The Journal of Transportation of the Institute of Transportation Engineers is circulated electronically twice yearly by the Institute of Transportation Engineers, 1627 Eye Street, NW, Suite 600, Washington, DC 20006 USA, phone: 202-785-0060. ©2012 Institute of Transportation Engineers. All rights reserved except for brief quotation with attribution. The views expressed in the articles are those of the authors and do not reflect official ITE policy unless so stated. ISSN 2162 1616. Journal of Transportation of ITE is distributed free to ITE members and is $25 per issue for nonmembers.
CONTENTS

- Predicting Multimodal Roundabout Operations by Combining Deterministic and Micro-Simulation Models
  By Michael A. Gingrich Sr. and Francois Dion  

- Using GPS to Locate No-Passing Zones on Straight Sections of Two-Lane Highways
  By Mehdi Azimi, Cameron L. Williams, and H. Gene Hawkins Jr.  

- Comparison of the Retroreflective Durability Between Rumble Stripes and Painted Lines
  By Stuart R. Mitkey, Thomas M. Brennan Jr., Stephen M. Remias, Alexander D. Davis, Gannon Grimmer, Alexander M. Hainen, Randy Morris, Paul Michael, and Darcy M. Bullock  

- Safety Impacts of Signing Delineation for Horizontal Curves on Rural Two-Lane Roads
  By Raghavan Srinivasan, Bhagwant Persaud, Kimberly Eccles, Daniel Carter, and Jongdae Baek
Predicting Multimodal Roundabout Operations by Combining Deterministic and Micro-Simulation Models

By Michael A. Gingrich Sr. and Francois Dion, Ph.D, P.E.

This article outlines a method to predict the operations of a high-capacity, multilane roundabout when light rail (with station), multiple bus stops, and heavy pedestrian traffic are added to form a truly multimodal environment. So far, no single model or micro-simulation software has proven effective in producing validated predictions in this type of multimodal environment. This article explores how a combination of existing models, in conjunction with micro-simulation, could be used to predict operations in such an environment. The article also examines calibration between existing models and micro-simulation, including the capabilities/limitations of various models and micro-simulation software in the context of recent North American capacity experiences.

Introduction

As more roundabout projects are included in multimodal transportation initiatives, the need increases for analysis beyond the limitation of deterministic models. For multimodal analyses, micro-simulation software is often the preferred alternative; however, over the course of a project, other models may also be applied. For example, many multimodal roundabout projects use both a deterministic capacity model and a micro-simulation model to evaluate their operations. Typically, the deterministic model is used during the planning and initial feasibility assessment, while the micro-simulation model is used in later project development stages and in public presentations. Results from these two models can vary greatly; they often produce different required lane configurations and predicted operations, which leaves review jurisdictions wondering which is correct. Discrepancies in capacity predictions most often occur in the middle of a project—after public presentations of services based on the outputs of deterministic models—which makes it difficult to justify late changes in project evaluations.
This article investigates how a multimodal project using both deterministic and stochastic micro-simulation models can minimize discrepancies in capacity prediction among different categories of models. To understand the challenges and issues involved in correlating the two models, we focused on differences in modeling approaches and model applications, situations in which micro-simulation models may prove beneficial, and variations in capacity prediction and observed capacity in the field. In conclusion, we describe a method for calibrating and correlating simulation results produced by deterministic and micro-simulation models.

**Operational Analysis Model Types**

Two model types are commonly used during a roundabout operational analysis to determine capacity and predictions for levels of service:

- **Macroscopic (deterministic) models.** These models (also referred to as static or steady state models) are based on mathematical formulas in which outcomes are precisely determined through known relationships between states and events; in such models, a given input will always produce the same capacity output. The two most commonly used approaches are based on (1) gap-based analysis and (2) empirical formulas. Most of these models use mathematical models that have been developed to replicate observed traffic behavior in real roundabouts. Rodel, Arcady, and Sidra are three of the macroscopic models used for roundabout evaluation.

- **Microscopic (micro-simulation) models.** These models attempt to replicate the movements of an individual vehicle, using mathematical models to capture the interactions among roadway elements, vehicles, and drivers. Each vehicle is moved through a transportation network or specific facility on a second-by-second basis according to the physical characteristics of the vehicle (e.g., length, maximum acceleration rate); the fundamental rules of motion (e.g., acceleration multiplied by time equals velocity, velocity multiplied by time equals distance); and rules of driver behavior (e.g., car-following rules, lane-changing rules). These models are typically stochastic in nature and can capture highly dynamic situations. Outcomes are affected by any change in the initial conditions and by what may happen during a simulation run. Vissim, Paramics, and Aimsun are three of the microscopic models.

Both macroscopic and microscopic simulation models have benefits and limitations. No model is perfect. The determination of which model to use to predict the operation of a roundabout in a specific project is often the result of a combination of factors, such as the analysts’ level of knowledge and comfort with a specific model, the reviewing jurisdiction’s instruction to use a specific model, available data, modes of transportation to be included in the analysis, and anticipated effects of a roundabout in a given roadway network. Table 1 provides general guidance on the use of the two models.

Table 1 shows that the models overlap into several applications. For example, it is becoming increasingly common for a review jurisdiction to perform or request both a deterministic and a micro-simulation analysis in the course of a roundabout development project to ensure a higher level of confidence in the predicted performance and operations. Sometimes micro-simulation models are used specifically to provide visual animation for public meetings.

**Situations in Which Micro-simulation Models May Be More Effective**

Deterministic models require less time and effort to calibrate the road network surrounding a roundabout, and they are not affected by random variations in simulation outputs. However, they are limited in their ability to analyze complex multimodal networks and situations. In the following situations, a micro-simulation model may provide benefits not readily achievable with deterministic models:
Table 1. Typical application of model types.

<table>
<thead>
<tr>
<th>Application</th>
<th>Desired Outcome</th>
<th>Input Data Available</th>
<th>Model Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning-level sizing</td>
<td>Number of lanes</td>
<td>Traffic volumes</td>
<td>Standard Calibration Methods section of this paper, HCM, deterministic software</td>
</tr>
<tr>
<td>Preliminary design of roundabouts with up to two lanes</td>
<td>Detailed lane configuration</td>
<td>Traffic volumes, geometry</td>
<td>HCM, deterministic software</td>
</tr>
<tr>
<td>Preliminary design of roundabouts with three lanes and/or short lanes/flared designs</td>
<td>Detailed lane configuration</td>
<td>Traffic volumes, geometry</td>
<td>Deterministic software</td>
</tr>
<tr>
<td>Analysis of pedestrian treatments</td>
<td>Vehicular delay, vehicular queuing, pedestrian delay</td>
<td>Vehicular traffic and pedestrian volumes, crosswalk design</td>
<td>HCM, deterministic software, micro-simulation</td>
</tr>
<tr>
<td>System analysis</td>
<td>Travel time, delays and queues between intersections</td>
<td>Traffic volumes, geometry</td>
<td>HCM, micro-simulation</td>
</tr>
<tr>
<td>Public involvement</td>
<td>Animation of no-build conditions and proposed alternatives</td>
<td>Traffic volumes, geometry</td>
<td>Micro-simulation</td>
</tr>
</tbody>
</table>

SOURCE: Adapted from Exhibit 4-4, NCHRP Report 672.

- Weaving analysis;
- Merging analysis;
- Effect of railways on capacity and operations;
- Evaluation of various railway gating scenarios;
- Effect of pedestrian crossings;
- Effect of bicycle traffic;
- Effect of bus loading/unloading zones;
- Signalized pedestrian crossings;
- Corridors of roundabouts;
- Interaction of traffic signals and roundabouts in close proximity with each other;
- Analysis of platooned pedestrian arrivals (e.g., bus loading/unloading zones, light rail stations, college campus);
- Planning and testing of event traffic schemes;
- Effect of access drives near the roundabout;
- Public presentations and visual educational materials; and
- Effects on operations due to access drives near a roundabout.
Variations in Capacity Predictions and Observed Capacities

As shown in Table 1, both macroscopic and microscopic models can be applied to estimate roundabout capacity, with each type offering certain benefits and limitations. Simulation results from different models can vary greatly; the choice of a model often depends on the application and stage of the project analysis/development process as well as professional preferences. In many cases, the analysis should also consider whether capacity in North America will grow to equal capacity observed elsewhere in the world.

Discrepancies in Capacity Estimations Between Model Types

Over the course of a project, it is not uncommon for both deterministic macroscopic and stochastic microscopic simulation models to be used for analysis. In fact, this has become commonplace for multimodal applications and complex networks. However, the use of both model types to evaluate a project often results in discrepancies in predicted operations, especially in delay and queue predictions. Practitioners and reviewing agencies then must decide which results are most probable and better reflect potential operations. Without knowing which of the simulation results is more accurate, decision-makers might rely on the lesser results to generate more conservative performance estimations. However, this approach might needlessly eliminate a beneficial alternative. In other cases, they might average the predictions from both models. Unfortunately, this approach can mean a false mixture of realistic and dramatically overestimated delay and queueing predictions. In many cases, discrepancies between the outcomes of macroscopic and microscopic models become increasingly apparent as the vehicle/capacity ratio of the entries become saturated.

Without a method to produce a realistic link between the two models, analysts might decide that model differences cannot be reconciled by changing model parameters. Consequently, different models are frequently not calibrated to reflect similar results. The differential outcomes can leave a review jurisdiction unsure of which prediction is correct and can undermine its confidence in both types of models, the practitioners, and the possibility of evaluating roundabout projects at all.

Figure 1 illustrates the differences that can occur between deterministic macroscopic and stochastic microscopic simulation models. It shows simulation outcomes for the evaluation of a planned multilane roundabout in an urban multimodal setting. The roundabout is to be located on a main arterial in the middle of the University of Michigan campus. The existing intersection serves many

Figure 1. Estimated average delay for 2035 design year for Fuller Road/East Medical Center Drive/Maiden Lane proposed roundabout in Ann Arbor, MI, USA.
users; it includes heavy pedestrian traffic, bus traffic, and many bicyclists. Because of the complexity of the operations, the review jurisdiction requested a simulation model. As the figure shows, significant variations in predicted delays were obtained for all approaches, with particularly wide variations for the westbound approach. The differences are especially significant in light of the tremendous efforts that have been made to calibrate the simulation models.

Discrepancies in Capacity Estimations Within Micro-Simulation Models

In addition to outcome differences between macroscopic and microscopic simulation models, differences exist among the various micro-simulation software programs. The University of Maryland conducted a study to compare the performance of Paramics, Vissim, and Aimsun—three of the most commonly used micro-simulation programs. The study analyzed and documented the queue and travel times for a single-lane roundabout. Figures 2 and 3 show some of the results for predicted queue and travel times. Note the wide variation produced by the different programs. These differ-

Figure 2. Comparison of queue estimates produced by leading microscopic simulation models.

![Figure 2](https://example.com/figure2)

Figure 3. Comparison of travel time estimates produced by leading microscopic simulation models.

![Figure 3](https://example.com/figure3)
ences are primarily the result of differences in the specific behavioral models used by each simulation software program. This situation can make it difficult for practitioners and review agencies to select software to use in roundabout analyses.

**Variations in Observed North American Roundabout Capacity**

Over the past decade, lengthy discussions have been held on roundabout capacity in North America. Much of the debate focuses on whether North American capacities can or will reach the level of those in international models. Research results for National Cooperative Highway Research Program (NCHRP) Report 672 concluded that current observed North American roundabout capacities (based on data collected in 2003) can be up to approximately 20 percent lower than those predicted by some of the international models, such as those of the United Kingdom and Australia. Several roundabout specialists have argued that North American capacities can be expected to increase over time owing to the following effects:

- North American drivers will become more familiar with roundabouts and gain experience driving through them;
- Roundabout design practice for North America continues to improve, resulting in more effective geometric designs; and
- Roundabout projects typically require a 20-year design period, so there should be ample time for drivers to adjust to the intricacies of driving through a roundabout before it needs to function at capacity. This anticipated increase in capacity was also mentioned in the recent NCHRP Report 672.²

---

**Figure 4. Change in observed roundabout capacity over time for the city of Bend, OR.**

---
Observed Increase in North American Roundabout Capacity

One explanation of the capacity differences between North American roundabout operations and predictions by international models is that drivers are still relatively unfamiliar with the concept. As they gain experience, capacities should increase. Increases are already being noted throughout North America. For instance, operational data collected at a roundabout in Bend, OR, USA,\(^5\) showed a substantial capacity increase over time. The first roundabout in Bend was constructed in 1999; since then, 27 more have been built. As drivers were exposed to an increasing number of roundabouts over almost a decade, their expertise in driving through them increased (see Figure 4).

Standard Calibration Methods, Variation, and Available Data

The two primary methods for calibrating micro-simulation models are—

1. Using capacity data collected from the same geographic location with similar saturation characteristics. For roundabouts, this is usually observed gap and headway data.

2. Visual calibration via parameter adjustments in the micro-simulation software. This method usually involves a technician adjusting software parameters to a point just before a collision occurs in the simulation.

A significant concern with regard to the use of micro-simulation models is calibrating a model to local conditions. This can be very difficult given the relatively few roundabouts in the United States and, thus, the lack of sufficient data.\(^6\) And even fewer U.S. roundabouts operate at capacity, which means very few opportunities to collect capacity data. To compensate for the lack of data, practitioners sometimes attempt to calibrate micro-simulation models by collecting data from one roundabout site for one day. This approach provides some general guidelines on driver behavior but does not usually provide enough statistically valid data to serve as a basis for calibration for a specific roundabout.

Visual calibration seeks to model the operation of a roundabout on the basis of observed differences. This approach can result in notable variety in capacity predictions from practitioner to practitioner—it is essentially a trial-and-error approach.\(^7\) Acting on their individual experiences, practitioners may attempt to seek a solution to observed differences by adjusting parameters. This may result in models producing somewhat similar results in considering some operational parameters but significantly different behavior in considering other parameters.

Variation in capacity results can be compounded by the use of different random seeds in stochastic microscopic models. A different random seed is commonly used with multiple simulation runs to minimize the stochastic nature of a microscopic model. However, using different random seeds can pose a challenge in fine-tuning a visual calibration. For example, a technician might adjust the software parameters to a point of maximum capacity before a collision occurs. Changing the random seed can cause changes in the sequence in which vehicles are entering the network, creating a new traffic dynamic that may result in different conflicts among vehicles and different operational performance. It is common practice to run 10 simulations with different random seeds, so the technician would have to visually calibrate all 10 runs to verify that the model is collision-free. This is not only time-consuming but often results in delay prediction results that deviate too much from their real value.

