Journal of Universal Computer Science, vol. 25, no. 8 (2019), 946-966
submitted: 26/12/18, accepted: 10/7/19, appeared: 28/8/19 © *J.UCS*

Planning of Urban Public Transportation Networks in a Smart City

Jonathan Frez
(Universidad Diego Portales, Santiago, Chile
jonathan.frez@mail.udp.cl)

Nelson Baloian
(Departamento de Ciencias de la Computación, Universidad de Chile, Santiago, Chile
nbaloian@dcc.uchile.cl)

José A. Pino
(Departamento de Ciencias de la Computación, Universidad de Chile, Santiago, Chile
jpino@dcc.uchile.cl)

Gustavo Zurita
(Management Control and Information Systems Department, Universidad de Chile
Santiago, Chile
gzurita@fen.uchile.cl)

Franco Basso
(Escuela de Ingeniería Industrial, Universidad Diego Portales, Santiago, Chile
franco.basso@udp.cl)

Abstract: Planning efficient public transport is a key issue in modern cities. When planning a route for a bus or a line for a tram or subway, it is necessary to consider people's demand for this service. In this work we present a method to use existing crowdsourced data (like Waze and OpenStreetMap) and cloud services (like Google Maps) to support a transportation network decision making process. The method is based on the Dempster-Shafer Theory to model transportation demand. It uses data from Waze to provide a congestion probability and data from OpenStreetMap to provide information about location of facilities such as shops, in order to predict where people may need to start or end their trips using public transportation vehicles. The paper also presents an example using this method with real data. The example shows an analysis of the current availability of public transportation stops in order to discover its weak points.

Keywords: Dempster-Shafer theory, transportation networks, smart cities, Origin-Destination planning problem.
Categories: H.1.0, H.4.0, J.2

1 Introduction

Many urban areas are quickly growing. Their decision-making problems are increasingly complex, and better methods to evaluate solutions are needed in order to support this growth [Heilig, 2012].

Many decision problems concerning cities are spatial. A typical problem is to define an area to support a certain requirement or service, e.g., space for a new road, an industrial area, or a hospital. Furthermore, cities are constantly changing, and they have dynamic problems, like in public transportation services: the routes can be dynamically defined to cope with new requirements and constraints. Some of the decisions may be to find a place for a new bus station, define a bus route, define multiple routes, or even plan a full transportation network.

According to Harrison et al. [Harrison,2010], a "smart city" is a city that monitors and integrates data and information of its critical infrastructures, including roads, bridges, tunnels, rails, subways, airports, and seaports to optimize its resources, and maximize services to its citizens [Basso, 2018], [Stylianou, 2019].

At the same time, citizens are using information technologies (IT) to provide data which can be used to support the decision making process for several cities requirements. They also consume data [Zurita, 2015]. Cloud Computing Services providing Software as a Service (SaaS) can support some of the IT used by citizens. Citizens often access the SaaS model of Cloud Computing and web interfaces using mobile applications [Chourabi, 2012]. Some of the SaaS services with spatial data properties are, e.g., Google Maps, OpenStreetMap and Waze. These services provide geo-localized data in a graphical way, they are free, and they share a singular characteristic: they use crowdsourced data to provide information.

The research question that inspired the work presented here is whether we can use the data freely available "in the cloud" to accomplish tasks for what is known as a "smart city" and would otherwise require an enormous effort for gathering the required data in terms of time and money. As an example, in this work we use free services provided by Google Maps, OpenStreetMap and Waze to download and use data to feed a Spatial Decision Support System for transportation network planning. More specifically, we use it for estimating the demand and congestion along a planned route of a bus line of the public transportation system, which is commonly known as the Origin-Destination (OD) route evaluation. For this purpose, we use an approach based on the Dempster-Shafer theory (DST) [Shafer, 1976]. This theory allows to model decisions based on uncertain and incomplete data, by studying the extent to which a hypothesis can be supported by data.

Santos et al. [Santos, 2011] describe a user-friendly web-based spatial decision support system (wSDSS) aimed at generating optimized vehicle routes for multiple vehicle routing problems that involve serving the demand located along arcs of a transportation network. The wSDSS incorporates Google Maps (maps and network data), a database, a heuristic and an ant-colony meta-heuristic developed by the authors to generate routes and detailed individual vehicle route maps. It accommodates realistic system specifics, such as vehicle capacity and shift time constraints, as well as network constraints such as one-way streets and forbidden turns. The wSDSS can be used for "what-if" analysis related to possible changes to input parameters such as vehicle capacity, maximum driving shift time, seasonal variations of demand, network modifications, and imposed arc orientations. The system was tested for urban trash collection in Coimbra, Portugal.

A crowdsourced database is the OpenStreetMap project [Haklay, 2008]. Worldwide, several volunteers are contributing information to this "free" geodatabase. In some cases this database exceeds proprietary ones by 27% [Neis, 2011], and for some authors, it is more complete than Google maps or Bing maps [Cipeluch , 2010]. OpenStreetmap data has been proposed to support traffic-related decisions by developing traffic simulations [Zilske, 2011], or solutions to achieve a new web-based trip optimization tool [Klug, 2012]. It has also been used to support transportation network planning [Joubert, 2013]. In addition, Boye et al. [Boye, 2014] describe a spoken-dialogue based prototype for pedestrian navigation using various grounding strategies based on OpenStreetMaps. Similarly, Jacob et al. [Jakob, 2009] present a web-based campus guidance system for pedestrian navigation aimed at providing support for visitors. They developed an OpenStreetMap based system to generate short paths using both outdoor walkways and indoor corridors between various locations.

Waze is generating another popular crowdsourced geodatabase. Waze is a mobile GPS application that allows to measure and report traffic conditions and events, e.g., it automatically detects traffic jams, users can report accidents, weather effects on the roads, and other alerts. In the literature, we did not find a decision support system using Waze data, maybe because it is hard to obtain it. However, we found traffic condition analysis systems [Silva, 2013] based on real time data obtained from Waze.com using a WebCrawler, and an accidents data mining analysis proposal [Fire, 2012] based on the same real time data from Waze.com. In our work, the data is obtained using the same technique: we developed a WebCrawler to reconstruct a historical database based on published data on waze.com.

We focus on spatial DSS using belief functions [Frez, 2014], in particular Dempster-Shafer theory. DST proposes to use sets of hypotheses regarding a variable (e.g. the number of people waiting for a bus at a particular bus stop and on a certain time is between n1 and n2) associated with a probability of being correct. Using belief functions, we can provide a "hypotheses support value" called belief. The belief can be assigned to a certain geographical area satisfying a hypotheses set.

Thus, the main contribution of this paper is to present and validate a method, which allows making a preliminary prediction about the passengers' demand and the traffic congestion along a planned bus route using only free available information and resources from the web.

