Effects of Controlling Parameters on Performance of a Decision-Rule Map-Matching Algorithm

Carola A. Blazquez[1] and Alan P. Vonderohe[2]

Abstract: Advanced map-matching algorithms resolve spatial ambiguities between differential global positioning system (DGPS) and roadway centerline data. Most of these algorithms need further research to assess their performances with respect to their controlling parameters and their relationships with spatial data and temporal resolution. This paper presents an analysis of the effects of three parameters controlled by the user and two variables dominated externally through simulated data on the performance of a postprocessing decision-rule map-matching algorithm previously developed by the writers. The algorithm is tested against three different digital roadway map scales from counties in Wisconsin and Iowa, and two automatic vehicle location (AVL)/DGPS technologies mounted on intelligent winter maintenance vehicles. Sensitivity analyses indicate that the algorithm is sensitive to controlling parameter values depending on the data being tested. The algorithm satisfactorily resolves spatial ambiguities given different spatial data qualities, AVL/DGPS technologies, and temporal resolutions. Statistical analysis suggests a direct relationship between data collection frequency and spatial mismatch resolution. Parameter values are presented for minimizing false negatives and maximizing solved cases, thus enhancing the performance of the map-matching algorithm.

DOI: 10.1061/(ASCE)TE.1943-5436.0000078

CE Database subject headings: Global positioning; Spatial data; Intelligent transportation systems; Algorithms; Mapping.

Introduction

The global positioning system (GPS), integrated with geographic information systems (GIS), are components of innovative advanced technology applied in intelligent transportation systems to make transportation more efficient, less congested, safer, and less polluting. Difficulties can arise with the use of GPS and wireless technology when two-dimensional or three-dimensional coordinates are transformed to one-dimensional linearly referenced locations to support transportation applications such as highway inventory (Vonderohe 1999). During this transformation, map-matching problems occur when GPS data points are snapped to incorrect roadway centerlines due to complexities of roadway networks (Chen et al. 2005), and error in GPS measurements and digital roadway maps.

Transportation applications such as in-vehicle navigation systems, dynamic route guidance, personal navigation assistants, travel speed and delay studies, congestion and traffic management, travel time studies, locating motor vehicle crashes, transit operations and management, commercial vehicle operations, fleet management, hazardous material management, infrastructure management, incident management, transportation policy analysis, and household travel surveys are in demand of an efficient map-matching algorithm associated with registration of GPS data points to a digital roadway map (Czerniak 2002; Noronha and Goodchild 2000; Syed and Cannon 2004). Associating vehicle locations and other collected data with correct roadway centerlines is essential for these applications.

Many different approaches for map-matching algorithms are described in the literature. According to Quddus (2006), these algorithms are classified as geometric, topological, probabilistic, and advanced techniques such as Kalman filters, fuzzy logic, and Bayesian statistics.

However, only a few existing map-matching algorithms have had performance assessments that evaluated success in resolution of spatial mismatches. For example, Syed and Cannon (2004) analyzed the performance of a map-matching algorithm by computing the percentage of correctly map-matched solution epochs out of the total number of epochs. Jagadeesh et al. (2004) assessed the performance of their proposed fuzzy rule set map-matching algorithm by computing a correct road-matching ratio in a custom-built simulation environment that reproduced on-road conditions. Chen et al. (2005) evaluated the performance of their hybrid map-matching algorithm based on errors in the roadway map, GPS positions, and distance and heading measurements. Wenk et al. (2006) assessed the performance between their adaptive clipping and existing incremental algorithms with respect to the quality of map-matching results (i.e., weak, strong, and average Frechet distance), running time, and database input/output (I/O) operations. The writers stated that their algorithm still needs to be evaluated under real-case scenarios. Li et al. (2005) tested overall performance of a road reduction filter with height aiding that tracks a vehicle on all road candidates in an error region, and eliminates those on inappropriate roadway centerlines. RMS values were computed for height and positional errors, assuming that real-time kinematic measurements represented true vehicle positions. A methodology for predicting the performance of a map-

[1]Assistant Professor, Dept. of Engineering Science, Universidad Andres Bello, Sazie 2315 Piso 2, Santiago, Chile (corresponding author). E-mail: cblazquez@unab.cl

[2]Emeritus Professor, Dept. of Civil and Environmental Engineering, Univ. of Wisconsin-Madison, 1204 Engineering Hall 1415 Engineering Dr., Madison, WI 53706. E-mail: vonderoh@engr.wisc.edu

Note. This manuscript was submitted on December 10, 2008; approved on June 16, 2009; published online on June 20, 2009. Discussion period open until May 1, 2010; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering***, Vol. 135, No. 12, December 1, 2009. ©ASCE, ISSN 0733-947X/2009/12-966–973/$25.00.

matching algorithm proposed by Karimi et al. (2006) indicated the probability that their algorithm will perform an accurate match. This map-matching algorithm employed a point-to-curve with a shortest distance matching approach. Further research is needed to determine how to apply the methodology on more complex algorithms and test it on real scenarios. Quddus (2006) quantified the effects of digital map quality and navigation sensor data on performances of different map-matching algorithms. Statistical values were computed for horizontal accuracy and correct identification of links using three map scales. Positional accuracies of data points along vehicle routes and error propagation are aspects of integrated automatic vehicle location and differential GPS (AVL/DGPS) or DGPS and dead reckoning (DR) (DGPS/DR) systems employed by some writers (Taylor et al. 2001; Quddus 2006) to estimate the performance of map-matching algorithms. However, the map-matching algorithm studied in this paper focused solely on correct vehicle route determination.

