

Map Matching for the public transport network in Xalapa City

Map Matching para a rede de transportes públicos na Cidade de Xalapa

DOI:10.34117/bjdv8n5-586

Recebimento dos originais: 21/03/2022

Aceitação para publicação: 29/04/2022

Abraham Toriz-Cruz

Bacharelato em Matemática

Institution: Universidad Veracruzana, Facultad de Matemáticas

Address: Unidad Arco Sur, Paseo 112, Nueva Xalapa, C.P. 91017, Xalapa, Ver, México

E-mail: categulario@gmail.com

Porfirio Toledo

Doutoramento em Ciência, Universidad Veracruzana

Institution: Facultad de Matemáticas

Address: Unidad Arco Sur, Paseo 112, Nueva Xalapa C.P. 91017, Xalapa, Ver, México

E-mail: ptoledo@uv.mx

Ligia Quintana-Torres

Doutoramento em Matemática

Institution: Universidad Veracruzana, Facultad de Matemáticas

Address: Unidad Arco Sur, Paseo 112, Nueva Xalapa, C.P. 91017, Xalapa, Ver, México

E-mail: liquintana@uv.mx

ABSTRACT

In Xalapa, Veracruz, Mexico, there was not a correct description of the more than 100 public transport bus routes that existed, for which a collective event was held to collect Global Positioning System locations. The trajectories obtained were not adjusted to the layout of the city's public road network, due to measurement errors caused by the variety of devices used for tracking. To recover the correct trajectories, a Map Matching algorithm was used based on the optimization of paths in graphs whose nodes belong to the public roads of the city. This work presents the description of the data collection methodology and the construction of the graphs for the implementation of the algorithm that finally recovered the real routes.

Keywords: global positioning system, graph theory, matching pursuit algorithms, mathematical modelling.

RESUMO

Em Xalapa, Veracruz, México, não houve uma descrição correcta das mais de 100 linhas de autocarros de transporte público que existiam, para as quais foi realizado um evento colectivo para recolher locais do Sistema de Posicionamento Global. As trajectórias obtidas não foram ajustadas ao traçado da rede rodoviária pública da cidade, devido a erros de medição causados pela variedade de dispositivos utilizados para a localização. Para recuperar as trajectórias correctas, foi utilizado um algoritmo de correspondência de

mapas com base na otimização dos percursos em gráficos cujos nós pertencem às vias públicas da cidade. Este trabalho apresenta a descrição da metodologia de recolha de dados e a construção dos gráficos para a implementação do algoritmo que finalmente recuperou os trajectos reais.

Palavras-chave: sistema de posicionamento global, teoria dos gráficos, algoritmos de busca de correspondência, modelação matemática.

1 INTRODUCTION

Different Map Matching algorithms are proposed to solve the problem of finding the best route within a road system that matches a succession of locations known in advance [1]. These locations are generally obtained by devices that use GPS (Global Positioning System) and the road system is represented as a digital graph whose edges correspond to passable avenues where a direction is established.

The accuracy of the measurements made by GPS systems is affected by the quality of the device and by other recurring factors such as weather and atmospheric conditions and even many times by some obstacles such as buildings, cables, trees or mountains; therefore, when a set of this type of measurements is obtained, the problem of data imprecision is always faced.

The Map Matching algorithms that have been used to solve problems in which there are few or imprecise measurements are mainly characterized by being based on geometric, topological or probabilistic arguments [2]. In this work, an improved probabilistic method based on the hidden model of Markov [3] has been implemented to define the routes followed by the passenger buses of Xalapa City. A group of project collaborators took charge of taking GPS locations with their cell phones while they traveled the entire route of each bus.

The devices used were different and there were no rules on when to take the measurement. The set of measurements obtained to define each route showed considerable imprecision.

The model presented by [3] was adapted to the data we had, and very good results were obtained since the routes were recovered.