The rest of the paper is organized as follows. Section 2 explains the OD route problem in detail. Section 3 introduces the "Belief Route" concept, which is our approach to tackle, this problem. Section 4 describes how to compute a "Belief Route". Section 5 describes a tool, which helps a user to do this computation in an easy and flexible way, which converts it in a powerful decision-making support tool for public transportation planning. Section 6 presents some experiments made for validating our approach using real data. Finally, Section 7 presents the Conclusions.

2 The OD Route Problem

People's transportation modeling has been largely based on the four-step model (FSM) that was developed in the 1950s [Wegener, 2004]. The four-step model is a method used to forecast transportation demand and it is primarily intended to be used for long term planning or for infrastructure development [Bhat, 1999]. It involves four major steps: trip generation, trip distribution, mode choice, and route choice (or assignment).

Trip generation establishes the propensity to travel by estimating how many trips are generated by analyzing possible origins (known as productions) and destinations (known as attractions) separately [Wang, 2015]. For example, how many trips would a residential building produce and how many trips would a shopping center attract. Trips are modeled at different aggregation levels (zonal, household, etc.) and sometimes at the personal level. The most common models used in trip generation are either category models or regression models combining many socioeconomic and land use related variables. In the next step, trip distribution [Ecans, 1976], the separate origins and destinations are combined to create origin-destination (OD) pairs. The most common way of creating OD pairs is using gravity models, the result of which would yield a matrix of OD pairs. The next step, mode choice, further groups the trips (OD pairs) into different modes of travel (car, public transport, cycling, etc.). The effect of time of day is integrated at this stage as well, depending if the data used has a time dependence or not (usually AM peak and PM peak periods or hours). The most common model form used during this process is discrete choice modeling, more specifically logit models [Ortuzar, 2011]. Route choice or route assignment places the generated trips onto the transportation network. Trips are injected into the network according to the OD matrix established in the distribution step and further refined in the mode choice step. Depending on the aggregation level, with zonal being common, household vs zonal, trips are injected into zone at the centroids and then spread to the road network. Trips are most commonly spread into the road network under the assumption of achieving system equilibrium (at first every user would choose the path of least resistance from origin to destination, and network equilibrium occurs when no user can decrease travel effort by shifting to a new path). Volumes of vehicles on the network because of route assignments are calibrated using real life traffic counts [Nagel, 1992].

Travel demand data is collected in a variety of ways. One popular method is the household travel survey, which involves surveying a sample of the population of varying socioeconomic status (income, number of members in household, vehicle ownership, etc.). Households participating in the survey provide their travel activity for a certain period (e.g., a random day of the week). The collected information includes origin, destination, travel mode, trip purpose, vehicle occupancy, etc. Nevertheless, recent trends have used passive data to determine real-world travel demand [Munizaga, 2012]. This information is then used with the trip production and attraction models from the trip generation step to create the OD matrix [Guy, 2005].

The first versions of the four-step model were limited in a number of ways, namely because of an aggregated approach of using zones as origins and destination, treating trips as a single origin and destination, and because of inadequate integration of temporal effects. The data was aggregated into zones (commonly known as traffic analysis zones-TAZ), OD pairs were split into two major categories: home-based (HB) and non-home-based (NHB) and intermediate stops within a trip or OD pair were not taken into account. Time of day was introduced late in the process and only as a percentage depending on time of day (a higher percentage of trips are injected into the system during morning and evening peak hours) [Bricker, 1992].

From the need of having a more behavior-based approach that better reflects real life transportation decisions rather than a trip-based approach, as well as the need for short-term traffic management strategies, the activity approach was developed. The activity based approach views demand derived from the need to pursue activities distributed in space. By focusing on sequences of patterns of activity behavior, the activity-based approach integrates how travelers modify their activity patterns due to changes in their life or the state of the transportation network (work schedule change, congestion, etc.) [Bricker, 1992].

With the advance of computer processing power, various simulation models have been created that can treat both the demand (trip generations and distributions) and supply (road infrastructure, traffic management actions, incidents, traffic signals, etc.) at different levels of detail. The three main levels of detail are: macroscopic (static, deterministic), mesoscopic (stochastic, dynamic traffic assignment), and microscopic (lane change rules, car following models) [Casas, 2011]. Often, a mix of aggregate and disaggregate approaches are used to model travel demand [Wu, 2013]. Currently, intelligent transportation systems (ITS) are used to manage traffic in real time with sensors placed on the road network for example [Barceló, 2007)]

Computational power and more detailed modeling is not the only change worth noting. Another fundamental component of modeling, how and where we get our data is also changing. With people more and more connected to social media and the advent of the Geospatial Revolution [Penn, 2016], we are essentially becoming sensors that are collecting all sorts of data, including data related to transportation. A recent example of using social media data to determine transportation activity patterns is the “A sense of place” project in Turku, Finland where Instagram posts near bus stops were collected and combined with other data to discover activity patterns [Turku, 2016]. In this paper, we propose a traffic assignment method and likelihood of travel between an origin and destination using OpenStreetMap and Waze data.

3 The OD and Belief Route

A public transportation system is typically a complex network. These networks are composed of various transportation lines designed to cooperate and complement an urban scale transportation solution [Yang, 2009]. Most metropolitan areas having more than 500.000 inhabitants would have at least bus lines, taxis, and depending on the development degree of the region, they will have some metropolitan railway system like subway and/or trams. When planning and evaluating the usefulness of a new bus line, which will serve passengers along a route, it is necessary to estimate at least two important characteristics. The first one is the estimation of the demand for that route, which could be represented by the number of passengers that would like to take the bus at a certain stop; the other one is the traveling time of the bus between all two consecutives stops of the route. Since the route does not exist yet, it is difficult to estimate with high certainty the demand that the service will have. Studies may be conducted based on data about population living along the route and facilities, services and amenities near it, but this requires data gathering which might not be very simple to perform. The traveling time is something, which is predictable with some certainty since the existing traffic is likely not to change too much with the introduction of a single new bus line. Nevertheless, it requires some work, which will require some tasks to perform.

It would be very useful if we could do at least some preliminary evaluation of the new planned route without having to spend too much time and/or money in the process, in order to decide the feasibility of a new route at an early stage. As we stated in the first section, we would like to find out the possibility of making a preliminary evaluation with data obtained exclusively from free sources on the web, specifically from OpenStreetMap and Waze. For this purpose, we propose to use "Belief Maps" [Frez, 2014]. Belief Maps are Suitability Maps [Hopkins, 1977] built with incomplete data, based mostly on beliefs rather than facts. Suitability maps are used to determine the appropriateness of a given area for a particular use. This process is called suitability analysis and its fundamental principle "is that each aspect of the landscape has intrinsic characteristics that are to some degree either suitable or unsuitable for the activities being planned. Suitability is determined through systematic, multi-factor analysis of the different aspects of the terrain" [Murphy, 2005].

