# A REAL-TIME METHODOLOGY FOR SOLVING MAP-MATCHING PROBLEMS IN GIS-BASED TRANSPORTATION NETWORKS

*Carola A. Blazquez**

*Engineering Science Department, Universidad Andres Bello, Santiago, Chile*

*Pablo A. Miranda*
*Jorge Barahona*

*School of Industrial Engineering*
*Pontificia Universidad Catolica de Valparaiso, Valparaíso, Chile*
*(*Corresponding Author´s e-mail cblazquez@unab.cl)*

## ABSTRACT

Vehicle locations do not exactly register with roadway centerlines due to errors in the GPS measurements and digital roadway maps, particularly in dense urban areas. This phenomenon may lead to a map-matching problem, where a vehicle may appear to be traveling on an incorrect roadway. This paper proposes a real-time methodology based on a previously developed post-processing map-matching algorithm, in order to rapidly and effectively identify the correct roadway centerline segment on which a vehicle is traveling and accurately estimate any transportation-related activities such as travel behavior, and traffic volume and travel speed computations. The proposed algorithm and different controlling parameter values were tested on three digital roadway maps and data collected every 5 and 10 seconds along highways and local roads in Valparaiso and Santiago, Chile, and Portage County, Wisconsin in the United States. The results indicate that the proposed methodology performs similarly to the post-processing algorithm. These results depend upon the spatial data quality and the temporal resolution employed. Best algorithmic parameter values are presented for maximizing solved spatial ambiguities, thus enhancing the performance of the real-time map-matching algorithm.

*Keywords: map matching, real time, transportation network, spatial data, GPS, GIS*

# INTRODUCTION

In recent years, there has been an increase of the transportation users implementing GPS technology for vehicle location. The applications vary from road use charging based on total mileage driven, computation of winter maintenance operations with accurate accumulated distances, trip planning behavior and route choice decision-making by knowing exact start and end trip times to the task of monitoring fleets of vehicles in real-time (Vonderohe et al., 2004; Grush, 2008; Papinski, 2009; Sheridan et al., 2011, Blazquez et al., 2012). However, the integration of these GPS measurements with a digital cartography in a GIS environment is complicated when the data points do not coincide with the transportation network particularly at diverging/converging roads or intersections. Thus, a map-matching problem (also referred to as spatial ambiguity or spatial mismatch) arises. Typically, data points are snapped to incorrect roadway centerlines due to lack of accuracy in the digital cartography, the GPS measurements, or both. For example, Figure 1 shows a vehicle path along Road 1 collecting GPS data points: GPS1, GPS2, and GPS3. Data point GPS2, located close to the at-grade intersection of Roads 1 and 2 snaps erroneously to Road 2 yielding a map-matching problem. As a consequence of this problem, the vehicle location is assigned to an incorrect roadway segment and, thus, affecting any subsequent process of transportation usage, computation, evaluation, analysis, planning, and decision-making.

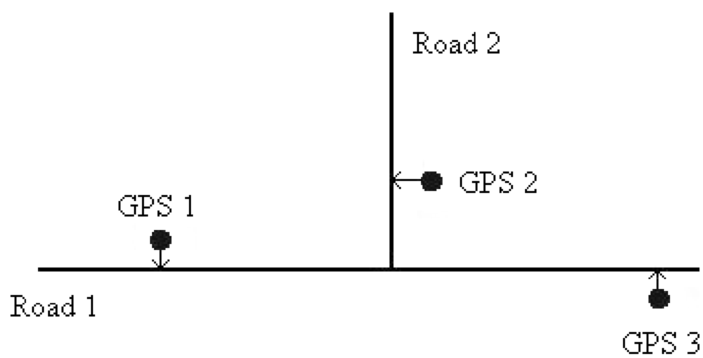



Figure 1 - Example of at-grade intersection ambiguity not resolved

Several approaches have been proposed in the literature to solve the map-matching problem or spatial ambiguity. These methodologies have different levels of complexity ranging from simple search techniques to advanced and sophisticated inference, filtering, and mathematical modeling approaches such as Kalman filters, fuzzy logic, and Bayesian statistics (Zhao, 1997; Blazquez and Vonderohe, 2005; Taylor and Blewitt, 2006; Quddus et al., 2007). Recently, topological map-matching algorithms are more inclined to be employed in solving spatial mismatches due to the rapid and simple implementation of these algorithms (Velaga et al., 2009), particularly in real-time mode. Thus, this type of algorithm is employed in this research.

The proposed real-time topological methodology presented in this paper is based on a modified version of the post-processing map-matching algorithm described in Blazquez and Vonderohe (2005). Instead of employing previous and subsequent snapped data points in the spatial mismatch resolution, the proposed methodology determines the association of data points to correct roadway centerlines in real-time by using solely previous snapped data points or historical data.

This paper is organized as follows. The first section describes the post-processing map-matching algorithm. Subsequently, the adapted real-time methodology and implementation is explained, and a computational application with real world spatial data sets, and the comparison analysis and results are presented. Finally, the conclusions are discussed.

