

Bouallouche, Dalicia; Vioix, Jean-Baptiste; Millot, Stéphane; Busvelle, Eric

Conference Paper

Address Matching Using Truck Tours Feedback

Provided in Cooperation with:

Hamburg University of Technology (TUHH), Institute of Business Logistics and General Management

Suggested Citation: Bouallouche, Dalicia; Vioix, Jean-Baptiste; Millot, Stéphane; Busvelle, Eric (2015) : Address Matching Using Truck Tours Feedback, In: Blecker, Thorsten Kersten, Wolfgang Ringle, Christian M. (Ed.): Operational Excellence in Logistics and Supply Chains: Optimization Methods, Data-driven Approaches and Security Insights. Proceedings of the Hamburg International Conference of Logistics (HICL), Vol. 22, ISBN 978-3-7375-4058-2, epubli GmbH, Berlin, pp. 391-418

This Version is available at:

<https://hdl.handle.net/10419/209293>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-sa/4.0/>

**Dalia Bouallouche, Jean-Baptiste Vioix, Stéphane
Millot and Eric Busvelle**

Address Matching Using Truck Tours Feedback



Address Matching Using Truck Tours Feedback

Dalicia Bouallouche, Jean-Baptiste Vioix, Stéphane Millot and Eric Busvelle

When researchers or logistics software developers deal with vehicle routing optimization, they mainly focus on minimizing the total traveled distance or time of the tours, and maximizing the number of visited customers. However, in real transporter situations, the actual data received is often of bad quality, particularly the irrelevance of addresses and address geocoding errors. Therefore, trying to optimize tours with impertinent customers' GPS-coordinates, which are the most important input data for solving a vehicle routing problem, will lead to an incoherent solution, especially if the locations of the customers used for the optimization are very different from their real positions.

Our work is supported by a logistics software editor Tedies (2013) and a transport company Upsilon (2009). We work with the company's real truck routes data to carry our experiments. The aim of this work is to use the experience of the driver and the feedback of the real truck tours to validate and correct GPS-coordinates to the next tours. Our method significantly improves the quality of the geocoding.

This study shows the importance of taking into account the feedback of the trucks to gradually correct address geocoding errors. Indeed, the accuracy of customer's address and its GPS-coordinates plays a major role in tours optimization. This feedback is naturally and usually taken into account by transporters (by asking drivers, calling customers, ...), to learn about their tours and bring corrections to the upcoming tours. Hence, we develop a method to do most of that automatically.

Keywords: Driver Experience Feedback, Geocoding Correction, Real Truck Tours, Address Matching

1 Introduction

Our study is in the context of a project of route optimization of collection/delivery vehicles. We work with a software development company that proposes transport and logistics software keys for transporters. During the development of this project, we noticed an important problem which does not fall in the scope of the vehicle routing optimization. However, it is an issue that has much impact on the quality of the optimization in real world situations. The problem is the inaccuracy of the customers' positions (GPS coordinates). The optimization of the vehicle tours, and other studies, strongly depends on the location of customers on the map. Indeed, transporters and transport and logistics software developers complain about the fact that geocoders do not always give pertinent address geocoding. Geocoding is a crucial step ahead many GIS (Geographic Information System) projects. Hence, transporters have to hire employees to check or correct all GPS coordinates of their customers' addresses. This takes a big part of their time. Transporters and transport & logistics software editors assert that geocoding errors coming from automatic geocoders and human errors could lead tours optimization or any other GIS project to failure.

A related problem is the address writing errors issue. Hence, a badly written, incomplete, or inaccurate (human errors) address, necessarily implies geocoding errors.

As well, an obvious geocoding issue is that the mapping used by the geocoders is not uploaded regularly. Also, in the case of larger companies or shops, their geocoding positions are usually different from the place where the vehicle should make its deliveries. For instance, in airports, a geocoder

returns GPS coordinates which do not necessarily indicate the freight area.

Clearly, the correctness of upstream customers' addresses and a rigorous address geocoding are crucial conditions to carry out the vehicle routing optimization and other GIS projects. Benefits resulting from optimization may be transformed in extra costs, painful for the driver, and may lead to give up the optimization software assistance. Foote and Huebner (2000, sec.1), claim that spatial data errors, inaccuracy, and imprecision can "make or break" a GIS project.

The aim of this work is to use the experience of the driver and the feedback of the real truck tours in order to first validate the GPS coordinates of an address which were pretty well geocoded, and, second, bring a geocoding correction to the rest of badly geocoded addresses. To achieve this, we rely on the real truck tours of the transporter. We retrieve tours data from truck GPSs to use the feedback of the trucks for validating/correcting our spatial data. Obviously, any GPS with data logging capabilities, including smartphones, may be used for this application.

In the next session, we will discuss on related works treating our problematic. In section 3, we will state the required tools to carry out our experiments. We will, thereafter, explain our algorithm in section 4. In section 5, we will present some results of our method. Finally, we will give in section 6 a conclusion and a preview of the further work of this paper.

