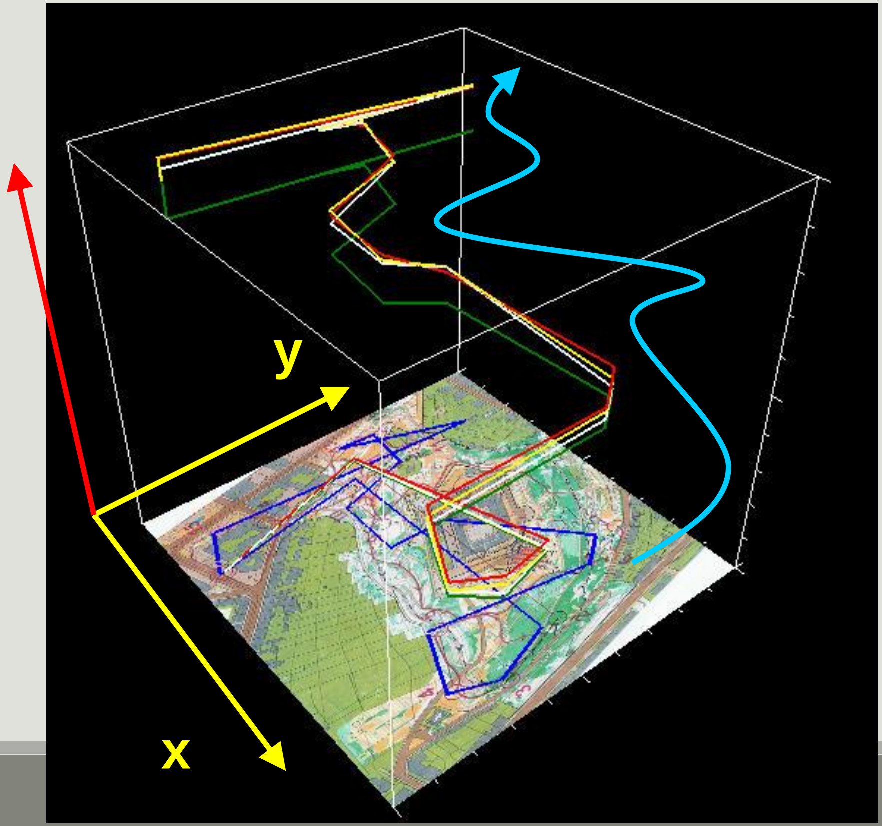
Method:
Time in 3D – Space-Time Cube and its variations

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

z axis – time t

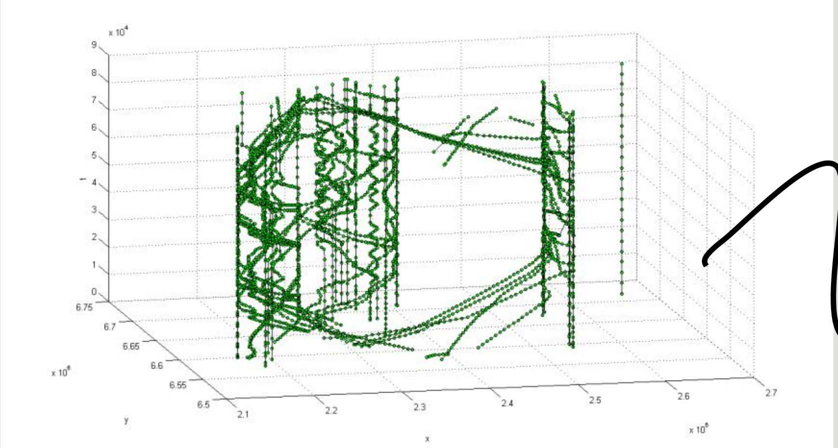
t



trajectories

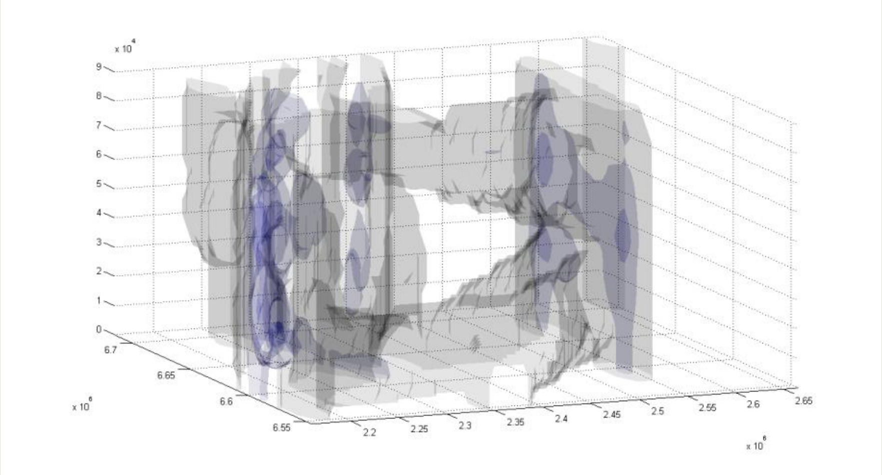
x-y plane = geographic map

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



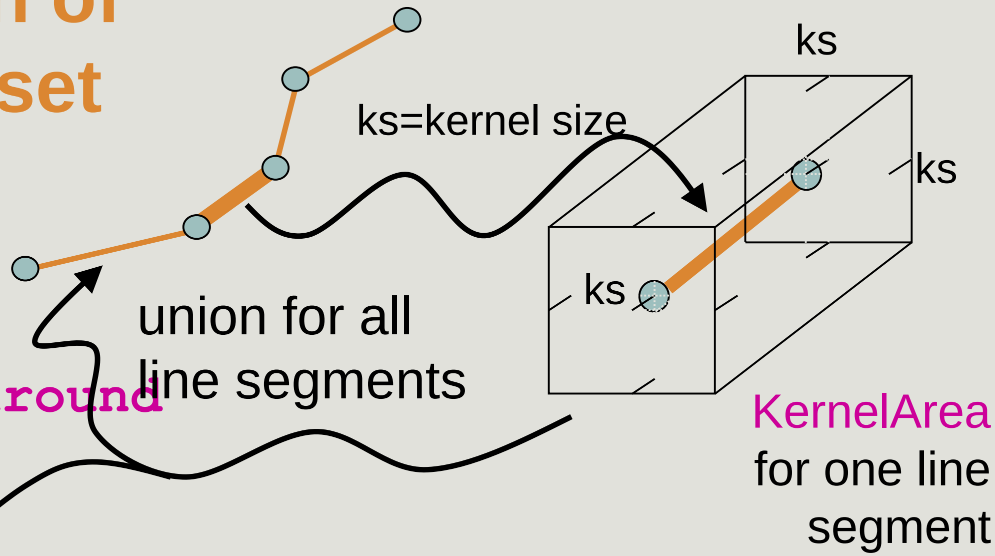
Solution:
Space-time density of trajectories

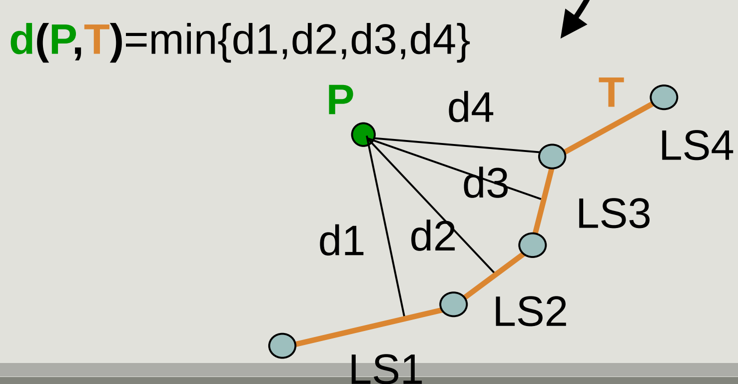
Volume visualisation of importance of each trajectory in space and time

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

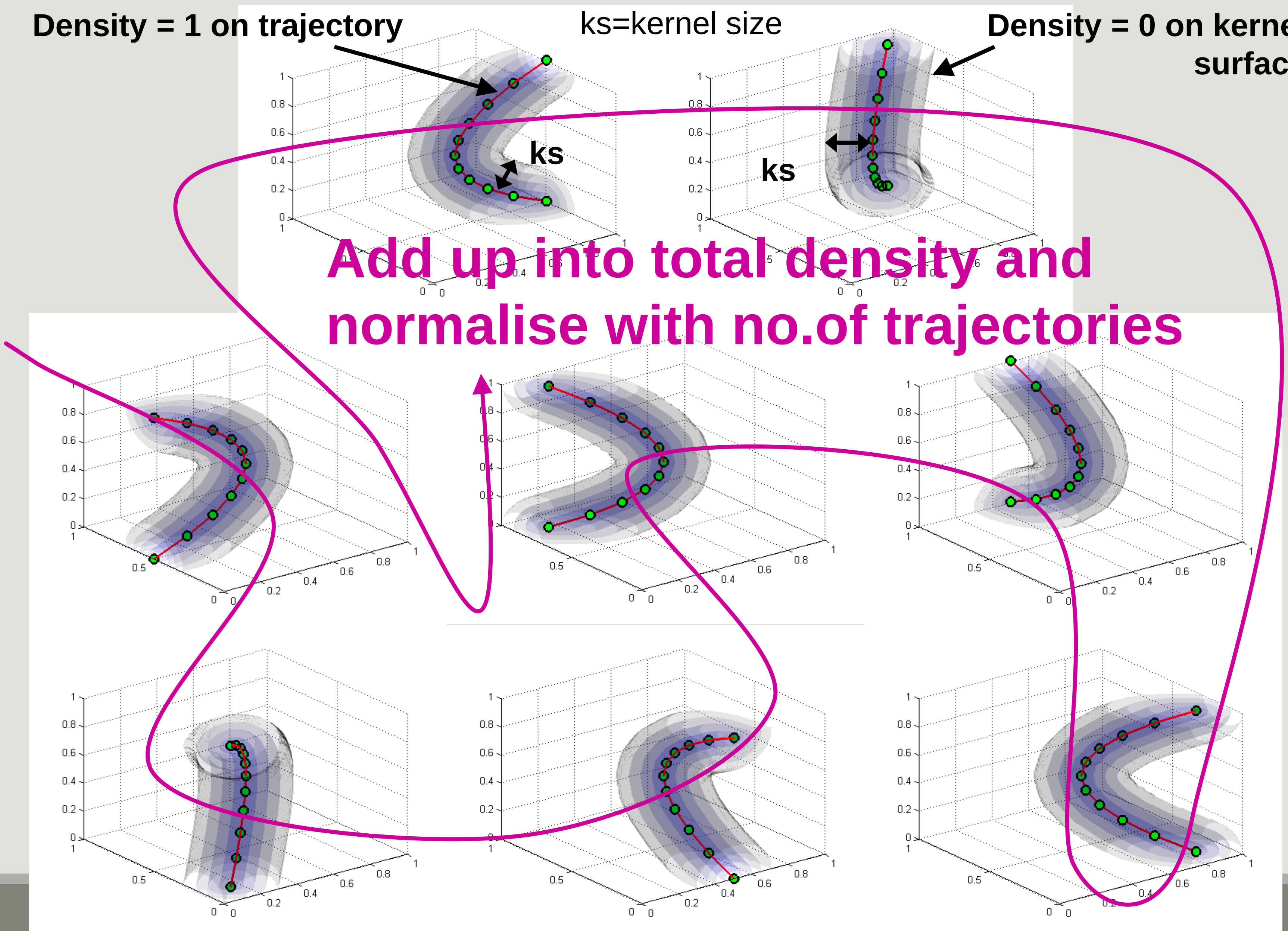
```
    TotalDensity = 0;
    for each trajectory T
         TrajectoryDensity = 0;
         calculate KernelArea around
the trajectory;
    for each voxel P in the
    KernelArea
              calculate
DistanceToTrajectory T;
         TrajectoryDensity = normalised
DistanceToTrajectory
         with kernel size;
         end

         TotalDensity = TotalDensity +
                  TrajectoryDensity;
    end
    Normalise TotalDensity with
number of trajectories;
```