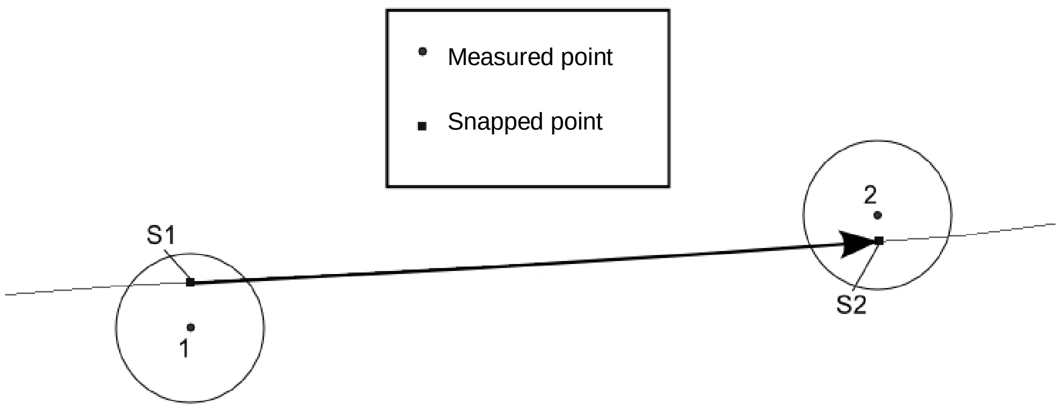



Figure 2 - Example of snapping process to the correct roadway for two GPS data points

# POST-PROCESSING MAP-MATCHING ALGORITHM

The post-processing map-matching algorithm previously developed by one of the authors (Blazquez and Vonderohe, 2005) solves spatial ambiguities by determining the correct roadway centerline on which a vehicle is traveling. First, the algorithm selects all roadways within a buffer around each analyzed GPS data point, and orthogonally projects it to the closest roadway by determining the minimum perpendicular distance between the point and each roadway centerline, as shown by Points 1 and 2 in Figure 2. Subsequently, the shortest traveled distance between the pair of snapped points (in this example, S1 and S2) is computed with Dijkstra´s algorithm using network topology and turn restrictions. This distance and the difference in time stamps for the points are employed to calculate the vehicle travel speed. If the average recorded speed at these points is within a speed range tolerance of the vehicle travel speed, then the obtained shortest path is viable and the snapped locations for both points are accepted as correct. If the path is rejected, then data points are snapped to alternative roadway centerlines contained within their buffers, shortest paths are recalculated, and speeds are compared once again. If no other roadway centerlines exist within the buffers or no feasible paths are obtained, then the algorithm tests for feasible paths between preceding and subsequent snapped data points until a predefined number of consecutive points are examined. Notice that the subsequent snapped data points have not been recorded yet in real-time mode. Thus, the proposed methodology is based on this algorithm and is applicable in real-time by only taking into account the preceding recorded GPS data points.

# REAL-TIME METHODOLOGY AND IMPLEMENTATION

The post-processing map-matching algorithm described in the previous section could tentatively be implemented in real-time. However, this algorithm would produce results in an inefficient manner by having to "wait" for every measured point, in order to continue executing the algorithm. Hence, the post-processing map-matching algorithm was modified to yield a real-time methodology that successfully identifies the correct roadway centerline for each data point as it is being measured by only employing historical data.

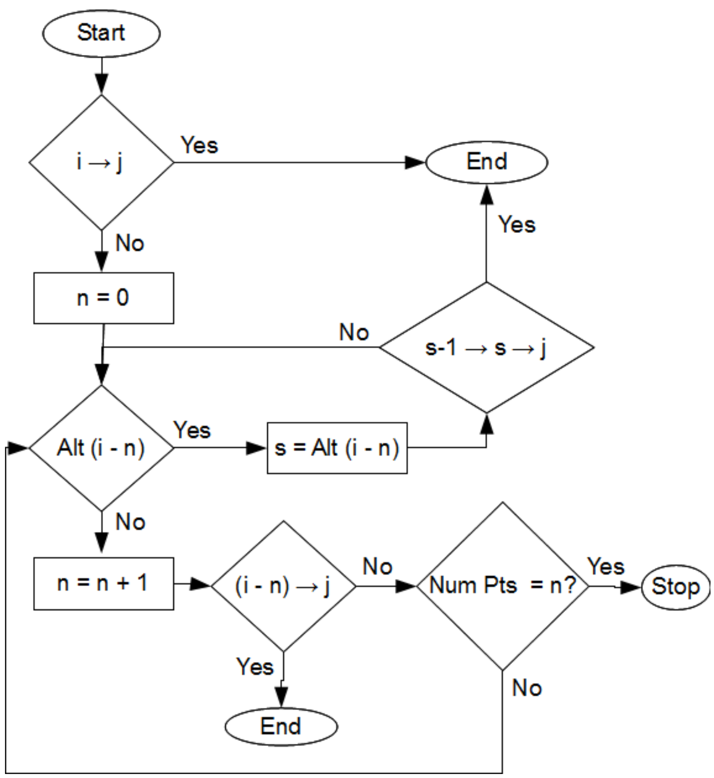



Figure 3 - Flow diagram of the real-time map-matching algorithm

Figure 3 presents a flow diagram of the proposed real-time map-matching algorithm. The algorithm is executed when a map-matching problem is encountered. In other words, no feasible shortest path is obtained between the pair of points **i** and **j**. First, the algorithm sets the counter **n** for number of consecutive points to zero. Subsequently, alternative roadway centerlines contained within the buffer of point **(i – n)** are searched. If an alternative roadway centerline is found, then the algorithm tests for feasible shortest paths between alternative point **(i – n)** or **s** and neighboring points **(i – n – 1)** or **(s – 1)** and **j** by performing a speed comparison. If the tested path is viable, then the spatial ambiguity is solved. Otherwise, the counter **n** is incremented in one unit and preceding points are examined broadening the test range backwards on a one-by-one step basis. The algorithm follows a recursive procedure until the spatial ambiguity is solved or terminates when the predefined number of consecutive data points has been reached, and registers the following measured data point.