In order to generate these maps, generators use data about a variety of factors like physical layout, political decisions, economic data, etc. The results are then displayed on a map, which shows in different colors the suitability of the geographical areas. This type of maps is used to answer questions like "Where is the best location to build a water reservoir?", "which are the most dangerous places where landslides may occur?", "which are the most probable places to find minerals" or "which is the best place for opening a retail store?" In this last example, a commercial developer may take into consideration distance to major highways and competing stores, then combine the results with land use, population density, and consumer spending data to decide on the best location for that store [Johnston, 2001]. In a Belief Map, we formally identify a set of hypotheses that may reinforce or weaken the plausibility and certainty about the occurrence of a certain fact, which will be used when computing the suitability of a certain location or area. Various hypotheses can be combined with different weights to compute what in DST is called the "accumulated mass" supporting a belief. For example, if we would like to generate a map showing the belief degree of finding people in a certain area, plausible hypotheses for this map would be: "Medium concentration of people can be found in commercial areas" and "There is a high concentration of people around schools at certain times of the day".

Based on similar principles as Belief Maps we now introduce the concept of Belief Route, which will be used to estimate the numbers of possible passengers that would like to use a public transportation bus serving that route in its various sections. Each section is delimited by two contiguous bus stops. A Belief Route is composed of three basic elements: 1) The origin and destination points; 2) The polyline connecting these two points, which describes the route followed by the bus in order to travel from origin to destination. 3) A set of hypotheses specifying a possible transportation demand of an OD. Besides estimating the possible demand for the route, in some planning scenarios it is important to estimate the traffic congestion along the whole route in order to evaluate if the traveling time between stops as well as from the origin to the destination is both predictable and it takes reasonable time [Fernandez, 2008], [Vuchic, 2002]. Using Belief Route and Belief Congestion Route maps, the decision maker can compare various paths and evaluate their ability to satisfy the transportation requirement of the population along the route.

4 Computing a Belief Route

Given the polyline representing the route and the hypotheses set, the first step for computing the Belief Route is to divide the area in 150x150 meters disjoint rectangles. The dimensions are based on the margin of error of the algorithm that associates the position of the users based on the GPS of the vehicles. Then we compute the mass for each rectangle the route goes through according to the set of hypotheses used as rules to predict the chance of finding people wanting to use the transportation service. To each hypothesis, we assign a weight that will determine the amount of mass assigned to the rectangle where the hypothesis holds.

The most critical aspect for the success of the Belief Route and Belief Congestion Route is a correct estimation of the hypotheses, which would generate the "mass" supporting the possibilities of the demand and congestion estimations along the planned route. In order to test the correctness of the hypotheses we should use real data from a public transportation system (see section 6). In this work, we will use the following hypotheses:

1. People require more transportation near shop areas. Weight: 50 %.
2. People require more transportation near amenity places. Weight: 50%.

The weights are preliminary guesses and they will have to be adjusted by comparing the predicted demand with the real data. After this step, the initial configuration for the predictor is complete and all contributions are correctly distributed, the next step is to combine the contributions in order to generate a mass assignment function and then compute the belief for each rectangle. At this point, it is assumed that the contribution of each rectangle is independent of the other rectangles and there is a spatial decay for the contribution to neighbouring cells. This decay is estimated with a normal cumulative distribution function, for which a μ and a σ as mean and standard deviation has to be previously defined according to the expert´s criteria, as seen in the next equation naming the contribution as p for compatibility.

$$p(d) = \{1\ 1 - \frac{1}{2}\left(1 + erf\left(\frac{d-\mu}{\sigma\sqrt{2}}\right)\right) \quad if\ d = 0\ (inside)\ if\ d > 0\ (outside)$$

In the Shafer extension to the theory, the main element are mass assignment functions, which are denoted as $m \to [0,1]$. These operates over subsets of a domain or frame of discernment X. In our case, the domain is people density, having two possible values high people density, and low people density. The mass m for the subset "high people density" (A) can be defined as the sum of all the amenities of a certain contribution (C).

$$m(A) = \sum_{e_i \in C} p(e_i)$$

The final value for people density within a rectangle can be measured using either the belief of A or the plausibility of A. The belief (bel(A)) is defined as the sum of all the masses of subsets of the set of interest and plausibility (pl(A)) as the sum of all subsets that contains A. Since the value of $p(d)$ depends on the distance, it is possible to choose a maximum value of distance to be evaluated, defining the set of interest as the subsets that are close to the cell.

When a cell is affected by more than one hypothesis (A,B and C subsets), the Dempster Rule is used to combine and compute the joint mass for the cell. See equation below:

$$m_{1,2}(\emptyset) = 0$$

$$m_{1,2}(A) = \frac{1}{1-K} \sum_{A=B\cap C \neq \emptyset} m_1(B)m_2(C)$$

$$K = \sum_{B\cap C = \emptyset} m_1(B)m_2(C)$$

K is only used for normalization purposes. The result of this process is a belief map, since the focus of the problem is the route; the next step is to find the subset of cells from the belief map that intersects with the route. The final result omits all other cells and only keeps those which represent a location of the route. Then we normalize the subset of cells in order to adjust the scale of belief to the new domain. The algorithm that allows generating Belief Routes includes the intersect operation before making the normalization is the following:

```
Input: a list of hypotheses and the historical data about paths segments records, the path to be evaluated
Result: grid, an array with the belief values for each square of the grid generated for each rectangle of the
considered area filtered by the geometrical intersection with the path.

/* obtain the hypotheses, the temporal and distance models, and generate a grid for computing the belief value
for each cell (quad) of the grid */
hypotheses = getHypos();
temp = getTemporalModel();
path = getPath();
dist = getDistanceModel();
quads = generateGrid();
/*a two-dimensional array for storing the mass for each cell of the grid according to each hypothesis given
by the user*/
quadrants = Array();
forall the hypotheses as hypo do
        //retrieving all features in the area which match the hypothesis
        feats = getFeaturesHypo(hypo);
        forall the feats as feat do
           forall the quads as quad do
              //obtain the mass associated to the hypothesis
              mass = feat.mass;
              /calculate its contribution to this cell
              mass = mass*temporalEval(datetime,hypo,temp[feat]);
              mass = mass*distanceEval(quad, feat.location);
              quadrants[quad][ feat].append(mass );
          end
```

```
        end
end

/* we apply the combination rules of Dempster-Shafer, thus obtaining a single mass for each hypothesis for
each cell */
quadrants = combineHypoteses(quadrants);

/* Return list of quadrants that intersects with path */
quadrants = intersects(quadrants,path)
/* values are normalized between 0 -1 */
quadrants = scaleValues(quadrants,getMin(quadrants),getMax(quadrants));

forall the quadrants as hypolist do
/*obtaining the mass of the composed hypothesis with the biggest value */
        bighypo = getBiggestHypo(hypolist);
        grid[quad] = bighypo;
end
```

We are going to use a basic example to explain the use of the Belief Route and Belief Congestion Route in a decision making process. We will analyze two different routes connecting a pair of origin-destination points using the Belief and Belief Congestion Routes. In Figs. 1 and 2, two options are shown (Route 1 and Route 2) both originating at point A and ending at point B. In this example route 1 is shorter than route 2, and the travel time is also shorter according to Google Directions API.