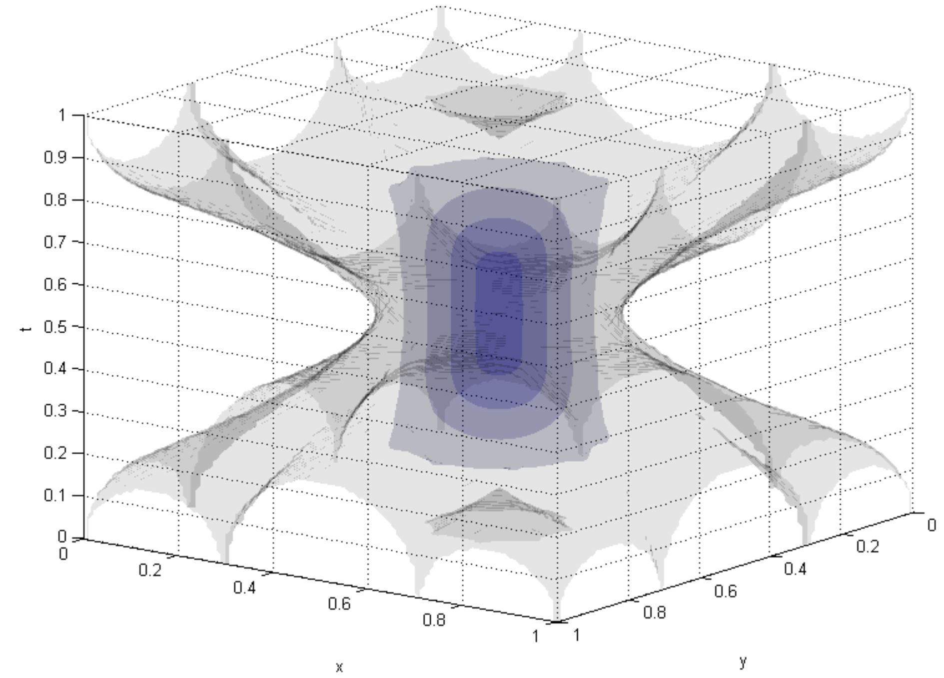
Densities around each of eight simulated trajectories

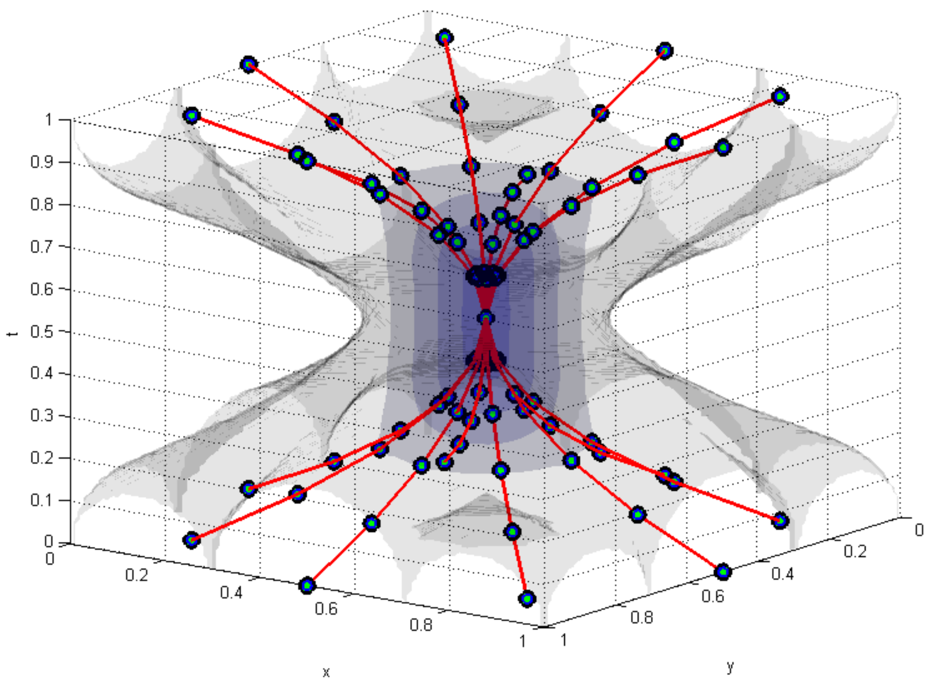


Total density of eight simulated trajectories

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

Same colour for all isosurfaces,
different transparency.





Density +
superimposed
trajectories

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

Two data sets:

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

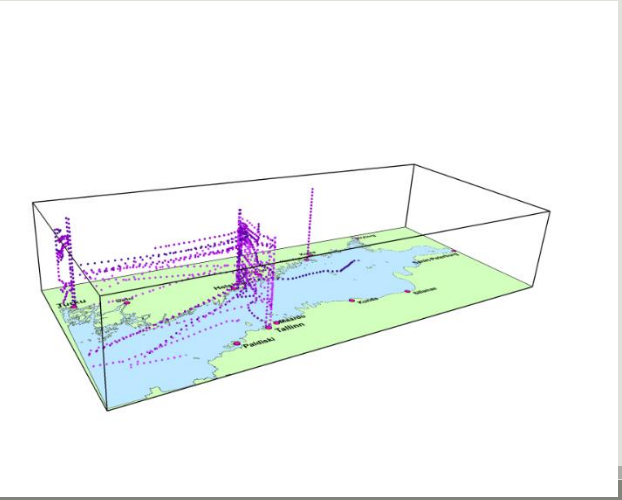
-one month AIS data (January 2008)

AIS = Automatic Identification System based on GPS positioning

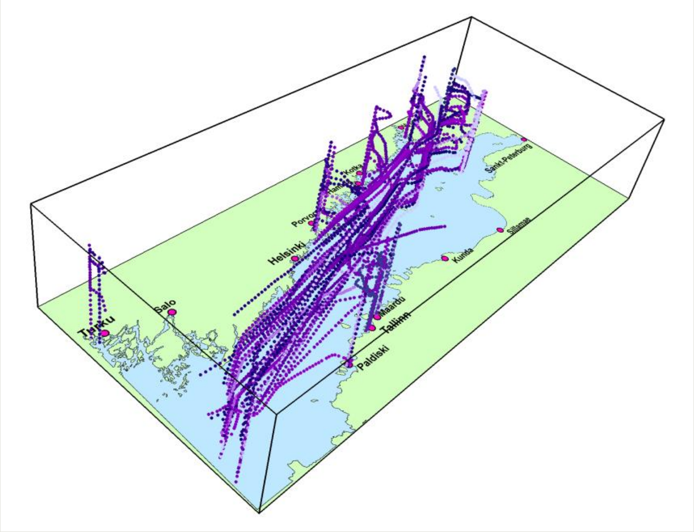
Densities for 2 vessel types were produced by using trajectory data of:

-passenger ships (ferries)

-tankers



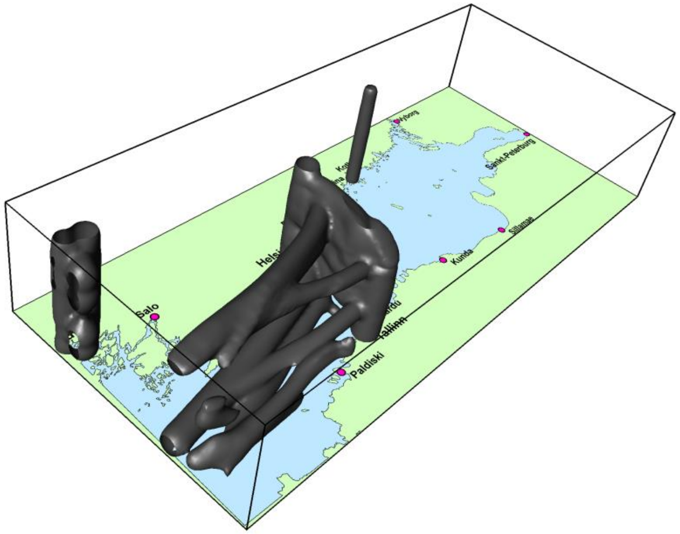
-trajectories



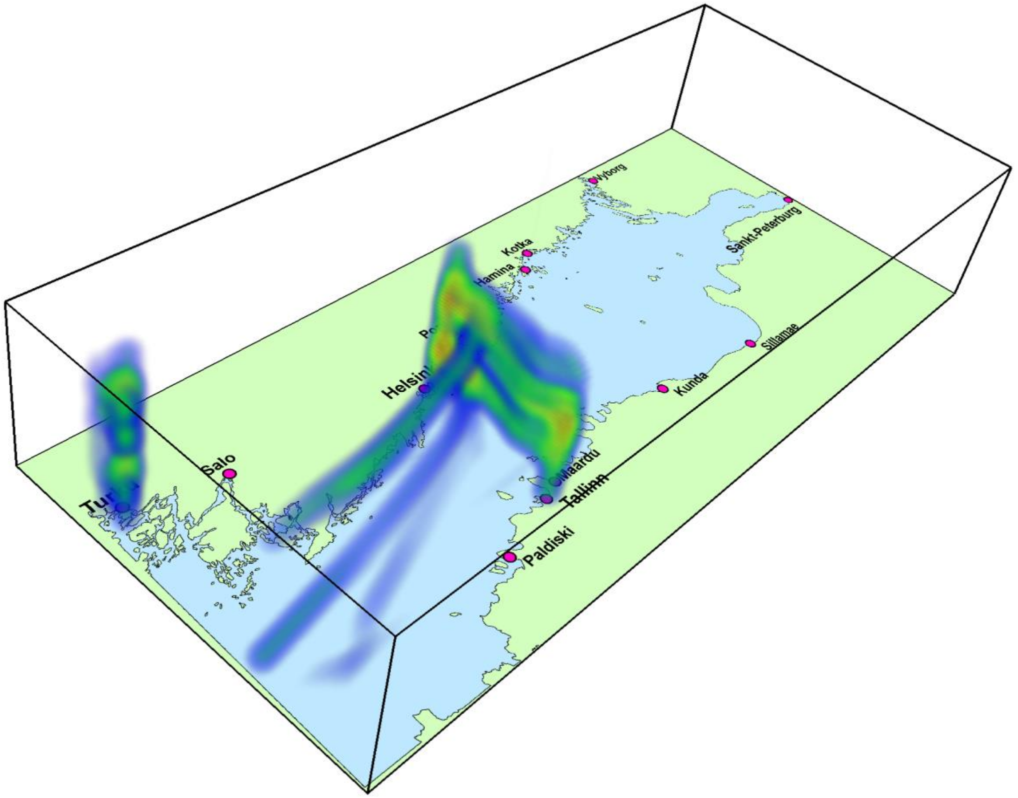
Visualisation possibilities and spatio-temporal patterns