## Example of a map-matching resolution

Figure 4 presents an example of a spatial ambiguity at a diverging roadway centerline when the new point 4 is measured. Points 1 and 4 are snapped to the closest roadway centerline corresponding to Spain Avenue and points 2 and 3 are snapped to Marina Street. A route between the pair of snapped points S3 and S4 is not possible. Therefore, the real-time methodology is initiated. Since no alternative roadway centerlines are contained within the buffer of point 3, then a feasible path is tested between previous snapped points S2 and S4. Once again there is no viable path, and an alternative roadway centerline is searched within the buffer of point 2. An alternative snapping location for point 2 is found, then feasible shortest paths are examined between this newly snapped point and snapped points S1 and S4. Finally, a feasible path is obtained. The intermediate point 3 is snapped along the obtained path and the map-matching problem is solved.

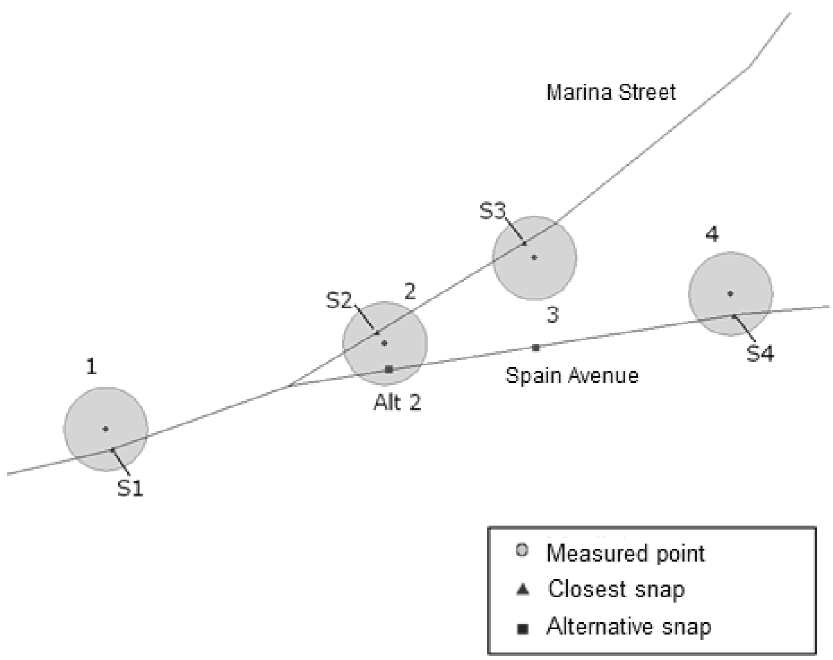



Figure 4 - Example of a spatial ambiguity resolution at a diverging road

## Implementation issues

The success of the real-time methodology depends directly on the parameter values. If buffer sizes are too small, then no alternative snapping locations are obtained and the algorithm is forced to explore additional preceding data points. If the number of tested predefined neighboring data points is reached, then the algorithm terminates without solving the map-matching problem.

Figure 5 illustrates an unsolved spatial mismatch when implementing the real-time methodology. A map-matching problem occurs when examining the path between snapped points S133 and S134. Points S130 through S133 snapped to Road 3, and point S134 snapped to Road 2. The true route corresponds to a vehicle traveling along Roads 3 and 1. Table 1 presents the step-by-step sequence of searches for feasible paths and alternative roadway centerlines in this example. Since there is no viable path between points S133 and S134, alternative roadway segments are searched within the buffer of point S133. The algorithm proceeds by testing for feasible paths using previous points and searching alternative roadway centerlines around each prior snapped point until five neighboring data points are employed and thus no solved solution is obtained. An algorithm implementation concern is the adequate number of points to analyze in the problem resolution.

