Spatiotemporal Characterization of Scalable Data Analysis and Mobility Stochastic Modeling. Application for Urban Planning and Transport Infrastructure

Chibli Joumaa

Année 2010
UNIVERSITÉ DE TECHNOLOGIE
DE BELFORT-MONTBÉLIARD

No d'ordre : 138

École Doctorale *Sciences Pour l'Ingénieur et Microtechniques*

THÈSE

Présentée pour obtenir le grade de

DOCTEUR de l'Université de Technologie Belfort-Montbéliard

Discipline : **Informatique**

Spatiotemporal Characterization of Evolutive Data Analysis and Mobility Stochastic Modeling – Application for Urban Planning

Par

Chibli JOUMAA

Jury

Rapporteur	ALEXANDRE Frédéric, Directeur de recherche INRIA, Nancy
Rapporteur	PERRONNE Jean-Marc, Professeur, ENSISA
Président	PANSIOT Jean-Jacques, Professeur, Université de Strasbourg
Examinateur	BECHERIF Mohamed, Maître de Conférences, UTBM
Directeur	CAMINADA Alexandre, Professeur, UTBM
Co-encadrant	LAMROUS Sid, Maître de Conférences, UTBM

Remerciements

Le travail présenté dans cette thèse a été mené au sein du laboratoire Systèmes et Transports (SeT), de l'Université de Technologie de Belfort-Montbéliard (UTBM), sous la direction d'Alexandre CAMINADA et de Sid LAMROUS. Je leur adresse ma profonde gratitude pour m'avoir permis d'intégrer leur structure dans le cadre de mon doctorat.

Apres avoir fini ce travail je tiens à remercier l'ensemble des membres du jury : Fréderic ALEXANDRE, Jean-Marc PERRONNE, Jean-Jacques PANSIOT pour le grand intérêt qu'ils ont porté à mon travail.

Je remercie aussi Eric CORNELIS et Dominique BADARIOTTI qui via leurs critiques m'ont permis de progresser tant sur le fond que sur la forme de ce manuscrit.

Je tiens à remercier particulièrement, tous les membres de notre équipe qui m'ont tout le long de ces années aidé dans mon travail et dans mon séjour en France; spécialement Hakim MABED, Jean-Noël MARTIN, Oumaya BAALA, Alexandre GONDRAN, Laurent MOALIC, Mohammed BECHERIF. Pour moi c'était mon autre famille.

Un grand merci à mon Oncle Safouh Nour et sa famille: Fida, Marina, Dany et William.

Enfin, je remercie mon père Chaker JOUMAA, ma mère Alice NOUR, mon frère Fadi et ma sœur Fida, mes neveux : Sammy, Noura et Ramsey et bien sur ma fiancée Farah qui ont toujours cru en moi, m'ont soutenu et encouragé dans tous les choix de ma vie.

CONTENT

Chapter 1. Introduction

The work presented in this thesis report can be situated in a major environmental and economical context which is the planning of town and country's people and goods displacement. The rise in demand of dimensioning and locating habitats, economic social services, transportation and telecommunication infrastructures and services among other, makes it necessary to have precise and up to date information on people's localization and displacement. Up to now, such information was derived from static data covering geographical information (databases with ground height and ground occupation), population distribution (number of people per building, and their social identity - marital status, education level, occupation, etc) and economical information (location of shops, industry, schools, government offices…).

In France, these data are mainly provided by governmental institutions, such as IGN (*Institut Géographique National*) and INSEE (*Institut National de la Statistique et des Etudes Economiques*), and are locally tuned by regional services for regular up-to-date results. These data are used by most decision makers for urban and suburban planning of services and infrastructures. However, these data include only few descriptions of flows of mobility of people and continuous variations of these flows over time. As a matter of fact, it is only at a later stage that the councils in charge of planning are acquiring specific studies on local displacement of people, in order to optimize their planning.

Numerous types of information are collected: personal or professional displacements, whether individually or in a group, transport mode, start and end time, start and end locations, frequency, number of people crossing a given street or place, etc. Such information is referred to as *Enquête Ménage/Déplacement* if the data summarize the way people are moving, or *Enquête Cordon* or *Comptage sur Sol* if the data summarize the number of people or vehicles crossing a given area. Unfortunately, the acquisition of these data is costly (sensors on ground, investigations…), lengthy (statements over several weeks) and very selective (some streets or axes, some specific buildings and social identity). However, town and country planning, and in particular infrastructure and transportation services, must be based on real time data, which translate the dynamic of the area. This information is necessary to detect bottlenecks which can occur at any time and any location which are very often correlated by stochastic successions of phenomena which govern flows.

In 2002, GPS (*Global Positioning System*) was introduced as a new source of information for planning. The data gathered by GPS make it possible to retrace the path and trajectory of vehicles using this system. Even though some smart phones also include GPS but they could not be used for understanding pedestrians' mobility. This is due to the fact that they are in little number and pedestrians use them for very short trajectories (very often it is for the last 100 meters). Despite the great advantage of the GPS in collecting very precise displacements data, many limitations still exist. Firstly, GPS is limited to motorized vehicles (cars, bus and trucks), therefore displacement of bicycles and motorbikes cannot be tracked. Moreover, GPS data covers a small number of vehicles, most of the time for professional usage. This is due to the fact that GPS does not send any information by itself. It needs a subscribed connection of the vehicle to a cellular network (GSM, GPRS, EDGE, UMTS...) to track and collect the information and send it back to a specific information system. Another limitation is that information collected by GPS about the route taken by the vehicle cannot be used unless a prior consent from the driver (CNIL protection). However, GPS is limited to specific professional usage and service, and cannot be used for global studies of population displacement.

More recently, radio communication was also incorporated in country planning, by collecting information about cellular phone subscribers' displacement. The radio communication mobile networks allow people to move while communicating. These networks produce a new type of observable information about the radio signal strength of the receivers, which can be used to understand human mobility and thus to propose much more effective models for town and country planning. Contrary to the other means of geolocation data collection, the information gathered from radio communication networks represents a huge population since more than 85% of the population uses this type of communication artifacts in France [76][77][78]. Moreover, results from several studies [79][80][81][82][83][84][85][86][87][88][89] showed that in terms of mobility, people with or without mobile phones tend to have the same behavior.

Another important advantage of this geolocation data source is that the mobile network does not restrict the location or the displacements of people inside a geographical area. This is because, nowadays, the cell phones are covering whole towns and subscribers can move in any way they want and remain connected to the network. They move where they want and put and receive calls at any time and place. Another advantage of this type of geolocation information collected from the radio communication mobile networks is that it is collected continuously without requiring additional means in the mobile networks, which was not the case in the previous years. Nowadays, the operators have the possibility to record the information related to subscribers' calls such as transmitters' identity and transmitters' changes during calls. With the transmitter's location known, it is then possible to identify globally the area in which the mobile phone is located. The drawback of this method is that idle phones information is not collected; only information about calls is collected.

The use of information coming from radio networks, along with the information gathered by conventional collection procedures, gives us the possibility to conceive and evaluate, more efficiently, mobility models for any application whether it is for policy planning or displacements survey. In this context this thesis work aims to propose a new approach to identify, model and finally understand the spatial and temporal factors of the individuals' displacement. Starting from the analysis of statistical evolutionary data and stochastic processes (Markov chains, queues…), we will use information from mobile network, transportation network and mobility surveys to propose new kind of approach for mobility understanding. Such information offers an important complex richness to understand the frequency, the spatial distribution and the temporal distribution of individual presence and mobility flows in a geographical area, and for a large population. In addition, it is possible to take into account the social identity of the individuals in the studied geographical area for a better mobility understanding, to associate to the mobility the reason of that mobility.

For this work we use variables from various sources (radio communication, geographical, socio-economical) and from various time periods, and this is a new point for mobility modeling. The evolution of the data during time requires the handling and treatment of several matrices simultaneously which represent a huge amount of data. This is done with the objective of generalizing and adapting the traditional methods of the multidimensional data analysis originally conceived for a single matrix. Four axes of development will be considered in our data analysis. The first axe relates to the multidimensional methods of exploratory analysis which will allow us to create and to emphasize the relevant indicators to characterize the individual presence and possible displacement over one precise period of time. The second will relate to the methods of evolutionary data insofar as the phenomenon which we study is related to the temporal variable. The third will particularly treat evolutionary methods of classification in the data analysis. The fourth will model the mobility while taking into account the extraction of the information previously characterized and will drive us to a new approach. These axes will be developed and detailed throughout the report.

The thesis report is divided into six chapters with chapter one being the Introduction and chapter six being the Conclusion. Chapter 2 gives an overview of existing mobility models from the literature with a particular axis on stochastic approaches which are the more general. Some of these models were a source of inspiration to the proposal of our new mobility model. These models are divided into six classes: basic mobility models, individual motion mobility models, group mobility models, path driven mobility models, target driven mobility models and hybrid mobility models. Basic mobility models provide a non realistic motion pattern but are simple to implement. Individual motion mobility models are used to simulate the motion of one individual. The motion pattern for these models can be completely random and can take into consideration the terrain characteristics of the simulation area. Group mobility models are used to simulate community or groups' behavior and habits with different motion pattern. For the path driven mobility models the motion pattern is restricted to prefixed paths or predefined itinerary. These restrictions reflect the influence of the environment on the motions. The aim of the target driven mobility models is to drive the simulated individual to a fixed point at which the simulation ends. Finally the hybrid mobility models are composed of two (or more) other kinds of existing models. Implementing such models consists on dividing the simulation into several concatenated processes. Also we do an overview of some other categories of model which are not based on these principles. The Multi-Agent models which were developed for mobility simulation by the use of separated interacting process, the Constructal Theory which can model the shape and the structure of a system, and the Four-Step model uses to simulate individual travel behavior between different zones.

Chapter 3 presents the terrain characterization procedure. The work presented in this chapter was a part of the work done by UTBM, during the *Territoire Mobile* project from the cluster *Véhicule du Futur*. This project was funded by the territorial collectivities and SMTC (*Syndicat Mixte de Transport en Commun*), a company in charge of bus management in the *Territoire de Belfort*. The aim of this characterization is to propose a realistic simulation environment for mobility analysis based on real city information. This is done after collecting data over several days from different sources of information such as the number, the location and the time of calls in mobile network (data issued from Orange Company), the number of times people use the bus as well as the location and time of each usage and data from different local surveys on mobility. All these information are coupled with the geographical, topographical and economical information of the area. The final set of information is then analyzed statistically in order to extract and visualize the simulation area behavior and evolution over time. In this chapter, the theoretical explanation of the used statistical methods is given. Then the applications of these methods on different types of data along with the results are presented to build the valid realistic simulation area in which the new mobility model will be tested and compared to others.

In Chapter 4 the new mobility model called the Mask Based Mobility Model (MBMM) is presented. The mobility model is based on Markov chains, so a quick review of the basic mathematical principles of these processes is given firstly. The simulation environment for this model is a grid cell matrix, where each grid cell contains a certain number of information as described in Chapter 3. The motion pattern of this model is bounded by a mask covering the current cell used by the individual and its neighboring grid cells. This is the displacement mask. The Markov chain states are the grid cells constituting this mask. Then a detailed explanation of the different aspects of the model validation, such as direction choosing, speed… along with validation tests results for the MBMM main features are presented. This model will permit the simulation of individuals in the simulation environment we have developed.

Real applications of MBMM are presented in Chapter 5. The first one is a comparison of the new mobility model with the mobility models that inspired it. The comparison is made after simulating all the models on the simulation area and is done through several new quantitative metrics to evaluate mobility. The second application is the use of MBMM in order to simulate the mobility of subscribers in a UMTS cellular network. The aim is to use these mobility traces in order to make a comparison between three distributed power control algorithms for UMTS and the standard UMTS power control algorithm. Two categories of mobile users in two case studies are considered: pedestrians and vehicles, where a pedestrian is a mobile user walking in a city at 5km/h and a vehicle is a car at 120 km/h. In both scenarios several individuals are simulated along with the concerned individual to represent the jammers in UMTS interference computation. The simulation scenario is a grid cell matrix composed of roadways and building blocks where the COST-231 propagation model is used to map the coverage area. For each grid cell a value of the received power is stored according to the coverage signal intensity over the area. The comparison of the 4 algorithms using MBMM allows the planner to get a better confidence in the simulation results.

Lastly a general conclusion is given summarizing all the thesis components, the contributions and the results, and proposing some perspectives of models developments and new applications.

Chapter 2. Overview of Existing Mobility Models

This chapter gives an overview of a number of mobility models present in the literature. After a general introduction, a classification of existing stochastic mobility models is presented in six major subsections. The first subsection deals with basic mobility models. The second lists individual mobility models. The third one gives some group mobility models. Path driven mobility models are described in subsection four. Target driven are then given in subsection five. The sixth deals with hybrid mobility models. In the second part of this chapter, other modeling techniques to simulate mobility urban planning and transport, are presented. And finally a general conclusion summarizes the major elements discussed in the chapter.

Content

2.1 Introduction

Mobility models are used to simulate human motion behavior. They allow studying how the displacement of humans changes over time, taking into consideration their velocity, acceleration, objectives, constraints... Human motion behavior may be modeled by a common part and a specific part, varying from an individual to another or between different individual classes. Therefore, different classes of mobility models exist.

The study of mobility consists of working on a very large set of research areas. The area of interest for this thesis report is mobility in urban and suburban areas. The study aims to observe and represent the motion of individuals in areas with high population presence such as cities. There exist two major types of observable mobility: macroscopic and microscopic.

Several mobility models have recently been proposed in order to simulate the individual motion and population flow. These models each have specific characteristics taking into account the uniqueness of individuals, environment, displacement speed, etc. According to these characteristics, some models may give a better mobility modeling for individual displacement and/or population flow.

Some of these models are totally random and do not offer any realism in their motion pattern and are mainly used to simulate the mobility in the worst case as Random Waypoint Mobility Model [5] or the Brownian Model for which motion pattern is based on that of particles in the theory of quantum mechanics.

Other models however, show a certain consistency in the trajectories such as Normal Walk Mobility Model [3][4]. This model is based on the assumption that an individual who has just a point is more likely to proceed forward than to go back in direction. Thus the trajectories are less chaotic.

The Smooth Mobility Model [1] follows the laws of kinematics. Its main advantage lies in eliminating sharp corners and sudden changes of speed. Thus it adds more realism to the individual motion pattern; however it does not model the effects of mass (population movements) which is its main limitation.

Unfortunately, none of these models mentioned above takes into account the characteristics of the environment. Thus we may see people moving on a watercourse or deserted areas, which produce a non realistic motion pattern for individuals moving in a city or any urban area.

The Markovian Mobility Model is the first who took into account the terrain characteristics by assigning weights to grid cells constituting the simulation area. However, it does not take into consideration that individuals make frequent ways and backs or long stays in places relatively more attractive than neighboring areas.

The Scalable Mobility Model [17] is based on the concept that an individual possesses some particular places of attendance at a day appointed *poles of gravity*. Indeed, the typical day for each person is based on a limited number of locations (home, work, shops...) and the displacement of people consists mainly on going to one of those categories of places.

However, the majority of these models imitate the common part of the human displacement behavior, and few are sufficiently generic to really cover all the aspects of mobility. Often, these models, studying the general human displacement behavior are used for specific needs or specific applications and we will try to define a more generic approach issued from data analysis on samples of people location.

Within this chapter, we firstly define six different classes of mobility models according to their main features linked to the approach we will develop based on stochastic process. They will be presented, along with their advantages and drawbacks. Then we present other approaches more dedicated to transport problems.

The first type of models presented is the basic mobility models. Such models provide basic non realistic motion simulations. Two of these models are described: the Brownian Mobility Model and the Column Mobility Model.

Then, individual mobility models are presented which, as the name implies, deal with the individual motion simulations. Different models of this category are described: the Random Waypoint Mobility Model, the Normal Walk Mobility Model, The Semi-Markov Model, Random Direction Mobility Model… Some special cases of Random Waypoint Mobility Model are also introduced.

The third type of models is the group mobility models such as the Nomadic Community Mobility Model, the Reference Point Group Mobility, the Reference Velocity Group Mobility Model, the In-Place Mobility Model... Such models are used when dealing with a population or a group. They are based on the definition of common displacement features for several individuals.

Then, Path Driven models, where the motion pattern is restricted to prefixed paths, are presented. Examples of such models are introduced: The Random Trip Model, the Freeway Mobility Model, The Manhattan Mobility Model…

The Target Driven models are presented in fifth section. Such models are very similar to the path driven model, but with fixed itineraries. Some of the models discussed are: the Pursue Mobility Model, the Contraction Mobility Model, the Modified Contraction Mobility Model…

The last class of mobility models presented is the hybrid mobility models. Examples of such models are the Hybrid Contraction and RWP Mobility Model.

However, mobility is a very vast subject of research. Therefore, other methods and techniques than the one used in the previously described models exist for mobility in urban planning and transport simulation. Among these techniques, the Multi-Agent System, the Constructal Theory and the Four-Step Model are described in details in the following sections. These approaches are based on splitting the system that needs to be modeled, into small entities. For each of these entities, a detailed description of behavior, rules and conditions is given. The interaction between these elements will allow the observation of the desired behavior. Based on these observations the system parameters and rules may be adjusted in order to get the optimal solution for the specific application. Most of models for transport are used for the estimation of infrastructure performance. We like to emphasize that our work is not dedicated to this kind of evaluation and is more linked to the models use for simulation in the case of communication networks where location of people is more important than car or people trajectory.

2.2 General Classification of Mobility Models

In this section a general description of major mobility models existing in the literature for our case study is presented. For a clearer understanding of their concepts and characteristics, the mobility models will be classified according to specific features. .

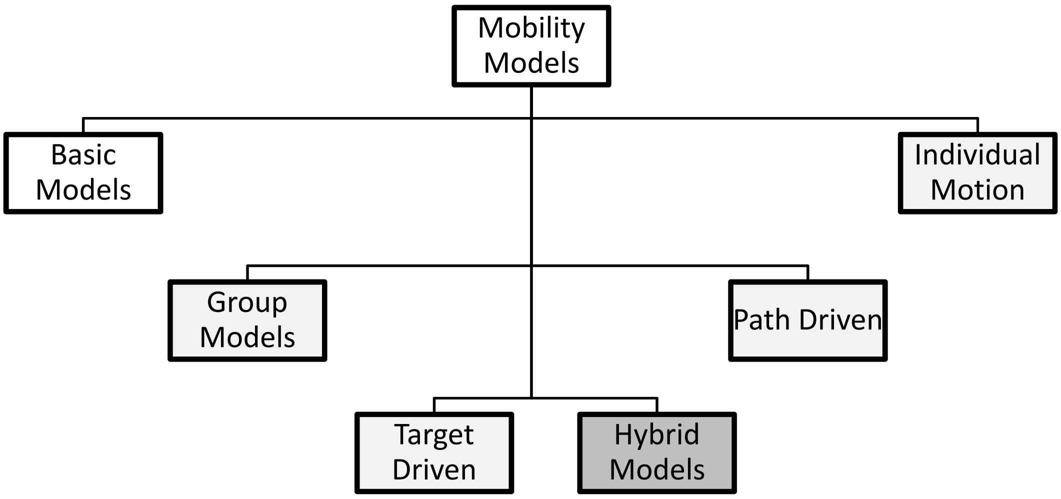


Figure 2.1. General mobility models classification

We have distinguished six classes: basic models, individual motion, group mobility, path driven, target driven and hybrid mobility models. This mobility model classification is presented in Figure 2.1.

2.2.1 Basic Mobility Models

Two basic mobility models, quoted in [8], are presented in this section. These models provide basic non realistic motion simulations. Figure 2.2 gives a listing of the models presented in this section.

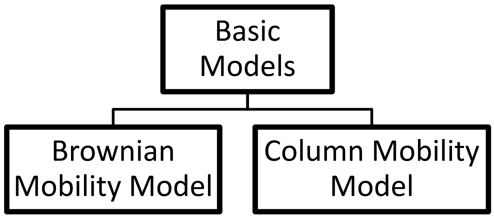


Figure 2.2. Basic models

The first model in this category is the Brownian Mobility Model. This model is a very basic model, which makes its motion pattern a non realistic one. The main advantage of the Brownian Mobility Model is its implementation simplicity; it is based on the following simple mathematical formula:

$$NP = OP + RV \tag{1}$$

Where NP is the new position of the individual, OP is the old position of the individual and RV is the chosen random displacement vector.

An individual starts from a position at a given point, then, in order to move to a second point, he randomly chooses a vector of displacement (angle and distance) which determines its next position. Therefore, the new position is the summation of the old position with the displacement vector.

The second and last model in this section is the Column Mobility Model. The aim of this model is to represent the motion pattern of a group of individuals aligned in a row, and moving in a given direction. The choice of direction is totally random.

In this model, the chosen direction is mainly represented by the chosen angle in $[0, 2\pi]$. This model is applied by defining a reference initial grid and then by implementing a motion pattern that performs movement around this fixed reference origin.

$$NP = NRP + RV \tag{2}$$

$$NRP = ORP + AV \tag{3}$$

Where NP is the new position of the individual, NRP is the new reference point, ORP is the old reference point, AV is the group advance vector and RV is the individual random displacement chosen vector.

2.2.2 Individual Motion Mobility Models

The second mobility models class deals with the individual motion simulations. Many models will be presented in this section. However, there is a few numbers of original models, and the rest are just adding small variation to the original model, to overcome a specific problem in motion simulation.

The first model presented in this category is the Random Waypoint Mobility Model [5] [6] [7] [12] [14]. The displacement in this model is done for a fixed time *t* or for a fixed distance *d*. Given a starting point, an individual chooses his destination point, direction and speed. The direction angle is uniformly distributed between 0 and 2π. Speed is *normally* distributed between 0 and *Vmax,* the maximum speed that can be taken by an individual in motion. After choosing the direction angle and speed the individual moves to his target point with constant speed, with no acceleration or deceleration.

The direction and speed choice are independent from each others. However, the individual in motion maintains its velocity to the chosen velocity value until the destination point is reached or as long as the direction is not changed.

At each destination point exists a pause time called *Tpause*. It is a random number describing the number of units of time the moving entity will have to stay at its current location. *Tpause* marks direction change.

The main drawback of this model is that it has no memory; past information, as direction or speed values chosen before, are not taken into consideration. This may produce a sudden important change in speed and/or direction angle. A detailed study of this model and its drawbacks can be found in [15].

The Random Walk Mobility Model [9] is a special case of the Random Waypoint Model. In this mobility model, an individual chooses the direction angle, defining the direction he is going to take, as well as the duration and speed of the trip.

The Random Walk Mobility Model with Reflection [9] is nothing but the classic Random Walk Mobility Model with modified border effect rule. When an individual reaches the boundary of the studied area, instead of being wrapped around, the individual is automatically redirected or the path is reflected back into the simulation area.

The Restricted Random Waypoint Mobility Model [9] is also a special case of the Random Waypoint Mobility Model where the simulation area is divided into several sub-domains. An individual moves from one sub-domain to another by Markov walk. Inside these sub-domains the motion pattern is that of the classical Random Waypoint Mobility Model.

The Space Graph Mobility Model [9] is a particular case of the Restricted Random Waypoint Mobility Model. The studied region is a collection of graph vertices which can be grouped to form a space graph. These graph vertices are associated to each point in the simulation zone space. A connectivity matrix is defined as edges between these vertices. So the domain is nothing but the union of the line segments defined by the graph edges. The path to move from vertices to another is the shortest path linking these two vertices.

Another individual motion model is the Semi-Markov Model [20]. A semi-Markov process is used because it allows for arbitrary distributed simulation times for each individual.

In [21] the author uses an m-th order Markov model to simulate mobility. The simulation area is divided into cells. The individual motion is described by a series of indices, *C1, C2… Ci…* where *Ci* denotes the identity of the visited cell. Since the future individual locations are probably correlated with his motion history, the sequence of cell identifier indices *C1, C2, C3… Ci…* is assumed to be generated by an m-th order Markov source. The probability that the individual passes from a cell to another cell depends on the location of the current cell in which he resides and the list of previously visited cells.

In [22] the authors propose a new mobility model for urban areas. The model contains motion channels computed by the multiple applications of Voronoï graphs. The resultant simulation area is then combined with a random-based mobility algorithm. The special points about this original model are the following:

- Node distribution: Nodes are uniformly distributed within the Voronoï channel area.

- Movement: Individuals are allowed to move within the channels. When the simulation starts, individual chooses to move with an initial speed v within the interval [0; *maxspeed*] and a given angle θ within the interval [0; 2.25]. The motion then continues for a given time t.

- Transition: There is a given probability for entering and leaving the Voronoï channels. This probability is expressed by the variable p_{ch}, this variable can take any value within the interval [0; 1]. When the p_{ch} value is greater than a predefined threshold probability $p_{chthres}$, the channel zone can be left. When $p_{chthres} = 1$, the individuals have to stay within the channel areas during the whole simulation.

- Pause Time: An individual can stop moving at a random point of time for a random waiting period. After this waiting period is over, motion is continued with a new angle θ and a new speed v.

This model can be extended and refined. Many parameters can be added such as street width and other factors. This will add more realism to the individual motion pattern.

The Random Landmark Model described in [23] is very close to the random waypoint model with the difference that destinations are randomly picked from a predefined set of locations instead of from the whole simulation area. When an individual in motion reaches the chosen destination, it stays there for a fixed time period; once the pause time is over the individual then chooses another destination from the set of locations and moves straightly towards it at a speed uniformly distributed between V_{min} and V_{max} which are respectively the minimum and maximum values of the displacement speed defined for the simulation.

Another well known individual motion mobility model is the Normal Walk Mobility Model [3] [4]. In this model, direction is defined by a drift angle normally distributed with 0 mean and standard deviation between [5°,90°]. This model is based on the concept that the majority of the steps taken follow the shortest path. This is done in order to add more realism to the movement pattern

Starting from a grid map with square or hexagonal cells, four or six directions are defined in this model respectively. Drift angles define the direction of displacement. The cell sides define the intervals of direction angle, six intervals for hexagonal cells and four for square ones.

If the standard deviation of the Normal Walk Mobility Model is equal to 71°, a special case scenario takes place: the Normal Walk becomes a Random Walk Mobility Model.

The major drawbacks of this model are sharp turns. This problem is solved in the Smooth Random Mobility Model [1] [2].

In comparison with other mobility models existing in the literature, the Smooth Random Mobility model contains two new major added features: acceleration, and the fact that the choice of velocity at each step depends on the velocity chosen at the previous step. In order to simulate real human behavior speed and direction change are incremental and not sudden producing a smoother motion pattern for the individual in motion.

From a given starting point a destination point is chosen anywhere on the studied map. Destination is characterized by the drift angle which is chosen in an interval between 0 to 2π. The displacement in this model is not done from one cell to another but rather by a vectorial displacement approach.

Speed is chosen at the beginning of the individual journey, and then speed is changed for each new destination point by drawing a value of the acceleration from a given interval.

The major advantage of this model is that all sharp turns are eliminated. The idea is to calculate the difference of the angle chosen at the current instant t and the previous instant t-1. If this difference is greater than the given threshold, the difference is divided by the threshold. This will give us the number of steps that the individual in displacement will have to do to compensate the sharp turn step.

Another type of models is the Random Direction Mobility Model [7]. Such a model forces mobile nodes to travel to the edge of the simulation area before changing direction and speed.

This model was created to solve the density wavers problem caused in the Random Waypoint Mobility Model simulations. A density wave is defined as the cluster or group of nodes in the simulation area, which is generally observed in the center of the simulation area for Random Waypoint.

In this model the traveling nodes or individuals choose their speed and direction randomly. When a node reaches the border of the simulation area following that direction, it stops for a specific pause time, and then chooses a new direction. Direction is chosen between 0 and 180°.

One popular mobility model is the Networked Game Mobility Model (NGMM) [28]. It is used to simulate mobility in First-Person-Shooter (FPS) network games. NGMM is an extension of Random Waypoint Mobility Model. It differs with what the authors call enhancements performed at two levels, macro and micro.

The macro-level enhancement is the popularity, where it simulates and mimics popular locations in the game, and the choice of next location are introduced.

The micro-level enhancement is the addition of noise in the stationary and in motion distribution direction choice states. The simulation results showed the NGMM flexibility to model different aspects of individual motion.

In all models described so far, an individual in motion is stopped or reflected once he reaches the boundaries of the simulation area. However, in the Boundless Simulation Area Mobility Model [7] the behavior is different.

In this model, motion is based on a relation between the previous chosen direction and the travel velocity of an individual. Velocity is described by the following vector $\bar{v} = (v, \theta)$. Given the individual's position by *(x,y)*, the position and velocity are updated every Δt by the following equations:

$$\mathrm{v}(\mathrm{t}+\Delta\mathrm{t}) = \min\left[\max(\mathrm{v}(\mathrm{t})+\Delta\mathrm{v}, 0), \mathrm{V}_{\max}\right] \tag{4}$$

$$\theta(\mathrm{t}+\Delta\mathrm{t}) = \theta(\mathrm{t}) + \Delta\theta \tag{5}$$

$$\mathrm{x}(\mathrm{t}+\Delta\mathrm{t}) = \mathrm{x}(\mathrm{t}) + \mathrm{v}(\mathrm{t}) * \cos\theta(\mathrm{t}) \tag{6}$$

$$y(t + \Delta t) = y(t) + v(t) * \sin\theta(t) \quad (7)$$

Where *Vmax* is the maximum defined velocity, Δv is the velocity change uniformly distributed between *[-Amax* Δt, Amax* Δt]*, *Amax* is the maximum acceleration of a given individual, $\Delta\theta$ is the change in direction uniformly distributed between *[-* $\alpha * \Delta t$, $\alpha * \Delta t$*]*, where α is the maximum angular direction change in the individual motion.

The Gauss Markov Mobility Model [7] is also classified as an individual mobility model. It is characterized by a tuning parameter making it capable of adaptation to different levels of randomness. An individual starts his journey by an initial speed and direction. For a fixed number of time steps *n*, the displacement occurs by updating the speed and direction according to the following equations: $s_n d_n$

$$s_n = \alpha s_{n-1} + (1 - \alpha)\bar{s} + \sqrt{(1 - \alpha^2)} s_{x_{n-1}} \quad \textbf{(8)}$$

$$d_n = \alpha d_{n-1} + (1 - \alpha)\bar{d} + \sqrt{(1 - \alpha^2)} d_{x_{n-1}} \quad \textbf{(9)}$$

where s_n and d_n are respectively the new speed and direction, α is a tuning parameter ($0 \leq \alpha \leq 1$), this parameter controls the randomness, $\bar{s}$ and $\bar{v}$ are the mean speed and direction values as $n \to \infty$, then $s_{x_{n-1}}$ and $d_{x_{n-1}}$ are random variables from a Gaussian distribution.

At each time step the next location is computed taking into consideration the current location along with speed and displacement direction. This motion position update is given by the following formulas:

$$x_n = x_{n-1} + s_{n-1} \cos d_{n-1} \quad \textbf{(10)}$$

$$y_n = y_{n-1} + s_{n-1} \sin d_{n-1} \quad \textbf{(11)}$$

Where (x_n, y_n) is the individual position coordinates at instant *n*, (x_{n-1}, y_{n-1}), s_{n-1} and d_{n-1} respectively represent the individual position coordinates, the displacement speed and the displacement direction at instant *n*-1.

When using this model individual is forced away from the simulation zone edges and borders by modifying the value of the average mean $\bar{d}$.

Another mobility model is the Probabilistic Version of the Random Walk Mobility Model [7]. This model uses a probability matrix to determine the individual position for the next time step. This is represented by three different states for the x position and three different states for position y.

The 0 state is the current position, the 1 state represents the previous positions and the 2 state represents the next or future position for x or y.

$$P = \begin{Bmatrix} P(0,0) & P(0,1) & P(0,2) \\ P(1,0) & P(1,1) & P(1,2) \\ P(2,0) & P(2,1) & P(2,2) \end{Bmatrix} \tag{12}$$

These values are in a transition probability matrix that is used to update both x and y position values. The matrix entries are $P(a,b)$ which is the probability to go from state a to state b and the transition probability values are given by a flow chart.

The last model of this class is the City Section Mobility Model [7]. In this model the simulation area is composed of a street network representing a city section where the individuals are moving. Streets and speed limits are predefined based on the simulated city model.

An individual begins his journey at a defined point on a given street. Then, he chooses a destination he is willing to reach which is another point on a street. The motion pattern that the individual follows to get from the origin to the destination point is based on the shortest travel time between these two points. When the individual reaches his destination he pauses for a specified time then randomly chooses another destination once the pause time is over.

2.2.3 Group Mobility Models

In this section the state of art models dealing with large or small population group are presented. In opposite to previous models, these one are in charge of several motions at the same time.

The first model we introduce is the Nomadic Community Mobility Model [8]; it is based on ancient human habits and way of life, still present in these days in certain regions of the terrestrial globe.

The community, represented by a group of points, moves around a reference point. A reference point represents the center of this community. This community moves from a position to another following the displacement of its reference point. When this group settles in the new position each point of this group creates and maintains a private area in which this individual moves randomly. This model gives a sporadic motion pattern. This walk is defined by:

$$NP = RP+RV \tag{13}$$

Where NP is the new position, RP is the reference position and RV is the random vector.

Another model based on reference is the Reference Point Group Mobility Model (RPGM) [10] [11] [12] [13] used to represent a group of individuals motion pattern. For each group, a logical reference center is defined that can be described as the group leader [11]. All individuals represented by mobile nodes are uniformly distributed around the group leader. The logical reference center is a point whose movement is followed by all nodes in the group. The central point is characterized by its position, the group motion and node-dependent random motion vectors. The group motion vector gives a mapping of the reference centers' location. The node-dependent random motion vectors, added to the group motion vector, give the positions of the nodes.

Also the Reference Velocity Group Mobility Model (RVGM) [10] simulates group motion. The group particularity on which this model is based is the similarity of movements of different individuals or members of the group. V is the velocity component of these individuals displacement. This velocity component is divided to two parts Vx and Vy which are respectively the velocity components in the x and y directions. This mobility model can be seen as the RPGM model with the characteristic group velocity.

Another group mobility model is the In Place Mobility Model [13], it is an alternative of the RPGM mobility model. This model divides a geographical given map into several equal size partitions. In each region there is a group of individuals. Different groups do not necessarily have the same behavior or motion pattern.

In the Overlap Mobility Model [13] the entire population is divided into different groups, each group having a specific objective to fulfill. Each group has its proper behavior and it makes the mobility motion pattern different for each group. Each group is characterized by its motion pattern, speed and object.

The SLAW mobility model presented in [24] captures the real human mobility patterns found in mobility traces. Both analytical and empirical information are used to show that the individuals' motion can be very well expressed with the use of gaps among fractal waypoints. The SLAW mobility model gives mobility patterns for peoples with common interests' motion traces or within a single community where people tend to share common gathering places.

Another group mobility model is presented in [25]. This model takes into consideration the changes in displacement velocity, which gives a better representation of the individual motion. The variation in velocity over time is called acceleration. Each group is characterized by a specific group velocity $V_i(t)$ and acceleration $A_i(t)$. Group velocity is defined as the mean velocity of the individuals within a group. The member nodes in the group have velocities close to the group velocity but deviate lightly from it. Group velocity $V_i(t)$, group acceleration $A_i(t)$ $)=dV_i(t)/dt$ and local velocity deviation $L_{i,j}(t)$ are random variables following any type of distribution, Normal, Uniform, etc. $N_j(t)$ is the node velocity of individual j at time t given by:

$$N_j(t) = V_i(0) + L_{i,j}(t) + A_i(t) \tag{14}$$

This mobility model has several advantages such as the mobility motion pattern and parameters for each mobility group.

The mobility model presented in [26] is based on the analysis of real motion traces. The trace data was collected from a military context in the U.S.A. The particularity is that the traces are on various level, infantry and aircrafts. This model takes into consideration interaction between different levels of mobile traces. In the model, individuals in motion are divided into many groups traveling along the same route but with different time schedules. When a group reaches its destination, it pauses for some time and then continues his journey towards the following destination. If there is no more predefined destination in the group destination queue, a new destination will be generated. Then this info will be spread to inform other groups. Thus there are communications between groups.

At last, the simulation area is divided into sub domains in the Convention Mobility Model [13]. In each sub domain there exists a group of individual carrying out a given objective. Each objective is presented by a specific motion pattern. In this model groups can migrate from a sub domain to another, which makes the domains interconnected.

Finally, there are less group mobility models than individual models. Modeling group motion is more complex. In that case it is necessary to define common behavior and interaction between individuals inside group. Very often, group mobility is simulated without common characteristics i.e. using several time individual mobility models.

2.2.4 Path Driven Mobility Models

In this mobility models category, the motion pattern is restricted to prefixed paths. These restrictions have for purpose to show the effect of the individuals neighborhood influence.

The Random Trip Model [9] is the first model described for this section. This mobility model is used to simulate individual and non group movements. The main characteristics of this mobility model are: a domain ***A*** composed of a set of paths, an initializing rule and a trip selection rule. According to this mobility model, a trip is defined as the combination of the given path and its duration where duration may be seen as a numerical speed factor for which the whole path is fulfilled.

An individual chooses a path according to the initializing rule. When this individual is gone along the chosen path, another path is chosen, but this time the choice is done according to the model's trip selection rule. The starting point of the new path is the ending point of the previous chosen path.

It can be noticed that this model is very general and that special cases or applications of this model can be seen as simulation of other mobility models such as Random Waypoint Mobility Model and Random Walk Mobility Model and alternatives of these models.

Another model in which geographical restrictions are very clear is the Freeway Mobility Model [11] [12] [14]. The studied region is divided into freeways on which individual displacement takes place. Individual moves on these lines so there is no real direction choice. Individuals choose their speed randomly. At each step the individual speed depends on that chosen in the previous step. Two individuals, close to each others, maintain a certain distance called Safety Distance. Furthermore the velocity relationship between two individuals is that the velocity of an individual cannot exceed the velocity of the preceding individual. Freeway Mobility Model is characterized by spatial and temporal dependences as well as geographical restrictions.

The Manhattan Mobility Model [11][12][14] is another path driven mobility model. A network of roads constitutes the simulation area. The network of roads is constructed by a grid of horizontal and vertical lines. An individual moves straight forward on a road without changing direction until reaching the intersection where the probability to keep moving straight forward is equal to 0.5 and the probability to turn left or right is equal to 0.25. At each step the individual speed depends on the one at the previous step as in the Freeway Mobility Model.

In the Expansion Mobility Model [14], individuals move from the logical center towards the simulation area edges and limitations. The individuals are uniformly distributed in the simulation area. It is preferred to restrain the initial individual distribution area to a small area around the logical center. Simulation time is divided into equal time intervals. At each time step a new destination is chosen with a randomly chosen speed within [*1, Vmax-1*]. Pause probability and pause time give the chance for an individual in motion to stand still for some time. The individual motion ends when he reaches a simulation area border.

The Circling Mobility Model [14] is a restricted version of the previous model. In this model individuals circle around a logical center. Each individual has a specific circle characterized by a unique radius. The individual moves along this circle. All circles have the same origin which is the logical center of the simulation area.

In the Street Unit Model [19] the individual moves on a rectangular Manhattan like grid only. The grid represents the street pattern of urban and/or suburban zones. Displacement direction is given by dx and dy which are respectively the distances between crossroads in x and y directions. Speed follows a Normal distribution and is either area dependent or updated at each time period. At every crossroads, each moving individual can choose to change his displacement direction. Each direction change is weighted by a probability value; so at each crossroads there exists up to four different direction change probabilities.

Another mobility model in this category is the Street Pattern Tracing Model [19]. In this model the individual moves on a predefined stretch only. This stretch represents a main street or a highway in which the probability for an individual to change his displacement directions is very low. When simulating this model the streets are represented by a set of consecutive straight lines. Speed in this model is chosen randomly either from a Uniform or from a Normal distribution.

The models we have seen in this section are more or less dedicated to vehicle motion. The speed parameters allow them to be adapted to any kind of vehicles. They are using graphs to define path mobility.

2.2.5 Target Driven Mobility Models

The models presented here are very similar, in the concept, to the path driven mobility models with one add on: there are not only predefined paths but also predefined itineraries. The goal of these model simulations is to carry the individual or group of individuals to a fixed target which, most of the times, ends the simulation when reached.

One of these models is the Pursue Mobility Model [8]. In this model one moving node is defined as the target to be reached. So despite the target point itself every moving point moves toward the target point. The target pursuit is defined by:

$$NP = OP + A(TOP) + RV \qquad \textbf{(15)}$$

Where *TOP* is the target old position and *A* is the acceleration function. *NP* is the new position and *OP* is the old position. Acceleration of the previous target position along with the random vector RV helps differentiating the motion pattern of each node in its pursuit of the target.

In the Contraction Mobility Model [14], another target based model, individuals move to a logical center (for example the center of the simulation area). Individuals are uniformly distributed over the simulation area. The motion pattern of individuals is very simple: the individuals move in a straight line towards the logical center. Simulation time is divided into equal time intervals. At each time interval the individual randomly chooses his new displacement speed within [*1, Vmax-1*]. Pause probability and max pause time permit to pause for one or more time intervals. An individual motion stops when he reaches the logical center.

An alternative of the previous model is the Modified Contraction Mobility Model [14]. This model adds some modifications to the Contraction Mobility Model [14]. The aim of these modifications is to add a degree of realism to the individual motion pattern. In this version, for each time interval, a square is drawn between the individual and the logical center. A destination point is chosen within this square. Speed for each time interval is chosen randomly within [*1, Vmax-1*].

2.2.6 Hybrid Mobility Models

The last section in the mobility models overview is about the hybrid models. Hybrid models are composed of at least two existing models. Implementing such models consists on dividing the simulation into several concatenated processes launched one after the other, when a fixed condition is satisfied.

Firstly we introduce the Hybrid Contraction and RWP Mobility Model [14]. In this model the simulation area is divided into two interconnected areas. The first one is a circle of center the simulation areas' logical center. The second area is the rest of the simulation area. In the first area is applied the Contraction Mobility Model [14]. In the second area Random Waypoint (RWP) Mobility Model [5] [6] [7] [12] [14] is used. Individuals are uniformly distributed throughout the simulation area. Simulation time is divided into time intervals. Pause probability and pause time characterize the rest time that may occur each time interval.

If at the beginning of the simulation individuals are placed inside the circle, they will move in a straight line towards the logical center. The individual trip ends when he reaches the logical center.

Otherwise, if the individuals are located in the second simulation area, the motion pattern of these individuals will follow the RWP Mobility Model till the moment they enter in the first area. In this case their motion pattern will change and it will be that of the Contraction Mobility Model.

The Hybrid Manhattan and Random Waypoint (RWP) Mobility Model [14] is another hybridization. The simulation area, for this model, is composed of horizontal and vertical streets. Between each two vertical and horizontal streets there is a building block represented by a square.

Individuals on the streets follow the Manhattan Mobility Model [11][12][14] and individuals inside the building blocks follow the RWP Mobility Model [5] [6] [7] [12] [14]. The individuals that are inside the building blocks have a probability to get to the street, and individuals on the streets have a probability to get into the buildings.

The Virtual Track based group mobility model [18] is used to simulate mobility patterns of military and/or urban MANET scenarios. This is a heterogeneous model because it handles several mobility motion patterns. It simulates groups and solitaire individuals in motion as well as static individuals. It models the concepts of group moving and group merging or splitting. Virtual tracks restrain the group mobility motion along given tracks. These groups can split or merge at *switch stations* or *crossing* between several tracks. Individual's motion pattern is not the same as the group motion pattern. Their motion is not limited to some tracks as for groups.

Another hybrid mobility model, called Scalable Mobility Model (SMM), is described in [17]. It makes use of topographical data about the area and, as its name suggests, is capable of working for different distance scales, which seems to be a great advantage. It is probably the more powerful but the more complex model.

In comparison to other models, SMM introduces a few concepts adding more realism to the simulated motion:

- Classes of Mobility: define individuals with the same mobility characteristics (e.g. Business, Residential)
- Area Zones: geographical locations that can be scaled to different granularities (e.g. urban/rural or working/residential/streets)
- Attraction Points: locations attracting users from a particular class of mobility (e.g. shopping, working)
- Time Periods: in which various motion patterns appear

However, the model's versatility is based first of all on its hybrid structure. SMM consists of three integrated sub-models:

- Physical model: It is responsible for constructing area zones (i.e. possible locations of subscribers) and connecting them into a network on the basis of the topographical data. It also characterizes these zones as attraction points of a specific type. The initial distribution of population within the zones is also calculated by the physical sub-model.
- Gravity model: The role of this sub-model is to assign users to the classes of mobility. It also calculates the steady-state values which ensures stability of the whole model and decides upon individuals' behavior by determining inter-departure times.
- Fluid model: The population motions are obtained explicitly from this sub-model in the form of a transition probability matrix. Factors such as: attraction values, distance, and population intensity in a particular location are taken into account.

The hybridization of models makes sense as it is not possible to associate the same behavior to all individuals if we want to simulate motions with different social and topographical context.

2.3 Other Approaches for Mobility Modeling

The models described previously in this chapter simulate the individual or group motion in space or on a graph taking sometimes into consideration the terrain characteristics of the simulation area and mainly based on stochastic process. Other approaches are also used in the literature with other fundamentals. In this section we present three other main approaches in the domain of mobility modeling: Multi-Agent Systems (MAS), Constructal Theory and the Four-Step Model.

2.3.1 Multi-Agent Systems

One of the existing approaches in modeling mobility and displacement is the Multi-Agent Systems (MAS). The models described using MAS are not simply simulations of individuals in a given geographical area but a general description of all existing interactions between all elements forming this geographical area. The whole environment which is usually a geographical area (city, state, etc) is modeled, with more or less details depending on the application for which the model is being conceived. The environment is composed of several agents, each representing an existing entity in the real environment to be reproduced in this model.

Ferber defined the environment architecture [90][91] with three principles: respect the environment integrity, the agent autonomy and distinguish the agent's mind from its body. Within an entity, a mobile agent refers mainly to a dynamic individual or a group of individuals. Mobile agents are active and communicate with other agents. These mobile agents can represent pedestrians as well as cars or any entity in motion in a city. A set of rules, descriptions and strategies are defined in order to represent the interactions between the different agents constituting the whole environment of simulation.

The MAS are used as a tool to represent the dynamic aspect of the interactions between different elements constituting the city and their mobility. In urban and regional planning, this type of models was proved to be well suited to model the behavior at the individual level, including the processes leading to the daily mobility and daily pattern of the mobile agents, but the major drawback is that all rules between agents have to be defined a priori and it is sometime very complex to correctly define the deterministic and stochastic parts of the agents.

Anyway the MAS is a candidate to model complex systems. It solves multidimensional problems like the urban transportation and displacement problem. Such problems have spatial, social, political, organizational, economical and financial considerations and dimensions. For this kind of problems, the solution does not result from a programmed method or algorithm but it emerges from interactions between mobile agents representing the individuals and the other agents representing the simulated environment. The advantage of the MAS is that it helps and gives the chance to understand real situations from the behavior of agents and interactions between each others. Several models using MAS for urban mobility and social interactions exist in the literature. In the following subsections, we will present static MAS models and dynamic MAS models that deal with urban mobility and social interactions.

2.3.1.1 **Static MAS Models**

An example of static MAS model is DSCMOD [92]. This model designed by David Simmonds [107] aims at integrating the transport model characteristics. This allows visualizing the changes that occur in a city or any urban area, caused by the variety of the existing transportation systems. In this model the configurations of urban space and transportation system are assumed to be known. The changes result from the use of multitude of transport systems from the accessibility produced by alternative transport strategies. Any variation of the parameter of accessibility to each and every transportation system, affects not only the system, but the other systems as well, and hence the whole city. All parameters of the problem must be clearly defined.

2.3.1.2 Dynamic MAS Models

In this section we focus on dynamic MAS models taking into account the spatial aspects and economical activities.