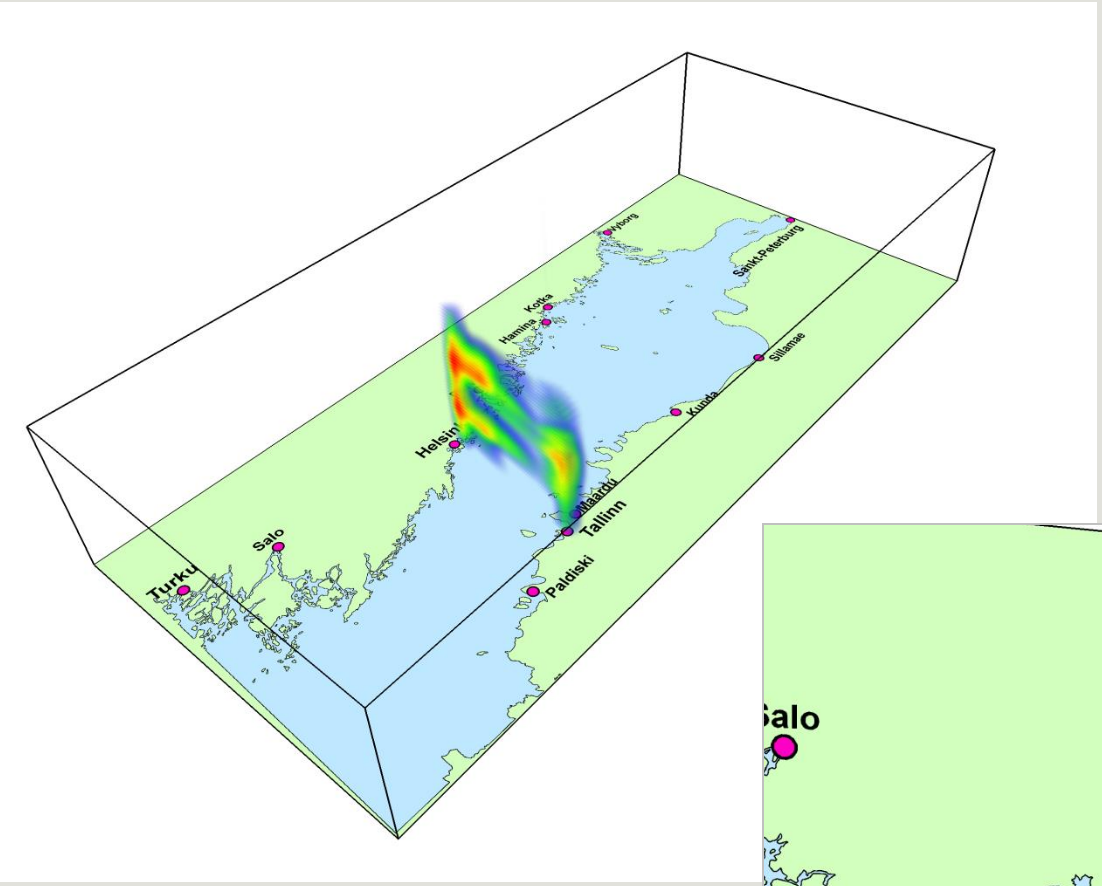
Same three possibilities as for simulated data

