A Decision-Rule Topological Map-Matching Algorithm with Multiple Spatial Data

Carola A. Blazquez

Universidad Andres Bello
Department of Engineering Science
Chile

1. Introduction

Intelligent Transportation System (ITS) applications such as congestion and traffic management employ Global Positioning Systems (GPS) technology to collect positioning data in two or three dimensions of events, incidents, or vehicles. This information is integrated with Geographic Information Systems (GIS) to determine the roadway upon which events and incidents occur, point features such as traffic signs are located, or vehicles are traveling.

Vehicle trajectories displayed on a digital map are not situated on top of the roadway centerlines, which represent the real world. Therefore, when both GPS measurements and roadway centerline maps are very accurate, a GPS data point is associated with the nearest roadway by calculating the minimum perpendicular distance between each roadway representation and the GPS data point. This process is called “snapping”. Unfortunately, a spatial mismatch occurs when a GPS data point is snapped to an incorrect roadway centerline due to roadway network complexities, inadequate GPS data collection procedures, and lack of accuracy in the digital roadway map and the GPS measurements, or combinations of them (Chen et al., 2005). Figure 1 shows an example where errors in the location of the measured GPS data point cause an incorrect snap to the nearest road 2 instead of snapping to road 1.

![Incorrect Snap](image.png)

Fig. 1. Measured GPS Data Point with Error Snapped to the Wrong Roadway Centerline

Generally, spatial mismatches or map-matching problems occur at overpasses and underpasses, converging and diverging roadways such as ramps and divided highways, or when roads are close together. Figure 2 presents GPS measurements of a vehicle traveling at
a major highway interchange containing ramps, overpasses, and underpasses. This example indicates that multiple spatial mismatches may occur at interchanges.

As a consequence of the map-matching problem, any subsequent usage, visualization, computation, evaluation, analysis, planning, and decision-making may be impacted negatively and produce erroneous perceptions. For example, the calculated cumulative distance traveled by a vehicle along a roadway network is incorrect and, therefore, calculated values for performance measures such as fuel consumption or decision management tools that depend upon cumulative distance are wrong. Additionally, any non-spatial data collected from vehicle sensors such as speed data or emission levels are associated with incorrect roadway centerlines. Furthermore, GPS data points might be incorrectly assigned to roadways along which no measurements were ever taken affecting transportation applications such as road use charging based on the total mileage driven by vehicle (Cozzens, 2009; Sheridan, 2011). The need to overcome spatial mismatches in ITS applications is a major motivation for implementing map-matching algorithms.

Fig. 2. GPS Data Points Collected by a Vehicle While Traveling at a Highway Interchange

Section 2 presents a literature review of map-matching algorithms developed to solve spatial ambiguities. Section 3 describes the proposed topological decision-rule map-matching algorithm and an example of its implementation. Results of the performance analysis with real spatial data are presented in section 4. Finally, section 5 presents a summary and the main conclusions of this chapter, and further research topics to be addressed.

2. Map-matching methods

The problem of resolving spatial ambiguities has been widely studied over the years. The following map-matching algorithms are described in the literature with different levels of complexity ranging from simple geometric techniques to complex, advanced approaches.

2.1 Semi-deterministic map-matching

The earliest map-matching algorithm, before GPS was developed in the 1970’s, followed a semi-deterministic model (French, 1989). This model assumes that the vehicle has an initial
location on a roadway and a given direction of travel. Conditional tests are applied to
determine whether the vehicle is traveling on the known road by comparing turns from the
vehicle location to a segment of the digital road map. A correction is performed whenever
the heading of the vehicle changes (Morisue & Ikeda, 1989). However, for this technique to
work, the vehicle is generally assumed to follow a predetermined road. There is
considerable uncertainty when the vehicle travels off-road because there is no longer any
way to correct for errors (Zhao, 1997; Czerniak, 2002).

2.2 Probabilistic map-matching

The probabilistic approach, described later, has the advantage of not assuming that the
vehicle is always on a road. Vehicle heading error is calculated with an elliptical or
rectangular confidence region and error models are developed within which the true vehicle
location can be determined. If the vehicle position within the region contains one
intersection or road segment, a match is made and the coordinates on the road are used in
the next position calculation. If more than one road or intersection lies within the region,
connectivity checks are made to determine the most probable location of the vehicle given
earlier vehicle positions. As a result, the algorithm yields the best match segment along with
the most probable matching point on the segment (Zhao, 1997; Czerniak, 2002).

2.3 Fuzzy logic map-matching

Fuzzy logic is an effective way to deal with tasks that involve qualitative terms and
concepts, vagueness, and human intervention. Expert knowledge and experiences employed
by a fuzzy logic based map-matching algorithm are represented as a set of rules to
determine vehicle location (e.g., if the difference between the orientation of the roadway
segment and the heading of the vehicle is small, then resemblance between the vehicle travel
path and the candidate route is high).

that identifies the roadway on which a vehicle is traveling by comparing C-measures
associated with each candidate roadway. These measures are membership functions that
represent the certainty of the existence of a vehicle on a specific roadway. After the roadway
is identified, the algorithm determines the vehicle position on the roadway by orthogonal
projection. The algorithm requires the distance between the vehicle’s GPS coordinates and
its projected position on the roadway to be small. Furthermore, the shape of the roadway
must be similar to the trajectory of the vehicle.