The first model in this category is MEPLAN [92]. In this model, the area is divided into zones of different sizes. The choice of location, mode of transportation and distribution over zones are determined from a structure of choice based on multilevel random utility. The behavior of localization of households and businesses are based on competitive markets. MEPLAN has been used in planning studies at all spatial scales. In most of these studies, the complete model of interaction between transport and urban development has been used. The MEPLAN model considers the housing market and its influence on the location of the population. It uses a bid-rent theory for the individuals to select their residential locations as a compromise between the quality of the residence location and their relative transportation costs. Once this procedure is done, the resulting data is used as input for a four-step transportation / land use model, that takes into consideration the retroactive effect of congestion, trip generation, trip distribution and residential location. The Four-Step Model is described in further section.

TRANUS [92] is an integrated model using MAS that is applied to cities. This model has two major goals: simulate the likely effects of planning policies and transportation projects on the city economical activity, and evaluate these effects on a social, economical and financial scale. The model incorporates several modeling techniques and theoretical foundations. In order for these techniques to be applied the TRANUS model requires the availability of databases with matrices containing all required information for the whole year. The model is composed of two major parts. The first part deals with the economical actors in a city, based on the demand of lands and services and their influence in the city. The aim is to minimize the transportation and house rent costs. The second part of the TRANUS model deals with the transportation network. Origin-Destination matrices are generated and assigned to the transportation the network. For each origin-destination pair, a non utility value of the transport network and services is calculated taking into account the generalized costs on each arc and for each mode. This disutility transport defines the demand for the transportation service.

The last model described in this section is URBANISM [93]. This model was developed in the United States to meet the intense need of federal governments to link planning of land to the use of transportation in the urban environment. It was also developed to monitor the effect of growth, such as urban congestion, housing, etc. URBANISM is a discrete choice model based on the theory of maximizing random utility. Its structure is integrated and based on the estimation of changes occurring in the city over short period of time. It determines the market demand of properties in each location and the actors and processes that impact the choices of urban space configuration and real estate market. URBANISM allows explicitly incorporating behavioral policies and rules to the modeled environment and then assessing their impact.

2.3.2 Constructal Theory

Adrian Bejan defined a novel principal deterministic geometric structure of natural systems. This principal is the Constructal Theory (CT) [94][95][96][97][98][99][100][101][102][103]. The CT assumes that the constraints and objectives of engineering systems are also those that govern the geometry of flows in nature. It aims at delivering a new way to ideally design objects, houses, machines, networks, etc. The theory for this modeling approach is not based on fragmentation but on construction and optimization. By organizing the scales in space, from the smallest to the largest, CT becomes much more "natural" than the fractal geometry. Rules and settings to model social networks and urban traffic are described in [97].

2.3.3 Four-Step Model

This model is used to simulate individual travel behavior between different zones for a given area of interest. The Four-Step Model (FSM) [104], as its name indicates, is divided into four steps: traffic generation, trip distribution, mode choice and trip assignment. In the following we will explain each step of this model.

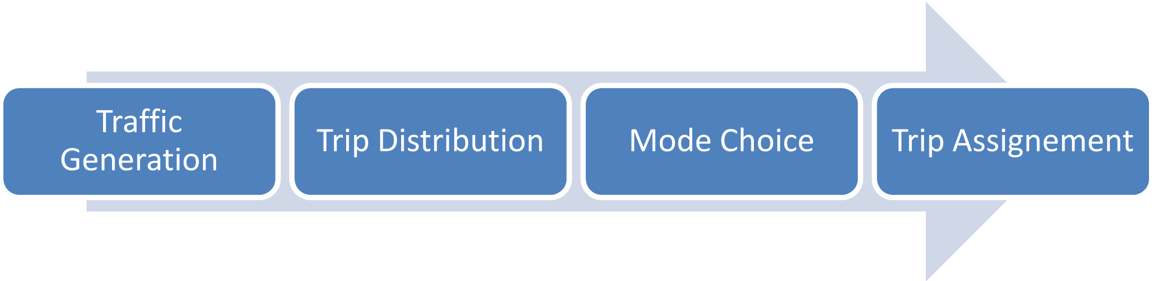


Figure 2.3. The Four-Step Model

The first step, “traffic generation” aims at quantifying the flow of individuals entering and leaving each zone of the area of interest. The idea is to consider all individuals possible displacement trajectories; this data can be detailed in an Origin/Destination matrix where each row represents an origin and each column represents a destination, and the elements of the matrix are the numbers of individuals that travel from their corresponding origin to their corresponding destination. The displacement trajectory is an Origin/Destination association, called trip.

For each trip two ends are defined: production end and attraction end. The production end is the trip’s starting point. And the attraction end is the area to which the individual will be attracted, in other words, the destination. The result of this step is an estimation of the number of production and attraction trip ends for each zone. The next step, “trip distribution” will create actual trips by matching up all the trip ends identified in the first step.

In the “trip distribution” step, the produced trip ends will be linked to the attracted trip ends, forming a list of complete trips. Several models exist for defining the link between two trip ends, production and attractions. Among these models we have for example the Gravity Model [105], which is intended to predict and estimate the geographical link between the trip ends. The model is inspired by Newton's law of universal gravitation: two bodies attract in proportional to their masses and inversely proportional to the distance between them. The mass for our case can be considered being a social or economical factor like residence area, shop, etc.

The link definition process, of two trip ends, relies on the general assumption that the more distant the destination is, the less probable the trip becomes. Most trips starting from a produced end in a given zone will be linked to attracted end in surrounding or nearby zones. A smaller amount of trips will be formed by links to moderately distant zones and very few will be formed by links to very distant zones. Once these models are applied, a set of trips is produced. These trips are stored in a trip table made from and to each zone in the region. Other methods can be also used to define the trips such as the maximum entropy [106].

After setting the trips, the next step would be the choice of a travel mode. The “mode choice” step consists of splitting the overall trip matrix into separate matrices each for a mode of transportation. The modes of transportation depend on the facilities available and/or we wish to visualize in the area of interest. Examples of mode of transportation are: bus, individual car, carpooling, bicycle, etc.

The last step for this modeling method is the “trip assignment”. The final step in the forecasting of travel behavior is to determine the routes travelers choose to reach their destinations. To be more accurate in the modeling, while building the paths, the real capacities of the road segments must be taken into consideration. This will assure not to overload roads with vehicles or any other mode of transportation type. This is needed in order to keep a certain degree of realism. As a result the remaining congested traffic is fed back to the trip distribution step in order for this load to be distributed.

2.4 Synthesis

In this chapter, we presented several mobility models. Some are based on stochastic process and have been classified in 6 classes. The basic mobility models are very simple but not realistic for simulation of real human world. The individual motion mobility models offer a very large variety of case for simulation. However, they do not allow the user to simulate a kind of common behavior between individuals. For that needs, the group motion mobility models are introduced. They allow defining common behavior and interaction between individuals inside groups. However, to have a mobility models applicable, they must be geographical and social context aware. The previously mentioned models lack such awareness. Other models allow the simulation to use predefined paths or targets defining these kinds of context. This is important to get realistic scenarios. The path defines the zones where motion is possible or not. The target defines objectives for mobility. Finally in order to have a large scale of simulation, it is necessary to define more complex models which are in fact a combination of several models, each of one having some particularity in simulation. The Scalable Mobility Model is a good example of one of these models.

Most of the time, the models we presented constitute adapted solution for a particular problem. Such models impose a lot of restrictions in order to simulate a particular environment. This makes them not reusable for other scenarios. To avoid these restrictions these models may be used in completion in order to eliminate all their drawbacks. By doing so, complex hybrid models are created which are corresponding to the association of mechanisms which were not dedicated to each other. However, these new models could be very heavy to implement. Moreover, it is sometimes impossible to identify the interactions between each component and which component to use in each scenario. Adding to this complexity, some models are incompatible for hybridization, some use a displacement grid while others use a vectorial displacement approach. For example the sharp turn problem is solved by the Smooth Mobility Model whereas the principle of dividing a motion step into several steps for path correction can be valid for vectorial displacement approaches but not for grid based displacement models.

Other mobility modeling alternatives were also presented in this chapter, these models are more complex and they were used for the simulation of motions in order to evaluate performances of infrastructure. We have restricted our analysis to three categories: the Multi-Agent System, the Constructal Theory and the Four-Step Model. These techniques require the definition of a lot of parameters to be used and seem to be complex to define in practice. Also some works require specific data which are not always available such as Origin/Destination matrix for the FSM. In our case, we do not want to use Origin/Destination matrix as input of the model but as an output, a result of the model and this is a different way to view the modeling problem. In fact one of our purposes is to conceive a model to generate traffic data and OD matrix which is based on the simulation of the temporal and spatial attraction of the territory, we want to start from a description of the population and a description of the territory to generate the traffic.

At the end of this chapter, and after detailing all the existing mobility models, it was obvious that for several applications, motion simulation should take into consideration social and topographical context. This presents a challenge to the existing models, and can only be achieved by complex hybrid models. Therefore, it is necessary to conceive a generic mobility model, in terms of its adaptability instead of complexity, and that can integrate several features from different models.

The aim of this thesis report is to conceive such a mobility model that we have called the Mask Based Mobility Model. This model detailed in the following chapters integrates basic features mainly from two different models: the Normal Walk and the Smooth Mobility model. The model does not use Origin/Destination matrix for the generation of mobility but a rich description of the territory based on attraction of social and economical activities along days. MBMM gives the individuals a grid based displacement in all directions, taking into account the correction of the sharp turns, and the environmental and neighborhood characteristics. The main objectives of this model is to be generic, applicable for inside and outside motion simulation, for individual and group features, and with different characteristics on transportation mode.

To define this model, the terrain that will be used for simulation will have to be introduced. Chapter 3 presents all the information and work done for terrain characterization, which will provide the information needed for modeling. The model itself, and the mathematics behind it are presented in chapter 4.

Chapter 3. Terrain Characterization

In this chapter the terrain characterization procedure is presented. This part is of major importance in the thesis. It will provide all the elements needed to build a good simulation environment for the mobility model presented in the next chapter which is the core of this research work. First, we give a theoretical explanation of the statistical methods applied in the analysis. Then the data analysis is presented. This analysis is divided in two parts: the first part deals with radio communication data and the second deals with bus data; both have been delivered by companies in charge of these services in the geographical area of interest. A view of the data we use for our analysis is showed. Then the statistical analysis results for each of these data types is presented and explained.

Content

3.1 Introduction

The terrain characterized in this chapter is the urban and suburban area of Belfort, a city in the East of France. This terrain is constituted of different structures: residential zones, city center, markets and malls, schools, universities, etc. The terrain can be divided into several groups by visual identification of these structures. This can be done for very small cities, but this task becomes harder as the size of the corresponding city gets bigger. Therefore, scientific methods and algorithms must be used for this issue.

The aim of this chapter is to describe the procedures and steps taken in the terrain characterization process, in order to build a simulation environment that reproduces the major characteristic of the terrain, needed for a mobility model simulation. Such a process starts by analyzing and detecting flows and presences of individuals and population. Since every terrain part has a characteristic that changes or evolves with time, the second step will be to identify different behaviors of geographical zones during weekdays and holidays. The result of such procedure gives a generic terrain model capable of reproducing real life events and scenarios for mobility simulations. The terrain characterization is a preamble work to the definition of a mobility model.

In addition to static data describing topography and population, two sets of data representing mobility were used and combined: data provided by a mobile network operator and data provided by a bus company. The mobile operator provides data about the presence of cellular phones users in the GSM/UMTS network cells for given periods during the day, and the transportation company provides data about the presence of individuals on bus stops for given periods during the day. These two types of data were used in *Territoire Mobile* which was a project from the cluster *Véhicule du Futur* funded by the territorial collectivities and the company SMTC (Syndicat Mixte de Transport en Commun) in charge of bus management in the Territoire de Belfort. The work presented in this chapter is a part of the work done by UTBM for this project.

These dynamic data represent a huge volume of information. So we have a multidimensional vector in space and in time with several variables to consider knowing that the value of these variables change continuously over time. To minimize the loss of information the collected data will be sampled at an appropriate rate.

After collecting the data, two multidimensional analysis and cluster analysis techniques were applied on the data sets to get the required terrain characterization. Those techniques were the Principal Component Analysis (PCA) and the k-means Clustering methods. They are described in details in this chapter.

Then, an in depth study of the collected data was done. This is shown in two major parts in this chapter. The first part studies and analyses the data collected from the mobile network operator. These data are covering the studied area representing the mobile communications on a radio electric network. The second part deals with the bus user's mobility within the area of interest. It is of major importance to state at this stage that the sets of the mobile operator data and the bus transportation data were delivered by the appropriate respective companies for the same dates and at the same sampling rate. The results of the analysis and the classification are presented in tables and graphs and they are visualized on the map of the studied area.

The identification of flows and population presence resulting from the analysis and classification over time for the chosen geographical area allows the creation of a simulation environment for the mobility model which is developed in chapter 4.

3.2 Overview of the Simulation Environment

This part briefly describes the steps taken to build the mobility simulation environment to understand the following parts. Everything will be given in more details progressively. The simulation environment is composed of several layers of information. The three major layers are the GIS information layer, the satellite view of the concerned area and the layer in which the analyzed information presented in this chapter will be stored. This last layer is called the Grid Layer and is the more complex one.

The aim of this simulation environment is to concentrate all characteristics of the real world into a grid on which the mobility model will be simulated in order to reproduce everyday real life motion patterns. GIS and satellite images are giving static information and needs to be enriched with additional information that will assign dynamic weight to each part and/or element of the geographical area of interest to drive the mobility simulation. This information is provided in the Grid layer. This Grid layer is characterized by the following elements:

— The grid: the grid is composed of equal square grid cells (25mx25m by default). Each of these grid cells contains a value called the attraction weight. The attraction weight is the value that makes each grid cell more attractive or less attractive relatively to other grid cells to drive the motion. The size of the square grid cells depends on the mobility detail level we wish to visualize in the mobility simulation. The smaller the grid cells are the more detailed mobility pattern we have. An example of grid along with the grid cells and their relative weights is given in Figure 3.1.

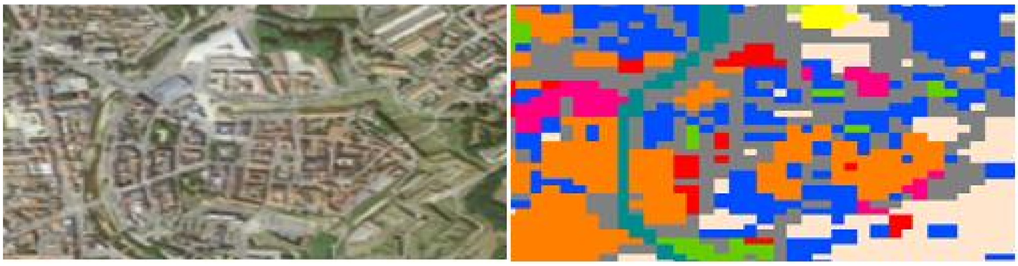


Figure 3.1. Sattelite aerial view of a geographical area and its proper grid and grid cell weights.

— Grid cell weight: it is well-known that individuals have different tastes and preferences. This is why grid cells are characterized by attraction weight. This weight is expressed by affecting a numerical value to each of the grid cells. Each grid cell level of attraction is relative to the geographical, topographical and economical structures and entities covered by this grid cell. The decision of the individual to move to one grid cell is related to the attraction weight that this grid cell has. An example of the grid cell weights representation is given in Figure 3.1 where the different attraction values are represented with different colors. The motion from one cell to another and the displacement policy is then defined by the mobility model presented in Chapter 4.

— Weight definition: the weights are issued from the analyzed data and extract all information for every sampling period. In fact one grid cell has one vector of weights for every time period which is considered along one day or along several days too. This procedure will be detailed later in this chapter.

— Time periods: the weight of each grid cell is not constant and varies during the day as each structure or entity of the map do not maintain the same level of attraction or interest all over the day. The simulation time is divided into periods such as sunrise, morning, noon, afternoon, evening and night, but it can be changed. Defining periods is important to change the grid cells over time. Further explanation and examples of this period definition and the day cutting are given in this chapter.

To situate this work among others, there are few studies taking into consideration data from GIS and/or mobile phone networks in the literature. This is quite a new concept because mobile phone information is available for few years and is very sensitive. One of these studies is the study of taxis mobility in the city of Shenzhen, in China [118]. This study takes the information from GPS and aims at describing the mobility patterns. Another study is [119], made in the city of Rome with the cooperation of Telecom Italia and the MIT where the collected information from the network operator give information about the presence of individuals on several parts of the city as well as the flow of individuals in motion from one zone to another with detected flow directions. Another study is about the mobility based on mobile telephone in Switzerland [120]. One additional study is [121], which deals with Geographical mapping of cell phone usage in the city of Milan, Italy. For every study, the time period of observation was quite short but gives elements on mobility for a very large number of people; this is the main interest in comparison with the inquests usually done on a small set of samples of the population. The information we got from one French mobile network operator was of this category but for several weeks, then with a better sensitivity to the mobility description and that was a good input for our model.

Now to use this information we need to apply algorithms to filter, analyze and synthesize the huge volume of data we had. The following parts are describing the choice and application of the methods.

3.3 Overview of the Methods

With the advances in technology, engineers are trying to study more and more complex phenomena which require the gathering of recordings of many different variables. Since the objective of our work is to study the relationship between those variables we need to resort to multivariate analysis.

Multivariate analysis includes different methodologies. The choice of any methodology is linked to the goal of the study, the nature of the used variables, and the dimension of the panel. Some of these methodologies are specific for small panels, while others are used with moderate to large panels. Among the former methodologies we have the ANOVA and MANOVA [32] [34] [39] [42]. These methods are used in the case of factor models where the factors are observable. They are related directly to regression analysis which aims to explain the total variability in a small panel of variables using their relations with observable factors. On the other hand, when dealing with moderate to large panels and when factors are not observable and when no information can be extracted from visual inspection of data plots (due to the multidimensionality of the data) sophisticated methods for extracting useful information are appealing.

From these methods we may mention the Principal Component Analysis (PCA), Factor Analysis and classification methods. These methods aim to reduce the dimensionality of the data and separate the fundamental from the noise (reduce the noise to signal ratio) and they use fewer variables to reformulate the data [29]. It should be noted that, however, classification methods do not aim to reduce the number of variables, but it aims to discriminate classes by classifying N variables into a given number of classes k ($k \ll N$) we may say that it reduces the dimensionality of the data. In our case we do not know in advance the nature of the classes to use, so we will talk about clustering instead of classification. In clustering the algorithm has also to discover the nature of the clusters.

In this study we analyze the presence and the flow of the population on an urban and extra urban area using socio-economical and geographical static data, and dynamic data from the mobility of cellular phones users on the radio-electric network and from the bus lines. The data matrixes, where one matrix is for one period of information, were provided by the network operator and the bus transportation company. The elements populating the operator data matrixes are the number of cellular phone calls (incoming and outgoing) at the base transceiver station i in the time slot t. Since $t \in \{1,...,71\}$, that is data from 6 am to 12 pm on a ¼ hour basis, and $i \in \{1,...,S\}$ stations, we may say that we are dealing with a large panel thus necessitating the use of dimensionality reduction methods in order to extract the information from our panel; in particular we use PCA and the clustering methodology k-means for this purpose.

3.3.1 Choice of the Methods

High dimensional spaces have singular mathematical properties that affect the behavior of methods used to manipulate data in these spaces. This problem is known as the "dimensionality curse". It refers to the difficulties that are encountered when managing and processing data from high dimensional spaces. This expression, "dimensionality curse", is found in the field of similarity search (Nearest Neighbor Search NNS).

Among the authors who have studied this problem, Beyer et al. presented in [122] one of the most interesting results. They show that under certain conditions, "a broad set of conditions", on the distribution of data, with the increase in dimensionality, individuals tend to become equidistant. This equidistance generates a strong volatility in the clustering results.

Several methods for dimensionality reduction have been developed in the fields of statistics and data analysis. These methods transform the data into the new space of reduced dimension while keeping the maximum amount of information carried by the data in their original space. To ensure such result, the dimensionality reduction methods exploit the correlation and/or dependence between the original dimensions.

Dimensionality reduction methods are presented in several works as a (makeshift) solution to problems caused by high dimensional spaces [123]. These studies suggest that, before it can be used for a study, the data must undergo a first stage of processing in order to reduce its initial dimension and to avoid the curse of dimensionality. Once it is done, the data, now in a space of fewer dimensions, is then partitioned by a traditional technique (K-means).

Principal component analysis (PCA) is probably the best known and most used technique for dimensionality reduction. It has its origins in the work of Hotelling, Karhunen and Loeve [124]. PCA is based on the study of the data's covariance matrix.

PCA consist on finding a new representation space with orthogonal axes that would insure that the dispersion of the data when projected on these axes is a maximum. These axes are called principal axes. The amount of information carried by each axis is relative to the variance of the data: the larger the variance of data along an axis, the more the information carried by it. In fact, the PCA consists of performing a translation followed by a rotation of the coordinate space. The dimensionality reduction is done by eliminating the axes that bear little information.

We can formulate the problem using the following mathematical notation: we have to find a lower dimensional representation, $s=(s_1,\dots,s_k)^T$, for a given j-dimensional random variable $y=(y_1,\dots,y_j)^T$ where $k \leq j$. The new dimensional representation, s, should, according to some criterion, capture the content of the original data y [134]; this is the main purpose of the reformulation. We call the *hidden components* the components of s. Several names are used for the description of the j multivariate vectors such as *variable* in statistics, but also *feature* or *attribute* in computer science domain.

There are numerous books and articles [110][111][112][113][114][115] in the statistical literature on techniques for analyzing multivariate datasets. Earlier survey papers, [116] review several methods, including principal components analysis, projection pursuit, k-means clustering, etc. HyvÄarinen's surveys results in independent component analysis [117].

Principal component analysis and factor analysis, the two most widely used linear dimension reduction methods, are based on second-order statistics. For standardized normal variables (with mean zero and variance 1), the correlation matrix contains all the information about the data. Second-order methods are relatively simple to code, as they require classical matrix manipulations. However, many datasets of interest may not be realizations from Gaussian distribution. For those cases, higher-order dimension reduction methods, using information not contained in the covariance matrix, are more appropriate. These linear higher-order methods include the projection pursuit, independent component analysis and random projections.

Principal component analysis is considered to be an efficient unsupervised method to reduce dimensionality in multivariate statistics [30] [112]. It is used for many problems and it gives good results whatever the "sense" accorded to the data. Due to Eckart-Young theorem, the usage of singular value decomposition (SVD) by PCA "gives the best low rank approximation to original data in L2 norm" [133]. It runs like a kind of noise reduction process which brings to the new data more expressive power than the original data set.

The objective of PCA is to determine linear combinations of the original variables that maximize the explained variation, subject to constraints of having unit length, and being orthogonal to previously identified principal components.

For the classification problem, the k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It attempts to find the centers of natural clusters in the data. Given a set of observations $(x_1, x_2\dots x_n)$, where each observation is a d-dimensional real vector, the k-means clustering aims to partition the n observations into k sets, $(k < n)$ $S = \{S_1, S_2\dots S_k\}$, so as to minimize the within-cluster sum of squares (WCSS):

$$\arg\min_{s} = \sum_{i=1}^{k} \sum_{x_j \in S_i} \left\| x_j - \mu_i \right\|^2 \tag{1}$$

Where μi is the mean of points in S_i.

The PCA aims to decompose the data matrix as:

$$X = S\Lambda \tag{2}$$

Where S contains the principal components and Λ are the loadings. The factor analysis decomposes the data as:

$$X = S\Lambda + U \tag{3}$$

Where S contains the common factors, Λ contains the factor loadings and U is the matrix of specific factors called the noises.

It has been shown [108] [35] that the relaxed solution of k-means clustering, specified by the cluster indicators, is given by the PCA principal components, and the PCA subspace spanned by the principal directions is identical to the cluster centroïd subspace specified by the between-class scatter matrix.

In general, if we have a high data dimension we can use the PCA firstly to reduce the dimension before applying the k-means clustering. In the PCA methods, the principal components and the loadings are computed in two steps. First, the components of S are computed as the eigenvectors of X and using them to compute the loadings. In the factor analysis, we can use the PCA to compute the factors and the factor loadings. But we can also use another optimization approach, for example we can assume the normality assumption and use the maximum likelihood estimator.

The main critics addressed to PCA and factor methods are their linearity and the fact that they are second order method, i.e. they use only the information in the second order moments (variance-covariance matrix of the data) to compute the components.

There are other methods that come over these shortcomings. We mentioned the projection pursuit and independent component analysis. The first one uses higher-order projection index which is based on the negative Shannon entropy [109]. The independent component analysis uses some cost function to extract the components. This cost function must nests the dependence property in some sense.

The process can be written as follows:

$$X = f(S, \Lambda, U) \tag{4}$$

Where f is any function relating X to the factors (S), factor loadings (Λ) and noise (U).

While, these methods capture higher order dependences and nonlinearities, it is not clear how to choose f to capture these nonlinearities. Also, their implementation is difficult because there are no guidelines to choose a good cost function; the choice of the cost functions is somewhat arbitrary. Also, the methods that use higher order moments suffer from the presence of outliers.

$$\mathrm{S}, \Lambda = \arg\min \mathrm{g}(\mathrm{U}) = \arg\min \mathrm{g}[\mathrm{X} - \mathrm{f}(\mathrm{S}, \Lambda, \mathrm{U})] \quad (5)$$

Where $g(U)$ is the cost function.

It is not clear how to choose f to capture nonlinearities. On the other hand, there are no clear guidelines to choose the form of the function $g(u)$ since the estimate of S and Λ depend on such choice. Those shortcomings (problems) render the interpretation of the components of S and Λ harder and cumbersome. For all these reasons we have used the k-means clustering and the PCA for the first step of data analysis. We know describe both methods.

3.3.2 The Principal Component Analysis method

The principal component analysis is a statistical technique for compressing or reducing the dimension of data but retaining most of the original variability in the data [30]. The principle is to take a set of correlated variables and produces a set of principal components, which is smaller in number than the original set of variables, which are uncorrelated i.e. they are orthogonal. This is done by doing a covariance analysis between factors [32]. The PCA is also known as discrete Karhunen-Loève transform (KLT), the Hotelling transform or proper orthogonal decomposition (POD).

Let the data set presented by the $M \times N$ matrix X which has N data vectors ($X_1 \dots X_N$) with M lines of data for each vector (vector dimension).

The first step is to produce a data set with 0 mean. In order to do so the mean of each dimension $m \in \{1,\dots,M\}$ has to be computed. The resulting $M \times 1$ vector U would be:

$$U[m] = \frac{1}{N} \sum_{n=1}^{N} X[m, n] \quad (6)$$

The mean is then subtracted from the main data set, in order to eliminate the scale effect, giving the matrix B, an $M*N$ matrix of centered data:

$$B = X - Uh \quad (7)$$

where h is a $1*N$ identity vector.

The new data set is used to find the covariance matrix C:

$$C = \frac{1}{N}\sum_{n=1}^{N} B_n B_n^T \quad (8)$$

where B^T is the transpose of B.

From the covariance matrix the matrix V of eigenvectors will be derived. The diagonal matrix of eigenvalues D of C results from the formula:

$$D = V^{-1}CV \quad (9)$$

D is an $M \times M$ diagonal matrix where $D[i, j] = \lambda_m$, the m^{th} eigenvalue of the covariance matrix C, for $i = j = m$ and 0 when $i \neq j$.

To choose the number of principal components for a study, the cumulative energy vector g is computed after rearranging the matrices D and V in decreasing order of eigenvalues.

$$g[m] = \sum_{q=1}^{m} D[p, q] \quad (10)$$

For $p=q$ and $m \in \{1,\ldots,M\}$. A subset of V is created choosing the first L columns, forming an $M \times L$ matrix called W. The choice of L is made on the basis of choosing as small of value as possible while achieving a reasonably high value of g on a percentage basis.

The first extracted component in a PCA accounts for a maximal amount of the total variance in the observed variables. The second component will be uncorrelated with the first component and accounts for a maximal amount of the variance in the data set that was not accounted for by the first component. And so on for the rest of the extracted principal components.

PCA is often used with other types of analysis, mainly as a dimension reduction technique. PCA may be used with classification methods [30]. In our case we are interested in the cluster analysis, to measure one type of dissimilarities but mainly as a visual tool in order to give a two dimensional image of the data with the projection of data on the axis of the first two principal components.

3.3.3 The *k*-means method

Classification analysis objective is to divide a data set into classes in a sensible way [30]. It creates subsets of homogeneous individuals, i.e. it groups subjects so that individuals of the same group are as similar as possible regarding to certain characteristic and the different groups are as dissimilar as possible [31].

One major step in any classification or clustering analysis is the choice of a dissimilarity measure that is used to quantify the distance between two subjects; the smaller the distance between two individuals the more similar they are.

When dealing with continuous data, the Euclidian distance is usually used [32]. The distance between two subjects i and j is then:

$$d(S_i, S_j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \cdots + (X_{ip} - X_{jp})^2} \tag{11}$$

In this study we use one of the most used and efficient [33] clustering methods, the k-means method. The clustering method instead of classification method was necessary as we do not know in advance the properties of the classes we have to consider for the study.

Like other clustering techniques, k-means aims to minimize the within variance in each cluster [34] [44]. k-means is a non-hierarchical clustering method which in contrast with hierarchical methods, the number of clusters has to be defined at the beginning. Then we will have to try different numbers of clusters to find the right one.

The first step in this methodology is to select k seeds that will be preliminary centroïds for the groups. The second step is to assign each observation to the nearest group, the group for which the Euclidian distance between the subject and its center is the smallest. Once this step is over, new centroïds for the clusters are calculated. This is done by minimizing the sum of squared errors:

$$J_K = \sum_{k=1}^{K} \sum_{i \in C_k} (x_i - m_k)^2 \tag{12}$$

Where $(x_i, \ldots, x_n)$ is the data matrix and $m_k = \sum_{i \in C_k} \frac{x_i}{n_k}$ the centroïd of cluster C_k and n_k is the number of points in C_k [35].

Once new centroïds are calculated the second step is repeated, i.e. assigning observations to groups, then new centroïds are calculated all over again. The process will be repeated until the centroïds are stable, that is changes become not significant or even 0, or until a predetermined stopping criterion is reached.

In this study we use a clustering quality criterion to assess the groups. The criterion is called the *silhouette* and notice S_i, the silhouette for the individual X_i. The *silhouette* is a number between -1 and +1 that measures the relevance of the membership of one subject to a cluster: the value close to +1 means an excellent relevance. The criterion is measured as follows [37]:

Let *a(i)* be the mean squared distance between subject i and the other subjects in the cluster. Let *b(i,k)* be the mean squared distance between subject i and the subjects of cluster k, then:

$$S_i = \frac{\min[b(i,1), b(i,2), \dots, b(i,n), 2] - a(i)}{\max\,[a(i), \min[b(i,1), b(i,2), \dots, b(i,n), 2]]} \tag{13}$$

The silhouette presented in equation (13) helps to define the number of clusters in each level of the clustering process. It is a measure of how close each subject in one cluster is to subjects in the neighboring clusters. This measure ranges from +1, indicating points that they are very distant from neighboring clusters, through 0, indicating points that are not distinctly in one cluster or another, to -1, indicating points that are probably assigned to the wrong cluster.

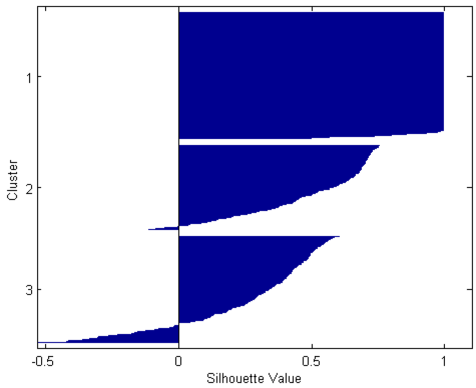
In other words, a positive value of silhouette expresses a good clustering of a subject i, while a negative value expresses a bad one. Interpretation of silhouette extreme values is then:

— s(i)=+1, the object is far away from the other clusters;

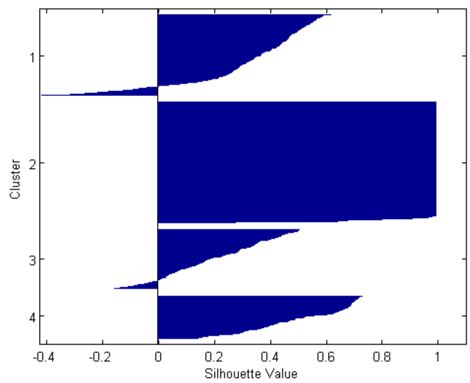
— s(i)=0, the object is not distinctly in one cluster or another;

— s(i)=-1, the object is badly ranked.

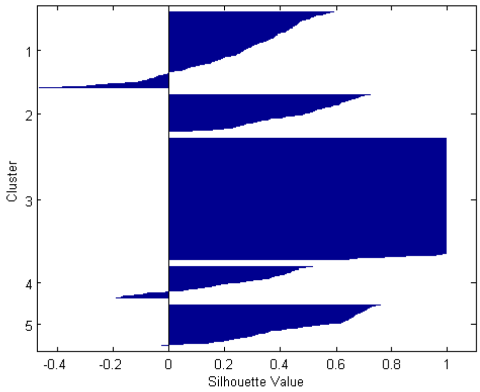
The study of the individual silhouette for various clustering with various numbers of clusters is a good mean of numerically identifying k, the right number of clusters.



(a) mean (silhouette) =0.6091



(b) mean (silhouette) = 0.5839



(c) mean (silhouette) = 0.5856

Figure 3.2. Example of Silhouette

Figure.3.2 shows an example of graphs obtained from the computation of silhouettes with different k value for a given set of subjects. In the graphs presented in Figure 3.2, silhouette values are along the x-axis and the clusters are along the y-axis.

From Figure 3.2 (a), we can see that most subjects in the first cluster have a large silhouette value, greater than 0.9, indicating that the cluster is somehow separated from other neighboring clusters. However, the third cluster contains many subjects with low silhouette values, and the second cluster contains few subjects with negative values, indicating that those two clusters are not well separated.

A more quantitative way to compare the results is to look at the average silhouette values for each one of the obtained results. Figure 3.2 (a) gives a better clustering; the silhouette indicates that three clusters are better separated than the fourth or fifth ones with greatest mean of silhouette value (0.6091).

After having introduced the algorithmic methods, we are now going to present the dynamic data we have used which is coming from radio communication network or from bus network.

3.4 Radio Data

3.4.1 Radio Data layers

The data provided by the mobile network operator is given as an individual/variable matrix, the individuals being the radio transmitters or base stations covering the studied area. Each base station is characterized by its identifier number and its load being fifteen minutes time periods computed after the division of the day from 6:00 till 23:45. Each base station corresponds to a given geographical zone called a cell where it is in charge of communications; that is any telephony call in the zone will be associated to the radio transmitter covering the cell. During one day, it is then possible to extract in continuous the number of calls for any cell, this information is called the *cell load*. This data is finally providing the dynamic number of persons which are located in a given area during a given time. The location is more or less precise since the size of the cell ranges from 400 meters of radius in downtown to 10 km in landscape.

For our study, we dispose, for each base station, of the following information for each cell and for one full week; it is corresponding to six layers of data:

— **Outgoing calls:** the number of persons present in a given radio cell and using their mobile phones to emit a call during a fifteen minutes period.

— **Incoming calls:** the number of persons present in a given radio cell using their mobile phones to receive a call during a fifteen minutes period.

— **Total number of calls:** the summation of outgoing calls and incoming calls data for every fifteen minutes period.

— **Outgoing HO:** the number of mobile phone users leaving a given cell to one of its neighboring cells, while in a call during a fifteen minutes period.

— **Incoming HO:** the number of mobile phone users entering a given cell from one of its neighboring cells, while in a call during a fifteen minutes period.

— **Total HO:** summation of the incoming and outgoing HO.

Here are some samples of the information we used. The Table 1 is the load of the mobile cells for one period of time. There is one column per cell and one line per category of call.

In Table1, for each base station identifier we have, for each time period, the information about the total number of calls, the incoming calls, the outgoing calls, the total HO (handover), the incoming HO and the outgoing HO. This table is a sample of the matrices provided by the mobile network operator for our work.

The Table 2 is the list of the 17 different files we used which are data collected between October 14th and October 30th 2006. It is important to know the day of collection in order to identify the different mobility behaviors for weekdays. Some collections of data were measured during holiday's period; like that we would be able to identify the difference in mobility behavior between holidays and working days.

Table 1. Sample of individuals (time periods) versus variables (base station identifiers) data matrix

Bases station Identifier	11173	11175	11110	11140	11141	11142
Start	6:00:01:541					

total number of calls	0	0	1	1	0	5
incoming calls	0	0	0	1	0	1
outgoing calls	0	0	1	0	0	4
total HO	0	0	0	1	0	2
incoming HO	0	0	0	0	0	1
outgoing HO	0	0	0	1	0	1

Table 1 (a). Sample of individuals (time periods) versus variables (base station identifiers) outgoing calls data matrix

Bases station Identifier	11173	11175	11110	11140	11141	11142
Start	6:00:01:541					
outgoing calls	0	0	1	0	0	4
Start	6:15:02:272					
outgoing calls	0	2	1	0	0	3
Start	6:30:03:181					
outgoing calls	1	1	0	0	2	0

Table 1 (b). Sample of individuals (base station identifiers) versus variables (time periods) outgoing calls data matrix

Bases station Identifier	6:00:01:541	outgoing calls	6:15:02:272	outgoing calls	6:30:03:181	outgoing calls
11173		0		0		1
11110		0		2		1
11175		1		1		0
11140		0		0		0
11141		0		0		2
11142		4		3		0

Table 2. Listing of days and dates in the data matrixes

date	day	holidays	week end
10/14/2006	Saturday	No	Yes
10/15/2006	Sunday	No	Yes

10/16/2006	Monday	No	No
10/17/2006	Tuesday	No	No
10/18/2006	Wednesday	No	No
10/19/2006	Thursday	No	No
10/20/2006	Friday	No	No
10/21/2006	Saturday	No	Yes
10/22/2006	Sunday	No	Yes
10/23/2006	Monday	No	No
10/24/2006	Tuesday	No	No
10/25/2006	Wednesday	Yes	No
10/26/2006	Thursday	Yes	No
10/27/2006	Friday	Yes	No
10/28/2006	Saturday	Yes	Yes
10/29/2006	Sunday	Yes	Yes
10/30/2006	Monday	Yes	No

3.4.2 Overview of Analysis and Clustering Steps

In this section we do the analysis and the clustering of the data issued from the radio communication network. Firstly a correlation analysis is performed in order to reduce the number of analyzed data. Then a PCA is undergone in order to clean up and reduce the size of the data matrixes. Finally, the clustering method k-means is applied to the data, splitting the information into groups, and the group division is evaluated using the computation of *silhouette*. All figures and results in this section correspond to October 3rd, 2006 which is a Tuesday. There is no particular reason to show that day, it is only an illustration example of the analysis that we did.

Different categories of results were obtained. The first result is the PCA for different data matrixes. Figure 3.3 is an output graph of the eigenvalues representation of the PCA on the outgoing calls matrix, where the individuals are the base stations and the variables are the time slots. The x-axis represents the principle components, and the y-axis gives the percentage of the variance explained by each component. In this graph on outgoing calls the first component explains more than 70% of the variables information. We got the same kind of results for other matrixes coming from the different categories of calls.

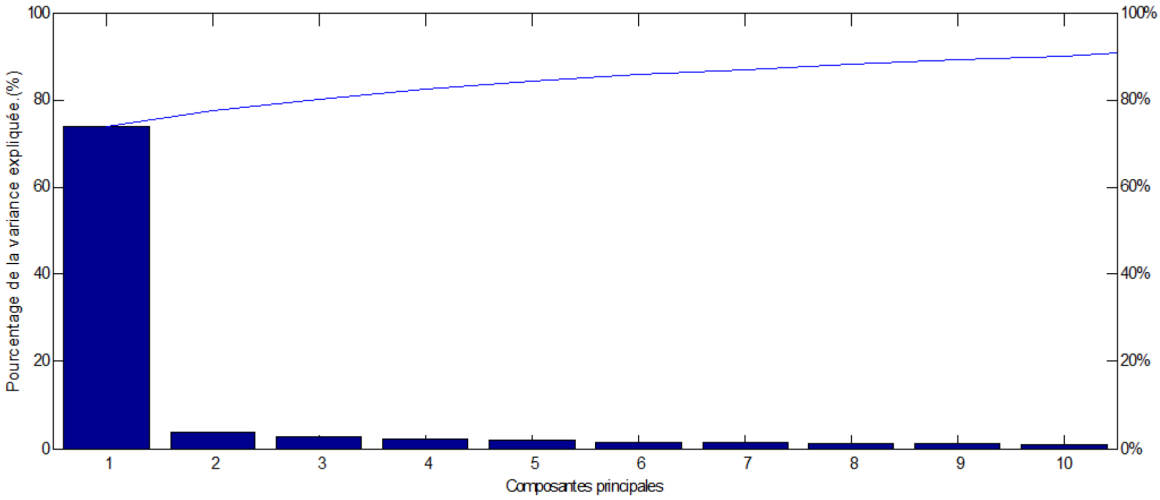


Figure 3.3. Eigen values representation of the PCA analysis of the outgoing calls matrix

A second result is the clustering output as illustrated in Figure 3.4. The clustering is done after applying *k*-means to create groups of base stations in function of their similarity in traffic load during days. The idea is to find the right number of clusters of base stations to use in order to have a good representation of the initial data. In Figure 3.4, we cluster the time periods of calls during the day. We want to find the similarity between time slots or periods of calls in terms of load. The different colors in a column indicate the different groups; it is again from 2 to 7 clusters. We only plot the global *silhouette* value on the column top. The best clustering is obtained with 3 clusters for silhouette equal to 0.71307. We see that there are two long periods of similarity, from 7:15 to 10:00 and from 17:30 to 23:30. Outside these periods the behavior of calls is more complex as the quarters of time are not adjacent.

In the Figure 3.5 each column is a possible clustering result with a given number of clusters from 2 to 7. Each of these clusters is characterized by its quality. On the top of each column there is the average quality value corresponding to the *silhouette*. Inside each column the color distinguishes the clusters and the value is the *silhouette* for the given cluster, so we know the contribution of the clusters to the global clustering. From this Figure it is also possible to see the successive refinement of data from 2 clusters to 7 clusters. We observe that three clusters of base stations give the best score with the higher *silhouette* (0.71307).

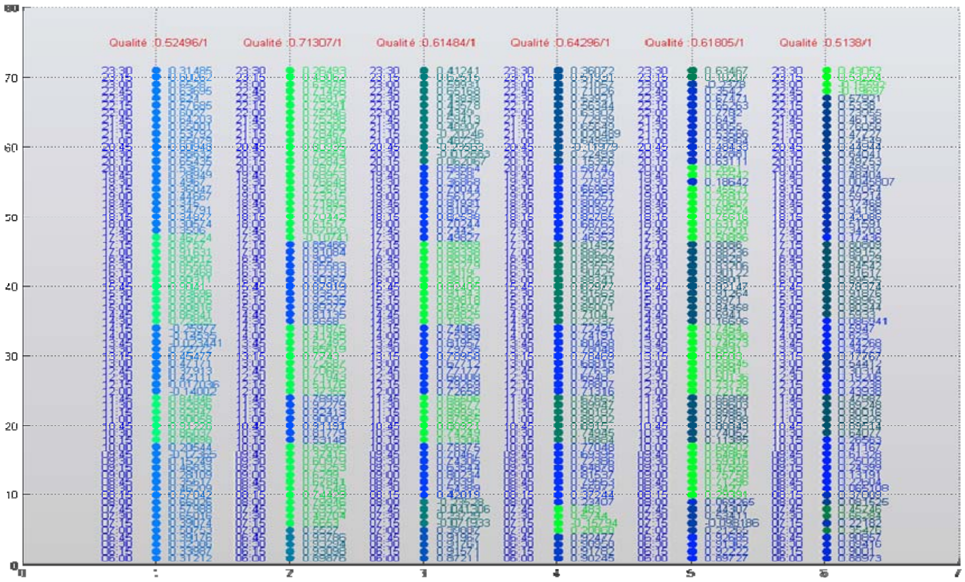


Figure 3.4. Clustering results of outgoing calls matrix/time periods (indiviuals = time periods, variables = base station identifiers) [table 1(a)]

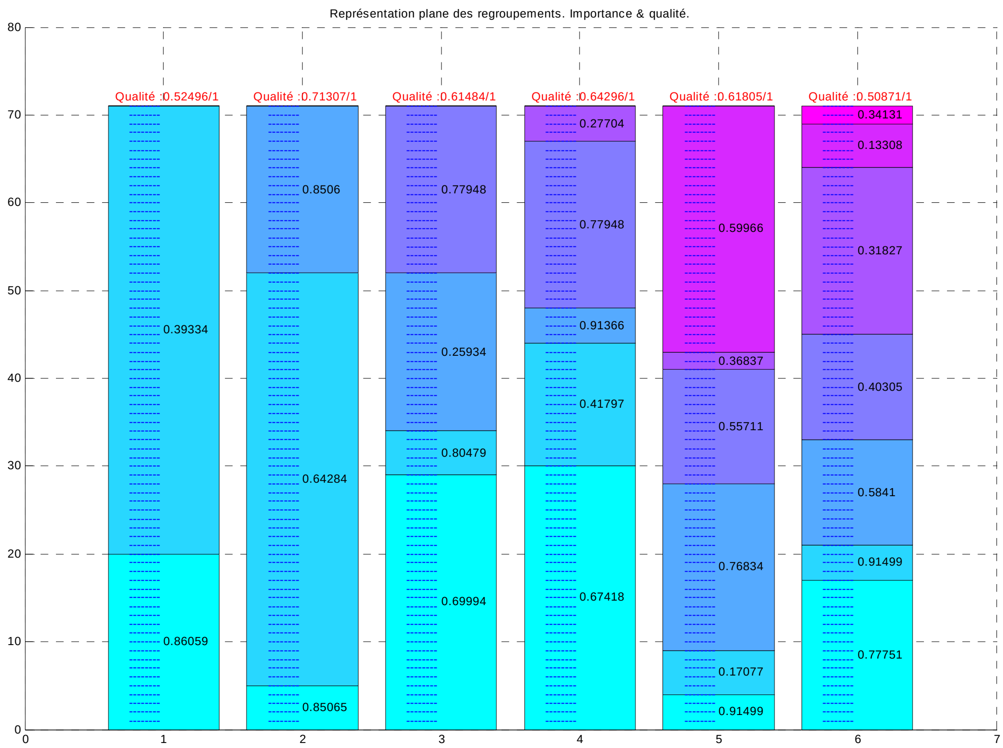
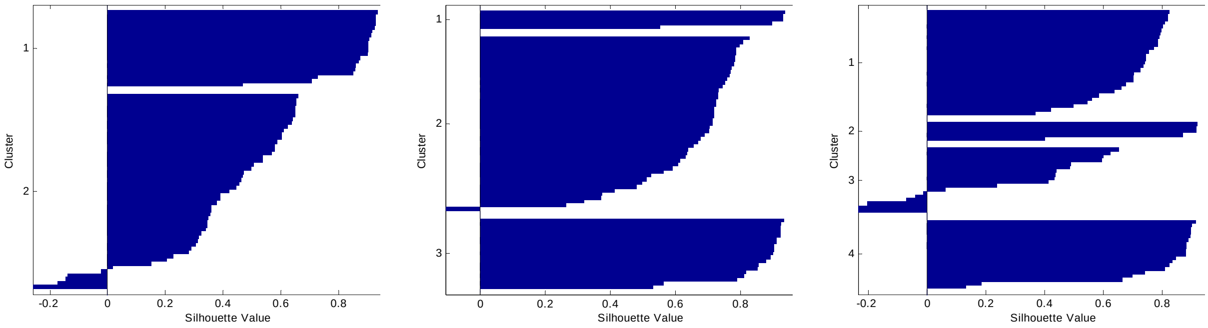


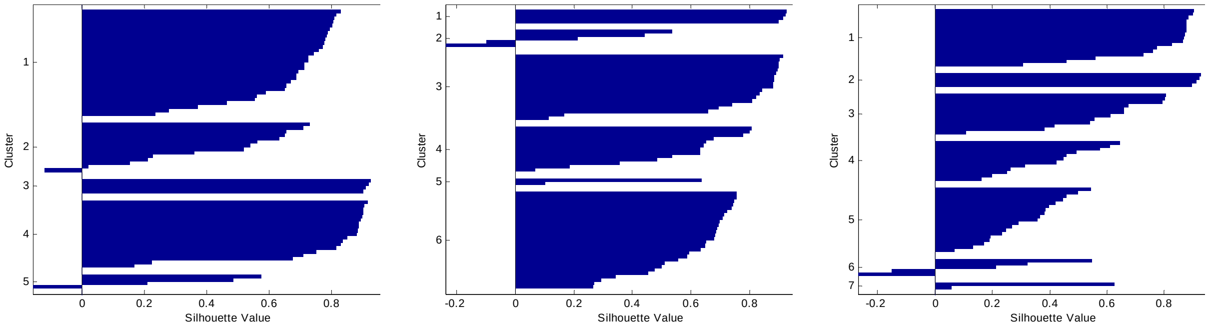
Figure 3.5. Clustering results of outgoing calls matrix/base stations(indiviuals = base station identifiers, variables= time periods) [table 1 (b)]



k=2 mean (silhouette) = 0.52496

k=3 mean (silhouette) = 0.71307

k=4 mean (silhouette) = 0.64296



k=5 mean (silhouette) = 0.64296

k=6 mean (silhouette) = 0.61805

k=7 mean (silhouette) = 0.50871

Figure 3.6. Silhouette representation for *k* = 2, 3, 4, 5, 6 and 7, for the outgoing calls matrix/base stations

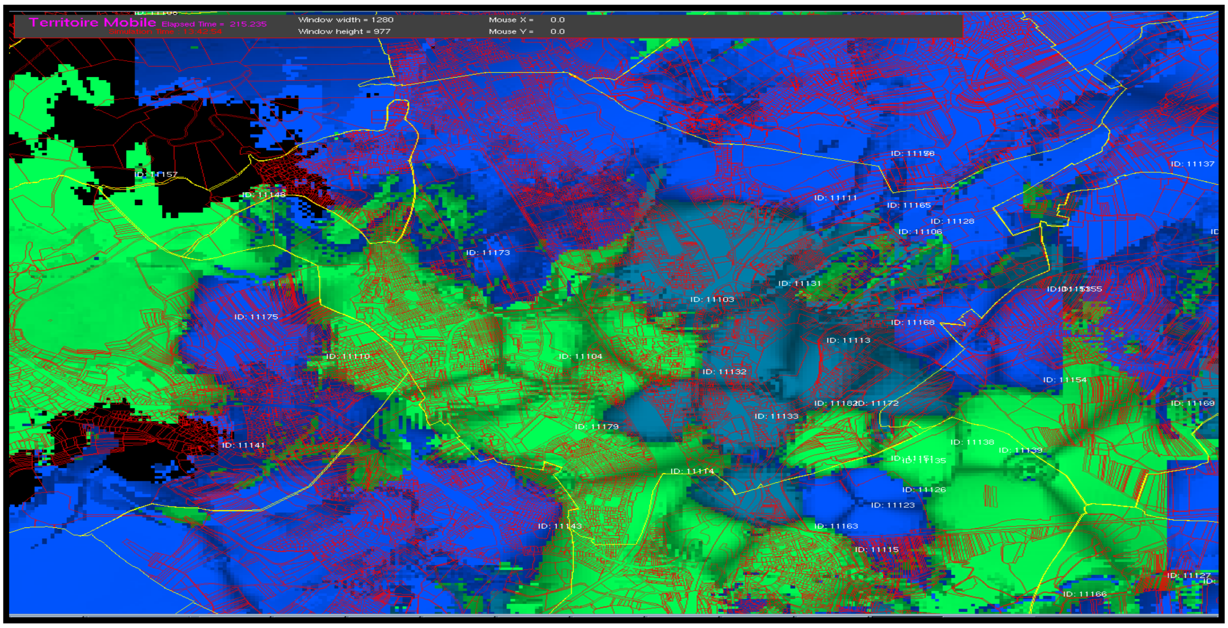


Figure 3.7. Interactive representation of the different groups on the map of the studied area

Figure 3.6 presents the silhouette for the different values of k. In Figure 3.7 a graphical representation of the clustering output is given to illustrate the result of clustering on a map. Each color in the figure refers to a cluster of cells (each identified by the base stations identifier) which have been identified as similar on their traffic load. The detected clusters use three colors on this map, grey blue, blue and green. Then it is possible to rapidly identify the zones of the territory which have similarity in their behavior.