Isosurfaces

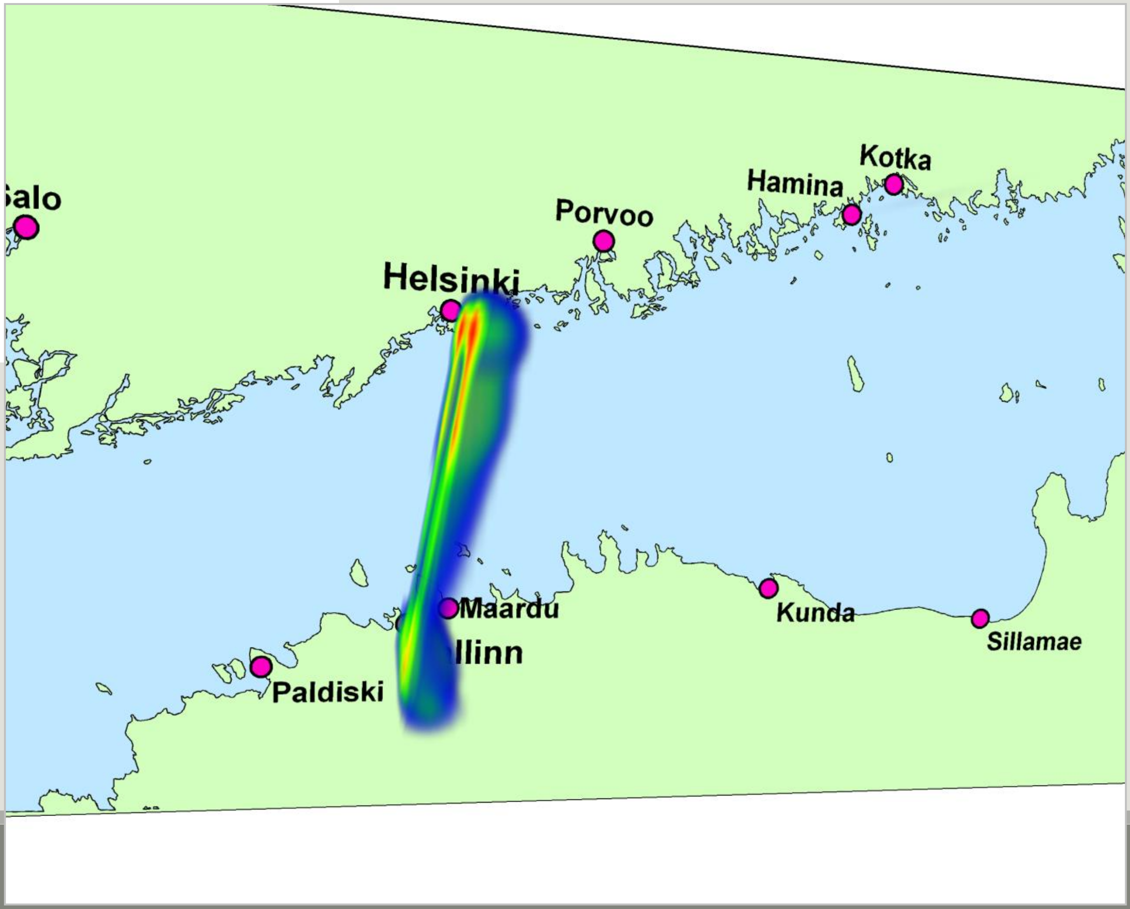


Direct volume rendering

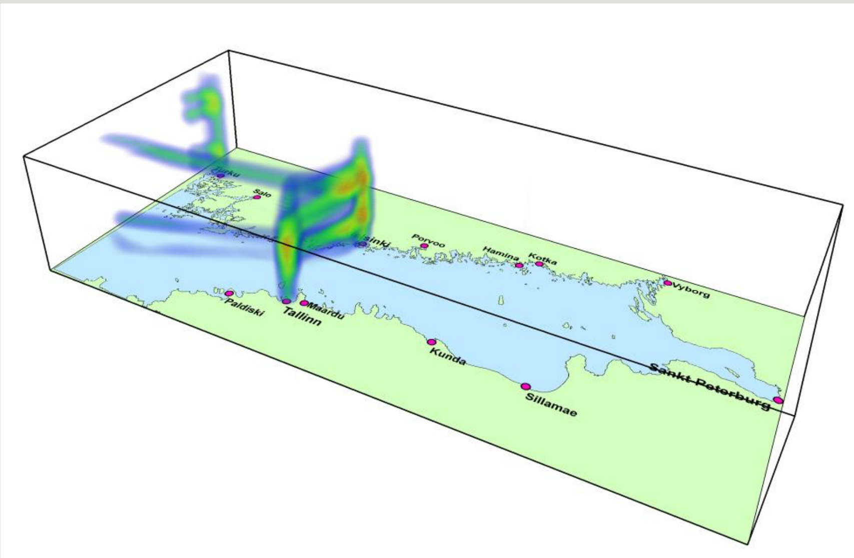


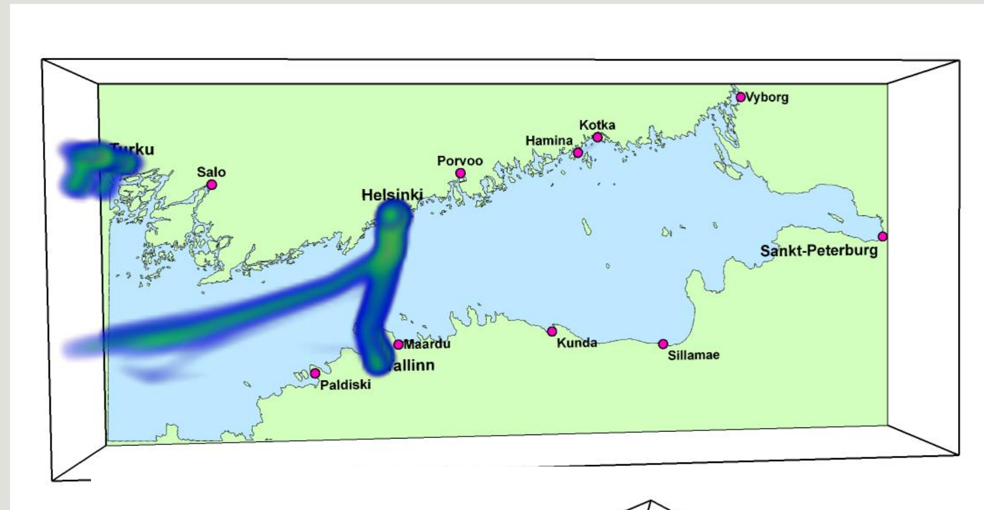


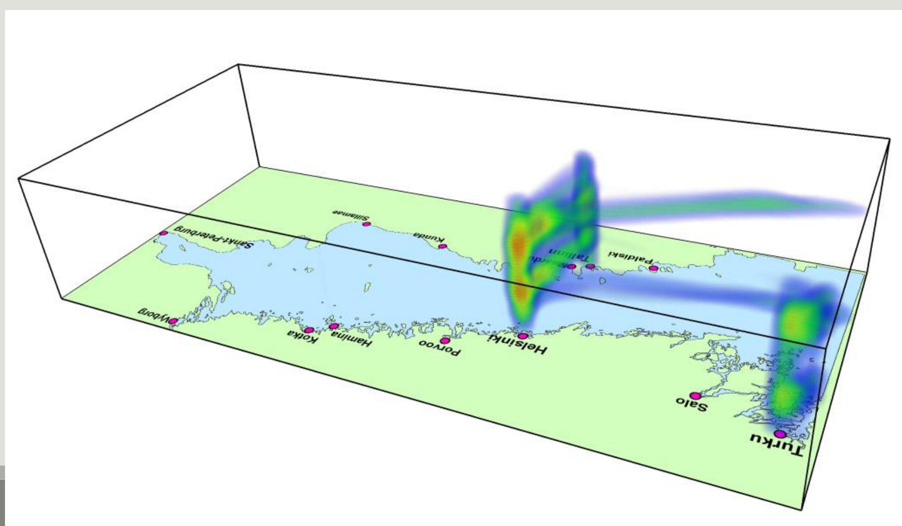
Clipping planes

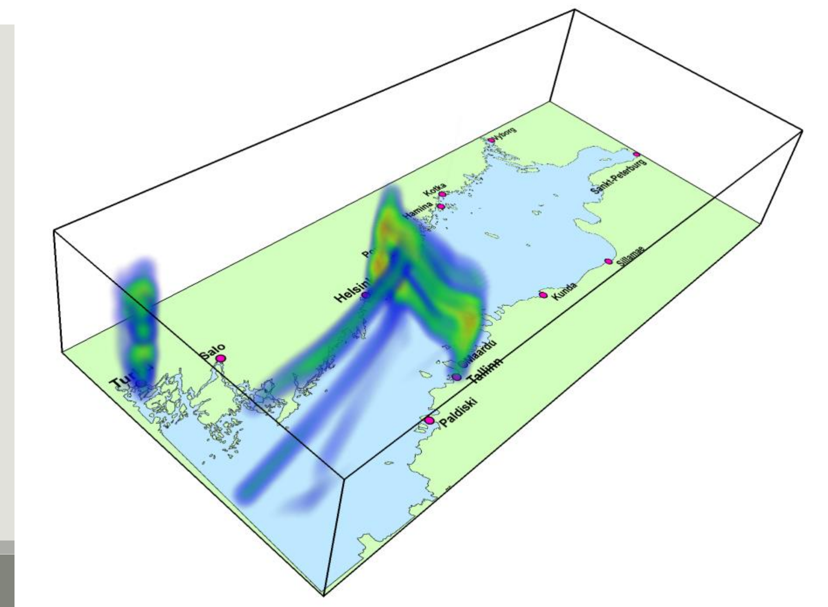


Results – passenger ships – monthly density

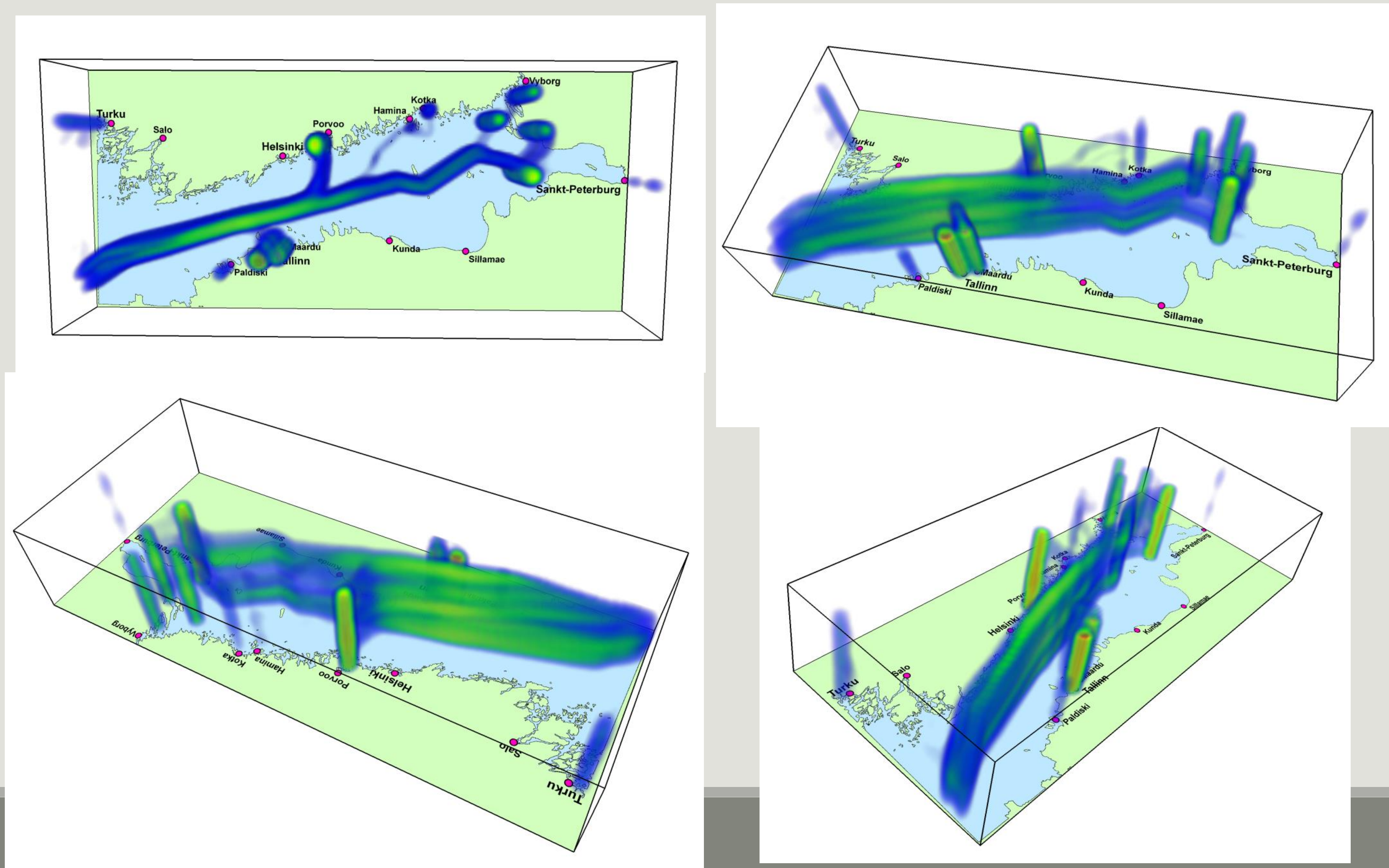




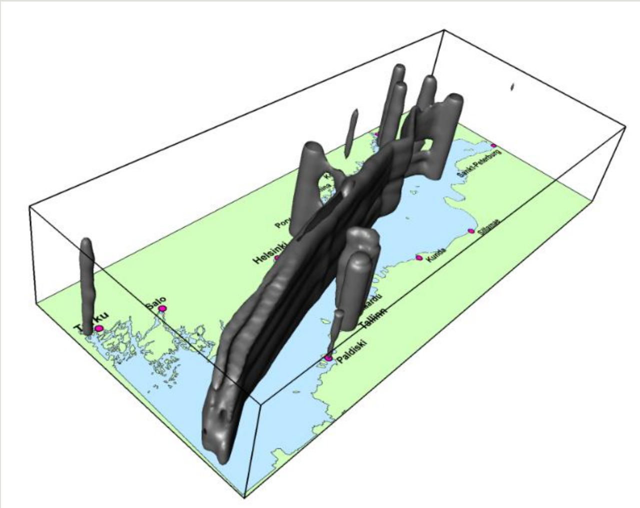


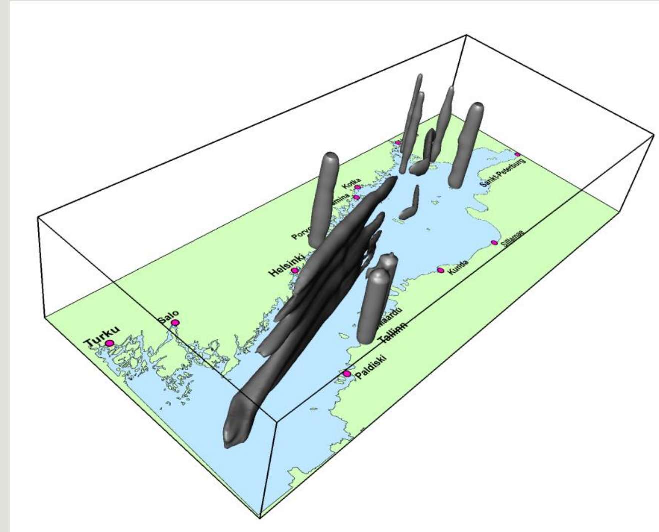


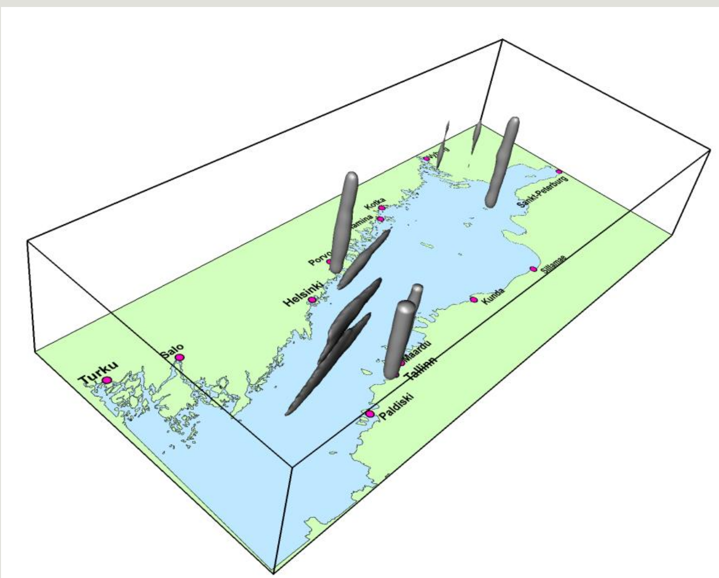
Results – tankers – monthly density – DVR