Jagadeesh et al. (2004) developed a map-matching algorithm based on the inferences and a
simple fuzzy rule set. This algorithm evaluates the likelihood of candidate roads to be the
actual traveled road. Three fuzzy rules are employed for this purpose, which include
heading comparison, road resemblance, and verification of off road vehicles. Test results
with simulated data indicate that the algorithm is capable of achieving high accuracy.

Quddus et al. (2006) describe a map-matching algorithm based on fuzzy logic theory. The
proposed algorithm employs an integrated navigation system and digital map data to
identify the correct link and determine the vehicle location on the selected link. Although
the algorithm was tested successfully in different road networks, the authors consider that
future evaluation of the algorithm is required under urban conditions.
Yet another map-matching algorithm based on fuzzy theory is proposed by Guo and Luo (2009). First, the algorithm compares the similarity degree between the trajectory curve of the road and all candidate roads to identify the road on which a vehicle is traveling. Subsequently, fuzzy preference relations are adopted to perform a multi-criteria decision and a look-ahead technique is employed to improve the matching accuracy. The algorithm requires testing and analysis with GPS data in addition to cell phone positions.

### 2.4 Kalman filter approach

There has been abundant research on application of Kalman filters in combination with GPS and dead-reckoning signals to solve spatial mismatches. This integrated technology improves positioning accuracy by estimating white noise and error in the GPS and then correcting the vehicle’s position (Jo et al., 1996; W. Kim et al., 2000; Zhao et al., 2003). For example, Quddus et al. (2003) present a general map-matching algorithm that integrates GPS and dead-reckoning sensor data (position, velocity, and time) through an extended Kalman filter and uses them as input to improve performance of the algorithm. The physical location of the vehicle on a roadway link is determined empirically from the weighted averages of two state determinations of the vehicle position based on topological information and external sensors.

Yang et al. (2003) present an improved map-matching algorithm that employs Kalman filtering to filter unreasonable GPS data and the Dempster-Shafer (D-S) theory to correctly snap GPS vehicle coordinates to the digital roadway map. The D-S theory allows explicit representation of ignorance and combination of evidence and operates with a smaller set of uncertainties. Although the authors report satisfying results, they suggest additional research to verify the accurate performance of the algorithm.

Nassreddine et al. (2009) describe a map-matching method based on D-S theory and interval analysis to compute accurate vehicle positions from an initial estimated position on a digital road network. The authors state that the proposed technique proves to be successful at junctions and parallel roads. However, real world data needs to be examined in addition to simulated data.

### 2.5 Particle filtering and map-matching

Particle filtering, based on a stochastic process, is another approach to the map-matching problem. Particle filters are recursive implementations of Monte Carlo-based statistical signal processing (Crisan & Doucet, 2002). Gustafsson et al. (2002) evaluate in real time a map-matching particle filter used to match a vehicle’s horizontal driven path to a digital roadway map. They conclude that the particle filter converged relatively rapid after a few iterations of the algorithm. The challenge of this map-matching technique is to find nonlinear relations and non-Gaussian sensor models that provide the most information about the vehicle’s position. The authors assert that research is still needed to seek a reliable way to detect divergence and to restart the filter.

Toledo-Moreo et al. (2009) present a multiple-hypothesis particle-filter based algorithm to solve the map-matching problem with integrity provision at the lane level. The proposed system joins measurements from a GPS receiver, an odometer, and a gyroscope along with road information in digital maps. A set of six experiments were conducted with real data for
a period of 30 minutes proving the feasibility of the approach for lane-level applications. The authors mention that outlier removal, multipath effect mitigation, and additional method validation are tasks that need to be addressed in the future.

2.6 Personal navigation assistants and map-matching

White et al. (2000) discuss solutions to the map-matching problem for personal navigation assistants (PNA). Four different map-matching algorithms were implemented and tested: 1) use of minimum distance (point-to-curve), 2) comparison of heading information with arc and trajectory, 3) use of topology to select roads that are reachable from the current road, and 4) construction of piece-wise linear curves from different paths, followed by comparison of them to centerline curves using points (curve-to-curve matching). The authors conclude that these algorithms performed better when the distance between the GPS point and the closest road is small and that correct matches tend to occur at greater speeds on straight roadways.

Freitas et al. (2009) explain the necessity of map-matching algorithms to correctly locate GPS positions on a map when using PNA, particularly for dynamic route guidance systems. The authors describe an approach to update digital maps through the use of GPS points, in order to identify map incongruence. The proposed system was designed as a prototype and lacks of extensive testing, however, it correctly processes and implements methods for map-matching and detecting discrepancies between the real network and digital maps.

2.7 Topological network-based algorithms

Taylor et al. (2001) describe an algorithm called “Road Reduction Filter (RRF)” that uses differential corrections and height aids. RRF identifies all possible roadway candidates while systematically removing incorrect ones. RRF is improved by using shortest path network analysis and drive restriction information. A shortest path network routine calculates the distance through the roadway network from a vehicle’s previous position to each potential present position offered by the algorithm. The drive restriction information routine selects roadways using direction and access information.