3.4.3 Characterization of the Mobility Space

Here we want to search the possible existing correlations between different data elements from the cells. The cells are mainly characterized by their different categories of load (incoming, outgoing and handover). Also they are identified by their own location on the map. So we want to look at the similarity we may find on a territory in terms of mobility trough the cell loads.

The results are presented in three parts:

- Correlation analysis: correlation between incoming and outgoing calls and between incoming and outgoing handovers.
- PCA results: determination of the adequate data for the study of traffic intensity per cell in function of the principal components.
- Clustering results: determination of the adequate clustering parameters to locate different areas on the map based on groups of cell.

3.4.3.1 Correlation Analysis of Call and Handover Data

Figure 3.8 shows a sample of the incoming/outgoing calls correlation analysis; the x axis in the figure corresponds to the incoming calls and the y axis corresponds to the outgoing calls.

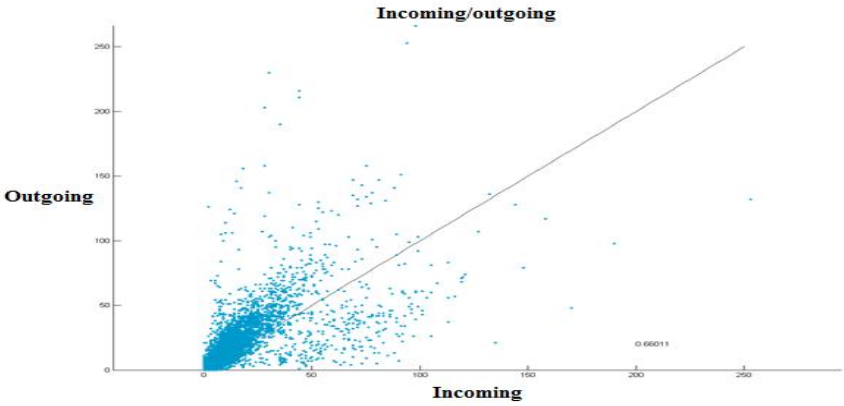


Figure 3.8. Correlation analysis incoming calls/outgoing calls

The correlation coefficients between both categories of calls are around 0.65 during week days and 0.55 on Sundays. In the analyzed data it was globally observed that the number of outgoing calls is greater than that of incoming calls in a proportion of sixty to forty percent. This observation was confirmed by the network operator traffic engineers. That means that the people are more using their phones to call instead of being called. For this reason the outgoing calls or the sum between incoming and outgoing data will be used because it will provide better results than the use of the incoming calls alone as it was planned at the beginning of the study.

Concerning handover data, the correlation coefficient between the incoming and outgoing handovers is around 0.96. We found it with the same kind of analysis plotted in Figure 3.6. This high correlation is coherent due to the fact that there is no important population flows in the area of interest during the studied days. Then the population motion is quite similar in all directions.

3.4.3.2 Analysis of PCA Results

From the previous step, the analyzed data are the calls matrixes corresponding to outgoing calls, incoming calls, and incoming plus outgoing calls. The individuals are the base stations and the variables are the three types of load along time. The different PCA gave the following results and observations plotted in Figures 3.9 to 3.11 (the observation day is a Saturday):

The first principal component represents the traffic intensity in the cell. It concentrates the traffic of the higher loaded periods during the global observation time and allows distinguishing the cells in regard with the global traffic load. The percentage of variance presented by the first principal component is higher for the outgoing calls than for the incoming calls or the outgoing + the incoming calls. Therefore analyzing the outgoing calls matrix will be sufficient to identify the traffic intensity periods in the different cells.

For the outgoing calls, 80% of the variance is given with the first three principal components. Visualization in a space with three dimensions of these data thus enables us to visualize in a precise way the behavior of the stations (individuals) and time periods (variable).

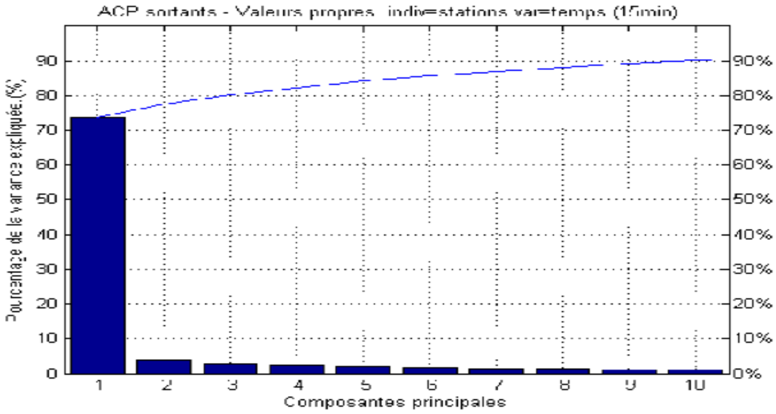


Figure 3.9. PCA results [eigen values] for the outgoing calls

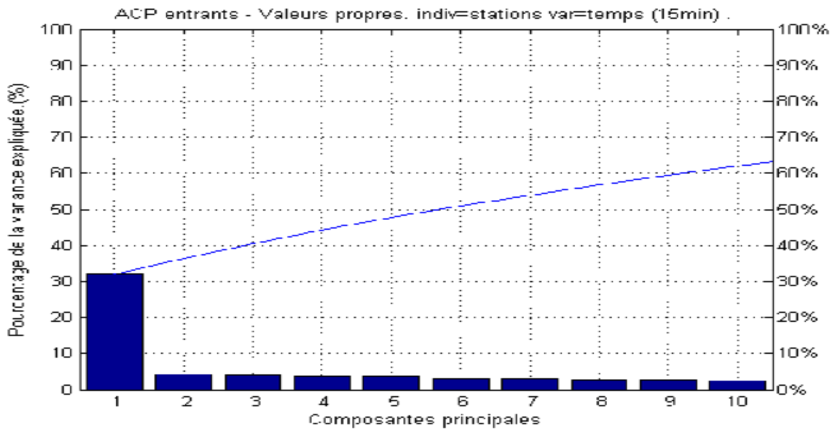


Figure 3.10. PCA results [eigen values] for the incoming calls

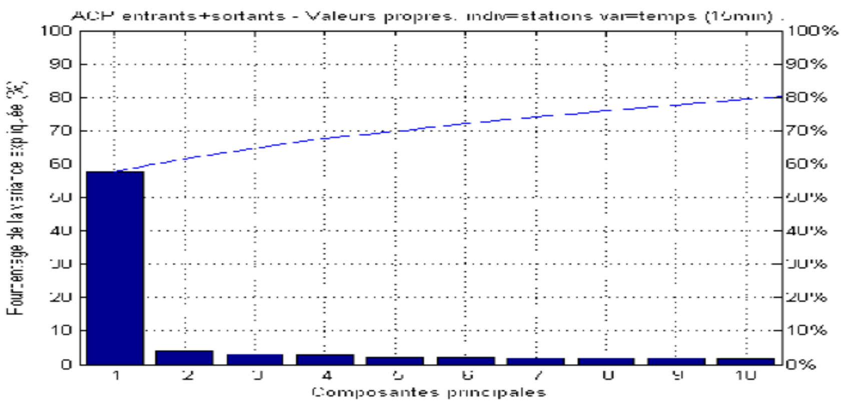


Figure 3.11. PCA results [eigen values] for the incoming+outgoing calls

3.4.3.3 Analysis of Clustering Results

Since the outgoing calls data carries more info than the other sets, clustering the PCA results for this data matrix will give a clear vision on base station traffic behavior.

We need to do several clustering with different numbers of clusters to check the nature of the traffic data. Each clustering result is then evaluated with the *silhouette* in order to make a cluster choice. The Figure 3.12 shows six different clustering for the analyzed outgoing calls data along its quality. The left side column is for two clusters then we add one more cluster from left to right up to seven clusters. Each cluster is represented by a different color. The ordinate values stand for the number of base stations in the group. The number inside each cluster is for the *silhouette* value of the group. Quality value upside each column is the global *silhouette* value for the set of clusters.

If the general quality standard of the clustering is higher for two and four clusters, two clusters are not sufficient to obtain a satisfactory visualization of the traffic, and with four clusters the quality of one of the cluster is very low (0.009) so the cluster itself is not well suited for data representation.

The clustering of the result of the PCA in 6 groups is the third result in terms of *silhouette* evaluation very close to the three clusters case. At that time we have the choice in getting more or less clustering precision; in fact the 1st cluster in 3-size case is divided into 4 subsets with 6-size case but the rest is very similar. And so using six clusters we have a deeper view of the data. The individual quality of each cluster is better with the 6-size case, i.e. the clustering representation is better.

The visualization of the different groups on the map, as shown in Figure 3.13, by coloring the cells associated with the stations, enables the detection of the zones having a particular behavior relative to the traffic intensity level. The figure is for the six clusters case.

The red cells are the ones having high traffic values relatively to the other cells. The yellow cells are the ones having less traffic. The blue and cyan lines on the figure show the *shapefiles* giving respectively the buildings and roads and the territorial limits. Thus, we can identify specifically zones with strong traffic and zones of weak traffic. For the mobility modeling it is then possible to identify the number of person that should be located in the different part of the city. The radio communications network gives a kind of measurements of real location data. It may be used as reference for modeling. We did this analysis for all files from the mobile network to get good information on space mobility (regardless whether it is a work day a holiday, a week day or a week-end day).

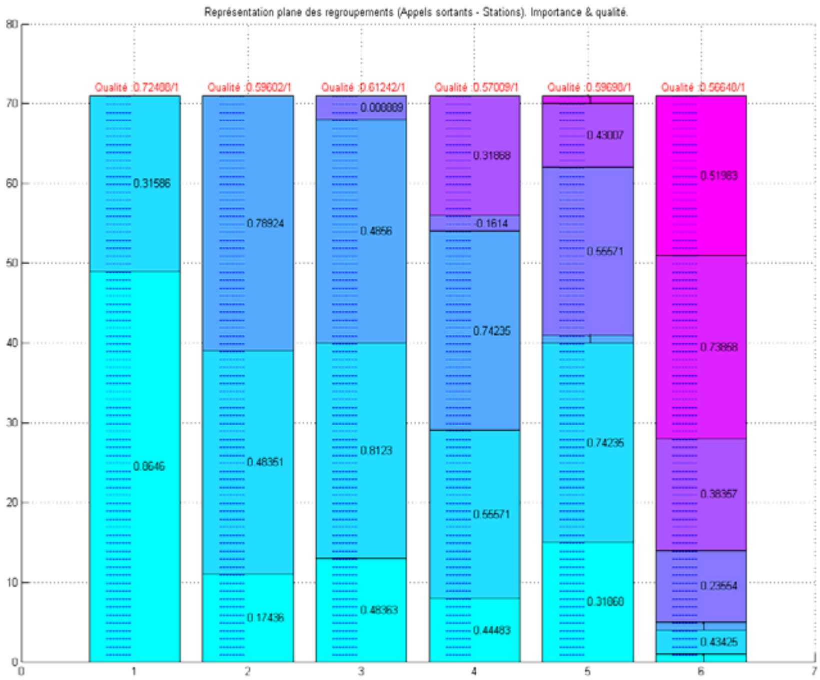


Figure 3.12. Clustering results and cluster quality for outgoing calls matrix

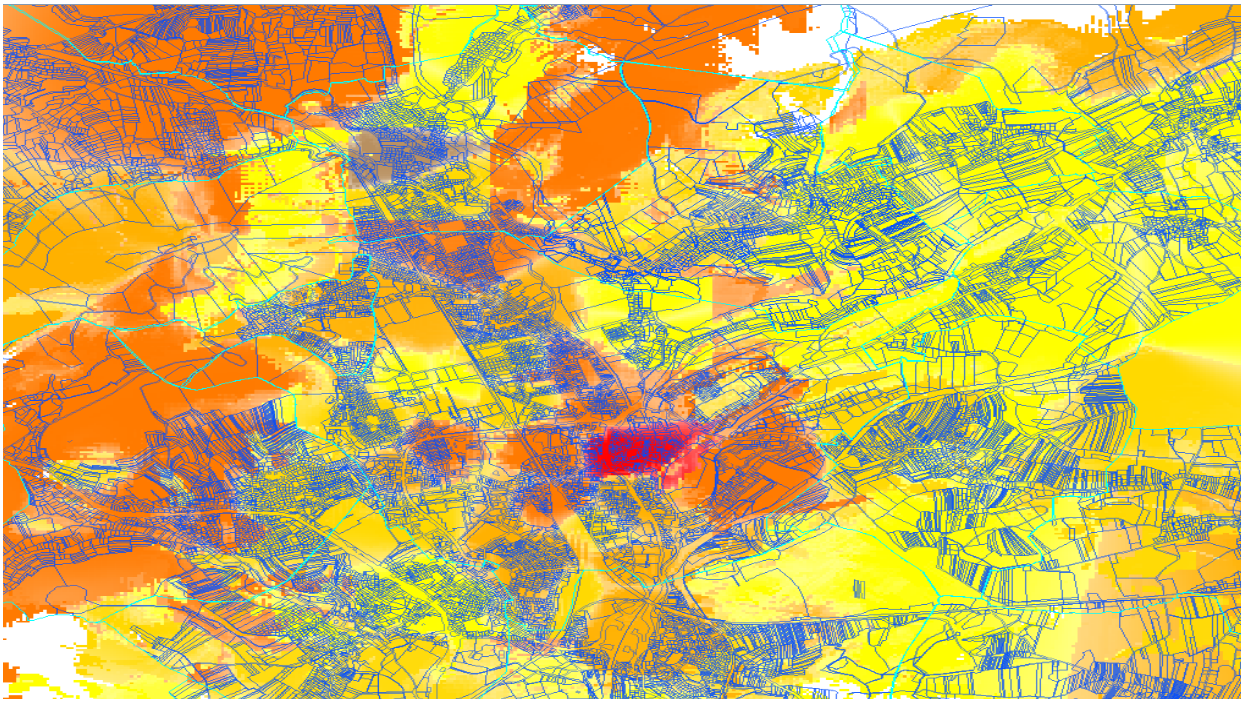


Figure 3.13. Visualization of the groups to identify the load behaviors on map

3.4.4 Characterization of the Mobility Time

In the previous part we analyzed the base stations in regard with their traffic load to identify mobility behavior inside the geographical area. This was a spatial analysis of the mobility. In this part we analyze the variables which are corresponding to the time slots. This is a temporal analysis of the mobility.

The goal of this analysis is to detect the time clusters which have a common influence on the mobility inside the cells, and to check if these clusters have a global coherence in the periods of the day once placed on the time scale. If this coherence is checked, we will be able to divide one day into several time slots having distinct behaviors of mobility.

The analysis of the time periods is done following the same steps as the analysis of the stations. The results are presented in two parts:

— PCA & Clustering results: selection of the most representative type of data on the target principle component and determination of the adequate clustering parameters, checking of the temporal coherence of the groups and determination of a standard splitting of the day.

— PCA and Clustering visualization: checking the presence of each new time group in the observed principal component and adjusting the day groups to be understandable for usage modeling.

All figures and results for this part correspond to October 14th, 2006 which is a Saturday.

3.4.4.1 PCA & Clustering Results

The PCA step is similar to that of the spatial analysis. The outgoing calls have been used for this part too in order to plan the time division because of the representativeness of these calls. We focus on the clustering of outgoing calls data to get the temporal day division. The results are specific to the day we analyzed and must not be generalized to other days as each day is different in behavior.

As shown in Figure 3.14a six clustering scenarios of time planning were carried out. For each of these scenarios the quality of the clustering is computed. The time is the reference here while the spatial distribution was the reference of the clustering in the previous part. The used colors (clusters) and values (*silhouette*) can be interpreted as before, we just have to change the result analysis going from space to time.

The four clusters scenario gives the best clustering quality among the six scenarios. In Figure 3.14b this four groups clustering result is verified on the time scale. The third column shows the four groups clustering. We clearly identify two major components that go together during the day, from 9:45 to 12:00 and from 14:00 to 17:30, and also three other periods inside the day but around the previous cluster, that is from 7:30 to 9:45, 12:00 to 14:00 and from 17:30 to 19:45. These periods correspond to difference in the number of calls that is the number of persons at that time in the coverage areas of cells.

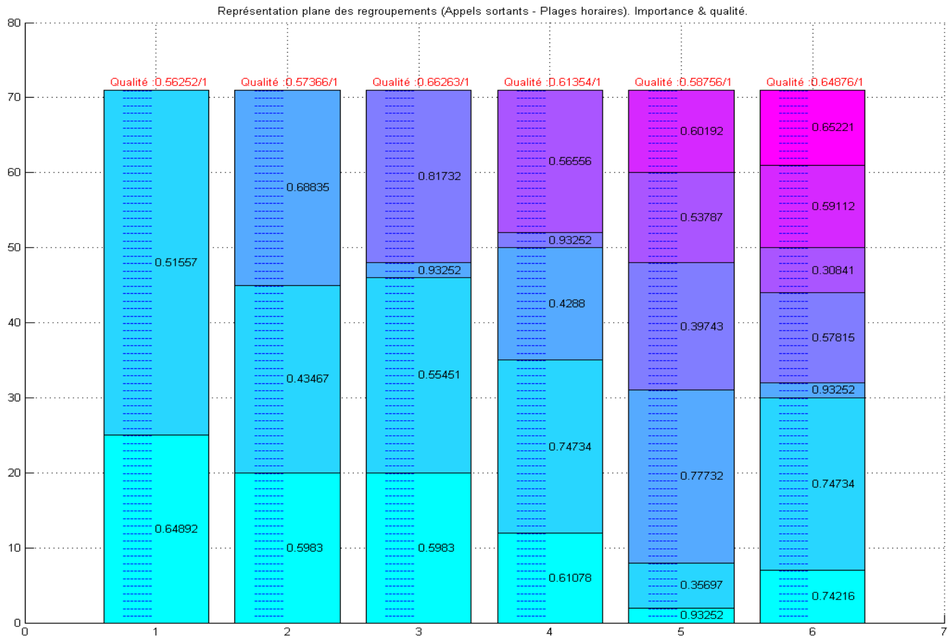


Figure 3.14a. Clustering results and cluster quality for outgoing calls matrix

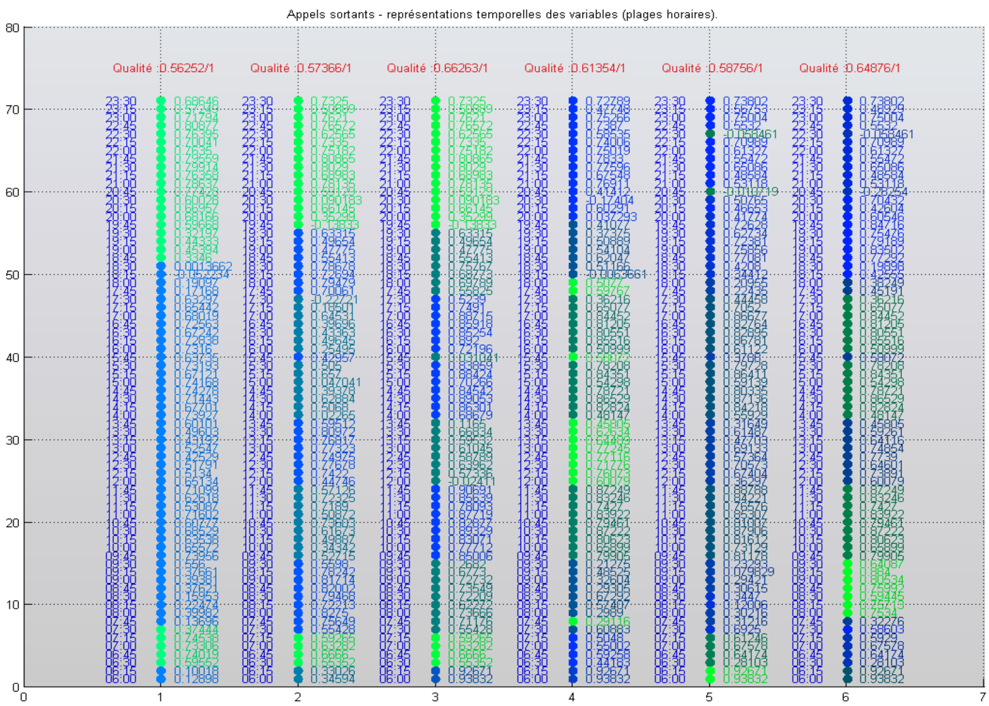


Figure 3.14b. Time periods representation of the “outgoing calls” matrix clustering results

In Figure 3.14b for each time period there is its position on a temporal axis, the group to which it belongs, where the groups are represented by different colors, and its quality within its corresponding group. We can thus divide the day into eight periods:

06:00→ 06:30	06:30→ 07:30	07:30→ 09:45	09:45→ 12:00
12:00→ 14:00	14:00→17:30	17:30→ 19:45	19:45→ 23:45

3.4.4.2 PCA and Clustering Visualization

The aim of this part is to visualize the importance of the traffic in the cells. It is only the first principal component which holds the information concerning this traffic. For that, a clustering is launched using the parameters previously given and we control the color of the groups to the first principal component. The representation of the results on the temporal axis is given in Figure 3.15.

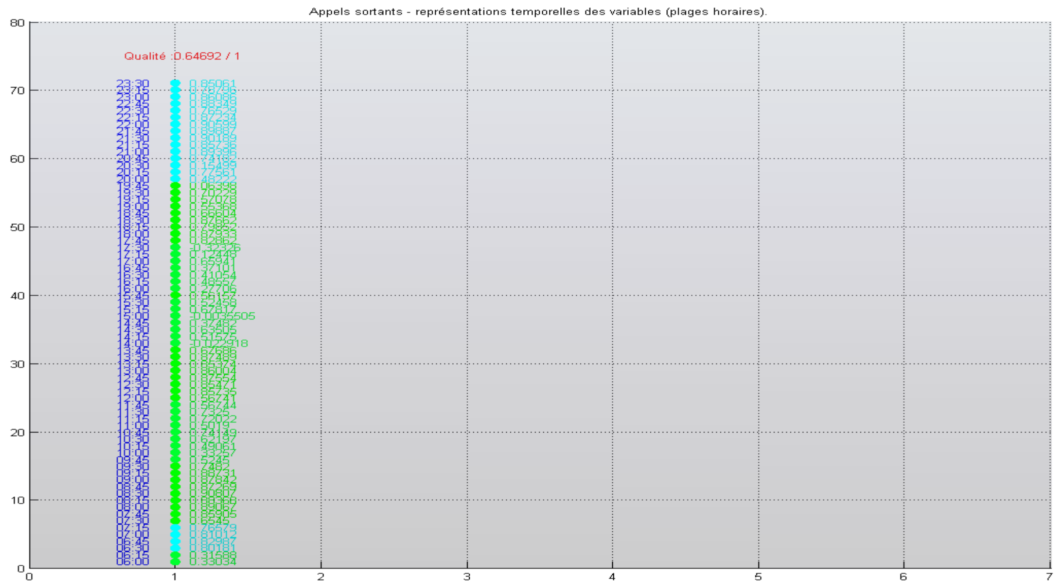


Figure 3.2. The temporal axis (color application according to 1st principal component)

The variables values on the first principal component, except the variables 06:00 and 06:15, are very close. Each variable thus contributes to the characterization of the traffic with the same intensity. Therefore it is coherent to analyze the traffic on each predetermined period and to compare the results of these analyses between them. After adjustment, the day is divided into 7 groups:

06:00→ 07:30 (6 variables)	07:30→ 09:45 (9 variables)	09:45→ 12:00 (9 variables)	12:00→ 14:00 (8 variables)
14:00→ 17:45 (15 variables)	17:45→ 19:45 (8 variables)	19:45→ 23:45 (16 variables)	

We did the same analysis for all days we got, with several files per day category (Table 3a) and we put the masks of cutting per category of day in Table 3b. We see that there are differences between days so it is not possible to consider one environment attraction as constant in time. Then our mobility model must tackle the attractivity variability of environment as parameter to make sense.

Table 3a. One day standard cutting

Type	Files available	References
Monday	3	16/10, 23/10, 30/10
Tuesday	3	03/10, 17/10, 24/10
Wednesday	2	18/10, 25/10
Thursday	2	19/10, 26/10
Friday	2	20/10, 27/10
Saturday	3	14/10, 21/10, 28/10
Sunday	3	15/10, 22/10, 29/10

Table 3b. Masks of cutting per category of day

Monday	6h→8h→ 19h→23 h45
Tuesday	6h→7 h30→ 10h→12h→ 14h30→ 17h30→ 19h30→ 23h45
Wednesday	6h→7h→ 10h→12h→ 14h30→ 17h30→ 20h→23h45
Thursday	6h→7h→ 10h→12 h15→ 14h15→ 17h30→ 20h→23h45
Friday	6h→7 h30→ 10h→12h→ 20h→23 h45
Saturday	6h→8h15→ 11h15→ 14h15→ 19h15→ 23h45
Sunday	6h→8h30→ 12h15→ 20h15→ 23h45
Week	6h→7 h30→ 10h→12h→ 14h30→ 17h30→ 19h45→ 23h45
Transitions – week	6h→9h30→ 11h→13h15→ 16h→19h→ 21h15→ 23h45

3.5 Generic Day Cutting

The studies carried out previously showed that there exist standard cuttings for the studied days and that it is coherent on one hand to separately analyze the identified periods and on the other hand to compare the results of these separate analyses. The aim of the study described in this part is to choose a standard day cutting and to carry out an analysis in principal component as well as a clustering of the cell areas on each period which constitutes one given day. We want to emphasize the variation of the distribution of the mobile population for a given period on a given place; this information will allow us to identify the attraction/repulsion areas along time. In this part we will use the data linked to the mobility inside the mobile network that is *handover* data. The handover is the changing of cells, areas, during a call.

3.5.1 Detection of Zones with Strong Population Density

The whole day analysis by cutting it into several periods not only allows us to study the evolution of the traffic distribution over the day, but also to identify phenomena or behavior related to the mobile phone usage. To illustrate that work we are going to consider some zones of the global area at some specific periods. The complete study has been done on a lot of cases which cannot be described here in totality.

3.5.1.1 Station Behavior Comparison from 10:00 to 12:00

For this study we used the behavior of the stations from 10:00 to 12:00 on weekend, Sunday 15/10/2006, and on workday, Tuesday 17/10/2006. The data were cumulated in the two hours period we like to observe.

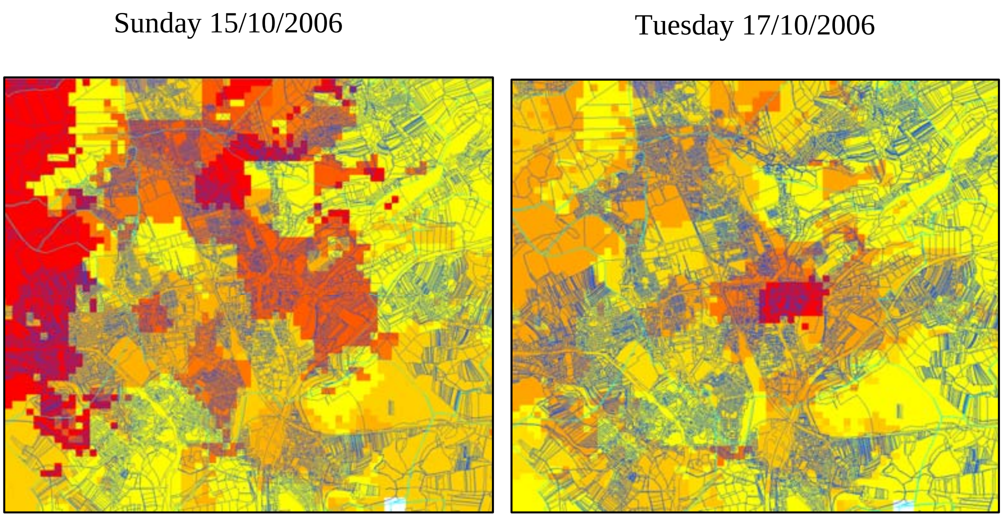


Figure 3.16. Difference of behavior on the same zone at the same time period for two different days

The first observation which can be made by looking at these two charts of distribution of the traffic is that there exist strong differences between these two days. Indeed, individuals seems to be much more distributed all over the map on Tuesday for which three colors of population density are used, excepted in one specific area downtown, whereas five colors of density are used on Sunday so with less homogeneity. The residential zones of the town are characterized by strong traffic on Sunday between 10:00 and 12:00; in the weekend more people tend to stay at home or in their neighborhood, which is not the case for Tuesday, where people are at work and children are at schools and/or universities. This is a working period for Tuesday and a non working period for Sunday. Then we see from the mobile network data for that given period that the identification of attractive areas not only depends on the time of the day but is also specific to the category of day very clearly. A mobility model must also be in charge of the management of day category.

3.5.1.2 Station Behavior Comparison from 19:45 to 23:45

For this study, the behavior of the cells from 19:45 to 23:45 over the same Sunday and Tuesday, as in the previous section is analyzed. Oppositely to the previous section, this is a non working period for both days. The geographical representation of the cell groups shows high similarities in the traffic distribution. We find the same number of colors for the traffic density distribution and we find that for both days the residential zones have high population concentration. This new period confirms the hypothesis we defined before about the necessity to introduce the notion of time inside the mobility models with at least two levels: days and hours. To complete our study we also used data from holidays listed in table 2. The day must be of two possible types *holiday or not* and *weekend or not*.

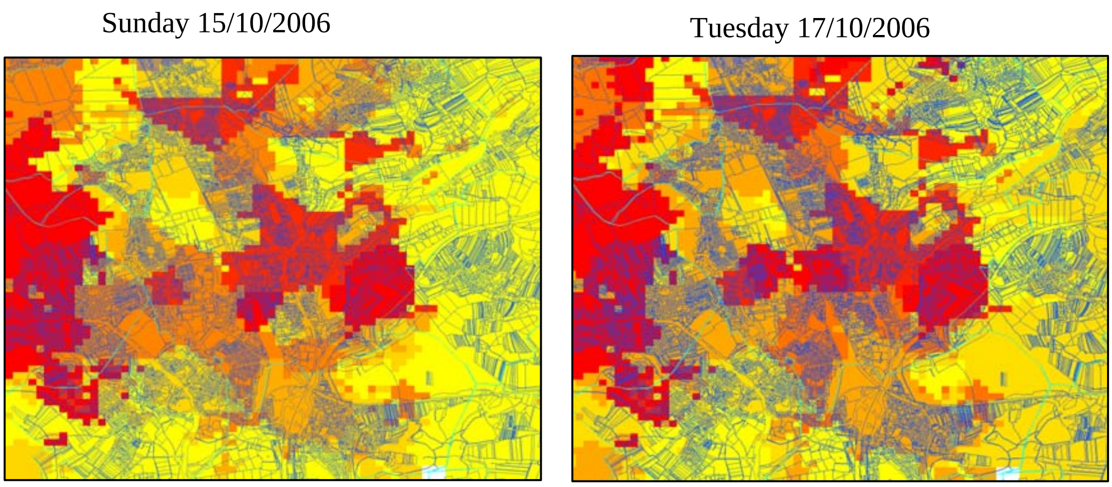


Figure 3.17. Behavior similarities for the same area at the same time period for two different days

3.5.2 Detection of Attraction/Repulsion Poles and Flow Paths

In this part we are going to use handover data to characterize the mobility during the calls. It allows us to measure the changing of zones for a given number of people as one handover is collected if one given call is moving from one area to a neighboring one. We will be able to identify attractive and repulsive zones from this new information.

3.5.2.1 Basic Principles of Handover

A representative example of handover mechanism in cellular networks is given in order to understand the origin of the data we used. We plot three steps.

A) User 1 in the cell A is calling user 2 in the cell B. In other words there is an outgoing call in cell A and an incoming call in cell B.

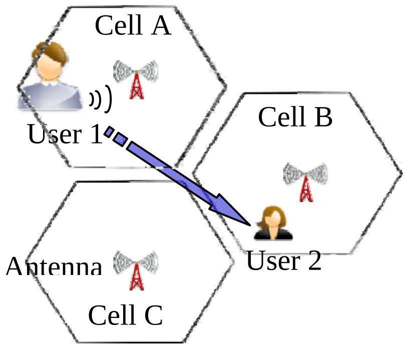


Figure 3.18. Outgoing call in cell A

B) User 1 moves to cell C during the communication. The mobile phone of user 1 must thus operate an intercellular transfer to preserve the communication quality. This event is identified as an outgoing HO (handover) for cell A and an incoming HO for cell C.

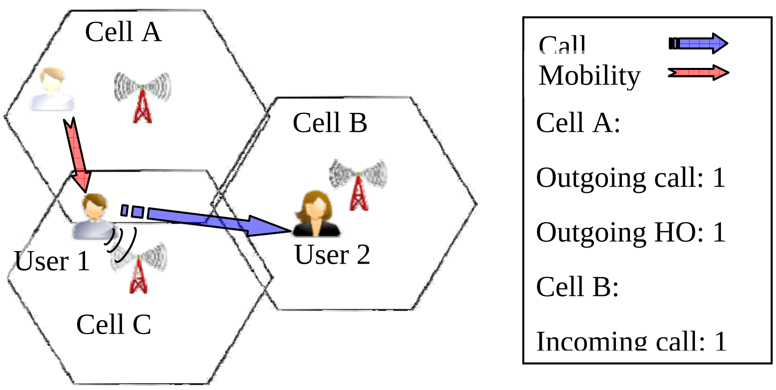


Figure 3.19. Mobility of call from cell A to cell C

C) When we have an incoming HO for cell C all we know is that someone having a communication has entered this cell but we do not know the origin of his displacement. And if we have for cell A an outgoing HO we can only say that someone in communications left the cell but we do not know his destination.

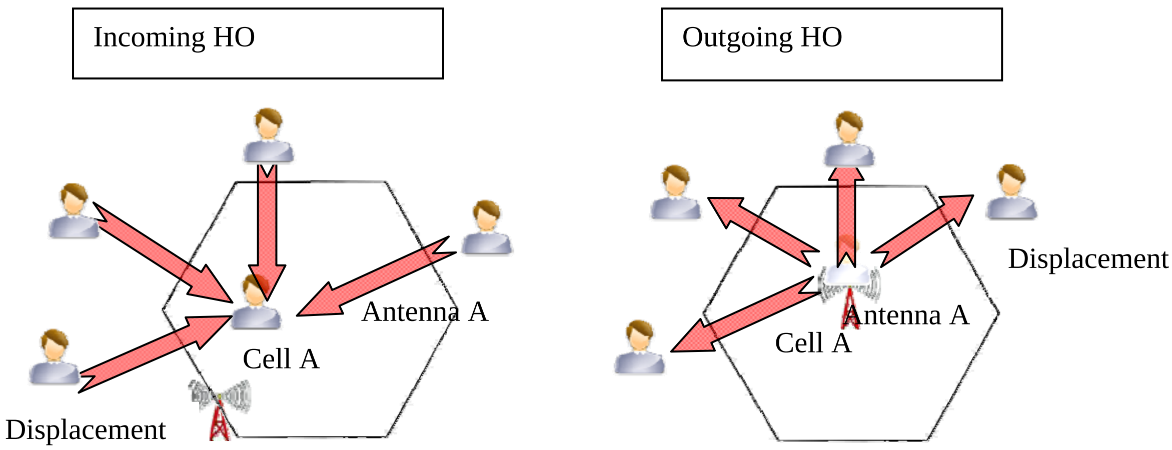


Figure 3.20. Categories of handover

3.5.2.2 Principle of Computation of Mobility from HO

Our study now consists of combining the standard day cut analysis done before on the basis of calls per cell, along with the incoming and outgoing HO in order to identify cells with particular behavior: see the flow between the cells, are they attracting or repulsing people?

The computation for the calls is applied to the HO with the difference that there are three types of data to analyze to characterize and visualize the attractive/repulsive poles along with the population flow. The data are:

— Incoming HO: this allows us the identification of attractive poles.

— Outgoing HO: this allows us the identification of repulsive poles.

— HO difference: the difference between outgoing and incoming HO shows the population flow all over the map.

When the incoming and the outgoing HO are combined into the HO difference it gives information about relative flow paths for each area. If incoming or outgoing HO is used alone, it gives information about absolute flow paths from and into respectively repulsive and attractive poles or cells. As shown in Figure 3.21 the zones with high entering flow can be directly observed. The progressive change of colors and the identification of the road axes allow us to trace the population flows on the map between the areas covered by cells.

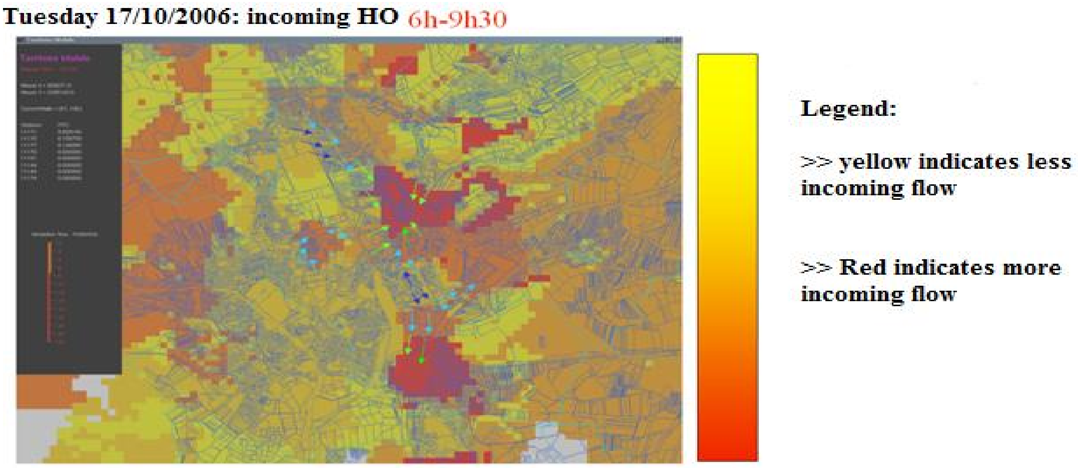


Figure 3.21. Incoming HO for Tuesday 17/10/06 (working day) between 6:00 and 9:30

The display system chosen to visualize the groups is a system known as relative because the color of a group is given according to its numerical position relatively to the position of the other groups on a user given period of observations and not in reference to an absolute criterion such as the maximum number of calls regardless of the cell (location) and time. Thus, for the visualization of the traffic over one day, the cells with the highest number of calls appear in red while those with the lowest number of calls appear in yellow. However, if we compare two red cells out of two different analyses, the number of calls is probably different. If the study aims to visualize the evolution of traffic for a whole day or week for example the solution would be to fix the red color to the highest traffic peak value.

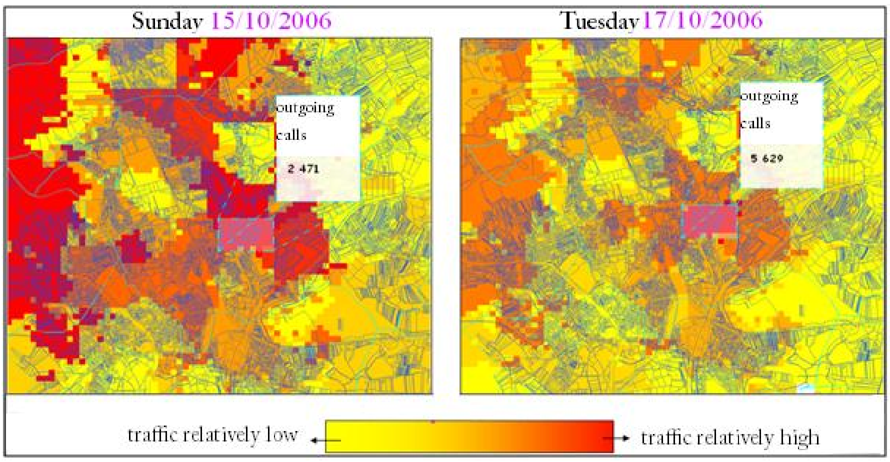


Figure 3.22. Color intensity relatively to the maximum value for this specific period

Finally this information allows us to define real, time varying origin/destination matrixes between one cell and its neighbors for different day categories to use in order to assess the quality of any mobility model. It gives attractive and repulsive areas which must be identified also in simulation conditions when using a model.

3.5.3 Radio Data Analysis Synthesis

We recall the description of the used data in order to clarify our approach in the assignment of weights to the grid cells of the simulation environment.

The weights numerical value proper to each grid cell reflects the relative presence of individual in this grid cell in the simulation environment. In the previous described analysis we have built groups of time periods and groups of antennas in order to assign for each time periods group a weight to the identified groups of antennas.

The data used are the traces of radio calls. These data gives for every ¼ hour period the number of calls for each antenna and the PPC for all the antennas covering each grid cell (*PPC: probability to get a communication*) as described in Figure 3.23. Unfortunately the computation formula is not public.

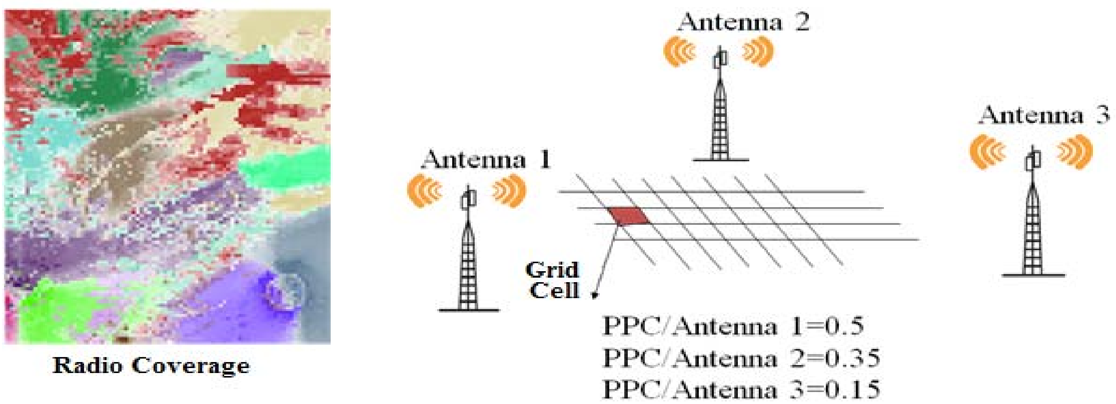


Figure 3.23. Example of PPC for a given grid cell

The radio data does not provide the location from which the call is initiated or received but the potential presence of an individual on a given grid cell, relative to the number of calls (for our case we used the outgoing calls data matrix). The purpose of the performed spatio-temporal clustering is to distribute this calls information on the grid cells of the simulation environment for each identified time period. For example, an area mostly covered by forest has a very low weight for any resulting time period, while an area covered by houses have an average weight for some periods of work (office hours) and a high weight in the evening, where the individuals in the city go back to their residences after work. The weight distribution model results from an experimental model developed in the lab by the research team where this work was done.

3.6 Bus Data

3.6.1 Introduction

In this section the bus company data is analyzed. This data is given for the same days as the network operator data. Two different data matrices are analyzed for each day. The first is the *subscription type/bus stop* which lists all the people with their specific subscription formulas that go in the bus at each bus stop. The second one is the *subscription type/time period* given for each time period, which is equal to 15 minutes in this study, the subscription formulas that go in the buses. These matrices are giving us a lot of new information about mobility but oppositely to the information from the mobile network, the bus network data is specific to a predefined place which is the bus stop. So the information contained in the bus data is for a given fixed position in time and space, while the radio network data was given for radio cells which are characterized by the area of each of the cells covering the geographical area of interest. The subscription type gives us new information about the category of people using the buses so been moving. This information is not available with the phone data. The analysis of the matrices is done to identify the social categories of users who tend to take the buses at the same bus stops (1^{st} analysis) or at the same time (2^{nd} analysis). All analyses are done with the *k*-means method with $k \in \{5,6,7\}$ to analyze and compare the clustering of information.