Although many micro-simulation software programs can model roundabouts, none has an internal roundabout-specific model, so practitioners have to replicate observed roundabout behavior by attempting to manipulate the driver behavioral parameters offered. Also, these programs do not have the capabilities to derive or directly correlate observed roundabout capacities or data. All these issues contribute to the subjective nature of micro-simulation models and highlight the need for a point of correlation between micro-simulation models and observed capacity models.
Finding Common Ground Among Models and Capacity Estimates

Multimodal roundabout projects often start with an initial operational analysis using a deterministic macroscopic model. As project development continues, a micro-simulation model is introduced to analyze the multimodal components. Correlating mixed capacity predictions from various models can be challenging, and the results are difficult to validate. We need a method of producing relatively similar capacity predictions throughout the course of a roundabout project. To achieve this, we need a process that can correlate the capacity predictions among the deterministic and micro-simulation models.

The following process is suggested to correlate the outputs of macroscopic and microscopic models and to allow consistency among capacity predictions from the different types of models. This process finds common ground among the models by addressing the variations associated with the following elements:

1. Discrepancies in capacity estimations among model types;
2. Discrepancies in capacity estimations within micro-simulation models;
3. Variations in observed North American roundabout capacities (need for calibration); and
4. Standard calibration methods and available data.

Method for Correlating Deterministic and Micro-Simulation Models

The following is a simplified description of the steps required to apply an adjusted Y-intercept to the capacity slope of a deterministic model, allowing for consistency of capacity prediction and incorporation into micro-simulation models.

1. As with the determination of predicted volumes for a design year, it is possible to determine a target capacity slope for a proposed design year. The information to consider in determining a target capacity slope should include—
   - Anticipated/projected growth rates for the design year;
   - Whether the roundabout location will be in an urban or rural setting;
   - The number of roundabouts anticipated to be built in the surrounding area by the design year; and
   - Available observed roundabout capacity data for the surrounding area.

The adjustment (calibration) is usually referenced as a percentage +/- the default capacity predictions produced by a particular model. For instance, the adjustment may consider a 5 percent increase in capacity from the predictions produced by the models described in NCHRP Report 672 or a 10 percent reduction from the empirical capacity predictions produced by models from the United Kingdom. These adjustments can be applied by moving the Y-axis of the slope on the capacity graph. Using a deterministic model, adjust the Y-intercept to the desired percentage adjustment. The adjusted (calibrated) Y-intercept slope will serve as the target (reference point) for capacity prediction for the rest of the project. Figure 5 illustrates the Y-intercept adjustment concept.

2. Using a micro-simulation software package, set up a roundabout simulation model for vehicular traffic only.

3. Run the micro-simulation model and extract vehicular delay and queuing results.

4. Compare these results with the delay results predicted by the calibrated deterministic macroscopic model. If the results are not similar, adjust the appropriate parameters in the micro-simulation model, then rerun the micro-simulation model and compare the results again. Continue this process until the results from the two models are in close proximity.
5. Once the micro-simulation model has been calibrated to reasonably match the deterministic model, other multimodal and network components can be added.

6. Rerun the micro-simulation model to extract results for multimodal operations.

Figure 5. Y-intercept adjustment.

Figure 6. Process to calibrate micro-simulation and deterministic models for multimodal application.
Although this method is not flawless, it can be used until the roundabout component of micro-simulation models is more fully developed. In particular, although the multimodal components added to a micro-simulation model in Step 5 might still produce some deviation, the results should be closer than they would be if no calibration process were used.

**Methodology Application Example**

In this section, we show how the suggested calibration method can be applied to a multimodal setting and how trends for project capacity might develop throughout the process. The figures below are in the form of an academic exercise, as the final capacity results were not available when this article was completed. However, the graphs reflect preliminary trends and observations, and illustrate the calibration method described in the article.

For this example, we will assume that a multilane roundabout is proposed to be built in an urban area. At the selected site, anticipated growth is expected to produce a steady increase in traffic volume of approximately 2 percent per year. The review jurisdiction has already constructed several roundabouts and is actively promoting the installation of more. Several bus routes with nearby stops are expected to use the roundabout, and a proposed light-rail transit service will run through the center of the roundabout, with a station on the edge. These transit services will create heavy pedestrian traffic, as transit riders walk around the roundabout to transfer between the bus stops and the light-rail station. Figure 7 shows the general site characteristics, geometric layout, light-rail station, and bus stop locations. The jurisdiction has ordered a micro-simulation performed to evaluate the effects of the multimodal components of the project; the jurisdiction is concerned about how the multimodal components will affect operations and wants to see a correlation of the deterministic and micro-simulation results.

The first step is to identify the Y-intercept that will be used as the capacity reference point throughout the project. The Y-intercept may be the model’s default or it may be calibrated, depending on which simulation model is used and the available site characteristics. In this example, we will assume that the UK model is used and the Y-intercept is calibrated to produce a 10 percent reduction in capacity predictions from the default UK model estimations. The Y-intercept will serve as the target to calibrate the micro-simulation model.

Figure 8 shows the micro-simulation model with vehicular traffic only. This scenario is run to extract vehicle delay and queuing for comparison, using the deterministic model predictions. Graphs 1 and 2 illustrate the delay and queuing results for the initial comparison. This allows for comparison of the results of the deterministic model and the initial simulation. Graphs 3 and 4 illustrate the delay and queuing after the micro-simulation model has been calibrated to the deterministic model. Several iterations may be required before the results of the two models converge. Figures 9–12 show the micro-simulation with the addition of the multimodal components. Graphs 5 and 6 illustrate the final results, showing the effect of the multimodal components on operations. This process allows for the isolated assessment of the effects of each component. The example shows how the two model types can be calibrated to produce similar results before multimodal components are added to the simulation. This approach provides a correlation of predicted capacity between the two model types, with a common
Figure 7. Site characteristics.

Figure 8. Micro-simulation, vehicular traffic only.

Figure 9. Multimodal components added (view 1).

Figure 10. Multimodal components, pedestrians.

Figure 11. Multimodal components, light rail.

Figure 12. Multimodal components, multiple bus routes.

NOTE: Full analysis results would include delay and queuing for SB right-turn bypass lane.

Images courtesy of Mark Lenters and Ourston Roundabout Engineering.
Graph 1. Initial comparison of vehicle traffic only.

Graph 2. Initial comparison of vehicle traffic only.

Graph 3. Vehicle traffic calibrated to match (similar).

Graph 4. Vehicle traffic calibrated to match (similar).

Graph 5. Result with multimodal components.

Graph 6. Result with multimodal components.

NOTE: Full analysis results would include delay and queuing for SB right-turn bypass lane.
reference point for capacity predictions. Subsequently, differences in predicted operation resulting from the multimodal elements can be observed as each element is added.

Conclusion

A demonstrated need exists for relatively consistent capacity predictions throughout the course of roundabout projects, especially between deterministic macroscopic and stochastic microscopic simulation models. This article has described a process to fill this need, at least as an interim solution until micro-simulation models are developed that can provide more direct correlations with observed capacity models. Consistent predictions between the two types of models will provide results that will allow decision-makers to have more confidence in their intersection control choices. As micro-simulation software continues to improve its ability to model roundabouts, the authors anticipate the development of more efficient and accurate ways to calibrate the roundabout to a specific Y-intercept.

The method presented here can also provide a correlation between the two models before multimodal components are introduced into the micro-simulation, thus isolating their effects on operations and predicted capacity. Existing deterministic models cannot do this.

The method is not flawless. It does not eliminate all discrepancies in capacity predictions between macroscopic and microscopic models—especially when multimodal components are added—because these discrepancies are inherent to the underlying behavioral models. However, the method should dramatically reduce overall discrepancies. More research and case studies are needed to determine to what extent this process minimizes the discrepancies between the two model types.

References


3. URS. Fuller Road/East Medical Center Drive/Maiden Lane Roundabout Simulation Summary. Prepared for the city of Ann Arbor, MI, USA, January 2010.


Michael A. Gingrich Sr. has been practicing traffic engineering for more than 15 years. His roundabout-related work to date includes approximately eighty-five projects. Mr. Gingrich has traveled extensively in England to study roundabout designs with international roundabout experts R. Barry Crown and Clive Sawers, and to consult with Transportation Research Laboratory senior researchers. Previously he managed and designed more than 250 traffic signal and intersection geometric design projects. Mr. Gingrich has been a guest speaker at numerous seminars and conferences including American Society of Civil Engineers, Institute of Transportation Engineers, and Transportation Research Board. Mr. Gingrich has taught analysis and design classes throughout the United States and Canada including classes on behalf of Northwestern University and the University of Wisconsin, Madison. His current focus includes all forms of alternative intersection control.

Francois Dion, Ph.D., P.E., is a senior development engineer with Partners for Advanced Transportation Technologies (PATH) at the University of California, Berkeley. Dr. Dion currently manages various projects for the California Department of Transportation, focusing on the integrated management of multimodal transportation corridors. His research interests include connected vehicle applications, transportation system modeling and simulation, traveler information systems, and traffic management systems. Prior to joining the University of California, Dr. Dion worked for the University of Michigan Transportation Research Institute, Michigan State University, and the Virginia Tech Transportation Institute. He received his Ph.D. in civil engineering from the University of Waterloo, in Canada, in 1999. He is a member of ITE.
This article describes a method of using global positioning system (GPS) coordinates to identify the location of no-passing zones in vertical curves. It details a new application of a mathematical technique to smooth GPS data and obtain a geometric model of the roadway, then describes an analytical algorithm for analyzing the availability of sight distance along the vertical profiles of two-lane highways. The algorithm is incorporated into a computer model that can use GPS data as the input and produce a more efficient, accurate, and safe method of locating no-passing zones compared with the current field measurement methods, which place crews on highways in the presence of moving traffic. The automated system processes GPS coordinates and converts them into easting and northing values, smooths GPS data, and evaluates roadway profiles for possible sight restrictions, which indicate where no-passing zones should be located according to the vertical alignment. In a comparison of these results with existing pavement markings, the model shows potential for evaluating available sight distance and locating no-passing zones.

Introduction

Two-lane highways are an important element of the transportation system, and they constitute a large percentage of total highway mileage. Over 62 percent of the 80,000 centerline highway miles on the Texas Department of Transportation (TxDOT) roadway system are rural two-lane highways. A unique feature of two-lane highways is that a faster moving vehicle must cross into the oncoming land to pass a slower moving vehicle (where adequate sight distance exists and no vehicles are oncoming). Pavement markings (solid centerlines) and traffic signs identify no-passing zones, in which driving to the left of the centerline is prohibited. In no-passing zones, the sight distance is less than the minimum passing sight distance specified in the Federal Highway Administration's
Identifying segments of highway that do not have adequate passing sight distance can be a time-consuming task and can be hazardous for those who work on the highway in the presence of moving traffic. Various methods exist for measuring passing sight distance and determining the location of no-passing zones. Most of the methods require work crews to physically evaluate sight distances; the weaknesses of these methods include the amount of time required, accuracy, and safety issues.

Global positioning system (GPS) coordinates provide an alternative means of evaluating sight distance and create the potential for an automated system to locate no-passing zones faster, more safely, and more accurately than possible using current methods. Implementation of such a system does not require detailed information about highway horizontal and vertical alignments and cross-sections. Previous researchers have proposed GPS-based approaches and developed finite element models for calculating sight distance. For example, Namala and Rys developed a model for measuring sight distance on two-lane highways using GPS and based on AASHTO’s standards, using projection on two-dimensional (2-D) planes. The model calculated the available passing sight distance (PSD) and the available stopping sight distance (SSD). The passing sight distance was calculated as a lower bound of PSD and SSD. In another work, Nehate and Rys presented a methodology to calculate SSD using projections on 2-D planes for vertical and horizontal alignments. Hassan, Easa, and Abd El Halim used a finite element method to model separate, complex alignments in 2-D and combined alignments in 3-D. They also studied an application of the models. Yan, Radwan, Zhang, and Parker applied a finite element model and evaluated the dynamic passing sight distance.

Calculating sight distance was the main purpose—or at least a part—of all of these studies. However, in some the researchers considered only special cases, while in others the approaches/models required extensive information about highway alignment, cross-section, and sight obstruction. In most cases (particularly for old roadways), this kind of information is not available or up to date. In this article, we describe a new application of a mathematical technique for smoothing GPS data to obtain a geometric model of the roadway. We also describe an algorithm that uses the geometric model to locate no-passing zones based on the vertical alignment of the roadway. The three main objectives of the article are (1) to determine a process to smooth GPS data and geometrically model roadway surface; (2) to create an algorithm for locating no-passing zones according to the vertical alignments; and (3) to calculate the no-passing zones from the modeled roadway surfaces (vertical alignments) and compare the results with existing no-passing zone pavement markings.

Background

Global Positioning System

The global positioning system is a satellite-based radio-navigation system that provides continuous (24-hour) reliable location information wherever there is an unobstructed line of sight to four or more GPS satellites. The satellites are at known locations at all times and transmit two L-band carrier signals (L1 and L2). A GPS receiver analyzes the coarse acquisition code (C/A) broadcast over the carrier signals and measures the time the signal was sent from the satellite and received by the receiver. The time is multiplied by the signal speed to determine a range (distance). Using ranging code from four satellites, the receiver can calculate its own position in three-dimensional space. Thus, GPS provides spatial coordinate triplets of longitude, latitude, and altitude for every position on earth. High-end GPS receivers, compared with autonomous receivers, reduce GPS errors and provide more accurate and reliable readings by using a differential signal broadcast from either known locations (reference stations) on earth or other sources (commercial satellite networks). In
this research, differential GPS survey methods were adapted to collect data to model the geometries of the roadway alignments using a high-end GPS receiver.

**GPS Applications in Highway Engineering**

GPS has been used extensively in research projects related to transportation engineering. In highway engineering, some researchers have applied GPS technology to their studies. Baffour combined GPS technology with kinematic vehicle operations to simultaneously collect roadway alignment, grade, and cross-slope data. In another research effort, Baffour investigated the use of a multi-antenna, single-receiver configuration of GPS to determine roadway cross-slope. Roh, Seo, and Lee determined road alignments based on collected GPS data using an RTK DGPS/GLONASS combination and compared their positioning accuracy with the values in the design drawings. Young and Miller developed methods to process more than 11 million GPS data points collected by the Kansas Department of Transportation, resulting in a geometric model of the state highway system. They showed that the spatial error from successive GPS data is highly correlated. Even though GPS error is widely acknowledged to be in the range of 1 to 5 m, the relative accuracy of sequential GPS data is much greater. If two successive GPS data points use the same constellation of satellites, the relative error between the two points is minimal. Assuming absolute errors of 2 m and 5 m, respectively, for horizontal and vertical axes, the relative error between successive readings is easily sub-meter in both dimensions.