Success in resolving spatial ambiguities depends on values assigned to each controlling or weighting parameter of a map-matching algorithm. Quality and geometry of the spatial data (e.g., GPS measurements and roadway centerline map) impact the appropriateness of values for these parameters and, thus, the optimal performance of a map-matching algorithm. The literature is scarce on indicators of the correct parameter, factor or threshold values adequate for multiple spatial data set qualities (Quddus 2006; Schlingelhof et al. 2008). Therefore, parameter value determination needs investigation to provide insight into the performance of existing map-matching algorithms (Taylor et al. 2006; Quddus et al. 2007).

The objective of this study is to examine the effects of the three controlling parameters (buffer size, speed range tolerance, and number of consecutive data points) for the decision-rule map-matching algorithm previously developed by the writers (Blazquez and Vonderohe 2005). This algorithm employs Dijkstra's algorithm to compute feasible shortest paths between pairs of data points, yielding a time complexity of $O(kN^2)$, where k is the number of DGPS data points, and N is the number of nodes of the roadway network. In addition, an analysis is presented of the impact of two parameters controlled externally through simulated data (DGPS positional error and temporal resolution) on the performance of the algorithm. Since different geospatial technologies and spatial qualities influence the performance of map-matching algorithms, each of these parameters was tested against three digital roadway network maps with different nominal scales (1:2,400, 1:24,000, and 1:100,000), and two different AVL/DGPS technologies. These technologies were mounted on intelligent winter maintenance vehicles that collected data every 2, 5, and 10 s along interstate, state, and county highways in rural and suburban areas of Wisconsin and Iowa during the 2002–2003 winter season.

When integrating DGPS data points and roadway network data in a GIS environment, spatial analysis software projects DGPS data points orthogonally onto nearest roadways by calculating the minimum distance between each roadway representation and each DGPS data point. This process is called "snapping" (Vonderohe et al. 2006). A methodology is presented for classifying snapped and unsnapped data points before and after applying the map-matching algorithm. Before applying the algorithm, data points are projected to the nearest roadway centerline yielding potential spatial ambiguities. After implementing the map-matching algorithm, most of these spatial mismatches are resolved by determining the correct roadway on which a vehicle was traveling. Average percentages for false negative (FN) data points, that initially fail to snap to a roadway centerline when in fact they should have snapped to one, and solved spatial ambiguities are calculated for each parameter by comparing true and computed vehicle routes.

Parameter values are presented for different spatial database qualities and AVL/DGPS technologies that minimize FN data points and maximize solved spatial ambiguities, thus enhancing the performance of the decision-rule map-matching algorithm. The algorithm performs in postprocessing mode and is applicable to map-matching problems that do not require real-time navigation data. For example, correct winter maintenance vehicle routes were obtained in this study, by postanalysis in a GIS environment, to obtain total traveled distances needed to compute performance measures (e.g., sand application rate by patrol section for a winter storm event) and decision management tools. Emergency response, pickup/delivery services, and urban solid waste collection are typical examples of logistics and transportation applications that employ postprocessing map-matching algorithms for strategic decision making and planning. In this type of application, a fleet of vehicles collects and stores GPS data for later retrieval, processing, and analysis. Optimal routes are determined that minimize operational and transportation costs (e.g., fuel consumption and maintenance), travel time, and air pollutant emissions while improving the overall service. Other transportation applications, such as video or photo logging and highway infrastructure inventory, require postanalysis map-matching algorithms to support infrastructure problem identification, highway safety control, resource allocation, and inventory maintenance.

Decision-Rule Map-Matching Algorithm

The decision-rule map-matching algorithm resolves spatial ambiguities by determining the correct roadway centerline on which a vehicle is traveling. First, the algorithm selects all roadways within a buffer around a DGPS data point and computes the orthogonal projection of the data point to the closest roadway by determining the minimum perpendicular distance between this point and each roadway centerline (Blazquez and Vonderohe 2005). For example, DGPS Data Points 1 and 2 are snapped to Ramp 2 because it is the closest roadway centerline contained within the buffers around each data point, as shown in Fig. 1. Subsequently, the shortest path is obtained between the two snapped DGPS data points (S1 and S2) using network topology and turn restrictions. A path is considered viable and snapped data point locations are accepted, if the difference between the computed travel speed and the average recorded vehicle speed for the data points is within a speed range tolerance. If a path is rejected, 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 data points.

Differential Global Positioning System Data Point Classification

Percentages of FN, and false positive (FP) data points are commonly employed to detect false alarms and misses, respectively, in algorithms and models in geospatial data (Chawla et al. 2001), geospatial health (Aron 2006), combinatorial chemistry (Zhang et

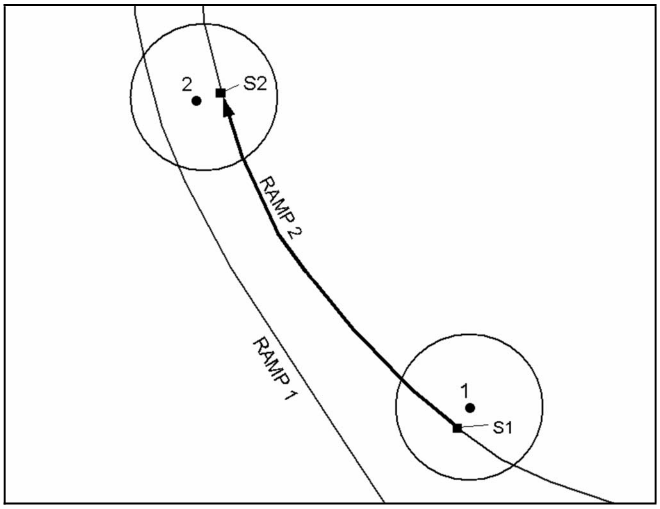

Fig. 1. Example of snapping to the correct roadway for two DGPS data points using the map-matching algorithm

al. 2000), pattern recognition (Haga et al. 2004), and feature extraction (Finlayson and Opitz 2004).