Greenfeld (2002) presents a map-matching procedure that consists of two algorithms. One algorithm assesses similarity between characteristics of the roadway network and the positioning pattern of the vehicle. The second algorithm performs topological analysis and applies a weighting scheme to match each GPS data point to the roadway network. The highest weighted score determines the most likely candidate for a correct match. The author indicates that further research is needed to determine the correct position of the vehicle along a roadway segment and to verify the accuracy performance of the algorithms.

Doherty et al. (2000) studied an algorithm that automatically matches GPS data points to roadway segments along a network. First, the algorithm joins GPS points to create a linear object forming the vehicle’s track. Subsequently, it creates a buffer zone around the linear object, and then identifies all the roadways that are totally included within the buffer to select the correct one.

Marchal et al. (2005) presents an innovative map-matching algorithm that relies on GPS measurements and network topology. The algorithm consists of maintaining a set of
candidate paths as GPS data are processed and computing matching scores for each path. The path with the best score represents the correct vehicle route. According to the authors further research is needed to improve the robustness of the algorithm.

Yet another topological map-matching algorithm is proposed by Wang and Yang (2009). The algorithm presents high accuracy and solves spatial ambiguities in complex roadway networks, specifically near intersections and parallel roads. Nevertheless, the topological algorithm was tested on only four road intersections with a 2-second sampling interval of GPS measurements.

Velaga et al (2009) describe an enhanced weight-based topological map-matching algorithm for ITS. The algorithm was tested with real data under different operational environments. However, the optimal algorithmic weights for different factors such as heading, proximity, connectivity, and turn-restriction still need to be estimated with a range of real-world field data from different road environments.

Blazquez and Vonderohe (2005) propose a topological map-matching algorithm that resolves spatial ambiguities that occur with intelligent winter maintenance vehicle data collected in Wisconsin. The algorithm computes shortest paths between snapped GPS data points using network topology and turn restrictions. If similarity exists between calculated and recorded vehicle speed values, then the path is feasible and snapped GPS locations are correct. If the path is not viable, then GPS data points are snapped to alternative roadway centerlines, shortest paths are recalculated, and speeds are again compared. The authors studied this problem further and published the effects of controlling parameters on the performance of the map-matching algorithm (Blazquez & Vonderohe, 2009). The current chapter discusses and describes in more detail the performance analysis of this map-matching algorithm.

2.8 Other map-matching algorithms

According to Zhao (1997), many pattern recognition methods (e.g., neural network) could be used for map-matching. Neural networks are dynamic systems that consist of many interconnected layered nodes (neurons). These networks need to be trained to arrange the layers and interconnections to model real-world applications. Other pattern recognition methods can be used to work with positioning sensors such as GPS. The underlying principle of these methods is that the digital map is used to filter out vehicle sensor errors and to determine the best position.

Schlingelhof et al. (2008) present a two-dimension map-matching algorithm based on a lane-level model. The output of this algorithm is the road segment identification number, the relative vehicle position along this segment, and the relative transversal vehicle position with respect to one of the border lines. The road selection algorithm consists of extracting candidate segments, computing positioning solution residuals, and selecting the most likely segment. The authors state that the first results obtained with real measurements are encouraging. However, these should be generalized to enhanced maps.

Li et al. (2005) present a novel map-matching method using least-squares position estimation, and digital mapping and height data to augment the vehicle position calculation. Experiment results indicate that combining the algorithm with height aiding improves the vehicle position accuracy when the number of visible satellites is reduced.
3. Decision-rule topological map-matching algorithm

3.1 Description

The decision-rule topological map-matching algorithm determines the correct roadway centerline for vehicle travel by obtaining feasible shortest paths between snapped GPS data points in post-processing mode. The algorithm selects all roadways within a buffer around a GPS data point and snaps the point to the closest roadway by obtaining the minimum perpendicular distance from the data point to each roadway. Figure 3 illustrates that GPS data points 1 and 2 (shown as circles) are snapped to ramp 2 because it is the closest roadway contained with the buffers around the points. Subsequently, the shortest path (displayed with a bold arrow) is obtained between the two snapped GPS data points S1 and S2 (shown as squares). Only paths that follow allowable traffic directions and allowable turns are employed. The travel speed between these two snapped GPS points is determined by the length of the shortest path and the difference in time stamps for the points. The computed speed is compared to the average of the speeds at the data points collected by the vehicle while traveling. If the computed speed is within a specified tolerance of the average recorded speed, then the obtained shortest path is viable and the snapped locations for points 1 and 2 are accepted as correct.

![Diagram](image-url)

Fig. 3. Example of Snapping to the Correct Roadway for Two GPS Data Points Using the Map-Matching Algorithm

The map-matching algorithm advances to GPS data point 3, snaps this point to the closest roadway centerline within its buffer, and calculates the shortest path between snapped point S2 and the newly-snapped GPS data point S3. If the path between S2 and S3 is not feasible because the speed comparison yields a large disparity, then the algorithm determines if feasible routes exist between the preceding and subsequent points bounding the GPS data points of concern, as illustrated in the example of Figure 4. This example shows that there is no feasible path between snapped points S2 and S3 when network topology and turn restrictions are employed. Therefore, the map-matching algorithm looks
ahead by snapping point 4 to the nearest roadway centerline within its buffer, and determines if the shortest path between snapped points S3 and S4 is possible. Since the tested path is not feasible, the algorithm snaps point 3 to the next nearest roadway centerline within its buffer obtaining point alt3, shown as a triangle.