**Geometric Modeling of Roadway**

Developing an automated method to locate no-passing zones requires the geometric definition of the roadway alignments. Roadway profiles, especially long segments, consist of multiple combinations of tangents and parabolic curves. In highway engineering, it is not possible to define a global function to fit the roadway data profile. In most cases, it is necessary to smooth the observation points before fitting the data, because roadway surfaces are continuous and the change in grade over a few feet is usually small. Measurement errors in the data are smoothed out so that the resulting curve is more accurate and smooth. Various mathematical curve-fitting and data-smoothing techniques have been used in several studies to define the best presentation of roadways based on the observation of data points. Easa, Hassan, and Karim developed an analytical method to establish the highway vertical alignment using profile field data. The method first determines the points separating straight and curved segments, then fits the straight lines (using linear regression) and clamped cubic spline curves to the respective segments. Easa also developed a linear optimization model to determine the best vertical alignment (parabolic curves and tangents) from field data to fit a given profile. Makanae examined an application of parametric curves used as spatial curves to highway alignment. The study showed that cubic and quadratic B-spline curves are suitable for application to highway alignment with a relatively large and small radius of curvature, respectively. Ben-Arieih, Chang, Rys, and Zhang applied piecewise parametric equations to GPS data and generated a B-spline approximation of the highway. Easa and Wang used the least squares method to develop an optimization model for estimating the parameters of vertical alignments. Castro, Iglesias, Rodríguez-Solano, and Sánchez used a parametric cubic spline to smooth GPS data to define the highway alignments. Young used a combination of techniques to create a three-dimensional geometric roadway model using historic GPS data.
In this article, locally estimated scatterplot smoothing (LOESS) is applied as a new application to smooth and fit the GPS data to obtain a geometric model of a highway. The advantage of LOESS over other methods is that it does not require a defined function to be fitted to all the observation data points. This method works well for the approach described in this article, because the sight distance algorithm does not require that the geometric model be defined by a function or equation.

**Data Collection and Experimental Work**

Data collection and experimental work included four main steps. After selecting the data collection sites, the first step was to collect the GPS data. Second, the appropriate transformation for GPS data was selected. Then a smoothing technique to fit the data and obtain a geometric model of the highway was selected. In the fourth step, a preprocessing program was developed to process the raw data, convert longitudes and latitudes to easting and northing values, smooth the data, and generate the input files for the no-passing zone computer model.

**Site Selection and GPS Data Collection**

A differential GPS (DGPS) survey method was adopted in this research to improve the accuracy of data collection. An instrumented vehicle equipped with a Trimble DSM 232 DGPS was used for data collection. This GPS receiver uses commercial satellite differential correction services provided by OmniSTAR and radio beacons. According to the DSM 232 specifications sheet, the entry-level model delivers sub-meter horizontal accuracy (typically better than 3 ft.). The collection rate capability of the unit is 10 hertz, which means the unit can record 10 separate data points every second. The instrumented vehicle has several other tools that were beneficial to the research. In addition to the GPS, the vehicle was equipped with a Dewetron computer, a Nu-Metrics Nitestar NS-60 distance measuring instrument (DMI), and video collection capabilities. These tools allow for the collection of not only GPS data but also video data of the roadway with DMI readings superimposed on the screen.

No-passing zones caused by sight restrictions in vertical profiles were the focus of this research; thus, testing sites had to have straight alignments and enough elevation change in the vertical profile to require no-passing zones. Roadway segments of five different two-lane highways in Texas were chosen for testing (Table 1). The test locations provided several different roadway lengths and a variety of vertical curves. The horizontal alignments of segments 1, 2, and 3 are straight and there are no horizontal curves. Except for a few isolated horizontal curves, segments 4 and 5 are also straight segments.

**Table 1. Data collection sites.**

<table>
<thead>
<tr>
<th>Name</th>
<th>Segment Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment 1</td>
<td>~ 8 miles</td>
</tr>
<tr>
<td>Segment 2</td>
<td>~ 4 miles</td>
</tr>
<tr>
<td>Segment 3</td>
<td>~ 3 miles</td>
</tr>
<tr>
<td>Segment 4</td>
<td>~ 8.5 miles</td>
</tr>
<tr>
<td>Segment 5</td>
<td>~ 7 miles</td>
</tr>
</tbody>
</table>

For each segment in Table 1, data were collected by driving on the highway and recording GPS data. The distance between collected data points was a function of the driving speed of the vehicle and the collection rate of the GPS receiver. For example, in this study, the test vehicle was driven at 60 mph and the data points were effectively collected every tenth of a second. This resulted in spacing
between data points equal to roughly 8.8 ft. At each test site, four runs were made in each travel direction (one of the runs in segment 2 was rejected because the collected data were corrupted). Each time, the instrumented vehicle took a running start. This was done for two major reasons. The first reason was safety. By entering the traffic stream, the instrumented vehicle did not have to sit in the middle of the road, which could have been hazardous. The second reason was that the vehicle could reach its target speed of 60 mph before it crossed the beginning point of the segment. When the vehicle crossed that point, the DMI was turned on and cruise control was set at 60 mph. Video data were collected along with the horizontal and vertical GPS data. This allowed for visual verification of the terrain and recording of pavement markings, as DMI distances were superimposed on the video. Dewesoft software enabled the researchers to collect video, DMI, and GPS data that were easily referenced to one another, and to record events at the beginning of the run, at the end, and at every change between passing and no-passing zone pavement markings.

**GPS Data Transformation**

Once the GPS raw data are collected in the format of longitudes, latitudes, and altitudes, a suitable map projection is selected to transform the terrestrial coordinates on the curved surface of the Earth to a planar Cartesian coordinate system, in which longitudes and latitudes (λ and φ) are converted into easting and northing coordinates (x and y), where x corresponds to the east-west dimension and y to the north-south. A map projection is a mathematical algorithm to transform locations defined on the curved surface of the earth into locations defined on the flat surface of a map. The conversion of the curved surface to the planar surface is always accompanied by some distortion owing to the spheroidal/ellipsoidal shape of the earth; however, map projection can preserve one or several characteristics of the surface at the cost of distorting other features. Since the data were collected in Texas, the Texas Centric Mapping System/Lambert Conformal (TCMS/LC) was chosen as the final coordinate system. The constants that are unique for the TCMS/LC and needed for the conversion process are shown in Table 2.22,23

<table>
<thead>
<tr>
<th>Variables</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>First eccentricity</td>
<td>e</td>
<td>0.081819191</td>
</tr>
<tr>
<td>Semimajor axis</td>
<td>a</td>
<td>6378137 m</td>
</tr>
<tr>
<td>Flattening</td>
<td>f</td>
<td>1/298.257222101</td>
</tr>
<tr>
<td>Latitude grid origin</td>
<td>φ₀</td>
<td>18° N</td>
</tr>
<tr>
<td>Longitude grid origin</td>
<td>λ₀</td>
<td>100° E</td>
</tr>
<tr>
<td>Northern latitude parallel</td>
<td>φₙ</td>
<td>35° N</td>
</tr>
<tr>
<td>Southern latitude parallel</td>
<td>φₛ</td>
<td>27.5° N</td>
</tr>
<tr>
<td>False easting</td>
<td>E₀</td>
<td>1500000 m</td>
</tr>
<tr>
<td>False northing</td>
<td>N₀</td>
<td>5000000 m</td>
</tr>
</tbody>
</table>


For inclusion in a no-passing zone location algorithm, all the GPS points collected in the previous step had to be converted into northing and easting values (see Wolf and Ghilani23 for a complete set of equations).
GPS Data Smoothing

The accuracy of GPS data, especially when it is collected from a moving vehicle, can vary drastically owing to satellite positions, and smooth profiles cannot be taken directly from a single GPS data collection run. However, multiple data collection runs result in unnecessarily repetitious data. The researcher must “clean” the data to eliminate repetition and errors. An important objective in this research was to smooth the GPS data and obtain the best curve representing the geometry of two-lane highways. The LOESS technique was used to smooth the GPS data. LOESS or LOWESS (locally weighted scatterplot smoothing), originally proposed by Cleveland, is a nonparametric regression technique to fit bivariate data. The technique is considered a local smoothing process because each smoothed value is determined by neighboring data points defined within a bandwidth (span). The LOESS procedure offers the simplicity of linear least squares regression as well as the benefits of the flexibility of nonlinear regression. The technique is computationally intensive and involves a number of steps. Assuming a set of n observations of x and z (the observation data points are not necessarily evenly spaced), \( \{(x_1, z_1), (x_2, z_2), (x_3, z_3), \ldots, (x_n, z_n)\} \) the LOESS technique first defines m equally spaced locations across the range of x values of observation data. The equally spaced locations are named \( s_j \) where \( j \) ranges from 1 to m. Each of the locations contains a bandwidth that includes q observation data points. Because the number of observation data points is constant within each bandwidth, the physical widths of the bandwidths corresponding to each \( s_j \) are different. A low-degree polynomial is fit at each of the bandwidths using weighted regression. The procedure is as follows:

Neighborhood weights are calculated for each \( s_j \) using the following tricube weight function (the weight function gives more weight to data points near \( s_j \) and less weight to points farther away):

\[
 w_i(s_j) = \begin{cases} 
 (1 - \left| \frac{\Delta_i(s_j)}{\Delta_q(s_j)} \right|^3)^2 & \text{for } \Delta_i(s_j)^* < 1 \\
 0 & \text{Otherwise} 
\end{cases} \quad \text{for } i = 1 \text{ to } n 
\]  

(1)

where

\[
 \Delta_i(s_j)^* = \frac{\Delta_i(s_j)}{\Delta_q(s_j)} 
\]

(2)

and \( \Delta_i(s_j) \) and \( \Delta_q(s_j) \) are the distances from \( s_j \) to the \( i^{th} \) value of x and the \( q^{th} \) farthest \( x_i \) (within the bandwidth), respectively:

\[
 \Delta_i(s_j) = |x_i - s_j| 
\]

(3)

\[
 \Delta_q(s_j) = |x_q - s_j| 
\]

(4)

The weighted least squares method is used to estimate coefficients \( b_{kj} \) in such a way that Equation (5) is minimized:

\[
 \sum_{i=1}^{n} w_i(s_j) \left( z_i - \sum_{k=0}^{\delta} b_{kj} x_i^k \right)^2 
\]

(5)

where \( \delta \) is the degree of the polynomial equation to be locally fitted to the observation data points. \( \delta \) gets the value of either 1 or 2, indicating linear or quadratic function, respectively.

The next step is to use a robust estimation procedure to update the coefficient derived from each regression. This procedure reduces the sensitivity of the LOESS to unusual data points and offsets outlier effects. The estimation procedure finds the robust coefficients \( b_k \) :

The robustness weight for each observation data point is calculated using the following bisquare weight function:

\[
 r_i = \begin{cases} 
 (1 - \left| e_i^2 \right|^2)^2 & \text{for } |e_i| < 1 \\
 0 & \text{Otherwise} 
\end{cases} \quad \text{for } i = 1 \text{ to } n 
\]  

(6)
$e_i^* = \frac{e_i}{\operatorname{Median}|e_i|}$

(7)

where $e_i$ the residual from the proceeding local regression for the $i^{th}$ observation data point:

$$e_i = z_i - g(s_j) = z_i - \sum_{k=0}^{\delta} b_{kj}s_j^k$$

(8)

The robust estimation procedure minimizes the influence of outliers by examining the residuals from the fitted curve and downweighting those that are relatively large. Using the robustness weights, the updated regression coefficients $b_{kj}$ that minimize Equation (9) are found.

$$\sum_{i=1}^{n} r_i w_i(s_j)(z_i - \sum_{k=0}^{\delta} b_{kj}s_j^k)^2$$

(9)

The robust estimation procedure is repeated until the estimated coefficients converge. The robust coefficients are used to calculate the vertical coordinates at each bandwidth, $g(s_j)$ by taking the predicted value of $Z$ at $s_j$:

$$g(s_j) = \sum_{k=0}^{\delta} b_{kj}s_j^k$$

(10)

Those coordinates are the fitted values:

$$\{(s_j, g(s_j))\} = \{(s_1, g(s_1)), (s_2, g(s_2)), (s_3, g(s_3)), \ldots, (s_m, g(s_m))\}$$

The smooth curve is produced by connecting the adjacent fitted points (i.e., $s_j$ and $g(s_j)$).

Two parameters had to be determined before the LOESS technique could be used: (1) the degree of the polynomial equation to be locally fitted to the observation data points ($\delta$) and (2) the bandwidth that included the neighboring data points to be locally weighted and fitted. A quadratic equation was selected in the smoothing technique to be locally fitted to the observation data points ($\delta = 2$) because it has the capacity to fit either a straight line tangent or a parabolic curve. This characteristic of the quadratic least squares method makes it more suitable for use in smoothing a roadway's vertical profile, as it contains both tangents and curves. On the other hand, selecting the bandwidth could be a frustrating task. Because the data were collected from a test vehicle moving at a constant speed, they were evenly spaced and the bandwidth lengths were the same. Early on in the creation of the smoothing process, the authors attempted to evaluate the effects of various combinations of smoothing iterations and bandwidth lengths; however, clear trends in the selection of an appropriate bandwidth did not appear. It was concluded that a bandwidth was needed that reflected typical lengths of design curves. A bandwidth of 900 ft. was selected for several reasons. First, a precedent was set in research conducted at Kansas State University. When Ben Arieh and colleagues wanted to clean the GPS data and use B-spline to fit the data, they selected 984 ft. because they had found that this distance constituted an “identifiable segment of road.” Second, in collecting data points once every second at 60 mph, this bandwidth distance gives approximately five data points to either side of the test point, which was determined to be an adequate number for smoothing. Finally, if the bandwidth is too small, there will not be enough data points to use in the smoothing process and the effect of smoothing will be lost. Thus, if a vertical curve is present, the smoothing process might not recognize it and could represent it as a tangent. Alternatively, if the bandwidth is too large, the process runs the risk of smoothing out the data too much and distorting the roadway geometry. A 900-ft. bandwidth seemed to provide a good compromise between these two extremes. By smoothing the GPS data using the LOESS regression smoothing process, the researchers generated new estimated elevation points at evenly spaced station values.
Development of GPS Conversion Program

A preprocessing program, the GPS Conversion Program, included a user-friendly procedure. The purpose of the program was to process raw data collected from straight alignments and generate the input file for the no-passing zone computer model. The GPS Conversion Program eliminates the extra data points and cleans the data, converts GPS coordinates (longitude and latitude values) to northing and easting values, calculates stations and elevations for the segment of the roadway, and generates the no-passing zone computer model input file. The GPS raw data for each run underwent preprocessing using the program, and the stations and elevations were calculated. Then the results of the preprocessing steps for all the runs in each direction of the roadway were combined and sorted according to station to create input files for the no-passing zone computer model. To achieve smaller station intervals, station values were generated every 10 ft. from station zero to the end of the roadway run.

Algorithm Development

According to the Manual on Uniform Traffic Control Devices (MUTCD), “The passing sight distance on a vertical curve is the distance at which an object 3.5 ft. above the pavement surface can be seen from a point 3.5 ft. above the pavement” (see Figure 1).

