Destination Choice Model including panel data using WiFi localization in a pedestrian facility

Loïc Tinguely

Antonin Danalet

Matthieu de Lapparent

Michel Bierlaire

EPFL

April 2015

STRC | **15th Swiss Transport Research Conference**

Monte Verità / Ascona, April 15 – 17, 2015

EPFL

Destination Choice Model including panel data using WiFi localization in a pedestrian facility

Loïc Tinguely, Antonin Danalet,
Matthieu de Lapparent, Michel Bierlaire
Transport and Mobility Laboratory,
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

phone: +41 21 693 25 32
fax: +41 21 693 25 32
loic.tinguely, antonin.danalet, matthieu.delapparent, michel.bierlaire@epfl.ch

April 2015

Abstract

This paper proposes a general methodology to model pedestrian destination choice from WiFi localization in multi-modal transport facilities (e.g., airports, railway stations). It is based on the output of Danalet *et al.* (2014) method to generate candidates of activity episode sequences from WiFi measurements, locations of activities on a map and prior information.

Destination choice is nested to the activity choice. An individual first chooses an activity (Danalet and Bierlaire, 2015), and then selects the destination where to perform it. We propose an approach to model destination choice accounting for panel nature of data. We compare static, dynamic strictly exogenous and dynamic with two different agent effect corrections models with inspiration from Wooldridge (2002) method.

In a case study using WiFi traces on EPFL campus, we focus on one activity: catering. The choice set contains 21 alternatives on campus (restaurants, self-services, cafeterias,. . .). Our models reveal that the choice of a catering facility especially depends on habits (e.g., where an individual ate the previous time), distance to walk from the previous activity episode (calculated with a weighted shortest path algorithm) and destination specific determinants. Price has a non-significant impact in this case study, most likely because the price range on campus is narrow. The models are successfully validated using the same WiFi dataset.

Keywords

Destination Choice Modelling, pedestrians, WiFi signatures, panel data, Wooldridge

Contents

List of Figures

List of Tables

1 Introduction

This paper proposes a framework to model pedestrian demand in multi-modal transport hubs such as airports or train stations. The use of those facilities increases, both for trains (425 million passengers in 2008 in Switzerland, 477 million in 2013, +12% (OFS, 2014)) and for planes (2 billion passengers in 2005 in the world, 3 billion in 2013, +50% (Worldbank, 2014)). Historically, paper-and-pencil or telephone surveys were conducted to collect data (information on people behavior and habits). They were expensive and could not be performed often. Nowadays, modern hubs all propose "free Wi-Fi". Localization data from cell phones, tablets and computers can thus be collected from access points all around these stations. These data are cheap and can cover the whole facility. Recently, Danalet *et al.* (2014) developed a methodology to use these data, where each measurement is associated to a point of interest (e.g., coffee shop, restaurant, ticket machine. . .) in time.

Knowing people location in time permits to generate probabilistic candidates of activity episode sequences. They can be used to develop both an activity choice model (this first step is discussed in Danalet and Bierlaire (2015)) and a destination choice model. These two models are sequential. Liu *et al.* (2014) suggest that they are explored together. Once an individual has chosen an activity, he selects the destination where to perform it (Bierlaire and Robin, 2009). This report discusses the second step of the sequence. It especially focuses on the development of a general methodology to describe and understand destination choice for pedestrians. We model and forecast people behavior based on WiFi data when visiting such a facility. These forecasts aim at optimizing multi-modal transport hubs, e.g., finding the optimal location for coffee shops or ticket machines.

To be more specific, the project is part of a collaboration between "Ecole Polytechnique Fédérale de Lausanne" (EPFL) and the Swiss Railway company (CFF) in the context of the project "Léman 2030" (CFF, 2014). It includes an increase in trains offer (100'000 travelers expected in 2030, 50'000 in 2013 – +100%) and huge changes in Geneva and Lausanne train stations. Furthermore, the CFF are one of the most important property owners in Switzerland. Their lands have become a major source of income. RailCity is the name given to the largest railway stations because of their similarities to cities: more and more, train stations offer the opportunity not only to travel but also to eat, drink, shop, or entertain oneself. In 2009, the revenue of these infrastructures was about 1.09 billion of Swiss francs (CFF, 2011). In order to optimize their stations, the company wants to know how people behave when they visit the facility.

A random pedestrian for example arrives at the station at 7:45 AM (it is the beginning of his activity episode sequence), buys a ticket at 7:49 AM, gets a sandwich at 7:58 AM and then moves to the platforms at 8:01 AM to take a train scheduled at 8:04 AM (it is the end of his activity episode sequence). Once the activity is chosen (i.e., buying a ticket or getting a sandwich), a specific destination needs to be chosen (i.e., a specific ticket machine or a specific luncheonette). If the pedestrian wants a sandwich, he has to visit a place where such a service is available (Subway, Polli, Coop. . .). He performs a destination choice. As soon as the destination is known, a path needs to be defined. These nested episodes represent pedestrians tactical and strategical behavior (Hoogendoorn *et al.*, 2002). Similar studies have already been made for a destination choice model in railway stations (e.g., Ton (2014); Liu (2013)) or airports (e.g., Kalakou and Moura (2014b)). However, only Ton (2014) is developed from WiFi traces. The others are based on stated/revealed preference surveys. Pettersson (2011) also studies the behavior of pedestrians in train stations but he focused on the factors (e.g., information, geometry, habits) influencing the waiting location of people on departure platforms. He used both video tracking and surveys.

The goal of this project is to develop a general framework to model destination choice and to apply it to the EPFL campus. Indeed WiFi traces from April 2012 to June 2012 are available (Danalet, 2015) and allow to track random people on the campus and to define their probabilistic activity episode sequences. The paper describes several destination choice models on eating establishments (e.g., restaurants, self-services, cafeterias. . .) on EPFL campus. Multinomial Logit Models and Mixed Logit Models are generated in order to explain the factors that influence pedestrians' choice for one catering destination compared to another. This paper also reviews the existing literature (see Section 2), explains the data requirement and the data merging, introduces new concepts (i.e., according to our researches, agent effect correction for pedestrian's panel data has never been explored before) (see Section 3) and discusses several examples of destination choice models for pedestrians (see Section 4).

2 Literature review

The literature review is separated into three parts. Section 2.1 explores methods to capture indoor pedestrians activities. In Section 2.2, destination choice models in multi-modal transport hubs are analyzed. Finally, Section 2.3 discusses the advantages and disadvantages of each methods.

2.1 Activity episodes detection

In Ton (2014), WiFi and Bluetooth traces are involved. The methodology to transform signatures into activity sequences is not revealed in detail. It is shortly discussed in Section 2.2. In Danalet *et al.* (2014), data requirement consists of timestamps and localization data coming from WiFi network traces and a semantically-enriched routing graph (SERG). A measurement is defined as

$$\hat{m} = (\hat{x}, \hat{t}) \tag{1}$$

where $\hat{x} \in \mathbb{R}^3$ is the position of the measurement and $\hat{t}$ is the timestamps. The accuracy ξ defines the distribution of the Euclidean distance between the location estimate $\hat{x}$ and the actual location å

$$\hat{x} = å + \xi \tag{2}$$

In order to associate activity episodes (including stop detection and semantics of the stop) to these measurements, a semantically enriched routing graph (Goetz and Zipf, 2011) is defined as a set of nodes corresponding to the type of potential destinations (room, restaurant, shop…, i.e., all points of interest).

The methodology to detect candidates of activity episode sequences performed by pedestrians from the previously calculated digital traces follows a Bayesian approach explained in Danalet *et al.* (2014). An activity episode is defined as

$$a = (x, t^-, t^+) \tag{3}$$

where x is the episode localization and $t^+ - t^- \geq T_{min}$ the time spent at that location. A minimum threshold T_{min} of five minutes is set like in Bekhor *et al.* (2013). It permits to only keep activity episodes longer than five minutes (and thus representing a destination and not only a crossing point). The output of this probabilistic method is defined as a set of L candidates of activity episode sequences $a_{1:K_i}$ which are specific to an individual i. Basically an activity episode sequence is a list of K_i activity episodes performed (i.e., visited points of interest) by one tracked individual i during one day. Each candidate activity episode sequence is associated with the probability of being the actual one. Danalet *et al.* (2014) define this probability as a Bayes formula (the subscript i is omitted to lighten the expressions):

$$P(a_{1:K}|\hat{m}_{1:J}) \propto P(\hat{m}_{1:J}|a_{1:K}) \cdot P(a_{1:K}) \tag{4}$$

It means that the activity probability $P(a_{1:K}|\hat{m}_{1:J})$ that $a_{1:K}$ is the actual activity episodes sequence given in the measurement $\hat{m}_{1:J}$ is proportional to the product of the measurement likelihood $P(\hat{m}_{1:J}|a_{1:K})$ with a prior knowledge $P(a_{1:K})$. As the goal is to compute the probability that the performed episodes generated the observed measurement sequence, the equation is decomposed

$$P(\hat{m}_{1:J}|a_{1:K}) = \prod_{k=1}^{K} \prod_{j=1}^{J} P(\hat{x}_j^k|x_k) \tag{5}$$

It is assumed that the only measurement error is a localization error. Similar to the land use planning concept (Miller, 2010), a prior is defined as:

$$S_{x,i}(t^-, t^+) = \int_{t=t^-}^{t^+} \delta_{x,i}(t) \cdot A_i(x,t)dt \tag{6}$$

The idea is that the potential attractivity measure $S_{x,i}(t^-, t^+)$ between a start time t^- and an end time t^+ for $x \in POI$ and individual i is time dependent. It depends on the instantaneous potential activity and a dummy variable δ for time-constraints (e.g., opening hours, schedules...). The attractivity $A_i(x,t)$ must define the potential of a place (e.g., number of seating places for restaurants or number of workers per room for an office) as explained in Danalet *et al.* (2014). Then the prior can be calculated as

$$P(a_{1,K}) = \prod_{k=1}^{K} \frac{S_{x_k,i}(t_k^-, t_k^+)}{\sum_{x \in POI} S_{x,i}(t_k^-, t_k^+)} \tag{7}$$

It assumes that consecutive activity episodes are independent.

Danalet *et al.* (2014) propose an algorithm to merge data from localization and pedestrian SERG to get candidates of activity episode sequences. The generation of activity episode sequences is divided in four steps. The first one introduces the concept of domain of data relevance (DDR) introduced in Bierlaire and Frejinger (2008). The DDR defines a physical area where a probabilistic measurement location linked to a *POI* is relevant. For each measurement $\hat{m}_j$, all possible activity episodes sequences are generated for each individual. It leads to a recursively built network.

The second step consists in generating activity episodes start and end times as soon as a sequence of potential episode locations is defined. The idea is to compare two consecutive measurements $\hat{m}_j$ and $\hat{m}_{j+1}$. Their timestamps and positions define a trip between them and thus a travel time. In that way, considering a maximum walking speed and a shortest path algorithm between both positions, bounds can be determined for the earliest and the latest start time and the earliest and the latest end time. Start and end times are considered to be uniformly distributed between these two bounds.

Third, once the distribution is known for the start and end times of each activity episode, the duration is estimated. Activity episodes with a lower bound smaller as T_{min} are rejected. The last part of the procedure is the sequence elimination procedure. As the number of path in a network growth exponentially with the number of measurement, there is a need for selection. Candidates with small probability of occurrence are rejected. The complete algorithm is available in Danalet *et al.* (2014). The methodology has been tested and validated on EPFL campus.

In Dalumpines (2014), the data requirement consists in GPS data. A GIS-based episode reconstruction toolkit (GERT) automatically extracts activity episodes from GPS data and derives information related to these episodes. This kit generates an input for route choice modeling. The methodologies of Danalet *et al.* (2014) and Dalumpines (2014) are similar but the latter classifies activity episodes into different types using multinomial logit models. Also, the first one deals with small scale problems (e.g., a multimodal facility, a campus. . .) whereas the second one fits better on a much larger framework (e.g., a transportation network).

2.2 Destination choice models for pedestrians

2.2.1 Influence of Space Syntax

Price of a ticket and distance are intuitive factors used to explain a destination choice. When it comes to a pedestrian destination choice model, more determinants have to be accounted for. Kalakou and Moura (2014a,b) study the influence of space syntax (SS). SS is a theory and a set of methods about space reflecting both the objectivity of space and the intuitive engagement with it (Hillier, 2005). Important characteristics about space are connectivity, integration and visibility. Connectivity is a factor that expresses the number of “neighbors” of each space. Integration is the relation of one space with all others. According to Zhang *et al.* (2012) visibility is one of the most influential factors in people's behavior when moving in commercial areas. Ueno *et al.* (2009) find out that the visibility, the number of turns and the distance affect pedestrians' path choice in railway stations.

2.2.2 An analogy with route choice modelling

Hoogendoorn and Bovy (2004) distinguish three levels of choices: the strategic level (Activity pattern choice and departure time choice), the tactical level (Activity scheduling, destination choice and route choice) and the operation level (walking behavior). An activity may be performed at multiple destinations.

According to Hoogendoorn and Bovy (2004), the choice of an activity area is based on factors such as the directness (number of sharp turns and rapid directional changes (Helbing, 1997)), the distance and the level-of-service of the route, the necessity of performing that activity (e.g., is it mandatory?) and personal preferences. Furthermore, the choice of a route and the choice of a destination are done simultaneously thus factors influencing both choices are considered.

2.2.3 Destination choice models in airports

Kalakou and Moura (2014b) made a survey in Lisbon Portela's airport and collected information about space syntax and travelers' habits. A discrete choice model was built to capture the significant parameters that influence the choice of a destination. Four coffee shops were selected as potential destinations. Space syntax parameters were introduced in the model. Visibility from a mandatory place to visit (check-in, entrance) has a significant impact on the choice of a destination. The integration level of the activity location adds value to a place for passengers who only choose one coffee shop. Similarly places having a good connectivity are more likely chosen after the check-in.

Liu (2013) also studied pedestrian behavior in an airport on the basis of both revealed and stated preference survey data. They develop an activity-destination choice model. Travel distance, congestion or the type of service have a significant impact in people's decisions. Models validated by Liu (2013) are used for forecasting: in more than 50% of the cases, the prediction fits the observation.

2.2.4 Destination choice models in a railway stations

Ton (2014) studies the route and activity location choice behavior of departing pedestrians in the Utrecht railway station in Netherlands. Using WiFi and Bluetooth traces, she builds both destination and path choice models. Her work is based on a framework proposed by Hoogendoorn *et al.* (2002). It focuses on the strategical and tactical levels when faced with discrete choices in a train station.