1.1 DEMOGRAPHIC DATA OF XALAPA CITY

The Inter-census Survey 2015 [4] in Mexico, was carried out by the National Institute of Statistics and Geography (INEGI) in the middle of the period between the

2010 and 2020 Census. According with this survey, the states of the Mexican Republic with the largest population were the State of Mexico, Mexico City and Veracruz, which had 16,187,608; 8,918,653 and 8,112,505 inhabitants, respectively.

While, at the state level [5], the three cities with the largest number of inhabitants in the state of Veracruz were: Veracruz, Xalapa and Coatzacoalcos, with 609,964; 480,841 and 319,187, respectively.

Xalapa City is the capital of Veracruz State. According to [6], it has an area of 124.38 km², that represents 0.17% of the total territory of the State, in 2015 the city had a population density of 3,858 inhabitants per square kilometer. It is located on the eastern foothills of the Cofre de Perote, so its ground is uneven. Its average altitude is 1,420 masl, with a minimum of 1,392 masl and a maximum of 1,522 masl in the Cerro de Macuiltepetl, which is almost in the geographic center of the city; there are other important elevations in Cerro de Acalotepetl and Cerro Colorado.

The INEGI also states that, in the same year, there were a total of 129,786 registered cars in circulation in Xalapa City, 1,427 were for official use, 7,524 for public use and 120,835 for private use. In addition, 1,631 of passenger trucks were registered in the city, 32 were for official use, 1,346 for public use and 253 for private use. To all the above we must add 41,431 trucks and cargo trucks and 12,606 registered motorcycles.

In total, there were 185,454 motor vehicles registered in the city, which implies that there was one vehicle in circulation for every 2.59 people.

The previous amounts have been increasing, because according to the projections of the National Population Council, the municipality of Xalapa would reach 513,443 inhabitants in 2020.

Considering also that vehicles from the metropolitan area circulate in the same city, made up of the municipalities of Xalapa, Banderilla and Tlalnehuayocan, there is an environment of intense traffic, especially at peak hours. Therefore, it is extremely important to have an efficient information about the public transport system.

1.2 CONTEXT OF THE PROBLEM

The public transport service in the city is concessioned to the private sector and does not offer the user information regarding the routes on which the buses circulate or the itineraries they have, which are usually very irregular. In addition to the above, the fact that some drivers make stops at unauthorized sites should be considered.

The growth of the population and the excess of private cars on narrow public roads make the transfer of users over long distances a problem, so it is necessary to have information that allows a reordering of public transport routes, and the travel time can be shortened.

Although the municipal government and private concessionaires may have information on public transport routes, it is scarce and partial, and it can only be consulted by these actors, without the possibility that the user public has free access for its use.

In order to improve the use of public transport in Xalapa City, at the initiative of the “Codeando Xalapa” group, the “Mapaton Xalapa 2016” project was launched in collaboration with the H. Xalapa City Council. This project consisted of a mapping of public transport routes carried out with the participation of civil society, the municipal government, higher education institutions and private companies, in a relatively short time. Free software tools from Open StreetMap, Mapillary and Transint Wand were used.

The objective of the project was to obtain the location and carry out the geo-referenced route, obtaining GTFS (General Transit Feed Specification) data for the route layout of public bus transport dealers from Xalapa City, in order to generate information that is open use to manipulate, export and represent them in maps, infographics, applications, among others, through the use of open data and open-source platforms.

The information collected would serve the public transport user to know precisely the different routes in the city as well as their schedules.

2 METHODOLOGY FOR DATA COLLECTION

The methodology followed in the “Mapaton Xalapa 2016” project is documented in [7], so that it can be replicated elsewhere.

The participating actors were the civil society organizations, companies with social responsibility, specialists in transportation, sustainable development and urban planning in addition with higher education institutions and the municipal government of Xalapa.

