

REPRESENTING UNCERTAINTY OF MOVING OBJECTS DATASETS INTRODUCED BY DENSITY-BASED AGGREGATION OF TRAJECTORIES

MANUEL GILBERTO GARCÍA ÁLVAREZ
March, 2013

SUPERVISORS:

Dr. Ivana Ivánová
Dr. Ulanbek Turduklov



REPRESENTING UNCERTAINTY OF MOVING OBJECTS DATASETS INTRODUCED BY DENSITY-BASED AGGREGATION OF TRAJECTORIES

MANUEL GILBERTO GARCÍA ÁLVAREZ
Enschede, The Netherlands, March, 2013

Thesis submitted to the Faculty of Geo-information Science and Earth
Observation of the University of Twente in partial fulfilment of the requirements
for the degree of Master of Science in Geo-information Science and Earth
Observation.
Specialization: GFM

SUPERVISORS:

Dr. Ivana Ivánová
Dr. Ulanbek Turduklov

THESIS ASSESSMENT BOARD:

Prof. Menno-Jan Kraak (chair)
Dr. Arta Dilo

Disclaimer

This document describes work undertaken as part of a programme of study at the Faculty of Geo-information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

In knowledge discovery from moving object datasets, aggregation is an approach that is typically used. This approach is necessary to extract meaning from large datasets. Uncertainties implicit in raw datasets are propagated to the outputs of trajectory density-based aggregation. In this research we address the problem of computing and representing positional uncertainties with focus on the outputs of trajectory representation, spatio-temporal aggregation and density based estimation. We identified, signal loss, device spatial accuracy, sampling rate, data cleaning, trajectory model, time window, tessellation and radius in density estimation, as factors that influence the level of positional uncertainty in knowledge discovery through density-based aggregation. Random errors were modeled using a stochastic approach to represent uncertainties as regular and irregular error bands, and as probability values. We proposed a system called PUTMO for computation of positional uncertainty of trajectories from moving objects dataset. Implementation of this system allowed adding a readability measurement, in the form of confidence intervals, to spatio-temporal aggregation of trajectories in combination with 2D visualization methods. Applicability and interpretation of our uncertainty model is movement type and context dependent. We demonstrate the applicability of uncertainty computation using hurricane and bird tracking data for the unconstrained case, and pedestrians' data for the constraint case.

Keywords

uncertainty, random errors, moving objects, spatio-temporal aggregation, density estimation, uncertainty visualization

ACKNOWLEDGEMENTS

Special thanks to Dr. Ivana Ivánová and Dr. Ulanbek Turduklov for their support and guidance during this research project.

We thank the Department of Urbanism of the Faculty of Architecture, Delft University of Technology, and MOVEBANK project of the Max Planck Institute for Ornithology, for providing some to the datasets used in this research.

This research was supported by the NICHE Project (GTM042) "Strengthening of Human Resources in Land Administration (LA) in Guatemala"

This thesis is dedicate to my families: my beloved family in Guatemala, and my 'international' family in The Netherlads, who took care of me during the last month of the research project.

"You cannot be certain about uncertainty" -Frank Knight, 1921.

TABLE OF CONTENTS

Abstract	i
Acknowledgements	ii
1 Introduction	1
1.1 Objectives	2
1.2 Research questions	3
1.3 Thesis Outline	3
2 Uncertainties in Movement Representation through Knowledge Discovery	5
2.1 Collection of Moving Object Data	5
2.2 Movement Representation	5
2.3 Knowledge Discovery from Moving Object Datasets	6
2.4 Factors Influencing Positional Uncertainty in Knowledge Discovery	8
2.5 Propagation of Positional Errors in Trajectory Aggregation	10
2.5.1 Errors in Data Collection	10
2.5.2 Errors in Data Pre-processing	11
2.5.3 Errors in Density-based Aggregation	15
2.6 Summary	16
3 Design: Representing Positional Uncertainties of Trajectories	19
3.1 Workflow for Considering Uncertainties in Trajectory Aggregation	19
3.2 Positional Uncertainty on Individual Trajectories	21
3.2.1 Box for Storing Probability Values	21
3.2.2 Algorithm for Error Band Along a Trajectory Segment	22
3.2.3 Algorithm Error Band for a Complete Trajectory	24
3.3 Positional Uncertainty in Aggregation of Trajectories	24
3.3.1 Total Probability for Overlaying Trajectories	26
3.3.2 Density-based Aggregation Considering Error Regions	27
3.4 Summary	28
4 System Implementation: Uncertainty in Different Types of Movement	29
4.1 Conceptual Data Model, Tools and Techniques	29
4.2 Uncertainty in Unconstrained Movement	30
4.3 Uncertainty in Constrained Movement	32
4.3.1 Pedestrian's Dataset	32
4.3.2 Positional Uncertainty in the Paths followed	33
4.3.3 Positional Uncertainty over time.	34
4.3.4 Positional Uncertainty in Density Aggregation	36
4.4 Summary	36

5	Use of Visualizations of Uncertainty as a Quality Indicator	39
5.1	The Purpose of Visualization in the Geographic Context	39
5.2	Application of Visualization of Uncertainties of Density-based Aggregation of Trajectories	39
5.2.1	Visualizing Uncertainties as Adjacent Maps	40
5.2.2	Visualizing Uncertainties as Contour lines	40
5.2.3	Color and Shading Visualizations	42
5.3	Application of Visualizations of Uncertainties in Geo-visual Analytics	43
5.4	Summary	43
6	Discussion, Conclusions and Recommendations	47
6.1	Discussion on the Uncertainty Representation in Aggregation of Trajectories . .	47
6.1.1	Movement and Context dependent Error-Bands Models	50
6.1.2	Visualization methods and Uncertainty in Moving Object Datasets . . .	50
6.2	Conclusions and Recommendations for Future Research	51
	References	53
A	List of Symbols and Variable in this Document	57
B	Database Schema for Storing and Computing Positional Uncertainties of Moving Object Datasets	59
C	Function Tree implemented in PUTMO	61
D	Source Code for Loading GPX files into PostgreSQL: Language DOS	63
E	Source Code of Functions Implemented in PUTMO: Language pl/pgSQL	65

LIST OF FIGURES

1.1	Possible effects on quality of moving objects datasets due aggregation types: spatial (S), temporal (T), and attributive (A); in the process of knowledge discovery.	2
2.1	Movement representations in the space-time cube: (a) Geospatial lifeline, and (b) lifeline bead. As presented in (Honsby & Egenhofer, 2002).	7
2.2	Factors that influence positional uncertainty in the workflow though knowledge discovery.	9
2.3	Positional error around a point considering a circular error ($\sigma_x = \sigma_y$), (Pfoser & Jessen, 1999).	11
2.4	Error ellipse around a trajectory segment as defined by Pfoser & Jessen, 1999.	12
2.5	Error band with error ellipses along a trajectory	13
2.6	Regular and irregular error bands applied to the representation positional errors around trajectories.	13
2.7	Effect of elimination of an invalid position on the error ellipse, and the area of possible movement.	14
2.8	Density probability functions along the trajectory segment. Ellipse representation case.	15
2.9	Error surface and probability distribution along a trajectory segment. Ellipse representation case.	15
2.10	Error band as defined by the isodensity countour and a confidence interval	16
3.1	Components and information flow of the system Positional Uncertainty over Trajectories in Moving Objects (PUTMO).	20
3.2	Buffer and raster box around a trajectory segment.	21
3.3	Estimation of probability parameters based on the geometry of a trajectory segment.	23
3.4	Volume over a pixel using the 'five point concept'. Adapted from Awasthi (2010).	24
3.5	Trajectory aggregation considering error regions (e) and confidence interval (α). Error regions are drawn by computing isodensity contours.	27
4.1	Result with a synthetic trajectories. (a) Error band, and (b) Error surface.	30
4.2	Error bands and probability values for hurricane data. (a) MITCH 1998, $\iota = 30\%$; (b) ISABEL 2003, $\iota = 50\%$; (c) KATRINA 2005, $\iota = 50\%$. Circles (in (b) and (c)) mark areas where irregular error bands are more evindet.	31
4.3	Error regions for oilbirds' flight around the Cueva del Guacharo National Park, Venezuela, 2007. Individuals 24 and 29. Artifacts are enlarge in the detailed inset.	32
4.4	Effect of data cleaning in the error bands and positional uncertainty in pedestrian's track. (a) Original dataset with 307 positions. (b) Cleaned dataset with 78 positions	34
4.5	Uncertainty contributions in two different time windows. Pedestrians shopping on November 19 th , 2009	35
4.6	Probability values of the paths followed by tourists on November 20 th , 2009. (a) Sum of probabilities, (b) Confidence intervals for places visited by all tourists, and (c) Confidence intervals for places visited by at least 4 tourists.	37

5.1	Categorization of uncertainty visualization methods. Only relevant static methods are highlighted. Adapted from Senaratne et al (2012).	40
5.2	Adjacent maps. (a) Density of pedestrians shopping on November 19 th , 2009 and (b) the corresponding spatial uncertainty.	41
5.3	Density values and uncertainties represented as contour lines. Building footprints and 'points of interest' were taken from OpenStreetMap	41
5.4	Visualization of Density of tourist and Positional probability values using color and shading. Building footprints and 'points of interest' were taken from OpenStreetMap	42
5.5	Visualization of Density of tourist and Positional probability values. Color and value representing densities and shading and contours representing probabilities. Building footprints and 'points of interest' were taken from OpenStreetMap	43
5.6	3D representation of density of tourists, November 20 th , 2009. Color represents density values and height represents probability values. (a) High density values and low probabilities close to <i>Sint Hippolytuskapel</i> . (b) Relative high density values and high probabilities around <i>Nieuwe Kerk</i> . Buildings from OpenStreetMap are modeled as 3D boxes	44
6.1	Effect of various size for radius and pixel in two density-based aggregation methods of trajectories. Tourist visiting Delft City, November 20 th , 2009.	48

LIST OF TABLES

4.1	Details of the synthetic trajectory used during the testing of the implementation.	30
4.2	Details of unconstrained moving object datasets.	31
4.3	GPS technical specification. Tracking Delft I dataset.	33

Chapter 1

Introduction

At a dataset level, basic representations of objects' movement include events and trajectories. Events represent locations visited by a moving object at specific time (Andrienko et al., 2011). Trajectories are used, for instance, to represent paths that tourists follow while visiting the highlights of a city. (Thompson, Rahayu, & Torabi, 2012) define a moving object dataset as a collection of spatial properties of an object or group of objects that change over time. "Moving object data typically includes trajectories of concrete objects: humans, vehicles, animals, and goods; as well as trajectories of abstract concepts: spreading diseases" (COST, 2009).

Reliability on moving object datasets finds its relevance in the context of knowledge discovery. Uncertainty influences quality, and this has an effect on the reliability of a dataset. (Shi, 2010) identifies data analysis or processing as a source of uncertainty. Factors that influence uncertainty in a moving object dataset can be present in several stages of the knowledge discovery process: spatio-temporal accuracy of collecting device, trajectory representation of raw data, and the possible aggregation type involved, Figure 1.1. In this context (Trajcevski et al., 2011) model the uncertainty in the database level by using a fixed radius around trajectories and *probability density functions*, in order to consider uncertainty as part of query formulation in moving object datasets. Other examples at database level are given by (Praing & Schneider, 2007; Liu & Schneider, 2011). However, uncertainty is also present during data processing, and a measure of it will contribute to determine the quality of the results. Moreover, summarization is the main goal of aggregation techniques applied to large moving object datasets, and therefore a different level of uncertainty is expected after one or more aggregation types are applied.

Large moving object datasets require aggregation methods to extract knowledge. Spatial, temporal and attributive aggregation (Fredrikson, North, Plaisant, & Shneiderman, 1999) are needed to deal with the problem of scalability in spatio-temporal datasets. In moving object data the development of algorithms to aggregate trajectories, identify movement patterns, visually explore the data, and represent spatio-temporal uncertainty remains as an active research field (e.g., (Demšar & Verrantaus, 2010; Orellana & Wachowicz, 2011; Yoon, Deutsch, Chen, & Goulias, 2012)). Density estimations are among the basic and most common methods to aggregate moving object datasets. Density is computed as the number of recorded locations (events), paths (trajectories), or attribute values per unit of area. Density estimations are also applied to the study of human mobility (Giannotti et al., 2011). Those are relevant for understanding human behavior, and have implications in many contexts: human geography, spatial epidemiology, location based services (LBS), transportation, urban planing, and tourism (Demšar & Verrantaus, 2010).

Geovisualization techniques can be used to communicate the output of aggregation methods and its uncertainty. (Goodchild, 2009) points out *visualization and communication* of uncertainty in the geospatial domain as a problem that is difficult to deal with, for technical and cognitive reasons. With this in mind (Senaratne & Gerharz, 2011) propose a model to automate the selection of visualization methods, where the main decision criteria include *data format, data type and uncertainty type*, and give general guidelines to find a suitable visualization method. Moreover, in the case of visualization of trajectories of moving object data; (Willems, van de Wetering, & van

Wijk, 2009) combined visual variable with kernel density estimation to improve geovisual analysis, and potentially similar approaches can be used to visualize spatial, temporal and attributive uncertainty.

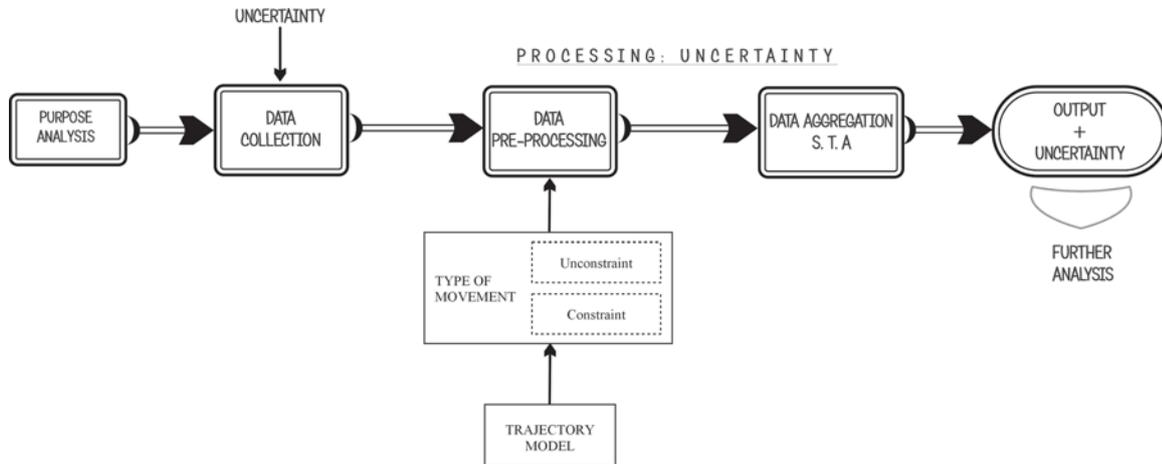


Figure 1.1: Possible effects on quality of moving objects datasets due aggregation types: spatial (S), temporal (T), and attributive (A); in the process of knowledge discovery.

In knowledge discovery from large moving object datasets, aggregation is one of the approaches typically used. Aggregation implies generalization in terms of space, time, attributes, or a combination of them (Andrienko & Andrienko, 2010), and these techniques potentially introduce errors to the original datasets. This approach is compulsory in many cases in order to extract meaning, but a measure of the quality of the output is needed in order to contribute with data analysis process. Uncertainty implicit in the raw datasets, and also introduced by aggregation types during processing has been ignored by data analysts, mainly due to the lack of an appropriate conceptual framework to represent and compute uncertainties related with the aggregation type. This has consequences in one or more of the context mentioned above. In this research we address the problems of how to assess uncertainties in raw datasets, how those uncertainties propagate due to the use of density computation applied for aggregation, and the appropriate methods for visualizing the uncertainty of moving object datasets in the contexts of constrained and unconstrained movement.

The novelty of this research is to apply uncertainty models in the assessment of data quality with focus on the outputs of trajectory representation, aggregation and density-based estimation methods used in knowledge discovery from moving object datasets, to derive a general conceptual framework to estimate uncertainty introduced by the aggregation of moving object datasets.

1.1 OBJECTIVES

The main objective of this research is to propose a framework to assess and visualize uncertainty, implicit in the raw datasets, introduced by the aggregation of trajectories, and applicable to moving object datasets.

For this purpose the following sub-objectives are defined:

1. To provide a conceptual framework for representing uncertainty influenced by spatio-temporal aggregation, in moving object data and human mobility context.

2. To develop algorithms to estimate the uncertainty introduced by representation and density estimation of trajectories.
3. To implement the approach using various moving object datasets case study.

1.2 RESEARCH QUESTIONS

Research questions to be answered are listed below.

Related with objective 1:

- What are the factors that influence positional uncertainty in trajectory representation of moving object data?
- Which factors influence the uncertainty in the spatio-temporal aggregation?
- Which statistic methods facilitate uncertainty measurement in trajectory representation and density aggregation methods, applicable to moving object datasets?

Related with objective 2:

- How to automate the estimation of uncertainty introduced by the representation and aggregation of trajectories using density-based methods?
- Which visualization techniques can be apply to represent the uncertainty introduced by aggregation types and density estimations, in a 2D geographic space?

Related with objective 3:

- How uncertainty propagates due spatio-temporal aggregation and density-based estimation of pedestrians' tracks?
- Is it possible to apply the approach to other moving object datasets?

1.3 THESIS OUTLINE

The content of the coming chapters reflects the methods, results and conclusions derived from this research. A conceptual framework for considering uncertainties in moving object data is given in Chapter 2. We give relevant literature related with the process of knowledge discovery, then describe the main factors influencing positional uncertainty in that process, and present a stochastic approach to derive a measurement for positional uncertainty in moving object datasets. Three main stages are considered in knowledge discovery: data collection, data processing, and aggregation. In chapter 3 we implement the conceptual framework. A system is designed in order to include reliability as part of the questions the knowledge discovery process intends to answer. We give the algorithms and functions to automate the uncertainty estimation along trajectories. Case studies using real moving object data appear in Chapter 4 . Hurricane, birds tracking and pedestrians' moving object datasets were used. We also demonstrate the effect of data cleaning, and the use of different time windows in the level of uncertainty. Chapter 5 industries the use of adjacent maps, contour lines representation, color and shading to visualize uncertainties. The use of other visualization techniques as part of Geo-visual analytics are also considered. Chapter 6 addresses the movement and context dependencies of the approaches in this research , and discusses the applicability of the error band models to represent positional uncertainty for constrained and unconstrained moving objects. We also give conclusions and recommendations, and some relevant guidelines for future research work.

Chapter 2

Uncertainties in Movement Representation through Knowledge Discovery

Data collection, movement representation, and data aggregation are part of the process of knowledge discovery. In the process of extracting knowledge from moving object datasets, uncertainties are introduced in various stages. The need of an approach to effectively understand the quality of the data in every stage has been recognized as relevant in handling spatio-temporal data (Andrienko et al., 2010). In this chapter an introduction to the process of knowledge discovery, and the sources and factors that introduce uncertainties in the case of unconstrained and constrained movement are presented. Finally, we present two methods to consider positional uncertainty in spatio-temporal aggregation of trajectories. An extensive list of symbols used in this chapter and in Chapter 3 is given in Appendix A.

2.1 COLLECTION OF MOVING OBJECT DATA

A moving object dataset is a set of locations over time, Section 1. The collection of position and time while the object is moving is usually acquired using GPS receivers.

Quality of collected data in terms accuracy depends partially on the choice of the device. GPS -*Global Position System*- receivers use a satellite network to establish the position and time of a moving object, and therefore any factors affecting the communication between receivers and satellites affect the measurements, those effects can be the result of biases in the satellite network, atmospheric conditions, and the reflection of signal on other objects, *e.g. buildings, mountains, etc.*, what is called multipath (Leick, 2004). In case of mobile phone, the quality of the data depends on more factors, as usually the position and time is determined by using a combination of positing methods, which can include: mobile networks, Wi-Fi networks, and sensor aid positioning. In the present study we limit to the case of GPS positing of a moving object, as it represents a simple case and the effects on the measurements have been widely studied.

GPS positioning provides location and time of a moving object. Position is recorded as a collection of geographic coordinates, longitude and latitude, or their equivalents (x, y) in a projected coordinate system. Time is recorded using the atomic clock on board of the satellite, and it is converted to local time. Then for every instance $[0 - n]$ the position of an object is recorded as a measurement of x, y and the time t , and the movement of an object is constructed by plotting its position in a chronological order. Then, we can define a moving object dataset (τ) , a *spatio-temporal* dataset, as:

$$\tau = \langle (x_0, y_0, t_0), (x_1, y_1, t_1), \dots (x_n, y_n, t_n) \rangle$$

2.2 MOVEMENT REPRESENTATION

The movement of a moving object can be represented as events or trajectories. If we consider a moving object as a discrete geographical entity (a point) that change its position over time,

we call this point a *moving point* (Schneider, 2009). A moving object dataset contains records of the positions visited by the object during its movement, records related to its historical movement (Schneider, 2009; Praing & Schneider, 2007). Andrienko and Andrienko (2010) proposed two representations to performed data analysis on moving object datasets: event-oriented or trajectory-oriented. In the present study we limit ourself to the trajectory-oriented representation.

Movement of an object can be classified as unconstrained or constrained. In the unconstrained case, objects's movement is not limited by other spatial objects, e.g. vessel moving in the sea, or hurricanes. Meanwhile in the constrained case, objects' movement is restricted by the environment in which the movement is realized, commonly this environment is defined by a spatial network, e.g. transportation networks in the case of car's movement (Schneider, 2009).

A trajectory represents the historical movement of an object in unconstrained and constrained environments. A trajectory is constructed by joining all the positions recored for a moving point using straight lines, and it represents the path that the moving point followed. This model for representing the movement assumes that the movement follows a straight path between records, which is not always the case if we consider that a moving object can exist in a free environment, in the unconstrained case. Then we formally define a trajectory as a set of *spatio-temporal* points (a moving object dataset) that belong to the same moving point (*oid*), and resembling the path followed, where every point is three dimensional (x-coord, y-coord, time). A trajectory can be represented as (Yan, Chakraborty, Parent, Spaccapietra, & Aberer, 2011):

$$\hat{\tau} = \{oid_i, (x_{i_1}, y_{i_1}, t_{i_1}), (x_{i_2}, y_{i_2}, t_{i_2}), \dots, (x_{i_n}, y_{i_n}, t_{i_n})\} \quad (2.1)$$

where oid_i identifies the object, and $t_{i_1} \leq t_{i_2} \leq t_{i_n}$. This model assumes that the objects move along straight lines, and at a constant speed (Trajcevski et al., 2011).

In the space-time cube a trajectory is also called a *lifeline thread*, when the position of the moving point is plotted against time, Figure 2.1 (a). The set of possible locations that the moving object could reach when a maximum speed is defined can be represented as two inverted cones, the so called *lifeline bead* (Miller, 1991; Hornsby & Egenhofer, 2002). This representation is based on the capability constraint and the idea of space-time prism by Hagerstrand (1970).