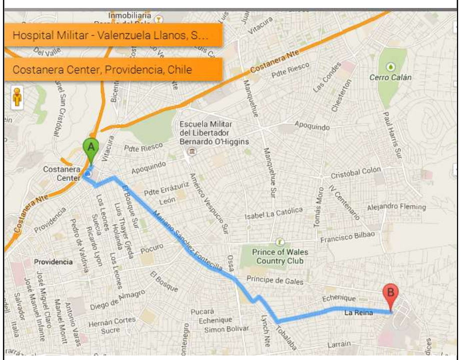


Figure 1: Route 1

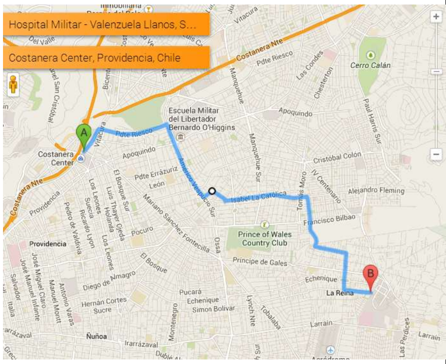


Figure 2: Route 2

In Figs. 3 and 4, the transportation demand is represented by a Belief Route. We can see that according to OpenStreetMap and the proposed hypothesis, route 2 has more demand than route 1.

Figure 3: Belief route 1

Figure 4: Belief route 2

The Belief Congestion Routes are computed in a similar way than the Belief Routes. The only difference is that now we use a single hypothesis: each report about a traffic jam in Waze adds mass to the rectangle where it was geo-referenced by the reporter. Figs. 5 and 6 show the BCR of both routes. According to the Waze information for both paths, route 1 has more belief of having congestions or traffic jams which implies less reliability.

From this example, we can note that route 2 has less congestion and more demand than route 1. However route 1 is shorter and the decision will depend of what kind of OD the decision maker is looking for. In order to support the decision, the visual evaluation of the Belief Route and the Belief Congestion Route is not enough. An evaluation metrics framework is needed and it will be part of our future work.

Figure 5: Congestion route 1

Figure 6: Congestion route 2

5 An Application for Developing Belief Routes

The prototype allows its users to define an OD pair and a polyline. It also lets the user specify hypotheses for transportation demand modeling, after which it can generate two types of visualization: The demand Belief Route and the Belief Congestion Route (Fig. 8). The application allows setting a transportation demand hypotheses set compatible with DST. It also allows including some model constraints, for example: avoid schools. After the hypotheses are included, the application allows choosing the type of 3D map, which will be generated and shown: Belief Route or Belief Congestion Route.

The application uses only data downloaded from sites offering free services available on the Web. In this version, we use the services of Google Maps to download the maps, which will be used to display the belief routes. Data about the facilities that are near the route are downloaded from OpenStreetMap. Data about the vehicular congestion on the streets is downloaded from Waze. The three sites have an API to request and download the data using the WebServices protocol. The application has been developed with an architectural design intended for easy and modular addition of other sites offering services for downloading data which could be used to include more information about the area where the OD route passes and which could help to make a better prediction of the transportation demand. The architecture of this application is explained in detail in [Zurita, 2014].

As seen in Fig. 7, the application interface is divided in two sections. The left one is for setting the parameters for computing the Belief Route and the congestion Belief Route. The user can choose one to compute using the toggle buttons at the bottom of this area. At the top portion of this area there are widgets for setting the parameters for computing the Belief Routes. The user should first select the model, which corresponds to what is going to be predicted. The model should be previously fed to the system as a program module developed in Java-Script (which is the language used to develop the whole application) and following a pre-defined interface. In the example of the figure the chosen model is called “People”. Beneath the model setting widget, we can see the widget used for stating the hypotheses. A pulldown menu displays all attributes available which can be used to define a hypothesis. The items of this menu are automatically generated according to the installed modules which download information about locations of facilities which could be useful for predicting the demand. For this work we used only information about objects provided by OpenStreetMap. At the right hand side of the selected attribute the user can define the weight with which this attribute will contribute to the calculation of the mass by clicking on a circular widget. The light-blue pushdown buttons allow stating additional hypotheses. In the figure we can see the attribute “shop” with a weight of 26%. Below this information, the current and previously defined hypotheses are displayed. In the figure, we can also see that the hypothesis amenity, with a weight of 50% was also stated. Underneath this information we can define attributes for which mass will be decremented or even suppressed if found on the route. For example, we can state that the probability of having people in a park may decrease or the probability of having people on a lake will be zero, even if there are shops and amenities nearby. Finally there is a widget that allows for setting the window of time for which the prediction will be generated. This is mainly for computing the Belief Congestion Route, which is very sensitive to the time of the day. For this case, data from Waze corresponding to that period will be considered.

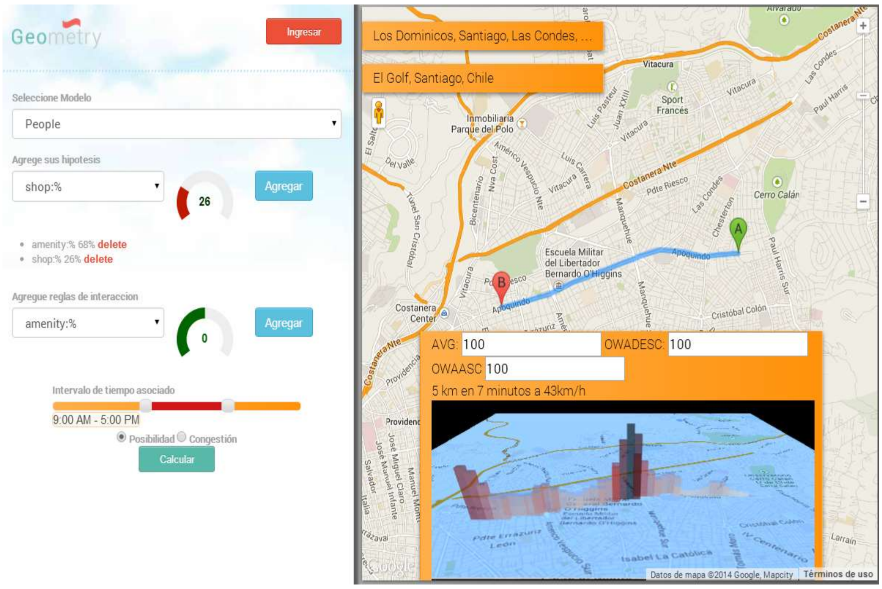


Figure 7: Evaluation of an OD using the developed application

The right hand side of the interface shows the map, downloaded from Google Maps, which allows the definition of the route to be evaluated and the display of the results. In Fig. 7 we can see a Belief Congestion Route is being displayed.