2 Background and Discussion

Address geocoding is considered in many fields. There is an increasing number of applications that rely on GIS. For instance, transport and logistics field, the field of health (Lewis-Michl et al., 1996; English et al., 1999; Rushton and Lolonis, 1996; Anselin, 1995; Geschwind et al., 1992; Kulldorff and Nagarwalla, 1995), etc.

Cayo and Talbot (2003, para.10) assert that very limited published information exists on address geocoding errors in automated street level geocoding. In the context of our work, no published papers were found about the problem that we had noticed in transporters real case situation. Especially, the problem of geocoding errors in the context of vehicle routing problem optimization, traveling salesman problem or other studies related to the transport and logistics field were not raised. Researchers of this field, mostly, consider that all the upstream data (in our case, customers positions on the map, or spatial data in general), are free from errors and all their optimization methods might work and give the same results as in theoretical case simulations.

Furthermore, we found some related research in other fields (different from transport and logistics). Most of them are related to the environment and health areas. Cayo and Talbot (2003) studied positional error in automated geocoding in the field of Environmental and Occupational Epidemiology. Authors match between geocoded points and true known locations. Authors acquired residential addresses from the NYSORPS (for New York State Office of Real Property Services) and evaluate GPS coordinates errors caused during automated address geocoding. Authors use the distance between each geocoded point and its true location, and then measure the

variation of these errors depending on population densities (urban, suburban and rural). Their conclusion is that errors increase as population density increases. Our method, also, deals with matching geocoded points with customers' addresses. However, in contrast to the study of Cayo and Talbot (2003), we do not rely on the true locations of these addresses because we don't know them. Instead, we construct them according to the real path of the truck tour. In the same paper, the authors use a method of geocoding using property parcel data (NYSORPS) instead of the traditional linear interpolation method. In this last, the interpolation error increases as the street segments are longer. In fact, Levine and Kim (1998, p.563) conclude that geocoding errors of interpolation algorithms vary with the street segment length and urban areas typically contain shorter segments compared to rural areas. Also, the geocoding software that uses interpolation algorithms supposes uniform intervals between street numbers along a street segment (Telogis, 2015; Google Maps, 2015; Cayo and Talbot, 2003, para.30), which is not the case in reality. Indeed, when addresses are not evenly spaced along a street, the probability to get an interpolation error increases.

An alternative solution is possible when the traditional geocoding methods do not give expected results. The use of property parcel points provides greater positional accuracy and reduces geocoding errors. In fact, individual house locations and separations are more accurate in property parcel data than in TIGER (Topologically Integrated Geographic Encoding and Referencing System) based files, where parcel centroids are rarely at the exact locations of the houses (Cayo and Talbot, 2003). In addition, parcel data is updated annually for tax purposes, so match rates are improved, unlike

with commonly used street centerline files which are less regularly updated (Cayo and Talbot, 2003, para.34).

Various researches are based on match rate statistics (Gregorio, Cromley, Mrozinski and Walsh, 1999; Howe, 1986; Levine and Kim, 1998; Yu, 1996). Authors found a positive dependency between computerized geocoding rates and population size and urbanity (Howe, 1986, p.1460). This could be justified by the fact that address information, in more densely populated areas, is often more complete in street reference and commercially enhanced files (Gregorio et al., 1999, p.177). Hence, geographic differences can alter study results.

Zinszer et al. (2010) examine the impact of address geocoding errors through the estimation of the spatial distribution of the disease. The authors evaluate address geocoding errors for a selected reportable disease in a large urban center in Canada (Zinszer et al., 2010, p.163). Researchers use an address verification algorithm on extracted data for all notifications of campylobacteriosis from the Montreal public health department to determine the accuracy of the residential address for each case and to suggest corrections for invalid addresses. Authors estimate address errors types as well as the resulting geocoding errors. For this, researchers calculate the distance gap between the original address and the correct address, like in Cayo and Talbot research (2003, para.2 and 11), as well as changes in disease density (Zinszer et al., 2010, p.164).

Communicable disease surveillance in public health practice has an equivalent problem with transporters on the received IDE (Information Data Exchange). Many errors are introduced on the informed addresses. Indeed, the way information (addresses and medical information) is carried from

diverse sources (hospital registries, laboratories reports, physical offices...) to the central databases of the public health department varies. It goes from a fully automated transmission to a fax to which manual data entry are added (Zinszer et al., 2010, p.164). The objective of the study is to examine address errors for a selected reportable disease in a large urban center in Canada and to assess the impact of identified errors on the estimation of the spatial distribution of the disease. To do that, Zinszer et al. (2010, p.164) compare the collected addresses of the public health dataset to the PCAD (the Postal Code Address Data) file of Canada Post. The authors determine if the street name, street number and postal code of an address from the public health data correspond to a correct street segment in the PCAD file. The resulting matchings between the public health dataset addresses and the PCAD are classified as exact matches, recoverable addresses or unprocessable addresses (Zinszer et al., 2010, p.165 and 166). The researchers based their study on an official file of good quality addresses.