In this research, FP data points are DGPS data points that snap to a roadway centerline when they should not have snapped to any centerline. FN data points occur when DGPS data points fail to snap to a roadway centerline when they should have snapped to one. Additionally, no solution, correct, and incorrect snapping cases are identified within the collected data points. The no solution case occurs when DGPS data points do not snap to any roadway centerline because the vehicle is traveling where there is no roadway represented in the spatial database due to missing information or generalization. Correct snapping cases are obtained when data points are projected to roadways that are on the TR of the vehicle, and incorrect snapping cases consist of data points snapped to roadways not on the true vehicle route.

The methodology for DGPS data point classification is described by the flow diagram in Fig. 2. Data points are classified as FN, FP, no solution, and incorrect and correct snapping cases by comparing computed and true vehicle routes. Computed vehicle routes are obtained from snapped data points, while true vehicle routes are determined by performing a visual inspection of the collected data. Computed routes (CRi) and true routes (TRs) are equal to zero if data points do not snap to any roadway centerline, and are not equal to zero if data points are snapped to roadway centerlines contained within their buffers. For example, if data point i does not snap to any roadway centerline (CRi=0) and the vehicle actually traveled on a roadway (TR< >0), then data point i is classified as FN.

The classification of DGPS data points is performed before and after applying the map-matching algorithm, as shown in Fig. 3. The top and bottom rectangles refer to snapped (e.g., FP, correct or incorrect snaps) and not-snapped (e.g., FN or no solution cases) data points before and after applying the map-matching algorithm, respectively. Before applying the algorithm, data points are projected onto the nearest roadway centerline yielding potential spatial ambiguities. After applying the map-matching algorithm, most spatial mismatches are resolved by determining the correct roadway on which the vehicle was traveling. If a data point snaps incorrectly before applying the algorithm and snaps correctly after applying the algorithm, then it is regarded as a solved case. If a data point is snapped incorrectly before and after

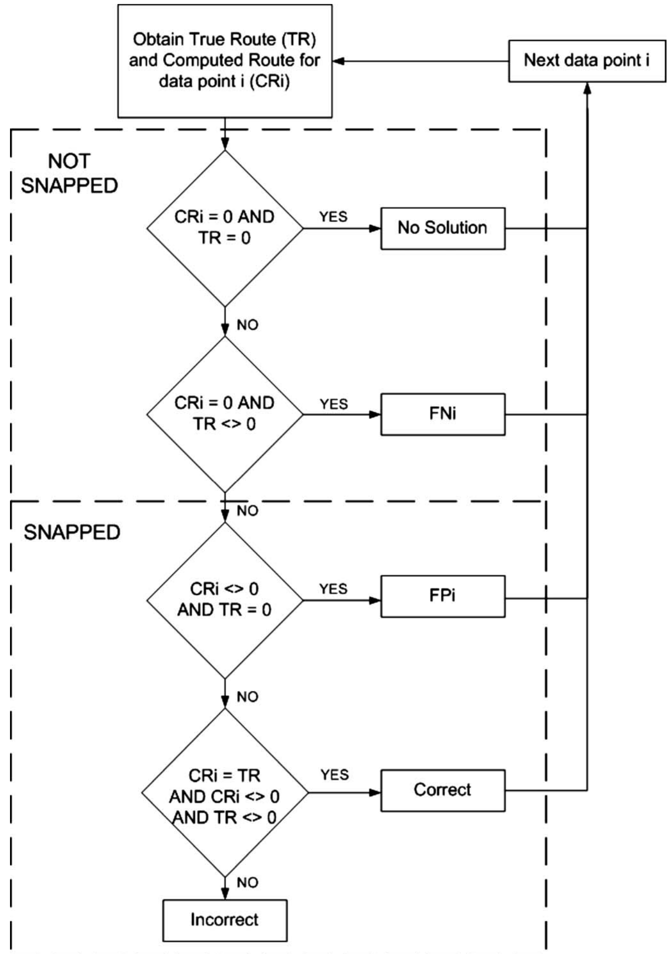

Fig. 2. Flow diagram for DGPS data point case classification

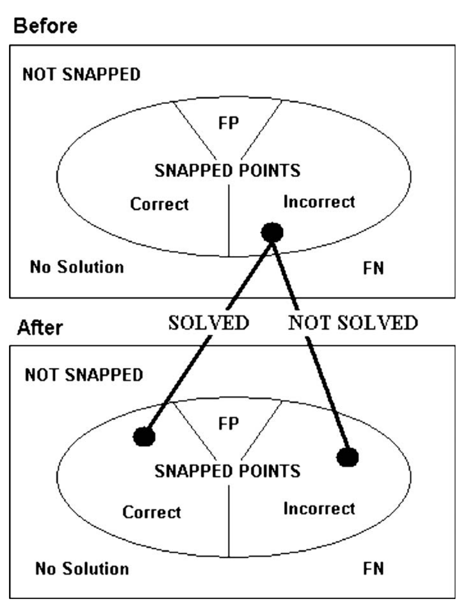

Fig. 3. Cases for snapped and not-snapped data points before and after applying the algorithm

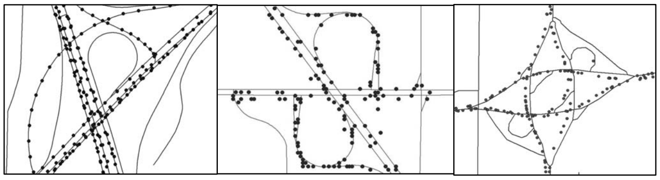

Fig. 4. Examples of spatial database and DGPS measurements for Columbia, Portage, and Polk Counties

applying the algorithm, then the spatial mismatch is not solved.