The information about the bus stops and lines is accompanied by the register of the users that takes the bus. It is a matrix in which the following information is shown:

- Date and time the individual took the bus
- The bus number
- The displacement direction
- Name of the bus stop
- Category of the service subscription
- Number of registered service categories

In order for this information to be analyzed, the data is presented in an individual / variable matrix with the following combinations:

- Bus stop / 15 minutes time period
- Bus stop / 30 minutes time period
- Service category / 15 minutes time period
- Service category / 30 minutes time period
- Bus stop / service category

3.6.2 Analysis of Matrixes Subscription Type/Bus Stop

By checking the type of subscription used by the company we have defined socio-professional categories (referred to as *reference groups* in the following tables) of subscriptions related to the users age and activity. The aim of analyzing this matrix is to check the users who tend to take the bus at the same bus stops. We want to compare these groups using the stops together with the socio-professional categories and to observe if we can establish a bond between them. The question is whether there exists a correlation between the socio-professional category of the users and the use of bus stops and lines of the transportation network of the city.

The table 4 presents twenty two different subscriptions in lines with the detail of the code subscription in column one, then its description, the associated social category and the *reference group* for that subscription. From that table, for the type of bus subscriptions used, we can intuitively define *reference groups* of users with similar social and professional characteristics: a group for subscriptions that are mainly school children, another group for the students (high school, college), a group for retired and disabled guests, a group for working people…

Table 4. Types of subscription and groups by socio-professional categories

CODE	TITLE	Comments	Social/professional category	Reference groups
1	Bus discount 20%	Chip owner	All (Working)	Workings
2	Bus discount 40%	Chip owner (discounted)	Schools / Retired / ….	Workings
11	Monthly TEMPO	Adult classic	Workings	Workings
12	Monthly JUNIOR	Young classic	Schools / Students	Schools / Students
13	Biannual JUNIOR+	Young (CAF)	Schools (-20 years & low revenue)	Schools
14	Monthly AMPHI	Student	Students	Students
15	Monthly UTBM	Student UTBM	Students	Students
16	Monthly SENIOR	Retired	Retired	Retired
17	Biannual SENIOR+	Retired non imposable	Retired	Retired
18	Monthly CONTACT	Jobless, trainees, RMI…	Working	Workings
19	CAT	« Centre d'aide par le travail »	Working / handicap	Workings
20	Subscription CG 90	?	?	? (Workings)
21	Monthly PRO	Working	Working	Workings

23	Biannual ACCES	Disabled person	Retired / Workings…	Workings / Retired
26	Monthly BM AMPHI	Student networks CTRB & CTPM	Students	Students
28	Monthly DEL JUNIOR	School from DELLE	School	Schools
29	Monthly DEL JUNIOR+	2nd children from DELLE	School	Schools
30	Subscription free move	All networks	Working	Workings
40	School CG 90	School peri-urban	School	Schools
60	MIMOSA	Veterans, widows…	Retired	Retired
80	Subscriber CG90 Test1		Test	Test
90	Subscriber CG90 Test2		Test	Test

In table 5 we give the 22 subscription types, the reference groups and the results for the 3 values of k with the k-means algorithm. The value in the last 3 columns of the table gives the reference to the cluster number where the subscription type is being clustered. The codes a, b, c are only used to separate the group names in the columns. The identified clusters contain the subscription types per bus stop.

Table 5. The results of the clustering

CODE	TITLE	Reference groups	5 Clustering groups	6 Clustering groups	7 Clustering groups
1	Bus discount 20%	Workings	2-a	2-b	2-c
2	Bus discount 40%	Workings	5-a	5-b	5-c
11	Monthly TEMPO	Workings	2-a	2-b	2-c
12	Monthly JUNIOR	Schools / Students	2-a	2-b	2-c
13	Biannual JUNIOR+	Schools	2-a	2-b	2-c
14	Monthly AMPHI	Students	5-a	5-b	7-c
15	Monthly UTBM	Students	5-a	5-b	7-c
16	Monthly SENIOR	Retired	2-a	2-b	2-c
17	Biannual SENIOR+	Retired	2-a	2-b	2-c
18	Monthly CONTACT	Workings	2-a	2-b	2-c
19	CAT	Workings	5-a	5-b	5-c
20	Subscription CG 90	? (Workings)	5-a	5-b	5-c
21	Monthly PRO	Workings	2-a	2-b	2-c

23	Biannual ACCES	Workings / Retired	2-a	2-b	2-c
26	Monthly BM AMPHI	Students	5-a	5-b	5-c
28	Monthly DEL JUNIOR	Schools	4-a	4-b	4-c
29	Monthly DEL JUNIOR+	Schools	4-a	4-b	4-c
30	Subscription free move	Workings	2-a	6-b	6-c
40	School CG 90	Schools	3-a	3-b	3-c
60	MIMOSA	Retired	2-a	2-b	2-c
80	Subscriber CG90 Test1	Test	1-a	1-b	1-c
90	Subscriber CG90 Test2	Test	1-a	1-b	1-c
		Clustering quality	0.75	0.71	0.57

In table 5, the first observation we can give, is that there are three persistent groups present in the three clustering:

— Group 1: It includes subscriptions Test of the General Council. This group is not very significant for our study.

— Group 3: It includes all school children domiciled outside the town of Belfort but close to Belfort.

— Group 4: It includes all school children using the bus network of the city of Delle. The city of Delle is outside the Community of Agglomeration Belfortaine, therefore outside our study area.

The other groups which are not the same from one clustering to another are:

— Group 2: It consists of retired people, working people without discount (or small discount) and school children inside Belfort (JUNIOR is for older than 14 years old). This group is the largest one. It seems to be the population moving mainly inside the town of Belfort.

— Group 5: This group seems to bring together people benefiting from highest cuts who are some working people and the students moving in Belfort and around Belfort.

— Group 6: Issued from Group 2, it consists of the subscription free move. These are some school children (under 14 years).

— Group 7: Issued from Group 5, these are the students doing studies in Belfort and close to Belfort (Sévenans at 5km) but not the student going to Montbéliard (15km away).

Thus, if we do not detect distinct clustering results of working and retired people in the choice of bus stops, we observe differences between schools (groups 2, 3, 4, 6), tertiary students (groups 5, 7) and the rest of the population mainly in the group 2. It might therefore be interesting to see people in the territory divided into three main groups of mobility having the same needs for bus stop location: schools (for under fourteen age) / students (for college and university students) / others (the remaining groups).

3.6.3 Analysis of Matrixes Subscription Type/Time Period

In the same way as for the analysis of the matrices *subscription/stop* we firstly grouped the users having close socio-professional characteristics as *reference groups*. A group for the subscriptions which relate to the schools and the students, a group for the pensioners and the people with reduced mobility, a group for the cards with credit, etc.

The analysis of the matrices *subscription/time* is done to identify the groups of users who tend to travel by the bus at the same time. The goal being to compare these groups with the socio-professional categories and to observe if one can establish a bond between them. All that to see whether there exists a correlation between the socio-professional category of people and the period in which they take the bus.

In Table 6 many distinct behaviors are detected between the different socio-professional categories of users. The table shows the clustering results with 5, 6 and 7 clusters. It can be clearly seen that for all the cases three main clusters are present.

- Group 1: It includes subscriptions Test of the General Council. It is not very significant for our study.
- Group 2: It seems to combine the population of elderly and users of the bus with 20% discount. We can infer from this that either the majority of users of credit buses are retired (which is not the case) or the retirees have the same behavior of the users of credit bus without the time constraints of the working hours.
- Group 5: It has mainly all the active people which behavior is correlated to the working hours. Their way of catching the bus seems to be strongly linked to their time constraints.

The other remaining groups are:

- Groups 3 and 4: They represent a majority of school students with variances often depending on the distance to their respective places of study. These two groups are also the group 6, students in college and universities, and the group 7, schools using the bus network of Delle.

Thus, we detect three different behaviors of mobility in time for the following groups: elderly, working people then students and schools. We finally notice that the clustering in 5 groups is the best (having the highest value for *silhouette*) for time period and bus stop analysis.

Table 6. Analysis of one working day, Tuesday, October 3, 2006

CODE	TITLE	Reference groups	5 Clustering groups	6 Clustering groups	7 Clustering groups
1	Bus discount 20%	Workings	2-a	2-b	2-c
2	Bus discount 40%	Workings	3-a	3-b	4-c
11	Monthly TEMPO	Workings	5-a	5-b	5-c
12	Monthly JUNIOR	Schools / Students	3-a	3-b	4-c
13	Biannual JUNIOR+	Schools	4-a	4-b	4-c
14	Monthly AMPHI	Students	3-a	6-b	6-c
15	Monthly UTBM	Students	3-a	6-b	6-c
16	Monthly SENIOR	Retired	2-a	2-b	2-c
17	Biannual SENIOR+	Retired	2-a	2-b	2-c
18	Monthly CONTACT	Workings	2-a	2-b	2-c
19	CAT	Workings	5-a	5-b	5-c
20	Subscription CG 90	? (Workings)	3-a	3-b	4-c
21	Monthly PRO	Workings	5-a	5-b	5-c
23	Biannual ACCES	Workings / Retired	5-a	5-b	5-c
26	Monthly BM AMPHI	Students	3-a	3-b	4-c
28	Monthly DEL JUNIOR	Schools	4-a	4-b	7-c
29	Monthly DEL JUNIOR+	Schools	4-a	4-b	7-c
30	Subscription free move	Workings	4-a	3-b	4-c
40	School CG 90	Schools	3-a	3-b	3-c
60	MIMOSA	Retired	2-a	2-b	2-c
80	Subscriber CG90 Test1	Test	1-a	1-b	1-c
90	Subscriber CG90 Test2	Test	1-a	1-b	1-c
		Clustering quality	0.587	0.561	0.452

3.6.4 Modeling of Displacements to the Bus Stops

In addition to the data analysis we did on the bus stops, we did a modeling of the displacement of people moving from residency areas to the bus stop. The question we tried to answer is the following: knowing a user travelling by bus, identified at the bus stop located in *(x, y)* at time *t*, where was this user located 5 minutes before he took the bus? This work is not used to describe the mobility model in the next chapter, but this analysis was of interest for the company in charge of the bus management to know the area of collection for each bus stop; this data is complementary to the knowledge associated with the bus stop load and is used in the planning of bus stop location. We define two criteria of decision for bus stop selection by the bus user:

1) The criterion of distance: how far is the user from the bus stops? The closer someone is to the bus stop the higher is the probability that he takes the bus at this stop.

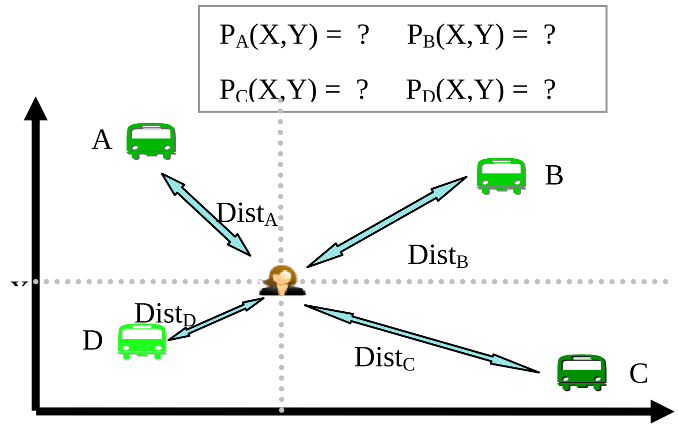


Figure 3.24. Probability to use a bus stop

2) The selection criterion: how a user is going to choose between different close bus stops? The selection criterion aims at determining the probability that a user chooses a stop rather than another. We propose two models to compute this probability.

The aim of this section is to present a mathematical model that can solve systems with multiple weights in order to model the criterion for bus stop selection.

The general formula of the model will be presented with the constraints formulation to respects some requirements. The mathematical demonstration of the model along with a simulation example will be presented too.

3.6.4.1 Model Properties

Let say that a user located in a system has N attraction poles or bus stops. The attraction pole n is located at the distance $dist_n$ of the user. The user has the P_n probability of choosing the attraction pole n. The user must choose a stop. Thus:

$$\sum_{i=1}^{N} P_i = 1 \tag{14}$$

If the user is on stop n, he chooses the stop n.

$$\text{if } dist_n = 0 \xrightarrow{yields} P_n = 1 \tag{15}$$

If a user is at equal distance of all stops, the probabilities to choose any of these stops are identical.

$$\text{if } dist_1 = dist_{2=\dots=}dist_N \Leftrightarrow P_1 = P_{2=\dots=}P_N = \frac{1}{N} \tag{16}$$

3.6.4.2 Presentation, Verification and Demonstration of Model #1

Model #1 is a mathematical model that respects (14), (15) and (16).

$$\forall\, N\ an\ integer;\ \forall\, n \in [1; N];\ P_n = \frac{\sum_{i=1}^{N} dist_i - (N-1)dist_n}{\sum_{i=1}^{N} dist_i} \tag{17}$$

Demonstration:

Assume that we have only two stops, since the probability to choose stop 1 is inertly related to its distance, one way to define the probability to not choose stop 1 is:

$$1 - P_1 = \frac{dist_1}{dist_1 + dist_2} \tag{18}$$

Where P_1 is the probability to choose stop one. Similarly, if we have N stops, then the probability to not choose stop 1 is:

$$1 - P_1 = \frac{dist_1}{\sum_{i=1}^{N} dist_i} + \dots + \frac{dist_1}{\sum_{i=1}^{N} dist_i} = \ (\mathrm{N}-1)\frac{dist_1}{\sum_{i=1}^{N} dist_i} \tag{19}$$

Thus, the probability to not choosing any stop n from N stops is:

$$1 - P_n = \ (\mathrm{N}-1)\frac{dist_n}{\sum_{i=1}^{N} dist_i} \tag{20}$$

Which lead to:

$$P_n = 1 - (\mathrm{N}-1)\frac{dist_n}{\sum_{i=1}^{N} dist_i} \tag{21}$$

Therefore, we have:

$$P_n = \frac{\sum_{i=1}^{N} dist_i - (N-1)dist_n}{\sum_{i=1}^{N} dist_i}, \quad \forall\, N\ an\ integer \text{ and } \forall\, n \in [1; N]. \tag{17}$$

With the following condition:

$$\forall n \in [1; N] \quad dist_n \leq \frac{\sum_{i=1}^{N} dist_i}{(N-1)} \xrightarrow{\boldsymbol{yields}} \forall n \in [1; N] \quad \max(dist_n) \leq \frac{\sum_{i=1}^{N} dist_i}{(N-1)} \tag{22}$$

If condition (18) is not verified this means that an attraction pole has a very small weigh with respect to the other attraction poles which implies physically that the bus stop represented by the weak attraction pole is very far from the actual possible position of the individual. So the probability that this individual chooses this attraction pole is close to zero, for this reason it is better to reduce the dimensions of space doing: N=N-1.

$$\text{if } dist_n > \frac{\sum_{i=1}^{N} dist_i}{(N-1)} \ \rightarrow\ P_n = 0 \,\&\, N = N-1 \tag{23}$$

Verification:

$$\text{for } N = 1 \quad P_1 = \frac{d-(1-1)d}{d} = \frac{d}{d} \tag{24}$$

$P_1=1$

(14) (15) (16)

$$\text{for } N = 2 \quad P_1 = \frac{d_1 + d_2 - (2-1)d_1}{d_1 + d_2} = \frac{d_2}{d_1 + d_2} \tag{25}$$

$$P_2 = \frac{d_1}{d_1 + d_2} \tag{26}$$

$$(14) \rightarrow \qquad P_1 + P_2 = \frac{d_2}{d_1 + d_2} + \frac{d_1}{d_1 + d_2} = \frac{d_1 + d_2}{d_1 + d_2} = 1 \tag{27}$$

$$(15) \ \rightarrow \qquad if\ d_1 = 0 \ \rightarrow\ P_1 = 1 \,\&\, P_2 = 0 \tag{28}$$

$$(16) \ \rightarrow \qquad if\ d_1 = d_2 = d \ \rightarrow\ P_1 = P_2 = \frac{d}{d+d} = \frac{1}{2} = \frac{1}{N} \tag{29}$$

$$(17) \rightarrow \forall\, N\ an\ integer;\ \forall\, n \in [1;N];\ P_n = \frac{\sum_{i=1}^{N} dist_i - (N-1)dist_n}{\sum_{i=1}^{N} dist_i} \tag{30}$$

$$(14) \rightarrow \quad \sum_{n=1}^{N} P_n = \frac{N\sum_{i=1}^{N} dist_i - (N-1)\sum_{n=1}^{N} dist_n}{\sum_{i=1}^{N} dist_i} = \frac{N\sum_{i=1}^{N} dist_i - N\sum_{n=1}^{N} dist_n + \sum_{n=1}^{N} dist_n}{\sum_{i=1}^{N} dist_i} \tag{31}$$

$$\sum_{n=1}^{N} P_n = \frac{\sum_{n=1}^{N} dist_n}{\sum_{i=1}^{N} dist_i} = 1 \tag{32}$$

$$\forall a, b \in [1, N];\ dist_a = dist_b = d \rightarrow P_a = P_b = \frac{\sum_{i=1}^{N} dist_i - (N-1)d}{\sum_{i=1}^{N} dist_i} \tag{33}$$

if $dist_1 = dist_2 = \cdots = dist_N = d \rightarrow \forall n \in [1;N]$

$$P_n = \frac{\sum_{i=1}^{N} dist_i - (N-1)d}{\sum_{i=1}^{N} dist_i} = \frac{N.d - (N-1).d}{N.d} = \frac{d}{N.d} = \frac{1}{N} \tag{34}$$

If the individual is on the n^{th} stop he will choose the n^{th} stop.

$$\forall n \in [1, N] \text{ so that } dist_n = 0:\ P_n = \frac{\sum_{i=1}^{N} dist_i}{\sum_{i=1}^{N} dist_i} = 1 \tag{30}$$

Checking the possibility that he chooses another stop:

$$\forall m \neq n \in [1, N] \text{ such that } dist_n = 0\ \&\ dist_m > 0 : P_m = \frac{\sum_{i=1}^{N} dist_i - (N-1)dist_m}{\sum_{i=1}^{N} dist_i} \tag{31}$$

And we know that: $dist_m \leq \frac{\sum_{i=1}^{N} dist_i}{(N-1)} \leftrightarrow P_m \geq \frac{\sum_{i=1}^{N} dist_i - (N-1)\frac{\sum_{i=1}^{N} dist_i}{(N-1)}}{\sum_{i=1}^{N} dist_i}$

Let $P_m \geq 0$

Now the sum of the probabilities minus the probability P_n=1 using (17):

$$\forall\, m \neq n \ \in [1, N] \text{such that } dist_n = 0 \ \& \ dist_m > 0 \tag{32}$$

$$\text{then } \forall\, i \ \in [1, N]\colon \sum_{i=1}^{N} P_i = \sum_{i=1}^{N-1} P_m + P_n$$

$$\sum_{i=1}^{N} P_i = \frac{(N-1)\sum_{i=1}^{N} dist_i - (N-1)\sum_{i=1}^{N} dist_i - dist_n}{\sum_{i=1}^{N} dist_i} + P_n \quad \text{but } dist_n = 0 \tag{33}$$

$$\text{so } \sum_{i=1}^{N} P_i = P_n = 1;\ \sum_{m=1}^{N-1} P_m = 0 \ \text{ but } \ \forall\ m \neq n \ \in [1, N] \ \ P_m \geq 0 \tag{34}$$

Which leads to the conclusion that: $\forall\, m \neq n \in [1, N],\ P_m = 0$.

This shows that (14), (15) and (16) are verified for all integers N.

3.6.4.3 Presentation and Verification of Model #2

This is another model respecting the three equations (14), (15) and (16).

$$\forall N \ an\ integer;\ P_n = \frac{\frac{1}{dist_n}}{\sum_{i=1}^{N} \frac{1}{dist_i}} \tag{35}$$

Unlike the preceding model, this model does not need a condition such as (17). Note that this model is in fact part of the Huff's gravitational model [47] [48].

Verification:

$$(14) \quad \rightarrow \quad \forall\, N \ an\ integer\,;\ \sum_{i=1}^{N} P_i = \sum_{i=1}^{N} \frac{\frac{1}{dist_i}}{\sum_{i=1}^{N} \frac{1}{dist_i}} = \frac{\sum_{i=1}^{N} \frac{1}{dist_i}}{\sum_{i=1}^{N} \frac{1}{dist_i}} = 1 \tag{36}$$

$$(16) \ \rightarrow \ \text{ if } dist_x = dist_y = d \ \ :\ \ P_x = P_y = \frac{\frac{1}{d}}{\sum_{i=1}^{N} \frac{1}{dist_i}} \tag{37}$$

$$\text{and if } dist_1 = dist_2 = \cdots = dist_N = d \ \ \forall\, N \ an\ integer\colon P_n = \frac{\frac{1}{d}}{\sum_{i=1}^{N} \frac{1}{dist_i}} = \frac{\frac{1}{d}}{N\frac{1}{d}} = \frac{1}{N} \tag{38}$$

(15) is verified for two cases:

If the individual is close to the attraction pole n, the probability of choosing the attraction pole n is close to one.

$$\forall\, N \; an\ integer: (dist_a \neq dist_n)$$

$$\lim_{dist_n \to 0} P_a = \lim_{dist_a \to 0} \frac{\frac{1}{dist_n}}{\frac{1}{dist_1} + \cdots + \frac{1}{dist_n} + \cdots + \frac{1}{dist_N}} \sim \lim_{dist_n \to 0} \frac{\frac{1}{dist_a}}{\frac{1}{dist_n}} = 0 \tag{39}$$

$$\text{and because } (dist_n > 0)\text{: therefore } \lim_{dist_n \to 0} \frac{1}{dist_n} = +\infty$$

The more the individual is close to the attraction pole a≠ n, the closer is the probability of choosing the attraction pole n is to zero.

$$\forall\, N \; an\ integer: (dist_n \neq 0)$$

$$\lim_{dist_n \to 0} P_n = \lim_{dist_n \to 0} \frac{\frac{1}{dist_n}}{\frac{1}{dist_1} + \cdots + \frac{1}{dist_n} + \cdots + \frac{1}{dist_N}} \sim \lim_{dist_n \to 0} \frac{\frac{1}{dist_n}}{\frac{1}{dist_n}} = 1 \tag{40}$$

$$\text{and because } (dist_n > 0)\text{: therefore } \lim_{dist_n \to 0} \frac{1}{dist_n} = +\infty \tag{41}$$

The three equations (14), (29) and (16) are verified for all integers N.

3.6.4.4 Comparative Study of both Models

The aim of this section is to present the differences between the previous models in order to determine which is the best adapted to the representation of the population distribution around the bus stops.

The distribution of the probabilities on a plane representation for both models is shown in Figure 3.25.

It can be noticed that if the dominant zones for the attraction poles are identical for both models, there is a faster decreasing slope for the probability values in this area for Model #2. The population distribution is proportional to the probability values. Thus a higher population concentration is observed in the vicinities of the attraction pole for Model #2 than for Model #1. If the company uses the Model #2 the system will propose a higher concentration of population closer to the bus stop itself. This hypothesis has to be checked on terrain in order to find the good representation of the population distribution.

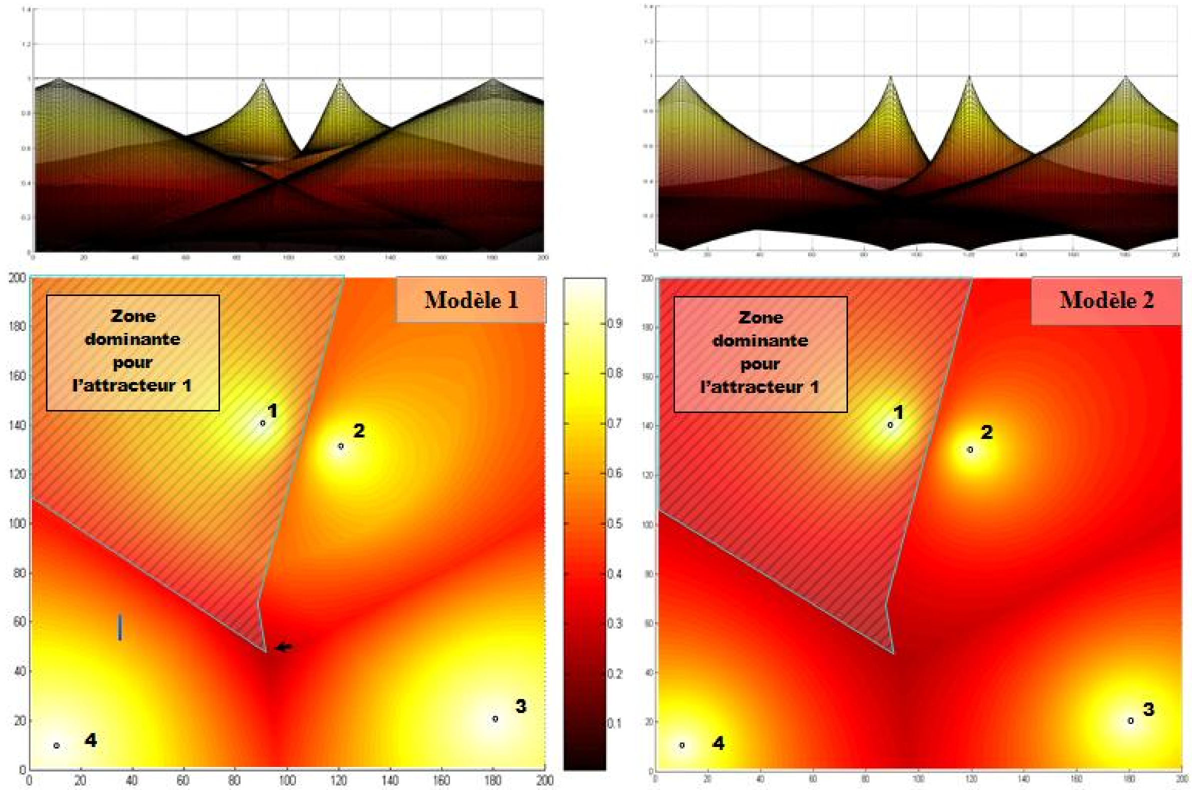


Figure 3.25. Probability distribution for 4 attraction poles with both models and projection on plane

3.7 Synthesis

All the undergone analysis permitted to successfully characterize the environment on which the mobility model simulation will be operational. The detection of the presence and the flows of population in urban and extra urban area helped identify attractive and repulsive zones in the studied area over time; therefore it made possible to create a time variant dynamic gridded simulation environment for the mobility model.

Five major steps were essential for the realization of the work presented in this chapter:

— First of all, the identification of the best suited analysis methods after one deep bibliographical studies.

— Secondly, the analysis and data processing to reduce the matrices dimensions. In This part we used the Principal Components Analysis, then a technique for non supervised and non hierarchical classification, the k-means, and a method to evaluate the quality of the clustering, the *silhouette* computation.

— Thirdly, the analysis of the mobile network operator data in order to divide each day into periods of identical behavior, to detect and visualize traffic flow from and onto areas and to identify the attractive and repulsive zones. These results helped in identifying and visualizing the population distribution on the studied area.

— Fourthly, the analysis of the bus transportation company data used to visualize the individual displacements throughout the studied zone

— Finally the study of the data of buses frequency enabled the visualization of the individual's displacement behaviors according to their socio-professional class.

At the end all the results and analysis were grouped together, to identify the flows and population presence in the chosen area. This helped in the creation of a good simulation environment for the mobility model targeted in this thesis.

Chapter 4. Mobility Model

In this chapter a new mobility model is presented: the Mask Based Mobility Model (MBMM). This model groups several mobility models concepts and ideas and uses Markov chains to choose direction when simulating human displacement. In this chapter MBMM is examined with all its aspects: direction, mask or path correction, speed, grid cell occupation, profile management and mobility leader concept. This chapter will show the major difference and advantages that this model has over other existing models. A set of verification tests along with some simulation scenarios such as 2D and 3D motion are presented to reinforce the theoretical observations.

Content

4.1 The Mobility Model

4.1.1 Introduction

Mobility is defined as the ability of a person to move freely, starting at a location, to a new chosen one, following a given itinerary, with a given velocity and acceleration. A study of individual motion pattern permits a more exact analysis in many fields of simulation like transportation networks, mobile communication networks, etc.

The purpose of our study is to conceive a new generic mobility model allowing a simulation of real human behavior. By real, it is meant that the model takes into account relevant information, dealing with terrain characteristics, urban infrastructure and their impact during the day on individuals and group mobility. The influence of terrain characteristics on the individual mobility is the main focus of our study. These characteristics are changing along the day, creating attraction and repulsion poles, and mobility models are often static in their perception of these periodic and daily variations.

Mobility models play a major role in evaluating performance and reliability of systems which include mobility features. A realistic representation of the individual's motion pattern makes it possible to determine the flow of entry and exit of individuals on a given zone, and consequently the dimensioning of the needed resources in those areas. For instance in mobile networks, the radio operator equipment must be aware of the number of connection people and the intercellular individual transfers. This is called handover, it happens when an individual is calling while in motion and passes from a cell to another. This is very difficult to parameterize, and it represents 70% of cancelled calls in current mobile networks, this value is provided from measurements done by Orange Labs. It could also cover the needs in urban planning such as the case of buses and other transportation networks, and simulate special case motion pattern such as disaster or emergency scenarios.

Territorial features affect on a periodic, often daily basis, the human individual behavior and motion. In order to get a model that shows a higher degree of realism in the motion pattern of a given individual or group, any given map of a city or a zone must be divided into pixels or cells inside a grid (small squares of equal size). In our model, this division is based on the Normal Walk Mobility Model presented previously in Chapter 2. This is done to predefine possible directions and describe trajectory in the territory. Any structure present on this map is represented by one or more grid cells, and individuals move from one cell center to another. In order to guide and draw a logical and realistic motion pattern for an individual or a group moving inside the given map, new layers are added to the map. In these layers the same cell distribution as in the map layer is kept in order to preserve homogeneity. To differentiate the cells and to add another level of granularity to the geographical data presented in the map layer, attraction weights are applied to all cells varying from low values for points or zones of poor or no attraction, and higher values for points or zones with more attraction power.

For the majority of the mobility models in which the individual displacement is the main feature of the model, terrain characteristics and geographical aspects are not taken into account. And in the models where these aspects are taken into account they are transformed into rules and restrictions imposing limitations to the displacement rules. In our approach the attraction weights are function of the individual needs in a given zone. So, one or several given attraction points on a map define the everyday most probable path taken by this individual, there are no specific displacement rules.

In order to add more realism and because individual attraction to some given zones vary with time, the attraction weights layer is dynamic. This is done by dividing a twenty four hours day into five periods; work (in the morning), lunch break (at noon), work again (may be different than the first work period), evening (diner and party) and finally the rest period. For each of these periods a redistribution of attraction weights all over the maps is done. Another period division, a more realistic one for a specific application, may be done after the data analysis. In this thesis report this was applied for radio communication data provided by a mobile network operator, and the analysis of the bus traffic information provided by a transportation company. These processes along with their results were presented in details in Chapter 3.

The mobility model presented in this chapter is called the Mask Based Mobility Model (MBMM). It is a time variant model and at the same time it may be considered as a mobility model with spatial and geographical constraints. The motion pattern in this model is based on Markov chains. MBMM differs from existing models in the fact that the terrain used in this model is dynamic and changes over the time periods. Another particularity of MBMM is that the terrain aspects and characteristics are the main factors in driving and guiding people in displacement. As a result to these characteristics, more realism is added to the mobility model. This comes from human nature giving an individual the ability to adapt to any environment. The map of the region of interest and the geographical socioeconomic data are collected and classified as shown in the previous chapter dividing the area into several classes. The map is hence devided into square cells of equal size. Each class is then assigned with a given weight: the time-attraction weight representing its power to attract people moving in its neighborhood. As a result, each cell in the map has its proper attraction weigh defined by the nature of the covered regions.

In the following subsections, the MBMM is presented in details with all its components. Afterwards, model verification tests are presented in order to verify the relevance of the theoretical model hypothesis and to show the advantages and drawbacks of this model. A practical comparison with other mobility approaches is given in the Chapter 5.

4.1.2 The Mask

4.1.2.1 Motion Principles

As noted before, the mask defines the vision range of individual in motion and the displacement Markov chain. The mask gets its weights values from the grid cells it covers. But this is not all; the major important feature of the mask is its memory. The mask stores the previous chosen direction. In addition to the Markov chain, this will allow biasing future displacement steps with respect to the previous one. This motion correlation will produce realistic motion patterns.

A given individual moves from a point to another with a step size equal to one cell. The simulation environment is divided into square grid cells of equal size. Displacement takes place from one grid cell center to another. In order to simulate all possible directions we suppose that we have inside each square cell an octagon which sides represent the eight possible movement directions, horizontally, vertically and diagonally.

In order to pave a map only three types of cells can be used: triangle cells, square cells and hexagonal cells. Theoretically, if square cells were chosen, and with the fact that a cell can only communicate with adjacent cells, i.e. which it shares a side with, only four displacement directions are possible. For hexagonal cells, following the same analysis as before they provide six possible displacement directions.

However, in order to express realistic motion, a large number of direction possibilities must be taken into account. For this reason, we choose a maximum displacement number of eight. Moving in one of the eight directions is done from one cell center to another as mentioned before. When passing from one cell to another the individual crosses the middle of the octagon side corresponding to this direction.

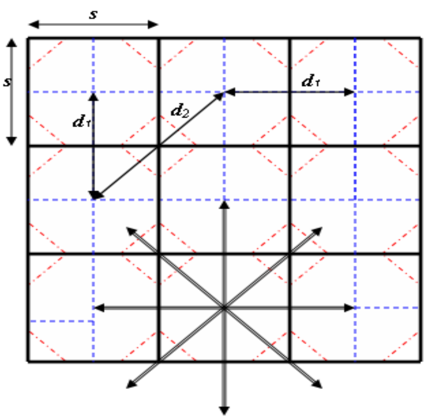


Figure 4.1. The eight possible displacement directions, horizontal, vertical and diagonal

In this model there exist two displacement distances: d_1 corresponding to displacement distances for horizontal and vertical displacements and d_2 corresponding to diagonal displacement.

Given the square side equal to s then:

$$d_1 = s \text{ and } d_2 = \sqrt{2}\, s$$

At each step, direction is chosen according to a Markov chain. For each individual located in a cell of the concerned area, there are nine possible displacement directions, as shown in figure 4.2. Therefore, for each step we define a Markov chain of nine states. A Markov process is used to analyze dependent random events; events whose likelihood depends on what happened before. In other words, Markov chain is based on the concept that all information useful for predicting the future is contained in the present state of the process.

The Markovian probability is defined by:

$$\Pr[X_{n+m} = j_{n+m} / X_{n+m-1} = j_{n+m-1}, \ldots, X_1 = j_1] = \Pr[X_{n+m} = j_{n+m} / X_{n+m-1} = j_{n+m-1}] \tag{1}$$

Where probabilities of transition from state k to state j are defined as follows:

$$\begin{cases} \Pr[X_0 = j] = p_j^0 \\ \Pr[X_{n+1} = k / X_n = j] = p_{jk} \end{cases} \tag{2}$$

And the transition matrix T is given by:

$$T = [p_{jk}]_{q\times q} \text{ and } (j,k) \in E^2 \tag{3}$$

Where P_n represents the instant n states occupation matrix, which is:

$$P_n = [\Pr[X_n=1]\Pr[X_n=2]\ldots \Pr[X_n=q]] \tag{4}$$

Knowing that T is stochastic i.e. the sum of the terms of any line always gives 1 (basic principle of probability definition):

$$\forall j \in E, \sum_{k\in E} p_{j,k} = 1 \tag{5}$$

The proper probability of each cell within the mask is calculated by dividing the weight of this cell i by the sum of the nine weights of the neighbor cells covered by the mask.

After classifying all the data and information regarding the terrain as described in Chapter 3, all grid cells having the same behavior are given same attraction weights as a first step. The next step is to integrate the geographical, topographical and socio-economical data to characterize each grid cell depending on the structure located on it. This result is in the *weight()* function used in calculating the probability to move from one state to another, p_i , as in the formula below:

$$p_i = \frac{weight(i)}{\sum_{k=0}^{9} weight(k)} \tag{6}$$

p_i is the probability to move from one state to another, where the state is a case of the mask, is equal to the weight of the grid cell corresponding to the state divided by the sum of the weight in the 9 grid cells.

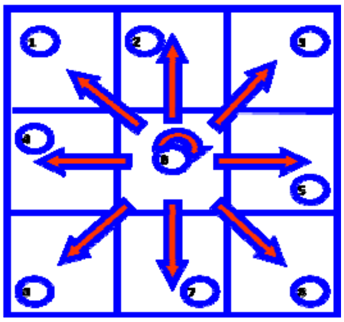


Figure 4.2. Possible displacement directions in the mask at rest

This neighborhood representation is called the displacement mask. Each individual in motion on the simulation area disposes of his own proper displacement mask. This mask represents a close vision range for the individual being on a given grid cell at a given moment. In Figure 4.3 an example of the mask transition between two consecutive steps is given.

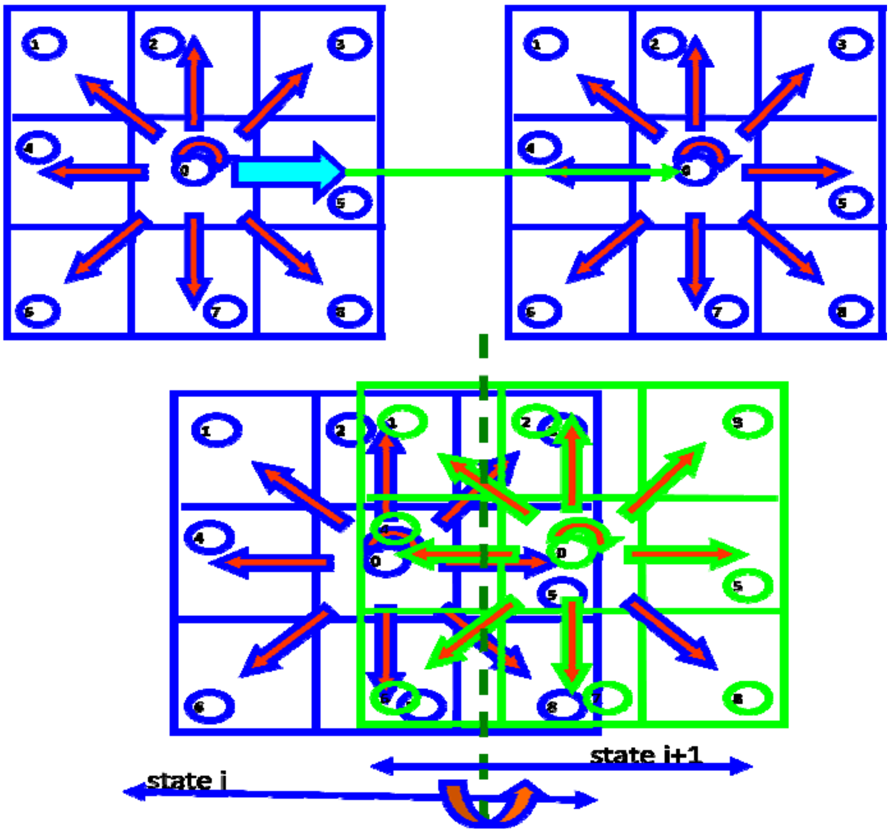


Figure 4.3. Mask sliding from step *i* to step *i*+1

The actual position of the individual is at the center of the mask, and as the mask slides with the individual displacement he will always be placed in the center of the mask. The blue arrow shown in the mask on the left indicates the direction chosen by the individual. This direction is chosen at state *i*. At state *i+1* the individual shifts to the new grid cell indicated by the previously chosen direction. Thus the mask shifts its center smoothly to the new position.

The mask sliding property gives a smooth continuity to the trajectory, but it is not sufficient to provide neither realistic motion patterns nor eliminate sharp angles turn. For that reason a set of mask policies is set:

If a direction is chosen at state *i*, the probability of choosing at state *i+1* a direction perpendicular to the previous direction is diminished by a factor of 0.25; this factor is a parameter calibrated later according to displacement speed.

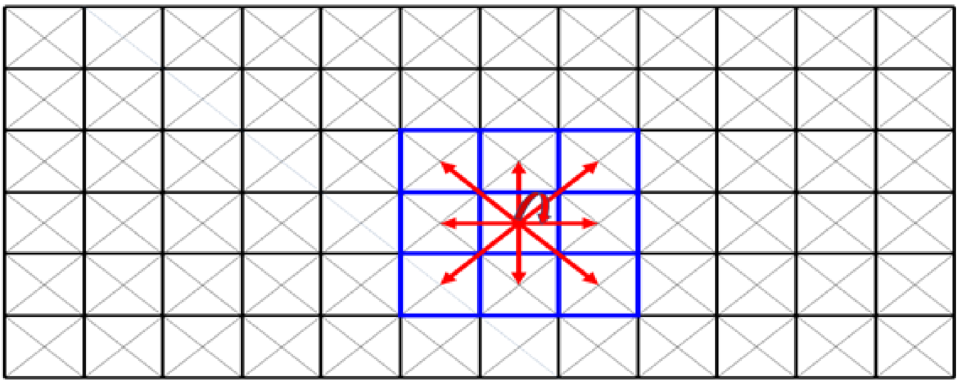
The probability that an individual staying in the same position between state *i* and state *i+1* is reduced by a factor of 0.5; again this is a parameter set according to the displacement speed value.

An individual moving in a certain direction is not permitted to move backwards right or backwards left. This is done by forcing the mask probabilities of these directions to zero.

The last displacement policy states that an in order to move backwards an individual must choose to stay in place. When the individual chooses to remain in the same position from state *i* to state *i+1* all previous parameters are set to 1 making all displacements allowed again.

4.1.2.2 2D Motion Example

The following pictures describe a simple example of direction choosing in MBMM. As shown in Figure 4.4, the individual has just started his journey or is at rest, so the nine possible displacements are allowed for him since his speed is relatively equal to zero. The mask is shown in blue, and the red arrows mark the possible displacement directions.



5	6	8	2	1	5	9	1	3	6	4	7
7	3	3	5	7	4	6	8	5	9	5	2
5	1	2	4	5	4	9	5	7	3	1	8
2	8	3	5	3	7	5	1	4	2	9	5
7	6	10	2	5	8	6	3	5	1	7	4
4	5	2	7	10	2	4	5	8	5	3	2

0,083	0,187	0,104
0,145	0,104	0,020
0,166	0,125	0,06

Figure 4.4. Individual at rest, grid cell weights and relative mask displacement probabilities

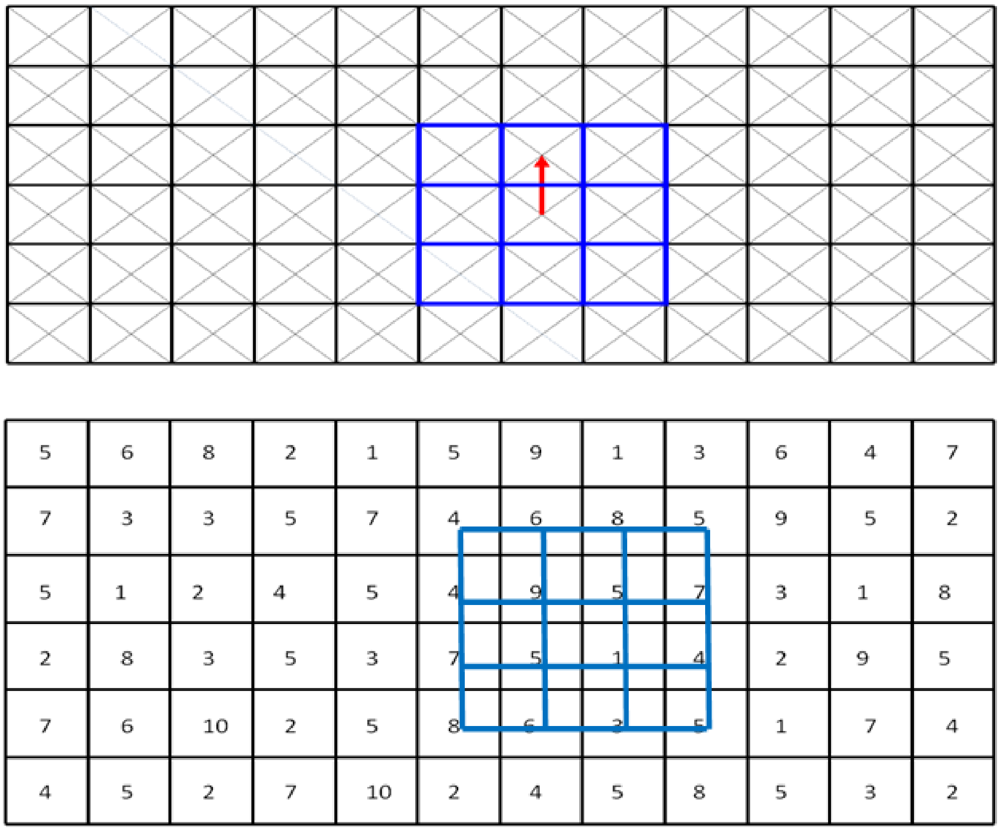
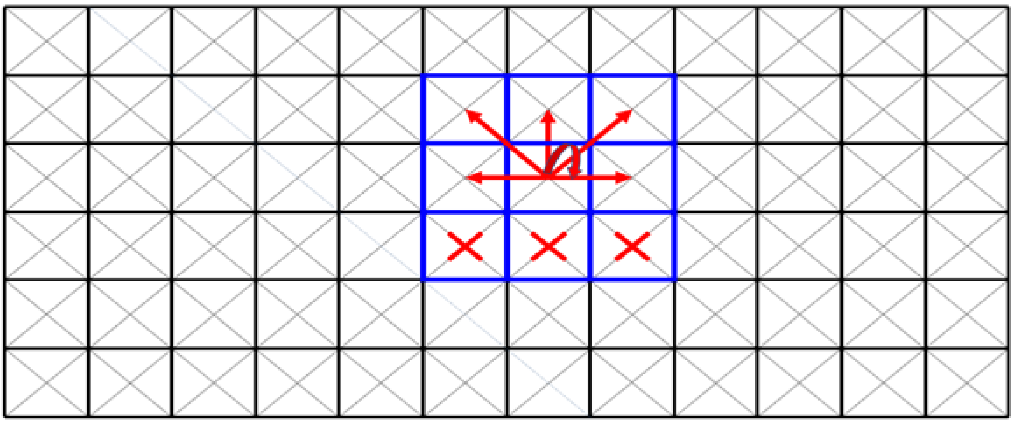


Figure 4.5. First step, the individual chooses to move upwards

In Figure 4.5 the individual chooses a direction which is moving upwards. Once the individual moves to the proper grid cell, the mask follows him and centers his position at the appropriate grid cell as shown in Figure 4.6. Once arrived to the new position new direction choices are possible. This is shown in Figure 4.7.