6 Belief Route Evaluation

The purpose of this evaluation it to test the correctness of the hypotheses that generate the “mass” supporting the beliefs for the demand and congestion estimations along the planned route and to show in detail how this work can be used to plan a new route. The accuracy of Waze to predict traffic jams has been already shown [Jämsä, 2013] so we will focus on evaluating the accuracy of the Belief Route. In general, an expert who has the necessary knowledge should state the hypotheses. In this case, for calculating Belief routes we used the hypotheses that people will use public transportation near areas where shops and services are located. We now want to test if these hypotheses were a correct guess. The testing method is simple: we will use real data from the public transportation system; if the prediction generated by the hypotheses set “matches”, the real data then we assume the hypotheses hold and the generated Belief Routes are valid. We used data about time and location where people start using each service from the Santiago integrated public transportation system, which uses exclusively plastic cards with magnetic bands pre-loaded with money as payment method. The trip cost is charged to the passenger’s card when the passenger takes the bus, and so, the transportation system registers the time and location of this event. We may notice this is not exactly the location of the demand, since this demand really occurs at the location where people live, work, shop, etc., but we assume people have done a short walk to the bus stop. On the other hand, the transportation system does not register the point where trips end since the trip costs are the same, regardless of the trip length and thus, passengers do not have to pass their cards through any device when leaving the vehicle. This data was obtained from the Chilean Ministry of Transportation.

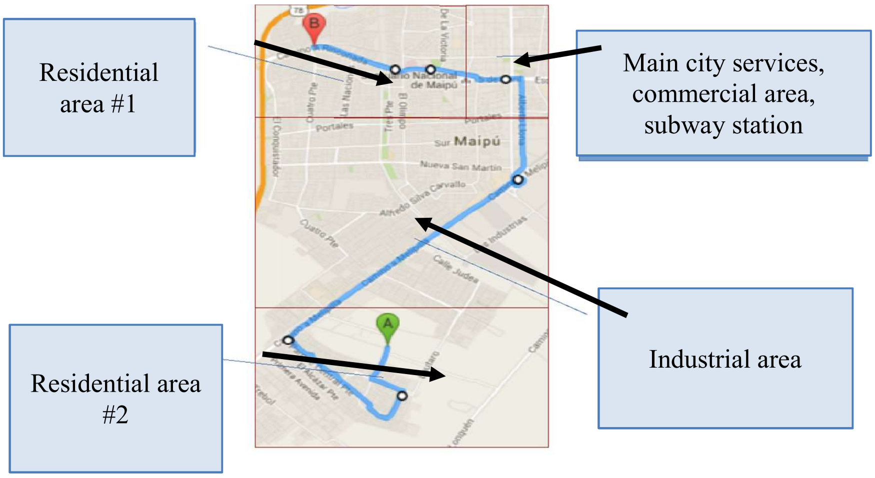


Figure 8: Zone types of the analyzed area.

We conducted a first test in a focalized area in order to make a preliminary evaluation of the suitability of the hypotheses and explore the possibilities of this method to plan a new route. The area selected for this test has a high transportation activity in the city of Santiago and is representative for having many shops in the city center and the other is representative for a residential area with high populationT he locations of the zone types are briefly described in Fig. 8. We will use the residential areas #1 and #2 as well as the industrial area to analyze the demand, since according to the data these are the areas where people start their trips.

First, Figs. 9, 10 and 11 show the places where people actually took the bus according to the data provided by the Ministry of Transportation. The colors show the concentration of people: light blue for few people to red for many people. In residential #1 area, it is possible to note two demand hotspots that distributed in five bus stops. In the Industrial zone, there is only one hotspot at the entrance of a small residential area. The transportation demand in the industrial zone tends to be more equally distributed as in the residential area and it shows only one small section with high demand. Finally, in residential area #2, the demand is distributed among several bus stops at the center of the area.

Now we will compare this data against the prediction we obtain when computing the Belief route using DST and information we can get from the web. As we already said, the tested hypotheses are that people require more transportation near shop areas, and that people require more transportation near amenity places.

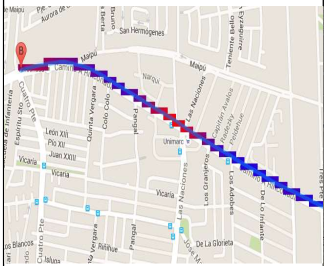
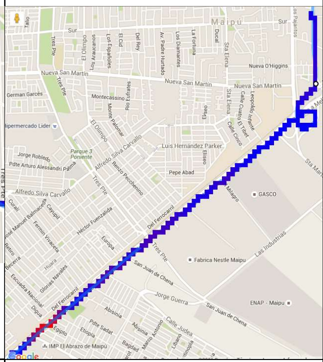
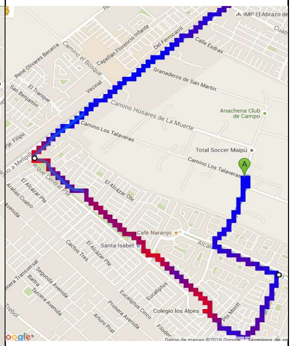
		

Figure 9: Transportation demand in Residential area #1	*Figure 10: Transportation demand in Industrial area*	*Figure 11: Transportation demand in Residential area #2*

These hypotheses are tuned to comply with the categories used by OpenStreetMap for the type of objects it stores. Apart from streets, we can get information about facilities classified as amenities and shops located in a certain area. Amenities refer to commercial places offering services like cafés, bars, restaurants, schools, universities, libraries, etc., while shops are related to commercial places selling goods like bakeries, convenience stores, supermarkets, medical supplies, etc. Figs. 12, 13 and 14 show the real demand vs. the predicted Belief route for the three selected areas.

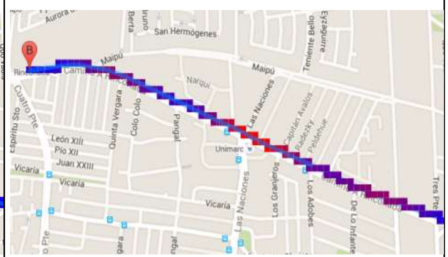
Real Demand	Predicted Demand
	

Figure 12: Real demand vs. predicted demand for residential area #1

For residential area #1 (Fig. 12), we verify the prediction was quite accurate for this route section. The Belief Route predicts a high concentration of the demand at the center of the route section, which almost fully matches the real demand. It also predicts a medium demand concentration in almost all the rest of the route which also matches the real data. The only noticeable difference is that the Belief route predicts a medium-high demand at the end of this route section, which is not backed by real data.