2.1 PRELIMINARY DATA ORGANIZATION

Contact was made with the Municipal Government of the City and concessionaire entrepreneurs to search for existing information on public transport routes, identifying the start and end points of the routes, the modal exchange centers and points of greatest concentration. The above with the purpose to have a point of reference and validation of

the data collected in addition to allowing the establishment and organization of work areas.

2.2 ORGANIZATION OF VOLUNTEER BRIGADES

Collaboration agreements were established with government institutions, higher education institutions with careers related to urban planning, engineering, architecture and systems; as well as with public and private sponsors which served to publicize the event.

For the recruitment of the volunteers, a registry was carried out, requesting contact details to specify the work mechanics and train them in the use of the information collection tools.

To participate it was necessary to be over 18 years old, have a smartphone or tablet whose operating system is Android 4.1 or higher with GPS enabled and with a camera.

Download the MapMap application developed by Codeando Xalapa and based on Transit Wand, and fill out the registration form on the web (<https://mapaton.org/mapaton-ciudadano-xalapa/>). Five hundred fifty registrations were received to be part of the project, mostly students who use public transportation.

2.3 WORK ORGANIZATION

The mapping was carried out starting at 6:00 hours on September 13, 2016, and until 23:00 hours on October 26, 2016.

Possible routes to map were established and the brigades were organized appointing a route coordinator and two people who would carry out the route through the mobile application.

The 80% of the volunteers were university students interested in civic participation and innovation, who attended the presentations of the project made in advance on campus or in civil organizations.

2.4 MAKING THE PLOT

It was tried that two people collected the information of the same route to corroborate the line, it was also seen the importance of repeating the survey in the 3 periods of the day, to identify not only the routes and the units but also the mobility time due to high, medium and low traffic flow: from 6:00 am to 9:00 am; from 9:00 am to 16:00 pm and finally from 19:00 pm to 23:00 pm.

For the Mapaton Xalapa 2016 exercise, the MapMap application allowed the geo-referencing of information, with a GTFS data structure for mapping public transport.

This application is a digital tool published on the website of the Code for Development initiative of the Inter-American Development Bank, which allows creating geo-referenced traces without the need to consume data from the mobile device. The information it collects is the following:

- Number of passengers getting on and off at each stop.
- Register of stops made by the bus.
- The trace each route.
- Start and end time of the tour.
- Photograph of the unit that makes the journey.

At the end of the tour, the user downloaded the information to the database as soon as the device was connected via WiFi to the Internet, using the section of the application indicated for that purpose.

2.5 DATA PROCESSING

Once the data was concentrated in the MapMap storage infrastructure, it was possible to download it in two formats (.shp, .csv). The information in Shapefile (.shp) format went through a process of conversion to Geojson format.

Immediately a problem was identified, the poor signal reception due to the use of various devices affected the geo-tagging, which motivates to carry out a Map Matching process on the city's public road network, this implementation is presented in this work.

The data that support the findings of this study are openly available in “Datamx por Codeando México” at <http://datamx.io>, id number f6332853-ea92-47ae-a7e1-093b7b97d51b.

According to the “Dirección de Buen Gobierno” and special projects of the Municipal Xalapa Council, information was obtained from 150 routes, 75 circuits and more than 19 thousand pedestrian photos.