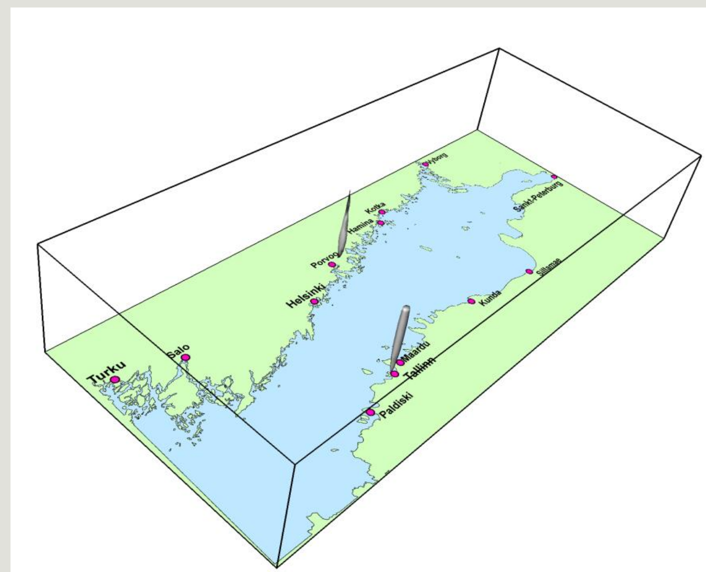


Results – tankers – monthly density - isosurfaces

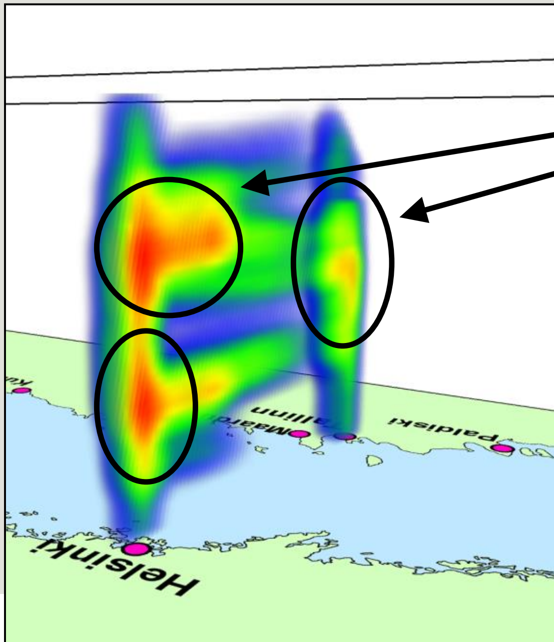






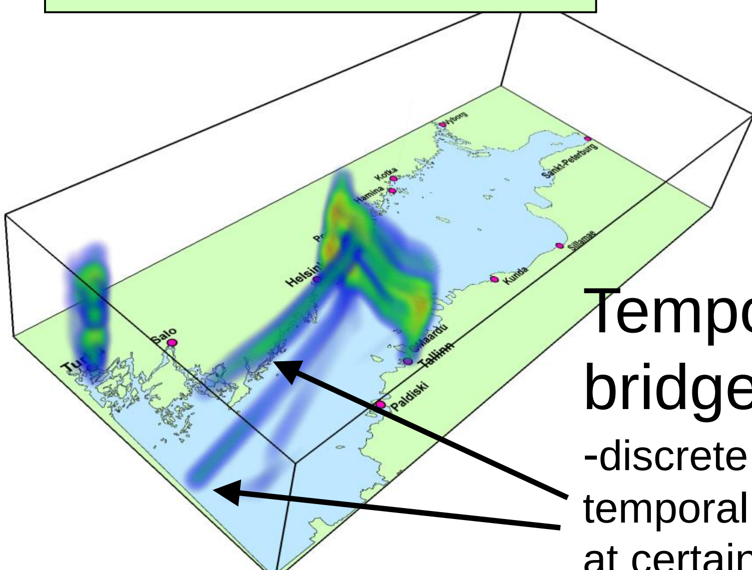


Spatio-temporal patterns to look for



Spatio-temporal **hotspots**

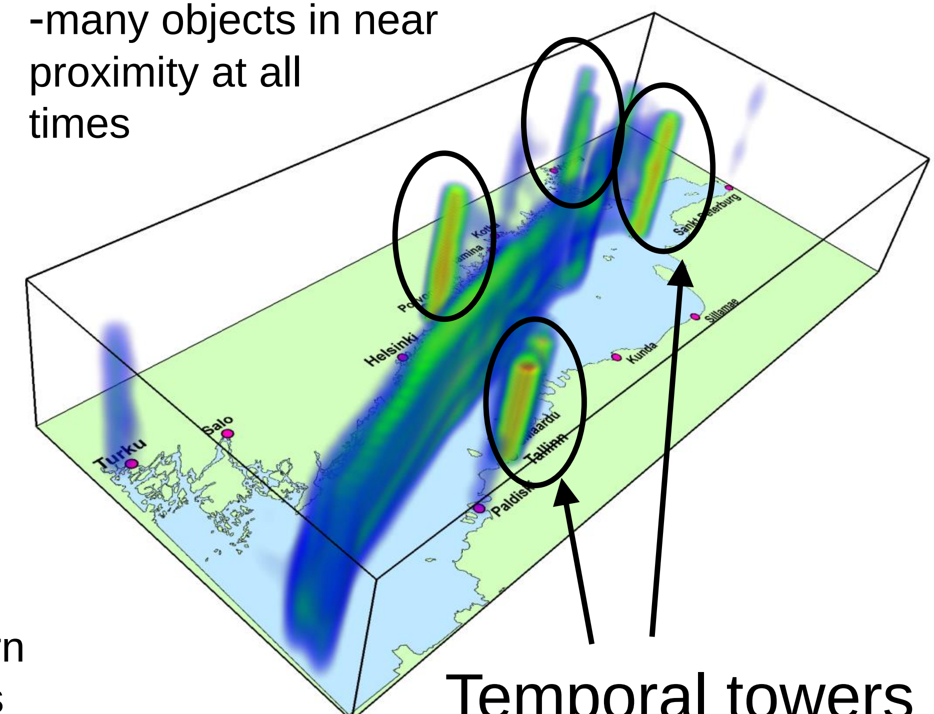
-convergence of trajectories in both space and time



Temporal bridges

-discrete temporal pattern at certain times

-many objects in near proximity at all times



Temporal towers

Early examples from Marc van Kreveld – algorithms

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

Marc van Kreveld was among the first researchers inside geospatial field who developed algorithms and theory for trajectory data mining

He developed simple algorithms. On the following pages some examples.

He also was the first (?) to try to create some taxonomy of geospatial trajectories.

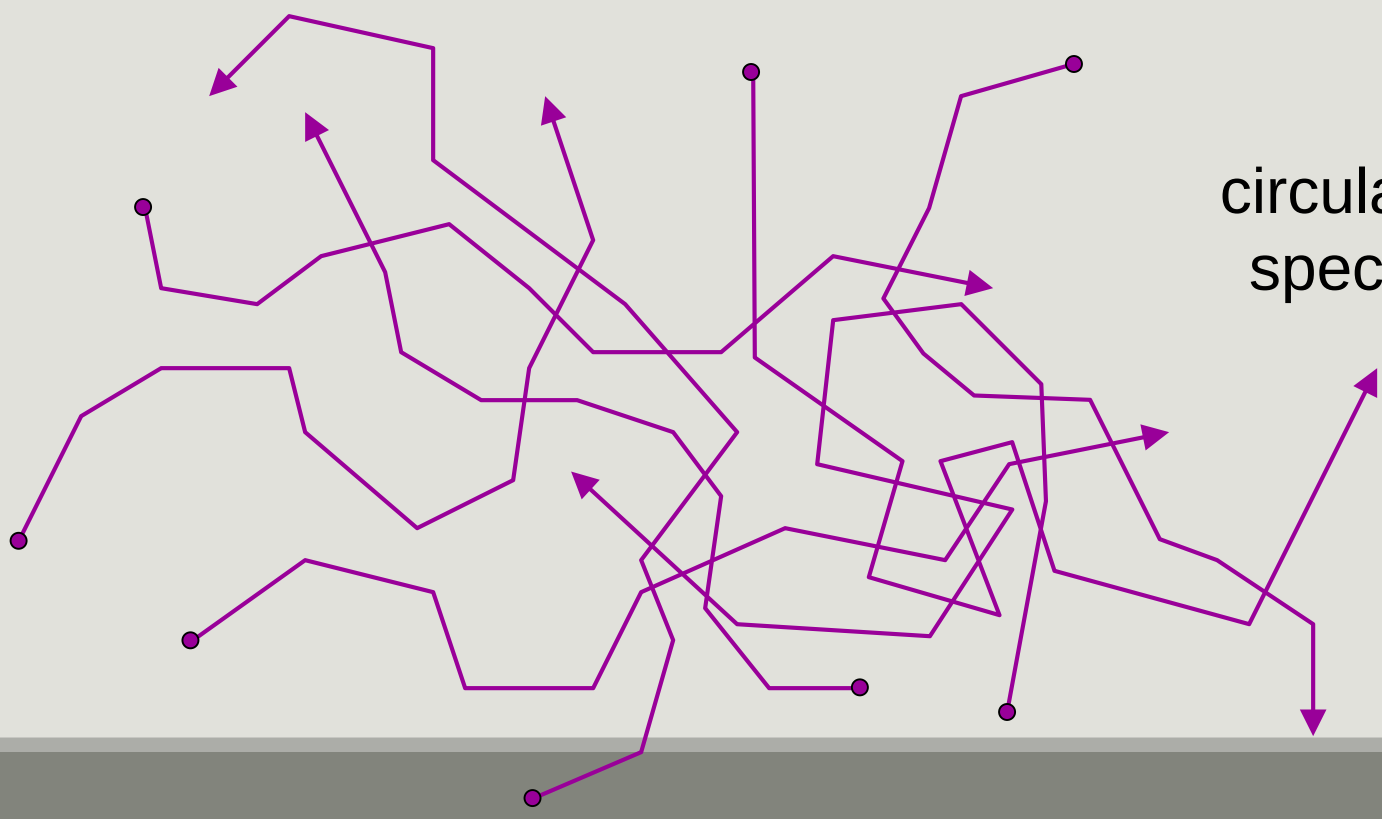
Exercise question

Look at the following slide:

Where is the most visited location?