2.3 KNOWLEDGE DISCOVERY FROM MOVING OBJECT DATASETS

In the process leading to knowledge discovery the purpose needs to be established. The purpose of the analysis conditions the selection of appropriate methods to extract the information we have in mind. Knowledge discovery is defined as a combination of methods to semi or fully automate the analysis of massive dataset (Andrienko et al., 2010), and in the present context we refer to moving object datasets. It takes advantage of data mining techniques for pattern identification and facilitates interpretation through simplification, for instance by aggregating moving object data based on space, time or attributes. Among those techniques: classification, clustering, aggregation, event detection, and visualization are mentioned (Andrienko et al., 2010, 2011).

Aggregation of moving object data facilitates interpretation of the movement. Aggregation involves to group the data in order to identify common patterns. The *grouping* technique can be based on any of the components of a moving object dataset:

- *space*, in which case the geographic space is usually divided using a regular grid and the data is aggregated based on common positions;

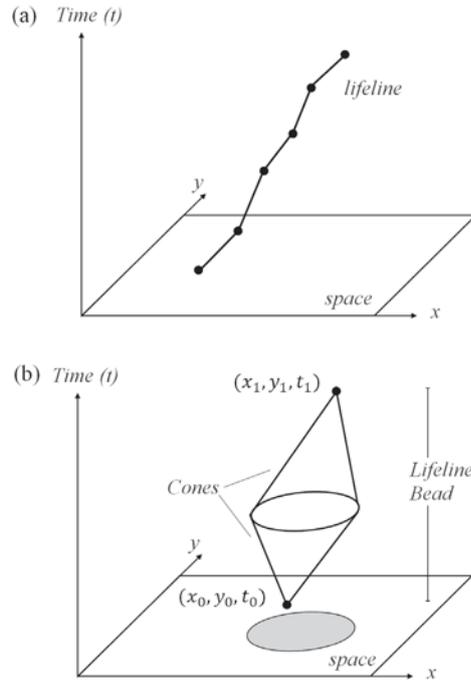


Figure 2.1: Movement representations in the space-time cube: (a) Geospatial lifeline, and (b) lifeline bead. As presented in (Honsby & Egenhofer, 2002).

- *time*, in this, data is aggregated using time windows or time intervals, and data is grouped independent of its position; and
- *attributes*, in which quantitative (e.g. speed) or qualitative (e.g. origin) properties are the ones which determine how the data is grouped.

More often a combination of those types of aggregations (Fredrikson et al., 1999) are used, since movement pattern and behavior of the object does not depend necessarily on its individual components but rather of a combination of them. Complex spatial, temporal, and attribute combinations to aggregate moving object datasets from a exploration-task oriented point of view are given and exemplified in Andrienko and Andrienko (2010).

Density estimation is a well-known technique to aggregate moving object datasets. Density estimation can be used as a tool for identification of major flows, routes, use of paths, and movement comparison (Andrienko & Andrienko, 2010), which in further analysis becomes evidence to answer some spatio-temporal questions, *questions that ask for where and when or how long*. The most simple estimation of density involves a count of the number of trajectories per unit of area, as described in Section 2.2, and represented by φ as follows:

$$\varphi = \frac{\sum_{i=1}^n \hat{\tau}_i}{A} \quad (2.2)$$

$$\text{for } t_i \in T$$

where $\hat{\tau}_i$ are individual trajectories, T is the *time window*, and A represents the area in any arbitrary squared units.

In aggregation of moving object data, density estimation is used to aggregate space and time at once. These two components cannot be considered independently if the idea of movement

needs to be maintained. This is not the same in the case of attributes, which can be disregarded if not relevant for the type of information that needs to be obtained. Space is aggregated by defining the extension of A , which in principle can be of any size and shape, but usually a circle is considered; such that $A = 2\pi k^2$, and then the definition of the radius (k) becomes relevant. Time is aggregated by defining T before the density is computed, and including only the data which match the condition given in equation 2.2. In the GIS domain a raster model is used to store the results from the density estimation, then φ is computed using the center of each cell as origin, and the neighborhood defined by k . The result from the density estimation is especially useful in identifying busy spots (routes, path ways, places, etc.); and its interpretation requires further analysis; in many cases contextual information is added or inference techniques are applied, a subject that is beyond the scope of this thesis. More advance methods to estimate densities, and examples of its application in the domain of moving object datasets can be found in Willems et al. (2009); Demšar and Virrantaus (2010); Giannotti et al. (2011); Scheepens, Willems, van de Wetering, and van Wijk (2012).

In the following sections we review the process of knowledge discovery presented in Figure 1.1, and point out the sources of uncertainties in every one of the stages. A compilation of methods to measure those uncertainties are proposed as well.

2.4 FACTORS INFLUENCING POSITIONAL UNCERTAINTY IN KNOWLEDGE DISCOVERY

Moving object datasets contain uncertainties in both spatial and temporal dimensions. Spatial uncertainty refers to the uncertainty in measurements in terms of position in the geographical context, x, y in *two dimensional space*. Similarly, temporal uncertainty refers to the time measurements. Moving objects are well defined objects and uncertainties can be represented as errors that can be estimated using stochastic approaches (Fisher et al., 2006; Shi, 2010). In the basic model for positional errors, any measurement is affected by three types of errors: gross, systematic, and random; which can be represented as:

$$\hat{x} = \mu_x + \varepsilon_g + \varepsilon_s + \varepsilon_r$$

where \hat{x} represents the measurement, μ_x the true value, and the epsilons (ε) the gross, systematic and random errors, respectively. Those errors contribute to the level of uncertainty in the measurements. Gross errors are considered as mistakes or blunders, systematic error have a magnitude that follows a pattern and do not change in sign, and random errors are irregular in magnitude and sign, and occur by chance (Shi, 2010).

For a well-defined moving object, uncertainty refers to errors (Fisher et al., 2006). Positional errors are introduced in different stages: data collection, movement representation, and data aggregation; and therefore factors influencing these errors can be found in those stages. By identifying where in the knowledge discovery process one or more of the errors play a role, it is possible to develop a framework to propagate positional uncertainty through the stages and considering the representation of movement and density-based aggregation. (See Figure 2.2).

In the data collection stage, factors related with the collecting device are more relevant. The final positional measurements are influenced mainly for the loss or poor quality of the signal, the spatial precision of the device, and the *sampling rate* at which the device is set up. Poor signal conditions have as major consequence that positions representing the movement are: on one hand omitted, and on the other hand recorded far away from the *true* position. Device precision is instrument dependent, for every measurement recorded part of the reliability of it is determined by the precision of the instrument, and it does not change from measurement to measurement. The frequency at which the location of the object is recorded, called *sampling rate* (Pfoser & Jessen,

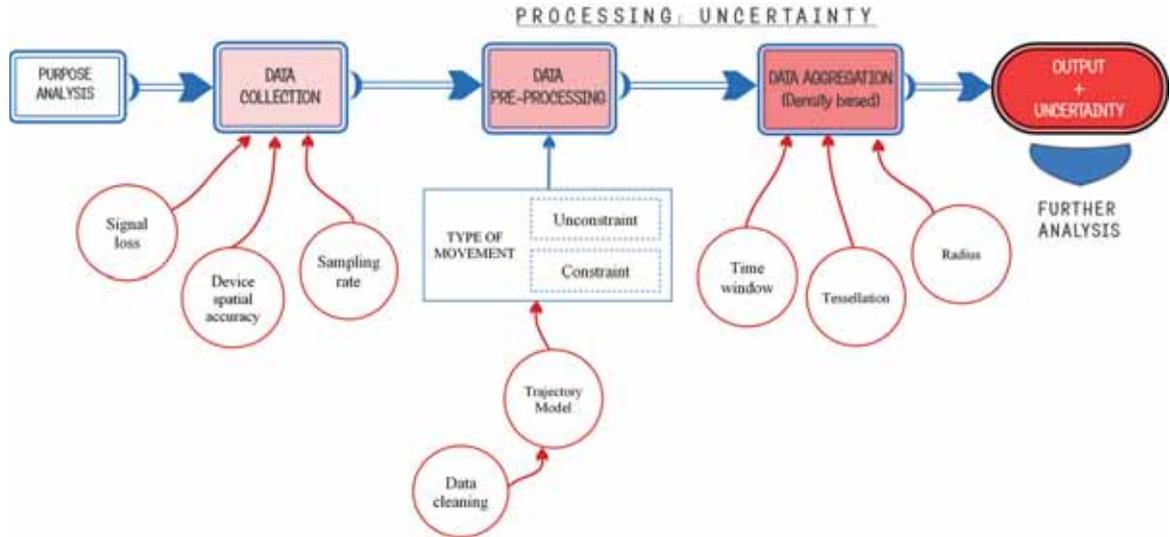


Figure 2.2: Factors that influence positional uncertainty in the workflow though knowledge discovery.

1999), influences the granularity at which the movement is recorded, the higher the sampling rate the more the dataset represent the 'true' movement of the object and therefore less uncertainty is expected (Liu & Schneider, 2011). This effect is only relevant in the case at which the *manoeuvrability* of the object allows sudden changes of positions between consecutive recorded positions. Sampling rate with origin in the data collection stage has a more relevant effect when considering error in movement representation.

In movement representation, data cleaning and trajectory representation introduce uncertainty. Data cleaning is a pre-processing step in which positions regarded as gross errors or invalid measurements are removed from the original dataset to reduce the level of error from the data collection stage, common criteria for identifying the invalid measurement are distance and/or speed thresholds. As a result a new dataset is produced in which 'wrong' measurements are eliminated, but this also implies that the location of the object at that specific time instance t , is unknown; as a consequence the gross errors are reduced, but random errors still need to be considered. Trajectories assume that the movement follows a straight path between recorded positions, in that representation the possibility of movement in other direction than a straight line is neglected; even though since the beginning of time geography, the definition of *capability constraint* by Hagerstrand (1970) recognized paths of possible movement between consecutive positions as spatio-temporal regions, by what is called space-time prisms. With this idea in mind, the model used to represent the movement of an object itself introduces uncertainty in the form of random errors to the dataset.

Density aggregation also contributes to the level of uncertainty. One of the factors is the selection of the *time window* T , as mentioned in Section 2.3. This factor does not affect the level of errors directly, but it conditions the final result as the level of errors, and therefore uncertainty, can be different (in location and values) if a time windows is defined, and in such time interval the measurements that belong to T contribute with more or less errors coming from the factors of previous stages. This means that a different level of uncertainty is expected for different time windows, which could be bigger or smaller. Space is aggregated using the radius, such that only the trajectories that are within the circle defined by R are considered in the density computation. Therefore the choice of R becomes important, using a big radius implies that trajectories falling

far from the origin are considered in the computation, which contribute with a bias in the computation. As a rule of thumb the smaller the radius, the more the accurate the results coming out of the density estimation. Another factor is the *tessellation* of the geographic space to store the result of φ , the major effect to this is the positional displacement (bias) of the results caused by the discretization of the space, which means that the outputs of the density estimations depend on the size of pixel used (Andrienko et al., 2010). This effect is known as the *modifiable areal unit problem*, see Openshaw (1984) for details. Then, the final output of the aggregation is usually stored using a raster model, in which according to the factors stated here inherits the uncertainty in the position for trajectories involved. As an effect of this transformation, the actual position of the trajectories is displaced. This ultimately affects any interpretation of the results of aggregation, and the conclusion drawn out of them.

2.5 PROPAGATION OF POSITIONAL ERRORS IN TRAJECTORY AGGREGATION

In this section, we focus on how to propagate and measure positional uncertainty of the factors described in the previous section. For this purpose we propose an approach that combines the ideas of representing the movement as lifeline threads and time beads (Hornsby & Egenhofer, 2002) using the space-time prism concepts (Miller, 1991), the representation of the position of a moving point and consequently the error round a line segment as an stochastic process by (Shi, 2010). This approach follows three main assumptions: (a) *positional random errors are more relevant, and the only ones considered*, (b) *the errors follow a normal distribution*, and (c) *no errors are considered in time measurements*.

2.5.1 Errors in Data Collection

For a trajectory $\hat{\tau}$ every position along itself is represented by the vector $P = [x, y, t]^T$ for a specific time instance t , in which x, y are the positional components of P , and the only ones considered for the meantime. Then equation 2.1 can be simplified to the set:

$$\hat{\tau} = \{oid; P_0, P_2, \dots P_n\} \quad (2.3)$$

Random errors around P are influenced by the accuracy of the device, e.g. given as the *circular error probable CEP*, which refers to the circle around a given position which contains the 'true' location with 50% of probability. Accuracy of the device is stated by the technical specifications that the manufacturer provides.

In this sense we can consider the total error in any individual position P as:

$$\epsilon_{(x,y)} = \epsilon_{r(x,y)} + \epsilon_i(x,y) \quad (2.4)$$

where ϵ_i represent the portion of the error that is influenced by other sources during the data collection process: gross error and systematics error, which according to our assumptions will be disregarded. Then equation 2.4 can be simplified as follows:

$$\epsilon(x, y) = \epsilon_r(x, y)$$

and,

$$\epsilon_r(x, y) \approx \sigma_{device}$$

where the accuracy of the GPS represents the positional error and can be approximated by the standard deviation of the device σ_{device} in both direction x and y . If we consider that the error

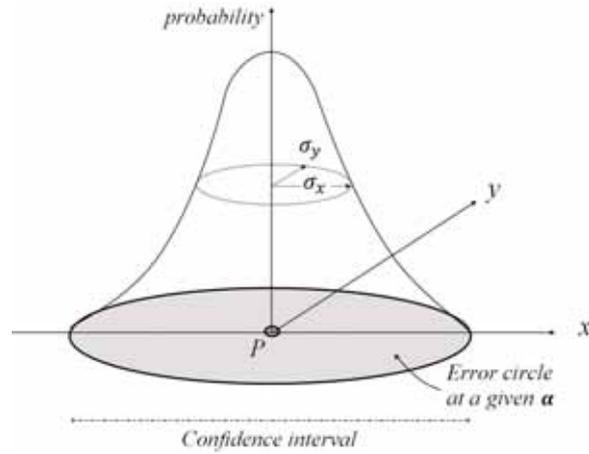


Figure 2.3: Positional error around a point considering a circular error ($\sigma_x = \sigma_y$), (Pfoser & Jessen, 1999).

follows a normal distribution (Pfoser & Jessen, 1999; Shi, 2010) and the correlation ρ between x and y is equals to *zero*. This leads to a circular representation of the error around a single position limited by a confidence interval α such that the radius of the circle is a multiple of σ , and the circular one contains the 'true' location of the point with a probability larger than or equals to a confidence level (Shi & Liu, 2000), Figure 2.3. The probability ϕ of the *error circle* to contain the true position of the point is given by the density function describe by the bivariate normal distribution (Shi, 2010; Pfoser & Jessen, 1999):

$$\phi_i(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\right\}\left\{\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \frac{(y-\mu_y)^2}{\sigma_y^2}\right\}} \quad (2.5)$$

This approach allows to represent the error around every position as an stochastic process (Shi & Liu, 2000), and to estimate the uncertainty in every location in terms of probability by considering the volume created by the error circle, and the surface of probability density function define by equation 2.5. In our case, the error around P 's in the trajectory, and its probability density function are considered the same for every known position, since σ_{device} does not change.

2.5.2 Errors in Data Pre-processing

Data cleaning and *trajectory representation* are the main factors influencing uncertainty in movement representation. In this section a method to compute and propagate the error from the previous stage is proposed. For this purpose the model of error representation proposed by (Pfoser & Jessen, 1999) and the probability model for error propagation along a line segment by (Shi, 2010) are combined.

The most probable movement between two consecutive position P_0 and P_1 is represented by a straight line segment (Liu & Schneider, 2011), its trajectory. This approximation of the 'actual' movement introduces positional uncertainty. Lifeline beads bind the movement of a moving object in the space-time cube. The projection on the geographic space of the region created by the intersection of the two cones that form the lifeline bead, is circular in the case the object is not moving and elliptical otherwise. This model allows representation of the probable positions of a moving object by defining error ellipses (Pfoser & Jessen, 1999) around trajectory segments, and to represent uncertainty regions around a trajectory by joining the error ellipses of individual segments.

Error ellipses around trajectory segments are estimated according to (Hornsby & Egenhofer, 2002; Pfoser & Jessen, 1999). In their approach the error ellipse is parameterized by: (a) the two end points P_0 and P_1 defining the trajectory segment and considered as the foci of the ellipse; (b) the semi-major axis a defined by the maximum speed v_{max} that the object can reach, and the time interval between positions $\Delta_t = t_1 - t_0$, such that $2a = v_{max} \times \Delta_t$; and (c) the semi-minor axis b , such that $2b = \sqrt{(2a)^2 + (2c)^2}$, where c is the distance between positions. Which means that the size of error ellipses is proportional to Δ_t . This is illustrated in Figure 2.4 taken from (Pfoser & Jessen, 1999).

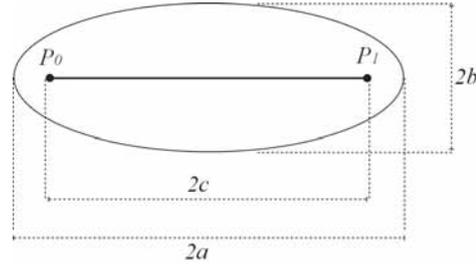


Figure 2.4: Error ellipse around a trajectory segment as defined by Pfoser & Jessen, 1999.

For a complete trajectory the joint of all the error ellipses defines the space of possible movement in $2D$, an *error band*, Figure 2.5. This representation is suitable for the case where the minor axis of the ellipse error is larger than the diameter drawn by the circular error around a single position, as it is explained further in this section.

In the representation of error bands around the trajectories, two cases are considered: regular and irregular error bands. In the first one, the sampling rate at which the positions are recorded, is *high* compared with the speed of the movement. This causes that the error *circles* at every position and modeled as proposed in Section 2.5.1 to be overlaid, and the minor axis $2b$ is *smaller* than the diameter of the error circles on the corresponding positions P_i ; in which case the error region around the trajectory is fairly represented by a *regular* error band around the trajectory segments that meet this condition. In the second case, the sampling rate in comparison with the speed of the object generates an ellipse for the trajectory segment, in which $2b$ is larger than the diameter of the error circles that belong to the same trajectory segment. In such case, the uncertainty regions around that segment is more fairly represented by an error ellipse and produce a *irregular* error band (See Figure 2.4). In a single trajectory both cases can be used to represent error regions (See Figure 2.6), if the sampling rate is high enough to use the first case, but in some of the trajectory segments this rate is altered by factors in data collection or by the data cleaning process.

Size of error ellipses is proportional to the time interval between the positions. These characteristics of the error ellipse allows to consider the case in which positions are removed from the original dataset, in a preparation step known as *data cleaning*. Data cleaning is used to eliminate outliers from the moving object datasets. The process removes wrong or *invalid* measurements from the dataset, but this raises the question: *Where did the object move at that instant of time?*; to look for a direct answer is beyond the scope of the present work. We focus on the effect of the removal of invalid positions in the case of error ellipses. The effect can be described as follows: errors along a trajectory segment are represented as an error bands around a straight line that join two endpoints (error ellipses), which contains the true location of the path followed. When a position is removed under the assumption that it is an invalid measurement, a gross and/or systematic error is removed from the dataset. But at the same time, the position of the moving object is unknown at that time interval, which makes the position of the moving object for the time between points P_i more uncertain.

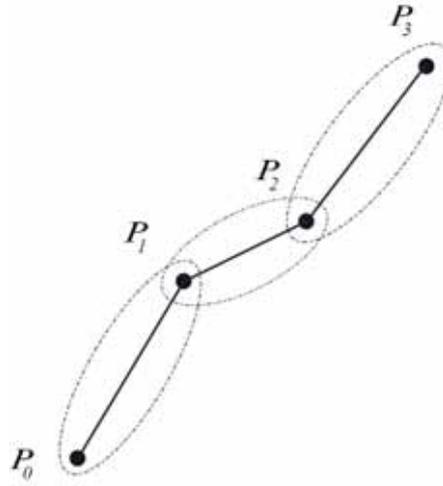


Figure 2.5: Error band with error ellipses along a trajectory

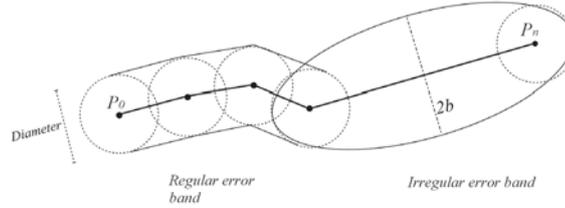


Figure 2.6: Regular and irregular error bands applied to the representation positional errors around trajectories.

For a trajectory, removal of outliers increase the time interval between positions. The removal of *invalid* positions decrease the sampling rate of the movement (Pfoser & Jessen, 1999). Common basic method to identify invalid positions is based on the speed of the object, and a threshold is set based on the type of movement. In GPS data collection the sampling rate is set to match a constant time interval Δ_t , if we consider the trajectory $\hat{\tau}_1 = \{1; P_0, P_1, P_2, P_3\}$ and a sample rate s_r ; then the removal of an inner position , e.g. P_1 in Figure 2.7, increases the time interval, such that:

$$\Delta_t = \begin{cases} s_r \times N_R & \text{if } N_R > 0 \\ s_r & \text{if } N_R = 0 \end{cases}$$

where N_R is the number of consecutive points that were removed, and the size of the error ellipse around the new trajectory segment is expanded, and its expansion is influenced by the new time interval between the new set of positions representing the corresponding segments of a trajectory.

Boundaries of the possible movement are described by the error ellipses and error circles. Positional uncertainty is represented at every position P_i by the bivariate normal distribution presented in Section 2.5.1 in terms of probability. To propagate the uncertainty from P_i to the trajectory segments, a probabilistic model of error propagation by Shi (2010) is adopted. In his approach the error bands and error region, represented in our case by the error ellipses, are intersected with an error surface define by the probability density function *pdf*, derived from the variance at the end points of trajectory segment. We modified this approach such that the (*pdf*) at a random point and $\rho = 0$, up to the middle point of the trajectory segment, is estimated by $\hat{\phi}$:

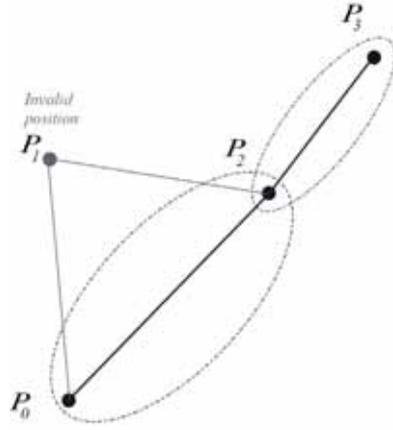


Figure 2.7: Effect of elimination of an invalid position on the error ellipse, and the area of possible movement.

$$\hat{\phi}_i(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{(x-\mu_x)^2+(y-\mu_y)^2}{2\sigma_x\sigma_y}} \quad (2.6)$$

where σ_r is the standard deviation along the trajectory segment in a random position r , which is maxima at the middle point between positions (*half of travel time between positions*), the correlation coefficient is zero as we consider x and y independent from each other, variance $\sigma_{rx}^2 = \sigma_{ry}^2$ increases from σ_P at P_0 up to the middle of the trajectory, and then decreases from that point to converge with the circle error at P_1 , Figure 2.8. The increment and posterior decrement of σ at the random point r , σ_r , is define by a function Γ , in which $\sigma_r = \Gamma(\sigma_P, t)$, is a function of the known σ at P_0 , and time. The definition of Γ is such that the distribution of the probability of finding the moving object around its more probable position, the trajectory segment itself, depends on the space and time available for the movement as in the *space-time prisms*, which is a function of time interval Δ_t between positions. Therefore $\Gamma = \sigma_P \pm \frac{\sigma_P dt}{dt}$, in which cases:

$$\sigma_r = \begin{cases} \sigma_P + \frac{\sigma_P dt}{dt} & \text{for } t_0 \leq \frac{\Delta_t}{2} \\ \sigma_P - \frac{\sigma_P dt}{dt} & \text{for } t_0 > \frac{\Delta_t}{2} \end{cases} \quad (2.7)$$

Then the probability of the moving point around the trajectory can be modeled by the volume defined by the error surface and the 2D plane (See Figure 2.8). Error surface is the result of the joint probability for all possible *pdf* over the trajectory segment in the movement direction, Figure 2.9. The 2D plane (Ω) represents the geographic space in two dimensions (x, y). Confidence interval α limits the error surface in perpendicular direction. Meanwhile, Γ shapes the ellipse error around a trajectory segment, and controls the height of the probability surface such that is minimum in the middle.