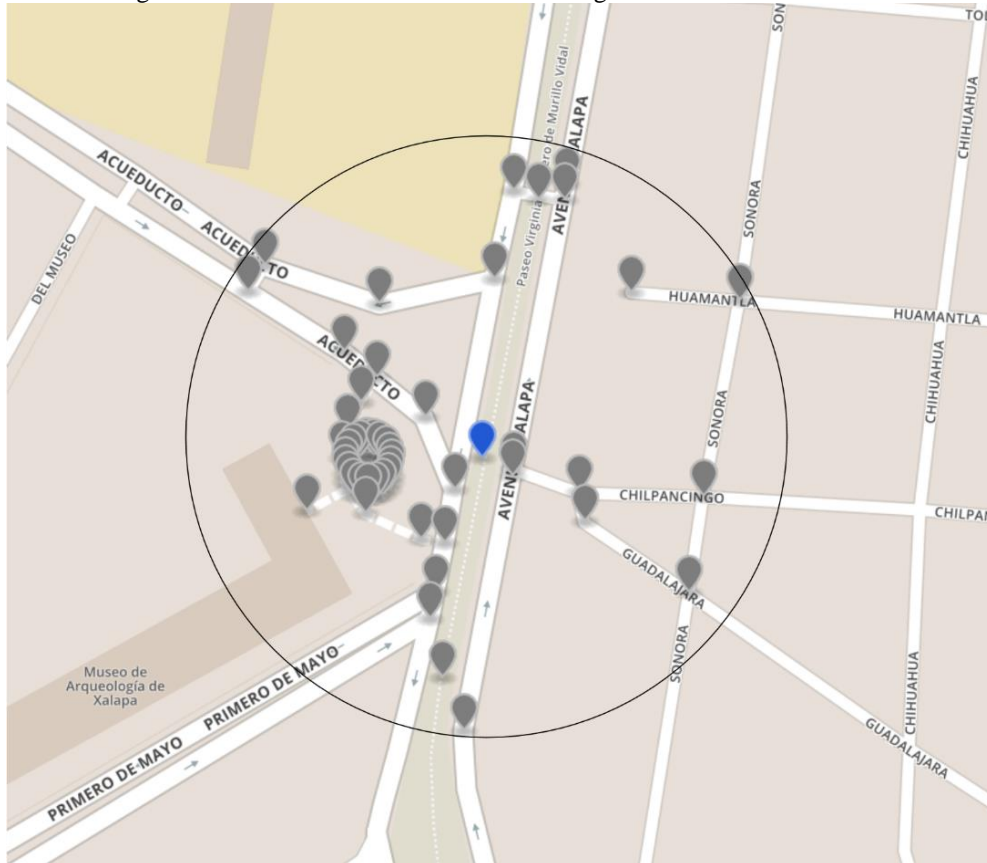
3 ALGORITHM

A Map Matching algorithm was used to determine which streets the vehicle was most likely to have traveled on and to generate a GPS trace that defined the route of each bus. The algorithm was raised by [8] and modified [3].

The street information was extracted from OpenStreetMaps and was used to generate the images (© OpenStreetMap contributors).

Consider the GPS trace of a bus as a succession of positions p_1, p_2, \dots, p_n and n the amount of them. For each point p_i we obtain the set C_i of streets that are in a radius γ of p_i , see Fig. 1.

Fig. 1. The blue point corresponds to the GPS position p_i . Inside the circle with center p_i , the gray points determine the changes in the behavior of the streets that belong to the circle.