For the industrial area, there is also a good matching between the real demand and the belief route: the real data shows a medium demand in almost all the route section except for the last part of the section, where a high demand is shown. The predicted Belief Route also shows a medium demand on almost the whole section except for the same high-level demand at the end. The only difference is that the Belief route predicts low demand at the middle of the section (colored with light blue) which is not backed by real data (Fig. 13).

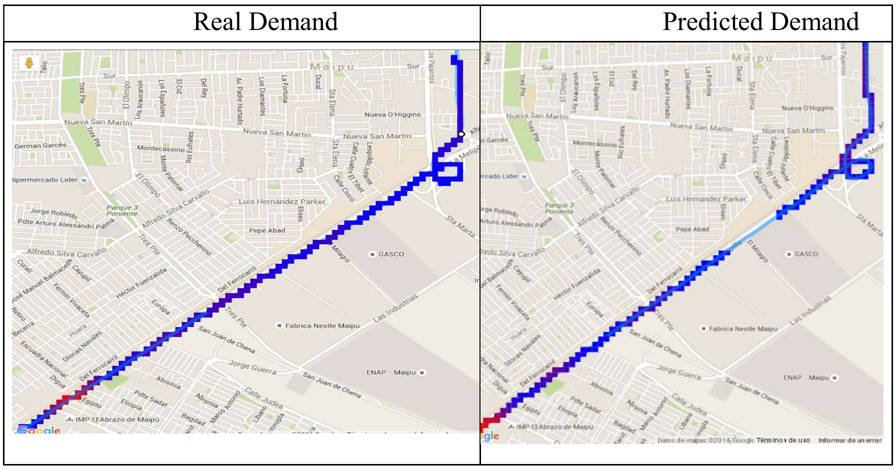


Figure 13: Real demand vs. predicted demand for industrial area

The prediction for the section going through the residential area #2 presents more differences than the other two areas, but it also shows many matches (Fig. 14). The prediction shows a small sub-section with medium-high demand at the beginning of the section whereas the real data shows only medium demand. After that, both figures show a medium demand until the point where the route makes a 90° left turn, where both, the real demand and the belief route show a high demand. After the turning point, there is a small section where the real data shows a medium-low demand whereas the Belief Route shows a medium demand followed by a sub-section where both figures show a medium-high demand. At the end of the section, the Belief route predicts some sections of low demand whereas the real data show medium-low level demand. The explanation may be as follows: the last part of the real route goes through an area with few shops and amenities but since it is the end of the route (or the beginning, depending on the direction people take the bus), it concentrates more customers coming from the nearby areas. This factor can be considered as an additional hypothesis in future analyses and may be incorporated to the set of hypotheses for obtaining a better prediction.

Summarizing, we can say that the Belief Route has a high level of matches with real data, and where it does not fully match it, the differences are not large. In fact, the Belief Route does never predict a high demand when the real data shows a low demand and vice versa. The most difficult place to predict the real demand seems to be the residential area where there are few shops and/or amenities and correspond to the beginning or end of the route.

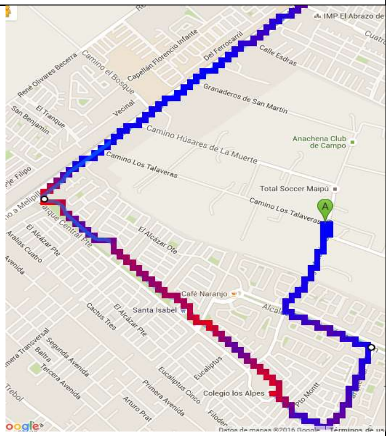
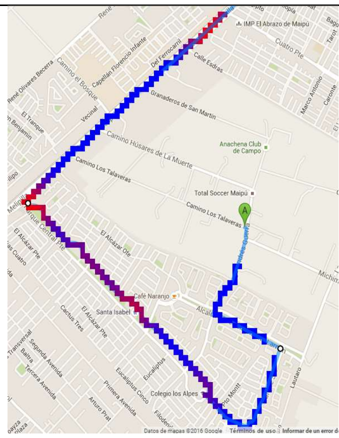
Real Demand	Predicted
	

Figure 14: Real demand vs. predicted demand for residential area #2

Now we will compute the Belief Route for the proposed new route (Route 2) and see if this one would have a higher demand than the original one according to the results of the prediction. In Fig. 15, we can easily note that for residential area #1, the new route presents a higher demand than the original one. This occurs mainly because the new route passes through areas where more shops and amenities are located.

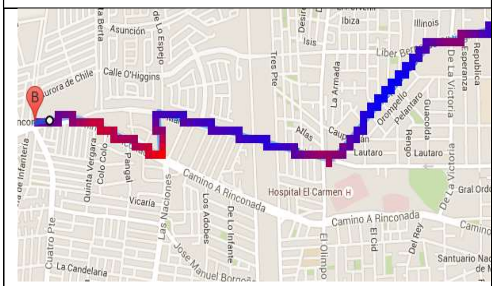
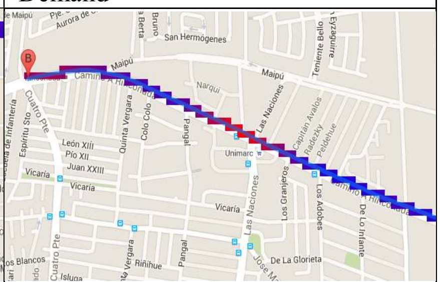
Original Route with Real Demand	Proposed Route with predicted Demand
	

Figure 15: Original route with real demand vs. new proposed route with its predicted demand for residential area #1

Fig. 16 shows the original route passing through the industrial area and the residential area #2 against the new proposed route (shown in a single figure). Here we also clearly note that the proposed route will have higher demand than the original route.

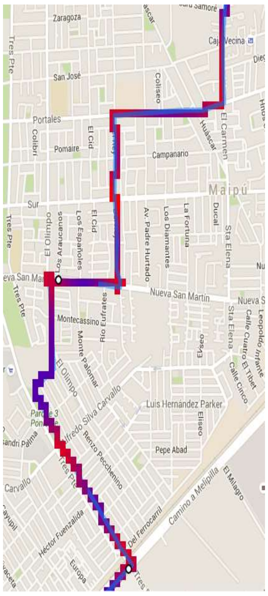
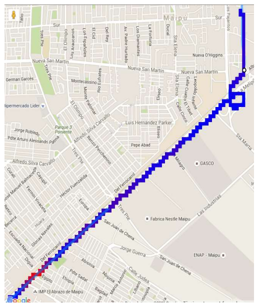
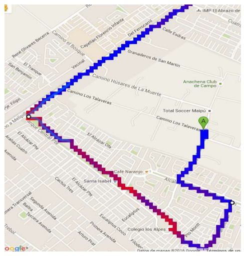
<table>
<tr><th>Original Route with Real Demand</th><th>Proposed Route with predicted Demand</th></tr>
<tr><td rowspan="2">
</td><td></td></tr>
<tr><td></td></tr>
</table>

Figure 16: Original route with real demand vs. new proposed route with its predicted demand for industrial area and residential area #2

In order to have more data about the accuracy of the predictions we conducted a large-scale experiment comprising the whole Santiago region. In a similar way to the previous experiments, the OpenStreetMap database was used to generate the belief map using hypotheses considered above. The belief map was calculated on a grid consisting of 1390 hexagons; each hexagon has an area of 700 square meters. Subsequently, the actual demand map was constructed based on the payment bearings that are georeferenced. We used 80 million markings, corresponding to two full weeks from July 5 until July 18, 2015. The demand map consists of the number of markings that exists in each zone (hexagon). Figs. 17 and 18 show the suitability maps for the real and predicted demands. At first glance, it is possible to see that there is a clear relationship between both maps (belief and demand), however this is not true for all cells.