Ton (2014) defines an activity as a punch. The movement of a pedestrian contains several punches (e.g., Enter the station, visit a Burger King, leave via platform…). Therefore the possible activities are caught in a punch card. However this list only tells if a pedestrian was seen at one place or not (binary observation). It means that the sequence cannot be directly derived from the punch card. Thus, the activity sequence must be determined. Ton (2014) does not explain how she defines the chronological order of the punches. One limit of the data is that the list of activities performed by an individual is only available for one day because everyone receives a new identification number everyday to respect privacy.

Using these activity sequences, Ton (2014) applies a binary discrete model to a choice of a coffeeshop. Two Starbucks are selling coffees in the railway station and the aim is to capture the factors that influence pedestrian destination choice. Travel time from entrance to coffeeshop, total distance covered and having to take a detour are robust parameters. It is interesting to note that the orientation is also significant. The fact that a coffeeshop is located on the right hand side of the railway station (from the main entrance) increases its utility because pedestrians are used to walk on the right.

2.3 Critics and comments

Given the nature of our data, we discuss how we are able to account for some of the ideas developed in reviewed literature:

- Liu (2013) and Kalakou and Moura (2014b) are based on both SP and RP surveys.
  - Socio-economic parameters can easily be taken into account with surveys, not with WiFi traces since the data are partially anonymized.
- Kalakou and Moura (2014b), Liu (2013) and Ton (2014) destination choice models were developed for destinations in only one building.
  - Impact of SS in larger facilities (e.g., a campus) is unknown.
- Liu (2013) used CCTV to emphasize the impact of congestion.

  - We have to find other indicators to take congestion into account.
- The methodology developed by Ton (2014) is limited because the route choice is dependent of the punch card's simplicity and does not measure pedestrian's habits.
  - WiFi traces are able to describe more accurately pedestrians movements using Danalet *et al.* (2014) candidates of activity episode sequences.
- Factors such as directness (Hoogendoorn and Bovy, 2004) or the works realized by Helbing (1997) and Ueno *et al.* (2009) are mainly discussing route choice.
- Alternative specific parameters (e.g., the price, the quality, the availability of services, the comfort or aesthetic indicators) are barely described in the reviewed papers but intuitively have an influence on people's choice.

The methodology developed by Danalet *et al.* (2014) is well fitted to create a destination choice model, but it has some limitations. The algorithm defined by Danalet *et al.* (2014) associates the WiFi measurements with *POI*[1] inside a zone. Points of interest are represented as points while they are areas in reality. It creates a problem when the accuracy of the measurement is good and the "zone of interest" is large. In this case, the point of interest might not be inside the domain of data relevance (*DDR*[2]). Thus, the actual point of interest, representing the possible activity performed by the receiver, might not be considered.

In the case of data collected with the Cisco Context Aware Mobility API with the Cisco Mobility Services Engine (MSE) (Cisco, 2011), the domain of data relevance is defined as a square around the measurement with sides of size $2 * cF$, where cF is called the confidence factor. The WiFi device is estimated to be in this square with 95% probability. The minimum observed cF is 16 meters (see Figure 2(c)) on EPFL campus. Some *POI* on campus clearly have a surface bigger than a $16 * 16$ square. In this case, the intersection between the *DDR* (i.e., the square with side $2 * cF$) and the point representing the *POI* might be empty, and so the actual activity episode is not detected.

This limitation is observed in the case study (see Section 4 for a detailed description). The data collected in the library of the Rolex Learning Center (RLC) are good due to the lack of walls or obstacles and due to the large number of WiFi antennas[3](Sen *et al.*, 2013; Nandakumar *et al.*, 2012). Figure 2(b) shows that the level of accuracy in the library is higher than on the rest of the campus. Figure 2(c) show that some points of the library effectively lead to an empty intersection between *DDR* and the *POI*. It is a limitation of the methodology but it can be corrected by using an area instead of a point for representing *POI*.

[1]POI: Point Of Interest, see Section 2.1

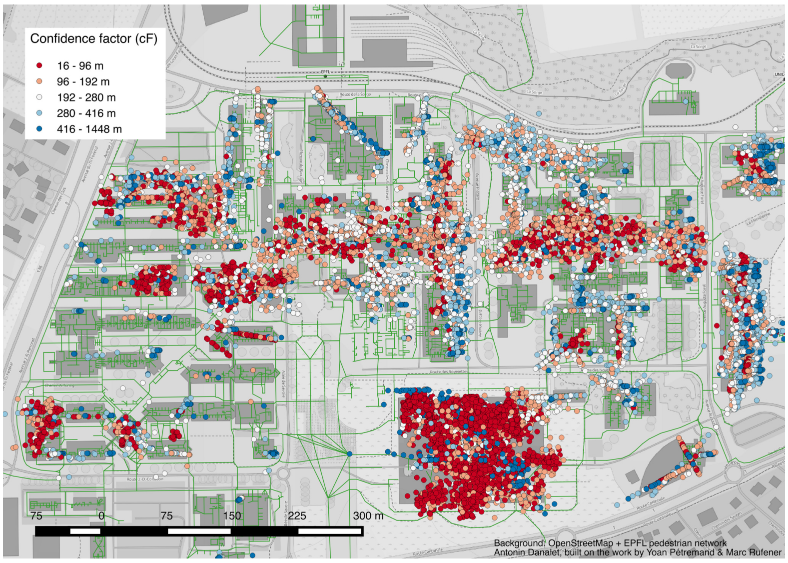
[2]Domain of Data Relevance, see Section 2.1

[3]On Figure 2(a), the density does not look higher than in other buildings on the campus. However, the RLC has only one floor, while all other buildings have more. This is just a visual effect due to projection on the map. In reality, density of WiFi access points is actually higher

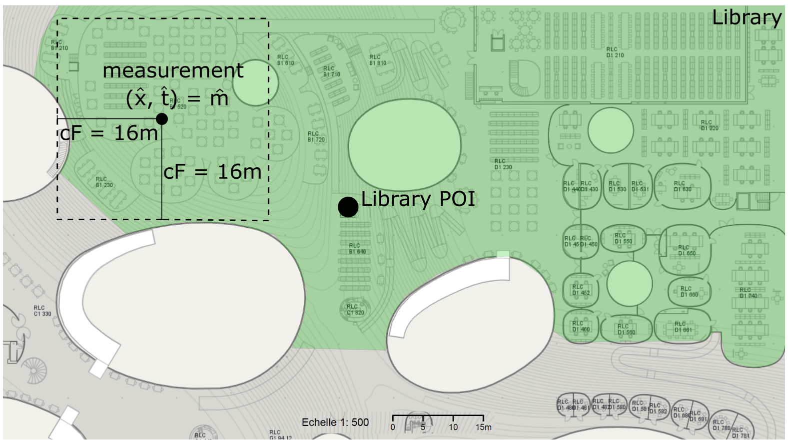
Figure 1: WiFi antennas and confidence factor (cF) on the EPFL campus

(a) WiFi antennas on the EPFL campus (map.epfl.ch)

(b) Confidence factor on the EPFL campus (Danalet, 2015)



(c) Empty intersection between DDR and POI (map.epfl.ch)



3 Methodology

Section 3.1 describes the output of the algorithm developed by Danalet *et al.* (2014) and characterizes activity episode sequences. Then, Section 3.2 presents our models' specifications and an approach to account for panel nature of data.

3.1 Activity episode sequences

3.1.1 Description of activity episode sequences

Danalet *et al.* (2014) develop a framework for detecting pedestrian mobility pattern from WiFi traces (see Section 2.1). The methodology explained in the paper is used to create candidate's lists of activity episode sequences from WiFi traces. They are then used to develop a destination choice model for pedestrians.

An activity episode sequence has several characteristics (sequence specific attributes). An example is described in Table 1 and Figure 3. Each sequence is associated to an individual (with a unique ID) tracked during one day and a probability of occurrence defined with its log-likelihood. Activity episode sequences also contain several socio-economic (e.g., age, gender, or typology of visitor) and time specific attributes (e.g., the day of the week and year of the sequence). As sequences may be calculated during a period of several months, each individual has potentially more than one observation.

Within the sequence, there are one or more activity episodes. Each activity episode is related to a point of interest (see Section 2.1). It is described by its start and end times bounds (following a uniform distribution). Each point of interest associated to an activity episode defines an activity and a destination. The activity is grouping destinations in categories. Typical categories, or activity types, are *working*, *maintenance*, *shopping*, etc. Destinations are more detailed. They have a name, coordinates and floor. Each type of destination is subject to an independent choice model.

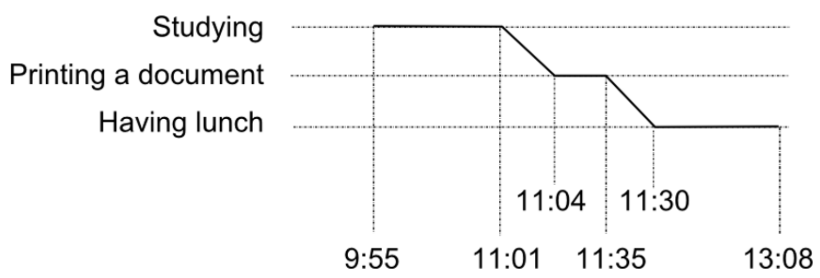
3.1.2 Characterization of activity episode sequences

Each activity type corresponds to several possible destinations. For each destination, three types of attributes exist: sequence attributes (it corresponds to attributes specific to the whole one-day

Table 1: This sequence taken from a second year bachelor student (ID=10001) in civil engineering contains 3 activity episodes caught the 29th of June 2012. This student has been seen 112 times by the Cisco WiFi device (only the destinations are kept). Each activity episode is related to a point of interest. In that case the student first visited the library, then printed something (still in the library) and finally went to eat at the library's self-service. These sequences are the input of our methodology. Each activity episode has an upper and lower bound for both start and end times (replaced by their mean on this figure).

Nb of observations: 112, Nb of activity episodes: 3, Date: 2012-06-29						
Start_time	End_time	Floor	Name	Type	X coordinate	Y coordinate
09:55:01	11:01:30	1	Library_name	Library	533226.888831	152274.939064
11:04:39	11:30:03	1	Printer_Lib	Printer	533229.919333	152284.564615
11:37:23	13:08:04	1	Self-service_Lib	Restaurant	533197.354323	152223.135494

Figure 3: This same sequence can also be represented graphically



sequence), activity episode attributes (it stands for attributes relative to one activity episode only) and alternative attributes (they are the destination specific attributes, they need to be collected). They are defined and imaged with examples in Table 2.

Table 2: Table of attributes

Sequence attributes	Activity episode attributes	Destination attributes
Day of the year	Activity-type	Capacity
Day of the week	Start/end times	Price/Quality
Socio-economic attributes	Coordinates	Integration
Individual specific attributes	Floor	Opening hours

3.1.3 Calculating distances

By comparing two consecutive activity episodes of a same activity episode sequence, one can calculate the distances between the two destinations (and similarly for all elements in the choice set in a discrete choice context). There are two possibilities to calculate the distances. First, one can simply compute an Euclidean distance between the consecutive activity episodes a_{t-1} and a_t using the (x, y) coordinates of the points.

$$d(a_{t-1}, a_t) = \sqrt{(a_{t-1,x} - a_{t,x})^2 + (a_{t-1,y} - a_{t,y})^2} \tag{8}$$

It means that we relax the assumption of anisotropy (Kim and Hespanha, 2003) and thus pedestrians can reach each point with a straight line path. A better way to calculate the distances is by using a shortest path algorithm. It may already have been constructed if the methodology explained by Danalet *et al.* (2014) has been strictly followed (indeed the path generation is based on it). It takes into account the network anisotropy and thus we obtain realistic distances.

3.2 Modelisation

3.2.1 A destination choice model

We develop a multinomial logit model with a linear-in-parameters formulation. The probability of choosing a destination d compared to the others is defined as:

$$P(d|D) = \frac{e^{\mu V_{dn}}}{\sum_{j=1}^{D} e^{\mu V_{jn}}} \tag{9}$$

We propose to split the utility function to explain the parameters one suggests to introduce. We use again the definition from Table 2. Activity episode sequences specific parameters are mainly represented by the distances between the consecutive performed activities. The distance beta should be a specific one if the destinations all offer the same type of offer (e.g., a same type of ticket machine). If the destination studied is more heterogeneous one should use alternative specific parameters (e.g., eating establishments).

Furthermore we propose to split the distance depending of the period of the day if time of the day may change the purpose of the visit (e.g., people visit a pub at 12 AM probably to eat but at 10 PM to drink a beer). Still from the sequence, socio-economic parameters are difficult to take into account because the data are usually partially anonymous. We suggest that the gender, the age and the type of visitor are collected and introduced in the model as dummy variables to alleviate the alternative specific constant.

The timestamps we propose to introduce in the distance function are activity episode specific parameters. There are only few factors from this category that are added in the utility function. The activity type permits to select a specific type of activity and the destination of the selected activity which represents the choice that the individual made. The floor of the destination is introduced in the case of a place without elevators.

Alternative specific parameters can be variables representing the congestion (capacity, queues), the quality/price ratio, the space syntax (visibility, integration, directness, detour), the type of services offered or the advertising (communication, information, directional sign). The case study presented in Section 4 gives an example in the context of catering destinations.

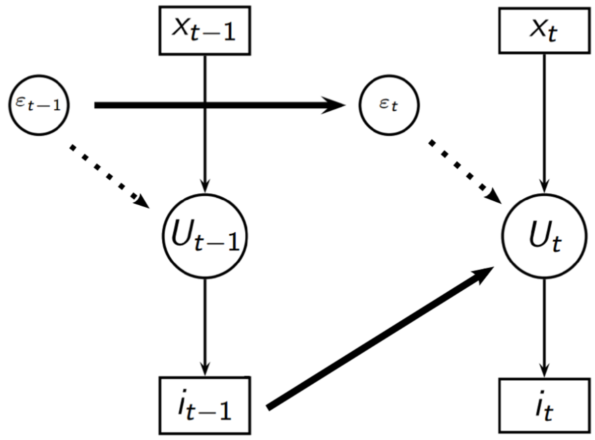
3.2.2 Accounting for panel nature of data

If the network traces are collected without anonymizing the identity of the individual too often, activity episode sequences are available for a long enough period to observe repeated destination choices for the same activity type and a same individual. Thus it is possible to take into account the habits of each individual $i \in I$ (where I is the total sampled population of individuals). Wooldridge (2002) develops a general methodology to deal with unobserved individual heterogeneity in dynamic panel data with discrete dependent variables. We apply it to our pedestrian destination choice model.

The habits of an individual i are considered as the previous choice for the same type of activity performed at a similar time of the day. It is represented as a dummy variable that takes the value 1 for the previously chosen alternative, 0 otherwise and -1 if no previous choice is available. There is no strict and regular periodicity between 2 consecutive choices: it can be one day, two weeks or several months, and it may change from individual to individual and from observation to observation. We improve this feature in future developments to make a consistent definition of temporal dimension. The difficulty of considering activity episode sequences over time is that the problem becomes dynamic (Bierlaire, 2014).