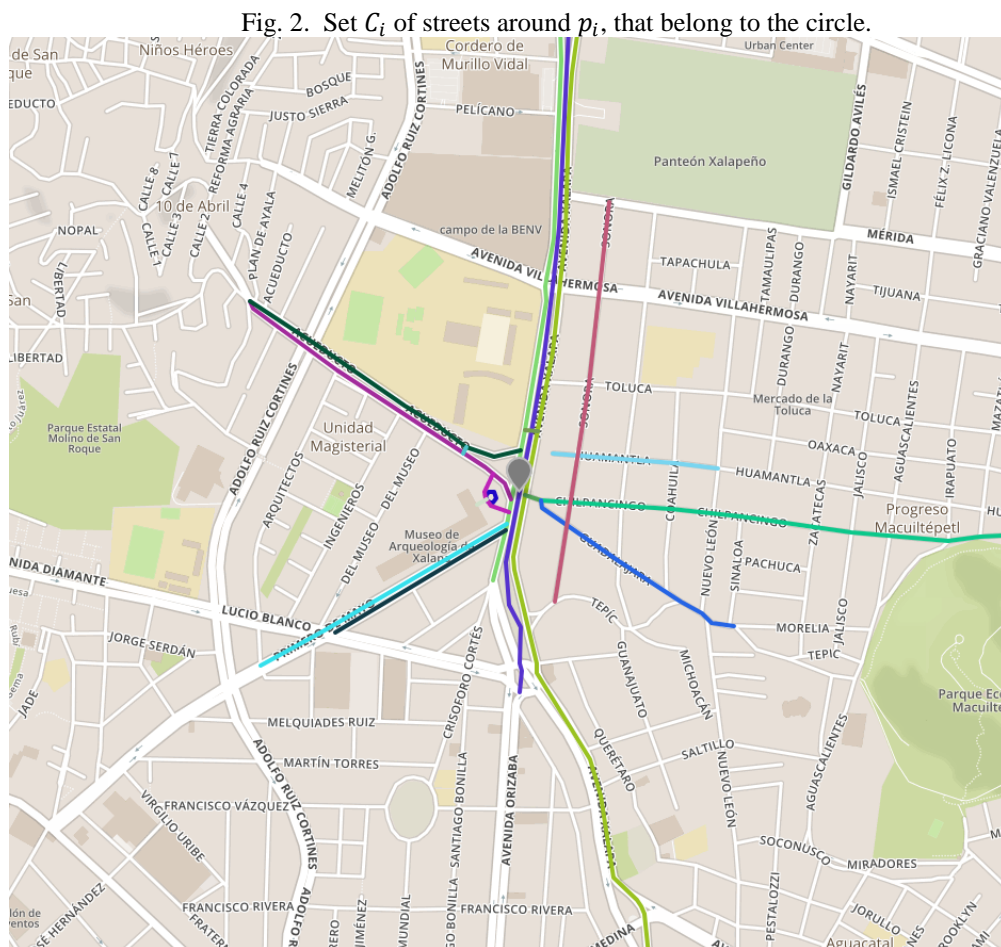
Between every two subsequent points p_i and p_{i+1} all pairs (a, b) with $a \in C_i$, $b \in C_{i+1}$ will be considered. Now we assign a cost to each pair using a cost function $f(p_i, a, p_{i+1}, b)$ that considers:

- The great-circle distance (orthodromic distance) between p_i and the closest point from a .
- The great-circle distance between p_{i+1} and the closest point of b .
- The great-circle distance between p_i and p_{i+1} .
- The shortest path between the point closest to p_i from a and the point closest to p_{i+1} from b . This in the graph of city streets.

3.1 DIRECTED GRAPH

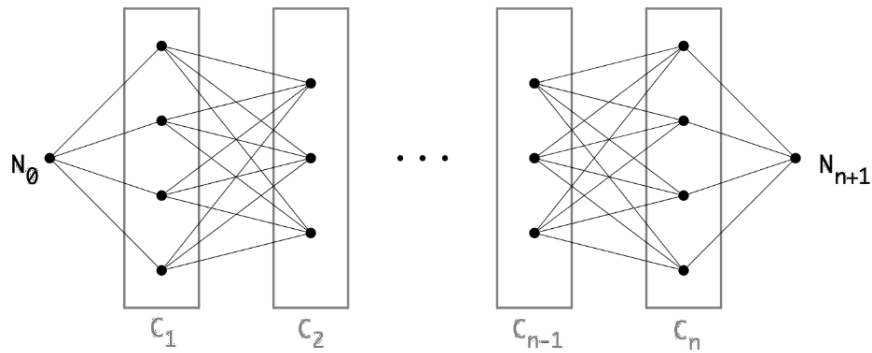
The directed graph G is defined as follows:

- Construction of the vertices:
 - We insert two vertices N_0 and N_{n+1} , with n the number of GPS positions in the trace.
 - For each p_i with $i \leq n$ we insert the streets $c \in C_i$ as vertices in graph G , see Fig. 1 and 2.