The difficulty in our study is that we only work with data (addresses) received from the IDE. Thus, customers requesting the delivery/collection services write the received addresses. Then, we build our customers' addresses real positions based on the real trucks data tours to compare with the received customers addresses from IDE.

A recently opened project, named BANO (for Base Adresses Nationale Ouverte) (2014), shows the importance of handling the problem of poor quality addresses and inaccurate geocoding. It is an open source project of OpenStreetMap® (2015) France. The objective of this project is to collect data (addresses) from hundreds of contributors, opendata sources, land

registry, etc., to construct from these gathered addresses the most complete and correct version of each address, then, match each address on the corresponding street in the map.

3 Experimentation Tools

3.1 The Context of our Study

The transporter includes all transport (collections and deliveries) of goods that require at least one stop on a platform to a sorting operation, collecting, or unbundling. It receives the packages to deliver from other transporters early in the morning, between 2:00 and 4:00AM. Meanwhile, received packages have to be sorted and dispatched after their recipients' addresses are geocoded and validated/corrected. Transport operators must do it very quickly so the drivers can make their deliveries, generally between 5:00 and 6:00AM. In average, the vehicles delivers 305 packages per day for 110 addresses, among them 30% of new addresses, with a fleet of 10 vehicles. The customers have to be delivered as soon as possible, by taking into account their collection/delivery constraints (time windows, collection/delivery equipment, etc.). Practically, the covered region by the transporter is within a radius of 100 km around the freight center. Each truck delivers some cities along its route. The addresses may vary from day to day and drivers may spend a lot of time with unknown addresses (especially new or inexperienced drivers).

The transport company performs other kind of transportations, like batch transport. However, a batch of packages is carried out permanently from a

single source to a single customer, so we are not interested in this simpler case.

Our tool helps the transport operators in the address geocoding phase (after receiving the addresses of delivery and before dispatching the corresponding packages). The validation and correction of address geocoding is done automatically in upstream. Thereby, it saves time for transport operators, and they can focus on the remaining tasks of the transport operations.

3.2 GPS Data Retrieving

The vehicles fleet of the transporter is equipped with TomTom® GPS units (2015) that continuously send trucks data to a remote database. It sends around one data message every 10 seconds. This message includes data related to the message (id, recording time and type, ...), data related to the truck (vehicle registration number, driver id, driver name, latitude, longitude, speed, odometer, ...), data related to the tachograph, FMS (Fuel Management System), vehicle order messages, etc. We recover these data from the database through TomTom Web Service. We select data which would be useful for our research and process it before being stored in a local database for use.

3.3 The Road Network

To carry out our work, we lean on OpenStreetMap® (2015) road network of the region of Burgundy (France). We only select this region because the transporter performs its collections/deliveries within this area. The road network is mainly composed of ways and nodes. We must know the

topology of the road network graph to, particularly, get the road intersections that will be used to identify the stops at intersections and differentiate them from the stops for deliveries/collections. For more details, refer to section 4.

3.4 Details on Address Geocoding

The company uses Google geocoding API and Bing Map Geocode Service API to geocode customers' addresses of its transporters. The geocoding precision is returned with each address geocoding request (Google Maps, 2015; Bing Maps, 2015). These geocoding precisions are saved in the software database and indexed in order to be used for our work. Values are given to addresses according to the precision of their geocoding. Geocoding precision 1 is the "ROOFTOP" geocoded addresses. It indicates that the GPS coordinates returned by the geocoding API are precise. Here, we have an accuracy down to the street address level precision. It is the best geocoding precision. Address geocoding of precision 2 is the "RANGE_INTERPOLATED" address geocoding precision. Here, the geocoder does not know the precise location of the address but knows the location of the street address and two precise points of addresses on the same street address. Hence, the geocoding of our address is done by interpolating between these two precise points. It is a geocoding of medium quality. Precision of geocoding 3 is the "GEOMETRIC_CENTER" precision. It returns GPS coordinates of the center of a region (the city of the address) or the middle of the street of the address (if the street is known). This geocoding

precision is bad. Finally, address geocoding of precision 4 is the “APPROXIMATE” precision. It returns an approximate geocoding result. It is often of very bad quality.

We introduced another address geocoding precision, we denote it geocoding precision 0. It is not returned by any geocoder. It indicates that the geocoding is corrected or done manually by the user (transport operator). In fact, an address could be not geocoded at all if it is not understandable by the geocoder (incorrect, incomplete or badly written address). It is equivalent to geocoding of precision 1 because the user must be very accurate in searching for and assigning GPS coordinates to a customer’s address.

We use these precisions to give priorities for validating and correcting address positions on a map (matching between addresses and real delivery points). Indeed, an address with a geocoding precision of 1 (accurate geocoding) and an address with a geocoding precision of 3 or 4 could not be matched with delivery points in the same manner. More details are given in the next section.