The utility function at time t takes into account the choice performed at time $t - 1$. It means that the observations and the error terms are not independent anymore. Figure 4 shows the interaction between error terms, utility functions and choices performed. It leads to serial correlation and agent effect issues (also known as one-way effect, i.e. time-invariant unobserved terms). We here consider that the error terms are defined as the sum of two unobserved components. The first is

Figure 4: Dynamic Markov model with correlation



Source: (Bierlaire, 2014)

a time-invariant unobserved effect (i.e., σ_i in Equation (12)) and the second is an error term that is independent and identically distributed over time and individuals (i.e., $u_{i,t}$ in Equation (12)). If we assume that current choices are influenced by past choices, the individual error terms are correlated over time. We thus need to correct for this correlation issue. According to Wooldridge (2002), it is possible to manage this issue by defining a function c_i that is that is (1) conditional to the initial choice and (2) time-invariant observed characteristics of the individual. We consider the following distribution:

$$c_i|y_{i,0}, z_i \sim Normal(\alpha_0 + \alpha_1 y_{i,0} + \alpha_2 z_i, \sigma_i^2) \tag{10}$$

We rewrite the function c_i as:

$$c_i = \alpha_0 y_{i,0} + \alpha_2 z_i + \sigma_i \tag{11}$$

σ_i is a parameter to be determined, normally distributed and independent of y_{i0} and z_i. y_{i0} is the first choice ever made by an individual i. z_i reveals the individual behavior among the past period (e.g., average distance covered, most frequently chosen destination…). Thus the choice of the alternative d at time t performed by i is rewritten as:

$$y_{d,i,t} = \beta z_{d,i,t} + \rho y_{i,t-1} + \alpha_0 y_{i,0} + \alpha_2 z_i + \sigma_i + u_{i,t} \tag{12}$$

Basically, the choice that the individual i does depends on some parameters observed at time t, his choice made at time $t-1$ and is corrected with his first choice ever performed, some observed habits among the past observations, a normally distributed zero centered error distribution and a common error term. The model is thus mixed in errors. It takes into account a panel effect specific to each individual. The parameters $\beta, \rho, \alpha_0, \alpha_2$ and σ_i are estimated. As suggested by Pirotte (1996), one has to consider short-term (between individuals variability) and long-term effects (within an individual variability). It means that some parameters that used to be significant in the short-term (without panel effect) should be left in the model even if they are not anymore.

Table 3: Definition of static and dynamic models. AE stands for Agent Effect. In the case study (see Section 4), one decides to split the dynamic model with agent effect correction into two submodels to fully understand the influence of each term of the Wooldridge (2002) correction (e.g., α_2 is equal to zero in one model).

Static model	Dynamic strict exogenous model	Dynamic with AE correction model
$\rho = 0$	$\rho \neq 0$	$\rho \neq 0$
$\alpha_0 = 0$	$\alpha_0 = 0$	$\alpha_0 \neq 0$
$\alpha_2 = 0$	$\alpha_2 = 0$	$\alpha_2 \neq 0$ or $\alpha_2 = 0$
$\sigma_i = 0$	$\sigma_i = 0$	$\sigma_i \neq 0$

We consider and compare three situations: a static model (no previous choice considered at all), a dynamic strict exogenous with the period model (previous choice considered but with the assumption that individuals have no memory on short observation periods. It means that the choice is exogenous within a short period, but endogenous over time) and a dynamic situation with panel data and agent effect model (previous choice considered and agent-effect issue corrected). These cases are explained in Table 3.

The strict utility function may have the following shape:

$$\begin{aligned} V_{i,d,t} = ASC_d + &\beta_{socio-eco} * SOCIOECO_i + \beta_{altspecific} * ALTSPECIFIC_d + \\ &\beta_{distance} * DISTANCE_d + \rho * CHOICE_{i,t-1} + \\ &\alpha_0 * CHOICE_{i,t_0} + \alpha_2 * SOMEHABITS_{i,\bar{t}} + \sigma_i \end{aligned} \tag{13}$$

where i is an individual, d a destination and t is the time. From an activity type to another (e.g., buying a ticket, visiting a shop, drinking a coffee…), a specific model must be developed with a specific panel of attributes. In this paper, we make the strong assumption that choices of destinations for different types of activities are independent: sequences of activities are series of independent choices.

4 A case study on EPFL campus

We perform a case study on the Ecole Polytechnique Fédérale de Lausanne (EPFL) campus (see Section 4.1). The methodology developed by Danalet *et al.* (2014) converts WiFi localizations collected from students and employees into activity episode sequences. These data are dated spring 2012. Due to privacy issues, they are partially confidential (see Section 4.2). In Section 4.3, some descriptive statistics on the data are reported. The models are presented and discussed in Section 4.4. A validation is proposed in Section 4.5.

4.1 The EPFL campus

We decide to work on the catering facilities destination choice and with the most likely candidate of activity episode sequences only as a first approach (see Section 3.1). It represents 21 possible alternatives (destinations). Their locations and types are represented on Figure 5. They are separated in 5 categories (restaurants, self-services, cafeterias, caravans and others) depending on the sort of service they propose (see Table 4). We use the methodology introduced in the previous chapter (see Section 3).

The activity episode sequences contain socio-economic information such as the individual anonymized and unique ID and the occupation (student or employee, see Section 4.2). They also collect the day of the year and the start and end times of the full sequence. Activity episodes contain start and end times and the location of the activity (destination).

We compare two consecutive activity episodes of a same day to calculate the distance between all the possible destinations (see Section 3.1.3). As People are tracked during a period of three months each individual has several observations (activity episode sequences). We use them to measure their habits (previous, first and most frequent choices as explained in Section 3.2.2).

More information is required in order to explain people destination choice. These factors are related to the destination (destination specific attribute) and not to the individual (socio-economic attribute). Services' availability is described in Table 5. Factors such as prices, outside/inside capacities, opening hours or quality surveys have been collected from the EPFL restauration service. Collected data are explained below.

Figure 5: Localization of destinations on the EPFL campus (map.epfl.ch)

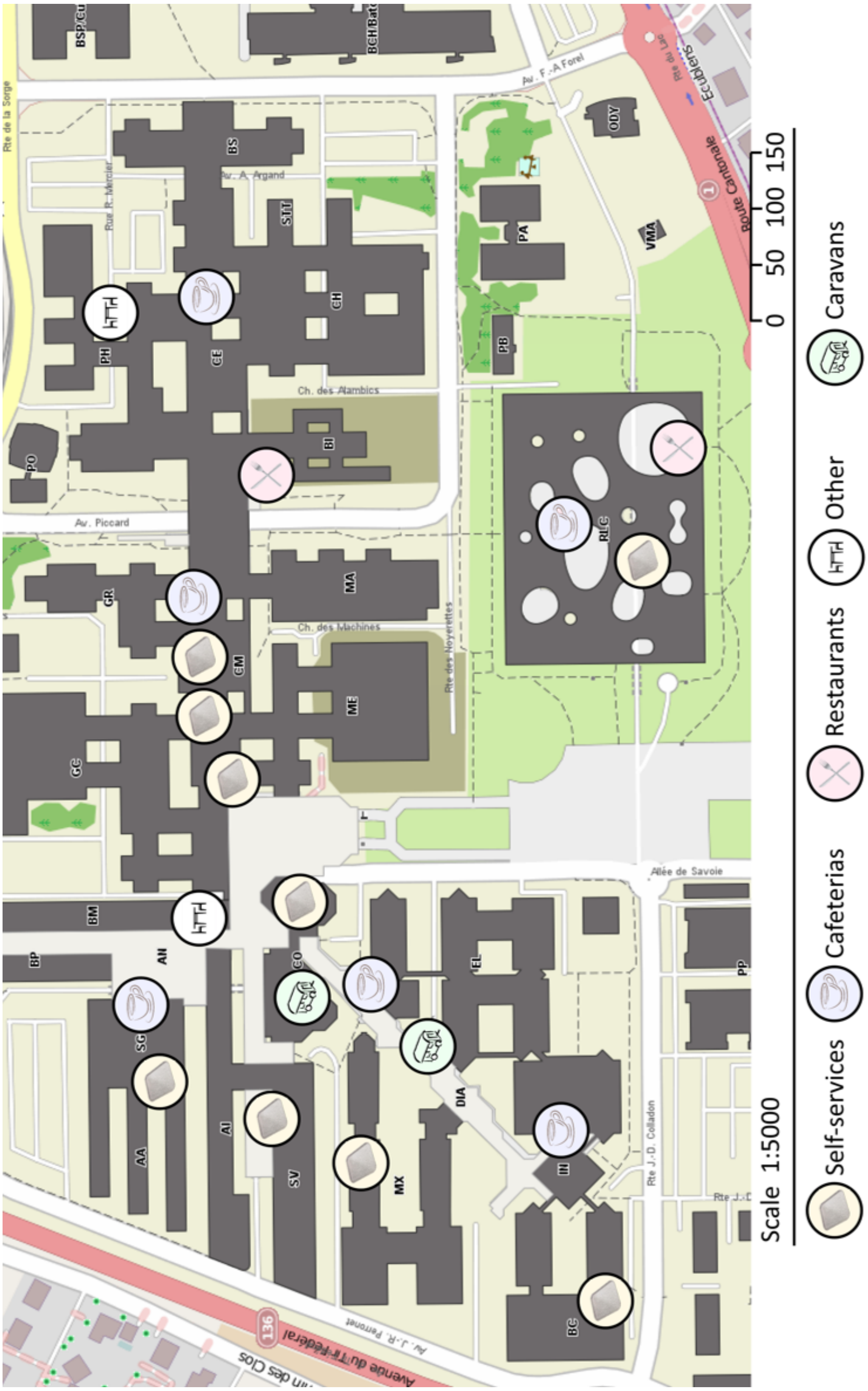


Table 4: Table of types of destinations

Destination	Type
Cafe Le Klee	Cafeteria
BC	Self-service
BM	Other
ELA	Cafeteria
INM	Cafeteria
MX	Cafeteria
PH	Other
L'Arcadie	Cafeteria
L'Atlantide	Self-service
Le Copernic	Restaurant
Le Corbusier	Self-service
Le Giacometti	Cafeteria
Le Parmentier	Self-service
Le Vinci	Self-service
L'Esplanade	Self-service
L'Ornithorynque	Self-service
Pizza	Caravan
Kebab	Caravan
Satellite	Cafeteria
Le Hodler	Self-service
Table de Vallotton	Restaurant

Cafeterias mostly offer coffee and sandwiches and can usually be used as workspaces outside lunch hours. Self-services have at least one hot lunch menu and may also propose pizzas, meat or pastas. Restaurants have several menus, propose a table service and are more expensive than the other catering destinations.

Caravans sell kebabs, pizzas and French-fries. They can be considered as fast-foods. The other catering areas are tables with an automatic coffee machine and a microwave. They are used for coffee breaks. Thus, catering destinations are not necessary visited with intent to have lunch. As Table 6 suggests, some of them are open all day as others are only open for a couple of hours during lunch time. The lunch period is the only moment of the day where all the eating establishments on the campus are open.

Table 7 shows the maximum and minimum prices for a hot meal at each destination. One can see that self-services all have a 7 CHF menu for students (Self-services get subsidies from EPFL) except for self-service L'Ornithorynque and self-service L'Atlantide who have a menu for about

Table 5: Table of services availability

Destination	Coffee	Hot meal	Tables service	Visibility	Terrace	Workspace	Green fork	Dinner	Sandwiches	Selecta	Food	Tap beer	Fidelity card	% Av.
Cafeteria Cafe Le Klee	X	0	0	X	0	X	0	0	X	X	X	X	0	53,8%
Self-service BC	X	X	0	0	X	X	X	0	0	X	X	0	0	53,8%
Other BM	X	0	0	0	X	X	0	0	0	X	0	0	0	30,8%
Cafeteria ELA	X	0	0	0	X	X	0	0	X	0	X	0	0	38,5%
Cafeteria INM	X	0	0	X	X	X	0	0	X	X	X	0	0	53,8%
Cafeteria MX	X	X	0	0	X	X	X	0	X	X	X	0	0	61,5%
Other PH	X	0	0	0	0	X	0	0	0	X	0	0	0	23,1%
Cafeteria L'Arcadie	X	X	0	X	X	X	0	0	X	0	X	X	0	61,5%
Self-service L'Atlantide	X	X	0	X	X	X	0	0	X	0	X	0	0	53,8%
Restaurant Le Copernic	0	X	X	X	X	0	0	0	0	0	X	0	0	38,5%
Self-service Le Corbusier	0	X	0	0	X	0	X	0	0	0	X	0	0	30,8%
Cafeteria Le Giacometti	X	0	0	X	X	X	0	0	X	X	X	0	0	53,8%
Self-service Le Parmentier	0	X	0	X	X	0	X	X	0	0	X	0	0	46,2%
Self-service Le Vinci	0	X	0	X	X	0	X	0	0	0	X	0	0	38,5%
Self-service L'Esplanade	X	X	0	X	X	X	X	X	X	X	X	0	0	76,9%
Self-service L'Ornithorynque	0	X	0	0	X	0	0	0	0	0	X	0	0	23,1%
Caravan Pizza	0	X	0	X	X	0	0	X	0	0	X	0	X	46,2%
Caravan Kebab	0	X	0	X	X	0	0	X	0	0	X	0	X	46,2%
Cafeteria Satellite	X	0	0	X	X	X	0	0	X	0	X	X	X	61,5%
Self-service Le Hodler	0	X	0	0	0	X	X	0	0	X	X	0	0	38,5%
Restaurant Table de Vallotton	0	X	X	0	0	0	0	0	0	0	X	0	0	23,1%
% Availability	57,1%	66,7%	9,5%	57,1%	81%	61,9%	33,3%	19%	42,9%	42,9%	90,5%	14,3%	14,3%	