In vertical curves, a no-passing zone should be provided any time the pavement surface restricts the sight line. The sight distance algorithm might be developed in a way that examines the intersection of the sight line that originated from points located 3.5 ft. above the centerline and the pavement surface.

An iterative process is used in the algorithm to measure the availability of required vertical sight distance and to locate no-passing zones. The algorithm has distinct iterative steps:

Figure 1. No-passing zone at vertical curve (Figure 3B-4, MUTCD).
The first step is to determine the observation points where the availability of sight distances are checked for those locations. Sight distances are checked for the observation points located every 10 ft. along the roadway. Those points are referred to as \( O_i, O_{i+1}, O_{i+2} \) and so on.

The second step pertains to the testing of iterative locations downstream of the first observation point (for example \( O_i \) in Figure 2a). A target point, referred to as \( T_{ij} \), is selected 50 ft. from the observation point. The interval between point \( O_i \) and point \( T_{ij} \) is tested for sight distance restriction. If a sight restriction is not found, \( T_{ij+1} \) (which is 50 ft. farther down the road and 100 ft. from the point \( O \)) is tested for sight restrictions. Assuming no sight restrictions are found from \( O_i \) to each successive target point, the process is repeated until the minimum required passing sight distance as set by the MUTCD (Table 3) is reached (point \( T_s \)). If no sight restrictions from point \( O \) are found, the algorithm assigns 1 as a passing attribute to this point (passing attribute = 1). If a sight obstruction is found in any interval, the iteration (related to this step) stops and the algorithm assigns 0 as a passing attribute to the observation point (passing attribute = 0). Figure 2a shows a sight distance obstruction between \( O_i \) and \( T_{ij+2} \).

The sight distance restriction is checked for each observation point, that is, \( O_{i+1}, O_{i+2}, O_{i+3} \) and so on. Figure 2b shows the process for observation point \( O_{i+1} \).

Table 3. Minimum passing sight distances for no-passing zone markings (Table 3B-1, MUTCD).

<table>
<thead>
<tr>
<th>85th Percentile, or Posted or Statutory Speed Limit</th>
<th>Minimum Passing Sight Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 mph</td>
<td>450 ft.</td>
</tr>
<tr>
<td>30 mph</td>
<td>500 ft.</td>
</tr>
<tr>
<td>35 mph</td>
<td>550 ft.</td>
</tr>
<tr>
<td>40 mph</td>
<td>600 ft.</td>
</tr>
<tr>
<td>45 mph</td>
<td>700 ft.</td>
</tr>
<tr>
<td>50 mph</td>
<td>800 ft.</td>
</tr>
<tr>
<td>55 mph</td>
<td>900 ft.</td>
</tr>
<tr>
<td>60 mph</td>
<td>1,000 ft.</td>
</tr>
<tr>
<td>65 mph</td>
<td>1,100 ft.</td>
</tr>
<tr>
<td>70 mph</td>
<td>1,200 ft.</td>
</tr>
</tbody>
</table>
In general terms, the determination of the theoretical sight line calculated for each interval from point $O_i$ to the points 50 ft. farther down the road is as follows. First, 3.5 ft. are added to the roadway elevations at point $O_i$ and point $T_{i,j}$. The line between these two points, which are 3.5 ft. above the pavement surface in space, is the theoretical sight line. Next, an equation of the line between these two points is written in which stationing is the independent variable and elevation is the dependent variable. After determining the equation of this line, 10-ft. station increments are identified between point $O_i$ and the last station of the theoretical sight line at point $T_{i,i'}$. Elevations are found for each of these stations on the sight line and compared with the corresponding roadway elevations at those precise stations as determined by the smoothing model. There are no sight restrictions in a given iteration if the sight line elevations are greater than the roadway profile. However, if at any station a roadway pavement elevation is greater than its corresponding sight line elevation, the sight distance is inadequate and a no-passing zone exists. This occurrence breaks the loop of checking $O_i$ and successively more distant points.

This process is repeated from the beginning to the end of the roadway segment. After examining the entire segment, the algorithm identifies the beginning and end of each no-passing zone on the basis of assigned passing attributes (i.e., no-passing zones would be required between consecutive points that have passing attributes of 0). After identifying the no-passing zone segments, the computer model checks adjacent no-passing zones to see if the distance between segments is less than 400 ft. If so, the last station of the first segment and the first station of the second segment are deleted and one no-passing zone is created. Closely located no-passing zones are joined in compliance with

---

**Algorithm 1: Vertical sight distance evaluation between $O_i$ and $T_{i,s}$**

```plaintext
Input i
Input s
Initiate \{O_i(x_{O_i}, z_{O_i}), T_{ij}(x_{T_{ij}}, z_{T_{ij}}), T_{ij+1}(x_{T_{ij+1}}, z_{T_{ij+1}}), \ldots, T_{is}(x_{T_{is}}, z_{T_{is}})\}
Generate \{O_i(x_{O_i}, z_{O_i}+3.5), T_{ij}(x_{T_{ij}}, z_{T_{ij}}+3.5), T_{ij+1}(x_{T_{ij+1}}, z_{T_{ij+1}}+3.5), \ldots, T_{is}(x_{T_{is}}, z_{T_{is}}+3.5)\}
Flag_problem ← false
j ← i + 1
repeat
Connect O_i to T_{ij}
k ← i + 1
repeat
if $z_{O_i,T_{ij}} > z_{O_k}$ then
  k ← k + 1
else
  Flag_problem ← true
  Break
end if
until $O_k = T_{ij}$
if Flag_problem = true
  Break
else
  j ← j + 1
until j > s
if j > s then
  return “There is adequate vertical sight distance between $O_i$ and $T_{i,s}$.”
else
  return “There is NOT adequate vertical sight distance between $O_i$ and $T_{i,s}$.”
end if
```
MUTCD guidelines: “Where the distance between successive no-passing zones is less than 400 ft., no-passing zone markings should connect the zones.” The vertical sight distance evaluation procedure is described in Algorithm 1.

“Where the distance between successive no-passing zones is less than 400 ft., no-passing zone markings should connect the zones.”

The algorithm was incorporated into a computer model to develop an automated method for calculating passing sight distance and locating no-passing zones on the basis of the GPS data. Compared with field methods, the computer model saves time and money, and eliminates human errors. The available passing sight distance on a two-lane highway depends on the travel direction, with the result that no-passing zones for traffic in both directions may or may not overlap. The traditional time-consuming method for determining the location of no-passing zones is to measure passing sight distances in the field for both travel directions. Algorithm 1 is based on the evaluation of available sight distance in one direction along the roadway alignment (one-directional sight distance algorithm). The computer model simply incorporates the algorithm twice (once in each direction) to evaluate sight distance for a segment of roadway.

Figure 3 illustrates the computational flowchart of the no-passing zone algorithm.

---

**Figure 3. Computational flowchart of no-passing zone algorithm.**
Data Analysis

To confirm the viability of the sight distance algorithm and the related GPS data smoothing procedure, the no-passing zone computer model was applied to the GPS data for the highway test segments, and no-passing zones were calculated on the basis of not only the individual runs but also the combined runs of collected data for each direction of the test roadways. Two methodologies were applied to compare the results of the computer model with existing no-passing zone markings: (1) safety factor methodology and (2) the absolute percentage and the root mean square error (RMSE) comparison methodology.

Safety Factor Methodology

Inaccurate GPS elevation readings at driver and sight object locations could affect no-passing zones. The critical points in establishing a no-passing zone are at the beginning and end of the zone; at these locations, inaccurate elevation data will have the greatest influence. One option to avoid this problem is to simply add a set distance to the beginning and end of a no-passing zone; however, this may not be the best approach. No-passing zones are of different lengths, so the effects of adding a uniform value to the beginning and end will be greater on a short no-passing zone than on a long no-passing zone. Thus, a percentage-based no-passing zone safety factor was tested. The researcher examined a 25 percent safety factor in which 12.5 percent was added to the beginning and end of the no-passing zones. The 25 percent safety factor was selected on the basis of engineering judgment, with a trade-off between safety and operational impact. Application of the safety factor to no-passing zone length increases the likelihood that a no-passing zone location does not go unmarked but no-passing zones are not extended to the point that they degrade traffic operations. (The authors plan to evaluate the safety factor in detail in future research.) After the safety factor distances are added, the no-passing zones are reevaluated and adjacent no-passing zones that are less than 400 ft. apart can be combined.

Figure 4 provides a visual comparison of existing no-passing zone markings, the results of the original no-passing zone computer model without a safety factor, and the resulting no-passing zones with safety factor added for the segment. For brevity, only the results with four runs combined are shown. To complete all possible combinations for each travel direction at each test site, the no-passing program would have to be run 90 times. To reduce the number of times the program was run, a single travel direction was selected for each test site: segment 3 eastbound (EB), segment 4 southbound (SB), and segment 5 northbound (NB). Additionally, a single scenario was randomly selected for each of the possible single, double, triple, and quadruple run combinations (see Table 4).

Table 4. Tested run combination.

<table>
<thead>
<tr>
<th></th>
<th>Segment 3 EB</th>
<th>Segment 4 SB</th>
<th>Segment 5 NB</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 runs</td>
<td>1234</td>
<td>1234</td>
<td>1234</td>
</tr>
<tr>
<td>3 runs</td>
<td>234</td>
<td>234</td>
<td>123</td>
</tr>
<tr>
<td>2 runs</td>
<td>13</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>1 run</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

In Figure 4, the benefits of a safety factor can be seen in no-passing zone 3, where the beginning and ending locations are extended beyond the existing markings. This is acceptable. Additional benefits of the safety factor can be seen in no-passing zone 6, where the original no-passing zone segments from the computer model have been combined to create a single no-passing zone.
However, the safety factor can come up short in its effectiveness; for example, in no-passing zones 7 and 8, the beginning locations do not extend to those of the existing markings. On segment 4, at least two significant horizontal curves were noted while driving the roadway. Those two horizontal curves occurred in no-passing zones 1 and 7 (identified by the circles in Figure 4). Zone 1 does not seem to be affected by the horizontal curves. As for zone 7, the horizontal curve does not have a significant impact because of the nature of the vertical geometry of the roadway at this location. However, it is difficult to quantify the effect of the horizontal curves because they occurred in conjunction with vertical curves.

Absolute Percentage and RMSE Comparison Methodology

A question that arose in the data analysis was whether a single run would be sufficient or whether data from two, three, or four runs combined would produce better results. The lengths of the existing no-passing zones were obtained by averaging DMI video results from the multiple runs in each travel direction at the test sites. These provided the basis for comparing the results of the no-passing zone algorithm. In processing the collected GPS values in the computer model, the results include several no-passing zones that, when grouped together, closely resembled the existing markings. Thus, when we analyzed the data to determine how to produce the best no-passing zone results, we decided to group algorithm-identified no-passing zones that seemed similar to existing markings. From these values, two measures were calculated. The first was the absolute percentage difference between existing no-passing zone lengths and the lengths of the grouped no-passing zones from the computer model. The second measure calculated was the root mean square error (RMSE) value of each of the no-passing zones in a given tested scenario. These measures were calculated for each of the single and combined runs.

Figure 4. Segment 4 (southbound) safety factor comparison.
To reduce the number of times the no-passing zone computer model was run, a single travel direction for each test site was selected, and a single scenario for each of the possible single, double, triple, and quadruple run combinations was randomly selected.

Figure 5 shows the average absolute percentage differences between the lengths of the grouped no-passing zones from the computer model and the existing zones for segments 3, 4, and 5. Figure 6 compares the RMSE value for no-passing zones at those three sites. The calculations were completed for each roadway and each scenario of a single run, two runs combined, three runs combined, four runs combined, and four runs combined with the safety factor for one direction on each of the three evaluated roadways. It is obvious that the variability in combined runs of segment 3 can be traced back to the GPS data.

Although Figures 5 and 6 show no clear trends, the researchers believe that the best method is to use at least four combined runs. As shown in these figures, four combined runs are not guaranteed to outperform three combined runs or even a single run; however, the combination run method does produce more stable results and avoids the problem of using a single run that is inaccurate.

Figures 5 and 6 show the average absolute percentage difference in no-passing zone lengths and the RMSE in no-passing zone lengths for four combined runs with safety factors applied. In general, the calculated no-passing zones were shorter than the existing no-passing zones. The safety factor improved the overall results but does not completely remedy the problem because the averages of the absolute percentage differences and the RMSE values are not zero. Furthermore, the research did not try to confirm the accuracy of the existing no-passing zones, so there could be errors in the placement of the existing zone markings.

**Conclusion and Recommendations**

In this article, a new application of a mathematical technique for smoothing GPS data was introduced and an analytical algorithm to use in evaluating the available passing sight distance along the vertical profiles of two-lane highways was developed. Processes for collecting, converting, and smoothing the GPS data used in the algorithm were described. The algorithm uses an iterative process to determine whether the pavement surface is above the sight line. The computer model captured the general
locations of the no-passing zones, and a safety factor was added to improve the model’s capabilities compared with existing field-measured no-passing zone markings. A potential continuation of this research would be to develop criteria for selecting the safety factors for the calculated lengths of no-passing zones. In most cases, the existing no-passing zones in the field were longer than necessary. However, it is not known whether the existing zone markings are in accordance with MUTCD sight distance criteria. Additionally, some adjacent no-passing zones are close enough that they are connected in the field, even though they may be far enough apart to exist as separate no-passing zones according to the MUTCD.

The evaluation of passing sight distance in the vertical plane as described in this article is one part of a research effort to examine the locations of no-passing zones on two-lane highways. As part of the overall effort, the authors are assessing alternatives for evaluating passing sight distance in the horizontal plane and possible methods of combining the two. The evaluation parameters for sight distance in the vertical plane are easier to address in an analytical model because all the physical parameters can be quantified using GPS coordinates. The evaluation of passing sight distance in the horizontal plane will require assumptions for parameters that cannot be defined by GPS coordinates.

Acknowledgments

Portions of this research were completed as parts of research projects sponsored by the Texas Department of Transportation and the Southwest Region University Transportation Center.

References


Mehdi Azimi is a Ph.D. candidate at Texas A&M University. He also works as a graduate research assistant in the Operations and Design Division at Texas Transportation Institute (TTI). He received his master degree in transportation planning and management from Texas Southern University. Mehdi’s research interests are roadway geometric design, traffic control devices, traffic operation, and signal control systems. He is a member of ITE.

Cameron L. Williams, P.E., PTP, received his master and bachelor of science in civil engineering degrees from Texas A&M University. Cameron is employed at Walter P Moore where his experience includes many different projects such as traffic studies, traffic modeling, traffic signal design, and safe routes to school programs. Cameron was a 2007 ENO Fellow and was awarded the 2008 Milton Pikarsky Memorial Award. Cameron has been actively involved in ITE since graduate school.