5	6	8	2	1	5	9	1	3	6	4	7
7	3	3	5	7	4	6	8	5	9	5	2
5	1	2	4	5	4	9	5	7	3	1	8
2	8	3	5	3	7	5	1	4	2	9	5
7	6	10	2	5	8	6	3	5	1	7	4
4	5	2	7	10	2	4	5	8	5	3	2

0,111	0,167	0,222
0,111	0,25	0,138
0	0	0

0,111	0,167	0,222
0,027	0,125	0,034
0	0	0

Figure 4.6. The possible direction choices for step 2 after backwards cancellation

(red X signs mark the position of the directions forced to zero)

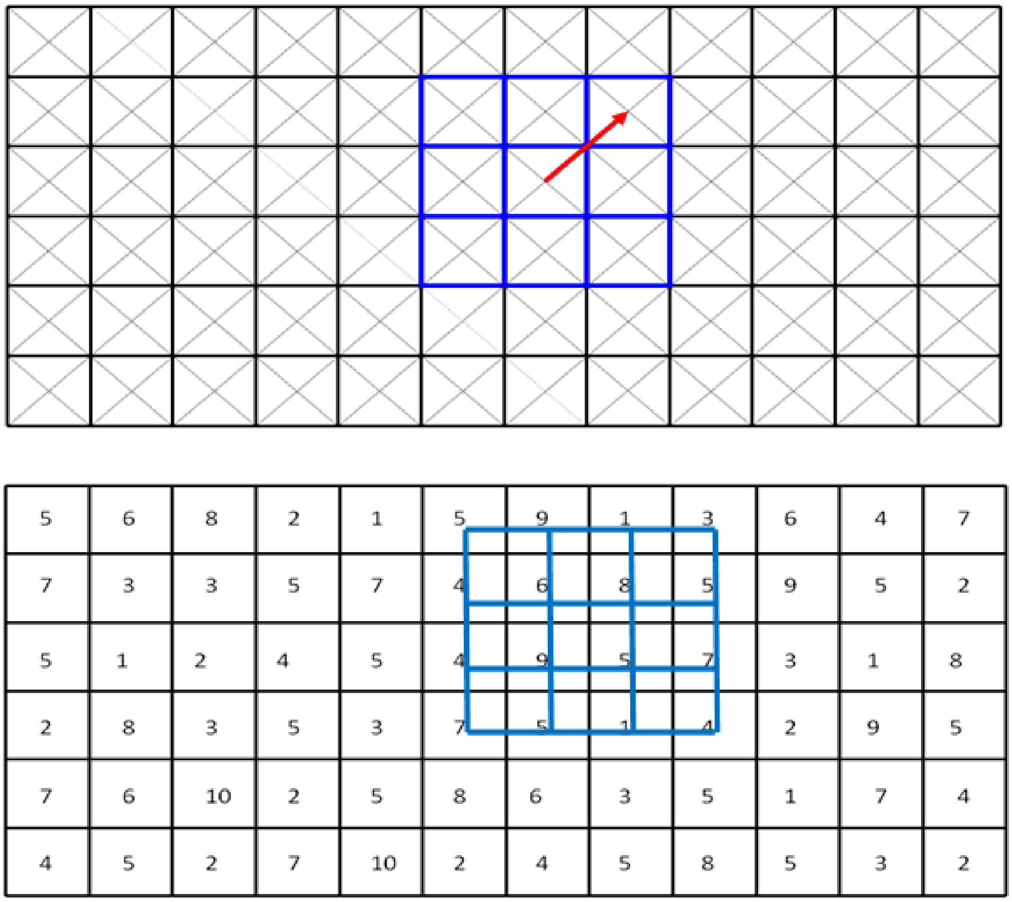


Figure 4.7. Step 3 the individual chooses to move in the up-right direction

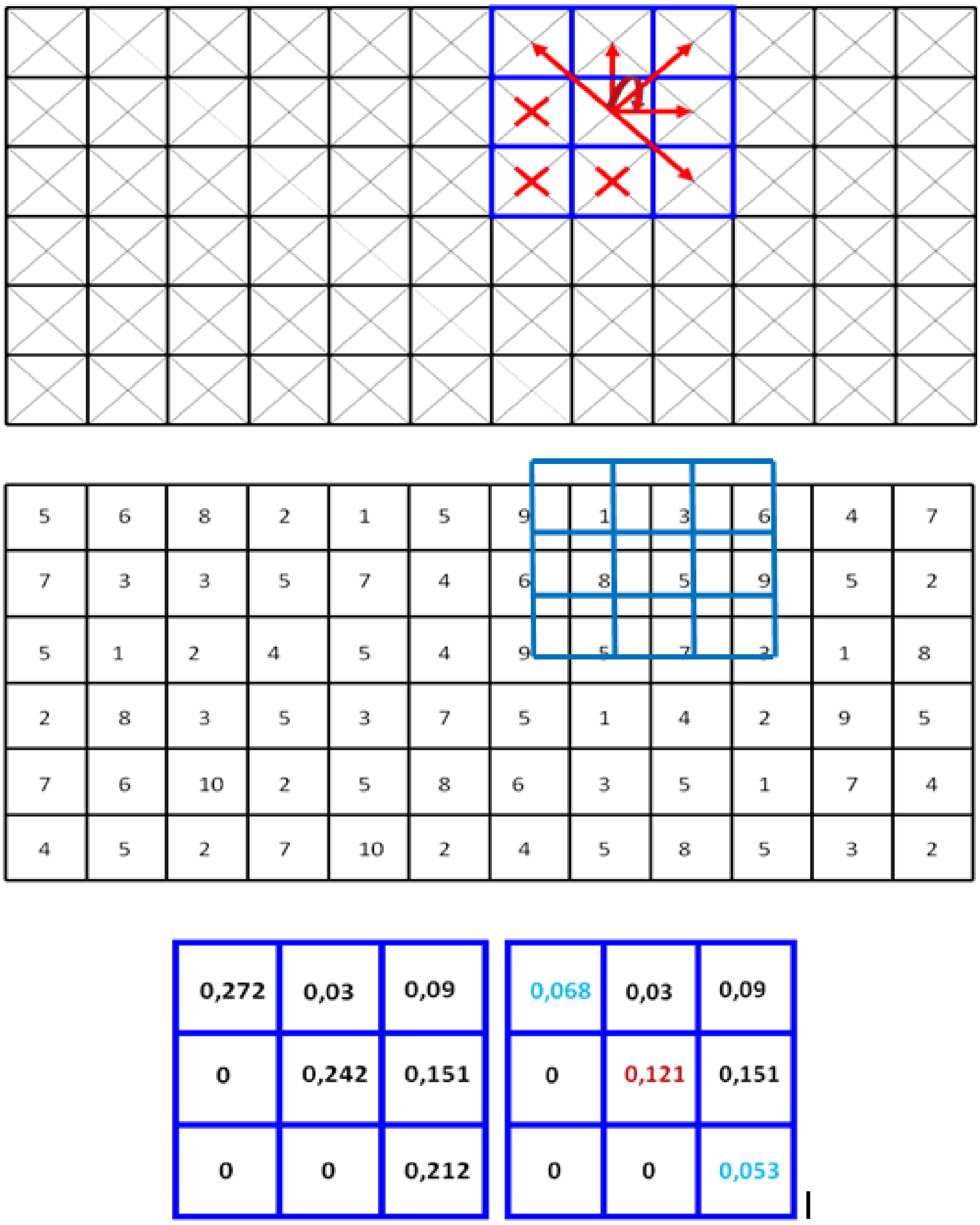


Figure 4.8. The possible direction choices for step 3

(red X signs mark the position of the directions forced to zero)

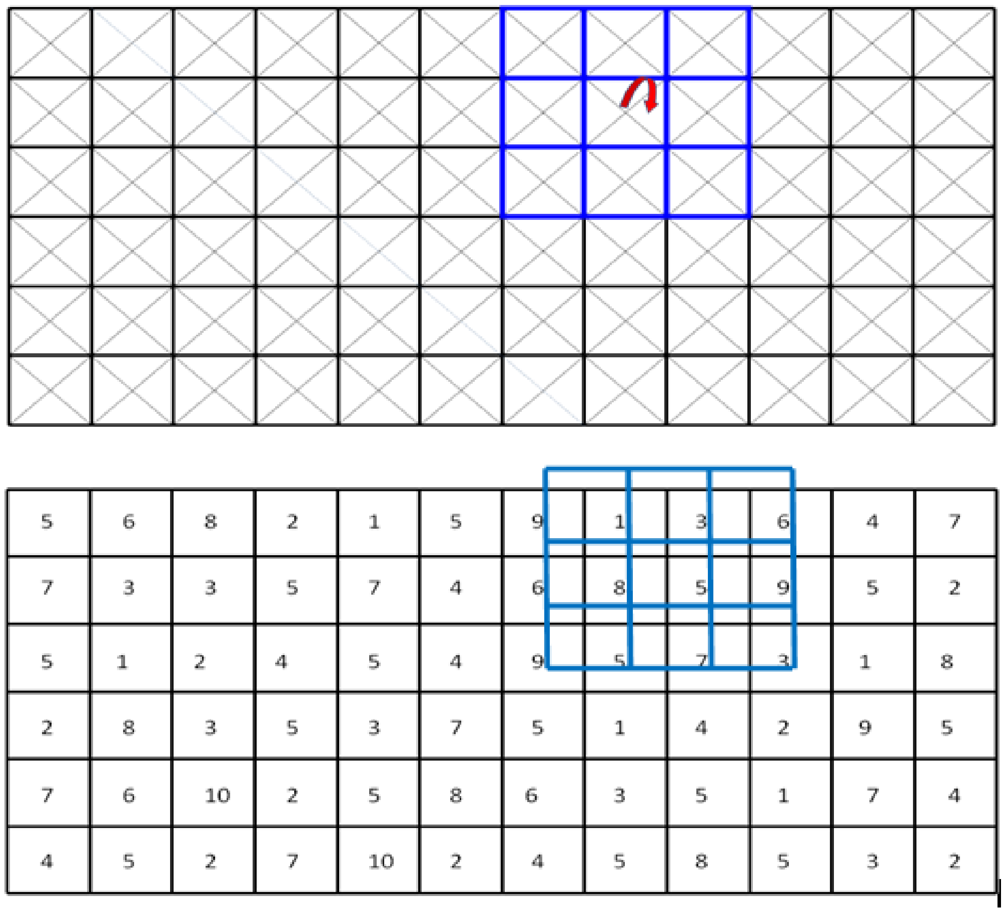
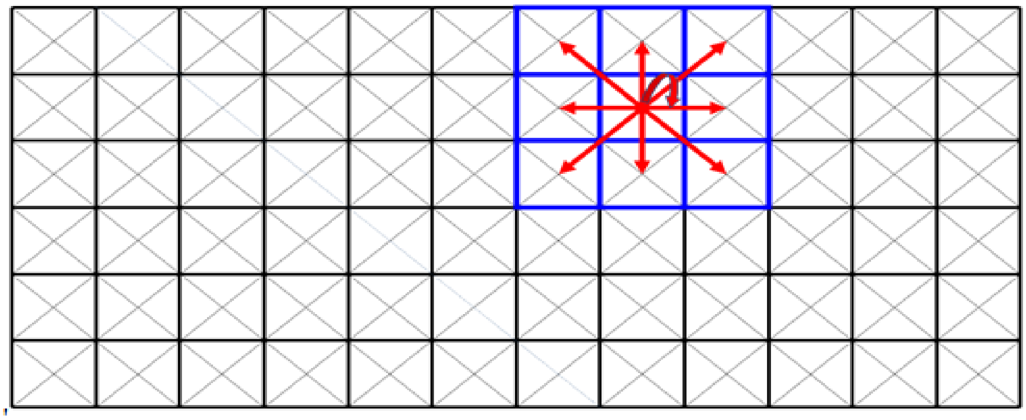


Figure 4.9. Step 4 the individual chooses to stay in place



5	6	8	2	1	5	9	1	3	6	4	7
7	3	3	5	7	4	6	8	5	9	5	2
5	1	2	4	5	4	9	5	7	3	1	8
2	8	3	5	3	7	5	1	4	2	9	5
7	6	10	2	5	8	6	3	5	1	7	4
4	5	2	7	10	2	4	5	8	5	3	2

0,173	0,019	0,057
0,115	0,153	0,096
0,173	0,096	0,134

Figure 4.10. The possible direction choices for step 4

(all directions are possible again since the individual chose to stay in place the previous step)

4.1.2.3 3D Motion Example

This is the case where indoor displacement is presented. This allows individuals to move within the same floor, or between different floors of a building. To allow an easy floor to floor transition, all the floors in a building are assumed to have grid cells of equal number and size. In other words, only one reference point is used for the building.

For 3D displacement a slight modification of the mask is needed. Two more states are added to the mask: going up or down, as shown in Figure 4.11.

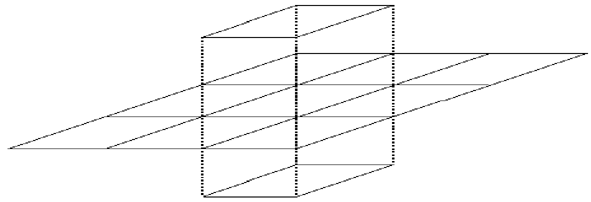


Figure 4.11. 3D mask, same as the 2D mask with two additional states up and down

Moving within a given floor is the same as the 2D case. But to transit between floors, a transition layer between each two floors is added. The transition layer carries the information about elevator openings or staircases. And the rest of the layer grid cells are filled with zeros. Once the individual chooses to go up or down he is transited automatically to the next floor. An example is shown in the following figure.

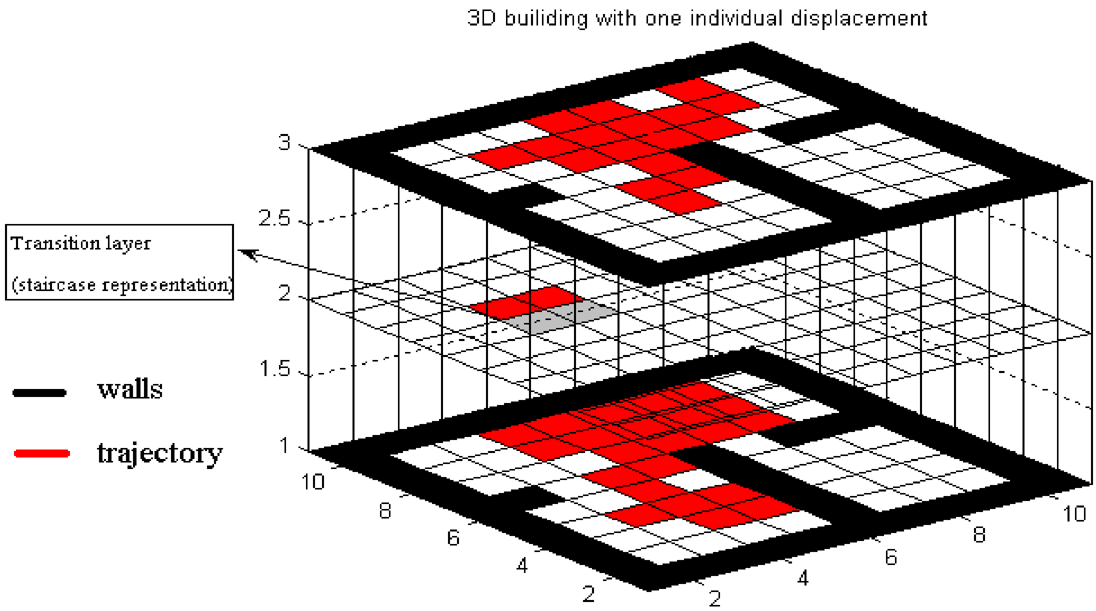


Figure 4.12. Example of 3D indoor displacement for one individual

In Figure 4.12 a simulation of one individual displacement in a two stories building is shown. Therefore only one transition layer is needed. The chosen grid cells, or the motion steps, are marked in red and the walls are marked in black. It can be clearly seen that the individual did move inside the building and transited two or more times between floors. For this simulation a random simulation area was chosen, no real building information was collected to establish a real 3D simulation. By random it is meant that the attraction weights of the cells are chosen randomly not considering any zone or structure within each floor, but it can be done with realistic attractive weights as in 2D model.

4.1.3 The Displacement Speed

Each grid cell presents several kinds of surface structure such as road, building… in a specific percentage for each structure. These structures are accessible by all individuals depending on their own profiles i.e. car will go on street and road, pedestrian will not go on a highway, etc. Thus, these structures impose an appropriate average speed proper for each individual in motion. For example the grid cells covering a highway will have an average speed of 120 Km/h, which is the average speed limit on the French highways, where cars are allowed to circulate with a maximum speed of 110 Km/h for rainy days and 130 Km/h for sunny days. Therefore, in our mobility model, speed declaration and use is specific for each cell and each individual.

For each grid cell, the mask reads the proper average speed values for each individual category. If a grid cell has a speed value of 0 then the grid cell is not accessible by the category of individual in motion; for example, the highway grid cells have their pedestrian average speed set to 0 as no pedestrians are supposed to move on highways. This behavior is represented by using speed coefficients along with the mask. These coefficients are multiplicative coefficients. Their value is multiplied by the value of the attraction weight of each corresponding grid cell. These coefficients can only have two values, 1 or 0. This will allow the passage or not of the individuals in motion on certain grid cells and if the individual is allowed to cross the model uses the speed dedicated to its category.

In order to handle the smooth passage between different speed zones, a test is made before choosing the destination grid cell. The difference between a, the speed value of the central mask position as shown in Figure 4.13 and the other values are calculated and compared to a threshold. If the difference for a given grid cell is greater than this given threshold then the value of the speed coefficient that gave this result will be forced to zero. The threshold value in our case is set to 40, which corresponds to 40 km/h maximum difference of speed between two adjacent grid cells, but it can be changed and adapted to the application purpose.

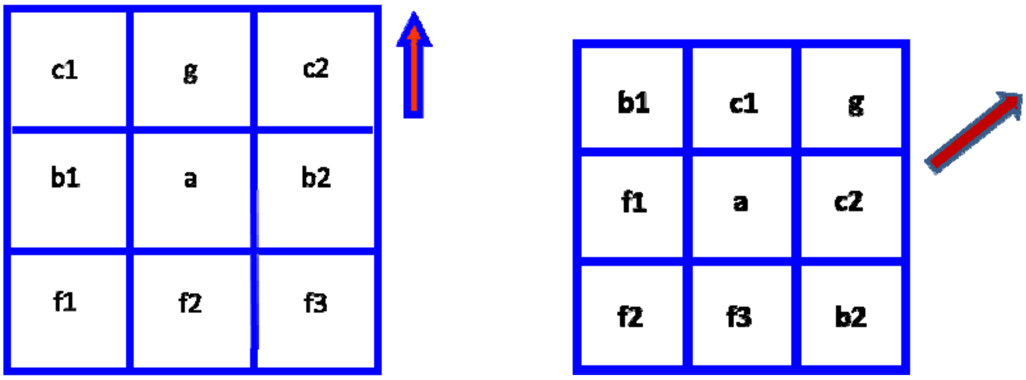


Figure 4.13. Speed coefficients in the mask with moving up (left) and diagonal (right)

It is clear from figure 4.13 on both cases that the *g* coefficient corresponds to move forward, the *a* coefficient corresponds to stay in place, coefficient *f2* corresponds to move straight backwards, *b1* and *b2* correspond to move left or right, etc. so the coefficient distribution is relative to the chosen direction.

For vehicle motion speed varies from one zone to another. Four main values of vehicle speed are identified, divided into two categories: 20 Km/h and 50 Km/h for urban motion and 90 Km/h and 120 Km/h for roads and highways. The different speed coefficients for vehicle motion are given in Table 7. For instance, it says that a vehicle can stop (column *a*) if its speed is 20km/h, it can move left or right at 20 and 50km/h, etc. For pedestrian mobility the speed coefficients do not affect the displacement choice with the exception of the case where a grid cell has an average speed of zero for pedestrians, indicating one inaccessible grid cell.

Table 7. Speed coefficients from cell *a* to the next one

Speed [Km/h]	a	b1	b2	c1	c2	g	f1	f2	f3
20	1	1	1	1	1	1	0	0	0
50	0	1	1	1	1	1	0	0	0
90	0	0	0	1	1	1	0	0	0
120	0	0	0	0	0	1	0	0	0

4.1.4 The Grid Cell Capacity

Grid cells of one simulation area are a representation of the real environment. Each of these grid cells has a certain capacity defining the maximum number of individuals inside it. This value depends on the following factors: the category of surface on the grid cell and the displacement speed of the individual profile. If the maximum capacity is reached for one cell, the cell cannot be used for displacement. In that case its attraction weight is distributed to all other possible directions in proportion to their own attraction.

4.1.4.1 Pedestrian Case

Pedestrians move at low and fixed speed varying between 4 Km/h for women and 5 Km/h for men. For simplification we assume that the pedestrian speed is constant and equal to 4.5 Km/h. This relatively low displacement speed means that for a given time step *t* several individuals can be present on a grid cell. Table 8 gives grid cell capacity for different grid cells volume. Obviously the table can be extended for larger size cells.

Table 8. Grid cell capacity for pedestrians

Grid Cell Area (A) [m^2]	Grid Cell Capacity (C)
$A \leq 1$	$C \leq 1$
$1 < A \leq 5$	$2 < C \leq 5$
$5 < A \leq 25$	$6 < C \leq 25$

4.1.4.2 Vehicle Case

Different factors are taken into consideration when defining vehicle speed, the area (city center, school zone…) and the type of road (low speed, high speed roads…) in which it is moving. In order to define the values for grid cell capacity in this case, presented in Table 9, the following assumptions are taken:

All vehicles in motion on a grid cell have the same speed.

A vehicle has an average length of 3 m.

A vehicle has an average width of 1.5 m.

Table 9. Grid cell capacity for vehicles

Speed (S) [Km/h]	Grid Cell Area (A) [m^2]	Grid Cell Capacity (C)
S = {20,50,90,120}	$1 < A \leq 5$	1
S = {20,50}	$5 < A \leq 25$	5
S = {90,120}	$5 < A \leq 25$	1

4.1.4.3 Example

The following example presents the case of grid cell occupation for the pedestrian case. The individuals are moving at 4.5 km/h. The grid cell area for this example is set to 25 m^2. One individual is moving forward as showed by the arrow on Figure 4.14. The attraction weights to go backwards are forced to 0. The numbers in the mask cells indicate the attraction weight of the corresponding grid cells. The grid cell marked in red already contains 25 individuals, which is the maximum capacity for a grid cell of this size. The weight of this grid cell is replaced by 0 and the original value of this grid cell weight is distributed proportionally to all the other grid cells covered by the mask. This redistribution is performed in a way to preserve the weight order that is the grid cell weight having reached its maximum capacity is proportionally distributed to the other grid cell weights in the mask. This process is presented with computation in Figure 4.15.

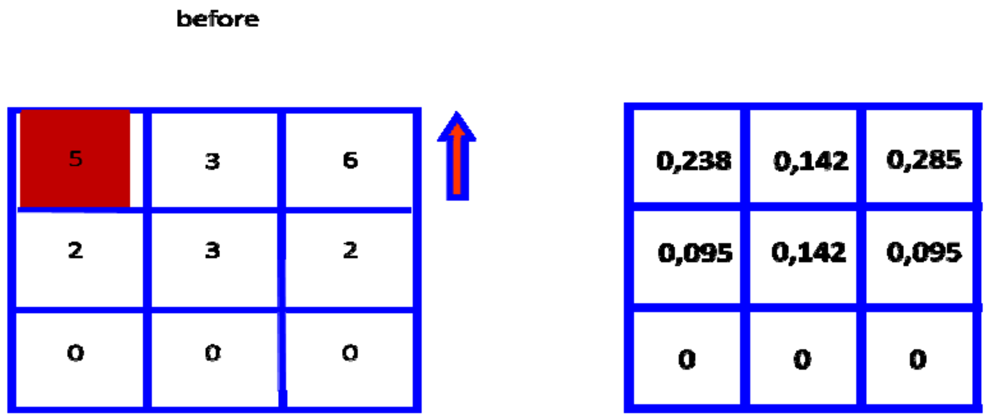


Figure 4.14. Pedestrian mask, with the crowded grid cell marked in red, and corresponding mask probabilities

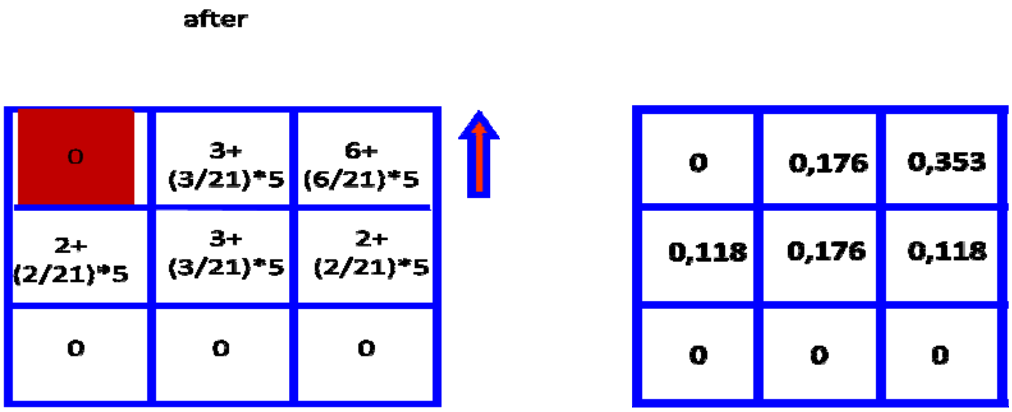


Figure 4.15. Pedestrian mask with the attraction weight of the crowded grid cells distributed to the other grid cells within the mask, and corresponding mask probabilities

4.1.5 The Profile and the Destination

The profile and the destination are two elements used to weight the initial displacement mask of the grid describing the environment according to the individual characteristics.

Individual profiles allow us to characterize different kind of individuals or transportation systems and then different responses of motion in a given environment. For instance a school is more attractive to students than to others. The individual profiles are managed through coefficients. This information is stored for each grid cell along with its attraction weight. Several profiles in the model correspond to several coefficients, therefore, each grid cell is defined by (cell attraction weight, coeff#1 for profile 1, coeff#2 for profile 2, etc.). Depending on his profile the individual will use the corresponding coefficient. This coefficient is multiplied by the cell weight giving the grid cell attraction weight for the correspondent profile. For instance, an individual from the second profile will use coeff#2, the attraction weight value for this grid cell = weight*coeff#2. After calculating the grid cell, each individual has its own displacement map. This allows the model to simulate a large application with different individual profiles. This is one of the major advantages of the MBMM. Most of the existing mobility models are only able to manage one profile at the time and then are not applied to complex cases.

The destination is also used to weight the initial displacement mask. It is linked to the individual profile. The profile defines a set of goals to achieve through its coefficient. These goals are attractive cells to reach because of the surface features, or environmental structures they are covering. Thanks to profile identification, any individual will only see the structures which are of interest to him. Then, each structure of interest for each individual profile is known in advance in the map and will drive the trajectory of all individuals of same profile. This is the main role of the destination information.

Each structure is characterized by its centers coordinates and its occupation percentage. Once defined, the number of persons from each profile that are supposed to go to this structure can be estimated.

Figure 4.16 shows an example of possible destinations for a given individual situated at a position (X_i, Y_i). The individual can choose between three destinations corresponding to its attractive structures, positioned at (X_{s1}, Y_{s1}), (X_{s2}, Y_{s2}) and (X_{s3}, Y_{s3}). Each destination has a maximal load given in number of people for each time period. When a structure occupation capacity is reached for a given period, it is no longer a valid destination for this period. The probability of going to structure #1, #2 or #3 is given by: $a/(a+b+c)$, $b/(a+b+c)$ and $c(a+b+c)$, where a, b and c are the occupation percentage of each structure. The structure is not included for computation if its percentage is already 100%, so one individual cannot choose to go to a fully occupied structure, he has to choose between the available ones.

With this computation, the more attractive structure, i.e. the one having the largest number of people, has the higher coefficient to get further visitors. Its attractiveness is reinforced from the initial value. It gives a dynamic behavior to the map which is computed from the pseudo-random trajectory. From one run to another, the map may have different views. This is very realistic, and is highly correlated to the real world where trajectories are changing with time.

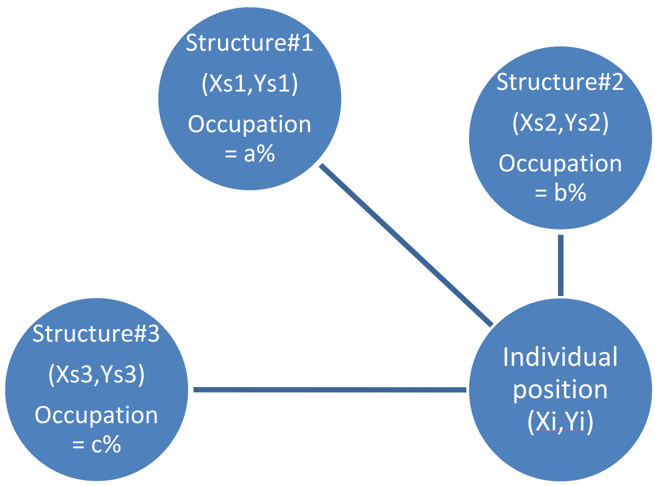


Figure 4.16. Illustration of the structure choice for one individual

After the structure choice is defined, for each individual the destination between his actual position at time t and each of the possible destination structures is computed. Direction is then defined by calculating the difference between the coordinates of the individual's current position and the chosen structure's center. This computed direction allows the biasing of the mask in order to drive the individual faster to the destination structure area. The biasing of the mask is done by multiplying the displacement probabilities by the biasing destination coefficients. The following is a listing of the biasing conditions and rules, where (X_i, Y_i) defines the individual position and (X_{sn}, Y_{sn}) defines the position of the chosen destination structure n.

If $X_{sn} - X_i = 0$ and $Y_{sn} - Y_i = 0$: the individual is at the center of the structure (position 0)

If $X_{sn} - X_i = 0$ and $Y_{sn} - Y_i > 0$: the individual should go up (to position1)

If $X_{sn} - X_i = 0$ and $Y_{sn} - Y_i < 0$: the individual should go down (to position 2)

If $X_{sn} - X_i > 0$ and $Y_{sn} - Y_i = 0$: the individual should go right (to position 3)

If $X_{sn} - X_i < 0$ and $Y_{sn} - Y_i = 0$: the individual should go left (to position 4)

If $X_{sn} - X_i > 0$ and $Y_{sn} - Y_i < 0$: the individual should go down right (to position 5)

If $X_{sn} - X_i > 0$ and $Y_{sn} - Y_i > 0$: the individual should go up right (to position 6)

If $X_{sn} - X_i < 0$ and $Y_{sn} - Y_i < 0$: the individual should go down left (to position 8)

If $X_{sn} - X_i < 0$ and $Y_{sn} - Y_i > 0$: the individual should go up left (to position7)

Now, the coefficients to bias the mask are assigned as input parameters to the model. If the biasing coefficients are high, they will rapidly drive the individuals to their goals, reducing the stochastic behavior, if not the trajectory will be more random. Figure 4.17 represents 2 examples of mask biasing. On the left the destination structure is on cell number 5 (down right) and on the right the destination structure is on cell number 3 (right). The biasing destination coefficients are initialized to 0.25. Then on cell destination it is overwritten by 1, i.e. no coefficient is applied, and on cell destination neighbors it is overwritten by 0.5, giving more interest to these directions. These values are used to multiply the initial displacement probability of the map and it is done for each individual. By applying this, individuals are entirely independent from each other, but the most attractive destinations become more and more attractive proportionally to the number of people already reaching them.

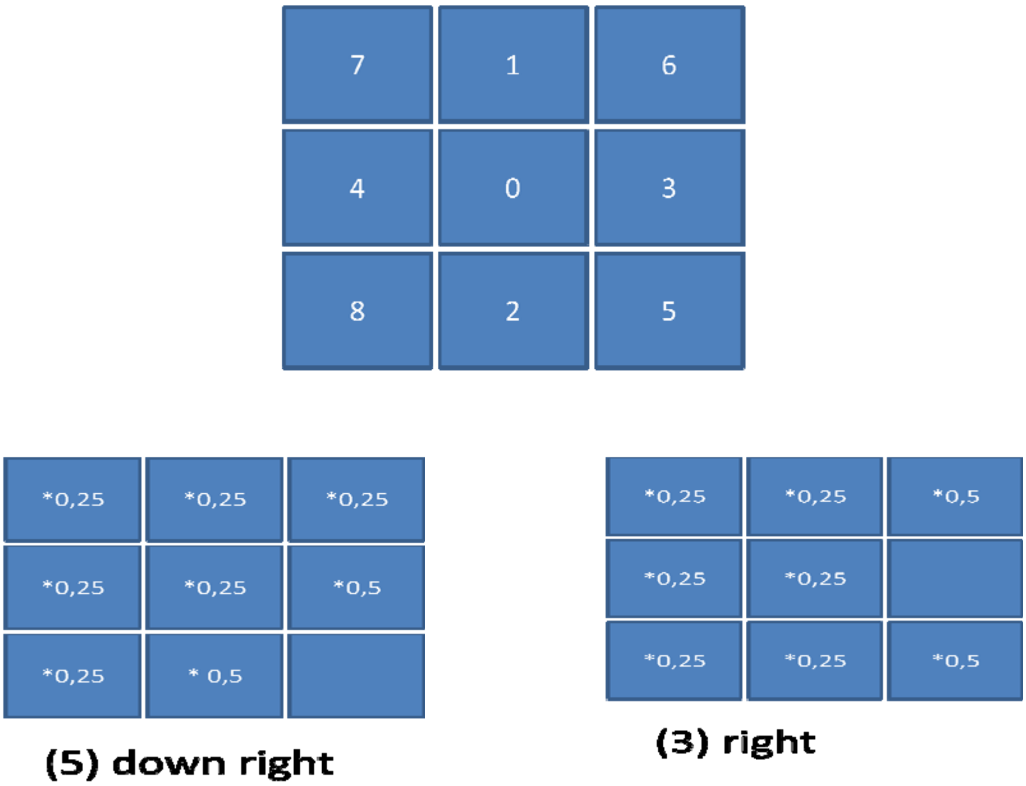


Figure 4.17. Mask biasing based directions and coefficients examples

4.1.6 The Mobility Leader

The mobility leader is defined to try to give some leading behavior to the whole population in motion. When a simulation starts, several individuals start their *journey* at t_0. The number of leaders is related to the population size. It is defined as a percentage of the total number of individuals. The generation zones are defined as any kind of structure present on the map, from which one or more individual will emerge at the beginning of the simulation period. The idea is to have at least one leader in each generation zone that is starting its *journey* at simulation beginning and to directly increment the attraction weights of the grid cells visited by the zone leader. To differentiate leaders using simulation, they will have one or more steps ahead of the other individuals starting in the same area. This weight increment is given by the following formula where p_i stands for grid cell weight: $p_i(t+1)=p_i(t)+p_i(t)*\mu$. The strength of μ will give more or less importance to the leader.

The expected result is to lead more people to follow the trajectory or a part of the trajectory of the leader. This is useful for mass or population simulations because it will divide the whole population of the city to more or less separate groups. This will reduce the population dispersion or distribution on the map and allow the simulation to create clouds of individuals. It is more or less a kind of ant colony behavior where pheromones or reinforcements are used to guide several ants in a given direction but without a strict rule defining the trajectory. It is also possible to say that each time a follower passes by a grid cell already visited by the leader, the attraction weight is further incremented. This will reinforce the leader trajectory in function of the number of followers.

This concept can be applied to each individual class separately by defining a kind of leader per profile. In this case, the grid cell weight is not increased but instead, the index proper to the individual profile for each of the visited grid cells is incremented. Each profile leader will be the leader for individuals of the same profile starting their journey at the same generation zone.

4.1.7 Model Parameters

In this subsection a list of the parameters of the Mask Based Mobility Model is presented. These parameters are indexes used to activate or deactivate features. A listing of all the data contained in a grid cell is also presented.

The input parameters of MBMM are:

- *io*: indoor or outdoor where 0 corresponds to outdoor and 1 corresponds to indoor; the displacement masks for the indoor, 3-dimension, and the outdoor, 2-dimension, cases are not the same.

- *np*: number of profiles to take into consideration to indicate the size of the vector of information stored in each grid cell at each given period. The size of this vector will be equal to *(1+np)*.
- *d*: destination, where 0 means that destination is not used for simulation and 1 means that direction is used for simulation.
- *r*: destination refresh rate, corresponds to the number of minutes after which the destination information is refreshed.

The grid cell parameters of MBMM are:

- The dynamic attraction weight relative to each structure covered by the cell. The computational procedure is described in Chapter 3. The attraction weight of each cell defines the probability that an individual crosses or not this grid cell and the value varies along time.
- A multiplicative factor for each existing profile applied to the initial attraction weights to specify the attraction by individual category. This permits to divide the population into different categories, not only vehicle and pedestrian but any groups within these two categories. For each group, a set of coefficients is given to associate the group interest to the environment structures i.e. roads, buildings, school, shops... These coefficients will generate a specific attraction map for that group.
- A multiplicative factor for each mobility leader applied to the initial attraction weights to specify the additional attraction due to the leader. This will generate a reinforcement of grid cell attractiveness to generate groups of displacement.
- The dynamic cell speed for each existing profile in the simulation. Each individual reads this information according to his profile knowing that for buildings and residential areas the speed value for vehicles is set to 0 and for highways the pedestrian speed is set to 0, i.e. they cannot go through these cells.

4.2 Model Verification

In this section a series of tests validating the basic functionalities of the model are described. The aim of these tests is to validate the characteristics that the model theoretically has.

4.2.1 Mask Statistics

The aim of this test is to validate the behavior of the mask in the choice of the different directions following the coefficients that modify the individual motion pattern.

The direction choices are:

- Forward: to move forward, keep moving in the same direction as the previous step.
- Forward RL: to move forward right or forward left.
- RightLeft: to move right or left.
- Stop: the individual stops. This happens when the individual chooses to stay in place after being in motion, i.e. in the previous step the individual was moving in a given direction.
- Still: the individual chooses to stay at the same grid cell. This is when the individual chooses to stay in place after a Stop or at the beginning of the simulation, i.e. no motion in any direction was made in the previous step.
- New Walk: this value is incremented by one each time an individual decides to restart walking after stopping. If the individual chooses to stand still this indicator is not incremented.
- Error: an error is observed when the individual chooses to move backwards.

An example of the mask statistics graph for one individual doing 1,440 steps is given in the following figure. The initial grid cells are uniformly set to 1. This is done to invoke the mask coefficients in a uniform environment where no guidance of the environment is provided. It can be clearly seen that the individual in motion respects the assumptions taken in the mask displacement policies and the coefficients applied to the mask as shown in Figure 4.17. The individual tends to move forward or forward right and left more than other directions, and there is no sudden backwards (then 0 error).

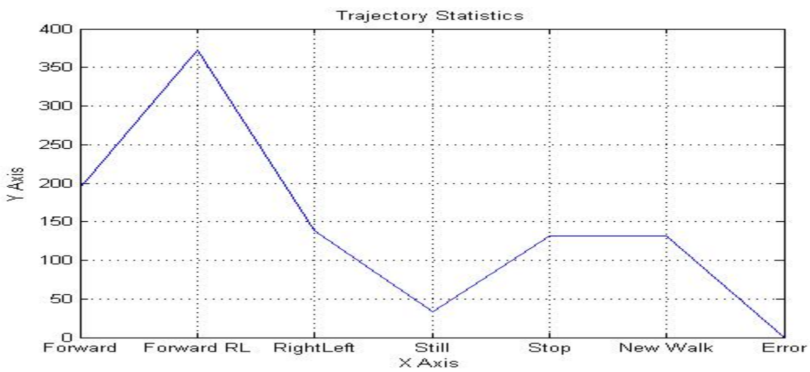


Figure 4.18. Mask statistics example on a trajectory

4.2.2 Direct Path

For the direct path test all grid cell weights are set to 0 and only a straight line has uniform values. The aim is to check if the individual will move on this path and if the mask statistics are adapted to this special scenario.

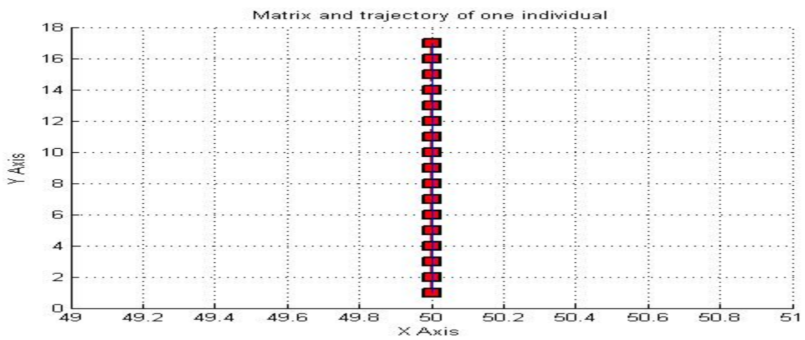


Figure 4.19. Individual following a direct path, red squares show the different individual steps

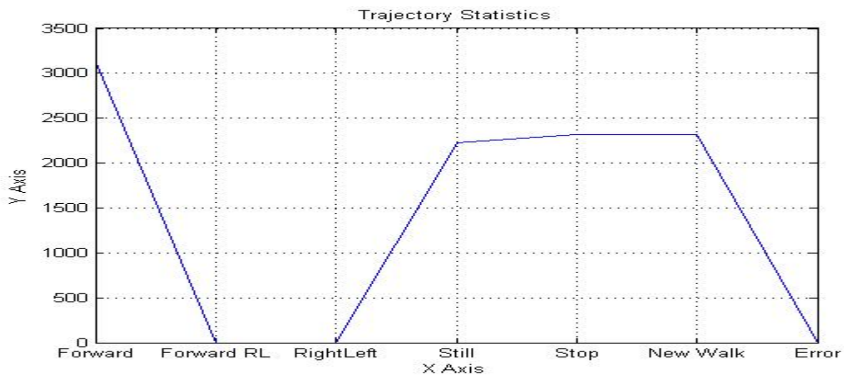


Figure 4.20. Mask statistics proper to the displacement presented in Figure 4.19

From Figures 4.19 and 4.20 it is obvious that the individual followed the straight line as expected, and the results of the mask statistics for this trajectory show that the individual either moved forward or stopped which are the only possible directions.

4.2.3 Noisy Path

This test is divided into two scenarios. In the first one we dispose nine vertical and parallel consecutive lines were the central one (number 50 in figure 4.21) has the maximum weight value. This value decreases gradually towards the borders of the zone (right and left). Figure 4.21 shows the motion pattern of the individual displacement in the scenario zone. It can be noted that the steps (red squares) are concentrated towards the center of the zone; number 51 is the mean just as expected. In Figure 4.22 the mask statistics for the trajectory show that the displacement follows the mask rules as in the general case.

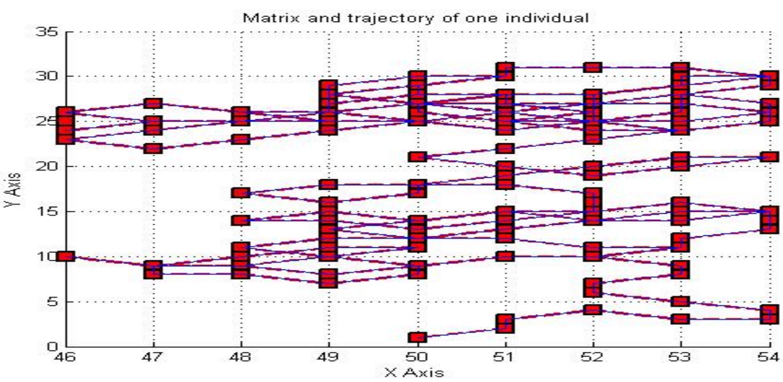


Figure 4.21. Noisy path scenario 1

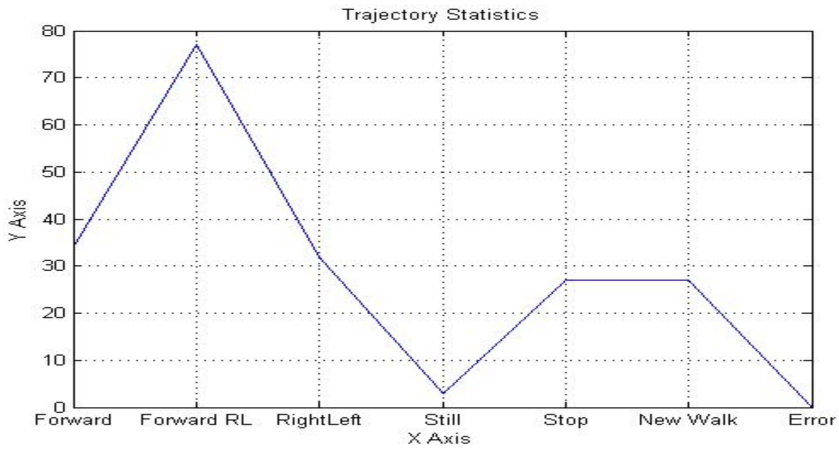


Figure 4.22. Noisy path scenario 1 displacement mask statistics

In the second noisy path scenario also nine vertical and parallel consecutive straight lines having non zero weights are introduced, just like in the first scenario, but the right and left borders of this active simulation zone have the highest weight values instead of the central one. The weight values decrease gradually to the center of the zone. This is why in Figure 4.23 it is obvious that there are more individual steps, shown by the red squares on the figure, concentrated on the sides rather than in the center. But it is still possible for the individual to cross the area from one side to the other even if the center attractiveness is lower.

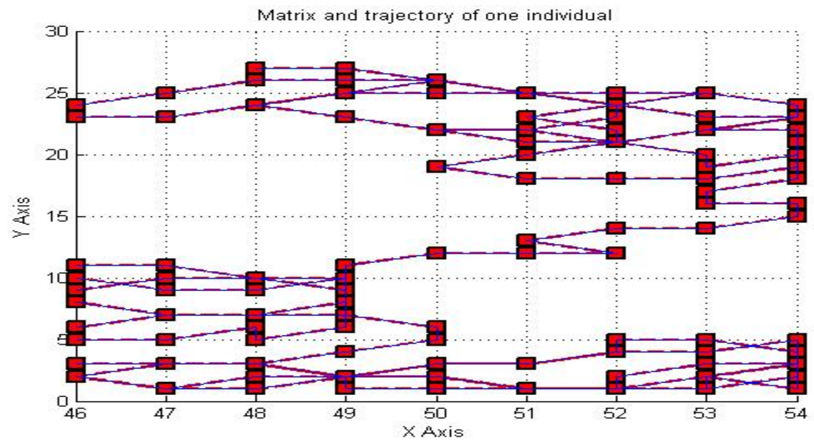


Figure 4.23. Noisy path scenario 2

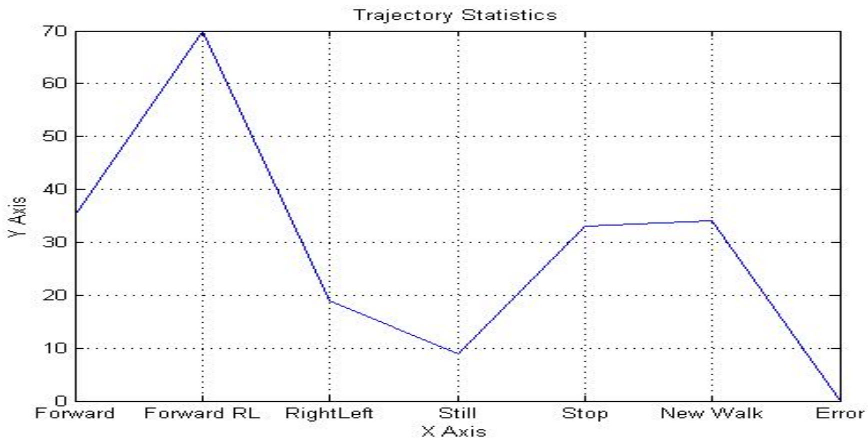


Figure 4.24. Noisy path scenario 2 displacements mask statistics

The displacement mask statistics shown in Figure 4.24 show that the individual motion pattern always follow the mask coefficients and are similar to the general case. Using both tests, it was proved that even inside a pseudo random motion, driven by the map, individuals follow some emerging pattern, they go several times from one side to another.

4.2.4 Convergence to a Zone

In this test several zones in the general simulation area are defined: active attraction pole represented in Figure 4.25 by the green squares, and non active attraction poles represented as white squares. The individual is positioned on the general map and starts its trajectory. The general matrix is indicated by blue dots, the individual steps are in red.