Table 6: Opening hours and availability of destinations

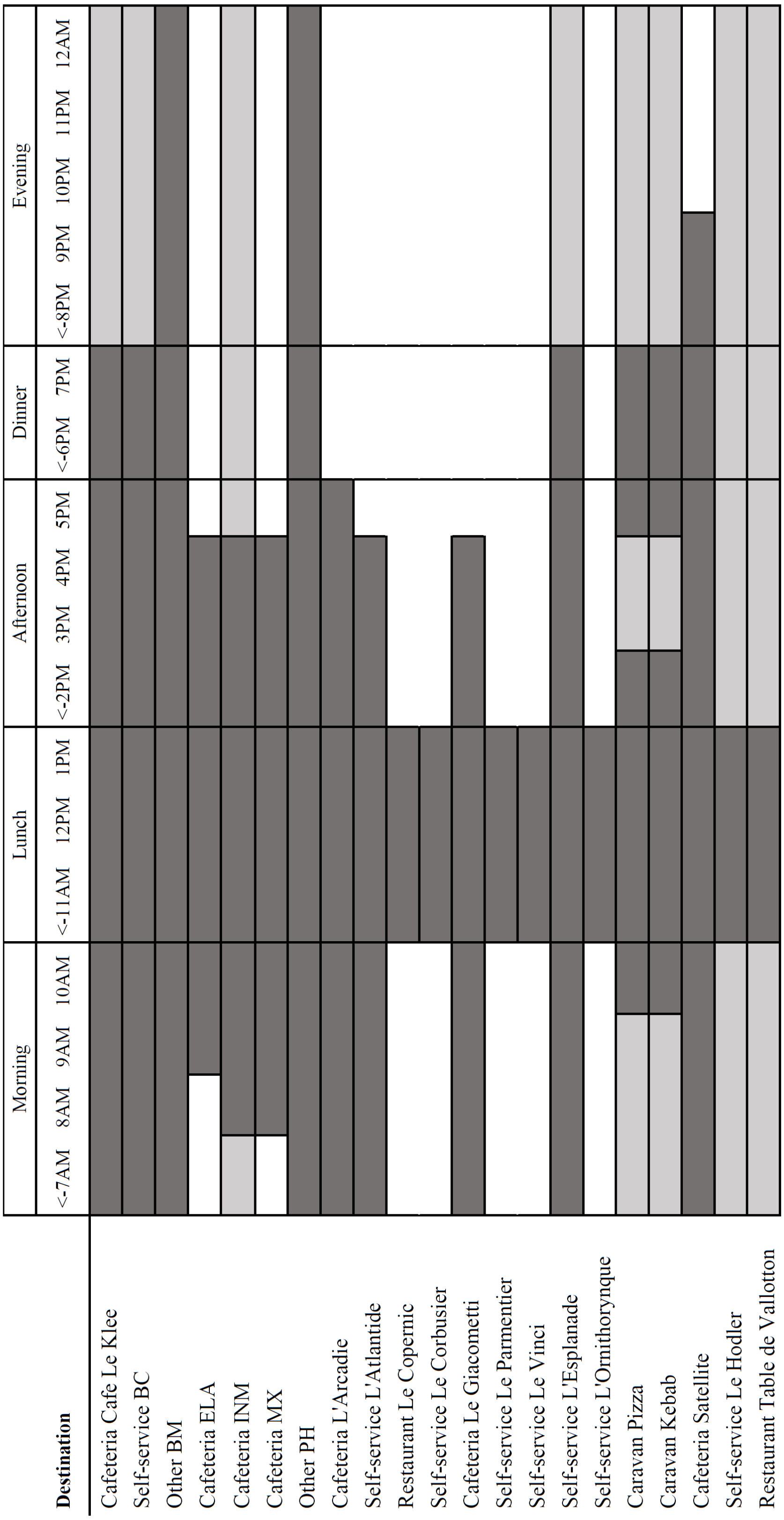


Table 7: Table of student prices

Destination	Cheapest	Most expensive
Cafeteria Cafe Le Klee	-	-
Self-service BC	7	12
Other BM	-	-
Cafeteria ELA	-	-
Cafeteria INM	-	-
Cafeteria MX	7	7
Other PH	-	-
Cafeteria L'Arcadie	9.9	9.9
Self-service L'Atlantide	9.8	9.8
Restaurant Le Copernic	18.5	27
Self-service Le Corbusier	7	11
Cafeteria Le Giacometti	-	-
Self-service Le Parmentier	7	12
Self-service Le Vinci	7	12
Self-service L'Esplanade	7	9
Self-service L'Ornithorynque	7.65	11.05
Caravan Pizza	8	12
Caravan Kebab	7	10
Cafeteria Satellite	-	-
Self-service Le Hodler	7	14
Restaurant Table de Vallotton	25	31

10 CHF. Restaurants are more expensive. Their cheapest meal is 18.5 CHF for Restaurant Le Copernic and 25 CHF for Restaurant La Table de Vallotton. Caravans sell Kebabs for 7 CHF and pizza (without fillings) for 8 CHF.

The only gap in these prices is between Restaurants and the rest of the destinations. Restaurants are mainly frequented by visitors, professors and employees. The maximum prices of self-services and caravans still are below the restaurants' cheapest menu's price. Employees must pay an additional amount of 1 CHF for self-services 7 CHF meals. There are no prices differences on the other menus. Students pay at least 7 CHF for a hot meal and personnel at least 8 CHF (except if they order a kebab).

The capacity (see Table 8) varies a lot between all the destinations. It is necessary to separate the inside capacity from the outside capacity as they are not available in winter or when it rains. Only caravans do not have an inside seating area. The inside capacity fluctuates between 25 and 320 seats. Furthermore some self-services offer up to 180 seats on their terrace. They are the

Table 8: Table of capacities

Destination	Inside	Outside
Cafeteria Cafe Le Klee	70	0
Self-service BC	82	119
Other BM	60	10
Cafeteria ELA	98	68
Cafeteria INM	20	14
Cafeteria MX	50	25
Other PH	15	0
Cafeteria L'Arcadie	60	100
Self-service L'Atlantide	125	50
Restaurant Le Copernic	105	50
Self-service Le Corbusier	228	100
Cafeteria Le Giacometti	90	30
Self-service Le Parmentier	320	52
Self-service Le Vinci	240	52
Self-service L'Esplanade	225	180
Self-service L'Ornithorynque	250	120
Caravan Pizza	0	15
Caravan Kebab	0	0
Cafeteria Satellite	200	30
Self-service Le Hodler	128	0
Restaurant Table de Vallotton	80	0

destinations with the highest capacities.

Since the campus is outside the city center, they need to accommodate all students and employees (about 12'000 people) for lunch with affordable menus and a large capacity. In 2012, the food service (restauration.epfl.ch) from EPFL made a survey (on both pen-and-paper and Internet supports) concerning the quality of the food on the campus. People were asked to grade the quality of food and to answer some questions about their habits and destination choice's factors. The results show that people choose their lunch destination because of determinants such as the proximity, the price, the meal itself (not taken into account in the model because it was not available) or the time they are willing to spend. These factors and the grades given to each destination are used in the model.

All the destinations got a grade superior to the mean (4). Furthermore, destinations with higher prices have a better evaluation which means that the price reflects the quality of the food and of the service. Small cafeterias also have good grades although they do not sell hot meals. According to the survey, these destinations have a good relation with customers.

Each destination has several additional services. They are summarized in Table 5. Most of the eating establishments have a terrace and sell food (of any kind) but only 67% offer a hot meal. The majority of them is selling coffee and proposes a workspace. Also, 57% of places are visible from the common sidewalk. Just 43% of the places sell sandwiches or have a Selecta (automatic vending machine). One third of the destinations are part of the "Green Fork" (a quality label) deal and only 14% of them sell tap beers or have a fidelity card.

Self-service L'Esplanade is the most complete catering destination. Nearly all services are available and it is located in the middle of the campus. Only table service, tap beer and fidelity card are missing. On the other hand, restaurants and self-service L'Ornithorynque only have half of all the presented services. There is not much heterogeneity between destinations of a same type.

4.2 WiFi traces on the campus

In their case study, Danalet *et al.* (2014) explain the nature of EPFL WiFi data (the data are available in Danalet (2015)). People working or studying on the campus can connect to the WiFi network (see Figure 2(a)) for free using their username. The authentication is made through WiFi Protected Access using a radius server. It processes accounting by allowing to associate a MAC address with the username.

In order to anonymize the data, the username and the MAC address are replaced by a individual and unique ID and a socio-economic attribute: the category of users. They are shown in Table 9.

Table 9: Category of traced individuals

Students			Employees
Section	Semester	Number of observations	Number of observations
Civil engineering	4	141	
Computer science	4	89	
Computer science	8	54	
Mathematics	2	109	
Life science engineering	2	152	
Physics	2	140	
Total		**685**	**1323**
Total number of observations: 2008			

Also, the number of observations per occupation is specified. Employees represent the majority of the total sample. The number of visits in eating establishments varies between 54 for students in master of computer science and 152 for life sciences bachelor students. Note that these activity episodes are performed by 192 different individuals.

4.3 Descriptive statistics on activity episodes

We compute some descriptive statistics about destination choice. The aim is to capture factors that reveal people's decision logic. Table 10 show that self-service L'Esplanade is the most visited eating establishment on the campus. It makes sense since this destination is strategically placed (in the middle of the school and surrounded with auditoriums). Then come the others self-services and cafeterias. They are followed by the caravans and the restaurants.

Catering facilities located in the Rolex Learning Center (RLC) do not have many visits. Danalet *et al.* (2014) explain that it is, in particular, due to the higher attractivity of the library (see Section 2.1 and Section 2.3). Indeed eating and working areas are (nearly) melted in the RLC and the seated capacity of the working area is about ten times bigger. Thus, activity episode sequences measured in the library are slightly biased due to the low precision of the attractivity measure in the library (the number of seats is used as an aggregate measure of occupation).

We present the catering destinations per period of the day (morning, lunch, afternoon, dinner, evening) in Table 11. Lunch time is the most attractive period in average. More than one third of the visits are made between 11 AM and 2 PM. Note that some destinations are less visited during this period. It is the case for self-service L'Atlantide, cafeteria Satellite, cafeteria MX or PH (others) which are destinations where it is common to take coffee breaks. Similar observations can be done in the afternoon. Destinations that are visited out of the lunch time all have a working space and/or additional services (e.g., coffee or tap beers). We consider now more specifically the lunch period. As students courses usually finish at 11 AM, 12 PM and 1 PM, one can expect several peaks in the demand. Destinations are aggregated by types (see Figure 6).

The lunch demand is separated into 3 peaks. There is one small peak between 11 AM and 12 PM because most of the self-services and restaurants only open at 11:30 AM. People reach a catering facility during this period to avoid queues and get a table more easily. The biggest peak is between 12 PM and 1 PM as the majority of students and people of the personnel lunch during this period. Then the third peak between 1 PM and 2 PM concerns students that finish their

Table 10: Observed choices per destination

Destination	Nb picks
Cafeteria Cafe Le Klee	4
Self-service BC	172
Other BM	47
Cafeteria ELA	145
Cafeteria INM	13
Cafeteria MX	86
Other PH	85
Cafeteria L'Arcadie	38
Self-service L'Atlantide	146
Restaurant Le Copernic	6
Self-service Le Corbusier	73
Cafeteria Le Giacometti	182
Self-service Le Parmentier	139
Self-service Le Vinci	2
Self-service L'Esplanade	448
Self-service L'Ornithorynque	102
Caravan Pizza	65
Caravan Kebab	68
Cafeteria Satellite	142
Self-service Le Hodler	36
Restaurant Table de Vallotton	9

Figure 6: Demand peaks during lunch hours (one hour periods)

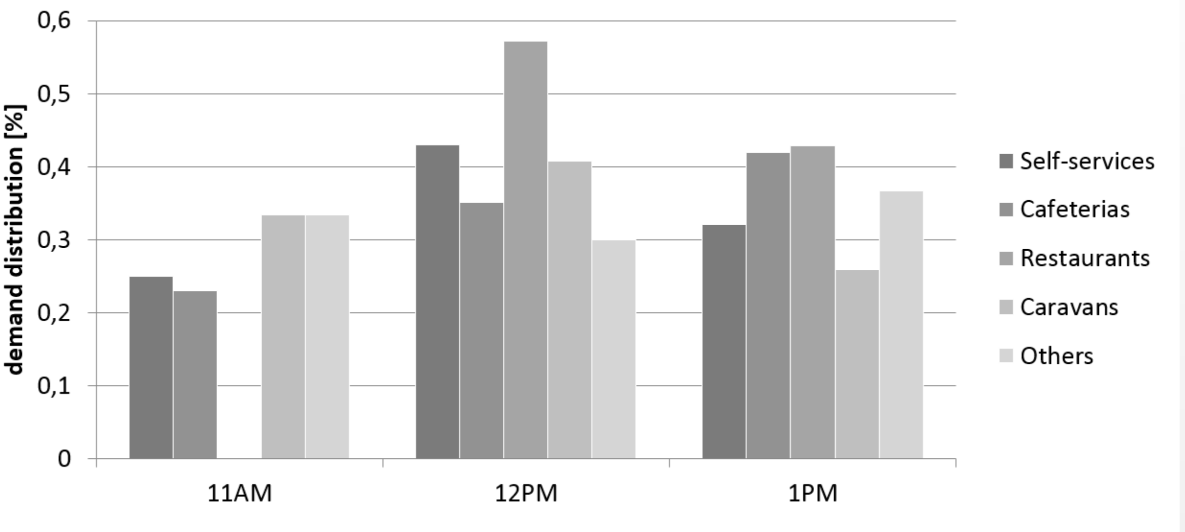
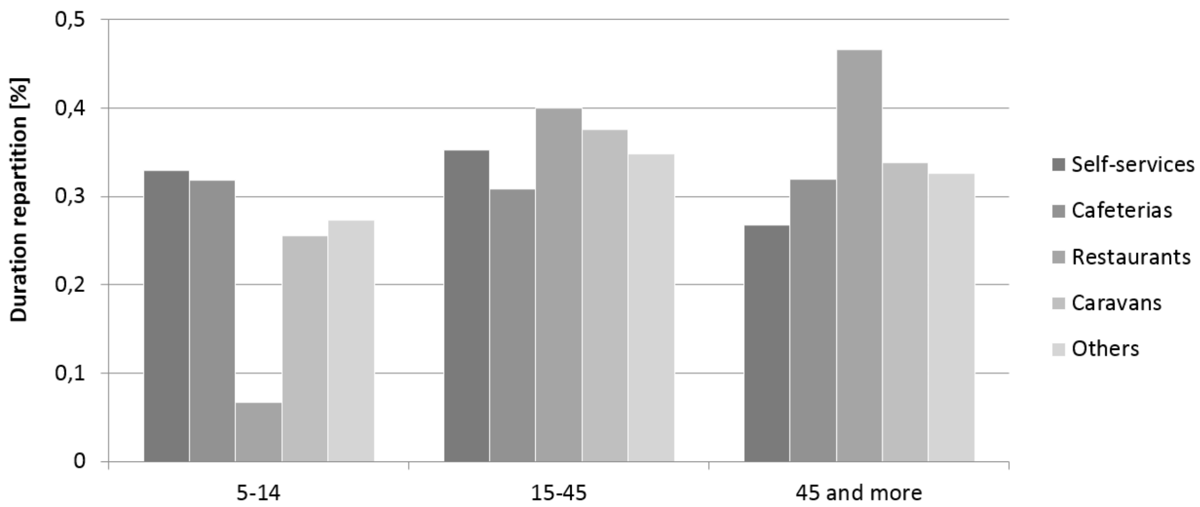


Table 11: Choices performed depending on the time of the day

	7AM-11AM	11AM-2PM	2PM-6PM	6PM-8PM	8PM-11PM	Total
Cafeteria Cafe Le Klee	1	1	2			4
Self-service BC	50	69	42	11		172
Other BM	11	14	16	5	1	47
Cafeteria ELA	37	55	53			145
Cafeteria INM	2	7	4			13
Cafeteria MX	38	22	26			86
Other PH	35	16	26	6	1	84
Cafeteria L'Arcadie	11	19	8			38
Self-service L'Atlantide	72	18	56			146
Restaurant Le Copernic		6				6
Self-service Le Corbusier		73				73
Cafeteria Le Giacometti	45	56	81			182
Self-service Le Parmentier		82		55	2	139
Self-service Le Vinci		2				2
Self-service L'Esplanade	95	148	162	44		449
Self-service L'Ornithorynque		102				102
Caravan Pizza	12	35	5	13		65
Caravan Kebab	11	19	24	14		68
Cafeteria Satellite	37	14	74	11	6	142
Self-service Le Hodler		36				36
Restaurant Table de Vallotton		8			1	9
Total	457	802	579	159	11	2008

Figure 7: Durations (in minutes) of observations depending on the type of destination



courses late and some employees. Cafeterias reach their maximum attendance during that period. It is possibly due to the fact that some people drink a coffee after their lunch. Also, individuals going to a restaurant do not move before 12 PM because their table is probably reserved.