Dr. H. Gene Hawkins, P.E., is an associate professor in the Zachry Department of Civil Engineering at Texas A&M University and holds a joint appointment as a research engineer with the Texas Transportation Institute (TTI). He received his Ph.D., M.E., and B.S. degrees in civil engineering from Texas A&M University. He is a Fellow of ITE.
Rumble strips have been used for many years to provide vibratory feedback to drivers who are drifting out of their lanes. Often these strips were placed next to the marked lines; recently agencies have started combining the markings with the strips. These painted rumble strips—known as “rumble stripes”—are emerging as centerline and edge line treatments that can alert drivers who are beginning to move outside the marked lane. A collateral benefit of painting edge lines in the rumble strip is the potential to increase the retroreflective durability of the lines, particularly in areas that have substantial snowplowing operations in the winter.

Introduction

Many states augment painted lanes to increase preview distance. Snowplowable raised pavement markers (SRPMs) are an example of a technology used in conjunction with painted lines to enhance nighttime visibility. However, SRPMs are expensive to install and maintain and are susceptible to dislodging as the roadway degrades. A dislodged SRPM can cause tire punctures and increased degradation of the pavement, and can become an airborne projectile that can penetrate windshields (Figure 1). The tort liability exposure associated with dislodged SRPMs has caused several states—including Iowa, Missouri, and Nebraska—to start removing all SRPMs.

Many agencies are using painted rumble strips—as also called rumble stripes—as an alternative to SRPMs for centerline and edge line treatments to enhance nighttime line visibility and provide vibratory feedback to drivers who are drifting out of their lane. A collateral benefit of painting edge lines along the rumble strips is the potential to increase the retroreflective durability of the lines, particularly in areas that have substantial snowplowing operations.
In this study, rumble stripes were installed next to the painted edge line to establish a side-by-side comparison of the two treatments. Data consisting of high resolution photography, video footage, and retroreflectivity measurements were collected before and after the 2010–11 winter season to determine the effects of snowplowing operations on retroreflectivity degradation. On the basis of qualitative and quantitative assessments, the rumble stripes outperformed the painted lines under both dry and wet conditions. The increased retroreflective performance and durability of the rumble stripe, combined with the auditory and vibratory feedback, led researchers to conclude that rumble stripes are a reasonable substitute for SRPMs on rural state highways.

Background
Over the past decade, shoulder, edge line, and centerline rumble strips have been widely installed in many states. Although a number of state standards exist for the use of centerline rumble strips, a national design and performance evaluation standard has yet to be established. A few studies have been conducted on the placement of paint on rumble strips, creating rumble stripes. A 2006 study by Lindly and Narci for the University Transportation Center for Alabama compared the retroreflectivity of a standard thermoplastic line with that of a rumble stripe using a mobile retroreflectometer to take measurements in dry and wet conditions as defined in ASTM 2177. Lindly and Narci concluded that the rumble stripe degraded more slowly than the thermoplastic line and remained retroreflective when wet. Wilder, in a 2010 study conducted for the New York Department of Transportation, noted the improved visibility of the edge line when it was placed within the rumble strip, particularly under wet nighttime conditions. As part of the National Cooperative Highway

Figure 1. Indiana Department of Transportation (INDOT) snowplow windshield damage caused by a dislodged SRPM.

[Image of a dislodged SRPM with windshield damage]

Photo courtesy Stacey Flick, INDOT.
Research Program’s (NCHRP’s) 2009 Report 641, a synthesis was conducted in North America that generally recommended installation of centerline and shoulder rumble strips on rural roadways to provide an auditory and vibratory warning of lane departure. The report briefly mentions the visibility/retroreflectivity of rumble stripes, saying that they offer improved visibility, particularly under wet nighttime conditions, but it does not define performance measures. Pike, Ballard, and Carlson recently outlined proper procedures for measuring the retroreflectivity of rumble stripes. Our study equaled or exceeded their minimum step size for moving the retroreflectometer as well as their recommended stabilization procedure for measuring the rumble stripes.

In 1992, Congress suggested the addition of a standard for minimum pavement retroreflectivity to the Manual on Uniform Traffic Control Devices (MUTCD). Although a standard does not yet exist, FHWA-sponsored research has produced numerous reports, notably the 1999 unpublished report by Turner and the 2007 update by Debaillon, Carlson, He, Schnell, and Aktan. In this article, we use the Turner numbers as the threshold. Debaillon et al. suggested a slight lowering of the Turner numbers, but their recommendation clearly acknowledges the limitations of their minimum and the importance of sound engineering judgment, both of which led us to compare our roadway with the Turner-recommended minimums of 100 and 65 mcd/m²/lux for white and yellow, respectively.

Study Objective

The objective of this study was to develop and execute an experimental test protocol to compare the retroreflective characteristic of rumble stripes and standard painted lines. The study was conducted after a season of winter snowplowing operations to identify any disparities in the durability of retroreflectivity between the rumble stripe and the painted line. We also compared standard glass beads with a blend of glass beads and retroreflective elements designed for enhanced performance in wet weather conditions. The paint, glass bead, and element application rates are shown in Table 1.

Study Area

To provide a field laboratory for the qualitative and quantitative evaluation of rumble stripes, we selected a site along a two-lane section of divided highway between mile marker (MM) 52–54 eastbound along Indiana State Route 28 (Figure 2). The rumble stripes were compared with the standard painted line using both standard glass beads and an element blend in waterborne paint. The corridor was divided into two sections:
Site 1 (MM 52–53)—type I glass beads

Site 2 (MM 53–54)—element blend

For a schematic of the rumble stripes and a rumble cross-section, see Figure 3a and Figure 3b, respectively. Photographs of the painted lines (glass beads) are shown in Figure 3c (yellow) and Figure 3d (white) (i and iii refer to the painted lines, while ii and iv refer to the adjacent rumble stripes).

Figure 4 includes photos taken from a vehicle traversing the study section with edge line rumble stripes and standard edge lines. Figures 4a and 4b show a daytime photo of the roadway with rumble stripes along the left (yellow) and right (white) sides. Figures 4c and 4d show how painted rumble stripes and traditional lines appear at night during dry conditions. Figures 4e and 4f compare the two under wet pavement conditions.

Table 1. Composition of glass beads and element blend.

<table>
<thead>
<tr>
<th></th>
<th>Paint Application Rate</th>
<th>Glass Beads Application Rate</th>
<th>Element Application Rate</th>
<th>Paint Thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I glass beads</td>
<td>16 gal/mile</td>
<td>104 lbs/mile</td>
<td>N/A</td>
<td>16 mil</td>
</tr>
<tr>
<td>Element blend</td>
<td>22 gal/mile</td>
<td>104 lbs/mile</td>
<td>84 lbs/mile</td>
<td>20 mil</td>
</tr>
</tbody>
</table>

Methodology

The study protocol included qualitative, macroscopic, and microscopic components. A qualitative and quantitative analysis compared standard white and yellow painted lines with rumble stripes using high resolution photography, comparative daytime/nighttime video analysis, and portable retroreflectometers. Evaluation was based on high resolution photographic data and retroreflectivity measurements collected along both white and yellow painted lines and rumble stripes. High resolution photographic data were collected before and after the 2010–11 winter season to assess glass bead and element loss on a microscopic scale. Retroreflectivity measurements of both the standard painted line and the rumble stripe were also analyzed before and after the 2010–11 snowplow season.

High Resolution Photography

In November 2010, before the snow season began, an agency crew drilled 6-in. cores in each section through the painted line, rumble stripe top, and rumble stripe bottom of both edge lines (a total of six cores per test section). High resolution photographs of the pavement marking on the surface of each core were analyzed in the lab. The pictures were orthorectified and then scaled to a 1/4-in. grid. This process of rectifying and gridding was repeated on high resolution field photographs taken in both test sections after the winter plowing season concluded. Bead and element counts were taken from the rectified and gridded images to compare the effects of plowing and general winter wear on the painted line, rumble stripe top, and rumble stripe bottom, as seen in Figure 5. In the figure, i and ii show glass beads in the paint a few weeks after the line was painted, while iv shows a missing glass bead in a section of the yellow line after a season of winter plowing operations.
Figure 3. SR 28 test site layout.

a) Plan view detail of rumble strip test layout on SR 28.

b) Cross-section detail of rumble strip.

c) Yellow test section:
   (i) painted line, (ii) rumble stripe.

d) White test section:
   (iii) painted line, (iv) rumble stripe.
Figure 4. Screen images from driving videos in test zone with standard INDOT glass beads.

a) Day dry yellow (January 2011).

b) Day dry white (January 2011).

c) Night dry yellow (February 2011).

d) Night dry white (February 2011).

e) Night wet yellow (April 2011).

f) Night wet white (April 2011).
Retroreflectometer Equipment

The Delta LTL-X is the current generation of pavement retroreflectometer, replacing the widely used LTL-2000. Both adhere to the ASTM E 1710 standard 30m geometry. However, the LTL-X observes an area approximately four times larger than the LTL-2000, as shown in Figure 6, and is capable of measuring milled rumble stripes. In accordance with the operator’s manual, the retroreflectometer was calibrated at least daily to the field calibration unit. At the beginning of each week, the retroreflectometer was calibrated by a certified operator in the lab to the reference lab calibration unit (black), and the field calibration unit (red) was checked for accuracy. As measured by the retroreflectometer, neither calibration unit deviated from the factory set 151 mcd*m\(^{-2}\)*lx\(^{-1}\) during the testing. The stability of the retroreflectometer when it was in a rumble was corrected by means of hand-leveling.

Data Collection

The 2010–11 winter season required more snowplowing operations than the typical Indiana winter. Local plowing operations consist of single- and double-rear-axle municipal plow trucks with standard carbide-edged steel plows rigged in front and in a wing configuration. Because it was close to the Frankfort substation, the studied section of SR 28 probably received more plowing than the typical Indiana state route.

Characterization of Rumble Strip Retroreflective Sensitivity to Rumble Strip Geometry

To evaluate the sensitivity of the retroreflectivity measurement traversing a rumble stripe, four 4-ft. areas of good pavement were selected within each 1,500-ft. section. In each area, three retroreflectivity measurements were taken at 1-in. increments on both the painted line and the rumble stripe in accordance with the procedures outlined in a vendor technical note. Figure 7 shows a time lapse (in 3-in. increments) of the Delta LTL-X retroreflectometer performing readings through 21 in. of the 48-in. section. The sensitivity study was reevaluated under conditions of continuous wetting designed to follow ASTM E 2176.
Painted line and rumble stripe sensitivity data were collected using the LTL-2000 and LTL-X in the same area in Site 1. Figure 8a shows the results of the LTL-2000 sensitivity test, with three measurements per location. Figure 8a illustrates some sensitivity to the location of the measuring device in the rumble. Figure 8b shows LTL-X results from the same location. Measuring a much larger area (Figure 6) makes the LTL-X less sensitive to point variations and location in the rumble stripe. Also characterized is the repeatability of the LTL-X. The graph in Figure 8b shows three measurements per location without moving the LTL-X between readings. Figure 8c shows the results from the same location obtained by taking three measurements per location while moving the LTL-X between readings. There was little variation between the two methods, so the first method was used to conserve time.
Figure 7. Time-lapse of Delta LTL-X through rumble strip sensitivity section; (i) center of measurement area (June 2011).

a) Delta LTL-X at d=0"
b) Delta LTL-X at d=3"
c) Delta LTL-X at d=6"
d) Delta LTL-X at d=9"
e) Delta LTL-X at d=12"
f) Delta LTL-X at d=15"
g) Delta LTL-X at d=18"
h) Delta LTL-X at d=21"
Figure 8. Comparison of Delta retroreflectometers and repeatability tests on white line.

a) Delta LTL-2000: 1 pass, 3 samples per position (May 2011).

b) Delta LTL-X: 1 pass, 3 samples per position (June 2011).

c) Delta LTL-X: 3 passes, 1 sample per position (June 2011).
Wet Retroreflectivity of Painted Line and Rumble Stripe

The results obtained from continuous wetting of both the white paint glass beads and the element blend are shown in Figure 9. (Note that the vertical scale in Figure 9a has been reduced from 600 to 60.) In both cases, the painted line lost most of its retroreflectivity. In Figure 9a, the retroreflectivity of the glass beads in the rumble stripe was drastically reduced, but they remained retroreflective.

Figure 9b shows that when the element blend is used, the wet retroreflectivity of the rumble stripe improves substantially and exceeds even the FHWA suggested replacement threshold for dry pavement (100 mcd·m⁻²·lx⁻¹ for white). These quantitative findings for wet conditions match the qualitative observations in wet conditions as shown in e and f, where the reflectivity of the painted line is almost imperceptible. Although the white standard line in Figure 4f can be seen during wet conditions, the actual retroreflectivity values were taken during a controlled continuously wetted condition, which may account for some of the difference between the qualitative and quantitative values. Without knowing how retroreflectivity measurements correlate with human ocular perception, which is outside the scope of this article, we do not know what level would be detectable when compared directly with the retroreflective values.

Collection of Retroreflectivity Data from Study Section

At both sites, a random 1,500-ft. section was chosen to collect retroreflectivity data. These data were collected to evaluate the variability along the lines as well as the sensitivity to measurement location within the rumble stripe.

Figure 10 shows the macroscopic effects likely to influence retroreflectivity, with portions of the line completely removed from the pavement.

To ensure representative random sampling, each 1,500-ft. test section was subdivided into 15-ft. increments, exceeding the sample size of 20 recommended by ASTM D7585. Regardless of paint condition, retroreflective data were collected on the following:

- White painted line—direction of travel;
- White rumble stripe top—direction of travel;
- White rumble stripe bottom—direction of travel;
Yellow painted line—both directions;

Yellow rumble stripe top—both directions; and

Yellow rumble stripe bottom—both directions.

Dry Retroreflectivity of Painted Line and Rumble Stripe

Three measurements per location were used to characterize the 15-ft. intervals along the 1,500-ft. sections in Site 1 and Site 2: one measurement of the painted line, one of the rumble stripe top, and one of the rumble stripe bottom. The results from the white and yellow glass bead sections are shown in Figure 11.

Figure 11a shows the white section;

Figure 11b shows the yellow section in direction of application; and

Figure 11c shows the yellow section against direction of application. All include callouts (i, ii, and iii) corresponding to the FHWA-suggested thresholds for replacement listed in Table 2, where (i = 100 mcd*m⁻²*lx⁻¹ for white and ii, iii = 65 mcd*m⁻²*lx⁻¹ for yellow). In all three cases, the rumble stripe exceeds the FHWA threshold.