By knowing the initial σ_P and σ_{rmax} at the middle of the segment, a value for $\frac{\sigma_P dt}{dt}$ depends on the distance and time from the initial position P_0 and a non linear function, in order to resemble an ellipse.

The proposed approach makes possible to estimate uncertainty in terms of probability for any point along a trajectory segment, and for every point around it. Then the error band around a trajectory and its probability (equation 2.8) is represented as an error surface, which is irregular in both the 2D plane and the height of the error surface (See Figure 2.9).

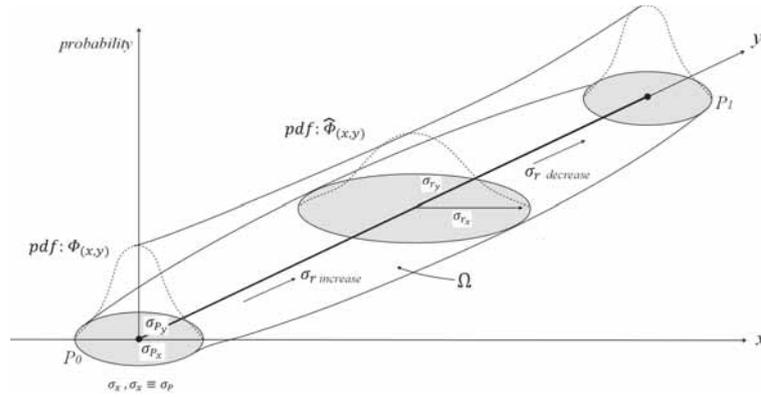


Figure 2.8: Density probability functions along the trajectory segment. Ellipse representation case.

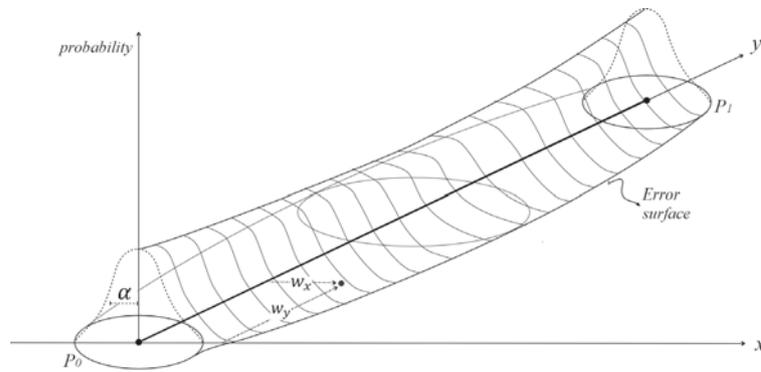


Figure 2.9: Error surface and probability distribution along a trajectory segment. Ellipse representation case.

$$\Phi = \int_{-\infty}^{\infty} \int_{\sigma_r=t_0}^{\sigma_r=t_1} \frac{1}{2\pi\sigma_r^2} e^{-\frac{(w_x)^2+(w_y)^2}{2\sigma_r^2}} \quad (2.8)$$

where $\sigma_r = \sigma_x = \sigma_y$; and w are the components of the euclidean distance between an arbitrary point under the error surface and the trajectory segment, in x and y directions.

The same approach can be applied consecutively to every segment j of a trajectory $\hat{\tau}$, then the probability of the moving point inside the error band along a complete trajectory S is represented by the union of all the probability values of the individual segments, such that $j \in \hat{\tau}$:

$$S_{1\dots m} = \cup_{j_1}^m \Phi_j \quad (2.9)$$

2.5.3 Errors in Density-based Aggregation

Positional uncertainty of the trajectories are propagated to the density estimation outputs. When density of trajectories is computed using equation 2.2, the position of the moving object along a trajectory is uncertain, as well as its location inside A . The definition of an error band around the trajectories and its corresponding probability will allow to represent the uncertainty in the density computation in two different ways: (a) as a *total* probability surface for the predefined time window, and (b) by relating the output to a confidence interval α at which we consider the presence of the moving object inside the neighborhood.

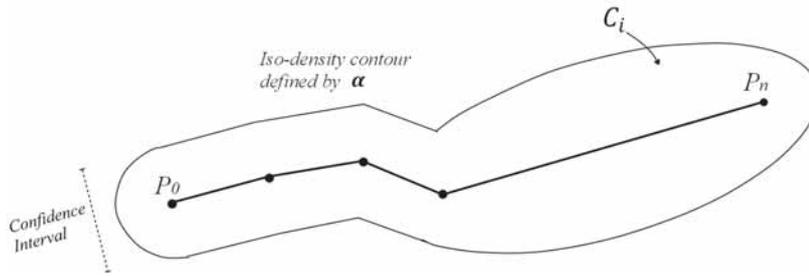


Figure 2.10: Error band as defined by the isodensity countour and a confidence interval

In the first case, probability theory is used to obtain a probability value that represents the likelihood of all the trajectories occurring as independent events (Shi, 2010), in the geographic extent defined by the time window. New probability values are computed by equation 2.10, which represents the likelihood of finding any point P_i in a trajectory at any specific location (Brunsdon, 1995).

$$Pr(B \cap C) = P(B)P(C) \quad (2.10)$$

For which B and C are independent “events” (probabilities that describe error bands in our case). Then by the product of the corresponding probability values for a set of trajectories, we obtain an aggregated probability value that represent the positional uncertainty of the set of $\hat{\tau}$.

In the second case, a confidence interval will define the reliability of density computation. When the error surface is computed to define an error band around trajectories by using the joint probabilities from equation 2.8 and confidence interval α . Aggregation by density estimation can be computed by the intersection between *error regions* or *isodensity contour*, which defines “the set of points for which the probability values are the same” (Gatignon, 2010). Then equation 2.2 can be redefined as:

$$\hat{\varphi} = \frac{\sum_{i=1}^n C_i}{A} \quad (2.11)$$

C_i represents the region bounded by a the isodensity contour which includes all the probability values equal or bigger than the contour itself (See Figure 2.5.3). C_i is defined by α , and determines if the trajectory is taking into account in the density estimation for that specific α . By considering the intersection of isodensity contours in the density computation the results of density inherits the confidence interval. This confidence interval becomes a quality indicator for the density aggregation.

2.6 SUMMARY

Data collection, movement representation and density aggregation are the main stages in the workflow for knowledge discovery presented here. Movement of a point can be modeled as trajectories in constrained and unconstrained environment. Trajectories represent the historical movement of the object, in which we assume the movement keeps a constant speed. Factors affecting positional uncertainties in that workflow are: (a) in data collection: *signal loss, device precision, and sampling rate*; (b) in movement representation: *data cleaning, and the trajectory model*; and (c) in aggregation by density estimation: *time window definition, radius in density estimation, and tessellation of the geographic space*. The ideas of capability constraints, lifeline beads, error ellipses, and a stochastic approach to represent error regions are combined to model the positional

uncertainty around trajectories as regular and irregular error bands. Uncertainties of the final outputs of the density estimation can be represented as: (1) a normalized probability surface of the positions of the trajectories, and (2) as an indicator of the reliability of the outputs, given by a confidence interval influencing the result of density aggregation.

Chapter 3

Design: Representing Positional Uncertainties of Trajectories

In this chapter the prototype of the system *Positional Uncertainty over Trajectories in Moving Objects -PUTMO-* is described. PUTMO is a system that allows of positional uncertainties over moving object datasets. A workflow is proposed to include a measurement of reliability to the outputs of the density-based aggregation of trajectories.

3.1 WORKFLOW FOR CONSIDERING UNCERTAINTIES IN TRAJECTORY AGGREGATION

PUTMO has three main components and the prototype computes positional uncertainty in density based aggregation as shown in Figure 3.1. In (a) the uncertainty for individual trajectories is computed, then the outputs can be used as either to (b) compute a surface that represent the total positional probability for a set of trajectories, or (c) to compute the density values under a predefined confidence interval using error bands. In advance a repository for storing the moving object dataset is necessary, in the present case a database management system is recommended as the volume of information is usually high. A database schema contains the data using a trajectory model, which requires to store an ID for every trajectory and a list of position over the time (See Appendix B).

Positional uncertainty for every single trajectory is computed in the first stage of the process. In this stage the system generates an empty raster with the predefined pixel size, and which dimensions are big enough to contain the spatial extent of the trajectory and the positional error related with it. Then, considering the set of positions that define the trajectory, the accuracy of the collecting device (e.g. GPS) given as the *probable circular error* and considered as $\sigma_{(x,y)}$, the error defined by σ_r along the trajectory segments, and δ_t as parameters, a surface error is defined and the uncertainty values are computed for every pixel. The output in this stage is a set of rasters containing the uncertainty values as defined by equation 2.6.

In case of choosing a total probability representation, two subprocesses are required. The computation of density-based aggregation of the original subset of trajectories, and in parallel the set of uncertainty rasters are combined by computing a new surface error using probability theory (See 2.5.3). This produce two different outputs: one raster containing the density values, as number of trajectories per unit of area, and the other as a raster with total probability values. Both outputs are combined using visualization methods.

As an alternative, the system will be able to compute density aggregation base on a predefined confidence interval. Then, by computing the *isodensity contour* for the probability value that correspond to α , a region can be drawn. Then, the density aggregations is estimated by equation 2.11. This produces as output a raster containing density values under an implicit confidence interval, which is considered a measurement of the reliability for the density aggregation.

The choice of a total probability representation or error region density is made by the GIS expert, and the system requires to be flexible enough to allow this choice, and in which complete automation is desirable.

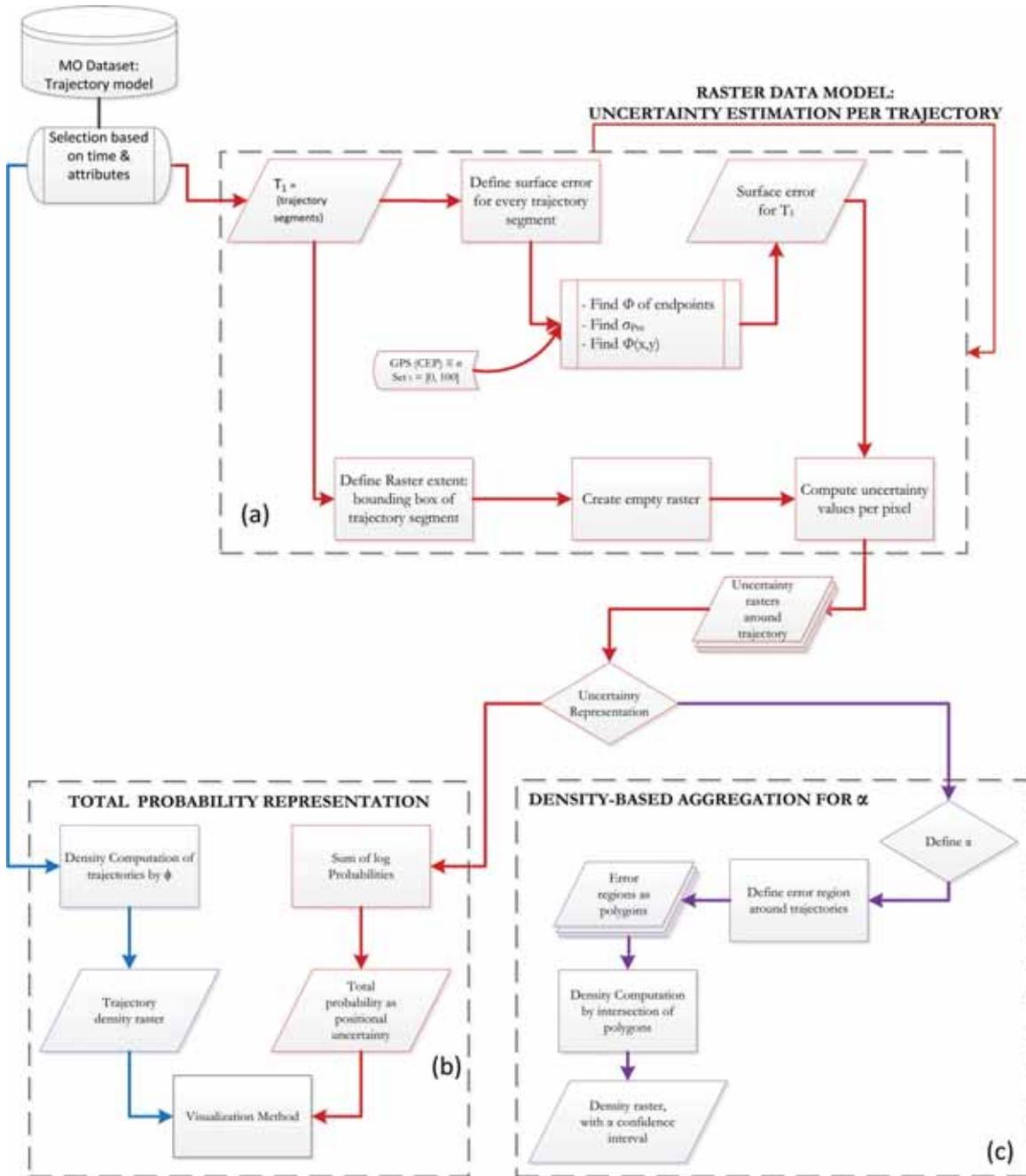


Figure 3.1: Components and information flow of the system Positional Uncertainty over Trajectories in Moving Objects (PUTMO).

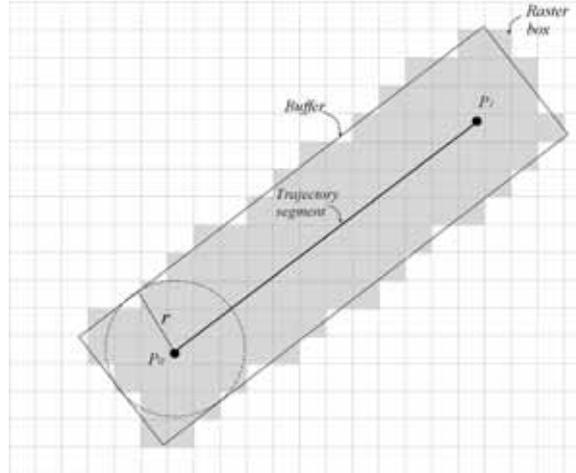


Figure 3.2: Buffer and raster box around a trajectory segment.

In the remaining parts of this chapter, the first stages of the systems are designed and the algorithms to compute the uncertainty rasters for individual trajectories are given.

3.2 POSITIONAL UNCERTAINTY ON INDIVIDUAL TRAJECTORIES

In this section we describe the algorithms used to calculate uncertainty values around individual trajectories. Positional uncertainty of a trajectory is represented by: (a) the probability of its position, and (b) the error band around it, as was explained in Section 2.5.2.

3.2.1 Box for Storing Probability Values

A trajectory segment represents the possible path followed by a moving object. This segment is represented by a straight line, and it is parameterized by: an initial point P_0 , an end point P_1 , and their corresponding times t_0 and t_1 . Besides, according to equations 2.6 and 2.7, a values for σ , μ_x , μ_y and $\frac{\sigma_P dt}{dt}$ has to be determined.

A raster model stores the probability values given by equation 2.6. Since the *pdf* has no boundaries, we limit the computation of probability by using a buffer with flat endings around the trajectory segment (See Algorithm 1). The radius r of the buffer was set to a distance determined by σ_{max} at the middle of the trajectory segment (ℓ):

$$r = Z\sigma_{max} + 0.1\sigma_{max} \quad (3.1)$$

where Z are the z-scores given for the normal distribution and considering a confidence level, e. g. $Z = 3.0$ for an $\alpha : 99.7\%$. An extra 10% is added to make the raster slightly bigger than the probability values to be computed. the value for σ_{max} is computed using a ratio in percentage $\iota \in [0, 100]$, at which σ expands in relation with the time interval Δ_t in seconds:

$$\sigma_{max} = \frac{\iota\Delta_t}{100} \quad (3.2)$$

Finally, the buffer was converted to raster. For a pixel p of size z in meters we used 1 as a default value for the pixels with its centroid inside the buffer, and 0 otherwise (considered in the present case as *no data*).

Algorithm 1: Raster Box for Trajectory Segment**Data:** Set $\{P_0, P_1\}$ such that $P_0, P_1 = (x, y, t)$.**Result:** Raster Box to bound probability computation.

```

1 begin
2    $\iota \leftarrow$  Expansion ratio  $\in [0, 100]$ ;
3    $\sigma \leftarrow$  CEP from tracking device;
4    $z \leftarrow$  pixel size;
5   if  $\iota \neq 0$  then                                     // Consider time interval in seconds.
6      $\Delta_t \leftarrow (t_1 - t_0)$ ;
7      $\sigma_{max} \leftarrow (\iota * \Delta_t) / 100 + \sigma_P$ ;
8   else
9      $\sigma_{max} \leftarrow \sigma_P$ ;
10  end
11   $r \leftarrow Z * \sigma_{max} + 0.1\sigma_{max}$ ;
12   $\ell \leftarrow$  Make line  $(P_0, P_1)$ ;
13  Box  $\leftarrow$  Make Buffer  $(\ell, r)$ ;
14  Raster Box  $\leftarrow$  Convert to Raster(Box,  $z$ );
15  if centroid of  $p$  inside Box then                   // Assign initial values to the pixels.
16     $p \leftarrow 1$ ;
17  else
18     $p \leftarrow 0$ ;
19  end
20 end

```

3.2.2 Algorithm for Error Band Along a Trajectory Segment

We approximate the probability values around a trajectory segment, by moving a bivariate normal distribution along the segments. A pure geometric and deterministic approach as the one described by Pfoser and Jessen (1999) will fail to propagate the error from the recorded position, since we model the error around a trajectory as an stochastic process (Shi, 2010). Therefore, regular or irregular error bands are not drawn directly, but they are a consequence of propagating the error under the principle described in Section 2.5.2. This requires to know μ_x , μ_y , and σ for every position around the trajectory segment that needs to be considered (See equation 2.6). These three parameters vary according to the position of the moving point, and they need to be updated when the object change its position. The work by Awasthi (2012) related with the uncertainty around a point was used in as part of Algorithm 2.

The probability value for a pixel p depends on its geographic coordinates. Let q represent the centroid of an arbitrary pixel around the trajectory segment, and q' the projection of q on the trajectory segment such that the distance $\|qq'\|$ is the closest distance. Then q' represents the mean position of the moving object, a *moving mean*. Then σ_r for an particular q' can be estimated according with equation 2.7, such that σ_r changes at a rate that shapes an error ellipse.

Error bands are approximated using a changing rate for σ_r and a quadratic function (See Figure 3.3). Here the general equation for a parabola $y = ax^2 + bx + c$ was used, where the quadratic coefficient a controls how open or close is the parabola, b and c are zero when the origin is at the vertex V . Then a is constant for a trajectory segment, and can be computed using the geometrical properties of the segment and the relative coordinates on the u and v axes. Let w_0 and w_1 be two symmetrical points on the parabola, V the vertex at the origin of an arbitrary coordinate system

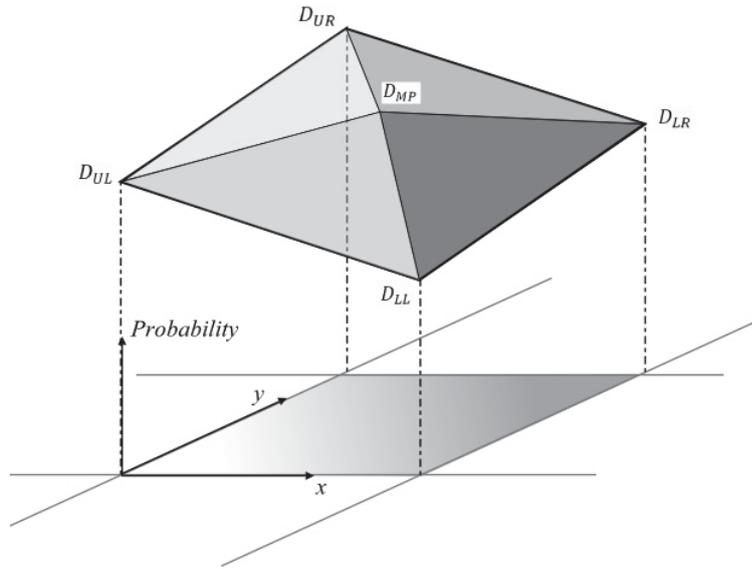


Figure 3.4: Volume over a pixel using the 'five point concept'. Adapted from Awasthi (2010).

$$p' = \frac{1}{6} 2z^2 (D_{UL} + D_{UR} + 2D_{MP} + D_{LL} + D_{LR})$$

and D are the probability values for every corner of the pixel and its centroid, computed using equation 2.6. This process is repeated for ever pixel in the raster for which the initial value for $p = 1$, and finally the probability value p' is assigned to p .

3.2.3 Algorithm Error Band for a Complete Trajectory

Algorithm 2 returns probability values that represent positional uncertainty. A raster containing these values can be generated for each individual trajectory segment by making a loop over all the segments, and the complete error band for a trajectory is composed by overlapping those rasters.

The problem of overlapping two segments was overcome using set operations. All the segments are overlapped and a union operation is computed. For pixels which actually overlap, specially around the positions that join two segments, a single value needs to be returned. In the simple case, when a trajectory does not cross itself the retrieval of *maximum* value allows to keep the continuity of the error surface over the trajectory and approximates the theoretical surface described in Section 2.5.2. This is mainly because the probability values for the end of one segment and the beginning of the next are the same, then selecting one or another is indifferent in such case since they were computed over the same point. Algorithm 3 describes the loop to compute the uncertainty rasters for all segment in a trajectory, and to produce a final output as a single raster.

3.3 POSITIONAL UNCERTAINTY IN AGGREGATION OF TRAJECTORIES

We describe the use of uncertainty raster in the aggregation of trajectories. The objective is to describe the tasks that are necessary to extend the concepts related with positional uncertainty, in relation with density-based trajectory aggregation.