4 Algorithm

Before starting to explain the algorithm, some definitions of basic aspects we use in this work are required. The matching is done between GPS coordinates of customers’ addresses (geocoded using geocoders) and their real positions on the map. We recover the vehicle’s path during its tour. The required data for our work are the speed of the vehicle and its GPS coordinates along its way. We use these data to determine the vehicle’s delivery

points. In fact, to deliver a customer, the vehicle needs to stop. Hence, we detect all the vehicle breakpoints (zero speed). They could be stops on road intersections, biological breaks or delivery/collection services. We suppose that, to deliver a customer, a vehicle takes at least a minimum period of time. Therefore, to determine delivery points on a tour, we select all the breakpoints where the vehicle is stopped at a location for a specified period of time. This downtime is the minimum delivery time (time threshold). It is a non deterministic parameter, hence, we can't entirely rely on this value. A way to select the good breakpoints that we search for, namely, the delivery points, is to get rid of other vehicle breakpoints of a different kind, especially the road intersection stops. Indeed, we identify all the intersections of the road network and eliminate all the breakpoints where the vehicle is close to the intersections. The breakpoints for biological breaks will not disturb us because they are relatively very few.

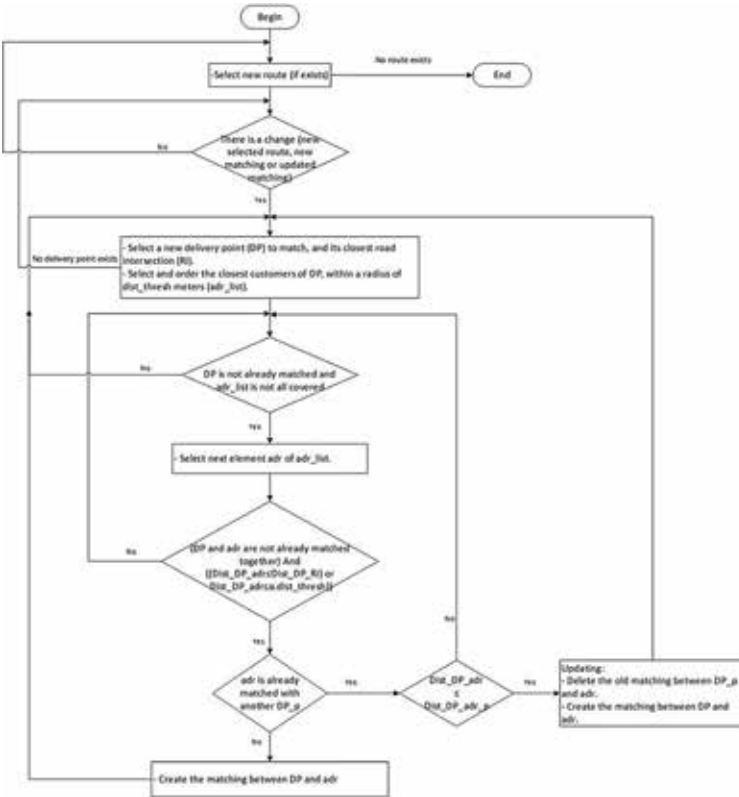
The matching is done route by route. For each route, we require addresses of the customers we had planned to deliver, the vehicle's breakpoints and the road intersections of the area where the vehicle made its trip. Vehicle breakpoints are unique to every real route, unlike customers' addresses where two planned routes (with different vehicles or with the same vehicle on different dates) might have customers' addresses in common. For instance, a customer can be delivered everyday by the same vehicle or with different vehicles, hence on the same area and, then, the same road intersections.

The general scheme of the Algorithm is shown on figure 1. For each vehicle route of the transporter, we select all the vehicle delivery points and their closest road intersections.

For each delivery/collection point and its closest road intersection, we search for all the closest customers' addresses within a specified radius (also stated as distance threshold). We store them in a list and in an increasing order of their distance to the delivery point. Initially, a delivery point is unmatched. While the delivery point has not been matched and we have not covered all the list of its closest customers (selected before), we search for the best customer's address to match with the delivery point. We first choose the first customer of the list (the closest one). If it does not comply

the conditions to be matched with the delivery point, we move to the next customer of the list, and so on.

Figure 1 General algorithm scheme



The selected customer's address is likely to be matched with the delivery point if the delivery point is closer to the customer than to its closest road intersection. There are cases where the selected delivery point and the customer's address could, pretty well, be matched, even if the precedent condition fails. This often happens in situations where the customer position is very close to the road intersection of the breakpoint. To relax our condition, we shorten the distance threshold. Thus, if the distance between the delivery point and the customer's position does not exceed a smaller distance threshold (by multiplying the last specified distance threshold by α , with $\alpha \in]0,1[$), which means that the delivery point is very close to the customer, therefore, the matching is possible.

Before matching, we have to ensure that the delivery point and the address have not been matched yet. Also, if the customer address is already matched with another delivery point, we check if the customer is closer to this new delivery point than to its precedent delivery point. In this case, we update the matching of the customer's address with the new delivery point. Consequently, this will give delivery points that will be free again (not matched). These latter should be matched at the forthcoming matching rounds, with other close and not yet matched customer addresses.