Figure 11c is particularly noteworthy—it shows that the glass beads are less retroreflective when observed against the direction of application. Had this been a nondivided roadway, the westbound drivers would...
Figure 11. SR 28 1,500-ft. glass bead test section.

a) White with glass beads in direction of travel and application (June 2011).

b) Yellow with glass beads in direction of travel and application (June 2011).

c) Yellow with glass beads against direction of travel and application (June 2011).
Figure 12. SR 28 1,500-ft. element blend test section.

a) White with element blend in direction of travel and application (June 2011).

b) Yellow with element blend in direction of travel and application (June 2011).

c) Yellow with element blend against direction of travel and application (June 2011).
have observed a less retroreflective line than their eastbound counterparts. Figure 12 shows results from the same tests conducted in the element blend section of SR 28. Here the entire white section (Figure 12a) measured above the FHWA threshold and would likely not require repainting for at least another year. The painted line in the yellow sections (Figure 12b and Figure 12c) hovered around the FHWA-suggested replacement threshold; however, the yellow rumble stripes measured several times the threshold.

Mean retroreflectivity measurements are summarized in Figure 13 for both glass beads and the element blend (designed for enhanced wet nighttime visibility). Figure 13a summarizes white line and rumble stripe retroreflectivity; Figure 13b summarizes yellow line and rumble stripe retroreflectivity. In both cases, the rumble stripe outperformed the standard line.

Table 2. 2000 FHWA-suggested retroreflectivity thresholds for replacement of pavement marking.

<table>
<thead>
<tr>
<th>Pavement Marking</th>
<th>FHWA-Suggested Retroreflective Threshold (mcd<em>m^-2</em>lx^-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nonfreeway ≤ 40 mph</td>
</tr>
<tr>
<td>White</td>
<td>85</td>
</tr>
<tr>
<td>White with lighting or RPM</td>
<td>30</td>
</tr>
<tr>
<td>Yellow</td>
<td>55</td>
</tr>
<tr>
<td>Yellow with lighting or RPM</td>
<td>30</td>
</tr>
</tbody>
</table>

RPM = raised pavement marker.

Results

The results of the research suggest that rumble stripes are effective in increasing the night and wet-night visibility of pavement markings along unlit rural roadways. Rumble stripes also increased the durability of pavement markings. As shown in Figure 5, the rumble protected the retroreflective glass beads and element—over 50 percent of the beads remained in the rumble (Figure 5c). These findings provide quantitative data that support a trend toward using rumble stripes rather than standard painted lines. As shown in Figure 13, under dry conditions the median coefficient of retroreflectivity for a rumble stripe with glass beads surpassed that of the standard painted line by approximately 95 percent for white and 80 percent for yellow. The study also observed that in a corridor with paint containing a blend of elements

Figure 13. Comparison of mean retroreflectivity measurements.
designed for enhanced retroreflectivity, the rumble stripe exceeded the edge line by approximately 90 percent for white and 260 percent for yellow.

A possible mechanism for the improvement in retroreflectivity of the rumble stripe is the upward-sloping painted surface at the back of the rumble, as seen in Figure 14. Should the retroreflectivity of the paint in the rumble be enhanced, even though we expect the top of the rumble to wear in the same manner as the standard line, the large measurement area of a current retroreflectometer (and likely of the human eye) will average the rumble tops and bottoms, leading to increased retroreflectivity along the entire rumble stripe. Although there is some concern that water pooling in these depressions may reduce retroreflectivity, our empirical findings did not substantiate this concern qualitatively (e and f) or quantitatively (Figure 13).

Conclusions and Recommendations

We suspect that snowplow wear is the primary cause of bead loss and retroreflectivity degradation, with tire contact a second order effect. On the basis of qualitative and quantitative data, rumble stripes offer better nighttime visibility than the standard painted line. In addition to the documented effectiveness of the rumble’s auditory and vibratory warning in reducing crashes caused by lane departures, rumble stripes enhance the visibility of the roadway under night and wet-night conditions. According to our findings, the two primary benefits of rumble stripes are:

1. Increased nighttime visibility in dry and wet conditions; and

2. Improved durability of pavement marking with regard to damage by winter snowplowing operations.

Additional studies should be considered to determine the optimal roughness of the milled rumbles and the optimal thickness of paint application, and research should be conducted to identify the conditions that justify the use of additional wet-night elements in the rumble stripe.

Acknowledgments

The authors would like to thank Will Woodfin and Jason Elsen of 3M for supplying the blend of elements designed for enhanced retroreflectivity under wet pavement conditions.

This work was supported by the Joint Transportation Research Program administered by the Indiana Department of Transportation and Purdue University. The article reflects the views of the authors, who are responsible for the facts and the accuracy of the data, and does not necessarily reflect the official views or policies of the sponsoring organizations. These contents do not constitute a standard, specification, or regulation.
References


Stuart R. Mitkey, E.I.T. is a master’s student in the School of Civil Engineering at Purdue University. Mitkey received his bachelor’s degree in biomedical engineering at Michigan Technological University and is currently an officer candidate in the U.S. Navy’s Civil Engineer Corps.

Stephen M. Remias, E.I.T. received his B.S.C.E. from Michigan State University and his M.S.C.E. from Purdue University. Remias is currently a doctoral student in the School of Civil Engineering at Purdue University. Remias has completed several projects in the areas of traffic signal systems, probe-vehicle travel time data, and airport security wait time measurements. He is a member of ITE.

Alexander D. Davis is an undergraduate civil engineering student at Purdue University. Davis worked for the Joint Transportation Research Program during the summer of 2011.

Gannon Grimmer is working as an undergraduate research assistant in the Joint Transportation Research Program while pursuing his B.S.C.E. at Purdue University. Grimmer is a member of the civil engineering honors fraternity Chi Epsilon.

Alexander M. Hainen, E.I.T. received his B.S.C.E. from Michigan Technological University and his M.S.C.E. from Purdue University. Hainen is currently a doctoral student in the School of Civil Engineering at Purdue University. Hainen has completed several projects in the areas of airport passenger studies, measuring techniques for origin-destination studies, and traffic signal systems. He is a member of ITE.

Randy Morris is a 33-year Indiana Department of Transportation (INDOT) employee in Crawfordsville Traffic. Morris is currently a traffic operations manager 2.

Paul Michael is a 34-year INDOT employee in traffic. Michael worked through traffic and is currently a highway maintenance supervisor 3.

Thomas M. Brennan Jr., Ph.D., P.E. is a senior research scientist in the Joint Transportation Research Program at Purdue University. Brennan is a registered professional engineer in Virginia. He is a member of ITE.

Darcy M. Bullock, Ph.D., P.E. is a professor of civil engineering and serves as the director of the Joint Transportation Research Program at Purdue University. Bullock is a registered professional engineer in Louisiana and Indiana. He is a member of ITE.
As part of its strategic highway safety effort, the Federal Highway Administration (FHWA) organized a pooled fund study of 26 states to evaluate low-cost safety strategies. One of the strategies chosen to be evaluated for this study was improving curve delineation. Geometric, traffic, and crash data were obtained at 89 treated curves in Connecticut, USA, and 139 treated curves in Washington, USA, to determine the safety effectiveness of improved curve signing. Treatments varied by site and included new chevrons, horizontal arrows, advance warning signs, and improvement of existing signs using fluorescent yellow sheeting. All sites were on two-lane rural roads. To account for potential selection bias and regression-to-the-mean, an empirical Bayes before-after analysis was conducted.

Introduction

In 1997, the American Association of State Highway and Transportation Officials (AASHTO) Standing Committee for Highway Traffic Safety—with the assistance of FHWA, the National Highway Traffic Safety Administration (NHTSA), and the Transportation Research Board (TRB) Committee on Transportation Safety Management—met with safety experts from various organizations in the field to develop a strategic plan for highway safety. The participants identified 22 key areas that affect highway safety; one of these areas is crashes on horizontal curves.

The National Cooperative Highway Research Program (NCHRP) published a series of implementation guides to advance the implementation of measures targeted to reduce crashes and injuries. Each guide addresses one of the 22 emphasis areas and includes an introduction to the problem, a list
of objectives for improving safety in that area, and strategies for each objective. Each strategy is designated as proven, tried, or experimental. Many of the strategies discussed in these guides have not been rigorously evaluated; about 80 percent of the strategies are considered tried or experimental.

FHWA organized a pooled fund study of 26 states to evaluate low-cost safety strategies as part of this strategic highway safety effort. The purpose of the study was to evaluate the effectiveness of several tried and experimental low-cost safety strategies through scientifically rigorous crash-based studies. Improving delineation on horizontal curves was one of the strategies evaluated as part of this effort.

Statistics from the Fatality Analysis Reporting System (FARS) indicate that 42,642 people were killed in 38,588 fatal crashes on U.S. highways in 2006. Approximately 27 percent of these crashes occurred along horizontal curves.

One strategy to improve horizontal curve safety is enhancing delineation along the curve. According to the NCHRP Report 500 Series Volume 7 on reducing collisions on horizontal curves, enhancing delineation along the curve is a tried but not proven strategy that can be implemented on most curves. Enhanced curve delineation provides visual cues to drivers that they are approaching a curve, as well as positive guidance while they navigate through the curve. Improved delineation can encourage drivers to decrease their speed into and through the curve, and thus reduce the frequency of run-off-road and head-on crashes. Improved delineation is especially helpful under low light or nighttime conditions.

Options for enhanced delineation include using better pavement markings—those of higher durability, all-weather quality and those that have a higher retroreflectivity. Other options include post-mounted delineators, chevrons, raised pavement markers, and wider edge lines. The photo in Figure 1 shows an example of enhanced curve delineation.

Enhanced signing is a potential treatment for curves that have some form of delineation or other safety treatment but continue to experience high crash rates. The installation or upgrade of pavement markings should follow the guidelines in the Manual on Uniform Traffic Control Devices (MUTCD).

**Literature Review**

According to the NCHRP report, although general conclusions can be drawn, no quantitative estimates of the effectiveness of enhancing delineation along curves can be made for conditions in North America. Thus it is not surprising that the 2006 FHWA report *Low-Cost Treatments for Horizontal Curve Safety* provides practical information on low-cost curve treatments but does not contain conclusive quantitative estimates of their safety effectiveness. Other North American studies have focused primarily on safety surrogates rather than crash-based evaluations, although...
one study by Fitzpatrick et al. indicates that post-mounted delineators lower the crash rate at night for relatively sharp curves.\textsuperscript{12}

Recently, a study by Montello evaluated the effectiveness of treatments aimed at improving horizontal curve delineation on 15 curves on the A16 Naples-Canosa motorway in Italy.\textsuperscript{13} All curves were characterized by small radius and high deflection angle, limited sight distance, and limited superelevation. Treatments involved installation of chevron signs, curve warning signs, and sequential flashing beacons along the road. An empirical Bayes (EB) observational before-after study found statistically significant reductions in total, nighttime, daytime, rainy, non-rainy, run-off-road, and property-damage-only crashes. Total crashes decreased by 39.4 percent. The treatment was more effective on curves with radii equal to or less than 300 m and with deflection angles greater than 54 degrees. The results of this study indicate that improving curve delineation can provide significant benefits, although it is not clear whether these results can be translated to conditions in North America. In addition, Montello’s focus was on access-controlled, four-lane divided roads rather than the two-lane rural roads investigated in this study.

The results of this literature review support the NCHRP report statement that this strategy cannot be considered proven because there are no truly valid estimates of its effectiveness based on sound before-after crash-based studies in North America.\textsuperscript{2}

**Study Objectives**

The primary objective was to examine the safety impacts of improved delineation through signing improvements on horizontal curves using the EB method. Target crashes included nonintersection crashes in the following categories: total, injury plus fatality, dark, and lane departure in the dark.
A second objective was to determine whether the effects varied by factors such as traffic volume, curve radius, and type of treatment; for example, installation of new signs versus replacement of old signs. The final objective was to estimate the overall cost-effectiveness of the strategy. To meet these objectives, researchers had to—

- Select a large enough sample size to detect, with statistical significance, what may be small changes in safety for some crash types;
- Identify appropriate reference sites for applying the EB method;
- Properly account for traffic volume changes; and
- Pool data from multiple jurisdictions to improve the reliability of the results and facilitate broader applicability of the products of the research.

**Methodology**

The EB methodology for observational before-after studies\(^{14}\) was used for the evaluation. This methodology is rigorous in that it properly accounts for regression-to-the-mean and has been included in the recently published *Highway Safety Manual*\(^{15}\) as a state-of-the-art method for conducting observational before-after studies. A tutorial on the EB methodology is available in Hauer et al.\(^{16}\)

In the EB approach, the change in frequency for a given crash type at a site is given by:

\[ \Delta \text{Crashes} = \lambda - \pi, \]  

Where \(\lambda\) is the expected number of crashes that would have occurred in the after period without the treatment and \(\pi\) is the number of reported crashes in the after period.

In the EB procedure, the following steps are used to estimate \(\lambda\):

1. Identify a reference group of untreated sites that is otherwise similar to the treatment group.
2. Estimate safety performance functions (SPFs) (mathematical equations) using the reference group, relating crashes of different types of traffic flow and other relevant factors for each jurisdiction.
3. In estimating SPFs, calibrate annual SPF multipliers to account for the temporal effects on safety (for example, variation in weather, demography, and crash reporting).
4. Use the SPFs, the annual SPF multipliers, and data on traffic volume and site characteristics for each year in the before period for each treatment site to estimate the number of crashes that would be expected each year at each site. Sum these annual estimates (called \(P\)).
5. If \(x\) is the count of crashes in the before period at a treatment site, calculate the EB estimate of the expected crashes in the before period at each site (called \(m\)) using the following equations (\(m\) is the expected number of crashes in the before period after correcting for possible bias owing to regression-to-the-mean):

\[ m = w_1(x) + w_2(P), \]  

The weights \(w_1\) and \(w_2\) are estimated from the mean and variance of the SPF estimate as:

\[ w_1 = \frac{Pk}{Pk + 1}, \]
where \( k \) is a constant for a given model and is estimated from the SPF calibration process with the use of a maximum likelihood procedure. In that process, a negative binomial distributed error structure is assumed, with \( k \) being the dispersion parameter of this distribution.

6. Estimate \( \lambda \) as the product of \( m \) and the sum of the annual SPF predictions for the after period divided by \( P \), the sum of these predictions for the before period (for each treatment site). The EB procedure also produces an estimate of the variance of \( \lambda \).

The estimate of \( \lambda \) is then summed over all sites in a strategy group of interest (to obtain \( \lambda_{\text{sum}} \)) and compared with the count of crashes during the after period in that group \( (\pi_{\text{sum}}) \). The variance of \( \lambda \) is also summed over all sites in the strategy group.