Subsequently, the upper part of the algorithm (shown in Figure 5) for alternative roadway centerline search and feasibility path check is initiated. This algorithm verifies if a path is feasible between the alternative snapped location for point 3 (where Ki = 3), and former and succeeding neighboring snapped points 2 and 4 (where Ki-1 = 2 and Kj = 4). If the shortest paths between these three points are not feasible because the speed comparison fails, then the algorithm searches for other roadway centerlines within the buffer around point 3 that have not already been used in a feasibility path check. When finding a new candidate, point 3 is then snapped to it and the feasibility of shortest paths between snapped points 2, 3, and 4 (Ki-1, Ki, Kj) is checked again. If these paths are feasible, then the spatial ambiguity is resolved, and the algorithm terminates. If no alternative roadway centerline exists within the buffer for GPS data point 3, then the algorithm continues by snapping data point 4 to alternative roadway candidates contained within its buffer, and the upper part of the algorithm is executed again. If no other roadway centerlines exist within the buffer of GPS data point 4 or no feasible paths are obtained, then the lower part of the algorithm is executed and feasible paths between preceding and subsequent data points are examined. If none of the consecutive data points aid in solving the spatial mismatch between the snapped points for 2 and 3, then it is likely that no roadway centerlines within their buffers yield a feasible path and larger buffers and/or more consecutive data points need to be utilized by the algorithm. Once a feasible path is obtained, the intermediate points not employed during the map-matching process are snapped to the roadway along that feasible path.
3.2 Example of an implementation of the algorithm

The example illustrated in Figure 6 includes a set of Differential GPS (DGPS) data points collected every five seconds by a winter maintenance vehicle during the 2002-2003 winter season in Columbia County, Wisconsin. The spatial mismatch, occurring at the diverging roadways in this figure, is resolved by implementing the decision-rule map-matching algorithm. Points 0, 2, 3, and 4 are snapped to the nearest roadway within their 35-foot buffers, resulting in points S0, S2, S3, and S4 (shown as rectangles). Points S0, S3, and S4 are on the Interstate 39 centerline, while point S2 is situated on the ramp centerline. Note that no roadways are contained within the buffer for GPS data point 1, thus, this point is not used in determining the feasible path.
The shortest path between points S0 and S2 is computed using network topology and allowable turns. Consequently, the speed comparison shown in Table 1 is performed to determine if this path is feasible. In this case, the obtained path is feasible since the difference between the average calculated and recorded speeds (26.8 and 31.5 mi/h, respectively) is within tolerance (25 mi/h). Therefore, the current snapped positions for points 0 and 2 are initially assumed to be correct. The main algorithm continues by finding the shortest path between the next pair of snapped points, S2 and S3. This path is not feasible when using network topology because if the vehicle was located at S2, it would have to exit the ramp and travel approximately 5,125.9 feet in 5 seconds at an average speed of 699 mi/h to reach snapped point S3. Hence, either point S2 or S3 or both were snapped to an incorrect roadway centerline. The map-matching algorithm now obtains the shortest path between points S3 and S4 and determines that the difference between calculated and average recorded speeds with values of 29 mi/h and 35 mi/h, respectively, is within tolerance. Therefore, an alternative roadway centerline is sought within the buffer around point 2. Interstate 39 is found to be the next nearest roadway, resulting in alternative point alt2, shown as a triangle in Figure 6. Consequently, feasibility is checked for alternative point alt2, and between alt2 and its successor, snapped point S3. As indicated in Table 1, both computed shortest paths are feasible. The calculated speeds along these paths are within 25 mi/h of their respective average recorded speeds for the vehicle. Therefore, the spatial ambiguity at the diverging roadway is resolved and the correct roadway for point 2 is Interstate 39. Data point S1 is then obtained by snapping point 1 to the Interstate 39 centerline.

Fig. 6. Example of Map-Matching Algorithm at Diverging Roadways
<table>
<thead>
<tr>
<th>Data Points</th>
<th>Shortest Path Distance (ft)</th>
<th>Calculated Speed (mi/h)</th>
<th>Average Recorded Speed (mi/h)</th>
<th>Is Path Feasible?</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0 → S2</td>
<td>392.6</td>
<td>26.8</td>
<td>31.5</td>
<td>YES</td>
</tr>
<tr>
<td>S2 → S3</td>
<td>5125.9</td>
<td>699</td>
<td>33</td>
<td>NO</td>
</tr>
<tr>
<td>S3 → S4</td>
<td>213</td>
<td>29</td>
<td>35</td>
<td>YES</td>
</tr>
<tr>
<td>S0 → alt2</td>
<td>392.8</td>
<td>26.8</td>
<td>31.5</td>
<td>YES</td>
</tr>
<tr>
<td>alt2 → S3</td>
<td>215.7</td>
<td>29.4</td>
<td>33</td>
<td>YES</td>
</tr>
</tbody>
</table>

Table 1. Speed Comparison for Determining Feasibility of Shortest Paths

4. Performance analysis of the decision-rule map-matching algorithm

Success in solving spatial ambiguities depends on the values assigned to each variable of the map-matching algorithm. The analysis in this chapter examines the performance of the map-matching algorithm as values of the following parameters vary: 1) buffer size, 2) speed range, 3) number of consecutive data points, 4) temporal resolution, and 5) DGPS positional error.