The matching round between delivery points and customers addresses of a route is done until there is no change in the situation of the selected delivery/collection and customers' addresses points, like new or updated matching between a delivery point and a customer's address position.

Our matching algorithm is first applied on customers' addresses that have been geocoded with precision 1 and 0. We prioritize them compared to

other address geocoding precisions because their geocoding is more accurate. After that, we move to match customers' addresses of geocoding precision 2. For matching these addresses, we give a slightly greater value for the distance threshold parameter. In fact, the positions of addresses could be somewhat far from their true location. Then, with a greater value, we can achieve matching these addresses with their corresponding delivery points. Finally, we try to match the rest of potential delivery points with customers' addresses geocoded with precision 3 and 4. Since these latter have bad geocoding precisions, they will be, mainly, far from the truck trajectory. So, matching rate will be very low.

The delivery points that have been matched with addresses with a geocoding precision of 1 or 0 will not be candidates for matching with customers' addresses of other geocoding precisions. Those that have been matched with addresses geocoding precision 2 will not be available for matching with customers addresses with a lower geocoding precisions, etc. Our method is used to automatically validate a big part of the customers' GPS coordinates. This was done manually by the transporter, before implementing our algorithm. In fact, after each back from tour, transport operators check the real delivery points of their customers and ask the drivers to validate or indicate the locations of the customers' addresses with which they had trouble. Generally, drivers have experience and could easily spot the real locations of the customers. At worst, they could ask the way to the customer if they have trouble with finding a customer's address location (they are on the field!).

To release transport operators from this heavy task, we develop our algorithm to do that automatically by detecting the potential delivery/collection points (breakpoints) of the vehicles. However, not all customers' addresses are matched with delivery/collection points. Indeed, for the remaining unmatched customers' addresses, this could be justified either by the fact that a service cancellation order has occurred during the path of the vehicle (the customer's position would be far from all delivery points or the whole ride of the vehicle), but this is still uncommon, or the customer's address point is simply unmatched because of a distance bigger than the fixed threshold. For precautionary measure, these unmatched points have to be matched manually by the transport operators. The rates of the remaining unmatched customers' addresses and delivery points are rather weak. The manual matching is, thus, not really constraining.

5 Results and Discussions

In our experimentations, there are some parameters to initialize. Namely, the maximum distance for choosing the closest customer's address to a delivery point (distance threshold) and the minimum delivery time. The distance threshold is set to 300 meters. In fact, it depends on the area of the deliveries. In urban areas, addresses are near each other, therefore, the delivery/collections points could be also near each other. Then, the distance threshold should be small enough to avoid matching confusion. In contrast to rural areas where addresses are spaced each other. Then, the distance threshold could be big enough not to miss matchings.



Figure 2 Example of a truck route with addresses, breakpoints and matchings

Considering the driver experience, we suppose that the minimum delivery time takes at least 70 seconds. This value might seem to be low but deliveries/collections could be very fast for small packages.

To display our vehicle tours, address points and vehicle breakpoints in a map, we use QGIS® (2015) (a free and open source geographic information system), in which we integrate OpenStreetMap® (2015) electronic map.

Figure 2 shows a vehicle real tour on the map. The solid line is the route crossed by this vehicle. This route line is composed of sequential line segments called Ways in the road network. We construct the path of the vehicle by selecting the Ways where the GPS of the vehicle has sent a data message (including its GPS coordinates) during its tour. We can observe missing segments (ways) on the truck path route. It is explained by the fact that the GPS had missed sending messages when it had passed on these Ways (short way, loss of GPS signal,...). The triangles are the potential truck delivery points (or breakpoints in a general way). As explained in section 4,

they are points belonging to the truck route where the vehicle downtime exceeds a minimum delivery/collection service time. The circles are the customers' addresses positions as they are geocoded and used in the software. For clarity reasons, we select only addresses with geocoding precision 1 and 0. For other precisions, the procedure is the same. The only difference is the parameter settings as stated in section 4. Small circles are the matched addresses and big circles are those that are unmatched. The stars indicate that matching is done between a customer address and a delivery/collection point.

We run our program for matching customers' addresses of precision 1 and 0, with delivery/collection points of a route in figure 2. For this last, we achieve 75% of matched addresses with geocoding precision 1 or 0. We have 9 matched customers' addresses over 12 addresses of geocoding precisions 1 or 0.