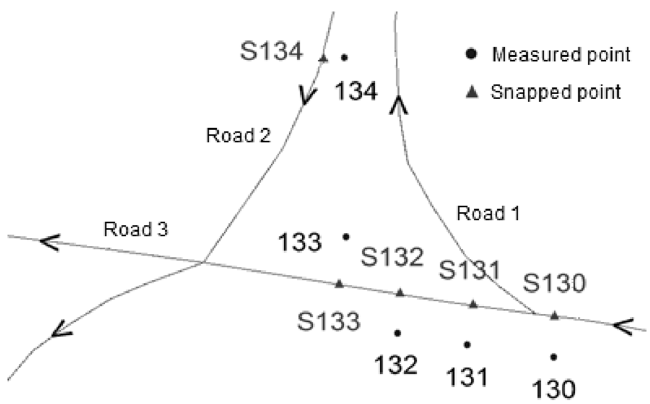



Figure 5 - Example of an unsolved spatial mismatch

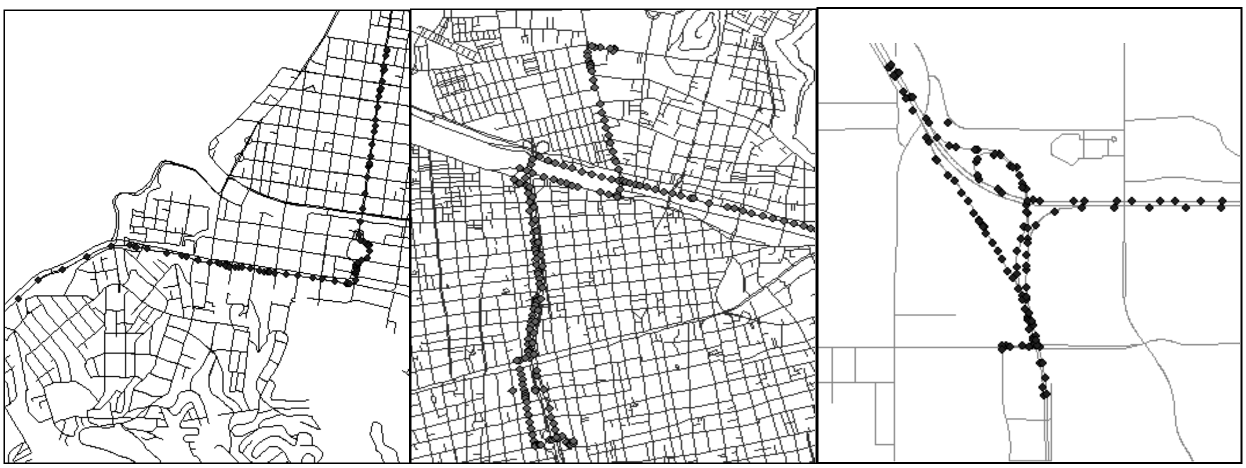
Table 1- Feasibility examination between pair of points

| Data Points | Feasible? |
|---|---|
| S133 → S134 | NO |
| Alt S133 | NO |
| S132 → S134 | NO |
| Alt S132 | NO |
| S131 → S134 | NO |
| Alt S131 | NO |
| S130 → S134 | NO |
| Alt S130 | NO |

# COMPUTATIONAL APPLICATION

## Spatial data sets

Since spatial data quality and temporal resolution impact the performance of map-matching algorithms, the proposed real-time map-matching methodology was tested with different controlling parameters on three digital roadway maps and on data sets collected every 5 and 10 seconds in Santiago and Valparaiso, Chile, and Portage County, Wisconsin in the United States. Six different routes (two routes in each city) with over 1,000 data points were collected along highways and local roads. Figure 6 presents the data samples employed in this study on roadway network accuracies with 1:5,000, and 1:24,000 nominal scales. Note that no heading information was available during the data collection process.



a) Valparaiso b) Santiago c) Portage County

Figure 6 - Examples of spatial database and GPS measurements for Valparaiso and Santiago, Chile, and Portage County, Wisconsin

A simulation process was employed with the complete data set (each route separately) to reproduce measuring conditions and implement the proposed methodology in real-time. Each data point is recorded, read, and entered into the GIS environment one at a time using a defined measuring frequency. As each point is entered, the point is snapped to the closest roadway centerline, a shortest path is pursued between this point and the prior snapped point, and recorded and computed speeds are compared. The real-time methodology shown in Figure 3 is executed to solve the map-matching problem when no feasible paths are obtained. Note that the time required execute the algorithm with the predefined number of consecutive data points should be less than the sampling interval.

## Algorithmic parameters

Two algorithmic parameters (buffer size, and speed range tolerance) were analyzed for different sampling frequencies (5 and 10 seconds). The percentage of solved spatial ambiguities is computed for each parameter by comparing true and computed vehicle routes.

The appropriate buffer size employed during the snapping process depends on the quality and geometry of the spatial data. This proximity parameter used to search and select roadway centerlines around data points is critical for solving the map-matching problem and, therefore, for the success of the algorithm. Buffers that are overly small in size might not include any roadways, while extremely large buffers make the algorithm less efficient since it needs to examine more roadways.

The feasibility of shortest paths between pairs snapped data points is sensitive to the allowable range tolerance utilized when comparing computed and recorded speeds. When small tolerances are employed, feasible paths are rejected leaving data points not snapped to any roadway centerline. On other hand, large speed range tolerances do not improve results significantly since all feasible paths are accepted (Blazquez and Vonderohe, 2009). Thus, best parameter values are sought for the analyzed spatial data, which maximize the percentage of solved spatial ambiguities.

Table 2 presents the parameter values and the respective increments for each route employed in the real-time methodological implementation. Different parameter values accommodate the tested spatial data used in this study. For example, the buffer size parameter was tested at 4 meter increments from 6 to 22 meters for data collected in Valparaiso. As different buffer sizes and speed range tolerances are analyzed and tested, the other variables are maintained constant with values of five consecutive data points for all routes, and 10 and 5 seconds for sampling intervals in Routes 1 – 4, and Routes 5 – 6, respectively.

Table2- Buffer size and speed range tolerance values and increments per route

| Routes | Buffer Size | | Speed Range Tolerance | |
|---|---|---|---|---|
| | Range Values | Increments | Range Values | Increments |
| Valparaiso | | | | |
| Route 1 | 6 – 22 m | 4 m | 20 – 50 km/h | 10 km/h |
| Route 2 | | | | |
| Santiago | | | | |
| Route 3 | 10 – 25 m | 5 m | 10 – 25 km/h | 5 km/h |
| Route 4 | | | | |
| Portage | | | | |
| Route 5 | 20 – 60 ft | 10 ft | 10 – 40 mi/h | 10 mi/h |
| Route 6 | | | | |

## Analysis results

The analysis results shown by the curves of the charts in Figures 7 through 12 indicate that the average percentage of solved spatial ambiguities increases as the buffer size increments tending to stabilize at a certain value. For example, approximately 69% of the spatial mismatches for Route 2 in Valparaiso are solved when using the smallest buffer radius, and this percentage reaches values over 96% for buffer sizes greater than 14 meters. On average, buffer sizes of 18 m, 25 m, and 35 ft for data collected in Valparaiso, Santiago, and Portage County solves approximately of 95%, 94%, and 97% of spatial ambiguities, respectively.