The computation of line density is not described. Density aggregation in general terms were described by Silverman (1992). Various implementations has been made in GIS software (e.g.

Algorithm 2: Computing probability values for all pixels in a raster

Data: Raster Box, Line Segment

Result: Raster representing positional uncertainty as probability values

```

1 begin
2    $a \leftarrow \frac{2(\sigma_{max} - \sigma)}{\|P_0P_1\|^2};$ 
3   for  $p \in Raster\ Box$  do
4     if  $p = 1$  then // Compute moving mean
5        $q \leftarrow CentroidOf(p);$ 
6        $q' \leftarrow ClosestPoint(q);$ 
7     end
8     if  $q' = P_0$  OR  $P_1$  then // Determine standard deviation for q
9        $\sigma_{q'} \leftarrow \sigma_{P_0};$ 
10    else
11       $\sigma_{q'} \leftarrow \|P_0P_{q'}\| \cdot (\|P_0P_1\| - \|P_0P_{q'}\|);$ 
12    end
13     $D_{UL} \leftarrow \hat{\phi}_q(X_{UL}, Y_{UL});$ 
14     $D_{UR} \leftarrow \hat{\phi}_q(X_{UR}, Y_{UR});$ 
15     $D_{LL} \leftarrow \hat{\phi}_q(X_{LL}, Y_{LL});$ 
16     $D_{LR} \leftarrow \hat{\phi}_q(X_{LR}, Y_{LR});$ 
17     $D_{MP} \leftarrow \hat{\phi}_q(X_{MP}, Y_{MP});$ 
18     $p' \leftarrow \frac{1}{6}2z^2(D_{UL} + D_{UR} + 2D_{MP} + D_{LL} + D_{LR});$ 
19     $p \leftarrow p';$ 
20  end
21 end

```

Algorithm 3: Uncertainty Raster for a Trajectory

Data: Trajectory = $\{j_1 \dots j_m\}$

Result: Trajectory Uncertainty Raster S_j

```

1 begin
2    $j_i \leftarrow$  Current trajectory segment;
3    $j_m \leftarrow$  Last trajectory segment;
4    $R_j \leftarrow$  Raster for current trajectory segment;
5   while  $j_i \geq j_m$  do // Loop over trajectory segments  $R_j \dots R_m$ 
6     Compute Uncertainty Raster using Algorithm 2;
7      $R_j \leftarrow$  raster for segment  $j_i$ ;
8      $S \leftarrow R_j$ ;
9      $j_i \leftarrow j_{i+1}$ ;
10  end
11   $S \leftarrow MAX \cup_{j_1}^m R_j$ ;
12 end

```

QuantumGIS, ArcGIS). Line density is related with our system, but we prefer to focus on describing the positional uncertainty related with the outputs of that computation. Therefore, we chose ArcGIS version 10.1 to compute density and do further processing related with the aggregation of trajectories (See Section 4.3).

3.3.1 Total Probability for Overlaying Trajectories

Total probability is used to estimate positional uncertainty in aggregation of trajectories. Every uncertainty raster is treated as an independent “event”, and the total probability of all trajectories in the time window are estimated as given by equation 3.5 in its logarithmic form to reduce the computation cost and avoid to produce too small values (Brunsdon, 1995):

$$Pr(S_1 \cap S_2 \cdots \cap S_m) = \sum_{i=1}^m \frac{1}{-\log P(S_i)} \quad (3.5)$$

where m represent the number of raster involved in the operation, and the inverse of the \log values is used to maintain consistency in the interpretation: the higher the values the more likely is to find the objects in that locations. To achieve this, probability values for individual rasters are converted to a logarithmic scale divided over one. Then every corresponding pixels are sum up, which results in a new raster containing one band (R_{log}) with total probability values (See Algorithm 4). Pixels containing no data or out of the extent of the error regions are considered as ‘null’ in the final result.

Algorithm 4: Total Probability Raster

Data: Set of uncertainty rasters S_j

Result: Total probability values R_{log}

```

1 begin
2   forall the  $p_i \in S_j$  do
3     if  $p_i <> 0$  then
4        $p_{log} \leftarrow \frac{1}{-\log(p)}$ ;
5     end
6   end
7   for  $S_j \in \text{time window}$  do
8     Sum  $p_{log}$  values;
9      $R_{log} \leftarrow$  Sum of  $p_{log}$ ;
10  end
11 end
```

Optionally, probability values produced by implementing algorithm 4 can be seen as uncertainty values. This can be done by computing the uncertainty contribution Υ as given in equation 3.6:

$$\Upsilon = 1 - \frac{p_{log}}{MAX(p_{log})} \quad (3.6)$$

in which the probability value in every pixel are normalized by the maximum log probability in the computation corresponding to a time window, and this result subtracted from one.

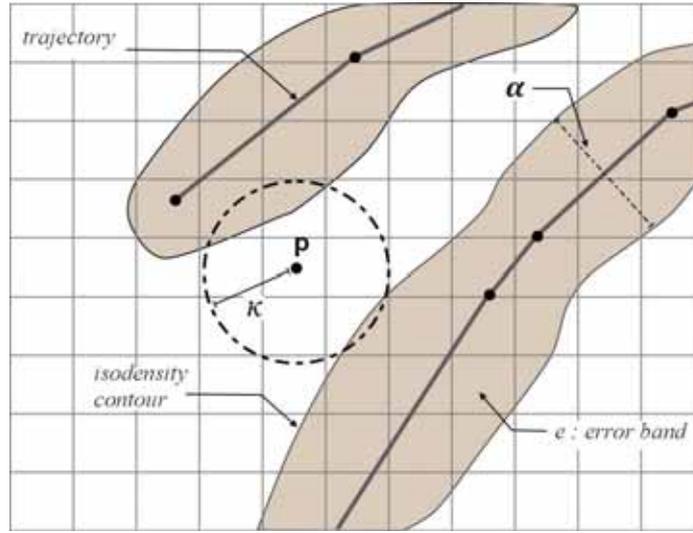


Figure 3.5: Trajectory aggregation considering error regions (e) and confidence interval (α). Error regions are drawn by computing isodensity contours.

3.3.2 Density-based Aggregation Considering Error Regions

Density aggregation can be computed using error regions. Instead of computing trajectory-density by considering only the case at which the trajectory is within a distance defined by a radius κ , density is computed using the error regions around every trajectory. Figure 3.5 illustrate the principle of the computation. For a pixel p a radius κ defines its neighbourhood, and the density is computed by counting the number of unique error regions e that intersect the neighbourhood.

Error regions are created computing isodensity contours that join equal probability values in equation 2.11. Probability values are chosen such that they represents a confidence interval for the density estimation. Only one error region per trajectory is considered.

Algorithm 5 shows the general steps for computing density aggregation based on error regions. For every pixel in the extent of the set of error regions $E = \{e_1, e_2, \dots, e_j\}$, and considering a radius κ , we center the radius in any of the pixels p and count the number of e 's that fall within the neighborhood bounded by κ , the total count is divided by the area of the circle defined by k which becomes the density value for that pixel.

Algorithm 5: Trajectory aggregation with confidence interval

Data: Set of error regions $E = \{e_1, e_2, \dots, e_j\}$

Result: Density Aggregation using Error Regions α

```

1 begin
2   foreach pixel  $p$  do
3      $\kappa \leftarrow$  radius to define neighborhood;
4     Total Count  $\leftarrow$  0 as initial value if  $e_j$  within radius  $\kappa$  then
5       Count;
6       Total Count  $\leftarrow$  Total Count + Count;
7     end
8      $p \leftarrow \frac{TotalCount}{\pi\kappa^2}$ 
9   end
10 end
```

3.4 SUMMARY

A three components system (*PUTMO*) is proposed to consider positional uncertainty in trajectory aggregation: 1.) Computation of uncertainty rasters for every trajectory; 2.) Total probability representation, which combines trajectory density outputs and probability values throughout visualization methods, and 3.) A density-based computation to aggregate trajectories by considering error regions and confidence intervals. We described the algorithms for implementing our system and depicted the general steps. In the next chapter we demonstrate the functionality of the system and the application of the theoretical foundation described in Chapter 2. We also describe the informatics tools for the implementation, and the use of the system in the cases constrained and unconstrained movement.

Chapter 4

System Implementation: Uncertainty in Different Types of Movement

The system described in the previous chapter was implemented and tested using different datasets. Details of three case studies are described in the following sections. Results in the case of unconstrained movement are presented using individual trajectories of hurricane and bird tracking data. Then we focus on constrained movement using pedestrians's tracking data, and go further to consider uncertainties in trajectory density-based aggregation.

4.1 CONCEPTUAL DATA MODEL, TOOLS AND TECHNIQUES

PostgreSQL¹ was chosen as a repository for storing and managing the datasets. Version 9.1.3 and the spatial extension PostGIS 2.0² were selected in order to handle potentially massive datasets, and the capabilities of PostGIS to support raster representation and processing. Rasters are suitable for the outputs of the uncertainty computations given by our system. Moreover, PostgreSQL can be extended by using some of the procedural language integrated in the platform. In this case, pl/pgSQL was used to code the functions needed to compute probability values and represent positional uncertainty. We used OGR2OGR³ driver to load GPX files directly into PostgreSQL. A batch file using DOS was created to automatically load and compile tables that contains all the positions for all the trajectories in the datasets (See Appendix D).

The database schema in Appendix B was implemented in PostgreSQL. The schema allows to store trajectories as a sequence of recorded positions. A *Track* is created by joining at least two *positions* using a linear function. Every position is spatio-temporal, which means that date and time are recorded as well as its spatial location. Every *Track* represents a *Moving Object* which can be a *Hurricane*, a *Pedestrian*, or a *Bird*, or can be extended to other moving objects. Pedestrians move with specific purposes: shopping, tourist, leisure or other. An *Uncertainty Raster* is produced by applying our uncertainty model to each of one of the tracks. Finally, an *uncertainty aggregation* is computed by combining the probability values of one or more uncertainty rasters. Positions and tracks are stored using the OGC Simple Feature specifications (OGC, 2010) for a *point* and a *linestring*. Meanwhile, we chose a raster model to store probability values related with the positional uncertainties of our system. This makes possible to perform uncertainty analysis over trajectories constructed using time windows and add thematic data as selection criteria. Other approaches exist for spatio-temporal data models and database design (Pelekis, Theodoulidis, Kopanakis, & Theodoridis, 2004), but we chose a simple one as our focus is not the database design itself.

We developed a set of functions to compute probability values around trajectories. The functions take as arguments: a set $\hat{\tau}$ which defines a trajectory as a collection of spatio-temporal position, σ as the error given by the location device and constant for every position in the trajectory,

¹<http://www.postgresql.org>

²<http://postgis.org>

³<http://www.gdal.org/ogr2ogr.html>

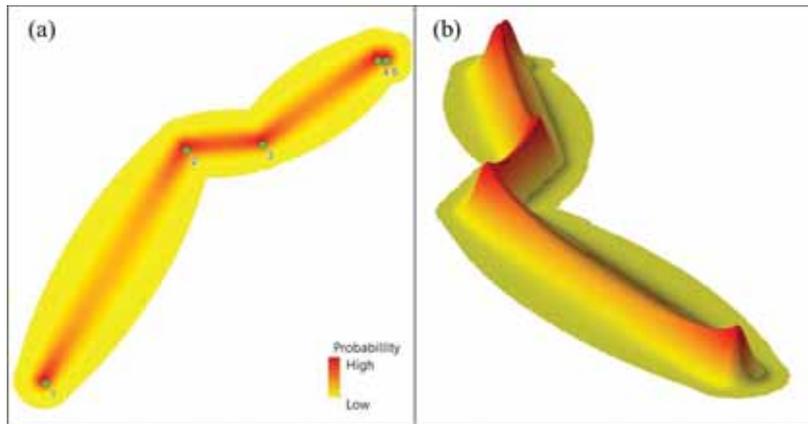


Figure 4.1: Result with a synthetic trajectories. (a) Error band, and (b) Error surface.

σ_{max} a trajectory segment dependent value, computed using equation 3.2, ι a ratio at which the error changes from σ to σ_{max} , and $pixel\ size$ which control the resolution of the uncertainty raster. A function three and the source code for the functions are given in Appendices C and E.

For visualization purposes Quantum GIS⁴ version 1.8.0 was used, and the plugin “Load Post-GIS Raster to QGIS” version 0.5.4 developed by (de Paulo, 2012) helped to render the raster data type into the Quantum GIS interface. For more advance visualization methods we relied on ArcGIS⁵ version 10.1.

We tested our implementation using a synthetic trajectory. The trajectory contains 5 positions and 4 segments (See table 4.1). Figure 4.1 shows the result for the error bands and error surfaces out of the implementation. The error band along the first segment is long enough to draw an ellipse error. Meanwhile, for the second segment the elliptical shape is less evident. This behavior is influenced by the size of the time interval between the positions and the distance traveled by the object in that time. Probability values reflect the level of uncertainty in the position of the object along the complete trajectory. The lower the probability values, the more uncertainty in the position of the object. Uncertainty is always less over the positions recorded by the tracking device.

Table 4.1: Details of the synthetic trajectory used during the testing of the implementation.

<i>POSITION</i>	<i>UTC TIME</i>	<i>X-COORD</i>	<i>Y-COORD</i>
1	2009-11-19 11:00:00+01	84380	447390
2	2009-11-19 11:01:40+01	84430	447476
3	2009-11-19 11:02:05+01	84457	447478
4	2009-11-19 11:03:02+01	84498	447509
5	2009-11-19 11:03:04+01	84501	447509

4.2 UNCERTAINTY IN UNCONSTRIANT MOVEMENT

Uncertainty estimations for unconstrained movement were computed using hurricane and wildlife tracking data. Hurricane tracking data was provided by UNISYS (2012) and collected by the US

⁴<http://www.qgis.org>

⁵<http://www.esri.com/software/arcgis>

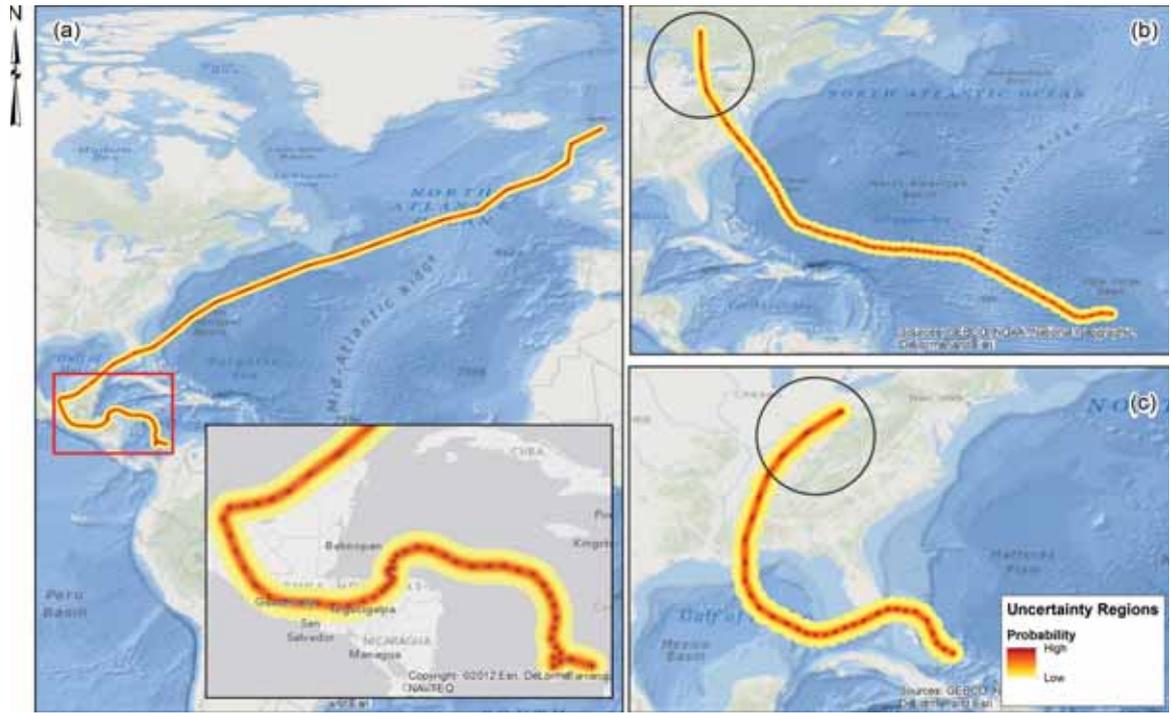


Figure 4.2: Error bands and probability values for hurricane data. (a) MITCH 1998, $\iota = 30\%$; (b) ISABEL 2003, $\iota = 50\%$; (c) KATRINA 2005, $\iota = 50\%$. Circles (in (b) and (c)) mark areas where irregular error bands are more evident.

National Weather Service. Wildlife tracking data collected on the first day of observations made by Holland et al. (2009) while tracking 'oilbirds' (*Steatornis caripensis*) was used⁶. Relevant details about both datasets are given in Table 4.2.

Table 4.2: Details of unconstrained moving object datasets.

DATA SET	SOURCE	COLLECTION YEAR	SAMPLE RATE	OBSERVATIONS
Hurricanes	US National Weather Service	1990 to 2006	6 hours	Time Zone was set to +00
Oilbirds	Movebank	2007	10/15 minutes	Time Zone was set to -04

Uncertainty along hurricane paths resulted in regular error bands in most of the cases. Here a $\sigma = 20 \text{ km}$, pixel size of 2 km and two different values for ι (30%, 50%) were used (See Figure 4.2. In our implementation, ι or *sigma ratio* is a value in percentage which determines how fast the error bands expand with respect to the time interval between positions. Even while change ι to match possible scenarios, results tended to produce regular error bands. Only when $\iota = 50\%$ and the distance between consecutive positions is large enough in comparison with the time interval, the error region takes the shape of an ellipse (circle marked areas in Figure 4.2 (b) and (c)).

In contrast, irregular error bands described uncertainty during *oilbirds* flights. The sampling rate of the dataset is 10 min. A conservative value of $\sigma = 100 \text{ m}$ was used (CLS, 2011) and ι was set to 40%, to consider the high level of maneuverability of the birds during their flight. When considering probability values and confidence intervals (CI), Elliptical error regions along the trajectories appear at 99% confidence interval in this case (Figure 4.3). Probability values are

⁶Available at <https://www.datarepository.movebank.org/handle/10255/move.270>

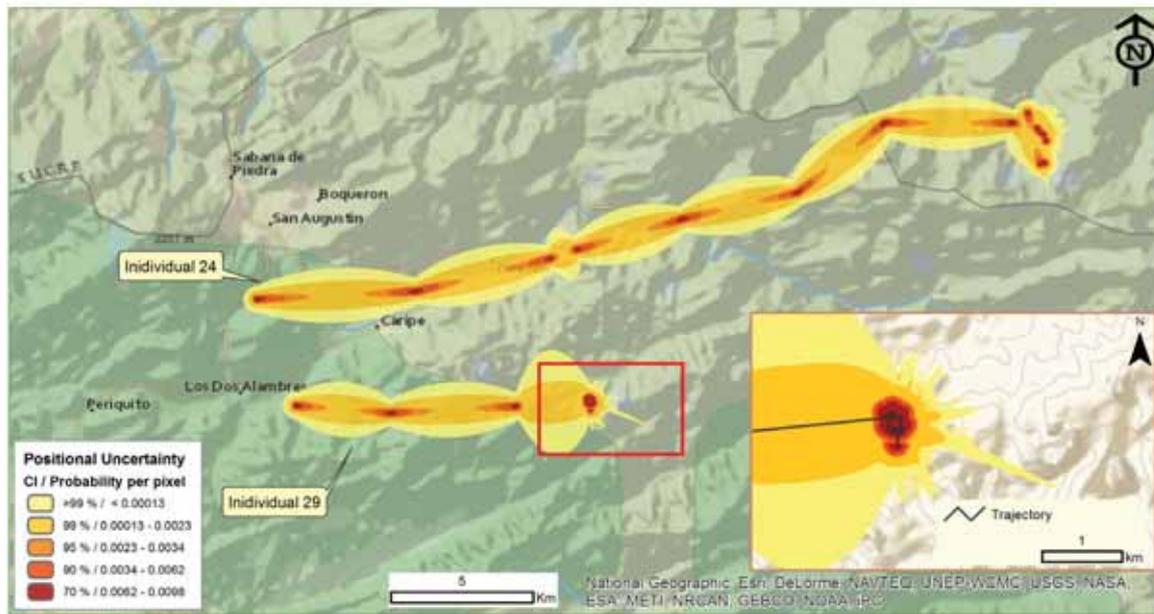


Figure 4.3: Error regions for oilbirds' flight around the Cueva del Guacharo National Park, Venezuela, 2007. Individuals 24 and 29. Artifacts are enlarge in the detailed inset.

given as the lower and upper limits of the confidence intervals, those are the values per pixel that define the limits of the error bands showed on the map. Strange artifacts appear when the distance traveled for the object between consecutive positions is small compared to the time interval, and errors bands are amorphous when trajectories cross or overlap.

4.3 UNCERTAINTY IN CONSTRAINED MOVEMENT

Which are the routes predominantly used by the visitors to the city center of Delft? This is one of the questions that motivated the collection of pedestrians data in Delft, in 2009 (van Lagelaar & van der Speck, 2010). In order to evaluate the experience of visitors to the city center and observe the use of the public space. Results for this research project are presented by Baltus et al. (2010) and van Lagelaar and van der Speck (2010). Here we extend the question and add a quality statement to it: Which are the routes predominantly used by the visitors to the city center, and what is its reliability? The first part of the question has been answered by applying simple (van Lagelaar & van der Speck, 2010) and complex (Scheepens et al., 2012) density techniques, which aggregate trajectories based on space, time and attributes.

By implementing our system and using the same dataset as in the previous works, our purpose is to add a reliability indicator to the answer derived from the density estimations of the paths followed by the visitors to Delft City.

4.3.1 Pedestrian's Dataset

In this case study the dataset *Tracking Delft 1* was used (Baltus et al., 2010). It contains 328 records of visitors to the center of Delft city, The Netherlands. It was collected over 4 days from November 18th to 21st, 2009. It includes individual GPS tracks in GPX format, and a MS Access database containing attribute information about the participants (visitors). Participants in the experiment

were pedestrians that came to visit the city for various purposes: shopping, tourism, leisure, among others.

Tracks were collected using GPS devices at a sample rate of 2 seconds. Devices were distributed to participant at two parking lots: *Phoenix and Zuidpoort*, and every participant filled a questionnaire which provided attributive information to every track (Baltus et al., 2010). Relevant technical specifications for the GPS device that was used in the data collection are provided in table 4.3 (Qstarz International Co., Ltd., 2012).

Table 4.3: GPS technical specification. Tracking Delft I dataset.

MODEL:	Qstarz BT-Q1000x(AGPS)
<i>GENERAL SPECIFICATIONS:</i>	<i>POSITIONAL ACCURACY</i>
Frequency: L1, 1575.42 MHz	2D-RMS (without aid): 3.0 m
C/A Code: 1.023 MHz chip rate	CEP (50%) without SA(horizontal) < 3.0 m
Channels: 66 CH performance tracking	DGPS (WAAS, EGNOS, MSAS): 2.5 m

Individual tracks were verified for inconsistency and classified as valid tracks. Verification was based on matching the tracks with the questionnaire information, readability and consistency of the track. Invalid points were removed or adjusted to produce a valid track (Baltus et al., 2010). This manipulation has an effect on the uncertainty of the movement as is mentioned below.