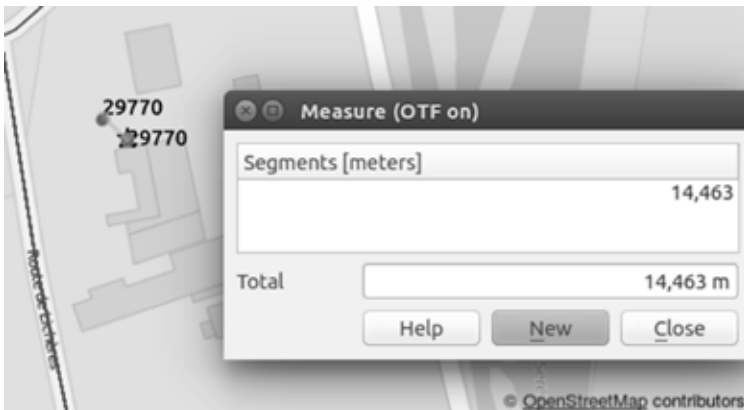


Figure 3 An example of a perfect address matching



Figure 4 An example of an address matching near a road intersection

In figure 3, we show an example of a perfect matching on the route in figure 2. The vehicle has joined the customer at his exact delivery location.

In figure 4, we see that the matching is done between the customer's address and the delivery point, even if the delivery point is very close to a road intersection. It is stated in section 4 that we relax our matching condition when a breakpoint is very close to the customer's address (refer to section 4 for more details).

Figure 5 shows the matching of a customer address of geocoding precision 1 and 0. We see that the position of the customer's address on the map is not precisely positioned because it is a bit far from the breakpoint of the truck (delivery/collection point). This is justified by the fact that the delivery is for this customer, but the freight area is in another special building. In this case, the delivery point and the customer's address are close enough to be matched.

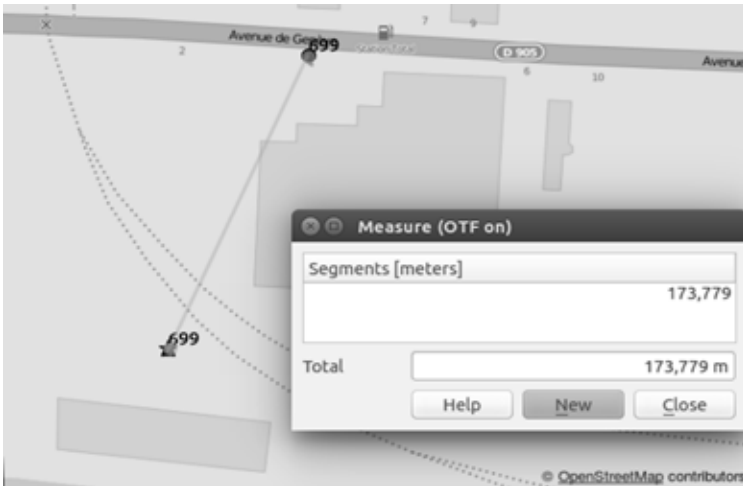


Figure 5 Another example of address matching

Figures 6 and 8 show two unmatched addresses. In figure 6 we have an unmatched address and a breakpoint at 238 meters from this customer's address. However, the matching is prevented because the nearest breakpoint is closest to a road intersection (87.5 meters far from it) (see figure 7). Hence, it is not considered as a delivery point but as an intersection stop. In figure 8, we see that the customer's address is far from the whole vehicle trajectory and at more than 2 km far from the closest breakpoint. The reason why the addresses of figures 6 and 8 are unmatched could, also, be explained by the fact that the addresses of the customers are not accurately geocoded. Indeed, geocoders are not exempt from errors (even with a geocoding precision 1) and human error can't be completely avoided (geocoding precision 0).

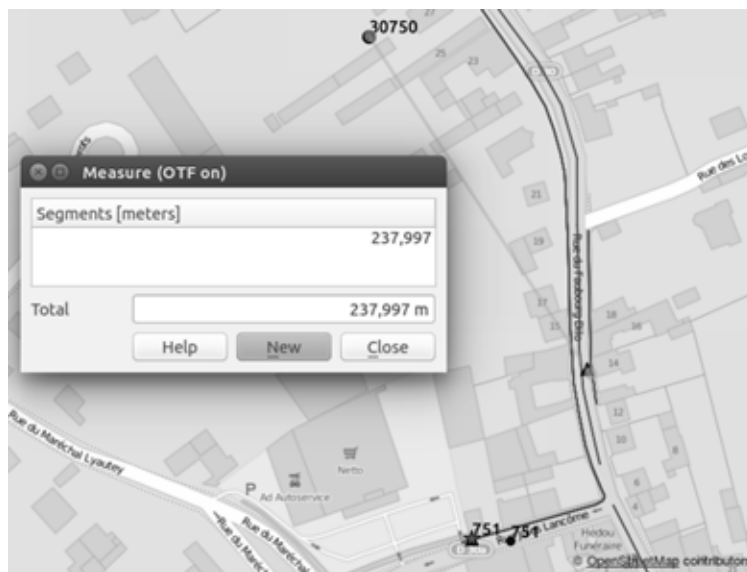


Figure 6 Example of avoided address matching because of the closeness of an intersection to a potential delivery point (breakpoint)

We have selected tours of a vehicle over a month, on which we had run our matching program. We achieve an average of 75% address matching per tour, including all addresses geocoding precisions. We have a total of 122 different delivered addresses by these tours and 60% of them were validated after running our program. Each pair of the vehicle tours could have common customers to deliver. Otherwise, we have 40% unmatched addresses per tour. This remains a rather high rate. It is mainly because of the address geocoding precisions 3 and 4. In fact, geocoding of these precisions are often far outside the vehicle path. Hence, matching could be unsuccessful at this stage. Each matched customer address will be positioned on its