The aim of this test is to check if the individual will be able to reach the active attraction pole, i.e. his destination, with a fixed number of steps. The number of steps is set to one step per second and the simulation period is two hours. This results is 1,440 steps. The success rate for these settings is 93%. It goes up to 100% if the number of steps is increased, and decreases if the number of steps is decreased.

The convergence is faster if the area of the attraction pole is enlarged by several radiation layers around its area. This will spread the information about the presence of an attraction pole to the individuals in motion. Obviously, this diminishes the number of steps because the attraction pole size grew. Convergence problems were mainly observed for random maps scenarios but were solved in the real cases scenarios after characterizing the terrain properly.

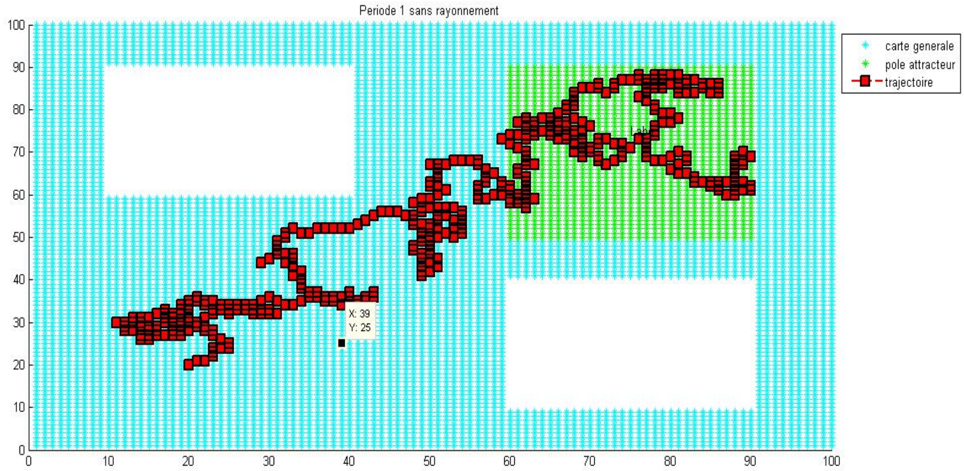


Figure 4.25. Individual in motion reaching his destination

The tests were performed on an early stage to validate the basic functionalities of the model. These tests are not to be considered as applications of the MBMM. Two real applications of the MBMM model are given in Chapter 5. The first one is theoretical application. It is a comparison of the MBMM with other mobility models existing in the literature. The second one is an application of mobility to UMTS cellular networks. MBMM is used to compare several algorithms dedicated to uplink power control. This is a practical application.

4.3 Conclusion

A new mobility model was presented in this chapter; the Mask Based Mobility Model (MBMM). This model combines the advantages of several other mobility models and creates a generic mobility model, suitable for transport planning, network dimensioning and many other applications

The basic concept of the mobility model is the displacement mask which was validated by a series of simulation tests and statistics. The motion principle in this model is Markovian with addition to a short memory of the last previous step. The individuals are divided into groups based on their needs on the simulation area during a day. This allows the user to set a series of displacement policies, when applied, make the simulated trajectories as close as possible to a real track. The main features of this model were validated by a series of tests with a simulation area based on the statistical analysis performed in Chapter 3. The model offers a wide range of application since it can handle individual and population motion and individual profiles and needs. Applications of this model are presented in the chapter 5.

Chapter 5. Applications

The aim of this chapter is to present applications in different fields using the Mask Based Mobility Model. Two major applications were developed throughout the thesis. The first application is a comparison of mobility models well known in the literature with the mobility model described in chapter 4. The simulation environment is a time variant real representation of a city based on GIS, integrating all geographical and socio-economical information relative to this city. Simulation results are the basis of the comparative analysis between the models. This comparison is made by setting metrics evaluating individual and population displacements and by quantifying the degree of realism of each model. The second one is an application of the Mask Based Mobility Model for the assessment of power control algorithms efficiency in UMTS networks. The aim of this study is to analyze the impact of mobility on the power control procedure in UMTS network through several power control mechanisms. After presenting the power control procedure and algorithms in UMTS, a description of the simulation environment, scenarios and results will be given. These works are published at NTMS'08 and SIMUTOOLS'09.

Content

5.1 Comparative Study of Mobility Models based on Simulation

5.1.1 Introduction

Mobility models are a key element in simulating human (individual/population) motion and displacement for telecommunications, transport, etc. These models become a necessity with the emergence of mobile networks in the last decade and all the mobile services proposed or to be proposed in the future. But with the large number of emerging mobility models; how can we choose the best adapted model for a given application?

The aim of the work presented in this paper is to perform a comparison between different well known mobility models called Random Waypoint Mobility Model [5] [6] [7] [12] [14], Random Walk Mobility Model [68], Normal Walk Mobility Model [3] [4], Smooth Mobility Model [2] [1] and Markovian Movement Model [74] [75], and both models we have developed, the Normal Markovian Mobility Model and the Mask Based Mobility Model.

Individual trajectories and population motion are simulated in a real environment representing an area of twenty per forty kilometers centered on the city of Belfort in France. Several simulation runs are performed for each of the mobility models enumerated previously.

Based on these simulations an evaluation of each model is done by quantifying its characteristics and evaluating it with several new metrics. These metrics are about the understanding of the individual motion pattern.

This chapter is divided into three major parts. Firstly we begin by presenting the simulation environment we used, which is a platform that we developed. Secondly simulation of the mobility models is done in this environment. And in the third section we present the evaluation metrics in details along with the obtained comparison results. The synthesis summarizes the major results of the simulation and some perspectives.

5.1.2 Simulation Environment

The environment we used for the simulations is a platform developed in C++ using the OpenGL library.

This platform is a representation of Belfort, a city in the East of France, along with its suburbs and surrounding villages, as shown in Figure 5.1. The area is about 20km*40km and the population size is around 140,000 people. The information used for environment description is the same than the information described in chapter 3. We remind the information below.

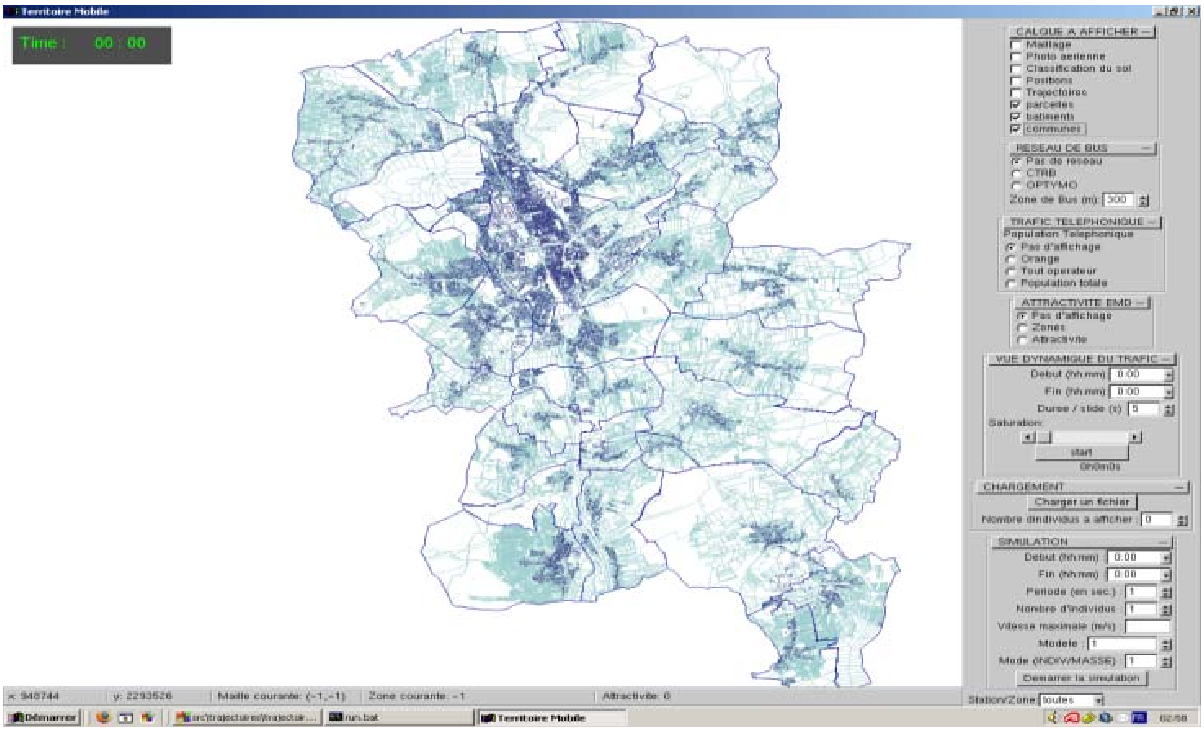


Figure 5.1. General view of the simulation platform

Several layers of data were used to reproduce a real environment, geographical data and socio-economical information.

Three display modes are available:

<u>Air view (Figure 5.2):</u> This is done using a geo-referenced satellite image of the region.

Figure 5.2. Air view mode of a part of the city of Belfort

GIS view (Figure 5.3): In order to use GIS data such as shapefiles, a specific module was developed. All shapefiles available along with their relative databases (dbf files) were used to get the closest to reality environment.

Figure 5.3. GIS view mode of a part of the city of Belfort

Ground Classification view (Figure 5.4): This layer of data is the aggregation of all the provided data. This layer is necessary for the simulation of mobility models taking into account the ground characteristics. The environment is divided into square grid cells, with a side length equal to 25 meters. Each of these grid cells is characterized by a value computed after the data classification based on the dominant structure present in this grid cell and all other sources of information, this value is called the attraction weight. This explains the different colors in Figure 5.4, where each color corresponds to a class of information. The grid cell values vary through the day; for example a grid cell containing a restaurant will have greater values between 12:00 pm and 2:00 pm and at night from 7:00 pm than for the rest of the hours of the day. The process of weight definition is detailed in chapter 3.

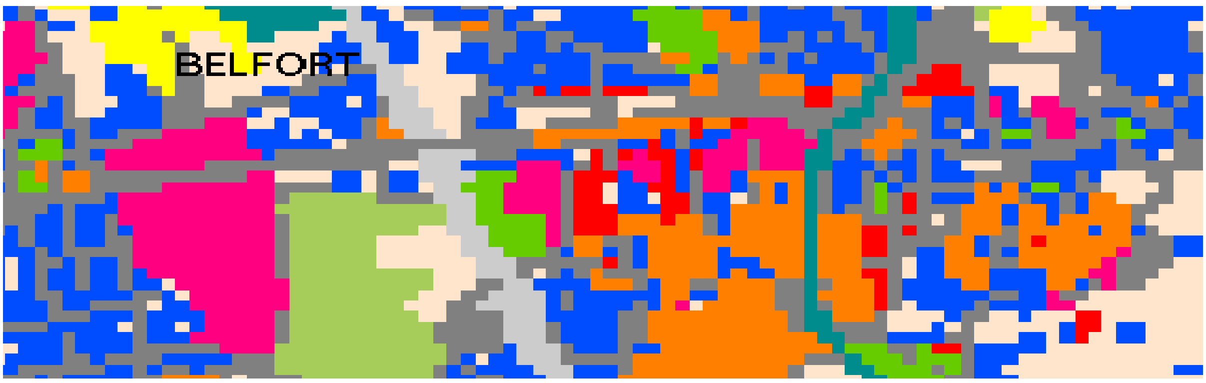


Figure 5.4. Ground classification view mode of a part of the city of Belfort where each color corresponds to a class of ground

Now that the simulation environment is presented we move on to the description of the implemented mobility models.

5.1.3 Mobility Models

The mobility models presented in this section may be classified in two major categories: random and terrain aware models.

Random models motion is based on random processes and random choices. They do not take into consideration any environment information. The second category may be described as terrain and environment aware models, i.e. terrain characteristics are taken into account for individual and population displacement.

For each of the models presented we will show two figures: individual trajectory, and population displacement. The trajectory corresponds to one individual moving along simulation area. The aim of presenting this figure is to visualize the degree of realism in individual displacement. For the population displacement, 75,877 individuals are simulated so that it is about 50% of the full population size of the area (this number corresponds to the maximum number of individuals that can be simulated without saturating the computer's memory). The aim of these figures is to visualize if the simulated population is logically distributed in a city: in buildings on roads, etc.

Visual evaluation of the simulated models will be given for trajectory and population simulations. These evaluations will be quantified in the next section with new evaluation criteria.

5.1.3.1 Random Mobility Models

5.1.3.1.1 Random Waypoint Mobility Model

The Random Waypoint Mobility Model is a totally random model. The displacement in this model is done for a fixed time t or for a fixed distance d.

Given a starting point an individual randomly chooses his destination point, direction and speed. The direction angle is uniformly distributed between 0 and 2π. Speed follows a truncated Normal distribution between 0 and *Vmax*. After choosing the direction angle and speed the individual moves to his target point with constant speed. No acceleration or deceleration is implied during the displacement. A pause time *Tpause* is chosen randomly at each destination reached. The random choice of direction angle and speed produces sudden change in displacement direction and speed which is a non realistic behavior for an individual in motion. A detailed study of this model and its drawbacks can be found in [15].

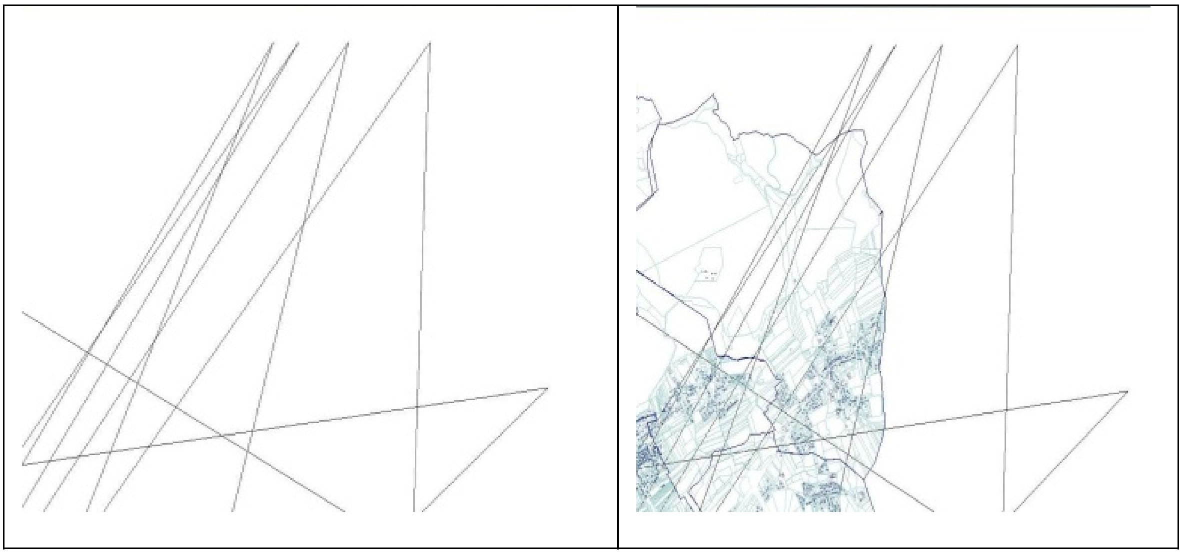


Figure 5.5. Random Waypoint Mobility Model (Trajectory)

Figure 5.5 shows a part of a trajectory simulated using the Random Waypoint Mobility Model. The left side of the figure shows the trajectory, as the right side shows the same trajectory with the environment.

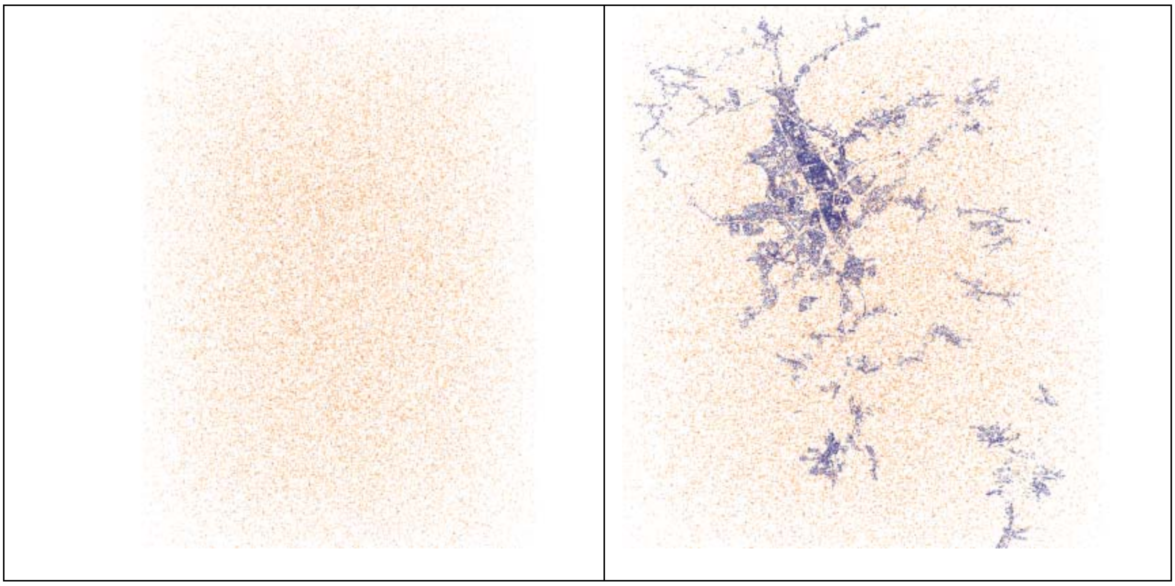


Figure 5.6. Random Waypoint Mobility Model (Population)

Figure 6 shows the Random Waypoint population simulation. The right hand side of the figure shows that the population distribution is not concentrated on the building and structures in the simulation area, and is quite uniformly distributed on all points of the simulation area.

5.1.3.1.2 Random Walk Mobility Model

The Random Walk Mobility model is a special case of the Random Waypoint Mobility Model. While for Random Waypoint a destination point is chosen, for the Random walk an individual chooses the direction angle, defining the direction he is going to take, as well as the duration and speed of the displacement. The other difference is that for the Random Walk there is no pause time.

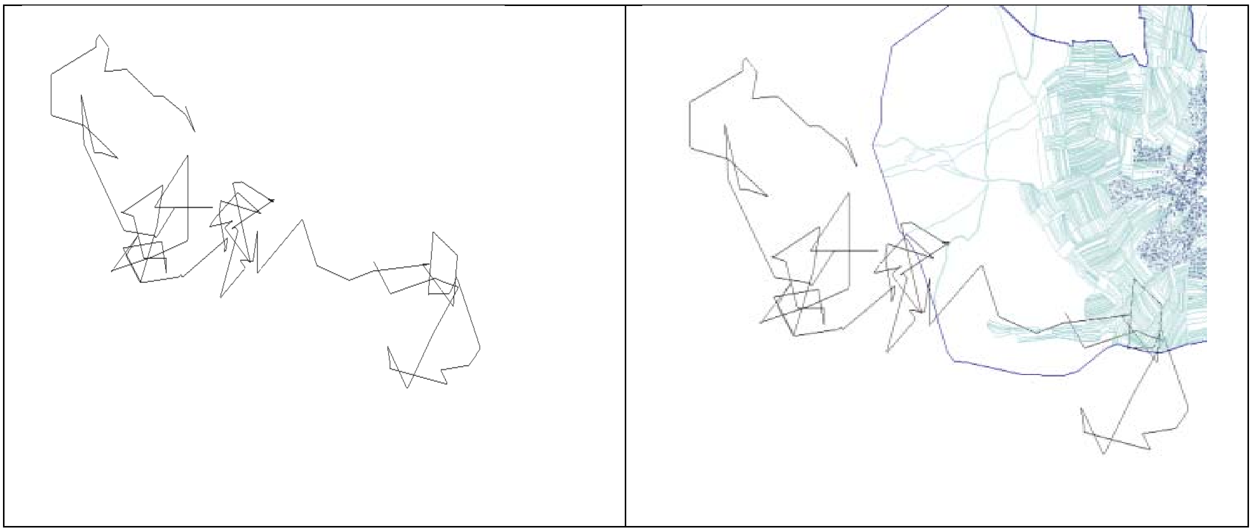


Figure 5.7. Random Walk Mobility Model (Trajectory)

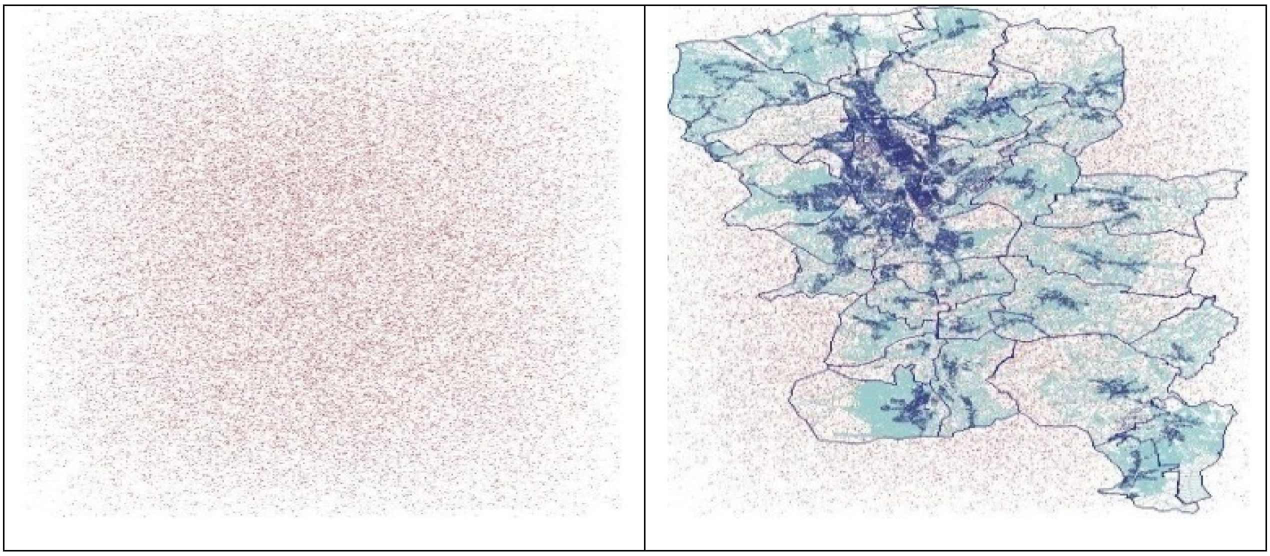


Figure 5.8. Random Walk Mobility Model (Population)

As shown in Figures 5.7 and 5.8 the same visual conclusions as for the Random Waypoint model can be drawn. These plots show the random distribution of the population.

5.1.3.1.3 Smooth Mobility Model

Figure 5.9. Smooth Mobility Model (Trajectory)

This model has two major features in addition to other random models, the first being the acceleration and the second is that the choice of velocity at each step depends on the previous step velocity chosen value. This leads to incremental speed change.

From a given starting point a destination point is chosen anywhere on the studied map. This destination point is characterized by the drift angle which is chosen in an interval from 0 to 2π. Speed is chosen at the beginning of the simulation, at each step it is either incremented or decremented by the acceleration value.

The major advantage of this model is the elimination of all sharp turns for individual trajectory simulation. As shown on Figure 5.9 we can see that the trajectory is smooth and there is no sharp turns.

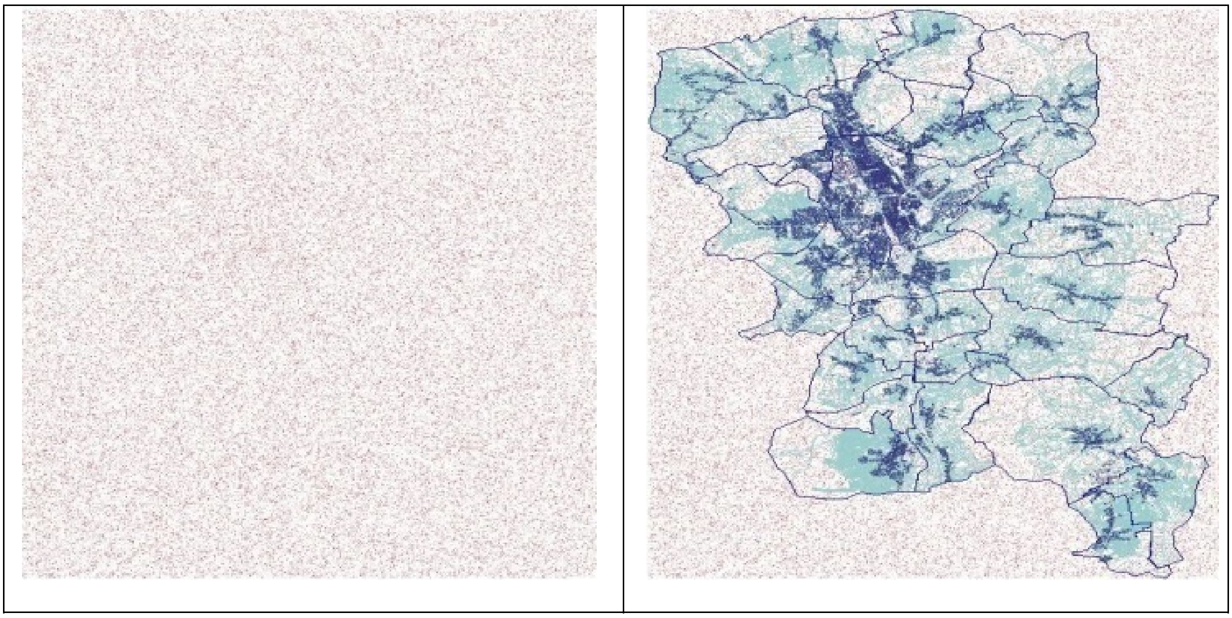


Figure 5.10. Smooth Mobility Model (Population)

For the population displacement simulation shown in Figure 5.10 the same conclusions as the previous models can be driven: lack of realism. These models cannot be used to simulate real system performance.

5.1.3.2 Terrain Aware Mobility Models

5.1.3.2.1 Normal Walk Mobility Model

The main idea of this model is that an individual in motion has a larger probability in moving forward than turning back. Direction is defined by a drift angle normally distributed with zero mean and standard deviation between [5°, 90°]. This model is based on the concept that the majority of the steps follow the shortest path. This is done in order to add to the movement pattern more realism.

Starting from a grid map with square or hexagonal cells four or six directions are defined in this model respectively; in our case we have implemented the square grid cell version, which suits better our environment. Drift angle define the displacement direction. The cell sides define the intervals of direction angle, six interval for hexagonal cells and four for square ones.

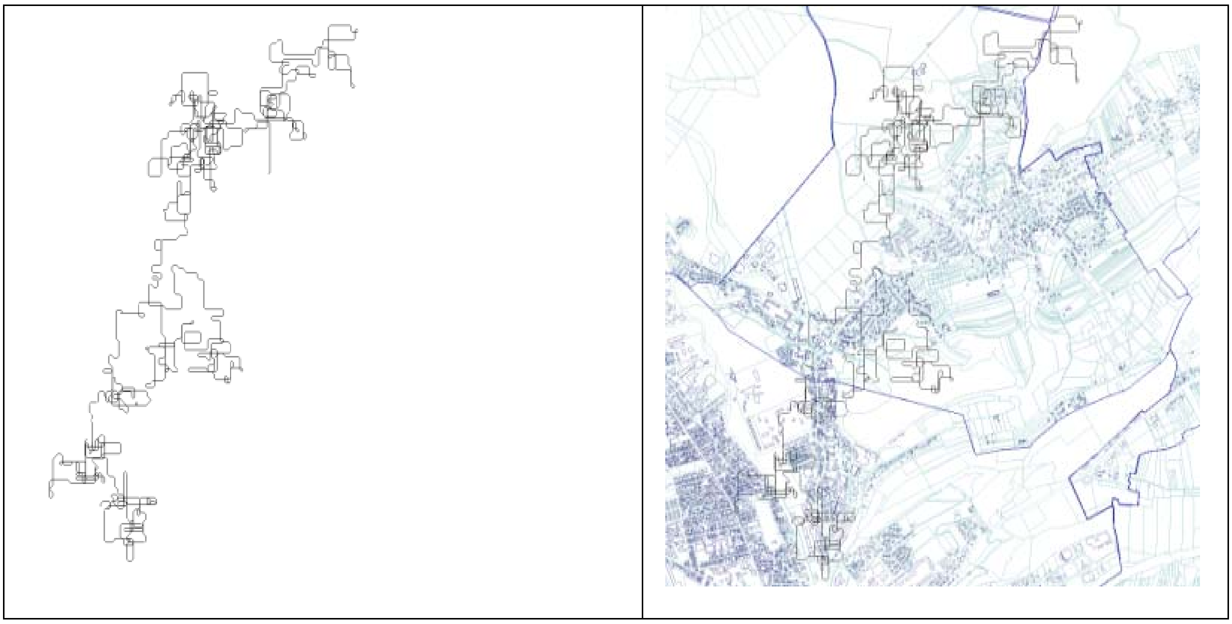


Figure 5.11. Normal Walk Mobility Model (Trajectory)

Figure 5.11 shows that the trajectory is formed by a series of looping and curls. But we can also see that a large part of this trajectory passes through different roads and building structures of the map. In Figure 5.12 it can be clearly seen that the population distribution is concentrated in the center of the map on and around structures such as roads and buildings.

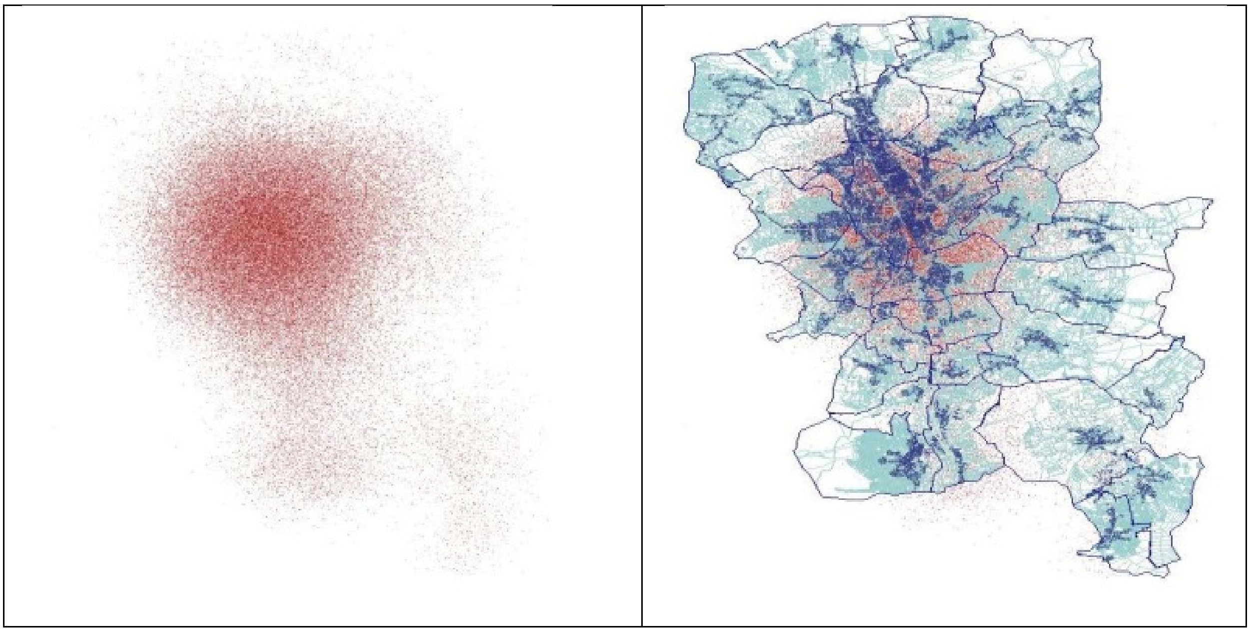


Figure 5.12. Normal Walk Mobility Model (Population)

5.1.3.2.2 Markovian Movement Model

This model uses the ground classification layer presented in Figure 5.4 for the individual and group displacement.

An individual present on a grid cell will have a certain probability for going to any of the neighboring cells or staying at the same position. This probability is function of the attraction weight ω_i of the cells involved in this choice. The probability function is given by:

$$p_i(t) = \frac{\varpi_i(t)}{\sum_{j=1}^{j=9} \varpi_j(t)} \tag{1}$$

Where $\omega_i(t)$ is the attraction weight of grid cell i at instant t.

Although some realism is added to the simulated displacement, this model does not provide logical succession of positions. This is due to the fact that each displacement choice at instant t is done independently from any previous direction choice, causing way and back motion. We have adapted the Markovian Movement Model to our simulation environment which uses square cells and not hexagonal cells. This increases the number of states for each step from six to nine. This leads to two displacement distances: one for horizontal and vertical displacement and the other for diagonal displacement.

Figure 5.13. Markovian Movement Model (Trajectory)

Figures 5.13 and 5.14 show a higher degree of realism in trajectories and population distribution than previously described models.

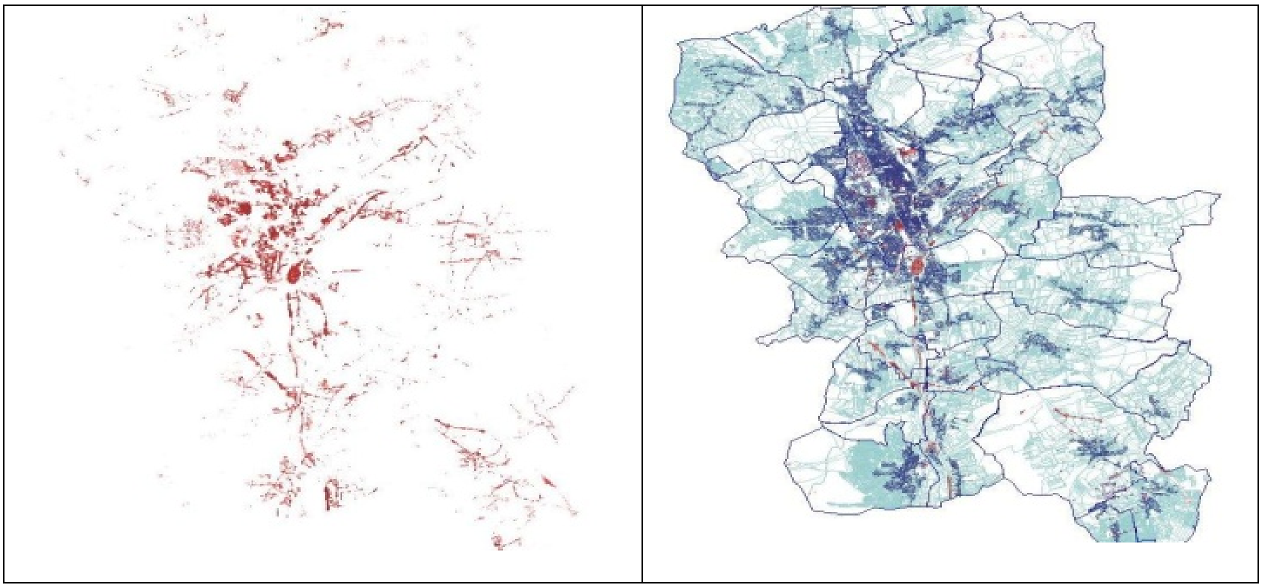


Figure 5.14. Markovian Movement Model (Population)

5.1.3.2.3 Normal Markovian Mobility Model

The Normal Markovian Mobility Model is a hybrid model that we proposed. It is created by merging the characteristics of two models: Markovian Movement Model and Normal Walk Mobility Model. The same displacement principle as the Markovian model but the possibility of moving backwards is eliminated.

From Figures 5.15 and 5.16 we can deduce that the resultant motion of this model is an improved version of the Markovian Movement Model from the fact that the trajectory is smoother and the population distribution is more concentrated.

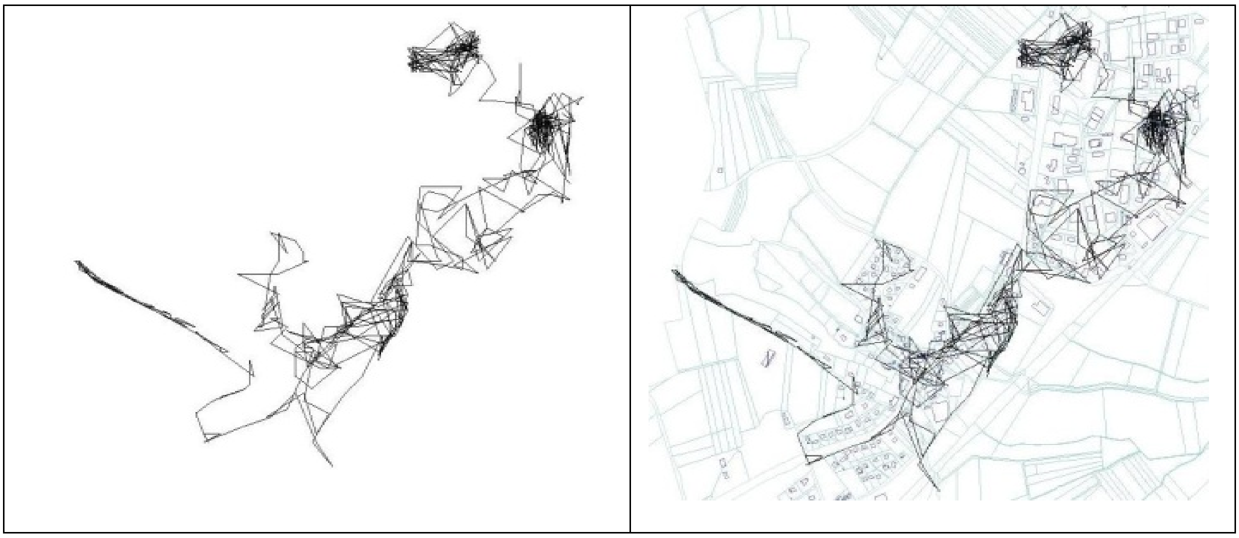


Figure 5.15. Normal Markovian Mobility Model (Trajectory)

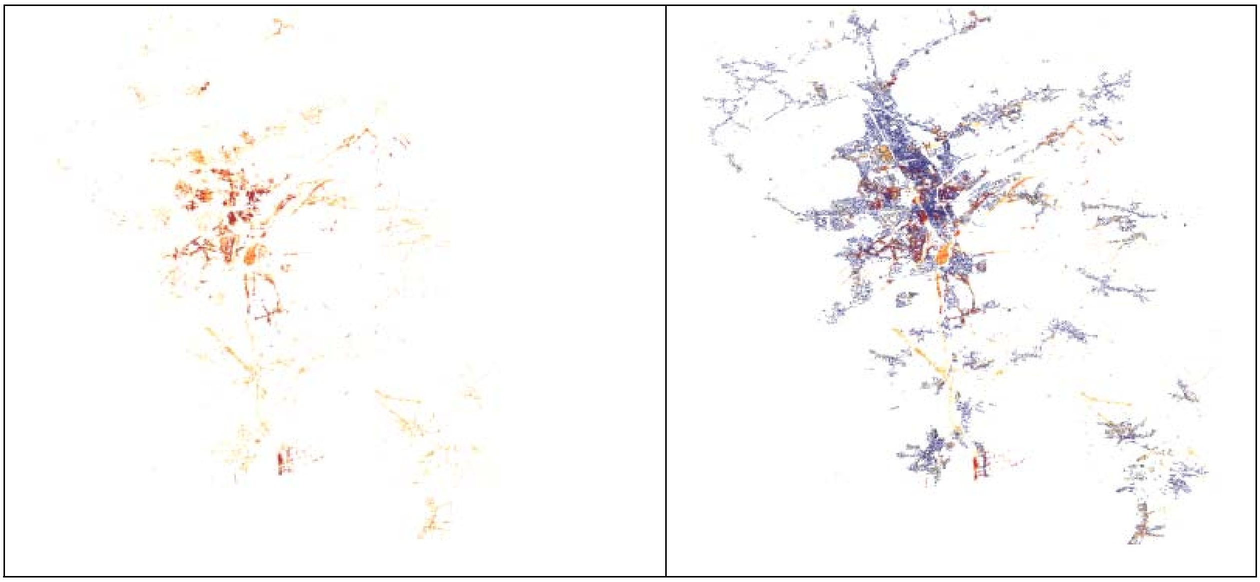


Figure 5.16. Normal Markovian Mobility Model (Population)

5.1.3.2.4 Mask Based Mobility Model

The motion in this model is driven by the grid cells attraction weights. For each individual in motion there is a displacement mask formed by his actual position and the eight adjacent grid cells. The displacement from one grid cell to another or staying in place is made according to the same equation presented for the Markovian Movement Model. For this model, when an individual starts his journey he has nine possible displacement directions. If he chooses to stay on the same grid cell, he will still have nine possibilities for the next step. But if he chooses to move in any possible direction, the possibility of going backwards is eliminated by forcing the weights in the mask to zero. The possibilities of staying in place or going straight right or left are diminished by a multiplicative factor.

Another layer of information is added which is the average speed, for each grid cell two values are stored {*pedestrian, vehicle*}. Other values can be also added. Each grid cell has an attraction weight at each instant t noted $\omega_i(t)$ and a value of the average speed for each type of individual.

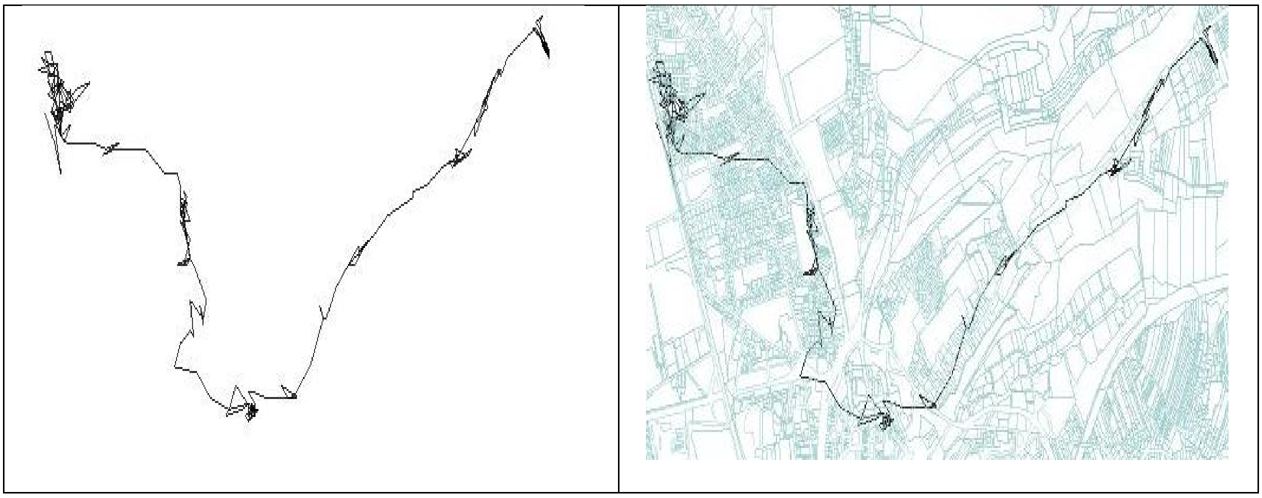


Figure 5.17. Mask Based Mobility Model (Trajectory)

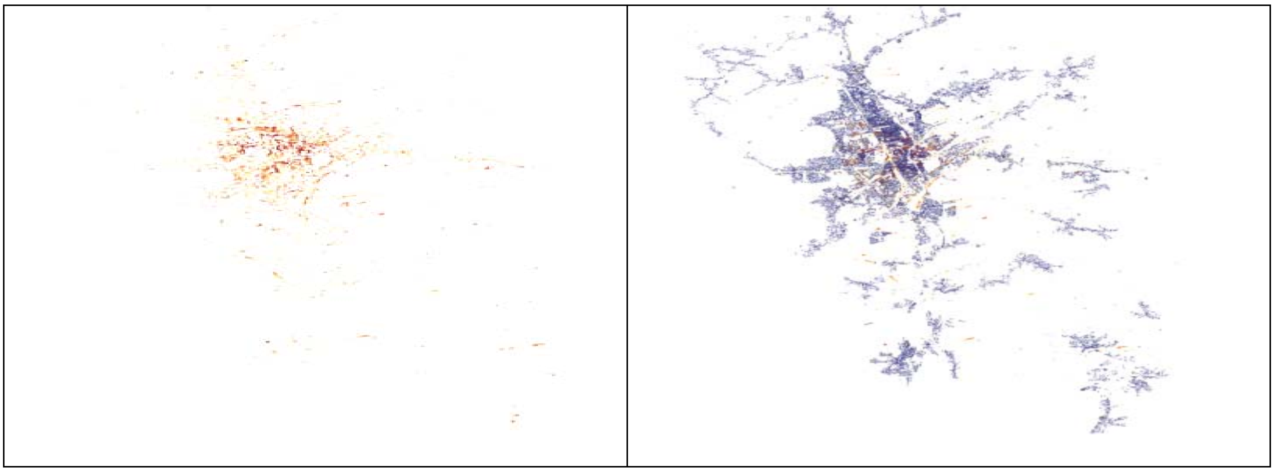


Figure 5.18. Mask Based Mobility Model (Population)

From Figures 5.17 and 5.18 we can deduce that the resultant motion pattern is better than that of the previous models, the trajectories are smoother and the population is highly concentrated on residential and buildings area on the map.

5.1.3.2.5 Synthesis

In this section we presented the mobility models and their implementations. It is visually clear that certain models have a higher degree of realism than others. The terrain aware mobility models are better than the other models in representing realistic motion, but how to differentiate and evaluate these models and their degree of realism?

We need to define metrics in order to quantify the models characteristics to get a rational comparison. Trajectory and population displacement metrics and evaluation for all the previously presented models are defined in the following section.

5.1.4 Metrics and Evaluation

In order to compare the presented mobility models we have conducted two series of tests: individual and population simulation tests. A series of metrics was developed in order to establish a full comparison of the results with numerical evaluations.

5.1.4.1 Existing Works

Several works on metrics related to mobility models exist in the literature. In general, mobility metrics are given to analyze the motion of mobiles for a given application where the metrics are linked to the application itself.

For example such application is the assessment of the quality of service that a communication network can provide to an individual in motion. The metrics can be used to analyze the influence of individuals' mobility simulated with different mobility models on the network service, like routing protocols [14], inside VANET networks [129] or ad hoc networks [130][131].

Another study proposes the use of metrics to visualize the quality of communications between nodes in ad hoc networks, using node mobility and topological dynamics [126]. An example of similar metrics would be the metrics used to validate generated mobile traces, by comparing them to real traces extracted from Wi-Fi devices moving in a given geographical area [125].

However, the analysis of the geometry of a displacement and the effects of motion changes on the information routing in a communication network [127], and the analysis of mobility parameters such as individual speed [128] or individual density in areas [132] do not evaluate the result of the mobility itself. They aim to study the impact of mobility on network performance and not to evaluate the mobility in terms of trajectory or motion pattern.

The metrics that were defined within this thesis report were proposed to evaluate the global trajectory of an individual and are not linked to a specific application. The aim is to quantify the visual aspect observed between simulated trajectories and population displacements.

In the following part, we define the metrics we propose for individual trajectory analysis and for the population analysis. In the following tests the individual trajectories and population flows are simulated using the MBMM model and other mobility models from the literature, and the results are compared to real traces.

5.1.4.2 Definitions

Before presenting the metrics we developed, we will define some of the terms used in individual and group evaluation.

- Trajectory (T): succession of displacement positions for a given individual during the whole simulation day.
- Position (P): P_i position at instant i.
- Curl (C): a concentration if a number of displacement positions of a trajectory in a zone form a curling or a yarn like shape.
- dist(P_i,P_j): Euclidean distance between positions P_i and P_j.
- time(P_i,P_j): time that the individual takes to go from P_i to P_j.
- L(P_i,P_j): is a part of the trajectory T. It is the path length of the trajectory between P_i and P_j.

5.1.4.3 Individual Metrics

The aim of individual metrics is to evaluate each and every trajectory T of any individual generated with any of the previously presented mobility models. The metrics are generic and may be applied to any other mobility models.