Conversely, variations in the speed range tolerance do not influence the analysis results for data collected in Valparaiso and Portage County, except for data collected in Santiago. Almost no changes exist in the shape of the curves since the speed ranges selected for the computational application are adequate. Thus, the methodology requires the minimum speed range tolerance value that stabilizes the curves.

The data collection environment in Santiago consists of a dense urban canyon with relatively tall building and narrow streets. Thus, the Position Dilution of Precision (PDOP) values are high. In addition, no GPS augmentation methods were applied yielding poor quality in the data measurements, which require larger buffer sizes and speed range tolerance in the implementation for best results.

Increasing the number of consecutive data points aids to solve the map-matching problem since more paths are sought within the vicinity of the spatial ambiguity. However, the computational time needed to test all paths among this number of consecutive points should not exceed the temporal resolution of the measurements. Further research is needed to test the algorithm with different number of consecutive data points and sampling intervals.

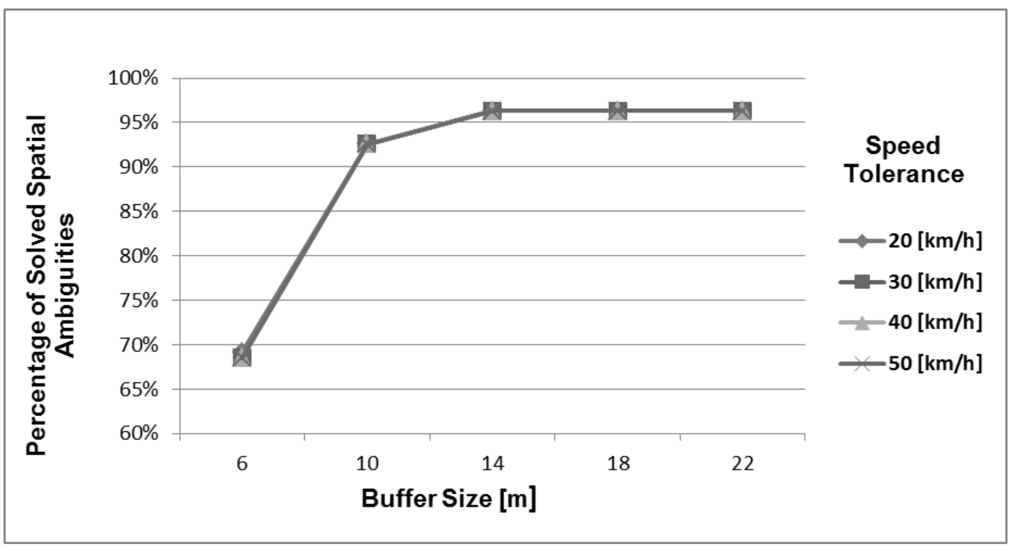



Figure 7 - Percentage of solved spatial ambiguities for different buffer sizes and speed range tolerances for Route 1 in Valparaiso

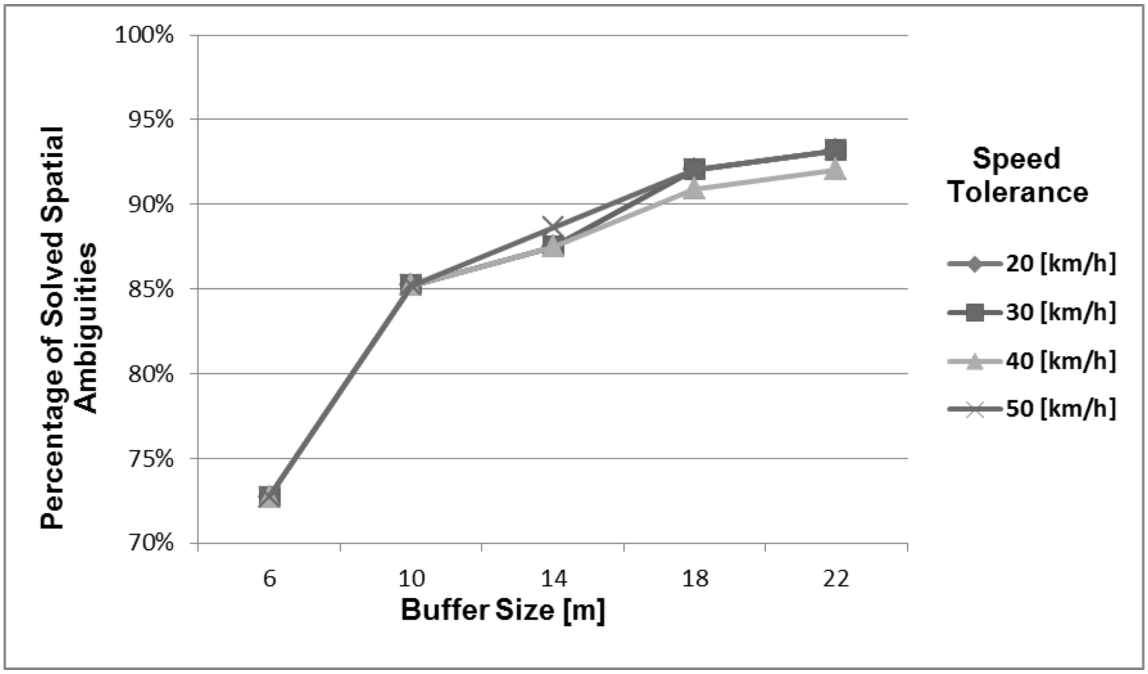



Figure 8 - Percentage of solved spatial ambiguities for different buffer sizes and speed range tolerances for Route 2 in Valparaiso