The durations of activity episodes depending on the destination type is shown in Figure 7. They have been separated into three nearly equal categories. The first one reflects short visits (between 5 and 14 minutes). They can be interpreted as short breaks or as visits to buy a snack or a drink. The second one represents long breaks (between 15 and 45 minutes) to perform activities such as having lunch or spend an hour to rest. The last one stands for long activities. One can see on the figure that visiting a restaurant may take more than 45 minutes. Also, studying for a course or spending free time in a cafeteria can take more than an hour.

We take a look at the choices performed by some individuals (see Table 12). Civil engineering students have some habits. Indeed nearly all individuals have a preference for one or several destinations. This is also true for students from other sections and for employees. The repetition of the same catering destination choice over time for a same individual motivates to consider habits.

According to the literature review, the distance to walk has a significant impact in both route and destination choices. On the campus, if a student finishes his course at the extreme east (CE) and decides to lunch at the extreme southwest (BC) he has to walk about 1200 meters if he takes the shortest path (only 700 in Euclidean distance). By looking at Euclidean distances, students and employees have a preference for short distances but may change their habits sometimes. Indeed the average Euclidean distance covered is 110 meters (109 for students and 100 for employees).

Since a pedestrian network is available, realistic shortest path can be calculated between two destinations. They are more realistic than Euclidean distances. We compare the Euclidean and real distances covered to reach the chosen destination.The Euclidean distances reduces all the non-null paths (i.e. paths shorter than 20 meters are omitted) by 90% in average compared to paths calculated with a weighted shortest path algorithm (the complete algorithm is available in Danalet *et al.* (2014)). However, the standard deviation is high (around 100%). Using such algorithms takes into account the anisotropy of the place (Kim and Hespanha, 2003). Figures 8 and 9 represent the distribution of Euclidean and real distances walked by the individuals.

The trends are similar as before except that the distances to reach a catering destination are longer. In average, both students and employees walk 175 meters to visit an eating establishment. 5% of individuals cover a distance longer than 500 meters to reach their catering destination.

Table 12: Choices performed by civil engineering students: the bold numbers represent the most frequently chosen destination of one individual and the italic numbers, its first chosen destination.

Choices made by individuals over a 3 months period													
ID	BM	ELA	MX	PH	L'Atlantide	Le Giacometti	Le Parmentier	L'Esplanade	L'Ornithorynque	Pizza	Kebab	Satellite	Le Hodler
10000		1	1				1	**9**	*3*	1		5	
10001			1				*5*	4				**6**	4
10003					1			***3***		1		2	1
10004							7	**10**			1	1	1
10006								***4***					
10007						***7***							1
10008							***1***						
10009		1	1				1	***5***		2			1
10010	1						*8*	4			1	**12**	
10013	1							***2***					
10017		1		1									***2***

Figure 8: Distribution of Euclidean distances

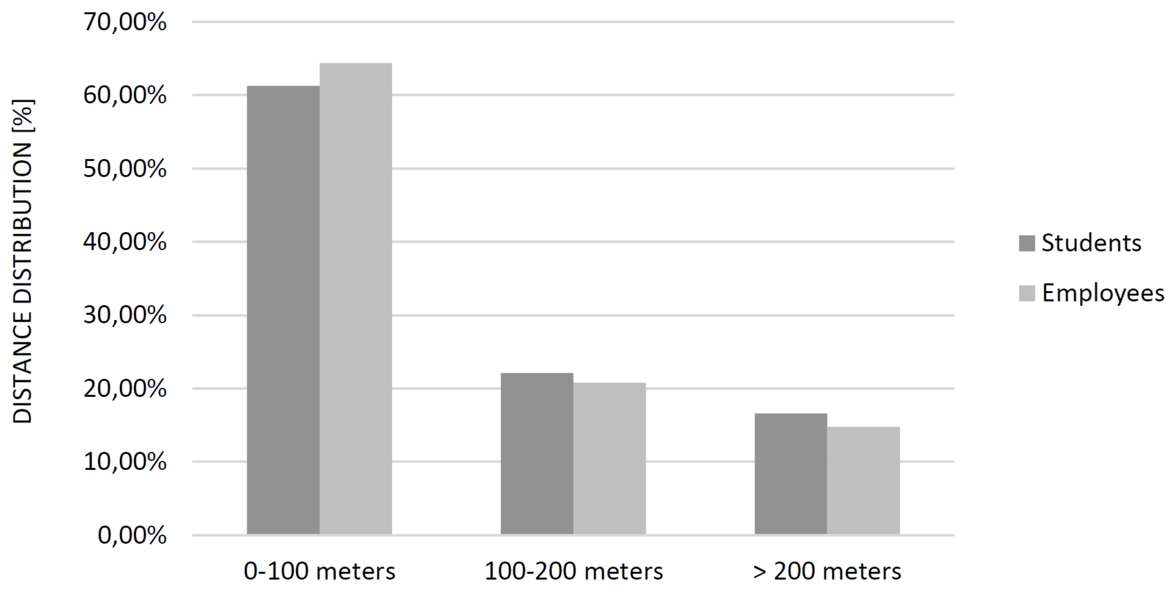
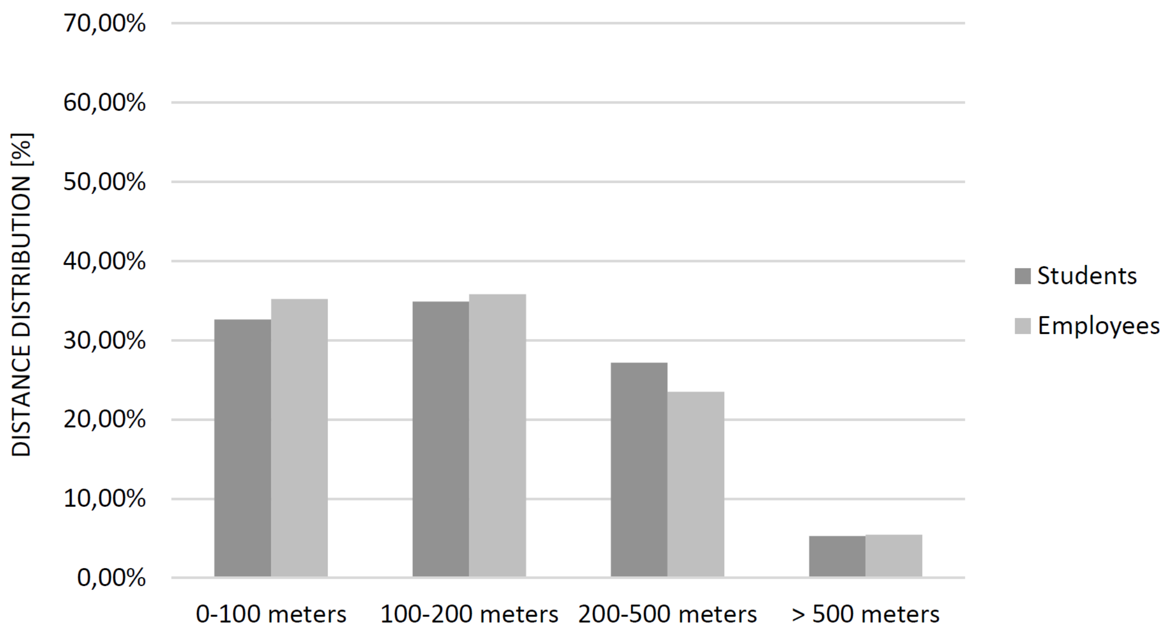


Figure 9: Distribution of real distances



This weighted shortest path algorithm does not provide distances between all points of the pedestrian network. This is due to the coding of the network (some doors need an access card and are assumed to be closed). This study uses a sample of 4,5 millions paths. About 10% of the possible distances are not calculated. However the distance to reach the chosen catering destination is always available.

We also consider the weather. We have daily data collected from Meteosuisse. During the case study, the average temperature was 15 Celsius degrees and two third of the days were shiny. It was a typical swiss spring.

Individual's choices are related to two important factors: the distance and the habits. Indeed people seem to prefer a catering destination close to their previous location and a destination they know well (they have already visited). Also, students and employees do not necessary visit a destination for the purpose of eating. More characteristics such as offering work places, coffee or tap beers may influence people's choice of catering destination.

4.4 Modelling of destination choice

4.4.1 Description of the models

Considering the points highlighted in Section 4.3, we develop linear in parameter Multinomial Logit Models. Before we describe these models in detail, one needs to define the dynamic variables. We decide to focus the dynamic on the lunch hours since it is the time of the day when catering destinations are the most frequented (40% of activity episodes) and when the purpose of the visit is obviously to have lunch.

One proposes that the previous choice is the previous catering destination visited by a same individual during the lunch period (11:30AM to 2PM). It means that the time interval between the activity episode sequences varies. It can be one day or weeks depending on the availability of information and the frequency of observations. Also, if this individual visits a catering destination out of the lunch hours, it is not considered as a previous choice.

Similarly, the first choice is the first catering destination ever visited by this same individual during the lunch period. Finally, we propose to use the most frequent choice to describe one individual average behavior among the past period (see Section 3.2.2). The most frequent choice stands for the most visited catering destination, during lunch time, before the actual choice. In the event of a tie, the most visited destination is randomly selected among the destinations with the same number of visits.

We consider three main variants as defined in Table 13 and we also add two submodels to the dynamic with agent effect correction's variant.

1. A static model with no previous choice considered at all where each observation is independent;
2. A dynamic strict exogenous model where the previous choice is considered but with the assumption that individuals have no memory on short observation periods (thus, the choice is based on exogenous factors within a short period, but on endogenous determinants over time);
3. Two dynamic models with panel data and agent effect correction. The previous choice is considered and two approaches are used to correct for agent effect issue using the principles described in Section 3.2.2:
   a) The first choice is considered to correct the agent effect;
   b) The first and most frequent choices are considered to correct the agent effect;

Table 13: Definition of static and dynamic models for the case study

Static model	Dynamic strict exogenous model	Dynamic models with agent effect correction	
		First choice	First and most frequent choices
$\rho = 0$	$\rho \neq 0$	$\rho \neq 0$	$\rho \neq 0$
$\alpha_0 = 0$	$\alpha_0 = 0$	$\alpha_0 \neq 0$	$\alpha_0 \neq 0$
$\alpha_2 = 0$	$\alpha_2 = 0$	$\alpha_2 = 0$	$\alpha_2 \neq 0$
$\sigma = 0$	$\sigma = 0$	$\sigma \neq 0$	$\sigma \neq 0$

There are 21 catering destinations on the EPFL campus, thus 21 utility functions. Table 14 clarifies the variables introduced in the models. We estimate the parameters for all four models using Python Biogeme software (Bierlaire, 2003; Bierlaire and Fetiarison, 2009). One shows a summary of the results in Table 15. The complete results (also containing Alternative Specific Constants (ASC) and σ values) and one typical utility function are available in Appendices A and B.

The values and signs of the short-term parameters are similar between all models (except for the price, see below). The static model is the *restricted* version of the dynamic strict exogenous model which also is the *restricted* version of both dynamic with agent effect correction models (see Table 13). The addition of the previous lunch's choice (at time $t-1$) decreases the t-test of the parameters related to the choice of the catering destination at time t. A similar effect is observed with the addition of both agent effect issue's corrections. As suggested by Pirotte

Table 14: Specification table: each variable has possibly 21 different values. The time periods are the following: morning hours are from 7 AM to 11:29 AM, lunch hours are from 11:30 AM to 2 PM, afternoon hours are from 2 PM to 6 PM, dinner hours are from 6 PM to 8 PM, night hours are from 8 PM to 11 PM. If the time period constraint is not respected, the variable is 0. Distances are measured in meters. A good weather stands for a dry day and at least a maximum daily temperature of 20 Celsius degrees. $\beta_{DISTANCE\_LUNCH_{TYPE}}$ and $\beta_{EVALUATION_{TYPE}}$ are type specific parameters (see Table 4). All the others are generic. Some variables are not available and take the value of -1 in the dataset. α_{FIRST_CHOICE}, α_{MOST_CHOSEN} and σ_d are only considered in the dynamic with agent effect correction models. Finally, $\rho_{PREVIOUS\_CHOICE}$ is null in the static model (see Table 13).