4.3.2 Positional Uncertainty in the Paths followed

When uncertainty computation is applied to a pedestrian’s track, it draws the error regions along the paths (See figure 4.4 (a)). In the case of a high sampling rate (one position every two seconds), the union of all error regions of the trajectory segment result in a regular error band.

Uncertainty in the movement may increase as a consequence of data cleaning. This increases the time and consequently the space of possible movement for a pedestrian, and therefore increases the uncertainty of the position of the pedestrian along the trajectory segment. In the extreme case when too many consecutive positions are removed and the remaining ones are relatively closed from each other, the error band expands abruptly and appears as peaks in perpendicular direction to the corresponding trajectory segment. This is, the evidence left by cleaning the data or other sources of errors, like signal loss. In Figure 4.4 (b), uncertainty in the position of the pedestrian is higher in the lighter areas, because the probability that pedestrian have visited that place is lower.

An increment in the time interval between the previous and the next known positions, potentially increases the distance between them. In that period of time the current position of the object is unknown, therefore less certain, in contrast to what the trajectory model assumes. If we have no evidence of where or why positions were removed, then the space of possible movement expands to the sides. This creates an ellipse error band inside which the position of the pedestrians is less certain as the probability of the pedestrian to have taken one specific path spreads out over the error band.

We compared the effect of cleaning the data by computing the error bands of a portion of the trip of pedestrian with ID: 101_101. A subset of 307 original positions was selected, and then compared to the same portion of the trip after Baltus et al. (2010) cleaned it applying the following criteria: date correctness, correspondence of the data with the thematic data provided by the participant, and reliability and consistency of the dataset (which is not explained how it was determined). In the cleaning process, the dataset was reduced to 78 positions: 139 consecutive positions were removed from the lower left corner of the trajectory, and another 90 in the middle of the trip. The time interval between the previous and the next positions increased up to 3

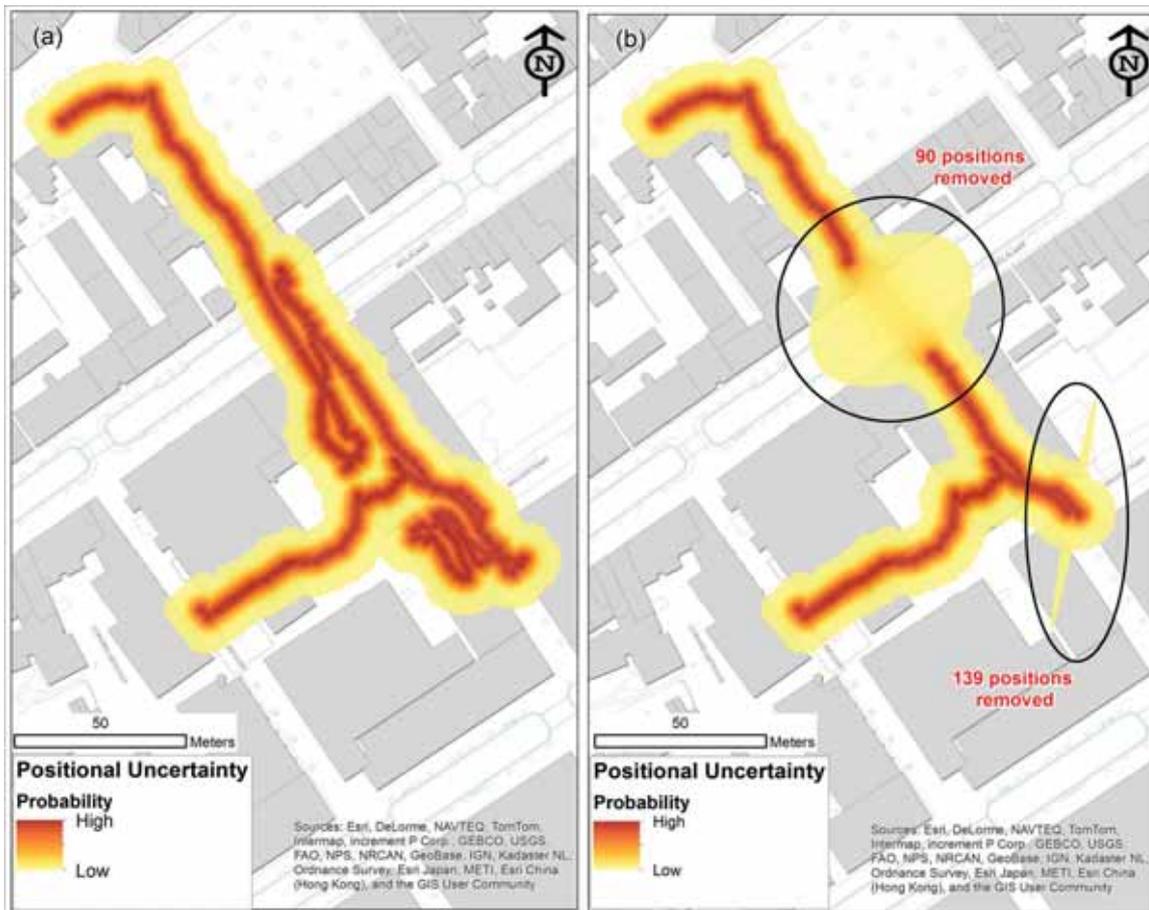


Figure 4.4: Effect of data cleaning in the error bands and positional uncertainty in pedestrian's track. (a) Original dataset with 307 positions. (b) Cleaned dataset with 78 positions

minutes as a consequence of the latter case. Error bands for both cases are presented in Figure 4.4. As mentioned above, the evidence of the cleaning process on the trajectory model is effectively represented by the error bands and probability values along the trajectory. Error regions does not necessarily contain the position that were removed. In those bands, the position of the pedestrian is more uncertain.

In both case presented in Figure 4.4 error regions expand beyond the limits of the street network. This suggest the possible path include the area inside the surrounding buildings, this might result improbable. Another situation is when higher probability values for the position of the pedestrian fall outside outside the street network, possible as a consequence of gross or systematic error in the collecting device. In cases like those the contextual information becomes essential to identify the most probable paths followed.

4.3.3 Positional Uncertainty over time.

What is the uncertainty of the places visited for shopping, on November the 19th? To answer this question we selected 29 pedestrians which declared shopping as main purpose and arrived at *Zuidpoort* parking lot, between 12 : 00 - 14 : 00 hours. From that selection we divided the group in two different time windows of 1 hour each one. During the first time window (12 : 00 – 13 : 00 hours),

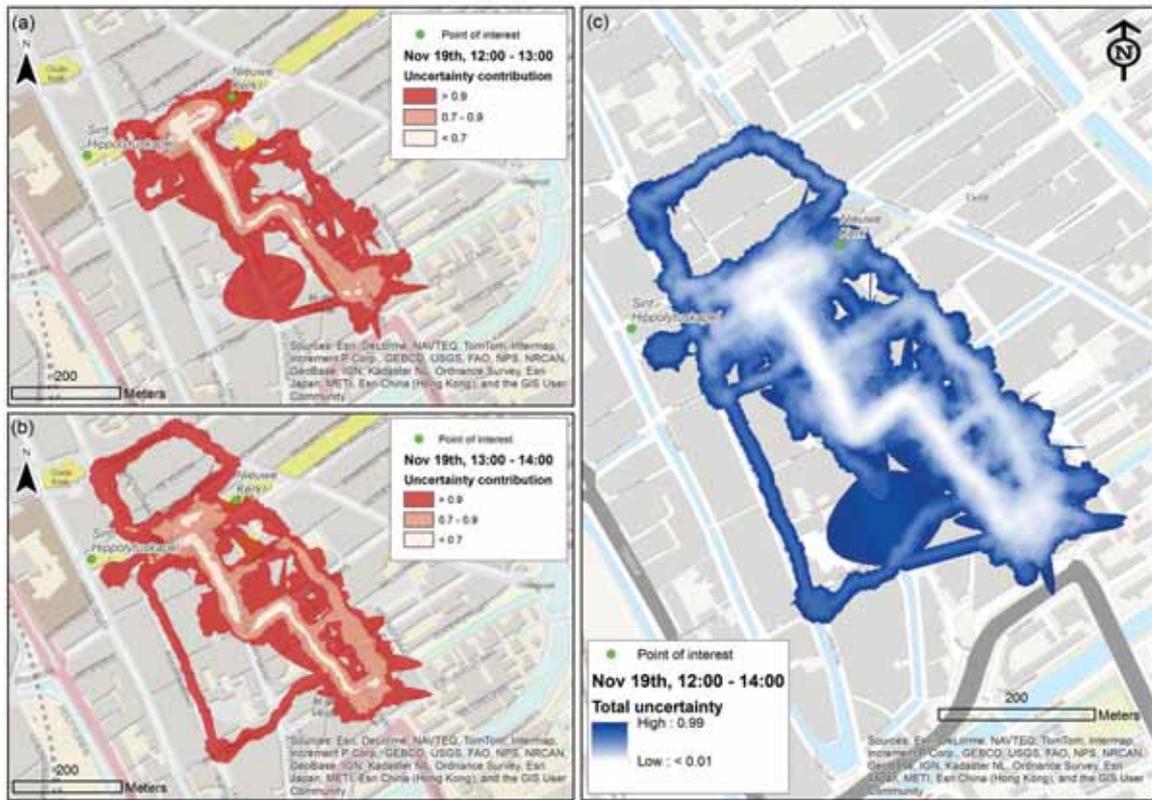


Figure 4.5: Uncertainty contributions in two different time windows. Pedestrians shopping on November 19th, 2009

23 pedestrians were present in the city center. Meanwhile, during the second time window (13 : 00 – 14 : 00 *hours*), 25 pedestrians were there, but not necessarily different from the ones in the first time window. Uncertainty contribution was computed by normalizing the probability values in each time window using equation 3.6, and considering the complete time interval 12 : 00 – 14 : 00 *hours*.

Uncertainty contribution varies spatially between time windows. The contribution is less around the *Nieuwe Kerk* in the first time window, and it is higher close to *Sint Hippolytuskapel* in the second time window. Contribution along the main shopping area (in light colors in Figure 4.5) is similar in both cases. Differences are given by the time interval we look at, and mainly because different pedestrians are considered, and pedestrians in both time windows are expected to have visited different places. Therefore, we say that in the aggregation of trajectories the level of uncertainty and its spatial distribution in the final output depends on the time window, as the time window determines the data used as input.

When time windows are used as selection criteria in the uncertainty analysis, this conditions which and how many trajectories are included in the aggregation process. Which according with the factors described in section 2.4, those have an influence in the level of uncertainty given in the final output.

4.3.4 Positional Uncertainty in Density Aggregation

Which are the most visited places by tourist? This question can be answer by computing the sum of probabilities as given by equation 3.5. We selected a subset of trajectories which represent 8 tourists visiting the Center of Delft, on 20th of November, 2009. Due to performance limitations in our implementation a relatively low resolution was used (1 m). High probability values show the areas where it is more likely to find the tourists, Figure 4.6 (a). Figure 4.6 (b) presents the common places visited by all tourists (8 tourists) with a confidence interval of 99%. And (c) presents the areas that were visited by at least 4 tourist, for both 95% and 95% confidence intervals (CI). Areas representing confidence intervals were identified by computing the inverse of equation 3.5, such that the probability value that bounds the area for a specific confidence interval is given by $[N / -\log(p_{CI})]$. p_{CI} is the probability value for the fartherst pixel in the case of positional uncertainty around a point. This value can be obtained form our system, by providing the same values for σ and pixel size used in the total probability computation. N is the number of trajectories (pedestrians is our case) are considered. N cannot be more than the number of trajectories considered in the total probability computation. Probability intervals showed in Figure 4.6 (b) and (c) were obtained by using the results of the formula presented here, as lower limits to the corresponding confidence intervals. An ultimate interpretation of the map in Figure 4.6 is that with the given confidence interval we can be sure that a number of tourist were in those location. This interpretation does not only give a measure of reability to the aggregation of tourist to identify common places, but also allow as to ask how many tourists were present in those places.

Oude Kerk and *Nieuwe Kerk* were common places visited by 8 tourists, with the confidence interval of 99% (See Figure 4.6 (b)). This is suggested by the size of the regions (in red) around these two places. One of the biggest regions is closed to *Zuidpoort*, this is explained by the fact that 7 out of 8 of the tourists arrived to the city center using *Zuidpoort* parking lot. Same explanation applies for the region within 95% confidence interval. Regions increase in size as we consider the case of the places visited for at least 4 tourists, Figure 4.6 (c), since more combinations of four or more tourists are possible.

We cannot draw conclusion about the tourist's preferences with such limited sample, as the present case serves as illustration of our implementation in the trajectory-based aggregation context, and it is not relevant for this study.

4.4 SUMMARY

The applicability of the method was demonstrated by applying it to datasets that represent constrained and unconstrained movement. Unconstrained movement is represented by Hurricane and birds tracking data. Hurricane data produced regular error bands along the trajectories, in most of the cases. Irregular bands are more evident for the oilbirds' tracks, this is influenced by a low sampling rate and a high sigma ratio (ι). Application in the case of constrained movement was illustrated by using pedestrian's tracks in the city of Delft. We have demonstrated the effect of data cleaning in the positional uncertainty represented by trajectories, the difference of the level of uncertainty in different time windows, and briefly discussed the limitations of our method in the constrained case and the importance of contextual information in such cases. The applicability of our implementation was demonstrated by computing and representing areas visited by tourist using confidence intervals as a measurement of reliability.

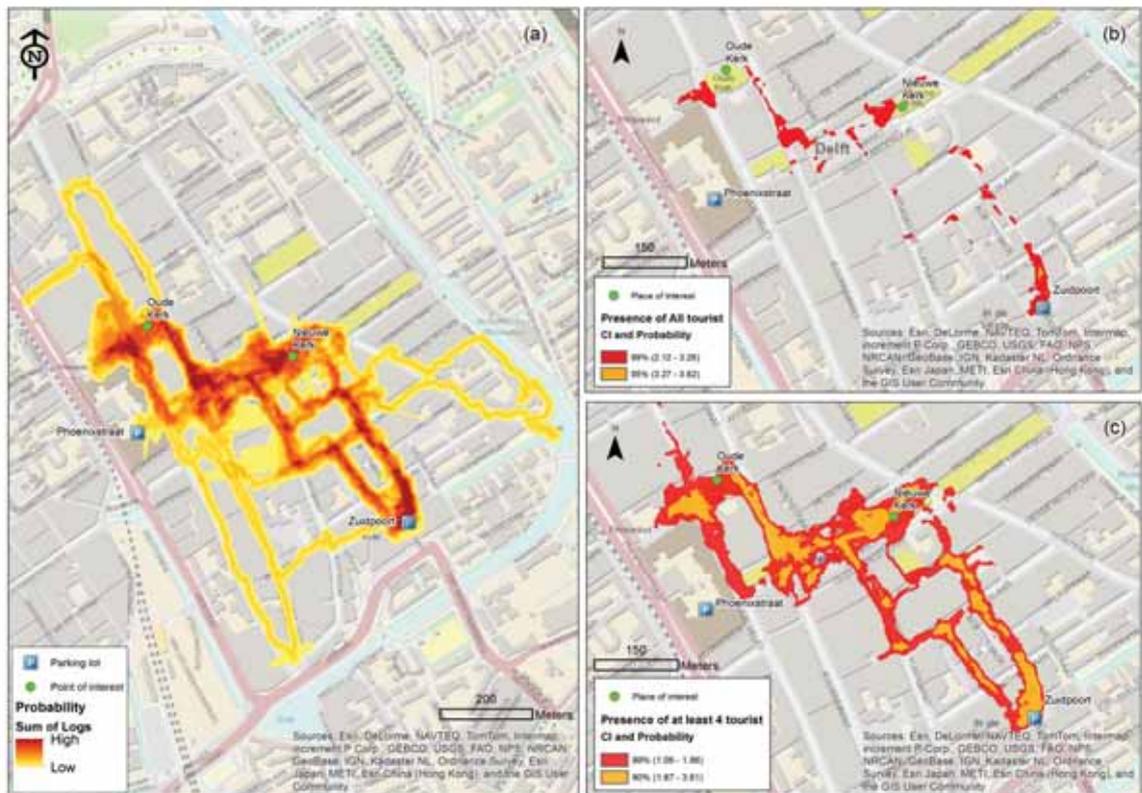


Figure 4.6: Probability values of the paths followed by tourists on November 20th, 2009. (a) Sum of probabilities, (b) Confidence intervals for places visited by all tourists, and (c) Confidence intervals for places visited by at least 4 tourists.

Chapter 5

Use of Visualizations of Uncertainty as a Quality Indicator

Uncertainty visualization helps to communication the quality of spatio-temporal aggregation. In the following sections, we apply several visualization methods to the outputs of aggregation and uncertainty measurement. Recommendations given by Senaratne and Gerharz (2011) and Senaratne, Gerharz, Pebesma, and Schwering (2012) were followed to explore the applicability of visualization methods to spatial uncertainties related with moving object datasets. In the last section, we apply a 3D representation to exemplify the use of uncertainty values in the field of Geovisual analytics.

5.1 THE PURPOSE OF VISUALIZATION IN THE GEOGRAPHIC CONTEXT

In the context of GIScience, visualization of geographic information has four main fields. Those fields are identified by Kraak and Ormeling (2010) and known as: exploration, analysis, synthesis and presentation. Exploration methods are used when the main goal is to understand the datasets. Analysis methods allow to relate different datasets by, for instance, overlaying one over another. Synthesis and presentation methods are used to summarize the results of an analysis and communicate geoinformation.

Limitations of the visualization techniques and the levels of cognition make communication of uncertainties difficult (Goodchild, 2009). To overcome some limitations advanced visualization methods, e.g. animation, has been applied, but there is not a visualization method that overcome all limitations. Then, the choice of the visualization method has to be made to meet one or at most two of the purposes mentioned above. Here, we focus on the synthesis and presentation purposes of geovisualization, and leave the remaining purposes as alternatives leading future research.

5.2 APPLICATION OF VISUALIZATION OF UNCERTAINTIES OF DENSITY-BASED AGGREGATION OF TRAJECTORIES

In order to choose appropriate methods to visualize spatial uncertainty, Senaratne and Gerharz (2011) propose a framework to select visualization techniques based on: data type, uncertainty type, and data format. See Figure 5.1.

Adjacent maps, symbols and contouring are effective static visualization methods to communicate spatial uncertainties. This, according to the usability test by Senaratne et al. (2012), the majority of the users (with a background in statistics, GIS, visualization, amongst others) recognized those visualization methods as the ones that were easier to interpret. Adjacent maps and contouring are static visualization methods that meet our purposes, after considering the selection criteria depicted in Figure 5.1, the usability test, and the characteristics of the output of our uncertainty estimation. We consider the outputs of our system PUTMO: density aggregation and probability/uncertainty estimations as continuous data type in raster format and representing

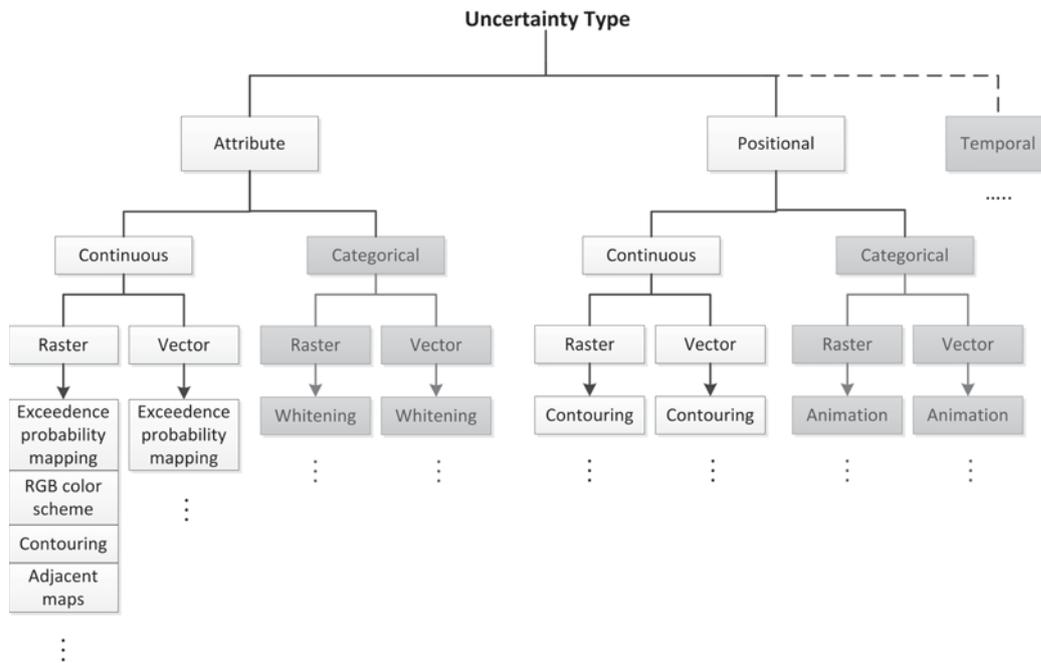


Figure 5.1: Categorization of uncertainty visualization methods. Only relevant static methods are highlighted. Adapted from Senaratne et al (2012).

positional uncertainty.

5.2.1 Visualizing Uncertainties as Adjacent Maps

Adjacent maps are useful to communicate continuous, attributive uncertainty in a raster format (Senaratne et al., 2012). In this method, two maps are presented next to each other, one representing the results of the computation (density aggregation in our case), and another representing the uncertainty values related with that computation. This is the case when for instance posing the question: *How certain are we that the Delft Market had a high density of pedestrians on November 19th, 2009 between 12:00 and 14:00 hours?* Figure 5.2 shows the adjacent maps for both density of pedestrians and uncertainty computations.

The area that corresponds to the *Market* presents a low uncertainty value. This means that that we can be certain about what the aggregation density shows. The *Market* was a place that most to the pedestrians visited for shopping on November 19th, 2009, and we can be more certain about this conclusion than in other cases, e.g. that they were shopping in the surroundings of *Sint Hippolytuskapel*.

5.2.2 Visualizing Uncertainties as Contour lines

Visualization using contouring is a static representation for positional uncertainty. This is suitable to represent continuous data type, in which case the uncertainty values are converted into contour lines and the thickness of the lines are proportional to the level of uncertainty (Senaratne et al., 2012).

Figure 5.3 shows contour uncertainty lines over density estimation. Density corresponds to tourists visiting the city center on November 20th, 2009. We represent uncertainty as contour lines. The type of question posted for this case is *how uncertain is the presence of tourist around Waalse Kerk?* An absolute answer is difficult to get even when using this visualization method.