Average percentages for FN and solved data points are presented in the following section for different algorithm parameter and spatial data values. Ideally, FN data points should be minimized and solved cases should be maximized after applying the algorithm. FP and no-solution cases occur due to spatial database incompleteness. Even though, these cases were identified when testing the field data, they amount to less than 0.5% of the total number of DGPS data points. Therefore, FP and no solution cases are not considered in the sensitivity analysis.

Sensitivity Analysis of Controlling Parameters

Parameters of the decision-rule map-matching algorithm (buffer size, speed range tolerance, and number of consecutive data points), and two variables controlled externally through simulated data (DGPS positional error and data collection frequency) were examined independently to determine their effects on performance of the algorithm. Average percentages for FN and solved cases were computed using data collected by winter maintenance vehicles every 2 and 5 s in Columbia County, Wisconsin and every 10 s in Portage County, Wisconsin and Polk County, Iowa. These vehicles collected speed data, environmental data (e.g., pavement and air temperature), equipment status data (e.g., plow up or plow down), and material usage data (e.g., salt application rate) along with DGPS measurements for different storm events and vehicle operators driving along interstate, state, and county highways during the 2002–2003 winter season.

Fig. 4 shows data samples on similar roadway network complexities with 1:2,400, 1:24,000, and 1:100,000 nominal scales, respectively, employed in the analysis. In this study, over 2,500 data points were collected in Portage County using AVL/DGPS technology provided by vendor Force America, while approximately 11,000 and 1,500 data points were collected in Columbia and Polk Counties, respectively, using Raven technology. None of the counties employed an integrated DGPS/DR system and heading information was not available during the data collection process.

Buffer Size

The appropriate buffer size depends on the quality and geometry of the spatial data. Very small buffer sizes might lead to selection of no roadway centerlines, whereas extremely large buffer sizes result in lower efficiencies since the algorithm must examine many alternative roadways.

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 et al. 1999). Considering the contributions of lane width, DGPS positional error, and map error, the buffer size parameter was tested at 3-m (10-ft) increments from 6 m (20 ft) to 18 m (60 ft) for data collected in Columbia and Portage Counties, and at 6-m (20-ft) increments from 6 to 30 m (~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 DGPS data and roadway centerlines.

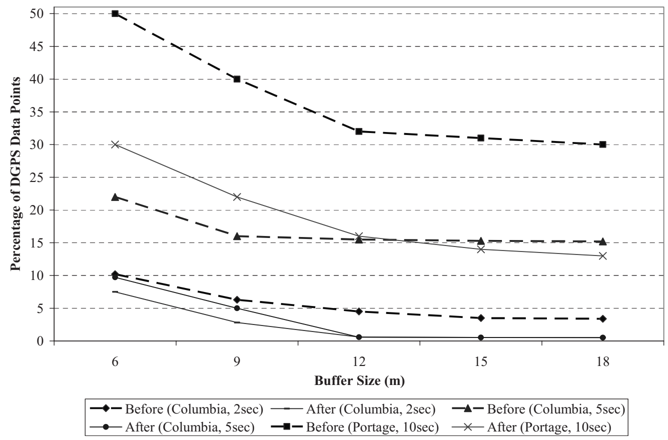

Fig. 5. Average FN percentages before and after applying algorithm by buffer size for Columbia and Portage County data collected every 2, 5, and 10 s

Fig. 5 shows a chart of average percentages of FN data points, before and after applying the algorithm, as buffer size varies for Columbia County data collected every 2 and 5 s, and Portage County data collected every 10 s. Lower FN percentages are obtained after applying the algorithm for all temporal resolutions (2, 5, and 10 s) and different roadway centerline map scales. Average percentages of FN data points diminish as buffer size increases because more data points are snapped to roadway centerlines. FN percentages stabilize at constant values after a certain buffer size is reached. For example, average FN percentage decreases from approximately 10 to 0.5% as buffer size increases from 6 to 12 m for Columbia County data collected every 5 s after applying the algorithm.

Over 90% and approximately 80% of incorrectly snapped DGPS data points collected every 2 s and 5 s, respectively, in Columbia County were solved by the algorithm. The largest percentage of solved spatial ambiguities for this county occurred when a 9-m buffer was employed for both 2- and 5-s sampling intervals. While only maximum values of 68% and 50% for solved cases occurred after executing the algorithm at 12- and 15-m buffer sizes for Portage and Polk Counties, respectively.