4.1 Spatial data description

The data employed in this study were collected by winter maintenance vehicles in Columbia and Portage Counties, Wisconsin, and Polk County, Iowa. These counties have different accuracy roadway centerline maps with 1:2,400, 1:12,000, and 1:100,000 nominal scales, respectively, and employ different AVL/DGPS systems for data collection. Selected data sets with sampling intervals of 2 and 10 seconds were collected for different storm events and vehicle operators driving through various routes over the 2000-2001, 2001-2002, and 2002-2003 winter seasons. These routes include federal, state, and interstate highways, and local roads. Figures 7, 8, and 9 display examples of data collected in Columbia, Portage, and Polk counties every 2, 10, and 10 seconds, respectively. Notice that none of the counties employed an integrated dead reckoning system and heading information was not available during the data collection process.

![Fig. 7. DGPS Data Points Collected in Portage County Every 10 seconds](https://www.intechopen.com)
4.2 DGPS data point classification

This section identifies different cases (i.e., false negatives, false positives, no solution, incorrect and correct snap, and solved spatial ambiguities) obtained from comparing snapping results to the true roadway centerline on which a vehicle is traveling. The true vehicle path was obtained by performing a visual examination of the collected data. Data points are classified in these cases before and after applying the map-matching algorithm.

4.2.1 False negatives and false positives

False Negatives (FN) occur when data points fail to snap to any roadway centerline when they should have snapped to one. False Positives (FP) are data points that snapped to some roadway centerline when they should have not snapped to any centerline. Figure 10 shows an example of three successive GPS data points (1, 2, and 3) considered as FN. They should have snapped to Interstate 39 east bound direction, however, their buffers with radius $r$ are too small to include any roadway centerline.

4.2.2 Solved / not solved cases

If roadway centerlines exist within the buffer of a data point, then a correct snap occurs when this point snaps along the true route of the vehicle. Conversely, an incorrect snap is obtained when a data point snaps to a roadway that is not on the true route of the vehicle.
Correct and incorrect snaps are computed before and after applying the map-matching algorithm.

Figure 11 presents the cases of snapped and not snapped data points before and after applying the map-matching algorithm. The group of data points that does not snap to any roadway contains either FN or points that have no solution. Data points that have roadway centerlines within their buffers are either snapped correctly or incorrectly, or are FP. A data point that snaps incorrectly before applying the algorithm and snaps correctly afterwards is

![Diagram](https://via.placeholder.com/150)

Fig. 10. Example of Three Consecutive GPS Data Points Considered False Negatives

Fig. 11. Cases for Data Points Snapped and Not Snapped Before and After Applying the Algorithm
regarded as a solved case. If a data point is snapped incorrectly before applying the algorithm and it is snapped incorrectly after applying the algorithm, then the spatial mismatch is not solved. If this occurs, then some neighboring data points may be left incorrectly snapped. Note that FN, FP, and no solution are not included in the solved and not solved case analysis.

In the following section of this chapter, FN data points are minimized and solved spatial mismatches are maximized after applying the algorithm. Although FP and no solution cases occur due to spatial database incompleteness, they amount to less than 0.5% of the total number of data points examined in this study. Therefore, these two cases were not taken into account in the analysis.

4.3 Analysis of the impact of variables on the performance of the map-matching algorithm

This section examines each algorithmic variable independently to determine its effect on the performance of the map-matching algorithm. These variables are classified into two groups. One group consists of parameters controlled by the user (i.e., buffer size, speed range, number of consecutive data points) and the other group comprises parameters controlled through the data (i.e., temporal resolution and DGPS error).

4.3.1 Buffer size

The appropriate buffer size employed during the snapping process when solving spatial ambiguities depends on the quality and geometry of the spatial data. This proximity parameter used to 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, many of which will not be correct.

Roadways are typically represented by centerlines that do not account for lane widths. Therefore, data points will almost always appear offset some distance from roadway centerlines in addition to being affected by errors in the DGPS measurements and digital roadway maps (Wolf & Ghilani, 1997). Hence, the buffer size parameter was tested at 10-ft increments from 20 ft to 60 ft for data collected in Columbia and Portage Counties, and at 20-ft increments from 20 to 100 ft for data collected in Polk County. The latter is due to the smaller scale of the Polk County roadway centerline map. These buffer size values were predetermined through the computation of average distance percentages between the data points and roadway centerlines. As different buffer sizes were analyzed and tested against the map-matching algorithm, the speed range tolerance and number of consecutive data points were maintained constant with values 25 mi/h and 5, respectively.