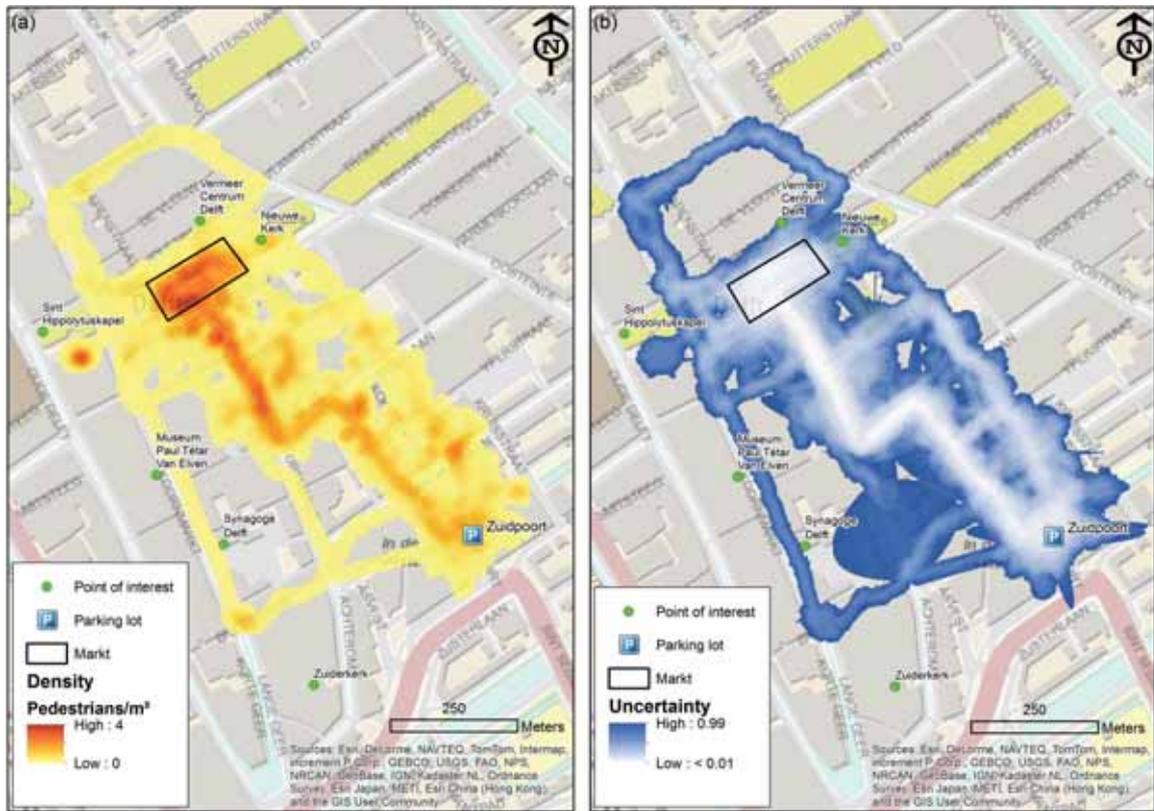


Figure 5.2: Adjacent maps. (a) Density of pedestrians shopping on November 19th, 2009 and (b) the corresponding spatial uncertainty.

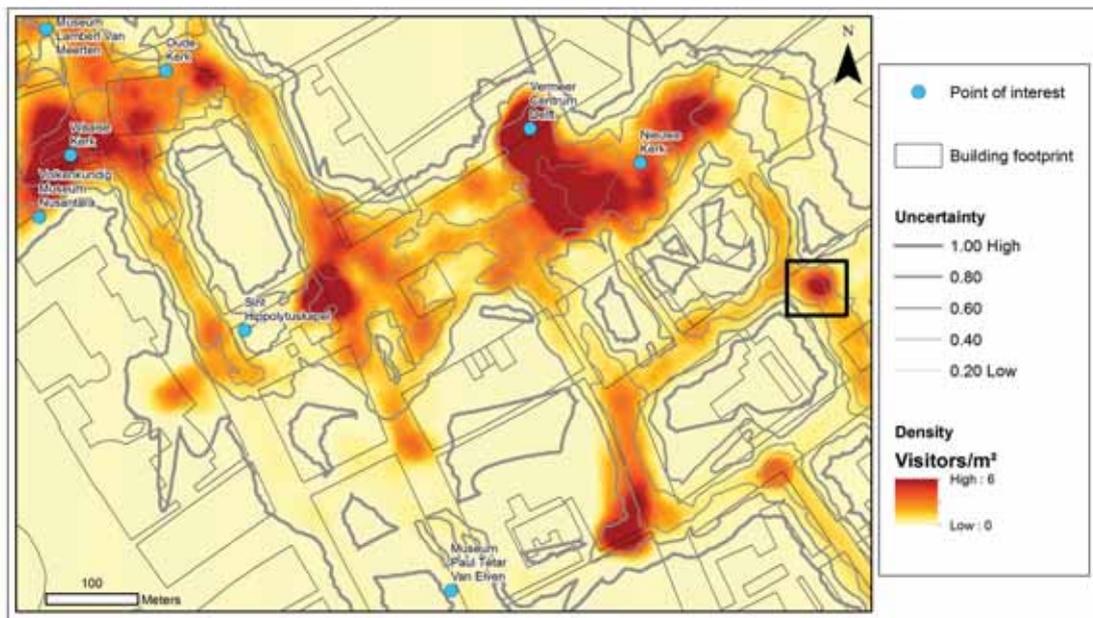


Figure 5.3: Density values and uncertainties represented as contour lines. Building footprints and 'points of interest' were taken from OpenStreetMap

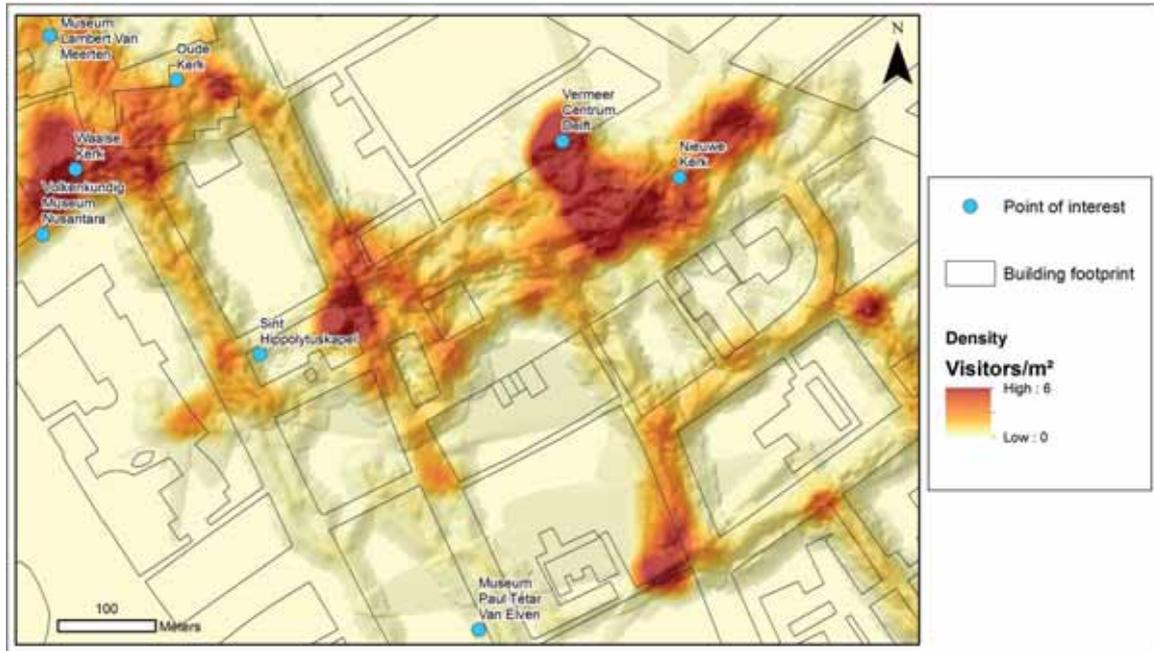


Figure 5.4: Visualization of Density of tourist and Positional probability values using color and shading. Building footprints and 'points of interest' were taken from OpenStreetMap

This is because the area of interest include contour lines that represent different levels of uncertainty, and only a general idea of the level of uncertainty is possible. Another question could be: *Where the Tourist Office should post more information to attract tourist to points of interest depicted on the map?* An answer is the place marked with a black box on the map. This place presents a high density value and a low uncertainty level.

5.2.3 Color and Shading Visualizations

Color and shading on maps are visual variable used extensively to represent of moving objects datasets (e.g., Willems et al., 2009; Scheepens, 2010; Scheepens, Willems, Van De Wetering, & Van Wijk, 2011; Scheepens et al., 2012). The map in Figure 5.4 presents density values as number of tourist per square meter using color and value. Meanwhile, the probability values are shown as shadows that resemble a relief surface in the case of DTM's. Density values are easy to understand, but the heterogeneous nature of the probability surface makes difficult to distinguish the small variations.

Interpretation becomes difficult when representing positional probabilities as shading. Even when it was successfully applied in cases mentions above, in our case this methods seems to fail due to behavior implicit in the movement in the case of pedestrians (erratic and highly unpredictable in speed and direction). The application to unconstraint movement dataset or to individual trajectories could give different results.

The map presented in Figure 5.5 combines contour lines and color and shading. This combination improves the visualization of relative big differences in the error surface. Small differences are still difficult to visualize and the addition of contours could make difficult to interpret density values.

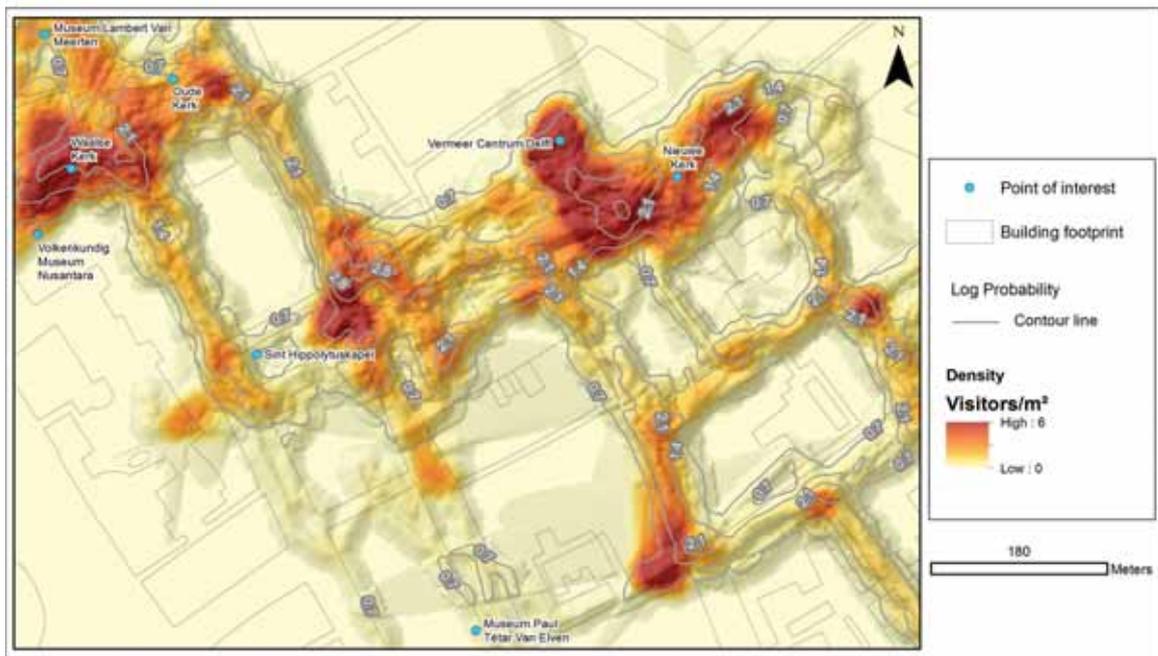


Figure 5.5: Visualization of Density of tourist and Positional probability values. Color and value representing densities and shading and contours representing probabilities. Building footprints and 'points of interest' were taken from OpenStreetMap

5.3 APPLICATION OF VISUALIZATIONS OF UNCERTAINTIES IN GEO-VISUAL ANALYTICS

Geovisual analytics combine visualization techniques and analysis tools to explore and derive knowledge from geographic datasets. It relies on interactive visualization environments and automated analysis to access and manipulate data, and facilitate reasoning and knowledge construction (Keim, 2010; Kraak & Ormeling, 2010). Combination of multiple geographic variables in static representation always has limitations. One major limitation is the level of interaction related with them. Previous subsections were limited to static visualization method from Senaratne and Gerharz (2011). Limitations can be overcome by adding interaction and/or using dynamic variables, e.g. display time and order to produce animations, (Kraak & Ormeling, 2010). Those types of visualizations comply with an exploratory purpose.

A 3D environment overcomes some of the limitations of the color and shading method (See Figure 5.6), and improves the visualization of probability values. This combined with a dynamic environment in which interaction options like filtering over the density and uncertainty values are enabled, can be used in the context of geovisual analytics. Interaction would allow to explore the dataset and realize about subtle changes in the probability values. For the present case, that not always high density values are associated with high probabilities 5.6 (a) and (b). And in the same time, makes us pose questions to understand the reason of that behavior, questions that are left as part of future work.

5.4 SUMMARY

The purpose of visualization of uncertainties in the geographical context is to facilitate the communication of the quality of raw or processed datasets. We have selected and applied three vi-

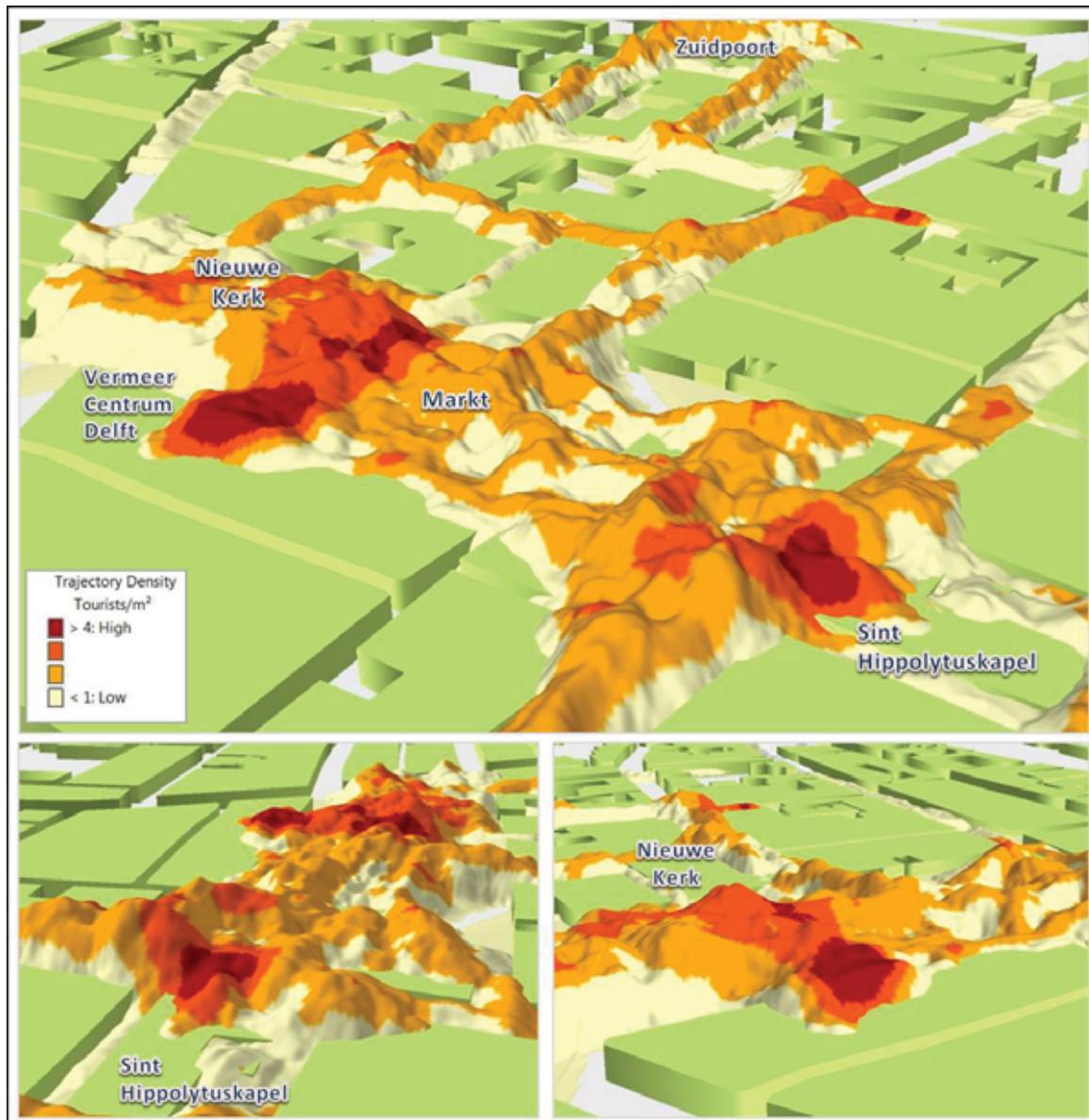


Figure 5.6: 3D representation of density of tourists, November 20th, 2009. Color represents density values and height represents probability values. (a) High density values and low probabilities close to *Sint Hippolytuskapel*. (b) Relative high density values and high probabilities around *Nieuwe Kerk*. Buildings from OpenStreetMap are modeled as 3D boxes

sualization methods to represent positional uncertainties of aggregation of trajectories. Adjacent maps are useful to communicate uncertainties related to specific places (e.g. Delft Markt). The use of contour lines facilitates the representation of positional uncertainty as a continuous surface. Color and shading proved to be less effective in visualizing density-based aggregation value in combination for positional probabilities. A slight improvement in visualization was obtained by using both shading and contour lines to represent the error surface. We also introduced the use of uncertainties of moving objects in the context of Geovisual analytics. A 3D representation of the outputs of our system overcomes some of the limitations of 2D representations and gives new insight.

Chapter 6

Discussion, Conclusions and Recommendations

In this chapter we discuss relevant points related with the system, and other approaches for both uncertainty in moving object datasets and trajectory-base aggregation. We conclude over the objectives of this research and give some guidelines for future research work.

6.1 DISCUSSION ON THE UNCERTAINTY REPRESENTATION IN AGGREGATION OF TRAJECTORIES

Uncertainty in moving object datasets is affected for several factors. Spatial and temporal uncertainties are of relevant interest as the main purpose of analyzing those dataset is to described their movement. In the present work we have limited our research to the spatial component and mainly addressed the modeling and computation of positional uncertainties. Nevertheless, we have described time as a factor that influence the level of positional uncertainties, and a key factor that controls the results of our computations.

Regular or irregular error bands around trajectories depend on the level of maneuverability of the moving object. In PUTMO the shape of band error is controlled by the value assigned to sigma ratio (ι). A sensible value for ι should be chosen accordingly with the type of movement. Figures 4.2 and 4.3 show that different values for ι has different result when the uncertainty computation is applied to different datasets. Therefore the value of ι is type of movement dependent, ad a good estimation should be made, and do not considered the choice of this value as an arbitrary task. The ideas by Pfoser and Jessen (1999) described in Section 2.5.2 makes possible to approximate a value for ι , as follows:

$$\iota = (1 - [(\frac{2a - 2c}{2}) / \frac{2b}{2}]) * 100$$

For the case of constrained movement the estimation of ι can be limited to the space that is available for movement, and the choice of a correct sigma ratio becomes less relevant, as in principle the object can not move out of the spatial network, a extreme case of this is a train moving on the railway.

In density aggregation of moving objects: time window, tessellation and radius are pointed as factor affecting positional uncertainty. Uncertainty contributions for the time window presented in Figure 4.5 shows the choice of time window for computation influences the level of uncertainty in the final output (density-based aggregation). The levels of uncertainty varies in value and location for different time windows as they include different objects in the computation.

Effect of radius and pixel size in the density computation seems not to be so relevant. We have pointed the choices of *radius* and *pixel size* in the density computation, as factors that increase the positional uncertainty of trajectory density-based aggregation method. Figure 6.1 shows that the effect of varying the size of the radius is mainly related with the size of the areas considered as 'hot spots'. A bigger radius takes into account trajectories that are farther apart, this makes the hot spots to spread out. Reduction of the radius has the opposite effects.

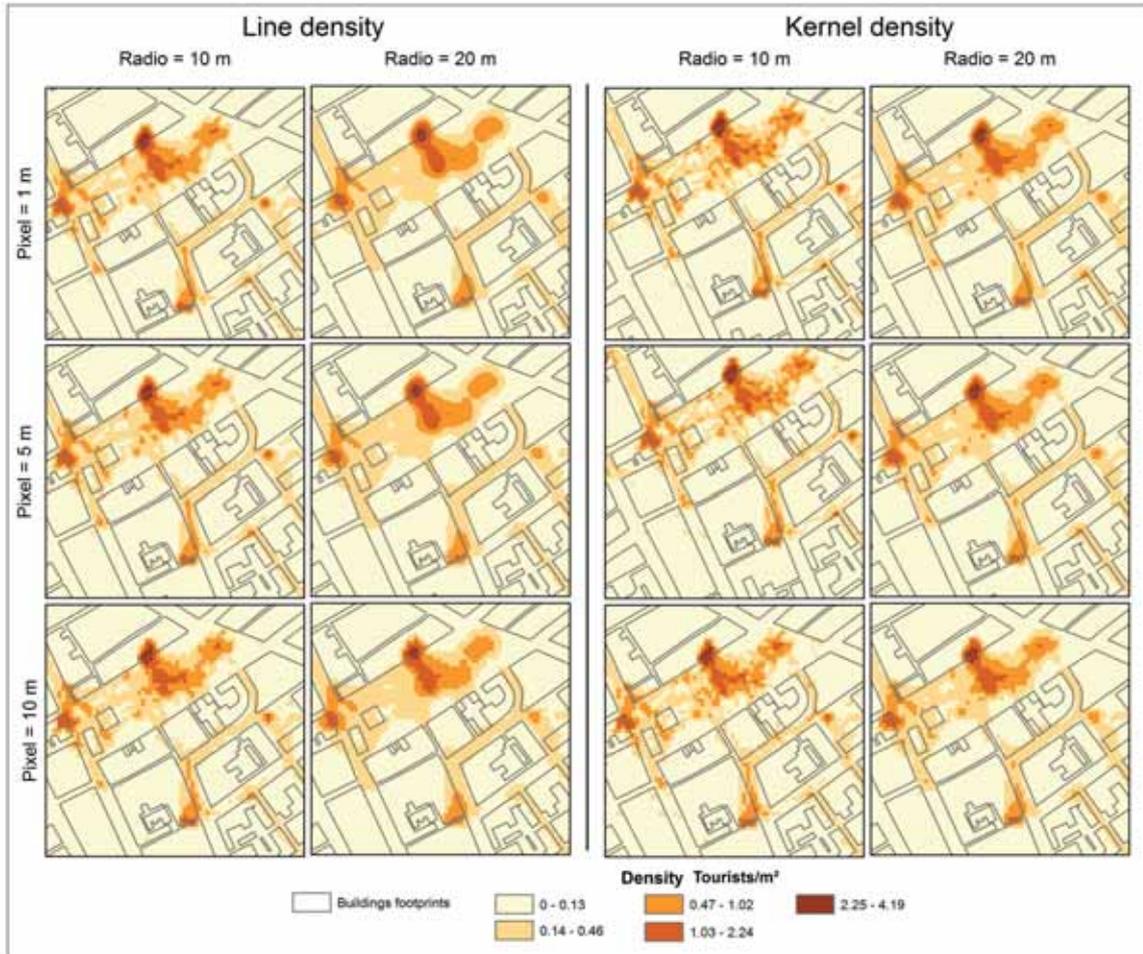


Figure 6.1: Effect of various size for radius and pixel in two density-based aggregation methods of trajectories. Tourist visiting Delft City, November 20th, 2009.