Speed

The feasibility of a shortest vehicle path between a pair of snapped DGPS data points is sensitive to the allowable speed range used during the computed and recorded speed comparison. Average recorded speed (v) is computed using recorded speeds (v_1 and v_2) of snapped data points 1 and 2, as shown in Eq. (1)

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

The computed speed (s) is computed with Eq. (2)

$$s = \frac{D}{(t_2 - t_1)} \quad (2)$$

where D is the shortest traveled distance calculated using Dijkstra's algorithm, and t_1 and t_2 are timestamps of the snapped data points. Subsequently, the algorithm accepts a tested path as feasible if the average recorded speed is within the equally distributed speed range [see Eq. (3)]

$$v \in s \pm \frac{\text{speed range}}{2} \quad (3)$$

FN curves were computed for multiple buffer sizes and different speed range tolerances from 8 to 56 km/h (5 to 35 mi/h) with increments of 8 km/h (5 mi/h) for Columbia and Portage Counties, and 16 km/h (10 mi/h) for Polk County. These controlling parameter values were obtained from average measurements of the AVL/DGPS system.

Feasible paths are rejected when small speed ranges are employed, leaving DGPS data points not snapped to any roadway centerline and, therefore, yielding FN data points. FN percentages decrease for different speed ranges as buffer size increases. FN percentages converge to stable minimum values for speed ranges equal to or greater than 24 km/h (15 mi/h) for Columbia County data and 40 km/h (25 mi/h) for Portage and Polk County data. Further speed range increase does not improve results significantly since all feasible paths are accepted. Additionally, solved case percentages remain relatively constant at approximately 90% and 80% for data collected every 2 and 5 s, respectively, in Columbia County for all buffer sizes and speed ranges equal to or greater than 32 km/h (20 mi/h). Spatial ambiguity percentages for Portage and Polk County data remained unsolved at 68% and 50%, respectively, with speed range tolerances greater than 40 km/h

Number of Consecutive Differential Global Positioning System Data Points

If no viable shortest path exists between two snapped DGPS data points, then the map-matching algorithm tests for feasible paths between preceding and subsequent data points. The decision-rule map-matching algorithm uses five consecutive data points, as described previously by the writers (Blazquez and Vonderohe 2005). In this analysis, the algorithm was modified to employ from three to eight consecutive data points with increments of one.

FN curves for this parameter behave similarly to the previous controlling parameters, converging to constant values as buffer size increases. No significant improvements were identified with Polk and Portage County data for greater than three and four consecutive data points, respectively. However, percentages of FN decreased considerably for Columbia County data collected every 2 and 5 s as the number of consecutive data points was increased from 3 to 8. Additionally, the percentage of solved spatial mismatches for this controlling parameter increased as the number of consecutive data points increased. However, the percentage of solved cases in Polk County remained constant at 50% as the buffer size and number of consecutive data points increased.

Global Positioning System Positional Error

GPS measurements are affected by both systematic and random errors. Corrections for systematic errors can be computed and applied to measurements to eliminate their effects (Wolf and Ghilani 2006). Random errors occur because of stochastic noise in the measurement process, meaning that a stationary GPS receiver produces different coordinates in every measurement. Random error in GPS measurements is assumed to be Gaussian and affects both latitude and longitude or X, Y coordinates.

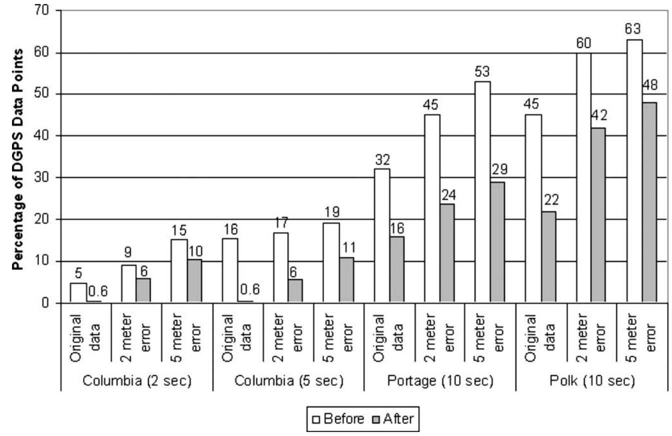

Fig. 6. Average FN percentages before and after applying algorithm for original data, 2 m, and 5 m error with a 12 m buffer by county

The accuracy of a GPS unit is determined by the sum of errors from several sources. According to Hofmann-Wellenhof et al. (1993), Czerniak (2002), and Kaplan and Hegarty (2006), these error sources are receiver clock and satellite clock error, satellite orbit error (ephemeris error), atmospheric refraction error, multipath error, loss of lock, and high dilution of precision value (poor geometric configuration of satellites, and/or number of visible satellites). DGPS is a method that increases the accuracy of coarse/acquisition (C/A) code measurements by canceling some of the inherent systematic errors.

The performance of the map-matching algorithm was assessed with artificially created lower quality spatial data by adding random noise to the original data collected by winter maintenance vehicles. These random errors were simulated assuming a normal distribution with zero mean and different standard deviations. The effects of any potentially remaining systematic errors were not modeled in this research.

Field experiments conducted during the execution of the Wisconsin Winter Maintenance Concept Vehicle Project in 2001 concluded that the absolute positional accuracy for DGPS data points was 2–5 m, RMS error (Vonderohe et al. 2001). Therefore, errors with a mean value of zero and SDs of ± 2 and ± 5 m were introduced in the DGPS data points using the Box and Muller method (Box and Muller 1958).

FN and solved case percentages were computed to compare the performance of the algorithm for original and perturbed DGPS data points. Fig. 6 presents changes in percentage of FN data points for original and perturbed data by county for a 12-m buffer before and after applying the algorithm. All FN percentages decrease after executing the algorithm, independent from spatial data quality. Average percentages of FN data points computed with original data present smaller values than data perturbed with 2- and 5-m error for both cases before and after applying the algorithm. For example, average FN percentages increase from 16 to 29% and from 22 to 48% for Portage and Polk County, respectively, when introducing 5-m error and after applying the algorithm.