To compare the relationship between both maps, we generated two series of values using the cell associated with each value as a common factor, and then we calculated the Pearson correlation between both series. The correlation between the values associated with each hexagon is 0.53, which indicates a high correlation between the real and the estimated magnitude of demand. However, for cells with less than 5000 passengers, the correlation is 0.09, so it is not possible to use it in areas of low demand. Between 5000 and 70,000 passengers is 0.50, between 50000 and 100000 passengers, the correlation is 0.73. To remind, Pearson's scale ranges from -1.0 to 1.0.

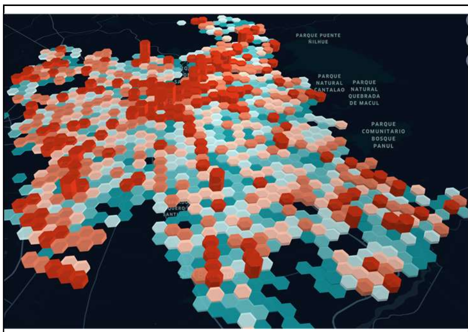


Figure 17: Real demand

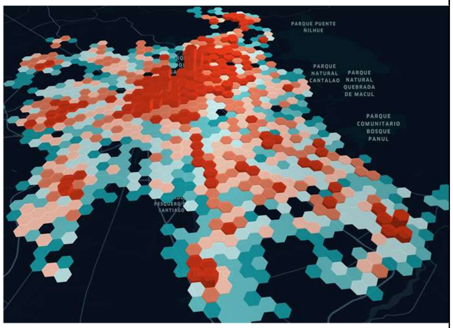


Figure 18: Predicted demand

7 Conclusions

In this work, we present a method to use existing crowdsourced data to support the transportation network decision-making process. The method uses the Dempster-Shafer Theory providing a framework to model transportation demand based on maps gathered from Google Maps, information about facilities located near the route from OpenStreetMap and information about the traffic from Waze. With this information, the Belief Route and the Belief Congestion can be computed. The Belief Route predicts the possible demand along an OD route and the Belief Congestion Route predicts the congestion so, the travel time for a bus on that route can be estimated. Our methodology allows us to make a preliminary evaluation of the route without investing time and resources to perform all the necessary tasks to gather the required information for a comprehensive assessment. Also, through an example based on real data, we illustrate how our procedure works making such preliminary evaluation possible.h prediction talk abstract.

We performed an experiment with a real existing route to test the validity of this approach. Specifically, we compare the prediction of the demand obtained with the Belief Route and the data collected from the Ministry of Transportation about the number of passengers getting in the buses serving that route in each bus stop. The obtained result showed that the computed Belief Route predicted the demand quite well (see previous section). To test the usefulness of this approach, we calculate the Belief Route for a new route with the same origin and destination points but traveling through other streets. We find that this new route would have a higher demand, thus being more resource efficient. The Belief Congestion Route could also be computed to check if the travel time would have also been acceptable. This paper does not cover the latter analysis.

The presented approach has some limitations. It is well-known that the route demand is dependent on the demand for other routes satisfaction [Peng, 2007]. Therefore, a large-scale application of the approach may not work well. Also, hypotheses such as "People require more transportation near shop areas" should be used with caution: transportation demand near shop areas may already be satisfied by other routes or other transportation modes.

As future work, we envisage the development of an indicator that could be used to measure, in a more objective way, the prediction performance of the Belief Route and the Belief Congestion Route when compared with real data. This indicator will help to calibrate the model and choose the hypotheses, which can better predict the demand as well as the travel time. It is also necessary to develop an indicator that could show the advantage of one route compared with another one in terms of satisfying a higher demand while maintaining acceptable travel times.

Acknowledgments

This paper is a revised and enlarged version of conference papers [Frez, 2014], [Baloian, 2015] and [Baloian, 2018].

References

[Baloian, 2015] Baloian, N., Frez, J., Pino, J. A., and Zurita, G. Efficient planning of urban public transportation networks. International Conference on Ubiquitous Computing and Ambient Intelligence, pp. 439-448. December 2015. Springer.

[Baloian, 2018] Baloian, N., Frez, J., Pino, J.A., Zurita, G., and Dolmajian, A. Preliminary Assessment of Public Transportation Routes using free available Cloud Services and Belief Routes. Collaborative Technologies and Data Science in Smart City Applications, pp. 82-100, Sep 2018. Logos Verlag, Berlin, Germany. ISBN 978-3-8325-4734-9.

[Barceló, 2007] Barceló, J., Delgado, M., Funes, G., García, D., Perarnau, J., and Torday, A. Online microscopic traffic simulation supports real-time traffic-management strategies. Siam News, 40(9), 2007.

[Basso, 2018] Basso, F., Basso, L. J., Bravo, F., and Pezoa, R. Real-time crash prediction in an urban expressway using disaggregated data. Transportation Research Part C: Emerging Technologies, 86, pp. 202-219, 2018.

[Bhat, 1999] Bhat, C. R., and Koppelman, F. S. Activity-based modeling of travel demand. In Handbook of transportation Science, pp. 35-61, 1999. Springer US.

[Boye, 2014] Boye, J., Fredriksson, M., Gotze, J, Gustafsson, J., and Konigsmann, J. Walk this way: Spatial grounding for city exploration, in Natural Interaction with Robots, Knowbots and Smartphones, pp. 59-67, 2014. Springer.

[Bricker, 1992] Bricker, D. and Cripe, J. J. W. An activity-based approach to early intervention. 1992. Paul H. Brookes Publishing.

[Casas, 2011] Casas, J., Perarnau, J., and Torday, A. The need to combine different traffic modelling levels for effectively tackling large-scale projects adding a hybrid meso/micro approach. Procedia-Social and Behavioral Sciences, 20, pp. 251-262, 2011.