Variation of the size has a less dramatic effect compared with the size of radius. Areas considered as hot spots expand slightly when increasing the size of the pixel. Another effect is the loose of information with respect to the details. Ideal pixel size for the computation cannot be recommended in this case. Similar conclusions were found by Hengl (2006).

Sections 4.3.4, 5.2.1 and 5.2.2 have showed the application of uncertainty computation in human mobility context. One limitation of the interpretation of the uncertainty aggregation in Figure 4.6, is that probabilistic values to bound the area which represent a confidence interval needs to be computed every time σ and pixel size changes, for the same dataset. Automation of this computation in parallel with uncertainty aggregation can be done by storing *z-scores* for common used confidence intervals and developing appropriate functions. Interpretation of log probability values makes difficult to interpret the result of the aggregation, and therefore a normalized uncertainty ratio as presented in Figure 5.2 is more meaningful. Log probability values are easily converted into uncertainty values by using equation 3.6. Nevertheless log probability values resulted to be useful in the 3D visualization method presented in Figure 5.6.

Results of positional uncertainty add a quality indicator to results presented by van Lagelaar and van der Speck (2010) and Baltus et al. (2010), and add reliability to the decision taken out of the result of density aggregation in general. This is demonstrated by the results presented in Figures 5.3 and 5.6.

Experimentation with different moving object datasets will only expand the possibilities of using PUTMO for computing positional uncertainties. Giannotti et al. (2011); Demšar and Virrantaus (2010) recognize the relevance of density aggregation method applied to moving object datasets in various contexts (e.g. human geography, spatial epidemiology, transport and LBS). Consequently, positional uncertainty computation also finds its relevance and applicability in those fields.

Artifacts in the error bands appear in the cases where trajectories cross or the continuity of the movement is interrupted, (Figure 4.3 makes the first case evident) this is driven by the maximum value for the case of overlaying error bands in the case of a 2D representation (See Algorithm 3). This produces a result that mask the probability values that correspond to different time intervals. Artifacts related with the interruption of the movement are presented in Figure 4.4 (b). Those discontinuities can correspond to: (i) the data cleaning process as was described in Chapter 2, and/or (ii) signal loss of the collecting device in the situation when the last and the next position recorded are not too far from each other.

Computation of positional uncertainty in PUTMO allows answering questions that involve a measure of reliability. A question like *Which were common places visited by tourist on November 19th, within 99% confidence interval?* finds an answer in our systems, in the form of areas that represent confidence intervals (See Figure 4.6). Similar questions cannot be answered with density estimation methods like the ones described in (Silverman, 1992) or the ones used or implemented by van Lagelaar and van der Speck (2010); Demšar and Virrantaus (2010); Willems et al. (2009); Scheepens (2010); Scheepens et al. (2012). Therefore, PUTMO adds a measure of reliability to spatio-temporal aggregation of trajectories from moving object datasets.

One limitation of our system is that it can only be used in the study of historical movement. Even when including a probabilistic model, it cannot be used as a predictive tool. An approach for modeling uncertainties as a prediction tool is given by Praing and Schneider (2007) and (Monreale, Pinelli, Trasarti, Giannotti, & Acm, 2009) propose a technique to predict object movement based on its historical movement.

6.1.1 Movement and Context dependent Error-Bands Models

Interpretation of error bands and its applicability is dependent of the type of movement. Section 4.2 demonstrated the applicability of PUTMO in the unconstrained movement. In unconstrained environments, interpretation is limited the idea that the error regions describe the space of possible movement, as any *authority* constraint (Hagerstrand, 1970) is disregarded. This means we assume the object is free to move in any direction and occupy any location. In the case of hurricane data the distinction of regular and irregular error bands demonstrated to be irrelevant, mainly because of the big size of the object, the high accuracy of the tracking method and the low maneuverability. When the standard deviation of the tracking device is small compare to the distance between recorded position, error bands adopt a regular shape even when *sigma ratio* is set to a high value (See Figure 4.2). The change of sigma along the trajectory is compensated by the distance travel by the hurricane between recorded positions, and in most of the cases regular bands are enough to represent the area of possible movement.

Meanwhile, in the case of constrained environments the movement is restricted. In the case of constrained movement (pedestrians) error bands could be limited to the area inside the street network, since the use of contextual information (e.g. buildings, canals, lakes) as a reference can help to reduce the extent of positional uncertainty of the moving object. The use of networks to reduce uncertainty in the movement was already described by Richter, Schmid, and Laube (2012).

Validation of the model needs to be done for both cases. In the case of unconstrained movement validation can be done by comparing location and extension of error band, and the common paths used for movement as reference data. This will only be possible in the case of those common paths are known in advance (e.g. maritime transport or air traffic). For the constrained case, validation of results becomes a more difficult task when obtaining reference data, for instance in the case of pedestrians, their movement is constrained to the street network but this network allows certain freedom in the movement and in some cases pedestrians can go out of the street network and inside the buildings. The movement constrains described by (Hagerstrand, 1970) could suggest ad-hoc solutions (e.g. authority constraints in the case of cyclists) for this case.

6.1.2 Visualization methods and Uncertainty in Moving Object Datasets

Shading as visual variable for probability values has some limitations (See Figure 5.3). These limitations are related with the highly heterogeneous probability surface produce by the aggregation of error bands and the high unpredictability in speed or direction of a pedestrian. Similar limitations where found by (Scheepens et al., 2012) when applying an innovative method for density visualization that showed to be more useful in the case of object in unconstrained environment and with lower maneuverability (vessels). Improvement to the visualization was done when using both shading and contour lines to represent probability values (See Figure 5.4). The development of tools to add interaction and the use of dynamic variables to emphasize the visualization of uncertainties could help to overcome some of the limitation of the visualization method presented in Section 5.2.

The work by Senaratne et al. (2012) showed the relevance of usability test in the communication of uncertainties related with geoinformation. In their test air quality data was presented using various visualization methods. A similar usability test using uncertainties of moving object might give different results, more over if the temporal component of the data is included.

6.2 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

In the process of knowledge discovery we identify and describe factor that introduce uncertainty in moving object spatio-temporal aggregation. In data collection: signal loss, device spatial accuracy and sampling rate. In data pre-processing: data cleaning process and trajectory model representation, and in data aggregation: time window selection, tessellation and radius as a parameter of density estimation.

In density-based aggregation of trajectories, the influence of the time window in the level of spatial uncertainty resulted to be more relevant than the choices of pixel size and radius as parameter for the density estimation.

Modeling of positional uncertainties of moving object as an stochastic process was effectively applied to propagate the random errors in the knowledge discovery process. Extension of the system for the case in which σ changes for every positions along a trajectory, the consideration of gross and systematic positional errors, and validation, can be seen as improvement to the system. An interested question for future research is related to how to extend the model to consider temporal uncertainties.

A complete automation of PUTMO to include density aggregation directly in the system will increase its applicability. This will be possible when PostGIS will be extended in order to include density estimations (currently in progress), and the limitations in performance in our implementation will be overcome. One way to do this is implement the system as a library using a low level programming language (e.g. C++). This will facilitate its applicability and extensibility from the technical point of view.

The result of total probability suggests that the model implemented in our system has the potential to become a density aggregation method itself. A probabilistic method, improvement in the interpretation of the results, specially related with log probabilities needs to be part of future work.

Application of the uncertainty model implemented in PUTMO is context and movement type dependent. Result as regular or irregular error bands around trajectories depend on the nature of the movement (constrained or unconstrained), and the choices of values for the parameter.

Artifact in the positional uncertainty computation for individual trajectories can be used to identify discontinuities in the movement of the object. Further work needs to be done to automatically separate this case from the case of crossing of trajectories.

Adjacent maps and contour lines were applicable to represent positional uncertainties in trajectory density-based aggregation, when applied to result from constrained moving object datasets. Color and shading were less effective, and a 3D representation of the probability makes evident that high density values not always correspond with high probabilities. Further research needs to be done in order to look for the reason of this behavior, and the applicability of color and shading in the case of aggregation of trajectories in the case of unconstrained movement.

Implementation and applicability of trajectory aggregation considering error region (See Algorithm 5), as an alternative to include confidence intervals directly in the density aggregation.

References

- Andrienko, G., & Andrienko, N. (2010). A general framework for using aggregation in visual exploration of movement data. *The Cartographic Journal*, 47(1), 22-40.
- Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., & Wrobel, S. (2011). A conceptual framework and taxonomy of techniques for analyzing movement. *Journal of Visual Languages and Computing*, 22(3), pp 214-215.
- Andrienko, G., Andrienko, N., Schumann, H., Tominski, C., Demšar, U., Dransch, D., & Dykes, J. (2010). Space and time. In D. Keim, J. Kohlhammer, G. Ellis, & F. Mansmann (Eds.), *Mastering the information age: Solving problems with visual analytics* (p. 66). Germany: Eurographics Association.
- Awasthi, D. (2012). *Design and realization of raster-based spatial uncertainty calculus around vector objects*. Master's thesis, University of Twente Faculty of Geo-Information and Earth Observation ITC.
- Baltus, M., Dirks, H., Esselink, S., Kwon, O., Langelaar, T. v., Rozemuller, M., & Saarloos, S. (2010). *Tracking delft* (Technical Report). Laboratory for Cities and Urban Landscapes, Technical University of Delft.
- Brunsdon, C. (1995). Estimating probability surfaces for geographical point data: An adaptive kernel algorithm. *Computers and Geosciences*, 21(7), pp 877-894.
- CLS. (2011). *Argos user's manual* (Manual). Collectie Localisation Satellites.
- COST. (2009). *Memorandum of understanding for the implementation of a european concerted research action designated as COST action IC0903: Knowledge discovery from moving objects MOVE* (Memorandum). Brussels, Belgium: European Cooperation in the field of Scientific and Technical Research.
- de Paulo, M. (2012). *Quantum GIS plugin repository* (Vol. 2012) (No. Nov 5th). Retrieved November 25th, 2012, from <http://plugins.qgis.org/plugins/wktraster/>
- Demšar, U., & Virrantaus, K. (2010). Space-time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*, 24(10), pp 1527-1542.
- Fisher, P., Comber, A., & Wadsworth, R. (2006). Approaches to uncertainty in spatial data. In R. Devillers & R. Jeansoulin (Eds.), *Fundamentals of spatial data quality* (p. 43-45). London: ISTE.
- Fredrikson, A., North, C., Plaisant, C., & Shneiderman, B. (1999). Temporal, geographical and categorical aggregations viewed through coordinated displays: a case study with highway incident data. In *Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM international conference on information and knowledge management* (p. 26). ACM.
- Gatignon, H. (2010). Multivariate normal distribution. In *Statistical analysis of management data* (p. pp 9-11). London: Springer.
- Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S., & Trasarti, R. (2011). Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal*, 20(5), pp 695-719.
- Goodchild, M. F. (2009). The quality of geospatial context. In (Vol. 5786 LNCS, p. pp 15-16).

- Hagerstrand, T. (1970). How about people in regional science? In *Regional science association conference* (p. 7-21).
- Hengl, T. (2006). Finding the right pixel size. *Computers and Geosciences*, 32(9), pp 1283-1298.
- Holland, R. A., Wikelski, M., Kummeth, F., & Bosque, C. (2009). The secret life of oilbirds: New insights into the movement ecology of a unique avian frugivore. *PLoS ONE*, 4(12), 8264.
- Hornsby, K., & Egenhofer, M. J. (2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, 36(1-2), pp 177-181.
- Keim, D. (2010). *Mastering the information age : solving problems with visual analytics : also as e-book*. Visual Analytics - Mastering the Information Age (VisMaster).
- Kraak, M. J., & Ormeling, F. (2010). Geographical information science and maps. In *Cartography: visualization of spatial data* (p. 1-3). GB: Pearson Education.
- Leick, A. (2004). *GPS satellite surveying* (Third ed.). New Jersey, US: Wiley & Sons.
- Liu, H., & Schneider, M. (2011). Balloon: Representing and querying the near future movement of predictive moving objects. In *19th ACM SIGSPATIAL international conference on advances in geographic information systems, ACM SIGSPATIAL GIS 2011* (p. p 9-16).
- Miller, H. J. (1991). Modelling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information Systems*, 5(3), 287-301.
- Monreale, A., Pinelli, F., Trasarti, R., Giannotti, F., & Acm. (2009). *Wherenext: a location predictor on trajectory pattern mining*. New York: Assoc Computing Machinery.
- OGC. (2010). *OpenGIS implementation standard for geographic information - simple feature access - part 1: Common architecture* (No. OGC 06-103r4).
- Openshaw, S. (1984). Introduction. In *The modifiable areal unit problem* (Vol. 38, p. 3-4). England: Geo Books.
- Orellana, D., & Wachowicz, M. (2011). Exploring patterns of movement suspension in pedestrian mobility. *Geographical Analysis*, 43(3), 241-260.
- Pelekis, N., Theodoulidis, B., Kopanakis, I., & Theodoridis, Y. (2004). Literature review of spatio-temporal database models. *Knowledge Engineering Review*, 19(3), 235-274. (Cited By (since 1996): 44 Export Date: 23 February 2013 Source: Scopus)
- Pfoser, D., & Jessen, C. (1999). Capturing the uncertainty of moving-object representations. In G. RH., P. D., & L. F. (Eds.), *Advances in spatial database. lecture notes in computer science* (Vol. 1651, p. 111-131). Berlin: Springer.
- Praing, R., & Schneider, M. (2007). Modeling historical and future movements of spatio-temporal objects in moving objects databases. In *Sixteenth ACM conference on conference on information and knowledge management* (p. 183-192).
- Qstarz International Co., Ltd. (2012). *Qstarz beyond navigation*. Retrieved January 23th, 2013, from <http://www.qstarz.com/Products/GPS%20Products/BT-Q1000X-S.htm>
- Richter, K., Schmid, F., & Laube, P. (2012). Semantic trajectory compression: Representing urban movement in a nutshell. *Journal of Spatial Information Science*, 4(1), pp 3-7.
- Scheepens, R. (2010). *Gpu-based track visualization of multivariate moving object data*. Master's thesis.
- Scheepens, R., Willems, N., van de Wetering, H., & van Wijk, J. (2012). Interactive density maps for moving objects. *Computer Graphics and Applications, IEEE*, 32(1), pp 64-65.
- Scheepens, R., Willems, N., Van De Wetering, H., & Van Wijk, J. J. (2011). Interactive visualization of multivariate trajectory data with density maps. In (p. 147-154).
- Schneider, M. (2009). Moving object in databases and gis: State-of-the-art and open problems. In G. Navratil (Ed.), *Reseach trends in geographic information* (p. 151-167). Springer Berlin Heidelberg.

- Senaratne, H., & Gerharz, L. (2011). An assesment and categorization of quantitative uncertainty visualisation methods for geospatial data. In S. Geertman, W. Reinhardt, & T. F (Eds.), *14th AGILE international conference on geographic information science - advancing geoinformation science for a changing world*. AGILE.
- Senaratne, H., Gerharz, L., Pebesma, E., & Schwering, A. (2012). Usability of spatio-temporal uncertainty visualisation methods. In J. Gensel, D. Josselin, & D. Vandenbroucke (Eds.), *Bridging the geographic information sciences* (p. 3-23). Springer Berlin Heidelberg.
- Shi, W. (2010). *Principles of modeling uncertainties in spatial data and spatial data analyses*. (W. Shi, Ed.). Boca Raton, CRC Press.
- Shi, W., & Liu, W. (2000). A stochastic process-based model for the positional error of line segments in gis. *International Journal of Geographical Information Science*, 14(1), 51-66.
- Silverman, B. W. (1992). Density estimation for statistics and data analysis. In *Monographs on statistics and applied probability* (Vol. 26). Chapman & Hall.
- Thompson, J., Rahayu, W., & Torabi, T. (2012). An interactive region-based filter for moving objects datasets: Making sense of chaos. In *26th IEEE international conference on advanced information networking and applications* (p. 199-206).
- Trajcevski, G., Tamassia, R., Cruz, I. F., Scheuermann, P., Hartglass, D., & Zamierowski, C. (2011). Ranking continuous nearest neighbors for uncertain trajectories. *VLDB Journal*, 20(5), 767-791.
- UNISYS. (2012). *UNISYS weather*. Retrieved November 11th, 2012, from <http://weather.unisys.com/hurricane/>
- van Lagelaar, T., & van der Speck, S. (2010). Visualizing pedestrian flows using GPS-tracking to improve inner-city quality. In *11th international Walk21 conference and 23rd international workshop of the international co-operation on theories and concepts in traffic safety* (p. 19). Walk21.
- Willems, N., van de Wetering, H., & van Wijk, J. (2009). Visualization of vessel movements. *Computer Graphics Forum*, 28(3), pp 959-966.
- Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., & Aberer, K. (2011). Semitri: A framework for semantic annotation of heterogeneous trajectories. In *4th international conference on extending database technology: Advances in database technology, edbt 2011* (p. 261).
- Yoon, S., Deutsch, K., Chen, Y., & Goulias, K. (2012). Feasibility of using time-space prism to represent available opportunities and choice sets for destination choice models in the context of dynamic urban environments. *Transportation, (Online first)*, pp 1-17.

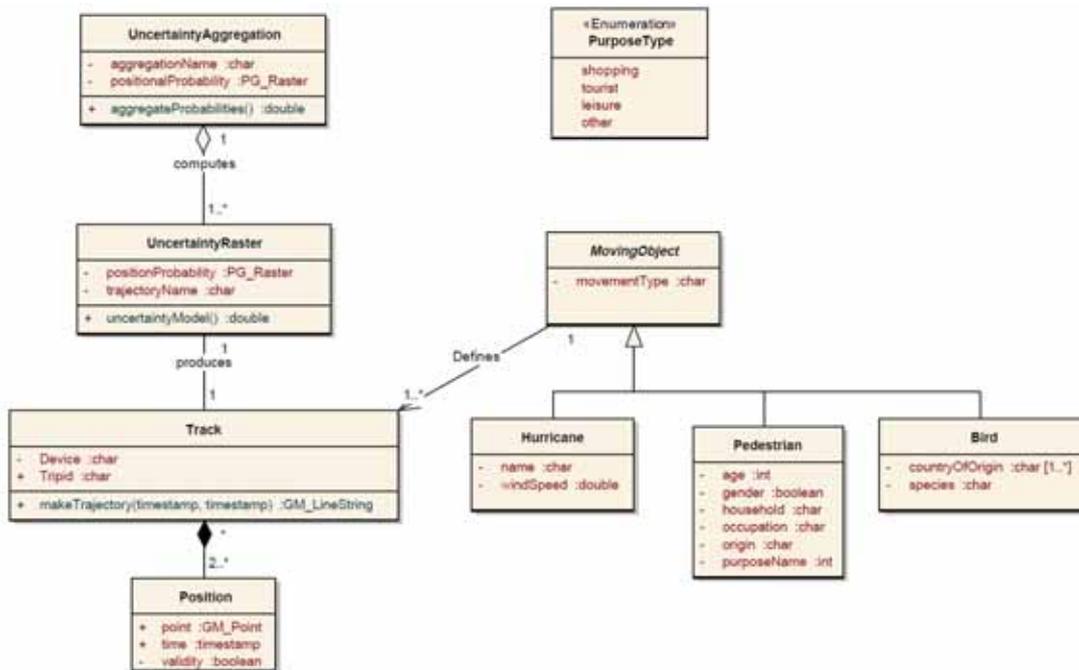
Appendix A

List of Symbols and Variable in this Document

Symbol	Name	Description	Symbol	Name	Description
A	Capital A	Area	y	Small Y	Quadratic function
a	Small A	Quadratic coefficient	\hat{x}	Small X.hat	Measurement
C	Capital C	Sigma ratio	Z	Capital Z	z-score
C_i	Capital C.I	Region isodensity contours	z	Small Z	pixel size
j	Small J	trajectory segment	α	Alpha	Confidence interval
k	Small K	Radius o	Γ	Gamma	Function change of σ
l	Small L	Trajectory segment	Δ_t	Delta.t	Time interval
N_R	Capital N.R	Number of removed positions	ϵ	Epsilon	Total error
P	Capital P	Total probability	ϵ_g	Epsilon.g	Gross error
P'	Capital P.log	Probability of pixel	ϵ_r	Epsilon.r	Random error
P_{log}	Capital P prime	Total probability of raster	ϵ_s	Epsilon.s	Systematic error
p	Small P	Pixel	μ	Mu	Mean value
q	Small Q	Centroid of a pixel	μ_x	Mu.x	True value
q'	Small Q prime	q on trajectory segment	ρ	Rho	Correlation between variables
R	Capital R	Raster	σ	Sigma	Standard deviation
r	Small R	Buffer radius	σ_{max}	Sigma	Maximum value for σ
S	Capital S	Trajectory error raster	τ	Tau	Moving object dataset
s_r	Small S.r	Sampling rate	$\hat{\tau}$	Tau.hat	Trajectory set of moving points
T	Capital T	Time Window	Φ	Phi	Trajectory segment error surface
w	Small W	Euclidean component	ϕ	Phi	Probability of error circle
Y	Capital Y	Uncertainty contribution	φ	Phi	Density by trajectory counting