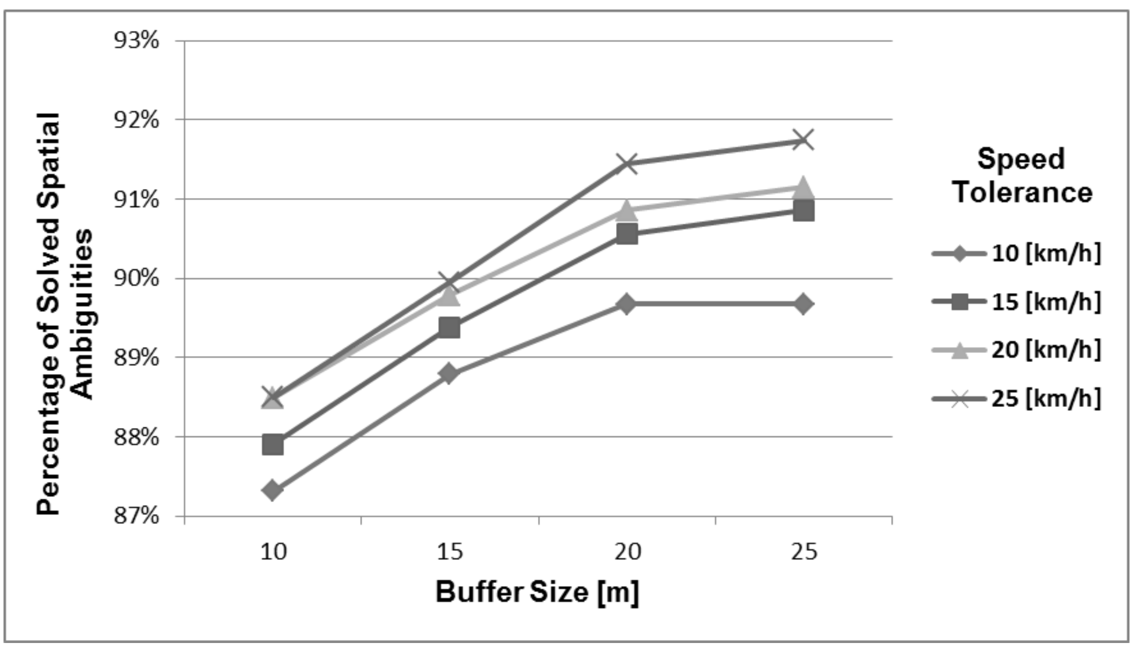



Figure 9 - Percentage of solved spatial ambiguities for different buffer sizes and speed range tolerances for Route 3 in Santiago

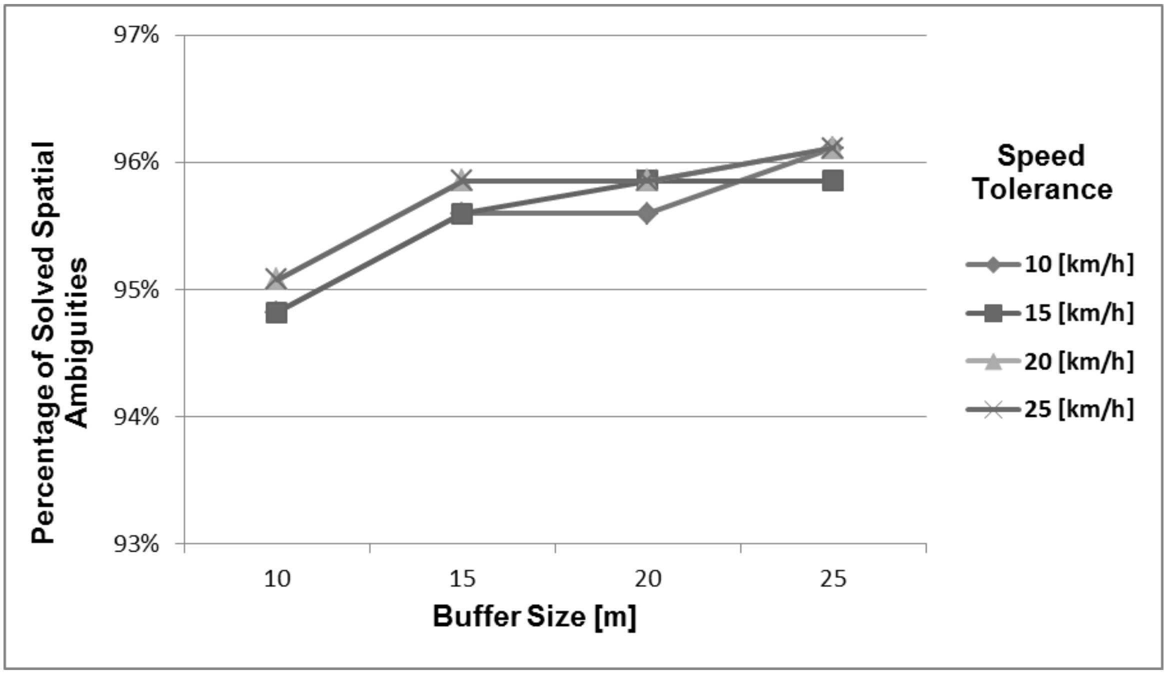



Figure 10 - Percentage of solved spatial ambiguities for different buffer sizes and speed range tolerances for Route 4 in Santiago