A good trajectory is a path that the individual takes to reach one or several destinations. We consider that, during the day, an individual has a limited number of destinations to reach. So a trajectory is composed of these destinations and the links between them. This can be seen in the previous trajectory simulation figures as the almost straight paths between the curls, where each curl marks a destination.

5.1.4.4 Detecting Curls

The main element in defining metrics for trajectory evaluation is the curl detection. For this we need to numerically define the curl description. Qualitatively a curl is defined as the set of positions between position P_i and position P_j such as: $time(P_i, P_j) \geq M$ where M is a given number of minutes.

We chose $M = 30$ minutes for our simulations.

$\forall i \leq k \leq j;\ \sqrt{(x_k - \bar{x})^2 + (y_k - \bar{y})^2} \leq L$ where L is a distance in meters. We chose $L = 200$ m, for our simulations. And $\bar{x}$ and $\bar{y}$ are the average coordinates values of all points between P_i and P_j. With this condition we impose a restricted space to identify a curl.

If other indexes i' and j' such that $i' \geq i$ and $j' \leq j$, and condition (2.) is also met for $P_{i'}$ and $P_{j'}$, then $i' = i$ and $j' = j$. This condition maximizes the size of the detected curl.

After the detailed presentation of the curl detection mechanism we will present the resulting simple and combined metrics and the possible evaluations.

5.1.4.5 Number of Curls

The aim of this metric is to compute the number of detected curls in each trajectory.

$$M_1 = number\ of\ curls$$

The results of this metric evaluation are shown in Figure 5.19. We can clearly see that the terrain aware models allow the generation of spatial-temporal stationary states in opposition to random models for which a very small number of curls is generated.

The following metrics will further evaluate the trajectories and the curls to check if the detected curls are real logical destinations.

5.1.4.6 Curl Proportion

The aim is to calculate the proportion of curl in a given trajectory, by dividing the curl time C_time by the trajectory time T_time. There is no rule about the expected value but inside a normal trajectory a significant curl proportion must be found, so relatively high values of M_2 are expected.

$$M_2 = \frac{C\_time}{T\_time} \tag{2}$$

The results of this metric evaluation are shown in Figure 5.20. We see that the models which are not terrain aware cannot generate curls in the trajectories which is not an expected results of displacement simulation.

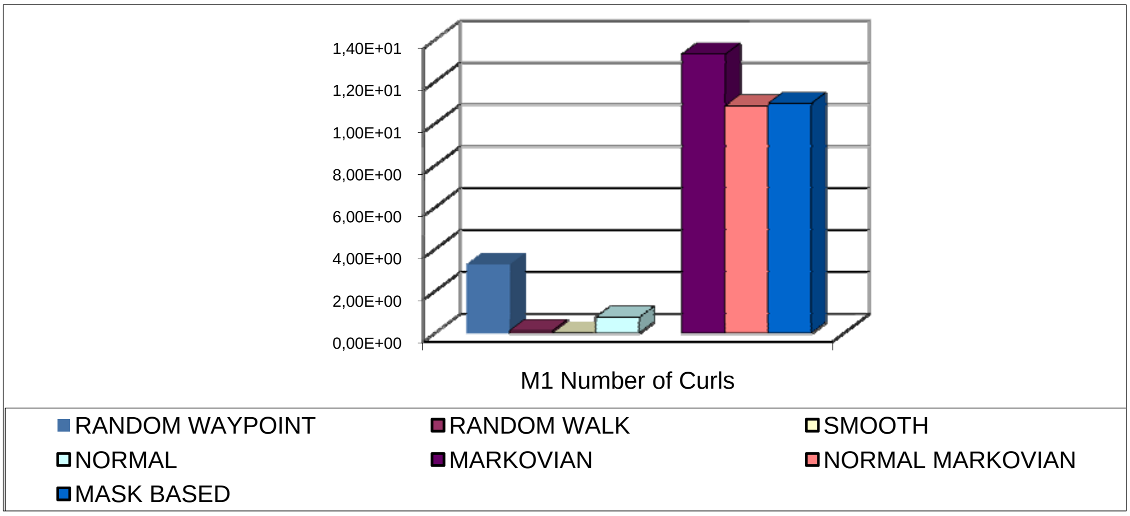


Figure 5.19. Number of curls on trajectory for random and for terrain aware mobility models

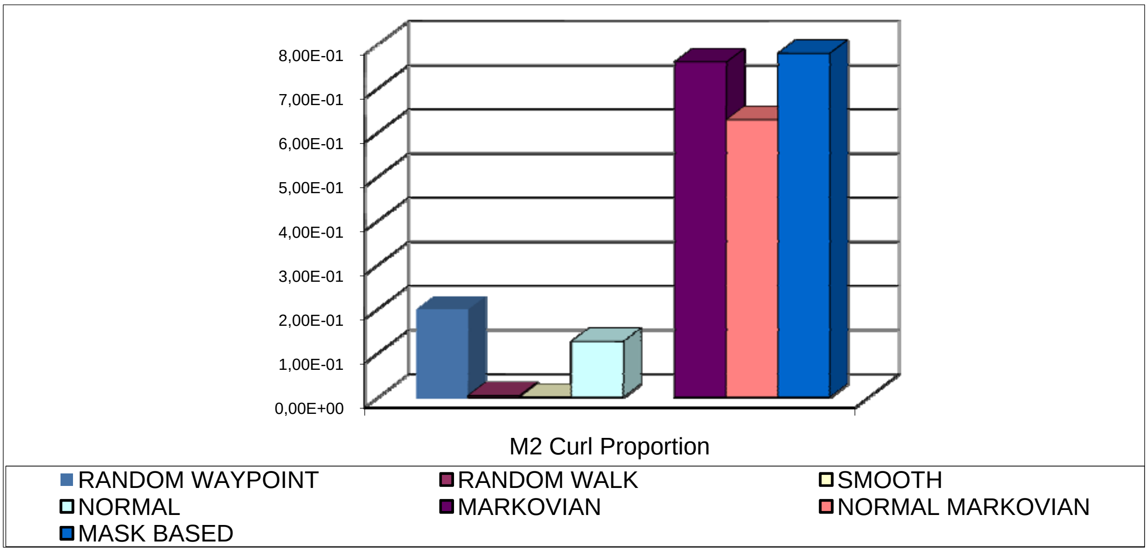


Figure 5.20. Curl proportion in trajectories for random and for terrain aware mobility models

5.1.4.7 Non-Curl Proportion

Oppositely we may calculate the proportion of non-curl in a given trajectory. The smaller the non-curl proportion is (small M_3 values), the better is the trajectory because in that case the individual in motion converges rapidly from his previous destination to his next destination.

$$M_3 = 1 - M_2 \tag{3}$$

Results are shown in Figure 5.21. We can see that the terrain aware models, especially the Mask Based Mobility Model, give better performances than random models.

5.1.4.8 On Curl Dispersion

The aim is to compute the average spatial dispersion of individual positions around the gravity center of a curl. The less dispersed a curl is, the better it is (small values for M_4). Where n is the number of steps and x and y are the positions coordinates and ¯,¯ are the coordinates of the curl center.

$$M_4 = \frac{1}{n}\sum_{i=1}^{n}((x-\bar{x})^2 + (y-\bar{y})^2) \tag{4}$$

Results are shown in Figure 5.22. Only the results of the terrain aware models are shown because the difference with the random models is too high. It can be seen that results are close, with better performance for the Mask Based Mobility Model.

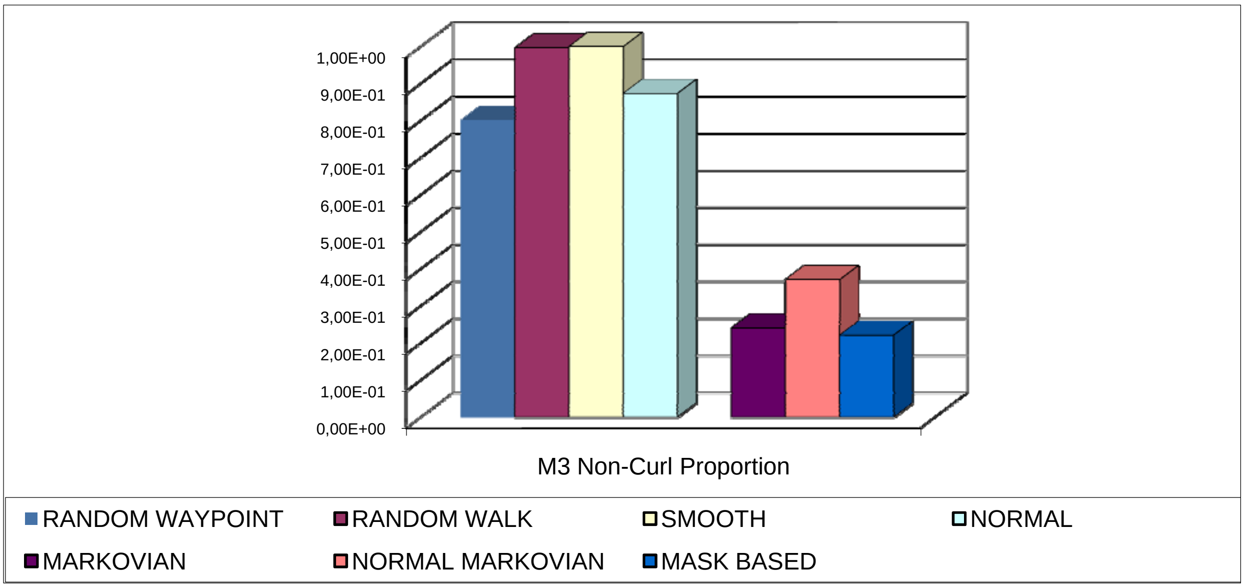


Figure 5.21. Non-curl Proportion in trajectories for random and for terrain aware mobility models

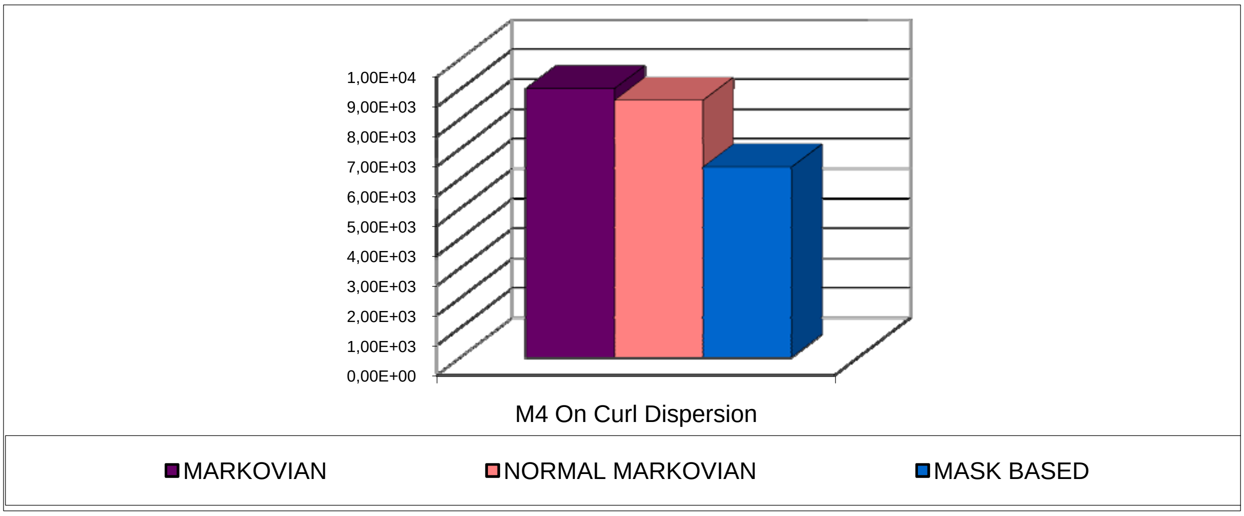


Figure 5.22. On curl dispersion in trajectories for terrain aware mobility models

5.1.4.9 Curl Activity Rate

This metric is set to compute the rate of presence of one *activity zone* on a curl. An activity zone is a structure (building, road…) or a set of structures that represent possible destinations of individuals for a given period. These structures are characterized by their numerical attraction weight for a given period making them a logically most probable destination for the individuals in motion if the attraction weight is high and a place to avoid if the weight is low. In that case, higher values of M_5 are preferred.

$$M_5 = \frac{activity\_length}{L(P_i, P_j)} \quad (5)$$

$$activity\_length = \sum_{k=i}^{j-1} dist(P_K, P_{k+1}) * \delta(activity, P_k) \quad (6)$$

$$\text{Where} \begin{cases} \delta(activity, P_k) = 1 & if\ P_k \in activity \\ \delta(activity, P_k) = 0 & elsewhere \end{cases}$$

Where P_i and P_j are defining one curl and δ is the Dirac function. The set called *activity* is the set of cells corresponding to a selected activity zone for which we like to assess the curl: is the curl within an activity zone or not?

The results of this metric shown on Figure 5.23 show that the curls in trajectories simulated with the Mask Based Mobility Model are slightly better than those of the trajectories simulated with the Normal Markovian and the Markovian mobility models.

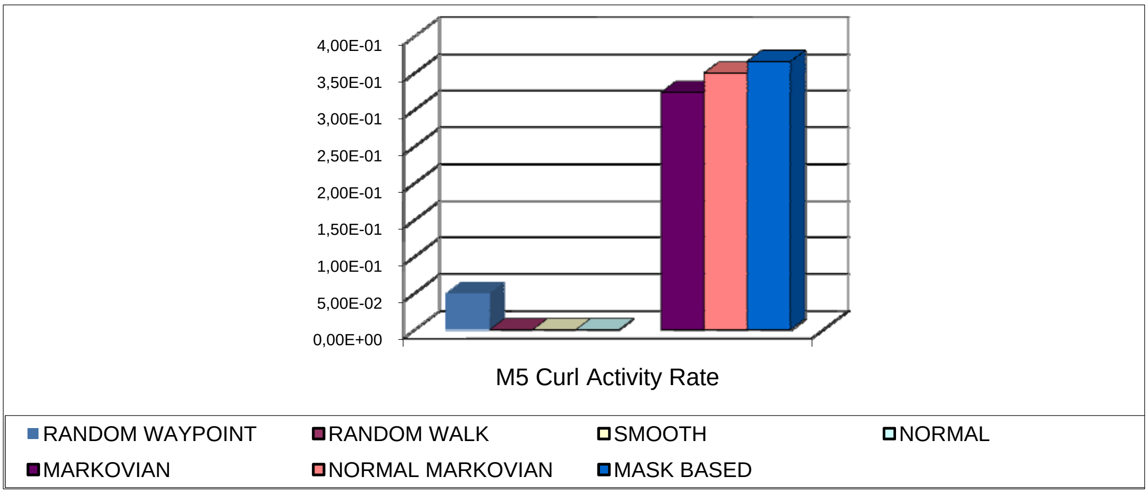


Figure 23. Curl activity rate in trajectories for random and for terrain aware mobility models

5.1.4.10 Detour Rate

This equation computes the ratio between the simulated path length and the Euclidian distance between two curls. In that case P_i is the ending point of the first curl and P_j is the starting point of the next one. The smaller the value of M_6 the faster an individual goes from a curl to another.

$$M_6 = \frac{L(P_i, P_j)}{dist(P_i, P_j)} \tag{7}$$

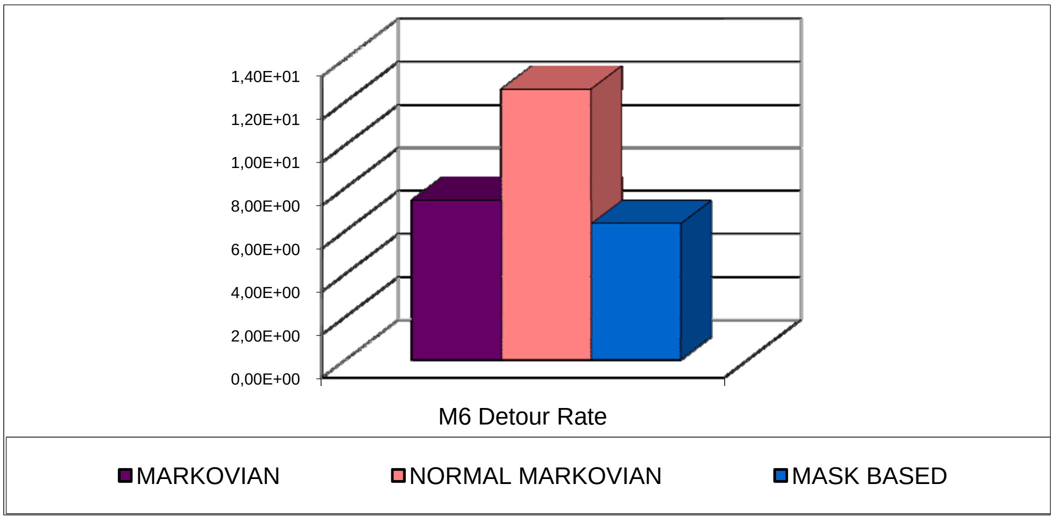


Figure 5.24. Detour rate in trajectories for terrain aware mobility models

5.1.4.11 Incoherence Rate

M_7 calculates the number of times the individual in motion chooses a different direction than that of going straight from one curl to another, over the path length between two curls. In that case P_i is the ending point of the first curl and P_j is the starting point of the next one. M_7 must be small.

$$M_7 = \frac{nb\ of\ changing\ directions}{L(P_i, P_j)} \tag{8}$$

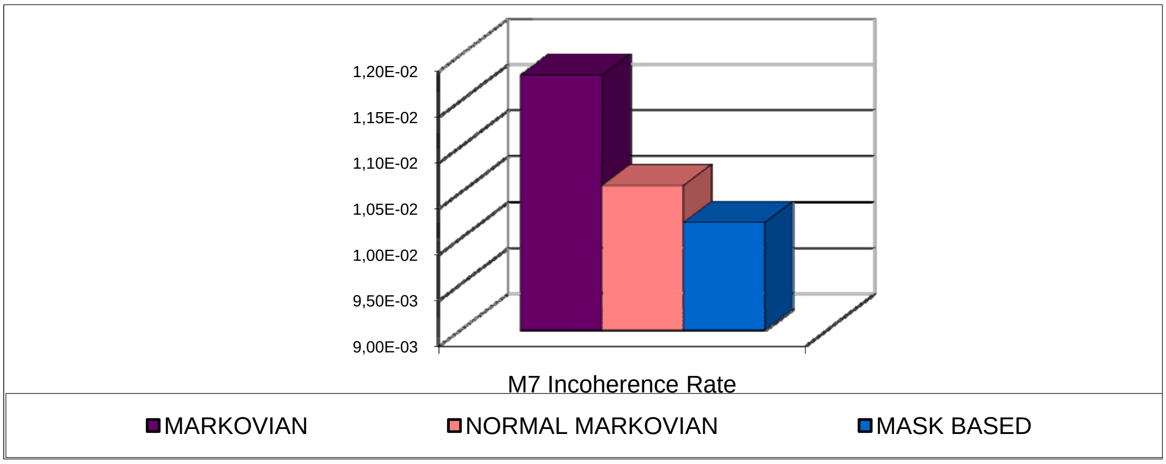


Figure 5.25. Incoherence rate in trajectories for terrain aware mobility models

From Figure 5.25 we can see that the link between two curls, simulated with the Mask based Mobility Models present less dispersion and this is why it converges faster when going from a curl to another.

5.1.4.12 Non-Curl Rate

Between two curls, M_8 computes the ratio between the simulated non-curl distance and the shortest real distance between these curls measured by the real distance between two points P_i and P_j. This gives the rate that one individual crosses through a grid cell of the shortest path between two curls. In that case P_i is the ending point of the first curl and P_j is the starting point of the next one. In that case, higher values of M_8 are preferred.

$$M_8 = \frac{path\_length}{L(P_i, P_j)} \tag{9}$$

$$path\_length = \sum_{k=i}^{j-1} dist(P_K, P_{k+1}) * \delta(path, P_k) \quad (10)$$

$$Where \begin{cases} \delta(path, P_k) = 1 & if\ P_k \in path \\ \delta(path, P_k) = 0 & elsewhere \end{cases}$$

Where, *path* is the set of ordered cells between two curls defining the real geographical path between them. P_k belongs to *path* means that the cell *k* used by the individual is within the real path.

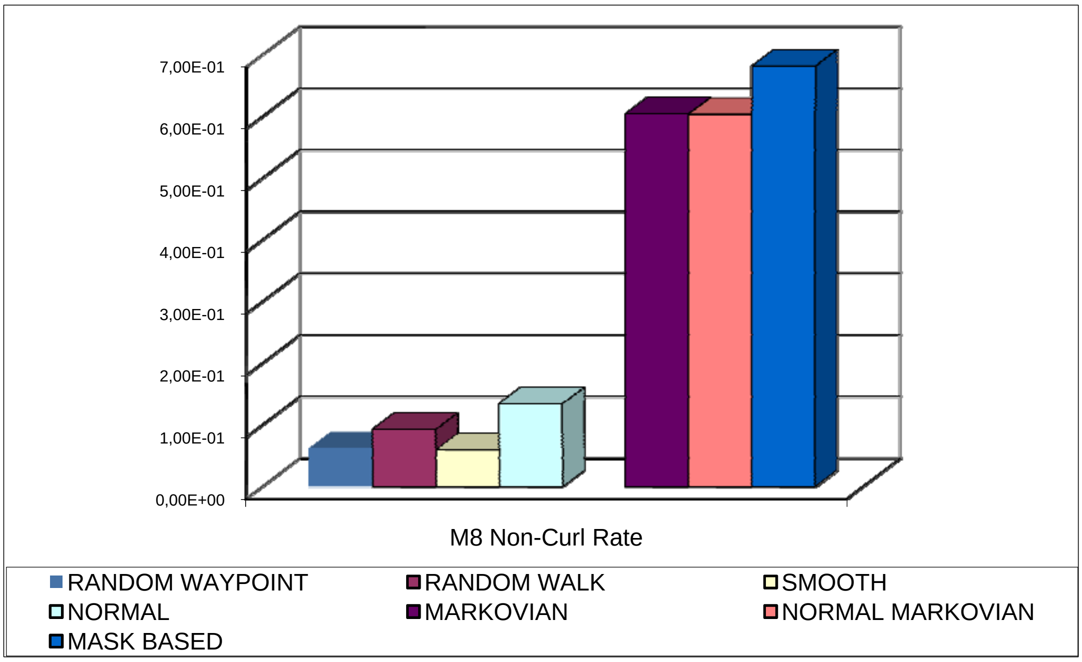


Figure 5.26. Non-curl rate in trajectories for random and for terrain aware mobility models

From Figure 5.26 we can see that individual simulated with random mobility models do not follow a logical track which is normal as the direction is chosen randomly. For terrain aware models we can see that the path followed by the individual from a curl to another is close to the real geo-referenced path a real individual would follow between these two areas. It can be noted that the best non-curl rate goes to the Mask Based mobility model trajectories.

5.1.4.13 Global Trajectory Evaluation

The previous metrics may be used alone to assess the trajectory on several criteria. In this section we propose to compute a global evaluation of trajectory.

An optimal trajectory is one that:

— Lets M_1 between 4 (minimum number of segments in one trajectory, that is go to work, go to lunch, go to work and go home) and 15 (maximum number of structure inside the environment description we define) for one day of normal activity (this interval is set through observations of real trajectories but it can be tuned for any application)

— Maximizes M_2: means that the individual while moving is passing by attractive zones and having several destinations.

— Minimizes M_3: diminish the number of steps between two consecutive destinations, which means a faster convergence.

— Minimizes M_4 on each curl: the more the curls are concentrated the better they are. This means that the individual in motion is within the area of the attraction pole and not going out and then in of this attraction pole all the time which leads to a higher number of steps (or time) from the total trajectory.

— Maximizes M_5 on each curl: this means that every detected curl is a valid possible destination.

— Minimizes M_6 means that the individual converges rapidly from a curl to another.

— Minimizes M_7 means that the individual goes straight forward to the destination with a minimum deviation.

— Maximizes M_8 on each path between two curls: this means that the individuals in motion are taking the shortest paths between two consecutive curls or destinations.

These conditions are combined to form the global trajectory evaluation metric MG. We added 1 to all the metrics because for some models there are metrics with zero value. This global metric must be maximized (the values to maximize are on the top and the others on the bottom of the global metric). Note that M_3 is not used as it is included in M_2.

$$M_G = \frac{(1+M_2)*(1+M_5)*(1+M_8)}{(1+M_4)*(1+M_6)*(1+M_7)} \tag{11}$$

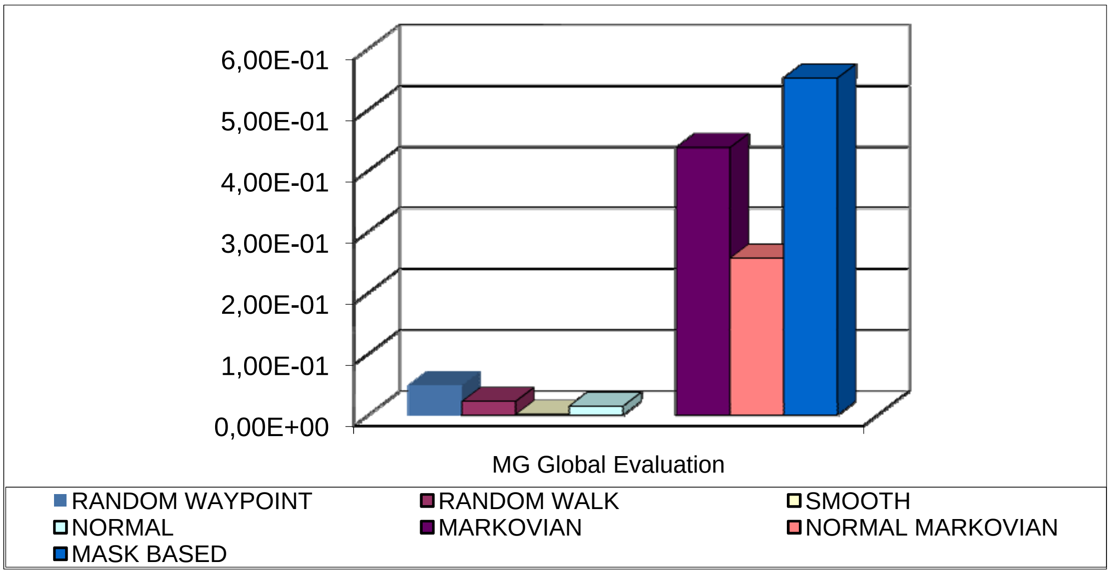


Figure 5.27. Global evaluation of trajectories for random and for terrain aware mobility models

In Figure 5.27 we show a global evaluation of the simulated trajectories using all the previously presented models. It is clear that the terrain aware models have a better evaluation; this is the result we were expecting from these models. And between the terrains aware models the MBMM gives the best results for trajectory simulation.

5.1.4.14 Population Metrics

The aim of the population metrics is to evaluate each and every cloud or mass concentration generated with any of the previously presented mobility models.

For each period convergence at time *t*, the positions of the generated population are compared with real traffic locations extracted from real survey data. The target value of *M(t)* is: *M(t)=0, ∀t.*

$$M(t) = \frac{1}{total\_ind(t)} \sum_{i=1}^{nb_{cells}} |nb\_indiv\_survey(t,i) - nb\_indiv\_model(t,i)| \tag{12}$$

In the formula, *nb_indiv_survey(t,i)* is the number of individuals observed by surveys on a specific position *i* and instant or period *t*, it is the target value. Whereas *nb_indiv_model(t,i)* is the simulation resulting number of individuals present on position *i* at instant *t*. This process of subtracting these two entities is done for all the cells in the simulation area. And the result is divided by the total number of individuals in the simulation zone at *t*.

Results presented in Figure 5.28 show the sum of *M(t)* for all *t* for the observation period. The final positions of the population simulation along the whole day are better for terrain aware models, with better performance for positions generated using the Normal Markovian Mobility Model. This was clear in the population simulation figures previously shown.

The MBMM is in third position but not far from the other terrain aware models. This is due to the fact that this model gives more liberty in motion which leads individuals to move freely not like the Markovian where individual can be absorbed by grid cells with high weight levels and the Normal Markovian where the displacement choice is limited and not dynamic as in the MBMM.

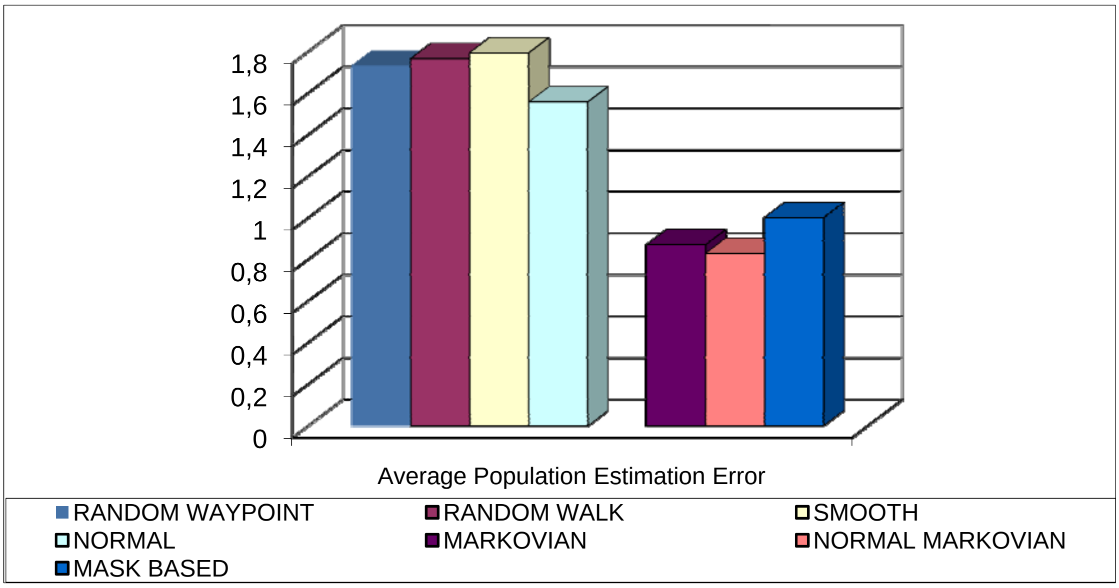


Figure 5.28. Average population estimation error for random and for terrain aware mobility models

5.1.5 Synthesis

The metrics in this study were created to evaluate and compare mobility models independently from their expected applications. The aim was to evaluate all the mobility features that each of the models gives.

Several mobility models were presented. These mobility models may be separated into two major categories: random models and terrain aware models. Within the first category we looked at the Random Waypoint mobility model, the Random Walk mobility model and the Smooth mobility model. The terrain aware mobility models used in this work were: the Normal Walk mobility model, the Markovian Movement Model, the Normal Markovian mobility model (which is a combination of the two previous models) and the Mask Based Mobility Model. We conceived both last models.

The evaluation we did has been possible with the use of the metrics we proposed in this work. These metrics are separated into two major categories: trajectory metrics and population metrics.

Trajectories are split into two main parts, curls and lines. The curls are constituted by the steps of an individual within a small area in which he remains for a certain number of steps. And the lines are formed when the individual decides to go straight through an area without generating a curl. The analysis begins with detecting these elements by their average number on the same terrain for each individual and by the number of steps they are formed with. Then the dispersion of the curls and the straightness of the trajectory lines are respectively computed by evaluating the distance between starting and ending curl points and the detour rate that an individual can perform while going in a certain direction. Then a global evaluation of the trajectory is made taking into consideration all previously defined metrics.

For the population metric the analysis is based on the comparison between the computed positions of individuals during time in the simulation, and real positioning information gathered and collected from surveys and other analysis.

The results came as expected: terrain aware models gave the best performance compared to random models. We proved it numerically by our metric computation. Between the terrain based models, the Mask Based Mobility Model, which is the model we developed in this thesis work, gave very good results on the global evaluation of trajectories with a very minimal error value on the full population metric on the test instance. These results should be confirmed by using another test instance on which the population displacement is available.

5.2 Mobility simulation for Evaluation of UMTS Power Control Algorithms

5.2.1 Introduction

Cellular networks are described as a deployment of antennas covering a given territory. These antennas are assimilated to Base Stations (BS). The coverage area of each BS is denoted by cell. In order to establish all the communications and provide a given quality of service for the users, several control functions are added to the system. One of these functions, the main subject of this paper, is the emitted Power Control of equipment [49].

The aim of power control is to immunize the cellular system against jamming. Power control is one of the most important existing features in CDMA systems and in particular UMTS standard. In the UMTS, the power control procedure is the solution to the near far problem, occurring when all received signal sequences are not equal within a margin of approximately 1dB [49]. The solution is that all mobile equipments (mobile users and base stations) vary their power in steps of 0.5, 1, 1.5 or 2 dB depending on uplink or downlink power control. This power reduction or increase, results in less channel interference, increasing also spectral efficiency and mobile equipments battery operating time [49].

In this paper we focus on uplink power control which is of major importance to allow mobile to base station connection. We will compare different power control algorithms from literature with the UMTS power control algorithm by testing them in mobility scenarios under urban and road conditions of motion. The mobility will be defined through the Mask Based Mobility Model (MBMM) that allows our simulation to control users speed and motion inside different environment: road, city, commercial areas, etc. Then efficiency of power control procedures in front of real motion conditions will be emphasized, that is the aim of this application based on MBMM work.

The paper is organized as follows. The second section presents a general overview of different classes of power control algorithms. Then a presentation of the algorithms used for the simulation is given. In the third section the choice of the mobility model is justified. Then the forth and the fifth parts present respectively the simulation scenarios and the comparison of results between the different algorithms in two mobility environment. Finally a general conclusion is given to summarize all the important features of this work.

5.2.2 Power Control

Three major classes of power control algorithms for mobile networks exist:

- Centralized [50][51][52]: For this type of algorithms, the base stations report to a central controller the information concerning the radio link. After computation, the controller spreads the power control signal decisions throughout the network. This type of algorithms provides results with optimal quality, but at the cost of an important time delay due to control signal exchange and computational complexity.

- Cooperative [53][54][55][56]: In Cooperative power control algorithms neighboring cells exchange limited control data. According to the information received by each station an appropriate power level is determined.

- Distributed [57][58][59][60][61]: For this class, in each cell the base station and the mobile stations control their transmission power. This is done with a limited knowledge of the radio link gain between mobile and base station. This procedure is the most efficient in terms of signal exchange. This type of algorithms delivers a high-speed computation which permits a real time evaluation of the power control in a network.

In our study we focused on the distributed power control algorithms to make a comparison of several approaches with UMTS power control system in order to analyze its performance in high speed and low speed mobility contexts. We selected four major distributed power control algorithms for this work: the UMTS procedure [73], the LI-SRA procedure (Limited Information Stepwise Removal Algorithm) [62][63][64][65][66], the AFM procedure (Augmented FDPC Method) [62][67] and the ACP procedure (Augmented CIPC Method) [62]. In the following, a presentation of the selected models is given. Let introduce the following notation for that work:

— $P_i(t)$ is the power transmitted by mobile i at time t (the transmitter);

— $\Gamma_i(t)$ is the signal to interference ratio measured at time t on uplink i at the base station (the receiver);

— and Γ_0 is the signal to interference ratio threshold value.

In order to implement these distributed algorithms we need to compute the signal to interference ratio at the considered receiver.

In the UMTS, the following algorithm gives the power control procedure where x is the power step (1 or 2 dB) after 1 or 5 decisional uplink Time-Slots to increase or decrease the power [73]:

$$\textbf{If } (\Gamma_i(t)/\Gamma_0) > 1 \textbf{ Then } P_i(t+1) = P_i(t) - x \tag{13}$$

$$\textbf{Else If } (\Gamma_i(t)/\Gamma_0) < 1 \textbf{ Then } P_i(t+1) = P_i(t) + x \tag{14}$$

$$\textbf{Else Then } P_i(t+1) = P_i(t) \tag{15}$$

This formula implies that when the measured signal to interference ratio $\Gamma_i(t)$ is higher than the threshold value Γ_0 then the emitted power for next step $P_i(t+1)$ is computed by decrementing a step value x from the actual emitted power value $P_i(t)$. Now when $\Gamma_i(t)$ is lower than the threshold value Γ_0 then $P_i(t+1)$ is computed by incrementing $P_i(t)$ by x. Otherwise $P_i(t+1)$ will be equal to $P_i(t)$.

In the three other algorithms, the power control is done at time *t* (TS time) by the computation of new power with the following equations:

$$\text{LI-SRA: } P_i(t+1) = P_i(t)\ [1+1/\Gamma_i(t)] \qquad (16)$$

$$\text{AFM: } P_i(t+1) = P_i(t)\ \Gamma_0/\Gamma_i(t) \qquad (17)$$

$$\text{ACP: } P_i(t+1) = (1/2)P_i(t)\ [1+\Gamma_0/\Gamma_i(t)] \qquad (18)$$

In AFM and ACP the values of $P_i(t+1)$ is increased or decreased proportionally with respect to the ratio between the threshold signal to interference ratio Γ_0 and the measured value of signal to interference ratio $\Gamma_i(t)$. The proportion is direct and the highest with AFM. With ACP the proportion is lowest (divided by 2) and is applied indirectly as an offset to the previous power divided by 2. For LI-SRA the power control only depends on the measured signal to interference ratio but not on the target value, it is supposing to bring sudden changing and then to be convenient for high speed mobile. AFM and ACP evolve directly in function of the ratio between the SINR and the threshold, while LISRA is a gradient-like algorithm.

After presenting the power control algorithms, a brief remind of the mobility model to deliver the scenarios is given below.

5.2.3 Mobility Model

The model we used in our simulations is the Mask Based Mobility Model (MBMM). Since the aim of the paper is to analyze the impact of mobility on power control a brief description of the used mobility model is given.

MBMM is a time variant model with spatial and geographical constraints. The motion in this model is based on a nine states Markov chain. The simulation area is divided into square cells of equal size and the mobile in motion moves from one cell center to another. Two displacement distances exist: d_1, for horizontal and vertical displacement, and d_2 for diagonal displacement, where $d_1 = s$ and $d_2 = \sqrt{2}\, s$ and s is the given square side. Also in MBMM, the simulation area is divided into weighted cells. For each individual located in a cell of the concerned area there are nine possible displacement directions. So for each step a Markov chain of nine states is defined. The state occupation probability of each cell is calculated by dividing each cell weight by the sum of the nine cell weights. In this application, the twenty four hours period used for motion simulation is divided into four periods, where each one corresponds to a general given behavior of the population. For each one of these periods the weight of every grid cell is chosen based on the geographical and socio-economical aspects of the structure covered by the cell.

A dynamic mask formed by the current cell in which the individual is standing along with its eight surrounding neighboring cells constitutes a correction of the Markov chain eliminating unrealistic behavior for the individual in displacement. Depending on the previously chosen direction the Markov Chain is reconfigured. For example if the individual chooses to move forward in the previous step, at the current one the cells representing the direct backward, the left and right backward directions are set to zero in order to eliminate go backs. To set the problem of sharp angles displacement directions, the probabilities of going right or left present in the cells corresponding to these directions are divided by four. For the cell in which the individual is standing corresponding to the stay in place action the weight is divided by two. Then, the new mask has a maximum of six cells which are not null. That reduces the number of steps in the displacement Markov Chain from nine to six. But if the individual chooses to stay in place all corrections are reactivated which gives nine states to the Markov chain. After the theoretical presentation of the motion we describe the simulation scenarios. For full MBMM description, please refer to Chapter 4.

5.2.4 Simulation Scenarios

Simulation area consists of a three layers matrix. The first layer contains the attraction weights proper to the mobility model. In the second layer, we have the received power computed using the COST-231 propagation model [72][73]. At each grid cell we get a value corresponding to the maximum value among all received power signals from all antennas covering this grid cell. This model (COST-231) is quite confident as it has been calibrated on real measurements. It is often used for the computation of simulation data sets. The third layer is a dynamic layer representing the power difference from one step to another.

Let us explain these layers in details using Figure 5.29. The first layer labeled Layer 1 shows a sample of the simulation area in which we can observe twenty-five grid cells filled with their proper attraction weights. Colored grid cells show a part of a simulated possible trajectory. The position i in grey is the position of the mobile at instant t. Layer 2 shows the same grid cells but with their appropriate maximum received power value computed from COST model. The values are in dBm. Layer 3 establishes a link between the Layer 1 representing the mobile motion at instant t and the Layer 2 recording power values.

Layer 3 changes at each instant because it computes again the difference of field strength at every step between all cells grid and the current cell been located by the mobile (that is the reference point). Figure 5.29 shows a representation of Layer 3 at instant t.

If the mobile is at position i at a given instant t, assuming that the motion started at this position then the corresponding power value is –85.3 dBm, as indicated in Layer 2, denoted $P_i(t)$. At this position the three power control algorithms presented earlier, that are LI-SRA, AFM and ACP, as well as the UMTS classical step by step power control algorithm, will run for ten cycles, this value is a parameter to set. Each of them will return the value of the controlled power. P_{ci} is the controlled power resultant from the UMTS algorithm at the end of 10 cycles.

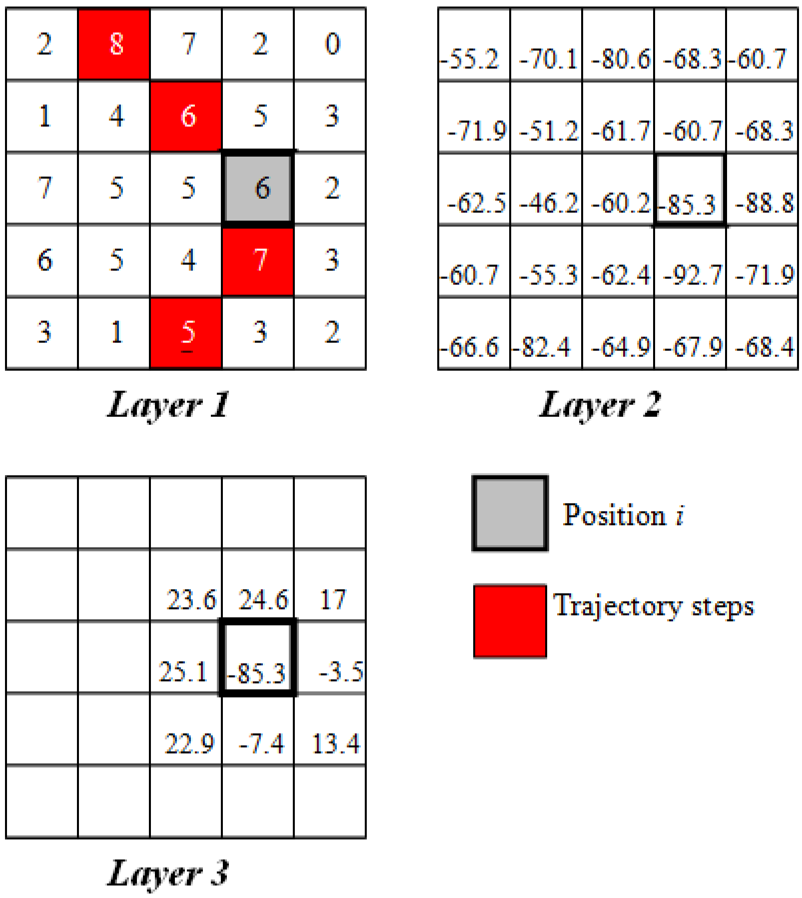


Figure 5.29. Simulation of the three layers

When moving to position $i+1$, if we follow the trajectory in Layer 1 from the bottom to the top, it corresponds to the colored cell with attraction weight 6 and corresponding power: P_{i+1}= –61.7dBm (Layer 2). To take into account the current power of the mobile at position i and the propagation loss on next cell, when transitioning from one grid cell to another we use the Layer 3 at instant t. This layer indicates the power difference between positions i and $i+1$. For the motion we are studying, the power difference from position i to position $i+1$ is +23.6dB. So the initial value of the power for position $i+1$ will be P_{i+1} = [P_{ci} + (+23.6)]. The same process is repeated for every step of the trajectory.

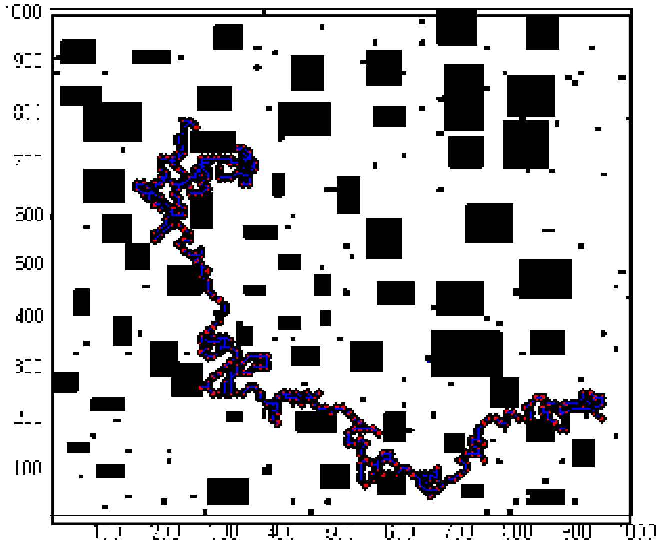


Figure 5.30. Downtown pedestrian trajectory (scenario 1)

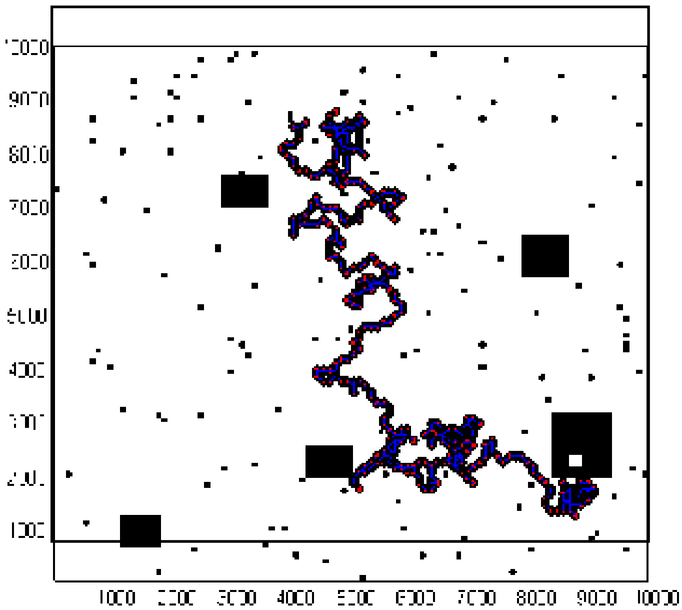


Figure 5.31. Vehicle trajectory on roads (scenario 2)

On this basis, several scenarios were created in order to undertake the simulations. In this paper we present two typical scenarios for urban and road motion. The areas of these scenarios are more or less large (from $1km^2$ to $100km^2$) and involve more or less power control during the terminal mobility. The first one is a representation of a city center of size 1kmx1km represented by a 100 by 100 grid. Each element of the grid is a square cell of size 10mx10m. The cells are filled with attraction weights able to bias the individual in motion and drive him throughout his displacement.

In this scenario there are a lot of repulsion blocks which represent the buildings. This scenario is illustrated in Figure 5.30 where the simulation of the mobility from MBMM underlines the motion around buildings (black boxes). This scenario is used to simulate pedestrian motion where the individual moves across the map between these blocks.

The second scenario plotted in Figure 5.31 intends to simulate a vehicle in motion on roads. In this area each grid cell represents 100mx100m. Repulsion blocks are considered as small village (large blocks) and isolated buildings (small blocks). In comparison with the previous one, this simulation scenario uses less repulsion blocks and the trajectory is straighter even if it is not that clear in the figure, this is caused by the fact that the scenario map and the grid cells are 10 times larger than in scenario 1.

In both tests we did not parameterize the MBMM to follow the shortest path from the initial point to the final one. The individual motion has being driven by the attraction weights of the simulation area map. For vehicle simulation, attractive points are on roads but the vehicle may use several paths. In the tests we chose the trajectory simulation looks like a postman one, going from house to house, doing some deviation from straight line between the starting and the ending point. We will now present the simulation results.