1) Example pattern in trajectories (MK)

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

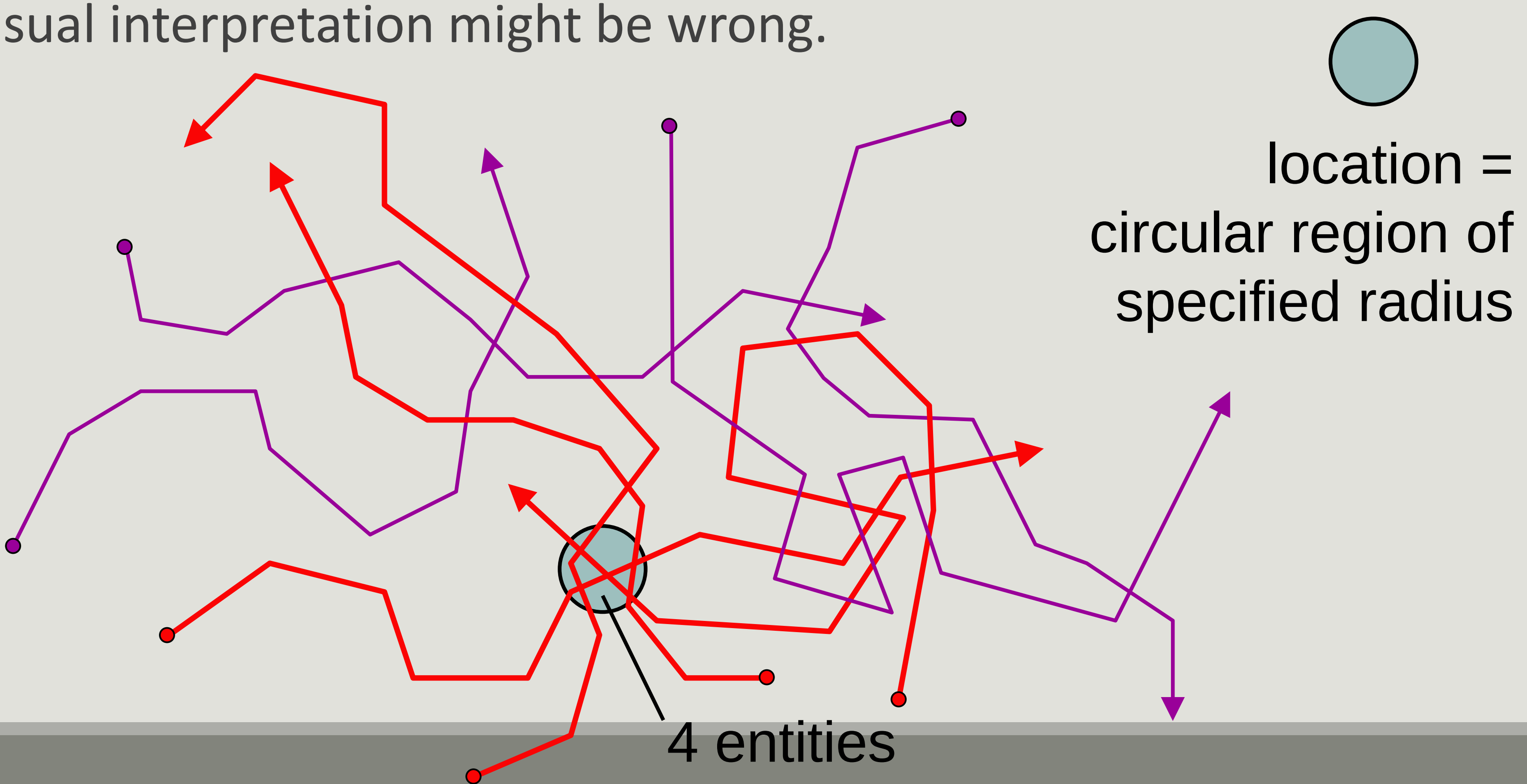


location = circular region of specified radius

Example pattern in trajectories (MK)

What is the location visited by most entities?

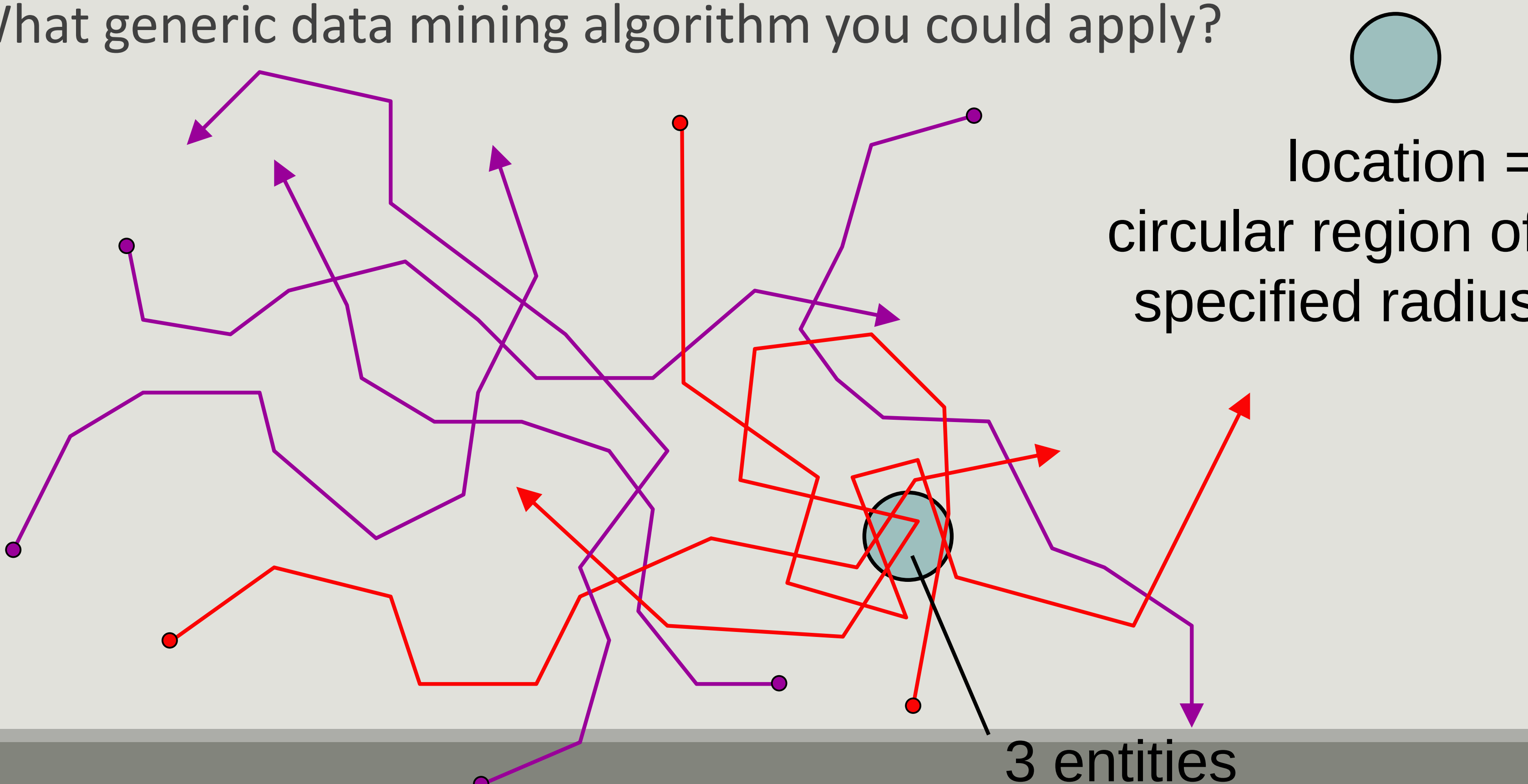
Visual interpretation might be wrong.



Example pattern in trajectories (MK)

What is the location visited by most entities?

What generic data mining algorithm you could apply?



2) Trajectory pattern types (by 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