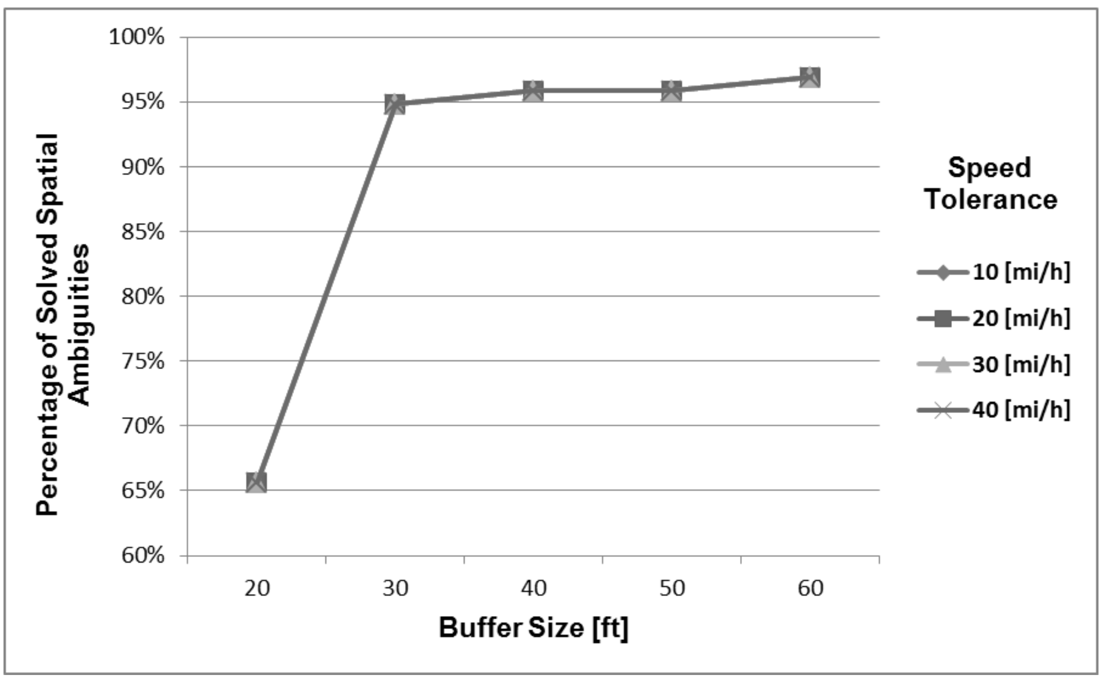



Figure 11 - Percentage of solved spatial ambiguities for different buffer sizes and speed range tolerances for Route 5 in Portage County

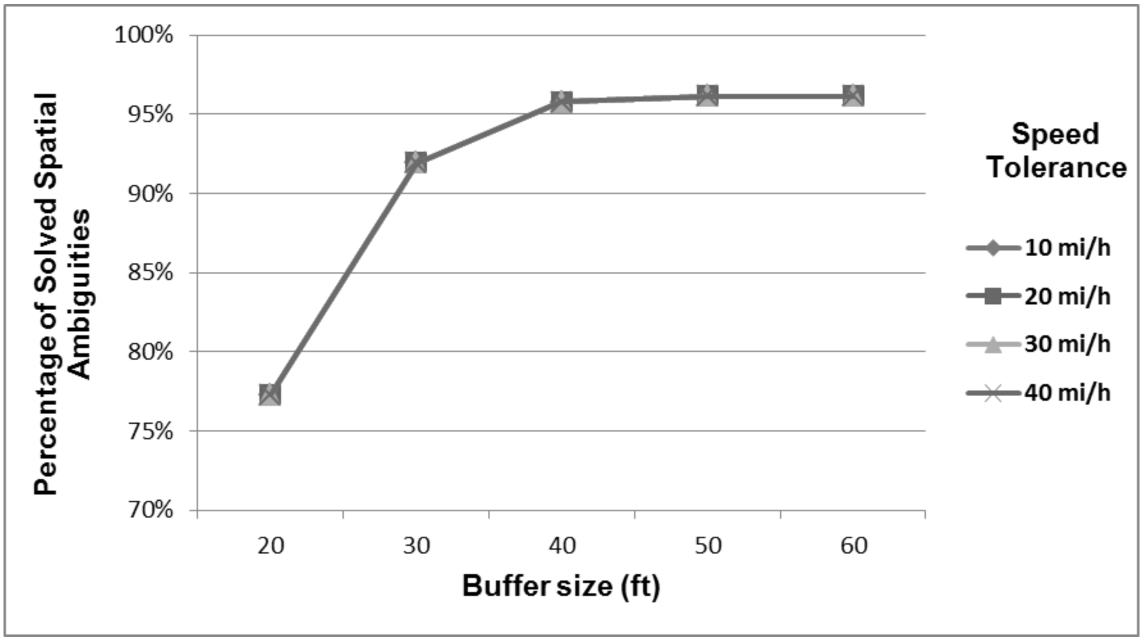



Figure 12 - Percentage of solved spatial ambiguities for different buffer sizes and speed range tolerances for Route 6 in Portage County

## Comparison analysis: post-processing versus real-time methodologies

The same algorithmic parameter values and ranges presented in Table 2 are employed in the comparison analysis. The analysis compares the largest percentage of solved spatial mismatches and the best associated parameters values for both post-processing and real-time methodologies. Table 3 shows the results of the comparison analysis.

Overall, the percentages of solved spatial mismatches are greater for the post-processing methodology since the complete data set is previously known. The last column of the table presents the difference in the results between the two methodologies with values less than 4% for all routes. The post-processing algorithm is capable of examining points ahead and behind during the map-matching problem resolution. Whereas, the proposed real-time methodology relies only on historical data and the newly measured data point for determining the correct roadway centerline. Interestingly, data collected along Route 4 in Santiago presents better results when implementing the real-time methodology. Further analysis is needed to comprehend this outcome.