Figure 12 shows a chart with the average percentages of FN before and after applying the algorithm, as the buffer size varies for Columbia, Portage, and Polk County. This figure indicates that lower FN percentages are obtained after applying the algorithm for all three counties. Portage and Polk counties present the largest decrease of FN percentages with an average difference of 20% before and after executing the algorithm. Overall, average percentages of FN data points diminish as the buffer size increases since more data points are snapped to roadway centerlines.
A Decision-Rule Topological Map-Matching Algorithm with Multiple Spatial Data

Fig. 12. FN Percentages Before and After Applying Algorithm by Buffer Size for Columbia, Portage, and Polk Counties

Fig. 13. Percentages of Solved Cases After Applying Algorithm by Buffer Size for Columbia, Portage, and Polk Counties
Figure 13 presents the percentage of solved spatial ambiguities after applying the map-matching algorithm for Columbia, Portage, and Polk counties. This chart indicates that over 90% of incorrectly snapped data points collected in Columbia County were solved by the algorithm when employing a 30-foot buffer size. Whereas, solved cases reached their maximum values (68% and 64%) for Portage and Polk counties with 50 and 60-foot buffers, respectively. As mentioned earlier, Polk County data was tested for buffer sizes every 20 feet, thus, there is no data for buffer sizes equal to 30 and 50 feet.

### 4.3.2 Speed

The map-matching algorithm determines the correct roadway centerline on which a vehicle is traveling by computing feasible shortest paths between snapped data points. This feasibility is sensitive to the allowable range utilized when comparing computed and recorded speeds. The analysis of this variable examines the effect that it has on the performance of the map-matching algorithm.

The average recorded speed \( (v) \) is computed using the recorded speeds \( (v_1 \) and \( v_2) \), as shown in Equation 1. Equation 2 presents the computed speed calculation \( (s) \) given the shortest distance traveled \( (D) \) and timestamps \( (t_1 \) and \( t_2) \) between a pair of snapped data points. Subsequently, the algorithm accepts a tested path as feasible if the average recorded speed is within the equally distributed speed range shown in Equation 3.

\[
v = \frac{v_1 + v_2}{2} \tag{1}
\]

\[
s = \frac{D}{(t_2 - t_1)} \tag{2}
\]

\[v \in s \pm \frac{\text{SpeedRange}}{2} \tag{3}\]

FN curves were computed for various buffer sizes and different speed range tolerances from 5 to 35 mi/h with increments of 5 mi/h for the three counties. Analysis results for this variable show that feasible paths are rejected when small speed ranges are employed leaving FN data points not snapped to any roadway centerline. On the contrary, as speed range increases, FN percentages diminish since feasible paths are found during the speed comparison process. Figure 14 shows FN curves for Columbia County with data collected every 2 seconds. These curves are approximately parallel as the speed range varies, and stabilize for speed ranges greater than 15 mi/h. Speed ranges equal to or greater than 25 mi/h are needed to minimize FN percentages in Portage and Polk counties. Further speed range increase does not improve the results because all feasible paths are accepted. In general, FN curves are steeper for small buffer sizes, and approach near-zero slope for buffer sizes equal to or greater than 40 feet.

Analysis results for this variable indicate that the percentage of solved cases increases as speed range also increases. The percentage of solved cases has the highest value of approximately 90% when the algorithm employs speed ranges equal to or greater than 20 mi/hr and a 30-foot buffer for Columbia County data. Conversely, there is no considerable
increase in the percentage of solved cases remaining at 68% for Portage County data when speed range values equal to or greater than 25 mi/h are employed. The percentage of solved cases for Polk County remained constant at 50% for speed ranges equal to or greater than 15 mi/h, independent of buffer size. Thus, the map-matching algorithm is sensitive to speed range values, particularly when small speed ranges are employed since feasible paths are rejected.

Fig. 14. FN Percentages After Applying Algorithm for Different Speed Ranges by Buffer Size for Columbia County

4.3.3 Number of consecutive GPS data points

If no feasible paths are obtained between a pair of snapped data points, then the algorithm tests for viable routes between preceding and subsequent data points, as described in Figure 5. If a small buffer size is utilized, several successive data points do not snap to any roadway centerline generating FN data points. Thus, the number of consecutive data points used by the algorithm needs to be incremented to consider adjacent data points that are correctly snapped and minimize FN percentages.

Although the map-matching algorithm may employ any number of consecutive data points, the performance of the map-matching algorithm was analyzed with a number of consecutive data points between three and eight. A previous test determined that this range of consecutive data points is sufficient for solving spatial ambiguities with the spatial and temporal data employed in this study.

Similar to the FN curve behavior due to speed range variations, FN curves for different number of consecutive data points are parallel for the three counties and converge to constant values as the buffer size increases. Figure 15 shows the percentage of FN data points as the
number of consecutive data points varies by county with a 40-foot buffer size. No significant improvements are identified in the percentage of FN for the three counties as the number of consecutive data points varies, except for Portage County data that presents a decrease in the amount of FN when increasing the number of consecutive data points from three to four.

![Graph showing percentages of FN after applying algorithm by number of consecutive data points and county](image)

**Fig. 15. Average Percentages of FN After Applying Algorithm by Number of Consecutive Data Points and County**

The percentage of solved spatial mismatches increased as the number of consecutive data points increased. Eight consecutive data points solve almost 100% of initial incorrect snaps for Columbia County data when employing a 40-foot buffer. The largest percentage of solved mismatches (over 70%) after applying the algorithm occurs with a 50-foot buffer for Portage County. While the percentage of solved cases in Polk County remained constant at 50%, as the buffer size and number of consecutive data points increased. The results of this analysis show that increasing the number of consecutive data points solves a larger number of spatial ambiguities. By increasing this number, the algorithm resolves ambiguities that arise when alternative roadway centerlines are equally viable.