5.2.5 Simulation Results

In this section we present the simulation results for both scenarios: pedestrian motion and vehicle motion. At first, the Table 1 illustrates the computation of the received signals with the COST model; it only gives 10 positions of both trajectories as samples. Along the trajectory, the COST model is applied to compute the signals received from the mobile on each surrounding Base Station (BS) depending on the environment. We used a maximum of 6 BS at a time which is a good representation of a network density. These signals are then used for the computation of the SINR (signal to interference ratio) and the power control from the fourth algorithms presented above on each point during the displacement.

Table 5.1. Power variation on different BS along the trajectory

Vehicle		*Pedestrian*	
BS Received Power (dB)	*Cell nr*	*BS Received Power (dB)*	*Cell nr*
-104.4311	*1*	*-74.3692*	*1*
-104.2182	*1*	*-74.9915*	*1*
-98.6396	*2*	*-69.3123*	*2*
-95.6185	*2*	*-64.8192*	*2*
-106.0714	*3*	*-76.2222*	*3*
-101.6978	*3*	*-70.9485*	*3*
-98.2938	*3*	*-68.4949*	*4*
-98.2773	*3*	*-71.9184*	*5*
-105.7350	*4*	*-71.9184*	*5*
-102.71553	*4*	*-65.3977*	*6*

Table 5.1 shows the evolution of the initial received power value, before power control, on 10 points along the undertaken trajectory of both studied cases. Also one column indicates the BS number receiving the best signal from the mobile. So it is possible to see the variation of best server along the trajectories. On these ten steps, the vehicle transits through four radio cells and the pedestrian walks throughout six cells, so the motion involves power control and handover between cells as well. These are only examples issued from both trajectories but sometime several successive points are not involved in BS handover. All cases are met.

In the tests we present, the requested SINR Γ_0 is equal to – 20dB, which is the threshold value for UMTS voice communication. This value can be changed to meet the requirements of any other UMTS service. Simulations were done with a noise value of – 117dBm and an average number of jamming mobiles of twenty for the pedestrian scenario and of five for the vehicle one; we assumed that there are more jammers in the city center than on the highway. All these parameters can be changed in order to check the algorithms robustness on different traffic conditions (hundreds of mobile for instance).

The jammers simulated in this application are nothing but other individuals present in the simulation area. The individuals we are tracking in the simulation scenarios are not the only individuals that are simulated. For the pedestrian case 50 individuals are simulated walking in the city at 5km/h and for the vehicle scenario we simulated 20 individuals at 120km/h. The power value of all the individuals changes along their motion. So for the target individual, a number of neighboring individuals or jammers are firstly identified. This can be done either by computing the Euclidean distance between the target individual position and all other individuals or by checking if the power value of the individuals is greater than a given threshold.

Then Figures 5.32 and 5.34 present the variation of the SINR levels in dB by using the different algorithms. On horizontal axis we represent the sequence of positions at which 10 cycles of power control have been applied and on vertical axis we represent the SINR value in dB after these 10 cycles of power control. The SINR_UMTS curve shows the variation of the algorithm used in UMTS. The SINR_Lisra, SINR_AFM and SINR_ACP represent respectively the SINR control using the LI-SRA, AFM and ACP distributed power control algorithms.

Figures 5.33 and 5.35 present the power variation in dBm relative to Figures 5.32 and 5.34 respectively. P_i is the power variation using the step-by-step UMTS power control algorithm (with 1dB step each time slot which is the most common feature). P_Lisra, P_AFM and P_ACP are the power variation curves using LI-SRA, AFM and ACP respectively. Finally the TJ, or Total Jamming, curve is the evolution of total jamming value perceived by the best server BS along the trajectory in each scenario, knowing that the average number of jamming mobiles in each scenario is different.

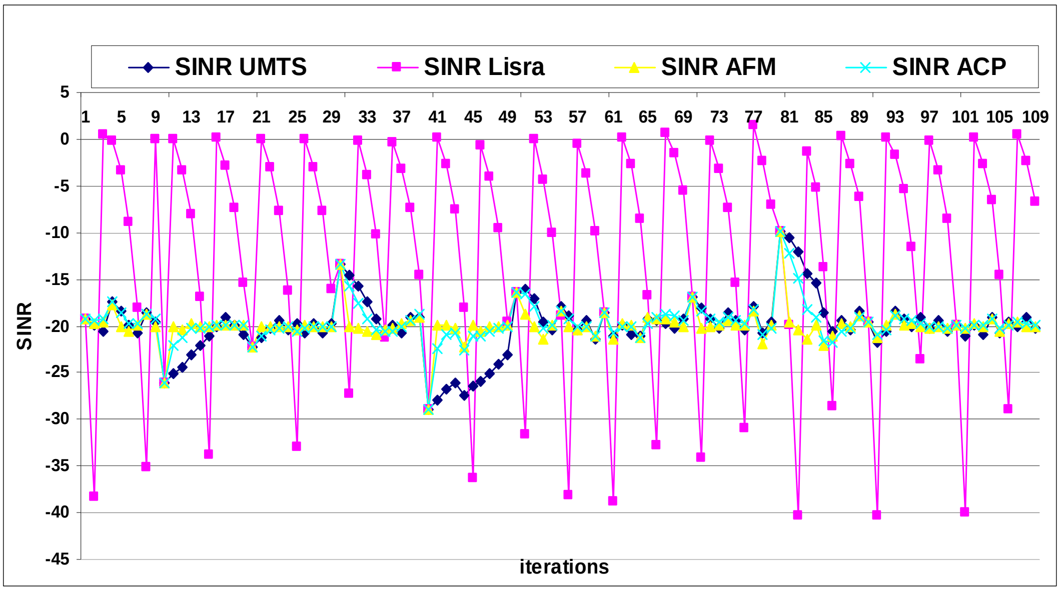


Figure 5.32. SINR variation for vehicle

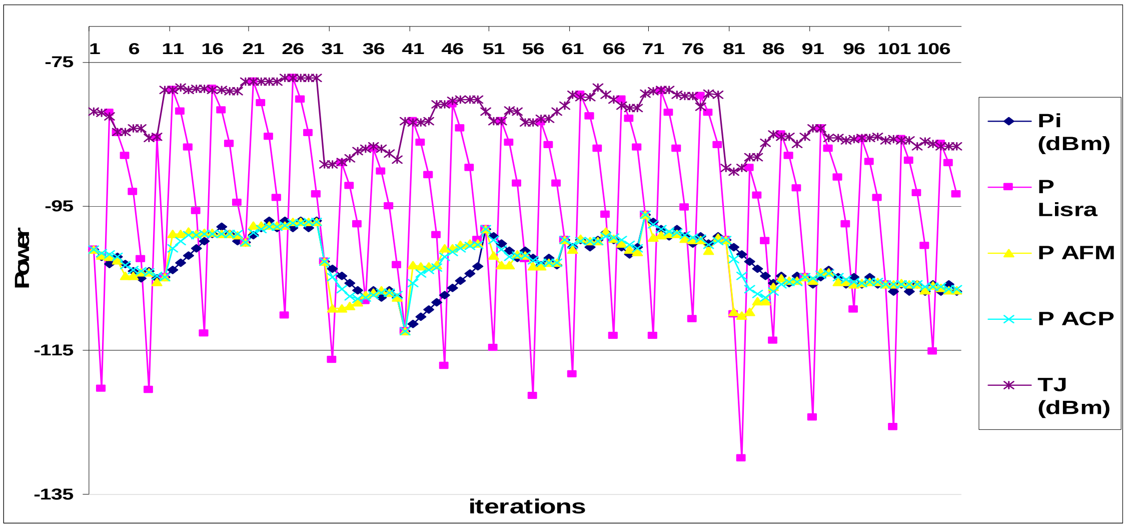


Figure 5.33. Power variation for vehicle

In Figures 5.32 and 5.34 we see that the control of the LI-SRA algorithm takes a long time to converge, which can be clearly explained by the important power variation shown in Figures 5.33 and 5.35. This algorithm is the worst one. AFM and ACP present better performances than the LI-SRA and also they are better that the UMTS power control algorithm. These two methods converge faster than the UMTS classical method to the SINR requested value (–20dB).

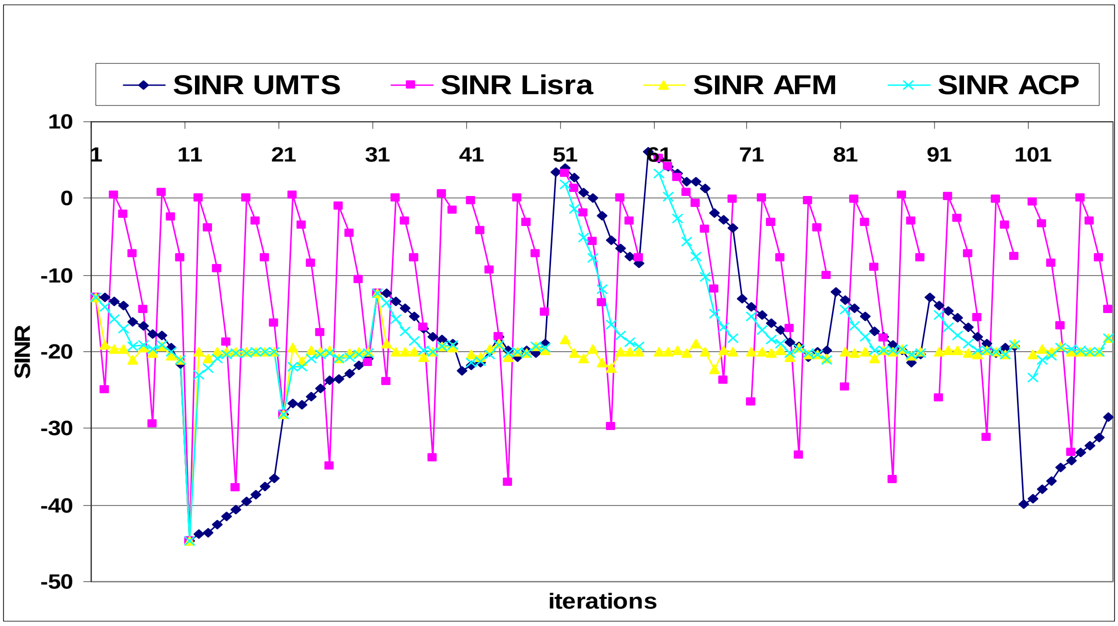


Figure 5.34. SINR variation for pedestrian

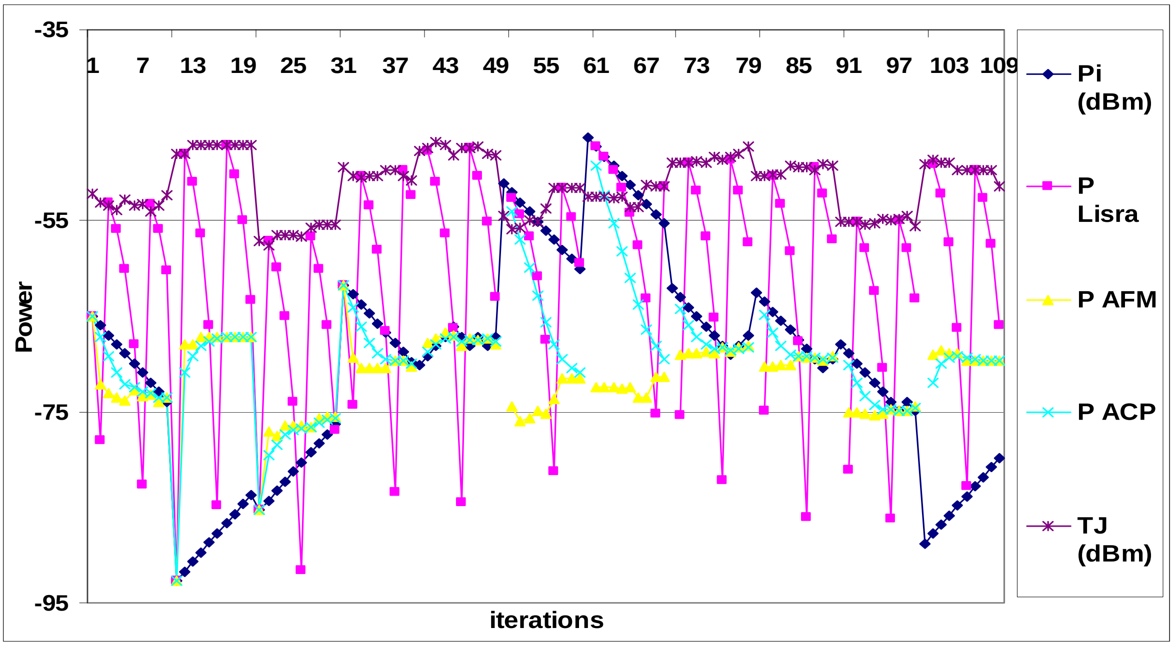


Figure 5.35. Power variation for pedestrian

These observations also hold in Figures 5.33 and 5.35 where the power variation is smoother for UMTS, ACP and AFM than LISRA. Globally the AFM algorithm has the best performance among all the simulated algorithms in both motion cases. We see in Figures 5.33 and 5.35 that the AFM power (P_AFM) and the total jamming (TJ) curves are parallel. It shows a perfect adaptation of the AFM algorithm to the jamming degree permitting a faster convergence to the requested value. Then the quality of the SINR regulation with AFM is very good, the result is always close to the SINR target. The UMTS control is the smoother one. It avoids ping-pong effect of signal around the SINR target and this is a good thing when the target is almost reached. But most of the time the mobility context we have tested brings high variations on SINR and then UMTS power control is not efficient to rapidly reach the target. Also when several mobiles are jamming as in pedestrian test, the step-by-step UMTS procedure is not so efficient than the gradient ones. Then the global system loses capacity: each mobile is over-jammed by the other one during a longer period until the target is reached.

5.2.6 Synthesis

In this paper we used the mobility model MBMM to present a study of three distributed power control algorithms, LI-SRA, AFM and ACP, and we compared them to the UMTS power control algorithm. To do that work, we calculated the trajectory simulated by the Mask Based Mobility Model which gives a non-random continuity of the power to control from one step to another. We defined two scenarios with different motions and speeds to check the algorithms in road for vehicle and city for pedestrian contexts. The results showed a better performance for the AFM procedure in converging and adapting the control and the SINR to the changing values of power and jamming. The UMTS control loses capacity in both vehicular and pedestrian cases.

The work mechanism presented in this paper might be applied to any other power control algorithm, not only the distributed ones but also the centralized and the cooperative models. It also provides a very flexible simulation environment due to the fact that any map or zone can be used as a simulation area. It can be divided into grid cells of different sizes in order to have a higher or less degree of precision on power control results. Also the simulation time can be divided into several periods. As for each of these periods a different weight value is affected for a grid cell, it makes the simulation area dynamic over time, making it the closest to real environment which is not possible with other mobility models encountered in literature.

One next step on this power control algorithm validation work may be to combine both scenarios in a more general one, having two users profile simultaneously: vehicle and pedestrian. The number of mobiles for each profile as long as the speed of each mobile terminal must be parameters to be taken into consideration. This will allow the network planner to study the effects of different motion and speed contexts on power control in a very sensitive way. This kind of simulations allows the engineers to have a very reliable and quick test of the systems, gaining a lot of time and money on experiments.

Chapter 6. Conclusions and perspectives

During the last decade, mobility modeling has become a major research topic. Research has shown that there is some kind of statistical pattern in human motion, which helped in building mobility models. The importance of such models lays in the fact that they help to predict the movement of individuals, pedestrians, cars, or any kind of human mobility, for a given period of time and over a given geographical area. Many applications prove that these models can play a major role in assessing transportation and communication networks, and navigation techniques, in terms of performance analysis. It is now possible to predict congested and deserted areas ahead of time, which will in particular allow the optimization of country and town planning.

In this thesis we aimed at developing a simulation environment and a generic mobility model to contribute to that research area by associating and analyzing several information coming from different sources (such as mobile network and bus transportation systems) which are varying along time and very complex. The work was done by the development of a whole process starting from data analysis and clustering to the proposition of one new model through the definition of several mobility components. This work was partly developed in cooperation with SMTC (*Syndicat Mixte de Transport en Commun*) which is the operator in charge of bus network in the *Territoire de Belfort* and Orange France which is the leading company in radio communications in France, as part of a project for the French *Pôle de Compétitivité "Véhicule du Futur"*, a cluster of companies and universities working on displacement and transportation problems.

The first step of our work was the definition of the simulation area. The simulation area is composed of the geographical area of interest divided into equal square grid cells. In this thesis, the simulation area used for the work was the city of Belfort and its suburbs. We used data from several sources on that city (GIS, mobile phone operator and bus transportation) to characterize and calibrate the simulation environment. The conjunction of this information in one place and one time is something very unique to study the mobility. The data was collected from 14 October 2006 till 30 October 2006. Each day was divided into periods of 15 minutes. For each of these periods, the mobile phone operator data supplied information for incoming calls, outgoing calls and handovers between radio cells, whereas the bus transportation network data provided information concerning individuals taking the bus at each bus stop.

Unfortunately, this collected data carries a lot of raw information from different sources, and could not be used to characterize a correct and useful simulation map, and that will be the case for any other place with such information. Therefore some data analysis methods were necessary to extract the useful information out of the raw data sets. In this thesis, we used the principal component analysis (PCA) and the *k*-means clustering method in order to extract and organize the information from the raw data. The PCA extracts the major information from the matrices of data. The PCA is a statistical technique used to reduce the dimension of data with minimum loss of information. The principle is to take a set of correlated variables and produces a set of principal components, which is smaller in number than the original set of variables, and which are uncorrelated i.e. orthogonal. Classification analysis objective was to divide a data set into classes. It creates subsets of homogeneous individuals of different category (radio cells, time periods…) where the individuals of the same group are as similar as possible regarding to certain characteristic and the different groups are as dissimilar as possible. The result of the analysis along with the GIS information was used to transform the geographical area into the simulation environment. At the end of this process each of the simulation environment grid cells is characterized by one or several numerical values. The values for each grid cell result from the classification of the analyzed data along with the information about the dominant structure present in this grid cell. The values are defining attraction weight for the given area.

After the simulation environment was set, the Mask Based Mobility Model (MBMM) was defined. The model is based on previous mobility models with the added value of new components, such as the displacement mask, on the grid cell simulation environment transforming each step to a Markov chain of up to 9 states. This model combined the advantages of several other general mobility models, and created a generic mobility suitable for urban planning, network dimensioning and many other applications. The motion principle in this model was Markovian with addition to a short memory of the last previous step. This allows setting, when applied, a series of displacement policies to make the simulated trajectories as close to real track as possible. The individuals were divided into groups based on their needs on the simulation area during a day. The simulation time was divided into regular periods and the simulation's environment changed regularly with different attractive and repulsive areas. The model is able to interpret these changes and to adapt the mobility to these changes. The main features of this model were verified by a series of simple verification tests.

Two applications of the mobility model in its proper simulation environment were undertaken. The first one was a comparison work between the new mobility model with other mobility models of the same category from the literature. All the models were simulated on the same simulation environment and several new quantitative metrics were proposed to evaluate different aspects of the induced mobility. The metrics were divided into two main categories: trajectory metrics for individual displacement and group metrics for population flow. For the individual metrics we were interested by the trajectory that we divided into two major parts, curls and non curls. Curls are the concentration of displacement steps in a relatively small area for a given period of time. The curls should be concentrated as it should constitute a potential destination for the individual in motion. The non curl part of the trajectory is the path linking two consecutive curls. This path is optimal when the individual in motion chooses to go from one curl to another straight without changing the direction. The population metrics compared the position of the simulated population with the potential presence information. This information was deducted from the data analysis and from real surveys. The metrics were applied to the different models and gave a very good ranking for the MBMM model. The trajectories simulated with MBMM were easily decomposed in curl and non curls where the curls corresponded to possible destination for an individual in motion. The convergence of the simulated population for every period of the time was compared to the real positions issued from the mobile network measurements and reports a small and acceptable error for simulation.

The second application of MBMM was a simulation of a power control algorithm in UMTS, where the mobility model was used to simulate the individual motion. Three distributed power control algorithms for UMTS, LI-SRA, AFM and ACP, were compared with the real UMTS power control algorithm. MBMM was used to simulate the displacement of the individuals. Simulating the individual trajectory using MBMM gave a non-random continuity of the power to control from one step to another. Two simulation scenarios were defined with different motions and speeds to check the quality of the power control on the mobile station on uplink radio link in road for vehicle at 120km/h and city for pedestrian at 5km/h. The results showed a better performance for the AFM procedure in converging and adapting the control and the SINR to the changing values of power and jamming. The UMTS control, defined as a standard, loses capacity in both vehicular and pedestrian cases. This application shows the importance of a good mobility modeling for system conception and engineering.

Just like any other model, the importance of the MBMM lays in the possible applications and future enhancements. We believe that there are several possible applications for the simulation environment and mobility models. The simulation of Vehicular Ad-Hoc Network (VANET) communication can be considered nowadays, as one of the most popular topics in the wireless networking research field. The aim of research for VANET is to create a vehicular communication system that permits a fast communication between different vehicles traveling on the same road. The most important parameter when simulating VANET is the vehicle mobility. This is why it is of major importance to have a realistic mobility model such as MBMM. The use of MBMM will allow the simulation of vehicle displacement on a real city or road network. The output of such simulation will be a valuable input for the evaluation of data transfer protocols between vehicles. As a matter of fact, this application is being carried out as the topic of a new thesis report in the lab that proposes an extended version of MBMM to simulate vehicular mobility on graphs which includes a lot of new elements like varying speed management, distance between cars, crossover stops, etc. This ongoing model called V-MBMM is more dedicated to car simulation with the integration of specific infrastructure parameters.

Another venue for future work on communication simulation is the comparison of UMTS handover algorithms and parameters settings over individual traces in city and highway. This will follow a similar procedure as that followed for power control presented in this thesis. A comparison between several handover algorithms and the original UMTS handover algorithm could be done. Two scenarios will be proposed with low speed displacement in city and high speed displacement for highways. The importance of such study is that nowadays about 70% of calls dropped inside UMTS networks are due to bad handover parameters settings due to bad simulation process. In the same area, another mobile network application could be the study of service availability and continuity for the GMS/UMTS networks covering the simulation area. The idea behind this study is to establish quality metrics for the mobile network services provided by the operator and evaluate them in geographical areas according to individual and group displacement and concentration for every day of the week. The individuals and groups mobility can be simulated using MBMM.

An additional future work to do is the development of real indoor simulation environment to run the model for indoor scenarios and applications. The area of interest for indoor scenarios will be characterized using the same data analysis procedure used for the outdoor simulation environment presented in this work but with smaller and 3D grid cells. These grid cells will contain the information about the structures and objects covered by the grid cell at any location. Additional information concerning the rate at which each structure is visited or occupied over time in hours, days and/or weeks is required, and then the model will require some enhancements to manage this information. After these two sets of information will be analyzed and clustered, a well characterized simulation terrain will be ready for MBMM run. Different case studies can be considered: resorts, supermarkets, offices... In a resort it is crucial to know the importance that each activity or attraction has, and how they should be arranged and presented for an optimal space management. In the supermarket the display of goods will attract individuals to go through the different brands and products that are offered. For a given supermarket different goods display possibilities can be tested in order to reach the optimal solution. In offices space management is very important to provide employees a good work environment.

One last future study using MBMM will be about the impact of the population displacement and flow on the economic activity in city, and the impact of the economic activities in a city on the mobility of its population. These economical aspects are not included in our model. People's behavior is a key factor for the economic development of several activity centers in city, and these centers of activity also have a great impact in controlling population behavior and habits. Adding or modifying the position or even removing a structure that has an economical activity in a city will have great impact on the population's mobility. The population is divided into individuals and groups with different means of transportation: bus, car, cycle, pedestrians... Economical activity structures are removed and/or created and for every possibility a new motion pattern and habits will appear for the different groups of the studied population. Quality metrics for the MBMM model will be developed in order to evaluate the impact of these changes on the population and the consequences of such changes will be evaluated. The model will be used in a further ANR project on city organization which will start by the end of 2010.

REFERENCES

[1] Christian Bettstetter, Mobility Modeling in Wireless Networks: Categorization, Smooth Movement, and Border Effects, ACM SIGMOBILE Computing and Communication Review, Volume 5, Issue 3, pp 55-56, ACM Press 2001.

[2] Christian Bettstetter, Smooth is better than Sharp: A random Mobility Model for Simulation of Wireless Networks, Proceedings of the 4th ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems, ACM press, 2001.

[3] Chiu-Ching Tuan, Chen-Chan Yang, A Normal Walk Model for Mesh PCS Networks, Proceedings of the 18th International Conference on Advanced Information Networking and Applications, Volume 2 AINA'04,IEEE Computer Society, March 2004.

[4] Chiu-Ching Tuan, Chen-Chan Yang, A compact Normal Walk Model for PCS Networks, Proceedings of the 18th International Conference on Advanced Information Networking and Applications, Volume 2 AINA'04, IEE Computer Society, March 2004.

[5] William Navidi , Tracy Camp, Stationary Distributions for the Random Waypoint Mobility Model, - IEEE Transactions on Mobile Computing, 2004.

[6] Fan Bai, Ahmed Helmy, "A Survey of Mobility Models in Wireless Ad hoc Networks". IEEE Transactions on Communications, 52(2):183–186, Febuary 2004.

[7] Tracy Camp, Jeff Boleng, Vanessa Davies, A Survey of Mobility Models for Ad hoc Network Researches, Wireless Communications F4 Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications, 2(5):483 502, 2002.

[8] M. Sanchez. Mobility models. http://www.disca.upv.es/misan/mobmodel.htm, 1998.

[9] Santashil PalChaudhuri, Jean-Yves Le Boudec, Milan Vojnovic, Perfect Simulation for Random Trip Mobility Models, Proceedings of IEEE Infocom 2005.

[10] Karen H. Wang, Baochun Li, Group Mobility and Partition Prediction in Wireless Ad-Hoc Networks, in Proceedings of IEEE International Conference on Communications (ICC), NYC, NY, April 2002.

[11] Chinmay Shete, Salil Sawhney, Supriya Herwadkar, Varun Mehandru, Ahmed Helmy, Analysis of the Effects of Mobility on the Grid Location Service in Ad Hoc Networks, IEEE International Conference on Communications, vol. 7, pp. 4341-4345, 2004.

[12] Fan Bai, Narayanan Sadagopan, Ahmed Helmy, IMPORTANT: A framework to systematically analyze the impact of Mobility on Performance of Routing protocols for Adhoc Networks, IEEE INFOCOM (The 22nd Annual Joint Conference of the IEEE Computer and Communications Societies), March/April, San Francisco, 2003.

[13] Xiaoyan Hong, Mario Gerla, Guangyu Pei and Ching-Chuan Chiang, A Group Mobility Model for Ad Hoc Wireless Networks, In Proceedings of ACM/IEEE MSWiM'99, Seattle, WA, pp.53-60, Aug. 1999.

[14] Yenliang Lu, Huier Lin, Yajuan Gu, Ahmed Helmy, Towards Mobility-Rich Analysis in Ad Hoc Networks: Using Contraction, Expansion and Hybrid Models, Communications, 2004 IEEE International Conference on, 2004.

[15] Jungkeun Yoon, Mingyan Liu, Brian Noble, Random Waypoint Considered Harmful, in Proc. IEEE INFOCOM, pp. 1312-1321, 2003.

[16] C. Siva Ram Mutrhy, B.S. Manoj, Ad Hoc Wireless Networks: Architectures and Protocols, Prentice Hall Communications Engineering and Emerging Techno, May 24, 2004.

[17] D. R. Basgeet, P. Dugenie, A. Munro, D. Kaleshi, J. Irvine, SMM: Mathematical Framework of a Scalable Mobility Model, 2003.

[18] Biao Zhou, Kaixin Xu, and Mario Gerla, Group and Swarm Mobility Models for Ad Hoc Network Scenarios using Virtual Tracks. Computer Science Department, University of California, Los Angeles, CA 90095, MILCOM–IEEE Military Communications, 2004.

[19] B. P. Quiles, An Enhanced Mobility Model, M.Sc. thesis, Inst. Fur Nachrichtentechnik und Hochfrequenztechnik, Tech. Univ. Wien, Austria, 2004.

[20] Jong-Kwon Lee, Jennifer C. Hou, Modeling Steady-state and transient Behaviors of user Mobility: Formulation, Analysis and application, International Symposium on Mobile Ad Hoc Networking & Computing, Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing, PP: 86 – 96, 2006.

[21] D. Sacko, Huang Benxiong, Wang Furong, Overview Of Reference Region Group Mobility Model For Ad Hoc Networks, Ubiquitous Computing and Communication Journal - Volume 3 Number 1 , Volume 3 No. 1, 2008.

[22] Hans-Martin Zimmermann, Ingo Gruber, Christian Roman, A Voronoi-based Mobility Model for urban environments, 11th European Wireless Conference - Next Generation wireless and Mobile Communications and Services, Nicosia, Cyprus, 2005.

[23] Guanha Yan, Hector D. Flores, Leticia Cuellar, Nicolas Hengartner , Vincent Vu, Stephan Eidenbenz, Bluetooth Worm Propagation: Mobility Pattern Matters!, ASIAN ACM Symposium on

Information, Computer and Communications Security, Proceedings of the 2nd ACM symposium on Information, computer and communications security, PP: 32 - 44, Singapore, 2007.

[7] Tracy Camp, Jeff Boleng and Vanessa Davies, A survey of Mobility Models for Ad Hoc Network Research, 2002.

[24] Kyunghan LEE, Seongik Hong, Seong Joon Kim, Injong Rhee, Song Chong, SLAW: A Mobility Model for Human Walks, IEEE INFOCOM, 2009.

[25] Wen-Tsuen Chen, Po-Yu Chen, Group Mobility Management in Wireless Ad Hoc Networks, Institute of Electrical and Electronics Engineers, 2003.

[26] Xiaofeng Lu, Yung-chih Chen, Ian Leung, Zhang Xiong, Pietro Liò, “A Novel Mobility Model from a heterogeneous Military MANET Trace”, Springer Berlin / Heidelberg, PP: 463-474, 2008.

[27] Bo Sun, Fei Yu, Kui Wu, Victor C.M. Leung, Mobility-Based Anomaly Detection in Cellular Mobile Networks, Workshop on Wireless Security, Proceedings of the 3rd ACM workshop on Wireless security, PP: 61 – 69, 2004.

[28] Swee Ann Tan, William Lau, Allan Loh, Networked Game Mobility Model for First-Person-Shooter Games, Network and System Support for Games, Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games, PP: 1 - 9, 2005.

[29] Lee, John A. and Verleysen, Michel “Nonlinear Dimensionality Reduction”, Information Science and Statistics. 2007.

[30] Jolliffe, I.T.’Principal Component Analysis’. Second Edition. Springer. 2002

[31] Aldenderfer, M.S and Blashfield, R.K. ‘Cluster Analysis’. Beverly Hills: Sage. 1984.

[32] Johson, Richard A. and Wichern, Dean W. ‘Applied Multivariate Statistical Analysis’ Fourth Edition. Prentice Hall, 1998.

[33] Hartigan, J. and Wang, M.’ A K-means clustering algorithm’. Applied Statistics, 28, 100-108. 1979.

[34] Krishnaiah, P.R. and Kanal, L.N.. ‘Handbook of Statistics 2. Classification Pattern, Recogniton and Reduction of Dimensionality’. North-Holland Publishing Company. 1982.

[35] Ding, Chris and He, Xiaofeng. ‘K-means Clustering via Principal Component Analysis’. Proceedings of the 21st International Conference on Machine Learning, Banff, Canada, 2004.

[36] Brian T. Luke. K-Means Clustering. http://fconyx.ncifcrf.gov/~lukeb/kmeans.html

[37] Peter J. Rousseuw. ‘Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis’. Computational and Applied Mathematics 20: 53–65, 1987.

[38] R. Lleti, M.C. Ortiz, L.A. Sarabia, M.S. Sánchez. "Selecting Variables for k-Means Cluster Analysis by Using a Genetic Algorithm that Optimises the Silhouettes". Analytica Chimica Acta 515: 87–100, 2004.

[39] B. Dubuisson.'Diagnostic et reconnaissances des formes', HERMES, 1990.

[40] A. Baccini et P. Besse : Data mining – Exploration Statistique, Laboratoire de Statistique et Probabilité – UMR CNRS C5583, 2004. mining

[41] J. B. MacQueen : "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297, 1967.

[42] Kaufman L. et P.J. Rousseuw : Finding Groups in Data: An Introduction to Cluster Analysis, Wiley, 1990.

[44] Andrew Moore. K-means and Hierarchical Clustering - Tutorial Slides. http://www-2.cs.cmu.edu/~awm/ tutorials/kmeans.html.

[45] J. Pagès : Eléments de comparaison entre l'analyse factorielle multiple et la méthode STATIS, Revue de Statistique appliquée, 1996.

[46] M. Bécue-Bertaut et J. Pagès : Analyse factorielle multiple intra-tableaux. Application à l'analyse simultanée de plusieurs questions ouvertes, 2000.

[47] D.L. Huff: In Probabilistic Analysis off Shopping Center Trades Areas - Economics Land, 1963.

[48] D.L. Huff: Parameter Estimate in the Huff Model - ESRI-ArcUser, 2003.

[49] 3GPP TS 25.101 V5.4.0, Power Control. 3rd Generation Partnership Project. 2002-2009.

[50] Qiang Wu, "Performance of optimum transmitter power control in CDMA cellular mobile systems", Vehicular Technology, IEEE Trans. on Vol. 48, Issue 2, March 1999.

[51] Tsern-Huei Lee; Jen-Cheng Lin; Su, Yu.T., "Downlink power control algorithms for cellular radio systems. Vehicular Technology, IEEE Trans. on Vol. 44, Issue 1Page(s):89 – 94, Feb. 1995.

[52] Grandhi, S.A. Vijayan, R. Goodman, D.J. Zander, J., "Centralized power control in cellular radio systems Vehicular Technology, IEEE Trans. on Vol. 42, Issue 4Page(s):466 – 468, Nov. 1993.

[53] Chi Wan Sung; Wing Shing Wong, "A cooperative algorithm for asynchronous distributed power control in cellular systems". GLOBECOM '96. 'Communications: The Key to Global Prosperity Vol. 3, 18-22 Nov. 1996.

[54] Chi Wan Sung, Wing Shing Wong, "The convergence of an asynchronous cooperative algorithm for distributed power control in cellular systems", Vehicular Technology, IEEE Trans. on Vol. 48, Issue 2, March 1999.

[55] Antoszkiewicz, A.; Gazor, S., “A cooperative algorithm for cellular assignment and power control”, Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on Vol. 3, 4-7 May 2003.

[56] C. W. Sung, W. S. Wong: “Performance of a Cooperative Algorithm for power control in cellular systems with a time-varying link gain matrix”. Wireless Networks, 2000.

[57] Uykan, Z., Jantti, R., Koivo, H.N., “A PI-power control algorithm for cellular radio systems”, Spread Spectrum Techniques and Applications, IEEE 6th Int. Symposium, Vol. 2, 6-8 Sept. 2000.

[58] Wang Hongyu, Hu Rong, Huang Aiping; Gu Weikang, “A distributed power control algorithm for cellular radio systems”. WCC - ICCT 2000, Vol. 1, 21-25 Aug. 2000.

[59] Won Jay Song; Byung Ha Ahn, “Distributed power control using the simultaneous mutation of genetic algorithms in cellular radio systems Information Technology”: Coding and Computing, 2002. Proceedings International Conference on 8-10 April 2002.

[60] Grandhi, S.A.; Zander, J., “Constrained power control in cellular radio systems”, Vehicular Technology Conf., IEEE 44th, June 1994.

[61] 13 Min Cai; Wei Wang; Huan Shui Zhang, “Power control algorithm for time-varying CDMA cellular systems”,Intelligent Mechatronics and Automation, 2004.

[62] Dongwoo Kim; Kun-Nyeong Chang; Sehun Kim, “Efficient distributed power control for cellular mobile systems”, Vehicular Technology, IEEE Trans. on Vol. 46, Issue 2, May 1997.

[63] Moscibroda, T.; Rejaie, R.; Wattenhofer, R., “How Optimal are Wireless Scheduling Protocols?”, INFOCOM 2007.26th IEEE International Conference on Computer Communications. May 2007.

[64] Qiang Wu, “Optimum transmitter power control in cellular systems with heterogeneous SIR thresholds”, Vehicular Technology, IEEE Trans. on Vol. 49, Issue 4, July 2000.

[65] Zander, J., “Performance of optimum transmitter power control in cellular radio systems”, Vehicular Technology, IEEE Trans. on Vol. 41, Issue 1, Feb.1992.

[66] Zander, J., “Distributed cochannel interference control in cellular radio systems”, IEEE Trans. on Vehicular Technology, Vol. 41, Issue 3, Aug.1992.

[67] Foschini, G.J. Miljanic, Z, “A simple distributed autonomous power control algorithm and its convergence”, Vehicular Technology, IEEE Trans. on Vol. 42, Issue 4, Nov. 1993.

[68] L. Lima, J. Barros, “Random Walks on Sensor Networks”, 5th WiOpt, Limassol, Cyprus, April, 2007.

[69] L. Lima, J. Barros, “Finding Coverage Bounds for Constrained Random Walks over Sensor Networks”.3rd Int. Workshop on Mathematical Techniques and Problems in Telecom., Portugal, 2006.

[70] C. Joumaa, A. Caminada, S. Lamrous, “Mask Based Mobility Model A new mobility model with smooth trajectories”, SpaSWiN 2007, Limassol, Cyprus, April, 2007.

[71] C. Joumaa, A. Caminada, S. Lamrous, “Mask Based Mobility Model: A 2-D indoor 3-D outdoor mobility model”, EMC Europe Workshop 2007, Paris, France, June 2007.

[72]M. Jeong, B. Lee, “Comparison between path-loss prediction models for wireless telecommunication system design”, Antennas and Propagation Society International, IEEE, Boston, USA, 2001.

[73] E. Damosso COST 231 (Digital mobile radio towards future generation systems), Final report, European Comission, Bruxelles, 1999.

[74] A. Burulitisz, S. Imre, and S. Szabo, On the accuracy of mobility modelling in wireless networks, presented at IEEE International Conference on Communications, 2004.

[75] A. Bar-Noy, I. Kessler, and M. Sidi, Mobile users: to update or not to update? Presented at INFOCOM '94. 13th Proceedings IEEE of Networking for Global Communications, Toronto, Ont, 1994.

[76] Observatoire sociétal du téléphone Mobile enquête. http://www.afom.fr/, 2005.

[77] Observatoire sociétal du téléphone Mobile enquête. http://www.afom.fr/, 2006.

[78] Observatoire sociétal du téléphone Mobile enquête. http://www.afom.fr/, 2007.

[79] O. S. Mitrea. « Understanding the mobile telephony usage patterns -The rise of the mobile communication 'Dispositif'», Thesis, Technischen Universität Darmstadt, 03/02/2006.

[80] « A comparative study of mobile phone use in public places in London, Madrid and Paris», DWRC, Vodafone, 09 August 2004.

[81] Lee Humphreys: Can you hear me now? A field study of mobile phone usage in public space, M.A. University of Pennsylvania, 2003

[82] International Telecommunication Union (2004): Social and Human Considerations for a more Mobile World Background Paper.

[83] Ling Rich: Adolescent Girls and young adult men: Two subculture of the mobile telephone Kjeller, Telenor Research and development R&D Report 34/2001).

[84] Ling, Rich : The diffusion of mobile telephony among Norwegian teens. A report after the revolution. Presented at ICUST 2001, in Paris, June 2001.

[85] Lorente, Santiago: Youth and Mobile Telephones: More than a Fashion. In: Revista de Astudios de Juventud 57, June 2002, pp. 9-24.

[86] Fox, K. (2004): Evolution, alienation and gossip. The role of mobile telecommunications in the 21th century. Social Issues Research Centre. Retrieved March 14, 2004

[87] Hans Geser: Are Girls (even) more addicted?, University of Zurich, June 2006

[88] Hans Geser: Towards a Sociological Theory of the Mobile Phone, University of Zurich, May 2004

[89] Hans Geser, Késia U. e S. Trench: Pre-teen cell phone adoption: consequences for later patterns of phone usage and involvement, University of Zurich, April 2006

[90] Fabien Michel. The IRM4S model: the inuence/reaction principle for multiagent based simulation. In Sixth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS07). ACM, May 2007

[91] J. Ferber, F. Michel, and J. Baez. AGRE: Integrating environments with organizations. In Danny Weyns, H. Van Dyke Parunak, and Fabien Michel, editors, Third International Workshop (E4MAS 2006), volume 4389 of Lecture Notes in Artificial Intelligence, pages 48{56. Springer, Hakodate, Japan, may 2006

[92] Simmonds D., Echenique M. et al. Bates J., Octobre 1999, “Review of Land-Use/ Transport Interaction Models”, DETR, 72p.

[93] “UrbanSim: Modeling Urban Development for Land-Use, Transportation and Environmental Planning”, 4 Waddell P., 2001.

[94] Bejan, A. 2006 Advanced engineering thermodynamics, 3rd edn. Hoboken, NJ: Wiley

[95] Bejan,A.,Lorente, S. & Lee, J. 2008 Unifying constructal theory of tree roots, canopies and forests. J. Theor. Biol. 254, 529–540. (doi:10.1016/j.jtbi.2008.06.026)

[96] Bejan, Adrian; Merkx, Gilbert W. (Eds.), “ Constructal Theory of Social Dynamics” 2007, XVII, 350 p., Hardcover ISBN: 978-0-387-47680-3,

[97] A. Bejan, “Advanced Engineering Thermodynamics”, Wiley, 2nd edition, 1997, 896 p. ISBN 0471148806

[98] A. Bejan, “Shape and Structure, from Engineering to Nature”, Cambridge University Press, Cambridge, UK, 2000, 324 p. ISBN 0521793882

[99] "Bejan's Constructal Theory of Shape and Structure" Edited by Rui N. Rosa, A. Heitor Reis & A. F. Miguel, Evora Geophysics Center, Portugal, 2004, ISBN 972-9039-75-5

[100] A. H. Reis, A. F. Miguel , M. Aydin, Constructal theory of flow architecture of the lungs, Journal of Medical Physics, May 2004, Volume 31, Issue 5, pp. 1135-1140.

[101] A. Bejan, S. Lorente, “La Loi Constructale”, L'Harmattan, Paris, 2005. ISBN 2-7475-8417-8

[102] A. Bejan, S. Lorente, “Constructal theory of generation of configuration in nature and engineering”, Journal of Applied Physics, 2006, Volume 100, no. 041301.

[103] A. Bejan, S. Lorente, “Design with Constructal Theory”, Wiley , 2008, ISBN: 978-0-471-99816-7

[104] McNally, Michael G., “The Four Step Model, center for activity systems analysis”, Institue of transportation study University of California Irvine 2008

[105Erlander, S. and N.F. Stewart, “The gravity model in Transportation Analysis Theory and Extensions: topics in Transportation “1990, ISBN 90-6764-089-1

[106] Ortuzar and Willumsen, “Modelling Transport”, Wiley, 1990

[107] D. Simmonds. Representing Planning Policy in Land-Use and Land-Use/Transport Modelling. http://www.davidsimmonds.com/, 2010.

[108] H. Zha, C. Ding, M. Gu, X. He and H.D. Simon. "Spectral Relaxation for K-means Clustering", Neural Information Processing Systems vol.14 (NIPS 2001). pp. 1057-1064, Vancouver, Canada. Dec. 2001.

[109] P.J. Huber. Projection pursuit. Annals of Statistics, 13(2):435-475, 1985.

[110] K.V. Mardia, J.T. Kent, and J.M. Bibby. Multivariate Analysis. Probability and Mathematical Statistics. Academic Press, 1995.

[111] D.J. Hand. Discrimination and Classiffication. New York: John Wiley, 1981.

[112] L.J. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. Classification and Regression Trees. Wadsworth, Belmont, California, 1984.

[113] K. Fukunaga. Introduction to Statistical Pattern Recognition. San Diego: Academic Press, 2nd edition, 1990.

[114] T.J. Hastie and R.J. Tibshirani. Generalized Additive Models, volume 43 of Monographs on Statistics and Applied Probability. Chapman & Hall, 1990.

[115] B.D. Ripley. Pattern Recognition and Neural Networks. Cambridge University Press, 1996.

[116] M.A. Carreira-Perpinan. A review of dimension reduction techniques. Technical report CS-96-09, Department of Computer Science, University of Sheffield, 1997.

[117] A. Hyvarinen. Survey on independent component analysis. Neural Computing Surveys, 2:94, CiteSeerX, 1999.

[118] Liu, L.,, Andris, C.,, Biderman, A.,, and Ratti, C. Uncovering Taxi Driver's Mobility Intelligence through His Trace. IEEE Pervasive Computing 2009

[119] Calabrese, F. and Ratti, C. Real Time Rome.Networks and Communication Studies - Official Journal of the IGU's Geography of Information Society Commission 2006, 20(3&4), 247-258

[120] Pulselli, R. M., Romano, P., Ratti, C. & Tiezzi, E. (2008) Computing Urban Mobile Landscapes through Monitoring Population Density Based on Cell-Phone Chatting. International Journal of Design & Nature and Ecodynamics, 3, 2, pp. 121-134.

[121] Ratti, C., Pulselli, R.M., Williams, S., and Frenchman, D.Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis 2006. Environment and Planning B, 33(5), 727-748.

[122] Beyer, K. S., Gildstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is "nearest neighbor" meaningful? In proceeding of the 7th International Conference on Database Therory, Jerusalem, Israel, pages 217-235. Springer

[123] Hotelling, H. (1933). Analysis of complex statistical variables into principal components. Journal of Educational Psychology, 24:417-441, 498-520.

[124] Karhunen, K. (1947). Uber lineare methoden in wahrscheinlichkeitsrechnung. Ann. Acad. Sci. Fenn. 37

[125] M. Kim, D. Kotz and S. Kim “Extracting a mobility model from real user traces”, Proc. IEEE Infocom, 2006

[126] Zhuoqun Li, Lingfen Sun and Emmanuel C. Ifeachor “GPS-Free Mobility Metrics for Mobile Ad-Hoc Networks”, IET Proceedings on Communications, vol.1, issue 5, pp. 970-976, October 2007

[127] Y.C Hu, D.B. Johnson, “Caching Strategies in on Demand Routing Protocols for Wireless Ad Hoc Networks”, In Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCom 2000), August 2000.

[128] P. Johansson, T. Larsson, N. Hedman and al. “Scenario_Based Performance Analysis of Routing Protocols for Mobile Ad-hoc Networks”. In MobiCom '99: Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking, pp. 195-206, 1999.

[129] A. K. Saha and D. B. Johnson, “Modeling mobility for vehicular ad hoc networks,” in Proc. of the ACM Workshop on Vehicular Ad hoc Networks (VANET), 2004, pp. 91–92.

[130] W.J. Hsu, K. Merchant, H.W. Shu, C.H. Hsu and Ahmed Helmy, “Weighted Waypoint Mobility Model and its Impact on Ad Hoc Networks”, SIGMOBILE Mobile Computing and Communications Review, January 2005

[131] S. R. Das, C.E. Perkins and E.M. Royer, “Performance Comparison of Two On-Demand Routing Protocols for Ad Hoc Networks”, MK Marina - IEEE INFOCOM, 2000

[132] J. Harri, F. Filali and C. Bonnet, “Mobility Models for Vehicular Ad Hoc networks: A survey and Taxonomy”, Research Report, Institut Eurecom, France, March 2006

[133] Chris Ding and Xiaofeng He, “Principal Component Analysis and Effective K-means Clustering”. Lawrence Berkeley National Laboratory University of California, Berkeley, CA 94720, http://www.siam.org/proceedings/datamining/2004.

[134] Imola K. Fodor, “A survey of dimension reduction techniques Center for Applied Scientific Computing”, Lawrence Livermore National Laboratory, Livermore, CA 94551, June 2002.