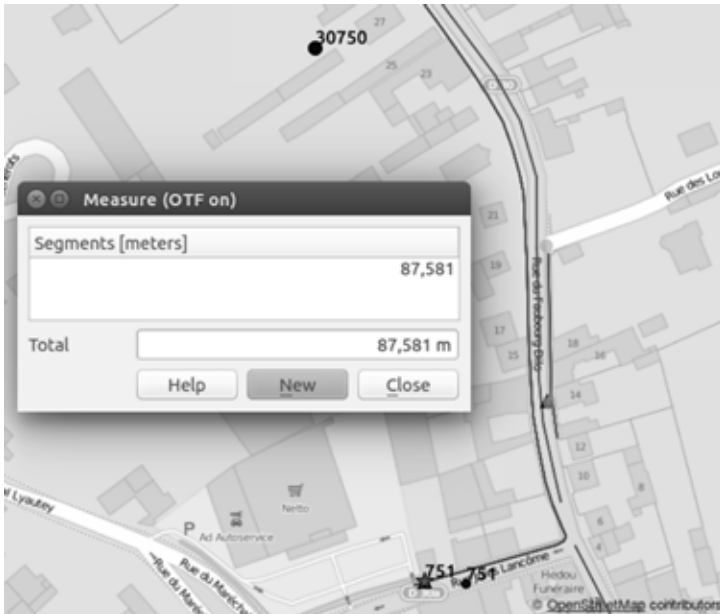


Figure 7 Example of a breakpoint near road intersection

corresponding delivery/collection point. It is considered as the real position (freight area) for serving the customer. The more we have matched good quality addresses (precisions 0, 1 and 2), the easier it would be to match the remaining few bad quality addresses (precisions 3 and 4).

The transporter delivers an average of 110 addresses per day, the transport operators have to check the validity of all the real position (GPS coordinates) of these addresses before sending the drivers to deliver them. To geocode, validate or correct an address, an experimented user spends 3 minutes in average (from 1 minute for a clear and quite simple address,

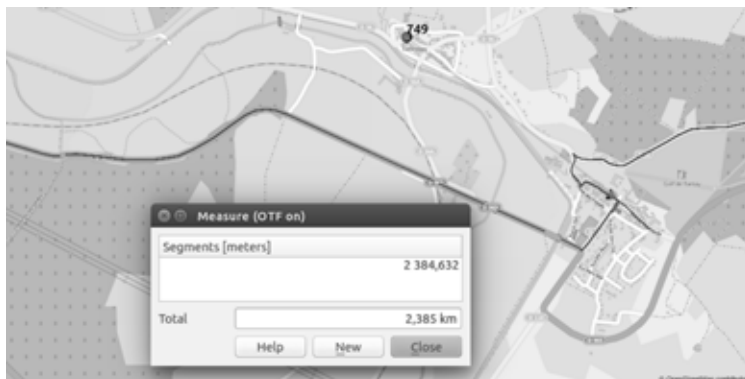


Figure 8 Customer's address out of vehicle path.

to 5 minutes for an unclear (badly written) and difficult to geocode address. It could be more if the transport operator have to ask an experimented driver to help him to find the real location of an address). As we correct an average of 75% of addresses for each tour in upstream, we save for the transporter 75% of the total time of address geocoding correction, in average. Precisely, if the users spend 3 minutes for geocoding each of the 110 addresses, instead of taking 5.5 hours for address geocoding correction, they will take only 1 hour and 22 minutes.

6 Conclusion and Future Work

We have emphasized an important issue that could break a vehicle routing optimization or any other GIS project. In fact, geocoders are not exempt from errors and they can give impertinent GPS coordinates. Also, even with a good geocoding, a poorly written, wrong or inaccurate address (human

errors) can lead to a bad or completely incorrect geocoding. We work with a company of editing transport and logistics software and a transport company. Both assert that they encounter big problems because of this address geocoding issue. We propose a method that takes into account the reality of the ground to correct address geocoding, and this, by means of the real truck routes feedback. Indeed, the best way to have the correct information is to go on the field. This is done manually and unconsciously by the transport operators before implementing our algorithm. In fact, the reflex to seek information from drivers after the return of the vehicles has always existed among transporters employees. Manually considering the feedback of the truck routes for address geocoding correction could become a heavy task. This is why having a tool for doing that automatically is of great use. Our collaborators confirm that.

Our algorithm achieves to match 75% of addresses per tour, in average. Transport operators will do the remaining unmatched addresses manually. Since we have identified the delivery/collection points on the vehicle path, the manual matching remains a fairly simple task.

With our geocoding validation and correction method, when a vehicle makes its tour, for each visited customer, it might have trouble with finding this customer's address (or the good warehouse entry of the customer) at most once. In other words, the vehicle would be wrong at most once for each customer's address (the first time the truck delivers it). As the vehicle stops near the customer for a delivery time, we will have the good GPS coordinates for this address at the next matching round.