The index of effectiveness \( (\theta) \) is estimated as:

\[
\theta = \frac{\frac{\lambda_{\text{sum}}}{\lambda_{\text{sum}}^{2}}}{1 + \left( \frac{\text{Var}(\lambda_{\text{sum}})}{\lambda_{\text{sum}}^{2}} \right)}
\]

The standard deviation of \( \theta \) is given by:

\[
\text{StDev}(\theta) = \sqrt{\frac{\theta^{2} \left( \text{Var}(\pi_{\text{sum}}) + \text{Var}(\lambda_{\text{sum}}) \right)}{\left(1 + \frac{\text{Var}(\lambda_{\text{sum}})}{\lambda_{\text{sum}}^{2}}\right)^{2}}}
\]

The percentage change in crashes is calculated as \( 100(1-\theta) \); thus, a value of \( \theta = 0.7 \) with a standard deviation of 0.12 indicates a 30 percent reduction in crashes with a standard deviation of 12 percent.

Data Collection

Data were collected in Washington and Connecticut. The two states provided installation data, including locations and dates for installation of improved signing on horizontal curves. The researchers obtained information on roadway characteristics, traffic volumes, and crash data from the Washington Highway Safety Information System (HSIS) and the Connecticut Department of Transportation (DOT).

The Connecticut DOT used fluorescent yellow sheeting to improve signing at horizontal curves between 2002 and 2006. This included installing new signs or replacing existing signs. The signs in question were warning signs (for example, CURVE AHEAD or suggested speed limit) and/or curve delineation signs (for example, chevrons or horizontal arrows). Post-mounted delineators were installed on some curves. Figure 2 shows the types of signs used in the treatment. Signs W1–1, –2, –3, –4, –5, and –10 were classified as warning signs; signs W1-6 and W1-8 were classified as curve delineation signs. The unit cost of a fluorescent yellow sign in Connecticut was $30 to $160, depending on size. Smaller signs (such as chevrons and advisory speed signs) were cheaper than larger signs (such as curve warning signs). In Washington, the treatments involved only the installation of chevrons (W1-8 signs).
The Connecticut treatment sites were selected through a program called the Suggested List of Surveillance Study Sites (SLOSSS). SLOSSS uses crash data, traffic volume, and roadway characteristics to identify intersections and road segments with higher than expected crash rates.

Table 1 summarizes the data collected for the treatment sites. Average annual daily traffic (AADT) is about 30 percent higher in Connecticut than in Washington, so it is not surprising that Connecticut has more crashes per mile per year than Washington. However, a higher percentage of crashes in Washington result in injury.

The total available sample of 117.3 mile-years in the before period and 116.69 mile-years in the after period are higher than the minimum sample size requirements estimated during the study design process, which is documented in the study report.17

To account for biases in treatment site selection, the study analysis required the identification of reference sites; that is, curves that were similar to the treatment sites but that did not receive the improved signing treatment. In Connecticut, video logs were used to identify reference sites by first identifying curves within 1.5 miles of the treatment curves. These curves were then reviewed to include only those that were similar to the treatment curves in terms of AADT and curve radius. The final group of reference sites included 334 horizontal curves. Similarly, in Washington, HSIS data were used to identify reference curves on rural two-lane roads that were similar to the treatment curves. The final reference group included 1,314 horizontal curves.
Table 1. Data summary.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Connecticut (89 curves, 7.08 miles)</th>
<th>Washington (139 curves, 14.06 miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mile-years</td>
<td>Before: 45.80, After: 21.60</td>
<td>Before: 71.50, After: 95.09</td>
</tr>
<tr>
<td>AADT (minimum)</td>
<td>895, 920</td>
<td>261, 354</td>
</tr>
<tr>
<td>AADT (average)</td>
<td>4,381, 4,741</td>
<td>3,305, 3,700</td>
</tr>
<tr>
<td>AADT (maximum)</td>
<td>19,945, 20,479</td>
<td>11,567, 14,790</td>
</tr>
<tr>
<td>Nonintersection crashes per mile per year (average)</td>
<td>9.01, 7.19</td>
<td>4.39, 3.80</td>
</tr>
<tr>
<td>Nonintersection lane departure crashes per mile per year</td>
<td>7.66, 6.08</td>
<td>3.57, 3.07</td>
</tr>
<tr>
<td></td>
<td>(average)</td>
<td></td>
</tr>
<tr>
<td>Nonintersection crashes with injury or fatality (K,A,B,C) per</td>
<td>3.43, 1.95</td>
<td>2.45, 1.88</td>
</tr>
<tr>
<td></td>
<td>mile per year (average)</td>
<td></td>
</tr>
<tr>
<td>Nonintersection crashes in the dark per mile per year (average)</td>
<td>3.99, 2.18</td>
<td>1.68, 1.36</td>
</tr>
<tr>
<td>Nonintersection lane departure crashes in the dark per mile</td>
<td>3.54, 1.86</td>
<td>1.47, 1.22</td>
</tr>
<tr>
<td></td>
<td>per year (average)</td>
<td></td>
</tr>
</tbody>
</table>

AADT = average annual daily traffic.

Development of Safety Performance Functions

As noted earlier, the EB methodology uses safety performance functions (SPFs) to estimate the effectiveness of this strategy. The researchers used generalized linear modeling to estimate model coefficients with the SAS software package; they assumed a negative binomial error distribution, which is consistent with the state of research in developing these models. The overdispersion parameter (k) was also estimated by SAS in the model calibration process. The overdispersion parameter relates the mean and variance of the SPF estimate—the smaller its value, the better a model is for a given set of data. It is a useful criterion for comparing candidate models.

The form of the SPFs is:

\[ \text{Crashes/year} = \alpha \exp\left( \sum_{i=1}^{n} \beta_i X_i \right) \]  \hspace{1cm} (7)

where the Xs are the independent variables, \( \beta \)s are estimated coefficients, and \( \alpha \) is the estimated constant term. The safety performance functions developed for Connecticut for five of the crash types (total, lane departure, injury and fatality, crashes in the dark, and lane departure crashes in the dark) are presented in Table 2, along with some goodness of fit statistics.

In estimating the SPFs, the researchers used a backward elimination procedure to eliminate variables one at a time that were not statistically significant. To account for the effect of changes in factors such as weather, crash reporting practices, and demography over time, annual factors were estimated for each year. The annual factor for a particular year is defined as the ratio of observed to predicted crashes for that year.
### Table 2. Safety performance functions for Connecticut reference sites.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Non-intersection crashes</th>
<th>Non-intersection lane departure crashes</th>
<th>Non-intersection crashes with injury or fatality</th>
<th>Non-intersection crashes in the dark</th>
<th>Non-intersection lane departure crashes in the dark</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>α (constant)</strong></td>
<td>17.2515 (S.E. = 10.6302)</td>
<td>4.5889 (S.E. = 0.8531)</td>
<td>11.2842 (S.E. = 0.9412)</td>
<td>4.4057 (S.E. = 0.9994)</td>
<td></td>
</tr>
<tr>
<td><strong>Ln (section length in miles)</strong></td>
<td>0.8161 (S.E. = 0.06768)</td>
<td>0.9898 (S.E. = 0.09439)</td>
<td>1.294 (S.E. = 0.09779)</td>
<td>0.9542 (S.E. = 0.09782)</td>
<td>0.8720 (S.E. = 0.1161)</td>
</tr>
<tr>
<td><strong>Ln (AADT 10,000)</strong></td>
<td>1.0047 (S.E. = 0.05998)</td>
<td>0.7003 (S.E. = 0.07719)</td>
<td>1.294 (S.E. = 0.09779)</td>
<td>0.9542 (S.E. = 0.09782)</td>
<td>0.8720 (S.E. = 0.1161)</td>
</tr>
<tr>
<td><strong>Radius (in meters)</strong></td>
<td>-0.00147 (S.E. = 0.00018)</td>
<td>-0.00206 (S.E. = 0.00024)</td>
<td>-0.00141 (S.E. = 0.00028)</td>
<td>-0.00125 (S.E. = 0.00027)</td>
<td>-0.00172 (S.E. = 0.00034)</td>
</tr>
<tr>
<td><strong>1 if roadside hazard rating (RHR) is 1–4, 0 otherwise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.9917 (S.E. = 0.5247)</td>
</tr>
<tr>
<td><strong>1 if RHR is 5, 0 otherwise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.7919 (S.E. = 0.5182)</td>
</tr>
<tr>
<td><strong>1 if RHR is 1–5, 0 otherwise</strong></td>
<td>-0.7523 (S.E. = 0.3313)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1 if RHR is 6, 0 otherwise</strong></td>
<td>-0.7386 (S.E. = 0.3478)</td>
<td>-0.6146 (S.E. = 0.5423)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RHR = 7 (reference case)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of driveways in the arc portion of the section</strong></td>
<td>0.04011 (S.E. = 0.01946)</td>
<td>0.06047 (S.E. = 0.03102)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1 if there is a cable barrier in the inside of the curve, 0 otherwise</strong></td>
<td>0.2219 (S.E. = 0.09202)</td>
<td>0.3343 (S.E. = 0.1136)</td>
<td></td>
<td></td>
<td>0.2741 (S.E. = 0.1642)</td>
</tr>
<tr>
<td><strong>1 if there is a cable barrier on the outside of the curve, 0 otherwise</strong></td>
<td></td>
<td></td>
<td></td>
<td>-0.3005 (S.E. = 0.1649)</td>
<td></td>
</tr>
<tr>
<td><strong>1, if there is a guardrail in the inside of the curve, 0 otherwise</strong></td>
<td>0.2077 (S.E. = 0.04093)</td>
<td>0.2446 (S.E. = 0.05538)</td>
<td>0.2625 (S.E. = 0.05517)</td>
<td>0.2164 (S.E. = 0.05444)</td>
<td>0.1827 (S.E. = 0.06614)</td>
</tr>
<tr>
<td><strong>1, if there is a guardrail on the outside of the curve, 0 otherwise</strong></td>
<td></td>
<td></td>
<td></td>
<td>-0.5141 (S.E. = 0.1511)</td>
<td></td>
</tr>
<tr>
<td><strong>Average distance to next curve (km)</strong></td>
<td>-0.8101 (S.E. = 0.4666)</td>
<td>-1.2513 (S.E. = 0.5683)</td>
<td>-0.7954 (S.E. = 0.5548)</td>
<td>-1.4775 (S.E. = 0.6586)</td>
<td></td>
</tr>
<tr>
<td><strong>1 if there is a curb and gutter in the shoulder, 0 otherwise</strong></td>
<td>0.4138 (S.E. = 0.1187)</td>
<td></td>
<td></td>
<td>0.3532 (S.E. = 0.1826)</td>
<td></td>
</tr>
<tr>
<td><strong>Grade in the middle of the arc portion of the section</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.04583 (S.E. = 0.02629)</td>
</tr>
<tr>
<td><strong>k (over-dispersion parameter)</strong></td>
<td>0.8535</td>
<td>1.3859</td>
<td>0.9321</td>
<td>1.2365</td>
<td>1.6375</td>
</tr>
<tr>
<td><strong>Observations in the reference group</strong></td>
<td>3290</td>
<td>3020</td>
<td>3020</td>
<td>3020</td>
<td>3020</td>
</tr>
<tr>
<td><strong>Crashes in the reference group</strong></td>
<td>1400</td>
<td>843</td>
<td>438</td>
<td>487</td>
<td>350</td>
</tr>
<tr>
<td><strong>Freeman Tukey R-square</strong></td>
<td>0.163</td>
<td>0.079</td>
<td>0.130</td>
<td>0.069</td>
<td>0.047</td>
</tr>
</tbody>
</table>

**NOTE:** Roadside hazard rating (RHR) varies from 1 through 7, with a higher number indicating a more hazardous roadside. RHS is based on the work of Zegeer et al.19
In Washington, the researchers estimated an SPF for total crashes. The SPFs for the other four crash types were assumed to be equal to the proportion of that crash type multiplied by the SPF for total crashes. A recalibrated overdispersion parameter was estimated for each crash type. Table 3 shows the parameter estimates and standard errors, along with the corresponding constant term and overdispersion parameter for each crash type.

Table 3. Safety performance functions for Washington reference sites.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate (S.E.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln (section length in miles)</td>
<td>1.1539 (0.0465)</td>
</tr>
<tr>
<td>Ln (AADT 10,000)</td>
<td>1.1995 (0.0776)</td>
</tr>
<tr>
<td>AADT 10,000</td>
<td>-0.7667 (0.1983)</td>
</tr>
<tr>
<td>Ln (1/radius in miles)</td>
<td>0.6106 (0.0480)</td>
</tr>
<tr>
<td>Shoulder width (feet)</td>
<td>-0.0565 (0.0131)</td>
</tr>
<tr>
<td>Lane width (feet)</td>
<td>-0.0701 (0.0218)</td>
</tr>
<tr>
<td>Percentage grade (%)</td>
<td>0.0249 (0.0124)</td>
</tr>
<tr>
<td>Terrain is mountainous (coded 1 for mountainous, 0 otherwise)</td>
<td>0.2477 (0.0774)</td>
</tr>
</tbody>
</table>

| Terrain is level or rolling (reference case)       | —               |
| Observations in the reference group                | 17078           |
| Total crashes in the reference group               | 1971            |
| Freeman Tukey R-square (for total crash SPF)      | 0.099           |

<table>
<thead>
<tr>
<th>α (constant) and k (overdispersion parameter) for different crash types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total crashes</td>
</tr>
<tr>
<td>Lane departure crashes</td>
</tr>
<tr>
<td>Injury and fatality crashes</td>
</tr>
<tr>
<td>Crashes in the dark</td>
</tr>
<tr>
<td>Lane departure crashes in the dark</td>
</tr>
</tbody>
</table>

Results

An aggregate analysis provided evidence for the general effectiveness of the strategy, while a disaggregate analysis provided insight on the situations in which the strategy may be most effective.

Aggregate Analysis

Table 4 shows the aggregate results for Connecticut, Washington, and the two states combined. The first measure of safety effect is the estimated percentage reduction due to the strategy, along with the standard error (S.E.) of this estimate; a negative value would indicate an increase in crashes. Bold text denotes the safety effects that are significant at the 95 percent confidence level. The second measure of safety effect is the change in the number of crashes per mile-year, which is calculated as the difference between the EB estimate of crashes expected in the after period and the number of observed crashes in the after period divided by the number of mile-years in the period.
Table 4. Aggregate results of curve signing improvements.