### 4.3.4 Temporal resolution

The outcome of the map-matching technique is not only affected by spatial inaccuracies, it is also influenced by the collection frequency of the data points. As temporal resolution increases, the tracking of the vehicle becomes more accurate. On the other hand, the sampling interval impacts the sizes of the data sets. Processing of large data sets takes significant CPU time, and increases storage requirements. Hence, there is a tradeoff between decreasing the sampling interval and quality of collected speed data.

Data sets collected in Columbia County with an original 2-second time interval were processed to generate data files with lower temporal resolutions varying from 2 to 30 with increments of 4 seconds. Similarly, data collected every 10 seconds in Portage and Polk counties were processed to create data files with temporal resolutions equal to 10, 20, and 30 seconds, respectively. The speed range and number of consecutive points remained constant with values 25 mi/h and 5, respectively.
Fig. 16. FN Percentages Before and After Applying Algorithm for Different Temporal Resolutions by Buffer Size for Polk County

Fig. 17. Percentages of Solved Cases for Different Temporal Resolutions by Buffer Size for Portage County
Figure 16 illustrates FN curves before and after applying the algorithm for different temporal resolutions with data originally collected every 10 seconds in Polk County. The graph presents relatively parallel FN curves for all data collection frequencies. These curves show that as temporal resolution increases, the percentages of FN data points decreases. FN curves after applying the algorithm show lower percentages of FN data points compared to before executing the algorithm. FN curves for Columbia and Portage counties behave similarly with different sampling intervals. All county cases illustrate that larger amount of FN points occur when using smaller buffers independent of data collection frequency. Figure 17 shows the variation of solved cases as temporal resolution increases in Portage County. Percentages of solved spatial ambiguities increase as data is collected at higher frequencies, being the largest at a 50-foot buffer with 68%. This percentage decreases in average for Columbia County data from approximately 80% to 20% as sampling intervals increase from 5 to 30 second for all buffer sizes. The same behavior is apparent for solved case percentages in Polk County as data is collected more frequently.

4.3.5 GPS error

GPS measurements are affected by both systematic and random errors. Their combined magnitudes will affect the accuracy of the positioning results. Systematic errors obey physical or mathematical law, and can be computed and applied to measurements to eliminate their effects (Ghilani & Wolf, 2006). Random errors occur because of stochastic noise in the measurement process producing different coordinates each time a measurement is achieved, even during short intervals. This type of error is assumed to be Gaussian affecting both latitude and longitude or X, Y coordinates. DGPS is a method that increases the accuracy of CA code measurements by canceling some of the inherent systematic errors. Any potentially remaining systematic errors were not modeled in this study, and only the effects of random errors were examined.

Random errors were simulated by using a normal distribution random number generator (Box & Muller, 1958) for known means and different standard deviations. If $U_1$ and $U_2$ are a pair of independent uniformly-distributed random numbers from the rectangular density function on the interval $(0, 1)$, then a pair of independent random numbers ($X_1$ and $X_2$) from a normal distribution with mean zero and standard deviation $\sigma$ are generated using Equations 4 and 5.

$$X_1 = (-2 \log U_1)^{1/2} \cos(2\pi U_2)$$ (4)

$$X_2 = (-2 \log U_1)^{1/2} \sin(2\pi U_2)$$ (5)

Experiments conducted by the Wisconsin Winter Maintenance Concept Vehicle project concluded that random DGPS errors were on the order of 2 to 5 meters, root-mean-square (Vonderohe et al., 2001). Therefore, a mean value of zero and standard deviations of ±2 and ±5 meters were employed in this analysis. Speed range and number of consecutive points values were held fixed as 2- and 5-meter standard deviation errors were introduced in the DGPS data points.

Percentages for FN and solved cases were computed to compare the performance of the algorithm for original and perturbed DGPS data points. Figure 18 presents variations in the percentage of FN data points for original and perturbed data by county for a 40-foot buffer.
Fig. 18. FN Percentages Before and After Applying the Algorithm for Original Data, 2 m, and 5 m Error with a 40-foot Buffer by County

Fig. 19. Percentage of Solved Cases After Applying the Algorithm for Original Data, 2 m, and 5 m Error by Buffer Size in Columbia County
before and after applying the algorithm. All FN percentages decrease after executing the algorithm, independent from the spatial data quality. Average FN percentages computed with original data present smaller values than data perturbed with 2- and 5-m error before and after applying the algorithm. For example, FN percentages increase from 22% to 48% for Polk County after executing the algorithm when introducing a 5-m error. In general, the percentage of data points that should snap to a roadway centerline increases when there is larger error in the DGPS data points.