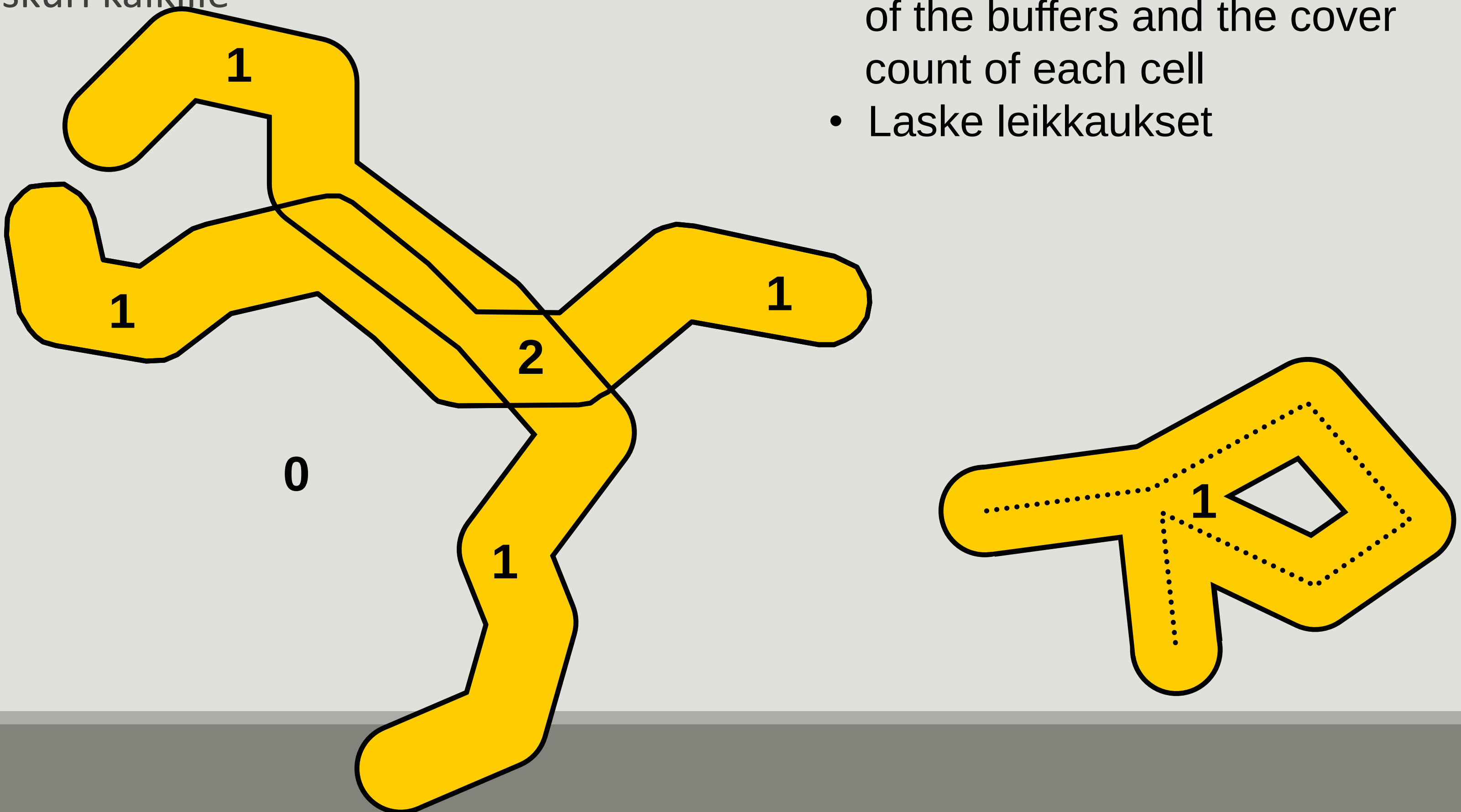
Compute buffer of each trajectory

Lasketaan puskureita käyttämällä

Example: Finding a flock

Compute buffer of each trajectory

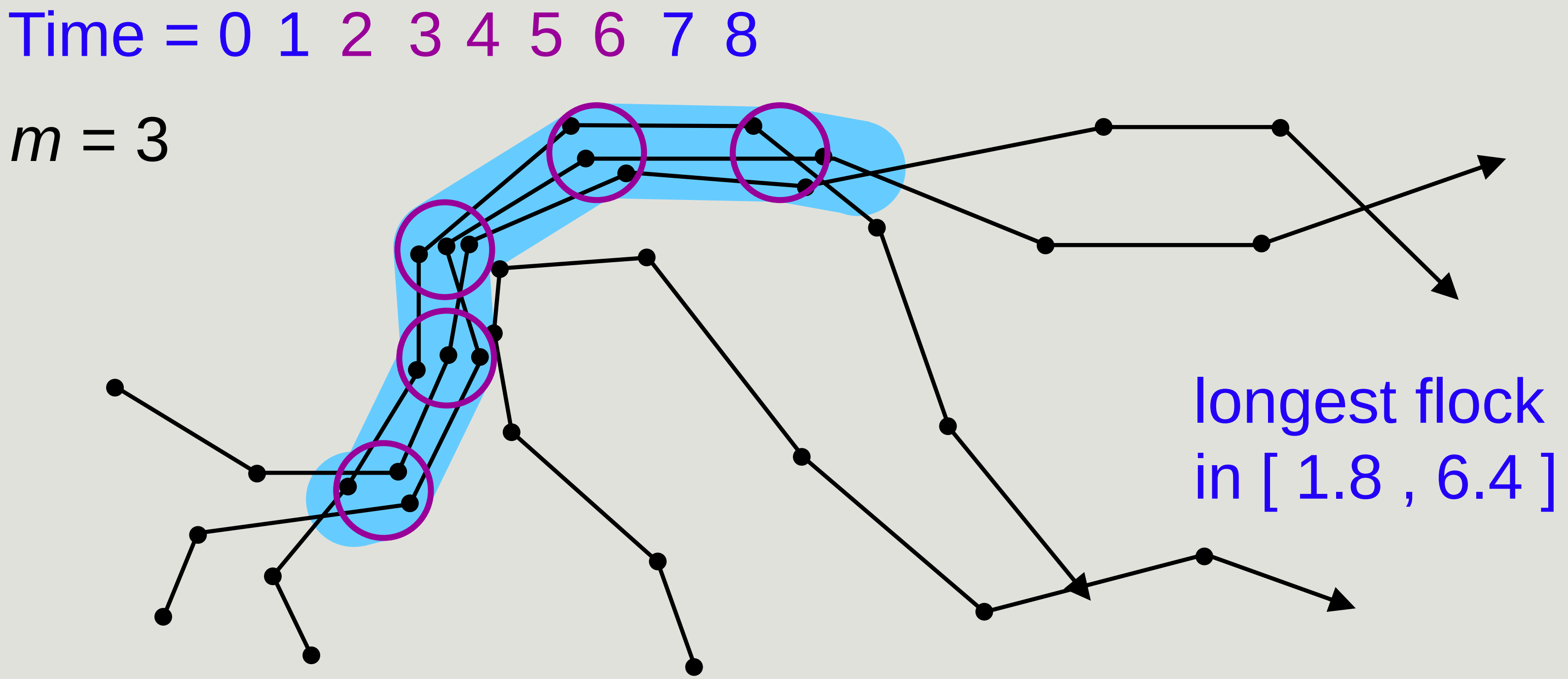
Laske puskuri kaikille



- 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*)



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

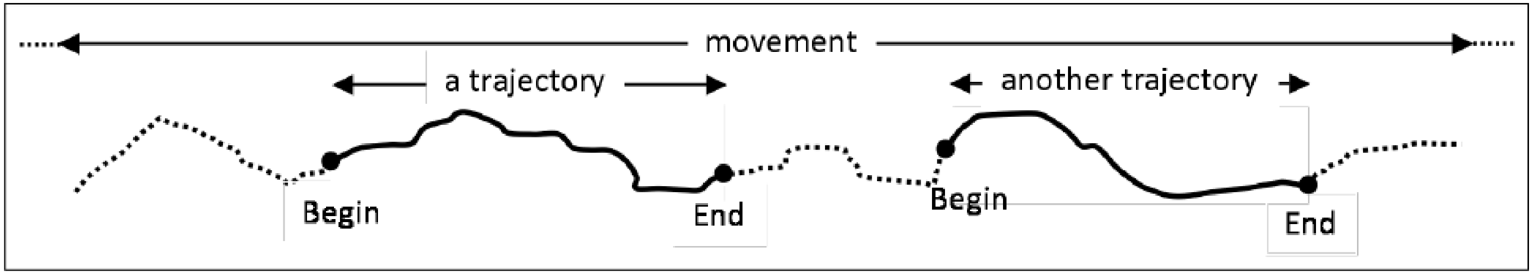


Figure 3 - Trajectories extracted from a movement track visualized as dotted line

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). □

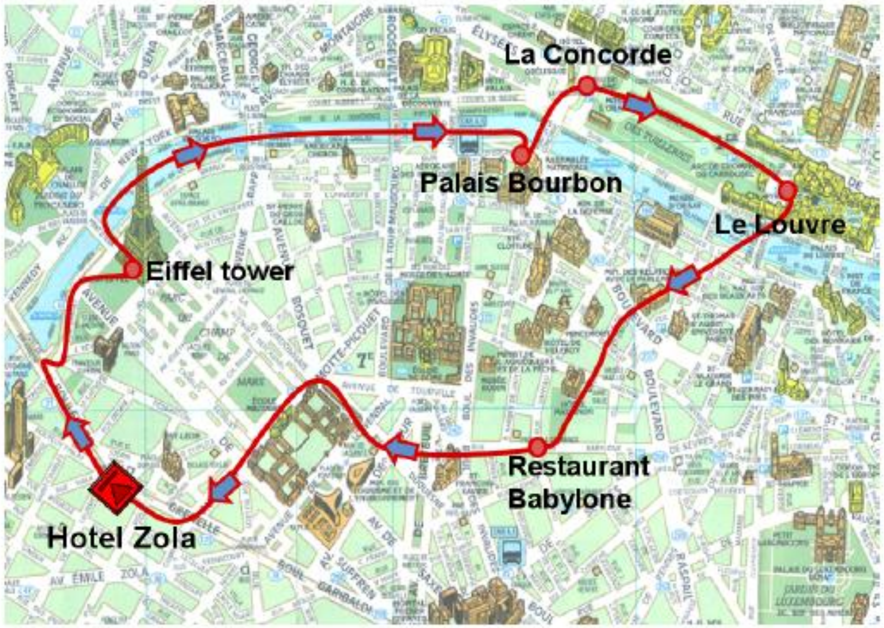


Figure 1 - 2D visualization of a one-day spatial trace left by a tourist visiting Paris – background map downloaded from Mappery.com, copyright unknown

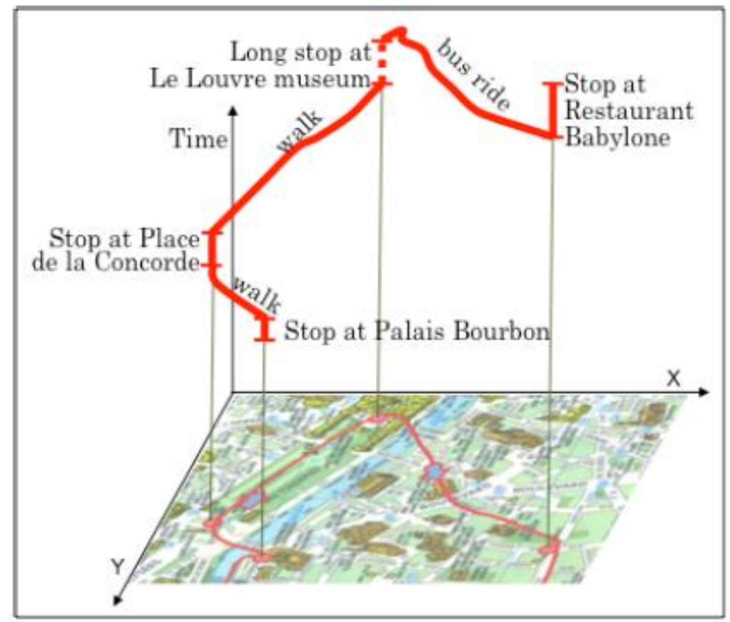


Figure 2 - A time-geography diagram showing part of the previous tourist track

Semantic trajectory

Semantic trajectory

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))) □

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)

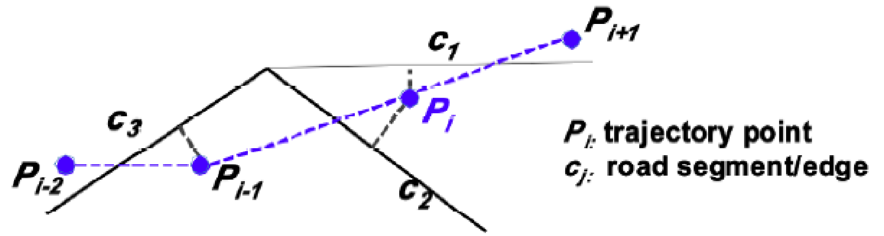


Figure 4 - The map-matching algorithm by [Brakatsoulas et al 2005]

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 semantic 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