Solved case percentages for original Columbia County data (almost 90%) were larger than those computed with perturbed

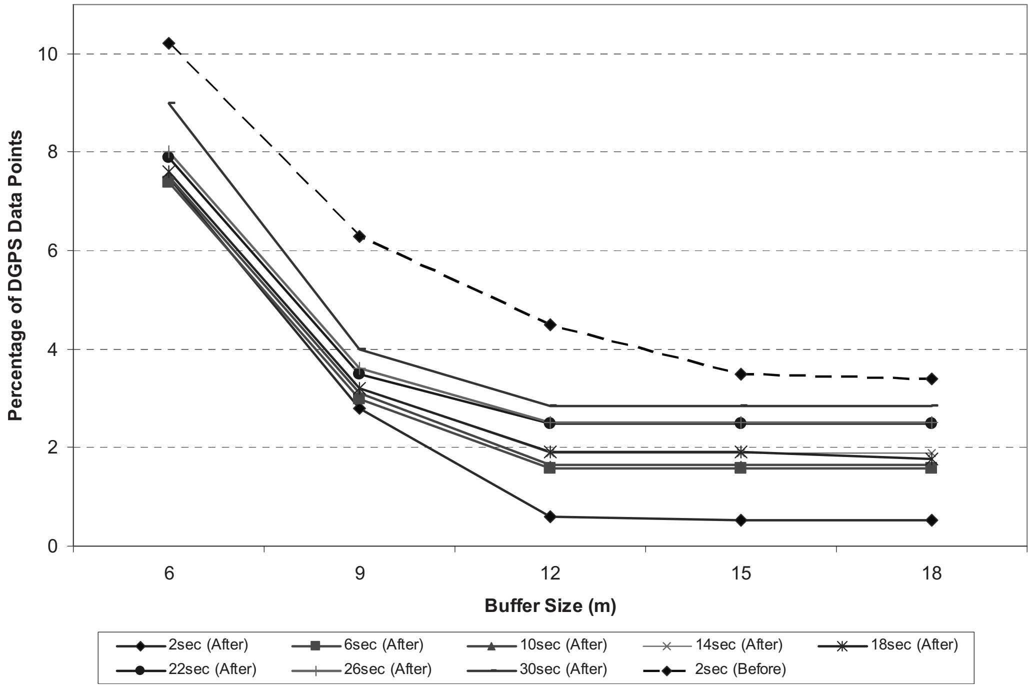

Fig. 7. Average FN percentages before and after applying algorithm for different temporal resolutions by buffer size for Columbia County data originally collected every 2 s

data. The average percentage of solved cases in Columbia County decreases from approximately 70 to 50% when simulated random error increases from 2 to 5 m, respectively. 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 both 2- and 5-m perturbed data.

Temporal Resolution

The outcome of map-matching is affected by not only spatial inaccuracies, but also data collection frequencies (Czerniak 2002; Yin and Wolfson 2004). As temporal resolution increases or data are collected more frequently, tracking of the vehicle becomes more accurate. However, temporal resolution impacts the sizes of the data sets, affecting processing time and storage requirements. Thus, it is important to examine the relationship between temporal resolution and algorithm performance.

Original data points were collected with sampling frequencies of 2 and 5 s for Columbia County and 10 s for Portage and Polk Counties. Lower temporal resolution data sets (i.e., from 2- to 30-s, 5- to 30-s, and 10- to 30-s sampling intervals for Columbia, Portage, and Polk County data, respectively) were generated from original data sets.

Fig. 7 presents FN curves before and after applying the algorithm for different temporal resolutions with data originally collected every 2 s in Columbia County. These curves show that as temporal resolution decreases the percentage of FN data points increases. These percentages decreased as the buffer size increased from 6 to 12 m and, thereafter, they stabilized. FN curves for the other counties with different temporal resolutions behaved similarly, showing larger FN percentage values and reaching zero-value slopes at larger buffer sizes.

Percentages of solved spatial ambiguities in Portage County decreased from approximately 67 to 40% with a 15-m buffer when increasing sample intervals from 10 to 30 s. Similarly, the percentage of solved cases for Columbia County data decreased in average from approximately 80 to 20% as sampling intervals increased from 5 to 30 s for all buffer sizes. The same behavior was apparent for solved case percentages in Polk County as data was collected more frequently.

Statistical Analysis

An ordinary least-squares (OLS) regression analysis was performed to explore the interactions between temporal resolutions and solved case percentages for different speed range tolerances using data aggregated by county at a 9-m buffer size. The model accounted for 70.5% in average ($SD=1.7\%$, $p<0.000\,5$) of the variance (see Table 1). As expected, all temporal resolution regression coefficients were negative and statistically significant ($p<0.000\,005$). Similar outcomes were obtained when using different roadway map scales and data aggregated by speed range tolerance. These results suggest that as temporal resolution increases the solution of spatial ambiguities also increases independent of speed range and roadway map quality.