Parameter	Variable	Variable description	Time period
ASC_d	1	-	
$\beta_{DIST\_LUNCH_{TYPE}}$	*lunch_distance*	distance from the previous activity episode 0 otherwise	lunch
$\beta_{DIST\_MORNING}$	*morning_distance*	distance from the previous activity episode 0 otherwise	morning
$\beta_{DIST\_AFTERNOON}$	*afternoon_distance*	distance from the previous activity episode 0 otherwise	afternoon
$\beta_{NO\_DISTANCE\_AV}$	*distance_not_av*	1 if no distance is available 0 otherwise	
$\beta_{EVALUATION_{TYPE}}$	*evaluation_survey*	quality evaluation on a [1;6] scale 0 otherwise	lunch
$\beta_{PRICE\_STUDENT}$	*price_min_student*	price for the cheapest hot meal if student 0 otherwise	lunch
$\beta_{PRICE\_EMPLOYEE}$	*price_min_employee*	price for the cheapest hot meal if employee 0 otherwise	lunch
β_{TAP_BEER}	*beer_av*	1 if tap beer is available 0 otherwise	> lunch
β_{DINNER}	*dinner_av*	1 if dinner is available 0 otherwise	dinner
$\beta_{CAPACITY\_TERRACE}$	*capacity_terrace*	outside number of seats if the weather is good 0 otherwise	lunch
$\beta_{CAPACITY\_INSIDE}$	*capacity_inside*	inside number of seats 0 otherwise	lunch
$\rho_{PREVIOUS\_CHOICE}$	*previous_choice*	1 if the destination was the previous destination 0 otherwise	lunch
α_{FIRST_CHOICE}	*first_choice*	1 if the destination was the first destination 0 otherwise	lunch
α_{MOST_CHOSEN}	*most_freq_choice*	1 if the destination was the most frequented 0 otherwise	lunch
σ_d	1	-	

(1996) (see Section 3.2.2), long-term parameters measure the variability within individuals and thus alleviate the weight of short-term significants (i.e., reduce their *t*-tests).

Table 15: Table of estimates. Number of observations = 1867

	Static		Strict exo		First choice		First and most freq	
Parameters	Value	*t*-test	Value	*t*-test	Value	*t*-test	Value	*t*-test
$\beta_{DIST\_LUNCH\_CAFET}$	-0.00703	-16.69	-0.00633	-14.81	-0.00398	-8.03	-0.00367	-7.41
$\beta_{DIST\_LUNCH\_REST}$	-0.00276	-2.18	-0.00256	-2	-0.00191	-1.11	-0.00225	-1.42
$\beta_{DIST\_LUNCH\_SELF}$	-0.00646	-19.99	-0.00578	-17.37	-0.00413	-10.88	-0.00375	-9.99
$\beta_{DIST\_MORNING}$	-0.00379	-5.97	-0.00395	-6.17	-0.00286	-3.65	-0.0029	-3.58
$\beta_{DIST\_AFTERNOON}$	-0.000606	-1.31	-0.00103	-2.2	-0.000782	-1.28	-0.00106	-1.74
$\beta_{NO\_DISTANCE\_AV}$	-4.89	-13.84	-4.5	-12.93	-3.72	-8.64	-3.38	-8.31
$\beta_{EVALUATION\_CAFET}$	1.79	9.98	1.76	9.53	2.21	9.03	2.02	8.84
$\beta_{EVALUATION\_SELF}$	1.88	9.66	1.84	9.19	2.26	8.61	2.09	8.38
$\beta_{PRICE\_STUDENT}$	-0.0681	-2.07	-0.057	-1.7	-0.00686	-0.14	-0.00488	-0.09
$\beta_{PRICE\_EMPLOYEE}$	-0.00537	-0.18	0.000645	0.02	0.03	0.65	0.0618	1.18
β_{TAP_BEER}	0.669	3.62	0.601	3.24	0.806	3.14	0.766	3
β_{DINNER}	0.943	3.35	0.977	3.47	0.633	1.7	0.654	1.78
$\beta_{CAPACITY\_TERRACE}$	0.00162	1.84	0.00152	1.71	0.00212	2.03	0.0012	1.11
$\beta_{CAPACITY\_INSIDE}$	0.00277	1.29	0.00308	1.43	0.00405	1.54	0.00647	2.37
$\rho_{PREVIOUS\_CHOICE}$	0	0	1.78	17.31	0.424	3.14	-0.118	-0.76
$\alpha_{MOST\_FREQ\_CHOICE}$	0	0	0	0	0	0	1.79	14.27
α_{FIRST_CHOICE}	0	0	0	0	1.23	10.65	0.985	8.16
$\mathcal{L}(0)$	−5035.429		−5035.429		−5035.429		−5035.429	
$\mathcal{L}(\hat{\beta})$	−3238.926		−3101.563		−2428.28		−2335.75	
ρ^2	0.357		0.384		0.518		0.536	

In each model, the opening hours are considered as the availability of the destination (closed catering destinations cannot be visited even if pedestrians could technically reach them). We examine parameters' sign and *t*-test to describe the results of the models. Capacities (number of seats) of terraces and inside spaces have a positive parameter sign. It means that people have a preference for catering destinations with a bigger capacity. It makes sense since having an important number of places increases the chance to find a seat. Also, the destinations with terraces are more likely to be visited when the weather is sunny.

The distance from the previous activity episode is significant in the choice of an eating establishment. The sign is negative independently of the period of the day which represents the fact that people prefer a close destination. In the morning, the main activity that can be performed in a catering destination is having a coffee. In the afternoon, it can be several things like having a coffee, working or drinking a beer. The comparison between the parameters of both these time periods shows that individuals prefer to walk less in the morning than in the afternoon. A possible explanation is that coffee is available nearly everywhere but descent workspaces or tap beers are much rarer so people accept to travel longer. Another possible reason is that people tend to have a coffee next to their following activity episode (instead of next to the previous activity episode). This has not been explored yet. Other possible explanations include looking for a sunny terrace or for a place selling ice creams, since the data collection took place in the beginning of the summer.

At lunch time, the distance covered from the last activity episode depends on the type of destination chosen. For example, individuals are more likely to walk when going to a restaurant as the choice set is small for this destination type (only two restaurants on campus). On the other hand, students and employees prefer a near self-service or cafeteria compared to a far one. The fact that this kind of destinations is distributed everywhere on the campus can be an explanation.

Note that individuals visiting a caravan or another catering destinations (PH and BM) are less sensitive to distance (i.e., the parameters are not significant). It is not a surprise since those places have their own distinctive offers. People accept to cover more distance if they want a specific type of meal. The parameter accounting for the non-availability of distances is negative as well. It means that catering destinations that are the least connected to the network are less likely to be visited.

The minimum price for a hot meal is not significant in dynamic models for both students and employees but we decide to keep it anyway because we expected it to be significant. As explained in Section 3.2.2, it may be the fact that the price is considered as a short-term determinant in our models. Moreover, prices have low variability on the campus; this also explains why cost is not significant in our models. We give an explanation to these parameters anyway. Price has a negative sign for students. It makes sense as they are not willing to spend 25 CHF to go to the restaurants and prefer catering destinations with 7 CHF meals or caravans. Employees look for eating establishments with higher prices because the price is connected with the food quality. Also, working people earn a salary and bills can be attributed to the company expenses.

Evaluations have a positive sign for both cafeterias and self-services. It means that individuals choose a cafeteria or a self-service as a destination depending on the average quality of the offer. Evaluations are not significant for caravans and restaurants. Eating establishment that offer dinner are more likely to be visited between 6 PM and 8 PM.

The availability of tap beer after midday increases the utility of a catering destination. Indeed, some individuals may want to relax more than work in the afternoon and the evening.

Only three destinations offer tap beers on the campus; the well-known Satellite bar, the cafeteria of the Rolex Learning Center (Klee) and cafeteria L'Arcadie.

Habits are significant in all dynamic models. The previous choice made by people at lunch time has a parameter with a positive sign. It means that students and employees have some habits when choosing for an eating destination. As an example, if the previous time they ate on the campus for lunch, they chose to eat at self-service Le Corbusier, they are more likely to pick this alternative again. Also, the correcting terms have a positive sign and a strong t-test. However, the previous choice becomes non-significant with the double agent effect correction which may mean that average behavior among the observation period is stronger as the previous choice (also short-term, long-term effects as explained above). We explore this topic in future research.

The most robust explanatory variables are the distances and the previous choice (except for the model with the double agent effect correction). Prices or services availability are less robust determinants. Probably because prices are relatively cheap and uniform (except for restaurants) and because a same type of catering destinations usually proposes the same services in every destinations. Also, both corrections of agent effect seem to improve the models. We verify this impression in the next chapter (see Section 4.4.2).

4.4.2 Comparison of the models

All four models shown in Section 4.4.1 have close values of parameters. We compare these models to find which one fits the data the best. A log-likelihood ratio test is performed. We can use this test because the models are nested. The static model is the *restricted* model of the dynamic strict exogenous model which is the *restricted* model of both dynamic with panel data and agent effect issue correction models. Also, the model considering the first choice is the *restricted* version of the one accounting for both the first and most frequent choices. The statistic

$$-2(\mathcal{L}(\hat{\beta_R}) - \mathcal{L}(\hat{\beta_U})) \tag{14}$$

is χ^2 distributed, with degrees of freedom equal to

$$K_U - K_R \tag{15}$$

with K, the number of parameters of each model ($Unrestricted$ and $Restricted$). If the result of Equation (14) is bigger than the percentile of the chi square distribution, then we can reject the null hypothesis (at a chosen level of confidence) and the *unrestricted* model is preferred to the *restricted* one. We perform the log-likelihood ratio test on each model according to the specification made in Table 13. Table 16 presents the results. Both models, accounting for panel nature of data and correcting agent effect, are statistically better (with more than 95% confidence) than the second one which is statistically better than the static one as well.

Table 16: Table of likelihood ratio test. DSE stands for Dynamic Strict Exogenous, DAEC stands for Dynamic with Agent Effect Correction.

	Static	DSE	DAEC first choice	DAEC first and most frequent choices
$\mathcal{L}(\hat{\beta})$	-3238.926	-3101.563	-2428.280	-2335.750
Nb of parameters	34	35	56	57

Loglikelihood ratio test

Static vs DSE:$-2(-3238.9 + 3101.5) = 275 > 3.84$

DSE vs AEC (first choice):$-2(-3101.5 + 2428.2) = 1347 > 33.92$

DAEC (first choice) vs DAEC (first and most frequent choices):$-2(-2428.2 + 2335.7) = 185 > 3.84$

4.5 Validation

We perform an aggregated validation on our models. The dataset is separated into two subsamples: one to calibrate the models, the second one to simulate the future destination choices and compare the output of the models with the actual choices. The first sample represents the past choices of individuals and the second sample contains their most recent observation. Basically we use people's past choices to estimate the models (first sample) and we forecast their most recent observation of a destination to have lunch (second sample). People with only one observation are removed because they do not fulfill the dynamic conditions (thus the dataset is not exactly the same that the one used for estimation in Table 15. We keep 1379 observations to calibrate the models and 121 to simulate future choices). An example of sample separation is given on Figure 10.

Dynamic models with agent effect correction are simulated as Mixed Logit Models because they have two error terms and one of them is normally distributed (see Section 3.2.2) whereas both static and dynamic strict exogenous models only have a single error term and are thus simulated as Multinomial Logit Models. Table 17 summarizes the results.

The trends are similar between observations and estimated choices. These results are positive since they show that even a basic static model simulates reasonable forecasting on a small validation sample. The errors mainly come from the estimation of self-services. The number of

Figure 10: Separation of the total sample for calibration and simulation: the black dots represent the activity episodes used for calibration whereas gray dots represent activity episodes used for simulation.

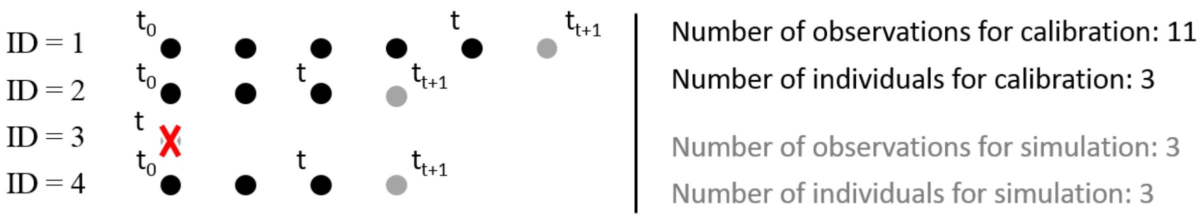


Table 17: Validation of the models. Observed and estimated choices performed by 121 individuals on their last activity episode

Observed			Predicted							
			Static		Strict exo		First choice		First and most freq	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
Cafeteria Cafe Le Klee	1	0.6%	0	0.1%	0	0.1%	1	1%	2	1.9%
Self-service BC	11	6.3%	7	6.2%	6	5%	6	5.1%	6	5%
Other BM	1	0.6%	3	2.3%	2	1.9%	2	1.2%	2	2.1%
Cafeteria ELA	9	5.1%	6	4.9%	6	4.7%	6	4.6%	6	4.8%
Cafeteria INM	0	0%	1	0.7%	1	0.6%	1	1%	1	1.1%
Cafeteria MX	4	2.3%	5	3.8%	4	3.7%	4	3.1%	4	3%
Other PH	3	1.7%	4	3%	3	2.8%	2	1.9%	2	1.9%
Cafeteria L'Arcadie	4	2.3%	1	0.9%	1	0.9%	1	0.9%	2	1.5%
Self-service L'Atlantide	3	1.7%	7	5.5%	6	5%	5	4.1%	3	2.5%
Restaurant Le Copernic	1	0.6%	1	0.9%	2	1.3%	2	1.4%	2	1.3%
Self-service Le Corbusier	5	2.9%	13	10.4%	10	8.3%	11	9%	11	8.7%
Cafeteria Le Giacometti	10	5.7%	8	6.6%	9	7.5%	9	7%	10	8.1%
Self-service Le Parmentier	10	5.7%	13	11%	14	11.5%	16	13.5%	12	9.9%
Self-service Le Vinci	1	0.6%	0	0.2%	0	0.2%	0	0.2%	0	0.2%
Self-service L'Esplanade	21	12%	18	14.6%	19	15.3%	18	14.6%	19	15.8%
Self-service L'Ornithorynque	15	8.6%	16	13.5%	18	14.8%	18	14.9%	18	14.6%
Caravan Pizza	7	4%	3	2.8%	4	3%	3	2.7%	4	3.6%
Caravan Kebab	4	2.3%	3	2.7%	3	2.7%	4	3.6%	3	2.3%
Cafeteria Satellite	3	1.7%	4	3.5%	4	3.6%	4	3.4%	6	4.8%
Self-service Le Hodler	7	4%	7	5.5%	7	6%	7	5.8%	7	5.8%
Restaurant Table de Vallotton	1	0.6%	1	1%	1	0.9%	1	1%	1	1.1%

destination type's choices (e.g., Self-service, cafeteria...) is accurate for each model. It means that our models are good at forecasting the destination type choice but then are less accurate to select a specific destination. The reason could be that the variability of services' availability for destinations of a same type is narrow. Also, the fact that catering destinations are relatively evenly distributed on the campus (see Figure 5) does that individuals usually have equidistant possible destinations of a same type. This latter point is especially true for self-services.