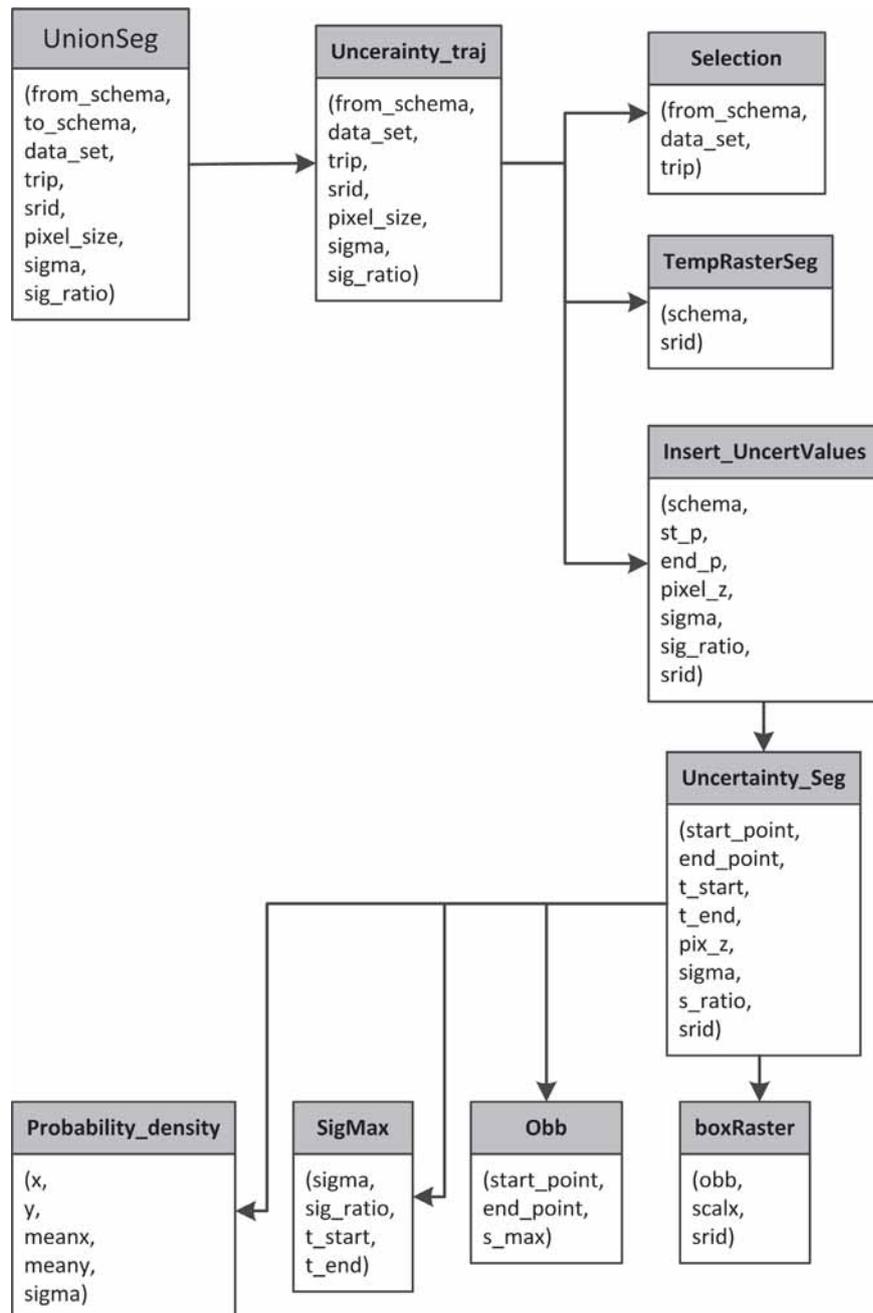
Appendix B

Database Schema for Storing and Computing Positional Uncertainties of Moving Object Datasets



Appendix C

Function Tree implemented in PUTMO



Appendix D

Source Code for Loading GPX files into PostgreSQL: Language DOS

```
set PGPORT='port'
set PGHOST='host'
set PGUSER=postgres
set PGPASSWORD='password'
set THEDB= 'database name'
set PGBIN= 'path to PostgreSQL\9.1\bin'
set OGR= 'path to ogr2ogr driver'
set DATADIR= 'data directory'
set RETRIVETYPE= track_points
set SRS= 'WKT file for SRS transformation'
set FLIST= 'list of file names'
set EXTFILE=.gpx
"%PGBIN%\psql" -d "%THEDB%" -c "CREATE SCHEMA temp"
set SCHEM=temp

for %%f in (%FLIST%) DO "%OGR%\ogr2ogr" -t_srs "%SRS%" -f "PostgreSQL"
PG:"host=%PGHOST% user=%PGUSER% port=%PGPORT% dbname=%THEDB%
password=%PGPASSWORD%"
"%DATADIR%\%%f%EXTFILE%" -overwrite -lco GEOMETRY_NAME=the_geom -lco
SCHEM=%SCHEM% %RETRIVETYPE% -nln "t%%f"

for %%f in (%FLIST%) DO "%PGBIN%\psql" -d "%THEDB%" -c "ALTER TABLE
"%SCHEM%.t%%f"
ADD COLUMN tripid character varying (50)"

for %%f in (%FLIST%) DO "%PGBIN%\psql" -d "%THEDB%" -c "UPDATE "%SCHEM%.t%%f"

SET tripid='%%f'"
"%PGBIN%\psql" -d "%THEDB%" -c "CREATE SCHEMA clean_data"
"%PGBIN%\psql" -d "%THEDB%" -f "C:\gisdata\new_table.sql"
for %%f in (%FLIST%) DO "%PGBIN%\psql" -d "%THEDB%" -c "INSERT INTO
"clean_data"."valid_tracks" (tripsoid, the_geom,track_seg_point_id,elev,
time_,desc_)
SELECT tripid, the_geom, track_seg_point_id, ele, time, 'desc' FROM
"%SCHEM%.t%%f"
"%PGBIN%\psql" -d "%THEDB%" -c "VACUUM ANALYZE"
pause
```


Appendix E

Source Code of Functions Implemented in PUTMO: Language pl/pgSQL

```
-----  
-- FUNCTION: UnionSeg --  
-----  
  
CREATE OR REPLACE FUNCTION pedestrians.mo_unionseg(from_schema text,  
  to_schema text, data_set text, trip text, srid int, pix_size double  
  precision, sigma double precision, sig_ratio real)  
  RETURNS void AS  
$BODY$  
  
DECLARE  
  _data_set text;  
  trip text:=trip;  
  table_name text;  
  return_val text:='MAX';  
  _from_schema text;  
  _to_schema text;  
  _srid int;  
  
BEGIN  
  _data_set:= data_set;  
  _from_schema:=from_schema;  
  _to_schema:= to_schema;  
  _srid:=srid;  
  
  --Compute rasters for each trajectory segment:  
  PERFORM pedestrians.mo_uncertainty_traj(_from_schema, _data_set,trip,  
    _srid, pix_size, sigma, sig_ratio);  
  
  --storage table:  
  table_name:='rast_'||trip;  
  
-----  
-- UNION OF ALL SEGMENTS AND RETURNING THE MAXIMUM PIXEL VALUE --  
-----
```

```
EXECUTE format ('DROP TABLE IF EXISTS %I.%I', _to_schema, table_name);
EXECUTE format('CREATE TABLE %I.%I (rid serial PRIMARY KEY, rast raster,
  track_id varchar)', _to_schema, table_name);
EXECUTE format('UPDATE %I.%I SET rast=st_setsrid(rast,$1)', _to_schema,
  table_name) USING _srid;
EXECUTE format ('INSERT INTO %I.%I (rast, track_id) VALUES ((SELECT
  St_union (rast, %L) FROM "rast_temp")::raster, $1::varchar)',_to_schema,
  table_name, return_val)
USING trip;
EXECUTE format('ANALYZE %I.%I',_to_schema, table_name);

--dropping temporal raster table:
DROP TABLE rast_temp;

END;
$BODY$
  LANGUAGE plpgsql VOLATILE
  COST 5000;

-----
-- FUNCTION: Uncertainty_traj --
-----

CREATE OR REPLACE FUNCTION pedestrians.mo_uncertainty_traj(from_schema text,
  data_set text, trip text, srid int, pix_size double precision, sigma double
  precision, sig_ratio real)
RETURNS void AS
$$
-- parameter: trip text, sigma double precision, sig_ratio double precision

DECLARE
  _data_set text;
  id_first integer;
  id_last integer;
  _confidence double precision;
  _from_schema text;
  _srid int;

BEGIN
  _data_set:= data_set;
  _from_schema:=from_schema;
  _srid:=srid;

  -- select all the point of a point in a trajectory:
  PERFORM pedestrians.mo_selection(trip,_from_schema::text, _data_set::text);

  -- create table to store probability values for individual segments:
  PERFORM pedestrians.mo_temprasterseg(_srid);
```

```
id_first:= (SELECT t.new_id2 FROM temp1 AS t ORDER BY t.new_id2 ASC LIMIT 1)
::integer;

id_last:= (SELECT t.new_id2 FROM temp1 AS t ORDER BY t.new_id2 DESC LIMIT 1)
::integer;

-- COMPUTE uncertainty along a segment:
WHILE id_first < id_last
LOOP

PERFORM pedestrians.mo_insert_UncertValues(id_first, id_first+1, pix_size,
sigma, sig_ratio, _srid);

id_first:= id_first+1;
END LOOP;

--dropping temporal table:
DROP TABLE temp1;

END;
$$
LANGUAGE plpgsql VOLATILE
COST 200;

-----
-- FUNCTION: Selection --
-----

CREATE OR REPLACE FUNCTION pedestrians.mo_selection(trip varchar,
from_schema text, data_set text)
RETURNS void AS

$$
DECLARE
trip_id varchar:=trip;
_data_set text;
_from_schema text;

BEGIN
_data_set:= data_set;
_from_schema:= from_schema;

EXECUTE format('CREATE TEMPORARY TABLE temp1 AS SELECT t.* FROM %I.%I AS t
WHERE t.tripid =%L',_from_schema, _data_set, trip_id);

ALTER TABLE temp1 ADD COLUMN new_id2 serial PRIMARY KEY;
--RETURN 'Selection table was created';

ANALYZE temp1;
```

```

END;
$$
LANGUAGE plpgsql VOLATILE
COST 100;

-----
-- FUNCTION: TempRasterSeg --
-----

CREATE OR REPLACE FUNCTION pedestrians.mo_tempRasterSeg(schema text, srid int)
RETURNS void AS

$$
-- Create a empty table to temporarily store individual raster segment.

DECLARE
_srid int;

BEGIN
_srid:=srid;
CREATE TEMPORARY TABLE "rast_temp"(rid serial PRIMARY KEY, rast raster);
UPDATE "rast_temp" set rast=st_setsrid(rast,_srid::int);

END;
$$
LANGUAGE plpgsql VOLATILE
COST 100;

-----
-- FUNCTION: Insert_UncertValues --
-----

CREATE OR REPLACE FUNCTION pedestrians.mo_insert_UncertValues(st_p int,
end_p int, pixel_z double precision, sigma double precision, sig_ratio real,
srid int)
RETURNS void AS
$$

-- r_table: name of an existing table where the values are will be inserted.
-- g_table: name of the table that contains the points.
-- po adn p1 refer to the id of two consecutive points.

DECLARE
po geometry;
p1 geometry;
t_o timestamp with time zone;
t_1 timestamp with time zone;
_srid int;

BEGIN

```

```
_srid:=srid;
t_o:= (SELECT p.time_ FROM "temp1" AS p WHERE p.new_id2=st_p);
t_1:= (SELECT p.time_ FROM "temp1" AS p WHERE p.new_id2=end_p);
po:= (SELECT p.the_geom FROM "temp1" AS p WHERE p.new_id2=st_p);
p1:= (SELECT p.the_geom FROM "temp1" AS p WHERE p.new_id2=end_p);

BEGIN
INSERT INTO "rast_temp"
("rast") VALUES (pedestrians.Uncertainty_Seg(
  po, --start point
  p1, --end point
  t_o, --start time
  t_1, --end time
  pixel_z, --pixel size
  sigma, -- standard deviation
  sig_ratio, -- S/time change in percentage
  _srid)::raster);

EXCEPTION WHEN internal_error THEN
RAISE NOTICE 'error at INSERTING VALUE TO TABLE, AT SEGMENT: % to %',
st_p, end_p;

END;
END;
$$
LANGUAGE plpgsql VOLATILE
COST 200;

-----
-- FUNCTION: Uncertainty_Seg --
-----

CREATE OR REPLACE FUNCTION pedestrians.uncertainty_seg(start_point geometry,
  end_point geometry, t_start timestamp with time zone, t_end timestamp with
  time zone, pix_z double precision, sigma double precision, s_ratio real,
  srid int)
  RETURNS raster AS
$BODY$

<<main>>
/*Compute the uncertainty along a line segment, given the endpoints,
  timestamps and error: sigma.
scale: pixel size
s_ratio, determine, s_max (the sigma maxima) half way between 2 consecutive
  time stamps; units are
given in percentage sigma over time in seconds, and it goes from zero (0) to
  infinity, when zero sigma does not change
along a trajectory segment*/

DECLARE
```

```
obb geometry;
output_ras raster;
s_max double precision; -- sigma maxima at the middle of the line segment
sigma double precision:= sigma; --sigma as given in function parameters
sig double precision; --sigma in the further computations, moving sigma
_confidence double precision;
pixel_val double precision;
scalx double precision:=pix_z;
scaly double precision:=pix_z;

--to fill in raster:
width smallint;
height smallint;
i smallint;
j smallint;

--SUPPORT to compute uncertainty value:
ULX double precision;
ULY double precision;
ULPD double precision;
LLPD double precision;
LRPD double precision;
URPD double precision;
MPPD double precision;
q geometry; --a pixel centroid
q_line geometry; -- q projected on the line (using closest projection)
qlx double precision; --moving x-mean
qly double precision; --moving y-mean
dist_qline double precision; --distance from q on the line to start point
l_leng double precision; --length of the line segment
pixel_init double precision;

--geometry constructors:
_srid int;
l_seg geometry; -- line segment

--for parabola computation:
a double precision; -- for smoothing ellipse error

BEGIN
_srid:= srid;
-----
-- COMPUTE CHANGE IN sigma using linear relation --
-----
IF st_equals(start_point,end_point) = false THEN
s_max:= mo_sigMax(sigma, s_ratio, t_start, t_end);
--Create geometry box around line segment:
obb:= pedestrians.mo_obb(start_point, end_point, s_max)::geometry;
```

```
--Convert geometry box to raster (empty raster):
output_ras:=pedestrians.mo_boxRaster(obb, scalx,_srid);

--find width & height of the empty raster:
width:= st_width(output_ras);
height:= st_height(output_ras);

--construct line segment:
l_seg:= st_setSRID(st_makeLine(start_point,end_point),_srid);
l_leng:= st_length(l_seg);

BEGIN
-- COMPUTE a, constant that defines parabola (behaviour) of sigma along
the line segment
-- Constant for line segment

a:= 2*(s_max - sigma) / (l_leng*l_leng);
END;

FOR i in 1..width
LOOP
FOR j IN 1..height
LOOP
IF st_value(output_ras,i,j)=1.0
THEN

--centroid of the pixel 'q'
--This function returns the x-coordinates of upperleft corner
ULX:= ST_Raster2WorldCoordX(output_ras,i,j) + scalx/2;
ULY:= ST_Raster2WorldCoordy(output_ras,i,j) - scaly/2;

q:= st_setsrid(st_MakePoint(ULX, ULY),_srid);
q_line:= st_closestPoint(l_seg, q); --project q to the closest point
on the line segment

-- coordinates of q' on the line = moving means:
qlx:= st_x(q_line);
qly:= st_y(q_line);
--distance start_point to q_line
dist_qline:= st_distance(start_point,q_line);

IF st_equals(start_point,q_line) = true OR
st_equals(end_point,q_line)=true THEN
sig:= sigma;

ELSE
```

```
--Compute new sigma based on the position along line segment:
sig:= sigma + a*(dist_qline)*(l_leng - dist_qline);

END IF;

--calculate probability density at each corner and middle point
of each cells

--For upperleft
ULPD:= pedestrians.mo_probability_density(ULX,ULY,qlx,qly,sig);
IF ULPD <> 0.0 THEN
-- For lowerleft
LLPD:= pedestrians.mo_probability_density(ULX,ULY-scalx,qlx,qly,sig);
IF LLPD <> 0.0 THEN
--For lowerright
LRPD:= pedestrians.mo_probability_density
(ULX+scalx,ULY-scalx,qlx,qly,sig);
IF LRPD <> 0.0 THEN

--For upperright
URPD:= pedestrians.mo_probability_density
(ULX+scalx,ULY,qlx,qly,sig);

IF URPD <> 0.0 THEN

--For middle point
MPPD:= pedestrians.mo_probability_density
(ULX+(scalx/2.0),ULY-(scalx/2.0),qlx,qly,sig);
IF MPPD <> 0.0 THEN

--Now calculating volume under each triangles of pyramid and
summing up them,

BEGIN
pixel_val:=1/6.0 *(scalx*scalx)*
(ULPD+LLPD+2*MPPD+LRPD+URPD)::double precision;

EXCEPTION
WHEN numeric_value_out_of_range THEN
pixel_val:=0.00000;
RETURN pixel_val::double precision;
END;
END IF;
END IF;
END IF;
END IF;

ELSE
```

```
pixel_val:=0.00000;
END IF;

output_ras := st_SetValue(output_ras,i,j,pixel_val);

END IF;
END LOOP;
END LOOP;
END IF;

RETURN output_ras;

END;
$BODY$
  LANGUAGE plpgsql IMMUTABLE
  COST 1000;

-----
-- FUNCTION: Probability_density --
-----

CREATE OR REPLACE FUNCTION pedestrians.mo_probability_density(x double
precision, y double precision, meanx double precision, meany double
precision, sigma double precision)
  RETURNS double precision AS
$BODY$
--This function takes parameter x and y as variable, meanx, sigmax are
parameter for x, similarly meany and sigmay for y, rho represents
correlation coefficient.

  DECLARE
    probability_density double precision;
    sigmax double precision:=sigma;
    sigmay double precision:=sigma;
    --rho double precision:=0; --DONT NEED THIS
    -- initial_val double precision;

  BEGIN
  BEGIN
    probability_density:= (1/(2.0*pi()*sigmax*sigmay))*exp
    (-((power((x-meanx),2) + power((y-meany),2))/(2*sigmax* sigmay)));

  EXCEPTION
  WHEN numeric_value_out_of_range THEN
    RAISE NOTICE 'UNDERFLOW';
    probability_density:=0.00000;
    RETURN probability_density;

  END;
```

```
RETURN probability_density::double precision;

END;
$BODY$
LANGUAGE plpgsql IMMUTABLE STRICT
COST 100;

-----
-- FUNCTION: SigMax --
-----

REATE OR REPLACE FUNCTION pedestrians.mo_sigMax(sigma double precision,
s_ratio real, t_start timestamp with time zone,t_end timestamp with time
zone)
RETURNS double precision AS
$BODY$

DECLARE
s_ratio real:= s_ratio; --chage in sigma given a change in time
s_max double precision;
int_sec integer;
sigma double precision:=sigma;

BEGIN
s_max:=sigma;

IF s_ratio <> 0::double precision THEN
t_int:= (t_end - t_start)::interval;
--convert time interval to seconds:

int_sec:= ((extract(hour from t_int))*3600 + (extract(minute from
t_int))*60 + (extract(second from t_int)))::integer;

s_max:= ((s_ratio/100 * int_sec) + sigma)::double precision;

END IF;

RAISE NOTICE 'time inverval in seconds: %; s_max: %', int_sec, s_max;

RETURN s_max;

END;
$BODY$
LANGUAGE plpgsql IMMUTABLE STRICT
COST 100;
```

```
-----
-- FUNCTION: Obb --
-----
```

```
CREATE OR REPLACE FUNCTION pedestrians.mo_obb(start_point geometry,  
end_point geometry, sigma_max double precision)  
RETURNS geometry AS  
  
/* Using the build-in BUFFER function to bound the extend the  
computation of the uncertainty values around a trajectory segment.  
Sigma represent the standar deviation and is used to comput the  
least max radius around the trajectory segment.  
line segment should be provided as line.  
*/  
  
$BODY$  
DECLARE  
z double precision;  
sigma_max double precision:= sigma_max;  
line_seg geometry;  
obb geometry;  
r double precision;  
  
BEGIN  
  
z:=3.4807;  
RAISE NOTICE 'z value set to: %', z;  
  
r:= (z*sigma_max)+ sigma_max*0.1;  
-- the constant 0.1 extends the radius of the buffer, 10% of sigma.  
  
obb:=ST_Buffer(ST_MakeLine(start_point, end_point), r, 'endcap=square');  
  
RETURN obb;  
  
END;  
$BODY$  
LANGUAGE plpgsql IMMUTABLE  
COST 100;  
  
-----  
-- FUNCTION: BoxRaster --  
-----  
  
CREATE OR REPLACE FUNCTION pedestrians.mo_boxRaster(obb geometry,  
pix_z double precision, srid int)  
RETURNS raster AS  
  
$body$  
DECLARE  
scalx double precision :=pix_z;  
scaly double precision :=pix_z;  
box geometry;
```

```
_srid int;
result raster;
--easting and northing to snap grid:
ULX double precision;
ULY double precision;

BEGIN
_srid:= srid;
-- SRS origins. Here a small list, it can be extended by finding origin
  values int the srs_ref_table:

IF _srid =28992 THEN -- for RD NEW
ULX:=155000;
ULY:=463000;

ELSIF _srid =31974 THEN -- for SIRGA 2000
ULX:=500000;
ULY:=0;

ELSIF _srid =31974 THEN -- for Plate Carre
ULX:=0;
ULY:=0;

ELSE
RETURN 'No SRS Match definiton, USE: RD_New, SIRGAS_200 or Plate Carre EPSG';
END IF;

box:= obb;
result:= (st_asraster(box, scalx::double precision, scaly::double precision,
ULX, ULY,'64BF'::text,1,0));
--1 is the starting value, and 0 as no data.

RETURN result;

END;
$body$
LANGUAGE plpgsql IMMUTABLE STRICT
COST 100;

=====
-- OTHER FUNCTIONS --
=====
-----
-- FUNCTION: Log (Computing log probabilities) --
-----
CREATE OR REPLACE FUNCTION pedestrians.mo_Log(to_schema text, new_table
text, input_ras raster) RETURNS text AS
$$
```

```
DECLARE
_to_schema text;
_rast_table text;
result raster;
_new_table text;
_in_raster raster;
expression text;

BEGIN
_to_schema:= to_schema;
expression:= '1/(-1*log([rast]))';
_new_table := new_table;
_in_raster:= input_ras;

--Prepare the new table
EXECUTE format ('CREATE TABLE IF NOT EXISTS %I.%I (rid serial PRIMARY KEY,
rast raster, rast_id varchar)', _to_schema, _new_table);

EXECUTE format ('UPDATE %I.%I SET rast=st_setsrid(rast,28992)',_to_schema,
_new_table);

result:= ST_MapAlgebraExpr(in_raster::raster,1,NULL,expression);

EXECUTE format ('INSERT INTO %I.%I (rast) VALUES (%L::raster)', _to_schema,
_new_table, result);

RETURN 'Raster was inserted successfully';

END;
$$
LANGUAGE plpgsql VOLATILE
COST 100;

-----
-- FUNCTION: SumLog (Total probability computation) --
-----

REATE OR REPLACE FUNCTION pedestrians.mo_SumLogs(logs_dataset text,
to_schema text)
RETURNS void AS
$$

DECLARE
_logs_dataset text;
output_tbl text;
return_val text:='SUM';
_to_schema text;

BEGIN
_to_schema:= to_schema;
```

```
_logs_dataset:= logs_dataset;
--table name to store results:
output_tbl:='sum_'||_logs_dataset;

EXECUTE format ('DROP TABLE IF EXISTS %I.%I', _to_schema, output_tbl);
EXECUTE format('CREATE TABLE %I.%I (rid serial PRIMARY KEY, rast raster,
  track_id varchar)', _to_schema, output_tbl);
EXECUTE format('UPDATE %I.%I SET rast=st_setsrid(rast,28992)', _to_schema,
  output_tbl);
EXECUTE format ('INSERT INTO %I.%I (rast) VALUES ((SELECT St_union (rast, %L)
  FROM %I)::raster)', _to_schema, output_tbl, return_val,_logs_dataset);

EXECUTE format('ANALYZE %I.%I', _to_schema, output_tbl);

END;
$$
LANGUAGE plpgsql VOLATILE
COST 200;
```