Table 1. OLS Regression Results for Solved Case Percentages

Speed range tolerance	Variables	Coefficient	t-statistic
16 km/h	Constant	69.86	11.28
	Temporal resolution	−2.18	−6.54
32 km/h	Constant	75.41	11.18
	Temporal resolution	−2.46	−6.97
48 km/h	Constant	76.87	10.24
	Temporal resolution	−2.51	−6.41
64 km/h	Constant	76.26	10.08
	Temporal resolution	−2.50	−6.33

Table 2. Best and Worst Parameter Values for FN and Solved Case Percentages by County

County	Buffer size (m)		Speed range (km/h)		Number of consecutive data points	
	Best	Worst	Best	Worst	Best	Worst
FN percentages						
Columbia (2 s)	≥12	6	≥24	8	≥6	3
Columbia (5 s)	≥12	6	≥24	8	≥5	3
Portage (10 s)	≥18	6	≥40	8	≥4	3
Polk (10 s)	≥32	6	≥40	8	≥3	≥3
Solved case percentages						
Columbia (2 s)	9	≥12	≥32	8	≥6	3
Columbia (5 s)	9	≥12	≥32	8	≥6	3
Portage (10 s)	12	6	≥40	8	≥6	3
Polk (10 s)	15	6	≥40	8	≥3	≥3

Summary Results

In summary, Table 2 presents parameter values for best and worst results in average FN and solved percentages by county after applying the map-matching algorithm. This table indicates that larger buffer sizes, speed range values, and numbers of consecutive data points yield better results in minimizing FN data points and maximizing solved cases.

Percentages for FN and solved cases approach stable values as both speed range and number of consecutive data 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, increasing the number of consecutive data points solves a larger number of spatial ambiguities and diminishes FN percentages since additional combinations between pairs of snapped data points are tested. This improvement is not apparent for Polk County data due to the lower quality of its spatial database. In general, higher parameter values yield better results as data are collected less frequently and snapped to lower quality roadway maps.

Table 3 summarizes the performance of the map-matching algorithm for different temporal resolutions and DGPS data point positional accuracies. Original data yielded better results than perturbed data in terms of percentages of solved cases. The largest decrease in percentage of solved cases occurred for Portage County with a drop from 68% without perturbation to 14% when a 5-m positional error was introduced in the data.

Conclusions

Transportation applications are in need of efficient map-matching algorithms when employing AVL/DGPS technology to accurately associate vehicle locations and other collected data with correct roadway centerlines. Scarce literature exists regarding the impact of controlling or weighting parameter values in the performance of map-matching algorithms. This research analyzed the effects of three parameters (buffer size, speed range tolerance, and number of consecutive data points) controlled by the user, and two variables (DGPS positional error and temporal resolution) controlled externally through simulated data, on the performance of a post-processing decision-rule map-matching algorithm developed earlier by the writers. This algorithm was tested against intelligent winter maintenance vehicle data collected in Wisconsin and Iowa during the 2002–2003 winter season.

A methodology was presented to classify snapped and not-snapped data points before and after applying the map-matching algorithm. Average FN percentages and solved spatial ambiguities contained within this classification were computed for multiple parameter values to evaluate the performance of the map-matching algorithm. The algorithm produces improvements in the overall snapping process by minimizing the percentage of FN data points and maximizing solved cases.

Sensitivity analysis indicates that the performance of the map-matching algorithm depends on the positional error and temporal resolution of DGPS data points. 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.

By increasing the number of consecutive data points, a larger number of spatial ambiguities are solved, especially when alternative roadway centerlines are equally viable, and FN percentages are reduced since more combinations are tested between pairs of snapped DGPS data points. However, the performance of the map-matching algorithm is not affected significantly by the number of consecutive data points when spatial data accuracies are low.

Parameter values that maximize solved spatial mismatches are presented for three spatial database qualities and different temporal resolutions. For example, approximately 90% of incorrect snaps for Columbia County data collected every 2 s are solved

Table 3. Percentages of Solved Cases with and without Perturbation in the Data by County for Different Temporal Resolutions and Best Buffer Sizes

County, roadway centerline map scale	Temporal resolution (s)	Best buffer size (m)	% solved cases		
				With perturbation	
			Original data without perturbation	$\sigma = \pm 2\ m$	$\sigma = \pm 5\ m$
Columbia, 1:2,400	2	9	90	72	49
	5	12	80	51	42
Portage, 1:24,000	10	15	68	16	14
Polk, 1:100,000	10	24	50	38	25

with a 9-m buffer size after implementing the algorithm.

Regression coefficients for temporal resolution regression have high levels of statistical significance with respect to solved cases. Thus, diminishing sampling intervals increases the success in solving spatial mismatches independent of speed range values and roadway map qualities. Further research is necessary to analyze the effects of controlling parameters on the performance of this algorithm for data collected on other types of transportation applications (such as travel speed studies) with different spatial data qualities and accuracies.

Acknowledgments

The writers gratefully acknowledge the support of the Wisconsin Department of Transportation, Bureau of Highway Operations, and the Midwest Regional University Transportation Center at the University of Wisconsin-Madison for funding this research. This research was also funded by the Chilean National Fund for Scientific and Technological Development (FONDECYT Grant No. 1070386) and Universidad Andres Bello Project No. DI-23-06/R. The writers also thank the Columbia, Portage, and Polk County Highway Departments for providing the data used in this research.

References

Aron, J. (2006). "Barriers to use of geospatial data for adaptation to climate change and variability: Case studies in public health." *Geospatial Health*, 1, 11–16.

Blazquez, C. A., and Vonderohe, A. P. (2005). "Simple map-matching algorithm applied to intelligent winter maintenance vehicle data." *Transp. Res. Rec.*, 1935, 68–76.

Box, G., and Muller, M. (1958). "A note on the generation of random normal deviates." *Ann. Math. Stat.*, 29(2), 610–611.