We expected the accuracy to be better for both dynamic models with panel data and agent effect correction as they are statistically better than both other models (see Section 4.4.2) but according to Table 17, it seems that it is not the case. We propose to use a least squares' method to measure objectively the accuracy of each model:

$$S_m = \sum_{d=1}^{21} (O_d - E_{d,m})^2 \tag{16}$$

where O_d is the percentage of Observations for destination d and $E_{d,m}$ is the expected number of visitors based on the choice probabilities for destination d and model m. The best model is the one that minimizes the least squares' method (S_m). The results are shown on Table 18.

Table 18: Least squares' method. DAEC stands for Dynamic with Agent Effect Correction

Static	Strict exogenous	DAEC first choice	DAEC first and most frequent choices
$S_{static} = 104$	$S_{strict\_exogenous} = 87$	$S_{first\_choice} = 112$	$S_{first\_and\_most\_frequent\_choices} = 75$

The gap between each model is small. The one that minimizes the difference between observations and estimated choices is the dynamic with both agent effect corrections (first and most frequent choices) which is also the one that fits the data the best (Section 4.4.2). The static and strict exogenous models show accurate forecasting as well. The "worst" model is the dynamic with only one agent effect correction. We think that the first choice may not be very representative of individuals' habits on short periods. Also, we emphasize some limitations that require further research:

1. The fact that one only considers dynamic during lunch hours does that some individuals do not have any prior and are thus removed from the dataset. It represents about one fourth of the total sample;
2. The first and most frequent choices are based on only 3 months of observations. Their efficiency at correcting agent effect may not be good (e.g., the fact that the first or most frequent choices fit the actual choice may be due to pure luck);
3. We suggest that lunch and out of lunch hours' observations are studied separately to have a clearly defined dynamic;

Despite of these highlighted problems, one considers that our models are successfully validated.

5 Conclusion

We propose a framework to model pedestrians destination choice from WiFi localization. One uses Danalet *et al.* (2014) method to generate candidates of activity episode sequences from WiFi measurements, locations of activities and prior information.

This paper describes a full methodology to develop a pedestrian destination choice model in a multi-modal facility from activity episode sequences. One activity type is selected and all the possible destinations to perform this activity are considered. The attributes to explain destination choice have been collected. These attributes are either sequence specific (e.g., ID, category, day), activity episode specific (e.g., location, start and end times) or destination specific (e.g., opening hours, prices).

These attributes are associated with additional determinants (e.g., habits, distance). Panel nature of data and how to correct agent effect issue are accounted for using Wooldridge (2002) approach. Three types of models are developed: a static model, a dynamic strict exogenous model and two dynamic with panel data and agent effect correction models (thus, a total of four models). They reveal the importance of past choices (the routine of an individual). We emphasizes that taking into account the previous choice and correcting for agent effect issue contribute to improve significantly the fit of a destination choice model for pedestrians but that a static model already performs accurate forecasts.

We present a case study on the EPFL campus where we generate, comment and validate our methodology. Eating is considered as the activity type. 21 eating establishments represent the destination choices for this activity type. These destinations are decomposed into types (i.e., cafeteria, self-service, restaurant, caravan or other) depending on the services they propose.

Our models reveal three major points. First, individuals prefer destinations close to their previous activity. It means that they reduce the distance to walk for reaching an eating establishment. This is especially observed when people need to chose for a destination to have coffee in the morning and lunch in a cafeteria or a self-service. Second, the choice of a catering destination at time t is connected to the previous catering choice performed at time $t-1$. Indeed, if one eating establishment has been visited before it is more likely to be chosen again. The results show that accounting for panel nature of data and correcting agent effect lead to accurate estimations. Third, ancillary services (e.g., selling sandwiches, having a fidelity card...) do not seem to influence people's choice because destinations of a same type all propose more or less the same range of services.

In future works, one should improve the methodology. It has revealed some limitations about the detection of points of interest (see Section 2.3). Also, as a first approach, we assumed in the case study that time interval between consecutive choices was undefined (mainly because of the nature of data). Time between activity episode sequences should be clearly defined to measure the impact of time. The choice of a destination performed 2 weeks or 2 years before the actual choice may not have the same impact (see Section 4.4.1). We should also consider more than one candidate of activity episode sequences since the input is generated considering a Bayesian approach (Danalet *et al.*, 2014). It involves using more than one candidate per person and including a measurement equation.

Destination choice models usually consider Space Syntax parameters (see Section 2.2.1). We did not implement such determinants in our models but we suggest that they may be significant. Furthermore, we should consider applying the methodology to develop a destination choice model to a multi-modal facility context. Railway stations, airports, stores or public buildings are as much new opportunities to understand and model pedestrian destination choice. Forecasting with the estimated models (e.g., what happens if a new destination opens?) may be explored as well.

6 References

Bekhor, S., Y. Cohen and C. Solomon (2013) Evaluating long-distance travel patterns in Israel by tracking cellular phone positions, *Journal of Advanced Transportation*, **47** (4) 435–446, June 2013, ISSN 01976729.

Bierlaire, M. (2003) BIOGEME: a free package for the estimation of discrete choice models, paper presented at the *Proceedings of the 3rd Swiss Transportation Research Conference*, Monte Verità, Ascona, Switzerland.

Bierlaire, M. (2014) Discrete panel data, *Mathematical modeling of behavior - EPFL course*.

Bierlaire, M. and M. Fetiarison (2009) Estimation of discrete choice models: extending biogeme, *Swiss Transport Research Conference (STRC)*, (September).

Bierlaire, M. and E. Frejinger (2008) Route choice modeling with network-free data, *Transportation Research Part C*, **16** (2) 187–198, April 2008, ISSN 0968090X.

Bierlaire, M. and T. Robin (2009) Pedestrians Choices, paper presented at the *Pedestrian Behavior. Models, Data Collection and Applications*, 1–26, ISBN 978-1-84855-750-5.

CFF (2011) La gare, une VIP: au Coeur des CFF, `http://www.cff.ch/content/dam/sbb/fr/pdf/fr_sbb-konzern/fr_ueber-die-sbb/fr_organisation/Nos+gares.pdf`.

CFF (2014) Léman 2030, `http://www.cff.ch/groupe/entreprise/projets/extension-du-reseau-ferroviaire/leman-2030/apercu/leman.html`.

Cisco (2011) Cisco MSE API Specification Guide - Location Service, Release 7.1, *Technical Report*, Cisco, April 2011.

Dalumpines, R. (2014) GIS-based episode reconstruction using gps data for activity analysis and route choice modeling, Thesis, McMaster.

Danalet, A. (2015) A Bayesian Approach to Detect Pedestrian Destination-Sequences from WiFi Signatures: Data, `http://zenodo.org/record/15798`.

Danalet, A. and M. Bierlaire (2015) Importance sampling for activity path choice, paper presented at the *15th Swiss Transport Research Conference (STRC)*, Monte Verità, Ascona, Switzerland.

Danalet, A., B. Farooq and M. Bierlaire (2014) A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures, *Transportation Research Part C: Emerging Technologies*, **44**, 146–170.

Goetz, M. and A. Zipf (2011) Formal definition of a user-adaptive and length-optimal routing graph for complex indoor environments, *Geo-spatial Information Science*, **14** (2) 119–128, January 2011, ISSN 1009-5020.

Helbing, D. (1997) Traffic Dynamics: New Physical Modeling Concepts (In German), *Springer-Verlag*.

Hillier, B. (2005) The art of place and the science of space, 24–34, ISSN 0956-9758.

Hoogendoorn, S. P. and P. H. L. Bovy (2004) Pedestrian route-choice and activity scheduling theory and models, *Transportation Research Part B: Methodological*, **38**, 169–190, ISSN 01912615.

Hoogendoorn, S. P., P. H. L. Bovy and W. Daamen (2002) Microscopic pedestrian wayfinding and dynamics modelling, in M. Schreckenberg and S. D. Sharma (eds.) *Pedestrian and Evacuation Dynamics*, 123–155, Springer.

Kalakou, S. and F. Moura (2014a) Bridging the Gap in Planning Indoor Pedestrian Facilities, *Transport Reviews*, **34** (May 2015) 474–500, ISSN 0144-1647.

Kalakou, S. and F. Moura (2014b) Effects of terminal planning on passenger choices, paper presented at the *14th Swiss Transport Research Conference (STRC)*, Monte Verità, Ascona, Switzerland.

Kim, J. and J. Hespanha (2003) Discrete approximations to continuous shortest-path: application to minimum-risk path planning for groups of UAVs, *42nd IEEE International Conference on Decision and Control (IEEE Cat. No.03CH37475)*, **2** (0), ISSN 0191-2216.

Liu, X. (2013) Activity-based pedestrian behavior simulation inside intermodal facilities, Ph.D. Thesis, Mississippi State University.

Liu, X., J. M. Usher and L. Strawderman (2014) An analysis of activity scheduling behavior of airport travelers, *Computers and Industrial Engineering*, **74**, 208–218, ISSN 03608352.

Miller, H. J. (2010) Measuring Space-Time Accessibility Benefits within Transportation Networks: Basic Theory and Computational Procedures, *Geographical Analysis*, **31** (1) 1–26, September 2010, ISSN 00167363.

Nandakumar, R., K. K. Chintalapudi and V. N. Padmanabhan (2012) Centaur: locating devices in an office environment, *Proc.18th Ann. Int. Conf. Mob. Comput. Netw. (MobiCom'12)*, 1–12.

OFS (2014) Prestations de transport de personnes, `http://www.bfs.admin.ch/bfs/portal/fr/index/themen/11/04/blank/01.html`.

Pettersson, P. (2011) Passenger waiting strategies on railway platforms - Effects of information and platform facilities, Master thesis, KTH.

Pirotte, A. (1996) Estimation de relations de long terme sur données de panel : nouveaux résultats, *Économie & Prévision*, **126**, 143–161, ISSN 0249-4744.

Sen, S., J. Lee, K.-h. Kim and P. Congdon (2013) Avoiding Multipath to Revive Inbuilding WiFi Localization, *Proc. of MobiSys*, 249.

Ton, D. (2014) NAVISTATION: a study into the route and activity location choice behaviour of departing pedestrians in train stations, Master thesis, Delft University of Technology.

Ueno, J., A. Nakazawa and T. Kishimoto (2009) An Analysis of Pedestrian Movement in Multilevel Complex by Space Syntax Theory - In the Case of Shibuya Station, *Proceedings of the 7th International Space Syntax Symposium*, 1–12.

Wooldridge, J. M. (2002) Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, 44, ISSN 1753-9196.

Worldbank (2014) Air transport, passengers carried, `http://data.worldbank.org/indicator/IS.AIR.PSGR/countries?display=graph`.

Zhang, L., Y. Zhuang and X. Dai (2012) A configurational study of pedestrian flows in multilevel commercial space. Case study Shanghai, *Eighth International Space Syntax Symposium*, 8044:1–16.

A Strict utility functions

$$
\begin{aligned}
V_d = ASC_d &+ \beta_{DISTANCE\_LUNCH_{TYPE}} * lunch\_distance\_d \\
&+ \beta_{DISTANCE\_MORNING} * morning\_distance\_d \\
&+ \beta_{DISTANCE\_AFTERNOON} * afternoon\_distance\_d \\
&+ \beta_{NO\_DISTANCE\_AV} * distance\_not\_av\_d \\
&+ \beta_{EVALUATION_{TYPE}} * evaluation\_survey\_2013\_d \\
&+ \beta_{PRICE\_STUDENT} * lunch\_price\_min\_student\_d \\
&+ \beta_{PRICE\_EMPLOYEE} * lunch\_price\_min\_employee\_d \\
&+ \beta_{TAP\_BEER\_AFTER\_LUNCH} * beer\_after\_lunch\_filter\_d \\
&+ \beta_{DINNER} * dinner\_filter\_d \\
&+ \beta_{METEO\_TERRACE} * meteo\_terrace\_filter\_d \\
&+ \beta_{CAPACITY\_INSIDE} * cap\_inside\_filter\_d \\
&+ \rho_{PREVIOUS\_CHOICE} * previous\_choice\_filter\_d \\
&+ \alpha_{MOST\_FREQUENT\_CHOICE} * most\_frequent\_choice\_filter\_d \\
&+ \alpha_{FIRST\_CHOICE} * first\_choice\_filter\_d + \mathcal{N}(0, \sigma_d^2)
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
V_{Espla} = ASC_{Espla} &+ \beta_{DISTANCE\_LUNCH_{SELF\_SERVICE}} * lunch\_distance\_Esplanade \\
&+ \beta_{DISTANCE\_MORNING} * morning\_distance\_Esplanade \\
&+ \beta_{DISTANCE\_AFTERNOON} * afternoon\_distance\_Esplanade \\
&+ \beta_{NO\_DISTANCE\_AV} * distance\_not\_av\_Esplanade \\
&+ \beta_{EVALUATION_{SELF\_SERVICE}} * evaluation\_survey\_2013\_Esplanade \\
&+ \beta_{PRICE\_STUDENT} * lunch\_price\_min\_student\_Esplanade \\
&+ \beta_{PRICE\_EMPLOYEE} * lunch\_price\_min\_employee\_Esplanade \\
&+ \beta_{TAP\_BEER\_AFTER\_LUNCH} * beer\_after\_lunch\_filter\_Esplanade \\
&+ \beta_{DINNER} * dinner\_filter\_Esplanade \\
&+ \beta_{METEO\_TERRACE} * meteo\_terrace\_filter\_Esplanade \\
&+ \beta_{CAPACITY\_INSIDE} * cap\_inside\_filter\_Esplanade \\
&+ \rho_{PREVIOUS\_CHOICE} * previous\_choice\_filter\_Esplanade \\
&+ \alpha_{MOST\_FREQUENT\_CHOICE} * most\_frequent\_choice\_filter\_Esplanade \\
&+ \alpha_{FIRST\_CHOICE} * first\_choice\_filter\_Esplanade + \mathcal{N}(0, \sigma_{Esplanade}^2)
\end{aligned}
\tag{18}
$$