As shown in Table 3, similar best parameter values are needed for both methodologies, in order to maximize the percentage of solved spatial ambiguities. Note that data collected along Routes 3 and 4 in Santiago should employ larger buffer sizes and speed range tolerances. This may be due to the low quality of the GPS measurements, as explained above.

Table 3 – Best parameter values and percentage of solved spatial ambiguities per route for the post-processing and real-time methodologies

| **Routes** | **Post-Processing Methodology** | | **Real-Time Methodology** | | **Difference Between Post-Processing and Real-Time** |
|---|---|---|---|---|---|
| | **Best Parameter Values** | **Solved Spatial Ambiguities** | **Best Parameter Values** | **Solved Spatial Ambiguities** | |
| Valparaiso | | | | | |
| Route 1 | Buffer size ≥ 14 m Speed ≥ 20 km/h | 98.8% | Buffer size ≥ 14 m Speed ≥ 20 km/h | 96.3% | 2.6% |
| Route 2 | Buffer size = 22 m Speed ≥ 20 km/h | 96.6% | Buffer size = 22 m Speed ≥ 20 km/h | 93.2% | 3.7% |
| Santiago | | | | | |
| Route 3 | Buffer size = 15 m Speed ≥ 10 km/h | 95% | Buffer size = 25 m Speed = 25 km/h | 91.7% | 3.6% |
| Route 4 | Buffer size = 20 m Speed ≥ 10 km/h | 87% | Buffer size = 25 m Speed ≥ 10 km/h | 96.1% | -9.5% |
| Portage | | | | | |
| Route 5 | Buffer size = 30 ft Speed ≥ 10 mi/h | 97.9% | Buffer size = 30 ft Speed ≥ 10 mi/h | 96.2% | 1.8% |
| Route 6 | Buffer size = 60 ft Speed ≥ 10 mi/h | 100% | Buffer size = 60 ft Speed ≥ 10 mi/h | 96.9% | 3.2% |

# CONCLUSIONS

This paper presents a real-time methodology using a map-matching algorithm that rapidly and efficiently identifies correct vehicle routes following network topology and turn restrictions. This methodology is based on a previously developed post-processing map-matching algorithm. Rather than seeking for feasible shortest paths between neighboring previous and subsequent snapped points, the proposed methodology solely employs historical data.

The implementation results suggest that the performance of the real-time methodology is sensitive to variations in the buffer size. Best buffer size values that maximize the percentage of solved spatial mismatches are presented for each route. If lower spatial data qualities and less frequent sampling intervals are used, then the algorithm requires larger buffer sizes and speed ranges to obtain best results.

The results of the comparison analysis are presented that indicate the best controlling parameter values to obtain the maximum percentage of solved spatial mismatches. As expected, the post-processing algorithm solves larger amount of map-matching problems since prior and future data point locations are known. However, on average the results of the real-time methodology are 3% less than the post-processing map-matching algorithm, excluding Route 4 in Santiago. Future research includes the performance examination of the real-time methodology with different spatial data qualities collected in various operational environments.

## ACKNOWLEDGEMENTS

This research has been supported by the Chilean National Fund for Scientific and Technological Development (FONDECYT 1070386).

## REFERENCES

Blazquez, C. and A. Vonderohe (2005). Simple Map-Matching Algorithm Applied to Intelligent Winter Maintenance Vehicle Data. Journal of Transportation Research Board, 1935, 68-76.

Blazquez, C. and A. Vonderohe (2009). Effects of Controlling Parameters on Performance of a Decision-Rule Map-Matching Algorithm. Journal of Transportation Engineering, 135(12), 966-973.

Blazquez, C., P. Miranda, and A. Ponce (2012). Performance of a New Enhanced Topological Decision Rule Map Matching Algorithm for Transportation Applications. Journal of Applied Research & Technology, 10, 929-940.

Grush, B. (2008) The Case Against Map-Matching. European Journal of Navigation, 6(3), 2-5.

Papinski, D., D. Scott, and S. Doherty (2009). Exploring the Route Choice Decision-Making Process: A Comparison of Planned and Observed Routes Obtained Using Person-Based GPS. Transportation Research Part F, 12, 347-358.

Quddus, M., W. Ochieng, and R. Noland (2007). Current Map-Matching Algorithms for Transport Applications: State-Of-The Art and Future Research Directions. Transportation Research Part C, 15, 312-328.

Sheridan, K., T. Dyjas, C. Botteron, J. Leclere, F. Dominic, and G. Marucco (2011). Demands of Roads. An Assisted-GNSS Solution Uses the EGNOS Data Access Service. GPS World, 22(3), 28.

Taylor, G. and G. Blewitt (2006). Intelligent Positioning GIS-GPS Unification. West Sussex, UK, John Wiley & Sons, Ltd.

Velaga, N., M. Quddus, and A. Bristow (2009). Developing an Enhanced Weight-Based Topological Map-Matching Algorithm for Intelligent Transport Systems. Transportation Research Part C, 17(6), 672-683.

Vonderohe, A., T. Adams, C. Blazquez, J. Maloney, and T. Martinelli (2004). GIS-Based Analysis of Intelligent Winter Maintenance Vehicle Data. Transportation Research Circular E-C063: Sixth International Symposium on Snow Removal and Ice Control Technology, Spokane, Washington, 348-360.

Zhao, Y. (1997). Vehicle location and navigation systems, Artech House, Inc. Norwood, MA.