- Construction of the edges:
 - We draw an edge with weight 0, origin at N_0 and directed to c for each $c \in C_1$.
 - We draw edges in graph G , starting from each $a \in C_i$ with direction to each $b \in C_{i+1}$ and weight considered by the cost function f , for $i = 1, 2, \dots, n$, see Fig. 3.
 - We draw an edge with weight 0, origin in c and direction to N_{n+1} for each $c \in C_n$.

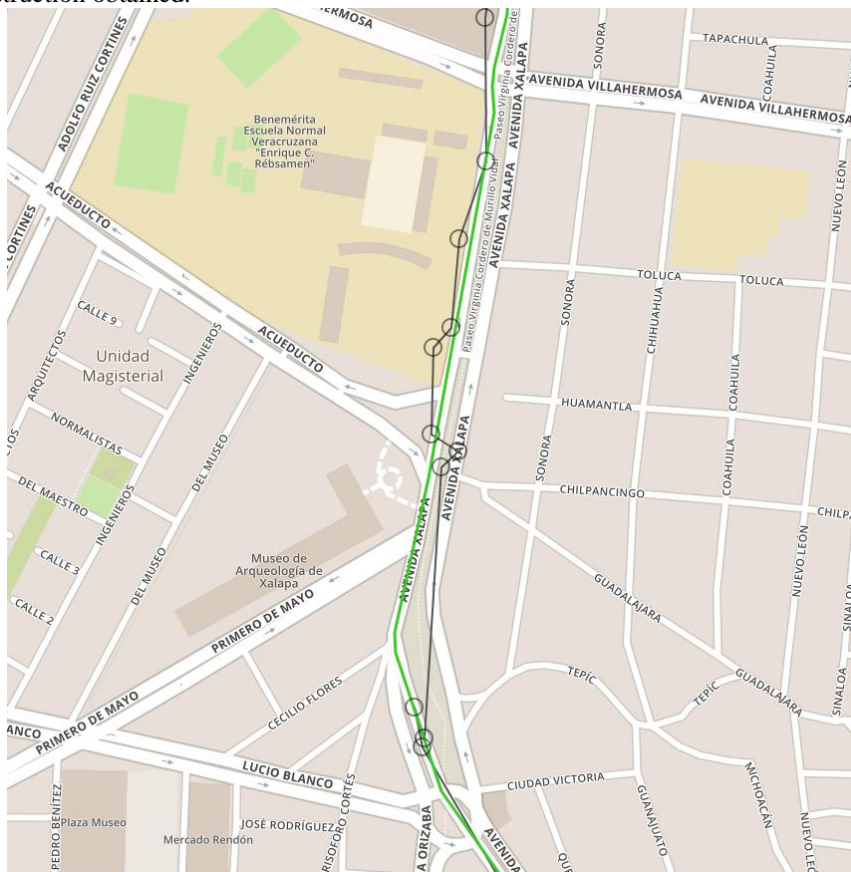
Fig. 3. Set of edges of the graph G .



3.2 BUS ROUTE

Finally, the minimum cost path is taken between N_0 and N_{n+1} , which will pass through the streets $c_k \in C_i$, for $i = 1, 2, \dots, n$. See Fig 4.

Fig. 4. The black line is determined by the set of GPS positions p_1, p_2, \dots, p_n . The green line corresponds to the reconstruction obtained.



4 DISCUSSION

The following difficulties were detected during the implementation of the algorithm.

4.1 THE RADIUS Γ

The search radius γ directly affects both the performance of the algorithm and its effectiveness. Larger radii can allow better results but also increase runtime considerably. A good value of γ must be related to the accuracy of the GPS device that generated the trace.

4.2 THE COST FUNCTION F

The hardest part of the algorithm is finding a balance between the weight we assign to:

- The distances between the obtained GPS points and the surrounding streets and
- The distance between the GPS points and the shortest route between these streets.