B Detailed results

Table 19: Static model

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC_ARC	-1.47	0.318	-4.60	0.00
2	ASC_ATL	-0.966	0.325	-2.97	0.00
3	ASC_BC	-0.369	0.397	-0.93	0.35
4	ASC_BM	0.666	0.324	2.06	0.04
5	ASC_COP	1.03	0.590	1.74	0.08
6	ASC_COR	-0.235	0.141	-1.67	0.10
7	ASC_ELA	-1.33	0.435	-3.06	0.00
8	ASC_GIA	0.204	0.392	0.52	0.60
9	ASC_HOD	-0.130	0.393	-0.33	0.74
10	ASC_INM	-2.92	0.608	-4.81	0.00
11	ASC_KEB	0.770	0.247	3.11	0.00
12	ASC_KLE	-3.34	0.647	-5.17	0.00
13	ASC_MX	-1.34	0.351	-3.81	0.00
14	ASC_ORN	-0.797	0.134	-5.93	0.00
15	ASC_PAR	-0.381	0.268	-1.42	0.15
16	ASC_PH	1.36	0.323	4.23	0.00
17	ASC_PIZ	0.980	0.237	4.14	0.00
18	ASC_SAT	-1.32	0.473	-2.79	0.01
19	ASC_VAL	1.49	0.734	2.02	0.04
20	ASC_VIN	-4.02	0.715	-5.62	0.00
21	BETA_CAPACITY_INSIDE	0.00277	0.00257	1.08	0.28
22	BETA_DINNER	0.943	0.289	3.26	0.00
23	BETA_DISTANCE_AFTERNOON	-0.000606	0.000545	-1.11	0.27
24	BETA_DISTANCE_LUNCH_CAF	-0.00703	0.000506	-13.88	0.00
25	BETA_DISTANCE_LUNCH_REST	-0.00276	0.00128	-2.16	0.03
26	BETA_DISTANCE_LUNCH_SELF	-0.00646	0.000418	-15.45	0.00
27	BETA_DISTANCE_MORNING	-0.00379	0.000826	-4.59	0.00
28	BETA_EVALUATION_CAFET	1.79	0.0929	19.26	0.00
29	BETA_EVALUATION_SELF	1.88	0.125	15.04	0.00
30	BETA_METEO_TERRACE	0.00162	0.000878	1.85	0.07
31	BETA_NO_DISTANCE_AV	-4.89	0.420	-11.66	0.00
32	BETA_PRICE_EMPLOYEE	-0.00537	0.0333	-0.16	0.87
33	BETA_PRICE_STUDENT	-0.0681	0.0369	-1.85	0.06
34	BETA_TAP_BEER_AFTER_LUNCH	0.669	0.180	3.71	0.00

Summary statistics

Number of observations = 1867

Number of estimated parameters = 34

$$\mathcal{L}(\beta_0) = -5035.429$$
$$\mathcal{L}(\hat{\beta}) = -3238.926$$
$$-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] = 3593.005$$
$$\rho^2 = 0.357$$
$$\bar{\rho}^2 = 0.350$$

Table 20: Dynamic strict exogenous model

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC_ARC	-1.44	0.316	-4.55	0.00
2	ASC_ATL	-0.882	0.330	-2.67	0.01
3	ASC_BC	-0.334	0.402	-0.83	0.41
4	ASC_BM	0.760	0.333	2.28	0.02
5	ASC_COP	1.02	0.579	1.77	0.08
6	ASC_COR	-0.221	0.155	-1.43	0.15
7	ASC_ELA	-1.22	0.438	-2.77	0.01
8	ASC_GIA	0.297	0.399	0.74	0.46
9	ASC_HOD	-0.0223	0.396	-0.06	0.96
10	ASC_INM	-2.74	0.606	-4.52	0.00
11	ASC_KEB	0.867	0.253	3.43	0.00
12	ASC_KLE	-3.12	0.648	-4.81	0.00
13	ASC_MX	-1.28	0.348	-3.67	0.00
14	ASC_ORN	-0.851	0.146	-5.82	0.00
15	ASC_PAR	-0.399	0.271	-1.47	0.14
16	ASC_PH	1.53	0.332	4.60	0.00
17	ASC_PIZ	0.974	0.234	4.16	0.00
18	ASC_SAT	-1.20	0.484	-2.48	0.01
19	ASC_VAL	1.59	0.750	2.12	0.03
20	ASC_VIN	-3.74	0.716	-5.22	0.00
21	BETA_CAPACITY_INSIDE	0.00308	0.00259	1.19	0.24
22	BETA_DINNER	0.977	0.287	3.41	0.00
23	BETA_DISTANCE_AFTERNOON	-0.00103	0.000549	-1.87	0.06
24	BETA_DISTANCE_LUNCH_CAF	-0.00633	0.000512	-12.35	0.00
25	BETA_DISTANCE_LUNCH_REST	-0.00256	0.00125	-2.05	0.04
26	BETA_DISTANCE_LUNCH_SELF	-0.00578	0.000430	-13.44	0.00
27	BETA_DISTANCE_MORNING	-0.00395	0.000837	-4.72	0.00
28	BETA_EVALUATION_CAFET	1.76	0.0938	18.78	0.00
29	BETA_EVALUATION_SELF	1.84	0.126	14.54	0.00
30	BETA_METEO_TERRACE	0.00152	0.000893	1.71	0.09
31	BETA_NO_DISTANCE_AV	-4.50	0.395	-11.40	0.00
32	BETA_PRICE_EMPLOYEE	0.000645	0.0342	0.02	0.98
33	BETA_PRICE_STUDENT	-0.0570	0.0376	-1.52	0.13
34	BETA_TAP_BEER_AFTER_LUNCH	0.601	0.180	3.34	0.00
35	RHO_PREVIOUS_CHOICE	1.78	0.109	16.38	0.00

Summary statistics

Number of observations = 1867

Number of estimated parameters = 35

$$\mathcal{L}(\beta_0) = -5035.429$$
$$\mathcal{L}(\hat{\beta}) = -3101.563$$
$$-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] = 3867.733$$
$$\rho^2 = 0.384$$
$$\bar{\rho}^2 = 0.377$$

Table 21: Dynamic model with agent effect correction (first choice only): here the results with 250 draws (results are similar with more draws).

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ALPHA_FIRST_CHOICE	1.23	0.209	5.90	0.00
2	ASC_ARC	-5.87	1.22	-4.81	0.00
3	ASC_ATL	-2.89	0.760	-3.80	0.00
4	ASC_BC	-2.01	0.680	-2.96	0.00
5	ASC_BM	-1.95	0.945	-2.06	0.04
6	ASC_COP	-0.679	1.28	-0.53	0.59
7	ASC_COR	-0.470	0.332	-1.42	0.16
8	ASC_ELA	-1.90	0.597	-3.19	0.00
9	ASC_GIA	-0.00900	0.492	-0.02	0.99
10	ASC_HOD	-0.135	0.494	-0.27	0.78
11	ASC_INM	-3.70	0.993	-3.72	0.00
12	ASC_KEB	1.45	0.296	4.89	0.00
13	ASC_KLE	-3.52	1.16	-3.03	0.00
14	ASC_MX	-4.39	0.688	-6.38	0.00
15	ASC_ORN	-1.17	0.251	-4.68	0.00
16	ASC_PAR	-0.630	0.293	-2.15	0.03
17	ASC_PH	-1.24	0.988	-1.25	0.21
18	ASC_PIZ	1.26	0.396	3.18	0.00
19	ASC_SAT	-2.17	0.657	-3.30	0.00
20	ASC_VAL	0.672	1.56	0.43	0.67
21	ASC_VIN	-5.65	3.82	-1.48	0.14
22	BETA_CAPACITY_INSIDE	0.00405	0.00280	1.45	0.15
23	BETA_DINNER	0.633	0.354	1.79	0.07
24	BETA_DISTANCE_AFTERNOON	-0.000782	0.000643	-1.22	0.22
25	BETA_DISTANCE_LUNCH_CAF	-0.00398	0.000653	-6.10	0.00
26	BETA_DISTANCE_LUNCH_REST	-0.00191	0.00150	-1.27	0.20
27	BETA_DISTANCE_LUNCH_SELF	-0.00413	0.000510	-8.09	0.00
28	BETA_DISTANCE_MORNING	-0.00286	0.000945	-3.03	0.00
29	BETA_EVALUATION_CAFET	2.21	0.161	13.75	0.00
30	BETA_EVALUATION_SELF	2.26	0.204	11.07	0.00
31	BETA_METEO_TERRACE	0.00212	0.00107	1.97	0.05
32	BETA_NO_DISTANCE_AV	-3.72	0.561	-6.62	0.00
33	BETA_PRICE_EMPLOYEE	0.0300	0.0491	0.61	0.54
34	BETA_PRICE_STUDENT	-0.00686	0.0514	-0.13	0.89
35	BETA_TAP_BEER_AFTER_LUNCH	0.806	0.256	3.14	0.00
36	RHO_PREVIOUS_CHOICE	0.424	0.161	2.63	0.01
37	SIGMA_ARC	4.34	0.752	5.77	0.00
38	SIGMA_ATL	2.04	0.285	7.15	0.00
39	SIGMA_BC	2.25	0.402	5.61	0.00
40	SIGMA_BM	4.23	0.806	5.25	0.00
41	SIGMA_COP	2.76	1.07	2.57	0.01
42	SIGMA_COR	1.09	0.377	2.88	0.00
43	SIGMA_ELA	-1.33	0.285	-4.65	0.00
44	SIGMA_GIA	1.19	0.107	11.18	0.00
45	SIGMA_HOD	0.750	0.602	1.24	0.21
46	SIGMA_INM	1.64	0.514	3.19	0.00
47	SIGMA_KEB	0.806	0.407	1.98	0.05
48	SIGMA_KLE	1.38	0.551	2.51	0.01
49	SIGMA_MX	2.47	0.375	6.58	0.00
50	SIGMA_ORN	0.921	0.230	4.01	0.00
51	SIGMA_PAR	-1.39	0.338	-4.12	0.00
52	SIGMA_PH	3.98	0.680	5.85	0.00
53	SIGMA_PIZ	-1.78	0.615	-2.90	0.00
54	SIGMA_SAT	1.76	0.281	6.27	0.00
55	SIGMA_VAL	-1.61	1.01	-1.59	0.11
56	SIGMA_VIN	-2.26	2.11	-1.08	0.28

Summary statistics

Number of observations = 1867

Number of estimated parameters = 56

$$\begin{aligned}
\mathcal{L}(\beta_0) &= -5035.429 \\
\mathcal{L}(\hat{\beta}) &= -2428.280 \\
-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] &= 5214.299 \\
\rho^2 &= 0.518 \\
\bar{\rho}^2 &= 0.507
\end{aligned}$$

Table 22: Dynamic model with agent effect correction (first and most frequent choices): here the results with 250 draws (results are similar with more draws).

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ALPHA_FIRST_CHOICE	0.985	0.201	4.90	0.00
2	ALPHA_MOST_CHOSEN	1.79	0.162	11.04	0.00
3	ASC_ARC	-5.70	1.45	-3.94	0.00
4	ASC_ATL	-2.72	0.828	-3.28	0.00
5	ASC_BC	-0.966	0.646	-1.50	0.13
6	ASC_BM	-1.42	0.985	-1.44	0.15
7	ASC_COP	0.833	1.15	0.72	0.47
8	ASC_COR	-0.551	0.255	-2.16	0.03
9	ASC_ELA	-1.02	0.568	-1.80	0.07
10	ASC_GIA	0.643	0.551	1.17	0.24
11	ASC_HOD	0.171	0.509	0.34	0.74
12	ASC_INM	-2.50	0.993	-2.52	0.01
13	ASC_KEB	1.87	0.431	4.33	0.00
14	ASC_KLE	-3.10	1.15	-2.70	0.01
15	ASC_MX	-2.93	0.739	-3.97	0.00
16	ASC_ORN	-1.10	0.244	-4.50	0.00
17	ASC_PAR	-0.753	0.292	-2.58	0.01
18	ASC_PH	-0.0867	0.822	-0.11	0.92
19	ASC_PIZ	1.53	0.385	3.99	0.00
20	ASC_SAT	-1.57	0.751	-2.09	0.04
21	ASC_VAL	1.09	1.57	0.69	0.49
22	ASC_VIN	-3.78	0.912	-4.14	0.00
23	BETA_CAPACITY_INSIDE	0.00647	0.00294	2.20	0.03
24	BETA_DINNER	0.654	0.366	1.79	0.07
25	BETA_DISTANCE_AFTERNOON	-0.00106	0.000632	-1.68	0.09
26	BETA_DISTANCE_LUNCH_CAF	-0.00367	0.000598	-6.14	0.00
27	BETA_DISTANCE_LUNCH_REST	-0.00225	0.00145	-1.55	0.12
28	BETA_DISTANCE_LUNCH_SELF	-0.00375	0.000480	-7.83	0.00
29	BETA_DISTANCE_MORNING	-0.00290	0.00105	-2.75	0.01
30	BETA_EVALUATION_CAFET	2.02	0.152	13.35	0.00
31	BETA_EVALUATION_SELF	2.09	0.196	10.67	0.00
32	BETA_METEO_TERRACE	0.00120	0.00125	0.96	0.34
33	BETA_NO_DISTANCE_AV	-3.38	0.540	-6.25	0.00
34	BETA_PRICE_EMPLOYEE	0.0618	0.0622	0.99	0.32
35	BETA_PRICE_STUDENT	-0.00488	0.0646	-0.08	0.94
36	BETA_TAP_BEER_AFTER_LUNCH	0.766	0.253	3.03	0.00
37	RHO_PREVIOUS_CHOICE	-0.118	0.188	-0.63	0.53
38	SIGMA_ARC	-4.49	0.870	-5.16	0.00
39	SIGMA_ATL	-2.09	0.316	-6.62	0.00
40	SIGMA_BC	-1.65	0.383	-4.32	0.00
41	SIGMA_BM	-3.93	0.775	-5.07	0.00
42	SIGMA_COP	-1.49	0.832	-1.79	0.07
43	SIGMA_COR	-1.02	0.208	-4.90	0.00
44	SIGMA_ELA	1.24	0.213	5.83	0.00
45	SIGMA_GIA	-1.21	0.120	-10.13	0.00
46	SIGMA_HOD	-0.192	0.572	-0.34	0.74
47	SIGMA_INM	-1.35	0.560	-2.42	0.02
48	SIGMA_KEB	-0.840	0.613	-1.37	0.17
49	SIGMA_KLE	-1.54	0.416	-3.71	0.00
50	SIGMA_MX	1.80	0.305	5.91	0.00
51	SIGMA_ORN	-0.606	0.238	-2.55	0.01
52	SIGMA_PAR	0.621	0.366	1.70	0.09
53	SIGMA_PH	-3.37	0.509	-6.62	0.00
54	SIGMA_PIZ	1.90	0.594	3.20	0.00
55	SIGMA_SAT	-1.67	0.291	-5.75	0.00
56	SIGMA_VAL	-1.25	1.30	-0.96	0.34
57	SIGMA_VIN	0.878	0.437	2.01	0.04

Summary statistics

Number of observations = 1867

Number of estimated parameters = 57

$$\mathcal{L}(\beta_0) = -5035.429$$
$$\mathcal{L}(\hat{\beta}) = -2335.750$$
$$-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] = 5399.357$$
$$\rho^2 = 0.536$$
$$\bar{\rho}^2 = 0.525$$