We are currently working on enhancing the matching rates, especially for addresses with low geocoding precisions (3 and 4), with which we still have

matching difficulties. Also, setting the distance matching threshold and the minimum delivery time parameters to fixed values is not appropriate. In fact, if the distance between the customer and the breakpoint is one centimeter greater than the distance threshold, matching will not be done. We, then, develop a method to adapt these parameters depending on the geocoding precision, the position of each customer's address compared to the breakpoints, the intersection, etc., in order to make matching decision.

References

- Anselin, L., 1995. Local indicators of spatial association-LISA. *Geographical analysis*, 27(2), pp.93–115.
- BANO, 2014. BANO(Base Adresse Nationale Ouverte), OpenStreetMap. [online] Le projet BANO. Available at: <<http://openstreetmap.fr/bano>>.
- Bing Maps, 2015. Bing Maps Geocode. [online] ProgrammableWeb. Available at: <<http://www.programmableweb.com/api/bing-maps-geocode>> [Accessed 25 Jun. 2015].
- Cayo, M.R., and Talbot, T.O., 2003. Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics*, 2(1), p.10.
- English, P., Neutra, R., Scalf, R., Sullivan, M., Waller, L., and Zhu, L., 1999. Examining associations between childhood asthma and traffic flow using a geographic information system. *Environmental Health Perspectives*, 107(9), pp.761–767.
- Foote, K.E., and Huebner, D.J., 2000. Error, Accuracy, and Precision. [online] Error, Accuracy, and Precision. Available at: <http://www.colorado.edu/geography/gcraft/notes/error/error_f.html>.
- Geschwind, S.A., Stolwijk, J.A., Bracken, M., Fitzgerald, E., Stark, A., Olsen, C., and Melius, J., 1992. Risk of congenital malformations associated with proximity to hazardous waste sites. *American Journal of Epidemiology*, 135(11), pp.1197–1207.
- Google Maps, 2015. The Google Geocoding API. [online] Google Developers. Available at: <<https://developers.google.com/maps/documentation/geocoding/>> [Accessed 25 Jun. 2015].
- Gregorio, D.I., Cromley, E., Mrozinski, R., and Walsh, S.J., 1999. Subject loss in spatial analysis of breast cancer. *Health & Place*, 5(2), pp.173–177.
- Howe, H.L., 1986. Geocoding NY State Cancer Registry. *American Journal of Public Health*, 76(12), pp.1459–1460.
- Kulldorff, M., and Nagarwalla, N., 1995. Spatial disease clusters: detection and inference. *Statistics in medicine*, 14(8), pp.799–810.

- Levine, N., and Kim, K.E., 1998. The location of motor vehicle crashes in Honolulu: a methodology for geocoding intersections. *Computers, environment and urban systems*, 22(6), pp.557–576.
- Lewis-Michl, E.L., Melius, J.M., Kallenbach, L.R., Ju, C.L., Talbot, T.O., Orr, M.F., and Lauridsen, P.E., 1996. Breast cancer risk and residence near industry or traffic in Nassau and Suffolk Counties, Long Island, New York. *Archives of Environmental Health*, 51(4), pp.255–265.
- OpenStreetMap, 2015. OpenStreetMap. [online] Available at: <<http://www.openstreetmap.org/>>.
- QGIS, 2015. QGIS. [online] QGIS A Free and Open Source Geographic Information System. Available at: <<http://www.qgis.org/en/site/>>.
- Rushton, G., and Lolonis, P., 1996. Exploratory Spatial Analysis of Birth Defect Rates in an Urban Population. *Statistics in Medicine*, 15(7-9), pp.717–726.
- Tedies, 2013. Tedies. [online] Edies. Available at: <<http://www.tedies.eu/>>.
- Telgis, 2015. Address Interpolation and Point Address Data. [online] Address Interpolation and Point Address Data. Available at: <<http://docs.geobase.info/html/7fd14468-fa46-4b9d-9908-09fdabbc621b.htm>>.
- Tomtom, 2015. Tomtom. [online] Fleet Management & Vehicle Tracking - Tomtom Telematics GB. Available at: <https://business.tomtom.com/en_gb/>.
- Upsilon, 2009. Upsilon. [online] Upsilon, le spécialiste du transport et de la logistique dans l'Yonne. Available at: <<http://upsilon89.com/>>.
- Yu, L., 1996. Development and Evaluation of a Framework for Assessing the Efficiency and Accuracy of Street Address Geocoding Strategies. State University of New York at Albany.
- Zinszer, K., Jauvin, C., Verma, A., Bedard, L., Allard, R., Schwartzman, K., de Montigny, L., Charland, K., and Buckeridge, D.L., 2010. Residential address errors in public health surveillance data: a description and analysis of the impact on geocoding. *Spatial and Spatio-Temporal Epidemiology*, 1(2-3), pp.163–168.