Chawla, S., Shekhar, S., Wu, W., and Ozesmi, U. (2001). "Modeling spatial dependencies for mining geospatial data." *Proc., 1st SIAM Int. Conf. on Data Mining*, Chicago, Society for Industrial and Applied Mathematics, Philadelphia.

Chen, W., Li, Z., Yu, M., and Chen, Y. (2005). "Effects of sensor errors on the performance of map-matching." *J. Navig.*, 58, 273–282.

Czerniak, R. (2002). "Collecting, processing, and integrating GPS data into GIS." *NCHRP Synthesis of Highway Practice Rep. No. 301*, Transportation Research Board, National Research Council, Washington, D.C.

Finlayson, C., and Opitz, D. (2004). "A complete building extraction system for elevation data." *Proc., Int. User Conf.*, ESRI, San Diego.

Haga, T., Sumi, K., and Yagi, Y. (2004). "Human detection in outdoor scene using spatio-temporal motion analysis." *Proc., 17th Int. Conf. on Pattern Recognition (ICPR'04)*, Vol. 4, IEEE Computer Society, Los Alamitos, Calif.

Hofmann-Wellenhof, B., Lichtenegger, H., and Collins, J. (1993). *GPS theory and practice*, Springer, Vienna, Austria.

Jagadeesh, G., Srikanthan, T., and Zhang, X. (2004). "A map-matching method for GPS based real-time vehicle location." *J. Navig.*, 57, 429–440.

Kaplan, E., and Hegarty, C. (2006). *Understanding GPS: Principles and applications*, Artech House, Norwood, Mass.

Karimi, H., Conahan, T., and Roongpiboonsopit, D. (2006). "A methodology for predicting performances of map-matching algorithms, web, and wireless geographical information systems." *Proc., 6th Int. Symp. Lecture Notes in Computer Science*, Vol. 4295, Springer, Berlin, 202–213.

Li, J., Taylor, G., and Kidner, D. (2005). "Accuracy and reliability of map-matched GPS coordinates: The dependence on terrain model resolution and interpolation algorithm." *Comput. Geosci.*, 31, 241–251.

Noronha, V., and Goodchild, M. (2000). "Map accuracy and location expression in transportation—Reality and prospects." *Transp. Res., Part C: Emerg. Technol.*, 8, 53–69.

Quddus, M. A. (2006). "The effects of navigation sensors and digital map quality on the performance of map-matching algorithms." *Proc., 85th Annual Meeting* (CD-ROM), TRB, National Research Council, Washington, D.C.

Quddus, M. A., Ochieng, W. Y., and Noland, R. B. (2007). "Current map-matching algorithms for transport applications: State-of-the art and future research directions." *Transp. Res., Part C: Emerg. Technol.*, 15, 312–328.

Schlingelhof, M., Betaille, D., Bonnifait, P., and Demaseure, K. (2008). "Advanced positioning technologies for co-operative systems." *J. Intell. Transp. Syst.*, 2(2), 81–91.

Syed, S., and Cannon, M. E. (2004). "Fuzzy logic based-map-matching algorithm for vehicle navigation system in urban canyons." *Proc., Institute of Navigation National Technical Meeting*, The Institute of Navigation, Manassa, Va.

Taylor, G., Blewitt, G., Steup, D., Corbett, S., and Car, A. (2001). "Road reduction filtering for GPS-GIS navigation." *Transactions in GIS*, 5(3), 193–207.

Taylor, G., Brunsdon, C., Li, J., Olden, A., Steup, D., and Winter, M. (2006). "GPS accuracy estimation using map-matching techniques: Applied to vehicle positioning and odometer calibration." *Comput. Environ. Urban Syst.*, 30, 757–772.

Vonderohe, A. (1999). "Adaptation of geographic information systems for transportation." *National Cooperative Highway Research Program Rep. No. 20-27*, TRB, National Research Council, Washington, D.C.

Vonderohe, A., Blazquez, C., Brezovar, N., Hong, S., Lauters, J., and Adams, T. (2006). "GIS tool to measure performance of winter highway operations." *Final Rep. Prepared for Project 04-01*, Wisconsin Dept. of Transportation, Madison, Wis.

Vonderohe, A., Malhotra, A., Sheth, V., Mezera, D., and Adams, A. (2001). "Report Wisconsin winter maintenance concept vehicle: Data management year 1." *Rep. Prepared for Dept. of Civil and Environmental Engineering Rep. No. 04–01*, Univ. of Wisconsin-Madison, Madison, Wis.

Wenk, C., Salas, R., and Pfoser, D. (2006). "Addressing the need for map-matching speed: Localizing global curve-matching algorithms." *Proc., 19th Int. Conf. on Scientific and Statistical Database Management*, Vienna, Austria, Springer, New York.

Wolf, J., Hallmark, S., Oliveira, M., and Guensler, R., and Sarasua, W. (1999). "Accuracy issues with route choice data collection by using global positioning system." *Transportation Research Record. 1660*, Transportation Research Board, Washington, D.C., 66–74.

Wolf, P., and Ghilani, C. (2006). *Adjustment computations: Spatial data analysis*, Wiley, Hoboken, New Jersey.

Yin, H., and Wolfson, O. (2004). "A weight-based map-matching method in moving objects databases." *Proc., 16th Int. Conf. on Scientific and Statistical Database Management*, Santorini, Greece, Springer, New York, 437–438.

Zhang, J.-H., Chung, T., and Oldenburg, K. (2000). "Confirmation of primary active substances from high throughput screening of chemical and biological populations: A statistical approach and practical considerations." *J. Comb. Chem.*, 2, 258–265.