<table>
<thead>
<tr>
<th>State</th>
<th>Non-intersection</th>
<th>Non-intersection In the Dark</th>
<th>Non-intersection Lane Departure</th>
<th>Non-intersection In the Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td>EB estimate of after period crashes without treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>188.1</td>
<td>55.9</td>
<td>72.2</td>
<td>60.4</td>
</tr>
<tr>
<td>Washington</td>
<td>374.8</td>
<td>211.8</td>
<td>169.5</td>
<td>147.7</td>
</tr>
<tr>
<td>Two states combined</td>
<td>562.9</td>
<td>267.7</td>
<td>241.7</td>
<td>208.1</td>
</tr>
<tr>
<td>Number of crashes observed in the after period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>155</td>
<td>42</td>
<td>47</td>
<td>40</td>
</tr>
<tr>
<td>Washington</td>
<td>361</td>
<td>179</td>
<td>129</td>
<td>116</td>
</tr>
<tr>
<td>Two states combined</td>
<td>516</td>
<td>221</td>
<td>176</td>
<td>156</td>
</tr>
<tr>
<td>Estimate of percentage reduction (S.E.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>17.8% (7.7)</td>
<td>25.2% (12.7)</td>
<td>35.3% (10.5)</td>
<td>34.2% (11.5)</td>
</tr>
<tr>
<td>Washington</td>
<td>4.3% (8.9)</td>
<td>16.4% (10.4)</td>
<td>24.5% (9.5)</td>
<td>22.1% (10.1)</td>
</tr>
<tr>
<td>Two states combined</td>
<td>8.6% (6.4)</td>
<td>18.0% (8.6)</td>
<td>27.5% (7.3)</td>
<td>25.4% (7.8)</td>
</tr>
<tr>
<td>Estimate of reduction in crashes per mile-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>1.54</td>
<td>0.64</td>
<td>1.17</td>
<td>0.95</td>
</tr>
<tr>
<td>Washington</td>
<td>0.15</td>
<td>0.35</td>
<td>0.43</td>
<td>0.33</td>
</tr>
<tr>
<td>Two states combined</td>
<td>0.40</td>
<td>0.40</td>
<td>0.56</td>
<td>0.45</td>
</tr>
</tbody>
</table>

NOTE: Bold type denotes results that are statistically significant at the 95 percent confidence level.

In Connecticut, all the evaluated crash types experienced a statistically significant reduction (at the 95 percent confidence level). Total nonintersection and nonintersection lane departure crashes were reduced by approximately 18 percent. Nonintersection crashes with injuries or fatalities were reduced by about 25 percent. The percentage reduction was larger (35 percent) for the crashes that occurred after dark.

In Washington, crashes after dark and lane departure crashes after dark experienced a statistically significant reduction—the reduction for both crash types exceeded 20 percent. The combined results for the two states indicated a statistically significant reduction in crashes involving an injury or fatality, crashes after dark, and lane departure crashes after dark.
**Disaggregate Analysis**

The disaggregate analysis focused on average AADT at a site before treatment, radius of the curve where the treatment was implemented, roadside hazard rating (RHR) at the treatment site, number of signs in advance of the curve that were added or replaced, and number of within-curve signs that were added or replaced. AADT and curve radius were available from both Washington and Connecticut, but only Connecticut provided information on RHR and number of signs.

When curve radius was examined using combined data from Washington and Connecticut, no clear trends emerged regarding the effectiveness of the treatment. When curve radius was examined separately in the two states, no clear trends regarding effectiveness emerged in Washington. However, in Connecticut, the treatments seemed to be more effective in sharper curves (radius less than 150 m or 0.093 miles) compared with flatter curves (radius equal to or exceeding 150 m or 0.093 miles).

When roadside hazard rating was examined using data from Connecticut, the treatments seemed to be more effective at sites with more hazardous roadsides (RHR rating of 5 or 6) than at those with less hazardous roadsides (RHR 2–4) for crashes after dark and lane departure crashes after dark.

When the effect of the number of signs added or replaced in advance of the curve was examined using Connecticut data, no clear safety trends emerged. However, when the number of within-curve signs added or replaced was examined, sites with more than seven signs added or replaced seemed to have experienced a larger reduction in crashes than sites where fewer signs were added or replaced.

Finally, the treatment seemed more effective at sites with a higher AADT, but these results are not shown because they are not definitive. The disaggregate results are presented in Table 5 for the Connecticut sites only, as there were no clear indications from the Washington results.
Table 5. Results of the disaggregate analysis of improved curve signing for Connecticut sites.

<table>
<thead>
<tr>
<th>Crash Type</th>
<th>Disaggregate group</th>
<th>EB estimate of crashes expected in the after period without strategy</th>
<th>Number of crashes observed in the after period</th>
<th>Estimate of percentage reduction (Standard error)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Radius category (meters)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane departure crashes</td>
<td>&lt; 150 m</td>
<td>92.5</td>
<td>70</td>
<td>24.7% (10.4)</td>
</tr>
<tr>
<td></td>
<td>&gt; 150 m</td>
<td>66.3</td>
<td>61</td>
<td>8.6% (13.7)</td>
</tr>
<tr>
<td>Crashes after dark</td>
<td>&lt; 150 m</td>
<td>44.6</td>
<td>26</td>
<td>42.3% (12.5)</td>
</tr>
<tr>
<td></td>
<td>&gt; 150 m</td>
<td>27.6</td>
<td>21</td>
<td>24.8% (18.1)</td>
</tr>
<tr>
<td>Lane departure crashes after dark</td>
<td>&lt; 150 m</td>
<td>38.6</td>
<td>22</td>
<td>43.6% (13.2)</td>
</tr>
<tr>
<td></td>
<td>&gt; 150 m</td>
<td>21.8</td>
<td>18</td>
<td>18.4% (21.2)</td>
</tr>
<tr>
<td><strong>Roadside hazard rating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane departure crashes</td>
<td>2, 3, 4</td>
<td>74.6</td>
<td>64</td>
<td>14.7% (12.3)</td>
</tr>
<tr>
<td></td>
<td>5, 6</td>
<td>84.2</td>
<td>67</td>
<td>20.9% (11.3)</td>
</tr>
<tr>
<td>Crashes after dark</td>
<td>2, 3, 4</td>
<td>29.5</td>
<td>27</td>
<td>9.5% (19.4)</td>
</tr>
<tr>
<td></td>
<td>5, 6</td>
<td>42.7</td>
<td>20</td>
<td>53.6% (11.2)</td>
</tr>
<tr>
<td>Lane departure crashes after dark</td>
<td>2, 3, 4</td>
<td>26.7</td>
<td>22</td>
<td>18.6% (19.2)</td>
</tr>
<tr>
<td></td>
<td>5, 6</td>
<td>33.7</td>
<td>18</td>
<td>47.2% (13.6)</td>
</tr>
<tr>
<td><strong>Within-curve signs added or replaced</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane departure crashes</td>
<td>&lt; 7 per curve</td>
<td>79.6</td>
<td>77</td>
<td>3.9% (13.0)</td>
</tr>
<tr>
<td></td>
<td>7 to 28 per curve</td>
<td>79.2</td>
<td>54</td>
<td>32.2% (10.5)</td>
</tr>
<tr>
<td>Crashes after dark</td>
<td>&lt; 7 per curve</td>
<td>35.4</td>
<td>27</td>
<td>24.5% (16.1)</td>
</tr>
<tr>
<td></td>
<td>7 to 28 per curve</td>
<td>36.8</td>
<td>20</td>
<td>46.3% (13.1)</td>
</tr>
<tr>
<td>Lane departure crashes after dark</td>
<td>&lt; 7 per curve</td>
<td>29.2</td>
<td>24</td>
<td>18.6% (18.5)</td>
</tr>
<tr>
<td></td>
<td>7 to 28 per curve</td>
<td>31.2</td>
<td>16</td>
<td>49.4% (13.7)</td>
</tr>
</tbody>
</table>

Note: Bold denotes results that are statistically significant at the 95 percent confidence level.
Economic Analysis

A combined economic analysis is provided, as well as separate analyses for the two states, as each had its own variation on the treatment. First, we computed the annualized cost of the treatment on the basis of information provided by Connecticut and Washington. Connecticut provided costs of the type of treatment, ranging from $30 to $160 for a fluorescent yellow sign with a service life of 5 years. The unit cost varies according to the size of the sign—chevrons and advisory speed signs are less expensive than larger signs such as curve warning signs. Washington provided an estimate of $100 for chevrons. We used the lower and upper limits from Connecticut to establish a range of cost estimates, but costs will vary in different states according to installation and maintenance practices and the size of the sign.

The formula to calculate the annual cost is:

$$\text{Annual Cost} = \frac{C \times R}{1 - (1 + R)^{-N}}$$

(8)

Where:

- $C$ = installation cost.
- $R$ = discount rate (as a decimal).
- $N$ = expected service life (years).

On the basis of information from the Office of Management and Budget,20 we used a discount rate of 2.4 percent to determine the annual cost of the strategy. Assuming installation costs of $30 per sign, the annualized cost of a sign is $6.44. Assuming 10 chevron signs installed per curve, the annual treatment cost per curve is $64. Assuming installation costs of $160 per sign, the annualized cost is $34.34 per sign and the annual treatment cost per curve is $343. Thus, the annual cost savings range from $128 to $686 per curve, which is a 2:1 benefit-cost ratio. This is the minimum benefit-cost ratio used by FHWA to recommend treatments for implementation.

We used the most recent FHWA mean comprehensive crash costs21 to estimate the cost of a lane departure crash (in 2001 dollars). The FHWA crash cost document does not directly report the cost of a lane departure crash, so we estimated it on the basis of the cost of lane-departure-related crashes and the percentage of these crashes. Lane-departure-related crashes include head-on, sideswipe, single-vehicle rollover, and single-vehicle fixed-object crashes. On the basis of Washington crash data, these crashes represent 2.6 percent, 1.6 percent, 25.9 percent, and 69.9 percent, respectively. The mean comprehensive crash costs for these crash types are $60,451, $16,019, $147,629, and $67,353, respectively. We computed a weighted average by combining the crash costs with the percentage of each crash type, resulting in a cost of $87,143 per lane departure crash.

Crash savings were computed using the results for nonintersection lane departure crashes after dark. The total crash reduction was calculated for each state by subtracting the actual crashes in the after period from the expected crashes in the after period if the treatment had not been implemented. Crashes per site-year were calculated by dividing total crash reduction by the number of site-years for each state. The benefit (that is, crash savings) is the product of the total crash reduction per site-year and the cost of a lane departure crash (that is, $87,143). The benefit-cost ratio indicates a range for which the lower and upper limits represent assumed installation costs of $160 and $30 per sign, respectively.

Table 6 summarizes the results of the economic analysis. Even using conservative assumptions only a very modest reduction in crashes is required to justify this strategy economically. The benefit-cost ratios shown in Table 6 far exceed a 2:1 ratio. Although the cost of the strategy will probably vary from state to state, it is not likely that the annualized costs will exceed annual crash savings.
Table 6. Summary of economic analysis results.

<table>
<thead>
<tr>
<th>State</th>
<th>Sites</th>
<th>Years</th>
<th>Total Crash Reduction (per year)</th>
<th>Total Crash Reduction (per site-year)</th>
<th>Crash Savings (per site-year)</th>
<th>Expected Range of Benefit-Cost Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>228</td>
<td>4.58</td>
<td>11.373</td>
<td>0.050</td>
<td>$4,347</td>
<td>12.7:1 to 67.9:1</td>
</tr>
<tr>
<td>Connecticut</td>
<td>89</td>
<td>3.05</td>
<td>6.686</td>
<td>0.075</td>
<td>$6,546</td>
<td>19.1:1 to 102.3:1</td>
</tr>
<tr>
<td>Washington</td>
<td>139</td>
<td>6.76</td>
<td>4.687</td>
<td>0.034</td>
<td>$2,938</td>
<td>8.6:1 to 45.9:1</td>
</tr>
</tbody>
</table>

Summary and Conclusions

The objective of this study was to evaluate the effectiveness in terms of safety of improved delineation of horizontal curves through signing enhancements. The results of an aggregate analysis using data from Connecticut and Washington indicate a statistically significant reduction in injury and fatal crashes (18 percent), crashes after dark (27.5 percent reduction), and lane departure crashes after dark (25.4 percent reduction). Limited data from Connecticut suggest that the treatment is more likely to reduce the number of crashes on sharper curves (curve radius less than 150 m) and in locations with more hazardous roadsides (RHR of 5 or higher). In addition, curves at which more within-curve signs were added or replaced (with a more retroreflective material) experienced larger reductions in crashes. In both states, the reductions were greater at locations with higher traffic volumes. An economic analysis revealed that improving curve delineation with signing enhancements is a very cost-effective treatment, with a benefit-cost ratio that exceeds 8:1.

Clearly, improving curve delineation is effective in reducing the number of crashes, especially after dark. States and local agencies can use these crash reduction estimates to quantify the expected benefit associated with improving curve delineation under various conditions. Future research could investigate the safety effects of adding or replacing specific warning and curve delineation signs, and how their effectiveness might depend on factors such as terrain, speed limit, and advisory speed.

Acknowledgments

This article is based on a study conducted for the FHWA Office of Safety under Contract DTFH61-05-D-00024. The FHWA COTM for this project was Roya Amjadi. Kimberly Eccles, P.E., of Vanasse Hangen Brustlin Inc. (VHB) was the principal investigator. Charles Hamlett from the University of North Carolina Highway Safety Research Center, Nancy Lefler of VHB, and Yusuf Mohamedshah of Lendis Corporation assisted with data collection, while Craig Lyon of Persaud & Lyon Inc. provided advice on the development of the SPF. Frank Gross of VHB assisted with the economic analysis. The project team gratefully acknowledges the following organizations and individuals for their assistance in this study: the Connecticut Department of Transportation, especially John F. Carey, Al Iallonardo, and Brad Overturf; the University of Connecticut, especially Hongmei Zhou and John Ivan; and the Washington State Department of Transportation, especially Rick Mowlds, Dan Davis, Mark Finch, and John Milton.
References


Raghavan Srinivasan, Ph.D., is a senior transportation research engineer at the UNC Highway Safety Research Center. He has more than fifteen years of experience conducting research in traffic safety and human factors for the Federal Highway Administration, National Cooperative Highway Research Program, National Highway Traffic Safety Administration, and many state agencies in the United States. He is a member of ITE.

Bhagwant Persaud, Ph.D., is a professor of civil engineering at Ryerson University. He is currently co-chair of the Transportation Research Board Committee ANB20: Safety Data, Analysis, and Evaluation.

Kimberly Eccles, P.E., is a principal with Vanasse Hangen Brustlin, Inc. (VHB). She has fifteen years of experience in transportation safety, specializing in safety evaluations, traffic control devices, and crash data analysis. She is the principal investigator for the Federal Highway Administration Evaluation of Low Cost Safety Improvements Pooled Fund Study. Ms. Eccles manages the VHB office in Raleigh, North Carolina. She is a member of ITE.

Daniel Carter serves as a senior engineering research associate with the UNC Highway Safety Research Center where he handles data collection and analysis for safety research projects. He has a B.S. and an M.S. in civil engineering (Transportation) from North Carolina State University. His research areas include safety treatment evaluation, crash modification factors, large truck safety, and pedestrian safety.

Dr. Jongdae Baek has research experience in the broad area of transportation engineering. Recent research projects have included work in building crash modification factors and conducting safety evaluations. Dr. Baek is currently involved with several projects in the area of environmentally-friendly roads.