Figure 19 presents the percentages of solved spatial ambiguities by the algorithm before and after perturbing the DGPS data points with original and simulated random errors (2- and 5-meter standard deviation) for Columbia County. This figure shows that the percentage of incorrect snaps solved after applying the algorithm for original Columbia County data are larger than those computed with perturbed data. On average, the percentage of solved cases decreases approximately 20% and 40% for data with 2- and 5-meter error for all buffer sizes, except for the 20-foot buffer. This small buffer is not able to accommodate the spatial ambiguities that arise with simulated data. Similarly, Portage and Polk counties present a drop in the percentages of solved data points from approximately 68% and 50% for original data to approximately 10% and 15%, respectively, for 5 m perturbed data.

5. Summary and conclusions

Transportation applications employ AVL/DGPS technology to collect vehicle positions and other sensor data. Normally, DGPS data points are associated with roadways by snapping to the nearest centerline in a GIS environment. The map-matching problem or spatial ambiguities arise during this association due to errors in DGPS measurements and digital cartography. Such ambiguities are common at underpasses and converging or diverging roadways. These can result in DGPS data points being snapped to incorrect roadway centerlines affecting the calculation of cumulative distance traveled by the vehicles along a roadway network, or the allocation of non-spatial data collected from vehicle sensors to incorrect roadways. Thus, this problem propagates to the computation of performance measures or decision management tools.

This study contributes with the development and implementation of a post-processing decision-rule map-matching algorithm that resolves many of these spatial ambiguities by examining the feasibility of paths between pairs of snapped data points. A viable path is the shortest-distance path between two snapped points that a vehicle can travel, while following network topology and turn restrictions, at a speed comparable to its average recorded speed. If a given shortest path is not feasible, then DGPS data points are related to other roadway centerlines within their buffers and new shortest paths are calculated; or adjacent DGPS data points are used to determine feasible paths. Examples were presented to describe the step-by-step process of the map-matching algorithm. Five variables were studied independently to analyze the performance of the map-matching algorithm. These variables are buffer size, speed range tolerance, number of consecutive points, temporal resolution, and positional error in the DGPS data points. Data collection frequency and DGPS error are variables controlled externally through the data, while buffer size, speed range, and number of consecutive data points are algorithm parameters are controlled by the user.
The results of this study indicate that the success of the map-matching algorithm in solving spatial ambiguities depends on not only by the variables employed by the algorithm, but also by the sampling interval and the quality of the spatial measurements and roadway map scale. If lower spatial data qualities and less frequent sampling intervals are used, then the algorithm requires larger buffers and speed ranges to obtain best results. On the other hand, if GPS data points collected more frequently are snapped to higher accuracy maps, such as the Columbia County case, then larger percentages of incorrect snaps are solved and smaller buffer sizes are adequate. By increasing the number of consecutive data points, a larger number of spatial ambiguities are solved, particularly when alternative roadway centerlines are equally viable, and FN percentages are reduced since more combinations are examined between pairs of snapped DGPS data points. However, no significant variations in the solved results for Polk County are apparent as the number of consecutive data points increases since lower spatial data accuracies were used in this county. Table 2 presents the best and worst variable values encountered when solving incorrect snaps after applying the map-matching algorithm by county. This table indicates that larger speed range values, and numbers of consecutive points provide better results in maximizing solved cases. Stable percentage values are reached as both speed range and number of consecutive points reach certain values. While small speed ranges tend to reject tested paths, larger speed ranges accept most of these paths without improving the performance of the algorithm. Similarly, larger percentages of solved cases are obtained as the number of consecutive points increases since additional combinations between pairs of snapped data points are examined. Overall, higher parameter values yield better results as data are collected less frequently and snapped to lower quality roadway maps.

<table>
<thead>
<tr>
<th>County</th>
<th>Buffer Size (ft)</th>
<th>Speed Range (mi/hr)</th>
<th>Number of Consecutive Points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best</td>
<td>Worst</td>
<td>Best</td>
</tr>
<tr>
<td>Columbia</td>
<td>30</td>
<td>≥50</td>
<td>35</td>
</tr>
<tr>
<td>Portage</td>
<td>50</td>
<td>20</td>
<td>≥25</td>
</tr>
<tr>
<td>Polk</td>
<td>40</td>
<td>20</td>
<td>≥15</td>
</tr>
</tbody>
</table>

Table 2. Best and Worst Variable Values for Solved Cases by County

Introducing positional error in the DGPS data points decreases the percentage of solved incorrect snaps and total number of snapped data points obtained before and after applying the algorithm. As the positional error increments from 2 to 5 meters in standard deviation, the percentage of solved cases decrease and FN percentages increase for all counties. Thus, larger buffer sizes and speed ranges are needed for lower quality data. Future research is required to explore these parameter values against additional spatial data qualities derived from multiple ITS applications. Further research may involve online implementation of the map-matching algorithm, in which spatial ambiguities are solved as GPS measurements are collected in real-time.

6. References


Global Navigation Satellite Systems (GNSS) plays a key role in high precision navigation, positioning, timing, and scientific questions related to precise positioning. This is a highly precise, continuous, all-weather, and real-time technique. The book is devoted to presenting recent results and developments in GNSS theory, system, signal, receiver, method, and errors sources, such as multipath effects and atmospheric delays. Furthermore, varied GNSS applications are demonstrated and evaluated in hybrid positioning, multi-sensor integration, height system, Network Real Time Kinematic (NRTK), wheeled robots, and status and engineering surveying. This book provides a good reference for GNSS designers, engineers, and scientists, as well as the user market.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