If much importance is given to the proximity of the streets to the GPS points and little to the shortest route, results can be obtained where large detours are included. If, on the other hand, the shortest route between the streets is overvalued, the answer could include segments not very close to the original points.

A good cost function must balance the possibility that a vehicle on the street has generated a certain GPS point and the difference in the distances of two consecutive GPS points and the shortest route between the streets associated with the two given GPS points.

A good cost function must balance the possibility that a vehicle on the street has generated a certain GPS point and the difference in the distances of two consecutive GPS points and the shortest route between the streets associated with the two given GPS points.

An example of a viable cost function that was used to match the bus routes is as follows: let p_i and p_{i+1} be two consecutive GPS points, g_i and g_{i+1} be the closest node to a and b in the street graph to each GPS point, and L be the length of the shortest path in the street graph between them,

$$f(p_i, a, p_{i+1}, b) = A \cdot \log|L - d(p_i, p_{i+1})| + B[d(p_i, g_i) + d(p_{i+1}, g_{i+1})].$$

In this case the logarithm allows us to give more importance to the distance between GPS points and street graph points than to the difference between the great circle distance between two consecutive GPS points and the shortest path connecting them in the street graph.

4.3 HEURISTICS

A fundamental element at the time of implementation is to consider some heuristics that would rule out implausible cases in real life, such as paths that include take a street, go one block and go back on it, etc.

5 CONCLUSION

As we know, the Map Matching is a very useful tool to solve real life problems in which data is scarce and imprecise as in our case. By implementing Map Matching algorithm proposed by [3] in this context, the routes of the passenger buses in Xalapa City were recovered.

The difficulties that we faced during the implementation of the algorithm were mainly due to the topography of the city, the misaligned streets and the variety of devices that originated different types of data. To solve these difficulties, it was necessary to adjust the algorithm parameters and find an adequate cost function. The impact produced by the selection of these parameters and their relationship with the precision of the GPS devices used is a subject that deserves to be studied.

REFERENCES

- [1] M. Kubicka, A. Cela, H. Mounier, and S.-I. Niculescu, “Comparative Study and Application-Oriented Classification of Vehicular Map-Matching Methods,” *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 2, pp. 150–166, 2018, doi: 10.1109/MITS.2018.2806630
- [2] M. A. Quddus, W. Y. Ochieng, L. Zhao, and R. B. Noland, “A general map matching algorithm for transport telematics applications,” *GPS Solut.*, vol. 7, no. 3, pp. 157–167, Dec. 2003, doi: 10.1007/s10291-003-0069-z
- [3] H. Koller, P. Widhalm, M. Dragaschnig, and A. Graser, “Fast Hidden Markov Model Map-Matching for Sparse and Noisy Trajectories,” in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Sep. 2015, pp. 2557–2561, doi: 10.1109/ITSC.2015.411
- [4] Instituto Nacional de Estadística y Geografía (INEGI), *Principales resultados de la Encuesta Inercial 2015 Estados Unidos Mexicanos*. México: Instituto Nacional de Estadística y Geografía, 2015.
- [5] Instituto Nacional de Estadística y Geografía (INEGI), *Principales resultados de la Encuesta Intercensal 2015. Veracruz de Ignacion de la Llave*. México: Instituto Nacional de Estadística y Geografía, 2016.
- [6] Instituto Nacional de Estadística y Geografía (INEGI), *Anuario estadístico y geográfico de Veracruz de Ignacio de la Llave 2016*. México: Instituto Nacional de Estadística y Geografía, 2016.
- [7] Codeando Xalapa, “Guía Metodológica Mapatón Ciudadano.Org,” 2018. [Online]. Available: <https://www.mapaton.org>.
- [8] P. Newson and J. Krumm, “Hidden Markov map matching through noise and sparseness,” in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09*, 2009, p. 336, doi: 10.1145/1653771.1653818