[Chourabi, 2012] Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J., Mellouli, S., Nahon, K., Pardo, T., and Scholl, H. Understanding smart cities: An integrative framework. 45th Hawaii International Conference on System Science (HICSS). 2012. IEEE.

[Cipeluch, 2010] Cipeluch, B., Jacob, R., Winstanley, A. and Mooney, P. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. University of Leicester, 2010.

[Fernández, 2008] Fernández, R., Del Campo, M., and Swett, C. Data collection and calibration of passenger service time models for the Transantiago system. European Transport Conference. 2008.

[Fire, 2012] Fire, M., Kagan, D., Puzis, R., Lior Rokach, L., and Elovici, Y. Data mining opportunities in geosocial networks for improving road safety. IEEE 27th Convention of Electrical & Electronics Engineers in Israel (IEEEI), pp. 1-4, 2012.

[Frez, 2014] Frez, J., Baloian, N., Zurita, G. and Pino, J.A. Dealing with Incomplete and Uncertain Context Data in Geographic Information Systems, IEEE Conference on Computer Supported Cooperative Work in Design (CSCWD), Hsinchu, Taiwan, pp. 129-134. 2014.

[Guy, 2005] Guy, B. P. and Fricker, J. D. Guidelines for Data Collection Techniques and Methods for Roadside Station Origin-Destination Studies, 2005.

[Haklay, 20008] Haklay, M. and Weber, P. OpenStreetMap: User-generated street maps. IEEE Pervasive Computing, 7(4), pp. 12-18, 2008.

[Harrison, 2010] Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J., and Williams, P. Foundations for smarter cities. IBM Journal of Research and Development, 54(4), pp. 1-16, 2010.

[Heilig, 2012] Heilig, G. World urbanization prospects the 2011 revision. United Nations, Department of Economic and Social Affairs (DESA), Population Division, Population Estimates and Projections Section, New York, 2012.

[Hopkins, 1977] Hopkins, L. D. Methods for Generating Land Suitability Maps: A Comparative Evaluation. Journal of the American Institute of Planners, 23(4), pp. 386-400, 1977.

[Jacob, 2009] Jacob, R., Zheng, J., Ciepluch, B. and Mooney, P. Campus guidance system for international conferences based on openstreetmap. Web and Wireless Geographical Information Systems. pp. 187-198, 2009.

[Jämsä, 2013] Jämsä, J., Pieskä, S., and Luimula, M. Situation awareness in cognitive transportation systems. Special Issue on Cognitive Infocommunications in Infocommun. J. 5(4), pp. 10-16, 2013.

[Johnston, 2001] Johnston, K., Ver Hoef, J. M., Krivoruchko, K., & Lucas, N. Using ArcGIS geostatistical analyst (Vol. 380). Redlands: Esri, 2001.

[Joubert, 2013] Joubert, J.W. and Van Heerden, Q. Large-scale multimodal transport modelling. Part 1: Demand generation, 2013.

[Klug, 2012] Klug, M. CS Transport-Optimisation–A Solution for Web-based Trip Optimization Basing on OpenStreetMap. In 19th ITS World Congress, 2012.

[Munizaga, 2012] Munizaga, M. A., and Palma, C. Estimation of a disaggregated multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. Transportation Research Part C: Emerging Technologies, 24, pp. 9-18, 2012.

[Murphy, 2005] Murphy, M. D. Landscape architecture theory: an evolving body of thought. Waveland Press, 2005.

[Nagel, 1992] Nagel, K., and Schreckenberg, M. A cellular automaton model for freeway traffic. Journal de physique I, 2(12), pp. 2221-2229, 1992.

[Neis, 2011] Neis, P., Zielstra, D., and Zipf, A. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. Future Internet 4(1), pp. 1-21, 2011.

[Ortuzar, 2011] Ortuzar, J., and Willumsen, L. G. Modelling transport. John Wiley & Sons, 2011.

[Peng, 1997] Peng, Z., Dueker, K., Strathman, J., and Hopper, J. A simultaneous route-level transit patronage model: demand, supply, and inter-route relationship. Transportation 24(2), pp. 159-181, 1997.

[Penn, 2016] Penn State University. The Geospatial Revolution. Pennsylvania State University Broadcastings http://geospatialrevolution.psu.edu/. 2016.

[Santos, 2011] Santos, L., Coutinho-Rodrigues, J., and Antunes, C.H. A web spatial decision support system for vehicle routing using Google Maps. Decision Support Systems 51(1), pp. 1-9, 2011.

[Shafer, 1976] Shafer, G. A mathematical theory of evidence (Vol. 1). Princeton: Princeton University Press, 1976.

[Silva, 2013] Silva, T., Vaz de Melo, P., Viana, A., Almeida, J., Salles, J., and Loureiro, L.Traffic Condition Is More Than Colored Lines on a Map: Characterization of Waze Alerts, in Social Informatics, pp. 309-318, 2013.

[Stylianou, 2019] Stylianou, K., Dimitriou, L., and Abdel-Aty, M. Big Data and Road Safety: A Comprehensive Review. In Mobility Patterns, Big Data and Transport Analytics, pp. 297-343. Elsevier, 2019.

[Turku, 2016] Turku City. Exploring the potentials and possible uses of Location Based Social Network Data for urban and transportation planning in Turku City Centre. http://www.spinunit.eu/portfolio/a-sense-of-place/, 2016.

[Vuchic, 2002] Vuchic, V. R. Urban public transportation systems. University of Pennsylvania, Philadelphia, PA, USA, 2002.

[Wang, 2015] Wang, C., Quddus, M., Enoch, M., Ryley, T., and Davison, L. Exploring the propensity to travel by demand responsive transport in the rural area of Lincolnshire in England. Case Studies on Transport Policy 3(2), pp. 129-136. 2015.

[Wegener, 2004] Wegener, M. Overview of land-use transport models. Handbook of Transport Geography and Spatial Systems, 5, 127-146, 2004.

[Wu, 2013] Wu, P. P. Y., and Mengersen, K. A review of models and model usage scenarios for an airport complex system. Transportation Research Part A: Policy and Practice, 47, pp. 124-140, 2013.

[Yang, 2009] Yang, L. and Wan, B. A Multimodal Composite Transportation Network Model and Topological Relationship Building Algorithm. International Conference on Environmental Science and Information Application Technology. IEEE, 2009.

[Zilske, 2011] Zilske, M., Neumann, A., and Nagel, K. OpenStreetMap for traffic simulation. 1st European State of the Map–OpenStreetMap conference, 2011.

[Zurita, 2014] Zurita, G., Baloian, N. and Frez, J. Using the cloud to develop applications supporting geo-collaborative situated learning. Future Generation Computer Systems 34, pp. 124-137, 2014.

[Zurita, 2015] Zurita, G., Pino, J.A., Baloian, N. Supporting Smart Community Decision Making for Self-governance with Multiple Views. Lecture Notes in Computer Science 9454, pp. 134-143, Springer, 2015.