Holes, semantic gaps, semantic enrichment

Holes

missing data because of malfunction of the positioning equipment

Semantic gaps

meaningful stop like lunch break during touristic trip

Semantic enrichment = giving annotations

Classification of methods and taxonomies for trajectories

Several researchers have (after van Kreveld) tried to develop “big pictures” on trajectory data mining methods and results

On the following slides two classifications of methods is presented

Mazimpaka & Timpf, 2016 also showing application areas

Yu Zheng gives an overview on research topics inside trajectory data mining

Dodge,S., Weibel,R., Lautenshuetz,A-K present a taxonomy and divides pattersn into generic ones and behavioral ones

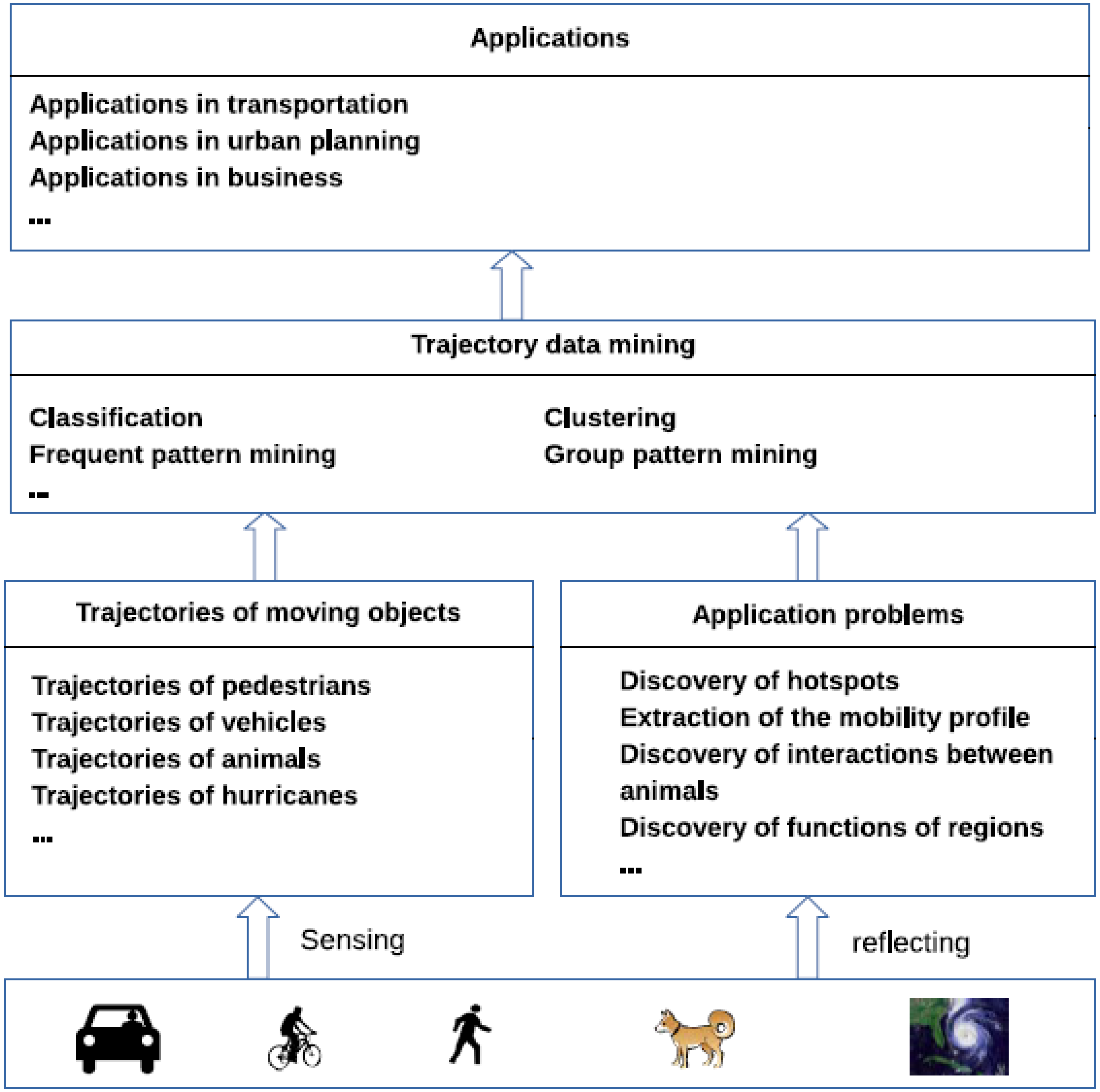


Figure 1: General framework of application-driven trajectory data mining.

(Mazimpaka & Timpf, 2016)

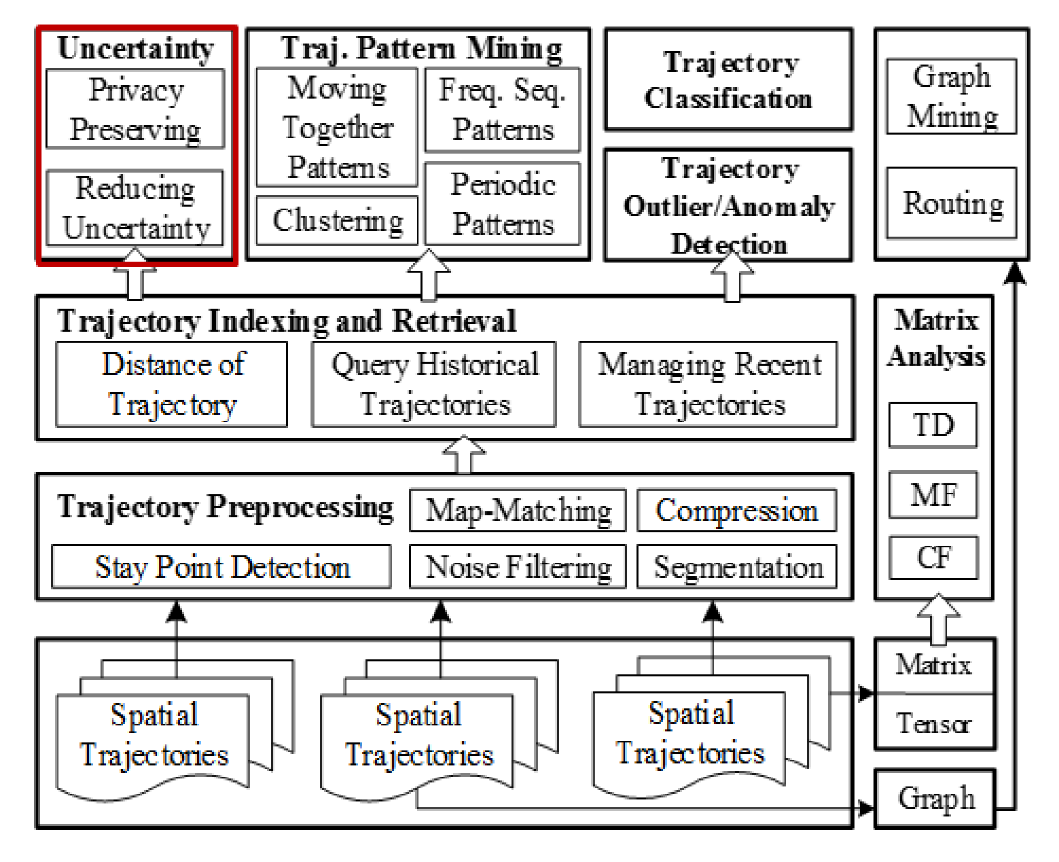
		Secondary methods		
		Pattern mining	Outlier detection	Prediction
Primary methods	Clustering	• Grouping periodically related locations [20, 85] • Grouping close trajectories [62, 129] • Extraction of places of interest for frequent pattern mining [49] • Aggregating close locations for sequential pattern 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)

Research problems 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

Taxonomy of trajectories

Paper by Dodge et al. on taxonomy of trajectories

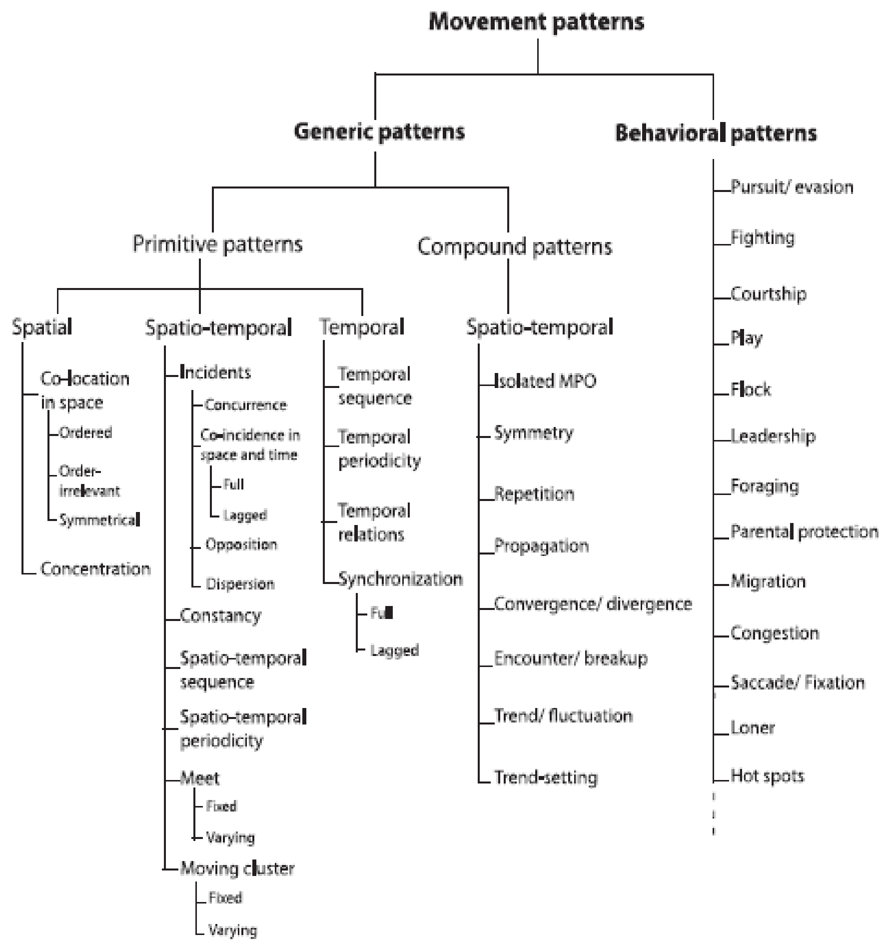


Figure 1: Classification of movement patterns

Literature/References

Aggarwal,C., Data mining, Ch 16.3 Trajectory data mining, 2015.

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